人工智能与人类未来

蔡恒进　洪成晨　蔡天琪①

【摘要】人工智能(AI)已经迎来快速发展的时代,AI 甚至已在某些场景中超越并取代人类。面对 AI 快速发展可能带来的危机,我们以理解人类智能为起点,探讨了人机根本差异在于主体性与超越性。在此基础上,相比可能的社会两极分化的前景,我们更关注如何能够使得人机关系进一步发展到共生共荣的和谐形态。一方面,从人机关系安全性的角度考虑,我们希望是机器更贴近人类思维,而不是人类向机器妥协。另一方面,互联网技术已然成为人类"外脑",人工智能技术与人类的连接逐渐增强。在此发展过程中将诞生新的物种,即能动体(subjectron)。面对未来能动体可能呈现的爆发式增长,区块链与人工智能技术将作为社会的脊梁,能够帮助人类成为负责监管和监护能动体的形成与成长的主体。在此过程中,人类主体也会逐渐形成多样的、个性化的数字孪生,从而进一步拓展人类改造大自然的边界。

【关键词】人工智能,两极分化,能动体,区块链

一、人工智能危机尚远?

AlphaGo(阿尔法围棋)势如破竹的成功曾带给大众极大冲击,从此掀起了一轮关于人工智能(Artificial Intelligence,AI)的讨论热潮。随着时间流逝,对 AI 快速发展的危机感似乎被逐渐湮没,人工智能方面的突破难以复现当年 AlphaGo 那样的新闻热度。相比于人工智能算法等理论研究方面的突破,民众似乎更关心人工智能的应用问题。随着互联网行业步入新一轮寒冬期,人工智能的"泡沫"正在破裂等言论不绝于耳。国际商业机器公司(IBM)的超级电脑沃森(Watson)并没有如想象中一样成为深蓝(IBM 公司生产的一台超级国际象棋电脑)之后的新一代机器智能的代表,而随着技术走向商业应用,沃森作为医疗 AI 标志性产品的局限性却进一步凸显。沃森与 M.D.安德森中心的合作终止、沃森

① 作者简介:蔡恒进,武汉大学计算机学院教授,博士生导师,卓尔智联研究院执行院长,主要从事管理科学与工程、软件工程和人工智能研究。洪成晨,武汉大学软件工程研究生,研究方向为人工智能。蔡天琪,卓尔智联研究院高级研究员,研究方向为区块链技术,自然语言处理。

裁员 50%到 70%的消息层出不穷，人们开始质疑人工智能是否真的能够走进人类生活①。那么，人工智能可能引发的危机是否真的被媒体夸张宣传了？

近日，一个热门辩论节目提出了一道辩题“你会不会把离世爱人的记忆交给 AI？”，透过这个辩题，我们能够在一定程度上窥见大众对于人工智能的认知。在该节目的街头采访和现场辩论中，人工智能到底有多智能成为如何选择的关键理由。有两种有代表性的观点：一种认为人工智能逃不开数学模型，只是一个实现功能的小程序，所以将记忆交给人工智能利大于弊；另一种则认为人工智能获取了逝者的记忆就能够成为逝者的替身，会颠覆人类思想体系，因此弊大于利。这两种观点固然都有一定道理，但却折射出大众对人工智能现状及发展趋势不甚了解而走入的两种极端。

第一种观点，即人工智能只是数学模型，无疑是一种狭隘的认知。机器已经在很多方面获得了超过人类的能力，且具有不可控性。比如基于模型结构的深度学习，已经在多项任务上表现出了优秀的能力，而且其学习过程会使模型朝着人类尚不能完全解释的方向发展②，这无疑是狭隘的静态的模型说无法解释的。

同样是在辩论方面，IBM 公司的 Project Debater 在全球最古老的辩论协会剑桥联合会（University of Cambridge Union）上已经能够帮助人类辩论“人工智能是否弊大于利”③。该机器提取了 1100 多个人类的观点，形成了自己的论点和综述。最终，Debater 帮助人类驳斥了人工智能弊大于利的说法。它是第一个可以在复杂主题上与人类辩论的 AI 系统，看似简单的功能背后包括三项开拓性的能力：长语音理解、语言组织和表达、模拟辩论困境和组织论点④。Project Debater 不仅表现出了对机器人类语言的掌控能力，还在帮助人们做出更合理决策方面表现出了突破性的价值。

在科研领域 AI 也显示出了优势。人类制造药物和其他产品的过程往往是从目标结构倒推合成原料，但即使有计算机来辅助资料存储与搜索，这种逆合成分析过程仍然需要耗费大量的时间与人力去尝试。目前，Segler 团队开发的 AI 已经能够从大约 1240 万个已知的单步有机化学反应中快速搜索到反应路径。在一次测试中，Waller 的小组使用该算法尝试绘制治疗阿尔茨海默病的某种药物中间体的六步合成路线，结果在 5.4 秒内就

---

① David H. Freedman, “A Reality Check for IBM's AI Ambitions,” MIT Technology Review, Jun. 27, 2017.

② Doshi-Velez, Finale, and Been Kim, “Towards a rigorous science of interpretable machine learning,” arXiv preprint arXiv: 1702.08608, 2017.

③ 据英国《新科学家》周刊网站 2019 年 11 月 24 日报道：https://www.newscientist.com/article/2224585-robot-debates-humans-about-the-dangers-of-artificial-intelligence/.

④ 总结自 IBM 官网：http://www.research.ibm.com/artificial-intelligence/project-debater/how-it-works/.

