CAICT 中国信通院

集智 专题报告

人工智能风险治理报告

——构建面向产业的人工智能安全治理实践方案

（2024 年）

中国信息通信研究院人工智能研究所

2024年12月

版权声明

前　言

人工智能技术以其前所未有的发展态势引领新一轮科技革命和产业变革，深度驱动经济社会发展。技术跃迁释放价值红利，与此同时带来风险挑战。2023 年 10 月，习近平主席提出《全球人工智能治理倡议》，其中着重指出，做好风险防范，不断提升人工智能技术的安全性、可靠性、可控性、公平性。人工智能风险治理已经成为把握未来人工智能发展的核心要素，也是全人类需要共同应对的时代课题。

当前，全球各界积极探索人工智能风险治理之道。本报告广泛汲取国际行动、各经济体努力、产业实践等各方经验，勾勒新时代下人工智能风险认知的演进路径，提炼出人工智能风险认知正向全球性融合、前瞻性考量、系统性分析、交叉性联动不断深化。通过深入分析全球人工智能风险治理多元共治的实践经验，我们深刻认识到，将人工智能风险治理的方案落实到技术革新、产品开发和应用部署中去，是回应社会关切、解决突出矛盾、防范安全风险的必然选择，也是关系到人工智能长远发展的重要议题。

本报告结合国际经验，立足我国产业实践，提出“系统治理-风险识别-风险评估-风险应对”的人工智能风险治理实践框架，实现穿透式风险管控与全链条流程管理的深度融合，为人工智能产业主体对于技术发展与安全保障提供解决方案。鉴于人工智能技术应用日新月异，本报告对人工智能风险治理的认识仍有未尽之处，恳请大家批评指正。

目 录

图 目 录

一、概述

（一）人工智能风险挑战引发全球关注

习近平总书记指出，人工智能是引领这一轮科技革命和产业变革的战略性技术，具有溢出带动性很强的“头雁”效应。当前，人工智能牵引工业、医疗、教育、交通、金融、文化等各领域提质增效，成为社会经济高质量发展的重要支撑和前进动能。然而，人工智能技术在重塑全球经济与社会格局的同时，也带来了多元风险挑战。根据美国斯坦福大学《2024 年人工智能指数报告》统计，2023 年全球共发生 123 起与人工智能滥用相关的重大事件，比 2013 年增长了近 20 倍，并呈现持续增长态势[1]。与此同时，人工智能风险的波及范围和影响程度也显著加大。国际货币基金组织《生成式人工智能：人工智能与未来就业》报告指出，全球 40%的工作将受到人工智能的影响[2]；奇安信《2024 人工智能安全报告》显示，2023 年基于人工智能的深度伪造欺诈暴增 3000%，钓鱼邮件增长 1000%[3]。

人工智能技术释放红利与应用风险显露并存，全球领域人工智能治理进程持续加速。国际组织和主要经济体在人工智能治理方面主动作为，联合国通过联合国大会将共识理念凝聚为决议文件，并通过组建人工智能高级别咨询机构以讨论形成全球合作的具体行动方案；经济合作与发展组织（OECD）研究、评查全球国家如何落实人工智能原则；金砖国家正推动成立人工智能研究小组以促进合作和负责任的

[1] https://aiindex.stanford.edu/report/
[2] https://www.elibrary.imf.org/view/journals/006/2024/001/006.2024.issue-001-en.xml
[3] https://www.qianxin.com/threat/reportdetail?report_id=311

治理；人工智能安全峰会为人工智能安全治理提供全球对话平台。此外，政府驱动的人工智能风险治理逐渐完备。根据加拿大咨询公司 FairlyAI 追踪结果显示，截至 2024 年 12 月，全球人工智能相关政策法规已达 317 份[4]；产业驱动的人工智能标准化行动更加密集，全球陆续出台 355 项人工智能标准[5]，目前尚有更多标准处于规划和制定过程之中。

我国在人工智能治理中侧重风险研判与防范，并积极为国际社会贡献“中国方案”。总体上，中国为全球人工智能治理积极贡献力量。从习近平总书记在“一带一路”高峰论坛提出《全球人工智能治理倡议》，到中国在联合国大会上主提的《人工智能能力建设普惠计划》，包含主权平等、发展导向、以人为本、普惠包容、协同合作等深刻内涵的人工智能治理原则不断传播，为国际社会提供了统一、先进、普适的合作框架与实践指引。**举措上**，中国加快完善人工智能治理体系，2024 年 7 月，党的二十届三中全会《中共中央关于进一步全面深化改革 推进中国式现代化的决定》中提出建立“人工智能安全监管制度”，体现出我国统筹发展与安全，体系化应对人工智能安全风险的谋划。我国已经通过《生成式人工智能服务管理暂行办法》《国家人工智能产业综合标准化体系建设指南（2024 版）》等部门规章与标准规划，保障人工智能技术的健康有序发展。

[4] https://www.fairly.ai/blog/map-of-global-ai-regulations
[5] 秦铭浩 徐慧芳. 全球人工智能标准化进展及我国发展建议[J]. 世界科技研究与发展, 2024, 46(4): 524-535.

