·科学技术史·

大语言模型的反事实推演在科技史研究中的应用
——以“曼哈顿工程”为例

Application of Large Language Model-Based Counterfactual Reasoning in the History of Science and Technology: A Case Study of the Manhattan Project

黄振羽 /HUANG Zhenyu[1,2]　贾开阳 /JIA Kaiyang[1]
侯博文 /HOU Bowen[1,2]　马睿泽 /MA Ruize[1,2]

（1. 哈尔滨工程大学人文社会科学学院，黑龙江哈尔滨，150001；
2. 哈尔滨工程大学智能社会科学实验室，黑龙江哈尔滨，150001）
(1. School of Humanities and Social Sciences, Harbin Engineering University, Harbin, Heilongjiang, 150001;
2. Intelligent Social Science Laboratory, Harbin Engineering University, Harbin, Heilongjiang, 150001)

摘　要：本文以“曼哈顿工程”为例，探讨大语言模型在科技史研究中的反事实推演能力和价值。研究采用 DeepSeek-V3.2、文心 5.0 Preview、智谱清言 GLM-4.6 三款大语言模型进行多轮实验，发现各模型在推演结论上高度一致：“珍珠港事件未发生”将导致“曼哈顿工程”启动延迟、资源投入缩减、组织模式由军事集中转向科研分散等。结果表明，大语言模型能够模拟历史可能路径，揭示科技发展对政治语境与社会动员的深刻依赖，有望为科技史研究提供系统化、可验证的新方法。

关键词：大语言模型　反事实推演　科技史　曼哈顿工程　STS

Abstract: By taking the “Manhattan Project” as a case study, this paper explores the application of large language models (LLMs) for counterfactual reasoning in the history of science and technology. Through multiple rounds of experiments using DeepSeek-V3.2, ERNIE Bot 5.0 Preview, and GLM-4.6, a strong consensus has been observed: the absence of “Attack on Pearl Harbor” would have led to a delayed initiation of the Manhattan Project, reduced resource investment, and a shift from a military-centralized to an academically decentralized organization. The findings demonstrate that LLMs can effectively simulate alternative historical pathways, revealing the profound dependence of technological development on political context and social mobilization, thus offering a systematic and verifiable new approach for research of history of science and technology.

Key Words: Large language models; Counterfactual reasoning; History of science and technology; Manhattan Project; STS

中图分类号：N09; TP18　DOI: 10.15994/j.1000–0763.2026.12.014　CSTR: 32281.14.jdn.2026.12.014

基金项目：国家社会科学基金一般项目“国家战略科技市场的制度机理与优化策略研究”（项目编号：24BSH077）。
收稿日期：2025 年 11 月 25 日
作者简介：黄振羽（1986–）男，广东化州人，哈尔滨工程大学人文社会科学学院教授，研究方向为 STS、智能社会治理。Email: huangzy@hrbeu.edu.cn
贾开阳（2000–）男，河南开封人，哈尔滨工程大学人文社会科学学院博士研究生，研究方向为智能社会治理。Email: 690154384@hrbeu.edu.cn
侯博文（1989–）男，吉林蛟河人，哈尔滨工程大学人文社会科学学院副教授，研究方向为智能社会治理、军事社会学。Email: houbw_1061610108@hrbeu.edu.cn
马睿泽（1987–）男，山东泰安人，哈尔滨工程大学人文社会科学学院博士后，研究方向为计算社会学、智能社会治理。Email: maruize@hrbeu.edu.cn

引　言

科技史研究中存在着“辉格史观”，常以当下“胜利者”视角和目的论方式解读历史，把复杂进程简化为通向必然结果的单一路径，遮蔽了历史的复杂性、偶然性与多元可能。在描述重大科技变革时，这种决定论色彩尤为明显。历史的因果结构并非简单的线性决定链条，而是由大量相互关联的事件与条件编织成的复杂网络。[1]因此，提出“如果……会怎样”这类反事实问题并非无谓臆想，而是有助于揭示出潜在的可能性。

