



Large language models at work in China's labor market[☆]

Qin Chen ^{a,*,1}, Jinfeng Ge ^{b,*,1}, Huaqing Xie ^{c,1}, Xingcheng Xu ^{c,*,1}, Yanqing Yang ^{d,*,1}

^a MetroData, China

^b Shanghai Maritime University, China

^c Shanghai Artificial Intelligence Laboratory, China

^d Fudan University, China

ARTICLE INFO

Keywords:

Large language models
Chinese labor market
Occupational exposure
Industry analysis
AI learning theory

ABSTRACT

This paper explores the potential impacts of large language models (LLMs) on the Chinese labor market. We analyze occupational exposure to LLM capabilities by incorporating human expertise and LLM classifications, following the methodology of Eloundou et al. (2023). The results indicate a positive correlation between occupational exposure and both wage levels and experience premiums at the occupation level. This suggests that higher-paying and experience-intensive jobs may face greater exposure risks from LLM-powered software. We then aggregate occupational exposure at the industry level to obtain industrial exposure scores. Both occupational and industrial exposure scores align with expert assessments. Our empirical analysis also demonstrates a distinct impact of LLMs, which deviates from the routinization hypothesis. We present a stylized theoretical framework to better understand this deviation from previous digital technologies. By incorporating entropy-based information theory into the task-based framework, we propose an AI learning theory that reveals a different pattern of LLM impacts compared to the routinization hypothesis.

1. Introduction

Recent remarkable progress in the field of generative AI and large language models (Bubeck et al., 2023; Zhao et al., 2023) has provoked many pressing questions about the effects of these powerful technologies on the economy. One of the most significant questions surrounding advances in generative AI and LLMs is the impact these technologies will have on the dynamics of the labor market due to the influence of LLMs on labor inputs. A branch of research emphasizing the disruptive labor market impacts of LLMs is emerging rapidly, however, it predominantly focuses attention on the labor market in developed economy, in particularly the U.S. (Brynjolfsson et al., 2023; Eloundou et al., 2023; Felten et al., 2023; Noy & Zhang, 2023; Peng et al., 2023). Nevertheless, countries differ in their labor market structures such as occupation and industry composition. Even for the same occupation, the detailed task composition or work content may exhibit great discrepancies across countries. Therefore, this paper analyzes the potential impacts of LLMs on China's labor market. To construct our primary exposure index, we use a recently developed methodology to systematically assess which occupations are most exposed to LLMs in China. Specifically, we employ three large language models — GPT4 (OpenAI, 2023), InternLM (InternLM-Team, 2023), and GLM (Du et al., 2022; Zeng et al., 2022) — as classifiers to determine the occupational exposure based on the detailed description for each occupation contained in the general

[☆] Supported by Key Program of the National Natural Science Foundation of China (No. 71933001), and Shanghai Artificial Intelligence Laboratory.

* Corresponding authors.

E-mail addresses: qinchen1986@hotmail.com (Q. Chen), jfge@shmtu.edu.cn (J. Ge), xiehuaqing@pjlab.org.cn (H. Xie), xingcheng.xu18@gmail.com (X. Xu), yanqingyang@fudan.edu.cn (Y. Yang).

¹ All authors contributed equally to this work.

code of occupational classification of the People's Republic of China. We also employ expert annotators to explore the impacts of LLMs, to make comparisons, and to shed more light on this issue. Our key empirical findings are twofold. First, we empirically analyze occupational exposure to LLMs, highlighting the implications of heterogeneous occupational exposure. Second, we leverage these findings to assess the labor market demand exposure to LLMs across industries and demographic groups in China.

The analysis reveals significant heterogeneity in occupational exposure, with more educated, higher paid, white-collar occupations being the most exposed to LLMs. These results align with recent studies on the US labor market (Eisfeldt et al., 2023; Eloundou et al., 2023), which also found that technological advancements disproportionately impact workers at the higher end of the wage spectrum. Beyond the positive correlation between wage, education, and occupational exposure, our findings indicate a positive correlation between experience premiums and exposure to LLMs, suggesting potential diminishing returns to “learning by doing” in the future.

To deepen these insights, we employed advanced language models (e.g., GPT-4, GLM, InternLM) to evaluate the quintet of occupations defined by Acemoglu and Autor (2011): Non-routine Cognitive (Analytical), Non-routine Cognitive (Interpersonal), Routine Cognitive, Routine Manual, and Non-routine Manual Physical. For each occupation, we calculated an aggregate exposure score by averaging sub-occupation evaluations. The results indicate that non-routine cognitive skills, particularly analytical ones, are most significantly affected by LLMs. Routine manual skills are also substantially influenced, whereas interpersonal and non-routine manual physical skills remain relatively unaffected.

This predominant impact on non-routine cognitive tasks deviates from the “routinization hypothesis” (Autor et al., 2003), which posited that ICT primarily automates routine, codifiable tasks performed by middle-wage workers. To explore the different occupational exposure patterns between traditional AI technologies and LLMs, we developed a theoretical model to shed light on the underlying mechanisms and policy implications.

Our analysis further indicates that the impacts of LLMs on China's labor market are pervasive and diverse. Industry-level analyses reveal that education and healthcare have a higher level of exposure, while manufacturing, agriculture, mining, and construction show lower exposure. In contrast to other developed countries, the uneven age distribution across industries in China amplifies the demographic exposure to LLMs, disproportionately affecting younger workers.

Furthermore, LLMs significantly impact labor demand. Using an online job posting dataset, we constructed an occupational vacancy index. A positive correlation between vacancy shares and occupational exposure scores suggests that labor demand structures may exacerbate the disruptive impacts of LLMs on China's labor market. The positive correlation between the growth rate of vacancy shares and exposure scores further indicates a potential reversal in labor demand trends. Contrary to expectations, China's economic and labor market structure intensifies rather than mitigates the disruptive effects of LLMs.

Motivated by the unique occupational exposure of LLMs compared to earlier technologies, we introduce a novel theoretical model to examine why different technologies produce different occupational exposure structures. This model incorporates entropy-based information theory into a stylized task-based framework, which is particularly effective for characterizing comparative advantage. By defining task complexity using entropy and modeling the relative productivity of skills across occupations, we can systematically assess the efficiency of AI technologies, including LLMs, across various tasks.

The model also integrates KL-divergence to represent the relative efficiency of AI models across occupations, distinguishing the automation logic of traditional AI from that of deep learning-driven LLMs. This represents a significant advancement in AI technology. Finally, we discuss the theoretical and policy implications of LLMs inferred from our model, providing critical insights for navigating the future of labor markets.

The paper is structured as follows: Section 2 reviews the related literature, Section 3 discusses methods and data collection, Section 4 presents the main results, Section 5 introduces a theoretical model to further discuss the impacts of LLMs, and Section 6 offers concluding remarks.

2. Literature review

Artificial Intelligence, like previous technologies, is poised to impact the economy in various ways, potentially fostering economic growth and reshaping the labor market structure (Furman & Seamans, 2019; Goldfarb et al., 2023). A substantial and expanding body of literature delves into the labor market consequences of artificial intelligence and automation technologies broadly defined. The skill-biased technological change framework (Acemoglu, 2002; Katz & Murphy, 1992), along with the task-based framework of automation (Acemoglu & Autor, 2011; Acemoglu & Restrepo, 2018; Autor et al., 2003) are often regarded as the foundational frameworks for comprehending technology's impact on the labor market. This line of research has introduced the concept of routine-biased technological change, indicating that workers engaged in routine tasks face a heightened risk of displacement due to technological advancements. Numerous studies have demonstrated that automation technologies have contributed to both income inequality and job polarization, driven by declines in relative wages and employment for workers specializing in routine tasks (Acemoglu & Restrepo, 2022; Autor et al., 2006; Van Reenen, 2011). The influence of AI on work is anticipated to be multifaceted. Recent studies have made distinctions between task-displacement and task-reinstatement effects of technology, whereby new technology introduces novel occupations that bolster labor demand (Acemoglu & Restrepo, 2018, 2019).

Historically, prior research has predominantly adopted the task-oriented approach to analyze the labor market impacts of artificial intelligence. Various methods have been employed to evaluate the similarity between AI capabilities and the tasks performed by workers across different occupations. These methodologies encompass aligning AI capabilities with diverse skills and abilities demanded by distinct occupations (Felten et al., 2018; Tolan et al., 2021), mapping AI patent descriptions to worker task descriptions (Meindl et al., 2021; Webb, 2019), employing machine learning classifiers to estimate the potential for

automation across all occupations (Frey & Osborne, 2017), devising innovative rubric the suitability of worker activities by machine learning (Brynjolfsson et al., 2018), and leveraging expert forecasts (Grace et al., 2018).

Nevertheless, this line of research is becoming increasingly challenging due to the evolving and advancing capabilities of AI. Following recent breakthroughs in Generative AI and LLMs, there has been a growing body of studies investigating the specific economic impacts and opportunities presented by LLMs. For instance, Peng et al. (2023) conducted a study where software engineers were enlisted for a specific coding task, revealing that those with access to GitHub Copilot completed the task twice as quickly. Similarly, Noy and Zhang (2023) conducted an online experiment to explore the displacement effects of Generative AI on professional writing tasks. Additionally, Brynjolfsson et al. (2023) examined the effects of Generative AI on customer support agents. Pertinent to this paper, Felten et al. (2023) explored the heterogeneity in occupational exposure, while Eloundou et al. (2023) introduced a novel rubric to assess the impacts of LLMs on labor forces. Concurrently with this line of inquiry, we aim to characterize the potential relevance of LLMs to China's labor market in particular.

Our theoretical model is based on the task-based automation framework (Acemoglu & Autor, 2011; Acemoglu & Restrepo, 2018; Autor et al., 2003). To better capture the unique characteristics of LLMs, we incorporate insights from information theory into this framework.

