# AI Bros? Gender Disparities in the Research on AI and its Workplace Footprint

**Sujin Lee, Tian Lou, Bruce A. Weinberg**[1]

Artificial Intelligence is widely seen as a new general purpose technology with the potential to radically alter the labor market. This paper contributes to our understanding of gender disparities in AI. First, we draw on unique data to show that women are underrepresented among the researchers working on AI and that most of this gap arises within observed categories such as fields, occupations, and the genders of principal investigators. Second, we analyze the workforces at the businesses that hire AI researchers as a way of identifying businesses that use cutting-edge AI. Here we find that the firms hiring AI researchers have lower shares of female employees and higher gender pay gaps than other firms in their detailed industries, including other firms that are hiring researchers outside of AI. This suggests wide differences in AI exposure of the workforce.

**Note that quantitative results have been produced but are redacted pending disclosure.**

[1] Ohio State University. Weinberg is also affiliated with the IZA Institute for the Study of Labor and the National Bureau of Economic Research.

The authors gratefully acknowledge funding from National Science Foundation (Award # 2332573). We also thank the Institute for Research on Innovation and Science (IRIS), the Ohio Longitudinal Data Archive (OLDA), The Ohio State University HR Department, and CHRR at the Ohio State University for their data support.

## 1. **Introduction**

Women are underrepresented in many science, technology, engineering, and mathematics (STEM) fields and even more so in the computer and information and communication technology fields (Ngo & Dustan, 2024; Adams & Kirchmaier, 2016; Lerchenmueller, & Sorenson, 2018).[2] How does representation in artificial intelligence compare to other STEM and computer fields and, as AI ripples out into business, how does the gender mix of the employees at AI-touched firms compare to that at other businesses? The answers to these questions have implications for access to some of the most technologically advanced jobs and perhaps for the development and implementation of a new general purpose technology.

This paper seeks to quantify gender differences in participation in and exposure to artificial intelligence.[3] Our primary analysis looks at gender differences in who is conducting research on artificial intelligence. Here we rely on unique data from the UMETRICS project, which provides information on all people employed on research projects at 45 universities with 103 campuses accounting for roughly 37% of total academic research and development in the United States.[4] Because these data are built on payment records from universities (as opposed to, for instance, authorship lists on articles), they provide visibility into everyone employed on projects, an important way of identifying the relevant population given that previous work shows a substantial gender imbalance in authorship attribution (Ross, et al., 2022).

[2] Blackburn, 2017; Sharma, 2023 provide some evidence that women's access to professional and STEM-based education has improved in recent years in the US.

[3] A few existing studies consider gender disparities in artificial intelligence research. For instance, Young et al. (2023) develops a unique cross-country data set to quantify gender gaps in artificial intelligence across countries. Nedungadi et al. (2024) find that women continue to lag men on "single-authored works and primary authorship positions" but are gaining on highly cited papers.

[4] While many universities are research intensive, our data also include many branch campuses, which are less research intensive. For additional work using UMETRICS, see Chang et al. (2019), Buffington et al. (2016), Lane et al. (2013), Lane et al. (2015), Weinberg et al. (2014), Zolas et al. (2015), and Ross et al. (2022).

Our estimates indicate that women are considerably less likely to be employed on AI research projects than other research projects. In other research, we have found that women principal investigators tend to employ more women than male principal investigators and that there are gender differences in the composition of occupations with, for instance, a larger share of staff than faculty being women (Berkes, Sikhova, and Weinberg, 2024). To quantify these effects, we decompose the gender gap in participation in AI projects relative to other projects by the gender of principal investigators and the occupation mix of teams. We find that differences in both account for a small minority of gender differences in participation in AI projects, indicating that most of the gender differences arise within occupations and are conditional on the gender of principal investigators. AI research cuts across many fields that differ in terms of gender compositions with fields such as health and biology having larger shares of female researchers while the physical sciences and engineering tend to have lower shares of female researchers. Decompositions by field indicate a substantial between effect, but that the majority of the gender gap is within fields. Thus, most of the gap in participation in AI projects cannot be explained by observable characteristics of projects.

