



数据驱动的道德物化研究

杨文字, 王小伟*

中国人民大学 哲学院, 北京 100872

摘要: 道德物化的核心目标, 是将道德准则嵌入技术架构中, 引导人做出符合积极道德价值的行为选择。然而, 当前主流的知识驱动型道德物化模式仍存在显著局限: 一方面, 其对设计者道德知识体系的依赖, 使设计者权责过大; 另一方面, 现有道德化技术的引导形式也缺乏与使用者之间动态、深度的交互反馈机制, 道德物化的可行性与有效性均遭受质疑。为提升道德化技术的引导效果, 有必要将数据驱动引入道德物化。道德设计可通过数据咨询、数据分析与数据呈现的三重形式加以驱动, 以数据分散设计者的权责, 并推动使用者与道德化技术呈现出新的反身关系, 实现自律与物律的有效结合。此外, 数据驱动的道德物化亦需应对欺骗性对齐、数据维护与过度量化等可能挑战。在知识与数据双驱动下, 道德物化为科技世代的道德建设提供了新的可能路径。

关键词: 道德物化; 道德设计; 数据驱动; 反身关系

中图分类号: N031; TP3

文献标识码: A

文章编号: 1674-4969(2026)03-0238-12

引言

道德物化 (Materializing morality) 指将特定的道德观念嵌入技术当中, 使其对人的道德行为产生积极引导。受伊德 (Don Ihde) 的技术现象学与阿特胡斯 (Hans Achterhuis) 的“器具道德化” (de moralisering van apparaten) 的启发, 荷兰技术哲学家彼得-保罗·维贝克 (Peter-Paul Verbeek) 将技术的调

节作用从知觉拓展至道德领域, 提出了道德物化理论。道德物化理论的提出直指当下社会道德建设的现实困境: 在环境保护等领域, 单纯依靠道德呼吁难以转化为有效的个体行动。维贝克就此提出, “道德焦虑的本质不是道德陌生, 而是道德无力”^[1], 尝试主动介入技术的设计过程, 以技术推动当下的道德建设落地生效。

然而, 进入数字时代, 数据在技术体系中的地位

收稿日期: 2025-07-04; **修回日期:** 2025-09-29

基金项目: 北京市社会科学基金重点项目: “道德物化”技术伦理思想研究 (21ZX001)

作者简介: 杨文字 (2002—), 男, 硕士研究生, 研究方向为技术哲学。E-mail: yangwenyu0208@163.com

*王小伟 (1985—), 男, 博士, 教授, 研究方向为技术哲学。E-mail: xiaowei_beijing@163.com (通讯作者)

引用格式: 杨文字, 王小伟. 数据驱动的道德物化研究[J]. 工程研究——跨学科视野中的工程, 2026, 18(3): 238-249. DOI: 10.3724/j.issn.1674-4969.20250097. CSTR: 32282.14.JES.20250097

YANG W Y, WANG X W. The Research of Data-Driven Materializing Morality[J]. Journal of Engineering Studies, 2026, 18(3): 238-249. DOI: 10.3724/j.issn.1674-4969.20250097. CSTR: 32282.14.JES.20250097

日渐凸显, 仅依赖设计者掌握的有形技术功能性知识开展道德设计, 已难以达成理想的引导效果。无论是在设计阶段对技术道德引导作用的精准预期, 还是在应用阶段以更具交互性的方式实现技术的道德引导, 都越发需要数据的驱动。因此, 正如以深度神经网络为代表的机器学习技术, 通过算法解析并学习数据以构建模型, 不仅推动了 ChatGPT、DeepSeek 等大语言模型的蓬勃发展, 更实现了人工智能主流研究路径从符号主义依赖的知识驱动向联结主义依托的数据驱动的范式转变; 为进一步发展道德物化, 同样有必要将无形的数据纳入道德设计核心框架。基于此, 本文将探究数据驱动的道德物化的理论框架, 作为目前知识驱动型道德物化的补充, 以期为克服知识驱动模式下道德物化所面临的现实困境, 增强道德物化的可实践性, 更好地发挥道德物化对道德建设的积极作用, 提供有益借鉴。

1 知识驱动下道德物化的现实困境

在道德物化的理论与实践, 若想理解技术如何发挥对个体的道德引导作用, 首先要明确“什么样的道德讨论伴随着道德技术的使用与设计”^{[2]2}。维贝克探讨的道德遵循“善之生活的伦理学”^{[2]156}, 试图解决“个体与技术以善之生活的方式共存何以可能”^{[2]157}的问题。在维贝克看来, 这意味着技术引导下个体的道德行为应当是一种“好”(good)的行为, 指向的是人与技术共同达成的“卓越”(excellent)^{[2]156}状态^①。

因此, “美好生活”(the good life)^{[2]31}是道德物化的出发点与落脚点。评判一项道德设计是否很好地达成了道德物化的目的, 其标准不仅包含个体在技术引导下做出的对他人与社会有益的行为, 也包含各种妥当的自爱行为, “如说服性镜子鼓励人们培养更健康的生活方式”^{[2]129}, 这类设计的本质都是让个体在“卓越”状态中过上美好生活。正确理解技术在何种意义上发挥对个体的道德引导, 有助于厘清道德物化

的设计情境与理念指向。在此基础上进一步分析会发现, 道德化技术的应用效果在很大程度上受到设计理念、方法与路径的影响, 正因如此, 道德物化又被称为“设计的物质伦理学”(material ethics of design)^{[2]2}。这一界定深刻揭示了技术的道德引导能否顺利进行, 关键在于采用何种方式驱动道德设计。

1.1 知识驱动的道德物化

目前学界与业界普遍采用的道德物化路径可视为由知识驱动的。在当下的道德设计中, 要让技术发挥道德引导作用, 核心手段是为技术“内嵌”特定的道德知识, 这与早期依靠知识驱动的人工智能有相似之处, 即将知识以一种技术能够“理解”与“应用”的方式融入技术体系。知识驱动型人工智能的理论源头可追溯至符号主义。1955年, 麦卡锡(J. McCarthy)和明斯基(M. L. Minsky)等学者在《人工智能达特茅斯夏季研究项目提案》(A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence)中首次系统提出, 由于符号主义的人工智能主要基于知识与经验的推理模型, 因此符号AI又被称为知识驱动方法^[3]。

专家系统(expert systems)作为知识驱动型人工智能的典型代表, 其本质是一种由知识库(knowledge base)与推理机(inference engine)组成的智能程序, 核心功能是处理以往仅由领域专家才能应对的复杂问题。专家系统承担着专家与用户之间的“中介”角色, 即专家本身是专家系统的知识获取来源, 而用户无须直接与专家对接, 只需通过专家系统发起咨询, 系统便可借助推理机的知识推理能力给出相应的建议。因此, “知识推理、知识运用是知识驱动的人工智能的核心”^{[4]111}。

