

生成式人工智能的技术风险与人的主体性维护

白一岑

西安交通大学马克思主义学院, 陕西 西安

收稿日期: 2026年6月3日; 录用日期: 2026年6月26日; 发布日期: 2026年7月8日

摘要

生成式人工智能的快速发展, 在显著提升生产效率的同时, 也引发了技术依赖与主体功能弱化等深层风险。在认知、实践与生存层面, 人的认知自主性、实践创造性与价值选择性面临潜在挑战。数字时代的主体功能弱化呈现出内在化、隐蔽化与微观化特征, 个体在无意识中倾向于将思考、判断与创造活动让渡于技术系统, 从而可能与自身的能动本质发生分离。应对这一风险, 需在技术发展与社会进步之间寻求动态平衡, 构建包括算法审计、数据伦理审查、人工智能素养教育及“人在回路”式人机协同机制在内的系统性策略, 从而在技术赋能与主体维护之间形成良性张力, 最终回归以“人”为中心的价值认同。

关键词

生成式人工智能, 人的主体性, 技术风险, 应对策略, 人机协同

Technological Risks of Generative Artificial Intelligence and the Preservation of Human Subjectivity

Yicen Bai

School of Marxism, Xi'an Jiaotong University, Xi'an Shaanxi

Received: June 3, 2026; accepted: June 26, 2026; published: July 8, 2026

Abstract

The rapid development of generative artificial intelligence, while significantly enhancing production efficiency, has also triggered deep-seated risks such as technological dependence and the weakening of human functions. At the cognitive, practical, and existential levels, human cognitive autonomy, practical creativity, and value-oriented choice face potential challenges. The weakening of human functions in the digital age exhibits characteristics of internalization, concealment, and micro-

level permeation, as individuals unconsciously tend to cede thinking, judgment, and creative activities to technological systems, potentially leading to a separation from their own capacity for agency. To address this risk, it is necessary to seek a dynamic balance between technological development and social progress, and to construct systematic strategies including algorithmic auditing, data ethics review, AI literacy education, and “human-in-the-loop” human-machine collaboration mechanisms. In doing so, a productive tension between technological empowerment and the preservation of human subjectivity can be formed, ultimately returning to a value identity centered on the “human”.

Keywords

Generative Artificial Intelligence, Human Subjectivity, Technological Risks, Coping Strategies, Human-Machine Collaboration

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来，网络上盛行的文化又出现了新变化，从风靡一时“摆烂文化”、“躺平哲学”进一步发展到“发疯文学”，人类似乎正步入一个被戏称为“发癫”的时代。当前数字社会语境下，青年群体中出现虚拟空间情绪表达、现实参与主动性弱化等社会心态现象，其背后是数字技术深度嵌入生活带来的行为模式转变。随着生成式人工智能技术的广泛应用，AI已被广泛应用于人们的工作、学习、创作等场景，人们逐步将脑力劳动、决策判断、内容生成等活动交由AI完成，形成技术依赖。这一过程中，人类主动让渡自主思考与实践创造空间，人工智能则在数据训练与算法迭代中持续强化内容生成与决策辅助能力，人机关系呈现人类能动性弱化、技术工具功能泛化的特征。

在人机共生的时代背景下，厘清技术风险、坚守人的主体地位、构建健康人机关系，成为哲学、社会学与科技伦理领域的重要研究议题。本文基于生成式人工智能发展现实，分析其对人的主体性的多重影响，探索风险防范与主体维护的可行路径，为技术向善与社会良性发展提供理论参考。

2. 核心概念与理论框架

在深入探讨生成式人工智能的影响之前，有必要对本文的两个核心概念“人的主体性”与“技术依赖下的主体功能弱化”进行清晰的界定，以明确分析的框架与边界。

(一) 人的主体性及其核心维度

马克思指出主体性“是对象性的本质力量的主体性”([1], p. 209)，人的主体性，在本研究的语境下，指的是人在与外部世界(包括技术系统)的互动中，所展现出的能动、自主和有尊严的地位与能力。它并非一个单一维度的抽象概念，而是可以具体分解为三个相互关联的核心维度。第一，认知自主性，即个体独立思考、批判性反思、自主判断和决策的能力，而非被动接受外部信息或指令。第二，实践创造性，即个体通过亲身实践、劳动创造来改变世界、实现自身目的的能动性，包括提出新问题、探索新方法、创造新事物的能力。第三，价值选择性，即个体基于自身信念、伦理准则和价值标准，在多元可能中进行自主选择和价值判断的能力，这构成了人的尊严和道德责任的基础。本文对生成式人工智能影响的评估，将主要围绕这三个维度展开。

