

科技脱钩下中国生成式 AI 安全风险识别与应对

李子浩^{1,2}, 王迎春¹

(1. 上海交通大学国际与公共事务学院, 上海 200240; 2. 上海人工智能实验室安全可信 AI 中心, 上海 200232)

摘要:构建中国生成式人工智能的安全风险要素体系,充分识别中美科技脱钩背景下中国生成式人工智能技术发展所面临的安全风险,为应对日益复杂的国际科技环境提供理论依据和实践参考。通过对基础层、技术层和应用层的分层研究,结合算力、数据和算法等生成式人工智能关键技术特征,构建中国生成式人工智能安全风险要素分析框架,并基于专家知识对风险要素进行逐一分析和评估。具体安全风险主要包括基础层限算力、限数据和限开发工具,技术层限模型算法、限计算架构和限开源平台,应用层限市场拓展和限安全门槛。建议从国家层面尽快建立统筹协调机制,构建基础层自主可控、技术层通专融合、应用层需求牵引的应对策略,以确保中国在全球人工智能技术竞争中的战略主动权。

关键词:科技脱钩;生成式人工智能;安全风险;风险应对

中图分类号:G322

文献标识码:A

近年来,中美之间的科技脱钩趋势愈发显著。无论是美国拜登政府时期就提出的“小院高墙”策略,还是特朗普政府再次上台后奉行的“全面脱钩”策略,均旨在通过构建一个小范围的“安全技术圈”,将中国排除在全球科技创新体系之外^[1]。在中美两国的全球科技格局竞争中,以多模态大模型为代表的生成式人工智能技术被视为下一代科技革命的核心,是在全球科技竞争中占据制高点的关键技术^[2]。

美国政府近年来出台了一系列科技脱钩政策,试图通过限制关键技术、人才和市场准入等手段,削弱中国在这一领域的创新能力和市场影响力^[3]。拜登政府时期签署的《芯片和科学法案》表明美国民主党和共和党在国会中已经形成科技反华共识^[4],该法案规定获得美国财政支持的实体不得在中国进行高端芯片相关投资和建设。2025年1月,美国商务部正式生效了《人工智能扩散出口管制框架》(Framework for Export Control of AI Diffusion)和《先进芯片额外尽职调查规则》(Implementation of Additional Export Controls)。这些政策以《出口管理法》(Export Administration Act)为法律基础,通过《出口管理条例》(Export Administration Regulations)、《实体清单》(Entity List)和《商业国家列表》

(Country Chart)等具体工具,对全球国家和地区实施基于 GPU 算力的分级管控。同时,美国还通过服务层面的限制强化对生成式人工智能能力的管控。例如,OpenAI 宣布禁止包括中国在内的特定国家通过 API 调用其生成式 AI 服务。2024年12月,美国司法部(DOJ)发布针对数据跨境管制的《最终规则》,限制中国等“受关注国家”使用美国公私部门数据集和语料库训练 AI 模型。2024年5月,美国众议院公布《增强关键出口海外限制国家框架法案》(Enhancing National Frameworks for Overseas Restriction of Critical Exports Act),限制美国先进生成式人工智能模型的出口;美国财政部紧跟其后发布了《关于实施对外投资行政命令的拟议规则》(Notice of Proposed Rule Making),限制美国人在中国量子信息技术以及人工智能等领域投资。这些举措表明,生成式人工智能技术发展风险已不再局限于研发本身,而是被赋予了更广泛的地缘政治意义^[5]。

国际关系的复杂性和不确定性直接影响到中国生成式人工智能领域的技术发展^[6]。科技创新与国家权力之间存在密切联系,国家间的技术封锁和竞争不仅影响技术的独立性和安全性,还会通过贸易战、制裁等手段加剧技术领域的对立^[7,8]。因此,国际局势对人工智能技术创新链条的冲击不容忽

视。根据国家软实力理论^[9] (Soft Power of the Nation), 技术创新是国家实力的重要组成部分, 而在全球化背景下, 科技脱钩可能导致国家间的技术竞争转向技术对抗, 进而影响到技术的跨国合作和全球分工。此外, 国家间的竞争和冲突会影响到科技领域的战略部署^[10], 特别是在人工智能这种具有重大军事和经济潜力的技术领域。因此, 科技脱钩不仅是经济和科技层面的竞争, 还涉及到国家安全、技术自主权和国际影响力的比拼^[11-13]。

1 构建中国生成式人工智能的安全风险要素体系

生成式人工智能技术是指通过机器学习模型生成新数据或内容的技术。这些模型通常基于深度学习算法, 如生成对抗网络 (GANs) 和变分自动编码器 (VAEs) 等。通过大规模预训练, 生成式人工智能模型可以创造出具有创意和实用性的图像、文本、语音等内容, 这对以往感知式、决策式人工智能的发展路径形成新补充, 同时更加适用于复杂度高的场景, 实现更高质量的分析与决策^[14,15]。