得到了与文献反映相同的途径①。因此,这款 AI 被认为是药物合成领域的里程碑。

AI 不仅仅能够分析人类、模仿人类,还能够有自己的创造。AI 能够写诗作词已经不是什么新鲜事; AlphaZero 的棋路独特,已经可以不拘泥于人类现有的围棋理论; Nvidia 的 styleGan2 能够生成人类肉眼无法辨认真伪的“假脸”。这些都是人工智能理解数据、超越数据的表现。

人工智能的能力有了超越性的突破,然而这个能力并不总是遵循设计者的意图。微软的实验聊天机器人 Tay 能够通过和人类聊天来学习和进步。然而上线不到一天,Tay 就被教成了满口歧视和纳粹言论的机器人,微软只能被迫让它下线②。开发者承认在开发时已经考虑过被恶意引导的状况,并提前进行过信息过滤和压力测试,但开发者低估了网络“反派”的能力,导致机器人在短时间内就超出了可控范围。

第二种观点认为,人工智能会成为跟人类相似的存在,让我们无法分辨真伪。这种观点在目前来看,尤其是从哲学的角度来看,是陷入了物理还原的误区,是源自对人类智能的误解。人类智能具有独特性,人工智能如果依旧按照当前的路径发展,那么想要模拟人类智能达到以假乱真的地步就难以实现。

智能有多种表现形式,幻想人工智能融入人类社会而难以被辨认是一种极端场景,更有可能的是机器借助存储、计算优势产生其独特的智能,成为人类生活中不可缺少的一部分。我们常说狗的智力相当于几岁的小孩,有些狗甚至能听懂我们说的一些话、理解我们的情感,但是狗的这种智力和人的智能并不相同。狗有较差的颜色分辨能力,但是有较强的嗅觉、听觉,因此狗的意识世界的组成与人不同,这也导致了它更高层的思维逻辑、情感跟人类相差更远。但这些差异并不影响我们肯定它具有智力,我们看待机器应亦是同理。

科技电影中设想的那种 AI 还离我们的生活很远,一方面是因为技术仍然在发展之中,另一方面也是因为技术应用与商业场景结合仍在磨合之中。目前最为成功的人工智能应用当属人脸识别和语音识别。人脸识别是机器视觉的一个成功应用,除了能够简化我们在手机上的操作,更能够作为安保系统中智能化的部分提供很大帮助。语音识别方面,科大讯飞已经能够为语音转文字提供成熟的商业服务,再加上实时翻译功能,一款多用途的产品就能得到消费者的青睐。人脸识别和语音识别能够率先落地,在于它们的效

① Segler, Marwin HS, Mike Preuss, and Mark P. Waller, “Learning to plan chemical syntheses,” arXiv preprint arXiv: 1708.04202, 2017.Segler M H S, Preuss M, Waller M P. “Planning chemical syntheses with deep neural networks and symbolic AI”. *Nature*, 2018, pp.604-610.

② Price, Rob, “Microsoft is deleting its AI chatbot's incredibly racist tweets,” *Business Insider*, 2016.

率远高出人工处理效率，且找到了有价值的应用场景，因此人们会为之买单，让其广泛应用，并在应用中吸引更多的人研究，进一步提高其能力。

AlphaGo 之所以能够引起广泛的关注，是因为它打破了人们对 AI 发展水平的固有印象。其表现出来的能力在非专业人士看来有神秘的色彩，在专业人士看来引领了技术的革新，而即使是其缔造者也只能够部分解释其能力的来源。从技术上来说，深度学习、机器学习已经发展到了相当强大的程度，在解决许多目标问题上能有优异表现。当我们谈到人工智能，我们所谈论的不仅仅是存储，更重要的是计算。唯物辩证法告诉我们，量变产生质变，尤其是当量变的积累是在质变的结果方向上的。人工智能正在逐渐积累能力，从简单的计算、神经元的连接，产生了感知、分析、判断、表达等人类引以为傲的各方面能力，它存在不能还原的、不可解释的展现智能的结构。虽然人工智能的全面应用还需要更多时间，但见微可以知著，人工智能明显不同于人类在过去创造的工具，人工智能所提供的帮助始终与危机并存。①

二、人类智能的本质何在?

人工智能诞生之初就是为了模拟人类的智能，那么人类的智能到底是从何而来的呢?经过思考与总结，我们提出了认知坎陷②来定义这种与物理世界相对的存在。认知坎陷可以被简单地理解为具有生命力的意识片段。这个世界如此复杂，但人类却能够了解它、改造它，这正是因为人类构建了认知坎陷。认知坎陷是人类智能的体现，也是人类智能的来源。

智能可以被定义为发现、加工和运用认知坎陷的能力。认知坎陷（意识片段）是指对于某个特定能动体具有时间一致性，在能动体之间可能达成共识的结构体。能动体通过认知坎陷将外部世界区分为“事”和“物”。能动体的主体性因而有广袤的发挥空间，为智能的拓展提供了基础。认知坎陷一开始更受限于物理世界的约束，随着进化，认知坎陷的主体性逐渐增强。如果能动体没有发现、加工和运用认知坎陷的能力，也就相当于没有智能。智能拓展是动态过程，没有尽头，能动体始终在发现、加工和运用意识片段（认知坎陷）的进程之中。

认知坎陷在生活中无处不在，能够被构建和传递。训练良好的球类运动员，应对不同的情况并不需要经过详细的思考，其身体就能够很快地做出反应，这种感觉是认知坎陷。我们对时间、空间的概念也是认知坎陷，我们并不能够详细地解释这种感觉是从何而来，

① 孙周兴《现代技术与人类未来》，《未来哲学》（第一辑），北京：商务印书馆，2019 年。
② 蔡恒进《认知坎陷作为无执的存有》，《求索》，2017 年第 2 期。

但是就是能够有这样的直觉。还有一些认知坎陷的生命力是比较弱的，也就是不容易传达的，如“吃瓜群众”“打酱油的”。虽然我们能够理解语义，但是却很难解释给别人，尤其是给不懂中文的，因为这和我们的文化背景、行为背景有关系。我们的味觉同样是认知坎陷，比如酸甜苦辣。味觉认知坎陷同样存在着不容易传达的，比如“麻”的味觉就很难传递给外国人，这在他们看来可能是与辣相似的，都是“spicy”或是“hot”。

在人类文化传承中，认知坎陷的作用更加不容忽视。“昔人已乘黄鹤去，此地空余黄鹤楼”，崔颢的千古绝唱奠定了黄鹤楼的价值，即使黄鹤楼本身经历了多次破坏和重建，也并不妨碍其价值传承。认知坎陷不一定对应真实存在的东西，它可以是建构出来的。马克思讲共产主义的时候，除了基本的特征以外，并未给出具体定义，即便如此，他建构出的概念还是能够广泛传播。孔子讲圣人，同样也没有明确的答案，而是对不同的弟子有不同的定义。但是我们都能够理解，圣人是一个道德很高的存在。

