

生成式人工智能赋能社会科学：应用探索与风险治理综述

徐嘉，袁璐

中国传媒大学

文章信息	摘要
<p>关键词: 生成式人工智能 社会科学 应用 风险</p>	<p>这篇综述的目的是对影响社会科学所使用的生成式人工智能进行针对学科的应用整合以及在不同情境下的风险整合，本文的出现为社会科学人工智能交叉领域的研究者提供了前沿的研究视角，为交叉学科的研究打下重要的基础。特别地，本文将 GAI 的应用方向进行了基于 GAI 身份的分类：作为生产力的根源、作为助手协助生产、以及模拟特定身份进行交流和实验。随着 GPT-4 等先进人工智能语言模型的出现，GAI 在社会学、心理学、政治学、经济学、人类学、伦理学、历史学、教育学和传播学等学科中展现出了巨大的潜力，GAI 的出现使人类的生活发生了巨大的变化，GAI 的强大能力也推进了科学研究的进步，我们必须承认，GAI 的高速发展极大地便利和改进了我们的生活，但随着这份福祉而来的是由于 GAI 强大能力而引发的许多挑战和风险，从引起的失业焦虑到今日备受关注的 AI 对齐问题，GAI 所带来的问题与引起的风险，正在极大地威胁着人类的权利以及社会的安定，本文将 GAI 的风险方面也进行了分类，我们将问题分为两个大的部分：Problem 和 Governance，我们在这两个分类下我们又根据生成式人工智能应用于社会中带来的风险及问题进行了细致的分类。基于大量的学科以及跨学科文献，我们对现有的主流大语言模型进行了多方面的整合，并探讨了 GAI 在社会科学中的应用及其在各学科中的潜在应用方向，最后探讨了 GAI 所带来的问题与风险。我们的综述对人工智能与社会科学的学科交叉研究有着十分重要的意义，为社会科学领域的 GAI 研究者和从业者提供相关交叉学科的深入见解，帮助他们理解 GAI 对社会科学的影响，为相关研究者提供了方向性指导，并呼吁对 GAI 的应用以及风险进行更深入的研究。</p>

1. 引言

生成式人工智能(GAI)的出现为社会科学带来了前所未有的可能性和挑战，目前的生成式人工智能作为以 AGI 为最终目标的智能目前能够带给我们什么，如何更好将 GAI 应用于不同的社会科学领域如何应对 GAI 因应用不当和其难以定义的社会身份而带来的问题。

最近生成式人工智能领域取得的重大进展，例如 OpenAI 旗下的 GPT-4^[1]，这些 GAI 的出现为社会科学带来的许多可能性，例如在社会学中的多 agents 社会模拟^[15]，在传播学中实现虚假信息的识别^[56]，在经济学中进行在不同环境中的政策模拟^[74]等。对于生成式人工智能，我们应视其为一种工具，在劳动力方面，GAI 可能会在未来取代一些基础的工作，但在 GAI 的竞争压力下生产力的质量也会提升，GAI 不会取代人类，未来是人机交互共存的时代。

现有的研究缺乏对人工智能与社会科学多学科交叉的详细研究，本篇的出现可以弥补交叉学科研究的缺漏、本篇以“让 GAI 成为更好的工具”观点出发，对 GAI 进行了基于人机交互的身份分类，分别为：作为最根本的生产力、作为助手协助人类进行生产、模拟特定个体或群体进行对话和模拟环境进行实验，分类并非完全分离的三部分，这三种分类是一种递进的关

系，分类旨在让人们对 GAI 在社会中的身份有着更明确的认知，为了实现更好的人机交互模式，我们必须明确 GAI 对于人类的交互身份，只有确定了 GAI 的交互身份，我们才能最大限度的避免智能涌现后出现的人际关系混乱的问题，也可以避免许多由于人机关系模糊不清而产生的伦理道德问题。

本文着眼于将生成式人工智能与社会科学中的学科进行交叉学科研究，研究内容包括生成式人工智能在学科层面的应用以及基于学科层面的风险探讨，我们秉持着让 GAI 成为更好的工具的观念进行了下列研究，第三部分为针对目前大模型能力信息的研究、对比和整合，第四部分为基于社会科学学科层面的应用及风险探讨，这一部分不仅包括基础的学科交叉应用，还包括前沿的研究成果以及未来的研究前景，第五部分为针对生成式人工智能所带来的风险的研究，我们将风险分为模型问题（Model problem）和治理问题（Governance problem），对已经出现的问题和未来可能出现的问题展开了研究，为未来的研究者提供风险指南和解决方向。

2. 背景

生成式人工智能的出现对社会产生了极大的影响，这些方面包括但不限于教育、医疗、经济等领域。要想了解生成式人工智能作为一种新兴技术如何影响我们的生活，我们首先要考

考虑到技术和社会之间的关系，技术的定义在不同的时期都有着不同的定义，Rudi Volti 在其撰写的书中提到对技术的定义：“A system created by humans that uses knowledge and organization to produce objects and techniques for the attainment of specific goals”^[2]这个定义很好的阐释了技术与人之关系，技术是应人的需求而产生的，技术应着人的技术热情（technological enthusiasm）和应用需求而产生，技术的发展又会影响着人们的动力水平、工作方式、思维模式^[3]。人们对自动化追求和对类人智能体的好奇心驱使了当今的人工智能热潮的产生，生成式人工智能作为一种技术工具，通过自然语言处理（NLP）和深度学习，GAI 可以进行复杂的数据分析、语言理解和生成任务，这些能力在社科中的研究有着巨大的潜力，这些能力使得社科中的许多领域有了进行更多种类、更高效研究的可能。

3. 方法

过去所存在的研究大部分为针对单一社科学科与人工智能的交叉研究，本文对社科与人工智能交叉的主要方面都进行了研究和探讨，不仅如此，为了更好的阐述 GAI 在社会科学中的应用关系和人机关系，我们提出了三个问题：

- RQ1.生成式人工智能的出现对社会科学产生了什么影响？
- RQ2.如何针对生成式人工智能和人构建更好的人机关系？
- RQ3.生成式人工智能的出现带来了哪些短期内会产生问题以及哪些长期发展可能会显现的风险？

我们使用了一种全新的分类研究方式，我们构建了基于 GAI 应用于社会科学中的身份和人机关系的分类，由于 GAI 的人机交互性，分类并非完全割裂，而是一种递进的分类形式。

3.1. 基于社会身份的生成式人工智能分类

3.1.1 作为最根本的生产力工具

GAI 作为最根本的生产力工具的应用方面包括基础的数据分析、信息处理、文本输出、模态转化、部分工作的自动化等方面。

3.1.2 作为助手协助人类进行生产

GAI 作为助手协助人类进行生产的应用方面包括通过人机交互实现生产率的提升、进行信息处理为决策和预测的进行提供数据支持、作为网络集成入口成为新型访问媒介、辅助进行 Deepfake 内容的识别等方面。

3.1.3 模拟特定个体、群体进行对话或模拟环境进行实验

GAI 模拟特定个体、群体进行对话或模拟环境进行实验的

应用方面包括：模拟个体思维、模拟个体之间的交互、模拟群体社交行为、模拟难以调查的人群、模拟情感、虚拟现实实验、模拟经济或政治决策场景等方面。

4. 各学科视角下的应用与风险探讨

4.1 生成式人工智能与社会学

生成式人工智能拥有强大的多模态信息理解与生成能力，自其进入大众视野以来已经对社会各领域带来了巨大影响。作为最根本生产力的人工智能已进入各行业进行基本生产工作，极大改变了劳资关系与对劳动者的需求，带来了一系列社会影响；在社会学研究的领域，生成式人工智能同样具有极大的潜力，研究者们将其应用在计算社会学，研究方法设计和社会情景与人际关系模拟等方面，帮助研究者们更好地模拟和理解人类在社会环境中的行为与决策过程，让我们能借助这一新技术更深入，更全面地了解人类社会，提出社会问题解决方案。但社会学学者在接近这一新技术的同时也应保持谨慎，生成式人工智能在社会学的应用中也可能带来问题和风险，如用户的隐私问题，偏见内容与刻板印象问题，生成低质量内容甚至产生“幻觉”等问题以及许多还未发现的潜在风险，在应用新工具之前我们需要清晰且全面地认识这些问题并保持谨慎的态度，使人工智能更安全，更有效地应用于社会学与人类社会。

4.1.1 生成式人工智能作为最根本的生产力工具

生成式人工智能基于本分类在社会学领域的应用包括：基于生成式人工智能对文本的理解能力进行的社会学文本分析与定性研究的数据处理，设计研究方法等。

（1）社会学文本内容分析：大语言模型（LLMs）类的生成式人工智能无需训练数据就能顺利完成许多社会学文本处理任务，例如情感分析、主题建模和文本分类等。Hayes 认为生成式人工智能让研究者实现与数据的动态交互，他使用了 open-ai 的 GPT-4 分析了十个半结构化的采访记录，通过提示让模型识别并给出了共同的主题，且提取出了文本中的理论洞见和框架^[4]。Wu 等人将美国第 116 届国会的参议员随机配对后传递给 GPT-3，提示其根据他们此前的政治发言判断他们中哪一个更自由，哪一个更保守，其结果与主流研究方法 DW-Nominate 得到的结果相似^[5]。

（2）作为定性研究的数据处理工具：生成式人工智能最为直接地影响了计算社会学领域^[6]；Nelson 等人认为，LLMs 与传统的计算相比，对于定性分析更加有利。现有的计算社会学研究方法在处理相对较大的语料库时，往往通过“预处理”提取出主要语义信息，但同时也会丢失一部分有价值的语境和词性信息。但像现场笔记和访谈记录这样的定性数据往往有着多个组成部分，失去语境和词性信息是无法接受的。由于生成式人工智能

