聚焦安全治理和人形机器人痛难点
推动人工智能产业高质量发展

人工智能是新一轮科技革命和产业变革的重要驱动力量，加快发展新一代人工智能是事关我国能否抓住新一轮科技革命和产业变革机遇的战略问题。目前，我国人工智能产业蓬勃发展，尤其是大模型涌现后，其强大的创新潜能给千行百业带来了新的发展机遇。但目前国内及北京市人工智能产业在安全治理及人形机器人等重点细分领域仍面临“大模型存在技术漏洞、价值对齐困难及数据价值开发利用不足”、“人形机器人技术路径不确定、通用泛化能力弱、应用场景不足和降本路径不清晰”等问题。

一、国内人工智能产业发展现状及存在的机遇

我国人工智能产业近年来蓬勃发展，北京市领跑全国。据中国信息通信研究院测算，2022 年中国人工智能核心产业规模达 5080 亿元。截至目前，10 亿参数规模以上的大模型已发布近 80 个。据《2022 年北京人工智能产业发展白皮书》显示，截至 2022 年，北京拥有人工智能核心企业 1000 余家，位列全国第一；核心区产业集聚能力全国第一，已经形成全栈式的人工智能产业链；工信部“揭榜挂帅”优胜项目数量居全国第一；人工智能算力发展排名居全国第一。诸多数据表明人工智能已成为北京“强项”，并有望在全国范围内发挥“头雁”效应，

引领“群雁”活力。另外，近五成大模型集中在北京。截至 10 月初，北京发布大模型数量达 115 个，其中通用大模型 12 个，垂类大模型 103 个。

大模型强大的创新潜能给千行百业带来了新的发展机遇。如在人工智能 Transformer 模型引入 CV 领域和**智能驾驶**领域后，智驾能力的提升明显加速；未来，基于 Transformer 有望出现感知决策一体化的大模型，进一步促进自动驾驶领域发展。**AIGC 领域**，代码生成、编程助手、软件助手、操作助手等产品显著提升大企业内部的生产、客户运营、市场营销等环节的效率，有望重塑企业生产力。**机器人**领域，把大模型当做“大脑”来辅助运行的人形机器人技术加速演进，已成为科技竞争的新高地、未来产业的新赛道、经济发展的新引擎，应用前景广阔。

二、国内人工智能产业发展面临的风险和挑战

（一）大模型面临技术漏洞、价值对齐及数据价值开发利用不足等挑战

AI 大模型的发展并非一片坦途，在孕育巨大机遇的同时，也面临着诸多的安全挑战。

一是**面临技术漏洞等内生安全问题和幻觉、滥用等**。当前许多大模型难以确保训练过程中数据质量和清洁度，模型设计的安全性、模型训练的稳定性都存在大的问题。因此大模型存在后门、数据隐私泄露等安全漏洞。例如，模型被黑客攻击、恶意注入病毒等问题时有发生。

代码实现不当也可能导致 AI 大模型出现安全问题，比如有些模型在实现过程中可能存在未经验证的功能或逻辑漏洞，给恶意攻击者留下可乘之机。由于大模型通常能够在互联网上可搜集到的数据上进行训练，因此不可避免地包含用户隐私信息，可能造成隐私泄露。除此之外，预训练大模型还存在大模型幻觉[1]、恐怖等不良信息滥用等新型安全问题，给人们带来潜在的安全威胁。

二是价值观对齐方面也面临两大挑战。一是对齐的基础-人类价值观，是多元且动态变化的。比如在不同的意识形态下，询问 chatGPT 和百度文心一言“中国的经济怎么样”，可能得到不同的回答。甚至模型在训练后会对具有不同宗教、种族、性别等特征的人群产生不一致的结果等。这对很多没有自主意识及判断能力的低龄人群存在危害。二是大模型有用性跟无害性之间目标不是完全一致，也会发生冲突。这导致对齐成为复杂的跨学科研究问题。

三是数据价值开发利用与保护数据安全之间存在冲突。目前国内 80%的数据实际上是在防火墙后面。数据利用新需求、新模式、新业态与保护数据安全之间存在天然冲突，形成了数据利用与保护国家数据资源、保护商业秘密、保护个人隐私三个主要矛盾。如何在不引发新的安全风险、隐私泄漏风险的基础上将这部分数据的价值开发出来仍是一项严峻的课题。目前国内已经在成立国家大数据局的基础上出台了促进数据发挥价值、保

