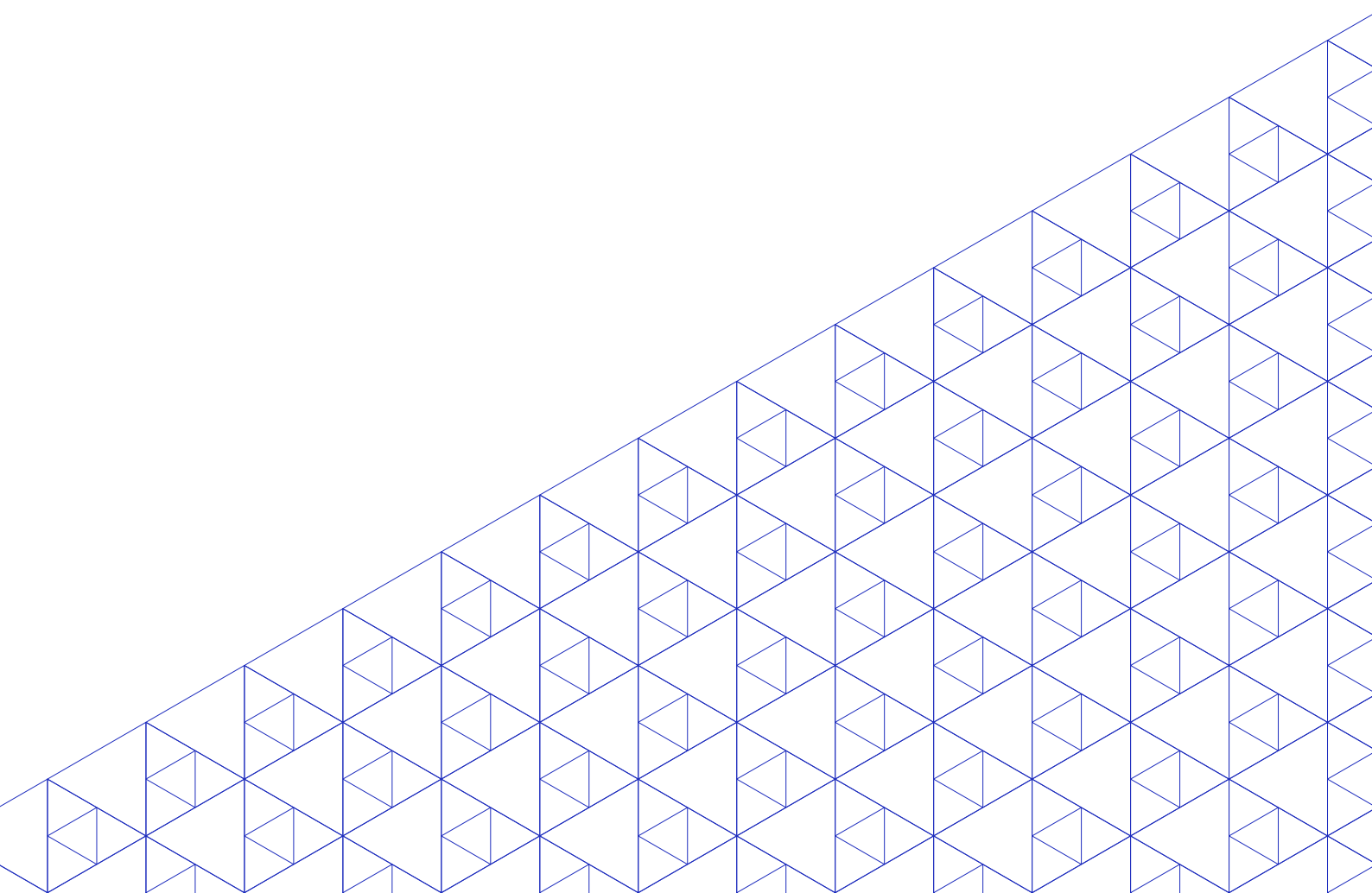




► Generative AI and jobs: A global analysis of potential effects on job quantity and quality

Authors / Paweł Gmyrek, Janine Berg, David Bescond





This is an open access work distributed under the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>). Users can reuse, share, adapt and build upon the original work, as detailed in the License. The ILO must be clearly credited as the owner of the original work. The use of the emblem of the ILO is not permitted in connection with users' work.

Attribution – The work must be cited as follows: Gmyrek, P., Berg, J., Bescond, D. *Generative AI and jobs: A global analysis of potential effects on job quantity and quality*. ILO Working Paper 96. Geneva: International Labour Office, 2023.

Translations – In case of a translation of this work, the following disclaimer must be added along with the attribution: *This translation was not created by the International Labour Organization (ILO) and should not be considered an official ILO translation. The ILO is not responsible for the content or accuracy of this translation.*

Adaptations – In case of an adaptation of this work, the following disclaimer must be added along with the attribution: *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.*

This CC license does not apply to non-ILO copyright materials included in this publication. If the material is attributed to a third party, the user of such material is solely responsible for clearing the rights with the right holder.

Any dispute arising under this license that cannot be settled amicably shall be referred to arbitration in accordance with the Arbitration Rules of the United Nations Commission on International Trade Law (UNCITRAL). The parties shall be bound by any arbitration award rendered as a result of such arbitration as the final adjudication of such a dispute.

All queries on rights and licensing should be addressed to the ILO Publishing Unit (Rights and Licensing), 1211 Geneva 22, Switzerland, or by email to rights@ilo.org.

ISBN 9789220395356 (print), ISBN 9789220395363 (web PDF), ISBN 9789220395370 (epub), ISBN 9789220395387 (mobi), ISBN 9789220395394 (html). ISSN 2708-3438 (print), ISSN 2708-3446 (digital)

<https://doi.org/10.54394/FHEM8239>

The designations employed in ILO publications, which are in conformity with United Nations practice, and the presentation of material therein do not imply the expression of any opinion whatsoever on the part of the ILO concerning the legal status of any country, area or territory or of its authorities, or concerning the delimitation of its frontiers.

The responsibility for opinions expressed in signed articles, studies and other contributions rests solely with their authors, and publication does not constitute an endorsement by the ILO of the opinions expressed in them.

Reference to names of firms and commercial products and processes does not imply their endorsement by the ILO, and any failure to mention a particular firm, commercial product or process is not a sign of disapproval.

Information on ILO publications and digital products can be found at: www.ilo.org/publns

ILO Working Papers summarize the results of ILO research in progress, and seek to stimulate discussion of a range of issues related to the world of work. Comments on this ILO Working Paper are welcome and can be sent to RESEARCH@ilo.org, berg@ilo.org.

Authorization for publication: Richard Samans, Director RESEARCH

ILO Working Papers can be found at: www.ilo.org/global/publications/working-papers

Suggested citation:

Gmyrek, P., Berg, J., Bescond, D. 2023. *Generative AI and jobs: A global analysis of potential effects on job quantity and quality*, ILO Working Paper 96 (Geneva, ILO). <https://doi.org/10.54394/FHEM8239>

Abstract

This study presents a global analysis of the potential exposure of occupations and tasks to Generative AI, and specifically to Generative Pre-Trained Transformers (GPTs), and the possible implications of such exposure for job quantity and quality. It uses the GPT-4 model to estimate task-level scores of potential exposure and then estimates potential employment effects at the global level as well as by country income group. Despite representing an upper-bound estimate of exposure, we find that only the broad occupation of clerical work is highly exposed to the technology with 24 per cent of clerical tasks considered highly exposed and an additional 58 percent with medium-level exposure. For the other occupational groups, the greatest share of highly exposed tasks oscillates between 1 and 4 per cent, and medium exposed tasks do not exceed 25 per cent. As a result, the most important impact of the technology is likely to be of augmenting work – automating some tasks within an occupation while leaving time for other duties – as opposed to fully automating occupations.

The potential employment effects, whether augmenting or automating, vary widely across country income groups, due to different occupational structures. In low-income countries, only 0.4 per cent of total employment is potentially exposed to automation effects, whereas in high-income countries the share rises to 5.5 percent. The effects are highly gendered, with more than double the share of women potentially affected by automation. The greater impact is from augmentation, which has the potential to affect 10.4 percent of employment in low-income countries and 13.4 percent of employment in high-income countries. However, such effects do not consider infrastructure constraints, which will impede the possibility for use in lower-income countries and likely increase the productivity gap.

We stress that the primary value of this analysis is not the precise estimates, but rather the insights that the overall distribution of such scores provides about the nature of possible changes. Such insights can encourage governments and social partners to proactively design policies that support orderly, fair, and consultative transitions, rather than dealing with change in a reactive manner. Moreover, the likely ramifications on job quality might be of greater consequence than the quantitative impacts, both with respect to the new jobs created because of the technology, but also the potential effects on work intensity and autonomy when the technology is integrated into the workplace. For this reason, we also emphasize the need for social dialogue and regulation to support quality employment.

About the authors

Pawel Gmyrek is Senior Researcher in the Research Department of the ILO.

Janine Berg is Senior Economist in the Research Department of the ILO.

David Bescond is Data Scientist in the ILO's Department of Statistics.

Table of contents

Abstract	01
About the authors	01
Acronyms	05
<hr/>	
▶ Introduction	07
<hr/>	
▶ 1 Methods and Data	10
1.1. ISCO data on occupations and tasks	11
1.2. Prompt design and sequence	12
<hr/>	
▶ 2 Assessment of the Predictions, Robustness Tests and the Bounds for Analysis	17
<hr/>	
▶ 3 Results	20
3.1. Automation vs augmentation: distribution of scores across tasks and occupations	24
<hr/>	
▶ 4 Exposed occupations as a share of employment: global and income-based estimates	30
4.1. Augmentation vs Automation: ILO microdata	30
4.2. Augmentation vs Automation: global estimate	32
4.3. The big unknown	36
<hr/>	
▶ 5 Managing the transition: Policies to address automation, augmentation and the growing digital divide	38
5.1 Mitigating the negative effects of automation	38
5.2 Ensuring job quality under augmentation	39
5.3 Addressing the digital divide	40
<hr/>	
▶ Conclusion	43
<hr/>	
Appendix 1. Countries with missing ISCO-08 4-digit data: estimation procedure	45
References	47
Acknowledgements and use of GPT	51

List of Figures

Figure 1. Mean automation scores by occupation, based on ISCO and GPT tasks	21
Figure 2. Tasks with medium and high GPT-exposure, by occupational category (ISCO 1-digit)	24
Figure 3. Box plot of task-level scores by ISCO 4d, grouped by ISCO 1d	25
Figure 4. Augmentation vs automation potential at occupational level	27
Figure 5. Occupations with high automation potential	28
Figure 6. Occupations with high augmentation potential	29
Figure 7a. Automation vs augmentation potential: shares of total employment, microdata for 59 countries	30
Figure 7b. Automation vs augmentation potential: shares of total employment in each sex (ILO microdata)	31
Figure 8. Country coverage based on the level of digits in ISCO-08 (ILO data)	33
Figure 9a. Global estimates: jobs with augmentation and automation potential as share of total employment	34
Figure 9b. Automation vs augmentation potential: shares of total employment for each sex (global estimate)	35
Figure 10. Occupations with high automation potential, by ISCO 4-digit and income group	36
Figure 11a. The “Big Unknown”: occupations between augmentation and automation potential	37
Figure 11b. The “Big Unknown”: share of total employment, by income group (global estimate)	37
Figure 11. Share of population not using the internet	41
Figure 12. A classic growth path: income and occupational diversification	42

List of Tables

Table 1. ISCO-08 Structure of occupations and tasks used in the study	11
Table 2. Sample of tasks and definitions from ISCO and predicted by GPT-4	14
Table 3. Sample of task-level scores (high-income country context)	15
Table 4. Test of score consistency (100 task-level predictions)	17
Table 4: Tasks with high automation potential clustered into thematic groups*	22
Table 5: Grouping of occupations based on task-level scores	26
Table 6: Microdata coverage by levels ISCO-08: number of countries	32

Acronyms

3G	Third Generation (referring to a generation of standards for mobile telecommunications)
Ada	A language model by OpenAI used to generate embeddings
AGI	Artificial General Intelligence
AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
ATMs	Automated Teller Machines
CPU	Central Processing Unit
DL	Deep Learning
DOLE	Department of Labor and Employment
ESCO	European Skills, Competences, Qualifications and Occupations
GPTs	Generative Pre-Trained Transformers
GPT-4	Generative Pre-Trained Transformer 4
GPU	Graphics Processing Unit
HIC	High-Income Countries
ICT	Information and Communications Technology
ILO	International Labour Organization
ISCO	International Standard Classification of Occupations
ISCO-08	International Standard Classification of Occupations 2008
K-Means	K-Means Clustering Algorithm
LFS	Labour Force Surveys
LIC	Low-Income Countries
LLMs	Large Language Models

LMIC	Lower-Middle-Income Countries
ML	Machine Learning
NLP	Natural Language Processing
OECD	Organisation for Economic Co-operation and Development
O*NET	Occupational Information Network
OpenAI	Open Artificial Intelligence (organization's name)
Python	High-level programming language
RL	Reinforcement Learning
SD	Standard Deviation
SMEs	Small and Medium-sized Enterprises
UMIC	Upper-Middle-Income Countries
US	United States
USD	United States Dollar
UMIC	Upper-Middle-Income Countries
US	United States

▶ Introduction

Each new wave of technological progress intensifies debates on automation and jobs. Current debates on Artificial Intelligence (AI) and jobs recall those of the early 1900s with the introduction of the moving assembly line, or even those of the 1950s and 1960s, which followed the introduction of the early mainframe computers. While there have been some nods to the alienation that technology can bring by standardizing and controlling work processes, in most cases, the debates have centred on two opposing viewpoints: the optimists, who view new technology as the means to relieve workers from the most arduous tasks, and the pessimists, who raise alarm about the imminent threat to jobs and the risk of mass unemployment.

