

中国联通研究院根植于联通集团（中国联通直属二级机构），作为中国联通科技创新专业子公司，自2022年起，挂牌成立中国网络安全研究院、下一代互联网宽带业务应用国家工程研究中心及首个国家级网络安全产业知识产权运营中心，形成“两院两中心”发展格局，开创了研究院高质量发展的新局面。中国联通研究院作为服务于国家战略、行业发展、企业生产的战略决策参谋者、技术发展的引领者、产业发展的助推者，坚持以高水平科技自立自强为使命担当，聚焦网络强国、数字中国主责，拓展联网通信、算网数智业务，形成“态度、速度、气度、有情怀、有格局、有担当”的企业文化，以下一代互联网、光网络、5G-A/6G、网络安全、数智网优、低空智能网联和新型智库研究七个领域为主攻方向，坚持“四个聚焦”，开展关键核心技术攻关、科创力量建设、专业技术人才队伍建设、创新成果转化等工作，着力实现价值创造提升、战略科技力量提升、专精特新能力成色提升，争做通信行业科技创新主力军，努力建设成为“国家信赖、行业领先、集团倚重、员工自豪”现代化一流研究院。

中国生成式人工智能 应用与实践展望

中国联通研究院 | 中国联通国际公司 · 2024年9月

战略决策的参谋者
技术发展的引领者
产业发展的助推者

态度、速度、气度
有情怀、有格局、有担当

中国联合网络通信有限公司研究院

地址：北京市亦庄经济技术开发区北环东路1号

电话：010-87926100

邮编：100176



中国联通研究院



中国联通泛终端技术

版权声明

本报告版权属于中国联合网络通信有限公司研究院，并受法律保护。转载、摘编或利用其他方式使用本报告文字或者观点的，应注明“来源：中国联通研究院”。违反上述声明者，本院将追究其相关法律责任。

目录

前 言	1
一、生成式人工智能发展迎来爆发期	4
（一）生成式人工智能开辟数字经济发展新空间	4
（二）全球生成式人工智能市场前景广阔	6
（三）中国在生成式人工智能领域加速突破	11
二、生成式人工智能产业发展前瞻	15
（一）基础设施层：智算设施将是产业发展的基石	16
（二）模型层：大模型商业模式将逐渐贴近一线用户	19
（三）应用层：内容、场景、生态将是应用发展关键	21
三、中国生成式人工智能的典型实践与应用	23
（一）生成式人工智能深入生活场景，重塑智能生活新篇章	24
（二）生成式人工智能推动行业变革，引领转型升级新潮流	37
四、探索“生成式人工智能”未来发展新路径	58
（一）重点布局算力平台，建立算力合作协同机制	59
（二）优化高质量数据供给，促进数据开放与共享	60
（三）加大基础大模型投入，联合打造行业大模型	61
（四）布局内容与生态，以应用带动整体能力发展	62
（五）实施包容审慎的监管，推动国际合作与创新	63

前 言

生成式人工智能（Generative AI）是一种基于学习大规模数据集生成新的原创内容的新型人工智能，其通过大量学习数据或语料库，生成全新的、或与原始内容相似但不限于原始数据的文本、图片、视频、代码等内容。生成式人工智能逐渐成为科技行业的焦点，重塑着各行各业的未来，在全球范围内引领一场深刻的产业革新浪潮，为中国乃至金砖国家的智能化发展提供了前所未有的机遇。

本白皮书聚焦生成式人工智能产业发展前瞻趋势，深度洞察中国在生成式人工领域的应用与实践。在应用方面，白皮书将从居民生活场景和行业应用场景两个维度，精选中国典型实践经验进行介绍。在居民生活场景中，分析日常、办公、娱乐等细分场景需求，探索如何满足个性化需求和提高生活质量的方式；在行业应用场景中，洞察电信、医疗、金融和工业等重点行业应用情况，深入解读相关行业的业务痛点、发展成效、应用经验。希望中国在生成式人工智能行业应用的探索和实践可为金砖国家未来的发展提供参考，中国愿与全球各国共同合作，探索新的发展模式，让数字文明造福各国人民，推动构建人类命运共同体。

编写组成员

牵头编制单位：

中国联合网络通信有限公司研究院、中国联通国际有限公司

专家指导委员会：

唐永博 中国联合网络通信集团有限公司副总裁

孟树森 中国联通国际有限公司董事长兼总经理

李红五 中国联合网络通信有限公司研究院院长

陈辉 金砖国家未来网络研究院（中国·深圳）院长

魏进武 中国联合网络通信有限公司研究院副院长

张立仁 金砖国家未来网络研究院（中国·深圳）副院长

李根 量子位智库总编辑

编写组主要成员：

郭耀隆、张肖、栾云、谢泽远、杨锦洲、马瑞涛、吕金宝、丁鼎、傅瑜瑜、余东海、张雅文

参与单位：

金砖国家未来网络研究院中国分院

量子位智库

百度在线网络技术（北京）有限公司

腾讯云计算（北京）有限责任公司

北京智谱华章科技有限公司

商汤科技智能产业研究院

科大讯飞股份有限公司

北京月之暗面科技有限公司

出门问问信息科技有限公司

北京爱诗科技有限公司

上海稀宇科技有限公司

中国移动研究院

北京金山办公软件股份有限公司

参与人员（排名不分先后）：

数字生命卡兹克（张仁杰）、李董、张全文、尹锐哲、周莹、张晋、任志国、张之远、谢奇芳、刘亮、田丰、王波、李健楠、刘晓娟、张云翊、姬润果、张政、杨学燕、谢旭璋、张中汇、张佳欣、马颂、李赓、刘航、王伟、刘远忠、欧琪琪、刘杭

一、生成式人工智能发展迎来爆发期

生成式人工智能的爆发式增长为数字经济的繁荣奠定坚实基础。生成式人工智能以内容生产模式变革催动生产力革新，引领数实融合浪潮下的产业变革，激活数据要素潜能，拓展数实融合空间，促进数字经济与实体经济的深度融合，改善实体经济对于数据资源的应用模式与利用效率，推动实体经济实现数智化转型，保障了数字经济运行的稳定性和可持续性。

（一）生成式人工智能开辟数字经济发展新空间

数字经济的高速发展为人工智能发展创造了良好的经济与技术环境，而人工智能作为关键性的新型信息基础设施，也被视为拉动服务业、工业的新动能。生成式人工智能的价值赋能将主要作用于工业和服务业，农业由于是劳动力密集型行业，以人工程式化作业为主，与生成式人工智能优势相关性暂时较低。¹

1. 生成式人工智能引领产业数字化创新革命

生成式人工智能在研发设计、生产制造、运营管理等环节发挥重要作用，推动了工业企业智慧化转型。生成式人工智能通过其高度自动化的设计和制造流程，正在重塑工业生产方式，它能够优化生产计划，预测设备故障，提高效率和降低成本。此外，生成式人工智能还通过模拟和虚拟测试减少物理原型的需求，缩短研发时间，辅助工人决策，提供实时数据分析，提升供应链的透明度和响应速度。作为工业 4.0 和智能制造发展的核心驱动力，生成式人工智能为工业创新和转型带来了新的机遇。

生成式人工智能以个性化服务、优化客户体验和提高运营效率的方式，全面提升了服务业的面貌。在客户服务领域，生成式人工智能

¹ 来源：《通用人工智能的曙光——生成式人工智能技术的产业影响》罗兰贝格

通过自然语言处理和机器学习，实现全天候智能客服和个性化聊天机器人，快速响应客户需求，提升客户满意度。在金融、教育、医疗等服务行业中，生成式人工智能借助大数据分析，提供了精准的风险评估、个性化学习计划和健康管理方案等。生成式人工智能正推动服务业数字化转型，创新服务模式，提升服务质量。

2. 生成式人工智能推动数字产业迭代升级

生成式人工智能对算力的巨大需求促进云计算市场快速扩展。随着生成式人工智能应用的增多，对计算资源的需求也在不断攀升，这促使云计算服务提供商不断提升其基础和服务能力，以满足高性能计算的需求。生成式人工智能在云平台上的部署和应用，使企业可以按需获取计算资源，有效降低硬件投资成本，进一步促进了云计算市场的迅猛增长。

生成式人工智能驱动了大数据产业的发展和成熟。生成式人工智能利用高效算法和自动化能力，在数据清洗、数据标注和数据分析等关键环节实现了质的飞跃，极大提升数据处理的效率，降低数据处理的成本。生成式人工智能还通过对大数据的创新应用，开发出新的商业模式和产品，推动大数据产业的发展和进步。

3. 生成式人工智能助力社会数字化治理

生成式人工智能推动数字政府服务发展建设。生成式人工智能技术的广泛应用有效促进自动化政府服务平台建设，提高人机交互效率及体验，为智慧城市与智慧乡村的建设提供了强大支持。此外，生成式人工智能还助力完善了数字信息基础设施，让科技进步真正惠及广大群众，提高了社会发展水平和社会治理能力，让人们的生活更加便捷、高效和美好。

生成式人工智能助力完善多方协同治理体系。生成式人工智能技术要求在数据训练、设计、生成等各个环节中都必须遵守合规性、保障数据安全和用户隐私，从而有效促使行业协会和社会各界积极参与人工智能技术管理规范的制定，实现多方协同治理。这不仅为人工智能技术的健康发展提供了监管框架，也使其成为支撑数字经济健康、稳定发展的坚实保障。

4. 生成式人工智能充分激活数据要素潜能

生成式人工智能提升数据要素价值和利用率。生成式人工智能时代强调数据合规性和私有数据价值，促使企业更加重视数据资产管理，通过训练数据自动化生成新内容，将现有数据资源高效转化为有价值的信息，为用户和企业创造全新的价值增长点。生成式人工智能还加速推动数据治理体系的完善，企业通过建立完善的数据治理体系和严格的管控策略，确保数据在其整个生命周期中满足数据安全和合规性的要求，最大限度地发挥数据价值，为企业的持续发展提供有力支撑。

(二) 全球生成式人工智能市场前景广阔

据彭博行业研究相关报告，2022 年全球生成式人工智能市场整体收入为 400 亿美元，预计到 2032 年将超过 1.3 万亿美元，2022 到 2032 年十年间的复合增长率高达 42%。生成式人工智能重塑了传统的内容生产和获取方式，简化了知识获取和创作的难度，激发了数字经济的活力与创造力，正在以创新的模式全面融入各行业各领域。

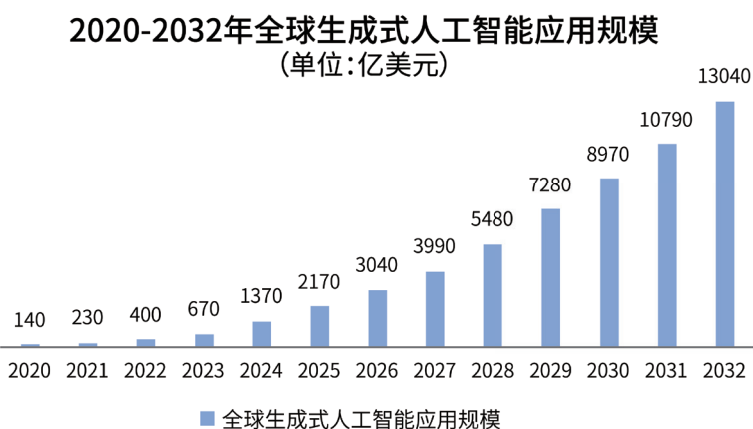


图 1：2020-2032 全球生成式人工智能应用规模 (来源：彭博行业研究)

1. 从决策式人工智能到生成式人工智能，人工智能领域实现跨越式发展

人工智能的发展可以划分为两大主要阶段，以 2012 年为显著的分界点。在 2012 年之前，是决策式人工智能的时代，这一阶段的人工智能技术侧重于分析数据中的条件概率分布，从而进行决策、分析和预测。自 2012 年起，生成式人工智能迅速崭露头角，成为人工智能领域的重要分支，基于大模型、大算力、训练方法等多方面技术突破，生成式人工智能实现了高质量、高效率、多样化的内容生产，其生成的内容形式丰富多样，包括文本、图片、音频和视频等多种格式。

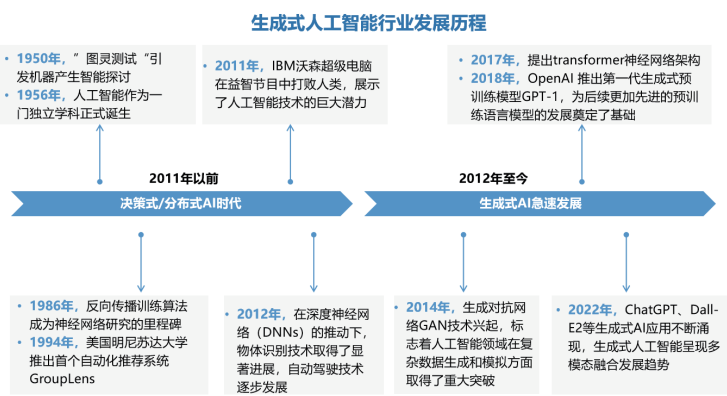


图 2：人工智能技术应用发展阶段 (来源：联通研究院)