(二) 技术依赖与主体功能弱化的时代特征

关于技术引发的“主体功能弱化”现象，学界常借用“异化”这一具有深厚学术传统的概念加以分析。最初，异化概念用于批判工业时代劳动者与其产品、劳动过程及自身本质的分离[1] (pp. 182-197)。而在数字时代，技术驱动的主体功能弱化呈现出新的形式与根源。工业时代的主体功能丧失主要表现为物质生产领域的外在强制与剥削，劳动者在物理层面上被机器所支配。与之相比，数字时代的主体功能弱化则更趋内在化、隐蔽化和微观化。它并非源于对身体的直接强制，而是通过吸引、诱导、习惯化和环境塑造等方式发生。其深层机制在于，当技术系统(如生成式人工智能)的设计与运行逻辑变得极度复杂，其输出结果又极其便捷、高效时，个体可能在无意识、非强迫的状态下，出于对效率与便利的追求，主动将原本属于自己的思考、判断与创造活动让渡给技术。这种“主动的让渡”导致了更深层次的能力退化与依赖，个体在享受便利的同时，也可能在认知、实践及价值选择层面与自身的能动本质发生分离。理解这种潜移默化的影响机制，是科学评估生成式人工智能风险并制定有效应对策略的关键前提。

3. 生成式人工智能与人的主体性关系分析

(一) 生成式人工智能的技术机理与发展现状

生成式人工智能是人类发挥主观能动性的产物，作为重要的辅助性工具，它已经从诞生、进化发展到一门日趋通用的技术，成为推动人类进步的一股重要力量。自 Open AI 公司推出聊天生成预训练转换技术(Chat Generative Pre-trained Transformer, 简称 Chat GPT)以来，以智谱清言、文心一言、通义千问、Sora 为代表的生成式人工智能技术被广泛应用于媒体、教育、娱乐等领域，成为当代人类工作和学习中重要的辅助性技术工具。生成式人工智能(Artificial Intelligence Generated Content, 简称 AIGC)是人工智能领域的重要分支，是一种基于算法和模型生成文本、图片、声音、视频、代码等内容的技术，不同于传统 AI 的分析功能，生成式人工智能可以学习并模拟事物的内在规律，根据用户的输入资料生成具有逻辑性和连贯性的新内容，运行原理见图 1。无论是撰写文章、设计图形还是制作视频，生成式人工智能都展现出了令人惊叹的能力。

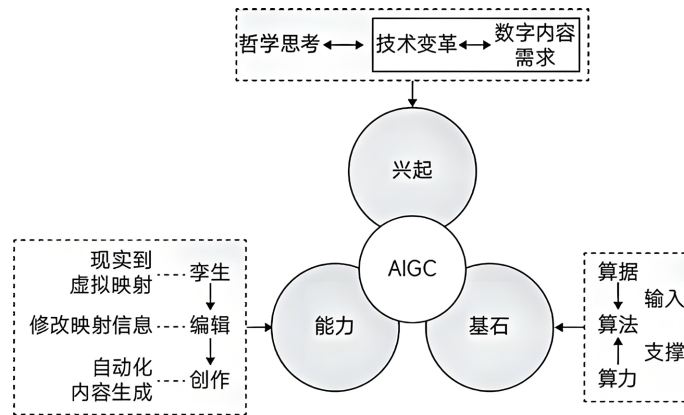


Figure 1. Tracing the origin of AIGC: emerging reasons, building blocks, driving capabilities [2]
 图 1. AIGC 溯源：兴起缘由、构建基石、驱动能力[2]

该技术依托计算机软硬件升级与智能算法突破，通过程序模拟人脑神经元结构、连接方式与信号传递机制，构建大规模数字模型，以人类生产生活过程中产生的海量数据为样本来训练数字模型，使其习得人类复杂的知识模式，展现出表达和决策的能力[3]。开放人工智能(open AI)公司发布的 GPT-3 模型总共包含了 1750 亿的参数，见表 1，而人脑神经元数量约为 1000 亿个，GPT-3 模型的参数与人脑神经元的数量已经达到了同一量级。

