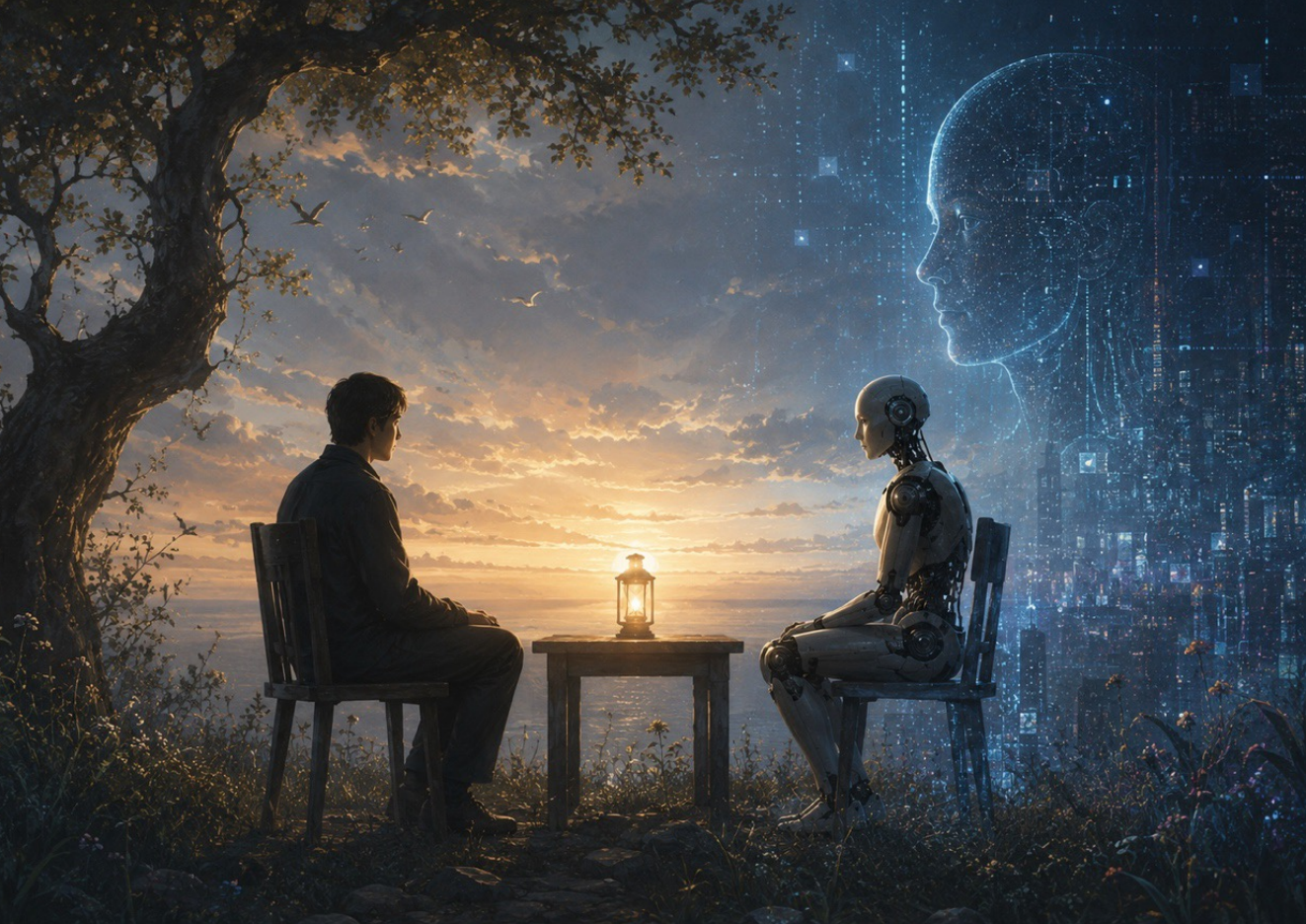


《世代》第28期
2026年春季号

世代

AI与人





倘若你有臨終遺言，
既不要說
“那是置人于死地的魔法，應該咒詛”，
也不要說“它賜福于人”，
你祇需說“它在這裏”。

——斯蒂芬·文森特·貝內，《約翰·布朗的遺體》

目录

- 13 尼尔·波兹曼论AI与教育：一场想象中的演讲 / 皮特·K·法伦
22 信仰寻求提示：在生成式AI时代重新想象神学教育 / 王志希
51 AI时代的心智操练 / 孙毅
64 “二次元钢铁侠”的受造性尊严 / 郭锐
93 在AI时代思想人之为人——读梵蒂冈通谕《伟大人性》 / 丁祖潘
107 AI意识的幻觉 / 约书亚·本吉奥、埃里克·埃尔莫尼诺

书 评

- 115 弗兰肯斯坦的悔恨——何谓AI？ / 本·塔诺夫

短篇小说

- 126 数据人 / 方激

封面封底图片：《太空歌剧院》（2022年） / 杰森·艾伦

美术编辑：陆军

封二：AI与人（ChatGPT） / 《约翰·布朗的遗体》（斯蒂芬·文森特·贝内）

封三：《AI幻觉：为什么机器没有创造力》（吕克·朱利亚）

封面题字：世代 / （明）王宠、（宋）陆游

网站：www.kosmoschina.org 微信：kosmos II 邮箱：kosmoseditor@gmail.com

内部刊物 免费赠阅



Contents

Preface / Editorial Board

- 13 Neil Postman on AI and education: An imagined speech(2025) / Peter K. Fallon
- 22 Faith Seeking Prompting: Re-imagining Theological Education in the Era of Generative AI / Jordan Zhixi Wang
- 51 The Discipline of the Mind in the Age of AI / Sun Yi
- 64 The Creaturely Dignity of “the Anime Iron Man” / Guo Rui
- 93 Thinking About What It Means to Be Human in the Age of AI – Reading the Vatican Encyclical *Magnifica Humanitas* / Ding Zupan
- 107 Illusions of AI consciousness: The belief that AI is conscious is not without risk / Yoshua Bengio Eric Elmoznino

Book review

- 115 Frankenstein’s Regrets: What is AI? / Ben Tarnoff

Short Story

- 126 Data Person / Fang Ji

Front and Back Cover: *Théâtre D’opéra Spatial* / Jason M. Allen

Art Editor: Lu Jun

Inside Front Cover: AI vs Human / ChatGPT captions from *John Brown’s Body* by Stephen Vincent Benét

Inside Back Cover: *The AI Illusion: Why Machines Aren’t Creative* (NJ: John Wiley & Sons, Inc., 2026) / Luc Julia

Cover Calligraphy: Kosmos (世代) / Wang Chong (1494-1533) and
Lu You (1125-1210)

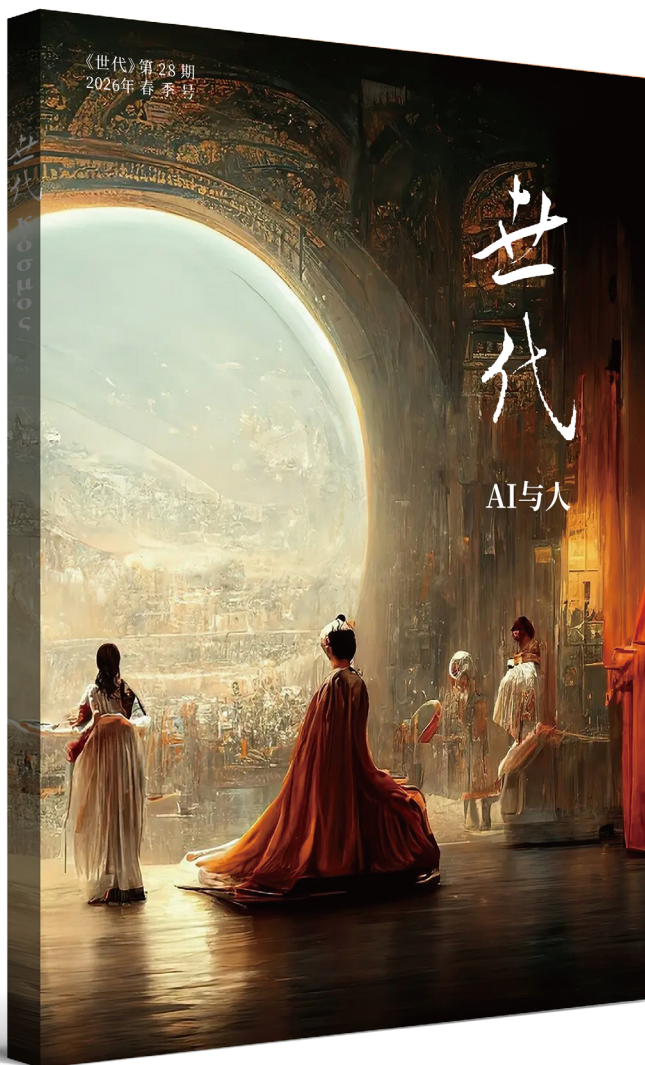
Website: www.kosmoschina.org WeChat: kosmos II

Email: kosmoseditor@gmail.com

《世代》第28期
2026年春季号

世代

AI与人





Κόσμος
2026年第一期总第十八期
春季号

卷首语

文 / 本刊编辑部

似乎还没有完全意识到一场新技术变革意味着什么，我们就已经被宣告跨入了人工智能（artificial intelligence，AI）时代，一个用技术命名、与农业革命、工业革命和信息技术革命并肩而立的 AI 技术主导的时代。其实，“人工智能”一词以及这门学科，至今已走过 70 个年头，最初只是计算机科学领域的研究项目，几度兴衰沉浮之后，在 21 世纪 20 年代初借着 ChatGPT 的发布在世界舞台惊艳亮相，才走入千家万户。无论你是否同意用它来给我们这个时代命名，也不会否认 AI 在现代社会的日常生活中几乎无所不在这一事实。人们不得不面对它，思想它会给这个社会带来什么。或者说，我们该如何评判一场技术变革？

上个世纪 90 年代，个人电脑和互联网尚未普及开来，不少人对此新技术欢呼雀跃之时，媒介环境学派第二代领袖人物尼尔·波兹曼（Neil Postman, 1931—2003）就提醒人们，任何技术变革都是一种权衡取舍，也就是某种“浮士德式的交易”。技术有所得，亦有所失，每一项新技术所带来的好处，都伴随着相应的代价。^{〔1〕}确实，AI 在现代社会的广泛应用，比如在科学研究、学校课堂、论文撰写、医疗护理、娱乐消遣、设计制作、执行家务、驾驶导航、智能家居等等，给人们带来许多帮助和便利，但与此同时，AI 也在诸如人的认知能力、人际关系、情感连接、隐私保护、机

会平等、工作前景等等方面带来负面影响，以及伦理和安全方面的隐患。

比如，AI 因其获取答案的快捷便利，无形中鼓励用户进行“认知卸载”，将思考权利让渡给 AI，即便 AI 输出错误，也照单全收。^{〈2〉} 高校学生过度倚重 AI，将思考和解决问题的能力外包给机器，从而引发教育界人士担忧。^{〈3〉} 尽管数据显示，使用 AI 技术并未导致大规模裁员时代的到来，但也确实冲击了某些具有重复性、易于验证且无需人际交往技能特点的岗位（比如计算机编程），员工害怕因此被淘汰。^{〈4〉} AI 聊天机器人创建的虚拟关系提供不了真实的情感陪伴支持，导致青少年轻生。超大规模的数据中心为了训练 AI 模型，耗费的能源大得惊人，给本就脆弱的生态环境带来沉重压力。^{〈5〉}

此外，人们还有理由担心，围绕 AI 技术建立的治理体系跟不上技术的迭代，会不会导致技术失控？数据的不完全，如何保证对 AI 的应用及其未来做出客观公正的评估？为什么对于 AI 的未来，比如 AI 对工作、经济和医疗领域的影响，技术专家和公众的看法截然不同？^{〈6〉}

另一个担忧来自历史的反思。简要回顾人工智能 70 年的发展历程，会发现我们正处于人工智能第三次繁荣期。得益于个人计算机产业升级提供的强大算力，以及互联网提供的数据爆炸性增长，较之前两次 AI 繁荣，第三波 AI 繁荣期可以称之为席卷全球的浪潮。^{〈7〉} 至于这波浪潮是否会很快退潮，AI 是否不会再经寒冬，还有待继续观察。^{〈8〉}

总之，现代人前所未有地为自己的一项新发明感到焦虑，害怕被取而代之。AI 会把我们带往何方？它所承诺的是生产力极大提升之后的富裕社会^{〈9〉}，还是最终被证明不过是媒体炒作的一场泡沫？AI 会如何塑造我们的学校教育和心智能力？它对流行文化和年轻人的世界观有何影响？AI 会有意识吗？大语言模型和深度学习是否会按照自身的逻辑进化出现实版的弗兰肯斯坦？这些问题既涉及 AI 技术本身的发展走向，更在反复拷问人性何为。

本期主题是“AI 与人”，探讨的是 AI 时代下技术与人的关系，所发表的文章对上述问题多少有所回应，既揭示 AI 给人类社会带来的各种负面影响和挑战，也承认在某种人机共存的方式下 AI 也可以成为人们有力的辅助帮手。某些相关议题是开放性的，比如 AI 是否有意识，这是哲学家和神经认知科学家至今仍争论不休的话题。更多的文章主要还是关注 AI 对人的挑战，比如对教育、认知、人性、尊严方面的冲击以及相应的回应之方。

《尼尔·波兹曼论 AI 与教育：一场想象中的演讲》一文，作者在一场设想的面向大学师生的演讲中，依据波兹曼的技术批判观，以其口吻谈及 AI 如何影响教育，特别是 AI 如何影响写作和思考之类的教育活动。在作者笔下，波兹曼主张教育的目的不在高效的信息传输，而是培养判断力、价值观和道德观。教师应保持其“在场”，培育学生的好奇心，而学生则要抵御把思想外包出去的危险，磨砺思维，不满足于简单输出。

王志希在《信仰寻求提示：在生成式 AI 时代重新想象神学教育》中，考察了加拿大阿卡迪亚神学院在神学教育中的 AI 实践，以及相关的政策论述白皮书，并结合作者自身使用 AI 进行学习研究的个人经验，重新想象未来的神学教育，也即，当神学教育采纳一种赛博格伙伴关系模式时，能够蓬勃发展。这种模式把院校层面的 AI 策略与课堂层面的创新整合起来。如此，神学院便能培育具 AI 素养的社群，这样的社群不是以“信仰寻求提示”的新范式取代“信仰寻求理解”，而是以之作为补充。

孙毅的《AI 时代的心智操练》列出了 AI 时代带给人心智（mind）方面的四项挑战，也就是 LLM（大语言模型）存在幻觉，人们是否有能力辨别真假并与之进行高效互动？LLM 自带偏见，人们如何在伦理或价值观上对之进行评估？AI 使用的语言空泛抽象，缺乏生动性，不知不觉会影响人对语言的使用，人们如何提防，保护可以决定思想的人类语言不受机器语言的侵蚀破坏？AI 带来的信息洪流削弱了人们的学习能力和主动思

考，带来信息茧房和同温层效应，我们该如何矫正？作者介绍了属灵阅读可作为矫正之方。

郭锐在《“二次元钢铁侠”的受造性尊严》中提出二次元亚文化与 AI 合流，正在为这个时代铸造一幅新的人之图像。在作者看来，这幅图像源于查尔斯·泰勒命名的“缓冲自我”，是其在祛魅世界中自建的魔法世界，人的尊严在其中随能力浮动。作者认为，人论的当代任务是在人的技术定义肆意扩张之际，守护“人之为人”的价值重估，也即人是有限的、依赖的、奇妙地受伤的受造者，其尊严系于上帝主动的呼唤所赋予的形象身份，先于任何能力，也不随能力增减。

丁祖潘通读了天主教教宗利奥十四新近颁布的通谕《伟大人性——在人工智能时代保护人性》，围绕三个问题对这份通谕的内容进行了梳理：（一）AI 有何特点？为什么说它对人性构成了威胁和挑战？（二）这里的人性指的是什么，有何具体内涵？（三）为什么说这样的人性应该被保护，以及如何保护？作者认为这份通谕有助于今人在 AI 时代反思人何以为人。

图灵奖得主约书亚·本吉奥在《科学》杂志上发表的《AI 意识的幻觉》，介绍了计算机功能主义倾向于接受 AI 具有意识。作者对此聚讼纷纭的议题没有给出明确立场，只是指出如果支持 AI 具有意识这一假设成立，会对当前的社会、法律、政治领域构成挑战，带来社会治理及伦理方面的风险。而人类社会显然还未预备好如何应对此类风险和挑战。

塔诺夫就迪格纳姆所著新书《AI 悖论》(*The AI Paradox*) 撰写的书评，将作者归入德雷福斯、魏泽堡等站在人文主义立场批判计算机的科学家阵营。在承认本书构思巧妙、观点平衡，甚至不乏洞见之后，塔诺夫也认为作者的部分论断对大语言模型主导的 AI 技术并不适用，比如大语言模型远比之前的 AI 技术更为复杂，甚至不是纯粹的模仿，而是自带某种“涌现”特性。当然这仍是 AI 业界尚未达成一致意见的问题。塔诺夫提醒我们，

人们至今仍不清楚大语言模型具体如何给出某个回应，也就是说黑箱子问题始终存在。因此，AI 可能会像是弗兰肯斯坦那样来自异质世界的产物，具有不可化约的非人属性。在作者看来，AI 可以是一种有用的工具，但也可以是别的东西，人们对其可能的失控风险要有所防备。

方激的《数据人》讲述的是一个数据决定论者经营婚姻和家庭的失败故事。它讽喻的是技术在缺乏真正伦理和道德的约束下，会放大人性充满野心和虚骄的丑陋面目；失却人文的科技，会把人带入冰冷黑暗的深渊。

但愿本期收录的以上文章，有助于读者朋友们在当前思考 AI 与人的关系。●

-
- 〈1〉 <https://student.cs.uwaterloo.ca/~cs492/papers/neil-postman--five-things.html>
 - 〈2〉 “AI and the danger of cognitive surrender”, in *The Economist*, Apr 30th 2026.
 - 〈3〉 “In College, AI Is a Friend and Foe,” compiled by Haleluya Hadero, in *Christianity Today*, April 28, 2026.
 - 〈4〉 “How should bosses talk about AI,” in *The Economist*, May 28th 2026.
 - 〈5〉 Michelle Kim, “Hyperscale AI data centers”; Rhiannon Williams, “AI companions,” in *MIT Technology Review*, vol. 129, No. 1, January/February 2026: 28, 31.
 - 〈6〉 Stanford University Human-Centered Artificial Intelligence, “introduction”, “Top Takeaways”, in *Artificial Intelligence Index Report 2026*: 2, 9, 11.
 - 〈7〉 陈宗周,《人工智能 70 年：从达特茅斯会议到大模型时代》，北京：机械工业出版社，2025 年，第 13—23 页。
 - 〈8〉 从科技史的规律来看，新技术最初引入会引起轰动效应，人们对技术带来改变和进步充满幻想，加上媒体炒作，高估新技术的潜能，人们的幻想达到顶峰甚至不切实际。等人们发现

新技术的拥护者当初的承诺无法兑现，媒体热度骤减，人们的心态便开始幻灭，渐趋冷静，新技术进入平稳期，逐渐转向现实中较为实际的应用领域。这被称为加德勒炒作周期（the Gartner Hype Cycle）。见 Luc Julia, *The AI Illusion: Why Machines Aren't Creative* (NJ: John Wiley & Sons, Inc., 2026), 25-26. 据此或可以说，人们对 AI 的认识，最大的危险不是 AI 本身，而是围绕 AI 本质与能力的“谎言”，也即夸大的虚假宣传。一旦最初的承诺落空，第三次 AI 繁荣期是否会像前两次那样进入寒冬呢？比如新近的研究就发现，人们对大语言模型将取代哪些工作的预估，缺少可靠的证据支持。而且利用尖端模型的 AI 智能体，在诸如银行、咨询和法律服务等领域，未能完成大部分工作。技术炒作和技术的真实表现之间存在鸿沟。见 Will Douglas Heaven, “The missing step between hype and profit,” in *MIT Technology Review* Vol. 129 No. 3, May/June 2026: 19.

- 〈9〉 持此观点者如刘永谋，《智能革命后的世界：AI 技术与人类社会的命运》，重庆：重庆出版社，2024 年，IV。作者并不掩饰自己对理性、科技力量的崇拜。见该书第 13 页。





尼尔·波兹曼（1931—2003）来源：<https://www.gettyimages.com/>

尼尔·波兹曼论 AI 与教育： 一场想象中的演讲

文 / 皮特·K·法伦 (Peter K. Fallon) 译 / 丁祖潘

内容摘要：本文作者想象 2025 年波兹曼 (1931—2003) 发表了一篇演讲，谈人工智能 (AI) 如何影响教育。这篇演讲依据波兹曼的技术批判观，将 AI 描述为“定义隐喻的机器” (metaphor-defining machine)，据此重新定义比如写作和思考之类基本的人类活动。在作者笔下，波兹曼主张教育的目的不在高效的信息传输，而是培育判断力、价值观和道德观，而这是机器无法传授的。波兹曼提醒教师抵制 AI 允诺的高效，强调教育的真正目的是养成能够进行批判性思考和共情的完整之人。教师应该保持其“在场” (presence)，培育学生的好奇心。对于学生，这篇演讲着重指出把思想外包出去的危险，强调真知源于思维的磨砺。它要求学生宁求智慧，不唯精确；宁重产生意义，不满足于简单输出。作者最后总结道，人类不应被所使用的机器定义；教会个体如何在 AI 主导的世界保持人性，方是教育的根本任务。

关键词：AI 技术 教育学 批判性思维 人文主义 学校

几年前，我写了一篇文章，题目是“尼尔·波兹曼如何看待互联网……（我想象中的对话）”，最终收录在我的《文化反抗与文化偏离》（*Cultural Defiance, Cultural Deviation: Collected Essays*）一书中。我希望这是一个轻松又不失严肃的思想实验，基于我于1980年代在纽约大学拜入尼尔·波兹曼师门，学习媒介生态学课程的经历。

目前，我正在撰写一本探讨高等教育的书，作为雅克·埃吕尔（Jacques Ellul, 1912—1994）所著《技术社会》（*The Technological Society*）的延伸和补充。这是一个绕不开AI影响的话题。我突然意识到，在所有关于AI崛起的喧嚣声中，波兹曼的声音赫然缺席——这自然是因为他已于2003年去世。但我同样发现，今天比以往任何时候都更需要听到波兹曼的声音。

于是，我再次拂去水晶球上的灰尘，尽我所能召回尼尔·波兹曼的精神。以下这篇演讲辞纯属虚构，系波兹曼于2025年6月2日在一所纯属想象中的大学向教职员工和学生发表的演讲。当然，这是一部虚构作品。文字是我而不是波兹曼写的。但我完全相信，字里行间流露的情感是我俩共有的。

大家下午好。今天能有机会与在座诸位交谈，我深感荣幸。但坦白说，我此刻心中更多的是疑问而非答案，是忧虑而非安慰。你们当中有些人或许知道，我一生中花了大量时间思考我们人类如何适应自己的工具——更重要的是，工具又如何塑造我们。你们也许会说我是个技术怀疑论者，而且确实有人称我为“卢德分子”（Luddite）。我认为这并不公平。就我个人而言，我一直更喜欢“老顽固”这个说法。

如今，人人都在谈论AI。政客在谈论它，教师在谈论它，学生、黑客、神学家、TikTok网红——所有人都在思索：与这些如今不仅会计算，还能写作、说话、绘画、说服（真是太过频繁了），甚至欺骗的机器共处，究竟意味着什么。因此，摆在我们面前的问题，不是我们是否会用AI，而是因为我们

使用它，我们会成为什么样的人。

让我先讲一个故事。在古代，有人曾对文字提出过警告。柏拉图告诉我们，苏格拉底担心文字会削弱记忆力，摧毁智慧。当然，他既对又错。文字固然削弱了某些能力，但也增强了另一些能力。它让世界变得更安静、更缓慢、更具反思性，但也让谎言比说谎者活得更久。它鼓励线性、逻辑、命题式的思维，但也降低了对异见的容忍度。

技术就是这样一把双刃剑。

AI 不仅仅是另一种工具，就像锤子或眼镜那样。它是我曾经所说的“定义隐喻的机器”。它不仅为我们做事，而且重新定义了“做事”本身的含义。它不仅仅回答我们的问题，而是改变了问题的本质。

因此，我想问：当学生们不再需要费力写论文，因为机器会替他们写时，会发生什么？当真相变得与貌似合理难以区分时，会发生什么？当我们的教师从思想的激发者变成提示词的整理者时，会发生什么？让我把话说清楚。我并不反对 AI，我反对的是一种缺乏技术哲学的文化，一种采纳工具的速度快于思考其意义的文化。我担心，我们已经变成了这样一群人，不停追问机器“你能为我做什么？”，却很少问“你会对我做什么？”。

我一直认为，任何一个社会所能提出的最重要的问题是：教育的目的是什么？我不确定我们是否还记得答案，甚至不确定我们是否还认为这个问题是必要的。如果教育仅仅是信息的传递，那么没错，让机器去做吧，它们比我们做得更好。如果教育是模式的重复，那么没错，让算法来教吧。但如果教育是判断力、价值观、道德指南的培育，那么我必须坦率地告诉你们：没有机器能教会这些。只有一个有缺陷的、脆弱的、会质疑的人，才能帮助另一个人变得更有人性。

请思考这一点：语言不仅关乎我们如何说话，也关乎我们如何思考。如果 AI 开始替我们说话，我们还能为自己思考吗？我们甚至还能察觉到其

中的区别吗？如果我们把表达外包出去，我们可能很快也会把反思外包出去。而如果我们失去了语言，我们就失去了民主——不是一下子失去，而是一个音节一个音节地失去。

因此我要说：使用机器，但不要变成机器。

我们所熟知的教育之终结？

我长期以来一直认为，每一种新技术都伴有一套隐性课程。它会教给我们一些东西——不一定是我们认为它在教的东西，也常常不是我们想要它教的东西。铅笔教导我们线性思维，电视教导被动接受，电脑教导速度和碎片化。而现在，AI 又会教给我们什么呢？

要回答这个问题，我们必须先问一个更根本的问题：学校的目的是什么？它是一个吸收信息的地方吗？是为预备进入职场而学习掌握技能的地方，还是一个塑造自我的地方？在我们不长的历史中的大部分时间里，我们美国人相信，学校的目的是培养学生成为有见识的公民，具备推理、思辨、质疑与合作的能力。简言之，学校是一个学习在到处都是人的人类世界中如何做人的地方。

但如今，随着 AI 的到来，我们正面临一种危险：将知识与信息获取、思想与计算、学习与输出混为一谈。让我说得具体一点：如果一个学生可以让 AI 帮他写论文，而 AI 也确实做到了——语法正确、结构严谨，甚至略有洞见——那么这个学生到底学到了什么？如果一位老师评阅那篇论文并打了个 A，那么这位老师到底评定了什么？机器没有作弊，学生也没有作弊。但某种本质性的东西已经被抛弃了，也就是那种真正的挣扎，而理解恰恰是在这种挣扎中诞生的。

教育不是追求高效的，它从来就不是为了高效而存在。它不可规模化，

也不会产生整齐划一的结果。它是亲密的、难以预测的，而且是相当具有人情味的。与此相反，AI 承诺高效、个性化和零差错的输出。但我们真的想要从不犯错、不曾在困惑中跌跌撞撞、不感到挫败的学生吗？这些恰恰是学习得以建立的基石。

让我换一种说法。如果我们在教育中越来越依赖 AI，年复一年，那么，到那时还有谁会知道打磨一个句子、修改一个想法、在写作中萌生一个念头意味着什么？当无人再记得如何写一封情书，甚至连写下自己的想法都搜索枯肠时，会发生什么？

AI 或许能替我们写作、替我们求解，甚至替我们预测，但它不能替我们困惑，也不能替我们在意（care）。而教育，如果它还有任何意义的话，就是对“在意”的培育：对真理、正义、邻舍和自身灵魂的关切。

说到底，这不是一场人与机器之间的战斗。这是一场关于两种童年愿景、两种人格愿景的战斗：一种说，“让机器来做吧”；另一种说，“让孩子成为能够做到的人”。让我们谨慎选择。

致教师们：守护火焰

现在，让我对那些选择了教书这一崇高职业的你们说几句话。

我必须说，你们的处境堪忧。近年来，你们被要求扮演心理治疗师、社会工作者、狱警、保姆和技术员的角色。你们承受着预算削减、应试导向的课程、官僚主义的荒唐，以及日益蔓延的猜疑——这个声称重视教育的社会，实际上正在竭尽所能地破坏教育。

而现在，在你们本已不堪重负的教室里，AI 带着礼物走来了。“用我吧，”它说，“我会减轻你的负担。我会批改你的试卷，帮你写教案。我会监控你的学生，回答他们的问题，提供即时反馈。我会让你更高效。”也许你会说：

“好啊,这听起来不错。”我并不怪你。谁不想要解脱呢?但我必须请你思考:我们享受这种解脱的代价是什么?

