



科学学研究  
*Studies in Science of Science*  
ISSN 1003-2053, CN 11-1805/G3

## 《科学学研究》网络首发论文

题目: 大语言模型对科研人员的潜在职业影响探究  
作者: 包芊颖, 李亚玲, 高金莎  
DOI: 10.16192/j.cnki.1003-2053.20240603.002  
收稿日期: 2024-03-26  
网络首发日期: 2024-06-03  
引用格式: 包芊颖, 李亚玲, 高金莎. 大语言模型对科研人员的潜在职业影响探究  
[J/OL]. 科学学研究. <https://doi.org/10.16192/j.cnki.1003-2053.20240603.002>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 大语言模型对科研人员的潜在职业影响探究<sup>1</sup>

包芊颖<sup>1,2</sup>, 李亚玲<sup>1,2</sup>, 高金莎<sup>1,2</sup>

(1.之江实验室智能社会治理研究中心, 2.浙江省哲学社会科学试点实验室——之江实验室智能社会治理实验室, 浙江杭州 311121)

**摘要:** 以 ChatGPT 为代表的大语言模型已在部分科研领域得到应用, 并引发了大语言模型对于科研人员职业影响的关注。本文探索利用 GPT-4 评估大语言模型或由大语言模型驱动的系统的应用对科研人员潜在职业影响, 并从知识、技能、受教育水平、工资水平探究职业特征与职业影响之间的关系。结果表明: 不同研究方向的科研人员会受到大语言模型及其驱动的系统不同程度的影响; 从职业任务来看, 大语言模型对于标准化、规则化的任务影响较大, 对于非标准化、具体化的实操任务影响较小; 从知识来看, 大语言模型对于人文社科、理学类学科影响较大, 对于机械工程类学科影响较小; 从技能来看, 大语言模型对于逻辑推理类技能影响较大, 对于涉及到人际交往、机械操控类技能影响较小; 从受教育水平和工资水平来看, 大语言模型对于高受教育水平、高工资水平的职业影响程度更大。由此可见, 大语言模型可以成为科研人员的生产工具应用到科研工作中, 从而提高科研人员的劳动效率、进一步加快科研进程。

**关键词** 大语言模型; 科研人员; 职业影响  
**中图分类号:** G3

**文献标志码:** A

## The potential career impact of large language models on researchers

Bao Qianying<sup>1,2</sup>, Li Yaling<sup>1,2</sup>, Gao Jinsha<sup>1,2</sup>

(1. Research Center for Intelligent Society and Governance,  
2. Zhejiang Pilot Laboratory of Philosophy and Social Science—Laboratory of Intelligent Society and Governance, Zhejiang Lab, Hangzhou 311121, China)

**Abstract:** The rapid advancement of artificial intelligence, particularly large language models (LLMs) such as ChatGPT, has garnered significant attention for its transformative potential across various sectors. This study focuses on the potential occupational impacts of LLMs on researchers, evaluating how these models may alter the research landscape and influence the careers of those within it. Using GPT-4, the study assesses the extent to which LLMs impact different aspects of researchers' work, including task performance, skill requirements, education levels, and salary scales. The study utilizes data from the "Occupational Classification of China" (2022 edition) and the U.S. Occupational Information Network (O\*NET) to map specific research tasks and assess

收稿日期: 2024-03-26; 修回日期: 2024-05-21

基金项目: 科技部国家重点研发计划“智能社会演化机理及其运转体系推演模型”(2022YFC3303103);

浙江省科技厅软科学计划研究项目“数字化赋能科技人才评价改革研究(2023C25074)”; 浙江省科技厅软科学计划研究项目“生成式预训练大模型引发的隐私泄露风险及治理路径”(2024C35035)

作者简介: 包芊颖(1990-), 女, 硕士研究生; 李亚玲(1991-), 女, 博士研究生; 高金莎(1996-), 女, 硕士研究生, 通讯作者, E-mail: g1693807389@163.com

their susceptibility to LLM influence. Tasks were evaluated based on their potential for automation using GPT-4, with a scoring system that quantifies the degree to which LLMs can reduce human labor input. The evaluation involved three rounds of scoring to ensure robustness and reliability of the results.

The findings reveal several key impacts of LLMs on the research sector:

Regarding research fields, the results indicate that researchers in different fields are affected by LLMs to varying degrees. Mathematical researchers experience higher impacts due to LLMs' capabilities in processing large volumes of textual data, facilitating tasks such as literature reviews, data analysis, and report generation. In these fields, the ability of LLMs to quickly synthesize information and generate comprehensive reports can significantly reduce the time researchers spend on these tasks. Mechanical researchers in this field are less impacted as their tasks often involve practical and hands-on activities that LLMs are not yet equipped to handle effectively.

Regarding skills, tasks requiring high levels of logical reasoning are more susceptible to LLMs, which can handle data analysis and modeling efficiently. Interpersonal communication and mechanical operations are less impacted as they involve nuanced human interactions and physical manipulations that LLMs cannot replicate.

Regarding education and salary levels, researchers with higher education levels are more impacted as LLMs can handle complex cognitive tasks that are typical at higher education levels. Roles offering higher salaries are more likely to be influenced by LLMs, reflecting the advanced nature of tasks and the significant potential for productivity improvements through automation.

LLMs have the potential to transform the research landscape by automating standardized tasks, thereby enhancing efficiency and accelerating research progress. However, the extent of this impact varies across different fields, skills, and demographic factors. Policymakers and educational institutions must adapt to these changes by integrating AI literacy and skills development into their programs to prepare researchers for an AI-enhanced future. Future research should continue to monitor the evolving capabilities of LLMs and their broader implications for the research community. In conclusion, while LLMs offer substantial benefits in terms of efficiency and productivity, their impact on the research sector is complex and multifaceted. It is crucial for stakeholders to develop strategies that maximize the advantages of LLMs while mitigating potential disruptions to researchers' careers. This study provides a foundational understanding of these dynamics, paving the way for more detailed investigations into specific impacts and mitigation strategies.

