Assessing AI's Disruptive Potential: A Novel Measure of Occupational Exposure Using Press Release Data

Georgi Demirev
Sofia University

June 9, 2024

Abstract

Understanding which occupations are most at risk of automation by Artificial Intelligence (AI) is a question of significant policy and academic interest, in light of recent rapid advances in the field. Previous studies addressing this issue have relied on benchmark performance and patent data to determine the similarity between AI capabilities and the tasks performed by human labor. However, these sources might not accurately reflect real-world AI applications. This study leverages a novel dataset of over 440,000 corporate press releases to analyze AI product launches and adoptions. Using advanced natural language processing (NLP) techniques and large language models (LLMs), I identify and categorize tens of thousands of specific AI capabilities. These capabilities are then mapped and compared to job-specific skills from the European Standardized Classification of Occupations (ESCO) to calculate an "AI Product Exposure Score" for over 3,000 individual occupations. The findings reveal significant variations in AI exposure across different job categories, with ICT professionals, managers, and business administration professionals being the most exposed. Overall, these findings show that higher-skilled and higher-paid professions are more exposed to AI, contrasting with previous automation waves that disproportionately affected lower-skilled workers.

JEL Classification: O33, J24, J23, O31, J31, C55, C81
Keywords: AI, job automation, technological change, occupational exposure

1 Introduction

Recent advances in Artificial Intelligence (AI) have sparked significant debate about the economic impacts these technologies will have as they spread across various industries. A particularly contentious issue is the potential of AI to automate jobs and displace human labor. In this respect, AI is often compared to previous waves of automation, such as software tools, industrial robots, and even electricity and the steam engine.

A natural question to ask in this context is to what extent current AI capabilities overlap with and substitute for tasks performed by humans. This task-based framework, which views the production process as a series of steps, some of which can be automated, is a common approach in theoretical models of job automation (Acemoglu and Autor, 2011). In that light, identifying which occupations are at high risk of being affected by AI is a natural first step in modelling and predicting the broader impacts of AI as it develops further.

Several influential studies have sought to address this question. For instance, Frey and Osborne (2017) combines expert judgments with a machine learning classifier to produce AI automation scores for tasks and occupations within the O*NET occupational dataset. Building on this, Felten et al. (2018) and Brynjolfsson et al. (2023) utilize objective measures of AI advancements to assess the automation susceptibility of various tasks. More recently, Webb (2022) applies text analytics to a large corpus of patent data to identify AI capabilities, while Eloundou et al. (2023) employs a state-of-the-art large language model to classify occupations by their exposure to AI.

In this paper, I propose a novel approach to estimate occupation-level exposure to artificial intelligence using data from company press releases. The key insight is that while academic papers, AI benchmarks, and patents might tell us what is feasible in principle, data on actual product releases and adoptions offers a more accurate picture of how AI is currently being implemented in the economy.

To achieve this, I collect a novel dataset of free-text corporate press releases from PRNewswire, a major press release aggregator, covering the period of August 2023 to April 2024 (prn, 2024). I then apply a combination of Natural Language Processing (NLP) techniques, such as word embeddings and vector similarity, alongside direct analysis by Large Language Models (LLM). This approach helps identify press releases that mention new AI product releases or adoptions and extract the specific capabilities of each product. These capabilities are then matched to a list of job-specific skills from the classification of European Skills, Competences, Qualifications and Occupations (ESCO, European Commission, Directorate-General for Employment, Social Affairs and Inclusion (2019)). By considering the relevance of each skill to each occupation in the ESCO taxonomy, I derive an "AI Product Exposure Score" for over 3,000 individual occupations.

The contributions of this paper are threefold: first, it introduces a novel dataset that has not been previously used in this field. Second, it employs an innovative combination of NLP and LLM techniques to analyze this data, which could be applied to other datasets as well. Third, it presents findings at the level of ESCO occupations, making them compatible with the classification system used by the International Labor Organization, unlike prior research that relies on the US-specific O*NET classification.

The rest of the paper is organized as follows. Section two reviews the related literature. Section three details the methodology of the analysis. Section four presents the main results, including the occupation exposure score and supplementary findings such as correlations with occupation-level statistics and comparisons with prior studies. Section five offers a discussion and concludes the paper.

2 Literature Review

Theoretical models of automation and labor displacement have long conceptualized economic production as a combination of outputs from individual tasks or jobs. Since Zeira (1998), it has been established that some tasks can be performed exclusively by labor, while others can also involve capital. In this framework, automation results from technological advances that reduce the number of labor-only tasks.

