Proceedings of the Eighth AAAI/ACM Conference on AI, Ethics, and Society (AIES 2025)

AI-OCI: A Novel Framework for Assessing AI's Workforce Impact Using LLMs

Frederick Awuah-Gyasi, Trilce Estrada

Department of Computer Science, University of New Mexico
fawuahgyasi5@unm.edu, trilce@unm.edu

Abstract

We introduce the AI Occupational Capability Index (AI-OCI), a novel methodology for quantifying the alignment between AI model capabilities and the tasks that define human occupations. Unlike prior automation risk metrics, which rely on expert heuristics or job-level generalizations, AI-OCI operates at the task level by embedding and comparing over 19,000 occupational tasks with 338 AI capabilities using state-of-the-art language models. The resulting scores reveal how well AI systems can perform specific human functions, enabling interpretable, task-aligned assessments of labor exposure. Empirical evaluations show strong correlations with benchmark indices such as AIOE and GPT-4 Beta exposure scores, while diverging from legacy automation risk measures. We demonstrate AI-OCI's utility through case-based analyses of employment and wage shifts across high-alignment occupations during the era of large language model adoption. The framework supports scalable, real-time tracking of AI's workforce impact and provides a foundation for integrating labor intelligence into education, policy, and economic planning.

Code — https://github.com/fawuahgyasi/AI-OCI

1 Introduction

Artificial intelligence (AI) is fundamentally reshaping the labor market, yet current methods fail to precisely and dynamically measure its impact. Policymakers, researchers, and businesses require clear insights into which occupations face the greatest exposure—not only to assess automation risks but to strategically guide workforce development, reskilling initiatives, and policy interventions. To address this need, researchers have introduced various AI exposure indices. Early efforts, such as Frey and Osborne's automation risk model (Frey and Osborne 2017), estimate the likelihood of automation at the occupation level. More recent indices, including the AI Occupational Exposure (AIOE) (Felten, Raj, and Seamans 2019) and GPT-generated AI Exposure (GENOE) (Benítez-Rueda and Parrado 2024), use expert surveys or heuristic scoring methods. However, these AI exposure assessments rely predominantly on human analysis or on static, occupation-level classifications, inadequately capturing the nuanced ways in which AI increasingly augments, rather than replaces, human labor.

Despite valuable contributions, these methodologies suffer critical limitations. Occupation-level assessments overlook significant task-specific variations within jobs, resulting in generalized and potentially misleading conclusions. Expert-driven surveys introduce subjectivity and scalability challenges, hindering timely updates as AI rapidly evolves. Approaches based on heuristic methods or job postings provide indirect approximations of AI capability rather than direct, measurable alignment with job tasks. Consequently, existing indices quickly lose relevance and cannot dynamically track how occupations integrate evolving AI technologies.

In this paper we propose the AI Occupational Capability Index (AI-OCI), a dynamic, task-level framework that quantifies AI's real-time alignment with occupational tasks. AI-OCI leverages NLP-based text embeddings and clustering techniques to directly measure semantic similarities between specific AI capabilities and individual job tasks. Our method continuously adapts as AI technologies advance, providing stakeholders with a precise, scalable, and dynamic assessment of AI's impact on the workforce.

AI-OCI is model-agnostic, allowing flexibility in selecting, refining, or updating NLP embedding models as technology evolves. Unlike high-level heuristics or subjective surveys, AI-OCI emphasizes explainability, enabling stakeholders to objectively pinpoint the exact reasoning behind each score. This transparency ensures accuracy and enhanced objectivity, making our approach suitable for real-time tracking of rapid AI advancements.

Contributions. This study makes four primary contributions:

- We introduce AI-OCI, a dynamic, task-level AI exposure framework leveraging NLP embeddings to quantify AI's alignment with occupational tasks without relying on subjective expert judgments.
- We propose a precise metric able to quantify AI's capability at both the individual task and aggregate occupation levels, enabling detailed and actionable workforce impact assessments.
- We evaluate AI-OCI against established benchmarks, including AIOE (Felten, Raj, and Seamans 2019), GENOE

(Benítez-Rueda and Parrado 2024), Frey's Automation Risk (Frey and Osborne 2017), OpenAI GPT-4 Exposure (Eloundou et al. 2023), and Webb's AI Score (Webb 2019), demonstrating comparable accuracy and superior dynamic adaptability.

- We analyze correlations between AI-OCI and economic indicators such as wages, employment trends, and workforce transitions, providing actionable guidance for policymakers and business leaders managing AI-driven labor market shifts.

The remainder of the paper is organized as follows. In Section 3, we introduce the AI-OCI framework, detailing our methodology for generating embeddings, clustering AI capabilities, computing similarity scores, and aggregating results at the occupational level. Section 4 presents our empirical findings, including benchmark validations, economic correlation analyses, occupational rankings, and case-based trajectory assessments. Section 5 discusses current limitations of the framework and outlines directions for future research. Finally, Section 6 concludes with a summary of contributions and potential applications of AI-OCI in labor market forecasting and education-to-workforce alignment.

2 Related Work

The rapid integration of AI into the workforce has driven substantial research on automation and labor displacement. However, existing methods often rely on static occupation-level predictions, heuristic scoring, or expert assessments, limiting their adaptability to evolving AI capabilities. AI-OCI significantly advances these approaches by introducing a dynamic, task-level framework that leverages text embeddings to quantify AI-job alignment with greater precision.

Occupation-Level Automation Risk Assessments. Frey and Osborne (Frey and Osborne 2017) pioneered one of the first large-scale studies on occupational automation risk. Using a Gaussian Process Classifier, they estimated the probability of automation across 702 occupations, categorizing jobs as *"susceptible" or "not susceptible"* to automation. While their study laid the groundwork for AI-driven labor analysis, it relied on binary classification and did not account for task-specific AI augmentation, limiting its relevance in an era where AI increasingly assists rather than fully replaces human labor. Webb (2019) (Webb 2019) introduced a novel approach to assessing AI's occupational impact by examining textual similarities between job descriptions and AI patents. By linking patent filings with occupational text data, Webb's work offers a unique patent-task matching methodology. While this approach captures AI-driven technological advancements, it does not directly quantify AI's ability to perform occupational tasks. AI-OCI complements Webb's framework by shifting from patent-based task exposure to embedding-based AI-task alignment, providing a direct AI capability-task assessment. Cazzaniga et al. (2024) (Cazzaniga et al. 2024) extended this research by examining the macroeconomic implications of generative AI on workforce transitions, highlighting the shift in labor demand due to task automation and augmentation. However, their study remains focused on broad labor trends rather than fine-grained occupational shifts. AI-OCI addresses this gap by providing a structured, task-level framework for evaluating AI-task alignment, enabling a more granular analysis of AI's role in augmenting, rather than replacing, human workers. Recent findings from the AI Index Report 2024 (Stanford University 2024) reinforce the importance of task-level assessments, showing that automation risk varies significantly across industries. Specifically, high-skill occupations are experiencing AI augmentation rather than outright job displacement, whereas lower-skill manual labor roles remain more resistant to automation due to their reliance on dexterity and contextual human judgment. These insights validate AI-OCI's approach, which models AI's gradual integration into the workforce as a continuum rather than a binary shift. Unlike previous work that categorizes jobs as "at-risk" or "safe," AI-OCI enables continuous tracking of AI impact at the task level, making it more adaptive to evolving AI capabilities.