What has changed in debates on technology and workers, however, is the types of workers affected. While the advances in technology in the early, mid and even late-1900s were primarily focused on manual workers, technological development since the 2010s, in particular the rapid progress of Machine Learning (ML), has centred on the ability of computers to perform non-routine, cognitive tasks, and by consequence potentially affect white-collar or knowledge workers. In addition, these technological advancements have occurred in the context of much stronger interconnectedness of economies across the globe, leading to a potentially larger exposure than location-based, factory-level applications. Yet despite these developments, to an average worker, even in the most highly developed countries, the potential implications of AI have, until recently, remained largely abstract.

The launch of ChatGPT marked an important advance in the public's exposure to AI tools. In this new wave of technological transformation, machine learning models have started to leave the labs and begin interacting with the public, demonstrating their strengths and weaknesses in daily use. The chat function dramatically shortened the distance between AI and the end user, simultaneously providing a platform for a wide range of custom-made applications and innovations. Given these significant advancements, it is not surprising that concerns over potential job loss have resurged.

While it is impossible to predict how generative AI will further develop, the current capabilities and future potential of this technology are central to discussions of its impact on jobs. Sceptics tend to believe that these machines are nothing more than “stochastic parrots” – powerful text summarizers, incapable of “learning” and producing original content, with little future for general purpose use and unsustainable computing costs (Bender et al. 2021). On the other hand, more recent technical literature focused on testing the limits of the latest models suggests an increasing capability to carry out “*novel and difficult tasks that span mathematics, coding, vision, medicine, law, psychology and more*”, and a general ability to produce responses exhibiting some forms of early “reasoning” (Bubeck et al. 2023). Some assessments go as far as suggesting that machine learning models, especially those based on large neural networks used by Generative Pre-trained Transformers (GPT, see Text Box 1), might have the potential to eventually become a general-purpose technology (Goldfarb, Taska, and 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.

As social scientists, we are not in position to take sides in these technical debates. Instead, we focus on the already demonstrated capabilities of GPT-4, including custom-made chatbots with retrieval of private content (such as collections documents, e-mails and other material), natural language processing functions of 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

¹ The three main characteristics of general-purpose technologies are pervasiveness, ability to continue improving over time, and ability to spawn further innovation (Jovanovic and Rousseau, 2005).

speech-to-text and text-to-speech generation, potentially expanding their interaction with different types of human tasks. Finally, the potential of interacting with live web content through custom agents and plugins, as well as the multimodal (not exclusive to text, but also capable of reading and generating image) character of GPT-4 makes it likely that this type of technology will expand into new areas, thereby increasing its impact on labour.

Departing from these observations, this study seeks to add the global perspective to the already lively debate on possible changes that may result in the labour markets as a consequence of the recent advent of generative AI. We stress the focus of our work on the concepts of “exposure” and “potential”, which does not imply automation, but rather lists occupations and associated employment figures for jobs that are more likely to be affected by GPT-4 and similar technologies in the coming years. The objective of this exercise is not to derive headline figures, but rather to analyse the direction of possible changes in order to facilitate the design of appropriate policy responses, including the possible consequences on job quality.

The analysis is based on 4-digit occupational classifications and their corresponding tasks in the ISCO-08 standard. It uses the GPT-4 model to estimate occupational and task-level scores of exposure to GPT technology and subsequently links 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 that have a high automation potential and discuss how the automating and augmenting effects will strongly depend on a range of additional factors and specific country context.

We discuss the results of this analysis in the broader context of labour market transformations. We put particular focus on the current disparities in digital access across countries of different income levels, the potential for this new wave of technological transformation to aggravate such disparities, and the ensuing consequences on productivity and income. We also give consideration to jobs with highest automation and augmentation potential and discuss gender-specific differences. The analysis does not take into account the new jobs that will be created to accompany the technological advancement. Twenty years ago, there were no social media managers, thirty years ago there were few web designers, and no amount of data modelling would have rendered a priori predictions concerning a vast array of other occupations that have emerged in the past decades. As demonstrated by Autor et al. (2022), some 60 per cent of employment in 2018 in the United States was in jobs that did not exist in the 1940s.

Indeed, the main value of studies such as this one is not in the precise estimates, but rather in understanding the possible direction of change. Such insights are necessary for proactively designing policies that can support orderly, fair, and consultative transitions, rather than dealing with change in a reactive manner. For this reason, we also emphasize the potential effects of technological change on working conditions and job quality and the need for workplace consultation and regulation to support the creation of quality employment and to manage transitions in the labour market.

We hope that this research will contribute to needed policy debates on digital transformation in the world of work. While the analysis outlines potential implications for different occupational categories, the outcomes of the technological transition are not pre-determined. It is humans that are behind the decision to incorporate such technologies and it is humans that need to guide the transition process. It is our hope that this information can support the development of policies needed to manage these changes for the benefit of current and future societies. We intend to use this broad global study as an opening to more in-depth analyses at country level, with a particular focus on developing countries.

► **Text Box 1: What are GPTs?**

Generative Pre-Trained Transformers belong to the family of Large Language Models – a type of Machine Learning model based on neural networks. The “generative” part refers to their ability to produce output of a creative nature, which in language models can take the form of sentences, paragraphs, or entire text structures, with characteristics often undistinguishable from that produced by humans. “Pre-trained” refers to the initial training on a large corpus of text data, typically through unsupervised or self-supervised learning, during which the model learns about the text structure by temporarily masking part of the content and trying to minimize errors in the prediction of the masked words. Following pre-training, such models are further fine-tuned with the use of labelled data and so-called “reinforcement learning”, making them more suitable for specific tasks. This part of training is often perceived as a specialized job, executed by a handful of technical experts. In reality, it is labour intensive and involves many invisible contributors (Dzieza 2023). Its prerequisite is the production of vast amounts of labelled data, typically done by workers on crowdsourcing platforms. “Transformers” refer to the underlying model architecture, which uses numerous mechanisms, such as attention and self-attention frameworks, to develop weights related to the importance of text elements, such as words in a sentence, which are subsequently used for predictions (Vaswani et al. 2017).

While GPT specifically refers to models developed by OpenAI (GPT-1, 2, 3 and 4), this type of architecture is used by many more language models already available commercially. The launch of ChatGPT on 30 November 2022 made GPTs more popular among the public, as it made it possible for individuals with no programming knowledge to interact with GPT-3 (and eventually GPT-4) through a chatbot function with a human-like tone. For research purposes and more complex applications, such language models are typically more powerful when used through an Application Programming Interface (API). An API is a developer access point that relies on a query-response protocol with the use of programming software. In our case, we rely on a Python script based on OpenAI library, designed to connect to GPT-4 model, provide a fine-tuned prompt and receive a response, which is subsequently stored in a database on our server. This enables bulk processing of large numbers of requests and relies on the GPT-4 model with more parameters than what is accessible through the public Chat function.

▶ 1 Methods and Data

There are two principal approaches to the analysis of automation of occupations (Georgieff and Hye 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 and 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 its 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 and Hye 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 (2013, 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, and Rock 2018; Felten, Raj, and Seamans 2018; Felten, Raj, and Seamans 2019; Acemoglu and Restrepo 2020; Fossen and Sorgner 2022) or even combining occupational and job posting data (Georgieff and Hye 2021). Some authors have also taken up the challenge of producing better estimates for developing countries (Balliester and Elsheikhi 2018), often by trying to link detailed occupational data and automation scores from the US with less structured datasets available for lower-income countries (Aly 2020; Carbonero et al. 2023).

Calculating occupational scores typically involves development of a rubric, which 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 very 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 LLMs with similar capabilities as the latest GPT models. We build upon the method recently demonstrated by Eloundou et al. (2023), and replicated by Eisefeldt et al. (2023), which relies on the use of sequential API calls to GPT-4 model for the purpose of estimating task and occupational-level automation scores concerning this particular technology. We observe that their study demonstrates an astonishing proximity of GPT-4 predictions to the judgements made by a group of AI experts (albeit with a hard-to-determine level of possible bias on the human side). Applying a similar approach to the [International Standard Classification of Occupations \(ISCO-08\)](#), 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 combine the resulting score matrix with what has long been the comparative advantage of the International

Labour Organization (ILO): the ability to translate expert knowledge about occupations into global, regional and country-level employment estimates.

1.1. ISCO data on occupations and tasks

The current ISCO-08 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 harmonized 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 are some 7,500 jobs titles mapped to ISCO at 4-digits, which we also use as a robustness test for our analysis (see Section 3).

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

ISCO-08 1-digit code	ISCO-08 1-digit full label	Nr of distinct 1-digit codes	Nr of distinct 2-digit codes	Nr of distinct 3-digit codes	Nr 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

To build the principal data frame of tasks and occupations, we use as our foundation Part III of the official ISCO-08 documentation, which provides detailed definition and description 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 that can be identified in LFS, other household surveys and censuses, and other non-statistical sources, such as data derived from

administrative records. This means that they are designed to provide a common denominator for the variability of tasks under a given occupation. The number of ISCO-08 tasks assigned to each occupation can vary anywhere from 4 to 14. The data frame with a full list of ISCO-08 occupations and tasks constitutes the starting point of our estimations.