Key words: Large Language Models; Researchers; Career Impact

习近平总书记强调，人工智能是引领这一轮科技革命和产业变革的战略性技术，具有溢出带动性很强的“头雁”效应。人工智能技术的快速发展，引发了学术界对于该项技术在劳动力和劳动力市场的应用的极大关注，涌现了大量关于自动化技术、人工智能技术等与劳动力市场变革主题的研究。从职业替代到人工智能暴露（AI Occupational Exposure），学者们初步推断出对于现有职业而言，规则性和重复性的职业易于受到冲击，甚至被替代，而工作任务突出情感性、复杂性和创新性的职业受影响较弱<sup>[1, 2]</sup>。

但随着以 ChatGPT 为代表的大语言模型（Large Language Models, LLMs）的迅猛进展，人工智能对劳动力市场的影响进一步深化。大语言模型开始对非常规型职业产生影响，一些简单认知类的职业也开始受到冲击，人们开始担忧此类工具对劳动力市场带来的负面影响，如纽约邮报专题发文称“ChatGPT 可能会让

这些工作过时：“狼来了””。于是学者们延续过往人工智能技术对于职业的影响研究，开始探索大语言模型对职业的影响。研究表明大语言模型的功能已然渗透到各种工作任务中，甚至影响劳动力市场的结构<sup>[3, 4]</sup>。

科学研究领域是颇受关注的职业领域之一，大语言模型在文本创作、知识生产、内容交互等能力的输出大大降低了科学研究的技术门槛，*Nature* 杂志也发文表示 ChatGPT 和其他大型语言模型已经应用于论文撰写、文献梳理、代码编写、统计分析、实验设计等任务<sup>[5]</sup>。国内外学者就大语言模型对科研人员的影响展开了理论探讨，但尚未有聚焦科研人员的定量分析研究。

因此本研究聚焦科研人员这一职业群体，研讨大语言模型对科研人员带来的影响，致力于回答以下问题：大语言模型究竟会对哪些科研职业和科研任务产生影响？哪些职业特征更容易受到大语言模型的冲击？本研究采用 Tyna 等<sup>[4]</sup>学者使用的创新方法，以自然科学和地球科学研究人员为例，利用 GPT-4 评估 LLM 或由 LLM 驱动的系统的应用对该类科研人员具体工作任务的影响。同时，通过与过往研究对比，证明本研究采用的评估方法具有稳定性。最后，通过分析职业要求、工资水平等职业特征与职业影响之间的关系，探究 LLM 在提高科研人员的劳动效率、辅助支持科研进程方面的潜力，从而为 LLM 在科学领域的深入应用提供参考。

## 1 文献综述

### 1.1 大语言模型进展

大语言模型是一种利用海量文本数据进行训练的预训练语料模型，其一般指的是包含数千亿或更多参数的模型，如 GPT-3、PaLM、Galactica 和 LLaMA 等。这类基于大规模参数的大语言模型主要表现出三种独特的涌现能力<sup>[6]</sup>：一是情境学习能力（In-context Learning），在给定自然语言任务描述或若干任务示例的条件下，在测试阶段便可生成解决方案，而不需要额外训练和梯度更新；二是指令遵循能力（Instruction Following），使用自然语言描述格式化的多任务混合数据集微调，LLMs 便可在不使用显式示例的情况下遵循新任务的指令；三是逐步推理能力（Step-by-Step Reasoning），使用思维链提示策略，LLMs 可以解决复杂的推理问题，如数学应用题。

随着 LLMs 逐步表现出与人类交流的卓越能力,越来越多互联网公司也着手开发属于自身的 LLMs,如 OpenAI 的 GPT-1 至 GPT-4,微软的 Copilot,谷歌的 Bard,Meta 的 Llama,百度的文心一言、华为的盘古等等。<sup>[6]</sup>这些大语言模型为人类提供了语言生成(代码合成、条件文本生成)、知识运用(知识回答、知识补充)、复杂推理(知识推理、符号推理、数学推理)等重要功能。

## 1.2 人工智能对职业的影响

大量文献探讨了人工智能(以下简称 AI)或自动化技术对职业环境的影响。最初学者们关注 AI 或自动化技术对职业总体情况的影响程度<sup>[7-9]</sup>,后续有学者开始分析 AI 对特定职业或具体任务的影响,认为 AI 所替代的是从事特定职业的劳动力群体,而并非整个行业或整个地区劳动力。Frey 和 Osborne<sup>[1, 10]</sup>最早提出了职业可替代风险测算方法,基于美国 O\*NET 数据库选取了 9 种 AI 难以替代的职业属性以预测美国职业可替代风险。不少学者沿用这一方法,拓展分析不同国家的职业替代情况<sup>[11-14]</sup>。之后,又有学者认为 AI 技术可能仅影响职业内的部分任务,Frey 和 Osborne 提出的测算方法可能高估了职业替代风险<sup>[15]</sup>,进而提出以任务方法测算可替代率作为修正<sup>[16, 17]</sup>。与上述研究不同,Felten 等人<sup>[18, 19]</sup>不局限于 AI 对职业的替代性,而是设计 AI 职业暴露(AI occupational exposure, AIOE)这一指标,分析不同职业受 AI 应用的影响,这种影响可能是替代性也可能是扩大就业量。

随着 ChatGPT 类大语言模型的兴起,学者们开始探讨 LLMs 对职业的影响情况。如通过供需模型来分析 ChatGPT 在短期和长期内对职业环境的影响<sup>[20]</sup>,计算 LLMs 下不同职业的 AIOE 值<sup>[3]</sup>,以及利用 GPT-4 作为辅助来计算 LLMs 下各职业的具体任务所对应的 AIOE 值等<sup>[4]</sup>。

总的来说,主流观点均认为 AI 或 LLMs 与职业之间存在替代与创造两种效应。从替代效应相关研究看, AI 或 LLMs 仅对部分职业造成替代影响,这些职业往往是具有明确规则性或者重复性较高的职业<sup>[2]</sup>。进一步看,这种替代效应还会受到受教育程度、技能等因素的影响,如受高水平教育的职业或职业所在的行业更易受到 AI 或 LLMs 的影响<sup>[4, 19]</sup>,在职业所需能力中,体力劳动、简单脑力劳动等常规能力较之思维、社交类能力更易受到 AI 或 LLMs 的影响<sup>[21]</sup>。而关于创造效应,其可能更需要长期分析,因此目前较少有研究能客观地得出结论。



