

大语言模型对员工效率的影响

——基于通信大数据的实证研究

陆迪 吴祁 张迎峰*

摘要: 本文以国产大语言模型“文心一言”向全社会开放为冲击,利用通信大数据构建企业层级员工加班率面板数据,采用双重差分方法,系统分析大语言模型对企业运营成本的影响。研究发现,受大语言模型冲击较大的公司员工加班率显著降低,劳动报酬增加,企业运营成本下降。进一步异质性分析表明,这种影响在不同股权结构和员工结构的公司间存在差异,且对青壮年和中低收入员工的影响更为显著。本文的研究结论为利用人工智能推动企业高质量发展提供了重要的经验证据和政策启示。

关键词: 大语言模型;企业运营成本;通信大数据

DOI: 10.13821/j.cnki.ceq.2026.03.12

一、引言

随着人工智能(Artificial Intelligence, AI)技术的迅猛发展,以 ChatGPT 为代表的大语言模型(Large Language Model, LLM)正以史无前例的速度重塑社会与经济格局。这些模型凭借其强大的语言理解和生成能力,大幅降低了技术应用的门槛,使得人工智能技术渗透到更广泛的工作场景中(Dell'Acqua et al., 2023)。在此背景下,探究 LLM 对组织和个人工作的影响,已成为学者、企业乃至公共部门亟须关注的课题(Berg et al., 2023)。然而,当前研究受数据限制,LLM 在企业微观运营层面的影响尚未得到充分发掘。尤其在企业日常运营中,LLM 如何改变员工工作模式以及影响企业运营成本等问题仍待研究。

现有文献大多考察了 LLM 之前的人工智能技术在宏观和微观层面的经济影响,对 LLM 的影响研究较少。在宏观层面,已有研究广泛探讨了人工智能对

* 陆迪,深圳大学微众银行金融科技学院、金融科技与数智管理研究中心;吴祁,北京大学光华管理学院;张迎峰,联通(广东)产业互联网有限公司。通信作者及地址:吴祁,北京市海淀区颐和园路 5 号北京大学光华管理学院 2 号楼 350 室,100871;电话:010-62747554;E-mail: qiwu@gsm.pku.edu.cn。本研究得到国家自然科学基金青年项目(72303008、72402140)以及专项项目(72342031)的资助。作者感谢联通(广东)产业互联网有限公司数据智能事业部对本文的数据支持。本文作者衷心感谢三位匿名审稿人和编委会提出的宝贵建议。文责自负。

经济增长的推动作用(林晨等,2020;陈彦斌等,2019)、对劳动力市场的重塑效应(Acemoglu and Restrepo, 2018)以及对不同技能劳动力产生的差异化影响(郭凯明等,2023)。这些研究为理解人工智能的经济效应奠定了坚实的理论基础。在微观层面,研究者们聚焦人工智能技术在特定行业或场景中的应用,如制造业中工业机器人对企业劳动力雇佣的影响(Mokyr et al., 2015; Acemoglu and Restrepo, 2018; 孔高文等,2020;王永钦和董雯,2020;李磊等,2021)以及企业数字化转型对劳动力需求的改变(吴非等,2021;肖土盛等,2022;叶永卫等,2022;姚加权等,2024;金星晔等,2024)。^①然而,这些研究大多基于企业年度财务数据或局限于特定的实验场景,难以全面捕捉人工智能技术在企业日常运营中的实际影响。

目前,上述研究尚未充分涵盖对LLM最新发展的探讨。这类模型凭借其持续增强的能力、显著提升的易用性以及对各行业产生的深远影响,已成为当前研究的重要议题。现有关于LLM经济效应的前沿文献主要采用实验或准自然实验方法,聚焦特定领域,包括咨询工作(Dell'Acqua et al., 2023)、中级专业写作任务(Noy and Zhang, 2023)、法学院考试(Choi and Schwarcz, 2023)、分析师预测(Michaely and Grennan, 2021; Bertomeu et al., 2025)、电话销售(Jia et al., 2024)、客服代理(Brynjolfsson et al., 2025)、战略决策(Doshi et al., 2025)以及常见职业任务(Li et al., 2024)等场景。但这些研究多集中于特定行业或任务,难以反映LLM在更广泛行业中的普遍影响。^②此外,LLM的通用性和广泛适用性使其对各行业的差异化影响仍待系统考察。

LLM凭借低使用成本与低操作门槛,可激发企业员工自发使用动机,通过提升任务效率与创造力,优化工作流程,减少加班,进而提高工作效率。进一步地,LLM通过在流程化岗位实现自动化处理以及在沟通岗位提供个性化策略建议等优化措施,降低人力与运营成本。对此,本文以国产大语言模型“文心一言”的发布与开放作为冲击,利用一个独特的通信大数据集,通过构造企业层面的员工日度加班高频数据,打开企业内部日常运营的“黑箱”,采用双重差分方法识别LLM在中短期对不同行业企业员工日常工作模式以及企业运营成本的影响。研究发现:第一,受LLM冲击较大的公司员工加班率在“文心一言”开放

^① 此外,陈琳等(2024)运用企业招聘大数据集发现,尽管人工智能降低了企业对常规职业劳动力的需求,却增加了企业对非常规职业劳动力的需求。Han et al.(2025)运用招聘数据发现人工智能增强了企业韧性。

^② 另有文献基于职位发布数据和企业电话会议的文本数据,探讨了生成式AI对金融职业(Eisfeldt and Schubert, 2024)、企业价值(Eisfeldt et al., 2023)以及企业决策(Jia et al., 2026)的影响。Eloundou et al.(2024)基于职业描述构建了LLM暴露度,并估计大约1.8%的工作岗位可能会受到通用LLM的影响。北京大学国家发展研究院与智联招聘研究团队于2023年联合发布了《AI大模型对我国劳动力市场潜在影响研究》,基于智联招聘平台大数据,构建不同职业的“大语言模型影响指数”,反映其易受大模型技术影响的程度,并在2024年更新了报告。此外,部分学者关注了生成式AI可能加剧的社会不平等问题(Capraro et al., 2024)。

后显著降低,动态分析显示这一效果在冲击发生后持续至少3个月,且在限制样本、控制干扰因素、更换度量方式及拓展到非上市公司样本后结果仍稳健。员工加班率的显著下降从侧面反映了LLM的广泛运用可能会提高员工的工作效率、沟通效率、团队合作效率从而降低企业运营成本。我们的企业问卷调查结果也支持大语言模型使用能节约员工工作时间、提高员工工作效率。第二,受LLM冲击程度越高的公司,支付给员工的劳动报酬显著增加,加班率降低与劳动报酬增加共同支持了员工工作效率提升机制,且LLM冲击最终降低了企业运营成本,表现为总营业成本率和销售费用率显著降低。第三,异质性分析表明,LLM冲击在国有企业与非国有企业、大公司与小公司间的差异不显著,但在管理层持股比例高的公司、机构持股比例低的公司、行政人员占比高的公司,以及青壮年员工和中低收入员工中,降低加班率的效应更显著。

本文的学术贡献主要体现在以下三个方面。第一,本文的研究建立在近年来关于人工智能技术的经济效应的广泛文献基础之上,对近期有关LLM的研究做出了有益的补充与贡献。鉴于LLM具有通用性和广泛适用性的特征,其对各行业的差异化影响亟待系统考察。基于此,本文通过量化分析LLM发布对全行业的不同冲击程度,评估其对不同行业员工工作模式的影响,从而对现有研究形成有益补充。第二,在研究数据上,本文创新性地利用高频通信大数据集,拓展了对LLM在企业运营成本影响上的机制分析,为数字技术与其他要素如何有机结合的研究提供了有益参考。已有研究证实,通信大数据能有效捕捉社会互动(Atkin et al., 2022)、劳动力流动(Barwick et al., 2023)、城市经济活动(Chen and Pope, 2020)等微观动态。本文通过追踪手机信令数据,识别出企业员工稳定工作地以及对应的公司,并构建较为准确且高频的员工加班率变量,从而拓展了通信大数据等新型数据要素在经济学研究中的运用,为现有文献提供了新的测度方式和分析视角。第三,在现实意义上,本研究在微观企业层面考察了如何通过人工智能技术融合以劳动力为代表的其他生产要素推动企业高质量发展,并为实现员工福利和企业发展的互利共赢局面提供了经验证据与政策支持。