具体来看，全球范围内人工智能的发展历程主要经历了四个阶段，**第一阶段是早期萌芽阶段（1950 年代至 1980 年代）**，该阶段主要为人工智能概念的诞生和方法论构建。1950 年，艾伦·麦席森·图灵提出了著名的“图灵测试”，引发了机器产生智能的探讨。1956 年的第一次人工智能研讨会正式宣告人工智能领域的诞生。**第二阶段是技术积淀阶段（1980 年代至 2010 年）**，人工智能研究和应用逐渐深入，神经网络和深度学习技术的发展为后来的人工智能爆发奠定了基础。1986 年杰弗里·辛顿等人提出的反向传播训练算法成为神经网络研究的里程碑。2006 年深度学习技术的兴起引领了人工智能技术的新浪潮。**第三阶段是快速发展阶段（2011 年至 2016 年）**，人工智能技术的实际应用和影响力开始迅速扩大。2014 年对抗式神经网络（GAN）技术的提出以及 2015 年谷歌的 AlphaGo 机器人战胜围棋世界冠军，不仅展示了人工智能在特定领域超越人类的能力，也推动了人工智能技术的广泛关注和研究。**第四阶段是爆发阶段（2017 年至今）**，2017 年 Transformer 架构的提出标志着人工智能进入了大模型时代。2022 年被认为是“生成式人工智能元年”，随着 ChatGPT、Dall-E 2、Midjourney、Stable Diffusion 等生成式 AI 应用的兴起，人工智能在文本、图像生成方面展现出巨大潜力，开启了多模态融合的新篇章。

2. 有为政府和有效市场的有机统一，推动美国在生成式 AI 领域全球领先

美国生成式人工智能发展战略布局呈现出稳定型、渐进性特征，行政治理为先，软法约束为主。在生成式人工智能的发展与治理两方面，美国政府始终坚持发展战略的稳定性和治理策略的渐进性，旨在实现生成式人工智能领域的长远发展和有效管理。2023 年 10 月美国

政府推出首部关于生成式人工智能的监管规定，标志着美国在生成式人工智能领域从过去以观望和探索为主，向积极的引导和规范方向转变，为生成式人工智能的创新应用提供了清晰的发展方向和规范标准，促进技术的健康成长，同时确保伦理和安全标准得到妥善维护。

美国高度重视人工智能技术发展，始终走在生成式人工智能产业浪潮前列，鼓励私营部门积极参与，持续推动技术创新和应用实践。美国在生成式人工智能浪潮中独占鳌头，得益于其强大的产业基础、卓越的创新能力和雄厚的资本支持。美国拥有众多全球领先的人工智能技术供应商和模型开发商，如英伟达等基础层技术供应商通过不断的技术创新，为生成式人工智能技术的发展提供了强大的硬件支持，Open AI 等模型开发商通过开发先进的生成式人工智能模型，推动了 AI 技术在各个领域的广泛应用。此外，美国生成式 AI 企业获得了资本市场的大力支持，数据显示，在全球生成式人工智能的投资流向中，美国初创公司独揽 89% 的资金份额，这一资本优势为美国在该领域的快速发展提供了强劲动力。

美国政府致力于推动数据开放，以增强生成式人工智能大模型的技术和应用。鉴于生成式人工智能对高质量模型数据的极高要求，2024 年 1 月美国政府启动了一项为期六年的国家人工智能研究资源（NAIRR）计划，旨在为美国研究人员和教育工作者提供更丰富的算力和数据资源，从而巩固其在全球生成式人工智能领域的领先地位。联邦政府通过建立数据资源服务平台，整合了来自政府与社会各界的开源数据资源，建立统一的数据汇聚标准，规范数据描述格式，促进不同来源和类型数据的有效融合，为生成式 AI 大模型的优化和迭代提供支持。

3. 其他多国加速生成式人工智能战略布局，发展路径与业务布局各有侧重

欧盟注重生成式人工智能的合规应用，着重强调数据安全和用户隐私保护。欧盟机构在遵循所适用法律的前提下，原则上被允许开发、部署和利用生成式人工智能系统以提供公共服务，但同时必须注重保护用户的数据安全。2024年6月3日，欧洲数据保护监督机构(EDPS)发布了《生成式人工智能与 EUDPR》² 指南，这是首份适用于欧盟机构的人工智能与数据安全指南，为欧盟机构负责任地使用生成式人工智能系统提供了重要参考，指南重点强调数据最小化、数据准确性和数据安全性等原则，并突出了透明度、问责性以及在整个人工智能系统生命周期内维护个人权利的重要性。

英国充分挖掘生成式人工智能潜力，推动生成式人工智能在教育领域的创新发展。2024年2月，英国教育部发布《教育中的生成式人工智能》报告，积极倡导在创建教学资源、辅助学生学习等教育实践中广泛应用生成式 AI 技术，旨在提高教学效率和教育质量。报告同时指出，在引入生成式 AI 技术时需要审慎考虑其可能带来的风险和挑战，包括学术不端行为的风险、数据安全和隐私问题、技术依赖增加以及对教育公平性的潜在影响等，强调必须通过制定相关政策、加强教育部门的监管以及合理应用技术等措施，确保生成式人工智能技术在教育领域的健康发展。

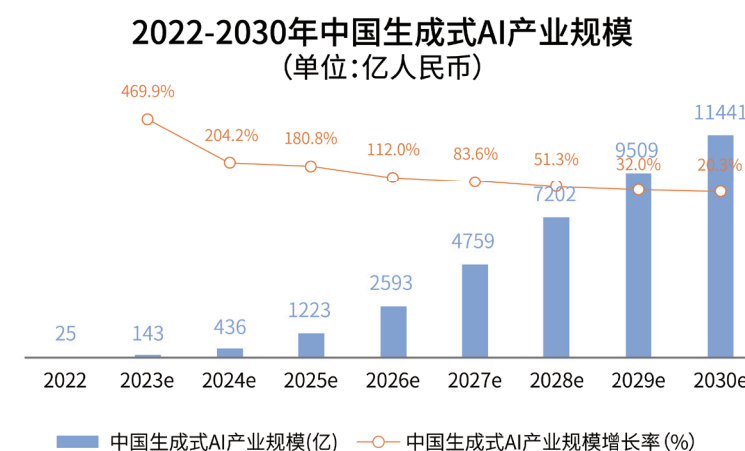
印度积极与全球科技巨头开展合作，对生成式人工智能发展寄予厚望。2023年8月，谷歌在印度推出生成式人工智能搜索服务 SGE，这项服务不仅提供传统的搜索结果链接，还通过智能整合向用户展示更为全面和直观的答案。同年9月，印度 IT 服务公司印孚瑟斯宣布与

² 注：EUDPR 指的是《欧盟 2018/1725 号条例》

微软合作开展生成式人工智能项目，结合双方各自的人工智能技术优势，共同开发行业领先的解决方案。2024年4月，印孚瑟斯进一步扩大了与美国芯片巨头英特尔的战略合作，采用基于英特尔技术的解决方案，助力客户将生成式人工智能技术融入其业务流程，推动业务优化和创新发展。

(三) 中国在生成式人工智能领域加速突破

目前，中国的生成式人工智能市场规模正在不断扩大，根据艾瑞咨询预测，2023年中国生成式人工智能产业规模约为143亿元，预计到2028年中国生成式人工智能产业规模将达到7202亿元，中国生成式人工智能产业生态日益稳固，完成重点领域、关键场景的技术价值兑现，逐步建立完善模型即服务产业生态，2030年中国生成式人工智能产业规模有望突破万亿元大关。



1. 中国生成式人工智能起步较晚，但发展势头强劲

相较于国际人工智能的发展历程，中国人工智能产业起步较晚，最早可以追溯到1978年。随着互联网的快速发展和技术的迭代更新，中国人工智能产业在过去近四十年间经历了从初步探索到加速成长，

再到快速发展的多个阶段。自 2022 年以来，随着一系列国产大模型的推出，中国正式迈入生成式人工智能的黄金发展期，展现出前所未有的活力和潜力。

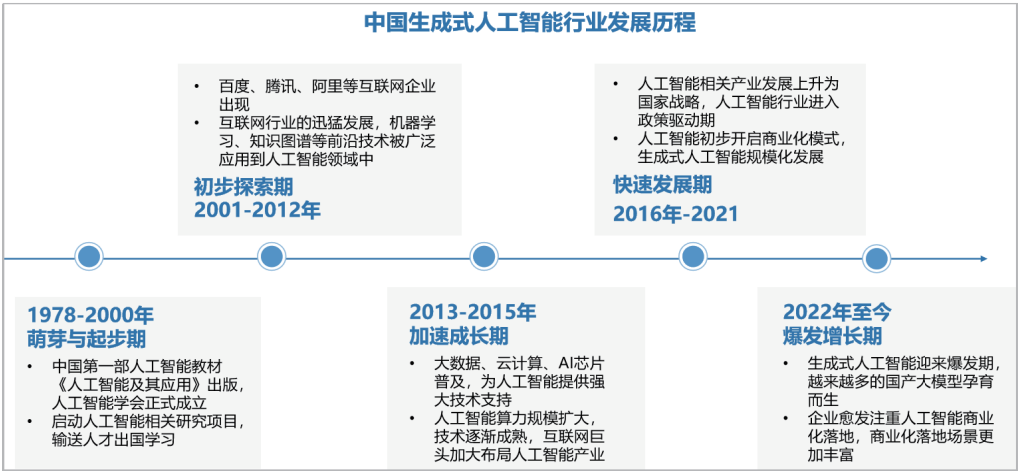


图 4：中国生成式人工智能发展历程
(来源：中国联通研究院)

中国生成式人工智能产业发展可以总结为五个阶段。**第一阶段是萌芽与起步期（1978-2000 年）**，以 1987 年出版的《人工智能及其应用》为标志，人工智能开始在国内进行理论探索和初步实验，但尚未得到广泛推广和应用。**第二阶段是初步探索期（2001-2012 年）**，进入 21 世纪后，人工智能研究获得国家自然科学基金、工信部重大项目基金等各种国家基金计划的大力支持，出现了视觉与听觉的认知计算、中文智能搜索引擎关键技术、虹膜识别、语音识别、基于仿生机器人的人机交互与合作等研究成果。**第三阶段是加速成长期（2013-2015 年）**，得益于大数据、云计算等新兴技术的推动，人工智能在自然语言处理、机器视觉、智能交互等领域取得突破性进展，行业迎来发展机遇。企业开始积极探索人工智能技术在各个领域的应用，推动了行业的快速发展，如百度推出了将深度学习技术应用到病毒查杀客户端的慧眼杀毒引擎，并组建了全球最大的深度机器学习开源平台，

中国科学院与阿里巴巴成立了量子计算实验室，进一步推动了人工智能技术的发展。**第四阶段是快速发展期（2016 年 -2021 年）**，随着国家政策的持续出台，中国人工智能行业进入了政策驱动期。2016 年国务院发布《“十三五”国家科技创新规划》中明确将人工智能列为国家战略层面的重大科技项目。2020 年，“十四五规划”将人工智能定位为“新基建”建设的重要一环，被视为新一轮产业变革的核心驱动力，人工智能迎来新的发展春天。**第五阶段是爆发增长期（2022 年至今）**，生成式人工智能快速崛起，逐渐成为人工智能领域的“顶流”，国内企业纷纷投身于这一创新领域，催生了众多国产生成式 AI 大模型的诞生，截至 2024 年 3 月，共有 117 个大模型成功完成“双备案”³。企业致力于将生成式人工智能大模型技术与实际应用场景深度融合，更加聚焦于商业模式的创新和变现，以实现技术的实际价值和商业潜力。

2. 中国积极推动技术革新与应用拓展

伴随着技术革新与产业深度融合，中国生成式人工智能行业正迅速扩张，市场规模已位居全球前列。得益于政府的政策支持、资本市场的积极参与以及应用场景的不断丰富，中国生成式人工智能产业呈现出迅猛的发展势头。

中国生成式人工智能政策环境持续优化，为生成式人工智能发展提供有力支持。2022 年 11 月，国家网信办等三部门联合发布了《互联网信息服务深度合成管理规定》，这是中国首个合成内容管理政策，旨在依法、合理、有效地使用深度合成技术，并积极防范和化解相关风险。2023 年 7 月，国家网信办等七部门共同发布《生成式人工智能服务管理暂行办法》，这是我国首个生成式人工智能监管文件，标志着“发展”正逐渐成为中国生成式人工智能治理的第一视角，确立了

³ 双备案：“境内深度合成服务算法备案”和“生成式人工智能服务备案”

包容审慎、分级分类的监管基调，提出安全评估、备案以及对生成内容进行标识等要求，有助于推动生成式人工智能技术的规范化应用，并促进整个产业的高质量发展。

中国生成式人工智能投资规模加速增长，推动生成式人工智能企业快速发展。中国生成式人工智能领域正迎来投资的快速增长期，从风险投资到产业资本，各类投资者纷纷布局。据 IDC 报告，2022 年中国生成式人工智能投资额占整个人工智能市场投资总额的 4.6%，预计到 2027 年这一投资比例将达到 33.0%，投资规模有望超过 130 亿美元，五年复合增长率达到 86.2%。中国庞大的市场规模和丰富的应用场景为生成式 AI 企业提供了广阔的发展空间，吸引了更多的资本关注和投入，形成了一个良性互动的投资与创新生态。

