

# 生成式 AI 驱动创新的数量和质量效应

——基于诺贝尔化学奖成果 AlphaFold 的发现

肖柏高 周阳 刘青\*

**摘要:** 本文以 2024 年诺贝尔化学奖成果 AlphaFold 为外生冲击, 研究生成式 AI 对创新的影响。研究发现, 能精准生成蛋白质空间结构的 AlphaFold 显著提升了蛋白质创新数量。但对创新质量影响出现分化: 其“创新生成”提升以知识组合、交叉为主的利用式创新质量, 同时也因“守旧训练”导致创新趋同, 未能提升以颠覆、引领为主的探索式创新质量。此外, AlphaFold 驱动科研体系转向“AI 范式”, 虽提升人均产出, 但带来算力数据依赖加深、企业短期研发效率下降和科研方向的 AI 路径收敛等挑战。

**关键词:** 生成式 AI; 科技创新; 数量和质量效应

**DOI:** 10.13821/j.cnki.ceq.2026.03.05

## 一、引言

“AI for Science”(AI 驱动科学)是 AI 的核心应用和研究领域,生成式 AI 的出现极大地增强了 AI 对科技创新的驱动能力。2024 年,因 AI 模型 AlphaFold 极大推动了生物化学领域研究,其两位主要开发者因此获得诺贝尔化学奖;2023 年 ChatGPT 成为《自然》(Nature)杂志年度十大科学人物的首位非人类学者,均标志着生成式 AI 驱动的科技创新范式正快速形成。当前,生成式 AI 已经快速融入科技创新之中并发挥重要作用。一方面,越来越多的研究者开始使用生成式 AI。诺贝尔化学奖颁奖词指出,AlphaFold 系列模型已被来自 190 个国家的 200 多万人使用。一项全球范围的调查也显示,已有约一半的受访者使用生成式 AI 辅助科研(Naddaf, 2025)。另一方面,生成式 AI 渗透到新药设计、材料发现和理论验证等各个学科和科研环节中。

\* 肖柏高、周阳,湖南大学经济与贸易学院;刘青,中国人民大学经济学院。通信作者及地址:刘青,北京市海淀区中国人民大学明德楼主楼 738,100872;电话:15810897360;E-mail: qingliu@ruc.edu.cn。作者感谢国家自然科学基金重大项目(23&ZD041)、教育部基地重大项目(20231JY0103)、教育部人文社会科学研究青年基金项目“智能算力封锁与 AI 行业发展:机制分析、效应评估和转型路径研究”(24YJC790195)、湖南省自然科学基金项目(2025JJ60456)的资助。感谢审稿专家的建设性意见。

生成式 AI 通过训练海量数据获得知识库与推理等高阶智能,并基于此生成有价值的新观点、新思路和新发现(即“创新生成”),有望将科技创新过程由传统的“人类思考再实验发现”变革为“AI 思考和直接发现”或“AI 思考发现后再实验验证”,极大提升创新来源、速度和效率。然而,其生成功能的训练是让生成内容尽可能地逼近已有数据中的规律和模式(即“守旧训练”),因此其输出内容的“新”可能只是超大训练数据中子模式的重组(Goodfellow et al., 2014; Brown et al., 2020)。这一特征尽管有发现遗漏或交叉创新的潜力,但其推动输出内容向已有数据规律或模式靠拢,训练方向与追求突破人类认知边界的原创性、颠覆性、引领性等创新的方向相悖。综上,理论上生成式 AI 对创新质量有复杂影响,本文旨在通过系统性实证分析,探究其对创新数量与质量的实际影响。

AlphaFold 的出现为本研究主题提供了一个极佳的实证环境:第一,AlphaFold 是生成式 AI 先驱之一,具备生成式 AI 基于“守旧训练”实现“创新生成”的核心功能与特征,故可借助它来研究生成式 AI 如何驱动科技创新。第二,AlphaFold 专用于蛋白质科技领域,影响范围清晰,非常有利于生成式 AI 与科技创新之间的因果推断。第三,AlphaFold 是 AI 技术在蛋白质结构领域的跨界应用成果,它的出现主要得益于 AI 领域的技术突破,而非受到医药等行业的逆向影响,而且 AlphaFold 是国外成果,与国内的关联更小。因此,当以国内企业为样本研究 AlphaFold 与蛋白质创新之间的因果关系时,可以极大减轻内生性问题。

本文以受 AlphaFold 影响更大的医药和化学行业为处理组,利用双重差分模型进行实证检验。上市企业的样本表明 AlphaFold 显著提高蛋白质相关的专利数量,这一结论通过了模型、变量、样本等方面的稳健性检验和事前平行趋势假设检验,说明 AlphaFold 提升了创新数量。然而,AlphaFold 对创新质量的影响则呈现出结构性分化。一方面,AlphaFold 显著促进侧重知识重组交叉的利用式创新质量,表现为跨学科合作与技术组合广度的提升;但另一方面,它并未带来探索式创新的实质突破,蛋白质专利相似度上升,且颠覆性与引领性指标未获提升。这表明 AlphaFold 虽然极大地优化了知识组合效率,但在推动范式突破与原始创新上仍面临局限。

上述现象折射出科研体系正经历从“人驱动”向“AI 驱动”的范式转型。本文进一步探讨了这一转型的特征与挑战:企业研发重心向 AI 赋能的领域集聚,人才需求结构也同步向 AI 与蛋白质研发岗位倾斜。这种新范式虽然通过降低门槛等因素提升了个人层面的研发生产力,但也带来了三重挑战:一是科研活动对算力与数据的依赖度显著加深,可能导致新的资源壁垒;二是由于转型成本等因素,企业整体研发效率在短期内面临不升反降的阵痛;三是可能导致研

发持续围绕少数 AI 高度适配的研究领域展开,抑制对当前尚无法被 AI 有效处理的研究方向的投入。

随着“AI for Science”加速成为科研新范式,生成式 AI 已从辅助工具演变为驱动创新的核心要素。然而,现有研究多聚焦于其带来的效率红利,却忽视了其底层技术逻辑可能引发的质量隐忧,亟需从学理层面辩证审视其对科技创新的双重效应。本文围绕生成式 AI 驱动创新过程中数量扩张与质量分化并存现象,系统刻画其内在机制与现实表现,揭示了潜在的同质化与路径依赖风险。基于此,本文贡献主要体现在三个方面:第一,在理论视角上,构建了“守旧训练”与“创新生成”共存的分析框架,为辨析生成式 AI 如何驱动创新提供新视角。第二,在研究内容上,突破了单一的数量视角,不仅揭示了 AI 对利用式创新与探索式创新的异质性影响,还捕捉到研发范式转型中效率的“J 曲线”调整特征及对算力设施的依赖,描绘了更立体的变革图景。第三,在因果识别上,利用 AlphaFold 这一垂直型 AI 构建了准自然实验,相比依赖问卷或泛化指标的现有研究,该设计有效规避内生性干扰,为评估 AI 驱动创新提供更严谨的因果证据。