二、研究假说

近年来,学术界对人工智能的经济效应进行了广泛研究。然而,当前LLM的技术范式与早期基于规则系统或专用机器学习模型的人工智能技术存在本质性差异。LLM通过海量数据预训练形成了通用任务处理能力,并且无需特定训练即可通过自然语言交互完成任务。在中国本土大语言模型刚开始对社

会开放早期,企业员工对大语言模型的自主采纳已形成独特的“技术下沉”现象。^①这类模型凭借其不断增长的能力、易用性以及对各行业的广泛影响脱颖而出,但其对各行业的深远影响亟待深入研究。因此,本文主要探讨在 LLM 对社会开放早期,尤其是中国国产大语言模型推出背景下,其在短期内对企业员工工作效率与企业运营成本的影响。

从技术角度来看,LLM 具有使用成本低和操作门槛低的特点,使得企业员工具备自发使用的动机。许多 LLM 提供免费版本或开源模型,允许用户以零成本完成基本的工作任务处理。^②这种边际成本趋近于零的技术供给,使得个体层面的自发使用成为可能。此外,在任务适配性方面,大语言模型的交互界面突破传统企业软件的功能限制,用户仅需通过自然语言指令即可完成文档处理、数据分析、多语种沟通等复杂任务。这种基于自然语言交互的特性,大幅降低了技术采纳门槛,使员工能够快速上手并将 LLM 应用于实际工作场景。

从员工个体角度来看,LLM 从三个方面提升员工的工作效率。一是语言生成效率提升。通过辅助撰写、自动答复等功能,减少员工在撰写报告、回复邮件等语言密集型任务上所花费的时间。例如,在撰写商务报告时,LLM 可以快速生成初稿,员工只需进行修改和完善,从而节省大量时间用于更具创造性和价值导向的工作(Brynjolfsson et al., 2017; Goldfarb et al., 2023)。二是任务聚焦能力增强。通过自动化处理重复性、规范性任务,释放员工的时间与精力,使其能够专注于处理高价值任务。例如,员工可将数据输入、文件整理等重复性工作交给 LLM 处理,而将更多时间和精力投入创新性工作或战略规划等高价值任务中,从而提升整体工作效率。三是沟通协同效率改善。LLM 通过提升员工在内部沟通与客户服务中的响应速度,提高团队协作效率和客户满意度。例如,在团队协作中,LLM 可以快速整理会议纪要、提炼关键信息;在客户服务中,LLM 能够快速理解客户需求并做出回应。基于以上分析,本文提出了假说 1。

假说 1 大语言模型有利于提高企业员工的工作效率。

LLM 通过提高员工工作效率,最终体现在降低企业的运营成本上。具体而言,LLM 可以优化企业内部流程,并在不同类型的工作任务中发挥作用。首先,在高度流程化的岗位(如运营、财务和合规)中,LLM 通过自动化文档处理和数据分析,减少了员工工作量,同时提高了任务的准确性和效率。例如,LLM

^① 附录 I 对大语言模型的发展以及国内外的应用做出介绍和总结。限于篇幅,附录未在正文列示,感兴趣的读者可在《经济学》(季刊)官网(<https://ceq.ccer.pku.edu.cn>)下载。

^② 例如,百度的“文心一言”提供了免费的 API Key,每个 API Key 每天可支持多达 10 000 次请求,新注册用户还可获得额外的 1 000 次免费调用。

可以快速处理重复性任务,如数据录入、报表生成和合规性检查,从而显著降低人力成本。其次,在需要语言沟通与信息处理的岗位(如销售)中,LLM通过个性化策略优化和流程改进,进一步降低了运营成本。销售工作通常依赖于灵活的语言沟通和精准的客户策略,而LLM能够分析市场数据和客户行为,为销售人员提供个性化的销售策略建议。例如,LLM可以帮助销售人员优化客户沟通脚本,提高客户获取和维护的效率,从而降低销售成本。相关研究(Jia et al., 2024)还表明,人工智能技术通过提升员工回答问题的创造力和效率,能够显著提高销售额。在销售岗位中,大语言模型的应用具有更高的灵活性和技术自主性,这使得其效果更加显著。此外,在LLM对社会开放的早期阶段,员工的自主使用行为也对企业的运营成本产生了积极影响。这种自主使用行为不仅减少了企业对正式培训和技术支持的依赖,还通过员工的主动探索和创新,进一步优化了工作流程。综上所述,LLM在流程化较高的部门(如管理和财务)中通过提高规范性和效率降低了人力成本和时间成本,而在技术应用灵活性较高的岗位(如销售)中则通过语言和策略的优化提升团队效率,降低了客户获取和维护的成本,从而实现企业多维度以及整体运营成本的降低。基于以上分析,本文提出了假说2。

假说2 大语言模型有利于降低企业的运营成本。

三、研究设计

(一) 样本选取与数据来源

本文将研究对象聚焦于2023年广东省所有A股上市公司^①。作为数字经济产业大省,广东近年来不遗余力发展人工智能。^②因此,本文选取的广东省上市公司可以被认为在大语言模型发展和应用初期即有较高概率接触并使用大语言模型。以这些公司作为研究样本,可以较好地反映出大语言模型对不同行业公司在中短期造成的冲击和影响,本文的结论与发现也可为研究其他地区类似的问题提供有价值的经验证据和参考。

本文共使用如下三套数据:

第一,加班数据。本文借助底层的脱敏通信大数据,构建了广东省全部上

^① 在稳健性检验中,我们对广东省的非上市公司进行了抽样,进一步收集并计算了非上市公司的加班率数据,详细结果请见附录II。

^② 2022年,广东全省人工智能核心产业规模达1500亿元;全省人工智能相关注册企业约17万家,核心企业900多家,居全国首位。2023年11月,广东出台“通用人工智能发展22条”,探索打造“粤港澳大湾区数据特区”(https://www.gd.gov.cn/zwgk/zcjd/mtjd/content/post_4283645.html,访问时间:2024年5月17日)。

市公司总部办公地址^①自2023年5月1日到2023年12月31日的日度员工加班比例数据。通信大数据提供了微观个体的高频轨迹信息,通过对匿名化的员工日度工作时间日志进行分析,并通过办公地址匹配对应公司信息进行加总,我们计算了每家上市公司每日总和以及分人口特征的加班员工数量和所占比例。^②

第二,公司层面数据。为衡量公司运营成本和效率,我们使用国泰安数据库(CSMAR)的中国上市公司财务数据,获得了上市公司办公地址,2023年季度层面支付给职工的现金、总营业成本率、销售费用率等指标。

第三,其他补充性数据。包括:①企业招聘数据。企业招聘文本数据来源于智联招聘网站,为2023年日度数据,包含公司名称、公司地址、行业、工作地点、岗位、职位描述、待遇、福利待遇、经验、学历、招聘人数、发布时间等字段。作者根据职位描述文本构建了企业层面的LLM暴露度指标。②公司问卷调查数据。2024年9月3日至2024年9月22日,作者对三家位于深圳的上市公司总部(分部)进行了问卷调查,对本文的研究结论进行了补充和支持。^③调查基本覆盖了被调查企业(总部或分部)不同岗位,涉及管理类、技术类、营销与销售类、财务类、人力资源类、行政类与客户服务类岗位,共计61名员工。

(二) 主要变量定义

1. 被解释变量

(1) 企业加班率。本文通过通信大数据构造了企业加班率变量。参考Barwick et al.(2023)具体步骤如下:首先,我们识别出员工稳定工作地以及对应的上市公司。如果一位员工在过去60天内早7点到晚7点时间段内,长时

① 本文考察的是上市公司总部的办公地址,主要基于以下两点原因:其一,上市公司的分公司和工厂的地址数据难以全面收集,而总部地址信息相对集中且更具可靠性;其二,总部员工大多负责企业的公司管理、治理和运营等职能,这些岗位的工作内容与大语言模型的潜在冲击更为密切相关。因此,大语言模型的影响也更能通过总部员工的加班率这一指标得以体现。

