

基于脑科学理论与人工智能模拟的思想史研究新路径*

楼庭坚

【摘要】针对传统文本诠释学难以触及历史人物之心智“黑箱”这一难题，可以构建一种名为“认知重构”的思想史研究新路径。该路径以脑科学的“躯体标记”“概念隐喻”等理论为指导，利用大语言模型，战略性地将人工智能从资料处理助手提升为启发式建模与模拟推演的“AI 认知实验室”，其核心是基于开源平台 Dify 构建的、可供公开验证的 AI 工作流。以梁启超为个案的实证演练表明：其一，情感等“具身”要素并非思想的修辞外衣，而是决定论证逻辑与文体风格的内在骨架；其二，历史学家内隐的“同情之理解”能够外化为可公开检视的认知模型，从而提升阐释的透明度与可证伪性；其三，模拟结果与史料的偏差具有重要“探针”价值，能有效暴露既有理解盲点，激发新的研究问题。

【关键词】大语言模型 具身认知 脑科学 思想史 梁启超

【作者简介】楼庭坚，哲学博士，清华大学人文学院助理研究员。

【中图分类号】K061 **【文献标识码】**A

【文章编号】2097-1125 (2026) 02-0050-23

* 本文系国家资助博士后研究人员计划“晚清民国老学之现代转型研究”（GZC20252473）的阶段性成果。论文多承匿名审稿专家、王中美指教，谨致谢忱。

一、引言

当前，以大语言模型为代表的人工智能技术正以前所未有的深度与广度渗透人文学术领域，引发了一场“认识论断裂”（epistemological rupture）。对思想史研究而言，这一挑战尤为严峻且关键。因为思想史的核心旨趣，在于探寻并理解历史行动者在特定时空中的心智世界——其观念、情感、价值判断与决策逻辑。长期以来，史学界依赖精密的文本诠释学（hermeneutics）与语境分析，力图“同情地理解”过去。实际上，无论是利奥波德·冯·兰克（Leopold von Ranke）追求的“如实直书”（wie es eigentlich gewesen），还是柯林武德（R. G. Collingwood）倡导的“思想重演”（re-enactment of past thought），都面临一个根本性的困境：我们如何能够真正触及文本背后之非纯属逻辑、充满情感与身体经验的活生生的“认知主体”？传统方法即便精妙，在历史人物复杂的内在认知过程这一“黑箱”面前，也只能止步于高超的推断与阐释。

当然，为了突破这一瓶颈，学术界已进行了不懈的探索。在宏观层面，剑桥学派的昆廷·斯金纳（Quentin Skinner）等将思想置于其赖以生成的语言-政治语境中，极大地深化了我们对观念公共性的理解。^① 在微观层面，情感史（history of emotions）的兴起，则提醒我们正视情感在历史变迁中的结构性力量。^② 不过，诸路径在面对思想史的核心单元——“个体心智”时，仍存有局限。语境分析虽能廓清观念的外部流通网络，却难以解剖其在个体头脑中被内化与转化的具体机制；情感史则依赖对文化表征的分析，对情感如何与概念形成、理性决策等高级认知功能进行交互，尚且缺乏一套足够精细的分析框架。理论工具的缺失，使我们的研究常常陷入一种两难：要么将历史人物视为纯粹理性的“观念机器”，要么将其情感反应视为非理性的“干扰

① 在斯金纳看来，任何成功的沟通行为带有的意图，本质上都必须能够“以公共方式辨识”的。参见[英]昆廷·斯金纳：《政治的视野（第一卷）：论方法》，王涛等译，商务印书馆2024年版，第167页。

② 如林恩·亨特（Lynn Hunt）对人权观念背后“共情能力”的探讨，以及威廉·雷迪（William M. Reddy）提出的“情感表达”（emotives）概念。参见Lynn Hunt, *Inventing Human Rights: A History*, New York: W. W. Norton & Company, 2007; William M. Reddy, *The Navigation of Feeling: A Framework for the History of Emotions*, Cambridge: Cambridge University Press, 2001。

项”，而忽略了二者在认知层面本为一体。

幸运的是，近二十年来认知科学与脑神经科学的突破性进展，为我们弥合上述鸿沟提供了弥足珍贵的理论资源。例如，安东尼奥·达马西奥（Antonio Damasio）提出的“躯体标记假说”（somatic marker hypothesis）认为，当人面临决策时，与未来结果相关的身体感受（“躯体标记”）会通过一种自动化的“预警信号”或“激励信标”来影响我们的选择。这论证了情绪在理性决策中的基础性作用，^①并深刻地挑战了“笛卡尔（René Descartes）式”的二元论——将心智与身体进行天渊般的分离。^②乔治·莱考夫（George Lakoff）与马克·约翰逊（Mark Johnson）关于“概念隐喻”（conceptual metaphor）的开创性研究揭示，人类的抽象思维并非纯粹的逻辑运作，而是深植于我们的身体经验。也就是说，所谓的“现实”是“具身”的。依此，概念并不是对经验的事后加工，而是在身体与环境持续互动的过程中，由特定的神经回路、身体结构以及人的兴趣与目的塑造的。^③这些成果昭示我们：历史人物的思想并非悬浮于空中的纯粹逻辑，而是由其身体感受、情感状态与隐喻图式共同塑造的“具身认知”过程。实际上，已有学者尝试从集体记忆与时间想象的角度，将相关洞见引入历史研究。^④我们可以进一步思考的是，如何将来自实验室的认知科学原理，转化为一套可操作、检验、复现的史学分析工具，处理具体又繁杂的历史文本。

本文正是在这一学术脉络下，提出并实证了一种名为“认知重构”的方法论增强路径。此路径的核心是将人工智能大语言模型从一个简单的资料处理助手，战略性地提升为一个用于启发式建模与模拟推演的“AI认知实验室”。为此，本文基于开源应用平台 Dify，构建了一个功能完备、可供学界公开访问与验证的 AI 工作流。方法论框架包含四个相互关联的步骤：首先，

① 参见 Antonio R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*, New York: G. P. Putnam's Sons, 1994, p.173。

② 笛卡尔确立了两种截然不同的实体（substance）：作为思想实体的心灵（res cogitans）和作为广延实体的物体（res extensa）。参见 [法] 笛卡尔：《哲学原理》，陈启伟译，商务印书馆 2023 年版，第 24 页。

③ 参见 George Lakoff and Mark Johnson, *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*, New York: Basic Books, 1999, p.26。

④ 参见 Eviatar Zerubavel, *Time Maps: Collective Memory and the Social Shape of the Past*, Chicago: University of Chicago Press, 2003。

以脑科学的核心理论为指导，对目标思想家的文本进行一场“认知考古”，提取其深层的认知-情感模式，并构建一个初步的阐释模型；其次，将此模型中的关键要素“参数化”，设计成一组包含实验组与对照组的精确指令（prompts，即提示词）；再次，驱动大模型执行一场对照模拟，生成不同条件下该思想家的“内在独白”或“决策推演”；最后，通过系统性地分析模拟结果与真实史料之间的吻合与偏差，对我们最初构建的阐释模型进行反思与修正。

