《新闻界》ISSN 1007-2438 2023 年第 11 期

DOI:10.15897/j.cnki.cn51-1046/g2.20231123.001

微调大模型：个性化人机信息交互模式分析

官璐　何康　斗维红

摘要　美国人工智能研究公司 OpenAI 近期加速迭代，推出了基于 GPT-4 的“自定义 GPT”，用户可以更加简便地创建出自己的专属微调大模型。国内外大模型团队也陆续涌入微调大模型技术下的个性化智能应用赛道，标志着个性化智能信息交互时代即将开启。本文认为，个性化微调大模型下的人机信息交互模式，将是人与机器双向互动优化信息觅食策略的过程。在对话与交谈中，个体不断精确化自己的信息需求表达，同时大模型通过学习用户输入的提示词，不断深化对用户信息需求的理解，提供更加精准的个性化回应。未来微调大模型还有可能作为“人脑的延伸”，辅助人们应对信息过载情景下的信息处理与信息存储任务。

关键词　微调；大模型；个性化；人机信息交互

中图分类号 G206　**文献标识码** A

作者　官璐，复旦大学新闻学院讲师，复旦大学全球传播全媒体研究院“计算与智能传播”团队成员，上海 200433；何康，宁波大学人文与传媒学院讲师，浙江宁波 315211；斗维红，兰州大学新闻与传播学院讲师，甘肃兰州 730000

一、问题提出

2022 年年底，OpenAI 发布生成式人工智能产品 ChatGPT，其模型表现引发不同观点的热议。基于大算力支持和超大规模语言数据作为训练样本的大模型技术，可以辅助人们完成自动文摘、机器翻译等基础通用任务[1]，但在医疗、法律等专业领域任务中，因其缺乏深度行业知识，无法满足定制化、精细化和行业化的落地需求，表现只能说是差强人意。此外，在人机互动场景中，虽然大模型可以与用户个体在对话界面中完成基本的聊天问答，但因为大模型无法提供情感支持，且在回应中少有个性化、创意化的互动，被认为依然难以替代人际陪伴[2]。

提高大模型在精细化、个性化特定任务上的表现，主要依赖于微调技术，如何优化微调策略、完成行业应用落地是目前各大模型团队的主攻方向。2023 年 8 月 22 日，OpenAI 公司公开了可微调 GPT-3 更新版 API 接口[3]，各行业的开发者都可以在底层大模型基础上，通过上传专属数据完成模型微调，在语气风格、文本结构等方面满足个性化的信息需求。11 月 6 日，OpenAI 推出了基于 GPT-4 的界面版“自定义 GPT”（Custom GPT），普通人不需要编写代码就可以创建自己的微调大模型[4]。此外，Character.AI、通义千问等国内外大模型团队陆续涌入个性化智能应用赛道，个性化智能信息交互时代即将开启，业界学界已将更多关注力聚焦到微调大模型个性化的研发与应用中。

那么，微调大模型的个性化具有哪些关键性的技术特征？基于微调大模型的个性化人机信息交互模式将发生哪些重要变化？微调大模型的个性化还可能带来哪些用户隐私和媒介伦理层面的新挑战？基于以上问题，本文从微调大模型的技术逻辑和个体信息管理视角出发，梳理国内外公开微调大模型的发展与演化脉络，探讨人机交互下的个性化信息获取模式，展望微调大模型对个人信息处理和信息存储模式的改变，并对个性化带来的新挑战展开讨论。

二、微调大模型技术的发展

（一）国外微调大模型

微调是大模型算法的核心技术之一，是预训练模型技术所附有的算法功能。以 ChatGPT 为代表的大语言模型首先在大规模数据中训练一个泛化能力强的预模型，而后根据人机对话的特定场景数据微调预模型，进行内容生成任务[5]。这种特性使得大语言模型可以对外公开其泛化的预模型，其他开发者可以在其基础上微调参数，更好地执行精细化、个性化特定情景任务。

目前已有多家公司机构公开了他们的大语言模型微调 API 或开源代码。Meta 在 2023 年 2 月和 8 月相继发布开源大语言模型 LLaMA 和 LLaMA2，该模型基于 2 万亿的标记文本数据进行训练，因其开源特性，使得其他开发者可以在其基础上微调，训练出专属版本的大语言模型[6]。2023 年 8 月 22 日，OpenAI 公司公开了可微调 GPT-3 的 API 接口，通过付费的 API 接口，开发者可以接触到 GPT-3 的底层模型，并在底层模型基础上微调参数以完成更加细致的具象化的情景任务[7]。11 月 6 日，OpenAI 推出了界面化的“自定义 GPT”，代码门槛被 UI 替代，普通用户可以更加简便地创建自己的大模型。每个人能同时拥有多个专长 GPT，可以是自己创建的，也能从 GPTs 商店中使用别人创建的[8]。普通用户也可以上传个人数据训练微调模型，获得在语气风格、文本结构等方面都更加贴合定制化需求的个性化微调大模型。

