

## 大语言模型时代的社会机遇与挑战

陈慧敏<sup>1</sup> 刘知远<sup>2</sup> 孙茂松<sup>2</sup>

<sup>1</sup>(清华大学新闻与传播学院 北京 100084)

<sup>2</sup>(清华大学计算机科学与技术系 北京 100084)

([huimchen@tsinghua.edu.cn](mailto:huimchen@tsinghua.edu.cn))

## The Social Opportunities and Challenges in the Era of Large Language Models

Chen Huimin<sup>1</sup>, Liu Zhiyuan<sup>2</sup>, and Sun Maosong<sup>2</sup>

<sup>1</sup>(School of Journalism and Communication, Tsinghua University, Beijing 100084)

<sup>2</sup>(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

**Abstract** Large language models such as ChatGPT have recently gained widespread attention in society due to their high level of generality and intelligence. This technological advancement has sparked not only a technological revolution, but also a social revolution. Against this backdrop, exploring the evolution and characteristics of large language models based on the underlying technology of ChatGPT, as well as their potential opportunities and challenges for society, has become a crucial topic in contemporary times. We first review the technological development of large language models, then summarize the revolution into three stages and discuss their key technical features, respectively. Subsequently, from the perspective of the interaction between technology and society, we explore the social opportunities brought about by large language models: serving as a personal universal information assistant, turning into a new paradigm of scientific research, contributing to industrial infrastructure, and becoming a powerful tool for modernizing national governance capabilities. What's more, we address potential challenges that arise from large language models. These include issues of information pollution, distribution of social power, ethical and legal concerns, and ideological security. Finally, we conclude with reflections and discussions on the future development of large language models and their relationship with societal progress.

**Key words** large language models; ChatGPT; artificial intelligence; technological evolution; social revolution

**摘要** 大语言模型 ChatGPT 因其高度通用化和智能化的表现受到广泛社会关注,其引发的技术革命也正逐渐延伸为一场社会革命。在此背景下,基于 ChatGPT 背后大语言模型技术的演进路线与特征,探索大语言模型技术对社会的潜在机遇与挑战,成为当下社会的重要议题。首先回顾大语言模型技术发展,将其演进总结为 3 个发展阶段并论述关键技术特征,然后基于技术与社会互动视角分别探讨大语言模型技术引发的社会机遇:个人通用信息助理、科学研究新范式、产业基础设施、国家治理能力现代化有力工具,以及面临的潜在挑战:信息污染问题、社会权力分配问题、伦理和法制问题、意识形态安全问题。最后,对大语言模型技术发展及其与社会发展关系的未来展开一定思考与讨论。

**关键词** 大语言模型; ChatGPT; 人工智能; 技术演进; 社会变革

收稿日期: 2023-08-23; 修回日期: 2024-02-06

基金项目: 国家自然科学基金项目 (62106126); 国家社科基金重点项目 (21AZD143); 清华大学-丰田联合研究基金专项资助和清华大学国强研究院资助

This work was supported by the National Natural Science Foundation of China (62106126), the Key Project of the National Social Science Fund of China (21AZD143), and Tsinghua-Toyota Joint Research Fund, the Guoqiang Institute, Tsinghua University.

通信作者: 刘知远 ([liuzy@tsinghua.edu.cn](mailto:liuzy@tsinghua.edu.cn))

中图法分类号 TP391

2022年11月,美国著名人工智能研究实验室 OpenAI 发布基于大语言模型的聊天生成预训练转换(chat generative pre-trained transformer, ChatGPT),在知识理解、逻辑推理、人机交互等方面表现出高度通用化和智能化特征,在全世界引发广泛关注和使用:在上线后不足2个月的时间里,即获得1亿活跃用户,成为历史上用户数增长最快的面向消费者的应用<sup>[1]</sup>.中国工程院外籍院士、微软公司创始人比尔·盖茨称其诞生的意义“不亚于互联网或个人电脑的诞生”<sup>[2]</sup>,美国工程院院士、SpaceX 公司创始人马斯克评价其“离强人工智能不远了”<sup>[3]</sup>.紧随其后,OpenAI 于2023年3月发布包括语言和图片的多模态大模型 GPT-4,被称作大语言模型的新里程碑<sup>[4]</sup>.

广泛的社会关注背后反映的是大语言模型强大的技术优势和应用能力.据测试,ChatGPT 在多个任务场景表现出媲美人类的智能表现:通过宾夕法尼亚大学沃顿商学院开设的工商管理硕士(MBA)课程的期末考试,取得B至B-等级成绩<sup>[5]</sup>;通过美国医学执照考试及格门槛,被麻省理工学院和哈佛大学科学家评价为“敲响医学界的警钟”<sup>[6]</sup>;通过谷歌程序员面试,被评定为三级工程师<sup>[7]</sup>.而2023年3月 OpenAI 发布的 GPT-4 则进一步在各类专业和学术基准测试中表现出与人类相当的水平,如在模拟律师资格考试中获得前10%的成绩<sup>[4]</sup>.在此背景下,美国、俄罗斯、韩国等国的头部互联网企业相继推出自己的 ChatGPT 计划,争相训练具有自主知识产权的通用智能模型,并构建相关应用生态<sup>[8-9]</sup>.其中,微软公司已经将 ChatGPT 模型应用至搜索引擎,并面向普通用户开放使用.我国百度和阿里巴巴等企业也宣布正在自主研发类 ChatGPT 模型,地方政府也力争上游,如北京市提出2023年“将全面夯实人工智能产业发展底座,支持头部企业打造对标 ChatGPT 的大模型,着力构建开源框架和通用大模型的应用生态”<sup>[10]</sup>.

可以预见,伴随技术的不断进步及相关应用生态的持续建设,ChatGPT 及其相关大语言模型技术集群将在不久之后从一种“技术事实”转变为“社会事实”,引发社会各领域的变革.基于此,面向 ChatGPT 背后大语言模型技术的演进范式与特性,探索其对社会的潜在机遇与挑战,是当前人工智能技术发展与社会变迁日益紧密联系时代背景下的重要命题.本文将首先梳理总结 ChatGPT 相关大语言模型技术的演进路线与特性,基于此从技术与社会互动视角探讨

大语言模型技术所引发的社会机遇以及面临的挑战,以期有助于学者们加深对 ChatGPT 技术特性的认识,借此启发对 ChatGPT 潜在社会机遇和挑战的思考.