More recent studies focusing specifically on AI build upon these foundational ideas. Notably, Acemoglu and Restrepo (2019) develop a comprehensive framework to examine the impacts of automation technologies such as industrial robots and AI. Their findings, echoed in Acemoglu and Johnson (2023), highlight the dual effects of technology on the labor market. On one hand, automation reduces labor demand and labor's share of income by substituting labor tasks. On the other hand, it can enhance productivity and output, potentially leading to the creation of new tasks and increased labor demand. This balance between displacement and complementarity underscores the necessity of examining AI's specific impacts at the task and occupation levels.

In another influential theoretical work using a task-based approach, Aghion et al. (2017) investigate whether inherent complementarities in the production function might limit AI's impact on the economy due to Baumol's cost disease (Baumol, 1967). They also explore the potential for AI to automate research labor in the ideas production function. Under certain assumptions, this automation of research can lead to significant increases in productivity and output growth, warranting a closer examination of AI's applicability to research-related tasks.

Several empirical studies have constructed occupational indices of AI exposure. Brynjolfsson et al. (2018) is considered pioneering in this area, following the initial Brynjolfsson and Mitchell (2017). These studies introduce the approach of evaluating occupations based on their constituent tasks and assessing each task's susceptibility to AI—a methodology that this paper also employs.

Frey and Osborne (2017) advances this field by developing a machine learning classifier to assess the AI exposure of individual occupations. This classifier uses a hand-labeled dataset of 70 occupations rated by machine learning experts and employs occupational characteristics from the O*NET database (National Center for O*NET Development, 2024) as predictor variables.

To introduce an objective measure of AI's technical progress, Felten et al. (2018) utilizes AI progress metrics compiled by the Electronic Frontier Foundation (Eckersley and Nasser, 2017-). They link these metrics to O*NET occupational abilities using hand-crafted weights. Felten's data is publicly available and is used in this paper to validate its results.

Webb (2022) uses text analytics to analyze patent titles and descriptions, extracting AI capabilities and matching them to occupational abilities. This study conceptually aligns with Webb's approach but uses product press releases instead of patent filings. While Webb relies on verb-noun pairs and hierarchical concept substitution, this paper utilizes vector similarity for a more nuanced measurement.

Eloundou et al. (2023) shares some methodological similarities with this study. The authors outsource the task of rating AI exposure to a Large Language Model (LLM). By crafting specific prompts for GPT-4 (OpenAI, 2024), they derive consistent scores across occupations. This paper also uses GPT-4, but only as a final step in text processing, as outlined in the methodology section.

3 Methodology

3.1 Data

This study utilizes three main data sources: the PRNewswire repository of press releases, used to extract AI product capabilities (prn, 2024); the European Skills, Competences, Qualifications and Occupations (ESCO) job taxonomy data, used to map these capabilities to occupational skills and capabilities (European Commission, Directorate-General for Employment, Social Affairs and Inclusion, 2019); and the Skills Intelligence Report from the European Centre for the Development of Vocational Training (CEDEFOP), used to examine job market and demographic statistics across occupations (European Centre for the Development of Vocational Training, 2024).

PRNewswire is a prominent online aggregator of corporate press releases, distributing and publishing them on its publicly accessible website. This website contains data for the last six months spread across 14 categories. For the purposes of this study, I collected the full text of all available press releases from August 8, 2023, to April 15, 2024, totalling 444,989 individual press releases.

The ESCO dataset contains a detailed multi-level hierarchical taxonomy of occupations. While the first four levels of classification align with the International Standard Classification of Occupations (ISCO) by the International Labour Organization (ILO), ESCO further subdivides the categories, resulting in five levels of occupational detail. For example, the occupation "bank teller" (ESCO code 4211.1) falls under "Tellers and money collectors" (ESCO code 421), within "Customer

Service Clerks" (ESCO code 42), which is part of "Clerical Support Workers" (ESCO code 4). The version of the ESCO database used (v1.1.2) includes 3,006 occupations and 10,821 job-relevant skills.

The Skills Intelligence Report is an online publication compiled by CEDEFOP based on several primary data sources, such as the European Labor Force Survey (EU-LFS) and the European Survey on Income and Living Conditions (EU-SILC). It contains occupational-level data for all EU countries aggregated at the 2-digit ESCO level. Key data used from this report includes the number of workers per occupation per country, unemployment rates, and the percentage of women employed in each occupation.