Expert Survey-Based AI Impact Assessments. Felten et al. (Felten, Raj, and Seamans 2019, 2021, 2018) introduced the AI Occupational Exposure (AIOE) metric, which estimates AI's impact on jobs through expert surveys. While useful, this methodology relies on subjective human assessments and does not dynamically adjust as AI capabilities evolve. Additionally, AIOE does not directly compute AI-task similarity but rather infers exposure based on broad occupational categories. Benítez-Rueda and Parrado (2024) (Benítez-Rueda and Parrado 2024) introduced the GENOE index, leveraging synthetic AI surveys with GPT models to estimate AI exposure across occupations. While this approach enables large-scale estimation, it remains dependent on LLM-generated judgments, which may introduce bias and lack interpretability. AI-OCI overcomes these limitations by using direct embedding-based similarity computations without requiring subjective survey responses.

Machine Learning Suitability and Job Automation. Brynjolfsson et al. (Brynjolfsson and Mitchell 2017) introduced the "Suitable for Machine Learning" (SML) index, an early attempt to assess job automation potential by examining task characteristics that could be performed by traditional ML models. While influential, this approach predates modern LLMs, making it less applicable to text-based AI capabilities such as ChatGPT and GPT-4. Unlike SML, AI-OCI directly quantifies AI-task alignment using embedding-based similarity, allowing for a more precise assessment of LLM-driven workforce transformation. Goldfarb et al. (Goldfarb, Taska, and Teodoridis 2022) extended this research by analyzing AI-related job postings, identifying shifts in AI skill demand. While this provides useful economic insights, it lacks direct AI-task matching, making it less precise than AI-OCI's dynamic embedding-based approach. The AI Index Report 2024 (Stanford University 2024) provides empirical evidence that AI adoption is driving employment shifts in high-skill occupations while reducing demand for repetitive, non-cognitive jobs. This supports AI-OCI's emphasis on task granularity, as LLMs and multimodal AI are increasingly integrated into professional workflows rather than simply automating predefined occupational categories. Wang et al. (2024) (Wang et al. 2024) advanced embedding optimization techniques for task-based

Authors	Year	Methodology	Key Findings	Evaluation Metric
Frey and Osborne (Frey and Osborne 2017)	2017	Gaussian Process Classifier	Estimated automation risk across occupations	Binary Classification
Brynjolfsson et al. (Brynjolfsson and Mitchell 2017)	2017	"Suitable for Machine Learning" (SML) index	ML's potential to transform workplace tasks	Task-Level Heuristic Scoring
Webb (Webb 2019)	2019	Patent-Task Matching	Uses AI patent data to estimate job task exposure	Patent-Based Exposure
Felten et al. (Felten, Raj, and Seamans 2019)	2019	AI Occupational Exposure (AIOE)	Differential AI impact across tasks	Expert Assessment
Tolan et al. (Tolan et al. 2020)	2020	Task-Cognitive-AI Mapping	Links tasks to cognitive benchmarks	Cognitive-AI Mapping
Goldfarb et al. (Goldfarb, Taska, and Teodoridis 2022)	2022	AI job posting analysis	Identifies shifts in AI skill demand	Job Posting Trends
Eloundou et al. (Eloundou et al. 2023)	2023	LLM labor task mapping	LLMs' impact on tasks	Heuristic Task Match
Pizzinelli et al. (Pizzinelli, Borup, and Fernandez-Cornejo 2023)	2023	Cross-country AI exposure	Measures substitution vs complementarity	Substitution vs Complementarity
MIT CSAIL (Thompson et al. 2024)	2024	Constraints on automation	Economic and technical barriers	Occupational AI Estimates
Cazzaniga et al. (Cazzaniga et al. 2024)	2024	IMF discussion on GenAI	Macroeconomic implications	Labor and Macro Impact
Wang et al. (Wang et al. 2024)	2024	Embedding similarity scoring	Task alignment via embeddings	Cosine Similarity
Benítez-Rueda et al. (Benítez-Rueda and Parrado 2024)	2024	GENOE (GPT-generated)	AI exposure from synthetic surveys	GPT-based Index
Colombo et al. (Colombo et al. 2024)	2024	TEAI index with LLMs	Multi-LLM consensus task evaluation	Consensus-Based Scores

Table 1: Comparative analysis of related works and their evaluation metrics.

AI evaluation, further reinforcing the effectiveness of text embeddings for workforce impact analysis. Their findings highlight that embedding-based task alignment significantly enhances AI capability assessments, aligning closely with AI-OCI's methodology. This suggests that embedding-based AI exposure indices, such as AI-OCI, will play a critical role in future labor market research.

Embedding-Based AI Task Analysis and LLM Workforce Impact. Eloundou et al. (Eloundou et al. 2023) assessed LLMs' labor impact by using heuristic-based task assessments. While valuable, their approach relies on expert annotations rather than computational embeddings, making it less scalable. Similarly, the MIT CSAIL (2024) study (Thompson et al. 2024) examined automation constraints but focused on computer vision tasks rather than text-based workforce transformation, limiting its applicability to LLMs. Colombo et al. (2024) (Colombo et al. 2024) introduced the Task Exposure to AI Index (TEAI), leveraging multiple open-source LLMs to evaluate occupational tasks. Their study provides a structured ranking of AI-task exposure but does not incorporate structured clustering of AI capabilities. Unlike TEAI, AI-OCI introduces a dynamic, task-specific framework by clustering AI capabilities and computing fine-grained AI-task alignment scores, making it more interpretable and adaptable. The IMF Discussion Paper (2024) (Cazzaniga et al. 2024) further reinforced the economic impact of AI on jobs, outlining key policy considerations. While their study provides a macroeconomic perspective, AI-OCI offers task-specific insights, allowing for a more granular understanding of how AI aligns with occupational duties.

