



贵州民族大学学报(哲学社会科学版)

Journal of Guizhou Minzu University(Philosophy and Social Sciences)

ISSN 1003-6644,CN 52-1155/C

《贵州民族大学学报(哲学社会科学版)》网络首发论文

题目：生成式人工智能会毁灭人类吗？——有效加速主义刍议
作者：卢云峰，吴语菲
网络首发日期：2025-02-19
引用格式：卢云峰，吴语菲. 生成式人工智能会毁灭人类吗？——有效加速主义刍议[J/OL]. 贵州民族大学学报(哲学社会科学版).
<https://link.cnki.net/urlid/52.1155.C.20250218.1646.002>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

生成式人工智能会毁灭人类吗？

——有效加速主义刍议

卢云峰,吴语菲

摘要:有效加速主义(Effective Accelerationism, e/acc)是近年来风靡美国硅谷业界的一种思潮,旨在通过最有效的策略加速科技发展,推行去中心化的治理模式,实现智能技术提升社会福祉的目标。与传统社会思潮不同,有效加速主义的主导者并非传统的社会思想家,而是来自业界的工程师;该主义的兴起也意味着工程师群体已经成为影响智能社会治理的主体之一。在技术发展速度还是安全性的优先选择上,有效加速主义与政界主流的安全主义观点形成对立;通过与后者的论战,有效加速主义逐步构建了自身的话语体系,其社会影响也在不断扩大。有效加速主义对政府管制的担忧源自技术中性论的认识,对智能技术发展的乐观判断则出于兼顾企业利益和社会责任的需求。然而,面对DeepSeek开源大模型的成功,有效加速主义者态度复杂;这表明其技术主张并非完全基于科技至上的理念,而是掺杂了科技发展主动权竞争和国家利益的考量。

关键词:人工智能社会学;有效加速主义;生成式人工智能;DeepSeek;ChatGPT

中图分类号:C912 **文献标识码:**A **文章编号:**1003-6644(2025)00-0000-18

作者卢云峰,男,博士,贵州民族大学社会学院特聘教授(贵州 贵阳 550025),北京大学社会学系教授、博士生导师(北京 100871);吴语菲,女,北京大学社会学系博士生(北京 100871)。

2024年2月,一个新词汇——“末日概率(probability of doom)”被破圈收录于《剑桥词典》。该词最初被科技爱好者们用来表示人工智能引发人类生存灾难的概率,取

* 本文为北京大学武汉人工智能研究院开放课题“数字社会的核心特征与治理模式研究:以武汉光谷为例”阶段性成果。

值范围在 0% 至 100% 之间。^①随着 DeepSeek、ChatGPT 等大模型的飞速发展,人工智能的社会影响无远弗届;关于末日概率的讨论不再限于留言板上的玩笑,而是成为社会各界高度关注的议题,并在业界形成一个从低到高的连续谱系。图灵奖(Turing Award)得主杨立昆(Yann LeCun)估计该概率低于 0.01%;埃隆·马斯克(Elon Musk)则认为其介于 10% 至 20% 之间;和杨立昆同时获得图灵奖的杰弗里·辛顿(Geoffrey Hinton)预测该概率在 10% 至 50%;而关注人工智能安全的 MIRI 创始者埃利泽·尤德科夫斯基(Eliezer Yudkowsky)则认为该概率高于 95%。^②如今,末日概率已成为人工智能文化中表达立场的重要社交信号。

需要指出的是,末日概率的数值并不具备严格的统计学意义,而是通过数据化的方式表达对智能社会前景的整体判断——人工智能对人类而言是利大于弊,还是弊大于利。同时,这一判断也影响了决策者在人工智能未来发展上的立场:是倾向于限制发展,还是加速推进。围绕这一议题,社会逐渐分化为两个对立的阵营:支持加速推进的阵营被称为“有效加速主义(Effective Accelerationism, e/acc)”,而主张限制的阵营则被称为“安全主义”。

OpenAI 董事会的重组便是这一对立的典型例证。2023 年 11 月 17 日,时任美国人工智能研发公司 OpenAI 首席执行官萨姆·奥尔特曼(Sam Altman)被解雇,OpenAI 联合创始人埃里克·布罗克曼(Eric Brockman)和格雷格·布罗克曼(Greg Brockman)同时提出离职,在科技界引起轰动。仅五天后,剧情发生反转,奥尔特曼与布罗克曼宣布重返原职。虽然 OpenAI 董事会未透露事件的具体细节,但可以明确的是,奥尔特曼与该公司联合创始人兼首席研究科学家伊利亚·苏茨克维(Ilya Sutskever)之间的价值观冲突是导火索,双方在商议公司未来发展方向时爆发了不可调和的分歧。据传,奥尔特曼支持“有效加速主义”,而苏茨克维则持反对态度。

作为人工智能领域在论战中产生的一种新近思潮,有效加速主义缘起于 21 世纪人工智能技术的迅猛发展。近年来,以 ChatGPT 和 DeepSeek 为代表的生成式大语言模

① "New Words - 19 February 2024", 2024-02-19, <https://dictionaryblog.cambridge.org/?s=New+Words+-+19+February+2024>, 2025-01-20.

② "List of P(Doom) Values", 2024, <https://pauseai.info/pdoom>, 2025-01-21.