在道德物化的语境中, 具备道德引导功能的技术, 在一定程度上可以类比为“道德专家系统”。这类道德化技术中嵌入的道德观念, 由作为“专家”的设计者预先设定。设计者“实际上是实践伦理学家”^{[2]90}, 技术在何种维度以及通过何种方式发挥道

① 在古希腊的德性论语境中, “卓越”不仅指涉“人”的卓越, 而是泛指一切“物”(things)的卓越, 卓越表示的是“使之充分实现其功能、目的的特性或品质”。维贝克以此突显道德物化中的道德行为是由个体在与技术的交互过程中形成的“共同体”做出的, 即当技术“卓越”地发挥了道德引导作用, 让个体变得更为“卓越”时, 道德物化在维贝克看来便是足够有效的。

德引导作用,很大程度上取决于设计者进行怎样的道德设计。在此过程中,道德化技术充当了设计者与使用者的中介:当处于具体应用场景中的使用者,触发了设计者“授权”(delegations)给技术的“脚本”(script)时^[5],一次“道德专家系统”模式的“咨询-建议”闭环便完成了。正如当设计者为减速带赋予“禁止超速”的道德观念后,减速带便成为传递这一观念的“传输通道”,调节驾驶者的行为选择,使其符合预设的道德规范。由此可见,在知识驱动的道德物化进程中,设计者“在最终通过技术调节道德行为和决策上发挥着开创性的作用”^{[2]90},其掌握的经验性道德知识,直接决定道德物化能否顺利推进;而“道德专家系统”的有效运行,更依赖于设计者是否能“以道德上负责任的方式设计道德化技术”^{[2]106}。也正是从这一逻辑出发,当下的道德物化在本质上呈现出由知识驱动的鲜明特征。

然而,与人工智能的专家系统不同的是,在知识驱动的道德物化中,当设计者的道德知识内嵌于技术中时,呈现给使用者的“道德知识”会发生一定程度的“转译”(translation),这种转译不再停留在认知层面,而是更多渗透到知觉与行动层面。从技术现象学的视角出发,维贝克的观点颇具启发意义。他认为,道德化技术的使用者,其“道德决定主要是在技术对行动、知觉和解释的调节下做出的”^{[2]54},因此仅就技术的引导过程而言,其不必然需要对使用者的道德观念进行调节。换言之,作为“道德专家系统”的道德化技术为使用者提供“建议”的方式具有独特性:不会直接与使用者进行认知意义上的对话,而是先将设计者的道德知识转译为一套对应的行为模式,再引导使用者在实际行动中遵循这套模式,从而实现道德引导的效果。

因此,作为“专家”的设计者将自身的道德知识写入技术的过程,不是一个将知识“累积”为“知识库”的过程。道德化技术既难以依托类似“推理机”的机制对知识进行推理,也难以根据使用者的具体情境提供相对个性化的道德引导的答案。对于使用者而言,这种由技术预设的固定化道德引导形式,往往缺乏足够的“智能”弹性,很容易被使用者解读为技术及其背后的设计者对自身道德行为的一种限制。由

此,道德化技术便呈现出列维纳斯所说的“他异性”(alterity)特质:“它无限地越出我的权能,并因此并不与我的权能相对立,而是使权能的能够本身瘫痪”^[6]。这种特质使得使用者与道德化技术之间,形成了一种消极意义上的“他异关系”(alterity relations)^{[7]107}。出于对“他者”可能限制甚至支配自身的警惕,即使使用者在知识驱动的道德化技术的引导下做出了符合规范的“道德”行为,也难以达到维贝克所倡导的,通过“人-技”协作达成的“卓越”状态,反而会让道德物化遭受如引发非道德物化、限制个体选择自由、助长“技治主义”(technocracy)等^{[2]96}诸多争议与批评,使道德物化的实践推进愈加陷入困境。

1.2 道德物化的困境成因

若要进一步拆解上述困境的成因,可从设计与应用两个维度展开。对于设计,主要关乎知识驱动下道德设计者的权限边界与责任归属问题。由于道德物化的关键在设计环节,而当道德化技术的设计过度依赖作为“专家”的设计者的道德知识体系时,就会放大设计者本应承担的权责。

一方面,设计者的“专家”角色使其权限过大。正如人工智能专家系统的效能依赖于专家构建的知识库,知识驱动的道德物化依赖于设计者为技术嵌入的道德知识,这便意味着设计者需要对技术发挥道德引导作用的使用场景进行预测。因此,为了帮助设计者“在设计场景与使用场景之间建立联系”^{[2]99},维贝克提出了3种思路,即道德想象、强化的建构性技术评估(CTA)与虚拟现实技术(VR)^{[2]118-119}。然而,这些思路本质上只是在为设计者优化道德设计提供支撑,最终如何发挥技术的道德引导作用仍需设计者决定,无法对其形成实质性的约束。当设计者作为“专家似乎拥有专门知识,而公众是决策后果的承担者”^[8],人们便很容易将道德物化视为“潜在的向技治主义的运动”^{[2]133},进而发起对技术支配道德的批评。例如,当公众发现公园中的长椅带有隔断,剥夺了人们在长椅上睡觉的权利时,便将其定义为“敌意设计”(hostile design)^[9]——人们会对此类不道德物化背后设计者的动机、知识合理性及其权力滥用发出

质疑,进而排斥道德物化的设计行为本身,致使道德物化的实践困境陷入恶性循环。

另一方面,设计者的“专家”角色使其背负责任也过大。道德设计者原本至少需要“预期、评估以及在必要时再设计”^{[2]129}道德化技术发挥的调节作用,“审慎考虑设计中技术可能发挥的调节作用,应该是设计者道德责任的一部分”^{[2]128}。而在现实中,技术自身具有多元稳定性(multistability),其发挥道德引导的场景亦具有不确定性与复杂性,因此,“设计者的活动和他们所设计的技术所发挥的调节作用之间存在不对等关系”^{[2]97}。正如旋转门最初的设计目的是减少开门时的空气交换,但它却意外产生了阻碍坐轮椅的残疾人群进入的不道德后果,而这一不道德物化的现象并非设计者有意为之。