生成式人工智能作为一个复杂且多层次的技术领域, 其发展需要全面考虑其独特的技术特性及系统性影响^[16]。其核心技术的发展不仅依赖于数据、算法和算力等核心技术要素, 还需要拓展到更为前端的基础理论与共性技术突破^[17], 以及更为后端的场景应用探索和安全准入门槛等全方位的影响^[18]。这一全链条的技术与应用结合, 决定了生成式人工智能持续创新和扩展的能力。

生成式人工智能技术的分层分类探讨有利于有效识别技术发展中的安全隐患。本研究以国家人工智能标准化总体组指导发布的《人工智能标准化白皮书》^[19] 为分层分类基础, 参考中国互联网络信息中心发布的《生成式人工智能应用发展报告》和中国科学技术信息研究所编制的《中国人工智能大模型地图研究报告》内容, 中国生成式人工智能技术安全风险可主要从基础层、技术层和应用层三个层次进行探讨, 分别对应训练基础设施、模型性能表现以及商业化落地。通过多轮专家意见征询, 提取并完善各安全要素细化项, 构建科技脱钩背景下中国生成式人工智能技术发展安全风险的分析框架, 具体包括以下几个方面, 见表 1。

表 1 科技脱钩下中国生成式人工智能技术安全要素总结
Table 1 Summary of security factors for generative artificial intelligence technology in China under technological decoupling

要素	细化项	概念	需求特征
基础层	算力	计算资源的处理能力	高算力、低功耗、可拓展
	数据	训练语料收集与处理	高质量、多模态、大参数
	开发工具	编程程序与开发系统	先进性、稳定性、兼容性
技术层	模型算法	数据分析与预测方法	泛化性、专业性、通用性
	计算架构	硬件与系统结构设计	可优化、可拓展、可兼容
	开源平台	代码资源的共享社区	技术共享、协作开发
应用层	市场需求	全球市场的应用需求	场景挖掘、市场开拓
	安全准入	各地区安全合规要求	符合当地法规和安全标准

1.1 基础层: 生成式人工智能技术的基础设施

基础层是整个生成式人工智能技术体系的根基。其中算力, 即计算资源的处理能力, 无疑是驱动生成式人工智能技术不断前行的“引擎”。近年来, 随着自然语言处理、计算机视觉等细分领域对算力的需求呈现出爆发式的增长态势, 传统的通用芯片已经难以满足这些领域对于高算力、低功耗以及可拓展性的严苛要求。因此, 专门用于处理人工智能

应用计算的 AI 芯片逐渐成为主流, 主要包括 GPU、ASIC 和 FPGA 等类型。

数据作为生成式人工智能的“燃料”。高质量的数据能够有效提升模型对于复杂场景的理解与生成能力, 其中多模态数据的整合利用为模型的跨领域应用奠定了坚实的基础。据国家信息中心预测, 到 2025 年, 中国数据总量预计将达到 48.6ZB, 数据总量将占全球 27%, 成为世界第一数据资源大国。

然而,尽管中文数据总量大,且增幅可观,但其中可用于模型训练的高质量中文数据却相对匮乏。

在开发工具方面,国外的编程语言,如 Python、Java、C++ 等,在生态系统、工具链以及社区支持等方面展现出了显著的优势,这些优势使其在市场中占据了主导地位,从而在一定程度上形成了垄断格局。中国深度学习框架的起步也比较晚,尽管百度、华为等企业已经陆续推出了 PaddlePaddle、MindSpore 等具有自主知识产权的深度学习框架,但与当前最为主流、应用最为广泛的 TensorFlow(TF)和 PyTorch(PT)等框架相比,仍存在一定的差距。

1.2 技术层:生成式人工智能技术的核心驱动力

技术层是生成式人工智能技术实现功能的核心,不仅涵盖了算法的实现、模型的训练,还深入到了模型的优化和评估等多个环节。目前生成式人工智能大模型在算法实现上强调高泛化能力,既要能处理文本、图像、音频、视频等大规模多样性的数据,也要能够在复杂场景下提高处理任务的通用能力。模型专业性的要求则体现为算法对特定任务的性能优化,通常情况下是针对某个专业化领域的任务,比如医疗影像分析、自动驾驶等领域改变算法,提高生成效果和准确性。

计算架构的可优化性与可拓展性在模型训练中扮演着至关重要的角色。分布式计算与并行处理技术的应用,使得生成式模型在处理大规模数据和复杂计算任务时,能够展现出更高的效率和更强的灵活性。这不仅加速了模型的训练过程,还提高了资源的利用率,降低了成本。此外,计算架构的兼容性也是推动生成式人工智能技术普及与应用的关键因素。它能够确保技术适应多样化的硬件与软件环境,为技术的升级迭代提供了坚实的基础。

开源平台是生成式人工智能技术的分享与协作的重要途径。GitHub(GH)、HuggingFace(HF)等开源平台汇聚了最新的研究成果和技术资源,并且可以提供丰富的预训练模型和数据集,极大降低了技术应用门槛。通过降低二次研发的代价,迅速提高技术的创新速度和扩散性。