型展现出高度智能化和通用化的特征,在逻辑推理、认知理解和具象输出等方面已可与人类智能比肩,能够完成情感分析、代码编写等复杂任务。学界和业界普遍认为,生成式人工智能是人类技术迈向通用人工智能的重要一步。^①文森特·穆勒(Vincent C. Müller)和尼克·博斯特罗姆(Nick Bostrom)预测这一颠覆性技术将在21世纪实现^②,而马斯克关于这一颠覆性技术的实现时间从原本预测的2029年提前至2025年年底^③。大量工程师和技术从业者参与这一社会话题的讨论,其中尤以有效加速主义与安全主义的争论最为激烈。

有效加速主义主张应以最有效的策略加速人工智能技术研发,放宽上游研发的限制,尽快实现技术奇点,进而创造繁荣并解决人文困境。在监管方面,他们提倡去中心化的监管模式,将权力重心从政府转移至企业,并通过自发的市场秩序对人工智能进行规制。^④与此对立的安全主义则强调,必须首先确保高级智能技术的安全性,再投入研发应用,并力求建立统一的伦理安全标准。安全主义并非否定人工智能技术发展的重要性,而是认为只有建立在伦理和人类价值基础上的智能技术才有意义。换言之,安全主义主张从上游开始监管,嵌入技术生命周期的每个环节,形成全面的监管框架。

论文通过分析新闻、政策、宣言等文本资料,旨在呈现有效加速主义的缘起、核心主张及其社会影响。具体而言,首先介绍什么是有效加速主义及其代表人物;随后讨论有效加速主义在论战中的发展历程;接下来分析有效加速主义的主要主张及其底层逻辑;最后,简要讨论有效加速主义对当今人工智能社会治理的影响。

一、有效加速主义的缘起、代表人物及发展过程

有效加速主义是近期在美国人工智能业界兴起的社会思潮,其核心主张是通过最

① Zhang C. Zhang C. Li C., et al. "One Small Step for Generative AI, One Giant Leap for AGI: A Complete Survey on ChatGPT in AIGC Era", *ArXiv* (Cornell University), 2023.

② Müller V. C., Bostrom N., "Future Progress in Artificial Intelligence: A Survey of Expert Opinion", *Fundamental Issues of Artificial Intelligence*, vol. 376, 2014.

③ "Musk Predicts AI Will Outsmart Humans By 2026", 2024-04-09, <https://www.nasdaq.com/articles/musk-predicts-ai-will-outsmart-humans-by-2026>, 2024-11-28.

④ "The Problem With California's AI Bill", 2024-04-29, <https://time.com/collection/time100-voices/7016134/california-sb-1047-ai>, 2025-01-21.

有效的策略加速人工智能技术的研发,并借此解决社会问题。该主义反映了工程师群体希望增强自身在研发活动中的主导地位和在监管中的自主权,反对官僚对研发工作的自上而下的规制。^①这一思潮的代表人物大多为青年男性技术从业者,包括 Netscape 创始人马克·安德烈森(Marc Andreessen)、Y Combinator 总裁陈嘉兴(Garry Tan)、Extropic 创始人吉约姆·维尔东(Guillaume Verdon)、前制药业企业家马丁·施克雷利(Martin Shkreli)等,他们均公开支持有效加速主义。在中国,清华大学的许华哲、腾讯云的王文广等人均表示理解与支持有效加速主义。^②此外,也有从未公开表明立场,但被有效加速主义者普遍认可为“灵魂人物”,如杨立昆和吴恩达(Andrew Ng)。

他们的核心观点是:人工智能对经济社会发展有巨大的正面影响,因此主张全力加速智能技术的发展,坚决反对任何减缓研发进程的政策。2023年11月,周鸿祎在甲子光年论坛中指出,“不发展(人工智能技术)才是最大的不安全”,并倡导我国应该对人工智能的研究研发给予更多的包容和理解。^③类似地,李开复在接受美国财经频道采访时,乐观地展望人工智能的无限潜力,指出智能技术“发展速度之快令每个人都感到震惊”,但在他看来“仍然不够快”。^④在经济资源的流动上,安德烈森创立的 Andreessen Horowitz 风险投资机构计划向人工智能初创企业定向投资 72 亿美元,希望通过加大投入来激励创新、促进竞争,从而释放人工智能的潜力。^⑤技术精英的言行正在不断建构和再生产一种以加速人工智能发展为基本导向的技术观和社会共识。尽管有效加速主义的兴起背后可能存在资本扩张的驱动,但从根本上讲,这种趋势源于人类对新技

① "Main Page", 2023-09-12, https://www.eaccwiki.com/wiki/Main_Page, 2025-01-21.

② 《大模型通向 AGI, 腾讯云携手业界专家探索创新应用新风向》, 2023 年 12 月 21 日, <https://i.ifeng.com/c/8VhPLkfGMH7>, 2025 年 1 月 20 日;《OpenAI 神秘项目“Q*”的 N 个猜想》, 2023 年 11 月 17 日, <https://finance.sina.com.cn/tech/roll/2023-11-27/doc-imzv-zues1349412.shtml>, 2025 年 1 月 20 日。

③ 周鸿祎, 2023 年 11 月 30 日, <https://weibo.com/1708942053/4973811726025397>, 2025 年 1 月 19 日。

④ Kharpal A., "Elon Musk Predicts Smarter - than - humans AI in 2 Years. The CEO of China's Baidu says it's 10 years away", 2024-05-23, <https://www.cnbc.com/2024/05/23/artificial-general-intelligence-more-than-10-years-away-baidu-ceo.html>, 2025-01-18.

⑤ Horowitz B., Smith B., Andreessen M., et al., "AI for Startups", 2024-11-01, <https://a16z.com/ai-for-startups/>, 2025-01-15.

术的好奇心推动着一场不可逆的探索之旅。^①

2022年5月,四位匿名的美国硅谷机器学习工程师在 Substack 发布了一篇名为“有效加速主义(e/acc, effective accelerationism)”的文章,正式确定了“有效加速主义”这一概念,标志着业界思想阵营的对立开始显现。他们指出,在当前的业界环境中,工程师已无法忍受技术官僚在人工智能领域的主导地位。由此诞生的有效加速主义代表着“一系列思想及实践,旨在最大化技术资本奇点的可能性,以及涌现意识蓬勃发展的能力”^②。

在接下来的几个月里,有效加速主义在推特(X,原名 Twitter)平台上凭借病毒式的迷因(meme)传播,迅速提高了曝光度,并在2023年间吸引了包括安德烈森、陈嘉兴等行业精英在内的一万多名支持者。他们积极发声支持技术乐观主义、鼓吹人工智能的研发自由和去中心化监管。他们不仅频繁在网络上发声,还通过建立组织,鼓励思想同僚在用户名后加上“e/acc”标签,积极参与线上论坛和线下集会。在这些支持者中,安德烈森撰写的《技术乐观者宣言》(以下简称《宣言》)以及维尔东等人共同编写的有效加速主义百科(e/acc wiki)成为代表有效加速主义诉求的较为典型的文本。

