

生成式人工智能的意识形态属性探究

余佳佳

华北电力大学，北京，102200；

摘要：生成式人工智能作为近年科技领域最具代表性的突破性技术，不仅深刻改变了知识生产、信息传播与社会交往的方式，而且在隐秘层面承载并放大了意识形态功能。本文在马克思主义意识形态理论的框架下，结合当代科技发展与资本逻辑的现实背景，探讨生成式人工智能意识形态属性的生成逻辑。文章首先指出，生成式人工智能的技术特性——包括数据依赖性、算法黑箱性与生成结果的拟人化表现——在知识生产与认知建构中天然带有意识形态偏向。其次，科技与资本的合谋进一步放大了其意识形态功能：资本逻辑通过平台垄断、流量控制和价值导向，将技术工具转化为维护既有社会结构与权力关系的“新型意识形态装置”。通过分析，可以发现生成式人工智能不仅是一种生产工具，更是社会关系与权力结构的镜像与再生产机制。最后指出，对生成式人工智能的意识形态反思，有助于在技术快速发展的当下保持主体性，推动其朝向更具公共性潜能的方向发展。

关键词：生成式人工智能；意识形态；技术特性；资本逻辑

DOI：10.64216/3080-1486.25.11.081

“人工智能是引领新一轮科技革命和产业变革的重要驱动力，正深刻改变着人们的生产、生活、学习方式，推动人类社会迎来人机协同、跨界融合、共创分享的智能时代。”生成式人工智能（Generative AI, GAI）则是近年来人工智能领域的重要发展方向之一，是指通过机器学习和人工智能技术来生成数据或内容的算法和模型，生成的内容涵盖文本、图像、音频、视频等多种形式。生成式人工智能与传统人工智能技术的区别在于其能够自动生成新的、未见过的内容，而不仅仅是对已有内容进行分类或识别。也是由于该特点，生成式人工智能也被认为可能是专用性人工智能走向通用式人工智能的关键转折点，目前已经广泛应用于机器翻译、文本编辑、海报设计等领域。然而生成式人工智能并非单纯的技术进步，其生成的原创性内容尽管不能被界定为严格意义上“人的劳动性成果”，但是作为一种生成式“产品”，其自身附着某种“价值”和“意识形态”倾向是毋庸置疑的。算法设计中的偏见、数据来源的选择以及背后资本逻辑的驱动，都可能在不知不觉中强化某些特定的意识形态，加剧我国意识形态领域的斗争，带来新的潜在意识形态风险。马克思曾指出：“如果从观念上来考察，那么一定的意识形态的解体足以使整个时代覆灭。”因此，对生成式人工智能技术带来意识形态安全风险的底层逻辑，即其自身的意识形态属性展开研究具有重要的理论和现实意义。

本文拟从两个维度来探讨生成式人工智能的意识

形态生成逻辑：其一是技术特性所内在蕴含的意识形态偏向，其二是科技与资本合谋所导致的意识形态放大效应。通过分析，可以更清晰地理解生成式人工智能如何在技术与资本的双重作用下，成为新的意识形态生产与传播机制。

1 生成式人工智能技术特性中的意识形态倾向

有学者指出“语言模型是意识形态的量化生产者，潜移默化中可能诱发更为隐蔽、多元和复杂的风险”，理解这一隐蔽风险的关键，在于深入审视生成式人工智能为何会具有意识形态倾向。生成式人工智能本质上是一个以大数据、算法与算力为根基，具备复杂语义理解与生成能力的大型网络，正是其以数据和算法为核心的这一技术本质，使得生成式人工智能不可避免地承载并再现了训练数据中蕴含的意识形态倾向。

1.1 数据依赖和历史偏见再生产

生成式人工智能模型通过海量数据进行训练，这些数据非无源之水，无根之木，其创造于人的行为，每一次点击、搜索、点赞都是一次微小的数字行为，这些行为被记录后，汇聚成了可分析的模式。因此，数据可以被看作人类行为的数字化。单个数据点的价值有限，但数十亿人的数据汇聚起来，就演生成了一种集体智慧或无意识的集体表达。它可以揭示宏观的社会趋势、群体情绪和文化变迁，其价值远超个体痕迹的总和。它反映了人类社会，但并不是完整和真实的反射。这些数据会

