

The Measurement of AI Exposure and Its Impact on Labor Demand in China: Evidence from Large Language Models

Zhang Dandan^{a,b}, Yu Hang^{a,b}, Li Lixing^{a,b}, Hu Jiayin^{a,b,c}, Mo Yiqing^b and Li Hongbo^b

(a. China Center for Economic Research, Peking University; b. National School of Development, Peking University; c. Institute of Digital Finance, Peking University)

Summary: This paper investigates the impact of rapidly developing Large Language Model Artificial Intelligence (LLM-AI) technology on the Chinese labor market. Utilizing a large dataset of online job postings, the authors construct an index of "LLM-AI exposure" for various occupations and analyze its effects on labor demand, as well as the education and salary requirements of these positions.

Much of the existing literature on technology exposure relies on data and occupational classifications from the United States, such as the O*NET database. However, applying these US-centric indices directly to analyze the Chinese labor market can be problematic due to potential inconsistencies in how occupations are defined and the specific tasks they entail across countries. To address these limitations and better understand the impact of new AI technologies on China's labor market, this paper adopts a novel approach by directly utilizing the detailed descriptions of Chinese online job posts. This allows us to construct an LLM-AI technology exposure index tailored to the Chinese context. This research provides timely empirical evidence on whether and to what extent AI, particularly LLMs, is replacing human jobs in China. The study period covers the rapid advancements and increased prominence of LLMs.

The primary data source is a randomly sampled dataset of 1.25 million online job postings from Zhaopin.com between January 2018 and May 2024. The authors deploy GPT-4 to assess the exposure of each detailed work activity and task to LLM-AI technology. They then correlate the DWAs and Tasks present in each job advertisement, weighted by their importance within the job, with these established exposure scores. By aggregating these scores at the occupational level, specifically using the O*NET 6-digit occupation codes, they construct the "DWA-based LLM-AI exposure index" and the "Task-based LLM-AI exposure index".

Several key findings of this study are as follows. (1) The average LLM-AI exposure of newly posted jobs show a decreasing trend over the sample period. (2) Occupations with the highest LLM-AI exposure are predominantly knowledge-intensive white-collar jobs such as accountants, editors, salespeople, and programmers. (3) There is a negative relationship between LLM-AI technology exposure and labor demand. (4) High LLM-AI exposure is associated with lower salary increases and a larger within-occupation salary gap. Occupations with higher LLM-AI exposure tended to have higher requirements for education and work experience. (5) City-level analysis shows that cities with higher average LLM-AI exposure experienced more significant contractions in overall labor demand, greater wage inequality, and increased demands for education and experience. The findings contrast with earlier research on AI adoption in the US, which finds an increase in demand for AI-related jobs without significant negative impacts on overall employment or wages.

The major innovations of this paper include the direct construction of LLM-AI exposure using Chinese online job post data, a timely analysis of the labor market impact of LLMs, an analysis of the adaptability of the Chinese labor market, and granular analysis at both occupational and city levels. The study's findings contrast with those in the US, suggesting that LLM-AI technologies may be having a more pronounced negative impact on labor demand for highly exposed occupations in China compared to the US. This highlights the importance of understanding the specific context of technological adoption in different countries. The authors emphasize the need to strengthen the adaptability of China's labor market and implement supportive policies to navigate the changes brought about by advancing LLM-AI technologies.

Keywords: artificial intelligence; large language models; labor demand; employment; ChatGPT

JEL Classification: E20, J01, J24, O3

中国人工智能技术暴露度的测算 及其对劳动需求的影响*

——基于大语言模型的新证据

张丹丹 于航 李力行 胡佳胤 莫怡青 李泓宇

摘要:大语言模型人工智能技术发展迅速。本文利用2018年1月~2024年5月之间发布的125万条在线招聘信息,基于对岗位工作任务的具体描述,构建了各职业“大语言模型人工智能技术暴露度”,估计了人工智能技术对劳动力需求以及对岗位学历和薪资等要求的影响。研究发现,在样本期内,中国劳动力市场上新增职位的大语言模型人工智能技术暴露度呈现降低的趋势;暴露度较高的职业主要是对受教育程度要求较高和薪资较高的白领职业,包括会计、编辑、销售及程序员等。基于职业层面的工具变量的回归结果显示,技术暴露度与劳动力需求呈现负相关关系,高暴露度的职业薪资增幅下降、职业内部的薪资差距变大,对教育和工作经验方面的要求也越高。本文强调,中国劳动力市场对新技术的适应性有待加强,建议在加大对大语言模型人工智能技术的研发投入的同时,降低应用门槛,鼓励企业和劳动者使用新技术。

关键词:人工智能 大语言模型 劳动需求 就业 ChatGPT

一、引言

人工智能(Artificial Intelligence,简称AI)被广泛视为继机械化、电力和信息技术之后的第四次工业革命。随着近年来大语言模型人工智能(Large Language Model Artificial Intelligence,简称LLM)技术的兴起,人工智能技术可能对高技能劳动者带来直接冲击,引发了关于岗位替代的广泛担忧。考察岗位替代问题,最常见的做法是对不同职业受到技术影响的程度进行测算,从而在职业层面构建“技术暴露度”,其核心是通过比较技术和人所具备的技能或者所能完成任务的相似度,从而估算出在多大程度上该技术能够完成原本由人工来完成的工作。美国劳动力市场上的职业具备较好的标准化基础,为构建指数提供了便利。例如,职业信息网络数据库基于SOC职业代码^①,从“任务”“工作活动”“技能”等不同维度对不同职业进行了详细的刻画。现有文献基于美国劳动力市场的职业特征,针对各种数字技术,构建了一系列的技术暴露指数(弗雷、奥斯本,2017;韦布,2020;费尔滕等,2021;埃隆杜等,2024)。以埃隆杜等(2024)为例,该文估计了新一代人工智能技术Generative Pre-trained Transformer(GPT)模型的技术暴露度,发现大约80%的美国工人至少有10%的工作内容可能受到影响。在研究技术进步对中国劳动力市场的影响时,文献大多是将中国的职业与美国职业名称进行匹配,利用基于美国数据构建的职业特征来分析中国劳动力市场的变化,因而可能存在不一致的问题。

为了克服这一弊端,并捕捉人工智能新技术对我国劳动力市场带来的影响,本文借鉴埃隆杜等(2024)的思路,采用来自中国的线上招聘大数据,根据岗位工作任务的具体描述,构建了各职业的大语言模型人工智能技术暴露指数。具体而言,本文利用从2018年1月~2024年5月期间发布在智联招聘平台的招聘广告中随机抽取的125万条信息,结合O*NET中的2087个“详细工作活动”(Detailed Working Activities,DWA),构建了职

收稿时间:2024-3-29;反馈外审意见时间:2024-6-6、2024-11-27;拟录用时间:2025-4-18。

*本文得到国家社会科学基金重点项目(22AJY002)、国家自然科学基金青年项目(72303009)、国家高端智库项目的资助。作者感谢匿名审稿人的宝贵意见,智联招聘提供的数据支持,刘诗尧对课题前期工作的重要贡献,以及蔡昉、姚洋、都阳、陆铭、梁建章、李强、Xin Meng、Alberto Rossi、陈沁、沈水生等提供的修改建议。李力行为本文通讯作者。

业层面的人工智能技术暴露指数。这一做法不需要对中国和美国的职业名称进行匹配,从而能够解决跨国职业匹配中潜在的不一致问题。随后,本文分别在职业和城市层面估计了大语言模型人工智能技术暴露指数带来的劳动力市场冲击,并使用美国相关暴露指数作为工具变量以克服潜在的内生性问题。

本文发现,第一,2019~2024年间新增职位的大语言模型人工智能技术暴露度呈现加速下降的态势。我们将职业按暴露度高低分为4组后,发现暴露度最低的职业组,其劳动需求呈现增长趋势,而暴露度最高的组出现下降趋势,这可能反映出新技术对高暴露度的劳动力产生了更强烈的替代效应。第二,暴露度最高的20个职业中包括了会计、编辑、销售及程序等知识密集型的白领工作,工作内容涉及大量的文本处理和资料整合,其被替代的风险较大。第三,回归分析显示,大语言模型人工智能技术暴露度越高,该职业需求缩减更为明显、内部薪资差异加剧,且可能会提出更高的针对受教育程度和工作经验的招聘要求。

本文的贡献主要体现在以下3个方面。第一,本文直接使用中国招聘广告文本中的详细工作内容描述来构建技术暴露指数,而此前的研究主要依赖中国和美国职业的匹配,并根据美国职业的任务内容进行分析。相比而言,本文的方法能够更准确地反映中国劳动力市场特征。第二,本文以月度数据展示了岗位需求变化的趋势,考察了大语言模型人工智能技术暴露指数对职业招聘需求以及相应岗位要求的影响,为2023年以来大语言/生成式人工智能流行之后“人工智能替代工作岗位”的相关讨论提供了相关证据。第三,本文提供了中国劳动力市场面对新技术进步的适应性反应的相关证据,为现阶段我国针对人工智能新技术的政策讨论提供了参考。

文章余下部分的安排如下:第二部分回顾相关文献;第三部分详细介绍本文使用的数据和技术路径;第四部分展示大语言模型人工智能技术暴露指数的计算结果,采用描述和计量分析的方法估计其对职业需求和职业招聘要求变化的影响,并描述了企业的应对状况;第五部分进行跨国对比分析,并提出政策建议。

二、文献综述

本文的工作主要包括两部分:一是构建基于中国劳动力市场的大语言模型人工智能技术暴露度;二是以该指标为基础估计人工智能技术对劳动力需求的影响。本节分别就这两个问题综述相关的文献并讨论本文与已有文献的关系。

(一)职业的技术暴露度指标

根据奥特等(2022)的分类,技术发展具有劳动增效(labor-augmenting)和劳动自动化(labor-automating)两种特性,即补充和替代。虽然理论上可以区分补充与替代,但实证研究中对二者的区分却困难重重,因为它们都与要素价格、市场结构等因素紧密相关。因此,文献一般都止步于“技术暴露度”的构建来衡量技术对工作任务或岗位的潜在影响。按照斯塔内娃和埃利奥特(2023)的总结,技术暴露度的测算存在5种思路,包括基于工作任务、专利、岗位招聘、技能标尺、技能等进行新技术影响的评估。这些思路的核心是结合标准化量表、专家评分、自然语言处理、模型预测、众包等手段,比较某种技术的潜在应用与人所具备的技能或者所能完成任务的相似度,从而估算出在多大程度上该技术能够完成人工进行的工作。