Our finding of lower representation of women in AI research holds when we characterize projects in terms of their AI intensity. Even after controlling for employee and grant characteristics such as occupation, principal investigator gender, and field, employees highly engaged in AI research are less likely to be female than those engaged in AI research but to a lesser extent and these researchers are, in turn, less likely to be female than researchers on non-AI projects. This gradient indicates that the gender gap in AI research increases as the AI-relatedness of the project increases.

To begin to understand how AI may be touching different portions of the broader workforce, we conduct a second (pilot) analysis. To do this, we seek to identify employers that are using AI. Doing so is far from trivial. Our approach draws on the insights of J. Robert Oppenheimer and others that early in their development and implementation, new knowledge is generally transmitted through flows of people (e.g., Zolas et al., 2014). Thus, we identify the private (i.e., non-academic, non-government) businesses that hire the AI and non-AI research staff, graduate students, and postdoctoral scholars in our UMETRICS data from Ohio State University after they completed working on research projects.[5] We further describe the gender mix and earnings of the businesses that employ these researchers. Our initial analysis here relies on Ohio Unemployment Insurance (UI) Wage Records and Quarterly Census of Employment and Wages (QCEW) to track workers and identify employers (our approach is akin to Alderucci et al., 2024). We define AI industries as industries where at least one firm hired a researcher that was an author on an AI research paper. Examples of AI industries are Custom Computer Programming Services, Software Publishers, and Media Streaming Distribution Services, Social Networks, and Other Media Networks and Content Providers. We then categorize employers in AI industries to four types: 1) firms that have only hired AI researchers; 2) firms that have only hired non-AI researchers; 3) firms that have hired both AI and non-AI researchers; and 4) firms that have never hired any researchers.

We find that in general, there are more male workers than female workers in AI industries. Looking across the four types of firms in AI industries, firms that hired both AI and non-AI researchers have the lowest average female percentage, followed by firms that only hired AI-researchers and firms that did not hire any researchers. Firms that only hired non-AI

[5] Our sample includes approximately 400 AI researchers and over 23,000 non-AI researchers although, as one would expect, many of these employees are not subsequently employed in Ohio.

researchers employ more women than men. We also observe considerable variations in gender compositions across firms, with the largest variations in firms that did not hire any researchers. We check the robustness of these findings by limiting the sample to full-year employees, i.e., people who worked at least six consecutive quarters in a firm. The results generally remain the same, except for even lower percentage of female workers in firms that only hired AI researchers and even higher percentage of female workers in firms that only hired non-AI researchers.

We further examine the earnings differences between female and male workers. Our results reveal that not only there are fewer female workers than male workers in AI industries, but female workers in AI industries also earn significantly less than male workers. Among the firms in AI industries, the average earnings of male workers in firms that only hired AI researchers are more double female workers' average earnings.[6] We also observe significantly higher earnings for male workers in firms that only hired non-AI researchers and firms that did not hire any researchers, but the earnings differences are smaller than the gender gap observed in firms that only hired AI researchers. The smallest earnings differences are in firms that hired both AI and non-AI researchers, where male workers still earn slightly more than female workers. When limiting the sample to full-year employees, the findings are consistent.

In summary, we find that there are more male workers than female workers in AI industries, and that the gender gap in earnings is significant. We find the most gender-related differences in firms that only hired AI-researchers and firms that hired both AI and non-AI researchers.

[6] The earnings are 2022 annual earnings. We limit the sample to people who earned at $1,000 in 2022. While our estimates focus on people employed for a full quarter (as identified by having earnings with that firm in at least one immediately preceding and succeeding quarter), they do not control for hours worked. Results are similar when we focus on full-year workers or on any worker.

While others have found that men are more likely to use artificial intelligence at work than women (Humlum and Vestergaard, 2024),[7] our estimates are different in that they show the gender mix and earnings gaps of the businesses that are using AI. One could, for instance find that men are more likely to use AI, but that they do so at businesses that are heavily female. (In ongoing work, we are estimating how the use of AI by businesses affects the demand for workers, which depends on the production technology (e.g., the elasticity of substitution between different inputs) and the elasticity of demand for final output and the implications for women. For now, we provide a "footprint" analysis.)