过度呈现活跃的、主流的、线上的群体的声音，而弱化甚至忽略弱势的、边缘的和线下的群体的声音。此外，在网络上，用户所留下的“痕迹”并非完全自然。它被平台设计、算法推荐、界面引导影响，可以说数据也被平台规则和商业利益所共同塑造。作为人的产物，数据也不可避免地包含固有的社会偏见、文化倾向和不平等现象。2024 年 3 月，联合国教科文组织发布《大型语言模型中对妇女和女童的偏见》的研究报告指出，当前的大语言模型存在性别偏见、恐同、种族歧视，这些偏见不仅难以减轻，还会被放大。当人工智能学习这些数据时，它不仅学习语法和事实，也潜移默化地吸收了这些内嵌的价值观和偏见，并会在其输出中再现甚至放大这些偏见。

开发人员针对这一数据本身固有的问题，在使用数据训练人工智能之前，会对数据进行清洗和标注，这个环节同样充满了人为的主观判断。为了去除有害内容，开发团队会制定过滤规则，例如删除暴力、仇恨言论、色情内容等。然而，“什么是有害的？”这个定义本身就因文化、法律和价值观而异。对“政治敏感内容”的界定，在不同国家之间存在巨大差异。一个团队依据其所在地的法律和规范清洗数据，必然会将其意识形态标准植入模型。

在数据标注环节，标注者需要根据指南对数据进行分类和打分。标注指南是制定者基于其所在社会的法律法规、文化规范、商业利益和自身价值观所做出的显性选择。这个过程将模糊的社会规范改写成机器可执行的明确规则，其本身就是一种意识形态的输出。即使有详细的指南，标注者在执行时依然拥有巨大的自由裁量空间，他们的个人背景会无形中影响判断。

1.2 算法黑箱与知识权威化

算法本是一种解决问题的工具，从算法的设计目的来看，算法的逻辑是基于现实问题的逻辑，脱离不开问题本身的价值合理性以及解决问题者的主观偏向。算法如果不能反映现实世界的真实逻辑，便是一种无意义的数学公式。不管是在政治、经济领域，还是在社会领域，算法应用场景必然包含了合目的性逻辑。例如，在政治场景中，算法本身就有可能变为政治操纵的工具。从算法的产生来看，算法是由现实的人设计产生的，而算法研发和技术实用化的过程都离不开资本的支持，而工程师作为在模型构建中，创造算法的设计过程不可避免地嵌入了人的价值立场与意识形态倾向。算法工程师通常代表广大用户作出统一的技术选择与决策，规划算

法实现目标的具体路径，并依据自身认知框架或历史用户行为惯例筛选训练数据、确立模型训练范式，从而界定何为“最优”的内容匹配方案。这一系列技术决策，实质上已预设了算法的价值排序与意识形态偏向。所以，“工程师自身的知识储备、价值信仰和思想观念，以及他们所处社会的伦理道德、意识形态和精神文化是影响算法设计的关键因素”。此外，算法的设计也需要数据的“喂养”，工程师们往往希望算法实现某种预期结果而精心选取某一类“指定”数据，或者在数据分析模型的设计中进行人为“扭曲”，甚至对数据的相关关系进行过度解读，这些都左右着算法的价值选择。更值得关注的是，算法不仅会承袭工程师们和历史数据中的价值立场，还可能随着自身迭代更新和数据体量的扩容而强化和放大这一价值立场，从而进一步固化、极化算法推荐的意识形态色彩算法黑箱与知识权威化

生成式人工智能的“黑箱效应”加剧了其意识形态功能。用户很难理解模型如何从数以亿计的参数中生成答案，但却在日常使用中习惯性地接受其结果。正如阿尔都塞所言，意识形态通过日常生活中的“显而易见性”运作，人们在“自然接受”的过程中被动接受了支配性的观念。生成式人工智能的权威化表现就在于：它用技术的复杂性掩盖了知识生产中的价值选择，使用户误以为生成内容是客观、科学的真理。这种技术权威替代了批判性思维，使意识形态通过技术合法化。

1.3 拟人化生成与主体性幻象

生成式人工智能通过自然语言生成、对话交互与多模态表现，营造出一种拟人化的“主体性幻象”。用户往往不再将其视为工具，而是倾向于与其对话、咨询甚至建立情感联系。齐泽克在分析幻想的意识形态功能时指出，意识形态常常通过想象层面的认同发挥作用。生成式人工智能正是通过“拟人”的表现形式，引导用户在幻想中接受其所传递的价值观与知识体系。这种拟人化效果降低了人们的批判性，反而在潜移默化中增强了技术对主体意识的规训。

此外，生成式人工智能可以快速生成真假难辨的文本与图像，这在一定程度上打破了传统的知识与信息鉴别机制。信息真实性的界限被模糊，使得“真假”不再依赖于事实本身，而是取决于平台的权力与技术的输出。这种机制符合鲍德里亚的“拟像”理论，即符号逐渐脱离现实，形成自我循环的意义体系。生成式人工智能的生成内容广泛传播，强化了拟像逻辑，使用户在被动接受的过程中进入一种新的意识形态场域。