其中,基于工作任务的做法最为常见,其借鉴了奥特等(2003)提出的“基于任务的工作模型”,把职业分解为若干所要完成的具体任务,在任务层面考察技术对人的替代性,然后在加总到职业层面。在这方面,O*NET数据已成为广泛使用的资源,它涵盖了大约1000种职业对应的2087种工作任务和19265种详细任务要求。利用O*NET数据,弗雷和奥斯本(2017)结合专家评估与量表匹配,估算了702种职业被自动化替代的概率。布林约尔夫森和米切尔(2017)基于23项评估标准来评估各工作任务对机器学习的适应性。韦布(2020)研究了人工智能专利与O*NET工作描述的能力匹配来评估各职业的人工智能敏感性。费尔滕等(2019)集中研究了自2010年开始的九大人工智能技术应用领域,并利用众包方式查找与O*NET的52种能力标准的关联,构建了每种职业的人工智能技术暴露指数。而在ChatGPT广受关注的2023年,埃隆杜等(2024)利用专家评分与ChatGPT大型语言模型构造了职业层面的大语言模型暴露指数。费尔滕等(2023)在费尔滕等(2019)的研究基础

上,进一步提出了大语言模型暴露指数,详细描述了774种职业受到的影响程度。本文主要就是根据这支文献,尤其是埃隆杜等(2024)的思路,基于中国招聘数据来构建技术暴露指数。

需要注意的是,所谓的人工智能技术暴露指数,仅表示某职业的工作内容与当前人工智能技术的匹配程度,并不能直接解读为该职业被人工智能技术替代的可能性。如奥特等(2022)所解释,新技术带来的影响是双面的,它既可能带来更多的工作机会,也可能替代某些工作。职业是否会被人工智能技术替代,很大程度上取决于在面对新技术时,企业和个体劳动力能够做出怎样的适应与调整。为此,要回答劳动力市场对人工智能技术的反应,需要将人工智能技术暴露度与实际的就业数据进行匹配,进而对其就业冲击进行判断。

此外,一些研究也采用了文本分析的方法,考察技术进步之下职业的变迁,这类研究一般需要利用到职业大典。例如,奥特等(2022)用专利文本与人口普查中用于界定新职业的工作内容微观描述的重合度衡量新技术对劳动需求产生的互补/增强效应;与此同时,该文也用专利文本与编撰职业大典过程中对宏观职业称谓的描述之间的重合度,度量新技术对劳动供给的替代程度。利用中国的职业大典,王永钦和董雯(2023)、胡涟漪等(2024)分析了职业大典中的职业文本描述或是职业在汉语词典中的释义文本,以此来考察职业的变迁。我们认为,尽管中国的职业分类大典对不同职业的具体工作内容进行了描述,但劳动力市场上实际使用的职业名称未必按照职业大典进行定义。相比而言,招聘广告文本是对具体工作内容更具时效性的描述。因此,本文选择使用招聘数据进行分析。

另外,除了构建技术暴露指数,文献还积极寻找采纳数字技术的一些外生冲击,在行业、地区等层面衡量技术的影响。这方面典型的研究是对工业机器人影响的考察。例如,在研究工业机器人对中国劳动力市场影响的文章中,文献往往利用Bartik工具变量的构造思路,将不同行业的工业机器人的运行数量与各地各行业的就业占比进行交乘,得到各地工业机器人数量的预测值,作为解释变量(陈媛媛等,2022);或是采用美国各行业的工业机器人渗透率来预测中国企业层面的机器人渗透度指标(王永钦、董雯,2020)。就人工智能技术而言,其应用的场景和载体比较多元,无法找到工业机器人这样的外生冲击,因此难以利用这类方法开展研究。

(二)技术进步对就业的影响

基于“任务”的技能衡量方式的提出为学术界研究技术革新对就业的影响提供了新的维度。从20世纪70年代开始,众多研究表明,计算机技术的进步主要替代了中等技能并主导“常规任务密集型”的劳动力,导致这部分劳动者的就业机会和收入减少(奥特等,2003)。与此相反,低技能劳动者,尤其是那些从事简单非常规劳动的劳动者,受此技术影响较为有限。更进一步地,高技能劳动者从事的复杂抽象任务得益于计算机的辅助,其薪酬和就业机会均有所增长。这种就业“极化”趋势在技术进步的背景下得到了更深入的探讨(阿西莫格鲁、奥特,2011;奥特、多恩,2013;古斯等,2014;阿西莫格鲁、雷斯特雷波,2019)。

工业机器人作为另一种广受关注的数字技术也受到了学术界的高度关注。阿西莫格鲁和雷斯特雷波(2019)、格雷茨和迈克尔斯(2018)、贝森(2019)和阿内利等(2021)等研究针对工业机器人技术对就业数量及工人福利进行了深入分析,并探讨了其对退休和生育行为的可能影响。不同于计算机技术,这些研究普遍认为工业机器人技术在替代就业方面的影响更为显著。关于工业机器人技术对中国劳动力市场的影响,可以参考王永钦和董雯(2020)、李磊等(2021)、闫雪凌等(2020)以及琼特拉等(2022)的相关研究。这些研究发现,机器人应用对劳动力需求产生明显的替代效应,显著减少了地区外来劳动力的迁入,降低了劳动者工资性收入,这些负面影响主要来自低技能劳动者。此外,机器人的应用也促进了劳动者的岗位转换,引致劳动者从繁重向非繁重、常规向非常规岗位转换(王林辉等,2023)。

就人工智能技术而言,弗雷和奥斯本(2017)的研究重点在于判定哪些职业更易受到人工智能和大数据驱动的技术进步影响。他们首先对O*NET数据库中的70种职业进行分类,然后利用高斯过程分类器估算其余632种职业的自动化风险,发现在美国约有47%的职业面临着高达70%的自动化风险。与此相对,阿恩茨等(2016)认为技术进步会改变工作性质和所需技能,而不仅仅是产生替代,因此弗雷和奥斯本(2017)可能高估了自动化的影响。该文认为,实际上只有9%的美国劳动者从事的工作高度可能被人工智能所替代。阿西莫

格鲁等(2022)使用2007~2018年的Burning Glass Technologies数据^②,探讨了职业受人工智能技术影响程度与相关职位空缺之间的关系。他们发现,随着某一职业受到人工智能影响的程度增加,与之相关的职位空缺也会随之增长。此外,这些职业所需技能的更新速度也更快。

随着大语言模型的出现,研究者得以更为灵活地评估人工智能新技术对职业的影响。OpenAI和宾夕法尼亚大学的团队(埃隆杜等,2024)利用大型语言模型GPT-4进行职业评分,发现约80%的美国劳动者的工作任务在一定程度上受到了大语言模型人工智能技术的影响。费尔滕等(2023)通过对费尔滕等(2018)中人工智能暴露指数的修正,构建了大语言模型人工智能暴露指数,并基于历史招聘数据,评估了大语言模型对美国劳动力市场的具体影响。

针对中国就业市场的研究发现,智能化对中国劳动就业产生了明显的替代作用,女性、低教育劳动者、大龄劳动者、移民等劳动力市场中相对脆弱的群体所受的冲击更大(周广肃等,2021);人工智能技术对职业的替代风险受制于任务属性,其中思维类属性对可替代风险影响最大,年龄越大与受教育年限越短的劳动者可替代风险越高(王林辉等,2022)。在企业层面,人工智能技术的应用通过提升企业市场占有率进而提升企业高学历员工、年轻员工、新员工占比(尹志锋等,2023),改变了企业对常规/非常规劳动力的需求结构(陈琳等,2024),通过促使企业增加高技能劳动力的使用从而提升了生产率(姚加权等,2024)。

三、招聘平台数据与暴露指数测算

(一)在线招聘数据

企业招聘行为反映了其经营模式转变,对新的发展趋势和技术变革有高度敏感性。国外研究常用燃烧玻璃科技公司(Burning Glass Technologies)提供的基于企业在线招聘信息的数据库来研究技术创新对就业的影响,比如赫什拜因和卡恩(2016)和阿西莫格鲁等(2022)。

沿用此前文献的做法,本文使用在线招聘平台数据进行分析。智联招聘成立于1994年,占据了超过30%的在线招聘市场份额。该平台拥有2.3亿活跃个人用户和约600万企业用户,对城镇劳动力市场具有一定代表性。智联招聘数据已被许多研究所采用,例如库恩和沈(2013)使用该数据分析了雇主对于员工的性别偏好,方等(2020)利用该数据分析了新冠肺炎疫情对全球供应链的冲击及其对企业招聘的影响。这些研究都凸显了利用智联招聘数据分析劳动力市场变化的价值。

值得注意的是,在线招聘广告往往更侧重城市中相对年轻化、高学历、中高工资的职位招聘需求,并不具备对整体劳动力市场的代表性^③。相对于常规的劳动力调查数据而言,在线招聘平台数据虽然代表性较差且无法获取供需双方的最终匹配结果,但却具有实时、高频、细颗粒度的优点。更重要的是,在线招聘平台数据能够捕捉劳动力供需双方在成功匹配前的动态过程,尤其是能及时刻画企业端的招聘需求变化,从而为我们探究新技术对劳动力市场的冲击提供了关键的先导变量。就我们所关注的大模型和生成式人工智能等新兴技术对“白领”工作的影响而言,智联在线招聘数据比劳动力调查数据能更好地满足我们的研究需要。

本文的主要分析数据为随机抽取的全国范围内的招聘广告样本,覆盖时间为2018年1月~2024年5月。样本来自于按年分层的随机抽样,加权调整后的总有效样本量约为125万^④,涵盖针对78个职业组别、346个职业的招聘广告^⑤。表1列出了各年度招聘量最大的5个职业及其市场占比。

(二)测算大语言模型人工智能技术暴露指数

参照现有文献特别是埃隆杜等(2024)的方法^⑥,我们利用O*NET数据库中对于工作内容描述的拆分和聚类指标,使用两种不同层级、互为补充的指标——“详细工作活动(DWA)”和“任务(Task)”——来构建大语言模型人工智能技术暴露指数。

图1展示了O*NET数据库中不同层级指标的含义和说明。以“电工”这一职业为例,其具体工作内容可以被拆解成若干个细分的“任务”,例如“准备草图或遵循蓝图,确定电线或设备的位置,确保其符合建筑和安全规范”,以及“提供材料或服务的初步草图或成本估算”等。这些职业下的“任务”均由O*NET通过实地调查问

表1 各年度招聘最多的5个职业及其市场占比(%)

年度排名	2018年	2019年	2020年	2021年	2022年	2023年	2024年1~5月
1	销售代表(批发和制造业) (11.8)	销售代表(批发和制造业) (11.8)	销售代表(批发和制造业) (11.8)	销售代表(批发和制造业) (10.5)	其他所有生产工人 (8.7)	其他所有生产工人 (11.3)	其他所有生产工人 (17.0)
2	销售经理 (8.1)	销售经理 (7.8)	销售经理 (7.2)	销售经理 (5.4)	销售代表(批发和制造业) (8.6)	销售代表(批发和制造业) (7.3)	销售代表(批发和制造业) (5.2)
3	总经理及运营经理 (5.0)	营销经理 (4.5)	营销经理 (4.3)	客户服务代表 (4.2)	客户服务代表 (5.4)	销售经理 (4.6)	销售经理 (4.0)
4	营销经理 (4.8)	总经理及运营经理 (4.3)	客户服务代表 (4.2)	营销经理 (4.0)	销售经理 (4.1)	客户服务代表 (4.5)	人力资源专家 (3.5)
5	执行秘书和执行行政助理 (4.2)	执行秘书和执行行政助理 (3.8)	其他所有生产工人 (3.7)	其他所有生产工人 (3.7)	营销经理 (3.5)	人力资源专家 (4.3)	供应链经理 (3.1)

