

# 中国教育财政

怀仁怀朴 唯真唯实

北京大学中国教育财政科学研究所

2025 年第 15 期（总第 280 期）

2025 年 7 月 11 日

## 中国大学的需求导向基础研究：概念演变、 基本事实与政策建议

闵亦杰 等\*

**摘要：**加强基础研究是实现高水平科技自立自强的迫切要求。大学是国家基础研究的主力军，大学基础研究可以分为自由探索性与需求导向性两类。其中，需求导向的基础研究将国家重大战略需求和经济社会发展目标作为驱动源头，强调在较短期内就能发挥作用、直接支撑技术创新过程。目前，我们对中国大学需求导向基础研究的基本情况缺少了解，难以回应社会公众对于“中国大学基础研究到底有没有用”的关切。研究团队在梳理相关概念演化的基础上，基于“国际专利引用近期学术论文数量/次数”设计了需求导向基础研究的代理性测量指标，发现如下基本事实：（1）相比于企业或科研院所，大学是中国需求导向基础研究的主导性力量；（2）中国大学需求导向基础研究规模不断扩大，在主要科技强国中地位不断上升；但历史累积存量仍偏低、领域间不平衡性较强；（3）中国大学需求导向基础研究

\* 闵亦杰，北京大学中国教育财政科学研究所助理研究员。本文数据收集整理计算以及部分写作工作由杨佳郑（北京大学中国教育财政科学研究所硕士研究生）完成；部分数据计算由侯万方（南京大学信息管理学院硕士研究生）完成；部分理论建构工作由史冬波副教授（上海交通大学国际与公共事务学院）与张闰龙副教授（北京大学光华管理学院）完成。

的本国吸收有限、国际溢出明显。基于这些事实，研究团队建议：（1）确立针对大学需求导向基础研究的“成效观”，引导大学建立兼容需求导向基础研究的激励体系，增强研究型大学在国家战略科技力量中的地位；（2）适应产业界吸收基础研究进行技术创新的需求，扩大博士生培养规模、优化培养模式；（3）由教育行政部门主导概念验证平台建设，帮助大学需求导向基础研究以更高的成熟度走出大学围墙。

**关键词：**研究型大学；需求导向；基础研究；高水平科技自立自强

## 一、概念与测量

“需求导向”与“基础研究”这两个概念在国家政策性话语中的结合存在一个演进过程。二者正式结合的一个标志是《“十三五”国家科技创新规划》（2016年），其中关于基础研究的表述为“坚持面向国家重大需求和世界科学前沿，坚持鼓励自由探索和目标导向相结合”。之后，教育部于2022年印发《关于加强高校有组织科研 推动高水平自立自强的若干意见》，强调“[在]持续开展高水平自由探索研究的基础上……更好服务国家安全和经济社会发展面临的现实问题和紧迫需求”。2023年习近平总书记在二十届中央政治局第三次集体学习时的讲话进一步明确了该概念的内涵，指出基础研究要“坚持目标导向和自由探索‘两条腿走路’，把世界科技前沿同国家重大战略需求和经济社会发展目标结合起来”。明确指出“需求导向”中“需求”一词内涵不再局限于国家战略需求，而是将经济社会发展需求也囊括其中<sup>①</sup>。中国大学的需求导向基础研究进行的如何？这是本报告尝试回答的核心问题。

“需求导向”概念形容的是科研工作的“出发点”和“目标”，难以直接测量，故研究团队采用代理性（proxy）测量的思路，站在科技企业等技术创新主体作为基础研究“阅卷人”的视角，通过国际专利对各国大学学术论文的引用，来测量基础科学研究是否被后续技术创新吸收，亦即是否“面向了后续技术创新的需求”。该方法是国际技术创新研究者常用的一个测量方法<sup>②</sup>。在实际操作中，研究团队将专利的范围限制在了国际专利范畴，这是代表全球高质量专利的一个集

---

<sup>①</sup> 随着政策界对需求导向基础研究的讨论逐渐升温，大量变体性概念逐渐涌现，包括“国家战略需求导向”“目标导向”“需求牵引”等等。本报告对其中的细微差异不做区分，统一使用“需求导向基础研究”一词。对相关政策中典型表述的摘录，见附录一。

<sup>②</sup> 更详细的合理性讨论，见附录二。

合，基于该集合所得结论一般是保守结论<sup>①</sup>。针对每项国际专利，研究团队识别了其近期学术论文的引用，即被其引用的、发表于其授权日期之前五年以内的学术论文<sup>②</sup>，并根据论文第一作者所属机构来确定论文国别；授权于 2001—2022 年的国际专利共引用全球大学论文 515,279 次，其中引用中国大学论文共 32,668 次<sup>③</sup>。

## 二、基本事实

### （一）大学是中国需求导向基础研究的主导性力量

研究团队首先对比了大学、科研院所、企业三种创新力量的需求导向基础研究规模，及其随年份变化的趋势（图 1a）。可以发现，在 2001 到 2022 年区间，大学是推动中国需求导向基础研究的最主要力量，其发表的学术论文被国际专利引用的占比始终高于科研院所与企业，并有随时间进一步提高的趋势，到 2022 年时，占比已达到 74.0%。科研院所占比呈下降趋势，从 2001 年时的 26.2% 下降到 2022 年时的 12.0%。企业占比最低，维持在 1.8% 到 10.5% 之间，且几乎不存在随时间变化而提高的趋势（2022 年时为 2.2%）。在此基础上，对比了中美两国需求导向基础研究的结构差异（图 1b）。从大学与企业的对比来看，中美两国都以大学作为需求导向基础研究的主导性力量。但是，中国大学力量占比在 2006 年后开始系统性高于美国，而中国企业力量占比一直低于美国。在 2022 年，中国大学与企业的需求导向基础研究规模之比约为 34：1，而美国约为 7：2。这种大学与企业的力量对比失衡，可能影响企业吸收大学基础研究成果的能力。

---

<sup>①</sup> 国际专利是对 PCT 专利的一种标签，PCT 即 Patent Cooperation Treaty，是一种专利申请形式。一般来说 PCT 专利追求全球保护，也愿意为此付出更高成本，这使得 PCT 专利成为定义国际上“好专利”的一个范围。

<sup>②</sup> 需指出的是，自由探索性基础研究也可能被专利引用，但滞后性一般较强。本文只将专利所引用的近期发表的论文纳入考虑，这在很大程度上排除了自由探索性基础研究。