1.2. 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 fine-tune 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 specialist3. 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)}
```

By comparing the result with official ISCO-08 definitions, we examine the model's “understanding” of the ISCO-08 structure. We observe that 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 specialized 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 able to engage in more complex exchanges and develop logical links between different types of occupational classifications and tasks – a surprising and useful ability 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 ten-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

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.”

⁴ We observe that some functionalities of the model could potentially be used to support specialized discussions on updating the standard classifications. For example, semantic clustering and content summarization 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.

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 ten 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.

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

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



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

ISCO-08	ISCO Tasks		Justification by GPT	GPT-predicted Tasks		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 their 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 personalized 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 conduct activities with the 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 organize events, activities, and educational programs.	0.55	GPT technology can aid in planning, communication, and organization, but human interaction and collaboration with other staff members is still essential to successfully implement events and programs.
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, summarizing 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.

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 settled 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.

► 2 Assessment of the 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 (Table 3).

We conduct an additional test of scoring consistency across tasks (whether the model predicts 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 5 tasks randomly selected from all tasks on 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.

► **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, galvanizing, 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

As a parallel robustness test, we use a slightly modified prompt to generate occupational-level scores for over 7,500 job titles that can be found in different national labour force surveys, and which aggregate to the 436 ISCO 4-digit occupations. These jobs do not have detailed tasks, but a comparison of occupation level scores with the mean occupational scores generated based on detailed tasks reveals a proximity across the board. In other words, whether we rely on individual tasks that aggregate to occupations or a much larger pool of job titles to generate predictions, GPT-4 is consistent in the way it scores 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 similarity of GPT-4 scores with human-based scoring by AI experts on task-level questions, demonstrated in Eloundou et al. (2023), we believe that our exercise is likely to be estimating the upper bound of the exposure to GPT. This is explained by multiple reasons.

First, as recently 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 at the workplace, in particular at a level that would allow for full elimination of the human component. This is 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, the 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 of GPT's tech-oriented training data and trainers' profiles, as well as the literature on automation that most certainly was part of its training data, GPT is likely to reflect techno-optimism and overstate some task-level scores. GPT-generated scores also do not account for job-level task variation, which can lower occupational-level scores (Arntz et al., 2017). Second, our prompts to GPT focus on technical feasibility and ignore important determinants of technological diffusion, including the feasibility of adoption in a given environment which is dependent on constraints such as access to electricity and internet in countries with lower income, or local market dynamics, such as relative cost of labour to technology, level of digital literacy or access to finance. Third, despite having generated score predictions from prompts specifying whether the job is in a low- or a high-income country, we find that the difference between the two is too small to justify the use of both datasets. Therefore, for purposes of this initial analysis, we use the high-income country scores, with the understanding that this further contributes to estimating the upper bound of global exposure, since technological deployment faces additional barriers in lower-income countries. Nevertheless, this theoretical approximation facilitates an initial global picture of the potential impact on occupations across the globe, for which more detailed and contextualized studies will be needed.

Since the tasks in ISCO-08 documentation and those generated by GPT-4 do not correspond directly, we cannot compare the values of automation scores at the individual task level in the two data sets. 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, as opposed to some of the more generic tasks generated by GPT. In other words, confronted with a higher complexity of tasks captured in the ISCO-08 documentation, GPT-4 seems to attribute lower automation scores, when compared to its own collection of tasks, for which it tends to be more generous with automation potential. We treat the scores related to ISCO-08 tasks as the basis for further analysis, since they are directly linked to an international standard and associated ILO employment statistics.

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), which also explains the missing points on ISCO-08 tasks in Figure 1 in the following section. As the catch-all character of these few occupations does not permit the assignation of specific tasks, we drop them from the final analysis.

Finally, a classic challenge in analysing occupational tasks concerns attributing the share of time needed to execute the individual tasks in a given occupation (Carbonero et al. 2023). Time distribution likely varies in different country contexts, but unfortunately, the labour force and other survey data do not provide enough information to make country-level distinctions. The problem of attributing time weights across task-level scores is not exclusive to our attempts and typically appears in the construction of composite indicators related to technology and occupations (e.g. Autor and Dorn 2013; Brynjolfsson, Mitchell, and Rock 2018). One of the reasons why many studies on automation focus on the USA is that the level of detail in the O*NET data facilitates such

estimations. For our case, we opt for the most straightforward solution especially given the global focus, which is to apply equal weights to each task-level sub-component or each occupation.

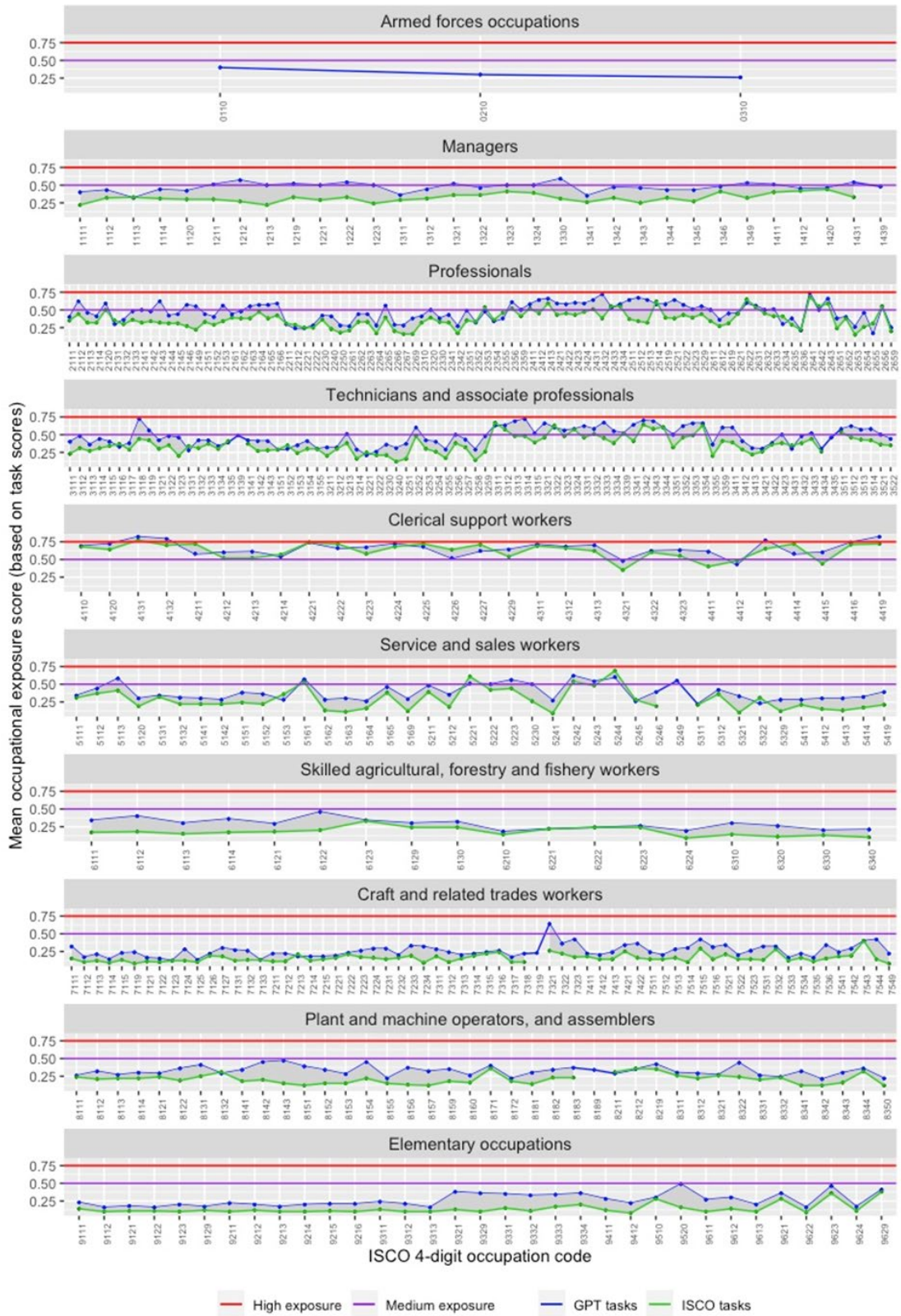
▶ 3 Results

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.

Figure 1 presents the breakdown, with the two upper limits of exposure marked with horizontal lines: 0.5 for medium exposure and 0.75 for high exposure. The grey area between the dotted lines represents the distance between the scores for each occupation based on ISCO-08 and GPT-predicted tasks. This illustration reveals a consistency among the scoring based on ISCO-08 and GPT-generated tasks, with highest exposure found amongst clerical support workers, followed by technicians and associate professionals, and by professionals. While these occupations have no official common category, they are broadly associated with “knowledge work” (Berg and Gmyrek 2023).

► Figure 1. Mean automation scores by occupation, based on ISCO and GPT tasks



— High exposure — Medium exposure — GPT tasks — ISCO tasks