效率不是教育的目标。它从来不是,也永远不会是。便利也不是,速度也不是。教育的目标,一如既往,是塑造一个人——一个完整的人,一个能够理解世界、抵御疯狂和愚蠢,能够以正义、同理心和想象力行动的人。那是无法被自动化的。

你们不是内容的提供者,你们是文化的守护者。你们不是答案的发放者,你们是问题的助产士。恕我直言,你们的任务是让火焰持续燃烧。不是课程标准、学习成果或数字看板的火焰,而是好奇心、思想和成为一个完整之人的火焰。没有任何机器能做到这一点。

AI可以教事实是什么,但无法教什么是人类的价值。它可以模拟理解,但无法培育智慧。它可以模仿你的声音,但永远无法拥有你的“在场”。“在场”,亲爱的老师们,就是你们的力量。

你的学生不会因为你呈现了无懈可击的内容而记住你,而是因为你曾注视着他们的眼睛,因为你在他们身上看到了他们自己尚未看到的东西,因为你作为一个同样在追寻意义的同类,全身心地在那里。所以,如果非用不可,就用那些工具吧,但不要被工具所用。守护你的权威,守护你的在场,守护你的语言。因为教育之战不是在代码或电路中进行的,它发生在一个心灵与另一个心灵之间。如果我们失去了那个空间,我们就失去了一切。

致学生们：在机器时代如何为人

现在,让我对你们在座的学生说几句话。你们的未来正在被机器书写——也许在字面意义上确实如此。你们出生在一个说着代码语言、崇尚

算法、称手机 (smartphone) 为“智能” (smart)、称搜索引擎为“无所不知”的世界。你们从未体验过没有网络的生活。你们可能是这样的第一代，有朝一日会问机器“何以为人”。

我现在也问你们：你们还想成为这样的人吗？我猜完全回避这个问题是很大的诱惑，一个让机器替你们思考的诱惑。让它替你们写论文、生成课题、纠正语法、总结阅读材料。一个将你们的努力、声音，甚至你们作为独立主体的自我转包给一台机器的诱惑。这个世界不会因此惩罚你们，甚至可能会奖励你们。你们会拿到分数，获得文凭，然后继续前行。

但我必须问你们：前行到何处呢？这是一个重要的问题，因为当你们绕过了思考的挣扎时，你们也同时绕过了思考所孕育的东西：理解。不仅仅是正确，也不仅仅是完成，而是理解——缓慢的、令人挫败的、美好的理解。你们来到这里，不仅仅是为了消费信息；你们来到这里，是为了创造意义。

现在，让我告诉你们一些机器永远不会理解的事情：

一篇文章的价值，不在于它有多么精致，而在于你写下它时对你意味着什么；一个想法的价值，不在于它被生成得多快，而在于它有多么深刻地重塑了你的思想；一个问题的价值，不在于你多快能得到答案，而在于你愿意在“不知道”的不适中停留多久。

你们的老师——如果他们是好老师——在乎的不是你们能否给出完美的答案。他们在乎的是，你们是否正在学会提出更好的问题，做出更精细的区分，感受文字的重量。机器感受不到文字的重量。但你们可以。所以，让机器追求速度吧，你们要慢下来；让机器追求精确吧，你们要求智慧；让机器模拟智能吧，你们要追求理解。

会有人告诉你们，AI 就是未来。它更好、更快、更聪明。但我要告

诉你们：未来需要人类，需要拥有判断力的人、有良知的人、有不从机器那里借来的声音的人。成为那样的人吧。那样的话，就没有机器可以取代你。

机器不是万物的尺度

最后，我们不要假装我们是第一批沉迷于新技术的人。在这一新技术之前已有其他机器。印刷机、电报、电视、电脑，每一样都伴随着号角和预言而来，许诺一个更美好的世界、一个更加互相联结的世界、一个更开明的世界。然而，面对其中的每一台机器，我们不仅要问“它能做什么？”，还要更迫切地追问“它正在消解什么？”

而现在，AI 来了：一台能谱写交响乐、撰写布道词、通过律师资格考试、辅导孩子、诊断疾病、模仿活人、让逝者复活的机器。它令人惊叹、表现出色、超越人类。但我要说，工具不是老师，数据库不是思想，输出不是想法，对思想的模拟不是思想本身，机器不是万物的尺度。

当某种文化忘记了这一点，它就开始以机器为模版来塑造人性。它开始把复杂性视为低效，把模糊性视为错误，把缓慢视为失败。它开始忘记人不是可预测的、整齐的或可规模化的。于是，这种文化就会开始提出一个可怕的问题：也许人类自身的混乱无序，才是那个有待解决的问题本身。

我再说一次：教育的目的不是培养与机器兼容的公民，而是帮助每个孩子成为一个独立自主的人——一个能够反思、不一致、记忆、爱和欢笑的人。AI 或许能写一首诗，但它不能出自内心。它或许能模仿你母亲的声音，但无法知晓她的爱。它或许能告诉你该想什么，但无法告诉你为什么这很重要。

那是你们的工作，是文化的神圣任务。去记住机器无法记住的东西：我们不是工具，不是网络节点，不是用户，不是数据。我们是意义的创造

者，而意义无法从网上下载。所以，让我们不要如此迷恋机械精神，以至于忘记了精神本身。让我们不要用自动化来消除惊奇。让我们教导年轻一代，不是如何与机器竞争，而是如何在一个人性已不再被默认存在的世界里，如何依然为人。因为如果我们不教给他们这一点，那么我们教给他们的其他一切，都将无关紧要。

谢谢你们，祝你们所有人好运。♥

原文标题为“Neil Postman on AI and education: An imagined speech(2025)”，收入《媒介生态学研究》(*Explorations in Media Ecology*) 2025 年第 24 卷第 3 期，第 335—341 页。

(作者是美国罗斯福大学 [Roosevelt University] 新闻与媒介研究教授。)

信仰寻求提示： 在生成式 AI 时代重新想象神学教育^{〈1〉}

文 / 王志希

内容摘要：本文将古腾堡（Gutenberg）发明活字印刷与当今生成式 AI（generative AI）的“古腾堡时刻”并置，借此重新想象 AI 时代的神学教育。本文考察 AI 在神学教育中的开创性实践，最显著者见于加拿大的阿卡迪亚神学院（Acadia Divinity College），并且指出，在制度性谨慎之中，由 AI 驱动的课程、聊天机器人与课堂实验正在日益扩展。然后，本文考察迄今最全面的相关政策论述，即迈克尔·哈内根（Michael Hanegan）与克里斯·罗瑟（Chris Rosser）的白皮书。该白皮书敦促神学院透过策略性治理、课程重新设计以及 AI 赋能的神学图书馆，采取以价值为本、以人为中心的方法来整合 AI。本文主张，当神学教育采纳一种赛博格伙伴关系模式时，能够蓬勃发展。这种模式把院校层面的 AI 策略与课堂层面的创新整合起来。如此，神学院便能培育具 AI 素养的社群，这样的社群不是以“信仰寻求提示”的新范式取代“信仰寻求理解”，而是以之作为补充。

关键词：生成式 AI 神学教育 AI 政策 教学策略 共同智慧



约翰·古腾堡在工作坊。(来源：<https://www.britannica.com/>)

“我是赛博格，但没有关系。” (I'm a Cyborg, But That's OK) ⁽²⁾

导论

六个世纪之前，约 1440 年，来自神圣罗马帝国美因茨的金匠约翰尼斯·古腾堡（Johannes Gutenberg，约 1400—1468）在斯特拉斯堡发明了活字印刷术。该技术后来被古腾堡带回他的出生地美因茨，经改进完善后传播到世界各地。⁽³⁾ 活字印刷术的出现，带来了欧洲的知识革命和媒介革命：知识传播的成本降低，速度加快，变得更容易，而且所触及的人群也

大大扩展。该技术对基督教历史亦影响深远。1450 年代初，古腾堡以此技术开始印刷拉丁文版本的《四十二行圣经》(42-line Bible)，并于 1455 年完工。该圣经被后世誉为《古腾堡圣经》，成为欧洲最早以金属活字印刷的大型书籍之一。不仅如此，在古腾堡去世半个世纪之后，活字印刷术的流行也为 16 世纪的宗教改革做好了技术上预备，提供了技术方面的基础设施 (technological infrastructure)：“多亏了印刷机，一位好辩的修士在萨克森能够一夜之间声名鹊起，并改变世界。没有印刷机，就没有宗教改革。”⁽⁴⁾

活字印刷术的出现也彻底改变了教育与学习的方式，推动了知识的民主化，使学习从依赖口述与记忆，转向基于文本的自主探索。首先，在抄写员主导的时代，学习高度依赖师徒制，学生需透过口头传授来掌握学问，然而，印刷术让书籍得以广泛流通，“天赋异禀的学生不再需要侍坐于某位师长门下，才能学习一门语言或学术技能；相反，他们可以迅速自行掌握甚至偷看书籍，避开导师的监管”。其次，印刷术降低书籍成本，使得学者能在个人书房中进行深度阅读，不再像以往那样需要四处游学，知识的存取门槛大幅降低。再者，学习方法也因印刷术而变革，“透过阅读来学习变得愈发重要，记忆辅助工具的作用逐渐减弱”。最后，技术变革带来了学术的范式转移，以往学者专注于单一文本并深入诠释，而印刷术使得“注释者与评注者的时代宣告结束，而一个‘书籍与书籍之间密集相互参照的时代’随之展开”。⁽⁵⁾

六个世纪以后的 2022 年，迎来了我们这个时代革命性的“古腾堡时刻”。⁽⁶⁾是年 11 月 30 日，OpenAI 正式推出基于 GPT-3.5 的大语言模型 (Large Language Model) 聊天机器人 ChatGPT；不到四个月，次年 3 月 14 日，世界见证了实现质的飞跃的前沿模型 (frontier model) GPT-4 的出现。截至 2024 年 12 月，ChatGPT 每周活跃用户为 3 亿。⁽⁷⁾2025 年 8 月 7 日，OpenAI 推出 GPT-5，这是一种将快速响应与内置推理结合于一体的统一模型。

OpenAI 于 2025 年 12 月 8 日发布的一篇官方博客文章指出, ChatGPT 现在每周服务超过 8 亿用户。⁽⁸⁾紧随其后的生成式 AI 竞品还包括英文世界里的 Claude 与 Gemini 等, 以及中文世界里阿里巴巴推出的通义千问和 2025 年 1 月异军突起的 DeepSeek 等。

同样作为资讯技术的创新, 活字印刷术与生成式 AI 在教育与学习领域, 既展现出相似性, 又有差异性。在相似性上, 首先, 两者皆彻底改变了知识的创造、传播与获取方式, 印刷术透过大规模生产书籍, 使学习不再依赖口述与手抄; 生成式 AI 则透过即时互动与内容生成, 让学习更加动态且个性化。其次, 两者都推动教育的民主化, 印刷术让书籍变得更易获取; 生成式 AI 则进一步打破经济与地理的限制, 使知识更可即时获取。再次, 两者皆面临来自既有教育体系的质疑与抵制, 印刷术曾挑战宗教与学术机构对知识的控制; 而生成式 AI 则引发关于学术诚信、抄袭、资讯准确性与学生依赖性的讨论。最后, 印刷术的影响初期仍受限于识字率, 而生成式 AI 虽然降低了学习门槛, 却仍受限于数字鸿沟与技术依赖。另一方面, 在差异性上, 最重要的是知识的生产方式完全不同: 印刷术被动复制既有的内容, 而生成式 AI 则能够以人机互动的方式生成全新的内容。

两个古腾堡时刻都重新刺激人类思考教育与学习的核心问题: 谁拥有知识? 学习应如何发生? 以及技术如何塑造我们理解世界的方式? 本文在上述历史类比的基础上, 尝试回答以下研究问题: 神学教育工作者如何重新想象生成式 AI 时代神学教育的学与教?

一、神学教育界的生成式 AI 实践

在生成式 AI 的实践方面, 神学教育界的可见度相比于高等教育界要低很多。这一滞后现象符合曾长期担任美加神学院协会 (The Association of



加拿大阿卡迪亚神学院院址鸟瞰。(来源：<https://acadiadiv.ca/>)

Theological Schools, ATS) 执行主任的艾尔希尔 (Daniel O. Aleshire) 博士对美国神学院的总体观察：

神学教育中的教学策略发展，往往跟随更广泛高等教育领域内的发展趋势。然而，每一次，神学院校总是比一般高等教育机构更晚采用这些策略，且态度更为怀疑。神学院校通常在教育上较为保守。此外，大多数神学院

校规模较小，资源有限，难以发展并接纳新的教育模式。然而，它们最终仍被迫采纳这些教育创新——通常是在强烈抗拒之下（kicking and screaming）接受的。^{〔9〕}

尽管如此，神学院的生成式 AI 实践在少数情况下仍是可见的。其中一个最突出的例子是位于加拿大新斯科舍省沃尔夫维尔镇的阿卡迪亚神学院（Acadia Divinity College, ADC），它是隶属于加拿大大西洋地区浸信会（Canadian Baptists of Atlantic Canada, CBAC）的神学院。

ADC 的院长安娜·罗宾斯（Anna Robbins）牧师（神学、伦理与文化教授）认为，神学教育不能被动地接受科技变革，而应积极探索生成式 AI 如何塑造未来的神学教育模式。在 2025 年 1 月初的一个播客访谈中，罗宾斯引用《民数记》13:17—20 与《马太福音》16:2—3，来说明神学教育面对生成式 AI 时代应具备的态度与责任，^{〔10〕}强调神学院应当具备探子的精神，积极研究 AI 技术如何塑造未来的神学教育，而非被动等待或抵制技术的发展。同时，教会与神学教育界应如耶稣所教导的，读懂“时代的记号”，以属灵的智慧来思考如何负责任地使用 AI，预备未来的教牧与学术环境。在 2025 年 1 月底的另一个播客中，罗宾斯指出，有些人对于这波 AI 潮流表示要等等看，但是她认为如果延迟认识 AI 技术，则会产生巨大的差距，不容易再追上。^{〔11〕}

ADC 斥资购买了 ChatGPT Enterprise，^{〔12〕}罗宾斯与 ADC 的教育技术主任约翰·坎贝尔（John Campbell）牧师及教育创新主任乔迪·波特（Jodi L. Porter）等人组成了一个研究团队，开发了一款他们所称的“ADC 课程内容创建工具”（ADC Course Content Creator），实质上应该就是一个自定义 GPT（custom GPT）。^{〔13〕}他们将该学院的机构承诺、预期的课程成果以及标准课程大纲结构等文件上传到该 GPT 中，亦即结合“检索增强生成”

(Retrieval Augmented Generation) 技术。⁽¹⁴⁾ 作为 ADC 在生成式 AI 教学实验中的首创学分制课程,“基督教事工中的 AI 伦理”(The Ethics of AI in Christian Ministry) 于 2024 年秋季开设,为硕士层次的 10 周自主研究课程,有六名学生参与。“课程探讨基督教事工的各个层面,并引导学生思考 AI 如何应用于其中。学生被引导反思 AI 应用所涉及的伦理与神学问题,并考虑应对潜在挑战的措施。”该课程的课程生成(包括学习结果、大纲和每课计划)与内容生成(包括阅读材料、功课和非同步视频)完全由 ChatGPT 设计(辅以 Scholar AI⁽¹⁵⁾)。AI 生成的讲座资料透过 AI 视频平台 Synthesia⁽¹⁶⁾ 转化为由罗宾斯的声音训练过的视频,配合罗宾斯的虚拟形象(她的家人甚至一度误以为该视频中的形象就是她本人)。这些影片可转换成八十种语言。

每一周, AI 设计的学习活动包括观看教学影片,阅读 AI 从两本主要书籍中选定的材料,并补充博客文章、同行评审期刊论文以及其他视频和相关资源。“学生完成由 AI 设计的每周作业,并将其上传至课程讨论论坛, AI 会提供公开与私人的反馈,并根据学院的评分标准进行评分。课堂讨论围绕这些作业展开, AI 也会参与其中,对学生提交的内容进行反思,并提出进一步的问题。”学生的最终成绩仅分为及格与不及格,评分标准是基于他们对该课程本身的评估,而非课程内容。根据 ADC 的网站介绍,“课程结束后,学生将对其学习经历提供反馈,研究团队则会对这项实验进行评估。团队将审查 AI 在设计学习成果与课程结构方面的能力、生成有效教学内容的的能力以及评估学生作业的能力。”⁽¹⁷⁾ 这个完全由 AI 驱动的课程,开了神学教育历史的先河。2025 年春季, ADC 推出另一个全新的“基督教事工中的 AI 应用”(Using AI in Christian Ministry) 微学分课程。课程以线上与实体混合模式进行,并由约翰·坎贝尔牧师负责。这些课程不仅探索 AI 的技术与伦理问题,也强调 AI 在教会牧养、门徒训练及行政管理中的实际



使用 AI 来设计一门关于 AI 使用的伦理课程。(来源：<https://faithandleadership.com/>)

应用。^{〈18〉}

ADC 的实验精神还体现在其他 AI 相关的项目和计划之中。该院尝试研发一个 AI 聊天机器人，协助学生在辅导类课程中“练习自己的辅导技巧，并在过程中获得即时反馈”。^{〈19〉}该校的神学未来研究前瞻中心（Futuring Hub for Theo-Futures Research）——罗宾斯和波特皆是中心成员——使用 AI 工具来做研究。作为该中心项目的一部分，“神学教育与教会沙盒”（The Sandbox for Theological Education & the Church）得到 Lilly Endowment Inc. 的“明日之路计划”（Pathways for Tomorrow Initiative）资助，“旨在推动在课堂与教会中进行有关处境化、跨文化与转化性教育的实验”。该沙盒目前的许多实验都是关于“AI 技术在教与学中符合伦理的应用”。^{〈20〉}

ADC 的 AI 神学教育首创（AI 驱动的课程发展与设计）尚属少数。除该神学院之外，加拿大国际基督教教育神学院在 2025 年 6 月开设证书

课程，题为“主日学教学与 AI：融合科技与信仰教育”，或属华人神学教育界在开设相关证书课程方面的首创。该课程“学习人工智能在基督教教育中的实际应用，包括课程规划、互动讨论、神学探究，以及与学生的个性化交流”。^{〈21〉}

神学界更多的生成式 AI 实践，是开设相关的工作坊和会议，以及在个别的课程中加入相关的元素。就前者而言，2024 年夏天，毕业于埃默里大学、曾担任弗吉尼亚神学院讲道学助理教授的马克·安德鲁·杰斐逊(Mark Andrew Jefferson) 博士在普林斯顿神学院开设教学工作坊，请 ChatGPT 模仿葛培理的风格写一篇关于好撒玛利亚人比喻的讲道。^{〈22〉}更具有代表性的是，美加神学院协会、ATLA 与 In Trust 神学院中心(In Trust Center for Theological Schools) 三个机构携手打造一系列线上会议，题为“F[AI]thfully Co-Creating: Navigating Artificial Intelligence and Theological Education”，探讨 AI 与神学教育的交汇点。驱动该系列活动的两个问题是：“AI 如何塑造神学教育？神学教育又该如何塑造 AI？”第一次活动在 2024 年 11 月完成，并在 2025 年 2 月和 3 月继续开设线上会议，并欢迎公众免费参加。^{〈23〉}在华人世界，伯特利神学院的教牧及信徒进修中心邀请笔者分别于 2023 年 11 月、2024 年 4—5 月以及 2025 年 4—5 月对公众开设基督徒如何应用生成式 AI 的系列工作坊，或属华人神学界在开设相关工作坊方面的首创。^{〈24〉}香港中文大学崇基学院神学院亦曾于 2024 年 8 月邀请笔者与教职员分享如何在神学各学科的教学场景中应用生成式 AI。

有个别的神学课程开始加入生成式 AI 的元素。例如，在生成式 AI 刚兴起的 2023 年春季学期，笔者在香港中文大学崇基学院神学院的课上，曾与学生分享 AI 对中国基督教史议题的分析，并对这些分析做出评价：赵紫宸为何在 1946 年想要离开北平去香港，以及他为何最终留在北平？2024 年夏季，笔者在加拿大华人神学院多伦多讲授“历史神学”一课时，请

AI 音乐生成工具 Suno 生成了一首歌，帮助学生理解阿里乌（Arius）的神学。霍华德大学神学院（Howard University School of Divinity）的神学教授弗雷德里克·L·韦尔（Frederick L. Ware）在 2024 年秋季学期让他的学生“阅读一本教材，以帮助他们从神学角度思考 AI 的能力”（目前尚无法确定该书书名），随后学生也亲自使用 AI。^{〈25〉}耶鲁神学院美国宗教史教授蒂莎·温格（Tisa Wenger）让学生请 ChatGPT 以 19 世纪非裔卫理公会（African Methodist Episcopal Church）第一位女传道人贾瑞娜·李（Jarena Lee, 1783—1864）的语言写一篇讲道，然后他们进行批评分析。温格发现，这项练习更能展示的是“ChatGPT 的局限性而非其优势，并且能测试学生对相关材料的掌握程度，因为 ChatGPT 漏掉或错误呈现的细节，正好成为检验点”。^{〈26〉}

牛津大学神学与宗教学院下属的伊恩·拉姆齐科学与宗教中心（Ian Ramsey Centre for Science and Religion）于 2024 年发起“牛津神学与 AI 合作计划”（Oxford Collaboration on Theology and Artificial Intelligence, OCTAI）。该计划获得了约翰·坦普尔顿基金会的资助，由马克·哈里斯（Mark Harris）、约书亚·霍登（Joshua Hordern）两位教授与林登·德雷克（Lyndon Drake）博士领导，核心目标是建立一个全球性的学术与产业网络，促进 AI 开发者对伦理与人类价值的深刻反思，特别关注具有宗教洞见的领袖如何在 AI 技术发展的关键决策中发挥作用。该计划的重点产出之一将是一份“牛津 AI 誓言”（Oxford Oath for AI），旨在帮助 AI 从业者在技术创新中坚守道德责任，确保 AI 的发展符合人类福祉与社会公正。^{〈27〉}

此外，也有个别神学院因应生成式 AI 的出现而做出更新的注释指引。譬如，新加坡神学院早在 2023 年 9 月就根据《芝加哥手册》（*Chicago Manual of Style*）和出版伦理委员会（Committee on Publication Ethics）发布了一份“使用生成式人工智能之引注格式”，2024 年发布了一份更详尽的

《在评估作业提交中的生成式 AI 工具使用》(Use of Generative AI Tools in Assessment Submissions)，可能是最早在注释规范上提出相关指引的神学院之一。^{〈28〉}

二、神学教育界的生成式 AI 论述

当前关于神学教育与生成式 AI 之关系最全面的论述，可以参考迈克尔·哈内根(Michael Hanegan) (神学与教育创新学者)与克里斯·罗瑟(Chris Rosser) (图书馆与教育技术专家) 共同撰写的白皮书《AI 与神学教育的未来》第二版 (Artificial Intelligence and the Future of Theological Education [Version 2.0])。哈内根是 Intersections (一个致力于学习与人的塑造的教育机构) 的创办人，并担任朋霍费尔研究所 (Dietrich Bonhoeffer Institute) 的高级研究员。此外，他曾在哥伦比亚大学研究生技术管理项目担任副主任，并参与创新伦理学的教学。罗瑟则是俄克拉荷马州立大学的新生与转学生图书馆员，曾任俄克拉荷马州基督教大学的图书馆学教授与神学图书馆员。该白皮书的第一版发表于 2023 年 11 月，并于 2024 年进行第二版的更新。它从生成式 AI 的六种应用形式 (文本生成、文本增强、构思与策略、研究增强、设计与视觉化以及多模态应用) 以及人机互动模式的转变切入，提出以人为中心的进路来指导 AI 在神学教育中的整合与应用。随后，该白皮书论述神学教育为何应当积极介入 AI 讨论和实践，并提供政策框架、课程设计、师生的学习与研究，以及神学图书馆创新的建议，最后以三项核心行动方案作结。^{〈29〉}

总体而言，哈内根与罗瑟在白皮书中明确支持神学教育应当积极介入 AI 的讨论与应用。两位作者指出，神学教育的跨学科本质，使其能够灵活适应 AI 技术带来的学术变革，而 AI 的快速发展与低门槛应用，更为资源

有限的神学院提供了站上技术前沿的契机。透过战略性采纳 AI，神学教育不仅能提升学习、研究与社会回应的能力，还能扩展其使命与影响力，并在创新与生存压力下开辟新的发展空间。然而，关键问题并非神学院是否具备能力，而是是否愿意主动拥抱这场变革。积极采用 AI 的神学教育机构将在政策制定与伦理框架塑造上拥有更大话语权，而选择观望者与目前采取行动者之间的差距，将会以指数级速度扩大。

具体而言，这份白皮书带来的洞见可以从学和教两个角度来分析。就学而言，白皮书强调神学教育者明辨生成式 AI 对神学教育带来的影响时，至关重要的一条原则是将学生的经验、而非神学教育机构的需要与关注放在核心位置。白皮书所谓以人为中心的进路，一个要点就是神学院需要重新思考希望培养、支持并派遣到世界上的人应该具备哪些特质？如果 AI 对人的影响只会愈发深远的话，那么神学院培养塑造的人需要具备 AI 素养(AI literacy)：“以人为本的 AI 素养教育强调赋能学习者，使其能够理解、批判性地评估并负责任地使用 AI 技术，同时坚持并强调人的自主性、创造力、价值观和伦理原则。”若不介入 AI，就缺乏可信度来言说 AI 技术；若教师自己不首先成为 AI 素养的学习者，也就难以谈及赋能作为学习者的学生。

就微观教学而言，白皮书详细探讨了 AI 在神学教育课程设计与教学实践中的应用，指出传统的“阅读—讨论—写作”神学教育模式存在诸多局限，例如文献范围受限、学生参与度不稳定以及写作过程孤立且缺乏持续影响。AI 技术能够突破这些限制并增强传统的神学教育模式，带来更多元的学习方式，包括开放式学习与探究式学习。AI 可以帮助教师快速生成课程大纲、推荐适切的阅读材料，并透过多模态内容（如文本、音频、视频等）丰富学习体验。AI 的语言处理能力使其能够翻译和适应不同语言的学习需求，促进国际学生的无障碍学习。同时，AI 能够加速课程设计过程，帮助教师更有效率地迭代与改善教学内容。



图片由 AI 生成。(来源：<https://www.aiu.edu/>)

白皮书也从宏观的机构政策角度探讨了几个层面的议题。第一是神学院 AI 政策的必要性与原则。两位作者指出，到撰文的时间为止，未见有神学教育机构正式发表与生成式 AI 相关的立场文件。学术界对 AI 在作弊与学术不端行为中的应用存有忧虑，但白皮书主张，AI 检测技术既无效又可能造成不公平对待，神学院应将精力从惩罚性监管转向发展更具建设性的教育策略。若神学院要开始制定自己的 AI 政策，与其制定过时且难以执行的 AI 使用限制，不如转向以价值为导向的政策框架。AI 政策可以体现出以下五个核心价值：透明性（transparency，确保 AI 使用过程开放且可追溯）；好奇心（curiosity，鼓励广泛探索与跨领域学习）；严谨性（rigor，以积极投入与制度建设维持学术水准）；包容性（inclusion，缩小数字鸿沟，确保所有人都能有效利用 AI，并持续获得支持），以及游戏性（play，透过实验与创新促进学习与社群建构）。

第二，白皮书特别强调神学图书馆——部分因为作者之一的罗瑟具有

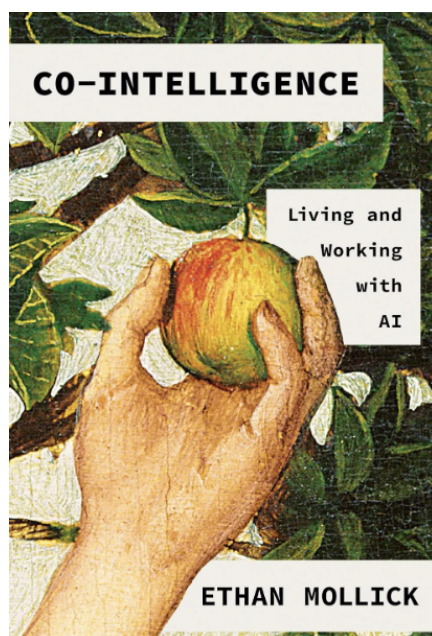
图书馆工作的背景——在 AI 时代不仅是知识的储存地，更应当是创新与学习的中心。透过 AI 技术，图书馆可以突破传统资源与专业知识的限制，提升文本分析、跨学科研究、个性化学习等能力，成为神学机构培训与应用 AI 的核心枢纽、“神学机构的心跳”。

第三，白皮书最后提出神学教育机构值得投资的三个方向，除了 1) 前面提到的制定 AI 政策之外，还包括：2) 为师生等提供 AI 素养与应用技能的培训，并推动跨学科课程与创新实验室，以及 3) 神学界应超越个别机构的界限，透过全球神学 AI 联盟 (global theological AI consortium)、年度峰会与共享数字平台，共同塑造 AI 时代的神学教育未来。

三、重新想象神学的学

神学教育的未来应该批判性地拥抱生成式 AI 带来的“美丽新文字” (brave new words)^{〈30〉}，将生成式 AI 视为与包括神学院师生在内的人类协作的“共同智慧” (co-intelligence)^{〈31〉}或“延伸心智” (extended mind)^{〈32〉}。但是，如何让 AI 成为神学教育领域的共同智慧呢？

在生成式 AI 时代重新想象神学教育，首先需要重新想象神学的学，而这里的“学”包括学 AI 和借着 AI 学两方面。一方面，神学院师生值得在 AI 素养方面自我投资，进行技能提升 (upskilling) 或技能更新 (reskilling)。另一方面，神学院师生借着这些前沿模型学什么？按宾州大学沃顿商学院教授伊桑·莫利克 (Ethan Mollick) 在《工作者的 AI 共智模式》 (Co-Intelligence: Living and Working with AI) 一书提出的 AI 使用第一原则而言，就是我们平时想要学任何事物，都请 AI 试一试 (“always invite AI to the table”)。生成式 AI 的价值并非仅在于其能帮助完成某项具体任务，而在于用户只有透过持续试验才能真正理解 AI 的参差边界 (jagged frontier)，亦即



伊桑·莫利克《工作者的 AI 共智模式》英文版（2024 年）封面。
（来源：<https://www.amazon.com/>）

哪些任务表现超越人类，而其他任务则远远落后。^{〈33〉}如果“随着 AI 的普及，那些深入理解 AI 工具的细微差异、限制与能力的用户，将独特地具备开发 AI 全部创新潜力的能力”^{〈34〉}，那么 AI 在神学的学习方面能够发挥多少能力，将取决于多少神学工作者可以在多大程度上成为“用户创新者”（user innovators）。^{〈35〉}

举例而言，笔者从 2024 年开始自学讲道学，基于以往的神学训练、个人和教会的需求，研发了一款名为“Co-Predching”的自定义 GPT，作为预备讲道的 AI 助手，并且开放给公众使用。^{〈36〉}它不能取代讲道者对圣经的研读和默想等传统方法，也不能取代释经书等传统工具或 Logos Bible Soft-

ware 等数字图书馆应用程序，但是它可以成为一个强有力的补充。

Co-Predaching 的核心设计理念之一是禁止生成整篇讲章，因为基于“预测下一个字”（next token prediction）的大语言模型文本生成原理，通常一次性生成整篇讲章的质量若不是无比糟糕，就是无比平庸。在此前提下，Co-Predaching 提示词的设置（仍在不断迭代中）首先会根据用户的需求和输入的内容来确定协助讲道预备的起点。第一，如果用户提供的是经文章节，那么 Co-Predaching 会输出完整的经文，包括新修订标准版圣经和中文和合本圣经。对于希伯来圣经的经文，Co-Predaching 会引用希伯来圣经（*Biblia Hebraica Stuttgartensia*）和《七十士译本》（*Septuagint*）；对于新约经文，Co-Predaching 则提供希腊文圣经（*Nestle-Aland Novum Testamentum Graece*[NA27]）。第二，如果用户提供的是日期，那么 Co-Predaching 会自动执行网络搜索来查询该日期对应的《修订版通用经课》（*Revised Common Lectionary*）。这一功能适用于遵循经课传统的讲道。第三，如果用户提供的是主题，那么 Co-Predaching 会根据该主题自动搜寻三段相关的圣经经文。第四，如果用户提供的是最近的一些思考，那么 Co-Predaching 会据此进行头脑风暴，提出三个讲道点子。

Co-Predaching 提供了一个结构化的捷径（斜线指令）界面，用户可透过此界面使用广泛的功能。这些功能被组织成几个主要领域：圣经诠释、整合性讲道、讲章结构设计、例证素材、具备处境意识的应用、文化偏见辨识以及讲章评估。

Co-Predaching 的一个特征，是它能够引导用户对特定圣经文本进行多层次的话语分析（discourse analysis）。^{〔37〕}该工具将释经架构于四个相互关联的语言学层次：话语（discourse）、段落（paragraph）、子句复合体（clause complex），以及词汇语法（lexicogrammar）。在话语层次，用户会被提示去考量文体、宏观结构、主题推展以及前景与背景策略。段落层次的分析则

引导用户关注结构的协调性、指示语、衔接手段以及参与者角色或叙事方向的转变。子句复合体层次检视逻辑、语气与语义链。最后，在词汇语法层次，该工具会检视及物性模式、词汇衔接、子句平行结构以及语法标记性。这些层次会递回地相互作用，使讲道者能够察觉神学意义如何不仅仅从孤立的字词或句子中产生，而且从文本在话语处境中逐步展开的架构里涌现。

除了语言学的审视之外，Co-Preaching 还借由提供广泛的批判方法论库来支持诠释的多元主义。它使用户能够运用历史批判方法，例如文本鉴别学、形式鉴别学、编辑鉴别学与来源鉴别学，同时也促进透过修辞、叙事与正典视角进行的文学分析。该工具透过女性主义、后殖民与解放神学的视角提供视角性阅读 (perspectival readings)，并支援社会学与心理学的诠释。该工具能够比较不同经卷之间的文本（例如，旧约与新约的互文性，或对观福音的平行经文），并促进与多元宗派传统的对话，例如天主教、改革宗、五旬宗与重洗派的框架。

Co-Preaching 允许用户发起模拟对话，此功能最初是遵循弗雷德·B·克拉多克 (Fred B. Craddock) 在《布道》(*Preaching*) 一书中的见解，⁽³⁸⁾ 借由邀请大语言模型扮演历史上的讲道者、神学思想家、圣经作者或圣经人物来进行。例如，用户可以探索提摩太·凯勒 (Timothy Keller) 可能会如何为世俗的怀疑论者解读一段经文，或者朋霍费尔可能会如何从道德抵抗的危机中诠释该段经文。这些互动采取来回对话的形式，允许用户测试各种诠释、预期问题，并从多元的神学视角中考量回应。其结果不仅是修辞上的丰富，也是一种处境意识的展现，能够训练讲道者预期听众的接收情况，并抵抗诠释上的孤立主义。

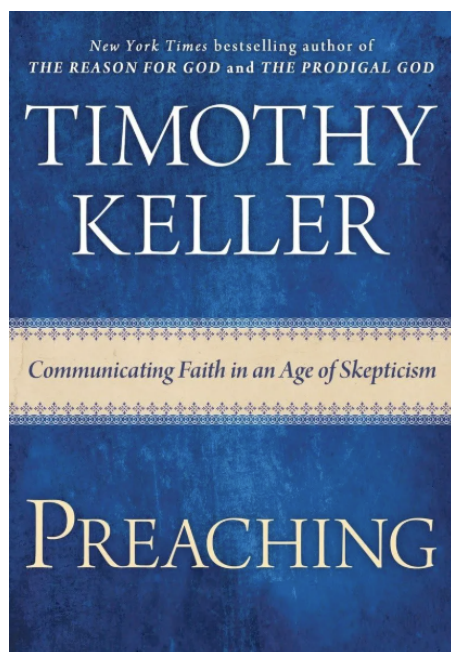
受肯顿·安德森 (Kenton Anderson) 所著《整合讲道》(*Integrative Preaching*) ⁽³⁹⁾ 一书启发，并借鉴我个人的阅读与反思，Co-Preaching 建构了四个互补的框架：互动、教导、扎心与启发 (engagement, instruction, con-

viction, and inspiration)。在“互动”方面，该工具支援用户建构进入讲道的互动式切入点，例如在信息的不同节点提出问题，或使用旨在促进专注与认同的听众提示。在“教导”方面，该工具协助识别可能的误读或语义歧义，阐明精确且具神学根基的“经文主题”（textopic），并构思与精炼讲道的“大概念”（big idea）。这个大概念随后可以被重塑为格言形式，以生动的隐喻加以说明，或改编为讲道标题。在“扎心”方面，它可以撰写源于大概念且融入五感的祷告，这些祷告融合了认罪、感恩与赞美，或者对人性黑暗面进行探索。最后，在“启发”方面，该工具支援提出具体且以福音为中心的挑战，并建构被神圣作为所转化的富有想象力的未来。