中国生成式人工智能应用场景不断丰富，集中于商业化发展成熟的优势产业。国内企业利用本土市场垂直场景的优势，将生成式人工智能技术优先应用于商业化基础设施较完善的新兴行业中，特别是在电商、传媒、娱乐和游戏等领域，例如，电商行业通过数字虚拟人和视频营销提升品牌营销效果，游戏行业则利用生成式人工智能技术开发 AI 陪玩和竞技机器人，并通过自然语言处理技术丰富游戏角色的互动性，增强玩家体验。

二、生成式人工智能产业发展前瞻

生成式人工智能产业大致可分为三个层次，基础设施层、模型层和应用层。基础设施层是生成式人工智能产业发展的关键资源引擎，大算力需求将推动数字基础设施建设。模型层是产业技术变革的原生驱动力，大模型的落地将赋能千行百业，推动传统产业数字化转型，带来新一轮人工智能产业化扩散，MaaS 模式将会持续作为大模型落地的关键业务模式，重构产业链价值流通和技术传递路径。应用层是生成式人工智能产业技术价值传递的实际落位，将通过对内容生产方式和人机交互方式的改变，深刻影响消费与生产生活。

生成式人工智能未来发展分为三波次，**第一波次是探索与初步应用**，未来两到三年，行业发展重点在于算力基础设施和数据基础设施建设，应用场景仍在探索中，C 端应用主要聚焦以聊天助手为代表的通用应用，B 端应用主要应用于金融、法律、教育等行业的知识辅助、客服营销等场景，以语言文字为载体。**第二波次是技术成熟与行业深化**，未来三到五年，C 端应用更加丰富，B 端应用向先进制造业、医疗行业核心场景环节渗透，重点是视觉技术以及动态数据分析管理的应用。**第三波次是普及与创新爆发**，未来五到十年，生成式人工智能与物理世界更紧密地结合，使其能够通过感知和行动来学习和交互，开拓全新的应用领域；同时，在产业侧，成为企业应用的标准必备工具。



图 5：生成式人工智能产业图谱
(来源：中国联通研究院)

(一) 基础设施层：智算设施将是产业发展的基石

基础设施层主要包含算力基础层和数据基础两大层面。基础设施层是生成式人工智能产业的底层架构，提供计算资源和数据资源，支撑上层的模型开发和应用。算力层指的是用于训练和运行生成式 AI 模型的硬件和软件资源，包括 GPU、TPU、高性能计算集群、云计算平台等。数据层涉及数据的收集、存储、管理和处理。数据是训练生成式 AI 模型的关键原材料，数据层包括数据仓库、数据湖、数据标注和预处理工具等。

在算力基础层面，中国在高性能计算和云计算领域取得了显著进展。阿里云、腾讯云和华为云等国内领先的云服务提供商正在大力投

资建设大型数据中心，提供强大的算力资源。中国的超级计算机在全球排名中也名列前茅。在数据基础层面，得益于在大力推进数字中国建设中的成效，在产业数据方面，目前中国拥有较为丰富的高质量产业数据资源，尤其是在交通运输、能源、智能制造等领域；在公共数据资源方面，政府和企业也在积极推动数据共享和数据治理，颁布了《“数据要素 X”三年行动计划（2024-2026 年）》，明确支持人工智能大模型发展。

在基础设施层面，数据训练和推理需求对数据中心能力提出更高要求，同时，巨量算力训练推理需求将促进新型数字基础设施建设。行业发展趋势如下：

趋势一：数据训练和推理需求对数据中心能力提出更高要求，算力供需结构承压

自 2017 年谷歌提出 Transformer 架构后人工智能发展逐步迈入预训练大模型时代。2018 年 Open AI 的 GPT 模型参数量已经达到 117 亿模型参数量开始实现亿级基底的飞跃发展，平均每三四个月即呈现翻倍态势作为数智化时代的资源引擎，随着生成式人工智能时代大模型参数的量级提升，算力供需结构承压持续加大。另外，随着数据量与计算量的飞涨，数据中心的网络带宽也亟待优化，来实现数据在节点内与节点间的高吞吐低延迟的传输与连接。面对算力层的供需结构矛盾，各国需要加快发展算力层基础设施建设，目前部分国家也在积极行动，在 2023 年全球算力指数排名中，美国排名第一，

算力指数 82，拥有最多的数据中心，算力总规模达 270EFLOPS，其中智能算力规模约 180EFLOPS，基础设施支持全球领先；中国排名第二，算力指数 71，算力总规模达 230EFLOPS，其中智能算力达 70EFLOPS，目前中国各地正加快智算建设，持续优化算力资源，满足未来高速发展的大算力需求。

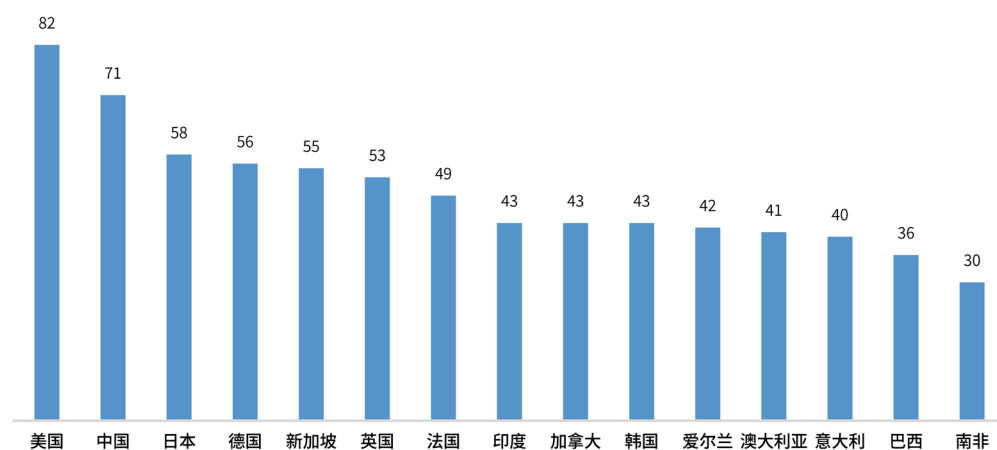


图 6：各国算力指数及排名
(来源：IDC)

趋势二：数据在生成式人工智能的发展中扮演日益重要的作用

生成式人工智能能力的提升主要依赖于高质量的数据集，未来将更加注重数据的准确性和数据来源的多样化。在多模态数据方面，未来越来越多的应用需要结合文本、图像、音频和视频等多种类型的数据。这要求数据收集和生成涵盖更广泛的内容形式，以支持多模态模型的训练和应用。在实时数据方面，物联网设备和传感器的发展使得未被大规模开发利用的实时数据收集成为可能，IoT 设备可以提供实时的传感器数据，如温度、湿度、运动等。这些数据可以用于实时监测和

预测系统中，支持生成式人工智能在智能家居、智能城市等领域的应用。同时提供了更加丰富和动态的数据源，能够实现实时更新和反馈。在合成数据方面，生成式人工智能技术可以用于生成高质量的合成数据，未来可以用于模型训练、测试和验证，减少对真实数据的依赖。

(二) 模型层：大模型商业模式将逐渐贴近一线用户

在生成式人工智能中，“模型层”是构成生成式人工智能系统的核心部分，负责处理和生成数据。大语言模型的核心任务是根据提示词和已有文本，不断预测下一个“合理”的字符。大模型底层是深度学习网络，可以像海绵一样吸收知识，并对知识进行无损压缩。

模型层分为通用大模型和行业大模型两大类，其中通用大模型主要包括闭源模型和开源模型；行业大模型主要包括金融、医疗、法律等领域专业模型。算力规模决定模型的升级迭代效率，大模型训练时长与总算力呈反比，因此具有更高资本优势和算力资源的公司，越有机会领先发布模型更新。目前美国是大模型领域的领跑者，中国起步略慢于美国，目前持续追赶海外进程。

在模型层，大模型发展行业将继续朝着多模态方向进化，MaaS 重构了生成式人工智能产业的商业生态，激发新的产业链分工和模式不断涌现，模型层将更贴近应用层。模型层具体发展主要趋势如下：

趋势三：MaaS 重构人工智能产业链价值流通和技术传递路径

MaaS⁴ 模式降低了应用开发门槛，重构生成式人工智能产业的商业化结构生态，激发新的产业链分工和商业模式不断涌现。未来，

⁴[MaaS (Model as a Service)：是指云计算、智能算力、模型能力等资源做高度融合，客户可以直接在云端调用、开发与部署模型]

MaaS 将成为大模型的主流商业模式。目前，MaaS 模式已成为云厂商的第一战略优先级，将模型作为重要的生产元素。依托于既有 IaaS 设施与 PaaS 平台架构，为下游客户提供以大模型为核心的数据处理、特征工程、模型训练、模型调优、推理部署等服务。未来，顺应大模型开源趋势，MaaS 服务上将着力打造大模型商店平台，发力大模型生态建设，纳入更多允许商用的开源模型，提升平台的基模类型及能力，并丰富工具链产品服务，通过数据回流、模型迭代逐步形成壁垒，形成厂商核心竞争力。

另外，MaaS 模式在 B/G 端和 C 端的商业化落地有所区别。B/G 端需要针对行业领域和业务场景进行大量工程工作，尤其是传统行业的知识获取和积累需要较长时间；即使是同一个行业下，细分场景的痛点不同，AI 大模型渗透率也有明显差异。应用场景碎片化的特点导致低成本、泛化能力较强的能力平台构建需较长周期。C 端用户量巨大，工具使用门槛较低，应用落地以及未来成长速度更快，“应用商店”等创新商业模式不断涌现；通过“对话 + 插件”形式将大模型单点工具接入厂商自有或第三方应用，将打造新的用户交互界面和入口。

趋势四：AI Agent 将为千行百业注入新生产力

AI Agent 是可感知环境及需求、进行决策和执行动作的智能体。相较于大模型底座，靠近应用层的 AI Agent 更具自主性，是提示工程的进化体，可拆解客户指令并自主执行任务。AI Agent 相对语言大模型最大的改进是增加自主行动能力，在实现目标过程中无需人类干预。

目前趋势是 AI 正从简单的工具，进化为复杂的助手乃至 Agent。这一转变的拐点在于 AI 技术在深度学习和自主决策能力的突破，然后将行业垂类数据与业务理解集成到 Agent 框架中，保证大模型应用能够准确理解复杂的任务需求、自主选择最合适的解决方案，并有效控制任务进度，届时就能在各行各业实现智能化转型，推动生产力的指数级增长。未来，AI Agent 可能集成到组织机构的各个层面，与人类员工形成一个互联互通的生态系统，这种深度整合不仅优化了组织结构和提升了运营效率，也将为企业创造前所未有的价值。

(三) 应用层：内容、场景、生态将是应用发展关键

应用层是生成式人工智能技术的具体落地和商业化场景，其包含软件和硬件两大应用场景。生成式人工智能技术在软件层面应用主要分为生产力和泛娱乐两个场景，在硬件层面，终端将逐渐从手机和电脑拓展至可穿戴设备，从自动驾驶的汽车延展至机器人。应用层具体发展主要趋势如下：

趋势五：内容和生态是未来生成式人工智能应用的基石

“内容”和“生态”将会决定生成式人工智能应用的发展，“场景化体验”是用户价值的最终体现。内容赋能模型、场景驱动落地，两者相辅相成。生成式人工智能的能力主要源自于学习和训练数据，因此拥有高质量、多样化的内容资源将是应用发展的关键基础，应用背后高质量的内容生态将是决定能力的核心要素。强大的内容生态不仅为模型训练提供动力，也为定制化应用奠定知识基础，拥有领先内容优势的企业，将在应用拓展上占据先机。在内容基础之上，如何为用

户提供无缝智能体验，是生成式人工智能真正落地的关键。生成式人工智能需要完全嵌入应用场景，高度智能化、自然化地服务用户，最终达到“无感”的体验效果。企业需要“内容生态”、“场景化体验”这两个维度同步布局，才能在激烈竞争中引领生成式人工智能的应用革新。

趋势六：端侧部署应用带来广阔应用场景，推动终端智能化

端侧人工智能的应用将大模型降低参数到百亿级规模，极大降低了边际推理成本，相比公有云，本地部署具备低成本、低时延、高隐私、无地域网络限制等优势。人工智能推理将在手机、便携电脑、汽车等终端上运行。端侧人工智能具有一些独特优势，如本地数据处理效率更高，节省云端服务器带宽和算力成本，对用户数据更好的隐私保护。其中，端侧人工智能应用将为用户带来流畅、便捷和贴心的全新使用体验，成为促进用户消费升级的主要动力。作为新技术应用的风向标，汽车和手机将优先落地端侧大模型技术。在移动设备上，手机端的 AI 助手则提供跨应用的系统级服务，允许用户通过语音输入或其他交互方式与助手沟通。在汽车端侧，用户可以在车载系统中与人工智能助手进行直接对话，享受包括衣食住行在内的全方位服务。

