GOSPODARKA NARODOWA

3(323)2025, 6–30
DOI: 10.33119/GN/203716
gnpje.sgh.waw.pl
SGH

Special Issue

The Polish Journal of Economics
Istnieje od / Published since 1931

Pawel Gmyrek
International Labour Organization, Switzerland

Janine Berg
International Labour Organization, Switzerland

David Bescond
International Labour Organization, Switzerland

Generative AI and Jobs: An Analysis of Potential Effects on Global Employment*

Generatywna sztuczna inteligencja a miejsca pracy – analiza potencjalnego oddziaływania na globalne zatrudnienie

Keywords:
artificial intelligence, AI, GPT4, ChatGPT, labour rights, workers' rights, employment, job quality

JEL classification codes:
J00, J01, J20, J21, J24

Article history:
submitted: November 11, 2024
revised: February 19, 2025
accepted: February 24, 2025

Abstract

This study presents a global analysis of the potential effects of generative AI on employment. Using the GPT-4 model, we estimate task-level exposure scores and assess their potential employment impacts globally and across country income groups. We find that clerical work is the only broad occupational category highly exposed to the technology, while other occupational groups such as managers, professionals and associate professionals exhibit much lower exposure levels. Consequently, the primary impact of generative AI is likely to be the augmentation of work rather than the full automation of occupations. Due to different occupational structures, employment effects vary across countries. In low-income countries, only 0.4 percent of total employment is potentially exposed to automation effects, compared with 5.5 percent in high-income countries. The effects are also highly gendered, with women more than twice as likely as men to be affected by automation. We find that 10.4 percent of employment in low-income countries has the potential to be augmented, compared with 13.4 percent in high-income countries. However, these estimates do not consider infrastructure constraints, which may significantly limit adoption in lower-income contexts.

Streszczenie

Badanie przedstawia globalną analizę potencjalnego wpływu generatywnej sztucznej inteligencji na zatrudnienie. Korzystając z modelu GPT-4, szacujemy wyniki na poziomie zadań i oceniamy ich potencjalne oddziaływanie na zatrudnienie w skali globalnej oraz w krajowych grupach dochodowych. Stwierdzamy, że praca biurowa jest jedyną spośród szerokich kategorii zawodowych w wysokim stopniu wystawionych na oddziaływanie technologii, podczas gdy inne grupy zawodowe, takie jak menedżerowie czy specjaliści, wykazują znacznie niższe poziomy ekspozycji. W związku z tym główny wpływ generatywnej sztucznej inteligencji będzie prawdopodobnie polegał na usprawnieniu pracy, a nie na pełnej automatyzacji

* This is an adaptation of an original work by the International Labour Organization (ILO). Responsibility for the views and opinions expressed in the adaptation rests solely with the author or authors of the adaptation and are not endorsed by the ILO.

Słowa kluczowe:
sztuczna inteligencja, SI, GPT4, ChatGPT, prawo pracy, prawa pracownicze, zatrudnienie, jakość pracy
Kody klasyfikacji JEL:
J00, J01, J20, J21, J24
Historia artykułu:
nadesłany: 11 listopada 2024 r.
poprawiony: 19 lutego 2025 r.
zaakceptowany: 24 lutego 2025 r.

zawodów. Ze względu na różne struktury zawodowe oddziaływanie na zatrudnienie różni się w zależności od kraju. W krajach o niskich dochodach tylko 0,4% całkowitego zatrudnienia może podlegać oddziaływaniu automatyzacji, w porównaniu z 5,5% w krajach o wysokich dochodach. Skutki automatyzacji są również silnie zróżnicowane ze względu na płeć, przy czym kobiety są ponad dwukrotnie bardziej podatne na wpływ automatyzacji niż mężczyźni. Ustaliliśmy, że 10,4 % zatrudnienia w krajach o niskich dochodach może zostać usprawniona w porównaniu z 13,4 % w krajach o wysokich dochodach. Szacunki te nie uwzględniają jednak ograniczeń infrastrukturalnych, które mogą znacząco ograniczać wdrażanie technologii w krajach o niższych dochodach.

Introduction

Each new wave of technological progress has reignited debates on automation and jobs. Current discussions on artificial intelligence (AI) echo those of the early 1900 s with the introduction of the moving assembly line, and those of the 1950s and 1960s following the introduction of early mainframe computers. While there have been some nods to the alienation that technology can bring by standardising and controlling work processes, most debates have centred on two opposing viewpoints: the optimists, who see new technology as a means of relieving workers from the most arduous tasks, and the pessimists, who warn of imminent job losses and mass unemployment.

What has changed, however, is the type of workers most affected. While technological advances throughout much of the 20th century primarily reduced the need for manual labour, the rapid progress of machine learning (ML) since the 2010s has centred on the ability of computers to perform non-routine cognitive tasks, potentially affecting white-collar or knowledge workers. In addition, these changes are occurring in an era of unprecedented global interconnectedness, making the exposure far greater than in earlier, factory-level applications. Yet despite these developments, the potential implications of AI remained largely abstract to most workers, even in highly developed economies—until very recently.

The launch of ChatGPT in November 2022 marked an important advance in the public's exposure to AI tools. In this new wave of technological transformation, machine learning models have left the laboratory and have begun interacting with the public, demonstrating their strengths and weaknesses in everyday use. The chat function dramatically shortened the distance between AI and the end user, while also providing a wide range of custom applications and innovations. Given these significant advancements, it is not surprising that concerns over potential job loss have resurfaced.

While it is impossible to predict how generative AI will evolve, both its current capabilities and future potential are central to debates on its impact on jobs. Sceptics argue that these systems are nothing more than "stochastic parrots"—powerful text summarisers incapable of genuine "learning" or producing original content, with limited prospects for general-purpose use and unsustainable computing costs [**Bender et al., 2021**]. In contrast, more recent technical studies testing the limits of the latest models point to growing capabilities in carrying out "novel and difficult tasks that span mathematics, coding, vision, medicine, law, psychology and more," and even to a general ability to produce responses exhibiting some early signs of "reasoning" [**Bubeck et al., 2023**]. Some assessments suggest that machine learning models, especially those based on large neural networks used by Generative Pre-trained Transformers (GPT), may develop into a general-purpose technology [**Goldfarb, Taska, Teodoridis, 2023**; **Eloundou et al., 2023**].[1] This would have multiplier effects on the economy and labour markets, as new products and services would likely spring from this technological platform.

1 The three main characteristics of general-purpose technologies are their pervasiveness, their capacity for continuous improvement, and their potential to spawn further innovation [**Jovanovic, Rousseau, 2005**].

As social scientists, we are not in a position to take sides in these technical debates. Instead, we focus on the demonstrated capabilities of generative AI, including custom-made chatbots with retrieval of private content (such as collections of documents or e-mails), natural language processing functions for content extraction, preparation of summaries, automated content generation, semantic text searches and broader semantic analysis based on text embeddings. Large language models (LLMs) can also be combined with other ML models, such as speech-to-text and text-to-speech generation systems, potentially expanding their interaction with different types of human tasks. Finally, their ability to interact with live web content through custom agents and plugins, along with the multimodal nature of generative AI (extending beyond text to image recognition and generation) makes it likely that this technology will expand into new areas, thereby increasing its impact on labour.

Building on these observations, this study seeks to add a global perspective to research on the exposure of occupations to generative AI. We focus on the concepts of "exposure" and "potential," which do not imply job losses, but rather serve to identify occupations that could be affected if the technology were to be fully deployed. The aim is to better understand the direction of possible changes, thereby facilitating the design of appropriate policy responses.

The analysis is based on 4-digit occupational classifications in the International Standard Classification of Occupations (ISCO-08) and their associated tasks. We use the GPT-4 model to estimate occupational and task-level scores of exposure to generative AI and subsequently link these scores to official ILO statistics to derive global employment estimates. We also apply embedding-based text analysis and semantic clustering algorithms to provide a better understanding of the types of tasks with high automation potential. We further discuss how automation and augmentation effects will strongly depend on a range of additional factors and specific country contexts.

We discuss the results of this analysis in the broader context of labour market transformation. We particularly focus on disparities in digital access across countries at different income levels and the risk that generative AI could exacerbate existing inequalities. We also explore gender-specific differences in potential exposure. The analysis does not attempt to account for new jobs created by technological change. As history shows, many occupations that are now widespread—such as social media managers or web designers—did not exist just a few decades ago. **Autor et al.** [**2021**] estimate that around 60 percent of US employment in 2018 was in jobs that had not existed in the 1940 s. We hope this research will contribute to policy debates on digital transformation in the world of work. While our analysis outlines potential implications for different occupational groups, the outcomes of technological transition are not pre-determined. Ultimately, it is human decisions—whether in workplaces, governments or institutions—that will shape the adoption of these technologies. We therefore hope our findings can support the development of policies that guide the transition in ways that benefit current and future generations.

Methods and data

There are two principal approaches to the analysis of automation of occupations [**Georgieff, Hyee, 2021**]. The first is to use data on job vacancies to understand how demand for specific skills evolves over time. Most studies using this approach harness data from online recruitment platforms [**Cammeraat, Squicciarini, 2021**; **Acemoglu et al., 2022**] to measure the frequency of references to AI (or to any other technology of interest) in the text of the job description. These references are then used as a proxy for the demand for specific skills and, by extension, a proxy for the rate of technological adoption at the enterprise level. This approach works well in countries with a high online presence in recruitment, though it does not always capture the industries affected as a result of subcontracting. The approach, however, is less well suited for a global study covering countries with less online presence, as most vacancies are not advertised on online platforms but recruited through other means of communication [**Georgieff, Hyee, 2021**].

The second approach is to focus on occupational structures, with the idea of estimating the automation potential of tasks or skills that make up a given job. The advantage of this method is that such occupational classifications can easily be linked to official labour market statistics, which is of particular importance for understanding global, regional and income-based differentials. This strand of literature is rich but frequently misunderstood, especially when it comes to communicating its findings to the public, as media interpretations tend to blur the distinction between automation potential and actual deployment in the workplace. For example, Frey and Osborne's [2017] influential study has been cited over 12,000 times, often for different types of doomsday pronouncements, even though the authors were clear about the distinction between potential and predicted effects. A range of studies follow this research tradition, attempting to calculate different types of occupational automation scores in OECD countries [Brynjolfsson, Mitchell, Rock, 2018; Felten, Raj, Seamans, 2018; Felten, Raj, Seamans, 2019; Acemoglu, Restrepo, 2020; Fossen, Sorgner, 2022] or even combining occupational and job posting data [Georgieff, Hyee, 2021]. Some authors have also taken up the challenge of producing better estimates for developing countries [Balliester, Elsheikhi, 2018], often by trying to link detailed occupational data and automation scores from the United States with less structured datasets available for lower-income countries [Carbonero et al., 2023].

Calculating occupational scores typically involves developing a rubric that defines a scoring method based on pre-established criteria to capture possible impacts from the technology of interest. The rubric is then applied to occupations or occupational tasks, to generate task- or occupation-specific scores. One of the challenges of this approach emerges in covering a wide range of technologies. While some tasks could be well suited for automation with a particular type of AI (for example, routine non-cognitive tasks in a factory setting), the same technology could be completely useless in other areas that require cognitive abilities. Attempting to cover the wide range of systems that currently fall into the AI category would require squeezing the assessments into one matrix of overall technological capabilities.