<sup>③</sup> 本报告使用到的数据库及其介绍，见附录三。

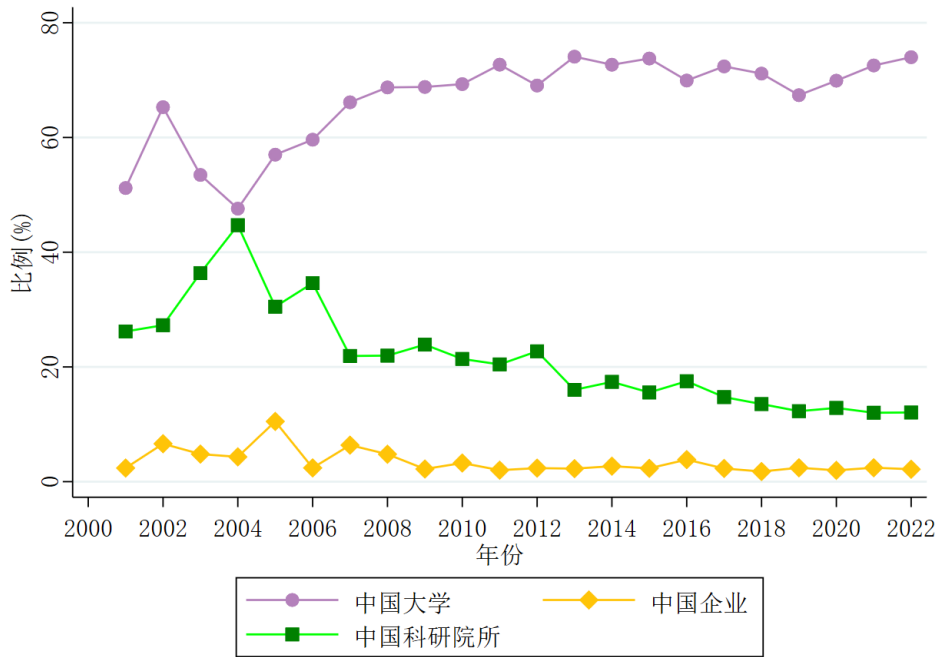


图 1a 中国各类型机构力量变化趋势

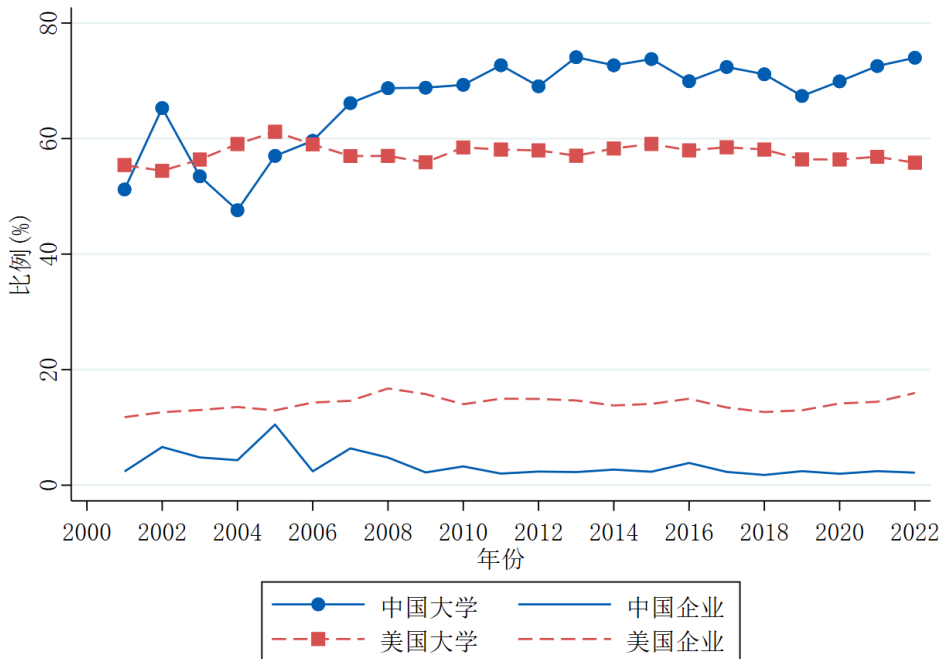


图 1b 中美两国大学与企业力量对比

注：Openalex 数据库提供了每篇论文所有作者所属机构的唯一 ROR 标识符（ROR 即 Research Organization Registry），这使得本文可以从 ROR 数据库中获得全部机构的类别。这其中，我们将“Education”定义为大学，“Company”定义为企业，“Facility”定义为科研院所，剩余分类归类为其他。在人工校验过程中，我们修正了少量中国机构的类别，如将中科院的分类由“Government”更正为“Facility”。本文以第一作者所属机构作为该论文的唯一所属机构进行统计。图中横轴为年份（2001-

2022), 纵轴为大学、科研院所、企业等力量分年份的占比。例如, 2022 年中国大学占比=2022 年授权专利引用中国大学论文的次数÷2022 年授权专利引用中国论文的次数×100%。需注意的是:(1) 本文另以最后一位作者所属机构做稳健性分析, 但由于两种方法所得结果具有高度的可比性, 故本文仅使用第一作者口径下的结论;(2) 使用 5 年口径与全口径指标所得结论亦具有高度可比性, 故此处仅展示使用 5 年口径之结论。

## **(二) 中国大学需求导向基础研究规模不断扩大, 在主要科技强国中地位不断上升; 但历史累积存量仍偏低、领域间不平衡性较强**

进一步考察中国与各科技强国大学之间的需求导向基础研究规模对比(图 2a)。国际专利引用中国大学论文占其全部引用的比例, 从 2001 年的 0.25% 上升到 2011 年的 5.16% 又上升到 2022 年的 13.16%; 22 年间涨幅超过 50 倍。不仅如此, 中国大学需求导向基础研究规模的国际排名在短时间内(2011—2015 年)超过英国、日本、德国等科技强国, 位居世界第二位。中美比较的结果显示中国大学的需求导向基础研究虽然在 2022 年依然落后于美国, 但差距在逐渐缩小。中美两国大学需求导向基础研究规模的比值从 2001 年时的 1:233 缩小到 2022 年的 1:3。考虑到中美两国基础研究的投入规模本身也存在差距(例如 2022 年美国国家科学基金会(NSF)的资助规模是中国自然科学基金委资助规模的 1.8 倍), 两国大学基础研究投入转化为国际专利引用的效率差距可能还要小于 1:3。此处国际比较存在可比口径选择等难点, 研究团队将继续探索这一问题。

从历史累积的角度来看(图 2b), 中国大学论文被国际专利引用的占比从 2001 年的 0.13%, 上升到 2022 年的 4.69%, 其涨幅也超过了 30 倍。在此口径下计算的中国需求导向基础研究规模超过了加拿大和法国, 但仍低于美国、英国、日本、德国等科技强国。相比五年口径, 历史累积口径下的中国大学指标普遍较低, 说明中国大学需求导向基础研究的存量追赶仍需一个较长的过程。此外, 研究团队还以论文年份为基准测算各国大学发表于同一年份的论文被引次数占比、单纯测算外国专利对各国学术论文的引用以排除近年来中国专利数量激增的影响, 前述主要结论依然成立。

