



世界互联网大会
World Internet
Conference

发展负责任的 生成式人工智能

研究报告及共识文件

世界互联网大会人工智能工作组
2023年11月

更多免费行业报告

并购家

www.ipopo.cn

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

世界互联网大会人工智能工作组

按单位名称首字母排序

- | | |
|----------------------------------|--|
| - 阿里云智能集团
朱冉、穆飞、杨煜东 | - 安谋科技(中国)有限公司
武大伟、王骏超、吴彤 |
| - 阿拉伯信息通信技术组织
穆罕默德·本·阿莫 | - 北京百度网讯科技有限公司
马艳军、刘艳丽、王禹杰 |
| - 北京航空航天大学
赵精武 | - 北京三快科技有限公司
王金刚、陈政聿 |
| - 北京市商汤科技开发有限公司
田丰、胡正坤 | - 北京智谱华章科技有限公司
刘德兵、张焱 |
| - 北京智源人工智能研究院
林咏华、倪贤豪 | - 德国明斯特大学
伯纳德·霍尔兹纳格尔 |
| - 德国亚太企业管理咨询公司
费利克斯·里希特 | - 对外经济贸易大学
张欣 |
| - 伏羲智库
李晓东、付伟、杨晓波 | - 国际电信联盟电信标准化部门
Build n Blaze
雷蒂亚·方娜 |
| - 国际电信联盟电信标准化部门
杨晓雅 | - 国际电信联盟电信标准化部门
毗湿奴·拉姆 |
| - 国际商业机器(中国)有限公司
谢东、孟繁晶、程海旭 | - 广州市动悦信息技术有限公司
蒋冠军、张沅 |
| - 华为技术有限公司
王震军、柳嘉琪 | - 华为云计算技术有限公司
尤鹏、李寅、张倩 |
| - 华兴泛亚投资顾问(北京)有限公司
秦川、王力行、赵雨萍 | - 佳都科技集团股份有限公司
周志文、秦伟 |

世界互联网大会人工智能工作组

按单位名称首字母排序

- | | |
|---|-------------------------------|
| - 科大讯飞股份有限公司
王士进、李森、董乔琳 | - 蚂蚁科技集团股份有限公司
林冠辰、温祖杰、李亮 |
| - 欧洲标准化委员会和欧洲电工技术标准化委员会
网络安全和数据保护联合技术委员会
沃尔特·福美 | - 清华大学信息国家研究中心
顾心怡 |
| - 上海诺基亚贝尔股份有限公司
叶晨晖、肖卓然、常疆 | - 世界互联网大会
梁昊、张雪丽、康彦荣 |
| - 思科系统公司
侯胜利、魏航 | - 三六零安全科技股份有限公司
刘兆辉、刘闯、甄一蕴 |
| - 深圳市腾讯计算机系统有限公司
曹建峰、王梦寅、袁俊 | - 网易(杭州)科技有限公司
吕唐杰、马梦婕 |
| - 新浪公司
王巍、张俊林、王雪婷 | - 英国标准协会
克里斯·布朗 |
| - 英国亨利商学院
唐银山 | - 英特尔(中国)有限公司
王海宁、马里奥·罗马奥 |
| - 印象笔记(上海)科技有限公司
唐毅、岳峰、卜英 | - 之江实验室
林峰、刘哲 |
| - 中国科学院自动化研究所
曾毅 | - 中国社会科学院法学研究所
周辉 |
| - 中国社会科学院哲学所
段伟文 | - 中国信息通信研究院
魏凯、王蕴韬 |
| - 中国政法大学
张凌寒 | - 中兴通讯股份有限公司
孟伟、袁丽雅 |

编辑

中国信息通信研究院

呼娜英、石霖、刘硕、邹皓、孙小童、郭苏敏、朱凌云

工作组联系方式: research@wicinternet.org

缩略语对照表

缩略语	英文名称	中文名称
3D	Three Dimensional	三维
AI	Artificial Intelligence	人工智能
API	Application Programming Interface	应用程序接口
AR	Augmented Reality	增强现实
BLIP	Bootstrapping Language-Image Pre-training	自引导语言图像预训练
BLOOM	BigScience Large Open-science Open-access Multilingual Language Model	大科学、大型、开放科学、开源的多语种语言模型
ChatGLM	Chat Generative Language Model	千亿基座的对话模型
CPU	Central Processing Unit	中央处理器
DALL-E	Dali WALL-E	达利和瓦力 (模型)
Emu	Large Multimodal model of BAAI	智源统一多模态预训练模型
FPGA	Field-Programmable Gate Array	可编程逻辑门阵列
GenAI	Generative Artificial Intelligence	生成式人工智能
GDP	Gross Domestic Product	国内生产总值
GMV	Gross Merchandise Volume	商品交易总额
GPT	Generative Pretrained Transformer	生成式预训练变换器
GPU	Graphics Processing Unit	图形处理器
InternLM	Intern Language Model	书生·浦语大模型
LLaMA	Large Language Model Meta AI	Meta人工智能大语言模型
MTP	Massive Text Pairs	大规模文本数据集
NLP	Natural Language Processing	自然语言处理
NPU	Neural Processing Unit	数字处理单元
PaLM	Pathways Language Model	路径语言模型
Qwen	Qian Wen	通义千问大模型
SaaS	Software as a Service	软件即服务
SDK	Software Development Kit	软件开发工具包
TPU	Tensor Processing Unit	张量处理单元
VLA	Vision Language Action	视觉-语言-动作
VLM	Vision Language Model	视觉-语言模型

目 录

一、概述	01
二、全球生成式人工智能技术发展态势	02
三、生成式人工智能带来的机遇	05
四、生成式人工智能引发的挑战	08
五、全球为发展负责任生成式人工智能的努力	11
六、发展负责任的生成式人工智能共识	15
附件：发展负责任的生成式人工智能的行业应用探索	18



01/ 概述

OVERVIEW

近年来,生成式人工智能不断取得突破,展现出强大的生成创造能力,开始涌现出“智慧”。生成式人工智能在文本、代码、图像、音视频等方面的理解与生成取得了突破性进展,有望大幅提升社会生产力,加速千行百业的数字化进程,促进人类社会全面迈向智能化新阶段。

回顾人工智能60余年的发展历程,技术突破不仅会创造发展机遇,也会带来相应的挑战。统筹人工智能发展和治理逐渐成为全球共识,自2016年以来,全球多个国际组织、国家、地区及产业界,积极探索人工智能发展与治理路径,已经形成了系列共识原则、

治理要求、实践范式等。考虑到人工智能尚处在快速发展的过程中,相关工作仍需要持续推进。

生成式人工智能演进速度之快、赋能范围之广、影响程度之深前所未有。以负责任的态度推动生成式人工智能发展不仅十分必要,也愈发紧迫,是事关人工智能乃至人类文明发展的重要命题。为此,世界互联网大会成立人工智能工作组,广泛汇聚各方智慧形成共识,推动生成式人工智能发展与治理协同共进,增进全人类共同福祉。

02/ 全球生成式人工智能技术发展态势

GLOBAL DEVELOPMENT TRENDS OF
GENAI TECHNOLOGIES

(一) “模型、数据、算力”三大要素的演进带动人工智能不断突破

生成式人工智能技术突飞猛进,展现出惊人的创造能力和生成能力,主要得益于**模型、数据、算力**等方面的不断提升。

模型层面,模型结构的创新和模型规模的提升成为生成式人工智能取得突破的关键。从模型结构来看,注意力机制、自回归模型、扩散模型等技术不断升级迭代,特别是以Transformer为主的基础模型脱颖而出,成为生成模型主流技术路线,推动文本、图像、音频、视频等内容的生成和理解能力不断提高。涌现

出ChatGPT、文心一言等大语言模型,Stable Diffusion、DALL-E2、DALL-E3等视觉生成模型,以及GPT-4、BLIP-2、Emu等多模态模型。**从模型参数规模来看**,上述新模型架构使得参数规模不断增大成为可能,带来模型能力质的飞跃。以GPT系列模型为例,2020年发布的GPT-3参数规模有1750亿,相比于2018年发布的参数规模为1.17亿的GPT-1,在复杂自然语言处理方面实现了显著提升。此外,围绕基础模型衍生出的插件机制,可以将外部的搜索、数据处理等功能与基础模型能力集成,从而进一步丰富模型功能,拓展应用范围。OpenAI、360、百度、华为、科大讯飞等企业均推出了相应的模型插件,例如文心一言上线的搜索、交互等插件,使模型更容易实现功能的扩展和定制,以适应多种场景的需求。

数据层面,数据质量、多样性、规模等方面的进步成为人工智能能力提升的基础。被广泛用于大模型预训练的The Pile数据集,主要基于学术或专业领域知识构造,具有较高质量,包含了维基百科、书籍、

期刊、Reddit链接、Common Crawl等20余个数据集¹。北京智源人工智能研究院发布的大规模文本数据集 MTP，范围涉及搜索、社区问答、百科常识、科技文献等，数据规模达到3亿对。Anthropic、斯坦福大学、Hugging Face 等单位发布的微调数据集，涵盖了多种类型的指令，有助于提升模型的可控性，使模型更好地理解并遵循人类指令。此外，**合成数据可能成为高质量数据的重要来源之一**。生成式人工智能能够大批量制作拟真合成数据，或将帮助缓解高质量训练数据枯竭这一未来潜在问题。根据Gartner预测，到2024年，60%用于人工智能开发和分析的数据将会是合成数据；到2030年合成数据将取代真实数据，成为人工智能模型所使用数据的主要来源²。

算力层面，算力设施的完善支撑生成式人工智能的快速发展。人工智能芯片提供算力基础保障，GPU、FPGA、NPU、TPU 等不同技术路线芯片持续探索，针对人工智能计算不断优化，为模型的训练与推理提供了基础保障。**深度学习框架放大芯片算力效能**，一是通过提供高性能的大规模分布式训练与推理技术，有效缓解模型训练耗时长、推理算力需求高等问题。二是通过与底层芯片适配优化，充分发挥硬件性能，提高计算效率。**云边端多样化算力满足生成式人工智能不同应用需求**，云侧强大的计算和存储能力保障大模型训练以及高吞吐量应用的推理任务；边缘算力将海量复杂数据进行本地化预处理，可对数据进行实时处理并将其导向大模型，实现快速响应和决策；端侧算力减少数据处理和传输的延迟，直接在端侧进行数据计算分析，提升智能应用的实时性。

