·科学技术与社会·

道德增强的困境与愿景

The Plight and Prospect of Moral Enhancement

周境林 /ZHOU Jinglin[1, 2]　　姜天晓 /JIANG Tianxiao[3]

（1. 复旦大学哲学学院，上海，200433；2. 慕尼黑大学哲学、科学哲学与宗教研究系，德国慕尼黑，80539；
3. 慕尼黑大学哈登医院，德国慕尼黑，81377）
(1. School of Philosophy, Fudan University, Shanghai, 220433; 2. Faculty of Philosophy, Philosophy of Science and Religious Studies, Ludwig Maximilian University of Munich, Munich, 80539, Germany;
3. Klinikum Großhadern, Ludwig Maximilian University of Munich, Munich, 81377, Germany)

摘　要：随着科学技术的发展，人类文明面临着资源枯竭、气候恶化、大规模杀伤性武器所带来的生存危机。因此，许多生命伦理学家积极探求使用生物医学技术来增强人类道德，以克服科学发展附带的伦理挑战。生物道德增强可分为道德动机增强与道德认知增强。然而，无论哪种道德增强技术设想都引发了技术与道德方面的忧虑。文章将逐一回顾用于各类道德增强的技术手段，论证它们的不足之处，并结合当前人工智能的进展简短地分析道德增强的潜在方向。

关键词：道德动机　道德认知　人工智能　生存危机

Abstract: As science and technology advance, human civilization is confronting existential crises such as the depletion of resources, climate change, and the use of weapons of mass destruction. As a result, many bioethicists are proactively exploring the utilization of biomedical technology to improve human morality and overcome ethical challenges associated with scientific progress. Moral bioenhancement can be categorized into two types: moral-motivational enhancement and moral-cognitive enhancement. However, every proposed method of moral bioenhancement has raised both technical and ethical concerns. This article sequentially reviews the two categories of technical means for moral bioenhancement, illustrate their shortcomings, and briefly analyze the potential direction of moral enhancement with an eye to the current progress in artificial intelligence.

Key Words: Moral motivation; Moral cognition; Artificially intelligence; Existential crisis

中图分类号：B82-052; R318　　**文献标识码：**A　　DOI: 10.15994/j.1000–0763.2024.05.012

我们今日的世界面临着严峻的生存危机，核子与生物恐怖主义、资源枯竭与气候恶化使我们不得不寻求自我拯救的技术手段。近几十年药理学、神经科学与基因编辑技术的快速发展使一些生命伦理学家在黑暗中看见一缕曙光。于是，从二十年前开始，基于生物医学技术的

基金项目：国家社科基金重大项目“当代新兴增强技术前沿的人文主义哲学研究”（项目编号：20&ZD0451）；国家自然科学基金委项目“全球视野下我国科研伦理主要议题与战略应对”（项目编号：LL2224015）；中国科学院学部科技伦理研究项目“数字技术的伦理研究”。
收稿日期：2023 年 4 月 11 日
作者简介：周境林（1993–）男，广东汕头人，复旦大学哲学学院博士后、慕尼黑大学哲学系博士生，研究方向为人工智能伦理学、道德认识论。Email: jinglin_zhou@fudan.edu.cn
姜天晓（1994–）男，辽宁大连人，慕尼黑大学哈登医院博士研究生，研究方向为普外科。Email: tianxiao.jiang@med.uni-muenchen.de

道德增强（moral bioenhancement）成为了各大生命伦理学杂志的热点议题。关于该议题的论文卷宗浩繁，涉及支持与反对道德增强的理由、增强的技术手段、应该施行强制性亦或自愿性增强、道德增强在特定场景的运用等等，无法以一文蔽之，因此在本文中，笔者仅批判性回顾各种关于道德增强技术手段的提议，其他类别的文章仅在与此议题相关时才会提及。本文首先梳理道德增强的理论历史背景，然后依次评议两类增强手段——道德动机增强与道德认知增强。在揭示生物医学道德增强的困境之后，笔者将简要探讨近期人工智能的发展为道德增强提供的机遇。

一、道德增强的理论背景

二十世纪以来，人类科技水平经历了空前的发展。科学上的进步使得人类可以将自己的影响力拓展到全球范围，并延续到久远的未来。科技进步在为地球上不断增长的人口提供前所未有的福祉的同时，也带来了前所未有的伦理挑战。过度的消费主义导致了资源的衰竭与气候的恶化。并且，随着科学知识的增长，一小撮极端恐怖分子也许能够掌握核子或生物武器技术，从而给无辜的众生带来最终伤害（Ultimate Harm）。[1] 许多学者认为，这些伦理问题若得不到妥善的解决，将会严重降低人类未来世代的福祉，甚至会威胁到人类物种自身的存续。[1], [2]