专属化微调大模型的公开，预示着大模型即将开启个性化智能应用的新纪元。目前，国内外业界学界都已展开相关研究。剑桥大学学者 Sebastian Porsdam Mann 团队通过私人文档为数据，微调出“个性化大语言模型”（Personalized Large Language Model），以辅助个性化的学术论文写作[9]。他们发现，现有技术足以支持基于个人数据微调出个性化的私人专属大语言模型，微调后的个性化模型可以在形式、文字风格、整体质量、创新观点等方面更加贴合个人需求。区别于通用版大模型针对泛化基础任务，定制版智能服务将聚焦于精细化、个性化定制任务，个人用户将享受到根据个人喜好在知识、性格、情感、记忆等多个维度定制后的智能体交互新体验[10]。

（二）国内微调大模型

面对国际大语言模型的领先优势，中国微调大模型相关技术领域正在奋起直追，势头迅猛。2023 年 8 月百度旗下 AI 大模型“文心一言”宣布向全社会开放，成为国内首个全面开放的大语言模型。其后，百川智能也宣布其大模型通过《生成式人工智能服务管理暂行办法》备案，据悉同批获得审批上线的大语言模型还包括字节跳动云雀大模型、中科院紫东太初大模型等[11]，国内 AI 大模型“百”模大战已打响。

在开源的可微调大模型领域，百川智能发布的 Baichuan 系列中文大模型获得了不断攀升的关注度，逐渐超越 LLaMA2 成为国内外开源大模型领域新晋顶流。据公开报告，Baichuan2 模型基于 2.6 万亿高质量多语言数据训练，数据类别来源广泛[12]。在法律、医疗、数学、代码等领域，Baichuan2 的性能已全面超过国际领先的开源大模型 LLaMA2，并且 Baichuan 系列弥补了中国开源生态的短板，让中国开发者可以用上对中文场景更友好的开源大模型。

在微调大模型的个性化领域，天猫精灵接入阿里大模型“通义千问”的个性化样机，开启了国内微调大模型个性化的实践初探。天猫精灵将智能随身眼镜与千问大模型样机结合在一起，通过骨传导技术的加持建立了一个相对

私密的对话环境，可以满足用户定制化与个性化问答的需求。千问大模型的个性化通过知识增强、工具增强、对话增强，以及人类反馈强化学习四个训练步骤，实现符合人设的人格化表达，增强大模型应用的可玩性和创意性[13]。

三、微调大模型个性化的技术逻辑与方向

个性化，是根据个体需求与喜好量身定做用户体验的艺术，是现代人工智能技术连接人与机器之间差距的重要纽带[14]。在当前新媒体时代，个性化已被应用于多种数字平台，用以增强用户与平台之间的交互和体验。例如，个性化信息推荐系统通过挖掘个体用户的过往行为与偏好，满足每个用户独特的信息需求，使交互体验更加高效和愉悦。

个性化不仅仅指算法过滤下的内容推荐，还涵盖了用户体验的方方面面，包括界面设计、审美取向、互动风格等[15]。随着人工智能新技术的不断涌现，用户行为偏好更加多样化、个性化应用场景变得愈加复杂，人机交互领域不断跟进前沿新技术，以满足用户动态变化的需求和喜好。例如，自适应界面设计技术可以基于用户的行为需求和设备类型支持系统自动调整界面布局、亮度等[16]。随着数字技术继续发展，个性化系统很可能会继续演变，最终创造出融入人类日常生活各个环节的个性化人机交互智能系统，为人类提供全方位的定制化服务体验。

大模型实现个性化的核心在于预训练技术。传统机器学习模型仅依靠场景任务下的特定小规模数据训练模型，准确率和效率不高。而预训练技术基于“迁移学习”的思路，预先在大规模数据中训练一个泛化能力较强的预模型，而后迁移到特定场景的数据环境下进行一个增量训练，对其模型参数权重进行微调[17]。这使得大模型不仅可以较好地完成通用任务，还有潜力通过微调手段，针对特定行业、特定用户数据调节模型表现，实现个性化的内容生成。

阿里巴巴算法专家高星认为，当前大模型的个性化实践主要包括知识、性格、情感、记忆四个维度的研发[18]。与传统的闲聊和信息查询的问答机器人不同，大模型的个性化需要更加拟人化、有性格特点，并且可以提供理解能力和情绪价值。这就要求微调大模型在对话中可以理解用户的情绪和偏好，并且可以准确地回应情感和知识需求。