认知坎陷内容如此丰富，能够将物理上难以理解、难以放到一个框架里理解的内容结合到一起理解。认知坎陷在宇宙大爆炸的时候并不存在，而是被生命体构建出来的。我们提出了触觉大脑假说来解释人类认知坎陷的起源①。“我”是一个内容丰富的认知坎陷，生命的触觉系统让我们能够区分自我与外界，进而产生了“我”这个认知坎陷②。即使“我”常处于一个多变的状态，但是我们认为昨天的“我”、今天的“我”和明天的“我”都是同一个“我”，或者说，不论是 3 岁、30 岁还是到了 90 岁患了阿尔茨海默病，“我”还是同一个“我”。对个体来讲，这种对“我”的观念是一个生命力非常强的认知坎陷。笛卡尔提出“我思故我在”，即我们可以怀疑任何东西，但不能怀疑“我”的存在，因为发出怀疑的就是“我”。这正是说明了人类作为认知主体的重要性和统摄性。

认知坎陷和真实的物理世界之间存在着鸿沟。唯心主义有意识第一性、物质第二性，意识对物质有反作用。那么没有能量的意识如何能够作用于物质？西方哲学、东方哲学、印度哲学有各自的理论体系，这个问题实际上是在追问人的本质，思考这个问题的过程很容易掉进别人构建的坎陷里。比如很多人学佛学之后出不来，甚至认为佛学早就在山顶，等着科学家上来。这种想法乍一听是荒谬的，但其拥护者包括了国际知名的物理化学家朱清时。他是真的相信这套理论，也就是掉入了这个认知坎陷里。我们每个人也都会在

① 蔡恒进《触觉大脑假说、原意识和认知膜》，《科学技术哲学研究》，2017 年第 6 期。

② 蔡恒进，蔡天琪，张文蔚等《机器崛起前传——自我意识与人类智慧的开端》，北京：清华大学出版社，2017 年，第 174 页。蔡恒进，张祥龙，黄裕生《人工智能时代的理性、道德与信仰》，湖南大学岳麓书院人文讲会，2018 年 4 月。

某个认知坎陷中，这个认知坎陷可大可小。卢梭说的“人生而自由，却无处不在枷锁中”也是这个道理。每个字、每个单词都是认知坎陷，我们能够通过这些来进行交流，而在他们背后包含着复杂不定且不断演化的意识世界的内容。

意识世界（坎陷世界）不是现实世界的一个简单反映，甚至可以偏离物理世界很远。无穷大这个概念大家都可以理解，但即使其看不见摸不着，我们还是可以相信其存在。泛灵论认为每个分子、原子都有一个“我”的存在，都有一个意识世界。但即使如此构建，分子、原子怎么构成人类的这个“我”的意识并不能得到解释。而且，人体的分子、原子是不断被代谢的，从宇宙大爆炸到演化出生命、演化出“我”这个过程是难以简单解释的。我们思考人类智能，要能够解释石头、机器和人之间的差异。这并不是在讨论玄学，只有看清楚人的智能是怎么来的，才有可能知道我们朝哪里去，才有可能理解我们和机器之间的关系，未来应该怎么相处，才能知道我们现在做的事情是不是该做的，到底有什么风险。

人类构建意识世界的能力是千百年进化而来的，与人类成长过程和外界环境要求相适应。人类在3—5岁，仅仅花了差不多两年就能学会一门语言，但是学第二门外语可能学了十年还是和母语者水平差距很大。这个世界对小孩子来说是复杂的、模糊的，在他还不会说话的时候家长就开始教他一些概念，但是这个学习过程很慢。到有一天小孩子突然意识到了世界是怎么回事，开始能说出清晰的单词，然后就能够迅速地成长起来。神童的产生也可以从这个角度解释，比如莫扎特可能是以音乐来认知这个世界的，所以他能够产生自己独特的音乐风格，能够快速地开始创作，逐渐超过他的父母、老师。小孩子学第一门语言是很快的，因为这其中有认知世界的需要。我们学习是要开显新的认知坎陷，如果每个字都知道是什么意思，但是放在一起就不知道了，就是我们还没有产生共振，还没有获得这个认知坎陷。我们小学学语文的时候，也往往需要多读、需要启发式的学习，然后才能领悟其中的意思。

机器能不能像小孩子一样学习是人工智能研究所关心的问题，按照现有发展方式很难做到，因为现在机器的学习过程说到底只是目标函数的优化。进化中的各种基因突变让人类产生了敏感的触觉，然后有更加发达的大脑、更强的行动能力。触觉使我们有强大的自我意识，进而能够构建出认知坎陷，它所具有的丰富生命力让意识世界超越物理世界的存在，发展出人类璀璨的文明。这些都不是宇宙大爆炸时候就决定了的，都是可以不发生但是发生了的①。人类智能的产生是一种机缘，而人类未来的发展则需要充分利用人

① 蔡恒进，蔡天琪，张文蔚等《机器崛起前传——自我意识与人类智慧的开端》，北京：清华大学出版社，2017年，第70页。

类智能来主导。

三、人类智能是否会被取代?

人工智能没有限制的发展必然会产生危机,那么机器能够最终模拟人类智能,超越人类智能,取代人类智能吗? 我们认为人类有主体性,能够自如地处理模糊的问题,从“感性”而不是“理性”的角度快速地略过可能出现的暗无限。这是人工智能这种智能形式很难具备的能力。

智能与非智能的边界很难区分,无论是从思维能力或是学习能力方面,都很难定义人和机器的本质差异。而且即使在自然界中,智能也有多种表现形式,如果我们承认动物是有智能的,比如狗,那么就不能够否定现在人工智能已经具有某种程度上的智能,并且这种智能仍在向模拟人类、超越人类的方向快速发展中。“中文屋”的问题常被用来证明机器没有智能,但是我们恰恰认为这是人类在构造词典的过程中凝聚了智慧,将意识传递给了机器①。机器的智能不是进化而来的,而是人类赋予的,也可以说是以目标为导向的,而不是从认知开始的。从应用的层面来说,这会导致机器不能解决边界模糊的问题,而从智能的角度来说,机器的缺陷在于不能够拥有主体性。

