

# 生成式人工智能（AIGC）技术应用及监管研究

魏晓光<sup>1</sup> 吕力<sup>1</sup> 祁俊贤<sup>1</sup> 徐彰义<sup>1</sup> 张倩<sup>2</sup>

1. 河北金融学院; 2. 河北软件职业技术学院

**摘要:** 生成式人工智能是一种基于海量数据训练的大型语言模型, 具备预训练语言模型、上下文学习和基于人类反馈的强化学习三项关键技术, 在教育、金融等领域有良好的应用前景。然而, 生成式人工智能技术的推广发展面临着技术、法律等多方面风险, 应遵循“分级分层”管理原则进行防范治理, 构建和谐稳定的技术发展环境, 推动生成式人工智能技术对社会各个领域的深度赋能。

**关键词:** 生成式人工智能; 自监督学习; 风险防范

【DOI】10.12252/j.issn.2096-627X.2023.10.111

## 一、引言

生成式人工智能是一种在大语言模型训练的迭代学习下, 自主生成新的文本、图形和音频等内容的人工智能技术。以ChatGPT为代表的生成式人工智能技术不断向通用化方向发展, 成为当今世界推动社会发展的新型动力。生成式AI具有预训练语言模型、上下文学习和基于人类反馈的强化学习三项核心技术, 在教育、金融等领域具有良好的应用前景。本研究将分析生成式AI结构原理及技术优势, 阐述生成式AI赋能社会发展的方向及风险防范措施。

## 二、生成式人工智能结构原理及技术优势

### 1. 预训练语言模型

生成式AI是一种结合对问题的理解生成类人化回答的大型语言模型, 通过大量数据投入及训练, 生成式AI创新能力不断提升。一方面, 生成式AI通过大量数据训练其文本自动生成能力, 并不断增加预训练数据, 扩大预训练模型规模来实现性能的进一步提升。另一方面, 通过大规模数据参数进行自监督学习, 使模型学习人类语言的框架结构和不同语境下的不同语义。预训练语言模型的研究涵盖了模型结构、训练方法、计算效率、理论分析和应用范围等多方面, 为生成式AI的发展提供了良好的技术支持。预训练语言模型通过自监督模式对所接受的海量数据进行语言学习, 在输入语料中随机掩盖部分内容, 训练模型对被掩盖内容进行预测。通过自监督学习, 预训练语言模型不仅将词语的语义描述从静态表示提升为上下文感知的动态表示, 而且为自然语言处理的任务提供了统一的建模框架<sup>[1]</sup>。随着预训练语言模型的不断完善, 生成式AI能针对复杂语言环境生成具有连贯性、条理性、层次性的文本内容, 为AI对自然语言的理解和生产奠定了更加坚实的技术基础。

### 2. 上下文学习

上下文学习技术是生成式AI通过对上下文语境、语义的理解, 给出最贴合使用者期望的智能回答, 是区别

于判别式AI的一大关键核心技术, 也是自然语言处理任务新常态下的一种新模式。在上下文学习过程中, 模型会在海量数据预训练的基础上, 结合历史数据分析当前输入数据的语料信息, 从而进行预测和学习。通过综合考量上下文的信息内容, 模型可以更精确地理解所输入的信息, 输出类人式回答, 同时做到在不同情境下结合具体上下文信息给出针对性答复, 推动生成式AI走向通用化道路。生成式AI的自监督学习模式直接在预训练语言模型基础上进行学习和预测, 不需要对模型数据进行更新, 节省了数据成本, 提高了模型使用效率。

### 3. 基于人类反馈的强化学习

生成式AI常常利用使用者的反馈进行强化学习, 目的在于通过与使用者进行交互交流得到有效反馈来优化模型训练, 从而在复杂任务中提高对话质量, 将专业知识运用于强化学习过程中以实现生成式AI的高效应用和发展。强化学习主要包括以下三个阶段: 一是生成式AI收集大量数据构建奖励模型。要先搭建数据集, 收集给定问题及其大语言模型生成的不同回答, 人类标注者需对特定问题生成的不同答案结合具体情境进行综合排序, 接着训练奖励模型预测符合使用者对于不同生成式回答的喜厌程度并进行排序, 进而搭建满足使用者喜好的奖励模型。二是生成式AI通过强化学习算法得到更优模型。生成式AI基于使用者的反馈信息进行学习算法强化, 最终训练得到满足使用者偏好的生成式AI模型。三是生成式AI根据奖励模型得到的回答进行评估优化。人工智能在进行自我学习、训练的基础上, 根据奖励模型的使用反馈不断优化自身产出, 在复杂的任务情境中不断提高自身性能来实现生成式AI交流的自我学习与优化。具体而言, 生成式AI在使用者进行提问后产生回答, 接着奖励模型对其回答进行评估, 之后生成式AI会根据评估去调整参数以便更好地回答下一次问题, 在生成式AI的不断学习、优化中, 减少了模型的不利输出, 实现了高质量的交流优化。

