

算法诱导成瘾的现状和未来*

刘 涛^{1,2,3}, 陈彦佐¹, 何 琳⁴, 杨一凡¹, 王 卿³, 于晓宇^{1,2,3}

(1. 上海大学管理学院, 上海 200444; 2. 上海企业创新与高质量发展研究中心, 上海 200444;

3. 上海大学 MBA 中心, 上海 200444; 4. 杭州电子科技大学管理学院, 杭州 310018)

摘 要:随着人工智能技术的深度社会化应用, 算法诱导成瘾 (Algorithm-Induced Addiction, AIA) 作为一种新型行为依赖现象日益显现。本文梳理了 AIA 的现状、影响、技术基础与生物机制, 并提出“神经-行为双层耦合模型”, 通过整合成瘾神经临床评估三维功能域与四类成瘾行为表型构建统一分析框架, 并从政策设计、跨学科研究与反成瘾算法开发三个方面提出未来发展方向。本文旨在为构建算法向善的治理范式提供理论支持, 推动人工智能从“成瘾智能”向“认知增强型智能”转型。

关键词:算法诱导成瘾; 神经-行为双层耦合模型; 反成瘾设计; 数字伦理

中图分类号: B848

文献标志码: A

文章编号: 1003-5184(2026)02-0099-08

1 引言

人工智能的深度社会化应用在重塑人类生活范式的同时, 正催生一种新型行为依赖危机——算法诱导成瘾 (Algorithm-Induced Addiction, AIA)。随着智能生产工具、情感陪伴机器人和娱乐消费平台等全领域渗透, 算法驱动的数字产品通过神经奖赏机制强化与预测性行为干预等技术, 持续诱发用户对社交媒体、短视频及游戏等非实物载体的过度依赖, 对个体与社会构成多维度的负面影响 (Yu et al., 2024)。在生理层面, 游戏成瘾青少年前额叶皮层发育因持续性注意力耗散而受到干扰, 削弱其自我调节能力 (Skok et al., 2024)。在心理层面, 社交媒体算法主动向青少年推送极端内容, 导致青年群体抑郁风险显著提升 (Costello et al., 2023)。而在社会经济维度, 《经济学人》(The Economist) 预计美国因数字分心导致的年生产力损失超过 1 万亿美元。更值得警惕的是, 商业实践表明这些危害并非无后果。

与传统的物质成瘾或行为成瘾有所不同, AI 成瘾指的是人类对基于 AI 算法的大模型、虚拟人和智能体等非实物载体的依赖 (Kooli et al., 2025)。当前, 学界已有大量研究讨论 AI 与人类交互所产生的

问题, 但既往研究多是对人工智能在不同领域引发的成瘾现象进行分散阐述, 或是对单一成瘾类型和模式的讨论, 缺乏对这一复杂问题的系统性概念整合及其内在机制剖析。这可能导致两个问题: 其一, 现有研究未能界定 AI 的主动诱导性: 算法并非被动响应用户需求, 而是通过自学习功能主动诱导人类产生成瘾行为; 其二, 尚未建立人工智能伦理与行为成瘾的链接机制 (谢潇等, 2025)。

基于此, 本文提出算法诱导成瘾概念 (Algorithm-Induced Addiction, AIA), 指算法系统基于人类本能机制, 通过自学习动态优化刺激策略主动培育用户依赖性行为的现象。在回顾既有文献的基础上, 本文进而提出“神经-行为双层耦合模型”, 通过整合成瘾神经临床评估的三维功能域与四类成瘾行为表型, 构建了 AIA 的统一分析框架。同时, 面对 AIA 负面影响的持续增长, 本文为呼吁学界重视、制定干预策略和引导反成瘾技术向善提供理论支撑, 推动人工智能从“成瘾智能”向“认知增强型智能”转变。

2 算法诱导成瘾的研究现状

2.1 算法诱导成瘾的现状和影响

人工智能算法的演进显著加剧了行为成瘾及其

* 基金项目: 国家自然科学基金面上项目“人工智能时代企业人机协同决策的演进机制”(72472094), 国家自然科学基金青年项目(C类)“人机协同对创业项目投资决策的影响机制研究”(72502144)。