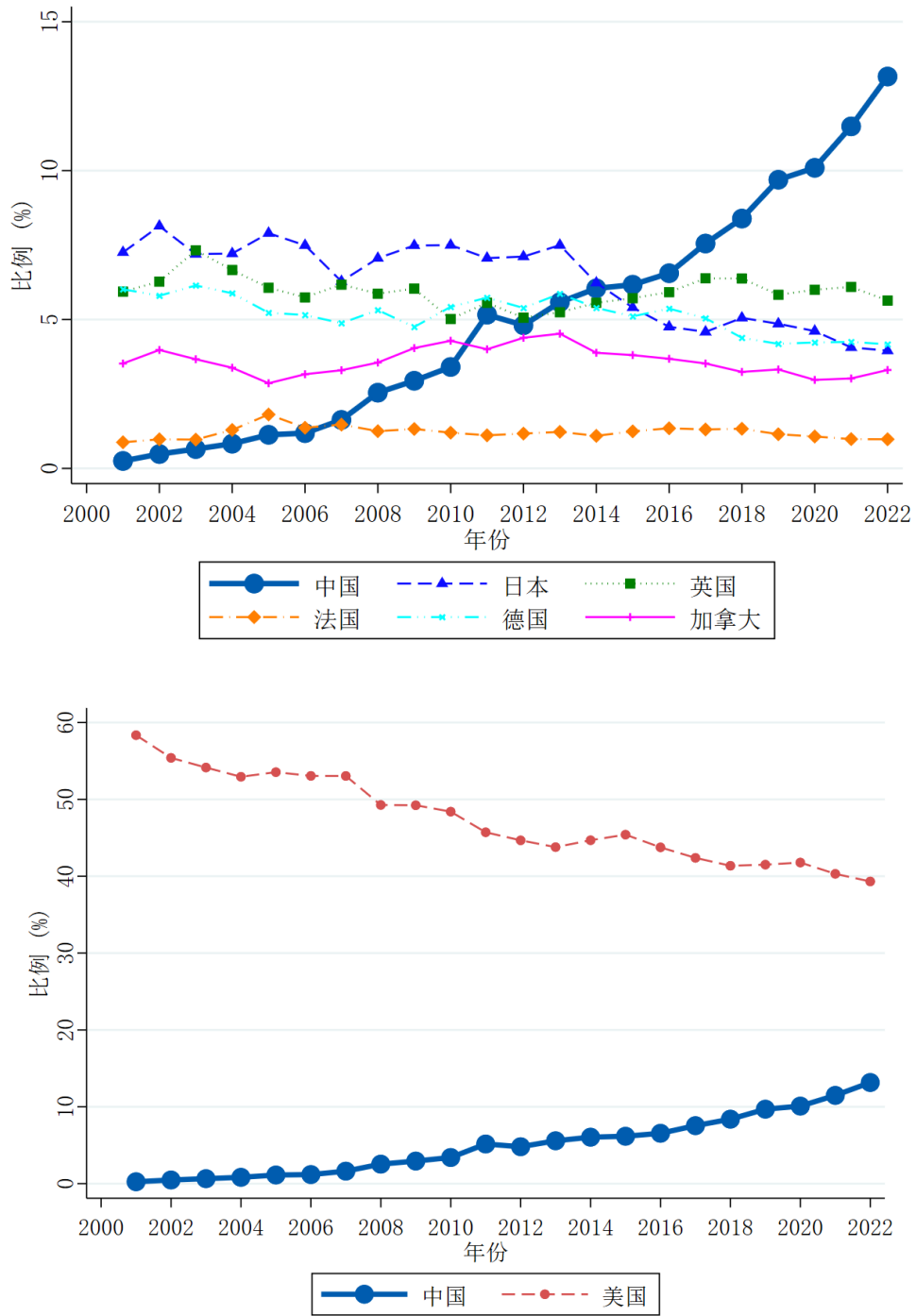


图 2a 各国大学需求导向基础研究能力对比（五年专利引用论文口径）

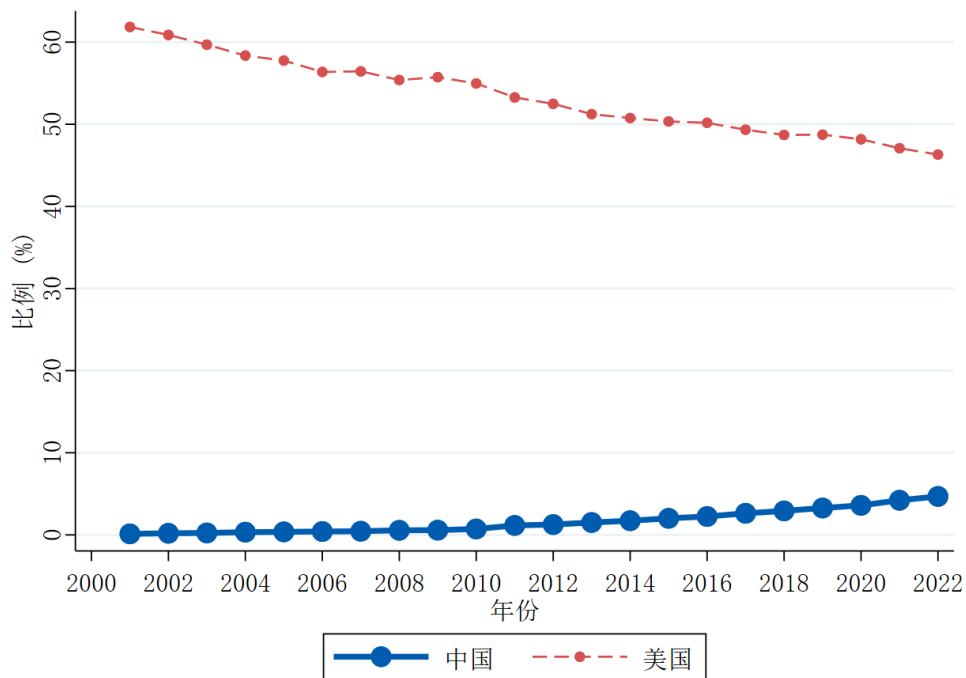
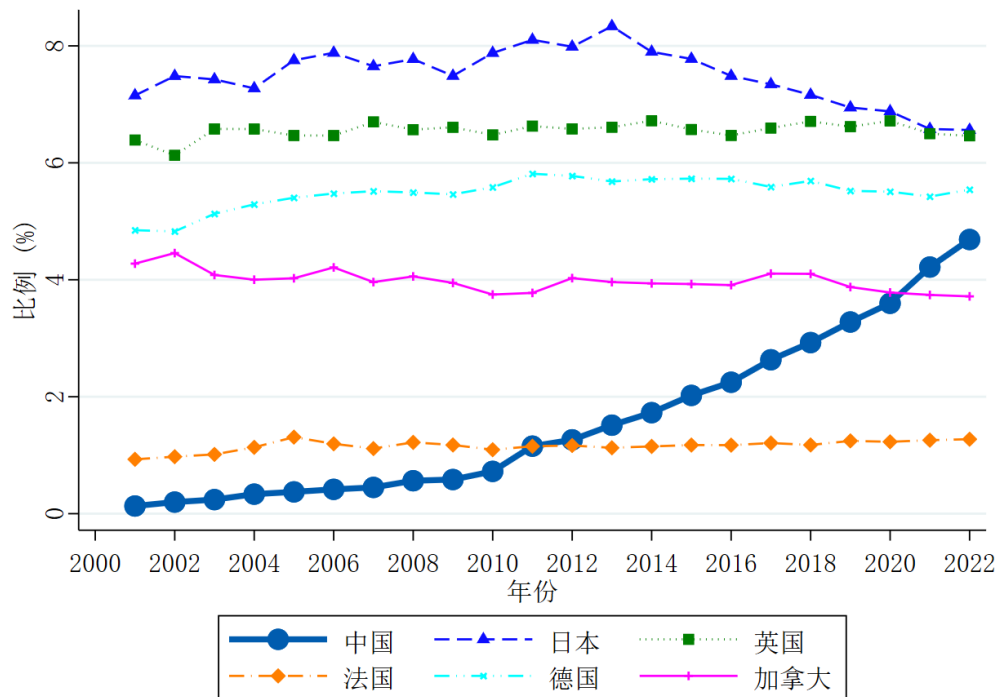