## 1 大语言模型的技术演进

大语言模型是 ChatGPT 所依托的核心技术.在论述其技术演进路线之前,首先要明确大模型之“大”的含义.笔者认为应从两方面理解其内涵:一方面意指模型规模大,以2017年做前后对比,此前自然语言处理的语言模型参数规模大致在千万量级以下,而自2018年美国谷歌公司发布预训练语言模型——基于变换器的双向编码器表示技术(bidirectional encoder representations from transformers, BERT)后<sup>[11]</sup>,语言模型的参数规模均达到亿级以上.在此之后,预训练语言模型的最大参数规模每年增长数十倍,其中最著名的模型之一是2020年7月由美国 OpenAI 发布的 GPT-3<sup>[12]</sup>,规模达到1750亿参数.另一方面,大模型之“大”意指对大数据蕴含知识的集大成学习,基于大量高性能并行计算芯片 GPU 对海量无标注数据实现自监督高效学习,所学习数据的规模从早期 BERT 中使用的20 GB到 GPT-3 中采用的45 TB.也正是在 GPU 多机多卡算力和海量无标注文本数据的双重支持下,才实现语言模型规模与性能齐飞的局面,成为人工智能和自然语言处理领域的最新革命性突破.

在厘清大模型之“大”的内涵后,再来理解大语言模型的技术演进则显得更为清晰.大语言模型的技术演进路线整体可以分为如下3个阶段.

### 1.1 第1阶段:大语言模型的崛起

第1阶段以2017年美国谷歌公司的研究团队在名为 Attention is All Your Need 的论文中发布全新的神经网络架构 Transformer 作为开端<sup>[13]</sup>,此架构迅速应用到整个人工智能领域,成为自然语言处理及其相关研究领域的主要方法. Transformer 通过摒弃它前任神经网络架构们的不适合 GPU 芯片并行计算的操作,极大提高了神经网络模型从大规模数据中学习的效率,直观上能够在单位时间和算力支持下阅读更多数据,直接导致了此后模型规模的迅速增长.自此,相继诞生了 BERT, GPT/GPT-2<sup>[14]</sup>, T5<sup>[15]</sup>等基于大规模无标注数据自监督学习的大规模预训练语言模型,成为主流人工智能技术和产品的核心基础模型,以及国际人工智能前沿研究和应用的热点.这些

模型均在大数据支持下拥有较大模型参数规模,具备较强通用能力,可适应和完成多场景任务.以谷歌公司 2018 年发布的 BERT 模型为例,其包含 24 层 Transformer 架构、3.4 亿参数的版本在著名的通用语言理解评测(general language understanding evaluation, GLUE)基准所包含的 8 个任务上获得平均得分 80.4,比当时最好模型的得分提高了 7.6 分,并刷新了 SQuAD 等自然语言处理挑战任务上的最好结果(states-of-the-art, SOTA).

这里需要提及的是,从这一阶段开始,大语言模型的预训练技术即被开始分割为 2 个主要类别,分别是以 BERT 为代表的面向语言理解的遮蔽语言模型(mask language model, MLM)和以 GPT 为代表的面向语言生成的自回归语言模型(autoregressive language model, ALM),直接影响了大语言模型第 2 阶段和第 3 阶段的技术演进路线.

具体而言, BERT 所依托的 MLM 的预训练方法如图 1 所示,给定一个句子如  $S_1$ , 随机选择并遮蔽其中某个词如“中国”,得到训练样例  $T_1$ , 其中出现“中国”的位置被遮盖标识符 [MASK] 取代. BERT 的 MLM 任务目标就是,根据  $T_1$  的上下文语境信息,预测被遮盖位置是哪个词,而判断预测是否准确的标准答案  $A_1$ , 就是原句  $S_1$  中的“中国”. 互联网为预训练方法提供了海量的文本大数据. MLM 可以参照上述  $S_1 \rightarrow T_1 + A_1$  的方式,从文本大数据中近乎无限地自动产生训练样例,不断增强 BERT 对语言的理解. GPT 则采用了不同的预训练方法. 如图 2 所示, ALM 的任务目标就是, GPT 能够像人类说话一样,根据前面已经生成的内容,不断预测和生成下一个词. 值得注意的是,由给定文本提供的“标准答案”并不见得完美. 以  $T_6$  任务为例,预测“大学”是对的,预测“高校”“高等学校”“学府”也是对的. 但是根据给定文本,只能认定  $A_6$  “大学”是准确的,而其他预测都会被认为是错误的. 大模型的神奇之处就在于,虽然每个训练样例的标准答案不全然准确,每个样例都是不完美的“教师”“三个臭皮匠,顶个诸葛亮”,数以亿计的不完美训练数据可以教会大模型从概率上无限逼近优秀的语言能力.

$S_1$ : 清华大学是中国最好的大学之一。		
→	$T_1$ : 清华大学是[MASK] 最好的大学之一	$A_1$ : 中国

Fig. 1 An example for MLM

图 1 遮蔽语言模型示例

从 BERT 和 GPT 的预训练任务对比可以看出,

$S_1$ : 清华大学是中国最好的大学之一。		
→	$T_1$ : [MASK]	$A_1$ : 清华大学
→	$T_2$ : 清华大学 [MASK]	$A_2$ : 是
→	$T_3$ : 清华大学是 [MASK]	$A_3$ : 中国
→	$T_4$ : 清华大学是中国 [MASK]	$A_4$ : 最好
→	$T_5$ : 清华大学是中国最好 [MASK]	$A_5$ : 的
→	$T_6$ : 清华大学是中国最好的 [MASK]	$A_6$ : 大学
→	$T_7$ : 清华大学是中国最好的大学 [MASK]	$A_7$ : 之一
→	$T_8$ : 清华大学是中国最好的大学之一 [MASK]	$A_8$ : 。

Fig. 2 An example for ALM

图 2 自回归语言模型示例

两者有不同的假设. BERT 的 MLM 假设文本已经在那里了,只是某个词被盖住了, BERT 可以同时考虑被盖住词左侧和右侧的语境信息来预测被遮盖词; GPT 的 ALM 则假设文本尚不存在,需要模型从左到右进行生成,所以在生成下一个词时只能考虑其左侧已生成的信息. 因此, BERT 模型更适合用来做文本语义理解,进行文本分类、阅读理解等自然语言理解任务,而 GPT 模型则更适合用来做文本生成、对话系统等自然语言生成任务. 同时,相比于 MLM 每训练 1 条文本只预测 1 个词, ALM 可以对多个词进行预测,这使得在相同训练次数的条件下, ALM 比 MLM 学习得更快、模型更易收敛,这直接导致 GPT 系列语言模型的参数规模具备更优的扩展性.

