Mapping Labor Market Vulnerability in the Age of AI: Evidence from Job Postings and Patent Data

Eun Cheol Choi[1,3], Qingyu Cao[1], Qi Guan[1], Shengzhu Peng[1], Po-Yuan Chen [2], Luca Luceri [2,3],

[1]Annenberg School for Communication and Journalism, [2]Viterbi School of Engineering, [3]Information Sciences Institute
University of Southern California
{euncheol, qingyuca, qiguan, pengs, poyuan}@usc.edu, lluceri@isi.edu,

Abstract

Emerging AI technologies are reshaping labor markets by potentially altering patterns of job openings, skill requirements, and offered salaries—changes that may disproportionately affect vulnerable populations and occupations at high risk of automation. This study uses a dataset of LinkedIn job postings and AI-related patent abstracts to investigate how AI innovations relate to labor demand. We propose a weighted AI Index (wAII), which measures job openings' exposure to AI by calculating the semantic similarity between job tasks extracted from job descriptions and relevant AI patents. Our analysis identifies industry sectors with high AI exposure—particularly those employing large shares of economically vulnerable workers. We also reveal thematic differences in job postings and show how higher exposure to current AI technologies may be associated with lower wages in some sectors. We validate the robustness of our findings through alternative configurations of task extraction and weighting schemes. Our results highlight AI's potentially uneven and multifaceted relationship with labor markets, underscoring implications for equitable AI and labor policy. We release our code and data to facilitate further research.

Code — https://github.com/EunCheolChoi0123/icwsm25-dwmv-ai-jobs

Introduction

The advent of artificial intelligence (AI) technologies, advanced as never before, has led to significant debate on their impact on labor markets. While automation may drive productivity, concerns persist about job displacement, wage polarization, and shifting skill demands. These shifts may disproportionately affect economically vulnerable workers with limited digital skills or access to retraining programs. A 2024 global survey found that 60% of respondents anticipate AI to transform their job functions within the next five years, while 36% fear it could lead to job displacement during the same period (HAI 2025). Studies indicate AI adoption benefits wages for highly skilled workers (Somjai, Jermsittiparsert, and Chankoson 2020). However, a growing gap may exist between the demand for advanced AI skills and the current workforce's abilities. Supporting this notion, a study of 23 OECD countries found that AI exposure benefits jobs requiring high computer use, leading to employment growth. In contrast, roles with low computer use are more likely to face reduced hours (Georgieff and Hyee 2021).

Given the dynamic nature of labor markets, job openings posted online offer a valuable lens through which to capture shifts in skill demand. Job postings reflect real-time labor market trends and are critical indicators of how workforce needs are evolving (Georgieff and Hyee 2021; Acemoglu et al. 2022; HAI 2025). Understanding these trends is essential for assessing the changing future of work and the broader impact of AI on occupations. Based on these premises, this paper addresses the following research questions:

RQ1: *Which industry sectors are most exposed to AI in job openings?*

RQ2: *What thematic differences exist between AI-highly-exposed and low-exposed job openings?*

RQ3: *How are current AI innovations associated with offered salaries across industries in job postings?*

Our study contributes to understanding the multifaceted relationship between current AI innovations and the labor market by measuring AI exposure in job openings using online job postings, patent data, and large language models. We release our code and data on https://github.com/EunCheolChoi0123/icwsm25-dwmv-ai-jobs.

Related Work

Previous research has assessed occupational exposure to AI technologies. One prominent approach involves breaking down occupations into skills or tasks and linking them to relevant AI areas (Felten, Raj, and Seamans 2021; Frey and Osborne 2017; Georgieff and Hyee 2021). These studies often rely on manually labeled datasets created by researchers or crowd workers using pre-defined categories. However, such annotation with fixed schemes may fail to capture the evolving nature of job markets and openings, where new tasks and skill sets continuously emerge.

An alternative method involves matching job tasks to AI-related patents to assess exposure to AI technologies. For instance, multiple studies explored the relevance of patents to specific job tasks (Montobbio et al. 2024; Autor et al.

2024; Kogan et al. 2023). Building on these studies, researchers proposed the AI Index (AII), which measures the exposure of existing jobs to AI by calculating the semantic textual similarity (STS) between job task descriptions and AI-related patents (Septiandri, Constantinides, and Quercia 2024). The AII aims to identify which jobs are more likely to be automated.

Our study primarily builds on AII (Septiandri, Constantinides, and Quercia 2024) but addresses key limitations by:

- **Capturing Emerging Tasks:** Unlike previous studies that rely on static task databases such as O*NET (updated yearly), our approach uses OpenAI's *GPT-4o-mini* to extract tasks from job descriptions, enabling us to detect emergent, fine-grained work activities beyond traditional occupational taxonomies.
- **Task Weighting:** The original AII assumed equal importance for all tasks within an occupation when calculating AI exposure. However, job openings typically consist of multiple tasks with varying levels of importance. To address this, we introduce weights that assign relative importance to each task, ensuring a more nuanced calculation of AI exposure.

Our weighted AI Index (wAII) approach combines job postings, patent data, and language models to offer a more dynamic and comprehensive way to measure AI exposure in the labor market.

Methodology

We used a LinkedIn job posting dataset and patent abstracts from WIPO to measure the exposure to AI technologies in various job openings. We calculated the wAII by measuring the textual similarity between job tasks extracted from job descriptions and AI-related patents. We then explored how wAII relates to industry sectors, job description topics, and offered salaries.