关于 LLMs 对科研人员职业任务的影响，学者们展开了大量讨论，研究发现 ChatGPT 的交互对话、文本生成等功能加快了科研人员的知识获取和研究进程<sup>[22-24]</sup>；但对于撰写论文等需要深入了解文献的职业任务，大语言模型容易产生错误或者有误导性的文本内容<sup>[5]</sup>。然而已有研究较多停留在理论探讨层面，缺少综合性的实证分析，使得我们无法全面判断 LLMs 对科研人员职业的哪些任务造成了何种程度的影响。因此，本文将在前人研究基础上，聚焦分析大语言模型对特定科研人员职业的影响，探究职业要求、工资水平与职业影响之间的关系，或为科研人员进一步聚焦核心任务、核心技能提供参考。

## 2 研究方法与数据来源

本文讨论的重点在大规模语言模型（LLM）的应用对科研人员职业影响。参照 Tyna 等<sup>[4]</sup>的创新方法，本文利用 GPT-4 对 LLM 的应用对科研人员职业影响进行评估。由于本文需要探讨使用此方法对于评估 LLM 应用对科研人员职业影响的可靠性和可推广性，因此先选取部分类型科研人员进行评估。《中华人民共和国职业分类大典（2022 年版）》（以下简称《大典》）科学研究人员中的自然科学和地球科学研究人员<sup>2</sup>具有细分的职业小类和详细的工作任务描述，而医学、农业等其他类科学研究人员没有细分职业小类和详细工作任务描述，不利于本文开展基于任务内容的职业影响评估，因此选择自然科学和地球科学研究人员作为评估对象。测算过程主要分为三步：首先，《大典》中关于职业的主要工作任务描述较为笼统，而美国职业评价系统 O\*NET 对职业的定义和划分相对比较细致，因此本文将《大典》中的职业及主要工作任务与 O\*NET 中职业和任务进行映射；然后，设置特定评分标准，利用 GPT-4 对匹配上的主要工作任务进行评估，并整合到职业上，得到不同科研人员职业受影响的可能性；最后利用自然语言处理中的 LDA 模型提取职业任务中的主题特征，从而获取职业中最容易受影响和最不容易受影响的任务内容。

### 2.1 测算数据说明

O\*NET 系统是由美国劳工部组织开发的工作分析系统，实现了对现有职业

---

<sup>2</sup> 《大典》中的科学研究人员中类包括 12 个小类，其中自然科学和地球科学研究人员中有最多职业小类，主要包括从事数学、物理学、化学、天文学、生物学和地球科学研究的专业人员

的科学定义和细致划分，其中职业特征包括任务、技能、知识、能力、职业兴趣、工作价值观、工作风格等，为本文提炼职业特征和利用 LLM 进行评价提供了权威参考。在该系统中，任务、技能和知识均拥有重要性（importance）评分，取值范围为 0-100，主要由相关领域的专家经过大量调研得出，其中知识类型共有 33 项，技能类型共有 35 项；职业所需要的受教育水平由被调查者的大量反馈获得，从低到高划分为 12 级，最低级别 1 为高中以下，最高级别 12 为博后培养。本文使用的数据主要来自于 O\*NET 系统中关于职业的详细描述包括职业名称、任务、技能、知识、受教育水平等数据；就业和工资数据则从美国劳工统计局提供的 2022 年职业就业系列中获取，该数据集包括职业类型、年平均工资、就业人数和 2021-2031 年该职业预计职位空缺数量。

由于《大典》中的细分职业与 O\*NET 上的职业分类存在一定差距。因此本文根据职业描述，将大典中的自然科学和地球科学研究人员小类中 6 类细分科研人员与 O\*NET 系统中的职业进行映射，共匹配到 56 种职业、1202 条任务内容。（表 1）

表 1 《大典》与 O\*NET 职业配对情况表  
Table 1 Table of Career Matching between Da Dian and O \* NET

| 《大典》职业小类 | O*NET 职业类型数量 | 任务类型 |
|----------|--------------|------|
| 数学研究人员   | 5            | 400  |
| 天文学研究人员  | 2            | 118  |
| 化学研究人员   | 6            | 322  |
| 物理学研究人员  | 10           | 86   |
| 地球科学研究人员 | 17           | 35   |
| 生物研究人员   | 16           | 241  |

2.2 评估方法设定

本文将职业影响定义为 LLM 或由 LLM 驱动的系统的应用对科研人员执行任务细项人工投入的减少程度。LLM 通常用于指代更广泛的模态，但本文中 LLM 驱动的系统限定在仅包括视觉和图像能力。因此，本文打分的前提是在可使用 LLM 以及任务中提到的任何其他现有软件或计算机硬件工具，同时也可通过笔记本电脑访问任何常用工具（如麦克风、扬声器等），但不能访问任何其他物理工具或材料。评分标准如下：

无影响：评分为 0，意味着在保证同等完成质量的前提下，该任务不能通过

应用 LLM 或 LLM 驱动的系统来减少人工投入。

部分影响：评分介于 0 和 100 之间，意味着在保证同等完成质量的前提下，通过应用 LLM 或 LLM 驱动的系统可减少相应数值的人工投入。

完全影响：评分为 100，意味着在保证同等完成质量的前提下，通过应用 LLM 或 LLM 驱动的系统可以减少 100% 的人工投入，即任务不再需要人工参与。

本文利用 GPT-4 按照评分标准，设定特定的提示词，对所有任务进行批量打分，并将任务得分聚合到职业层次。为保证打分稳健性，采取 3 轮打分取平均的方式获得任务最终评分结果。根据 GPT-4 和人工返回的任务评分，每个职业对应的评分 $s_i$ ，由任务 $t_j$ 的评分乘以该职业内任务重要性（importance）评分 $I_j$ 转化后的权重为：

$$S_i = \sum_j \frac{I_j}{\sum_j I_j} t_j \quad (1)$$

### 3 实证结果

#### 3.1 大语言模型对职业影响的初步分析

从职业来看，6 个职业小类中影响最大的是数学研究人员，平均得分为 61.9 分。（表 2）从 56 个细分职业来看，超过 50 分的共有 17 类职业，比例为 30.4%，分别是数学家、运筹分析师、统计学家、地理学家、大气和空间科学家、环境工程师、地理信息系统技术专家等，其中数学家的影响得分高达 71.5 分；介于 30 到 50 分之间的共有 29 类，比例为 51.8%，主要为高等化学教师、核工程师、石油工程师、高等生物教师、遥感科学家等；而低于 30 分的共有 10 类职业，比例为 17.9%，分别是神经学家、组织技术学家化学技术人员、细胞遗传技术人员、细胞学家等。（表 3）