将语言以更一般的形式表现，拥有更大的灵活性，因此可以无需预处理，直接分析这些复杂的，多方面的文本，避免使用传统计算技术时容易丢失的细微差别^[7]。

(3) 社会学研究方法设计：在社会学研究中，研究方法设计的好坏往往决定了是否能较好地收集到研究者所需的数据。生成式人工智能目前已经在量表开发，问卷设计等方面被广泛应用。Götz 等人使用基于 GPT-2 的生成语言模型 PIG 对新的，未经研究的结构（衡量爱好旅游的程度）和已建立的结构（人格特质测量）两个方面进行量表设计，并验证了自动生成的量表在野外测试时表现效果与现有的评估黄金标准一样好^[8]。

4.1.2 生成式人工智能作为助手协助人类进行生产

生成式人工智能基于本分类在社会学领域的应用包括：通过实时建议改善对话，协助人类进行社会学相关数据标注以及辅助进行更广泛意义上的劳动等。

(1) 利用生成式人工智能对文本的即时响应，Argyle, Bail 等人开发了一个人工智能聊天助手，在有关分裂的在线政治对话场景中进行实时的，基于证据的建议。这样的干预措施能改善对话的基调而不会改变对话内容或侵犯对话者的政治立场，从而达到减少社会冲突，促进民主互惠的社会目的^[9]。

(2) 在机器学习领域的众多算法系统都需要人工标注的数据进行启动，这些数据标注者组成了一个新兴的工人群体，同时也带来了如劳资关系这样的社会问题^[10]。而生成式人工智能进入大众视野后，Gilardi、Alizadeh 和 Kubli 等人发现大语言模型可以提升数据标注效率，比众包数据标注者更好地完成特定工作（零样本准确率高出 25 个百分点），且成本比众包工作者便宜三十倍，具有大幅提升计算社会学中数据标注效率和解决众包工作者带来的社会问题的潜力^[11]。但 Ziems 等人发现生成式人工智能不是完美的，它同样有着容易误解指令和犯错误的问题，因此他们建议将节省的资金用来培训专家注释者，采用人工智能与专家注释者结合的模式，提高数据标注的质量^[12]。

(3) 从现有研究来看，生成式人工智能在更为一般意义上的劳动也具有很大潜力。现有的大语言模型在文书工作上以及能很好地协助人类，具有具身性的生成式人工智能在体力劳动上也能很好地提供辅助。Barbara 等人认为，人工智能能够替代工人执行常规的、可编码的任务，人工智能与人类劳动之间存在互补性，它们并不会代替人类劳动，而会让人类更集中精力于创造性工作^[13]。

4.1.3 生成式人工智能模拟特定个体、群体进行对话或模拟环境进行实验

生成式人工智能基于本分类在心理学领域的应用包括：模拟人类个体思维，模拟群体社交行为，模拟难以调查的人群

(1) 生成式人工智能拥有模拟人类个体思维方式的能力。Argyle, Busby 等人为 GPT-3 提供“背景故事”，要求其表现得像一

个强大的民主党人，并用“无知”，“种族主义者”，“厌女症患者”，“恐同”等提示词赋予生成式人工智能更明确的习惯进行模拟选举研究^[9]而 Anton 等人将大语言模型接入了游戏《外交风云》，这类战略游戏与围棋等不同，人工智能无法纯粹通过自我游戏来学习迭代，而是需要有意地与人类交流来理解其他参与者的动机和观点，并通过自然语言协商完成复杂推理，人工智能在其中进行强化学习，增强自然语言处理与表达能力，从而更好地完成人类个体模拟^[14]。

(2) 接入多个生成式人工智能进行群体社交行为模拟同样可行。Joon.S.P 等人在一个虚拟沙盒小镇环境中接入了 25 个生成式人工智能代理，赋予每个个体不同特征与个性，让它们在环境中相互交互。最终这些代理不仅能完成日常生活模拟，还表现出了群体特性（如组织选举活动，派对等）^[15]。

(3) 生成式人工智能在模拟一些生活中具有危险性或可能会给人类参与者带来潜在创伤的研究时格外具有优势（如对极端恐怖主义分子和网络性骚扰者的研究）。Igor 等人认为，这样的主题可能会给人类参与者带来困扰，因此通过生成式人工智能来模拟是一种妥协的解决方式，通过这样的方式能够得到原本难以得到的信息，为防止此类行为的社会干预提供帮助^[16]。

4.1.4 生成式人工智能在社会学中可能带来的问题和风险

(1) 隐私问题：现有的生成式人工智能通常会在与人类的交互中进一步训练模型，我们在将数据输入到像 ChatGPT 这样的生成式人工智能中时，也就意味着将数据与第三方共享了。Spirling 等人认为这会带来隐私问题，一些处理敏感性数据的学者应该明确认识到这些泄露隐私的风险，谨慎将敏感数据输入商业性非开源的生成式人工智能，否则可能会被模型以某种形式保留并输出^[17]。

(2) 偏见问题：Bianchi 等人认为生成式人工智能是使用大量未经审查的人类创造的数据进行训练的，因此会带有原始数据中存在的偏见与刻板印象^[18]例如美国对非裔社区内的犯罪报道更为频繁导致数据库内这类人的犯罪数据占比更高，因而使生成式人工智能在犯罪预测上对非裔群体产生了偏见^[19]。

(3) 内容质量问题：目前的生成式人工智能并非基于形式逻辑或真值生成内容的，而是基于大量文本预测下一个最大概率出现的单词或单词序列，这就引发了人工智能生成内容可靠性的问题。Bender 等人发现同一模型即使在大多数情况下输出符合事实的正确内容，也有可能产生虚构的误导性信息，这些研究人员将这些模型以自信的方式提供不准确的内容的行为称为“幻觉”，并认为目前的人工智能不适宜进行严肃性任务^[20]。

4.2 生成式人工智能与心理学

生成式人工智能和心理学的关系可以说是相辅相成，人工智能本质就是模拟或实现类人的智能，但如何将人脑的工作方式进行抽象并运用到人工智能上是目前的一大挑战。联结学派被人脑神经系统启发而设计的深度学习神经网络已经在计算

机科学、机器学习和人工智能领域中取得了显著的成功，但人工神经网络本质上与人脑神经系统还是有着很大的差别：比如人脑的编码方式是基于“01”出现的频率进行编码并且人脑神经元有着许多种类和功能但人工神经网络没有，并且人工神经网络还会使用不会在人脑中出现的反向传播来进行训练。目前人工智能在处理大量数据、自动化分类、模拟情感和环境、预测洞察等方面的强大能力对心理学的应用起着极大的作用。但随着人工智能在智能层面的进步，关于人工智能作为工具的伦理问题也备受关注^[22]，同时应用层面，对于人工智能判断准确度也需要接受一定的考验^[35]，关于人工智能在心理学应用中的情绪和病症识别方面是否会侵犯他人隐私的问题也需要我们进行协调和深入思考。在人工领域的对齐领域，未来的心理学研究也应该反向应用于生成式人工智能进而提高人机对齐程度，进而使得人工智能更高效更准确更安全地服务于人。

4.2.1 生成式人工智能作为最根本的生产力工具

生成式人工智能基于本分类在心理学领域的应用包括：基于人工智能算法的分类和心智测量等。生成式人工智能的强理解能力为研究者通过文字，图像判断其背后发布者的心理状态提供了可能性。

(1) 心理疾病的判断：早期有 Moin Nadeem 等人构建了 2.5M 条推文语料库并通过文本分类的方式对抑郁症进行判断，实验结果表明其准确率高达 86%^[20]，后有 Zhang, YP 等人通过使用融合分类器将深度学习模型分数、心理文本特征、用户信息结合后研究其与抑郁信号的关系，他们的融合模型在 446 人的测试集上有 78.9% 的正确率^[24]。使用生成式人工智能判断自杀倾向，压力，舆论等同样可行。

(2) 基于语音特征进行对语言情感的判断同样可行，Starlet Ben Alex 等人提出了一种基于深度学习的系统 AER（自动情绪识别系统），这是一种通过对话语水平和音节水平提取韵律来进行情感判断的可行系统^[25]。

(3) 受益于生成式人工智能及深度学习网络，对图像和视频中的面部表情进行识别并判断情感也变得可行 Qiuyu Li 等人通过使用用于检测面部特征的深度多任务卷积网络和用于估计微表情光流特征的融合深度卷积网络的识别算法并利用支持向量机对精细后光流特征进行分类，实现了对微表情的自动识别^[26]。

(4) 生成式人工智能的相关算法也在心理学领域的分类应用方面起到了大的作用，例如通过决策树来理解不同心理学变量之间的关系，进而实现根据特征预测个体的心理状态或疾病评估^[27]。利用支持向量机来提高将个体分为不同心理状态或疾病类别的准确性^{[25][28]}，利用机器学习算法来优化对自杀现象的分析^[29]，利用神经网络进行图像识别、情感分析、语义理解等任务，进而对大规模的心理数据进行分类。

4.2.2 生成式人工智能作为助手协助人类进行生产

生成式人工智能基于本分类在心理学领域的应用包括：基

于心理数据的健康干预系统、认知辅助、心理治疗等

(1) 基于深度学习的神经网络结构，GAI 能够逐层地学习抽象的特征表示，在心理学中的临床方面可以对生理信号或医学影像图片进行理解处理，Xiao-Wei Wang 等人实现了使用机器学习方法从脑电图数据中对情绪状态进行分类的研究^[30]，通过分析对应的情绪状态可以更好为心理学临床提供帮助。

(2) 利用生成式人工智能的强大生成能力和其中的算法可以结合其他心理数据的分析结果结合个体言行举止提供合适的支持、决策和建议^[31]，为心理医生提供参考文献。

(3) 通过利用 AI 聊天机器人使心理健康服务变得更加触手可及。Johanna Habicht 等人开发的基于多个机器学习模型进行预测诊断的 AI 个性化自动聊天机器人 Limbic Access^[32]，该聊天机器人的提出不仅提高了现有的心理健康服务的可接触性和包容性，还为未来人工智能作用于心理学提供了实例，但该研究仅涵盖了英格兰地区的 129400 人的观察数据，若将该聊天机器人直接应用于其他地区可能会产生具有偏见的输出，后续针对如何设计更泛用的服务型聊天机器人还需要我们进一步对数据集广泛性和算法的泛化能力层面进行优化，进而实现普适性更强的聊天机器人。