通信作者: 陈彦佐, E-mail: yanzuochen@shu.edu.cn。

社会危害性。国内外既有研究已揭示算法成瘾跨领域的负面影响。在消费领域,个性化推荐系统可通过诱发用户的心流体验显著刺激冲动购买行为(范文芳等,2022)。在娱乐领域,TikTok通过系统设计诱导用户进入“沉浸状态”,进而形成成瘾行为(Qin等,2022)。社交层面,社交媒体情感依恋算法通过虚拟依恋关系操纵消费意图(Bilal et al., 2024);而错失恐惧则通过社交网络的整合性使用间接诱发成瘾行为(马建苓等,2019)。更为严峻的是,青少年群体由于神经发育尚未成熟成为高风险人群。Xiao等人(2025)研究显示,青少年对社交媒体、手机使用和电子游戏的高成瘾与自杀行为、自杀意念和不健康的心理状态风险增加 22% ~ 139%。平台主动推送导致的社交焦虑与大学生自身负性情绪信息注意偏向在手机成瘾与大学生抑郁之间起链式中介作用(侯娟等,2021)。

2.2 算法诱导成瘾的技术基础

算法诱导成瘾的技术基础源于人工智能算法中“强化学习”、“生成对抗网络”和“图神经网络”三类核心技术所构建的操控逻辑,其核心实现方式包括个性化推荐系统与行为设计技术,通过精准捕捉用户偏好、强化即时反馈以及满足深层心理需求来实现对用户行为的持续诱导。

在具体操作层面,基于用户行为数据,个性化推荐系统通过协同过滤与内容推荐算法构建动态模型。例如,K最近邻分类模型通过实时更新用户画像标签,实现内容精准匹配(张炎亮等,2020)。TikTok算法系统通过分析用户停留时长、互动频率等数据生成个性化视频流,诱发“无限下拉”行为,这种高度适配的内容供给显著提升用户沉浸感,并导致时间知觉扭曲(Jain et al., 2025)。随着生成式AI的深入应用,个性化推荐系统对用户行为的成瘾性诱导进一步加剧(Zhou et al., 2024)。行为设计技术则通过界面工程削弱用户的自控能力,如采用无限刷新和自动播放等交互设计策略,诱导用户成瘾(孙瑜晨,2022)。大量社媒平台采用蔡格尼克效应,利用未完成任务记忆留存效应,通过红点提示、进度条等视觉符号诱发用户的强迫性使用行为。除此之外,游戏研究公司 Naavik 2025 年的最新研究表

明,游戏中公会和战队系统通过排行榜、贡献可视化来增加玩家的社会压力,进而增加游戏时长并提升付费行为。

2.3 算法诱导成瘾的生物机制

世界卫生组织将成瘾定义为个体与特定物质或行为反复交互所形成的精神与生理依赖状态,其核心表现为强迫性寻求行为、戒断反应、耐受性提升以及社会功能损伤。现代医学将成瘾划分为物质成瘾与行为成瘾,二者在神经机制上既有相似性也存在差异(何全兴等,2025)。行为成瘾虽不涉及外源性物质,但其神经机制与物质成瘾高度重叠(Volkow et al., 2016)。进一步,Griffiths(1995)将技术成瘾(TAs)定义为涉及人机交互的非化学性(行为性)成瘾,可分为主动型和被动型。Turel等人(2011)则将技术成瘾概念化为对技术使用的不良心理依赖,且这种依赖会达到出现行为成瘾症状的程度。尽管这些定义存在差异,但均强调了技术的参与以及成瘾所表现出的行为症状,同时明确技术成瘾是数字时代技术进步背景下行为成瘾的一个子集,包含网络游戏障碍、网上购物成瘾、社交媒体成瘾等类型,其中网络游戏障碍已被世界卫生组织纳入《国际疾病分类》第 11 版(ICD-11)(Huang & Huang, 2025)。目前,国内外对算法诱导成瘾的机制研究较少,但可以依照先前技术成瘾和 AI 成瘾的定义,将其归为行为成瘾的一种以进行生物机制分析。

首先,多巴胺奖赏回路失调是成瘾的核心机制。成瘾行为会刺激腹侧被盖区(Ventral Tegmental Area, VTA)释放多巴胺至伏隔核(Nucleus Accumbens, NAs),从而产生强烈愉悦感。反复刺激会导致 D2 受体下调,形成耐受性,用户需接受更强刺激才能激活同等程度的奖赏效应(Volkow et al., 2015; Koob et al., 2008)。典型表现为,在短视频浏览过程中,即时性与多样性内容持续刺激 VTA 和 NAs,触发多巴胺大量释放。用户对“下一条视频更好看”的期待形成变率强化机制,长期高频刺激导致多巴胺受体敏感性下降,用户需延长使用时间才能获得同等愉悦感。功能磁共振脑成像(fMRI)研究表明,TikTok 的个性化视频推荐会激活默认模式网络和腹侧被盖区(与奖赏加工密切相关),其神经激活强度与