Our research relates to a long line of research studying how technologies, from computers to the internet to robots, have transformed work and the economy (Acemoglu and Restrepo, 2018; Agrawal et al., 2019; Frank et al., 2019) and have had widely disparate impacts across groups. Although there is considerable debate around its medium-run effects (e.g., Acemoglu, 2024 versus Briggs and Kodnani, 2023, and McKinsey Global Institute, 2024), AI has the potential to continue transforming labor demand and the labor market. Recent work has pointed to a substantial gender gap in the use of artificial intelligence at work in favor of men (Humlum and Vestergaard, 2024). If so, that could represent a departure from many digital technologies which, if anything, have appeared to favor women relative to men (Weinberg, 2020 and Welch 2020).[8]

[7] Previous studies examined gender disparities in AI research and occupations, but authors rely on the existing standard classification of occupations related to data science, AI exposure indices, or the use of specific AI tools such as ChatGPT (Young et al., 2023; Nedungadi et al., 2024; Conde-Ruiz et al., 2024). Brynjolfsson et al. (2018) develop the AI Suitability for Machine Learning (SML) index and Webb (2019) develops the AI Exposure score, but none apply these to study gender differences.

[8] Aksoy et al., (2021) argue that robots are increasing the gender gap. Conde-Ruiz et al. (2024), for instance, finds that women in Spain tend to study in fields that are more routine intensive, exposing their jobs to automation.

## 2. Data

### UMETRICS

The UMETRICS data contain administrative information on all grant-related transactions at 45 universities covering 103 campuses accounting for roughly 37% of total academic research and development in the United States. While these institutions are not a random sample, they provide a valuable window into the academic research enterprise, especially at research intensive institutions. The inclusion of many branch campuses broadens the population covered.

The UMETRICS data provide detailed transaction records for each employee who received compensation or contributed time to an award during a payment period, typically a month. The data is well-suited for studying research activities in AI as it enables the identification of all individuals involved in sponsored projects, including trainees (such as graduate students and postdocs), staff, and those who may not be listed as authors on publications.

Key information includes employees' occupational classifications, divided into six categories: faculty, undergraduate students, graduate students, students at an unknown level, postdocs, and staff. Additionally, UMETRICS provides data on employee-award characteristics, such as principal investigator (PI) status and department affiliation for PIs, as well as employee-specific attributes like gender, which can be linked to transaction data. Gender is imputed by IRIS using a large language model applied to employees' first names.[9] The dataset also includes award-level details, such as funding sources and award titles.

[9] In separate work, we compared a similar imputation to self-reported data on gender for 12,867 faculty from one participating institution and, for that population, gender is imputed for 95% of people with a precision of 93% (overall, for men, and for women).

The full UMETRICS dataset comprises 45 universities with 103 campuses, covering 37% of total academic R&D. For our analysis, we begin with the 41 UMETRICS universities that provided employee transaction data that cover 33% of academic R&D. We then excluded 7 universities due to the absence of data identifying PIs and another 7 universities that lacked employee demographic information. This resulted in a final sample of 27 universities, which represents 22.5% of academic R&D. We restrict the grant receipt data to the period from 2015 to 2022, as AI research activities have emerged relatively recently.

Several employee-specific characteristics may contribute to gender disparities between AI and non-AI research. For example, AI research might be concentrated in occupational groups or fields with low female representation. Additionally, the gender of the Principal Investigators (PIs) on a project may influence the likelihood of hiring employees of a particular gender (Berkes, Sikhova, and Weinberg, 2024, find that female PIs tend to employ more women than men). To explore these factors, we construct a number of characteristics for each employee.