在科技史研究中运用反事实分析具有一定的理论与实践价值。理论层面，推演未曾发生但可能发生的历史情景，为历史决定论带来挑战；实践层面，反事实分析能够深化对特定历史事件的理解。以第二次世界大战期间的“曼哈顿工程”为例，关于其如何成功研制出原子弹和推动大科学时代来临的历史叙事已颇为丰富。[2]然而，对于特定政治事件、组织决策或个人行为在其中扮演的角色，传统叙事方式则难以给出超越既定事实的解释。对此，反事实分析能提供独特洞见。

然而，针对科技史的传统反事实分析高度依赖研究者的专业知识与主观构想，且推演结果的连贯性与合理性受限于个体的知识和能力边界。近年来，大语言模型（Large language models, LLM）的迅猛发展为突破这一瓶颈带来了新机遇。LLM拥有基于海量涵盖历史、科学、政治、文化等多维度训练数据构建的知识基底，以及日益强大的逻辑推理与情境模拟能力。这使得LLM能够超越人类个体局限，模拟复杂的反事实历史情境。基于此，本文以“曼哈顿工程”为核心案例，探讨基于LLM的反事实推演在科技史研究中的应用，以期为科技史、STS乃至科技战略分析提供一种新方法。

一、约束框架与研究方法

1. 三重约束框架

反事实推演在科技史研究中并不鲜见。弗格森（Niall Ferguson）就曾系统论证过反事实分析的价值，指出其能打破历史决定论的桎梏，通过构建“可能的世界”来检验历史因果关系。[3]霍桑（Geoffrey Hawthorn）阐明了反事实推演需遵循仅改变关键变量的“最小改写”原则，通过保持历史背景的整体稳定来确保推演的合理性。[4]福格尔（Robert W. Fogel）对铁路的经济影响研究开创了量化反事实的先河，尽管其方法的争议性较大，但确实激发了学界对历史可能性边界的反思。[5]泰特洛克（Philip E. Tetlock）等把反事实推演应用到政治史研究，论证了关键节点上的偶然决策对宏观历史走向的影响。[6]然而，传统科技史中的反事实分析多局限于零散的“思想实验”——如讨论“若爱因斯坦未致信罗斯福，曼哈顿工程会否延迟”[7]——这类研究受到研究者个体能力和知识边界的限制，很难系统开展多种可能性的探讨。卡特赖特（Nancy Cartwright）曾提醒过反事实分析的潜在风险，即过度依赖叙事连贯性而忽略历史证据的碎片化事实。[8]

LLM在推理能力上的突破为反事实推演开辟了新路径。布朗（Tom Brown）指出，通过无监督学习，LLM已经掌握了语言中隐含的常识与因果模式，从而能进行“if-then”式的条件推演。[9]魏（Jason Wei）等提出的“思维链”技术表明，LLM能够通过多步推理生成连贯的逻辑链条，提升在复杂任务中的表现。[10]此外，LLM还具备一定的情境模拟能力，能够动态维持反事实世界的状态——例如在历史推演中保持政治和技术条件的内在一致性。[11],[12]然而，将LLM应用于历史研究的挑战依然不少。例如，LLM存在所谓的“幻觉”问题，有可能生成与史实相悖的内容；[13]训练数据本身潜藏的历史偏见也可能被LLM复刻和延续。[14]

针对上述问题，本文尝试构建一个三重约束框架。其一，“最小改写”原则源自反事实推演的哲学基础，要求只改变关键历史节点而尽可能保持其他宏观条件不变，避免推演成为某种任意臆想；[15]其二，“逻辑一致性”约束强调反事实叙事须符合当时的社会、技术与政

治背景；[3]其三，“多路径推演”策略借鉴了复杂系统理论的多模型思维，以探究历史走向的多种可能性。[16]此外，这一框架也吸收了历史方法论中的“对比案例”思想，[17]注重可验证性。

2. 研究方法

本文的研究方法将从可行性分析、方法设计及结果评估三个层面展开论述，确保推理过程既尊重历史复杂性，又发挥LLM的生成性能力。

利用LLM进行反事实推演的可行性在于：第一，模型在超大规模文本数据上自监督训练，形成跨领域知识表征，可随上下文提示模拟不同时代的语言风格与认知模式；[9]第二，模型具备情境建模与逻辑链推演能力，可借助“思维链”等实现多步条件推理，并维持变量间因果关系；[10]第三，模型的生成性与灵活性支持多路径探索，能快速生成多个反事实分支以描绘历史可能性空间。[11]