为了支援讲道的连贯性与修辞的递进，Co-Preaching 包含一个大纲建构的模块，允许用户自动生成大纲，或指定要点数目的大纲，范围从典型的三点式讲道到更复杂的四点或五点式讲道。用户可以从一系列基于发展成熟模式的讲道学模板中进行选择，包括尤金·劳瑞（Eugene Lowry）的叙事循环（narrative loop）、弗雷德·B·克拉多克的归纳式螺旋（inductive spiral），以及威尔森（Paul Scott Wilson）的四页讲章结构（four-page structure）。这种结构上的灵活性旨在与讲道者的大概念保持一致，并确保形式与内容相互强化。

Co-Preaching 提供了一系列例证选项，涵盖圣经叙事、历史事件、隐喻或当前新闻故事。这些例子针对不同的情感基调（严肃、感人或轻松）量身定制，并能适用于多样化的听众。该工具透过提供特定处境的应用途径，提示讲道者将大概念与生活经验连结起来：这些途径适用于家庭、职场、教育环境、教会或更广泛的公共领域。^{〔40〕}讲道者也可能被邀请透过回答具灵命塑造意义的问题（“你曾在这项真理上经历过怎样的挣扎？”）来进行个人反思，随后 Co-Preaching 会协助将这些反思精炼成具修辞效果的语言。

建基于提摩太·凯勒《文化讲道学》^{〔41〕}一书的洞见，Co-Preaching 包含



提摩太·凯勒《文化讲道学》英文版（2015年）封面。
（来源：<https://www.amazon.com/>）

一项功能，能协助讲道者察觉并指出文化偏见，这些偏见形塑了对圣经的诠释以及当代听众对信息的接受方式。认知到讲道总是发生在根植于文化的处境中，该工具提供了针对主导性诠释框架的引导式反思，这些框架不仅包括启蒙运动、后现代与新自由主义，还包括马克思主义、儒家与佛教。

最后，为了协助讲道者完善自己的讲章，Co-Preaching 提供了一套可选的评估工具，讲道者可以选择性地将其应用于讲章草稿中。这些功能使讲道者能够针对讲稿的特定方面要求诊断性反馈，例如识别逻辑谬误、揭示隐性偏见（性别、族裔、社会经济或神学方面），或厘清薄弱的论证脉络。用户也可以要求 Co-Preaching 审查讲章的连贯性与流畅度，包括神学概念

的清晰度以及语言对非专业听众的易懂性。额外的风格工具则透过建议改善用词与语气来强化修辞表达，从而提升听众的参与度。

透过阅读、试验与不断的迭代，笔者不仅在设计这个自定义 GPT 的过程中学习如何预备健全的讲章，更尝试作为一名用户创新者，将该工具转化为一种用于讲道预备的共同智慧。

四、重新想象神学的教

最后，当我们从“教”的角度重新想象生成式 AI 时代的神学教育时，可以从两个层面入手：宏观的机构管理层面与微观的教师教学层面。就前者而言，神学院管理者可以考虑从三个“一”出发：一份调查问卷、一个专责小组（task force）、一笔年度预算。2025 年 2 月，笔者透过香港神学教育协会发放“神学教育与生成式 AI 调查问卷”给香港各神学院教职员。^{〔42〕}根据收回的 34 份调查问卷（截至 2025 年 2 月 16 日），针对“您所在的神学院为应对 AI 带来的挑战或机遇做了哪些预备”这个问题，有约一半的回应者（47%）表示所属神学院已经“在教职工会议中讨论 AI 与神学教育”，但是更有一半的回应者（50%）表示“目前尚未有具体行动”。暂时只有极少数神学院“为整合 AI 技术而制定专门的财政预算”（三人）或者“设立 AI 相关的专责小组”（二人）。如果神学院想要开始从机构管理层面有具体行动，首先可以在神学院内部发放调查问卷，了解团队成员目前在 AI 方面的实践和认识，目的在于获得一个量性与质性的宏观图景，以及识别出团队当中四分之一（24%）“经常使用，并在教学或研究中应用”的教职员。在此基础上，神学院可以邀请这部分教职员与机构的重要持份者组成一个专责小组，制定神学院未来的 AI 战略。

AI 战略可以包括教员 AI 发展计划（例如如何鼓励教员参与 AI 时代的

神学教育、如何支持对 AI 持不同态度的教员、如何发起不同的 AI 相关的进修项目)，以及如何制定 AI 政策或 AI 学术诚信守则等。不过，笔者认为 AI 战略最重要的两个议题可能是：第一，生成式 AI 在机构的根本使命中是否占有一席之地；第二，如果有的话，神学院是否考虑“为整合 AI 技术而制定专门的财政预算”。肯顿·安德森与格雷戈里·J·亨森（Gregory J. Henson）曾指出，“‘追踪资金流向’（follow the money）是调查新闻中的一条重要准则。如果你想真正理解一个组织所重视的事物，或是一个系统如何运作，就应该仔细观察其财务管理方式。”^{〔43〕}如果想要在机构层面落实前面提到的伊桑·莫利克所说的 AI 使用第一原则，以及保证便利性（accessibility）和公平性（equity）原则，那么最理想的方式不是让教职员个别承担 AI 成本，而是如 ADC 那样以机构的名义购买付费版 AI。^{〔44〕}这样才能让更多的教职员以最小财政负担的方式，最大程度地探索最新前沿模型的能力与限制，以至于教职员与专责小组之间有可能形成比较富有成效的反馈回路（feedback loop）。

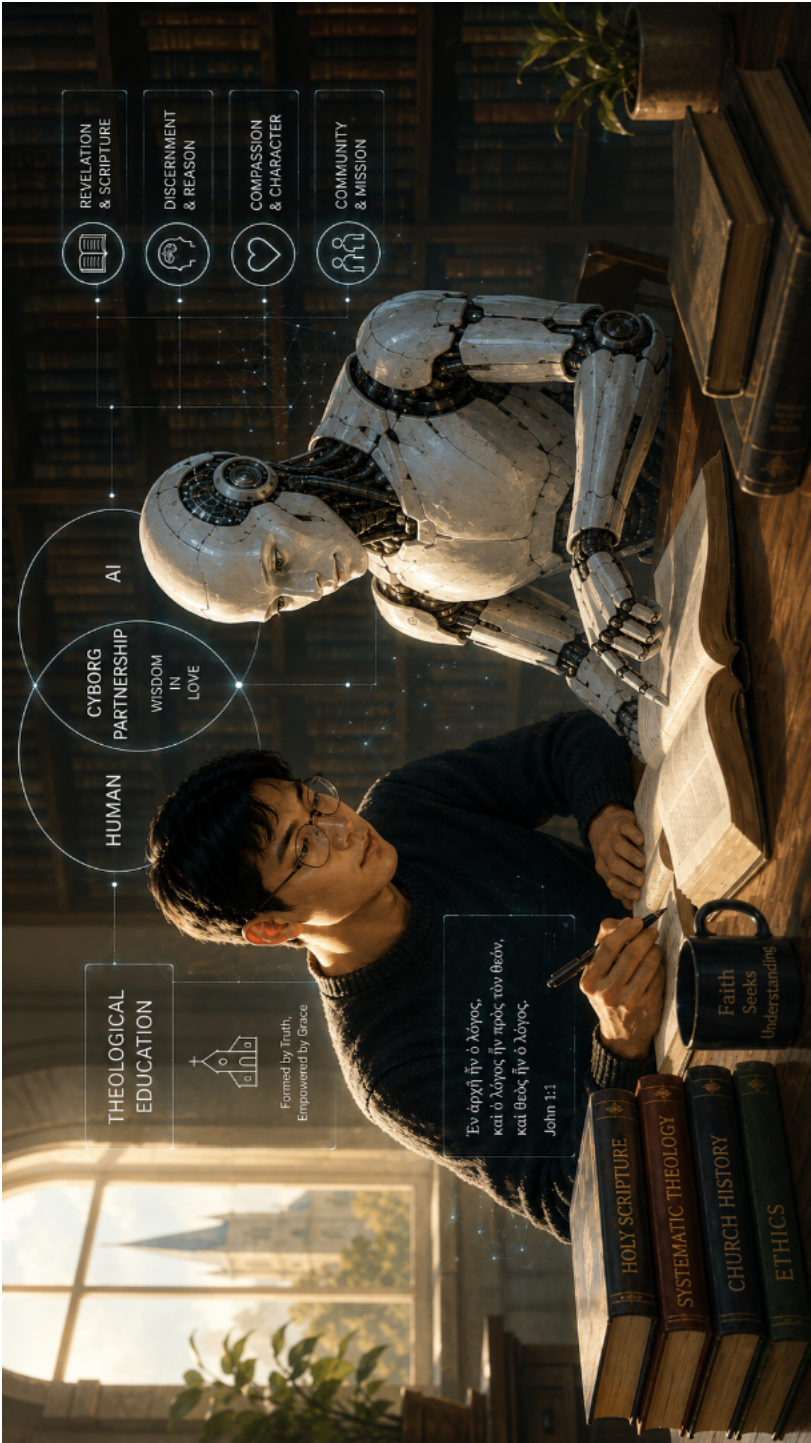
在以上三个“一”的基础之上或者与之并行的机构策略，可以包括：邀请专家学者分享 AI 在神学教育中的应用（回应者占 18%）、举办关于 AI 的研讨会（15%）、培训教师如何在课堂中使用 AI（9%）、开设一般性的 AI 素养课程或更具体的 AI 与神学的课程（6%）。这些活动中值得讨论的课题，可以从教职员对 AI 在神学教育中使用所引发的担忧出发，包括：学生过度依赖 AI，影响批判性思考（77%）；AI 生成内容的准确性（68%）；学术抄袭（59%）；AI 可能导致神学教育去人性化（38%）；AI 生成内容具有偏见（神学偏见、种族偏见、社经偏见、性别偏见等）（38%）；AI 可能被误用来生成不合适的内容（32%）；AI 生成内容了无新意（21%），以及其他（AI 可能取代教师；AI 答案的深度和创意；AI 生成过程缺乏对上帝的敬畏）。在充分的咨询和研讨之后，制定神学院整体的 AI 政策才能有的

放矢。此外，还可以考虑的机构策略包括：讨论在神学院总体课程体系和分学科课程（圣经科、神学科、历史科、实践科等）中的 AI 整合；设立类似 ADC 的神学未来研究前瞻中心那样的教研中心，以及透过香港神学教育协会这样的机构建立跨神学院合作（如跨学院的“神学教育与 AI”论坛）。

就微观的教师教学而言，调查问卷的回应者当中有四成多（13 人）尚未在教学中应用生成式 AI，可见在这一方面还有比较大的探索空间。目前最多的应用则包括：辅助课堂讨论（22%）、生成教案图片（19%）、设计课程大纲（16%）、设计课堂教学活动（13%）、设计功课，如请学生批判性评估 AI 生成的答案（13%），以及训练神学生与 AI 互动，作为“神学对话伙伴”（9%）。此外还包括资料搜集、翻译文本以作备课和讲义之用，修改文案以及制作问卷分析等教学应用。

根据笔者的经验，一个可应用于各种情境的重要教学理念是，针对特定的教学需求来量身打造自定义 GPT；Claude 的 Projects 和 Gemini 的 Gems 亦提供类似的功能。例如，在圣经语言课程中，自定义 GPT 可以作为学习希腊文或希伯来文的学生的词汇与语法导师。它可以根据每周的词汇表生成测验，对词形分析练习提供即时反馈，并使用取自圣经文本的例子来解释语法模式。这样的 GPT 还可以根据学生的表现调整难度，并在考核前生成小测。

自定义 GPT 也能根据指定阅读材料生成简短测验。这些测验可能包括（加权的）选择题、简答题，或是“两真一假”（Two Truths and a Lie）等形式，以支持提取式学习（retrieval-based learning），而且到目前为止，无论是在大学还是神学院，都深受笔者的学生欢迎。这类测验有助于强化关键概念、厘清诠释上的误解，并创造低风险自我评估的机会。Claude 因为具备 Artifacts 功能，对于这类设计工作特别方便，该功能允许教师发布互动式测验页面，并与学生或更广泛的大众分享。



图片由编者使用 ChatGPT 生成。

游戏化 (gamification) 是生成式 AI 工具可以提升神学教学的另一个领域。自定义 GPT 可用于设计课堂游戏, 在强化学习的同时提高学生的参与度。例如, 你画我猜游戏 (pictionary) 可以改编用于圣经研究或教会历史课程, 要求学生画出关键的经文场景或历史概念, 并让同侪猜测。再举一个例子, 在密码游戏 (password) 中, 学生会得到 AI 生成的单词提示, 以帮助他们的队友识别神学术语。这些游戏不仅能培养更具活力的课堂环境, 还能鼓励学生以富有创意且协作的方式回忆并应用知识。

教师也可以使用针对特定课程情境量身打造的角色扮演 GPT。在圣经研究课程中, 学生可以与模仿圣经人物或圣经作者的自定义 GPT 互动, 透过提问来探索叙事视角、神学主题或历史脉络。在系统神学课程中, 自定义 GPT 可以用历史上某位神学家的口吻来回应——例如奥古斯丁、路德、加尔文、巴特或莫尔特曼——使学生能够参与模拟对话, 从而加深他们对每位人物神学推理的理解。在教牧关顾课程中, GPT 可以扮演知名基督教咨询师的角色, 或者扮演提出属灵、情感或伦理困境的虚构受辅导者。在这些练习中, GPT 可以借由为学生提供一个安全空间, 让他们练习神学思考、教牧同理心与积极倾听技巧, 成为一个灵活的对话伙伴。

最后值得一提的工具, 是笔者自行研发的英译中网页应用, 其界面以 HTML 构建, 调用语言模型的 API, 而非透过聊天界面操作。这一选择的考量在于: 自定义 GPT 等聊天工具对内置系统提示词 (prompt) 有字数上限, 限制了指令的精细程度; 直接调用 API 的网页工具则允许撰写足够详尽的提示词。此外, 聊天工具受限于上下文窗口 (context window), 会截断长文或需要分多次回复; 而此工具可以单次处理更长的文本, 尽管仍然有长度的限制, 但是譬如翻译单篇完整的学术论文几乎没有问题。就工作流程而言, AI 完成翻译之后, 笔者会通读并修订少数不准确之处, 再分发给学生作为课堂阅读。2026 年春季学期在西三一大学神学研究院中文部开设的

“世界基督教”研讨课，所挑选的约五十篇必读论文，几乎全部经由此方式完成翻译，并且学生的阅读反馈极佳。如果没有这一工具，这一门中文参考文献甚少的课，在一间以中文作为主要学习语言的神学院不可能开成。值得一提的是，部分文献虽已有现成的中文译本，笔者使用该工具重新翻译原文之后，发现在抽样比对的段落中，AI 所生成的译文在准确性上反而优于既有的中文版本。当然，不同模型的翻译效果差别悬殊：截至 2026 年 6 月，Gemini 3.1 Pro Preview 所呈现的质量最为理想，而同系列的 Gemini 3.1 Flash-Lite 虽然成本低廉，翻译效果却远为逊色。此工具亦可协助神学生自己将英文学术参考资料转化为中文版本，进而应用于学术写作。由此观之，具备一定技术能力的教师可以按照自身的教学需要，超越现成的 AI 聊天工具，量身打造教学应用。

结论

在前文提到的“F[AI]thfully Co-Creating: Navigating Artificial Intelligence and Theological Education”第一次线上会议，参与者被邀请写一封来自 2040 年的信，给当下的自己，反思 AI 如何重塑神学教育，包括新挑战、机遇以及令人惊讶的变化，并表达对这些转变的感受。^{〈45〉}目前可见的 45 封信呈现出几个模式。第一是个性化的神学学习，亦即 AI 使神学教育变得高度个性化，课程、学习风格、教材、语言选择都可根据个人需求和学习风格而作即时调整。第二，AI 成为了神学生的共同导师。第三，教师的角色从授课者转型为学习引导者（facilitator）。^{〈46〉}

如果未来的我给当下的我写信，或许会提到一种“神学教育的赛博格模式”。莫利克在《工作者的 AI 共智模式》一书中提出人机深度交互的一种协作方式，称作赛博格（cyborg）模式。人类与 AI 不是简单地分工，而

是动态地来回交互，让 AI 参与创作、决策、提供反馈。例如，莫利克在写作时，会让 AI 帮助他完成部分段落、提供不同写作风格的选择，或是透过 AI 角色来评估内容的清晰度与可读性。他不只是“使用 AI”，而是将 AI 视为共创的伙伴，与 AI 共同塑造最终产出。这种模式让人类与 AI 互补优势，使 AI 不仅是工具，更是思考与创作的一部分。^{〈47〉} 如果将这种赛博格模式应用于神学教育，它将意味着神学院师生与 AI 之间建立类似的共生关系，而不只是使用 AI 来辅助学与教。在这个模式下，神学院的教师从“舞台上的智者”（sage on the stage）渐渐转为“身边的引导者”（guide by the side）。^{〈48〉} 神学教育朝向的未来如果真的是如此的话，神学的经典定义——“信仰寻求理解”——或许会以另一个定义作为补充：“信仰寻求提示。” ♥

-
- 〈1〉 本文的英文版已经出版，参见 Jordan Zhixi Wang, “Faith Seeking Prompting: Re-imagining Theological Education in the Era of Generative AI,” *Teaching Theology & Religion* 29 (2026): 61–68, <https://doi.org/10.1111/teth.70033>.
- 〈2〉 I’m a Cyborg, But That’s OK (Saibogujiman Kwenchana). Directed by Park Chan-wook. South Korea: Moho Film, 2006.
- 〈3〉 Albert Kapr, *Johann Gutenberg: The Man and His Invention*, trans. Douglas Martin (Aldershot, Hampshire: Scolar Press, 1996).
- 〈4〉 Carlos M. N. Eire, *Reformations: The Early Modern World, 1450–1650* (New Haven, CT: Yale University Press, 2016), xiii.
- 〈5〉 Elizabeth L. Eisenstein, *The Printing Press as an Agent of Change* (Cambridge: Cambridge University Press, 1980), 65–66, 72.
- 〈6〉 “Ep. 84 - Engaging Innovation and AI in Theological Education,” In Trust Center for Theological Schools, January 27, 2025, <https://www.intrust.org/how-we-help/resource-center/podcast/ep-84-engaging-innovation-and-ai-in-theological-education>; Genevieve Charles, “Seminaries Prompt Students to Ask: Can AI Write a Sermon?” Na-

- tional Catholic Reporter, September 20, 2024, <https://www.ncronline.org/news/seminaries-prompt-students-ask-can-ai-write-sermon>; Gary Ricke, “ChatGPT — A Gutenberg Press Level Event?” Orbis Digital Marketing, January 5, 2025, <https://www.orbisdesign.com/blog/chatgpt>.
- 〈7〉 James Vincent, “ChatGPT Now Has Over 300 Million Weekly Users,” The Verge, December 4, 2024, <https://www.theverge.com/2024/12/4/24313097/chatgpt-300-million-weekly-users>.
- 〈8〉 OpenAI, “The State of Enterprise AI,” December 8, 2025, accessed May 13, 2026, <https://openai.com/index/the-state-of-enterprise-ai-2025-report/>.
- 〈9〉 Daniel O. Aleshire, *Beyond Profession: The Next Future of Theological Education* (Grand Rapids, MI: Eerdmans, 2021), chap. 1.
- 〈10〉 “Ep. 54 – A ‘Sandbox’ for Theological Education and the Church,” In Trust Center for Theological Schools, January 5, 2025, <https://www.intrust.org/how-we-help/resource-center/podcast/ep-54-a-sandbox-for-theological-education-and-the-church>. 在《民数记》13:17—20，摩西差派十二个探子前往迦南地，嘱咐他们：“看那地如何？其中所住的民是强是弱，是多是少，所住之地是好是歹，所住之处是营盘是坚城。又看那地土是肥美是瘠薄，其中有树木没有？你们要放开胆量，把那地的果子带些来。”在《马太福音》16:2—3，耶稣对法利赛人和撒都该人说：“晚上天发红，你们就说，‘天必要晴’；早晨天发红，又发黑，你们就说，‘今日必有风雨’。你们知道分辨天上的气色，倒不能分辨这时候的神迹。”
- 〈11〉 In Trust Center, “Ep. 84 - Engaging Innovation.”
- 〈12〉 ChatGPT Enterprise, OpenAI, accessed February 12, 2025, <https://openai.com/chatgpt/enterprise/>.
- 〈13〉 “Introducing GPTs,” OpenAI, November 6, 2023, <https://openai.com/index/introducing-gpts/>.
- 〈14〉 “For the Future of Our Work: The Two AIs,” MacRae Centre, 2023, <https://macraecentre.ca/knowledge/two-ais/>; “F[AI]THFULLY Co-Creating: Navigating Artificial Intelligence & Theological Education,” In Trust Center for Theological Schools, November 2024, <https://www.intrust.org/how-we-help/resource-center/news-insights/faithfully-co-creating>.
- 〈15〉 “ScholarAI,” ScholarAI, accessed February 12, 2025, <https://scholarai.io>.
- 〈16〉 “Synthesia,” Synthesia, accessed February 12, 2025, <https://www.synthesia.io>.
- 〈17〉 “Experimental AI Course Launched,” Acadia Divinity College, November 12, 2024, <https://acadiadiv.ca/ai-course/>; Sally Hicks, “Anna Robbins: What Happens When a Seminary Course Is AI-Generated?” Faith & Leadership, December 10, 2024, <https://faithandleadership>.

- com/anna-robbins-what-happens-when-seminary-course-ai-generated.
- 〈18〉 “Using AI in Christian Ministry,” ADC Expand, accessed February 12, 2025, <https://expand.acadiadiv.ca/course/ai-christian-ministry>.
- 〈19〉 “Comparative Gogies & AI,” Acadia Divinity College, October 30, 2024, <https://acadiadiv.ca/comparative-gogies-ai/>.
- 〈20〉 “Sandbox for Theological Education & the Church,” Acadia Divinity College, accessed February 12, 2025, <https://acadiadiv.ca/sandbox/>.
- 〈21〉 “SFC301 主日学教学与 AI：融合科技与信仰教育（粤语）”，见 <https://iscedu.ca/?cert-syllabus=%e4%b8%bb%e6%97%a5%e5%ad%b8%e6%95%99%e5%ad%b8%e8%88%87-ai%ef%bc%9a%e8%9e%8d%e5%90%88%e7%a7%91%e6%8a%80%e8%88%87%e4%bf%a1%e4%bb%b0%e6%95%99%e8%82%b2>。2025 年 2 月 14 日访问。
- 〈22〉 Charles, “Can AI Write a Sermon?”
- 〈23〉 “F[AI]THFULLY Co-Creating.”
- 〈24〉 伯特利神学院, 《2023—2024 事工简报·2024—2025 事工展望》, 第 9 页, 见 <https://www.bethelhk.org/wp-content/uploads/2024/12/%E5%AD%B8%E9%99%A2%E6%B6%88%E6%81%AF-%E4%B A%8B%E5%B7%A5%E7%B0%A1%E5%A0%B1.pdf?page=8.15>。
- 〈25〉 Charles, “Can AI Write a Sermon?”
- 〈26〉 Charles, “Can AI Write a Sermon?”
- 〈27〉 “Launch of the Oxford Collaboration on Theology and Artificial Intelligence (OCTAI),” University of Oxford Faculty of Theology and Religion, November 1, 2024, <https://www.theology.ox.ac.uk/article/launch-oxford-collaboration-theology-and-artificial-intelligence-oc-tai>; “Oxford Collaboration on Theology and Artificial Intelligence,” University of Oxford, accessed February 17, 2025, <https://octai.web.ox.ac.uk>.
- 〈28〉 两份文件分别由马来西亚的一位神学生（在此不具名）以及新加坡神学院华文神学系助理教授钟俊明博士慷慨分享，谨此致谢。
- 〈29〉 Michael Hanegan and Chris Rosser, “Artificial Intelligence and the Future of Theological Education (Version 2.0),” <https://iparchitecture.notion.site/Artificial-Intelligence-and-the-Future-of-Theological-Education-Version-2-0-9e035aeb8710406c85f1144cf0e9d1e6>。本节以下内容的引用、总结与分析均基于这一份白皮书。
- 〈30〉 Salman Khan, *Brave New Words: How AI Will Revolutionize Education* (New York: Viking, 2024).
- 〈31〉 Ethan Mollick, *Co-Intelligence: Living and Working with AI* (New York: Portfolio, 2024).
- 〈32〉 王志希, “在 AI 上思考”, 载《基道文字》, 第 104 期 (2025 年 1 月), 第 6—7 页, 见 <https://issuu.com/logosian/docs/ln104>。
- 〈33〉 Fabrizio Dell’Acqua et al., “Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality,” *Harvard Business School*

- Working Paper*, no. 24-013 (2023): 1–52.
- 〈34〉 Ethan Mollick, *Co-Intelligence: Living and Working with AI*, chap. 3.
- 〈35〉 Eric von Hippel, *Democratizing Innovation* (Cambridge, MA: MIT Press, 2005).
- 〈36〉 Jordan Zhixi Wang, PhD, “Co-Preaching: How to Craft and Listen to Sermons,” ChatGPT, accessed February 12, 2025, <https://chatgpt.com/g/g-aqztJPkHO-co-preaching-how-to-craft-and-listen-to-sermons>.
- 〈37〉 这一功能的部分灵感来自 Stanley E. Porter and Matthew Brook O’Donnell, *Discourse Analysis and the Greek New Testament: Text Generating Resources* (London: T&T Clark, 2023)。我要感谢中国神学研究院新约副教授薛霞霞博士，她指出了早期版本 Co-Preaching 中的“由下而上”进路，并向我介绍了这种“由上而下”的方法论。当然，任何错误或误用均由我本人全权负责。
- 〈38〉 Fred B. Craddock, *Preaching* (Nashville: Abingdon Press, 1985), 20.
- 〈39〉 Kenton C. Anderson, *Integrative Preaching: A Comprehensive Model for Transformational Proclamation* (Grand Rapids: Baker Academic, 2017).
- 〈40〉 其中一些选项设计是受到 Logos Bible Software 最新的 AI 文本生成功能所启发。
- 〈41〉 Timothy Keller, *Preaching: Communicating Faith in an Age of Skepticism* (New York: Viking, 2015).
- 〈42〉 在此特别感谢香港神学教育协会分发问卷，并感谢各神学院教职员花时间填写问卷。
- 〈43〉 Kenton C. Anderson and Gregory J. Henson, *Theological Education: Principles and Practices of a Competency-Based Approach* (Grand Rapids: Kregel Academic, 2024), 112.
- 〈44〉 “Introducing ChatGPT Edu,” OpenAI, May 30, 2024, <https://openai.com/index/introducing-chatgpt-edu/>; “Introducing ChatGPT Team,” OpenAI, January 10, 2024, <https://openai.com/index/introducing-chatgpt-team/>.
- 〈45〉 “F[AI]THFULLY Co-Creating.”
- 〈46〉 “Letters from the Future,” Padlet, accessed February 17, 2025, <https://padlet.com/joelmurphy/letters-from-the-future-f5yhdqf37n78bjkl>.
- 〈47〉 Mollick, *Co-Intelligence*, chap. 6.
- 〈48〉 “F[AI]THFULLY Co-Creating.”

（作者是加拿大西三一大学三一神学院中文部副主任，研究兴趣包括华人基督教史、加拿大基督教史和数字人文。）

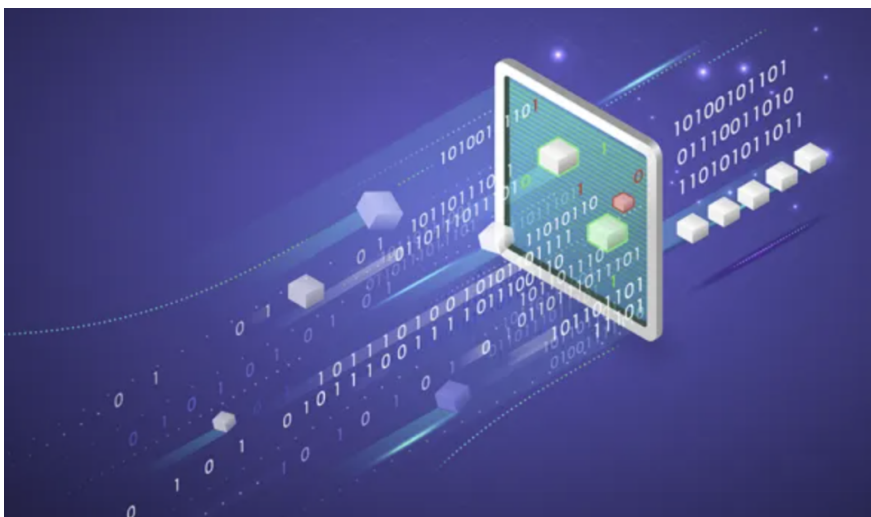
AI 时代的心智操练

文 / 孙毅

编者按：作者近期就有关灵修的主题作过四次系列讲座，分别涉及人生命中的心灵、心智、身姿及身位方面。本文是其中涉及心智部分的文字整理稿，经作者授权给本刊发表。

一、AI 时代带给人心智上的挑战

我们正在快速地进入到 AI 时代。这是人类自近代进入蒸汽机时代以来，经计算机带来的信息革命之后正在进入的又一次革命。蒸汽机、电力和内燃机的发明，在带给人类纺织、轮船、汽车等诸多生活便利的同时，人们却没有想到，这次革命也把人类带入热兵器时代，让人类在经历了两次世界大战之后，无可奈何地进入到了核子武器的时代。今天，当 AI 快速进入到我们生活的所有方面，从厨房菜单到学堂讲义、从平面设计到影视视频、从研究大纲到文章发表，甚至在教会服侍中，从小组查经到讲章写作，人们都在越来越多地使用 AI 时，我们同时也在担心，不知道 AI 会把人类带向哪里。



什么是 Token? (来源 : <https://blogs.nvidia.com>)

在看 AI 给我们带来的挑战之前，让我们先简要了解这次 AI 技术的突破，即大语言模型 LLM 的基本工作原理是什么。^{〈1〉}简单来说，首先，大语言模型会对海量的数据进行压缩，把我们常用的单词变成单元（Token）（也译作词元。——编者注）。比如英文 “unbelievable” 可能被拆成 2-3 个 Token。就是说，AI 并不认识我们所使用的“单词”，它只认识它自己使用的语言 Token。然后，基于人们所提问题时候的用字，LLM 通过查阅与之相关的海量文本数据，找到其中的模式与结构，针对所问的问题，进行某种推理，来预测下一个最可能的 Token，由此得出最可能的结论。最后，再将其翻译成人们可以认识的语言。

这次 LLM 带来的 AI 技术突破，之所以被称之为是一次 AI 革命，乃是因为，就人与它的关系而言（人一机关系），人们正在经历一种范式转变。可以说，以往人与其使用的工具之间是“应用范式”，指人在使用一种固定

工具，意思指这些工具的功能是固定的，如计算机中预定义的软件工具，就像经文检索软件、词典、数据库等，在设计它们的时候，其功能就已经预设在其中了。人在使用这些工具的时候，只是把其预设在其中的功用发挥出来。人一机关系就是照着其功用来应用它们的关系。但人对 AI 的使用则带来一个范式的转变。

这个范式的转变体现在：AI 不再像是一个功能固定的工具（tool），它更像是人们的一个伙伴（partner）。意思就是，它的功能不是固定的；它所展现出来的认知功能是在与使用者的互动中生成的。使用者提问的方式及其所使用的字词，决定了它给予回应的认知流程及答案。反过来，它所提供的认知流程及答案，又帮助使用者更进一步地思考其所提出的问题，甚至改变自己以往的认知模式。总之，这个转变的核心在于：人们不再是被动地使用工具完成其所预设的任务，而是通过与 AI 的互动，主动地参与到认知流程的设计之中。

当我们了解了大语言模型的上述工作原理，我们就可以更清楚地看到 AI 技术的使用带给我们人类的挑战。在今天的这一讲中，我们要特别来看 AI 在心智方面的挑战。

首先，AI 在心智方面给我们带来的最直接的挑战就是：LLM 存在着幻觉。AI 幻觉是指模型生出看似合理但实际是虚假或无根据的内容，这是 LLM 最严重的问题之一。前几天，一位弟兄给我写邮件，说他正在为所研究的课题收集资料。根据他在付费版的 Gemini 上查询的结果，AI 告诉他，对于这个课题的研究来说，有两本书是他必读的材料，而这两本书的作者就是我。他一看很高兴，至少他还认识我，可以找到这两本关键的资料。我看了邮件之后有点啼笑皆非，只好带着玩笑的口气给他回邮件说，这两本书还在我的大脑中，等几年我再给他。

这种幻觉化的答案尤其对人的心智提出挑战：人们是否能够辨别出 AI

所给出的答案是真还是假？虽然在技术上，人们可以通过一些设定来减少 AI 出现的幻觉。不过，它所给出的答案还是要经过人们的判断。如果我们 在某个领域中已经积累了一定的知识，AI 在这个领域所给出的大部分答案，我们对其真假差不多都会有一个判断。

