

生成式人工智能的交互主体性探析

陈依婷

北京交通大学马克思主义学院, 北京

收稿日期: 2026年3月4日; 录用日期: 2026年3月23日; 发布日期: 2026年4月2日

摘要

生成式人工智能的崛起打破了传统人机单向交互模式, 催生了全新的交互主体性问题, 成为数字时代技术与人文交叉研究的核心议题。本文以哲学、社会学、马克思主义理论为多学科研究视角, 先对自主性、主体性、交互主体性及生成式人工智能等核心概念进行界定辨析, 明确研究的理论边界与基础; 再从人机、社会、自然三个维度剖析生成式人工智能的交互机制, 系统探讨其交互主体性的表现特征、核心本质, 同时深入分析其在价值导向、信息安全、技术依赖等方面的局限性与潜在风险; 进而从哲学层面对人机关系、生成式人工智能的主体性地位展开深度反思, 并结合其技术本质展望未来发展趋势与哲学内涵。本文的研究厘清了生成式人工智能交互主体性的核心内涵与发展边界, 明晰了数字时代人机关系的本质逻辑, 为实现生成式人工智能与人、社会、自然的和谐共生, 推动技术向正向价值发展提供了理论参考与实践思路。

关键词

生成式人工智能, 交互主体性, 技术异化, 人机关系

An Analysis of the Interactive Subjectivity of Generative Artificial Intelligence

Yiting Chen

School of Marxism, Beijing Jiaotong University, Beijing

Received: March 4, 2026; accepted: March 23, 2026; published: April 2, 2026

Abstract

The rise of generative artificial intelligence has broken the traditional one-way human-computer interaction model, giving rise to entirely new issues of interaction subjectivity, and has become a core topic at the intersection of technology and humanities in the digital age. From the multidisciplinary research perspectives of philosophy, sociology, and Marxist theory, this paper first defines and

analyzes the core concepts such as autonomy, subjectivity, interactive subjectivity, and generative artificial intelligence, clarifying the theoretical boundaries and foundations of the study; it then analyzes the interaction mechanisms of generative artificial intelligence from the three dimensions of human-computer, society, and nature, systematically exploring the characteristics and essential nature of its interactive subjectivity, while deeply analyzing its limitations and potential risks in value orientation, information security, and technological dependency; furthermore, it engages in a philosophical reflection on the human-computer relationship and the subjectivity status of generative artificial intelligence, and, in combination with its technological essence, looks forward to future development trends and philosophical implications. This study clarifies the core connotations and developmental boundaries of the interactive subjectivity of generative artificial intelligence, elucidates the essential logic of human-computer relationships in the digital age, and provides theoretical references and practical approaches for achieving harmonious coexistence between generative artificial intelligence and humans, society, and nature, as well as promoting the development of technology toward positive values.

Keywords

Generative Artificial Intelligence, Interactive Subjectivity, Technological Alienation, Human-Computer Relationship

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

人工智能技术的迭代升级正在深刻重塑人类的生产方式、生活方式与思维方式，从早期的运算型人工智能到感知型人工智能，再到如今的生成式人工智能，技术对“人机关系”的定义不断被改写。生成式人工智能以深度学习、大语言模型、生成对抗网络等为技术核心，能够自主生成文本、图像、音频、视频等多元内容，实现了从人机应答到人机共创的交互转型，这种转型使得“主体性”这一原本属于人的哲学范畴，开始向人工智能领域延伸，进而引发了生成式人工智能的交互主体性问题。

2. 相关概念辨析

概念是研究的逻辑起点，生成式人工智能的交互主体性研究涉及哲学、人工智能科学等多领域的核心概念，自主性、主体性、交互主体性其内涵与外延存在交叉与差异，人工智能与生成式人工智能的技术范畴也存在差异，若不加以明确辨析，极易导致研究边界模糊与逻辑混乱。

2.1. 自主性、主体性、交互主体性

2.1.1. 自主性

自主性是主体性的基础前提，指主体在认识世界和改造世界的过程中，摆脱外部强制与依附，凭借自身的意志、理性与能力进行自主判断、自主选择与自主行动的属性。在哲学层面，自主性是人的本质特征之一，其核心是“自我决定”，即人作为实践主体，能够根据自身的需求与目的，能动地作用于客观世界，而非被动地接受外部环境的支配。自主性的存在以主体具备独立的意识、理性与实践能力为前提，其本质是客观规律性与主观能动性的统一[1]。需要明确的是，哲学意义上的自主性具有鲜明的人类性，其与人类的意识、情感、价值判断等主观精神世界密不可分，是人类在长期的社会实践中逐步形成的本

质属性。而在人工智能领域，“自主性”是指机器在无需人类实时干预的情况下，根据预设的算法、模型与数据，自主完成感知、决策、行动与内容生成的能力，如生成式人工智能自主生成文本、图像的行为。但这种“自主性”与人类的自主性存在本质区别，人工智能的自主性是“算法化的自主性”，其所有的决策与生成行为均基于人类预设的逻辑与数据，缺乏独立的意识、意志与价值判断，是一种“伪自主性”，这也是区分人类自主性与人工智能自主性的核心边界。