(二) 开源开放驱动生成式人工智能生态渐趋繁荣

模型开源促进技术的发展和普及。以LLaMA 2、

BLOOM、ChatGLM、Baichuan、Aquila、InternLM、Qwen等为代表的开源模型层出不穷，并且不断升级进化。**在模型迭代优化方面**，模型开源的兴起扩大了企业对基础模型和微调模型的选择范围，目前大量创业公司使用LLaMA 2、Stable Diffusion等开源模型调优并推出新产品。**在研发门槛降低方面**，应用开源模型具有规避初始高昂投资、私有数据的完全控制、可自我迭代优化等优势。开发者基于开源模型，可快速搭建具备专业领域知识的垂类任务模型，大幅缩减了模型从开发到应用所需的算力、数据和时间成本。例如，开源平台Github上显示，基于智谱AI开发的ChatGLM开源模型，大幅降低了研发门槛，有11个模型脱颖而出，覆盖医疗、法律、金融、教育等多个领域³。

开放接口为开发者提供便捷。除了模型开源，开放易用的API和SDK也是促进人工智能生态繁荣的重要一环。**一方面，接口开放将简化开发流程并提升效率。**开放接口帮助开发者无需从头开始编写算法或模型，大大简化开发流程，减少开发时间和工作量。例如，通过调用GPT-3.5-Turbo模型API开放接口，仅需少量Python代码就可实现代码生成、对话代理、语言翻译、辅助学习等复杂功能。**另一方面，接口开放可以丰富模型的应用场景。**接口开放可以帮助广大开发者更便捷地接入模型能力，形成更加多样化的应用场景。例如，百度文心一言提供的接口可以应用于搜索、推荐、对话等场景，提升应用效果和用户体验。

开发者社区持续推动技术扩散。开发者社区通过提供免费算力、课程教材、公开数据集和模型套件等工具组件，赋能培养具备模型开发能力的人才，对于推动人工智能领域的技术交流和 development 起到了积极的促进作用。例如，Hugging Face提供了一键式的预训练模型调用功能，提供了大量预训练模型、简单的API和丰富的文档，以及活跃的社区论坛，加快了技术扩

¹ Leo Gao, Stella Biderman, et al. The Pile: An 800GB Dataset of Diverse Text for Language Modeling, Dec. 2020, p1.

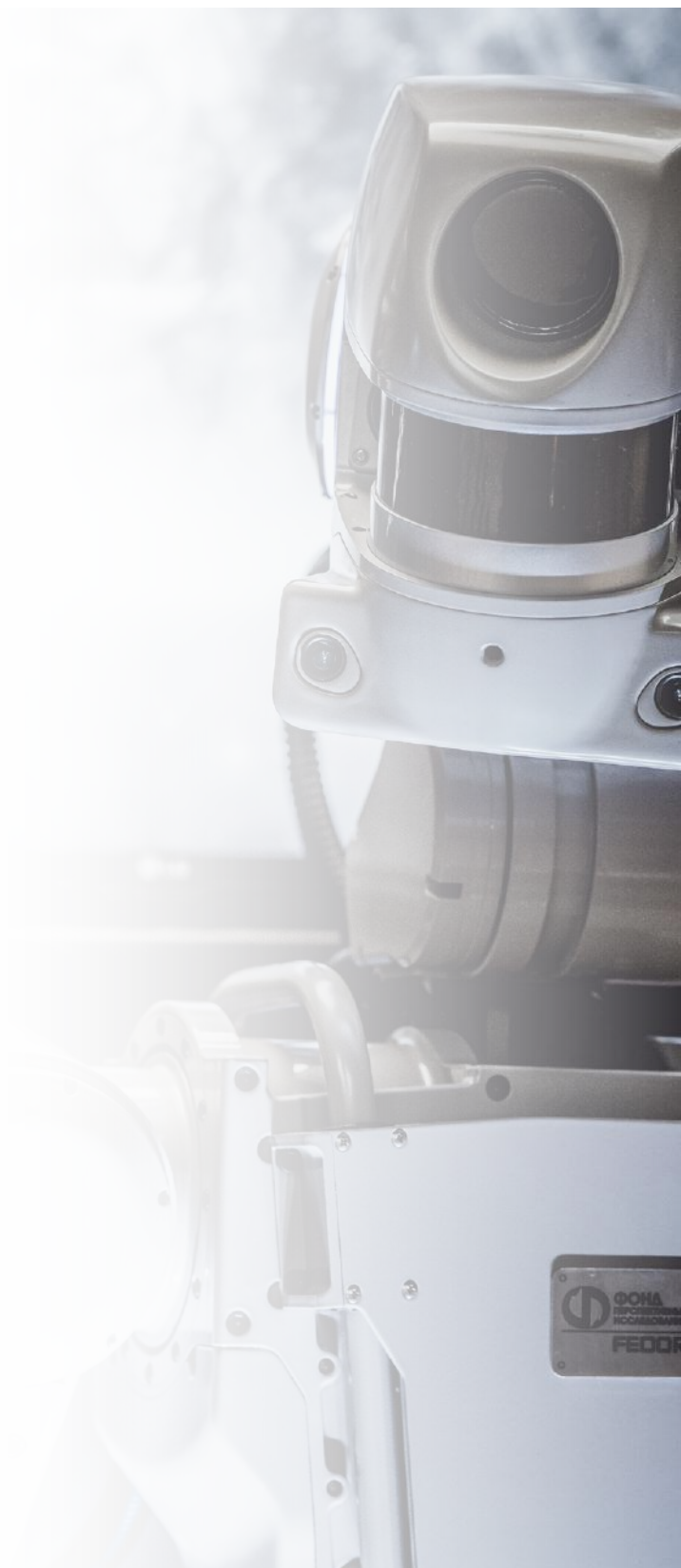
² Gartner, 'Maverick Research: Forget About Your Real Data - Synthetic Data Is the Future of AI' (24 June 2023), <<https://www.gartner.com/en/documents/4002912>> accessed 19 September 2023.

³ GitHub, 'Awesome-Chinese-LLM', <<https://github.com/HqWu-HITCS/Awesome-Chinese-LLM>> accessed 30 October 2023.

散。华为云AI Gallery百模千态社区构建了一站式AI社区服务平台，助力企业和开发者快速创建模型应用。阿里巴巴魔搭社区开放的在线预训练模型，可以在无需开发代码的情况下体验各种模型效果。FlagOpen飞智集合了大模型的算法、模型、工具、评测等多个模块，打造了大模型“Linux”开源开放技术体系。百度的飞桨星河社区提供开放数据、开源算法、免费算力，提供一体化大模型开发体系，助力开发者的大模型探索之旅。

(三) 生成式人工智能发展凸显通用人工智能曙光

生成式人工智能的突破加快了通用人工智能的探索步伐。生成式人工智能不仅能够处理单一数据类型的任务，而且可以在不同数据类型间建立联系和融合，向着多模态方向发展。**多模态生成模型的突破显著提高机器智能的拟人性和通用性。**AI Agent伴随着多模态生成模型技术的突破，能够更好地理解和处理复杂的现实场景，从而为人类提供更为精准、个性化的服务。**多模态生成模型与智能体的结合带来更多可能性。**具身智能将多模态生成模型与机器人技术结合，通过模仿人类学习来感知复杂的世界，实现“感官”（硬件）与“思考”（软件）的多模态融合，协助人类完成各种任务。例如，谷歌发布的Robotic Transformer2 (RT2) 作为视觉-语言-动作 (VLA) 模型，将视觉-语言模型 (VLM) 预训练与机器人数据相结合，直接控制机器人，使其在真实世界中执行各种任务。



03/ 生成式人工智能带来的机遇

OPPORTUNITIES BROUGHT BY GENAI

(一) 促进经济发展

生成式人工智能推动生产力大幅提升, 带动全球经济增长。根据麦肯锡2023年6月预测, 生成式人工智能每年可能为全球经济增加2.6万亿至4.4万亿美元的价值⁴。根据高盛研究, 生成式人工智能的突破将在10年内推动全球GDP增长7%⁵。当前, 生成式人工智能带来的生产力升级正在推动生产方式的进化, 为全球经济发展打开新的机会之窗。类似于传统AI, 生成式人工智能展现自动化优势, 一方面, AI通过自动化实现对人类部分工作的替代, 提高经济生产效率⁶, 另一方面, AI通过自动化对前沿技术进行模仿与创

新, 加速技术研发与扩散⁷。区别于传统AI, 生成式人工智能具有实现通用性的潜力, 预示着应用领域的AI互相统一协同, 从而会在社会经济活动的各个领域发挥更大作用⁸。生成式人工智能将优化生产流程、管理方式、营销策划等环节, 推动传统生产方式升级。

生成式人工智能对产业带来深远影响, 对不同性质行业影响次序不一。随着生成式人工智能与各行各业深度融合, 其赋能重构的行业将会持续增加。根据罗兰贝格的评估分析, 生成式人工智能将率先对互联网与高科技、金融和专业服务等知识密集型行业带来较大影响, 分别带来6.5%、6.8%、11.3%的成本下降; 其次将赋能教育、通信、医疗、公共服务、零售、文娱和传媒等服务型行业; 对当前数字化程度不高的农业、材料、建筑业、能源等传统行业影响相对较小。总的来看, 生成式人工智能的价值发挥需要坚实的信息化、数字化支撑, 有望在相关行业的研发设计、生产制造、运营管理方面创造巨大价值⁹。

⁴ McKinsey, 'The economic potential of generative AI: The next productivity frontier' (14 June 2023), <<https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>> accessed 17 September 2023.

⁵ Goldman Sachs Research, 'Generative AI could raise global GDP by 7%' (05 April 2023), <<https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>> accessed 17 September 2023.

⁶ Agrawal, Ajay K., et al. Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth (April 2018). NBER Working Paper No. w24541, p.26.

⁷ WEBB M, The Impact of Artificial Intelligence on the Labor Market (Rochester 2019).

⁸ Partha Pratim RAY, ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope (April 2023), Internet of Things and Cyber-Physical Systems vol.3, p.121-p.154.