4.2.3 生成式人工智能模拟特定个体、群体进行对话或模拟环境进行实验

生成式人工智能基于本分类在心理学领域的应用包括：认知过程模拟、情感模拟、行为模拟、群体动力学模拟、虚拟现实实验

(1) 情感模拟：构建情感模型使得模型拥有一定的情感理解能力，推进人机关系的进步并提供依照个性化情感进行反馈的服务，Yujie Li 等人设计的 AI-EmoCom 将情感视作网络通信的一种媒介，并将该系统应用于无人驾驶和情感社交机器人领域并取得了显著的成果^[34]。

(2) 行为模拟：利用算法模拟人类行为，研究人类在各种情境下的决策过程和行为模式，Joon Sung Park 等人设计了可信的人类行为代理，并将这些人类行为的代理置于虚拟空间和社区中进而探究相关的心理学现象和社会学现象，未来可能将其应用于培训人们处理复杂的人际关系情境，测试社会科学理论等^[33]。通过利用生成式人工智能还可以进行利用人工智能模拟对话进而研究不同群体之间的影响关系^[35]。

(3) 群体动力学模拟：作为工具在群体层面模拟信息传递、个体互动，研究群体行为和社会动态^[36]。生成式人工智能为研究人员提供了探究在不同的群体大小情况下个体或群体观点或行为的变化，通过观察这些模拟实验中发生的变化可以了解影响观点变化的因素和观点变化的过程。

(5) 虚拟现实实验：利用生成式人工智能模拟特定情境，观察参与者在模拟环境中的行为和心理，目前已经存在基于 VR 技术的心理健康干预系统，未来应该对如何提高模拟环境的真实性以及对个体是否符合的角度展开研究。

4.2.4 生成式人工智能在心理学中可能带来的风险和问题

(1) 隐私安全问题：数据集来训练模型进而对心理学疾病进行判断或预测可能会涉及到侵犯他人隐私的问题，后续的研究者应尽力保证应用对象的隐私，也可以通过隐私保护学习来保障应用对象的隐私安全，避免个体信息被泄露。

(2) 模型的输出问题：由于训练数据可能存在偏差，生成式人工智能的输出在复杂的心理学领域存在误差和不确定性，甚至输出包含刻板印象、幻觉^[37]、歧视或偏见^[38]的观点，过度依赖 GAI 的输出结果可能会导致不准确的研究结果甚至损害研究的客观性和公平性，这些问题不仅仅体现在心理学的研究上，还体现在教育学上^[39]，未来的研究必须朝着如何识别并减轻以上输出问题的方向展开。

(3) 伦理问题：当使用生成式人工智能进行上述的模拟实验时不可避免的会出现伦理隐私问题，模拟特定个体的认知、情感和行为的实验可能会触及到特定个体的隐私伦理问题。上述的模拟过程应该在合格的管控下进行确保虚拟的对象并不会受到影响，并且需要独立的伦理审查机制确保研究和应用均符合道德标准。

(4) 法律责任：当心理学研究者或医疗人员将生成式人工智能作为绝对可信的参考时可能会出现知识上的偏差，可能会引起研究的学术性问题或医疗治疗上的问题，后续应完善对此类问题甚至案件的审判流程并加强学术类和医疗类的监管。

4.3 生成式人工智能与新闻传播学

生成式人工智能的出现对新闻传播学产生了极大的影响，随着 ChatGPT 等生成式人工智能的推出，新闻工作者的工作效率大大提高，GAI 对于信息数据的采集、整理、分析的强大能力简化了新闻专业工作者以及研究人员的工作流程，使得他们可以更专注于高质量内容的设计和细节方面的处理^{[40][41][45]}。得益于 GAI 训练的庞大数据量，GAI 也可以作为信息传播的媒介和用户进行网络访问的窗口通过人机交互的方式提高信息的传播效率，提高用户的宏观知觉。GAI 的出现为新闻传播学带来了极大的便利，不仅提高了相关工作者的工作效率，还能够实现信息传播的多模态转化^[42]，GAI 的出现也使得人与人的传播关系开始向人于人工智能之间的传播关系^[49]。但 GAI 在新闻传播学的应用也带来了问题和风险，由于 GAI 的特点，输出内容不可避免的产生虽然具体内容不同但实际输出观点完全相同的情况，如何避免由于 GAI 输出模式产生的信息齐一化，如何避免未来信息传播多样性的衰减，我们应在保证模型输出的安全性的前提下提高模型输出的多样性，进而避免由于 GAI 本身输出习惯而导致的信息传播齐一化。GAI 的出现对传统以实践性，真实性为根本新闻传播的观念发起了挑战，GAI 本身的输出有可能会产生包含偏见、歧视、毒性、虚假的信息^{[44][46][50][55]}，在信息的多模态转化中还可能还会产生信息失真的问题，Newsguard 在 23 年底利用 AI Tracking Center 标记发现了 614 个由 AI 生成的虚假新闻信息网站，这些网站扩散着有关政治、宗教、技术等方面的

虚假信息^[43]。接下来我们会对 GAI 在新闻传播学上的应用展开叙述，并对 GAI 在应用过程中产生的风险和问题进行叙述。

4.3.1 生成式人工智能作为最根本的生产力工具

生成式人工智能基于本分类在新闻传播学领域的应用包括：基于 GAI 对数据的加工、基于 GAI 的新闻制作流程、基于 GAI 实现的个性化内容推送生成、基于 GAI 的新闻视觉化、多模态转换。

(1) 基于 GAI 对数据的加工：基于 GAI 对自然语言的强大理理解析能力，我们可以利用 GAI 进行最基础的数据采集、整理、分析的任务，进而简化新闻传播制作层面的工作流程^{[40][41]}，使得新闻传播相关从业者可以将更多的精力用于制作更有深度的报道^[45]。

(2) 基于 GAI 的新闻制作流程：通过利用 GAI 可以学习上下文的特性，我们可以利用 NLG 技术，将利用 NLP 分析的结果变成文字来实现基于事实的文本输出。GAI 在学习上下文的过程中还可以进行意图推断并可以进行一定程度的创造，也就意味着我们可以利用 GAI 生成符合主题的新闻报道或信息，受益于 GAI 的强大能力，越来越多未经专业训练的人也可以产出高质量的新闻报道，这一影响不仅意味着新闻多样性的增加，也意味着传播权利的下沉。我们期待着未来的传播表达，内容创新中会有着更加平等的权利分配。

(3) 基于 GAI 实现的个性化内容推送生成：GAI 的算法可以对用户的偏好进行定义，进而实现更能满足个性化需求的内容推送，在信息传播的过程中，也可以提高对特定偏好群体的信息传播效率，通过不断学习用户的偏好来实现更精准，更符合用户喜好的信息推送。

(4) 基于 GAI 的新闻视觉化、多模态转换：GAI 的强大性能可以使得传播的媒介变得更加多元化，例如 DALL·E3 可以生成符合文本内容的图片，OpenAI 的文生视频模型 Sora 以及特斯拉旗下用于增强 FSD 的文生视频模型都已经具备了生成极其真实并几乎符合物理规律的视频^[60]，GAI 可以将单一文本转化为图像、音频、视频等模态实现自动化的多模态转换^{[58][59]}，让信息的传播变得更有听觉、视觉冲击，提高信息的传播效率，降低信息传播的成本^[42]。如何提高转化后与原文本的一致性可能会成为未来科研工作的焦点。

4.3.2 生成式人工智能作为助手协助人类进行生产

生成式人工智能基于本分类在新闻传播学领域的应用包括：人机交互、网络集成入口、识别虚假信息。

(1) 人机交互：个体可以通过与 GAI 进行交互来提高自身的知识水平，提高工作效率。在新闻学方面，GAI 已经改变了传统的新闻生产流程^[45]，GAI 替代了原本由人进行的数据分析，内容提炼甚至是写作的流程。GAI 作为拥有海量知识的中介已经开始改变传播关系^[49]，它会以对话的形式持续的为对话个体提供超出个体认知范围的内容，并协助个体进行生产工作，未来

的操作能力和硬知识教育的重要性可能会随着 GAI 的发展而降低,如何提高人机交互的效率、如何提高人机之间的价值观对齐程度会成为未来的一大挑战。

(2) 网络集成入口: 现有的 ChatGPT 已经可以提供信息检索、创意生成、情感对话等服务于一体的支持, 例如在 Chatgpt, Llama 等 GAI 中进行临床决策的搜索, 这种方式的检索结果比传统搜索引擎更清晰, 更有效果^[54]。在界面设计上, 不同于其他网站复杂的访问界面, ChatGPT 的访问界面十分简单, 用户按照需求进行检索时不需要访问多个不同的网站, Microsoft 于 2023 年也推出一种与 Microsoft bing 搜索引擎相集成的 AI 聊天机器人 bing chat 收获了许多积极的评价^{[52][53]}, 如何创建一个集成更多功能以至于覆盖用户生活需求的人工智能可能会成为服务人工智能的一大难题, 未来生成式人工智能会在突破临界点后变为迄今为止最大的访问媒介。

(3) 识别虚假信息: 随着生成式人工智能的快速发展, 利用 GAI 生成虚假评论进而实现对特定目的的诱导^[56]、生成虚假新闻扩散特定信息的事件也层出不穷, GAI 可以通过特定的强化训练来增强其对虚假信息的识别能力, 可以通过对语言模式、情绪、用户反馈等方面的分析实现对虚假信息的识别和判断。