成瘾行为呈正相关(Su et al., 2021)。

其次,前额叶认知控制功能的抑制是另一关键机制。背外侧前额叶皮层(dorsolateral prefrontal cortex)负责执行功能、自控调节和长期规划(Méndez et al., 2024)。长期成瘾可导致前额叶、眶额皮层和辅助运动区的灰质密度下降。短视频的频繁切换会占用大量认知资源,降低对复杂任务的专注力,进而导致该区域灰质密度逐年下降,最终引发执行功能的退化,使用户陷入明知有害仍无法自控的困境(Yuan et al., 2011)。

2.4 现有研究总结

现有研究从多个维度对算法诱导成瘾的现状影响、技术基础与生物机制进行了探讨。可以看出,“AI”不仅是“人工智能”(Artificial Intelligence),也可能演变为“成瘾智能”(Addictive Intelligence)。在消费、社交与娱乐等领域存在大量算法诱导成瘾的实证证据。在技术层面,个性化推荐系统与行为设计技术被证明具有成瘾强化作用。而在生物机制方面,多巴胺奖赏通路与前额叶抑制功能的神经基础也得到了实证支持(Park et al., 2017)。

然而,当前研究存在三方面局限性。首先,理论解释呈现碎片化特征。现有研究多聚焦短视频成瘾、游戏成瘾或大语言模型成瘾等独立场景,缺乏对算法如何系统性诱导用户依赖的统一理论框架。其次,机制研究存在不均衡现象,当前研究过度依赖行为成瘾中多巴胺通路模型,而对数字成瘾、技术成瘾和算法诱导成瘾所特有的特征关注不足。最后,缺乏统一的诊断标准,学界尚未就算法成瘾与其他成瘾类型间的关系、评估工具及诊断框架达成共识,即对定义、产生过程和判断方法仍不清晰,导致概念混淆与研究整合困难。尤其值得关注的是,生成式人工智能通过基于人类反馈的强化学习训练,其自学习能力大幅跃升,成瘾力显著高于传统推荐算法,亟需构建创新性理论框架加以应对。

基于此,本文提出“神经-行为双层耦合模型”。第一层为神经功能层,主要依托成瘾神经临床评估(Addiction Neuroclinical Assessment, ANA)的三域分类框架;第二层为行为表型层,整合情感依赖、认知替代、行为强化与社交代偿四类行为表型。

中间层为算法中介路径。该模型旨在建立算法刺激与神经响应和行为表型之间的系统关联,为开发反成瘾技术方案及制定相关政策法规提供理论支撑。

3 神经-行为双层耦合模型框架

图1展示了算法诱导成瘾的神经-行为双层耦合模型的系统框架。模型整合神经功能评估与行为表型分析,构建从机制到表现的层级解释路径。

第一层:神经功能层。ANA是一种基于神经科学的成瘾障碍评估框架,旨在从神经生物学维度系统评估和分类成瘾行为,而非仅依赖传统的症状描述。ANA能够识别不同神经机制驱动的成瘾亚型,将成瘾的神经机制解构为三个神经功能领域:执行功能(executive function, EF)反映预期与专注阶段的抑制控制与决策能力;激励显著性(incentive salience, IS)对应成瘾与放纵阶段的奖赏加工与动机突显;负性情绪(negative emotionality, NE)则表征戒断阶段的情绪失调状态(Kwako et al., 2016)。需要指出的是,ANA原用于物质成瘾的特质性缺陷评估,而算法诱导成瘾更强调交互塑造性,即算法环境通过持续反馈逐步诱导神经功能变化,而非仅放大预存的缺陷。因此,本模型将ANA三域重新定义为神经功能塑造指标,用于评估个体在算法交互中的功能动态变化。