表 2 职业小类影响评估结果

Table 2 Impact assessment for occupational categories

| 职业小类     | 职业影响评估 |
|----------|--------|
| 数学研究人员   | 61.9   |
| 天文学研究人员  | 57.5   |
| 地球科学研究人员 | 47.6   |
| 物理学研究人员  | 42.3   |
| 生物学研究人员  | 41.1   |
| 化学研究人员   | 33.9   |



表 3 细分职业影响评估结果

Table 3 Impact assessment of occupational subcategory

| 职业小类     | 细分职业            | 职业影响评估 |
|----------|-----------------|--------|
| 数学研究人员   | 数学家             | 71.5   |
| 数学研究人员   | 运筹分析师           | 69.9   |
| 数学研究人员   | 统计学家            | 68.1   |
| 地球科学研究人员 | 地理学家            | 67.7   |
| 地球科学研究人员 | 大气和空间科学家        | 65.7   |
| 地球科学研究人员 | 高等环境教师          | 60.4   |
| 地球科学研究人员 | 环境工程师           | 59.8   |
| 地球科学研究人员 | 地理信息系统技术人员      | 59.7   |
| 化学研究人员   | 化学工程师           | 59.1   |
| 物理学研究人员  | 高等工程学教师         | 59.0   |
| 天文学研究人员  | 航空航天工程师         | 58.9   |
| 地球科学研究人员 | 采矿和地质工程师        | 58.3   |
| 物理学研究人员  | 物理学家            | 56.2   |
| 天文学研究人员  | 天文学家            | 56.1   |
| 地球科学研究人员 | 水文技术员           | 55.0   |
| 数学研究人员   | 高等数学教师          | 53.5   |
| 地球科学研究人员 | 高等地理教师          | 52.1   |
| 化学研究人员   | 高等化学教师          | 49.9   |
| 物理学研究人员  | 核工程师            | 49.0   |
| 化学研究人员   | 石油工程师           | 48.0   |
| 生物学研究人员  | 高等生物教师          | 47.3   |
| 地球科学研究人员 | 遥感科学家           | 47.0   |
| 物理学研究人员  | 电气工程师           | 46.8   |
| 物理学研究人员  | 高等物理教师          | 46.7   |
| 数学研究人员   | 计算机和信息研究科学家     | 46.3   |
| 地球科学研究人员 | 大气、地球、海洋和宇宙科学教师 | 46.1   |
| 地球科学研究人员 | 环境学家            | 46.1   |
| 生物学研究人员  | 生物化学家和生物物理学家    | 40.4   |
| 生物学研究人员  | 遗传学家            | 39.1   |
| 生物学研究人员  | 生物信息技术人员        | 38.9   |
| 生物学研究人员  | 微生物学家           | 38.8   |
| 生物学研究人员  | 生物与生物医学工程师      | 38.2   |
| 物理学研究人员  | 纳米系统工程师         | 38.2   |
| 生物学研究人员  | 分子和细胞生物学家       | 37.6   |
| 生物学研究人员  | 生物信息科学家         | 37.0   |
| 地球科学研究人员 | 地球科学家           | 36.5   |
| 地球科学研究人员 | 水文科学家           | 36.3   |
| 物理学研究人员  | 纳米技术工程技术人员      | 35.1   |
| 化学研究人员   | 燃料电池工程师         | 34.9   |
| 生物学研究人员  | 生物学家            | 34.9   |
| 物理学研究人员  | 光子学工程师          | 34.7   |

|          |             |      |
|----------|-------------|------|
| 生物学的研究人员 | 动物学家        | 31.9 |
| 物理学研究人员  | 电子和电器工程技术人员 | 31.1 |
| 地球科学研究人员 | 遥感技术人员      | 30.6 |
| 地球科学研究人员 | 地质技术人员      | 30.3 |
| 地球科学研究人员 | 大地测量师       | 30.1 |
| 生物学的研究人员 | 生物技术人员      | 29.7 |
| 化学研究人员   | 化学家         | 28.8 |
| 生物学的研究人员 | 组织学技术人员     | 28.8 |
| 生物学的研究人员 | 细胞技术学家      | 27.6 |
| 生物学的研究人员 | 细胞遗传技术人员    | 27.4 |
| 地球科学研究人员 | 环境工程技术人员    | 27.4 |
| 物理学研究人员  | 机械工程技术人员    | 26.7 |
| 化学研究人员   | 化学技术人员      | 26.0 |
| 生物学的研究人员 | 组织技术学家      | 23.8 |
| 生物学的研究人员 | 神经学家        | 21.3 |

### 3.2 基于 LDA 的职业特征分析

为能更好的识别自然科学和地球科学研究人员中影响较高和影响较低的职业特征的异同，本文采用文本挖掘中的潜在狄利克雷分配模型（Latent Dirichlet Allocation, LDA）对职业中的各条任务文本进行主题建模，挖掘出隐藏的语义主题。

LDA 主题模型是一种文本内容的量化分析方法，该方法假设各任务是若干个主题的概率分布，而每个主题则可以表现为若干个主题词语的概率分布，从而能够从海量文本数据中提取出具有语义信息的主题。在文档建模过程中，每个任务被解释成各个基于主题的一个分布。基于 LDA 主题模型的职业任务主题分析框架如图 3 所示。首先，我们将任务影响得分处于 $[0, 30]$ 、 $(30, 70]$ 和 $(70, 100]$ 区间划分为低、中、高影响三组，对应的任务数分别是 467、632 和 103。然后，分别对高、低影响组的职业任务进行分词、去停用词等文本预处理。最后，在数据预处理基础上进行 LDA 建模，采用主题一致性得分（Coherence Score）确定主题数，分别提取高低两组的共性特征和主题差异。（图 1）

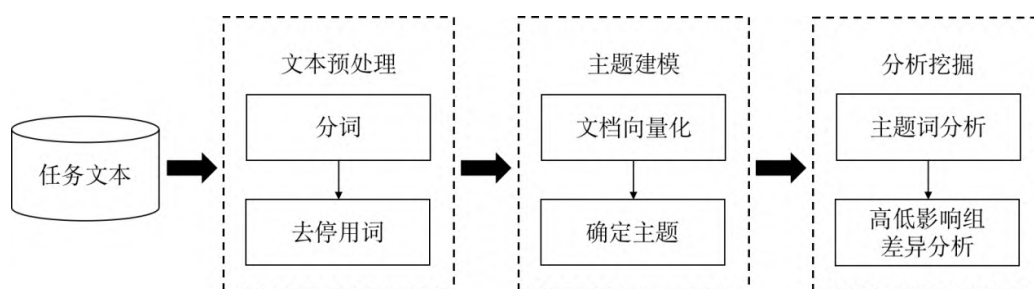


图 1 LDA 主题模型分析框架

Figure 1 LDA Theme Model Analysis Framework

### (1) 任务文本的主题分析

本研究使用主题一致性得分提炼出文本中的高频研究主题。研究发现，对于高影响组主题数为 2 时一致性得分（Coherence Score）最高为 0.461，且随着主题数的增加一致性得分趋于稳定；对于低影响组主题数为 8 时一致性得分最高为 0.433，且随着主题数增加一致性得分逐步下降。（图 2）因此分别选择 2 和 8 作为高影响组和低影响组的最佳主题数，并进行主题建模。