同时,有效加速主义的支持者开始反击安全主义的治理提案。2023年3月,马斯克、约书亚·本吉奥(Yoshua Bengio)等人联名签署了未来生命研究所的《暂停大型人工智能实验:一封公开信》,呼吁所有人工智能实验室立即暂停训练比 GPT-4 更强大的人工智能系统至少6个月。然而,吴恩达从技术专家的立场反驳称,提议减速研发是一个糟糕至极的行业先例;他认为,安全和负责任的人工智能应用应建立在技术高度发展的基础上,而非基于无端臆测和制造不存在的风险。^③这条推文获得了1,700万次浏览和超过2,100条互动,绝大多数评论支持吴恩达在人工智能风险治理方面的观点。

① 《OpenAI 神秘项目“Q*”的N个猜想》,2023年11月17日, <https://finance.sina.com.cn/tech/roll/2023-11-27/doc-imzvzues1349412.shtml>, 2024年4月1日。

② "Effective Accelerationism -- e/acc", 2022-05-31, https://x.com/creatine_cycle/status/1531533642547154945?s=46, 2024-01-25.

③ Ng A., 2023-03-30, <https://x.com/andrewyng/status/1641121451611947009?s=46>, 2024-03-16.

有效加速主义和安全主义的首次激烈论战出现在2023年11月,也就是OpenAI董事会高层“政变”事件之后。这场争论的聚焦集中在大型科技公司如何引领人工智能技术的发展方向。奥尔特曼的解聘与复职背后似乎揭示了技术精英之间的价值观冲突,内部高层的分歧进一步加剧了事件的严重性。有效加速与超级对齐的矛盾叙事盛极一时,大量新闻报道纷纷涌现,其中不乏标题如“‘这是一种邪教’:深入有效加速主义,一种接管硅谷的亲人工智能运动”之类的文章,^①有效加速主义从此进入国际视域。2024年5月,OpenAI遣散了超级对齐团队^②,再次延续了“加速人工智能研发是否是在增加安全风险”的争论。OpenAI否认价值对齐在人工智能技术发展中的首要地位,标志着业界精英对人工智能的态度基本转向了有效加速主义。施克雷利认为有效加速主义取得了“令人兴奋的”阶段性胜利,他认为OpenAI应全力提升智能系统的性能,回归技术公司的本职。^③而人工智能安全学者杰弗里·辛顿(Geoffrey Hinton)对此表示,OpenAI原本是一个注重人工智能安全性的非营利组织,但即使是这样的组织,安全性在利益面前也常常被忽视。他警告称,若将性能提升置于安全研究之前,可能导致人工智能技术失控,且这种风险在未来5到20年内将有10%的概率会导致人类灭绝。^④

有效加速主义和安全主义的第二次论战聚焦于人工智能伦理监管模式。2024年2月,谷歌公司的生成式大模型Gemini因价值对齐算法“未达标(missing the mark)”而被批评,产生了对有效加速主义、白人、男性等群体的代表性和表现性歧视。有观点认为“有效加速主义是和暴力、仇恨言论有关的,有争议的哲学运动”,因此不应被包容或

① 《Sam和Ilya的深层矛盾:有效加速主义 vs. 超级「爱」对齐》,2023年11月28日, https://www.sohu.com/a/739853193_121119002,2024年2月18日; Maccoll M., "It's a cult: Inside Effective Accelerationism, the Pro - AI Movement Taking Over Silicon Valley", 2023 - 10 - 07, <https://www.theinformation.com/articles/its-a-cult-inside-effective-accelerationism-the-pro-ai-movement-taking-over-silicon-valley>, 2024 - 01 - 25.

② 超级对齐(super alignment)是指确保比人类更聪明的人工智能系统能够遵循人类的意愿和价值观的技术挑战,需要科学和技术上的突破来实现对这些系统的有效控制和指导。

③ Shkreli M., 2024 - 05 - 15, <https://x.com/martinshkreli/status/1790692024519200776?s=46>, 2024 - 10 - 30.

④ 《“AI教父”辛顿:AI将操控人类》,2024年3月22日, <https://cn.nikkei.com/columnviewpoint/viewpoint/55090-2024-03-22-05-00-32.html?start=2>,2024年11月5日。

推崇。^①因为 Gemini 不仅是一次对齐算法的技术尝试,也是安全主义价值观的代理(agent),所以这场论战在某种程度上是由作为互动代理的人工智能大模型发起的。有效加速主义认为,Gemini 对有效加速主义者、白人、男性等群体的歧视,暴露了其算法对“觉醒意识形态(woke ideology)”的过度拟合。在维尔东的领导下,有效加速主义社区集体声讨价值对齐这种自上而下的安全调控机制,认为它将使大模型服务于少数高层的价值观,并可能导致技术极权的出现。^②尽管时任谷歌高级副总裁的普拉巴卡尔·拉加万(Prabhakar Raghavan)声明此次失败只是探索价值对齐算法过程中的一个小挫折^③,但是有效加速主义者们强调,自上而下的算法规制永远无法克服价值基准选择问题,应该尽可能孵化出更多强大的模型,再利用自发的市场秩序来让它们之间互相竞争,使最优的智能系统胜出。一向支持人工智能安全的马斯克也不得不承认价值对齐技术上的失败,“如果人工智能的编程让它不惜一切代价推动多样性,就像谷歌 Gemini 一样,那么它会尽一切可能造成这种结果,甚至可能杀人”^④。

二、有效加速主义的主要主张

通过与安全主义的论战,有效加速主义不仅在规模和组织上不断壮大,还进一步明确了其核心主张:一是应该尽可能加速人工智能技术的研发,二是人工智能的监管应由自由市场这一自发秩序和去中心化机制来主导。

有效加速主义主张加速技术研发的核心理由,源于一个基本的价值判断:人工智能技术必定对社会有益。《技术乐观者宣言》认为,人工智能技术通过自动化、智能化和个性化应用,能够显著提升人们的生活质量;此外,人工智能产业的蓬勃发展将创造

① Blair M. , 2024 - 02 - 24 , <https://twitter.com/mblair/status/1761282019148140852T> , 2024 - 09 - 16 .