图 2b 各国大学需求导向基础研究能力对比（全周期专利引用论文口径）

注：图 2a/b 中的横轴为专利授权/论文发表年份，图 2a/b 纵轴为当年授权专利引用论文中各国大学论文被引次数的比例，例如 2022 年中国占比=2022 年授权专利引用中国大学论文的次数÷2022 年授权的专利引用世界大学论文的次数×100%。

研究团队还分析了中国大学基础研究对全球技术创新贡献的领域间差异(图

3), 发现不同技术领域引用中国大学论文的比例差异显著。其中, 运输工程、环境技术、化学工程、材料科学与冶金、数字通信、视听技术、电气机械与装置等技术领域国际专利引用中国大学论文的比例显著更高。相对而言, 基础通信原理、半导体、光学、生物材料分析、医疗技术、生物技术、制药技术、家具与游戏产品等领域所引用中国大学论文占其引用全球论文的比例均低于 6%, 表明中国大学基础研究响应这几个技术领域需求的程度较低。

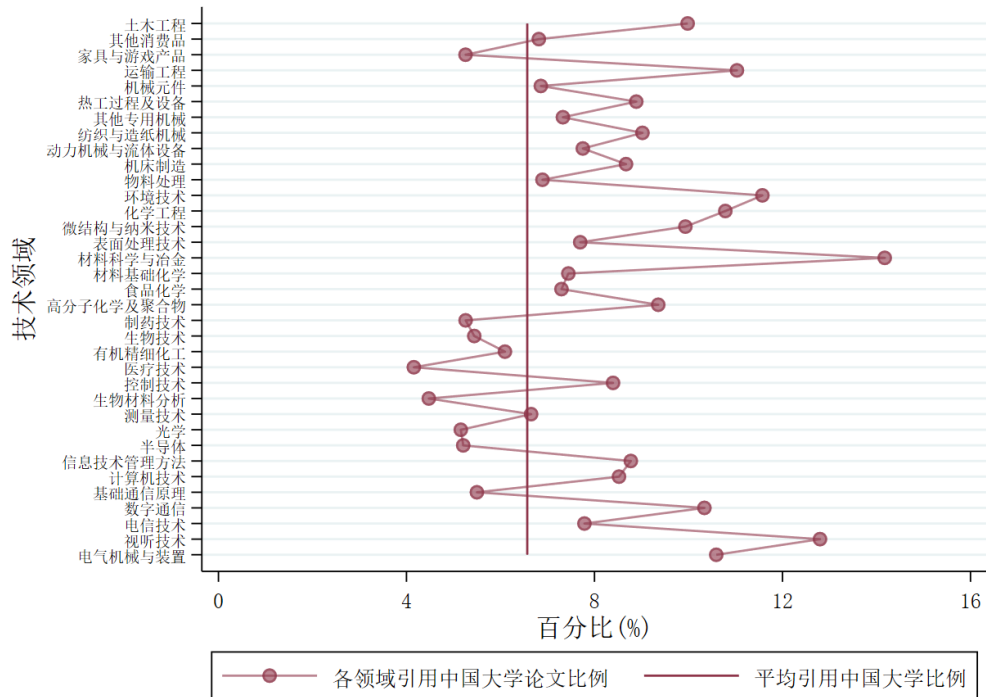


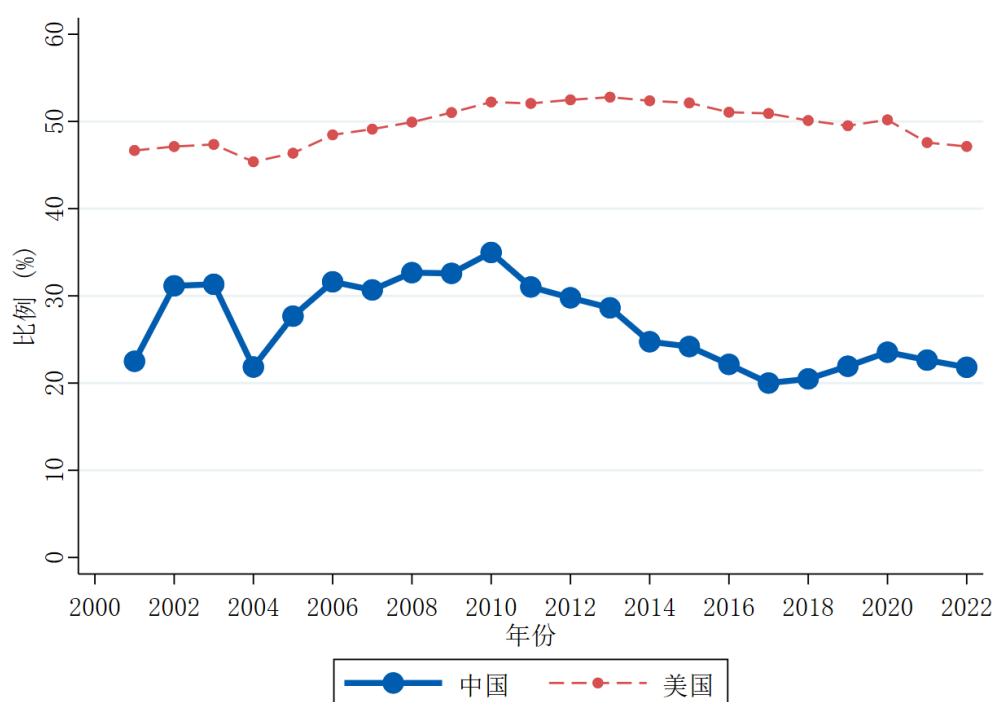
图 3 各技术领域引用中国大学论文比较

注: 根据世界知识产权组织 (WIPO) 发布的专利 IPC-行业对照表, 将专利根据 IPC 分类号分为 35 个行业类别。一个专利往往具有多个 IPC 分类号, 首先将每个 IPC 分类号映射到行业中, 然后计算每个行业类别所占比重 (使用该行业在专利 IPC 中出现次数除以总 IPC 数), 若某专利 A 行业权重为 0.8, B 行业权重为 0.2, 则该专利所有的引用记录乘以 0.8 后被归类为 A 行业, 乘以 0.2 后被归类为 B 行业。图 3 为 35 个行业所引用的论文中中国大学论文被引次数占比和在考虑行业分类时 (即所有专利) 中国大学论文被引占比 (平均值=全部授权专利引用中国大学论文的次数÷全部授权的专利引用大学论文的次数×100%, 半导体领域比例=全部半导体领域授权专利引用中国大学论文的次数÷全部半导体领域授权专利引用大学论文的次数×100%), 图为五年口径。