Information theory, originating from Shannon's seminal work (Shannon, 2001), has been widely applied across various disciplines e.g. Cover (1999) and MacKay (2003). The emergence of deep learning has further expanded its relevance by enabling the representation of complex, nonlinear, and hierarchical data structures. Shannon's maxim of "information first, then computation" inspires the integration of information theory into AI research, exemplified by the information bottleneck framework (Tishby et al., 2000; Tishby & Zaslavsky, 2015). This framework formulates a learning objective grounded in information theory and prescribes algorithms to optimize it. Information theory has significantly advanced the understanding of deep neural networks, from practical applications like the variational information bottleneck (Alemi et al., 2016) to theoretical investigations of generalization bounds using mutual information (Xu & Raginsky, 2017). Recent work by Schwartz Ziv and LeCun (2024) unifies research on self-supervised and semi-supervised learning through an information-theoretic perspective, discussing optimal representations for neural networks. Notably, information theory has been employed to analyze LLM performance through "neural scaling laws" (Barnett & Besiroglu, 2023; Hoffmann et al., 2022; Kaplan et al., 2020), which describe the relationship between compute, data, and model performance.

In econometrics, statistics, and economics, information theory has also proven valuable. Its core concept, entropy, quantifies uncertainty in random variables and complexity in transmitted information. Sims (2003) introduced a dynamic model of rational inattention, where agents face entropy-based costs in processing information. This approach, further developed by Maćkowiak et al. (2023), provides a flexible framework for modeling information constraints, with entropy serving as a tractable benchmark.

By integrating information theory into our task-based framework, we establish deeper insights into task automation and artificial intelligence. This approach allows us to characterize the relative productivity of skills and AI models across tasks using key concepts such as entropy, KL-divergence, and neural scaling laws. These tools enable a rigorous analysis, akin to the Cobb–Douglas production function, capturing the fundamental distinctions between traditional and advanced AI-driven automation.

3. Methodology of exposure scoring and data collection

3.1. Data on occupations in China

In order to assess occupational exposure scoring in China and facilitate standardized comparisons, a consistent occupational classification system is essential. The Occupational Classification Dictionary of the People's Republic of China (OCD) version 2022, published by the National Bureau of Statistics of China (NBS) and the Ministry of Human Resources and Social Security of the People's Republic of China (MOHRSS), provides such a classification system that serves as a standardized analytical tool for occupations. In particular, we utilize the general code of Occupational Classification of the People's Republic of China 2022, which encompasses comprehensive information across 8 large categories, 79 medium categories, 449 small categories, and 1636 fine categories. This information, includes the definition of each occupation, the content and format of work activities, as well as specific description of the scope of work activities. We leverage the detailed descriptions of various occupations to facilitate the classification of online job vacancies into distinct occupational categories. To provide further clarity, a sample of occupations and their exposure to LLMs, categorized by medium categories, is presented in Appendix A.

3.2. Data on wages and vacancies

To acquire both vacancy and wage information, we leverage two datasets. Our primary data source is an online job posting dataset collected by the City Data Group. The dataset compiles online job postings spanning from January 2017 to December 2022, originating from major online job market platforms in China including zhaoping.com, 51job, 58.com, Ganji.com, Lagou.com, and Kanzhun. This comprehensive compilation encompasses over 800 million job openings across nearly 400 cities and 5.2 million employers. For each job vacancy entry within this database, we have access to a range of information: including the posting date, position type, occupation titles, the quantity of workers to be recruited, wage ranges (if specified), education requirements (if applicable), work experience prerequisites (if indicated), the name of the employing firm, the work location for the position, and the textual content of job descriptions. Following the classification of each job posting distinct occupations, we derive corresponding statistics encompassing the number of vacancies, the typical educational qualifications required for entry, and the wage structure within each occupation. The City Data Group frequently reviews the representativeness of the job vacancy postings it scrapes, to

ensure the information renders an accurate picture at least the online recruitment labor market. Obviously, the online recruitment labor market does not equal to the whole labor market; to gauge the nuance difference between online recruiting labor market and the overall labor market, the one precondition is a representative aggregate recruitment dataset. Since in China there is no representative recruitment dataset such as JOLTS in United States, therefore, we cannot speak too much about representativeness of online job recruitment compared with the overall recruitment. Although the potential unknown biased feature of job posting dataset, we believe large proportion of our main empirical results are immune to this potential bias. First, our key empirical task which is eliciting occupational exposure does not use online job posting data at all. We use online posting data to deliver three key information: vacancy numbers, the average wage rate and experience premium for each occupation. If different sectors are different reliance on online recruitment, indeed the more online recruitment reliant sector's occupation will overestimate. However, the calculation of the average wage rate and experience premium for each occupation are robust to heterogeneous sectoral reliance on online recruitment.

Our second data source is the China Labor-force Dynamic Survey (CLDS) 2016, conducted by the Social Science Survey Center of Sun Yat-sen University. The CLDS surveys the working-age population to explore aspects such as education, employment, labor rights, occupational mobility, health, and well-being. The survey encompasses comprehensive industry and occupation information for each employment entry. Consequently, we can harness this dataset to construct an occupational intensity index for each industry. This index facilitates the acquisition of an exposure score at the industry level derived from occupational exposure scores. Further details regarding the occupational intensity index across 15 industry categories are presented in Appendix C.

3.3. Methodology of exposure scoring

To evaluate the likelihood of an occupation in China undergoing a disruptive shock due to the widespread availability and utilization of LLMs, we continue to adopt a task-based approach. We gauge the exposure of each occupation to LLMs based on the comprehensive occupation descriptions in Chinese, as documented the Occupational Classification Dictionary of the People's Republic of China 2022. Building upon the exposure scoring methods proposed by Eloundou et al. (2023) which conceive an occupation as a collection of tasks, assess whether a given occupation can be executed more efficiently using ChatGPT or analogous LLMs. Our methodology employs three prominent large language models to determine exposure of various occupations. Specifically, we utilize Open AI's GPT-4 model (OpenAI, 2023), the InternLM model developed by Shanghai AI Laboratory and Sense Time (InternLM-Team, 2023), and the GLM (Du et al., 2022; Zeng et al., 2022) to categorize occupations based on their complete set of occupational descriptions. Each of these models operates based on a comprehensive rubric for scoring LLMs exposure. We, then, submit an occupation's description, in conjunction with its title, to each model. In response, each model provides an exposure score. These scores effectively capture whether the time required to complete a task could be halved while maintaining consistent quality, assuming a worker has access to ChatGPT-like LLMs. The Scores are divided into four categories: E0, E1, E2, and E3, the details of each category are presented in Appendix A.

Although we share similar methodologies with Eloundou et al. (2023), several caveats should be noted. First, all the occupational description and prompt are in Chinese, our exercise relies on capabilities of the Chinese large language models. To accomplish better Chinese language performance, we use two top Chinese large language models-GLM and InternLM-in addition to GPT4. Second, the Occupational Classification Dictionary of the People's Republic of China contains only detailed work content and descriptions for each occupation. Therefore, it is infeasible to calculate exposure score at the task level as Eloundou et al. (2023). Instead, we choose to calculate exposure score directly at the occupation level.

OpenAI has pointed out several weaknesses of the method, such as the validity of the task-based framework, relative versus absolute measures, as well as forward-looking and changing nature of the scores. Another limitation we would like to discuss here is the randomness of LLM scoring. The same prompt can still yields different results from large language models, even with a higher temperature setting. In order to compensate for this issue, we first had each LLM model label each occupation 8 times and calculated the scores. We then took the average as the final scoring result for each LLM model for each occupation.

To compare the consistency of LLM scoring with human scoring, we invited a group of experts in economics and artificial intelligence to serve as judges. We provided them with descriptions of medium-category occupations in China and asked them to score each occupation according to the rating criteria. After collecting the scores from all of the experts, we calculated the average score for each occupation as the final human score. The rating criteria represented the proportion of labor input that large language models could save in each occupation. More details on expert evaluation is presented in Appendix A.

4. Results and assessment of impacts of LLMs on China's labor market

4.1. Summary statistics

We employed the aforementioned methodology to gather results from GPT-4, GLM, and InternLM, subsequently assigning scores to the fine categories occupations. We designated E1 as 1 point, E2 and E3 as 0.5 points each, and E0 as 0 points. It is worth noting that the OCD encompasses 8 large categories, which also include 3 distinctive categories: "National institutions, party and mass organizations leaders", "Military personnel", and "Other personnel not classified elsewhere". In the subsequent sections, our results exclude these three categories of occupations, thereby focusing on 1606 fine-category occupations. The scores assigned to small-category, medium-category and large-category occupations represent the averaged scores of the occupations they encompass

Table 1
Summary statistics of LLMs exposure.