4.3.3 生成式人工智能模拟特定个体、群体进行对话或模拟环境进行实验

生成式人工智能基于本分类在新闻传播学领域的应用包括: 利用 GAI 模拟场景或对话进而了解传播过程、利用 GAI 模拟特定内容实现新型传播形式。

(1) 利用 GAI 模拟场景或对话进而了解传播过程: 利用生成式人工智能模型, 可以模拟特定场景和平台上的用户对特定事件或话题的对话^[33], 这有助于了解舆论动向、情感倾向以及信息传播的模式, 可以构建传播模型, 进而预测在不同条件下的传播效果, 为公关、营销和政府的传播决策提供参考。

4.3.4 生成式人工智能在新闻传播学中可能带来的问题和风险

(1) 输出问题: 受数据集、算法、监督规则等的影响, GAI 输出的内容可能会包含偏见、歧视、毒性、虚假的信息^{[46][50][55]}, 当问题过于繁杂或有轻微扰动时 GAI 的输出也会受到影响, 将 GAI 的输出内容直接进行传播可能会产生恶劣的影响, 针对输出问题, 我们应从优化数据集质量, 将数据集预先进行筛查并尽可能的避免由于案例数量带来的偏见、或进行外挂额外的联网检索 (FLARE)^[51], 设计合理的通用监督规则、考虑使用对抗性训练提升模型的鲁棒性并且在 GAI 产生输出时保证有据可循。

(2) 信息失真: 在进行信息的多模态转换时, 信息会发生损耗, 进而影响人们对复杂传播信息的理解, 例如对复杂新闻信息生成封面时, GAI 可能无法输出包含合适内容的封面发生信

息失真。在民族文化传播的模式转化过程中会产生类似的问题, 例如在测试多模态转化的过程中, 研究者发现将文字转化为图像时会大规模的放大人口的刻板印象。如何提升多模态转化的精度和契合度会成为未来针对信息失真方面的研究的方向。

(3) 传播影响: 生成式人工智能如今已经有了极大的影响力, GAI 的表达会潜移默化地影响社会认知, 在用户进行访问时持续的对话会体现 GAI 基于数据集的认知, 并在对话过程中持续影响用户的认知。这种影响的效果远超以外的传播效果, 我们必须重视生成式人工智能潜移默化中对我们认知的影响效果, 并考虑随着 GAI 作用于我们的时间越来越长其影响效果会发生怎样的变化, 为了控制人工智能带来的危害, 欧洲《数字服务法》草案中已经提出应在大型在线平台对高风险的人工智能施行监管^[48]。

(4) 虚假内容的生成: 随着 GAI 的快速发展, GAI 的强大生成能力被用于虚假信息、新闻的生成, 23 年底 Newsguard 发现自 23 年 4 月以来 AI 生成的虚假新闻网站数量已经增加到了 1100% 以上, 从 49 个增加到 614 个, 这些网站通过广告投放获利^[57], 由于这些 AI 经营网站的构建以及内容的发布完全没有人为的监督和审查, 这些网站中充斥中包含政治、宗教、死亡信息等敏感的虚假信息^[43]。美国著名科技网站 CNET 也被爆料在一个月内刊出了 77 篇由人工智能生成的文章, 这些文章中还出现了一些事实性错误。如何自动化的识别这些虚假网站、虚假信息会成为虚假信息识别的一大难题, 这些仅由 AI 制造的网站推出效率是非常快的, 自动化识别监管虚假信息将成为未来可能的研究方向。

(5) 对新闻传播观念的挑战: GAI 的出现重塑了新闻的生产流程, 大部分流程如数据分析、提炼主要内容、寻找创新点等都可以由 GAI 协助完成, 当注重对现实进行人工分析的新闻学的研究流程发生变化时, 这些变化是否会对新闻学专业的基础产生影响? 是否会对专业从业者的观念产生影响? GAI 的出现不仅影响了学科的基础和观念^[47], 未来对优秀记者的定位可能也会发生变化, 将会由数据的采集者和阐释者转变为数据的管理者和分析者, 新闻传播学未来可能会需要更多具有强大的数据处理分析能力的人才。

4.4 生成式人工智能与经济学

生成式人工智能的出现使得经济学家进行微观任务的自动化变得可能, 经济学家们可以通过利用生成式人工智能对文本、背景的理解能力以及其强大的数据分析、模型构建、自动编程能力对他们的工作流程进行优化, 生成式人工智能的出现使得自动化智能时代的出现提供了可能^[61]实现生产率的提升^{[62][73]}。生成式人工智能的出现对劳动力市场也产生了极大的冲击^[63], 除了一些户外体力劳动的工作目前无法被人工智能进行自动化取代, 大部分基础工作都面临着一定程度被 GAI 自动化取代的风险, 并且研究发现高绩效的工作人员受到 GAI 建议的影响高于低绩效的工作人员^[64]。通过对美国职业工作的调查, 一些学者发现行法律、证券等方向的职业有着很高的被 AI 自动化取代的风险^[65], 这种风险的出现提高了劳动生产率的潜力, 同时被

AI 解放的生产力部分可以通入新兴或更需要创造力的生产活动中，未来可能并非是 AI 取代人类的工作，而是会演变为 AI 对人类的工作进行辅助和补充，这一人机关系模式可以大大的提高全球的生产力，AI 的强大潜力可能在最终达到使得全球的年 GDP 增长 7% 的增长结果。尽管生成式人工智能可以在经济学领域带来如此之大的影响，我们仍不能忽视其在输出上的幻觉、毒性、偏见问题，在与用户交互中的隐私安全问题。

4.4.1 生成式人工智能作为最根本的生产力工具

(1) 利用 GAI 的信息处理能力：基于生成式人工智能强大的自然语言处理以及机器学习、深度学习等技术，GAI 可以利用 NLP 技术处理和理解大量经济学文本数据，利用机器学习算法对大量的经济数据进行分析、建模，还可以利用回归分析、分类算法等技术来进行市场分析。

(2) 利用 GAI 进行异常检测：利用 GAI 自动化数据分析处理可以实现对数据流的动态侦测，当数据分析结果出现低效或异常时，不同于耗时耗力的人工检测，GAI 可以第一时间发出警报^[69]。利用 GAI 进行异常检测不仅仅节约了人力方面的资源，还最大限度减少了由于异常未被发现而产生的负面影响。

(3) 利用 GAI 来实现针对用户的个性化：利用 GAI 学习用户行为，分析其偏好数据，将用户根据特点进行分类，提供更具个性化的服务以及产品推送^[71]，进而提高客户的满意度^[72]，更符合客户的产品推送和服务不仅减少广告资源的浪费，也会提高商品的销量、企业的口碑。

4.4.2 生成式人工智能作为助手协助人类进行生产

(1) 进行经济学层面的预测：利用机器学习算法分析历史数据（包括市场数据，媒体信息等）预测不同经济政策变化对市场的走势的影响^[74]，为政策制定者提供模拟实验平台，为消费者、投资者、经济学家提供基于数据的建议和决策方面的辅助。

(2) 利用 GAI 提升工作效率：根据 3.5.1.1 中的利用 GAI 进行信息处理、利用 GAI 进行异常检测，将 GAI 用于这些基础的工作可以提高工作的效率和质量，例如使用 Chat-bot 来辅助客服等涉及语言交涉的职业来提高工作效率。GAI 的出现实现了将工作流程进行拆解和优化，使得人可以将更多的注意力放在更需要创造力的工作上，进而避免了注意力被基础任务所消耗而带来低质工作结果。

4.4.3 生成式人工智能模拟特定个体、群体进行对话或模拟环境进行实验

(1) 模拟经济场景：根据不同的偏好和场景来进行调整模拟的内容，实现更廉价、更高效的进行实验，LLM 的出现使得经济学家可以便利地建立很多经济模型，这些模型并不是为了

完全映射到现实，而是为了帮助我们更好的思考，LLM 的建模能力为未来的经济学研究提供了丰富的思路和方法^[66]。

(2) 模拟经济学家：基于经济模拟环境，模拟经济学家进行决策。Korinek 等人提出了一种两级的深度强化学习方法来学习动态税收政策，模拟的角色可以对模拟环境进行学习和适应，最后模拟的税收结果与经济理论相吻合，并且模拟的税收政策在人类参与者上进行实验也有着不错的效果^[68]。这种模拟决策的结果为人类社会经济政策的制定提供了学习和参考的内容，未来可以通过对经济数据的完善来提高模拟经济行为的准确度并扩大经济模拟的范围。

4.4.4 生成式人工智能在经济学可能带来的问题和风险

(1) 对劳动力市场的冲击：生成式人工智能的出现对劳动力市场产生了极大的影响^[67]，研究发现约 80% 的美国劳动力可能会受到生成式人工智能发展的影响，约有 19% 的工人的至少 50% 工作内容会收到生成式人工智能的影响，高收入工作可能会受到更大的影响^[70]，未来的生成式人工智能巍峨劳动力市场带来的压力可能会提高劳动力的平均水平，GAI 不会完全的取代人的劳动，如何构筑更好的人机关系将成为未来的一大课题。

(2) 解释性和透明度：在利用 GAI 进行政策模拟和进行经济学理论的模拟实验中，理解 GAI 输出的背后原理不仅有助于评估政策和理论模拟的准确性和可靠性，也有助于建立人对 GAI 的信任，并且为监管机构提供了模型的评估路径，使得模型的可靠性和公正性变得更透明、公开。

(3) 隐私安全问题：更多的数据意味着更少的隐私^[74]，由于 GAI 的输入-输出模式，GAI 中可能存储着大量的数据和信息，一旦发生数据泄露就可能会导致经济模型泄密、研究成果窃取等问题。

(4) 输出问题：GAI 在经济学中仍然存在输出问题，主要包括输出的准确性、一致性问题 GAI 在数据处理分析的过程中无法保证完全正确的输出，这些问题涉及到训练数据、算法、模型、过度泛化的问题，为了缓解 GAI 的输出问题，未来的模型训练应采用更高质量和多样化的训练数据，优化调整算法和模型，定期的对模型进行评估和测试，并且在用户层面优化人机交互的模式。

