From Job Displacement to Task Reallocation: Evidence from Temporal Analysis of Data Science Job Postings

Gunasai Muppala
gunasai58@gmail.com
Independent Researcher, United States

Abstract

Recent advances in artificial intelligence (AI), particularly generative AI, have increased public concern regarding the potential displacement of data science professionals. Popular discourse frequently frames AI as a direct substitute for analytical labor, fueling fears of widespread job loss. However, jobs are bundles of tasks, and technological change often reshapes work internally rather than removing occupations outright. This study adopts a task oriented perspective to examine whether AI diffusion corresponds with job displacement or, instead, task reallocation within the data science profession.

Using a comparative analysis of U.S. based Data Scientist job postings from a pre-diffusion baseline derived from a Glassdoor scraped dataset (published in the late 2010s) and a post-diffusion dataset of 2023 postings curated by Luke Barousse, this paper analyzes changes in task and skill related language as proxies for evolving employer expectations. The analysis indicates substantial shifts in the composition of emphasized skills and responsibilities rather than a collapse in demand. In particular, several traditional tools and routine signals have declined in relative share, while signals related to orchestration and modern production workflows have increased. These findings are consistent with a task reallocation framework and motivate task level measurement as a complement to occupation level narratives.

Keywords: artificial intelligence, labor markets, task reallocation, job postings, data science

1. Introduction

Advances in artificial intelligence have repeatedly triggered anxiety about the future of work. In recent years, the rapid diffusion of generative AI tools, capable of writing code, producing analyses, and generating text, has amplified such concerns among knowledge workers in analytical fields such as data science. Public narratives often imply that AI will replace data scientists altogether. Yet occupations are not indivisible units; they are bundles of tasks (Autor, Levy, & Murnane, 2003; Autor, 2015). If technology changes the cost or speed of certain tasks, the job can evolve without disappearing.

This paper reframes the question from "Will AI replace data scientists?" to "How are employers reallocating tasks and skill expectations within Data Scientist roles as AI tools diffuse?" To address this question empirically, I use job postings as a proxy for employer demand and expectations, and compare a pre-diffusion baseline (Glassdoor scraped postings) to a post-diffusion snapshot from 2023. The goal is descriptive: to measure shifts in the relative emphasis of tasks and skills, not to establish causality.

2. Data

The empirical analysis uses two publicly available datasets representing different periods.

Pre-diffusion baseline. The baseline dataset is a Glassdoor scraped job postings dataset distributed via Kaggle. It includes job titles, employer job descriptions, and salary estimate strings. I filter postings to titles containing "Data Scientist, " resulting in a baseline sample (n = 358) used for text based task and skill indicators.

Post-diffusion period (2023). The post period uses the Hugging Face dataset lukebarousse/data_jobs, which contains standardized job titles, posting dates, geography, and extracted skills for data related roles in 2023. I restrict the analysis to U.S. based postings with job_title_short = "Data Scientist, " yielding n = 58, 830 postings.

Occupational grounding and context. Task statements for Data Scientists from O*NET (SOC 15-2051.00) are used to ground the paper's task taxonomy. BLS Occupational Outlook Handbook and OEWS materials are used for contextual framing only.

Because the baseline dataset contains full job description text while the 2023 dataset provides structured skills, cross period comparisons focus on normalized shares and are interpreted conservatively (Barousse, 2023; National Center for O*NET Development, n.d.; U.S. Bureau of Labor Statistics, n.d.).

3. Methodology

Each posting is coded using binary indicators that capture whether a given task or skill category is mentioned at least once. In the baseline dataset, indicators are computed by keyword pattern matching on job descriptions. In the 2023 dataset, indicators are computed primarily from extracted skills fields (and additional keyword checks where applicable). To compare periods, I compute the share of postings that mention each indicator and report changes in percentage points.

Task taxonomy. I organize indicators into a small set of interpretive task categories anchored to O*NET descriptions: (1) Data Preparation and Cleaning, (2) Feature Engineering, (3) Model Development and Training, (4) Experimentation and Validation, (5) AI and Automation Tool Usage, (6) Deployment and Monitoring, and (7) Business and Stakeholder Communication. The taxonomy is used to summarize patterns rather than claim definitive boundaries between tasks.

This study is descriptive. Observed differences coincide with AI diffusion but may also reflect other changes in tooling, market conditions, and dataset coverage. Accordingly, the results are reported as associations.