Fine categories occupation level exposure			
	GLM	InternLM	GPT-4
count	1606	1606	1606
mean	0.44	0.18	0.24
std	0.26	0.18	0.21
Fine categories occupation level exposure corr.			
GLM	1.0***	0.284***	0.1915***
InternLM	0.284***	1.0***	0.2887***
GPT-4	0.1915***	0.2887***	1.0***
Medium categories occupation level exposure			
	GLM	InternLM	GPT-4
count	63	63	63
mean	0.40	0.14	0.22
std	0.15	0.10	0.18
Medium categories occupation level exposure corr.			
GLM	1.0***	0.5938***	0.306*
InternLM	0.5938***	1.0***	0.4807***
GPT-4	0.306*	0.4807***	1.0***

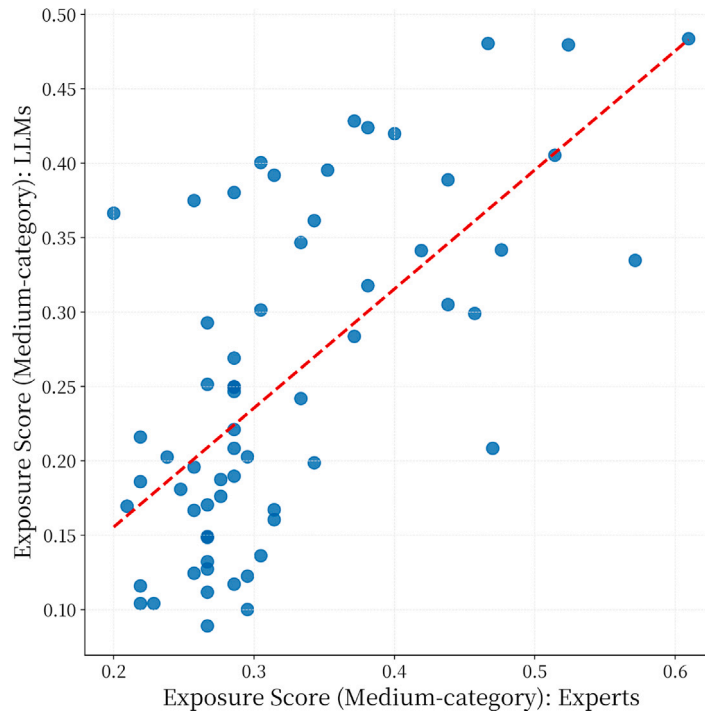


Fig. 1. Exposure score of LLMs and expert: corr. = 0.65.

from the higher level. A comprehensive summary of these metrics, along with the correlations between different models, is provided in the [Table 1](#).

When examining the scoring across fine-categories occupations, GLM demonstrated an average of 0.44, while InternLM and GPT-4 yielded averages of 0.18 and 0.24 respectively. This suggests that, on average, GLM ascribes stronger capabilities to LLMs and deems occupations to be more susceptible to LLMs exposure. Despite variations in the models' scoring and comprehension of occupations, the scoring outcomes from three models exhibited noteworthy positive correlations. This held true across fine-category, small-category, and medium-category occupations. We synthesized the outcomes from the three models, and throughout the subsequent

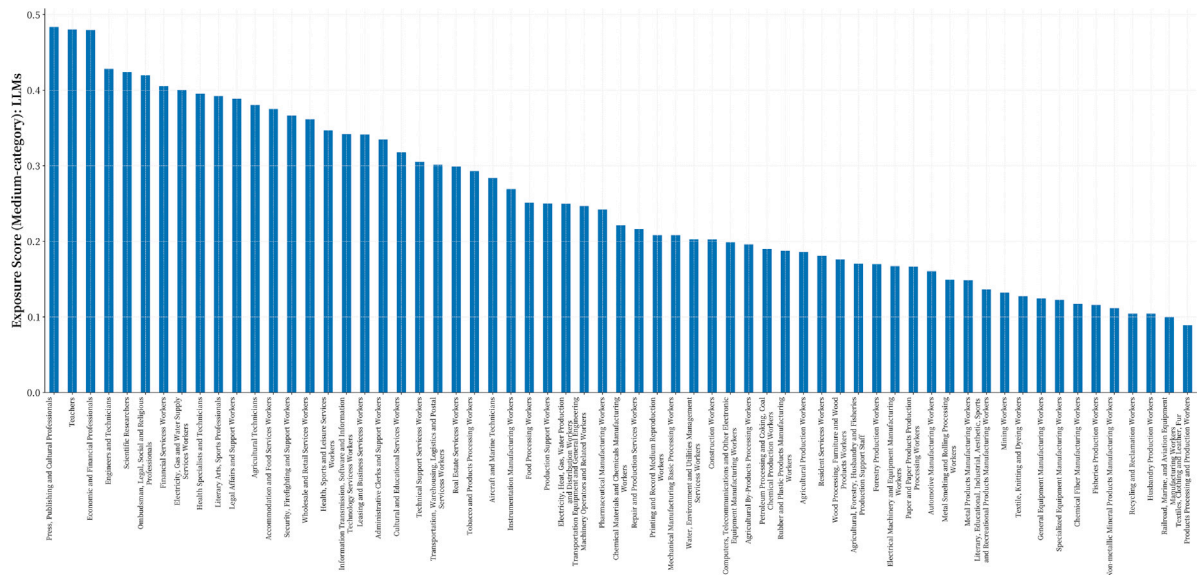


Fig. 2. Exposure score on medium-category level: LLMs.

analysis, unless otherwise specified, we relied on the average scores generated by the three models based on medium-category occupations.

Upon comparing the model scores with the results of human expert scoring in Fig. 1, we observed substantial positive correlations between the model scores and human expert scores.

4.2. Occupational exposure

Fig. 2 provides an overview of the medium-category occupation exposure scores by LLMs. It is worth noting that within the highest exposure occupations, such as “Teachers”, “Economic and financial professionals”, and “Press, publishing and cultural professionals”, a strong resonance emerges with recent technological advancements in LLMs, particularly in their capacity to handle a broad spectrum of intricate language-based tasks and even generate functional code based on high-level descriptions of a programming task. Moving to a broader perspective, the large-category occupation exposure scores by LLMs and expert are represented in Appendix B.

The results of our LLMs exposure scores indicate significantly different mechanisms and corresponding impacts on labor market, when compared with previous waves of computer-based automation. Prior research on computer-based automation posited that routine work was most susceptible to replacement by computers (Acemoglu & Autor, 2011; Autor et al., 2006). This “routinization” hypothesis assumed that “computers and computer-controlled equipment are highly productive and reliable at performing the tasks that programmers can script - and relatively inept at everything else”. The displacement of routine jobs by computers led to a heightened demand for skilled workers in “abstract” jobs, resulting in wage inequality and job polarization. However, Fig. 3 suggests that the labor market repercussions resulting from recent advances in LLMs may diverge. Specifically, this wave of technological change distinguishes itself from earlier waves by potentially substituting numerous tasks within non-routine cognitive analytical jobs that were previously considered immune to automation. In essence, this development suggests a shift in the labor market landscape, where the winners and losers associated with general-purpose technology evolve.

Beyond the distinct impacts on routine and non-routine occupations, LLMs also reshape the experience premium profile. LLMs have the capacity to learn tasks requiring tacit knowledge, knowledge that was once attainable only through experiential learning or learning by doing. As depicted in Fig. 4, LLMs could potentially alter the experience premium and the profile of returns from learning. Our analysis aligns with the insights gleaned from an experimental study (Brynjolfsson et al., 2023).

4.3. Composite occupational exposure

In our study, we utilized advanced language models, including GPT-4, GLM, and InternLM, to evaluate the quintet of occupational categories defined by [Acemoglu and Autor \(2011\)](#). These composite occupational categories are: Non-routine Cognitive (Analytical), Non-routine Cognitive (Interpersonal), Routine Cognitive, Routine Manual, and Non-routine Manual Physical.

To derive the composite exposure scores, the models were tasked with repeatedly scoring various sub-elements within the occupational descriptions. The aggregate score for each composite occupation was then computed as the mean of these evaluations.

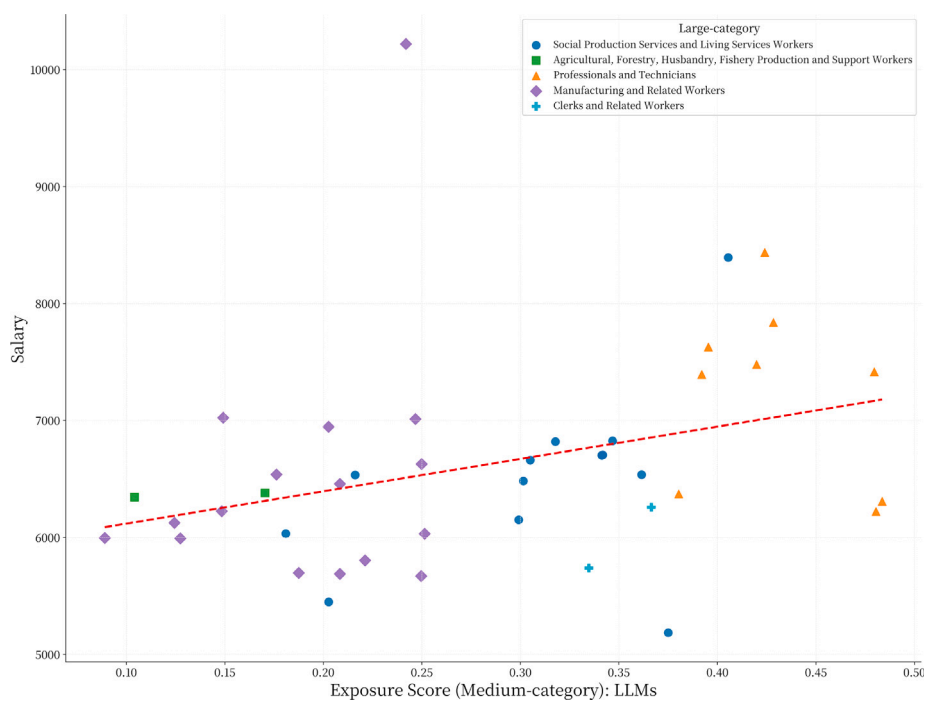


Fig. 3. Salary and exposure score (Medium-category): LLMs.