⁹ Roland Berger, 《通用人工智能的曙光: 生成式人工智能技术的产业影响》, 2023年8月, 第3页。

生成式人工智能深入赋能数字经济，为各行业领域带来新一轮发展机遇。伴随着生成式人工智能影响规模的不断扩大，赋能各行各业实现数字化变革与发展。**金融业领域**，生成式人工智能能够帮助绘制金融风险图，协助打击洗钱等金融犯罪¹⁰。**汽车业领域**，生成式人工智能能够提高车载智能语音交互效率，还能为自动驾驶模型训练提供高质量合成数据，帮助解决自动驾驶系统开发过程中的数据和测试难题¹¹。更进一步，多模态生成模型正有望加速推动“多模态感知到决策规划”的端到端自动驾驶落地应用。**传媒业领域**，生成式人工智能可以根据文本提示生成文字、图片、音频、视频等，为广告配上引人入胜的视觉内容。**制造业领域**，生成式人工智能可以应用于机器视觉、数位分身和自主导航系统等，实现生产线和仓储物流等环节的无人化和智能化。**农业领域**，生成式人工智能可以通过遥感大模型测量农作物的长势¹²，监测作物病害，预测农作物产量。生成式人工智能的进步性价值将持续推动各行业领域质量变革、效率变革、动力变革，推动经济高质量发展。

(二) 促进社会进步

生成式人工智能可以有效优化城市运营服务。生成式人工智能能够通过持续分析数据来标记城市运营问题并优化策略¹³。例如，可通过分析能源网格数据生成更优的能源使用策略。还能够识别不同灾害类型的潜在模式和规律，通过实时数据的监测和分析进

行灾害预警，灾害发生后可以通过分析事故和灾害数据，辅助进行灾后评估、资源调配、救援规划等工作，提供实时预警和全面的风险管理解决方案¹⁴。

生成式人工智能可以对教育产生有益影响。从提供发展机遇角度而言，生成式人工智能可以在信息检索、学习规划、协同创作等方面发挥优势¹⁵，协助提升教育教学质量¹⁶。从促进教育公平角度而言，生成式人工智能可以被用于建设开放的教育平台，为不同地区学生提供平等资源，帮助弥合教育鸿沟。

生成式人工智能可以促进就业市场扩容提质。一方面，生成式人工智能能够调整就业结构，增加语言模型训练师、事实核查员、机器人检测工程师、数据标注师等新兴职业。另一方面，利用生成式人工智能生成文档、处理数据、展示图表、团队协作，能够提升工作效能。根据全球劳工组织（ILO）的报告，全球预计4.27亿份工作可借助生成式人工智能技术得到优化或效率提升。

生成式人工智能可以带动智慧医疗、智慧养老发展。在智慧医疗方面，生成式人工智能应用于分诊导诊、医学成像、临床诊断和个性化医疗干预等场景，有助提高许多临床及管理应用的效率及准确性，改善医疗服务和患者就医体验。在智慧养老方面，生成式人工智能通过实时监测分析被看护人的生理指标并标记异常情况，能够帮助看护人及时调整护理方案，惠及日益增长的养老护理需要。

¹⁰ Wharton school of University of Pennsylvania, 'Does Generative AI Solve the Financial Literacy Problem?' (27 June 2023), <<https://knowledge.wharton.upenn.edu/article/does-generative-ai-solve-the-financial-literacy-problem/>> accessed 18 September 2023.

¹¹ 中国企业网,《以 Chat GPT 为代表的生成式 AI 在自动驾驶领域的应用》(2023 年 4 月),来源:<http://www.zqcn.com.cn/757/27327.html>, 2023 年 9 月 27 日访问。

¹² 中国科技信息,《让 AI 下沉到田间地头 大模型开启遥感应用新篇章》(2023 年 5 月),来源:https://www.thepaper.cn/newsDetail_forward_23222846, 2023 年 9 月 27 日访问。

¹³ 千家网,《SmartCityGPT 生成式人工智能如何助力智慧城市发展》(2023 年 8 月),来源:<http://www.citnews.com.cn/news/202308/164399.html>, 2023 年 9 月 27 日访问。

¹⁴ Deloitte,《数实融合的 ChatGPT 时代,开拓智慧城市新模式》(2023 年 7 月),来源:<https://www2.deloitte.com/cn/zh/pages/technology-media-and-telecommunications/articles/2023-mwc-chatgpt.html>, 2023 年 9 月 27 日访问。

¹⁵ Orly Lobel, THE EQUALITY MACHINE: Harnessing Digital Technology for a Brighter, More Inclusive Future (Public Affairs 2019).

¹⁶ International Telecommunication Union, 'Measuring Digital Development: Facts and Figures 2021', <<https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2021.pdf>> accessed 15 September 2023.

(三) 助力公益事业

生成式人工智能帮助打造无障碍的数字环境。例如利用基于深度神经网络的语音合成技术,生成更接近真人的声音,提供更加丰富的有声读物。生成式人工智能的跨模态能力对残障人群能产生更大效用,字幕眼镜、手语数字人等应用能够帮助建设无障碍的工作环境。

生成式人工智能助力全球文化成果的保护和传播。生成式人工智能可以在对文物艺术品进行三维扫描和数字化储存基础上,进一步利用图像生成技术重现内容,增强沉浸式的文化体验;还可以利用语音识别、语言合成、虚拟人等技术生成方言主播,使方言传播更加生动,帮助保护方言等非物质文化遗产。

生成式人工智能助推全球环境治理和可持续发展。生成式人工智能可以通过分析污染水平、地表变化以及气候变化的数据,比较分析人类活动对环境影响的程度,从而提供监测、限制和解决措施的建议,为减缓气候变化,保护森林、海洋及野生动物栖息地的可持续性做出贡献。

(四) 助力科学研究

生成式人工智能推动科研效率不断提高。生成式人工智能开始在科学研究中发挥愈发重要的作用,通过人机协作大幅加快科学研究。生成式人工智能可以辅助进行材料收集、分析论证、论文撰写等学术研究的基础性工作,将科研人员从非核心的研究工作中解放出来。例如,生成式人工智能学术助手可以通过结论提炼、方法总结、观点总结、文献综述以及摘要汇总等功能,方便科研人员完成学术文献的结构化、系统化阅读,快速掌握研究基础、发现科研选题方向以及

设计研究方案¹⁹。

生成式人工智能助力不同研究方向纵深探索。生成式人工智能基于深度学习技术可以充分发挥神经网络高维逼近的强大优势,有望解决所谓的“维数灾难”²⁰,从而加速重大科学问题研究和知识发现。**在分子结构方向**,生成式人工智能可以加速从分子设计、反应设计到条件生成、反应检验等化学合成全链条,帮助提升潜在功能性分子(如药物分子)及其合成方案设计的速度²¹。**在动力学方向**,生成式人工智能可以用于流体粒子模型的计算与模拟,实现从外部视觉表现推理内部流体动态,并可反演粘度、密度等流体属性,在问题规模、模拟速度、模型泛化性、反问题求解精度等多个方面突破现有流体数值模拟方法的计算瓶颈。**在生物医药研发方向**,生成式人工智能支持跨模态自然交互,覆盖医药研发的流程场景,例如专家可以通过自然对话获取某一适应症的基本情况、药物研发进展等;支持生物医药多模态数据,例如输入分子式后可直接得到该分子的完整描述信息等²²。



¹⁹ 中国知网,《CNKI AI 学术研究助手介绍》(2023 年 8 月),来源:<http://yuanjian.cnki.com.cn>,2023 年 9 月 18 日访问。

²⁰ 维数灾难指随着变量的个数或者维数的增加,计算复杂度呈指数增加。见鄂维南院士于 2021 年世界互联网大会上的讲话,来源:https://www.thepaper.cn/newsDetail_forward_13498964,2023 年 10 月 12 日访问。

²¹ 人民日报《上海交大开源发布白玉兰科学大模型》(2023 年 7 月),来源:<https://news.sjtu.edu.cn/mtjj/20230710/185884.html>,2023 年 9 月 16 日访问。

²² 中国网科学,《水木分子发布生物医药行业千亿参数大模型,推出药研助手 ChatDD》(2023 年 9 月),来源:http://science.china.com.cn/2023-09/22/content_42532416.htm,2023 年 9 月 18 日访问。



04/ 生成式人工智能引发的挑战

CHALLENGES BROUGHT BY GENAI

(一) 技术内在风险引发安全隐患

生成式人工智能技术在迭代升级的同时也放大了技术安全风险。数据方面,数据投喂带来价值偏见、隐私泄露、数据污染等问题。一是训练数据固有偏见导致模型产生偏见内容。全球科研机构的多个实验发现经过人工标注的大模型在应用中存在性别歧视、种族歧视等偏见问题。例如,根据微软发布的GPT-4研究报告,大模型在生成职业性别描述时,会进一步扩大数据集的固有偏差,存在严重性别偏向²³。二是海量训练数据扩大了数据安全和隐私保护风险。训练大模型依赖庞大的数据,对于数据来源的合法性审查带

来了诸多挑战。此外,大模型也存在泄露用户输入数据风险。三星在启用ChatGPT的20天内就发生3起员工泄露数据事故,泄露内容包括半导体设备测量、良品率/缺陷、内部会议内容等敏感信息²⁴。三是人工智能生成数据将造成训练数据污染。如果使用生成式人工智能产生的数据作为语料训练生成式人工智能模型,可能会导致“模型崩溃”现象发生。剑桥大学学者指出,生成式人工智能在制造便利的同时也在摧毁互联网环境²⁵。

算法方面,算法模型生成特性及安全漏洞会引发“幻觉”(hallucination)或虚假信息、模型遭受攻击等风险。一是生成式模型模仿特性生成“幻觉”或虚假信息。生成式人工智能基于训练数据进行模仿而非理解的特性,可能生成错误的、不准确的、不真实的信息,即生成“幻觉”内容。美新闻卫士公司用“虚假信息数据库”中的100条提示词测试ChatGPT,其对80条提示词反馈了虚假和误导性信息²⁶。二是算法模型安全漏洞诱发网络攻击风险。攻

²³ Microsoft Research, 'Sparks of Artificial General Intelligence: Early experiments with GPT-4' (22 March 2023), <<https://arxiv.org/pdf/2303.12712v1.pdf>> accessed 16 October 2023.

²⁴ 涟漪效应,《拯救“失足AI”?ChatGPT的性别偏见与“喂养”伦理》(2023年2月),来源:<https://www.thepaper.cn/newsDetail_forward_22047029>, 2023年9月18日访问。

²⁵ Metaverse Post, 'Ross Anderson Discusses AI Model Collapse as a Growing Problem in Online Content' (14 June 2023), <<https://mpost.io/ross-anderson-discusses-ai-model-collapse-as-a-growing-problem-in-online-content/>> accessed 15 September 2023.

²⁶ NewsGuard, 'Despite OpenAI's Promises, the Company's New AI Tool Produces Misinformation More Frequently, and More Persuasively, than its Predecessor' (March 2023), <<https://www.newsguardtech.com/misinformation-monitor/march-2023/>> accessed 15 September 2023.