## 二、文献综述与理论分析

“AI for Science”的实践和话题是近十年随着 AI 技术不断突破才逐渐兴起的,因此相关研究较少也较新。“AI for Science”的研究可以划分为以下三支脉络:第一支文献从科技范式变革的角度探讨基于 AI 的科研范式的特点和趋势,他们普遍认为“AI 驱动”或者“AI 范式”将成为“经验驱动”、“理论驱动”、“计算驱动”、“数据驱动”后的第五代科研范式(李建会和杨宁,2023;李伦和刘梦迪,2024;政光景和吕鹏,2023)。第二支文献研究 AI 在科学研究和科技创新中的具体应用方式,研究发现,AI 在数据处理、假设验证、因果推断等环节中展现出显著的辅助与扩展功能(Merchant et al., 2023; Van Noorden and Perkel, 2023;洪永森和汪寿阳,2023;刘景江等,2023;肖柏高等,2025)。第三支文献构建 AI 影响科技创新的理论机制,部分研究进行了实证检验,普遍发现 AI 会提升科技创新数量(Bianchini et al., 2022; Cockburn et al., 2018; Rammer et al., 2022; 胡蓉宁和王跃堂,2023;李玉花等,2024),这支文献与本课题最相关。

尽管上述文献为理解 AI 的创新效应提供了重要基础,但多聚焦于传统 AI,尚未深入剖析生成式 AI 的独特影响机理。与传统 AI 不同,生成式 AI 实现从“判别”到“生成”的变革,其“创新生成”与“守旧训练”并存的矛盾特性,预示其对创新产出的影响更为复杂。因此,本部分将结合生成式 AI 的技术特征,分别从创新数量的扩张机制和创新质量的结构变化两个维度展开理论分析。

### (一) 生成式 AI 影响创新数量的理论分析

生成式 AI 相较传统 AI 驱动创新的能力显著提升,主要来自其“创新生成”能力。科技创新最核心的环节是“创新”,传统 AI 一般是“判别式 AI”,即主要对已有结果和数据进行判别,输出“是否”之类的标签或简单数值,如图像识别、回归等,这决定“判别式 AI”在主要承担事后验证的辅助工作。而生成式 AI 基于其庞大知识库和高阶智能,可以生成新想法、新问题乃至新发现,有望将科学发现过程从传统的“实验获得发现”变革为“AI 发现再验证”,从而大幅拓宽创新构思的源头,直接增加潜在创新起点的数量。例如解析蛋白质三维结构是深入了解其功能的前提,而蛋白质结构测量要求冷冻电镜等昂贵设备和大量人力,是一个典型且艰难的“实验获得发现”过程。而 AlphaFold 可以直接“生成”高精度的蛋白质结构预测,这一过程本身就是一项由 AI 完成的“新发现”。这使得科学家不再受限于漫长的蛋白质结构探索,而是可以直接利用 AI 提供的海量结构预测作为创新起点,进行下游的药物设计、功能验证和机理研究,从而极大地拓宽创新源头、增加潜在创新数量。因此中国科学院院士、结构生物学家施一公称赞 AlphaFold 是 AI 对科学领域最大的一次贡献。<sup>①</sup>

生成式 AI 的超大知识库等特征也有助于将“知识负担”转为“巨人肩膀”。随着人类知识和成果不断累积,原本作为“巨人肩膀”的已有知识日益庞大,以至于开始成为研究人员的负担(Park et al., 2023)。传统 AI 由于训练样本和模型特征所限,能吸收的知识较少,而生成式 AI 利用超大算力从超大数据中进行学习,将庞杂分散的知识库转化为能够主动生成新洞见的动态引擎。AlphaFold 对人类积累的数十万蛋白质结构进行学习后掌握从氨基酸序列到三维结构的复杂规律,将过去仅限于结构生物学家的专业知识,甚至是结构生物学家都难以掌握的庞大分散知识,通过按要求生成相应结构的方式迅速赋能给医学、材料科学等众多领域的专家,从而催生大量的潜在创新点。基于上述分析,本文提出如下假设:

**假设 1** 生成式 AI 显著促进创新数量的增长。

### (二) 生成式 AI 影响创新质量的理论分析

相较于对创新数量较为明晰的促进作用,生成式 AI 对创新质量的影响则更为复杂。本文引入 March(1991)关于“利用式创新”(exploitation)与“探索式创新”(exploration)的经典组织学习理论框架来进行分析。利用式创新侧重于在已有知识边界内、现有技术轨迹上进行精细化利用与跨模块重组,其质量体现为知识、技术融合的广度与效率。探索式创新则侧重于对知识边界外部未知

<sup>①</sup> 参见 <https://www.mittrchina.com/news/detail/9754>, 访问时间:2025年10月5日。

领域的搜寻、试验与范式突破,其质量体现为原创性突破与路径开辟。结合生成式 AI 的技术机理,本文认为其对这两类创新质量具有截然不同的影响。

首先,生成式 AI 可以促进基于现有知识重组的利用式和组合式创新,其核心机制在于降低跨学科知识融合的壁垒。不同知识模块的重新连接与组合是创新的重要来源之一(Fleming, 2001; Strumsky and Lobo, 2015; Verhoeven et al., 2016)。AlphaFold 利用深度学习将结构生物学领域积累的海量蛋白质折叠规律封装为一种低门槛的通用工具,从而极大地降低了结构生物学知识的获取与应用成本,使得原本局限于该领域的专业知识被轻松迁移并重组到药物化学、合成生物学、材料科学等其他学科的知识体系中。换言之,AlphaFold 充当不同学科间的连接器,更推动蛋白质知识与医学、化学等异质性知识的深度交叉与融合。因此,在利用现有知识存量进行组合式创新的维度上,生成式 AI 可带来显著的质量提升。

其次,生成式 AI 可能无助于追求颠覆性、突破性的探索式创新,其中“守旧训练”模式构成关键制约。诚然,学习已有知识、规律和经验是人类与 AI 开展知识生产的共同基础,是生成式 AI 充分高效利用已有知识实现“创新生成”、推动利用式创新的关键。然而,两者在生成新知识环节的底层逻辑存在本质差异:人类科学家学习是为了获得基础知识,并锻炼因果推断、逻辑演绎甚至直觉等思维能力,从而突破知识边界(即“跳出分布”);而当前生成式 AI 的训练目标是最大程度地逼近训练数据中隐含的统计规律和概率分布(即“拟合分布”)。例如 AlphaFold 的训练目标是逼近已有蛋白质空间结构数据。这种训练目标的设定决定了 AlphaFold 等生成式 AI 的输出内容是拟合已知,而非探索未知,其生成内容的“新”更多的是对训练数据中规律和模式的概率性排列组合,而非科学规律的发现或灵感迸发。