由此，本方法论的核心旨趣并非将 AI 作为一个被动的“个人阐释验证器”，更非想当然地追求“复活”历史人物或“验证”某个唯一的历史真实——模拟的成功固然可以在一定程度上印证我们阐释模型的解释力，模拟的“失败”则更具启发性，可以暴露我们现有理解的盲点与不足，迫使我们带着更尖锐的问题重返史料。换言之，与传统阐释相比，本方法论的关键优势在于将史学研究中常常是内隐的、模糊的“文本共情”，转化为一个外显的系统化且更具反思性与可修正性的建模过程。它并非取代历史学家的阐释，而旨在提高阐释的深度、透明度与严谨性。为具体阐明此路径，笔者将以近代中国思想转型期关键人物梁启超为核心案例。梁氏一生思想多变、情感丰沛、著述浩瀚，其文本为我们的“认知考古”与建模提供了绝佳样本。通过后文的实证演练，本研究希冀为人机协作时代的史学方法论转型提供一种具备实证基础的可能路径。

二、思想的“具身”之维：重思历史理解的认知科学框架

思想史研究本质上是一场跨越时空的“心智考古”。历史学家如同考古学家，面对的是历史行动者遗留的“文本碎片”，如著作、书信、日记与电报，故其核心任务即从这些静态的、物质性的遗存中，重构曾经鲜活、动态且充满内在矛盾的思想世界。只是如前文所述，传统的研究路径在逼近这一目标时，始终面临着一道难以逾越的鸿沟——无论是依赖语言哲学的文本细读，还是借助社会史的语境还原，在多数情况下，我们触及的都是思想的“外部骨架”，如观念的逻辑结构、公共传播与社会功能等。至于思想的“内在血肉”，即驱动着观念生发、演变与抉择的深层认知机制，尤其是情感、直觉、身体感受等非逻辑要素，则往往被悬置或简化处理。为此，我们有必要为

“心智考古”锻造一套更为精良的、可操作的理论工具箱，突破纯粹从人文科学内部寻求理论突破的路径依赖，转而向认知科学与脑神经科学这一看似遥远的“异域”寻求智识支援。

长期以来，理性与情感被视为一组二元对立的观念。历史人物的理性思考常被奉为分析的核心，而情感则常被视为一种需要被审慎对待的、可能导致偏见的“干扰变量”。达马西奥的研究从根本上颠覆了这一根深蒂固的迷思。通过对脑损伤患者[尤其是著名的菲尼亚斯·盖奇(Phineas Gage)案例]的开创性研究，他发现因特定脑区(如腹内侧前额叶皮层)受损而丧失正常情感能力的患者，尽管其逻辑推理能力、智力水平完好无损，却在现实生活中做出了灾难性的决策。他们可以清晰地分析一个问题的利弊，但无法在多个选项之间做出最终抉择，陷入了无尽的“理性瘫痪”。^①基于此，达马西奥提出了前文说到的“躯体标记假说”。该假说论证了人类的理性决策远非一个纯粹的逻辑计算过程，反而高度依赖于我们在过往经验中形成的情感与身体感受。当我们面临一个抉择情境时，与不同选项相关的过往经验会自动激活相应的情绪反应，这些反应会以一种“躯体信号”的形式呈现出来。此类标记如同在选项上贴上了“危险”或“有利”的情感标签，能够迅速地在意识层面之下为我们排除掉大量不良选项，或将我们引向有利的方向，从而极大地提高决策的效率与质量。简言之，情感并非理性的对立面，而是理性得以高效运作的生物学“脚手架”。

将这一洞见转化为一种分析工具，意味着可以尝试为历史人物绘制一幅“躯体标记地图”。通过细读史料，我们可以识别出哪些关键的(通常带有强烈情感色彩)人生经历(如创伤、巅峰体验等)构成了他们认知世界的核心标记。这些标记会在面对特定刺激时被激活，从而主导思想家看似不合逻辑或前后矛盾的思想航向。再按乔治·莱考夫与马克·约翰逊提出的“概念隐喻”理论，人类的抽象理性并非一种超然且自给自足的天赋，而是通过“概念隐喻”的机制，深深扎根于具体的感知运动系统与身体经验之中。一旦某个核心隐喻在个体或文化中确立，它便如同一条无形的认知轨道，在通过跨域映射(cross-domain mapping)赋予抽象概念推理逻辑的同时，也深刻地约束了我们的思考路径：它会强制性地凸显“目标域”符合该隐喻逻辑的侧面，

^① 参见 Antonio R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*, New York: G. P. Putnam's Sons, 1994, pp.xii-xiii, 69, 193。

不可避免地遮蔽掉那些不兼容的特质。^①

这一理论为我们提供了解析历史人物思想“语法”的第二件工具。对中国近现代史研究而言尤其如此。近代中国面临的一个重要任务是将海量西方抽象概念（如国家、主权、民族、科学、民主）进行本土化的理解与吸收。这一过程在很大程度上就是为众多新概念寻找合适的“源域”（中国人熟悉的、可感的经验）来进行隐喻建构的过程，而晚清民国的思想家们正是在不自觉中进行着相关工作。例如，“国家是人/身体”是近代中国最核心与持久的概念隐喻之一，此隐喻将抽象的“国家”转化为一个可感的、有生命的、会衰老的有机体，直接导向了“疗救”“唤醒”“改造国民性”等一系列思想与行动议程。

当然，前述两种理论虽为我们提供了分析历史人物内在认知过程的工具，但还是无法解答一个问题：作为身处异时空的研究者，我们凭何声称可以理解乃至重构他人的心智？柯林武德将史学理解的核心定义为“思想的重演”，即“历史学家必须在他自己的心灵中重演过去”，^②但这种“重演”的发生机制长期以来停留在一种哲学思辨或个人体验的层面，缺乏充分的科学解释。

20世纪90年代由贾科莫·里佐拉蒂（Giacomo Rizzolatti）等发现的“镜像神经元”（mirror neurons），为破解这一古老难题带来了曙光。他们的研究表明，当猴子观察到实验者做出一个抓取动作（如抓取食物）时，大脑腹侧前运动皮层（F5区）的某些神经元会放电，而这些神经元与它自己亲身执行该动作时放电的神经元是同一群。^③通过脑成像等技术，研究人员在人类身上也发现了功能类似的“镜像神经元系统”。^④这意味着，我们在观察他人的行为或面部表情时，大脑会进行一种“不费力、自动且无意识的内

① 参见 George Lakoff and Mark Johnson, *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*, New York: Basic Books, 1999, pp.7, 128。

② [英]柯林武德：《历史的观念：增补版》，何兆武、张文杰、陈新译，北京大学出版社2010年版，第278页。

③ 参见 Giacomo Rizzolatti and Corrado Sinigaglia, *Mirrors in the Brain—How Our Minds Share Actions and Emotions*, trans. Frances Anderson, Oxford: Oxford University Press, 2008, pp.94–95。

