

人工智能最终会把人类带向何方？

原创张月红 [知识分子](#) 2024 年 12 月 31 日 10:05 北京

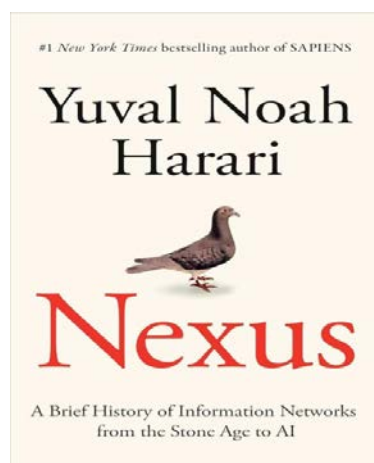
12.31 知识分子 The Intellectual

撰文 | 张月红

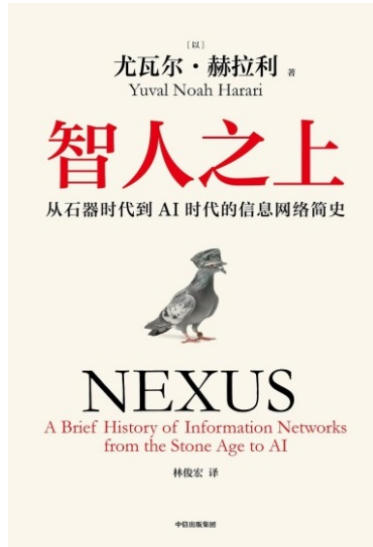
我在岁末听了 2024 几位诺贝尔奖得主的一场思想座谈会，主要议题是人工智能的发展与风险，物理获得主，人工智能教父辛顿教授在采访中率先发言说 GPT4 这类语言机器的知识储备已超过任何个体的人类，如人类大脑的连接点有一百万个；而这种机器已有 1 万亿个连接。因此我们人类面临着潜在的长、短期风险，短期是人工智能将被不良行为所利用，如虚假信息，网络犯罪，军方开发致命的自主武器等；长期风险如生成式人工智能一旦在某个时刻产生了自主学习和决策目标的能力，或将以人类为目标等。为防止其趋势，呼吁国际层面上的合作，研究其安全问题。为此，几位诺贝尔获奖者们均强调人类自身并不完美，呼吁强化科学与哲学的互补，提升公众对科学的信任，伦理学或在人工智能时代将成为最重要的一门学科，其中哲学家的作用不可缺失。

这两年里，我也一直在关注和阅读这一领域的文章，但尚未得到关于人工智能的进化对人类社会意味着什么的观察，以

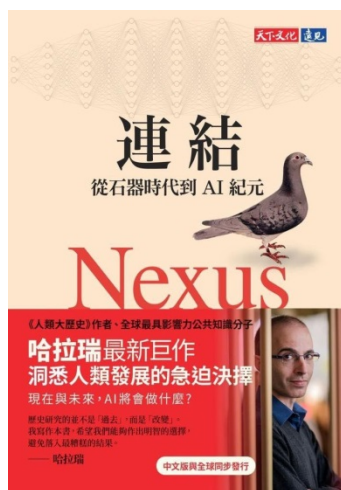
及它将带来特殊威胁的系统答案。但不可思议的是这几天被畅销书《人类简史》的作者尤瓦尔·赫拉利 (Yuval Noah Harari)在 2024 年 9 月出版的 ,并被译成多种文字全球“热”销的《Nexus》彻底‘黏’牢了。或许是职业的使然，我几乎也饶有兴致地翻看着一书三地、不同文种、不同出版社几乎同时出版的 3 个版本，详见三个版本的封面(下图)信息：



● 原版书名：《Nexue—A brief History of Information Netwroks from the Stone Age to AI》由美国企鹅兰登出版商 (Penguin Random House)于 2024 年 9 月 10 日出版 ;作者尤瓦尔·赫拉利(Yuval Noah Harari)于 1976 年出生于以色列海法，2002 年在牛津大学取得博士学位，目前是耶路撒冷希伯來大学历史系讲师，也是剑桥大学风险研究中心 (CSER) 特聘研究员。



- 中文简译书名为《智人之上：从石器时代到 AI 时代的信息网络简史》由中信出版集团于 2024 年 9 月 1 日出版；译者 林俊宏



- 中文繁体译书名为：《连接》從石器時代到 AI 紀元；由 2024 年 9 月 10 日遠見天下文化出版股份有限公司出版；译者林俊宏

有趣的是该书的中文简、繁体的译者同为台湾师范大学翻译研究所的林俊宏博士。他也是赫拉利的忠实粉丝，翻译了赫拉利的几部热著，如《人類大歷史》、《人類大命運》、《 21