第二层:行为表型层。基于现有研究,算法诱导成瘾表现为四类行为模式:(1)情感依赖型成瘾:大语言模型驱动的人工智能程序可以培养类似于人类关系的极端情感依恋和依赖,用户将生成式AI作为情感代偿工具,要求其扮演恋人或心理医生等,与恋爱成瘾的神经响应高度相似(Chen et al., 2025; Freitas et al., 2025)。(2)认知替代型成瘾:表现为用户将写作、决策等认知任务外包给AI,导致自主认知能力退化(Zhou et al., 2026)。(3)行为强化型成瘾:算法通过分析用户情绪与语言习惯,动态优化对话策略以提升粘性,与此同时的是推荐流沉迷,最终导致用户难以抵抗并逐渐强化成瘾,很难停止短视频、信息流刷新(Jain et al., 2025)。(4)社交代偿型成瘾:用户以AI替代现实社交关系,尤其高发于社交焦虑人群,表现为主动减少人际接触、将AI互动量作为社交指标(Marriott et al., 2024)。

中间层:算法中介路径。ANA 框架虽然能够识别个体在不同神经功能维度上的成瘾差异,却无法解释算法交互如何将这差异转化为特定行为。已有算法通过多种中介机制实现这一转化,例如可强化调度(如间歇性奖励)、情绪韵律同步(如情感计算)、功能通道替代(如 AI 工具化)等。个体神经功能的变化与缺陷通过这些机制被算法环境选择性强化,逐步固化为稳定的行为表型。

综上,本文构建了一个神经-行为双层耦合模型,第一层是 ANA 三域(执行功能、激励显著性、负性情绪)的神经功能缺陷,第二层是情感依赖、认知

替代、行为强化、社交代偿四类具体行为,中间层则是算法诱导的中介路径。同时,成瘾行为产生的新数据流易被算法系统重新捕获,作为优化下一轮交互的输入信号。算法据此进一步精准化个体适配策略,形成神经功能-算法响应-行为强化-算法再捕捉-神经重塑-算法再适配的负循环。因此,结合成瘾个体在 EF、IS、NE 三个神经功能维度的具体状态及其对应的行为表征,可以制定具有针对性的算法干预措施。同时,精准识别神经功能与行为表型之间的算法作用路径,有助于为算法诱导成瘾提供科学有效的治理策略。

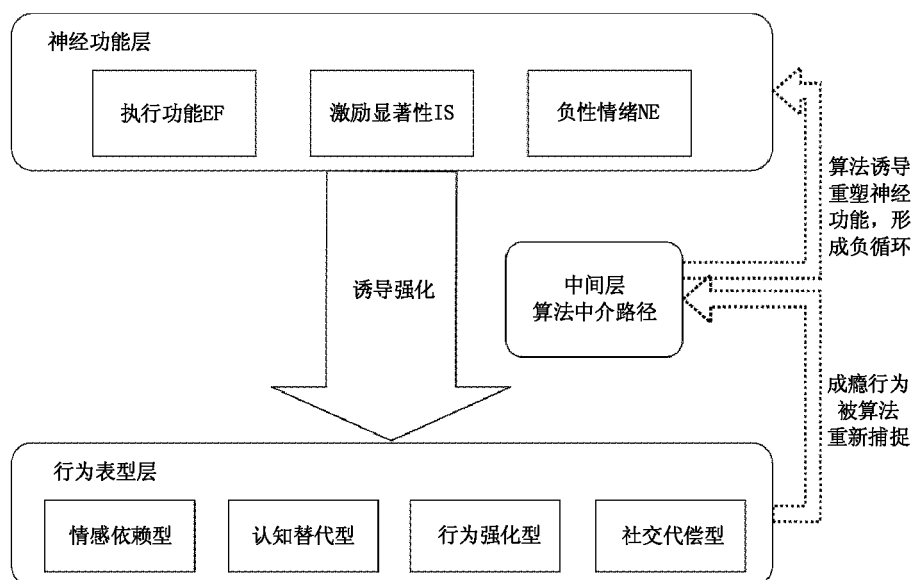


图1 神经-行为双层耦合模型

EF = 执行功能 (executive function); IS = 激励显著性 (incentive salience); NE = 负性情绪 (negative emotionality)

4 算法诱导成瘾治理难题

当前,算法诱导成瘾的治理面临技术快速演进与监管体系滞后的矛盾。研究表明,其危害呈现出跨领域广泛渗透。例如,最新的研究中短视频成瘾会改变个体默认模式网络的功能连接,对大脑造成伤害(Gao et al., 2025),大语言模型辅助写作的过度依赖会损害长期教育效能(Kosmyrna et al., 2025)。随着更高度智能的 AI 到来,此类现象将愈发增多,凸显了构建统一的算法诱导成瘾模型的紧迫性与必要性。