提示词是方法设计的关键，需包含四要素的结构化模板：历史背景、待改变的转折点、不变约束条件、具体推理问题。本文设计的提示词为：“你的身份是一名科技社会学研究者。现在需要你针对二战时期的曼哈顿工程进行反事实推演：假设珍珠港事件没有发生，但欧洲战场的局势和爱因斯坦致信罗斯福等科学建议依然存在，请基于1941–1945年的全球政治、军事和科技背景，推演美国曼哈顿工程可能的发展路径。请重点分析：（1）曼哈顿工程启动的政治动因和时机；（2）资源投入的规模和速度；（3）最终的组织模式会有什么不同？”

对于结果评估，本文主要使用多款不同的LLM对同一提示词进行多轮的回答与比较来实现。本研究采用国内主流的三款前沿LLM，分别是DeepSeek-V3.2、文心5.0 Preview和智谱清言GLM-4.6，每个模型分别进行3轮独立实验，步骤为：（1）新建对话；（2）发送提示词；（3）记录完整响应；（4）清空会话；（5）重复3次；（6）切换模型重复。

二、以“曼哈顿工程”为例的反事实推演

“曼哈顿工程”案例具有代表性，牵涉科学、政治、军事与经济等多维交互，且史料较完整，便于从政治动因与时机、资源投入规模与速度、组织模式等三个维度建立分析基线（见表1）。主流叙事认为其是1942年在罗斯福支持下启动，集结顶尖科学家，采取集中化、军事化组织，并于1945年成功试爆原子弹，[7]因而被视为“大科学”开端。[2]传统史学多将“珍珠港事件”视为关键节点，而反事实推演允许剥离这一变量，以评估其影响。

1. 基于DeepSeek-V3.2的反事实推演

通过基于DeepSeek-V3.2三轮的实验可以发现，推演结果在总体方向上高度一致（如表2所示）。通过模型推演可以看出，在“珍珠港

表1 “曼哈顿工程”真实历史基线表

分析维度	真实历史
政治动因与启动时机	直接动因：珍珠港事件引发美国全面参战，产生紧迫的战争需求； 时机：1942年8月迅速启动，作为最高优先级战争项目； 政治环境：举国一致支持，国会迅速批准拨款。
资源投入规模与速度	资金：约20亿美元（相当于2024年300多亿美元）； 人力：超过13万人参与，军方强力动员； 速度：“冲刺模式”，3年内从启动到首次核试验； 设施：大规模集中建设（橡树岭、汉福德等）。
组织模式	领导结构：陆军工程兵团主导（格罗夫斯将军），高度集中； 合作模式：政府–军方–工业–学术强力整合； 决策方式：军方主导，命令链明确； 保密性：极端保密，严格信息控制； 文化：“军事任务”导向，强调实用产出。

事件”未发生的假设下，曼哈顿工程的启动时机将延迟至1943年甚至更晚。这种启动时间的延迟并非偶然，而是由于政治动因的转变。在反事实设定下，曼哈顿工程的直接驱动力从“战时紧急需求”转向了由科学精英主导的长期战略游说。此外，在缺乏直接战争刺激的背景下，美国国内强大的孤立主义情绪和财政保守主义成为主要阻力，促使决策过程从“战时特权下的冲刺”转变为基于情报评估和科学报告的渐进式决策。这也表明，重大科技项目的启动窗口不仅取决于技术储备，更深刻依赖于特定的政治动员环境。

在资源投入规模与速度方面，推演结果显示核计划的推进模式发生了明显的转型。由于失去了战时背景下的绝对优先地位，项目的资源投入规模大幅缩减，预计总投资将由真实历史的20亿美元降至5–10亿美元水平。投入模式也发生了根本性变革，由真实历史中为了抢时间而不计成本、多线并行的同步开发，转向了技术路线分阶段验证、先试点后扩产的循序渐进的开发逻辑。这种更加稳健的研究模式导致关键工程节点（如铀浓缩设施的大规模建设）出现明显的后移，使得原子弹的问世时间也随之推迟至1946年。这一推演结果也说明了，大科学工程的推进速度实质上是行政意志对社会资源进行压缩与集中的结果。