Additionally, data from the European Commission's Wage Determinants report (European Commission. Statistical Office of the European Union., 2021) is used to assign a wage premium to each occupation. This report contains an econometric model analyzing wage determinants, using 2-level ESCO code dummies. The coefficients on these dummies are interpreted as relative wage premiums by occupation.

3.2 Text Analysis

The text analysis of the press releases involves two main steps: 1) identifying press releases about new AI product launches or the adoption of AI-powered products, and 2) identifying the capabilities of these AI systems.

Firstly, since PRNewswire is a general-purpose publication, I apply a keyword filter to limit the number of documents for analysis. After standard preprocessing (lowercasing, removing stop words and punctuation, stemming), I check for the presence of phrases related to artificial intelligence (e.g., "AI," "machine learning," "gpt," "openai"). This filtering process identifies 69,661 potentially relevant articles. I denote this document corpus as P.

Next, I use a semantic similarity comparison to determine which articles pertain to AI product launches and introductions. Using an open-source NLP machine learning model (Honnibal et al., 2020), I perform text embedding on all product releases and benchmark phrases [1].

Formally, if $E()$ is the embedding function that acts on an individual word w to produce a K-dimensional embedding vector, the document-level embedding of each document $r \in P$ is denoted as $\bar{E}(r)$, whose the k-th element is given by:

$$\bar{E(r)}_k = \frac{\sum_{w \in r} E(w)_k}{\sum_{w \in r} 1} \tag{1}$$

i.e. an element-wise average of all individual word embeddings in the document.

I define several phrases indicative of AI product launches or adoptions[2] and derive their embeddings in an analogous fashion. I then compute the cosine similarity between the press release embeddings and benchmark phrase embeddings. Using the $\bar{E}()$ function defined above, the cosine similarity between any two documents (in this case a press release and a key phrase) is defined as:

$$S(i,j) = \frac{\bar{E}(i) \cdot \bar{E}(j)}{\|\bar{E}(i)\| \, \|\bar{E}(j)\|} \tag{2}$$

Having computed the similarity measure between all press releases in P and each key phrase, I filter out press releases with similarity scores below 0.7 across all key phrases, resulting in 8,507 relevant press releases. I denote this document corpus as R.

[1]The particular model used is version 3.7.1 of the model en_core_web_lg of the open-source package SpaCy, which produces 300-dimensional embeddings.

[2]See Appendix

The next phase of the analysis involves identifying the specific capabilities of the AI tools mentioned in the press releases. I apply a similar semantic similarity exercise at the sentence level to identify relevant text segments. I.e. for each document $r \in R$ I retain only those sentences s for which $S(s, j) \geq 0.8$ for at least one key phrase j. This is needed to reduce the overall text size and hence the processing costs for the next step of the analysis.

Finally, I pass the condensed version of each press release (containing only the sentences scored as relevant) to a large language model (GPT-4). The model is prompted to read the text and output short, comma-separated strings detailing the AI system's capabilities (e.g., "reads and analyzes documents, issues treatment recommendations")[3]. This format mirrors the occupation capability strings in the ESCO database. The model outputs an empty string if the text is not relevant or if no specific capabilities are identified.

Formally, if $U(r)$ is the set of AI capability phrases, identified by the Large Language Model based on input document r, then the set of AI capability phrases is defined as:

$$C_{ai} = \bigcup_{r \in R} U(r) \tag{3}$$

This process results in 28,908 AI capability strings coming from 5,300 press releases.

3.3 Deriving Occupational Exposure Scores

The identified AI capabilities are compared to the 10,821 individual occupational skills in the ESCO database using cosine similarity between the mean word embeddings of each string. This generates nearly 313 million pairwise semantic similarity scores.