这种“守旧训练”驱动的输出特性,与原创性和颠覆性等高质量科技创新存在着根本性矛盾。原创性创新要求打破现有思维框架,提出新的问题、概念或解决方案,颠覆性创新往往源于对现有范式、基础假设或核心原理的挑战与突破,这需要敢于突破常规、基于深刻洞见而非数据拟合的能力。然而,“守旧训练”受限于其拟合已知数据的本质,其“创造力”被严格框定在已有数据所定义的空间之内。它擅长在已知路径上优化或组合,却难以主动开辟全新的、数据中未曾体现的探索方向。此外,生成式 AI 的“守旧”特性使其天然倾向于维护和强化训练数据中体现的主流观点和共识(因为它们在数据中占比最高、统计权重最大),也不利于识别和挑战颠覆性问题。所以如果过度依赖生成式 AI 的发现或观点,大量“创新”实质上是对已有模式的变体重组,整个领域的创新生态将陷入同质化,难以涌现出超越现有范式的颠覆性成果。基于上述分析,本文提出如下假设:

**假设 2a** 生成式 AI 可提升知识组合、融合和交叉能力,从而有助于利用式创新质量的提升。

**假设 2b** 生成式 AI 的“守旧训练”与颠覆性和引领性等高质量创新相悖,甚至可能引致创新趋同,可能无助于提升探索式创新质量。

需明确的是,虽然 AlphaFold 是专注于特定科学领域的垂直型生成式 AI,不具备 ChatGPT 等通用型生成式 AI 的通用能力,但两者在底层架构(均基于 Transformer 变体)和生成逻辑(拟合蛋白质或语言数据分布)上具有高度同源性(Jumper et al., 2021)。因此,本文以 AlphaFold 为对象的研究虽捕捉的是垂直型生成式 AI 的效应,但对理解广义生成式 AI 的创新机制仍具重要启示。当然,鉴于“通用”等特征是 AlphaFold 等垂直型生成式 AI 所不具备的,本文结论不应被无条件外推至所有通用型生成式 AI。

### 三、基准模型和结果

#### (一) 基准模型

在实证研究中,如何准确衡量 AI 并确立其与创新的因果关系是核心挑战。现有文献采用多元化的衡量与识别方法:第一是利用企业创新调查数据,直接获取公司的 AI 应用水平以及其如何驱动新产品研发(Rammer et al., 2022)。第二是采用文本分析法,通过构建 AI 词典并分析上市公司年报、专利文本或新闻报道,量化企业在 AI 方面的应用水平并研究 AI 应用对专利的影响(李雨花等,2024)。第三是从人力资本投入视角出发,分析员工简历或招聘信息中的 AI 技能数据,以间接衡量企业 AI 投资(Babina et al., 2024)。这些 AI 衡量指标虽然在数据可得性和操作性上具有优势,但普遍存在主观偏误、测量噪声和识别精度不足等问题。在内生性问题上,由于更具创新性的企业更可能采用 AI,实证识别中不可避免地存在反向因果和遗漏变量问题。上述研究尽管尝试增加控制变量和引入工具变量的方法,但由于 AI 指标的内生性,因果识别难题尚未得到较好解决。而本文所使用的 AlphaFold 场景凭借其突发性强、应用边界清晰及独立于国内研发的外生特征,有效规避了预期效应、混淆偏误与反向因果等内生性干扰,从而能精准识别生成式 AI 对科技创新的净效应。

本文选取医药和化学行业为处理组,其余行业为对照组。从技术角度而言,AlphaFold 的核心功能是预测蛋白质的三维结构,并非通用的降本增效工具,而是解决特定科学难题的垂直型技术。因此,AlphaFold 对与蛋白质研究密切相关的医药制造、有机化学、化学材料等领域的科技创新影响最为显著。样本数据显示,医药和化学行业的蛋白质专利占总专利比例为 4.3%,而对照组行业的数值为 0.35%,仅为前者的 1/12。处理组行业蛋白质专利的高集中度意味着该行业聚集了大量蛋白质相关的研发活动,是 AlphaFold 技术冲击的主要承载者。综上,我们认为 AlphaFold 的技术冲击主要由医药与化学行业承载。本

文根据证监会《上市公司行业分类指引(2012年修订)》，将医药制造业(C27)、化学原料和化学制品制造业(C26)的企业(下文简称医药化学企业)选取为处理组，其余行业的企业选取为对照组。

基于上述分析，本文构建的模型具体设定如下：

$$\ln Pro_{i,t} = \beta AlphaFold_i \times Treat_i + \Gamma X + \delta_i + \sigma_t + \alpha + \epsilon_{i,t},$$

其中， $AlphaFold$  是代表 AlphaFold 系列第 2 代模型出现的时间哑变量，这代模型发表和开源的 2021 年及之后取 1，否则取 0。AlphaFold 第 1 代模型于 2018 年诞生，但效果不佳，未得到实际应用。改进后的第 2 代模型于 2020 年年底推出，并于 2021 年开源，其生成结果达到非常高的准确率，且速度快、效率高。因此 AlphaFold 第 2 代模型一经诞生就得到广泛使用(如无特别说明，本文中 AlphaFold 特指第 2 代模型)。由于蛋白质是核心生命物质，该模型为生物化学领域带来颠覆性影响，构成考察生成式 AI 影响科技创新的理想外生冲击。 $Treat$  是处理组的指示变量，受益更多的医药和化学企业取 1，否则取 0。

$\ln Pro_{i,t}$  表示企业  $i$  在  $t$  年的蛋白质专利数量加 1 后的对数值，代表企业的蛋白质创新数量。<sup>①</sup> 蛋白质物质名称主要以“某某蛋白”的形式出现，如胶原蛋白等，因此本文将发明名称或摘要中包含关键词“蛋白”的发明专利定义为蛋白质专利，如果专利只是提及“蛋白质”而非某特定蛋白，以“蛋白”进行识别也不会遗漏此情况。发明专利审核周期长，从申请到公开最长有 18 个月的时间滞后，而本文专利数据的收集时间为 2024 年 10 月，这就导致 2023 年的蛋白质发明专利申请数量可能公开不全。因此，本文借鉴吴超鹏和严泽浩(2023)的做法，根据 1990—2016 年专利从申请日到公开日的时间分布情况，对本文所有的发明专利申请数量进行调整。调整系数计算公式见附录 I。<sup>②</sup>