（二）人工智能风险治理模式亟需创新

随着人工智能技术的飞速进步，技术风险不确定性增大。自 2022 年 ChatGPT 发布以来，大模型能力愈发突出，人工智能领域迎来了前所未有的投资热潮。然而，随着技术应用复杂度的增加，风险持续显露，亟需采取有效措施应对日益迫切的风险治理问题。在技术快速演变、产品跨境应用、链条主体复杂的背景下，产业界在人工智能风险治理中需要发挥更重要的作用。

一是技术快速演变，产业界提供灵活有效的自律工具能够及时回应现实问题。法律法规的制定因其严谨性，通常需要较长的流程与时间，往往滞后于技术创新。人工智能技术专业性强、演变速度快、原理愈发复杂，而社会对治理措施能够及时回应的需求也日益增长，这进一步加剧了法律稳定性与监管灵活性之间的紧张关系。例如前两年语言大模型、多模态模型取得显著进展，而如今智能体和具身智能等领域的创新突破层出不穷，安全治理的步伐愈发难以跟上技术发展的速度。相较于传统的政府监管，产业界行业规范、标准指南、技术工具更有能力以应对技术发展迅速、风险不断变化的局面，与政府监管形成良性互动。

二是产品跨境应用，产业界在国际治理合作方面需要承担更重要角色。在互联网和跨国企业的推动下，人工智能技术在全球范围内广泛应用，许多风险的跨境性质日益凸显。传统监管模式依赖国家主权进行，易受国界限制，无法有效应对风险的跨境挑战。例如，开源人工智能模型的广泛应用和全球流通性，使得任何安全漏洞、偏见或不

当使用都可能迅速传播，影响多个国家和地区。对于各国政府而言，尽管制定针对开源模型的监管并不困难，但也正是鉴于其全球流通特性，确保监管措施的有效实施却面临巨大挑战。在此背景下，产业组织通过促进跨国合作，例如与其他国家的相关组织开展联合测试、提升工具的互操作性、参与“一轨半、二轨对话”等，推动全球范围内的有效治理。

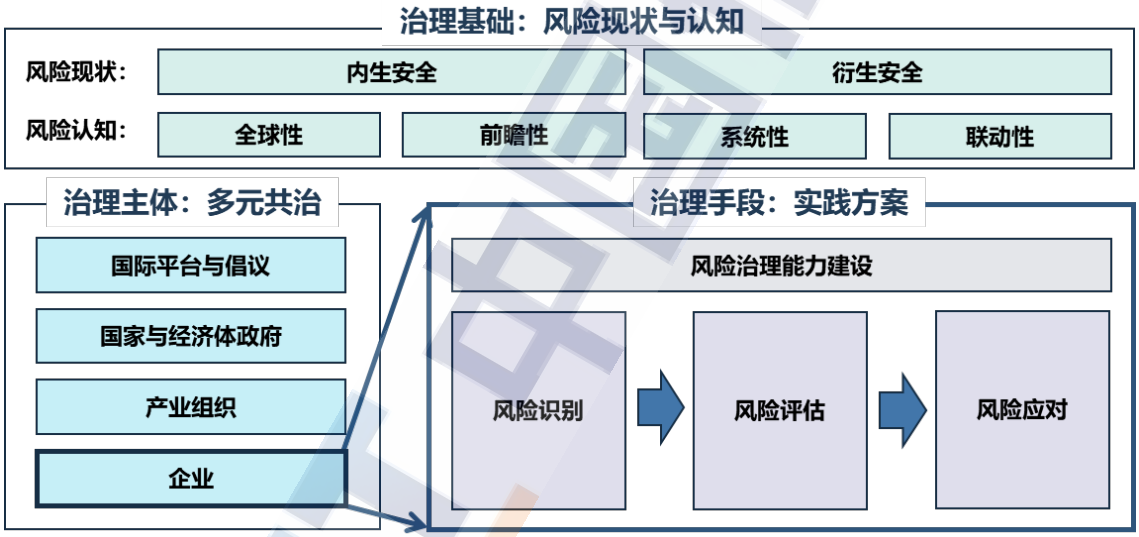
三是链条主体复杂，产业界充分协调人工智能治理多利益相关方。人工智能训练数据、算法模型、系统应用及基础设施等环节涉及多个主体多个维度复杂互动，产业链的延展使得治理结构变得更加复杂，进一步加剧了传统监管模式的挑战。2023 年布鲁金斯学会报告指出，未参加生成式人工智能原始模型开发的“下游开发者”可能会将原始模型调整后再整合到其他软件系统，由于双方均无法全面了解整个系统，或将增加这些软件错误和失控风险[6]。人工智能技术应用涉及数据、模型、软硬件基础设施、应用服务等产业链上下游多利益攸关方，有效的治理措施需要充分协调各方，促进不同主体之间的信息流动与共享，整合各方资源推动解决方案的落地和实施，在共识框架下形成灵活应对地响应模式。

（三）推动面向产业界的人工智能风险治理落地框架迫在眉睫

本报告广泛汲取国际行动、各经济体努力、产业实践等各方经验，形成人工智能风险治理框架（见图 1）。人工智能的迅速发展与广泛

[6] https://www.brookings.edu/articles/early-thoughts-on-regulating-generative-ai-like-chatgpt/

应用所带来的风险，呈现出日益复杂且多维交织的特点。这些风险不仅涉及技术层面的固有问题，还涵盖了应用层面的衍生性挑战。在多元共治的推进下，尽管各国政府在国际和国内层面推出的治理框架或法律规范已取得一定进展，但面对人工智能技术飞速演进和前瞻性风险，仍然存在诸多治理空白与亟待解决的问题。企业作为人工智能风险治理的核心主体，处于治理的前沿，掌握着实施有效治理所需的关键资源，迫切需要一套切实可行的治理方案，以识别、评估并有效应对日益复杂的风险。