In addition, the broad category of managers, which for the most part falls underneath the 0.5 cut-off, nonetheless approximates the line of medium exposure. The results for service and sales workers are more mixed, with some occupations surpassing the threshold but most others falling below. Plant and machine operators and assemblers, elementary occupations, craft and related trades workers and skilled agricultural, forestry and fishery workers have more limited exposure.

What drives these results? To answer this question, we apply machine learning techniques to the analysis of ISCO-08 tasks that have been classified as having a high level of exposure. First, we 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.⁶ We then perform semantic clustering of the tasks, based on the K-Means algorithm and a visual inspection of 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 4 presents the result of this exercise, with the corresponding tasks in each cluster and their individual scores.

► **Table 4: 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 optimize physical and chemical processes for several processing units;	0.80
	Preparing invoices and sales contracts and accepting payment;	0.80

⁶ Embeddings are a vectorial high-dimensional representation of the text, generated by an LLM. Standard Ada embeddings have 1536 dimensions.

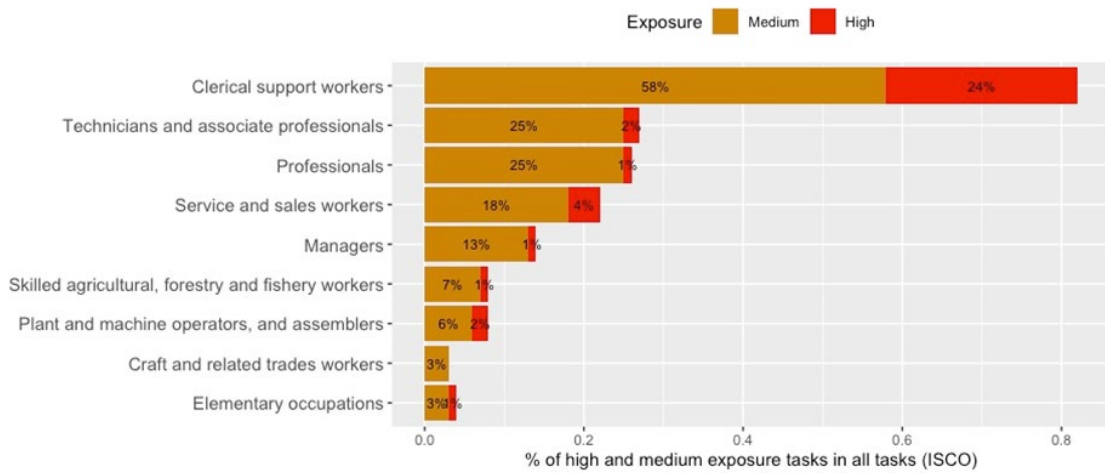
Thematic Group	Sample Tasks	Score
Information Processing and Language Services	Taking dictation and recording other matter in shorthand;	0.80
	Translating from one language into another and ensuring that the correct meaning of the original is retained, that legal, technical or scientific works are correctly 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

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

As the next step, we calculate the share of tasks with high and medium exposure in each ISCO 1-digit grouping. Figure 2 reveals in stark terms the degree of exposure among clerical support workers, among whom some 24 per cent of all tasks fall into the highly exposed category. If we also account for tasks with medium-level exposure (58 per cent of all tasks), a full 82 per cent of clerical job tasks are exposed at an above-average level. This stands in contrast to the other occupational groups, in which the highest share of highly exposed tasks oscillates between 1 and 4 per cent, and where the medium-exposed tasks do not exceed 25 percent.⁷ Even assuming large margins of error, the result is still striking.

⁷ Armed forces are absent from the figure, since they do not have any tasks scored at the level of medium and high exposure.

► **Figure 2. Tasks with medium and high GPT-exposure, by occupational category (ISCO 1-digit)**

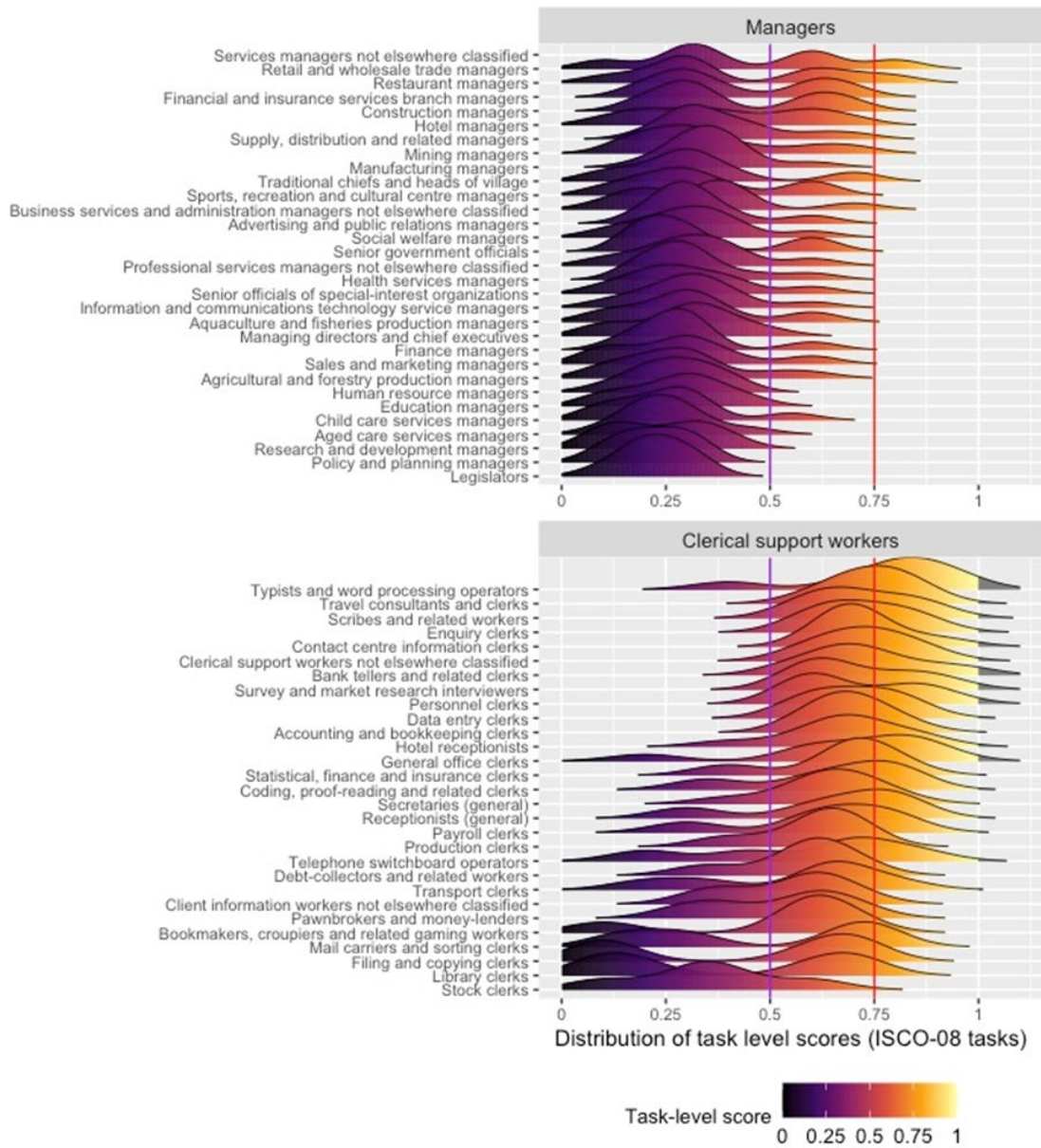


3.1. Automation vs augmentation: distribution of scores across tasks and occupations

In this next section, we analyse how the exposure to GPT-like technology could potentially affect occupations. Will the technology replace most tasks within an occupation, provoking job loss? 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 3 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 3. Box plot of task-level scores by ISCO 4d, grouped by ISCO 1d



To determine whether the technology has a greater potential for automation or augmentation across all ISCO-08 4-digit occupations, we use a method similar to Carbonero et al. (2023). Considering an occupation as a collection of tasks with different levels of exposure to a particular technology, we focus on two principal parameters of the task scores distribution: (i) the mean score for a given occupation, and (ii) its standard deviation (SD). Jobs with a high mean score and a 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 a 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 5).

► **Table 5: 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

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

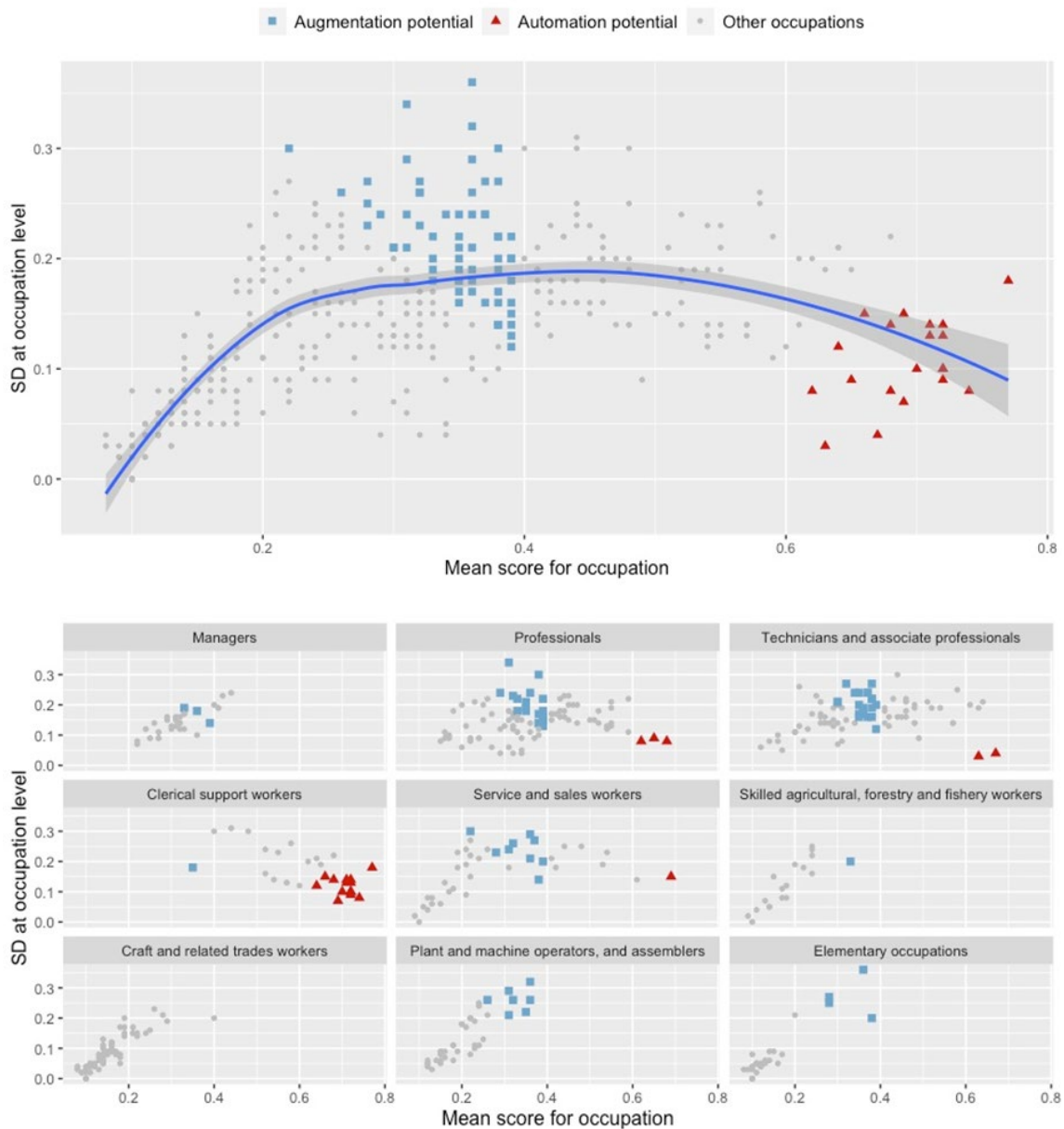
$$0.4 > \mu_i \text{ and } \mu_i + \sigma_i > 0.5 \quad (1.1)$$

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

$$\mu_i > 0.6 \text{ and } \mu_i - \sigma_i > 0.5 \quad (1.2)$$

Figure 4 provides two visual representations of this grouping: the top panel pools all occupational scores into one sample, while the bottom panel provides a more detailed breakdown by occupational category at ISCO-08 1-digit level. The blue trend line 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 the 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.

► **Figure 4. Augmentation vs automation potential at occupational level**



Based on these definitions, we produce two separate lists of occupations, one with a high automation potential and one with a high augmentation potential. Figure 5 lists occupations that not only have a high mean score across their tasks, but which also have a low SD, suggesting that the tasks' scores do not move far from the overall mean. This means that such jobs are mostly composed of tasks that could eventually be automated, provided that other conditions are in place.

► Figure 5. Occupations with high automation potential

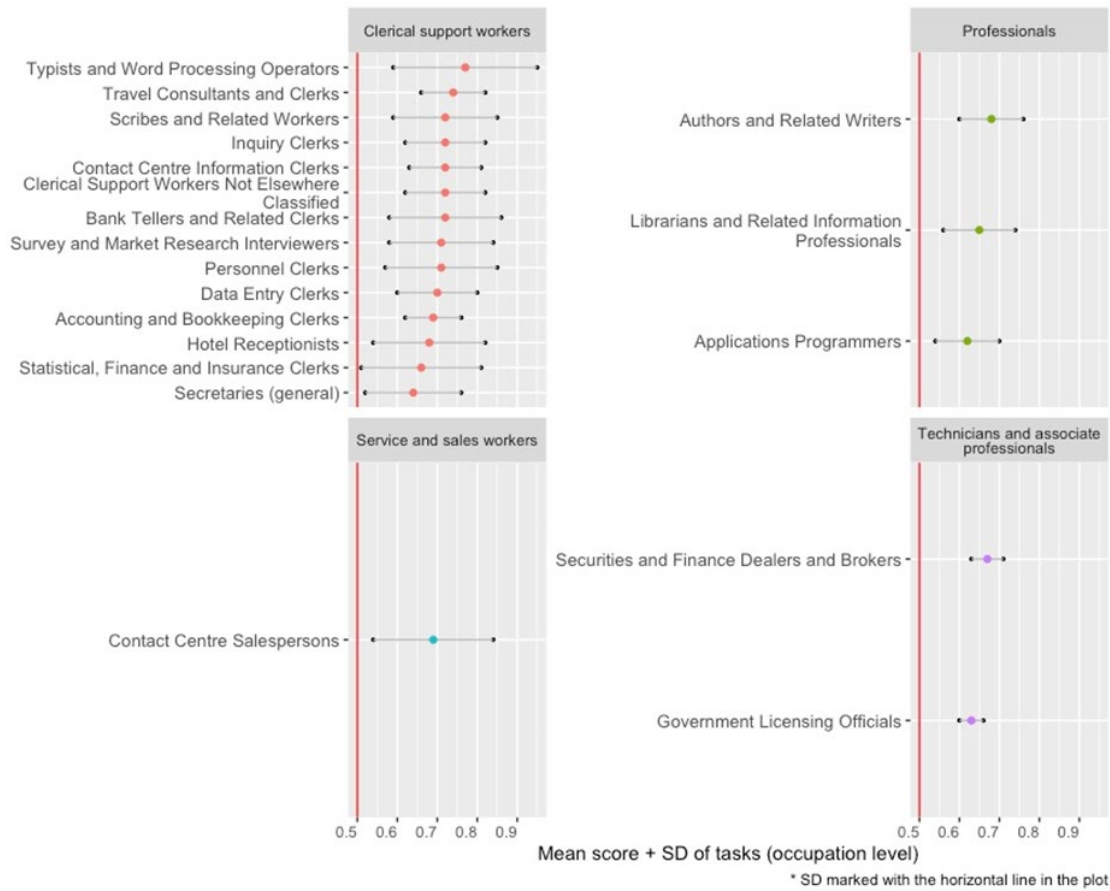


Figure 6, in turn, presents the occupations that have a low mean score and a high SD, with the sum of the mean and the SD reaching above the limit of medium exposure. Such jobs are most likely to experience an augmenting effect of GPT technologies, while still retaining an important human component.

► Figure 6. Occupations with high augmentation potential

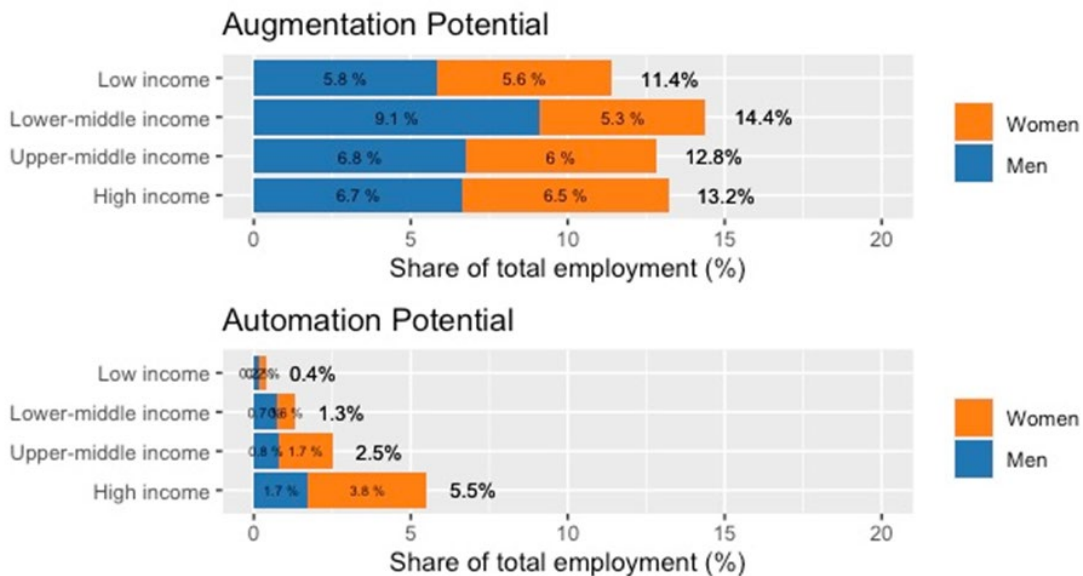


► 4 Exposed occupations as a share of employment: global and income-based estimates

4.1. Augmentation vs Automation: ILO microdata

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 enables extracting detailed country-level employment information. We use microdata for 59 countries that report 4-digit microdata in ISCO-08 format: 8 low-income countries (LIC), 24 lower-middle-income countries (LMIC), 19 upper-middle-income countries (UMIC) and 8 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 the 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 visualized in Figure 7a.⁸

► Figure 7a. Automation vs augmentation potential: shares of total employment, microdata for 59 countries



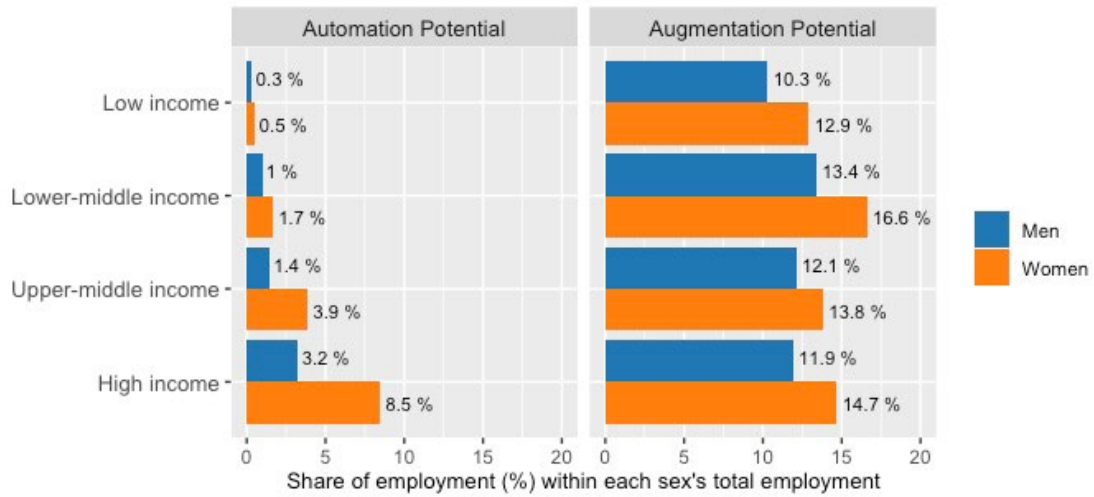
Several elements stand out in this comparison. First, occupations with high augmentation potential constitute a significantly larger share of the total employment in each income group than the jobs with high automation potential. In the LMICs, such jobs have the highest share of the employment distribution, with 14.4 per cent of total employment classified in this category. Second, augmentation-related jobs have a fairly equal gender distribution, with the shares of such jobs being held by men visibly higher only in the LMICs.

⁸ 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 weighted-median. Since the results are stable and very similar in both cases, we keep the weighted mean as the main calculation method.

Contrasting with that, occupations with high automation potential show significant differences across income groupings of countries and the visible trend is that they increase their share in the overall employment together with the countries' income levels. In the LICs, only some 0.4 per cent of total employment falls into this category, whereas in the HICs the share of such occupations rises to 5.5 per cent. In addition, the share of female participation in these occupations also grows with countries' income levels, and in the HICs it is more than double the male share of total employment.

This effect becomes even more apparent if we present the jobs with high automation and augmentation potential as a share of total employment for each sex. As demonstrated in Figure 7b, in high-income countries, jobs with high automation potential constitute 8.5 per cent of female employment, compared to 3.9 per cent of male employment. In addition, the share of jobs with high augmentation potential is visibly higher among women than among men in all income groups.

► **Figure 7b. Automation vs augmentation potential: shares of total employment in each sex (ILO microdata)**



4.2. Augmentation vs Automation: global estimate

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 lower-digit ISCO-08 levels. We thus exploit this greater data availability and move up the cascading structure of ISCO-08 system with each stage of estimations (see Table 6).

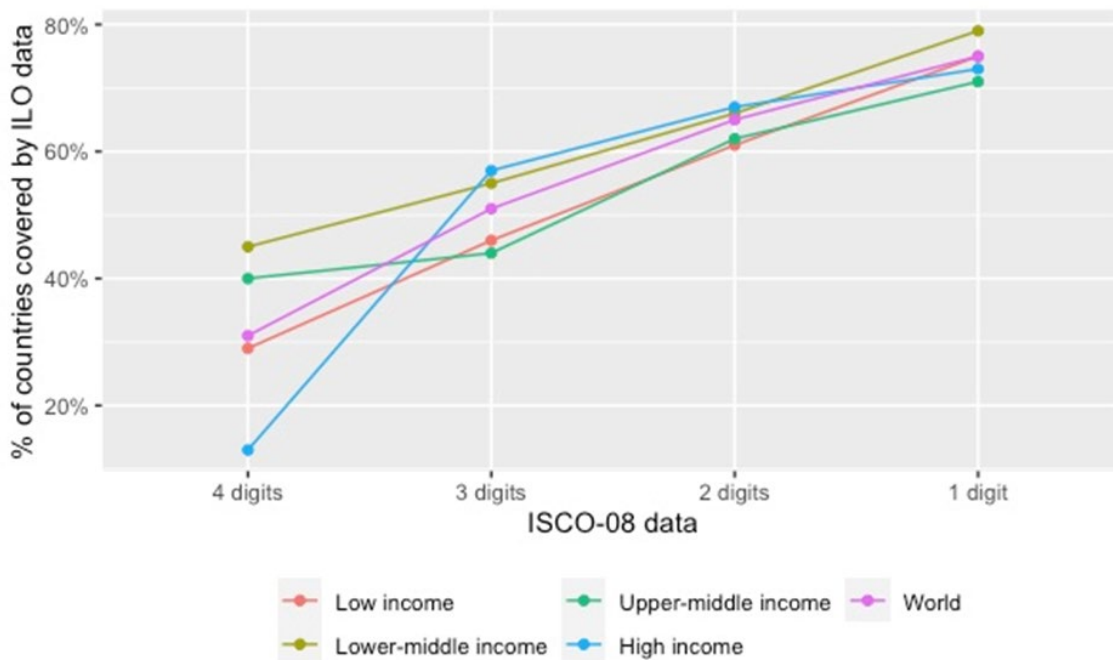
► **Table 6: Microdata coverage by levels ISCO-08: number of countries**