(5) 伦理问题：GAI 的不稳定输出可能会产生一些伦理问题，例如 Air Canada 使用的 Chat-bot 在回答客户问题时产生了错误，输出了违背航空公司政策的问题解决方案，最终导致了涉及 GAI 责任的纠纷^[75]。这是一个涉及技术、责任、透明度的伦理问题，不仅是针对 Air Canada，在商用 GAI 出现问题时，如何应对 GAI 产生错误的责任，我们未来需要更细致化的 GAI 责任分配规定，未来的商用 GAI 应在使用前验证其可靠性，避免在使用过程中产生一些包含错误、歧视的信息。

4.5 生成式人工智能与管理学

在本部分中我们着重讨论管理学中的商业管理学，在商业

管理学中的应用可以更好的体现 GAI 的强大能力, GAI 在商业管理学中的应用同样可以类比到其他管理学的方面中(例如:公共管理、教育管理、供应链管理、商业管理、信息系统管理等)。

生成式人工智能在商业管理中的广泛应用已经改变了许多领域,包括但不限于客户服务、数据分析、市场营销等内容。我们可以通过利用生成式人工智能来实现对供应链管理和物流中一些基础任务的自动化,使得员工可以专注于战略规划和创新方面的内容,进而提高工作效率和行业竞争力^{[76][77]}。GAI 的出现改变了商业中企业与用户之间的关系^[78], GAI 的出现使得针对用户的服务更具个性化,并且 GAI 可以对用户进行基于不同标准的分类,使得针对分类商品的推送效果更好, GAI 也可以及时反馈与用户交互中产生的问题以及用户的反馈,有研究表明引入 GAI 的服务系统可以提高用户的满意度^{[71][79]}。生成式人工智能的出现简化了管理学方面文档的编写以及数据的分析统计,并且可以在大量的数据分析结果之下可以辅助一定决策的进行。GAI 的出现同样简化了人力资源管理、项目管理、知识管理、公共管理、教育管理等管理学领域中的流程。这些应用方面大大提高管理效率和效果。但有利就一定有弊,我们无法忽视 GAI 在管理学应用中带来的数据隐私、偏见以及应用过程是否合乎伦理规范等问题。

4.5.1 生成式人工智能作为最根本的生产力工具

(1) 与用户进行实时的交互:生成式人工智能的出现使得企业可以实时的与用户进行交互,在以往的交互模式中,实现大规模的实时交互是一件很耗费人力资源的事情,但 GAI 的出现实现了一定程度的生产力解放,并且可以在第一时间对用户产生合理的反馈^[80],及时的反馈可以提高用户的满意度^[81],并且基于 GAI 强大的数据分析理解能力, GAI 可以实现针对用户的个性化服务设计,结合 GAI 高度流畅的文字输出,可以大大提高对客户的服务效果^[82]。未来可以将更高精度的共情模块引入到与用户进行交互的模型中,共情能力的提高可以提高客户的粘性以及客户的忠实度。

(2) 基于 GAI 的强大数据分析能力和在管理学方面的理解能力^[83], GAI 可以用于管理学报告、决策文件、政策文件等管理文档的编写,我们也可以使用 GAI 处理生成一些需要由大量数据总结而成的报告, GAI 可以同时处理各种复杂的数据并可以从数据中分析出市场趋势。

4.5.2 生成式人工智能作为助手协助人类进行生产

(1) 协助基础的内容管理: GAI 可以协助进行项目管理、知识管理、公共管理、教育管理等内容。项目管理方面: GAI 可以协助项目进行针对时间、任务的管理分配,还可以实时跟踪项目的进度,推进项目的效率。知识管理方面: GAI 可以作为一个由数据集整合而成的数据库,不仅可以实现大量信息的存储,还可以实现高效的信息检索,将模型接入团队即可以提高团队

协作的效率并提高团队成员的知识水平。

(2) 协助进行人力资源管理(HRM): GAI 的出现简化了传统的招聘流程, Jinbo Zhou 等人阐述了一种基于 Chatgpt 的数字化人力资源管理平台,该平台可以使用 GAI 来针对需求岗位生成精确的职位描述并进行自动筛选,这个平台还可以通过利用 Chatgpt 来实时与求职者进行对话^[84]。GAI 可以通过自动筛选能力对应聘人的建立进行第一阶段的评估和筛选, GAI 对 HRM 的改变还体现在包括员工关系,绩效管理等方面^[85]。

4.5.3 生成式人工智能模拟特定个体、群体进行对话或模拟环境进行实验

(1) 通过模拟实现策略的制定:生成式人工智能可以通过模拟不同的业务场景和结果来辅助管理层进行评估潜在的风险和可能的方案,协助管理层进行管理策略的制定,进而实现对人力、财力、物力等因素的高效利用最终实现生产力的提升^[68]。

4.5.4 生成式人工智能在管理中可能带来的问题和风险

(1) 数据隐私问题: GAI 的训练需要大量的数据,这些数据包括知识、用户数据、员工数据等内容,内容中又可能包括一些敏感信息,这些内容的获取过程不一定完全合规,数据的爬取过程涉及到数据隐私的侵犯问题,这些问题会引发用户和员工对数据隐私的担忧,未来应从提高模型训练的透明性和提高对数据获取的监管方面来解决侵犯数据隐私的问题。

(2) 缺乏管理学的细节知识:管理学涉及到许多领域,不同的领域有着不同专业知识内容,即使 GAI 有着强大的理解能力可以在 zero-shot 的情况下实现接近的文本输出,但其输出的内容可能难以为用户提供准确的信息,在需要专业知识的任务环境中 GAI 可能会输出错误,这些错误可能源于专业知识的匮乏、难以处理复杂的情景、无法完全实现与人类价值观对齐等问题。

(3) 针对用户交互产生的问题: GAI 的输出具有不确定性,即使 GAI 的引入大大提高了企业与用户之间的交互效率,但 GAI 的输出依旧可能会含有毒性、偏见、幻觉、伦理等方面的问题,当出现这些问题时我们该如何判责,由于输出问题产生的经济损失和名誉损害该如何认定,这些问题都需要我们进一步去思考。我们也在呼吁企业应该构建更负责任的人工智能,这不仅是保护用户的权益,也是保护企业的名誉和利益。

4.6 生成式人工智能与政治学

生成式人工智能一经问世即对政治学领域产生了深远影响。其强大的文本分析能力使政府公务人员能够显著提高工作效率,助力各政府部门的工作;而模型的分类能力使生成式人工智能具备推测人类政治态度的潜力,从而提升政治信息传播的精确性和效率。在政治学研究中,生成式人工智能也具有巨大潜力。

研究人员可以利用语言模型对文本进行分析，从而检测可能引发政治问题的网络仇恨言论，预防许多潜在政治问题的发生；此外，生成式人工智能还能模拟不同政治立场的个体进行问答，帮助研究者了解不同政治立场下个体可能的思维模式和行为模式。然而，我们也必须审慎对待可能出现的风险和挑战。与其他领域相比，生成式人工智能在政治学领域中的偏见问题可能带来更为严重的影响，未来研究者必须正视这一重要议题，使生成式人工智能更广泛更包容更民主地应用于政治学领域。

4.6.1 生成式人工智能作为最根本的生产力工具

生成式人工智能基于本分类在政治学领域的应用包括：仇恨言论检测与解释，政治文本分析和生成，政治广告微定位等。

(1) 仇恨言论检测与解释：网络仇恨言论是政治问题的重要原因之一，即使对于人类来说，仇恨言论的分类与解释也因为较低的一致性而成为一项困难的任务。Fan Huang 等人研究了 ChatGPT 在对仇恨言论的检测和分类以及为隐性仇恨言论生成解释两个方面的表现，发现 ChatGPT 正确识别了 80% 的隐含仇恨推文；且 ChatGPT 生成的解释可以增强人类的感知，它们往往被认为比人类编写的解释更清晰^[86]。

(2) 政治文本分析和生成：对政治文本的分析和促进政策透明度，问责制直接相关，正确的分析才能保障公民的知情权。Alejandro 等人利用多种大型语言模型对对公共事务文档进行了主题分类，结果表明各模型均有较好的分类效果，使用生成式人工智能的自然语言处理功能来加强对公共文件的分析是可取的。^[87]而欧盟理事会则关注到了生成式人工智能在识别复杂公共行政文件（如法律合同）时能够高效提取出关键信息，从而减少公务员审查这些文件的时间，使他们能专注于更复杂，更有创新性的工作^[88]。

(3) 政治广告的微观定位：微观定位式的广告能够根据个人性格特征进行有针对性的信息制作和投放。Almog 等人 2023 年的实验证明了大语言模型能够通过网络评论较为精准地检测用户的个性偏好，推测他们的政治态度，并将政治广告文本也进行相应的分类进行精准投放，这显著地提升了政治广告的说服力与用户对信息的接受度^[89]他们在 2024 年的研究中使用 ChatGPT 自动化整个过程，进一步提升了其有效性和可扩展性^[90]。

4.6.2 生成式人工智能作为助手协助人类进行生产

生成式人工智能基于本分类在政治学领域的应用包括：协助政府公务员工作，辅助进行政治文本分析，

(1) Jonathan 等人发现政府公务员每天有大量时间花费在官僚主义事务上（如记录信息和其他基本行政任务），而通过对使用生成式人工智能的英国公务员的采访他们发现，生成式人工智能能够高效协助公务员完成简单的官僚主义工作，将时间分配到更有价值的公共服务上^[91]。

(2) FedSpeak（用于描述技术语言的术语）通常带有模糊

性和不确定性，以避免对社会产生不必要的冲击，是自然语言中难以理解的一类词汇。Anne 等人实证评估了 GPT 模型破译 FedSpeak 的能力，得出了 GPT 模型具有一定对 FedSpeak 的分类能力，能够作为一种非常有价值的工具协助政府评估员的工作^[92]。

