

面对“AI危崖”，人类应有所作为

■刘永谋

有人指出，20世纪下半叶新科技发展会导致意想不到的后果：单个人有能力实施危害巨大的恐怖行为，巨大灾难可能由无心的技术错误引发。

鉴于新科技可以毁灭整个人类文明的惊人破坏力，21世纪以降，很多思想家都开始关切当下的文明是否存在全局性的崩溃，甚至灭绝的生存性风险，使得人类社会如跌下悬崖一般，突然陷入黑暗甚至永夜之中？这便是近年来尤其是新冠疫情暴发之后，全球广为讨论的“文明危崖问题”。

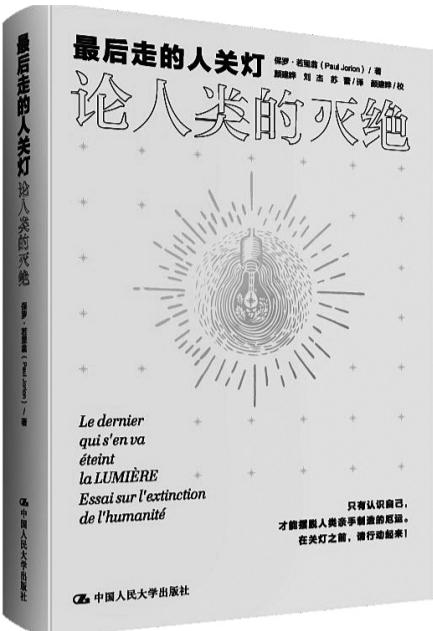
的确，人类从未如此切实地感受到技术末世的逼近。历经70多年的世界大体和平之后，“第三次大战”被人越来越多地提及。绝大多数人相信，它将是人类最后一战，全球核战之后文明将被荡平。一些人则认为，气候变化已经到了十万火急的程度，很快将迫使人类搬到地下，然后灭绝。另有一些人认为，穷国的“核武器”即廉价生化武器，迟早会造成本无法挽回的毁灭性灾难。还有一些人表示，超级AI很快会出现，之后硅基文明将取代碳基文明——很多人不相信这种观点，认为在这之前人类更有可能用核武器、AI武器、生化武器相互残杀殆尽。

法国人类学家、人工智能专家保罗·若里翁却认为“AI危崖”很重，并写了《最后走的人关灯：论人类的灭绝》一书加以讨论。该书的核心问题是，“为何人类的智慧可以制造出那些就算有一天人类消失了还能存在的智能机器，而不能同时保证人类自己的生存？”

若里翁的提问方式，包含着两个成见。第一，他相信人类在两三百代人之后即21世纪末很可能就会灭绝，并表示这是科学家尤其是物理学家、化学家和古生物学家经过计算得出的科学结论。

第二，他认为人类没有准备好面对“AI危崖”，不可能避免自身的灭绝。人类能做的事情就是分析人类悲剧如何发生。因此本书只能是一篇悼词，根本改变不了结局。

为了支持技术末世说，若里翁总结了“人类三大危机”。第一是环境危机，根源在于人类不可持续和不可再生的生存方式；第二是复杂性危机，根源在于世界联系越来越紧密，变得越来越复杂，人类难以认识和掌控；第三是经济与金融危机，即经济系统只从商业视角考虑问题，成为无



视物种未来和灭绝风险的敛财机器。

总之，“地球毁灭的原因就在我们身边，就是我们自己的行动”。

应该说，上述危机并非新见解，早已为各种社会批判理论所详述。何况这三大危机并不能得出人类很可能在21世纪末灭绝的结论。而且，人类灭绝不等于地球毁灭。换句话说，当代人的能力足以灭绝自己，却完全谈不上毁灭地球。