数据来源:智联招聘抽样数据。

卷和访谈收集整理,总共约有2万项。随后,O*NET再将类似的任务聚类到一起,形成约2000项“详细工作活动”。例如,前述两项任务可以被聚类到同一个详细工作活动——“创造建筑或安装图”之下。在这之上,还有约300项“中间工作活动(IWA)”和约41项“广义工作活动(GWA)”等更为宽泛的指标层级。由此可见,详细工作活动和任务这两个层级是更细化、信息量更丰富、能更好地区分不同职业间差异的细分层级。

图2呈现了本文的技术路线。我们的目标是根据招聘广告中的职位描述来判断每个职位受大语言模型人工智能技术影响的程度,进而加总到职业大类层面,构造一个科学性、标准化、通用型的指数

体系。以路线1为例,我们选择O*NET的2087个详细工作活动(即DWA)作为定义工作内容的基础^⑦。首先,在GPT-4的辅助下,我们对每一个详细工作活动(d)的暴露度进行评分^⑧,记为 s^d 。其次,我们将每个招聘广告 j 中的职位描述文本与全部详细工作活动比对,筛选出所有包含在广告 j 中的详细工作活动(d_1, \dots, d_n)及它们在该职位工作中的重要性权重($w_1, \dots, w_n, \sum_{x=1}^n w_x = 1$)。将所有与 j 有关的DWA暴露度评分(s^{d_1}, \dots, s^{d_n})按重要性加权求和,得到每个岗位 j 的暴露度。最后,对每个职业或职业组别中所有招聘岗位的分数取平均值,得到该职业(组别)的暴露度,从而构造出“基于DWA的大语言模型人工智能技术暴露指数”。

为了优化计算效率,我们从全样本中按月分层随机选取了77000个广告(每月1000个)作为测度样本,在此样本上执行了上述两种技术路线以计算每种职业或职业组别的大语言模型人工智能技术暴露指数。

值得注意的是,这两种技术路线都是从招聘广告职位自身的详细工作内容出发,并不依赖于中国和美国之间职业名称的直接对应关系,因此克服了关于同一职业名称在不同国家有不同内涵、因而不完全可比的挑战,可以为中国的各种职业提供独立的评估。此外,作为参照,我们也将数据中的中国职业名称直接对应到美国标准化职业名称,进而与费尔滕等(2023)中的指数进行匹配,形成了依赖中美职业名称对应关系的Felten-LLM暴露指数^⑨,以便与前述基于DWA和基于任务所构建的指数进行对比。Felten-LLM暴露指数与这两个指数之间的相关系数分别为0.75和0.71(详见《管理世界》网络发行版附录附图1),而基于DWA和Task的两个暴露指数相关系数为0.89。本文将主要展示基于DWA的大语言模型人工智能技术暴露指数的分析结果,基于

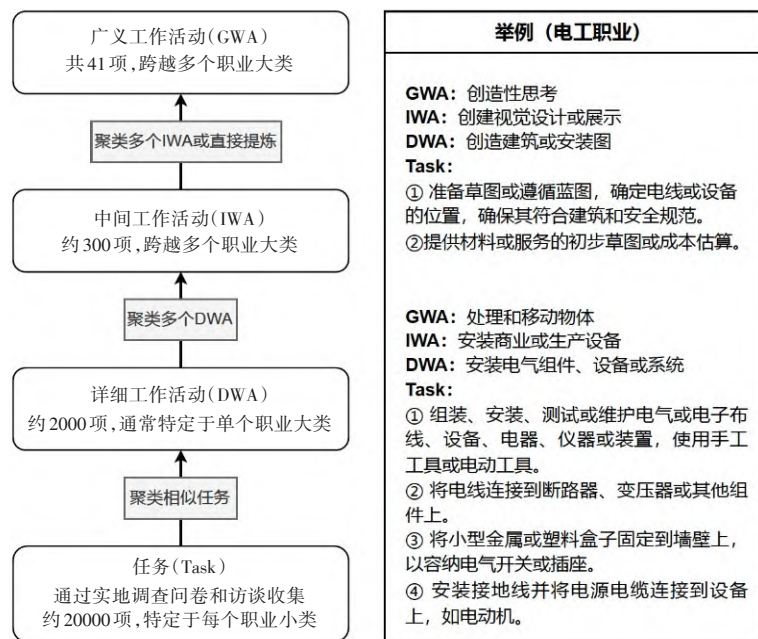


图1 O*NET不同层级指标定义和示例

工作任务的大语言模型人工智能技术暴露指数的分析结果包含在《管理世界》网络发行版附录3中。

四、评估大语言模型人工智能技术对劳动力市场的影响

(一)大语言模型人工智能技术暴露指数的跨时间变化

为了分析劳动力市场需求对新的人工智能技术的适应性变化,我们对上述构建的3个暴露指数逐月进行了加权平均,计算出每个月新增招聘职位对应的暴露指数,并进行了标准化处理和季节性调整^⑩。如图3所示,在过去6年多时间内,中国劳动力市场新发布职位的大语言模型人工智能技术暴露指数均呈现持续下降的趋势,并在2021年初至2022年下半年间加速下滑,2024年又再次出现加速下降。总体来看,新发布职位越来越集中于暴露度低的领域,而暴露度相对高的职位正在逐步减少,显示出大语言模型人工智能技术对劳动力的替代效应可能占据主导作用。

为了考察不同受影响程度的职业在适应性上的差异,图4按照基于DWA的大语言模型人工智能技术暴露指数的高低,将所有职业分为四组别,分别描述归属各组别的新增工作占比的动态走势。结果显示,相对而言暴露指数最低的职业,其劳动需求呈现相对较强的增长态势,而在暴露指数较高的职业中,劳动需求均呈现相对逐渐下降的趋势,特别是在暴露度最高的一组中下降最为显著。这表明主要是由于劳动技术暴露度低的职业需求上升导致了图3中的下行趋势。

(二)大语言模型人工智能技术暴露指数的跨职业差异

图5展示了基于DWA的大语言模型人工智能技术暴露指数最高和最低的20个职业。其中,新技术暴露指数最高的职业包括会计、编辑、销售及程序员等,这些主要是知识密集型的白领工作,工作内容涉及大量的文本处理和资料整合,其任务属性多属于规则化和重复性的任务,对创造性的需求较低。其中,销售类职业暴露度较高,这与人们的常规认识

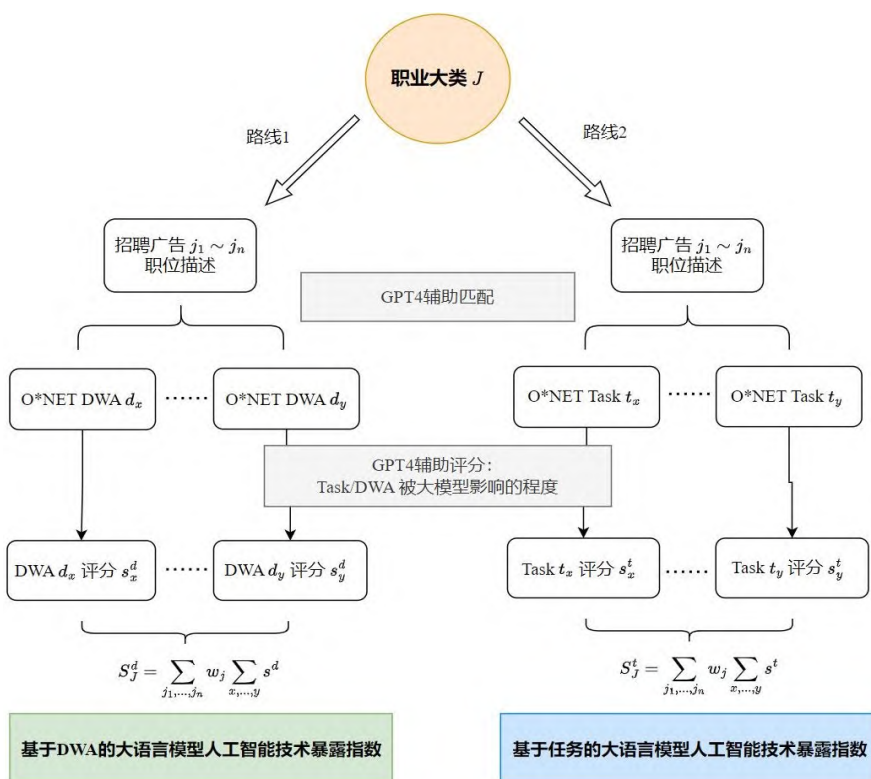


图2 大语言模型人工智能技术暴露指数构建策略

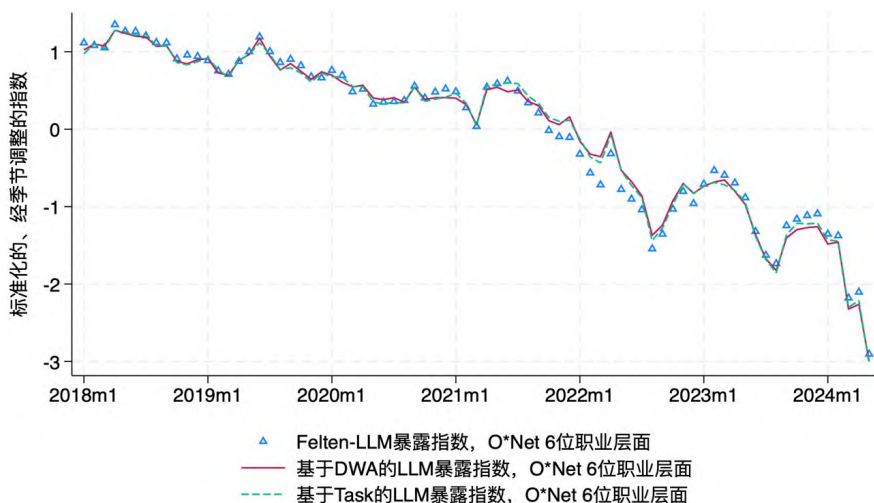


图3 新发布职位的大语言模型人工智能技术暴露指数月度趋势

可能有所不同。实际上,尽管销售通常需要很多“软技能”和直接与人沟通,但很多销售工作,如客户信息收集、电话沟通、信息咨询等,都属于常规性质任务,因此其暴露度较高。值得一提的是,相对于以往的数字技术,大语言模型属于通用人工智能,其应用场景更广,对自然语言的处理能力和对规则化任务的执行能力更强,因此受其影响的职业会更多,且包含大量的白领工作。《管理世界》网络发行版附录4描述了客服职业被人工智能技术所替代和改造的一个案例。