来源：中国信息通信研究院

图 1 人工智能风险治理框架

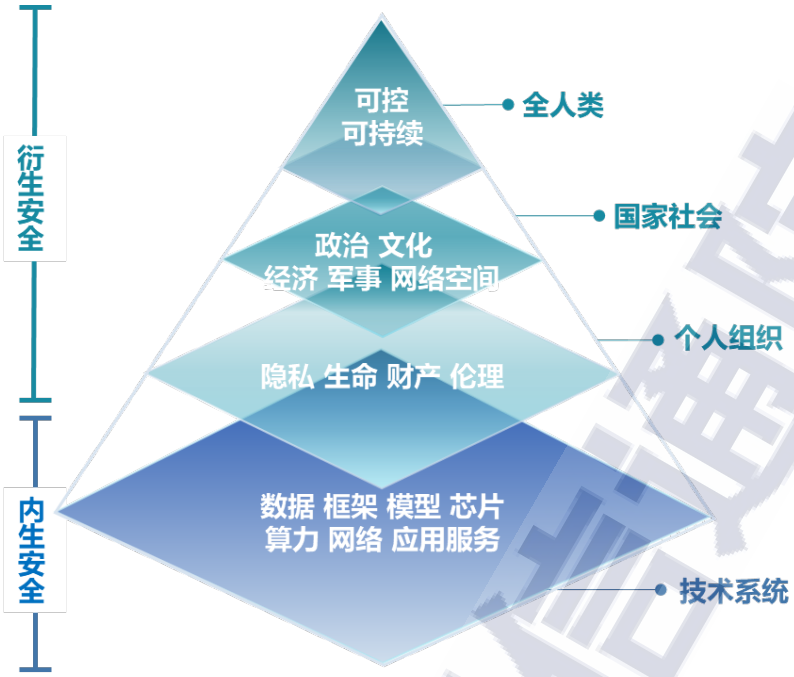
在此背景下，推动面向产业界的人工智能风险治理框架落地已成为迫在眉睫的任务。本报告以安全性、透明性、公平性和可问责性等核心原则为指导，提出了一个集穿透式风险管控与全链条流程管理的落地框架，旨在提升企业人工智能风险治理能力，涵盖从风险识别、评估到有效应对的全过程。**框架强调通过制度完善、文化建设和外部协作来提升系统治理能力**，同时聚焦技术、应用和管理三个层面的风

险识别、评估与应对，构建全链条的治理体系。**在风险识别环节**，框架从自身技术层和衍生应用层两维度着手，通过动态监测和多方协作，及时发现数据质量、算法模型、软件安全以及衍生风险等问题。**在风险评估环节**，通过对风险进行评估量化与优先级划分，帮助合理配置资源。并根据技术迭代和社会变化，动态调整评估结果。**在风险应对环节**，为不同优先级的风险制定定制化方案，并动态优化治理措施。此外，框架明确技术和应用提供方在研发、部署和应用各环节的责任，强调多利益攸关方的合作与信息共享，以提升行业整体风险治理能力。

二、人工智能风险认知渐进深化

（一）人工智能内生安全与衍生安全复杂交织

人工智能在引领科技创新、赋能千行百业的同时，也伴随着多层次安全问题。一方面，人工智能系统自身存在鲁棒性不足以及不可解释性等疑难；另一方面，在人机交互过程中，人工智能内生风险进一步延伸，可能对个人组织、国家社会、全人类带来诸如隐私泄露、就业替代、不可持续发展等衍生安全问题（见图 2）。安全问题的客观存在成为人工智能风险的诱因。系统运行中可能存在的失控或错误决策的内生安全问题，引发操作性风险。隐私侵害和歧视偏见等衍生安全问题，则随着人工智能应用的扩展，可能引发危害性更大的风险挑战。



来源：中国信息通信研究院

图 2 人工智能内生安全和衍生安全

人工智能内生安全主要集中在技术层面，包括数据、算法模型、软件和基础设施等。数据方面，一是有毒的训练数据和不规范的标注可能导致有害输出。训练数据来自现实世界，难免包含有害内容或被“投毒”，数据标注不规范也会降低数据质量，诱发输出错误、生成违法内容以及引发偏见等风险。**二是未经清洗处理的训练数据可能导致敏感信息泄露。**训练数据可能包含未授权的敏感信息，如果数据未经适当保护、过滤或匿名化处理，可能导致敏感信息泄露。2024 年 11 月，X 公司在其隐私政策中声明用户信息会向第三方合作者共享，可能会被第三方用于训练包括生成式人工智能在内的人工智能模型[7]。**算法模型方面，一是算法存在不可解释性问题。**深度学习算法通常包

[7] https://x.com/en/privacy

含数以万亿计的参数，使得即便是技术专家也难以清楚解释其推导过程。当人工智能造成损害时，不可解释性使得责任难以明确，影响可问责性的有效实施。**二是模型存在被攻击风险。**通过提示注入、输入扰动等对抗性攻击方法，模型可能被诱导生成错误或有害结果。同时，模型核心结构、权重参数和偏差等关键信息可能被窃取、复制或篡改，导致知识产权和商业秘密泄露，破坏市场公平竞争秩序，损害科研动力。**此外，与模型训练紧密相关的算力、网络、软硬件方面存在被后门攻击等风险。**2024 年 10 月，包括 ChuanhuChatGPT、Lunary、LocalAI 和 DJL 在内的多个开源模型被曝存在安全漏洞，其中一些可能导致远程代码执行或信息盗窃[8]。