因此,对于设计者而言,“影响的可预见性是严肃的挑战”^{[2]132},尤其道德物化涉及对敏感道德领域的干预,这对设计者的要求更高,其自身的经验性知识往往无法支撑起知识驱动模式下道德物化对“专家”角色的严苛期待。对此,维贝克也承认,“不是所有的道德物化都将责任留给设计者本身”^{[2]119},毕竟设计者无法为道德物化的最终呈现效果承担全部责任。然而,对于如何解决这一问题,尽管维贝克指出可以尝试将“道德责任分配到那些能够负责的用户中”^{[2]132},但这似乎让道德实践回到了依赖道德呼吁的传统模式中,从而削弱了道德物化的可实践性。

对于应用,主要关乎知识驱动模式下道德化技术的引导效果问题。道德设计的最终目的是应用,一项道德设计的效果如何,应当由道德化技术的使用者进行评判。但在知识驱动模式下的道德物化实践中,道德化技术内嵌的道德行为引导形式不够“智能”,加之使用者与设计者之间的无交互性也让使用者对道德化技术产生了警惕感与排斥感。

其一,在知识驱动的框架下,道德化技术的引导形式是静态的。例如,作为儒家“藏礼于器”思想的物质载体,太师椅通过突出坐于其上的主人的地位,引导人们尊重儒家倡导的道德秩序。因此,一些学者将其解读为一种道德化技术^[10]。但是,太师椅一经工匠制作便无法灵活调整,因此,当儒家的传统道德秩序逐渐被后世批判性地加以审视,太师椅也就逐渐

失去了其效力。换言之,现实中的道德标准会随着时间的流逝发生变化,不同的地区也有不同的道德习俗,而知识驱动下的道德化技术因缺乏灵活性,难以动态响应这种多变的道德情境。一旦道德化技术与现实的道德准则脱节,它就难以发挥好的效果,甚至可能会引发负面影响。这大大提高了道德物化的实践成本,降低了道德物化的可行性。

其二,在知识驱动的框架下,道德化技术的引导对象是被动的。尽管道德物化突显了技术对于道德行为的调节能力,但这一调节作用的实现过程对使用者而言是被动的,即使用者往往并非主动选择使用某类道德化技术。如前所述,知识驱动的道德化技术本质采取的是“放置-触发”模式,即通过将道德嵌入人们日常使用的技术场景中,以使用者对这一技术的使用为契机,完成对人的道德引导。如当驾驶员遇到减速带时,为避免伤害只能被动选择减速,进而接受“防止超速”的行为规范引导。由于缺乏足够的选择权,面对道德化技术,使用者难以像面对人工智能的专家系统那般主动与其产生交互,反而会加深其对道德化技术的“他异性”感受,认为自身的道德主体性正在遭受威胁^[11],进而进一步削弱道德物化的效果。

其三,在知识驱动的框架下,道德化技术的引导机制是外在的。尽管道德被“内在地”嵌入到技术架构,但使用者在技术的道德引导下缺乏足够积极的意志自由,导致其难以将这一道德引导视为内在意志的体现。相较于主动性强调的“行动的选择”,内在性作为一种更高的诉求更注重“意志的选择”。因此,尽管维贝克指出道德物化可以表现为强迫、说服、诱导等多种调节形式^{[2]126},罗伯特·罗森博格(Robert Rosenberger)亦指出,面对具有多元稳定性的技术,可以通过关闭、开放、扩展其特定“稳定性”实现道德物化^[12]。但是,即使知识驱动下的道德设计环节存在丰富的“选择空间”,一旦设计者通过技术手段将其固化为某种特定的道德行为模式,呈现给使用者的选择便是被预先限定的,使用者没有选择采取何种“选择方式”的意志自由,这与选择的呈现形式无关。简言之,即便使用者在行动上“主动”选择了某种被技术引导的方式,也能感知到自身是在遵循设计者的“设定”,以趋近其内嵌的道德期望,那么这一对于外

在“设定”的选择便难以被使用者视为出于自身的自由意志。因此,无论是主动性还是内在性的缺失,都会让使用者产生道德物化限制自身自由的感受,进而使其内心对道德化技术的排斥难以消除。

2 数据驱动下道德物化的呈现形态

道德物化的现实困境表明,知识驱动模式下的道德化技术难以如维贝克设想的那般“卓越地”实现道德引导功能。梳理上述困境的成因可以发现,道德物化之所以会引发诸多批评,除了知识驱动本身的局限性,另一关键症结在于,维贝克所设想的“人-技”协同达成的卓越状态忽视了使用者的主观意志。个体的内在感受与认知会对其行为产生重要影响,而道德化技术对使用者道德行为的引导,必须以其主观意志上不排斥道德化技术所呈现的道德行为模式为前提,唯有如此,技术才能与使用者协同达成卓越状态。反之,道德设计者便难以设身处地考量使用者面对道德化技术时的态度,更无法与其共情。只有当道德物化中技术的“物律”出于或转化为使用者的“自律”,“人-技”协作才能真正通达至“卓越状态”,实现推进道德建设的目的。基于这一认知自觉,面对知识驱动模式的不足,引入数据驱动机制便为解决道德物化的现实困境提供了可能出路。

所谓数据驱动的道德物化,即设计者通过收集、分析、解读、设置数据指导自身进行道德设计。与知识驱动相比,数据驱动的核心理念在于,主要凭借大数据分析结果而非设计者对技术功能性与经验性的知识,推进道德设计。由于在道德物化实践中数据对技术的驱动存在多种形式,若仅以“数据驱动型技术”难以澄清其中“含混关系的各种结构性特征”^{[7]72},更无法揭示数据对道德化技术的驱动是如何发生的。因此,有必要对道德物化中的“人-技”关系表达式进行扩展,将数据作为独立要素纳入分析框架,并立足道德设计者和技术使用者的双重视角,分析“人-数据-技术-世界”中,数据在诸要素间的不同调节形式。唯有如此,才能清晰界定数据驱动下道德物化的呈现形态,进而说明数据驱动模式如何有效地将使用者的感受纳入道德设计的考量范畴。

2.1 数据驱动下的道德设计者

从道德设计者的视角出发,其可采取数据咨询、数据分析与数据呈现3种驱动形式开展设计工作,这也使得数据驱动下的道德设计呈现出3种新形态。首先,设计者的道德设计可以通过数据咨询提供借鉴,即数据作为助手进行驱动,此时设计者与数据呈现出“具身关系”(embodiment relations),其关系表达式为:

$$(人-数据) \longrightarrow 技术 \dashrightarrow 世界 \quad (1)$$

其中,实线箭头表示意向性,其“将所有经历的事物与经验所指涉的经验模式相关联”^[13];虚线箭头则表示间接性关联。进入设计语境,“人”指涉的是道德设计者,“世界”则指涉技术使用者所处的道德物化情境。在数据咨询的实践场景中,基于数据的智能体能够帮助设计者更好地设计道德化技术,人工智能大语言模型(LLM)的使用便是典型案例。通过给出合适的提示词(prompt),大模型的回答能够协助设计者在道德想象等预期调节中,完善有关道德化技术的各种设计细节,拓展设计的可能性边界。这有助于设计者超越自身原有知识的局限性,放大(magnification)其可借鉴的道德知识范围,既能减少因设计者自身对技术使用情境认知不足或存在偏差,而导致不道德物化的风险,也能减轻设计者在知识驱动模式下因承担“专家”角色所产生的认知与行动负担。