Advancing AI-Labor Alignment with AI-OCI. AI-OCI builds upon these foundational studies by introducing the first dynamic, embedding-based framework for AI-labor market analysis. Unlike occupation-wide assessments, AI-OCI computes direct AI-task alignments, offering a higher degree of granularity. By leveraging state-of-the-art NLP models such as Text-Embedding-3-Large, AI-OCI provides a scalable, adaptable, and fine-grained workforce impact assessment. The structured clustering approach allows AI capabilities to be grouped functionally before alignment calculations, enabling a more interpretable framework for policymakers and researchers. Overall, AI-OCI represents a paradigm shift from static AI exposure metrics to a continuously evolving and dynamic model. With increasing AI adoption, embedding-based methodologies such as AI-OCI will become indispensable in tracking workforce transitions and guiding AI policy interventions.

3 Methodology

The AI Occupational Capability Index (AI-OCI) quantifies AI's ability to perform occupational tasks by computing the semantic similarity between occupational task descriptions and AI capability descriptions. The framework consists of five key stages: 1) Embedding Generation: converts textual descriptions of occupational tasks and AI capabilities into numerical vector embeddings. These embeddings preserve the semantic meaning and allow for quantitative comparisons between textual descriptions; 2) AI capability clustering: organizes AI capabilities into distinct groups based on their functional similarities, which simplifies subsequent similarity computations; 3) AI-task similarity computation: each occupational task is assigned to the most relevant AI capability cluster based on semantic similarity; then it computes the exact similarity between each task and individual AI capability within the assigned cluster, identifying the closest AI-task alignments; 4) AI-OCI score computation: task-level similarity scores are aggregated by occupation, providing a concise and interpretable measure of AI alignment; and 5) Occupation-level aggregation: ranks occupations based on their computed AI-OCI scores. This allows policymakers and stakeholders to readily compare occupations, assess AI's workforce implications, and identify occupations most aligned with current AI capabilities.

Algorithm 1: AI-OCI Computation

Input:
Occupational task descriptions $OT = \{t_1, t_2, ..., t_n\}$
AI capability descriptions $C = \{c_1, c_2, ..., c_m\}$
Step 1: Use an embedding model $f()$ to map textual descriptions into vector embeddings $f(t_i) \to z_{ti} \in R^d$ and $f(c_j) \to z_{cj} \in R^d$ where $OT \to Z_{OT}$ and $C \to Z_C$
Step 2: Group AI capability embeddings Z_C into k sets of embeddings $ZS = \{ZS_1, ZS_2, ..., ZS_k\}$ using k-means, where k is selected via silhouette score.
Step 3: For each $ZS_i = \{z_{cp}, z_{cq}, ...z_{cr}\} \in ZS$ build its corresponding mapping set $S_i = \{c_p, c_q, ...c_r\}$ in the textual space.
Step 4: For each cluster S_i, compute a *cluster embedding* z_{Si} by concatenating all AI capability textual descriptions assigned to S_i and generating a global embedding

$$z_{Si} \leftarrow f(cat(c_p, c_q) | \forall c_p, c_q \in S_i, p < q)$$

Step 5: For each occupational task $z_{tj} \in Z_{OT}$ assign it to the most similar AI capability cluster AI_{Si} by finding the cluster embedding z_{Si} that maximizes its cosine similarity:

$$AI_{Si} \cup \{z_{tj} \textit{ iff } \arg\max_{Si} \cos(z_{tj}, z_{Si}) | \forall z_{tj} \in Z_{OT}\}$$

Step 6: For all tasks within an occupation; compute the maximum similarity score between the task embedding and each AI capability within its assigned cluster:

$$score_{tj} \leftarrow \max(\cos(z_{tj}, z_{cp})) | \forall z_{cp} \in ZS_i$$

Step 7: Compute the AI-OCI score for the occupation OT by aggregating its individual task scores:

$$\text{AI-OCI}_{OT} = \frac{1}{n} \sum_{j=1}^{n} score_{tj}$$

Output: $AI - OCI_{OT}$ scores for occupation OT.

Embedding Generation

In order to build a semantic embedding space encompassing task-level descriptors, we begin by constructing two distinct corpora: one representing occupational tasks and another representing AI capabilities. We extracted occupational task descriptions from the *O*NET* database, which provides standardized, task-level documentation for a wide range of occupations. These tasks reflect domain-relevant skills and responsibilities in clear natural language. In addition, we generated AI capability descriptions using outputs from large language models, including *ChatGPT* and *Gemini*. The resulting dataset includes AI functionalities that span natural language processing, reasoning, problem solving, and perception-based capabilities. We applied semantic similarity checks to remove redundancies and ensure that the final capability set was representative and diverse.

Given the two corpora, we convert occupational textual task descriptions and textual AI capability descriptions from natural language into numerical representations using a pre-trained NLP embedding model. We evaluated multiple embedding models for this task, including well-established transformer-based architectures. Each model was assessed on its ability to preserve semantic distinctions and support clustering. After experimentation, we selected *Text-Embedding-3-Large* for its ability in capturing finer task granularity and producing coherent capability clusters.

Formally, our method maps textual data x into a high dimensional vector space $f(x) \in R^d$. The goal being exposing semantic relationships between human and AI skills associated with higher order tasks. Given a list of occupational task descriptions $OT = \{t_1, t_2, ..., t_N\}$ and a list of AI capability descriptions $C = \{c_1, c_2, ..., c_M\}$, the embedding model maps each element in the list into its respective vector embedding. Thus, $f(OT) \to Z_{OT}$ and $f(C) \to Z_C$ where $Z_T = \{z_{t1}, z_{t2}, ..., z_{tN}\}$ and $Z_C = \{z_{c1}, z_{c2}, ..., z_{cM}\}$ are sets of d-dimensional vectors residing in the same vector space.