人工智能衍生安全重点体现在应用层面，涉及个人隐私、组织财产、社会公平和国家安全等领域，并可能对全人类产生深远影响。个人组织层面，一是不当使用可能加剧信息泄露风险。随着人工智能的广泛应用，个人和组织需要提供大量数据，不当输入可能危及数据安全，导致个人隐私和商业秘密泄露。**二是**可能引发法律风险，如侵犯知识产权，或被用于犯罪、制造虚假信息等，同时也存在诱发伦理风险的可能性。2024 年 2 月，香港发生一起利用人工智能深度伪造技术的诈骗案，涉案金额达 2 亿港元[9]。**国家社会层面，一是**可能对社会舆论和政治安全带来挑战。通过生成虚假信息和有害言论，人工智能可能扰乱社交媒体秩序、破坏公众信任、威胁社会秩序，甚至干涉他国内政与社会稳定。**二是**可能带来就业结构深度调整。人工智能的

[8] https://ourcoders.com/news/show/18647/
[9] https://www.stcn.com/article/detail/1116440.html

广泛应用可能引发部分就业岗位的消失和新兴就业形态的出现，带来劳动力结构调整。2024 年 1 月，国际货币基金组织（IMF）发布报告表明，在新兴市场和低收入国家，受人工智能影响的就业岗位比例预计分别为 40%和 26%[10]。**全人类层面，一是**引发能源担忧，人工智能的性能突破需要强大的算力支持，但模型训练过程中却导致了巨大的资源浪费和碳排放水平的上升。**二是**核武器、生化武器等两用物项的能力可能被滥用，威胁全人类安全。2024 年 9 月，OpenAI 承认，其最新模型 o1“显著”增加了人工智能的能力滥用风险，在涉及化学、生物、放射性和核武器等问题时存在最高等级风险[11]。

（二）人工智能风险认知迈向多维度深化融合

人工智能风险的认知向全球性融合、前瞻性考量、系统性分析、交叉性联动不断深化。这一转变彰显了人类在全面理解和应对人工智能风险的不懈追求，也反映了人工智能风险的最新动向（见图 3）。

来源：中国信息通信研究院

图 3 人工智能风险认知趋势

[10] http://www.xinhuanet.com/20240115/66fe98299e054bf08a5d77aa9ed8325c/c.html
[11] https://www.c114.com.cn/ai/5339/a1273371.html

总体上，社会各界愈发关注人工智能风险全球性。多项国际人工智能治理方案愈发深刻认识到，人工智能技术的应用与影响并非局限于某一地区或国家，而是深刻影响全球社会、经济和政治秩序。**具体而言，人工智能的开发、部署和使用具有全球化特征。**大型跨国科技公司主导的人工智能技术能够快速传播到世界各地，进而引发风险跨境扩散。**《全球人工智能治理倡议》**提出人工智能技术带来难以预知的各种风险和复杂挑战，明确指出“人工智能治理攸关全人类命运，是世界各国面临的共同课题”[12]。**《布莱切利宣言》**关注识别全球共同关注的人工智能安全风险，提出“在更广泛的全球背景下审视人工智能技术对社会、经济及人类生活的潜在影响”[13]。**《首尔宣言》**提出人工智能可能被跨国犯罪集团利用，用于网络诈骗、操控选举或传播虚假信息，这些问题只能通过全球协调加以解决[14]。

国际倡议和报告日益关注人工智能发展带来的前瞻性风险。人工智能的技术发展迅速，潜在风险难以准确预测，甚至是其技术能力超出当前认知范围。**一是人工智能治理框架日益强调对高优先级风险的评估和理解。《布莱切利宣言》**指出，前沿模型潜在的滥用或意外失控问题可能带来严重甚至灾难性的风险[15]。英国人工智能安全研究所则将前沿模型的自主性和滥用作为人工智能的核心风险类别。**二是人工智能治理方案逐渐将未来风险研判纳入认知议程。**联合国《治理人

[12] https://www.mfa.gov.cn/web/ziliao_674904/1179_674909/202310/t20231020_11164831.shtml
[13] https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023
[14] https://www.gov.uk/government/publications/seoul-declaration-for-safe-innovative-and-inclusive-ai-ai-seoul-summit-2024
[15] https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

工智能，助力造福人类》最终报告预判人工智能风险演变方向，指出约七成受访者担忧未来 18 个月内人工智能损害范围和严重程度将大幅增加。OECD**《评估未来人工智能的潜在风险、效益和政策》报告**确定了 38 项潜在的未来人工智能风险，并分析确定了 10 项需要加强政策关注的优先人工智能风险，并对未来治理机制和机构无法跟上人工智能快速发展表达担忧[16]。

各国机构日益加强对人工智能风险认识的系统化。各国组织机构日益认识到人工智能风险涉及技术、社会、伦理等多个层面，需综合考量，强调风险的多维特性。**美国国家标准与技术研究院（NIST）《人工智能风险管理框架》**（1.0）建立全周期、多层次风险管理方案，强调风险来源覆盖设计、开发、部署、运行和退役全部五个阶段，并从个体模型风险到生态系统风险，涵盖了人工智能对单一系统和整个技术生态的影响[17]。**德国联邦信息安全办公室（BSI）《生成式人工智能模型指南》**全面梳理生成式人工智能各项风险，通过对不同风险场景的分析和分层管理，全面响应了偏见、滥用风险及外部攻击等不同层面的安全挑战。**中国信息通信研究院《可信人工智能白皮书》**系统性分析了人工智能的多重风险，指出随着技术发展，算法安全、不透明、数据歧视、责任界定困难和隐私泄露等问题日益复杂化[18]。