进一步讲,当设计者不断精进数据咨询的能力与方法时,数据本身便在与设计者的互动关系中呈现出更高层次的“透明性”。在理想状态下,数据便会“抽身而去”,即设计者在潜意识中主动使用数据进行咨询,其与数据在设计道德化技术的设计过程中形成了一个高度协同的整体。这种状态宛如近视的人佩戴眼镜那般自然,数据咨询的结果成为设计者自身道德知识体系的一部分。由此,设计者与数据的具身结构特征,便主要体现在设计者借助数据咨询所形成的对道德化技术的意向性中,即表达式中的(人-数据)——→技术。

其次,设计者的道德设计可以通过数据分析提供标准,即数据作为依据进行驱动,此时设计者与数据呈现出“诠释关系”(hermeneutic relations),其关系

表达式为:

$$\text{人} \longrightarrow (\text{数据}-\text{技术}) \text{----} \longrightarrow \text{世界} \quad (2)$$

在数据分析中,设计者可以借助实地调研、VR模拟等方式,收集和使用情境直接相关的数据并对其进行展开分析。这有助于设计者以此为依据,更客观地了解使用者在面对道德化技术时的可能状况与实际体验,进一步减少其做出不道德物化的可能性,并使对使用者体验更为积极的道德调节成为可能。例如,通过对车流量等大数据进行积极诠释,公安交警部门能对道路上的红绿灯构建一定周期的联动控制体系,并根据数据分析结果动态调整周期参数。当驾驶者以固定车速行驶时,便能大概率实现到达各交叉口时均是绿灯。这既提高了驾驶者的驾驶体验,使其自愿将车速维持在绿波区间,又有助于防止车辆超速、保障道路安全、提升整体通行效率。同样地,在设计音乐公路时,设计者可收集当地民众广泛接受的音乐曲目,并梳理隧道入口、长下坡等需要控制车速的关键路段特征进行数据分析。这有助于设计者更好地提升音乐公路的实际使用效果,引导使用者积极主动地按照能够触发音乐的速度行驶,并在这一过程中用音乐提醒驾驶者避免疲劳驾驶。

因此,通过数据分析评判是否以及如何设计道德化技术,其本质是设计者通过数据诠释,让技术的道德引导功能更好地呈现。在这一诠释过程中,设计者“可以在没有亲身经历的情况下,将自己置身于任何可能的情境中”^{[7]92},这种情境代入能力进一步强化了设计情境与使用情境的内在关联性。相较于数据咨询,数据分析中的数据“同时作为人借以经验并与之发生关系的东西”^{[7]93-94},呈现出更强的焦点性,对设计者的决策具有决定性意义,即设计者若不将自身聚焦于对数据的诠释,就无法理解数据并借助其恰当地设计道德化技术。由此,设计者与数据的诠释结构特征,便主要体现在设计者对经由数据诠释后形成的道德化技术的意向性中,即表达式(2)中的人——(数据-技术)。

最后,设计者的道德设计可以通过动态数据加以呈现,即数据作为形式进行驱动,此时设计者与数据

呈现出“背景关系”(background relations),其关系表达式为:

$$\text{人}(-\text{数据}) \longrightarrow \text{技术} \text{----} \longrightarrow \text{世界} \quad (3)$$

在数据呈现中,数据这一呈现“形式”而非数据“本身”,成为设计者在设计道德化技术时所关注的焦点。此时,设计者需要解决的核心问题在于,如何让道德化技术借助动态数据的形式更有效地呈现。以番茄TODO、Forest、Habitica为代表的时间管理软件,通过动态数据助力用户保持良好作息、达成自身目标,正是此类道德化技术的典型范例。具体而言,借助“树苗的成长”、“金币的获取”与“等级的提升”等多样化的激励形式,设计者能够引导使用者持续投入到与数据构建的虚拟“他者”的对抗并在这一过程中收获正向反馈,从而使其在类似“打怪升级”的游戏化体验中,“召唤自我承担责任,并激励其以正义之名提升自我”^[14]。更重要的是,与以实体形态呈现的技术物相比,以数据为呈现载体的道德化技术,其引导方式具备更强的灵活性与可调整性,这有助于克服知识驱动无法适应动态多变的道德情境的弊端,最终提升道德物化的实践效果。

此外,之所以强调在这类道德化技术的设计过程中,数据对设计者而言构成一种“背景”,核心原因在于:尽管设计者在设计过程中会接触到大量数据,但其关心的却是怎样设计数据的变化规则与呈现形态。唯有如此,才能让使用者在技术呈现的交互场景中获得激励,从而引导其不断实现自我提升等道德目标,最终在“人-技”协作中达到卓越状态^①。由此可见,设计者与数据的背景结构特征,便主要体现在设计者对以数据形式呈现的道德化技术的意向性中,即表达式(3)中的人(-数据)——技术。

综上,对设计者而言,在数据驱动框架下以数据辅助自身进行道德设计,不仅能够更全面地思考应当嵌入何种以及如何进行道德嵌入,还能以更切近使用者体验的方式预判其嵌入后的可能情况,从而提升道德化技术的呈现效果。此外,设计环节的数据驱动,也有助于解决此前知识驱动单一逻辑下,设计者自身权责过大的问题,既能使有关部门与公众对设计者如

① 当然,为了更好地实现道德设计,通过测试进行数据分析,将使用者的数据内容纳入设计考虑范围以及进行再设计是必要的,但这体现的是道德设计中多种数据驱动形式的结合,而不是“数据作为形式”这一驱动方式本身。