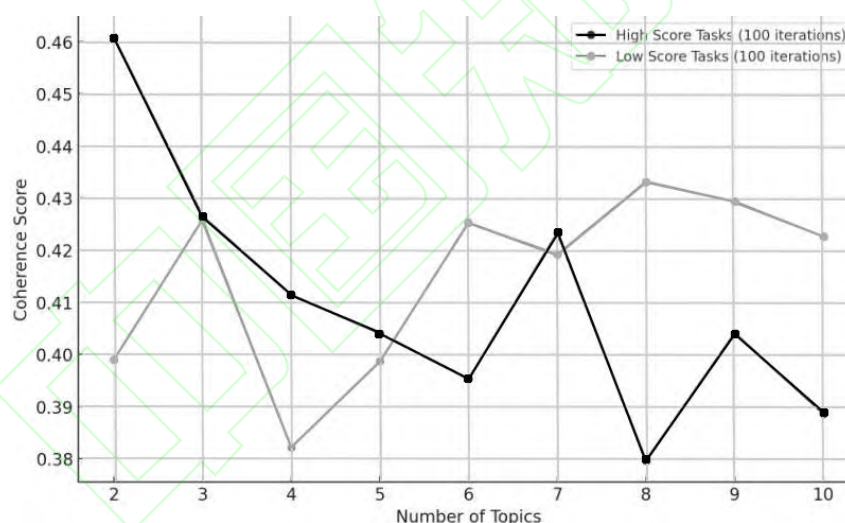


图 2 LDA 模型在多个主题上的一致性得分

Figure2 Consistency score of LDA model on multiple topics

针对高影响组，从提取出的 2 类主题的关键词中可以得出（图 3），受 LLM 影响较大的主要是与数据清洗、分析、建模及材料记录整理等相关的任务内容，具有一定的重复性和逻辑性，正是 LLM 擅长处理的任务类型。针对低影响组，从提取出的 8 类主题的关键词可以得出（图 4），受 LLM 影响较小的主要是与社会活动、研究实验、工程实操、实地调研等相关的任务内容，具有一定的随机性和特殊性，LLM 不擅长处理这样的任务类型。



图 3 高影响组主题中关键词云图

Figure 3 Keyword cloud map in high impact group themes



图 4 低影响组主题中关键词云图

Figure 4 Keyword cloud map in low impact group themes

## (2) 高低影响组差异分析

基于上述提炼出的高影响组和低影响组的主题，本研究认为：

LLM 影响比较高的任务主要集中于数据分析建模、材料梳理和报告撰写等研究工作。LLM 在对于文本的理解、处理和生成方面表现非常出色，能够适应不同的语言和语境，具有很强的泛化能力，因此在文献搜集与整理、科研论文辅助撰写以及辅助分析数据方面具有很强的能力。

LLM 影响比较弱的任务更集中在实验室设备操作、监测和维护、实地调研与数据搜集以及日常运营和管理等实务操作任务上。与之前机器及自动化发展对于职业的影响不同，LLM 对于一些需要身体力行的劳动替代性相对较弱，尤其是对于一些实验室数据的采集整理、标本制作等此类非常具体且多样化的科研

任务，LLM 的优势无法很好的发挥。但是未来，随着 LLM 及其驱动的系统的不  
断发展，这些任务的劳动效率也存在提升空间。

3.3 稳健性分析

本文通过使用线性回归方法，与 Frey 和 Osborne<sup>[1]</sup>、Webb<sup>[25]</sup>、Felten 等<sup>[3,19]</sup>  
基于自动化、AI、机器人、大语言模型等技术发展测算的职业暴露度进行比较。  
其中，Frey 和 Osborne<sup>[1]</sup>利用专家标记部分职业在自动化技术影响下的暴露度，  
再利用概率分类器对其余职业进行暴露度估计；Webb<sup>[25]</sup>通过将任务描述映射到  
专利中技术进步的描述，建立软件（Software）、机器人（Robot）和 AI 对职业  
暴露度的衡量方法；Felten 等<sup>[3,19]</sup>通过建立 AI 能力与人类能力匹配的衡量方法形  
成 AI 对于职业暴露度的评估（AIOE）；Felten 等<sup>[3]</sup>后续加入大语言模型发展的  
AI 对于职业暴露度进行评估（LLM）。

回归分析发现，本文使用的 GPT-4 评分与前人针对人工智能对职业影响的  
度量之间通常存在显著正相关，与自动化、机器人对职业影响的度量存在负相关，  
这与多项研究发现保持一致，即此前自动化技术的发展主要影响的是重复性体力  
劳动，此轮人工智能技术的发展则会对脑力劳动造成冲击。本文的评分与 Felten  
等<sup>[3,19]</sup>评估 AI 和大语言模型的职业暴露度评分在 1%的水平上呈现显著正相关，  
与 Frey 和 Osborne<sup>[1]</sup>、Webb<sup>[25]</sup>的 Robot 评分在 1%的水平上呈现显著负相关，与  
Webb<sup>[25]</sup>的 Software 评分在 5%的水平上呈现显著负相关，这说明本文使用的测  
量方法与前人针对人工智能的度量方法的结果存在一致性。（表 4）