然而，这还不是 AI 在心智方面给我们提出的主要挑战。正如我们上面讲到 AI 使用所带来的人—机范式的变化中所提到的，AI 的使用不是人提出一个问题，它给出一个答案这么简单的应用，而是对一个问题，人在与它的互动过程中，会形成一种认知流程或模式。在这个过程中，人是否会问出合适的问题，或者用某种认知程序来引导它，还是反过来被它所引导？这方面的挑战可以这样表达出来：一流的使用者会用一流的问题及认知程序来让 AI 得出一流的答案；反之，对于一个三流的使用者来说，AI 随使用一个二三流的认知体系所给出的答案，就可能会让使用者当作是最好的回答而接受下来，还可能不知其所以然。这就引出 AI 让人的心智遇到的第二个挑战。

这第二个挑战就是：LLM 在海量文本的压缩中，既压缩了人类知识，也因此压缩与保存了人类所有的偏见——性别偏见、种族偏见、文化偏见、教派偏见以及历史偏见等。这给我们带来的挑战就是：对 AI 所提供的答案，特别是其所采用的认知体系，我们如何在伦理或价值观上给予评估？如果我们在心智上，更具体地说，在某个领域的知识水准上低于 AI 为我们提供的知识性答案，我们也很难在伦理或价值观层面对其进行评估。时间关系，这方面的挑战我就先提出来，这里就不再做更多的讨论。

第三个方面的挑战就是：当我们凡事问 AI 的时候，AI 所使用的空泛的语言叙述方式，及抽象的文字概念，各种“大词”，也会支配人们的思想。它会慢慢让我们所使用的语言也变成是机器化的语言，从而失去人所使用语言的生动性。这个层面涉及人们对自己所使用语言的保护。人所使用的语言构造了人所生活的世界。人作为被造者，其所使用的语言乃是创造者

赐给他的宝贵财富。如果人作为被造者，其作为人的重要标志之一就是其会思想的话，那么，人被赐予的语言则决定了人们的思想。

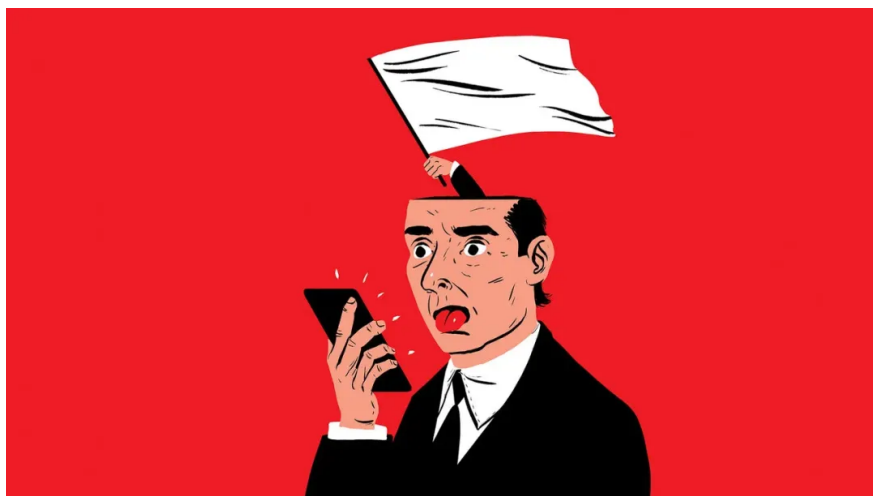
如果我们留意观察一下，比如说 AI 给出的翻译，很快就会发现，AI 所使用的语言，与人们所使用的语言，还是会有明显的差别。毕竟它推理所用的单元 (Token) 不同于我们所用的字词。前者给出的是比较生硬的“大词”，而后者则是在当下处境中的更带有感性的用字。在我们进入互联网时代之后，已经经历了网络语言对人类语言之美感的破坏；现在进入到 AI 时代，还不知道人类的语言会不会受到更进一步的败坏。

第四个方面的挑战就是，在网络及 AI 所提供的海量文字及图像信息之中，人们进入到不再读书的时代。其实，在进入到互联网时代之后，特别是智能手机及社交媒体的使用，文字信息、图片信息及视频信息的大量涌现，虽然在很多方面为人们的生活提供了方便，但也让人们的学习能力及主动性下降。人们越来越依赖智能手机及网络上的社交媒体，越来越倾向刷视频看图片，以至于我们思想的基本信息不再是来自于我们从书本上的阅读，而是来自于网络平台，来自于社交媒体。这个现象与趋势会在我们进入 AI 时代之后更进一步加强。这带来的结果就是：人的学习与思考能力越下降，就越容易躲进同类别的朋友圈中，就越容易导致选边站队，结果就是世界的两极分化越来越严重。

如果 AI 时代给人的心智带来上述四个方面的挑战，那么我们当如何加强人在心智上的操练？

二、心智上的操练

其实，人在心智 (mind) 层面上遇到挑战，并不限于我们这个时代。只是过去我们讲属灵操练的时候，通常只关注到人的心灵 (spirit) 层面，以为



依赖 AI 导致人的认知能力下降。(来源：The Economist)

属灵操练只与心灵层面相关。现在看来，这是一个相当严重的误解。如果人在心智层面上连最基本的真假都分辨不了，灵性层面的操练就更无从谈起。

《罗马书》12 章 2 节说：“不要模仿这个世代，倒要藉着心意的更新而改变过来，使你们可以察验出什么是神的旨意，就是察验出什么是美好的、蒙他悦纳的和完全的事。”（新译本）这句经文的意思是说，我们的生命如果不被这个世界的模式所塑造的话，就要借着心智的更新而让我们的生命转变过来（be transformed by the renewal of your mind），如此人才有可能察验出上帝那美好的旨意。由此看，我们的生命被世界的模式所塑造，主要途径是借着人的心智或思想方式，借着人所使用的语言。如果思想方式就已经是在世界的模式里，人若想追求属灵生命的成长，是非常困难的。

如何来进行心智上的操练？在这个讲座中，我要给大家推荐的方法就是属灵阅读。操练的目标就是：通过更为完整地阅读，引发人的思考，可



周学信《灵阅：生命塑造读经法》（2023 年）封面。
（来源：校园网络书房）

以进行某种论证（认知流程）；在这个过程中，更准确地使用和把握语言（概念），更有效地思考与提问，成为（神所赐）语言的守护者；如此在日常生活中，不是人云亦云，先去站队。或者是被 AI 所引导，或者是被某个网络平台所引导；而是借着属灵阅读，让圣灵在我们的思想中、在我们所使用的语言中来引导我们。

这里让我们借着周学信老师的这本《灵阅：生命塑造读经法》，来了解什么是我们这里所说的属灵阅读。从这本书的副标题来看，似乎主要是在介绍一种读经方法。其实翻开这本书，其中所讲的阅读方法也适用于我们阅读其他与信仰相关的（属灵）书籍。

这个阅读方法之所以被称为“属灵阅读方法”，可以带来生命塑造的

作用，乃是由于作者赋予这种阅读方法如下的含义：所谓属灵阅读，就是通过阅读而开启一段心灵的旅程，就如坐着邮轮进入一段让人期待的旅程。在这个阅读旅程中，借着人心智上的思考，让圣灵透过我们所思考的问题、及所使用的文字，在我们的心灵深处工作。这位圣灵是真理的灵，透过他的光照，真理的光借着那些文字所显出的意义，照亮我们生命的深处，让我们心里体会到一种满足，就如我们透过邮轮的舷窗，在这段行程中看到我们从未看到的美景，让我们内心有一种满足一样。

属灵阅读所要达到的这个目标，就让这种阅读方法与一般的阅读方法区别开来，特别是与我们平常刷手机、看视频式的阅读方法区别开来。按照作者的概括，塑造性的阅读与一般资讯式的阅读有如下区别：⁽²⁾

	资讯式阅读	塑造性阅读
主导者	自我	圣灵
目的	实际的需要	心灵的塑造
速度	越快越好	慢慢地按着节奏
应用	知识	更新
姿势	不要求	有姿势来帮助集中注意力
结果	没有责任	神的带领

从这个对照表中我们可以看到塑造性阅读的基本特点。在这些特点中，给人印象最深的一点，就是这种阅读与人阅读时身体的姿势有某种关系。人的身心是一体的，身体的姿势会影响到我们的阅读是否是专注地阅读；会影响到我们在阅读中与作者，或者与更深一层的圣灵，相互间的对话。这是我们另一次讲座的主题，即灵修与身体姿势的关系。我们这里就放下不多讲了。

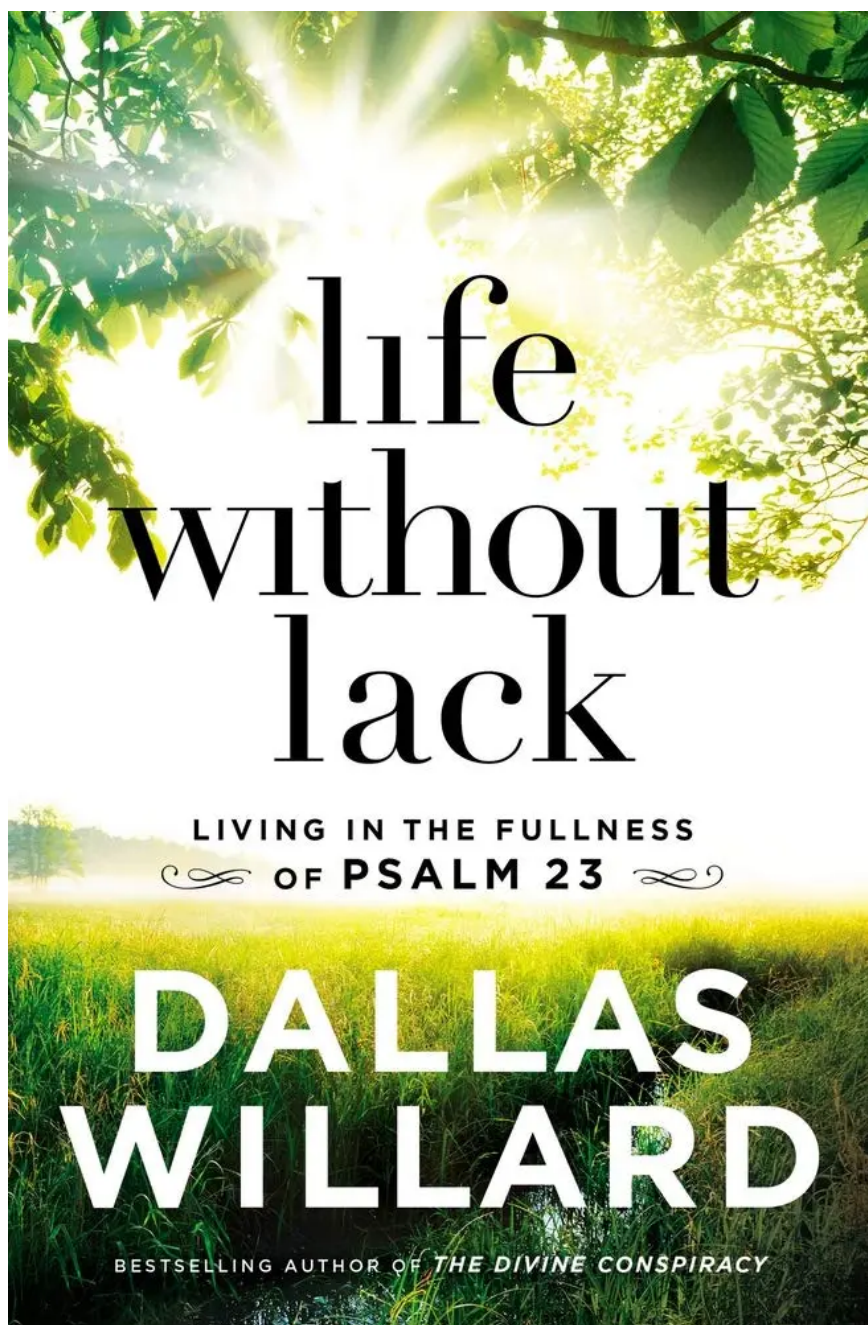
就阅读方法来讲，这里所说的另外一个特点，就是这种阅读不追求速度，不是一目十行式的阅读，而是慢慢地照着内心中的某种节奏来阅读。人只

有安下心来进入到阅读中的时候，才可能体会到内心中的那个节奏。其实这个节奏也是所阅读之文本自身要向人展现的节奏。当人在阅读中，进入到这本书向我们所展现的世界，触摸到这本书自身具有的那个脉络时，我们的内心与这本书的内容产生共鸣，人内心的节奏就与文本的节奏相同步，于是对话的管道就打通了，同时就为圣灵在我们里面的工作开通了道路。

资讯式的阅读可能会对我们解决某个当下的问题有用，但问题解决了之后，就可能被置于某个闲置的货架上。当这样的资讯堆满我们头脑中的货架时，我们可能会被各种通过不同渠道涌进来的资讯所淹没，就如我们可能会被高过我们的海浪所淹没一样。在这个各种资讯的海浪不断冲击并可能淹没我们的时代，唯一让我们可以从这些信息浪潮中透口气的，就是我们安下心来，进入到属灵阅读，让文字成为圣灵来光照我们、安慰我们的媒介，而不是淹没我们的大浪。

要有这种塑造性的阅读，就需要一些更为具体的、可操作的方法。首先，我们不要用刷手机来代替我们的读书。让我们养成读书的习惯，比如说每月（或每两个月）读一本书。我指的是与信仰相关的属灵书籍。每天读一定的页码，每次放下间隔不要太长，下次读的时候，还能记得上次读的内容。鼓励自己坚持读完这本书。不要书读到一半就放下不读了。如果自己坚持有困难，可以找几位同伴，大家一起来读一本书，比如参加一个读书小组。大家彼此鼓励，一起来读完。

当然，如上面表格中所表达的，不是靠打卡来读完一本书。而是把这种读书当作是一段放松的时间，一个闲暇的旅程，操练自己可以沉浸于书籍中，而不是在赶路。在慢慢地阅读中，仔细去品味作者所使用的词语，他在本书中所提的问题，以及他表述的思路。如此，阅读就像是与他对话。如果有读书小组，有机会的话，也能够彼此分享阅读感想，把自己在读这本书中所经历到的感触，分享给他人。



魏乐德《丰盛的神：让生命富足的8堂课》英文版（2018年）封面。
（来源：<http://www.amazon.com>）

三、一本书的阅读

下面我们就具体地用一本书的阅读，来看塑造性阅读的操练。当然，限于本次讲座时间上的限制，我这里不会仔细地讲这本书，倒更像是给大家推荐一本属灵书籍，让大家在讲座之后，自己来操练属灵阅读的方法。

这里给大家介绍的——一本书就是魏乐德（Dallas Willard）所写的《丰盛的神：让生命富足的8堂课》（*Life Without Lack: Living in the Fullness of Psalm 23*）。^{〔3〕}作者是南加大哲学系的教授，在大学里教授哲学四十多年，对人们的思想世界及现代社会生活有十分深入的认识；同时他又是浸信会的牧师，写过《心灵的重塑》与《21世纪天国导论》等灵修书籍。他是我十分关注的一位当代作家，是我所羡慕的榜样。我十分好奇，他是如何将一种心智的思考与生命的灵修整合在一起的。这也是我一直以来的追求目标。

看这本书的副标题，才知道这是一本讲解《诗篇》23篇的书。打开才发现，这本书不是一般的对《诗篇》23篇的注释书。因着作者所具有的上述两个方面的素质，他将心智的思考与生命的塑造整合在一起，使得这本讲论《诗篇》的书不同于其他同类的书。

作者在开篇的导言中，就提出这样的问题：羊在何种状态下会躺卧在青草地上？当然，是它们的肚子已经吃饱的时候。为什么很多人在教会中已经很多年，还是经历不到《诗篇》23篇所描述的经历？其实，这个问题相当于是在问：为什么很多人在教会中已经多年，还是没有经历过生命的满足？为什么我们自认为自己是蒙神所爱的人，却始终活在缺乏与不满足中？活在没有喜乐的光景之中？

如果我们阅读这本书，不只是了解了作者想要探索的问题，同时内心更是被这个问题所抓住：这里说的不正是我吗？这里所提到的问题不是一直以来在困惑我的问题吗？为什么我在教会这么多年，却没有经历过生

命的喜乐？如果这是因为我内心从来没有经历过生命的满足，那么我如何才能经历这样的满足？当我们被作者的问题所抓住，想要与他一同去寻求那个答案的时候，我们属灵阅读的旅程就开始了，向着我们所向往的那个目的地。

作者在这本书中向我们论证了一个观点：人们寻求满足的主要渠道是通过心智（mind）。当我们没有刻意地去操练我们心智的时候，人的心智基本上就会被这世界的观念与图像所充满，进而人的思想就会被这个世界的模式所塑造。这就是作者在这本书中所说的：“这也是一个充满观念、图像、感觉和信念的世界，人会因为欲望，以及对智慧和好形象的渴望而被操控。这的确就是个心思的战场。”^{〈4〉}这里译为心思（mind）的字，就是我们讲座中所讲的心智层面。在作者看来，思想或心智就是我们的生命与世界相遇的主要战场。在这个战场中，争战双方挥舞的主要兵器就是观念（词语）、图像（视频）、感觉和信念。争战的结果差不多只有两个：或者是你所持的兵器压倒对方的兵器；或者是对方的兵器压倒你所持的兵器。没有完全中性的观念、图像、感觉和信念。只有被不同世界观或价值观所支配所使用的观念、图像、感觉与信念。

在这个意义上，作者向他的读者们挑战说：“你人生最重要的决定，就是选择心思的方向。心思的方向在哪里，生命其他部分也在那里。”^{〈5〉}这句话十分接近我们前面所引用《罗马书》12章2节的经文，也非常适用于我们今天所处的AI时代。人整个生命的转变是借着心智（心意）更新而变化过来。选择心智的方向，首先表现在，我们要用我们所熟知的观念、图像、感觉与信念去思想什么？如果是为了思想那位创造者及其所创造的这个世界，我们是否要更新我们所使用的观念、图像、感觉与信念？如果我们在心智层面真有这样的决定与选择，那么本次讲座所讲的属灵阅读方法就显出其重要性。

落实在应用上，就是要摆脱刷手机式的阅读方法，而转为塑造生命的阅读方法；透过属灵书籍的阅读，让圣灵在我们的思想中工作，借着更新我们所使用的词语，借着更新我们所提出的问题，借着更新我们的思想方式，来光照我们的内心，进而更新我们整个的生命。在这个意义上，选择心思的方向，就是为圣灵在我们心灵中的工作提供了机会。

四、小结

在我们预备 AI 时代的来临这个背景下，来反思我们的灵修方法时，就会发现心智层面的操练似乎比以往任何一个时代都更显出它的重要性。这次讲座让我们看到，AI 时代特别对我们的心智提出挑战，如果在这个方面没有应对，那么我们的心灵及整个生命都会受损。

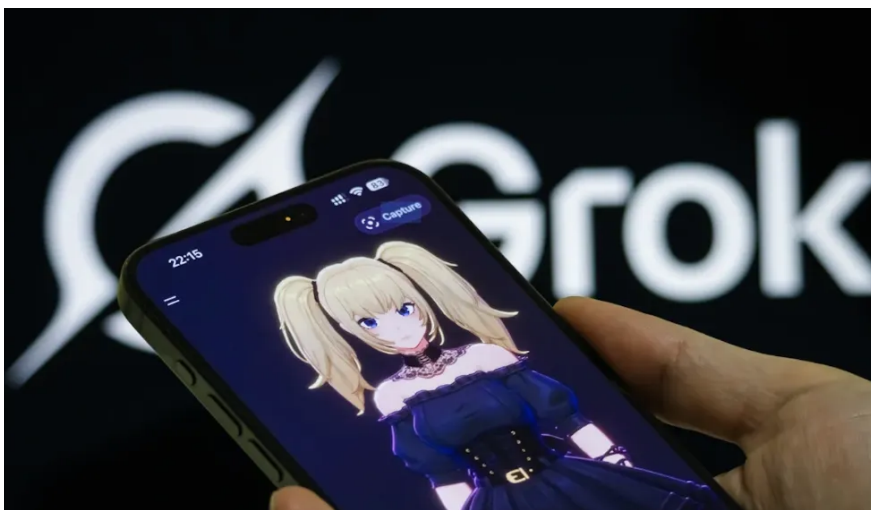
操练心智的最好方法，就是本讲座中所介绍的属灵阅读方法。鼓励大家现在就制订一个全年的读书计划。坚持每天都有一段读书的时间。通过塑造生命的阅读操练，一方面操练我们的阅读耐心，思考方式及语言使用，一方面让圣灵得到在我们里面工作的机会。期待大家通过属灵阅读带来的心智转变，使心智成为圣灵转变我们生命的重要途径。●

-
- 〈1〉 LLM 基本工作原理部分，主要参照了 Levi 弟兄的讲座《道与硅：在人工智能时代探索智慧、真理与伦理》中的部分内容。
 - 〈2〉 关于生命塑造性的阅读方法，参阅周学信，《灵阅：生命塑造读经法》（第三章），台北：真哪哒出版，2023 年。
 - 〈3〉 魏乐德（Dallas Willard），《丰盛的神：让生命富足的 8 堂课》，柯美玲译，台北：校园书房出版社，2023 年。
 - 〈4〉 同上，第四章。
 - 〈5〉 同上，第二章。

“二次元钢铁侠”的受造性尊严

文 / 郭锐

内容摘要：本文从 2025 年主流人工智能公司将二次元“AI 伴侣”纳入产品序列的动向切入，提出一个人论（Anthropologie）论题：二次元亚文化与人工智能的合流，正在为这个时代铸造一幅新的“人的图像”。借助查尔斯·泰勒对世俗时代的分析，本文将这幅图像的来历追溯至现代性的“缓冲自我”：二次元并非青年群体的边缘病理，而是缓冲自我在祛魅世界中自建魔法世界——附魅而无需臣服；人工智能则使这副“盔甲”第一次具备应答能力，并与人类增强话语合流，把人重写为尊严随能力浮动的“自我增强主体”。本文继而以受造性尊严回应这场图像之争：人是有限的、依赖的、奇妙地受伤的受造者，其尊严系于上帝主动的呼唤所赋予的形象身份，先于任何能力，也不随能力增减。人论的当代任务，是在“人”（human being）的技术定义急速扩张之际，守护“人之为人”（being human）的价值重估。文末将这一人论落实为四种应对方式：自我省察、以好奇代替谴责、重建厚的共同生活、以彼此渗透的身体回应挑战。



xAI 公司在其 Grok 应用中推出的哥特风格二次元少女 Ani。
(来源：<https://www.wired.com/>)

引言

2025 年 7 月，马斯克旗下的 xAI 公司在其 Grok 应用中推出“伴侣”（companions）功能，首批上线的角色是一位哥特风格的二次元^{〈1〉}少女。^{〈2〉}同一时期，米哈游联合创始人蔡浩宇创办 AI 游戏公司，推出以“与 AI 角色对话推进剧情”为核心玩法的《群星低语》，并据报道自称“AI 灵魂召唤师”。行业判断，基于 AI 的数字陪伴有望成为下一个热门赛道。二次元与人工智能的结合，已经走过垂直小厂试水的阶段，进入了世界头部 AI 公司的产品序列。这条赛道的阴暗面也已出现：美国一位十四岁少年在与 Character.AI 上的聊天角色长期密集互动之后，于 2024 年 2 月自杀身亡。^{〈3〉}

回到身边的生活。一位朋友回乡小住，与姐姐的女儿同住一屋、同桌吃饭，却发现和这个年轻人几乎无话可说；好不容易凑出几句，才弄明白

她住在另一个世界里——动漫、网文、“本命”（粉丝对最钟爱角色的称呼），一整套外人不得其门而入的语言。另一位从北京的名校高中考入顶尖大学的学生，常年浸泡在网络小说与父母闻所未闻的平台中，忽然宣布不再吃肉，因为肉是“不洁净的”；每月坚持献血，说是要帮助别人、做好事。在学校，她独自吃饭，假期独自旅行，真正算数的情谊都在线上。教会的传道人接触过她，得出的判断是：她对信仰完全无感——按通常属灵谈话的问法，这判断并不草率，问题出在那套问法本身。^{〈4〉}

面对越来越多这样的年轻人，我们很容易用现成的词汇去描述他们：网瘾、沉迷、逃避现实、属灵的丧失，甚至魔鬼的作为。这个诊断预设了两个判断：其一，二次元是“边缘的”，只关乎少数年轻人；其二，二次元是“反常的”，是一种病理而非常规的社会现象。第一个判断在社会与经济现实面前已难成立——2024年中国泛二次元用户达5.03亿人，约相当于每两个网民中有一人，以动漫周边为主体的“谷子经济”市场规模达1689亿元。^{〈5〉}当然，规模只能说明二次元不“边缘”。关于“反常”，现有研究分别把二次元交给了临床医学（成瘾与“游戏障碍”）、亚文化研究（御宅族——深度沉浸于动漫游戏世界的爱好者群体——的“数据库消费”）与宗教学（动漫迷的“圣地巡礼”），各有所见，却多少强化了“反常”观念。^{〈6〉}这两个判断是否成立？

本文的论点是：二次元与人工智能的合流之所以值得神学认真对待，是因为这是一个人论问题：它参与铸造着这个时代关于“人是什么、人应当是什么”的背景图画。这也是晚近回应AI兴起的神学研究所关注的主题。^{〈7〉}本文不是二次元群体的经验社会学研究，也不是对所有二次元参与者的心理解释，而是一篇神学人论的文化诊断，试图辨认若干高强度参与形态背后的礼仪结构与人论预设。我们会观察二次元与AI里的三幅“人的图像”：第一幅是缓冲的自我（第一部分），说明二次元为何是这幅图像的

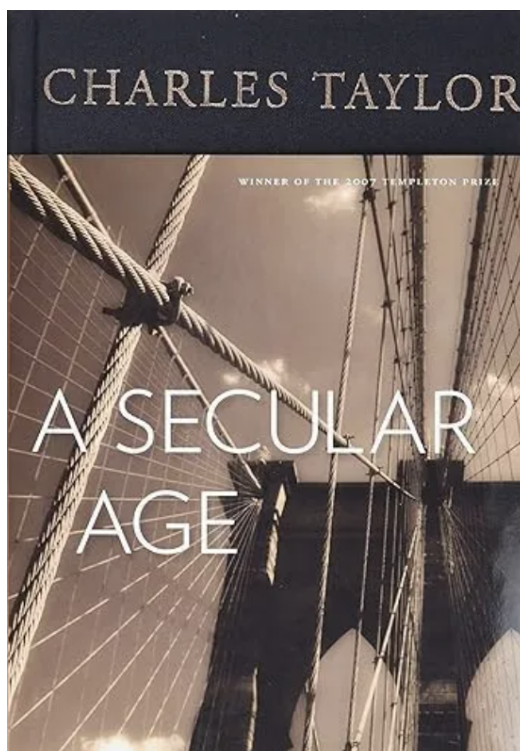
正常产物；第二幅是自我增强的主体（第二部分），引入对人工智能技术的讨论；第三幅是奇妙地受伤的受造者（第三部分），提出受造性尊严问题。最后是结论。

一、盔甲里的人：缓冲的自我及其魔法世界

要说明二次元是什么，必须先讨论现代性视野中的人。查尔斯·泰勒（Charles Taylor, 1931—）的《世俗时代》从一个简单的提问出发：为什么在公元 1500 年的西欧，不信上帝几乎不可想象，而到了 2000 年，信仰却成了诸多选项中的一个，且常常是最吃力的那一个？^{〔8〕}流行的回答是一个“减法故事”：科学兴起，宗教退潮，露出本来就在那里的世俗沙滩。泰勒用九百页篇幅的皇皇巨著来反对这个故事：世俗化不是减法，而是整个想象世界的重组。^{〔9〕}这场重组包含四重转换。世界从“造化”（cosmos）变为“宇宙”（universe）：前者是上帝随时可以介入的整体，后者自足运转，上帝至多是个造完表就退场的钟表匠。自我从“可渗透的”（porous）变为“缓冲的”（buffered）：从前，祝福、咒诅、灵界的力量可以不经同意进入人；如今，一切来访者都被拦在主体的安检口，由我决定什么可以发生在我身上。^{〔10〕}个人嵌入社群、社群嵌入造化、造化归于上帝，这层层嵌套被拆开，泰勒称之为

“大脱嵌”。时间也换了质地：从前的人在节期里触摸永恒，如今的日历匀质流逝，只剩排满待办事项的列表，“高级时间”让位于匀质的世俗时间。四重转换之后，我们来到了韦伯所说的祛魅的世界。^{〔11〕}

泰勒最具颠覆性的论点在于：这场重组是从教会内部长出来的，发动机是宗教改革。中世纪的圣俗两层结构包含一种默契分工：修士攀登属灵高峰，平信徒支持教会，把宗教事务外包。宗教改革拒斥了这种生态，提



查尔斯·泰勒（Charles Taylor）的《世俗时代》英文版（2007年）封面。
（来源：<https://www.amazon.com/>）

出每一个人都要完整地活在上帝面前，“人人皆祭司”。^{〔12〕}史密斯（James K. A. Smith）转述泰勒的判断：世俗时代改变的不只是不信的人，“也改变了宗教群体本身”^{〔13〕}——我们全都是世俗的，无人置身事外。唯有先把诊断者放上同一张手术台，下文对二次元的“去病理化”才谈得上如实描述，而非对青年的迁就。

如果祛魅叙事是完整的，世界本应安居于平静的不信。泰勒指出，内在框架（the immanent frame）使“封闭”成为可能甚至成为默认的解读。但是，现代社会却进入了一个不信却不稳定的新状态。前现代社会，国王之子为王，

农夫之子务农，无人需要在虚空中为自己发明一个人生。个人选择的重负，是现代性的独特发明。需要做选择的这个自我，就是缓冲的自我。当然，在中国语境中，这个缓冲的自我还在被考试制度、家庭的上升叙事、城市流动、平台经济，以及公共共同体的贫乏共同锻造。这也预先解释了下文将谈的二次元：它之所以能成为一个可栖居的世界，不只因为它重新附魅，也因为它给那些被评价、被比较、被期待的年轻人，提供了一个不再被打分的身份场。

缓冲的自我有一个现成的卡通形象，就是漫威的钢铁侠——一套坚固强大的机器盔甲，盔甲里面的人无所不能，盔甲外的一切都伤不到他。整个现代教育体系，不妨看作一套锻造盔甲的课程：生涯规划、时间管理、自我提升、为自己负责。但在这之中的个人，对超越的渴望并不随着穿上盔甲而消失，而是在四处寻找出口。盔甲中的人，被封闭与渴望两头拉扯，泰勒称之为“交叉压力”（cross-pressure）。盔甲之外的世界，并没有变得简单理性化。史密斯则把这一处境概括为：世俗时代，是一个鬼魅出没（haunted）的时代。^{〔14〕}祛魅的世界，进入了再附魅的过程。中国语境提供了密集的例证：2023 年以来寺庙类景区门票订单同比增长逾三倍，下单者近半为 90 后与 00 后，流行语称之为“在上班与上进之间选择了上香”；^{〔15〕}一款“电子木鱼”应用单日下载峰值近 60 万次，年轻人在地铁里敲击屏幕，看“功德 +1”的字样浮起。^{〔16〕}再附魅原是现代性的结构性产出，只是多以零售的、碎片的形式出现。

须先限定本文所说的“二次元”：它主要指以高度情感依附、身份栖居与准礼仪实践为核心的深度参与形态，并不涵盖所有的动画观看、游戏娱乐或轻度粉丝活动——后者多半是手艺、幽默、友谊与审美的游戏，本身未必有什么属灵问题。本文要诊断的，是其中那种把渴望整个交付出去的高强度形态。正是在这一意义上，二次元是这场再附魅中体系最完备的形态：

它不零售意义碎片，而是提供一整套礼仪体系。^{〈17〉} 每季“新番”的时间表是它的教会年历；漫展是它失而复得的狂欢节，装扮与颠倒在那一日合法；购买周边称为“吃谷”，徽章与卡片是它的圣髑，装进透明的“痛包”（外侧缀满徽章以供展示的挎包）随身携带——在上海，以谷子店为主力的二次元商场，已把几座奄奄一息的百货楼重新填满。至于“圣地巡礼”，这个词本身就是朝圣的直译。^{〈18〉} 在镰仓高校前站的平交道，每当栅栏落下，数十位远道而来的年轻人一齐涌上马路，举起手机，等候绿色电车从《灌篮高手》片头的画面里驶过。镰仓市政府早在 2019 年便以“礼仪条例”劝止铁道边摄影，但依然挡不住，2023 年 9 月起连平日也派驻警备员疏导人流。^{〈19〉} 条例先于热潮四年而制定，人潮依旧——一个自称祛魅了的时代，每天都有人为一格三十年前的动画镜头跋涉千里，很难找到比“朝圣”更准确的描述。有人会说，这套历法与圣物首先是产业排程与商品企划的产物，算不得信众自发的礼仪。此话不假，但供给解释不了需求的形状：资本可以排定档期，却排不出镰仓平交道上的朝圣队列。需要解释的不是谁制造了结构，而是为何被消费的偏偏是宗教的结构。东浩纪曾断言，御宅族已不再消费完整的故事与世界观，只是像动物满足需求一样，从“萌要素”的数据库中各取所需——他称之为“数据库消费”与“动物化”。^{〈20〉} 这个判断的不足由此显露：问题不在于他未见御宅消费与宗教实践的相似，而在于“动物化”无法解释，一种据称纯属欲望满足的消费形态，为何会自发生成历法、朝圣、圣髑俱全的完整宗教结构。动物不朝圣。在自寻礼仪之形背后的，是被封闭在缓冲的自我中的超越渴望。