世紀的 21 堂課》、以及《人類大歷史：知識漫畫》與《少年讀人類的故事》几套叢書。

史上最大的一场信息革命

赫拉利不仅是畅销书的作家，也是历史学家与哲学家。

赫拉利在本书研究的一个关键点是，信息至少需要两点之间的联系，因此，信息在本质上被组织成连接的网络，一些网络变得越大，越占主导地位，最终导致了机器智能和人工智能的发展。赫拉利与我们一样，对此既敬畏又害怕。故而他写本书的初衷是围绕着历史上最大的一场信息革命，即生成式人工智能展开，用意是希望人们要理解这场革命，必须与过去的信息革命做比较。而历史研究的真正要点也并非单纯看“过去”，而是看“变化”。历史能告诉我们什么是不变的、什么是变的，以及事物是如何变化的。这不仅与信息革命相关，也与其他类型的历史变革都息息相关。

依照赫拉利的提示，我粗读地读了几个版本后，略感作者有两条写作脉络：一是以历史视角佐证每一次新信息技术的发明总能促成重大的历史变革，因为信息最重要的作用就是编织网络，而非呈现既有现实。从石器时代到《圣经》，从欧

洲的猎巫行动、现代的斯大林主义、纳粹主义，以及今天民粹主义复苏等，可以启发读者在历史教训中思考信息与真相、官僚与神话、智慧与权力间的复杂关系均与信息网络脱不了干系；二是为人类将与机器一起“进化”描绘了一条探索之路，换句话说，通过了解人工智能是如何工作的物理原理，人们可以开发出有益的应用，避免出现反乌托邦的未来。也如作者简言之：‘写这本书的目的是希望为人工智能革命提供一个比较准确的历史观。’

我自知欠缺把控全书要点的能力，但略有收获的是我捕捉到作者从人工智能论及科学出版的一些观点，如我在全书搜索到 15 处谈及同行评审的期刊，若用 Journal 一词搜索英文原版，那就更多了。更有意思的是这两天我还从美国学术出版学会的学术厨房中看到了 2 位行业大佬将此书评为“本年度最新最好的图书，并极尽所能地捧《Nexue》是一本了不起的综合著作，对从事学术交流的人来说更是一本非常重要的基础著作。” [1]

01 “故事”理论的说服力

我仔细品味了一位美国同行的评论，他说本书的“故事”理论很有说服力。顺此我很快囫圇吞枣地看了第 3 章，赫拉利

的确提出了一个独到观点，即人类创造的第一个信息技术就是故事（the first information technology developed by humans, is the story），并佐证很多国家最早都出现在诗人的想象之中。为了验证他的理论，我特地查溯了公元前 25 世纪埃及金字塔的铭文；中国先秦时代的《诗经》；西亚美索不达米亚的《吉尔伽美什史诗》；梵语文学中的《吠陀》；荷马史诗，包括《伊利亚特》和《奥德赛》等等。不禁被作者的故事理论惊讶到人类分享故事信息有一种遗传优势，不仅利于交流，也助于人类进步，还帮助人类从家庭、部落和物种层面取得成功，让人类成为地球上最先进、最具统治力的物种。

赫拉利顺次探索了从写作到出版，再到互联网的技术是如何加速了人类交流故事的能力。然后他在故事和事实之间话锋一转，迅速指出一个重要的区别，即事实远不如故事那么有影响力，如我们人类能处理事实，但我们更能被很多故事所驱动。如一个人可以记住整部《古兰经》、整部《圣经》、数千页的《伊利亚特》，但一般不能记住会计报告或人口普查中的几页表格，这本质上是信息处理的过程，于是一种非生物的信息技术，即书面文件，到平板电脑的标记，后来发展成为电脑领域，计算机比人类在数百万、数十亿、数万亿个数据点中追踪和计算发现模式成功演化为现实。这也是他

提出故事理论的哲学要点，即人类常常被故事所驱动，这反映出我们人类的内在特征，如此类推，故事及我们传达故事的方式对我们的生活比事实更有吸引力。

接着，赫拉利将信息结构转移到人工智能新世界，他认为与印刷等先前的信息技术不同，生成式人工智能不仅是人类交流的被动工具，也将成为系统中的全面参与者。以前，将一个文档连接到另一个文档的唯一方法是通过人工进行管理。如今机器能够创造内容，吸收内容，并对其做出反应。然后，他将这一观点扩展到对政治与社会讨论的范畴，这对媒体、治理和更广泛的社会潜在影响是令人着迷的。