2.1.2. 主体性

主体性是哲学的核心范畴，指人作为实践与认识的主体，在与客体的相互作用中所表现出的能动性、创造性与主导性。从马克思主义哲学视角来看，主体性并非先天存在的属性，而是人在社会实践中逐步形成并发展的。“人的本质不是单个人所固有的抽象物，在其现实性上，它是一切社会关系的总和。”^[2]主体性的本质是人的社会实践性。主体性的存在具有两个基本前提：一是主体具有独立的意识与自我意识，能够区分“自我”与“他者”；二是主体能够通过社会实践与客体发生相互作用，实现对客观世界的改造与自我的发展。这两个前提决定了传统哲学中“主体”的范畴仅限于人，因为只有人才具备真正的意识、自我意识与社会实践能力，而包括机器在内的一切客观事物，均只能作为被人类认识与改造的客体而存在。

2.1.3. 交互主体性

交互主体性，又称为主体间性，由胡塞尔首次提出，后经海德格尔、哈贝马斯等哲学家发展，成为现代哲学的重要概念，其核心是打破传统主客二元对立的思维模式，强调主体与主体之间通过沟通、理解、互动、交往形成的相互建构、相互依存的关系。

在胡塞尔的现象学中，交互主体性是解决“他人存在”问题的关键，即通过“移情”作用，实现自我与他人的相互理解，确认他人的主体性，从而构建起主体间的世界^[3]。海德格尔从存在论视角出发，将交互主体性理解为“此在”的共在，认为人作为此在，其存在本身就是与他人的共在，交互主体性是人的存在方式。哈贝马斯则将交互主体性与交往行为理论结合，认为主体间通过符合有效性要求的言语交往，能够达成相互理解与共识，实现社会的整合与发展。

交互主体性具有三个核心特征：一是主体的平等性，交互主体之间并非主客关系，而是具有平等的主体地位；二是互动的双向性，主体间的作用是相互的，而非单向的支配与被支配，每个主体的主体性都在与其他主体的互动中得到体现与建构；三是关系的建构性，交互主体性并非先天存在，而是主体在交往实践中逐步建构的，主体间的互动不仅塑造了彼此的认知与行为，也构建了新的社会关系与意义世界^[4]。需要明确的是，传统哲学中的交互主体性始终是“人的交互主体性”，其主体是具有独立意识与社会实践能力的人，而生成式人工智能的出现，使得交互主体性的主体范畴面临拓展，人机之间的交互是否能构成真正的交互主体性，成为亟待探讨的新问题。

2.2. 人工智能、生成式人工智能

2.2.1. 人工智能

人工智能是一门以让机器模拟、延伸和扩展人类智能为目标的技术科学，核心在于使机器具备感知、认知、决策、学习、推理等类人智能能力，进而替代或辅助人类完成各类任务。1956年的达特茅斯会议，让“人工智能”这一概念被正式确立。自此，该技术沿着符号主义、连接主义、行为主义等方向迭代演进，从最初的逻辑推理、专家系统，到中后期的机器学习、深度学习，再到当下的生成式人工智能，其智能程度持续精进，应用边界也不断拓宽。按照智能水平与应用特征划分，人工智能可分为弱人工智能、强人工智能与超人工智能三个阶段，目前人类仍处于弱人工智能阶段，这一阶段的机器仅能在特定领域

模拟人类的某项智能能力，像语音识别、图像分类、内容生成等均属此类，其核心是“模拟智能”，并不具备独立的意识与自我意识。强人工智能则意味着机器拥有与人类相当的通用智能，能够在任意领域完成人类可实现的各类任务，且具备独立的意识、思维与价值判断。超人工智能则是指机器的智能水平远超人类，能够实现自我进化与自我发展，其能力范畴已超出人类的认知边界。

从本质来看，人工智能是人类智能的延伸与模拟，其所有智能能力均源于人类的算法设计、数据训练与技术研发，是人类社会实践的产物。而从马克思主义唯物论的视角分析，人工智能作为人类创造的工具，归属于生产力的范畴，其发展能够推动生产力的进步，进而带动生产关系的调整与人类社会的整体发展。

2.2.2. 生成式人工智能

生成式人工智能是人工智能技术发展的新阶段，属于弱人工智能的范畴，能基于海量的训练数据，自主生成符合人类认知与使用需求的文本、图像、音频、视频、代码等多元内容的人工智能技术体系。与传统的判别式人工智能不同，判别式人工智能的核心是“识别与判断”，即根据输入的数据，判断其所属的类别或做出相应的应答。而生成式人工智能的核心是“创造与生成”，即能够根据人类的简单指令，自主生成全新的、具有一定逻辑性与创造性的内容，其交互模式是“人类引导-机器生成-人类反馈-机器优化”的双向循环模式，这种双向交互模式是生成式人工智能区别于传统人工智能的核心特征。