Table 1. Task taxonomy used for coding job postings.

Task category	Description	Example phrases/signals
Data Preparation & Cleaning	Cleaning, joining, and transforming raw data; ETL and data quality checks.	ETL, data pipeline, data cleaning, data validation
Feature Engineering	Deriving variables and features for modeling and analytics.	feature engineering, feature selection
Model Development & Training	Building statistical or ML models and running training workflows.	modeling, training, ML algorithms
Experimentation & Validation	Evaluating models and hypotheses using tests and metrics.	A/B testing, cross-validation, metrics, and evaluation
AI & Automation Tool Usage	Using or integrating AI systems and tools, including generative AI.	LLM, GPT, prompt, RAG, LangChain
Deployment & Monitoring	Deploying models and monitoring performance in production.	deployment, MLOps, monitoring, production

Business & Stakeholder Communication	Communicating insights and translating requirements across teams.	stakeholders, communicate, storytelling, requirements

4. Results

4.1 Posting volume in 2023

Figure 1 shows the monthly distribution of U.S. Data Scientist postings in 2023 (post-diffusion dataset). Posting volume remains consistently high across the year, with month to month variation but no sustained collapse.

4.2 Pre vs post shifts in skill emphasis

Figure 2 and Table 2 summarize changes in the share of postings that mention selected tools and signals between the baseline dataset and the 2023 dataset. Several foundational skills remain prevalent (e.g., Python and SQL), while other tools shift in share. In Table 2, the largest increases among the tracked indicators include: mlops +3.47 pp (3.63% → 7.10%); azure +3.03 pp (8.66% → 11.69%); power bi +1.24 pp (8.38% → 9.62%). The largest decreases among the tracked indicators include: r -14.43 pp (58.66% → 44.23%); spark -10.28 pp (27.09% → 16.81%); ab testing -9.10 pp (9.22% → 0.12%). Because the baseline uses full job descriptions while the 2023 dataset uses extracted skills, these changes should be interpreted as conservative proxies for shifting emphasis rather than precise task replacement.

4.3 Emergence of AI mediated terms

Figure 3 plots the share of 2023 postings with explicit generative AI terms (e.g., LLM/GPT related keywords). Although the overall share is small in 2023, the trend is upward over the year, consistent with gradual diffusion.

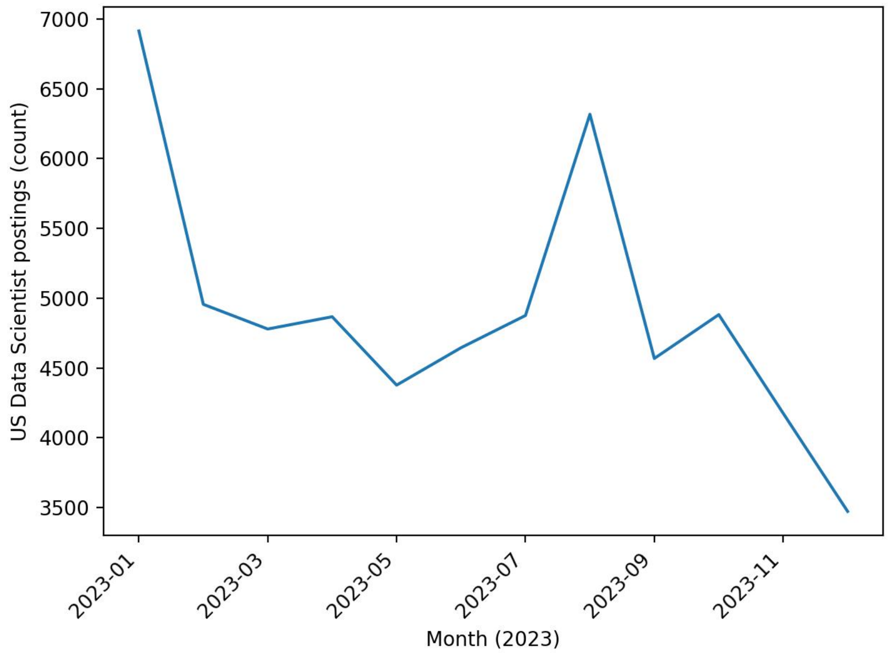

Figure 1. Monthly volume of U.S. Data Scientist job postings in 2023 (post-diffusion dataset).