People can be employed under multiple job titles.[10] For each employee's occupation, we identify each employee's most frequently held occupation. People work on multiple grants at a point in time or over time. When we assign a principal investigator gender, we first assign a gender to each grant based on the most frequently observed gender among the principal investigator(s) on that project. We then determine whether each employee has been employed by more male or female principal investigators. To assign fields to each grant, we utilize multiple sources of information, including principal investigator's department affiliation, the funding source, and award title. Fields are classified into AI, non-AI but related, agriculture, biological

[10] For instance, a staff person who is enrolled as a student (e.g., because of tuition benefits) could appear both as a staff member and as a student. A graduate student could subsequently be hired as postdoc, staff member, or as a faculty member. In addition, some occupations are ambiguous and may misclassified. See Ikudo et al. (2019) for a discussion of these issues.

science, physical science, engineering, health, social science, humanities, and others. Similarly, we determine each employee's specific field based on the most frequent field of the grant they have worked on.

To measure how involved an employee is in AI research, we use data from Elsevier, which provides detailed information about the authors of AI publications. This is linked to UMETRICS transaction records using author names. Key variables include percentile ranks of an author's productivity in AI-related publications (prolificity), their engagement with diverse AI concepts (breadth), the concentration of AI within their overall body of work (density), and the consistency of their work on AI over time (dedication).

We first measure the AI-relatedness of each grant based on the attributes of principal investigator(s) identified as AI authors. Specifically, we calculate the mean percentile rank for each author attribute. For grants without PIs, we use the mean author attribute of all employees on the grant identified as AI authors. Then, we measure the AI-relatedness of each employee based on the AI-relatedness of the grants they have worked on. To be specific, an employee's AI-relatedness, $AI_i$, is defined as the weighted sum of the AI-relatedness of the grants ($A_{ij}$), where the weights are the total number of months ($m_{ij}$) the employ was employed on that grant:

$$AI_i = \frac{\sum_j AI_{ij} \times m_{ij}}{\sum_j m_{ij}}$$

For our purposes to examine gender disparities in AI research activities, we categorize employees into three groups: those who have never been involved in AI research ($AI_i = 0$), those with low involvement in AI research (conditional on $AI_i > 0$, bottom 50%), and those with high involvement in AI research (conditional on $AI_i > 0$, top 50%).

In the next section, we discuss the administrative data used to examine AI researchers' labor market outcomes. To link the UMETRICS data with other data sources, we need additional

person identifiers. The Ohio State University's (OSU) human resources department provided us with person-level records, which contain unique person identifiers and demographic information, such as gender and race. The missing rates for identifiers and demographic variables are very low.

**Ohio Administrative Records**

We use Ohio administrative records to track UMETRICS researchers' labor market outcomes after their employment on research projects end.[11] We also utilize the rich information in the administrative records to further explore gender compositions and wage differences between female and male workers. The UMETRICS sample is limited to researchers who were research staff, graduate students, and postdoctoral scholars at The Ohio State University. We focus on people by private business rather than universities or government.

The administrative data comes from Ohio Longitudinal Data Archive (OLDA), which manages a variety of datasets from state agencies.[12] In this project, we use the Unemployment Insurance (UI) wage records, firm-level Quarterly Census of Employment and Wages (QCEW) data, and Bureau of Motor Vehicles (BMV) records. As of the time of writing the manuscript, we have access to Ohio administrative data from 2010 through the first quarter of 2024.

The UI wage records cover earnings and employment for people working in Ohio and in jobs covered by the UI system. The data contains a person's quarterly earnings from an employer, weeks worked for that employer during a quarter, and the employer's identifier. Information in the UI data is from employers' payroll records and has been found to be more

[11]Technically, we focus on the end date of the last pay period on which we observe them employed.
[12] The Ohio Longitudinal Data Archive is a project of the Ohio Education Research Center (oerc.osu.edu) and provides researchers with centralized access to administrative data. The OLDA is managed by CHRR at The Ohio State University (chrr.osu.edu) in collaboration with Ohio's state workforce and education agencies (olda.ohio.gov), with those agencies providing oversight and funding.

accurate and reliable than self-reported data from surveys (Hotz and Scholz, 2002). However, the data have their own limitations. The UI data do not cover jobs outside of Ohio. Nor do they cover certain types of workers, such as self-employed, military, federal employees, and independent contractors. Therefore, if a person has worked in other states or has worked in jobs not covered by the Ohio UI system, we are not able to accurately measure his/her earnings and employment.