Table 1. ChatGPT series generative artificial intelligence models [4]**表 1.** ChatGPT 系列生成式人工智能模型[4]

模型名称	训练参数	学习材料	发布时间
GPT-1	1.17 亿	5 GB	2018
GPT-2	15 亿	40 GB	2019
GPT-3	1750 亿	45 TB	2020
ChatGPT (GPT-3.5)	同上	同上	2022 年 11 月
GPT-4	1.8 万亿	未公布	2023 年 3 月

大模型的构建和人工智能算法的创新，是技术飞速发展背景下计算机科学、认知科学、神经科学等理论应用的结晶，是当前时代最前沿的科技成果。部分先进的生成式人工智能模型已接近甚至在某些维度上超越了人脑的某些功能，但这并不等同于其具备了与人类相同的主体性。

生成式人工智能的兴起与发展，是人类在数据处理与知识挖掘领域的一次重大飞跃。通过深度学习、自然语言处理、生成模型等先进算法，生成式人工智能能够高效地分析、理解、甚至创造信息，这不仅极大地便利了人类的学习过程，也深刻改变了工作模式，提高了生产效率与创造力。生成式 AI 在当今社会多个领域协助人们完成重复性高、繁琐复杂的任务，释放人类劳动力去从事更高层次的创造性工作，实现人职匹配的最优化，并展现出在科学研究、艺术创作、医疗健康等多个领域的巨大潜力，不断拓展着人类认知与实践的边界。

(二) 人的主体性地位的理论澄清

在人与生成式人工智能的关系定位上，需要从多重视角加以澄清。首先，从历史与哲学的视角审视，人工智能并非自然演化的结果，亦非独立于人类智慧之外的存在，而是人类长期探索与创新的智慧结晶。人类作为创造主体，将原本抽象的概念逐步转化为具体的技术形态，从而推动了人工智能领域的新兴发展。生成式人工智能所生成的内容同样具有“属人性”。其次，在人机互动与共生的关系框架内，人类不仅是创造者，更是使用主体。人工智能技术的每一次迭代与升级，都是基于人类在使用过程中遇到的新挑战、新需求而展开的。人类通过不断实践发现现有技术的局限性，进而提出新的需求，驱动着人工智能技术的持续进步。这种由人类需求驱动的发展模式，体现了人类作为使用者的主观能动性。最后，从目的论的角度来看，人工智能技术的终极目标应是服务于人类，提升人类的生活质量与社会福祉。生成式人工智能的设计初衷与应用场景均围绕着人类的需求与利益展开。因此，从本质上说，人是居于主体性地位的，生成式人工智能的发展与应用始终应以为人服务为根本导向。

(三) 主体功能让渡与技术功能越位的现实表现

作为辅助性工具，人工智能的出现越来越成为一种通用技术，其异化趋势越来越明显。在部分人类主动将自己的主体性地位让位出来之后，生成式人工智能便展现出巨大的异己力量，人在“物的依赖性”中逐渐丧失了自己，被技术“物”所操控([5], p. 52)。

然而，随着以人工智能为代表的数字技术不断迭代进化，实践中出现了人的认知自主性、实践创造性和价值选择性被削弱的风险。这种风险具体表现为两种相互关联的倾向。一是人类能动性的不自觉让渡，即面对复杂或枯燥任务时，越来越多的人尤其是年轻人倾向于选择让人工智能去处理，在高度肯定技术能力的前提下形成了一种不加批判的依赖习惯。长此以往，这种习惯可能会在潜移默化中弱化人类独立解决复杂问题的意愿与能力。二是技术功能的潜在越位，即作为辅助性工具的人工智能在广泛应用中衍生出多元风险。在认知层面，人们若过度追求工具效率而将批判性思维任务交由机器，可能导致认知自主性的降低。在实践层面，算法歧视、隐私泄露、信息过载等问题日益凸显，不仅干扰了正常的数

字实践活动，也可能使得个体的价值选择性受到无形框架的束缚，例如被平台算法塑造的观点和偏好。

4. 技术依赖对人类主体性地位的多维挑战

生成式人工智能技术的飞速发展，引发了深刻的技术与社会变革。当人们对其依赖度逐渐提高，一系列由技术依赖引发的潜在风险愈发显著，对人类主体性地位构成严峻挑战。这些挑战不仅体现在认知领域、实践领域，还深刻影响着人类的生存领域。