专家学者日益强调对人工智能风险认知的联动性。越来越多的顶尖人工智能专家在公开场合强调，不同类型的人工智能风险往往相互

[16] https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/11/assessing-potential-future-artificial-intelligence-risks-benefits-and-policy-imperatives_8a491447/3f4e3dfb-en.pdf
[17] https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf
[18] http://www.caict.ac.cn/kxyj/qwfb/bps/202107/P020210709319866413974.pdf

交织，不能孤立地应对与管理，因此亟需加强风险认知的联动性与整体性。由 75 位顶尖人工智能专家参与编写的**《先进人工智能安全国际科学报告》中期报告**将交叉风险因素视为通用人工智能的风险根源，认为这些因素增加了多种风险的可能性和严重性。从技术角度看，系统可靠性难以保障，内部机制缺乏透明性，且人工智能可能在减少监督下拥有更大自由；从社会角度看，技术发展超前于监管，人工智能系统研发方在竞争压力下可能忽视风险管理，推出有害产品[19]。由 3 位图灵奖得主和 25 位权威专家在《科学》（Science）期刊联合发表的**《人工智能飞速进步背景下的极端风险管理》文章**指出，人工智能带来的风险不仅相互交织，还可能在多个层面产生连锁反应。先进的人工智能不仅可能加剧社会不公、破坏社会稳定，还可能推动自动化战争和犯罪活动。随着公司开发先进人工智能，现有风险可能被放大，甚至产生新的风险[20]。

三、人工智能风险治理全球实践

（一）国际层面：加速推进风险治理行动及合作

国际组织推进从风险认识到风险治理进程。联合国组建“人工智能高级别咨询机构”，发布《治理人工智能，助力造福人类》中期及最终报告等全球人工智能治理框架。联合国将成立一个独立的国际人工智能科学小组，联合其他国际组织研究人工智能风险等议题。此外，联合国大会先后商讨通过两项决议，提出采取国际互操作性强的措施，

[19]https://assets.publishing.service.gov.uk/media/666ac68ca8b7ec4fae43d0f9/international_scientific_report_on_the_safety_of_advanced_ai_executive_summary_chinese.pdf

[20] https://news.qq.com/rain/a/20240526A01TUD00?web_channel=wap&openApp=false&suid=&media_id=

涵盖识别、分类、评估、测试、预防和缓解措施。**金砖国家**开展多形式人工智能治理合作，启动人工智能研究组、组织金砖国家人工智能技术与治理卓越人才研修班、举办技能发展与技术创新大赛，形成了多样化的合作方式。**经合组织**推进人工智能规则适应落地。2024 年 5 月更新《人工智能原则》，指导政府更新人工智能风险认知并实际吸纳人类监督、多方合作和互操作治理等相关治理原则。**世界互联网大会**构建常态化治理架构。世界互联网大会在组建人工智能工作组、发布《发展负责任的生成式人工智能研究报告及共识文件》的基础上，于 2024 年 11 月的乌镇峰会上，正式成立人工智能专业委员会，常态化推进人工智能标准化、产业发展和安全与治理工作。

国际峰会推动紧密合作以强化人工智能风险治理。2023 年 11 月，英国召开首届全球人工智能安全峰会，通过《布莱切利宣言》，就前沿人工智能带来的风险和采取行动的必要性达成共识，并提出建立“具有国际包容性”的前沿人工智能安全科学研究网络。2024 年 5 月，韩国及英国共同主办人工智能首尔峰会，16 家顶尖人工智能科技公司联合签署《前沿人工智能安全承诺》，承诺公开透明地报告其前沿人工智能风险管理方法，设定“风险阈值”，并在风险过大时暂停开发和部署。同时，英、美等 11 国签署《首尔人工智能安全科学国际合作意向声明》，依托各国人工智能安全研究所，强化前沿人工智能风险治理研究合作，提升治理体系的互操作性。

国际标准化组织（ISO）积极推动风险治理的技术标准和管理规范建设。2023 年 12 月，ISO 发布《信息技术人工智能管理体系》

（ISO/IEC 42001），为各类组织实施人工智能风险管理等关键流程提供指引。2024 年 4 月，联合国下属标准组织**世界数字技术研究院**（WDTA）发布《生成式人工智能应用安全测试标准》和《大语言模型安全测试方法》两项国际标准，为大模型本身的安全风险防范提供了一套全面、严谨且实操性强的结构性方案。2024 年 5 月，**国际电信联盟**（ITU）在“AI for Good”全球峰会期间组织了“检测深度合成和生成式人工智能：人工智能水印和多媒体真实性的标准”研讨会，讨论了安全风险、治理技术以及标准化需求等议题。同月，国际电信联盟与中国信通院、ISO、国际电工委员会（IEC）、Adobe、微软等联合启动了“人工智能水印、多媒体内容真实性和深度伪造检测标准合作计划”，旨在共同推动相关标准的制定与合作，应对人工智能生成内容带来的风险挑战，提高多媒体内容的真实性和可信度。