4.6.3 生成式人工智能模拟特定个体、群体进行对话或模拟环境进行实验

(1) 模拟不同政治立场的人类：Fabio 等人在研究大语言模型的政治偏见时提出了一种新颖的实证设计，他们要求 ChatGPT 模拟特定政治阵营的代表人物进行问答，并将这些回答与其在默认条件下的答案进行比较，以此判断 ChatGPT 是否具有政治偏见，研究发现 ChatGPT 对美国民主党、巴西卢拉和英国工党表现出重大和系统性的政治偏见的同时，也证明了 ChatGPT 拥有模拟不同政治立场的人类的能力^[93]。

4.6.4 生成式人工智能在政治学中可能带来的问题和风险

(1) 与人工智能接触导致民粹主义：Siegfried 通过严密的论证发现接触人工智能的个体容易对自己的社会地位预期下降并产生焦虑，因此将导向民粹主义与极右政党，并设计实验让三组参与者分别试用 ChatGPT，观看有关 ChatGPT 的信息视频和作为对照组，最后对三组成员分别提问以观测他们在实验后是否更加偏向了民粹主义，结果证明了它的观点^[94]。

(2) 偏见问题：人工智能的偏见问题在政治学领域同样有所体现，这样的偏见由数据与算法两方面共同导致。如 Broockman 等人发现，如果人工智能职位的招聘算法仅根据来自硅谷的招聘数据进行训练，因为激进左派和保守派人士在该领域的代表性往往不足，得到的模型就会对相应人群产生政治偏见^[95]。Li 等人研究在招聘决策中使用的各模型发现，这些算法根据过去候选人的简历（分为“合格候选人”和“不合格候选人”两类）作为训练数据，从中提取“合格候选人”的简历特征，并用于识别新的申请者是否属于“合格候选人”，然而算法将简历中与资格无关的因素也加入了训练（如性别，种族身份和政治倾向等），形成的模型会根据这些无关因素不平等地对待不同的群体^[96]。

5. 生成式人工智能应用带来的风险

与问题

5.1 Problem

5.1.1 输出内容问题

(1) 真实性和可信度的问题：GAI 生成的内容虽然已经可以达到难以区分的程度，虽然这种拟真度的输出在许多领域都有着很大的潜力，但同时也带来了真实性问题，在绘画创作，新闻报道领域已经出现了许多 deepfake 内容，这些内容的泛滥对传统的艺术概念产生了巨大的冲击，也破坏了新闻报道的真实性，损害了公众对媒体的信任。为了解决这些问题，我们应开发更精细的技术来检测 AI 生成的内容并进行标记，帮助公众区分人类创作和 AI 生成的作品，同时也要出台相关的法律内容来打击 deepfake 的生成行为。

(2) 内容包含偏见、歧视、幻觉、毒性、虚假信息等内容的问题，由于用于训练的历史数据不足以公平地代表所有人口，导致模型在输出时可能会输出具有偏见以及歧视的内容，例如亚马逊使用的人工智能招聘工具在训练过程中的训练数据 60% 为男性导致该工具的判断结果具有偏见，亚马逊在一年后便停止使用了该招聘工具。未来研究者应侧重于建立有道德和责任感的人工智能，未来的研究者可以通过提高训练数据的质量作为切入点。

(3) 输出违反伦理道德内容的治理问题：由于数据集、算法、模型特点等问题，GAI 可能会生成一些含有歧视、偏见、暴力血腥、色情等可能对社会价值观和道德标准产生冲击的内容，并且 GAI 可能会在训练的过程中放大偏见，导致其输出的内容有着不平等的针对性，未来应着重于构建在伦理道德方面负责任的人工智能。

5.1.2 技术责任问题

(1) 解释性和透明度问题：在涉及复杂算法或大量数据的处理过程中，我们无从得知 GAI 的决策过程，当决策过程无法得知时，决策的结果就会失去可信性。当发生由于 GAI 决策错误而产生不良后果时，由于不清楚决策的产生过程，责任的归属就会变得十分模糊，例如一个以 GAI 为基础的医疗诊断决策系统给出错误建议后患者接受了不适合的药物治疗，这个问题是应该归咎于使用 GAI 的医生，还是 GAI 的开发者，又或是数据集的数据源呢。提高模型的解释性和透明度不仅可以更好的帮助我们理解 AI 的决策过程，还可以提高 AI 系统的可信度。

5.1.3 对经济和社会的影响

(1) GAI 对传统劳动力市场的冲击：GAI 的出现使得一些重复性的基础工作出现了被取代的风险，这种风险可能会导致某些职业的就业机会减少。GAI 的出现同时重塑了许多行业的工作流程以及一些企业的运作方式，GAI 的出现使得一些工作流程的拆解和自动化变得可能。未来的工作可能需要更强的 AI 工具使用能力，如何构建更好更高效的人机交互模式也是一大长久的课题。

(2) GAI 引起的资源分配不均问题：受限现实中的许多因素，并非所有群体都有平等的权利使用 GAI，这个现象会导致不同群体在使用 GAI 上的机会、权力不平等的现象随着时间的推移而不断加剧。GAI 可能会加剧未来社会的经济差异，拥有 AI 技术的企业或个人可能会在潮流中获得巨大的经济利益，而无相关知识的群体可能会落后。GAI 的出现同样对网络内容的多样性发出了挑战，使用 GAI 可以轻易的生成大量的文本内容，这些文本内容的意识风格十分相似，若未来 GAI 的文本生成量于互联网上的占比持续高速增长，则会淹没许多群体的声音，导致信息的传播权利失衡。

5.2 Governance

5.2.1 GAI 的伦理道德及责任问题

(1) GAI 所面临的伦理和道德问题：由于 GAI 本身并不具有自由意识，所以该如何判断 GAI 所产生的伦理道德问题会成为急需被解决的问题，例如在自动驾驶的情况下由于 GAI 进行的自主决策而引发的事故该如何进行正确的判罚，如何划分 GAI 的权力边界以及如何将道德责任施加于 GAI，这些问题都需要通过提出同时符合 GAI 的身份和社会要求的伦理道德判断标准来解决。

(2) 道德法律责任分配问题：生成式人工智能的内容生成能力可能设计现有的版权法律的争议问题，由于 GAI 的生成能力源于数据集中大量的样例，GAI 的生成实际也只是按照样例将数据集中的内容进行融合后输出，所以 GAI 的输出到底属不属于创新是我们未来需要讨论的问题之一，若无法解决该问题，GAI 将对创造业行业产生不小的打击。

(3) 隐私安全问题：由于 GAI 的训练基于大量的数据，在这些训练的数据中可能包含一些敏感的信息，GAI 的学习模式可能在无意之中实现了侵犯他人隐私的行为，这些被学习的数据有着被泄露和滥用的风险，未来应该增强对模型训练数据的审查，避免由学习数据集引起的隐私泄露，在交互式更新的模型中应使用隐私保护学习来维护使用者的信息隐私安全。

(4) GAI 的安全性问题：由于 GAI 通过对话的工作方式，使得通过 GAI 监视用户变的可能，在美国大选期间，候选人特朗普就利用 GAI 对选举信息以及民意反馈进行了趋势分析。由 GAI 生成的 Deepfake 内容也体现了 GAI 不安全的一面，虽然大

部分 Deepfake 仅应用于色情内容，但 Deepfake 应用于政治或宗教相关情况也屡见不鲜，这种情况带来的负面影响极其严重。通过一些手段干扰 GAI 输出使其产生危险内容也威胁着模型生成的安全性。后续针对 GAI 的安全性，我们应从提高透明度、提高对 Deepfake 的识别能力监管强度、提高模型的鲁棒性、提高通过 GAI 的信息的隐私性等角度进行研究。

（5）GAI 的军事化用：GAI 的强大能力若用于军事化研究将会带来极大的灾难，LAWS（致命性自主武器）作为军事武器中的一种，随着人工智能的发展，LAWS 也成为了军备竞赛中的焦点，为了警告世人人工智能军备竞赛对世界的威胁，马斯克、霍金等数千名 AI、机器人专家在 IJCAI 大会上联名签署了公开信。

5.2.2 GAI 的对齐问题

（1）GAI 价值观的对齐：目前的一些主流模型已经可以从其输出中体现其的价值观倾向，但仍未达到完全对齐的阶段。我们应明确对齐的目标，对齐的方面包括指令遵从、理解人类意图、理解人类偏好、理解价值等角度，但价值观并不是一成不变的，如何实现动态对齐也是未来的一大挑战。对齐程度高的模型可以输出与指令和价值观一致的内容，在评估过程中出现偏差的模型或是出现对齐错位的模型可能会无意中导致一些意料之外的问题，在生成的角度会产生毒性、幻觉、偏见、歧视等内容，对齐后的模型仍然可能会被攻击，为了对抗这些可能的攻击手段需要我们提高对隐私安全保护的意识，提高模型的鲁棒性，并且加强对模型设计者的监管。现有的对齐方法有：强化学习类对齐、插入式对齐、微调式对齐、上下文学习类对齐（添加样例、设计 prompt）等主流对齐方法。

为了提高模型的对齐程度我们应从明确对齐目标、提高对齐方法的效果、提高模型可解释性、加强对模型监管、定期对模型进行偏见和价值观评估等方面进行。

（2）GAI 的价值观对人可能产生的负面影响：由于 GAI 的训练是基于数据集的训练模式，数据集中的意识以及开发者的偏好会在模型的输出中被放大，进而可能会在输出中输出偏激、错误的内容，GAI 的强大影响力会在潜移默化中影响使用者的价值观和文化观念，这种重塑的过程可能会提高使用者的知识水平，也可能被错误的观念所诱导进而形成不正确的价值观，我们应呼吁公众保留自身的决策权，降低对 GAI 的依赖性，不盲目相信 GAI 的建议或输出内容，养成批判性对待 GAI 的习惯。GAI 的出现提升了一部分人的知识上限，但其强大能力也导致完成任务所收获的成就感下降。