AI Capability Clustering

To expose similar traits among the different AI capability descriptions, and to provide a humanly-interpretable view of the embedding space, we used clustering to organize AI capabilities into well defined groups of skills. To this end, we applied K-means clustering to the embedding representation of AI capabilities to produce a multi-set $ZS = \{ZS_1, ZS_2, ..., ZS_k\}$ of functionally coherent clusters. The optimal number of clusters, k, was determined by maximizing the clustering silhouette score, ensuring well-separated groupings. Each AI capability was assigned to one of these clusters. Figure 1 shows a T-SNE depiction of the embedding space, labeled by the cluster grouping found in this step. The visualization confirms that AI capabilities form well-separated, semantically coherent clusters when projected into the embedding space. To condense the information in each cluster, we compute a *clustering embedding* as follows: for each $ZS_i = \{z_{cp}, z_{cq}, ...z_{cr}\} \in ZS$ we build its corresponding mapping set in the textual space $S_i = \{c_p, c_q, ...c_r\}$. Then we compute the *cluster embedding* z_{Si} by concatenating all AI capability textual descriptions assigned to S_i and generating a global embedding

$$z_{Si} \leftarrow f(cat(c_p, c_q)|\forall c_p, c_q \in S_i, p < q)$$

. This approach summarizes the semantic richness of the AI capability set, making our subsequent computations more computationally efficient.

Similarity Computation

For each task in an occupation, our framework identifies the most similar AI-capability that can be associated to the task, and produces a per-task similarity score. This is a two-step process: (1) cluster assignment and (2) within-cluster similarity computation.

Cluster Assignment. The goal of this step is to identify the AI capability cluster that more closely relates to each specific occupational task. By identifying first the closest cluster rather than only the closest task, we ensure that similarity is assessed considering the full context of the cluster and not just the words involved in a task. This approach allows us to focus on a smaller set of semantically relevant AI-capabilities. Thus, for each occupational task $z_{tj} \in Z_{OT}$ we assign it to the most similar AI capability cluster AI_{Si} by finding the cluster embedding z_{Si} that maximizes its cosine similarity:

$$AI_{Si} \cup \{z_{tj} \textit{ iff } \arg\max_{Si} \cos(z_{tj}, z_{Si})|\forall z_{tj} \in Z_{OT}\}$$

Within-Cluster Similarity Computation. Once the framework assigns a task to its most relevant AI capability cluster, it computes the cosine similarity between the task embedding and all individual AI capabilities within that cluster. The highest similarity score within the assigned cluster determines the AI alignment score for that task. Thus, for a given task z_{tj} that was assigned to an AI-capability cluster AI_{Si} we compute the maximum similarity score between the task embedding and each AI capability embedding in ZS_i:

$$score_{tj} \leftarrow \max(\cos(z_{tj}, z_{cp}))|\forall z_{cp} \in ZS_i$$

This two-step similarity computation reduces the risk of assigning a task to an AI capability with little to no semantic relation, but that shares a large fraction of non-functional words with the task. It also enhances efficiency by reducing the number of comparisons from the full set of AI capabilities to only those within the most relevant cluster; ensuring that the task alignment process is computationally efficient and semantically meaningful.

AI-OCI Score Computation

Finally, we aggregate the AI-aignment scores for all the tasks within an occupation and produce our *AI Occupational Capability Index (AI-OCI)* as a metric designed to understand what percentage of an occupation can be performed or enhanced through AI existing capabilities. The AI-OCI is computed by averaging the highest similarity scores across all tasks within each specific occupation (OT) :

$$\text{AI-OCI}_{OT} = \frac{1}{n}\sum_{j=1}^{n} score_{tj}|\forall t_j \in OT$$

where N represents the number of tasks in the occupation. Occupations with higher AI-OCI scores exhibit stronger AI-task alignment, while those with lower scores remain less compatible with current AI capabilities.

4 Evaluation and Use of AI-OCI

Experimental Setup

Data. We developed the AI-OCI pipeline by integrating two data sources: (1) occupational task descriptions from the O*NET 27.2 database, which covers 923 U.S. occupations and approximately 19,281 unique tasks; (2) curated 338 AI capability descriptions by prompting various commercial LLMs (ChatGPT, Gemini) to describe tasks that can be performed by state-of-the-art AI. To evaluate AI-OCI use in the context of the labor market, we use wage and employment data from the U.S. Bureau of Labor Statistics (BLS), spanning from 2001 to 2024. For the AI impact analysis, we focus on the period from 2022–2024, aligned with the public adoption of LLMs.

Embeddings. To ensure semantic alignment between tasks and AI capabilities, we generated text embeddings using several state-of-the-art models, including BERT (bert-base-uncased), text-embedding-ada-002, multilingual-e5-large-instruct, and OpenAI's text-embedding-3-small and -3-large. We assessed each model's effectiveness based on clustering quality and correlation with the AI Occupational Exposure (AIOE) index. We ultimately selected text-embedding-3-large due to its superior cluster separation (48 optimal clusters) and the initial highest observed correlation with AIOE (r = 0.70). Each task and capability was embedded into a 3,072-dimensional space using the OpenAI API. We normalized all embeddings and applied K-means clustering to the AI capabilities (optimal $k = 48$, silhouette-based). The task-to-capability alignment was computed using cosine similarity. For interpretability, we retained both the average similarity per occupation (AI-OCI score) and the top-matching capability per task.

Figure 1: T-SNE Visualization of 338 AI capabilities embedded using OpenAI's Text-Embedding-3-Large model. Each point represents an AI capability. Similar capabilities cluster together forming distinct groups (e.g., language understanding).

We evaluated the results across both detailed and major occupation groupings. The full pipeline was implemented in Python and executed on an Linux instance (64GB RAM, NVIDIA T4 GPU).

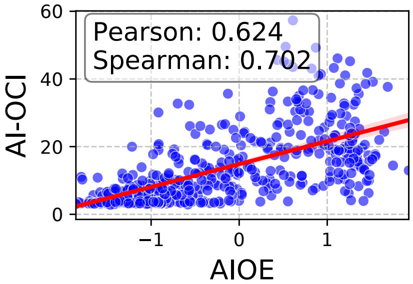
Benchmark Correlation Analysis

To validate the reliability of AI-OCI compared to established approaches, we compared it against three leading benchmarks that estimate occupational exposure to AI: the AI Occupational Exposure (AIOE) index (Felten, Raj, and Seamans 2019), OpenAI's GPT-4 Beta task performance scores (Eloundou et al. 2023), and Frey and Osborne's automation risk index (Frey and Osborne 2017). As shown in Figure 2, AI-OCI exhibits strong positive correlations with AIOE (Pearson: 0.624, Spearman: 0.702) and GPT-4 Beta (Pearson: 0.639, Spearman: 0.744), and a negative correlation with Frey (Pearson: −0.500, Spearman: −0.403). While we focus our analysis on these three key metrics, we also evaluated AI-OCI against additional benchmarks, including GENOE and Webb's AI keyword indices. These results are shown in Table 2.