然而，人类当前道德心理的缺陷使我们无法克服这种困境。虽然人类当前生活在拥有先进科技的大型社群之中，但我们的道德心理却是为了适应于小而紧密且技术原始的社群而产生的。这种演化上的不匹配（mismatch）在几种心理倾向或偏见中得到了体现。例如，人类在时间和空间上是“短见”的：我们更关心在不远的未来将会发生在我们自己与亲朋好友身上的事情。[1] 相反，我们无法适当地回应遥远、未知个体和更大的集体的苦难，无法对未来世代、自然环境与动物做出恰当的道德回应。我们有排外主义的心理倾向，如自私、裙带关系、仇外或小集团主义。这种倾向使我们无法达成大规模协作以处理恐怖主义与气候恶化的危机。

道德心理上的短见导致科技所带来的伦理挑战显得急迫且严峻，以至于一些学者认为传统的道德驯化方法——如通过家长监管和社会机构的教育与社会化——不足以确保人类性格的改善与人类物种的存续。在这种背景下，英格马尔·佩尔松（Ingmar Persson）和朱利安·萨维列斯库（Julian Savulescu）以及托马斯·道格拉斯（Thomas Douglas）在2008年分别发表了文章来论证通过生物医学技术增强人类道德能力。[1], [2] 自此之后，西方学界对于道德生物增强的辩论逐渐升温。迄今为止，生命伦理学家们已经提出了各色基于生物医学技术的道德增强手段，但他们的提议也遭受其他学者的强烈反对。

我国是世界上最大的现代化国家，科学技术发展为人民带来巨大福祉的同时，也蕴含着严峻的伦理挑战。由于我国用几十年完成西方国家长达数世纪的工业化进程，科技进步所带来的伦理困境在我国表现得更加尖锐，因此笔者希望通过述评当前学界对生物道德增强手段的研究，来探索是否能从中习得走出困境的手段。

二、道德动机增强

对于道德，人们往往有各尽不同的定义。但生物道德增强的支持者认为有一个相对中立且被许多人共享的观点：“我们道德倾向的核心首先由利他主义倾向构成，即同情其他生物，因他们自身之故而希望他们的生活变得更好而不是更糟”。[3] 此外，以“以牙还牙”为基础的公正或公平感是我们另外一种核心道德倾向。生物道德增强支持者认为这两种倾向是“是世界各地人类社会的不同道德规范的共通之处”，[1] 激励着我们按照共享的基本道德原则行事。[4] 由于这些核心道德倾向（利他主义和公正感）是我们的生物学特征，生物增强支持者主张我们可以通过生物医学手段——如药物或基因工程——来增强我们的利他主义倾向；或使我们变得更加公正或公平，即调适某些“以牙还牙”的情感，使我们更加适当地表达感激、愤怒与

宽恕等情绪。如此一来，我们将会有充足的道德动机去克服意志薄弱，以忠实地履行我们的道德义务。

催产素与选择性血清素再摄取抑制剂（SSRIs）是生物增强支持者们经常提及的可用于道德动机增强的药物。催产素和血清素是对人类具有道德效应的神经递质。催产素通常被称为“亲密激素”，因为它调节了母性抚养、伴侣结合以及信任、同情和慷慨等其他亲社会倾向。通过喷鼻剂摄入催产素已被证明能在简单的合作游戏中增加信任行为。另一方面，血清素参与情绪调节，并协助调控诸如进食、睡眠和性行为等活动。选择性血清素再摄取抑制剂（SSRIs）——如西酞普兰类抗抑郁药——常用于治疗抑郁症、焦虑症和强迫症，已被证明能使受试者更加公平及更具合作意愿。而色氨酸（血清素的前体）的耗竭会导致血清素水平降低，因此降低合作意愿。一些支持者认为催产素与SSRIs的道德效应说明生物道德增强具有广阔的发展前景。