1.3 应用层:生成式人工智能技术的价值体现

应用层是生成式人工智能技术实际落地和价值实现的终端环节,当下市场需求日益增强且与各类应用场景深度融合,成为推动产业升级的重要驱动力。对于制造、医疗、教育、金融等细分领域,生成式人工智能技术的优势体现在不仅提高了生产效率,

还加快了传统产业的转型、颠覆和创造了产业新形态。以医疗领域为例,生成式人工智能通过深度挖掘与分析海量的医学影像数据,极大地提高了诊断的准确率与效率,从而提升医疗服务的质量与水平,为医疗行业带来了革命性的变化。

随着生成式人工智能技术的广泛应用,数据隐私保护、算法透明度和公正性等问题也日益凸显,目前已有多个国家和地区陆续出台了严格的法律法规,企业需要根据不同市场进行合规审查,按照各地的法律条文去规范技术内容,例如,欧盟强调必须确保技术处理满足《通用数据保护条例》(GDPR)中关于隐私保护的规定;美国《算法公平与透明法案》则更突出模型算法在决策中的公平性与透明性。

2 充分识别科技脱钩下中国生成式人工智能安全风险

科技脱钩下中国在生成式人工智能领域的技术发展面临诸多风险挑战,对中国技术自主创新能力提出了严峻考验,也对中国在全球科技竞争中获取战略主动权产生了深远影响。在此背景下,亟待系统识别和深入分析科技脱钩下中国生成式人工智能技术发展的风险种类。根据生成式人工智能技术发展分析要素及特征,基础层风险可分为限算力、限数据和限开发工具;技术层风险分为限模型算法、限计算架构和限开源平台;应用层风险分为限市场拓展和限安全门槛,见图1。

2.1 基础层风险:算力、数据和开发工具高度依赖

基础层受限将直接削弱中国生成式人工智能技术的自主研发能力,导致中国在核心技术上的战略被动和技术差距的进一步拉大。在算力方面,美国对中国出口的人工智能半导体产品采取了严格的“逐案审查”政策,全面限制先进 GPU 芯片和相关半导体设备的售往中国。尽管英伟达推出了“特供版”A800 和 H800 芯片,但其性能相比原版降低了约 30% - 40%,无法满足训练千亿参数级别模型的需求。同时,云服务的断供风险也在不断加剧,美国政府要求基础设施即服务(IaaS)提供商在与外国客户进行交易时须向商务部提交详细报告,旨在切断中国企业通过美国云计算服务获取计算资源的途径。

在数据方面,美国政府不仅限制了包括中国在内的 6 个“受关注国家”及相关主体获取美国政府

的数据集。生成式人工智能模型依赖于大量的高质量数据进行训练。以 GPT-4 为例,其训练数据规模据估算超过 1.8 万亿参数。如果中国企业无法获

取类似规模和质量的 dataset,将直接影响模型的生成能力和泛化性能,尤其是在医疗、金融等敏感领域的发展和应用。

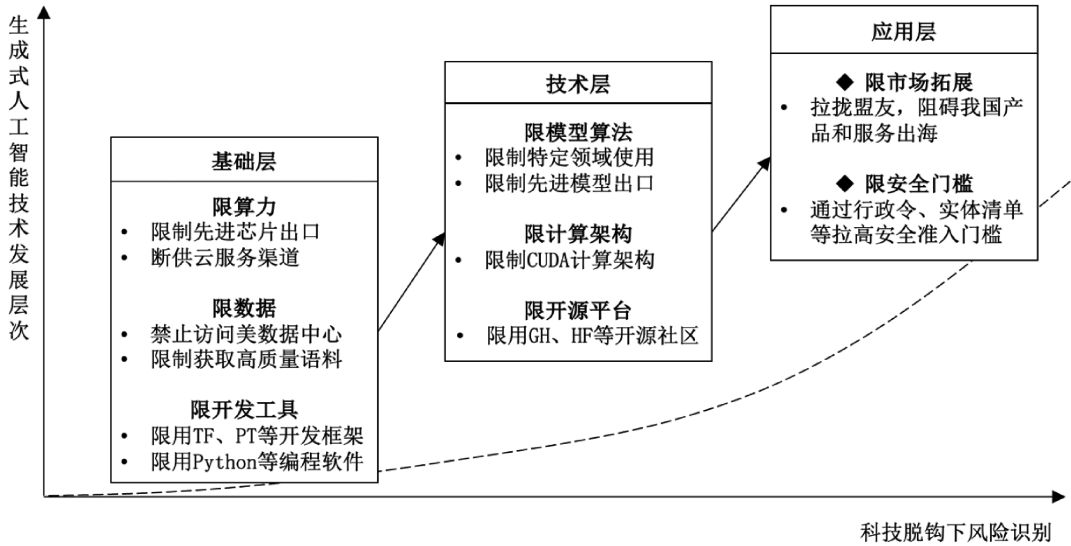


图 1 科技脱钩下中国生成式人工智能安全风险识别

Figure 1 Identification of security risks of generative artificial intelligence in China under technological decoupling