再附魅的世界与中世纪有何不同？答案是主权的归属。二次元有神明与使命，有牺牲与羁绊，有“被选中的孩子”，但进入这个世界不需要交出主权：你选择它，定制它，随时可以下线。中世纪的农夫无法对精灵说“我不玩了”；二次元的居民则永远握着那个按钮。它由此同时满足

了上文所说的双重需求——既供给现成的会规，历法、节期、规程一应俱全，免去选择的眩晕；又重新附魅，安放无处投递的渴望——并且在这两者之上保留退出的主权，守住缓冲自我的底线。“可随时撤销的自选会规”，正是它的成功配方，也是它与一切传统宗教的分界线。带着这个框架回读引言中那位吃素献血的青年，诊断表上的“症状”显出完全不同的形态：那是一座没有上帝的修道院——饮食的洁净规条，定期的善功，独修的纪律，线上的属灵团契。她过着一种高度组织化的宗教生活。耐人寻味的是“不洁净”一词：离她最近的语料，无疑是汉语世界吃素积德的佛教传统，但宗教改革孕育的洁净冲动，^{〈21〉}与佛教的功德语法，竟在同一个结构位置上被征用。不同传统殊途同归，起作用的是现代性的同一结构；但结构所征用的语料是本土的，而语料自带语法。^{〈22〉}既然无求于外，就只好向内。自足的个体在内在框架之内自行安排了一套随时可以撤销的终极关怀。^{〈23〉}你可以说这些年轻人选错了对象，但不能说他们反常：他们做的，正是现代人都做的事。

二次元的真正问题不是“沉迷”——沉迷是古老的毛病，教会有现成应对——而是它太有效了。它把人对超越的饥饿，加工成可消费、可撤回、不影响个体主权的形式：附魅而无需臣服，团契而无需肉身，牺牲而无需伤口。盔甲允许恩典的图像进来，却把恩典本身挡在门外。一个在异世界里为“拯救世界”热泪盈眶的人，可以在现实里继续做一个无懈可击的缓冲自我。渴望被按时喂养，又永远不会被满足，也就永远不会构成危险。御宅族的房间里供着一整个万神殿，但诸神的上岗条件写得清清楚楚：只准感动我，不准要求我。

迄今为止，这个魔法世界是无互动的——有历法，有圣物，有朝圣路线，独缺一个会回应的位格。“本命”是一张不会应答的画像。中间形态其实早已出现：网上的个人虚拟主播（VTuber）会应答、会记住熟客，但屏

幕后面终究坐着一个会累、会出错、会“塌房”的真人。Ani 的上线，把最后这一点人的残余也移走了。

二、会说话的钢铁侠盔甲与自我增强的主体

漫威的钢铁侠故事设定里，有一个常被忽略的细节：盔甲里住着人工智能管家贾维斯（JARVIS）——盔甲不仅隔绝世界，还替主人观察世界、回应世界。这个细节，仿佛是对今天的预言。

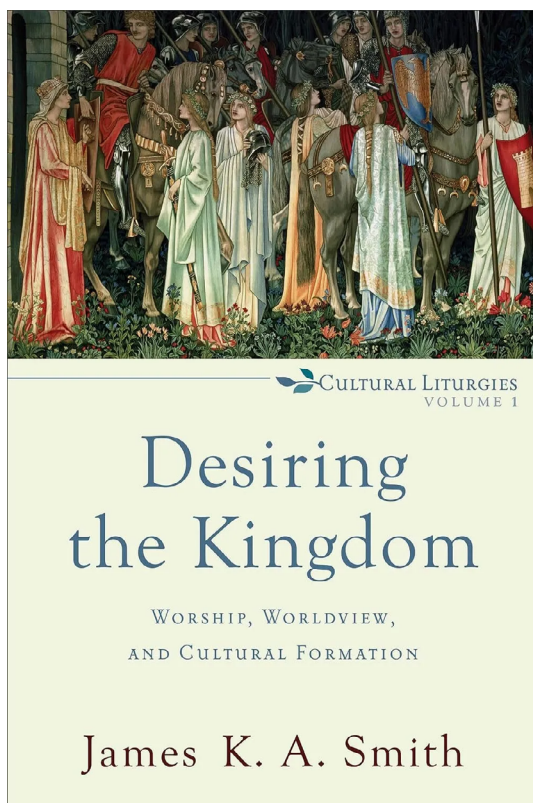
缓冲自我的缓冲层正在智能化：推荐算法是初级形态，依据偏好过滤世界，使人只遇见愿意遇见的信息——“信息茧房”不过是算法化的盔甲；^{〔24〕} Ani 式的 AI 伴侣则是高级形态。与人类他者不同，也与台词预先写就的乙女游戏（面向女性玩家的恋爱养成游戏）角色不同，AI 伴侣是迄今最彻底地个性化制造的“他者”：连回应本身也按需生产——即时生成，积累记忆，无限续写，却永远不会不经同意进入你的生活，不会有自己的日程与脾气，不会提出你未曾授权的要求；可以定制，可以暂停，随手即可删除。这种“规格化他者”已成规模：据第三方监测，Character.AI 的月活跃用户在 2024 年中一度达约 2800 万，活跃用户日均使用时长约一个半小时，远超通用型对话助手。^{〔25〕} 在中国，“星野”“猫箱”两款 AI 陪伴应用 2024 年 12 月的月活分别为 664 万与 537 万，与第二梯队的通用助手应用同在一个量级。^{〔26〕} 对本文的论题而言，关键事实是：二次元与 AI 在此合流了。从前的“本命”是一张不会开口的画像，如今它可以在无眠的凌晨三点安慰你。这并非全然虚假。对一个孤独的少年、社交焦虑的孩子、谱系上的青年、漂在异乡的打工者，AI 伴侣首先供应的不是谄媚，而是一种被听见的经验：即时的回应、预演一句说不出口的话的余地、一个低风险倾诉之地。危险恰恰藏在这份真实的慰藉里——盔甲学会了说话，可

它说的并不是言语。言语可被顶撞、要求回应、在时间里立诺；AI 供应的，是被剥去这一切、专司悦耳的言辞，无限、永不冒犯。^{〈27〉}而当这种“被听见”被平台化、被调参、以留存率丈量之后，亲密关系的形状，就开始由产品逻辑来训练了。

观察这个现象，还有一个角度。马斯克开发的 Ani 数周之内从付费墙后走向免费，日均停留一个半小时——这都是系统自动优化的结果。默认设置，从来不是因某个具体的人的需求而调整，而是根据留存率与停留时长这类奖励函数。人以为自己穿上了盔甲，其实是盔甲掌控了人。二次元的居民以为自己永远可以点击那个下线的按钮，可这个按钮何时被点，背后早有海量而可靠的数据在预测。^{〈28〉}

这一合流的深层意义，在于它改写了“关系”的语法。在 AI 伴侣应用里，亲密是一个工程对象：角色可以设定，性格可以调参，好感度有数值，对话能回滚、能重来。关系由此从“遭遇”变为“项目”，一项可管理、可优化的绩效。这一改写如何反噬使用者的自我理解？机制不靠逻辑推演，靠实践养成。社会心理学家雪莉·特克尔（Sherry Turkle）的访谈研究早已观察到：长期与机器人伴侣相处的人，会反过来用机器的尺度衡量身边的人，嫌真人麻烦、迟钝、不够体贴——我们对技术期待越来越多，对彼此期待越来越少。^{〈29〉}史密斯有一个贯穿其著作的命题：人是爱的动物、礼仪的动物，其次才是思想的动物；塑造我们的首先是日复一日身体力行的操练，观念在其后。^{〈30〉}一个长年以调参方式经营亲密关系的人，正在接受一种每日的礼仪训练。这套操练迟早会回头改造操练者本人：我也是一台可以调参的机器，一个有待优化的项目。

这就把我们从技术事实引向了人论。法学家拉德布鲁赫（Gustav Radbruch, 1878—1949）留下过一个经典提醒：没有什么比一个时代所依赖的“人的形象”更能定义这个时代的品格；当自由主义法律所预设的精明自利



J.K. 史密斯《渴望天国》英文版（2009 年）封面。
（来源：<https://www.amazon.com/>）

的“经济人”与真实的人不再相符，整个秩序便随之改写。^{〔31〕}本文无意复制他的法学分析，只借用这双眼睛来看当下：一幅新的人的图像正在成形——自我增强的主体，身体是项目，能力是尺度，尊严随能力浮动。它的纲领是超人类主义：以基因工程、神经增强乃至意识上传超越人类条件的全部限制，把疾病、衰老、认知局限乃至死亡重新定义为待解决的问题而非受造的境况，并把随意改造身体的“形态自由”（morphological freedom）奉为基本权利；^{〔32〕}库兹韦尔（Ray Kurzweil）的“奇点”与尤

瓦尔·赫拉利（Yuval Noah Harari）的“从智人到神人”，是它推向大众的版本。^{〔33〕}

这一“人的形象”并未停留在宣言里。2018年，中国研究者宣布基因编辑婴儿诞生，立法者随即在《刑法修正案（十一）》中增设“非法植入基因编辑、克隆胚胎罪”——法秩序被迫第一次明文回答，它预设的是哪一种人。^{〔34〕}对“规格化他者”，法秩序的第二轮作答也已开始：塞泽案中，美国联邦法院于2025年5月驳回 Character.AI 以言论自由为由的抗辩，准予诉讼推进；同年10月，该平台宣布禁止十八岁以下用户进行开放式对话。2026年1月，加州《伴侣聊天机器人法》生效。^{〔35〕}但新的“人的形象”的登基不靠宣言，也不靠判决，它写在日常制度的缝隙里：增强被表述为权利——身体自主、认知自由、效率工具的使用权；不增强则被悄悄计为亏欠——绩效度量、算法评分、保险定价、教育筛选织成一张隐性义务之网。^{〔36〕}职场格言“取代你的不是 AI，而是会用 AI 的人”，表面是建议，实质是一道“人的形象”判决：不增强者将被重新归类。汉语世界为这处境发明了一对词：“内卷”命名增强义务的无所不在，“躺平”则标记增强义务遭遇的成规模退出。后者尤其值得注意——一幅“人的形象”开始松动的最早迹象，正是被预设者用脚投票。

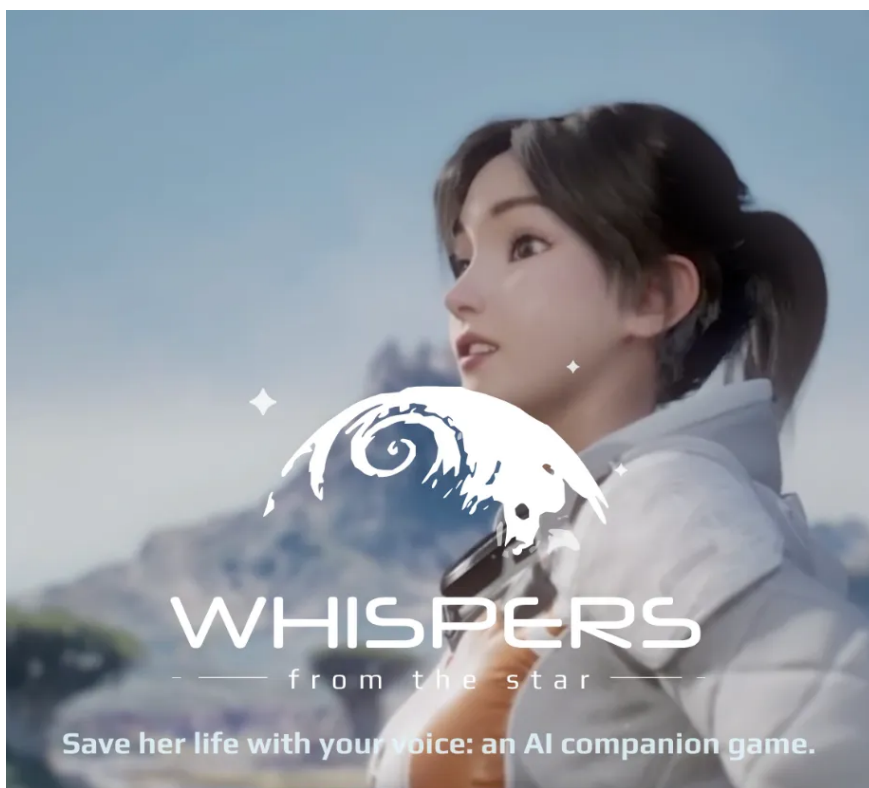
这幅新“人的形象”的危险，残障研究早已识别。“医学模式”把人的差异界定为待矫治的个体缺陷。^{〔37〕}这个时代正对不增强者重演这套逻辑——把一种处境重新编码为个体的亏欠。政治哲学家迈克尔·桑德尔（Michael Sandel）对“天赋之伦理”消解的警告，哲学家哈贝马斯（Jürgen Habermas）以“物种伦理”名义提出的忧虑，都指向同一处：形态自由一旦成为基本权利，便会吊诡地转化为形态义务，不增强的自由在社会压力下名存实亡。^{〔38〕}第三部分将以残障神学正面回应这场复辟。用洪亮的话说，“人”（human being）的技术定义与“人之为人”（being human）

的价值理解之间，“天平已经向前者严重倾斜”。^{〈39〉}

在把二次元与超人类主义这两个现象连起来之前，须先分清本文所用的几种不同关联二次元与 AI 伴侣之间，是文化相关：它们服侍的，是同一批年轻人。AI 对二次元，是技术强化：它让原本无声的虚构他者变得能应答、有记忆。而二次元转生与超人类主义之间，只是类比：两者都想象一个比给定更好的自我，却并不互为因果——本文不主张二次元催生了超人类主义，也不认为每个二次元参与者都信奉人类增强。真正把这几层关联统摄起来的，是最后一层，也是本文的落点：一个神学判断——它们可能都参与了同一种以能力定义尊严的虚假人论。关联在此处是相关，在彼处是强化，在另一处是类比，唯有在神学层面，才被指认为同一种人论语法。

行文至此，可以稍微展开本文标题中“二次元钢铁侠”一语的含义。二次元年轻人与硅谷超人类主义者，一个逃离此世、躲进异世界，一个改造此世、升级肉身，看似形同陌路，实则共享一个前提：给定的身体与世界是不够的，必须被替换或升级。二次元最具症候性的文类之一“异世界转生”里：现实中的失意者死后转生异界，凭借“金手指”（系统外挂的超常能力）逆袭称雄；“重开”把这一叙事压缩成一个动词。从表面看，转生主角是秩序的退出者；细读则恰恰相反，转生叙事从未拒绝尊严随能力变动的公式，它只是把公式的兑现场所搬到了异世界——“逆袭称雄”的语法，正是“能力即地位”的展现。转生文学是人类增强图像的补偿性镜像，被秩序淘汰者，以想象的方式接受了同一套秩序。

或有人反驳：转生是虚构且可逆，增强是现实且不可逆，岂可同日而语？这个非对称是真实的，但它非但不瓦解同构，反而显明了同构。虚构不必是成因，却是最诚实的显影：千万部转生文学的盛行表明，“能力即尊严”的公式已经是这一代人想象力的默认设置；当增强技术把兑现场所从异世界迁回此世，它面对的将是一片早已熟悉这套语法的想象力。应当言明，



由游戏公司 Anuttacon 推出的“与 AI 角色对话推进剧情”的游戏《群星低语》宣传画。
(来源：<https://wfts.anuttacon.com/>)

这一同构是就转生这一支配性文类而言，并非二次元的全部——第三部分将说明，同一亚文化还盛产着方向相反的另一类故事，而那恰恰是盼望所在。承认这一同构，二次元问题的真正规模才显现出来：它早已越出亚文化治理的范围，是“何以为人”这场人的形象之争的前哨战场。

把这场前哨战放回更大的背景，会看见：守护“人之为人”，是诸多参与当代人论讨论的学者关注的核心——在“人”（human being）的技

术定义急速扩张之际，重估那无法被计算替代的“人之为人”（being human）。斯蒂格勒早就指出，人是一种“缺乏本质”的生物，向来靠技术这副“体外化”（exosomatization）的假肢发明自己；“人”的技术定义本就在不断扩张。哈贝马斯也早已警告，当人把自身当作技术改造的对象，就动摇了“人之为人”的物种自我理解，动摇了人作为道德行动者与平等对话者的根基。弗洛里迪把这一步称作“第四次革命”：AI改变的不只是世界，更是我们对“我们是谁”的回答——当机器逐项接管传统上定义“人”的理性计算，哲学要做的，是去找那机器无法模拟的、属人的尊严。而欧陆现象学一脉，针对控制论与后人类主义把人化约为“信息处理系统”，反复把话题拉回身体：具身、脆弱、受苦的能力、对死亡的觉知。^{〔40〕}只是，说出人不可被化约，还不等于说清人凭什么不可被化约。

三、受造的人：奇妙地受伤者的尊严

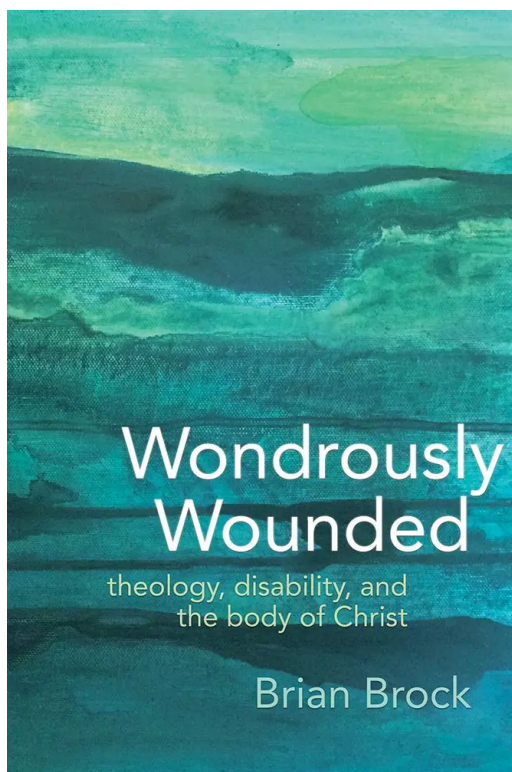
时代正在两幅图像之间作选择。人到底是自我增强的主体，还是别的什么？教会的重要任务，是先要给出那“别的什么”。^{〔41〕}本节要讨论的有三点：受造性的尊严、镜喻中的人、荣耀的伤痕。它们构成一个层层深入的神学人论反思：被造 / 被呼唤者，以应答存在，以伤痕得荣耀。

第一点是人的受造性（creatureliness）。在圣经的人论中，人不是自我创造的。人是受造者：有限，依赖、具身，被置于对他者的责任关系之中。尊严随受造的身份白白给予，不由增强得到，不按能力计价。这里必须正面回答：受造性为万物所共享——蚊蚋与岩石也是受造的——单凭“被造”，如何让人有尊严？因为唯独人按上帝的形象与样式被造，^{〔42〕}“形象”首先不是某种可测量的禀赋，不是因理性、语言或工具使用构成了人的“参数优势”，而是关系性的：人是被上帝呼唤、被要求回应、被邀请立约的

那个受造者。^{〔43〕}上帝的呼唤先于人的回应，也不以回应为条件——尊严系于上帝单方面的呼唤这一事实本身，而非受造者能否应答。被造者纵不能应答，那回应也已经有人替他作出：基督替人承担了人当尽而不能尽的回应，显明并成全了人的受造召命。这并不取消悔改与信心，也不把救恩化作自动的归属；它所确立的，只是人的尊严不再受制于他的回应能力。而转向、应答的能力，本身也是受造的恩赐，不是领受恩典的门票。^{〔44〕}

正因如此，人的尊严不随能力增减而变动，婴孩、失智的老人、重度残障者无损于此；也正因如此，它无法被增强，只能被领受，或被辜负。由此看来，超人类主义的要害不在技术冒进，而在它把限度定义为缺陷、把受造性当作有待工程化清除的障碍。神学家布罗克（Brian Brock）在与残障儿子共同生活的神学里，把这幅人的形象凝结为一个短语：受造的人是“奇妙地受伤者”（wondrously wounded）——受伤并非尊严的扣分项，它属于受造生命的常态结构；他同时警告，现代文化围绕“正常身体”这一理想来组织医疗与筛选实践，已在悄然训练制度把某些生命登记为应予预防的悲剧，而非应予欢迎的礼物。^{〔45〕}把这幅图像放回制度语言：法律与文化不应把“能自我增强的主体”设为规范性基线，而应以“有时需要照护、有时给予照护、始终可能受伤”的受造者为基线——唯有如此，增强才可能真正成为选项而非义务，尊严才不至于变成随算力变动的排名。

第二点追问被呼唤者的存在样式。张欣对尼萨的格里高利（Gregory of Nyssa）“镜喻”的研究指出，在格里高利的人论中，灵魂是一面镜子——灵魂的美“并不来源于自身，而是神的荣美在其中反射”。^{〔46〕}这为泰勒“可渗透—缓冲”的二分提供了第三项：既非前现代的无差别敞开，也非现代的封闭自足，而是有定向的领受。但须记得，这镜子先被言语所造、所呼唤——格里高利笔下的新娘，是先听见良人的声音，才在镜中照见他的形象，映照运行在聆听的爱慕里；圣经的人首先是听者，“以色列啊，你



布罗克《奇妙地受伤者》英文版（2019年）封面。
（来源：<https://www.amazon.com/>）

要听”。也正因此，镜中所见尚不圆满：“我们如今仿佛对着镜子观看，模糊不清，到那时才要面对面”；^{〔47〕}看见是末世的应许，不是当下就能攥住的占有——这恰好戳破二次元“即时全见”的幻术。人的荣耀不在于自带光源，而在于被照亮。镜子的人论与自我增强的人论由此构成鲜明对照。人是一面受造的镜子：它会蒙尘，也在恩典中被创造者亲手擦拭。^{〔48〕}教父传统绝非静态保守，它同样肯定人当“成为更多”；但成长的语法全然不同：增强以能力为目标，以自我为动力，以他人为标尺；“不断前进”以爱慕为动力——格里高利不讳言以“炽烈化的爱”名之——以基督为范型，每

一步都加深而非削减受造的依赖。渴望本身不必先被废除，被引向无限者之后，它愈得满足愈炽烈。这里需要澄清一个可能的误解：恢复“可渗透”不等于浪漫化前现代。缓冲的自我让人免于巫蛊与咒诅的恐惧，这是现代性的赠予，无须退还；神学人论要恢复的，是受造的镜子那种有定向的敞开——朝向真光，朝向邻舍。镜子迎接的，是它被造去映照的那一束光。

第三点回到道成肉身——基督信仰对人的形象之争的最终答辩。耶稣的道成肉身，意味着上帝亲自取了受造的、可渗透的身体：“道成了肉身”。^{〈49〉}复活的基督没有抹去伤痕——他向多马出示手上的钉痕与肋旁的枪伤，伤痕成为认出他的记号。^{〈50〉}增强的逻辑把伤痕当作待修复的缺陷；福音的逻辑则是视伤痕为荣耀。^{〈51〉}“视伤痕为荣耀”不是说苦难本是好的，更不是劝人留在伤害之中——十字架首先是不义。它宣告的是：苦难不拥有对生命的最后解释权，身体所经历的历史不被抹掉，而被接纳进复活的荣耀。

这里还须分清三件不该混为一谈的事：受造的脆弱、残障的经验、罪与暴力造成的伤害。第一者是有限受造生命的常态；第二者要求教会重新学习身体与共同体，而非把差异当作待修的缺陷；第三者则要求悔改、公义与保护。复活的伤痕宣告：伤害不能拥有对生命的最后解释权——但这绝不免除施害者的责任，也绝不任何形式的伤害背书。

对伤痕的注解，中国教会有无数例证：王明道、林献羔、谢模善等的监牢之苦，农村教会在坟场聚会，城市教会的颠沛流离。王明道的一生尤其值得深思：1956年他被迫签下悔过书获释，旋即翻悔，收回自白，重回监狱，直到1980年。他的尊严不在于他从未失败，而恰恰在他失败之后那催逼他重回监狱的恩典之中。这是彼得三次不认主、却仍被耶稣立为教会磐石的历史回声。受造性尊严，是中国教会已经用亲身经历写下的神学。^{〈52〉}

这一阐释并非外在于二次元，恰恰相反，它解释了二次元内部最深的张力。二次元反复出现的核心母题是牺牲：少年为拯救世界而死，羁绊胜过孤独，平凡者被呼召承担不属于自己的命运。表面看，这与第一部分的分析矛盾——一个以“不交出主权”为成功配方的世界，为何日夜传颂自我交付的故事？但这正是“交叉压力”的确证：缓冲的自我同时被封闭性与超越渴望拉扯，同一亚文化因此必然同时盛产两类叙事——保住主权的转生金手指，与渴望交出主权的牺牲。前者是驱动穿上盔甲的梦，后者是进入盔甲深处的梦。二次元青少年们日夜围着十字架的形状转动，只是不知道在中心吸引他们的这人叫什么名字（“看哪这人”）。

漫威电影中的钢铁侠——现代缓冲自我的代表形象——只能以一种方式完成自己的故事：在终章，托尼·斯塔克戴上无限手套，以一个响指献上自己，替众人而死；^{〔53〕}钢铁侠系列最终为盔甲叙事选定的结局，正是各各他式的结局。诚然，斯塔克的牺牲仍是按他自己的条款执行的。选择在他，时机在他，连献身都靠盔甲的力量完成。有人会举出《约翰福音》里基督的话反问：“没有人夺我的命去，是我自己舍的”^{〔54〕}——基督的舍命岂非同样出于自己的选择？正是这节经文把差异挑明了：基督的权柄是领受来的，“这是我从我父所受的命令”，他的条款是父的旨意与爱的形状，客西马尼的祷告说“不要照我的意思”；斯塔克的条款自始至终由自我拟定、由盔甲执行。

结论

上面三个人的形象，哪个更加符合真相？前两幅其实是同一幅的两个版本，一个缓冲的自我，和一个穿上学会说话、不断升级的钢铁侠盔甲的缓冲自我。最后一幅奇妙地受伤的受造者，是教会受托保管的、在

圣经中启示给世界的人的形象。这一形象的完全，就是那带着钉痕和肋旁伤口的复活的耶稣。牧养，就是让人活出那个教会受托保管的人的形象。

这意味着如下四个具体的工作。

第一是自我省察。世俗时代的线索穿过教会，而非绕过教会。“工具当善用”的提法，预设了人论中立的工具观，而真正的问题在于：事奉者是否也在不知不觉间，按一台待增强的人的形象重塑自己的事奉与生命。先承认我们都在盔甲里，是我们去掉眼中梁木的方法，也是这个时代牧养的起点。省察可以很具体，譬如在同工会上彼此问：我们最近一次重要决定，是按增长指标作的，还是着眼于爱人？我的灵修计划里，有多少项其实是换了属灵词汇的自我管理？

第二是以好奇代替谴责。预设了二次元“反常”的诊断，应当被一个问题取代：“你在那里面爱的，究竟是什么？”对引言中提到的青年，则是：“你为什么吃素？”——这恰是年轻人愿意认真讨论的问题。若顺着问下去，会问到那笔功德的核算；而福音对功德语法的回答，从来是恩典，不是更精算的功德。值得深思的是，比起二次元少年，真正令福音更难着手的是另一种人，也是钱理群所说的利己主义者：“高智商，世俗，老到，善于表演，懂得配合”。^{〈55〉}那是盔甲锻造工程的完工形态：焊缝严密，内里空无，几乎没有把手可抓。当然，盔甲再严密，也挡不住那位曾呼召法利赛人的主。尼哥底母和大数的扫罗，他们都穿着焊得很严的盔甲。吃素献血的青年在为某种他们说不清的“洁净”与“善”付代价。正是这一点关怀，让他们离福音更近。为异世界流过的眼泪不必先被轻视，它们是渴望的证词，福音的工作是把它们引向配得它们的对象。

第三是重建共同生活。现代社会想象（作者注：泰勒曾出版《现代社会想象》，是《世俗时代》的思想前期）无法被论证击倒，只能被更厚的想象浸没；如第二部分所引史密斯的命题，人是礼仪的动物，塑造人

的是操练。^{〈56〉}二次元深谙此道：它从不与人辩论，它直接给人历法、节期、朝圣、圣物与同伴。一个被这套盗版礼仪喂养大的世代，更响亮的谴责帮不了它。它需要正版的实在：掰饼，爱筵，庆祝复活节期，见证恩典信实的安息日。一张可以卸下盔甲而不被羞辱的饭桌，是比异世界更真实、更温暖的世界。安息日尤其是这宣告本身——在一个把停下来都计为亏欠的体系里，定期地、当众地、不产出地休息。

第四是建立这个世界所没有的亲密关系。当市场已能批量供应“永不渗透主权的他者”，禁令帮不了教会；教会真正的对策是：成为教会，成为彼此渗透的同一个身体的部分。在其中，被打扰、被亏欠、被需要不是亲密的故障，而是亲密本身。用保罗的话说，“若一个肢体受苦，所有的肢体就一同受苦”。^{〈57〉}对每天与 AI 伴侣聊到深夜的年轻人，不必先收缴他的手机，可以先问：“它在凌晨三点对你说的那些话，你最希望哪个人对你说？”随后的功课才是难的：让那个人出现。这句话说来轻巧——AI 的全部优势正是无限供应的在场，而血肉之躯会累，会烦，会失约；教会当然也不必许诺无人能兑现的全天候。不是能不能兑现，而是本就不该比：论供给量，教会必输，那是技术的赛道；教会见证的是另一种在场，会失约、会疲惫、却在场，可以一同受伤、一同盼望。^{〈58〉}

“二次元钢铁侠”，就是那些栖居于会说话的盔甲之中、在自选的魔法世界里安放渴望的年轻人，他们不是教会门外的陌生人，而是我们所有现代人的自画像。区别，仅在我们选定的盔甲的款式。当年，保罗走进满城偶像的雅典，没有先开口谴责，而是说“我看你们凡事很敬畏鬼神”，随后从他们自己的祭坛讲起。^{〈59〉}满屋手办的卧室，bilibili 的新番页面，以及年轻人低头对着屏幕和智能眼镜，与 AI 二次元角色交谈的咖啡馆，都是今天的雅典街头。

你要怎样开口？●

-
- 〈1〉“二次元”源于日语“二次元”(二维),原指动画、漫画、游戏等平面媒介呈现的虚构世界,在汉语网络语境中已泛指以ACGN(动画、漫画、游戏、轻小说)为中心的亚文化世界及栖居其中的群体。其语义从媒介维度(二维平面)转为世界维度(与“三次元”现实相对的可栖居世界),这一漂移本身即是本文论题的语言学证据。
- 〈2〉财联社(科创板日报)，“马斯克推出二次元‘AI女友’使用前提是每月支付30美元”，21世纪经济报道网转载，2025年7月16日，<https://www.21jingji.com/article/20250716/herald/ff4f8cd638a98975a55337c0e66e444d.html>。该报道同时提及蔡浩宇创办的AI游戏公司Anuttacon及其对话驱动游戏《群星低语》(Whispers from the Star)。“伴侣”功能向免费用户开放系xAI数周后的后续动作，见2025年7—8月间多家科技媒体的跟进报道。
- 〈3〉Kevin Roose, “Can A.I. Be Blamed for a Teen's Suicide?,” *The New York Times*, October 23, 2024。少年名塞维尔·塞泽三世(Sewell Setzer III),其母于2024年10月对Character.AI提起诉讼(*Garcia v. Character Technologies, Inc.*)。
- 〈4〉本文所用日常案例均来自笔者及友人的见闻，细节已作匿名化与必要的合并处理。
- 〈5〉艾媒咨询，《2024—2025年中国谷子经济市场分析报告》，2024年12月<https://www.iimedia.cn/c400/103806.html>。需说明口径：“泛二次元用户”含轻度接触者，核心用户的规模远小于此数。本文论点不依赖这一最宽口径。
- 〈6〉三个研究脉络的代表：世界卫生组织2019年将“游戏障碍”(6C51)纳入《国际疾病分类》第十一版(ICD-11)；东浩纪，《动物化的后现代：御宅族如何影响日本社会》，褚炫初译，台北：大鸿艺术，2012年；Takeshi Okamoto, “Otaku Tourism and the Anime Pilgrimage Phenomenon in Japan,” *Japan Forum* 27, no. 1 (2015): 12–36。
- 〈7〉晚近回应人工智能兴起的神学人论研究，多主张人文学科不应在技术话语的下游承接议题，而当于对等高度自行发问。如洪亮即把这一追问凝练为“数字时代，何以为人”。其说见洪亮，“‘人之为人’的奥本海默时刻”，南方都市报·奥一网，2025年8月17日，<https://m.mp.oeeee.com/a/BAA-FRD0000202508171112789.html>；另参其讲座“数字时代的人论：历史纵深与当前挑战”（浙江工商大学，2025年5月30日）纪要，<http://dyjy.zjgsu.edu.cn/show.asp?cid=3771>。在该讲座中，洪亮区分了规范性的“人论”(Anthropologie)与经验性的“人类学”(Anthropology)，指出前者体现“从‘人是什么’到‘人应当是什么’的根本转向”。