赫拉利特别指出，本书所关注的不是人工智能系统的积极潜力，因为已有许多支持者在关注人工智能系统对生产力提高的好处。国际同业人认为本书对探索学术信息发布和人类理解新领域至关重要，如作者提到诸多挑战的一个例子是，在公共对话的生态系统中，故事的力量有能力克服理性和寻求事实的探索，就如今天，当我们看到世界各地的公众对科学探索真理的信任度下降时，这个问题尤为突出。随之而来，人工智能系统创造了看似合理但缺乏事实依据的内容与能力，或将传为故事而产生更广泛重大影响，所以说事实与故事始

终相伴而行，真不是非黑即白能说的清的。

美国同行又说“在一个以机器为媒介的内容时代，反思《Nexus》中的信息框架如何影响 STM 学术出版区，让我们看到了几个相互联系的线索。学术出版与其他领域的区别在于它基于科学方法追求事实记录。学术出版探索和传播新思想，就像科学本身一样，寻求建立在一种内在的自我纠正方法之上。在我们的学术出版社区中，国际出版伦理委员会（COPE）、国际科学技术医学学会（STM）、美国国家信息标准局（NISO）等其他机构都在改进撤稿流程以维护学术诚信的努力就是例证。也如赫拉利在书中最有力地强调，人类在历史进程中不断地治愈和反思，然后在认识错误的过程中调整路线，是人类最强大的能力。在他的框架中，自我修正的能力已经并将继续阻止灾难的发生” [1]

02 科学的自我修正机制

‘自我修正’ 这个短语在本书中或是除了信息，人工智能外，出现几率很高的词组，我查过几个版本中都过百次，我也突然顿悟出作者的最终用意，无论人类多会讲故事，历史有多少次的无奈反省，超越人类记忆的智能储备有多么能干，人类社会最终都有自我纠错的能力。无论是社会体制，企业

文化，尤其是科学界。按照作者的说法，科学群体是最具有纠错能力的群体。科学革命扮演的重要角色就是筛选机制。如此类推，人工智能的飞跃发展，自我修正模式也在相向而行中。

作者用了大量篇幅谈了科学的自我修正机制。尽管我对他的有些说辞不太认同，但我偏爱这个论题的阐述，如果用照我自己的话去诠释其内容，不如他直接引用他的原文更有深度，故此选编了几段原文如下：

“科学要想加快步伐，科学家就得相信远方同仁发布的信息。从某种机制中都能看到，科学家虽然素未谋面，却愿意相信彼此的研究成果。第一是各种科学协会的作用，例如 1660 年成立的英国皇家学会和 1666 年成立的法国科学院。第二是各种科学期刊的创办，如《皇家学会哲学学报》（*Philosophical Transactions of the Royal Society*, 1665）、《皇家科学院刊》（*Histoire de l'Académie Royale des Sciences*, 1699）。第三是多家科学出版商的出现，如策划了《百科全书》（*Encyclopédie*, 1751—1772）这类作品的机构。这些机制是以实证为基础来整理信息，当一篇论文投向《皇家学会哲学学报》时，编辑主要的问题不是会

有多少人愿意付费阅读这篇文章，而是有什么证据来证明这是真的。

《皇家学会哲学学报》的编辑不像那些猎巫专家，法国科学院也不像天主教会，并没有庞大的领地或预算。但这些科学机构就是因为这种特别的原因而得到信赖，逐渐积累其影响力。科学机构之所以能取得权威，是因为它们有强大的自我修正机制，能够揭露并修正自身的错误。科学革命真正的引擎正是这些自我修正机制，而不是印刷技术！

换句话说，人类是因为发现了自己的无知，才推动了科学革命。那些信奉某本经典的宗教，会觉得它们已经取得了无懈可击的知识来源。基督徒有《圣经》，穆斯林有《古兰经》，印度教徒有《吠陀经》，佛教徒有《佛经》，但科学文化并没有这样的神圣经典，也从未宣称某位科学泰斗是绝不会犯错的先知、圣人或天才。科学革命从一开始就不相信有绝对正确这种事，为其打造的信息网络也认为错误本就不可避免。当然，大家经常谈到哥白尼、达尔文与爱因斯坦如何有天赋，但并不会说其中任何一位绝对完美无缺。这些科学家都犯过错误，即便是最著名的科学著作肯定也会有错误与疏漏。即便是天才，也难免受到证真偏差的影响，所以并不能相信天才肯定能纠正自身的错误。科学是一项团队工作，需要团