The QCEW is a comprehensive dataset that offers firm-level information, such as monthly employment, total payroll, firms' locations, industry classifications (via the North American Industry Classification System (NAICS) codes), and establishment types (single-unit or multi-unit) on a quarterly basis. Similar to the wage records, the coverage of the QCEW data includes nearly all employers subject to state UI laws, making it possible for us to obtain work locations and industries for workers in the UI wage records. The drawback of the QCEW data is that if a person works for a multi-unit firm, such as Walmart or McDonalds, or a firm that operates in multiple industries, such as Amazon or Procter and Gamble, we cannot accurately identify the person's actual work location and industry. In the analysis for this project, we use the master-units' locations and industries for multi-unit firms.

We obtain workers' gender information from Ohio Bureau of Moter Vehicle (BMV) data. Ohio residents need to report their gender and birth dates when registering for state license and driver's licenses and when they renew their licenses. We use the demographic information in their most recent records. Only a very small percentage of workers have missing gender information in our data.[13]

[13] The majority of the missing gender information is due to the absence of BMV records. A very small percentage of the missingness is because some individuals did not report their gender in the BMV data.

We construct the sample for labor market analysis by linking UMETRICS researchers to UI wage records and identifying their employers after they left the research projects at OSU.[14] We then match these employers to their records in the QCEW data to determine their industries. Our ultimate goal is to compare the firms and employees of firms that hire AI researchers to other firms in the same 6-digit NAICS codes that do not hire AI researchers. The industries of employers that hired any AI researchers are classified as "AI industries". These include "Custom Computer Programming Services", "Research and Development in the Physical, Engineering, and Life Sciences", etc. Some researchers take jobs in "Universities and Colleges", and we exclude these employers, as well as the "Colleges, University, and Professional Schools" industry, from the sample.

Next, we identify all employers in the AI industries and all their employees in 2022. We categorize these employers into four categories based on whether they hired any UMETRICS researchers from OSU: (1) firms that have only hired AI researchers; (2) firms that have only hired non-AI researchers; (3) firms that have hired both AI and non-AI researchers; and (4) firms that do not hire any researchers.

As with most UI wage records, our data unfortunately do not include measures of hours worked or weeks worked. To obtain the cleanest measure of earnings abstracting from variations in weeks worked, our primary analysis focuses on "full-quarter" employees, who worked for a firm for at least three consecutive quarters in 2022 and earned at least $1,000.[15] To measure earnings, we drop earnings in the first and last quarter we observe an employee paid by a firm and focus on earnings in the "middle" quarter(s) of their employment. We also conduct an

[14] The current analysis does not threshold the employer sample based on measures such as duration of employment or earnings from an employer. We will limit the sample in the future analysis.
[15] See the Appendix for the complete definition of full-quarter employee.

analysis for all employees who worked at least one quarter in a firm and earned at least $100 during a quarter. This measure can be defined for more workers but includes people who were only employed at a firm for part of a quarter increasing the variation in hours and weeks worked for the firm. We also identify full-year employees, who worked six consecutive quarters, including all of 2022, the fourth quarter in 2021, and the first quarter in 2023 and earned at least $1,000.[16] These workers are the highest-attachment workers, but the stringent restrictions reduce the size of the population.[17] Earnings are calculated as the sum of quarterly earnings in calendar year 2022.

## 3. Methods

### 3.1. Decomposition of Gender Disparities in AI research

To quantify the difference in gender representation between AI and non-AI research, we employ an Oaxaca-Blinder-Kitigawa decomposition to break down the overall disparity into an occupational composition effect (a "between effect") and the within-occupation gender disparity effect (a "within effect") using the following equation:

$$F^{AI} - F^{NAI} = \sum_o S_o^{AI} F_o^{AI} - \sum_o S_o^{NAI} F_o^{NAI} = \sum_o (S_o^{AI} - S_o^{NAI}) F_o^{AI} + \sum_o (F_o^{AI} - F_o^{NAI}) S_o^{NAI} \quad (1)$$

where $F^{AI}$ and $F^{NAI}$ are the shares of female in AI and non-AI research, respectively and $S_o^{AI}$ and $S_o^{NAI}$ give the share of workers on AI and non-AI projects who are in occupation *O*. The first equality in equation (1) shows that $F^{AI}$ and $F^{NAI}$ are essentially weighted averages of the shares of female workers in each occupation weighted by the share of the AI and non-AI researchers in each occupation. The second equality gives our decomposition. The first term gives the