相对来说,暴露指数低的职业主要有餐饮服务人员、产业工人、家政服务人员等,这些工作多涉及体力劳动,且需要与人直接互动。但需要注意的是,这些职业尽管受大语言模型的影响小,但可能受到机器人和其他数字技术的冲击。全面评估职业的可替代性和未来的发展前景,需要结合各类不同新技术发展的趋势,做出综合评判。

表2使用普通最小二乘回归(OLS)方法分析了72个职业组别中大语言模型暴露指数与受教育程度、工作经验年限要求和薪资水平3个重要的职业特征之间的相关关系。结果表明,大语言模型暴露指数越高,职业招聘广告对应聘者的学历要求更高(第1列),但是对工作经验的要求则越高(第2列);给定教育和经验的要求,薪资水平更高的职业大语言模型暴露指数更高(第4列)。说明大语言模型暴露指数主要影响高教育程度、高工作经验

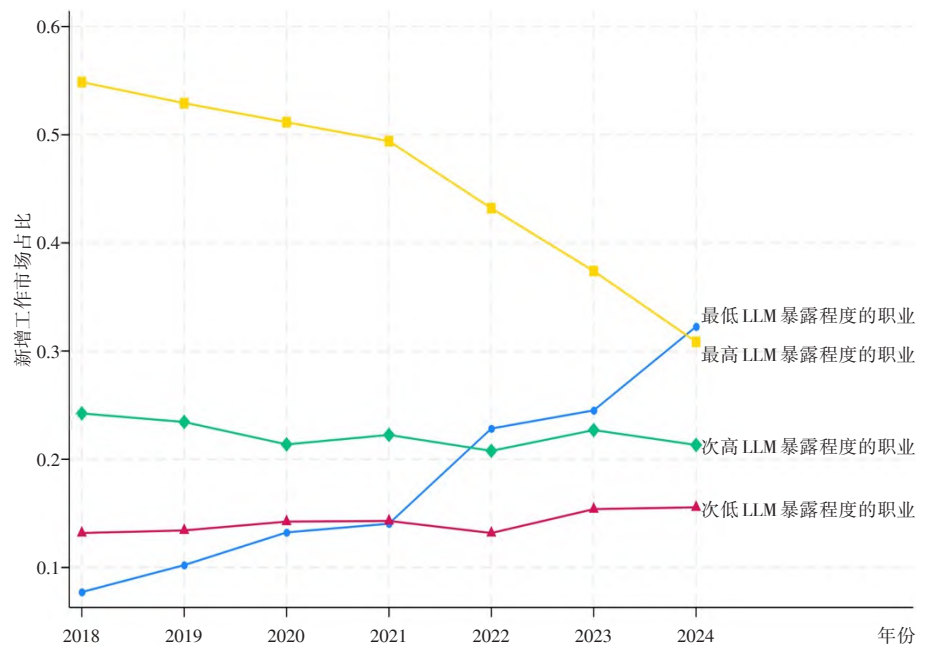


图4 大语言模型人工智能技术暴露指数四分位的市场占比变化趋势

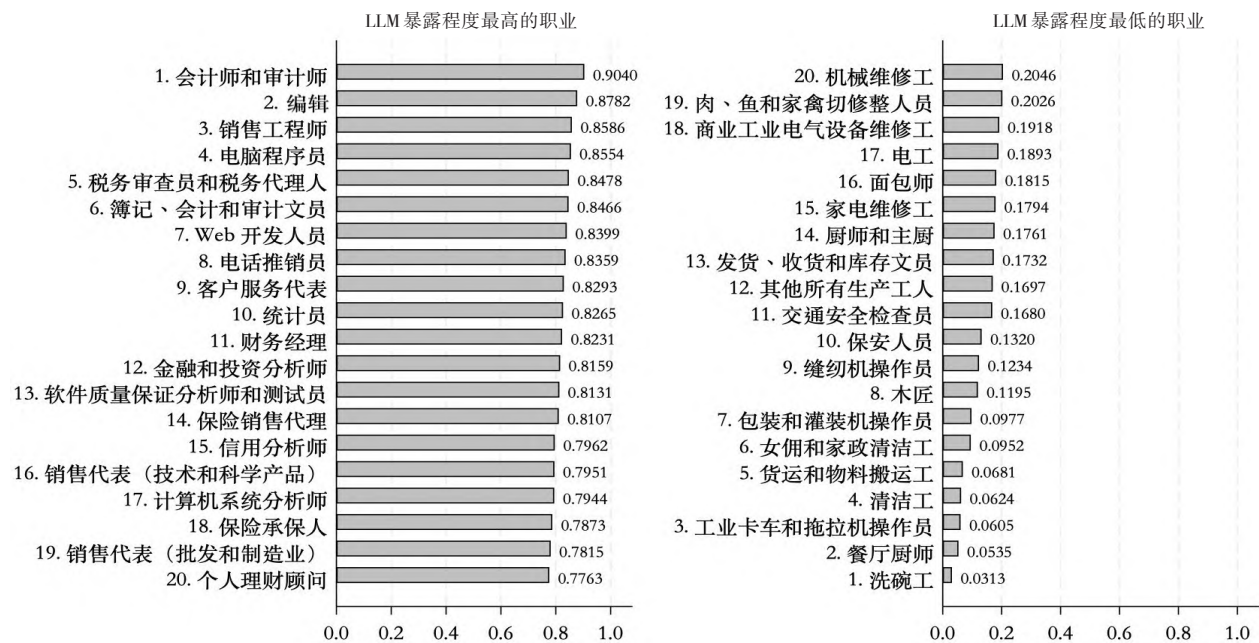


图5 大语言模型人工智能技术暴露指数最高和最低的职业

值和高薪的职业,这与埃隆杜等(2024)分析美国就业市场的发现基本一致。

(三)大语言模型人工智能技术暴露指数与新增职位的关系

图6展示了各大职业新发布职位在全部职位中占比的变化情况与基于DWA大语言模型人工智能技术暴露指数之间的关系^①。图6中的每个圆圈代表一个职业组别,其大小反映了该职业2018年的招聘职位总量。纵轴表示2022年6月~2024年5月(数据中的最近24个月)该职业招聘数量占样本中全部新增招聘数量的比例与2019~2020年之差,横轴表示标准化处理后的暴露指数。从图6左图(全部职业组别)展示的整体趋势来看,2019~2024年间,暴露指数越高的职业,市场新增职位越少;而暴露程度越低的行业,新增职位越多。相应的回归分析显示,某职业的大语言模型人工智能技术暴露指数每增加1个标准差(大致相当于从“工程师”职业组别到“信息与记录文员”职业组别之间的差距),其新增职位市场占比多降低(或少增加)1.9个百分点(在1%的水平上统计显著)。考虑到数据中的各职业大类招聘数量的占比平均仅为1.4%,1.9个百分点的差别意味着需求出现大幅下滑。

在图6左图中,“其他生产职业”的市场份额变化较大,变化的绝对值超过了5个百分点。为排除极端值的影响,我们在图6右图中去除了这个职业。可以看出,大语言模型人工智能技术暴露指数与新增用工需求的变化之间的负相关关系依然稳健地存在。

为了进一步识别大语言模型暴露指数和劳动力需求变化之间的因果关系,我们采用该职业所对应的美国职业的人工智能技术暴露指数作为中国劳动力市场上该职业大语言暴露指数的工具变量进行估计。为此,我们基于费尔滕等(2023)构建的指数,通过对中美两国的职业进行匹配,形成了Felten-LLM暴露指数。美国作为大语言模型发展成熟的先发国家,基于其职业的能力水平要求构建的人工智能暴露指数反映全球技术发展趋势,与中国相应职业受到的技术冲击具有很强的相关性(详见《管理世界》网络发行版附录2)。同时,美国的职业层面的大语言模型暴露指数不会对中国的劳动力市场产生直接的冲击,满足了工具变量的外生性假设。

为了在职业和城市层面对大语言模型暴露指数和各类劳动力需求变化之间的因果关系进行说明,我们采用了两种实证策略。首先,

表2 基于DWA的LLM暴露指数与招聘广告职业特征之间的关系

被解释变量	(1)	(2)	(3)	(4)
	基于DWA的LLM暴露指数(标准化)			
受教育年限	0.150*** (0.0180)			0.173*** (0.0247)
工作经验(年)		0.354** (0.175)		-0.586*** (0.197)
薪资水平(千元)			0.193*** (0.0621)	0.123** (0.0518)
观测值	72	72	72	72
R-平方	0.438	0.042	0.147	0.528

注:括号中为Bootstrap1000次后计算出的标准误差。***代表在1%的水平上统计显著,**代表在5%的水平上统计显著,*代表在10%的水平上统计显著。表中3个解释变量的计算基于2018年1月1日~2024年5月1日的招聘广告按72个“职业组别”计算的平均水平。

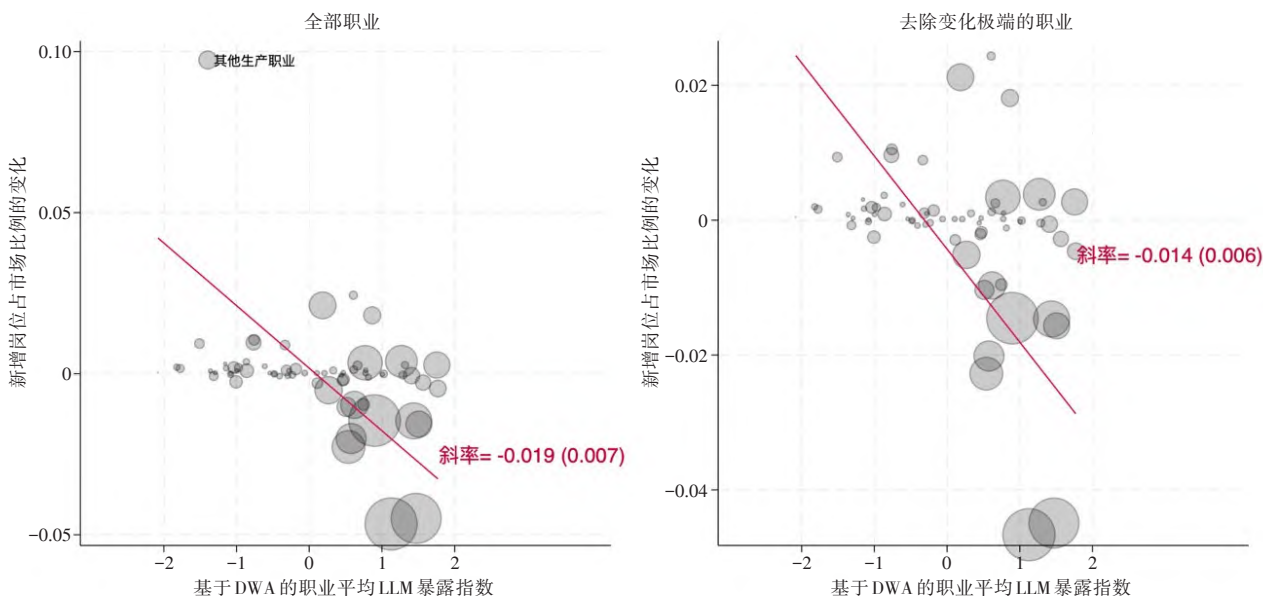


图6 各职业新发布职位市场占比变化与大语言模型人工智能技术暴露指数之间的关系
注:图中汇报的斜率为回归方程占比变化=α+β暴露指数+ε的估计系数β,括号中汇报了相应的稳健标准误差,回归方程以2018年的招聘广告数为权重加权。