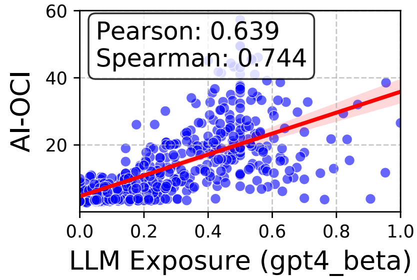
The AIOE and GPT-4 Beta benchmarks reflect model-informed assessments of how well current AI systems, especially LLMs, perform occupational tasks. AI-OCI aligns closely with these state-of-the-art metrics, highlighting its accuracy in identifying tasks within occupations that align with AI capabilities. Recall that while AIOE relies on intensive manual analysis and expert judgment, our AI-OCI

Benchmark Index	Pearson	Spearman
AIOE	0.624	0.702
GPT-4 Beta	0.639	0.744
Frey Index	−0.500	−0.403
GENOE (1-Year)	−0.088	−0.055
GENOE (5-Year)	−0.163	−0.126
GENOE (10-Year)	−0.188	−0.169
Webb AI Score	0.224	0.217
Webb Software Score	−0.144	−0.142
Webb Robot Score	−0.434	−0.557

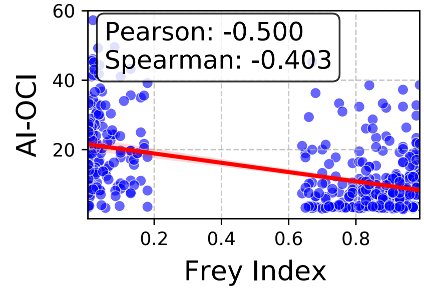
Table 2: Correlation of AI-OCI with various AI exposure benchmarks. Positive values indicate alignment with augmentation-based indices (e.g., AIOE, GPT-4 Beta), while negative values reflect divergence from substitution-centric or robotics-focused measures.



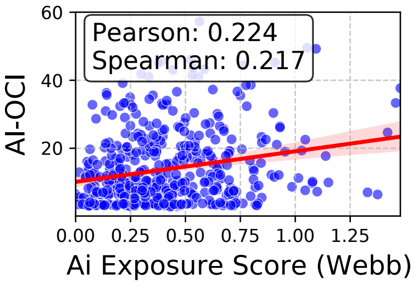
(a) AI-OCI vs AIOE



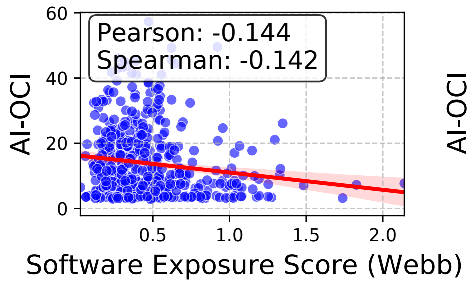
(b) AI-OCI vs GPT-4 Beta



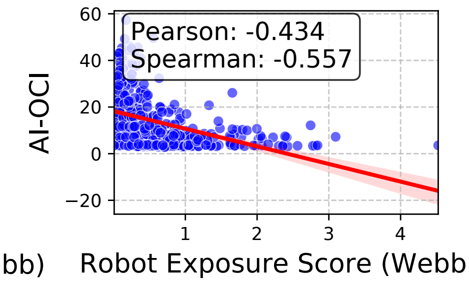
(c) AI-OCI vs Frey Index



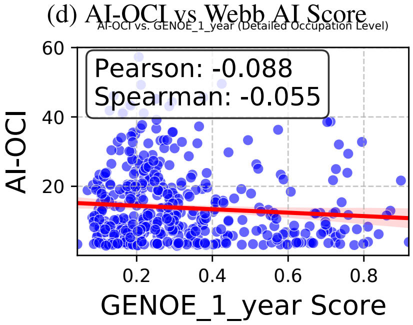
(d) AI-OCI vs Webb AI Score



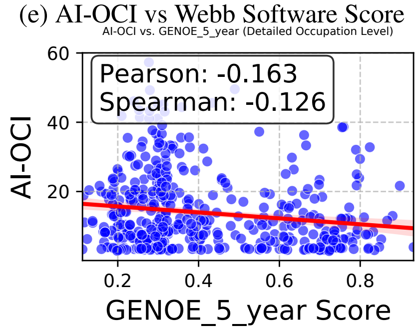
(e) AI-OCI vs Webb Software Score



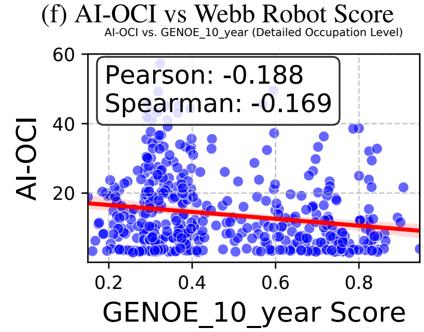
(f) AI-OCI vs Webb Robot Score



(g) AI-OCI vs GENOE (1-Year)



(h) AI-OCI vs GENOE (5-Year)



(i) AI-OCI vs GENOE (10-Year)

Figure 2: AI-OCI correlation with key benchmarks. (a) AI-OCI vs AIOE, (b) vs GPT-4 Beta scores, and (c) vs Frey and Osborne's automation risk index. Webb's AI, Software, and Robot exposure scores (d–f), and GENOE's 1-year, 5-year, and 10-year exposure estimates (g–i). Each point is an occupation. The red line shows the linear regression fit with a shaded confidence interval.

is fully automated. To further strengthen this alignment, we replicated the wage binning strategy from Eloundou et al. (2023), who validated AIOE as an economic signal. Our replication, as seen in Figure 3 shows a consistent positive correlation (0.78) between AI-OCI and mean wages across 20 quantiles, further validating AI-OCI as a reliable economic indicator.

In contrast, the Frey index was developed in the pre-LLM era and emphasizes binary automation risk based on task routineness and manuality. Its negative correlation with AI-OCI exposes a conceptual shift: from binary substitution toward nuanced augmentation. While Frey emphasizes job elimination, AI-OCI captures AI's growing role in enhancing professional roles across knowledge-based domains. Furthermore, when the Frey index was developed, the worldview of AI in the labor market was very limited, and it was assumed that AI would be mostly used to automate boring, repetitive tasks. However, with the invention of more sophisticated generative AI systems, creative tasks are increasingly being supported, if not replaced, by AI.