$X_{it}$  为一系列控制变量，主要包括研发投入( $R\&D$ )、生产率水平( $Productivity$ ，营业收入除以员工人数)、流动资产占比( $Current\_asset$ )、股权集中度( $Concentration$ ，第一大股东持股数除以总股数)、高管激励( $Executive$ ，高管持股数除以总股数)、企业年龄( $\ln Age$ )、企业规模( $\ln Size$ ，企业总资产)、资产负债率( $Leverage$ )、资产报酬率( $ROA$ )、营业收入增长率( $Growth$ )。主要变量的描述性统计见表 1。 $\delta_i$  为个体固定效应， $\sigma_t$  为时间固定效应， $\epsilon_{it}$  为随机扰动项， $\alpha$  为截距项。为避免自相关和异方差对估计结果的影响，本文所有回归均采用行业层面的聚类标准误。

本文样本时间区间选定为 2017—2023 年，样本对象为中国沪深 A 股上市公司。2017 年，国务院印发《新一代人工智能发展规划》，国家和企业开始更加重视人工智能技术的发展和應用，因而本文选取 2017 年为样本起始年份。目

① 如无特殊说明，下文中以专利数量为因变量时，均采用相同的对数化处理。

② 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。

前为止本文能获取的最新财务数据截至2023年,因此选择2023年为样本结束年份。企业财务数据和专利数据分别来源于CSMAR数据库和国家专利局。本文剔除了金融业、ST类及退市企业、关键变量缺失及单年度样本,并对所有连续变量进行双侧0.5%缩尾处理。

表1 主要变量的描述性统计

变量	观测值	均值	标准差	最小值	最大值
<i>lnPro</i>	26 764	0.0470	0.252	0	2.079
<i>AlphaFold</i> × <i>Treat</i>	26 764	0.0660	0.249	0	1
<i>R&amp;D</i>	26 764	14.84	7.060	0	22.80
<i>Productivity</i>	26 764	198.4	770.9	-64.89	60940
<i>Current_asset</i>	26 764	0.582	0.200	0.0170	1
<i>Concentration</i>	26 764	32.97	14.76	1.844	89.99
<i>Executive</i>	26 764	0.107	0.170	0	0.879
<i>lnAge</i>	26 764	2.985	0.314	1.946	3.689
<i>lnSize</i>	26 764	22.30	1.317	19.62	26.85
<i>ROA</i>	26 764	0.0330	0.0760	-0.418	0.249
<i>Leverage</i>	26 764	0.409	0.203	0.0440	0.962
<i>Growth</i>	26 764	0.156	0.447	-0.660	3.909

## (二) 基准结果

基准模型的回归结果如表2列(1)所示,在其他因素不变的情况下,相较于对照组企业,AlphaFold使得处理组企业的蛋白质专利申请数量增加2.5%,这说明AlphaFold对企业科技创新的净效应为正,其“创新生成”功能可显著提升科技创新数量,假设1得到验证。

表2 基准回归和稳健性检验结果

变量	基准结果	Qwen 识别	词库扩充	生物学 词典	删除 2023年	连续型 DID
	(1)	(2)	(3)	(4)	(5)	(6)
<i>AlphaFold</i> × <i>Treat</i>	0.025** (0.012)	0.019*** (0.007)	0.042** (0.020)	0.018* (0.010)	0.016** (0.007)	0.016** (0.008)
控制变量	是	是	是	是	是	是
企业固定效应	是	是	是	是	是	是
年份固定效应	是	是	是	是	是	是
观测值	26 764	26 764	26 764	26 764	21 853	26 764
R <sup>2</sup>	0.618	0.642	0.718	0.721	0.657	0.618

注:括号中为聚类到行业的稳健标准误,其余控制变量回归结果已略去。\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ 。如无特殊说明,下表与此相同。

## 四、事前平行趋势和稳健性检验

### （一）事前趋势检验

在 AlphaFold 出现之前,医药化学企业和对照组企业在蛋白质创新数量方面无显著趋势差异,否则意味着可能存在样本选择偏差等内生性问题。本文采用事件研究法,构建以下模型检验事前趋势:

$$\ln Pro_{i,t} = \beta_{\tau} \sum_{\tau=-3}^2 AlphaFold_i \times Treat_i + \Gamma X + \delta_i + \sigma_t + \alpha + \varepsilon_{i,t},$$

其中, $\tau$ 为距离 AlphaFold 冲击发生年的时间间隔, $\tau$ 为负表示冲击发生之前 $\tau$ 年, $\tau$ 为正表示冲击发生之后第 $\tau$ 年, $\tau$ 为0时表示冲击发生当年。回归时以样本区间的第一年即2017年作为基期,因此 $\tau$ 取值区间为-3到2。 $\beta_{\tau}$ 即为本文关注的动态效应系数,如果在 AlphaFold 出现之前该系数接近于0且不显著,则说明医药化学企业和对照组企业在 AlphaFold 出现之前蛋白质创新趋势无显著差异。其余变量定义与基准模型保持一致。如图1所示,在冲击发生之前, $\beta_{\tau}$ 接近于0并且均不显著,未拒绝事前趋势平行的假设。而在冲击之后, $\beta_{\tau}$ 显著大于0。在 AlphaFold 前后处理组和对照组趋势的显著差异,直观表明 AlphaFold 对企业蛋白质创新产生实质性的外生冲击。

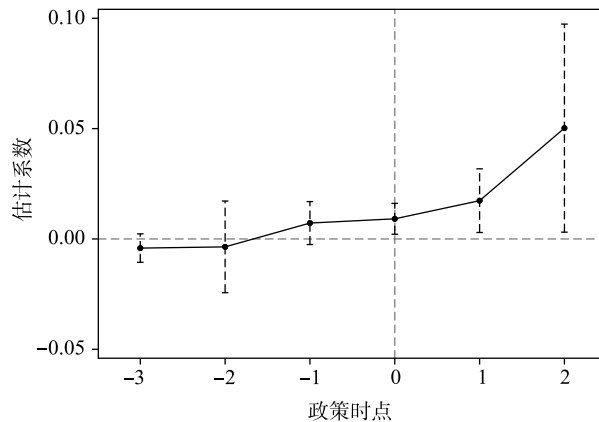


图1 事前趋势检验

注:图中虚线为90%置信区间。

### （二）稳健性检验

#### 1. 蛋白质专利再识别

基于关键词识别蛋白质专利可能存在某些专利只是提及“蛋白”而并不是以其为研究主题的误差,因此需要进行更严谨的识别。本文主要使用三种方法

对蛋白质专利进行再识别。第一,在使用关键词识别蛋白质专利的基础上,使用阿里巴巴的大语言模型 Qwen-Turbo,根据专利的摘要文本进行再识别。第二,以“蛋白”为关键词可能无法充分识别蛋白质专利,一些与蛋白质具有一定相关性的物质其名称并不包含“蛋白”。我们通过阅读文献和专利等资料,将与蛋白质具有一定相关性但并不包含“蛋白”的术语纳入关键词库,具体包含“肽”(组成蛋白质的片段)、“抗体”(免疫球蛋白)、“酶”(绝大多数为蛋白质)等术语。第三,上述术语是蛋白质物质的精炼概括,包含大部分蛋白质物质,但也有部分蛋白质名称不符合上述命名规律,例如干扰素、白介素等。为尽可能全面找出蛋白质,我们整理了科学出版社于2017年出版的《生物学大词典》,它包含25 000余条常用、基础和重要的生物学词目,我们根据定义和解释从中筛选出2 370条蛋白质词目进一步扩充词库。上述三种稳健性检验的结果分别见表2的列(2)至列(4),结果稳健。

