DISCUSSION PAPER Nº IDB-DP- 01076

AI and the Increase of Productivity and Labor Inequality in Latin America:

Potential Impact of Large Language Models on Latin American Workforce

Oliver Azuara
Laura Ripani
Eric Torres

Inter-American Development Bank
Labor Markets and Social Security Division

September 2024

AI and the Increase of Productivity and Labor Inequality in Latin America:

Potential Impact of Large Language Models on Latin American Workforce

Oliver Azuara
Laura Ripani
Eric Torres

Inter-American Development Bank
Labor Markets and Social Security Division

September 2024

AI and the Increase of Productivity and Labor Inequality in Latin America

Potential Impact of Large Language Models on Latin American Workforce

Oliver Azuara Laura Ripani Eric Torres

September 3, 2024

Abstract

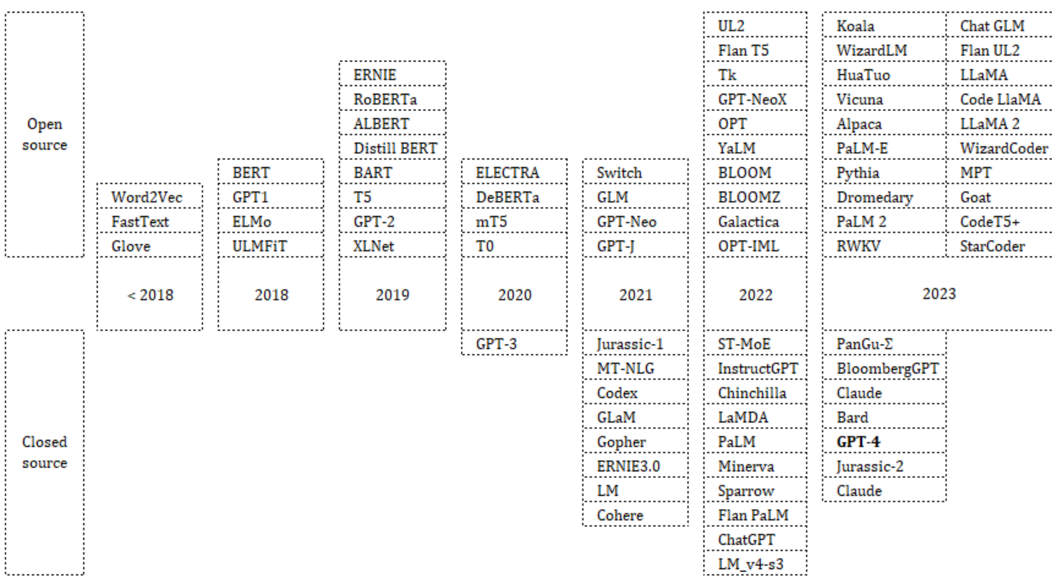
We assess the potential effect of large language models (LLMs) on the labor markets of Chile, Mexico, and Peru using the methodology of Eloundou et al. (2023). This approach involves detailed guidelines (rubrics) for each job to assess whether access to LMM software would reduce the time required for workers to complete their daily tasks. Adapting this methodology to the Latin American context necessitated developing a comprehensive crosswalk between the Occupational Information Network (O*NET) and regional occupational classifications, SINCO-2011 and ISCO-2008. When we use this adaptation, the theoretical average task exposure of occupations under these classifications is 32% and 31% for each classification. Recognizing the unique characteristics of each country's labor market, we refined these ratings to reflect better each nation's capacity to adopt and effectively implement new technologies. After these adjustments, the task exposure for SINCO-2011 drops to 27% and for ISCO-2008 to 23%. These adjusted exposure ratings provide a more accurate depiction of the real-world implications of LLM integration in the Latin American context. According to this methodology, the LLM-powered exposure using GPT-4 estimates suggests that the percentage of jobs with task exposure exceeding 10% is 74% in Mexico, 76% in Chile, and 76% in Peru. When we raise the exposure threshold to 40% or more, the proportion of affected occupations significantly decreases to 9% in Mexico, 20% in Chile, and 6% in Peru. The exposure is close to zero after this threshold. In other words, the exposure would only affect less than half of the total labor force in these countries. Further analysis of exposure by socioeconomic conditions indicates higher exposure among women, individuals with higher education, formal employees, and higher-income groups. This suggests a potential increase in labor inequality in the region due to adopting this technology. Our findings highlight the need for targeted policy interventions and adaptive strategies to ensure that the transition to an AI-enhanced labor market benefits all socio-economic groups and minimizes disruptions.

Keywords: Large language models (LLMs), GPTs, labor markets, exposure.
JEL Codes: C45, J23, J40, O33, O54

1 Introduction[1]

A large language model (LLM), such as Generative Pre-trained Transformers (GPTs) is a type of artificial intelligence (AI) algorithm that utilizes deep learning techniques and big data to process and generate text with a coherent sense of communication, capable of generalizing across multiple tasks (see Naveed et al. (2023) and Vaswani et al. (2023)). The evolution of LLMs has been rapid in recent years, particularly following the launch of ChatGPT in November 2022, illustrating an exponential growth trajectory as depicted in Figure 1. Prior to this era, pivotal innovations such as Word2Vec laid the foundational blocks, affecting contemporary uses like the Inter-American Development Bank's Labor Market Observatory, which incorporates that model into the classification of job vacancies and skills through Google Careers' virtual platform[2] Mikolov et al. (2013); IDB (2023). Recent advancements in LLMs have significantly broadened the spectrum, introducing an array of models that span from open to closed source. These breakthroughs have embraced specialized architectures, including decoder-only and encoder-decoder frameworks, showcasing the versatility and innovation in the field Devlin et al. (2019); Brown et al. (2020). They also represent an essential opportunity to increase productivity in the economy by reducing differences across workers in specific sectors. However, the practical deployment of these innovations necessitates a thorough understanding of their operational dynamics, potential, and limitations, which is crucial for leveraging their capabilities across different domains and tasks Yang et al. (2023).

Figure 1: Release of LLM technologies in chronological order. Those above the timeline are open-source, while those below the timeline are closed-source.



Source: The list has been prepared based on three sources of information. The work by Yang et al. (2023), which elaborates on an evolutionary tree towards modern large language models but differentiates them into four categories: transformed-based, decoder-only, encoder-only, and encoder-decoder. We have also used the curated list of papers, articles, tools, and communities that work extensively with LLMs, most of them related to Chat GPT of S. (2023), and finally, the curated list of efficient LLMs of Xinyin (2024). Developments existing before the end of the year 2023 have been taken into account.

We assess the potential effect of adopting LLMs in Latin American countries by following the methodology developed by Eloundou et al. (2023). This methodology offers ratings for assessing occupational exposure to LLMs within the U.S. labor market. Their research was not merely to uncover levels of direct exposure to LLMs but also to explore LLM-powered applications to have more comprehensive measures of their effects. In other words, the methodology facilitates estimating LLM-powered

[1]Special thanks to Pamela Mishkin, co-author of the paper that serves as the basis for this work, who has provided us with the complete list of ratings for each of the tasks/occupations.

[2]https://observatoriolaboral.iadb.org/en/

systems, which refers to any application or service that utilizes a Large Language Model. These models are types of artificial intelligence algorithms designed to understand, generate, and work with human language in a way that can mimic human-like understanding and response. They are "large" in the massive datasets needed for their training and complex neural network architectures, allowing them to process and produce language in diverse ways. Such systems can power a wide range of applications, including but not limited to chatbots and virtual assistants, content creation, language translation, information retrieval and summarization, education and tutoring, accessibility services, analysis, and insight generation, among others.

Eloundou et al. (2023) applied the labels to the United States labor market survey, using the Labor Force Demographics BLS (2023). Despite its limitations, such as the lack of occupational diversity among annotators or the sensitivity to changes in the wording of the rubric, this methodology provides a foundational framework for understanding the intersection of LLMs with labor markets O*NET (2023a). This innovative method offers a systematic approach to evaluate the implications of LLMs on various occupations and tasks. Incorporating human judgment and advanced AI tools bridges the gap between traditional labor market analysis and the emerging technological landscape. The dual use of human annotators and GPT-4-generated rubrics also allows for a comprehensive assessment, combining human intuition with AI's scalability and data-processing capabilities. Nevertheless, it's essential to note that absolute certainty remains elusive when determining whether human assessments of exposure more accurately mirror reality compared to those produced by GPT-4 or vice versa. This uncertainty stems from the fact that both sets of evaluations derive from criteria established by humans. However, a significant correlation between them enables the formulation of comparable conclusions from both data sets.

Given these initial considerations, we aim to use these rubrics to analyze employment surveys across Latin American countries like Chile, Mexico, and Peru. We ascertain how the emergence of LLMs could affect Latin American workers. We look into how using new technologies is not just a single experience but requires its adoption. However, when this adoption increases in intensive and extensive margins, the gains in productivity are significant Bessen (2002). The effects of technology adoption are complex, and they are challenging to estimate ex-ante. They can appear in various unexpected places and ways, affecting different groups and industries. These new inventions are exciting but also mix things up, changing how we work, the economy, and how people interact. Innovations across numerous scientific domains typically possess a disruptive nature, with their impacts manifesting in diverse ways. In that context, this research only aims to provide specific policy recommendations because of the unique circumstances of each Latin American country. Nonetheless, the findings we share could serve as a pivotal moment, offering an opportunity to shift the discourse from high-level, often abstract engagement with technology to a more nuanced understanding of its real-world impact and how these technologies are reshaping the day-to-day realities of the workforce.

2 Literature review

2.1 A brief review of different approaches to the impact of AI on labor markets

The effects of artificial intelligence can be approached from different perspectives, given its disruption in many areas. Ghosh et al. (2024) conduct an extensive review on the recent literature concerning the relationship between exposure to AI and labor markets, highlighting that among the documents they cite, it is notable that the majority of them indicate effects that are not entirely clear because it depends largely on how the response is approached both from firms and economic policy. On the one hand, Frey and Osborne (2017) suggests that approximately 47% of employment in the U.S. is at risk due to its likelihood of computerization. Autor et al. (2022) discuss the destruction of jobs in labor-intensive sectors that operate with outdated technologies. Acemoglu and Restrepo (2020) studies how automation and AI are displacing human labor with capital, which has impacted slower growth in total productivity in recent years. Acemoglu (2002); Duch-Brown et al. (2022); Moll et al. (2022) also talks about how technological change has generated wage inequalities originating both from the skill side and the greater increase generated by returns to wealth. Aghion et al. (2019) examines the impact of robotization and AI on employment in France, showing a decrease in aggregate employment at the zonal level. The negative impact is more pronounced for uneducated workers, challenging the notion that AI and automation only affect low-skill jobs.

Seen from another angle, Damioli et al. (2023) have a more friendly outlook than works that weigh more on labor displacement, finding a positive impact of AI families on employment. At the regional level, Guarascio et al. (2023) studied the effects of AI in Europe in the period 2011-2018, finding positive employment effects in the region characterized by a high participation of AI-exposed occupations. Acemoglu and Restrepo (2020) also found that the displacement effect of labor by capital is accompanied by an increase in labor demand, especially for jobs with non-automated tasks. Moreover, it is worth mentioning the works that highlight that any positive effect of AI on labor markets must be accompanied by training and updating of human capital, changes in organizational design, and investment in retraining and skill improvement during the labor transition Lu et al. (2021); Widuckel and Bellmann (2023); George et al. (2023). Although the direction of the impacts depends on a diverse set of circumstances and actions that each economy can take, these should also consider the differentiated impact within each economy, such as in industries, context, and specific sectors Fatun and Pazour (2021); Xie et al. (2021); Zhou et al. (2020); Holm and Lorenz (2022).