### (三) 中国大学需求导向基础研究的本国吸收有限、国际溢出明显

进一步地, 研究团队考察中国产业界是否有效使用 (吸收) 了中国大学产出

的需求导向基础研究成果，其结果并不乐观。对比中美两国引用学术论文的国际专利占全部国际专利的比例（图 4a），其在美国样本中保持在 50%的水平上下，且在 2001—2022 年间比较稳定；相比之下，引用至少一篇学术论文的中国国际专利占比则在 20%~35%范围波动，且在 2010 年以后存在下行趋势，在 2022 年时下降到了 21.8%。这说明中国产业界在技术创新的过程中吸收基础研究知识的能力还比较弱，而该能力是实现原创性创新的重要影响因素之一。进一步具体到两国国际专利对中国大学需求导向基础研究的吸收（图 4b），可发现美国国际专利对中国大学基础研究的吸收存在明显的随时间变化增强的趋势，美国国际专利中引用中国大学学术论文的比例在 2022 年达到了 10.2%；相比之下，中国产业界对本国大学基础研究的吸收反倒不足，在 2011 年交叉点后增幅有限，2022 年中国国际专利中引用中国大学学术论文的比例在 6.1%，仅是美国水平六成。



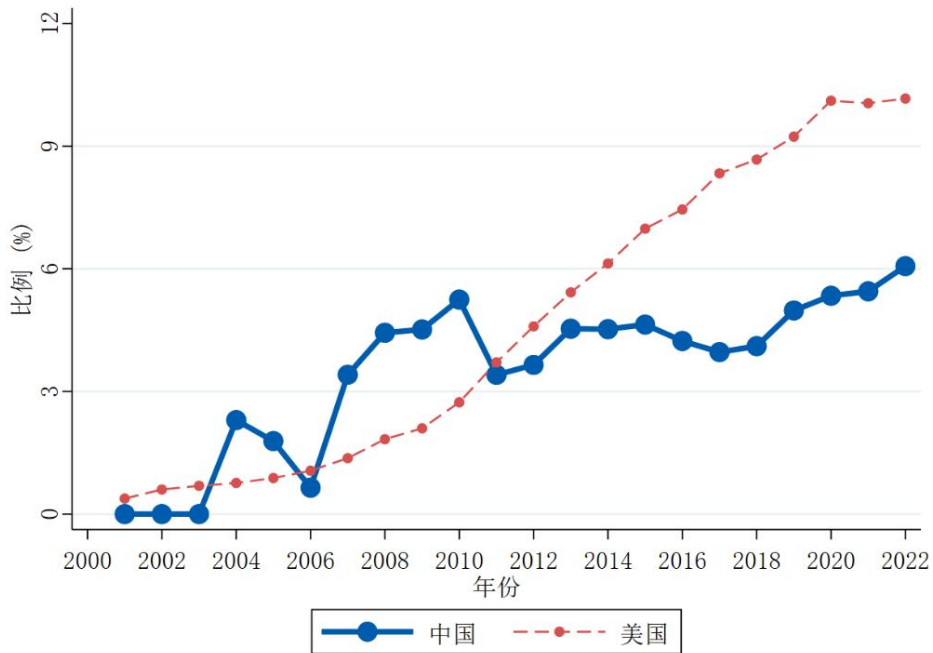


图 4a/b 中美两国国际专利引用论文和引用中国论文的比例

注：图 4a/b 横轴为专利授权年份，图 4a 纵轴为中美两国专利引用学术论文的比例，例如 2022 年中国比例=2022 年授权的引用了至少一篇论文的中国国际专利数量/2022 年授权的中国国际专利数量，图 4b 纵轴为中美两国专利引用中国大学学术论文的比例，例如 2022 年中国比例=2022 年授权的引用了至少一篇中国大学论文的中国国际专利数量/2022 年授权的中国国际专利数量，图为全周期口径。

图 4c/d 展示了一种美国产业界吸收中国大学基础研究的极端情况。图 4c 使用洛克希德、波音、诺斯罗普这三家具有代表性的美国战略性军工科技企业的专利样本；图 4d 使用美国能源部和美国国家航空航天局（NASA）下辖的 27 家美国国家实验室作为美国战略性科技力量的专利样本。可以看到，这两类美国战略性力量吸收中国大学基础研究的程度逐年上升，峰值分别达到 14.76% 与 13.93%。研究团队提示，虽然这两类美国科技力量不一定将其技术创新全部注册为专利，但前述极苛刻条件下的分析仍足以呈现中国大学面向战略需求的基础研究能力提升迅速但国际溢出效应明显的特征。当本国产业界难以及早、充分吸收本国基础研究成果时，后者就可能成为其他国家技术创新的科学基础。

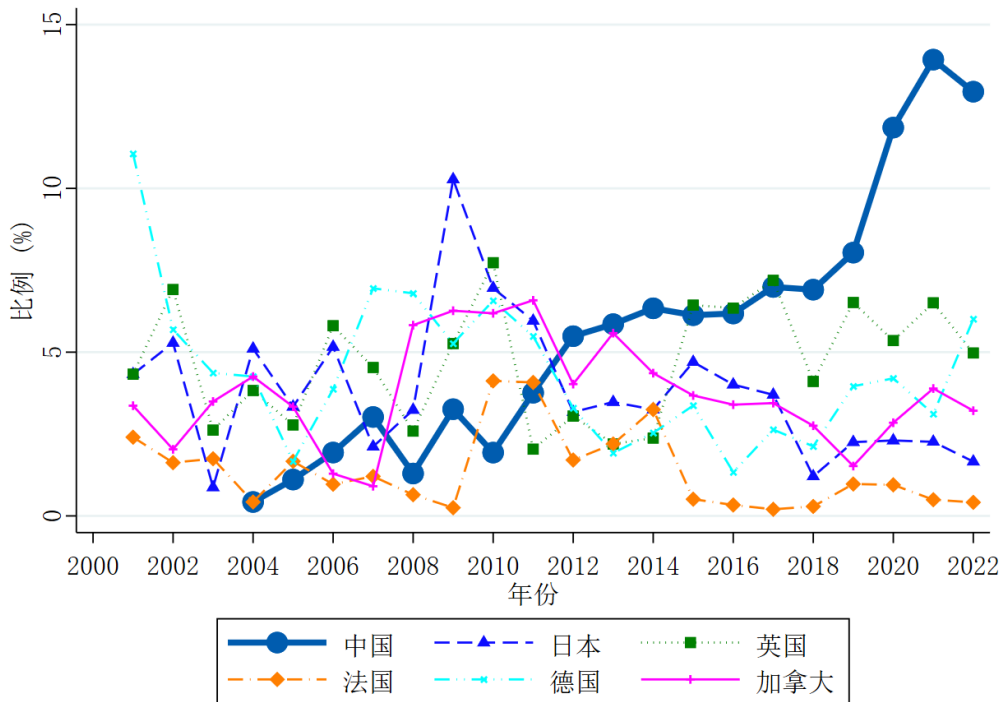
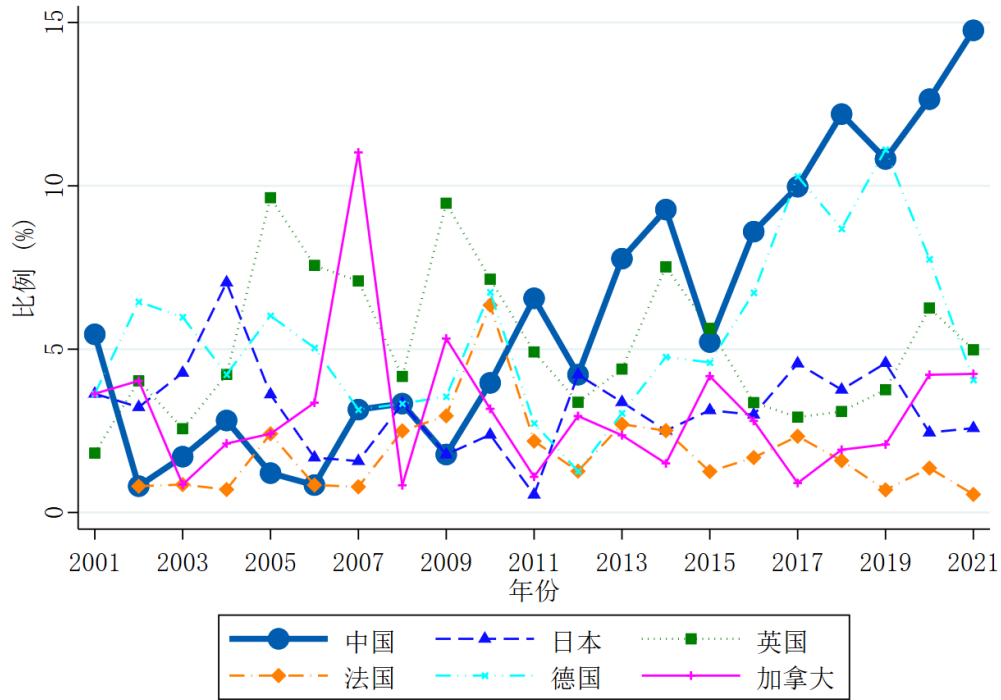