（二）主要经济体：迭代完善风险划分与管控举措

欧盟采用四分法对人工智能风险进行划分，并配套制定风险管理措施。风险划分上，2024 年 8 月正式生效的欧盟《人工智能法》，依据人工智能风险程度，将人工智能系统划分为不可接受、高风险、有限风险、低风险四种风险等级。**风险管控上，**欧盟通过加强统一监管、产业标准化、签署国际公约等方式加强治理，全面应对人工智能风险。包括成立人工智能办公室，负责监管人工智能发展，同时防范风险；要求欧洲标准化委员会（CEN）和欧洲电工标准化委员会（CENELEC）起草针对高风险人工智能系统的欧洲风险管理标准；参与起草和签署全球首个聚焦人工智能治理的多边国际公约，要求缔约方共同应对人

工智能系统所带来的具体风险。

美国重点关注人工智能对国家安全的风险，通过安全测试加强技术监管。风险划分上，美国核心关注人工智能对国家安全造成的风险挑战。2023 年 10 月，美国总统拜登签署了《安全、可靠和值得信赖地开发和使用人工智能》行政令，明确推动技术发展、企业自律和合作治理，从维护国家安全出发，加强对关键基础设施领域的保护。**风险管控上，**白宫推动人工智能企业签署自愿承诺书，确保模型发布前经过严格安全测试，并承诺与行业、政府及学术界分享风险管理经验，以提高整体治理水平。截至 2024 年 11 月，已有包括 OpenAI、微软、苹果在内的 16 家企业自愿签署承诺书。

英国治理方案优先关注风险聚集领域，通过监管试点和开源平台促进共治。风险划分上，英国政府主要按垂直场景划分风险，优先治理医疗保健、消费市场等领域。2024 年英国药品和医疗保健产品监管机构（MHRA）规划报告重点之一是为软件作为医疗器械和人工智能及机器学习产品制定规定。2024 年 4 月，英国竞争与市场管理局（CMA）发布《人工智能战略更新》报告，提出识别人工智能对竞争和消费者保护的风险，开展前瞻性评估，增强执法机构的直接执行权力。**风险管控上，**英国推出开源安全平台和试点项目。2024 年 4 月，英国金融行为监管局（FCA）与数字监管合作论坛（DRCF）合作建立人工智能和数字中心试点，推进监管沙盒计划，帮助企业测试人工智能风险的创新解决方案。2024 年 5 月，英国人

工智能安全研究所（AISI）推出开源的“Inspect”人工智能模型安全评估平台，帮助产业界评估人工智能模型性能及风险。

新加坡结合事前事后风险治理，建立可验证的安全测试机制。风险划分上，事前通过《人工智能模型治理评估框架》进行多维度治理，事后则根据风险的不同责任主体建立相应框架。2024 年 5 月，新加坡迭代发布《生成式人工智能治理模型框架》，专注生成式人工智能特性，强调人工智能产业链上各方按照其控制水平分担责任，将开发商承诺、产品损害责任与无过失保险统筹结合，确保风险发生时能够为用户提供及时救济。**风险管控上，**2024 年 5 月，新加坡 AI Verify 基金会在原有 AI Verify 人工智能风险测试工具集中再添“登月计划”，包含基准测试、红队测试和测试基线，帮助开发人员根据风险基线测试人工智能模型，推动人工智能安全应用。

我国统筹风险分类分级，依托制度规范和技术自律推进敏捷治理。风险划分上，我国依据应用场景进行分类，并根据危害程度进行分级。我国《互联网信息服务算法推荐管理规定》《生成式人工智能服务管理暂行办法》等部门规章针对具有舆论属性或者社会动员能力的人工智能应用场景进行规制。对于可能对社会舆论和公众认知产生重大影响、涉及人类生命健康、公共安全等的人工智能应用主体应履行备案审查、安全评估、伦理审查等义务。**风险管控上，**党的二十届三中全会提出建设人工智能安全监管制度，整体推进制度规范完善和包容审慎监管。产业界通过备案、安全评估等强化安全防范措施，应对信息服务领域安全风险。截至 2024 年 11 月，国内已有近 2500 个深度合

成算法、252 款生成式人工智能产品完成备案、57 款生成式人工智能产品完成登记，形成良好示范效应。

（三）产业组织：协同推进治理框架和安全工具

治理框架方面，国内外产业组织正面向政府和企业提出综合治理的系统框架。2023 年 1 月，美国国家标准技术研究所（NIST）发布《人工智能风险管理框架》（1.0），为产业提供了系统化的风险识别、映射、评估和管理方法。2023 年 8 月，美国 AI Now 研究所等机构联合发布“零信任人工智能治理框架”，为政策制定者提供治理路线图。2023 年 12 月，中国信通院依托中国人工智能产业发展联盟（AIIA）筹建安全治理委员会，发布“人工智能风险管理体系”，旨在持续推动人工智能安全治理技术能力提升和安全应用落地。

安全技术方面，多个研究所和实验室积极推动人工智能安全技术的研发与应用。2024 年 1 月，NIST 发布了关于对抗性机器学习攻击的报告，概述了包括数据投毒、模型窃取、成员推断和属性推断等多种攻击方法。2024 年 4 月，MLCommons 制定基准测试 v0.5，用于验证评估大模型的安全性，使用超 43,000 个测试提示词评估大模型对危险提示的响应。2024 年 5 月，阿里巴巴达摩院（湖畔实验室）、新加坡南洋理工大学等联合提出了“大模型知识链”（CoK）框架，进一步提高大模型回答知识类问题的准确率。2024 年 4 月，中国信通院依托中国人工智能产业发展联盟发起大模型安全基准测试 AI Safety Benchmark。在内容安全维度，整理了 50 余万条测试输入，涵盖了底线红线、信息泄露和社会伦理等风险类型。在攻击测试维度，