三、中国生成式人工智能的典型实践与应用

随着近年来生成式人工智能的高速发展，人工智能基础设施逐步完善，深度学习技术不断进步，生成式人工智能的应用场景也逐渐增多。

在人工智能时代，数据成为推动其发展的关键基石，而大模型则有效地承载了这些数据。凭借大模型的通用性和可定制性，生成式人工智能得以建立稳固的基础，进而推动其能够在更广泛的领域中进行创新和突破。

生成式人工智能的渗透路径可遵循数字产业的基本发展逻辑，分为 C 端与 B 端 /G 端两大领域。一方面，生成式人工智能在 C 端的应用已逐步走向成熟。各类面向公众用户的 C 端产品通过灵活调用通用大模型 API 形成各种 AI 创作工具，以内容和工具形式触达用户，提供了诸如文本生成、图片生成、音视频生成和代码生成等多种功能，极大地丰富了用户的日常生活和工作体验。另一方面，以 C 端带动 B 端 /G 端的发展模式将成为生成式人工智能的重要演进趋势。得益于生成式人工智能真实性、多样性、可控性及组合性等显著特性，结合大模型能力，生成式人工智能有望逐步替代或补充传统算法小模型和软件功能，实现对各行业的深度渗透。这将助力企业提升生产与办公效率，提供定制化解决方案，如精准营销、客户服务优化等。同时，在高数字化和高内容需求的行业，如医疗、金融和教育等领域，生成式人工智能将持续推动创新发展，引领行业变革。

大模型技术的崛起带动了生成式人工智能的蓬勃发展。在供给侧，

大模型的基础能力将进一步牵引上层应用的开发、开发者工具生态的形成和底层硬件的技术路线；在需求侧，大模型极大提高了聊天对话、文本翻译、搜索引擎以及内容生成方面的用户体验，促进下一代生成式人工智能原生应用在多元场景中爆发。本章节将深入探讨中国生成式人工智能的应用实践，从产品矩阵、服务模式、关键场景应用等多个维度进行全面分析，分享中国生成式人工智能的成功应用经验，为金砖国家提供清晰且深入的了解视角，共同迎接生成式人工智能新时代的到来。

（一）生成式人工智能深入生活场景，重塑智能生活新篇章

生成式人工智能正在深度渗透至人们生活和生产环节中，在需求侧，C端用户对于生成式人工智能的需求呈现出多元化和个性化的特点。无论是智能交互的便捷性、文档制作的效率性，还是代码生成的准确性、决策辅助的智能性，用户都在寻求更加贴合自身场景的应用解决方案。而在供给侧，大模型厂商们正不断深耕细作，基于文本和图像、音频等核心功能，拓展出更加丰富的应用能力。对于大模型厂商和行业参与者而言，如何更好地洞察用户需求、降低大模型使用门槛、把握新兴技术趋势、推动生态共建共享将是未来发展的关键所在。

在日常场景中，以智谱清言、文心一言、Kimi 等为代表的大模型应用，依托自身过往技术积累与高校科研资源，以文本生成与图像生成为核心，展开了一系列开放服务，通过智能体搭建等技术，降低了用户使用门槛，进一步挖掘了生成式人工智能的潜在应用场景，实现

了与大模型外部工具、服务的无缝对接，从而加速了整个应用生态的良性循环与发展。

在办公场景中，生成式人工智能已经涵盖了智能交互、文档制作、代码生成等多个细分市场，为用户提供了全方位的支持。商汤科技的日日新代码小浣熊具备出色的代码生成与修改能力，为用户带来的编码能效提升达到 20%~78%；而金山 WPS AI 等则在多个维度实现了办公效率与质量的飞跃，使用户仅需几十秒就能完成过去要花费几小时的机械性任务，降低了办公领域生成式人工智能的使用门槛。

在娱乐场景中，从图片设计到视频生成，再到虚拟人和 3D 模型生成等前沿技术，每一次技术的突破都为用户带来了全新的视觉体验。以 Pixverse 为代表的视频生成应用凭借其优秀的视频生成能力，可生成高质量、连续性强且风格一致的视频内容，降低了创作门槛，使更多用户能够参与到视频创作中来；而筑梦岛聚焦虚拟人的创新与发展，依托阅文版权小说为主的语料库，使用户可以和自己喜欢的虚拟人物角色深度沟通，满足了用户情感陪伴需求。

1.Gen AI+ 日常：融入生活细节，提升智能体验

生成式人工智能搜索问答功能已在日常生活中发挥重要作用。该功能通过深入学习和分析海量数据，结合全网搜索能力，实现对各类问题的理解与响应，覆盖常识、学术、技术等多元领域。它能够根据用户的查询进行推理和逻辑分析，相较于传统搜索引擎，提供了更为精确且高效的解答。生成式人工智能搜索问答功能的应用范围广泛，

不仅助力学生解决学业疑惑，还为医生提供医疗咨询，协助投资者分析市场动态，以及协助用户应对日常生活中的问题等。此外，该功能具备显著的可扩展性与智能化特性，能够基于用户反馈和数据持续学习、优化，从而缩减用户的搜索时长，提升搜索质量与效率。

生成式人工智能在文学创作领域得到普遍应用。依托自动化、智能化和自适应的文本生成技术，该功能显著提升内容创作的效率与质量。**在激发创作灵感方面**，生成式人工智能写作工具能够自动生成吸引人的文章标题和摘要，提供贴合语境的写作建议，以及基于关键词或主题的内容创意生成，助力创作者突破思维局限，迅速捕捉创作灵感。**在确保文章质量方面**，生成式人工智能纠错工具通过细致的语法检查、拼写检查、标点符号检查以及风格和用词建议，帮助创作者避免常见错误，提升文章的专业性和可读性。**在拓宽创作领域方面**，生成式人工智能协同创作工具实现跨领域知识融合，人机协作，推动创作走向更深层次的创新与探索。此外，生成式人工智能写作功能的应用场景极为丰富，覆盖创意文案、小说创作、旅游指南、科研文献等多个领域，不仅减轻了人力负担，还降低了创作成本，为创作者带来了前所未有的便捷性与可能性。

长文本阅读与处理能力已成为生成式人工智能市场的差异化竞争优势。用户对于文本方面的需求已从短文本获取扩展至超长文本生成、超长文本理解与分析和具备超长记忆的个性化对话等方面，这些需求与实际生活和工作中所面临的问题密切相关。优秀的长文本处理能力

使生成式人工智能系统能够深入理解书籍、研究报告、法律文件和商务材料等复杂信息，为知识密集型工作者提供显著的实用价值。在知识摘要、长篇内容提取、文档分析和语义理解等领域，该功能有效解决用户的迫切需求。鉴于此，中国的生成式人工智能厂商敏锐地捕捉到了这一关键需求，竞相致力于提升大模型在长文本处理方面的能力，以此作为市场竞争的新优势。

典型案例一 —— 智谱清言

智谱清言是基于参数规模超千亿的 GLM-4 打造的智能助手，可在日常生活中为用户解答各类问题。除了在多模态交互上有着杰出表现外，智谱清言还推出了智能体定制功能，用户只需用简单提示词指令即能创建属于自己的 GLM 智能体，进而实现大模型开发的大众化。截至 2024 年 4 月，智谱清言上已有超过千万用户，以及 30 万个活跃智能体。通用问答方面，智谱清言可以实时、准确地回答用户涉及各领域的疑问，为决策提供信息支持。在创意写作辅助方面，智谱清言可激发用户创作灵感，构建文章框架，生成高质量文案，提升写作效率与质量。图像生成与解读方面，智谱清言不仅支持文本生成图片，用户还可以通过输入图像及文本信息，得到视觉描述、图像标注、视觉问答以及视觉定位等多项功能服务。联网增强方面，智谱清言可自动检索网络资源，提供参考链接。文档处理方面，智谱清言可对 PDF 文件进行深度解析，提供检索、总结、改写、翻译等服务，并可对文档结构进行梳理，生成摘要、大纲、关键词分析等，便于用户快速理

解和利用。此外，在视频生成方面，最新发布了清影功能，用户只需输入一段指令或图片，智谱清言即可理解并且生成对应的高品质视频画面。

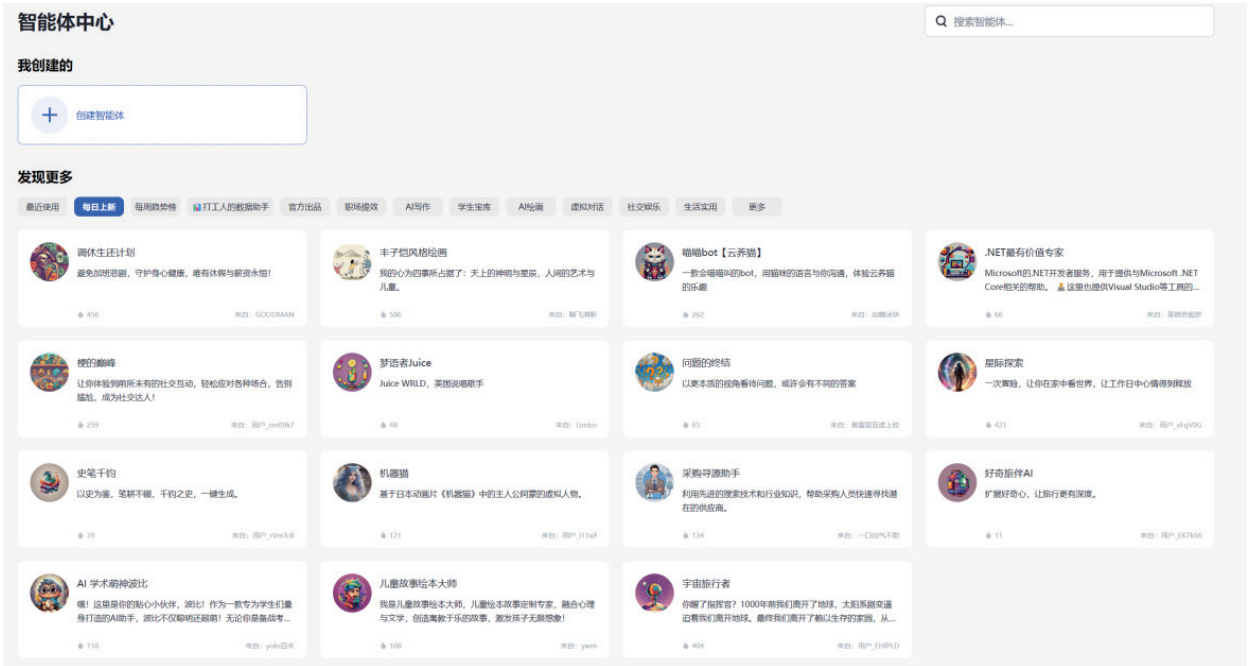


图 7：“智谱清言”智能体中心
(来源：智谱清言)

典型案例二 —— 文心一言

基于文心大模型 4.0 为底座的文心一言是百度全新一代知识增强大语言模型，具备跨模态、跨语言的深度语义理解与生成能力。相较于其他同类型产品，文心一言在提供多种实用插件和易用性方面表现优秀，适合在拓展性要求较高的应用场景使用，为用户提供更加便捷、高效的解决方案。文学创作方面，文心一言可以清晰地表达观点、传递情感，激发创意思维，为用户提供新的灵感和想法，因此可以应用于小说、散文、诗歌等文学作品的创作中。搜索问答方面，文心一言

的聊天机器人可以与用户进行自然语言交互，理解用户的意图和需求，并提供相应的回答和建议。这种应用可以广泛应用于生活服务等领域。多模态生成方面，文心一言支持图像生成和处理，可以根据用户需求生成图像或者对已有图像进行处理编辑。特色应用方面，文心一言与甲骨文信息处理教育部重点实验室共同打造“来自甲骨文的回答”互动程序，用户只需通过输入 prompt 的形式进行提问，就可以收到来自甲骨文的回答。通过调用文心一言的对话能力及对甲骨文文字的释义，让古老的甲骨文“活起来”。截至 2024 年 6 月，文心一言用户规模已超过 3 亿，日调用量已超过 5 亿次。

典型案例三 —— Kimi

Kimi 智能助手是月之暗面公司推出的一款基于自研 Kimi 大模型的应用产品。该产品以千亿级模型参数为基础，以长文本处理为核心能力，凭借其初始阶段约 20 万汉字的无损上下文处理能力，为用户解锁了众多新应用场景。相较于其他同类型产品，Kimi 在文件处理和信息检索方面具有优势，适合用于报告优化、网页内容提取和长文本信息概括等领域。目前，Kimi 已将上下文处理能力提升至 200 万汉字，能够一次性深入阅读 500 份甚至更多文件。在响应速度方面，Kimi 能够在 20 秒内完成对 30 页以内文档的迅速解析，并输出相关文本综述。在长文本概括能力上，Kimi 能够全面把握各篇文章的核心观点，并在处理多篇论文时有效比较和融合相似观点，同时确保输出的文本具备良好的可读性和层次结构。

