

README for the Reproducibility Package for Working Paper

“Labor Demand in the Age of Generative AI: Early Evidence from the U.S. Job Posting Data”¹

Yan Liu², He Wang³ and Shu Yu⁴

October 3, 2025

Table of Content

Reproduction Instructions.....	1
Data Statement.....	3
List of Exhibits.....	3

Reproduction Instructions

Software and platform parameters

Stata: Stata 18.0 MP—Parallel Edition

Databricks:

Runtime: 16.4 | Spark 3.5.2 | Scala 2.12

Driver: Standard_D16ds_v4 | 64GB | 16 Cores

Worker: Standard_D16ds_v4 | 64GB | 16Cores

Code

Two sets of code have been used for this project.

- The first consists of a Stata do-file run on a PC to process occupation-level datasets.
- The second consists of Python and R notebooks used on the Databricks cloud platform, where the proprietary LightCast job posting data are hosted, to process the data, conduct analysis, and generate the main figures and tables.

Execute the code

¹ We would like to thank Aart Kraay, Daniel Lederman, Franziska Lieselotte Ohnsorge, Jonah Matthew Rexer, and participants at various seminars for helpful comments and suggestions. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

² Development Effectiveness, Digital VPU, The World Bank. yanliu@worldbank.org. Corresponding author.

³ Development Effectiveness, Digital VPU, The World Bank. hwang21@worldbank.org.

⁴ Office of the Chief Economist for the Prosperity Practice Group, The World Bank. syu2@worldbank.org.

1. Do-files

Users can run the do-file “OccupationLevelData.do” for processing the occupation-level datasets, as the input for the second step on Databricks.

To run the scripts, new users need to change the directory in line 7. After changing the directory, running the do-file will produce the dataset “occupation_scores.dta” under folder “root/workdata”.

2. Notebooks

Before running the notebooks, the user should have access to the raw LightCast job posting data on Databricks. Then the user can run the do-files and upload the processed dta file “occupation_scores.dta” to Databricks, under folder “/processedData/”. Also, the user should upload datafiles “ILO2025_Final_Scores_ISCO08_Gmyrek_et_al_2025.xlsx” and “GBB-scores-2024-10-16.csv” to folder “/rawdata/” on Databricks.

- **Notebook 1. “LightCast Data Aggregation”:** This notebook processes raw job postings data from LightCast and produce the state-industry-occupation-quarter level aggregated dataset for empirical analysis. Raw job posting data stored as Delta Data Lake at “/Volumes/prd_datascience_data/volumes/data/He/RawData_202507/”
- **Notebook 2. “Prepare Data for Analysis”:** This notebook uses the aggregated LightCast data , and the occupation-level indicators to generate working data for DiD and event-study analysis.
- **Notebook 3. “Results”:** This notebook produces main figures and tables in the paper.

Output files

Output figures and tables from the notebook are under folder “root/output/”.

Data Statement

The project uses three sets of data.

1. Publicly available data

Dataset	Located at	Data URL and additional information	Date acquired
GBB-scores-2024-10-16.csv	"root/rawdata/GBB-scores-2024-10-16.csv"	https://pgmyrek.shinyapps.io/AI_Data_Portal_Research/	10/16/2024
ILO GenAI exposure	"root/rawdata/ILO2025_Final_Scores_ISCO08_Gmyrek_et_al_2025.xlsx"	https://github.com/pgmyrek/2025_GenAI_scores_ISCO08/blob/main/Final_Scores_ISCO08_Gmyrek_et_al_2025.xlsx	06/15/2025
JOLTS job vacancy data	"root/rawdata/JOLTS.xlsx"	https://download.bls.gov/pub/time.series/jt/jt.data.2.JobOpenings (JTS000000000000000JOL series)	07/05/2025

2. Data are available upon request

- The AI complementarity score from Pizzinelli et al. (2023) is provided by the authors. As the research team is not authorized to share the dataset publicly, interested users should contact the authors directly for access.