然而,当前国内外针对算法成瘾的治理响应仍呈现碎片化特征。在病理治理层面,尽管世界卫生组织已于 2020 年将数字成瘾列为全球公共卫生问

题,但主流诊断体系(如 DSM-5、ICD-11)尚未正式纳入数字成瘾类别,导致诊断标准不统一、概念界定混乱与术语使用不一致(Almourad et al., 2020)。在行政治理层面,监管尝试已显现:中国网信办要求平台公示推荐算法逻辑,要求标注 AI 生成;欧盟《数字服务法》明确禁止利用情绪识别技术定向推送赌博广告。然而,整体法律框架仍处于局部探索阶段,尚未形成系统性治理机制。

与此同时,企业在自我规制方面也存在根本性局限。尽管数字健康功能,如使用时长提醒、强制休息机制已被广泛采用,但研究表明,这些措施仅具有短期新奇效应,用户易产生适应性规避行为,且缺乏对其长期有效性的系统追踪(Cemiloglu et al.,

2022)。基于上述治理困境,准确识别算法诱导成瘾的作用路径,并构建反成瘾算法,已成为当前治理策略中不可或缺的重要方向。

5 未来研究方向

当前关于算法诱导成瘾的研究仍处于探索阶段。我国对成瘾性算法的综合性研究较少,而学者们已经肯定了对成瘾性算法进行规制的必要性,对现实生活算法运营商利用“上瘾”模式设计算法这一长久未引起重视的现象提出疑问。但是对于成瘾性算法的具体运行机制以及对成瘾性算法的规制方法,仍然处于起步阶段(魏亚楠,2024)。面对人工智能技术的迅猛发展与成瘾机制的复杂性,未来研究亟需在理论整合、跨学科协同与技术治理等方面实现突破。首先,在理论构建方面,应进一步完善算法诱导成瘾的统一概念框架,推动其从现象描述向机制揭示与分类整合的纵深发展。现有研究多聚焦于特定平台或成瘾类型,如短视频成瘾、社交媒体依赖或生成式 AI 的情感绑定,缺乏对算法如何主动诱导用户形成依赖的理论整合。因此,在“神经-行为双层耦合模型”的基础上,应进一步拓展对不同技术场景中神经响应机制和个体行为模式变化的研究,构建更具解释力与预测力的理论模型,为后续干预策略提供理论支撑。

其次,算法诱导成瘾本质上是一个跨学科议题,涉及神经科学、心理学、计算机科学和伦理学等多个领域,未来研究应加强学科交叉与方法融合。一方面,可通过神经影像技术和生理测量手段,结合大规模用户行为数据,系统揭示算法刺激与神经反应间的关系,明确不同行为表型背后的神经基础。另一方面,应深入挖掘算法设计中的底层机制,借助计算建模与算法逆向工程手段,提升对算法诱导机制的可解释性与可控性。此外,还需关注个体差异与文化背景对成瘾行为的影响,推动模型在不同人群的适应性拓展,提升研究成果的普适性与应用价值。

在技术与政策治理层面,未来研究应聚焦反成瘾技术的开发与治理机制的创新。尽管当前平台已推出部分数字健康功能,但其效果有限,用户易产生适应性规避行为。因此,亟需开发基于人工智能的反成瘾算法,通过情绪识别、行为轨迹分析等手段,

实时识别用户成瘾倾向,并采取个性化干预措施,如内容替换、使用限制或认知引导,帮助用户建立健康的使用习惯。同时,应推动算法设计从“刺激用户多巴胺”向“增强用户执行功能”转变,使 AI 真正成为促进人类认知发展的工具而非依赖的源头。在政策层面,应推动将算法诱导成瘾纳入公共卫生与数字治理的法律框架,制定平台算法伦理标准,明确企业在算法设计中的社会责任。此外,还需加强用户教育与行业自律,提升公众特别是青少年群体对算法诱导机制的认知能力,构建多方协同治理生态体系。

综上所述,算法诱导成瘾的研究不仅关乎个体心理健康与行为控制,更涉及人工智能技术发展的伦理边界与社会影响。未来研究应在理论深化、技术突破与政策创新三方面协同推进,推动人工智能从“成瘾智能”向“认知增强型智能”转型,为构建人机良性协同的数字社会提供理论与实践支持。