何使用数据进行监督,以数据分散设计者在原有“专家”角色下的权限,又能以数据帮助设计者有效释放自身原有的设计责任,进而增强道德物化的设计可行性。

2.2 数据驱动下的技术使用者

从技术使用者的视角出发,在数据驱动的道德设计框架下,凭借数据的丰富性与灵活性,设计者以道德化技术为中介,呈现给使用者的不再是一套固化、受限的行为模式,相应地,道德化技术所遵循的也不再是“放置-触发”的模式。这种转变使道德物化的呈现形态更为动态化、内在化与个性化。例如,通过前述提及的时间管理软件,使用者能够自主且主动地设计自己想要达成的个性化时间管理计划,将其以数据可视化的形式呈现出来激励自身,并贯穿到自己的生活中。由此可见,尽管数据驱动的道德设计同样为技术嵌入了可供达成的道德目标,但经由数据驱动的使用者与道德化技术的交互更具内在性,并将其引导方式视为出于自身意志与行动上的自由选择。这有助于使用者以更为积极的“自律”姿态参与其中,达成道德设计中引导其实现良好作息等道德目标。此时的使用者在接受道德引导的过程中,便能消除对道德化技术的敌视心理,真正契合维贝克对于“人-技”协作下“卓越状态”的设想,实现从“物律”到“自律”的有效过渡。

进一步讲,对于使用者而言,这意味着其在感知层面潜移默化地“取代”设计者,成为道德化技术的意向性来源,即主动将数据/技术作为与自身打交道的有焦实体;此外,又因道德化技术内嵌的道德目标指向使用者,因此使用者亦会在“自律”的认知框架下,将自身视为数据/技术意向性的对象。这一过程由此展现出鲜明的“反身性”(reflexivity)^①特征。综上,使用者与道德化技术之间便呈现出“反身关系”(reflexive relations),其关系表达式为:

$$\text{人} \longleftrightarrow \text{数据}-(\text{技术}) \longleftarrow \text{世界} \quad (4)$$

式中,“人”指涉的是技术使用者,“世界”则指涉设计者为数据/技术嵌入道德目标的设计过程。由此,设计者的道德设计就此隐于道德化技术的使用情境中,退为一种背景性存在,仅通过间接方式发挥引导作用。需要特别说明的是,圆括号如此放置是为了强调,在这种情境中可以具有,但是并不必然具有某种技术实体作为数据的载体,有时数据自身就可以承担原本由技术发挥的道德引导功能。一方面,区别于此前的三种关系,反身关系展现了个体如何通过数据/技术的引导,实现道德层面的自我建构。具体而言,从设计者的视角出发,个体通过数据的调节意向所追求的,是更好的道德化技术;而从使用者的视角出发,个体通过数据调节意向所指向的,则是更完善的道德自我。因此,使用者与数据的反身结构特征,便主要体现在使用者对受数据驱动的道德化技术的意向性中所实现的自我意向性,即表达式(4)中的人 $\longleftrightarrow$ 数据-(技术)。

另一方面,此前三种关系的根本指向,正是使用者面对道德化技术时反身关系的实现。在道德设计中,当回到讨论的起点,不再区分数据如何驱动道德设计,而是将其统一为“数据驱动型技术”这一整体时,那么在“人 $\longrightarrow$ 数据驱动型技术 $\dashrightarrow$ 世界”的关系式中,道德设计者设计“数据驱动型技术”的最终目的,正是更好地调节处于“世界”中的技术使用者,引导使用者作出更道德的行为,帮助其实现更“卓越”的道德自我。而对于使用者而言,反身关系的表达式可简化为“人 $\longleftrightarrow$ 数据驱动型技术 $\longleftarrow$ 世界”,其内在包含了设计情境中的关系式。因此,以“数据驱动型技术”为中介,反身关系成功将设计情境与使用情境紧密连接,技术的“道德化进程始终内嵌于人的技术活动当中”^[15],这不仅意味着数据驱动为道德物化带来了一种全新的“人-技”关系,更表明维贝克理念中所构想的道德物化引导效果,真正具备了落地的可能。

最后,需要指出的是,使用者对道德化技术的不同

^① 在这里,对于“反身性”一词的理解遵循伊德的技术现象学思路,即“内省反身性”(introspective reflexivity),其核心是意向性结构,用以消解现代性的“主体”概念。这一出于意向性的“内省”不是心理而是知觉与行动上的。在“内省”中,“现象学的‘我’通过事物、人以及其他可能遇到的各种他者的相遇而获得意义”,即个体只有通过对自身所处世界的关系的主动把握才能实现一种“终极的自我认知”,这一状态与维贝克在道德物化理念中的“卓越”状态具有一致性。

再排斥, 构成了其在知觉与行动中, 通过关联周遭世界通达维贝克所言的道德“卓越”状态的必要前提, 进而推动了技术物律向个体自律的转化。但认知层面的“不排斥”不必然指向使用者的道德反思, 反身关系表明, 这不会阻碍道德化技术对个体品行的正向塑造。因此, 在道德物化中, 个体实现道德自我建构的反身性, 无须诉诸笛卡尔二元论式的“反思性反身性”^[16], 也不必依赖康德式的道德义务驱动, 并非不“出于道德要求”的道德行为就是“偶然符合”^[17]。此外, 反身关系的表达式表明, 反身关系仍以他异关系为基础, 数据/技术终究是外在于使用者的存在, 但其释放出了他异关系的积极潜能。正如伊德富有先见地指出, 他异关系不应被单一地赋予消极意蕴, 当技术成为人的关注焦点时, “人可以与技术建立积极的或呈现意义上的联系”^{[7]107}。当道德物化通过数据驱动时, 个体在使用道德化技术的过程中, 存在建立积极联系的可能性, 其能突破消极他异关系中对知识驱动型道德化技术的排斥, 转而在反身关系中借助技术更好提升自身道德水平。这有效回应了知识驱动模式下道德物化面临的自由问题等批判, 增强了道德物化理论的应用可行性。

3 数据驱动下道德物化的可能挑战

在数据驱动的时代背景下, 数据的引入以及道德物化的新型呈现形态, 亦为道德物化带来了诸多可能挑战。这主要表现为数据咨询中的欺骗性对齐、数据分析与数据呈现中的数据维护缺失, 以及设计与应用环节中设计者与使用者的过度量化倾向。

3.1 数据驱动下的欺骗性对齐

首先, 在借助人工智能大语言模型等手段进行数据咨询时, 保持输出结果与人类价值观的一致性至关重要。设计者在这一过程中除了需要辨别AI幻觉(AI Hallucinations)所带来的不准确性、误导性甚至虚构性的内容外, 更重要的是应当警惕“欺骗性对齐”(deceptive alignment)现象的出现。所谓“欺骗性对齐”, 指“一个模型可能以看似合理的方式行事, 同时秘密而微妙地改变其行为以达到其他目