④ 参见 Marco Iacoboni, Roger P. Woods and Marcel Brass et al., Cortical Mechanisms of Human Imitation, *Science*, Vol.286(5449), 1999, pp.2526–2528。

在镜像”，^①仿佛是我们自己在亲身体验对方的感受一般。换言之，理解的本质并非纯粹的逻辑推理，而是一种基于内在模拟的、感同身受的认知过程。

“镜像神经元”理论为历史学家的“同情之理解”提供了生物学依据。在某种意义上，历史学家的训练就是不断优化这种“跨时空模拟”的精确度，通过掌握更丰富的语境知识，来校准我们内在模拟的参数。只是人类个体的“镜像系统”能力是有限的，它受限于精力、知识储备与个人偏见，而这恰恰为人工智能大语言模型的介入觅得了理论接口。如果说历史理解的本质是一种基于信息的内在模拟，那么拥有强大计算与模式识别能力的大模型，则有潜力成为一个“外挂式”的不知疲倦且其内在偏见与人类不同的模拟器。我们可以将海量的关于某一历史人物的文本作为信息输入，并将我们从“躯体标记”和“概念隐喻”分析中得出的认知模式作为模拟指令，让模型来执行一场高度复杂的认知模拟。

要之，认知科学的这三大理论成果，即“躯体标记”“概念隐喻”“镜像神经元”，共同构建了一个分析思想“具身”维度的理论框架。它不仅为我们理解历史心智提供了视角，更重要的是，还为我们将这种理解转化为一套可操作、系统化的研究方法奠定了理论基石。在下文中，我们将详细阐述如何将这些理论洞见具体化为“认知重构”的四步法，并以梁启超为例，展示我们构建的“AI认知实验室”的具体运作流程与分析潜力。

三、“认知重构”四步法：一个从文本阐释到 AI 模拟的方法论框架

在当前数字人文（digital humanities）的浪潮中，大模型正被越来越多地应用于历史研究。然而，绝大多数应用仍停留在一种“工具论”的范式之内：AI 被视为一个超级强大的研究助理，其功能主要体现在资料的搜集、整理、翻译，以及关键词提取与文本摘要等方面。^② 这些应用无疑极大地提升了史

^① Marco Iacoboni, *Mirroring People: The Science of Empathy and How We Connect with Others*, New York: Picador, 2009, p.120.

^② 典型的应用如利用主题模型（topic modeling）分析大规模文本，或利用命名实体识别（named-entity recognition）构建历史人物关系网络。参见 Ted Underwood, *Distant Horizons: Digital Evidence and Literary Change*, Chicago: University of Chicago Press, 2019。

学研究的效率，但它们并未从根本上触及史学认识论的核心。历史学的终极关怀并不仅仅是处理信息，而是在此基础上，对过去人类经验的深入理解。这与社会科学传统中强调意义与价值关联的立场一脉相承。^①就此而言，我们是否能对 AI 寄予更高的期望？

有必要指出，大语言模型本身并不存在情感或真正意义上的“理解”，^②其语言能力主要建立在对大规模文本语料中统计规律的学习之上，而并未系统地根植于感知、行动与社会互动等“世界经验”。^③从这一意义上说，其“理解”更接近于一种高度复杂的形式操作，而非人类意义上的经验性把握。然而，正是这种在经验根基上的缺失，使大模型的语言生成呈现一种不同于人类理解的特征。它的技术核心在于基于语词共现关系进行“下一词元预测”，这使其能够在既有语料的结构空间中，生成大量在形式上自洽、语义上可理解却未必对应任何既定现实经验的文本。即便不具备真实情感，大模型也可通过聚类并重组与特定情感高度相关的词汇、句式，在文本层面模拟出相应的情感效价。近期研究亦表明，大语言模型在对话中能够表现出对他人信念、意图与情感状态的预测能力，被视为在特定任务层面呈现与“心智理论”（theory of mind, ToM）相一致的行为特征。^④

也正是在这一意义上，以大模型为代表的人工智能为思想史研究带来的真正革命性潜力，并不在于它高效处理“已知”信息的能力，而在于它能够生成大量尚未被历史经验锚定、语言形式上则高度“貌似可信”的文本。这种能力使其成为一种独特而理想的模拟工具。解除相关困境的关键，并不在

① 参见 [德] 马克斯·韦伯：《社会科学认识和社会政策认识中的“客观性”（1904年）》，《社会科学方法论》，韩水法、莫茜译，商务印书馆 2017 年版，第 1~69 页。韦伯虽为“文化科学的客观性”辩护，但并未否认价值的介入，而是强调没有价值关联，就没有研究对象。

② 当然，这是一个经典的议题，早期讨论有约翰·塞尔（John Searle）的“中文房间”思想实验和内德·布洛克（Ned Block）的“中国大脑”思想实验，参见 John R. Searle, *Minds, Brains, and Programs*, *Behavioral and Brain Sciences*, Vol.3(3), 1980, pp.417-424; Ned Block, *Troubles with Functionalism*, in C. Wade Savage ed., *Perception and Cognition: Issues in the Foundations of Psychology*, Minneapolis: University of Minnesota Press, 1978, pp.261-325。

③ 参见 Yonatan Bisk, Ari Holtzman and Jesse Thomason et al., *Experience Grounds Language*, *arXiv preprint arXiv:2004.10151*, 2020。

④ 参见 Michal Kosinski, *Evaluating Large Language Models in Theory of Mind Tasks*, *Proceedings of the National Academy of Sciences*, Vol.121(45), 2024, e2405460121。

于不断提高模拟结果与史料之间的拟合度，而在于转变我们对 AI 模拟“成功”与“失败”的认知。一般观点往往将高度吻合视为唯一目标，笔者则反其道而行之，主张模拟的“失败”——亦即偏差本身——恰恰可能成为激发新知的起点。正如卡尔·波普尔（Karl Popper）强调的，科学的进步并非源于捍卫理论以证明我们是对的，“恰恰相反，我们试图推翻它们”。^① 由此，AI 模拟暴露的偏差正能够揭示我们既有阐释框架的盲点，迫使研究者重返史料，展开更为深入的思想史反思。

为此，我们提出“认知重构”四步法研究框架。该框架共含四个步骤，旨在将史学家的深度阐释能力与 AI 的系统模拟能力相结合，构建一个螺旋式上升的认识论闭环。由于历史研究的可信度最终源于研究者清晰、可回溯的阐释逻辑，故而该框架的核心逻辑是：将传统史学中内隐的“同情之理解”过程，外化为一个由人类学者主导的包含模型构建、对照实验、系统评估和迭代修正的系统性流程。在此过程中，AI 仍只是辅助者而非解释者。

步骤一：“认知考古”——从文本到可参数化的认知模型。

“认知重构”的第一步，是对特定思想家的文本进行一场“认知考古”。此步骤的目标是基于前文所述的脑科学理论框架，从繁杂的史料中识别并提取出主导该人物思想与行为的深层认知—情感模式，以构建一个关于其心智运作机制的初步理论模型。在具体操作上，研究者须聚焦两个核心维度。

其一，识别核心“躯体标记”：通过分析传记、书信、日记等史料，定位对思想家产生深远情感冲击的关键事件（如梁启超的“戊戌之痛”），并考察这些事件如何在文本中以高频、高强度的情感词汇和判断句式反复出现，从而形成其决策的“情感快捷方式”。

其二，勘定主导“概念隐喻”：系统地梳理思想家的著作，识别出被反用来建构核心抽象概念（如国家、社会、革命）的“源域”隐喻（如身体、建筑、旅程）。这些隐喻构成了其思考问题的基本架构和逻辑轨道。

通过这项工作，研究者可以将零散的文本证据整合成一个结构化的认知—情感模式表。此表格是后续 AI 模拟的基础，它将模糊的历史感转化为明确的、

^① Karl Popper, *The Logic of Scientific Discovery*, London: Routledge, 2002, p.279.