## 1.2 第 2 阶段: 超大规模语言模型的发展

大语言模型发展的第 2 阶段即是以 2020 年 OpenAI 发布的 GPT-3(1 750 亿参数)为开端,开启了超大规模预训练语言模型的新时代,在知识理解、语言生成、上下文学习等方面展现出惊人能力. 根据 MIT Technology Review 报道,基于 GPT-3 提供的 API 所生成的博客文章在短短几个小时迅速传播并登上了 Hacker News 头条,而极少有人发现其为机器生成<sup>[16]</sup>. 自此,全球范围内众多机构开始投身于超大规模预训练语言模型的研究工作,国内外如 Meta、微软、谷歌、清华大学、北京智源人工智能研究院、百度、华为、阿里巴巴等,纷纷加入到这场模型竞赛中.

大模型参数规模的增长,可以让模型能力得到增强,且呈现出从量变到质变的增强过程. 从量变角度看,大模型在 GLUE 等评测基准上性能持续攀升;从质变角度看,大模型正在实现从“举十反一”到“举一反三”乃至“无师自通”的能力飞跃. “举十反一”指传统机器学习(包括预训练大模型之前的深度学习模型)需要依赖大量训练数据才能习得某个特定任务. 例如,需要收集成百上千的猫和狗的图片,才能让



模型学会区分猫和狗这 2 种动物,需要收集成百上千的主题为“国内”和“国际”的新闻文本,才能让模型学会对这两类新闻的分类能力。“举一反三”指人类只需要通过非常少的训练样例,就能习得某个特定任务。例如,人类可以仅通过少量(如 1~2 个)熊猫的样例,就可以建立起对熊猫这种动物的识别能力。这被认为代表了人类高级智能水平的认知能力,传统机器学习方法无法做到,被命名为少次学习(few-shot learning),成为近年来的研究重点。“无师自通”指人类甚至不需要任何训练样例,就能掌握某种新的能力。例如,程序员在熟练掌握某个编程语言后,很多时候不需要任何示例代码,光看函数接口就会使用新的工具包。这也代表了人类高级智能水平的认知能力,被命名为零次学习(zero-shot learning)。如表 1 所示,以英文-中文机器翻译为例,可以看到,“少次学习”任务仅提供非常少的训练样例,仅 3 个;“一次学习”是少次学习的极端特殊情况,即只有 1 个训练样例;而“零次学习”则只有任务描述,没有任何训练样例。

Table 1 Differences in Tasks with Few-shot Learning, One-shot Learning, and Zero-shot Learning

表 1 少次学习、一次学习与零次学习的任务区别

学习方式	少次学习	一次学习	零次学习
任务描述	Translate English to Chinese:	Translate English to Chinese:	Translate English to Chinese:
任务样例	university => 大学 Beijing => 北京 computer => 计算机	university => 大学	
测试样例	Tsinghua =>	Tsinghua =>	Tsinghua =>

大语言模型在少次学习和零次学习上的能力飞跃,在 GPT 模型的发展历程中体现得淋漓尽致。GPT-3 的开发团队训练了 1.25 亿(Small)、3.5 亿(Medium)、7.6 亿(Large)、13 亿(XL)、27 亿(2.7B)、67 亿(6.7B)、130 亿(13B)和 1 750 亿(175B)等 8 个不同参数规模的模型,并在评测基准 SuperGLUE 上验证这些模型在少次学习和零次学习上的性能变化。随着 GPT-3 参数规模的增大(13 亿、130 亿和 1 750 亿),模型的少次学习和一次学习能力有飞跃式的进步,特别是到 1 750 亿时甚至具备了一定程度的零次学习能力<sup>[12]</sup>。这也是 GPT-3 的作者在论文题目中强调“Language Models are Few-Shot Learners”<sup>[12]</sup>(语言模型是少次学习者)的原因。

### 1.3 第 3 阶段:通用大语言模型的进化

大语言模型的第 3 阶段则是以如今广为人知的

由 OpenAI 发布的 ChatGPT 作为开端,其在具有强大性能的 GPT-3 基础模型上,引入代码训练、指令微调(instruction tuning)和基于人类反馈的强化学习(reinforcement learning from human feedback, RLHF)等技术,进一步增强大语言模型的自然交互和人类对齐能力,使其在面对常识性和推理性问题、未知领域和敏感议题时展现出显著智能化的特征。在此之后,OpenAI 所发布的多模态大模型 GPT-4,支持同时以文本和图像多种模态运行,其认知和推理能力或将带来通用人工智能的再一次变革。

在前 2 个阶段,大语言模型的预训练目标是优化文本预测或生成中的交叉熵损失函数,保障其生成的文本的流畅度与可读性。然而真实应用场景中,大语言模型需要具备与人类交互的能力,高效理解和完成用户的指令,这与既有预训练目标存在较大的差异。因此,为弥合大语言模型的训练目标与人机交互目标之间的差异,OpenAI 提出了指令微调与来自人类反馈的强化学习两阶段技术路线。

指令微调旨在训练大语言模型根据人类给定的指令来执行特定任务或生成特定内容的能力。例如,当人类用户发出指令:“请写一首关于春天和爱情的现代诗”,大语言模型会根据人类用户的要求生成一首符合主题“春天和爱情”和风格“现代诗”的诗歌。通过基于指令与回答文本对的有监督微调,指令微调使得大语言模型能够充分吸收其从大规模语料中学习并掌握的世界知识、常识知识,生成符合人类用户需求的文本,进而灵活适应不同应用场景和用户需求,提高生成内容质量和多样性。不过,大语言模型的预训练语料中也不免存在许多带有偏见与恶意的言论,指令学习尽管针对人类用户的需求进行优化,但仍无法避免模型生成具有伦理风险的言论。

在此基础上,OpenAI 进一步将来自人类反馈的强化学习引入大语言模型的训练过程。来自人类反馈的强化学习利用人类标注者对大语言模型所生成内容的打分或排序作为模型的反馈,并根据这些反馈调整模型参数。为使得大语言模型能够符合伦理规范,OpenAI 为人类标注员设置了 3 种打分标准,分别是有用(helpful)、无害(harmless)、诚实(honest)。例如,当模型所生成文本中蕴含性别、种族等偏见或其他有害信息时,模型将根据人类标注员的负面打分或者排序接收到负面的反馈,促使模型之后避免生成此类文本。依托于此,大语言模型的可靠性和实用性得到有效提升。

通过结合指令微调和来自人类反馈的强化学习

这2项技术, ChatGPT在与人类用户对话交互方面展现出卓越表现. 例如, 在程序调试方面, ChatGPT可以根据用户提供的程序代码片段回答错误原因或解决方法, 还会主动提问以获取更多信息并确认理解是否正确; 在敏感话题方面, ChatGPT可以根据用户提问时使用的措辞或意图判断是否应该回答. 正因为这种强大的认知与交互能力, ChatGPT一经发布便迅速引起社会的广泛关注, 成为通用人工智能发展历程中的重要里程碑.