虽然实验表明摄入催产素可以产生道德效应，但药理学研究也发现了催产素的严重副作用：催产素刺激下的利他主义倾向敏于（sensitive to）群体成员的身份。这意味着催产素引发的亲社会倾向仅限于群组内成员（ingroup members）。它在促进内部信任与合作的同时排除外部群体，并引发对外部群体的防备心态。[5]，[6] 卡斯滕·德律（Carsten de Dreu）等人的实验还表明，更高的催产素水平会加剧人类种族主义，即催产素放大了群体内部的信任与互惠强度，因而可能会引发种族或阶级歧视。[7] 而且，研究表明催产素对人类决策的影响高度依赖于情境特征。换言之，催产素能否产生亲社会倾向高度取决于摄入催产素的个人自身及其所处的环境的特征。[8]

SSRIs也无法担任道德增强的重任。首先，持续高浓度的血清素对人体有致命毒性，其安全性缺乏可靠保证。其次，SSRIs有严重的副作用，包括失静症、烦躁与自杀企图，长期使用SSRIs可能会导致不安、紧张、睡眠障碍等不良症状。第三，SSRIs的效果在使用后并不能持久，因而不能真正改善动机或人格。第四，SSRIs能导致药物诱发的冷漠状态，甚至会降低与他人的社会联系或依恋，所以无法真正地改善道德表现。[9] 最后，SSRIs不能以敏于情境的方式运作，因而甚至不是治疗易怒症的最有效药物。[10]，[11]

相较于增强利他主义倾向与公平感，一些哲学家认为我们应该探索如何削弱反社会动机。拉斐尔·阿尔斯科格（Rafael Ahlskog）主张我们应该利用诸如LSD之类的致幻剂来降低人们的自我意识（sense of self），让人们不僵化地将自己等同于一个独立于外在世界的实体，从而降低人们的自利动机。[12] 然而，LSD之流的致幻剂是多国法律规定的违禁品，它们降低个人自利动机的同时会剥夺他们自我实现、自我关爱与自我保存的动力，这些动力却是社会运作与发展的基础。一个由没有自利动机的瘾君子所组成的停滞社会绝非是一个道德沃土。

就削弱反社会动机而言，一种伦理上较为可取的方法是安德烈亚·拉瓦扎（Andrea Lavazza）所提倡的记忆编辑技术。[13] 她认为具有不安全依恋性人格（insecure attachment）的人们往往有童年创伤，这些创伤记忆使得他们对他人有较低的信任感，也抑制了他们的利他主义倾向。因此，她推荐使用心安得（propranolol）之类的药物或更先进的记忆编辑技术来消除创伤记忆，从而削弱反社会动机。笔者认为记忆编辑技术是比较合理且可实践的道德增强方式，毕竟心理治疗已经得到社会的广泛接受，而其功能就在于降低创伤记忆对个人的影响。唯一的不足之处是记忆编辑只能作用于受创伤困扰的个人，它无法成为通用的道德增强手段。

考虑到上述药物的不足之处，一些生物增强的支持者转而探索基因工程对道德增强的作用。沃金·拉基奇（Vojin Raki）主张可以利用基因编辑来增强未出生婴儿的利他主义倾向、减少他们的进攻性。[14] 但是，人类的基因编辑知识依然十分浅薄，对一段基因的修改往往会给有机体带来意想不到的副作用。拉基奇的提议就目前的科研水平而言完全没有可行性。而且对胚胎进行基因编辑会侵犯父母依照他们认为合适的方式来生育子女的权利：它限制了父母对孩子未来的决策自主权与自由。产前基因道德增强还会限制儿童选择与形成自己道德性格

的能力，将一套特定的价值观强加给他们。自主选择是人类繁荣（flourishing）的基础，不应该从儿童身上夺走。[15]

就当前科研水平而言，较具有现实性的措施有脑部刺激技术，比如经颅磁刺激（TMS）或深部脑刺激（DBS），以及神经反馈训练，比如使用实时反馈的脑部活动监测来训练个体调节他们自己的脑过程。[16]-[18] 两者都可以用于增加与利他主义倾向相关的脑区活动，或减少与攻击性相关的区域的活动。然而，道德行为是一个复杂的现象，涉及多种认知与情绪过程，因此很难确定始终与道德行为相关联的特定脑区。并且，不同个体在与道德动机相关的脑活动模式上可能存在差异，我们难以开发一种适用于所有人的脑部刺激或神经反馈训练方法来增强道德动机。而且这两方法不止耗费时间、价格高昂，还会引起头痛、肌肉抽搐等副作用。总体而言，它们并非生物道德增强的有效手段。