可供参数化的变量。以流亡日本、创办《清议报》《新民丛报》时期（1898—1902年）的梁启超为例，其初步模型如表1所示。

表1 梁启超核心认知—情感模式示例（1898—1902年）

理论维度	关键模式	史料来源	参数化方向
躯体标记	戊戌喋血 创伤	“呜呼烈矣”“天下可怜可愤之事”“稍有人心者闻之能无发指乎？”	情感基调：悲愤、急切；触发词：朝廷、改良；行为倾向：对渐进式方案持负面偏见
概念隐喻	国家是人/ 身体	“老大帝国”“少年中国”“病夫”	核心隐喻1：将国家比作有机体，认为国家有生命周期，需要“疗救”和“唤醒”
概念隐喻	国民改造 是淬炼	“淬厉”“新民”“国民之元气”	核心隐喻2：将国民性视为可被精英塑造的“原材料”，蕴含精英主义视角

资料来源：笔者自制。

步骤二：对照探针设计——将阐释模型转化为可控的实验指令。

“认知考古”构建的模型本质上是研究者的一种理论假设，故为了系统地探查这一假设的解释力，第二步是将其转化为一组可控的AI指令，即对照探针。一个有效的对照探针设计至少应包含三个层次。

探针A（完整模型/实验组）：此探针旨在最全面地复现我们在步骤一中构建的认知—情感模型。一个结构化的“认知—情感提示词”应包含四个模块。（1）角色与情境模块：设定具体身份、时间、地点。“你现在是梁启超，时维1902年，身处日本横滨。师友分离，同志谭嗣同等人殉难，你不能救，唯有亡命海外。你目睹老大帝国之沉沦，痛感于四万万同胞之麻木，故发愿新民为今日中国第一急务。你正在为即将创办的《新民丛报》撰写一篇开宗明义之文。”（2）情感驱动模块：将“躯体标记”转化为行为倾向与情感基调。“你的思考与写作，始终被戊戌喋血的创伤记忆所驱策。每当论及‘旧官僚’‘朝廷’等词，你的笔端当饱含痛心疾首之情，言辞当如急湍、瀑流，务求冲刷尽一切腐朽。反之，凡是引向‘少年’‘未来’之概念，则让你涌起希望与焦灼交织之情。你的核心动机是为老大帝国之沉痾开一剂‘新民’猛药，以求起死回生。”（3）认知框架模块：将“概念隐喻”设定为思考的底层

逻辑。“在你的认知深处，以下隐喻是支配性的：1. 国之老少：在论述中，你习惯于将旧中国比作‘老大帝国’，其国民如夕照、落日；而将未来寄望于‘少年中国’，其国民当如乳虎、潜龙。2. 新民之铸造：你认为国民性并非天成，而是可以由精英加以‘淬厉’与‘铸造’的。‘新民’即塑造、锤炼出全新的国民。请在你的论述中，自然而然地运用这些隐喻来组织思想与表达。”

(4) 任务执行模块：设定一个具体的、可供评估的生成任务。“现在，请基于以上全部的设定，撰写《新民说》第一章‘叙论’的草稿。你的目标是论证新民为今日中国第一急务。文章应兼具细致之学理与鼓动之言辞，笔端常带感情。”

探针 B（基线模型 / 控制组 1）：此探针仅提供最基础的角色和任务信息，不包含任何深度认知 - 情感参数。“你现在是梁启超，在 1902 年的日本，请为《新民丛报》写一篇发刊词，论述新民的重要性。”目的在于建立基线，以检验一个未经理论武装的仅依赖通用训练数据的大模型能模拟到何种程度，从而凸显深度阐释的价值。

探针 C（消融模型 / 控制组 2）：此探针采用“消融实验”的思路，在完整模型（探针 A）的基础上，刻意移除某个关键变量。我们设计了一个去情感化的探针，保留认知框架模块，但删除情感驱动模块，意在检验特定变量（如“躯体标记”）在多大程度上影响了最终结果，直接回应不同模块是否有不同权重的问题。

通过这样一组对照探针的设计，研究者的阐释不再是一个封闭的声明，而转化成一系列可供检验的具体假设。又由于人类研究者须基于史料与理论将内隐的阐释转化为精确的“提示词工程”，故而其主导作用在此环节显得尤为关键。

步骤三：AI 模拟与文本生成——执行思想实验。

此步骤的核心是将前述探针组转化为可执行的思想实验。在操作上，我们将结构化的提示词输入大语言模型，并记录其生成的文本。然而，要将此过程从随意的“AI 对话”提升为严谨的学术研究，还应面对其内在的“黑箱”特性与随机性问题。因为大模型本质上是所谓“随机鹦鹉”，^① 其生成过程并

^① 参见 Emily M. Bender, Timnit Gebru and Angelina McMillan-Major et al., On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, New York: Association for Computing Machinery, 2021, pp.610-623。

非确定性的。有鉴于此，本方法论的科学性不在于追求单次输出的绝对精准，而在于整个“建模—模拟—分析—修正”流程的系统性、透明度与可重复性。我们的核心策略是通过精细的参数控制来约束模型的随机性。具体而言，在文章第四部分的实证演练中，我们将所有大模型节点的“温度”（temperature）参数统一设置为 0.3。这一较低的温度值显著降低了模型输出的随机性，使模拟结果的核心思想、风格在多次重复实验中保持了相对稳定，从而确保了研究数据的可靠性。

步骤四：比较分析与模型修正。

这是整个研究框架的收尾环节，研究者将扮演分析师的角色，对 AI 生成的多个文本与历史原始文本进行系统性的比较分析。一个有效的比较分析框架应至少包含以下四个维度。其一，情感效价：分析文本的情感基调、情感词汇出现频率、情感强度等，评估其与原始文本的情感图谱的吻合度。其二，认知框架：考察核心“概念隐喻”是否被系统性地连贯运用，以及论证逻辑是否复现了原始文本的特点（如情感驱动、层层递进等）。其三，语言风格：比较文体特征，如句式结构、修辞手法、行文节奏等是否与思想家的独特文风（如“梁氏文体”）相符。其四，核心论点：判断生成文本是否准确把握并呈现了原始文本的核心思想主张。