## 2. 剔除通用生成式 AI 的影响

2021—2022年,蛋白质和其他行业尚未出现类似于 AlphaFold 这样具有突破性且大规模投入应用的生成式 AI 模型。虽然在 AlphaFold 之后迅速出现了基于 AlphaFold 进行修改优化的其他替代型生成式 AI 模型,但它们在原理上与 AlphaFold 高度同源且时间同步,可将其视为 AlphaFold 冲击的一部分,能更完整地反映 AlphaFold 的技术范式效应。且 AlphaFold 在实际应用中占据绝对主导地位,具有最强的代表性。相比之下,ChatGPT 等通用生成式 AI 的效应会混杂在所有行业的专利产出中,为了确保回归结果捕捉的是 AlphaFold 这一垂直型生成式 AI 在特定科学领域的纯净效应,而非通用工具带来的普遍效率提升,因此需要进行区分处理。2023年是通用生成式 AI 元年(ChatGPT 于2022年年底发布),2023年及以后的创新可能混杂通用生成式 AI 的影响。此外,由于发明专利从申请到公开通常需要15—18个月,本文收集专利数据的时间截至2024年10月,此时2022年及以前的专利基本已全部公开,而2023年的专利尚未完全公开。虽然基准模型中已经根据已有研究对此情况进行考虑,但仍可能存在未知偏误。综合考虑上述因素,本文将2023年数据从样本中剔除后重新进行回归,表2列(5)的结果显示结论保持稳健。

## 3. 处理变量的稳健性分析

为进一步规避人为划分二值处理组(0/1)可能存在的边界粗糙、行业特征影响结果等问题,本文构建连续型双重差分模型,以各行业在 AlphaFold 问世前的蛋白质专利占比作为连续型暴露指标进行检验,表2列(6)的结果依然稳健。此外,计算2017—2020年企业年报中蛋白质词频,并取对数衡量企业层面的 AlphaFold 冲击暴露度。结果仍然稳健,详见附录 II。

#### 4. 其他稳健性检验

本文还进行了以下稳健性检验：第一，使用 ChatGPT 替换 Qwen-Turbo 识别蛋白质专利以防止单个大模型的潜在偏误。第二，使用泊松模型进行回归以处理计数型因变量存在大量 0 值可能导致的偏误。第三，控制新冠疫情冲击。我们一方面剔除跟疫情相关的医药制造业企业，另一方面加入年份与城市的交互固定效应以控制各地不同的疫情情况。第四，使用 PSM-DID 进行回归，减轻样本自选择偏差。第五，剔除无蛋白质专利的完全无关行业以减少行业异质性噪声。第六，为减轻同期其他宏观或行业性冲击干扰结果的潜在担忧，进行处理时间和处理组的混合安慰剂检验。以上结果均稳健或符合预期，详见附录 II。

## 五、生成式 AI 与创新质量

正如理论分析所指出，生成式 AI 对创新质量的影响并非单一维度的提升或抑制，而是呈现出有利于利用式创新但可能无助于探索式创新的分化特征。因此，本部分将深入探讨 AlphaFold 对创新质量的真实影响：一方面，检验其“创新生成”是否通过降低知识重组门槛，促进代表利用式创新质量的组合式创新与跨界合作；另一方面，考察其“守旧训练”机制是否导致创新趋同，以及是否无助于代表探索式创新质量的颠覆性与引领性提升。

### （一）利用式创新

表 3 检验生成式 AI 对利用式创新质量的影响，重点考察其在促进知识重组交叉与组织跨界协作方面的作用。根据 March (1991)，利用式创新主要表现在对现有不同知识要素的整合效率与连接广度上，本文选取技术组合广度 (IPC) 与组织协作深度 (合作专利) 作为衡量利用式创新质量的核心指标。首先，IPC 分类号代表专利所涵盖的技术领域，其数量越多，意味着该项创新整合越多元的技术知识要素 (Lerner, 1994)。列 (1) 的回归结果显示，AlphaFold 显著提升企业蛋白质专利的技术组合广度。这证实 AlphaFold 充当高效的知识连接工具，帮助研发人员将蛋白质结构知识与其他领域的异质性技术模块进行重新排列与组合，从而丰富创新的技术内涵。

其次，从组织合作网络维度来看，合作主体与跨行业合作反映了打破组织与认知边界的能力，表征了知识的“融合深度”。如果专利的“申请人”字段中包含除本企业以外的其他机构 (包括高校、科研院所、其他企业)，则定义为“合作专利”。列 (2) 和列 (3) 的结果显示，AlphaFold 显著增加由多个企业或组织合作研发的蛋白质专利数量以及蛋白质专利中的合作组织数量，说明生成式 AI 降

低知识传递与沟通的成本,从总体上活跃了研发协作网络。更进一步,本文通过匹配专利和工商注册数据获得企业行业信息,进而可以将合作的蛋白质专利细分为“行业间合作”与“行业内合作”进行异质性检验。列(4)结果显示,AlphaFold对行业间合作研发的蛋白质专利具有显著的促进作用;而列(5)则显示其对行业内合作研发的蛋白质专利的影响并不显著。这一核心发现有力地证明:AlphaFold的核心价值在于打破不同学科和行业间的知识壁垒,促进异质性知识的互补与融合。综上所述,表3的结果从技术组合的广度和跨界合作的深度两个维度,证明生成式AI显著提升利用式创新的质量,从而有力支持假设2a。

表3 生成式AI与利用式创新

变量	专利IPC数量	合作专利数量	合作者数量	行业间合作专利	行业内合作专利
	(1)	(2)	(3)	(4)	(5)
<i>AlphaFold</i> × <i>Treat</i>	0.040*** (0.014)	0.011*** (0.002)	0.005** (0.002)	0.005** (0.002)	-0.004 (0.004)
控制变量	是	是	是	是	是
年份固定效应	是	是	是	是	是
企业固定效应	是	是	是	是	是
观测值	26 764	26 764	26 764	26 764	26 764
R <sup>2</sup>	0.462	0.467	0.423	0.399	0.543