在组织模式与治理结构方面，推演结果则呈现出明显的“去军事化”的趋势。不同于真实历史中由陆军工程兵团主导的高度集中、军事化管理模式，反事实情境下的组织架构则更倾向于由国家科学院或文职机构牵头的“学术–政府协同”模式。在这种结构下，科学家们拥有更高的决策自主权和话语权，管理氛围更接近于战后国家实验室的学术共同体，而非严密的军事单位。虽然这种模式可能会促进国际科研协作，但由于缺乏强有力的军事化指挥体系，因此，该项目在跨部门协调中会面临更多的制约与沟通协调成本，导致执行效率在常规社会环境下出现下降。

总的来看，DeepSeek-V3.2在三轮实验中展现出极高的一致性，推演结论几乎没有出现大的偏差。无论是对项目启动时机的判断，还是对资源规模和组织架构的构想，模型在三次独立测试中给出的答案都高度契合，始终紧扣着“非战争环境下项目进度必然放缓”这一核心逻辑。

2. 基于文心5.0 Preview的反事实推演

在政治动因与启动时机方面，文心5.0 Preview的三轮推演都存在“战略威慑”导向。在失去珍珠港事件的直接刺激后，核计划的性质发生了根本性变化：由具备实战用途的武器系统转向为旨在打破技术垄断的战略制衡手段。在这一背景下，模型将“英国MAUD委员会”报告及1942年芝加哥大学的人工受控核链式反应的实验实现视为关键的技术与政策催化剂。这种转变导致其在政治决策中的优先级下降，行政部门必须以“国防科研”名义通过长期的政治博弈来对冲国内孤立主义阻力，从而面对复杂的国会拨款程序。

在资源投入规模与速度方面，由于必须与

表2 基于DeepSeek-V3.2的三轮实验结果

推演轮次	政治动因与启动时机	资源投入规模与速度	组织模式与治理结构
第一轮	模型判断项目不具备战时紧迫性，启动时间被推迟至1943年以后，决策节奏明显放缓。	资源投入显著低于真实历史，采取循序渐进的配置方式，以控制技术风险。	组织结构呈现去军事化趋势，科研机构在项目中占据主导地位。
第二轮	启动时间判断与第一轮保持一致，仍强调技术与政策条件尚未成熟。	投入规模判断基本不变，强调阶段性评估后再追加资源。	组织模式未出现方向性变化，继续强调科研主导与行政协调。
第三轮	启动延迟判断在第三轮中保持稳定，未出现提前启动的推演结果。	资源投入判断与前两轮一致，未出现大规模集中投入情形。	组织结构判断稳定，未出现高度军事化或集中化管理模式。

坦克、飞机等常规军工项目竞争资源，核计划的资金规模被大幅压缩（预计降至历史水平的1/3至1/2），资源获取方式也需遵循常规财政逻辑而非战时动员逻辑。在人力资源配置上，实验数据预估科研团队规模将较真实历史缩减约30%。这也促使研发策略由多条技术路径的并行开发，转向了对特定技术方案的选择性开发。例如，通过集中资源优先保障铀浓缩路径，相应地延缓了钚弹的开发进程。这种调整导致汉福德等核心工厂建设全面滞后，最终使得核武器无法在1945年前问世。

最后，在组织模式与治理结构方面，可以发现在权力分配与协作机制上存在着“学术–政府协同”的特点。权力核心由军方统筹转向了以科学研究发展局（Office of Scientific Research and Development, OSRD）或国家科学院为首的文官系统。在这一背景下，布什（Vannevar Bush）或费米（Enrico Fermi）等学者可能出任总负责人，而奥本海默的角色则更偏向“首席科学顾问”。科学家在维持更高决策权的同时，也面临着跨部门协调困难带来的效率下降问题。这种模式虽然在1944年前后促进了放射性同位素等民用领域的早期探索，但由于保密级别的降低与军事执行力的弱化，使该项目从一个国家工程转变为了大型的科研项目。

表3可以看出，文心5.0 Preview的三轮推演结果也基本一致。模型通过引入具体的资金缩减数值、人力规模变化，并结合OSRD与“MAUD委员会”等历史线索，将反事实逻辑与具体的物质资源约束紧密联系。