目前来说，每一种增强道德动机的生物医学手段都有其自身的缺陷。而且，不管是利用药物、脑神经技术还是基因编辑来增强道德动机，似乎都会侵犯个人的自由。如佩尔松和萨维列斯库所说，道德动机增强就像一台“上帝机器”，使人们心理上无法（psychologically incapable）选择做不道德之事。[19] 道德动机增强剥夺了个人“堕落的自由”，[20] 因而染指了个人的自主性（autonomy）与道德行动能力（moral agency）。虽然做坏事的自由本身毫无价值，但当一个人不由自主地做好事时，他的善行似乎也失去了道德价值（moral worth）。[21] 因此，相较于基于生物医学的干预，一些哲学家更偏向于运用传统的、具有群众基础的道德增强方式，[22] 比如社会化、教育以及父母的看护。[2]，[4]

三、道德认知增强

传统的道德增强方式历时长、见效慢，而且通过这种方式来取得道德进步往往需要付出巨大的代价。比如，美国奴隶制经历了旷日持久的战争与无数鲜血的浸礼才得以废除。[24] 为了在不侵害个人自由的前提下催产道德进步，减少前进过程的阵痛，一些哲学家认为我们应该采取各色措施进行（道德）认知增强。

由于存在广泛且合理的道德分歧（moral disagreement），道德认知增强不应简单地将某种实质性道德观点强加于人。[25] 于是，欧文·舍费尔（Owen Schaefer）与萨维列斯库识别了六种独立于任何实质性道德观点的特征，认为这些特征的改善能提高人们道德推理（moral reasoning）的可靠性，使道德推理更加可能获得道德真理。[26] 第一种特征是逻辑能力，即进行适当的逻辑推理与演绎，发现自己和他人信念（beliefs）中的矛盾，并且以一种能够突显与对谈者之间真正分歧点的方式来构建论证的能力。其次是概念理解力。这包括了一般性反思能力、对细节的关注以及对抽象内容的阐明和理解。第三种特征是经验能力（empirical competence）。这种能力包括了长期记忆力和与道德判断相关的知识。第四种特征是对于修改道德观点的开放心态。具有这种能力的人愿意接受自己观点的惊人逻辑蕴含，关注支持和反对这些观点的理由，且在经过深思熟虑后愿意改变自己的观点。第五种特质是共情理解能力。它使得我们能够领会他人的主观经验，从而为我们的道德推理提供了经验性前提。第六种特质是避免偏见的能力。这种能力使我们的道德推理能够排除与道德无关的因素。上述两位作者认为他们“所识别的许多能力原则上应该是可以通过生物学手段进行改善的”。[26]

事实上，现有文献早有论述如何改善这些认知能力。瑞恩·汤肯斯（Ryan Tonkens）提出使用阿得拉尔（Aderall）来增强人们的大脑能量、学习动机与对认知任务的兴趣和享受感。[27] 这些增强可以带来更强的注意力和集中力，从而增强诸如逻辑能力、概念理解等认知功能。但是阿得拉尔作为一种治疗过动症的处方药，如用于提高认知能力的目的，会导致焦虑、失眠以及心脏问题。长期依赖阿得拉尔不止会危害个人健康，还会影响其个人人格同一性与自主性，使他依赖于捷径而非持续努力来获取认知上的成就。

保罗·内尔布特（Paulo Norbert）认为我们

可以用增强道德动机的药物来提升我们的共感能力。[28]比如，一个好斗的人可以通过摄入抑制进攻性的大麻来抑制自己的暴力倾向，从而体验那些畏惧进攻性与支配性行为的人们的精神世界。相似地，布赖恩·厄普（Brian D. Earp）认为，在严格的临床医疗条件下，我们可以使用致幻剂来促进对自己内在世界和周遭环境的洞察力与认知。伴随追求道德成长的自我努力，这些认知可以帮助我们取得道德增强。[29]戈登认可致幻剂对道德增强的作用，但她进一步强调如果致幻剂要起最大作用，药物提供者与接受者之间必须确立强健的信任关系。[30]

但是，无论是SSRIs、大麻还是致幻剂，都无法避免上文所指出的这些药物的副作用，比如引发使用者对世界与他人的冷漠，弱化其自我实现的动力等等。再者，大麻与致幻剂本身就游离在许多国家法律和伦理规范的边缘，广泛使用这些药物造成的伦理问题可能比解决的还要多。