## (二) 探索式创新

表4检验生成式AI对探索式创新质量的影响,结果显示AlphaFold加剧创新趋同,且未能显著提升颠覆性与引领性。理论上,生成式AI的“守旧训练”机制可能导致生成内容向训练数据的统计规律靠拢,引发同质化。为检验这一猜想,本文参考Kelly et al.(2021)的TF-IDF算法,基于专利摘要文本计算语义相似度(值越大代表独特性越低),并将相似度排序前10%的专利定义为高相似度专利。<sup>①</sup>列(1)和列(2)显示AlphaFold显著增加企业的高相似度蛋白质专利数量和占比,同时我们也计算企业所有蛋白质专利的平均相似度,列(3)表明AlphaFold同样提升这一指标。上述发现与Doshi and Hauser(2024)的结论相呼应,即对生成式AI依赖度越高,产出的区分度越低。

针对趋同现象,必须承认在研发密集型领域,不同主体沿着同一技术前沿开展并行探索也可能会导致相似度上升,这本身是技术迭代的特征之一。然

① 我们也尝试以相似度排序前20%和50%为高相似度专利的标准,结果同样稳健,详见附录III。

而,在创新经济学的经典语境下,专利文本相似度的显著、系统性上升,普遍被视为创新多样性下降、探索空间收窄的一个强相关信号(Kelly et al., 2021; Park et al., 2023; 孙雅慧等, 2024)。本文认为这种趋同既可能是创新向 AI 主导的研发范式集中,同时还是研发活动被引导至模型预测概率最高的领域,从而挤占了边缘性、试错性探索的空间。因此,本文将相似度上升解释为创新生态系统面临的同质化风险,而非对在单项技术上密集研发的价值之否定。

表 4 生成式 AI 与探索式创新

变量	高相似专利 数量	高相似专利 占比	专利平均 相似度	专利颠覆性	专利引领性
	(1)	(2)	(3)	(4)	(5)
<i>AlphaFold</i> × <i>Treat</i>	0.145*** (0.022)	0.124*** (0.028)	0.208*** (0.034)	0.049 (0.064)	-0.394 (0.347)
控制变量	是	是	是	是	是
年份固定效应	是	是	是	是	是
企业固定效应	是	是	是	是	是
观测值	876	876	876	876	876
R <sup>2</sup>	0.249	0.136	0.465	0.504	0.054

接着,本文参考 Funk and Owen-Smith(2017)、Park et al.(2023)的经典算法构建专利颠覆性指数,该指标通过分析引文网络结构,衡量一项专利是强化现有技术路线(巩固型)还是开启全新方向(颠覆型)。同时,参考主流文献利用相对被引次数(剔除时间趋势后的被引量)来衡量专利的引领性。表 4 的列(4)和列(5)中颠覆性和引领性指标的系数均不显著,说明 AlphaFold 无助于提升创新的颠覆性与引领性。<sup>①</sup> 这警示我们,当前的“AI 驱动”范式可能存在“路径锁定”风险,即创新更多是在已知路径上的优化模仿,而非涌现出超越现有范式的原始创新。综上所述,表 4 的经验证据一致支持假设 2b:生成式 AI 在推动颠覆性与引领性等探索式创新方面尚无显著建树,并且加剧创新的趋同。

## 六、“AI 驱动”创新范式的转型和挑战

前文揭示生成式 AI 驱动创新的数量和质量效应,其背后是科技创新正经历从“人驱动”向“AI 驱动”的范式转型。为进一步揭示这一变革,本部分从创

<sup>①</sup> 我们还进行了以下稳健性检验:第一,尝试剔除企业“自引”带来的潜在问题。第二,考虑到表 4 回归系基于拥有蛋白质专利的子样本(N=876),较小的样本量可能影响估计的稳定性。为此,本文参考基准回归构建了全样本下(N=26 764)的企业年度高相似度、高被引及高颠覆性专利数量指标。第三,利用千问大模型(Qwen)将专利文本向量化后计算相似度。上述结果依然稳健,详见附录 III。

新体系的转型及其面临的挑战两方面展开分析。

### (一) 创新范式的“AI 驱动”转型

生成式 AI 的出现并非仅在工具层面引入新手段,而是推动科研体系的底层结构性转型。这一转型主要表现在研究方向、资源配置和人才结构的重新分布,构成科技创新体系的“AI 驱动”重构。

#### 1. 创新重心转移

当前生成式 AI 本身已成为科研活动的主导要素之一,驱动企业在创新项目选择上更加倾向于具有 AI 参与潜力的技术方向,形成“AI 导向”的研发投资结构。表 5 列(1)和列(2)的结果表明 AlphaFold 促使蛋白质专利占比显著上升,而非蛋白质专利数量下降,结合基准结果说明企业创新活动的重心明显向 AI 可赋能的蛋白质领域集中。这种资源集中效应一方面提升 AI 赋能领域的创新产出,另一方面也对非 AI 赋能领域产生一定的“挤出效应”,导致创新更加偏向 AI 赋能的方向。这一结果也有助于因果识别,它反向证明我们观测到的蛋白质创新增加并非源于可带来创新普遍增加的普惠政策,而是源于 AlphaFold 导致的蛋白质领域的相对优势。

表 5 创新体系的“AI 驱动”转型

变量	蛋白质专利 占比	非蛋白质 专利数量	AI 招聘岗位	蛋白质专利 发明人数	人均专利 数量
	(1)	(2)	(3)	(4)	(5)
<i>AlphaFold</i> × <i>Treat</i>	0.002* (0.001)	-0.079** (0.040)	0.248*** (0.057)	0.026*** (0.010)	0.003*** (0.001)
控制变量	是	是	是	是	是
年份固定效应	是	是	是	是	是
企业固定效应	是	是	是	是	是
观测值	26 764	26 764	23 791	26 764	4 921 796
R <sup>2</sup>	0.375	0.811	0.714	0.581	0.312

#### 2. 人才需求结构变化

随着 AI 技术在创新中的嵌入加深,企业需要更多的 AI 人才从事模型部署、算法优化、数据清洗等工作,因此对 AI 人才的招聘可能会增加。同时,为配合企业在 AlphaFold 赋能的蛋白质研究领域的扩张,蛋白质研发人员数量或许也会同步增长。为验证这一可能性,本文从数百万的企业招聘数据中基于词典法构建 AI 人才招聘岗位(姚加权等,2024)<sup>①</sup>,基于上市企业蛋白质专利库构建