Given the diverse perspectives on the impact of artificial intelligence on labor markets, its role is complex and multifaceted. Substantial evidence suggests risks and challenges, including job displacement and widening wage inequalities. Similarly, there's optimism about AI's potential to create jobs and demand new skills, especially when coupled with appropriate training and organizational adjustments. What emerges from this analysis is not a one-size-fits-all outcome but a nuanced picture that demands a proactive approach from policymakers, businesses, and workers alike. The future of AI in the labor market hinges on our ability to navigate its challenges while seizing the opportunities it presents for innovation and growth. This balance will be crucial in ensuring that the benefits of AI are broadly shared across society, mitigating the risks of displacement and inequality. The discussion underscores the importance of a collaborative effort involving investment in education, skills development, and the thoughtful integration of AI technologies into the economic fabric, aiming for a future where technology and human labor coexist, complement, and enhance each other's value.

2.2 Relevant studies on LLMs and labor markets

The effects of language models and labor markets are determined on the capacity of these models to solve tasks ranging from the simplest to the most complex. The scalability of the models from small language models to LLMs defines the capacity since it becomes becomes larger in terms of training, parameters, or data quantity, which requires emergent abilities Wei et al. (2022). New studies on the effects of LLMs on occupations have emerged to explore the development of new skills in recent years. For example, Chen et al. (2023a) conducted a thorough examination of job listings from BOSS Zhipin, China's leading online job platform, revealing that approximately 28% of current job roles demand skills related to ChatGPT. They used an extensive occupation-centered knowledge graph to enhance a predictive analysis through a semantic information-enhanced collaborative filtering approach. Their findings suggest an anticipated increase, where an additional 45% of job roles would use ChatGPT competencies, especially within tech, product, and operational sectors. Conversely, areas like manufacturing, services, education, and health sciences will likely demand fewer skills. Huseynov (2023) analyzes how AI discussions impact US students' labor market outlooks. He finds that positive and negative AI narratives can affect earnings expectations, with a more substantial impact of the negative information. Non-STEM and non-male students are more likely to show pessimistic shifts in their beliefs. These findings highlight the importance of educational adjustments to prepare students for an AI-influenced future.

Zarifhonarvar (2023) examines the potential impacts of ChatGPT on the labor market. It finds that 32.8% of occupations could be fully impacted, 36.5% might face partial impacts, and 30.7% could remain unaffected. The study categorizes the level of impact across different occupations, indicating significant shifts in job requirements and skill sets due to AI integration. This analysis is a reference on how generative AI technologies, like ChatGPT, may reshape the workforce and employment patterns. Kreitmeir and Raschky (2023) offer a more detailed examination of the impacts of LLMs on productivity, exploring the immediate aftermath of Italy's ban on ChatGPT due to privacy concerns. The paper utilizes GitHub data to analyze software developers' productivity changes in Italy compared to control groups in Austria and France. While focusing on the ban's impact, the study reflects broader themes on how abrupt regulatory actions can influence individual productivity and technological adoption. It discusses using GitHub and Google Trends data as empirical tools for measuring labor activity and public interest in AI technologies post-ban. Additionally, Liu (2023) explored the potential effects of LLMs on the economic cycle of the Chinese economy and its labor markets, which were significant in both instances. Concerning economic cycles, it enhances precision in predicting economic trends, improves data processing efficiency, and holds the potential to identify new business opportunities. In labor markets, it aids in better identifying the evolution of job vacancies, job searches, and job requirements. It also contributes to a deeper understanding of the implications of changes in regulations and minimum wages. Finally, Chen et al. (2023b), who have developed their own rubrics following Eloundou et al. (2023) methodology to explore the potential impact of LLMs on job roles in China. They found a significant occupational exposure variation and suggested higher paying, educated jobs are more susceptible to LLM impacts. It also explores industry-specific exposures and demographic impacts, indicating young workers face greater disruption risks.

In this context, our work sheds light on the nuanced implications of LLMs across various demographic segments, examining potential disparities in impact across genders, age groups, income, and levels of educational attainment. Additionally, we endeavor to explore the ramifications of these technological advancements in the formal and informal sectors, thereby contributing to a more comprehensive understanding of the digital transformation's reach within diverse labor ecosystems.

3 Exposure rubric methodology

Following Eloundou et al. (2023), exposure indicates whether having access to a Large Language Model (LLM) would decrease the time needed for a human to complete his daily work tasks. However, while some individuals find it relatively easy to use LLMs, others may be required to modify the task to adapt to the technology. This type of exposure is defined as LLM-powered software. Exposure is a measure of access to an LLM or an LLM-powered system that could reduce the time required by a human to complete a Detailed Work Activity (DWA) [3] or at least 50% of a task. DWAs provide information on the common work activities required across occupations. These cross-occupational descriptors are less specific than occupationally specific Tasks, and they correspond with general Work Activities associated with O*NET (2023a).

Exposure data was generated for human and GPT-4 annotations using a newly designed rubric for assessing LLM capabilities and their possible effects on employment. Human annotations were conducted at the DWAs level, while GPT-4 classifications were performed at the task level, revealing a strong correlation between the two. In this study, following the approach taken by the authors in the case of the US labor market, we will use both human and GPT-4 scores for exposure analysis, to be more specific, the rubric number one that the authors created.[4] The authors set a threshold at a potential 50% reduction in the time required to complete a DWA. They also note that the choice of this number is arbitrary, primarily chosen for its ease of interpretation for annotators. Furthermore, to ensure a reasonable understanding of occupational exposure, they expressed the variables of interest using three parameters: The first is α corresponding solely to E1, which in this case would be considered the lower bound of exposure. The second one is β, which, in addition to E1, adds 0.5 times E2, representing the potential exposure to LLMs if complementary technology is developed to enable their effective

[3] A DWA represents a detailed activity involved in the execution of a task. Conversely, a task is defined as a specific work associated with an occupation, which might encompass none, one, or several DWAs.

[4] The authors acknowledge that the GPT-4 model's assignment of exposure ratings can be sensitive to word usage or emphasis variations. Consequently, they use two different rubrics to present their findings. Moreover, the rubrics provided to human evaluators slightly deviate from those supplied to the GPT-4 model. This strategic choice ensured that the model generated ratings without being unduly influenced by the human annotations.

use—previously referred to as the *LLM-powered software*. Finally, ζ, an upper bound represented as the unweighted sum of E1 and E2. The following box extracted from the authors' paper clearly expresses the concept of exposure addressed in this document[5]. The paper will primarily focus on the α/β exposure.

No exposure (E0) if:

- using the described LLM results in no or minimal reduction in the time required to complete the activity or task while maintaining equivalent quality[a] or using the described LLM results in a decrease in the quality of the activity/task output.

Direct exposure (E1) if:

- using the described LLM via ChatGPT or the OpenAI playground can decrease the time required to complete the DWA or task by at least half .

LLM+ Exposed (E2) if:

- access to the described LLM alone would not reduce the time required to complete the activity/task by at least half, but additional software could be developed on top of the LLM that could reduce the time it takes to complete the specific activity/task with quality by at least half. Among these systems, we count access to image generation systems

[a]Equivalent quality means that a third party, typically the recipient of the output, would not notice or care about LLM assistance

$$GPT-4,\ Rubric\ 1 \begin{cases} \alpha & \text{if E1} \\ \beta & \text{if E1 + (0.5)E2} \\ \zeta & \text{if E1 + E2} \end{cases}$$

4 Classification of occupations for Latin America: a crosswalk from O*NET to ISCO-2008 and SINCO-2011

The ratings we use in this document for occupations in the Latin American labor markets were initially created by assigning exposure values to LLMs for each task/occupation in the O*NET classification. O*NET serves as the primary source of information on occupations in the United States of America, containing data on hundreds of standardized occupation-specific descriptors. This database is continuously updated through surveys of a diverse range of workers in various occupations O*NET (2023b). Notably, the first six digits of O*NET correspond to the Standard Occupational Classification (SOC). [6] The only distinction between the two is that the O*NET system incorporates two additional digits, refining the categorization of occupations into more specialized groups in specific instances, as illustrated in Table 1. This table describes how the original code 29.1141 corresponding to registered nurses serves as the base for disaggregating it into four additional occupations DLR (2023).

[5]The authors of these rubrics argue that exposure levels could be mild but significantly lower in the real world. That's why they chose relatively high thresholds.

[6]The SOC system is the framework we will consistently employ throughout the remainder of this document to align the occupations of Latin America with those in the United States labor market. Nonetheless, given that O*NET is essentially an extension of SOC, we will employ the terms interchangeably when referring to the North American system.

Table 1: Correlation between SOC and O*NET

SOC Code	**SOC Title**	**O*Net Code**	**O*Net Title**
29.1141	Registered Nurses	29.1141.00	Registered Nurses
		21.1141.01	Acute Care Nurses
		21.1141.02	Advanced Practice Psychiatric Nurses
		21.1141.03	Critical Care Nurses
		21.1141.04	Critical Nurse Specialists

Source: South Dakota Department of Labor and Regulation.

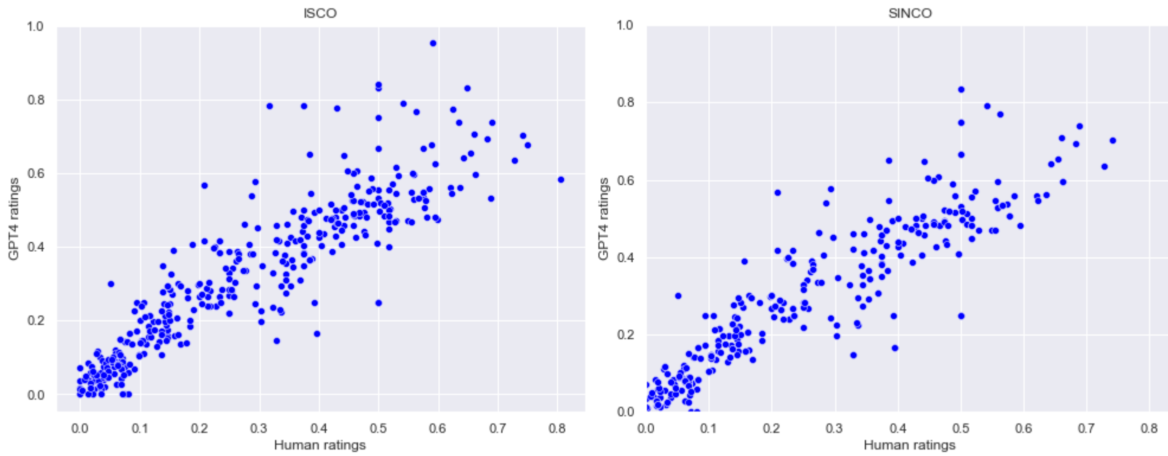
This paper delves into the challenge of occupational classification systems in Latin American countries. Most of them use ISCO, and having a reliable crosswalk with O*NET would be ideal. Unfortunately, there is no official instrument available for this purpose. Alternatively, we use the official crosswalk between ISCO-2008 and the SOC-2010. This crosswalk shows no perfect correspondence since few occupations have been successfully matched, and some others show weak matches. The reason for these flaws is that some tasks in occupations defined in SOC do not have a clear counterpart in the ISCO definitions, leading to the use of an approximate match BLS (2012).