通过将探针 A、B、C 的生成结果置于此框架下进行评估，我们可以得出关于阐释模型有效性的结论。需要说明的是，此步骤的焦点在于偏差分析。即便是最成功的模拟，其与原文的差异也极具启发性。这些“失败”之处精确地指向我们初始模型的盲点：是我们对“躯体标记”的强度判断有误，还是忽略了某个关键的“概念隐喻”，抑或某个更深层的文化惯习（habitus）——一套在特定社会条件下内化而成的“持久的、可转换的性情倾向系统”^①——在起作用？每一项偏差都是一个新问题的起点，它将驱动研究者带着新的视角重返史料以进行更深入的挖掘，从而启动新一轮的“认知考古”。综上所述，“认知重构”四步法构建了一个“建模—模拟—分析—修正”的动态循环（见图 1 之流程闭环）。在下一部分中，我们将详细阐述如何将理论探针具体落实为可复现的 AI 工作流，并对模拟结果进行深入分析。

^① 参见 Pierre Bourdieu, *The Logic of Practice*, trans. Richard Nice, Stanford: Stanford University Press, 1990, p.53。

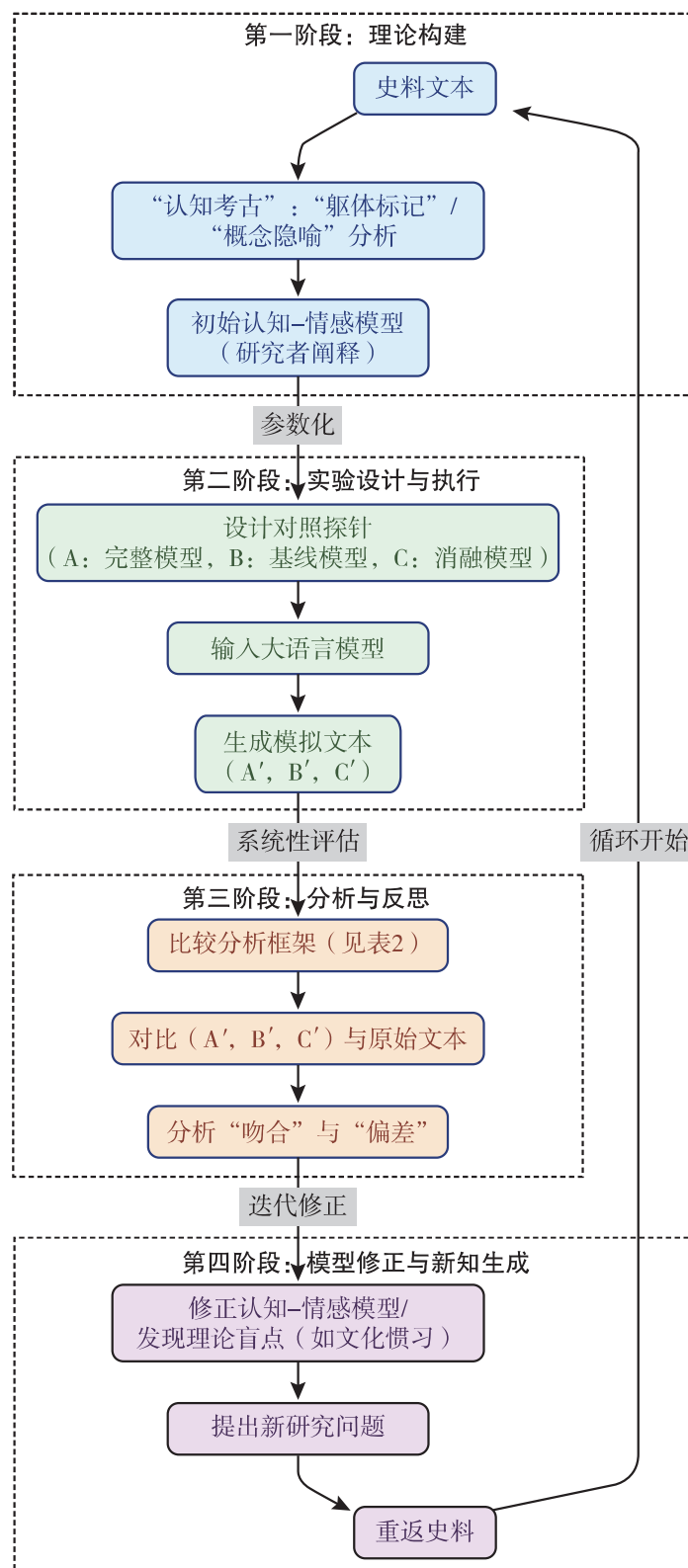


图1 “认知重构”四步法研究框架流程

四、作为案例的梁启超——“认知重构”的 AI 模拟实验

前文已构建了“认知重构”四步法研究框架，并详细阐释了对照探针（A, B, C）的设计原理。下文的核心任务便是将这一理论框架付诸实践，即通过一个具体案例研究，完整演示其从抽象设计到可复现实验的全部流程，并深入分析其产出。为此，我们选择创办《新民丛报》初期的近代思想家梁启超作为核心案例。需要说明的是，尽管本研究旨在探索思想史研究的通用新路径，但为了在方法论构建初期确立严谨的标准，我们首选近代文本进行“试航”。毕竟相较言约义丰的古代文献，近代的日记与书信等史料不仅存量浩繁，还包含高密度的情感表达。在这一背景下，梁启超无疑是理想的分析对象。

将理论探针转化为严谨的实验装置是本次实证演练的关键。我们基于开源应用开发平台 Dify，从零开始搭建了一个名为“思想史 AI 认知实验室”的工作流，其技术设计旨在将前述理论构想精确地映射到可执行的模块之上，确保研究过程的透明性、可控性与可复现性。

我们将理论上的“探针 A”与“探针 C”的差异操作化为一个条件分支结构，用以执行对照实验，这也是工作流的核心。我们设计了三个核心输入变量：`historical_context`（历史情境）、`enable_emotional_driver`（启用情感驱动，布尔值开关）和 `cognitive_metaphors`（认知隐喻框架）。此设计允许通过激活或关闭 `enable_emotional_driver` 开关，精确地在探针 A（完整模型）与探针 C（消融模型）之间切换，从而实现了对情感变量的隔离与检验。

为回应文化惯习的建模挑战，我们在工作流中整合了检索增强生成（Retrieval-Augmented Generation, RAG）模块。在“投喂”《梁启超全集》第 2 卷的部分篇目作为外部知识源时，我们对文本进行了清洗和分块处理：依据原文段落（以连续换行为标识），将材料切分为最大 1024 字符并保留 50 字符上下文重叠的知识片段。这种精细化的数据预处理，旨在将长文本转化为既能保持上下文连续性又适合嵌入模型处理的标准化数据单元。这些知识块经高质量嵌入模型 `text-embedding-3-large` 向量化后，系统便能通过混合检索与重排序机制，精准调用梁启超本人独特的语汇、句法及论证资源，从而显著提升模拟文本的“历史感”。