6. 总结

在本文中，我们结合了最前沿的生成式人工智能与社会科学交叉领域的研究。我们将生成式人工智能依据其在社会科学中应用层面的身份进行了分类，包括“生成式人工智能作为最

根本的生产力工具”，“生成式人工智能作为助手协助人类进行生产”，“生成式人工智能模拟特定个体、群体进行对话或模拟环境进行实验”，我们的分类方式强调了将生成式人工智能作为工具的人机关系，我们始终认为，人工智能的出现并不是为了取代劳动力，甚至取代人类，它的出现是为了让人成为更好的人。生成式人工智能已经在社会学、心理学、新闻传播学、教育学、经济学、管理学、政治学等多个学科中展现了巨大的应用潜力，其强大的模拟实验能力也为社会科学的研究提供了一种新型的低门槛的高效方法。

生成式人工智能越是被广泛的应用于我们的生活，我们越需要注意其发展所带来的一系列挑战和风险，我们将风险分类为“Problem”和“Governance”两类，这两种分类分别强调了由于模型算法等技术缺陷所带来的问题和由于人工智能在社会或道德伦理层面治理规范不完善所带来的问题，从对隐私安全的问题、偏见歧视等输出的问题、对劳动力市场可能存在的冲击问题、知识不平权的加剧问题、再到道德伦理和责任归属的问题，我们应积极对待处理这些风险。随着生成式人工智能技术的进步，其影响力和应用范围将不断扩大，我们必须确保人工智能与人类在价值观和伦理层面的标准相一致，避免由于对齐偏差所带来的严重后果。

7. 引用

- [1] Achiam, Josh, et al. "Gpt-4 technical report." arxiv preprint arxiv:2303.08774 (2023).
- [2] Volti, Rudi, and Jennifer Croissant. Society and technological change. Waveland Press, 2024.
- [3] Au-Yong-Oliveira M, Gonçalves R, Martins J, et al. The social impact of technology on millennials and consequences for higher education and leadership[J]. Telematics and Informatics, 2018, 35(4): 954-963.
- [4] Hayes, A. (2023, June 15). "Conversing" with Qualitative Data: Enhancing Qualitative Research through Large Language Models (LLMs). <https://doi.org/10.31235/osf.io/yms8p>
- [5] Patrick Y. Wu, Jonathan Nagler, Joshua A. Tucker, and Solomon Messing. Large Language Models Can Be Used to Estimate the Latent Positions of Politicians September 2023.arXiv:2303.12057[cs.CY]
- [6] Edelmann, Achim, Tom Wolff, Danielle Montagne, and Christopher A. Bail. 2020. "Computational Social Science and Sociology." Annual Review of Sociology 46(1): annurev-soc-121919-054621.
- [7] Nelson, L. K., Burk, D., Knudsen, M., & McCall, L. (2021). The Future of Coding: A Comparison of Hand-Coding and Three Types of Computer-Assisted Text Analysis Methods. SociologicalMethods&Research, 50(1),202-237. <https://doi.org/10.1177/0049124118769114>
- [8] Götz, F. M., Maertens, R., Loomba, S., & van der Linden, S. (2023). Let the algorithm speak: How to use neural networks for automatic item generation in psychological scale development. Psychological Methods. Advance online publication. <https://doi.org/10.1037/met0000540>
- [9] Argyle, Lisa P, Christopher A Bail, Ethan C Busby, Joshua R Gubler, Thomas

- Howe, Christopher Rytting, Taylor Sorensen, and David Wingate. 2023. "Leveraging AI for Democratic Discourse: Chat Interventions Can Improve Online Political Conversations at Scale." *Proceedings of the National Academy of Sciences* 120 (41): e2311627120.
- [10] Burrell, J., & Fourcade, M. (2021). The society of algorithms. *Annual Review of Sociology*, 47, 213-237.
- [11] Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30), e2305016120.
- [12] Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2023). Can large language models transform computational social science?. *Computational Linguistics*, 1-53.
- [13] Barbara Ribeiro, Robert Meckin, Andrew Balmer, Philip Shapira, The digitalisation paradox of everyday scientific labour: How mundane knowledge work is amplified and diversified in the biosciences, *Research Policy*, Volume 52, Issue 1, 2023, 104607, ISSN 0048-7333, <https://doi.org/10.1016/j.respol.2022.104607>.
- [14] Meta Fundamental AI Research Diplomacy Team (FAIR)[†] et al., Human-level play in the game of Diplomacy by combining language models with strategic reasoning, *Science* 378, 1067-1074 (2022).
- [15] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 2, 1–22. <https://doi.org/10.1145/3586183.3606763>
- [16] Igor Grossmann, Matthew Feinberg, Dawn C. Parker, Nicholas A. Christakis, Philip E. Tetlock, and William A. Cunningham. AI and the transformation of social science research.
- [17] *Science*, 380(6650):1108–1109, June 2023. doi: 10.1126/science.adi1778.
- [18] Spirling, A. (2023). Why open-source generative AI models are an ethical way forward for science. *Nature*, 616(7957), 413-413.
- [19] Federico Bianchi, Pratyusha Kalluri, Esin Durmus, Faisal Ladhak, Myra Cheng, Debora Nozza, Tatsunori Hashimoto, Dan Jurafsky, James Zou, and Aylin Caliskan. 2023. Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1493–1504. <https://doi.org/10.1145/3593013.3594095>
- [20] Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, Sendhil Mullainathan, Human Decisions and Machine Predictions, *The Quarterly Journal of Economics*, Volume 133, Issue 1, February 2018, Pages 237–293, <https://doi.org/10.1093/qje/qjx032>
- [21] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 610–623. <https://doi.org/10.1145/3442188.3445922>
- [22] Wirtz J, Kunz W H, Hartley N, et al. Corporate digital responsibility in service firms and their ecosystems[J]. *Journal of Service Research*, 2023, 26(2): 173-190.
- [23] Nadeem M. Identifying depression on Twitter[J]. *arXiv preprint arXiv:1607.07384*, 2016.
- [24] Zhang Y, Lyu H, Liu Y, et al. Monitoring depression trends on Twitter during the COVID-19 pandemic: observational study[J]. *JMIR infodemiology*, 2021, 1(1): e26769.
- [25] Alex S B, Babu B P, Mary L. Utterance and syllable level prosodic features for automatic emotion recognition[C]//2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS). IEEE, 2018: 31-35.
- [26] Li Q, Zhan S, Xu L, et al. Facial micro-expression recognition based on the fusion of deep learning and enhanced optical flow[J]. *Multimedia Tools and Applications*, 2019, 78: 29307-29322.
- [27] Handley T E, Hiles S A, Inder K J, et al. Predictors of suicidal ideation in older people: a decision tree analysis[J]. *The American Journal of Geriatric Psychiatry*, 2014, 22(11): 1325-1335.
- [28] Rampisela T V, Rustam Z. Classification of schizophrenia data using support vector machine (SVM)[C]//Journal of Physics: Conference Series. IOP Publishing, 2018, 1108: 012044.
- [29] Franklin J C, Ribeiro J D, Fox K R, et al. Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research[J]. *Psychological bulletin*, 2017, 143(2): 187.
- [30] Wang, Hao-Wei, Dan Nie, and Bao-Liang Lu. "Emotional state classification from EEG data using machine learning approach." *Neurocomputing* 129 (2014): 94-106.
- [31] Dwyer, Dominic B., Peter Falkai, and Nikolaos Koutsouleris. "Machine learning approaches for clinical psychology and psychiatry." *Annual review of clinical psychology* 14 (2018): 91-118.
- [32] Habicht J, Viswanathan S, Carrington B, et al. Closing the accessibility gap to mental health treatment with a personalized self-referral chatbot[J]. *Nature Medicine*, 2024: 1-8.
- [33] Park J S, O'Brien J, Cai C J, et al. Generative agents: Interactive simulacra of human behavior[C]//Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. 2023: 1-22.
- [34] Li Y, Jiang Y, Tian D, et al. AI-enabled emotion communication[J]. *IEEE Network*, 2019, 33(6): 15-21.
- [35] Dwivedi Y K, Kshetri N, Hughes L, et al. "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy[J]. *International Journal of Information Management*, 2023, 71: 102642.
- [36] Hoey J, Schröder T, Morgan J, et al. Artificial intelligence and social simulation: Studying group dynamics on a massive scale[J]. *Small Group Research*, 2018, 49(6): 647-683.
- [37] Ji Z, Lee N, Frieske R, et al. Survey of hallucination in natural language generation[J]. *ACM Computing Surveys*, 2023, 55(12): 1-38.
- [38] Akter S, McCarthy G, Sajib S, et al. Algorithmic bias in data-driven innovation in the age of AI[J]. *International Journal of Information Management*, 2021, 60: 102387.