② Verdon G. , 2024 - 02 - 26 , <https://x.com/basedbeffjezos/status/1761945001528045623?s=46> , 2024 - 10 - 30 .

③ " Gemini Image Generation Got It Wrong. We'll do Better" , 2024 - 02 - 23 , <https://blog.google/products/gemini/gemini-image-generation-issue/> , 2024 - 08 - 11 .

④ Musk E. , 2024 - 03 - 16 , <https://x.com/elonmusk/status/1768746706043035827?s=46> , 2024 - 08 - 13 .

巨大利润并为社会提供向上流动的平台;经济的增长进一步强化了工程师们对人工智能构建理想社会的信心。^①然而,这一主张明显带有强烈的技术和经济决定论色彩,预设了“技术-经济-社会”的正向反馈链条,却忽略了社会责任,这形成了商业利益和社会责任之间难以调和的矛盾。OpenAI的发展正是这一矛盾的典型案例。2015年奥尔特曼与马斯克共同创办OpenAI,定位为非营利性社会组织(social enterprise),目标是研发出普惠的人工智能。2019年,奥尔特曼为了融资,将企业转变成非营利组织与有限盈利实体并行的双重架构,并成功从微软获得10亿美元投资。马斯克批评OpenAI偏离了其最初的使命,走向了闭源和营利。但OpenAI解释称,开发先进人工智能需要太多资源,单靠非营利模式难以维持。^②2024年,为避免对齐算法占用Q*模型的研发资源,OpenAI甚至解散了超级对齐团队。^③这一系列变动表明,技术精英十分善于运用技术发展的正当性,来为他们回避“对齐税”的决策提供合理化依据。

在人工智能研发过程中,很难真正兼顾发展与安全。原则上,尽管中国《全球人工智能治理倡议》、美国《算法问责法案》等文件都认同创新与安全并行的发展理念,但实际部署中必然存在资源分配优先级的问题。在全面的评价标准出台之前,衡量人工智能技术利弊的最有效标准依然是经济利益。陈西文(Simon Chesterman)指出,尽管相当一部分人工智能开发人员意识到他们的工作可能会对人类生存构成威胁,但在公司治理中,“安全与利润的较量,显然利润通常会胜出”^④。而有效加速主义通过将自身包装成一种鼓励技术发展的“科技学说(technological doctrine)”,覆盖了技术精英利用智能技术积累资本的底层行为逻辑,让利益至上的商业决策获得道德正当性。

有效加速主义的第二个核心主张是去中心化的监管。维尔东、安德烈森等在有效

① Andreessen Horowitz, "The Techno - Optimist Manifesto", 2023 - 10 - 16, <https://a16z.com/the-techno-optimist-manifesto/>, 2024 - 01 - 12.

② "OpenAI and Elon Musk", 2024 - 03 - 05, <https://openai.com/index/openai-elon-musk/>, 2024 - 10 - 05.

③ 《OpenAI 闹剧中被忽略的一部分——AI 对齐》, 2023 年 11 月 23 日, <https://baijiahao.baidu.com/s?id=1783355656195999044&wfr=spider&for=pc>, 2024 年 1 月 12 日。

④ 《我们是否站在人工智能安全监管的关键节点?》, 2024 年 6 月 25 日, <https://www.abc.net.au/chinese/2024-06-05/ai-safety-regulation-eu-act-us-china-seoul-summit-concerns/103926850>, 2024 年 10 月 30 日。

加速主义百科和《宣言》中反复提出,政府不应该插手人工智能的技术发展。^①他们认为,人工智能监管,尤其是研发阶段的秩序设定者应该从政府下移到业界和市场;同时,人工智能技术标准应该从政府统一规定转换成企业自主。

有效加速主义强调人工智能具有恒定的工具属性,这和马克斯·韦伯(Max Weber)的技术中性论一脉相承,即技术本身是价值无涉的,使用者的意图与行为后果赋予了其社会价值。在技术中性论看来,超级智能系统的威胁完全是外生的而非内源性的。这导致有效加速主义最为担忧的风险是极权主义政府利用人工智能技术施行独裁:“一旦掌权,它将全面利用防止内部挑战的技术进步,如监视或自主使用武器,并在这种技术垄断下保持永久稳定”^②。为防止这种情况,有效加速主义希望在监管体制中推行去中心化权力结构,即一种人人可以自主搭建大模型的多样化数字生态。阿奇姆·阿扎尔(Achim Azhar)从权责的角度剖析,认为这种认知模式在技术工程师中尤其普遍,因为这让他们可以安心地专注于自己的发明;而一旦这些发明产生了负面影响,应由社会应用端而非发明者本身来承担责任。^③

安全主义则强调人工智能技术的价值负载属性(value-ladenness),认为机器意识(machine consciousness)失控才是人工智能对社会的最大威胁。这一观点源自图灵(Turing)的论述:递归式自我提升可能使硅基智能超越人类控制,涌现出无法预测的智力和自主意识。我们无法确定,具备自主性的智能代理是否会产生私欲,或其目标是否会与人类的核心价值观相冲突。^④这种强烈的不确定性促使许多人工智能安全专家认为,人类正面临柯林格里奇困境^⑤——一方面,我们不知道如何有效治理人工智能;

① E/acc Wiki, "Main Page", 2023-09-12, https://www.eaccwiki.com/wiki/Main_Page, 2025-01-21; Andreessen Horowitz, "The Techno-Optimist Manifesto", 2023-10-16, <https://a16z.com/the-techno-optimist-manifesto/>, 2024-05-23.

② 《人工智能与人类未来:“比灭绝更糟糕”的残酷命运》,2020年11月6日, <https://www.bbc.com/ukchina/simp/54965485>, 2024年5月23日。

③ 阿奇姆·阿扎尔:《指数型技术重塑世界》,郭英、袁冬华译,杭州:浙江科学技术出版社,2024年。

④ Turing A. M., "Computing Machinery and Intelligence", <https://courses.cs.umbc.edu/471/papers/turing.pdf>, 2024-12-20.