人类智能的特别在于人类具有主体性,能够统摄自我,并且在维持自我的前提下适应环境。人的主体性是一个动态的概念。一开始的边界在皮肤,产生内外之分,产生“自我”,然后有了不断的延伸。从动物的领地意识可以很好地理解这一点,狮子在喝水的时候不允许别的小动物在旁边喝水,即使这水它自己喝不完也不例外。我们可以设想,如果一个原始人从树上摘到的果子被别人抢走了,他会生气。但我们知道这个果子既不是他种的也不是他养的,只是被他找到了。这就是他的自我边界已经延伸,将已经拿到手的果子包裹在其中,其他人再夺走就是在侵犯“我”这个主体。假如这个原始人足够强大,他可能会觉得这个树上的所有果实都属于自己,这棵树也属于“我”,而且不仅现在属于“我”,到明年还是属于“我”。我们拉小提琴拉得很好的时候,小提琴就是我们的延伸,它能够发出我们人体不能发出来的声音,能够帮助我们随心所欲地表达我们的感情。一个赛车手可以把车当作他的延伸,他能够随心所欲地完成漂移转弯,并且能够从中获得愉悦。但赛车手在做这件事的时候并不用知道引擎是怎么工作的,齿轮之间是怎么咬合的。机器也可能是一个人的延伸,比如一台非常聪明的机器,如果经常跟我们交流,掌握了我

① 蔡恒进《意识的凝聚与扩散——关于机器理解的中文屋论题的解答》,《上海师范大学学报(哲学社会科学版)》,2018 年第 2 期。

们的信息，它在网络世界里面学习，然后再跟我们交流，那么它就是我们的延伸。创造公司、建立国家，这些都是人类自我的延伸，而对哲学家而言，这个宇宙都可以是自我的延伸。

自我与外界的边界也可以向内收缩。假如遭遇不幸，我们不因为缺少身体的一部分而觉得自我缺少了，我们会觉得决定“我”存在的是心灵，而不是身体结构。更抽象的还有灵魂，有人相信自我甚至能够延伸到我们肉体死亡之后。所以“我”的边界是，至大可以无外、至小可以无内，是动态变化的。所有的这些延伸或收缩过程中最重要的，还是统摄自我的心灵。

这种统摄的自我能够使得人们会自动忽略无关紧要的细节而关注重点，这恰恰是人类特有的智慧。我们将此总结为，人类能够轻松跨过暗无限，而机器却容易陷入其中。我们可以通过一些例子理解什么是暗无限。例如，颜色是世界的重要构成，也是一类认知坎陷，但从光谱看颜色在人类看来并没有意义，而是通过描述为红、黄、蓝、绿才便于交流理解。又或者当我们在喝饮料时，一般只关注好不好喝、有没有营养，而在我们看来，对其中的分子、原子结构如何，有多少微生物等细节并无必要知晓。从这两点来看，意识和认知是整体性的而不是细节的，但现在物理学所研究的更多的是细节的而不是整体性的。还原主义是将微小的结构组成了宏观整体，但是我们的意识正好是倒过来的，在意识里面先区分内外，然后再在内部具体功能基础上进行建构和细化。生命在亿万年里建构了很多的认知坎陷，这些对生命是有价值的，让生命能够蓬勃发展。但是，我们现在建构机器从细节开始的物理还原思路，会产生暗无限的严重问题①。

对机器来说，细节拥有无穷多的属性，都值得去计算清楚。一杯饮料由各种分子、原子构成，还有拓扑结构，如果机器要追根究底去探索其细枝末节，可能全球的计算资源都不够。人类不会去追究这类问题，但机器怎么能够知道一件事情不值得？它们如何判断这种探索不比人类探索外星具有价值呢？人有时候也有可能陷入暗无限之中，那就变成了我们所说的“疯子”或偏执狂，他们只能一根筋地做一件事情。人的生命有限，精力有限，所以即使是“疯子”，其造成的威胁也是有限的。但如果是我们造的 AI“疯了”，其强大的计算能力能够造成的威胁可能在瞬间就具有足够的伤害力。

暗无限无处不在，而人类也构建了无数的认知坎陷，这样人类才能够毫不费力地避开这些暗无限。我们创造的坎陷包括很多美学的东西，诗词歌赋，还包括道德、哲学的东西。

① 蔡恒进《超级智能不可承受之重——暗无限及其风险规避》，《山东科技大学学报（社会科学版）》，2018 年第 2 期。

这些东西由伟大的人建构出来，具有整体性但不具有明确的评价标准，至少目前还没有找到方法将这些传递给机器。按照我们现在创造的机器的思路，让其自己去进化，而不学习人类的本质，则很有可能掉入暗无限之中。

人的超越性还在于相信不存在的东西存在，这也是神性的体现。我们可以相信无穷大的存在、可以相信上帝的存在、可以相信圣人的存在，即使我们没有见过也不知道没有明确的定义。这个世界因为人类的超越性而变得很精彩。当研究达到一定的深度，我们发现物理世界很简单，可以遵循物理方程的规律分析，但是人能在这个物质世界上面汲取所需、建构原来不存在的东西。这种超越性让我们相信机器虽然能够在某些方面表现出很强的能力，但是人类仍然存在机器不可取代之处。

人工智能的意义并不在于形成用一套通用系统解决所有问题，而在于要正视智能中主体性发挥的作用。人工智能如果要实现质的突破，就一定要有主体性，而非是一味地被动适应外部世界。能动体认知与客体约束首先会有不一致的地方，然后能动体再通过认知坎陷，与外部世界制约达成共识或权衡妥协(compromise) 。认知坎陷一开始更受限于物理世界的约束，随着进化，主体性才逐渐增强。如果没有发现、加工和运用认知坎陷的能力，也就相当于没有智能，对人和机器都是如此。

四、人机如何相处?