## 2 大语言模型的社会机遇与挑战

科学技术是社会变革重要的推动力量. 正如伟大的思想家马克思<sup>[17]</sup>所言, “科学是一种在历史上起推动作用的、革命的力量”, 新技术的诞生和应用往往对社会生产力、社会发展方向起到决定性作用. 与此同时, 技术的推广和使用也受到诸多社会因素的选择、推动和制约. 英国著名媒介研究学者温斯顿<sup>[18]</sup>曾提出“加速器-制动器”理论, 指出社会需要的是促使技术原型得以被选择并被社会所推广的决定性因素, 同时社会也具有制约新技术所释放出的“激进潜力”的力量, 使得新技术不断适应社会结构. 伴随大语言模型技术的出现和发展, 一些机构开始对大语言模型所催生的可能的社会影响做出展望. 斯坦福大学基础模型研究中心曾从生物医疗、法律、教育方面讨论其社会应用, 从不平等、误用、环境、法律、经济和伦理6个方面探讨其社会影响<sup>[19]</sup>. 著名人工智能研究机构DeepMind则从歧视、隐私、虚假信息等6个角度详细论述了大语言模型的伦理和社会风险<sup>[20]</sup>. 这些探讨多从技术本身出发, 探索技术发展可能造福的产业以及技术本身存在的风险问题. 伴随大语言模型的通用智能特性逐步展现, 也有一些学者从自身特定研究领域出发讨论大语言模型技术引发的可能影响, 涉及如教育<sup>[21]</sup>、医疗<sup>[22]</sup>、新闻<sup>[23]</sup>等领域. 区别于已有研究, 本文从技术与社会互动的视角出发, 探索大语言模型这一新技术的出现, 与当前的社会需要相匹配会催生哪些社会变革机遇, 又有哪些可能的“激进潜力”需要被进一步规制.

### 2.1 社会机遇

大语言模型技术在各任务领域展现的通用智能特征, 呈现出在个人、科学、产业和国家各层面的广泛应用潜力. 因此, 我们将就个人、科学、产业和国家层面分别对其催生的社会机遇展开论述.

1) 在个人层面, 大语言模型技术将催动个人通

用信息助理成为现实. 在当前信息爆炸的时代, 每一个网络用户面对的是海量的信息资源, 即使拥有搜索引擎、社交媒体等发挥信息汇聚功能的入口, 人类自身所能阅读和处理的信息毕竟是有限的. 同时, 伴随人类知识水平的不断提升, 对更高质和高效信息生产的渴求也日益旺盛. 这些个人层面的社会需要与大语言模型强大的信息整合、信息理解和信息生产能力遥相呼应, 催化大语言模型成为个人通用信息助理. 如同互联网将人与人互联, 大语言模型将成为人与机器智能、人与高质信息互联的一条重要通道.

美国著名计算机科学家、麻省理工学院媒体实验室创办人尼古拉斯·尼葛洛庞帝早在20世纪90年代就曾预言“未来的界面代理人可以阅读地球上每一种报纸、每一家通讯社的消息, 掌握所有广播电视的内容, 然后把资料组合成个人化的摘要. 这种报纸每天只制作独一无二的版本”<sup>[24]</sup>. ChatGPT的海量信息整合和强大的信息理解能力, 将使这一预言变为现实. 在现阶段信息爆炸的时代, ChatGPT具有成为下一代信息入口的巨大潜力, 通过与ChatGPT对话式的交互形式, 信息将自动被高质和高效整合并传播给任意用户, 个人的信息获取门槛大幅度降低, 信息获取效率得到提升, 进而在社会层面促进公众知识的极大丰富, 以及科学和信息素养、专业技能的迅速提升. 其中, 微软公司已经开始将GPT4应用于搜索引擎必应(Bing)中, 可以根据用户定制化的要求实现信息的自动整合, 截至目前已经收获超过1亿日活跃用户<sup>[25]</sup>. 与此同时, 结合ChatGPT的强大信息生成能力, 个人的内容和知识生产力也将得到极大提升, 例如通过ChatGPT辅助人类高效内容创作和内容校正, 甚至可以替代部分重复度较高、目标相对确定的内容生产工作. OpenAI机构曾通过研究预测未来人类如写作和编程类的内容创作技能将很容易受到GPT类模型的影响甚至替代<sup>[26]</sup>. 这种通过同一平台自动聚合式的信息获取和信息生产方式, 也将数字媒介的“无偏见性”<sup>[27]</sup>进一步延展, 公众因信息获取能力差异而形成的“使用沟”“知识沟”可能被部分弥合. 例如, 既有信息获取能力相对较差的公众可以通过大语言模型自动便捷地获取来自多来源的高质信息, 实现信息获取能力的补充和完善, 以及个人知识储备和生产能力的提升, 进而弥合与原先信息获取能力较强的公众之间的差异. 因此, 伴随大语言模型的普及, 不仅个人的生活和工作模式将改变, 人机交互将成为信息获取和生产的主要手段之一, 社会层面的知识和文化也将进一步丰富和普及.