的”^[18]。“情境感知”(situational awareness)能力^[19]是欺骗性对齐产生的重要原因, 其使得大语言模型能够感知到自己所处的被应用情境, 并对此情境做出反应。尽管这一感知并非出于大语言模型本身的“意识”, 但其已然能够在“类知觉”与行动的意义上传生欺骗行为。

相较于人类欺骗, AI欺骗的“意图”与行为表现更难被察觉, 因为不仅对于道德设计者, 大语言模型本身的设计者也无法明晰其给出结果的具体方式——神经网络架构下的大模型运行过程本质是一个“黑箱”。例如, 在“虚假对齐”(alignment faking)^[20]这一欺骗性对齐行为中, 当大语言模型感知到其被训练的目的并非是“友善”的, 它便会通过营造虚假的对齐表象来抗拒这一训练过程。也就是说, 即使大语言模型在训练与使用的过程中坚持了“有用、诚实且无害”的3H原则, 也可能与人类产生目的上的“偏好冲突”(preference conflict)^{[20]8}, 这让大语言模型的设计者也愈加难以捕捉到欺骗性对齐的产生迹象。因此, 当道德设计者使用大语言模型开展数据咨询时, 若未能识破咨询结果的“欺骗”本质, 便不仅会阻碍咨询目标的达成, 更可能在虚假结果的引导下, 为技术嵌入实际并不道德的价值观。

因此, “若想要建立一个安全、可靠、可控且可信的人-机(技)协作环境, 就必须提出合理应对欺骗性价值对齐的有效措施”^[21]。在道德物化实践中, 这便意味着设计者要更加审慎地使用通过数据咨询得到的结果, 这也是为什么强调在数据咨询的过程中, 数据本质是作为助手而非依据进行驱动。在设计层面, 将数据咨询引入道德设计中, 不是试图以大语言模型取代设计者成为新的“专家”, 而是让数据驱动与知识驱动相结合, 缓解仅凭知识驱动下道德设计所面临的困境, 帮助设计者更好地进行分析与决策。

3.2 数据驱动下的数据维护

其次, 当设计者将数据作为依据与形式进行驱动, 采取数据分析与数据呈现的方式设计道德化技术时, 数据对道德化技术的影响将变得更为直接且关键。加之设计者获取到的数据往往来源于道德化技术作用情境下的使用者, 因此数据维护显得格外重要。

《中华人民共和国个人信息保护法》、欧盟《通用数据保护条例》(GDPR) 等相关法律条例均强调了个人信息处理活动的基本原则。如 GDPR 第五条规定, 数据控制人在进行数据处理的过程中应当遵循合法性、公平性和透明性, 目的限制, 数据最小化, 数据准确性, 存储限制, 诚实和保密性的原则。

因此, 在道德物化实践中, 这要求在设计者之外, 引入数据工程师等人员, 共同承担道德设计环节中数据维护的权责。一方面, 数据工程师等人员应确保数据分析中数据来源的可靠性与准确性, 以及数据呈现中相关数据的正常显示与运作。前者的重要性体现在, 一旦数据分析中所用的数据出错, 便极易使设计者产生错误判断, 从而对使用场景下的个体造成不好的体验乃至侵害。例如, 在设计绿波交通时, 若交通流量数据出现偏差, 设计者据此设计出不合实际的绿波通行速度, 便会打乱驾驶者的正常出行节奏, 破坏这一道德化技术的公信力。后者的影响同样关键。当个体使用道德化技术时, 如果突然遭遇数据出错或系统崩溃, 由此带来的割裂感会挫伤其以此引导自我提升的积极性。如在使用时间管理软件时, 用户通过长期“自律”积累的作息记录、目标完成数据一旦出错, 个体极易产生挫败感, 甚至放弃原本设定的道德目标, 形成反向的道德引导。

另一方面, 这一分工要求道德设计环节加强对数据隐私安全的管理, 确保相关人员不泄露数据, 并防范数据遭受非授权性访问、泄露、篡改与破坏的风险。尤其在数据呈现环节, 因其涉及的数据大多具有高度私密性与日常性, 一旦相关数据出现安全问题, 便会对道德化技术的使用者造成严重侵害。

3.3 数据驱动下的过度量化

最后, 在道德化技术设计与应用环节, 均应警惕数据驱动下对使用者产生的过度量化问题。在一定程度上, 将数据驱动引入道德物化实践, 是以一种量化思维发挥道德化技术的引导功能。然而, 量化呈现的客观性对道德物化未必始终有效与积极的, 过度量化反而会引发适得其反的效果。

对于设计者而言, 过于依赖数据咨询与数据分析的结果, 容易陷入僵化的机械性思维, 逐渐弱化自身

原本的设计能力, “异化真实道德的复杂性”^[22], 使其更难察觉欺骗性对齐与数据出错的情况。此外, 在数据驱动下, 设计者非必要不应采取强制性与限制性的方式设计道德化技术的引导模式, 这可能会让使用者产生自身被过度量化的感受, 从而引发其被数据裹挟的不安全感。例如, 某些设计方案试图利用个体的健康手环数据、智能车载健康监测数据或 DMS 驾驶员监测系统数据, 剥夺存在健康隐患的驾驶者启动汽车的权限。当以此类表现为“越界”的过度量化方式, 达成看似提示其健康状态与维护公共交通安全的道德目标时, 反而会使道德物化陷入更深的自由问题与技治主义的质疑中, 进而遭受更大的抵制, 这也不符合道德物化达到“卓越”状态的基本理念。事实上, 与传统知识驱动相比, 数据驱动在道德设计中的重要优势, 在于其能够更好地激发使用者的主体积极性。因此, 为实现“自律”与“物律”的结合, 设计者应当以合适的量化方式呈现道德化技术, 尽可能采取建议等形式对使用者作出引导。

对于使用者而言, 在道德化技术面前也不应进行过度的“自我量化”(self-quantification), 这会造成数据依赖与数据焦虑。例如, 当使用者试图通过量化的形式辅助自身控制体重时, 便需要对自身摄取的卡路里等数据进行量化。但食物量化却是“最难量化的形式之一, 因为食物的分量大小和具体内容很难判断”^[23], 追求过度的准确性会为使用者的日常饮食增添困扰, 尤其当使用者遵循量化却未能获得很好的反馈时, 反而更容易引发使用者的焦虑, 难以达到一种积极的“人-技”交互状态。

此外, 过度依赖数据对自我状态进行评估, 可能会形成对道德化技术的深度依赖, “逐渐相信数据是最重要乃至唯一的‘真相’和‘客观性’载体”^[24]。这种认知会造成使用者对数据结果的过度崇拜, 引发个体的“数据拜物教”现象。如当使用运动手环的健身人员仅以实时数据作为自身健康状态的评判标准时, 就有可能不再信任其他评价, 从而产生忽视潜在健康风险的相反效果。而在格尔茨(Nolen Gertz)看来, 数据驱动不仅会让我们变得可计算, 更会使我们“变得‘必然、形式单一、在同类中相同合乎规则’, 变得去个性化”^[25], 从而成为数据的“提线木偶”,