We use this crosswalk on the employment variables of the household surveys collected for Chile, Mexico, and Peru. In the case of Chile, the National Socioeconomic Characterization Survey (CASEN 2022) employs an adaptation of the International Standard Classification of Occupations ISCO-08.CL.[7] This version has been designed to enhance the survey's quality, comparability, and statistical standardization by adopting an international framework, ensuring alignment with the specificities of the local labor market MDSF (2022); INE (2018). Our approach involves aligning this adaptation with the original ISCO-2008 and applying the match to SOC-2010. In Peru, the Permanent National Employment Survey (EPEN) adopts the National Classification of Occupations 2015 (CNO-2015), similar to Chile, as an adaptation of ISCO-2008. Then, it will be aligned to the original ISCO-2008 and matched to SOC-2010. Finally, in the case of Mexico, we use the National Occupation and Employment Survey (ENOE). This survey has its own occupational classification system, SINCO, in its most recent version from 2019. The challenge arises from the lack of a direct mapping between SINCO-2019 and SOC-2010, although it exists for SINCO-2011. Therefore, our initial approach will be to convert data from SINCO-2019 to SINCO-2011 and then align it with the SOC-2010 standards. Finally, once the crosswalks are in place, the specific labels (ratings) associated with each occupation in O*NET will be seamlessly integrated into the corresponding occupations in ISCO-2008 and SINCO-2011.

The procedure described in the preceding paragraph yields the list of occupations for both classification systems, along with their respective exposure levels. We average the exposure of the occupation by using the exposure values per task collapsed at the level of each occupation. In other words, the values reflect the average percentage of exposure across all tasks within a particular occupation. Figure 2 represents the close correlation between human rubrics and those of the GPT-4 model using the coefficient β for both occupation classification systems after using the crosswalk values.

[7] Initially, working with the National Employment Survey (ENE) would have been preferable. However, we faced a challenge: the variable that includes the occupational classification code is limited to just one digit, unlike the CASEN survey, which provides the complete four-digit code.

Figure 2: Correlation between human and GPT-4 ratings for each occupational classifier



Note: On the x-axis, we present the human ratings, while the GPT-4 ratings are displayed on the y-axis. Each point represents the average exposure value assigned to a particular occupation, calculated as the aggregate. average exposure of all tasks within that specific occupation.

Table 2 shows the descriptive statistics of exposition for both occupational classifiers, comparing them with those originally sourced from O*NET/SOC.[8] These average exposure levels can differ between human and model-generated rubrics. Although we focus the results on the α and β coefficients, the figure helps highlight the consistency of the ratings after the crosswalk process and the similarity of values between classifiers in Latin America and SOC. According to this, the results can be interpreted as follows: at the occupation level, in the case of ISCO, the lower bound α indicates that on average, approximately 13% of tasks within an occupation are exposed to LLMs, both in human annotations and the GPT-4 model. On the other hand, in the SINCO classifier, the GPT-4 model generates slightly lower exposure (11%) compared to the human annotations. The values consistently increase when considering LLMs powered system, reflected in the coefficient β, which averages 32% exposure for the GPT-4 model under the ISCO classifier, four points above human annotations. For SINCO, the average exposure of the model maintains the same difference as that of the human labels.

Table 2: Average task exposure of occupations using the crosswalk

	Human ratings			**GPT-4 rubrics**		
	O*NET/SOC	ISCO	SINCO	O*NET/SOC	ISCO	SINCO
α	0.14	0.13	0.13	0.14	0.13	0.11
	(0.14)	(0.14)	(0.13)	(0.16)	(0.14)	(0.12)
β	0.30	0.28	0.27	0.34	0.32	0.31
	(0.21)	(0.19)	(0.19)	(0.22)	(0.20)	(0.19)
ζ	0.46	0.42	0.41	0.55	0.51	0.50
	(0.30)	(0.27)	(0.26)	(0.34)	(0.33)	(0.33)

*Standard deviations in parentheses

Now, let's see which specific occupations exhibit the greatest direct and indirect exposure. Table 3 reveals some interesting points in this regard. For instance, the human assessments for α exposure suggest a focus on professions that traditionally rely heavily on language and communication skills. Notably, 'Public relations professionals', 'Journalists', and 'Authors and related writers' rank high, indicating a belief in the potential for direct LLM application in creating content or managing communications. In contrast, the model-based exposure rates 'Data entry clerks' and 'Telephone

[8]These ratings assign double the weight to tasks designated as "core" by O*NET compared to tasks designated as "supplementary." The task weights add up to one within an occupation.

switchboard operators' higher, reflecting an expectation that LLMs can automate or facilitate tasks involving structured, routine information processing. The consistency across human and model assessments for clerical and support roles entails a consensus on their susceptibility to LLM integration. The α exposure shows an interesting spread across both human and social science roles, with humans presumably weighing the nuances and unpredictability of such positions. The model's emphasis on clerical roles suggests an algorithmic bent toward recognizing the ease of automating repetitive tasks. For β exposure, there is an intriguing acknowledgment of the potential for LLMs to alter the work landscape without necessarily eliminating jobs. For example, roles such as "Credit and loans officers" and "Advertising and marketing professionals" indicate a belief in the potential for LLMs to augment complex decision-making and creative processes.

The table also reflects the socio-technical dynamics in the labor market's integration of LLMs. In some areas, the alignment between human and model assessments denotes a shared understanding of automation trends. At the same time, divergences highlight areas where human judgment still finds complexity beyond the reach of current models. The discrepancies between human and GPT-4 ratings could also imply a gap in technological literacy and understanding of the capabilities of LLMs. It indicates a need for better communication and education around AI's potential and limitations. For a detailed review of the top 50 occupations with the highest exposure levels, please refer to the Annex.

Table 3: Top 10 Occupations with the Highest Exposure to LLMs According to the ISCO-2008 Occupational Classification System

Human		**Model**	
		α exposure	
Public relations professionals	66.7	Data entry clerks	77.8
Administrative and executive secretaries	59.0	Telephone switchboard operators	70.1
Translators, interpreters and other linguists	56.4	Typists and word processing op...	70.0
Journalists	52.4	Clerical support workers not...	68.7
Authors and related writers	50.8	Payroll clerks	68.2
Legal secretaries	50.0	Accounting and bookkeeping clerks	63.3
Contact centre information clerks	47.7	Statistical, mathematical and rel...	63.2
Enquiry clerks	47.2	Medical secretaries	58.8
Clerical support workers not els..	47.2	Contact centre salespersons	58.3
Economists	45.5	Enquiry clerks	58.1
		β exposure	
Public relations professionals	80.6	Payroll clerks	84.1
Legal secretaries	75.0	Data entry clerks	83.3
Translators, interpreters and other linguists	74.1	Clerical support workers not else...	83.1
Administrative and executive secretaries	72.7	Contact centre salespersons	79.2
Authors and related writers	68.9	Accounting and bookkeeping clerks	78.3
Mathematicians, actuaries and statisticians	68.2	Accounting associate professionals	78.3
Credit and loans officers	66.2	Medical secretaries	77.7
Advertising and marketing professionals	66.1	Statistical, mathematical and rel...	77.3
Transport clerks	65.4	Telephone switchboard operators	76.9
Clerical support workers..	64.8	Typists and word processing operators	75.0

Note: Occupations with no participation in employment surveys have not been considered, as due to changes over time in the types of occupations, it is possible that the classifiers we use are outdated with jobs that are no longer in demand today.

5 Data sources

Our estimates of the exposure of occupations are based on four different databases. They include, first, the ratings generated by Eloundou et al. (2023). Second, the occupational classification tables with their respective crosswalks, which include the Occupational Information Network (O*NET) 27.2 database, the Standard Occupational Classification (SOC) 2010, the International Standard Classification of Occupations (ISCO) 2008, and the National Occupational Classification System (SINCO) 2011 and 2019. Third, employment and household surveys which encompass the National Socioeconomic Characterization Survey (CASEN) for Chile, the Permanent National Employment Survey of Peru (EPEN), and the National Occupation and Employment Survey of Mexico (ENOE). In addition to classifying the occupations, we use sociodemographic characteristics of subjects in the surveys to adjust the exposure. We finally use World Development Indicators and the Global Information Technology report[9], both from the World Bank also for individual adjustment.

6 Exposure estimates

The exposure ratings are primarily assigned based on the tasks of the individuals, which is inherently tied to the labor market's structure. In our selected cases, only the proportion of the workforce in transformative and digital sectors is reduced compared to more developed countries, and the share of workers in the primary is bigger, with limited exposure levels. Consequently, when these labels are applied within employment surveys, we essentially attribute a value to each employed individual corresponding to their occupational classification code. Table 4 illustrates the intensity of exposure across each economy, utilizing both labeling methods. For direct exposure both human labels and the model indicate that fewer than 5% of occupations in the three countries involve a task exposure of 50% or higher. This means that less than 5% of the occupations demonstrate an α_{50} and β_{50} level of task exposure.

Table 4: Exposure intensity by country (share of occupations with α/β exposure)

LLM's direct exposure		α_{10}	α_{20}	α_{30}	α_{40}	α_{50}	α_{60}	α_{70}	α_{80}	α_{90}	α_{100}
Human	Mexico	50.7	28.3	5.4	3.5	1.5	0.0	0.0	0.0	0.0	0.0
	Chile	41.3	22.7	8.1	6.4	4.5	0.1	0.0	0.0	0.0	0.0
	Peru	38.6	24.3	8.3	4.4	2.3	0.0	0.0	0.0	0.0	0.0
GPT-4	Mexico	40.0	13.2	7.0	3.8	0.7	0.6	0.6	0.0	0.0	0.0
	Chile	40.4	11.9	6.8	3.4	2.2	0.5	0.1	0.0	0.0	0.0
	Peru	41.5	13.1	5.9	4.3	3.4	2.9	0.2	0.0	0.0	0.0
LLM-powered exposure		β_{10}	β_{20}	β_{30}	β_{40}	β_{50}	β_{60}	β_{70}	β_{80}	β_{90}	β_{100}
Human	Mexico	68.8	54.8	44.5	19.7	8.2	2.6	1.2	0.0	0.0	0.0
	Chile	66.7	48.8	38.6	20.2	10.4	3.4	1.5	0.1	0.0	0.0
	Peru	63.2	41.6	34.7	15.4	9.7	3.7	0.0	0.0	0.0	0.0
GPT-4	Mexico	77.4	61.4	47.3	34.4	11.0	2.6	0.8	0.4	0.0	0.0
	Chile	77.1	59.9	37.0	27.2	11.7	5.3	2.4	0.3	0.0	0.0
	Peru	81.8	49.8	27.0	16.6	8.6	4.8	3.5	0.4	0.0	0.0

Note: The figures represent percentages. The employment surveys taken into account correspond to the end of the year 2022. Each value represents a minimum, meaning that, for example, the first value in the first row should be understood as 50.7% of the employed population having at least 10% of their tasks directly exposed to LLMs. This way of interpreting exposure data is adopted from the foundational study of this paper.