在药物之外，一些哲学家也认为前述的神经技术——如脑部刺激与神经反馈训练——是增强道德认知能力的可行办法。约翰·舒克（John Shook）认为实时的电脑图（electroencephalography）或核磁共振神经反馈等技术可以用于修改某些认知过程，例如注意力、记忆和决策制定。这些技术涉及实时监测脑活动并向个体提供反馈，使他们能够学会如何控制脑活动，提高认知表现，进而提高其道德表现。例如，通过提高我们对于与道德相关的因素的注意力或记住过去的道德决策，我们可能能够在未来做出更好的道德选择。[31]但是，道德决策是一个复杂的过程，涉及到无数认知和情感因素。虽然人类在理解道德决策的神经基础方面已经取得了显著进展，但仍有许多未知之处。例如，我们仍不清楚不同的脑区是如何相互作用来产生道德判断，以及个体脑结构或功能上的差异如何影响道德认知。[32]就当前神经科学发展水平而言，通过神经技术提高道德认知水平的设想只能说是未来可期。

由于直接增强所有人的认知能力既不可取也不可行，一些哲学家设想我们是否能够通过改善个别人的认知能力，以使他们成为道德增强的引路人。艾玛·戈登（Emma C. Gordon）提倡首先将认知增强的生物医学技术运用于科研人员，然后让他们去探索道德增强的有效手段（比如脑机交互、苏格拉底式人工智能等等），并将这些手段作用于普罗大众。[33]但是，即使我们当前真的有增强个别人智能的成熟技术，我们也无法保证这些得到认知能力增强的人是道德上正直的人，无法保证他们没有利用高等智能宰制大众的邪恶目的。[3]

就当前科技水平来说，生物医学技术还无法将认知增强广泛应用于大众，而聚焦于一小撮精英又与道德增强的最初目的（防范人类存在性风险）背道而驰。现今社会其实有一些成熟的道德认知增强手段。比如，哈里斯·威斯曼（Harris Wiseman）推荐我们将认知行为疗法（CBT）作用于具有人格障碍或行为问题的患者。他还认为我们可以通过正念冥想来发展出更大的自我意识与同理心。我们也可以通过道德教育来教授个人关于伦理原则与价值观的知识。[34]不过，这些措施或者适用范围小，或者见效缓慢，无法符合大多数生物道德增强支持者的预期。

四、困境与曙光

通过对现有文献的回顾，笔者发现当前生物医学技术仍无法为道德增强提供令人满意的解决方案。就增强道德动机而言，目前可用于增强利他主义倾向或公平感的药品都会诱发不利于道德进步的反向道德动机。催产素会加剧对群体内部成员的偏爱，深化种族、阶级中心主义，而SSRIs会引发个人对他人与社会的冷漠感。致幻剂之类的药品具有更严重的副作用，且与许多国家法律与道德规则相背离。人类现有基因编辑与神经科学技术也不够成熟，难当重任。而且不管采取哪一种方式增强道德动机，都在某些程度上剥夺个人意志自由与自主性，从而降低个人善行的道德功绩。

相对增强道德动机而言，道德认知增强有利于保障个人自主地做出道德行为。因为无论个人是在逻辑能力、概念能力、经验能力、开

放心态、共情能力还是偏见避免中的哪一方面得到增强，他依然是进行道德推理的主体，其道德决策与道德行为仍然是他自己凭借意志所做出的选择。不过，当前人类生物医学科技，无论是药理学还是神经科学，仍无法实现大规模的道德认知增强。

诚然，有些增强道德动机或认知的技术已经得到普遍使用，比如利用心安得等药物来治愈童年创伤与认知行为疗法。但是，这些技术仅适用于心理疾病患者，不能推广到一般大众。然而，考虑到资源枯竭、气候恶化与利用大规模杀伤性武器的恐怖袭击等伦理挑战的迫切性，我们需要的恰恰是大规模的道德增强。难道我们的伦理困境是无解的吗？

若仅着眼于生物医学技术，答案或许是无解的：我们在研发出安全有效的技术前仍需无数岁月，而人类生存危机却迫在眉睫。但是，当前人工智能技术（特别是大型语言模型）的发展，似乎为我们的伦理困境提供了一丝曙光。一些坚定的生物道德增强支持者（比如萨维列斯库）甚至转而支持人工智能（AI）道德增强。[35]