Income Group	ISCO-08 1-digit	ISCO-08 2-digit	ISCO-08 3-digit	ISCO-08 4-digit
HIC	44	40	34	8
UMIC	34	30	21	19
LMIC	42	35	29	24
LIC	21	17	13	8
World	141	122	97	59

We start by calculating the share of jobs in categories of automation and augmentation potential in total employment for each of the 59 countries with available 4-digit data. We then calculate the weighted mean for each income group, as previously done for Figure 7. As the next step, we calculate for these countries the share of these isolated jobs in the total jobs covered by a higher-digit category, in this case ISCO-08 at 3-digit level. Subsequently, we calculate the weighted mean of these shares at ISCO-08 3-digit for each of the income groups and apply these to estimate the number of jobs in the countries for which we have ISCO-08 3-digit data, but for which ISCO 4-digit data was missing. We then repeat an analogical procedure moving up the data coverage ladder, that is, from ISCO 3-digit to 2-digit and, finally, from 2-digit to 1-digit. At this level we arrive at an estimation that relies on data available for 141 countries, which ensures a broad coverage of data points from ILO's repository (Figure 8). The final batch of 48 countries still missing at this point is estimated using the same method, thereby aligning our calculations with the total employment figures in the official global employment estimates of the ILO for 2021, available for 189 countries.⁹

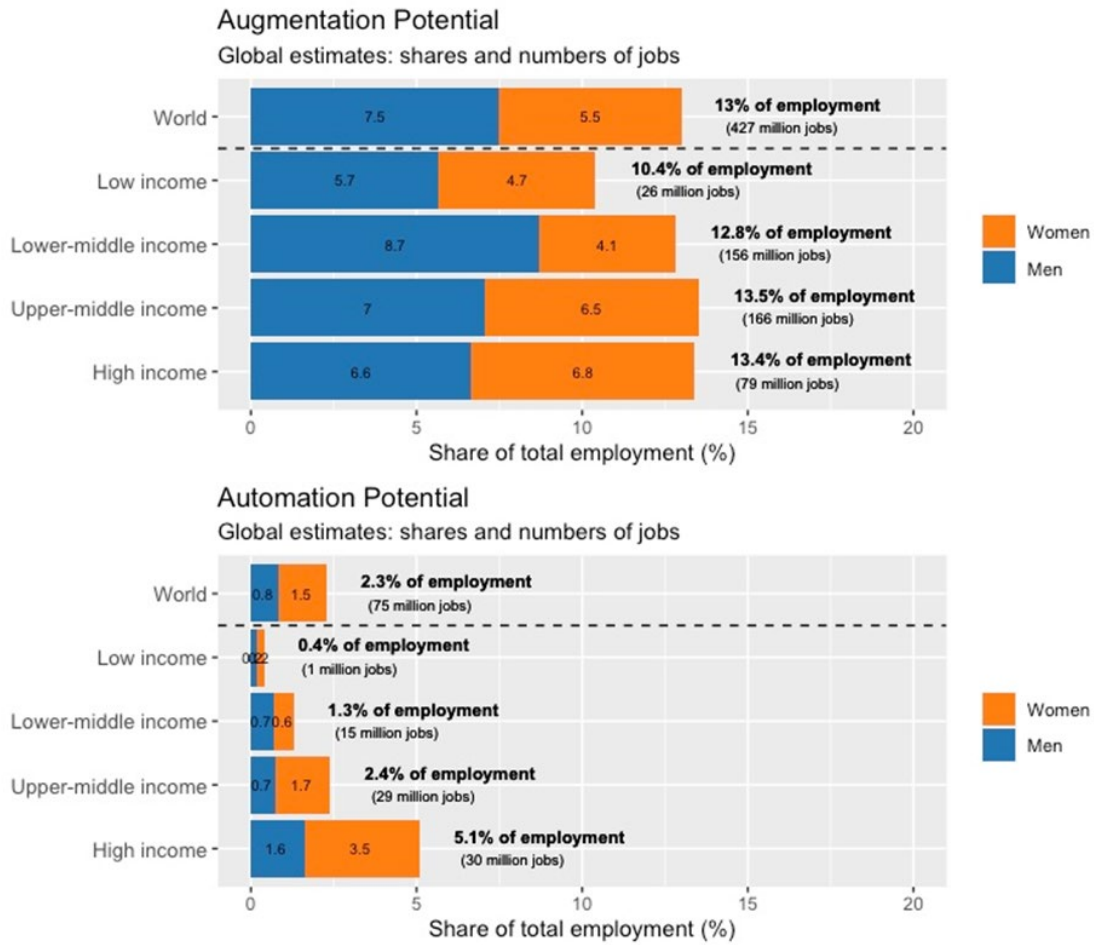
⁹ See Appendix for details.

► Figure 8. Country coverage based on the level of digits in ISCO-08 (ILO data)¹⁰



¹⁰ Total refers to countries and income groupings used in ILO modelled estimates (<https://ilostat.ilo.org/resources/concepts-and-definitions/classification-country-groupings/>).

► **Figure 9a. Global estimates: jobs with augmentation and automation potential as share of total employment**

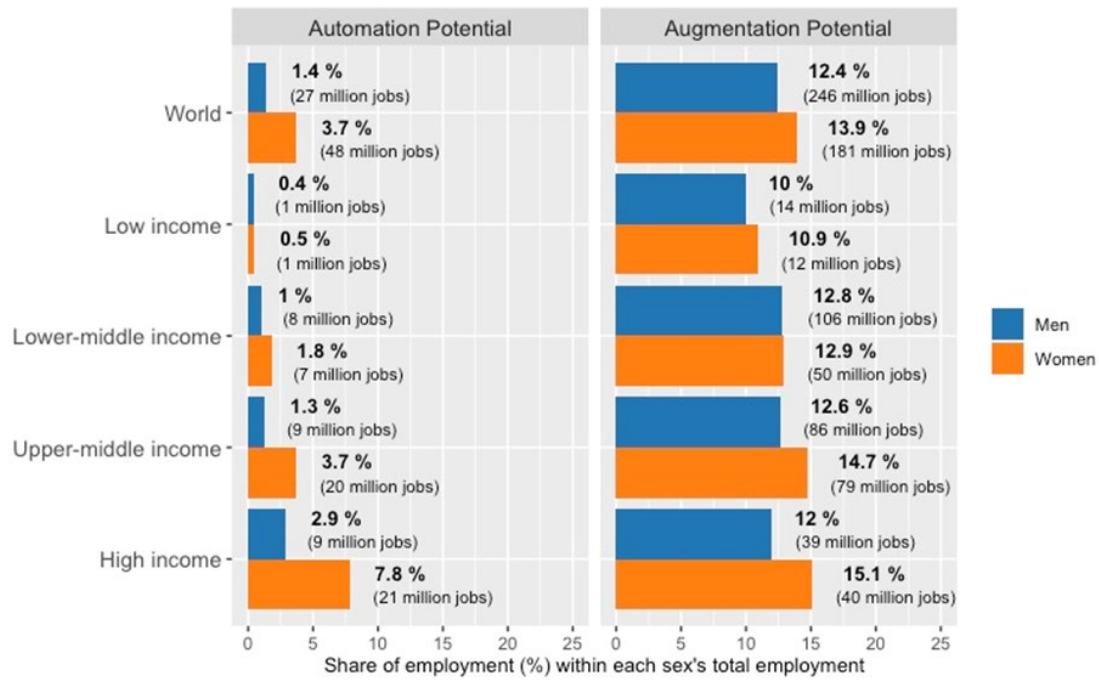


Given the data limitations, the exact numbers presented in Figure 9a should be read as an indication of a general trend, based on the best employment estimate that can be produced at the global level for a selection of 4-digit ISCO-08 occupations. More importantly, the global estimate confirms the trends already observed based on the analysis of microdata for 59 countries (Figures 7a-b). Specifically, it confirms that the number of jobs in the augmentation category is significantly higher than the number of jobs that have a 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 per cent respectively) and the LICs having the lowest share (10.4 per cent). This means that, once the size and employment distribution aspects of individual countries are considered in the estimate, globally, the share of jobs potentially exposed to automation with generative AI of similar properties as the current GPT technology grows with income, but so does the share of jobs that have a 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. We discuss these differential effects in more detail in section 6.1.

The global estimates also confirm the strong gender effect observed in the microdata (Figure 7b). When we disaggregate the estimate to shares of female and male employment (Figure 9b), we observe that 3.7 per cent of all female employment in the world is in jobs that are potentially automatable with generative AI technology, compared with only 1.4 per cent of male employment. In high-income countries, the share of potentially affected female jobs is 7.8 per cent, more than double the 2.9 per cent of male jobs for that income group. At the same time, the share of jobs

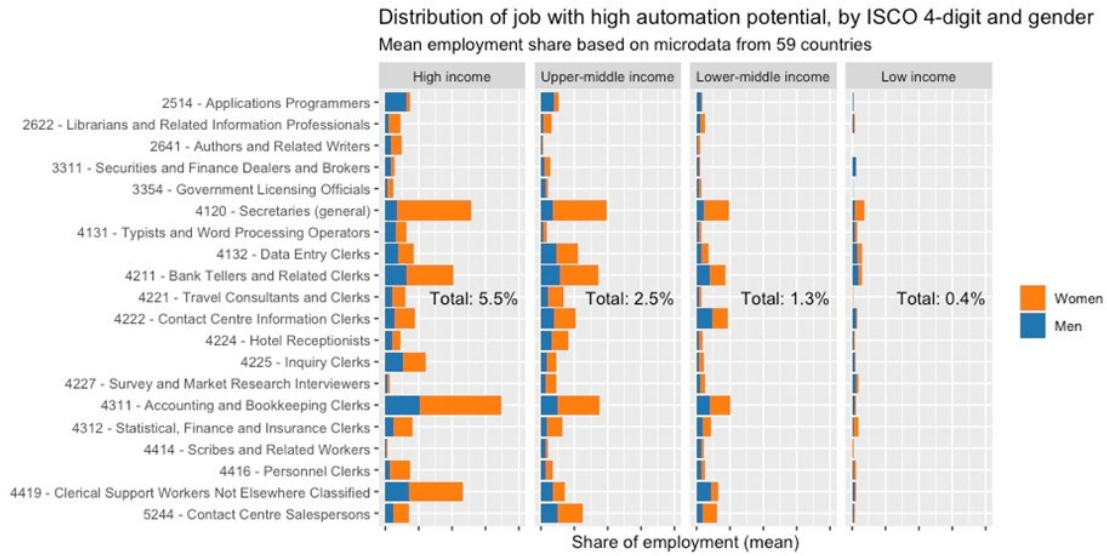
with high augmentation potential is also greater among female than male jobs across all income groups. This suggests that any form of technological transition would have a strongly gendered effect, with a badly managed process disproportionately harming women, and a well-managed transition potentially creating important opportunities in terms of women's empowerment.

► **Figure 9b. Automation vs augmentation potential: shares of total employment for each sex (global estimate)**



To further illustrate the origins of these discrepancies, it is helpful to consider a 4-digit breakdown of the occupational structures across country groups. Figure 10 presents a selection of ISCO 4-digit occupations with high automation potential, based on the mean share of each occupation in total employment, for each income group. While the low number of responses underpinning some of the bars would not qualify this breakdown as statistically representative, it still provides useful insight into the overall differences in the employment structures of countries with different income levels.

► **Figure 10. Occupations with high automation potential, by ISCO 4-digit and income group**

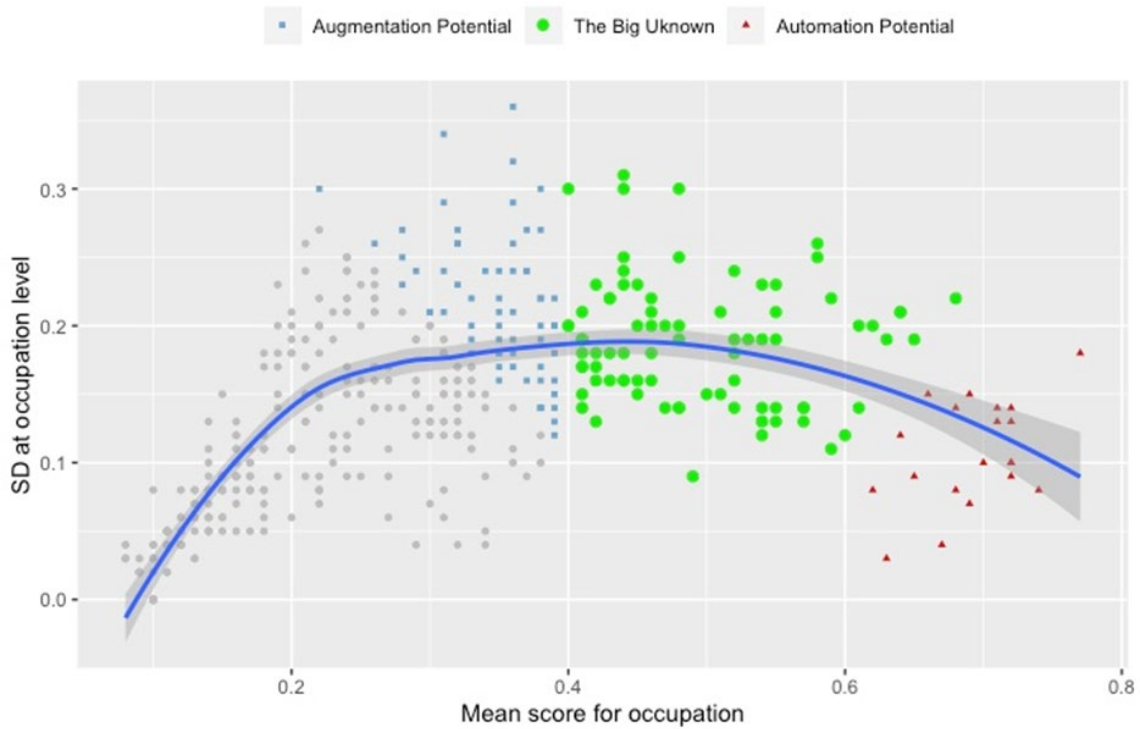


We can observe that the general trend is for the share of clerical occupations to grow with income, which explains the disproportionately higher potential automation effects in wealthier economies. For example, jobs of secretaries, accounting and bookkeeping clerks, or bank tellers and cashiers enjoy a nearly linear relationship between the country’s income and the share of employment they take. This clearly reflects the general trend of the last decade, which saw many call centre and client service jobs outsourced to locations outside high-income countries. In addition, as previously discussed, such jobs are disproportionately held by women and this pattern remains visible across occupations even at the very detailed breakdown to ISCO-08 4-digits. There are, however, a few notable exceptions to this rule. For example, occupations of contact centre salespersons and data entry clerks are relatively more present in the middle-income countries than in the high-income countries, while the jobs of application programmers are strongly dominated by men.

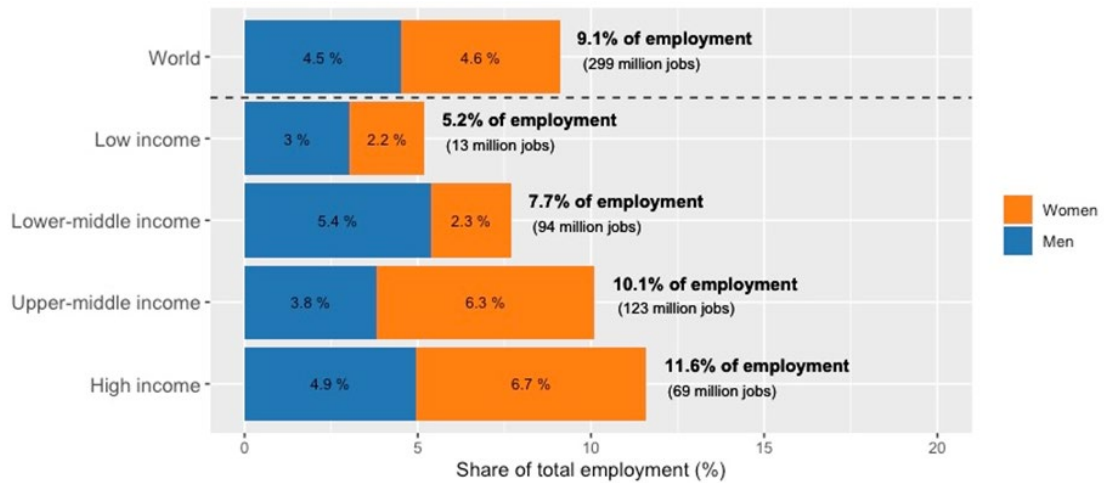
4.3. The big unknown

The breakdown of occupations into high automation and augmentation potential provided a helpful framework to discuss the extremes of scores’ distribution, thereby minimizing 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 11 a-b 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. 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.

► Figure 11a. The “Big Unknown”: occupations between augmentation and automation potential



► Figure 11b. The “Big Unknown”: share of total employment, by income group (global estimate)



As illustrated in Figure 11b, these occupations constitute a nontrivial share of the global employment, with some 8.6 per cent and 281 million workers falling into this category. While in the 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.

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

5.1 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 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 work¹¹, figure prominently in some of these countries, particularly India and the Philippines, which dominate the world's call centre industry. In the Philippines, a half million people were employed in call centres in 2016, of whom 53 percent were women (DOLE, 2018).¹²

The challenges, and consequences, of such adjustments should not be underestimated. For example, a study of the effects of automation on Dutch workers during 2010-2016, found that workers made redundant as a result of automation experienced a 5-year cumulative wage income loss of 9 per cent of an annual wage (Bessen et al., 2019). The losses were only partially offset by various benefits systems, despite the relatively robust Dutch 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 is critical for managing the transition process as it encourages redeployment and training over job loss. The ILO's Employment Protection Convention (No. 158, 1982) includes provisions on the termination of employment for technological reasons. It advocates, particularly in cases of collective dismissals, special procedural requirements including consultations of the employer with workers' representatives, notifications to the competent authorities, undertaking measures to avert or minimize terminations and to mitigate their effects, and establishing criteria for selection for termination and

¹¹ (4222) contact centre information clerks, (4227) survey and market research interviewers, (5244) customer contact salesperson