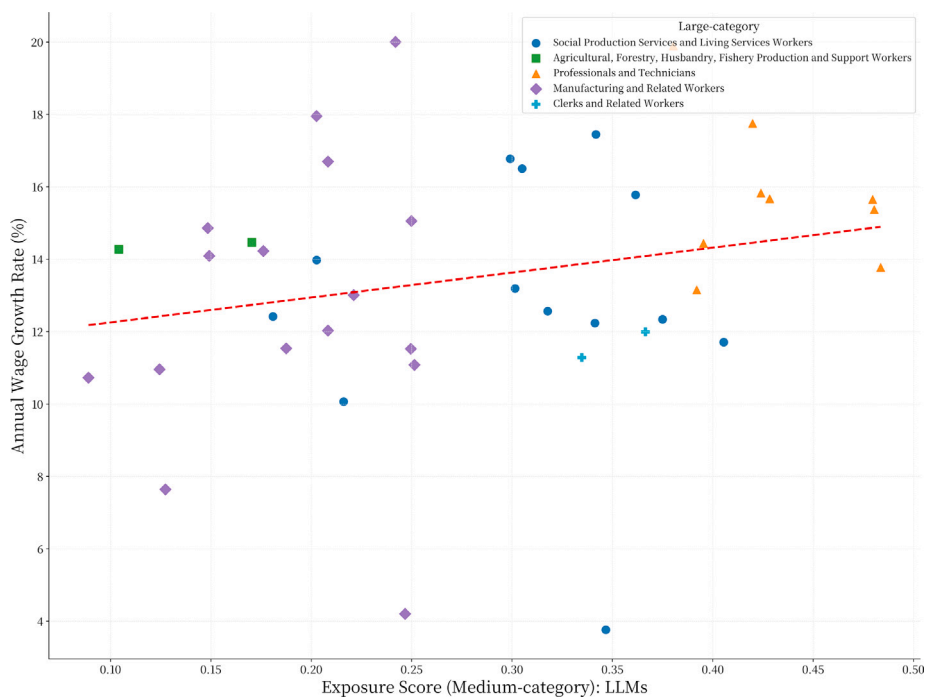


Fig. 4. Annual wage growth rate and exposure score (Medium-category): LLMs.

The analytical results, illustrated in Fig. 5, reveal that cognitive tasks, particularly those of a non-routine nature, are the most significantly impacted by large language models (LLMs). Routine manual tasks are also expected to experience considerable influence, while tasks classified as Interpersonal and Non-routine Manual Physical are less affected by these advanced models.

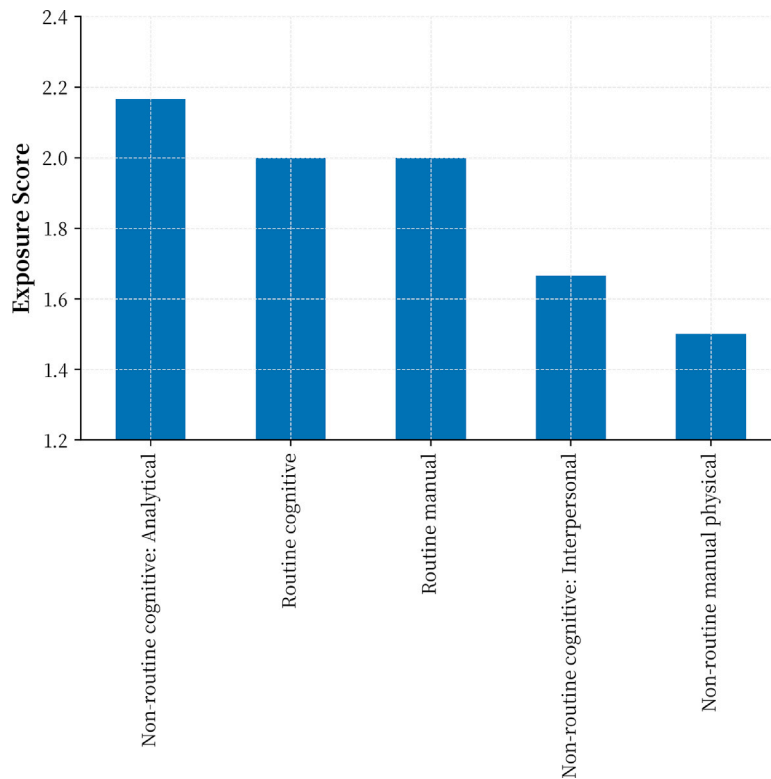


Fig. 5. Composite occupation and LLMs exposure.

This differential impact underscores the nuanced effects of LLMs across task categories. Notably, our findings diverge from the “routinization hypothesis” (Autor et al., 2003), which posits that information and communication technologies (ICT) primarily automate “routine” codifiable tasks traditionally performed by middle-wage workers. In Section 5, we develop a theoretical model to further investigate the mechanisms driving the different occupational exposures of traditional technologies and LLMs.

4.4. Industrial and demographic exposure

As highlighted in the preceding Data section, we leveraged the 2016 CLDS dataset, which encompasses comprehensive industry and occupational information for each individual in employment. Utilizing this dataset, we derived the occupational distribution within each industry. This distribution, when coupled with the exposure scores attributed to the corresponding occupations, enabled the computation of a weighted occupational intensity index for each industry.

Fig. 6 provides insights into substantial variations in the exposure levels of LLMs across distinct industries. Notably, sectors with the highest scores, including education, culture and arts, radio, film and television, exhibit heightened susceptibility. This observation underscores the pronounced influence of the current technological landscape on these domains. Conversely, industries with lower scores, such as agriculture, forestry, husbandry, fisheries, mining, manufacturing and construction—persist in displaying a range of distinctive complexities. Despite the overarching advancements in artificial intelligence across diverse domains, these sectors present enduring challenges. Their intricate nuances, customized requirements, and the enduring necessity for human judgment contribute to their current resilience against full automation and replacement in the foreseeable future.

Armed with industrial exposure data, we proceed to compute demographic exposure by integrating it with the demographic distribution across industries. Relying on the tabulations of census data from 2020 in China, we observe a notable concentration of youth employment within the tertiary sector. To illustrate, during 2020, 42.5% of individuals aged 16–19 and 34.9% of those aged 20–24 were engaged in consumer service industries. Fig. 7 showcases our demographic exposure scores, revealing that LLMs wield a more pronounced influence on the employment prospects of young individuals.

4.5. Vacancies

Figs. 8 and 9 offer a comprehensive overview of potential impacts on future labor demand within China. The presence of a positive correlation between occupational exposure and vacancy share, as depicted in Fig. 8, suggests that occupations with a higher vacancy share may potentially experience more pronounced disruptive effects from LLMs in the future. Contrary to the prevailing

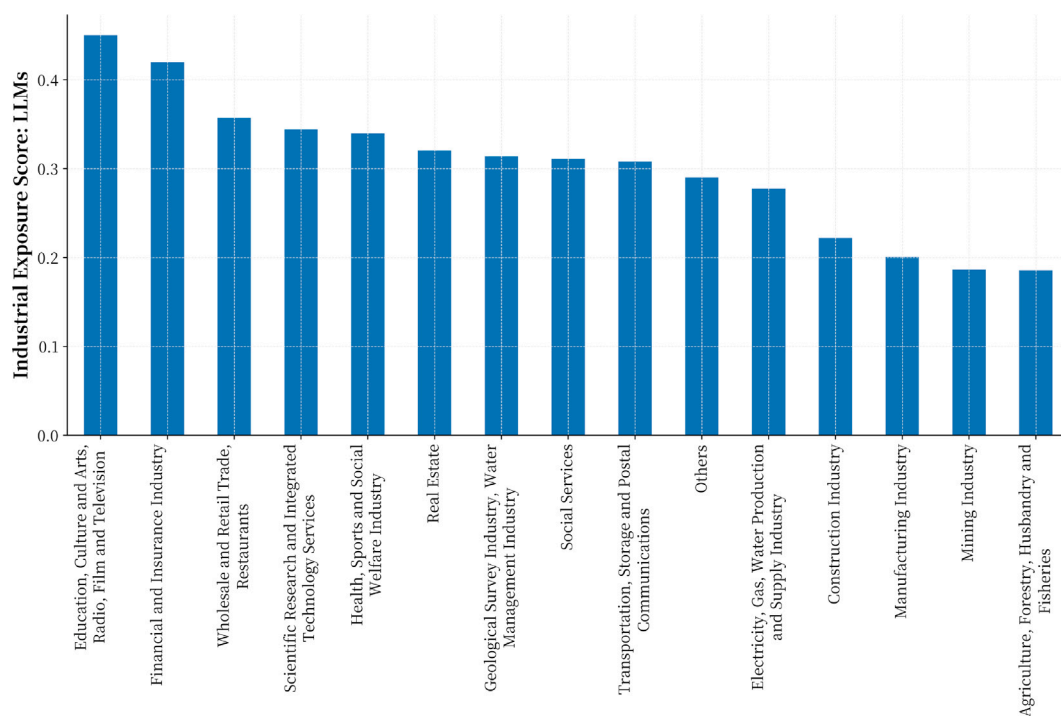


Fig. 6. Industrial exposure score: LLMs.

