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

Georgi Demirev

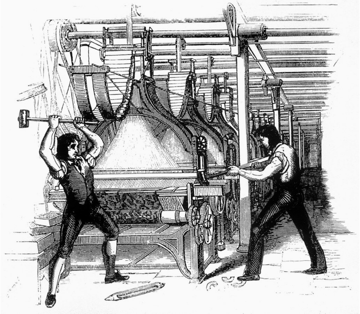
Sofia University "St Kliment Ohridski"

gdemirev@uni-sofia.bg

May 28, 2024

Overview

GPTs



Past eras of rapid technological innovation have been associated with waves of labor displacement and job automation (Acemoglu and Johnson, 2023). As modern incarnations of artificial intelligence (AI) are often lauded as being a general purpose technology (GPT) (Eloundou et al., 2023), a natural question that arises is to what extent will jobs be affected and which occupations are most exposed to disruption.

History Repeating?

Pinpointing Occupational Exposure to AI

- Brynjolfsson and Mitchell (2017):
  - Pioneered the task-based approach to evaluate AI susceptibility
  - Examined occupations through their constituent tasks
- Frey and Osborne (2017):
  - Combined expert judgments with a machine learning classifier
  - Derived AI automation scores for tasks and occupations using O*NET data
- Felten et al. (2018):
  - Used objective AI progress metrics from the Electronic Frontier Foundation
  - Linked these metrics to O*NET occupational abilities with hand-crafted weights
- Webb (2022):
  - Applied text analytics to patent data to elicit AI capabilities
  - Matched AI capabilities to occupational abilities using verb-noun pairs
- Eloundou et al. (2023):
  - Directly asked GPT-4 to classify occupations based on AI exposure

Main Idea

Use data on real world applications of AI by examining announcements of AI product launches and adoptions, instead of relying patent data or academic achievements.

Contributions

- Introduce a novel data set of corporate press releases
- Utilize a new methodology for linking AI capabilities and job skills based on advanced NLP and modern LLMs
- Derive Occupational exposure scores based on the internationally adopted ISCO/ESCO classification, instead of the US-centric O*NET

Main Data Source - PRNewswire

Resources Blog Journalists | Log In Sign Up Data Privacy Send a Release

PR Newswire | News | Products | Contact | Search

News in Focus | Business & Money | Science & Tech | Lifestyle & Health | Policy & Public Interest | People & Culture

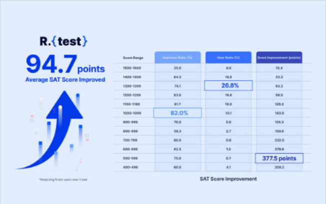
R.test, AI-Powered SAT Learning platform, Boasts Score Improvements of 94.7 points on Average

NEWS PROVIDED BY
Riiid →
20 Feb, 2024, 09:00 ET

SHARE THIS ARTICLE

SAN RAMON, Calif., Feb. 20, 2024 /PRNewswire/ -- Riiid, a leading AI education technology company, has revealed impressive score enhancements achieved by users of its AI-driven SAT learning platform, R.test. Through comprehensive data analysis from over 70,000 premium users over the past year, R.test has demonstrated its effectiveness in elevating SAT performance.

Launched in January 2023, R.test stands as the world's foremost AI-based digital SAT diagnostic service. With the complete transition to a digital SAT format in 2023 for various countries, followed by the United States in 2024, the platform has garnered significant global attention from students and educators alike, seeking to adapt to the new digitized test format. The imminent first digital SAT administration in the U.S. on March 9th further underscores the importance of effective preparation tools like R.test.