Similarly to the Frey index, the low or negative correlations observed between AI-OCI and the Webb and GENOE metrics—and even between those benchmarks and AIOE—suggest that these alternative measures fail to capture the evolving nature of modern AI systems. This discrepancy may stem from their reliance on outdated assumptions, indirect proxies like patents or keyword matching, or subjective heuristic judgments rather than direct task-level alignment. As a result, they may underrepresent augmentation effects, overlook semantic task complexity, or misclas-

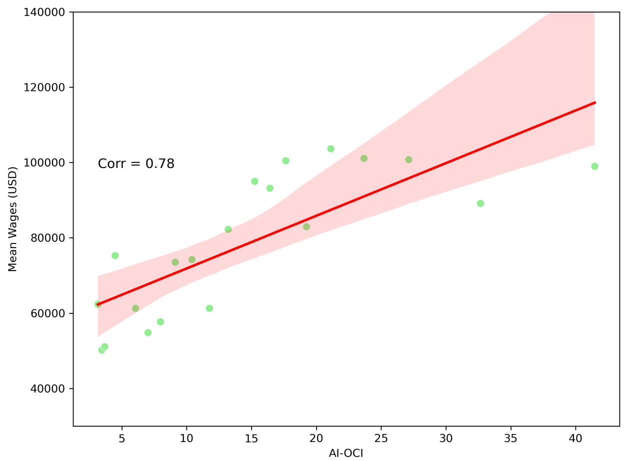


Figure 3: Relationship between AI-OCI scores and average wages in 2024. Occupations are binned into 20 quantiles based on AI-OCI scores. The linear regression line (red) with a shaded 95% confidence interval illustrates a positive association, consistent with findings by Eloundou et al. (2023)

sify roles that are now increasingly influenced by generative and multimodal AI capabilities.

Occupational Ranking and Drill-Down Capabilities

AI-OCI supports both high-level and fine-grained analysis, allowing labor market analyst:wqs to navigate between major occupational groups and individual job roles. At the most detailed level, the framework computes AI-task similarity scores at the task level, which are then aggregated to produce occupation-level AI-OCI scores. Figure 4 presents a ranked view of the top and bottom 20 occupations. High-scoring roles—such as *Market Research Analysts*, *Financial Risk Specialists*, and *Interpreters* exhibit strong alignment with language-based AI capabilities. Conversely, roles such as *Meat Packers*, *Tire Builders*, and *Laundry Workers* fall at the bottom, reflecting the current limitations of AI in manual and physically intensive work.

The metric also scales upward to support broader comparisons. Figure 5 displays average AI-OCI scores across major occupational groups, with fields like Legal, Engineering, and Life Sciences showing the strongest AI alignment. This hierarchical design allows AI-OCI to power both drill down diagnostics and roll up summaries, providing a flexible tool for in-depth labor analysis.

Divergent Labor Trajectories

To analyze occupational shifts during the era of large language model (LLM) adoption, we constructed a six-dimensional feature vector for each occupation based on year-over-year (YoY) percentage changes in both wage and employment. Specifically, we calculated the percentage change in average wage for each occupation from 2021 to 2022, 2022 to 2023, and 2023 to 2024. We performed the same set of calculations for employment over the same three

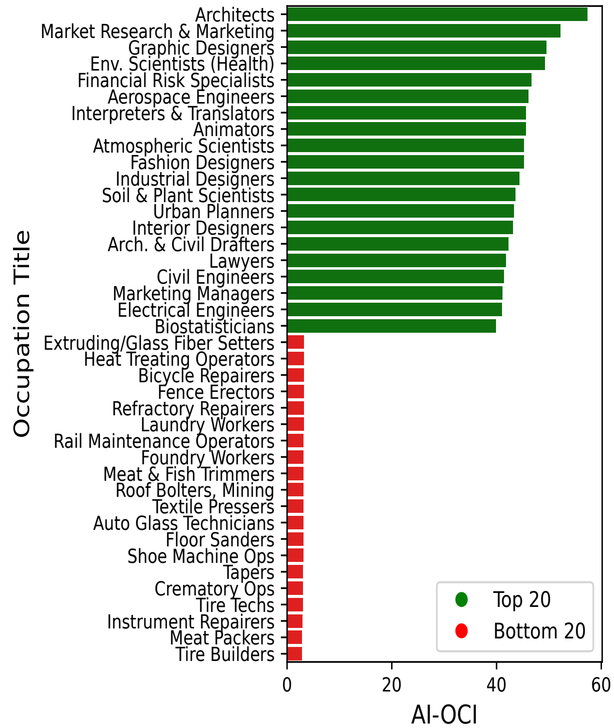


Figure 4: Top and bottom 20 occupations ranked by AI-OCI. Green occupations (e.g., Interpreters) are highly aligned with AI capabilities. Red occupations (e.g., Meat Packers)show low AI alignment.

periods. This resulted in three wage change values and three employment change values per occupation, capturing a continuous view of labor market dynamics over the 2021–2024 period. We then summed these six values to produce a single combined score representing the overall magnitude and direction of recent economic change for each occupation. This score was used to examine the relationship between economic trajectory and AI-OCI alignment, allowing us to identify occupations undergoing either augmentation or displacement as AI capabilities expand.

This analysis, as shown in Figure 6, reveals a group of "bipolar outlier" occupations with high AI-OCI scores (above 30%) but combined change scores beyond ± 1 standard deviation from the mean. These outliers exemplify how alignment with AI capabilities can translate into sharply contrasting economic trajectories. **Positive outliers:** Occupations such as *Marketing Managers* and *Fashion Designers* experienced strong wage or employment growth alongside high AI-OCI scores. These roles appear to benefit from AI augmentation, likely incorporating generative tools and LLMs to enhance creativity, productivity, and market relevance. **Negative outliers:** In contrast, roles like *Order Clerks*, *Medical Transcriptionists*, and *Environmental Engineering Technicians* also exhibit high AI-task alignment but face wage stagnation or declining employment. These trends suggest displacement effects, where AI is able to effectively