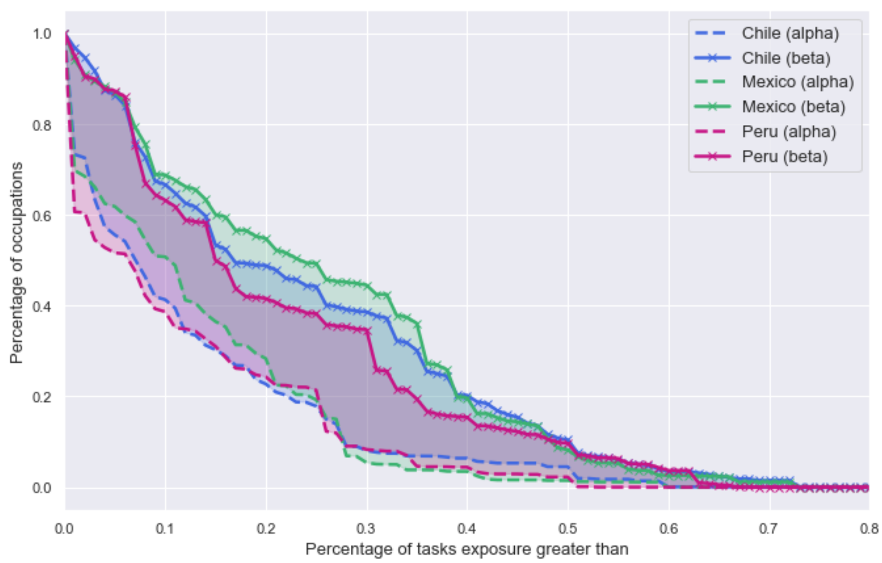
Although the exposure values across the three countries remain low, they show important variations, particularly for task exposure levels below 80%. On the other hand, if we include LLM-powered applications, the percentage of occupations with α_{50} and β_{50} task exposure increases significantly in both labeling methods. For instance, according to the model, Mexico shows the greatest growth potential, as it jumps from 0.7% to 11%.

These figures are highly informative, yet their graphical representation provides other insights. Figures 3 and 4 show which country stands in a more favorable position based on its direct exposure

[9] Prosperity Data 360. World Economic Forum.

(lower bound). They also highlight the countries where LLM-powered systems significantly influence the proportion of occupations subject to higher task exposure beyond the mere consideration of LLMs alone. On the x-axis each data point indicates the percentage of tasks subject to each exposure level α and β. The y-axis represents the proportion of occupations within a country subjected to that particular level of exposure. In both charts, the intensity of direct exposure (α) of tasks in the share of occupations among countries behaves very similarly for task exposure levels above 20%. However, the differences become more pronounced when LLM-powered software is included, especially in the ratings provided by the model (see Figure 4). Observing the α exposures, all three countries show a steep decline in the percentage of occupations as exposure increases, with the lines converging around the 20% mark. This suggests that up to this point, occupations within these economies share a similar level of exposure to tasks that could be affected or require significant adaptation due to technological advancements. The graph reveals an even starker contrast when we introduce the β exposure lines. Mexico's β exposure line initially runs parallel to its α line but then diverges, indicating that integrating LLM-powered applications may increase the proportion of occupations facing significant task exposure. On the other hand, Peru's β exposure line diverges from its α line at lower levels of exposure. It remains consistently below Mexico's, suggesting that Peruvian occupations are not adopting LLM-powered tools at the same rate or that these tools are not as applicable to the tasks within the Peruvian labor market.

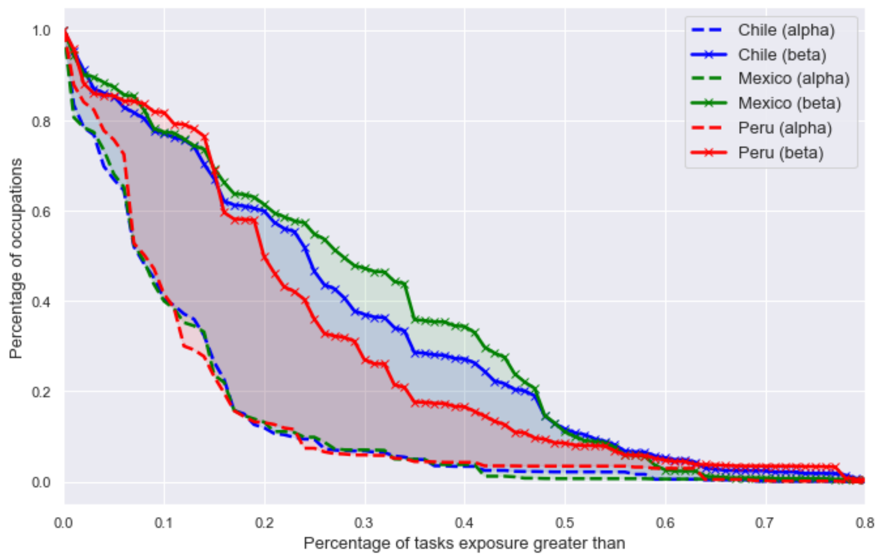
Figure 3: Percentage of occupations by level of task exposure (Human ratings)



The consistency of the alpha lines among the three countries at lower exposure levels and their divergence at higher levels point to a nuanced interplay between the types of tasks performed within these economies and the potential for technology to reshape them. This underscores the varying degrees to which each country's labor market might either harness the benefits or face the challenges posed by the advent of LLM-powered applications. The β exposure lines also show a broader spread between the countries in Figure 4 compared to Figure 3. This suggests that the GPT-4 model may predict a more significant impact of LLM-powered systems on task exposure than human raters do.

The convexity of the curve in our analysis represents the depth of a labor market in terms of quantity for potentially incorporating LLMs efficiently. In countries with a higher proportion of workers in highly specialized activities, the convexity is lower, indicating that the impact on employment significantly depends on the structure of the labor market and the level of specialization and adaptability of the workforce. This insight is crucial in understanding the potential for LLMs in different labor markets. However, we must also consider that these outcomes are influenced by the step-by-step crosswalk process from O*NET to each respective classification of occupations.

Figure 4: Percentage of occupations by level of task exposure (GPT-4 model)



In summary, while human raters and the GPT-4 model share similarities in their task exposure assessments, the GPT-4 model tends to predict a more significant impact of LLM-powered systems, particularly in Mexico. These differences highlight the importance of considering various evaluative models and perspectives when assessing the potential effects of LLMs on labor markets.

6.1 Exposure by sociodemographic group

While general exposure metrics provide valuable insights into the potential for occupational adoption of new technologies, it is equally crucial to assess who will ultimately harness and maximize the capabilities afforded by these advancements. Table 4 displays the average α/β exposure values to LLMs in each country according to socioeconomic characteristics, including gender, age, education, formality condition, and salaries. Key findings include, on average, higher exposure for women than men for human and GPT-4 labels. In Peru, this indicator slightly favors men by around one point under the metric of the model. The increased exposure of women could suggest a couple of trends. First, the job roles typically occupied by women may evolve to incorporate LLM tools at a quicker rate, whether in administrative, educational, or other service-oriented sectors. This correlation could imply that this technology can have long-term effects on job security and the need for skills training. The notably higher β exposure for women in all countries could shows that while direct use of LLMs isn't substantially higher than men, the indirect use through applications is significant, highlighting a potential area for capacity-building and targeted education. Regarding age, a bell-curve pattern emerges for both labeling methods, showing that middle-aged individuals (25-45 years) report the highest levels of LLM interaction. The exposure within the youngest cohort across human and GPT-4 ratings reflect the digital native status of this group, potentially indicating a greater comfort with and access to technology. We also identify greater exposure rates among individuals possessing higher educational degrees. Individuals with postgraduate education have the highest value, aligning with the expectation that more specialized, high-skill jobs would leverage advanced technologies for efficiency and innovation. This trend emphasizes the growing importance of advanced education in the digital economy but also raises concerns about increasing the digital gap.

Table 5: Average α/β exposure levels by socioeconomic characteristics

Category	Human ratings			GPT-4 ratings		
	Chile	Peru	Mexico	Chile	Peru	Mexico
Gender						
men	8.7/20.2	8.4/19.6	9.9/22.4	10.1/24.5	12.2/24.2	10.8/27.3
women	15.8/28.1	13.4/22.9	15.3/27.8	11.9/29.3	10.8/23.0	10.7/29.7
Age group						
[15,24]	10.9/21.2	10.7/19.2	11.7/21.1	11.3/23.6	10.9/21.2	10.7/23.8
[25,45]	12.9/25.9	11.2/22.5	12.8/26.1	11.7/29.0	12.3/25.3	11.1/29.8
[46,64]	10.7/21.7	10.3/20.7	11.9/25.0	10.2/24.6	10.9/23.1	10.5/28.7
Education						
none	6.7/13.6	3.9/10.9	6.6/15.1	7.1/15.6	7.9/15.7	6.7/18.5
primary	7.6/15.8	5.3/12.2	7.7/16.8	7.1/17.9	8.2/15.7	7.5/20.2
secondary	6.4/13.6	7.9/15.8	10.6/20.9	7.0/15.9	10.4/18.4	10.6/24.6
technical	10.5/21.3	13.1/25.1	15.5/29.5	11.4/24.6	14.0/27.4	13.2/33.1
graduate	18.4/36.9	17.8/33.9	17.7/36.3	12.4/39.7	14.9/36.0	12.9/40.0
post graduate	22.5/43.7	20.0/41.5	20.1/41.0	11.3/44.6	10.9/42.4	11.5/42.3
Formality						
informal	10.7/20.5	7.9/16.1	10.0/21.1	9.1/22.7	9.9/19.0	8.6/24.6
formal	12.2/25.2	16.6/32.1	14.9/29.4	11.9/28.6	15.1/34.1	13.5/33.2
Wage quintile						
q1	11.5/20.4	7.1/14.6	10.2/19.3	9.2/21.3	9.2/17.3	8.1/21.4
q2	9.0/18.3	9.7/18.3	11.6/21.3	9.9/21.2	11.2/20.0	10.9/24.1
q3	10.1/21.1	10.4/20.1	10.9/21.3	11.6/24.8	12.4/22.4	11.0/24.7
q4	12.1/25.5	11.2/22.8	11.5/24.6	12.9/29.8	12.6/25.7	11.3/29.0
q5	16.9/35.6	16.2/33.1	14.1/31.2	11.7/39.1	13.9/35.6	11.5/35.8

Note: The displayed values represent simple averages calculated across the entire employed population. However, for the "age group" category, these averages exclude employed people younger than 15 years and older than 64 years. Each educational category comprises not only those who have completed their studies at that level but also individuals who started but did not finish it for several reasons. Employment formality is clas-. sified according to the definition by SIMS (2023), identifying formal employment as any employed individual contributing to the social security system. Salary quintiles are determined based on the monthly labor earnings in the local currency of each respective country.

Formal workers exhibit significant differences compared to informal, with this gap amounting to approximately 16 percentage points in Peru, 9 in Mexico, and 6 in Chile. This difference may reflect technological integration within structured employment sectors. Specifically, considering Peru's informality rate, which hovers around 70%, there seems to be a correlation where high levels of precarious occupations are associated with a more significant exposure gap between both types of jobs. The functioning and organization of formal jobs make it easier to integrate LLMs compared to informal ones. This includes better access to technology, infrastructure, organizational support for technology adoption, and roles that benefit from automation and AI assistance. Lastly, regarding salaries, there appears to be a positive correlation between quintiles and exposure levels. Specifically, the fifth one averages an exposure rate exceeding 30% across all instances, in contrast to the first quintile, which averages 14%. It aligns with the presumption that higher-wage jobs are more complex, cognitive, and non-routine. These are the types of tasks where LLMs can serve as powerful tools for augmentation rather than replacement, hence the increased exposure.

7 Adjustment of exposure using sociodemographic characteristics

Eloundou et al. (2023) introduced exposure estimators for both individual tasks and broader occupational categories. These occupational estimators are constructed by aggregating data from individual tasks. However, our estimates for the selected cases are derived from the the process of crosswalking—aligning different occupational classification systems—results in expo- sure estimates for Latin American contexts solely at the occupational level, not at the task level. Up to this point, it's clear that the exposure ratings applied to O*NET initially targeted tasks, and then the exposure of an occupation was determined as the average exposure of the tasks within that occupation. These crosswalks are not perfect but represent the most commonly used most used tool when analyzing occupations across countries with different occupational classifiers. Therefore, it is evident that the purpose of this crosswalk is not to equate the realities of both regions but rather to approximate the methodology while acknowledging the significant differences that exist.