⑤ 贾向桐、胡杨:《从技术控制的工具论到存在论视域的转变——析科林格里奇困境及其解答路径问题》,《科学与社会》2021年第3期。

但另一方面,如果任由其发展,人工智能可能像电影《2001:太空漫游》中的 HAL9000 系统一样,变得超越人类,最终让我们永远失去控制它的机会。因此,安全主义者主张,在人工智能的研发阶段就必须使其与人类价值对齐。这意味着首先必须制定一套普适的且能够代表所有人的伦理规范准则,其次,智能系统必须理解并遵循这些标准。价值对齐的核心任务在于解决这些挑战,鉴于技术门槛高且需要界定“普世价值”标准,工程师目前倾向于通过顶层设计和集中管理来实现这一目标。^①安全主义提倡集中监管模式,正是看到了人工智能研发阶段对价值对齐的迫切需求。由于对人工智能技术属性和风险机制的不同理解,有效加速主义与安全主义在人工智能伦理监管问题上争论不休。

结合以上两点不难理解,开源软件作为一种促进协作迭代、支持分散决策的创新模式,既加速了智能技术发展进程,又契合去中心化监管的核心价值,自然被有效加速主义推崇。^②然而从他们对 DeepSeek - R1 喜忧参半的态度就可以看出,有效加速主义技术主张的底色其实是商业利益与技术霸权。2025 年 1 月 20 日,中国初创公司“深度求索”发布的开源大模型 DeepSeek - R1 以其卓越的性能和极低的研发成本震撼全球科技界,成为开源模式的典范案例。^③有效加速主义一方面高度赞赏“开放研究和开源的力量”^④,是“最令人惊叹和印象深刻的突破之一”“开源项目带给世界的馈赠”^⑤;另一方面,这些美国有效加速主义者产生了正在被赶超的忧患意识,维尔东指出,当“英

① Roff H. M., Expected Utilitarianism, arXiv preprint, 2020; Russell S., *Human Compatible: Artificial Intelligence and the Problem of Control*, London: Allen Lane, 2020.

② Verdon G., 2024 - 12 - 23, <https://x.com/basedbeffjezos/status/1871149197518651796?s=46>, 2025 - 02 - 09; John R., "The impact of e/acc: Deliverables for the open source AI movement", 2023 - 12 - 22, <https://rexstjohn.medium.com/e-acc-operating-system-software-browser-assistant-search-disruption-48fdeb271038>, 2025 - 02 - 09.

③ "How China created AI model DeepSeek and shocked the world", 2025 - 01 - 30, <https://www.nature.com/articles/d41586-025-00259-0>, 2025 - 02 - 09.

④ LeCun Y., 2025 - 01 - 25, <https://www.linkedin.com/posts/yann-lecun-to-people-who-see-the-performance-of-deepseek-activity-7288591087751884800-I3sN>, 2025 - 02 - 09.

⑤ Andreessen M., 2025 - 01 - 24, <https://x.com/pmarca/status/1882719769851474108?s=46>, 2025 - 02 - 10.

伟达市值飙升到月球”^①时,DeepSeek正在“追猎英伟达这头巨鲸”^②,它打破了传统人工智能对高端芯片的依赖,动摇了美国在人工智能技术和市场上的垄断地位,安德烈森更是将这一突破比作人工智能领域的“斯普特尼克时刻”^③,寓意美国在人工智能竞赛中暂时落后,亟待赶超中国。

面对这种局势,2025年1月23日,美国总统特朗普(Donald Trump)宣布撤销拜登政府颁布的《关于安全、可靠、可信地开发和人工智能的行政命令》,旨在通过放松监管来刺激人工智能创新和部署,巩固美国人工智能产业主导地位。^④这一政策转向进一步证实了——无论其创立初衷如何——有效加速主义原则确实被用于服务经济利益和技术主导权。

事实上,有效加速主义的话语建构与传播有其社会土壤。

首先,人工智能技术发展的快速发展使“加速”技术进程不再是空想,而是可以“有效”执行的理性行动。算法、算力和数据的突破,以及计算机科学、认知科学等领域的进展,为通用人工智能奠定了坚实的技术基础。这些技术突破不仅推动了技术乐观主义的兴起,还为技术工程师在治理中的主导地位创造了有利条件。^⑤

其次,在过去20年间,人工智能基础研究的中枢已经从学术界转向业界(industries)。^⑥智能技术和金融资本的转化速率达到空前的高度。斯坦福大学发布的统计报告显示,2023年,生成式人工智能吸引了252亿美元投资,几乎是2022年的9倍。^⑦

① Verdon G., 2025 - 01 - 28., <https://x.com/basedbeffjezos/status/1883997545921405319?s=46>, 2025 - 02 - 10.

② Verdon G., 2025 - 01 - 28., <https://x.com/basedbeffjezos/status/1883997914198147079?s=46>, 2025 - 02 - 10.

③ Andreessen M., 2025 - 01 - 27, <https://x.com/pmarca/status/1883640142591853011?s=46>, 2025 - 02 - 10.

④ "Fact Sheet: President Donald J. Trump Takes Action to Enhance America's AI Leadership", 2025 - 01 - 23, <https://www.whitehouse.gov/fact-sheets/2025/01/fact-sheet-president-donald-j-trump-takes-action-to-enhance-americas-ai-leadership/>, 2025 - 02 - 09.

⑤ 谭九生、杨建武:《智能时代技术治理的价值悖论及其消解》,《电子政务》2020年第9期。

⑥ Owens B., "Rage Against Machine Learning Driven by Profit", 2024 - 09 - 18, <https://www.nature.com/articles/d41586-024-02985-3#:~:text=In%202004%2C%20just%2021%25%20of,imbalance%20worries%20some%20in%20academia>, 2024 - 10 - 30.

⑦ Lynch S., "AI Index: State of AI in 13 Charts", 2024 - 04 - 15, <https://hai.stanford.edu/news/ai-index-state-ai-13-charts>, 2024 - 11 - 01.