In this study, we focus exclusively on generative AI. We build on the method recently demonstrated by Eloundou et al. [2023] and replicated by Eisfeldt et al. [2023], which employs sequential API calls to the GPT-4 model to estimate task- and occupation-level automation scores for this technology. Their findings reveal a remarkable alignment between GPT-4 predictions and the judgements of AI experts. Applying a similar approach to ISCO-08, we conduct some 25,000 high-frequency API calls, fine-tuned at the level of occupational definitions, job titles, tasks and country income classifications. We then combine the resulting score matrix into global, regional and country-level employment estimates.

ISCO data on occupations and tasks

The ISCO-08 classification relies on a hierarchical structure, reflected in a system of digits. The highest 1-digit level covers 10 different types of occupational groups that can be further broken down into lower-level sub-groups, each time represented by an increasing number of digits. The most detailed, 4-digit level captures 436 occupations (See Table 1).

While the publicly available ILO statistics are at the 2-digit ISCO-08 level, the ILO holds a wealth of additional information from labour force surveys (LFS) and other national surveys in the ILO Harmonized Microdata collection. Its statistical repository contains microdata on employment at the 4-digit ISCO level for some 73 countries and 3-digit employment data for over 117 countries. This gives us access to a sizeable repository of harmonised survey data that can be used to analyse labour market information in a wide range of countries, including the detailed distributions of employment across occupations. The internal processing of LFS data also captures additional parameters of interest, such as variations in job titles that belong to each ISCO 4-digit category across different countries. As of 2023, there were some 7,500 jobs titles mapped to ISCO at 4-digits, which we also use as a robustness test for our analysis.

To build the principal data frame of tasks and occupations, we use Part III of the official ISCO-08 documentation, which provides detailed definitions and descriptions of tasks for each of the 436 ISCO-08 4-digit occupations [ILO, 2023b]. These tasks are devised with a global perspective and used to describe similar occupations identified in LFS, other household surveys and censuses, as well as non-statistical sources such as administrative records. In practice, they provide a common denominator for the variability of tasks within a given occupation. The number of tasks assigned to each ISCO-08 occupation ranges from 4 to 14. This data frame, with the complete list of ISCO-08 occupations and tasks, constitutes the starting point of our estimations.

Table 1. ISCO-08 Structure of occupations and tasks used in the study

ISCO-08 1-digit code	ISCO-08 1-digit full label	No. of distinct 1-digit codes	No. of distinct 2-digit codes	No. of distinct 3-digit codes	No. of distinct 4-digit codes	Total ISCO tasks	Total GPT tasks
0	Armed forces occupations	1	3	3	3	0	30
1	Managers	1	4	11	31	236	310
2	Professionals	1	6	27	92	751	920
3	Technicians and associate professionals	1	5	20	84	580	840
4	Clerical support workers	1	4	8	29	163	290
5	Service and sales workers	1	4	13	40	269	400
6	Skilled agricultural, forestry and fishery workers	1	3	9	18	141	180
7	Craft and related trades workers	1	5	14	66	503	660
8	Plant and machine operators, and assemblers	1	3	14	40	280	400
9	Elementary occupations	1	6	11	33	200	330
	Total	10	43	130	436	3,123	4,360

Source: Authors' own elaboration.

Prompt design and sequence

We develop a Python script that uses the OpenAI library to loop over the ISCO-08 task structure and conduct a series of sequential API calls to the GPT-4 model, using a range of prompts that we adjust for specific queries. Before predicting task-level scores, we run several initial tests of the GPT-4 model on the overall ISCO dataset, to determine its capacity for processing detailed occupational information. As a first step, we use the GPT-4 model to generate an international definition for each of the ISCO 4-digit codes, and to mark the level of skills required for each job, according to the same classification as used in ISCO-08 (1 for low-level skills, 4 for the highest). We design the first GPT-4 API prompt as follows:[2]

{"role": "system", "content": "You are a skills specialist[3]. You will provide job definitions based on a job title and ISCO code. Follow instructions closely."},

{"role": "user", "content": "Look at this ISCO code and job title and provide an international standard definition of this job:" + "Do not provide any other content, just the definition of some 100 words that describes what the job is about and which level of ISCO skills it requires (1-4)." + "ISCO code:" + str (ISCO_08) + "Job Title:" + str (Title)}

2 The prompt in the example is used as part of ChatCompletion.create () function in OpenAI Python library.

3 We use a "skills and AI specialist" for the system role, which is understood by GPT-4 as "someone who has in-depth knowledge of artificial intelligence technologies, such as machine learning, deep learning, and AI architectures like GPT. They understand how AI systems work, what they are capable of, and the limitations they have. Moreover, they can assess the skills required for certain jobs or tasks and evaluate the potential for these tasks to be automated by AI. They could provide insights into the extent to which AI might be able to replace or augment human roles in various fields, as well as advise on how people might need to adapt or acquire new skills to remain competitive in an increasingly automated job market."

By comparing the result with official ISCO-08 definitions, we examine the model's "understanding" of the ISCO-08 structure. The generated definitions are largely consistent with ISCO-08 and often contain more detailed information, which could potentially be a helpful feature in complementing some of the definitions so far created by humans specialised in this domain.

As the next step, we move our tests to the level of tasks. It is likely that the training data of GPT-4 included publicly available information from the O*NET occupations and their corresponding tasks, as well as the European Skills, Competencies and Occupations (ESCO) and ISCO occupational classifications at the 4-digit level, as the model demonstrates familiarity with the details of these different systems. Yet beyond simply reciting the content of these databases, GPT-4 seems capable of engaging in more complex exchanges and developing logical links between different types of occupational classifications and tasks—a surprising and useful feature that has been documented in other domains of application [**Bubeck et al., 2023**].[4]

We therefore adjust the prompt and request GPT-4 to generate a set of 10 typical tasks for each of the 436 ISCO-08 4-digit occupations, which we append to the main data frame alongside the official ISCO-08 tasks and definitions. Generating a uniform set of tasks across all occupations provides some analytical benefits. First, considering that GPT-4 has detailed ISCO-08 information already in its training data, the 10-task requirement helps to avoid a situation where the responses simply mirror what GPT-4 already knows about ISCO-08, but rather pushes the model to provide its "own" perception of tasks that belong to each occupation. Second, in the next step of prediction, a uniform set of tasks allows us to compute mean occupational automation scores based on a more statistically balanced list across all ISCO-08 occupations and compare the result to scores based on a varied number of tasks per occupation in the official ISCO-08 documentation.

Table 2 shows a detailed comparison of ISCO-08 tasks with the tasks derived by GPT-4, using the example of primary school teachers (ISCO 2341). As can be seen, the definition and tasks predicted by GPT-4 are comparable to the expert description provided by the ILO. Some of the GPT-generated tasks are more generic than those given in ISCO, which could partially be a result of our API request that requires 10 tasks, as opposed to the varied task numbers in the ISCO documentation.[5] We suspect that the model, by being prompted to provide a list of 10 tasks, might be deriving some additional tasks in order to complete the list, or it may be breaking down more complex tasks into simpler ones in order to match the required parameters of response.

Table 2. Sample of tasks and definitions from ISCO and predicted by GPT-4

ISCO-08	Definition ISCO	Definition predicted by GPT4
2341	Primary school teachers teach a range of subjects at the primary education level.	Primary School Teachers (ISCO code: 2341) are education professionals responsible for planning, preparing, and delivering lessons and activities to facilitate the cognitive, social, and emotional development of children aged between 5 and 12. They assess students' progress, provide a supportive and inclusive learning environment, and collaborate with parents and other education professionals to ensure the overall well-being and success of students. Primary School Teachers require ISCO skill level 4, which includes a bachelor's degree in education or a related field, and they must possess pedagogical skills, communication skills, and creativity in teaching methodologies.

4 We observe that some functionalities of the model could potentially be used to support specialised discussions on updating the standard classifications. For example, semantic clustering and content summarisation could be used to extract core tasks from a multitude of currently unprocessed country-level inputs, which is a tedious and overwhelming task for humans.

5 The full list of ISCO 4-digit descriptions by ISCO and GPT is available from the authors.

cont. Table 2

ISCO-08	Tasks ISCO	Tasks predicted by GPT4
2341	Preparing daily and longer-term lesson plans in accordance with curriculum guidelines;	Develop lesson plans and daily schedules for a variety of subjects at the primary education level;
	Instructing children individually and in groups, using various teaching methods and materials (e.g. computers, books, games), adapting to children's varying needs;	Teach basic skills such as reading, writing, and arithmetic to students using various teaching methods and tools;
	Maintaining discipline and good working habits in the classroom;	Establish and maintain classroom rules to create a positive and structured learning environment;
	Planning and conducting activities with children such as sporting activities, concerts and excursions;	Collaborate with other teachers and school staff to organise events, activities, and educational programmes;
	Assigning and grading class work and homework;	Provide individualised attention and support to students who are struggling or require additional help;
	Preparing, administering, and grading tests and assignments to evaluate children's progress;	Assess students' performance and progress through assignments, quizzes, and standardised tests;
	Observing and evaluating children's performance and behaviour;	Prepare students for transition to higher education levels by helping them develop necessary study habits and organisational skills;
	Supervising children during classes and at other times in the school day, including the playground during breaks;	Adapt teaching methods and instructional materials to meet students' varying needs and interests;
	Participating in staff meetings and other sessions, and conferring with other teachers concerning educational issues;	Participate in professional development sessions and workshops to stay updated with new teaching techniques and educational reforms;
	Preparing for and attending parent meetings to discuss children's progress and problems.	Communicate regularly with parents and guardians to discuss students' progress and address any concerns.

Source: Authors' own elaboration.

As the final step in the data generation process, we run another set of sequential API calls at the level of individual tasks. We request GPT-4 to generate a score between 0 and 1, representing potential automation with GPT-based technology for each task in the ISCO task collection and in the GPT-generated set of tasks. We provide the occupation's ISCO 4-digit code, specify whether the job is located in a high-income or a low-income country and ask the model to justify its decision. After several rounds of fine-tuning, we settle on the following prompt:

> *{"role": "system", "content": "You are a skills and AI specialist." + "You will provide a score of potential automation with GPT technology for a given task. Follow instructions closely."},*
>
> *{"role": "user", "content": "Look at this job task:" + str (Tasks_GPT) + "It is related to ISCO code:" + str(ISCO_08) + "Provide a score of potential automation of this task with GPT technology, given that the job is located in a high [low] income country:" + "The score should range 0–1. Provide a score in one line, and a justification in next line. Do not provide any other commentary, only the score and justification." + "Do not give any ranges just one score for each task."}*

This exercise results in an ISCO-08 4-digit level data frame, with automation scores predicted for each ISCO-08 tasks and for GPT-predicted tasks, with separate scores for low- and high-income countries. Each of the task-level scores is accompanied with a short justification generated by GPT-4. Table 3 shows the results for primary school teachers (ISCO-08 2341) in a high-income country.

