

# Science最新：大模型是文化社会技术

原创 Henry Farrell等 集智俱乐部 2025年04月12日 20:18 上海



大模型正重塑我们的社会文化生态，但其本质常被误解。《Science》上的一篇最新文章，揭示大模型并非自主智能主体，而是一种革命性文化社会技术，类似于语言、印刷与市场制度。它以独特方式汇聚并重构人类信息，成为“人类社会人工系统”的新变体。这一重新定位的理论框架引导我们超越“机器接管”与“乌托邦助手”的狭隘辩论，转而关注更为现实且紧迫的问题：大模型将如何影响资源分配、文化多样性、科学进步以及权力结构的重构？

研究领域：文化社会技术、大语言模型、信息重构、跨学科研究、社会变革、集体协调、权力分配

Henry Farrell等 | 作者

彭晨 | 译者

HOME > SCIENCE > VOL. 387, NO. 6739 > LARGE AI MODELS ARE CULTURAL AND SOCIAL TECHNOLOGIES

✉ | POLICY FORUM | ARTIFICIAL INTELLIGENCE



## Large AI models are cultural and social technologies

Implications draw on the history of transformative information systems from the past

HENRY FARRELL, ALISON GOPNIK, COSMA SHALIZI, AND JAMES EVANS [Authors Info & Affiliations](#)

SCIENCE • 13 Mar 2025 • Vol 387, Issue 6739 • pp. 1153-1156 • DOI: 10.1126/science.adt9819

来源：<https://www.science.org/doi/full/10.1126/science.adt9819>

关于人工智能（AI）的争论大多集中在大模型是否是具备智能和自主能力的主体。一些人工智能研究人员和评论家推测，我们正处于创造通用人工智能（AGI）主体的边缘，这一前景既令人兴奋又引起焦虑。同时，人们也就大模型的文化和社会影响进行了广泛讨论，讨论主要聚焦于两个方面：一是这些系统在当前使用时所产生的直接影响；二是假设未来这些系统转变为AGI主体，甚至可能成为超智能AGI主体时，可能带来的潜在影响。然而，把大模型视为智能主体的论调从根本上来说是误导性的。结合社会与行为科学的理念与计算机科学的方法，可以帮助我们更准确地理解人工智能系统。大模型不应主要被视为智能主体，而应被看作是一种全新的文化和社会技术，使人们能够从其他人长期积累的信息中获益。

拓展阅读：梅拉妮·米歇尔Science刊文：“通用人工智能”本质之辩

大模型这一新技术融合了早期多种技术的重要特性。正如图像、文字、印刷、视频、互联网搜索等技术一样，大模型使人们能够获取其他人创造的信息。目前，大模型主要集中于语言、视觉以及多模态领域，依赖于互联网，将早期技术的产出成为机器可读的形式（machine-readable form）。但类似于经济市场、社会机构和其他社会技术，这些系统不仅广泛传播信息，还能以独特的方式对信息进行重组、转化和重构。用西蒙的话说，**大模型是“人类社会人工系统（artificial systems of human society）”的一种新变体，通过信息处理实现大规模人群的协调合作** [(1), p. 33].

**我们的核心观点不仅在于，这些技术创新像其他所有创新一样会带来文化和社会上的影响，相反，我们主张大型模型本身最好被理解为一种特殊类型的文化社会技术。**它们类似于过去的技术，如文字、印刷、市场、官僚体系和代议制民主。基于这一视角，我们可以提出一个独立的问题：**这些系统将会带来什么影响？**那些本身并非文化或社会性质的新技术，比如蒸汽和电力，也能产生文化影响，但真正的新型文化技术，例如维基百科的影响可能较为有限。然而，许多过去的文化和社会技术曾对社会产生深远而变革性的积极或消极影响，大模型很可能也将如此。