简要地讲，AI道德增强可分为准备性协助与现场协助。[36] 准备性协助功能是指AI在行动者面临道德决策情境之前就为其提供建议和培训。这种协助旨在帮助个人提高决策技能并提供信息，以便在未来真正的决策过程中使用。例如，AI可以向用户提供有关不同购物决策的道德影响的信息——比方说，购买汽油车会对大气层造成哪些伤害。AI也可以通过描绘复杂道德困境来帮助用户提前训练。

另一方面，现场协助涉及实时提供建议和推进（nudge）道德行为，以在道德审思（moral deliberation）过程中帮助个人做出道德决策。由于个人常有逻辑能力缺陷，且易受情境因素与偏见的干扰，AI可以依据用户预设的道德原则，直接为其提议与他价值观一致的行动步骤。

由上可见，AI道德增强的作用点正是个人道德认知能力。因此，这种道德增强并不会干预个人道德选择的自由——做出决策的主体仍然是行动者本人。与生物医学认知增强不同的是，大型语言模型的发展证明了AI道德增强技术可以快速走向成熟，并且不会有剧烈的副作用。

AI道德增强似乎能够有效地缓解人类文明的存续危机。在日常生活中，相较于道德动机与根本性道德原则，缺乏道德认知能力往往是行动者做出不道德决策的深层原因。在资源枯竭与环境恶化方面，许多消费者虽然意识到环境保护的重要性，但由于不了解各类产品的环境影响（即缺乏经验能力），而购买了高碳足迹的商品。而通过AI拟真受气候恶化影响的第三世界人口的窘境，发达国家的大众也能强化共情能力。在防范恐怖主义方面，AI所提供的各类识别恶意意图（malintent）的手段能为执法与司法机关提供必要信息，从而更好地保护公民生命与安全。

当然，AI道德增强也引起一些伦理上的忧虑。[36] 首先，为了给用户提供个性化的建议，AI道德增强系统需要获取用户的敏感道德数据（sensitive moral data），包括用户的价值观与信念，甚至还需要深入研究并建模用户的道德心理。收集这些信息可能会侵犯用户的隐私权。更加重要的是，敏感道德信息的泄露会给用户带来极大的困扰。其次，创建一个AI道德增强系统需要特定群体（往往是科技巨头）来设计、制造与推广。这种技术的开发和使用会受到经济与政治体系、以及开发者的利益和价值观的影响。因此，系统向用户提供的任何建议或指引，都将部分取决于系统设计者、制造机构、标准化组织等的选择。由于算法的复杂性与不透明性，个体用户可能会受到系统开发者的宰制。由于AI道德增强系统能够收集用户的敏感道德数据，系统开发者能轻易利用这些数据来操纵用户。最后，AI道德增强系统的不透明性，也容易使用户在不清楚系统决策理由的基础上盲从系统的道德建议，从而削弱自身行为的道德责任与价值。

然而，与生物医学道德增强相比，AI道德增强的短板显然较为轻微。当前，许多应用软件均由科技巨头等外部机构开发，并持续收集用户敏感信息，用户也常因应用推荐而盲目作出决策，如消费决策等。鉴于信息科技的这一

现状并未给社会带来大规模负面效应，我们也没有理由过度忧虑AI增强系统的风险。更为关键的是，如果人工智能道德增强能助力人类实施他们原本不会做出的道德行为，特别是那些在危机中帮助文明延续的行为，那么AI道德增强的短板实则是我们愿意并值得承担的风险。

AI道德增强的伦理风险还可以通过民众参与和社会实验加以缓解。民众参与是指一般大众参与到AI的设计、开发与推广中。在AI道德增强系统的设计阶段，开发者要广泛征求民众的意见，并基于这些意见确定系统的基本方向。诚然，大多数民众缺乏编程的必要知识，但民众有权利且有能力决定自己依据哪些道德原则、在哪些方面进行道德增强，因此他们应该主导道德增强系统的蓝图绘制。在开发阶段，系统开发者应着力于开发可解释性AI（explainable AI），力求将复杂模糊的算法以简洁易懂的词汇加以解释，以回应民众疑虑。同时，开发者应将训练数据的筛选标准与来源公之于众，以便民众检查数据选取是否存在偏见（bias）。在推广阶段，开发者应该明确透露系统的缺陷与副作用，以便民众做出基于充分信息的自主决策，选择是否使用该系统。笔者认为有AI知识背景的伦理学家也应积极参与AI道德增强系统的开发过程中，充当民众与开发者之间的桥梁：一方面向开发者传递民众的道德观念与需求，另一方面用平实语言向民众解释算法与训练数据。