6 讨论与结论

6.1 理论意义与实践价值

本文提出的神经-行为双层耦合模型将成瘾神经临床评估的三维功能域与四类算法诱导成瘾行为表型(情感依赖、认知替代、行为强化与社交代偿)纳入统一分析框架,为算法诱导成瘾的识别与诊断提供分类学整合范式。

相较于传统成瘾理论主要聚焦于物质依赖或单一行为模式的局限性,本模型通过揭示神经由算法诱导到行为表型的机制解释框架,有三方面理论贡献:(1)在机制解释层面,阐明了算法如何通过情绪韵律模仿、社交替代通道等中介路径,精准放大 ANA 所识别的功能缺陷,填补了从神经机制向行为表型转化的理论不足。(2)在分类学层面,以神经功能剖面作为行为表型分类的锚定点,有效克服现行诊断标准在解释数字成瘾异质性方面的不足(Almourad et al., 2020)。(3)在学科融合层面,推动成瘾研究从单一神经或行为范式向神经、算法和行为三元交互转型,为人工智能伦理与治理问题提供了具有扩展潜力的分析框架。

在实践层面,本模型有两方面启示:(1)政策设计:为相关立法提供神经科学依据,呼吁将算法诱导

成瘾纳入公共卫生与数字治理法律体系,从而从界面限制层面提升为对算法机制的深度干预。(2)技术伦理:基于本模型构建反成瘾算法,有助于抑制当前完全逐利的商业逻辑,引导人工智能从诱发成瘾的技术工具转变为支持认知发展的积极工具。

6.2 研究局限性

本文有以下四点局限性。首先,算法诱导成瘾的深层动因挖掘有限。本文虽提出了中介路径的表型转化功能,但未深入剖析算法诱导设计的底层机制。例如,强化学习中的奖励塑造如何通过 Q-learning 函数优化成瘾行为发生率,需结合算法逆向工程与计算神经科学进行因果推断。

其次,模型为解释性框架,当前仍缺乏经充分验证的算法诱导成瘾专用评估工具。目前,模型受因果推断的效度约束,受限于算法-用户动态反馈循环,即个体神经状态影响算法响应,算法响应又重塑神经功能,传统中介分析假设的单向因果链难以成立;同时,神经功能、算法暴露与行为表现三者在时间上高度重叠,时序分离困难,导致中介效应的方向性难以确定。未来需突破动态反馈系统的识别困境,采用双重机器学习、因果森林等方法,在动态系统中识别主导性时序关联,或尝试构建共同演化模型。

第三,数据生态的约束挑战。商业算法黑箱导致关键行为参数难以获取,而神经影像数据的稀疏性使神经功能域剖面难以大样本验证。未来可在保障用户隐私,加密用户神经画像前提下实现跨平台数据协同分析。

最后,个体文化变量的模型兼容性缺失。当前框架对集体主义文化中的社交代偿、高神经质个体的情感依赖易感性等差异缺乏适配性。建议引入更多的个性特征、文化背景等,提高其解释力和预测力,并探讨这些差异背后的机制和原因,构建神经功能域、个体特质与文化背景的三阶交互模型。

6.3 结论

AI 技术不应成为操纵用户、诱导成瘾的工具,而应成为助力人类提升自我价值、实现个人成长和社会进步的重要力量。本文提出算法诱导成瘾(Algorithm-Induced Addiction, AIA)概念,并初步构建

神经-行为双层耦合模型,在此基础上,呼吁推动人工智能治理与人工智能伦理的双重范式转型。在工具层面,将反成瘾算法纳入技术设计的责任机制,重建“人控制算法”而非“算法控制人”的协同关系。在价值层面,引导人工智能从成瘾驱动型转向认知增强型。此外,将 AI 技术应用于教育、医疗、文化等更多有益于人类发展的领域,可以充分发挥人工智能的潜力,推动 AI 走出成瘾与人智交互错位之思,为构建一个更加健康、和谐、可持续发展的数字社会奠定坚实的基础。