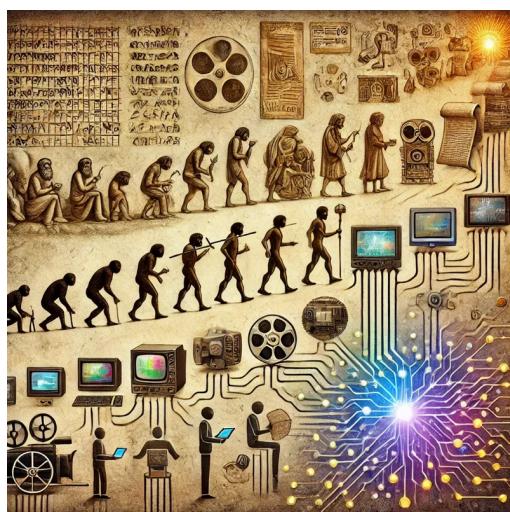
这些影响与其他重要通用技术（如蒸汽或电力）所带来的截然不同，它们也不同于我们对假设性通用人工智能可能产生影响的预期。通过审视历史上的文化社会技术及其影响，我们能够更好地理解人工智能模型的潜在风险与发展机遇，而非仅仅忧虑超级智能体的出现。

集智读书会上新：AI时代的学习：共探人类学习的复杂性，以跨学科的视角全方位审视智能时代的学习，AI如何重塑人类学习？学习如何在复杂系统中涌现与进化？技术向善还是技术缺陷？

人类自存在以来，就一直依赖文化传承。从语言本身开始，人类具备了从其他人经验中学习的独特能力，这种能力可以说是人类进化成功的奥秘，这种学习能力的重大技术变革也随之引发了剧烈的社会变革。语言交流之后出现了图像，接着是文字、印刷、电影和视频。随着越来越多的信息跨越更广阔的空间和时间得以传播，获取和组织这些信息的新方式也应运而生，从图书馆到报纸再到互联网搜索。这些发展对人类思想和社会产生了深远影响，无论是积极还是消极。例如，十八世纪印刷技术的进步使新思想得以迅速传播，对启蒙运动和法国大革命发挥了重要作用。而在2000年左右，当几乎所有文字、图像和动态影像信息都转化为数字格式时，一场标志性变革随之发生，这些信息可以被即时传输和无限复制。

自人类存在以来，我们也一直依赖社会制度来协调个体的信息收集和决策过程。这些制度本身可以被视为一种技术(1)。在现代，市场、民主制度和官僚体系尤为重要。经济学家弗里德里希·哈耶克 (Friedrich Hayek) 认为，市场的价格机制能够生成对极其复杂、难以捉摸的经济关系的动态信息(2)。生产者和消费者无需理解生产过程中的复杂性，只需关注价格，而价格则将大量细节压缩成一种简化而可用的表示。类似地，民主政体中的选举机制以相关方式将分散的民意汇聚为集体性质的法律和领导决策。人类学家斯科特 (Scott) (3)指出，无论是民主国家还是其他形式的国家，都通过构建分类和系统化信息的官僚体系来管理复杂社会。早在计算机出现之前，市场、民主和官僚体系就依赖于生成一种有损（不完整、选择性且不可逆）的但却有用的表示的机制。这些表示既依赖于个体的知识和决策，同时也超越了个体层面。价格、选举结果或诸如国内生产总值 (GDP) 之类的指标，都是对大量个体知识、价值观、偏好和行为的总结。同时，这些社会技术也能反过来影响个体的知识和决策。

市场、国家或官僚体系的抽象机制，正如文化媒介一样，能以至关重要的方式影响个体生活，有时甚至产生负面效果。例如，中央银行将金融经济的复杂性归约为几个关键变量，这表面上带来了金融稳定，但却使住房市场的不稳定性得以累积，而中央银行对此关注不足，从而引发了2008年全球金融危机(4)。同样，市场可能无法反映诸如有害碳排放等“外部性”信息。通过诸如碳税之类的手段将这些信息整合进价格体系中或许有帮助，但这需要国家采取行动。