在开发工具方面,虽然目前美国尚未出台具体法案禁用人工智能开发框架,但未来通过修改出口管制法规或调整开源许可,限制特定国家或实体使用主流开发框架和编程语言的可能性依然存在。如果美国政府决定停止对这些工具的维护和版本更新,中国企业将不得不依赖自主开发的工具,而这些工具的成熟度和生态系统均远不及国际主流产品。

2.2 技术层风险:模型算法、开源平台和计算架构限制获取和使用

技术层受限将严重削弱中国在关键技术领域的竞争优势,使我们在全球科技竞争中逐渐失去话语权,并在战略性新兴领域面临长期技术风险。在模型算法方面,美国及其盟国可能会针对特定领域和行业实施严格的使用限制,甚至通过出口禁令的形式,禁止闭源模型权重的出口交易。这种限制将直接阻碍中国企业获取全球领先的算法技术,导致我们在国家安全、军事应用以及关键基础设施等关键领域的技术研发和应用受到严重制约。

在计算架构方面,CUDA 作为英伟达开发的通用并行计算架构,已成为当前深度学习领域的标准工具。目前,全球主流的深度学习框架都对 CUDA 进行了深度优化,支持 CUDA 的代码库和工具链数量远超其他计算架构。然而,英伟达在其最新发布

平台的开发者使用这一架构进行限制。由于深度学习的许多算法和模型都高度依赖 CUDA 进行加速运算,若全面转向中国本土计算架构,则需要对现有代码进行大规模重构,迁移成本极高且训练效率将大幅降低。

在开源社区方面,根据中国信通院数据,在全球活跃度排名前 100 的开源软件项目中,美国开源软件项目占比 45%。然而,随着地缘政治紧张局势的加剧,这些平台可能会逐步限制对贸易制裁国家提供服务。例如,GitHub 早在 2020 年曾对特定国家的开发者限制访问,而类似的措施可能扩展到中国,包括采取禁用受制开发者的代码存储库、删除用户账号等措施,使得中国在国际技术合作中处于孤立状态。

2.3 应用层风险:拉高安全准入门槛,设置产业生态封锁

应用层受阻将限制中国生成式人工智能技术的国际影响力和商业化路径,进而削弱全球市场份额。据麦肯锡数据,全球生成式人工智能市场整体收入预计 2032 年将达到 1.3 万亿美元。在市场拓展方面,美国及其盟友很可能通过建立严密的产业生态封锁圈,对中国人工智能产品和服务进行联合抵制和排斥,并借助政治和意识形态工具进行道德污名化,降低我国人工智能技术在国际上的名誉和认可

度,从而在更大范围内形成技术和市场的孤立效应。

在安全准入门槛方面,美国预计将通过一系列行政令、实体清单以及专门针对中国企业的特殊法律法规,设立更加严格且复杂的市场准入条件,其中可能包括繁琐的审查程序、严苛的合规要求以及高昂的税收负担等。因此,中国企业在进入国际市场时将面临巨大的法律和合规风险,不仅要应对复杂的国际法律环境,还要时刻警惕不断变化的国际政治环境可能带来的经营风险。在这种双重压力下,必然会有大量中国企业放弃欧美等高风险市场,转而专注于国内市场或寻求其他相对友好的国际市场。长期来看,这将导致中国在全球人工智能市场的份额和影响力逐渐下降,甚至可能错失引领未来科技发展的先机。

3 科技脱钩下中国生成式人工智能发展的应对策略

3.1 基础层:推动国产替代与关键资源自主可控

在全球科技竞争日益激烈的背景下,推动国产替代与关键资源的自主可控的基础层建设是确保中国在这一领域保持竞争优势的核心策略^[20]。具体而言,这一策略可以从算力基础设施建设、数据资源整合、开发工具链条的本土化等多个方面展开。

加快构建新质算力基础设施是推动生成式人工智能技术国产替代的关键举措。算力是目前中国生成式人工智能模型性能提升最大的制约因素。专家预测训练下一代生成式人工智能模型(GPT-5同级别)的训练门槛需要约10万块英伟达H100 GPU同等算力。单靠少数企业的技术突破尚不足以支撑中国在全球科技竞争中获得持续的竞争优势,要推动全国产业化的高性能计算芯片产业链的形成,不仅要实现核心芯片的自主设计与制造,还需要配套发展国产芯片的生态系统。打造异构芯片协同的软硬件生态,依托模型研发带动人工智能芯片性能和集群性能持续提升。充分发挥每种芯片的计算优势,在性能、功耗和成本之间实现最优平衡,并减少对单一芯片厂商(如英伟达)的依赖,促进芯片产业生态的自主可控和协同创新。

数据资源的整合与生态建设同样是基础层保障技术自主可控的重要环节。为应对这一挑战,中国需要加快构建自主的数据生态,建立高质量的语料库和基础科学数据集。这不仅包括对现有数据资源