Dataset

Our study utilizes two datasets to measure the potential exposure to AI technologies in job openings: LinkedIn job postings from 2023 to 2024 and AI-related patent abstracts from 2016 to 2024.

LinkedIn Job Posting Dataset. The LinkedIn job posting dataset (Koneru and Zou 2024) contains essential information such as job descriptions, industry sectors, experience levels, and salaries offered. To achieve a more standardized classification of industry sectors, we mapped the LinkedIn industry codes to the North American Industry Classification System (NAICS) codes, as provided by LinkedIn API reference table (Walia 2023).

For this study, we used the *sectors*, or two front digits of NAICS codes, to classify the industry sectors (Bureau 2025). The NAICS codes comprise 20 distinct sectors, but we dropped two sectors with fewer than 100 job postings to account for outliers and ensure the robustness of our analysis (Agriculture, Forestry, Fishing, and Hunting; Management of Companies and Enterprises). After filtering, the final dataset comprises 109,247 job descriptions, each assigned to one of the 18 distinct NAICS industry sectors.

AI Patents Dataset. We utilized patent data from Lens.org, a patent search engine open to the public. We focused on the patents filed between 2016 and 2024, as these technological innovations may impact job postings in 2023–2024. We filtered the query to include only patents jurisdicted by WIPO. WIPO has identified a set of Cooperative Patent Classification (CPC) codes that correspond to AI technologies (WIPO 2019). We queried the patent data to retrieve relevant records using these CPC codes. After deduplication, the resulting dataset consists of 48,819 AI-related patent abstracts, providing detailed information on AI innovations and technological advancements.

To assess how recent innovations may differ from those in earlier years, we split the patent dataset into two periods: a *past* period (March 2016 to November 2022) and a *current* period (December 2022 to September 2024). This split reflects ChatGPT's release and public adoption as a key marker of increased attention to AI. The *past* set includes 39,005 patent abstracts, while the *current* set includes 9,814.

Weighted AI Index

To quantify the exposure of various occupations to AI technologies, we employed *GPT-4o-mini* (OpenAI 2025) to automatically extract up to five key job tasks from each job description. This automated step was crucial in capturing each role's technical and skill-based expectations at scale. The model was guided with a structured prompt (Figure 1) to focus on *actionable tasks and technical requirements* directly tied to job performance while filtering out soft skills or generic responsibilities. The prompt instructed the model to merge similar tasks into a single representative task to ensure consistency and reduce redundancy.

For each job description, *GPT-4o-mini* generated a structured **JSON array** of tasks, where each task entry included:

- A unique `task_num` identifier based on its first appearance,
- A concise summary of the task emphasizing hard or technical skills,
- A `weight` score (a floating number summing to 1 across all tasks) reflecting the task's relative importance and prominence within the job description.

The AI Index measures the AI exposure of each job opening by computing the textual similarity between job tasks and the nearest AI-related patent abstracts. To achieve this, we embedded both the extracted tasks and patent abstracts using the *nomic-embed-text-v1.5* model, an open-source text embedding model capable of handling extended contexts of up to 8,192 tokens (Nussbaum et al. 2024).

To efficiently identify the nearest patents for each extracted task, we utilized the *Facebook AI Similarity Search* (FAISS) library (Douze et al. 2024). FAISS is a scalable similarity search library that enables fast nearest-neighbor searches in large datasets. Its architecture is optimized for high-dimensional vector searches, making it well-suited for

System	You are an assistant specialized in analyzing job descriptions. Your task is to extract up to 5 key tasks, responsibilities, or requirements essential for the role from the provided job description. Focus only on actionable tasks and skills that are critical to the candidate's performance. Combine similar tasks or requirements into a single task to avoid redundancy. For each task: 1. Assign a 'task_num' starting from 1. This number should reflect the order of the task based on its first appearance in the job description. 2. Provide a 'task' summary focusing on technical or hard skills. 3. Assign a 'weight' (a float up to two decimal places) reflecting the task's importance and relevance within the job description, considering factors such as role focus and key responsibilities. The total should sum to 1. Output the result as a JSON array of dictionaries. Each dictionary should contain: 'task_num': Integer. 'task': String. 'weight': Float.
Input	Job Description: We are seeking a data analyst to join our analytics team. The candidate will be responsible for cleaning and analyzing large datasets using SQL and Python. They should be able to build dashboards in Tableau and present insights to stakeholders. Experience with A/B testing and statistical modeling is a plus. The role also involves collaborating with product managers to define KPIs and track business metrics.
Output	[{"task_num": 1, "task": "Clean and analyze large datasets using SQL and Python.", "weight": 0.35}, ...]

Figure 1: Prompt used to extract and weight core technical tasks from job descriptions, with illustrative examples for input and output

our task of matching job tasks to relevant patents. By clustering similar patents, FAISS reduces the need to iterate through all 48,819 AI-related patents for up to 5 tasks extracted from each of the 109,247 job descriptions, significantly improving computational efficiency when calculating wAII.