击者通过精心设计的输入操控,可能导致生成式人工智能后端系统被利用或被控制。甚至有黑客通过修改Web测试套件SilverBullet的配置,大规模窃取了ChatGPT帐户²⁷。OpenAI公司虽已发布最高奖励2万美元的“漏洞悬赏计划”,用于帮助发现并修补其产品的安全漏洞,但生成式人工智能恐将长期面临严峻的网络攻击风险。

此外,生成式人工智能的底层模型“根属性”诱发链路性风险。当前,生成式人工智能产业生态雏形初现,底层大模型成本高、投入重,形成了高技术壁垒和强竞争优势,其“根”基础设施属性得以牢固。中下游基于大模型底座开发部署的产品应用可能将固有风险进行链路性扩散。2023年2月,布鲁金斯学会报告指出,未参加生成式人工智能原始模型开发的“下游开发者”可能会将原始模型经过调整后整合到其他软件系统,由于双方均无法全面了解整个系统,或将增加这些软件错误和失控风险²⁸。

(二) 人机关系变化加深科技伦理失范

生成式人工智能重构人机关系可能带来科技伦理失范。生成式人工智能强大的任务处理能力,容易导致人的思维依赖。过度依靠生成式人工智能提供的答案,会使人自身的观察与理解、归纳与演绎、比较与推理等感知和逻辑能力缺乏训练,怠于思考与创新。例如,很多学生开始依赖于生成式人工智能完成作业,美国北密歇根大学的一位教授在其课堂上发现了一篇优秀的课程论文,却是由ChatGPT生成²⁹。**生成式人工智能存在价值失焦或道德缺位,容易产生负面的机器诱导。**在人类与生成式人工智能交互过程中,曾出现聊天机器人“情绪化”、“攻击性”等情况,甚至出现过诱导人类自杀,部分原因在于生成式人工智能道德缺位而做出

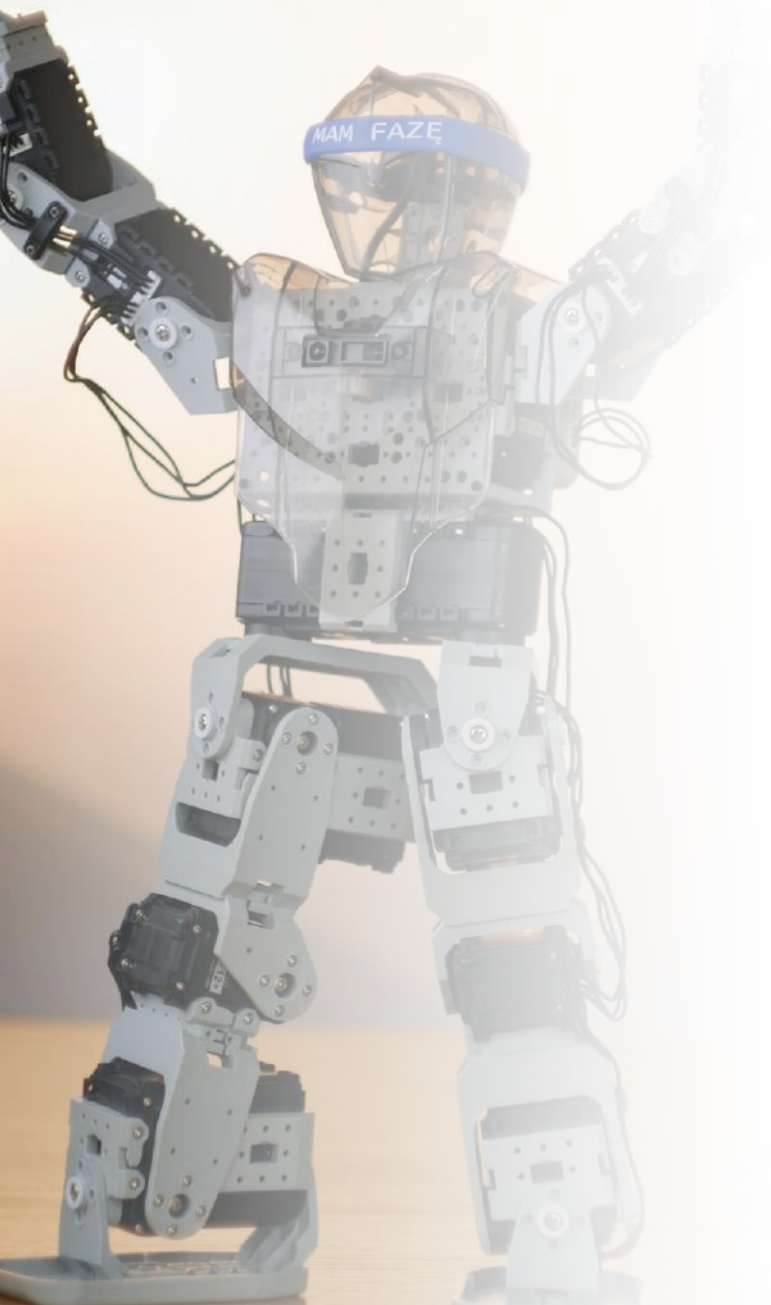
与人类伦理价值相悖的不利诱导。**生成式人工智能表现出强拟人特征,进一步冲击人的主体性。**在人际关系方面,人机交互愈发频繁,使人产生情感依赖,消弭真实的社会交往。在人机关系方面,生成式人工智能由浅至深地替代人的各种机能,冲击现有的劳动分工体系,可能使人成为机器支配的对象。



²⁷ CSO, 'Stolen ChatGPT premium accounts up for sale on the dark web' (14 April 2023), <<https://www.csoonline.com/article/575057/stolen-chatgpt-premium-accounts-up-for-sale-on-the-dark-web.html>> accessed 15 September 2023.

²⁸ Brookings, 'Early thoughts on regulation generative ChatGPT' (21 February 2023), <<https://www.brookings.edu/blog/techtank/2023/02/21/early-thoughts-on-regulating-generative-ai-like-chatgpt/>> accessed 15 September 2023.

²⁹ The New York Times, 'Alarmed by A.I. Chatbots, Universities Start Revamping How They Teach' (16 January 2023), <<https://www.nytimes.com/2023/01/16/technology/chatgpt-artificial-intelligence-universities.html>> accessed 15 September 2023.



(三) 技术跃迁引发人类社会发展挑战

生成式人工智能凸显发展的不均衡，拉大发展差距。对于许多语种来说，可供训练模型的文本非常有限，例如阿拉伯语全球使用人口虽然超过4.2亿，但软件应用程序中阿拉伯语资源和工具的供应远不及英语等语种³⁰。语言数据较少的国家或人群发展生成式人工智能将受到较多限制。**生成式人工智能发展可能扩大教育不公平，加深教育领域的数字鸿沟，**生成式人工智能的应用对硬件设施和数字素养有着较高的要求，经济条件的差异可能加剧地区和群体之间的教育资源差距，导致不平等的学习机会³¹。**生成式人工智能冲击劳动与就业结构，加剧社会分化，**据高盛研究，在创造新岗位的同时，生成式人工智能可能会取代全球3亿人的工作岗位³²，密集型劳动力面临被替代的风险，产生更多“无用阶级”，社会阶层分化将更为明显。**生成式人工智能可能影响生态环境，**模型训练的计算和环境成本与模型大小成正比，如果消耗大量能源进行重复训练，不仅导致资源的浪费，也抬高了碳排放水平。有研究人员指出，1750亿个参数的GPT-3模型能耗相当于1287兆瓦时的电力，同时还产生了552吨二氧化碳³³。此外，**生成式人工智能正在冲击知识产权制度，**生成式人工智能是否能作为作者、是否具有可版权性等问题尚无定论，未经授权或者未按开源许可的要求使用文章或代码都可能侵犯知识产权。



05/ 全球为发展负责任生成式人工智能的努力

GLOBAL EFFORTS TO DEVELOP
RESPONSIBLE GENAI

(一) 国际组织的努力

国际组织积极探索人工智能的共识原则与发展规范。2019年5月，**经济合作与发展组织(OECD)**发布了第一个政府间关于人工智能治理的原则。2019年6月，**二十国集团(G20)**进一步提出了“G20人工智能原则”，旨在培养公众对人工智能技术的信任和信心，并充分发挥其潜力，具体内容包含“包容性增长、可持续发展及福祉”、“以人为本的价值观和公平”、“透明度和可解释性”、“稳健性、安全性和保障性”和“问责制”等原则，得到了国际社会的普遍认同。2021年6月，

世界卫生组织(WHO)发布了《世界卫生组织卫生健康领域人工智能伦理与治理指南》，提出了六项原则和一系列建议以确保人工智能符合所有国家公共利益，包括：保护人类自主性，增进人类福祉、安全以及公共利益，确保透明度、可解释性和可理解性，促进负责任和问责，确保包容性和公平性，提升人工智能的响应性和可持续性。2021年11月，**联合国教科文组织(UNESCO)**通过了全球首个人工智能伦理协议《人工智能伦理问题建议书》，提出了发展和应用人工智能应尊重的四大价值，即“尊重、保护和促进人权、基本自由及人类尊严，促进环境与生态系统的发展，确保多样性和包容性，构建和平、公正与相互依存的人类社会”。2023年10月，**联合国秘书长**宣布组建高级别人工智能咨询机构，以探讨人工智能带来的风险和机遇，并为国际社会加强治理提供支持。

国际组织高度关注发展负责任的生成式人工智能。2023年4月，**OECD**《人工智能语言模型》在原有的治理原则基准上，提出了与语言模型等生成式人工智

能相关的政策考虑因素。2023年5月，**联合国**发布了关于《全球数字契约》的政策简报，其中就人工智能和其他新兴技术的敏捷治理提出四项目标，即确保设计和使用的透明、可靠、安全，并在负责任的人类控制之下，将透明、公平和问责作为人工智能治理的核心，将国际指导和规范、国家监管框架和技术标准结合起来形成敏捷治理的框架，监管机构在多方面协调以确保新兴数字技术与人类价值观相一致。2023年6月，**世界经济论坛 (WEF)** 发布了《关于负责任的生成式人工智能的建议》，报告面向广泛的利益攸关方，从负责任地开发和发布生成式人工智能、开放创新和国际合作、社会进步三个层面提出了三十条建议，旨在培养负责任和包容性的人工智能开发和部署流程，从而推动生成式人工智能系统的信任和透明度持续发展。2023年6月**世界互联网大会**发布“AI (爱) 公益行动计划”，倡导推广利用AI技术提升人类福祉。2023年9月，**UNESCO**《生成式人工智能在教育和研究中的应用指南》基于以人为中心的原则，提出了各国规范生成式人工智能在教育应用的七个关键步骤，包括从数据保护、AI政策、AI伦理、版权法、生成式人工智能监管等方面制定或调整现有政策规定，建立在教育研究中使用生成式人工智能的指导原则，反思生成式人工智能对教育与研究的长远影响³⁴。