生成式人工智能的出现，打破了人类在内容创作领域的主导地位，实现了人机协同创作的新形态，其应用已渗透到文案创作、艺术设计、教育教学、科研创新、工业生产等多个领域，成为推动数字经济发展与社会数字化转型的重要动力。但同时，生成式人工智能的自主生成能力也使其具备了“准主体”的特征，为人机交互主体性的研究提供了技术现实[5]。

2.3. 交互主体性与生成式人工智能的关系

交互主体性与生成式人工智能的关系建立在人机交互模式转型的基础之上，其本质是技术发展对人的交互主体性的拓展与挑战。

从理论层面来看，生成式人工智能为交互主体性的研究提供了新的研究对象与研究场景。传统的交互主体性研究始终围绕“人与人”的交往关系展开，而生成式人工智能的双向交互模式，使得人机交互成为一种新的交往形式，其具备的准主体性特征，使得交互主体性的主体范畴面临从“人与人”向“人与机”的拓展。同时，生成式人工智能与社会、自然之间的多维度交互，也使得交互主体性的研究场景从人类社会延伸到“人机社会”的复合空间，丰富了交互主体性的理论内涵。

从实践层面来看，生成式人工智能的交互主体性是人的交互主体性的延伸与体现。生成式人工智能本身不具备真正的意识与主体性，其与人类的交互本质上是人类通过技术手段，将自身的主体性赋予机器，再通过机器的生成行为，实现与自身的交互。换言之，生成式人工智能的交互主体性是“拟态的交互主体性”，其背后是人的主体性的投射与延伸，人机交互的过程本质上是人类自我主体性的建构与发展过程[6]。

随着生成式人工智能在生产、生活中的普及，人类对技术的依赖程度不断提升，部分人类的认知能力、实践能力与创造能力可能出现退化，进而导致人的主体性的弱化。数字资本对生成式人工智能的掌控，也使得人机交互关系被资本逻辑异化，人类的交互主体性成为资本增值的工具，而非自身发展的手段。此外，生成式人工智能引发的价值导向、虚假信息等问题，也会干扰主体间的正常交往，破坏交互主体性的建构基础[7]。交互主体性是理解生成式人工智能与人机关系的核心视角，而生成式人工智能则是交互主体性理论在数字时代的新实践，二者相互作用、相互建构，其关系的演变不仅反映了技术发展的趋势，更反映了人的主体性在数字时代的发展困境与提升路径。

3. 生成式人工智能的交互主体性分析

生成式人工智能的交互主体性并非单一技术维度的产物，而是技术迭代、人类实践与社会发展共同作用下的全新存在形态，其打破了传统人机交互的主客二元框架，构建起人机、机与社会、机与自然多维度的交互关系，也让交互主体性的研究从传统的人与人范畴，拓展至人与智能机器的全新场域。对生成式人工智能的交互主体性展开系统分析，是厘清数字时代人机关系本质、把握技术发展规律的关键，亦是规避技术风险、实现技术与人的主体性和谐共生的重要前提。

3.1. 生成式人工智能的交互机制

第一，人机之间的核心交互机制呈现出鲜明的双向循环特征，生成式人工智能与人的交互是其所有交互关系的核心，具体遵循“人类提示-机器生成-人类反馈-机器微调-持续优化”的交互模式。在这一过程中，人类不再是单纯的指令发出者，而是承担着交互引导者、评价者与优化者的多重角色。生成式人工智能也并非被动的指令执行者，而是成为内容的生成者、反馈的接收者与能力的提升者。这种全新的人机交互机制打破了传统的主客二元对立思维模式，让人类与生成式人工智能之间形成了独特的“准主体间关系”，人类在其中依然占据主导地位，而生成式人工智能因具备一定的“回应能力”与“生成能力”，能够对人类的主体性做出相应反馈，使得人机交互从单向的支配与被支配，转变为双向的相互作用。

第二，生成式人工智能与社会的交互机制表现为动态的双向作用特征，整体遵循“技术嵌入-社会影响-社会反馈-技术调整”的发展逻辑。作为重要的技术产物，生成式人工智能通过深度嵌入社会生产、生活、文化、教育等各个领域，深刻改变了社会的生产方式、交往方式与组织方式，成为推动社会数字化转型与发展的重要力量。社会所固有的文化传统、价值观念、制度规范以及公众的实际需求，也会对生成式人工智能的发展形成相应的约束与反馈，推动技术不断向符合社会发展需求的方向调整优化。