¹² In 2023, the ITBPA (IT and Business Processing Association) of the Philippines stated that the sector employed 1.5 million full-time equivalent employees in 2022 (ITBPA, 2023).

priority of rehiring. The aim of such requirements is to minimize the negative externalities from dismissal, especially when collective, as well as to better internalize the cost of such dismissals and support an orderly process that balances the needs of workers, employers, and societies at large (Aleksnyska and 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. As Figure 9 showed, the potential exposure to automation disproportionately affects the share of women's employment by more than two-fold in high-income countries (7.9 per cent vs 2.9 per cent) and upper-middle-income countries (2.7 per cent v 1.3 per cent). Concentrated job losses in female-dominated occupations could threaten advances made in the past decades in increasing women's labour market participation.

The care economy, comprising both health care and education, traditionally employs a greater share of women, yet these are also sectors that suffer from underinvestment. According to the ILO (Addati et al. 2018), achieving the SDG targets would more than double employment in these sectors from 206 million in 2015 to 475 million in 2030. In addition, some care occupations, such as in long-term care, whose demand is projected to increase substantially in the next decades due to ageing populations, are often characterised by poor working conditions. Meeting the demand for workers in this sector and improving their job quality so that they are a decent source of employment, would be a means to not only provide a potential source of decent work for displaced workers, but would also help meet societies' need for more care work. Shifting to these opportunities, however, will require greater investment in the sectors, in addition to training and income support during the transition.

Another source of policy intervention is to ensure quality of the new jobs created as a result of technological change. The development of AI relies on tagging and repetitive feedback done by humans, in what is known as "microtask" work (Irani 2015; Tubaro, Casilli, and Coville 2020). For LLMs in particular, human workers train, mould and evaluate the systems through "reinforcement learning", in order to ensure the safety of such systems as well as improve accuracy (Xu et al., 2023). While no global figures exist on the number of microtask workers, estimates from the mid-2010s suggested 9 million workers from across the globe (Kuek et al. 2015). This figure has most certainly grown since then and is likely to continue to expand, as new and often small players enter the market of LLMs. A recently leaked note from Google's engineers noted that "*the barrier to entry for [LLM] training and experimentation has dropped from the total output of a major research organization to one person, an evening, and a beefy laptop*" (Patel and Ahmad 2023). This dramatic decrease in the cost and ease of entry to the LLM market points to an increase in demand for domain-specific labelled datasets curated by microtask workers.

Much microtask work has been conducted on digital labour platforms, either through crowdsourcing websites or through businesses processing firms that directly hire workers. Microtask jobs mediated through crowdsourcing platforms, are paid by the task and regulated by civil contracts, meaning that the workers have none of the labour protections or social security benefits that come with the employment relationship. The poor working conditions of much platform work prompted ILO constituents to agree to a two-year standard setting discussion beginning on 2025 with a view to crafting an international labour standard on decent work in the platform economy that can guide national regulation (ILO 2023d).

5.2 Ensuring job quality under augmentation

Technology can also affect job quality in its application at the workplace. While the technology can allow the more routine tasks that one does to be automated, potentially leaving time for more engaging work, it can also be implemented in a way that limits workers agency or accelerates work intensity. Concerns over AIs integration at the workplace has focused on the growth of

algorithmic management, essentially work settings in which *“human jobs are assigned, optimized, and evaluated through algorithms and tracked data”* (Lee et al. 2015). Algorithmic management is a defining feature of digital labour platforms, but it is also pervasive in offline industries such as the warehousing and logistics sectors. In warehouses an automated, “voice-picking” system directs warehouse staff to pick certain products in the warehouse, while using data collection to monitor workers and set the pace of work (Matopoulos 2011). Besides lacking autonomy to organize their work or set its pace, workers also have little ability to provide feedback or discuss with management about the organization of work (Wood, 2021). The integration of generative AI into other fields such as banking, insurance, social services, and customer service more broadly may have similar effect.

Technological advancements are often felt more immediately at the workplace level and are usually best addressed at the workplace. As a result, whether the effect of technology on working conditions is positive or negative depends in large part on the voice that workers have in the design, implementation and use of technology. Having such voice relies in turn on the opportunities for worker participation and dialogue. This can take place either through formalized settings, such as works councils or guidance provided in collective bargaining agreements, or less formally, in workplaces where there is a high degree of employee engagement, such as in organizational structures that support teamwork, problem-solving and decentralized decision-making (ILO 2023c). Studies on Europe have shown that it is the countries with stronger and more cooperative forms of workplace consultation, essentially the Nordic countries, followed by Germany, where workers are more open to technological adoption at the workplace. Yet even in Denmark, focus group discussions with workers on digital integration reveal a desire for greater attention to the implementation and organization of technology at the workplace so as to better meet the needs of end users (Refslund and Borello, 2023).

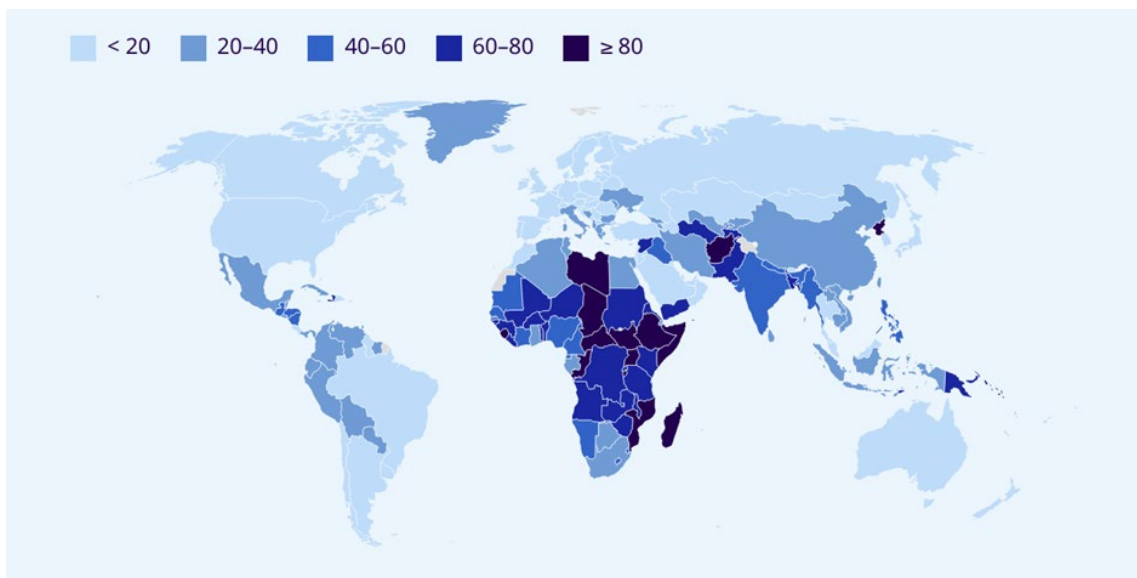
In addition to consultation at the workplace, there is also need for laws that regulate AI’s application at the workplace. To date, much of the discussion on regulation of AI has ignored its possible effects on working conditions (Moore 2023). Where there has been discussion, the focus has overwhelmingly been on voluntary standards of AI ethics, ignoring the uneven power relations inherent in working relationships (Cole et al. 2022). AI tools may aggravate power relations at the workplace, especially if workers cannot have access to the data used to survey their activities, if there are no mechanisms in place to assess the ex-post use of the technology in the workplace, or if decisions on dismissal are taken without proper recourse to conflict resolution mechanisms. Adams-Prassl et al. (2023) advocate for a prohibition of worker monitoring and data collection outside of work (temporally or geographically) or in contexts where it poses risks to human dignity or the exercise of fundamental rights, in addition to other limitations. The design and application of such regulations is best crafted through tripartite systems, in which workers’, employers’ and governments representatives engage with equal voice. The negotiations should build on existing tripartite consultation mechanisms and structures and use the already existing labour rights and norms as the point of departure. Giving the quickly evolving nature of AI and its iterative learning process, mechanisms for ex-post evaluation and tripartite governance need to be built into the regulation.