The FAISS indexing and retrieval were conducted on a single L4 GPU and configured with the following hyperparameters:

- Number of indices created: 300
- Number of nearest indices to search: 10

For each job opening, we assigned a relative weight to each task (up to five tasks from the job description), multiplied each weight by the textual similarity between the task and its closest AI-related patent, and summed the results to calculate the job opening's wAII. Thus, we define wAII as:

$$\text{wAII}_j = \sum_{i=1}^{n} w_i \cdot \text{sim}(t_i, p_i)$$

Where:

- wAII_j: Weighted AI Index for job opening j.
- w_i: Relative weight assigned to task i.
- t_i: Task i extracted from the job description.
- p_i: Closest AI-related patent to task t_i.
- $\text{sim}(t_i, p_i)$: Textual similarity between task t_i and patent p_i.
- n: Number of tasks considered (up to 5).

This weighted AI index (wAII) builds on the AII proposed by (Septiandri, Constantinides, and Quercia 2024) but introduces two critical innovations. First, instead of relying on a fixed set of manually curated tasks from occupational taxonomies like O*NET, we dynamically generate tasks using a large language model (*GPT-4o-mini*) applied directly to job descriptions. Second, we assign relative importance to each extracted task through model-generated weights, enabling a more nuanced estimation of AI exposure. Together, these improvements allow wAII to reflect work's evolving and varied nature in real-world job postings.

We calculated the index separately for job tasks matched against "past" (March 2016–November 2022) and "current" (December 2022–September 2024) patent corpora to account for shifts in technological innovation. For robustness, we removed outliers based on the 1.5 interquartile range (IQR) rule after computing wAII. After outlier removal, the distribution of wAII remained approximately normal. The overall distribution is illustrated in Figure 2. A paired samples t-test revealed that the average wAII is slightly lower in the current period ($M = 0.72$, $SD = 0.02$) compared to the past period ($M = 0.74$, $SD = 0.02$), $t(109246) = 531.16$, $p < .001$, possibly due to the smaller number of AI-related patents in the current corpus.

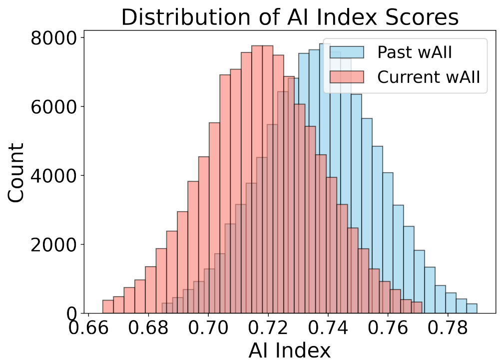

Figure 2: Distribution of past and current weighted AI Index

Results

RQ1: *Which industry sectors are most exposed to AI in job openings?*

To identify and characterize meaningful differences in AI exposure levels across industry sectors, we first conducted a one-way ANOVA for each outcome variable. For the past AI index, the omnibus ANOVA was significant, $F(17, 106906) = 234.69$, $p < .001$, indicating significant differences in AI exposure across NAICS sectors. Similarly, for the current AI index, the ANOVA revealed a significant main effect of sector, $F(17, 106906) = 336.29$, $p < .001$.

We then applied Tukey's Honest Significant Difference (HSD) test to conduct all pairwise post-hoc comparisons while controlling for the family-wise error rate (FWER = 0.05). Each comparison yielded adjusted p-values and 95% confidence intervals, allowing us to determine which pairs of sectors had statistically distinguishable AI exposure means.

To synthesize these pairwise findings into a compact and interpretable structure, we treated sectors as nodes in a directed acyclic graph (DAG), ordering them by decreasing average wAII to reflect their relative exposure levels. A directed edge was drawn from sector i to sector j if $i < j$ and the pairwise comparison between the two indicated a significant difference. We then used dynamic programming to identify the longest possible sequence of sectors in this ordered space where each step corresponded to a statistically significant decrease. Table 1 captures the longest sequence of industry sectors where each step shows a statistically significant decrease in AI exposure.

Figure 3 highlights the industry sectors with the highest and lowest average wAII scores. Notably, job openings in Wholesale, Transportation/Warehousing, Information, and Manufacturing sectors consistently exhibit higher past and current wAII values, indicating greater exposure to AI innovations and/or potential automation. Additionally, Retail Trade shows elevated exposure in the past period, while Health Care and Social Assistance emerge among the top-exposed sectors in the current period. In contrast, sectors such as Education and Real Estate consistently show lower average wAII values, reflecting limited exposure to AI technologies in the context of job openings.

Table 1: Chains of pairwise comparisons for past (top) and current (bottom) AI Index; Tukey's HSD; *: $p < .05$, **: $p < .01$, ***: $p < .001$

Sector i	**Sector j**	**Mean Diff.**
Wholesale Trade	Manufacturing	0.004^{***}
Manufacturing	Retail Trade	0.001^{*}
Retail Trade	Health Care	0.002^{***}
Health Care	Admin. Support	0.001^{**}
Admin. Support	Accommodation	0.003^{***}
Accommodation	Construction	0.002^{***}
Construction	Education	0.002^{***}
Health Care	Information	0.002^{***}
Information	Admin. Support	0.002^{***}
Admin. Support	Prof. Sci. & Tech.	0.003^{***}
Prof. Sci. & Tech.	Accommodation	0.002^{***}
Accommodation	Education	0.003^{***}
Education	Real Estate	0.004^{***}

