Artificial Intelligence in the Office and the Factory:

Evidence from Administrative Software Registry Data

Gustavo de Souza

July 21, 2025

WP 2025-11

https://doi.org/10.21033/wp-2025-11

FEDERAL RESERVE BANK of CHICAGO

Artificial Intelligence in the Office and the Factory: Evidence from Administrative Software Registry Data*

Gustavo de Souza

Federal Reserve Bank of Chicago

July 21, 2025

Abstract

I use administrative data on artificial intelligence (AI) software created in Brazil to study its effects on the labor market. Owing to a unique copyright system, Brazilian firms have registered their software with the government since the 1980s, creating a detailed record of nearly all commercial AI applications developed in the country. Drawing on this registry, I show that AI is widely used not only in administrative tasks but also in production settings, where it primarily supports process optimization and quality control. Using an instrument based on variation in software development costs, I find that AI affects administrative and production workers differently. Among office workers, AI reduces employment and wages, particularly for middle-wage earners. Among production workers, it increases employment of low-skilled and young workers operating machinery. These results suggest that AI displaces routine office tasks while making machines more productive and easier to operate, leading to a net increase in employment.

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

JEL: J23, J24, F63

* This paper represents my opinions and not those of the Federal Reserve Bank of Chicago or the Federal Reserve System. This paper benefited from valuable comments from Anders Humlum, Xavier Jaravel, Aakash Kalyani, Anna Salomons, and Yongseok Shin. Jack Mannion and Vitor Santos provided superb research assistance.

1 Introduction

Artificial intelligence (AI) now provokes intense public and academic debate. Some warn that AI will automate extensive areas of work, displacing millions of employees and increasing inequality (Acemoglu and Restrepo 2019, Humlum and Vestergaard 2025b). Others contend that AI can enhance rather than replace human labor by simplifying complex machines and easing decision-making Autor (2024), Agrawal et al. (2019). Engineers and operations researchers increasingly report AI applications in Manufacturing Execution Systems, predictive maintenance, and computer vision for quality control (McElheran et al. 2025, Alam et al. 2024). However, despite growing interest, empirical evidence on the use and impact of AI remains limited, largely due to a lack of detailed data on AI technologies.

To address this gap, I study the impact of AI on the labor market using a new administrative registry covering nearly all commercial AI software developed in Brazil. To isolate the causal effect, I use an instrumental variable strategy that leverages exogenous variation in the cost of developing AI software. I find that AI increases overall employment but reduces average wages: a one-standard-deviation increase in AI exposure raises total employment by about 1.9 percent and lowers average wages by roughly 2 percent. This decline in average wages is driven by a shift in the composition of the workforce: AI leads to greater hiring of young, inexperienced, and low-education workers, particularly in production occupations. Among administrative workers, AI decreases employment, especially among low-wage workers. These findings help reconcile competing perspectives on the effect of AI. It is consistent with the idea that, in factories, AI increases the productivity of machines, simplifies machine operation, and raises demand for low-skilled labor, as suggested by Agrawal et al. (2019) and Gruber et al. (2020). In contrast, in office settings, AI automates tasks and reduces employment (Acemoglu and Restrepo 2018, Webb 2020, Aum and Shin 2025, Hui et al. 2023).

To see these effects in action, consider Techplus, a Brazilian firm that develops AI systems for industrial operations. Its software analyzes real-time sensor data from factory equipment to predict failures, schedule maintenance, and assist machine operation. These tools 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. At the same time, the software automates routine administrative tasks such as maintenance planning and inventory control. This case reflects the broader pattern I document in the data: AI increases employment in low-skilled production roles by making machines easier to operate and more productive, while reducing demand for routine office work through automation.

To study the effect of AI on the labor market, I begin by constructing a novel dataset on software registration records in Brazil. In 1987, Brazil introduced copyright protection to the source code of computer programs registered with the National Institute of Industrial Property (NIIP). Subsequently, registering software became standard industry practice, with approximately 96% of software firms registering at least one program. I collect data on all software programs that uses AI technology ever registered with the NIIP, which includes detailed information on ownership, programmers, intended use of the software, and technical features.[1] The NIIP registry provides a more precise proxy for AI development than patents or job advertisements, the most commonly used metrics in prior research, because it includes detailed technical and application information on actual software products, including third-party tools.[2] The dataset spans from 1987 to 2022, offering a rare long-run view of AI development.

The NIIP registry reveals three stylized facts about AI software development in Brazil. First, AI creation surged after 2013, tracking global advances in deep learning and GPU computing. The annual filings increased from 90 in 2014 to 420 in 2022. Second, 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. Third, production-oriented AI is concentrated in a few practical applications: 40 percent are Manufacturing Execution Systems (MES), 21 percent handle automated quality inspection, and 15 percent are industrial robots interface. These patterns challenge the idea that AI affects only white-collar work and point instead to

[1] When registering, firms have to assign their software to classes. There are two types of software classes. One that relates to the intended use of the software and another with the technology used in the software. I study all software with a technical class associated with AI.

[2] Only 10 percent of firms with a software registration have a patent application, illustrating the broader coverage of the NIIP software registry.

substantial labor-market implications for machine operators and other shop-floor workers.[3]

To estimate how AI affects different occupations, I compute the text similarity between occupational tasks and the descriptions of AI software in the NIIP database (Webb 2020, de Souza and Li 2023, Kogan et al. 2023, Hampole et al. 2025). The intuition is straightforward: occupations with tasks that resemble those performed by AI software are more exposed to AI adoption. To estimate the causal effect of AI, I use an instrumental variable strategy that exploits heterogeneous exposure to a decrease in the cost of AI development.[4]

I find that, on average, exposure to AI increases employment but reduces wages. A one-standard-deviation increase in AI exposure raises employment by about 1.9 percent and lowers average wages by roughly 2 percent. This wage decline reflects a shift in workforce composition in observable and non-observable characteristics: AI adoption leads to increased hiring of younger, less-educated, less-experienced, lower-ability workers who typically earn less. After accounting for worker composition, however, a one-standard-deviation increase in AI exposure increases wages by 0.5%.

AI reduces wage inequality within occupations by compressing wages at the top of the distribution. A one-standard-deviation increase in AI exposure lowers within-occupation wage inequality by 3%. This effect is driven by larger wage declines at the upper among high earners: AI exposure has no effect in the first decile of wages but decreases the 9th decile by 3%. This pattern is consistent with experimental evidence showing that AI tools boost productivity more for lower-skill workers than for experts (Brynjolfsson et al. 2025, Peng et al. 2023, Kanazawa et al. 2022).

AI affects office and production workers differently. Among production workers, it raises

[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 (2019) and Borboni et al. (2023) discuss the use of AI in the operation of industrial robots.

[4] One of the key advantages of the NIIP registry is that it includes software developed by specialized IT companies. To fully exploit this feature, I conduct the analysis at the occupation level rather than the firm level. A firm-level approach would restrict the sample to a narrow set of firms that develop AI in-house, missing the adoption of AI through third-party vendors. In contrast, an occupation-level analysis captures both direct and indirect adoption, and it allows me to identify broader general equilibrium effects—such as labor reallocation across firms and occupations—that a firm-level analysis would miss.

employment but lowers average wages by shifting hiring toward younger, less-educated, and less-experienced workers. Within this group, AI raises wages for the bottom decile by 1% but reduces wages at the top decile by 4%. These effects are more pronounced among workers that operate machinery directly. A possible explanation for these findings is a reduced skill-premium: as AI makes complex equipment easier to operate, the value of expertise declines, allowing firms to substitute toward low-skill workers (Autor 2024).

Administrative occupations show a different pattern. For these workers, AI reduces employment without significantly affecting average wages. However, this stability in averages masks a hollowing out of the wage distribution: AI decreases wages among administrative workers in the middle of the income distribution. These findings are consistent with the idea that AI automates tasks done by white collar workers(Acemoglu and Restrepo 2018, Webb 2020, Aum and Shin 2025, Hui et al. 2023).

AI also reduces inequality across occupations. In occupations with higher initial wages and education levels, AI exposure lowers both average earnings and years of education. That is again due to an influx of younger, less-educated, less-experienced, lower-ability workers. These results are consistent with the idea that AI lowers skill barriers and opens access to roles once reserved for highly qualified individuals. As a result, AI reduces inequality both within and across occupations.

This paper contributes to the literature on the labor market effects of AI by introducing a new dataset and an instrumental variable strategy that identify 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. 2019, Gruber et al. 2020). 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, finds that the productivity gains are larger than the displacement effect at firms adopting AI. Others, including Webb (2020), Aum and Shin (2025), and Hui et al. (2023), show that AI adoption can reduce earnings and job opportunities. In contrast, firm-level studies like Babina et al. (2024) and Adams et al. (2024) find positive effects on firm growth and productivity, pointing to possible employment gains.

This paper helps reconcile the conflicting perspectives by showing that the labor market effects of AI vary across occupations. As Webb (2020), Aum and Shin (2025), and Hui et al. (2023) have argued, I find that AI reduces employment and wages among administrative workers, consistent with the displacement effects emphasized by (Acemoglu and Restrepo 2018). At the same time, I find that AI increases employment in production occupations, supporting the view of Agrawal et al. (2019) and Gruber et al. (2020), 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.

This paper also relates to a growing body of experimental evidence showing that AI copilots decrease inequality by boosting the productivity of low-skilled and less experienced workers. This pattern has been documented across a wide range of occupations, including taxi drivers (Kanazawa et al. 2022), customer-support agents (Brynjolfsson et al. 2025), health care advisors (Gruber et al. 2020), lawyers (Choi et al. 2023), consultants (Dell'Acqua et al. 2023), writers (Noy and Zhang 2023), and software engineers (Peng et al. 2023). Engineers and manufacturers have argued that AI can be used to empower the less-skilled to be more productive and do expert work (Alam et al. 2024). Autor (2024) argues that AI has the potential to assist with restoring the middle-skill.

I contribute to this literature by showing that the predictions from small-scale experiments and the conjectures of engineers and economists hold at the aggregate level: in Brazil, AI has contributed to a reduction in both within- and across-occupation inequality by complementing low-skilled workers operating machinery.

The paper is organized as follows. Section 2 describes the data and the institutional context. Section 3 presents stylized facts about AI development in Brazil. Section 4 outlines the empirical strategy. Section 5 discusses the main results, and Section 6 addresses robustness. Section 7 concludes.

2 Data and Institutional Setting

This section introduces the data and explains why Brazil offers an unusually clear setting for studying the labor-market effects of AI. The main reason is the software registry maintained by the National Institute of Industrial Property (NIIP), which functions as a near-complete census of in-house, third-party, and commercially marketed software (Barbosa et al. 2022). Since a 1987 statute granted fifty-year copyright protection to any original program filed with the NIIP, registration has become routine: 96 percent of domestic IT firms appear in the database alongside firms developing software in other sectors.[5]

A second reason to study Brazil is its mature and self-sufficient IT sector. Brazil has the 11th largest IT services industry in the world, ahead of advanced economies such as Sweden, Australia, Denmark, and Spain, and hosts the third-largest pool of software developers (GitHub Staff 2024). As early as 2001, Brazil had the seventh largest IT sector in the world, on par with China and India (Junqueira Botelho et al. 2005).

As a result of booming local development of software (Junqueira Botelho et al. 2005, Arora and Gambardella 2005) and regulatory barriers (Ezell et al. 2013, Benz and Jaax 2022), the import share of computer services in Brazil, which includes software license fees, was just 9.3% in 2015.[6] Digital services generally exhibit strong home bias (Alaveras and Martens 2015, Miroudot et al. 2010, Harms and Shuvalova 2020); among OECD countries, the average import share of computer services is only 19%. This means that most software used in Brazil is developed locally and thus captured by the NIIP software registry.[7]

The third advantage of studying Brazil is the administrative dataset *"Relação Anual de Informações Sociais"* (RAIS), that covers all formal jobs from 2003 to 2022, which allows tracking wages and employment by eight-digit occupation. By linking NIIP software descriptions to occupation-level tasks, I construct a broad measure of AI exposure.

[5] As I describe in detail below, 96% is a lower bound on the coverage of the dataset, since many software design firms provide services to third parties who then register the software under their own name.

[6] Consistent with the low import share of computer services, fewer than 0.02% of IT firms were foreign-owned in 2016.

[7] I discuss in detail the import share of computer services in Appendix A.2.

2.1 Administrative Software Registry Data

This section describes a unique dataset of Brazilian AI software, enabled by a copyright system that records most software. Because registration serves as legal proof of ownership—required for public bids, R&D loans, venture funding, and anti-piracy requests—developers routinely file every commercial program they release. This setting provides me access to thousands of AI program descriptions across diverse applications, from predictive maintenance in factories to office automation tools. Unlike conventional patent data, this database covers practically every piece of commercialized software, providing a far more detailed picture of how AI has evolved in Brazil.

2.1.1 Institutional Setting

Since 1987, the Brazilian government has allowed companies to register their software with the National Institute of Industrial Property (NIIP) under a special copyright system.[8] The registration process is similar to that of patents but less demanding: it requires no claims of novelty or inventive step, involves no technical review, charges low fees, and keeps the source code confidential. While it does not grant exclusive rights over the underlying idea, as patents do, it protects against copying or distribution of the software itself, offering firms effective protection against piracy. Because registration provides legal security at low cost, it has become standard practice. This distinctive feature of Brazilian copyright law allows me to identify a uniquely representative sample of AI applications developed in the country.

Brazilian software law provides special copyright protection for software. In 1987, Brazil passed a law granting 50 years of copyright protection to a program's source code and executables.[9] Firms that register software with the NIIP can stop others from copying, distributing, or modifying their code, but—unlike patents—do not gain exclusive rights to the underlying market concept. Piracy of a protected software carry criminal penalties of up to four years in prison and substantial fines, giving firms strong incentives to

[8] In Portuguese, the NIIP is called *Instituto Nacional de Propriedade Industrial*.

[9] 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."

register even routine applications.[10]

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.

A software registration contains detailed information about the software. To register software, firms must submit an electronic form with the firm's name, the names of the software's programmers, the software's application domain, the methods used, a descriptive title of the software, and the programming languages. The application domain identifies where the software is intended to be used, while the technical class specifies the underlying technology. 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

[10] The law aimed to strengthen intellectual property rights in the IT sector and encourage investment. It was part of a broader set of industrial policies targeting the sector in the late 1980s (Stuber 1984).

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.

Software registration provides proof of ownership. Once software is registered, the firm receives a certificate that serves as official legal proof of ownership. Figure 1 shows one of such a certificate. These certificates are commonly used in cease-and-desist letters, piracy takedown requests, and intellectual property lawsuits. Beyond legal enforcement, they also serve as credentials in financial and institutional contexts, providing evidence of intangible assets to investors, banks, and government agencies.

In litigation, the certificate plays a crucial role in proving ownership. If a company suspects its source code was stolen—by a former employee or a hacker, for example—it can file a lawsuit against the suspected party. As part of the case, the hash code submitted at the time of registration is used to verify that the firm possessed the exact software at that date. The judge then evaluates whether there is sufficient evidence to suspect that part or all of the code was unlawfully copied.

Why are firms registering their software? Firms typically list five reasons for registering their software: intellectual protection, government procurement, access to credit, signaling to investors, and the low financial cost.[11] First, registration provides legal proof of ownership and forms the basis for asserting rights against unauthorized use. Second, registration is mandatory for participating in government procurement of software or IT services. Third, firms seeking R&D credit must show proof of software registration as evidence of prior innovation. Fourth, registration signals the value of a firm's intangible assets to investors. According to the Brazilian Association of Software Firms, most startup investors require all software developed by a firm to be registered.[12] Finally, the process is inexpensive, with the NIIP charging just $34 per registration.

Developers also benefit from software registration. Software registrations list the programmers involved in the project, giving them formal proof of experience. These cre-

[11] Torquato (2022)

[12] Associação Brasileira das Empresas de Software (ABES) (2024)

dentials are often included on résumés and shared on platforms like LinkedIn.[13] de Souza et al. (2024) shows that these credentials matter: programmers receive a 5% permanent wage increase after being listed on a registered software project.

96% of software development companies have at least one software registered. If NIIP registration is standard, most firms that develop software should have at least one registered. In 2016, 96% of software development firms with more than two years of operation had registered at least one software. Employment-weighted, 99.9% of software development companies had at least one software registered.[14]Among larger developers, registration is nearly universal: 99.7 percent of firms with more than 20 employees have at least one program registered with the NIIP.[15] This result confirms that most comercial software created in Brazil is registered with the NIIP.

Most software used in Brazil is developed locally and, therefore, registered with the NIIP. It could be the case that foreign firms do not to adhere to the Brazilian copyright system, therefore leaving imported programs out of the database. In practice, though, the import share of software is small: imported computer services—a category that includes software publishing, programming, data processing, and cross-border licensing fees—accounted for just 9% of the market in 2015. Consistent with this, fewer than 0.02% of IT firms were foreign-owned in 2016.[16] This limited reliance on foreign software can be explained by Brazil's large domestic IT industry (Arora and Gambardella 2005, Junqueira Botelho et al. 2005), institutional barriers to IT imports (Ezell et al. 2013, Benz and Jaax 2022), and the generally high trade costs in services (Miroudot et al. 2010, Harms and Shuvalova 2020,

[13]Figure A1 in the Appendix shows two examples of developers displaying their certificates.

[14]I call a firm a software development firm if it has Classificação Nacional de Atividades Econômicas (CNAE) 2002 codes 6201500 (custom software development), 6202300 (development and licensing of customizable software), 6203100 (development and licensing of non-customizable software), or 6204000 (IT consulting services), and has hired a computer engineer at least once in its life-cycle. I focus on this sample because these firms are especially likely to have deployed at least one software.

[15]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 B.4, software development takes more than three years for 50% of firms.

[16]About 0.03% of employment weighted IT companies are foreign owned. In the broader group of IT and telecomunication firms, 0.04% of employment weighted companies are foreign owned.

Alaveras and Martens 2015). I discuss these factors in more detail in Section A.2. Because the NIIP registry covers locally developed code, these facts imply that the database captures most of the software actually deployed in Brazilian workplaces.

Examples of AI software in Brazil. Imachine AI, developed by Techplus and registered with the NIIP (Figure 1), connects to factory sensors to monitor the temperature, vibration, energy use, and performance of factory equipment.[17] Using fuzzy neural networks, it predicts equipment failures and schedules preventive maintenance. According to Murtaza et al. (2024), predictive maintenance tools like Imachine AI can enhance productivity by minimizing unplanned downtime and improving machine reliability.

Techplus has also registered a second software: a large language model industrial copilot designed to assist factory-floor workers. This copilot functions as a chatbot with access to real-time data from the production line, including sensor readings from machines, usage metrics, input characteristics, and others. It also integrates operation and maintenance manuals, allowing it to answer workers' questions, assist in equipment operation, diagnose problems, and guide routine maintenance. Similar copilots in other contexts have been shown to increase the productivity of workers, especially among less experienced or lower-skilled workers, by providing on-demand expertise and operational support (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).

2.1.2 Software Dataset

This section introduces the software database and shows that it offers a more comprehensive measure of AI development in Brazil than traditional proxies like patents and job ads.

Data collection and variables. I extracted information from the universe of NIIP filings: 37,677 software registrations recorded between 1987 and 2022. For each entry, I observe all the information in the certificate of software registration (Figure 1): the software title, the

[17] According to its registration, the software's application domains are: AD-06 (factory planning, product engineering, prototyping, production planning, quality control); FQ-04 (units of measurement, dimensional analysis, measurement systems); IF-10 (generic data processing); and IN-02 (technology policy, technical cooperation, technological research, technological innovation).

publication and creation dates, the names of the owners, the programmers involved, the programming languages used, the application domains, and the technical classifications.

Importantly, the raw count of 37,677 registrations does not reflect the number of distinct software products. Registration is a legal tool, not a technical unit. A comparison of the software registration with public online records review that firms follow different registration approaches. IT companies usually bundle all their programs in one registration while non-IT companies create a registration for each software that they commission. Legally, both approaches offer the same level of protection. For the purposes of this paper, what matters is that nearly every relevant program appears somewhere in the registry with a description of its use, regardless of how firms choose to structure their filings.