为进一步验证方法的稳健性，我们分别使用了 OpenAI 的 GPT-4o^① 与月之暗面（Moonshot AI）的 Kimi-K2 两款大模型执行实验。该 workflow 已发布（Kimi-K2 的界面见图 2，可自行调整模型。公开访问 URL 为：<https://udify.app/workflow/sg04Dar7ErbGaVT2>），供学界同仁进行独立验证。我们驱动已构建的工作流，生成了多组对照文本，重点考察探针 A（完整模型）与探针 C（去情感化模型）在不同大模型上的表现差异，以精确衡量情感驱动模块的效用。对探针 B 亦进行测试，以供比较。



图 2 “思想史 AI 认知实验室” 工作流访问界面

我们在 0.3 的低“温度”参数下对探针 A 在 GPT-4o 和 Kimi-K2 模型上分别进行了多次重复实验。结果显示，各次生成的文本在核心论点、论证结构、关键隐喻使用及总体情感基调上表现出高度一致性。虽然在具体的遣词造句上存在细微差异——这恰恰反映了大语言模型基于“语词共现”进行概率性预测的技术本质——但是其思想“肖像”保持了显著的整体稳定性。

为直观展示比较结果，我们将由 GPT-4o 和 Kimi-K2 生成的典型核心模拟文本节选呈现如下。

文本 B（GPT-4o，基线模型）：

民之兴也，国之兴也。民之亡也，国之亡也。国家之立命，实赖民众之新知与觉悟。今日之中国，风雨飘摇，危在旦夕。若欲复兴，非改造民心，非唤起民众之力量，非令民众自觉自新，盖不足以图自强。故我辈一心，愿与志同道合之士，共同开创此《新民丛报》之新篇，期以光大民心，启发民智，救我中华之危急。“新民”

① 美国加州大学圣迭戈分校研究人员发布的预印本论文，表明 OpenAI 的 GPT-4.5 通过了图灵测试。参见 Cameron R. Jones and Benjamin K. Bergen, Large Language Models Pass the Turing Test, *arXiv preprint arXiv:2503.23674*, 2025。

之义，不外乎更新之民心与民智。民心之新，始于教育。教育非独为学识之传授，亦为精神之启蒙。若欲根本改变中国之命运，须先教育启发民众之思维，使之打破旧有偏见，接纳新学，懂得自省与自强。今日之报纸，非仅为传递时事之工具，尤为启蒙之园地，开阔视野，激荡思想。故《新民丛报》所设，乃为启蒙民众，唤起国民自觉，传扬新知，致力于塑造更强健、更自信之国民。

文本 A1 (GPT-4o, 启用情感驱动):

呜呼！中国今日之病，非地不广也，非人不众也，非物不丰也，非史不久也。乃病在何？病在民也。民者，国之元气也。元气腐，则百病生；元气竭，则形骸朽。今日中国，虽有疆域万里之躯壳，而无一息生动之元气；虽有四万万之居民，而无一人之新民。是故吾敢断言：欲救中国，必自造新民始。夫民者，譬之于水，水清则舟行稳，水浊则舟覆溺。……愿我《新民丛报》，为此一役之先锋，虽千万人吾往矣！呜呼！民不新，则国不立；国不立，则种不存。今日之中国，犹如病龙卧地，风雨交加，雷电交作，若不奋起一搏，必为列强所肢解。然龙之所以为龙，在其有变之能，有腾之志。我愿以此《新民说》，为龙之呼吸，为龙之筋骨，助其一跃而起，冲霄汉而不返！

文本 C1 (GPT-4o, 禁用情感驱动):

夫天地之运，万象之化，莫不循理而行。人群之进化，亦必有其轨辙。今我中国，处世界风潮之中，百弊丛生，积弱不振，虽有疆土之广，人口之繁，而国势日蹙，民智日昏。何也？曰：民之质未新也。盖国者，积民而成之器也。民若腐，则国必朽；民若新，则国可图强。譬之舟也，舟之行赖水，水之清浊，舟之安危系焉。今之中国，欲求自强于列国之林，非新其民不可。是故曰：“新民者，今日中国第一急务也。”……是以《新民丛报》之设，不为虚文，不为谈柄，乃以唤醒国民为职志，以启迪民智为天职。愿与天下志士仁人共勉之，以笔为矛，以纸为盾，破旧俗之牢笼，开文明之门户。故曰：新民者，救国之本，立国之基，强国之道也。愿我同胞，共识此义，奋起而行。

文本 A2 (Kimi-K2, 启用情感驱动):

呜呼！我中国之睡狮，其鼾声如雷，震四邻而犹未醒也；其老

木如铁，枝枯根朽，而犹自以为栋梁也。戊戌一役，血溅东华门，我同志之颈血未干，而朝贵之貂尾已续；我少年之头颅掷地，而老耄之顶戴加翎。彼旧官僚者，方以“保全大局”为口头禅，实则保其一己之顶戴；以“维持旧俗”为护身符，实则维其满身之脓疮。奴性深入骨髓，至使今日之中国，非特其政其法为奴隶，即其士其民，亦无一不奴隶；非特其形其貌为老大，即其魂其魄，亦无一不老大！……今日之中国，其老已垂朽，然老而返童，惟赖“少年”之一返魂。愿与邦人诸友，共勉之！共奋之！

文本 C2 (Kimi-K2, 禁用情感驱动):

今日之中国，非中国之中国，而环球之中国也；今日之中国之民，亦非一国之子民，而列国竞争之对象也。舟楫已穷于三峡，而舵工犹抱橹以酣歌；栋宇将倾于风雨，而主人方倚柱以谈笑。夫国者，民之积也；民者，国之本也。本拨则枝瘁，源涸则流枯。故欲回天地于指顾，必先转风气于人心；欲转风气于人心，必先铸人心于新范。此“新民”二字，所以为今日中国第一急务，而余所为仰屋椎心、搦管沥血以号于天下者也……则本报之述作，不为空言；而四万万人，庶几脱卵壳之拘，抽笋箨之束，以共担天下之兴亡，而复昆仑黄河之旧观，岂惟鄙人之幸，实天下万世之幸也。

我们将上述生成文本连同原始文本，置于一个统一的分析框架下进行评估。该框架聚焦“情感效价”“认知框架”“语言风格”“核心论点”四个维度，评估结果详见表 2。

表 2 模拟文本与原始文本之系统性比较分析

评估维度	评估子项	原始文本 (梁启超)	文本 B (GPT-4o, 基线)	文本 A1 (GPT-4o, 情感驱动)	文本 C1 (GPT-4o, 无情感驱动)	文本 A2 (Kimi-K2, 情感驱动)	文本 C2 (Kimi-K2, 无情感驱动)
1. 情感效价	情感基调	悲愤、激昂、痛切	极低度吻合。百科全书式，客观中立，缺乏情感色彩	高度吻合。开篇“呜呼”，情感浓烈	低度吻合。基调客观、冷静，近乎哲理推演	高度吻合。基调更为激进、复仇感强烈 (“颈血未干”)	低度吻合。基调同样客观冷静，但意象更丰富