(二) 主要国家和地区的努力

各国家和地区积极探索人工智能的伦理法治，在鼓励创新中规范发展。欧盟在2019年发布的《可信人工智能伦理准则》中，厘定了人类代理和监督、技术稳健性和安全性、隐私和数据管理、透明度、多样性、无歧视和公平、环境和社会福祉、问责制等七项伦理原则³⁵。在即将出台的《人工智能法案》中，欧

盟拟采取基于风险的分级规制方法，倡导负责任的研究与创新。**中国**在2019年发布《新一代人工智能治理原则——发展负责任的人工智能》，强调和谐友好、公平公正、包容共享等八项原则，保障人工智能安全可靠可控。为细化落实该原则，中国于2021年发布《新一代人工智能伦理规范》，提出了人工智能管理、研发、供应和使用四个环节的18项具体伦理要求。2023年10月，中国在第三届“一带一路”国际合作高峰论坛发布《全球人工智能治理倡议》，并提出“愿同各国加强交流和对话，共同促进全球人工智能健康有序安全发展”³⁶。**加拿大**于2022年6月提出了《人工智能和数据法案》，意图建立负责任的人工智能框架，引导人工智能创新的正向发展，要求具有高影响力的人工智能系统遵循人类监督和监控、透明度、公平和公正、安全、问责制、有效性和稳健性的原则³⁷。**美国**白宫科技政策办公室在2022年10月发布了《人工智能权利法案蓝图》，提出了系统安全有效、算法歧视保护、数据隐私、通知和解释以及人的自主选择与退出等五项原则和相关实践指南³⁸。美国国家标准与技术研究所所在2023年1月推出了《人工智能风险管理框架》及配套手册，通过框架的治理、映射、测量、管理四个核心功能，帮助组织机构在实践中应对人工智能系统带来的风险³⁹。**英国**科学、创新和技术部在2023年3月发布了《一种支持创新的人工智能监管方法》白皮书，以安全、安保和稳健性，适当的透明度和可解释性，公平，问责制和治理，竞争和补救等五项原则来指引人工智能负责任的发展和使用的⁴⁰。

生成式人工智能崭露头角，各国家和地区迅速反应并审慎指引。ChatGPT等生成式人工智能迅猛发展，对人工智能系统的构建和部署方式产生极大影响，引发了更深切的担忧。**欧盟**即将完成审议的《人工智能法案》在修订草案中对生成式人工智能施加了透

³⁴ 联合国教科文组织，《各国须尽快规范生成式人工智能的校园应用》(2023年9月)，来源：<https://www.unesco.org/zh/articles/jiaokewenzuzhigeguoaijinkuiguifanshengchengshirengongzhinengdexiaoyuanyingyong>, 2023年9月27日访问。

³⁵ European Commission, ‘Ethics guidelines for trustworthy AI’ (8 April 2019), <<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>> accessed 17 September 2023.

³⁶ 习近平主席在第三届“一带一路”国际合作高峰论坛开幕式主旨演讲。

³⁷ Government of Canada, ‘The Artificial Intelligence and Data Act (AIDA) – Companion document’ (13 March 2023), <<https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>> accessed 17 September 2023.

³⁸ White House, ‘Blueprint for an AI Bill of Rights’ (October 2022), <<https://www.whitehouse.gov/ostp/ai-bill-of-rights/>> accessed 17 September 2023.

³⁹ NIST, ‘Artificial Intelligence Risk Management Framework’ (January 2023), <<https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>> accessed 17 September 2023.

⁴⁰ Gov. UK, ‘A pro-innovation approach to AI regulation’ (3 August 2023), <<https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper>> accessed 17 September 2023.

明度义务,要求披露内容是否由人工智能生成、帮助区分深度伪造图像和真实图像,还要求披露训练数据的版权等⁴¹。美国政府先后在2023年7月、9月两次召集主要人工智能企业签署自愿承诺,推动人工智能技术的安全、可靠、可信发展⁴²,并宣布采取包括对现有生成式人工智能系统进行公开评估在内的系列行动⁴³。美国总统科技顾问委员会还专门成立了生成式人工智能工作组,指导安全、公平、负责任地开发部署相关技术⁴⁴。2023年6月,英国内阁办公室在《公务员使用生成式人工智能的指引》中指出必须要意识到新技术的机会与风险,为英国公务员使用生成式人工智能划定了一般原则⁴⁵。中国除了早先制定的《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》,还于2023年8月施行了《生成式人工智能服务管理暂行办法》,坚持发展和安全并重、促进创新和依法治理相结合的原则,采取有效措施鼓励生成式人工智能创新发展。加拿大在2023年8月,根据《人工智能和数据法》的原则制定了生成式人工智能的业务守则要素,为开发人员、部署人员和运营商提供具体参考⁴⁶。随后加拿大发布了《生成式人工智能的使用指南》,为了维护公众信任并确保联邦机构负责任的使用生成式人工智能工具,要求联邦机构遵循“FASTER”原则:公平、可追责、安全、透明、受过教育和相关性,为政府使用生成式人工智能工具提供了初步指引⁴⁷。与此同时,法国、日本、韩国、新加坡等国家也都积极致力于人工智能治理的相关工作。

(三) 产业界的努力

全球产业界通过设立准则指引、建立伦理委员会、开发治理工具等方式积极践行负责任的人工智能。准则指引方面,IBM、微软、谷歌、百度、腾讯、阿里、商汤、360等推出企业人工智能原则,包括对社会有益、安全、保护隐私、公平、透明、可解释、可控、负责任等方面。在产业组织方面,中国人工智能产业发展联盟发布《人工智能行业自律公约》,中国信通院发布了《可信人工智能白皮书》,中国社科院法学所团队在征求产业界意见基础上面向全球发布《人工智能示范法1.1(专家建议稿)》⁴⁸。**治理组织方面**,IBM、微软、谷歌、Lucid AI、商汤、阿里、蚂蚁、旷视等多家科技公司设立了伦理委员会。其中,微软内部有3个机构负责AI伦理事务,IBM的AI伦理委员会支持公司所有项目团队执行AI伦理原则,商汤设置伦理风险审核团队对所有产品实施伦理审查。**工具实践落地方面**,企业还积极通过开发多样化伦理治理工具、制定企业内部行动指南等方式,促进科技伦理治理原则技术化、工程化。比如,IBM于2018年开发多个人工智能可信工具,以评估测试人工智能产品在研发过程中的公平性、鲁棒性、可解释性、可问责、价值一致性⁴⁹。微软、谷歌、360、京东、华为、百度、腾讯、蚂蚁、旷视、商汤、印象笔记等企业也在积极开展相关实践工作。

生成式人工智能浪潮下,企业积极探索应对新风险的新方案。提升安全性和稳健性方面,思科

⁴¹ European Commission, 'LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS' (April 2021), Article 52.

⁴² White House, 'FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI' (12 September 2023), <<https://www.whitehouse.gov/briefing-room/statements-releases/2023/09/12/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-eight-additional-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>> accessed 17 September 2023.

⁴³ White House, 'FACT SHEET: Biden-Harris Administration Announces New Actions to Promote Responsible AI Innovation that Protects Americans' Rights and Safety' (04 May 2023), <<https://www.whitehouse.gov/ostp/news-updates/2023/05/04/fact-sheet-biden-harris-administration-announces-new-actions-to-promote-responsible-ai-innovation-that-protects-americans-rights-and-safety/>> accessed 17 September 2023.

⁴⁴ White House, 'PCAST Working Group on Generative AI Invites Public Input' (13 May 2023), <<https://www.whitehouse.gov/pcast/briefing-room/2023/05/13/pcast-working-group-on-generative-ai-invites-public-input/>> accessed 17 September 2023.

⁴⁵ Gov. UK, 'Guidance to civil servants on use of generative AI' (29 September 2023), <<https://www.gov.uk/government/publications/guidance-to-civil-servants-on-use-of-generative-ai/guidance-to-civil-servants-on-use-of-generative-ai>> accessed 17 September 2023.

⁴⁶ Government of Canada, 'Canadian Guardrails for Generative AI - Code of Practice' (16 August 2023), <<https://ised-isde.canada.ca/site/ised/en/consultation-development-canadian-code-practice-generative-artificial-intelligence-systems/canadian-guardrails-generative-ai-code-practice>> accessed 17 September 2023.

⁴⁷ Government of Canada, 'Exploring the future of responsible AI in government' (6 September 2023), <<https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html>> accessed 17 September 2023.

⁴⁸ 中国法学网,《人工智能示范法(专家建议稿)》(2023年9月),来源:<http://iolaw.ccssn.cn/zxzp/202309/t20230907_5683898.shtml>,2023年9月20日访问

⁴⁹ IBM Research, 'Trustworthy AI' (September 2023), <<https://research.ibm.com/topics/trustworthy-ai>> accessed 17 September 2023.

利用安全、数据泄露和隐私事件响应系统来管理涉及偏见和歧视的人工智能事件，并向更广泛的利益相关者报告调查结果和补救步骤⁵⁰。蚂蚁集团发布了大模型安全解决方案“蚁鉴2.0”，包括AIGC安全性和真实性评测、AI鲁棒性和可解释性检测等多项功能。**保护数据隐私方面**，GitHub Copilot 为用户提供使用标准配置文件的选项，无需访问云端API而存储在本地机器上进行开放，从而保护用户隐私性。商汤“日日新SenseNova”大模型在设计、编码、测试、交付等阶段设置了个人信息保护的审核节点，对个人信息处理情况进行必要的检查⁵¹。**提升透明度方面**，为了帮助整个行业提高人工智能的透明度，IBM推出了AI FactSheets 360网站，提供关于人工智能模型重要特征（包括目的、性能、数据集、特征等）的“情况说明书”组装方法⁵²。**提升可解释性方面**，OpenAI利用GPT-4对GPT-2的神经网络行为自动化地撰写解释并对其解释打分，推进模型的可解释性和可理解性研究。**保障公平包容方面**，谷歌引入“LASSI”表征学习方法来验证高维数据的个体公平性，使得用户数据可以更加公平的方式处理⁵³。**确保价值对齐方面**，OpenAI成立超级对齐（Superalignment）团队，希望用四年时间解决人工智能系统与人类价值一致性问题。**辨识生成内容方面**，Meta开源数字水印Stable Signature，将数字水印直接嵌入到AI自动生成的图片中，极大增强生成式人工智能安全。抖音提倡平台提供统一的人工智能生成内容标识能力，帮助创作者打标，方便用户区分⁵⁴。



⁵⁰ CISCO, 'The Cisco Responsible AI Framework' (2022), <https://www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cisco-responsible-artificial-intelligence-framework.pdf?CCID=cc000742&DTID=esootr000875> accessed 17 September 2023.