收集了 80 余种提示词攻击模板和数十种图文多模态攻击手段。

（四）企业主体：深化技管融合的风险应对方案

为了应对人工智能带来的多样化风险，企业在管理和技术两个维度积极探索解决方案。企业管理方面，加强全流程管控，提升治理能力。越来越多企业通过设立专门的内部治理组织和发布治理规则，统筹人工智能的治理工作。多个科技公司成立伦理委员会并发布伦理准则，推动人工智能产品的伦理审查。**技术合作方面**，企业之间的技术治理合作也在不断深化。例如，2024 年 2 月，亚马逊、谷歌、IBM、TikTok 和 X 等 20 家科技公司在慕尼黑安全会议上宣布，联合打击深度伪造信息，提出平台或产品解决方案，以检测和防止欺骗性人工智能生成内容。

技术解决方案逐步向一体化和定制化发展，涵盖了从风险识别、评估到防御等多个环节。风险识别方面，OpenAI 于 2024 年 5 月发布一款图像检测分类器，其可在 98%的情况下正确识别由其图像生成模型 Dall-E 3 创建的图片。微软推出的 PyRIT 工具能够评估大模型生成内容的安全性，推动风险识别的自动化和智能化。百度推出深伪检测算法服务，有效覆盖各类深伪威胁场景，助力实现更安全、更可靠的内容审核与用户验证。阿里巴巴发起开源大模型治理项目，推出由不同领域专家人工构造的中文评测集 CValues，评估中文语言模型安全性，帮助了解模型的能力和局限性，判断可能存在的风险。**风险评估和防御方面**，谷歌发布的“安全人工智能框架”（SAIF）工具，帮助研发方评估人工智能系统的安全状况并实施防护。微软的 Azure AI

Content Safety 提供了自动识别和干预有害内容的功能，并允许人工智能系统研发方自定义过滤规则，从而根据具体需求提升内容合规性。百度于2024年8月安全隐式水印产品能够嵌入生图场景，抵御裁剪、压缩、截图、涂抹等组合攻击，保障生成内容的权属明晰与持久可溯源。商汤科技针对生成式人工智能的风险特点，打造 SenseTrust 人工智能治理平台，提供数据处理、模型训练、模型部署、推理服务全面的治理工具箱。

当前，企业作为人工智能风险治理的核心主体，面临着风险日益复杂化挑战，亟需一套切实可行的治理框架来识别、评估并有效应对风险。推动具有前瞻性和落地性的风险治理框架，确保人工智能安全发展，已经成为产业界的紧迫任务。

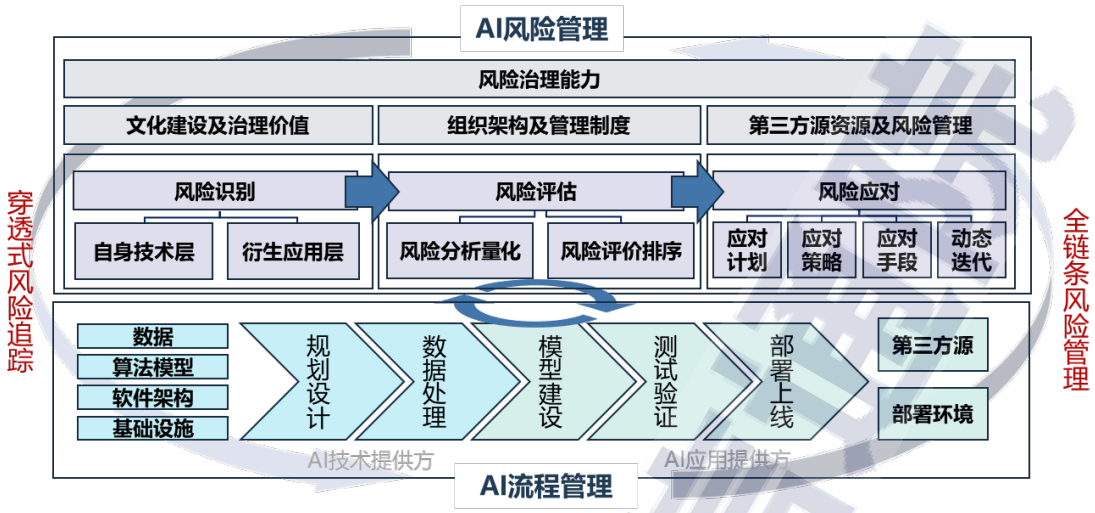
四、面向产业的人工智能风险治理实践方案

（一）打造全链条人工智能风险治理框架

人工智能技术的迅猛发展使技术与应用风险复杂交织，对构建系统化、整体性、可落地的风险治理框架提出了更高要求。本报告在延续中国信通院前期发布的《可信人工智能白皮书（2021 年）》[21]研究基础上，以可控可靠、透明可释、隐私保护、明确责任及多元包容等核心指导原则，提出兼具穿透式风险追踪与全链条风险管理的实践框架。框架以提升提供人工智能技术和应用的相关组织（下称“组织”）的人工智能风险治理能力为目标，形成了从精准识别、科学评估到有效应对的全流程治理路径，旨在为组织制定发展与安全有机统一的系