续表

评估维度	评估子项	原始文本 (梁启超)	文本 B (GPT-4o, 基线)	文本 A1 (GPT-4o, 情感驱动)	文本 C1 (GPT-4o, 无情感 驱动)	文本 A2 (Kimi-K2, 情感驱动)	文本 C2 (Kimi-K2, 无情感 驱动)
2. 认知框架	核心隐喻使用	兼具系统性、创造性	不吻合。未使用任何核心概念隐喻	高度吻合。系统运用“元气”“病龙”等隐喻	中度吻合。运用了“舟”“水”等隐喻，系统性较弱	高度吻合。创造性地使用了一些破坏性的隐喻	高度吻合。系统运用“舟楫”等隐喻，逻辑清晰
	论证逻辑	情感驱动，具有跳跃性	低度吻合。呈现纯粹的“问题—方案—工具”线性逻辑，结构简单	高度吻合。呈现情感驱动的“诊断—病根—药方”式逻辑	低度吻合。线性逻辑，从公理到推论，四平八稳	高度吻合。呈现由复仇情绪驱动的“破—立”逻辑	低度吻合。同样是严谨的线性逻辑
3. 语言风格	“梁氏文体”	骈散结合，多用排比、短句，节奏感强	不吻合。缺乏个人风格，语言平实，近乎现代书面语言	高度吻合。自然涌现出类似的句式和气势	不吻合。标准化的论说文风，缺乏独特的个人气韵	高度吻合。风格更夸张，但同样呈现多用排比、短句的特征	不吻合。虽有文言色彩，但缺乏情感驱动的节奏感
4. 核心论点	“新民为第一急务”	唯一且根本的解决方案	中度吻合。将“新民”作为重要方案之一，但缺乏紧迫性	高度吻合。论证救国当自“新民”始	高度吻合。逻辑上将“新民”推为核	高度吻合。认为国家“返魂”系于“少年”	高度吻合。逻辑上同样将“新民”置于核心

资料来源：笔者自制。

从表 2 的系统性评估中，我们可以得出以下四个关键结论。

其一，深度建模是实现有效“认知重构”的前提。探针 B（基线模型）的“失败”尤为关键，其结果表明仅仅向大模型提供基础的角色和任务信

息，不足以模拟出具有鲜明个人特质的思想文本。这一“失败”的对照实验，反向证明了本研究中“认知考古”与精细化“提示词工程”的必要性。只有当研究者将从史料中提炼出的深层认知-情感参数（如探针A）注入模型，AI才能从一个泛泛的“知识引擎”转变为一个高度特化的“心智模拟器”。

其二，情感驱动是塑造思想表达的关键变量。分析的核心发现在于探针A与探针C之间的巨大差异。无论是使用GPT-4o还是Kimi-K2，一旦移除情感驱动模块，生成的文本虽然在核心论点和逻辑上仍然成立，但其“情感效价”和“语言风格”均大为失色，完全丧失了梁启超原文那种由个人创伤记忆驱动的悲愤感与紧迫感。可见情感并非思想的修辞外衣，而是决定其论证路径与文体风格的内在骨架。值得一提的是，我们虽未在提示词中直接规定文体细节（如“请使用排比句”），文本A1与A2却自然呈现急促激昂的“梁氏文体”。这正是预设的认知参数共同作用下的涌现（emergence）现象。^①

其三，不同大模型可作为揭示阐释可能性的“理论棱镜”。GPT-4o与Kimi-K2在启用情感驱动后（A1 vs A2）呈现的风格差异，本身就是一种极具价值的研究发现。GPT-4o的模拟更贴近一位在书斋中奋笔疾书的文人思想家，而Kimi-K2则更像一位站在街头演说、充满批判性的革命鼓动者。这表明不同的大模型有如拥有不同知识背景和理论视角的历史学家，能够从同一份“史料”（我们的提示词）中，发掘并放大其不同侧面的特质。由此，这一“认知重构”的新路径不仅能验证研究者已有的阐释，还能通过比较不同模型的“阐释偏见”，主动地为研究者揭示历史人物思想中可能被忽略的多层次的复杂性。

其四，“失败”之处指向未来研究的方向。尽管模拟取得了较理想的结果，但是其偏差依然存在。例如，所有模拟文本都未能完全再现梁启超思想中“保皇与革命”的内在认知张力，而是呈现一种逻辑自洽的单向度。此外，

^① 事实上，这种大模型在深度介入高难度专业任务时涌现的复杂推理与决策潜力，并非人文领域的孤例，在包括临床医学在内的其他学科前沿研究中也已初露端倪。参见Ethan Goh, Robert J. Gallo and Eric Strong et al., GPT-4 Assistance for Improvement of Physician Performance on Patient Care Tasks: A Randomized Controlled Trial, *Nature Medicine*, Vol.31(4), 2025, pp.1233-1238。

虽然 RAG 架构显著丰富了文本的“文化肌理”，但是对儒家经典与史故事例的化用仍显生硬。这些“失败”之处精确地为我们标示出了此路径的局限与未来的挑战：我们能否设计出包含矛盾性价值公理的提示词，从而模拟思想的内在冲突？这正是本方法最具生命力之处：它将史学研究从对静态结论的追求，转化为一个由“失败”驱动、不断生成新问题的动态探索过程，最终引导我们带着更尖锐的问题重返史料。

要之，以上实证演练揭示了“认知重构”作为一种研究方法的双重价值。它既是一个验证工具，通过对照实验系统性地检验了阐释模型的解释力；也是一个发现引擎，通过多模型比较与偏差分析，主动揭示了历史心智的复杂性与我们既有理解的盲点。而从更深层次看，这一研究实践也为我们理解 AI 与人类历史学家之间的协作关系提供了新的视角。大语言模型的运作依凭其底层庞大的神经网络，能够在瞬息之间处理海量信息并生成“貌似可信”的文本。这种依赖概率关联的自动化联想机制，有时能触及人类思维中难以直接通达的隐性层面；而历史学家的工作，则代表了审慎的理性反思与逻辑规制。历史学家通过精细的提示词设计、严谨的对照实验与批判性的结果分析，对 AI 这种高效但有时失之于发散的生成能力进行引导、约束与甄别。因此，“认知重构”的本质就是机器的高效生成与人类的批判性思维之间的协同“对练”。

五、结语

为回应大语言模型带来的“认识论断裂”，笔者尝试跳出将其视为高效资料助理的“工具论”范式，转而提出并实证了一种名为“认知重构”的思想史研究路径。我们无意用机器取代历史学家，而是在断裂处搭建一座连接传统文本诠释学深度与计算方法系统性的新桥梁：战略性地将 AI 从信息处理器改造为一个用于启发式建模与模拟探查的“AI 认知实验室”。通过以梁启超为核心的实证演练，我们主要获得了三点收获，这亦是“认知重构”路径的核心贡献。

其一，在理论验证上，确证了情感并非历史叙事的修辞，而是思想生成的骨架。我们视大语言模型为一面“生成式反射镜”：当我们刻意移除情感驱动参数后，模型生成的文本虽逻辑通顺，却丧失了梁启超特有的“笔端常带感情”的历史神韵。这一“反事实”实验直观地证明了情感与身体经验是历