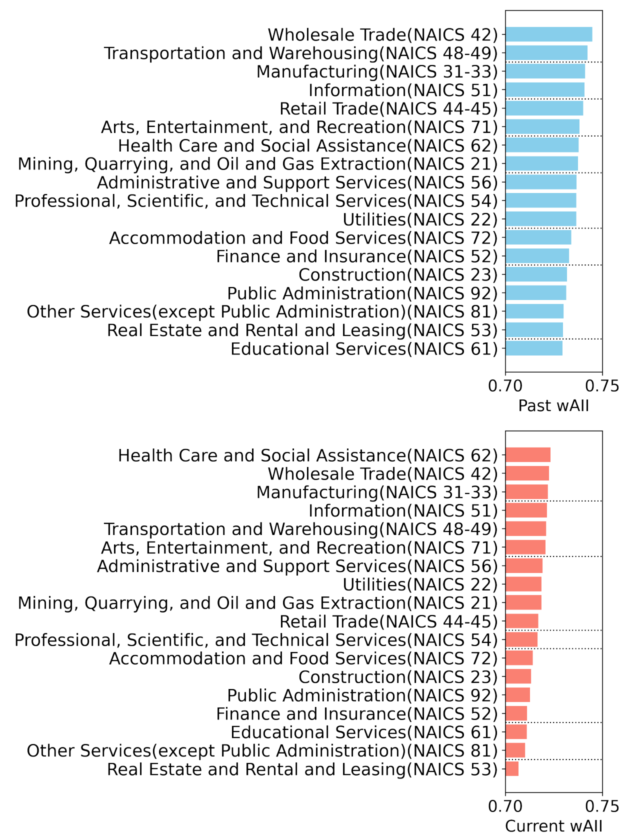

Figure 3: Average weighted AI Index across NAICS sectors—past (top) and current (bottom) periods—ordered by decreasing average wAII. Dotted lines segment sectors where the top of each group differs significantly from the top of the next group, as in Table 1 (Tukey's HSD, $p < .05$).

Topic	Words	Theme
1	construction, legal, labor, claim, civil, administrative, resume, public, human, file, proposal, payroll, agreement, estimate, government, contractor, administration, california, recruitment, vendor	Legal, HR, and public sector roles
2	nursing, nurse, clinical, hospital, therapy, register, treatment, physician, clinic, child, bls, compassionate, assignment, assistant, medication, heart, licensure, referral, accredit, certify	Nursing and clinical healthcare
3	accounting, credit, bank, finance, loan, audit, monthly, banking, investment, transaction, estate, property, reporting, cash, quarterly, executive, consulting, regulatory, capital, invoice	Finance and accounting
4	manufacturing, electrical, plant, mechanical, repair, technician, industrial, energy, manufacture, supplier, environmental, machine, chain, inspection, component, installation, specification, drawing, construction, troubleshoot	Manufacturing and technical repair
5	medium, prospect, digital, territory, revenue, campaign, executive, sell, creative, retail, content, incentive, social, director, insight, empower, partnership, inspire, commission, love	Marketing and sales
6	cloud, user, architecture, integration, developer, infrastructure, data, sql, enterprise, agile, database, web, configuration, server, automation, analytic, python, digital, analyst, hardware	Tech and data engineering
7	clinical, university, laboratory, lab, hospital, appointment, assistant, clinic, medicine, physician, instruction, counseling, collection, collect, technician, teach, college, disease, treatment, academic	Academic healthcare and labs
8	merchandise, vehicle, warehouse, retail, pound, truck, hand, floor, occasionally, walk, supervisor, item, repair, frequently, room, representative, pull, stoop, bend, load	Warehousing and physical labor

Table 2: Topic modeling results for job descriptions.

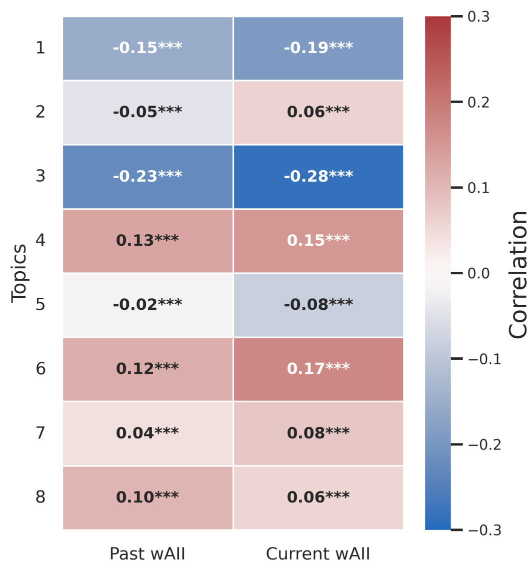

Figure 4: Correlation between the topic probabilities of job descriptions with past and current wAII; ***: p<.001

RQ2: *What thematic differences exist among job openings with high and low levels of AI exposure?*

To discern the thematic differences among job openings with high and low levels of AI exposure, We computed the Pearson correlation between each topic's document-level probability (i.e., its prevalence across job descriptions) and the wAII scores, using *gensim* for topic modeling (Řehůřek and Sojka 2010). We searched for the optimal number of topics (from 3 to 10) in terms of coherence (Röder, Both, and Hinneburg 2015), and then learned asymmetric priors from the corpus for deciding α (Wallach, Mimno, and McCallum 2009). Balancing different metrics suggests eight optimal topics, as shown in Table 3. Topics, top words, and an overarching theme for each topic are illustrated in Table 2.