[16] We do not have hours worked in the UI wage records and therefore cannot know whether a person is a full-time or part-time employee. Worker categories are only based on earnings and quarters worked. See the Appendix for the complete definition of full-year employee.

[17] We note that these three populations are not mutually exclusive with all workers being a proper subset of full-quarter workers who are, in turn, a proper subset of full-year workers.

"between" term reflecting differences in the share of workers in each occupation between AI and non-AI research ($S_o^{AI} - S_o^{NAI}$), multiplied by the female share within AI research in each occupation ($F_o^{AI}$). This captures how differences in the distribution of workers across occupations contribute to the gender disparity. For example, if AI research is more likely to be conducted by researchers in occupations with lower female representation compared to non-AI research, this between effect will increase the overall disparity.

The second term measures the effect of differences in the female share within occupations between AI and non-AI research ($F_o^{AI} - F_o^{NAI}$), weighted by the share of workers in that occupation within non-AI research ($S_o^{NAI}$). This term captures the contribution of gender representation gaps within the same occupations. If women are underrepresented in AI research relative to non-AI research within a given occupation, this within effect would have a large contribution to the overall disparity. The same approach can be applied by substituting occupation with PI gender and field of study.

An alternative approach to the decomposition involves estimating the following linear probability model:

$$Female_{ik} = \beta_0 + \beta_1 AI_{ik} + \beta_2 Occ_{ijk} + \beta_3 PIGender_{ijk} + \beta_4 Field_{ijk} + \delta_k + \epsilon_{ijk}$$

The dependent variable, $Female_{ijk}$, is a binary variable indicating that employee $i$ in institution $k$ is female. The main explanatory variable, $AI_{ik}$, is a categorical variable that represents publication-based AI-relatedness of an employee. It is classified into none, low AI research activity, and high AI research activity. We control for $Occ_{ijk}$, a vector of categorical variables representing the occupation of employee $i$ on grant $j$ , $PIGender_{ijk}$, which captures the gender of the PI on grant $j$ that employee $i$ worked on, and $Field_{ijk}$, a categorical variable for academic fields of grant $j$ . The model also includes institution fixed effects, $\delta_k$, to account for institution-

specific factors. This specification allows for a detailed decomposition by isolating the contribution of occupational differences, PI gender, and field-specific effects to the observed gender disparity.

## 4. Results – PENDING DISCLOSURE REVIEW

## 5. Conclusion and Future Work

Artificial intelligence has the potential to serve as a general purpose technology, which might reshape the economy. This paper contributes to our understanding of gender disparities in AI. First, we draw on unique data to show that women are underrepresented among the researchers working on AI. To some extent this reflects the fields in which AI research is being conducted, but it is primarily due to differences within fields, researcher occupations, and the gender of principal investigators on research projects. Second, we analyze the workforces at the businesses that are implementing cutting-edge AI, as measured by hires of AI researchers. Here we find that the firms hiring AI researchers have lower shares of female employees and higher gender pay gaps than other firms in their industries, including other firms that are hiring researchers outside of AI. This suggests wide differences in AI exposure of the workforce.

We see our work on the AI research workforce as a novel window into who is advancing this potentially transformative new technology. We see our work on the broader workforce as a first step toward a more causal analysis of the impact of hiring AI workers on the employment and earnings of women and men.

## 6. References


Acemoglu, Daron. 2024. "The Simple Macroeconomics of AI." NBER Working Paper 32487.

Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. In *The economics of artificial intelligence: An agenda* (pp. 197-236). University of Chicago Press.