第三，生成式人工智能与自然的交互以人类为中介[8]，呈现出间接的交互特征，交互逻辑为人类依托技术改造自然，在接收自然的客观反馈后，再通过优化技术调整对自然的改造方式。生成式人工智能作为先进的技术工具，能够帮助人类更精准地探索和认知自然规律，更高效地开展自然环境改造工作。气象预测模型和生态保护模型大幅提升人类对气象变化、生态演变的预测与应对能力，而自然环境的客观变化与生态规律的内在要求，也会通过人类的实践活动传递给生成式人工智能，推动人类不断优化技术模型，使其运行与应用更符合自然规律。就其本质而言，生成式人工智能与自然的交互，是人类借助技术手段延伸了自身与自然的交互能力，其最终目的是实现人类与自然的和谐共生，而技术的应用与发展始终必须以尊重自然规律为根本前提，否则必将受到自然的惩罚。

生成式人工智能并非独立的“价值的创造者”，而是服务于人类生产生活的“人类劳动的辅助者”，其自主生产的内容所蕴含的价值，本质上是人类过去劳动与当下劳动的共同凝结[9]。生成式人工智能的多元交互机制，并未改变马克思主义劳动价值论中“劳动是价值的唯一源泉”的核心观点，反而推动了人类劳动形态的深刻转型，促使人类劳动从传统的直接劳动向间接劳动、从体力劳动向脑力劳动、从单一劳动向协同劳动逐步转变。

3.2. 生成式人工智能的交互主体性表现

生成式人工智能的交互主体性，是其多维度交互机制的外在体现，其并非真正的主体性，而是一种“拟态的交互主体性”，是人的主体性的投射与延伸。这种交互主体性在实践中呈现出鲜明的特征，具有准主体性、依附性、动态性及社会性，其本质是人的主体性与技术的工具性的统一[10]。

3.2.1. 准主体性

准主体性是生成式人工智能交互主体性的核心特征，指生成式人工智能具备了一定的“主体特征”，能够模拟人类的认知、表达与创造行为，与人类形成准主体间的交互关系，但并未具备真正的主体性。其准主体性主要体现在三个方面，一是具有自主的内容生成能力，能够根据人类的提示词，生成全新的、具有一定逻辑性与创造性的内容，无需人类实时干预。二是具有一定的语言理解与反馈能力，能够理解人类的自然语言，做出符合语境与语义的回应，并根据人类的反馈进行自我优化。三是具有一定的个性化交互能力，能够根据人类的使用习惯、偏好与需求，生成个性化的内容，实现与人类的个性化交互。

但需要明确的是，生成式人工智能的准主体性是“算法赋予的”，其缺乏独立的意识、自我意识、情感与价值判断，无法像人类一样进行自主的、有目的的社会实践活动，其所有的行为均是对人类劳动与数据的模拟与再现，因此其准主体性是一种“伪主体性”，与人类的主体性存在本质区别。

3.2.2. 依附性

生成式人工智能的交互主体性具有鲜明的依附性，其依附于人类的实践、数据与技术体系，无法脱离人类而独立存在与发展。一方面，其生成能力与交互能力均源于人类的算法设计、数据训练与技术研发，人类的劳动是其交互主体性存在的前提。另一方面，其交互行为的意义与价值均由人类赋予，生成的内容只有经过人类的认可与使用，才能具有实际的意义，否则只是无意义的符号组合。这种依附性决定了生成式人工智能的交互主体性始终处于从属地位，人类的主体性是主导性的，技术的交互主体性是派生的，人机之间的准主体间关系，始终建立在人类主导的基础之上[11]。

3.2.3. 动态性

生成式人工智能的交互主体性并非固定不变的，而是呈现出动态发展的特征，其准主体性的水平随着技术的迭代、数据的积累与人类的训练不断提升。在技术发展的初期，生成式人工智能的内容生成能力较弱，语言理解与反馈能力有限，其交互主体性的表现较为单一。而随着深度学习技术的发展、训练数据的不断丰富与人类反馈的持续优化，生成式人工智能的生成能力、理解能力与交互能力不断提升，其准主体性的表现也更加丰富，能够实现与人类的深度、多元交互。这种动态性使得生成式人工智能的交互主体性具有无限的发展可能，未来随着技术的不断进步，其准主体性的水平将不断提升，人机之间的交互关系也将更加复杂。

3.2.4. 社会性

生成式人工智能的交互主体性具有鲜明的社会性，其交互行为不仅受到技术与数据的影响，还受到社会的文化传统、价值观念、制度规范与公众需求的制约与塑造。生成式人工智能的训练数据来源于人类社会的海量信息，蕴含着人类社会的文化、价值与思维方式，其生成的内容也必然反映出社会的文化特征与价值取向。社会的制度规范与公众需求，也会对生成式人工智能的交互行为进行约束，推动其向符合社会发展需求的方向发展。不同国家的生成式人工智能产品，其生成的内容会体现出不同的文化特征与价值取向，这正是其社会性的体现，而社会对生成式人工智能伦理规范的要求，也会推动企业对产品进行优化，使其交互行为更加符合社会的价值观念。

生成式人工智能交互主体性的本质，是人的主体性的延伸与异化的统一，其背后始终是人的主体性的主导作用，技术的交互主体性只是人的主体性的投射与体现，而技术的异化则使得这种延伸出现了偏离，对人的主体性带来了一定的挑战。