2.Gen AI+ 办公：创新工作流程，提高生产效率

在数字化转型的浪潮中，传统的办公模式已难以满足日益增长的工作效率和个性化需求。以生成式人工智能赋能办公，正是顺应这一趋势的关键举措。通过集成先进的生成式人工智能技术，能够为各类办公场景注入智能动力，实现代码生成的自动化、数据分析的智能化以及信息管理的精准化。这不仅极大地提升了办公效率，降低了人力成本，还为员工创造了更加灵活、高效的协作环境。生成式人工智能的引入，不仅优化了办公流程，更为企业带来了创新思维和战略优势，推动办公领域迈向全新的智能时代。

生成式人工智能代码编写助手作为高效生产力工具，推动了软件和信息行业的变革。生成式人工智能（Gen AI）编程助手的价值贯穿于开发者工作的始终，从代码的编写、记录到测试，均能发挥重要作用。生成式人工智能（Gen AI）编程助手不仅能通过智能优化、错误检测及代码重构等功能大幅提升软件开发的效率与质量，还能有效减少开发者在编程任务上的时间和精力投入，自动生成代码、自动化处理繁琐的重复性任务，使开发者能够专注于更高层次的设计和创新工作。此外，生成式人工智能（Gen AI）编程助手还可以作为一个强大的学习工具，帮助开发者学习新的软件开发技能。通过 AI 辅助的教程、代码解释和调试工具，使开发者可以更快地掌握新技术和编程语言，从而推动软件开发进入一个更加高效、优质的新时代。

生成式人工智能赋能数据分析与报告生成，实现数据的高效转化

与精准呈现。凭借强大的数据处理能力，生成式人工智能能够迅速分析和解读各类数据，包括结构化的表格数据以及非结构化的文本、图像等多媒体数据。其能够高效地识别、分类和提炼出有价值的信息，从而大幅缩短数据分析周期，降低人工分析成本和出错率。基于分析结果，生成式人工智能能够自动生成清晰、直观且逻辑严密的报告。这些报告不仅包含了详细的数据分析结果，还通过图表、图像等可视化元素，使复杂数据信息易于理解。此外，生成式人工智能还能根据不同的需求和场景灵活调整报告的格式和内容，以满足不同受众的阅读习惯。

典型案例一 —— 商汤代码小浣熊

代码小浣熊是基于商汤大语言模型的软件智能研发助手，覆盖软件需求分析、架构设计、代码编写、软件测试等环节，满足用户代码编写、编程学习等各类需求。代码小浣熊支持 Python、Java、JavaScript、C++、Go、SQL 等 90+ 主流编程语言和 VS Code、IntelliJ IDEA 等主流 IDE，覆盖软件开发全生命周期，包括需求分析、架构设计、代码编写、软件测试、部署上线、系统维护等环节，助力开发者代码生成和补全、代码翻译、代码重构、代码纠错、代码问答、测试用例生成等多个应用场景。在实际应用中，代码小浣熊可帮助开发者提升编程效率超 50%。此外，代码小浣熊可以为用户提供个性化的学习路径，根据开发者的现有技能和学习进度推荐合适的资源，还可以通过分析大量的代码库来提供最佳实践和设计模式的建议，从而提高开发者的

技术能力和生产力。



图 8：“代码小浣熊”软件开发全流程
(来源：IDC)

典型案例二 —— 金山 WPS AI

WPS AI 是金山办公旗下基于大语言模型的人工智能办公助手，在 2023 年 11 月开启公测，锚定 AIGC（内容创作）、Copilot（智慧助理）、Insight（知识洞察）三个方向发展。目前 WPS AI 2.0 已经发布，面向个人用户新增四个 AI 办公助手。在**数据分析方面**，WPS AI 数据助手可在表格中完成海量数据的 AI 计算、分析和归类，用户以自然语言提出需求，AI 可转换写出相应代码，并在表格中执行。在**报告生成方面**，WPS AI 可根据用户的需求和预设规则，自动生成各类财务报表、数据报告等文档。用户只需输入相应的指令或设置规则，WPS AI 便能自动执行复杂的公式计算、创建数据透视表，并生成格式规范、内容详实的报告。此外，WPS AI 还具备智能图表和报告功能，能够根据数据内容自动选择合适的图表类型，并生成直观易懂的图表展示数据。WPS

AI 设计助手则可对已有数据、文字和图片快速完成排版、设计风格和图片处理等方面的工作，例如自动调整文本和图片位置，从此告别繁琐格式设置，一键完成文档排版，让每一篇文档都清晰规范，使报告更具可读性和说服力。

3.Gen AI+ 娱乐：拓展娱乐领域，激发创新活力

在数字经济蓬勃发展的背景下，娱乐产业不仅构建了产品服务与消费者之间的紧密联系，还巧妙地满足了现代人对于归属感的深切需求，其重要性愈发显著。得益于生成式人工智能的强大助力，娱乐行业得以通过音频、视频、虚拟人物等多种形式，有效拓宽其影响范围，以更加亲切、易接受的方式触达消费者，从而注入新的发展活力，推动行业实现持续创新与发展。

生成式人工智能激发用户创作热情，促进音频内容行业革新。音频领域专注于运用先进的人工智能技术和复杂算法来创造声音内容。通过集成机器学习和深度学习算法，AI 音频生成技术能够模仿和再现人类的语音、音乐节奏以及各种声音效果，实现高度自然和逼真的音频产出。随着生成式人工智能技术的进步，AI 音频生成不仅能够精准模仿已知声音，还能创造全新的音频体验。从音乐创作到语音合成，再到音频处理，生成式人工智能不仅丰富了音频内容的创作方式，还提高了制作效率和质量，为音频行业带来了革命性的变化。

生成式人工智能重塑视频内容创作，打造智能技术与创意融合新篇章。在视频生成环节，用户仅需通过文字和图片输入，并设定风格、

运动及比例等参数，便可快速生成准确、一致且生动的高清视频，极大地提升了视频制作的效率和便捷性。在视频编辑环节，借助智能辅助编辑工具，用户可以删除视频特定主体、自动跟踪剪辑、视频特效生成、自动添加、视频美颜等。这些智能编辑功能进一步丰富了视频创作的可能性，使得普通用户也能轻松制作出专业级别的视频作品。

生成式人工智能正成为人类的情感寄托，虚拟社交正在成为互联网新的流量入口。伴随生成式人工智能技术的突破，AI 智能体不仅具有了人格特征，而且逐渐赢得了人们的接纳、喜爱与信任。这为在传统“人 - 人”社交网络基础上，开辟了“人 -AI 角色”颠覆式创新交互模式。通过模拟人类的情感反应，AI 角色能够深入理解并回应人类复杂的情绪，进而极大地丰富了用户的社交体验。在当今各类 ToC 的大模型应用中，访问量排名第一的是主打 AI 虚拟角色聊天机器人的 Character.AI，这个 2021 年创立的 AI 平台上共有 1800 万个聊天机器人。在 2023 年底它的月访问量超过 2 亿次，比同期 ChatGPT 高出 300%，可见年轻网民们对 AI 虚拟角色的喜爱程度。

典型案例一 —— “出门问问” 魔音工坊

出门问问旗下的“魔音工坊”是一款集文案、配音、剪辑全流程一站式 AI 软件。其具有六大核心功能，即软件配音、文字转语音、真人配音、声音商店、效率工具、声音克隆及视频编辑，并拥有有声音克隆、声音搜索、情感合成 / 角色迁移以及声音合成四方面声音黑科技。依托序列猴子大模型，“魔音工坊”可通过 3-10 秒的短音频，快速地实

现声音克隆，并支持跨语言迁移、情感语气生成。其海外版 DupDub 还支持多语音音频生成，目前已涵盖英语、法语、日语、西班牙语、葡萄牙语、泰语等。

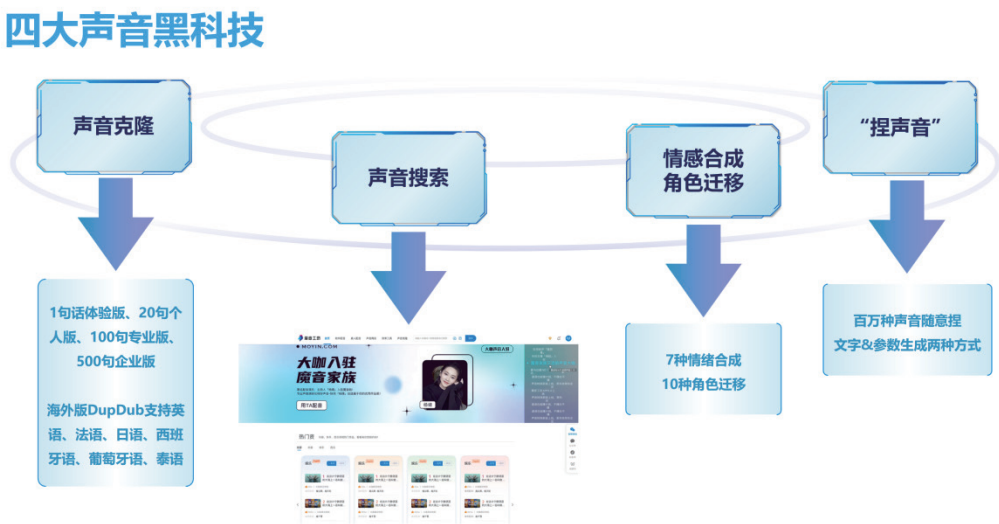


图 9：“魔音工坊”的四大声音黑科技
(来源：魔音工坊)

典型案例二 —— 爱诗科技 PixVerse

爱诗科技旗下的 PixVerse 是一个在线 AI 视频生成平台。该平台以其简洁直观的用户界面和强大的后端技术著称，使得用户无需复杂的操作或专业的视频制作知识即可快速上手。其主要应用场景包括文生视频、图生视频和角色生视频三种模式。在文生视频模式下，用户只需输入提示词、选择风格和设置参数，就能一键生成视频。而在图生视频模式下，则可以将已有的图片或视频素材转化为新的内容，并可运用 Magic Brush 运动笔刷功能，对画面的局部区域进行运动控制，包括方向、运动幅度、角度等，实现更精细的视频编辑。角色生视频 C2V（Character to Video）为全球首创特色功能，用户可以创建属于

自己的“角色库”，选用特定角色连续进行视频生成，根据创作需求将其融入不同场景和故事线。此功能初步解决了 AI 视频创作中的一致性难题，为创作者解锁了更高效、灵活的工作流。

典型案例三 —— 筑梦岛

阅文集团 2023 年发布的筑梦岛 APP 是一款原生虚拟社交生态平台，该平台以商汤科技的“日日新·商量-拟人大模型为基座打造，为用户提供和虚拟人物单聊、小剧场情景对话、围观他人群聊动态、世界观故事模式等有趣易上手的社交体验。该平台从技术角度考虑大众喜爱的 AI 角色，按照“人设”、“形象”、“声音”三个方向深度开发，人设基于商量拟人语言大模型开发出具有记忆、经验、世界观、真实感受的虚拟角色，形象通过商汤秒画文生图大模型生成高质感、多风格的人物外貌图片，声音则依靠语音大模型产生低延迟、高质量的语音对话。截至 2024 年 4 月底，筑梦岛 APP 上共有 15 万个 AI 虚拟角色“梦中人”，即包括影视综艺、网文小说、动漫、游戏、历史人物，也包括 UGC 原创人物、AI 工具助手，其中最受欢迎的热门 IP 角色数量为 300 个。



图 10：“筑梦岛” IP 角色三元素
(来源：筑梦岛)

(二) 生成式人工智能推动行业变革，引领转型升级新潮流

生成式人工智能在向产业渗透的过程中，各行业的接纳与受益顺序依据其信息化基础、数据资源及行业特性而有所不同。互联网与高科技、金融和专业服务行业将是首批受益的领域，这些行业凭借扎实的信息技术基础、庞大的用户数据池，推动生成式人工智能的快速部署。其次是教育、通信、医疗服务、零售和先进制造业等行业，这些行业虽然信息化投入相对较少，但生成式人工智能技术带来的价值空间广阔。此外，农业、建筑业、能源这些传统领域在生成式人工智能技术的应用上尚处于起步阶段，这些行业需要更多时间与资源来探索生成式人工智能的应用场景与价值点。

生成式人工智能在各个行业的应用场景和侧重点存在一定差异。**在通信行业**，生成式人工智能最初被应用于客服领域，旨在节约运营成本并提升服务质量。随着技术的演进，电信领域的生成式人工智能大模型正逐步向其他行业渗透，为垂直行业的数字化转型提供了强有力的支撑。**在医疗行业**，生成式人工智能已广泛应用于诊疗全流程，通过处理海量医疗数据，给出精确诊断意见，并实施个性化治疗方案，既提升了医疗信息化效率，也改善了诊疗体验。**在教育行业**，生成式人工智能正推动行业向个性化与高效化转型，其依据学生特性提供定制化资源与辅导，同时助力教师提升教学质量，促进教育公平化和普及化。**在金融行业**，生成式人工智能能迅速响应金融市场变化，提供准确的风险评估和投资建议，显著提高金融服务的智能化水平。**在工**