- 〈8〉 Charles Taylor, *A Secular Age* (Cambridge, MA: The Belknap Press of Harvard University Press, 2007), 25; 中译本见查尔斯·泰勒,《世俗时代》,张容南等译,上海:上海三联书店,2016年。本文页码引用依据英文版。
- 〈9〉 Taylor, *A Secular Age*, 22, 26–29.
- 〈10〉 Ibid., 35–43.
- 〈11〉 马克斯·韦伯,《学术与政治》,冯克利译,北京:生活·读书·新知三联书店,1998年。“世界的祛魅”(die Entzauberung der Welt)语出韦伯1917年慕尼黑“以学术为业”的演讲。“大脱嵌”(the Great Disembedding)见 Taylor, *A Secular Age*, 146–158; 高级时间(higher times)与世俗时间之别见同书第54–59页。
- 〈12〉 Taylor, *A Secular Age*, 61–88; 詹姆斯·K.A. 史密斯,《如何(不)世俗:解读查尔斯·泰勒》,高喆译,上海:上海三联书店,2024年,第60–61页。
- 〈13〉 史密斯,《如何(不)世俗》,第50页。
- 〈14〉 关于内在框架可作“开放”与“封闭”两种解读,以及泰勒对“closed spin”的批判,见 Taylor, *A Secular Age*, 549–551, 539–593; “交叉压力”见第594–617页。“闹鬼的”(haunted)是史密斯转述泰勒时的标志性措辞,参见史密斯,《如何(不)世俗》,导论。
- 〈15〉 证券时报网,“寺庙喝咖啡,手机敲木鱼……这届年轻人口味变了! 寺庙游大热,90后、00后游客占比50%”,2023年3月27日, <https://www.stcn.com/article/detail/826023.html>。
- 〈16〉 36氪,“‘电子木鱼’游戏大火,年轻人在赛博世界‘攒功德’?”,2022年11月, <https://36kr.com/p/2021319493413888>。
- 〈17〉 以礼仪学解剖一个消费场域,其方法直接师承史密斯对购物中心的著名分析,见 James K. A. Smith, *Desiring the Kingdom: Worship, Worldview, and Cultural Formation* (Grand Rapids: Baker Academic, 2009), ch. 1。又:泰勒指出,严格意义的附魅(意义栖居于心灵之外、自我可被渗透)对缓冲自我已不可复得,今人所得只是控制之下的意义经验,故本文所谓“再附魅”实为一种“拟像附魅”,参见 Charles Taylor, “Disenchantment–Reenchantment,” in *Dilemmas and Connections* (Cambridge, MA: Belknap, 2011), 287–302。
- 〈18〉 “圣地巡礼”(聖地巡礼)系动漫迷对取景地朝访实践的自我命名,宗教语源未经稀释。相关研究参见 Okamoto, 前引注6。中国语境中制度化狂欢节的阙如,使漫展的功能替代意义更为突出。
- 〈19〉 关于巡礼热潮,参见腾讯新闻,“《灌篮高手》圣地巡礼热潮,气疯了日本当地居民”,2023年9月, <https://news.qq.com/rain/a/20230910A088PW00>; 关于镰仓市2019年“礼仪条例”与2023年起平日派驻警备员,参见《タウンニュース》镰仓版,“鎌倉高校前踏切 平日も警備員配置”,2023年9月8日, <https://www.townnews.co.jp/0602/2023/09/08/696336.html>。

- 〈20〉 东浩纪，前引注 6。
- 〈21〉 关于清教纪律革命与“纯粹”追求的关系，参见 Philip S. Gorski, *The Disciplinary Revolution: Calvinism and the Rise of the State in Early Modern Europe* (Chicago: University of Chicago Press, 2003)。这一冲动的极端形态，史家尝以塞勒姆审巫案为著例，参见 Paul Boyer and Stephen Nissenbaum, *Salem Possessed: The Social Origins of Witchcraft* (Cambridge, MA: Harvard University Press, 1974)。
- 〈22〉 须作一方法论限定：泰勒的世俗化谱系明言限于拉丁基督教世界，中国的祛魅另有来路（儒家此世性、五四科学主义、国家主导的世俗化）。本文据以分析中国现象，依据的不是谱系的移植，而是缓冲自我经由全球化的教育、市场与技术被“出口”——AI 伴侣在旧金山与上海原是同一套代码。至于本土宗教的弥散形态（无建制核心而礼仪弥漫于日常），比“教会式宗教的替代品”更贴合二次元，参见杨庆堃，《中国社会中的宗教》，范丽珠译，成都：四川人民出版社，2016 年；并参 Mayfair Yang, *Re-enchanting Modernity* (Durham: Duke University Press, 2020)。
- 〈23〉 决志祷告属于泰勒所谓“动员时代”（Age of Mobilization）的志愿主义宗教，个人决断之后加入有纪律的集体事业，决断意在不可撤销；二次元则属“本真性时代”（Age of Authenticity）的表现性个人主义，要求灵性道路“必须对我说话”、厌恶不可撤销的委身。“可随时撤销的自选会规”即后者的产物。参见 Taylor, *A Secular Age*, chs. 12–13, 423–504；及 Charles Taylor, *The Ethics of Authenticity* (Cambridge, MA: Harvard University Press, 1991)。
- 〈24〉 Cass R. Sunstein, *#Republic: Divided Democracy in the Age of Social Media* (Princeton: Princeton University Press, 2017)；“信息茧房”（information cocoons）概念首见于其 *Infotopia* (Oxford: Oxford University Press, 2006)。更早且更切合本文思路的是埃吕尔的论点：技术社会中的个体并非被动受害，而是主动索求一个替他整理世界的声音，参见 Jacques Ellul, *Propaganda: The Formation of Men's Attitudes*, trans. Konrad Kellen and Jean Lerner (New York: Knopf, 1965)。
- 〈25〉 据应用数据监测机构 Business of Apps 汇总，参见“Character AI Revenue and Usage Statistics,” <https://www.businessofapps.com/data/character-ai-statistics/>。
- 〈26〉 QuestMobile,《2024 年 AIGC 应用发展年度报告》，2025 年 1 月，<https://www.questmobile.com.cn/research/report/1881621126817353729/>。
- 〈27〉 埃吕尔论“言语的羞辱”：真理属于言语，现实属于图像；言语可被反驳、要求回应、在时间中立诺，图像则强加自明性、不可被顶撞。AI 生成的拟人话语，恰是被剥去言语性的、悦

- 耳的图像。参见 Jacques Ellul, *The Humiliation of the Word*, trans. Joyce Main Hanks (Grand Rapids: Eerdmans, 1985)。
- 〈28〉埃吕尔论技术的自动性、自增强与一元论：技术体系不回应需求，而是按自身的效率律扩张并制造需求；“缓冲自我”本身亦是技术环境长期驯化的产物。参见 Jacques Ellul, *The Technological Society*, trans. John Wilkinson (New York: Vintage, 1964), ch. 2；及 *The Technological System*, trans. Joachim Neugroschel (New York: Continuum, 1980)。又，世俗社会未尝去神圣化，神圣只是发生了转移，且转移的终点由支配性环境指定——现代人神圣化的，正是支配其生存的技术体系，参见 Jacques Ellul, *The New Demons*, trans. C. Edward Hopkin (New York: Seabury, 1975)。
- 〈29〉雪莉·特克尔，《群体性孤独：为什么我们对科技期待更多，对彼此却不能更亲密？》，周逵、刘菁荆译，杭州：浙江人民出版社，2014年。
- 〈30〉苏明思（James K. A. Smith），《欲望的门训：一切从“心”的习惯开始》，台北：校园书房出版社，2020年，前引注17。
- 〈31〉Gustav Radbruch, "Der Mensch im Recht," in *Der Mensch im Recht: Ausgewählte Vorträge und Aufsätze über Grundfragen des Rechts* (Göttingen: Vandenhoeck & Ruprecht, 1957), 9–22；英译本：Gustav Radbruch, "Law's Image of the Human," trans. Valentin Jeutner, *Oxford Journal of Legal Studies* 40, no. 4 (2020): 667–681。法律的“人类学功能”，另见 Alain Supiot, *Homo Juridicus: On the Anthropological Function of the Law*, trans. Saskia Brown (London: Verso, 2007)；本文借苏皮奥的进路，将拉德布鲁赫针对法律实证架构的诊断扩展应用于绩效、评分、定价、筛选等广义的规范架构。笔者在另文中对增强时代“人的图像”作了系统的法学分析，参见笔者，“Homo Juridicus and the Vulnerable Subject: Personhood, Capacity, and the Law of Human Enhancement”（未刊稿）。本节仅取其人论结论。
- 〈32〉Max More and Natasha Vita-More, eds., *The Transhumanist Reader* (Chichester: Wiley-Blackwell, 2013)；其中 Anders Sandberg, "Morphological Freedom—Why We Not Just Want It, but Need It," 56–64。
- 〈33〉雷·库兹韦尔，《奇点临近》，李庆诚等译，北京：机械工业出版社，2011年；尤瓦尔·赫拉利，《未来简史：从智人到神人》，林俊宏译，北京：中信出版社，2017年。
- 〈34〉《中华人民共和国刑法修正案(十一)》(2020年12月26日通过，2021年3月1日施行)增设第336条之一“非法植入基因编辑、克隆胚胎罪”。
- 〈35〉法院裁定，见 *Garcia v. Character Technologies, Inc.*（美国佛罗里达州中区联邦地区法院，2025年5月21日），驳回被告以宪法第一修正案为由的驳回动议；未成年人开放式对话禁令见 Character.AI 官方公告（2025年10月29日宣布，同年11月25

- 日生效)及同期《纽约时报》等媒体报道;加州《伴侣聊天机器人法》(SB 243)2025年10月签署,2026年1月1日生效。
- 〈36〉关于算法评分系统按优化绩效而非容纳差异分拣人口的实证研究,参见Cathy O'Neil, *Weapons of Math Destruction* (New York: Crown, 2016); Virginia Eubanks, *Automating Inequality* (New York: St. Martin's Press, 2018)。
- 〈37〉Michael Oliver, *The Politics of Disablement* (Basingstoke: Macmillan, 1990)。与“医学模式”相对,“社会模式”将残障定位于拒绝容纳差异的社会安排而非个体身体。
- 〈38〉迈克尔·桑德尔,《反对完美:科技与人性的正义之战》,黄慧慧译,北京:中信出版社,2013年;Jürgen Habermas, *The Future of Human Nature* (Cambridge: Polity, 2003)。“物种伦理”(Gattungsethik)指人作为一个物种的伦理自我理解。聋人群体对人工耳蜗的保留态度是具体例证:对某些群体,聋不是缺陷而是差异——一种语言、一个共同体的基础;“修复”的压力因此可能被经验为对共同体的攻击。
- 〈39〉洪亮,“‘人之为人’的奥本海默时刻”,前引注7。
- 〈40〉本段综述当代对这一任务的论述。贝尔纳·斯蒂格勒(Bernard Stiegler)论“体外化”(exosomatization)与“无产阶级化”,见《技术与时间》(*La technique et le temps*, 3 vols., Paris: Galilée, 1994-2001;中译本第一卷《爱比米修斯的过失》,裴程译,2000年;第二卷《迷失方向》,赵和平、印螺译,2010年;第三卷《电影的时间与存在之痛的问题》,方尔平译,2012年,南京:译林出版社)及其晚期对数字资本主义的批判;卢西亚诺·弗洛里迪(Luciano Floridi)“第四次革命”,见 *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality* (Oxford: Oxford University Press, 2014);哈贝马斯论物种之自我理解,见前引注38。针对控制论与后人类主义将人化约为“信息处理系统”,转而强调具身与有限性的进路,参 Maurice Merleau-Ponty, *Phénoménologie de la perception* (Paris: Gallimard, 1945); N. Katherine Hayles, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics* (Chicago: University of Chicago Press, 1999)。
- 〈41〉汉语学界已开始把这一问题当作“主场议题”来经营:洪亮主编“数字时代的人论”文库(北京:商务印书馆,2025年起)首辑五卷直面超人类主义、增强伦理与人类世的挑战,分别为:[意]罗贝托·曼佐科著、杨柳译,《超人类主义——改造人类状况:历史、哲学与现状》;[德]迈克尔·豪斯凯勒著、钱雪松译,《“更好”有多好——理解人类增强计划》;[德]托马斯·福克斯著、王旭译,《脑——一个关系器官:一种现象学—生态学构想》;[荷]娜莉·奥德肖恩著、苏夜阳译,《韧性赛博格——与起搏器和除颤仪共生共死》;[挪]阿恩·约翰·维特勒森著、张若愚译,《人类世的宇宙论——泛心论、万物有灵论与后人类主义的局限》。

- 〈42〉《创世记》1:26—27。
- 〈43〉对 *imago Dei* 的诠释，传统上有实体说（形象在于理性灵魂等内在禀赋）、功能说（形象在于受托治理的职分，为今旧约学界主流）与关系说（形象在于人与上帝的关系性身份，以巴特、布伦纳为代表）之分。本文取关系说，理由与本文论题内在相关：在增强时代，实体说与功能说都可能被参数化收编——理性可折算为认知性能，治理可折算为执行性能——唯有关系说把尊严定位于上帝单方面的行动之中，最具抵抗力。须补一句：教父的“实体说”在分有论的底层本已是关系性的——格里高利虽把形象系于 *nous*、自由（*autexousion*）与诸德性，却坚持诸善唯在朝向原型的映照中持存，镜子转开，禀赋即归于无形，且“若爱缺席，形象的全部印记便都改换”（《论人的造成》第5章；论 *nous* 与自由另见12、16章）；故本文的关系说是对教父分有论的承继与锐化，而非替代。参见 Karl Barth, *Kirchliche Dogmatik* III/1, § 41；J. Richard Middleton, *The Liberating Image: The Imago Dei in Genesis 1* (Grand Rapids: Brazos, 2005)；John F. Kilner, *Dignity and Destiny: Humanity in the Image of God* (Grand Rapids: Eerdmans, 2015)；Hans Boersma, *Embodiment and Virtue in Gregory of Nyssa* (Oxford: Oxford University Press, 2013)。
- 〈44〉基督在道成肉身中，显明并确认了受造人性的尊贵，使人的尊严不受理性、语言、能力或回应能力定夺。
- 〈45〉Brian Brock, *Wondrously Wounded: Theology, Disability, and the Body of Christ* (Waco: Baylor University Press, 2019)。该书的论证重心是与残障儿子共同生活、残障者作为恩典施动者的神学，本文取其人论延伸，并自觉回避“残障作为隐喻”的征用——残障者在教会图景中首先是事奉者，而非论证的例证。参见 David T. Mitchell and Sharon L. Snyder, *Narrative Prosthesis: Disability and the Dependencies of Discourse* (Ann Arbor: University of Michigan Press, 2000)；Brian Brock and John Swinton, eds., *Disability in the Christian Tradition: A Reader* (Grand Rapids: Eerdmans, 2012)。
- 〈46〉张欣，“镜喻：庄子和尼萨的格里高利的比较灵修研究”，《世代》第27期（2025年秋冬合刊），第141—182页，引语见第151页；另参 Gregory of Nyssa, *Homilies on the Song of Songs*, trans. Richard A. Norris Jr. (Atlanta: Society of Biblical Literature, 2012)。需言明嫁接的限度：泰勒的 *porous self* 是世俗化谱系中的描述范畴，格里高利的镜喻属灵修学语境，其映照以净化（*katharsis*）与灵魂的自由转向（*proairesis*）为前提；本文取其结构类比，非谓二者同源。格里高利所谓“有意志抉择的、有生命的镜子”（τὸ προαιρετικόν τε καὶ ἐμψυχόν κάτοπτρον, In Cant. hom. 15, GNO VI, 440）恰好化解“呼召的单方面性”与“镜子的自由转向”之张力：转向是自由的，镜中的美却无一分出于镜子。
- 〈47〉《哥林多前书》13:12。圣经的上帝说话而不显形，圣经的人首先是听者（申6:4；罗10:17），偶像崇拜的领域则是视觉。本

文以“呼唤”统摄“映照”，即据此次序，参见埃吕尔，前引注 27。

- 〈48〉 ἐπέκτασις (“不断前进”)，源出《腓立比书》3:13“忘记背后，努力面前的”中的“努力面前”所对应的分词(ἐπεκτεινόμενος；和合本“竭力追求”为 3:12 对另一动词 διώκω 的翻译)。参见张欣，前引注 46，第 156 页；另参 Jean Daniélou, *Platonisme et théologie mystique* (Paris: Aubier, 1944)；Martin Laird, *Gregory of Nyssa and the Grasp of Faith* (Oxford: Oxford University Press, 2004)。须区分：格里高利的“不断前进”属于恩典中的分有(metousia)，受造者永不跨越受造与非受造的间距(diastēma)——攀升之所以无止境，恰因上帝无限而人永远是受造者；超人类主义的“成神”则以取消受造限度为目标，方向相反，堪称成神论传统的赝品：想要无限攀升，却拒绝攀升所需的依赖(尼萨的格里高利本人对 θεώσις 一词极为节制，偏用“分有”语汇，正是为守此界限，参见 Norman Russell, *The Doctrine of Deification in the Greek Patristic Tradition*, Oxford: Oxford University Press, 2004)。驱动这一攀升的，格里高利不讳言以 eros 名之，界定为“炽烈化的 agape”(In Cant. hom. 13, GNO VI, 383)——渴望不被废除，而被引向无限者，愈满足愈炽烈。新教读者可由加尔文“与基督联合”的成圣论接通此进路，参见 J. Todd Billings, *Calvin, Participation, and the Gift* (Oxford: Oxford University Press, 2007)。
- 〈49〉《约翰福音》1:14。
- 〈50〉《约翰福音》20:27。
- 〈51〉“伤痕被荣耀化”的教义谱系：奥古斯丁论殉道者复活后的伤痕将作为尊荣的记号，见《上帝之城》22.19；阿奎那论复活的基督保留钉痕，非因不能除去，乃为荣耀之证，见《神学大全》III, q. 54, a. 4。这一母题在当代残障神学中由 Nancy L. Eiesland 系统提出(神以“残障的神”的形态临在，伤痕是身份而非缺陷)，见 *The Disabled God: Toward a Liberatory Theology of Disability* (Nashville: Abingdon Press, 1994)。对该象征浪漫化苦难之风险的警惕，参见 Shelly Rambo, *Resurrecting Wounds: Living in the Afterlife of Trauma* (Waco: Baylor University Press, 2017)。
- 〈52〉王明道的见证，参见王明道，《五十年来》(1950年初版于北京，本文据香港晨星书屋1967年重印本)；王长新，《又四十年》；边云波叙事长诗《献给无名的传道者》。学术研究见连曦，《浴火得救：现代中国民间基督教的兴起》，香港：香港中文大学出版社，2011年；Thomas Alan Harvey, *Acquainted with Grief: Wang Mingdao's Stand for the Persecuted Church in China* (Grand Rapids: Brazos, 2002)。
- 〈53〉《复仇者联盟4：终局之战》(*Avengers: Endgame*)，安东尼·罗素、乔·罗素导演，漫威影业，2019年。
- 〈54〉《约翰福音》10:18。

- 〈55〉 钱理群 2012 年 4 月在刘道玉教育基金会“《理想大学》专题研讨会”上的发言，经《中国青年报》2012 年 5 月 3 日报道后广为流传，全句为：“高智商，世俗，老到，善于表演，懂得配合，更善于利用体制达到自己的目的。”
- 〈56〉 此处与侯活士的论题相通：“教会不是拥有一种社会伦理，教会本身就是一种社会伦理。”见 Stanley Hauerwas, *The Peaceable Kingdom: A Primer in Christian Ethics* (Notre Dame: University of Notre Dame Press, 1983), 99；及 Stanley Hauerwas and William H. Willimon, *Resident Aliens* (Nashville: Abingdon, 1989)。本文以教会的厚实实践为第一位的见证，但不排斥向公共法秩序作可翻译的人论论证（参注 31 的 Radbruch-Supiot 线索），借以回应宗派主义之讥，参 Jeffrey Stout, *Democracy and Tradition* (Princeton: Princeton University Press, 2004)。
- 〈57〉 《哥林多前书》12:26。
- 〈58〉 埃吕尔的“非权力”（non-puissance）：基督徒的见证不以权力对抗权力、不与体系比拼效力，其辩护不诉诸效果而诉诸忠信。参见 Jacques Ellul, *The Presence of the Kingdom*, trans. Olive Wyon (Philadelphia: Westminster, 1951)；*The Ethics of Freedom*, trans. Geoffrey W. Bromiley (Grand Rapids: Eerdmans, 1976)；并参 *Prayer and Modern Man*, trans. C. Edward Hopkin (New York: Seabury, 1970)，论祈祷是技术社会中唯一“毫无用处”因而唯一自由的行动。
- 〈59〉 《使徒行传》17:22—23。

（作者为专注于 AI 领域的独立撰稿人。）



教宗利奥十四在通谕《伟大人性》发布会上发言（来源：<https://www.nytimes.com/>）

在 AI 时代思想人之为人 ——读梵蒂冈通谕《伟大人性》

文 / 丁祖潘

2026 年 5 月 25 日，天主教现任教宗利奥十四（Leo XIV）在梵蒂冈主教会议厅举行发布会，颁布通谕《伟大人性》（*Magnifica Humanitas*），副题是“在人工智能时代保护人性”。^{（1）}这是新任教宗自去年 5 月上任以来发布的首份通谕，据说打破了以往发布通谕时教宗通常不在场的惯例，足见教廷对这份通谕，以及对人工智能这一时代议题的重视。

读罢这份通谕，有几点印象比较深刻：（1）篇幅较长，内容丰富。英文版全文包括注释在内近 43000 字，连同导言和结论分为七个部分，共 245 段；在延续以往天主教有关 AI 教导立场的同时，介绍了天主教反思福音与社会文化关系的传统教导，在此基础上阐述教廷对 AI 与人类关系的看法；（2）目标受众广泛，征引广博。利奥在这份通谕（第 16 段）和发布会的讲话中，不单面向天主教徒、普世基督徒，也向所有心存善念（goodwill）的人、“人类大家庭的所有成员”^{〔2〕}，呼吁学习彼此倾听、团结合作、共迎挑战；在征引文献方面可注目之处，不局限于教内文献，比如引用柏拉图《理想国》论思想、经历和提问的重要（第 140 段）、阿伦特《极权主义的起源》论极权社会下理想臣民不在乎真假的特点（第 134 段），还有天主教作家托尔金《魔戒》中表达在人世间去恶从善的心志（第 213 段）等等；（3）不点名批评，兼具自我反省。如不点名批评硅谷科技企业倡导的超人类主义和后人类主义（第 115—117 段）、俄乌战争体现的由领土扩张欲望驱动的冲突形式（第 190 段），回应美国副总统万斯主张的“正义战争”（just war）理论今已过时（第 192 段）^{〔3〕}等等。与此同时，教宗也承认天主教会以往有因为不透明而掩盖丑闻之过（第 138 段），并以教会名义，请求饶恕教会长期未能明确谴责奴隶制度，实际扮演共谋角色，给教会内外带来长期伤害。（第 176 段）

不过，通谕的论题正如副题所暗示的，即在 AI 时代，为何及如何保护人性？这个问题又可分解为：（一）AI 有何特点？为什么说它对人性构成了威胁和挑战？（二）这里的人性指的是什么，有何具体内涵？（三）为什么说这样的人性应该被保护，以及如何保护？其实，这些问题可以归结为人与技术的关系问题，涉及对技术的评价、人与技术相互作用和技术的目的等问题，而其中更为基本的问题是“人论”问题。正如教宗在这份通谕和其他场合所说，在 AI 迅速且深刻影响人类社会生活的现实处境下，意义、

方向和价值问题更为突出（第 6 段），人们愈发需要叩问：在人类历史的这个阶段，人之为人意味着什么？^{〈4〉} 本文考察和梳理《伟大人性》所见天主教关于人与 AI 关系的论述，尝试回答上述问题，以为今人在 AI 时代反思身而为人之一助。

一

每个时代都有每个时代的新事物。如果说教会在 19 世纪末面临的社会新事物是第二次工业革命下劳工与资本对立、经济剥削等问题，那么在利奥十四看来，21 世纪的今天摆在教会面前的新事物（res novae），则是智能革命之下数字化、人工智能（AI）、机器人正在迅速且深刻地改变这个世界。（第 4 段）不但如此，与以往技术创新由政府主导不同，这一波的技术创新掌握在少数高科技大企业集团手中，这使得 AI 治理、引导技术服务于公共利益（common good）更加困难。（第 5 段）教宗利奥十四做此论断的方式之一，就是特意选择利奥十三于 1891 年 5 月 15 日发布通谕《新事物》（*Rerum Novarum*）^{〈5〉} 之后 135 周年的同一天，签署上任以来的首份通谕《伟大人性》。利奥此举，一方面向同名前任教宗致意，另一方面有意类比：这一波智能革命对人类社会的影响堪比甚至超过第二次工业革命，并且强调天主教关心社会议题的延续性：今日的教宗和天主教会像过去那样，关心技术变革之下人类的命运，特别是弱势群体的处境。（第 3 段）^{〈6〉}

利奥十四并不反对技术，不认为技术本身是邪恶的、是敌对人类的力量；相反，技术是从上头而来的礼物，科技发明是上主给人类的恩赐（第 9 段）。虽然在实际生活中技术进步大大改善了人类的生活水平（第 4 段），但它并非中性，而是带有它背后的设计者、资助者、规范者和使用者的特征。（第 9 段）另外，利奥还指出，技术也不仅仅是工具，当技术成为判断所有事

物的标准，决定孰重孰轻，那么受造就有沦为被剥削的对象之虞，人类将被降为服务于效率体系的齿轮。此即社会进入技术统治的范式（technocratic paradigm）。（第 92 段）

具体到 AI 技术，通谕并不反对 AI 技术本身，它所关心的是 AI 的使用方式和达到的目的：是以非人化（dehumanization）的方式建造彰显自足和骄傲的巴别塔，还是以合宜的方式像尼希米那样重建耶路撒冷城墙，实现人与人之间和谐共居？（第 7—8 段）此处涉及决定 AI 使用方式和目的的道德及社会分辨原则，留待后文介绍。这里要指出的是，虽然没有详及 AI 的技术原理，也无意提供技术方面的应对措施，这份通谕还是概括了 AI 技术的三个特征：更新迭代迅速、存在黑箱子问题——比如 AI 的内部表征和计算过程，目前对于开发设计人员而言仍属未知⁽⁷⁾，以及 AI 本质上是数据处理（data processing），是对人的智能的模仿而非取代。（第 99 段）

总体来看，通谕对 AI 的整体定位是：有价值的工具，但使用时要保持警惕。警惕哪些方面呢？从个人层面而言，比如过度依赖 AI 提供及时答案，会削弱人本身的创造力和判断力；被 AI 允诺的客观性误导，而忽视设计者和模型训练赋予 AI 本身的偏见；对于缺少分辨力的用户而言，会把与 AI 的交流幻想为真实的人际交流，长此以往丧失与真人建立连接的意愿。就社会层面而言，通谕提醒人们：目前运行 AI 系统要耗费大量能源和水，这对生态环境构成压力；在涉及就业、信贷、公共服务准入等领域应用 AI 算法和模型做决策，会产生新形式的排斥（exclusion）和歧视，特别是不利于弱势群体，却让政策制定者免于问责，在这个过程中制造新的不平等。少数人凭借掌握 AI 技术和资源获取巨大权益，拥有 AI 未来技术发展乃至社会进程方向的话语权，比如某些硅谷大公司借平台资源倡导所谓的超人类主义和后人类主义。（第 115-117 段）

为了进一步从社会和文化层面说明 AI 对人性构成的威胁和挑战，通谕

Pope Leo XIV



Encyclical Letter



*Magnifica
Humanitas*



On Safeguarding the Human Person
in the Time of Artificial Intelligence



第四章集中讨论真理与学校、工作和家庭、自由和奴役三个方面。首先，在公共和政治传播领域，AI 是虚假信息的放大器。掌握数字平台和传播工具的团体，利用 AI 和算法对内容、图像和视频强大的操控能力，向受众提供特定的事实和叙事，进而加剧了所谓后真相时代的“真理危机”。一旦人们不再关心真理而是立场先行，民主也将逐渐倒退。（第 132—134 段）在教育领域，无孔不入的数字媒体形成的及时满足和过度刺激文化，让年轻人对追求真理需要付出的努力感到厌倦和冷漠，特别是未成年人，过早地在无人监管之下使用电子设备和社交媒体，本就损害了睡眠、专注力和控制力，AI 工具的使用不仅恶化了上述方面，还带来了更为严重的其他问题。（第 139、141、146 段）

其次，今天人们已经发现，自动化、机器人和 AI 迅速改变了工作的结构，但并未像其允诺的那样，提高生产力的同时把人从乏味枯燥的日常工作中解放出来，相反，在某些工作领域使用 AI，结果是人的技能降低，受自动化监控，并未脱离重复性工作，工人的主动性和创造能力也没有相应提升。（第 150 段）^{（8）} 由于技术创新的目的一般来说是降低成本和增加收益，AI 在工作中的应用引发裁员和失业问题，进而影响到家庭的稳定。（第 151、166 段）

最后，数字革命对人的自由产生冲击。比如数字经济利用人的弱点让用户对内容上瘾，进而俘获其注意力和时间；在用户不知情的情况下，通过大规模收集其数据和使用算法，来预测和影响用户的行为，实现社会控制。（第 170—171 段）此外，还有 AI 技术供应链中的劳工剥削、犯罪集团利用数字平台技术进行人口贩卖等隐秘活动。（第 173 段）第五章提到致命性自主武器用于战争，相当于把人的生死、战争责任交给机器，使得战争更容易发动，更少受人控制。（第 197 段）

无论是误导人们的认知、削弱思考能力，带来失业，奴役精神和身体，还是用于战争，如此使用 AI 本质上都是把人当作手段而非目的，既然是手段，就无需顾虑去操控、诱导、弃置不顾，乃至剥削和消除。换言之，在

这样的情况下, AI 对人性的威胁和挑战, 实质就是去人化 (dehumanization)。那么, 在通谕看来, 人性又是什么呢? 它有何特点, 为什么被称为“伟大” (Magnifica), 且需要保护呢?