(一) 认知领域的挑战：自主思考能力的潜在退化

在 ChatGPT 问世之初，人们对于人工智能所展现出的潜能普遍持有一种审慎的怀疑态度。经过初步的实践，人工智能以其卓越的性能逐渐赢得了人们的信赖，进而在工作场景中被有选择性地采纳与应用。这一趋势随着时间的推移而不断深化，人们对于生成式人工智能的信任度持续攀升，加之长期的使用习惯逐渐养成，使得人们在面对复杂繁琐的任务时，自然而然地倾向于首先考虑并诉诸于 AI 寻求解决方案。时至今日，已有相当一部分人群将自身的工作职责与任务完全托付给了 AI，对其输出结果保持着无条件的信任。在这一演变过程中，AI 在获得广泛信任的同时，也呈现出一种强大的影响力，它逐渐从一种辅助工具转变为一种可能深刻影响甚至塑造着人们思维习惯的外部力量。

此外，人工智能技术的快速发展也加剧了信息的过载和碎片化。尤瓦尔·赫拉利在《未来简史》中深刻指出：“数据的流动量已经大到非人所能处理，人类无法再将数据转化为信息，更不用说转化成知识或智能。于是，处理数据的工作应该交给能力远超人类大脑的电子算法。”^[6]人们每天面临着海量的信息输入，但往往缺乏足够的时间和精力去筛选、整合这些信息。在海量信息面前，人们筛选、整合、批判信息的精力被大量消耗，维持独立、清晰的思维和判断变得更加困难。

(二) 实践领域的挑战：机械麻木感知的固化趋势

在实践领域，对 AI 的依赖导致人类直接感知机会的大幅缩减。随着生成式人工智能的发展，数据的监控和收集直接绕过了我们的身体感知，形成了巨大的传播和交换网络，并在此基础上构成了不依赖于人类身体的机器感知^[7]。生成式人工智能技术的飞速进步，数据的采集与监控过程日益超越了传统身体感知的界限，构建起一个庞大而复杂的传播与交换网络。这一网络不仅极大地扩展了信息流通的范围与速度，更重要的是，它孕育了一种全新的感知模式——机器感知，这种感知模式不再局限于人类身体的直接体验，而是依托于算法与数据的深度分析。人类日益习惯于与智能机器进行交互，这种交互模式的便捷性与高效性，使得在面对需要实践验证的任务时，往往因惰性而倾向于信赖拥有海量数据处理能力的人工智能系统，从而主动或被动地放弃了通过亲身实践去探索未知、验证真理的机会。这一行为模式的普遍化，不仅削弱了人类对于自然环境和人类社会变迁的直接感知与深刻体验，更在无形中削弱了人们作为实践主体的创造力与灵活性。

长此以往，这种机械麻木感知的固化趋势，无疑将对人类的实践活动产生深远的负面影响。一方面，它可能导致问题解决路径的单一化，人们习惯于寻找机器提供的“最优解”或“标准答案”，从而限制了探索多样路径和试错创新的勇气与能力。另一方面，对机器方案的过度依赖，也可能降低社会整体的适应性与韧性，使人们面对超出 AI 训练数据范畴的新挑战时，显得手足无措。

(三) 生存领域的挑战：劳动生产关系与自我价值认知的变革

研究指出，美国所有工人任务中约有 15% 的工作可以通过 LLMs (大型语言模型) 来完成，而且在 LLMs 的帮助下，工人可以保质保量地加快完成任务。如果将未来建立在 LLMs 之上的软件和工具纳入其中，被人工智能影响的工作任务，比例将增加到 47% 到 56% 之间^[8]。随着人工智能技术的迅猛发展与广泛应用，我们正见证着传统劳动生产关系经历着一场前所未有的根本性变革。这场变革的核心在于，机器与智能系统在生产流程中的角色日益凸显，它们以高效、精准、不知疲倦的特性，逐步取代了人类在生产