For each individual skill, I take the highest score from these comparisons, and assign it as that skill's exposure to AI product innovations. Formally, if $i \in C_{jobs}$ is an individual jobs skill string, and $j \in C_{ai}$ is an individual AI capability string, and $S(i, j)$ is the similarity between i and j as defined above, the skill-level similarity score $S^*(i)$ is given by:

$$S^*(i) = \max_{j \in C_{jobs}} S(i, j) \tag{4}$$

Finally, I aggregate these similarity scores up to the 5-digit occupation level. I do this by taking the list of skills associated with each occupation in the ESCO database and computing a weighted average of the individual skills' exposure scores. I weigh skills labeled as "essential" twice as heavily as skills labeled as "optional" [4]. Formally:

$$ES(o) = \frac{\sum_{i \in JS(o)} w_{o,i} S^*(i)}{\sum_{i \in JS(o)} w_{o,i}} \tag{5}$$

Where $ES(o)$ denotes the (non-standardized) occupational exposure score of occupation o, $JS(o)$ is the set of job skills relevant to occupation o, and $w_{o,i}$ is the importance weight of the skill for the occupation (either 1 for non-essential or 2 for essential skills).

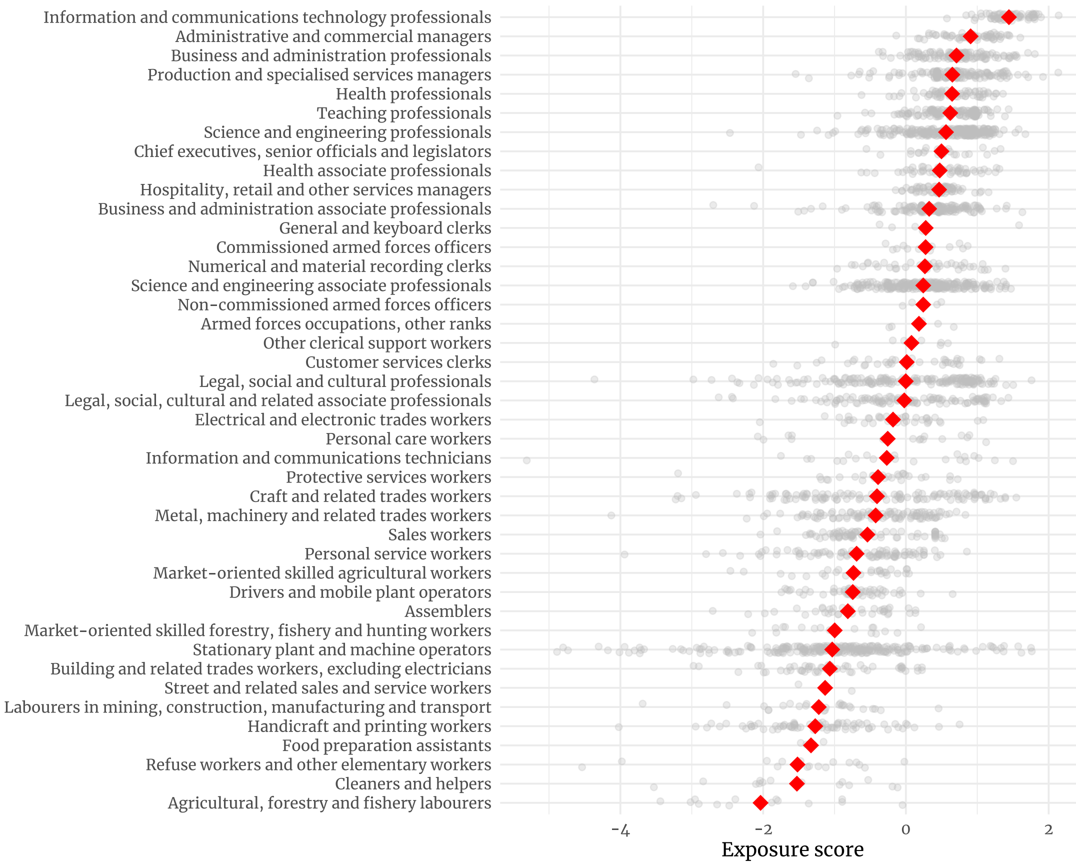
Since this is essentially a dimensionless index, I standardize the result by subtracting the mean and dividing by the standard deviation. The resulting quantity $ES^*(o)$ is the **occupation-level exposure to AI product innovations score**, the principal output of this study.

When presenting the exposure scores for broader ESCO categories, I use a simple average. For example, the scores at the 3-digit occupational level are the average of the corresponding 4-digit

[3] The prompt can be found in the Appendix

[4] This weighting is admittedly arbitrary, but the results do not meaningfully change if I employ an unweighted average instead

Figure 1: Occupational Exposure to AI Product Innovation



occupations. I also apply a crosswalk file provided by ESCO to convert the scores to the United States Standard Occupational Classification for comparisons with prior work. All final scores are available through the open-source repository associated with this paper.