地球自诞生以来，历经多次物种大灭绝，以及超级火山爆发、小行星撞击、冰雪覆盖，不仅没有被毁灭，连此间的生命也未完全灭绝。所以，若里翁的言辞颇为夸张和武断。

更重要的是，若里翁没有仔细论证，就得出了人类根本应对不了灭绝风险的悲观结论。为什么人类不具备避免自身灭绝的能力？他的理由主要是人性使然，比如人类短视、毫无斗志、虚度光阴，所以他“对扭转形势也不抱任何期望了”。

如他所说，人们面对即将到来的灭绝可能

“跨越文明危崖需要数个世纪，这相对于地球历史只是一刹那，但对于人类文明却是一次长期艰苦的考验。在行过危崖的漫长旅程中，面临许多危险的歧路，不断考验人类的智慧和德行。”

《最后走的人关灯：论人类的灭绝》，[法]保罗·若里翁著，颜建晔、刘杰、苏蕾译，中国人民大学出版社2023年3月出版，定价：68元

有4种态度：否认和忽视危险；承认但毫不在意；积极面对，加速应对；承认但无能为力，“只能为人类哀悼”。若里翁坚持的是第四种态度，并自认为这是“对人类的命运做出的一种现实和真实的描述”，而“我们陷入对人类物种的过分眷恋是不明智的”。

如果说提醒人们重视即为贡献，本书的确敲响了技术末世论的警钟。但是，若里翁的丧气话一再显示出他对人性的武断。

古人类学研究表明，现代智人是一种具有强烈进取心的种族。在长期的生存竞争中，现代智人灭绝了地球上大量的物种，甚至灭绝了尼安德特人、弗洛勒斯人等其他同类。

当然，以此可以指责现代智人是残暴的物种，但残暴的另一面是进取、奋斗和勇气。面对灭绝风险，人类必定不会完全无所作为。仅仅哀叹“你们最好不要出生”，绝不是现代智人的做派。

三

总的来说，AI至少有两个问题直接与文明危崖相关。第一，AI特别是超级AI在总体上会

不会威胁人类文明？第二，对于全局性生存性风险的应对，AI是有所助益，还是只能加剧潜在的危险？

AI可能导致的文明危崖风险可以分为两类：一是AI灭绝，即AI可能灭绝人类；二是AI衰退，即AI可能导致文明衰退。

虽然在流行文化尤其是科幻文艺中，类似的问题非常“吸睛”，但是严肃的思想家并不关注，或者说认为“文明的AI危崖”远比不上气候变化、核大战和新发未知病毒等的威胁。一些人认为，在生存性风险中，超级AI对人类文明的威胁完全排不上号。

进一步而言，若里翁对人类能力的否定，并无令人信服的科学证据和推导，只是某种哲学思辨或价值观的选择。同样，相反的结论也无法被科学所证明。没有人能断定现代智人会安然跨过文明危崖，更没有人能找到一劳永逸的万全之策。

很多人认为，跨越文明危崖需要数个世纪，这相对于地球历史只是一刹那，但对于人类文明却是一次长期的艰苦考验。在行过危崖的漫长旅程中，面临许多危险的歧路，不断考验人类的智慧和德行。

应对文明危崖，首先需要直面危崖，而不是视而不见、掩耳盗铃。放任自流会错过应对风险的良机，莽撞冒进同样会加大悲剧发生的可能性。对于AI的发展，既不要小觑它的冲击，也不能夸大它的风险。尤其不能将还没有实现的“超级AI神话”当作已经发生的东西，结果是自乱阵脚，甚至误入歧途。

因此，人类必须保持勇气、下定决心，认真研究和反思可能出现的文明危崖。

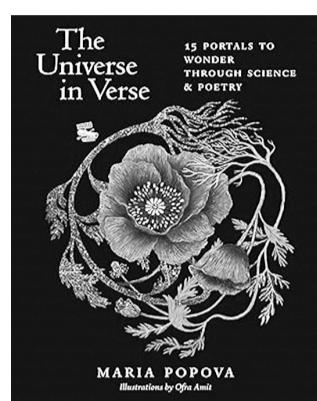
具体到超级AI问题上，既然超级AI可能奴役人类，甚至灭绝人类，我们为什么要发展超级AI呢？仅仅为了某种好奇心，就置AI产生意识的生存性风险于不顾，这难道不是完完全全的莽撞和不负责任吗？因此，AI发展的有限主义，反对将道德问题交给AI处理，反对在无法控制的情况下放任AI意识的出现。