② 然而,需要承认的是,通信大数据的特性带来了一定的客观存在的局限性。例如,在中央商务区等高楼密集区域,一个地址可能同时对应到多家企业。这种情况下,不同企业的员工可能被归为同一地址某个企业的员工,导致度量误差并可能造成估计偏误。针对这一问题,本文运用了广义双重差分估计,通过比较处理组与控制组的差异,尽可能降低这一局限对估计结果的影响。如果一个地址对应的其他多家企业是相似行业的企业,该设定在一定程度上可以缓解这一担忧。如果一个地址对应的其他多家企业属于完全不同的行业,则可能将受大模型冲击小的企业和受大模型冲击大的企业混杂在一起,从而拉低处理组内受到大模型冲击大和小的企业的效应差异。这意味着,模型的估计效应可能会偏低。因此,数据的局限性使得本文的估计结果更偏向保守。尽管如此,这种偏误在本文的研究背景下应当较为有限。主要原因在于,中国上市公司总部通常规模较大,且其办公地址往往独立于规模较小的非上市企业或其他组织。这种规模差异使得同一地址内的混淆效应对加班率测算的影响可能较小。在附录II稳健性检验中,我们依据上市公司详细的办公地址,排除了公司地址为非独栋办公楼的企业(如云润大厦15—19层)。子样本的回归结果仍然是稳健的。

③ 这三家上市公司分别为:大晟时代文化投资股份有限公司总部(股票代码600892,主要业务集中在影视投资与制作、游戏开发等板块)、中国工商银行深圳某支行(股票代码:601398)以及深圳市瑞凌实业股份有限公司总部(股票代码:300154,主要业务为逆变焊割设备供应)。前两家企业属于LLM暴露度高的行业,最后一家企业属于LLM暴露度低的行业。LLM暴露度指标详见下一节介绍。

间停留在公司 A 所在地点且超过 30 天以上,我们将该员工定义为公司 A 的员工。相似地,我们通过晚 10 点到第二天早 7 点的员工长时间停留地点识别出员工的居住地。接下来,我们对该员工是否加班进行定义。如果该员工在晚上 7 点之后仍停留在工作地且持续超过 4 小时,则识别其为当天加班工作。最后,将个人数据加总到公司层面,用该公司当天的加班人数除以总员工数,计算出该公司当日的员工加班率。^①

此外,我们还对不同人口学特征的员工加班率进行了汇总,包括加班员工中女性员工数量及其在总员工中的占比,加班员工中不同年龄段(16—22 岁,23—35 岁,36—45 岁,46—60 岁)员工数量及占比,不同收入群体(收入前 10%,10%—25%,25%以后)员工数量及占比。其中员工收入水平用其居住地住房市场价的估值作为代理变量。

(2) 企业运营成本。参考以往相关研究(Ang et al.,2000;杨继生和阳建辉,2015,温素彬等,2022),本文分别考察了企业的总营业成本率、销售费用率、管理费用率、财务费用率以及销售期间费用率。

2. 解释变量

(1) 行业受 LLM 冲击指数。参考最前沿的相关研究 Eloundou et al. (2024)的方法,我们将中国上市公司与 Eloundou 等(2024)测度的 LLM 冲击程度按行业进行匹配,构造出上市公司所在行业受 LLM 冲击指数。第一步,我们将北美行业分类(NAICS 3)与国际标准行业分类(ISIC Rev. 4.0)^②进行匹配,再与中国《国民经济行业分类》^③进行匹配。第二步,根据《国民经济行业分类》与《上市公司分类与代码》^④的行业门类代码、行业名称和细分门类名称,我们将国民经济行业分类代码与上市公司行业分类代码进行对应匹配,从而得到中国上市公司行业所对应的 LLM 冲击指标。

图 1 列出了匹配后的上市公司行业平均 LLM 冲击程度。尽管使用 GPT4 和人工对职业进行标注,得到的测度值略有不同,但二者之间也存在相当程度的共性,反映出测度的代表性。按照 GPT4 对职业任务标注测度,冲击程度最高的行业为互联网和相关服务,其次为软件和信息技术服务业,广播、电视、电影和影视录音制作业,保险业和资本市场服务业。按照人工对职业任务标注测度,上述行业也依次为冲击程度最高的前五行业。

^① 值得注意的是,如果员工在夜间出现在公司并持续超过 4 小时,但其日度总工作时长并没有超过 8 小时,那么本质上并不属于加班工作,因此有可能会把一些上班时间晚或者上夜班的职业误判为加班。对此,我们进一步计算了加班率的替代指标“日度工作时长超过 8 小时的员工占比”,即公司每日工作超过 8 小时员工数除以总员工数,作为日度加班率的稳健性测度,用于稳健性检验。

^② 资料来源:加拿大统计局, <https://www.statcan.gc.ca/en/statistical-programs/document/naics-2022-istic4>, 访问时间:2024 年 2 月 25 日。

^③ 资料来源:《国民经济行业分类》(GB/T 4754—2017)附录 C。

^④ 2005 年发布并实施《上市公司分类与代码》,2024 年修订后更名为《上市公司行业统计分类与代码》。本文使用的是 2024 年修订前版本。

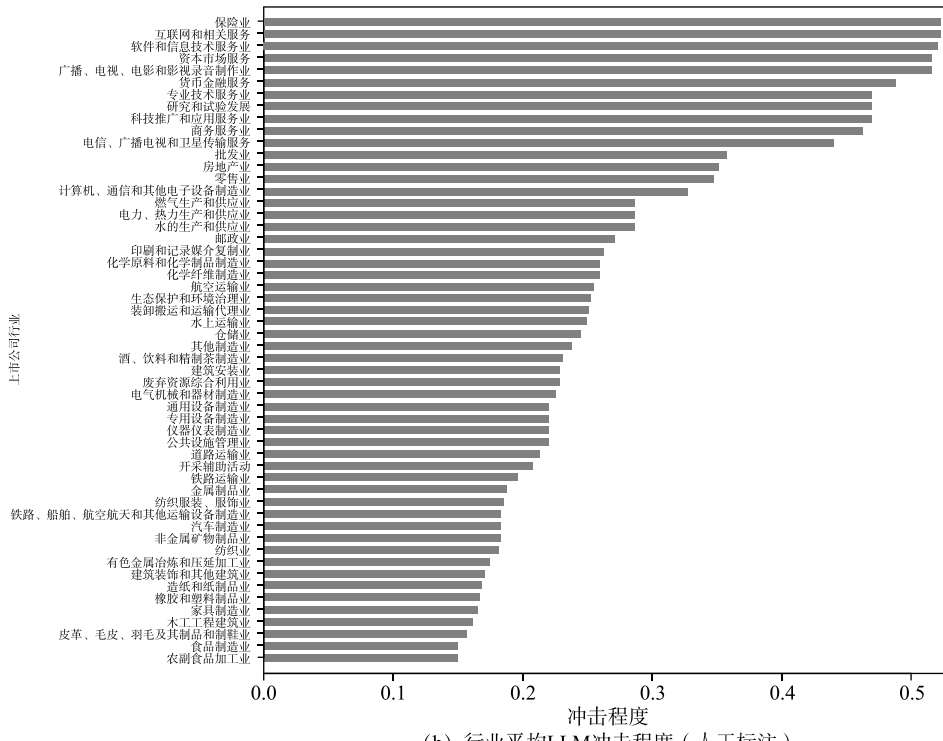
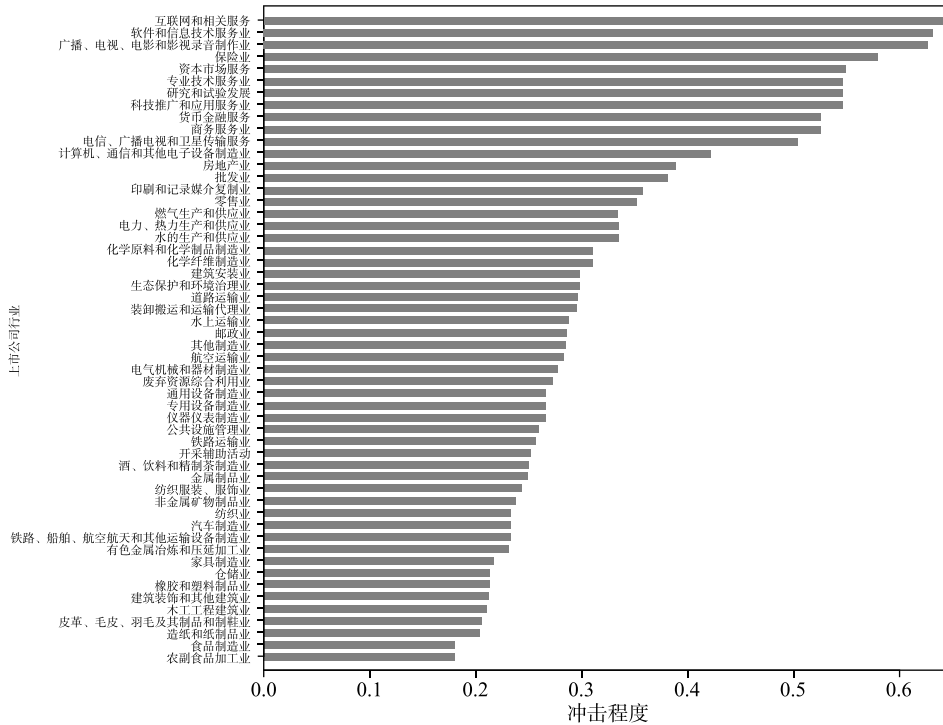