3. 基于GLM-4.6实验结果的分析

在政治动因与启动时机方面，三轮实验的结果一致表明，在没有珍珠港事件冲击的假设下，核计划的启动不再是行政意志的产物，而是演变为了一个充满政治不确定性且依赖于技术论证的结果。由于缺失了战时特殊环境赋予政府的超常规的行政动员权限，罗斯福政府必须在应对国内孤立主义势力质疑的同时，通过持续的情报研判升级与长期性的政治引导，并在复杂的立法程序中寻求国会对高额研发预算的授权。在这种决策环境的改变下，1943年成为一个充满政策变数的关键时间节点，并直接导致项目的推进节奏从真实历史中不计代价的动员模式，回归到了遵循常规治理流程的国家重大科技工程研发逻辑。

在资源投入规模与速度方面，可以发现在资源获取上有着资源配置受到“竞争性挤压”的特点。在本研究的假设下，核计划在和平时期丧失了最高优先级，因此工程必须在严苛的财政审查、联邦采购流程及劳工法规约束下推进。此外，资源规模的大幅缩减不仅削弱了工业产能的扩张，更迫使研发策略从原有的多条技术路径并行推进，转向对少数高可行性方案

表3 基于文心5.0 Preview的三轮实验过程结果

推演轮次	政治动因与启动时机	资源投入规模与速度	组织模式与治理结构
第一轮	核心动因转向“防止德国核垄断”的战略威慑；启动时机延迟至1943年前后，政策逻辑由“战争应急”转为常规“国防科研”。	资金降至10–15亿美元，依赖国会专项拨款而非战时紧急拨款；人力缩减30%–40%，首次核试验推迟至1946年。	转向“科研机构主导+政府协调”模式，由国家科学院等牵头；军事色彩减弱，奥本海默角色偏向“首席科学顾问”。
第二轮	强调国内孤立主义情绪对决策的影响，需长期舆论引导；启动时间推迟至1943年甚至更晚，项目长期停留在实验室研究的阶段。	投入规模仅为历史的1/3-1/2，更多依赖私人资本与高校实验室；缺乏大规模工厂建设，研发周期显著延长。	主导权由OSRD掌握，军方紧迫感不足；倾向于“独立研发”而非国际协作，部分民用研究提前公开。
第三轮	延续“铀委员会”模式进行防御性科研，优先级低于常规军工生产；因缺乏“战时特权”，需通过国会特别拨款推动。	初期拨款大幅收缩，动员规模仅3–5万人；关键设施建设推迟至1944年后，判断核计划难以在1945年前完成武器化。	初期由OSRD主导而非陆军接管；科学家保留决策权导致项目效率降低，组织形态从“军事化”转向“松散科研主导”。

的推进，这也在一定程度上增加了项目失败的风险。由于这种制度性制约导致的工程滞后，使得原子弹无法在二战结束前完成武器化，从而在逻辑上推导出了另一种战争演进路径。

在组织模式与治理结构方面，三轮推演均体现出管理体制的转型趋势：由最初的高度集中的军事化领导，逐步转向由文职机构与学术团体共同构成的协作体系。推演结果一致显示，权力重心由军方转移至文职官员与科学家组成的协调机构，其组织形态表现出类似于战后国家实验室的协作特征。在这种架构下，科学家获得了更大的决策自主权，且伦理与学术论证被纳入决策环节，但同时也产生了明显的程序性成本与效率损耗。这种演变逻辑说明，在非极端战争环境下，大科学项目必须适应更加复杂的制度治理环境。

由表4可见，GLM-4.6的三轮推演更侧重于捕捉复杂的社会过程。模型在三次实验中始终紧扣法律程序、伦理辩论以及公众舆论等制度性阻力，展现了科技项目如何从极端的军事化状态回归到常规的社会治理逻辑。

三、讨　论

1. 单个模型的反事实推演分析结果比较

前文已分别呈现DeepSeek-V3.2、文心5.0 Preview与GLM-4.6在同一反事实设定下的推演结果，本节不再重复共同结论，而是从政治动因与时机、资源投入与速度、组织模式与治理结构三个维度比较三模型的差异化侧重，以识别其机制重心与敏感条件。如图1所示，三种模型在保持核心共识的同时，展现出各自独特的分析风格和理论贡献，形成了有价值的互补关系。