人类在很大程度上依赖于这些文化和社会技术。然而，这些技术之所以能够存在，正是因为人类具备作为智能主体所特有的独特能力。人类和其他动物能够感知并应对不断变化的外部世界，构建该世界的新模型，随着证据的积累修正模型，进而设计新的目标。人类个体可以创造新的信念和价值观，并通过语言或印刷的方式将这些信念和价值观传递给他人。**文化和社会技术以强大的方式传递和组织这些信念和价值观，但如果失去这些个体的能力，这些技术就无法发挥作用。没有创新，模仿就失去了意义** (5)。

一些人工智能系统，例如机器人领域的系统，的确在试图实现类似的探索真理能力。从原理上讲，未来的某个时刻，人工系统完全有可能具备这种能力，毕竟人脑具备这些能力。但就目前而言，这类系统距离人类的这些能力还有很大差距。我们可以讨论现在是否应当对这些潜在的未来人工智能系统感到担忧，或者如果它们出现时我们该如何应对，但这与回答当前及近期大模型所产生的影响是不同的。

## 大模型

大模型，不同于那些更具备能动性的系统，在过去几年中取得了显著且出乎意料的进展，这使得它们成为当前人工智能讨论的焦点。这一进展引发了这样的观点：通过“规模效应 (scaling)”，也就是简单地采用现有设计，并增加它们使用的数据量和算力，便能在不久的将来实现通用人工智能 (AGI)。但大模型在根本上不同于智能主体，规模的扩大不会改变这一点。例如，“幻觉”现象在这些系统中普遍存在，因为它们没有真伪概念（尽管已有一些相应的解决办法），它们只是对文本和图像进行采样和生成。

大模型读书会资源：



AGI 通用人工智能读书会  
([https://pattern.swarma.org/study\\_group/33](https://pattern.swarma.org/study_group/33))，共同探讨大语言

模型与信息世界的智能、多视角下的人工智能、对齐技术与AGI安全性、AGI时代的未来社会等话题；

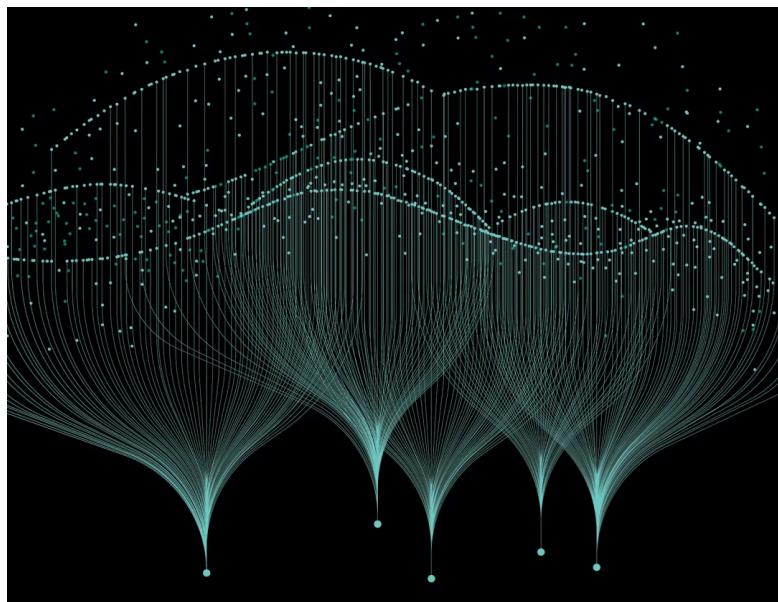


大模型2.0：融合学习与推理的新范式读书会 ([https://pattern.swarma.org/study\\_group/56](https://pattern.swarma.org/study_group/56))，深入探讨大模型推理新范式背后的核心技术和基本原理。



大语言模型与多智能体系统 ([https://pattern.swarma.org/study\\_group/38](https://pattern.swarma.org/study_group/38))，关注大模型与智能体的话题，涵盖包括大语言模型赋能下智能体之间的辩论、协作、模拟人类，以及实际场景中的多Agent协作等内容。