图 1 上市公司行业平均 LLM 冲击程度

(2) 企业受 LLM 冲击指数。文献中构建企业层面人工智能影响测度的方法有两种。第一种为 Eisfeldt et al.(2023) 基于 LinkedIn 员工数据, 加总计算出企业层面的 LLM 暴露度。由于中国企业员工在 LinkedIn 上披露信息较少且存在选择偏误^①, 这一方法很难应用于中国上市公司。第二种方法是采用公司招聘文本数据构造企业层面的 LLM 暴露度, 这一方法在考察人工智能影响的研究中较为常见(姚加权等, 2024)。^② 本文借鉴第二种思路, 采用企业招聘文本数据构建企业层面的 LLM 暴露度指标。

具体地, 参考 Alekseeva et al.(2021) 和姚加权等(2024), 我们首先构造 AI 以及大语言模型相关技能词典。词典中的词汇为与 AI 和大语言模型知识或使用相关软件直接相关的词汇, 例如“人工智能”“深度学习”“大数据”等。^③ 之后, 我们从智联招聘平台获得招聘数据, 根据企业招聘广告中的职位描述, 匹配 AI 大语言模型技能词典, 识别出各企业大语言模型相关技能和职位。^④ 最后, 我们通过以下公式计算企业层面的 LLM 暴露度(LLM Share):

$$LLMshare_{i,t} = \frac{\text{Number of job postings requesting AI/LLM skills}_{i,t}}{\text{Total number of job postings}_{i,t}}$$

其中, 分子为企业 i 在时间 t 发布的要求 AI 和大语言模型技能的职位空缺数量, 分母为企业 i 在时间 t 发布的总职位空缺数量。

3. 控制变量

在研究大语言模型如何影响企业员工工作效率的部分, 由于使用的是月度数据, 我们通过加入月份固定效应和公司固定效应来控制加班工作的季节性以及消除观测不到的其他冲击和公司特征的影响。具体变量定义见表 1。

表 1 主要变量说明

变量类型	变量名称	变量符号	变量说明
被解释变量	日度加班率	<i>overtime_ratio</i>	公司员工日度加班比例
	支付给职工以及为职工支付的现金 [#]	<i>compensation</i>	公司实际支付给职工, 以及为职工支付的现金(百万)
	总营业成本率	<i>operate_cost_ratio</i>	营业总成本/营业总收入
	销售费用率	<i>sales_cost_ratio</i>	销售费用/营业收入
	管理费用率	<i>overhead_ratio</i>	管理费用/营业收入
	财务费用率	<i>finance_cost_ratio</i>	财务费用/营业收入

① 某些行业或企业以及某些职业的工会更倾向于在 LinkedIn 上披露信息。

② 具体到 LLM 领域, 北京大学国家发展研究院与智联招聘研究团队(2023) 基于平台大数据, 构建了职业层面的“大语言模型影响指数”, 并于 2024 年更新了测算结果。

③ 具体词典参见附录 III。

④ 这里我们假设: 招聘更多人工智能相关岗位的公司员工, 有更多机会接触到人工智能和大模型相关的知识, 也更有可能在短期内了解并使用大模型。

(续表)

变量类型	变量名称	变量符号	变量说明
被解释变量	销售期间费用率	<i>sales_period_ratio</i>	(销售费用+管理费用+财务费用)/(营业收入)
	女员工日度加班率	<i>over_f_ratio</i>	公司加班女员工占总员工比例
	男员工日度加班率	<i>over_m_ratio</i>	公司加班男员工占总员工比例
	16—22岁员工日度加班率	<i>over_age1_ratio</i>	公司加班16—22岁员工占总员工比例
	23—35岁员工日度加班率	<i>over_age2_ratio</i>	公司加班23—35岁员工占总员工比例
	36—45岁员工日度加班率	<i>over_age3_ratio</i>	公司加班36—45岁员工占总员工比例
	46—60岁员工日度加班率	<i>over_age4_ratio</i>	公司加班46—60岁员工占总员工比例
	收入前10%员工日度加班率	<i>over_incgr1_ratio</i>	公司加班收入前10%员工占总员工比例
	收入10%—25%员工日度加班率	<i>over_incgr2_ratio</i>	公司加班收入10%—25%员工占总员工比例
	其他中低收入员工日度加班率	<i>over_incgr3_ratio</i>	公司加班其他中低收入员工占总员工比例
解释变量	LLM冲击之后	<i>post</i>	日期在“文心一言”向全社会公开之后取值为1,否则为0
	行业受LLM冲击指数	<i>exposure_gpt4</i>	公司所在行业受LLM冲击的指数(GPT4标注)
		<i>exposure_hl</i>	公司所在行业受LLM冲击的指数(人工标注)
	高冲击行业	<i>high_exposure_gpt4</i>	公司所在行业受LLM冲击高于中位数(GPT4标注)
<i>high_exposure_hl</i>		公司所在行业受LLM冲击高于中位数(人工标注)	
控制变量	是否在两周后有季报或年报披露	<i>within_2_weeks</i>	日期在季报或年报披露前两周之内取值为1,否则为0
	公司季度营业总收入	<i>gross_revenue</i>	公司季度营业总收入(百万)

注:“支付给职工的现金,即企业直接发给职工本人的现金,如工资、奖金、津贴、补贴等;为职工支付的现金,即企业以职工为受益人、但支付给第三方机构的现金,如社会保险费、住房公积金、工会经费、职工教育经费等。

本文主要变量的描述性统计如表 2 所示。

表 2 主要变量描述性统计

变量名称	样本数	均值	标准差	最小值	中位数	最大值
Panel A						
<i>overtime_ratio</i>	132 781	0.12	0.11	0.00	0.12	1.00
<i>exposure_gpt4</i>	132 781	0.37	0.13	0.18	0.36	0.64
<i>exposure_hl</i>	132 781	0.31	0.11	0.15	0.33	0.52
<i>post</i>	132 781	0.52	0.50	0.00	1.00	1.00
<i>high_exposure_gpt4</i>	132 781	0.50	0.50	0.00	0.00	1.00
<i>high_exposure_hl</i>	132 781	0.52	0.50	0.00	1.00	1.00
<i>within_2_weeks</i>	132 781	0.05	0.21	0.00	0.00	1.00
<i>over_f_ratio</i>	132 781	0.05	0.07	0.00	0.04	1.00
<i>over_m_ratio</i>	132 781	0.08	0.09	0.00	0.07	1.00
<i>over_age1_ratio</i>	132 781	0.01	0.03	0.00	0.00	1.00
<i>over_age2_ratio</i>	132 781	0.02	0.02	0.00	0.02	0.50
<i>over_age3_ratio</i>	132 781	0.07	0.08	0.00	0.06	1.00
<i>over_age4_ratio</i>	132 781	0.02	0.05	0.00	0.02	1.00
<i>over_incgr1_ratio</i>	132 781	0.01	0.01	0.00	0.00	0.17
<i>over_incgr2_ratio</i>	132 781	0.01	0.02	0.00	0.00	1.00
<i>over_incgr3_ratio</i>	132 781	0.11	0.11	0.00	0.10	1.00
Panel B						
<i>compensation</i> (百万)	17 718	66.65	205.31	0.83	13.92	1 652.30
<i>gross_revenue</i> (百万)	17 732	661.25	2 249.81	2.89	89.31	17 544.78
<i>operate_cost_ratio</i>	16 880	1.81	50.84	0.09	0.94	4 139.87
<i>sales_cost_ratio</i>	16 880	0.09	1.49	0.00	0.03	162.92
<i>overhead_ratio</i>	16 539	0.59	25.47	-0.18	0.06	1 751.44
<i>finance_cost_ratio</i>	16 533	0.03	3.13	-209.10	0.00	135.14
<i>sales_period_ratio</i>	16 533	0.71	27.26	-0.38	0.12	1 886.58
<i>exposure_gpt4</i>	15 837	0.34	0.13	0.17	0.29	0.64
<i>post</i>	17 739	0.35	0.48	0.00	0.00	1.00