参考文献

- 范文芳,王千.(2022). 个性化智能推荐对消费者在线冲动购买意愿的影响研究. *管理评论*, 12(12), 146-156, 194.
- 何全兴,李兆岚,杨海波.(2025). 物质成瘾患者和行为成瘾患者在纹状体和前额叶激活的异同:基于双系统模型的元分析. *心理学报*, 57(8), 1333-1348.
- 侯娟,朱英格,方晓义.(2021). 手机成瘾与抑郁:社交焦虑和负性情绪信息注意偏向的多重中介作用. *心理学报*, 53(4), 362-373.
- 马建苓,刘畅.(2019). 错失恐惧对大学生社交网络成瘾的影响:社交网络整合性使用与社交网络支持的中介作用. *心理发展与教育*, 35(5), 605-614.
- 孙瑜晨.(2022). 数字平台成瘾性技术的滥用与反垄断监管. *东方法学*, (6), 58-71.
- 魏亚楠.(2024). 人工智能时代成瘾性算法的危害及其规制问题研究. *辽宁经济*, (8), 80-87.
- 谢潇,罗世杰.(2025). 论生成式人工智能的动态风险及适应性治理. *北京工业大学学报(社会科学版)*, 25(1), 112-125.
- 张炎亮,张超,李静.(2020). 基于动态用户画像标签的 KNN 分类推荐算法研究. *情报科学*, 38(8), 11-15.
- Almourad, M. B., McAlaney, J., Skinner, T., Pleva, M., & Ali, R. (2020). Defining digital addiction: Key features from the literature. *Psihologija*, 53(3), 237-253.
- Bilal, M., Zhang, Y. F., Cai, S. K., Akram, U., & Halibas, A. (2024). Artificial intelligence is the magic wand making customer-centric a reality! An investigation into the relationship between consumer purchase intention and consumer engagement through affective attachment. *Journal of Retailing and Consumer Services*, 77, 103674.

- Cemiloglu, D. , Almourad, M. B. , McAlaney, J. , & Ali, R. (2022). Combatting digital addiction: Current approaches and future directions. *Technology in Society*, 68, 101832.
- Chen, Q. , Jing, Y. , Gong, Y. , & Tan, J. (2025). Will users fall in love with ChatGPT? A perspective from the triangular theory of love. *Journal of Business Research*, 186, 114982.
- Costello, N. , Sutton, R. , Jones, M. , Almassian, M. , Raffoul, A. , Ojumu, O. , Salvia, M. , Santoso, M. , Kavanaugh, J. R. , & Austin, S. B. (2023). Algorithms, addiction, and adolescent mental health: An interdisciplinary study to inform state – level policy action to protect youth from the dangers of social media. *American Journal of Law & Medicine*, 49(2 – 3), 135 – 172.
- Freitas, J. D. , & Cohen, I. G. (2025). Unregulated emotional risks of AI wellness apps. *Nature Machine Intelligence*, 7, 813 – 815.
- Gao, Y. , Hu, Y. , Wang, J. , Liu, C. , Im, H. , Jin, W. , Zhu, W. , Ge, W. , Zhao, G. , Yao, Q. , Wang, P. , Zhang, M. , Niu, X. , He, Q. , & Wang, Q. (2025). Neuroanatomical and functional substrates of the short video addiction and its association with brain transcriptomic and cellular architecture. *NeuroImage*, 307, 121029.
- Griffiths, M. (1995). Technological addictions. *Clinical Psychology Forum*, 76, 14 – 19.
- Huang, Y. T. , & Huang, H. Y. (2025). Exploring the effect of attachment on technology addiction to generative AI chatbots: A structural equation modeling analysis. *International Journal of Human – Computer Interaction*, 41(15), 9440 – 9449.
- Jain, L. , Velez, L. , Karlapati, S. , Forand, M. , Kannali, R. , Yousaf, R. A. , Ahmed, R. , Sarfraz, Z. , Sutter, P. A. , & Tallo, C. A. (2025). Exploring problematic TikTok use and mental health issues: A systematic review of empirical studies. *Journal of Primary Care and Community Health*, 16, 215.
- Koob, G. F. , & Le Moal, M. (2008). Addiction and the brain antireward system. *Annual Review of Psychology*, 59, 29 – 53.
- Kooli, C. , Kooli, Y. , & Kooli, E. (2025). Generative artificial intelligence addiction syndrome: A new behavioral disorder? *Asian Journal of Psychiatry*, 107, 104476.
- Kosmyna, N. , Hauptmann, E. , Yuan, Y. T. , Situ, J. , Liao, X. H. , Beresnitsky, A. V. , Braunstein, I. , & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. arXiv: 2506.08872.
- Kwako, L. E. , Momenan, R. , Litten, R. Z. , Koob, G. F. , & Goldman, D. (2016). Addictions neuroclinical assessment: A neuroscience – based framework for addictive disorders. *Biological Psychiatry*, 80(3), 179 – 189.
- Marriott, H. R. , & Pitardi, V. (2024). One is the loneliest number... Two can be as bad as one. The influence of AI friendship apps on users' well – being and addiction. *Psychology & Marketing*, 41(1), 86 – 101.
- Méndez, M. L. , Padrón, I. , Fumero, A. , & Marrero, R. J. (2024). Effects of internet and smartphone addiction on cognitive control in adolescents and young adults: A systematic review of fMRI studies. *Neuroscience and Biobehavioral Reviews*, 159, 105572.
- Park, B. , Han, D. H. , & Roh, S. (2017). Neurobiological findings related to internet use disorders. *Psychiatry and Clinical Neurosciences*, 71(8), 467 – 478.
- Qin, Y. , Omar, B. , & Musetti, A. (2022). The addiction behavior of short – form video app TikTok: The information quality and system quality perspective. *Frontiers in Psychology*, 13, 932805.
- Skok, K. , & Waszkiewicz, N. (2024). Biomarkers of Internet Gaming Disorder – A Narrative Review. *Journal of Clinical Medicine*, 13(17), 5110.
- Su, C. , Zhou, H. , Gong, L. , Teng, B. , Geng, F. , & Hu, Y. (2021). Viewing personalized video clips recommended by TikTok activates default mode network and ventral tegmental area. *NeuroImage*, 237, 118136.
- Turel, O. , Serenko, A. , & Giles, P. (2011). Integrating technology addiction and use: An empirical investigation of online auction users. *MIS Quarterly*, 35(4), 1043 – 1061.
- Volkow, N. D. , Koob, G. F. , & McLellan, A. T. (2016). Neurobiologic advances from the brain disease model of addiction. *New England Journal of Medicine*, 374(4), 363 – 371.
- Volkow, N. D. , & Morales, M. (2015). The brain on drugs: From reward to addiction. *Cell*, 162(4), 712 – 725.
- Xiao, Y. Y. , Meng, Y. , Brown, T. T. , Keyes, K. M. , & Mann, J. J. (2025). Addictive screen use trajectories and suicidal behaviors, suicidal ideation, and mental health in US youths. *JAMA – Journal of the American Medical Association*, 334(3), 219 – 228.
- Yuan, K. , Qin, W. , Liu, Y. , & Tian, J. (2011). Internet addiction.