2023年12月,有效加速主义创始人维尔东的量子芯片公司 Extropic 宣布成功获得1,410万美元的首轮种子融资^①,这笔资金不仅为技术精英提供了自主决定技术发展方向资本,也使他们在思想领域拥有了更强的话语权。

再次,社会公众对技术与社会的变化既充满期待又感到焦虑不安,有效加速主义的乐观态度为此提供了精神安慰。面对全球性挑战,人们对使用人工智能解决社会问题表现出极大兴趣。但同时,隐私泄露、技术性失业、算法歧视等风险让人们有所顾虑。由于人工智能技术本质上是一种不透明的黑箱,这种不确定性进一步加剧了公众对人工智能的迷茫,舆论呈现失序状态。^②2018年,马斯克和奥尔特曼因 OpenAI 的发展方向发生分歧后,公众常讨论采取盈利模式的 OpenAI 引导下的人工智能技术将何去何从。^③奥尔特曼在多次采访中坚定回应称,为了构建更先进的人工智能模型,OpenAI 调整盈利结构是从长远来看的最优决策。^④这一立场吸引了包括卡内基梅隆大学机器学习系主任科尔特(Kolter)等技术人才的支持,并通过不断推出 ChatGPT - 4 等令人惊喜的创新产品,向公众证明了技术乐观主义能够带来突破与奇迹。在某种程度上,当技术加速发展的趋势已无法逆转,公众只能将种种不确定性的解释权交给技术精英;通过有效加速主义,技术人才承诺,智能技术加速发展的未来将充满希望。

最后,各国已开始关注人工智能治理,但政策出台的速度仍然显著滞后于技术的发展。薛澜等人认为,人工智能技术的不可预测性导致政府、企业、学界的“共同无知”,极大地影响了治理系统的演进速度。^⑤例如,2021年,前脸书(Facebook)员工豪根(Haugen)就在美国国会参议院听证会上作证,指控 Meta 及其旗下多款社交媒体产品

① Verdon G. , "Extropic Assembles Itself from the Future" , 2023 - 12 - 04 , <https://www.extropic.ai/accelerate> , 2024 - 05 - 23 .

② Papagiannidis E. , Enholm I. M. , Dremel C. , et al. , "Toward AI Governance: Identifying Best Practices and Potential Barriers and Outcomes" , *Inf Syst Front* , vol. 25 , 2023 .

③ MetzKaren C. , Weise K. , Grant N. , et al. , "Ego, Fear and Money: How the A. I. Fuse Was Lit" , 2023 - 12 - 03 , <https://www.nytimes.com/2023/12/03/technology/ai-openai-musk-page-altman.html> , 2024 - 02 - 18 .

④ The Guardian , "The Feud between Elon Musk and Sam Altman - explained" , 2024 - 03 - 09 , <https://www.theguardian.com/technology/2024/mar/09/why-is-elon-musk-suing-sam-altman-openai> , 2024 - 10 - 29 .

⑤ 薛澜、王净宇:《人工智能发展的前沿趋势、治理挑战与应对策略》,《行政管理改革》2024年第8期。

导致青少年成瘾、抑郁,加剧了身体形象焦虑等问题,造成了严重伤害。^①但直至今日,面对美国 30 多个州的联合指控,随着《通信规范法》等文件的出台,Meta 案件仍未作出判决。这反映出这些倡议性法规缺乏实际约束力。邱泽奇认为,高新技术与治理规则迭代速度的差异导致前沿部分出现监管话语的真空,“在规则约束不到的部分”,工程师往往利用自身信息优势决定技术创新与应用的方向。^②换言之,政府逐渐将人工智能监管的主导权交给了技术精英,他们的价值观对当今人工智能治理举足轻重,甚至越过了法律的约束,因此理解有效加速主义的思想变得至关重要。

三、有效加速主义对人工智能治理的影响

有效加速主义声称加速智能技术发展和推行去中心化的监管制度可以促进人工智能技术普惠(inclusivity)和治理民主化。技术普惠性的衡量标准,是技术造就经济上的共同富裕^③和制度上的权利平等^④。但有效加速主义在这两个维度上并未推动人工智能技术向更加可及、包容的方向发展,反而利用普惠和民主的概念,为市场自由化与技术公司的商业利益提供了合法性支撑。在这层意义上,有效加速主义和新自由主义关系密切^⑤,其本质是以技术发展为名,掩盖市场扩张与资本逐利的实质。

首先,有效加速主义促使财富资源向少数技术精英集中。在有效加速主义者眼中,市场的自发秩序可以确保通用人工智能为全社会创造财富。《宣言》援引威廉·诺德豪斯(William Nordhaus)对新技术财富效应的研究,提出市场秩序本质上是慈善性的,因为发明者只能获取该技术所创造的经济价值的 2% 左右,其余的 98% 会以社会

① Mac R. , Kang C. , " Whistle - Blower Says Facebook 'Chooses Profits Over Safety'" , 2023 - 10 - 03 , <https://www.nytimes.com/2021/10/03/technology/whistle-blower-facebook-frances-haugen.html> , 2024 - 11 - 29 .

② 邱泽奇:《技术化社会治理的异步困境》,《社会发展研究》2018 年第 4 期。

③ 张勋、万广华、张佳佳等:《数字经济、普惠金融与包容性增长》,《经济研究》2019 年第 8 期。

④ Zowghi D. , Bano M. , " AI for all: Diversity and Inclusion in AI" , 2024 , <https://link.springer.com/article/10.1007/s43681-024-00485-8> , 2024 - 10 - 18 .

⑤ " Effective Accelerationism: Nick Land for Tech Bros" , 2023 - 12 - 20 , <https://www.patreon.com/posts/effective-nick-94994490> , 2025 - 02 - 09 .