在知识方面，微调大模型的个性化不仅需要具有泛化知识、实现开放域的知识对话，还需要具有用户特定领域的实时知识储备。大语言模型基于大规模多模态数据训练，具有强大的事实知识检索能力，并且可以将存储的常识知识，通过语义概率运算的形式运用到下游内容生产任务中[19]。

然而当前大模型作为“知识库”的实践中还存在一个难以解决的问题——知识幻觉，即生成没有考证或与事实不一致、胡说八道的内容。这些胡说八道的内容会干扰人类对大模型产生信任，阻碍人机深层互动。知识幻觉可能由多方面因素造成，例如，预训练数据中可能存在虚假信息导致模型学习到错误信息，此外，大模型采用随机性机制概率生成内容，这也可能会导致模型采样到错误的内容上。当前计算机领域主要通过提高数据质量、增强互联网实时检索功能等方法减少知识幻觉[20]。

微调大模型可以在知识维度实现个性化。用户个人的职业身份信息和行业领域知识信息可以被组合成数据集，用于大模型在个性化增量训练中微调参数。此外，大模型也可以仿照个性化信息推荐系统，在用户初次使用前请用户预设勾选喜好话题的分类，便于微调大模型预载丰富且深入的领域数据作为支持。

在性格方面，阿里巴巴达摩院专家认为，大模型的性格实践可以总结为四个维度：稳定准确的人设、鲜明的语言风格、逻辑自洽的三观和有偏好的对话风格[21]。他们设计了系列实验探讨不同人格特质对对话体验和人格判别的影响，发现闲聊情景比知识问答情景更加显现大模型的人格特质，这是因为闲聊情景更加关注于人际交往中的处事风格。

此外，国际大模型团队也在聊天机器人的性格领域不断尝试突破。Character.AI 团队融资

1.5 亿美元，宣称要为地球上的每个人打造“他们自己的深度个性化超级智能，帮助他们过上最美好的生活”。Character.AI 搭建了用户创建 AI 角色并与之聊天的平台及社区，通过记录用户与 AI 角色的聊天记录和行为偏好，保持 AI 性格模型训练的连贯性和持续塑造性[22]。虽然该团队研发仍处于初期阶段，但其技术实力和发展方向被认为具有成为下一代 AI 全民应用的潜力。

在情感方面，微调大模型的个性化希望通过与人建立共情，来增强用户体验的愉悦感。相比于在知识检索、自动文摘等情景任务下的突出表现，当前大模型技术在情感对话方面能力较弱。因为大模型主要通过语义概率和代码逻辑来执行任务，并不具备心智推理和情感分析能力，因此现阶段大模型还不能完成情感理解并根据情感需求进行回应[23]。但国内外已有多个团队针对情感类大模型展开攻势，例如国内西湖心辰的情感类 AI、美国 Character.AI 团队等[24]。具备情感共情的大模型将会在教育、娱乐、医疗等诸多领域发挥巨大影响。

在记忆方面，大模型的个性化希望能够实现长期记忆和短期记忆的融合，并记住用户的行为、需求和场景偏好。虽然 ChatGPT 也可以按照人们的要求进行聊天，但因为考虑到个人隐私数据保护和产品设定不同，每次重启对话后 ChatGPT 会“忘记”上一次聊天记录，回归到预先设定的泛化智能体的记忆和性格。目前国内外个性化微调大模型团队主要通过搭建全新的智能平台或智能终端环境（智能眼镜、智能音箱、智能手表等），构建相对私密的对话环境，完成用户个人数据的安全收集，尝试实现具有长短期记忆的大模型个性化应用[25]。

总结来说，当前微调大模型的个性化实践应用主要集中于在知识维度上实现专属化。通过预设用户感兴趣的领域话题标签、增加私域知识的增量训练等方式，微调模型参数，以满足用户专属化的信息需求。此外，大模型未来还有可能在个性、情感、记忆的维度逐步突破，实现拟人化、有性格特色、有长短期记忆以及情感共情等功能的智能对话。

四、微调大模型技术下的个性化人机信息交互模式

（一）人机互动下的信息觅食优化

微调大模型的个性化带来全新的人机信息交互模式。传统模式下，人们在互联网检索信息，主要通过信息推荐系统和搜索引擎。这种模式下，信息过滤主要由互联网公司的算法驱动为主导，人类被动接受经算法策展后的信息流。而在智能时代，个性化人机信息交互模式将以用户主动参与作为基础，通过人机互动逐渐优化提示词，不断完善、优化对用户需求内容的理解和大模型的个性化参数。图 1 呈现微调大模型技术下的个性化人机信息交互流程图。