4 Results

4.1 Occupational Exposure Scores

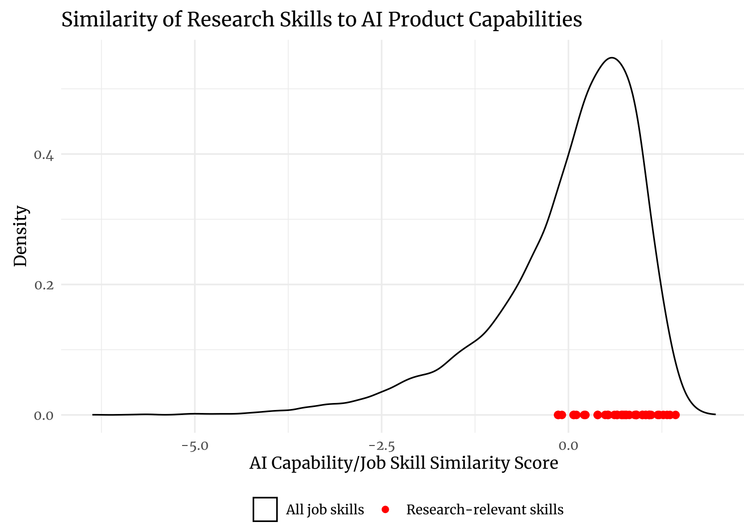
The main result of this study, the occupational exposure scores, is summarized in Figure 1. Each labeled row corresponds to an occupational group at the 2-digit level of the ISCO hierarchy. The gray dots represent individual (5-digit) occupations within each group, while the red rhomboids represent group means.

ICT professionals are the most exposed group with an exposure score of 1.44, followed by Administrative and Commercial Managers (0.907) and Business and Administration Professionals (0.709). Also notable near the top of the table are Health and Teaching Workers, Science and Engineering Professionals, and Senior Executives.

At the opposite end, Agricultural, Forestry, and Fishery Laborers are the least exposed group, followed by Cleaners and Helpers, Refuse Workers, and Food Preparation Assistants.

At the individual occupation level, the most exposed jobs are "Data Analyst," "Chief Data Analyst," "ICT Information and Knowledge Manager," "ICT Business Analyst," and "Software

Figure 2: AI Capabilities and Research Skills



Analyst," suggesting that the engineers developing AI systems are first focusing on improving the efficiency of their own work. The full list of 3,007 scored occupations is available in the online repository of this paper.

A specific group of occupations and skills that merits a closer look is research-related professions. As mentioned above, AI advances in research and knowledge production have the potential to bring long-lasting productivity and welfare gains (Aghion et al., 2017). ESCO provides a list of skills especially relevant for research work. According to the measure developed in this paper, those skills have a high susceptibility to AI disruption, with a mean score of 0.93 compared to 0.0 (by construction) for all other skills. Similarly, research-related professions score highly on the occupation exposure level, with Science and Engineering Professionals having a mean score of 0.562.

4.2 Comparison to Prior Work

To validate the results and compare them to prior work, I downloaded and joined the final result datasets of Felten et al. (2018) and Webb (2022). Since their work uses the O*NET occupational classification, I used a crosswalk provided by ESCO to partially match the two sets of occupations.

The results of this paper exhibit a strong correlation with Felten's occupational exposure ($r = 0.702$), with a Pearson product-moment correlation test yielding $t = 10.531$ on 114 degrees of freedom ($p < 2.2e - 16$).

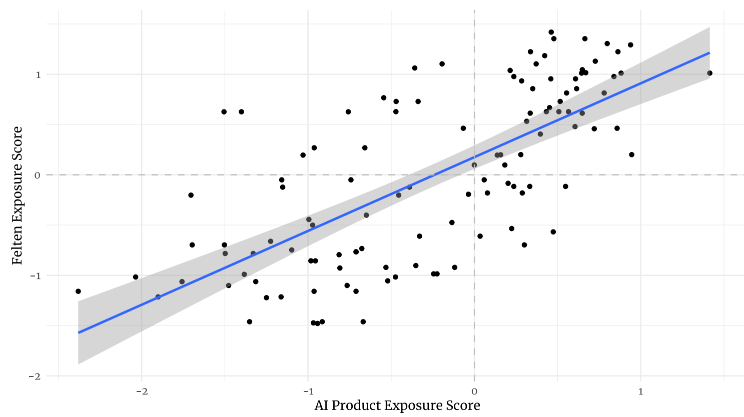
Conversely, the correlation coefficient with Webb's AI score is small and insignificant ($r = 0.06$, $t = 0.637$, $p = 0.52$). Notably, Felten's and Webb's measures also do not correlate with each other.