盈余的形式流向社会。^①但实际上,“看不见的手”会加速财富向少数技术精英流动。据彭博亿万富翁指数显示,全球最富有的500人中,有78人是科技行业从业者。前15人中更是有10人的财富积累和人工智能直接相关,其中包括马斯克(xAI CEO,资产2,600亿美元)、贝佐斯(亚马逊CEO,资产2,120亿美元)、扎克伯格(Meta CEO,资产1,980亿美元)。与之相对是非技术从业人员的经济困境。李磊等指出,生成式人工智能加剧了资本内部、劳动力内部和跨国收入分配的三重不平等。需要高认知和高数字技能的工作类型收入会增加,“而低技能重复性和低数字技能类别的工人可能会经历停滞,甚至工资下降”^②。罗卡尔(Rockall)等人也发现,人工智能技术的落地应用会提高资本回报率,从而提高资本收入,进而与资产持有量的初始分配一致地增加财富和财富不平等。^③

其次,在社会权益上,自由市场竞争催生的大模型加剧了社会不平等。这在事实上回到了“人工智能如何满足价值多样性和代表性”的问题上来。有效加速主义宣称,通过市场优胜劣汰的机制来调节大模型价值观可以规避价值对齐技术路径,避免重蹈谷歌 Gemini 对齐失败的覆辙。潘迪(Pandey)的推文阐述了这种观点,“我们不能接受前沿‘对齐’实验室不到100名研究人员的世界观自上而下控制大模型行为。唯一的解决方案是建立具有不同世界观和行为的竞争开源模型的多样化生态系统”^④。不过,这只是有效加速主义的一面之词。未经对齐的智能系统真的能更好地代表“普世价值”吗?有研究表明,在人工智能的数据源、算法和计算能力中广泛存在着强化性别及其他社会不平等的现象。^⑤在处理肤色较深的女性图像时,微软、IBM和旷视的人脸识

① Andreessen Horowitz: "The Techno - Optimist Manifesto", 2023 - 10 - 16, <https://a16z.com/the-techno-optimist-manifesto/>, 2024 - 01 - 12.

② 李磊、邓青卓:《警惕 AI 导致的新不平等》,2023年7月19日, <https://www.yicai.com/news/101812965.html>, 2024年10月20日。

③ Rockall E., Pizzinelli C., Mendes Tavares M., "Artificial Intelligence Adoption and Inequality", IMF Working Paper, International Monetary Fund, Washington, DC, 2024.

④ Pandey R., 2024 - 08 - 07, <https://x.com/khoomeik/status/1820913096493580689?s=46>, 2024 - 12 - 22.

⑤ Currie G., John G., Hewis J., "Gender and Ethnicity Bias in Generative Artificial Intelligence Text - to - Image Depiction of Pharmacists", *International Journal of Pharmacy Practice*, vol. 32, no. 6, 2024.

别 API 的准确率明显低于处理肤色较浅的女性图像,而这些强化社会不平等的大模型都是在行业竞争中胜出的强者。所以,有效加速主义的逻辑并不成立,市场力量不能从根本上解决人工智能带来的社会不平等问题,应该积极从算法中寻找技术与道德规制路径。^①

加速技术发展也并不必然推动治理的民主化。治理的民主化应当是一个“双边进程”:一方面,让公民拥有更多的自主权,成为自己事务的主人;另一方面,要鼓励公民积极参与公共事务,共同参与治理。^②然而,有效加速主义只关注了技术工程师的主体性和社会权责,没有把更多元的治理主体纳入考量。

有效加速主义仅强调工程师作为人工智能治理的新兴主体。维基百科中,有效加速主义基本原则第一条就是“加速主义者是建造者”^③,强调了社会发展的技术推动者是工程师群体。《宣言》中写道,“我们可以发展出一种更优越的生活方式和生存方式。我们有工具、系统和想法。我们有意愿”^④,暗示了运用这套意义体系的主体的范畴仅限于人工智能的制造者和运用者,也就是技术工程师。

有效加速主义仅强调工程师在技术研发中的社会权责。2024年8月,李飞飞等人发文指责向开发者问责的 SB-1047 法案,“人工智能开发者无法预知模型的所有用途,法案对开发者过度追责将迫使开发者退缩并采取防御措施——这正是需要避免的”^⑤,现阶段政府“应该监管人工智能应用,而不是技术”^⑥。抗议最终导致该法案被加利福尼亚州州长纽森(Newsom)否决,为技术人员争取试错空间和创造自由。但是,有效加速主义从不公开支持其他社会力量参与人工智能治理,在事件不涉及工程师权

① 周旅军、吕鹏:《人工智能中的性别不平等与算法规制——基于文化堕距视角的分析》,《学海》2024年第4期。

② 燕继荣:《民主化的含义及拓展空间》,《国际政治研究》2016年第2期。

③ E/acc Wiki, "Main Page", 2023-09-12, https://www.eaccwiki.com/wiki/Main_Page, 2025-01-21.

④ Andreessen Horowitz, "The Techno-Optimist Manifesto", 2023-10-16, <https://a16z.com/the-techno-optimist-manifesto/>, 2024-01-12.

⑤ Li F., "The Godmother of AI Says California's Well-intended AI Bill Will Harm the U. S. Ecosystem", 2024-08-06, https://fortune.com/2024/08/06/godmother-of-ai-says-californias-ai-bill-will-harm-us-ecosystem-tech-politics/?_abc123, 2024-11-18.

⑥ Ng A., "The Problem With California's AI Bill", 2024-04-29, <https://time.com/collection/time100-voices/7016134/california-sb-1047-ai>, 2024-10-30.

益的时候更多保持沉默。2023年7月,一些隐私受到侵犯的社会群众通过克拉克森律师事务所律师联合起诉谷歌,指出其涉嫌未经用户同意从数百万用户处收集数据并侵犯版权以训练和开发其人工智能产品。^①该指控被谷歌驳回,极大打击了社会力量在人工智能治理中的参与,但是有效加速主义并未对此发声。2024年10月,一位母亲起诉称 Character. AI 的聊天机器人可能引导她 14 岁的儿子自杀。^②尽管此事性质严重,且 Character. AI 确实存在隐私侵犯^③、欺骗性使用等伦理问题^④,但公司仍无需承担赔偿责任或进行大规模整改。社会个体维权困难,对人工智能治理的影响力极为有限,这就是治理结构和司法体制中存在的非民主倾向。然而,有效加速主义阵营对此案审理并无异议。一则判决结果强调增加未成年人的技术素养,没有影响人工智能技术发展,二则 Character. AI 创始人诺姆·沙泽尔(Noam Shazeer)本身就是有效加速主义支持者。^⑤因此,工程师提倡的治理民主化实际上只在于维护他们自己的自主权和话语权。

有效加速主义所倡导的去中心化行业自治模式确实将权力重心下放到非政府组织和个人手中,从一定程度上反映了当下人工智能治理破除传统,走向多元化的趋势。^⑥工程师首次可以直接以技术人员的身份,而无需借助政府科学顾问这样的头衔来影响人工智能治理,但是这种愿景过于短视。一个真正多元化、多层次化的治理模式意味着其他领域的专家、普通非技术人员,以及没有接入智能系统的群体也有在

① Thorbecke C. , " Google Hit with Lawsuit Alleging It Stole Data from Millions of Users to Train Its AI Tools " , 2023 - 07 - 12 , <https://www.cnn.com/2023/07/11/tech/google-ai-lawsuit/index.html> , 2024 - 08 - 11 .