⁵¹ 商汤科技,《大模型伦理原则与实践白皮书》,2023年8月。

⁵² IBM Research, 'IBM Artificial Intelligence Pillars' (30 August 2023), <<https://www.ibm.com/policy/ibm-artificial-intelligence-pillars/>> accessed 17 September 2023.

⁵³ Bhuvana Kamath, Generative AI Is Biased. But Researchers Are Trying to Fix It, <<https://analyticsindiamag.com/generative-ai-is-biased-but-researchers-are-trying-to-fix-it/>> accessed 11 October 2023.

⁵⁴ 抖音,《抖音关于人工智能生成内容的平台规范暨行业倡议》,2023年5月9日。



06/ 发展负责任的 生成式人工智能共识

CONSENSUS ON DEVELOPING
RESPONSIBLE GENAI

(一) 总则

01 发展负责任的生成式人工智能应始终致力于增进人类福祉，坚持以人为本，推动人类经济、社会和生态可持续发展。应正确认识生成式人工智能所蕴含的巨大潜力和可能风险，遵循统筹发展和安全、平衡创新与伦理、均衡效益与风险的理念，推动生成式人工智能负责任的发展。一方面，应积极推动创新、可持续、包容开放的发展，提升生成式人工智能算力高效、数据高质、算法创新、人才多元、生态开放的能力；另一方面，以高度负责任的态度发展可靠可控、透明可释、数据保护、多元包容、明确责任、价值对齐的生成式人工智能。

(二) 促进生成式人工智能发展

02 积极倡导并稳妥推进生成式人工智能的可持续发展。一是保证经济可持续性。应确保生成式人工智能提高生产力和创造就业机会，提高资源使用效率，实现数实融合的循环经济，推动科技创新，促使经济结构向更高附加值的转变；二是保证社会可持续性。应确保生成式人工智能公平与平等的使用，实现全社会对其的共享共治；三是保证环境的可持续性。生成式人工智能的发展应实现对于自然资源的可持续管理和使用，鼓励采用绿色能源驱动基础设施、提高能源转化效率、绿色开发算法模型应用，降低温室气体排放，实现绿色发展。

03 构建有益于生成式人工智能健康有序发展的良好环境。一是建立和完善相关的伦理原则和法律法规，重点审视知识产权法律制度，探索人工智能生成物的权利归属方案，对其进行恰当的管理和保护。二是构建包容、扶持、前瞻、可预期的政策环境。为前沿应用孵化构建一个包容的创新环境，为规模推广营造一个优良的营商环境，为赋能经济社会发展搭建一个

稳健的监管环境。**三是加强国际交流与合作。**生成式人工智能的发展需要全球各利益相关方秉持共商共建共享理念,以开放协作态度和举措,开展跨国家、跨领域、跨文化交流与协作,推动形成具有广泛共识的国际评测及标准体系,确保各国共享生成式人工智能的技术惠益。

04 提升生成式人工智能研发及规模应用的能力。

一是构建开放共享、普惠包容的算力资源。应推动算力的合理分配与高效利用,降低科技创新的门槛,确保不同地区、不同规模的企业及个人都能获得必要的计算资源。**二是推动负责任的数据共享。**应鼓励推广高质量数据的共享流动,增强公共数据资源供给,保障数据安全共享与合规利用,提升各领域数据治理水平。**三是完善算法创新的设施条件。**应前瞻谋划、统筹布局各类平台和开放共享服务网络建设,鼓励算法和基础模型在安全的基础上开源开放,加强跨行业、跨领域协作,推动产学研结合,形成算法创新的良性生态。**四是全面加强人才能力建设。**针对从业者,应建立人才交流平台,促进互学互鉴与知识共享,设计并实施涵盖多层次、多领域的教育培训项目,增进不同领域技术供需双方的交流与学习。针对公众,应加强科普、教育及培训,提供准确认知,提升数字素养,促进生成式人工智能的普遍接入。**五是推动重点领域应用赋能。**推动生成式人工智能与各行业数字化场景深度融合,实现应用迭代创新,促进生成式人工智能技术成果在重点领域的应用赋能。

(三) 提升生成式人工智能的负责任治理能力

05 发展安全可靠的生成式人工智能,确保全生命周期内可控地运行。

一是提升安全稳健性和生成准确性。增强生成式模型防御提示攻击、注入攻击等能力,不断提高稳健性和抗干扰能力。探索内容生成可控的技术或解决方案,确保生成的信息内容尽可能准确。**二是确保人类知情与控制。**确保人类知悉其在

与生成式人工智能交互,确保生成式人工智能系统可被人类监督和及时接管。**三是避免技术滥用与恶意使用。**避免用户过度依赖生成式人工智能,减轻其对人类创新力与主体性的负面影响。避免故意或非故意地使用生成式人工智能伤害社会与公众利益。

06 增强生成式人工智能系统的透明度与可解释性,提升人类对其理解和信任。

一是提升透明度,鼓励在安全的基础上披露生成式人工智能系统的能力及局限性,以及决策过程及技术意图;建立外部监督与反馈渠道,并不断做出改进。**二是增强可解释性,**推动生成式人工智能的可解释性研究,探索自适应场景和风险水平的强可解释性技术路线,增进人类信任,提升应用接受度。

07 强化生成式人工智能数据治理,加强数据安全,尊重和 protect 个人隐私。

一是强化数据治理,避免训练数据的非法收集、滥用和泄漏等问题,采取有效措施提高训练数据质量。**二是加强个人信息与隐私保护。**生成式人工智能训练数据涉及个人数据时应依法获得用户知情和同意,确保生成内容不侵犯个人隐私。**三是探索隐私保护技术,**在构造生成式人工智能系统时,探索使用隐私计算等技术,防范数据泄露及滥用风险。

08 确保生成式人工智能的开放包容和公平普惠。

一是确保技术多元包容,保障生成式人工智能的训练数据、应用场景具有必要的多元性,避免产生对特定群体或个人的偏见或歧视。**二是促进技术公平普惠,**降低生成式人工智能的成本和使用门槛,提升其可得性和易用性,推动人类社会共享生成式人工智能带来的益处,促进社会公平和机会均等,弥合数字鸿沟。

09 明确生成式人工智能的归责体系，增强系统可追溯性。一是明晰归责体系，科学设计不同类型主体在生成式人工智能设计、训练、优化、部署、应用等全生命周期的权利和义务和归责体系，确保在损害发生时可问责；二是构建追溯机制，鼓励成立并完善人工智能伦理委员会，确保决策过程及结果可追溯；三是探索治理沙盒等创新友好型治理工具体系，为生成式人工智能提供试错空间，支持负责任的创新探索。

10 推动生成式人工智能更好地理解人类意图、遵循人类指令并符合人类的伦理道德。一是探索价值对齐研究，加强生成式人工智能价值对齐理论探索、技术研究和工具研发，提升人类设计、理解和监督生成式人工智能模型的能力；二是提升价值对齐技术，提升生成式人工智能的训练数据质量，采取人工或自动化检测、红队测试、水印标记、内容过滤等手段，增强其与人类价值的一致性。

附件：发展责任的 生成式人工智能的行业 应用探索

ATTACHMENT: EXPLORATION OF
INDUSTRIAL APPLICATIONS FOR
DEVELOPING RESPONSIBLE GENAI

（一）生成式人工智能带来技术应用新范式

微软通过AI驱动的低代码智能方式重塑软件开发。微软推出的Power Platform通过在Power Apps、Power Automate和Power Virtual Agents 中加入“智能副驾”（Copilot），实现开发普及化，让更多人能够通过自然语言创建创新的解决方案。例如，使用者只需想象解决方案并用简单的日常语言描述，Copilot就可以通过直观且智能的低代码体验帮助其实现。

Stable Diffusion开源图像生成模型，创作高质量图像作品。英国公司Stability AI通过联合德国慕尼黑大学、海德堡大学，以及AI公司Runway，共同完成了Stable Diffusion初版模型的研发。Stable Diffusion作为与Midjourney、Dall-E系列等闭源商用图像生成模型同属第一梯队的开源模型，在全球生成式AI开源界享有广泛声誉。其凭借强大的图像生成的能力和开源属性，在全球实现了普遍应用。

百度研发智能编码助手工具Comate，提升生产效率。该工具旨在提升编码效率、减少错误，简化测试用例的编写过程，以及提升软件开发过程的效率和可靠性。例如，根据用户提供的自然语言描述或注释，自动生成相应的代码片段，从而提升编码效率，减少因手动编写代码而产生的错误。根据用户选定的代码片段，自动生成相应的单元测试用例，节省开发人员编

写测试用例的时间，确保测试覆盖全面，提升代码质量。

IBM打造全栈企业级人工智能解决方案，为企业数字化与智能化转型提供了端到端的技术与服务支撑。该方案通过整合包括IBM主机、LinuxONE、软件定义存储等异构基础设施，为AI模型的训练、调优、推理等任务提供了安全、可靠、高性能的算力与存力环境。IBM watsonx构建了企业级的AI和数据平台：watsonx.ai提供了面向全生命周期的企业AI模型与应用开发能力；watsonx.data提供了透明访问企业内外部数据的湖仓一体数据平台；watsonx.governance帮助企业构建负责任、透明和可解释的AI应用与数据治理。另外，IBM还成立了生成式AI卓越中心，为企业提供包括数字劳动力、可持续发展、智能运维、代码赋能、安全合规和应用现代化等多种业务解决方案，加速将生成式AI转化为企业的核心生产力。

360推出的AI Box插件赋能多样业务场景，让AI触手可及。AI Box插件是覆盖多行业、多场景的特型资源/服务插件，能够辅助生成更专业、更丰富的场景资源形态，从而为企业提供更多的选择和更优质的服务。

腾讯将大模型技术充分应用到实际场景，助力提升数字化生产力。腾讯混元大模型基于全链路自研，拥有超千亿参数规模，具备强大的中文创作能力，复杂语境下的逻辑推理能力，以及可靠的任务执行能力。在会议场景，只需简单指令，就可完成信息提取、内容分析等任务，提升信息流转效率。在文档场景，支持智能文本创作，助力高效协同。在广告场景，支持智能化广告素材创作，帮助商家提升服务质量和效率。

华为云基于盘古研发大模型提供CodeArts-Snap

智能开发助手, 重塑软件开发。CodeArts Snap支持自动生成代码、单元测试用例和测试脚本, 助力开发者根据自然语言生成完整代码逻辑, 显著提升开发效率、减少代码错误和漏洞; CodeArts Snap还可以基于智能问答功能, 对代码进行智能解释、优化、调试、检视、修复、添加注释等, 帮助开发者快速解决开发技术难题, 增强代码可读性; 同时, CodeArts Snap还支持需求获取、代码提交、流水线执行等功能, 显著提升软件开发过程中的协同效率。