Table 3. Sample of task-level scores (high-income country context)

ISCO-08	ISCO Tasks	Score	Justification by GPT	GPT-predicted Tasks	Score	Justification by GPT
2341	Preparing daily and longer-term lesson plans in accordance with curriculum guidelines;	0.60	GPT technology can aid in the generation of lesson plans and offer suggestions based on curriculum guidelines, but a human teacher's expertise and nuanced understanding of students' needs will still be important for crafting effective plans.	Develop lesson plans and daily schedules for a variety of subjects at the primary education level	0.70	GPT technology can help in generating content and providing suggestions for lesson plans, but human guidance is still required for contextual understanding and tailoring the lessons to suit the specific needs of the students at the primary education level.
2341	Instructing children individually and in groups, using various teaching methods and materials (e.g. computers, books, games), adapting to children's varying needs;	0.30	GPT technology can assist in providing instructional materials and adaptive learning approaches, but the physical presence, emotional connection, and real-time adaptability of a human teacher are essential for effectively teaching young children.	Teach basic skills such as reading, writing, and arithmetic to students using various teaching methods and tools	0.30	GPT technology can assist in teaching basic skills by providing content and exercises, but it cannot fully replace a human teacher needed for personalised guidance, classroom management, and social-emotional development.
2341	Maintaining discipline and good working habits in the classroom;	0.15	GPT technology can assist in monitoring and providing feedback, but it cannot fully automate maintaining discipline and good working habits in the classroom because human interaction and physical presence are essential for effective discipline and enforcing rules.	Establish and maintain classroom rules to create a positive and structured learning environment	0.20	Establishing and maintaining classroom rules involves understanding the unique social dynamics of a specific group of students, which GPT technology may struggle to assess comprehensively and adapt to.
2341	Planning and conducting activities with children such as sporting activities, concerts and excursions;	0.25	GPT technology can contribute to idea generation and planning for activities, but it cannot physically conduct activities or interact with children effectively in real-life situations.	Collaborate with other teachers and school staff to organise events, activities, and educational programmes.	0.55	GPT technology can aid in planning, communication, and organisation, but human interaction and collaboration with other staff members is still essential to successfully implement events and programmes.
2341	Participating in staff meetings and other sessions, and conferring with other teachers concerning educational issues;	0.15	GPT technology can potentially assist in identifying meeting agendas, summarising discussion points, and providing insights on issues, but it cannot replace human interaction and collaboration required in staff meetings and conferring with other teachers.	Participate in professional development sessions and workshops to stay updated with new teaching techniques and educational reforms	0.30	GPT technology can partially provide information and resources for professional development, but human engagement and interaction are essential for proper learning and understanding of new teaching techniques and educational reforms.

Source: Authors' own elaboration.

Assessment of predictions, robustness tests and the bounds for analysis

We approach our predicted task-level scores with scepticism. However, following a manual review, at a large scale of 3,123 tasks across all ISCO-08 occupations, we find no evidence of bias in one direction: highly automatable tasks such as typing consistently get a high score (above 0.7), whereas tasks requiring manual dexterity consistently get low scores. Moreover, GPT-4 provides a reasonable written explanation of differences across the scores attributed to similar categories (see Table 3).

We conduct an additional test of scoring consistency across tasks (whether the model predicts a similar level of scores for different types of tasks across multiple runs, based on the same input) and score variability at task level (the range of scores predicted for the same task across multiple runs, based on the same input) by making 100 predictions for five tasks randomly selected from all tasks on the ISCO-08 list. We then calculate the mean score and standard deviation (SD) for each of the tasks, as shown in Table 4. The scores are highly consistent across different types of tasks, with SDs not exceeding 0.05. This is likely because the random element in scoring is lower than what it would be in the case of scoring by human respondents, who typically struggle with score uncertainty (e.g. whether a score of 0.2 would be more adequate than 0.15 or 0.25) and tend to have greater variability of opinions [**Gmyrek et al., 2025**].

Table 4. Test of score consistency (100 task-level predictions)

ISCO_08	Task	Mean ± SD
5141	Cutting, washing, tinting and waving hair;	0.06 ± 0.03
8122	Operating and monitoring equipment which cleans metal articles in preparation for electroplating, galvanising, enamelling or similar processes;	0.11 ± 0.04
2264	Recording information on patients' health status and responses to treatment in medical records-keeping systems, and sharing information with other health professionals as required to ensure continuing and comprehensive care;	0.64 ± 0.05
3313	Verifying accuracy of documents and records relating to payments, receipts and other financial transactions;	0.73 ± 0.05
4411	Maintaining library records relating to the acquisition, issue and return of books and other materials.	0.73 ± 0.05

Source: Authors' own elaboration.

As a parallel robustness test, we use a slightly modified prompt to generate occupational-level scores for over 7,500 job titles found in various national labour force surveys, which aggregate to the 436 ISCO 4-digit occupations. These jobs do not have detailed task descriptions, but a comparison between their occupation-level scores and the mean occupational scores derived from task-level data reveals close alignment. In other words, whether predictions are generated from individual tasks that aggregate to occupations or from a much larger pool of job titles, GPT-4 produces consistent scores of automation potential.

This obviously has to do with its training data, both in terms of originally ingested textual sources and further human-based fine-tuning of the model. Given the proximity between GPT-4 scores and human-based scoring by AI experts on task-level questions, as demonstrated in **Eloundou et al.** [**2023**], we believe our exercise is likely to be estimating the upper bound of exposure to GPT. Several factors account for this.

First, as shown by **Karger et al.** [**2023**], tech experts tend to overstate technological capacities and risks in questions concerning broader applications. We believe this is also likely to be true when it comes to full-scale deployment of GPT technology in the workplace, in particular at a level that would allow for full elimination of the human component. This was well illustrated in earlier automation studies, which often assigned high scores of displacement potential to routine tasks and even entire occupations, including in garment production. In practice, however, work continues to be performed by humans due to the challenges of handling highly pliable fabrics and the complexity of skills and dexterity involved in the stitching process [**de Mattos et al., 2020**]. Because research studies on automation were most certainly part of its training data, GPT is likely to reflect techno-optimism and overstate some task-level scores. Moreover, GPT-generated scores do not capture job-level task variation, which can lower occupational-level scores [**Arntz et al., 2017**].

Second, our prompts focus on technical feasibility but omit some important determinants of technological diffusion. These include contextual factors such as access to electricity and the internet in lower-income countries, as well as market dynamics such as the relative cost of labour and technology, levels of digital literacy, and access to finance. Third, although we generated predictions separately for high- and low-income countries, the differences between the two sets of scores were too small to justify retaining both datasets. For the purposes of this initial analysis, we use high-income country scores, with the understanding that this contributes to estimating the upper bound of global exposure, since technological deployment faces additional barriers in lower-income countries. Nonetheless, this approach offers an initial approximation of the global distribution of exposure, which can later be refined by more detailed, context-specific studies.

Since ISCO-08 tasks and those generated by GPT-4 do not correspond directly, we cannot compare automation scores at the individual task level in the two datasets. Instead, we focus on the occupation level and examine the similarity of the occupational scores, calculated as an arithmetic mean of the task-level scores for each ISCO-08 4-digit occupation. We find that, in general, scores based on tasks previously generated by GPT tend to be higher than those attributed to tasks coming directly from ISCO-08. We attribute this differential to the more refined character of ISCO-08 tasks, compared to the more generic tasks generated by GPT. When confronted with the more complex tasks from the ISCO-08 classification, GPT-4 tends to assign lower automation scores. By contrast, its own, more general task descriptions invite higher scores. We treat ISCO-08 scores as the basis for further analysis since they are directly linked to an international standard and associated ILO employment statistics.[6]

Finally, a classic challenge in analysing occupational tasks lies in estimating the share of time needed to execute individual tasks within a given occupation [**Carbonero et al., 2023**]. Time distributions likely vary across country contexts, but labour force and other survey data lack sufficient detail to allow for country-level distinctions. The problem of attributing time weights across task-level scores is not exclusive to our study and typically appears in the construction of composite indicators for technology and occupations [e.g. **Autor, Dorn, 2013**; **Brynjolfsson, Mitchell, Rock, 2018**]. One reason many automation studies focus on the United States is that the level of detail in O*NET data facilitates such estimations. For our global study, we adopt the most straightforward solution: applying equal weights to each task-level component within occupations.

Results

Exposure levels

To further address any potential score imprecision, we establish generous margins for classifications in the calculations that follow, focusing on the extremes of the scoring scale, and interpret most results at a higher level of aggregate ISCO-08 1-digit categories. Given the range of the estimated index (0-1), we consider scores below 0.25 as representing very low exposure and those between 0.25 and 0.5 as low exposure. Medium exposure is captured in scores with the range of 0.5–0.75, while tasks with scores above 0.75 are considered as highly exposed. The same cut-off points are applied to the occupation-level scores, calculated as a mean score of the tasks that belong to each occupation.

We then group and sort all tasks with the highest exposure scores and use the OpenAI Ada model to assign embeddings for each task through sequential task-level API calls.[7] We perform semantic clustering of the tasks, based on the K-Means algorithm and a visual inspection of the results, which suggests five principal thematic clusters. Once the clusters have been attributed, we engineer another set of API calls to GPT-4 and request the model to provide the common semantic denominator for each thematic cluster. Table 5 presents the result of this exercise, with the corresponding tasks in each cluster and their individual scores.

[6] Since ISCO-08 documentation does not provide any tasks for the first major group of "Armed Forces Occupations," we use GPT-predicted tasks and scores to include this category in further analysis. In addition, ISCO-08 does not provide tasks for occupations with codes 1439 (Services Managers Not Elsewhere Classified), 3139 (Process Control Technicians Not Elsewhere Classified), 3435 (Other Artistic and Cultural Associate Professionals), 5249 (Sales Workers Not Elsewhere Classified), 7319 (Handicraft Workers Not Elsewhere Classified) and 8189 (Stationary Plant and Machine Operators Not Elsewhere Classified). As the catch-all character of these few occupations does not permit the assigning of specific tasks, we drop them from the final analysis.

[7] Embeddings are a vectoral high-dimensional representation of the text generated by an LLM. Standard Ada embeddings have 1,536 dimensions.