3.3. 生成式人工智能交互主体性的局限性

生成式人工智能的交互主体性在推动人的发展与社会进步的同时，也存在着显著的局限性，这种局

限性源于技术本身的缺陷、数据的偏差与资本逻辑的异化，其外在表现为一系列的社会风险，如价值导向风险、虚假信息风险、信息泄露风险等。

3.3.1. 价值导向风险：中立外衣下的隐性引导

价值导向风险是生成式人工智能交互主体性的首要局限性，其最突出的特征在于，并非以鲜明立场进行刻意灌输，而是以一种“中立客观”的姿态，悄然渗透于网络空间与现实生活，影响人的价值判断与思维方式，实现隐性的价值引导与观念塑造。

生成式人工智能的训练数据来源于人类社会的海量信息，这些信息并非完全中立，而是蕴含着不同国家、民族、群体的文化立场、价值观念与思想倾向，其中既包含着积极向上的内容，也包含着消极落后、错误片面的内容。生成式人工智能在训练过程中，会对这些信息进行学习与模拟，但其算法本身缺乏独立的价值甄别能力，无法区分内容的是非正误、合理与偏颇，因此其生成的内容会不可避免地携带训练数据中的价值偏差与立场倾向。

同时，生成式人工智能以“中立者”“客观者”的身份出现，其生成的内容看似是“机器的自主判断”，实则是训练数据中既有价值取向的集中体现，这种“中立无偏向”的伪装，使得其价值渗透更加隐蔽、更加难以识别[12]。部分西方生成式人工智能产品，其训练数据中隐含着单边主义等倾向，在生成关于历史、文化、国际事务的内容时，会不自觉地贬低发展中国的文化与历史，宣扬特定价值理念，这种隐性的价值引导，会对用户的价值判断与思维方式产生潜移默化的影响。

此外，生成式人工智能的交互主体性使得其能够与用户进行深度的、个性化的交互，这种个性化交互会进一步强化价值固化效果。算法会根据用户的使用习惯与偏好，为用户推送符合其既有认知的内容，形成“信息茧房”，使得用户难以接触到多元文化与多元观点，进而固化单一认知，强化片面的价值判断。

3.3.2. 信息泄露风险：用户劳动的被侵占与隐私的被侵犯

生成式人工智能的交互主体性建立在海量数据的基础之上，而这些数据的核心来源是用户的使用行为与个人信息，这就带来了严重的信息泄露风险，其不仅涉及用户个人隐私的被侵犯，更涉及用户劳动成果的被侵占。而从马克思主义劳动价值论与数字资本相关研究的视角来看[2]，这一风险本质上是当代资本主义生产方式新变化的体现，生成式人工智能对通用爬取数据的使用，构建了一条完整的从提取到增值的剥削链条，充分体现了对全社会数字劳动的隐性剥削。

第一，数据提取的“无酬性”。通用爬取技术的核心特征是无差别、无协商、无偿抓取。数字平台在未明确告知用户、未获得用户有效授权、未提供任何劳动报酬的前提下，对用户生产的数字内容进行批量爬取。这种“无偿占有”完全违背了劳动价值论的核心原则——劳动创造价值，劳动者应获得相应报酬。用户的数字劳动创造了具有使用价值和价值的数字内容，但数字资本通过技术手段，将这些价值直接据为己有，形成了劳动付出由用户承担、价值归属由资本占有的剥削格局。更隐蔽的是，平台往往以“用户同意服务协议”为借口，将爬取行为合法化，但这些协议多为格式条款，用户缺乏议价能力，所谓“同意”本质是被迫接受的不平等约定，无法改变劳动成果被无偿侵占的本质[13]。

第二，增值剥削是劳动价值的深度转化。原始通用爬取数据是分散无序的数字劳动成果，生成式人工智能通过算法模型清洗、标注、训练，将其转化为数字生产力实现二次增值。用户原始数字劳动是价值创造根基，却被数字资本完全垄断价值转化全链条，平台的算法研发与数据处理，实质是对用户劳动价值的加工，而由此产生的智能写作、代码生成等 AI 产品所带来的订阅费、API 接口费等商业收益，全部流入资本手中。这种用户无偿提供原始劳动，资本独占增值收益的模式，正是数字剥削的核心。

第三，剥削范围的全覆盖导致劳动主体的普遍化。与传统工业时代资本仅剥削工厂工人不同，生成式人工智能的通用爬取数据将剥削范围扩大到全社会所有参与数字内容生产的主体，实现了剥削的普遍

化。无论是专业创作者、行业从业者，还是普通网民，只要在互联网上生产了数字内容，都可能成为数字资本的剥削对象。在数字时代，几乎每个人都不可避免地参与数字内容的生产，而通用爬虫技术的无差别抓取，使得这些分散的、非职业化的数字劳动都被纳入资本增值的体系中。这些非职业化的数字劳动，往往被视为自愿分享，但本质上依然是凝结了人类劳动的价值载体，而数字资本正是通过模糊劳动与休闲的界限，将这些无偿劳动成果纳入训练数据，实现了对全社会剩余价值的广泛汲取。