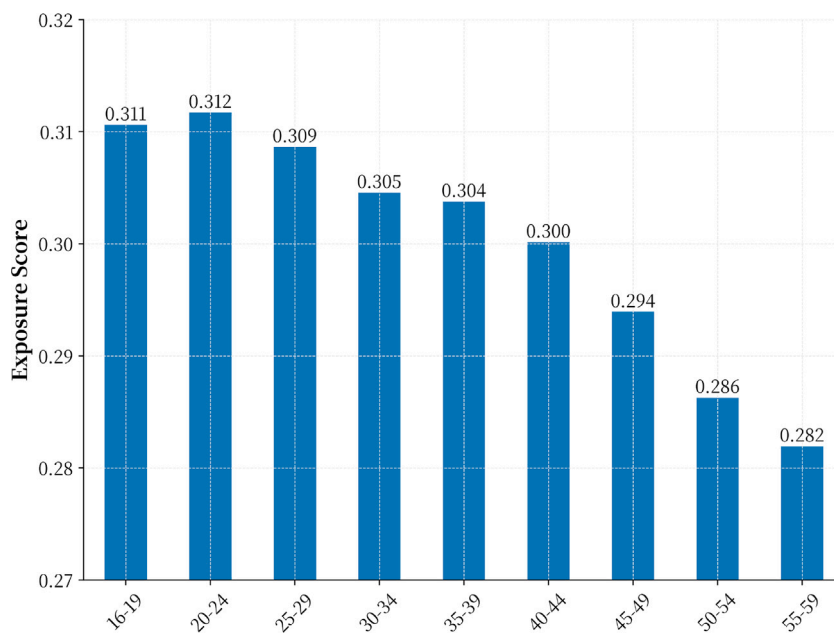


Fig. 7. Demographic exposure score: LLMs.

assumptions, the occupational structure in China tends to exacerbates rather than ameliorate the disruption caused by LLMs. Fig. 9 unveils an additional layer of concern related to the disruptive effects of LLMs. The positive correlation between occupational exposure and growth rate of vacancy share implies that the advancements of LLMs could unexpectedly reverse the prevailing labor demand trend, thereby intensifying the structural unemployment predicament in the future.

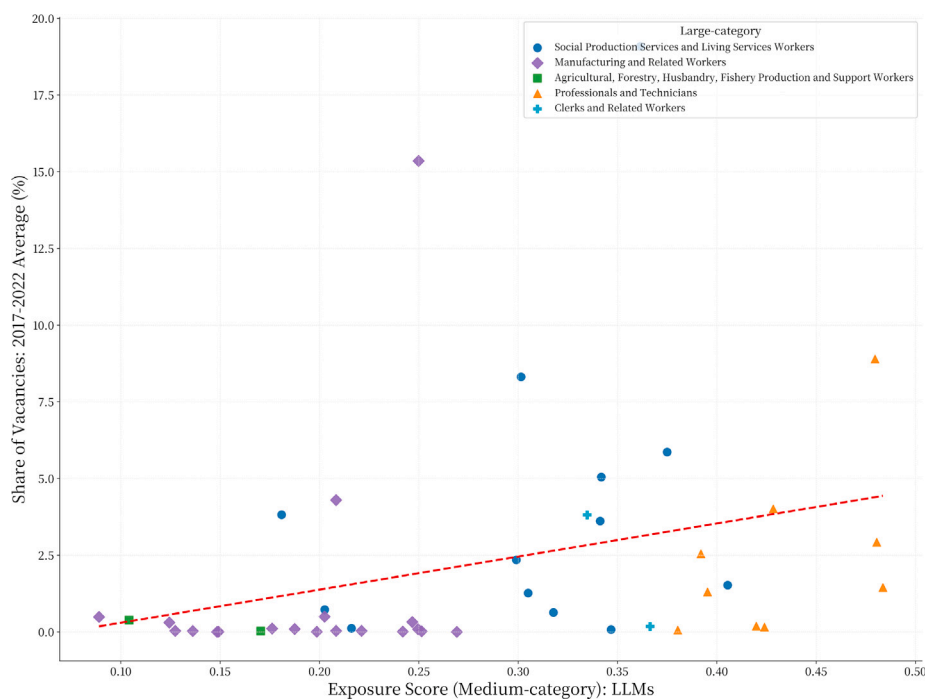


Fig. 8. Share of vacancies and exposure score (2017–2022): LLMs.

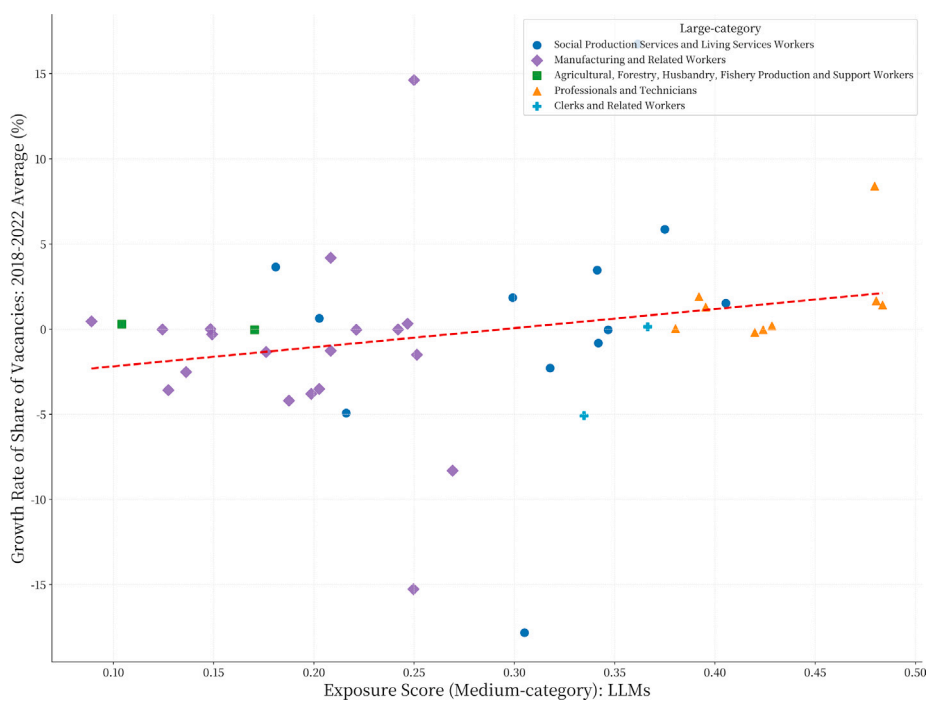


Fig. 9. Growth rate of share of vacancies and exposure score: LLMs.

5. The model

This section introduces a theoretical framework designed to enhance understanding of task automation and its relationship with large language models (LLMs). Drawing on prior empirical findings, we have discerned a predominant impact of automation

on occupations involving non-routine cognitive tasks. This observation marks a departure from earlier conclusions, such as those posited by Autor et al. (2003), which suggested that automation primarily affects occupations characterized by routine tasks. A key insight from our analysis is the pronounced effect of automation on occupations heavily reliant on non-routine cognitive analytical tasks, in contrast to the relatively minimal impact on manual physical tasks. Occupations involving interpersonal tasks demonstrate a moderate degree of susceptibility to LLMs, positioned between cognitive and manual tasks in terms of exposure. Additionally, our findings indicate a correlation between occupations with higher wage and experience premiums and their increased vulnerability to LLMs. This aligns with the recent research by Eisfeldt et al. (2023) and Eloundou et al. (2023), who have identified that technological advancements disproportionately affect workers in the higher echelons of the wage spectrum. To elucidate the varied patterns of occupational exposure to different technological advancements, our proposed theoretical model offers a nuanced perspective on the distinct mechanisms driving occupational automation. This model aims to bridge the gap in understanding the differential impacts of technology across various occupations.

The central thesis of our analysis hinges on the distinct technological attributes inherent to the traditional AI, deep learning AI, and LLMs AI. Essentially, our theoretical model posits that the unique patterns of occupational exposure observed are directly attributable to the underlying technological principles governing automation processes. To effectively articulate this proposition, we integrate the principles of information theory into our task framework. This integration facilitates a more comprehensive understanding of the disparities between various AI technological paradigms, particularly highlighting the differences in productivity between models based on traditional AI and those developed using deep learning AI and LLMs AI technology.

The structure of the theoretical section of our study is methodically organized as follows: Section 1 delineates the environments and competitive equilibrium within a stylized task model. This foundational subsection lays the groundwork for understanding the basic dynamics of task allocation and performance. Section 3 seamlessly integrates traditional AI technology into our task framework. This integration is instrumental in reexamining and substantiating the routinization hypothesis, which asserts that traditional AI technologies are predominantly applied to routine tasks. Section 4 is pivotal as it illuminates the distinct patterns of occupational exposure. It accomplishes this by elucidating the scaling laws characteristic of deep learning AI and LLMs AI technology. This subsection is critical in contrasting the logic of these advanced technologies with that of traditional AI, thereby offering a nuanced perspective on how technological advancements influence various occupational sectors differently.

5.1. The benchmark model

The benchmark model shares common basic settings with the other models, but it lacks task automation originated from the different technologies.

5.1.1. The environment

The unique final good is produced by combining a continuum of tasks represented by the unit interval $[0, 1]$. In particular,

$$Y = \exp \left[\int_0^1 \ln y(i) di \right], \quad (1)$$

where Y denotes the output of a unique final good and $y(i)$ is referred to as the output of task i . We assume all markets to be competitive. Throughout, we choose the price of the final good as the numeraire.

Each task is produced by the following technology function

$$y(i) = A_L \alpha_L(i) l(i) + A_H \alpha_H(i) h(i) + A_M \alpha_M(i) m(i), \quad (2)$$

where A terms represent a factor-augmenting technology, and α terms are the task productivity profiles, designating the productivity of low or high skill workers and models in different tasks. It is important to see that this production function for task services implies that each task can be performed by low or high skill workers and a technological model, but the comparative advantage of skill groups and different model differs across tasks which is captured by the α terms. The differences in comparative advantage play a central role in a Ricardian model of the labor market.

In our analysis, the deliberate use of the term “model” instead of “capital” accentuates our focus on the intangible aspects of AI automation technologies. These models are distinguished by their unique technological logics. To deepen our understanding of the varying technological logics underpinning these automation technologies, we delve into the prominent discourse on “Software 1.0 vs. 2.0”, notably advanced by Andrej Karpathy, Sr. Director of AI at Tesla, in his 2017 Medium post titled “Software 2.0”.