5.3 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 the high- and low-income countries. Larger shares of jobs falling into the augmentation category suggest that, at least in near future, generative AI systems similar to GPT 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 per cent of employment in these countries is found in the potential augmentation category, in practice potential benefits of GPT 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, one-third of the global population, corresponding to some 2.7 billion people, still did not have access to the internet (Figure 11). Among the two-thirds that do have access, many would not be able to use GPT technologies 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.¹³

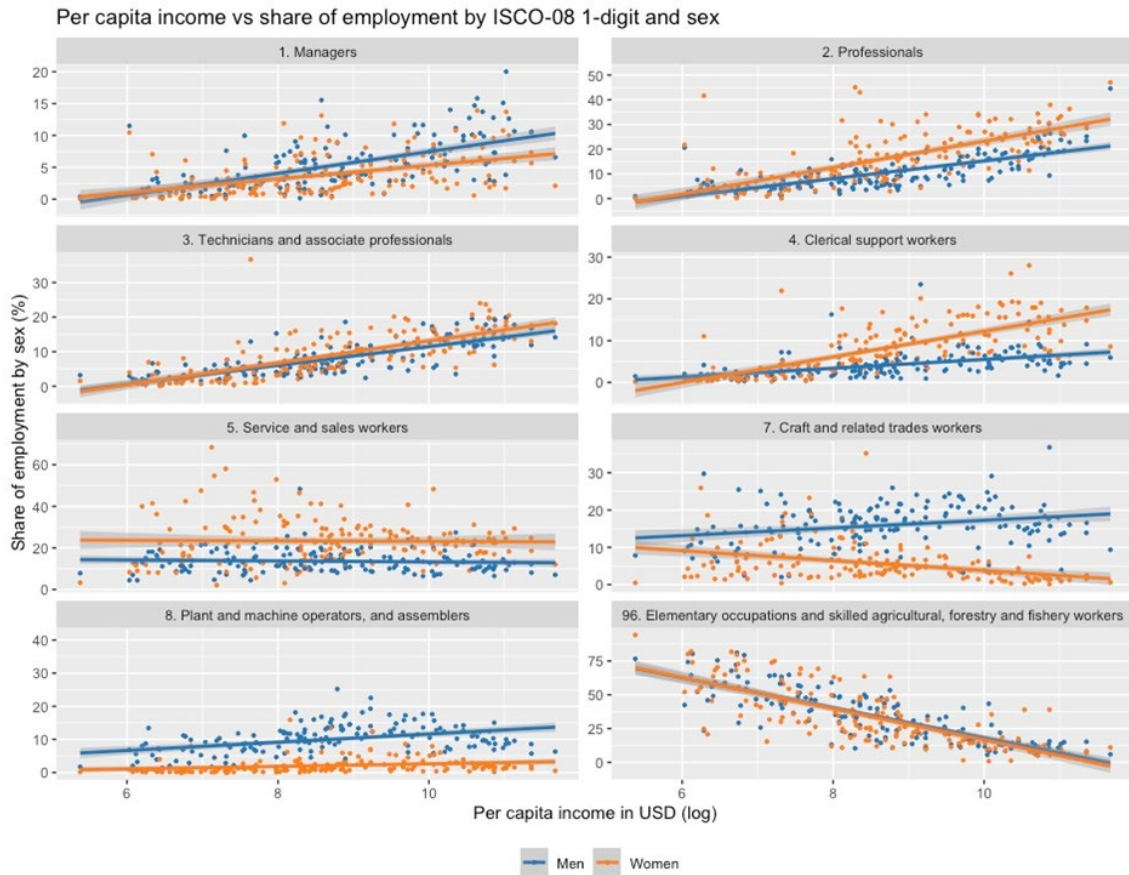
► **Figure 11. Share of population not using the internet¹⁴**



¹³ <https://www.enterprisesurveys.org/en/data/exploretopics/infrastructure>

¹⁴ Authors' calculations based on available country data for the most recent year (ITU 2023). Map created with Datawrapper. The boundaries shown, designations used, and any other information shown does not imply official endorsement or acceptance by the International Labour Organization.

► **Figure 12. A classic growth path: income and occupational diversification**



On the other hand, 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 such 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 creation of a network of 110,000 agents, 40 times the number of bank ATMs in Kenya (Buku and 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 the 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 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 the developing countries, further distinction needs to be made. While middle-income countries, are more exposed to the automating effects of GPT technologies, 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. We also highlighted the possible consequences for job quality, in order to draw attention to this important effect on the world of work that has too often been ignored in discussions of digital technologies' impact on labour.

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 ignores the more limited potential for deployment in lower-income countries, where the necessary infrastructure is typically of lower standard, 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 when it comes to ease of deployment, that in practice is difficult to operationalize. Third, without being able to make reliable predictions on future technological progress, we focused on the potential of task automation as of today, without speculating on the numbers of new jobs that might emerge. This approach might have been expected to generate alarming estimates of net job loss – but it did not. Rather, 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 LLMs, and GPT more specifically. 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 and 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 and 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 and 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 are the clerical jobs, where the majority of tasks fall at least into medium-level exposure, and about a quarter of tasks are highly exposed to potential automation. As a result of technological progress, many such jobs might never emerge in developing countries, where they traditionally served as a vehicle for increasing female employment. For other types of “knowledge work”, exposure is only partial, suggesting a stronger augmentation potential and productivity benefits, rather than job displacement.

These findings align with some of the most recent literature on generative AI systems with a global focus. A recent study by 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. WEF's global survey, focussed on large enterprises, also lists clerical and administrative jobs among occupations with fastest expected declines (WEF 2023). Estimates provided 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 character than the US-oriented O*NET database. Second, the application of ILO’s 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 numbers of jobs in both the augmentation and automation categories identified in our paper, our analysis suggests that the general 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 elements for managing AI’s deployment 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 brutal. 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.

Appendix 1. 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, denote the total employment as T_{IG} . The total employment in each income group is the sum of the 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 $J_{3_{di}}$ is given by:

$$J_{3_{di}} = \sum_{k \in D_{3_{di}}} J_{4_{ki}}$$

where $J_{4_{ki}}$ 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 $S_{3_{di}}$ 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:

$$S_{3_{di}} = \frac{A_{di}}{J_{3_{di}}}$$

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

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

$$WMS_{IG} = \frac{\sum_{i \in IG} E_i * S_{3_{di}}}{\sum_{i \in IG} E_i}$$

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 the 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.

References

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo. 2022. "Artificial Intelligence and Jobs: Evidence from Online Vacancies." *Journal of Labor Economics* 40 (S1): S293–340. <https://doi.org/10.1086/718327>.

Acemoglu, Daron, and Pascual Restrepo. 2020. "Robots and Jobs: Evidence from US Labor Markets." *Journal of Political Economy* 128 (6): 2188–2244. <https://doi.org/10.1086/705716>.

Adams-Prassl, Jeremias, Sangh Rakshita, Halefom Abraha, and M. Six Silberman. 2023. "Towards an International Standard for Regulating Algorithmic Management: A Blueprint." In . Vol. Paper prepared for presentation at the "8th Conference of the Regulating for Decent Work Network" on Ensuring decent work in times of uncertainty at the International Labour Office Geneva, Switzerland.

Addati, Laura, Umberto Cattaneo, Valeria Esquivel, and Isabel Valarino. 2018. "Care Work and Care Jobs for the Future of Decent Work." Report. http://www.ilo.org/global/publications/books/WCMS_633135/lang--en/index.htm.

Aleksnyska, Mariya and Muller, Angelika. 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, Melanie, Terry Gregory, and Ulrich Zierahn. 2017. "Revisiting the Risk of Automation." *Economics Letters* 159 (October): 157–60. <https://doi.org/10.1016/j.econlet.2017.07.001>.

Autor, David 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, David H., and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review* 103 (5): 1553–97. <https://doi.org/10.1257/aer.103.5.1553>.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 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*, 610–23. FAccT '21. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>.

Berg, Janine, and Pawel Gmyrek. 2023. "Automation Hits the Knowledge Worker: ChatGPT and the Future of Work." In . UN Multi-Stakeholder Forum on Science, Technology and Innovation for the SDGs (STI Forum) 2023. <https://papers.ssrn.com/abstract=4458221>.

Brynjolfsson, Erik, Tom Mitchell, and Daniel Rock. 2018. "What Can Machines Learn and What Does It Mean for Occupations and the Economy?" *AEA Papers and Proceedings* 108 (May): 43–47. <https://doi.org/10.1257/pandp.20181019>.

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, et al. 2023. "Sparks of Artificial General Intelligence: Early Experiments with GPT-4." arXiv. <http://arxiv.org/abs/2303.12712>.

Buku, Mercy W., and Michael W. Meredith. 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, Bernardo, Marco Grazzi, Martina Occelli, and Marco Sanfilippo. 2022. *Mobile Internet, Skills and Structural Transformation in Rwanda*. ILO Working Paper. ILO. <https://doi.org/10.54394/XSTK4695>.

Cammeraat, Emile, and Mariagrazia Squicciarini. 2021. "Burning Glass Technologies' Data Use in Policy-Relevant Analysis: An Occupation- Level Assessment," April. <https://doi.org/10.1787/cd75c3e7-en>.

Carbonero, Francesco, Jeremy Davies, Ekkehard Ernst, Frank M. Fossen, Daniel Samaan, and Alina Sorgner. 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*, February. <https://doi.org/10.1007/s00191-023-00809-7>.

Cherry, Miriam 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>.

Cole, Matthew, Callum Cant, Funda Ustek Spilda, and Mark Graham. 2022. "Politics by Automatic Means? A Critique of Artificial Intelligence Ethics at Work." *Frontiers in Artificial Intelligence* 5. <https://www.frontiersin.org/articles/10.3389/frai.2022.869114>.

Dzieza, Josh. 2023. "AI Is a Lot of Work." *The Verge*, June 20, 2023. <https://www.theverge.com/features/23764584/ai-artificial-intelligence-data-notation-labor-scale-surge-remotasks-openai-chat-bots>.