对AI取代人类的可能前景，有些人甚至大为欢迎，高喊“爱那些机器人吧，我们应该爱这些我们设计的机器人而不是爱我们自己”。这多么荒唐！人类的使命不是迎接智能机器的来临，而是种族的繁衍、进取和幸福。说得严重一点，爱机器人而不爱人，多少有点反人类的倾向吧？

(作者系中国人民大学吴玉章讲席教授)

域外

可是，诗歌能够通过情感的后门，将一些观念偷运到我们的意识中，将我们日常观察世界和关联世界的方式，以及我们的偏见和先入之见都抛在一边，打开感受世界的另一扇大门。



2005年，北京搞过一场“照亮黑夜的烛光——科学经典名篇朗诵会”，我曾参加了该朗诵会的一点前期策划，也欣赏了朗诵会的演出。这场活动非常成功，可惜，后来就没有后来者。

12年之后的2017年，保加利亚裔美国学者玛利亚·波波娃首次策划组织了一场题为“Universe in Verse”（诗篇中的宇宙）的大型诗歌朗诵会。后来，同一题目的朗诵会年年举办，只在新冠疫情期间停办过一次，2024年4月举办的朗诵会已经是第七季了。为了不显单调，朗诵会中间会穿插一些音乐表演和魔术表演节目。

每次朗诵会，波波娃及其合作伙伴如美国诗人学会，都会精心选择一些涉及科学主题的诗歌，邀请一些著名的作家、音乐家、科学家朗诵，或是邀请当代著名诗人到现场朗诵自己的作品。

波波娃还会邀请著名科学家与她一起组织活动，如天体物理学家珍娜·列文。2018年第二季“诗篇中的科学”朗诵会就安排在列文的“先驱者工坊”。

这些活动非常火爆，总是一票难求。迄今为止，已经有成千上万的受众聚集到“诗篇中的宇宙”活动现场，一起聆听，一起思考，一起感受。他们都非常关注真与美的关系、爱与死亡的关系、有限与无限的关系。

由于“诗篇中的宇宙”已经产生了品牌效应，美国诗人学会决定设立“Universe in Verse Fund”（诗篇中的宇宙基金），支持诗人创作传播科学、讴歌大自然的作品。

现在，波波娃又在美国Storey出版社出版了一本同标题的图书——“The Universe in Verse: 15 Portals to Wonder through Science & Poetry”（本文作者译为“诗篇中的宇宙：通过科学与诗歌获得讶异感的15条门户”）。该书虽标注的正式出版日期是今年10月1日，其实已在亚马逊网站上销售多时。

波波娃在本书“引言”中写道：“我们的人生处于多种空隙之中——真相与意义之间的空隙、客观现实与带有感情的主观意义构建之间的空隙。我们的所有向往、我们的所有绝望、我们对所处困境的所有考量，无不瞄准空隙的这一端或那一端。在这瞄准的过程中有我们称为创造性的东西，即如何接触美——无论是定理之美还是商籁体诗歌之美。”

她接着说，诗歌似乎不太像是通往实在之本性的门径。可是，诗歌能够通过情感的后门，将一些观念偷运到我们的意识中，将我们日常观察世界和关联世界的方式，以及我们的偏见和先入之见都抛在一边，打开感受世界的另一扇大门。

通过诗歌、时间、空间和显著性（它们都是科学研究的原材料）等其他维度可以更充分、更忠实地融入我们的世界观，使得我们回归日常生活任务和存在性向往时，带着恢复了的弹性和平抱可能性的激情。