图 4c/d 美国军工科技企业与国家实验室吸收中国大学需求导向基础研究的情况

注：图 4c/d 中，为保证用以分析的专利数量，这里不再局限于 PCT 专利。图 4c/d 横轴为专利授权年份，纵轴为各国大学论文被引占比，例如 2022 年中国占比=2022 年授权专利引用中国大学论文的次数 ÷ 2022 年授权专利引用世界大学论文的次数 × 100%，图为五年口径。

### 三、政策建议

#### （一）确立针对大学需求导向基础研究的“成效观”，引导大学建立兼容需求导向基础研究的激励体系，增强研究型大学在多支国家战略科技力量中的地位

传统观点认为基础研究转化周期漫长、（未来是否有用的）不确定性较高，故应由学术共同体判断其质量，并采用高被引论文数、“自然指数”等指标作为评价抓手。近年来，新一轮科技革命和产业变革深入发展，科学研究范式发生深刻变革，基础研究转化周期明显缩短，社会公众开始期待中国大学基础研究能够直接支撑国家高水平科技自立自强，这些变化挑战了传统观点成立的前提。建议教育行政部门确立针对需求导向基础研究的“成效观”，以此引导大学建立能够兼容需求导向基础研究的激励体系，作为回答“中国大学在科技全局中起何种作用”等社会关切的理论基础。

具体来说，针对“评价谁（什么）”的问题，建议将大学视为一个生态系统而非一个固定范围，将视野延展到由大学衍生的、溢出大学围墙之外的更大范围。一个典型例子是大学科学家创办的企业，这类企业呈现了跨越大学与产业两界的场域、涵盖“从科学到技术再到产业化”的完整转化。根据研究团队的初步统计<sup>①</sup>，2020—2022年间，政府引导基金投资笔数的至少25%流向了大学科学家创办的企业<sup>②</sup>。采用生态系统的视角，将大学科学家创业企业等纳入分析大学基础研究成效的视野之中，能够更全面地回答大学科学活动的“有用性”。另一个典型例子是“有组织科研”问题。该概念要求大学资源划分与绩效评估单元从学科转向需求（如集成电路或人工智能），建立针对需求导向基础研究的“成效观”，有机会作为一个评估抓手推动“有组织科研”理念得到更彻底的贯彻。针对“如何评”问题，建议构建基于知识图谱的动态分类评价体系，科学界定在国家重大科技工程中大学（或大学科学家团队）针对“硬骨头”开展基础研究、进行核心技术攻关的贡献度；区分自由探索性与需求导向性基础研究的差异化考核评价标准，

<sup>①</sup> 相关数据库的建设仍在进行之中，研究团队根据预研究中进行的多个样本切片，得出上述比例。

<sup>②</sup> 根据研究团队对长三角国家技术创新研究院、国投创新投资管理有限公司的访谈，相关人员反映其投资笔数的约40%流向了大学科学家创业的企业。该数字涵盖了大学科学家股权由家人或学生代持的情况。

对需求导向性基础研究强调技术转化的产业适配度，在国防科技等领域实行“成果替代论文”的柔性评价机制。创新“知识图谱+同行评议”双轨制，通过技术影响力追踪与学术贡献量化分析，明确科研人员在跨学科协作、关键技术攻关中的角色权重，既破除“五唯”倾向，又确保评价客观性。

## **(二) 适应产业界吸收基础研究进行技术创新的需求，扩大博士生培养规模、优化培养模式**

“从科学到技术”转化的顺利实现既依赖于“科学端”需求导向基础研究的成果，也依赖于“技术端”吸收基础研究成果的能力。目前看，“中国大学写论文能力高于中国企业读论文能力”现象还比较明显，一个根本性的原因在于产业界所拥有的、具备基础研究能力的高质量人才数量不足。一个数据是，在美国，拥有博士学位的人员在高校与非高校领域（主要为企业）的就业比例大致为 1：1；相较之下我国该比例约为 10：1<sup>①</sup>。研究团队建议：第一，以企业能够高效利用大学基础研究为目标，进一步扩大博士生培养的规模。扩大博士生培养的规模，可以缓解产业界技术创新的人才缺口，提高产业界对需求导向基础研究的吸收能力。从高水平科技自立自强全局来看，增强产业吸收能力可以直接提高基础研究转化效率，改善教育与科技财政资金的使用效率。第二，更充分地考虑博士毕业生在企业技术创新中的角色，优化培养模式。大学需要接触企业技术创新过程，与之合作收集技术创新过程中的痛点并提炼为科学研究问题，作为一部分博士生的研究课题。大学基础研究与企业技术创新的更深结合，可以增强博士生吸收、重构基础研究知识并形成技术创新成果的能力，使其更有意愿和能力在毕业后进入企业技术创新部门工作。

## **(三) 由教育行政部门主导概念验证平台建设，帮助大学需求导向基础研究以更高的成熟度走出大学**

我国科技交易市场与早期创业投资市场的成熟均需要一个过程，在此期间基础研究的初步转化或者说“概念验证”环节尚难以交给市场机制进行，需要教育

---

<sup>①</sup>刘庆. 在高校开展行本教育，培养产业技术创新人才，中关村论坛，2025-04-01，<https://wxredian.com/art?id=c4b432ff3096a4395a9b21542159c380>.