The significant difference between labor markets within Latin America and the U.S. labor market is key in this regard. These differences are not merely statistical but deeply rooted in many historical, economic, and societal factors that shape each region's unique employment landscape. For instance, structural differences in economic sectors, varying levels of technological adoption, and distinct labor regulations contribute to the diverse labor market dynamics. Additionally, cultural values and educational systems also play pivotal roles in shaping the workforce and its adaptability to new technologies. Therefore, when comparing these labor markets, one must consider these multifaceted backgrounds to fully understand the disparities and similarities in occupational exposure to technological advancements. This comprehensive approach enhances our understanding of each labor market's unique characteristics and provides valuable insights into the broader implications of global technological shifts on employment. Then, Beyond the broad historical differences that have led the United States and Latin America down distinct development paths, as outlined in Trebat (2012), our focus will narrow to examine specific indicators relevant to this study. In particular, we will delve into aspects such as productivity, which significantly contribute to the disparities between these regions. As listed in Zeileis (2023) and Feenstra et al. (2015), a marked productivity disparity exists between North and South America. The Gross Domestic Product (GDP) per hour worked (measured in international dollars) stands at \$73.7 in the United States, compared to \$28.4 in Chile, \$20.3 in Mexico, and \$11 in Peru[10]. However, productivity is just one dimension of a broader spectrum of differences, including infrastructure, education levels, and the quality of public services. These discrepancies significantly influence the applicability and impact of labor market technologies such as Large Language Models (LLMs). As explored in Wang et al. (2023), leveraging LLMs in public services could notably bridge access and efficiency gaps in lower to middle-income nations. Nonetheless, the real-world application of such technologies in these regions is often hampered by infrastructural deficiencies, prevalent low literacy rates, and limited digital literacy among the population.

To lessen the overestimation of the exposure, we adjusted the exposure ratings by tailoring the weightage of the E1 and E2 parameters of the selected methodology (see Box 1) to better align with the socio-economic realities of Latin American countries. These estimations is informed by the significant body of research addressing regional disparities in economic development, technological infrastructure, and educational attainment levels. According to Kafka (2021) and Reinhold et al. (2010) and Iammarino et al. (2018), integrating such regional disparities into economic models enhances their precision and relevance. This way, to refine the exposure parameters, we initially sought a reliable data source containing indicators associated with or have significant implications for the development and use of new technologies. These indicators needed to be relevant and reasonably comparable across countries in Latin America and the United States. For this purpose, we selected eight indicators from the World Development Indicators database WBG (2024); Dutta et al. (2017). Then we categorized these indicators into two primary groups: Education and Access & Skills .

[10]Data sourced from Our World in Data, "Productivity: output per hour worked," 2019.

Indicators by group

Education

- Literacy rates, adult total (% of people ages 15 and above)
- Average years of schooling for adults (aged 25 years and above)
- PISA math scores
- Government expenditure on education, total (% of GDP)

Access and Skills

- Individuals using the Internet (% of population)
- Fixed broadband subscriptions (per 100 people)
- Digital skills among the population
- Firm-level technology absorption

We use these indicators to build our adjustment factor for the three studied countries. We estimated the adjustment factor through a two-tiered composite index. Initially, the first tier combines four indicators within each category to form a composite index. Then, in the second tier, these composite indices from each category are averaged, using specific weights, to produce the final adjustment factor. Before developing this factor, we standardized each indicator, which is essential to accurately account for the existing disparities and facilitate the comparison between the United States and selected cases. Finally, weights were assigned to each indicator and group using the Equal Weights (EW) methodology. This approach is extensively employed in quality-of-life and well-being studies mainly because of its straightforward computation, interpretability, and desirable statistical properties. Moreover, it has been demonstrated to be more robust compared to others Shi and Land (2021). The Inter-American Development Bank employs this method to develop the Better Jobs Index, ensuring equal emphasis is placed on the index's quality and quantity aspects IDB (2017). The equation 1 straightforwardly illustrates the construction of the factor using each indicator within every group:

$$F^c = \frac{1}{2}(\sum_{i}^{4} E_i^c * \frac{1}{4}) + \frac{1}{2}(\sum_{i}^{4} A_i^c * \frac{1}{4}) \tag{1}$$

The term F^c denotes the final adjustment factor for the country c. E_i^c represents the i^{th} indicator from the education group for country c, and A_i^c refers to the i^{th} indicator from the access and skills indicator group for country c. The fractions represent the weights assigned within each indicator and between the groups. Upon completing the calculations, we have ascertained the final adjustment factors, which are a testament to the comparative analysis. The outcomes reveal a value of 0.86 for Chile, which suggests a relatively low level of adjustment in the context of our study. For Mexico, the factor is 0.76, indicating a moderate adjustment. Lastly, Peru's factor is established at 0.71, indicating a more substantial adjustment. These factors are pivotal in understanding the regional variances and will be instrumental in the subsequent comparative analysis.

7.1 Results obtained from adjusted ratings

Tables 6 and 7 show the results using the adjusted rubrics for both types of annotations before and after the adjustments of the coefficients. Compared to the original estimates included in Table 2, the direct exposure coefficient, α, shows a reduction of 3 points for human labels, whereas the model displays a less pronounced adjustment. The exposure coefficient β results in a more pronounced difference across the three countries. Peru, which experienced the most significant adjustment, the difference is

8 points. So, the proportion of tasks exposed to LLM of Peruvian workers, initially at 28%, is reduced to merely 20% for human annotations and decreases from 32% to 23% for model predictions. Standard deviations show a consistent decrease after adjustments in both tables. This contraction suggests that the adjustment leads to more consistent estimations across occupations within the countries, reducing the disparity between occupations with high and low initial exposure estimates.

Table 6: Adjusted occupation exposure by country (Human annotations)

	ISCO			SINCO	
	Chile/Peru (before)	Chile (after)	Peru (after)	Mexico (before)	Mexico (after)
α_{adj}	0.13	0.11	0.09	0.13	0.10
	(0.14)	(0.12)	(0.10)	(0.13)	(0.10)
β_{adj}	0.28	0.24	0.20	0.27	0.21
	(0.19)	(0.17)	(0.14)	(0.19)	(0.14)
ζ_{adj}	0.42	0.37	0.30	0.41	0.32
	(0.27)	(0.23)	(0.19)	(0.26)	(0.20)

*Standard deviations in parentheses

The adjustment factor has been applied directly to the occupation classification tables (ISCO and SINCO), so the numbers we observe in these tables reflect the average exposure of the tasks within an occupation for any of the classifiers. These will subsequently be re-aggregated to the microdata from the employment surveys. Figures 5, 6 and 7 illustrate the distribution of alpha (α) and beta (β) exposure for Mexico, Peru, and Chile with both human annotations and GPT-4 model predictions, before and after adjustments. Each graph presents two subgraphs, one for each exposure type (α on the left and β on the right), showing the percentage of tasks with exposure greater than a certain point on the x-axis and the percentage of occupations with that exposure on the y-axis.

Table 7: Adjusted occupation exposure by country (GPT-4 model)

	ISCO			SINCO	
	Chile/Peru (before)	Chile (after)	Peru (after)	Mexico (before)	Mexico (after)
α_{adj}	0.13	0.11	0.09	0.11	0.09
	(0.14)	(0.12)	(0.10)	(0.12)	(0.09)
β_{adj}	0.32	0.27	0.23	0.31	0.23
	(0.20)	(0.18)	(0.14)	(0.19)	(0.15)
ζ_{adj}	0.51	0.44	0.36	0.50	0.38
	(0.33)	(0.28)	(0.23)	(0.33)	(0.25)

*Standard deviations in parentheses

In our results, the beta (β) ratings are more affected in every case despite the uniform application of correction factors to both ratings. This result suggests a potential underestimation of the direct exposure but is affected by the job exposure with other occupations. These direct exposure metrics may require additional refinement to accurately reflect varying susceptibility to LLM influences within different labor markets.

Table 8, like Table 4, shows the intensity of occupational exposure to LLMs in the three countries, but now considering the adjusted exposure ratings. The adjustments applied reveal a nuanced landscape of occupational exposure, where the direct exposure coefficients (α) are generally lower compared to the initial estimates presented in Table 4. For example, the α_{10} values, representing the percentage of occupations with at least 10% of their tasks directly exposed to LLMs, have decreased notably, especially in Peru. This reduction suggests a more conservative estimate of the potential impact of

LLMs, which aligns with the less advanced technological infrastructure and lower digital skills prevalent in these regions. Similarly, the β coefficients, which account for LLM-powered applications, show a significant decrease, particularly in Peru, where the β_{10} value drops from 63.2% to 55.3%. This adjustment highlights the limited scope for LLM-powered applications to disrupt tasks in the Peruvian labor market compared to the other two countries. This is indicative of the broader challenges facing Peru in adopting such advanced technologies, likely due to its lower levels of formal employment and technological infrastructure.

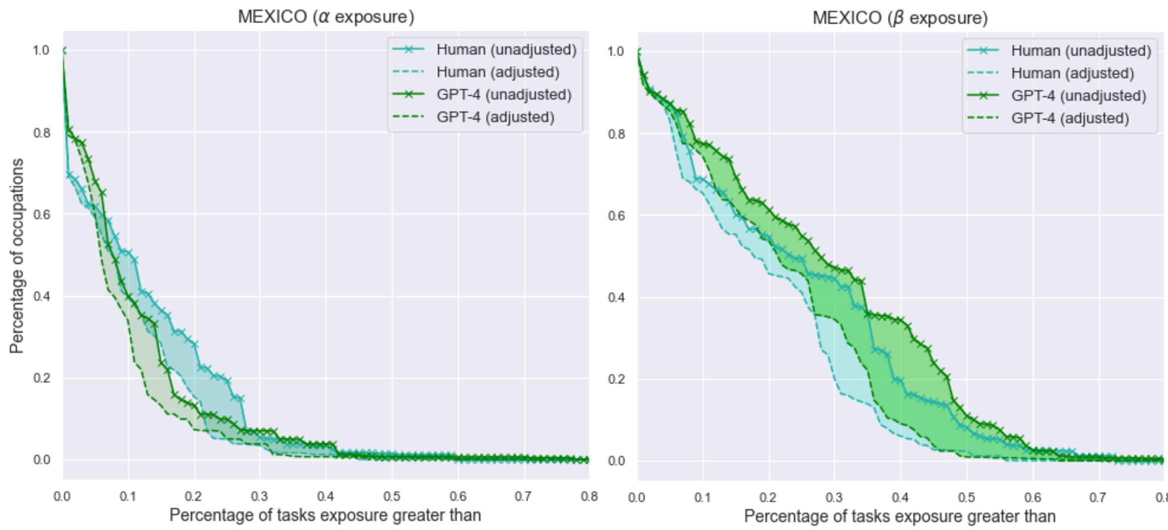
Interestingly, despite the adjustments, Mexico's β exposure remains relatively high, particularly for the GPT-4 model predictions. This suggests that, even after considering socio-economic adjustments, Mexico's labor market could experience significant changes due to the integration of LLM-powered applications, possibly driven by its more developed technological infrastructure compared to Peru.