我们在职业组别层面进行回归分析,展示职业组别的大语言模型暴露指数与2019~2024年间职业需求变动的关系,并使用其所对应的美国相关职业的 Felten-LLM 指数作为工具变量。回归方程如下:

$$\begin{aligned} \Delta Y_i &= \beta_0 + \beta_1 X_i + \epsilon_i, \\ X_i &= \gamma_0 + \gamma_1 Felten_i + u_i \end{aligned} \quad (1)$$

其中,被解释变量 ΔY_i 代表职业 i 的特征 Y ,最近两年(2022年6月~2024年5月)相比2019~2020年的变化,包括该职业招聘广告的市场占比、平均受教育水平要求、平均经验要求、平均薪资以及职业内薪资水平的变异系数。主要的解释变量 X_i 为基于 DWA 或基于任务构建的大语言模型暴露指数,汇总到职业 i 层面的平均水平^①。

简单线性模型可能存在内生性问题,比如一些观察不到的职业特征或者基于职业的宏观需求冲击可能与大语言模型暴露指数有关,同时影响了该职业的需求变化、薪资变化和技能要求。为了解决潜在内生性问题,我们使用职业层面对应的美国相关职业的 Felten-LLM 暴露指数($Felten_i$)作为工具变量。

其次,为了说明我国大语言模型在城市层面(可视作局部劳动力市场)的冲击,我们在城市层面重新构造了大语言模型暴露指数,并分析了其与城市职位招聘需求和特征的关系。城市层面的平均大语言模型暴露指数 X_c 为:

$$X_c = \sum_{i=1}^I v_{ci} X_i \quad (2)$$

其中, X_i 为职业 i 的大语言模型暴露指数,权重 v_{ci} 为职业 i 的招聘广告占城市 c 总招聘广告的比例。基于城市层面的分析,使用 Bartik 的方法构建城市—职业层面的大语言模型暴露指数工具变量,即采用美国的人工智能技术发展作为全国层面的一个共同冲击,再基于每个中国城市不同的职业构成,将该冲击分解到各个城市并加权平均一个城市层面的大语言模型冲击程度变量。博鲁西亚克等(2022)论证了上述构造方法的科学性,即仅冲击的外生性可保证工具变量的有效性。具体构建方法如公式(3)所示:

$$Predict_c = \sum_{k=1}^K w_{ck} Felten_k \quad (3)$$

其中, $Predict_c$ 代表城市 c 的“预测大语言模型暴露程度”,构建方法为,以2015年全国人口1%抽样普查中城市 c 中职业 k 就业占该城市总就业人口的比例(w_{ck})作为权重,对城市 c 中职业 k 的 $Felten$ 大语言模型暴露指数求均值。该工具变量融合了全国层面基于职业的外生技术冲击和 AI 技术大规模发展之前的城市不同职业就业比例,与当下的城市的大语言模型暴露情况高度相关,而与该城市在2019~2024年间的职业需求变动无直接相关关系,因此具有更好的外生性。城市层面的工具变量回归方程组如下:

$$\begin{aligned} \Delta Y_c &= \delta_0 + \delta_1 X_c + \epsilon_c \\ X_c &= \psi_0 + \psi_1 Predict_c + v_c \end{aligned} \quad (4)$$

表3汇报了基于 DWA 的大语言模型暴露指数与2019~2024年间职业在多个维度的变化之间的关系。为方便解读,在回归中我们对“基于 DWA 的大语言模型暴露指数”做了标准化处理。表3上端 A 部分报告了 OLS 回归结果,下端 B 部分报告了工具变量回归结果。工具变量第一阶段的 F 统计量高达 104.5,显示出高度的相关性。

工具变化回归结果显示,大语言模型暴露指数与该职业在过去的6年间劳动力总需求负相关。暴露指数每增加一个标准差^②,招聘广告的市场占比份额平均将下降 2.3 个百分点。尽管我们没有发现暴露指数对职业平均薪资水平的显著影响,但高暴露度职业的薪资

表3 基于 DWA 的 LLM 暴露指数与招聘广告特征要求之间的关系
(职业组别层面回归)

	(1)	(2)	(3)	(4)	(5)	(6)
	第一阶段	Δ 该职业招聘广告占市场总额%	Δ 受教育年限	Δ 工作经验	Δ 薪资水平	Δ 薪资水平的变异系数
A. OLS 回归						
基于 DWA 的 LLM 暴露指数		-0.0197*** (0.00695)	0.596*** (0.146)	0.0886** (0.0359)	284.5 (189.2)	0.0570** (0.0260)
B. 工具变量回归						
Felten 的 LLM 暴露指数	0.845*** (0.0827)					
基于 DWA 的 LLM 暴露指数		-0.0234*** (0.00467)	0.324 (0.198)	0.0400 (0.0393)	27.74 (227.0)	0.101*** (0.0293)
第一阶段 F 统计量	104.5					
观测值	70	70	70	70	70	70

注:OLS 回归在括号中汇报 robust 标准误;工具变量回归的标准误经过小样本调整。所有回归分析均按各职业在2018年的市场份额数进行了加权处理。***代表在1%的水平上统计显著,**代表在5%的水平上统计显著,*代表在10%的水平上统计显著。表中(1)列为第一阶段回归,采用的工具变量为美国相关职业的 Felten 大语言模型暴露指数;表中(2)~(6)列采用不同的被解释变量,均为该“职业组别”2022年6月~2024年5月(即数据中最近的24个月)相较于2019~2020年的变化量。

水平变异系数近年来显著提高,显示 AI 新技术有可能加剧了职业内部的薪资不平等,同一职业的薪资差距存在拉大趋势。此外,尽管在工具变量回归中统计上不显著,我们发现高暴露度的职业对教育水平和工作经验的要求更高。

表4中报告了城市层面的回归结果,与职业层面方向一致。一个城市的暴露指数越高,其在过去的6年经历了更为严重的劳动力市场需求萎缩,城市暴露指数提高一个标准差^⑧,其劳动力市场需求份额下降1.2个百分点。高职业暴露度的城市正在经历更加剧的薪酬不均,其薪资水平的变异系数更高。此外,高暴露度的城市对新招聘岗位的受教育程度和工作经验有了显著更高的要求。

(四)职业内部的适应性调整

到目前为止,我们的分析都假设同一职业的大语言模型人工智能技术暴露程度不随时间变化。如果职业的工作内容在发生变动,且变动与该职业的岗位需求以及技术暴露度都相关,就可能引致遗漏变量偏误问题。在本小节中,我们放松该假设,考察隶属同一职业的招聘职位的工作内容所呈现出来的暴露程度随时间的变化情况。我们使用了包含77000个招聘职位的测度样本,直接分析每个招聘职位的工作描述,并在职位层面给出大语言模型人工智能技术暴露程度的打分,然后再按照其所标注的“职业组别”逐月取平均值。针对数据中出现频数最多的16个职业组别,图7描绘了基于DWA的大语言模型人工智能技术暴露指数的变化趋势。

如果一个职业的具体工作中更多地采用了新技术,并由此改变了职业的内涵,则隶属该职业的招聘职位所体现出来的技术暴露指数应逐渐走高;反之,暴露指数会出现走低的态势。在图7中,16个职业大类按其2018~2024年平均暴露程度由高到低排序。总的来说,我们没有看到职业内部系统性的调整;大部分职业组别的大语言模型人工智能技术暴露指数保持稳定,说明这些职业在对工作任务的要求方面尚未出现明显的、针对人工智能新技术的调整。少数职业组别,如信息和记录文员、艺术和设计工作者和运营专业经理等,其大语言模型人工智能技术暴露程度近两年呈现出随时间下降趋势,可能出现了规避新技术的态势。

五、跨国比较和政策含义

从计算机、互联网、工业机器人,再到人工智能,数字技术的进步深刻重塑了生产组织形式,引发了劳动力市场的变迁。与此前的数字技术明显不同,人工智能技术,特别是生成式/大语言模型人工智能技术的崛起,对白领和高薪职业产生了相对来说更为显著的影响。本文基于在线招聘平台的数据,构建了适用于中国职业的大语言模型人工智能技术暴露指数,对“人工智能替代人工”的现状和变化趋势进行了初步描述,发现近年来中国的劳动力市场已经出现大语言模型人工智能技术开始替代人工的趋势,而劳动力市场尚未对人工智能新技术做出明显的适应性调整。

阿西莫格鲁等(2022)的研究指出,自2010年起,美国人工智能相关职位的需求出现持续上升趋势,在2015~2016年间增速尤为明显,且人工智能暴露度高的职业需求增长更为迅速。在企业层面,人工智能暴露度和人工智能职位呈正相关关系,人工智能暴露度增加一个标准差——大约是金融与采矿和石油行业企业间的差异,对应人工智能职位空缺占比增加15个百分点。同

表4 基于DWA的LLM暴露指数与招聘广告职业特征要求之间的关系(城市层面回归)

	(1)	(2)	(3)	(4)	(5)	(6)
	第一阶段	Δ该职业招聘广告占市场总额%	Δ受教育年限	Δ工作经验	Δ薪资水平	Δ薪资水平的变异系数
A. OLS 回归						
基于DWA的LLM暴露指数		-0.00335*** (0.00111)	0.604** (0.246)	0.0691** (0.0347)	-67.33 (175.7)	0.0533* (0.0278)
控制变量		Y	Y	Y	Y	Y
B. 工具变量回归						
Felten的LLM暴露指数	1.384*** (0.388)					
基于DWA的LLM暴露指数		-0.0119*** (0.00382)	3.000*** (0.948)	0.389*** (0.127)	-741.1* (411.4)	0.212** (0.0989)
控制变量		Y	Y	Y	Y	Y
第一阶段F统计量	12.74					
观测值	95	95	95	95	95	95