Dataset	Located at	Data URL and additional information	Date acquired
AI complementarity from Pizzinelli et al. (2023)	"root/rawdata/AIOE_CAIOE_theta_for_sharing.xlsx"	The authors of Pizzinelli et al. (2023) shared the data file via email. This dataset should not be shared to public without approval from the authors of Pizzinelli et al. (2023).	11/15/2024
Occupation description and skills	"root/rawdata/onet_occ.xlsx"	The authors of Martins-Neto et al. (2025) shared the data file processed from O*NET.	12/09/2024

3. Proprietary dataset

- LightCast job posting data is a proprietary dataset provided by the company. The dataset is hosted on cloud computing platform Databricks, with access limited to the research team.
- Monthly postings for the US is stored in folder, which is shareable to the public

Dataset	Located at	Data URL and additional information	Date acquired
Monthly postings for the US	"root/rawdata/LC_USA_MonthlyPostings.csv"	Processed from LightCast	07/05/2025

List of Exhibits

Exhibit name	Output filename	Script	Note
Figure 1	"LC_JOLTS.xlsx" under folder "root/output"	-	This figure is inside the xlsx file, with JOLTS data downloaded from the website as indicated above, and merged with LightCast monthly job posting numbers.
Figure 2	"/Version2025Sep/occ_scatter.png"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Figure A2	"/Version2025Sep/two_occupation_trend.png"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Figure A1	"/Version2025Sep/job_postings_raw.png"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Figure 3	"/Version2025Sep/job_postings_normalized.png"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Table A1	"/Version2025Sep/sumstat_industry.csv"; "/Version2025Sep/sumstat_edu.csv"; "/Version2025Sep/sumstat_exp.csv"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Table 1	"/Version2025Sep/did_results_baseline.csv"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Figure 4	"/Version2025Sep/event_baseline.png"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Figure A3	"/Version2025Sep/event_robust_nowt.png"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Figure A4	"/Version2025Sep/event_robust_cl_occ.png"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Figure 9	"/Version2025Sep/event_robust_ilo25.png"; "/Version2025Sep/event_robust_AIOE.png"; "/Version2025Sep/event_robust_noctrl.png"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Figure 8	"/Version2025Sep/event_robust_interperson.png"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Figure 7	"/Version2025Sep/event_robust_sub1_com.png"; "/Version2025Sep/event_robust_sub2_res.png";	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"

Exhibit name	Output filename	Script	Note
	"/Version2025Sep/event_robust_sub3_phy.png"; "/Version2025Sep/event_robust_sub4_cri.png"; "/Version2025Sep/event_robust_sub5_rou.png"; "/Version2025Sep/event_robust_sub6_ski.png"		
Figure A6	"/Version2025Sep/event_robust_GBBtreatment.png"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Figure A5	"/Version2025Sep/event_3groups_50_90.png"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Figure 10	"/Version2025Sep/did2_coeff_byeduexp.png"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Figure 11	"/Version2025Sep/did2_coeff_byindustry.png"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Table A2	"/Version2025Sep/did_result_s_byedu_shortlong.csv"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Table A3	"/Version2025Sep/did_result_s_byexp_shortlong.csv"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Figure 5	"/Version2025Sep/placebo_did.png"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"
Figure 6	"/Version2025Sep/placebo_eventstudy.png"	"3 Results.ipynb"	Found on Databricks, under the root folder "/Volumes/prd_datascience_data/volumes/data/He/AllLabor/figures"

Reference

Pizzinelli, Carlo, Augustus J Panton, Ms Marina Mendes Tavares, Mauro Cazzaniga, and Longji Li (2023). *Labor market exposure to AI: Cross-country differences and distributional implications*. International Monetary Fund.

Martins-Neto, A., Liu, Y., Khurana, S., & Porras Lopez, J. M. (2025). Click, Code, Earn: The Returns to Digital Skills. Juan Manuel and PORRAS LOPEZ, Juan Manuel, Click, Code, Earn: The Returns to Digital Skills (September 25, 2025).

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5529280