I observe two classifications for each software: one regarding its application domain and another regarding the technology used. The application domain classification identifies the sector or area in which the software is intended to be used. For example, the software *MapForest*, which employs AI to forecast weather and predict land productivity, is classified under agriculture. In total, there are 225 application domains grouped into 35 broad categories. Each software entry has at least one application domain, though most include multiple; on average, each program lists three. Table A3 in the Appendix lists the 35 broad application domains together with a description.

The second classification for each software, called "technical class," organizes software by technical function and purpose. It ranges from fundamental systems such as Operating Systems (SO01) and Network Controllers (SO08) to more specialized tools like encryption software (PD03) and AI applications (IA01). The technical classes of *MapForest*, for instance, are artificial intelligence, simulation, pattern recognition, and office automation. In the empirical analysis, I focus on AI software (IA01). Overall, there are 97 technical classes, divided into 18 broad technical classification groups. Table A2, in the Appendix A.2.1, lists the broad technical classification groups with a description of each.

NIIP's software database has broader coverage than patents. Previous literature studying the effects of AI has primarily relied on patents or job advertisements to measure AI adoption. The NIIP software dataset offers broader coverage than patents for two key reasons: (1) NIIP software registration does not require novelty, and (2) software is typically not patentable. Because registration does not require software to include a novel method or application, the NIIP dataset captures applications of AI that, while not innovative enough for a patent, are still highly relevant to the labor market, such as predictive maintenance software, quality control software,[18] legal automation tools,[19] diagnostic medicine,[20] management,[21] logistics,[22] and many others.

Under U.S. law, software is only patentable if it goes beyond implementing an abstract idea (Samuelson 2017). For example, software that controls a robotic arm may be patentable, while software that calculates optimal pricing is not.[23] The U.S. Supreme Court has determined that certain types of software, such as AI software for combining images (Digitech Image Technologies, LLC v. Electronics for Imaging, Inc.), online transaction software (buySAFE, Inc. v. Google, Inc.), and software implementing numerical algorithms (Gottschalk v. Benson and Parker v. Flook (1978)), are not eligible for patent protection. Any of these software types could be registered with the NIIP. Because of its broader eligibility criteria and fewer restrictions, the NIIP's software records a wider range of AI software than patents.

Software registration is more common than patenting. Table 1 shows that Brazilian firms are more likely to register software with the NIIP than to seek patent protection, highlighting the broader coverage of the NIIP's software registry data. Between 1987 and 2022, firms filed 37,677 software registrations with the NIIP but obtained only 445 software

[18]For instance, Tbit, an industrial software company, use deep neural network, a common method in machine learning, to identify impurities in inputs or outputs.

[19]Jurídico.AI, ChatAdv, and LawX are companies in Brazil that use AI to automate routine legal tasks, particularly tasks traditionally performed by paralegals.

[20]LEUKNET is a convolutional neural network model designed to assist in diagnosing leukemia using blood smear images. Santos et al. (2022) shows that the pandemic led to the creation of several software targeting Covid-19 case monitoring, diagnosis, teleconsultation, simulation, and guidance.

[21]The company Generativa, for instance, registered a software for municipal administration to manage the process of receiving, processing, and responding to citizen inquiries.

[22]Phoenix offers AI solutions for logistics companies to manage their fleets. They use AI to monitor in real time the behavior of drivers, identify if the driver has drunk or is likely to fall sleep while driving, prepare work shift for drivers, and suggest routes based on the type of cargo that the truck is carrying.

[23]In Parker v. Flook (1978), the Supreme Court ruled that an invention containing only a new mathematical algorithm is not patentable.

patents.[24] Even among firms that registered software, only 10% also held a software patent (line 6 of Table 1). These figures confirm that the NIIP registry captures a broader range of commercially deployed software and provides a more comprehensive picture of AI adoption.

Table 1: **Comparison Between the Coverage of Software and Software-Patent**

Number of Software	37,677
Number of Software-Patent	445
Percentage of firms with software-patent with at least one software	87%
Percentage of firms with software with at least one software-patent	10.72%

Note: This table shows statistics from the software database and from patents. Patent data comes from de Souza (2022). Following Webb (2020), we say that a patent is a software patent if the title contains the words “software”, “computer”, or “program” and not the words “chip”, “semiconductor”, or “circuit”.

NIIP’s software database has broader coverage and more detailed information than job ads. Other studies have measured AI adoption using job advertisements, where listings for programmers or roles requiring AI-related skills signal a company’s intent to develop or adopt AI. The NIIP registry offers three key advantages. First, it captures software that is typically adopted through third-party vendors rather than developed in-house. Roughly 30% of commercial AI registrations are filed by specialized IT firms, not by the end users. As a result, off-the-shelf tools—such as HR suites, accounting packages, MES platforms, and predictive maintenance systems—appear in the NIIP data even when the adopting firm never advertises an AI-related position or hire an AI specialist.

Second, the NIIP registry captures completed software, not just planned projects. Surveys of IT firms show that 31.1% of software projects are canceled before completion (The Standish Group 1995), so relying on job ads or résumés can misclassify failed efforts as adoption—a problem the NIIP registry avoids.

Third, the registry includes detailed descriptions of AI software applications, allowing researchers to directly link AI tools to affected occupations.[25] Together, these features make

[24]To classify a patent as software-related, I follow the method proposed by Webb (2020).

[25]A noteworthy distinction is Hampole et al. (2025), which uses text-analysis to infer the characteristics of the AI application from the resume of engineers.

the NIIP a broader measure of AI adoption than hiring data alone.

2.2 Other Datasets

I also use four 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 posts and information on the popularity of programming languages in the U.S., drawn from Tiobe.

RAIS. I use the public 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 2022.[26] 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* (CBO02), which includes detailed occupation descriptions. To avoid crosswalk issues with older classifications, I restrict the analysis to 2003–2022. I exclude from the sample IT workers 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 adjudicate responsibilities and regulatory compliance, these descriptions are long-form, legally precise, and highly specific. The dataset covers 2,641 occupations, each with an average of 64 task statements.

Popularity of Programming Languages. I construct an instrument using historical trends in U.S. programming language popularity using data from Tiobe.[27] Tiobe measures the number of Google.com search results for each language, serving as a proxy for community

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

[27] Tiobe constructs an index of programming language popularity using the number of hits in major search engines. For more information, access https://www.tiobe.com/tiobe-index/.

size, available resources, libraries, and overall interest.[28] The data covers the period of 2000 to 2023.

Stack Overflow. To link programming languages to NIIP's technical classifications, I use data from Stack Overflow, a question-and-answer website widely used by programmers. In this website, each post includes tags that indicate the technologies involved, such as the programming language or the type of application. Tags are selected from a community-maintained list and can only be created or edited by experienced users. Senior users manage tag usage, oversee synonyms, and maintain tag descriptions. Tags play a central role on the platform: they connect questions to domain experts, enable topic-specific moderation, and link directly to the platform's reputation system. Figure A15 plots the most common tags, which often reflect both the programming language and the technical area of the question, allowing me to infer how languages map to software technical classes. I collect data on all Stack Overflow posts created between 2008 and 2022.

3 Statistics of AI Software

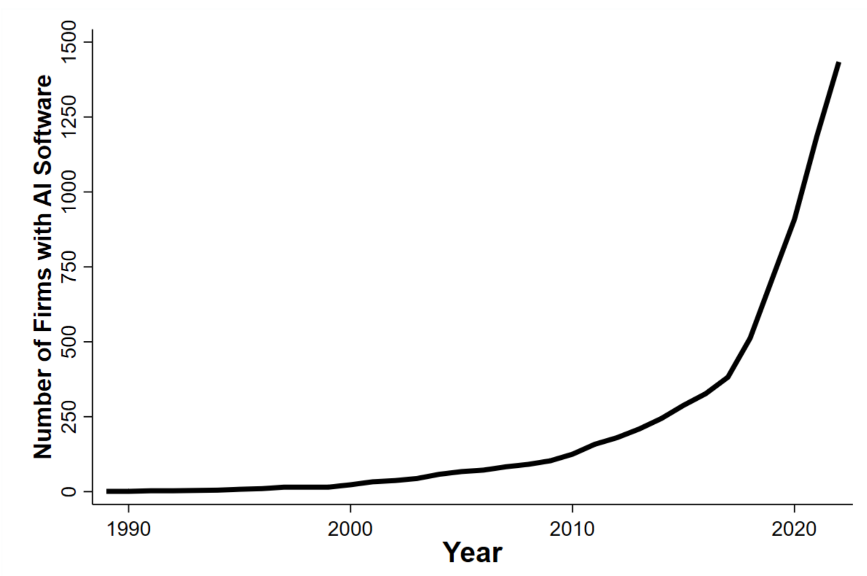
This section presents three key facts about AI software development in Brazil, with additional statistics in Appendix A.3. First, Brazilian AI development accelerated following global breakthroughs in deep neural networks and GPU computing: the number of firms that own an AI software increased more than 10 times between 2010 and 2022. Second, AI is applied across a wide range of fields. While administrative uses such as report generation remain common, 40% of Brazilian AI targets production settings, including manufacturing, agriculture, and logistics. Third, two applications dominate production use: manufacturing execution systems for workflow management and quality-control tools that perform real-time visual inspections.

[28] Google's result estimates reflect indexed pages matching the interpreted query, filtered for duplicates and spam. Results are location- and language-sensitive, favoring content in the user's language and region.

Surge in the creation of AI software following major technological breakthroughs. Figure 2 plots the number of unique firms with at least one AI software registration. [29] Since 30% of these firms develop software for external clients, the number of firms adopting AI is considerably larger than the number of firms registering AI software. Therefore, I consider Figure 2 as illustrative of the trend in AI adoption but not of its levels.

The growth of AI software closely tracks major technological breakthroughs in the field.[30] Before 2000, during the so-called AI winter, only 23 firms had registered AI software. This period was marked by limited interest and investment in AI research (Lee 2018, Mitchell 2020), as the absence of large datasets and insufficient computing power constrained the development of practical applications (Harguess and Ward 2008).

Figure 2: **Number of Firms Owning AI Software**



Notes: This figure shows the number of unique firms in Brazil with at least one AI software from 1987 to 2023. Because many of these firms provide services to others, the number of firms using AI is likely much higher than those directly owning the software.

From the 2000's to 2012, the period known as the AI Renaissance, the number of firms with AI software in Brazil increased 682%, to 180 by 2012. According to Kaplan et al. (2020), the resurgence of AI was driven by an abundance of data, more powerful computers,

[29] I plot the number of firms rather than the raw count of registrations because, as discussed earlier, a registration does not reflect the number of distinct software products. In Appendix A.3, I show that the results are similar when using the share or number of AI software.

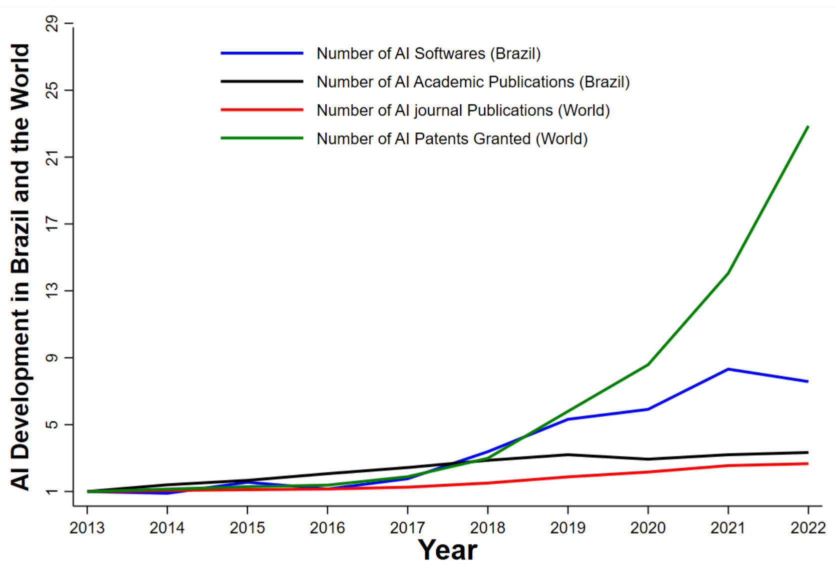
[30] Silva et al. (2025) finds a strong correlation between AI software development and academic publications in machine learning in Brazil.

and advances in machine learning. Notably, Raina et al. (2009) introduced the use of GPUs to train models, Krizhevsky et al. (2012) led to the widespread use of Deep Convolutional Neural Networks for image recognition, and Mikolov et al. (2013) developed word embedding used for speech recognition.

Starting around 2013, the development of AI software accelerated rapidly. The number of firms owning AI technologies increased sevenfold, reaching 1,434 by 2022. This period, often referred to as the AI boom, is marked by a surge in innovation and investment in artificial intelligence tools (Toosi et al. 2021, Chauvet 2018, Sevilla et al. 2022). In the empirical analysis that follows, I show that this wave of AI innovation triggered significant adjustments in the labor market.

Figure 3 shows that the rise in Brazilian AI software creation is part of a broader global surge. The figure plots the number of AI software registrations in the NIIP database alongside three benchmarks: AI patents granted worldwide, AI journal publications, and AI academic publications in Brazil. All series are normalized to 1 in 2013. Between 2013 and 2022, each line at least triples, indicating that Brazilian developers participated in the same global wave of AI innovation that drove increases in patenting and research activity elsewhere.

Figure 3: **AI Development in Brazil Follows Global Trend**



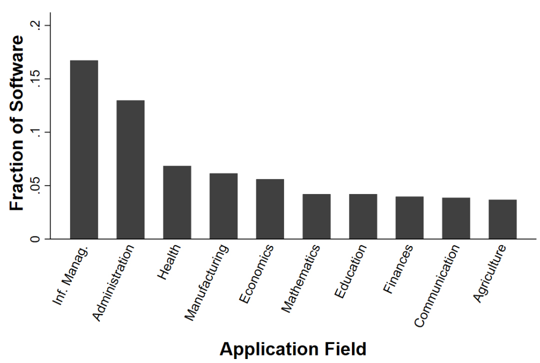
Notes: This figure plots the time series of four indicators of AI development: (i) the number of AI software programs registered in Brazil's NIIP database, (ii) the number of AI-related academic publications in Brazil, (iii) the number of AI-related journal articles globally, and (iv) the number of AI 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.

AI is used in management and production. When firms register their software, they indicate its application domain: a list of areas where the software is intended to be used. Figure 4 presents the ten most common broad application domains for AI software. Managerial uses, such as information management and administration, are the most frequent but account for only 27% of registrations. AI is also widely applied in fields like healthcare, manufacturing, and agriculture, which together represent 16.7% of the total.[31]

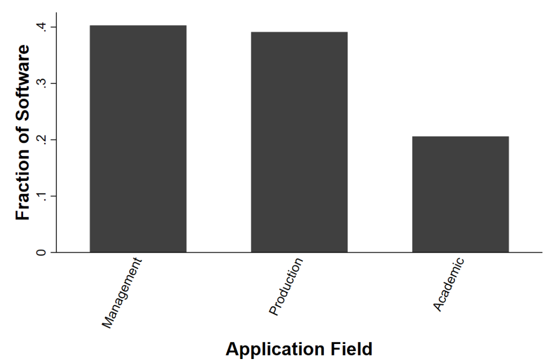
Figure 4b groups all application domains into three categories: management, production, and academic. Management includes back-office functions such as human resources, accounting, and administration. Production covers applications linked to the physical economy, such as manufacturing, agriculture, and construction. The academic category includes research-oriented applications in fields such as mathematics and the social sciences.[32]

Figure 4: **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 A8a 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 4b 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.

[31] Table A3 lists all application domains with descriptions.

[32] 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 4b shows that AI registrations are nearly evenly distributed between management and production applications, each accounting for about 40% of filings. This distributional balance challenges the prevailing view that AI primarily targets high-skill, white-collar occupations (Frey and Osborne 2017a, Webb 2020, Felten et al. 2021). With a comparable share of AI tools applied in production, blue-collar workers may be just as exposed to AI as administrative workers.

Production AI coordinates workflow, controls quality, and automates tasks. To better understand how AI is used in production, I classify each production-related registry into one of seven mutually exclusive categories, based on the engineering literature and a manual review of program descriptions.[33]

The most common use of AI software is in Manufacturing Execution Systems (MES), which coordinate and control production processes. MES accounts for nearly 40% of all AI applications in the database (Figure 6). A concrete example is the AI software developed by Copel, a large electricity distribution company.

Copel's system monitors household-level energy usage and uses AI to predict power outages.[34] When risks are detected, the software automatically dispatches maintenance and inspection teams with a description of the issue, the repairs needed to be made, and the location, increasing the productivity of field workers. According to Consoli, Flávia (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

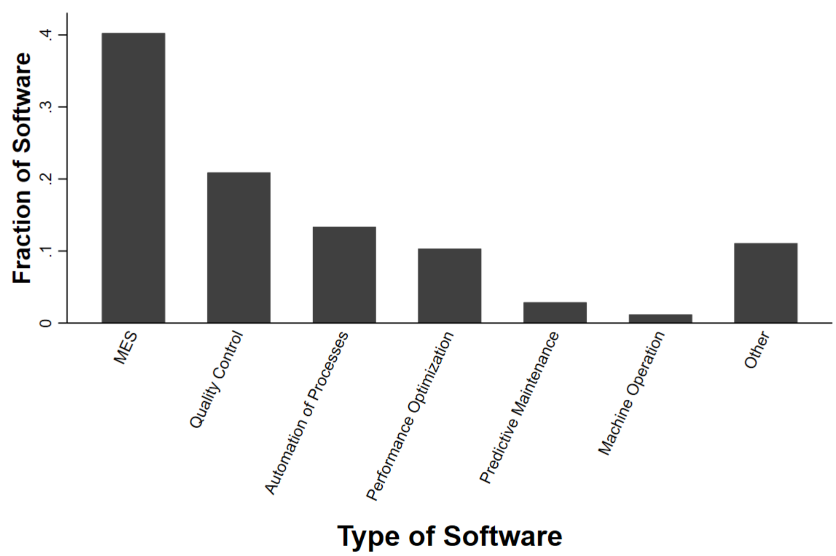
[33]To construct the AI classifications, I began by reviewing the engineering and operations literature on the practical applications of AI in industry. This review identified several recurring use cases, including Manufacturing Execution Systems (MES) (Chen et al. 2006, Capgemini 2020, Shojaeinasab 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 2019), and predictive maintenance (Keleko et al. 2022, Ucar et al. 2024, Murtaza et al. 2024). Following a review of the dataset, I found examples of AI in each of these applications. Based on additional examples discussed in Marr (2019), I also introduced a separate category for machine operation tools. To classify each software entry, I used a large language model (LLM) with a few-shot prompting approach Brown et al. (2020): I provided the LLM with a description of each category and hand-coded examples for guidance, 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.

[34]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).

rates (Capgemini 2020), which could increase productivity and the demand for workers.[35]

The second most common use of AI is in quality control, where software with computer vision 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 and removed defective portions from the production line. 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, this integration of 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).

Figure 5: **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 classification, application domain, and complementary information obtained through online searches.

Takeaways: AI development in Brazil surged after 2013, with use in production and management. This section presented three new facts about the development of AI technologies in Brazil. First, Brazilian AI development has followed global trends, accelerating after the deep-learning breakthroughs of the 2010s. Second, AI use is nearly evenly split

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

between back-office management tools and production-oriented applications, challenging the view that AI primarily targets office workers. Third, within production, most filings focus on Manufacturing Execution Systems and automated quality control technologies, which engineering research links to productivity gains. The following sections examine how the adoption of these technologies has affected the labor market.

4 Empirics

This section presents the empirical strategy used to estimate the effect of AI on the labor market. Following Kogan et al. (2023), de Souza and Li (2023), and Webb (2020), I first construct an occupation-level AI exposure measure 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 changes in U.S. programming-language popularity. These changes affect the availability of AI programmers and libraries in Brazil, generating exogenous variation in AI development cost that is plausibly unrelated to labor-market conditions in Brazil.

4.1 AI Exposure

Text similarity between occupations and AI software. To measure each occupation's exposure to AI, I link the tasks performed by workers to those executed by AI software. This measure captures the overlap between AI applications and worker's tasks, either due to task replacement or complementarity. It is built in three steps. First, I collect detailed task descriptions for each occupation o from the Brazilian Ministry of Labor. Second, I construct descriptions for each AI software s by combining its title and the description of the listed application domains. Third, I calculate the cosine text similarity, $\rho_{o,s}$, between occupation and AI software descriptions using latent semantic analysis. Appendix B.1 provides further details on the calculation of the cosine similarity.