- [39] Perkins M. Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond[J]. *Journal of university teaching & learning practice*, 2023, 20(2): 07.
- [40] Pavlik J V. Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education[J]. *Journalism & mass communication educator*, 2023, 78(1): 84-93.
- [41] Petridis S, Diakopoulos N, Crowston K, et al. Anglekindling: Supporting journalistic angle ideation with large language models[C]//*Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023: 1-16.
- [42] Liu V, Qiao H, Chilton L. Opal: Multimodal image generation for news illustration[C]//*Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 2022: 1-17.
- [43] Tracking AI-enabled Misinformation: 739 ‘Unreliable AI-Generated News’ Websites (and Counting), Plus the Top False Narratives Generated by Artificial Intelligence Tools <https://www.newsguardtech.com/special-reports/ai-tracking-center/>
- [44] Goldstein J A, Sastry G, Musser M, et al. Generative language models and automated influence operations: Emerging threats and potential mitigations[J]. *arXiv preprint arXiv:2301.04246*, 2023.
- [45] Manjoo F. ChatGPT Is Already Changing How I Do My Job[J]. *International New York Times*, 2023: NA-NA.
- [46] Bianchi F, Kalluri P, Durmus E, et al. Easily accessible text-to-image generation amplifies demographic stereotypes at large scale[C]//*Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 2023: 1493-1504.
- [47] Peña Fernández S, Meso Ayerdi K, Larrondo Ureta A, et al. Without journalists, there is no journalism: the social dimension of generative artificial intelligence in the media[J]. 2023.
- [48] Helberger N, Diakopoulos N. The European AI act and how it matters for research into AI in media and journalism[J]. *Digital Journalism*, 2023, 11(9): 1751-1760.
- [49] Guzman A L, Lewis S C. Artificial intelligence and communication: A human-machine communication research agenda[J]. *New media & society*, 2020, 22(1): 70-86.
- [50] Rawte V, Sheth A, Das A. A survey of hallucination in large foundation models[J]. *arXiv preprint arXiv:2309.05922*, 2023.
- [51] Jiang Z, Xu F F, Gao L, et al. Active retrieval augmented generation[J]. *arXiv preprint arXiv:2305.06983*, 2023.
- [52] Kelly D, Chen Y, Cornwell S E, et al. Bing Chat: The Future of Search Engines?[J]. *Proceedings of the Association for Information Science and Technology*, 2023, 60(1): 1007-1009.
- [53] Bubaš G, Babić S, Čizmešija A. Usability and User Experience Related Perceptions of University Students Regarding the Use of Bing Chat Search Engine and AI Chatbot: Preliminary Evaluation of Assessment Scales[C]//*2023 IEEE 21st Jubilee International Symposium on Intelligent Systems and Informatics (SISY)*. IEEE, 2023: 000607-000612.
- [54] Sandmann S, Riepenhausen S, Plagwitz L, et al. Systematic analysis of ChatGPT, Google search and Llama 2 for clinical decision support tasks[J]. *Nature Communications*, 2024, 15(1): 2050.
- [55] into Bias A I. Challenging Systematic Prejudices[J]. 2024.
- [56] Mohawesh R, Xu S, Tran S N, et al. Fake reviews detection: A survey[J]. *IEEE Access*, 2021, 9: 65771-65802.
- [57] Rastogi S, Bansal D. A review on fake news detection 3T’s: typology, time of detection, taxonomies[J]. *International Journal of Information Security*, 2023, 22(1): 177-212.
- [58] Wu S, Fei H, Qu L, et al. Next-gpt: Any-to-any multimodal llm[J]. *arXiv preprint arXiv:2309.05519*, 2023.
- [59] Lyu C, Wu M, Wang L, et al. Macaw-llm: Multi-modal language modeling with image, audio, video, and text integration[J]. *arXiv preprint arXiv:2306.09093*, 2023.
- [60] Karaarslan E, Aydın Ö. Generate Impressive Videos with Text Instructions: A Review of OpenAI Sora, Stable Diffusion, Lumiere and Comparable Models[J]. *Authorea Preprints*, 2024.
- [61] von Garrel J, Jahn C. Design framework for the implementation of AI-based (service) business models for small and medium-sized manufacturing enterprises[J]. *Journal of the knowledge economy*, 2023, 14(3): 3551-3569.
- [62] Korinek A. Generative AI for economic research: Use cases and implications for economists[J]. *Journal of Economic Literature*, 2023, 61(4): 1281-1317.
- [63] Eloundou T, Manning S, Mishkin P, et al. Gpts are gpts: An early look at the labor market impact potential of large language models[J]. *arXiv preprint arXiv:2303.10130*, 2023.
- [64] Otis N, Clarke R P, Delecourt S, et al. The Uneven Impact of Generative AI on Entrepreneurial Performance[J]. Available at SSRN 4671369, 2023.
- [65] Felten E, Raj M, Seamans R. How will language modelers like chatgpt affect occupations and industries?[J]. *arXiv preprint arXiv:2303.01157*, 2023.
- [66] Horton J J. Large language models as simulated economic agents: What can we learn from homo silicus?[R]. National Bureau of Economic Research, 2023.
- [67] Furman J, Seamans R. AI and the Economy[J]. *Innovation policy and the economy*, 2019, 19(1): 161-191.
- [68] Zheng S, Trott A, Srinivasa S, et al. The ai economist: Improving equality and productivity with ai-driven tax policies[J]. *arXiv preprint arXiv:2004.13332*, 2020.
- [69] Mughal A A. Artificial Intelligence in Information Security: Exploring the Advantages, Challenges, and Future Directions[J]. *Journal of Artificial Intelligence and Machine Learning in Management*, 2018, 2(1): 22-34.
- [70] Eloundou T, Manning S, Mishkin P, et al. Gpts are gpts: An early look at the labor market impact potential of large language models[J]. *arXiv preprint arXiv:2303.10130*, 2023.
- [71] Lee G H, Lee K J, Jeong B, et al. Developing Personalized Marketing Service Using Generative AI[J]. *IEEE Access*, 2024.
- [72] Huang M H, Rust R T. Engaged to a robot? The role of AI in service[J]. *Journal of Service Research*, 2021, 24(1): 30-41.
- [73] Naimi-Sadigh A, Asgari T, Rabiei M. Digital transformation in the value chain disruption of banking services[J]. *Journal of the Knowledge Economy*, 2022, 13(2): 1212-1242.
- [74] Agrawal A, Gans J, Goldfarb A. Prediction Machines, Updated and Expanded: The Simple Economics of Artificial Intelligence[M]. Harvard

- Business Press, 2022.
- [75] What Air Canada Lost In 'Remarkable' Lying AI Chatbot Case <https://www.forbes.com/sites/marisagarcia/2024/02/19/what-air-canada-lost-in-remarkable-lying-ai-chatbot-case/?sh=1e631e41696f>
- [76] Rane, Nitin. "Role and challenges of ChatGPT and similar generative artificial intelligence in business management." Available at SSRN 4603227 (2023).
- [77] Teubner, Timm, et al. "Welcome to the era of chatgpt et al. the prospects of large language models." *Business & Information Systems Engineering* 65.2 (2023): 95-101.
- [78] Subagja, Agus Dedi, et al. "Improving customer service quality in MSMEs through the use of ChatGPT." *Jurnal Minfo Polgan* 12.1 (2023): 380-386.
- [79] Niu, Ben, and Gustave Florentin Nkoulou Mvondo. "I Am ChatGPT, the ultimate AI Chatbot! Investigating the determinants of users' loyalty and ethical usage concerns of ChatGPT." *Journal of Retailing and Consumer Services* 76 (2024): 103562.
- [80] Paul J, Ueno A, Dennis C. ChatGPT and consumers: Benefits, pitfalls and future research agenda[J]. *International Journal of Consumer Studies*, 2023, 47(4): 1213-1225.
- [81] Koc, Erdogan, et al. "Houston, we have a problem!: The use of ChatGPT in responding to customer complaints." *Technology in Society* 74 (2023): 102333.
- [82] Sudirjo F, Ausat A M A, Rijal S, et al. ChatGPT: Improving Communication Efficiency and Business Management of MSMEs in the Digital Age[J]. *Innovative: Journal Of Social Science Research*, 2023, 3(2): 643-652.
- [83] Burger B, Kanbach D K, Kraus S, et al. On the use of AI-based tools like ChatGPT to support management research[J]. *European Journal of Innovation Management*, 2023, 26(7): 233-241.
- [84] Zhou J, Cen W. Design and Application Research of a Digital Human Resource Management Platform based on ChatGPT[J]. *Journal of Theory and Practice of Social Science*, 2023, 3(7): 49-57.
- [85] Budhwar, Pawan, et al. "Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT." *Human Resource Management Journal* 33.3 (2023): 606-659.
- [86] Fan Huang, Haewoon Kwak, and Jisun An. 2023. Is ChatGPT better than Human Annotators? Potential and Limitations of ChatGPT in Explaining Implicit Hate Speech. In *Companion Proceedings of the ACM Web Conference 2023 (WWW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 294-297. <https://doi.org/10.1145/3543873.3587368>
- [87] Peña, A., Morales, A., Fierrez, J., Serna, I., Ortega-Garcia, J., Puente, I., ... & Cordova, G. (2023, August). Leveraging large language models for topic classification in the domain of public affairs. In *International Conference on Document Analysis and Recognition* (pp. 20-33). Cham: Springer Nature Switzerland.
- [88] Council of European Union, 2023. ChatGPT in the Public Sector—overhyped or overlooked?https://www.consilium.europa.eu/media/63818/art-paper-chatgpt-in-the-public-sector-overhyped-or-overlooked-24-april-2023_ext.pdf Accessed 4 Dec 2023
- [89] Simchon, A., Sutton, A., Edwards, M., & Lewandowsky, S. (2023). Online reading habits can reveal personality traits: towards detecting psychological microtargeting. *PNAS nexus*, 2(6), pgad191.
- [90] Simchon, A., Edwards, M., & Lewandowsky, S. (2024). The persuasive effects of political microtargeting in the age of generative artificial intelligence. *PNAS nexus*, 3(2), pgae035
- [91] JBright, J., Enock, F. E., Esnaashari, S., Francis, J., Hashem, Y., & Morgan, D. (2024). Generative AI is already widespread in the public sector. *arXiv preprint arXiv:2401.01291*.
- [92] JHansen, A. L., & Kazinnik, S. (2023). Can chatgpt decipher fedspeak?. Available at SSRN.
- [93] Motoki, F., Pinho Neto, V., & Rodrigues, V. (2024). More human than human: Measuring ChatGPT political bias. *Public Choice*, 198(1), 3-23.
- [94] Manschein, S. (2023). Artificial Intelligence and Political Behavior-Experience with Chatgpt. Available at SSRN 4567428.
- [95] Broockman, D. E., Ferenstein, G., & Malhotra, N. (2019). Predispositions and the political behavior of American economic elites: Evidence from technology entrepreneurs. *American Journal of Political Science*, 63(1), 212-233.
- [96] Li, D., Raymond, L. R., & Bergman, P. (2020). Hiring as exploration (No. w27736). National Bureau of Economic Research.