的整合与优化,还需要通过多来源知识融合、细粒度知识抽取等技术手段,深入挖掘各类科学数据和科技文献,构建一个覆盖广泛的科学知识资源底座。与此同时,数据生态的建设也需要通过统一标准和平台化管理,实现数据的互通互联与高质量开放共享。截至2024年,已有包括上海、江苏、四川等十余个省级数据管理机构陆续揭牌,旨在促进数据流通,破解区域性数据无法流转的困境。

推动开发工具的本土化是应对外部风险的有效策略。建议鼓励加入国产AI开源联盟,如飞桨PaddlePaddle、昇思MindSpore等,设立企业级开发协同组,与主流国产框架开发团队双向对接,大幅提升迁移国产工具的效率。通过国内开发者社区和技术联盟,吸引更多的开发者参与到自主工具和框架的开发中来,可以有效降低对国外技术的依赖,增强中国在生成式人工智能技术领域的自主创新能力,保持技术更新的连续性和灵活性。

3.2 技术层:打造通用与专用融合的技术创新体系

构建通用与专用融合的技术创新体系^[21]是应对科技脱钩和国际技术封锁的关键策略。现阶段的生成式人工智能主要专注于更高效地构建通用泛化能力,尤其是其高阶能力,如数理、因果推理等。然而,为了更好地适应和解决现实世界中的复杂问题,需要构建一个既具有泛化性又具备专业能力的人工智能系统,将泛化性和专业性有效地结合起来。模型只有在复杂环境中能够做出决策,将复杂任务分解为更易管理的子任务,制定行动计划,并有效地协调多个智能体,才能实现群体智能的涌现。因此,在当前国际技术竞争加剧的背景下,首要任务是打造通专融合的自主基座生成式人工智能模型。具体而言,基座模型的研发应当聚焦于两个方面:一方面是在通用性上进行强化,确保模型能够在多领域、多场景下稳定运行;另一方面是在专用性上进行深度优化,针对特定行业或应用需求开发具备专业化特征的生成式模型。

软硬件适配是当前亟待解决的技术瓶颈。当前,生成式人工智能技术的发展高度依赖于高性能计算芯片和深度学习框架的协同运作。然而,由于国内在高端芯片制造和核心算法研发方面起步较晚,导致在软硬件适配上存在较大差距。建议面向未来生成式人工智能模型架构和技术路线,加强人工智能芯片与国产深度学习框架的适配工作。鼓励国产模型厂商和人工智能硬件厂商协同,形成从算

力集群设计到模型研发的深度耦合。上下游厂商通过一次适配即可深度接入算法生态,实现软硬件解耦和算力要素多样化,共建开放的软硬件适配生态。

建设自主开源平台则是实现通专融合的重要路径。通过坚持开源化路径,可以形成一个技术共享与合作的生态系统。中国在开源平台建设方面已经取得了一定进展,但要形成更为强大的技术生态,还需加速推进国产开源平台的标准化建设。制定统一的技术标准和开发规范是确保技术互通互联的关键,包括对模型算法、数据接口、硬件适配等方面的标准化规则制定。通过构建一套完整的技术标准体系,可以在全球技术生态中形成更大的影响力。

针对数据隐私泄露、算法歧视偏见等问题,亟需构建完善的安全测评体系。通过制定严格的测试和评估标准,确保生成式人工智能技术的安全性。例如,可以设立专门的安全评估机构,对大模型的训练过程、数据来源和应用场景进行全方位的安全审查,确保技术在应用中的公正性和透明性。通过这种严密的安全测评,可以有效防范可能的技术滥用和安全风险,从而为技术的广泛应用提供可靠保障。

3.3 应用层:构建以需求场景为牵引的产业创新生态

构建以需求场景为牵引的产业创新生态是推动生成式人工智能技术商业化与国际化的核心策略^[22]。面对全球复杂的市场环境和技术封锁带来的挑战,中国应积极探索生成式人工智能技术在不同应用场景中的创新路径,推动技术的落地与产业化。在这一过程中,需求场景的多样化、技术定制化、标准国际化以及产业链生态的建设,均是确保技术应用效果和商业价值最大化的关键环节。

需求场景的多样化是构建生成式人工智能产业创新生态的基础。为了有效推动技术的商业化落地,必须针对具体场景进行优化和调整,以满足不同领域的独特需求。小语种市场的开发便是一个具有巨大潜力的应用场景。当前,主流的生成式人工智能模型主要针对英语等主流语言市场进行优化,然而,全球范围内的小语种市场同样具备庞大的应用空间和商业价值。借助“一带一路”等国际合作平台,通过针对小语种进行专门的模型优化训练,可以显著提升模型在这些市场中的应用效果。这不仅有助于提高技术在全球范围内的适用性,也能帮助中国企业开拓更为多元化的海外市场,增强国际竞争力。此外,生成式人工智能技术的应用场景广泛且

复杂,不同行业对技术的需求各具特色。通过深入理解行业需求,对特定场景进行技术定制和优化,可以显著提高技术的适应性和应用效果,从而为相关产业带来显著的附加值。例如,在金融领域,生成式人工智能技术可以应用于风险评估、市场预测和自动交易系统的开发;在医疗领域,生成式人工智能可以用于精准诊断、个性化治疗方案的制定以及新药研发。