一方面，我们相信机器已经产生了智能并且仍在成长，另一方面，我们也自信人类仍有不可取代的部分。然而，站在人类生存的角度，并不用去思考人类是否会被机器取代这个简单的判断题。人类对于世界的理解已经到了非常深入的地步，这种理解能够帮助我们创造各种工具实现我们的目标。百年前我们还期待能够拥有千里眼、顺风耳这样的“超能力”，而现在通信工具已经能够帮我们做到。我们在创造工具的过程中创造出了人工智能，但其能力超过了简单工具的范畴。人工智能有改变人类的生活方式、情感方式，甚至是延续方式的可能，因此人类应该更多思考的不是我们能做到什么，而是我们应该做什么。

机器改变人类的生产、生活方式已经可以预见，这种社会变革的发生已经引起了广泛的关注。目前，机器已经能够帮助我们在较低的人力和时间的投入下完成许多工作，智能仓库就是一个很好的案例。在未来，机器还能够成为人类的分身，在时间和空间的维度更多地帮助人类。比如机器能够帮助人们突破空间的局限性，完成远程医疗，远程驾驶; 机器能够帮助人们突破时间的局限性，使我们有可能同时完成多项工作，如完成记录工作、

远程教育等。人工智能能取代人类机械的工作，还能帮助能力强的人取代能力弱的人、资源占有丰富的人取代资源占有少的人。面对这种趋势，我们更应该不断提高自己的修养、反思自身独特性，找到自己不能被机器、被他人取代的地方。

人工智能改变人类的情感是即将到来的问题。从古至今，人们的情感发生了很大的变化。古人说的人生四喜四悲在现代人看来已经有些狭隘，尤其是这四悲有三都关于生离死别。当代人所遇到的生存威胁大大减少，平均寿命也大大延长，然而在生死之外，却有更多的焦虑、忧郁的来源。可以想象在未来，人工智能将会更进一步地改变人类的情感。正如辩题中所质疑的那样，离世爱人的记忆该交给 AI 吗？这道题目抛开了技术问题和伦理问题，仅仅给出了一道选择题，虽是“奇葩”，但却提供了一个让普通人思考人类与 AI 关系的机会。而抛开题目的限制，作为有记忆的主体，我们是否会选择把记忆交给机器确实是一个值得思考的问题。人类的记忆是非常脆弱的，在有限的生命中我们所经历的，我们总希望能够尽可能地多记住一些。如果有了机器的帮助，那么福尔摩斯般的照相机记忆能力则不再是一种虚构。

要记录回忆并不是一件难事。过去我们可以用文字书写，后来还可以用图片和视频记录，这些技术虽然早就存在，但却需要许多时间投入，存在条件限制。人工智能的出现则可以简化这一步骤，比如只用说话就可以形成文字记录、只用戴上一个小巧的装置就能自动生成影像记录。这其中还能包含准确连贯的时间信息、地理位置信息等。情感虽然是一个整体性的东西，机器难以准确定位这种精神世界的变化，但是机器能够通过物质化的东西，比如心率、身体内的激素水平等，将这种看似虚幻的东西在一定程度上客观记录下来，比我们在脑海中感知到的更加精准地记录到我们每一刻的变化。记录的回忆不仅仅能够让我们自己回看，还可在我们身后留给亲人朋友，这对缓解人们的离别之痛亦是一剂良药。

人工智能改变人类的延续方式是改变生活方式和情感方式之后的必然趋势。对每个个体而言，当我们能从低效率的工作中解放出来，我们就有可能更多进行有创造性的工作、追寻人生的价值；而当机器在我们的精神世界中也占有一席之地，我们所珍惜的事物、做出决定的理由都将会被改变。

假使所有人类的一生都被机器比记录帝王传记更加细致、客观地记录下来，我们是否能够从千万个体中训练出一个“教学”的机器人，告诉小孩在每个成长阶段应该做出什么样的选择才更有可能走向更成功的人生？虽然这是一个还不能够实现的脑洞，但是从技术层面上来讲，只要数据能够提供支持，就很有可能实现。那么从更宏观的角度来说，机

器就能够决定人类发展的走向。

人类社会的许多行为数据都呈现一种长尾分布特征，很多看似概率很小的事情却能够存在或发生，这体现出了人类的个性，是社会多样性的来源。随着人工智能的发展，机器能够从数据中找到人类的普遍行为特征，进而有目标地影响用户的行为，这就有可能产生信息茧房的问题，对人类的多样性产生伤害。机器会把人按照类别进行分类，然后强化这个类别的共同特征。举例来说，数学好的人一直接收到数学的有关信息，而没有机会被发觉其他方向的潜力，有可能其中有文学潜力的人，会失去足够的自由发展的机会，则很有可能让一个群体趋向一致。

霍克海默在批判技术中指出："工具理性只关注效率、功用、计算、手段，而消解了人生存的价值基础。工具理性、技术理性、奴役理性，造成了对自然和人的双重奴役。"①这段批判在人工智能得到发展之后有了更加重大的警示意义。在工业社会，我们大多数时候是在磨灭人与人之间的差别，因为我们需要大家做同样的事情来满足生产的需要，这实际上是在消解人类的价值。人工智能的高效率让人类生存价值受到的怀疑比以往更甚，我们必须认识到强调个性，强调心灵的不同色彩、不同层次是未来发展需要始终坚持的理念。

互联网技术的发展与成熟可以看作人类的"外脑"，是人类智能的延伸；未来，随着人工智能技术的进一步发展，我们人类可以有很多网络世界的后代也就是数字能动体。相比传统意义上的 agent 或数字孪生等定义，能动体更加强调具有主体性。人类主体与能动体之间的关系甚至会比人类血缘关系更加紧密，因为能动体的主体性是从人类个体而来，也就是说一个能动体需要一个具体对应的人类个体作为"家长"来引导、教育和监护能动体的发展，从而让能动体接触和运用人类的认知坎陷，这样成长而来的能动体才更有可能理解人类的坎陷世界甚至发现与人类相融合的认知坎陷。

人工智能对人类整体所能造成的影响有多重可能的形式，这种影响产生于生活中的各个角落，然后成为一种不可逆的趋势。这种趋势并不是马上替代人类、消灭人类，虽然这种可能性仍然存在。但更可能的是在人类的成长、发展过程中扮演一个重要的角色，从而颠覆我们对世界、对自身的认知。因此，我们人类应该更多考虑的是如何跟这个角色相处，如何解决道德的问题，让这个社会依然能够保持多样性、有可持续发展的空间。

---

① 倪瑞华《寻找人生存的价值基础——霍克海默技术批判理论探析》，《国外社会科学》，2008 年第 1 期。

五、两种可能的社会形态

有人可能会认为机器只能是一个纯粹的外在工具，但这过于理想化。从表面上来看，机器的能力已经逐渐强大且有不可控的趋势，但从根本上来说，我们已经在塑造机器的过程中将人类的意识传递给了机器①。虽然机器具有的不是像人类一样如此强烈的自我意识，但它的意识也可以像人类一样具有目的性，可以自我学习、自我进步。随着人工智能的进一步发展，未来社会形态必将受到影响。