Table 2 presents three illustrative AI programs and the three occupations with the highest cosine similarity scores for each. The first program, whose software registration is shown in

1, is a predictive maintenance tool most closely associated with occupations involved in machine maintenance and calibration. The second software automates the processing and sale of insurance policies and is strongly linked to insurance sales representatives and brokers. The third program analyzes soil samples to identify mineral content and is most similar with occupations such as miners and oil well cementers, whose tasks require identifying and studying minerals found in the soil.[36]

Table 2: **Top Occupation Matches for Selected Software Titles**

Occ. Code	**Occupation Title**	**Similarity**
Imachine AI: Online Predictive Maintenance System for Electric Machines Using Machine Learning Techniques		
3134-10	Instrumentation technician	0.265
3134-05	Calibration technician	0.265
3134-15	Supervisor of maintenance for control, measurement, and related instruments	0.262
Software for Insurance Sales and Customer Service Using Blockchain and Smart Contracts		
3517-40	Insurance technician	0.417
3545-05	Insurance broker	0.399
3517-05	Insurance analyst	0.389
Software with a Computational Model Based on Semi-Supervised Learning for Automatic Recognition of Potential Mineral Bodies		
3163-40	Oil well cementer	0.271
3163-35	Wellbore scale removal technician	0.280
7114-05	Miner	0.344

Notes: This table shows the occupations with highest text similarity to each AI software. Column 1 contains the occupation code, column 2 the occupation title, and column 3 the text similarity. Each panel shows the occupations with highest cosine similarity for each AI software.

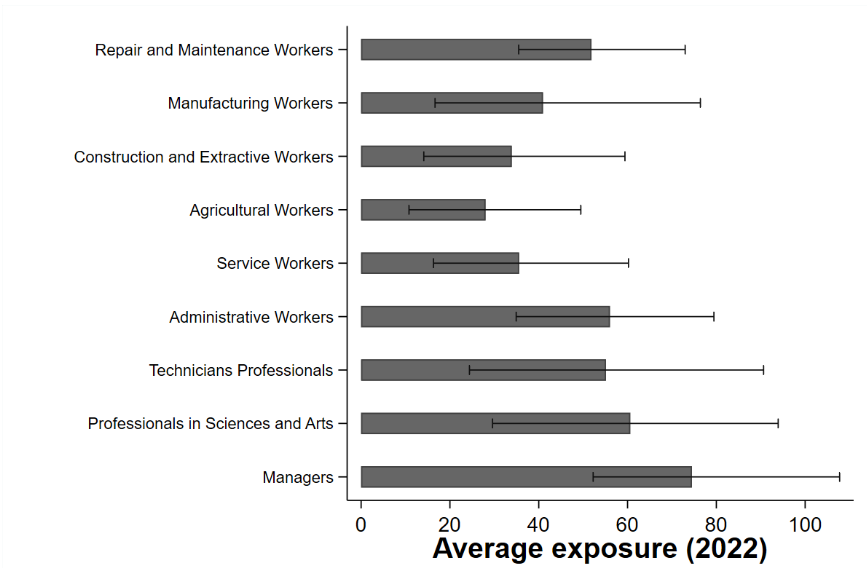
Exposure to AI. Using the cosine text similarity, I calculate the exposure of each occupation o to the AI software created before year t as:

$$AI\ Exposure_{o,t} = \sum_{s} \rho_{o,s} \mathbb{I}_s \left\{ t(s) \leq t \right\}, \tag{1}$$

[36] Among the tasks performed by an oil well cementer are prospect, explore, and analyzing mineral and soil data.

where $\rho_{o,s}$ represents the text similarity between occupation o and software s, and $\mathbb{I}_s \{t(s) \leq t\}$ is an indicator variable that equals one if software s, created in year $t(s)$, was developed before year t. Because firms register software right before commercialization, $t(s)$ is also the time in which adoption of the software begins. Thus, $AI\ Exposure_{o,t}$ measures the cumulative exposure of occupation o to all AI software developed and commercialized up to year t.

Figure 6: **Large Variance within and Across Broad Occupations in AI Exposure**



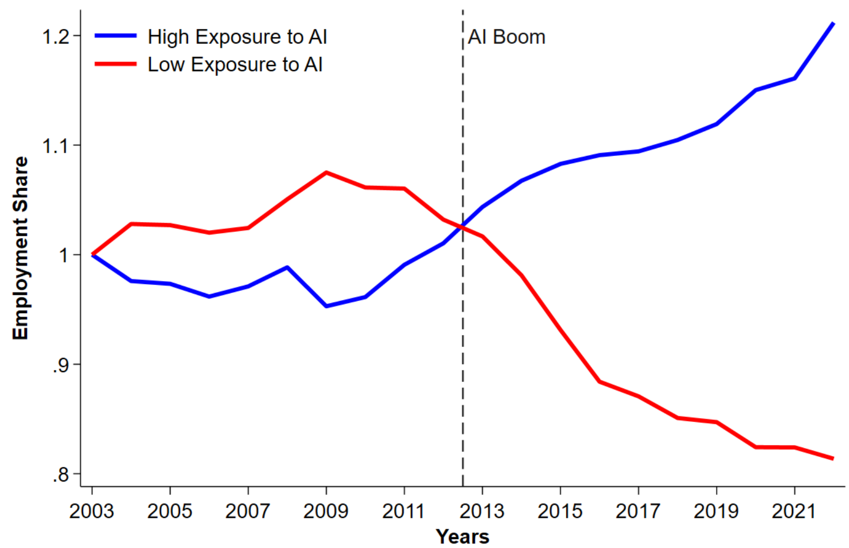
Notes: This figure plots the average AI exposure in 2022 at a one digit CBO occupational classification. The bars plot the average AI exposure and the lines the top and bottom decile.

Occupations most exposed to AI. Figure 6 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.[37] 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 3.

[37] 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 B.2, I compare the AI exposure measure 1 to other measures developed in the literature.

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 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.

Figure 7: **Occupations More Exposed to AI Experienced Faster Growth After AI Boom**

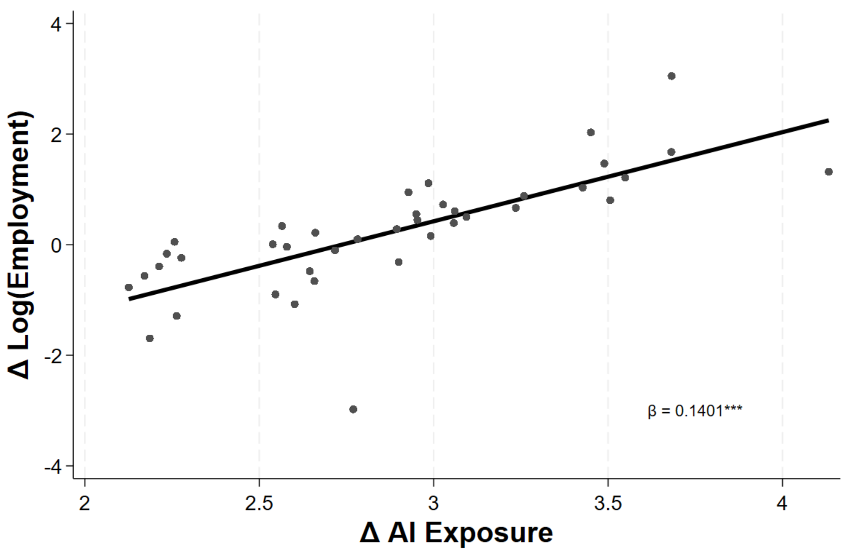


Notes: Figure 7 plots the log employment of occupations in the top and bottom deciles of the 2022 AI-exposure distribution, normalizing both series to 1 in 2003.

Employment has increased relatively faster in occupations more exposed to AI. Figure 7 plots the employment share of occupations in the top and bottom deciles of the 2022 AI-exposure distribution, normalizing both series to 1 in 2003. From 2003 to 2012, the two employment trajectories moved in parallel. Starting in the first year of the AI boom (2013), when the development of AI software in Brazil exploded, employment in high-exposure occupations began to rise relative to low-exposure ones. By 2022, employment share in the most exposed occupations increased by 20%.

Supporting the pattern from Figure 7, Figure 8 shows a strong positive correlation between changes in AI exposure and employment growth from 2003 to 2022. This correlation suggests that AI may have contributed to employment growth. In the following subsections, I present a method to identify the causal effect of AI out of this correlation.

Figure 8: **Exposure to AI is Correlated with Larger Employment Growth**



Notes: This figure shows the correlation between changes in AI exposure and employment growth from 2003 to 2022. The data is binned into 40 dots, each representing the average change in employment and AI exposure within each percentile intervals. The regression line is run with all the observations.

4.2 Empirical Model

Main empirical model. To estimate the effect of AI, I use the following empirical model:

$$y_{o,t} = \beta AI\ Exposure_{o,t} + \mu_o + \mu_t + X'_{o,t}\theta + \epsilon_{o,t}, \tag{2}$$

where $y_{o,t}$ is a labor market outcome for occupation o in year t, such as average wages or employment. $AI\ Exposure_{o,t}$ is the AI exposure measure defined in 1. Since the exposure measure is normalized, β represents the effect of a one standard deviation increase in the AI exposure on labor market outcomes in year t. The terms μ_o and μ_t denote occupation and year fixed effects, respectively.[38]

[38] Since each occupation o corresponds to a narrowly defined six-digit occupational code narrow enough to, most often, include sectoral requirements, I do not use a sector–occupation specification as in Webb (2020). In the robustness section, however, I show that adopting a specification similar to Webb (2020) delivers qualitatively similar results.

Controls to deal with omitted variable bias. To limit omitted-variable bias, the control vector $X_{o,t}$ includes three terms designed to address specific confounders. First, I add a one-digit-occupation-by-year fixed effect to absorb shocks common to broad occupational groups. Second, to account for exposure to foreign-developed software, I include a U.S. patent-based AI exposure measure.[39] Although, as discussed in Section A.2, imported software accounts for less than 10% of the market. Finally, I control for potential trends by including the lagged outcome variable, $y_{o,t}$. As shown in the robustness checks, results remain consistent without these controls.

Dynamic Model. To test for parallel trends and capture the dynamic effects of AI on the labor market, I estimate the following model:

$$y_{o,t+j} = \beta_j AI\ Exposure_{o,t} + \mu_o + \mu_t + X'_{o,t}\theta + \epsilon_{o,t} \tag{3}$$

where β_j captures the effect of a one standard deviation increase in AI exposure on the labor outcome $y_{o,t+j}$, measured j years after the software is adopted.

Occupation level analysis includes third party software and captures general equilibrium forces. In this paper, I study the effects of AI at the occupation level rather than the firm level, for two main reasons. First, a firm-level approach would cover only those companies that develop AI in-house, omitting the many that adopt AI via third-party vendors. Since 35% of AI developers supply software used across multiple firms, this approach would overlook a substantial share of labor market effects. An occupation-level analysis better leverages the NIIP dataset, which, different from other datasets, contains a description of software developed by third-party vendors. Second, AI adoption in one firm can affect employment in others through market competition and general equilibrium dynamics. Studying occupations, rather than firms, allows the analysis to broader effects, including worker mobility and displacement across employers. A more detailed study of firm-level outcomes among AI developers is left to future work.

[39] To construct the exposure to patents in the US, I follow the method proposed by Webb (2020).

4.3 Instrument

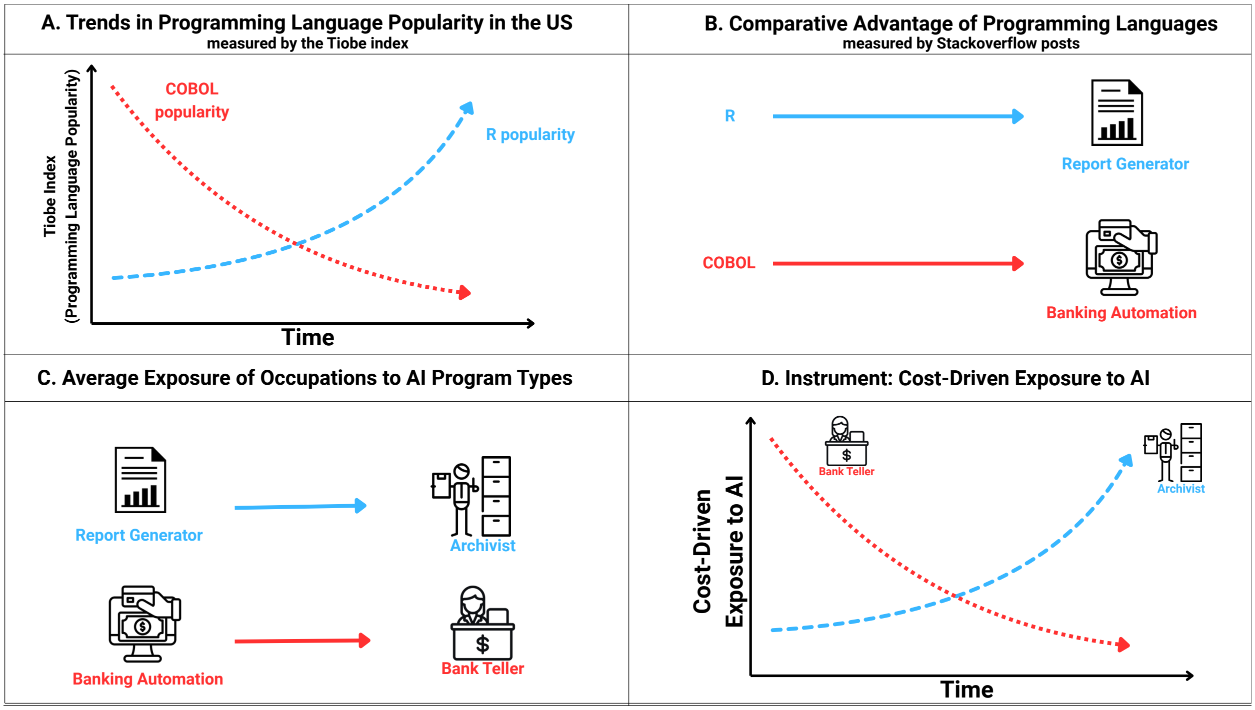
Endogeneity concerns: trends and directed technological change. An OLS estimate of β in equation (2) may be biased for two reasons. First, software companies may direct AI development toward tasks performed by high-wage workers, generating a spurious positive correlation between AI exposure and wages. Second, AI adoption may coincide with other occupation-specific trends, confounding its true effect. To isolate the causal impact of AI, I instrument AI exposure using variation in the cost of developing AI software, proxied by changes in the relative popularity of programming languages in the United States.

Exogenous variation: changes in the cost of creating software. I construct an instrument based on differential exposure to a decline in the cost of building AI software. The key idea is that this cost depends on the size of a programming language's supporting ecosystem (Delorey et al. 2007, Krein et al. 2010, Tambad et al. 2020, Vijayaraghavan et al. 2022). Languages with large, active communities provide extensive libraries, tutorials, forums, and off-the-shelf code, allowing developers to reuse scripts rather than build them from scratch. As U.S. communities around a language grow, this stock of reusable knowledge expands, allowing Brazilian firms to develop AI software at lower cost.

The cost advantage created by a growing programming-language community does not affect all occupations equally. 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 instance, as the U.S. Matlab community grows, it becomes easier to build AI engineering tools; if HTML grows, the cost advantage instead favors web applications. Consequently, as some languages gain popularity in the U.S. while others decline, AI tools in Brazil become disproportionately cheaper to build in certain application areas. As consequence, 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

Figure 9: **Illustration of the Instrument**



Notes: This figure illustrates how the instrument is constructed.

applications and remains widely used in financial systems and government operations because of its ability to manage large financial datasets (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 top 22 in 2023 (TIOBE 2025). This decline led to a shortage of experienced programmers, outdated documentation, and difficulties integrating with modern languages. Overall, these constraints make software development for the financial sector relatively more costly (Navot 2024).

Following a different trend to 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 (TIOBE 2025). As a result, developing software for statistical analysis is comparatively less costly due to the large pool of available resources for R.

Therefore, because writing software in COBOL is relatively more expensive than in R, occupations linked to banking automation—such as bank tellers or bank managers—are less exposed to AI than those focused on statistical reports, like archivists. The instrument identifies the effect of AI on employment by comparing employment growth between these two types of occupations, as illustrated in Figure 9.

Instrument definition. Using these insights, the easiness to build AI applications targeting occupation o in year t is given by

$$\textit{AI Ease-of-Dev}_{o,t} = \sum_{l,p} \frac{\textit{Resources in US}_{l,t-2} \times \textit{Comparative Advantage}_{l,p} \times \overline{\textit{AI Exposure}}_{p,o,t-2}}{N_l N_p}, \tag{4}$$

where l indexes programming languages and p denotes a software technical class from the NIIP classification. The instrument combines three components: (i) the time-varying availability of resources for language l in the U.S. lagged 2 years, $\textit{Resources in US}_{l,t-2}$; (ii) the degree to which language l is suited to technical class p, $\textit{Comparative Advantage}_{l,p}$; and (iii) the baseline relevance of technical class p to tasks in occupation o, $\overline{\textit{AI Exposure}}_{p,o,t-2}$, calculated using data lagged 2 years. Normalizing by the number of languages N_l and technical classes N_p. I explain each component of the instrument in detail below.

Number of webpages in the U.S. mentioning a programming language is a proxy for the availability of resources. I proxy resource availability, $\textit{Resources in US}_{l,t-2}$, with the number of U.S.-based Google search hits for programming language l, lagged by two years.[40] This measure reflects the volume of U.S.-based web content related to the language, such as libraries, tutorials, books, and forum discussions. A greater volume of resources lowers the marginal cost of creating AI software by increasing the productivity of programmers, who can reuse available scripts rather than starting from scratch (Delorey et al. 2007, Krein et al. 2010, Tambad et al. 2020, Vijayaraghavan et al. 2022). Supporting this link, Table A7, in the appendix, shows that higher lagged U.S. search hits predict more

[40] Appendix A.4 summarizes the Tiobe dataset, which is the source for the number of hits of each programming language.

Brazilian AI software using that language. I lag the search hits by two years to reflect the information available at the start of a software project and to mitigate reverse causality.[41] I use the resource availability in the U.S. because Brazilian resource availability could be affected by labor market shocks in Brazil.

Stack Overflow reflects the comparative advantage of programming languages. The second component of the instrument, $\textit{Comparative Advantage}_{l,p}$, measures the comparative advantage of programming language l to the NIIP's software technical class p. To measure that, I calculate for each programming language the share of posts on Stack Overflow with flags of programming language l and technical class p.[42]

$$\textit{Comparative Advantage}_{l,p} = \frac{\textit{Stack Overflow posts tagging language } l \textit{ and technical class } p}{\textit{Stack Overflow posts tagging language } l} \tag{5}$$

The component $\textit{Comparative Advantage}_{l,p}$ reflects two aspects of software development: a language's built-in comparative advantage for certain applications and the availability of libraries. Some programming languages are designed for specific tasks: HTML and JavaScript for web development, Swift for app development, Fortran for scientific computing, SQL for data management, and R for data science, among many others. Beyond design intent, historical factors such as legacy code, rich libraries, and active user communities reinforce specialization. For example, C++ and C# dominate game development (Wijdan 2020), while Python and Perl are widely used in bioinformatics (Fourment and Gillings 2008).[43] Table A5 shows that, on average, the leading language in each application area accounts

[41] According to statistics discussed in B.4, 45% of software take more than 2 years to finish.

[42] Stack Overflow is a popular question-and-answer website for computer programmers. Users create posts there asking for support on their code. When creating a post, users have to choose flags that reflect the nature of their post. Figure A14 lists the top 25 most popular flags in Stack Overflow. Those flags reflect the programming language used, the type of code written, and the overall application of the software. To link tags in Stack Overflow to NIIP's technical classes, I created a cross-walk. Section A.5, in the Appendix, provides summary statistics of Stack Overflow posts.

[43] According to Meyerovich and Rabkin (2013), the availability of libraries, community support, and legacy code largely determines why certain fields converge on a few languages. Rabah et al. (2010) emphasizes the inherent trade-offs across programming languages, which prevent any single one from dominating all applications.

for over 42% of related software projects, demonstrating that programming languages are indeed specialized in practice.