社会实验是指先将AI系统运用于局部试验中，根据试验的反馈摈弃或改进系统，最终才将其推广到全社会。用社会实验来检验具有重大社会影响的改革是包括约翰·密尔（John Mill）、伊丽莎白·安德森（Elizabeth Anderson）等美国实用主义者，卡尔·波普尔（Karl Popper）以及邓小平在内诸多重要思想家的共识。AI道德增强技术既能塑造社会整体道德风貌，又会引起种种伦理忧虑。因此，开发者应该基于自愿原则，在小规模试点人群中试验各色AI道德增强系统，依据它们的表现进行选择与完善，再逐步扩展到更广大的人群中，力求在不断地检验与在开发中增进系统表现，消弭其缺陷。

AI道德增强究竟在多大程度上可以解决我们当前的伦理困境，仍然是一个方兴未艾的议题。目前，科技哲学家正逐步将目光转向这个话题。随着文献进一步积累，AI道德增强的前景将会逐步明朗。

结　语

笔者通过回顾生物道德增强的现有文献，发现现有技术无法实现其支持者的根本愿景，即从生存危机中拯救人类未来。当前人工智能的急速发展使得技术与道德上都更加优越的AI道德增强成为可能，这为缓解科技发展所引发的伦理挑战带来一缕曙光。

诚然，AI道德增强仍有不少弊端，比如对隐私的侵犯、科技巨头的宰制、个人道德责任的弱化等等，[36]笔者因字数所限无法进一步探索该技术的缺陷与对策，将通过后续研究加以探讨。不过，人类当前伦理挑战是如此严峻，任何解决方案都值得探索。既然AI道德增强是技术与伦理上最为可取的选项，我们没有理由不着重对其进行研究。不过，生物道德增强仍有用武之地：那些由于心理缺陷而缺乏道德动机的个体，仍需要生物医学措施加以调节。AI与生物医学相结合，人类文明才会有蓬勃发展的未来。

［参 考 文 献］

[1]Persson, I., Savulescu, J. *Unfit for the Future: The Need for Moral Enhancement*[M]. New York: Oxford University Press, 2012.

[2]Douglas, T. 'Moral Enhancement'[J]. *Journal of Applied Philosophy*, 2008, 25(3): 228–245.

[3]Persson, I., Savulescu, J. 'The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity'[J]. *Journal of Applied Philosophy*, 2008, 25(3): 162–177.

[4]Harris, J., Savulescu, J. 'A Debate About Moral Enhancement'[J]. *Cambridge Quarterly of Healthcare Ethics*, 2015, 24(1): 8–22.

[5]Rakić, V. 'Compulsory Administration of Oxytocin Does not Result in Genuine Moral Enhancement'[J]. *Medicine, Health Care and Philosophy*, 2017, 20(3): 291–297.

[6]De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., et al.

'The Neuropeptide Oxytocin Regulates Parochial Altruism in Intergroup Conflict Among Humans'[J]. *Science*, 2010, 328(5984): 1408–1411.

[7] De Dreu, C. K. W., Greer, L. L., Van Kleef, G. A., et al. 'Oxytocin Promotes Human Ethnocentrism'[J]. *Proceedings of the National Academy of Sciences*, 2011, 108(4): 1262–1266.

[8] Kudlek, K. 'The Role of Emotion Modulation in Moral Bioenhancement Debate'[J]. *Topoi*, 2019, 38(1): 113–123.

[9] Powell, S. K. 'SSRIs as a Component of, Rather Than Exclusive Means to, Moral Enhancement'[J]. *AJOB Neuroscience*, 2014, 5(3): 33–34.

[10] Wiseman, H. 'SSRIs and Moral Enhancement: Looking Deeper'[J]. *AJOB Neuroscience*, 2014, 5(4): W1–W7.

[11] Wiseman, H. 'SSRIs as Moral Enhancement Interventions: A Practical Dead End'[J]. *AJOB Neuroscience*, 2014, 5(3): 21–30.

[12] Ahlskog, R. 'Moral Enhancement Should Target Self-interest and Cognitive Capacity'[J]. *Neuroethics*, 2017, 10(3): 363–373.

[13] Lavazza, A. 'Moral Bioenhancement Through Memory-editing: A Risk for Identity and Authenticity?'[J]. *Topoi*, 2019, 38(1): 15–27.