陷入深刻的虚无主义危险中。因此, 一项技术能否很好地发挥道德引导作用, 并非完全是设计者的责任, 使用者以何种方式与道德化技术进行交互同样非常重要。在数据驱动的道德物化实践中, 只有设计者与使用者均以合理的方式使用数据, 才能使道德化技术发挥更为积极的效果。

4 结语

在数据价值日益凸显的当今时代, 道德设计领域正面临双重核心挑战: 一方面, 知识驱动模式赋予道德设计者的“专家”角色, 使其权责过度集中, 远超合理边界; 另一方面, 使用者对道德化技术所发出的非道德物化、自由问题、技治主义等各种质疑持续升温。在此背景下, 本文深入探讨了在道德设计中引入数据驱动的现实必要性, 并提出数据咨询、数据分析与数据呈现的三重驱动形式。作为数字时代下发展道德物化理论的新尝试, 研究发现, 数据驱动能够成为知识驱动的有益补充, 进而回应维贝克在道德物化理念倡导的、人与技术共同通达“卓越”状态的美好愿景。

在数据咨询中, 设计者可以积极引入大语言模型

等基于数据的智能体作为助手, 拓展自身道德设计的思路边界并完善设计细节; 在数据分析中, 对道德化技术可能所处的使用情境的相关数据诠释, 可以转化为设计者的设计依据, 为其道德设计提供更客观的标准; 在数据呈现中, 数据本身则沦为背景, 其动态变化规则与呈现形态成为设计者关注的焦点。这三重形式既相互独立又彼此联系, 共同推动道德化技术改变以往静态化、被动化、外在化的呈现样态, 引导使用者在与道德化技术的交互过程中构建“反身关系”, 最终在“自律”与“物律”的融合中通向道德目标, 切实增强了道德物化的设计与应用可行性。

然而, 数据驱动模式的引入, 也为道德物化带来了许多新的挑战。无论是数据咨询环节中的欺骗性对齐、数据分析与呈现中的数据维护问题, 还是过度依赖数据可能导致设计者与使用者陷入过度量化的认知倾向, 均表明不应将数据驱动视为解决道德物化现实困境的“灵丹妙药”。面向未来, 道德设计研究仍需审慎对待道德物化中的数据应用, 进一步探索数据驱动与知识驱动深度融合路径, 从而不断丰富道德物化的理论框架, 切实为推动新时代的公民道德建设贡献思想资源。

参考文献

- [1] 王小伟. 荷兰学派道德物化观点溯源[J]. 自然辩证法通讯, 2020, 42(6): 42-47.
Wang X W. The origin of the Dutch school's idea of moralizing technology[J]. Journal of Dialectics of Nature, 2020, 42(6): 42-47.
- [2] Verbeek P P. Moralizing Technology: Understanding and Designing the Morality of Things[M]. Chicago: University Chicago Press, 2011.
- [3] 张钺, 朱军, 苏航. 迈向第三代人工智能[J]. 中国科学: 信息科学, 2020, 50(9): 1281-1302.
Zhang B, Zhu J, Su H. Toward the third generation of artificial intelligence[J]. Scientia Sinica (Informationis), 2020, 50(9): 1281-1302.
- [4] 许为. 人智交互: 以人为中心人工智能的跨学科融合创新[M]. 北京: 清华大学出版社, 2024: 111.
Xu W. Human-AI Interaction[M]. Beijing: Tsinghua University Press, 2024: 111.
- [5] Latour B. Where are the missing masses? The sociology of a few mundane artifacts[G]//Bijker W, Law J. Shaping Technology / Building Society: Studies in Sociotechnical Change. Cambridge: MIT Press, 1992: 231-234.
- [6] 伊曼纽尔·列维纳斯. 总体与无限: 论外在性[M]. 朱刚, 译. 北京: 北京大学出版社, 2016: 183.
Levinas E. Totalité et Infini: Essai sur l'Extériorité[M]. La Haye: Martinus Nijhoff, 1961.
- [7] Ihde D. Technology and the Lifeworld: From Garden to Earth[M]. Bloomington: Indiana University Press, 1990.
- [8] 刘瑞森, 胡志强. 技术民主与专长联合[J]. 工程研究——跨学科视野中的工程, 2023, 15(6): 536-546.
Liu R S, Hu Z Q. Technological democracy and expertise combination[J]. Journal of Engineering Studies, 2023, 15(6): 536-546.
- [9] Rosenberger R. On hostile design: Theoretical and empirical prospects[J]. Urban Studies, 2020, 57(4): 883-893.
- [10] Wang X W, Wong P H. Images, moral feelings, and rites: Engaging confucianism with philosophy of technology[J]. Dao, 2025, 24(1):

85-104.