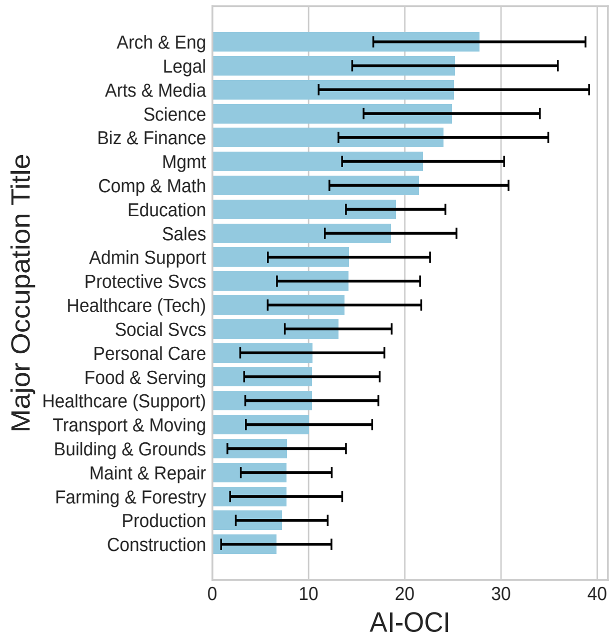


Figure 5: Average AI-OCI scores across 22 major occupation groups. Higher alignment is observed in analytical-intensive sectors like Legal and Engineering.

substitute human labor in routine documentation, transcription, or clerical tasks.

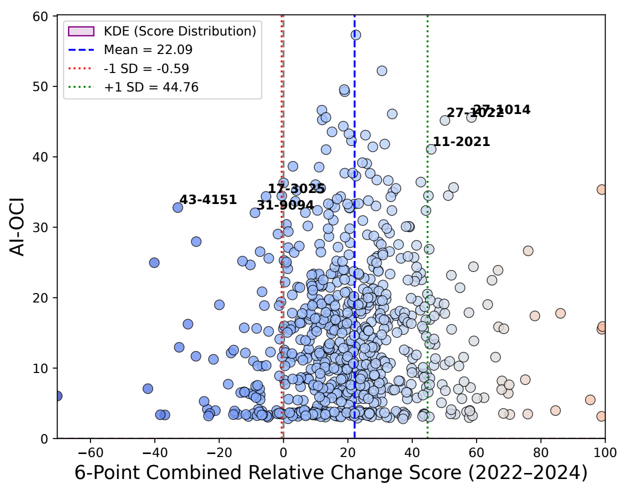


Figure 6: AI-OCI vs. 6-point Combined Score. KDE and SD bands reveal occupations undergoing significant wage/employment changes.

To better understand the trajectories of bipolar outliers, we traced year-over-year wage and employment changes for these outlier roles from 2013 to 2024. Figure 7 illustrates diverging trajectories; while *Fashion Designers* show strong post-2022 growth, *Order Clerks* and *Transcriptionists* display a marked decline coinciding with the rise of LLM adoption. An interesting case is exemplified by the *Animators* where both, wages and employment increased drastically right after COVID-19, but exhibited a similarly drastic decline after 2022, where generative AI tools became pervasive in this industry. Through the outlier analysis AI-OCI enables stakeholders with practical economic and policy insights. It demonstrates that high AI alignment does not uniformly predict positive or negative outcomes. Instead, outcomes depend on whether AI augments or replaces task performance. Policymakers can use these insights to identify at-risk occupations for re-skilling, while employers can better target roles primed for augmentation and innovation in the AI-driven economy.

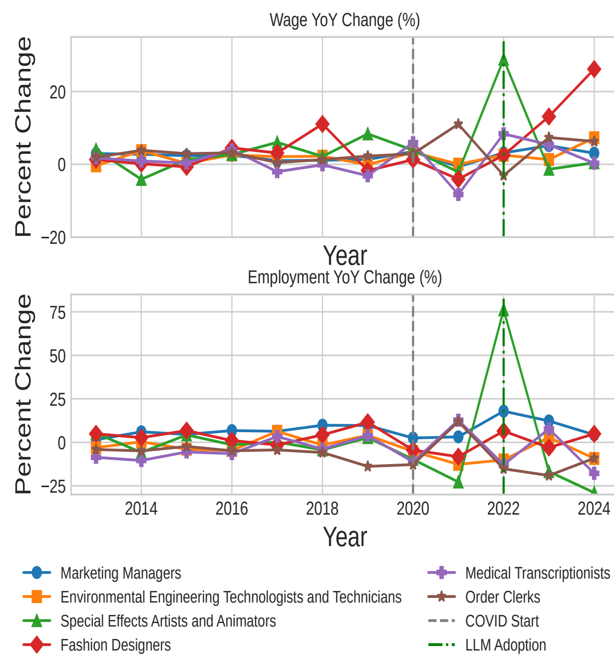


Figure 7: Year-over-year wage and employment trends (2022–2024) for high-AI-OCI outliers. Highlights roles with notable recent shifts.

5 Limitations

While AI-OCI offers a reliable, task-level framework for quantifying AI's impact on the workforce, several limitations remain: First, the framework relies on structured datasets—namely, O*NET for occupational tasks and BLS data for employment trends. Although these sources are comprehensive and widely used, they may lag behind real-time labor shifts or fail to capture emerging occupations and informal sectors. Future iterations could incorporate alternative sources such as online job postings, global taxonomies, or employer-submitted competency data to improve coverage and timeliness. Second, AI-OCI uses pretrained language models that may encode systemic biases present in their training corpora. These biases can affect which occupations appear more AI-aligned, potentially skewing assessments of vulnerability or opportunity. Mitigating such effects will require debiasing strategies, including contrastive learning, counterfactual data augmentation, and adversarial fine-tuning. Third, although the curated set of 338 AI capabilities spans NLP, reasoning, planning, and code generation, it remains skewed toward text-centered tasks. Current capabilities under-represent domains such as robotics, perception, and embodied interaction. Consequently, AI-OCI may underestimate alignment for occupations that rely on physical dexterity, spatial reasoning, or motor control. Expanding the capability taxonomy to include multi-modal and embodied systems would strengthen real-world applicability.

6 Conclusions and Future Work

This paper introduces the AI Occupational Capability Index (AI-OCI), a novel, task-level framework for measuring how well current AI capabilities align with human occupational tasks. This fully automated framework uses state-of-the-art text embeddings, semantic similarity, and clustering-based alignment. Our proposed metric produces robust and interpretable scores that quantify AI exposure and AI augmentation potential at multiple scales of granularity.