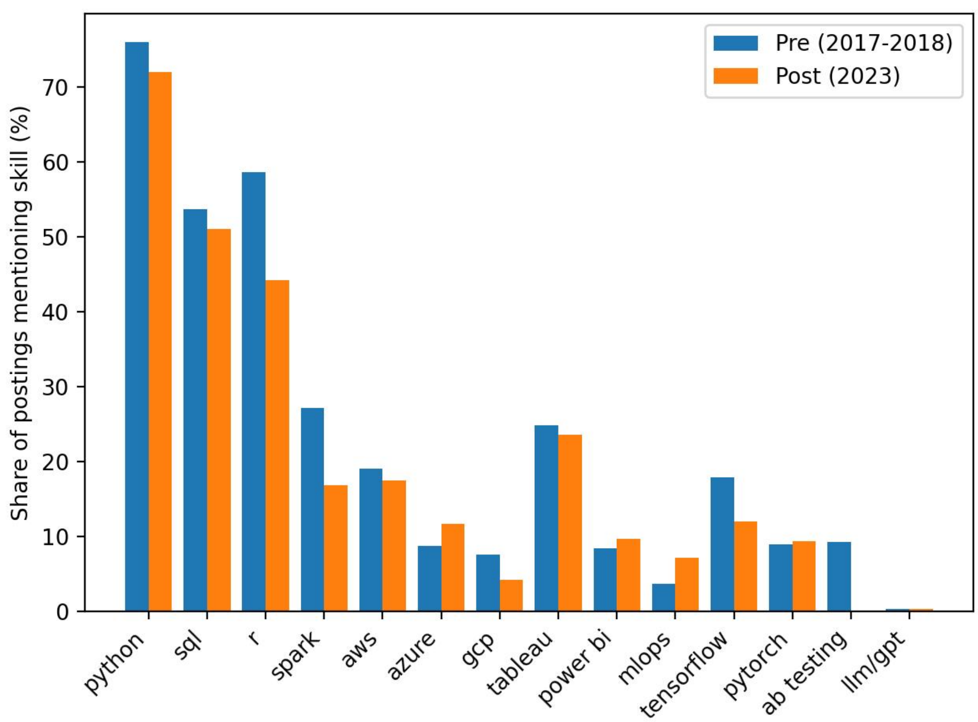

Figure 2. Pre vs post comparison of selected skill/indicator shares (baseline dataset vs 2023 dataset).

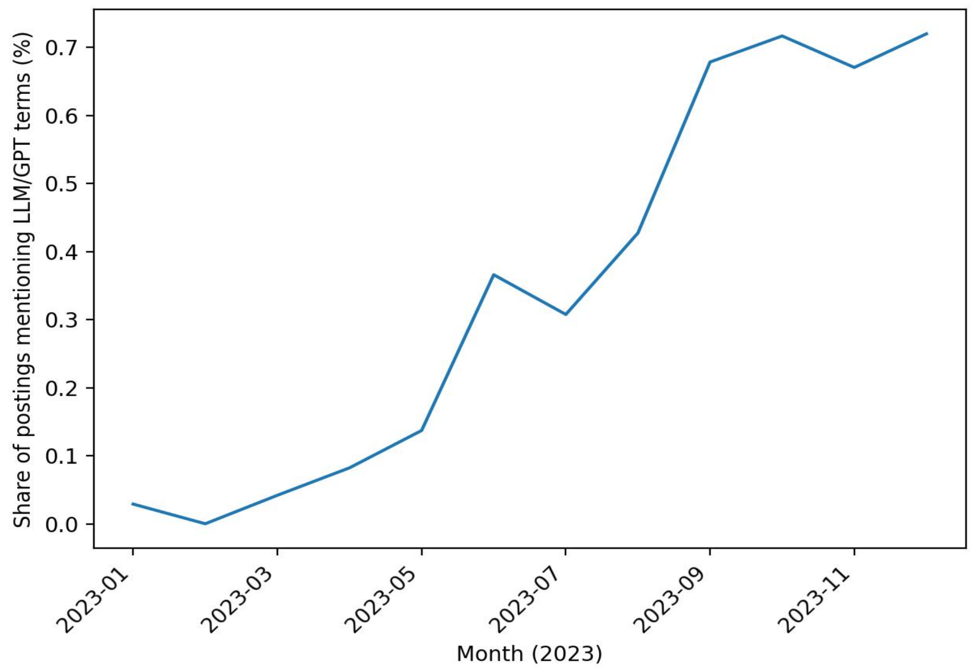

Figure 3. Monthly trend of explicit LLM/GPT related term mentions in 2023 postings.