Adams, R. B., & Kirchmaier, T. (2016). Women on boards in finance and STEM industries. *American Economic Review*, *106*(5), 277-281.

Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Artificial intelligence: the ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives*, *33*(2), 31-50.

Aksoy, C. G., Özcan, B., & Philipp, J. (2021). Robots and the gender pay gap in Europe. *European Economic Review*, *134*, 103693.

Alderucci, Dean, Sagar Baviskar, Lee Branstetter, Nathan Goldschlag, Eduard Hovy, Andrew Runge, Prasana Tambe, and Nikolas Zolas. 2024. "Quantifying the Impact of AI on Productivity and Labor Demand: Evidence from U.S. Census Microdata." Working Paper.

Berkes, Enrico, Aiday Sikhova, and Bruce A. Weinberg. 2024. "Better Together? Faculty, Gender Match, and Credit in Science." Working Paper.

Blackburn, H. (2017). The status of women in STEM in higher education: A review of the literature 2007–2017. *Science & Technology Libraries*, *36*(3), 235-273.

Buffington, Catherine, Benjamin Cerf, Christina Jones, and Bruce A. Weinberg. 2016. "STEM Training and Early Career Outcomes of Female and Male Graduate Students: Evidence from UMETRICS Data linked to the 2010 Census." *American Economic Association Papers and Proceedings* 106 (No. 5, May): 333-38. *http://dx.doi.org/10.1257/aer.p20161124*

Briggs, Joseph and Devesh Kodnani. 2023. "The Potentially Large Effects of Artificial Intelligence on Economic Growth." Goldman Sachs Report.

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018, May). What can machines learn and what does it mean for occupations and the economy? In *AEA papers and proceedings* (Vol. 108, pp. 43-47). 2014 Broadway, Suite 305, Nashville, TN 37203: American Economic Association.

Chang, Wan-Ying, Wei Cheng, Julia Lane, and Bruce Weinberg. 2019. "Federal Funding of Doctoral Recipients: Results from new Linked Survey and Transaction Data." *Research Policy* 46 (No. 6, July): 1487-1492. https://doi.org/10.1016/j.respol.2019.03.001

Conde-Ruiz, J. I., Ganuza Fernandez, J. J., García, M., & Victoria Lanzón, C. (2024). AI and digital technology: gender gaps in higher education. *CESifo Economic Studies*, ifae020.
Gomez-Herrera, E., & Köszegi, S. T. (2022). *A gender perspective on artificial intelligence and jobs: the vicious cycle of digital inequality* (No. 15/2022). Bruegel Working Paper.

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., ... & Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences*, *116*(14), 6531-6539.

Humlum, Anders and Vestergaard, Emilie, The Adoption of ChatGPT (April 25, 2024). University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2024-50.

Hotz, V. Joseph, and John Karl Scholz. "Measuring employment and income for low-income populations with administrative and survey data." *Studies of welfare populations: Data collection and research issues* (2002): 275-315.

Ikudo, Akina, Julia Lane, Joe Staudt, and Bruce A. Weinberg. 2020. "Occupational classifications: A machine learning approach." *Journal of Economic and Social Measurement* 44 (Issue 2-3): 57-87. DOI: 10.3233/JEM-190463.

Lane, Julia, John King, and Lou Schwarz (2013), "The creation of new administrative data."Available at SSRN 2213916.

Lane, Julia, Jason Owen-Smith, Rebecca Rosen, and Bruce A. Weinberg. 2015 "New linked data on science investments, the scientific workforce and the economic and scientific results of science" *Research Policy* 44 (Issue 4, November): 1659-71. http://www.sciencedirect.com/science/article/pii/S0048733315000025

Lane, Julia I., Jason Owen-Smith, Bruce A. Weinberg. 2024. "How to track the economic impact of public investments in AI." *Nature* 630 (June 13): 302-304. https://www.nature.com/articles/d41586-024-01721-1

Lerchenmueller, M. J., & Sorenson, O. (2018). The gender gap in early career transitions in the life sciences. *Research Policy*, *47*(6), 1007-1017.