In the preceding year, R.test amassed a user base of

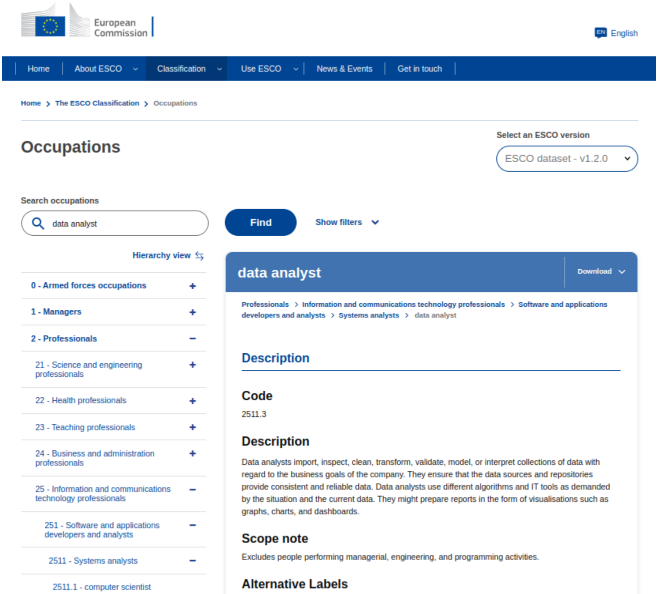
Data Collection

- Scrapped all press releases from the site for the period August 8-th 2023 - April 15-th 2024
- A total of 444989 individual press releases.
- Sanitize the text (lowercasing, stop-word removal, stemming) and filter AI-related press releases using a list of 30 relevant key phrases.
- 69661 press releases related to AI.

Identify AI Capabilities

- Calculate mean text embeddings of each press release using SpaCy (Honnibal et al., 2020) and calculate cosine similarity with the embeddings for 15 key phrases related to product launches or adoptions.
- 8507 press releases classified as relating to product launches or adoptions
- Pass the text of the releases to GPT-4 (OpenAI, 2024) to extract concrete AI capability strings
- 28908 AI capabilities associated with new product launches or adoptions

Job Skills Data



Calculating capability-skill similarity

If $E(i)$ is the set of vector embeddings of any job skill i, and $E(j)$ is the set of embeddings for any AI capability j, then the similarity between skill i and capability j is:

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

where:

$$\bar{E}(i) = \frac{1}{n_i} \sum_{k=1}^{n_i} E(i)_k$$

$$\bar{E}(j) = \frac{1}{n_j} \sum_{l=1}^{n_j} E(j)_l$$

and $E()$ denotes the (multi-valued) embedding function (in this case, the SpaCy model "en_core_web_lg").

Job Skill Exposure Scores

For any individual job skill i, the Skill AI Product Exposure Score is calculated as the highest cosine similarity between that skill and all AI capabilities.

$$S^*(i) = \max_j S(i,j)$$

In total, there are 10821 job-relevant skills in the ESCO database, meaning a little over 313 million pairwise similarity scores were calculated, which were then aggregated back to the skill level to derive one final score for each skill.

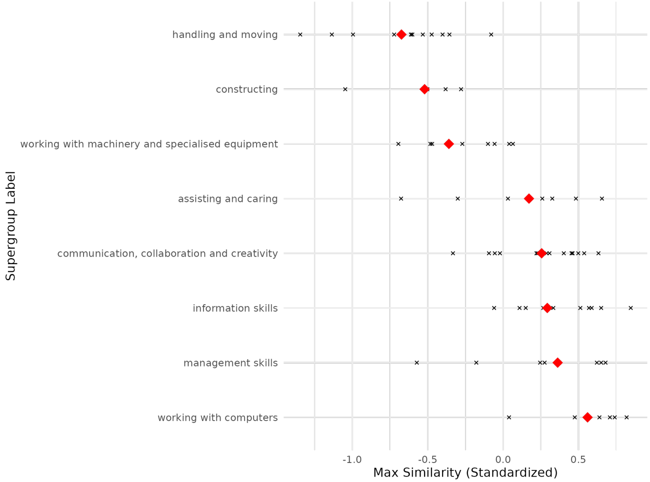
Calculating Job Exposure

To calculate the AI Product Exposure Score for each of the 3008 occupations in the ESCO database, I take a weighted average of the skill-exposure scores of all skills listed as relevant for that occupation, weighting "optional" skills half as much as "essential" skills.