Table 8: Adjusted exposure intensity by country (share of occupations with α/β exposure)

LLM's direct exposure		α_{10}	α_{20}	α_{30}	α_{40}	α_{50}	α_{60}	α_{70}	α_{80}	α_{90}	α_{100}
Human	Mexico	39.5	15.4	3.5	1.2	0.0	0.0	0.0	0.0	0.0	0.0
	Chile	37.8	18.7	6.9	5.3	1.4	0.0	0.0	0.0	0.0	0.0
	Peru	32.7	9.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GPT-4	Mexico	34.0	7.3	3.8	0.7	0.6	0.0	0.0	0.0	0.0	0.0
	Chile	37.9	9.7	4.9	2.2	1.6	0.1	0.0	0.0	0.0	0.0
	Peru	27.7	6.1	3.5	3.4	0.1	0.0	0.0	0.0	0.0	0.0
LLM-powered exposure		β_{10}	β_{20}	β_{30}	β_{40}	β_{50}	β_{60}	β_{70}	β_{80}	β_{90}	β_{100}
Human	Mexico	65.4	45.7	19.9	6.0	2.3	0.0	0.0	0.0	0.0	0.0
	Chile	63.5	45.7	30.1	13.9	4.9	1.5	0.0	0.0	0.0	0.0
	Peru	55.3	35.3	13.4	5.2	0.0	0.0	0.0	0.0	0.0	0.0
GPT-4	Mexico	74.4	53.7	34.5	9.0	0.9	0.5	0.0	0.0	0.0	0.0
	Chile	75.7	55.0	28.6	19.9	6.5	2.4	0.3	0.0	0.0	0.0
	Peru	76.5	32.0	13.7	5.8	3.5	0.0	0.0	0.0	0.0	0.0

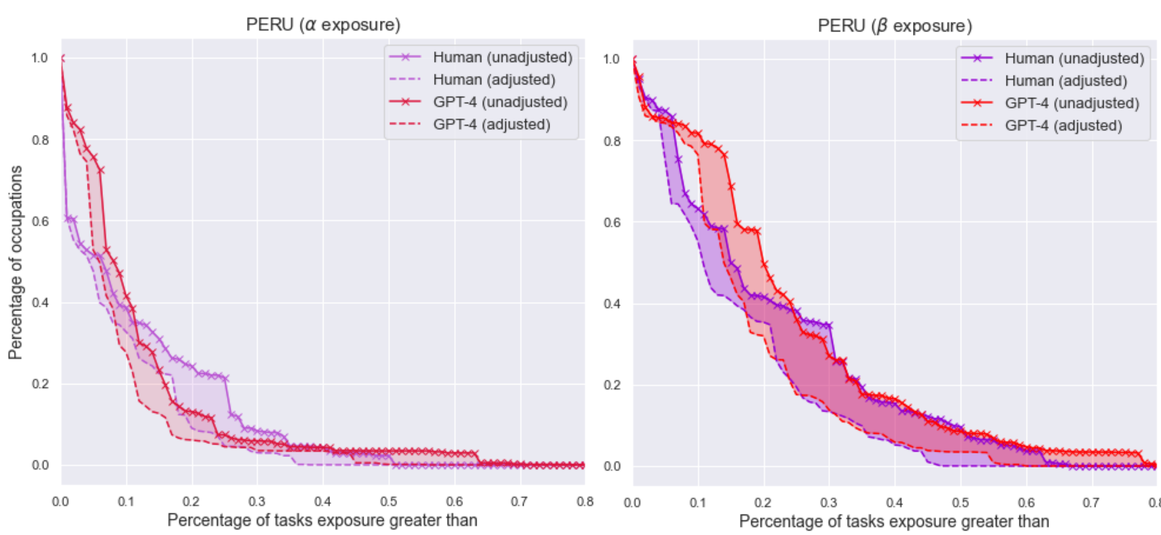
Note: The figures represent percentages. The employment surveys taken into account correspond to the end of the year 2022. Each value represents a minimum, meaning that, for example, the first value in the first row should be understood as 39.5% of the employed population having at least 10% of their tasks directly exposed to LLMs. This way of interpreting exposure data is adopted from the foundational study of this paper.

Figure 5: Percentage of occupations by level of adjusted task exposure (Human ratings)



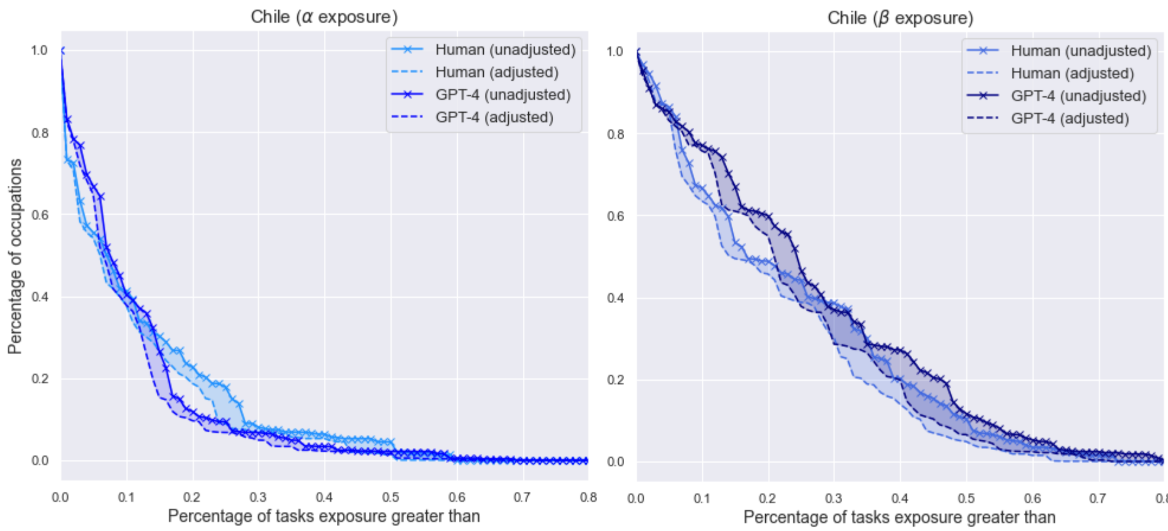
Another aspect worth noting is that while Peru's adjustment factor is larger than Mexico's, the applied adjusted alpha (α) and beta (β) exposure to each labor structure is similar in both countries. Interestingly, an observed effect is slightly more pronounced in Mexico (as seen in the right panel of Figure 5). This could indicate that, despite the greater adjustment in Peru, factors specific to the Mexican labor market may cause it to respond more to the adjustments when considering the potential implications of LLM-based technology exposure.

Figure 6: Percentage of occupations by level of adjusted task exposure (GPT-4 model)



In Chile, the lines for human and GPT-4 (unadjusted) followed each other before the correction, suggesting that human annotators' initial estimates closely mirrored the model's predictions. Post-adjustment, human, and GPT-4 lines dip slightly, indicating a consensus on reduced impact after considering the correction factors. The overlap between the adjusted and unadjusted lines for human and GPT-4 annotations implies that the changes made are relatively modest in scale. The adjustments do not significantly alter the estimated exposure distributions, which suggests that the original estimators were already well-aligned with the labor realities in Chile or that the correction factors employed did not dramatically shift the exposure estimates.

Figure 7: Occupational exposure before and after adjustment by country



The adjustment process in Chile's case is a meticulous and significant step, leading to a more conservative estimation of LLM exposure, particularly regarding the indirect one. The nuanced adjustment done with great care, is crucial, as it ensures that predictions align more closely with the realities of the Chilean labor market. It provides stakeholders with more accurate insights for policy-making and strategic planning in the face of technological disruption.

8 Conclusions

The paper shows the interaction between adopting emerging technologies, exemplified by Large Language Models (LLMs), and the existing labor structure in Latin America. Differences in labor structure, educational level, and access to technology among the countries analyzed influence the susceptibility and adaptability of labor markets. This interaction suggests that the impact of LLMs extends beyond simple task automation, affecting labor dynamics, employment distribution, and professional trajectories in a multifaceted and heterogeneous manner.

We model labor exposure to disruptive technologies using a new methodology but adjusting it to each country's specific realities. Conventional modeling approaches must be more accurate when applied to less developed labor markets. So, there is a need to develop flexible methodological frameworks incorporating local and sectoral variables and differences in technological absorption and labor market adaptability.

This work underscores the role of LLMs in shaping gender and educational dynamics within labor markets. Technology can increase the existing inequalities in the labor markets of Latin America. Evidence of greater exposure among women and those with higher education in some contexts raises critical questions about the future of work and equity in the digital age. Our findings also suggest that roles requiring high levels of human interaction, creativity, and complex decision-making, despite their exposure, may not be fully automatable in the near term. Hence, educational and training programs should promote equal opportunities to access the benefits of technology and enhance skills that complement LLM capabilities, such as critical thinking, emotional intelligence, and complex problem-solving.

Moreover, the distinction between formal and informal employment emerges as a crucial factor in determining the impact of LLMs, with significant implications for employment and development policies. The study challenges prior narratives about automation and offers a more nuanced view of how formal labor and income levels can mediate the relationship between technology and employment, highlighting the complexity of designing inclusive and effective adaptation and mitigation strategies. Therefore, as LLMs continue to evolve, their integration into the workplace will likely follow a nonlinear path, necessitating adaptive strategies that can accommodate both the opportunities and challenges posed by this technology.

This research underscores the urgent need for strategic planning and the development of public policies informed by a deep understanding of technological trends and their potential repercussions on labor markets. The call for adaptive policies and training strategies, focusing on digital skills, critical thinking, and adaptability, is a key takeaway. It reflects a sophisticated understanding of preparing for the future of work, essential for navigating the challenges and opportunities presented by LLMs and other emerging technologies. Furthermore, the relative insensitivity of direct exposure to the realities of the labor market underscores the need for a more granular, perhaps localized, approach to understanding how technology disrupts or integrates with existing work structures. This implies that the beta measure, which might be more related to shifts in human labor demand due to the adoption of LLMs, is a critical factor in assessing the resilience or vulnerability of occupations to technological advances. Such insights are invaluable for policymakers and economists tasked with preparing the workforce for the inevitable integration of LLMs and other advanced technologies, ensuring that interventions are tailored to the specific economic context and labor conditions of each country.

9 Appendix

Table 9: Adjustment indicators

Group	**Indicators**	**US**	**Chile**	**Mexico**	**Peru**
Education	Literacy rates (% of people ages 15 and above)	99.0	97.0	95.0	94.0
	Average years of schooling (people 25+ years)	13.8	10.6	8.9	9.7
	Government expenditure on education (% GDP)	5.4	5.6	4.6	3.9
	PISA Math score	465	412	395	391
Access and skills	Individuals using the Internet (% of population)	92.0	90.0	76.0	75.0
	Fixed broadband subscriptions (per 100 people)	37.6	22.7	19.5	9.3
	Firm-level technology absorption, 1-7 (best)	6.1	5.2	4.6	4.5
	Digital skills among population	5.3	4.3	3.8	3.4

Source: World Development Indicators, World Bank. Global Information Technology Report, World Economic Forum. .