① AI 招聘岗位以及本部分中其他变量的具体构造过程请参考附录 IV。

企业蛋白质研发人员数量指标,表 5 列(3)和列(4)的结果支持上述假设,说明企业的人才需求也向“AI 驱动”的创新范式靠拢。

### 3. 个人研发效率提升

AlphaFold 通过大幅减轻蛋白质结构测定难度、加快测定速度、提供更多创新可能,可以提升个人产出效率。我们从企业单独申请的专利的发明人字段获取企业蛋白质研究者名单,经过去重和统计可知每位研究者的蛋白质专利产出。<sup>①</sup>表 5 的列(5)显示 AlphaFold 的出现显著提高研究者的人均蛋白质专利数量。

## (二)“AI 驱动”创新范式的转型挑战

尽管生成式 AI 为科技创新赋予了强大能力,但“AI 驱动”范式在现实运行中仍面临多重挑战与约束,主要体现在算力与数据依赖、企业研发效率的短期阵痛和科研方向的 AI 依赖等方面。

### 1. 深化的算力和数据依赖

生成式 AI 的运行高度依赖于大规模算力与高质量数据资源。为考察这一转变,表 6 列(1)指出 AlphaFold 出现后,处理组企业与互联网、信息技术服务等算力密集型行业的合作专利数量显著上升。尽管这一指标形式上反映了跨行业协作的增强,但在“AI for Science”的语境下,它很有可能映射出企业为克服自身算力瓶颈而向外部寻求算力资源。逻辑在于,运行 AlphaFold 所需的算力基础设施并非传统生物医药企业的核心资产。因此,与拥有充裕算力的组织合作成为获取计算资源的高效途径。现实案例也佐证了这一逻辑:例如山东大学基于高性能计算云平台建成 AlphaFold 批量预测平台,北京大学联合华为推出基于 AlphaFold 的结构预测工具。<sup>②</sup>因此,合作专利数量的上升,间接表明算力已成为“AI 驱动”创新范式下企业必须解决的关键要素约束。

表 6 “AI 驱动”创新范式的转型挑战

变量	与 IT 企业合作专利	有数据积累	无数据积累	企业研发效率
	(1)	(2)	(3)	(4)
<i>AlphaFold</i> × <i>Treat</i>	0.006*** (0.002)	0.023*** (0.005)	0.027 (0.019)	-0.028** (0.010)
控制变量	是	是	是	是
年份固定效应	是	是	是	是

<sup>①</sup> 我们还考虑不同的研究者蛋白质专利产出指标,例如以发明人位次加权、只考虑第一发明人等情况,结果均稳健,详见附录 V。

<sup>②</sup> 详见 <https://www.view.sdu.edu.cn/info/1022/155635.htm>; <https://www.chem.pku.edu.cn/kjz/132966.htm>, 访问时间:2025 年 10 月 5 日。

(续表)

变量	与 IT 企业合作专利	有数据积累	无数据积累	企业研发效率
	(1)	(2)	(3)	(4)
个体固定效应	是	是	是	是
观测值	26 764	24 125	25 836	21 924
R <sup>2</sup>	0.392	0.624	0.600	0.246

高质量的数据积累是生成式 AI 发挥效能的另一关键约束。在“AI 驱动”范式下,数据已成为生成模型训练与推理的核心燃料。本文从企业内部数据积累的视角考察了这一机制。企业积累的私有蛋白质数据能够用于 AlphaFold 针对企业研究领域的微调和优化,从而提升其生成的精准度。本文以 AlphaFold 出现前企业是否拥有蛋白质专利作为数据积累的代理变量。表 6 的列(2)和列(3)显示,AlphaFold 的创新促进效应仅在具备蛋白质数据积累的企业样本中显著。综上所述,算力协作需求的上升与数据门槛的存在,共同揭示了“AI 驱动”范式下科研资源依赖性的深化。需要警惕的是,当前算力与数据愈发集中于头部企业的趋势使创新活动更容易被少数企业垄断,从而可能加剧科研资源的不均衡。

## 2. 企业研发效率的下降

表 6 列(4)显示,AlphaFold 显著降低每单位研发支出带来的专利产量,即企业整体研发效率下降。这一发现与前文的个人研发效率上升并不矛盾,代表的是研发范式从“人驱动”向“AI 驱动”转型过程中资本深化的结果。具体而言,AlphaFold 作为强有力的科研辅助工具与研发人员形成互补,直接推动了劳动生产率的跃升;但在宏观企业层面,“AI 驱动”范式要求企业在算力设备、数据清洗及人员培训等方面进行资本开支。因此,当研发资本投入(分母)的增速超过创新产出(分子)的增速时,企业层面的资金回报效率出现短期下降。这符合 Brynjolfsson et al.(2021)提出的“生产率 J 曲线”效应,即企业在转型初期因大量投资被计入当期成本而面临的暂时性效率阵痛,反映构建长期 AI 竞争力所需的转型成本而非创新能力的实质衰退。当然,在长期中“AI 驱动”的科技创新体系效率如何变化值得进一步探究。

## 3. 科研方向的“AI 依赖”风险

需要注意的是,AlphaFold 驱动企业研发重心向蛋白质领域转型并非仅源于蛋白质领域重要性提升,更多的是科研方向从科学问题驱动向 AI 可行性驱动的转变。AlphaFold 的突破性成功,关键在于蛋白质领域长期积累的高质量结构数据以及相对成熟的生物信息学研究。这一数据与方法基础,使得相关问题更容易被形式化为可由 AI 模型学习和验证的预测任务,从而显著提高 AI 模

型性能。在此条件下,蛋白质研究在生成式 AI 赋能下呈现出更高的技术确定性和成果可预期性,因而在不确定性较高的创新环境中,更容易被企业视为风险较低、回报更可控的研发方向。这种由 AI 技术能力边界所诱发的方向集中,本质上强化了科研活动对既有数据和技术路径的依赖,形成新的“路径锁定”机制。它虽有助于提升研发效率,但若企业研发持续围绕少数 AI 高度适配的研究领域展开,可能压缩科学探索的多样性,抑制对数据基础薄弱、高不确定性问题的投入,进而使那些当前尚无法被 AI 有效处理、却可能孕育颠覆性突破的研究方向被系统性忽视。

## 七、政策含义

本文的实证发现表明,生成式 AI 在提升创新产出、重塑科研范式的过程中,伴随着创新趋同、关键要素依赖加深等问题。基于此,政策制定应当超越单纯的生成式 AI 技术推广,转向构建更加包容、多元的创新生态。第一,构建合理的人-AI 协作科研机制,规避科研同质化风险。本文发现生成式 AI“守旧训练”带来创新趋同风险,因此一方面应重视对 AI 模型难以有效预测或适配科研领域的支持;另一方面,在 AI 高效驱动的研究领域,探索 AI 负责高效验证、人类负责突破边界的互补格局,防止科研生态因过度依赖 AI 而丧失演化活力。第二,强化普惠共享的关键要素建设思路,打破“AI 驱动”范式的准入壁垒。算力与数据日益成为 AI 时代创新的关键要素,应进一步加强科研算力基础设施和高质量科学数据资源建设,探索更加开放、协同的资源供给方式,降低高校、中小企业和初创团队进入“AI for Science”领域的基础门槛,缓解要素获取能力差异带来的创新机会不均问题。