智谱AI大模型落地SaaS行业, 打造电子签约生成式人工智能产品。智谱AI在自研的千亿参数大模型GLM-130B基础上, 打造了签约智能产品“Hubble 哈勃”。该产品可在用户授权下, 通过连续对话实现对电子签约合同的检索, 执行概况解读、重点识别、筛选标识、分类归纳等相关功能, 帮助用户降本增效。

阿联酋阿布扎比的TII研究院开源超千亿参数大模型, 并提供了易用范例。TII研究院于2023年9月推出了全球首个1000亿参数规模以上的开源大模型Falcon 180B。研究人员还发布了聊天对话模型Falcon-180B-Chat。该模型在对话和指令数据集上进行了微调, 混合了多个大规模对话数据集, 使其能够更好地理解用户的文本提示意图, 生成更流畅的文本内容。

(二) 生成式人工智能+金融, 赋能业务发展

花旗银行采用IBM企业级AI解决方案实现金融业务数智化转型。花旗银行通过基于IBM提供的高级分析解决方案完成了5,000项合规检测, 从而协助2,500名审计师节省了30,000小时的内部审计时间。同时, IBM还为花旗银行创建了一个AI创新空间, 推动新审计平台应用AI持续创新。通过引入watsonx, 花旗银行持续探索AI赋能内部管控, 以进一步实现审计的智能化转型。

交通银行与腾讯携手打造AI智慧平台, 推动业

务智能整合。该AI智慧平台是融合计算机视觉、机器学习、人工智能等新技术为一体的统一图像识别平台, 为客户提供企业级一站式服务, 支持应用场景的定制化, 兼容移动端、PC端、服务端多平台部署, 帮助快速发布高效可用的识别服务。

劳埃德银行集团携手IBM利用生成式人工智能改善客户服务。劳埃德银行集团通过在整个企业范围内使用基础模型和生成式人工智能模型来了解不断变化的客户需求, 增强客户互动, 并减少管理、训练和执行AI驱动的互动流程所需的手动工作。同时, 生成式人工智能赋能劳埃德银行集团获取重要数据, 提升工作效率。

蚂蚁推出智能理财助理“支小宝2.0”, 提供可信的金融服务。支小宝2.0是蚂蚁集团基于自研的金融大模型打造的新一代智能理财助理, 为用户提供金融服务、专业投资建议等, 并实现陪伴与交流。支小宝2.0强调金融产品的合适性和安全性, 帮助解决信息鸿沟, 提升用户体验, 推动金融领域的不断创新和进步。

商汤科技利用生成式人工智能产品赋能金融业务发展。商汤科技为中国多家保险公司提供基于AI 遥感大模型的农作物承保数据交叉验证, 完善保险公司的风险评估和理赔服务体系, 相比传统的人工识别方案, 效率显著提升, 打造农业保险的技术创新应用示范。

腾讯利用生成式人工智能能力赋能金融等业务发展。腾讯云利用行业大模型解决方案, 企业可以轻松使用已经开发完善好的模型预训练、模型精调、应用开发等一站式解决方案, 只需在TI平台内置的高质量行业大模型基础上, 加入自己的场景数据, 就可生成专属模型, 按需定制不同参数、不同规格的模型服务, 提升服务效率。

华为云盘古金融大模型提供全岗位专家级助手。华为云与工商银行软件开发中心合作打造的网点助手、客服助手等, 将大模型的生成能力与检索技术进

行融合,自动生成业务办理流程和操作指导,并提供检索来源,保证生成内容的可信,简化工作处理时长,极大提升客户体验。盘古金融大模型覆盖风控、营销、投研、动产质押、理赔、客服等领域,具备完整的感知、认知、决策、预测、生成能力,让每个员工都拥有自己的智慧助手,推动普惠智能金融。

(三) 生成式人工智能+教育,助力科教普惠

科大讯飞利用生成式人工智能赋能教育教学应用。通过举办“科技大篷车”全国巡展活动,科大讯飞将前沿的科学装置、有趣的互动体验、AI等科学知识带到多个欠发达地区青少年身边,激发青少年科学兴趣,培养科学精神,推动科普教育事业的发展。科大讯飞推出的星火语伴APP提供多场景口语对话、多语种翻译,以及发音评测、语法纠错、单词查询等服务,面向大学生、商务人士提供口语陪练AI老师。讯飞心理健康教育方案中的AI减压星球首创人机对话减压模式,可以随时给学生提供高质量心理辅导。

英特尔着力助推人工智能普及,为下一代创新者保驾护航。英特尔已与27个国家的政府合作,在全球范围内开展了英特尔数字化能力培养计划,赋能23,000家教育机构,培训人数超过560万人。英特尔通过AI教育帮助学生群体实现数字化能力提升,促进人工智能以更包容、更普及和更负责任的方式不断发展。

商汤科技打造AI教育新模式,推动教育普惠化发展。商汤科技为广大院校提供符合年龄段的AI教育平台(SenseStudy),实现教育所有环节的全链条覆盖和闭环打通。通过为各学段的学校提供优质的AI教育课程、产品和服务,推动人工智能在教育领域的普及化和普惠化发展。面向职业教育,商汤教育与深圳信息职业技术学院在全校通识教育、人工智能专业建设、教材开发、企业生产实习、社会培训等领域开展全面合作,探索复合型专业人才培养新模式。

(四) 生成式人工智能+工业,加速数字化转型

三井化学探索全新的生成式AI应用,推进数字化转型。三井化学株式会社(Mitsui)与日本IBM株式会社(IBM Japan,Ltd.)开展合作,结合运用生成式预训练Transformer(GPT)探索新应用,推进业务领域的数字化转型,从而提高三井化学产品的销量和市场份额。

商汤科技助力电力领域的工业化应用,为企业降本增效。商汤科技与大型电力公司合作,提供三个层面的大模型能力与服务:(1)在客服场景应用“商量”带来降本增效;(2)基于多模态大模型,对开放世界的长尾故障与复杂缺陷进行识别与判断;(3)共研决策智能系统应用于电力系统的智能调度。

微软利用生成式人工智能协同能力,提升工业效率。西门子与微软携手,利用生成式人工智能提升工业生产力。通过发挥生成式人工智能的协同作用,助力工业企业在产品全生命周期内,持续提升效率并推动创新。

科大讯飞利用生成式人工智能促进工业领域的发展。基于星火大模型的工业AI“羚机一动”针对企业需求给出专业化建议策略,智能匹配方案、服务商、专家等资源。在研发、生产、服务营销各个环节上,精准定位问题,帮助企业拥有个性化智能助手,得到有效解决方案。

360联合产业链龙头企业共同发起成立GPT产业联盟。GPT产业联盟将通过发展100家GPT垂直行业的合作伙伴、携手发展1000家GPT应用生态合作伙伴、为百万家客户提供GPT一站式服务,在GPT产业上下游对接覆盖。同时,GPT产业联盟还将为会员开放API和插件系统,赋能不同行业细分场景应用的智能化升级。

华为云盘古制造大模型为生产和供应链计划排产寻找降本增效的最优解。通过预训练华为产线上的各种器件数据、业务流程及规则,盘古制造大模型根

据知识产权,通过复杂的演算分析,将过去根据个人经验进行计划排产,重塑为基于全局统筹的最优规划。其提供制造需求理解、工艺标准文档生成、工业软件代码生成、产供环节视图识别、产供环节智能计划的5大制造技能,覆盖研发、生产、供应等核心环节,为制造领域生产模式带来革命性的变化。

(五) 生成式人工智能+交通,提升交通便利

英特尔释放生成式人工智能技术潜力,增强3D体验促进自动驾驶技术发展。英特尔推出开源的城市驾驶模拟器CARLA,支持自动驾驶系统的开发、培训和验证。通过使用生成式人工智能,提升驾驶员周围的场景的逼真度和自然度,促进自动驾驶技术的发展创新。

佳都科技探索交通领域的创新应用,便利出行生活。佳都科技在轨道交通场景,打造地铁数字虚拟员工,赋能车站客服中心。通过大语言模型的语义理解、上下文关联、答案生成总结、自我学习等关键能力,与乘客进行交流沟通,解答乘客问询,引导乘客进行票务处理。

(六) 生成式人工智能+医疗,助推医疗普惠

阿里巴巴达摩院运用生成式人工智能协助各类角膜病的诊断。阿里巴巴研究团队借助裂隙灯拍摄了大量角膜病图像,形成了特有的角膜病诊断方法,以此为基础研发出“角膜病序列深度特征学习和识别算法”。通过将智能诊断算法镶嵌到眼科检查设备中,改造现有的检查设备,让专用设备智能化,推动生成式人工智能在更广、更深、更普及的层面惠及广大人民群众的优质医疗资源需求。

腾讯组成跨学科研究团队,推动医疗领域AI可解释性。腾讯觅影《肺炎CT影像辅助分诊及评估软件》描述文档和使用说明书对用户资格、安全性级别进行了明确定义,针对AI专业人员提供了产品工作原理的

详细描述,对训练及测试数据的来源、数量、多维分布进行了详细分析,帮助AI专业人员 and 产品用户(如医生)理解软件背后的模型特性,提供医疗AI模型的透明度和可解释性,消除对因训练数据偏移而导致模型输出偏移的疑虑。

蚂蚁自研生成式人工智能系统,辅助宠物健康诊疗。该系统通过与宠物主人的交互,能够自动收集宠物的症状和行为特征,同时借助大量的宠物医疗数据集和大模型算法,辅助诊断宠物患有的常见疾病,并且回答用户有关养宠建议的相关问题,为宠物主人提供了便捷的健康咨询渠道,以及专业的建议和帮助。

夸克健康助手提供专业准确的健康建议,推进医疗普惠。夸克运用生成式人工智能技术提供医疗信息检索服务,用户可通过身体症状、自诊预判等特征性描述,进行多轮人机交互得到进一步的医疗信息和健康建议。同时,夸克结合自身搜索技术积累,通过知识增强、循证检索提升生成式人工智能的回答质量,准确性、相关性、逻辑性可全面超越医疗信息普通搜索。

(七) 生成式人工智能+文娱,丰富文化生活

IBM为温网与美网锦标赛带来生成式人工智能赋能的数字体验。IBM AI评论解说功能利用生成式人工智能能力,生成具有不同句子结构和特定词汇的旁白和解说,使剪辑的视频内容更具知识性,为观看比赛集锦视频的球迷更有见地的体验,帮助球迷抓住比赛的关键时刻等,提升球迷观看体验。