注：Panel A 是企业日度层面；Panel B 是企业季度层面（对应表 6）。

（三）模型设定

为了探究大语言模型对员工工作模式的影响，我们首先利用独特的企业-日度员工加班数据，采用广义双重差分模型考察 LLM 冲击对企业员工加班比

例的影响,具体回归模型如下:

$$Overtime_{ijt} = \beta_0 + \beta_1 Post_t \times IndExpo_j + X_{it} + \delta_i + \mu_{m(t)} + \epsilon_{ijt}. \quad (1)$$

被解释变量为从事行业 j 的公司 i 在日期 t 的加班率。关键解释变量为 LLM 冲击 ($Post_t$) 与公司所在行业受冲击程度 ($IndExpo_j$) 的交乘项。 $Post_t$ 是一个哑变量, LLM 冲击发生前的日期为 0, 发生后的日期为 1。对于大语言模型对中国各行业的冲击事件, 我们选择 2023 年 8 月 31 日“文心一言”全平台开放作为主要的外生冲击。 $IndExpo_j$ 是一个连续指标, 反映了公司 i 所在行业 j 受 LLM 冲击的程度大小。 X_{it} 是公司日度的控制变量, 包括公司是否在两周后有季报或年报披露, 以控制因季报、年报工作带来的加班趋势。 $\mu_{m(t)}$ 为月份固定效应, 进一步控制加班工作的季节性。 δ_i 为公司的固定效应, 控制不可观测的公司特征的影响, 比如, 在短期, 员工使用大语言模型更多地属于自发行为, 可能和观测不到的企业文化有关。

四、实证分析

(一) 基准回归结果

表 3 展示了式(1)的回归结果。第(1)列将“文心一言”向全社会公开之后的哑变量 $Post$ 和基于 GPT4 标注的行业冲击度交乘, 可以看到, 受 LLM 冲击高的行业的日度加班率有所下降, 且估计效果显著不同于零。如果 LLM 冲击指数增加 1 个标准差(0.13), 日度加班比则下降 0.002(0.0189×0.13), 相当于平均日加班率的 2.05% ($0.002/0.12$)。第(2)列采用人工标注的行业冲击指数, 估计结果与第(1)列相似。

第(3)—(4)列将“文心一言”向全社会公开之后的哑变量 $Post$ 和高冲击度行业的哑变量交乘。我们将高冲击度行业定义为冲击指数高于中位数的行业。第(3)列的交乘项系数为 -0.005, 且在 0.05 的水平显著不同于 0, 这表明在高冲击行业的公司, 大语言模型工具可以进一步使得加班率降低 0.005, 相当于平均日加班率的 4.2% ($0.005/0.12 = 0.042$)。^①

为确保研究结论的可靠性, 我们进行了多维度稳健性检验, 包括①通过剔除金融业、高冲击行业、ST 公司及极端值样本, 限定办公地址为独立办公楼的子样本; ②拓展至广东省非上市公司抽样数据; ③控制行业-月固定效应以排除行业之间的趋势差别和政策干扰; ④采用加班率的替代指标“日度工作时长超过 8 小时的员工占比”。所有检验结果均与主结论一致, 详见附录 II。

^① 值得一提的是, 2023 年 8 月 31 日, 中国首批通过《生成式人工智能服务管理暂行办法》备案的国产大模型正式向全社会开放。因此, 我们发现的结果是以“文心一言”为首的国产大模型“井喷”的冲击下, 中国国产通用大模型对企业员工工作模式与运营成本的影响。

表 3 大语言模型对员工日加班率的影响

因变量：日度加班率 (<i>overtime_ratio</i>)	(1)	(2)	(3)	(4)
<i>Post</i> × <i>exposure_gpt4</i>	-0.0189** (0.0094)			
<i>Post</i> × <i>exposure_hl</i>		-0.0219** (0.0107)		
<i>Post</i> × <i>high_exposure_gpt4</i>			-0.0054** (0.0024)	
<i>Post</i> × <i>high_exposure_hl</i>				-0.0054** (0.0024)
<i>within_2_weeks</i>	0.0056*** (0.0021)	0.0056*** (0.0021)	0.0056*** (0.0021)	0.0056*** (0.0021)
Constant	0.1269*** (0.0018)	0.1268*** (0.0017)	0.1247*** (0.0006)	0.1248*** (0.0007)
Observations	132 781	132 781	132 781	132 781
R ²	0.081	0.081	0.081	0.081
Firm FE	是	是	是	是
Month FE	是	是	是	是

注：观测单位为企业日度层面。因变量为公司日度加班比率，第(1)列的主要解释变量“*Post* × *exposure_gpt4*”是“文心一言”公开测试之后和行业冲击指数(*gpt4* 标注)的交乘项。类似地，第(2)列的主要解释变量为“文心一言”公开测试之后和行业冲击指数(人工标注)的交乘项。第(3)、(4)列把公司分为是否属于高冲击，即行业冲击指数高于中位数，并把这一哑变量与“文心一言”公开测试哑变量交乘。第(1)–(4)列控制了公司固定效应和月份固定效应。***、**、* 分别表示在 1%、5%、10% 水平下显著，括号内为公司层面的聚类标准误(下同)。

(二) 事前趋势检验与动态效应

本文双重差分法的识别策略的一个重要假设是，行业间企业加班率随时间变化的差异仅由“文心一言”大模型的公开所致，而非源于企业预先存在的趋势差异。为了验证这一假设，我们将模型中的行业冲击程度与后冲击期的交互项，替换为企业是否属于大语言模型高冲击行业的哑变量与模型开放前后各月(除基准月 8 月)的交互项之和。具体来说，在“文心一言”开放日期(2023 年 8 月 31 日)后的 0 到 29 天内，我们定义代表事件后 1 个月的哑变量取值为 1，否则取值为 0；之后 30 到 59 天，事件后 2 个月的哑变量取值为 1，否则取值为 0；之后 60 天及以上，事件后 3 个月的哑变量取值为 1，否则取值为 0。同样地，定

义一系列事件前1,2,3,4月的哑变量。我们用式(2)估计动态效应:

$$Overtime_{ijt} = \beta_0 + \sum_{s=-4, s \neq -1}^3 \beta_s \mathbf{1}\{t - t_0 = s\} \times HighExp_{oj} + X_{it} + \delta_i + \mu_m(t) + \epsilon_{ijt}, \quad (2)$$

其中, t_0 为“文心一言”向全社会公开的时间,即式(1)中 $Post_{it}$ 取值为1的第一天。我们选择冲击前1个月为基期参考时间。所有的系数解读都是与基期进行对比。为了控制加班的季节性变动,我们加入了月份固定效应,同时加入公司固定效应来控制不可观测的公司特征。

图2展示了基于式(2)的估计结果。首先,在大语言模型开放前1到4个月,相对应的事前哑变量的系数 $\beta_s (s = \{-4, -3, -2\})$ 估计与0无显著差异,这表明受到不同冲击程度的公司的日度加班率差异在“文心一言”全面开放前并无显著变动,数据分析结果与平行趋势假设一致。更重要的是,我们发现,在大语言模型开放后30天内($s=0$)日加班率显著下降,这一效果持续到3个月之后至样本观测期最后。由此可见,人工智能和新技术的引入对劳动时间分配有着较为持续的作用。

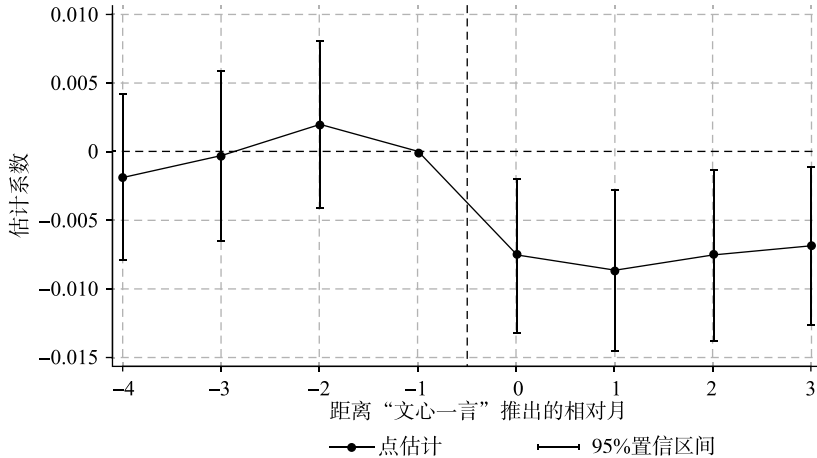


图2 大语言模型对企业员工日加班率影响的动态效应

注:图2考察“文心一言”大模型对全社会开放前后对企业员工加班率的影响。图中的圆点为式(2)中系数的点估计,上下环绕着95%置信区间。

五、进一步研究

(一) 异质性分析

考虑到LLM在不同公司对不同员工的影响可能不同,本部分探究LLM冲击对员工工作模式的异质性效应。

1. 公司层面特征的异质性

(1) 公司类型的异质性。我们依据实际控制人属性和总资产中位数将企业分为国有企业与非国有企业、大企业与小企业进行异质性分析。表4第(1)—(2)列的结果表明,国有企业与非国有企业、大企业与小企业员工加班率改变无显著差异。^①可能的原因在于通用大语言模型使用门槛低、投资成本小且易于操作,能快速应用于多种工作场景,降低了企业规模对技术普及的影响。且短期内大语言模型使用依赖员工自发行为,同行业内不同规模上市公司员工对大语言模型的了解与需求程度差异不大。^②对于国有企业与非国有企业差异,可能因国有企业流程审批文书多而有更多应用场景,但同时存在涉密问题或员工使用新技术动机不足限制应用,最终平均来看国有企业与非国有企业间加班率无显著差异。