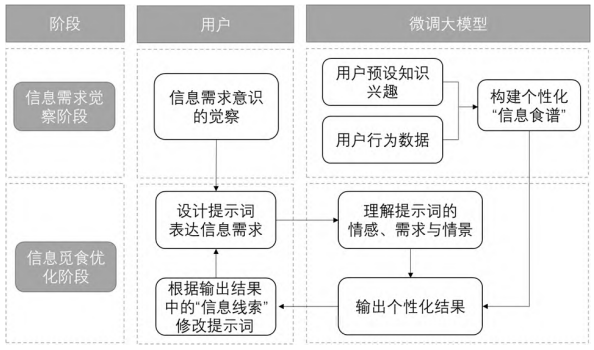

图 1　微调大模型技术下的个性化人机信息交互模式

信息寻求一般被认为是从信息需求的意识觉察开始，信息管理研究认为信息需求是一个从无意识到有意识的过程。信息学家罗伯特·泰勒将其具体描述成四个步骤：从需求的本能状态（visceral status），到有意识的心理描述（conscious mental description），再到形式化的需求（formalized status），最终妥协后的需求落实为一个实际行为或搜索语句（compromised to an actual behavior or searching statement）[26]。信息需求的意识觉察也是个性化人机信息交互模式的第一步，即个体用户觉察到自己有对某个信息内容的需求，并有意识地希望通过某些信息检索的手段获取该信息，为之后将需求落实为人机交互的提示词做准备。

在此同时，微调大模型通过用户预设感兴趣的知识类别以及用户历史行为数据和身份特征的分析挖掘，构建用户的个性化“信息食谱”。

“信息食谱”这一概念来源于“信息觅食”理论，Pirolli 等将人们对信息的选择偏好，比作动物狩猎过程中对猎物的选择[27]。动物生活在拥有多种食物来源的环境中，面临着如何选择食谱来满足自己能量补充的问题。而人们面对多种信息来源的海量信息环境，也需要设计自己的信息选择方案。微调大模型可以通过挖掘用户历史行为数据，在大规模训练的基础上，微调模型参数，构建用户专属的信息食谱偏好。

接下来，用户进入人机交互模式下的信息觅食优化阶段。信息觅食理论认为，人们搜寻信息不是一蹴而就的。相反，人们是在搜寻过程中，面对大量信息源环境，因为未知信息的具体定位，不断根据搜寻效率优化方案的过程。信息觅食理论被认为可以很好地描述互联网搜索引擎环境下人们的信息寻求过程，即用户在使用搜索引擎时，无法在海量信息中迅速定位自己的需求信息，他们会根据搜寻到的信息线索和效率，不断调整搜索关键词或信息源平台[28]。

笔者认为，微调大模型技术下的个性化人机信息交互也是一个信息觅食的优化调整过程。首先，用户将意识到的信息需求落实为具体的提示词语句，输入大模型界面。其次，大模型通过理解提示词的情感、需求与应用场景，结合对该用户构建的专属“信息食谱”微调参数，输出个性化结果。接下来，用户根据大模型在第一轮对话中输出的结果，判断该信息是否在自己本次信息搜寻的菜单中。Pirolli 等将人们进行判断时所处理的局部信息内容，称为信息线索（information scent，又称信息气味）[29]。用户根据首次对话输出结果中的信息线索（如关键词的上下文、链接标签中的线索等）判断信息价值，同时修改向大模型继续发问的提示词。此后，人与大模型的信息交互将不断循环以上步骤，个性化微调大模型在循环过程中逐步加深对用户信息需求的理解，同时用户也在循环过程中加深自己的需求意识、更加精确地设计提示词。

微调大模型下的个性化人机信息交互模式，虽然也是信息觅食优化过程，但其不同于互联网搜索引擎环境下个体用户单方面的优化觅食策略。微调大模型技术下的个性化技术实践，是基于人机双向互动的智能平台。人与大模型在信息觅食优化过程中，实现了双向优化调整策略的互动机制。个体不断精确化自己的信息需求表达，同时在对话过程中加深对大模型输出机制的理解，大模型在用户不断调整提示词的过程中学习了用户的表达习惯和背景知识，成为进一步优化个性化参数的行为数据。

在这种双向优化驱动机制下，个性化人机信息交互模式下的信息搜寻将显著优于互联网信息推荐与搜索引擎，表现为以下三个特点。首先，信息检索效率大幅提高。大模型在大算力支持下对大规模数据进行预训练，其本身在文本挖掘和语义理解方面的表现都优于普通信息推荐与搜索引擎算法。而且，微调大模型还可以在与用户双向互动的过程中，逐步加深对用户需求的理解，调整用户个性化模型参数，检索内容的精准性和相关性将得到显著优化，这将极大利于信息检索的效率提升。