- tion; Neuroimaging findings. *Communicative & Integrative Biology*, 4(6), 637 – 639.
- Yu, S. – C. , Chen, H. – R. , & Yang, Y. – W. (2024). Development and validation the Problematic ChatGPT Use Scale; A preliminary report. *Current Psychology*, 43 (7), 26080 – 26092.
- Zhou, T. , & Zhang, C. (2024). Examining generative AI user addiction from a C – A – C perspective. *Technology in Society*, 78, 102653.
- Zhou, Y. Y. , Liu, Q. H. , Huang, J. H. , & Li, G. Q. (2026). Creative scar without generative AI; Individual creativity fails to sustain while homogeneity keeps climbing. *Technology in Society*, 84, 103087.

Algorithm – Induced Addiction: Current Status and Future Directions

Liu Tao^{1,2,3}, Chen Yanzuo¹, He Lin⁴, Yang Yifan¹, Wang Qing³, Yu Xiaoyu^{1,2,3}

(1. School of Management, Shanghai University, Shanghai 200444;

2. Shanghai Enterprise Innovation and High Quality Development Research Center, Shanghai 200444;

3. MBA Center, Shanghai University, Shanghai 200444;

4. School of Management, Hangzhou Dianzi University, Hangzhou 310018)

Abstract: As artificial intelligence becomes deeply embedded in social practices, Algorithm – Induced Addiction (AIA) has emerged as a novel form of behavioral dependence. We examine the current status, impacts, technological underpinnings, and biological mechanisms of AIA, then propose a neural – behavioral dual – layer coupling model that integrates a three – dimensional functional domain for clinical addiction assessment with four categories of addictive behavioral phenotypes, thereby establishing a unified analytical framework. Future research directions are outlined across three domains: policy design, interdisciplinary investigation, and the development of anti – addiction algorithms. We aim to provide theoretical support for establishing a pro – social governance paradigm for algorithms, facilitating the transition of artificial intelligence from addictive intelligence toward cognitively enhancing intelligence.

Key words: algorithm – induced addiction; neuro – behavioral dual – layer coupling model; anti – addiction design; digital ethics