Table 5. Tasks with high automation potential clustered into thematic groups

Thematic Group	Sample Tasks	Score
Administrative and Communication Tasks	Making appointments for clients;	0.80
	Dealing with routine correspondence on their own initiative.	0.80
	Arranging to buy and sell stocks and bonds for clients;	0.80
	Photocopying and faxing documents;	0.80
	Addressing circulars and envelopes by hand.	0.80
Customer Service and Coordination	Issuing tickets for attendance at sporting and cultural events;	0.80
	Selecting area for fishing, plotting courses and computing navigational positions using compass, charts and other aids;	0.80
	Taking reservations, greeting guests and assisting in taking orders;	0.80
	Determining most appropriate route;	0.80
	Making and confirming reservations for travel, tours and accommodation;	0.85
Data Management and Record Keeping	Maintaining records of stock levels and financial transactions;	0.80
	Initiating records for newly appointed workers and checking records for completeness;	0.85
	Importing and exporting data between different database systems and software;	0.80
	Operating electronic or computerized control panel from a central control room to monitor and optimise physical and chemical processes for several processing units;	0.80
	Preparing invoices and sales contracts and accepting payment;	0.80
Information Processing and Language Services	Taking dictation and recording other matter in shorthand;	0.80
	Translating from one language into another and ensuring that the correct meaning of the original is retained, that legal, technical or scientific works are correctly rendered, and that the phraseology and terminology of the spirit and style of literary works are conveyed as far as possible;	0.80
	Converting information into codes and classifying information by codes for data-processing purposes;	0.80
	Keying in processing instructions to programme electronic equipment;	0.80
	Recording, preparing, sorting, classifying and filing information;	0.90
Providing Information and Responding to Inquiries	Responding to inquiries about problems and providing advice, information and assistance;	0.80
	Describing and providing information on points of interest and exhibits and responding to questions;	0.80
	Preparing and reporting short-term or long-term weather maps, forecasts and warnings relating to atmospheric phenomena such as cyclones, storms and other hazards to life and property and disseminating information about atmospheric conditions through a variety of media including radio, television, print and the Internet;	0.80
	Determining customer requirements and advising on product range, price, delivery, warranties and product use and care;	0.80
	Responding to inquiries concerning services provided and costs for room and equipment hire, catering and related services;	0.80

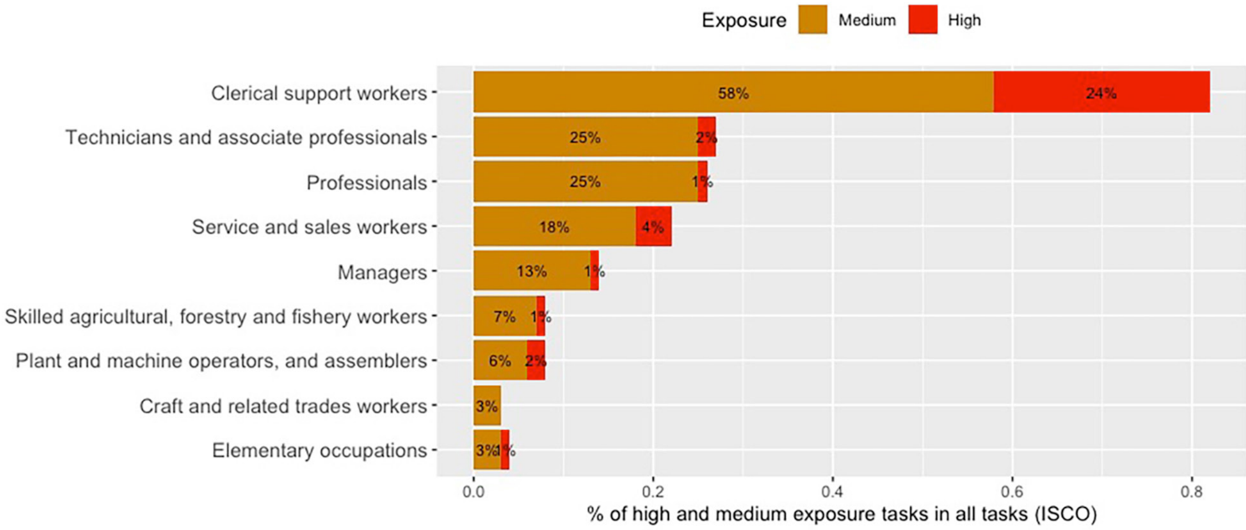
Notes: Clustering relies on semantic proximity, based on K-means clustering of task embeddings. Cluster names have been assigned by sending all tasks within a cluster to GPT4 API and requesting a common group heading and identification of similarities.

Source: Authors' own elaboration.

As the next step, we calculate the share of tasks with high and medium exposure in each ISCO 1-digit grouping. Figure 1 reveals in stark terms the degree of exposure among clerical support workers, where some 24 percent of tasks fall into the highly exposed category. If we also account for tasks with medium-level exposure (58 percent of all tasks), 82 percent of clerical job tasks are exposed at an above-average level. This stands in contrast to other occupational groups, in which the highest share of highly exposed tasks ranges between 1 and 4 percent, and where the medium-exposed tasks do not exceed 25 percent.[8] Even assuming large margins of error, the result is still striking.

[8] Armed forces are absent from the figure since they do not have any tasks scored at the level of medium and high exposure.

Figure 1. Tasks with medium and high GPT exposure, by occupational category (ISCO 1-digit)



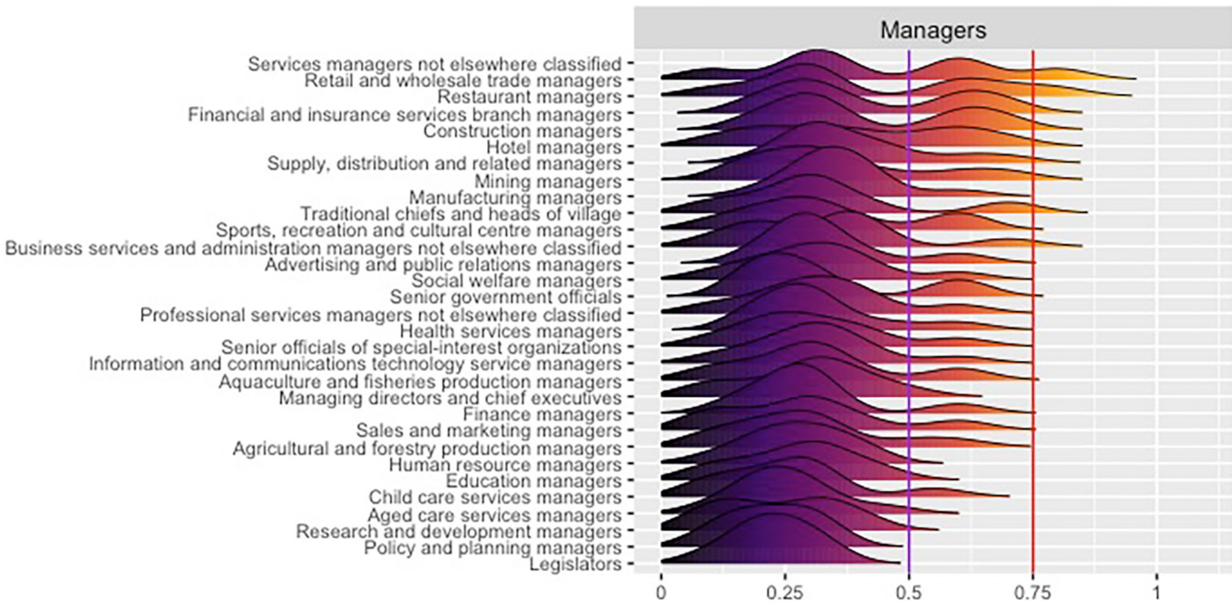
Source: Authors' own elaboration.

Automation vs augmentation: distribution of scores across tasks and occupations

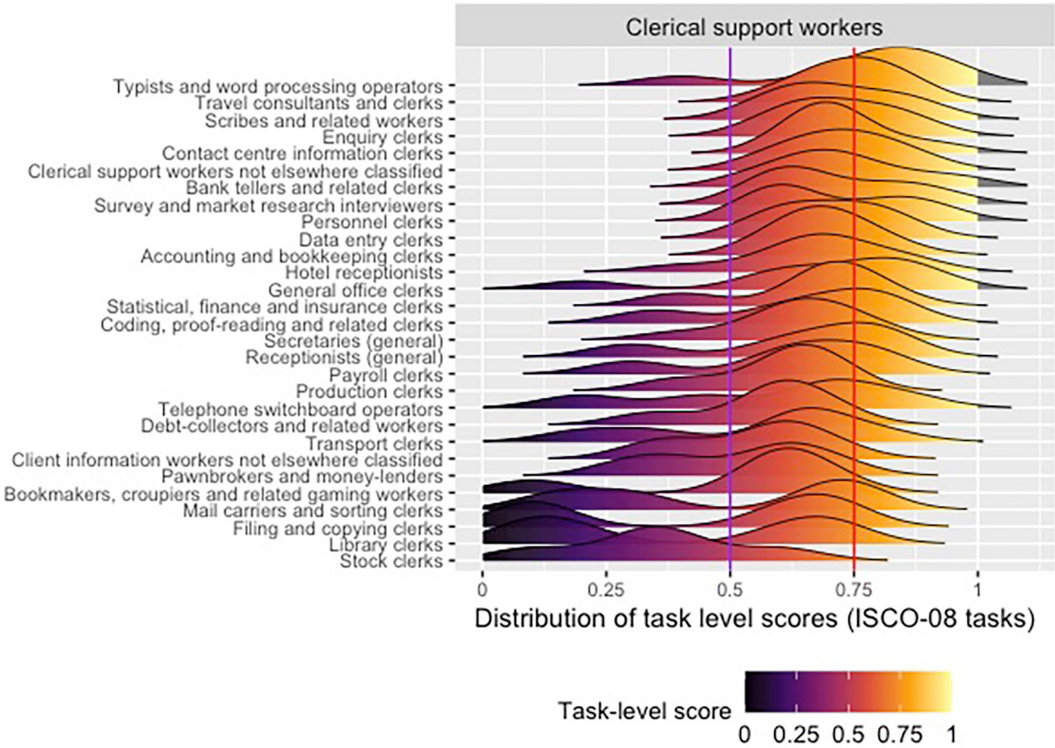
In this next section, we analyse how exposure to GPT-like technology could potentially affect occupations. Will the technology replace most tasks within an occupation, provoking job losses? Or could it be used to automate the more routine tasks, leaving time for more gratifying activities?

To probe these questions, we turn to the analysis of the distribution of tasks for each of the 4-digit ISCO-08 occupations. Figure 2 provides a visual representation of task scores for the ISCO 1-digit group of managers and clerical support workers. It shows that for the manager category, most occupations have a task-level score distribution somewhere on both sides of the medium exposure line of 0.5, with more tasks falling into low-level exposure. In contrast, for clerical support workers, many occupations have an entire task distribution that falls to the right of the medium exposure threshold of 0.5.

Figure 2. Distribution of task-level scores by ISCO 4d, managers and clerical support workers



cont. Figure 2



Source: Authors' own elaboration.

To determine whether the technology has greater potential for automation or augmentation across all ISCO-08 4-digit occupations, we use a method similar to that applied by Carbonero et al. [2023]. Considering each occupation as a collection of tasks with varying levels of exposure to a given technology, we focus on two principal parameters of the task score distribution: (i) the mean score for a given occupation, and (ii) the standard deviation (SD). Jobs with a high mean score and low standard deviation fall into the category of high automation potential, as the majority of the occupation's tasks have high exposure scores. Jobs with high augmentation potential are at the other extreme as they have a low occupation-level mean score but a high standard deviation of the task scores. These jobs are composed of some tasks that are difficult to automate, and others that can be automated more easily. In such cases, technology is likely to have an augmenting effect, taking away some of the more exposed tasks, but still requiring the human element for the overall performance of the job (Table 6).

Table 6. Grouping of occupations based on task-level scores

	Low Mean	High Mean
High SD	Augmentation potential	The big unknown
Low SD	Not affected	Automation potential

Source: Authors' own elaboration.