行政部门在这一环节发挥更大作为。概念验证被用来确定一项基础研究成果能否变成技术或产品“原型（prototype）”，保留进一步转化的可能性；这一步通常不要求具有完整的功能，也不关注原型的可市场化程度，而是专注于验证核心的科学发现是否有效。建议教育行政部门主导建设概念验证平台，为大学科学家提供更好的概念验证条件，帮助大学需求导向基础研究以更高的成熟度走出大学围墙。概念验证平台建设涉及项目、资金和管理三个要素，最大难点在于管理，即组建一个既懂基础研究又懂产业逻辑的技术经理人团队。其权衡在于，如果平台在各细分技术领域均配备技术经理人，会导致团队规模太大，成本上升，反之专业性则可能不足。目前来看，依区域建立综合性概念验证平台，搭配全国性质的专业性概念验证平台是较为可行的方案。省一级的高等研究院、各大学管理的国家重点实验室均是概念验证平台可能的依托组织。

## 附录一：主要政策文件中关于需求导向基础研究概念的表述

表 1 政策文件与领导人讲话中与需求导向基础研究相关的表述

时间	来源	表述
2016 年	《“十三五”国家科技创新规划》	“坚持面向国家重大需求和世界科学前沿，坚持鼓励自由探索和目标导向相结合”“面向我国经济社会发展中的关键科学问题”“聚焦国家重大战略任务部署基础研究”
2019 年	《2019 年度国家自然科学基金项目指南》	“鼓励探索，突出原创”“聚焦前沿，独辟蹊径”“需求牵引，突破瓶颈”“共性导向，交叉融通”
2020 年	《新形势下加强基础研究若干重点举措》	“把握基础研究与应用研究日趋一体化的发展趋势”“加强重大科学目标导向、应用目标导向的基础研究项目部署”“强化目标导向，支持自由探索，突出原始创新，强化战略性前瞻性基础研究”
2020 年	习近平总书记在科学家座谈会上的讲话	“坚持面向世界科技前沿、面向经济主战场、面向国家重大需求、面向人民生命健康”
2022 年	《关于加强高校有组织科研推动高水平自立自强的若干意见》	“要在继续充分发挥好自由探索基础研究主力军和主阵地作用，持续开展高水平自由探索研究的基础上，加快变革高校科研范式和组织模式，强化有组织科研，更好服务国家安全和经济社会发展面临的现实问题和紧迫需求”
2022 年	《国家自然科学基金“十四五”发展规划》	“科学问题源于国家重大需求和经济主战场，且具有鲜明的需求导向、问题导向和目标导向特征，旨在通过解决技术瓶颈背后的核心科学问题，促使基础研究成果走向

		应用”
2023 年	习近平总书记在二十届中央政治局第三次集体学习时的讲话	“坚持目标导向和自由探索‘两条腿走路’，把世界科技前沿同国家重大战略需求和经济社会发展目标结合起来”“有组织推进战略导向的体系化基础研究、前沿导向的探索性基础研究、市场导向的应用性基础研究”
2024 年	《关于加强自然资源领域基础研究的若干举措》	“围绕国家战略、资源能源安全和经济社会高质量发展重大需求，形成自上而下的自然资源目标导向和自下而上的科学实践需求导向相结合的上下联动科学问题凝练机制”

## 附录二：对需求导向基础研究测量方法的讨论与延伸

科技成果转化遵循从基础科学到技术再到产品的变化过程。技术从基础科学和其他技术中演变而来，会吸收已有的基础科学知识并将之转化为具有更高应用价值的形态。技术发明者通过申请专利的方式将其技术公开并进行产权保护，其有义务在专利文本中自行披露或由专利审查员加入该专利所参考的已有专利和非专利文献（主要为论文）。因此，本报告通过专利对论文的引用关系来测量基础科学知识向技术的流动，以此衡量基础科学的“面向需求”程度。在后文，本报告还介绍了两种更细颗粒度的测量方案。

在实际操作中，本报告使用论文被 PCT 专利引用（后文称国际专利引用）这一范围，来衡量科学论文对技术创新的贡献。《专利合作条约》（Patent Cooperation Treaty）简称 PCT，是世界各国在巴黎公约下缔结的一份专利合作协议，目的在于对同一发明创造在多个国家申请时简化流程。虽然 PCT 的本意是对国际专利申请流程进行简化，但 PCT 提交形式本身释放了专利质量较高的信号。第一，高质量专利更可能具有全球性的商业影响，其所有者更有动机在多个国家申请保护，进而采用 PCT 形式。第二，相较于只在单一国家申请保护的专利，PCT 专利的保护费用更高，能够起到筛选专利质量的作用。第三，PCT 专利申请的来源国十分多样，中国专利占比较低（2001—2022 年期间的平均占比为 4.14%，2022 年单年为 11.84%），能够较好克服因为关系网络等因素产生的专利本国引用偏误，进而得出相对保守的统计结果。

本文采用五年引用与全周期引用这两个口径衡量需求导向基础研究，并以五年口径为主。第一，专利引用论文通常存在滞后性。第二，相对于世界主要科技

强国，中国的科技发展起步较晚，2000 年以前中国发表的论文数量远低于其他科技强国。第三，新一轮科技革命和产业变革深入发展，基础研究转化周期明显缩短，这使得有转化价值的论文发表后在较短时间内就可能被专利引用。因此，选择五年引用口径，能够更好反映发生在当下的、正在支撑新一轮科技革命的基础研究成果；而选择全周期口径，则能更好反映历史上的、累积性的基础研究能力。

本报告采用国际专利引用论文这一方法测量基础研究面向需求的程度。该方案最大的优点是所观测的现象真实发生且解释直观，论文与专利这两种知识载体能够被明确区分。该方案的缺点在于未将“从科学到技术”视为一个连续光谱（此时，应用学科论文引用基础学科论文也是后者的一种转化），也难以评价最近一段时期发表学术论文面向需求的程度（因为论文被引用需要时间）。针对此，本报告进一步介绍两种先进方法以供后续研究应对上述局限。

**基于主题共现的方案。**前文测量方法的一个隐含假定是论文代表基础研究，专利代表技术创新，即科学与技术之间存在截然性的区分。但在现实中，一部分更类似于技术的知识也会以论文形式呈现，这在工程学科或计算机科学等领域较为常见。在此情况下，就需要将“科学—技术”视作一个光谱，采用更细颗粒度的方法识别其中相对更偏向（不偏向）基础研究的子领域。共词分析与语义分析技术的结合有机会解决上述问题。共词分析法根据科学主题与技术主题中相同词汇的出现频率来刻画科学与技术间的联系，并在社区化过程后将整个领域分为若干子领域，而 SAO 语义分析则用来刻画知识演化路径。这种组合方法先将一个大的研究领域分为若干子领域，然后根据子领域之间的语义逻辑识别其上下游关系，位于网络上游的知识要素更为基础，位于网络下游的知识要素则更偏向应用（技术）。这样，每个子领域就不必被二分入基础研究或非基础研究的类别，而是拥有一个连续性的“基础研究”程度分数。