(2) 公司治理结构的异质性。基于已有文献^③,我们依据上一年度企业的管理层持股比例和机构持股比例,将样本划分为高、低比例两组进行异质性分析^④。表4第(3)—(4)列的结果显示,LLM冲击降低员工加班率的效应在管理层持股比例高的公司和机构持股比例低的公司中更显著。良好的公司治理结构能通过激励机制、工作环境和员工归属感等方面促进员工积极使用LLM提升效率。而机构投资者持股比例高的公司,可能因信息不对称、决策谨慎和关注短期财务表现等问题,导致大语言模型使用滞后。

(3) 公司员工构成结构的异质性。利用2022年A股上市公司员工构成数据,我们构造了“行政人员占比”指标,以中位数为界划分为高占比与低占比企业。表4第(5)列的结果表明,在行政人员占比较高的公司中,LLM冲击对员工加班率的降低效应更显著。因行政人员工作任务重复性高、流程规范,易被大语言模型替代或辅助优化,能有效减少员工工作负担。

① 这一发现与研究LLM技术之前的AI技术影响的研究结果不同。比如,姚加权等(2024)对企业年报的文本分析发现,人工智能对国有企业和非国有企业的全要素生产率均有提升作用,并且对国有企业的提升作用更显著。此外,大公司可能会从人工智能技术中获益更多(冀云阳等,2023; Babina et al., 2024; Mihet and Philippon, 2019; Farboodi et al., 2019)。比如 Babina et al.(2024)发现大公司更多地投资AI,相应地业绩增长也更快。

② 我们的问卷调查结果也表明,截至2024年9月22日,使用LLM的员工均使用免费版,且没有公司为员工购买付费LLM。问卷调查的具体内容见附录IV。

③ 已有研究表明,在一定范围内,增加管理层持股可以更好地协调股东与管理层的利益,从而有利于公司的业绩表现(Meckling and Jensen, 1976; Hermalin and Weisbach, 1991)。而对于机构投资者,尤其是金融机构,作为公司的“局外人”以及许多客户的代理人,这些情形都增加了委托监督的相关成本(Dasgupta et al., 2021)。

④ 考虑到不管是内部管理层还是外部机构投资者,只有持股比例相对较高的时候他们才有足够的动机干预企业管理,我们将高低两组划分的界限定义为管理层持股比例高于75%分位数以及机构投资者持股比例高于75%分位数。

表 4 大语言模型与员工日加班率:基于公司异质性分析

因变量:日度加班率 (<i>overtime_ratio</i>)	(1)	(2)	(3)	(4)	(5)
<i>Post</i> × <i>exposure</i>	-0.0186*	-0.0297**	-0.0080	-0.0262**	0.00746
	(0.0101)	(0.0136)	(0.0109)	(0.0108)	(0.0185)
<i>Post</i> × <i>exposure</i> × <i>SOE</i>	-0.0010				
	(0.0315)				
<i>Post</i> × <i>SOE</i>	0.0026				
	(0.0114)				
<i>Post</i> × <i>exposure</i> × <i>Large firm</i>		0.0270			
		(0.0186)			
<i>Post</i> × <i>Large firm</i>		-0.0110			
		(0.0077)			
<i>Post</i> × <i>exposure</i> × <i>high_Mngmhldn</i>			-0.0360*		
			(0.0216)		
<i>Post</i> × <i>high_Mngmhldn</i>			0.0133		
			(0.0091)		
<i>Post</i> × <i>exposure</i> × <i>high_InsInvestorProp</i>				0.0506**	
				(0.0203)	
<i>Post</i> × <i>high_InsInvestorProp</i>				-0.0208***	
				(0.0079)	
<i>Post</i> × <i>exposure</i> × <i>high_admin share</i>					-0.0469**
					(0.0237)
<i>Post</i> × <i>high_admin share</i>					0.0236**
					(0.0097)
<i>within_2_weeks</i>	0.0057***	0.0059***	0.0054***	0.0060***	0.0051**
	(0.0021)	(0.0021)	(0.0020)	(0.0021)	(0.0022)
Constant	0.1270***	0.1290***	0.1250***	0.1290***	0.120***
	(0.0019)	(0.0029)	(0.0022)	(0.0022)	(0.0041)
<i>Observations</i>	127 048	131 569	129 724	130 889	117 241
R ²	0.081	0.080	0.081	0.081	0.081
Firm FE	是	是	是	是	是
Month FE	是	是	是	是	是

注:在表 3 基础上,第(1)列“*Post* × *exposure* × *SOE*”为“*Post* × *exposure*”与国有企业哑变量的交乘项,第(2)–(5)列分别引入“*Post* × *exposure*”与是否为大公司(公司规模高于 50%分位数)、管理层持股比例高(比例高于 75%分位数)、机构投资者持股比例高(比例高于 75%分位数)、是否在 2022 年为高行政人员占比(比例高于 50%分位数)的交乘项。第(1)–(5)列控制了公司固定效应和月份固定效应。

2. 员工人口特征的异质性

接下来,我们将式(1)中的因变量根据员工的人口特征进行了进一步的细分,考察不同类型的员工受到 LLM 的影响程度。

表 5 中第(1)列和第(2)列的结果表明,LLM 冲击对于男女员工的影响较为接近。这一结果可能是由两个相反的因素共同导致的。一方面,LLM 目前在材料搜索、文本撰写等重复性较强工作上的应用更为广泛,而女性员工更可能从事此类工作(如行政秘书、销售等),使她们的工作任务更易被 LLM 辅助完成。另一方面,虽然女性的工作任务对 LLM 暴露度可能更高,但相比男性,女性对新技术的使用通常是滞后的。^①

第(3)—(6)列考察不同年龄员工受到的影响。从年龄层来看,23—35 岁以及 36—45 岁的员工加班率在 LLM 影响下显著降低,而其他年龄段员工的加班率变化则不显著。这一结果可能是由于青壮年对新技术接受度更高,从而更愿意将 LLM 和自己日常工作任务相结合,提高工作效率。

第(7)—(9)列考察不同收入水平员工受到的影响,分别为收入前 10% 员工、收入水平排在 10%—25% 的员工,以及其他收入水平员工。第(9)列结果表明,收入水平低于前 25% 的员工加班率显著下降了 0.019 个百分点,其他高收入员工则没有明显的下降。一个可能的解释是,由于 LLM 目前在工作任务上的应用多在材料搜索汇总、文本撰写和一些重复性较强的工作,与之相对应的是行政秘书、销售等职业,而女性、中低收入群体相对来说更可能接触到此类工作任务,相应地他们的工作内容更可能让人工智能来辅助完成。

表 5 大语言模型与员工日加班率:基于员工人口特征异质性分析

因变量:各类 员工日度 加班率	女员工	男员工	16—22 岁 员工	23—35 岁 员工	36—45 岁 员工	46—60 岁 员工	收入前 10% 员工	收入 10%—25% 员工	其他中 低收入 员工
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Post</i> × <i>exposure</i>	-0.009* (0.005)	-0.010 (0.006)	-0.002 (0.002)	-0.005** (0.002)	-0.011* (0.007)	0.001 (0.004)	-0.001 (0.001)	0.001 (0.002)	-0.019** (0.009)
<i>within_2_weeks</i>	0.002 (0.001)	0.003** (0.002)	0.001** (0.000)	0.002*** (0.000)	0.001 (0.002)	0.002* (0.001)	0.000* (0.000)	0.001* (0.000)	0.005** (0.002)
Constant	0.049*** (0.001)	0.078*** (0.001)	0.006*** (0.000)	0.022*** (0.000)	0.068*** (0.001)	0.024*** (0.001)	0.006*** (0.000)	0.010*** (0.000)	0.111*** (0.002)
Observations	132 781	132 781	132 781	132 781	132 781	132 781	132 781	132 781	132 781