表 4 GPT-4 评分稳健性分析  
Table 4 Robustness analysis of GPT-4 scoring

|                              | GPT-4 评分             |                       |                        |                       |     |     |
|------------------------------|----------------------|-----------------------|------------------------|-----------------------|-----|-----|
|                              | (1)                  | (2)                   | (3)                    | (4)                   | (5) | (6) |
| LLM（Felten）                  | 0.8182***<br>(0.076) |                       |                        |                       |     |     |
| AIOE（Felten）                 |                      | 0.8662***<br>(0.0647) |                        |                       |     |     |
| Automation<br>（Frey&Osborne） |                      |                       | -0.4324***<br>(0.1234) |                       |     |     |
| Software(Webb)               |                      |                       |                        | -0.3211**<br>(0.1336) |     |     |



|                            |          |          |          |          |            |          |
|----------------------------|----------|----------|----------|----------|------------|----------|
| Robot(Webb)                |          |          |          |          | -0.3913*** |          |
|                            |          |          |          |          | (0.1300)   |          |
| AI(Webb)                   |          |          |          |          |            | -0.1449  |
|                            |          |          |          |          |            | (0.1394) |
| _cons                      | -0.007   | -0.007   | -0.007   | 0.0069   | 0.0069     | 0.0069   |
|                            | (0.0753) | (0.0641) | (0.1223) | (0.1324) | (0.1288)   | (0.1381) |
| <i>N</i>                   | 53       | 53       | 53       | 54       | 54         | 54       |
| adj. <i>R</i> <sup>2</sup> | 0.688    | 0.774    | 0.178    | 0.083    | 0.132      | 0.002    |

Standard errors under estimate

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

### 3.4 职业要求、工资水平与职业影响评估

#### (1) 职业要求

本文通过回归分析探讨职业所需知识、技能和受教育水平与职业影响之间的关联强度。其中，知识、技能的类别均在 30 类以上，重要性评分为百分制，因此进行归一化处理，并采用多元线性回归中的逐步回归法去除多重共线性影响进行分析。受教育水平则通过计算不同职业不同级别受教育水平的人数占该职业人群的比例，作为不同级别受教育水平在该职业中的权重，通过级别与该级别受教育水平的权重相乘加总，计算出每个职业平均所需的受教育水平，再通过一元线性回归法进行分析。

从知识要求看，传媒、设计、经济与会计、美术、数学、物理、公共安全、社会学与人类学与职业影响程度呈显著正相关，且系数越高受到 LLM 的影响越大，其中社会学与人类学的系数最高为 0.7419，说明涉及这些知识的科研人员更容易受到 LLM 的影响。相反，地理、法律与行政、机械、人力资源、通信与职业影响程度呈显著负相关，且系数越小受到 LLM 的影响越小，其中机械的系数最小为-0.7594，说明涉及这些知识的科研人员不太容易受 LLM 的影响。（表 5）其他未入选模型的知识与职业影响程度无显著相关性。

表 5 GPT-4 评分与知识的回归分析

Table 5 Regression Analysis of GPT-4 Scoring and Knowledge

| 知识 | GPT-4 评分             |
|----|----------------------|
| 化学 | -0.2219<br>(0.1384)  |
| 传媒 | 0.3485**<br>(0.1415) |

|         |                        |
|---------|------------------------|
| 客户和个人服务 | 0.1923<br>(0.1544)     |
| 设计      | 0.3212*<br>(0.1849)    |
| 经济和会计   | 0.4102***<br>(0.1279)  |
| 美术      | 0.2072*<br>(0.1147)    |
| 地理      | -0.5860***<br>(0.1787) |
| 法律和行政   | -0.2387*<br>(0.1370)   |
| 数学      | 0.3300**<br>(0.1456)   |
| 机械      | -0.7594***<br>(0.2126) |
| 医学和牙科   | 0.2231<br>(0.1559)     |
| 人力资源    | -0.3322***<br>(0.1143) |
| 物理      | 0.3650*<br>(0.1987)    |
| 心理学     | -0.3934<br>(0.2617)    |
| 公共安全    | 0.4558***<br>(0.1422)  |
| 社会学与人类学 | 0.7419***<br>(0.2080)  |
| 通信      | -0.4163***<br>(0.1148) |
| 治疗及咨询   | -0.3696<br>(0.2195)    |

Standard errors under estimate

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

从技能要求看，协调能力、学习策略、物资管理能力、数学能力、服务导向能力、故障排除能力与职业影响程度呈显著正相关，且系数越高受到 LLM 的影响越大，说明涉及这些技能的科研人员更容易受到 LLM 的影响。相反，倾听能力、设备选择能力、指导能力、监控能力、操控能力以及社会感知能力与职业影响程度呈显著负相关，且系数越小收到 LLM 的影响越小，这些技能涉及到亲身操作和实践受 LLM 的影响相对较小。（表 6）其他未入选模型的技能与职业影响程度无显著相关性。

表 6 GPT-4 评分与技能的回归分析  
Table 6 Regression analysis of ratings and skills

| 技能   | GPT-4 评分               |
|------|------------------------|
| 倾听   | -0.2168**<br>(0.1050)  |
| 协调能力 | 0.2177*<br>(0.1183)    |
| 设备选择 | -0.3910**<br>(1.621)   |
| 安装   | -0.9444<br>(0.0783)    |
| 指导   | -0.4781**<br>(0.2039)  |
| 学习策略 | 0.6670***<br>(0.2221)  |
| 物资管理 | 0.3343***<br>(0.0899)  |
| 数学   | 0.2057**<br>(0.0765)   |
| 监控   | -0.2155**<br>(0.1042)  |
| 操作控制 | -0.5296***<br>(0.1585) |
| 操作监控 | -0.3124<br>(0.1854)    |
| 说服   | 0.3089**<br>(0.1199)   |
| 服务导向 | 0.2556**<br>(0.1099)   |
| 社会感知 | -0.8129***<br>(0.1165) |
| 故障排除 | 0.3442**<br>(0.1633)   |

Standard errors in parentheses

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

从受教育水平看，自然科学和地球科学研究人员平均受教育水平为硕士，最低也需要受过大学教育，其中平均受教育水平达到博士级别的有神经学家、天文学家和物理学家。（图 5）。从回归结果来看，受教育水平与职业影响程度呈显著正相关，受教育水平越高的科研人员更容易接触到 LLM 或 LLM 驱动的系统，因此受教育水平越高越容易受到 LLM 的影响。（表 7）