## 参考文献

- [1] Agrawal, A., J. Gans, and A. Goldfarb, “Prediction, Judgment, and Complexity: A Theory of Decision-Making and Artificial Intelligence”, *The Economics of Artificial Intelligence: An Agenda*, 2018, 89-110.
- [2] Babina, T., A. Fedyk, A. He, and J. Hodson, “Artificial Intelligence, Firm Growth, and Product Innovation”, *Journal of Financial Economics*, 2024, 151, 103745.
- [3] Bianchini, S., M. Müller, and P. Pelletier, “Artificial Intelligence in Science: An Emerging General Method of Invention”, *Research Policy*, 2022, 51(10), 104604.
- [4] Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., “Language Models Are Few-Shot Learners”, *Advances in Neural*

- Information Processing Systems*, 2020, 33,1877-1901.
- [5] Brynjolfsson, E., D. Rock, and C. Syverson, "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies", *American Economic Journal: Macroeconomics*, 2021, 13(1), 333-372.
- [6] Cockburn, I. M., R. Henderson, and S. Stern, *The Impact of Artificial Intelligence on Innovation*. Cambridge, MA: National Bureau of Economic Research, 2018.
- [7] Doshi, A. R., and O. P. Hauser, "Generative AI Enhances Individual Creativity but Reduces the Collective Diversity of Novel Content", *Science Advances*, 2024, 10(28), eadn5290.
- [8] Fleming, L., "Recombinant Uncertainty in Technological Search", *Management Science*, 2001, 47(1), 117-132.
- [9] Funk, R. J., and J. Owen-Smith, "A Dynamic Network Measure of Technological Change", *Management Science*, 2017, 63(3), 791-817.
- [10] Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets", *Advances in Neural Information Processing Systems*, 2014, 27, 2672-2680.
- [11] 洪永淼、汪寿阳, "人工智能新近发展及其对经济学研究范式的影响", 《中国科学院院刊》, 2023年第3期, 第353—357页。
- [12] 胡蓉宁、王跃堂, "人工智能的技术创新效应及其实现机制", 《江海学刊》, 2023年第6期, 第99—105页。
- [13] Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, and D. Hassabis, "Highly Accurate Protein Structure Prediction with AlphaFold", *Nature*, 2021, 596(7873), 583-589.
- [14] Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy, "Measuring Technological Innovation over the Long Run", *American Economic Review: Insights*, 2021, 3(3), 303-320.
- [15] Lerner, J., "The Importance of Patent Scope: An Empirical Analysis", *The RAND Journal of Economics*, 1994, 25(2), 319-333.
- [16] 李建会、杨宁, "AI for Science: 科学研究范式的新革命", 《广东社会科学》, 2023年第6期, 第81—92页。
- [17] 李伦、刘梦迪, "人工智能驱动的科学范式革命: 态势与未来", 《探索与争鸣》, 2024年第10期, 第143—151页。
- [18] 李玉花、林雨昕、李丹丹, "人工智能技术应用如何影响企业创新", 《中国工业经济》, 2024年第10期, 第155—173页。
- [19] 刘景江、郑畅然、洪永淼, "机器学习如何赋能管理学研究? ——国内外前沿综述和未来展望", 《管理世界》, 2023年第9期, 第191—216页。
- [20] March, J. G., "Exploration and Exploitation in Organizational Learning", *Organization Science*, 1991, 2(1), 71-87.
- [21] Merchant, A., S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, "Scaling Deep Learning for Materials Discovery", *Nature*, 2023, 624(7990), 80-85.
- [22] Naddaf, M., "How Are Researchers Using AI? Survey Reveals Pros and Cons for Science", *Nature*, 2025.

- [23] Park, M., E. Leahey, and R. J. Funk, "Papers and Patents Are Becoming Less Disruptive Over Time", *Nature*, 2023, 613(7942), 138-144.
- [24] Rammer, C., G. P. Fernández, and D. Czarnitzki, "Artificial Intelligence and Industrial Innovation: Evidence from German Firm-Level Data", *Research Policy*, 2022, 51(7), 104555.
- [25] Strumsky, D., and J. Lobo, "Identifying the Sources of Technological Novelty in the Process of Invention", *Research Policy*, 2015, 44(8), 1445-1461.
- [26] 孙雅慧、时省、彭飞, "研发补贴与渐进式创新锁定: 基于机器学习的专利文本分析", 《经济研究》, 2024年第11期, 第89—105页。
- [27] Van Noorden, R., and J. M. Perkel, "AI and Science: What 1,600 Researchers Think", *Nature*, 2023, 621(7980), 672-675.
- [28] Verhoeven, D., J. Bakker, and R. Veugelers, "Measuring Technological Novelty with Patent-Based Indicators", *Research Policy*, 2016, 45(3), 707-723.
- [29] 吴超鹏、严泽浩, "政府基金引导与企业核心技术突破: 机制与效应", 《经济研究》, 2023年第6期, 第137—154页。
- [30] 肖柏高、周艺珮、谢锐, "生成式 AI 如何变革经济学文本分析范式", 《湖南大学学报(社会科学版)》, 2025年第3期, 第63—71页。
- [31] 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, 第101—116页。
- [32] 政光景、吕鹏, "生成式人工智能与哲学社会科学新范式的涌现", 《江海学刊》, 2023年第4期, 第132—142页。

## The Quantity-Quality Effect of Generative AI-Driven Innovation: Evidence from the Nobel Prize-Winning AlphaFold

XIAO Baigao ZHOU Yang

(Hunan University)

LIU Qing\*

(Renmin University of China)

**Abstract:** Using the Nobel-winning AlphaFold as an exogenous shock, we try to study how generative AI affects innovation. We find AlphaFold significantly increases the quantity of protein innovations, but its impact on quality is mixed. It improves exploitative innovation

---

\* Corresponding Author: LIU Qing, Room 738, Main Building, Mingde Building, Renmin University of China, Haidian District, Beijing 100872, China; Tel: 86-15810897360; E-mail: qingliu@ruc.edu.cn.

---

through knowledge combination. However, limited by its training data, it causes innovation to become similar, failing to promote disruptive explorative innovation. Furthermore, while shifting research to an “AI paradigm” increases per-capita output, it creates challenges like heavy reliance on computing and data, lower short-term corporate R&D efficiency, and narrower research directions.

**Keywords:** generative AI; technological innovation; quantity-quality effect

**JEL Classification:** O31, O32, D83