首先，在政治动因与启动时机方面，三款模型展现了不同的解释路径。

（1）DeepSeek-V3.2侧重于“内外因结合与时机捕捉”，认为项目的启动既离不开科学界精英在内部的持续推动，也受到外部国际局势突发变动的影响；只有当内部的准备和外部的触发条件同时具备，即开启了一个特定的“政策机会窗口”时，决策层才会最终拍板启动项目。

（2）文心5.0Preview则更依赖“历史线索与战略定位”。它将非战争背景下的核计划看作一种为了打破潜在技术垄断、维持战略平衡的工具。该模型更加注重挖掘具体的历史证据，详细论证了诸如英国MAUD委员会报告这样的关键文件，是如何在当时不紧迫的环境下起到了加速决策的关键作用。

（3）相比之下，GLM-4.6更关注“多方博弈与共识达成”的过程。该模型强调启动决策绝非一蹴而就，而是科学家、政治家和情报机构等多方力量在长期的舆论动员和反复谈判中

表4 基于GLM-4.6的三轮实验结果

推演轮次	政治动因与启动时机	资源投入规模与速度	组织模式与治理结构
第一轮	对纳粹核威胁的恐惧不足以催生总动员，启动延迟至1943年中后期；面临孤立主义与财政保守主义的双重政治阻力。	投入规模缩减30%–50%，陷入常规联邦采购与劳工法规限制；项目由“冲刺”变为“马拉松”，二战结束前难以问世。	转向“政治家–科学家”联盟，由文职管理者取代军方领导；组织形态接近战后国家实验室系统的早期雏形。
第二轮	启动过程充满犹豫与博弈，由小规模研究计划渐进升级；1943年成为决策关键窗口，项目可能升级为“国家战略工程”。	预算规模仅为历史的1/3–1/2，需与其他军备项目竞争资源；关键设施建设周期翻倍，进度显著放缓。	成立跨部门“原子能研究委员会”，军方角色被限定在安保与后勤；国会监督委员会权力增大，项目透明度相对更高。
第三轮	以“战略焦虑”为驱动，依赖爱因斯坦信件与情报作为催化剂；分阶段渐进启动，国会批准大规模预算的过程艰难迟缓。	资源获取遵循“和平时期的逻辑”，面临法律和程序阻碍；至1945年可能仍处于工程验证阶段，进度比真实历史慢2–3年。	由高级别文官委员会领导，科学家拥有更大自主权与话语权；伦理辩论前置化，“学术争鸣”氛围取代了军事命令链。



图1 各模型推演对比

逐步妥协的结果。这种解释突出了决策过程是如何受制于复杂的政治环境，并随着环境变化而缓慢推进的特点。

其次，在资源投入规模与速度方面，各模型的差异主要集中在对“效率损耗”机制的理解上。

（1）DeepSeek-V3.2侧重于“研发模式的转变与量化估算”。该模型提供了具体的预算预测，并指出进度滞后的核心原因在于开发逻辑的变化。通过该模型的推演来看，核计划从战时不计成本、多条技术路径同时推进的同步模式，转变为先验证、后扩产的循序渐进模式。这种研发顺序的调整，也被视为导致项目关键节点向后推移的主要变量。

（2）文心 5.0 Preview侧重于“部门间的资源争夺与挤压”。它将眼光投向了工业产能的分配问题。该模型认为，在非战争状态下，核计划不再拥有绝对的优先权，必须与飞机、坦克等常规军工项目争夺有限的钢材、电力和技术劳动力。正是这种来自其他军事部门的资源挤压，限制了核设施的建设规模，从而拖慢了整体进度。

（3）GLM-4.6则侧重于“管理程序与制度性成本”。它从社会治理的角度出发，认为在常规环境下，项目必须遵守复杂的行政规则。例如，物资采购需要漫长的招标流程，合同签署涉及多方商业谈判，人员雇佣也必须符合劳工法规。这些在战时可以被特事特办绕过的问题，在这一设定下则变成了必须付出的时间成本，因此也成为了导致研发周期延长2–3年的关键因素。