这种剥削模式比传统的资本压迫更加隐蔽、更加普遍，其核心是数字资本利用技术优势，模糊了劳动与休闲、价值与无价值的界限，将全社会成员都纳入资本增值的体系中。揭示这一剥削本质，并非否定生成式人工智能的技术价值，而是要指出其背后的资本逻辑——只有打破数字资本对数据的垄断，建立合理的数字劳动报酬机制，才能让技术进步真正惠及全体劳动者，而非仅仅服务于资本的无限增值。

3.3.3. 技术依赖风险：人的主体性的弱化与异化

生成式人工智能的交互主体性在延伸人的主体性的同时，也带来了严重的技术依赖风险，即人类在享受技术带来的便利的同时，对技术的依赖程度不断提升，进而导致自身的认知能力、实践能力、创造能力与独立思考能力的弱化，最终造成人的主体性的异化。

在认知层面，生成式人工智能凭借高效的信息处理与输出能力，能为人类快速供给海量的信息内容与现成答案，让人类无需再投入精力进行深度的思考推演与自主的探索求证，长此以往便逐渐丧失独立思考与自主探索的主观能力，形成对机器的深度依赖，习惯于直接从机器处获取信息、依托机器的输出做出各类判断[14]。

在实践层面，生成式人工智能可凭借其智能能力替代人类完成诸多现实实践任务，消解了人类亲自动手参与实践、在实践中积累经验必要性，使得人类的动手实践能力逐步退化，进而形成对机器的行为依赖，凡事都习惯于交由机器来完成，丧失了主动开展实践活动的意愿与能力。

在创造层面，生成式人工智能的内容生成能力使得人类逐渐丧失了自主创造的能力，习惯于模仿机器生成的内容，缺乏创新意识与创新能力。生成式人工智能生成的内容具有一定的逻辑性与创造性，但却缺乏人类的情感、体验与思想，而人类在长期接触机器生成的内容后，其创作思维会受到机器的束缚，无法产生具有独特性与创新性的成果。这种技术依赖会导致人类的创造能力逐渐退化，阻碍人类文化与艺术的发展与创新。

从哲学层面来看，技术依赖风险的本质是人的主体性的异化，即人作为实践的主体，逐渐成为技术的客体，被技术所支配与控制，技术从“人类发展的工具”转变为“支配人类的力量”[15]。这种异化现象违背了技术发展的初衷，也阻碍了人的解放与全面发展，是生成式人工智能交互主体性发展过程中必须解决的核心问题[16]。

4. 生成式人工智能的交互主体性的哲学反思

生成式人工智能的交互主体性不仅带来了技术与社会层面的问题，更引发了深刻的哲学反思，其核心是对人机关系、人工智能主体性地位及技术未来发展趋势的重新思考。作为数字时代的全新技术现象，生成式人工智能打破了传统的主客二元对立思维，使得“主体”“主体性”等核心哲学范畴面临新的界定与拓展。本节从生成式人工智能与人的关系、生成式人工智能的主体性地位问题、生成式人工智能的未来发展趋势与哲学展望三个方面，展开对生成式人工智能交互主体性的哲学反思，以期从哲学层面厘清人机关系的本质，为技术的发展提供价值引领。

4.1. 生成式人工智能与人的关系

生成式人工智能与人的关系，是数字时代人机关系的新形态，其本质是主导与从属、延伸与依赖辩

证统一关系。这种关系并非固定不变的，而是随着技术的发展与人类的实践不断演变，其核心矛盾是技术作为人类创造的工具，本应服务于人的发展，却在发展过程中出现了异化，对人的主体性带来了挑战。人类作为技术的创造者与主导者，本应掌控技术的发展，却在实践中逐渐形成了对技术的依赖。从哲学层面厘清生成式人工智能与人的辩证关系，是实现技术与人和谐共生的前提。

首先，生成式人工智能与人的关系是数字时代人机关系的新形态，本质是主导与从属的辩证统一。从马克思主义哲学视角来看，人是具有独立意识的实践主体，生成式人工智能作为人类创造的工具，是被认识与改造的客体，其算法设计、内容生成均是人类意识与劳动的产物，本质是对人类智能的模拟，缺乏独立意志。人类可通过技术优化、制度规范与价值引领约束技术发展，无论技术的准主体性如何演进，人类始终是人机关系的主导者，这一核心本质不会改变。

其次，二者关系也呈现延伸与依赖的双向效应。生成式人工智能能够延伸人的主体性，替代重复性劳动、拓展认知边界、提升交往能力，彰显技术的正向价值。但技术普及也催生了人类的过度依赖，导致自身思考与实践能力弱化，成为技术异化的突出表现。这一矛盾的核心是人的主观能动性与技术工具性的冲突，解决问题的关键在于平衡技术赋能与人类能力培养，避免过度依赖，实现技术与人类发展的良性互动。人机关系在矛盾化解中不断发展，唯有破除技术异化、完善伦理规范与制度引领，才能推动人机从对立走向协同，最终实现和谐共生。