To ensure a clear separation of the occupations with high augmentation and automation potential, we apply a simple formula focused on the extremes of this distribution. Let μ_i and σ_i denote the mean and standard deviations of the task-level scores for occupation i respectively. We define an occupation to have "augmentation potential" if the following conditions are satisfied:

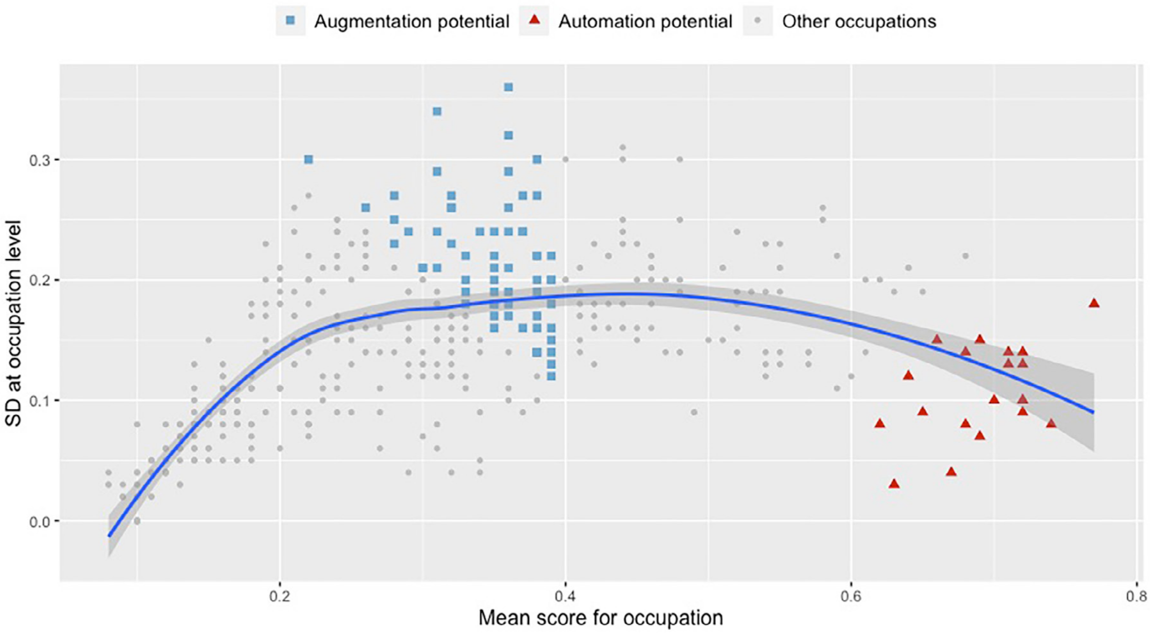
$$0.4 > \mu_i \text{ and } \mu_i + \sigma_i > 0.5 \tag{1.1}$$

Similarly, an occupation is said to have "automation potential" if it fulfils these criteria:

$$\mu_i > 0.6 \text{ and } \mu_i - \sigma_i > 0.5 \tag{1.2}$$

Figure 3 provides a visual representation of this grouping. The blue trend line (polynomial fit) illustrates the relationship between the two plotted variables: the occupation-level mean on the horizontal axis and the SD of task-level scores on the vertical axis. Close to the start of the axes, mean scores and SD grow simultaneously, but the scores in this group have a low overall mean and hence low exposure. As the SD begins to plateau in the middle section around 0.2, the mean scores reach levels closer to 0.5, meaning that the sum of these two components starts to significantly exceed the middle exposure threshold of 0.5. As the SD begins to drop to some 0.1, the occupational scores arrive at the level of 0.6 and higher, meaning that the difference between the mean and the SD would still put such scores well above the middle exposure limit of 0.5. These definitions allow us to identify professions with high augmentation potential and those with high automation potential.

Figure 3. Augmentation vs automation potential at occupational level



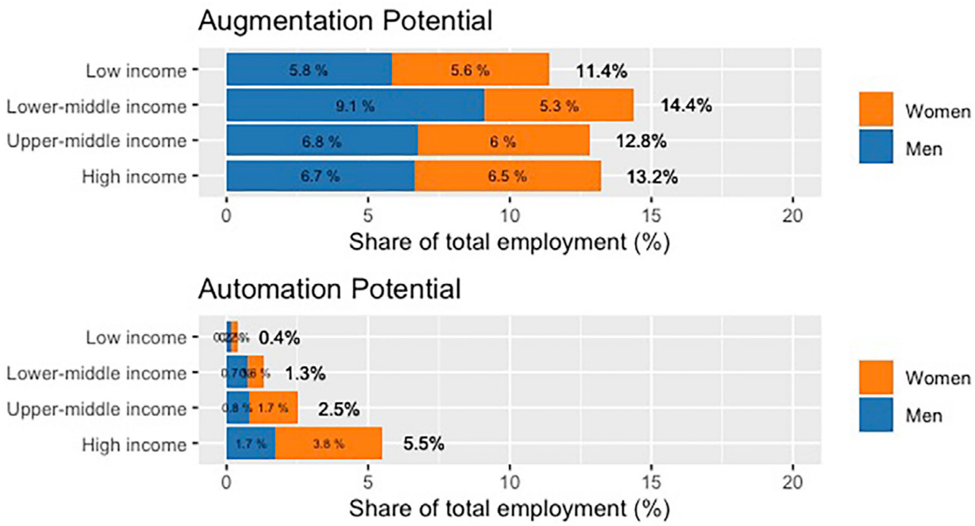
Source: Authors' own elaboration.

Exposed occupations as a share of employment: global and income-based estimates

Now that we know which occupations have the greatest potential for automation and augmentation from generative AI technology with similar properties as GPT, we can proceed with deriving employment estimates globally and by country income groups. To do this, we use the ILO Harmonized Microdata collection, which allows extracting detailed country-level employment information. We use microdata for 59 countries that report 4-digit microdata in ISCO-08 format: eight low-income countries (LIC), 24 lower-middle-income countries (LMIC), 19 upper-middle-income countries (UMIC), and eight high-income countries. We take the latest year available for each country and calculate the share of each occupation belonging to our automation and augmentation categories in total employment in that country, with further disaggregation by sex. Subsequently, we construct income-group profiles, by calculating the weighted mean of those automation and augmentation shares within each income group, as visualised in Figure 4.[9]

9 We rely on weighted means as our instrument of choice for the most balanced approach to country-level differences within groups (see Appendix for detailed formulas). To ensure that the results are not affected by extreme differences in the distribution of values within groups, we also test calculations based on the weighted-median. Since the results are stable and very similar in both cases, we keep the weighted mean as the main calculation method.

Figure 4. Automation vs augmentation potential: shares of total employment, microdata for 59 countries



Source: Authors' own elaboration.

Global estimates

Our next step is to expand this initial estimation to the global level, with the same type of income-based country groupings. For this, we benchmark to the ILO modelled estimates data series, which includes employment estimates for 189 countries [ILO, 2023a]. One of the main challenges of producing this type of global employment figure concerns the sample representativeness for each income group. Since only 59 countries report occupational data disaggregated at the 4-digit level of ISCO-08, data for other countries needs to be estimated. Fortunately, the availability of country microdata increases significantly at higher-digit ISCO-08 levels.[10] We thus exploit this greater data availability and move up the cascading structure of ISCO-08 system with each stage of estimations.

We start by calculating the share of jobs with automation and augmentation potential in total employment for each of the 59 countries with available 4-digit data. We then compute the weighted mean for each income group. As the next step, we calculate, for these countries, the share of such jobs within broader occupational categories, defined at the ISCO-08 3-digit level. Subsequently, we obtain the weighted mean of these shares by income group and apply these to estimate the number of jobs in countries for which we have ISCO-08 3-digit data but where ISCO 4-digit data are missing. We then repeat the procedure as we move up the data coverage ladder—from ISCO 3-digit to 2-digit and, finally, from 2-digit to 1-digit. At the 1-digit level, we obtain estimates for 141 countries, ensuring broad coverage of data points from the ILO's repository. The remaining 48 countries are estimated using the same approach, thereby aligning our calculations with the ILO's official global employment figures for 2021, which cover 189 countries.[11]

Given the data limitations, the exact numbers presented in Figure 5 should be read as an indication of a general trend, based on the best employment estimate that can be produced at the global level for a selection of 4-digit ISCO-08 occupations. More importantly, the global estimate confirms the trends already observed based on the analysis of microdata for 59 countries (Figure 4). Specifically, it confirms that the number of jobs in the augmentation category is significantly higher than the number of jobs with high automation potential. Calculating the global figures leads to an adjustment in the ranking of income groups in the augmentation category, with UMICs and HICs having the largest share of employment with high augmentation potential (13.5 and 13.4 percent respectively) and the LICs having the lowest share (10.4 percent). This means that, once the size and employment distribution aspects of individual countries are considered in the estimate, globally

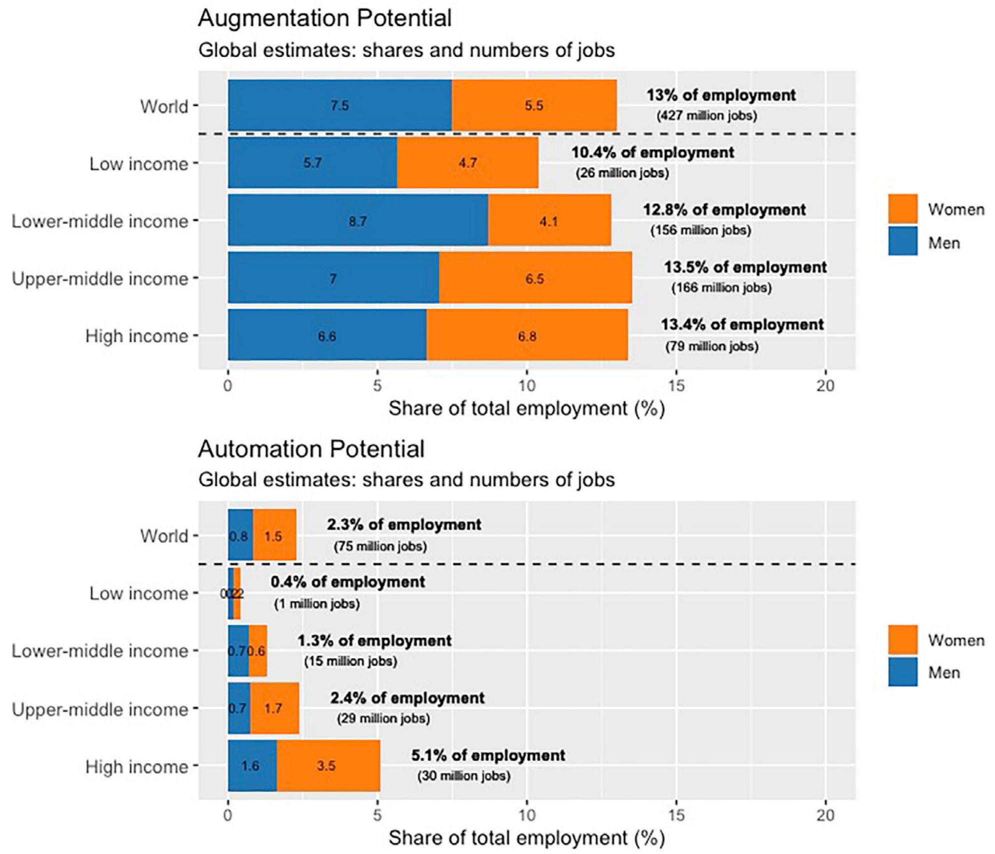
---

10 There are 59 countries at ISCO 4-digit, 97 countries at ISCO 3-digit, 122 countries at ISCO 2-digit, and 141 countries at ISCO 1-digit.