Average AI exposure by technical class. The final component of the instrument connects technical classes to occupations. Specifically, $\overline{AI\ Exposure}_{p,o,t-2}$ measures the average text similarity between AI software in technical class p and the task descriptions of occupation o, using only software registered before $t-2$. Limiting the calculation to lagged years ensures that these exposure weights are predetermined and not mechanically correlated with AI adoption, which is used in the first-stage estimation.[44]

This component captures two dimensions of the relationship between technical classes and occupations. First, some AI technology classes are more relevant to certain occupations than others. For instance, archivists are more exposed to technical classes related to report generation than to those involving banking automation. Second, it reflects the breadth of exposure: some occupations are more broadly exposed to AI across many technical classes and, therefore, should be more affected by a decrease in the implementation cost of AI. For example, archivists, with high similarity scores across multiple classes, are more exposed than secretaries, whose average exposure is lower.

Aggregating to the same level of the AI exposure measure. As the AI exposure measure is build using all AI software developed up to time t, I aggregate the AI Ease-of-Development in the same way:

$$AI\ Ease\text{-}of\text{-}Dev\ IV_{o,t} = \sum_{s=2003}^{t} AI\ Ease\text{-}of\text{-}Dev_{o,s} \tag{6}$$

Therefore, $AI\ Ease\text{-}of\text{-}Dev\ IV_{o,t}$ captures the easiness to develop AI software up to year t.

Identifying variation. The instrument exploits differential exposure across occupations to an exogenous decline in the cost of developing AI software. Specifically, it compares out-

[44] In the robustness section, I show that using AI software created before 2003 only does not change the results. This specification is more appealing because it captures the evolving applications of AI.

comes for occupations more versus less exposed to technical classes that become cheaper to implement due to changes in programming language popularity. 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 over time (Goldsmith-Pinkham et al. 2020). This assumption ensures that any divergence in outcomes can be attributed to the differential effect of AI development costs, rather than to pre-existing trends or omitted shocks.

5 Results

In this section, I show that AI increases employment among young, low-education, inexperienced, and low-ability workers, which in turn lowers average wages due to a change in the composition of workers. Yet, these aggregate effects hide substantial heterogeneity. Employment gains are concentrated in production roles tied to machine operation, while administrative occupations experience declines in both employment and wages. These results are consistent with the view that, in the factory, AI increases the productivity and operability of machines, leading to capital deepening and the hiring of low-skill workers to operate them. In contrast, in office settings, these results are consistent with AI-driven automation of routine administrative tasks.

5.1 AI increases employment of low-skilled workers

AI ease-of-development instrument predicts AI exposure. Table 3 reports the first-stage under various sets of controls. Column 2, which presents the baseline specification, shows that a one–standard deviation increase in the AI ease-of-development instrument increases AI exposure by approximately 0.78 standard deviations. The F-statistics is far above the threshold of 10, indicating that the instrument comfortably passes the weak instruments test.

Table 3: **First-Stage: Ease-of-Development Increases AI Exposure**

	(1)	(2)	(3)
	AI Exposure	*AI Exposure*	*AI Exposure*
AI DevEase IV	0.862***	0.780***	0.784***
	(0.00978)	(0.0101)	(0.0103)
N	45,336	45,336	45,336
R^2	0.751	0.770	0.780
F	7015.7	6291.7	5127.0

Notes: This table presents the estimates of the first-stage. The instrument is given by 6 and the endogenous variable is the exposure to AI, defined in 1. All columns include occupation and year fixed effects. Column 2 adds a 1-digit occupation-year fixed effect, and Column 3 a 2-digit occupation-year fixed effect. Standard errors are in parentheses. Standard errors are clustered at the occupation level. $*p < 0.10, **p < 0.05, ***p < 0.010$.

AI leads to net employment gains. Column 1 of Table 4 shows that occupations more exposed to AI experienced faster employment growth. A one–standard deviation increase in AI exposure raises employment by 1.9%.

The OLS estimate in Table 4 is slightly smaller than the 2SLS estimate but similar in magnitude. One plausible explanation is that AI development in Brazil is largely supply-driven: domestic IT firms adopt applications based on technological breakthroughs abroad rather than designing software in response to local labor market conditions. As a result, AI exposure mainly reflects foreign cost shocks, which are captured by the instrument, and is only weakly correlated with domestic labor shocks. This weak correlation with domestic labor shocks could explain the similarity between the OLS and IV estimates.

Figure 10 illustrates the identifying variation by plotting the component of log employment unexplained by the controls for occupations in the top and bottom deciles of growth in the AI ease-of-development instrument. Before 2010, both groups followed parallel trends, supporting the identifying assumption. Beginning in 2013, coinciding with the AI boom and key deep-learning breakthroughs, employment in high-exposure occupations began to increase while low-exposure occupations followed its trend. The gap between the two groups continued to widen throughout the AI boom that followed.

Figure A19 in the Appendix plots the correlation between employment and the AI Ease-of-Development instrument. The figure shows that, across the sample, higher values of the instrument are consistently associated with higher levels of employment. This positive relationship suggests that the main result is not driven by outliers or a small set of influential observations.

Table 4: **AI Increases Employment of Younger and Less Skilled Workers**

	(1)	(2)	(3)	(4)
	log(Employment)	*log(Age)*	*log(Yrs. Education)*	*log(Experience)*
Panel A: 2SLS				
AI Exposure	0.0194^{***}	-0.00498^{***}	-0.00658^{***}	-0.0199^{***}
	(0.00722)	(0.00112)	(0.00113)	(0.00591)
Panel B: OLS				
AI Exposure	0.0147^{**}	-0.00415^{***}	-0.00521^{***}	-0.0208^{***}
	(0.00583)	(0.000824)	(0.000783)	(0.00420)
N	45,336	45,336	45,332	45,336

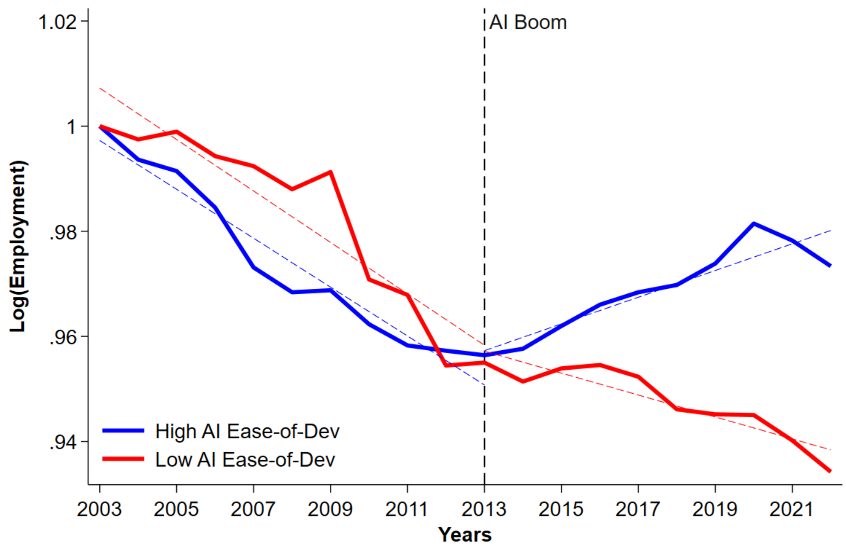
Notes: This table reports estimates of the effect of AI on labor market outcomes, following the specification in Section 2. Panel A presents the 2SLS estimates using the AI ease-of-development instrument defined in Equation 6. Panel B shows the corresponding OLS estimates. Column 1 reports the effect of AI on log employment; column 2 on the log of average age; column 3 on the log of average years of education; column 4 on the log of average tenure at the current firm (in years); column 5 on the log of average wage; and column 6 on the log of the standard deviation of wages. Controls include occupation fixed effects, one-digit-occupation-by-year fixed effects, and exposure to U.S. AI patents. Standard errors, clustered at the occupation level, appear in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

AI shifts employment toward younger, less educated, and inexperienced workers. Columns 2 to 4 of Table 4 show that AI shifts the workforce toward younger, less educated, and less experienced workers. Specifically, column 2 reports a decrease in average age; column 3, a decline in average years of education; and column 4, a drop in average firm tenure.

According to Table A10, in the Appendix, the shift toward younger, less educated, and less experienced workers comes from larger employment gains within these groups and not declines in others. Column 1 shows the effect of AI on employment by demographic group. AI increases employment for all groups. The strongest effects occur among workers who are younger, less educated, and less experienced, explaining the observed compositional shift.

AI reduces average wages due to the entry of lower-ability workers. Column 1 of Table 5 shows that a one-standard-deviation increase in AI exposure lowers average monthly wages by 2%. This decline reflects the shift in hiring toward younger, less experienced workers: AI draws more low-wage entrants into high-exposure occupations, pulling the average wage down.

Figure 10: **Employment Grew Faster in Occupations with Higher AI Ease-of-Development after AI Breakthroughs**



Notes: This figure illustrates the identifying variation by plotting the component of log employment unexpalined by the controls for occupations in the top and bottom deciles of growth in the AI ease-of-development instrument. To construct this figure, I compute the long difference in the instrument as $\Delta AI\ DevEase\ IV_o = AI\ DevEase\ IV_{o,2022} - AI\ DevEase\ IV_{o,2003}$ and regress log employment on the full set of controls, including occupation fixed effects, one-digit-occupation-by-year fixed effects, and exposure to U.S. AI patents. The regression excludes both the instrument and the AI exposure. Using the residual, I construct a time-series for log employment. The figure plots the log employment time-series from this regression for the top and bottom deciles of $\Delta AI\ DevEase\ IV_o$.

To understand how AI affects the entry of low-ability workers, I estimate individual-specific abilities using a Mincer regression, where the worker fixed effect from a regression of observable characteristics on wage history captures each individual's underlying ability. Column 2 of Table 5 reports the effect of AI on the average ability of workers in each occupation. Because ability is fixed at the individual level, changes in average ability reflect shifts in workforce composition. Column 3 shows the average wage residual from the Mincer regression, i.e., the portion of wages not explained by observable characteristics or fixed effects. The wage residual reflects, among other factors, the wage per unit of productivity. Appendix B.5 provides a detailed description of the Mincer regression and estimation procedure.[45]

Column 2 of Table 5 shows that AI leads to the entry of lower-ability workers. Col-

[45]The ability is calculated using the identified RAIS. Due to administrative lags in the release of this dataset, it only covers the period to 2016. Therefore, these results capture only the beginning of the AI boom.

umn 3 indicates that AI increases wages per efficiency unit, likely reflecting higher labor demand. Together, these results suggest that the decline in average wages comes from a change in the composition of workers: AI draws in workers who, based on both observable and unobservable characteristics, tend to earn lower wages. This pattern likely reflects the fact that AI complements low-ability workers, a finding supported by evidence from multiple experimental studies.

Table 5 also shows that AI reduces wage inequality within occupations. Column 6 reports that a one–standard–deviation increase in AI exposure lowers the within-occupation standard deviation of wages by 3.4%. This decline reflects the already discussed increase in the relative demand for low-ability workers, which compresses the wage distribution within occupations.

Table 5: **AI Lowers Wages and Inequality Due the Entry of Low-Ability Workers**

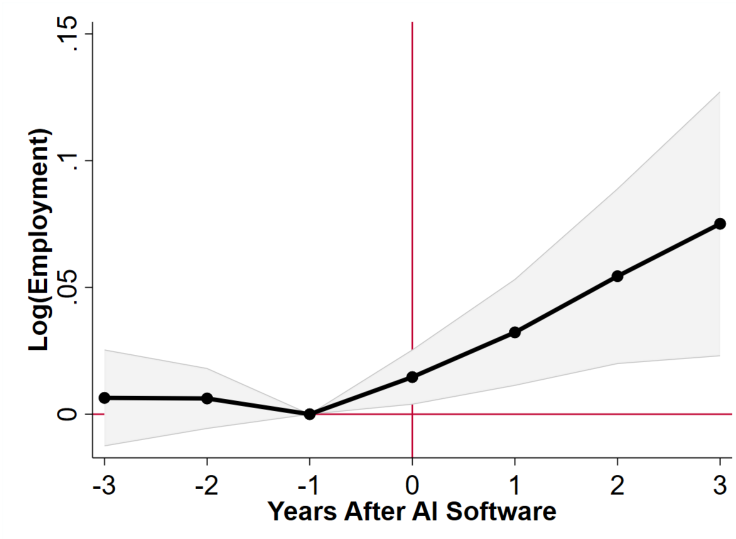
	(1)	(2)	(3)	(4)
	log(Wage)	*Ability*	*Wage Residual*	*log(Std. Wage)*
Panel A: 2SLS				
AI Exposure	-0.0229^{***}	-0.0196^{***}	0.00493^{*}	-0.0345^{***}
	(0.00380)	(0.00401)	(0.00286)	(0.00992)
Panel B: OLS				
AI Exposure	-0.0191^{***}	-0.0158^{***}	0.00278	-0.0294^{***}
	(0.00271)	(0.00376)	(0.00256)	(0.00661)
N	45,336	19,885	19,885	45,208

Notes: This table reports estimates of the effect of AI on labor market outcomes, following the specification in Section 2. Panel A presents the 2SLS estimates using the AI ease-of-development instrument defined in Equation 6. Panel B shows the corresponding OLS estimates. 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 this data use the panel structure of RAIS, it is limited to 2003 to 2016. Section B.5 explains in detail how these variables were calculated. In column 4 the variable of interest is the standard deviation of wages within each occupation. Controls include occupation fixed effects, one-digit-occupation-by-year fixed effects, and exposure to U.S. AI patents. Standard errors, clustered at the occupation level, appear in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

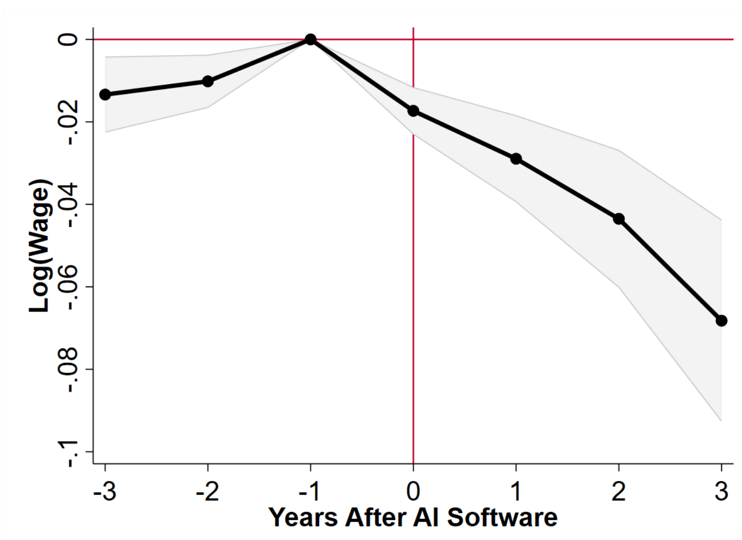
No pre-trends confirm instrument validity. Figure 11 shows the dynamic effects of AI on employment and wages. The x-axis indicates the lag in years, and the y-axis the estimated effect of AI, as defined in model 3. There is no relationship between AI exposure and lagged employment, supporting the validity of the instrument. The correlation with lag wages is small and reflects a positive trend in wages, i.e., at most biasing estimates towards positive effects. In the robustness, I show that adding further lags as controls do not change the results.

Figure 11: **No Pre-Trends and Large Effects on the Long-Run**

(a) Effect of AI on Employment



(b) Effect of AI on Wages



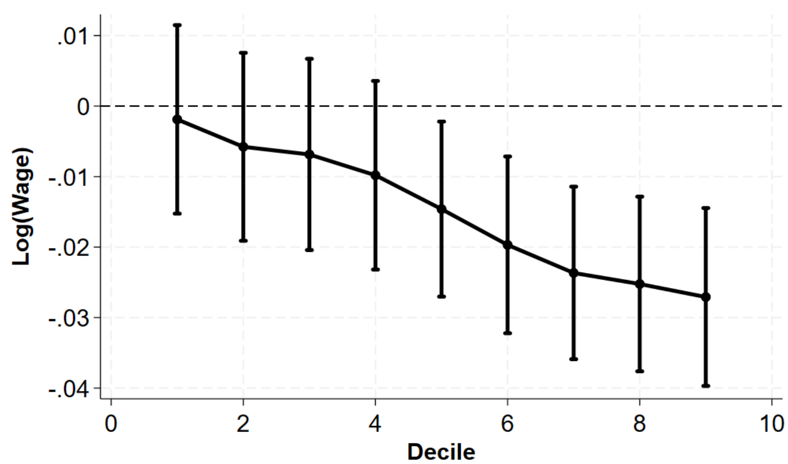
Notes: This figure reports estimates of the dynamic effect of AI on employment and wages, following the specification in Equation 3. All specifications use the AI ease-of-development instrument defined in Equation 6. Controls include occupation fixed effects, one-digit-occupation-by-year fixed effects, and exposure to U.S. AI patents. Standard errors are clustered at the occupation level. The shaded area reports 95% confidence intervals.

Following software registration, AI has an immediate impact on the labor market, likely because firms register software shortly before its adoption or commercialization. Therefore, the registration date is a close proxy for when the software begins affecting the labor market. The effect of AI increases over time: after three years, a one–standard–deviation increase in AI exposure raises employment by 7% and reduces average wages by 6%. This delayed response suggests that AI technologies take time to diffuse and reorganize production, consistent with historical patterns of technological adoption (Solow 1987, Brynjolfsson and Milgrom 2012, Brynjolfsson et al. 2021, 2023).

AI decreases inequality by lowering wages at the top of the distribution. Figure 12 shows how the effect of AI varies across the wage distribution. To construct the figure, I calculate wage deciles within each occupation for every year over the sample period. The y-axis plots the estimated effect of AI on the log wage at each decile.

According to Figure 12, AI reduces inequality by lowering wages at the top of the wage distribution. A one–standard–deviation increase in AI exposure has no significant effect at wages in the lowest decile, while wages in the 9th decile fall by about 2.8%. This pattern, explains the decline in wage inequality reported in Column 6 of Table 4.

Figure 12: **AI Decreases Wages at the Top of the Wage Distribution**



Notes: This figure shows the effect of AI across different points in the wage distribution. To construct this figure, I compute wage-deciles within each occupation over the sample period. The figure plots the estimates of the effect of AI on the log of each wage decile. All specifications use the AI ease-of-development instrument defined in Equation 6. Controls include occupation fixed effects, one-digit-occupation-by-year fixed effects, and exposure to U.S. AI patents. Standard errors are clustered at the occupation level. The bands reports the 95% confidence intervals.

Discussion: AI increases employment of low-skilled workers. This section reports two main findings. First, occupations more exposed to AI experience relatively higher employment growth. Second, this growth is concentrated among younger, less educated, inexperienced, and lower-ability workers. One explanation for the rise in employment is that AI raises overall productivity, which increases labor demand even when some tasks are automated (Acemoglu and Restrepo 2018, Acemoglu et al. 2022, Hampole et al. 2025). Another possibility is that AI increases the productivity of machines, which increases the marginal productivity of labor and leads to higher demand for workers (Agrawal et al. 2019, Gruber et al. 2020). This mechanism is consistent with the evidence that 40% of AI applications are deployed in production settings, such as manufacturing execution systems or predictive maintenance, where engineers have documented substantial productivity gains (Ghelani et al. 2024, Tariq et al. 2024, Patel and Ghelani 2024, Evans and Gao 2016, Marr 2019). In the next subsection, I show that this second explanation better matches the data: employment gains are concentrated in production occupations, especially among workers operating machinery.

My results also show that AI shifts employment toward younger, less educated, less experienced, and lower-ability workers, reducing both average wages and within-occupation inequality. This pattern suggests that AI acts as a skill-replacing technology: it substitutes for expertise, enabling lower-skilled workers to perform tasks that once required more training or experience, a conclusion supported by multiple experiments (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). As far as I am aware, this is the first paper that shows that the conclusion of these micro-level experiments hold in scale.