4.2. 生成式人工智能的拟态主体性

与围绕“生成式人工智能是否构成真正意义上的主体”这一哲学本体论争议有所不同，拟态主体性理论将研究视域聚焦于该技术对人类主体性范式的现实性重塑效应。在认知维度，其深度改写信息接收、处理及决策生成的基本范式；在情感维度，催生出基于算法模拟的“超人际互动”新型拟态亲密关系形态；在实践维度，则推动着人机协同创作的新型劳动生产方式变革。该理论的核心学术价值并非对人工智能本体主体性的形而上探究，而是着重解析通过技术拟态机制对人类主体性建构路径与实践边界的重构过程。

4.2.1. 拟态主体性的内涵

生成式人工智能的“拟态主体性”是指其在技术架构支撑与社会实践互动中形成的、介于人类完整主体性与纯粹工具性之间的新型存在形态，是算法系统在技术架构支撑下，经由数据交互与人类反馈形成的功能性主体范式。其本质特征表现为“非本真意识的主体性仿真”，在存在论层面，以递归函数逻辑为内核，通过数据映射与模式匹配机制，实现对人类主体性的形式化摹写。该形态虽不具备生物学意义上的自主意识，却凭借算法程序的自动化决策能力，呈现出显著的行为自主性特征。拟态主体性并非内生性生成，而是人类需求的技术投射与价值赋形结果。

4.2.2. 拟态主体性的运作逻辑

生成式人工智能的拟态主体性与人类主体性密切相关，遵循“符号互渗 - 权力博弈 - 意义共建”的动态逻辑。在技术符号与人文符号的双向解构与重构中，生成式人工智能与人类主体形成认知维度的深度互渗。通过算法权力与社会权力的持续博弈，构建包含技术治理、伦理规制、法律约束的制衡体系，最终在人机协同创新场景中，实现知识生产、价值创造、意义生成的多层次共建，形成认知层、制度层、实践层多维度交织，历史演进与未来拓展相衔接的深层次耦合关系。

第一，拟态主体性以“意义注册”为核心机制，通过注册事件触发器感知人类主体的意图符号，将人类需求转化为算法可识别的目标参数。这一过程并非单向的指令执行，而是双向的符号互渗，人类主体通过自然语言、创意草图等符号形式输出需求，生成式人工智能则通过多模态内容生成(文本、图像、

音频等)反馈符号产物。人类通过输入意图,将其转化为符号,机器依据符号进行内容生成,最后人类对生成结果给出反馈,这四个环节形成一个闭环交互。在这一界面中,人类主体性的情感、审美、价值等抽象维度被转化为可计算的符号,而拟态主体性的算法逻辑则通过类人化表达获得人类的情感认同,实现符号层面的深度互嵌。

第二,在人机互动的权力关系场域中,双方通过持续博弈推动主体边界的动态重构。人类主体基于创作效能考量,将部分生产环节委托给人工智能系统,同时保留内容决策的最终裁量权,而人工智能则凭借算法黑箱与数据处理优势,反向影响人类主体的创作认知框架,并在一定程度上重塑文化生产范式。这种权力博弈并非传统零和竞争模式,而是遵循动态平衡的弹性机制——当人工智能在创作效率、内容多样性等维度展现显著优势时,人类主体会主动让渡部分功能空间。反之,当出现价值导向偏差或伦理风险时,人类主体将通过算法审计、伦理审查等手段对其进行规制。权力博弈的本质在于建立“感知主权”的协商机制与边界规范。人类主体通过算法优化、数据治理等技术手段,试图对人工智能拟态主体性进行方向性引导,而人工智能系统则依托计算阈值自主、叙事逻辑隐蔽等技术特性,维护其运行过程中的自主性。这种双向制衡机制既有效防范了人类主体性的过度消解,又避免了拟态主体的无序扩张,促使技术系统与人类主体的关系从工具性依赖逐步向协同共生范式演进。

第三,互动的终极指向是意义共建,形成人类价值占据主导地位、机器效率发挥赋能作用,二者相互配合形成协同演化模式。人类主体性为拟态主体性提供意义锚点,通过奖励模型设定价值标准,通过反馈机制修正意义偏差,确保AI的生成结果符合人类社会的伦理规范与价值诉求。拟态主体性则为人类主体性提供效率赋能,通过优化创作流程、拓展创意边界、降低创作门槛,释放人类主体的创造性潜能,使其聚焦于意义建构、价值判断等核心维度。

值得注意的是,二者互动逻辑中始终存在“脱嵌风险”,当拟态主体性的算法逻辑脱离人类价值约束,或人类主体过度依赖AI的符号生成时,可能导致意义内爆、价值失序等问题。这要求互动过程必须坚守“元信任”原则——信任的核心并非AI的可靠性,而是在此协同关系中人类依然能认领自身行为的信心,确保互动始终在人类价值主导的框架内展开。