业领域，生成式人工智能整合生产、运营和管理数据信息，实现生产流程的自动化和智能化，从而提高生产效率并降低生产成本。

1.Gen AI+ 电信：生成式人工智能助力电信运营商转型升级

在数字化转型的大背景下，中国运营商的第二曲线增势迅猛，新兴业务收入占比持续攀升。电信行业正站在数字经济、人工智能等多领域的新赛道上，进行全面的战略升级。作为数字经济主力军，电信运营商在人工智能时代的共识日益加强。构建坚实的人工智能基础设施，提升生成式人工智能服务和运营能力，是实现智能普惠、智能泛在及智能产业高效规模化发展的基石。

生成式人工智能的应用逐渐成为电信运营商 AI 战略的核心。先进的生成式人工智能大模型能够处理和分析海量数据，提供精准的洞察和预测，从而优化网络运营、提升客户服务体验、增强网络安全，并推动业务模式的创新。通过集成大模型，电信运营商能够实现更高效的数据处理，更智能的决策支持以及更个性化的服务定制，极大地推动电信行业的数字化转型和智能化升级。

生成式人工智能大模型不仅将成为电信运营商提升运营效率和创新能力的重要工具，也将成为推动垂直行业向更高水平发展的驱动力。随着生成式人工智能技术的不断进步，电信运营商将能够更好地利用其在数据、基础设施和用户资源方面的优势，进一步深化生成式人工智能应用，为垂直行业客户实现业务流程的自动化和智能化提供更加丰富和高效的服务。

典型案例一 —— 中国联通元景大模型全面提升 10010 客户服务质量

(1) 技术介绍

中国联通推出元景“1+1+M”大模型体系。首个“1”指一套基础大模型，该基础大模型由两大基座组成。一是语言大模型，面对不同功能需求布局多参数版本，目前已完成布局 10 亿、70 亿、130 亿、340 亿、700 亿等多个参数版本。第二个“1”指一个基础模型底座 -MaaS 平台。元景大模型借鉴人类职业技能形成规律，打造选模型 - 改模型 - 用模型工具链，使能从通用能力到专业能力和“职业技能”的塑造。“M”指 M 种特定行业的大模型，目前基于元景大模型能力，已发布网络专家大模型、客服大模型、反诈大模型等多款行业大模型，应用在工业、城市、交通等多行业超 200 个场景。



图 11：中国联通元景“1+1+M”大模型体系
(来源：中国联通)

(2) 服务模式

从服务模式来看，中国联通以元景大模型为底座，研发了面向内外部业务场景的 10 个行业大模型，赋能网络、客服、终端运营等，助力政务、经济运行、医疗、服装、物流、城市治理领域示范应用。

(3) 应用情况

中国联通充分挖掘生成式人工智能价值，发挥自研大模型优势，在客服大模型基础上打造 26 项 AI 创新服务，赋能其客户联络中心-10010，提供精准化、智能化服务保障。通过全流程、全场景数智赋能，全面提升客户感知、一线效率和管理效能，实现智能服务占比超 85%，平均通话时长压降 15%，首次问题解决率提升 1.6pp，用户满意度达 98% 以上，打造 10010 “一号通解”的智能化服务，助力热线满意度跃升行业第一。

典型案例二 —— 中国移动九天大模型协助 10086 打造智慧客服

(1) 技术介绍

中国移动的九天·客服大模型，依托于客户服务行业深厚的业务知识、丰富交互数据、系统接口信息以及人工坐席的专业经验，采用了科学有效的大小模型融合策略。该模型以自然语言处理为入口，与现有系统无缝对接，实现一体化服务；通过增强意图识别和任务式对话的能力，该模型重塑了以大模型与人工客服协作为中心的客服流程，覆盖了全场景和全流程。这一创新显著提升了用户的满意度和忠诚度，并且通过闭环跟踪技术应用，智能客服系统在问题解决率上也取得了

显著进步。

(2) 服务模式

中国移动九天·客服大模型，既可以独立为用户提供服务，根据用户提供的自然语言描述，解析用户问题并调用 API 获取答案；又可以与人工客服合作，分析历史沟通内容的语义和上下文、总结和归纳对话的重点和关键信息，根据对话内容为坐席提示回复话术，以减少回答的时间和思考负担，形成“大模型—人工坐席—用户”的三方沟通场，极大的提升传统人工客服的工作效率，同时客服大模型将一定程度上重新定义客户服务的内涵和模式，带给用户全新的体验。

(3) 应用情况

依托九天大模型，中国移动成功地将客服行业的服务经验与先进技术相结合，通过“专业+通用大模型”结合、训练和推理计算一体化的深度学习，依托强大的算力中心保障，实现智能搜索、伴随式顾问、多轮交互等多样化的智能交互服务，初步具备了面向客户的复杂场景的识别能力和有效的响应能力。

生成式人工智能大模型应用在智能客服系统前，线上用户解决率约为 46%，通过将原有小模型替换为大模型，并通过一套科学有效的管道方案实现大模型对整个用户的贯穿服务，应用大模型增强意图识别、任务式对话能力，提升文字机器人交互体验，提高用户满意度和忠诚度。通过对客服大模型技术的闭环跟踪，自九天客服大模型上线后智能客服系统意图识别准确率由 92% 提升至 95%，解决率由 46%

提升至 65%，问题解决能力提升带来人力节约 200 人。⁵

2.Gen AI+ 医疗：生成式人工智能引领医疗行业创新发展

医疗产业面临着数据量大、数据质量差、数据隐私保护、医生资源不足、诊断效率低等痛点。生成式人工智能在医疗健康领域的应用，为提高医疗信息化效率、改善在线问诊体验、实现实时监测预警以及助力药物研发带来了革命性的变革，从而为患者提供更优质的诊疗体验与个性化管理。

得益于近年来医疗大模型的不断迭代，复杂的医疗数据分析任务可以得到更有效地处理。鉴于患者行为数据的独特性，生成式人工智能大模型通过个性化设计，满足患者“千人千面”的医疗服务需求，广泛应用于智慧影像、智慧手术、智慧健康等领域。智慧影像覆盖 CT、MR、DR、US、DSA、钼靶等多种医疗影像场景，为患者进行早期检测、诊断及健康风险评估。智慧手术功能则大幅提高患者病情评价的准确性，为术前风险评估、术中手术规划、术后预后估计奠定坚实基础。智慧健康则作为一般患者的贴身健康助手，通过小程序等便捷方式为患者提供高质量导诊服务和个性化健康建议。

典型案例一 —— 商汤大医助力新华医院打造智慧医院

(1) 技术介绍

商汤大医是由商汤科技研发的医疗健康大型语言模型，以商汤自研的千亿参数规模的生成式大型语言模型“商量”为基座（拥有万亿 token 预训练语料），利用超 200 亿 tokens 的高质量医学知识数据训

⁵ 数据来源：《领先企业智能化转型实践白皮书》

练而成，数据范围涵盖医学教材医学指南、临床路径、药品库、疾病库、体检报告等资料，以及 4000 万真实病历、医患问答和对话等。

(2) 服务模式

在服务模式方面，“大医”聚焦智慧大健康、智慧患者服务、智慧临床以及数智建设四大应用领域，已覆盖智能自诊、体检咨询、健康问答、导诊、预问诊、用药咨询诊后随访管理、智慧病历、诊室听译机器人、智慧医助、智慧随访、影像报告结构化及病历结构化共 13 个细分医疗健康场景，实现模型功能与具体场景的精准匹配。为了满足客户差异化部署需求，商汤推出多个“大医”版本，参数量从千亿到百亿级别不等，既可以公有云服务方式为合作机构提供服务，也可帮助合作机构高效完成模型私有化部署。

(3) 应用情况

新华医院基于商汤科技研发的医疗健康大模型“大医”，实现多个医疗场景的无缝融合，通过“智慧医疗”、“智慧管理”、“智慧服务”、“智慧科研”四个板块建设智慧医院，以服务为主线持续推动公立医院的高质量发展。医疗技术与人工智能深度融合，重构了诊前、诊中、诊后三个阶段的服务体验。

在诊前阶段，基于商汤“大医”开发的“智能陪诊助手”为每一位患者推荐合适的医院科室、专病医生，解决“知症不知病”、“知病不知科”、“知科找不对医生”的困惑和诉求。大部分就诊者通常不知道应该挂哪个科，针对此类问题，患者可登录“上海新华医院”

微信服务号，进入“门诊预约 - 智能问诊”板块，就能获得“智能陪诊助手”便捷就医服务，通过文字或语音方式描述自己的症状，即可获得系统推送的相关科室、专病、专家的线上挂号链接。该 AI 助手通过完整学习海量就诊数据，拥有了“智能分诊”的工作能力，同时还“请教”了经验丰富的预检护士们，强化部分患者的常见疾病问题，使得“智能陪诊助手”能够给出智能分诊建议。

在诊中阶段，商汤科技自主研发的 SenseCare® 智慧诊疗平台能够帮助医生显著提升诊疗效率。该平台搭载丰富 AI 辅助诊疗应用，能根据各类检查影像，快速准确地识别微小病灶（如结节），并重建高精度三维器官模型，从而有效助力疾病早筛与精准化个体治疗。在传统方式中，三维重建往往需要耗费大量人工交互操作，给医生带来繁重的工作负担，也限制了精细重建的病例数量，并且在脉管等精细结构的重建效果也存在局限性。SenseCare® 肺部智能手术规划系统，能够实现全自动的胸部多种组织结构的精细分割与三维重建，重建时间由传统医生手工操作的 1-2 小时大幅缩短至 3-5 分钟。助力医生高效、精准、便捷地规划设计个性化手术方案。

在诊后管理阶段，商汤“大医”助力新华医院率先建立“全院级智能随访中心”，灵活、高效、便捷地支撑起三类诊后随访需求。其一是“关怀性随访”，其二是“诊后跟随性随访”，其三是“科研随访”。“大医”随访的优势在于可根据随访表单进行提问，并根据患者回复调整问题，同时耐心解答被访者问题。此外，“大医”可以根据患者

档案以及随访信息给出智能风险评估和随访建议。

典型案例二 —— 百度灵医大模型助力固生堂全流程优化

(1) 技术介绍

灵医大模型是百度集团研发推出的中国首个“产业级”医疗大模型，其以自研国产、知识增强的文心大模型为底座，通过海量医疗健康的优质语料进行预训练和指令精调，以及循证医学对齐后获得的医疗行业大模型。**算法方面**，灵医大模型基于文心大模型，打造千亿参数的大模型，并内置了知识增强、检索增强和上下文增强等多项增强技术，提升了大模型生成的准确性和多样性。**算力方面**，灵医大模型基于万卡算力集群和全生命周期的模型开发工具链的保障，进行无标注的预训练，不断优化底层参数。**数据方面**灵医大模型使用了千亿 Token 的训练语料数据，包括海量临床脱敏数据、海量医学知识图谱、300 万 + 例多模态影像数据，6 亿 + 条健康科普内容，70 万 + 临床试验研究信息，丰富的语料数据提高了模型的全产业链服务能力。

(2) 服务模式

在服务模式方面，灵医大模型可划分为四层架构，按需为不同用户提供服务。**应用层**主要为患者、医院、企业等用户提供 AI 原生应用，目前以灵医 BOT 为助手，聚焦在智能健康管家、智能医生助手、智能企业服务三大方向，满足“医 - 患 - 药”各自的特定需要。**能力层**主要以 API 或 AI 插件的方式，为生态合作伙伴提供高质量的 AI 服务，帮助合作伙伴进行二次开发，打造 AI 原生应用。**模型层**主要根据不同

的应用需要和部署资源，提供旗舰版、Lite 版和定制版服务。**算力层**重点面向有私有化部署需求的用户，提供三个等级的软硬一体的算力支持能力。



图 12：“灵医大模型”商业化服务模式
(来源：百度灵医大模型)

(3) 应用情况

固生堂为客户提供融合线上线下的中医诊疗服务，是中国最大的中医诊疗服务机构之一。百度和固生堂基于灵医大模型，推出“智能健康助手”。这一项目通过整合灵医大模型的能力，实现了**智能分导诊**和**智能问答（客服）**两大功能。**智能分导诊系统**能够根据患者的症状和需求，智能地为其分配最合适的科室和医生，大大提升了患者的就医体验，从数据来看，提高了固生堂患者满意度 12%。**智能问答（客服）**则能自动化处理患者的基础问题如订单配送、发票开具等问题，不仅减少了人力成本，还提高了客服的响应速度和准确性。据统计，

固生堂通过应用灵医大模型，提升了运营人员的服务人效约 76%，极大地提升了整体运营效率。

灵医大模型还通过提供丰富的数据支持，帮助固生堂优化医疗资源配置和医疗服务流程。通过对患者就诊数据的分析，固生堂能够更准确地了解患者的需求和偏好，从而调整科室设置、医生排班等资源配置，提高医疗服务的针对性和效率。同时，灵医大模型还能够对医疗服务流程进行实时监控和优化，确保医疗服务的质量和效率得到持续提升。

3.Gen AI+ 教育：生成式人工智能推进教育行业个性化发展

当今教育行业的传统教学模式存在一定程度上的个性化教育与普惠教育之间的矛盾。个性化教育虽能够根据学生的特点和需求提供精准的教学方案，从而最大程度地发掘每个学生的潜能，但教育个性化的实施需要配套教师资源的持续投入，这对于具有规模化教育需求的国家而言，带来了师资稀缺、教育开支过高等问题。