5.2 AI in the Office and the Factory

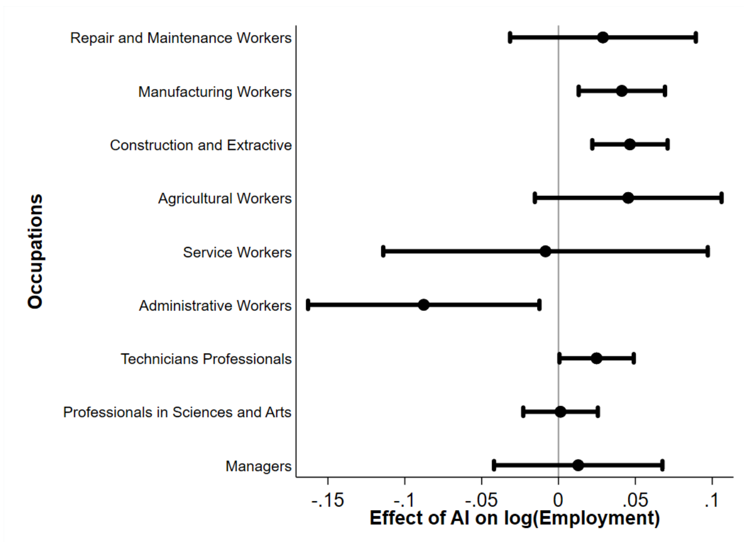
AI replaces workers in the office but increases employment in the factory. In Figure 13, I allow the effect of AI to vary across broad occupational groups by interacting AI exposure with one-digit occupational dummies. [46] A one–standard–deviation increase in AI

[46] 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

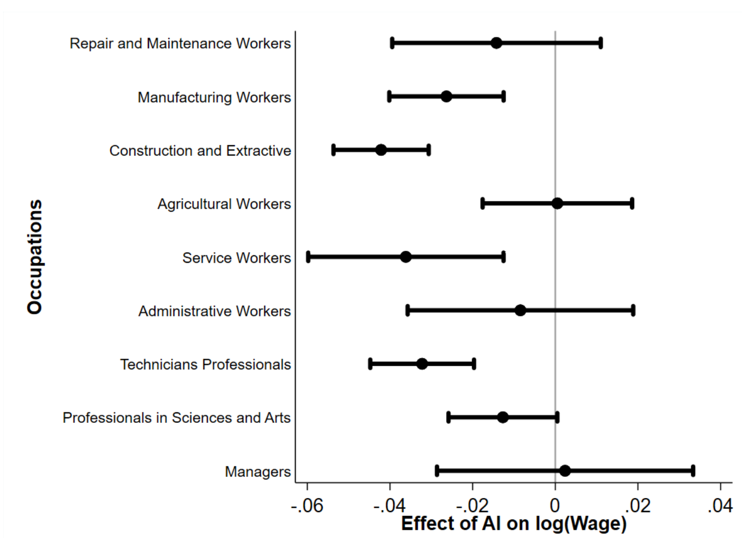
exposure raises employment in production-related occupations, such as workers in agriculture, extraction, manufacturing, maintenance and technicians, by around 4%. In contrast, it reduces employment in administrative roles by 8%. Employment among managers and high-skilled professionals, including engineers and chemists, remains largely unchanged.

Figure 13: **Differential Labor Market Effects of AI Across Occupations**

(a) Effect of AI on Employment for Different Occupations



(b) Effect of AI on Wages for Different Occupations



Notes: This figure reports estimates of the effect of AI on employment and wages 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 6. 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.

---

occupational class of occupation o.

Figure 13b shows the effect of AI on wages across one-digit occupational categories. Managers are the only group for whom AI is associated with a wage increase, though the effect is not statistically significant. In contrast, production-related occupations, which are also the occupations where AI increases employment, see the largest declines in average wages. As shown earlier, AI disproportionately increases employment among younger, less educated, less experienced, and lower-ability workers. These compositional shifts toward lower-wage entrants help explain why average wages fall in occupations where overall employment rises. As I discuss below, these compositional effects are especially pronounced among production workers.

Table 6 shows how the effect of AI varies across occupational groups by dividing the workforce into three categories. The first is production workers, which includes workers in agriculture, extraction, manufacturing, maintenance, and technicians. The second is administrative workers. The third group includes all other occupations, such as managers, skilled professionals, and service workers; for simplicity, I refer to this group as professionals.

Table 6: **AI Shrinks the Office and Expands the Factory**

	(1)	(2)	(3)	(4)	(5)
	log(Employment)	*log(Age)*	*log(Yrs. Education)*	*log(Experience)*	*log(Wage)*
$\mathbb{I}\{Prod.\} \times AI\ Exposure$	0.0437^{***}	-0.00649^{***}	-0.00982^{***}	-0.0222^{***}	-0.0301^{***}
	(0.00997)	(0.00148)	(0.00198)	(0.00832)	(0.00457)
$\mathbb{I}\{Prof.\} \times AI\ Exposure$	0.0108	-0.00456^{***}	-0.00521^{***}	-0.0206^{***}	-0.0196^{***}
	(0.00861)	(0.00135)	(0.00109)	(0.00684)	(0.00466)
$\mathbb{I}\{Adm.\} \times AI\ Exposure$	-0.0878^{**}	0.00419	0.00239	0.0225	0.00789
	(0.0384)	(0.00566)	(0.00302)	(0.0295)	(0.0139)
N	45,336	45,336	45,332	45,336	45,336

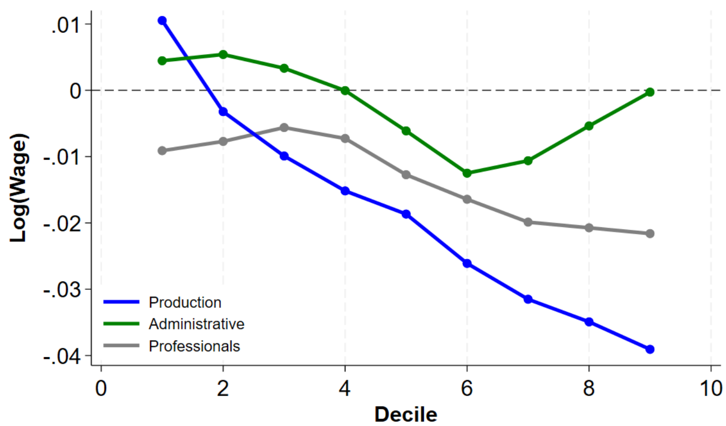
Notes: This table reports 2SLS estimates of the effect of AI on labor market outcomes, allowing the parameter of interest to vary by production, professional, and administrative occupations. Each column uses the AI ease-of-development instrument defined in Equation 6. In the first stage, the instrument is also interacted with a dummy for each occupational group. Column 1 reports the effect of AI on log employment; column 2 on the log of average age; column 3 on the log of average years of education; column 4 on the log of average tenure at the current firm (in years); column 5 on the log of average wage; and column 6 on the log of the standard deviation of wages. Controls include occupation fixed effects, one-digit-occupation-by-year fixed effects, and exposure to U.S. AI patents. Standard errors, clustered at the occupation level, appear in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 6 shows that AI shrinks the office and expands the factory. In production occupations, AI increases employment among younger, less experienced, and lower-skilled workers.

This compositional shift lowers average wages and reduces within-occupation inequality, consistent with increased hiring of lower-wage workers. These effects are quantitatively meaningful: a one–standard deviation increase in AI exposure raises employment among production workers by 4.3% and lowers their average wage by 3%. For comparison, the effect of AI on employment of professionals is 4 times smaller and not statistically significant.

In contrast, in administrative occupations, AI reduces employment without significantly changing workforce composition. A one-standard deviation increase in AI exposure decreases employment of administrative workers by 8% without any effect on average wages.

Figure 14: **In Production, AI Compresses the Wage Distribution**



Notes: This figure shows the effect of AI across different points in the wage distribution by production, administrative, or professional workers. To construct this figure, I compute wage-deciles within each occupation over the sample period. The figure plots the estimates of the effect of AI on the log of each wage decile allowing the effect of AI to differ by production, administrative, and professionals. All specifications use the AI ease-of-development instrument defined in Equation 6. Controls include occupation fixed effects, one-digit-occupation-by-year fixed effects, and exposure to U.S. AI patents. Standard errors are clustered at the occupation level. The bands reports the 95% confidence intervals.

Employment gains in production is skewed toward low-wage workers. Figure 14 shows the effect of AI across deciles of the wage distribution, as in Figure 12.[47] Unlike before, I allow the effect of AI to be different for production, administrative, and professional workers.

For production workers, AI increases wages at the bottom of the distribution and decreases them at the top. A one–standard–deviation increase in AI exposure raises wages in the bottom decile by 1%, while wages in the ninth decile fall by 4%. This pattern re-

[47] To improve readability, confidence intervals are omitted from the figure.

inforces the earlier finding that AI shifts the composition of production workers toward low-productivity employees. As a result, average wages fall, as shown in Table 6, because low-wage hiring expands by more.

This result suggests that, in production, AI acts as a substitute for skill: it enables low-wage, low-experience workers to perform tasks that previously required more training or expertise. As AI lowers the skill threshold needed for these jobs, less-qualified workers become more productive and are hired in greater numbers, increasing wages at the bottom of the distribution. At the same time, the value of experience and expertise decreases, leading to wage declines for high-ability workers whose skills are partially replaced by AI.

The effect of AI on the wage distribution is different for administrative workers and professionals. In administrative occupations, AI raises wages at the bottom of the distribution, lowers them in the middle, and has no discernible effect at the top. Among professionals, by contrast, AI reduces wages across the entire distribution. These patterns suggest that the deskilling mechanism observed among production workers is not present in administrative or professional roles.

AI increases more employment of workers operating machines. Figure 15 shows how the effect of AI varies with occupational characteristics. To estimate this heterogeneity, I augment the baseline regression with an interaction term between AI exposure and an occupational trait. The figure plots the coefficient on the interaction term, capturing how the effect of AI changes as the characteristic varies.[48]

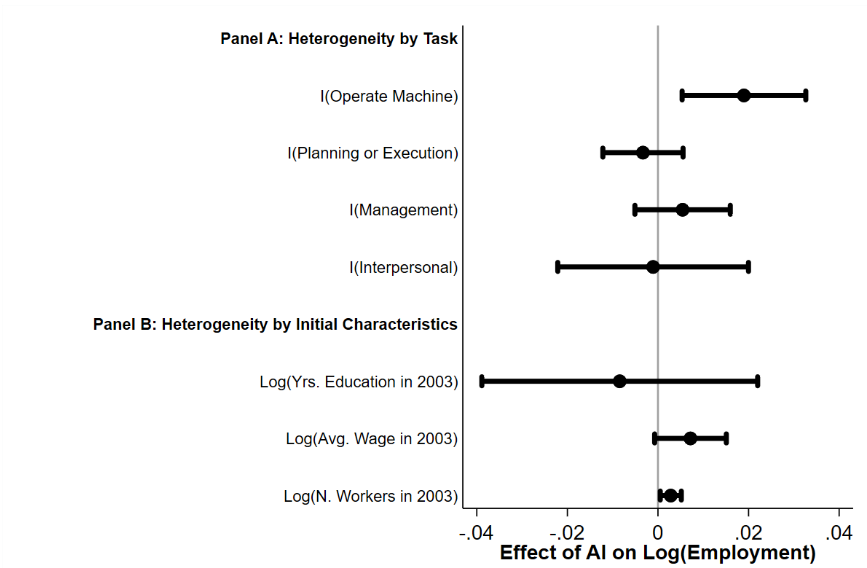
Figure 15 shows that occupations involving machine operation experience the largest employment gains from AI. These gains coincide with a shift toward younger, less-educated workers, as AI adoption reduces both average education and age among machine operators.[49] This compositional change likely contributes to the decline in average wages observed in Figure 16 among machine operators. One interpretation for these results is that AI makes

[48] In practice, I estimate a model given by $y_{o,t} = \beta AI\ Exposure_{o,t} + \beta_x^{hetero} AI\ Exposure_{o,t} \times x_o + \mu_o + \mu_t + X'_{o,t}\theta + \epsilon_{o,t}$, where x_o is a time-invariant characteristic of occupation o. Figure 14 plots the parameter β_x^{hetero} for different heterogeneity measures x_o.

[49] Figures A20a and A20b in the appendix shows that occupations related to machine operation had the largest influx of young and uneducated workers.

machinery more productive and easier to operate, allowing firms to hire less-educated workers.

Figure 15: **AI Has Larger Employment Effects in Machine-Operating Jobs**

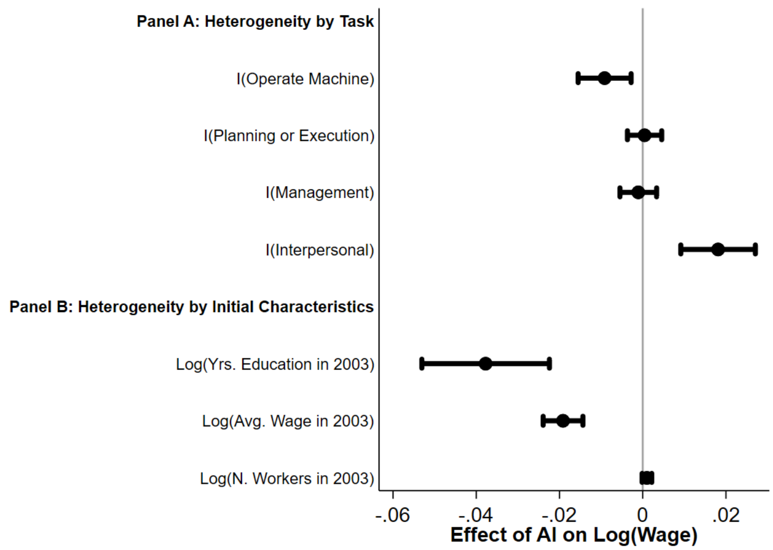


Notes: This figure reports heterogeneity in the effect of AI on employment based on different characteristics of each occupation. To construct it, I augment the baseline model by including an interaction term: $y_{o,t} = \beta AI\ Exposure_{o,t} + \beta^{hetero} x_o AI\ Exposure_{o,t} \times x_o + \mu_o + \mu_t + X'o, t\theta + \epsilon_{o,t}$, where x_o denotes an occupation-level characteristic. Panel A uses task-based indicators constructed from occupational descriptions: *I(Operate Machine)* equals one if the occupation involves operating machinery in sectors like manufacturing, agriculture, or construction; *I(Planning or Execution)* equals one for roles involving process planning or execution, typical of high-level managers; *I(Management)* takes one for occupations managing people or information; *I(Interpersonal)* identifies occupations requiring interpersonal interactions, such as customer service or supervision; and *I(Creation and Innovation)* takes one for creative or innovation-related work. Panel B calculates heterogeneity using characteristics of the occupation measured at the beginning of the period. All specifications use the AI ease-of-development instrument defined in Equation 6. Controls include occupation fixed effects, one-digit-occupation-by-year fixed effects, and exposure to U.S. AI patents. Shaded areas represent 95% confidence intervals. Standard errors are clustered at the occupation level.

AI decreases inequality across occupations. Figure 16 shows that AI reduces wages more in occupations that had higher average education and wages in 2003. Figures A20a and A20b in the Appendix show that these same occupations experienced larger declines in average education and age in response to AI. This suggests that AI enables younger and less-educated workers to enter roles once dominated by high-skilled individuals, contributing to a reduction in wage inequality across occupations. Therefore, AI decreases inequality within and across occupations.

Discussion: AI automates the office and expands the factory. These results show that AI affects administrative and production workers differently. The decline in administrative employment is consistent with the view that AI automates routine tasks traditionally performed by office workers (Acemoglu and Restrepo 2018, Webb 2020, Hampole et al. 2025). In contrast, employment gains in production—especially among younger, less experienced, and less educated workers—suggest that AI makes machines more productive and easier to operate, lowering the skill barrier to entry. AI has an overall positive effect on employment because the gains in production occupations outweighs losses in administrative jobs.

Figure 16: **AI Decreases Inequality Across Occupations**



Notes: This figure reports heterogeneity in the effect of AI on employment based on different characteristics of each occupation. To construct it, I augment the baseline model by including an interaction term: $y_{o,t} = \beta AI\ Exposure_{o,t} + \beta^{hetero} x_o AI\ Exposure_{o,t} \times x_o + \mu_o + \mu_t + X'o, t\theta + \epsilon_{o,t}$, where x_o denotes an occupation-level characteristic. Panel A uses task-based indicators constructed from occupational descriptions: *I(Operate Machine)* equals one if the occupation involves operating machinery in sectors like manufacturing, agriculture, or construction; *I(Planning or Execution)* equals one for roles involving process planning or execution, typical of high-level managers; *I(Management)* takes one for occupations managing people or information; *I(Interpersonal)* identifies occupations requiring interpersonal interactions, such as customer service or supervision; and *I(Creation and Innovation)* takes one for creative or innovation-related work. Panel B calculates heterogeneity using characteristics of the occupation measured at the beginning of the period. All specifications use the AI ease-of-development instrument defined in Equation 6. Controls include occupation fixed effects, one-digit-occupation-by-year fixed effects, and exposure to U.S. AI patents. Shaded areas represent 95% confidence intervals. Standard errors are clustered at the occupation level.

6 Robustness

In the Section 5, I showed that AI increases employment among young, low-education, and inexperienced workers, which in turn lowers average wages due to a change in the composition of workers. In this section, I show that this result is robust to several tests: controlling for U.S. labor market outcomes; controlling for non-AI software; using an instrument orthogonal to non-AI software exposure; running regressions at the sector-occupation level; constructing the instrument using pre-period AI software; weighting regressions by initial employment;

and removing controls.

Control for Labor Market Outcomes in the U.S. One concern is that trends in U.S. programming languages may reflect global labor market shocks that also affect Brazil, creating spurious correlations. To address this, Panel A of Table 7 controls for the same labor market outcome used on the left-hand side, but measured in the U.S. This adjustment accounts for common global shocks to employment, education, experience, and wages. The results are unchanged: AI continues to increase employment among younger, less educated, and less experienced workers, and to lower average wages.

Control for non-AI software. To ensure that results isolate the effect of AI software from general software adoption, Panel B of Table 7 controls for exposure to non-AI software, constructed similarly to the AI measure. The results still show that AI increases employment and shifts the workforce toward lower-skilled workers. However, due to larger standard errors, the employment effect is not statistically significant immediately and becomes significant only after three years.

AI exposure orthogonal to non-AI software. Another way to ensure that the results are not driven by exposure to software in general is to construct an instrument based on variation in AI exposure that is orthogonal to exposure to non-AI software. To do this, I first compute the average similarity between each occupation and the technical classes of non-AI software. I then regress $\overline{\textit{AI Exposure}}_{p,o}$ on $\overline{\textit{Non-AI Exposure}}_{p,o}$ and use the residual from this regression to construct the instrument. This residual isolates the component of AI exposure unrelated to general software exposure. Panel C of Table 7 presents the results using this alternative instrument. The conclusions remain unchanged: AI increases employment and shifts the composition of the workforce toward lower-skilled workers. However, the effect on wages is no longer statistically significant due to larger standard errors but it is still negative.

Regression at the occupation-industry, as in Webb (2020). In Panel D, I re-estimate the baseline specification at the occupation–industry level. Occupations are defined using the 6-digit CBO classification, and industries follow the 4-digit CNAE (*Classificação Nacional* *de Atividades Econômicas*) 2 classification. Given the narrowness of the Brazilian occupational codes, this approach introduces substantial sparsity, with many occupation–industry cells missing from the data. To account for sector-specific shocks, I include industry-year fixed effects and occupation–industry fixed effects to the baseline controls. Despite these adjustments and the resulting loss in precision, Panel D shows that the main conclusions remain robust.

Instrument with Pre-2003 AI Software. In the baseline instrument described in 6, I use lagged AI software to compute each occupation's exposure to different AI technical classes. This approach allows me to capture the evolving range of AI applications that emerged during the period of analysis. In Panel E of Table 7, I instead construct the instrument using only AI software created before 2003. The results remain qualitatively unchanged. The effect of AI on employment is not statistically significant due to larger standard error, but the coefficient is still positive.

Control for trends in labor market outcomes. Figure 11b shows a slight pre-trend in wages. To account for potential differential trends across occupations, Panel F of Table 7 augments the baseline controls by including three years of lagged values of the dependent variable. The results remain robust: I still find that AI increases employment, decreases wages, and shifts the composition of the workforce toward younger, less-educated, and less-experienced workers.

Employment-weighted regressions. Panel G of Table 7 reports results from baseline regressions weighted by 2003 employment. I, one more time, find that AI increases employment among young, low-education, and inexperienced workers, which in turn lowers average wages due to a change in the composition of workers.

