

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。

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

2024 中国人工智能 伦理风险白皮书

中国联通研究院 | 中国联通国际公司 · 2024年9月

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、人工智能方兴未艾，风险防范未雨绸缪	4
（一）数字经济发展孕育人工智能全面成长	4
（二）人工智能高歌猛进，融入百业大显身手	6
（三）中国 AI 快速应用，发展举世瞩目	9
（四）伦理规范有利于人工智能长治久安	13
二、他山之石可以攻玉，先进经验学习借鉴	20
（一）国际 AI 伦理研究现状	20
（二）国际 AI 伦理研究贡献及未来走向	21
（三）全球主要国家 AI 政策	23
（四）全球智者的观点	26
（五）全球头部企业的选择和态势	29
三、居安思危如履薄冰，杜绝风险伦理安全	34
（一）伴随社会和技术进步，AI 伦理安全理念与时俱进	34
（二）加强监管，防范失控性风险	35
（三）完善法规，杜绝侵权性风险	39
（四）确保透明公正，规避歧视性风险	42
（五）加强培训教育，缓解人类失业风险	45
（六）促进政策公平和广泛合作，抹平数字鸿沟风险	48
（七）构建指导原则，减少人类道德价值观受损风险	52
（八）展望未来伦理演进，探索 45°平衡律	56
四、金砖国家扩展加盟，伦理安全展望新篇	59
（一）金砖国家应正视 AI 伦理风险	59
（二）金砖国家需合作应对共同挑战	60
参考文献	63
附录	66

前言

2023 年，我国经济社会发展遭遇国内外多重超预期因素冲击，风险挑战前所未有。中国的数字经济克服了重重困难，取得了长足的发展，人工智能更是如耀眼的明珠，在各行各业中广泛应用，凸显了新质生产力对经济的推动、带动了人类文明的发展。人工智能的突破性成就带动了人工智能应用的快速蔓延，数字经济的快速增长有了新的引擎，通信产业等数字基础设施有了长足的进步。同时，数字经济也给人工智能技术进步带来新的需求和宽广的市场。

人工智能已经成为我们生活中不可或缺的一部分。从智能手机、自动驾驶汽车到医疗诊断和金融交易，AI 的应用已经深入到社会的各个角落。然而，人工智能技术的应用在给人们带来生活和工作便利的同时，也带来了一系列伦理风险与挑战。AI 伦理研究，就是在这样的背景下应运而生，旨在探索和定义 AI 的道德边界。

人工智能伦理研究是一个跨学科的领域，它探讨 AI 带来的伦理问题及风险，研究解决 AI 伦理问题，以促进 AI 向善并引领其健康发展。这种研究领域涉及到哲学、计算机科学、法律、经济等众多学科的交汇碰撞。

中国新一代人工智能发展规划推进办公室于 2019 年专门成立了新一代人工智能治理专业委员会，全面负责开展人工智能治理方面的政策体系、法律法规和伦理规范研究和工作推进。这体现了我国对人工智能伦理及其治理的密切关注程度和积极推进决心，同时也突显了这一问题的重要性。

前瞻研究和制定人工智能伦理规范有助于更好地发挥其“头雁”效应，对产业健康发展、人民幸福生活具有重要意义。在智能时代，人与机器的关系将会被重新定义。因此，国务院关于印发新一代人工智能发展规划的

通知中专门强调要探索建立无人驾驶、在线医疗、金融科技、智能配送等监管框架，完善相关法律法规和伦理审查规则。

国家新一代人工智能治理专业委员会 2021 年 09 月 25 日发布了《新一代人工智能伦理规范》，旨在将伦理道德融入人工智能全生命周期，为从事人工智能相关活动的自然人、法人和其他相关机构等提供伦理指引。

本文首先描述了 2023 年人工智能的数字化环境，这既是人工智能发展推动的结果，也是人工智能发展的广阔需求，人工智能的伦理问题，也在其具体的应用中发生了警示和碰撞。接下来本文介绍了常见的人工智能应用，呈现了 2023 年中国人工智能欣欣向荣的社会画卷。本文还介绍了人工智能伦理的内涵、研究范围、国外人工智能的发展情况、各国政策、智者和头部企业等在人工智能伦理方面的观点，以资参考、借鉴。