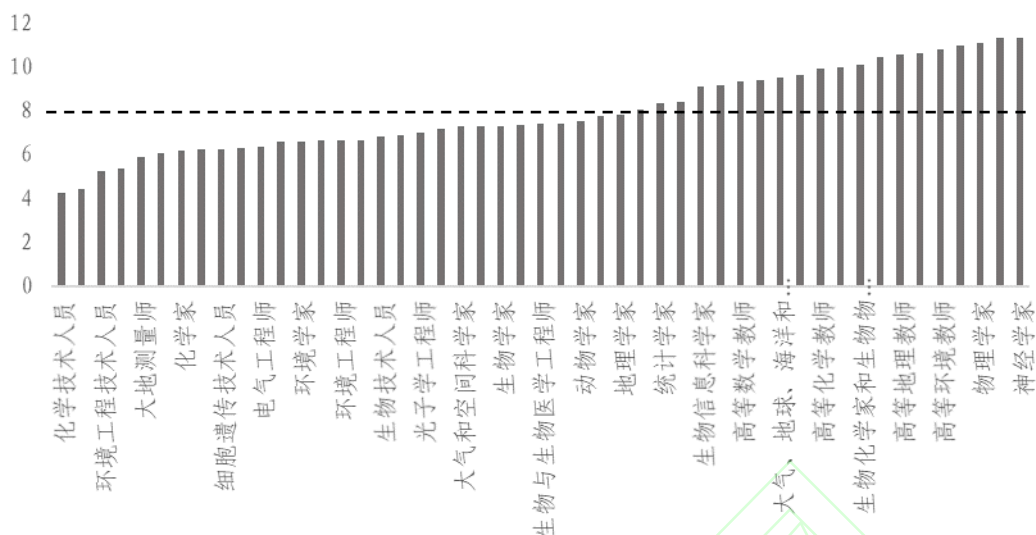


图 5 GPT-4 评分与受教育水平的关系

Figure 5 The relationship between GPT-4 score and educational level

表 7 GPT-4 评分与受教育水平的回归分析

Table 7 Regression analysis of GPT-4 score and educational level

| 受教育水平 | GPT-4 评分             |
|-------|----------------------|
| 受教育水平 | 0.3345**<br>(0.1375) |

Standard errors in parentheses

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## (2) 工资水平

由于国内各细分职业的工资水平较难获取，本文参考美国劳动统计局对不同职业开展就业调查，获得该职业年平均工资、工作人数和 2021-2031 年该职业预计职位空缺数量。首先，参照 Tyna 等<sup>[4]</sup>的处理方式，将年平均工资、就业人数、预计职位空缺数量进行对数处理。其次，采用多元线性回归中的逐步回归法，探讨职业工资水平与职业影响之间的关联强度。

从散点图和回归结果来看，在自然科学和地球科学研究人员中，年平均工资与职业影响程度呈显著正相关，高薪职业更容易接触到 LLM 或 LLM 驱动的系统，因此薪资越高越容易受到 LLM 的影响；由于就业人数、预计职位空缺与职业影响程度的趋势非常接近，因此在变量选择时，保留了就业人数而剔除了预计职位空缺，即就业人数与职业影响程度呈显著负相关，说明就业人数较多的职业的工作内容还无法被 LLM 或 LLM 驱动的系统影响。（图 6，表 8）

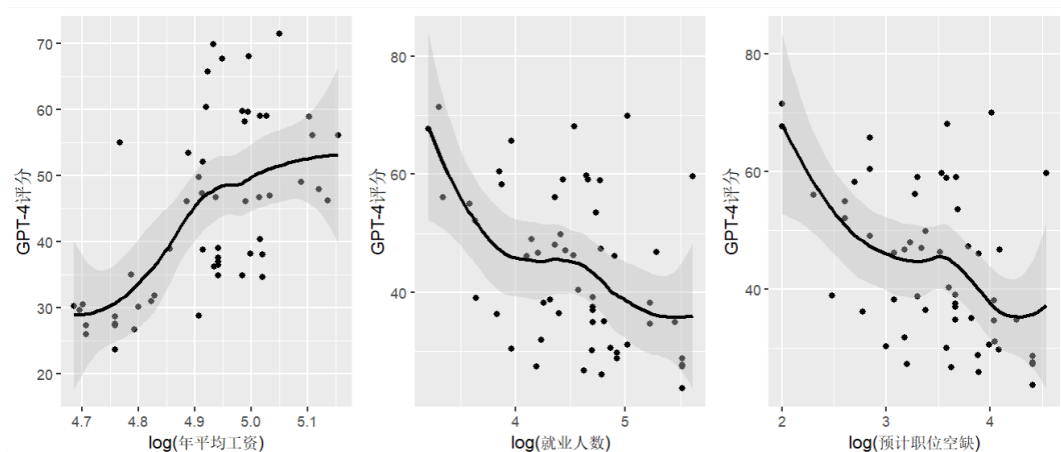


图 6 GPT-4 评分与工资水平的关系

Figure 6 The relationship between GPT-4 score and salary level

表 8 GPT-4 评分与工资水平的回归分析

Table 8 Regression analysis of GPT-4 score and wage level

| 工资水平       | GPT-4 评分              |
|------------|-----------------------|
| log (工资水平) | 58.201***<br>(11.063) |
| log (就业人数) | -7.542***<br>(2.282)  |

Standard errors in parentheses

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

#### 4 结论与建议

本文以中国职业分类大典中的自然科学和地球科学研究人员为例，利用 GPT-4 评估科研人员主要工作任务受大语言模型的影响，并验证该方法的稳健性后，得到研究结论如下：（1）大语言模型对不同职业都存在影响，受影响较大的职业占比为少数；（2）大语言模型在处理大规模的文本检索、理解生成以及标准化的分析建模具有较强的优势，因此需要简单脑力劳动的任务受影响较大，而对于需要社会交往、亲身实践的任务影响较小；（3）具有一定逻辑推理类特征的知识技能更容易受大语言模型的影响，而偏实务型、互动型、情感性的技能受影响较小；（4）受教育程度越高、工资水平越高越容易受大语言模型的影响，大语言模型在一些需要大量知识储备和高级技能的工作内容的支撑方面表现出了巨大的潜力。

基于上述研究结果，本文提出如下建议。首先，加强人工智能新技术在科研发现中的应用。随着以大语言模型为代表的人工智能的蓬勃发展，人工智能技术



在加速科学发现上已出现诸多典型案例，如利用 GNoMe 发现新材料、利用 GraphCast 提供更为准确的天气预报等，这也充分说明智能技术对于科研效能提升的重要作用。其次，巧用智能技术红利生成科研人才红利。一方面，智能时代高等教育改革，要把人工智能技能作为科研人才的培养的重要内容；另一方面，探索人工智能技术在弥合跨学科壁垒的应用，提高学科交叉带来的科学发现的可能性。最后，需要注意使用人工智能技术的局限和伦理问题。大语言模型的发展不仅存在一些技术上的局限亟待突破，如存在非真实性及偏见性输出、缺乏实时自主学习能力以及强依赖于数据集质量和数量等，还存在一些安全上的风险如数据隐私、信息安全、数据伦理等，因此在技术应用时也要加强评估。