注: OLS 回归在括号中汇报 robust 标准误; 工具变量回归的标准误经过小样本调整。所有回归分析均按各职业在2018年的市场份额数进行了加权处理。所有回归中都控制了以下城市层面控制变量(2018年取值): 人均GDP, 第一产业占比, 第二产业占比, 该城市招聘广告占市场总额比例, 该城市招聘平均受教育年限要求, 该城市招聘平均工作经验要求, 该城市招聘平均薪资水平。其中人均GDP和产业占比来自城市统计年鉴, 其他变量根据智联招聘样本计算。***代表在1%的水平上统计显著, **代表在5%的水平上统计显著, *代表在10%的水平上统计显著。表中的样本为2018年招聘广告数占全国市场比例超过0.1%的城市, 共95个。表中(1)列为第一阶段回归, 采用的工具变量为美国相关职业的 Felten 大语言模型暴露指数; 表中(2)~(6)列采用不同的被解释变量, 均为该城市2022年6月~2024年5月(即数据中最近的24个月)相较于2019~2020年的变化量。

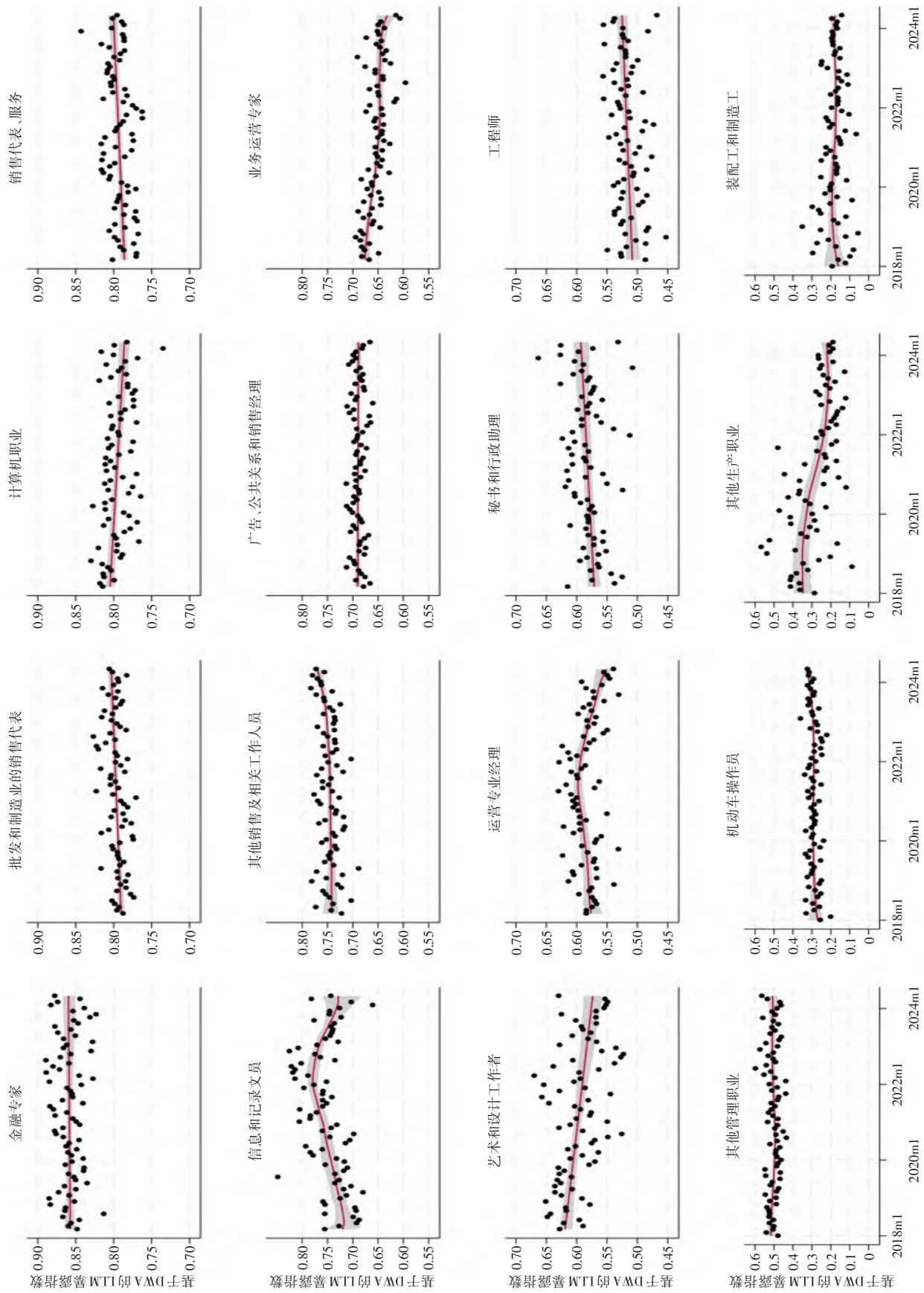


图7 各职业大语言模型人工智能技术暴露指数变化时间趋势

时,在行业层面或职业级别的整体劳动力市场上并未检测到人工智能暴露程度与就业或工资之间的关联性。作者认为,尽管人工智能技术的采用在机构中明显增多,但这种新技术的影响相比美国劳动力市场的规模仍显微不足道。他们因此得出结论:除了人工智能相关工作以外,新技术尚未对就业模式产生明显影响;从整体劳动力市场来看,人工智能对生产力的促进和就业的互补效果远大于就业的替代效果。

这与中国近期的劳动力市场表现形成了鲜明的对比。本文研究显示,在中国的劳动力市场上,具有高技术暴露度的职业,其需求相对下降,且对应聘者的教育程度和工作经验提出了更高的要求。另外,我们还发现了大语言模型人工智能技术会拉大职业内部薪资差距的证据。这些发现尽管可能受到其他经济基本面因素的影响,但值得持续追踪。新技术发展初期,就业市场上的技术适应性应当引起政府相关部门的重视,以避免产生国家间新的技术鸿沟。政府应当通过建立数据、算力等基础设施,构建技术转让和应用的良好市场环境,鼓励人工智能技术的研发和采纳。只有新技术与生产过程充分结合,才能扩大市场需求、创造就业。另外,针对初步出现的“人工智能替代人工”的趋势,应当对就业冲击进行预判,做好三方面的准备工作。一是在教育和人才队伍建设方面发力,努力提升劳动者技能,让更多的劳动者适应技术变革,解决好人力资源供需不匹配的结构性矛盾;二是在技术冲击引发的岗位转型调整的过程中,应当积极采取措施,通过创业支持、新业态帮扶等手段,帮助高技术暴露度的就业者实现就业转变;三是做好社会民生保障,完善就业援助制度,加强就业困难群体的兜底帮扶,加强失业人群保障工作,扩大社会保障体系的覆盖面^⑤。

(作者单位:张丹丹、于航、李力行,北京大学中国经济研究中心、北京大学国家发展研究院;胡佳胤,北京大学中国经济研究中心、北京大学国家发展研究院、北京大学数字金融研究中心;莫怡青、李泓宇,北京大学国家发展研究院)

注释

①职业信息网络(O*NET, Occupational Informal Network)是美国劳工部支持的一个全面的在线职业信息数据库,它为上千种职业提供了详细的标准化分类。SOC(The Standard Occupational Classification Code)即标准职业分类代码,是在包括美国在内的多个国家使用的、根据工作内容对不同职业进行分类和归类标准。

②Burning Glass Technologies(现已更名为Lightcast)是一家劳动市场分析公司,专门提供基于大数据的劳动力市场信息。它通过收集和分析数百万在线招聘广告,为企业、政府机构、教育机构和工作寻求者提供了深入的劳动力市场见解。其数据源主要是各种在线招聘网站、公司网站和社交媒体平台,覆盖了广泛的行业和职业。这些数据不仅提供了关于工作机会、所需技能和资质的信息,还包括了薪酬、工作地点、经验要求等关键细节。

③库恩和沈(2013)将智联招聘的职位样本与2005年1%人口抽样调查中的城市人口样本进行了对比,证实了智联招聘用户更年轻、学历更高且工资更高,在该平台上发布招聘广告的公司也更多地集中在互联网等快速扩张或员工流失率较高的行业。考虑到我们关注的是数字化技术对白领工作的影响,招聘平台数据在这些职位类别上的高暴露度可以很好地满足我们的研究需要。

④具体抽样和赋权重方法详见《管理世界》网络发行版附录1。相比于采用网络爬取数据的研究而言,本文采用的是招聘平台的全样本抽样数据,这保证本文样本的代表性,避免了爬取数据过程中因为各种原因导致的样本疏漏。

⑤我们根据招聘平台内部使用的“职业大类”“职业中类”和“职业小类”等标签,把每条招聘广告对应到O*NET体系中的6位数职业(Detailed Occupation),此处和下文中的“职业”无特别说明均指O*NET的6位数职业。在O*NET体系中,相近的职业进一步逐层向上归集为“广义职业”(Broad Occupation)、“职业组别”(Minor Group)和“职业大组”(Major Group)。本文中使用较多的“职业组别”(Minor Group)的分类层次大致对应《中国职业分类大典》分类体系中的“中类”;“职业”(Detailed Occupation)大致对应《中国职业分类大典》分类体系中的“职业”。

⑥尽管我们的方法和埃隆杜等(2024)存在一定的重合度,但他们的研究中并没有提供具体的DWA和Task层面的暴露指数数值,因此需要我们自行测算。

⑦类似的,图2的右半部分(也即路线2)显示了我们选取O*NET中最为细致的19265个任务(Task)作为定义工作内容的层级构建暴露指数的思路。在这里,每个招聘岗位的工作内容文本信息可以与最多20个O*NET任务进行关联,从而建立起每个招聘广告职位与O*NET任务的一对多匹配。接着,我们使用GPT-4评估这些任务受大语言模型人工智能技术的影响程度,并进而在职业或职业组别层面对这些评分进行加权平均,从而定义“基于任务的大语言模型人工智能技术暴露指数”。

⑧埃隆杜等(2024)研究的另一个贡献是他们分别采用人工和GPT-4两种不同的评分方式,并进行了系统比较。最终证实了使用GPT-4进行人工智能暴露度评估的准确性与稳定性。

⑨正如引言中所指出的,这种方法的局限在于两点。一是中美两国职业内涵不同,即同一职业类型包含的具体工作任务存在较大差异,研究中国劳动力市场与新技术的关系需要根据中国各职业的具体工作任务估算相应的技术影响指数;二是此前的研究尚未考虑不同职业的工作任务在与新技术结合后可能产生的变化,即职业名称不变但是内涵发生改变。

⑩其中,Felten-LLM暴露指数使用美国的职业分类来匹配中国的职业,其趋势与基于中国职业构建的大语言模型人工智能技术暴露指数相似。在后文中,我们将使用该指数作为基于中国职业构建的暴露指数的工具变量,以帮助识别其对中国劳动力市场职业需求的影响。

⑪我们也进行了城市—行业层面的回归分析,以控制地区、行业类型等的固定效应,其结果与图6中展示的结论一致。在此不再

赘述,如有需要可向作者索求相关分析结果。

⑫由于基于DWA或基于Task大语言模型暴露指数高度相关,正文中仅报告以“基于DWA的LLM暴露指数”为主要解释变量的回归结果,基于工作任务的大语言模型暴露指数的回归结果表包含在《管理世界》网络发行版附录3中,其结果与基于DWA大语言模型暴露指数的分析结果基本一致。