科学和诗歌，是让人们更深入地了解世界和更深切地热爱世界的工具。我们需要科学，帮助我们依据实在的词汇去面对实在，我们需要诗歌，帮助我们扩展和深化我们的词汇，以面对我们自身和互相面对。

本书收入了波波娃的15篇散文，讨论了暗物质的神秘、圆周率的无穷位数、树木的韧性、章鱼的智慧等五花八门的科学主题。这些文章生动地讲述了科学探索和科学发现的故事，穿插了一些科学家和诗人的生活经历。

书中的科学家中包含多位女性，她们中有许多并不是家喻户晓的。她介绍的这些诗人的创作灵感源泉和揭示的美并无二致。

就像诗配画一样，本书是“诗配文”，每篇文章都配上一首思考同一主题的诗歌。这些诗歌的作者包括美国天才诗人艾米丽·迪金森，英裔美国诗人W.H.奥登，美国历史上首位获得普利策诗歌奖的女诗人埃德娜·圣·文森特·米莱，美国黑人女诗人玛雅·安吉罗，曾获斯蒂芬·霍金科学传播奖的美国女诗人、散文家和博物学者Diane Ackerman，曾获普利策诗歌奖的黑人女诗人特蕾西·K·史密斯，等等。本书的插图是由以色列著名插画家Ofra Amit女士创作的。

沉浸式感受科学巨匠平凡而伟大的人生

■本报记者 甘晓

科学哲学家、科学史家范岱年曾说：“要了解科学的方方面面，阅读科学家的传记是一条重要途径。”

近日，由《自然辩证法通讯》编辑部与大象出版社共同策划的“中外科学家传记丛书”发布，该丛书30册全部出版。这套丛书凝聚了两三代专业研究者的心血，兼具专业性、原创性、思想性和可读性。丛书涵盖了200多位中外著名科学家传记，像是搭建的一条“时光隧道”，引领读者“穿越”至那些科学巨匠的生活与思想世界中，沉浸式感受他们平凡而伟大的科学人生。

40年专业积累

丛书内容编写的背后是《自然辩证法通讯》杂志40年的积累，收录的传记全部来自该杂志常年开设的“人物评传”栏目，从文艺复兴时期的欧洲科学家到远渡重洋将最初的西方近代科学知识带到中国的欧洲传教士，再到现代科学家。

其中，数学领域包括希尔伯特、哈代、陈省身、吴文俊等，物理学领域有玻尔、普朗克、薛定谔、海森伯、钱三强、束星北、王淦昌等，还有天文学、地学、生物学、计算机科学和若干工程领域的科学家。

《自然辩证法通讯》于1979年在中国科学院创刊，是国内科学史研究领域的权威杂志之一。2018年前后，作为杂志的副主编，该丛书主编之一、中国科学院大学人文学院哲学系教授王大明萌生将“人物评传”栏目集结出版的想法。

王大明的想法很快得到了科技史研究同行的支持，他们开始为这个选题寻找合适的出版方。经科技史学者、清华大学教授刘兵的推荐，他们找到时任大象出版社社科文献编辑部主任李光洁。

“相当于出版社用40年的时间邀请国内顶尖科学史专家，精心撰写了200多位科学家传记，出版社很难做到如此宏大和长远的内容积淀。”现任大象出版社副总编辑李光洁介绍，“我深知这个选题的分量。”

这些传记文章的作者大都是在相关领域学有专长的专家学者。例如，写过多篇数学家传记的胡作玄是中国科学院原系统科学研究所研究员；写过多篇物理学家传记的戈革是12卷本《尼尔斯·玻尔集》的译者，等等。

该丛书“总序”写道：“这些传记文章从专业和普及两个角度看，其数量之多、涉及领域之广、内容质量之上乘、可读性之强，在国内的中外科学家群体传记中都可以说是无出其右。”