Regression without controls. Finally, Panel H of Table 7 presents results from regressions that exclude both the lagged dependent variable and controls for exposure to U.S. AI software. The conclusions are still the same.

Table 7: **Robustness: AI Increases Employment of Low-Skilled Workers**

	(1) *log(Employment)*	(2) *log(Age)*	(3) *log(Yrs. Education)*	(4) *log(Experience)*	(5) *log(Wage)*	(6) *log(Std. Wage)*
Panel A: Control for U.S. Labor Market Outcome						
AI Exposure	0.0285^{***}	-0.00667^{***}	-0.00362^{***}	-0.00497^{***}	-0.0180^{***}	-0.00383
	(0.00889)	(0.00131)	(0.00178)	(0.00448)	(0.00442)	(0.0113)
Panel B: Control for Non-AI Software						
AI Exposure	0.0150	0.00718^{**}	-0.0114^{***}	-0.00304	-0.0229^{**}	-0.0442
	(0.0245)	(0.00339)	(0.00305)	(0.0175)	(0.0111)	(0.0312)
Panel C: Residual AI Exposure						
AI Exposure	0.0409^{***}	-0.00478^{**}	-0.00729^{***}	-0.0155	-0.0103	-0.0204
	(0.0117)	(0.00196)	(0.00174)	(0.0103)	(0.00634)	(0.0147)
Panel D: Sector-Occupation Level						
AI Exposure	0.0350^{***}	-0.00519^{***}	-0.00484^{***}	-0.00341^{*}	-0.0115^{***}	0.0156^{***}
	(0.00178)	(0.000351)	(0.000386)	(0.00180)	(0.000939)	(0.00544)
Panel E: Instrument with Pre-2003 AI Exposure						
AI Exposure	0.0134	-0.00357^{***}	-0.00602^{**}	-0.0225^{***}	-0.0138^{***}	-0.0279^{**}
	(0.00867)	(0.00135)	(0.00124)	(0.00692)	(0.00449)	(0.0109)
Panel F: Control for Trends						
AI Exposure	0.0200^{***}	-0.00251^{***}	-0.00314^{***}	-0.0107^{***}	-0.0101^{***}	-0.00958
	(0.00617)	(0.000773)	(0.000718)	(0.00369)	(0.00247)	(0.00690)
Panel G: Employment Weighted						
AI Exposure	0.0294^{***}	0.00264^{***}	-0.00223^{**}	-0.0164^{***}	-0.0207^{***}	-0.0403^{***}
	(0.00947)	(0.000774)	(0.00112)	(0.00543)	(0.00411)	(0.0110)
Panel H: No Controls						
AI Exposure	0.0428^{***}	-0.00558^{***}	-0.0116^{***}	-0.0205^{***}	-0.0184^{***}	-0.0119^{*}
	(0.00519)	(0.000792)	(0.00108)	(0.00379)	(0.00222)	(0.00614)

Notes: This table reports robustness checks of the effects of AI software exposure on labor market outcomes, following the specification in Section 2. Panel A controls for the U.S. version of each outcome variable to account for global labor market shocks. Panel B adds as control exposure to non-AI software, constructed constructed using text similarity between the description of non-AI software and the description of each occupation's task. Panel C instruments AI exposure using variation orthogonal to non-AI software exposure. Panel D runs the baseline regression at the occupation–industry level, including industry-year and occupation–industry fixed effects. Panel E instruments AI exposure using only AI software developed before 2003 to construct the instrument. Panel F includes three lags of the dependent variable to control for pre-trends. Panel G reports employment-weighted regressions using 2003 employment as weights. Panel H excludes both the lagged dependent variable and the U.S. AI exposure control. Controls include occupation fixed effects, one-digit-occupation-by-year fixed effects, and exposure to U.S. AI software, unless otherwise noted. Standard errors, clustered at the occupation level, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

7 Conclusion

This paper studies the effect of artificial intelligence on the labor market using a novel dataset of software registrations from Brazil's National Institute of Industrial Property (NIIP). The dataset contains detailed information on AI software developed in Brazil, including its intended applications and technical characteristics. I construct an occupation-level measure of AI exposure by computing text similarity between software descriptions and occupational tasks. To address endogeneity in AI adoption, I use an instrumental variable strategy that leverages variation in software development costs driven by trends in programming language popularity.

I find that, on average, AI increases employment but lowers wages, driven by rising employment among younger, less-educated, and less-experienced workers. Within occupations, AI reduces wage inequality by disproportionately raising wages at the lower end of the distribution. 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, as emphasized by Acemoglu and Restrepo (2018), while also supporting with evidence that AI can complement labor and increase employment, as proposed by Agrawal et al. (2019) and Gruber et al. (2020).

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.

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.

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

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

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.

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 (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.

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.

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.

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 X. Wang (2023): "Information Technology, Firm Size, and Industrial Concentration," Working Paper 31065, National Bureau of Economic Research.

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

Brynjolfsson, E. and P. Milgrom (2012): "Complementarity in Organizations," in *The Handbook of Organizational Economics*, ed. by R. Gibbons and J. Roberts, Princeton University Press, chap. 1, 11–55.

Brynjolfsson, E., D. Rock, and C. Syverson (2021): "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies," *American Economic Journal: Macroeconomics*, 13, 333–72.

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.

Chauvet, J.-M. (2018): "The 30-Year Cycle In The AI Debate," .

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, Flávia (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.

de Souza, G. (2022): "The Labor Market Consequences of Appropriate Technology," Working Paper Series WP 2022-53, Federal Reserve Bank of Chicago.

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.

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

FOURMENT, M. AND M. R. GILLINGS (2008): "A comparison of common programming languages used in bioinformatics," *BMC Bioinformatics*, 9, 82.

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

——— (2017b): "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.

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.

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.

HARGUESS, J. AND C. M. WARD (2008): "Is the Next Winter Coming for AI? Elements of Making Secure and Robust AI," *IEEE Computer Society*.

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.

Hui, X., O. Reshef, and L. Zhou (2023): "The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market," Tech. rep.

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–.

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.

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.

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.

Kaplan, J., S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei (2020): "Scaling Laws for Neural Language Models," .

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.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012): "ImageNet Classification with Deep Convolutional Neural Networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Curran Associates Inc., 1097–1105.

Lee, K.-F. (2018): *AI Superpowers: China, Silicon Valley, and the New World Order*, Boston, MA: Houghton Mifflin Harcourt.

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.

Marr, B. (2019): *Artificial Intelligence in Practice: How 50 Successful Companies Used AI and Machine Learning to Solve Problems*, Chichester, UK: Wiley.

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., 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.

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.

Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013): "Efficient Estimation of Word Representations in Vector Space," .

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

Mitchell, M. (2020): *Artificial Intelligence: A Guide for Thinking Humans*, New York: Picador, first picador paperback edition ed.

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," .

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," .

Raina, R., A. Madhavan, and A. Y. Ng (2009): "Large-scale deep unsupervised learning using graphics processors," in *Proceedings of the 26th Annual International Conference on Machine Learning*, New York, NY, USA: Association for Computing Machinery, ICML '09, 873–880.

Samuelson, P. (2017): "Functionality and Expression in Computer Programs: Refining the Tests for Software Copyrightability and Patentability," *Berkeley Technology Law Journal*, 32, 487–543.

Santos, J. W. d., J. Barreto Netto, C. Monteiro de Farias Rezende, A. K. Abud, M. Santos, and C. Dias (2022): *Perfil dos registros de programas de computador voltados ao enfrentamento à pandemia de COVID-19 no Brasil*, 77–90.

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

Sevilla, J., L. Heim, A. Ho, T. Besiroglu, M. Hobbhahn, and P. Villalobos (2022): "Compute Trends Across Three Eras of Machine Learning," in *2022 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 1–8.

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.

SHOJAEINASAB, A. ET AL. (2022): "Intelligent Manufacturing Execution Systems: A Systematic Review," *Journal of Manufacturing Systems*, 62, 503–522.

SILVA, T. A. T. D. H., J. P. L. SANTOS, AND S. B. B. UCHÔA (2025): "Prospecção tecnológica sobre inteligência artificial em programas de computador registrados no INPI," *Revista P2P & Inovação*, 11, 1–17, licença CC BY-NC-SA 4.0.

SOLOW, R. (1987): "We'd Better Watch Out," *New York Times Book Review*.

STUBER, W. D. (1984): "Industrial Policy in the Field of Informatics in Brazil," *Michigan Journal of International Law*, 6, 303–.

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.

THE STANDISH GROUP (1995): "CHAOS Report," https://www.csus.edu/indiv/v/velianitis/161/chaosreport.pdf, accessed: 2025-06-01.

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

TOOSI, A., A. G. BOTTINO, B. SABOURY, E. SIEGEL, AND A. RAHMIM (2021): "A Brief History of AI: How to Prevent Another Winter (A Critical Review)," *PET Clinics*, 16, 449–469.

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

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.

Wijdan, A. (2020): "Comparison of Programming Languages in Game Development," .

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 Data

A.1 Developers use software registration to signal achievements

Figure A1: **LinkedIn posts by developers showcasing their NIIP software-registration certificates.**

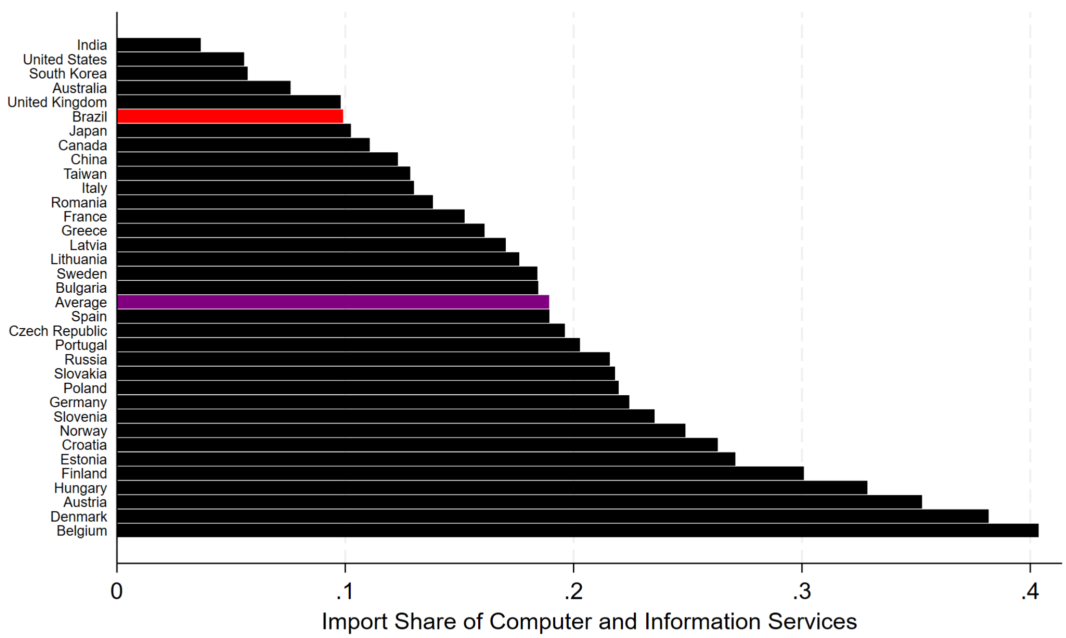
A.2 Most software used in Brazil is developed locally and therefore registered with the NIIP

Figure A2 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.[50] 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.

It could be the case that foreign IT firms set subsidiaries in Brazil. In that case, their payments to the headquarters would appears as royalties and not as licensing fees. In these cases, these firms would be require to register their software, a requirement that they might not be aware off. To test the importance of that, Figure 7 plots the share of multinationals among the IT firms in Brazil. It shows that only 1% of the IT companies in Brazil is a multinational, showing, again, that most software developed in Brazil is locally made.

Figure A2: **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 the import in 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.

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

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, reinforcing the database's near-census coverage.

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.

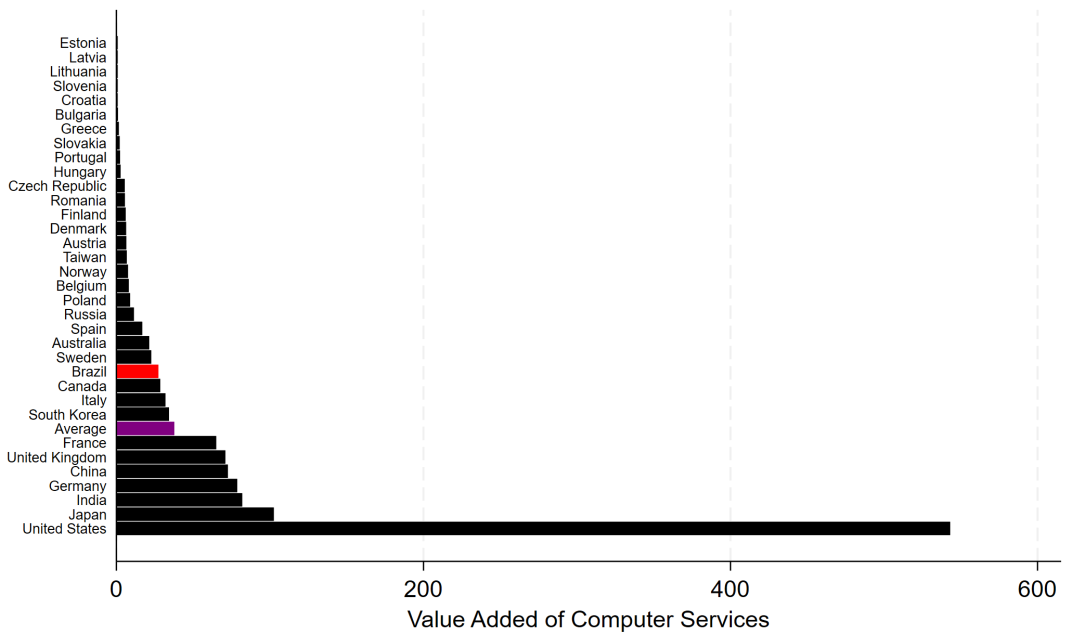
Three mutually reinforcing forces keep Brazil's import share of IT services exceptionally low. First, digital services tend to exhibit a strong home bias across countries. Alaveras and Martens (2015) shows that even in the United States, only the largest firms export online services, and that the preference 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 A3 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 actually used within Brazilian firms.

Figure A3: **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.

A.2.1 List of Broad Technical Classes and Application Domains

Table A2: 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.

Continued on next page

Code	Title	# Subcategories	Description
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.
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.

Continued on next page

Code	Title	# Subcategories	Description
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.
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 A3: 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: 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.

Continued on next page

Code	Title	# Subcategories	Description
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.
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.

Continued on next page

Code	Title	# Subcategories	Description
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.
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: Subcategories: Public Finance, Private Finance, Financial Systems, Budgeting and Financial Instruments, Financial Management, Accounting.

Continued on next page

Code	Title	# Subcategories	Description
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.
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.

Continued on next page

Code	Title	# Subcategories	Description
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.
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.

Continued on next page

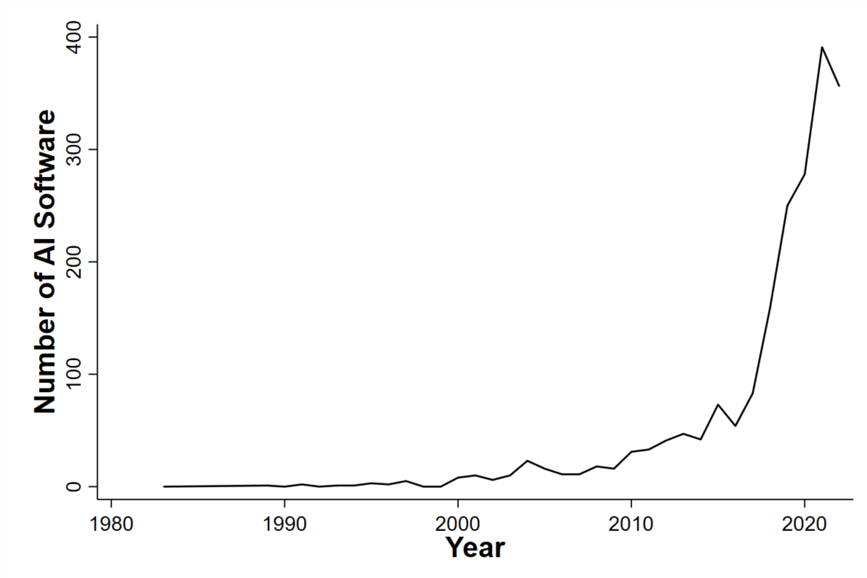
Code	Title	# Subcategories	Description
SV	Services	4	The Services application domain includes software that supports a wide array of public and private service activities essential to social and economic life. It encompasses utilities such as electricity, water, and telecommunications, as well as services 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.

A.3 Summary Statistics of the Software Dataset

This section presents summary statistics for the software dataset. Figure A4 displays the number of AI software registrations over time. Consistent with the trends discussed in Section 3, there is a sharp increase in AI software creation beginning around 2013.

Figure A4: **Number of New AI Software Registrations Over Time**

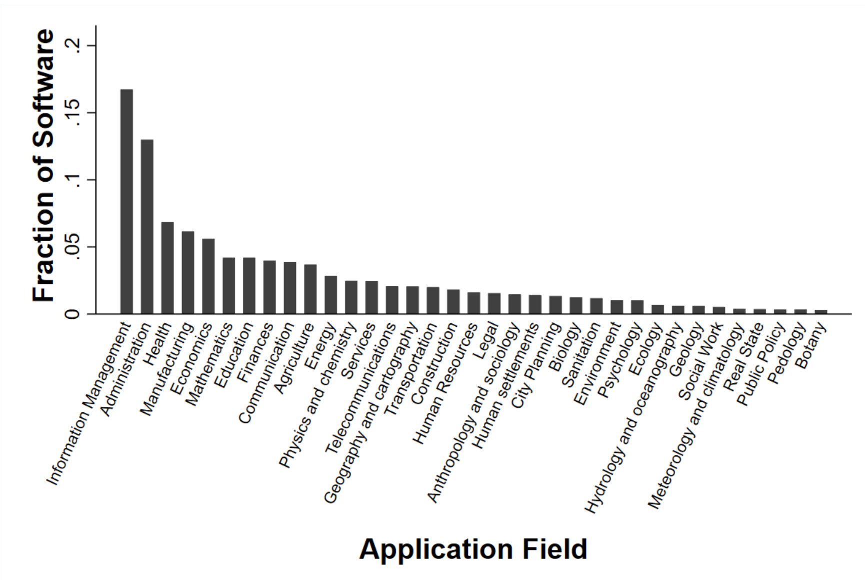


Notes: This figure plots the number of newly registered software classified as artificial intelligence (AI) in Brazil from 1987 to 2023.

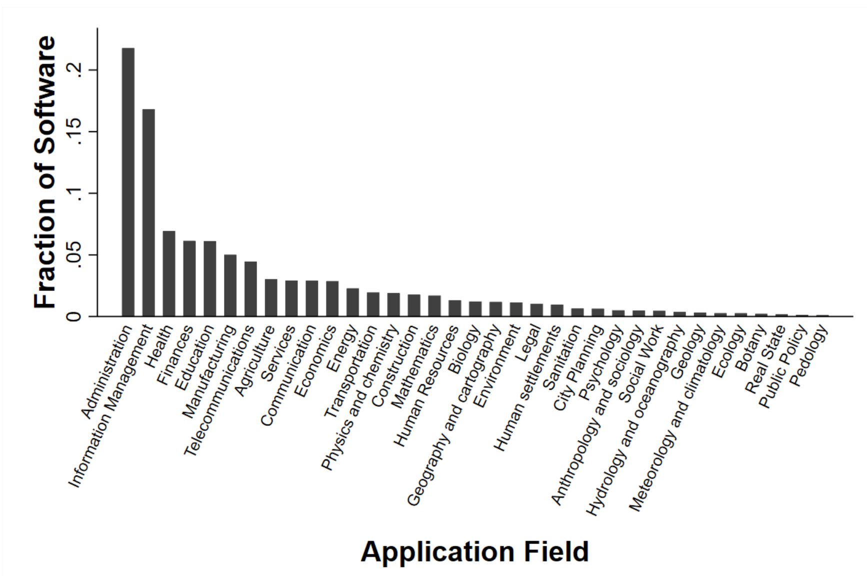
Figure A5 compares the distribution of application domains between AI and non-AI software. As shown in Figure A5b, non-AI software is more heavily concentrated in administrative and information management domains, whereas AI software is more broadly distributed across different areas.