Figure 8: ISCO, top 50 occupations with highest exposure to LLMs (Human ratings)

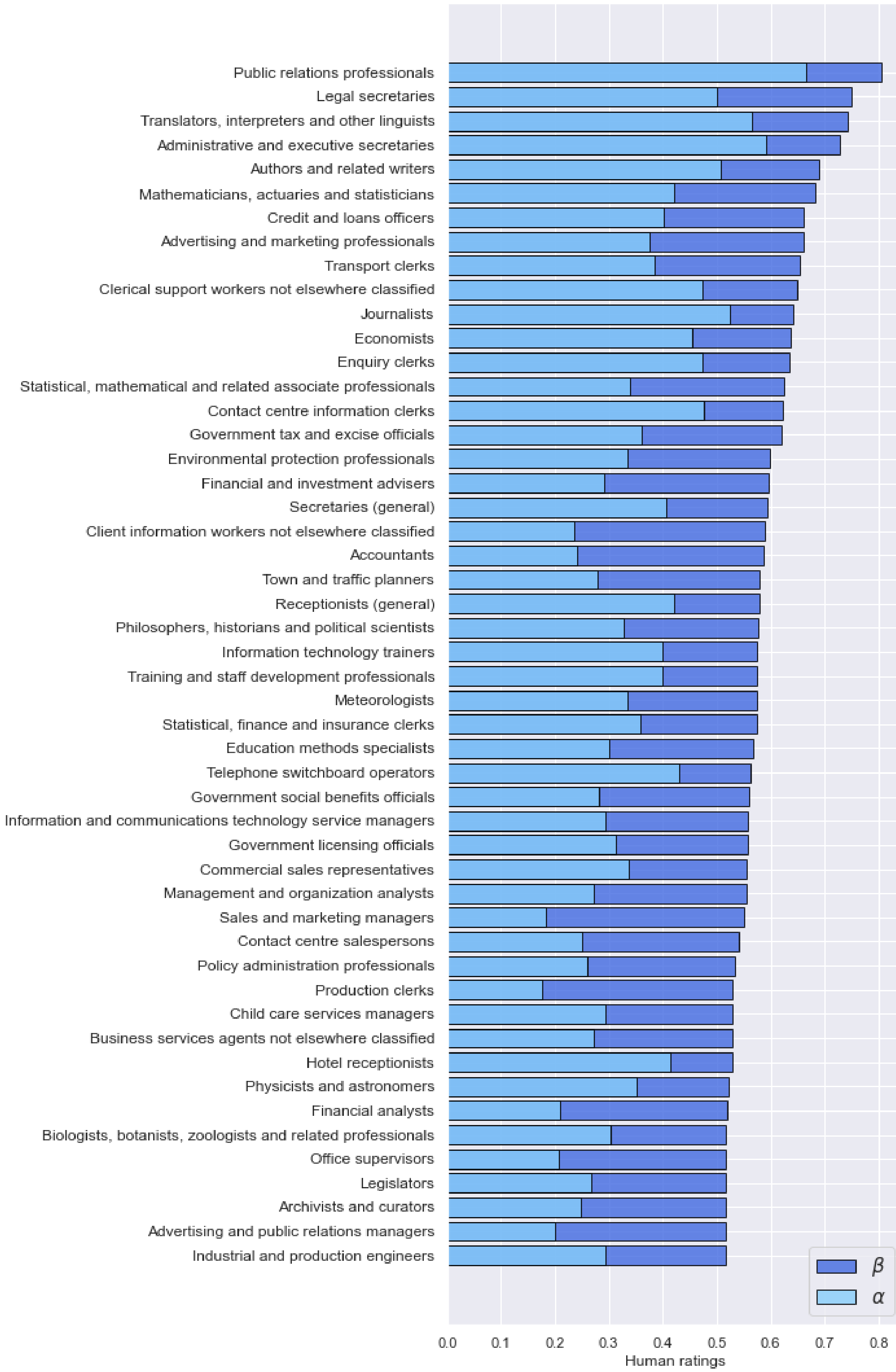


Figure 9: ISCO, top 50 occupations with highest exposure to LLMs (GPT-4 model)

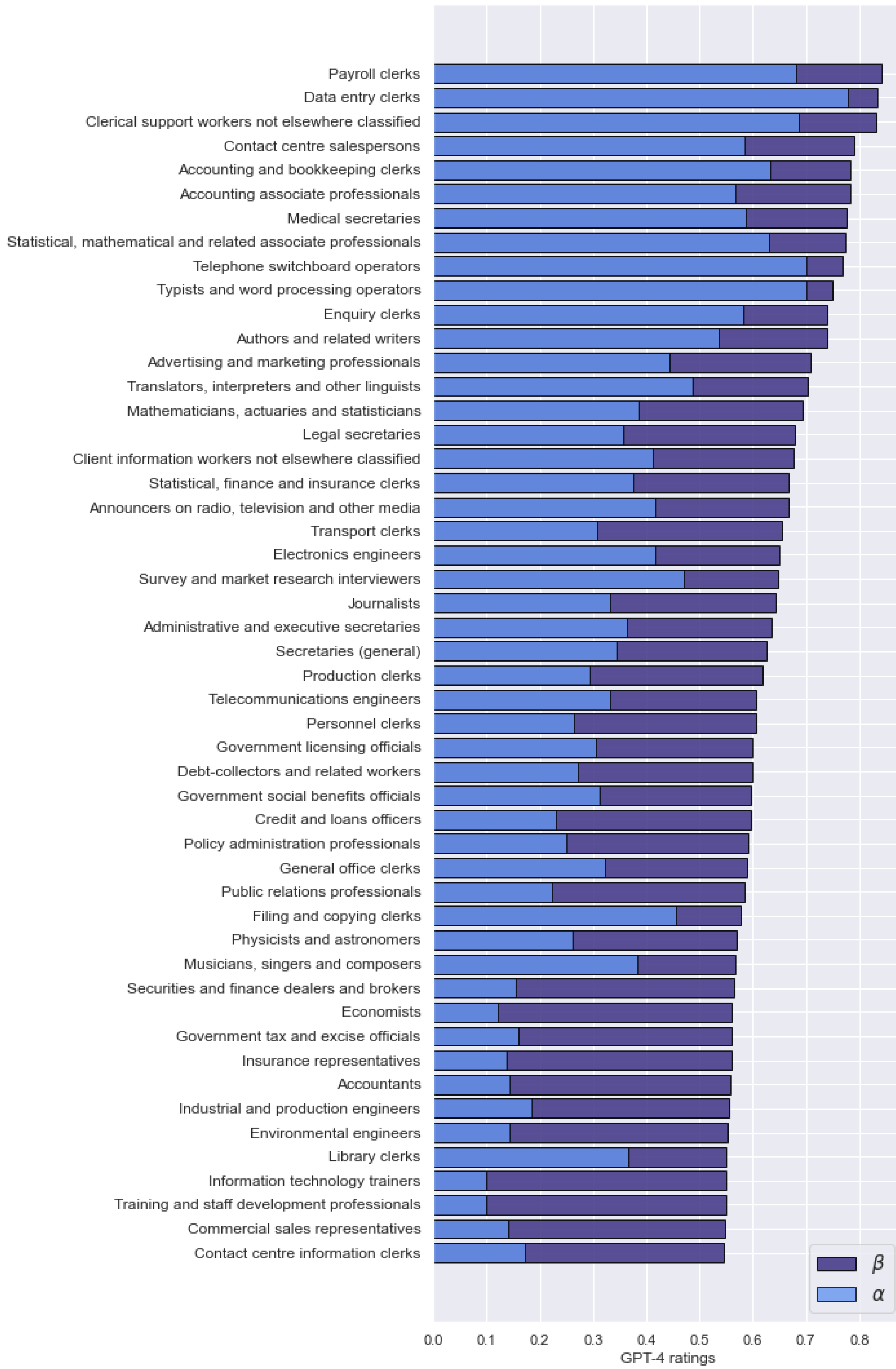


Figure 10: SINCO, top 50 occupations with highest exposure to LLMs (Human ratings)

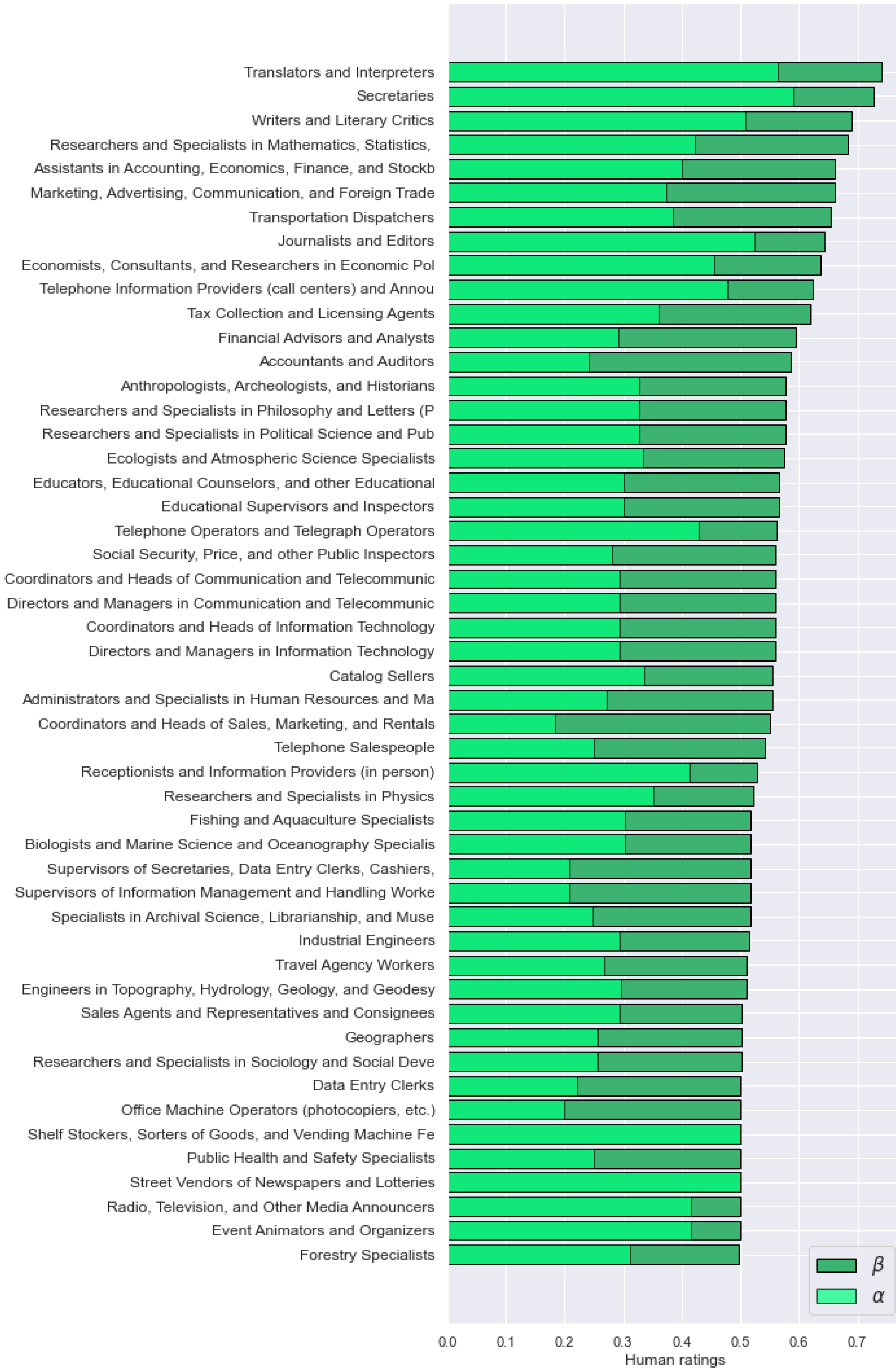
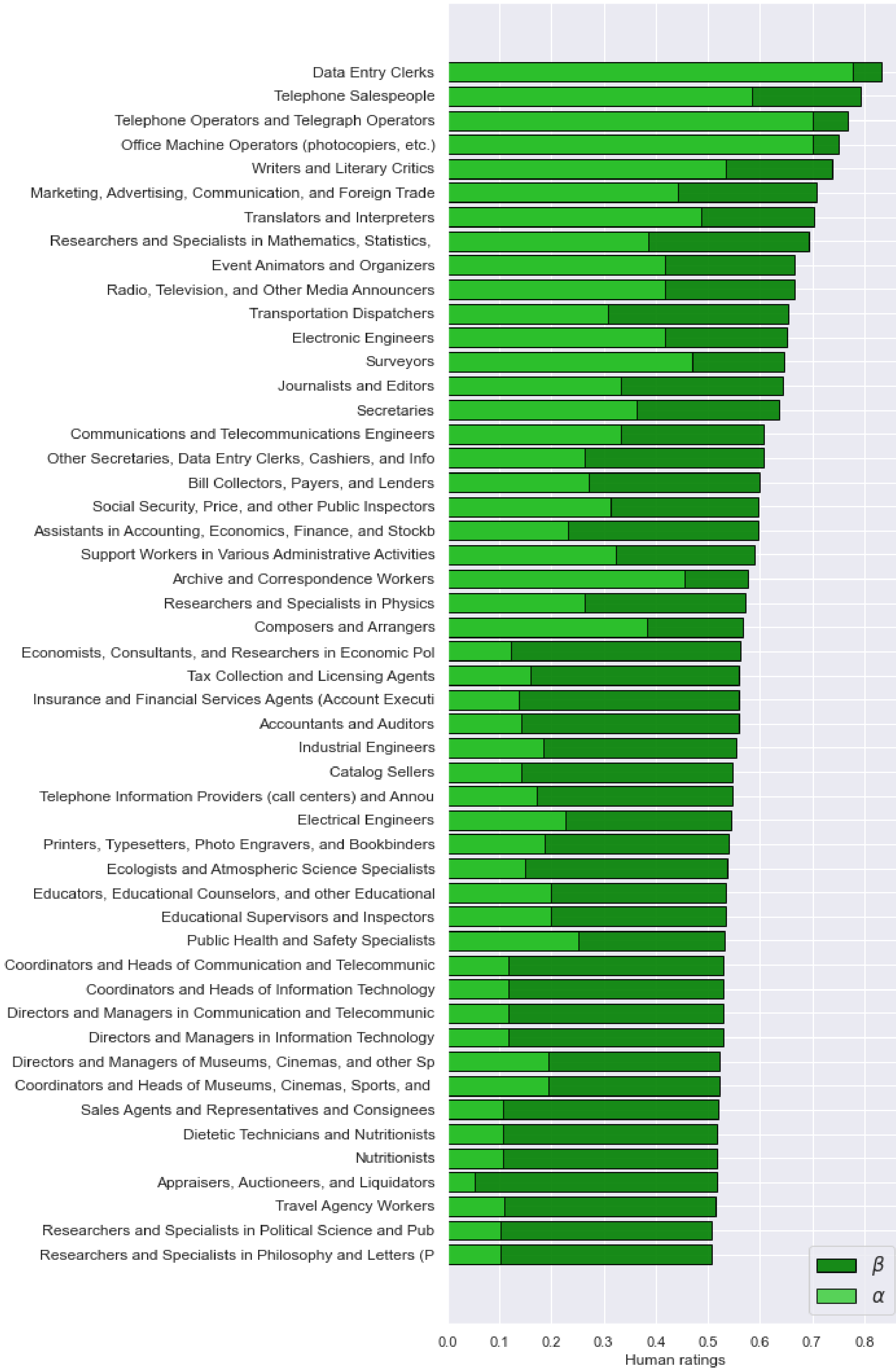