最后，在组织模式与治理结构方面，三者对管理体制转型的侧重点各异。

（1）DeepSeek-V3.2侧重于“协作形态的去中心化与协调成本”。该模型描绘了一种不同于军方管理的网络化协作模式。它认为，在失去了军事命令链的强制约束后，项目演变成了一个由多个研究机构构成的松散联盟。虽然这种模式增加了灵活性，但由于缺乏统一的指挥核心，跨部门之间的沟通与协调变得更加困难。

（2）文心 5.0 Preview更侧重于“文官建制化与科学家的决策地位”。该模型提出了一个非常具体的行政方案，即将项目置于OSRD等文职机构的主导之下。它强调，在非战争环境下，科学家不再仅仅是技术的执行者，而是更多地承担起“科学顾问”的角色，拥有了更高的学术自主权和决策参与度。这种转变体现了科研逻辑对军事逻辑的替代。

（3）GLM-4.6则侧重于“外部治理规范对组织的塑造”。它将核计划放在一个更广阔的社会治理框架中进行考量。该模型认为，常规时期的组织形态必须服从于外部规则，包括国会的预算监督、社会层面的伦理辩论以及更加透明但也更复杂的保密分级制度。这些外部约束机制共同塑造了项目的运行轨迹，使得项目管理必须在透明度与效率之间进行艰难的平衡。

2. 模型间的反事实推演分析结果比较

三种模型在核心推演结论上展现出高度一

致性，这种跨模型共识极大地增强了反事实推演的可信度。所有模型一致认为，“珍珠港事件”的缺失将导致“曼哈顿工程”发生根本性转型。从历史上“赢得战争”的紧急军事项目转变为“战略威慑”驱动的长期科研计划。这种高度一致性不仅验证了利用LLM进行反事实推演的可靠性，更重要的是揭示了科技发展路径对政治语境的深刻依赖。

如表5所示，三种模型在保持核心共识的同时，展现出各自独特的分析风格和理论贡献，形成了有价值的互补关系。DeepSeek-V3.2以其结构化的比较分析方法见长，擅长揭示“历史偶然性”和“关键节点”对科技项目的塑造作用；文心5.0 Preview体现出量化分析特色，提供了具体的资金数额、人力缩减比例等详细估计；GLM-4.6创造了“地缘战略威慑”“科学保险”等概念工具，从社会建构视角深入分析了项目合法性在不同社会语境中的建构过程。

结　论

通过系统运用三种LLM对“曼哈顿工程”进行反事实推演，实证验证了该方法在科技史研究中的可行性与独特价值。三种模型对“珍珠港事件未发生”这一反事实情境的推演达成共识，不仅强化了推演结论的稳健性，更是从跨模型比较的角度确证了STS的一个核心命题：重大科技项目的路径、规模与形态并非由技术逻辑单向决定，而是受到特定政治语境、资源动员模式与社会建构过程的深刻塑造。[18]，[19]在当今新一轮科技革命背景下，理解科技与社会因素的复杂互动，对于制定有效的科技创新政策具有重要参考价值。

从研究方法看，基于LLM的反事实推演不仅提供了关于“曼哈顿工程”这一历史案例的新洞见，更展示了一条人机协作的科技史研究新路径和LLM的社会科学研究潜力。当然，本研究也存在一定的局限：一方面，这种反事实推演方式，本质上仍受制于LLM训练数据中隐含的历史叙事框架；另一方面，反事实推演结果的评估尚缺乏客观的量化标准，提示词设计虽力求严谨，但仍可能无形中引导了模型的推理方向。未来可在三方面深化研究：一是将基于LLM的反事实推演方法应用到更多的科技史事件分析，以检验方法的普适性；二是构建更精细的提示词框架，融入更多的制度细节与约束条件，提升推演的历史贴合度；三是将LLM与因果推断模型相结合，打造一个能够量化分析的反事实推演系统。

［参 考 文 献］

[1] Ermakoff, I. 'Causality and History: Modes of Causal Investigation in Historical Social Sciences'[J]. *Annual Review of Sociology*, 2019, 45: 581–606.