Software 1.0 represents the paradigm prevalent in traditional software development, employing languages such as Python, C++, among others. In this paradigm, programmers craft explicit instructions in the form of code, which the software executes. Each line of code is a directive that specifies a certain aspect of the program, guiding it towards a desired behavior.

Karpathy posits that the advent of neural networks, which are foundational to modern deep learning AI technology, heralds a paradigmatic shift in software development. According to him, Software 2.0 centers on compiling a dataset that defines the “desirable behavior” or the problem to be solved. This is followed by establishing the basic structure or “skeleton code” of the neural network’s architecture. The network is then allowed to optimize itself for the solution. Crucially, Software 2.0 diverges from Software 1.0 in that it does not involve code written by programmers in the traditional sense. Instead, the “code” emerges from computations performed on large datasets.

This insightful discussion about the technological distinctions between traditional AI and deep learning AI or LLMs AI informs our categorization of models’ technologies into three paradigms: traditional AI, deep learning AI, and the LLMs AI.

Our theoretical framework also incorporates two pivotal concepts from information theory. The first is entropy, which aids in defining the complexity inherent in each occupation. The second concept is the Kullback–Leibler divergence (KL-divergence), which is instrumental in specifying the productivity of different models. These concepts are critical for a nuanced understanding of the impact of various technological paradigms on different occupations.

Let us utilize concept of entropy to define complexity of different tasks and comparative advantage of skills. Let X be a discrete random variable with probability mass function $P(x)$. The entropy (or Shannon entropy) of X is

$$H(P) = E_P \left[\log \frac{1}{P(X)} \right] = \sum_x P(x) \log \frac{1}{P(x)}. \quad (3)$$

Entropy is a quantification of uncertainty in a distribution. The higher the entropy of the distribution, the more information is contained in the realization of a random variable it governs. Let us assume that there exists a distribution function P_i which characterize the relevant information in task i . Therefore, the complexity of a task i is defined by entropy $H_i(P)$. Furthermore, we assume that skilled worker has a comparative advantage in the more complex tasks. Specifically, we assume that H_i is evenly distributed on the domain $[0, \bar{H}]$ and $\bar{H} = \max\{H_i\}$. Now we can index task i by their correspondent relative complexity H_i/\bar{H} .

Assumption 1. $\alpha_L(i) = (1 - H_i/\bar{H})^\eta = (1 - i)^\eta$ and $\alpha_H(i) = (H_i/\bar{H})^\eta = i^\eta$, $\eta > 0$.

The [Assumption 1](#) characterize the comparative advantage of skills across tasks. The implication of the assumption is that compared to unskilled workers skilled worker have a competitive edge on the complex tasks.

5.1.2. Competitive equilibrium without task automation

We will characterize the competitive equilibrium without task automation in the following. There will exist a threshold I such that all tasks $i < I$ will be performed by low skill workers, and all tasks $i > I$ will be performed by high skill workers.

Lemma 1. *In equilibrium, there exists a threshold I such that $0 \leq I \leq 1$ and for any $0 \leq i < I$, $h(i) = 0$, for any $I < i \leq 1$, $l(i) = 0$.*

The competitive equilibrium can be characterized by the following four conditions:

1. Law of one price for skills

$$W_L = p(i) A_L \alpha_L(i) \quad \text{for any } i < I, \quad (4)$$

$$W_H = p(i) A_H \alpha_H(i) \quad \text{for any } i > I. \quad (5)$$

2. Goods market (or task market) clearing condition demands:

$$p(i) y(i) = p(i') y(i') = Y. \quad (6)$$

In the interval $i < I$, the following equation is satisfied

$$p(i) A_L \alpha_L(i) l(i) = p(i') A_L \alpha_L(i') l(i').$$

Combining with the condition of the law of one price for skills, we obtain

$$l(i) = l(i') \quad \text{for any } i < I, \quad (7)$$

and

$$h(i) = h(i') \quad \text{for any } i > I. \quad (8)$$

3. No arbitrage condition across skills: the threshold task I can be profitably produced using either skilled or unskilled workers:

$$A_L \alpha_L(I) l(I) = A_H \alpha_H(I) h(I). \quad (9)$$

4. Labor market clearing condition:

$$\int_0^I l(i) di = L, \quad (10)$$

$$\int_I^1 h(i) di = H. \quad (11)$$

By (7), (8), (10), and (11), we obtain that

$$l(I) = L/I, \quad (12)$$

$$h(I) = H/(1 - I). \quad (13)$$

Plugging (12), and (13) into (9), we can obtain

$$I = 1 / \left(1 + \left(\frac{A_H H}{A_L L} \right)^{\frac{1}{1+\eta}} \right). \quad (14)$$

Wage difference is determined by the productivity difference at the threshold

$$\frac{W_H}{W_L} = \frac{A_H \alpha_H(I)}{A_L \alpha_L(I)} = \left(\frac{A_H}{A_L} \right)^{1/(1+\eta)} \left(\frac{H}{L} \right)^{-\eta/(1+\eta)}.$$

The following proposition summarizes the equilibrium allocation and corresponding wage premium.

Proposition 1. *The competitive equilibrium is characterized by a threshold condition I*

$$I = 1 / \left(1 + \left(\frac{A_H H}{A_L L} \right)^{\frac{1}{1+\eta}} \right),$$

and labor allocations across occupations. In the domain $i \in [0, I]$, L/I measure of unskilled workers are allocated to each task, meanwhile, in the domain $i \in [I, 1]$, $H/(1-I)$ measure of skilled workers are allocated to each task.

The corresponding wage premium is

$$\frac{W_H}{W_L} = \left(\frac{A_H}{A_L} \right)^{1/(1+\eta)} \left(\frac{H}{L} \right)^{-\eta/(1+\eta)}. \quad (15)$$

The two terms on the right hand of previous Eq. (15) correspond to two forces in Tinbergen's race: technology and education. The terms $\left(\frac{A_H}{A_L} \right)^{1/(1+\eta)}$ represents the impacts of the skill-biased technology change on the relative demand of skills, otherwise the terms $\left(\frac{H}{L} \right)^{-\eta/(1+\eta)}$ embodies the impacts of the relative supply of skills which is determined by education.

5.2. The model with different AI technologies

In this subsection, let us introduce the automation model with different AI technologies to perform a task.

5.2.1. The comparative advantage of traditional AI

The primary innovation of our theoretical construct is the nuanced elucidation of the comparative productivity advantage of various models in specific tasks, as informed by information theory. This study systematically considers models grounded in three technological paradigms: traditional AI, deep learning AI, and large language models (LLMs) AI. Each model, underpinned by these divergent technologies, has been rigorously trained to manifest optimal task-specific behavior, achieving a level of precision that renders any distinction between the model's performance and an idealized version of task execution virtually undetectable. Consequently, the accuracy of a model's performance is directly proportional to its productivity in automated tasks. Hereafter, we denote all endogenous variable in this economy with an upper dagger.

A fundamental postulate of this research is that the inability to distinguish between a model's output and the theoretically perfect execution of a task denotes competence. This principle is a cornerstone in the behavioral evaluation of machine intelligence, epitomized by seminal tests such as Turing's Imitation Game.

Specifically, we resort to the foundational concept in information theory. KL-divergence, also known as relative entropy, quantifies how different two distributions P and Q are. It is arguably a central pillar of information theory. The KL-divergence between P and Q is

$$D(P \parallel Q) = \sum_x P(x) \log \left(\frac{P(x)}{Q(x)} \right). \quad (16)$$

Intuitively, larger log likelihood ratios $\log \left(\frac{P(x)}{Q(x)} \right)$ reflect distributions that are more different. KL-divergence aggregates these log likelihood ratios by weighting them with respect to their probabilities under a reference distribution. Now we can define a model's KL-divergence D_i where P_i is the information distribution embodied in a model performing task i and Q_i is the information distribution that a best model can achieve.

Now we can resort to we KL-divergence to characterize the productivity profile $\alpha_M(i)$. Let assume that for each task, we have a productivity upper bound \bar{A} , and productivity in task i for a model with traditional AI is

$$\alpha_M(i) = F(\tilde{D}_i) = (1 - \tilde{D}(i)/\bar{D})^\mu, \quad A_m = \bar{A},$$

Let us assume that \tilde{D}_i is evenly distributed on the domain $[0, \bar{D}]$ and $\bar{D} = \max\{\tilde{D}_i\}$, therefore,

$$\alpha_M(i) = (1 - i)^\mu,$$

which implies that the model tends to have a higher productivity in less complex tasks. Furthermore, to ensure an internal solution, we assume that

$$\bar{A} > A_L, \mu > \eta.$$

We also impose a exogenous price of models M as F .

The aforementioned assumption elucidates the underlying technical rationale inherent in traditional AI. Drawing an analogy with the paradigm of software 1.0, this technology facilitates task automation by encoding explicit rules pertinent to a specific task. It is evident that the process of encoding rules is considerably more straightforward in tasks of lesser complexity. In contrast, for more complex tasks where rules are either excessively intricate or ambiguously defined, the codification process becomes substantially more challenging. Hence, our postulation aligns with the routinization hypothesis, which posits that traditional AI models demonstrate a comparative advantage in automating routine or less complex tasks. Consequently, in line with these assumptions, our analysis reaffirms the routinization hypothesis, indicating a tendency for models based on traditional AI to predominantly automate routine tasks.

Proposition 2. *In competitive equilibrium, there exists two threshold \tilde{I}_1 and \tilde{I}_2 such that $0 < \tilde{I}_1 < \tilde{I}_2 < 1$ and for any $i < \tilde{I}_1$, $\tilde{l}(i) = \tilde{h}(i) = 0$, for any $\tilde{I}_1 < i < \tilde{I}_2$, $\tilde{h}(i) = \tilde{m}(i) = 0$, for any $\tilde{I}_2 < i \leq 1$, $\tilde{l}(i) = \tilde{m}(i) = 0$. Furthermore, the wage premium increases.*

Proof. see appendix.