综上,生成式人工智能的技术特性天然具有意识形态潜势。它不仅在数据与算法层面再生产历史偏见,也在交互与表现层面建构新的认知模式,从而为意识形态的运行提供了技术基础。

2 科技与资本合谋下的意识形态放大

除了技术本身的偏向性,生成式人工智能的意识形态功能更在于其与资本逻辑的深度结合。“从技术研发到应用,是一个政治的过程,即社会权力参与其中为实现自身的意图展开斗争的过程。”生成式人工智能的研发和运用过程往往被商业资本裹挟,服务于特定的资本目的或政治目的。

2.1 资本驱动与平台垄断

生成式人工智能的研发与运行需要巨额的算力与数据资源,这使得其发展几乎完全掌握在少数科技巨头手中。这种垄断格局不仅限制了技术的公共性,也强化了资本在技术背后的主导权。资本逻辑主导下的生成式人工智能,往往被设计为服务于利润最大化与市场扩张的工具,其生成内容、使用规则与传播方式,都在无形中塑造并维护现有的社会秩序。

2.2 流量逻辑与注意力经济

在平台资本主义的语境下,生成式人工智能往往被用来吸引注意力、制造流量与推动消费。算法生成的内容不仅追求用户粘性,还在价值取向上强化消费主义、娱乐化与碎片化趋势。这种逻辑使得技术成为注意力经济的加速器,将意识形态的功能深植于用户的日常生活与文化体验之中。

2.3 技术治理与意识形态规训

资本与国家权力往往通过对生成式人工智能的技术治理,实现对公共舆论与社会认知的规训。例如,生成式人工智能在内容审核、信息分发等环节,隐含着对价值观与意识形态的筛选与塑造。用户在与技术交互的过程中,事实上已处于一种隐形的意识形态规训之中。

因此,生成式人工智能的意识形态属性,不仅仅是技术偏见的被动投射,更是资本逻辑与权力结构的主动塑造。它构成了一种新的“意识形态装置”,在社会再生产中发挥深远影响。

3 结语

生成式人工智能作为一种新兴技术形态,正在重塑社会的知识生产方式与文化传播机制。然而,从马克思

主义意识形态批判的角度来看,这种技术并非中性的生产力,而是与社会关系、资本逻辑与权力结构紧密耦合的复杂装置。其意识形态属性的生成逻辑,可以从技术特性与资本合谋两个维度来理解:前者使其天然带有社会偏见与认知偏向,后者则放大了其在维护既有秩序、规训主体意识方面的功能。

因此,对生成式人工智能的批判性研究,需要超越单纯的技术视角,而应放在社会关系与历史逻辑的广阔语境中加以审视。这不仅有助于揭示其潜在的意识形态风险,也为探索其公共性潜能提供了理论路径。唯有在技术治理、资本约束与社会批判的共同作用下,生成式人工智能才能真正服务于人的解放,而非成为新的意识形态束缚。

参考文献

- [1] 习近平向国际人工智能与教育大会致贺信[N]. 人民日报,2019-05-17(001).
- [2] GoogleCloud. Introduction to generative AI. [2023-07-28]/[2023-08-11]. https://www.cloudskillsboost.google/course_templates/536.
- [3] 陈永伟. 超越 ChatGPT: 生成式 AI 的机遇、风险与挑战. 山东大学学报(哲学社会科学版), 2023(3): 127—143.
- [4] 中共中央马克思恩格斯列宁斯大林著作编译局. 马克思恩格斯文集: 第8卷[M]. 北京: 人民出版社, 2009.
- [5] 王海威. 人工智能诱发隐性意识形态话语风险的逻辑机理及化解策略[J]. 马克思主义与现实, 2024(4): 124—133.
- [6] UNESCO IRCAI. Challenging systematic prejudices: an investigation into gender bias in large language models [EB/OL]. [2014-03-13] (2023-05-06). <https://unesdoc.unesco.org/ark:/48223/pf0000388971>.
- [7] 李静辉. 推荐算法场域思想政治教育的危机与变革[J]. 甘肃广播电视大学学报, 2021(4): 59.
- [8] 达拉斯·斯迈思. 自行车之后是什么: 技术的政治与意识形态属性[J]. 开放时代, 2014(5): 96.

作者简介: 余佳佳, 2000-06, 女, 汉族, 江苏徐州人, 硕士研究生, 华北电力大学, 研究方向: 马克思主义理论。