[1] 大模型幻觉问题是指一些人工智能模型在面对某些输入时，会生成不准确、不完整或误导性的输出。

护数据知识产权、隐私以及安全的数据资产入表等相关规划；作为数字经济时代的第一生产要素，数据有望成为政企报表及财政等收入的重要支撑。未来仍需要国家层面出台更多关于数据要素确权、定价、交易流通、收益分配、试点等框架性的规定促进数据价值的开发利用。

（二）人形机器人面临技术路径不确定、通用泛化能力弱、应用场景不足和降本路径不清晰的问题

自 2023 年起，随着以 ChatGPT 为代表的 AI 大模型风靡全球，人形机器人产业加速。10 月，工业和信息化部印发《人形机器人创新发展指导意见》，指出人形机器人有望成为继计算机、智能手机、新能源汽车后的颠覆性产品，发展潜力大、应用前景广，是未来产业的新赛道。但国内人形机器人产业目前仍处于 0-1 的阶段，发展中仍面临着诸多难点。

一是基础技术能力薄弱、技术路径不确定。首先，人形机器人的**外形问题**上产业界仍存在分歧。如有的认为人形机器人发展过程中还可能出现其它中间态，而有的认为通用机器人的终极形态一定是人形机器人，因为其更适应人类社会的环境。外形设计也是机器人研发的基础和关键所在，既要考虑到机器人的独特性，与人类的文化基因、社会背景相匹配，还必须符合机器人本身的结构、功能、使用方法，机器人的安全性、机器人与人沟通等问题。**其次，**从**核心零部件**看，目前国内外人形机器人的技术路径处于百花齐放、百家争鸣的状态。

各科技企业、主机厂在量产前会大量试用各种零部件组装方案，该阶段对各种路线无法证伪。如特斯拉的人形机器人采用了 14 个线性执行器、14 个旋转执行器，采用无框力矩电机、力传感器、行星滚柱丝杠、谐波减速器、行星减速器和空心杯电机等零部件。国内厂家技术路线则不同。但总体来看，国内在核心关键零部件虽然具有了一定的基础，但在人形机器人专用传感器、高功率密度执行器、专用芯片，以及高能效专用动力组件等**核心零部件方面技术仍较为薄弱**。依靠现有的产业链支持，实现人形机器人的产品创新仍具有一定难度。

二是智能化、通用泛化是人形机器人发展的最大难点。长期以来,人形机器人的用户痛点主要集中在人机交互、智能感知、行为控制等方面。目前国内外科技巨头（而非工业机器人巨头）和初创企业在用大模型等技术解决这些问题方面已经做了一些探索，让大家看到了通用机器人的曙光。如决策规划算法是人形机器人智能化的关键，**特斯拉**将自动驾驶的 FSD 系统复用到机器人领域，使机器人拥有了智能属性。国内小鹏汽车也参照特斯拉的思路，迁移其自动驾驶算法、灵犀大模型等进入人形机器人赛道。AI 大模型赋能从技术层面直接提升了人形机器人的人机交互能力，使机器人通用泛化成为可能。例如**谷歌** Palm-e 大模型可将自然语言转化为机器人可执行的任务步骤，从而补全其推理决策能力并泛化至更通用的场景；**Deepmind** 推出了自我改进的 AI 智能体 RoboCat 和 RT-2 模型，大幅提升机

器人对新事物的适应性，并展现出对未出现指令的解释和推理能力。**上海稚晖君“远征 A1”**的 WorkGPT 应用使其具备了理解人的指令并执行任务的能力。但总体在通用机器人“大脑”方面仍有许多待解决的问题，而且国外巨头凭借强大的资金和 AI 能力在该领域展现出更强大的竞争力。

三是应用场景和商业化降本路径仍不明晰。成本下降是推动人形机器人落地的必备条件。人形机器人最早出现在上世纪七十年代，但成本高企一直是其难以商业化的重要原因，如波士顿动力的 Atlas 成本达到 200 万美元，小米 Cyber One 单台造价目前是 60-70 万元人民币。目前已有的人形机器人囿于技术、成本等因素应用场景较为有限，主要在工业场景。未来一方面仍应在技术进步的基础上大幅降低软硬件成本，另一方面仍需寻找能推动人形机器人大规模降本落地的场景，通过爆款场景带动成本摊薄，进而推动人形机器人的商业化和产业爆发。业界预期，人形机器人产业将类似自动驾驶产业发展呈现出类似 L0-L5 等分阶段发展的趋势。目前，特斯拉凭借其优秀的汽车配套产业链、汽车降本经验以及自动驾驶 FSD 算法协同提出使人形机器人降本至 2 万美元，使人形机器人落地出现了可能。国内智元机器人远征 A1 也提出要将整体硬件成本控制在 20 万元以内，并将产品应用于比亚迪的应用场景。但目前两家公司的产品仍均在研发中，落地时间仍需进一步跟踪。其它相关厂商暂时没有提出明确的降本路径。