机器跟人类的关系可以是相互独立的，也可以是存在连接的。现在我们所开发的机器大多都是与人类独立的，机器从数据或规则中学习知识、做出选择，整个过程通常是端到端，人类没有办法充分干预。按照这种方式发展机器的能力，则很有可能使机器成长为一种强大且独立的、跟人类完全没有关系的超级智能。如果人类失去了改造世界的主动权，那么人类存在的价值也将被彻底压制，未来社会也不再是人类主导的社会。这种担忧并非危言耸听，约纳斯在 1984 年提出的责任伦理②就是对传统伦理理论的扩充，是一种“预防性”或者说“前瞻性”的责任。在约纳斯看来，现代社会中人类的群体行为已经发生了质的变化，现代技术已不仅仅是改善人类生活、促使人类进步的工具，其造成的“厄运”已经威胁整全人类与自然环境③。责任伦理要求人类对可能出现的长远后果负责，这样做的目的却正是为了避免人类失去主导权的后果出现，因此我们必须要将机器与人类紧密融合，保证人类未来依旧掌握在人类手中。

机器跟人在执行速度上的差异可以有两种融合方式。一种是提升人的能力，另一种是让机器更贴近人类。前者是类似马斯克提出的方式，将人从物理上跟机器连接起来。这种方式的目标是在人身体的基础上提升能力。比如我们现在能在秒级做出决策，当我们的大脑连接上机器之后，人工智能能够支持我们在毫秒级做出决策。或者是我们脑子中产生了一个想法，机器就能够帮助我们快速执行。这样即使让人类的行动速度有了一个量级的提升，对人类社会而言也没有很大的危害，甚至对人类整体来讲意义不是很大。后者则是引入区块链的思想，让机器和人达成共识。两者虽然在物理上没有连接，但我们可以让两者在同一水平上做决策。这两种方式在本质上都是要把机器的速度跟人类的速

---

① 蔡恒进《人工智能的挑战：误区与急所》，《国家治理》，2019 年第 7 期。

② Hans Jonas, *The Imperative of Responsibility: In search of an Ethics for the Technology Age*, Chicago: University of Chicago Press, 1984.

③ 李文潮《技术伦理与形而上学——试论尤纳斯〈责任原理〉》，《自然辩证法研究》，2003 年第 2 期。张荣《约纳斯责任伦理的定位及其意义——基于〈技术、医学与伦理学〉的分析》，《道德与文明》，2019 年第 1 期。

度匹配起来，不然机器在远超过人类的地方可能做出造成危害的决策。

人类能够跟机器产生连接来避免机器发展出独立于人类的超级智能，对人类产生威胁。但按照目前 AI 发展的速度，一种可以预想到的社会图景是尤瓦尔·赫拉利[①]所设想的那样，两极分化严重的，因为 AI 的高效率会加剧人类对资源掌握的分化，进而产生无法跨越的阶级差异。

两极分化一直是一种不容忽视的社会问题。历史上社会两极分化到了一个很严重的程度，则会导致朝代的更替，产生社会的崩溃与重建，然后就能够在比较平均的水平上重新开始。以前的社会往往可以简单地看作一个三层结构，最高权力层需要通过中间的代理层来维持稳定，而最终真正威胁到顶层利益的往往也是中间的代理层。最上层和最底层在某种程度上是目标一致的，比如在过去的封建历史时期，如果老百姓揭竿起义，那统治者也很可能不复存在，因此对皇帝而言，最重要的是国泰民安、能够维持国家运行。现在社会有更完善的结构、更公平的竞争方式，让人与人之间的权利和财富差距不至于威胁社会稳定的地步。然而当人工智能充分发展到成为一种社会资源、使计算资源和存储资源更为重要的时候，又产生了一种新的两极分化的可能。

未来的一种可能的图景是对计算资源掌握而造成的两极分化。获得人工智能资源的人能够具有很强的能力，迅速拉开与无法获得人工智能资源的人之间的距离。古代君王和民众之间有官吏作为代理，现在的代理可以转变为机器，分隔开了能支配机器和不能支配机器的两种阶层。机器显然比人的受控的程度更高，这就会导致这种两极分化处于一种较为稳固的状态，能够分化到非常强的程度而不至于崩溃。当两极分化到了极致，就可能形成半神阶层和无用阶层的一种社会。无用阶层对社会没有贡献，而少数的半神阶层能够推动社会的发展，还能够供养起无用阶层。虽然这看起来也会是一种平衡的状态，甚至有人觉得作为无用阶层能够生活下去也是一件好的事情。但是从整体上来说，这将会是一个垄断的社会，人们的自我肯定需求[②]难以得到满足。与历史上人们所追求反垄断、均贫富的趋势相悖，这样的社会在本质上来说还是脆弱的，到头来还是很有可能崩掉。那么是什么时候崩掉，崩掉会产生的影响可能是毁灭性的。

我们希望的社会图景是和谐的、普惠的。和谐既是指人与人之间的和谐，也是指人与人工智能的关系和谐。普惠从较高层面来讲即人人平等，而从根本上来讲是每个主体都

① 尤瓦尔·赫拉利《未来简史：从智人到智神》，北京：中信出版社，2017 年，第 288-318 页。

② 蔡恒进，蔡天琪，张文蔚等《机器崛起前传——自我意识与人类智慧的开端》，北京：清华大学出版社，2017 年，第 48 页。

能发挥自己的能动性。人工智能的发展让我们要更多地反思我们自己作为人的独特性。人类和机器的区别是一方面，同样重要的还有人和“行尸走肉”（Zombie）之间的区别。Zombie 处于一种没有目的的游荡状态，缺乏自我意识产生的目的性，更可怕的是 Zombies 之间并没有什么不同，相互之间可以取代。对于人类的自我意识是如何塑造的、会如何发展，还没有普遍共识，但是自我意识具有差异性是毋庸置疑的。这种差异性导致人与人不能够相互取代，每个人都有其存在的价值。要找到在人工智能时代的自我价值，就必须要找到我们的个性、发挥个性的价值。