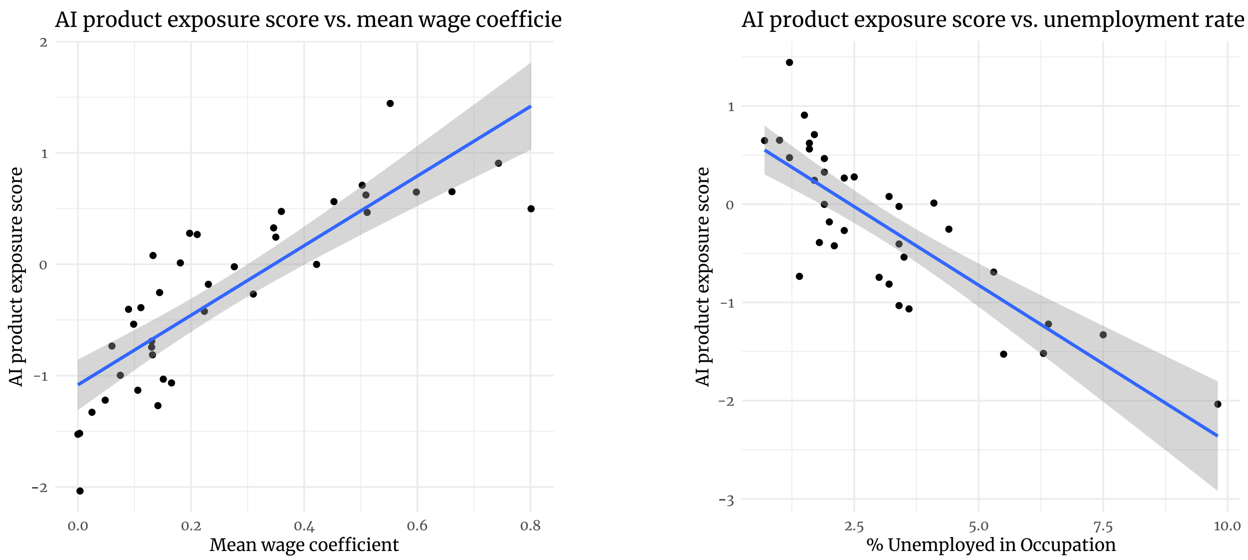
4.3 Relationship to Job Market Outcomes

The list of occupations presented above suggests that higher-skilled and higher-paid professions are more exposed to Artificial Intelligence. To confirm this, I compare the exposure score to data on average unemployment and occupational wage premium (available at the ISCO 2-digit level). The results are presented in Figure 4.

As evident from the plot, the AI product exposure score is strongly negatively correlated with the unemployment rate within the occupation ($r = -0.82$, $t = -8.158$, $df = 33$, $p = 2.04e - 9$),

Figure 3: Comparison with Felten et al. (2018)





(a) AI Exposure vs Wage Premiums

(b) AI Exposure vs Unemployment

Figure 4: AI Exposure and Job Market Outcomes

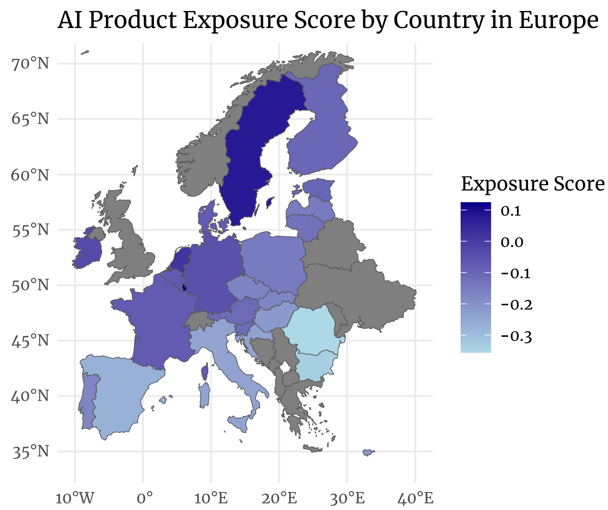
and strongly positively correlated with relative wages ($r = 0.84$, $t = 9.38$, $df = 37$, $p = 2.53e - 11$) [5]. This is perhaps the most intriguing finding of this paper, as previous waves of automation have typically disproportionately affected lower-skilled and lower-paid workers (Acemoglu and Johnson, 2023).