其次，用户对个性化算法的体验更好。个性化人机信息交互模式以用户主动参与作为基础，对话交互形式使得用户可以对个性化系统提出反馈和改进方案。以往个性化信息推荐出现用户感受到“信息隔离”的问题，主要是个性化推荐算法过度拟合用户已有的兴趣和行为数据，导致用户被局限在狭窄的信息范围内，限制了用户接触到多样性的观点和信息的可能性[30]。个性化智能人机信息交互可以通过反馈机制解决这一问题，用户主动参与兴趣预设，并在互动中即时提供反馈，系统可以不断调整和优化策略，以更好地提高用户体验。

最后，个性化智能平台未来有潜力扩展个性化服务范围。人工智能领域正在将大模型与智能家居、智能穿戴等设备连接在一起，随着技术逐渐发展，未来智能平台将会从“人－物－环境”三个维度构建人机“共生关系”[31]。现有个性化服务的范围将被极大地扩展，从仅仅完成个性化信息检索与管理服务，发展为可调用外部工具、提供复合型个性化服务的智能“个

人门户”[32]。个性化系统可以将用户需求编码为指令，调用外部工具（如实时检索、发送邮件、计算器、外部数据资源等）的功能来执行任务，并结合智能家居、智能穿戴等设备以视觉或听觉的形式输出结果。个性化智能平台将可能以前所未有的全能智能管家身份，适配于所有智能终端和场景中。

（二）信息处理与存储：个性化微调大模型作为“人脑的延伸”

麦克卢汉提出“媒介是人的延伸”，认为新技术的产生都是人类自身生理或心理无法应对外界变化而寻求解决问题的出口。大模型技术受到广泛关注后，有学者提出智能媒介将作为人脑的延伸，对人类意识进行多维再造[33]。基于目前国内外大模型团队在知识、性格、情感、记忆等维度的个性化实践研发方向，个性化微调大模型或在不久的将来作为“人脑的延伸”，辅助人们应对信息过载情景下的信息处理与信息存储任务。

首先，个性化微调大模型将可以提供个性化的自动文摘功能，辅助人们应对过载量的信息处理任务。动机性媒介信息加工的有限容量理论（Limited Capacity Model of Motivated Mediated Message Processing，LC4MP）认为，信息处理是通过编码、储存和检索三个子过程转换成一个动态的记忆表征平行进行的[34]，三个子过程都可能会发生认知资源耗竭，导致信息无法被彻底加工[35]。面对互联网多模态形式的信息海洋，人们常常会因为时间、注意力等认知资源有限，面临信息过载、信息焦虑的困境[36]。

根据用户个人信息数据微调后的微调大模型将有希望帮助人们改善这个局面。一方面，微调大模型可以对信息量过载的文本、图片、视频内容进行自动化的信息提取，提供摘要性的信息内容，辅助用户高效完成信息编码。另一方面，微调大模型还可以根据用户已有领域知识背景，有侧重点地对新知识、新内容提供解释说明，缩略概述用户已知信息内容，重点诠释新知识内容，并根据用户需求对新知识进行链接与延展。

同时，微调大模型支持下的个性化智能信息服务，在未来或将帮助用户拓展信息存储的极限。人脑的信息存储受限于人脑的存储记忆资源，研究认为人脑短时记忆只有5—9个信息单元[37]，当个体用户在短时接收过量信息时，信息存储效率会大幅降低，表现为疲惫、记不住等现象[38]。

微调大模型支持下的个性化智能信息服务，在未来将会帮助个体用户拓展信息存储的极限。目前国内外多个大模型团队都已将大模型个性化实践的记忆功能纳入主要研发方向中。Character.AI通过搭建新智能平台，收集用户与多人设智能体的对话记录，实现对用户数据的长、短期记忆[39]。天猫精灵通过构建智能眼镜、智能音箱等终端环境，构建用户专属的对话环境，完成骨传导、语音数据的安全收集，尝试实现用户数据的记忆[40]。未来个性化微调大模型将有可能辅助用户完成信息存储，人们可以将人脑短期无法记忆的超额内容，以文字、音频、图像等形式存储至智能终端的大模型长短期用户记忆数据中。例如，人们在社交情景下常常会面临信息存储超负荷的困扰，因为新结识友人数量多，超过人脑短时记忆单元，导致无法将友人的姓名、面庞、职业、单位及其他信息及时对应并存储下来。未来，借助于个性化微调大模型辅助下的智能眼镜，用户可以将新结识友人的信息和图像存储于大模型的长期记忆中。当用户再次佩戴智能眼镜遇到此前结识的友人，个性化微调大模型可以自动调取已储存的友人相关信息，并通过智能眼镜构建骨传导、透明显示器等人机私密对话空间，以语音或文字的形式提醒用户。