11 See Appendix for details.

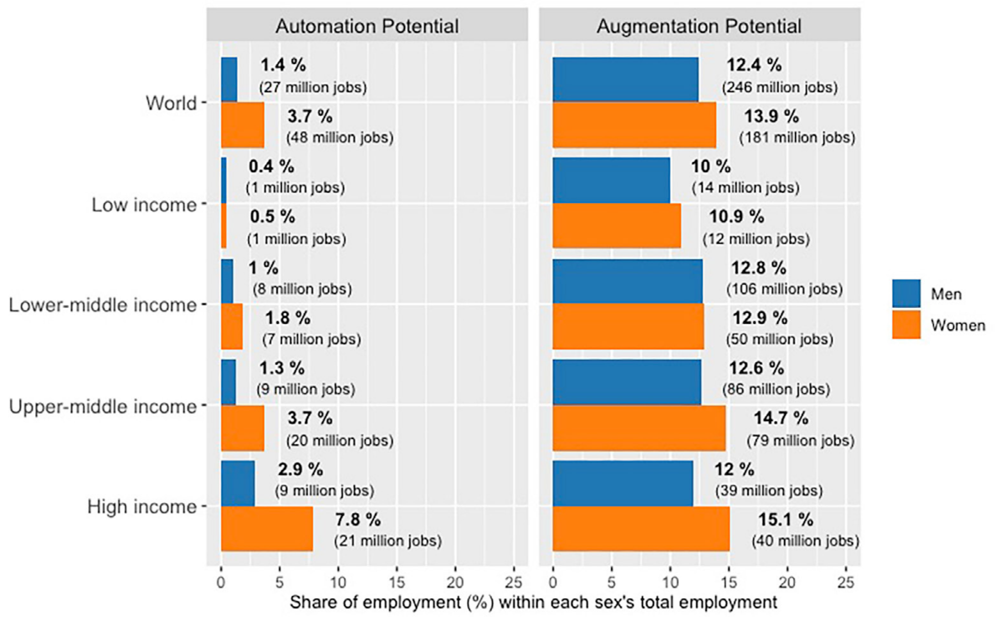
the share of jobs potentially exposed to automation with generative AI grows with income, but so does the share of jobs with high potential of experiencing augmenting effects. In other words, wealthier countries are likely to face both more disruptive effects in the technological transition and higher net gains from the process.

Figure 5. Global estimates: jobs with augmentation and automation potential as share of total employment



Source: Authors' own elaboration.

Figure 6. Automation vs augmentation potential: shares of total employment for each sex (global estimate)



Source: Authors' own elaboration.

The global estimates also confirm the strong gender effect observed in the microdata. When we disaggregate the estimate to shares of female and male employment (Figure 6), we observe that 3.7 percent of all female employment in the world is in jobs that are potentially automatable with generative AI technology, compared with just 1.4 percent of male employment. In high-income countries, the share of potentially affected female jobs is 7.8 percent, more than double the 2.9 percent of male jobs for that income group. At the same time, the share of jobs with high augmentation potential is also greater among female than male jobs across all income groups.

The big unknown

The breakdown of occupations into high automation and augmentation potential provided a helpful framework to discuss the extremes of score distribution, thereby minimising the risk of statistical overlaps between the two groups. Nevertheless, this left an important group of occupations, located between the automation and augmentation out of focus of the discussion. We refer to these jobs, illustrated in Figure 7 with green points, as "the big unknown" since our framework and data do not allow for a clear-cut classification of this group. In general, such jobs have a high occupational mean score, and a high variance of tasks-level scores, which means that their exposure to GPT technology can have varied and idiosyncratic effects.

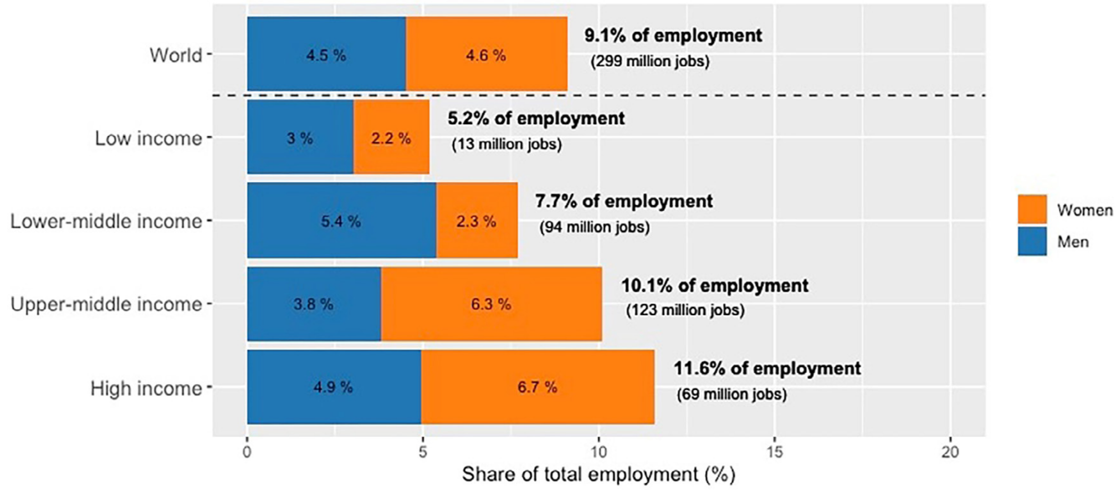
Figure 7. The "Big Unknown": occupations between augmentation and automation potential

Source: Authors' own elaboration.

Depending on the technological progress of generative AI, as well as the applications built on top of the technology, some of the tasks might become more automatable, while new tasks could emerge in these professions, pushing them closer to the augmentation or automation cluster or, the more likely scenario, having them evolve into new occupations. While we refrain from speculating on the direction of this evolution, we find it important to quantify the share of employment belonging to this group.

As illustrated in Figure 8, these occupations constitute a nontrivial share of global employment, with some 16.2 percent, or 553 million workers, falling into this category. While in low-income and middle-income countries, such jobs are to a larger extent held by men, in UMICs and HICs, women dominate this share of total employment.

Figure 8. The "Big Unknown": share of total employment, by income group (global estimate)



Source: Authors' own elaboration.

Managing the transition: policies to address automation, augmentation and the growing digital divide

The estimates presented in the preceding section suggest that the recent progress in machine learning, in particular developments around LLMs, are likely to have disruptive effects on labour markets, with larger effects in high-income countries and specific occupational groups. Still much remains unknown with respect to the progress and limitations of this and similar technologies, which will ultimately determine its overall impact. Taking the current capabilities of GPT at face value and applying it to the distribution of labour markets around the world gives us an indicative picture that suggests greater potential for job augmentation as opposed to automation. This finding represents a continuum with previous waves of technological progress, despite recurring bouts of anxiety [Autor, 2015, Cherry, 2020].

Nevertheless, policies are needed to manage the transition of workers affected by automation, in addition to managing the potential effects on job quality for those affected by augmentation. Indeed, both scenarios require building and strengthening systems of social dialogue, including workplace consultation. Policy attention is also needed for countries that lack the requisite physical infrastructure and skills to benefit from the new technology.

Mitigating the negative effects of automation

The analysis revealed that higher-income countries will experience the greatest effects from automation as a result of the important share of clerical and para-professional jobs in the occupational distribution. Middle- and low-income countries will be less exposed, though certain occupations that are potentially exposed to automation, such as call centre work[12], figure prominently in some of these countries, particularly India and the Philippines, which dominate the world's call centre industry. In the Philippines, half a million people were employed in call centres in 2016, of whom 53 percent were women.[13]

The challenges and consequences of such adjustments should not be underestimated. For example, a study of the effects of automation on Dutch workers from 2010 to 2016 found that workers made redundant as a result of automation experienced a five-year cumulative wage income loss of 9 percent of the annual wage [Bessen et al., 2023]. The losses were only partially offset by various benefits systems, despite the country's

12 (4222) contact centre information clerks, (4227) survey and market research interviewers, (5244) customer contact salesperson.

13 In 2023, the IT and Business Processing Association (ITBPA) of the Philippines reported that the sector employed 1.5 million full-time equivalent employees in 2022 [ITBPA, 2023].

relatively robust unemployment insurance system. Workers experiencing such effects in countries with less developed insurance systems and which lack job training and job placement services, or where there are high levels of unemployment, are more vulnerable.

Consultation and negotiation between employers and workers are critical for managing the transition process, as they encourage redeployment and training over job loss. The ILO's Employment Protection Recommendation (No. 166, 1982) includes provisions on termination of employment for technological reasons. In particular for cases of collective dismissals, it calls for special procedural requirements, including consultation with workers' representatives, notification of competent authorities, adoption of measures to avert or minimise terminations and mitigate their effects, as well as the establishment of fair criteria for selection for termination and priority in rehiring. The aim of these requirements is to minimise the negative externalities from dismissals, especially collective ones, as well as to better internalise the cost of such dismissals and support an orderly process that balances the needs of workers, employers, and societies at large [**Aleksynska, Muller, 2020**]. Social dialogue is also useful for designing and instituting social protection and skills development programmes that can help mitigate the negative effects of automation.

One issue that will require specific attention is the gendered effects of the automation. The potential exposure to automation disproportionately affects the share of women's employment by more than two-fold in high-income countries (7.9 percent vs. 2.9 percent) and upper-middle-income countries (2.7 percent vs. 1.3 percent). Concentrated job losses in female-dominated occupations could threaten advances in increasing women's labour market participation, unless these are compensated by increases in demand for labour in other professions where women are well represented.

Addressing the digital divide

A potentially more significant consequence of a wider adoption of generative AI products could be an increased divergence in productivity between high- and low-income countries. Larger shares of jobs falling into the augmentation category suggest that, at least in the near future, generative AI systems are more likely to become productivity tools, supporting and speeding up the execution of some tasks within certain occupations. The digital divide will influence how the benefits of such productivity tools are distributed among societies and countries, with high-income countries and privileged groups likely to reap the biggest rewards.