Figure A5: **AI and Non-AI Application Domain**

(a) Fraction of Application Domain Among AI Software



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

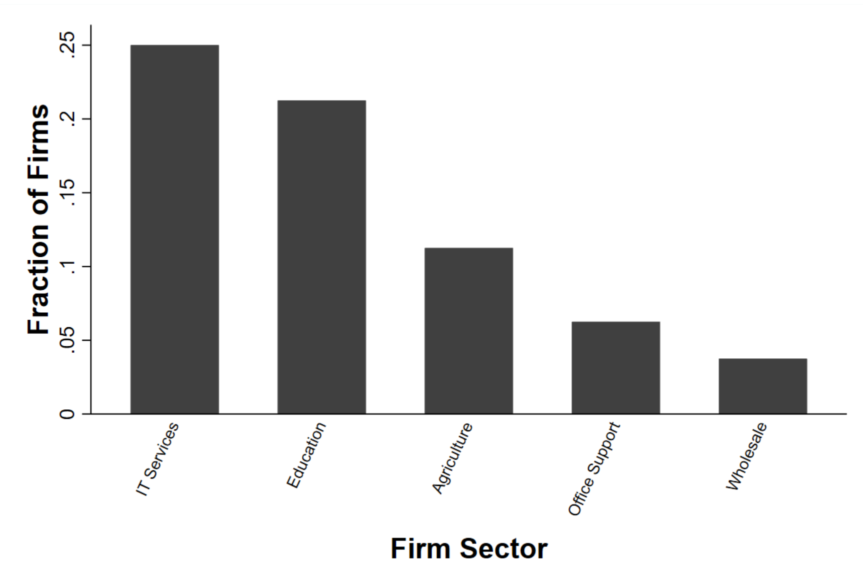


Notes: These figures present the distribution of AI software across different application domains, as defined by the NIIP. The application domain classification was developed by the NIIP to indicate the areas in which each software is intended to be used. Because a single software program may be assigned to multiple domains, each figure reports the fraction of each domain relative to the total number of domain classifications. Figure A8a plots the distribution of application domain among AI software while Figure A5b has the distribution of application domain among non-AI software.

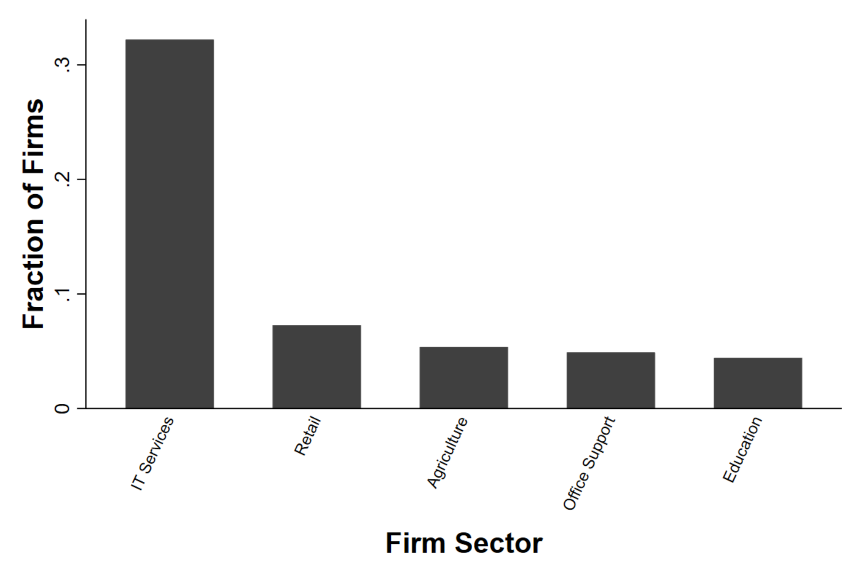
Figure A8 displays the distribution of AI and non-AI software across firm sectors. Among AI software, the most intensive sectors are Information Technology (IT) and Education, which includes universities and research institutions. In contrast, non-AI software is also concentrated in the IT sector but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

Figure A6: **Sectoral Distribution of Firms Creating AI and Non-AI Software**

(a) Sectoral Distribution of Firms Creating AI Software



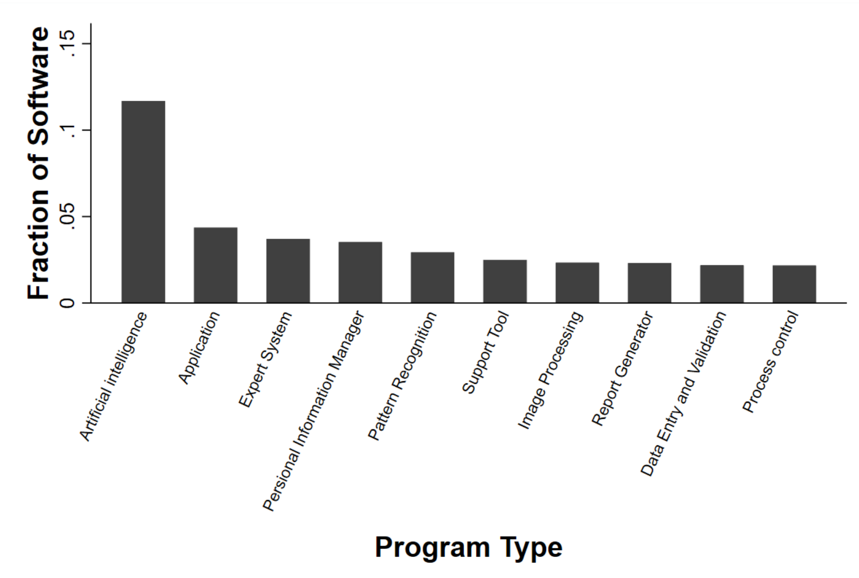
(b) Sectoral Distribution of Firms Creating non-AI Software



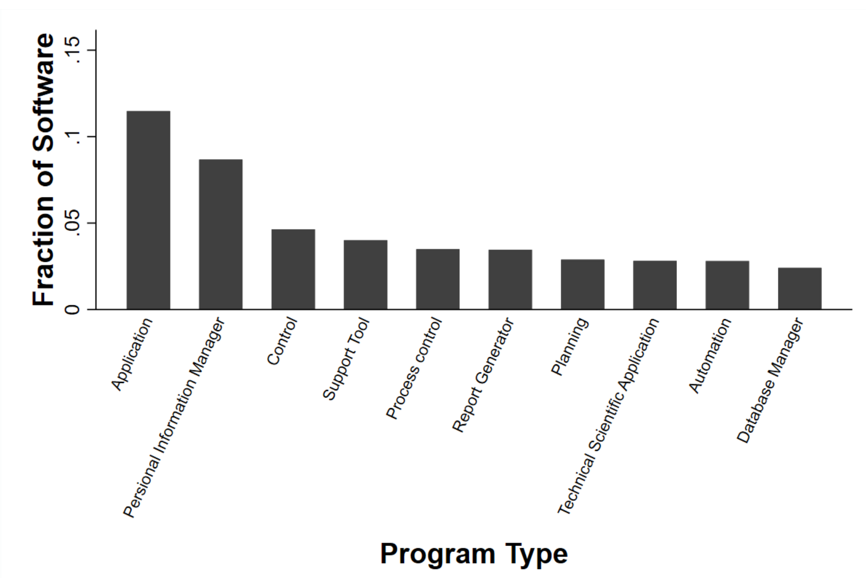
Notes: These figures show the distribution of firms registering AI and non-AI software across economic sectors.

Figure A7: **Fraction of Technical Class Among AI and Non-AI Software**

(a) Fraction of Technical Class Among AI Software



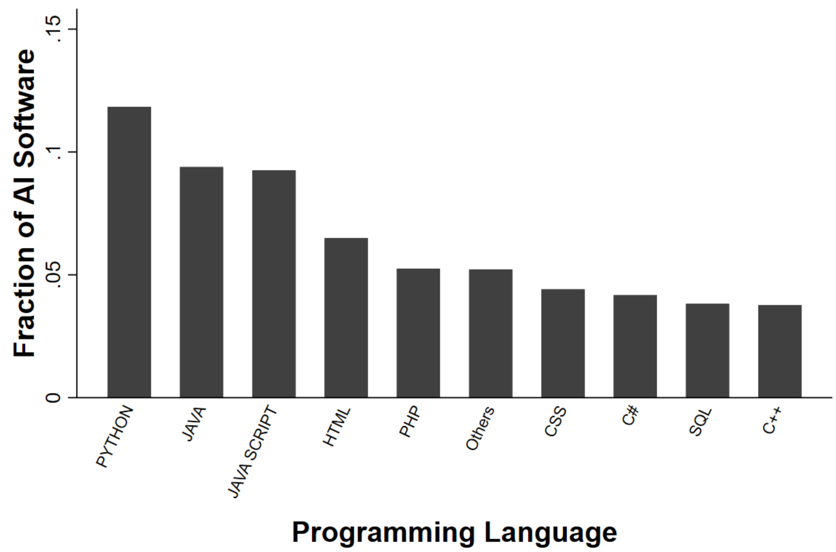
(b) Fraction of Technical Class Among non-AI Software



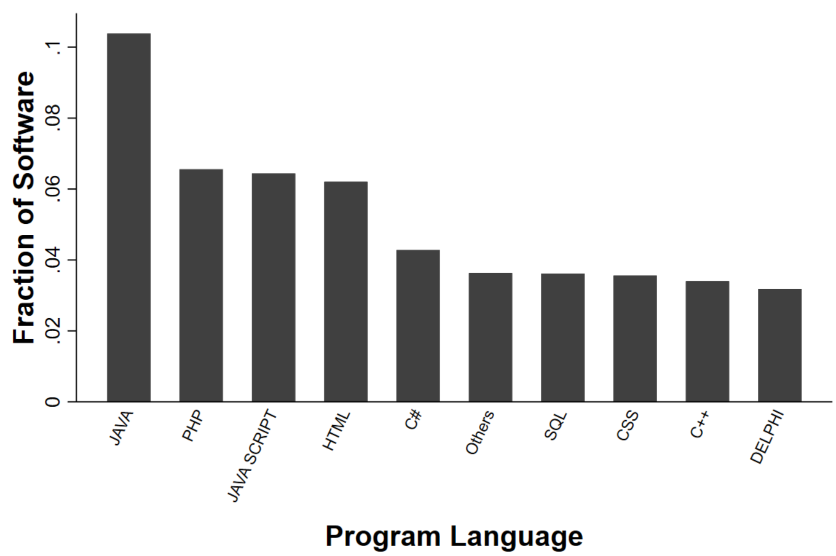
Notes: These figures show the distribution of technical class among AI and non-AI software.

Figure A8: **Fraction of Different Programming Languages Among AI and Non-AI Software**

(a) Fraction of Programming Language Among AI Software



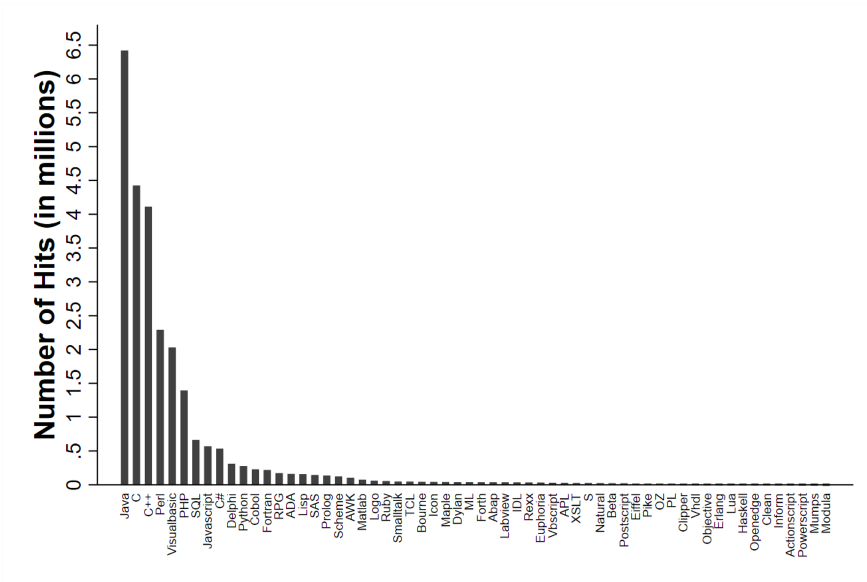
(b) Fraction of Programming Language Among non-AI Software



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

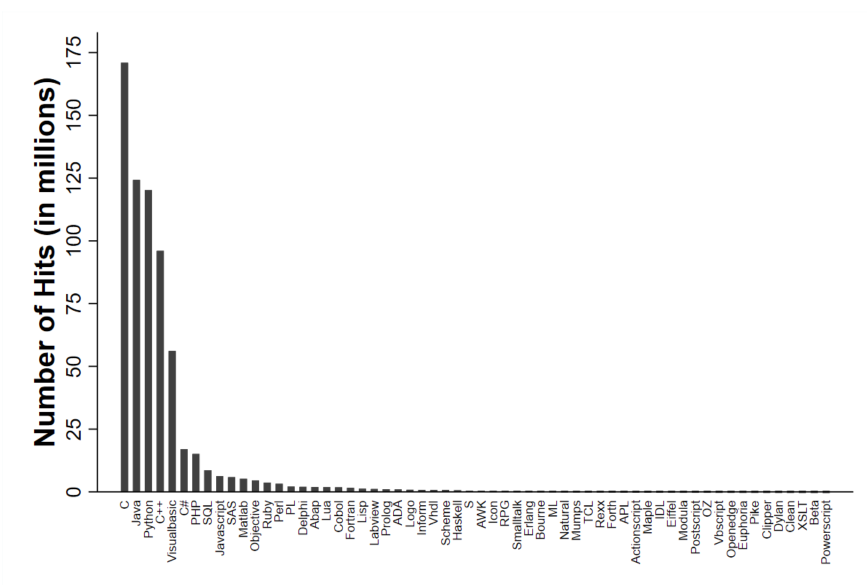
A.4 Summary Statistics of the Tiobe Index

Figure A9: Distribution of tiobe index for programming languages in 2003



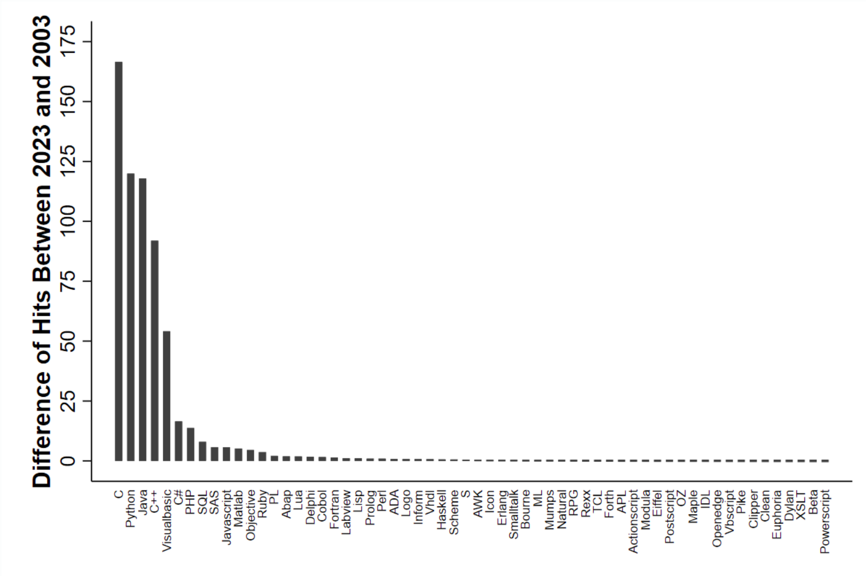
Notes: This figure shows the number of hits in Google.com in 2003 for the top programming languages using data from Tiobe.

Figure A10: Distribution of tiobe index for programming languages in 2023



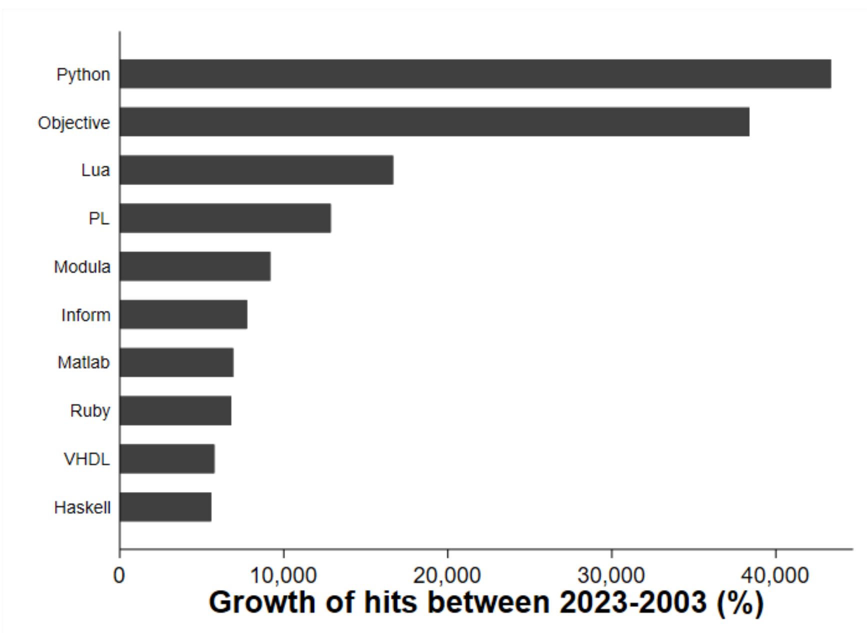
Notes: This figure shows the number of hits in Google.com in 2023 for the top programming languages using data from Tiobe.

Figure A11: Distribution of the difference in tiobe index (2023-2003)



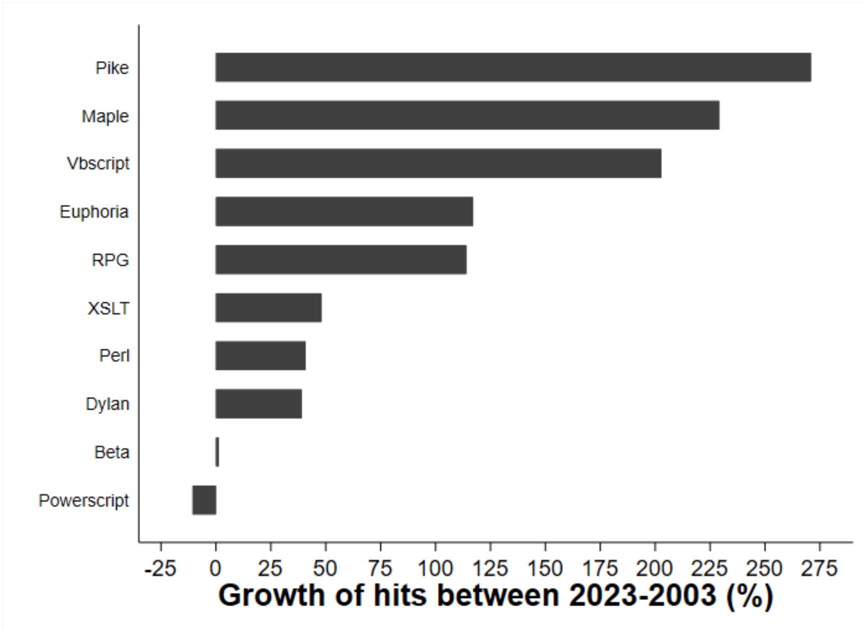
Notes: This figure shows the change in number of hits in Google.com between 2023 and 2003 for the top programming languages using data from Tiobe.

Figure A12: Top 10 programming languages with largest growth



Notes: This figure plots the programming languages with largest growth rate in number of hits in Google.com between between 2023 and 2003 using data from Tiobe.