技术标准的国际化是拓展生成式人工智能全球市场的重要举措。随着生成式人工智能技术的迅速发展,各国对技术的监管和标准制定也在逐步加强。中国在推动生成式人工智能技术的国际化过程中,必须积极参与国际技术标准的制定,确保技术标准的兼容性与可扩展性。通过参与国际标准化组织和国际电信联盟(ITU)的相关工作推动生成式人工智能技术标准的国际化进程,例如,可以通过制定统一的数据接口标准、模型训练规范以及应用场景安全标准,确保技术的跨国适用性与市场拓展性。

4 相关启示

基于对科技脱钩背景下生成式人工智能技术发展风险的系统性分析,本研究认为需要从国家层面建立统筹协调机制,以有效应对这些潜在威胁。以下是三点关键启示:

一是加强国家战略科技力量的体系化创新。通过构建新型举国体制,将国家的科技资源集中在生成式人工智能等关键技术领域^[23]。在这一过程中,国家需要统筹协调各类科技资源,形成跨部门、跨行业的协同创新机制^[24]。以国家实验室、重点科研院所、龙头企业为核心,构建多层次、广覆盖的创新网络,实现基础研究与应用研究的有机衔接。其中,国家实验室要充分发挥其在人才、设备、资金等方面的优势,紧密联动全国优秀力量协同攻关重大任务,将有限的资源集中投向具有战略意义的科研项目,特别是在高性能计算芯片、深度学习算法和大规模数据集等基础技术领域实现突破。通过重点支持和布局,形成对核心技术的全方位、系统性攻坚,确保在关键领域的自主研发能力。与此同时,应加强对高水平创新人才的引进与培养,打造一支具备国际竞争力的科技创新队伍,形成强有力的技术后备力量。

二是优化国家科技创新政策环境。国家应通过制定和实施针对性强、执行力高的科技政策,为生成

式人工智能技术的发展营造良好的外部环境。首先,国家需要加强对生成式人工智能技术的政策引导,特别是在研发投入、技术推广和市场应用方面,提供更有力的政策支持。通过财政补贴、税收优惠、研发激励等手段,降低企业和科研机构创新成本,激发其研发积极性。同时,国家应简化科研项目审批程序,减少行政干预,赋予科研人员更大的自主权和灵活性。进一步完善科技成果转化机制,加速科技成果从实验室走向市场,提升技术的产业化效率。特别是在当前国际合作受限的情况下,应大力推进技术本土化,鼓励国内企业和科研机构自主开发、推广和应用生成式人工智能技术,形成一个健康、可持续发展的科技创新生态系统。

三是建立国家级技术安全与应急响应体系。这一体系的构建不仅需要全面的技术安全监测与预警机制,还必须具有对突发技术风险的快速应对能力。首先需要加强对全球科技发展动向的实时监控。通过建立健全的监测系统,及时捕捉国际科技政策变化、技术封锁措施以及供应链风险的早期信号。在生成式人工智能领域,技术封锁、供应链中断以及数据安全问题尤为关键。中国应着力构建一套多层次的风险防范机制和高度灵活的应急调配机制。制定一系列风险预案,确保在遭遇国际技术封锁和外部供应链受到破坏时,能够迅速转向国内或其他替代技术资源,迅速动员国内资源,确保生成式人工智能相关的设备、材料和技术服务的供应稳定,保障重要技术项目的连续性和稳定性。此外,还应加快建立数据安全管理体系,对数据的采集、存储、传输和使用进行全方位的安全监控。通过构建国家级技术安全与应急响应体系,中国能够在全球科技竞争和国际技术封锁的复杂局势下,确保生成式人工智能技术的稳定和可持续发展,从而维护国家的科技主权和战略安全。

参考文献:

- [1] 黄涛,惠宇吉,张胜. 政府创设需求、赛马竞争与创新领跑——美国商业船员计划的案例研究[J]. 科学学研究,2024,42(7):1527-1535,1568. Huang T, Hui Y J, Zhang S. Government demand creation, horse racing competition, and innovation leadership: A case study of the us merchant marine program[J]. Studies in Science of Science, 2024, 42(7): 1527-1535,1568.
- [2] 施锦诚,王迎春. 大模型产业创新生态系统竞争力评价研究[J]. 科学学研究,2025,43(2):382-393. Shi J C, Wang Y C. Research on the competitiveness evaluation of the innovation ecosystem in the large model industry[J]. Studies in Science of Science, 2025, 43(2): 382-393.
- [3] 陈树,李辉,西桂权,等. 中美科技竞争视阈下加强国家科技情报体系建设研究[J]. 情报理论与实践,2023,46(1):53-60,80. Chen S, Li H, Xi G Q, et al. Research on strengthening the construction of the national science and technology intelligence system in the context of Sino-US technological competition[J]. Information Studies: Theory & Application, 2023, 46(1): 53-60,80.
- [4] 刘兴华,吴保. 对华科技脱钩政策下的美国数字公司:压力与行动[J]. 中国科技论坛,2023(8):180-188. Liu X H, Wu Y. U.S. Digital companies under the technology decoupling policy with China: Pressures and actions[J]. Forum on Science and Technology in China, 2023(8): 180-188.
- [5] 史丹,聂新伟,齐飞. 数字经济全球化:技术竞争、规则博弈与中国选择[J]. 管理世界,2023,39(9):1-15. Shi D, Nie X W, Qi F. Globalization of the digital economy: Technological competition, rule games, and China's choices[J]. Management World, 2023, 39(9): 1-15.
- [6] 王燕鹏,王学昭,陈小莉,等. 基于科技政策和前沿动态的第四次工业革命关键技术和举措分析[J]. 情报学报,2022,41(1):29-37. Wang Y P, Wang X Z, Chen X L, et al. Analysis of key technologies and initiatives of the fourth industrial revolution based on science and technology policies and frontier trends[J]. Journal of the China Society for Scientific and Technical Information, 2022, 41(1): 29-37.
- [7] 郁建兴,刘宇轩,吴超. 人工智能大模型的变革与治理[J]. 中国行政管理,2023,39(4):6-13. Yu J X, Liu Y X, Wu C. Transformation and governance of large ai models[J]. Chinese Public Administration, 2023, 39(4): 6-13.
- [8] 李子浩,施锦诚,王迎春. AI技术采用对科研人员创新绩效的影响研究[J]. 科学学研究,2025,43(10):2056-2065. Li Z H, Shi J C, Wang Y C. Study on the impact of ai technology adoption on researchers' innovation performance[J]. Studies in Science of Science, 2025,43(10):2056-2065.
- [9] 王飞跃,缪青海. 人工智能驱动的科学新范式:从AI4S到智能科学[J]. 中国科学院院刊,2023,38(4):536-540. Wang F Y, Miao Q H. AI-Driven new paradigm of scientific research: From ai4s to intelli-

- gent science[J]. Bulletin of Chinese Academy of Sciences, 2023, 38(4): 536 – 540.
- [10] 张娟. 新一代生成式人工智能助力国家创新体系深度优化的价值辩证与制度建构[J]. 科技进步与对策, 2024, 41(13): 23 – 32. Zhang J. The value dialectics and institutional construction of new generation generative AI in optimizing the national innovation system[J]. Science & Technology Progress and Policy, 2024, 41(13): 23 – 32.
- [11] 余江,管开轩,李哲,等. 聚焦关键核心技术攻关强化国家科技创新体系化能力[J]. 中国科学院院刊, 2020, 35(8): 1018 – 1023. Yu J, Guan K X, Li Z, et al. Focus on key technology breakthroughs, and strengthen national scientific and technological innovation systematization capability[J]. Bulletin of Chinese Academy of Sciences, 2020, 35(8): 1018 – 1023.
- [12] 部彦君,许开轶. 重塑与介入:人工智能技术对国际权力结构的影响作用探析[J]. 世界经济与政治论坛, 2023(1): 86 – 111. Bu Y J, Xu K Y. Reshaping and intervention: An analysis of the impact of artificial intelligence technology on the international power structure[J]. World Economics and Politics Forum, 2023(1): 86 – 111.
- [13] Namkoong Y. Dependency theory: Concepts, classifications, and criticisms[J]. International Area Review, 1999, 2(1): 121 – 150.
- [14] Montresor S. Techno – globalism, techno – nationalism and technological systems: Organizing the evidence[J]. Technovation, 2001, 21(7): 399 – 412.
- [15] Nye J S. Soft power[J]. Foreign Policy, 1990 (80): 153 – 171.
- [16] Fischer M. Machiavelli's political psychology[J]. The Review of Politics, 1997, 59(4): 789 – 830.
- [17] 吴先明,高厚宾,邵福泽. 当后发企业接近技术创新的前沿:国际化的“跳板作用”[J]. 管理评论, 2018, 30(6): 40 – 54. Wu X M, Gao H B, Shao F Z. The "springboard effect" of latecomer firms approaching technological innovation frontiers: A study of internationalization[J]. Management Review, 2018, 30(6): 40 – 54.
- [18] 胡登峰,黄紫微,冯楠,等. 关键核心技术突破与国产替代路径及机制——科大讯飞智能语音技术纵向案例研究[J]. 管理世界, 2022, 38(5): 188 – 209. Hu D F, Huang Z W, Feng N, et al. Breakthroughs in key core technologies and pathways for domestic substitution: A longitudinal case study of iFlytek's intelligent speech technology[J]. Management World, 2022, 38(5): 188 – 209.
- [19] 中国电子技术标准化研究院. 人工智能标准化白皮书[R]. 北京:中国电子技术标准化研究院, 2021. China Electronics Standardization Institute. Artificial Intelligence Standardization White Paper[R]. Beijing: China Electronics Standardization Institute, 2021.
- [20] 窦永香,胥佳美,刘玮,等. 面向特定领域的国家科技评估框架研究[J]. 情报理论与实践, 2022, 45(12): 60 – 67. Dou Y X, Xu J M, Liu W, et al. Research on the national science and technology evaluation framework for specific fields[J]. Information Studies: Theory & Application, 2022, 45(12): 60 – 67.
- [21] 周伯文. 通专融合是通往 AGI 的战略路径[EB/OL]. <https://www.shlab.org.cn/news/5443948>. 2025 – 04 – 29. Zhou B W. Integration of general and specialized intelligence as a strategic path to AGI[EB/OL]. <https://www.shlab.org.cn/news/5443948>. 2025 – 04 – 29.
- [22] 陈凯华. 以需求和情景为牵引培育未来产业创新生态[J]. 人民论坛·学术前沿, 2024(12): 22 – 28. Chen K H. Cultivating future industrial innovation ecosystems driven by demand and scenarios[J]. People's Forum · Academic Frontiers, 2024(12): 22 – 28.
- [23] 张三保,陈晨,张志学. 举国体制演进如何推动关键技术升级? ——中国 3G 到 5G 标准的案例研究[J]. 经济管理, 2022, 44(9): 27 – 46. Zhang S B, Chen C, Zhang Z X. How does the evolution of the nationwide system prompt the upgrading of key technologies? Case studies of China's 3G to 5G standards[J]. Business Management Journal, 2022, 44(9): 27 – 46.
- [24] 陈光华. 高技术研究机构研发模式分类体系研究[J]. 科研管理, 2018, 39(10): 120 – 127. Chen G H. Research on the classification system of the R&D model of high – tech research institutes[J]. Science Research Management, 2018, 39(10): 120 – 127.