Proposition 2 in our study reveals that models leveraging traditional AI technology are inclined to replace routine tasks, typically executed by unskilled workers. This leads to a reduction in the task domain available to unskilled workers, subsequently widening the wage disparity between skilled and unskilled labor.

5.2.2. The comparative advantage of deep learning AI and LLMs AI technology

In this subsection, let us introduce the automation model with deep learning AI and LLMs AI technology to perform a task. We use the hat operator to indicate endogenous variables in this economy.

At present, in the realm of artificial intelligence, Large Language Models (LLMs) represent the most advantageous form of generative artificial intelligence. These AI systems, designed to predict subsequent words based on preceding text, are further refined through fine-tuning to align with human instructions and preferences. LLMs are distinguished by their foundation in deep neural networks, boasting an expansive range of billions, or even trillions, of parameters in cutting-edge models. This positions both deep learning AI and LLM AI models firmly within the realm of the “Software 2.0” paradigm, characterized by a shift towards data-driven learning and adaptive algorithms.

The advent of transformer models, as introduced by Vaswani et al. (2017), marks a pivotal development in the evolution of LLMs. Transformers incorporate an “attention mechanism” that dynamically assigns varying levels of significance to different words within a text. This innovation significantly enhances the model’s ability to decipher complex patterns and dependencies in language, thereby improving the efficiency and interpretative capabilities of language models.

Now let us assume that deep learning AI and LLMs AI models represent a distinct technological logic. Still we assume that KL-divergence is a measurement of productivity loss. Let assume that for each task, we have a productivity upper bound \bar{A} , and productivity in task i for an AI model are

$$\alpha_M(i) = F(\hat{D}(i)) = (1 - \hat{D}(i))^\eta, \quad A_m = \bar{A},$$

where

$$\hat{D}(i) = L(N, G) = C + \frac{a}{N(i)^\alpha} + \frac{b}{G^\beta}. \quad (17)$$

Let us assume there is a different data accumulation pattern across task. Now assume that $N(i) = \bar{N}$, if $i \in [\underline{I}, \bar{I}]$, and $I < \underline{I} < \bar{I} < 1$ and $N(i) = 0$, if $i \notin [\underline{I}, \bar{I}]$. A key conceptual contribution from the field of AI, particularly relevant to this discussion, is encapsulated in (17), which represents a well-established empirical regularity known as the scaling law. This concept, highlighted in recent AI literature by researchers such as Barnett and Besiroglu (2023), Hoffmann et al. (2022) and Kaplan et al. (2020), articulates a parametric form for Kullback–Leibler divergence. This divergence is expressed as a power law in terms of data (N) and parameter count (G), with C denoting the irreducible loss, and a , b , α , and β are constants. The robustness of these relationships is underscored by their persistence across various orders of magnitude. Furthermore, power-law scaling laws have been identified as a ubiquitous presence in the deep learning AI and LLMs AI domain, with their application extending to areas like auto-regressive modeling and reinforcement learning, and have been the subject of extensive theoretical investigation.

Given that deep learning AI and LLMs AI models exhibit similar power scaling laws, our focus can be directed specifically towards LLMs AI without a significant loss of generality. In the ensuing analysis, we conceptualize technological changes in LLMs as a substantial increase in the parameter count G , facilitating the execution of tasks within the specified domain $[\underline{I}, \bar{I}]$ by LLMs. This approach allows for a nuanced understanding of the impact and potential of LLMs AI in transforming various task domains.

5.3. Discussions

In this subsection, we discuss our model with specific AI Applications. And we also elaborate theoretical and policy implications from theoretical model furthermore.

5.3.1. Discussions with AI applications

The healthcare industry has long been a dynamic testing ground for cutting-edge AI capabilities. Recent advances in large language models (LLMs) have enabled these systems to acquire broader competencies, allowing for application across a wide range of healthcare tasks. Given the pervasive use of text in healthcare practices, healthcare-specific LLMs, trained on extensive medical language data, have demonstrated exceptional performance in processing medical texts, conducting dialogues, making diagnoses, and supporting education and consultation. We focus here on two particular applications: medical diagnosis and report generation.

LLMs have been employed to predict the most likely diagnoses based on medical tests and patients' personal descriptions. These models have proven effective in enhancing diagnostic processes, showing strong generalist abilities across various diseases. Despite their diagnostic potential, LLMs still face significant challenges. The lack of transparency — where clinicians are unable to fully trace the reasoning behind a model's decisions — remains a key concern in medical practice.

LLMs have also demonstrated remarkable potential in generating medical reports, such as radiology reports, discharge summaries, and referral letters. These models excel at synthesizing information from a variety of sources, including electronic health records (EHRs), medical literature, and clinical guidelines, to produce coherent, informative reports. Medical report generation is often a tedious and time-consuming task for doctors; thus, the use of LLMs can significantly alleviate their workload.

While these applications confirm that both medical diagnosis and report generation require highly skilled expertise, LLMs are showing the potential to empower a broader group of workers to perform these expert tasks. As AI continues to evolve, it could ultimately complement and enhance the judgment of professionals, broadening the scope of tasks that non-experts can manage. This trend is applicable beyond healthcare, suggesting that LLMs can enable more workers to perform complex, high-stakes tasks by augmenting their skills and supporting their professional judgment.

Our theoretical model aligns with this vision. We predict that LLMs will reshape the value and nature of human expertise, helping to narrow the income inequality between skill levels. Several recent empirical studies corroborate this perspective. For example, a controlled experiment by [Peng et al. \(2023\)](#) found that GitHub Copilot, a generative AI programming aid, significantly boosted productivity: the group using Copilot completed programming tasks approximately 56% faster than the control group without access to the tool. Similarly, a study by [Noy and Zhang \(2023\)](#) found that access to ChatGPT improved the speed and quality of writing tasks among professionals such as marketers, grant writers, consultants, and managers. The ChatGPT group reduced the time spent on tasks by 40%, and even the least skilled writers in this group performed as well as the median writer in the control group—highlighting a significant boost in writing quality. Moreover, ChatGPT allowed the most capable writers to work faster and enabled the less capable to not only write faster but also to improve the quality of their output. This narrowing of the productivity gap underscores the potential of LLMs to democratize expertise.

To understand the fundamental difference between traditional AI and modern deep learning-based LLMs, we need to focus on the underlying technological logic. Both AI models and machines require software code to execute tasks. In traditional AI, this code follows a rule-based structure: developers must explicitly define each step in the process, particularly for tasks that are repetitive or relatively simple. This makes routine, less complex tasks more susceptible to automation.

In contrast, deep learning AI and LLMs follow a data-driven methodology. Developers no longer need to predefine the rules; instead, they feed vast datasets into an algorithm, which then learns to perform tasks. For instance, ChatGPT does not rely on specific, preset rules for generating poetry; it learns this skill through exposure to an enormous amount of language data. This data-driven approach enables deep learning AI and LLMs to tackle tasks that lack clear, codifiable rules—typically found in more complex, skilled domains. Consequently, LLMs are more suited to automating tasks traditionally performed by skilled workers, rather than simple, rule-based jobs.

5.3.2. Theoretical and policy implications

Our model incorporates entropy-based information theory into a stylized task-based framework. This task-based model is well-suited to characterizing comparative advantages. To analyze the comparative advantage of skills and technologies, we need a consistent way to define the productivity profiles of skills and technologies across occupations. Specifically, we characterize task complexity using the concept of entropy and define the relative productivity of skills across occupations with varying complexity levels.

Furthermore, the introduction of KL-divergence allows us to model the relative efficiency of AI models across occupations. The central idea is that cognitive tasks, at their core, involve making predictions—such as forecasting GDP growth, predicting stock prices, or generating responses in the case of LLMs. Intuitively, the productivity of cognitive tasks is determined by prediction accuracy, which we represent through KL-divergence. KL-divergence is also used in the “neural scaling laws” to measure an AI model's prediction accuracy, which gives task performance a solid foundation by directly borrowing well-established AI technological principles and bridging them with information theory.

Our theoretical framework uniquely integrates information theory into a stylized task model. This integration accomplishes dual objectives. Firstly, it enables the application of concepts like entropy and KL-divergence to quantify the complexity of tasks and the efficacy of algorithms in specific tasks. Secondly, information theory acts as a conduit, linking economic models with the body of literature in computer science. This synthesis provides a robust micro-foundation for the understanding of deep learning AI and LLMs AI technology.

In our theoretical model, the key distinction between traditional AI and LLMs as automation technologies lies in their operational logic: traditional AI depends on pre-programmed rules to perform tasks, while LLMs leverage neural scaling laws and data-driven methods. This fundamental difference in technological logic results in varying comparative advantages across tasks. Traditional AI

excels at performing less complex tasks typically carried out by unskilled workers. The automation of such tasks increases wage premiums between skilled and unskilled workers by eroding the position of unskilled workers.

In contrast, LLMs, powered by neural scaling laws and vast amounts of data and computational resources, exhibit a comparative advantage in data-intensive and complex tasks, usually performed by skilled workers. This shift erodes the expertise position of skilled workers and leads to a more flattened wage distribution between skill levels.

The overarching insight from our theoretical model is that algorithms, based on distinct technological logic, are predisposed to disrupt different types of tasks. These variances in technological underpinnings will inevitably result in divergent outcomes in terms of productivity and income distribution. This distinction is why the progression of LLMs AI technology will generate outcomes that deviate from those predicted by the routinization hypothesis. Fundamentally, the influence of LLMs AI technology on labor markets is anticipated to be distinctly different from that of traditional AI. In essence, this era of technological advancement represents a departure from past trends.