活动中的直接参与，促使人类劳动力从基础、重复的劳作中解放出来，转向更为复杂、高端、创造性的劳动形态。

然而，这一变革也引发了一系列新的问题与挑战。从就业市场的角度来看，机器与智能系统的广泛应用无疑会导致大量传统劳动岗位的消失，特别是那些依赖于简单重复劳动的行业。这种就业结构的调整，要求劳动者必须不断适应新的就业环境，积极学习新技能，以应对市场需求的快速变化。这一过程不仅充满了不确定性，还可能加剧社会的就业压力与不平等现象。另一方面，生产方式的变革也促使人类与机器之间的关系发生了微妙而深刻的变化。在过去，人类是生产过程中的绝对主体，机器只是辅助工具。但在人工智能时代，机器与智能系统的智能水平不断提升，它们在某些领域甚至超越了人类的能力，这使得人类不得不重新审视自己在生产体系中的位置与角色。

如何在与机器的共生中保持人类的独特价值，如何避免因技术替代而产生自我价值的困惑与迷失，成为亟待解决的时代课题。面对人工智能带来的劳动生产关系变革，人们不仅需要关注技术进步带来的生产效率提升，更要深入思考如何平衡技术发展与社会稳定、人类尊严之间的关系。

5. 维护人的主体性的系统性策略

为了有效应对生成式人工智能对人类主体性构成的挑战，需要构建一个多层次的系统性应对策略。这一策略的核心目标是回归“科技增进人类福祉”的根本宗旨，在拥抱技术进步的同时，始终将人的发展、尊严与主体地位作为衡量技术进步的根本尺度。

(一) 价值理念层面：正视技术工具属性，重申人的核心地位。

在探讨人工智能与人类文明的关系时，应当认识到技术是一把双刃剑。尽管人工智能的兴起带来了诸多挑战与争议，但其无疑也在人类文明各个方面的进步和发展中扮演了至关重要的角色，起到了历史性的推动作用。从科学研究到工业生产，从医疗健康到社会服务，人工智能的应用极大地提升了效率，拓宽了人类能力的边界，为人类社会的可持续发展注入了新的活力。然而，在享受人工智能带来的便利的同时，必须坚定人的主体性地位。人工智能尽管强大，但终究是人类智慧的结晶，是人类利用自身知识和技术创造出来的工具。因此，我们不应因技术的先进而贬低自己，更不应过分崇拜、依赖人工智能，以至于丧失了对自身能力的信心。

科技与人文二者并非对立，而是相辅相成的。人与数字技术需要保有既相互是目的又相互是手段的关系，展开“协同性”的对话，实现主体间的交互性合作，构建一种交互主体性的普遍网络，才能实现有机共生与融合共进的有机组织新生态([5], p. 49)。科技的发展离不开人类的智慧与创造力，而人类社会的进步也需要科技的支撑与推动。因此，我们应该正视人工智能的出现，将其视为人类智慧的一部分，在主体间性的交互对话中实现有机共生与融合共进。

(二) 个体行为层面：发挥主观能动性，培养批判性人机协作习惯。

在探讨人工智能大语言模型对人类生产创造活动的影响时，每个人应当发挥主观能动性，有选择地利用先进技术作为辅助工具。

人工智能大语言模型基于大量数据快速生成文本、分析语言模式，为人类提供前所未有的便利，但这种便利不应成为放弃思考和创造的借口。相反，应明确的一点是，人工智能只是工具，真正的创造力和智慧仍然源于人类自身。个人在面对人工智能大语言模型时，应发挥主观能动性，有选择地利用这些技术来辅助自己的生产创造活动。具体而言，首先，需要保持认知警觉，对 AI 的输出结果应保持验证与质疑的态度，不盲从、不轻信，将 AI 视为“副驾驶”而非“自动驾驶仪”；其次，要有意识地进行思维训练，主动进行独立思考、深度阅读、与人辩论等“低科技”活动，以对抗可能的思维惰性；此外，在使用 AI 完成何种任务、在何种程度上依赖 AI 结果等问题上，进行主动选择而非被动接受。同时，应该警