本文还存在以下局限。本文重点研究了应用 GPT-4 进行职业影响评估的稳健性，因此只选取了中国职业分类大典中的自然科学和地球科学研究人员进行研究，存在一定的片面性。

#### 参考文献

- [1] Frey C. B., Osborne M. A. The future of employment: How susceptible are jobs to computerization?[J].Technological Forecasting and Social Change, 2017(114): 254-280.
- [2] 刘涛雄, 潘资兴, 刘骏. 机器人技术发展对就业的影响——职业替代的视角[J]. 科学学研究, 2022, 40 (03): 443-453. Liu T X, Pan Z X, Liu J. The impact of robotics development on employment: a perspective from occupation replacement[J]. Studies in Science of Science, 2022, 40 (03): 443-453.
- [3] Felten E., Raj M., Seamans R. How will language modelers like ChatGPT affect occupations and industries?[J/OL]. <https://arxiv.org/pdf/2303.01157>, (2023-03-18)[2024-05-20].
- [4] Eloundou T., Manning S., Mishkin P., et al. GPTs are GPTs: an Early Look at the Labor Market Impact Potential of Large Language Models[J/OL]. <https://arxiv.org/abs/2303.10130>, (2023-08-21)[2024-05-20].
- [5] Van Dis E. A., Bollen J., Zuidema W., et al. ChatGPT: five priorities for research[J]. Nature, 2023, 614 (7947): 224-226.
- [6] Zhao W., Zhou K., Li J., et al. A Survey of Large Language Models[J/OL]. <https://arxiv.org/abs/2303.18223>, (2023-11-24)[2024-05-20].
- [7] Acemoglu D., Restrepo P. Robots and Jobs: Evidence from US Labor Markets[J]. Journal of Political Economy, 2020, 128 (6): 2188-2244.
- [8] 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[J]. 经济研究, 2020, 55 (10): 159-175. Wang Y Q, Dong W. How the rise of robots has affected China's labor market: evidence from China's listed manufacturing firms[J]. Economic Research Journal, 2020, 55 (10): 159-175.
- [9] 韩民春, 韩青江, 夏蕾. 工业机器人应用对制造业就业的影响——基于中国地级市数据的实证研究[J]. 改革, 2020 (03): 22-39. Han M C, Han Q J, Xia L. The impact of industrial robot

application on manufacturing employment: an empirical study based on the data of prefecture level cities in China.[J].Reform, 2020 (03): 22-39.

[10] Frey C. B., Osborne M. A. The future of employment: How susceptible are jobs to computerisation?[J/OL].Oxford Martin School Working Paper , [https://www.oxfordmartin.ox.ac.uk/publications/the-future-of-employment\(2013-09-01\)\[2024-05-20\]](https://www.oxfordmartin.ox.ac.uk/publications/the-future-of-employment(2013-09-01)[2024-05-20]).

[11] Brzeski C., Burk I. Die Roboter kommen:Folgen der Automatisierung für den deutschen Arbeitsmarkt. [J].INGDiBa Economic Research, 2015 (30): 1-7.

[12] David B. Computer technology and probable job destructions in Japan: An evaluation[J].Journal of the Japanese and International Economies, 2017 (43): 77-87.

[13] 李磊,何艳辉.人工智能与就业——以中国为例[J].贵州大学学报(社会科学版), 2019, 37 (05): 13-22.Li L , He Y H. Artificial intelligence and employment: the example of China.[J].Journal of Guizhou University(Social Sciences), 2019, 37 (05): 13-22.

[14] 龚遥,彭希哲.人工智能技术应用的职业替代效应[J].人口与经济, 2020(03): 86-105.Gong Y, Peng X Z.A Research on occupational substitution effect of artificial intelligence[J].Population & Economics, 2020 (03): 86-105.

[15] Autor D. H., Handel M. J. Putting Tasks to the Test: Human Capital, Job Tasks, and Wages[J].Journal of Labor Economics, 2013, 31 (S1): S59-S96.

[16] Arntz M., Gregory T. Zierahn U. The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis[J].OECD Social, Employment and Migration Working Papers, 2016, NO.189:1-34.

[17] Tolan S., Pesole A., Martínez-Plumed F., et al. Measuring the Occupational Impact of AI: Tasks, Cognitive Abilities and AI Benchmarks[J].Journal of Artificial Intelligence Research, 2021 (71): 191-236.

[18] Felten E. W., Raj M., Seamans R.A Method to Link Advances in Artificial Intelligence to Occupational Abilities[J].AEA Papers and Proceedings, 2018, (108): 54-57.

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

[20] Zarifhonorvar.“Economics of ChatGPT: a labor market view on the occupational impact of artificial intelligence[J].Journal of Electronic Business & Digital Economics, 2023.

[21] 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[J].管理世界, 2022, 38 (07): 60-79.Wang L H, Hu C M. Artificial intelligence technology, task attribute and occupational substitutable risk: empirical evidence from the micro-level[J].Journal of Management World, 2022, 38 (07): 60-79.

[22] 游俊哲.ChatGPT 类生成式人工智能在科研场景中的应用风险与控制措施[J].情报理论与实践, 2023, 46 (06): 24-32.You J Z.Risk and control measures of ChatGPT-like generative artificial in research[J].Information Studies: Theory & Application, 2023, 46 (06): 24-32.

[23] 王树义,张庆薇. ChatGPT 给科研工作者带来的机遇与挑战[J].图书馆论坛, 2023, 43 (03): 109-118.Wang S Y, Zhang Q W. ChatGPT's opportunities and challenges for researchers[J].Library Tribune, 2023, 43 (03): 109-118.

[24] Graf A., Bernardi R. E. ChatGPT in Research: balancing ethics, transparency and advancement[J].NeuroScience, 2023(515): 71-73.

[25] Webb M. The impact of artificial intelligence on the labor

market[J/OL].[https://www.michaelwebb.co/webb\\_ai.pdf](https://www.michaelwebb.co/webb_ai.pdf), (2020-01-01)[2024-05-20].