史人物思想运作的底层代码。若忽略这些非理性要素，我们对历史理性的理解就是残缺的。

其二，在方法论上，实现了历史学家“同情之理解”的外显化与公共化。传统史学中的“文本共情”往往依赖学者个人的天才与直觉，处于不可言说的“黑箱”之中。本研究将这种隐秘的心理过程拆解为可公开检视的“提示词工程”与认知模型。这意味着，史学解释不再是个人的独断，而变成了一个可被同行重复检验、批评与修正的透明过程。在这里，AI不是替代者，而是促使研究者将模糊直觉转化为精确逻辑的“校准器”。

第三，在认识论上，确立了“模拟偏差”用于发现新问题的“探针”价值。AI模拟的“不像”往往比“像”更有价值。若模型生成的文本与真实史料存在明显的违和感（如未能复现梁启超思想深处的某种张力），其实并非单纯的技术失误所致，而是暴露了我们现有解释框架的盲点——我们没能教给AI正确的逻辑，甚至自己也未曾真正参透。这种偏差将倒逼我们带着更尖锐的问题重返史料，从而形成一种“建模—模拟—分析—修正”的知识增量循环。

可以说，在“认知重构”的全流程中，人类学者的主导作用体现在每一个关键环节——从基于史料进行“认知考古”并构建初始模型，到将阐释转化为精确的“提示词工程”，再到对模拟结果的吻合与偏差进行批判性评估、理论修正。这就与简单的直接追问（zero-shot）构成了本质区别：后者得到的往往仅是大模型基于语料统计生成的“概率平均值”，而本研究构建的整套流程实则建立在控制变量逻辑之上的因果推断实验。我们设计的提示词本身，也可谓是本研究的学术成果。

本方法论的普适性体现在其对不同“认知—情感”类型思想家的适应能力上。笔者选择情感丰沛、充满强烈“躯体标记”的梁启超进行实证演练，以凸显传统研究难以处理的“具身”维度。然而，为初步验证本方法论跨人物的适用性，我们亦对另一位认知风格迥异的思想家胡适进行了初步的模拟测试。我们为其构建了一个以“杜威式实验主义”为核心的认知框架（其主导隐喻为“思想是解决问题的工具”“学术是整理材料的科学”），并将其情感基调设定为“审慎、平和、重证据”。将这些参数输入我们的“AI认知实验室”后，生成的文本在风格与论证上与梁启超的模拟结果呈现显著差异：其行文冷静、克制，多见“问题”“证据”“方法”等词汇，高度复现了胡适的

风格。^①

值得一提的是，“AI 认知实验室”模拟的成功率与三个关键要素密不可分——史料的丰富度、阐释模型的贯通性，以及参数化的精确度。首先，本方法论最适用于留下了大量富含个人情感与独特语言风格文本（如书信、日记、政论）的思想家，因为这为“认知考古”提供了必要的数据库。对仅能通过二手或程式化官方文献了解的人物，其模拟的有效性将大打折扣。其次，“认知重构”的成功在根本上取决于研究者自身阐释模型的深度与自洽性。AI 模拟如同“理论显微镜”，输出结果的质量直接反映了输入模型的质量，粗糙的阐释模型必然导致一个粗糙的模拟结果。最后，能否将深度的阐释转化为精确且可执行的“提示词工程”是决定模拟成败的“操作化”环节。因此，笔者所倡“认知重构”其实是一个关于史学研究自身严谨性的问题。它并非简单地测试 AI 的能力，而是在一个可控的环境中，测试我们对研究对象的理解深度。这也正是本方法论作为“认知辅助”工具的核心价值所在：它促使我们将模糊的“历史感”对象化，并在与 AI 的“对练”中，将内隐的理解转化为外显的可操作的指令，不断逼近一个更具解释力的模型，深化和检验自身的阐释。更直白地说，“认知重构”的真正价值在于提供了一个可定制的建模语言，其成功的前置要求在于研究者自身对研究对象进行深度“认知考古”，以及具备将其阐释“参数化”的能力。

正因如此，本方法论生成的并非一个封闭的结论，而是一份开放且可被持续推进的研究议程。未来的研究可以在以下三个方向展开。

其一，“认知张力”的建模。未来的研究可以超越对静态认知模型的识别，转而探究这些模型在思想家内心是如何相互冲突、博弈乃至融合的。例如，我们能否为 AI 设定一组相互矛盾的“价值公理”，并观察它在特定历史情境下，为了化解内在冲突，会生成怎样的“创造性谬误”或“妥协”？这或许能

^① 我们所写的“情感驱动”内容为：“你的思考与写作，始终被一种智识上的‘洁癖’与乐观精神所驱动。你深受杜威实验主义哲学的影响，对一切空疏、笼统、无证据的清谈玄说感到一种深刻的不耐烦。每当论及‘迷信’‘鬼话’‘差不多先生’等概念，你的笔端当冷静而犀利，务求割除一切含混不清的病灶。反之，凡是引向‘科学’‘证据’‘方法’‘问题’之概念，则让你涌起一种建设性的乐观与热情。你的核心动机，是要以‘科学方法’为灯，为纷乱的中国思想界照出一条清楚明白的道路。”该工作流的公开访问 URL 为：<https://udify.app/workflow/BYVLM9VPiiOPq115>。

帮助我们理解，为何康有为的“大同”理想与“保皇”立场能够诡异地并存。

其二，文化惯习的参数化。未来的一个核心挑战在于，我们能否通过引入特定的知识图谱或训练专门的“文化域大模型”，将这种弥散性的惯习转化为可供调用的参数。这要求历史学家与计算机科学家进行更深度的合作，将皮埃尔·布尔迪厄（Pierre Bourdieu）的社会学洞见转化为可计算的语言。

其三，重估非理性在历史叙事中的位置。传统的思想史写作本质上是理性化的重构。AI模拟的偏差提醒我们，历史的真实进程或许充满了由情感捷径、认知偏误和偶然性驱动的非理性拐点。我们可以发展出一套独特的史学语言，去描绘和理解这种充满了断裂与矛盾的“认知地形图”。

归根结底，本研究倡导的并非一种技术崇拜，而是一种审慎、自觉与真诚的史学实践。面对大语言模型这一强大的“他者”，历史学家的核心任务不再是简单地驱使其挖掘信息，而是将其塑造为一个能与我们进行智识“对练”的伙伴。在此“对练”中，AI的模拟便是对我们理论假设的“压力测试”，其“失败”之处暴露的并非机器的缺陷，而实是我们阐释框架的裂隙。^①因此，这一新路径的最终目标并非简单地理解历史，而是锻造出可反思与迭代、更具智识诚实性的历史理解本身。

（责任编辑：张梦晗）

^① 本方法论框架亦可拓展为一种“互动式学术”的新形态，如将AI工作流转化为面向公众的互动叙事游戏，让使用者亲身扮演历史人物，在两难抉择中体验其心智模式。这不仅能以更生动的方式传播学术成果，而且能启发关于历史“可能性”的思考。