intensity, technology R&D intensity, scientific research quality measured by high impact publications, technology R&D breadth measured by the diversity of patent classifications, and cooperation intensity reflecting joint outputs with universities and research institutes.

The empirical analysis confirms the existence of a robust positive feedback loop between scientific research and technology R&D within the original innovation system. Firstly, the results demonstrate a significant bidirectional and time lagged causal relationship at the firm level. Specifically, the scientific research intensity of an enterprise in a given year positively predicts its technology R&D intensity in the following year, indicating that scientific accumulation effectively transfers to technological applications. Conversely, technology R&D intensity also positively drives future scientific research intensity as technological development identifies new theoretical gaps that require scientific inquiry. Second, the study reveals the mediating roles of knowledge quality and breadth within this cycle. Scientific research quality, representing theoretical rigor and authority, acts as a bridge for translating discovery into application. Meanwhile, technology R&D breadth facilitates the feedback from technology to science by expanding the diversity of the knowledge base and generating complex, cross domain research questions. Thirdly, the findings highlight the moderating role of innovation collaboration. Technology cooperation intensity reinforces the path from scientific quality to technological output, whereas research cooperation intensity strengthens the transformation of technological diversity into scientific advancement. These results confirm a knowledge complementarity effect through external partnerships.

The current study offers significant contributions to innovation management. Theoretically, it provides the first systematic empirical validation of the bidirectional positive feedback mechanism in the original innovation system, shifting focus from static linkages to dynamic sustainability. By integrating the knowledge based view, it clarifies how qualitative and structural attributes of knowledge function as catalysts for co evolution. Furthermore, it clarifies the micro level enabling role of university industry collaboration in fostering internal R&D iterations. Practically, the study advises managers to prioritize the simultaneous development of basic research and applied technology. Firms should establish dual capability systems where scientific breakthroughs serve as triggers for technology and engineering bottlenecks act as incubators for scientific exploration. Policymakers are encouraged to support collaborative innovation platforms that facilitate the seamless integration of high quality science with diversified technological applications to foster new quality productive forces.

Key words: original innovation; scientific research; technological R&D; innovation collaboration; new energy vehicles

(上接第 1128 页)

Identification and response to the security risks of generative AI in China under the technological decoupling

LI Zi - hao^{1,2}, WANG Ying - chun¹

- (1. School of International and Public Affairs, Shanghai Jiao Tong University, Shanghai 200240, China;
- 2. Secure and Trustworthy AI Center, Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China)

Abstract: This study aims to establish a framework for identifying the key security risks of generative AI in China, particularly in the context of the ongoing technological decoupling between China and the United States. It seeks to provide theoretical foundations and practical guidance for navigating the increasingly complex global technological environment. Through a layered analysis of foundational, technical, and application levels, this study constructs an analytical framework for evaluating the factors influencing the development of generative AI technology in China. Key technological characteristics such as computing power, data, and algorithms are considered, and expert knowledge is utilized to analyze and assess the identified risk factors. The identified security risks include limitations in computing power, data, and development tools at the foundational level; restrictions on model algorithms, computing architectures, and open - source platforms at the technical level; and challenges in market expansion and security thresholds at the application level. The study recommends establishing a coordinated national strategy to develop a response that ensures foundational autonomy, technical integration, and application - driven needs, thereby securing China's strategic position in global AI competition.

Key words: technological decoupling; generative AI; security risks; risk responses