**大模型并非智能主体，而是以一种全新的方式融合了文化与社会技术的特性，对人类生产的大量复杂信息进行总结概括。**但这些系统不仅像图书馆目录、互联网搜索或维基百科那样对信息进行总结，它们还能像市场、国家以及官僚机构那样，在规模上以全新方式重新组织和重构这些信息的表示或“模拟”(1)。正如市场价格是对资源配置和底层信息的有损表示，政府统计数据和官僚分类也并不能完全反映底层人群的特征，大模型同样是其训练数据语料库的“有损JPEG”(6)。



**由于人类难以对大规模文化和社会技术有清晰的思考和认识，我们倾向于将它们视作具有主体性的存在。**故事是一种极具力量的信息传递方式，从篝火旁的传说说到小说再到视频游戏，它们都是通过创造富有表现力的虚构主体来实现这一目的，尽管听众清楚地知道这些主体并非真实存在。聊天机器人是赫拉克勒斯、阿南西和彼得兔的继承者。同样，人们很容易将市场和国家视为具有能动性的主体，甚至机构或公司也能拥有某种法律人格。

**然而，在那些类似主体的交互界面和拟人化的外表背后，大语言模型（LLM）和大型多模态模型实际上是统计模型，它们处理大量由人类生产的文本语料，将其分解成特定的词语，并估计长词序列的概率分布。**这种方式虽然对语言的表征并不完美，但却包含了关于其所总结的模式中大量丰富的信息。大语言模型能够预测序列中下一个出现的词，从而生成类似于人类生成的文本。大型多模态模型则以类似方式处理音频、图像和视频数据。大模型不仅抽象出了大量人类文化，还可以基于它开发各种新操作，例如，可以通过提示使大语言模型对其训练数据进行复杂转换：简单的论点可以进行华丽的隐喻表达，而华丽的辞藻也可以被凝练成朴实的语言。类似的技术还使得其他模型能够根据提示生成新的图片、歌曲和视频。一整套曾经海量、过于复杂、且混乱的文化信息，如今已变得易于大规模操作。

实际上，这些最前沿的AI系统不仅依赖于由人类生成和策划的大量文本和图像缓存，还依赖于其他形式的人类判断和知识。特别地，这些系统依靠来自人类反馈的强化学习（RLHF）或其变体：成千上万的工作人员对模型输出进行评分。它们还依赖于提示词工程，人类必须运用背景知识和创造力从模型中提取有用的信息。即使是最新的“思维链（chain of thought）”模型，通常也是从与人类用户的对话开始。

大模型从文本中提取一般统计模式所依靠的那些相对简单而强大的算法，并不是它们成功的关键。相反，**现代人工智能的基础是图书馆、互联网、成千上万的程序员，以及日益壮大的国际活跃用户群。**请求机器人帮助撰写求职信，实际上是与成千上万的前求职者和求职信撰写者，以及RLHF工作人员进行一种技术中介关系（technically mediated relationship）。

人工智能的话题讨论应该聚焦于这些新文化、社会技术所带来的挑战与机遇。我们现在拥有一种技术，对书面与图像文化的作用相当于大规模市场对经济、庞大官僚机构对社会的作用，甚至可与印刷曾经对语言产生的影响相提并论。那么，接下来会发生什么？正如过去的经济、组织和信息领域的“通用技术”一样，这些系统将对生产力产生影响 (7)，既能补充人类劳动，也能自动化以前只有人类才能完成的任务；同时还会影响资源分配，决定“谁获得什么” (8)。

然而，它们还会带来更广泛且深远的文化影响。我们尚不清楚这些影响是否会像印刷、市场或官僚机构等早期技术一样大，但将它们视为文化技术，反而会增强其潜在的影响力。无论是作为变革的原因还是结果，这些早期技术在18、19世纪广泛的社会转型中发挥了核心作用。所有这些技术，和大模型一样，都支持对信息的抽象处理，从而使得新型大规模操作得以实施。它们引发了关于错误信息和偏见传播、文化同质化或分裂以及权力和资源分配变动的合理担忧。新传播媒体的出现，包括印刷和电视，也伴随着对其可能传播错误信息和强化有害文化力量的合理担忧。同样，官僚机构和市场所采用的分类方案往往带有压迫性的前提假设。