4.4 Geographical Exposure

Automation-inducing technologies tend to have uneven impacts across geographical regions (Acemoglu and Restrepo). To identify potential regional imbalances, I calculate a weighted national exposure score using data from CEDEFOP on the number of people employed in each ISCO 2-digit occupation across EU countries. The results are presented in Figure 5.

[5] I also computed the correlation between the AI exposure score and total employment and the percentage of women in each occupation. Both these results were not significant and are not presented here but rather relegated to the online repository

Figure 5: Occupational AI Exposure by Country



The countries with an occupational mix most susceptible to AI disruption, as per this measure, are Luxembourg (0.12), Sweden (0.08), and the Netherlands (0.02). In contrast, Romania (−0.36), Bulgaria (−0.33), Spain (−0.27), and Italy (−0.24) have occupational structures that are relatively less exposed to AI [6].

It is worth noting that this is a preliminary look at the exposure of regional labor markets to Artificial Intelligence, considering only the occupational mix and disregarding the demographic structure and overall economic conditions of each region.

5 Conclusion

The results presented in this paper suggest that AI's impact on the labor market may differ significantly from previous technological revolutions, as it is more likely to affect well-paid and prestigious occupations. While it remains to be seen to what extent jobs will be automated and whether new task creation and productivity gains will outpace displacement effects, the occupational exposure score developed in this study provides a valuable tool for future explorations.

As an example of the type of study that can utilize the exposure measure, Acemoglu et al. (2022) use the metrics developed by Webb and Felten as treatment variables in a macroeconomic model of job automation, tested against data about online job vacancies. They find significant hiring differences in firms with more exposed occupational structures, showing an increase in hiring of professionals likely to implement AI systems and a decrease in hiring for other occupations. However, they do not detect aggregate employment effects. Using the measure proposed here, which is based on actual product deployments, could provide additional insights and precision.

Future research could also explore the gap between what is theoretically possible to automate (as measured by Felten's method) and what is actually implemented (as measured in this study). The strong correlation with wage premiums and unemployment rates indicates that entrepreneurs

[6] A more informative analysis would also look at sub-national data; however, I was unable to find regional data on employment by ISCO occupation

are most interested in innovating where labor is relatively scarce and expensive. Developing a more robust model of directed technical change and entrepreneurial activity, perhaps within the framework of Aghion and Howitt (1992), could enhance our understanding of this behavior.

Beyond academic research, these findings can inform policymakers as societies adjust to the potential disruptions brought by AI. The occupational exposure scores highlight potential areas of skill mismatch and can help identify targets for retraining programs or other policy interventions. When combined with national and sub-national data on occupational composition, the exposure scores can aid in regional planning and development, as previous impacts of automation technologies have often been concentrated in specific regions or metropolitan areas (Acemoglu and Johnson, 2023).

Furthermore, as highlighted in the Results section, the current wave of AI automation is particularly effective at automating and accelerating parts of research and development. This is significant because research productivity is crucial for long-term growth in endogenous and semi-endogenous growth models, such as those by Romer (1986) and Jones (2021). Investigating how these productivity-enhancing tools affect research output could help us better understand and anticipate AI's impact on the knowledge production function and long-term productivity growth.

References

PRNewswire: press release distribution, targeting, monitoring and marketing — prnewswire.com. `https://www.prnewswire.com/`, 2024. [Accessed 09-06-2024].

D. Acemoglu and S. Johnson. *Power and Progress: Our Thousand-year Struggle Over Technology and Prosperity.* PublicAffairs, 2023. ISBN 9781541702530. URL `https://books.google.bg/books?id=ttxfzwEACAAJ`.

Daron Acemoglu and David Autor. Chapter 12 - skills, tasks and technologies: Implications for employment and earnings. volume 4 of *Handbook of Labor Economics*, pages 1043–1171. Elsevier, 2011. doi: https://doi.org/10.1016/S0169-7218(11)02410-5. URL `https://www.sciencedirect.com/science/article/pii/S0169721811024105`.