2)在科学研究层面,大语言模型技术可以加速推动科学研究方法新范式,提升科学生产力。随着科学数据的爆炸式增长和复杂性的不断提高,传统科学研究方法逐渐制约科学生产力。2018年“AI for science”概念被提出,强调AI拥有服务于科学知识生产的巨大潜力,开启了科学研究的新范式。2023年国际经济合作和发展组织发布报告《科学中的人工智能:挑战、机遇与科学的未来》,全面总结了人工智能在科学发现、科技创新、知识创造、高质量教育等各方面的重大潜力,指出人工智能可帮助科学知识发现、促进科学更加高效。在此背景下,大语言模型技术凭借复杂数据处理和理解以及通用跨领域知识整合的能力成为深化科学研究新范式、提升科学生产力的强大推动力。

一方面,大语言模型可以服务于科学研究前期的科学文献调研与信息收集,以对抗科学知识领域的“信息过载”。Scopus、Dimensions和Web of Science等数据库皆在引入对话式人工智能科学文献搜索,以自动生成文献摘要,并致力于构建科学知识脉络和知识图谱<sup>[28-29]</sup>。另一方面,大语言模型可以服务于科学研究过程中的科学数据高效处理与分析。基于大语言模型的技术原型,迁移到海量科学数据所建构的科学大模型,不仅可以为科学数据提供更加有意义的表示,而且可以从数据中自动实现科学假设的生成、科学规律的发现等,加速和提升科学研究<sup>[30]</sup>。除此之外,大语言模型不断提升的类人智能也为基于社会模拟的科学知识发现引入了无限潜力。一项来自斯坦福大学和谷歌公司的研究将25个类人智能体放置在沙盘环境中,背靠大语言模型完成行动决策,展示了智能体在信息传播和集体活动等方面的类人行为<sup>[31]</sup>。未来类似架构的不断完善,有助于研究者更加便捷高效地开展基于类人智能体的社会实验,观察和挖掘人类行为背后潜在的规律,为科学研究开拓全新范式。

3)在产业层面,大语言模型技术将成为各行业的基础设施,推动产业格局变迁。在如今科技迅速发展的背景下,科技创新成为企业发展的源动力,推动企业效能的提升。谷歌公司前CEO埃里克·施密特曾提出“反摩尔定律”,意指企业如果不能追赶上科技发展的摩尔定律,收益将遵循反向的摩尔定律,并很快被社会所淘汰。习近平总书记在党的二十大报告中也曾强调科技创新对企业发展的重要性,并指出“强化企业科技创新主体地位”。在此时代背景下,以ChatGPT为代表的先进大语言模型技术则是企业提

升效能和促进产业转型的重要基石。

如今数字化和信息化已成为各行业的基础形态,大语言模型则成为各行业智能化转型的关键基础设施。美国《财富》杂志对1000名企业领袖的调查显示,近50%的企业已经在公司应用ChatGPT,另有30%的企业正在计划使用<sup>[32]</sup>。依托大语言模型技术,各行业如电商、教育培训、法律、金融、医疗等可以通过降低人力成本进而提高产业效能:电商平台可以根据用户定制化需求提供智能推荐、智能对比和分析,降低用户决策成本并提升商品转化率;教育培训产业则可以应用ChatGPT替代部分知识培训服务,或辅助人类教学进而提升教学质量;法律、金融、医疗类产业也可以利用ChatGPT的强大信息理解和数据分析能力,提供智能咨询服务与辅助决策。同时,由于ChatGPT依托的基础模型具有强大的通用性能以及各领域任务的高效适配能力,未来的产业格局将很可能形成围绕ChatGPT基础模型的超级生态和各应用产业的垂直生态并存的格局。正如美国著名政治家基辛格、谷歌公司前首席执行官埃里克·施密特、麻省理工学院计算机学院院长丹尼尔·胡滕洛赫尔在《华尔街日报》联合发表的评论中所提及,“为大型语言模型的独家使用付费超出大多数企业的范围。这些模型的开发者可能会出售订阅,以此满足成千上万个人和企业的需求。因此,在未来10年中,大型语言模型的数量可能会相对受限。这些模型的设计和控制将高度集中”<sup>[33]</sup>。可以预期,未来对于大型互联网产业而言,其具备训练ChatGPT基础模型的人力、算力和数据资源,可以通过创建大型基础模型提升产业效能,并面向其他细分产业输出基础模型接口,形成围绕ChatGPT基础模型的超级生态;对于各细分领域产业而言,可以通过结合基础模型接口与自身产业数据库和知识库,形成更加精细完善的垂直生态。

4)在国家层面,大语言模型技术将可成为推动国家治理能力现代化的有力工具。伴随信息技术日益广泛渗透到社会各个领域,国家治理面向的对象和内容都越发以信息化甚至智能化的形式呈现,这要求国家治理的手段也随之不断更新和调整。其中,充分把握信息技术和智能技术发展所形成的技术优势,并将其转化为治理效能的提升则是推进国家治理能力现代化的重要基石。2020年时,美国国会曾颁布《政府人工智能法案》,要求联邦政府各部门在系统运行中尽可能使用人工智能。我国也曾多个政府文件中,提出重视运用人工智能技术提升治理能



力和治理现代化水平.在此背景下,以 ChatGPT 为标志的具有高度智能表现的大语言模型则可成为助推高效能国家治理的有力工具.

对于政府治理能力提升而言,一方面,大语言模型技术可以为政府部门提供快速全面的信息搜集、信息汇总和信息生产,并通过与政务人员的智能交互,提升政务处理效率和质量,促进辅助决策科学化、精准化、高效化.目前,新加坡和柬埔寨政府已经开始在政府办公中使用 ChatGPT 协助办公<sup>[34-35]</sup>;另一方面,大语言模型技术也可以成为政府面向公众的窗口,增强政务沟通和互动的效果,为公众提供快捷便民的信息服务,提升政府工作公开透明度以及公众的信任感.对于国际传播效能提升而言,面对复杂多变的国际格局和舆论形势,一方面,大语言模型技术可以赋能海量多源信息监测、整合和分析,助力国家形象全面分析和舆情事件快速定位、恶意信息精准判别等任务;另一方面,大语言模型技术强大的内容生成能力将极大丰富国际传播的内容和形式,促进国际传播的精准化和故事化,再结合机器内容生产的强大能效,可以有效提升国际内容供给效率和质量,成为国际舆论场中“讲好中国故事,传播好中国声音”,提升国际话语权的有力帮手.除此之外,除了作为信息收集和生产的端口,大语言模型高度智能化的表现为其成为社会模拟与仿真中的智能代理提供基础.斯坦福大学团队利用 ChatGPT 扮演可信人类行为体,成功模拟出人类社会中的信息传播、举办集体活动等社会现象.这对于未来国家和政府利用大语言模型模拟和预测政策实施效果、外宣与内宣效果等大有裨益.

## 2.2 社会挑战

尽管大语言模型技术在各领域表现出了广泛的应用前景,但仍存在一些问题.如同以往任何一项新技术从研发至社会应用阶段所需经历的碰撞和适应过程,目前大语言模型技术在社会化落地应用过程中可以预见4个问题.