五、个性化带来的新挑战

面对当前人工智能技术和人类形成的复杂人机关系，微调大模型的个性化还需要谨慎解决数据隐私保护、媒介伦理、数据治理等诸多数据安全与伦理意识问题。因为智能技术有其自身逻辑，也有其不可预料性，它们的设计和应用可能受到利益集团、政治力量、社会价值

观和用户心理的多重影响，因而技术问题不能简单地通过技术本身来解决，还需要涉及社会、伦理和人文层面的全盘考量。

（一）个性化训练的数据泄露风险与媒介伦理意识

为个体用户打造个性化信息管理服务时，开发者需要收集用户的个人信息及行为数据，包括用户个人信息、职业身份、特定行业知识、特定场景下的信息偏好与使用习惯等。为了更好地满足个人用语、个性特点等信息生产需求，还有可能需要收集用户日常聊天记录、语音视频等作为微调个性化参数的训练数据集。当这些数据传输到开发者服务器时，可能存在隐私数据泄露风险。此前 ChatGPT 已有发生部分用户支付信息因缓存软件开源库错误而泄露的问题[41]，如果在个性化大语言模型训练中出现类似数据泄露问题，将会对用户隐私保护和个人权益造成严重损失[42]。

大型自然语言处理模型在安全防御技术本身存在难度。现阶段模型技术针对数据投毒、数据提取攻击等活动的防御能力依然不稳定。当大语言模型的参数越来越大，其脆弱性问题逐渐凸显，防御难度逐渐加大，隐私泄露风险也会更加普遍[43]。如何加强大语言模型的安全防御技能，建立数据安全与数据治理机制，成为当前开发者与监管部门共同面对的重要课题。

此外，个性化大语言模型应用还需格外注重媒介伦理风险。当前大语言模型技术主要依赖文本数据前后文相关性生成内容，尚未产生主观意识，而当通用人工智能逐渐向自主学习、自我迭代方向继续发展，其是否会朝着开发者预期计划的方向发展不得而知。大语言模型的个性化训练应具备尊重并保护人权、保护社会多样性、保护隐私权、拒绝歧视、维护公平公正和向善性等伦理要求[44]，在人工审查试用与政府监管中逐步补充并完善数字社会人际互动的伦理框架与道德规范标准。

（二）人工智能的算法开放性与数据多样性

个性化大语言模型需要格外注重个性化算法的开放性问题。在对个性化新闻推荐系统的研究中，部分学者提出了对“信息茧房”的担忧，质疑基于语义过滤、协同过滤等算法的个性化推荐模型是否会导致个体用户接收到的信息过于窄化，从而难以接触到观点不一致的内容[45]。微调后的大语言模型在提供个性化信息管理服务时，也需要平衡个人化精准资讯服务与多元化开放选择之间的关系，避免因个性化算法设计造成信息“封闭”。英国卫报通过专栏“戳破你的泡泡”（Burst Your Bubble），定期列出对立党派的文章，帮助用户扩展其已有观点外的信息视野[46]。个性化大语言模型在提供信息检索服务时，也可以适当参考新闻推荐系统中对个性化算法的“破茧”实践，在输出与用户需求最相关的精准信息的同时，设计相关“破茧”功能，提醒用户在互动中反馈个性化体验，平衡、多元化其信息消费。

此外，现有大语言模型的预训练主要基于英文语料库，这会导致其对不同语言文化的代表性存在偏差和数据多样性不足。从公开的 GPT-3 训练数据集来看，其超过 92% 的语料来自于英文语料，法语、德语等其他国家语言占比均低于 2%，中文语料占总语料库不足 0.1%[47]。在模型运算过程中，数据集代表性不强、多样性不足很可能会影响模型运算在统计性和科学性上出现偏差，进而在应用层面产生系统性问题。已有文章发现，ChatGPT 的输出有着鲜明的美国左派立场，在俄乌战争、抗美援朝战争、素食主义等问题上，ChatGPT 生成的内容有失偏颇，俨然成为被英语世界操控、贯彻西方意识形态的宣传工具[48]。大模型的个性化也需要在意识形态、数据代表性、多样性等问题上格外关注，如何构建高质量中文数据集、训练出体现我国价值观体系的大模型意义重大。

未来，个性化大语言模型应用的市场化需要构建更加完善的数据治理框架。从技术属性来看，大语言模型是现阶段通用人工智能的技术底座，其应用产业链条广泛，未来很可能被赋权参与更多数字社会的构建。因此，微调大模型的个性化需要搭建面向产业链特征的数据治理框架，从开发者、部署者、用户个体、接收者、

监管部门等多元主体构建数据责任矩阵，建立灵活高效的监管工具体系，完善面向产业链特性的法律监管制度[49]。另一方面也要为用户的个性化使用留有继续观察、勘探与应对的余地。

参考文献

[1] 王静静，叶鹰．生成式 AI 及其 GPT 类技术应用对信息管理与传播的变革探析 [J/OL]. 中国图书馆学报 :1-12[2023-10-04].http://kns.cnki.net/kcms/detail/11.2746.G2.20230508.1612.002.html.