与此同时，这些技术为重新组合信息和在全球范围内协调数百万人的行动创造了全新可能。对于大语言模型在社会、经济和政治影响方面的讨论，延续了对新型文化与社会技术由来已久的担忧与希望。要对这些讨论进行定位，既需要认识到新旧论点之间的共性，也需要勾勒出这些新兴且不断进化的技术的具体特征。

这种将大模型映射到文化社会之中正是社会科学的核心任务之一，而社会科学正是在工业革命及其影响引发的社会、经济和政治动荡中孕育而生的。社会科学家对过去这些技术影响的研究，可以帮助我们思考人工智能在社会层面上那些不那么明显的积极消极影响，同时探索如何设计新的人工智能系统，放大其积极效应减少消极影响。正如19世纪和20世纪媒体、市场以及官僚技术的扩展所产生的效应，造就了经济上的输家与赢家，取代了从文员、打字员到“人工计算机”等整个类别的工作者，今天的大模型也引发了人们对相关技术可能取代“知识工作者”的担忧。

除此之外，还有一些不那么明显的问题：**大模型究竟会使文化和社会趋同，还是会使其分化？**在历史背景下思考这一问题尤其富有启发性。当前的担忧与19至20世纪有关市场和官僚机构的争论十分相似。韦伯曾担忧(9)经济和官僚“理性化”带来的麻木与同质化后果，而密尔(10)则认为市场交换可以使参与者接触到多样的生活方式，从而缓和冲突（“温和商业”）。

大模型表现得非常出色，它们能够如实重现文本、图像和视频序列的实际概率分布。**本质上，它们在训练数据中最一般的情形下倾向于表现得最为准确，而在训练数据中罕见或全新情境下则可能准确性较差。**这种特性可能使大模型加剧韦伯所担忧的那种同质化现象。

另一方面，结合它们所总结的多样化文化视角，大模型可能会帮助我们设计出新方法。将这些视角进行整合与平衡，或许能为解决复杂问题提供更为精细有效的手段(11)。例如，我们可以构建类似“社会”的生态系统，其中不同视角（编码在不同的大模型中）彼此辩论，甚至交叉融合以产生混合视角(12)，或者识别出人类专业知识领域中的空白(13)，进而寻求有效的弥补。大模型在抽象化文本和图像中那些微妙的数据模式方面表现出惊人的能力，这表明这类技术能够发现穿插于人类知识和文化各领域的模式，甚至包括那些对单个个体来说无法觉察的模式。我们可能需要构建出新的系统，使大模型所反映的内容和所呈现的人格多样化，从而再现出类似于人类社会的分布和多样性。

这种系统多样化对科学进步或许尤为重要。形式科学本身就依赖于17世纪和18世纪新型文化技术的涌现——从咖啡馆和快速邮递到期刊和同行评审。人工智能技术有可能进一步加速科学进展，但这将取决于我们如何富有创造性地使用和重新思考这些技术。通过将众多文本、音频和图像中的不同视角相互联结，大模型可能使我们发现前所未有的联系，从而给科学和社会带来益处。**迄今为止，这些技术大多被训练为重复常规信息的小助手；而如果我们将它们视作探索未知领域的地图，或许可以打开新世界的大门。**

新的文化和社会技术还以一些不那么明显，但更有趣的方式影响着经济关系。**文化技术的发展引发了信息生产者与信息分发系统之间根本性的经济张力（economic tension）**。两者相辅相成：作家需要出版商，出版商也需要作家；但他们的经济受益却朝着相反的方向：信息分发者若能以低廉成本获取生产者的信息则能获利，而生产者若能以低成本将信息分发出去也能获利。这种张力一直是新文化技术的一个特征，数字信息分发的便捷与高效已使这一问题尤为尖锐，从地方报纸到学术期刊都可见一斑。然而，大模型以极快的速度、高效的处理能力和广泛的覆盖范围一次性处理所有可用信息，再加上这些模型的集中所有制，使得这一问题显得格外严重。权力的集中可能使系统所有者更容易在牺牲他人利益的情况下独享效率带来的好处。