Unlike legacy frameworks that rely on static occupation-level heuristics or expert manual analysis, AI-OCI performs direct comparisons between AI functionalities and over 19,000 real-world job tasks. The resulting scores offer dynamic, model-agnostic insights into how AI intersects with labor at the task level. Our empirical validation shows that AI-OCI correlates strongly with leading benchmarks such as AIOE and GPT-4 Beta exposure scores, while diverging from outdated automation-centric measures like the Frey index. Case-based analyses highlight AI-OCI's practical value: revealing early signals of substitution (e.g., Graphic Designers), augmentation (e.g., Interpreters), and forecast mismatches (e.g., Market Research Analysts). We envision that its application to education-to-workforce pipelines will enable valuable insights for labor policy, and skills planning in the AI era.

We plan to extend AI-OCI into a dynamic, real-time index that continuously reflects evolving labor and AI dynamics. A possible direction involves integrating retrieval augmented generation (RAG) to support continuous updates to both task and capability descriptions. This enhancement will enable the framework to ingest new job postings, technical documentation, and research outputs, allowing AI-OCI to track shifting task requirements and technological developments in near real time. We also intend to broaden the AI capability layer by incorporating multimodal models such as GPT-4V, CLIP, and Perceiver IO. These additions will better represent vision, robotics, and sensorimotor capabilities, improving alignment assessments for occupations beyond language-based tasks.

Beyond the labor market, we are exploring how AI-task alignment influences education and skills development. Early findings suggest that AI-OCI could predict shifts in college admissions trends, degree program demand, and curricular redesign. These extensions position AI-OCI as a bridge between labor intelligence and educational strategy.

Ethics Statement and Broader Impact

This research introduces the AI Occupational Capability Index (AI-OCI), a task-level framework for measuring alignment between AI capabilities and occupational functions. While our methodology advances the measurement of AI's potential impact on labor, several ethical considerations are relevant.

Bias and Fairness. The AI-OCI framework relies on language model embeddings, which may inherit biases present in the training data of those models. Such biases could unintentionally favor or penalize certain occupations, particularly those with gendered or culturally specific descriptions. We acknowledge this risk and recommend that future iterations integrate bias correction techniques or fairness-aware embeddings.

Socioeconomic Impact. AI-OCI highlights which occupations are most aligned with AI capabilities, potentially influencing decisions around hiring, automation, and reskilling. Although this information is valuable for workforce planning, it could also lead to unintended consequences, such as accelerating job displacement in already vulnerable sectors. We encourage responsible use of AI-OCI scores alongside human-centered economic planning to mitigate harm.

Use of Public Data. The study exclusively uses publicly available occupational task data (O*NET) and labor statistics (BLS), ensuring compliance with ethical data sourcing. AI capability descriptions were generated through LLM prompting and curated by the authors.

Responsible Deployment. We do not release an automated AI-OCI scoring tool for public use at this stage, as unintended use without sufficient context could result in oversimplified policy or employment decisions. Future deployments will prioritize explainability and human oversight.

Broader Impact. AI-OCI offers actionable insight for educators, policymakers, and labor economists by quantifying how AI aligns with real-world tasks. It supports transparent workforce forecasting and ethical AI adoption strategies. We also envision extending AI-OCI into education and upskilling domains to guide program design and long-term planning in the AI era.

References

Autor, D. 2023. The Labor Market Impacts of Technological Change. *Journal of Economic Perspectives*.

Benítez-Rueda, M.; and Parrado, E. 2024. Mirror, Mirror on the Wall: Which Jobs Will AI Replace After All? A New Index of Occupational Exposure. Working Paper IDB-WP-1624, Inter-American Development Bank (IDB).

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

Cazzaniga, A.; et al. 2024. Gen-AI: Artificial Intelligence and the Future of Work. IMF Staff Discussion Note SDN/2024/001, International Monetary Fund, Washington, DC. Accessed: 2024-07-20.

Colombo, E.; Mercorio, F.; Mezzanzanica, M.; and Serino, A. 2024. Towards the Terminator Economy: Assessing Job Exposure to AI through LLMs. Working Paper WP 2401, DISEIS, Università Cattolica del Sacro Cuore.

Eloundou, T.; Manning, S.; Mishkin, P.; and Rock, D. 2023. GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. *arXiv preprint arXiv:2303.10130*.

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*.

Felten, E. W.; Raj, M.; and Seamans, R. 2018. A Method to Link Advances in Artificial Intelligence to Occupational Abilities. *AEA Papers and Proceedings*, 108: 54–57.

Felten, E. W.; Raj, M.; and Seamans, R. 2019. The Occupational Impact of Artificial Intelligence: Labor, Skills, and Polarization. Technical report, SSRN Electronic Journal.

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.

Goldfarb, A.; Taska, B.; and Teodoridis, F. 2022. Could Machine Learning be a General Purpose Technology? A Comparison of Emerging Technologies Using Data from Online Job Postings. Technical Report w29767, National Bureau of Economic Research.

Pennington, J.; Socher, R.; and Manning, C. D. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.

Pizzinelli, C.; Borup, R.; and Fernandez-Cornejo, J. 2023. Labor Market Exposure to AI: Cross-country Differences and Distributional Implications. IMF Working Paper WP/23/216, International Monetary Fund.

Stanford University. 2024. Artificial Intelligence Index Report 2024. https://aiindex.stanford.edu/2024. Accessed: February 6, 2024.

Thompson, N.; et al. 2024. Rethinking AI's impact: MIT CSAIL study reveals economic limits to job automation. https://www.csail.mit.edu/news/rethinking-ais-impact-mit-csail-study-reveals-economic-limits-job-automation. Accessed: 2025-05-20.

Tolan, S.; Binns, R.; Van Kleek, M.; and Shadbolt, N. 2020. Measuring the Occupational Impact of AI: Tasks, Cognitive Abilities and AI Benchmarks. *Journal of Artificial Intelligence Research*, 69: 365–416.

Wang, L.; Yang, N.; Huang, X.; Yang, L.; Majumder, R.; and Wei, F. 2024. Improving Text Embeddings with Large Language Models. *arXiv preprint arXiv:2401.00368*. 20 pages, 15 tables. Subjects: Computation and Language (cs.CL); Information Retrieval (cs.IR).

Webb, M. 2019. The Impact of Artificial Intelligence on the Labor Market. Stanford University Working Paper.

World Economic Forum. 2024. AI - Artificial Intelligence at Davos 2024: What to Know. Accessed: 2024-07-20.

Yildiz, C. 2023. AI Revolution: Large Language Models (LLMs) Shaping the Future of Branding and Marketing. LinkedIn article. Accessed: March 1, 2024.