**论文被引预测方法。**专利对论文知识的吸收具有明显的时滞，因此专利对论文的引用无法准确评估当前发表的论文面向需求的程度。最新学术研究使用 LLM 模型和神经网络方法开发了一套用于预测论文被专利引用概率的方法，这使我们能够在同一时间节点比较发表于不同年份的论文的需求导向能力。对于发表于  $t$  年份的论文，选取发表于  $[t-14, t-5]$  时段的论文和  $[t-14, t-1]$  时段的专利作

为训练数据集。该方法依赖于大型语言模型和自然语言处理技术（NLP）进行训练，首先提取每篇论文的摘要，然后使用 SciBERT 对选取的训练论文的摘要进行训练，在此基础上利用训练模型预测发表于  $t$  年份的论文是否会被专利引用。对该方案的测试结果表明基于 2000—2020 年数据的 21 个模型的 AUROC 在 0.80-0.84 之间，充分验证了该训练预测模型的准确性。这种方法使我们可以对新近发表论文的需求导向程度打分，而不必等待若干年后再看实际的引用情况。

### 附录三：数据处理流程与数据库介绍

本报告的数据处理如下：根据专利号确认 PCT 专利样本，将专利号链接到 reliance on science 数据库，匹配 patents view 数据库得到专利的年份，保留专利年份在 2001—2022 年的样本，识别出这些专利所引用的论文。然后我们从 OpenAlex 网站上爬取所有论文的信息，得到了这些论文的发表年份、作者、作者隶属机构、机构国籍以及其他与论文主题、被引次数相关的信息。剔除专利与论文年份缺失以及专利年份小于论文年份的样本，得到了 15,386,853 条记录，包含 2,695,623 篇论文和 788,111 个专利。根据第一作者隶属机构的国籍来判断该论文的国家归属，有 494,806 篇论文（约 18.4%）因缺乏国籍信息而被删除，最终得到 2,200,817 篇论文与 704,121 个专利形成的 12,717,758 条引用记录。本文的重点工作是识别中国论文，其定义为第一作者所属机构在中国大陆的论文（不含中国港澳台地区）。根据机构名称对所有国别信息为中国的论文进行数据清洗，通过人工查找机构所属国别的方式更改了少量来源于中国香港和其他地区的论文的归属信息，最终我们识别出中国论文 79,443 篇，共被引 226,646 次。本文以此作为全样本数据集，在此基础上仅保留机构类型为“Education”即大学的样本得到由 1,328,011 篇论文与 596,249 个专利构成的 6,840,591 条引用记录，其中中国论文 55,697 篇，共被引 150,910 次。本文选择五年期作为主口径，因此进一步去除专利年份减论文年份大于 5 年的样本，得到 252,900 篇论文与 173,638 个专利形成的 515,279 条引用记录，其中中国论文 17,419 篇，共被引 32,668 次。

本报告所使用的数据库详细介绍：

1. **Reliance on science** 数据库：由于数据的可获得性和算法识别偏误，获取专利引用的论文数据一直以来是一个难题。最新学术研究构建的算法识别出了

全球专利所引用的非专利文献 (patent citation to science, PCS), 并将论文元数据链接到开源网站 [openalex](#) 中, 即每篇论文都有一个供唯一区分的论文 id, 同时我们可以根据专利号唯一识别专利。所有数据可以从 [reliance on science](#) 数据库中公开获取, 本文所使用的版本为 2023 年 7 月 20 日发布的第 40 版本, 我们可以从中获取 1947—2022 年美国与非美国专利引用的来自于 1800—2022 年的科学论文。本文剔除了数据库中置信度小于 3 分的样本 (全样本中, 83% 的记录置信度为 10 分)。

2. PCT 专利: 根据世界知识产权组织 (WIPO) 制定的 WIPO ST.3 标准, PCT 专利 (即由 WIPO 受理的国际专利) 的专利公开号前两位字母通常为“WO”。参照“专利公开号以 WO 开头”这一规则, 本文从 [Patstat](#) 专利库中筛选出 PCT 专利。基于欧洲专利局 (EPO) 提供的专利家族信息 (属于同一专利家族的专利在其内容上本质是同一项发明), 本研究将 PCT 专利的公开号与美国专利商标局 (USPTO) 的专利公开号进行了匹配。通过这一关联过程, 我们筛选并确认了授权于 2001—2022 年的 1,998,941 件同时在世界知识产权组织 (WIPO) 和美国专利商标局公开的发明专利。前述操作的核心目的均在于筛选一个高质量的专利集合, 以防止近年来中国专利数量激增现象导致对中国需求导向基础研究水平的高估。

3. OpenAlex 网站: 2022 年, 在继承 MAG (微软学术年谱) 的基础上, [OurResearch](#) 在 [Arcadia](#) 的资助下建立了 OpenAlex 网站, 这是一个类似于 [Goolger Scholar](#)、[Scoups](#) 的学术研究系统, 提供了海量的学术数据元, 并且免费开源, 任何用户都能直接调用其 API 获取论文数据元, 仅有每天访问 10 万次以内、每秒不超过 10 次访问的限制。对比 [Scoups](#) 的 8700 万论文、[Dimension](#) 的 13500 万论文, OpenAlex 的 24500 万论文数据量更大、涵盖范围更广。学者的研究发现 OpenAlex 相比于 [Scopus](#)、[Web of science](#)、[PubMed](#) 对论文的覆盖率更高。本报告通过调用 API 从 OpenAlex 网站中爬取了所有样本论文的信息, 包括论文发表期刊、发表年份、主题、作者、作者所属机构与机构所属国家、被引次数等我们在之后分析中会使用的变量。

4. Patents view 数据库: 该数据库由美国专利商标局 (USPTO) 官方建立, 涵盖了 1976 年以来美国专利商标局授予的所有专利信息。Patents view 数据库提

供了成块整合后的美国专利数据下载,可以直接获取美国专利商标局授予的全部专利的详细权威数据,我们从中获得了每个专利的授予年份、IPC 分类号、专利权人国籍等信息。

## 上期回顾

2025 年第 14 期（总第 279 期）

### 为数字赋能教育而筹资——美国中小学人工智能教育财政投入政策介绍

**摘要：**人工智能作为颠覆性的教育技术变革，设计配套政策须考虑教育投入原则及保障机制。基于对美国人工智能教育财政政策的考察，笔者分联邦及州两个层级分析其设计意图、拨款方及经费使用侧重。最后，简要讨论对我国的启示。

《中国教育财政》由北京大学中国教育财政科学研究所主办；旨在反映本所最新的学术科研活动；相关内容仅体现作者本人观点，并不必然代表本所的立场。

文章内容仅供参考，如需转载须事先征得本研究所同意。

本期印发：2000 份

下载网址：<http://ciefr.pku.edu.cn>

---

**主办单位：**北京大学中国教育财政科学研究所

**邮箱：**[workingpaper@ciefr.pku.edu.cn](mailto:workingpaper@ciefr.pku.edu.cn)

**责任编辑：**毕建宏

**传真：**010-6275-6183

**地 址：**北京市海淀区颐和园路 5 号

**微信公众号：**中国教育财政

北京大学教育学院大楼四层（100871）