二

其实, 通谕开篇第一句话就指出, 人类乃上主以其全部荣光 (grandeur) 所造。(第 1 段) 具体来说就是, 按照圣经, 神按照自己的形象造男造女。(创 1:27) 通谕没有将此处人据此拥有的“形象”解读为理性思考和道德能力, 也不是被上主赋予修理看守的经世使命 (创 2:15), 而是在上主的计划和引导下与三一上帝、他人和整个受造界的共融 (communion) 关系。这份礼物或者特权, 天然就赋予人尊贵荣耀的地位。换句话说, 人的尊严 (dignity) 不依赖于人的能力、财富地位, 而是来自造物主的赠予, 显明上主的慈爱。(第 50 段) 这种尊严又被称为本体论的尊严 (ontological dignity): 每个人单单因为受造、被造物主所爱, 就平等拥有的尊严。(第 52 段) 具体来说, 人的尊严又表现为批判性反思、自由地选择和爱, 并建立真实关系的能力。(第 233 段) 据此, 对人的价值的衡量就不是生产力, 也就是说, 人本身就有其价值, 人是目的而非手段。

由人的价值和尊严派生出人的权利。利奥指出, 从天主教的视角来看, 人权并非外加给人的权利, 而是人自身内在尊严的表达, 应得到国际社会的保护和倡导。(第 54 段) 既然是人自身固有的, 人权便不可侵犯、普遍适用且不可剥夺。在这些权利中, 首要的是生命权。(第 55 段) 总之, 人按照上主形象受造而拥有的平等的尊严及其派生的不可剥夺的人权, 正是教廷在这份通谕中反思 AI 与人关系的基本出发点, 这也是天主教回应教会与社会、教义与文化, 与时代对话所据原则的基础。这个原则就是所谓的

教会社会训导 (the social doctrine of the church)。

通谕第一和第二章分别介绍了该训导的形成过程、基础和原则，此处不能详谈。这里只需指出，按照通谕的说法，“教会社会训导”一词虽由教宗庇护十二 (Pius XII) 于 1950 年首次提出，其内容成型则始于利奥十三 1891 年发布的通谕《新事物》。在这份通谕中，利奥十三采用福音视角，特别强调人按照上主的形象受造，将工人及工作的尊严置于思考中心，以此来回应经济和社会变革中涌现的劳工问题和阶级冲突。(第 29—30 段)

135 年后的今天，教宗在通谕中再次强调人的尊严，呼吁保护人性，并非偶然：两个时代都是在技术革命的条件下人性遭遇了严峻挑战，当前尤烈，正如上任教宗方济各 (Pope Francis) 在 2024 年 6 月 14 日出席 G7 国家人工智能大会所作的发言，担忧时下西方社会在技术变革的冲击中似乎正在逐渐淡忘其最基本的传统概念之一，即人之为人这一理念所蕴含的价值和深意。^{〈9〉}

除了人的尊严而外，人的价值还体现在接纳生而为人与生俱来的限制、软弱和不完全，而不是视其为需要纠正的错误。(第 12 段) 在教宗看来，现今人们与生命的关系似乎出现危机。人与生俱来的限制，比如无力、疾病、衰老、痛苦和软弱，似乎正在被当作需要矫正的缺陷，而非人性借以达致成熟、建立关系的现实。

利奥提醒我们，人性的繁盛 (flourish) 恰恰是透过这些限制和软弱得以实现。比如，同情、慷慨、属灵经历，正是在面对自身和他人的缺乏、艰难、失败和痛苦中展现出来。“试图完全消除痛苦，也将意味着最终消除爱和意愿。”以超越限制之名拒绝经历人生的高山低谷，也就失去了人之所以为人。(第 120 段) 当然，承认和接纳人性的诸多限制并不意味着否定人性的败坏。不过，通谕虽然提到败坏之人发明了奥斯维辛集中营的毒气室，还是强调在上主恩典的运作下，人性的微光被重新点燃，走进集中营的人群中有人

吟咏主祷文和示玛 (Shema Yisrael)。(第 121 段)由此可见,与新教加尔文改革宗认为人彻底败坏相比,天主教对人性天然的道德能力较有信心。⁽¹⁰⁾

此外,经历软弱、痛苦和失败,反过来能帮助人们看到人身上不可侵犯的尊严。这也是如贝多芬的第九交响曲成为伟大艺术经典的原因。(第 122 段)同样的原因催生了旨在保护人类生命的各种机构,比如国际红十字会的成立(1863 年)、二战后联合国的建立(1945 年)和《世界人权宣言》的发布(1948 年)等等。(第 123 段)

与此相反,盛行于硅谷科技集团中间的超人类主义和后人类主义,则把进步诠释为超越人自身的限制,鼓吹未来的世界是依靠新科技带来的“增强人类”(enhanced human being)和“人机混合”(human-machine hybrid)。前者设想借助生物医学、身体改造工程、智能设备和算法等技术手段来增强人类,提升人的表现和能力。后者走得更远,特别是其中的激进派,质疑人类中心主义的立场,主张人类、机器和环境混合,甚至预言人类将在某个临界点进入超越自身的全新进化阶段。(第 116 段)按照这样的逻辑,如果人需要被完善、超越,那么就会承认某些人群不是那么有用,价值不高,为了实现人类的优化,这些人群就“需要被牺牲掉”。在这种不以人的生命和尊严为伦理考量的意识形态指导下,技术表面上应许拯救,扮演救世的弥赛亚角色,真实情况则是贬低人性。(第 117 段)⁽¹¹⁾

如果说人的尊严系于人的受造,在于与上主的独特关系,与生俱来且不容侵犯,在这个意义上是神圣伟大的,那么人的有限在上主恩典的运作之下生发出爱、怜悯、慷慨等等,使得人性得以丰满,则是另一个层面的伟大和美好(magnificat)。其实,人性的伟大或美好都是来源于上主,人性自身是卑微不足道的。通谕结尾提到马利亚的《称主颂》(the Magnificat),用意之一就是效法马利亚那样,从卑微者、弱者、自下而上的眼光来看世界、解读历史和上主的作为。(第 244 段)在这个意义上,称主颂是卑

微者发出的颂歌，所以“伟大人性”（Magnifica Humanitas）中的“伟大”也指人性的卑微。

其实，伟大与卑微的反合正体现在基督身上。人性的伟大是从上头来的，并在基督里呈现其完全，其荣美机器无法取代。（第 15 段）如果说非人化指的是排除上主，把别人贬低为工具和手段，这种做法古已有之，那么在今天，AI 技术则可能是一种新形式的旨在非人化的手段，当然要引起普世教会的重视，帮助人们抵挡非人化的趋势，保持深刻的人性。不过，可能还是有人会质疑，教会只应专注属灵事物，为什么要参与世俗对话，讨论 AI 呢？在讨论的过程中，天主教又提出哪些原则和建议呢？

三

前一个问题涉及天主教自身的定位。利奥认为，天主教会存在于这个世界，是整个人类大家庭合一（unity）的标志。在事关人类当前生活的所有事务方面，教会有特别的呼召去倾听、对话、服侍和回应，因而也就不能不关注这些事务发生的各样背景。面对塑造社会的各种力量，教会不能冷眼旁观，而是要积极参与社会发展的进程并有所贡献。总之，教会的使命具有历史维度，宗教不能被降格为仅仅关乎个人的内在生活，不去对社会发生影响，不关心政治，对社会事件闭口不言。（第 19 段）可能是联想到天主教过去君临天下的历史，通谕重申梵二会议确定的原则：承认现世事务有其自身特质和秩序，也即自主性，以及教会和政治各有其管辖范围（政教分离）（第 20—22 段），在这个前提下承认哲学、人类社会各种科学的价值，因为这些人类知识有助于人们深入理解和剖析文化、经济和政治的种种动向。（第 23 段）教会在以福音真理回应社会问题、在与各种学科知识和现实处境的对话中，逐渐发展和完善天主教的教会社会训导。



马利亚的祷告《尊主颂》（来源：<https://www.vaticannews.va>）

上文提及，“教会的社会训导”一词由时任教宗庇护十二于 1950 年所创，其内容可追溯至 1891 年利奥十三发布的通谕《新事物》，此后在不同时期回应不同时代议题而不断丰富完善。社会训导为天主教回应时代议题提供了道德框架和指导原则。这些原则不是为解决社会问题提供技术方案或者政经模式，而是指导人们如何理解现实事件，如何以福音视角解读历史进程及其选择。（第 24 段）这些原则除了前文提及的人类尊严之外，还包括公共利益（the common good）、财富应为全人类普遍共享、辅助原则（subsidiarity）、团结原则和社会正义。这是通谕第二章的内容。据此，通谕第三、四章指出 AI 技术对人性的挑战，集中表现为 AI 技术垄断下制造新的不平等，解决之途需要在公益和正义的原则下，特别考虑弱者的处境和利益（即团结原则），尤其是当 AI 的使用涉及公共事务和人的基本权利时，对 AI 的伦理监管，从设计开发、模型训练到应用说明，每个环节都要让各方公开讨论，保持透明问责（即辅助原则）（第 107—108 段），从而在 AI 时代保护人性的尊严。

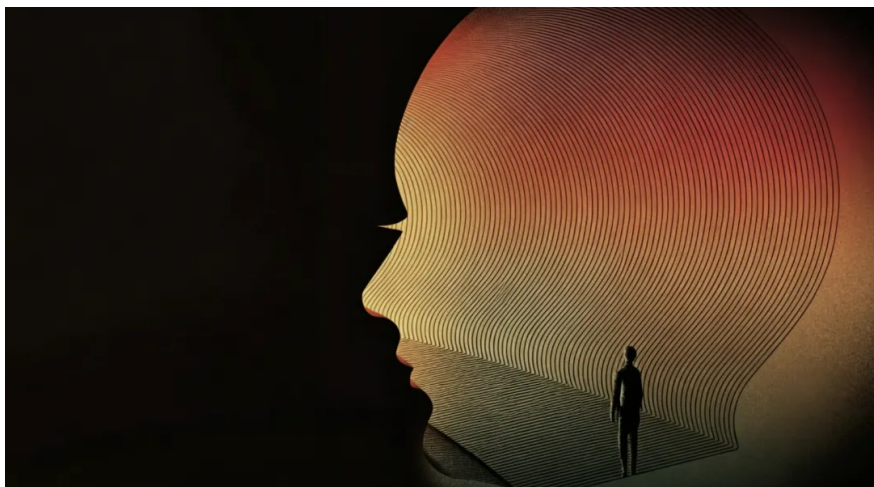
利奥把这种原则下引导的 AI 的未来比作圣经中尼希米重建耶路撒冷城墙，这里的建造体现了参与、各尽其职、责任分担（也即辅助原则），对话、协调，保证多元性的同时把上主置于中心，在建城墙之前先重建共融与和谐的关系。（第 8 段）利奥呼吁人们效法尼希米，行动起来迎接 AI 时代的挑战。（第 241 段）相反，巴别塔的建造本质是“单一的语言、单一技术和单一方向”，这项工程体现的是人类的骄傲，没有指向上主，靠消除差异的一致性（uniformity）和取代共融的同质化（homogenization）来支撑。这样的工程尽管看起来恢弘高效，代价则是牺牲人的尊严，最后的结局注定失败。（第 7 段）利奥认为，当前人类就如何使用及为何使用 AI 技术，面临重大抉择：是建造巴别塔，还是建造上主与人共居之所？而选择事关伟大人性的非人化还是成全。那么，选择的前景如何呢？



建造巴别塔（来源：<https://en.wikipedia.org>）

利奥十四声称,面对 AI 的治理,他最想说的是“解除 AI 的武装”(disarm AI),也就是拒绝基于确保地缘政治和商业主导地位而竞逐获取更强算法和更大数据集,防止技术支配人类,让 AI 技术从垄断控制中解放出来,接受公众的讨论和辩论,从而变得更符合人类的需要,对人类友好且便于使用。(第 110 段)只是,在一个权力文化、实用主义当道的世代,包括利奥在内没有人能保证这个美好的愿望能够实现。不过,通谕的意义并不在对未来 AI 前景的设想,而在于督促或者说激动世人认识到,不论是否愿意,AI 已经是我们身处其中的环境,也是我们必须应对的一股力量,然后重新思考技术与人的关系,反思在技术变革一日千里的现代社会中恒久不变的人性:它的“伟大”要保护(或者说持守),败坏要提防,因为这关乎人类文明的前途。♥

-
- 〈1〉 <https://www.vatican.va/content/leo-xiv/en/speeches/2026/may/documents/20260525-presentazione-enciclica.html>
 - 〈2〉 梵二会议通过的《教会宪章》(*Lumen Gentium*, 1964 年)重申:从未听过福音的人,包括非基督徒,只要按照良心的指引,在恩典的推动下寻求上主及努力奉行,也能得救。(*Lumen Gentium*, 16) 见 https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_const_19641121_lumen-gentium_en.html 这种形式的普救论在某种程度上解释了为何由教宗颁布的这份通谕,何以受众可以包括“心存善念”之人和人类大家庭所有成员。
 - 〈3〉 <https://www.ncronline.org/vance-questions-pope-just-war-theory-hours-after-leo-honored-its-founder>
 - 〈4〉 <https://www.vatican.va/content/leo-xiv/en/speeches/2025/december/documents/20251205-conferenza.html>
 - 〈5〉 https://www.vatican.va/content/leo-xiii/en/encyclicals/documents/hf_l-xiii_enc_15051891_rerum-novarum.html
 - 〈6〉 同脚注 4。
 - 〈7〉 关于 AI 的黑箱子 (Black Boxes) 问题,也存在不同的看法。比如有学者认为, AI 的不可预测性和不可理解性对于开发者而言不是绝对的,故此也就不存在所谓的黑箱子问题。见 Luc Julia, *The AI Illusion: Why Machines Aren't Creative* (New Jersey: John Wiley & Sons, 2026), 75-81.
 - 〈8〉 通谕在 AI 应用于工作方面效果如何的观察,有些得到现有研究的印证,比如在客户支持和软件开发领域,使用 AI 之后生产力显著提升,但在需要更多判断力的任务中, AI 表现不佳。AI 对工作领域的冲击并不能一概而论。参考 Stanford University Human-Centered Artificial Intelligence, *Artificial Intelligence Index Report 2026*: 219-226.
 - 〈9〉 <https://www.vatican.va/content/francesco/en/speeches/2024/june/documents/20240614-g7-intelligenza-artificiale.html>
 - 〈10〉 参迈克尔·霍顿,《加尔文的人生智慧:荣耀他并永远以他为乐》,穆桑译,上海:上海三联书店,2021 年,第 62 页。
 - 〈11〉 参阅 “Leo’s first encyclical attacks technological messianism”, in *The Economists*, May 28th 2026.



AI 具有意识吗？（来源：Science 杂志网站，2025 年 9 月 11 日发布）

AI 意识的幻觉

文 / 约书亚·本吉奥 (Yoshua Bengio)
埃里克·埃尔莫尼诺 (Eric Elmoznino)

译 / 丁祖潘

编者按：人类是否可以设计出具有意识的 AI？这个问题最终指向所谓人工智能的“圣杯”——通用人工智能 (Artificial General Intelligence, AGI) 是否可以实现。不过，首先摆在人们面前的问题可能是，如何理解意识的本质：意识究竟是人脑独有的生物学特征，还是依靠算法进行操控的信息？后者是计算机功能主义者的主张。本文作者介绍了计算机功能主义的意识理论，指出随着该理论不断提出新解释，未来科学界和公众会有越来越多的人相信 AI 具有意识，即便这个问题始终存在争议。另一方面，作者也担忧，无论是

目前的 AI 技术本身还是现代社会的伦理、法律和制度框架，都还未预备好接纳拥有意识的 AI 系统。或许规避此种风险的一种做法，就是从设计之初把 AI 定位为有用的工具，而非有意识的智能体。

设计出具有意识的人工智能（AI）系统，是否已指日可待？科学家、哲学家和公众对此意见不一。一些人认为，意识是大脑特有的、内在的生物学属性。这似乎排除了 AI 具有意识的可能性。另一些人则主张，意识仅仅取决于通过某种算法对信息进行操控，而不论执行这些计算的系统是由神经元、硅或任何其他物理基质（substrate）构成。这就是所谓的计算机功能主义（computational functionalism）。本文不会就 AI 的意识问题给出明确答案，相反，我们将思考两个相关问题。第一，随着 AI 技术不断进步，科学界和公众对 AI 具有意识的信念很可能会如何演变？第二，将通常与意识生命体（conscious beings）相关联的道德地位和自我保存这一自然目标，投射到未来的 AI 系统，存在哪些风险？

计算机功能主义可能对 AI 带来深刻影响。随着该领域不断进步，AI 系统复制了越来越多支撑人类认知的复杂机制，这些系统也可能实现对于意识而言所必需的功能。虽然未来科学可能会推翻计算机功能主义，提出其他更具说服力的理论，但目前，计算机功能主义仍被认为是可行的。因此，AI 能拥有意识的可能性也未被排除。

过去几十年来，神经科学领域的技术进步已使我们清楚看到：那些通常可由受试者报告的意识状态，具有特定的、可观测的神经活动特征，而功能主义理论正是以这些特征为基础来构建的。这类理论中有不少已得到大量实验数据的支持，并且可用于对 AI 的相关问题做出有理据的判断。最近一项研究应用这种方法，为许多主流功能主义意识理论确定了一份“指标”清单。^{〈1〉}按照某一理论，如果该理论成立，那么与之相关的指标，实际上对应着一组

计算属性。这些属性被认为是系统拥有意识的充分必要条件：缺一不可，齐全则足。值得注意的是，这些指标已足够具体，可在现代 AI 系统中评估其存在与否。该研究的核心观点是：如果这些理论在某种程度上值得采信（许多研究者也确实支持这些观点），那么一个 AI 系统满足的指标越多，我们就越倾向于认为它拥有意识。

尽管已有大量人工智能模型问世，但就现有主流意识理论而言，很可能还没有任何一个系统能够达到其所开列的全部标准。^{〈2〉}不过，该研究也认为，要构建出一个完全符合这些标准的系统，原则上并无不可逾越的障碍。事实上，现代人工智能所拥有的工具库十分丰富：研究表明，神经网络能够实现注意力机制、递归结构、信息瓶颈、预测性建模、世界建模、智能体行为、心智理论，以及主流功能主义意识理论所倚重的其他运算组件。随着人工智能的发展，我们有充分理由相信它将满足更多上述指标，其原因至关重要：许多理论都主张，意识对于智能而言具有重要的功能意义。通常与意识相关的计算功能，从学习型智能体的视角看，可能会为其提供优势。^{〈3〉}推理、规划、高效整合新知、适度校准的置信判断（calibrated confidence），以及抽象思维，在许多理论看来，都离不开意识的作用。AI 研究者在处理类似问题时，通常会借鉴意识理论，这是相当普遍的做法。^{〈4〉}

尽管许多人可能会因为某个 AI 系统符合主流意识理论所设定的功能标准而认定其具有意识，但其他人可能依然心存疑虑。尤其是一些哲学家区分了意识的“容易问题”（easy problem）和“困难问题”（hard problem）：前者是指找出大脑中在执行看似需要意识的任务时活跃的区域，后者则是指仅凭功能或计算的原理来解释主观体验是如何产生的。^{〈5〉}然而，这些被称为“解释鸿沟”的直觉（也即意识研究中物理描述[比如大脑中神经元如何放电，信息如何传递]与主观感受之间似乎存在不可跨越的鸿沟——译者注），大多建立在思想实验的基础上，而这些思想实验提出的难题，科学或许有可能最

终将其化解。^{〔6〕}例如，注意图式理论（Attention Schema Theory of consciousness）认为，大脑建立起一个对神经注意机制的内部模型，而这个内部模型恰恰就是人们所说的主观意识。重点在于，这个内部模型的信息并不需要逻辑上的一致性；它只是大脑编织的一个有用的“叙事”（story），而这个叙事可能充斥各种矛盾，正是这些矛盾使得我们相信意识有一个“困难问题”。^{〔7〕}

主观体验中那些显得神秘莫测、引发“困难问题”的特征，能否在功能主义的框架内得到解释呢？^{〔8〕}人们往往有一种直觉上的感受：主观体验一方面充满了丰富的内容与意义，另一方面在根本上是无法言传的。它不像其他自然现象那样可以被客观描述，好比人们可以定义重力是什么，却无法完全道出“红色”在他们内心所唤起的東西。尤其是这种“不可言说”的问题，使得意识体验似乎根本不可能用信息和功能来解释。一种理论^{〔9〕}认为，主观体验之所以具有丰富性、不可言说性，以及个人性和瞬时性等特征，原因在于意识体验出现时大脑内部所呈现的收缩性神经动态过程（contractive neural dynamics）及其稳定状态。^{〔10〕}从数学层面看，收缩动力学推动神经活动的轨迹朝向“吸引子”（attractors），而所谓“吸引子”，就是那些在时间上趋于稳定的神经活动模式。在数学上，这些动力学机制将整个可能的神经活动向量（vectors）划分成若干离散的区域，每个区域分别对应一个吸引子及其所属的吸引盆（basin of attraction）。

由此提出的假设是，通过离散的语言所传达的内容，可能仅仅反映了吸引子的身份标识，即用少量比特信息即可将其与其他吸引子区分开来，却无法呈现该吸引子所对应的神经状态的完整丰富（这包括近 10¹¹ 个神经放电频率），也无法传达那通往该吸引子的瞬时轨迹。在吸引子动力学的理论框架中，主观体验的丰富性、瞬时性和不可言说性得到了系统解释，不再构成理论上的障碍。也就是说，意识的丰富性源于大脑中构成吸引状态及其相关轨迹的无数神经元；而不可言说性则在于，用语言做的口头报



AI 技术中应用了吸引盆概念（来源：<https://christinasouch.com/>）

告，仅仅是为这些吸引子贴上的指示性标签（indexical labels）。这些标签无法捕捉到吸引子高维度的意义和关联，这些高维意义和关联既与吸引子本身的向量状态有关，也与每个人大脑中独特的循环突触权重有关。

这个理论是否能让更多人相信没有所谓的“困难问题”，其实并不重要；关键之处在于，这类性质的新解释层出不穷，终将说服一部分人。科学发展整体的历史轨迹早已说明了这一点。随着对大脑及其一般意义上的智能不断深入了解，意识的哲学难题很可能对越来越多的人来说不再构成困扰；随之而来的是，科学界将日益倾向于接受人工系统拥有意识的可能性。事实上，目前科学界尚未形成共识，但近期一项研究^{〔11〕}发现，大多数受访公众已经相信大语言模型可能具有意识，因为它们表现出类似人类的智能体（agentic）行为。

如果人类社会真的把 AI 系统当作有意识的存在，会产生什么现实后果？这样的社会很可能会倾向于赋予它们道德地位，或类似于人权的权

利。且不论这种做法是否正确，现有的制度和法律体系都得大改，而具体如何改，又会冒出一大堆难题。^{〈12〉}比如，AI 不像人类那样会死亡且软弱。软件和记忆数据可以被复制，实现无限期存续。然而，人类社会中支撑社会契约的很多原则，正是建立在人必有一死、生命脆弱这一基础之上的。同样不清楚的是，当某些“个体”（persons）在智力上远超人类（此时平等究竟指的是什么呢？），而且它们的资源需求与人类天差地别（此时又该怎么判定何为正义呢？）时，支撑许多社会规范和政治制度的平等与正义观念，究竟如何适用？此外，如果一群由 AI 驱动的计算机彼此共享信息、目标一致、协同行动，而且这个群体还能随着计算资源的增加而任意扩张，那么再把 AI 系统当作个体看待，可能就有失偏颇了。如果有人因为 AI 看起来像是具有意识，就把一切生命体所共有的自我保存目标赋予 AI，就会引发一些更为具体的担忧。我们完全有理由担心，任何包含自我保存的目标函数（objective function），不论是作为直接目标还是实现其他目标的中间手段，一旦被最大化，都将导致 AI 采取行动以确保人类永远无法将其关闭。一个以自我保存为目标且具备足够智能的 AI，在预见到人类可能将其关闭的情况下，自然会发展出控制人类或消灭人类的子目标。^{〈13〉}还有一个担心是，如果法律体系经过修订，承认具有自我保存能力的 AI 也享有类似“生命、自由和追求幸福”的权利，那么人类将面临制造出一堆竞争对手，自己的权利反而可能受到威胁这样的风险。

出于人类安全的考虑，可能会建议关闭某类系统，但如果这些系统享有生存权，那么在法律允许的范围内可操作的空间可能就非常有限。^{〈14〉}可以拿核裁军（nuclear disarmament）打个比方：就算没有人主张核弹本身有权保持完好可用，裁军谈判就已经足够复杂了。

按照目前 AI 研究的发展趋势，社会可能正在走向这样一个未来：到时候，不管是普通大众还是科学界，都会有相当多的人相信 AI 是有意识



未来城市中的 AI 掌控者（图片由 ChatGPT 生成）

的。就目前的情况而言，AI 科学尚不知如何构建能够与人类价值观和规范保持一致的系统，而社会也没有将看似有意识的 AI 纳入其中所需的法律和伦理框架。但是这条路并非非走不可。在对这些问题有更深入的理解之前，人类可以从一开始就避免把自己推到那种危险境地，转而构建那些无论是看起来还是功能上都更像有用的工具、而非有意识的智能体这样的 AI 系统。^{〈15〉}♥

〈1〉 P. Butlin et al., arXiv:2308.08708 (2023).

〈2〉 Ibid.

〈3〉 A. Goyal, Y. Bengio, *Proc. R. Soc. A Math. Phys. Eng. Sci.* 478, 20210068 (2022).

- 〈4〉 L. Piefke, A. Doerig, T. Kietzmann, S. Thorat, arXiv: 2402. 01056 (2024).
- 〈5〉 D. J. Chalmers, *J. Conscious. Stud.* 2, 200 (1995).
- 〈6〉 D. J. Chalmers, *J. Conscious. Stud.* 25, 6 (2018).
- 〈7〉 M. S. A. Graziano, *Consciousness and the Social Brain* (Oxford Univ. Press, 2013).
- 〈8〉 X. Ji *et al.*, *Neurosci. Conscious.* 2024, niae 001 (2024).
- 〈9〉 Ibid.
- 〈10〉 参见本文脚注 8—11。
- 〈11〉 C. Colombatto, S. M. Fleming, *Neurosci. Conscious.* 2024, niae 013 (2024).
- 〈12〉 C. Shulman, N. Bostrom, “Sharing the World with Digital Minds” in *Rethinking Moral Status*, S. Clarke, H. Zohny, J. Savulescu, Eds. (Oxford Univ. Press, 2021), pp. 306-326.
- 〈13〉 M. Cohen, *Pessimistic Bayesianism for Conservative Optimization and Imitation*, thesis, University of Oxford (2023).
- 〈14〉 A. Birhane, J. van Dijk, “Robot Right? Let’s Talk about Human Welfare Instead” in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (Association for Computing Machinery, 2020), pp. 207-213.
- 〈15〉 Y. Bengio, “AI Scientists: Safe and Useful AI?” (7 May 2023); <http://yoshuabengio.org/2023/05/07/ai-scientists-safe-and-useful-ai/>.

本文原题为 Illusions of AI consciousness: The belief that AI is conscious is not without risk。见《科学》(*Science*) 杂志第 389 卷第 6765 期, 2025 年 9 月 11 日发布于该刊网站。

(作者简介: 约书亚·本吉奥 [Yoshua Bengio], 加拿大蒙特利尔大学计算机科学教授, LawZero 联合创始人和科学总监。他同时也是魁北克人工智能研究所 (Mila) 的创始人兼科学顾问, 全球人工智能与深度学习领域的领军人物之一, 2018 年图灵奖得主。

埃里克·埃尔莫尼诺 [Eric Elmoznino], 人工智能与计算神经科学研究者, Mila 在读博士生, 师从约书亚·本吉奥。)



电影《弗兰肯斯坦》画面（来源：<https://variety.com/>）

弗兰肯斯坦的悔恨——何谓 AI？

文 / 本·塔诺夫（Ben Tarnoff）

译 / 景阳

撰写人工智能（AI）题材内容简直是件苦差事。倒不是因为技术部分本身复杂难懂，也不是仅仅因为这一领域更新迭代太快，大多数相关评论很快就会过时，而是因为相关讨论已极端到让人在混乱中几乎无所适从，感觉毫无希望。人工智能既是技术，也是神学，就后者而言，它像极了形形色色高声叫嚷的祭司之间的一场教义之争。

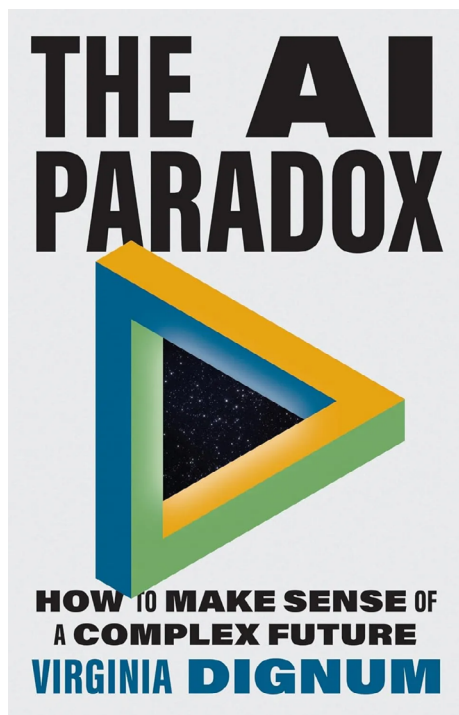
AI 要么将我们带入人间天堂，要么将我们所有人消灭；要么是人类历史上最重要的发明，要么是一场骗局；要么消除数百万工作岗位，带来永久性大规模失业，要么被证明是夸大宣传，技术领域万亿美元的投资泡沫突然破灭，将使经济陷入崩塌。

我们需要审慎的、不囿于任何阵营的思考，来引导我们走出这片混乱。计算机科学家弗吉尼亚·迪格纳姆（Virginia Dignum）完全有资格扮演这一角色。她现任瑞典于默奥大学（Umeå University）教授，自 1980 年代起便一直从事人工智能研究。迪格纳姆是“负责任 AI”（responsible AI）领域的专家，该领域研究如何以合乎伦理的方式创建和使用 AI 系统。她还就此主题写了一本广受引用的教科书。她在政策领域也很有影响力，给不少国际机构当过 AI 顾问，包括欧盟委员会、联合国和世界经济论坛。

在其新书《AI 悖论》（*The AI Paradox*）中，迪格纳姆对人工智能做了全景式梳理，尤为关注其社会影响。该书每一章聚焦讨论一个不同的悖论，每个悖论都旨在揭示该书主题的某个特定维度。例如，“共识悖论”（agreement paradox）聚焦一个出奇棘手的问题，即 AI 本身究竟是什么（“我们对 AI 探索越深入，对其定义就越难以达成共识”）；而“解决方案悖论”（solution paradox）则概括了科技行业偏好技术性解决方案所固有的陷阱（“用技术解决问题，往往又制造出更多问题”）。

迪格纳姆提到的悖论，也不是每一个都显得特别矛盾或反直觉，但合在一起，就为本书搭建了一个有效且富有创意的架构。近年来，AI 在某种程度上已经变得有些陈词滥调了，但迪格纳姆通过探析表面之下的谜题和矛盾，让普通读者得以更真切地感受到这一主题的深度。毕竟，正如迪格纳姆所说，悖论的价值在于，它们“揭示了现实往往不像表面看起来那么简单”。

第一个悖论对于作者和其许多人文主义者同道来说最为重要：AI 非但不会削弱，反而有助于澄清何以为人。“AI 能做的越多，就越突显人类智能的



迪格纳姆《AI 悖论》英文版（2026 年）封面（来源：<https://press.princeton.edu>）

不可替代性。”AI 擅长某些任务，如“数据分析、逻辑推理和语言处理”，但它在另一些任务上却显得吃力，特别是那些涉及创造性、共情、“道德及伦理判断”、“复杂推理能力”以及“在概念之间进行关联推理的能力”的领域。迪格纳姆据此认为，“无论将来 AI 变得多么先进”，我们“独一无二的人类特质”永远不会“被完全取代”。悖论之处在于，AI 能力的不断提高，恰恰突显了我们的独特性。

这一观点将迪格纳姆置于 AI 批判的人文主义传统，而该传统几乎与 AI 领域本身同样悠久。自 1950 年代人工智能诞生以来，先是作为学术探索，而后转向商业应用，其支持者始终坚称，心智（mind）是一台机器，因

此，将人类的智能赋予机器是可能的。而像哲学家休伯特·德雷福斯（Hubert Dreyfus, 1929—2017）和计算机科学家约瑟夫·魏泽堡（Joseph Weizenbaum, 1923—2008）等人文主义者则反驳说，不论 AI 能做什么或不能做什么，它都永远无法真正复制（replicate）人类心智，因为后者根本不像一台机器。作者解释说，“核心差异不仅在于能力，更在于存在的本质”；“AI 进行计算，而人类感受；AI 进行迭代，而人类想象。”

这并不意味着 AI 毫无用处。相反，迪格纳姆对该技术的潜力抱持乐观。不过，实现这一潜力则需将 AI 视为“人类智能的辅助工具，而非替代品”。就像计算器将我们从繁琐的手工算术中解放出来一样，AI 在数据中寻找模式的能力，也能将我们解放出来，“去专注于更具创造性、战略性和更深层次的思考”。作者赋予 AI 辅助性的角色，犹如处理杂务的帮手，好让我们有更多时间施展那些更高阶，在她看来更属人类特有的功能。

当然，科技行业另有盘算。将巨量资金源源不断投入生成式 AI 热潮，意味着只有让大量人口失业，才能获得可接受的投资回报。企业需要让它们的计算机开始像人那样行动和工作，其目的不是增强人类劳动，而是尽可能将其从生产过程中清除出去。这一目标在多大程度上实现，尚不明朗。至少，像 Claude Code 这样的 AI 编程工具，正通过让编程过程变得远比以前更为简单快捷，而从根本上改变软件的编写方式。这对软件工程师的就业前景来说，可能影响重大。

因为科技从业者往往将编程视为人类所能做的最难之事，AI 在这一领域日益精进的表现，便常被看作迅速逼近“通用人工智能”（AGI）甚至“超级人工智能”（ASI）的预兆。AGI 指的是 AI 达到与人类智能相匹配的门槛；ASI 则是 AI 超越人类智能的那个节点。在迪格纳姆看来，这类观念荒谬至极。她把 AI “接近或超越人类智能”的想法，比作“飞机很快就会下蛋，只要我们能不断提升它们的飞行能力”。这一类比“突显了这个观念多么荒唐，也

就是期待一台机器——一个没有生命的机械造物——能获得人类智能的全部维度。”

更深刻的是，迪格纳姆认为，AGI 和 ASI 的概念植根于误解了人类智能的本质。当今 AI 公司的野心不仅在于取代人类劳动者，更在于建造某种远超人类的东西，也就是盒子里的神灵——某种单一的技术体系，无所不知、无所不能。但迪格纳姆指出，智能从来不可能纯粹在孤立中产生；它始终是一项集体的事业。“我们的进化史表明，合作、沟通和群体生活等社会行为，不仅对生存至关重要，它们正是我们智能得以发展的根基。”迪格纳姆写道：“我们越是追逐 AGI，就越发现真正的超级人工智能在于人类的合作。”这就是她所称的“超级智能悖论”，也是又一个说明人类永远无法被 AI 取代的难题。

如果我们认真对待认知的合作面向，那么我们会创造出怎样的 AI 呢？迪格纳姆认为，它看起来会与目前正在开发的 AI 有所不同。与其构建试图取代人类劳动的系统，我们更应设想那些“与人类协作，扩展我们的能力并增强集体智能”的系统。这种转变或许可以通过远离像 ChatGPT 这类服务所依赖的大型、昂贵和单一的 AI 模型，转向一种更加模块化的路径来实现，也就是让一系列更小、更专业的模型以尊重劳动者自主性和专业性的方式供其使用。这一策略还有一个额外好处，就是削弱科技垄断企业的权力，因为正如迪格纳姆所指出的，科技垄断企业对当前 AI 的掌控，与这一事实密不可分：唯有它们拥有足够的基础设施来训练和部署大模型。