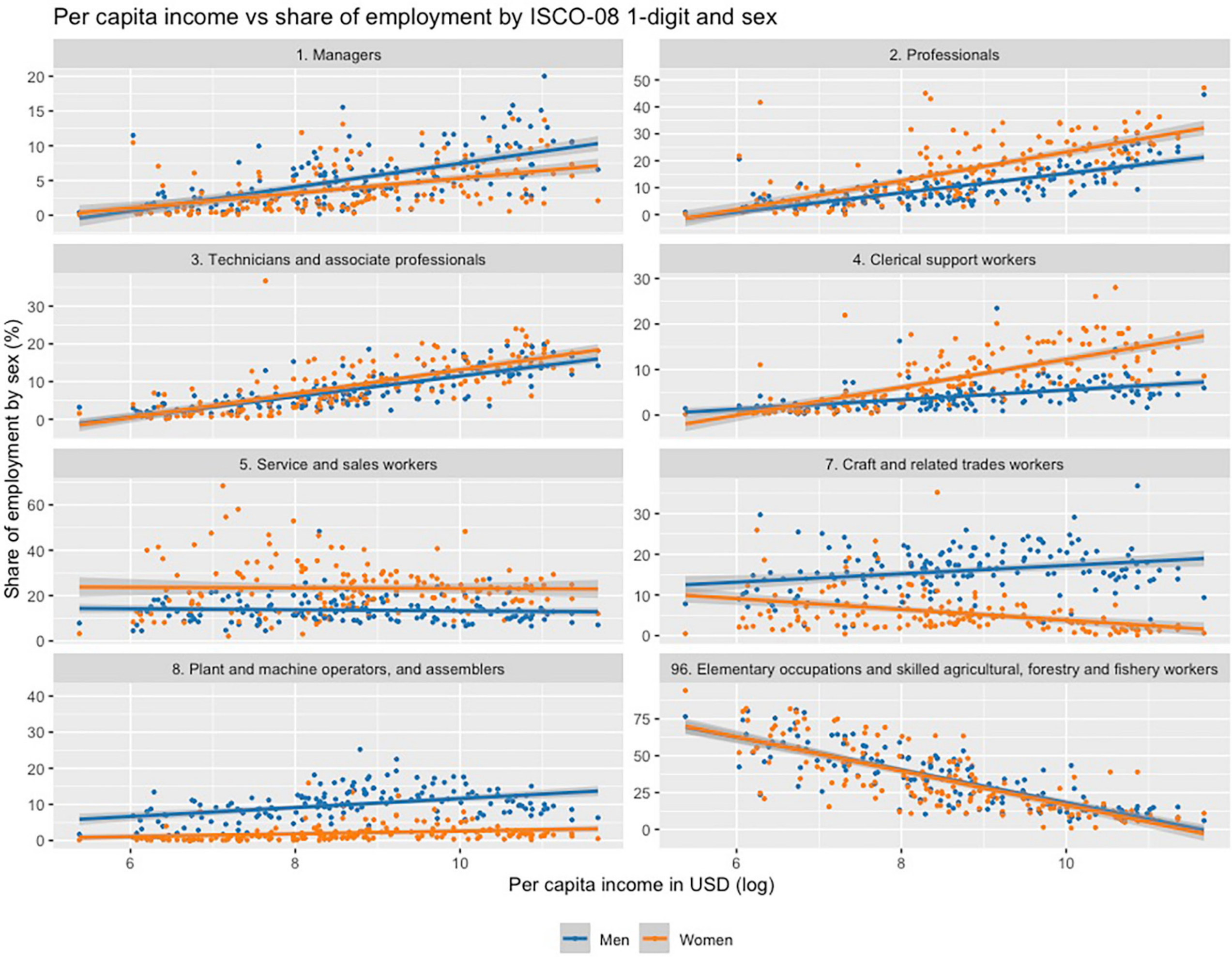
Low-income countries, in particular, are at risk of falling behind. While up to 13 percent of employment in these countries is found in the potential augmentation category, in practice potential benefits of generative AI technologies are likely to be limited, as the lack of reliable infrastructure will constrain its application. To begin with, such technology is dependent on access and cost of broadband connectivity, as well as electricity. In 2022, a third of the global population, corresponding to some 2.7 billion people, still did not have access to the internet. Among the two-thirds that do, many would not be able to use generative AI due to the limitations in the quality of their connection or the cost of the service. Even more fundamental than the internet, reliable electricity provision is often a challenge. According to the World Bank Enterprise Survey, 49 percent of registered firms in developing countries experienced electrical outages, averaging 4.5 days per month and lasting 4 hours on average.[14]

There is a risk that the technology could affect negatively the classic path of labour market diversification that is associated with economic development. Figure 9 shows the share of employment by 1-digit ISCO groups for countries at different levels of income. As countries get richer, employment distributions evolve according to a visible log-linear pattern: the share of elementary occupations plummets, while the share of managers, professionals, technicians and clerical support workers increases. The automation potential of clerical work could disturb this growth trajectory, making such jobs less available to workers, as countries' income level rises. In that sense, the potential impact of generative AI on jobs in developing countries has less to do

[14] https://www.enterprisesurveys.org/en/data/exploretopics/infrastructure, (accessed on 31.07.2023).

with actual job destruction, but rather the potential that certain jobs will never be created. Moreover, since women are well represented in clerical support occupations, the potential implications of these "missing jobs" will be more pronounced for women.

Figure 9. A classic growth path: income and occupational diversification



Source: Authors' own elaboration.

With the right conditions in place, a new wave of technology could fuel growth opportunities. In the past, technological advancements have spurred new and successful industries in many developing countries. One example is the M-Pesa money service, which relied on the diffusion of mobile telephones in Kenya. The service, in turn, increased financial inclusion thus helping to propel the growth of SMEs and led to the creation of a network of 110,000 agents, 40 times the number of bank ATMs in Kenya [Buku, Meredith, 2012; de Soyres et al., 2018]. Similarly, a study of the diffusion of 3G coverage in Rwanda between 2002 and 2019 found that increased mobile internet coverage was positively associated with employment growth, increasing both skilled and unskilled occupations [Caldarola et al., 2022]. Hjort and Poulsen [2019] also find positive employment effects from the arrival of the internet in 12 African countries, albeit with a slight bias towards skilled occupations. These gains are attributed to increases in productivity and growth of markets that followed increased connectivity.

Among developing countries, further distinction needs to be made. While middle-income countries are more exposed to the automating effects of generative AI, their digital infrastructure and skilled workforce can also be an asset for spawning the growth of complementary industries. Although India and the Philippines are at risk of losing some call centre work, their dominance in business process outsourcing may provide the needed foundation for the development of new industries.

Conclusion

In this paper, we attempted to quantify some of the potential effects of generative AI on occupations from a global perspective. Our study provided a global estimate of the number of jobs in the categories that are most exposed to technologies with similar capabilities as GPT-4, by relying on the international standard of ISCO-08 and linking the task-level scores to employment distributions reflected in official ILO statistics. We subsequently discussed the consequences of these findings in the context of differential impacts that can be expected depending on countries' income levels.

The analysis was based on the top threshold of current technological possibilities and relied on three bold assumptions. First, we assumed that the tasks for which automation scores were estimated would be executed in the context of a high-income country. This disregards the more limited potential for deployment in lower-income countries, where infrastructure is typically less developed, unreliable and often more expensive, and where lower skill and wage levels make the costs of technological adoption relatively high. Second, we relied on GPT-4 to predict the scores, which is likely to reflect an apex of technological optimism in terms of ease of deployment—an approach that in practice is difficult to operationalize. Third, without being able to make reliable predictions on future technological progress, we focused on the current potential for task automation without speculating on the numbers of new jobs that might emerge. While this approach might have been expected to produce alarming estimates of net job loss, it did not. Instead, our global estimates point to a future in which work is transformed, but still very much in existence.

Our findings largely align with the evolving body of academic literature concerning previous waves of technological transformations, but some of the trends we identify are new as a result of our exclusive focus on generative AI. While early studies of potential AI adoption identified low-skill, repetitive and routine jobs as those with the highest potential of automation [e.g., **McKinsey, 2016**; **Frey, Osborne, 2017**], in which a computer-based system could be coupled with a machine to replace a human in manual production jobs [**Autor, 2015**; **Acemoglu, Restrepo, 2020**], more recent literature has highlighted the ability of machine learning systems to improve their performance in non-routine tasks [**Brynjolfsson et al., 2018**; **Ernst et al., 2019**; **Webb, 2019**; **Lane, Saint-Martin, 2021**]. We argue that the emergence of GPT reinforces this shifting picture, due to its refined ability to perform cognitive tasks, such as analysing text, drafting documents and messages, or searching through private repositories and the web for additional information. As a consequence, our study indicates that—at least in the short run—this new wave of automation will focus on a different group of workers, typically associated with "knowledge work" [**Surawski, 2019**].

The occupational group with the highest share of tasks exposed to GPT technology is clerical work, where most tasks fall into at least a medium level of exposure, and roughly a quarter are highly exposed to potential automation. As a result, many such jobs may never materialise in developing countries, where they have traditionally provided an important pathway to increasing female employment. For other types of "knowledge work," exposure is only partial, suggesting stronger potential for augmentation and productivity gains rather than outright job displacement.

These findings align with some of the most recent literature on generative AI systems with a global focus. **McKinsey** [**2023**] points to a similar group of "knowledge work" occupations and tasks as having the highest level of exposure, though with a significantly higher suggested level of displacement. The WEF's global survey, focussed on large enterprises, also lists clerical and administrative jobs among occupations with the steepest expected declines [**WEF, 2023**]. Estimates by **Goldman Sachs** [**2023**] suggest a slightly higher level of potential automation than our calculations, but with the general conclusion aligning with our main finding that "most jobs and industries are only partially exposed to automation and are thus more likely to be complemented rather than substituted by AI."

The more moderate effects observed in our estimations stem from several factors. First, we rely on ISCO-08 as the source of tasks and occupations, which is more adequate for a study with a global scope than the US-oriented O*NET database. Second, the application of ILO country-level employment statistics adds important nuance to the actual number of jobs that exists in those categories, bringing out income-based differences that affect the final employment effects at the global level. Third, we do not attempt to make predictions on the evolution of the technology. While the growing capabilities of generative AI and the range of secondary applications that can be built on top of this technology are likely to increase the number of jobs in both the augmentation and automation categories identified in our paper, our analysis suggests that the broad contours of transformation identified in this study will remain valid for the coming years.

Ultimately, we argue that in the realm of work, generative AI is neither inherently good nor bad, and that its socioeconomic impacts will largely depend on how its diffusion is managed. The questions of power balance, voice of the workers affected by labour market adjustments, respect for existing norms and rights, and adequate use of national social protection and skills training systems will be crucial in managing the deployment of AI in the workplace. Without proper policies in place, there is a risk that only some of the well-positioned countries and market participants will be able to harness the benefits of the transition, while the costs to affected workers could be severe. Therefore, for policy makers, our study should not read as a calming voice, but rather as a call for harnessing policy to address the technological changes that are upon us.

References

Acemoglu D., Autor D., Hazell J., Restrepo P. [2022], Artificial Intelligence and Jobs: Evidence from Online Vacancies, *Journal of Labor Economics*, 40 (S1): S293–S340, https://doi.org/10.1086/718327.

Acemoglu D., Restrepo P. [2020], Robots and Jobs: Evidence from US Labor Markets, *Journal of Political Economy*, 128 (6): 2188–2244, https://doi.org/10.1086/705716.

Aleksnyska M., Muller A. [2020], *The regulation of collective dismissals: Economic rationale and legal practice,* Working Paper, http://www.ilo.org/global/publications/working-papers/WCMS_745125/lang--en/index.htm.

Arntz M., Gregory T., Zierahn U. [2017], Revisiting the risk of automation, *Economics Letters*, 159: 157–160, https://doi. org/10.1016/j.econlet.2017.07.001.