本文随后具体描述了伴随人工智能的应用产生的潜在人工智能风险，在应用性能效益和风险的二维平面上，我们审慎地认为，兼顾二者的 45 度角发展方向可能会是未来人类的选择，并提出了预防和应对的建议以供读者参考。

最后，考虑到金砖国家人工智能方兴未艾，提出了几点简单的建议，我们相信，在这飞快变化的时代，金砖国家正不断壮大发展，在世界政治经济的地位日益突出，在人工智能这个科技高地上的共识和合作，将有效规避人工智能的伦理风险，极大地促进金砖国家的经济增长和社会进步。

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

编写成员：

康旗、刘亚琪、金春花、吴钢、吕金宝、傅瑜瑜、余东海

参与编制单位：

金砖国家未来网络研究院中国分院

一、人工智能方兴未艾，风险防范未雨绸缪

2023 年人工智能（下文简称 AI）的快速发展，是人类生产力的一次跳跃。一方面，AI 给数字基础设施的建设带来直接的刺激和需求，数字经济的发展空间也得到大大提升，另一方面，数字经济的强大需求拉动了 AI 的研发，数字经济的广泛市场也给 AI 带来一望无垠的试验田，AI 的伦理问题研究就是在这样的背景下展开的。

（一）数字经济发展孕育人工智能全面成长

2023 年是数字中国战略深入实施的关键一年，数字中国建设的基础更加稳固，赋能经济社会发展的效果显著。

1、数字基础设施扩容提速

网络基础设施进入提速升级新阶段。5G 网络覆盖面更广、技术更先进，并加速向重点场所深度延伸。截至 2023 年底，5G 基站数达 337.7 万个，同比增长 46.1%；平均每万人拥有 5G 基站 24 个，较上年末提高 7.6 个百分点；5G 移动电话用户数达 8.05 亿，在移动电话用户中占比 46.6%；5G 虚拟专网数量超 3 万个。基础电信企业 IP 骨干网、城域网、接入网 IPv6 改造全面完成，全国网络基础设施已全面支持 IPv6。截至 2023 年底，IPv6 活跃用户数达到 7.78 亿，移动网络 IPv6 流量占比达到 60.88%，固定网络 IPv6 流量占比达到 19.57%。

算力基础设施达到世界领先水平。截至 2023 年底，中国在用数据中心标准机架超过 810 万架，算力总规模达到 230EFLOPS，居全球第二位。8 个国家算力枢纽节点进入落地应用阶段。算力供给结构逐步优化，

包括数据中心、智算中心、超算中心等多种类型。截至 2023 年底，智能算力规模达到 70EFLOPS，增速超过 70%。全国累计建成国家级超算中心 14 个，全国在用超大型和大型数据中心达 633 个、智算中心达 60 个（AI 卡 500 张以上），智能算力占比超 30%。