Table 3: LDA topic model evaluations

N_{topics}	PP	C_v	U_{mass}	C_{uci}	C_{npmi}
3	-7.054	0.427	-2.422	-0.179	0.017
4	-7.142	0.427	**-2.513**	-0.194	0.021
5	-7.179	0.485	-2.346	0.002	0.030
6	-7.198	0.589	-2.094	0.529	0.068
7	-7.241	0.484	-2.318	0.077	0.038
8	-7.236	**0.601**	-2.117	**0.639**	**0.074**
9	-7.271	0.531	-2.196	0.228	0.043
10	**-7.289**	0.570	-2.148	0.513	0.067

Figure 4 presents the correlation between the topic probabilities of job descriptions with past and current wAII. Among the eight job description topics, several show distinct patterns about AI exposure. Topic 3 (Finance and Accounting) exhibits the strongest negative correlation with past and current wAII, suggesting that occupations in this area might be less exposed to AI innovation. Topic 1 (Legal, HR, and Public Sector Roles) also shows a slightly negative correlation, implying limited AI exposure in administrative and compliance-related domains.

In contrast, Topic 6 (Tech and Data Engineering) shows the most positive correlation, underscoring its central role in building and maintaining AI systems and infrastructure. Topic 4 (Manufacturing and Technical Repair) also displays a modest positive correlation, reflecting the growing use of AI in industrial automation and diagnostics. These patterns highlight where AI integration is advancing versus where its relevance remains constrained.

RQ3: *How are current AI innovations associated with offered salaries across industries in job postings?*

Table 4: Linear regression (OLS) results predicting job salary ($\log_{10}$); mean centered (Model 2); *: $p < .05$, **: $p < .01$, ***: $p < .001$

Predictor	**Model 1**	**Model 2**
Past wAII	-2.98***	-2.97***
Current wAII	1.54***	1.55***
Contract Work Type	0.02***	0.02***
Company Size	0.00	0.00
Wholesale Trade	-0.07***	-0.07***
Transport/Warehousing	-0.07***	-0.07***
Information	0.11***	0.11***
Manufacturing	0.01**	0.01**
Health Care/Social Assistance	-0.03***	-0.03***
Retail Trade	-0.15***	-0.15***
C. wAII × Wholesale Trade		-2.28**
C. wAII × Transport/Warehouse		-1.19**
C. wAII × Information		0.95***
C. wAII × Manufacturing		0.06
C. wAII × HC/SA		-0.02
C. wAII × Retail Trade		0.34
Intercept	6.05***	4.96***
Adjusted R²	.079	.080
F-statistic	318.9***	202.4***
Number of Observations	36,821	

To analyze how AI exposure relates to salary levels across job openings, we first transformed the salary data to make it suitable for statistical modeling. Salary was $\log_{10}$-transformed to normalize skewed values ($M = 4.964$, $SD = 0.217$). To ensure comparability, we standardized all salary types by converting hourly, weekly, biweekly, and monthly wages into annual salaries using conventional multipliers (e.g., 40 hours/week × 52 weeks/year for hourly wages). We then filtered the dataset to include only jobs with annual salaries between $15,080 (the federal minimum) and $520,000 (equivalent to $250/hour) to remove extreme outliers. As postings typically provided salary ranges, we used the midpoint between the reported minimum and maximum values. After removing entries with missing data, the final analytical sample included 36,821 job openings.

The regression results in Table 4 show a nuanced relationship between current AI exposure and salary. To isolate the association between current AI exposure and salary better, we controlled for several confounding variables, including past wAII, contract type (contract vs. full-time), and company size (based on reported employer metadata). We also

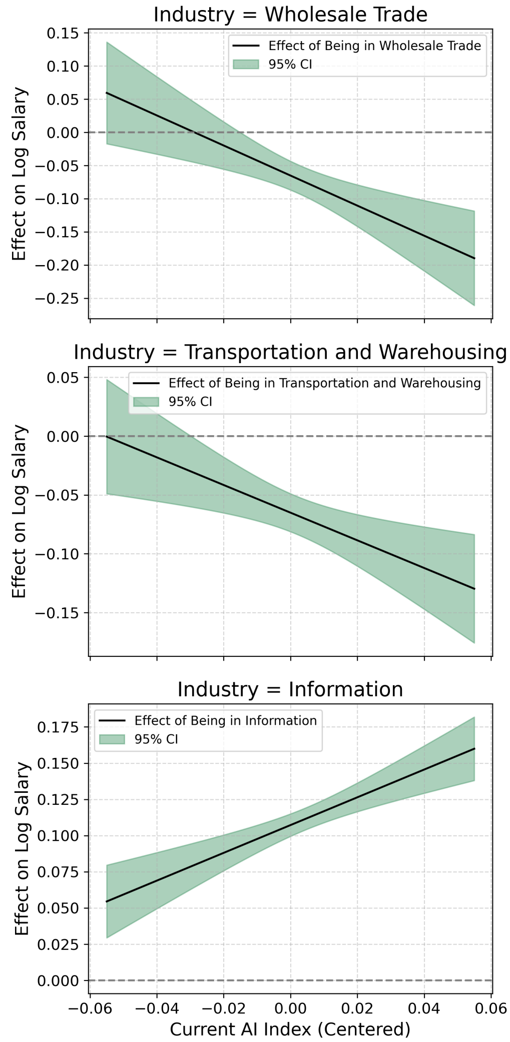

Figure 5: Johnson-Neyman Plot for three significant interaction terms in Table 4

add several industry sectors showing high levels of past/current wAII in Figure 3 as dummy variables. Analysis shows that past AI exposure (past wAII) is negatively associated with salary, whereas current AI exposure (current wAII) is positively associated with salary overall. This contrast suggests that compared to earlier AI-related transformations, current AI adoption is on average more associated with competitive compensation—possibly due to demand for newly emerging skill sets.