Autor D.H. [2015], Why Are There Still So Many Jobs? The History and Future of Workplace Automation, *Journal of Economic Perspectives*, 29 (3): 3–30, https://doi.org/10.1257/jep.29.3.3.

Autor D.H., Dorn D. [2013], The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market, *American Economic Review*, 103 (5): 1553–1597, https://doi.org/10.1257/aer.103.5.1553.

Autor D.H., Salomons A., Seegmiller B. [2021], *New Frontiers: The Origin and Content of New Work, 1940–2018*, MIT Working Paper.

Balliester T., Elsheikhi A. [2018], *The Future of Work: A Literature Review*, Research Department Working Paper No. 29, International Labour Organization.

Bender E.M., Gebru T., McMillan-Major A., Shmitchell S. [2021], On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, in: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*: 610–623, https://doi. org/10.1145/3442188.3445922.

Bessen J., Goos M., Salomons A., Berge W. [2023], What Happens to Workers at Firms that Automate?, *The Review of Economics and Statistics*, 107 (1): 125–141.

Brynjolfsson E., Mitchell T., Rock D. [2018], What Can Machines Learn and What Does It Mean for Occupations and the Economy?, *AEA Papers and Proceedings*, 108: 43–47, https://doi.org/10.1257/pandp.20181019.

Bubeck S., Chandrasekaran V., Eldan R., Gehrke J., Horvitz E., Kamar E., … Zhang Y. [2023], *Sparks of Artificial General Intelligence: Early experiments with GPT-4*, http://arxiv.org/abs/2303.12712.

Buku M.W., Meredith M.W. [2012], Safaricom and M-PESA in Kenya: Financial Inclusion and Financial Integrity Mobile Money Symposium 2013, *Washington Journal of Law, Technology & Arts*, 8 (3): 375–400.

Caldarola B., Grazzi M., Occelli M., Sanfilippo M. [2022], *Mobile internet, skills and structural transformation in Rwanda*, https:// doi.org/10.54394/XSTK4695.

Cammeraat E., Squicciarini M. [2021], *Burning Glass Technologies' data use in policy-relevant analysis: An occupation- level assessment*, https://doi.org/10.1787/cd75c3e7-en.

Carbonero F., Davies J., Ernst E., Fossen F.M., Samaan D., Sorgner A. [2023], The impact of artificial intelligence on labor markets in developing countries: a new method with an illustration for Lao PDR and urban Viet Nam, *Journal of Evolutionary Economics*, 33: 707–736, https://doi.org/10.1007/s00191-023-00809-7.

Cherry M.A. [2020], Back to the future: A continuity of dialogue on work and technology at the ILO, *International Labour Review*, 159 (1), 1–23, https://doi.org/10.1111/ilr.12156.

Eisfeldt A.L., Schubert G., Zhang M.B. [2023], *Generative AI and Firm Values*, SSRN Scholarly Paper, https://doi.org/10.2139/ ssrn.4436627.

Eloundou T., Manning S., Mishkin P., Rock D. [2023], *GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models*, http://arxiv.org/abs/2303.10130.

Ernst E., Merola R., Samaan D. [2019], Economics of Artificial Intelligence: Implications for the Future of Work, *IZA Journal of Labor Policy*, 9 (1), https://doi.org/10.2478/izajolp-2019-0004.

Felten E.W., Raj M., Seamans R. [2019], *The Occupational Impact of Artificial Intelligence: Labor, Skills, and Polarization*, SSRN Scholarly Paper, https://doi.org/10.2139/ssrn.3368605.

Fossen F.M., Sorgner A. [2022], New digital technologies and heterogeneous wage and employment dynamics in the United States: Evidence from individual-level data, *Technological Forecasting and Social Change*, 175: 121381, https://doi.org/10.1016/j.techfore.2021.121381.

Frey C.B., Osborne M.A. [2017], The future of employment: How susceptible are jobs to computerisation?, *Technological Forecasting and Social Change*, 114: 254–280, https://doi.org/10.1016/j.techfore.2016.08.019.

Georgieff A., Hyee R. [2021], *Artificial intelligence and employment: New cross-country evidence*, https://www.oecd.org/sti/artificial-intelligence-and-employment-c2c1d276-en.htm (accessed on 31.07.2023).

Gmyrek P., Lutz C., Newlands G. [2025], A technological construction of society: Comparing GPT-4 and human respondents for occupational evaluation in the UK, *British Journal of Industrial Relations*, 63: 180–208, https://doi.org/10.1111/bjir.12840.

Goldfarb A., Taska B., Teodoridis F. [2023], Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings, *Research Policy*, 52 (1), 104653, https://doi.org/10.1016/j.respol.2022.104653.

Goldman Sachs [2023], *Global Economics Analyst The Potentially Large Effects of Artificial Intelligence on Economic Growth (BriggsKodnani)*, https://www.key4biz.it/wp-content/uploads/2023/03/Global-Economics-Analyst_-The-Potentially- Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf (accessed on 31.07.2023).

Hjort J., Poulsen J. [2019], The Arrival of Fast Internet and Employment in Africa, *American Economic Review*, 109 (3): 1032–1079, https://doi.org/10.1257/aer.20161385.

ILO [1982], *Convention C158 - Termination of Employment Convention*, 158, https://www.ilo.org/dyn/normlex/ en/f?p=NORMLEXPUB:12100:0::NO:12100: P12100_INSTRUMENT_ID:312303: NO (accessed on 31.07.2023).

ILO [2023a], *ILO Modelled Estimates (ILOEST database)*, https://ilostat.ilo.org/resources/ concepts-and-definitions/ilo-modelled-estimates/ (accessed on 20.07.2023).

ILO [2023b], *ISCO documentation: part III. Definitions of major groups, sub-major groups, minor groups and unit groups*, https://www.ilo.org/public/english/bureau/stat/isco/docs/groupdefn08.pdf (accessed on 20.07.2023).

Jovanovic B., Rousseau P.L. [2005], *General Purpose Technologies*, NBER Working Paper Series, National Bureau of Economic Research, https://www.nber.org/system/files/working_papers/w11093/w11093.pdf (accessed on 20.07.2023).

Karger E., Rosenberg J., Jacobs Z., Hickman M., Hadshar R., Gamin K., … Tetlock P.E. [2023], *Forecasting Existential Risks: Evidence from a Long-Run Forecasting Tournament*, Forecasting Research Institute, https://static1. squarespace.com/static/635693acf15a3e2a14a56a4a/t/64abffe3f024747dd0e38d71/1688993798938/XPT.pdf (accessed on 31.07.2023).

Lane M., Saint-Martin A. [2021], *The impact of Artificial Intelligence on the labour market: What do we know so far?*, https://doi. org/10.1787/7c895724-en.

Mattos F.B. de, Dasgupta S., Jiang X., Kucera D., Schiavone A.F. [2020], *Robotics and reshoring - Employment implications for developing countries*, http://www.ilo.org/emppolicy/pubs/WCMS_751599/lang--en/index.htm.

McKinsey [2016], *A future that works: automation, employment and productivity*, https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet (accessed on 31.07.2023).

McKinsey [2023], *Economic potential of generative AI, McKinsey*, https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#/.

Soyres F. de, Jelil M.A., Cerruti C., Kiwara L. [2018], *What Kenya's mobile money success could mean for the Arab world*, https://www.worldbank.org/en/news/feature/2018/10/03/what-kenya-s-mobile-money-success-could-mean-for-the-arab-world (accessed on 31.07.2023).

Surawski B. [2019], Who is a "knowledge worker" – clarifying the meaning of the term through comparison with synonymous and associated terms, *Management*, 23: 105–133, https://doi.org/10.2478/manment-2019-0007.

Webb M. [2019], *The Impact of Artificial Intelligence on the Labor Market*, SSRN Scholarly Paper, https://doi.org/10.2139/ssrn.3482150.

WEF [2023], *The Future of Jobs Report 2023*, https://www.weforum.org/reports/the-future-of-jobs-report-2023 (accessed on 31.07.2023).

Declaration of generative AI and AI-assisted technologies in the research

As explained in the text, GPT-4 API was used to generate alternative occupational definitions, tasks and task-level scores. We placed some 25,000 API requests to GPT-4 and used the Ada model to generate task-level embeddings for 7,482 tasks. We used GPT-4 API to summarise the content of these task clusters. We also used ChatGPT to generate the list of abbreviations based on our final text. OpenAI provided us with a research credit in API tokens with a total value of USD 1,000, out of which some USD 600 were used for this research. We are grateful to Elizabeth Proehl and Pamela Mishkin from OpenAI for their openness about the methods applied in **Eloundou et al.** [**2023**], for responding to our request for GPT-4 API access, and for the research credit of GPT tokens.

The authors of this text have no affiliation to OpenAI and no vested interest in that regard. GPT was not used to generate the main text of this article, which is fully of our own authorship, along with any mistakes.

Appendix

Countries with missing ISCO-08 4-digit data: estimation procedure

To illustrate our estimation method, we use the example of jobs identified as having high automation potential. For an income group *IG*, let total employment be denoted as T_{IG}. The total employment in each income group is the sum of total jobs J_i in all the countries i that belong to the income group *IG*:

$$T_{IG} = \sum_{i \in IG} J_i$$

For each country i, denote A_i as the number of jobs with high automation potential and J_i as the total number of jobs. The share of automation jobs S_i is then calculated as:

$$S_i = \frac{A_i}{J_i}$$

The weight W_i for each country i in income group *IG* is defined as the share of the country's employment in the total employment of that income group:

$$W_i = \frac{J_i}{T_{IG}}$$

The weighted mean M_{IG} for each income group *IG* is then the sum of the product of the weights W_i and the automation job shares S_i for all countries i in income group *IG*:

$$M_{IG} = \sum_{i \in IG} W_i S_i$$

For each ISCO-08 3-digit category d, in country i where 4-digit ISCO-08 data exists, the total number of jobs $J3_{di}$ is given by:

$$J3_{di} = \sum_{k \in D3_{di}} J4_{ki}$$

where $J4_{ki}$ is the total number of jobs in the 4-digit category k that falls under the 3-digit category d in that country. The share $S3_{di}$ of automation jobs in 4-digit category d to the total jobs in the corresponding 3-digit category d in country i is given by:

$$S3_{di} = \frac{A_{di}}{J3_{di}}$$

where A_{di} is the number of automation jobs in the 4-digit category d, and $J3_{di}$ is the total number of jobs in the 3-digit category d in country i.

As the next step, each 3-digit share $S3_{di}$ is weighted by total employment E_i in country i relative to total employment E_{IG} in income group IG. The weighted mean WMS_{IG} for income group IG is then calculated as:

$$WMS_{IG} = \frac{\sum_{i \in IG} E_i * S3_{di}}{\sum_{i \in IG} E_i}$$

For each country i with missing 4-digit data but available 3-digit data, the estimated number of automation jobs A_i can then be calculated using the weighted mean share WMS_{IG} of the corresponding income group and total employment E_i in country i:

$$A_i = WMS_{IG} * E_i$$

We then repeat an analogical procedure moving up the data coverage ladder, that is, from ISCO 3-digit to 2-digit, from 2-digit to 1-digit, and finally to global coverage.