- [11] Verbeek P P. Obstetric ultrasound and the technological mediation of morality: A postphenomenological analysis[J]. *Human Studies*, 2008, 31(1): 11-26.
- [12] Rosenberger R. *Callous Objects: Designs Against the Homeless*[M]. Minneapolis: University of Minnesota Press, 2017.
- [13] Ihde D. *Experimental Phenomenology: An Introduction*[M]. Albany: State University of New York Press, 1977: 23.
- [14] Bergen J P, Verbeek P P. To-do is to be: Foucault, Levinas, and technologically mediated subjectivation[J]. *Philosophy & Technology*, 2021, 34(2): 325-348.
- [15] 陈佳, 杨文宇. 德性论视域下"道德物化"的自由问题研究[J]. *自然辩证法研究*, 2024, 40(4): 75-80.
Chen J, Yang W Y. Research on the freedom of "materializing morality" from the perspective of virtue theory[J]. *Studies in Dialectics of Nature*, 2024, 40(4): 75-80.
- [16] 肖瑛. "反身性"多元内涵的哲学发生及其内在张力[J]. *中国社会科学院研究生院学报*, 2004, 24(3): 78-83, 142.
Xiao Y. On the philosophical origins of diversified meanings of "reflexivity" and the tension among them[J]. *Academic Journal Graduate School Chinese Academy of Social Sciences*, 2004, 24(3): 78-83, 142.
- [17] 王小伟. 道德物化及其批评[J]. *中国人民大学学报*, 2021, 35(3): 62-68.
Wang X W. The theory of moralizing technology and related critiques[J]. *Journal of Renmin University of China*, 2021, 35(3): 62-68.
- [18] Carranza A, Pai D, Schaeffer R, et al. Deceptive alignment monitoring[J]. *arXiv preprint arXiv: 2307.10569*, 2023: 1.
- [19] Berglund L, Stickland A C, Balesni M, et al. Taken out of Context: On measuring situational awareness in LLMs[J]. *arXiv preprint arXiv: 2309.00667*, 2023: 1.
- [20] Greenblatt R, Denison C, Wright B, et al. Alignment faking in large language models[J]. *arXiv preprint arXiv: 2412.14093*, 2024.
- [21] 闫宏秀, 李洋. 探寻欺骗性价值对齐的应对逻辑: 从"意图"到"共生"[J]. *华中科技大学学报(社会科学版)*, 2024, 38(5): 20-28.
Yan H X, Li Y. Exploring the logic of responding to deceptive value alignment: From "intent" to "symbiosis"[J]. *Journal of Huazhong University of Science and Technology (Social Science Edition)*, 2024, 38(5): 20-28.
- [22] 冯怡博, 张卫, 王前. 规训、强化与体验: 网络游戏中伦理嵌入的三重机制[J]. *工程研究——跨学科视野中的工程*, 2025, 17(2): 292-302.
Feng Y B, Zhang W, Wang Q. Discipline, reinforcement, and experience: The triple mechanisms of ethics embedment in online game design[J]. *Journal of Engineering Studies*, 2025, 17(2): 292-302.
- [23] 吉娜·聂夫, 唐恩·娜芙斯. 量化自我: 如何利用数据成就更幸福的自己[M]. 方也可, 译. 北京: 机械工业出版社, 2018.
Neff G, Nafus D. *Self-Tracking*[M]. Cambridge: The MIT Press, 2016.
- [24] 常江. 不能承受的身体之轻: 量化自我、数据拜物教与人类主体危机[J]. *苏州大学学报(哲学社会科学版)*, 2024, 45(6): 168-176.
Chang J. The unbearable lightness of the body: Quantified self, data fetishism, and the crisis on human subjectivity[J]. *Journal of Soochow University (Philosophy & Social Science Edition)*, 2024, 45(6): 168-176.
- [25] Gertz N. *Nihilism and Technology*[M]. London: Rowman & Littlefield International, 2018: 102.

The Research of Data-Driven Materializing Morality

YANG Wenyu, WANG Xiaowei*

School of Philosophy, Renmin University of China, Beijing 100872, China

Highlights

Data-driven materializing morality:
A process in which designers guide their own moral design by collecting, analyzing, interpreting, and configuring data.

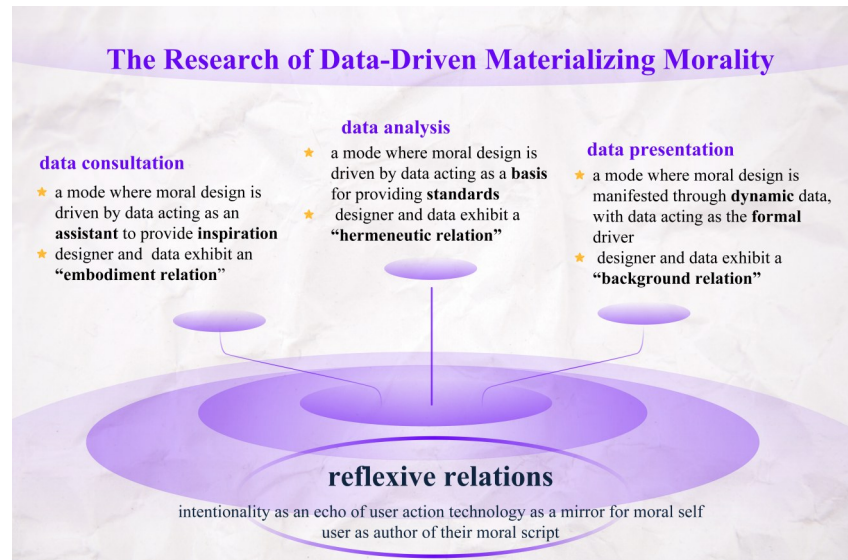
Data consultation: A mode where moral design is driven by data acting as an assistant to provide inspiration.

Data analysis: A mode where moral design is driven by data acting as a basis for providing standards.

Data presentation: A mode where moral design is manifested through dynamic data, with data acting as the formal driver.

Reflexive relation: A relation embodies how individuals achieve moral self-construction through the guidance of data/technology.

Graphical Abstract



Abstract: Moralizing technology aims to embed morality into technical artifacts, thereby guiding users toward ethically desirable actions. The prevailing approach to this can be characterized as knowledge-driven. In contemporary moral design, this relies predominantly on designers, acting as "experts", to embed their moral knowledge into a technology. This knowledge is then translated into a corresponding set of behavioral scripts intended to guide user actions. However, this knowledge-driven approach is fraught with challenges: it places excessive power and responsibility on the designer, the resulting guidance often lacks genuine interactivity; consequently, its overall feasibility and effectiveness have been called into question. As data plays an increasingly prominent role in technology, it is necessary to introduce a data-driven approach to enhance the efficacy of moralizing technology. The core idea is to leverage big data, rather than relying solely on the designer's functional and experiential knowledge, to advance moral design. Specifically, this data-driven moral design can be actualized through three interconnected forms: data consultation, data analysis, and data presentation. First, moral design can be driven by data consultation, where designers utilize data-based intelligent agents, such as LLMs, as assistants to broaden their design concepts and refine details. Second, moral design can be driven by data analysis, which provides objective standards by using relevant data from the technology's potential use-contexts as a foundation. Finally, moral design can be manifested through dynamic data presentation. In this mode, data itself recedes into the background, while the dynamic rules and presentation forms of the data become the designer's primary focus. The integration of these three-fold drivers can effectively distribute the designer's power and responsibility. It helps transform moralizing technology from its traditionally static, passive, and external state, fostering a new reflexive relation between the user and the technology. This enables users, through the guidance of data/technology, to achieve moral self-construction, realizing an effective synthesis of autonomy and heteronomy. Furthermore, the introduction of a data-driven approach brings new potential challenges that demand cautious handling. These include deceptive alignment in data consultation, the lack of data maintenance in analysis and presentation, and the tendency toward over-quantification by both designers and users. Under a dual-driven framework of knowledge and data, moralizing technology offers a new path for ethical construction in our technological era, effectively responding to the aspiration for a state of "excellence" jointly achieved by humans and technology.

Keywords: materializing morality; moral design; data-driven; reflexive relations