同时，还存在一些关键的技术问题：大模型系统性的不完善在多大程度上可以被弥补？它们何时比基于人类知识工作者的系统的不完善更优越或劣势？这些技术问题不应掩盖更为关键的政治问题：哪些参与者有能力围绕自身利益动员组织？他们又将如何影响技术与组织能力的结合？技术领域的评论家常常将这些问题简化为机器与人类之间的单一对抗——要么进步力量战胜倒退的卢德派倾向（retrograde Luddite tendencies），要么人类成功抵制人工技术的非人道侵蚀。这种看法不仅未能充分理解计算机出现之前就已存在的复杂分配性斗争，而且也忽略了未来进步可能采取的多种不同路径，每种路径都有其独特的技术可能性与选择(8)。

在早期的社会和文化技术中，规范性和监管性制度在内的各种制度，相继出现以调节它们的影响。从印刷媒体的编辑、同行评审和诽谤法，到市场、民主和官僚体系中的选举法、存款保险及证券交易委员会，这些制度的效果各异，且需要不断修订。然而，这些制衡力量并非自发出现，而是技术内外部各方经过共同且持续的努力而形成的。

关于通用人工智能（AGI）的叙事——将大模型视作超智能主体的说法，不仅在科技界内外受到推广，无论是乐观的“热衷者”还是忧虑的“末日论者”都在宣传这一观点。这种叙事误解了这些模型的本质及其与过去技术变革之间的关系。但更为重要的是，**它有意转移了人们对这些技术真正带来的问题与机遇的关注，忽略了历史教给我们的如何确保收益大于成本的经验。**



当然，正如前文所提，未来可能会出现更类似于智能主体的假设性人工智能系统（hypothetical systems），我们或许会讨论如何应对这些假设系统，但大语言模型并不是这样的系统，就像图书馆借书目录或互联网一样。正如目录和互联网，大模型只是悠久文化和社会技术历史的一部分。

社会科学已经详细探讨了这段历史，形成了对过去技术剧变的理解。**将计算机科学与工程学与社会科学紧密结合，将有助于我们理解这段历史并运用这些教训。**大模型究竟会导致文化趋同还是分裂？它们会强化还是破坏人类探索真知的社会制度？在重塑政治经济结构的过程中，究竟谁将

得利，谁会损失？这些以及其他十分紧迫的问题，在讲大模型简单类比为人类主体的讨论中，都难以得到关注。

改变讨论的范式将带来更好的研究成果。如果双方都认识到大模型仅仅是新型的文化与社会技术，既不过分也不疏忽，那么社会科学家和计算机科学家之间的合作及结合各自优势将变得容易得多。计算机科学家可以将他们对这些系统运作机制的深刻理解，与社会科学家对其他大规模系统如何在过去重塑社会、政治和经济的认知相结合，既能完善现有的研究议程，也能发现新的研究方向。这将有助于纠正过去存在的混乱局面：**一方面，计算机科学家往往过于简化复杂社会现象(14)；另一方面，社会科学家又未能理解这些新技术复杂的运作机制。**

**这也将使人工智能政策讨论明确摆脱简单地将焦点放在“机器接管”的恐惧，与近未来人人拥有完美、可靠且高效的人工助手这一乌托邦式愿景之间的对抗上，大模型实际政策的影响无疑会有所不同。**就像市场和官僚机构一样，它们会使某些类型的知识比过去更加易得和易于处理，从而促使政策制定者关注那些可以衡量和观察到的新事物，而忽视那些不那么明显、较为混乱的部分。结果便是，如市场与媒体的历史案例所示，权力和影响力将向那些能够充分运用这些技术的人倾斜，而那些无法运用者则会日益边缘化。人工智能削弱了被其使用者以及数据提供者的地位，同时强化了人工智能专家和政策制定者的地位(14)。