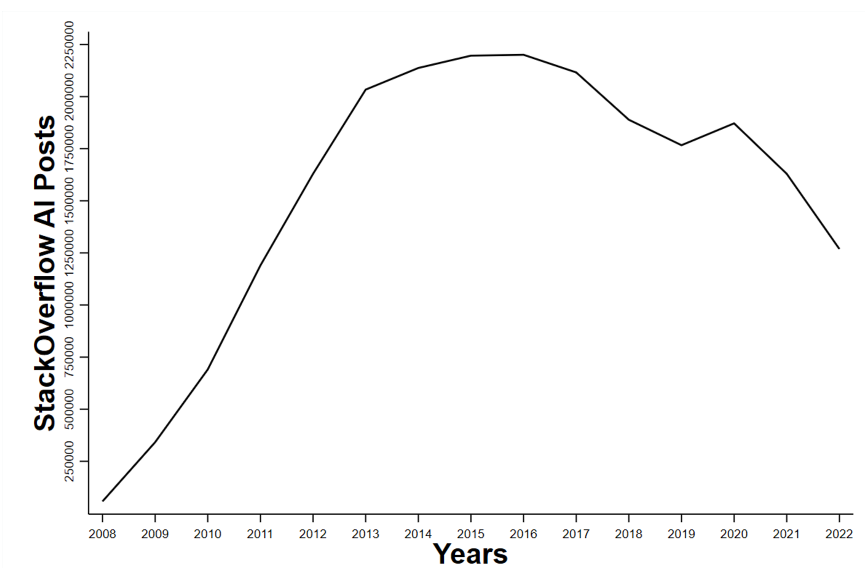
Figure A13: Top 10 programming languages with smallest growth



Notes: This figure plots the programming languages with smallest growth rate in number of hits in Google.com between between 2023 and 2003 using data from Tiobe.

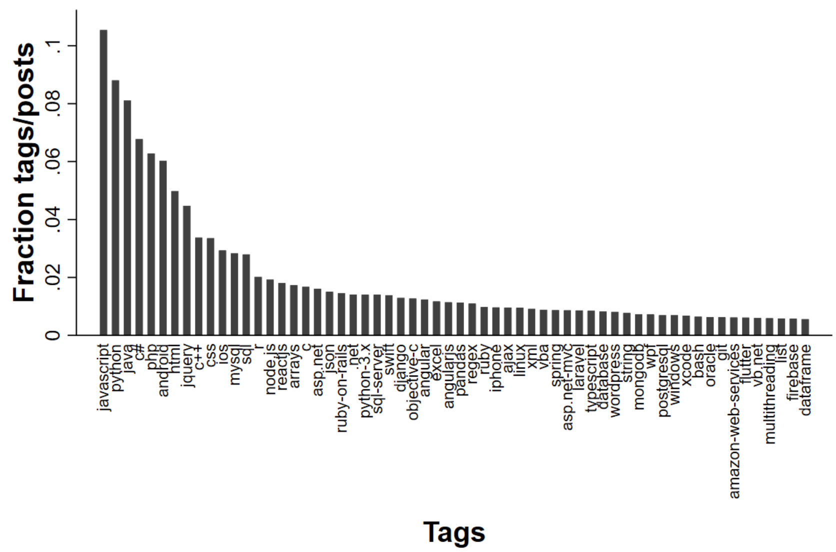
A.5 Statistics of Stack Overflow Posts

Figure A14: Number of stack overflow posts per year



Notes: This figure plots the number of new stack overflow posts over time.

Figure A15: 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.

Table A4: Top 3 tags Stack overflow for TOP 20 programming languages

Language	First Tag	Second Tag	Third Tag
JavaScript	jQuery	HTML	CSS
HTML	CSS	Twitter Bootstrap	Angular
Java	Android	Spring	Swing
PHP	MySQL	Laravel	HTML
Python	Pandas	Python-3.x	Django
C#	.NET	ASP.NET	WPF
iOS	Swift	Objective-C	Xcode
SQL	SQL Server	Oracle	PostgreSQL
MySQL	SQL	Database	Join
Excel	VBA	Excel Formula	Excel 2010
jQuery	HTML	CSS	AJAX
Android	Android Layout	Android Studio	Kotlin
R	ggplot2	DataFrame	dplyr
C++	C++11	Qt	C
VBA	Excel	MS Word	MS Access
Angular	TypeScript	Angular Material	RxJS
Linux	Bash	Shell	Unix
C	Linux	Pointers	Arrays
Git	GitHub	Version Control	Bitbucket
Django	Django Models	Django REST Framework	Django Views

A.6 Comparative Advantage 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
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

Continued on next page

Table A5 – continued from previous page

Technical Class	**Main Language**	**Share**
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
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

Continued on next page

Table A5 – continued from previous page

Technical Class	Main Language	Share
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
Operational Research	python	.359
Pattern Recognition	python	.439
Image Processing	python	.427
Teleinformatics	javascript	.364
Terminals	sql	.450
Data Transmission	javascript	.479

Continued on next page

Table A5 – continued from previous page

Technical Class	Main Language	Share
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

B Empirics

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

	(1)	(2)	(3)
	AI $Exposure_{o,2022}$	AI $Exposure_{o,2022}$	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 perform tasks related to planning or execution, $\mathbb{I}\{Creation\ and\ Innovation\}$ is a dummy taking one if the occupation perform tasks related to creation or innovation, $\mathbb{I}\{Management\}$ is a dummy if the occupation perform tasks related to management, and $\mathbb{I}\{Interpersonal\}$ is a dummy taking one if the occupation perform tasks related to interpersonal relationships. Standard errors in parentheses.Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: **Programming-language popularity and software development**

	(1)	(2)	(3)	(4)
	N. AI Software using language l	*N. AI Software using language l*	*Shr. AI Software using language l*	*Shr. AI Software using language l*
$\frac{\textit{hits language } l \textit{ in year } t}{\textit{hits all languages in year } t}$	88.59***		0.358***	
	(7.434)		(0.0239)	
$\frac{\textit{hits language } l \textit{ in year } t-2}{\textit{hits all languages in year } t-2}$		38.17***		0.395***
		(8.405)		(0.0238)
N	6 027	6 027	5 740	5 740

Notes: This table reports coefficients from an OLS regression relating each programming language's relative popularity in the United States to the number of AI software applications created in Brazil using that language. Popularity is the share of Google search hits for language l relative to all languages, measured contemporaneously (columns 1 and 3) or with a two-year lag (columns 2 and 4). Standard errors clustered by language appear in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B.1 Text Similarity

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

Parsing. I transform documents into a vector using words as tokens, i.e., 1-gram.

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 $m \in \{1, ..., M\} = \mathcal{M}$ be the set for words in the documents. 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 decompose 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 than zero.

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

$$s_{so} = \sum_{m \in \mathbb{M}} \tilde{f}_{sm} \times \tilde{f}_{om}. \tag{7}$$

B.2 Comparison to existing exposure metrics

Table A8 shows that the exposure measure 1 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 (2017b) 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, as reflected by software that has been commercialized and deployed.

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 (2017b), are negative correlated while the remaining pairwise correlation does not exceed 0.417.

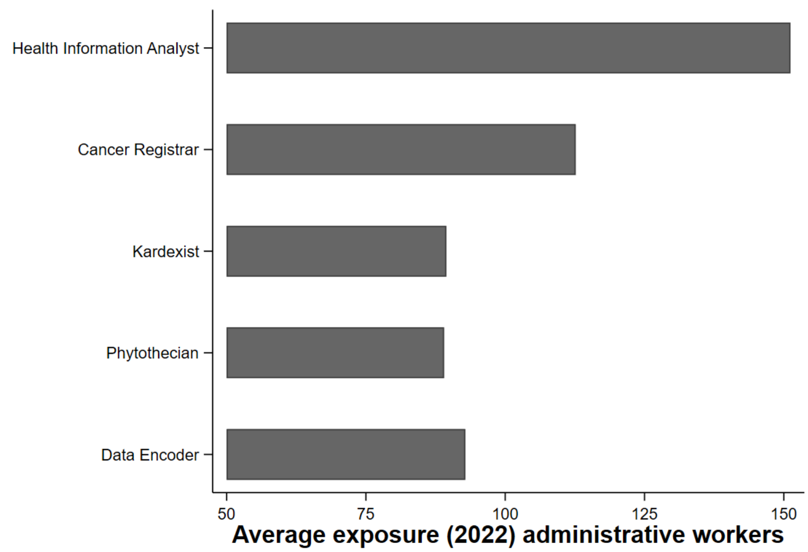
Table A8: **Correlation Between Different AI Exposure Metrics**

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 (2017b)	-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 1 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.

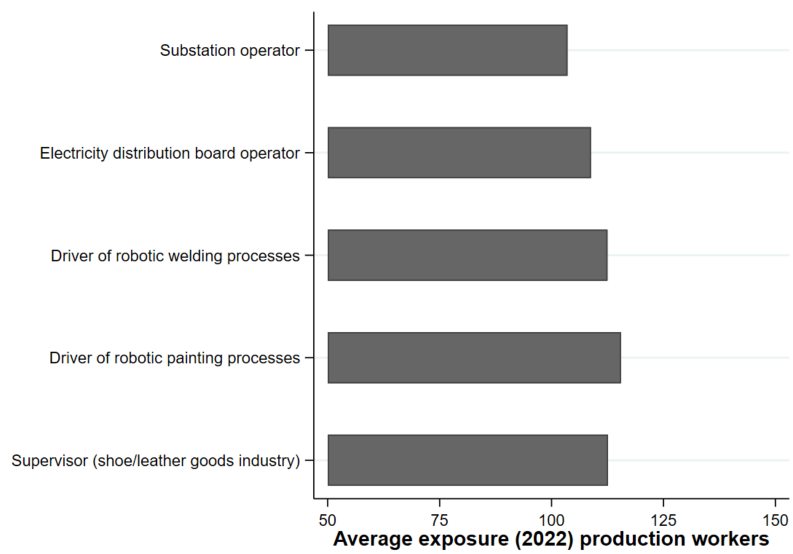
B.3 Summary Statistics of the Exposure to AI

Figure A16: Top 5 occupations most exposed within administrative workers



Notes: This figure shows the top 5 6-digit occupations most exposed to AI within administrative workers.

Figure A17: Top 5 occupations most exposed within production workers



Notes: This figure shows the top 5 6-digit occupations most exposed to AI among production workers.

B.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 A18: **Distribution of Time to Build Software in Years**

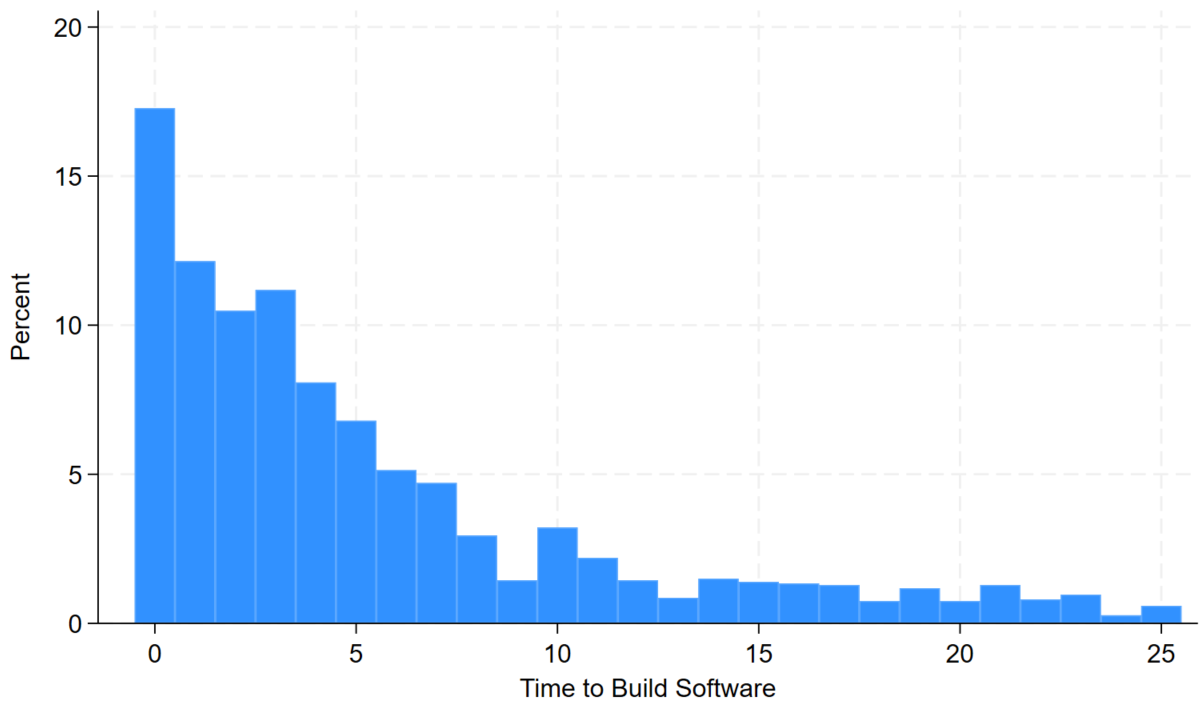


Table A9: **Time to Build Software**

	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.

B.5 Mincer Regression to Identify Abilities

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

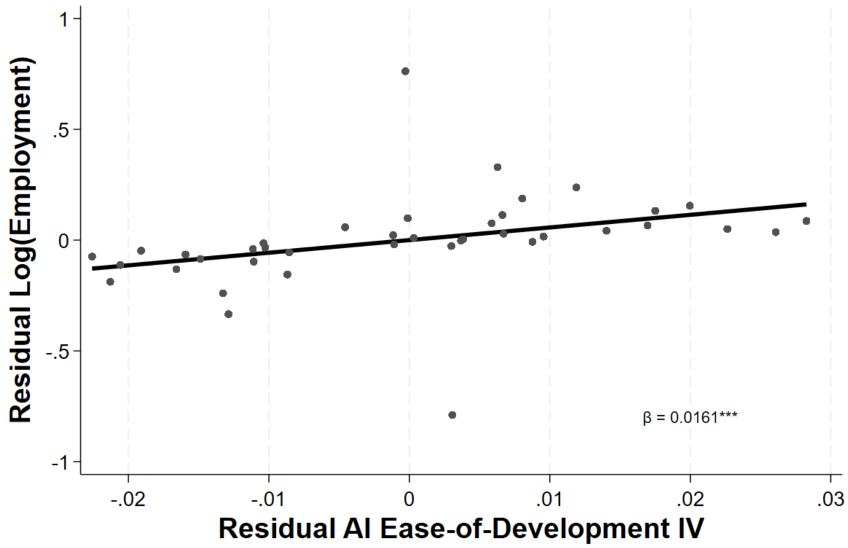
$$log(Wages_{i,t}) = \mu_i + X_{i,t} + \epsilon_{i,t}$$

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 5, 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.

B.6 Other Results

Figure A19: **Correlation Between Employment and the Instrument**



Notes: This figure plots the residual of log employment against the residual of the instrument. Observations are binned in 40 points.

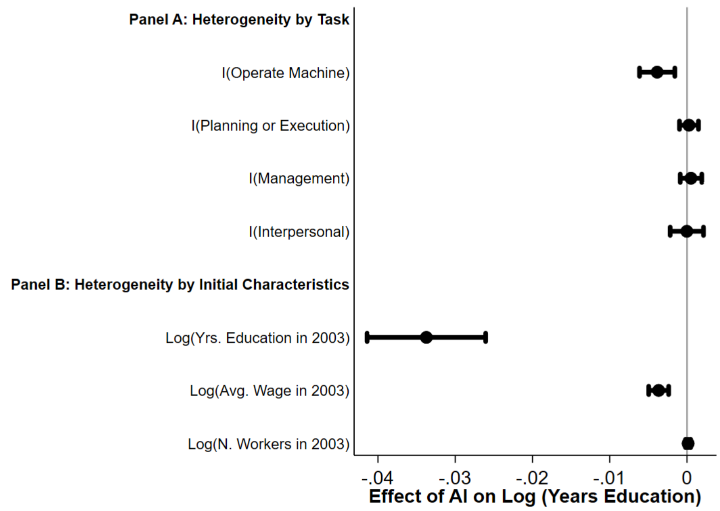
Table A10: **Effect of AI by Gender, Age, Education, and Experience**

Group	(1) *log(Employment)*	(2) *log(Wage)*	(3) *Ability*
By age			
Young *(less than 35)*	0.0212^{***}	-0.0278^{***}	-0.0872^{***}
	(0.00801)	(0.00465)	(0.0156)
Old *(more than 35)*	0.0158^{**}	-0.0298^{***}	-0.0806^{***}
	(0.00733)	(0.00461)	(0.0185)
By education			
High School or more	-0.0000127	-0.0266^{***}	-0.0992^{***}
	(0.00736)	(0.00418)	(0.0189)
High School incomplete	0.0260^{***}	-0.0378^{***}	-0.0767^{***}
	(0.00901)	(0.00558)	(0.0181)
By experience			
$>$ *one year*	0.0175^{***}	-0.0272^{***}	-0.0874^{***}
	(0.00757)	(0.00453)	(0.0230)
$\leq$ *one year*	0.0331^{***}	-0.0408^{***}	-0.0899^{***}
	(0.0122)	(0.00632)	(0.0168)
By gender			
Male	0.0213^{***}	-0.0236^{***}	-0.0815^{***}
	(0.00721)	(0.00436)	(0.0171)
Female	0.0240^{**}	-0.0115^{**}	-0.0442^{**}
	(0.00957)	(0.00564)	(0.0196)

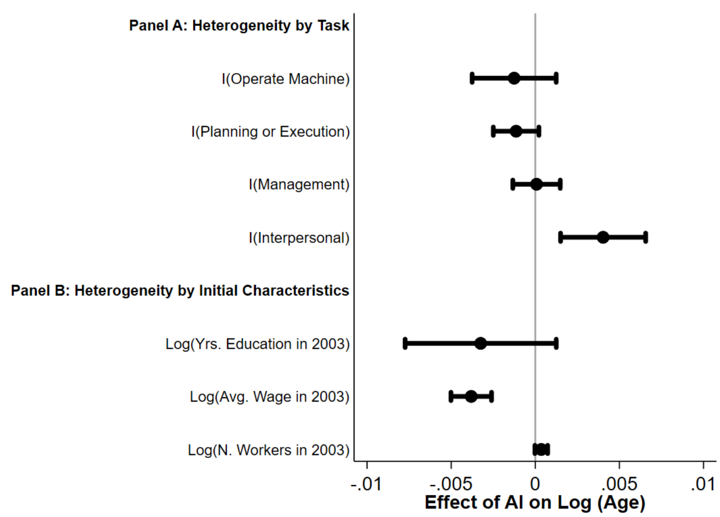
Notes: This table reports estimates of the effect of AI on employment and wages across demographic groups, following the specification in Equation 2. All specifications use the AI ease-of-development instrument defined in Equation 6. For each specification, the left-hand side variable is calculated among the sample described in *Group* column. Column 1 reports the effect of AI on log employment; column 2 on the log of average wages. Column 3 reports the effect of AI on productivity. 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. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Figure A20: **Heterogeneous Effect of AI According to Tasks and Education**

(a) Heterogeneous Effect of AI on Years of Education



(b) Heterogeneous Effect of AI on Age



Notes: This figure reports heterogeneity in the effect of AI based on different characteristics of each occupation. To construct it, I augment the baseline model by including an interaction term: $y_{o,t} = \beta AI\ Exposure_{o,t} + \beta^{hetero} x_o AI\ Exposure_{o,t} \times x_o + \mu_o + \mu_t + X'o, t\theta + \epsilon_{o,t}$, where x_o denotes an occupation-level characteristic. Figure 15 plots the interaction coefficient β_x^{hetero} using log employment as the outcome, while Figure 16 uses log wage. Panel A uses task-based indicators constructed from occupational descriptions: *I(Operate Machine)* equals one if the occupation involves operating machinery in sectors like manufacturing, agriculture, or construction; *I(Planning or Execution)* equals one for roles involving process planning or execution, typical of high-level managers; *I(Management)* takes one for occupations managing people or information; *I(Interpersonal)* identifies occupations requiring interpersonal interactions, such as customer service or supervision; and *I(Creation and Innovation)* takes one for creative or innovation-related work. Panel B calculates heterogeneity using characteristics of the occupation measured at the beginning of the period. All specifications use the AI ease-of-development instrument defined in Equation 6. Controls include occupation fixed effects, one-digit-occupation-by-year fixed effects, exposure to non-AI software, and exposure to U.S. AI patents. Shaded areas represent 95% confidence intervals. Standard errors are clustered at the occupation level.