Table 2. Selected indicator shares in the baseline vs 2023 dataset.

Indicator	Baseline share	2023 share	Delta (pp)
MLOPS	3.63%	7.10%	+3.47 pp
AZURE	8.66%	11.69%	+3.03 pp
POWER BI	8.38%	9.62%	+1.24 pp
PYTORCH	8.94%	9.36%	+0.42 pp
LLM/GPT	0.28%	0.33%	+0.05 pp
TABLEAU	24.86%	23.56%	-1.30 pp
AWS	18.99%	17.49%	-1.51 pp
SQL	53.63%	51.05%	-2.58 pp
GCP	7.54%	4.21%	-3.34 pp
PYTHON	75.98%	72.04%	-3.94 pp
TENSORFLOW	17.88%	11.96%	-5.91 pp
AB TESTING	9.22%	0.12%	-9.10 pp
SPARK	27.09%	16.81%	-10.28 pp
R	58.66%	44.23%	-14.43 pp

5. Discussion

The results are consistent with a task reallocation interpretation (Autor, 2015; Acemoglu & Restrepo, 2020). Foundational analytical skills remain central, while employer signaling shifts toward orchestration and modern workflow responsibilities (e.g., MLOps). Rather than indicating the disappearance of Data Scientist roles, the evidence suggests that organizations are reshaping the internal mix of responsibilities, automating or abstracting some routine work while increasing emphasis on integration, evaluation, and cross functional communication.

This perspective helps reconcile two observations that often appear in tension in public discourse: (1) AI tools can automate pieces of analytical work, and (2) demand for data science talent persists. The task level lens supports both: the job persists, but the allocation of work shifts.

6. Limitations and Future Work

Job postings reflect employer expectations and signaling, not realized day to day work. Dataset structure differs across periods (full descriptions vs extracted skills), which limits direct comparability. The study focuses on U.S. Data Scientist roles only, so generalization to other occupations is not guaranteed. Finally, the analysis is descriptive and does not attempt causal identification; observed shifts coincide with AI diffusion but may also reflect other changes in the labor market and tooling.

Future work could incorporate consistent full text postings across periods, expand to adjacent roles (e.g., Data Analyst, ML Engineer), and add firm level or longitudinal designs to better isolate mechanisms.

7. Conclusion

Using two public job postings datasets, this paper documents shifts in the composition of tasks and skill signals within Data Scientist roles between a pre-diffusion baseline and a 2023 post-diffusion snapshot. The evidence is more consistent with task reallocation than job displacement: core analytical skills remain widespread, while signals related to orchestration and production workflows gain relative prominence. Task level measurement provides a practical complement to occupation level narratives and offers a more grounded way to discuss AI's impact on analytical work.

References

Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from U.S. labor markets. Journal of Political Economy, 128(6), 2188-2244. https://doi.org/10.1086/705716

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. Journal of Economic Perspectives, 29(3), 3-30. https://doi.org/10.1257/jep.29.3.3

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. Quarterly Journal of Economics, 118(4), 1279-1333. https://doi.org/10.1162/003355303322552801

Barousse, L. (2023). data_jobs: Real-world data analytics job postings from 2023 [Data set]. Hugging Face. https://huggingface.co/datasets/lukebarousse/data_jobs

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn, and what does it mean for occupations and the economy? AEA Papers and Proceedings, 108, 43-47. https://doi.org/10.1257/pandp.20181019

Kaggle. (n.d.). Data Scientist Jobs (Glassdoor-scraped postings) [Data set]. https://www.kaggle.com/datasets/andrewmvd/data-scientist-jobs

National Center for O*NET Development. (n.d.). O*NET OnLine: Data Scientists (15-2051.00). https://www.onetonline.org/link/summary/15-2051.00

U.S. Bureau of Labor Statistics. (n.d.). Data Scientists: Occupational Outlook Handbook. https://www.bls.gov/ooh/math/data-scientists.htm

U.S. Bureau of Labor Statistics. (n.d.). Occupational Employment and Wage Statistics (OEWS): Data Scientists (15-2051). https://www.bls.gov/oes/2023/may/oes152051.htm