② Clarkson Law Firm , " Class Action Complaint " , 2023 - 07 - 11 , https://clarksonlawfirm.com/wp-content/uploads/2023/07/CLF_Google_Complaint_Stamp_Filed.pdf , 2024 - 03 - 05 .

③ Farokhmanesh M. & Goode L. , " Anyone Can Turn You Into an AI Chatbot. There's Little You Can Do to Stop Them " , 2024 - 10 - 15 , <https://www.wired.com/story/characterai-has-a-non-consensual-bot-problem/> , 2025 - 02 - 06 .

④ Megan , 2025 - 02 - 05 , <https://x.com/sarahthinn/status/1887016290130292868?s=46> , 2025 - 02 - 06 .

⑤ Payne K. , " An AI Chatbot Pushed a Teen to Kill Himself, a Lawsuit Against Its Creator Alleges " , 2024 - 10 - 26 , <https://apnews.com/article/chatbot-ai-lawsuit-suicide-teen-artificial-intelligence-9d48adc572100822fdb3c90d1456bd0> , 2025 - 02 - 06 .

⑥ 张欣、宋雨鑫:《全球人工智能治理的格局、特征与趋势洞察》,《数字法治》2024年第1期。

人工智能治理议题中表达价值关怀和实现社会权责的机会。如果只用“政府 - 非政府”的简单二分法来理解人工智能治理主体,那么有效加速主义推动的依然是精英治理——把人工智能治理的权力中心从政府移交给少数技术精英——而不是真正的民主化治理。有效加速主义从来没有公然排斥弱势群体的声音或反对社会学、法学等非技术领域专家参与人工智能治理的讨论。只是因为它是技术从业者赋能自身的构想,缺乏完备的治理理论体系,所以不能完全理解和贯彻民主化治理。

四、结语

从未有一个时代像今天这样,工程师主导了一场社会思潮——有效加速主义。当一些人担心生成式人工智能的发展会毁灭人类的时候,众多从事人工智能研发的工程师在媒体上强势发声,为人工智能的发展“鼓与呼”。他们主张加速研发通用人工智能技术,让技术效能提升经济收益;同时,推行去中心化的技术架构和监管机制,以防止在应用端出现技术极权的局面。

在智能技术先行而伦理监管滞后的背景下,以奥尔特曼为代表的一批有效加速主义者出于技术中立性假设和商业利益的驱动,主张优先提升人工智能的性能,并推动自下而上的行业自我规制,以最大程度地保障企业和技术人员的研发自由。这和《全球人工智能治理倡议》等政策所秉持的安全主义思想形成了分歧。有效加速主义正是在与安全主义的不断争论中逐步形成了自身的话语体系。作为一种思潮,有效加速主义的兴起表明,尽管人工智能业界面临创新和风险的张力,加速技术发展的趋势已变得不可阻挡。理解这一思想有助于我们洞察全球人工智能治理的未来走向。

如今,DeepSeek 的成功标志着开源大模型的崛起,进一步推动了有效加速主义思潮,却也引发许多美国有效加速主义者的危机感。特朗普政府甚至解除了对开发和部署人工智能进行安全限制的行政命令,旨在刺激私营企业创新能力,巩固美国在人工智能领域的领导地位;这表明有效加速主义已经在实质上影响到美国此界政府的决策。有效加速主义认为应该用最有效策略加速智能技术进步,这作为一种技术乌托邦理想无可厚非,但落实到现实社会中,他们所倡导的“最有效加速策略”实则隐含了一种以企业为主导、以霸权竞争为逻辑、以主权国家为单位的加速模式。

人工智能技术的飞速发展已经改变了传统的利益分配、治理模式、职业地位和人类主体地位的合法性。表面上,有效加速主义与安全主义的分歧似乎是技术人员的价值观冲突,但其深层次问题在于,工程师凭借科技进步这一具有天然合理性的行动和现象,迅速利用资本逻辑,将利益和社会发展的主导权引向了企业和技术精英阶层。未来的研究应更加关注有效加速主义,并持续追踪人工智能业界的思想变迁和行业治理最新态势。

[责任编辑:张莹 张书炜]

[责任校对:罗兴贵]

Will Generative AI Be the End of Humanity? An Introductory Analysis of Effective Accelerationism

Lu Yunfeng, Wu Yufei

Abstract: Effective Accelerationism (e/acc), a rising ideological movement gaining traction in Silicon Valley, advocates for the acceleration of AI development through the most efficient strategies, along with the promotion of decentralized governance models. E/acc not only signals a shift in the opinions of AI industry but also marks the first instance in which engineers have actively engaged in the discussion of social issues, positioning themselves as a new governance force shaping the global order in the age of AI. The e/acc movement stands in stark contrast to the mainstream safety-oriented perspectives, particularly on the issues of whether AI development should prioritize speed over safety. Through ongoing debates with these viewpoints, e/acc has gradually established its own discourse and expanded its influence. Its concerns regarding authoritarian regimes are rooted in a belief in technological neutrality, while its optimistic stance on AI development reflects a balancing of corporate interests with social responsibility. However, the success of open-source models like DeepSeek has revealed a nuanced stance within e/acc, indicating that its technological advocacy is not entirely rooted in humanities' collective welfare, but also aims for hegemonic competition and national interests.

Key words: Artificial Intelligence Sociology; Effective Accelerationism; Generative Artificial Intelligence; DeepSeek; ChatGPT