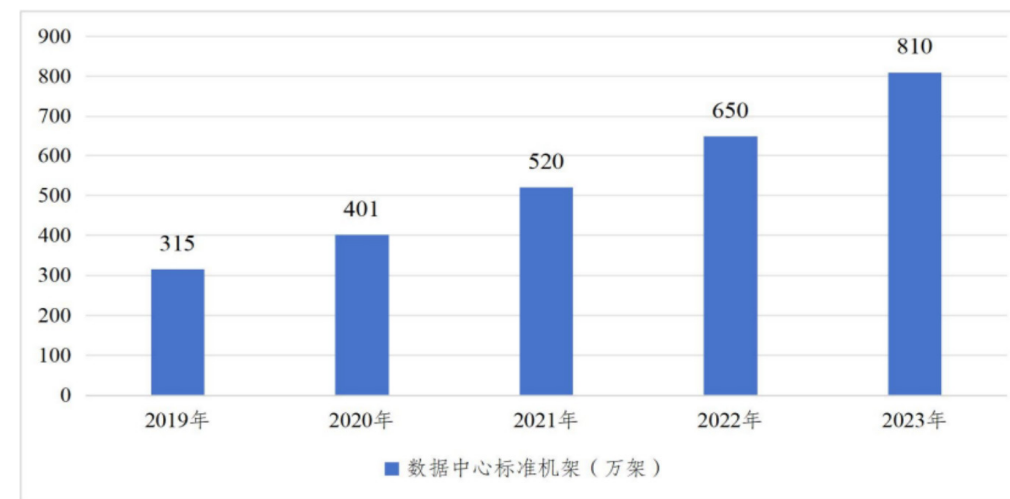


图 1 2019 年—2023 年数据中心发展情况

数据来源：工业和信息化部

2、数字技术创新活力涌现

基础数字技术能力持续增强。数字技术领域保持较高的创新热度。

先进计算、人工智能等关键核心技术不断取得突破。高性能计算持续处于全球第一梯队。智能芯片、通用大模型等创新成果加速涌现。人形机器人进入提速发展阶段，专利申请数量增长较快。人工智能核心企业数量超过了 4500 家。**量子计算机、脑机接口等前沿技术的研发进度不断加快。**

算力网络实现阶段性跃升。积极落实国家东数西算战略，持续完善全国性算力网络布局，截至 2023 年底，三家基础电信企业为公众提供

的数据中心机架数达 97 万架，比上年末净增 15.2 万架，净增量是去年的近两倍，可对外提供的公共基础算力规模超 26EFlops(每秒万亿亿次浮点运算)。

构建人工智能（AI）基础设施体系。内部应用与外部赋能相结合探索 AI 应用新模式，推动“云网智”融合发展，全面向“AI+”战略转变。在网络规划 - 建设 - 维护 - 优化等网络运营各环节引入 AI 和大数据技术，提升网络智能化水平，减低网络故障率、节约基站能耗、优化信号覆盖。三家基础电信企业积极构建“1 个通用 +N 个专用”的通专结合大模型体系，通用大模型达千亿级参数级别，行业大模型可面向政务、应急、文旅等多个领域。

(二) 人工智能高歌猛进，融入百业大显身手

2023 年，人工智能领域发生了许多重大事件，深刻地影响了我们的生活和社会，2023 年被誉为“AI 时代开启元年”，人工智能技术的创新与应用突飞猛进。其中，生成式人工智能的出现是最具代表性的进展，例如 OpenAI 推出的 ChatGPT 订阅版、NVIDIA 的 BioNeMo 平台，以及 AWS 的‘AI Ready’培训计划。这些技术改变了人工智能技术与应用的发展轨迹，加速了人与 AI 的互动进程，是人工智能发展史上的新里程碑。特别是 ChatGPT 的火爆反应，让全球见证了人工智能技术的突破性发展和广泛应用。

以下是人工智能（AI）在全球范围内的应用取得的进展：

1、 AI 与各行各业的融合：AI 技术在金融、医疗、教育、交通等领

域的应用日益广泛，为各行业带来了更高效、智能的解决方案。

2、大模型的发展：随着深度学习技术的不断进步，大型神经网络模型在各领域取得了显著成果，提高了 AI 的准确性和泛化能力。

3、跨领域融合与创新：AI 技术与其他领域的融合进一步加深，如生物学、物理学、心理学等，推动了各行各业的创新发展。

4、强人工智能的发展：AI 逐渐向强人工智能发展，使得 AI 能够理解、思考和感知，甚至在某些方面超越人类的能力。

5、AI 与人类协作：AI 技术正逐渐与人类形成协同作战的局面，如在医疗、法律、教育等领域，AI 辅助专业人士进行工作，提高工作效率和质量。

总之，2023 年以来，AI 在全球范围内取得了重要进展，为人类带来了诸多便利和价值。

AI 能力一览表		
序号	应用领域	说明
1	ChatGPT	主要应用在对话系统中，为用户提供问题解答和信息查询等服务。据统计，ChatGPT 的访问量高达 146 亿次，占总流量的 60% 以上。
2	GAN	主要用于图像生成，可以创建出逼真的人脸图像。
3	语音识别	将人的语音转化为文字，广泛应用于智能助手、语音输入等领域。
4	自然语言处理	使计算机能理解和生成人类语言，被广泛应用于搜索引擎优化、机器翻译等场景中。