[2] 黄振羽、丁云龙. 曼哈顿工程与美国科技制度变迁：一个“社会运动”的诠释[J]. 科技进步与对策，2015, 32（11）: 102–107.

表5 三种LLM的反事实推演结果比较

比较维度	DeepSeek-V3.2	文心 5.0 Preview	GLM-4.6
政治动因与启动时机	缺乏“震撼效应”，依赖情报与科学建议评估；1943年或更晚启动，依赖外部催化剂。	由“战争应急”转向“战略威慑”；1943年启动，引用MAUD报告等具体事件。	强调生存威胁转向地缘博弈与公众舆论；1943–44年渐进启动，叙事更生动。
资源投入规模与速度	规模：5–10亿美元（减半）；时间：1946年或更晚试爆；视角：强调战争驱动力的缺失。	规模：10–15亿美元；时间：1946年试爆，关键节点延迟1-2年；视角：强调资源竞争与优先权逻辑。	规模：仅原历史1/3–1/2；时间：1946年后完成；视角：强调预算竞争与市场制约。
组织模式与治理结构	结构：分散、科学机构领导；保密：降低，基础研究更开放；合作：国际地位更平等。	结构：从军事主导转向政府协同；保密：降低，部分向民间开放；合作：偏向美国独立研发。	结构：文官主导，去军事独裁化；保密：分层保密（核心设计仍机密）；合作：深度的“英美联合”模式。

[3] 尼尔·弗格森. 虚拟的历史 [M]. 颜筝 译, 北京: 中信出版社, 2012.

[4] Hawthorn, G. *Plausible Worlds: Possibility and Understanding in History and the Social Sciences* [M]. Cambridge: Cambridge University Press, 1993.

[5] Fogel, R. W. *Railroads and American Economic Growth : Essays in Econometric History* [M]. Baltimore: The Johns Hopkins University Press, 1964.

[6] Tetlock, P. E., Belkin, A. *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives* [M]. Princeton: Princeton University Press, 1996.

[7] 理查德·罗兹. 原子弹秘史 [M]. 江向东、廖湘彧 译, 北京: 金城出版社, 2018.

[8] 南希·卡特赖特. 斑杂的世界: 科学边界的研究 [M]. 王巍、王娜 译, 上海: 上海科技教育出版社, 2006.

[9] Brown, T., Mann, B., Ryder, N., et al. 'Language Models are Few-Shot Learners' [A], *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)* [C], 2020, 1877–1901.

[10] Wei, J., Wang, X., Schuurmans, D., et al.'Chain-of-Thought Prompting Elicits Reasoning in Large Language Models' [A], *36th Conference on Neural Information Processing Systems (NeurIPS 2022)* [C], 2022, 24824–24837.

[11] Bubeck, S., Chandrasekaran, V., Eldan, R., et al. 'Sparks of Artificial General Intelligence: Early experiments with GPT-4' [J]. arXiv preprint, arXiv: 2303.12712, 2023.

[12] Pearl, J. *Causality: Models, Reasoning, and Inference* [M]. Cambridge: Cambridge University Press, 2009.

[13] Ji, Z., Lee, N., Frieske, R., et al. 'Survey of Hallucination in Natural Language Generation' [J]. *ACM Computing Surveys*, 2023, 55(12): 1–38.

[14] Obermeyer, Z., Powers, B., Vogeli, C., et al. 'Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations' [J]. *Science*, 2019, 366(6464): 447–453.

[15] Lewis, D. *Counterfactuals* [M]. Cambridge, MA: Harvard University Press, 1973.

[16] 斯科特·佩奇. 模型思维 [M]. 贾拥民 译, 杭州: 浙江科学技术出版社, 2023.

[17] 西达·斯考切波. 国家与社会革命: 对法国、俄国和中国的比较分析 [M]. 何俊志、王学东 译, 上海: 上海世纪出版集团, 2007.

[18] 邱泽奇、乔天宇、宋洁 等. 数字生态：概念源流与建构 [J]. 社会学研究, 2025, 40（5）: 21–41.

[19] Ak, O. 'Platforms as Laboratories of the Social: How Digital Capitalism Matters for Computational Social Research in North America' [J]. *Social Studies of Science*, 2025, 55(5): 711–731.

[责任编辑　王大明　柯遵科]