惕盲目追求智能化和自动化的倾向,保持对技术的警觉和批判性思维,不断审视和反思自己的使用行为,以确保技术真正服务于人类的福祉和进步。

(三) 制度与社会层面: 构建风险防范与能力培养体系

第一,建立精细化的监管与伦理审查机制。在监管方面,应当建立针对生成式人工智能的专门监管框架,明确其应用边界和伦理准则。尤其需要推行强制性的算法审计制度,对于用于对个人机会有重大影响的 AI 应用,应由第三方独立机构定期进行算法审计,检查其是否存在歧视性、不公平性或系统性偏见,并将审计结果向社会公开。同时,在 AI 产品开发生命周期中嵌入数据伦理审查机制,评估其在数据收集、隐私保护、内容生成等方面的合规性与伦理风险,明确算法透明度与可解释性要求,法律规定 AI 系统在做出可能影响个人重要权益的决定时,需提供用户可理解的、非技术黑箱式的解释。

第二,推动技术创新与伦理价值的融合。鼓励研发聚焦于“以人为本”的 AI 技术,例如支持可解释性 AI 的发展,使 AI 的决策过程更加透明。发展用于辅助而非替代的技术,同时,加强 AI 伦理与安全技术的研究,如用于检测 AI 生成内容、防范偏见和滥用的技术。

第三,强化面向全民的人工智能素养与人文教育。在数字时代,提高公众的数字素养和批判性思维能力至关重要。提高公众对 AI 的基本原理与固有局限的理解,如大数据、机器学习、生成式 AI 的工作原理及其“幻觉”、“偏见”等问题;探讨 AI 应用中的伦理与社会影响,包括算法歧视、隐私保护、知识产权、深度伪造等议题;培养高阶思维技能,增强人们在 AI 辅助下的问题定义、信息甄别、批判性验证和创造性重构能力;进行元认知能力训练,引导人们反思自身在使用 AI 时的认知习惯与决策过程,培养在人机互动中保持自主意识的元认知能力。此外,还应关注技术失业和就业转型问题,政府与企业合作提供职业培训和再就业支持,帮助受影响的群体适应新的就业环境。

第四,探索人机协同的新型合作模式。在各个领域主动设计并推广“人在回路”的协作模式,确保关键决策环节由人类做出最终判断与承担责任。同时,鼓励创新应用,利用人工智能辅助,激发人类创造力,提升人的主体性。建立反馈机制,让人类能够参与到人工智能系统的优化和改进中,确保技术始终服务于人类的需求和利益。

综合来看,生成式人工智能技术一方面作为一种新的生产要素,推动生产力呈幂次方增长,极大地解放了人,为人的生产生活带来了极大的便利;另一方面,失控的技术依赖引发的潜在风险在人们的认知领域、实践领域以及生存领域对人的主体性带来了挑战。生成式人工智能作为人类的产物、辅助人类的工具,要坚持强调人的主体性地位,正确对待人机共生关系,重塑良性的人机交互关系以及从“物”的价值重新回归到“人”的认同上[9],对物的认同和自我的定位要以人为中心,在享受技术红利的同时,守护并砥砺人类独有的能动性与创造力,从而真正实现“科技更好增进人类福祉”。

参考文献

- [1] 马克思,恩格斯. 马克思恩格斯文集 第1卷[M]. 北京: 人民出版社, 2009: 182-197+209.
- [2] 马阅欢, 方佳明, 杨慧颖, 等. 机器的觉醒: 生成式 AI 溯源、演进与展望[J]. 南开管理评论, 2024, 27(4): 66-77.
- [3] 马昕. 生成式人工智能技术对网络意识形态话语权的影响、挑战及其应对[J]. 陕西师范大学学报(哲学社会科学版), 2024年, 53(4): 170-176.
- [4] 王薇, 刘大椿. ChatGPT 的进化与人的主体性[J]. 科学技术哲学研究, 2024, 41(5): 22-29.
- [5] 阴雅婷. 从主体性、主客间性到主体间性: 人与数字技术的交互伦理关系[J]. 新闻爱好者, 2024(10): 49-52.
- [6] [以]尤瓦尔·赫拉利. 未来简史: 从智人到神人[M]. 林俊宏, 译. 北京: 中信出版社, 2017: 335.
- [7] 蓝江. 从身体感知到机器感知——数字化时代下感知形式的嬗变[J]. 西北师大学报(社会科学版), 2023, 60(3): 13-20.

-
- [8] Eloundou, T., Manning, S., Mishkin, P., *et al.* (2023) GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models.
- [9] 王晓丽, 李伟鑫. 生成式人工智能对人的主体性的挑战及应对[J]. 江苏大学学报(社会科学版), 2025, 27(4): 91-101+124.