6. Discussions and conclusions

The utopian vision of the information age anticipated that computers would democratize information and flatten economic hierarchies. However, this vision has not materialized; instead, the opposite has occurred. Labor income inequality has widened significantly, with computerization facilitating the unprecedented concentration of decision-making power among elite experts, who now have access to abundant and inexpensive information (Deming, 2021). Our theoretical model predicts that artificial intelligence, by leveraging vast datasets and computational power to synthesize information and rules, could enable a broad range of workers equipped with foundational training to perform high-stakes decision-making tasks. These tasks, traditionally monopolized by elite experts — such as doctors, lawyers, software engineers, and college professors — may become accessible to more individuals. In this regard, our theoretical perspective suggests that artificial intelligence has the potential to reduce income inequality, contrary to the trends observed with earlier technologies.

Ge and Yang (2014) documented that between 1992 and 2007, during a period of rapid economic growth in China, the average real wage increased by 202%, accompanied by a sharp rise in wage inequality. While job polarization did not occur in China, the wages of elite experts rose dramatically, indicating that information technology has played a significant role in exacerbating income inequality. Our theoretical model applies to this context as well, reflecting the mechanisms by which technological advancements affect wage distribution. These results align with the broader discussions on technological impacts, as noted by Autor (2024). To summarize, our theoretical model aims to interpret the differing implications of various technologies on income inequality based on their distinct features. In this sense, we believe the model captures at least part of the momentum shaping labor market demand in China.

In terms of policy implications, our theoretical findings for artificial intelligence should be viewed as scenario analyses rather than definitive predictions. The impact of any technology depends not only on its inherent features but also on institutional responses. The first policy implication of our model is the disruptive impact of advanced AI technologies, such as LLMs, on expert labor. AI has the potential to reduce scarcity by empowering more workers to perform expert-level tasks. However, this potential will not be realized automatically. For example, a recent experiment found that providing AI assistance to radiologists did not improve diagnostic quality, even though AI predictions were at least as accurate as those of two-thirds of the doctors studied (Agarwal et al., 2023). This outcome occurred because doctors often overrode AI predictions, highlighting their lack of understanding of how to use the tools effectively. Thus, a key policy orientation for the future is to develop training programs that enable workers to effectively utilize AI tools.

The second policy implication concerns the substantial uncertainties surrounding the technological impacts on labor markets, as evidenced by the unrealized utopian vision of the information age. To design effective policy tools, it is crucial to systematically track labor market demand conditions. Unlike the U.S., which has a system like ONET, China currently lacks a comprehensive statistical framework to monitor occupational demand. Establishing such a system should be a priority, given the impending disruptive waves of AI technologies.

As a general-purpose technology, the introduction and proliferation of large language models constitute a substantial technological upheaval with significant implications for the overall economy. This paper employs measures of occupational exposures to LLMs, in conjunction with aggregate assessments of occupational composition, to evaluate the potential impact of LLMs on labor market within China's economy. Our findings suggest that the potential effect of the release of LLMs could have a substantial effect on the Chinese labor market, potentially altering previous labor demand trends. LLMs indeed represent a noteworthy shock to the Chinese labor market. Furthermore, we develop a theoretical model to provide a deeper understanding of the reasons behind occupational exposure to LLMs, which diverges from the prediction of the routinization hypothesis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful to Jin Li and Xiaodong Zhu for their insightful comments and to the anonymous reviewers for their constructive feedback that helped improve the paper. Qin Chen and Jinfeng Ge acknowledge financial support from the Key Program of the National Natural Science Foundation of China (Grant No. 71933001). Xingcheng Xu and Huaqing Xie are supported by the Shanghai Artificial Intelligence Laboratory. The views expressed in this paper are those of the authors and should not be interpreted as reflecting the views of the National Natural Science Foundation of China or other affiliated institutions.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.chieco.2025.102413>.

Data availability

Data will be made available on request.

References

- Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of Economic Literature*, 40(1), 7–72.
- Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. vol. 4, In *Handbook of labor economics* (pp. 1043–1171). Elsevier.
- Acemoglu, D., & Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6), 1488–1542.
- Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2), 3–30.
- Acemoglu, D., & Restrepo, P. (2022). Tasks, automation, and the rise in us wage inequality. *Econometrica*, 90(5), 1973–2016.
- Agarwal, N., Moehring, A., Rajpurkar, P., & Salz, T. (2023). *Combining human expertise with artificial intelligence: Experimental evidence from radiology: Technical report*, National Bureau of Economic Research.
- Alemi, A. A., Fischer, I., Dillon, J. V., & Murphy, K. (2016). Deep variational information bottleneck. arXiv preprint arXiv:1612.00410.
- Autor, D. (2024). *Applying AI to rebuild middle class jobs: Technical report*, National Bureau of Economic Research.
- Autor, D. H., Katz, L. F., & Kearney, M. S. (2006). The polarization of the US labor market. *American Economic Review*, 96(2), 189–194.
- Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, 118(4), 1279–1333.
- Barnett, M., & Besiroglu, T. (2023). Scaling transformative autoregressive models.
- Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work: Technical report*, National Bureau of Economic Research.
- Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn and what does it mean for occupations and the economy? vol. 108, In *AEA papers and proceedings* (pp. 43–47). American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., et al. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv preprint arXiv:2303.12712.
- Cover, T. M. (1999). *Elements of information theory*. John Wiley & Sons.
- Deming, D. J. (2021). *The growing importance of decision-making on the job: Technical report*, National Bureau of Economic Research.
- Du, Z., Qian, Y., Liu, X., Ding, M., Qiu, J., Yang, Z., & Tang, J. (2022). GLM: General language model pretraining with autoregressive blank infilling. In *Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: long papers)* (pp. 320–335).
- Eisfeldt, A. L., Schubert, G., & Zhang, M. B. (2023). *Generative ai and firm values: Technical report*, National Bureau of Economic Research.
- Eloundou, T., Manning, S., Mishkin, P., & 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. W., Raj, M., & Seamans, R. (2018). A method to link advances in artificial intelligence to occupational abilities. vol. 108, In *AEA papers and proceedings* (pp. 54–57). American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.
- Felten, E. W., Raj, M., & Seamans, R. (2023). Occupational heterogeneity in exposure to generative ai. Available At SSRN 4414065.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280.
- Furman, J., & Seamans, R. (2019). AI and the economy. *Innovation Policy and the Economy*, 19(1), 161–191.
- Ge, S., & Yang, D. T. (2014). Changes in China's wage structure. *Journal of the European Economic Association*, 12(2), 300–336.
- Goldfarb, A., Taska, B., & Teodoridis, F. (2023). Could machine learning be a general purpose technology? a comparison of emerging technologies using data from online job postings. *Research Policy*, 52(1), Article 104653.
- Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2018). When will AI exceed human performance? Evidence from AI experts. *Journal of Artificial Intelligence Research*, 62, 729–754.
- Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D. d. L., Hendricks, L. A., Welbl, J., Clark, A., et al. (2022). Training compute-optimal large language models. arXiv preprint arXiv:2203.15556.
- InternLM-Team (2023). *InternLM: A multilingual language model with progressively enhanced capabilities*. GitHub, <https://Github.Com/InternLM/InternLM-Techreport/Blob/Main/InternLM.Pdf>.
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling laws for neural language models. arXiv preprint arXiv:2001.08361.
- Katz, L. F., & Murphy, K. M. (1992). Changes in relative wages, 1963–1987: supply and demand factors. *The Quarterly Journal of Economics*, 107(1), 35–78.
- MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge University Press.
- Maćkowiak, B., Matějka, F., & Wiederholt, M. (2023). Rational inattention: A review. *Journal of Economic Literature*, 61(1), 226–273.
- Meindl, B., Frank, M. R., & Mendonça, J. (2021). Exposure of occupations to technologies of the fourth industrial revolution. arXiv preprint arXiv:2110.13317.
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Available At SSRN 4375283.
- OpenAI (2023). GPT-4 technical report. ArXiv abs/2303.08774.
- Peng, S., Kalliamvakou, E., Cihon, P., & Demirel, M. (2023). The impact of ai on developer productivity: Evidence from github copilot. arXiv preprint arXiv:2302.06590.

- Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1), 3–55.
- Shwartz Ziv, R., & LeCun, Y. (2024). To compress or not to compress—self-supervised learning and information theory: A review. *Entropy*, 26(3), 252.
- Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3), 665–690.
- Tishby, N., Pereira, F. C., & Bialek, W. (2000). The information bottleneck method. arXiv preprint physics/0004057.
- Tishby, N., & Zaslavsky, N. (2015). Deep learning and the information bottleneck principle. In *2015 IEEE information theory workshop (ITW)* (pp. 1–5). IEEE.
- Tolan, S., Pesole, A., Martínez-Plumed, F., Fernández-Macías, E., Hernández-Orallo, J., & Gómez, E. (2021). Measuring the occupational impact of ai: tasks, cognitive abilities and ai benchmarks. *Journal of Artificial Intelligence Research*, 71, 191–236.
- Van Reenen, J. (2011). Wage inequality, technology and trade: 21st century evidence. *Labour Economics*, 18(6), 730–741.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Webb, M. (2019). The impact of artificial intelligence on the labor market. Available At SSRN 3482150.
- Xu, A., & Raginsky, M. (2017). Information-theoretic analysis of generalization capability of learning algorithms. *Advances in Neural Information Processing Systems*, 30.
- Zeng, A., Liu, X., Du, Z., Wang, Z., Lai, H., Ding, M., Yang, Z., Xu, Y., Zheng, W., Xia, X., et al. (2022). GLM-130B: An open bilingual pre-trained model. arXiv preprint arXiv:2210.02414.
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., et al. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223.