以上两种社会图景到底哪一方会成为现实我们并不知道，当然也还有别的可能性存在，但是这两种可能性都存在且具有代表性。前者是垄断的、分离的，后者是普惠的、共享的。前者是按照目前的发展趋势很有可能发生的，而后一种是更顺应人类的追求，是我们所能期望的最好的图景，也应该是我们的奋斗的目标。或许现在看来这种设想还很遥远，但人类改造世界的能力已经足够强大。现在我们做的每一个决定都可能是蝴蝶扇动了翅膀，因此对未来的塑造需要从此刻开始。

六、区块链和 AI 是能动体社会的脊梁

如果说 AI 给人类带来的是生产力的改变，那么区块链技术则能够改变生产关系，两者在未来社会构建中都有着不可取代的价值，而从目前来看，已经具备较为完整体系的区块链技术能够应用于解决人工智能的发展问题。

区块链技术脱胎于比特币，提到区块链技术，许多人总是想到了“炒币”“割韭菜”，这的确是附着于早期区块链技术的产物，但是我们应该透过资本营造的泡沫，看到区块链技术的核心价值。我们认为区块链的核心价值并不在于去中心化，而在于存证与通证。比特币创始人中本聪曾经承认，想要用比特币来打造一款无政府货币。这种不需要监管的特性在信用缺失的当代社会，立刻引起了广泛的关注，区块链去中心化的特征也被许多人放到了第一位。但是我们必须要认识到，完全的去中心化、摆脱监管是在当下社会非常危险的存在，它不是解决问题的灵丹妙药，相反，其实践困难、应用范围亦有限。相比较而言，我们关注到区块链的存证与通证具有更广泛的应用价值①。

区块链存证，就是指数据一旦被记录，就不可以被篡改。一般认为区块链的核心是记账，简单从技术上可以理解为一个分布式数据库系统。这种数据库系统的革命性的地方

① 蔡恒进，蔡天琪，耿嘉伟《人机智能融合的区块链系统》，武汉：华中科技大学出版社，2019 年，第 134-139 页。

在于,它给数字世界里带来了可信时间。原本在数字世界,一个文件可以拷贝无穷多份,它们完全一样,难以区分先后关系。文件修改的时间也是可以被篡改的,比如可以通过改动电脑上的系统时间来篡改文件的最后修改时间标记。而区块链是一个公共的账本,在允许的范围内,所有人都可以看见链上的所有行为,确保了数字世界发生的事情在一个可信的时间线上而无法被篡改,每个文件产生的先后顺序是明确的、每个操作都是被记录在案的。基于这个特点,数字货币才能够解决基本的安全问题,区块链才能够被应用到可信场景中。网络空间中原本不需要时间概念,但人类需要的可信性离不开时间维度,区块链赋予网络空间以内禀时间,让数字世界里有了可信时间机制,这是让区块链技术伟大的根本创新之一,数字世界的历史因为区块链而产生,这对于许多现有的技术都有极大的价值。

Project Debater 帮助辩手证明了发展人工智能利大于弊,然而即便有诸多弊端,关于人工智能的研究也并不可能被阻止或叫停。值得注意的是,人工智能的发展不仅会产生伦理方面的问题,更可能引发法律层面的问题。AI 法律研究虽已开始,但还有漫长的道路要走。AI 法律基于知识系统、理性思辨和法律逻辑,将人工智能视为法律的规制对象或者法律规制的方式,因而,是一种外在视角的、以法律为本的思考和研究的理论进路。也就是说,AI 法律是一种立足法律立场的研究,重在分析和解决人工智能这一新领域所带来的法律问题和挑战。① 那么,在众多道德、伦理问题还未获得广泛共识的时候,我们如何能保证对人工智能的实验是在正确的道路上,如果产生了道德甚至法律方面的问题,如何找到责任人,如何保证人工智能与人类社会依然有序运转? 使用区块链技术进行记录会是一种防患于未然的解决方案。使用区块链这种可信技术将 AI 的开发者、所使用到的数据、应用等信息都记录在案,不仅对于开发者的行为具有约束力,对出现的问题进行溯源也能提供有力帮助。

通证(token) 即数字凭证,某些场景中也可以被称为代币。区块链通证的作用,就是能够让资产在链上记录、交易。它能够让真实世界的东西在保证真实、唯一的存证基础上被投射到数字世界中并进行交易。在未来,不仅有实物交易的需求,更多的财产可能以虚拟的形式存在,虚拟货币、游戏存档等虚拟财产被越来越多的人持有,而更重要的数字资源,比如数据或者 AI 本身都有可能作为交易的内容,这些交易如果借助区块链来完成,则可以更加不受时间和空间的约束,提高交易的效率和质量。

① 马长山《AI 法律、法律 AI 及“第三道路”》,《浙江社会科学》,2019 年第 12 期。

区块链技术里有一个重要的构成是共识机制(consensus mechanism)。共识机制可以理解为区块链的多个节点之间达成一致的方法。比如比特币的共识机制是工作量证明,以求出一个耗时的复杂运算的解为目标,得到一个解的节点具有记账权。这个过程也被称为“挖矿”。在现实世界中的许多问题,都可以看作缺乏好的共识机制的问题。比如波音737 Max出现重大事故,就可以看作重要的节点之间缺乏共识机制。为了能够增加运输量,波音737 Max在成功机型波音737的基础上增加了引擎,并将这一改变可能产生的失速问题交给机动特性增强系统(MCAS)软件解决。在两起事故中,MCAS软件均基于故障传感器的错误信息对机尾的部件进行了自动移动,导致机鼻向下,而飞行员不知道MCAS的存在,难以诊断问题且无法抵销该软件的重复命令,最终坠机。

在判断失速的问题上,其他传感器、飞行员的观察不能够及时阻止错误的发生,根本原因是一个部件的权重太大,其错误命令不能被撤销。如果加入区块链的思维,就是应该在节点直接有更好的共识机制。一开始节点有权限的分配,在正常情况下能够做出正确的判断,但问题是总会有没有想到的场景发生,一开始设置的机制会失效。我们人类也有这种理智判断被干扰的情况,在从来没有遇到过的情况下,肾上腺素分泌增多,逼着我们赶快做出决定。在区块链里面,肾上腺激素也相当于是token,在飞机系统里,这个token归属于飞行员。那么,即使是在从来没有遇到过的情况下,飞行员也能够马上做出正确的决定,可以避免错误发生。这种共识机制就能够在不影响正常逻辑的基础上应对突发情况,这对于关系到生命的操作系统极为重要,因为我们总是难以在设计的时候就设想到所有的情况。