Daron Acemoglu and Pascual Restrepo. Robots and jobs: Evidence from us labor markets. *Journal of Political Economy.*

Daron Acemoglu and Pascual Restrepo. Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30, May 2019. ISSN 0895-3309. doi: 10.1257/jep.33.2.3. URL `https://pubs.aeaweb.org/doi/10.1257/jep.33.2.3`.

Daron Acemoglu, David Autor, Jonathon Hazell, and Pascual Restrepo. Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, 40 (S1):S293–S340, April 2022. ISSN 0734-306X, 1537-5307. doi: 10.1086/718327. URL `https://www.journals.uchicago.edu/doi/10.1086/718327`.

Philippe Aghion and Peter Howitt. A model of growth through creative destruction. *Econometrica*, 60(2):323, March 1992. ISSN 00129682. doi: 10.2307/2951599. URL `https://www.jstor.org/stable/2951599?origin=crossref`.

Philippe Aghion, Benjamin F Jones, and Charles I Jones. Artificial intelligence and economic growth. 2017.

William J. Baumol. Macroeconomics of unbalanced growth: The anatomy of urban crisis. *The American Economic Review*, 57(3):415–426, 1967. URL `http://www.jstor.org/stable/1812111`.

Erik Brynjolfsson and Tom Mitchell. What can machine learning do? workforce implications. *Science*, 358(6370):1530–1534, December 2017. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aap8062. URL `https://www.science.org/doi/10.1126/science.aap8062`.

Erik Brynjolfsson, Tom Mitchell, and Daniel Rock. What can machines learn and what does it mean for occupations and the economy? *AEA Papers and Proceedings*, 108: 43–47, May 2018. ISSN 2574-0768, 2574-0776. doi: 10.1257/pandp.20181019. URL `https://pubs.aeaweb.org/doi/10.1257/pandp.20181019`.

Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond. Generative ai at work, 2023.

Peter Eckersley and Yomna et al. Nasser. Eff ai progress measurement project. `https://eff.org/ai/metrics`, 2017-. Accessed on 2017-09-09.

Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. Gpts are gpts: An early look at the labor market impact potential of large language models, August 2023. URL http://arxiv.org/abs/2303.10130. arXiv:2303.10130 [cs, econ, q-fin].

European Centre for the Development of Vocational Training. Skills intelligence tool. Web Page, 2024. URL https://www.cedefop.europa.eu/en/tools/skills-intelligence. Accessed: 2024-06-09.

European Commission, Directorate-General for Employment, Social Affairs and Inclusion. European skills, competences, qualifications and occupations (esco) handbook. European Commission, 2019. URL https://esco.ec.europa.eu/en. Accessed: 2024-06-09.

European Commission. Statistical Office of the European Union. *Wages determinants in the European Union: evidence from structure of earnings survey (SES 2018) data : 2021 edition.* Publications Office, LU, 2021. URL https://data.europa.eu/doi/10.2785/62601.

Edward W. Felten, Manav Raj, and Robert Seamans. A method to link advances in artificial intelligence to occupational abilities. *AEA Papers and Proceedings*, 108:54–57, May 2018. ISSN 2574-0768, 2574-0776. doi: 10.1257/pandp.20181021. URL https://pubs.aeaweb.org/doi/10.1257/pandp.20181021.

Carl Benedikt Frey and Michael A. Osborne. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114: 254–280, January 2017. ISSN 00401625. doi: 10.1016/j.techfore.2016.08.019. URL https://linkinghub.elsevier.com/retrieve/pii/S0040162516302244.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. spaCy: Industrial-strength Natural Language Processing in Python. 2020. doi: 10.5281/zenodo.1212303.

Charles I Jones. The past and future of economic growth: A semi-endogenous perspective. 2021.

National Center for O*NET Development. O*net online, 2024. URL https://www.onetonline.org/. Retrieved April 24, 2024.

OpenAI. Gpt-4 technical report, 2024.

Paul M. Romer. Increasing returns and long-run growth. *Journal of Political Economy*, 94 (5):1002–1037, October 1986. ISSN 0022-3808, 1537-534X. doi: 10.1086/261420. URL https://www.journals.uchicago.edu/doi/10.1086/261420.

Michael Webb. The impact of artificial intelligence on the labor market. 2022.

Joseph Zeira. Workers, machines, and economic growth. *The Quarterly Journal of Economics*, 113(4):1091–1117, 1998. URL https://EconPapers.repec.org/RePEc:oup:qjecon:v:113:y:1998:i:4:p:1091-1117.