⑬一个标准差的差异大致相当于其LLM暴露度从“工程师”职业组别(暴露度处于中位)的水平提高到“信息和记录文员”职业组别(暴露度高于职业中位水平一个标准差)的水平。

⑭一个标准差的差异大致相当于其LLM暴露度从宁波(以制造业为主的城市,暴露度处于全国中位)的水平提高到北京(以高新技术和服务业为主的城市,暴露度高于全国中位水平一个标准差)的水平。

⑮中外文人名(机构名)对照:弗雷(Frey);奥斯本(Osborne);韦布(Webb);费尔滕(Felten);埃隆杜(Eloundou);奥特(Autor);斯塔内娃(Staneva);埃利奥特(Elliott);布林约尔夫森(Brynjolfsson);米切尔(Mitchell);阿西莫格鲁(Acemoglu);古斯(Goos);多恩(Dorn);雷斯特雷波(Restrepo);格雷茨(Graetz);迈克尔斯(Michaels);贝森(Bessen);阿内利(Anelli);琼特拉(Giuntella);阿恩茨(Arntz);赫什拜因(Hershbein);卡恩(Kahn);库恩(Kuhn);沈(Shen);方(Fang);博鲁西亚克(Borusyak)。

参考文献

- (1)陈琳、高悦蓬、余林徽:《人工智能如何改变企业对劳动力的需求——来自招聘平台大数据的分析》,《管理世界》,2024年第6期。
- (2)陈媛媛、张竞、周亚虹:《工业机器人与劳动力的空间配置》,《经济研究》,2022年第1期。
- (3)胡连漪、盖庆恩、朱喜、郭士祺:《中国职业技能的结构转型:基于任务内容的视角》,《经济研究》,2024年第1期。
- (4)李磊、王小霞、包群:《机器人的就业效应:机制与中国经验》,《管理世界》,2021年第9期。
- (5)王林辉、胡晟明、董直庆:《人工智能技术、任务属性与职业可替代风险:来自微观层面的经验证据》,《管理世界》,2022年第7期。
- (6)王林辉、钱圆圆、宋冬林、董直庆:《机器人应用的岗位转换效应及就业敏感性群体特征——来自微观个体层面的经验证据》,《经济研究》,2023年第7期。
- (7)王永钦、董雯:《机器人的兴起如何影响中国劳动力市场?——来自制造业上市公司的证据》,《经济研究》,2020年第10期。
- (8)王永钦、董雯:《中国劳动力市场结构变迁——基于任务偏向型技术进步的视角》,《中国社会科学》,2023年第11期。
- (9)闫雪凌、朱博楷、马超:《工业机器人使用与制造业就业:来自中国的证据》,《统计研究》,2020年第1期。
- (10)姚加权、张银澎、郭李鹏、冯绪:《人工智能如何提升企业生产效率——基于劳动力技能结构调整的视角》,《管理世界》,2024年第2期。
- (11)尹志锋、曹爱家、郭家宝、郭冬梅:《基于专利数据的人工智能就业效应研究——来自中关村企业的微观证据》,《中国工业经济》,2023年第5期。
- (12)周广肃、李力行、孟岭生:《智能化对中国劳动力市场的影响——基于就业广度和强度的分析》,《金融研究》,2021年第6期。
- (13)Acemoglu, D. and Autor, D., 2011, “Skills, Tasks and Technologies: Implications for Employment and Earnings”, in *Handbook of Labor Economics*, 4b(12), pp.1043~1171.
- (14)Acemoglu, D., Autor, D., Hazell, J. and Restrepo, P., 2022, “AI and Jobs: Evidence from Online Vacancies”, *Journal of Labor Economics*, 40(S1), pp.293~340.
- (15)Acemoglu, D. and Restrepo, P., 2019, “Automation and New Tasks: How Technology Displaces and Reinstates Labor”, *Journal of Economic Perspectives*, 33(2), pp.3~30.
- (16)Anelli, M., Colantone, I. and Stanig, P., 2021, “Individual Vulnerability to Industrial Robot Adoption Increases Support for the Radical Right”, *Proceedings of the National Academy of Sciences*, 118(47), e2111611118.
- (17)Arntz, M., Gregory, T. and Zierahni, U., 2016, “The Risk of Automation for Jobs in OECD countries: A Comparative Analysis”, *OECD Social Employment and Migration Working Papers*, No.189.
- (18)Autor, D., Chin, C., Salomons, A. M., et al., 2022, “New Frontiers: The Origins and Content of New Work, 1940–2018”, *NBER Working Paper*, No.30389.
- (19)Autor, D. and Dorn, D., 2013, “The Growth of Low-skill Service Jobs and the Polarization of the US Labor Market”, *American Economic Review*, 103(5), pp.1553~1597.
- (20)Autor, D., Levy, F. and Murnane, R., 2003, “The Skill Content of Recent Technological Change: An Empirical Exploration”, *Quarterly Journal of Economics*, 118(4), pp.1279~1333.
- (21)Bessen, J., 2019, “Automation and Jobs: When Technology Boosts Employment”, *Economic Policy*, 34(100), pp.589~626.
- (22)Borusyak, K., Jaravel, X. and Spiess, J., 2022, “Revisiting Event Study Designs: Robust and Efficient Estimation”, *CEPR Discussion Paper No.17247*, CEPR Press, Paris & London, Available at <https://cepr.org/publications/dp17247>.
- (23)Brynjolfsson, E. and Mitchell, T., 2017, “What Can Machine Learning Do? Workforce Implications”, *Science*, 358(6370), pp.1530~1534.
- (24)Eloundou, T., Manning, S., Mishkin, P. and Rock, D., 2024, “GPTs are GPTs: Labor Market Impact Potential of LLMs”, *Science*, 384(6702), pp.1306~1308.
- (25)Fang, H., Ge, C., Huang, H. and Li, H., 2020, “Pandemics, Global Supply Chains, and Local Labor Demand: Evidence from 100 Million Posted Jobs in China”, *NBER Working Paper*, No.28072.
- (26)Felten, E., Raj, M. and Seamans, R. C., 2018, “A Method to Link Advances in Artificial Intelligence to Occupational Abilities”, *AEA Papers and Proceedings*, 108, pp.54~57.
- (27)Felten, E., Raj, M. and Seamans, R. C., 2019, “The Effect of Artificial Intelligence on Human Labor: An Ability-Based Approach”, *Academy of Management Proceedings*, 2019(1), p.15784.
- (28)Felten, E., Raj, M. and Seamans, R. C., 2021, “Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and Its Potential Uses”, *Strategic Management Journal*, 42(12), pp.2195~2217.

- (29) Felten, E., Raj, M. and Seamans, R. C., 2023, "How will Language Modelers like ChatGPT Affect Occupations and Industries?", Available at SSRN: <https://ssrn.com/abstract=4375268>.
- (30) Frey, C. and Osborne, M., 2017, "The Future of Employment: How Susceptible Are Jobs to Computerisation?", *Technological Forecasting and Social Change*, 114, pp.254~280.
- (31) Giuntella, O., Lu, Y. and Wang, T., 2022, "How Do Workers and Households Adjust to Robots? Evidence from China", NBER Working Paper, No.30307.
- (32) Goos, M., Manning, A. and Salomons, A., 2014, "Explaining Job Polarization: Routine-biased Technological Change and Offshoring", *American Economic Review*, 104(8), pp.2509~2526.
- (33) Graetz, G. and Michaels, G., 2018, "Robots at Work", *Review of Economics and Statistics*, 100(5), 753~768.
- (34) Hershbein, B. and Kahn, L., 2016, "Do Recessions Accelerate Routine Biased Technological Change", NBER Working Paper, No.22762.
- (35) Kuhn, P. and Shen, K., 2013, "Gender Discrimination in Job Ads: Evidence from China", *The Quarterly Journal of Economics*, 128(1), pp.287~336.
- (36) Staneva, M. and Elliott, S., 2023, "Measuring the Impact of Artificial Intelligence and Robotics on the Workplace", in Shajek, A. and Hartmann, E. A. (Eds.), *New Digital Work*, Chapter 2, pp.16~30.
- (37) Webb, M., 2020, "The Impact of Artificial Intelligence on the Labor Market", SSRN Working Paper 3482150, Available at: <https://ssrn.com/abstract=3482150>.

**The Measurement of AI Exposure and Its Impact on Labor Demand in China:
Evidence from Large Language Models**

Zhang Dandan^{ab}, Yu Hang^{ab}, Li Lixing^{ab}, Hu Jiayin^{abc}, Mo Yiqing^b and Li Hongbo^b

(a. China Center for Economic Research, Peking University; b. National School of Development, Peking University;
c. Institute of Digital Finance, Peking University)

Abstract: The Large Language Models (LLM) and generative Artificial Intelligence (AI) are developing rapidly. Using 1.25 million on-line job posts published between January 2018 and May 2024, this paper constructs the LLM-AI exposure index for each occupation based on the detailed descriptions of job tasks, and estimates the impact on labor demand. The study finds that the average AI exposure for new jobs in the Chinese labor market shows a decreasing trend during the sample period; occupations with higher exposure are mainly white-collar occupations with relatively higher education requirements and higher salaries, such as accountant, editor, salesperson and programmer. Regression results based on instrumental variables at the occupation level show a negative relationship between exposure and labor demand, with high exposure occupations experiencing lower salary increases, larger within-occupation salary gap, and higher education and work experience requirements. This paper highlights the need to strengthen the adaptability of China's labor market to new technologies, and suggests that the threshold of AI application should be lowered to encourage enterprises and workers to adopt new technologies.

Keywords: artificial intelligence; large language models; labor demand; employment; ChatGPT

=====

(上接第 58 页)

**Large Language Models Driving the Transformation of Knowledge Production
and the Reconstruction of Decision-Making Paradigms**

Mi Jianing^a, Li Dayu^b and Dong Changqi^c

(a. School of Economics and Management, Beijing University of Posts and Telecommunications; b. School of Finance and Public Administration, Harbin University of Commerce; c. School of Management, Harbin Institute of Technology)

Abstract: Large Language Models (LLMs) are initiating a revolutionary transformation in knowledge production methodologies, profoundly influencing how humans perceive and transform the world. This paper aims to explore how knowledge production transformations driven by LLMs reconstruct decision-making paradigms, establishing an integrative analytical framework that reveals their intrinsic connection and providing theoretical guidance for navigating the intelligent revolution and advancing human-machine collaborative management decisions. The research systematically examines the co-evolutionary trajectory of knowledge production and decision-making paradigms, and analyzes in depth the internal mechanisms through which LLMs revolutionize knowledge production methodologies and the consequent pathways of decision paradigm transformation. Findings reveal that LLMs, through technical processes including knowledge representation, mapping, generation, and modeling, drive a paradigm shift from discrete symbols to continuous vectors and from local features to global representations, fostering large-scale knowledge integration and interactive knowledge retrieval capabilities. In terms of reshaping decision paradigms, this knowledge production revolution leads organizations toward openness, complex adaptation, and dynamic evolution through reconstructing knowledge supply mechanisms, enabling human-machine hybrid intelligence, and fostering satisficing evolutionary decision-making. Simultaneously, the research provides technical approaches to mitigating the "Hayek Problem"—the structural contradiction between distributed knowledge and centralized decision-making. The study also dialectically analyzes the technical risks and social impacts of LLM applications, ultimately proposing multi-level application strategies and practical recommendations that expand new possibility spaces for human-machine collaboration, collective wisdom emergence, and decision paradigm transformation.