McKinsey Global Institute. (2024). "The state of AI in early 2024: Gen AI adoption spikes and starts to generate value" McKinsey Global Institute Report. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai#/

Nedungadi, P., Ramesh, M., Govindaraju, V., Rao, B., Berbeglia, P., & Raman, R. (2024). Emerging leaders or persistent gaps? Generative AI research may foster women in STEM. *International Journal of Information Management*, *77*, 102785.

Ngo, D., & Dustan, A. (2024). Preferences, access, and the stem gender gap in centralized high school assignment. *American Economic Journal: Applied Economics*, *16*(4), 257-287.

Sharma, M. (2023). STEM education and gender income parity in USA, 2019. *Social Sciences & Humanities Open*, *8*(1), 100541.

Ross, Matthew B.; Britta Glennon; Raviv Murciano-Goroff; Enrico Berkes; Bruce A. Weinberg; and Julia I. Lane. 2022. "Women are credited less than men in science." *Nature* 608, pages 135–145 (2022). Accelerated publication, June 22. https://doi.org/10.1038/s41586-022-04966-w.

Webb, M. (2019). The impact of artificial intelligence on the labor market. *Available at SSRN 3482150*.

Weinberg, Bruce A. (2020). "Computer Use and the Demand for Female Workers," *Industrial and Labor Relations Review* 53 (No. 2, January): 290-308.

Weinberg, Bruce A., Jason Owen-Smith, Rebecca Rosen, Lou Schwarz, Barbara McFadden Allen, Roy Weiss, Julia Lane. 2014 "Science Funding and Short-Term Economic Activity." *Science* 304 (April 4): 41-43. DOI:10.1126/science.1250055.

Welch, Finis. (2020). "Growth in Women's Relative Wages and in Inequality Among Men: One Phenomenon or Two?" *American Economic Review* 90 (May, No. 2):444-449.

Young, E., Wajcman, J., & Sprejer, L. (2023). Mind the gender gap: Inequalities in the emergent professions of artificial intelligence (AI) and data science. *New Technology, Work and Employment*, *38*(3), 391-414.

Nikolas Zolas, Nathan Goldschlag, Ron Jarmin, Paula Stephan, Jason Owen Smith, Rebecca Rosen, Barbara McFadden Allen, Bruce Weinberg (Corresponding Author) and Julia Lane. 2015. **"**Wrapping it up in a person: Examining employment and earnings outcomes for PhD recipients." *Science* 350 (December 11, No. 6266): 1367-1371. http://www.sciencemag.org/content/350/6266/1367

## Appendix

**Table A.1. Possible Employment Status for Full-Quarter Employee**

| Full-Quarter Employee Employment Status | $t-1$ (2021) | $t$ (2022) | | | | $t+1$ (2023) |
|---|---|---|---|---|---|---|
| **1** | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 |
| **2** | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 |
| **3** | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 |
| **4** | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 |
| **5** | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 |
| **6** | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 |
| **7** | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 |
| **8** | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 |
| **9** | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 |
| **10** | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 |

***Notes*:** The colors in this table represent a worker's employment status. No color means a worker didn't receive any earnings from a firm during a quarter. Light grey and dark grey both represent that a worker had positive earnings from a firm during a quarter. A full-quarter employee needs to work for a firm for at least three consecutive quarters. It's possible that a worker is only partially employed by a firm during the first and the last quarter. Therefore, we only use the middle quarter(s) to calculate earnings.

**Table A.2. Possible Employment Status for Full-Year Employee**

| Full-Year Employee Employment Status | $t-1$ (2021) | $t$ (2022) | | | | $t+1$ (2023) |
|---|---|---|---|---|---|---|
| | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 |

***Notes*:** The colors in this table represent a worker's employment status. No color means a worker didn't receive any earnings from a firm during a quarter. Light grey and dark grey both represent that a worker had positive earnings from a firm during a quarter. A full-year employee must have positive earnings in a firm during all four quarters of a calendar year, as well as the last quarter of the prior year and the first quarter of the following year, accounting for the possibility of partial work during the last and first quarters.