### 三、生成式人工智能赋能社会发展

#### 1. 生成式AI助力金融改革

当前，生成式AI已经被落实于多种应用场景，例如国内多家银行、券商等已经宣布接入通用大模型，并结合行业知识进行微调。生成式AI在个性化金融服务、自动化客户服务、风险预测和管理、产品创新和市场预测等方面发挥了积极作用。

##### (1) 个性化金融服务

生成式AI通过分析客户的偏好、行为和需求为客户提供个性化的金融建议、投资组合和理财规划。通过模拟和学习客户的行为，生成式AI能够为个体客户提供更符合其需求的金融服务，从而改善客户满意度。例如中国的衍生品交易量呈逐年递增的发展趋势，场内交易量增大意味着需要面对衍生品交易风险的客户逐渐增加，生成式AI可以模拟客户的风险承受能力为客户提供交易建议，最大程度满足客户交易需求。在场外市场中客户交易需求更加巨大，生成式AI的作用也更加显著化。

##### (2) 自动化服务客户

生成式AI可以为客户提供自动化服务，为客户提供即时、个性化的支持和解决方案，有助于提高服务客户的效率，降低金融机构的运营成本，并能够更好地应对客户需求的快速变化。对于期货交易的中介市场而言，交易双方标的种类、交易价格需求、交易利率需求不同，中介公司利用AI精确客户需求并在最大限度地降低公司运营成本。

##### (3) 风险预测和管理

生成式AI通过分析大量的金融数据和市场信息，帮助金融机构和个体客户进行风险预测与管理。不同市场存在不同的风险差异，生成式AI通过模拟不同的市场情景和交易策略，有助于发现潜在风险，并针对市场中的不同风险提供风险规避的差异化建议，从而提高金融机构的风险抵御能力，为个体客户最大程度降低风险损失制定风险对冲方案或风险规避策略。

生成式AI也可用于分析和识别金融欺诈行为，通过学习客户的交易模式和行为特征，识别异常交易和欺诈风险，同样有助于金融机构及时发现并遏制欺诈行为，维护金融市场的公平和健康。

##### (4) 产品创新和市场预测

生成式AI不仅是对已有数据的复制粘贴，还可以在原有基础上进行再生成，帮助金融机构开发新的金融产品和服务，挖掘新的市场机会。通过分析市场趋势和客户需求，生成式AI能够预测市场变化并提供相关建议，有助于金融机构更准确地洞察市场和客户行为。生成式AI在金融领域的广泛应用，有助于提高金融服务的个性

化程度、改善客户体验、提升企业效率、强化风险管理和预测能力，并促进金融产品与服务的改革创新。

#### 2. 生成式AI辅助教学升级

在教学领域，生成式AI利用其技术优势可以打破原有的信息壁垒，提升教育教学质量，促进师生协同发展。

##### (1) 生成式AI双向交流模式促进师生共同发展

在数字时代的大背景下，教师作为课堂主导者为推进智能化课堂建设，应积极利用ChatGPT等生成式人工智能技术打造“共赢课堂”，提升教育教学治理能力。生成式AI可辅助教师提高备课质量，建立个性化、多元化、包容化的多层次教学模式，培养学生自主探索学习的意识，挖掘自身潜能的同时促进教师教学质量的提升和教师自身素养的提高。ChatGPT为代表的生成式人工智能在促进学生个性化学习的同时，提高了教师数字化能力素养，实现了师生共同发展。

##### (2) 生成式AI智能交互促进个性化教学

将理论知识运用于实践是教育教学的真实目标，但传统单一化教学模式无法满足学生的个性化需求，生成式AI的智能交互突破原有教学模式单一的限制，促进个性化学习，生成适合每一位学习者认知习惯的教学内容。在传统教育体系中中学生素养提升与知识获得存在严重脱节，教学模式单一、教学内容固定，受教育者丧失对学习实践的主动性和创造性，导致我国创新型人才培养受阻，创新氛围严重缺失局面的形成。生成式AI应用于教育领域，可以使受教育者清晰地认识到自身不足，精准了解知识储备缺陷；为受教育者提供精准、个性化分析，创建智能化学习系统，增强受教育者的学习获得感、体验感；使受教育者对学习产生主动探索兴趣，提升情感认同达到情感共鸣的效果，深化受教育者对世界的理解。此外，生成式AI个性化学习模式的可选择性极大地提高了受教育者的学习兴趣，使受教育者由被动学习转变为主动学习，使学习模式的选择性由单一向多元化转变，为我国教育数字化高质量发展提供人才输出。

生成式AI通过分析与学生对话内容，结合上下文语境、观点等给出针对性意见，为学习者提供适合学习者认知结构的作业，形成差异化作业体系，实现受教育者的个性化发展。生成式AI提供适合学习者作业的同时进行作业批改反馈，实现育人模式的高质量发展，提高作业的精准程度。