However, the relationship between AI exposure and salary varies substantially by industry. Several interaction terms between current wAII and industry dummies are statistically significant, indicating that the association is not uniform across sectors. Specifically, the interaction coefficients for current wAII were negative for Wholesale Trade ($B = -2.28$, $p < .01$) and Transportation/Warehousing ($B = -1.19$, $p < .01$), and positive for the Information sector

($B = 0.95$, $p < .001$), indicating sector-specific salary patterns. For instance, the negative interaction in Wholesale Trade suggests that jobs with higher AI exposure in this sector tend to offer lower salaries.

This pattern is visualized in the Johnson-Neyman plot (Figure 5), which illustrates how the relationship between industry affiliation and log salary varies across levels of current wAII. For sectors like Wholesale Trade and Transportation and Warehousing, the slope becomes significantly negative only at higher levels of AI exposure while remaining statistically non-significant at lower exposure levels. This may imply the emergence of a salary penalty in high-AI-exposure contexts. In contrast, the Information sector shows the opposite pattern: the relationship between industry and salary is non-significant at low exposure levels but becomes significantly positive as wAII increases—suggesting that workers in this sector may benefit more from increased AI exposure.

Robustness Check. To ensure the robustness of our results, we experimented with variations in task extraction and weighting settings. Specifically, we tested scenarios involving extracting up to seven tasks per job description and an alternative weighting scheme. The results from different configurations are detailed in this section.

- **Number of tasks extracted**: We tested extracting either up to *5* or *7 tasks* per job description.
- **Relative weighting methods**: In addition to using *GPT-4o-mini*-generated weights, we tried an alternative weighting method based on the textual similarity between the job opening description and each extracted task. The resulting cosine similarity scores were normalized using a softmax function to create a probability distribution summing to 1, and used as weights.

We evaluated four configurations:

- **Config 1 (Main)**: 5 tasks with weights generated by *GPT-4o-mini*.
- **Config 2**: 7 tasks with weights generated by *GPT-4o-mini*.
- **Config 3**: 5 tasks with softmax weights.
- **Config 4**: 7 tasks with softmax weights.

The wAII from different configurations remained highly correlated with one another (Figure 6). Although specific numerical results varied across configurations, the overall trends and insights remained consistent at the aggregate level (Table 5 and Figure 7).

Discussion

Our analysis revealed significant sectoral differences in AI exposure. These differences offer key insights into the ongoing AI-driven transformation of workforce demand.

A high AI index may suggest a trend toward automation or AI assistance within certain industries (Muro, Maxim, and Whiton 2019; GAO 2019):

- **Wholesale**: Companies in this sector leverage AI for predictive analytics and optimize inventory management (Albayrak Ünal, Erkayman, and Usanmaz 2023).

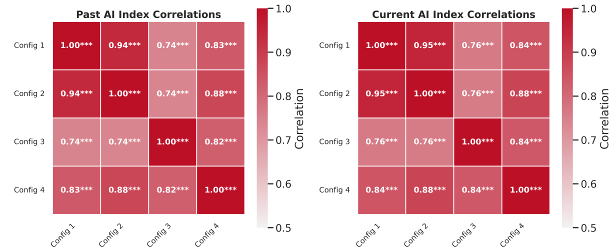

Figure 6: Correlation among weighted AI Index for different configurations; ***: p<.001

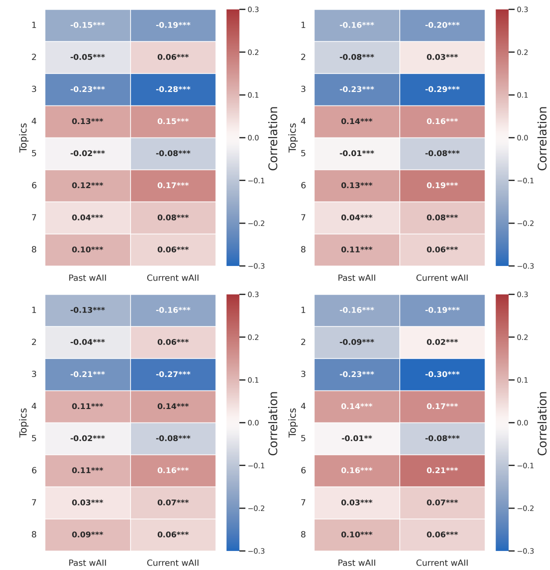

Figure 7: Correlation between the topic probabilities of job descriptions with past and current wAII; Configuration 1-4; ***: p<.001

- **Transportation and Warehousing**: Self-driving and assisted driving technologies are being developed to improve logistics efficiency (Mirindi 2024). AI is also used for route optimization, real-time tracking, and predictive maintenance to reduce operational costs and improve service quality (Mallouk, Sallez, and Abou El Majd 2021).
- **Manufacturing**: The sector increasingly relies on automation and robotics to enhance production processes. This shift towards automation and robotics is a key component of smart manufacturing, and Industry 4.0 (Evjemo et al. 2020).
- **Information**: This sector plays a role in developing and deploying AI innovations across other industries.
- **Retail**: AI-powered chatbots and virtual assistants are used to improve customer service (Leung and Yan Chan 2020). At the same time, predictive analytics are applied for in-store operations, such as shelf management and loss prevention (Sekhar 2022).