^① Humlum and Vestergaard (2024)对丹麦的员工大规模的调查问卷结果显示,对大模型的使用倾向存在显著的性别差异。尽管大模型被认为能够提高工作效率,但与男性相比,女性使用大模型的可能性较小。我们的问卷调查结果也显示,被调查的员工中,男性占 34.43%,女性占 65.57%。但在使用大模型的员工中,男员工占 41.95%。

(续表)

因变量:各类 员工日度 加班率	女员工	男员工	16—22岁 员工	23—35岁 员工	36—45岁 员工	46—60岁 员工	收入前 10%员工	收入 10%—25% 员工	其他中 低收入 员工
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
R ²	0.051	0.048	0.037	0.206	0.048	0.048	0.212	0.142	0.075
Firm FE	是	是	是	是	是	是	是	是	是
Month FE	是	是	是	是	是	是	是	是	是

注:表5采用表3的回归模型,因变量替换为公司不同类型员工的日度加班比率,其中第(1)、(2)列为公司加班女员工人数与公司员工总人数的比值和公司加班男员工人数与公司员工总人数的比值;第(3)—(6)列为加班员工中不同年龄段员工人数与公司员工总人数的比值;第(7)—(9)列为加班员工中不同收入层次员工人数与公司员工总人数的比值。

(二) LLM 冲击对公司运营成本的影响

为进一步探讨人工智能大语言模型的引入对企业劳动生产率和运营成本的潜在影响,我们开展了问卷调查,以提供动机性证据。具体的问卷调查背景、员工对大语言模型的使用现状及其对工作效率的影响参见附录IV。

基于调查问卷结果,我们进一步对大语言模型如何对员工生产率和公司运营成本产生影响进行实证分析。我们使用全国A股上市公司在2023年全年的季度报表数据,将其中关于员工收入、公司各项成本率指标作为因变量,并使用式(1)进行双重差分回归分析。表6展示了回归结果。第(1)列考察的是LLM的引入对职工劳动报酬的影响。我们控制了营业总收入,结果显示在营业收入相同的公司中,被LLM冲击程度越高的公司,支付给员工的劳动报酬增加。每增加10%的冲击比例,员工劳动报酬支出比冲击前增加360万,相当于平均支出水平的5.4%($0.1 \times 36.15 / 66.65 = 5.4\%$)。由于数据为一年内2—3个季度的数据,员工数量变动较为有限^①,员工劳动报酬支出的增加在一定程度上可以反映劳动生产率的提高。

在企业层面,LLM可以通过提高员工工作效率进一步降低运营成本。^②参考相关文献(温素彬等,2022),我们利用总营业成本率和销售费用率作为生产成本和销售成本的代理变量,以度量生产运营流程优化。表6第(2)—(6)列具体考察了大语言模型冲击对企业运营成本的潜在影响。回归结果显示,

^① 作者计算了全国上市公司在2017—2019年及2020—2022年的年度员工变动比率,75%公司的年度员工变动率在22%以下,而50%的公司年度员工数量变动在11%以下。

^② 以运营部门为例,大量重复性任务(如数据录入、报告生成和文档整理)可通过大语言模型实现自动化处理,从而减少员工的工作量。而对于销售岗位,通过自动化和优化销售流程,大模型可以帮助减少销售团队在客户获取和维护上的时间和资源投入,从而降低销售成本。综合来看,LLM通过优化运营与销售流程,可以提升员工效率,最终降低企业运营成本。

LLM冲击与总营业成本率和销售费用率的下降显著相关。当企业受到的LLM冲击比例增加10%时,总营业成本率平均降低0.007,相当于样本均值的约0.4%($0.007/1.81=0.004$);销售费用率平均降低0.002,相当于样本均值的约2%($0.002/0.09=0.022$)。此外,我们还考察了大语言模型对企业管理费用率、财务费用率和销售期间费用率的影响,尽管相关估计系数均为负值,但其与零的差异未达到统计显著性水平,表明这些成本指标的降低幅度较小。^①

表6 “文心一言”开放与上市公司劳动收入和成本

	支付给职工以及为职工支付的现金(百万)	总营业成本率	销售费用率	管理费用率	财务费用率	销售期间费用率
因变量:	<i>Compensation</i>	<i>Operate_cost_ratio</i>	<i>Sales_cost_ratio</i>	<i>Overhead_ratio</i>	<i>Finance_cost_ratio</i>	<i>Sales_period_ratio</i>
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Post</i> × <i>exposure</i>	36.15*** (7.188)	-0.074*** (0.020)	-0.022*** (0.004)	-0.370 (0.350)	-0.061 (0.066)	-0.455 (0.416)
Constant	26.39*** (2.146)	0.971*** (0.003)	0.065*** (0.001)	0.481*** (0.043)	0.043*** (0.008)	0.590*** (0.051)
Observations	15 754	15 033	15 033	14 697	14 697	14 697
R ²	0.971	0.893	0.958	0.999	0.988	0.998
Firm FE	是	是	是	是	是	是
Month FE	是	是	是	是	是	是

注:本表样本为2023年全国上市公司季度数据,因变量为公司季度支付给员工的现金(百万),成本费用率(百分比)等测度。主要解释变量“*Post* × *exposure*”是“文心一言”公开测试之后和行业冲击指数(*gpt4*标注)的交互项。第(1)列控制了公司季度营业总收入。第(1)—(6)列控制了公司固定效应和季报月度固定效应。

六、政策含义、研究局限和未来拓展

本文有以下政策含义。首先,本文发现大语言模型的低成本、通用性和易操作性在短期内就会显著影响企业员工的工作效率。因此,企业在数字化转型过程中,应充分运用以大语言模型为代表的新一代人工智能技术,探索如何协同

^① 然而,需注意的是,由于冲击后观测时间段较短且成本和费用率数据频率相对较低,变量间的内在关系可能受限。同时,受限于数据颗粒度,模型中无法纳入行业-月份固定效应,因此某些行业特定因素可能未被完全捕捉。因此,我们的结果应被视为相关性的初步证据,而非因果关系的直接证明。不过,这些发现与已有文献(温素彬等,2022)一致,支持了我们的假设2,即大语言模型有助于降低企业的运营成本,尤其是在运营和销售相关环节的成本。未来研究应依赖更精细的成本数据及更丰富的控制变量,以进一步验证大语言模型对企业运营成本的影响。

人工智能与劳动力要素,降低企业的运营成本和数字化转型成本,提升企业生产力。其次,在企业层面,及时更新管理层和股东对新技术的认知以及协调不同利益相关者的利益异质性至关重要。建议在政策层面对企业进行相关领域和技术的培训与推进数字化人才队伍的建设,帮助企业及时抓住机遇,应用新的人工智能技术进行高质量发展。

由于数据的可得性,本文的研究存在以下两点局限。第一,由于数据限制,目前的研究没有进一步考察企业内部具体经营管理流程上的差异。未来我们计划使用其他企业文本数据以及更丰富的企业调查数据来做进一步拓展。第二,我们目前只考察了大语言模型对企业中短期员工加班率的影响,而2025年DeepSeek等新一代模型的推出,技术应用从员工自发使用向企业主动部署转变。未来我们可以进一步拓展,考察大语言模型技术较为长期的影响,并利用更精细的企业数据进一步验证大语言模型对企业经营决策与运营成本等多方面的影响。

参 考 文 献

- [1] Acemoglu, D., and P. Restrepo, "The Race Between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment", *American Economic Review*, 2018, 108(6), 1488-1542.
- [2] Alekseeva, L., J. Azar, M. Giné, S. Samila, and B. Taska, "The Demand for AI Skills in the Labor Market", *Labour Economics*, 2021, 71, 102002.
- [3] Atkin, D., M. K. Chen, and A. Popov, "The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley", National Bureau of Economic Research, 2022.
- [4] Babina, T., A. Fedyk, A. He, and J. Hodson, "Artificial Intelligence, Firm Growth, and Product Innovation", *Journal of Financial Economics*, 2024, 151, 103745.
- [5] Barwick, P. J., Y. Liu, E. Patacchini, et al., "Information, Mobile Communication, and Referral Effects", *American Economic Review*, 2023, 113(5), 1170-1207.
- [6] 北京大学国家发展研究院、智联招聘, "ChatGPT如何影响我们的工作? ——AI大模型对我国劳动力市场潜在影响研究", (2023—07), <https://hrssit.cn/Uploads/file/20241011/1728635352230865.pdf>.
- [7] Berg, J. M., M. Raj, and R. Seamans, "Capturing Value from Artificial Intelligence", *Academy of Management Discoveries*, 2023, 9(4), 424-428.
- [8] Bertomeu, J., Y. Lin, Y. Liu, et al., "The Impact of Generative AI on Information Processing: Evidence from the Ban of ChatGPT in Italy", *Journal of Accounting and Economics*, 2025, 101782.
- [9] Brynjolfsson, E., D. Li, and L. Raymond, "Generative AI at Work", *The Quarterly Journal of Economics*, 2025, qjae044.
- [10] Brynjolfsson, E., D. Rock, and C. Syverson, "Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics", NBER Working Paper, 2017 (w24001).