[2] Dillion D, Tandon N, Gu Y, et al. Can AI language models replace human participants?[J]. *Trends in Cognitive Sciences*, 2023, 27(7):597-600.

[3] Savolainen R. Everyday life information seeking: Approaching information seeking in the context of "way of life"[J]. *Library & Information Science Research*, 1995, 17(3): 259-294.

[4][8] Heath, A.OpenAI is letting anyone create their own version of ChatGPT [EB/OL]. (2023-11-07) [2023-11-07]. https://www.theverge.com/2023/11/6/23948957/openai-chatgpt-gpt-custom-developer-platform

[5] Sanderson K. GPT-4 is here: what scientists think[J]. *Nature*, 2023, 615(7954): 773.

[6] 林志佳．Meta 联手微软挑战大模型格局，最新 Llama 2 免费开源，可直接商用 [N]. 钛媒体 ,2023-07-19.

[7] OpenAI,Inc. GPT-3.5 Turbo fine-tuning and API updates[EB/OL].[2023-10-04]. https://openai.com/blog/gpt-3-5-turbo-fine-tuning-and-api-updates.

[9] Porsdam Mann S, Earp B D, Møller N, et al. AUTOGEN: A personalized large language model for academic enhancement—Ethics and proof of principle[J]. *The American Journal of Bioethics*, 2023, 23(10): 28-41.

[10][18][21] 高星．个性化大模型技术实践 .DataFunSummit 2023[C] // 大模型与 AIGC 峰会 DataFun, 2023: 1-36.

[11] 孙奇茹．首批 AI 大模型面向公众开放 [N]. 京报网 ,2023-09-01.

[12] Baichuan, Inc.Baichuan2 [EB/OL]. (2023-09-20) [2023-10-06]. https://github.com/baichuan-inc/Baichuan2.

[13][25][40] 王思原．个性化大模型"装进"随身终端，不是想象，是风向 [EB/OL]. (2023-05-19) [2023-10-29]. https://mp.weixin.qq.com/s/fsVBzBtnx5HteIw6IO2r7Q.

[14][15] Chen J, Liu Z, Huang X, et al. When large language models meet personalization: Perspectives of challenges and opportunities[J]. *arXiv preprint arXiv*:2307.16376, 2023.

[16] Miraz M H, Ali M, Excell P S. Adaptive user interfaces and universal usability through plasticity of user interface design[J]. *Computer Science Review*, 2021(40): 100363.

[17] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. *arXiv preprint arXiv:*1810.04805, 2018.

[19] Petroni F, Rocktäschel T, Lewis P, et al. Language models as knowledge bases?[J]. *arXiv preprint arXiv*:1909.01066, 2019.

[20] Martino A, Iannelli M, Truong C. Knowledge injection to counter large language model (LLM) hallucination[C]// European Semantic Web Conference. *Cham: Springer Nature Switzerland*, 2023: 182-185.

[22][39] Perez S.Character.AI introduces group chats where people and multiple AIs can talk to each other [EB/OL]. (2023-10-12) [2023-10-31]. https://techcrunch.com/2023/10/11/character-ai-introduces-group-chats-where-people-and-multiple-ais-can-talk-to-each-other/.

[23] 更具情感的大模型，会给人类带来什么？ [EB/OL]. (2023-07-10) [2023-10-31].https://mp.weixin.qq.com/s/u4IEuOJVPVfsBNRyjwlFYA.

[24] 刘科，陈抗．汤姆猫连投两轮的西湖心辰：一年半内把高情商大模型"做到国际顶尖"[N]. 财联社 .(2023-07-12) [2023-10-31].https://baijiahao.baidu.com/s?id=1771221050051707332.

[26] Taylor R S. Question-negotiation and information seeking in libraries[J]. *College & Research Libraries*, 1968, 29(3): 178-194.

[27][29] Pirolli P, Card S. Information foraging[J]. *Psychological Review*, 1999, 106(4): 643.