Topic modeling analysis reveals notable thematic contrasts between job openings with high and low AI exposure.

Table 5: Linear regression (OLS) results predicting job salary ($\log_{10}$); Configuration 2–4; mean centered; *: $p < .05$, **: $p < .01$, ***: $p < .001$

Predictor	Config 2	Config 3	Config 4
wAII	-2.96***	-3.34***	-3.37***
Current wAII	1.59***	2.02***	2.01***
Contract Work Type	0.01***	0.01***	0.01***
Company Size	0.00	0.00	0.00
Wholesale	-0.07***	-0.06***	-0.06***
Tran./Ware.	-0.07***	-0.07***	-0.07***
Information	0.11***	0.10***	0.10***
Manufacture	0.01**	0.01*	0.01*
Health/Social Assist.	-0.03***	-0.03***	-0.04***
Retail	-0.15***	-0.15***	-0.15***
wAII × Wholesale	-2.43***	-2.64***	-2.94***
wAII × Tran./Ware.	-1.02*	-0.91*	-0.89*
wAII × Info.	1.09***	0.84***	1.04***
wAII × Manu.	0.10	0.10	0.02
wAII × Health/Social	-0.30	0.04	-0.09
wAII × Retail	0.37	0.42	0.64
Intercept	4.96***	4.96***	4.96***
Adjusted R^2	.079	.078	.077
F-stat	197.9***	194.0***	190.1***
No. of Obs.	36,821	36,415	

Topics more strongly associated with higher wAII—such as Tech and Data Engineering and Manufacturing and Technical Repair—tend to involve roles foundational to AI development or amenable to automation. In contrast, topics correlated with lower wAII—such as Finance and Accounting and Legal, HR, and Public Sector Roles—emphasize more procedural, regulated, or human-centric tasks. These patterns suggest that AI exposure varies meaningfully across job types, aligning with the roles' technical intensity and/or automation potential.

Our regression analysis indicates a sector-specific relationship between AI exposure and offered salaries. Sectors such as Wholesale and Transportation/Warehousing are associated negatively with salary when coupled with higher AI exposure, likely reflecting the automation of routine tasks and competition between AI resources and workforces, which are reflected in lower salaries offered to job applicants. Our analysis suggests that sectors experiencing wage suppression and high AI exposure—such as Wholesale and Transportation—might also be those employing large shares of low-income and frontline workers (Filippi, Bannò, and Trento 2023). This intersection indicates a potential amplification of socioeconomic inequality due to technological disruption. In contrast, the Information sector exhibits a steeper positive slope when coupled with high current wAII, indicating that current AI innovations may benefit the workers in this sector.

These findings carry implications for economically vulnerable populations. Sectors such as wholesale, warehousing, and transportation—associated with high AI exposure and downward salary trends—employ a disproportionate share of low-income, frontline workers. This raises concern that AI-driven displacement or wage compression may exacerbate labor inequalities. While AI innovations offer efficiency gains, they may also deepen structural disadvantages unless supported by targeted policy interventions in upskilling, job redesign, and labor protections.

Our study has multiple contributions. First, our work contributes to understanding the exposure of AI technologies in the labor market by providing a refined measure of wAII. The use of FAISS for efficient similarity search ensured the scalability of our analysis. Second, we demonstrated the robustness of our findings by experimenting with different configurations. Finally, our findings highlight the need to address the uneven exposure of AI across sectors. The negative salary associations observed in specific sectors may suggest potential wage suppression due to automation.

Limitation. Although our study demonstrated the robustness of the weighted AI index (wAII) across various numbers of tasks and weighting schemes, further experimentation with different generative and embedding models is necessary to overcome potential biases from individual model architectures. The models of choice should support long-context inputs, as LinkedIn job descriptions (max 3,400 words in our data) and patent abstracts (max 726 words) tend to be lengthy. Moreover, while semantic similarity provides a scalable approach to estimate AI exposure, it may not fully capture the functional relevance between job tasks and AI capabilities described in patents, potentially leading to under- or overestimation of exposure in certain domains. Future iterations of wAII could benefit from incorporating knowledge graphs, annotated datasets, or human-in-the-loop validation to improve relevance estimation. Additionally, relying solely on LinkedIn postings introduces systemic bias, as it may not fully represent the labor market across all industries or regions. Jobs not posted online—often in small businesses or filled via union hiring halls—remain unaccounted for (HAI 2025). Finally, although wAII builds upon the original AI Index (AII) by introducing task extraction and differential weighting, further work is needed to more explicitly articulate its added value over the original AII approach. In particular, future analyses should empirically contrast the two indices across sectors to clarify whether the weighting reveals patterns of exposure that the unweighted AII may miss. To deepen our understanding of AI's labor impact, future studies should incorporate methods such as experiments, interviews, and participatory observation.