- [11] Capraro, V., A. Lentsch, D. Acemoglu, et al., “The Impact of Generative Artificial Intelligence on Socioeconomic Inequalities and Policy Making”, *PNAS nexus*, 2024, 3(6), pgae191.
- [12] 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, 第 74—93 页。
- [13] Chen, M. K., and D. G. Pope, “Geographic Mobility in America: Evidence from Cell Phone Data”, National Bureau of Economic Research, 2020.
- [14] 陈彦斌、林晨、陈小亮, “人工智能、老龄化与经济增长”, 《经济研究》, 2019 年第 7 期, 第 47—63 页。
- [15] Choi, J. H., and D. Schwarcz, “AI Assistance in Legal Analysis: An Empirical Study”, Available at SSRN 4539836, 2023.
- [16] Dasgupta, A., V. Fos, and Z. Sautner, “Institutional Investors and Corporate Governance”, *Foundations and Trends © in Finance*, 2021, 12(4), 276-394.
- [17] Dell’Acqua, F., E. McFowland, E. R. Mollick, H. Lifshitz-Assaf, K. Kellogg, S. Rajendran, ... and K. R. Lakhani, “Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality”, Harvard Business School Technology & Operations Mgt. Unit Working Paper, 2023 (24-013).
- [18] Doshi, A. R., J. J. Bell, E. Mirzayev, et al., “Generative Artificial Intelligence and Evaluating Strategic Decisions”, *Strategic Management Journal*, 2025, 46(3), 583-610.
- [19] Eisfeldt, A. L., and G. Schubert, “AI and Finance”, National Bureau of Economic Research, 2024.
- [20] Eisfeldt, A. L., G. Schubert, M. B. Zhang, “Generative AI and Firm Values”, National Bureau of Economic Research, 2023.
- [21] Eloundou, T., S. Manning, P. Mishkin, et al., “GPTs Are GPTs: Labor Market Impact Potential of LLMs”, *Science*, 2024, 384(6702), 1306-1308.
- [22] Goldfarb, A., B. Taska, and F. Teodoridis, “Could Machine Learning Be a General Purpose Technology? A Comparison of Emerging Technologies Using Data from Online Job Postings”, *Research Policy*, 2023, 52(1), 104653.
- [23] 郭凯明、王钰冰、龚六堂, “劳动供给转变、有为政府作用与人工智能时代开启”, 《管理世界》, 2023 年第 6 期, 第 1—21 页。
- [24] Han, M., H. Shen, J. Wu, et al., “Artificial Intelligence and Firm Resilience: Empirical Evidence from Natural Disaster Shocks”, Information Systems Research, 2025.
- [25] Hermalin, B. E., and M. S. Weisbach, “The Effects of Board Composition and Direct Incentives on Firm Performance”, *Financial Management*, 1991, 101-112.
- [26] Humlum, A., and E. Vestergaard, “The Adoption of ChatGPT”, University of Chicago, Becker Friedman Institute for Economics Working Paper, (2024-50).
- [27] Jensen, M. C., and W. H. Meckling, “Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure”, *Journal of Financial Economics*, 1976, 3(4), 305-360.
- [28] 冀云阳、周鑫、张谦, “数字化转型与企业创新——基于研发投入和研发效率视角的分析”, 《金融研究》, 2023 年第 4 期, 第 111—129 页。
- [29] Jia, N., N. Li, G. Ma, et al., “Corporate Responses to Generative AI: Early Evidence from Conference Calls”, *Review of Accounting Studies*, 31, 649-703 (2026). <https://doi.org/10.1007/s11142-025-09916-1>.
- [30] Jia, N., X. Luo, Z. Fang, and C. Liao, “When and How Artificial Intelligence Augments Employee

- Creativity”, *Academy of Management Journal*, 2024, 67(1), 5-32.
- [31] 金星晔、左从江、方明月、李涛、聂辉华,“企业数字化转型的测度难题:基于大语言模型的新方法与新发现”,《经济研究》,2024年第3期,第34—53页。
- [32] 孔高文、刘莎莎、孔东民,“机器人与就业——基于行业与地区异质性的探索性分析”,《中国工业经济》,2020年第8期,第80—98页。
- [33] 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,第104—119页。
- [34] Li, N., H. Zhou, W. Deng, J. Liu, F. Liu, and K. Mikel-Hong, “When Advanced AI Isn’t Enough: Human Factors as Drivers of Success in Generative AI-Human Collaborations”, Available at SSRN 4738829, 2024.
- [35] 林晨、陈小亮、陈伟泽、陈彦斌,“人工智能、经济增长与居民消费改善:资本结构优化的视角”,《中国工业经济》,2020年第2期,第61—83页。
- [36] Michaely, R., and J. Grennan, “Artificial Intelligence and the Future of Work: Evidence from Analysts”, working paper, 2021.
- [37] Mokyr, J., C. Vickers, and N. L. Ziebarth, “The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different?”, *Journal of Economic Perspectives*, 2015, 29(3), 31-50.
- [38] Noy, S., and W. Zhang, “Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence”, *Science*, 2023, 381(6654), 187-192.
- [39] 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,第159—175页。
- [40] 温素彬、张金泉、焦然,“智能制造、市场化程度与企业运营效率——基于A股制造业上市公司年报的文本分析”,《会计研究》,2022年第11期,第102—117页。
- [41] 吴非、胡慧芷、林慧妍、任晓怡,“企业数字化转型与资本市场表现——来自股票流动性的经验证据”,《管理世界》,2021年第7期,第130—144+10页。
- [42] 肖土盛、孙瑞琦、袁淳、孙健,“企业数字化转型、人力资本结构调整与劳动收入份额”,《管理世界》,2022年第12期,第220—237页。
- [43] 杨继生、阳建辉,“行政垄断、政治庇佑与国有企业的超额成本”,《经济研究》,2015年第4期,第50—61+106页。
- [44] 姚加权、张锐澎、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,第101—116+133+117—122页。
- [45] 叶永卫、李鑫、刘贯春,“数字化转型与企业人力资本升级”,《金融研究》,2022年第12期,第74—92页。

The Effect of Large Language Models on Employee Work Efficiency: An Empirical Study Based on Mobile Data

LU Di

(Shenzhen University)

WU Qi*

(Peking University)

ZHANG Yingfeng

(Unicom (Guangdong) Industrial Internet Co., Ltd.)

Abstract: We examine the impact of the “Wenxin Yiyan (Ernie Bot)” Large Language Model (LLM) on corporate operating costs via changes in employee work patterns. By constructing a firm-level panel dataset of employee overtime rates from mobile data and applying the Difference-in-Differences (DID) methodology, we find that employees in firms highly exposed to LLMs experience significant reductions in overtime rates, along with increases in labor compensation and decreases in corporate operating costs. Heterogeneity analysis indicates that these effects vary across different shareholding and employee structures, with more pronounced impacts on young and low-to-middle-income employees. Overall, this study offers empirical evidence and policy insights for leveraging artificial intelligence to promote high-quality development of enterprises.

Keywords: Large Language Model; corporate operating costs; mobile big data

JEL Classification: D24, G14, M12

* Corresponding Author: WU Qi, Guanghua School of Management, Peking University, No. 5 Yiheyuan Road, Haidian District, Beijing 100871, China; Tel: 86-10-62747554; E-mail: qiwu@gsm.pku.edu.cn.