Figure 11: SINCO, top 50 occupations with highest exposure to LLMs (GPT-4 model)



References

Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of Economic Literature*, 40(1):7–72.

Acemoglu, D. and Restrepo, P. (2020). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 34(2):3–30.

Aghion, P., Antonin, C., and Bunel, S. (2019). Artificial intelligence, growth and employment: The role of policy. *Economie et Statistique / Economics and Statistics*, 510-511-512:149–164.

Autor, D., Chin, C., Salomons, A., and Seegmiller, B. (2022). New frontiers: the origins and content of new work, 1940–2018. NBER Working Paper w30389, National Bureau of Economic Research.

Bessen, J. E. (2002). Technology adoption costs and productivity growth: The transition to information technology. *Review of Economic Dynamics*, 5(2):443–469.

BLS (2012). Crosswalk between the international standard classification of occupations (isco-08) and the 2010 standard occupational classification (soc).

BLS (2023). Demographic characteristics (cps). Accessed: [02-20-2024].

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901.

Chen, L., Chen, X., Wu, S., Yang, Y., Chang, M., and Zhu, H. (2023a). The future of chatgpt-enabled labor market: A preliminary study in china.

Chen, Q., Ge, J., Xie, H., Xu, X., and Yang, Y. (2023b). Large language models at work in china's labor market.

Damioli, G., Van Roy, V., Vertesy, D., and Vivarelli, M. (2023). Ai technologies and employment: micro evidence from the supply side. *Appl Econ Lett*, 30(6):816–821.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT*, pages 4171–4186.

DLR (2023). Standard occupational classification system. https://dlr.sd.gov/lmic/soc_overview.aspx. Accessed: 2024-02-27.

Duch-Brown, N., Gomez-Herrera, E., Mueller-Langer, F., and Tolan, S. (2022). Market power and artificial intelligence work on online labour markets. *Res Policy*, 51(3):104446.

Dutta, S., Geiger, T., and Lanvin, B. (2017). Insight report: The global information technology report 2017. icts for inclusive growth. World Economic Forum. Cornell University, World Economic Forum, INSEAD.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2023). Gpts are gpts: An early look at the labor market impact potential of large language models.

Fatun, M. and Pazour, M. (2021). Modelling the impact of artificial intelligence on the labour market in czechia. *Central European Journal of Public Policy*, 15(2):31–42.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the penn world table. *American Economic Review*, 105(10):3150–3182.

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

George, A. S., George, A. H., and Martin, A. G. (2023). Chatgpt and the future of work: a comprehensive analysis of ai's impact on jobs and employment. *Partners Univ Int Innov J*, 1(3):154–186.

Ghosh, D., Ghosh, R., Roy Chowdhury, S., and Ganguly, B. (2024). Ai-exposure and labour market: a systematic literature review on estimations, validations, and perceptions. *Management Review Quarterly*. Published on 15 January 2024.

Guarascio, D., Reljic, J., and Stöllinger, R. (2023). Artificial intelligence and employment: a look into the crystal ball. GLO Discussion Paper No. 1333, GLO Discussion Paper.

Holm, J. and Lorenz, E. (2022). The impact of artificial intelligence on skills at work in denmark. *New Technology, Work and Employment*, 37(1):79–101.

Huseynov, S. (2023). Chatgpt and the labor market: Unraveling the effect of ai discussions on students' earnings expectations.

Iammarino, S., Rodriguez-Pose, A., and Storper, M. (2018). Regional inequality in Europe: evidence, theory and policy implications. *Journal of Economic Geography*, 19(2):273–298.

IDB (2017). Índice de Mejores Trabajos: Índice de Condiciones Laborales de América Latina. Nota Técnica IDB-TN-1326, Banco Interamericano de Desarrollo.

IDB (2023). Labor market observatory. https://observatoriolaboral.iadb.org/es/. Last accessed: [02-20-2024].

INE (2018). *CIUO 08.CL Clasificador Chileno de Ocupaciones*. Subdirección Técnica - Departamento de Infraestructura Económica - Sección de Nomenclaturas.

Kafka, K. I. (2021). *Economic Theory and Economic Reality: A Continuously Dialectic Relationship*, chapter 1, pages 1–26. IGI Global.

Kreitmeir, D. H. and Raschky, P. A. (2023). The unintended consequences of censoring digital technology – evidence from italy's chatgpt ban.

Liu, N. (2023). Assessing the potential impact of large language models on labor markets and business cycles in china: A preliminary study. *South Asian Journal of Social Studies and Economics*, 18(4):1–12.

Lu, L., Zhang, P., and Zhang, T. (2021). Leveraging "human-likeness" of robotic service at restaurants. *Int J Hosp Manag*, 94:102823.

MDSF (2022). Encuesta de caracterización socioeconómica nacional (casen) 2022. Informe técnico, Ministerio de Desarrollo Social y Familia, Gobierno de Chile, Santiago, Chile.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality.

Moll, B., Rachel, L., and Restrepo, P. (2022). Uneven growth: Automation's impact on income and wealth inequality. *Econometrica*, 90(6):2645–2683.

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., and Mian, A. (2023). A comprehensive overview of large language models.

O*NET (2023a). O*net online help. https://www.onetonline.org/help/online/summary. Last accessed: [02-20-2024].

O*NET (2023b). O*net online help. https://www.onetonline.org/help/onet/. Last accessed: [02-22-2024].

Reinhold, T., Brüggenjürgen, B., Schlander, M., et al. (2010). Economic analysis based on multinational studies: methods for adapting findings to national contexts. *Journal of Public Health*, 18:327–335.

S., X. (2023). Awesome llm interpretability. https://github.com/JShollaj/awesome-llm-interpretability. Software repository.

Shi, C. and Land, K. C. (2021). The data envelopment analysis and equal weights/minimax methods of composite social indicator construction: a methodological study of data sensitivity and robustness. *Applied Research in Quality of Life*, 16:1689–1716.

SIMS (2023). *Labor Market and Social Security Information System (SIMS). Inter-American Development Bank IDB.*

Trebat, T. J. (2012). Review of falling behind: Explaining the development gap between latin america and the united states by francis fukuyama. *Political Science Quarterly*, 127(1):150–152.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2023). Attention is all you need.

Wang, X., Sanders, H., Liu, Y., Seang, K., Tran, B., Atanasov, A., Qiu, Y., Tang, S., Car, J., Wang, Y., Wong, T., Tham, Y., and Chung, K. (2023). Chatgpt: promise and challenges for deployment in low- and middle-income countries. *The Lancet regional health. Western Pacific, 41, 100905.*

WBG (2024). World development indicators. https://data.worldbank.org/indicator. Accessed: 2024-02-27.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., and Fedus, W. (2022). Emergent abilities of large language models.

Widuckel, W. and Bellmann, L. (2023). Employment effects and changes in work organisation arising from ai. In *Work and AI 2030: challenges and strategies for tomorrow's work*, pages 195–201. Springer, Wiesbaden.

Xie, M., Ding, L., Xia, Y., Guo, J., Pan, J., and Wang, H. (2021). Does artificial intelligence affect the pattern of skill demand? evidence from chinese manufacturing firms. *Economic Modelling*, 96:295–309.

Xinyin, M. (2024). Awesome-efficient-llm. https://github.com/horseee/Awesome-Efficient-LLM. Software repository.

Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Yin, B., and Hu, X. (2023). Harnessing the power of llms in practice: A survey on chatgpt and beyond.

Zarifhonarvar, A. (2023). Economics of chatgpt: a labor market view on the occupational impact of artificial intelligence. *Journal of Electronic Business & Digital Economics*. Available at SSRN: https://ssrn.com/abstract=4350925 or http://dx.doi.org/10.2139/ssrn.4350925.

Zeileis, A. (2023). pwt10: Penn world table (version 10.x). R package version 10.01-0.

Zhou, G., Chu, G., Li, L., and Meng, L. (2020). The effect of artificial intelligence on china's labor market. *China Economic Journal*, 13(1):24–41.