##### (3) 生成式AI打破交流壁垒，完善教学评价体系

生成式AI可以实现教授者和学习者的双向高效沟通，使学习者从意见的被动接受者转变为与教育者意见平等交换者，实现教学评价体系向高阶综合方向迈步，

推动静态评价体系向动态化转变。在教育领域中生成式AI会更有针对性的为每一位受教育者提供适合的个性化教学方案，以此推动教学评价智能化转型。生成式AI赋能教育数字化转型，使教学工作由“单向传导”逐渐转变为“双向互通”的意见交流方式，打破交流壁垒态势下所呈现的信息差。生成式AI推动教育数字化高质量发展，推动教育技术革新、理念更新、模式创新，实现教育高质量发展。由此可见，生成式AI赋能教育转型升级为教育行业的发展提供了无限可能，是教育智能化建设的重要基石，为推动我国教育事业高质量发展，走可持续发展道路提供了技术支持。

### 3. 生成式AI赋能医疗健康服务

以ChatGPT为代表的生成式人工智能推动了众多领域的变革，也将影响医疗健康领域的发展。通过收集大量数据对人的语言行为进行模仿是生成式AI的重要功能，其语序结构、用词方式都与真人高度相似。如果将大量的医疗数据信息输入语言模型形成有关医疗健康的语料库，实现健康、医疗与生成式AI的连接，使生成式AI成为“线上医生”代替真人医生与患者进行健康问题的解答互动。生成式AI根据用户提供的信息在丰富的数据库和语料库中寻找相似病例，对用户身体健康状况进行分析，辅助医生制定适合用户自身特点的健康计划，帮助用户了解更多健康信息。生成式AI辅助医疗诊断，为医生减少了大量的时间损耗，医生可以将更多的时间与经历投入到医学研究等方面，使医疗资源得到充分应用。

### 四、生成式人工智能技术应用的风险监管

生成式人工智能技术为社会各个领域发展赋能的同时，也带来了一些风险和挑战，如何防范化解这些风险是当前推进生成式AI走向通用化道路必须解决的问题。

#### 1. 规范数据使用，严格把控技术风险

近年来，随着生成式AI深度学习的不断发展，生成式AI在文本、图像、音频等方面取得了显著成效，同时技术风险也不断显现，需要采取相应措施进行治理防范。其一，用户隐私数据泄露一直是互联网时代下的重点安全问题，生成式AI需要大量数据进行训练，这可能导致用户信息的滥用或者在模型生成内容时泄露敏感信息。其二，随着算法的不断更迭创新，生成式AI容易在行为人的故意训练、引导下输出虚假信息。生成式AI凭借其强大的学习功能，模仿不同的人的对话风格，生成极尽相似的内容，成为谣言散播者的强有力“工具”，引导社会舆论走向。

生成式AI首先要建立严格的数据使用和审查机制，

对用户隐私数据使用加密、脱敏等手段，积极保护使用者权益。研究人员要在不断优化算法结构的同时，提高自身素养，树立正确道德伦理观、价值观和社会责任意识，引导公众舆论正向发展，提供良好的技术使用环境。

#### 2. 科学立法，防范法律风险

在当今法治社会的背景下，生成式AI的运用应当遵循法律的框架和原则，确保使用过程合法、公正、透明。但在实际应用过程中生成式AI存在以下问题：其一，由于生成式AI会根据使用者的个人数据生成相关敏感信息，造成隐私泄露等问题。其二，生成式AI的智能化可能放大社会偏见、歧视，导致生成不公平、不公正的内容。其三，生成式AI在其自主生成过程中，由于法律意识的不明晰极易容易导致侵犯他人知识产权等法律问题，造成多方利益共同受损。政府应完善相关法律法规对生成式AI算法进行限制和规范<sup>[2]</sup>，同时有关部门应加强立法监管明确生成式AI的适用范围，明晰责任归属，为技术的正向发展提供法律保障。

#### 参考文献

[1] 张熙, 杨小汕, 徐常胜. ChatGPT及生成式人工智能现状及未来发展方向[J]. 中国科学基金, 2023, 37(05): 743-750.

[2] 侯东德, 张丽萍. 生成式人工智能背景下网络信息生态风险的法律规制[J]. 社会科学研究, 2023, (06): 93-104.

项目来源:

2023年度河北金融学院大学生创新创业训练计划项目(《生成式人工智能对教育方向转型的影响——以ChatGPT为例》项目编号202311420009)

2023河北金融学院第三期师生学习共同体项目(《生成式人工智能支持师生科研探究——基于百度文心一言的实践》)

作者简介:

魏晓光, 男, 河北定州人, 博士研究生, 河北金融学院副教授, 研究方向: 金融科技、智能传播。

吕力, 女, 河北金融学院21级金融工程专业大学生。

祁俊贤, 女, 河北金融学院21级金融工程专业大学生。

徐彰义, 男, 河北金融学院22级金融科技专业大学生。

张倩, 女, 河北软件职业技术学院, 讲师, 硕士, 研究方向: 智能传播、文化传播。