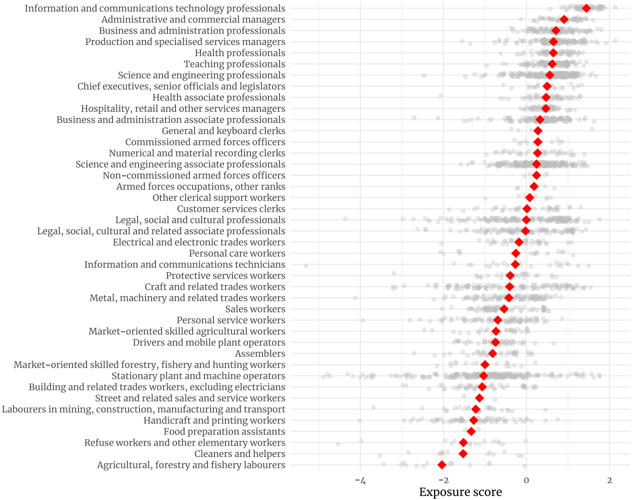
Formally, if $JS(i)$ are all skills relevant to job i, and $w_{i,j}$ is 2 if skill j is essential for job i, 1 otherwise, the exposure score of occupation i denoted as $E(i)$ is

$$E(i) = \frac{\sum_{j \in JS(i)} w_{i,j} S^*(j)}{\sum_{j \in JS(i)} w_{i,j}}$$

Skill Exposure Scores



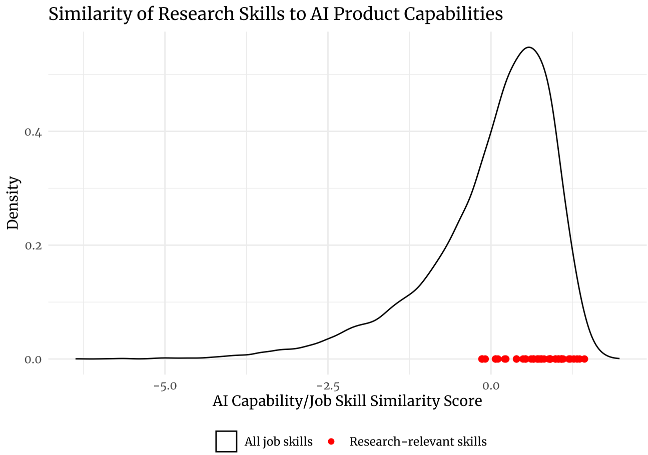
Occupational Exposure Scores



Most Exposed Occupations

- data analyst
- chief data analyst
- ICT information and knowledge manager
- ICT business analyst
- software analyst