队互相配合,不可能只靠单个科学家或者所谓绝对正确的书。当然,机构也可能出错。但科学机构与宗教机构的不同之处在于科学激励怀疑与创新,从不激励遵循与顺从。科学制度与阴谋论的不同之处在于,科学鼓励的是怀疑自己,而阴谋论者常常对众人既有的共识表示怀疑,认为他们自己的信念是毋庸置疑的,从而落入了证真偏差的陷阱。科学的标志不仅是怀疑,而且自我怀疑,在所有科学机构的核心都能看到强大的自我修正机制。科学机构确实对某些理论(如量子力学或演化论)的正确性达成了广泛共识,但只是因为这些理论成功顶住了一波波强力挑战,而且质疑者除了外部人士,也有机构的内部成员。’

另外,在出现重大错误与罪行时,科学机构制度愿意承认是制度本身出了问题。如在19世纪、20世纪,生物学、人类学、历史学的科学研究常常有制度性的种族与性别歧视,现在的大学课程、专业期刊都在诚实地揭露这些问题。

科学机构的聘任与晋升,遵守的原则是“不发表就淘汰”,尽管这个默认的规则常被诟病,但要想在优秀的期刊上发表论文,你必须揭露现有理论的某种错误,或是发现某些前辈不知道的内容。没有人会因为忠实地重复过去学者的话、反

对所有新的科学理论，就能拿到诺贝尔奖。

科学家也是人，同样会有各种各样偏见，这点说得没错，但因为科学机构制度拥有自我修正机制，从而让这些偏见得以克服。只要能提供充分的实证证据，常常只需要数十年，就能让非正统理论推翻传统概念，成为新的共识。（例如：谢赫特曼因为他的发现荣获 2011 年诺贝尔化学奖。尽管他的发现极具争议，但最终迫使科学家重新思考他们对于物质本质的看法。”）

03 我们需要时间学习

年末看到美国科学和技术政策办公室（OSTP）提出 2025 年 7 项优先研究事项（Multi-Agency Priority Guidance），第一条就是‘推进可信赖的人工智能（AI）技术，保护人类的权利和利益，安全，并利用它来加速国家的进步。’这也旁证我们与美国的科技竞争中人工智能的研究已在首选中。

如本文开篇所述，如今全球科技界都把生成式人工智能的发展作为第一议题来讨论。如此联想今年 5 月看到《科学美国人》的编辑克里斯蒂 Aschwanden 主持的一场讨论“不确定性是科学的超能力（Uncertainty Is Science's

Superpower) [2]。直觉作者赫拉利也正是从 GPT 的不确定性中找到了写作《Nexus》的灵感和创造力。醉翁之意是启发我们从历史到哲学，最后落脚在科学是一个强大的工具，帮助我们理解周围世界复杂性，它在做决定时非常有用。如果我们越能接受所有科学背后的不确定性，如面临生成式人工智能或某天将超越我们人类的意识，我们就越能更好地面对这种不确定性，根据新证据更新我们的信念。

所有新东西都是从一个未知的地方起始，研究人员，包括科学出版人员应该知道如何驾驭这些不确定性。例如《生物设计与制造》2024 年 8 月在日本东京大学召开的国际学术年会上，主编崔占峰院士直接在大会报告中向学者与期刊抛出了 2 个尖锐的问题“研究和期刊的发文缺少了什么 (What is missing BDM?) ; 人工智能和分布式生物制造是否为先进制造的一部分? (Intelligent and distributed bio-manufacturing, part of Advanced Manufacturing?) 这或许就是科学家与时俱进，反向思维的模式，即人工智能的知识渐进中总要受到新证据的影响，但我们可以中级的理解去尝试做很多事情。

总之，我们不是无所不知，以开放的心态，尝试中可以帮助我们弄清楚一些问题。

人工智能对生产力的提高已是毋庸置疑，而符合人类社会价值观与利益的人工智能时代呼吁健全伦理学这门最重要的学问。

参考文献：

[1]<https://scholarlykitchen.sspnet.org/2024/12/04/chefs-selections-best-books-read-and-favorite-cultural-creations-during-2024-part-3/>

[2]<https://www.scientificamerican.com/author/christie-aschwanden/>

作者简介：

张月红，浙江大学《生物设计与制造》负责人