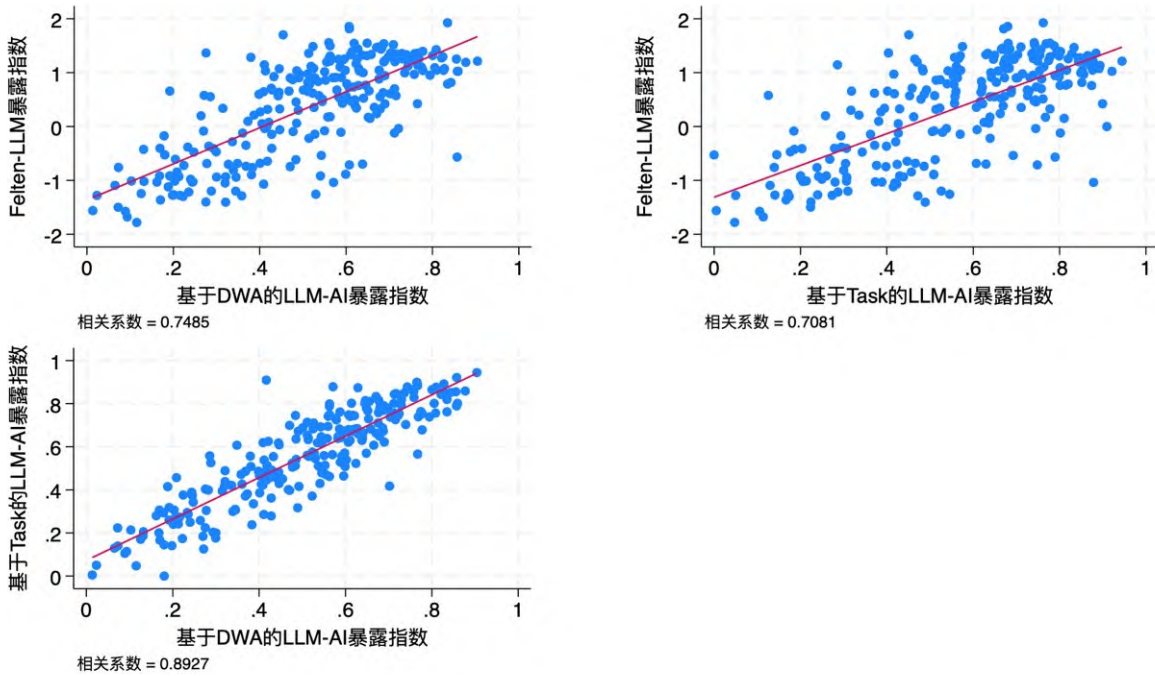
Keywords: large language models (LLMs); knowledge production methods; Hayek Problem; human-machine collaborative decision-making; decision-making paradigm

附录1 抽样方法

智联招聘为本研究提供了招聘广告抽样数据,时间范围涵盖2018年1月1日~2024年5月31日。抽样按年分层,每年随机抽取样本约20万条。特别地,2024年仅5个月的总抽样数也达到了约20万条;为平衡样本数量,我们为2024年的每条观测都赋予权重0.39,该权重来源于2018~2023年,每年5月31日前的招聘广告总数占全年的39%,加权后,样本权重总和约125万。

鉴于我们使用的按年分层的抽样方法,职位招聘总量在不同年份间的不具备可比性,但同一年内不同月份间的总量可比。

附录2 不同测度的LLM-AI暴露指数之间的相关性



附图1 两个原创LLM暴露指数、Felten-LLM暴露指数之间的相关性

附图1展示了247个O*Net6位数职业中的3种暴露指数之间的两两相关关系。三者显示出较高的相关性,特别是通过中国的招聘数据构建的两个原创LLM-AI暴露指数之间的相关性更强,佐证了基于DWA和Task构建中国的LLM-AI暴露指数这两个技术路线的合理性。此外,两个基于中国招聘数据的原创基于DWA和Task的LLM-AI暴露指数与美国数据构建的Felten-LLM暴露指数,在职业大类上匹配后也体现出较强的正相关关系,这个部分与正文中的工具变量第一阶段重合,即论证了使用Felten-LLM暴露指数作为工具变量与我国的两个LLM-AI暴露指数的强相关性和工具变量第一阶段回归的显著性。

尽管全样本招聘广告共涵盖346个O*Net6位数职业,我们用于构建LLM暴露指数的“测度样本”(供77000条观测)仅涵盖其中的247个职业,其他职业因市场占比太小而未被“测度样本”包含。构建了LLM暴露指数的247个职业占全样本招聘广告的比例为99.05%。

附录3 基于工作任务(Task)的LLM-AI暴露指数的回归分析

附表1 基于Task的LLM暴露指数与招聘广告职业特征要求之间的关系

被解释变量	(1)	(2)	(3)	(4)
	基于Task的LLM暴露指数(标准化)			
受教育年限	0.153*** (0.0191)			0.153*** (0.0242)
工作经验(年)		0.580*** (0.170)		-0.246 (0.184)
薪资水平(千元)			0.203*** (0.0621)	0.104** (0.0506)
观测值	72	72	72	72
R-平方	0.455	0.114	0.162	0.491

注:括号中为Bootstrap1000次后计算出的标准误。***代表在1%的水平上统计显著,**代表在5%的水平上统计显著,*代表在10%的水平上统计显著。表中3个解释变量的计算基于2018年1月1日~2024年5月1日招聘信息按72个“职业组别”计算的平均水平。

附表2 基于Task的LLM暴露指数与招聘广告职业特征之间的关系(职业层面回归)

	(1)	(2)	(3)	(4)	(5)	(6)
	第一阶段	Δ该职业招聘广告占市场总额%	Δ受教育年限	Δ工作经验	Δ薪资水平	Δ薪资水平的变异系数
A.OLS回归						
基于DWA的LLM暴露指数		-0.0198*** (0.00681)	0.433*** (0.130)	0.0720** (0.0319)	258.4 (181.7)	0.0627** (0.0256)
B.工具变量回归						
Felten的LLM暴露指数	0.912*** (0.0663)					
基于DWA的LLM暴露指数		-0.0217*** (0.00434)	0.301 (0.190)	0.0371 (0.0370)	25.72 (210.7)	0.0940*** (0.0266)
第一阶段F统计量	188.7					
观测值	70	70	70	70	70	70

注:OLS回归在括号中汇报robust标准误;工具变量回归的标准误经过小样本调整。所有回归分析均按各职业在2018年的市场份额数进行了加权处理。***代表在1%的水平上统计显著,**代表在5%的水平上统计显著,*代表在10%的水平上统计显著。表中(1)列为第一阶段回归,采用的工具变量为美国相关职业的Felten大语言模型暴露指数;表中(2)~(6)列采用不同的被解释变量,均为该“职位组别”2022年6月~2024年5月(即数据中最近的24个月)相较于2019~2020年的变化量。

附表3 基于Task的LLM暴露指数与招聘广告职业特征要求之间的关系(城市层面回归分析)

	(1)	(2)	(3)	(4)	(5)	(6)
	第一阶段	Δ该城市招聘广告占市场总额%	Δ受教育年限	Δ工作经验	Δ薪资水平	Δ薪资水平的变异系数
A.OLS回归						
基于Task的LLM暴露指数		-0.00446*** (0.00136)	0.768*** (0.235)	0.0943*** (0.0345)	-90.29 (187.9)	0.0601** (0.0268)
控制变量		Y	Y	Y	Y	Y
B.工具变量回归						
Felten的LLM暴露指数	1.388*** (0.389)					
基于Task的LLM暴露指数		-0.0118*** (0.00348)	2.990*** (0.897)	0.388*** (0.119)	-738.8* (406.4)	0.212** (0.0973)
控制变量	Y	Y	Y	Y	Y	Y
第一阶段F统计量	12.71					
观测值	95	95	95	95	95	95

注:OLS回归在括号中汇报robust标准误;工具变量回归的标准误经过小样本调整。所有回归分析均按各职业在2018年的市场份额数进行了加权处理。所有回归中都控制了以下城市层面控制变量(2018年取值):人均GDP,第一产业占比,第二产业占比,该城市招聘广告占市场总额比例,该城市招聘平均受教育年限要求,该城市招聘平均工作经验要求,该城市招聘平均薪资水平。其中人均GDP和产业占比来自城市统计年鉴,其他变量根据智联招聘样本计算。***代表在1%的水平上统计显著,**代表在5%的水平上统计显著,*代表在10%的水平上统计显著。表中的样本为2018年招聘广告数占全国市场比例超过0.1%的城市,共95个。表中(1)列为第一阶段回归,采用的工具变量为美国相关职业的Felten大语言模型暴露指数;表中(2)~(6)列采用不同的被解释变量,均为该城市2022年6月~2024年5月(即数据中最近的24个月)相较于2019~2020年的变化量。

附录4 案例——人工智能技术重塑客服职业

课题组对一家国内头部在线服务企业进行了调研,了解该公司使用人工智能技术改造客服岗位的情况。该公司的一大特点,是在2023年底以ChatGPT为代表的大语言模型人工智能爆火之前,就已经广泛采纳了人工智能技术,大致分为两个阶段。第一阶段是2017~2021年,主要聚焦语音识别、智能质检与推荐系统优化,语音转写准确率提升至94%,实现了自动工单分类和多语言翻译;第二阶段是2021~2022年,人工智能技术进一步拓展至多语种支持和自助服务优化,在线咨询自助率从不足50%提高至75%。

目前,人工智能技术在该公司已广泛替代了重复性任务(如语音转写、工单生成),自助服务覆盖率显著提升(电话场景达35%,在线场景达75%),并通过动态推荐系统和多模态语音识别技术显著提高了服务效率。与此同时,人工客服转向处理复杂场景,如紧急协调、情绪安抚及高净值客户服务,岗位技能从操作性任务转向以“沟通力、服务力、解决力”为核心的综合能力。

该公司在引入人工智能技术后,系统性重构了客服岗位的功能与组织结构。在人工智能技术应用之前,客服团队规模超过2万人;而在人工智能技术全面部署后,团队规模缩减至约1万人,人力成本显著下降。尽管岗位总数有所减少,该公司客服团队仍保持较高的稳定性与薪酬水平(平均工龄6年,月流失率2%,一线城市薪酬位于行业75分位以上)。公司通过人机协同优化岗位结构,推动服务业由“成本中心”向“价值中心”转型。