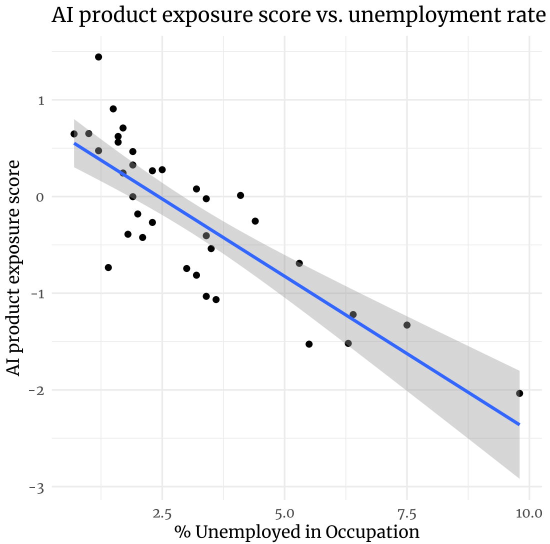
Research Skill Exposure



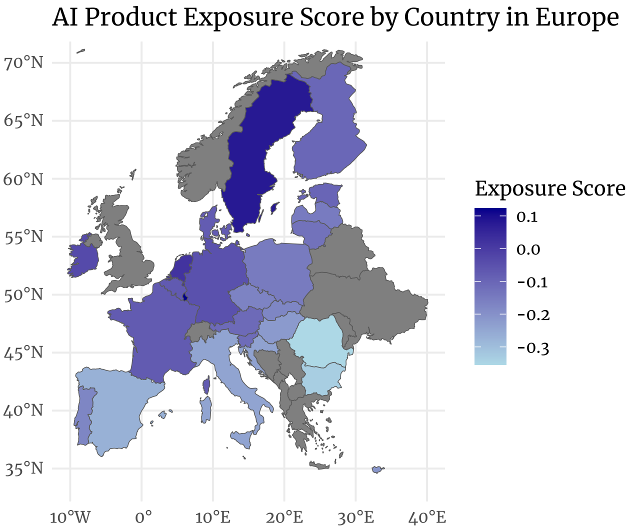
Wages


AI product exposure score vs. mean wage coefficie
AI product exposure score
Mean wage coefficient
1
0
−1
−2
0.0
0.2
0.4
0.6
0.8

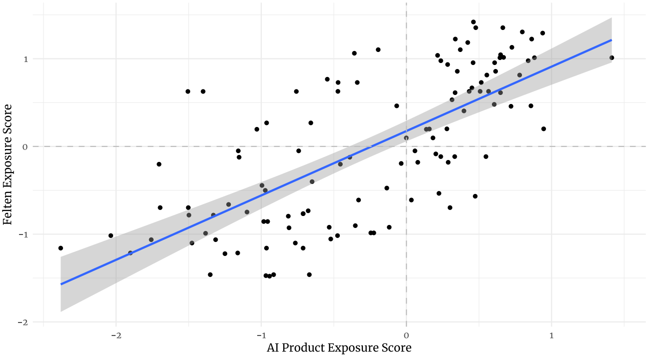
Unemployment



Geographical Exposure



Comparison with Felten et al.



The Theory of Job Automation

According to widely used models of job automation (e.g. Acemoglu and Restrepo (2019)), technology can impact production in several ways:

- Increase the per-unit productivity of either labor or capital
- Create new tasks for labor in the production function
- Substitute capital for labor

Only the last effect amounts to true automation and job displacement. A highly exposed profession might be at risk of being automated, but a high exposure score may just as well indicate the potential for productivity gains or task creation. Therefore, exposure scores do not tell us which jobs will get automated by themselves.

Directions for Future Research

- Use the exposure score as an explanatory variable when analyzing aggregate per-occupation employment or vacancy data (similar to Acemoglu et al. (2022)).
- Understand firm-level decisions for product launches or adoptions and their interplay with hiring and labour decision.
- Try to measure productivity impact of AI-powered technologies .on the firm or worker level.
- Examine the relationship between AI tools and research productivity.

Thank You!

Contact Information:
Email: gdemirev@uni-sofia.bg
Website: https://github.com/demirev/ai-products

References

Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, 40(S1):S293–S340.

Acemoglu, D. and Johnson, S. (2023). *Power and Progress: Our Thousand-year Struggle Over Technology and Prosperity*. PublicAffairs.

Acemoglu, D. and Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30.

Brynjolfsson, E. and Mitchell, T. (2017). What can machine learning do? workforce implications. *Science*, 358(6370):1530–1534.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2023). Gpts are gpts: An early look at the labor market impact potential of large language models. arXiv:2303.10130 [cs, econ, q-fin].

Felten, E. W., Raj, M., and Seamans, R. (2018). A method to link advances in artificial intelligence to occupational abilities. *AEA Papers and Proceedings*, 108:54–57.

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

Honnibal, M., Montani, I., Van Landeghem, S., and Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python.

OpenAI (2024). Gpt-4 technical report.

Webb, M. (2022). The impact of artificial intelligence on the labor market.