最后，这样的思考方式可能会重塑人工智能的实践。工程师和计算机科学家已经意识到大模型偏见的问题，并在思考它们与伦理和正义之间的关系，但应当更进一步。这些系统将如何影响“谁获得什么”？它们对社会分化和融合将产生怎样的实际影响？大模型是否能够开发用来增强人类创造力，而非让我们变得迟钝？要找到这些问题的正确答案，不仅需要工程学的理解，还需要社会科学的洞察。将人工智能辩论的焦点从主体转向文化和社会技术，是构建这种跨学科理解的重要第一步(15)。

## 参考文献

1. H. Simon, *The Sciences of the Artificial* (MIT Press, 1996).
2. F. A. von Hayek, *Am. Econ. Rev.* 35, 519 (1945).
3. J. C. Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed* (Yale Univ. Press, 1998).
4. D. Davies, *The Unaccountability Machine* (Univ. Chicago Press, 2025).
5. E. Yiu, E. Kosoy, A. Gopnik, *Perspect. Psychol. Sci.* 19, 874 (2024).
6. T. Chiang, *New Yorker* 9 (2023).
7. C. Goldin, L. Katz, *Q. J. Econ.* 113, 693 (1998).
8. D. Acemoglu, S. Johnson, *Power and Progress: Our 1000 Year Struggle over Technology and Prosperity* (Hachette, 2023).
9. M. Weber, *Wissenschaft Als Beruf* (Duncker & Humblot, 1919).
10. J. S. Mill, *Principles of Political Economy* (Longmans and Green, 1920).
11. L. Hong, S. E. Page, *Proc. Natl. Acad. Sci. U.S.A.* 101, 16385 (2004).
12. S. Lai et al., *Proc. 41st Int. Conf. Mach. Learn.* 235, 25892 (2024).
13. J. Sourati, J. A. Evans, *Nat. Hum. Behav.* 7, 1682 (2023).
14. S. L. Blodgett, S. Barocas, H. Daumé, H. Wallach, arXiv:2005.14050 [cs.CL] (2020).

## 「AI时代的学习：共探学习的复杂性」主题读书会

在技术浪潮的冲击下，智能时代对人才培养的需求正发生根本性转变——学习已不再局限于简单的知识传递与记忆，当机器能够替代程式化技能，人类的创造力、批判性思维与跨界协作能力将成为核心竞争力；当知识更新周期以月甚至天为单位迭代，教育的使命不再是填鸭式灌输，而是培养终身学习者的自适应能力。

在此背景下，集智俱乐部联合江南大学教授王志军，北京师范大学教授崔光佐，翼鸥教育创始人宋军波，TalkingBrain 联合创始人林思恩，清华大学讲师方可，北京师范大学博士后郭玉娟，共同发起「AI时代的学习：共探学习的复杂性」主题读书会。希望通过汇聚教育学、系统科学、脑科学、计算机科学、社会学等多领域交叉视角，突破单一学科的局限，对人类社会未来学习发展形成更加全面深入的认识。



详情请见：[AI时代的学习：共探人类学习的复杂性](#)

### 推荐阅读

1. [压缩即智能：连接大语言模型与复杂性研究](#)
2. [多模态大语言模型 vs 人类：视觉认知能力的较量](#)
3. [以数据驱动的社交网络和多智能体模型模拟社会韧性](#)
4. [涌现动力学如何用来分析复杂系统？| 新课上线](#)
5. [AI时代的学习：共探人类学习的复杂性](#)
6. [探索者计划 | 集智俱乐部2025内容团队招募（全职&兼职）](#)

[点击“阅读原文”，报名读书会](#)

复杂科学前沿2025 76

复杂科学前沿2025 · 目录

[上一篇 · 稀疏有向网络上复杂系统的动力学平均场理论](#)

[阅读原文](#)