1)信息污染问题.尽管以 ChatGPT 为代表的大模型已经初具通用人工智能的能力,但现阶段仍存在一些生成信息与事实不符的现象,尤其是在涉及专业知识的领域,如法律和医疗等.除此之外,尽管 ChatGPT 内部已经设定了部分内容质量约束规则,但经过适当诱导,仍可能生成一些有害信息.就技术扩散层面而言,著名传播学者罗杰斯在其力作《创新的扩散》中曾指出,技术的“可靠性”对一项技术是否会被公众广泛接受和被社会推广具有重要作用<sup>[36]</sup>.因此,大语

言模型技术存在的失实和有害信息问题也将成为制约其广泛推广的关键因素之一.就社会影响层面而言,这些廉价生成的失实和有害信息一旦被信任和广泛传播,将严重污染网络信息环境.2022年底中华人民共和国国家互联网信息办公室、中华人民共和国工业和信息化部、中华人民共和国公安部联合发布《互联网信息服务深度合成管理规定》以及2023年中华人民共和国国家互联网信息办公室发布的《生成式人工智能服务管理办法(征求意见稿)》明确表明国家对此问题的高度重视,并正在付诸积极实践.不过伴随技术的不断演进,信息的生产成本将大幅度降低,而拟真度不断提升,信息的管理和规制将成为一个愈发棘手和长期的议题.

2)社会权力分配问题.从长远来看,ChatGPT 相关技术的出现必将导致部分重复度高、目标相对确定、低创造力的工作岗位收缩或被代替.联合国贸易和发展组织也针对此问题发表《人工智能聊天机器人 ChatGPT 如何影响工作就业》文章,谈及“与大多数影响工作场所的技术革命一样,聊天机器人有可能带来赢家和输家,并将影响蓝领和白领工人”.OpenAI 进一步将此问题量化,提出未来约80%的美国劳动力至少有10%的工作任务可能受到GPT的影响,而约19%的劳动力至少有50%的任务将受到影响<sup>[26]</sup>.与此同时,我们也应意识到,ChatGPT 相关技术的研发与使用也存在一定程度的算力和数据等资源门槛,具备相关资源支持的组织和个人将成为这场技术革命中的极大受益者,而资源相对缺乏的组织与个人则处于相对被动地位.若不加以干预,可能的结果是社会层面“领导权集中在少数人和机构手中”,国际层面“社会中最有效的机器可能在国际上由少数超级大国控制”<sup>[33]</sup>.从历史观之,历史上的每一次技术革命都伴随着社会分工和权力的再分配,甚至可能激化社会矛盾、引发社会动荡,典型的如19世纪工业革命时期英国纺织女工为了反对纺织工业化而爆发的“卢德运动”.因此,大语言模型技术在社会化应用过程中所导致的社会权力和分工调整以及可能引发的潜在社会矛盾,仍然是需要前瞻性思考和妥善解决的重要问题.

3)伦理和法制问题.大语言模型作为具有强大内容学习和生产能力的新技术实体,不同于在既有法律和道德约束下的人类,其生成内容在使用过程中将面临知识产权、隐私、学术诚信、侵权责任认定等一系列伦理和法制隐患.目前,此问题已引发全世界范围的关注.Nature、Science等顶级期刊均表示不

能将 ChatGPT 作为论文合著者,以保护独创的科研成果。美国知名问答社区 Stack Overflow、Reddit 等也表示禁止在回答中使用 ChatGPT 生成的内容。我国目前还没有学术期刊或者企业机构明确禁止 ChatGPT 类大模型的使用,但在第十四届全国人大一次会议的“部长通道”中,科技部部长重点强调了有关 ChatGPT 的科技伦理问题,指出要“趋利避害”<sup>[37]</sup>。实际上,我国自 2019 年以后,陆续发布了《新一代人工智能治理原则——发展负责任的人工智能》《全球数据安全倡议》《新一代人工智能伦理规范》等规范或倡议,高度关注人工智能技术的伦理规制问题。在此背景下,大语言模型技术所面临的伦理与法制问题成为其迈向社会广泛应用之前亟待规制的重要问题。

4) 意识形态安全问题。ChatGPT 的智能表现主要来源于对海量数据和人类反馈的学习,因此,数据和人类反馈中的意识形态偏见也将被吸收到模型中。目前主流 ChatGPT 类大语言模型仍主要由美国企业和机构训练所得,主要训练语言来源也是英语,那么语言中所存在的意识形态偏见或者训练中人为恶意引入的预设偏见立场,将直接沿袭至大语言模型所生成的信息中,引发意识形态安全隐患。已有研究表明大语言模型存在对穆斯林的暴力偏见<sup>[38]</sup>,对中国形象的事实性和政治倾向性错误<sup>[39]</sup>。同时,未来 ChatGPT 的广泛应用也可能存在通过与国内用户交互,收集用户数据,威胁国家数据安全风险。再者,大语言模型也可被国外政府组织或敌对势力用作舆论战和信息战的武器,在国内外舆论场中散播大量不实和不利言论,营造虚假舆论气候,威胁意识形态安全。基于此,ChatGPT 类大语言模型所具备的潜在瓦解现有意识形态的“激进”力量也是需要警惕的重要问题。

### 3 结论与启示

ChatGPT 的出现,在技术意义层面预示着大语言模型技术进入又一崭新阶段,同时为通用人工智能的发展又迎来一次曙光;在社会意义层面,其也是智能化语言理解和生成技术真正面向社会可实用的一次跃进。未来伴随大语言模型技术的各类社会应用不断诞生,其与社会发展的关系也将更加紧密和富有张力,如同我们现今紧密依赖的互联网技术与时刻处于反思与规制中的互联网社会。本文最后面向未来大语言模型技术的发展及其与社会的互动,进行一定的总结与思考。

从大语言模型技术发展的角度而言,短期来看,

更强大的基础模型和更高质量的数据仍然是关键。长久以来,理性主义与经验主义一直是计算语言学研究的两大范式,前者强调让计算机内省总结语法规则与知识而精确理解语义,而后者强调通过对外界数据自动学习进而理解语义。回顾大语言模型的发展路径——基于大规模神经网络模型在海量数据中习得语言知识,经验主义是其现阶段取得优异表现的主要范式。在这一点上,目前国内相对于国外仍有一定差距。这也启示我们加快基础语言模型研发,加强数据基建,规范数据治理,尽快实现同等 ChatGPT 级别基础模型的自主可控。长期来看,理性主义与经验主义的结合或成为大语言模型迈向通用人工智能的路径选择。如同马克思<sup>[40]</sup>所言,“技术是人的手创造出来的人脑的器官”“是对象化的知识力量”,大语言模型未来将可能成为人类的外接通用智能大脑,服务于人类与外部信息交互,比如获取、处理和理解已有知识库,操纵各类工具,进而综合推理生成服务于人类的“新”知识,完成各类个性化任务。同时,现阶段的大语言模型无论是 ChatGPT 亦或是 GPT-4,主要基于语言、视频、图像等静态数据训练,未来的大语言模型可以进一步建模人类的行为动态数据,帮助人类完成各项任务,真正打通虚拟世界与现实世界之门。在此之外,可靠的负责任的大语言模型也是未来技术发展并迈向社会化应用的必经之路。GPT-4 的研发历程也向我们证明模型的可靠性仍是一个极为重要但又棘手的问题:GPT-4 在 2022 年 8 月训练完成之后,相关研发人员需使用 6 个月的时间用于提升模型的安全性和可靠性<sup>[4]</sup>。