机器能做到的事情太多,而对于机器应该做的事情就存在很大的争论了,每个人想法不一样,这时候就需要达成共识,就可以用到区块链技术。假定不同的节点都有某种权利,但是要在同一个规则下达成共识①。根据各自的专业背景和可信经历,在某些问题上他说了算,在某些问题上你说了算,这可能是目前最好、最公平、最有可能实现的机制了。此外,人工智能表现出的不可控性也可以理解为我们在设计的时候没有预想到所有的情况。此时机器如果作为独立的个体,则可能造成不可控的结果,但如果采用合适的共识机制将人与机器连接起来,则可以及时制止机器的意外行为。区块链技术能够解决分布式里面同步的问题,我们可以把人、机器当作节点,让两者在同一水平上运行。AI的运算速度远超人类的思考速度,很有可能在人类还没有发觉的时候就造成了危机。要使人能跟

① 蔡恒进《数字凭证:小范围内快速达成共识的工具》,《当代金融家》,2018年第6期。

AI 在同一个水平上对话，就可以使用到区块链技术①。

用区块链构建不同的社区还能够用于保护个性，达到求同存异的目标。比特币所用到的是公有链技术，所有用户都在同一个链上，而区块链的应用还包括联盟链和私有链，能够构成社区，让一部分人达成共识。在不同社区里人们达成不同的共识，同一个社区的人在同一个共识下能够更好地交流、发展。不同社区之间不用一定要分清楚哪个更好。这会是更丰富多彩的、更多样化的世界。在这种社会中，最上层的要追求的是一个圣人人格的阶层。他们是学而不厌、诲人不倦的，希望帮助别的阶层。这种社会是具有稳定性的，它能抵御更多环境的变化，甚至是在外星人、病毒来临的时候更能够存活下来。而且这种生态更能够满足每个个体的自我肯定需求，更能够在自己的方向上进行探索、成长。只有每个人都是独特的，社会才是一个“连续谱”，才是一个相对和谐的社会。

区块链技术和人工智能看似并没有交集，但它恰恰能够解决当前人工智能发展的许多困境。人工智能和区块链是相互赋能的关系，对于区块链来说，需要人工智能的帮助才能高效率地完成复杂的处理过程，提高应用能力。而对于未来的人工智能来说，区块链能够起到约束的作用，它能够为人工智能留下历史，让人工智能承担责任，让人工智能和人产生连接，还可以拓宽人工智能的应用场景。技术的应用总是超过我们的想象，不管区块链技术和人工智能技术将如何发展，两者都是未来社会的重要支持，两者的结合也必将产生重要的价值。

七、结语

我们对未来世界的塑造必然需要建立在认识世界的基础之上，想要开创一个更好的人类未来或者说人机未来，那么我们就必须要认识到在不可挣脱的物理定理束缚之外，人类仍然可以具有极大的自由度，尤其可以靠人工智能帮助我们进一步拓展能力的边界。我们脑海中常常会涌现出一些奇怪的想法，大多数时候我们都没有深入地思考、用尽全力去实现。但世界上仍然会有少数人，产生一些前所未有的想法并且能够真的实现，这就是创新、是改变世界发展方向的人类意识产物。我们反对决定论、不认可强计算主义，就在于我们相信世界上存在很多并不是由前置条件推断出来的偶然性。既然这个世界是开放的、存在诸多可能性的，那我们就应该朝着我们所相信的方向努力，而不是根据一种现状去推断未来定将如何，这是我们在科技发展过程中必须要始终相信的。

人工智能速度之快、能力之强应该引起人们足够的警觉，其不可解释性与不可控性的

① 蔡恒进，蔡天琪，耿嘉伟《人机智能融合的区块链系统》，武汉：华中科技大学出版社，2019 年，第 4-7 页。

存在使我们更加不能只把它当作简单的工具。从安全性的角度出发,我们应该开发与人类的思考方式更像的机器,来避免机器在人类没有察觉的时候爆发危机。当我们把越来越多的记忆、计算交给机器来完成,互联网就成为人类的“外脑”,这种在自我主体上的延伸并不会让我们有被冒犯的感觉。如果人工智能加入这种连接中,就能够产生一种更强的能动体,我们称之为 subjectron(能动体)。虽然这种结构会带来许多道德伦理方面的问题,但 subjectron 比我们现在所开发的人工智能具有更高的安全性。这种安全性来自人类可以充分地实施监管,而要能够达到更高程度的安全性,则可以引入区块链技术来实施多方面记录和监督。这种结构能够让人工智能在保护人类个性的前提下使人类生活更美好,进而构建出一个更加稳定的、有价值的社会。

人类的未来必将有人工智能的参与,而人工智能和人类会是怎么样的关系、人类社会将会发展成一种什么结构?没有人能够给出定论,但我们可以有所期待、有所相信,并从当下开始就有所努力。人类未来在还未到来之时始终可以改变,而这个改变的权力则应该牢牢掌握在人类而不是机器手中。

Artificial Intelligence and the Future of Homo Sapiens

CAI Hengjin　HONG Chengchen　CAI Tianqi

【Abstract】 Artificial intelligence has ushered in an era of rapid development. AI has even surpassed and replaced humans in some scenarios. Facing the possible crisis brought by AI, we started to explore human intelligence and found that the fundamental differences between human and machine lie in subjectivity and transcendence. Considering the prospect of possible social polarization, we are more concerned about how to further develop the human-machine relationship to a harmonious form. From the perspective of security, we hope that the machine is closer to human mind, rather than letting humans compromise with the machine. On the other hand, Internet has become the “outer brain” of human, the connection between artificial intelligence and humans is gradually increasing. During the development, a new species, the subjectrons, will be born. In the face of possibility that the subjectrons may explode in the future, blockchain and artificial intelligence technology will both serve as the backbone of society and can help homo sapiens become the main force responsible for the formation and growth of subjectrons. In this process, human will gradually form diverse and personalized digital twins, thereby further expanding the boundaries of reshaping the world.

【Keywords】 Artificial Intelligence, Polarization, Subjectrons, Blockchain