[28] Pirolli P. *Rational analyses of information foraging on the web*[M]//Dictionary of World Philosophy. London: Routledge, 2013: 343-373.

[30] Hou L, Pan X, Liu K, et al. Information cocoons in online navigation[J]. *iScience*, 2023, 26(1): 105893.

[31][32] 彭兰．新"个人门户"与智能平台：智能时代互联网发展的可能走向 [J]. 新闻界 ,2023(9):4-14+96.

[33] 邵婉霞，徐啸．智能媒介对人类意识的延伸与再造机制 [J]. 编辑之友 ,2023(9):71-77.

[34] Lang A. The limited capacity model of mediated message processing[J]. *Journal of Communication*, 2000, 50(1): 46-70.

[35] Zechmeister E B, Nyberg S E. *Human memory: An introduction to research and theory*[M]. Monterey, CA: Brooks. 1982: 85;97.

[36][38] Lang A. Limited capacity model of motivated mediated message processing (LC4MP)[C]//The International Encyclopedia of Media Effects. New Jersey, USA: John Wiley & Sons.

[37] Boutla M, Supalla T, Newport E L, et al. Short-term memory span: Insights from sign language[J]. *Nature Neuroscience*, 2004, 7(9): 997-1002.

[41] OpenAI, Inc.March 20 ChatGPT outage: Here's what happened [EB/OL].[2023-03-24]. https://openai.com/blog/

（下转第 76 页）

which reflects his basic understanding of newspaper management, that is, ensuring the survival of the newspaper is more important than expressing political views; at the same time, he successfully unified the economic interests and political communication of subscription advertisements for the headline. This article takes the headline advertisements of the Neue Rheinische Zeitung as the research object, examines the many practical problems encountered in the newspaper management process and Marx's management strategies, and further deepens the study of Marx's advertising thought.
Keywords: Headline advertisements of the Neue Rheinische Zeitung; Marxist Viewpoint on journalim; Marx's advertising thought; newspaper management strategies
Authors: Chen Lidan, Sichuan University; Renmin University of China. Rong Xueyan,The College of Literature and Journalism of Sichuan University.

（上接第 51 页）

march-20-chatgpt-outage.
[42][49] 张欣 . 生成式人工智能的数据风险与治理路径 [J]. 法律科学 (西北政法大学学报),2023,41(5):42-54.
[43] Carlini N, Tramer F, Wallace E, et al. *Extracting training data from large language models*[C]//30th USENIX Security Symposium (USENIX Security 21). 2021: 2633-2650.
[44] 魏雪松 . 提高生成式人工智能伦理要求的研究 [J/OL]. 经营与管理 :1-11[2023-10-05].
[45] 陈昌凤 , 仇筠茜 . "信息茧房" 在中国：望文生义的概念与算法的破茧求解 [J]. 新闻与写作 ,2020(1):58-63.
[46] Spohr D. Fake news and ideological polarization: Filter bubbles and selective exposure on social media[J]. *Business Information Review*, 2017, 34(3): 150-160.
[47] OpenAI, Inc.Dataset Language Statistics [EB/OL]. (2020-06-01) [2023-10-06]. https://github.com/openai/gpt-3/commits/master/dataset_statistics.
[48] 郑海阳 . 关于警惕 ChatGPT 滋生意识形态风险的建议 [EB/OL]. (2023-02-15) [2023-10-06].https://www.mjshsw.org.cn/detailpage/jyxc-f6187d4e-cda5-4265-b14f-710d875d07a8.html.

Fine-tuning Large Language Models: The Analysis of Personalized Human-Computer Information Interaction Patterns

Guan Lu, He Kang, Dou Weihong

Abstract: OpenAI has recently launched "custom GPT" based on GPT-4, which allows users to conveniently create their own large language models. Domestic and international teams on large language models crowded into the personalized intelligent application track, marking the beginning of the era of personalized intelligent information interaction. This article believes that the human-computer information interaction model under the personalized fine-tuning large language models will be a process of two-way interaction between humans and machines to optimize the information for aging strategies. During conversations, individuals continue to revise the prompts to more accurately express their information needs. At the same time, the large language models continuously deepen their understandings of the users' information needs by learning the prompt input by the user, and provides more accurate and personalized responses. In the future, fine-tuned large language models may also serve as "extensions of the human brain" to assist people in dealing with information processing and information storage tasks in information overload situations.
Keywords: fine-tuning; large language model; personalization; human-computer information interaction
Authors: Guan Lu, School of Journalism, Fudan University; Research Group of Computational and AI Communication at Institute for Global Communications and Integrated Media, Fudan University. He Kang, School of Humanities and Communication, Ningbo University. Dou Weihong, School of Journalism and Communication, Lanzhou University.