“科学家传记是科学史上最具有传播价值的一种作品类型。”刘兵指出。

然而，目前国内科学家传记创作面临的一个问题在于，传记内容往往缺乏深度研

历时5年艰苦出版

2019年，大象出版社为该丛书立项。主编王大明、刘兵及中国科学院大学人文学院副教授、《自然辩证法通讯》编辑部主任李斌和多位作者参与其中，与大象出版社的编辑通力合作，得以完成对传记文章的编辑。

编辑团队对原刊物的传记文章进行了重新整理编辑。主要工作集中在4个方面，一是在总体设计上，丛书共分30册，每册收录8篇人物传记；二是基本按照学科领域来划分各个分册；三是每分册中的人物大致参考历史顺序或学术地位来编排；四是为照顾阅读的连续性，将原刊物文章中的所有参考资料一律转移到每分册的最后，并增加人名对照表。

这一过程历时5年，其间艰辛与挑战，非亲历者难以言喻。从书编者之一、中国科学技术史学会工程史专业委员会副秘书长陈印政告诉《中国科学报》，许多科学家在新中国成立前就作出了重要的科学贡献，但因为年代久远，很多当事人已经去世，如何能够精准还原史实成了问题。

陈印政回忆，在中国科学院院士叶培大的传记编写中，有文献记载其在1947年就采用杜黑特电路，主持设计、安装、调试了中国第一部100千瓦的大功率广播发射机。“由于叶培大先生已经去世，该广播发射机的使用单位是位于南京的中央广播电台，新中国成立后已经不复存在，了解此事的当事人也难以寻找。”他说。

为准确还原叶培大的科技贡献，陈印政多次到南京第二历史档案馆、国家图书馆查阅资料，并与中国工程师学会取得联系查寻历史文献，争取家属支持，查找遗物中的重要线索，最终找到了当时留下的研究论文和笔记资料，再加上文献佐证，终于弄清楚当时的技术细节和学术价值。

而对于中国历史上有重要成就的科学家，丛书编辑则需要扎进故纸堆，查阅大量的古代文献进行对比分析。这不仅要求他们有扎实的古代汉语基础，还要能理解古代数学符号体系，这对他们来说都充满挑战。

播撒科学种子

“科学家传记是科学史中最具有传播价值的一种作品类型。”刘兵指出。

然而，目前国内科学家传记创作面临的一个问题在于，传记内容往往缺乏深度研



“中外科学家传记丛书”(共30册)，王大明、刘兵、李斌主编，大象出版社出版

差、内容简化问题频出。这直接导致科学家传记实际的科学传播效果有待提高。

对此，刘兵相信，《自然辩证法通讯》杂志在科学史研究领域的学术地位与学术传承，对该丛书将严谨的科学史研究成果转化科普资料具有举足轻重的作用。“原刊物发表的文章不仅确保了科学传播的严肃性和准确性，也为未来科学家传记编写树立了一个高标准的典范，提供了极有价值的参考框架。”他强调。

该丛书在策划之初，就把科学普及作为重要目标。正如丛书“总序”提及的，从书编辑和出版的初衷，在于发扬理性思考、刻苦钻研、求真务实、勇于创新的科学精神，营造科学、爱科学，支持创新、尊重人才的文化氛围。

“希望当下的科研工作者能够从科学家的成长经历中得到启发，深刻了解他们在重大历史转折关头如何作出伟大贡献，更加真切地感受他们的科学思想和创新方法、科学精神和品格作风。”王大明表示。

青少年是该丛书的主要读者群体之一。自推出以来，该丛书先后入选“科学家（精神）进校园行动”优秀图书、全国中小学图书馆推荐书目和农家书屋推荐目录，获得了多项荣誉，显示了其广泛的影响力和认可度。

编辑团队期待更多青少年能够阅读这套丛书，因为书中科学家的传记故事所传达的科学家精神，有助于青少年感受科学家探索未知、追求真理的激情与执着，引导他们热爱科学、崇尚科学，培养科学素养，树立远大志向。