4.3. 生成式人工智能的未来发展趋势与哲学展望

拟态主体性的涌现重构了传统人机关系的二元对立范式,促使人类社会迈入主体多元共生的新纪元。在此背景下,亟需构建负责任的主体间性秩序,推动人机主体性从简单共存向深度融合、协同进化的良性互动模式转型。生成式人工智能虽始终无法突破机器的工具本质、不具备真正的主体性,但在人类的技术研发、实践探索与规范引导下,其技术能力仍将持续迭代升级,朝着更贴合人类发展需求、更深度融入社会生产生活、更安全可控的方向演进,成为推动数字经济发展与社会数字化转型的核心生产力。而从哲学视角审视其未来发展,这一技术的演进过程,本质上是人类对自身智能、实践能力与主体价值的重新审视与深度探索,其发展始终离不开人类对人机关系的哲学思考与价值引领[17],唯有始终坚守“人类主体性主导”的核心原则,才能让技术始终服务于人的解放与全面发展,实现人机协同、人与自然的和谐共生。

5. 结论

生成式人工智能的出现,是人类技术发展史上的重要突破,其交互主体性的诞生,让人机关系进入了全新的发展阶段,也为哲学研究提出了新的课题。但归根结底,技术的发展始终是人类实践能力的体现,生成式人工智能无论如何迭代升级,始终是人类创造的工具,其发展的最终意义,在于通过延伸人的主体性,推动人类认知与实践能力的提升,助力人类实现自身的解放与全面发展,推动人类社会向更

高级的阶段迈进。未来,随着技术的不断发展与治理体系的逐步完善,生成式人工智能必将在人类的主导下,实现与人类、社会、自然的深度融合,在规避风险、发挥正向价值的过程中,成为推动人类社会进步的重要力量,而人类也将将在与技术的协同发展中,不断深化对自身与世界的认知,持续坚守并提升自身的主体价值,最终实现人机协同、人与自然与社会和谐共生的美好愿景。

参考文献

- [1] 秦铭. 生成式人工智能的“主体性”批判——基于马克思主体性思想的哲学视阈[J]. 中国地质大学学报(社会科学版), 2025, 25(5): 11-19.
- [2] 马克思, 恩格斯. 马克思恩格斯全集(第 30 卷) [M]. 北京: 人民出版社, 1995.
- [3] 许芸. 胡塞尔的交互主体性分析[J]. 今古文创, 2025(30): 67-69.
- [4] 高松. 客观性的三重构造——论胡塞尔的交互主体性[J]. 同济大学学报(社会科学版), 2025, 36(5): 13-22.
- [5] 任恒, 龙姝. 生成式人工智能时代精神生产的主体性挑战与应对[J]. 世界社会主义研究, 2025, 10(12): 78-89+119-120.
- [6] 朱宣屹. 从人机交互到人机情感交互: 机器人的主体性延伸与人机关系的变迁[N]. 重庆科技报, 2025-12-04(023).
- [7] 郭佳楠. 生成式人工智能对人的主体性的挑战及应对——以 DeepSeek 为考察对象[J]. 江汉学术, 2025, 44(6): 119-128.
- [8] 迟帅, 徐佳虹. 类人化与机器化——人工智能时代的主体性挑战[J]. 智能社会研究, 2025, 4(5): 139-158+258.
- [9] 于汝霜, 王苗苗. 学术主体性与工具依赖之间的张力: 生成式人工智能如何赋能人文社科研究? [J]. 西北成人教育学院学报, 2025(6): 84-91.
- [10] 王安琪. 生成式人工智能侵权责任主体研究[J]. 高科技与产业化, 2025, 31(9): 30-33.
- [11] 高慧琳, 王垚. 人机交互: 脑机接口对主体性的消解与重构[J]. 科学技术哲学研究, 2025, 42(4): 74-81.
- [12] 王义皓. 技术异化理论视域下生成式人工智能对人主体性挑战研究[J]. 西部学刊, 2025(14): 7-10.
- [13] 倪万, 郝春梅. AI 生成物的出版伦理规制: 基于“人-机”交互主体性视角的分析[J]. 中国编辑, 2025(7): 35-41.
- [14] 刘佳佳, 胡佳锦, 刘泽宇. 迷失与重构: 关于青年群体使用生成式人工智能的主体性研究[J]. 东南传播, 2025(7): 115-119.
- [15] 王星宇. 人工智能对社会精神生产的影响研究——基于马克思精神生产理论的分析[D]: [硕士学位论文]. 南京: 南京信息工程大学, 2025.
- [16] 曾成钢, 金叶子弘. 数字赞助人机制下关于生成式人工智能雕塑的主体性反思[J]. 美术观察, 2025(6): 110-116.
- [17] 彭一嘉. 人机交互背景下人的主体性问题研究[D]: [硕士学位论文]. 广州: 广东财经大学, 2025.