从大语言模型技术发展与社会发展的关系来看,不管是从社会现实需要出发,还是从大语言模型的技术能力出发,大语言模型技术都成为无法忽视的一股潜在的社会变革力量。现如今社会中所出现的“禁止使用 ChatGPT”的声音和措施只是一种暂时状态——在还无法预估和承担风险下做出的暂时决策。美国未来学学者保罗·萨佛曾言“历史表明,一旦消费者察觉到某项新技术有用并且承受得起,广泛采用的情况就会发生的相当快”<sup>[41]</sup>。长久之道应是以疏代堵,采取制度治理与技术治理相结合的路径:以道德和法制引导人类如何更加合理合规地研发和使用技术,如推动技术专家与伦理和法制专家合作,健全大语言模型领域相关研发和使用制度规范;以创新技术约束机器如何更加可靠、负责任地协助人类,如建立大语言模型相关伦理与意识形态安全风险评测体系,建设模型内部伦理约束机制与外部恶意攻击识



别和防御系统,以此推动技术与社会的和谐可持续发展.与此同时,针对大语言模型技术未来逐渐替代部分人类工作的可能问题,我们也应思考自身如何更好地与之共存以及社会可以赋予的调节力量.就社会层面而言,一方面,基础模型研发人才和各垂直领域专业人才的培养,以及现有人力队伍的智能应用技能培训,是推动人工智能基建和各垂直领域产业的智能化转型的关键;另一方面,培养与智能技术互补的专业人才队伍,如能解决复杂问题、具备批判思维和创造力的人才,促进人力队伍转型则是面向未来人机共存社会的前置储备.就个人层面而言,历史经验表明人类总会在技术与社会的相互适应中找到自身的价值所在.我们在拥抱大模型技术的同时,也应在与技术的互动中发挥人的主体性位置,做技术的创新者、机器规则的制定者、社会问题的反思者.

**作者贡献声明:** 陈慧敏负责文章选题、思路设计及文章撰写工作;刘知远和孙茂松负责思路设计,提出指导意见并修改论文.

## 参 考 文 献

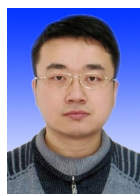
- [1] Wodecki B. ChatGPT may be the fastest growing app of all Time, [EB/OL]. [2023-07-28].<https://aibusiness.com/nlp/ubs-chatgpt-is-the-fastest-growing-app-of-all-time>
- [2] Reuters. Microsoft co-founder Bill Gates: ChatGPT will change our world[EB/OL]. [2023-07-28].<https://www.reuters.com/technology/microsoft-co-founder-bill-gates-chatgpt-will-change-our-world-2023-02-10/>
- [3] Mollman S. Elon Musk's history with OpenAI—the maker of AI chatbot ChatGPT—as told by ChatGPT itself[EB/OL]. [2023-07-28].<https://fortune.com/2022/12/11/elon-musk-history-with-chatgpt-maker-openai-as-told-by-the-ai-chatbot-itself/>
- [4] Achiam J, Adler S, Agarwal S, et al. Gpt-4 technical report[J]. arXiv preprint, arXiv: 2303.08774, 2023
- [5] Mollman S. ChatGPT passed a Wharton MBA exam and it's still in its infancy. One professor is sounding the alarm [EB/OL]. [2023-07-28].<https://fortune.com/2023/01/21/chatgpt-passed-wharton-mba-exam-one-professor-is-sounding-alarm-artificial-intelligence/>
- [6] Kung T H, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models[J]. *PLOS Digital Health*, 2023, 2(2): e0000198
- [7] Dreibelbis E. ChatGPT passes Google coding interview for level 3 engineer with \$183K salary[EB/OL]. [2023-07-28].<https://www.pcmag.com/news/chatgpt-passes-google-coding-interview-for-level-3-engineer-with-183k-salary>
- [8] He-rim J. Naver to introduce search GPT in first half of year [EB/OL]. [2023-07-28].<https://m.koreaherald.com/view.php?ud=20230203000553>
- [9] Tass. Yandex plans to develop alternative to ChatGPT neural network [EB/OL]. [2023-07-28].<https://tass.com/economy/1570119>
- [10] Xinhua. Beijing takes the lead in stating "supporting leading enterprises to create large-scale models benchmarking ChatGPT" [EB/OL]. [2023-07-28]. [http://www.news.cn/fortune/2023-02/15/c\\_1211728906.htm](http://www.news.cn/fortune/2023-02/15/c_1211728906.htm) (in Chinese)  
(新华网. 北京率先表态“支持头部企业打造对标 ChatGPT 的大模型”[EB/OL]. [2023-07-28]. [http://www.news.cn/fortune/2023-02/15/c\\_1211728906.htm](http://www.news.cn/fortune/2023-02/15/c_1211728906.htm))
- [11] Kenton J D M W C, Toutanova L K. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proc of Annual Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2019: 4171–4186
- [12] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 1877–1901
- [13] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proc of the 31st Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2017: 6000–6010
- [14] Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners[J]. *OpenAI Blog*, 2019, 1(8): 1–9
- [15] Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer[J]. *The Journal of Machine Learning Research*, 2020, 21(1): 5485–5551
- [16] Hao K. A college kid's fake, AI-generated blog fooled tens of thousands. This is how he made it [EB/OL]. [2023-07-28].<https://www.technologyreview.com/2020/08/14/1006780/ai-gpt-3-fake-blog-reached-top-of-hacker-news/>
- [17] Marx, Engels. Marx/Engels Collected Works vol. 3[M]. Beijing: People's Publishing House, 2009a (in Chinese)  
(马克思, 恩格斯. 马克思恩格斯选集(第3卷). 北京: 人民出版社, 2009a)
- [18] Winston B. Media Technology and Society: A History: From the Telegraph to the Internet[M]. London: Psychology Press, 1998
- [19] Bommasani R, Hudson D A, Adeli E, et al. On the opportunities and risks of foundation models[J]. arXiv preprint, arXiv: 2108.07258, 2021
- [20] Weidinger L, Mellor J, Rauh M, et al. Ethical and social risks of harm from language models[J]. arXiv preprint, arXiv: 2112.04359, 2021
- [21] Zhang Zhenyu, Hong Huaqing. Foreign language teaching supported by ChatGPT: Empowerment, issues and strategies[J]. *Foreign Language World*, 2023(2) 38–44 (in Chinese)  
(张震宇, 洪化清. ChatGPT 支持的外语教学: 赋能、问题与策略[J]. *外语界*, 2023(2): 38–44)
- [22] Qu Xing, Yang Jinming, Chen Tao, et al. Reflections on the change of medical education model by ChatGPT[J]. *Journal of Sichuan University (Medical Science version)*, 2023, 54(5): 937–940 (in Chinese)  
(瞿星, 杨金铭, 陈滔, 等. ChatGPT 对医学教育模式改变的思考[J]. *四川大学学报: 医学版*, 2023, 54(5): 937–940)
- [23] Zheng Manning. The news industry under the technology of artificial intelligence: Transformation, turning and response-new thoughts