Ethical Statement. The job posting and patent data used in this study were collected from publicly accessible platforms, minimizing ethical concerns related to data privacy and consent. Because the data is anonymized and aggregated, we consider the likelihood of direct harm to individuals low. Nonetheless, we recognize the potential for misinterpretation or bias and have presented our findings with appropriate context and limitations.

While our analysis does not include individual-level demographic data, it indirectly surfaces structural inequalities—particularly in sectors highly exposed to AI that disproportionately employ vulnerable or precarious workers. These findings underscore the need for future research to incorporate qualitative methods and disaggregated demographic data to capture the lived experiences and social dynamics behind these labor market shifts. Although some degree of misclassification may exist due to potential measurement error in our weighted AI Index (wAII), our robustness checks suggest that such error has minimal influence on the overall results.

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.

Albayrak Ünal, Ö.; Erkayman, B.; and Usanmaz, B. 2023. Applications of artificial intelligence in inventory management: A systematic review of the literature. *Archives of Computational Methods in Engineering*, 30(4): 2605–2625.

Autor, D.; Chin, C.; Salomons, A.; and Seegmiller, B. 2024. New frontiers: The origins and content of new work, 1940–2018. *The Quarterly Journal of Economics*, qjae008.

Bureau, U. C. 2025. North American Industry Classification System (NAICS). Accessed: 2025-01-14.

Douze, M.; Guzhva, A.; Deng, C.; Johnson, J.; Szilvasy, G.; Mazaré, P.-E.; Lomeli, M.; Hosseini, L.; and Jégou, H. 2024. The Faiss library.

Evjemo, L. D.; Gjerstad, T.; Grøtli, E. I.; and Sziebig, G. 2020. Trends in smart manufacturing: Role of humans and industrial robots in smart factories. *Current Robotics Reports*, 1: 35–41.

Felten, E.; Raj, M.; and Seamans, R. 2021. Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, 42(12): 2195–2217.

Filippi, E.; Bannò, M.; and Trento, S. 2023. Automation technologies and their impact on employment: A review, synthesis and future research agenda. *Technological Forecasting and Social Change*, 191: 122448.

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/.

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.

GAO. 2019. Emerging Technologies: Information on Workforce Automation. Technical Report GAO-19-257.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Georgieff, A.; and Hyee, R. 2021. Artificial intelligence and employment: New cross-country evidence. Technical report, OECD.

HAI. 2025. The 2025 AI Index Report.

Kogan, L.; Papanikolaou, D.; Schmidt, L. D.; and Seegmiller, B. 2023. Technology and labor displacement: Evidence from linking patents with worker-level data. Technical report, National Bureau of Economic Research.

Koneru, A.; and Zou, Z. Y. 2024. Linkedin job postings (2023 - 2024). https://www.kaggle.com/datasets/arshkon/linkedin-job-postings.

Leung, C. H.; and Yan Chan, W. T. 2020. Retail chatbots: The challenges and opportunities of conversational commerce. *Journal of Digital & Social Media Marketing*, 8(1): 68–84.

Mallouk, I.; Sallez, Y.; and Abou El Majd, B. 2021. Machine learning approach for predictive maintenance of transport systems. In *2021 Third International Conference on Transportation and Smart Technologies (TST)*, 96–100. IEEE.

Mirindi, D. 2024. A Review of the Advances in Artificial Intelligence in Transportation System Development.

Montobbio, F.; Staccioli, J.; Virgillito, M. E.; and Vivarelli, M. 2024. Labour-saving automation: A direct measure of occupational exposure. *The World Economy*, 47(1): 332–361.

Muro, M.; Maxim, R.; and Whiton, J. 2019. Automation and Artificial Intelligence: How Machines Are Affecting People and Places. Technical report, Brookings Institution.

Nussbaum, Z.; Morris, J. X.; Duderstadt, B.; and Mulyar, A. 2024. Nomic embed: Training a reproducible long context text embedder. *arXiv preprint arXiv:2402.01613*.

OpenAI. 2025. GPT-4o Mini: Advancing Cost-Efficient Intelligence. Accessed: 2025-01-14.

Řehůřek, R.; and Sojka, P. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50. Valletta, Malta: ELRA. http://is.muni.cz/publication/884893/en.

Röder, M.; Both, A.; and Hinneburg, A. 2015. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, 399–408.

Sekhar, C. 2022. Optimizing Retail Inventory Management with AI: A Predictive Approach to Demand Forecasting, Stock Optimization, and Automated Reordering. *European Journal of Advances in Engineering and Technology*, 9(11): 89–94.

Septiandri, A. A.; Constantinides, M.; and Quercia, D. 2024. The potential impact of AI innovations on US occupations. *PNAS nexus*, 3(9): 1–13.

Somjai, S.; Jermsittiparsert, K.; and Chankoson, T. 2020. Determining the initial and subsequent impact of artificial intelligence adoption on economy: a macroeconomic survey from ASEAN. *Journal of Intelligent & Fuzzy Systems*, 39(4): 5459–5474.

Walia, A. K. 2023. LinkedIn Industry Codes v2 (NAICS). https://learn.microsoft.com/en-us/linkedin/shared/references/reference-tables/industry-codes-v2-naics".

Wallach, H.; Mimno, D.; and McCallum, A. 2009. Rethinking LDA: Why priors matter. *Advances in neural information processing systems*, 22.

WIPO. 2019. PATENTSCOPE Artificial Intelligence Index. https://www.wipo.int/tech_trends/en/artificial_intelligence/patentscope.html.