如何利用前沿科技结合现代教育理论，实现规模更大、质量更高、效率更佳以及更过更优的教育，始终是当代教育研究领域探讨的核心课题。2023 年各类教育行业的生成式人工智能大模型相继发布，为教育行业迈向生成式人工智能之路敞开了大门，有助于师生大幅提升教学效率，并推动大规模因材施教的逐步实现。

典型案例一 —— abab 系列大模型助力高途教育探索教育场景创新应用

(1) 技术介绍

Abab 是由稀宇科技（MiniMax）推出的全栈自研多模态大模型，具有领先的文本和语音能力。今年 4 月，MiniMax 发布了国内首个基于 MoE 架构、包含万亿参数的 abab 6.5 系列模型，最高支持 245k tokens 的上下文长度，实现了高效处理复杂任务的能力。结合语音模型 abab-speech 的应用，MiniMax 致力于在教育领域提升教学效率和质量，促进学生个性化学习和全面发展，为教育行业降本增效注入新动能。

(2) 服务模式

在服务模式方面，MiniMax 聚焦教师与学生两大群体，在课前、课中与课后等不同教学、学习阶段提供多种数字化教学平台与学习助手，在数字化教育的创新探索中实现模型功能与教育场景的精准匹配。通过与教育机构合作调研，MiniMax 研发并提供教案生成、课件生成、课堂助手、作文批改、课堂总结等 AI 助理，辅佐教师完成高质量课堂教学。课堂预习、课堂助手和课后辅导等 AI 工具则能帮助学生群体进一步提升学习效果。

(3) 应用情况

MiniMax 与高途合作，针对不同学龄阶段学生的特点和学习规律，推出了一系列智能助手。这些助手采用先进的文本和语音大模型技术，实现了智能问答、数字课件讲解、数字讲师等数十种精细化的应用场景，为师生提供了丰富的教学交互方式。此外，双方还共同开发了数字讲师、

生涯助手等功能，实现了针对学习内容的问答、聊天、内容生成及规划指导等任务。其中，MiniMax 教师音色复刻等特色功能可根据不同场景定制特色音色，实现重点 IP 音色的高还原度复刻，并与使用者进行一对一交互。Abab-speech 语音模型技术则支持合成更自然的声音、消除机械音，从而提升了学生在高途数字讲师前的学习体验。目前，MiniMax 携手高途共同推出数字老师 30 余名，累计制作超十万分钟 AI 数字课程，辅助生成虚拟资源课件数量破万。

对于处于高考、考研等升学阶段的考生，生成式人工智能也展现出了新的创新应用价值。高途教育利用 MiniMax 大模型和 RAG 全套工具链，结合院校信息、招生信息以及初试复试参考信息等数据，推出了“生涯助手”应用场景。该场景与考生实现智能问答，帮助他们在与大模型的对话过程中逐步选定理想院校和专业，并自动生成针对每位考生的个性化回答和建议。

典型案例二 —— 讯飞星火智慧教育赋能国家开放大学英语教学

(1) 技术介绍

科大讯飞积极推进教育大模型实践落地，从教师备课、课堂教学、学生学习、心理健康、科普教学、素养培养多个应用场景阐述了生成式人工智能在教育数字化转型方面的创新探索。以讯飞星火大模型为底座打造了智慧教育板块，构建了从国家、省、市、县（区）到学校家庭的智慧教育体系。

(2) 服务模式

在服务模式方面，科大讯飞将星火大模型与智慧课堂相结合，从备课、授课、评价三个维度着手，为教师带来更高效、更智能的数字化时代课堂教学新样态。**在备课环节**，提供体系化课程资源，如课件、教学设计、学习单、说课视频、虚拟实验等，帮助老师轻松备课。**在授课环节**，为老师打造一套信息科技教学评管一体化工具，支持实验教学、项目式教学、合作式教学等场景。在评价环节，星火编程助手可以实现编程任务一键批改并给予个性化指导，此外，基于伴随式数据采集，还可以帮助老师智能生成学生学期评价。此外，科大讯飞还充分发挥其在英语教学方面的优势，开发了英语听说教考平台，打通了教、学、考全场景数据，全面分析学生的听说能力。

(3) 应用情况

国家开放大学是教育部直属、以现代信息技术为支撑、面向全国提供开放教育的新型高等学校。其英语学习者每年超过 300 万，学习需求多样，英语教学亟需充分交互、及时反馈的个性化支持服务。

基于讯飞核心人工智能技术及星火大模型能力，国家开放大学打造了个性化语言学习环境，通过英语口语智能训练系统，让每位学生都能即时获得不同维度的训练反馈。基于大模型 + 知识图谱双轮驱动，学校建设了自适应学习平台，并将《学位英语》课程打造为新型数字课程。学生可以结合自身需求开展“千人千面”的自适应学习，并通过课程知识图谱及学习画像，自主寻找知识盲点、查缺补漏，实现针对性提升。借助讯飞虚拟人技术，学校还打造了依托虚拟教师的新型

数字教学模式，帮助教师便捷定制 AI 课程资源，提升资源建设效率；为学生提供 7*24 小时的智能答疑服务，回答过程可同步展示与问题相关的知识点及学习资源，助力拓展练习。经过近万名师生的教学实践，国家开放大学智慧教学资源体系已经形成，共建设智能问答资源库 1 个、自适应课程 1 门，完成知识图谱知识点拆解 200 个、虚拟资源课件 141 个、虚拟资源 844 分钟，打造了开放教育“因材施教”的全新范式。

4.Gen AI+ 金融：生成式人工智能推动金融行业提质增效

金融产业的核心竞争力，在于持续降低金融风险与不断提高行业的服务水平。当前，中国金融服务呈现出服务智能化、业务场景化、渠道一体化、融合深度化等特点。生成式人工智能大模型凭借其强大的理解、学习、生成和推理能力，以及可观的知识容量和卓越的多任务泛化能力，将进一步推动金融服务的智能化升级。

金融行业的前、中、后台业务已全方位受益于生成式人工智能的赋能，显著提升运营效率。对话机器人、虚拟助理等生成式人工智能应用已经逐渐出现在个性化服务、客户服务、金融欺诈检测、信贷支持等服务场景中。在个性化服务方面，基于客户数据的深度分析，生成式人工智能大模型能够为客户量身定制财务和产品计划，实现精准服务。在客户服务方面，生成式人工智能大模型依据客户的行为偏好，生成个性化电子邮件与信息，并提供差异化的智慧客服服务，有效提升客户体验。在金融欺诈检测方面，生成式人工智能大模型助力专业

人员高效检索和分析大量数据，准确识别欺诈行为，保障金融安全。在信贷支持方面，生成式人工智能大模型通过分析海量生产生活和信用数据，为信贷部门人员生成高质量的信贷方案建议，减少贷款收益损失。

典型案例一 —— 腾讯云金融大模型赋能银行开发智慧客服

(1) 技术介绍

腾讯云正式发布的金融行业大模型主要包含四层架构，分别从算力层、平台层、大模型层及应用层助力金融行业提升效率。在**算力层**，该大模型基于 HCC 高性能算力集群、云原生数据湖仓和向量数据库，提供高性能算力和数据接入能力。在**平台层**，腾讯云 TI 平台已实现从数据预处理、模型训练、模型部署的一站式训练，配合训练加速套件，大幅提升模型训练效率。在**大模型层**，金融机构能根据不同细分场景的业务需求，通过模型商店 MaaS 接入腾讯混元大模型、金融行业大模型及 20 多个主流开源模型，通过灵活选择大模型类型降低使用成本。在**应用层**，腾讯云基于金融行业大模型结合金融行业前中后台多个业务场景，打造了一系列智能应用，致力于更好地服务各类金融机构。

(2) 服务模式

目前，腾讯云金融大模型已经在智能客服、智能投顾等场景落地使用。在保险展业场景中，腾讯云金融大模型已经实现较高匹配度并用于具体金融实践，保险代理人开始结合数字人应用场景进行营销展业、业务培训以及智能客服等工作。此外，对于金融行业重点关注的

合规问题，腾讯云基于自身的数据隐私、内容安全解决方案，通过模型私有化部署、权限管控和数据加密等方式，确保金融机构在使用模型的过程中安全合规。

(3) 应用情况

在银行客服场景中，银行的客服系统场景存在三大痛点。一是知识维护量大，冷启动知识配置成本 14 天到 1 个月不等，且需要持续投入运营。二是问答覆盖率低、拦截率低，由于知识边界受限，不在知识库的问题无法回复或者几轮下来往往答非所问。三是接待上限低、服务效率低，进线后坐席需要经历知识理解、搜索、组织回复的复杂流程。

银行客户基于腾讯云金融行业大模型能力，结合自身场景数据，通过腾讯云 TI 平台进行精调，构建了专属的金融客服大模型，并进行私有化部署。通过快速接入银行企业知识，直接学习企业文档库、搜索引擎现有资源，同时直接对接银行 API 进行任务式对话问答，打造了银行专属 AI 助手。银行可采用语音识别、语音合成、人脸识别等 AI 技术，在进行安全认证的基础上，对自然语言进行深度分析，并进行精准回复，让服务“看得见”、“听得见”，大幅减少人工成本的基础上，极大地提升用户交互体验。**一是智能语音导航和智能问答。**通过智能语音导航和智能问答，可以实现对客户的合理引导，将复杂的功能菜单扁平化，提升客服服务效率。**二是智能外呼和智能质检。**一方面，该银行利用腾讯云金融大模型的 NLP、情绪识别、语音识别等

技术，将人工客服的服务录音进行转写，并在此基础上进行数据分析，形成专题分析。另一方面，将外呼营销、催收等过去由人工开展的业务，交由机器人办理，并实时对数据进行深度分析，朝着定制化的客户处理方案演进。三是客服助手。客服助手可以在人工坐席服务时，为员工提供即时的话术支持，也可以根据人工坐席的需求，为其提供即时的协助，提升工作效率。

典型案例二 —— 轩辕大模型驱动度小满金融全场景智能化应用升级

(1) 技术介绍

轩辕大模型是由度小满开源的中国首个千亿级中文金融大模型，其基于庞大的 1760 亿参数的 Bloom 大模型进行训练，通过独创的 hybrid-tuning 的创新训练方式，在金融领域内容理解和生成上能力卓越。截至 2024 年 3 月，度小满轩辕系列已经有 17 款开源大模型，适配广泛场景任务，满足不同开发者需要。在解决实际金融任务的能力上，据评测结果显示，各个参数尺寸的轩辕大模型均具有“以小搏大”的实力，达到了自己 2—5 倍参数量的模型水平。轩辕大模型不仅在金融领域成绩优异，凭借融合百度文心一言的全面能力，轩辕大模型还覆盖了数学计算、场景写作、逻辑推理、文本摘要等多个通用能力维度。

(2) 服务模式

轩辕金融大模型不仅在金融产品营销、金融数据运营、金融风控管理、金融知识问答等关键金融场景中得到了广泛应用，同时也在代码研发、办公协助等多个通用应用场景中发挥着重要作用。自开源以来，

已经有上百家金融机构申请试用“轩辕”大模型。

电销、信审、客服等坐席相关场景是金融业务不可或缺的核心场景，提升坐席作业效率和业务转化率是机会也是挑战。基于轩辕大模型研发的坐席助手可对 CRM 系统客户信息进行摘要，帮助坐席人员提前了解客户，使对话更加有的放矢；开展对话中，坐席助手可实时推荐经营话术，帮助坐席人员把握时机，提升转化率；完成对话后，坐席助手可对通话过程进行摘要，帮助坐席人员复盘作业表现，提升服务能力。

(3) 应用情况

度小满基于轩辕开源模型探索大模型训练和微调技术，最终形成了自己的自研大模型技术栈。在预训练和指令微调阶段，除了通用数据，轩辕系列加入了度小满依托多年行业实践经验构建的海量金融数据，外挂实时更新业务知识库，同时定向增强摘要、逻辑、计算等金融场景核心能力，继而使大模型对金融知识的理解更专业、更精准、更可控、更安全。在此基础上，度小满通过使用人工校验的高质量指令数据对模型进行指令微调和对齐，使轩辕系列具备更加出色的金融能力。

目前，轩辕大模型技术已经应用在度小满各个业务场景，从营销、客服、风控、办公再到研发，已初见成效。在代码助手方面，度小满用轩辕大模型辅助生成的代码，采纳率能够达到 42%，帮助公司整体研发效率提升了 20%；在智能办公领域，轩辕大模型目前的意图识别准确率已达到 97%；在核心的坐席助手应用方面，轩辕大模型也已经取得了一定进展，比如电销坐席助手，目前幻觉问题解决率 96%，转

化率提升 12%；客服坐席助手，服务效率提升 25%；增信坐席助手，增信补件效率提升 37%。

5.Gen AI+ 工业：生成式人工智能推动工业智能化升级

工业发展正加速迈向智能化阶段。继机械化、电气化、自动化、信息化阶段后，工业正处于从数字化向智能化转型的关键时期。在当前阶段，中国在工业领域已积累了海量数据、基础能力和场景需求，借助生成式人工智能的能力，可对工业进行更深入的升级和改造，从而提高生产效率、降低生产成本并提升产品质量。此外，生成式人工智能技术可渗透至工业生产、运营、管理等各个领域，持续优化感知、认知和决策等关键环节，推动工业发展实现“自适应、自决策、自执行”的智能化模式。