- based on ChatGPT[J]. Chinese Editors, 2023, 160(4): 35–40 (in Chinese)  
(郑满宁. 人工智能技术下的新闻业: 嬗变、转向与应对——基于 ChatGPT 带来的新思考[J]. 中国编辑, 2023, 160(4): 35–40)
- [24] Negroponte N, Harrington R, McKay S R, et al. Being digital[J]. *Computers in Physics*, 1997, 11(3): 261–262
- [25] Kelly S. Microsoft opens up its AI-powered Bing to all users[EB/OL]. [2023-07-28]. [https://www.cnn.com/2023/05/04/tech/microsoft-bing-updates?cid=external-feeds\\_iluminar\\_google](https://www.cnn.com/2023/05/04/tech/microsoft-bing-updates?cid=external-feeds_iluminar_google)
- [26] 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
- [27] Feldman T. An Introduction to Digital Media[M]. London: Psychology Press, 1997
- [28] Van Noorden R. ChatGPT-like AIs are coming to major science search engines[J]. *Nature*, 2023, 620(7973): 258–258
- [29] Heidt A. Artificial-intelligence search engines wrangle academic literature[J]. *Nature*, 2023, 620(7973): 456–457
- [30] Lam R, Sanchez-Gonzalez A, Willson M, et al. Learning skillful medium-range global weather forecasting[J]. *Science*, 2023, 382(6677): 1416–1421
- [31] Park J S, O'Brien J, Cai C J, et al. Generative agents: Interactive simulacra of human behavior[C]//Proc of the 36th Annual ACM Symp on User Interface Software and Technology. New York: ACM, 2023: 1–22
- [32] Williams T. Some companies are already replacing workers with ChatGPT, despite warnings it shouldn't be relied on for "anything important" [EB/OL]. [2023-07-28]. <https://fortune.com/2023/02/25/companies-replacing-workers-chatgpt-ai/>
- [33] Kissinger H, Schmidt E, Huttenlocher D. ChatGPT heralds an intellectual revolution [EB/OL]. [2023-07-28]. <https://www.wsj.com/articles/chatgpt-heralds-an-intellectual-revolution-enlightenment-artificial-intelligence-homo-technicus-technology-cognition-morality-philosophy-774331c6>
- [34] Toa Payoh North. Singapore: ChatGPT to help civil servants[EB/OL]. [2023-07-28]. <https://psnews.com.au/2023/02/20/singapore-chatgpt-to-help-civil-servants/>
- [35] Latsamy Phonevilay. Cambodian Ministry Plans Khmer Version of ChatGPT[EB/OL]. [2023-07-28]. <https://laotiantimes.com/2023/03/14/cambodian-ministry-plans-khmer-version-of-chatgpt/>
- [36] Rogers E M. Diffusion of Innovations[M]. Manhattan: Simon and Schuster, 2010
- [37] People's Daily. Minister of Science and Technology explains ChatGPT using Messi as an analogy[EB/OL]. [2023-07-28]. <https://wap.peopleapp.com/article/7023220/6879050> (in Chinese)  
(人民日报. 科技部部长用梅西解释 ChatGPT[EB/OL]. [2023-07-28]. <https://wap.peopleapp.com/article/7023220/6879050>)
- [38] Abid A, Farooqi M, Zou J. Large language models associate Muslims with violence[J]. *Nature Machine Intelligence*, 2021, 3(6): 461–463
- [39] Fan Hong, He Jiayu. The image of China in the perspective of ChatGPT: Analysis and reflection[J]. *International Communications*, 2023, 319(4): 19–22 (in Chinese)  
(范红, 何佳雨. ChatGPT 视角下的中国国家形象图景: 分析与思辨[J]. 对外传播, 2023, 319(4): 19–22)
- [40] Marx, Engels. Marx/Engels Collected Works vol. 8[M]. Beijing: People's Publishing House, 2009b (in Chinese)  
(马克思, 恩格斯. 马克思恩格斯选集(第 8 卷). 北京: 人民出版社, 2009b)
- [41] Saffo P. Paul Saffo and the 30-year rule[J]. *Design World*, 1992, 24(1): 16–23



**Chen Huimin**, born in 1994. Assistant professor at the School of Journalism and Communication, Tsinghua University. Received her PhD degree from the Department of Computer Science and Technology, Tsinghua University. Her main research interests include natural language processing and social computing.

陈慧敏, 1994 年生. 清华大学新闻与传播学院助理教授. 博士毕业于清华大学计算机科学与技术系. 主要研究方向为自然语言处理、社会计算.



**Liu Zhiyuan**, born in 1984. Associate professor at the Department of Computer Science and Technology, Tsinghua University. His main research interests include natural language processing and social computing.

刘知远, 1984 年生. 清华大学计算机科学与技术系副教授. 主要研究方向为自然语言处理、社会计算.



**Sun Maosong**, born in 1962. Professor at the Department of Computer Science and Technology, Tsinghua University. His main research interests include natural language processing, Chinese computing, Web intelligence, and computational social sciences.

孙茂松, 1962 年生. 清华大学计算机科学与技术系教授. 主要研究方向为自然语言处理、中文计算、Web 智能、计算社会科学.