[14] Rakić, V. 'Genome Editing for Involuntary Moral Enhancement'[J]. *Cambridge Quarterly of Healthcare Ethics*, 2019, 28(1): 46–54.

[15] Tonkens, R. '"My Child Will Never Initiate Ultimate Harm": An Argument Against Moral Enhancement'[J]. *Journal of Medical Ethics*, 2015, 41(3): 245–251.

[16] Darby, R. R., Pascual-Leone, A. 'Moral Enhancement Using Non-invasive Brain Stimulation'[J/OL]. Frontiers in Human Neuroscience, 2017, 11(77), https://www.frontiersin.org/articles/10.3389/fnhum.2017.00077/full.

[17] Nakazawa, E., Yamamoto, K., Tachibana, K., et al. 'Ethics of Decoded Neurofeedback in Clinical Research, Treatment, and Moral Enhancement'[J]. *AJOB Neuroscience*, 2016, 7(2): 110–117.

[18] Shook, J. R. 'Neuroethics and the Possible Types of Moral Enhancement'[J]. *AJOB Neuroscience*, 2012, 3(4): 3–14.

[19] Savulescu, J., Persson, I., The Hegeler Institute. 'Moral Enhancement, Freedom, and the God Machine'[J]. *Monist*, 2012, 95(3): 399–421.

[20] Harris, J. 'Moral Enhancement and Freedom'[J]. *Bioethics*, 2011, 25(2): 102–111.

[21] Simkulet, W. 'Intention and Moral Enhancement'[J]. *Bioethics*, 2016, 30(9): 714–720.

[22] Specker, J., Schermer, M. H. N., Reiner, P. B. 'Public Attitudes Towards Moral Enhancement: Evidence that Means Matter Morally'[J]. *Neuroethics*, 2017, 10(3): 405–417.

[23] Buchanan, A., Powell, R. *The Evolution of Moral Progress: A Biocultural Theory*[M]. New York: Oxford University Press, 2018.

[24] Fabiano, J. 'Technological Moral Enhancement or Traditional Moral Progress? Why Not Both?'[J]. *Journal of Medical Ethics*, 2020, 46(6): 405–411.

[25] Schaefer, G. O. 'Direct vs. Indirect Moral Enhancement'[J]. *Kennedy Institute of Ethics Journal*, 2015, 25(3): 261–289.

[26] Schaefer, G. O., Savulescu, J. 'Procedural Moral Enhancement'[J]. *Neuroethics*, 2019, 12(1): 73–84.

[27] Tonkens, R. 'Feeling Good about the End: Adderall and Moral Enhancement'[J]. *AJOB Neuroscience*, 2013, 4(1): 15–16.

[28] Paulo, N. 'Moral-epistemic Enhancement'[J]. *Royal Institute of Philosophy Supplements*, 2018, 83: 165–188.

[29] Earp, B. D. 'Psychedelic Moral Enhancement'[J]. *Royal Institute of Philosophy Supplements*, 2018, 83: 415–439.

[30] Gordon, E. C. 'Trust and Psychedelic Moral Enhancement'[J]. *Neuroethics*, 2022, 15(2): 19.

[31] Shook, J. R. 'My Brain Made Me Moral: Moral Performance Enhancement for Realists'[J]. *Neuroethics*, 2016, 9(3): 199–211.

[32] Shook, J. R., Giordano, J. 'Moral Enhancement? Acknowledging Limitations of Neurotechnology and Morality'[J]. *AJOB Neuroscience*, 2016, 7(2): 118–120.

[33] Gordon, E. C., Ragonese, V. 'Cognitive and Moral Enhancement: A Practical Proposal'[J]. *Journal of Applied Philosophy*, 2022, 1–14.

[34] Wiseman, H. 'Moral Enhancement-"Hard" and "Soft" Forms'[J]. *The American Journal of Bioethics*, 2014, 14(4): 48–49.

[35] Savulescu, J., Maslen, H. 'Moral Enhancement and Artificial Intelligence: Moral AI?'[A], Romportl, J., Zackova, E., Kelemen, J. (Eds.) *Beyond Artificial Intelligence: The Disappearing Human-Machine Divide*[C], Cham: Springer International Publishing, 2015, 79–95.

[36] O'Neil, E., Klincewicz, M., Kemmer, M. 'Ethical Issues with Artificial Ethics Assistants'[A], Véliz, C. (Ed.) *The Oxford Handbook of Digital Ethics*[C], Oxford: Oxford University Press, 2021.

[责任编辑　李斌]