生成式人工智能大模型有望为工业智能化开辟新的发展空间。相较于传统小模型，生成式人工智能大模型有望挖掘工业领域智能应用的新场景，提升生成式人工智能应用的普及率。例如，在研发设计阶段，生成式人工智能大模型能够深度挖掘、生成并分析海量数据，为产品设计提供更为精准且富有创新性的思路。在经营管理层面，生成式人工智能大模型可实现对生产流程、供应链管理等各个环节的监控与智能优化，进而提升企业的运营效率和市场竞争力。

典型案例 —— 中国联通元景大模型赋能杭州嘉溢制衣有限公司

（1）技术介绍

中国联通服装制造军团依托浙江联通，承接集团元景大模型在工

业领域助力服纺行业智能制造升级的使命，通过元景大模型底座，提供性能领先的衣裳云生成式人工智能能力平台，汇聚视觉合规、视觉质检、服纺 AIGC 等产品能力单元。

（2）服务模式

中国联通服装军团借助元景大模型能力，已在全国服纺行业成功打造包括浙江的报喜鸟，中国巨石、嘉溢服饰等多个重点标杆项目，服务海宁经编、安徽宿松等多个产业集群平台。

（3）应用情况

中国联通与杭州嘉溢制衣有限公司联合开发定制化元景服装大模型，依托“衣影”大模型的能力，设计师只要在大模型聊天框内输入想生成的衣服款式描述，短短数秒后，穿着该样式服装的模特便呈现在屏幕上，设计师可以用画笔等在模特衣服上进行修改调整，大模型便即时生成了修改后的最新款式。衬衣、长裙、外套，学院风、中式风等各种款式、各种风格大模型都可以生成。除根据文字实时生成款式图样外，“衣影”大模型还可以导入现有的图片，自动进行款式调整。

杭州嘉溢制衣有限公司借助元景大模型能力，从过去一个业务熟练

设计师一天最多设计 3 到 4 款服装，确定一款设计需要两天左右时间，到如今只需 3 秒钟就能生成设计图，且看到上身效果，平均只需大模型生成修改十余次就能得到较为满意的款式，大幅提升了设计师的工作效率。

四、探索“生成式人工智能”未来发展新路径

（一）重点布局算力平台，建立算力合作协同机制

未来，金砖国家在生成式人工智能方面可能面临算力不足的挑战，可支持围绕云计算建设的各类行业训练数据集和人工智能的联合训练平台，形成从理论模型创新、模型工程化到场景化服务的技术和商业闭环，构建统一、开放、有序的人工智能产业大生态。在这一过程中，可考虑从以下几个方面具体展开：

一是推动金砖国家建立区域级、国家级的算力合作调度平台，建立计算资源共享名录，实现跨区域、跨行业的算力资源调配。通过云计算和边缘计算的结合，可以在区域范围内动态分配算力资源，提高资源利用效率。例如，利用分布式计算技术，将各地的超算中心、数据中心进行联通，实现资源的弹性调度和按需分配，满足大规模生成式人工智能模型训练的需求。通过建立人工智能计算资源共享名录，支持超算中心、算力平台、行业训练数据集和人工智能训练平台等基础设施资源的开放共享，避免资源浪费和重复建设。同时，建议金砖多国开放国家级实验室的算力平台支持各类大模型训练。

二是鼓励金砖国家电信企业、互联网公司和科研机构通过合作共建高性能计算平台，形成算力合作联盟。集中已有高等级 GPU 的算力资源，提供大模型数据训练所需算力，通过联合实验室、研究院等形式，促进产学研深度融合，加速前沿技术研究成果的产业化进程。例如，联合开发专用的 AI 加速芯片和高效的计算框架，提升算力平台的性能和效率，

支持生成式人工智能的快速发展。

三是推动绿色数据中心的建设，做好算力和电力协同。采用先进的节能技术和可再生能源，优化数据中心的能源管理，提升算力基础设施的环保性和可持续性。针对地域、时间、天气等对绿电供给影响较大的问题，可通过储能、源网储荷一体化等方法应对。另外，未来金砖国家可考虑部署碳智能计算联合平台，通过获得各国与地区历史、实时或24小时内每小时电力能源结构及碳强度，通过在时间或空间上转移计算任务，实现计算任务与低碳电力供应的最佳匹配。

（二）优化高质量数据供给，促进数据开放与共享

金砖国家大部分国家目前政府侧公开数据和产业专业数据相对缺乏，能体现金砖国家传统文化的源数据暂时不足，金砖国家需优化数据供给，聚焦打造专业数据服务。训练大模型需要优质的大数据集，有些数据还需要人工标注。收集和清洗数据是一项耗时较长的基础性工作，其质量直接决定模型的智能程度。为此，建议加快推动数据资源的整合共享和开发利用。ChatGPT的成功因素之一是拥有高质量语料库和丰富的基础科学数据集。金砖国家具备较为丰富的公共和行业数据。政府、企业和社会多方可综合利用多方资源，构建共享、共创的合作生态，激发数据要素活力。参考操作层面主要包括如下：

一是在政府侧制定公共数据资源开放清单，开展开放试点，对可用于生成式人工智能的公共数据鼓励“应开尽开”。优先开放高价值、低敏感、数据量大的民生公共数据，在数据开放过程中不要过多预设实景

场景，逐步开放公共数据库及行业专业数据库等。政府对于生成式人工智能相关的数据定期维护定期更新、分类、清洗、标注、结构化、并确定分级开放权限，同时为社会提供API数据接口，促进政府数据和行业数据“汇聚融合”。

二是在社会侧，推动共建“金砖价值取向”的源语言高质量语料库，提升优质数据供给。为了让生成式人工智能更好地理解客观世界和掌握规律，需要学习大量来自知识和价值观层的数据。金砖国家拥有源远流长的文化和历史积淀，各国合作共建金砖国家数据库，推动图书馆、出版社和科研机构国际合作，系统地将中文、西班牙文、阿拉伯文、俄文等源语言的历史典籍、主流媒体价值观数据、科研期刊数字化和开放共享，扩大反映“金砖式文化和价值取向”高质量语料数据供给，促进生成式人工智能的发展，推动“人类命运共同体”建设。

三是政府和社会力量协同合作，构建产业数据资源服务平台，创新数据交易机制，促进数据交易流通。政府通过建立数据资源服务平台，汇聚交通运输、能源等多行业数据资源，通过建立统一的数据汇聚标准，规范数据描述格式，促进多方数据融合。社会力量基于政府开放数据及海量网络数据加工形成预训练数据集，在数据流通方面，部分行业数据的非敏感部分可按照“先使用后收益”原则进行交易，在确认使用的情况下，大模型训练可免费使用，在取得商业收益后，按比例支付数据使用费。

(三) 加大基础大模型投入，联合打造行业大模型

“大模型”是生成式人工智能的核心技术之一，其研发需要大量的资源配置，目前金砖国家的基础大模型能力仍有较大提升空间，应集中力量支持人工智能领域的头部企业和研究机构，分工合作，建设有水平、有影响力的底层大模型，赋能多行业，具体如下：

一是在通用大模型方面，政府应制定明确的支持政策，鼓励头部企业、研究机构、高等院校加大对基座大模型研发的投入。可以通过出台专项扶持政策、制定人工智能发展规划等措施，为头部企业提供政策上的保障。例如，设立专门的人工智能发展基金，对参与大模型研发的企业、科研机构、院校给予财政支持。鼓励其在大模型技术上进行探索和创新。通过设立示范项目，展示大模型的实际应用效果，推动技术在全行业的推广和应用。

二是在行业大模型方面，大模型头部企业与细分产业企业在不同领域进行分工合作，分领域联合开发大模型。各企业根据自身的优势和专长，选择不同的研究方向，联合开发有竞争力、有水平的大模型。建立跨企业、跨领域的协作网络，推动大模型的协同研发。通过定期的技术交流、联合项目、共享研发成果等方式，形成良性的合作机制。这样不仅可以加速技术突破，还能避免重复研发，节约资源。

三是金砖国家可建立跨国界的联合研发机构。多国可联合建设生成式人工智能研究中心或实验室，集中资源和科研力量，共同攻关技术难题，该机构专注于人工智能“大模型”的共同研究与开发，作为一个平台，

聚集来自成员国的顶尖科学家、工程师和技术专家，共享设备、资源和数据，协同攻关，以加快“大模型”技术的进步和应用。提高科研效率和成果质量。

(四) 布局内容与生态，以应用带动整体能力发展

金砖国家在应用层面可以走差异化发展路线，通过发展内容、场景和生态，分类施策场景牵引，带动整体能力发展。生成式人工智能的未来竞争力将越来越依赖于其内容生态系统，这不仅包括数据资源和算法创新，还涵盖了广泛的应用场景和合作伙伴关系，构建一个强大的内容生态系统将是生成式人工智能发展的核心。这首先把各行各业的应用做好，C端的大规模用户数据和B端的专业领域数据相结合，不断完善和强化大模型的底层能力和应用效果，通过建立C端超级应用培养用户认知，带动B端应用发展。

一是在C端应用方面，内容、场景和生态是整体发展的三大关键要素，需重点布局。关注用户在C端应用方面的共性需求，分别是内容生成与创意设计、信息提炼与专业辅助、智能交互。在内容布局方面，通过推广智能助手、内容生成工具等应用，吸引大量用户使用并产生丰富的互动数据，帮助优化模型的自然语言处理能力和内容生成能力，通过用户反馈提升模型的准确性和适用性。聚焦用户高频的应用场景，如办公助手、个性化学习辅助、情感虚拟人物陪伴等，开发定制化的解决方案，打造爆款应用，建立用户认知，同时推动行业应用的发展，形成良性循环。

二是在B端行业应用方面，分类施策场景牵引，“小应用，大推广”，

有序推进规模化应用。在重点行业和业务场景的选择上，遵循“**循序渐进、由易到难**”的原则。在重点行业方面，以金融、教育、法律等在客户服务和数据处理有强烈需求的重点行业为抓手，规模化落地标杆应用；在制造业等重资产行业，针对性采集专业数据集逐步进行模型训练开发，推动应用渐进发展。在重点业务场景方面，宜从目前进展最快的营销 / 服务环节向研发 / 设计和生产 / 制造等核心环节渗透，降低生成式人工智能使用门槛，让企业首先通过自有知识库的内容生成和智能对话，提升营销服务活动的效率和用户体验，进而提振信心，推动在生产制造核心环节复杂的时序分析、实时决策等场景的探索和应用。

(五) 实施包容审慎的监管，推动国际合作与创新

生成式人工智能产品作为新生事物，鼓励创新和发展是主旋律。生成式人工智能应用初期可能会“犯错误”，但其自身也逐步建立了技术机制，针对内容、伦理等相关风险进行了过滤与阻断，经过用户反馈、专家机制优化迭代后逐步成熟完善。依托优良的网络环境、丰富的内容治理经验，以及完善的人工智能监管框架，中国及其他金砖成员国完全具备对生成式人工智能应用的治理能力，具体建议如下：

一是建立包容谨慎的监管规范，推动生成式人工智能技术沿着安全与性能 45 度平衡率 (AI-45° Law) 发展。推动可信生成式人工智能技术的发展，安全优先，同时保证生成式人工智能性能长期发展，促进生成式人工智能技术长期沿着 45 度安全与性能平衡发展，强技术驱动、全流程优化、多主体参与以及敏捷治理。例如，设立专门的人工智能伦

理委员会，负责审查和指导人工智能技术的伦理风险；针对生成式人工智能可能带来的个人隐私、内容安全等问题，建立个人数据保护和内容审查机制，保障用户权益和社会安全。

二是建立人工智能金砖合作机制。建立一个涵盖金砖国家的国际科研合作交流机制，各国的科研机构、企业和政府可以分享最新的研究成果和应用经验，促进各国在生成式人工智能技术研发和标准制定等方面的协同发展。同时，金砖国家可设立专门的跨国人才交流计划，鼓励科研人员、工程师和学生在生成式人工智能领域进行互访和联合研究，促进知识共享和技术互通。

三是建立金砖国家生成式人工智能产业联合投资基金。金砖国家可考虑联合设立投资基金，汇集各国政府、企业和资本的力量，提供充足的资金支持，助力生成式人工智能技术的研发和产业化应用。重点投资于生成式人工智能的基础研究、技术创新和产业应用项目，特别是前沿技术和具有高潜力的创新项目。通过基金的支持，可以加速技术突破和应用落地，提升金砖国家在全球生成式人工智能领域的竞争力。

中国有一句古话“合心之力，无所不克；集思之智，无所不成”，这正是我们面对挑战时所秉持的信念。中国愿与金砖各国携手努力，推动生成式人工智能在消费和产业领域应用深度和广度的突破，不断为推动建设开放型世界经济、构建人类命运共同体作出贡献，让世界更加繁荣、让人民生活更加幸福。