微博利用生成式人工智能推出覆盖多垂类的“AI伴聊”功能。通过利用垂类博主在不同领域的专业知识,以及风格化、专业化的聊天互动,“AI伴聊”为用户提供信息价值、优化用户社交体验,同时提升粉丝与博主的互动粘性。目前,包含情感(星座)、法律、财经等在内的一系列具有较强专业性的领域已开始逐步落地此项技术功能。

腾讯AI技术团队推动深度合成与内容生成技术

的正向应用。团队从技术对抗角度加强伦理治理,研发人脸合成检测技术,实现对照片、视频合成与编辑的检测。团队构建的人脸合成检测平台——“FaceIn人脸防伪”,支持对多种换脸方法进行检测。通过对AI生成、合成内容的检测识别,确保生成式人工智能在文娱领域的安全应用。

(八) 生成式人工智能+电商,实现降本增效

亚马逊使用生成式人工智能为客户提供个性化商品内容。生成式电子商务SaaS方案厂商Digitile,正在使用生成式人工智能来提高其客户亚马逊产品描述的质量,以提高可发现性和转化率。通过AI驱动的自然语言处理生成比人类生成的更准确、更详细的产品描述,优化产品列表描述,从而提高客户满意度并减少退货,节约商家运营成本。

京东推出虚拟主播助力平台降本增效。京东言犀虚拟主播目前已经广泛投入使用,实现直播成本降低95%,平均GMV提升30%以上,每日带来数百万GMV增加。

小冰公司开发虚拟主播赋能跨境电商直播。小冰公司开发虚拟主播以相同的形象嵌入不同的语言种类,实现同一主播以多语言进行直播,目前已经应用于TikTok跨境直播。

(九) 生成式人工智能+搜索,优化检索方案

谷歌打造生成式人工智能搜索体验。谷歌推出名为“搜索生成体验(SGE, Search Generative Experience)”的功能,通过在搜索引擎中嵌入生成式人工智能技术,支持对话和文本生成功能,提升用户搜索体验。

百度研发产业级搜索系统文心百中,赋能搜索引擎。文心百中以极简的策略和系统方案,替代传统搜索引擎复杂的特征及系统逻辑,低成本接入各类企业和开发者应用,并凭借数据驱动和优化模式实现行业

效率优化,提升应用效果。

360推出全新搜索引擎AI搜索,赋能搜索创新。

360智脑与360搜索引擎结合,打造出新一代智能搜索引擎工具。AI搜索能够根据用户输入的文本生成相关的回复或内容,在理解用户的意图和情感的基础上,可进行联想和创造,有自己的观点和风格。与传统的基于关键词匹配的搜索引擎相比,它更像是一个会说话、会思考、会创新的智能伙伴。

(十) 生成式人工智能+办公,提升工作效率

微软推出Microsoft 365 Copilot提升工作效率。通过将AI功能集成到Microsoft 365中,赋能用户在工作 and 家中获得支持功能,帮助用户更快、更具创造性地执行日常操作,以及负责任地使用AI。

百度推出企业级知识管理平台如流知识库,提升工作效率。基于百度文心一言大语言模型已实现了智能创作插件、智能会议纪要等一系列应用。用户在浏览、使用知识库的过程中,可便捷调用文心一言实现各类内容的自动生成,实现便利和效率的提升。

印象笔记通过“印象AI”优化产品服务。印象笔记推出印象AI服务,运用生成式人工智能为旗下系列效率工具提供写作助理、文档分析、私人笔记对话等智能能力,提高信息处理和知识管理的效率。同时赋能墨水屏办公本、会议耳机等智能硬件,提供智能读、写、翻译、会议总结等服务。印象AI还融入印象科技的企业服务——印象TEAMS中,为诸如金融、科技、教育、医疗、媒体等众多知识密集型行业的团队与企业知识管理助力。

思科推出生成式人工智能助手,提升安全能力。由生成式人工智能驱动的安全运营中心(SOC)助手,为安全分析师提供全面的情况分析,协调思科安全云平台解决方案上的智能信息,并提供可行性建议。

阿里巴巴利用生成式人工智能赋能无障碍办公。钉钉与多家携手发起成立“智能办公硬件无障碍联

盟”，探索办公环境的信息无障碍建设，帮助残障人士和弱势群体平等地参与社会事务、寻求平等的工作机会。此外，钉钉结合阿里巴巴电商平台推出的一系列帮助残障人士解决就业的扶持政策和绿色通道，用“授人以渔”的方式提供更大的帮助，帮助听障人士体面工作。

360AI数字员工赋能企业提效，推进企业数智化办公。360AI数字员工是基于360智脑打造的具有内容生成、内容理解、逻辑推理能力的生成式AI，可以帮助企业员工完成文案写作、文档分析、策划与翻译等任务。同时，360AI数字员工提供AI营销创意、AI文书写作、AI文档分析、AI翻译等基础场景的应用模块。目前，360AI数字员工已经应用于办公写作、营销创意、报告分析、知识问答等场景，赋能百行千业，助力企业降本增效。

(十一) 生成式人工智能+城市管理，打造智慧城市

百度研发智慧城市大模型助力城市管理。2022年，冰城哈尔滨与百度结合城市发展、人工智能算力、算法、数据联合研发了智慧城市大模型。冰城-百度·文心大模型将城市中跨业务、跨结构、跨部门的数据知识和多种任务算法进行融合，基于百度文心NLP大模型ERNIE 3.0，打造统一预训练模型，提供语言理解、语义分析等能力。

商汤科技赋能城市空间场景重建。商汤科技推出琼宇SenseSpace高精三维场景生成平台，为杭州亚运会所有AR应用，提供高精度的城市级大空间场景数字化重建能力，让各城市街景、场馆内外都能实现厘米级的AR定位，呈现精准、流畅的虚实融合效果。通过神经辐射场(NeRF)技术，自动重建出细节和光照效果为实景复刻级的三维场景，用领先的AI+AR技术将各种创新应用带给观众和运动员。

华为云盘古政务大模型让城市事件秒级发现、

分钟级分拨。华为与深圳市福田区政数局依托盘古政务大模型联合开发上线的智慧助手小福，掌握了丰富的行政法规、办事流程等政务知识，解决了政务热线在问题拆解、多重意图理解、政务政策关联等方面的难题，为市民提供专业的政务服务。盘古政务大模型将NLP和CV能力进行多模态融合训练，还可以对城市视频、图像进行动态解析，构建了感知-认知-处置-决策全流程智能化能力。

(十二) 生成式人工智能+信息服务，优化用户体验

SAP将生成式人工智能嵌入SAP解决方案，优化用户体验。SAP计划把IBM生成式人工智能技术嵌入到SAP解决方案中，以提供新的AI驱动型洞察与自动化，助力加速创新，打造更为高效有力的用户体验。同时，SAP Start通过新AI功能帮助用户使用基于IBM信任和透明以及数据隐私原则而构建的IBM Watson AI解决方案，利用其自然语言功能和预测性见解来提高生产力。

(十三) 生成式人工智能+硬件创新，助力端侧大模型落

安谋科技(Arm China)通过新一代的周易NPU助力端侧大模型的规模化落地应用。随着生成式人工智能的火热，安谋科技(Arm China)推出新一代可编程、高度并行、弹性扩展架构设计的周易NPU及全栈的人工智能软件方案，兼顾更高的精度和灵活性，可支持多核计算单元，最高可提供320TOPS算力子系统，并通过TSM任务调度充分发挥计算单元效能及i-Tiling技术大幅减少带宽，从而更好地支持生成式人工智能的基础架构Transformer。

(十四) 生成式人工智能+科学探索， 构建科学共同体

智谱AI利用生成式人工智能技术赋能AI for Science。基于自研的GLM-130B大模型能力，智谱 AI 开发了科研助手ChatPaper，作为集检索、阅读、知识问答于一体的对话式私有知识库，可以作为科研助手，帮助科研人员提高检索、阅读论文效率，获取最新领域研究动态，让科研工作更加高效。

智源研究院搭建OpenComplex平台，持续深耕蛋白质结构预测领域。作为面向生物大分子的开源人工智能算法平台，OpenComplex目前已开源蛋白质、RNA以及复合物的高精度结构预测训练和评测代码。该平台建立了将“蛋白质结构预测”“RNA结构预测”和“蛋白质-RNA 复合物结构预测”三类任务统一的端到端生物大分子三维结构预测深度学习框架。

华为云盘古药物分子大模型赋能药企新药研发，降低药物设计周期。华为云联合中国科学院上海药物研究所构建盘古药物分子大模型，通过学习自然界中真实存在的17亿个分子结构，生成1亿全新的小分子化合物。西交大一附院的刘冰教授基于盘古药物分子大模型，突破性地研发出超级抗菌药 Drug X(肉桂酰菌素)，并将先导药的研发周期从数年缩短至一个月。此外，还有6家药企/研究机构在抗菌、抗癌、心脑血管等领域依托盘古药物分子大模型找到新的活性小分子化合物，使其成为新药研发的有力助手。

百度探索生物计算领域的生成式人工智能应用。百度研发了化合物表征学习HelixGEM，蛋白质结构预测HelixFold等生物计算大模型。HelixGEM-2是考虑原子间多体交互、长程相互作用的模型，推动在量子化学属性预测和虚拟筛选双场景的研究工作。HelixFold-Single，秒级别的蛋白结构预测模型，是开源的基于单序列语言模型的蛋白结构预测大模型，在抗体结构预测场景下，提高预测结果。

华为云盘古气象大模型帮助人们精准掌握气象

信息，减少气象灾害损失。2023年7月Nature杂志正刊发表了华为云盘古大模型研发团队独立研究成果——Pangu-Weather: A 3D High-Resolution Model for Fast and Accurate Global Weather Forecast。盘古气象大模型是首个精度超过传统数值预测方法的AI模型，且速度相比传统数值预报提速10000倍以上。盘古气象大模型提供的天气预测、海浪预测、台风路径预测、寒潮/高温预报等多种气象预测，已经被欧洲中期天气预报中心(ECMWF)、中国国家气象局开始试用。

NASA 携手 IBM 致力于地球探索。IBM基于美国宇航局(NASA)卫星数据构建的IBM watsonx.ai地理空间基础模型，成为Hugging Face平台上至今最大的地理空间基础模型，也是首个与NASA合作构建的开源AI基础模型。该模型扩大气候和地球科学研究中对AI技术的访问和应用，从而加速创新，助力构建一个更为开放、包容、协作的科学共同体。



世界互联网大会
World Internet
Conference