三、推动人工智能产业高质量发展的对策建议

（一）统筹大模型安全与发展，加大科研投入和技术推广

一是大模型监管要兼顾安全与发展。今年7月，国家互联网信息办公室发布了关于《生成式人工智能服务管理办法》，旨在促进生成式人工智能健康发展和规范应用。这也意味着国家已经开始出手应对大模型火热带来的一系列的安全问题。未来应通过进一步细化大模型相关制度与规则、拓宽管辖范围、加大监督落实力度等方式，为大模型的发展探索划定原则底线。**二是**更前置全面考虑大模型的安全问题，防御思路从“被动"变为“主动”，加大科研投入，完善基础设施，建立示范样本。支持各大安全厂商开发适应新的安全需求的产品。**三是**加强数据资产管理、数据全生命周期管控，提高数据质量，积极寻求多种手段以增强模型的抵御能力，并鼓励在模型的应用和布署阶段实施多重保护策略，推广拟态防御、“AI识别AI”等技术应对大模型的内生安全挑战。

（二）探索保障数据安全的数据资源价值利用，推动数据财政体系构建

一是进一步完善数据要素评估、定价、入表、流通、价值分配的政策工具箱。在政策框架下，鼓励政府、企业、投资机构多方参与数据要素生态建设，推动数据要素产业发展。**二是**选择数据要素市场化运营水平相对较高的地区作为试点，推动政府进行数据要素财政的探索。从国内看，今年11月，衡阳市政务数据资源和智慧城市

特许经营权出让项目交易公告发布，项目起始价为180244.12万元，已经开始了国内公共数据特许经营权出让交易的探索。未来仍将通过试点或者市场化案例的方式探索数据财政的可行性。并利用财政政策逐步引导社会资本参与公共数据运营产业投资和经营。**三是**不断改进和优化数据治理手段，提升数据质量分级分类处理效率，保障数据财政顺利推进和落实。积极探索基于保障数据安全的技术手段在数据财政运行过程中的作用路径。

（三）强化头部引领与产业协同，推动人形机器人技术跃升

一是鼓励科技巨头加紧布局人形机器人“安卓”平台。结合目前国际科技巨头占据产业生态位的实际情况、特斯拉复用自动驾驶FSD系统后的显著优势、人形机器人研发与落地均需要巨量资金等因素考虑，再加上国内科技巨头已在资金、算力、算法、数据、大模型垂直应用场景等方面拥有丰富的积累，建议聚焦政策和资源支持国内在AI领域具有较强实力的科技巨头加紧在该领域布局或进一步加大研发投入，着力打造机器人“大脑”。鼓励巨头企业通过开源平台，加快产品研发和提升。同时，对北京市新涌现的优异的专注开发机器人通用大模型以及致力于将机器人硬件与行业垂直大模型结合的初创和中小企业进行精准扶持，打造该领域优异的鼓励创新环境。**二是**鼓励以科技巨头为引领打造国内人形机器

人创新联合体，以协同创新的方式推动核心零部件攻关和成本降低。引导中小企业关注产业前沿进展，做好公司业务数据积累，主动与科技巨头进行链接，共同研发。三是鼓励多种技术路线研发，建立现有重点企业及新进者的技术路线及产业化进程的跟踪目录，并关注成熟产业迁移到人形机器人产业过程中可能发生的技术复用现象，通过资金扶持、政府采购等方式加速人形机器人产业“0-1”拐点的到来。

（四）通过虚拟仿真开发及场景应用等途径，推动人形机器人降本

一是鼓励打造人形机器人产业虚拟仿真平台或数字孪生平台等，通过对机器人系统进行仿真和编程减少企业研发成本，加快产业技术迭代速度。二是加快探索适合人形机器人落地的爆款场景，尤其是能明显发挥人形机器人优越性而其它机器人不适用的场景等，如机器人长尾场景或者柔性制造场景等，提升人形机器人需求，以应用牵引推动成本下降，进而推动商业化落地。强化人形机器人整机的批量化生产制造能力，持续提升整机产品的质量和可靠性。