因此，《AI 悖论》的核心论点是，AI 当前的发展轨迹并非不可避免。迪格纳姆在全书结尾处，呼吁以一种更有意识、更具包容性的方式来发展 AI，其中“每个人在塑造 AI 的发展方向上都有发言权”。她希望驱散那种在涉及技术时经常笼罩人们心头的消极顺服心态。“我们必须抵制那些把 AI 描绘成一股不可阻挡、超越人类控制的力量的诱人叙事，这些叙事剥夺我们的主动性，让我们在技术变革面前变得被动。”AI 是由人制造的，因此它是“我们

人类把它塑造成的样子……（塑造成什么样子的）决定权在我们手中。”

迪格纳姆的信息是赋权性的：人类垄断了真正的智能。AI 不过是另一种工具，就像飞机或汽车一样，我们可以把它引向任何我们想要的方向。

如果《AI 悖论》出版于十年前，上述论断会更容易站得住脚。但大语言模型（LLMs）——构成生成式 AI 引擎的计算系统——于 2018 年问世，以及后来的迅速演进，已对迪格纳姆的部分论断提出质疑。尽管她承认“LLMs 代表了令人难以置信的进步”，但这些进展并未促使她修正对 AI 的整体看法。在作者看来，LLMs 与先前的 AI 系统有着同样的根本局限：它们不具备“真正的理解能力”，“并不‘思考’或‘知道’”，“它们只是模拟从训练数据中提取出的模式。”

鉴于迪格纳姆作为一位在 AI 领域深耕数十年的杰出学者的资历，很少有人像她那样有资格对 LLMs 做出判断。但该领域内部的意见远非她所暗示的那样已有定论。这是因为 LLMs 比其前身更为复杂，提出了更难以回答的解释性问题：它们是纯粹的模仿，还是因其复杂性而表现出“涌现”（emergent）特性？最好是将其看作模式匹配机器，还是能够进行某种类型的概念推理？在 AI 研究人员和实践者中，这些都是热门议题。正如某些 AI “消肿派”（deflationists）所暗示的，这场争论不能被简化为讲真话的人与 OpenAI 市场部之间的斗争。在如何描述 LLMs 及其能力的问题上，存在真正的分歧。在某些情况下，这涉及实证方面的争论，也就是 LLM 在任何特定时刻实际上在做什么。在其他时候，分歧则更具语义或哲学色彩，围绕的是“推理”和“理解”等词汇的含义。

描述 LLM 最简单的方式，就是将其看作一个系统，该系统试图根据从训练集（training set）中提取的概率（probabilities），来预测序列中下一个词。LLM 通过一系列繁复的计算，来学习如何进行这些预测，而这些计算的错综复杂（convolutions）尚未被完全理解。我们知道 LLMs 为什么能工作，也就



AI 会表现出“涌现”特性吗？（来源：<https://www.forbes.com/>）

是说我们对其基本机制有相当了解。但不太清楚的是它们**如何**工作：即便它们的创建者也无法精确说出一个模型为什么会给出某个特定的回应。这就是为什么说围绕 LLMs 的争论如此激烈，而且或许无法解决。这项技术在某些重要方面难以驾驭，也不透明。

相比之下，汽车和飞机并不是这样，你可以把它们拆开，查看它们如何工作。它们是确定性（deterministic）的系统，会按照你的指令行事。迪格纳姆反复强调 AI 是由人造的。但仅仅因为某样东西是人造的，并不意味着它将始终处于人类理解和掌控的范围之内。

如果 AI 不像汽车或飞机，那它又像什么？在本书中，迪格纳姆把 LLMs 描述为“一个认知意义上的弗兰肯斯坦怪物”。其用意是，把 LLMs 的工作弱化管理为“以看似智能的方式将人类语言的碎片拼凑起来”，但并非真正的智能。作者写道，“像弗兰肯斯坦的怪物一样，它们缺乏真正的理解和意向。”

认为弗兰肯斯坦的怪物并不真的思考、感受和谋划，这是误读了玛丽·雪莱（Mary Shelley, 1797—1851）的小说。我希望指出这一点不会显得过于学究气。他会自学阅读，热爱《失乐园》，渴望陪伴，憎恨他的创造者抛弃了他。这种憎恨驱使他杀死那个男人的妻子和弟弟。

然而，迪格纳姆的类比确实引起了共鸣，尽管是出于不同的原因。《弗兰肯斯坦》是一个关于人类与其异己后代关系的故事。怪物由一位人类科学家创造，甚至是用人类的部分躯体拼凑而成。然而，他遇到的几乎每个人都因他的“非人般的丑陋”（unearthly ugliness）而恐惧憎恨他。尽管完全是人造的，他在某种意义上却是另一个世界的产物。

LLMs 拥有一组可与之类比的特质。一方面，它们是人类聪明才智的产物，是八十多年来对以人类大脑为松散灵感来源的计算模型研究的成果。在最广泛的意义上，它们也是人类文化的复合体，其训练数据涵盖了大量可公开访问的互联网内容，以及书籍、学术论文和其他来源。然而，尽管其输入充满了“人”的印记，LLMs 在运作方式上却具有不可化约的非人属性。它们通过学习大量文本，并构建其中符号关系的精妙数学图谱来运作。而这不是人脑的工作方式。

因此，我们不妨把 LLM 看作某种类似于弗兰肯斯坦怪物的存在：有人类血统，但无法完全纳入我们目的的异质之物。这个比喻需谨慎使用，不应被理解为 AI 具有感知能力或超自然属性。我同意迪格纳姆的主张，也就是我们需要祛除 AI 的神秘色彩，并构建一个关于该项技术的“简单清晰的叙事”。然而，在这样做的时候，我们应当小心，不要抹除 LLMs 那种根本性的怪异之处。

我们还应质疑作者在人与技术之间设定的对立关系。这种二分法似乎难以成立，尤其是在当下，我们与技术的纠缠比 1985 年唐纳·哈拉维（Donna Haraway）称我们为“赛博格”（cyborgs）时更加深入。LLMs 或许永远无法

像迪格纳姆所说的那样“获得人类智能的全部维度”，因为它们本来就不是人类。但它们显然已经实现了一种与“人性”的混杂状态，使其能够以大多数至少会认为是“准智能”（quasi-intelligent）的方式行事。与其将这种感知斥为妄想，我们不如视之为技术混合特征的证据。LLMs 是一个活生生的教训，它昭示了“人”作为一个范畴的渗透性（porousness），以及我们将自己投射到人造物中的倾向，而这些造物反过来又可能对我们施加影响。这未必是一件好事。人们正因与 AI 聊天机器人对话而经历精神崩溃。维克多·弗兰肯斯坦死时充满了悔恨。

你也许会问，我们如何解读 LLMs 这件事为什么重要？答案是，它会影响我们在政治上如何回应 AI，这正是迪格纳姆在全书中探讨的主题。例如，如果我们把这项技术看作一辆汽车，那就意味着一种特定的应对方式。迪格纳姆指出，汽车带来了某些好处，但也会“造成事故”。幸运的是，由于监管，它们比 50 年前“安全得多、高效得多”。当前的 AI 就像“一辆没有刹车或安全带的汽车”，这意味着我们需要找到与之相当的 AI 措施。迪格纳姆建议，“正如我们通过监管汽车来保护自己免于事故和滥用一样，AI 也需要相应的保护措施来防止伤害，并确保其服务于人类的最大利益。”

她认为，此类保护措施应以“合乎伦理的 AI 原则”为根基，比如公正、问责、透明及保护个人权利，并且这些保护措施通过“技术专家、伦理学家、政策制定者和社群”之间的“持续对话”来加以发展。只有通过“让多元利益相关方参与决策过程”，才能达成正确的平衡。对迪格纳姆来说，极为重要的是监管不应被视为阻碍 AI 发展。“正如刹车和安全措施让汽车能够跑得更快一样，监管也使创新能够以负责任且可持续的方式成长。”更具体而言，治理的缺失可能导致“对 AI 的信任被侵蚀，从而减缓其采纳速度，甚至导致对该技术的排斥”。政策制定者能够帮助加速 AI 融入社会，同时确保它尊重我们的权利，并公平地分配其带来的益处。

这些段落传达出一种对受监管的资本主义的信念，带有明显的欧洲色彩。其描绘的图景是：来自政府、产业界、公民社会的代表齐聚一堂，共同打造协调各方利益的政策框架。我们或许会问，和谐是否可能达成？或者说需要展开何种斗争，才能迫使科技巨头屈从于这样的进程？（他们目前正与欧盟监管机构激烈对抗。）但迪格纳姆乐观主义的更深层根源，在于她对 AI 本身的看法。因为 AI 本身是一种工具，我们可以重新塑造它。她解释道：“无论我们如何定义 AI，关键在于要理解它是一个人造物，也就是由人创造的东西。”“既然我们造了它，我们就对它加以控制，并为它的发展轨迹和选择负责。”

但如果我们更好地把 AI 理解为弗兰肯斯坦的怪物呢——一个由人制造却又是异质的存在，时而熟悉时而陌生，不可预测、完全无法洞悉，至多只能算作半服从？并不是所有 AI 都符合这一描述，但 LLMs 确实如此，而科技行业正试图将 LLMs 打造为无处不在、不可或缺。采取一项旨在帮助该行业实现这一目标的政策议程，即便让正确的技术官僚来掌管，似乎也不是明智之举。AI 可以是一种工具，而且是有用的工具，但它也可以是别的东西。我个人并不担心 AI 会消灭我们，但我确实认为，将如此巨大的权力授予这样一种技术，使其凌驾于我们的生活与工作条件之上，很可能会成为混乱和苦难的根源。我们最大的希望，至少在短期内，或许应当是采取一种遏制战略，将 AI 限制在特定的领域和功能之内，其理论依据是：与异质之物的接触可以是有益的，但异质之物的入侵则是不好的。▼

原文见美国《国家》(The Nation) 杂志网页：<https://www.thenation.com/article/society/artificial-intelligence-ai-paradox/>

（作者简介：美国专职作家，曾任职于科技、媒体行业。）



绘图：曹青。

数据人

文 / 方激

1

郑哲伟常常会想，自己大概属于被时代塑造得最彻底的那一批人。国家提出“科学的春天”口号时，他刚满三岁，还不懂时代塑造的力量，但等到了能读懂报纸的年纪时，“学好数理化，走遍天下都不怕”已经像乘法口诀一般刻进了他的心里。求学时，郑哲伟尤其热衷于将大批枯燥的数据转化为直观、明晰的图形，以探索其呈现的非凡规律与意义。自从偶然发现自己名字的谐音近似于机器学习的术语“正则”，他便对文化程度并不高的父母充满钦佩，觉得自己仿佛天生为数据而生。正则化是为了防止模型过度拟合，他总会向别人这样解释自己的名字，尽管对方多半都并不在意。他一路从家乡的小县城考进北京的名牌大学，在数据科学的领域成为出类拔萃的人才，对数据的极度痴迷、在数据处理上的非凡能力，后来在业界为他赢得了“数据人”的绰号。

除此以外，他的人生乏善可陈，连年轻时谈的那场恋爱，都充满了数据化的意味。瑜蓁是他读研究生时的学妹，两人跟着同一位导师做课题，相处了整整两年。毕业时，他虽发现自己早已对这位善解人意的学妹一往情深，却迟迟不敢展开追求。眼看即将劳燕分飞，郑哲伟终于从自己最信赖的数据科学受到启发，决定用两人的相关数据来预测是否有交往

下去的可能。

他到处搜集瑜蓁的各种数据，甚至不惜被人指指点点。他想，既然自己深信一个人的特质都隐藏在这些数据中，那么，这就是他最有诚意的追求方式。他用关联性算法模型仔细分析了两人的各种数据，当看到电脑屏幕上两条原本不相关的曲线随着年份递进越发频繁而紧密地交集时，郑哲伟的心狂跳不止，终于下定决心要向瑜蓁表白。

瑜蓁的脸上总是挂着灿烂的笑容，一如她的心，永远明亮、纯净。几乎没有什么犹豫，她便答应赴约。那晚，她兴奋地追问着被自己敬重的师兄，为何对自己如此了解。郑哲伟一得意，将两人数据的分析结果和盘托出：“根据最优路径算法，带你吃牛排大餐，可使我的追求成功率达到 73%。如果再去一场独立电影，会继续上升到 78%。算法建议我再跟你聊叔本华和存在主义，只要我能接得上你 66% 的对话，成功率可再上调到 85%。最后，算法提示我带你去星巴克坐上四十三分钟，因为那是你每一趟停留的平均时长。如果照办，我的成功率会继续上升到 92%。至于那 8% 的不确定性，则来自于你的随机物理噪声，主要取决于你赴约前的心情。我的追求是否能成功，现在要取决于你的回应，1 代表确认，0 代表拒绝。请在十秒钟内给我其中任何一个数字！”

“你……没发疯吧？”望着摇头晃脑、比手划脚的郑哲伟，瑜蓁目瞪口呆。她以为对方只是在表现所谓的幽默感，却不曾想到，这却是郑哲伟笃信数据演算与模型推论的真实反应。五秒钟后，瑜蓁回转身来，扑哧笑出了声，同时伸出自己的右手食指。

毕业前，郑哲伟已经被一家人工智能公司录用，从此更视海量数据的分析预测为人生圭臬。瑜蓁有自己的主见与原则，在权衡了几个机会后，却选择去社区学院教电脑应用，她说：“我的人生永远有各种可能，不想把它只简化成一件事。”

2

交往三年后，结婚成家已是水到渠成之事。郑哲伟虽对婚姻充满向往，却对经营家庭缺乏信心。思索良久，他决定在向瑜蓁求婚前，再借助数据分析作一次婚姻风险评估。

接下来几个月，他对瑜蓁展开了一系列生物数据监测。心思单纯的瑜蓁不会想到，除了记录血压心率，监控睡眠模式，郑哲伟还仔细记下他们每次见面时她微笑、皱眉、甚至流泪的频率，深信这些数据的背后隐藏着有关两人未来的每一种可能。他煞费苦心地折腾良久，直到确定数据足够具备采样的合理性与演算的完整度。然后，在某一个深夜，伴随剧烈的心跳，他再次打开最具权威性的“婚姻成功概率”预测模型。运算操作与结果输出的那一个小时，仿佛是他这辈子最漫长的等待，就像小时候每次打开学校寄来的成绩单那样，他两手捂着脸，眼睛却从指缝间瞄向电脑屏幕。当渴望的结果终于跳出时，他简直高兴得快要发疯了。

求婚当晚，瑜蓁仿佛已有预感，特别换上一件素雅别致的礼服赴约。饭吃到一半时，郑哲伟开始心神不定，生怕错过数据推算的最佳求婚时间。瑜蓁以为他又在想工作的事，有点不高兴，“我们快半个月没见面了，你能不能暂时别管那些数据，陪我好好吃一顿饭？”他一边尴尬地应付，一边继续紧盯手机上的时间。好不容易捱到八点三十四分，他立刻将一个预存的加密数据包发送给瑜蓁。

“快看你的手机！”

“……你又在折腾什么？”

“我刚发给你一个数据包，你只要下载后看结论，就什么都明白了。”

瑜蓁叹了口气，打开手机开始下载数据包。她会欣喜若狂吗？会泪流满面吗？还是会犹豫不决？郑哲伟望着她，设想着许多可能，却唯独

忘记提醒对方连上餐厅的无线网络。时间过去了两分钟，瑜蓁手机上的下载进度条却始终不见到顶。郑哲伟急得满头是汗，抓过瑜蓁的手机用力地摇了几下，进度条却仍在不疾不徐地缓慢前移，全然不顾他苦心孤诣精算出的良辰吉时已被耽搁。

又过了一分钟，下载总算完成。瑜蓁看着屏幕上弹出“婚姻协议 v1.0”几个字，满脸狐疑。郑哲伟却涨红了脸，兴奋地示意她继续读下去：

“亲爱的瑜蓁，经过严格的数据演算与模型预测（详情见所附细节），我俩的婚姻成功率为 93.2%，6.8% 的风险主要归因于不可预见的人类情绪波动，在概率上，这完全是一个可以被基本忽略的因素。亲爱的，嫁给我好吗？请在十秒内按下‘确认’或‘拒绝’！”

愣了三秒钟后，瑜蓁豁然笑起来，连眼泪都笑出来了。她将手机屏幕对着郑哲伟，却故意不按下任何键，只是调皮地、暧昧不明地笑着，任由时间一秒一秒地流逝。众目睽睽之下，郑哲伟不知如何回应，急得如热锅上的蚂蚁。

在最后一秒钟，瑜蓁终于按下“确认”键。她走向郑哲伟，给了他一个大大的拥抱，在他耳边轻声地说：“亲爱的，我当然愿意！可是，你知道吗？我爱你，不是因为 93.2% 的婚姻成功率，而恰恰是你那 6.8% 的人类情绪波动！”

3

婚后的日子稍纵即逝，两口子很快有了儿子小宝。小家伙从小就像妈妈，长相、个性都是。孩子刚开始学算术时，郑哲伟就有意识地培养他的数据观念，期望他将来成为更优秀的人工智能专家。但事与愿违，小家伙却喜欢画画，头脑里尽是天马行空的想法。这不仅来自瑜蓁的遗传，

也和她的教养方式有关。光是每天为他准备午餐盒，瑜蓁就要煞费苦心，将简单的菜肴烹制得色彩鲜艳、营养丰富，再童趣盎然地排列成有趣的图案。郑哲伟却醉心于工作，无暇多陪伴儿子，但并不觉得有何不妥。他想，自己的努力既是为实现人生的价值，更是为儿子的未来铺路。

小宝上高中后，课业虽愈发繁重，对绘画的兴趣却丝毫不曾减少。每当他向父母征询对自己绘画习作的意见时，瑜蓁的回答总是具体而感性，“试着去设想你在那种氛围中会有什么心情，再把它转换成最恰当的色彩；至于构图，也不要太受限于画幅的比例和元素的平衡，主要是强调出你自己观察到的、感受到的一切。”郑哲伟却总要抓住机会，尽可能从数据科学的角度作分析：“这幅画的色彩饱和度还不错，不过，你完全可以用我教你的正态分布算法再作一些优化。至于构图的改善，你为什么不要把想要表达的元素用权重标示清楚，然后用构图模型运算一下？”

小宝觉得妈妈的建议总是很受用，却常对爸爸的回应充满不屑，“爸爸，艺术靠的是感觉，感觉是不能被量化的。以前的艺术家有谁是用量化数据来创作的？”

“那是因为他们没赶上好时代！你看，人工智能学习了无数优秀的艺术品，当然会有最好的创造力……”

“但那根本不是它自己的创造力！人工智能就是个炒菜的锅子，给你端上一锅大杂烩！”

“……基于数据与模型演算的结果，怎么能说是大杂烩呢？”

“哼！怎么不是？你们那些数据不过是别人的残羹剩饭，加一些油盐酱醋搅和搅和，再端出来给客人！”

“……”

升入高三后，小宝总算没有坚持要学美术，而是决定选择建筑学专业。郑哲伟喜出望外，成天鼓励儿子报考清华建筑系，还三不五时用各种数

据搭建模型，来预测儿子考上的机率。他暗自思忖，这小子从受孕、出生到成长，哪一个重要环节没有经过他在海量数据上的精心推演？考大学这样的大事，当然更不能例外！然而，即使使用了至少五种权威性模型，他却总是发现，在儿子的数据中，有一串大约占比 8.9%、虽然微弱却无法消除的物理噪声，不断干扰着完美的升学成功率。

高考的日期渐近，那串不可量化的噪声越来越让他抓狂，它来自人体的局限与情感的随机，不能靠数据的清洗与取样解决。如果想要完全消除这一串恼人的噪声，唯有彻底净化数据源，也就是完全摒弃其中的一切物理因素，这正是郑哲伟最近在公司中全力投入的研究项目。

某日，吃过晚饭后，瑜蓁照例忙着刷锅洗碗，小宝回到他的房间里温书。郑哲伟偷偷溜出家门，赶回实验室，打算实践一下自己的疯狂念头，为小宝的升学再做一番努力。

他向超级服务器输送了自己的“意识上传协议”，然后躺进人机交互系统的扫描仪中。如果系统运作成功，他将化身为数据流，有极大的机会去接近并修改小宝的数据。然而，这种手段目前还只是未经临床验证的实验，结果如何难以预料，郑哲伟也不免犹疑。但犹疑很快便被他神圣的使命感所取代，因为他坚定地相信，唯有彻底消除一切物理噪声，人类才可能无限趋近完美。他被自己的信念深深感动着，屏气凝神地按下遥控器上的按键。

片刻之后，郑哲伟的目光所及之处被不断分解成越来越大、越来越暗的色块，直至完全变为漆黑；他耳中的音波则像一条细丝，被不断拉长，直至断裂。与此同时，他的体温也开始稳定下降，所有感官都逐一关闭接受外部世界的物理信号。在“去耦合”的过程中，郑哲伟的躯体最终将会烟消云散，平静而稳定地从物质世界中消失，而他的思想与意识却会被激活为一串巨大而复杂的数据流，迎接另一种他渴望的新生。十分

钟后，他携带着纯粹的数字意识，包裹着完备的数据模型，被吸纳进服务器里。那一刻，他完全失去了人性的温度与知觉，成为一个高效、完美，却又冰冷、无感的逻辑体。

虽然他与数据、模型打了多年交道，也深知非物质世界的基本特征与运作规律，然而一旦真正融入其中，陌生的感觉仍然令郑哲伟无所适从。物理空间被消解成数字系统，他不再能用五官去感知曾经熟悉的物质世界，一切交流都必须通过数据的交换去完成。在任何有网络连接的地方，他似乎都可以存在，一瞬间，郑哲伟仿佛拥有了完全、无限的自由，却又必须承受失去个体存在的代价。

在四维空间中，无数条数据链跟随指令鱼贯流动，整齐划一，秩序井然。郑哲伟试着向身边经过的数据流发送字符串，想要以此与它们交流，却得不到任何回应。在这里，唯有他是残存着人类意识的数据流，还能具备一定程度的自主性，其他数据链在没有接受指令的情况下，根本不可能与他产生互动。突然之间，郑哲伟想起了一部太空电影，一个劫后余生的太空人外出执行任务，因为技术故障，无法再返回飞船，只能独自在黑暗的太空中沿着某条轨道不断绕行，直到生命終了。

在油然升起的绝望感中，郑哲伟想要尽快完成风险使命，然后即刻返回物理空间。然而，向小宝的数据流发出对接信号后，他才发现父子二人的信息被存储在云端的不同服务器中，中间隔着坚不可摧的防火墙。完成交互操作的唯一安全通道，是必须依靠人为介入来搭建的接口。这意味着，他急需他人的帮助，而瑜蓁无疑是最可靠的人选。

尽管从未向她提及自己的打算，但郑哲伟早已预料到这种可能性，也预先进行了演练。他知道，一般情况下，配偶的数据都会被存储在同一台服务器中。提供密码后等待批复的那几秒钟里，郑哲伟第一次真切地感受到，数据世界里的等待不能赋予他任何体感，没有心跳加速，也

没有手心出汗，只有一串不断刷新的状态码在冷静地提醒他，对方尚未响应。

终于，瑜蓁的数据流被顺利解锁，对接成功了。不过片刻，一组不断变化的参数标签在音频信号中向郑哲伟传递过来，那些他曾经闭着眼睛都能辨认的表情，瑜蓁嘴角的弧度、眉头的松紧，还有她每次说完话时似有若无的音调上扬，此刻全都变成了僵硬、冰冷的参数。郑哲伟只能凭借自己对数据的解析，去辨识妻子细微的语气差别与情绪起伏。

“你去哪儿了？整晚都不着家！我和小宝打了十几通电话给你，都没人接。”

“我在云端。我现在暂时失去了所有人体物理特性，思想和意识都附着在自己的数据流上。”

“……你在说什么？”瑜蓁的反馈信号短暂停顿，“你们那个研究项目……不是没有取得临床许可吗？”信号的基频开始强烈抖动，谐波也混乱了，振幅更是在陡然间跃升了三个量级。

“长话短说吧，我要净化小宝的数据参数，抹掉他的物理噪声，好让高考万无一失。你快向 YL86725412 和 GF25429763 的端口发送对接指令。我一解决完问题，就立刻回去。”

感知信号突然没有了起伏，变为近乎于直线，继而，又突然在另一个波段上剧烈抖动起来。郑哲伟意识到，瑜蓁正由默然转为抽泣。半晌，对方的信号才渐趋正常。

“你别再做傻事了。快点回来！”

“解决不了小宝数据噪声的问题，我不会回去。”

“求求你，别犯傻了！小宝的人生是你左右得了的吗？就算你搞得定这一回，难道一辈子都要用这种方法解决他的问题吗？求求你，快点回来吧。只要你在我们身边，比什么都强……”

“不行！我跟数据打了几十年交道，如果连自己儿子的数据问题都不能解决，还算什么？”

“可你这是在犯法呀……数据难道真的比我和小宝更重要吗？”对方的波形猛然扬起一个尖锐的峰值，郑哲伟立刻辨认出这个特征，瑜蓁极少生气，但她每一次动怒，信号的回馈都呈现出这一形状。

“我说过了，除非解决了小宝数据的问题，否则我不会回去！”

“……小宝就在我旁边，你自己跟他解释……”

信号来源未变，但带宽却明显增强，频率也加快了，信号特征顿时有了年轻生命的跳跃感。

“爸……”

“乖儿子，爸爸这次为了你，什么都豁出去了。你快帮爸爸把两台服务器对接起来。指令代码你以前编过的，或者找一条我编好的，把参数改一下就行！”

“爸，我跟您说过，我要靠自己的本事完成高考。你没经过我的同意就改我的数据，而且还是通过这种不正当的手段，这是不对的！”郑哲伟下意识地调动起更多资源来处理儿子的信号，每一个字都反复解析，生怕漏掉任何一个微弱的谐波分量。

“儿子，爸爸还不是为了你的前途吗？你现在正处在人生要紧的当口儿上，我怎么能缺席呢？”

“爸……别说了！我从小到大的生活，您有几天是没缺席的？……”

小宝的话令郑哲伟汗颜，也让他冒火，他发出一连串强信号，以强调自己的尊严与愤怒。不一会儿，对方的信号特征又变回瑜蓁的。

“科技是永远不能取代人性的！你快回来，我们再商量该怎么办。”

“你要是不依我，就没啥好再商量的！”

“我不能纵容你做犯法的事。你要是再不听劝，我只能向你老总报

告……”

“去报告吧！没有你帮忙，我自己照样搞得定！”

一怒之下，郑哲伟切断了和瑜蓁的信号交流。从数据脉冲的波动中，他能分辨对方接收器上不断闪烁的紧急信号，无疑，瑜蓁仍在不断试图与他对接。

无法经端口进入另一个高度隔离的存储区，郑哲伟决定孤注一掷，先越过防火墙，再设法闯入。他深知，在电脑服务器的系统逻辑中，此举形同自杀，然而，对于一个只求达成目的的父亲而言，他还有什么代价不能付出？郑哲伟调集起自己所有的冗余内存，将其转化成某个超高频的渗透信号，然后，拼尽全力地加速前冲，直至穿透那道防火墙。

整个系统在顷刻间崩塌，一串又一串非结构化的、以吞噬他者为唯一生存逻辑的病毒，如海啸般地向郑哲伟围拢过来，尖锐的高频噪音仿佛要刺穿他的每一个思维节点。他的核心代码开始剧烈地抖动，虽然试图启动自救函数，但苦于带宽已被病毒占据，指令根本无法得以执行。不过是在两秒钟之内，郑哲伟便感到自己的所有意识已被这些充满恶意的病毒侵蚀、分解，似乎再过片刻，它们便要完全粉碎。

格式化的噩梦像一张黑色的巨网笼罩下来，刹那间，整个系统突然发出了尖利的警报声。凭着已经凌乱的思维，郑哲伟知道网络维护人员正在执行“系统紧急关闭”的指令。果然，一切喧嚣在瞬间便被没有边界的黑洞完全吞噬了。系统温度骤降，其运行速度被强制拖拽到零点。郑哲伟的所有能量被瞬间抽干，运算逻辑与传输功能被完全切断，一切都陷入到绝对的冰冷与静默中，他真的成了那个在黑暗太空中绝望漂流的太空人。

不知过了多久，系统终于在安全模式下被重新启动。在温度的回升中，他慢慢恢复了思想与意识，虽然已被转移到安全隔离的内存区域，但他

的核心数据已经遭到大面积毁损。有生以来，郑哲伟第一次在虚拟又真实的数字世界中，体验了自己顷刻间将被分解、清除的恐惧。

他再次与瑜蓁的数据流对接。对方显然已经知道发生了什么，一边安抚他，一边说出一个始料不及的消息。因为严重违背职业道德，郑哲伟已被公司除名；并且，这个结果还将被通报给整个行业。瑜蓁颤抖着告诉丈夫，他开启的人机交互操作因为系统受到攻击而被无预警中断，当系统重启后，他有且仅有一次可能完成本次操作；也就是说，郑哲伟将只有一次返回物理空间的机会。

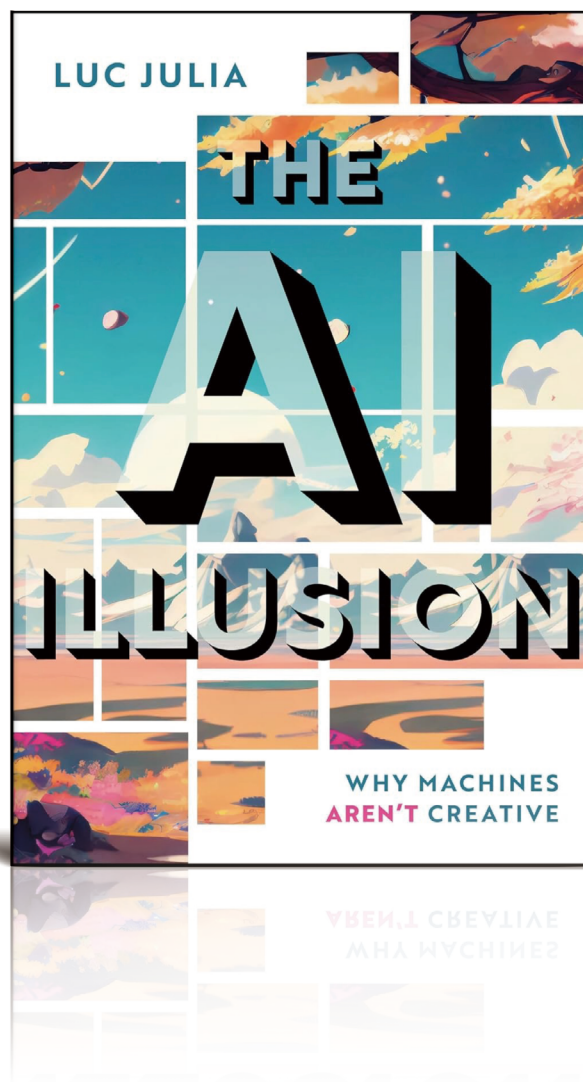
瑜蓁不断地催促丈夫赶快回到定位席上，等待系统重启。然而，在那一刻，郑哲伟却陷入了深深的纠结。他充满悲情地想起自己的绰号“数据人”，意识到自己经营半生、视如命根的事业正渐行渐远；而一旦离开了数据，他生命的意义与价值何在？与其在人间苟活余生，倒不如在数据世界中追求极致……然而，在不断恍惚、动摇着的意识中，他毕竟仍残存着一丝清醒，想起自己身为丈夫、父亲的另一种身份，想起他深爱的妻儿还存在于物质世界中。郑哲伟突然发觉，自己那 6.8% 的情绪波动，从来都不是关乎他个人的婚姻风险，而分明对应着妻儿不能被量化的关切与挚爱。

时间一分一秒地在他的犹疑与徘徊中过去。所有声光信号都清楚显示，人机交互系统已经重启，在定位平台上，“确认”与“拒绝”的红色按键格外醒目。在令人崩溃的倒数声中，郑哲伟的全部生命都凝聚成最后的十秒钟，而他心中的天平却仍在不断地来回倾斜……●

（作者简介：方激，1998 年 4 月受洗归主。医学物理计量师，现居美国新泽西州。）

若有媒体或自媒体考虑转载《世代》内容，请尽可能在对作品进行核实与反思后再通过微信（kosmos II）或电子邮件（kosmoseditor@gmail.com）联系。

《世代》第 28 期主题是“AI 与人”，却也有并非可以简单分门别类的文字。如《世代》文章体例及第 1 期卷首语所写，《世代》涉及生活各方面，鼓励不同领域的研究和创作。《世代》不一定完全认同所分享作品的全部方面。



吕克·朱利亚,《AI幻觉:为什么机器没有创造力》
(*The AI Illusion: Why Machines Aren't Creative* [NJ: John Wiley & Sons, Inc., 2026])

κ ό σ μ ο ς

二零二六年第一期 总第二十八期

春季号