Eisfeldt, Andrea L., Gregor Schubert, and Miao Ben Zhang. 2023. "Generative AI and Firm Values." SSRN Scholarly Paper. Rochester, NY. <https://doi.org/10.2139/ssrn.4436627>.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. 2023. "GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models." arXiv. <http://arxiv.org/abs/2303.10130>.

Ernst, Ekkehardt, Rossana Merola, and Daniel Samaan. 2019. "Economics of Artificial Intelligence: Implications for the Future of Work." *IZA Journal of Labor Policy* 9 (1). <https://doi.org/10.2478/iza-jolp-2019-0004>.

Felten, Edward W., Manav Raj, and Robert Seamans. 2019. "The Occupational Impact of Artificial Intelligence: Labor, Skills, and Polarization." SSRN Scholarly Paper. Rochester, NY. <https://doi.org/10.2139/ssrn.3368605>.

Frey, Carl Benedikt, and Michael A. Osborne. 2017. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114 (January): 254–80. <https://doi.org/10.1016/j.techfore.2016.08.019>.

Georgieff, Alexandre, and Raphaela Hye. 2021. "Artificial Intelligence and Employment : New Cross-Country Evidence | En | OECD." OECD Social, Employment and Migration Working Papers No. 265. <https://www.oecd.org/sti/artificial-intelligence-and-employment-c2c1d276-en.htm>.

Goldfarb, Avi, Bledi Taska, and Florenta Teodoridis. 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)." Global Economics Analyst. https://www.key-4biz.it/wp-content/uploads/2023/03/Global-Economics-Analyst_-The-Potentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf.

Hjort, Jonas, and Jonas Poulsen. 2019. "The Arrival of Fast Internet and Employment in Africa." *American Economic Review* 109 (3): 1032–79. <https://doi.org/10.1257/aer.20161385>.

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

———. 2023a. "ILO Modelled Estimates (ILOEST Database)." ILOSTAT. 2023. <https://ilostat.ilo.org/resources/concepts-and-definitions/ilo-modelled-estimates/>.

———. 2023b. "ISCO Documentation: Part III. Definitions of Major Groups, Sub-Major Groups, Minor Groups and Unit Groups." 2023. <https://www.ilo.org/public/english/bureau/stat/isco/docs/groupdefn08.pdf>.

———. 2023c. "World Employment and Social Outlook 2023: The Value of Essential Work." 2023. <https://www.ilo.org/digitalguides/en-gb/story/weso2023-key-workers>.

———. 2023d. "Decision Concerning the Agenda of the Future Sessions of the International Labour Conference." Record of GB decisions. March 22, 2023. http://www.ilo.org/gb/GBSessions/GB347/ins/WCMS_873027/lang-en/index.htm.

Irani, Lilly. 2015. "Difference and Dependence among Digital Workers: The Case of Amazon Mechanical Turk." *South Atlantic Quarterly* 114 (1): 225–34. <https://doi.org/10.1215/00382876-2831665>.

ITU. 2023. "Measuring Digital Development: Facts and Figures 2022." ITU. 2023. <https://www.itu.int:443/en/ITU-D/Statistics/Pages/facts/default.aspx>.

Karger, Ezra, Josh Rosenberg, Zachary Jacobs, Molly Hickman, Rose Hadshar, Kayla Gamin, Taylor Smith, Bridget Williams, Tegan McCaslin, and Philip E. Tetlock. 2023. "Forecasting Existential Risks: Evidence from a Long-Run Forecasting Tournament." FRI Working Paper #1. Forecasting Research Institute. <https://static1.squarespace.com/static/635693acf15a3e2a14a56a4a/t/64abffe3f024747d-d0e38d71/1688993798938/XPT.pdf>.

Kuek, Siou Chew, Cecilia Paradi-Guilford, Toks Fayomi, Saori Imaizumi, Panos Ipeirotis, Patricia Pina, and Manpreet Singh. 2015. "The Global Opportunity in Online Outsourcing." World Bank Publications - Reports 22284. The World Bank Group. <https://econpapers.repec.org/paper/wbk-wboper/22284.htm>.

Lane, Marguerita, and Anne Saint-Martin. 2021. "The Impact of Artificial Intelligence on the Labour Market: What Do We Know so Far?," January. <https://doi.org/10.1787/7c895724-en>.

Lee, Min Kyung, Daniel Kusbit, Evan Metsky, and Laura Dabbish. 2015. *Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers*. <https://doi.org/10.1145/2702123.2702548>.

Matopoulos, Aristides. 2011. "Warehouse Technologies in Retail Operations: The Case of Voice Picking." In *Intelligent Agrifood Chains and Networks*, 195–207. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781444339895.ch12>.

Mattos, Fernanda Bárçia de, Sukti Dasgupta, Xiao Jiang, David Kucera, and Ansel F. Schiavone. 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>.

———. 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#/.

Moore, Phoebe. 2023. "The AI Act." *Phoebe V Moore* (blog). May 5, 2023. <https://phoebevmoore.wordpress.com/2023/05/05/2393/>.

Patel, Dylan, and Afzal Ahmad. 2023. "Google 'We Have No Moat, And Neither Does OpenAI.'" May 4, 2023. <https://www.semianalysis.com/p/google-we-have-no-moat-and-neither>.

Soyres, François de, Mohamed Abdel Jelil, Caroline Cerruti, and Leah Kiwara. 2018. "What Kenya's Mobile Money Success Could Mean for the Arab World." World Bank. October 2018. <https://www.worldbank.org/en/news/feature/2018/10/03/what-kenya-s-mobile-money-success-could-mean-for-the-arab-world>.

Surawski, Bartosz. 2019. "Who Is a 'Knowledge Worker' – Clarifying the Meaning of the Term through Comparison with Synonymous and Associated Terms." *Management* 23 (June): 105–33. <https://doi.org/10.2478/manment-2019-0007>.

Tubaro, Paola, Antonio A. Casilli, and Marion Coville. 2020. "The Trainer, the Verifier, the Imitator: Three Ways in Which Human Platform Workers Support Artificial Intelligence." *Big Data & Society* 7 (1): 2053951720919776.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." In *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc. https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Webb, Michael. 2019. "The Impact of Artificial Intelligence on the Labor Market." SSRN Scholarly Paper. Rochester, NY. <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/>.

Wood, Alex. 2021. "Algorithmic Management: Consequences for Work Organisation and Working Conditions." JRC124874. EU JRC. https://joint-research-centre.ec.europa.eu/publications/algorithmic-management-consequences-work-organisation-and-working-conditions_en.

Xu, Jing, Da Ju, Joshua Lane, Mojtaba Komeili, Eric Michael Smith, Megan Ung, Morteza Behrooz, et al. 2023. "Improving Open Language Models by Learning from Organic Interactions." arXiv. <https://doi.org/10.48550/arXiv.2306.04707>.

Acknowledgements and use of GPT

We acknowledge support from several colleagues at the ILO. We are particularly grateful to Bálint Náfrádi and Sergei Soares for their comments on the content and their independent review of our calculations. We thank Lara Badre, who provided a clean mapping of different ISCO levels to detailed tasks as well as a mapping of over 7,500 job titles found in Labour Force Surveys (LFS) to ISCO 4-digit. Lara also gave us numerous useful suggestions regarding the ISCO system and conducted some spot checks of the predicted job definitions and tasks. We are grateful to Steve Kapsos, for his suggestions on the methodology for calculating the global employment estimates, to Daniel Samaan, who shared his expert knowledge on different types of occupational scores, to David Kucera for his helpful comments on the final draft, and to several colleagues in the ILO's Research Department for their constructive feedback during initial internal presentations of this work.

GPT-4 API was used to generate alternative occupational definitions, tasks and task-level scores, as discussed in the text. We placed some 25,000 API requests to GPT-4 and used Ada model to generate task-level embeddings for 7,482 tasks. We used GPT-4 API to summarize 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 US\$ 1,000, out of which some US\$ 600 have been 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 are full time staff of the ILO, with 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.

► Advancing social justice, promoting decent work

The International Labour Organization is the United Nations agency for the world of work. We bring together governments, employers and workers to improve the working lives of all people, driving a human-centred approach to the future of work through employment creation, rights at work, social protection and social dialogue.

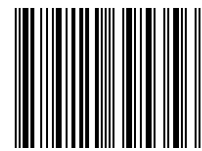
Contact details

Research Department (RESEARCH)

International Labour Organization
Route des Morillons 4
1211 Geneva 22
Switzerland
T +41 22 799 6530
research@ilo.org
www.ilo.org/research



ISBN 9789220395356



9 789220 395356