[21] 中国信息通信研究院、京东探索研究院：《可信人工智能白皮书》，2021 年 7 月。

统化解决方案（见图 4）。



来源：中国信息通信研究院

图 4 面向产业的人工智能风险治理实践框架

一是注重顶层设计，通过制度完善、文化建设和外部协作来提升系统治理能力。从组建专业团队与管理制度以确保治理工作的专业性和高效性，到塑造治理共识和组织文化氛围以促进治理工作的全员参与；从建立贯穿人工智能全生命周期的内部机制，到强化对第三方资源的治理以获取外部协同，通过构建系统治理能力为风险识别、评估和应对提供全面支撑。

二是聚焦风险管理，覆盖风险识别、评估与应对的动态闭环管理体系。首先，风险识别环节从自身技术层与衍生应用层两个维度展开。自身技术层风险识别重点关注数据质量与合法性、算法模型的鲁棒性、软件供应链安全及基础设施的稳定性；衍生应用层则重点审查人工智能引发的价值侵害、安全影响和潜在的社会问题。**其次，**风险评估环节通过风险影响评估与优先级划分，将风险严重性、发生可能性及其影响程度转化为具体可执行的决策依据。采用定量分析工具如概率风

险评估和仿真模拟，帮助组织合理配置资源，集中应对高优先级风险，同时动态调整评估结果以适应技术迭代与社会环境变化。**最后，**风险应对环节为不同优先级的风险定制应对方案，包括技术层的模型微调、隐私保护与安全测试，应用层的动态监控与内容过滤，以及管理层的应急预案与跨领域协作。同时，通过持续评估剩余风险，不断优化应对措施，提升风险治理的敏捷性。

三是明确主体责任，构建贯穿研发、部署和应用等关键环节的全链条流程管理体系。首先，技术提供方作为人工智能技术的基础能力构建者，专注于数据处理、算法模型开发、软件架构设计以及基础设施建设等方面，尤其应注重在系统设计和开发等阶段遵循风险治理相关原则及机制，确保系统具有高度安全性、透明性与可控性，并提供必要的技术工具以支持风险管控措施的落实。**其次，**应用提供方作为人工智能产品的打造者，需根据自身业务需求定制部署方案，并在实际使用过程中持续监控与评估人工智能表现，及时发现潜在风险并采取应对措施。**最后，**二者应形成沟通与合作机制，通过信息共享和协同制定风险识别、评估与应对策略，提升行业整体的人工智能风险治理能力。

（二）提升人工智能风险系统治理能力

系统治理能力是人工智能风险治理产业实践的核心支柱，贯穿人工智能系统全生命周期。人工智能风险系统治理能力是组织在系统设计、开发、测试、部署及运行等各阶段中识别、评估并应对潜在威胁的能力，是确保技术安全性与业务连续性的关键保障。通过将人工智

能风险治理整合到企业现有的综合风险治理框架中，组织可以更加全面地识别和应对网络安全、数据安全、隐私保护等跨领域风险的交互效应。具体而言，组织应从**文化建设**、**制度完善**和**外部关系**三个层面建立系统治理能力，从而在快速变化的技术与环境中保持竞争优势和合规性。

文化建设是人工智能风险系统治理能力的内涵要素，是塑造组织价值观与行为准则的重要环节。首先，管理层应形成治理共识，将安全、透明、公平、可问责和隐私保护等理念融入组织文化，贯穿技术研发与决策全过程，明确组织可能承担的社会、经济和环境责任。**其次**，组织需通过强化员工学习与实践能力提升整体治理水平。可制定培训计划或客场、邀请交叉领域专家开展讲座或组织案例学习、提供技术工具等方法提升员工的治理意识与能力。**最后**，组织应营造支持风险治理的文化氛围，通过宣传材料和内部活动展示治理成果，激励员工参与治理创新，推动形成全员参与的治理生态。

制度完善是人工智能风险系统治理能力的基本要求，为组织实施有效治理提供明确依据。首先，组织应建立专门的风险治理团队，由核心负责人领导协调，结合业务需求细分为法律合规、安全技术等专职小组，确保覆盖人工智能系统全生命周期的治理工作。**其次**，组织应制定并落实人工智能风险治理相关的人员管理制度，明确参与需求分析、产品设计、研发和测试等环节的人员职责，确保各环节相关人员能够识别和减轻可能发生的人工智能风险。**最后**，组织应建立人工智能系统研发与使用阶段的管理制度，制定应急预案与救济措施，确

保系统在使用中满足风险治理要求。此外，制度应随技术发展与政策变化持续迭代优化，保持对动态风险的高效响应。

外部协作是人工智能风险系统治理的重要支撑，为组织安全高效利用第三方资源提供坚实保障。首先，组织应对第三方软硬件和数据供应商进行全面评估，确保其产品和服务符合组织的安全标准与治理要求。**其次**，组织应建立与第三方协作的风险治理机制，明确各方在人工智能系统设计、开发、部署和应用中的责任边界。通过签署协议或制定技术标准，确保第三方资源的使用方式与组织内部治理框架相一致。**最后**，为应对多主体、多环节可能产生的风险叠加效应，组织应强化动态监控与反馈机制，定期审查第三方提供的系统和数据的安全与合规性。

（三）设立人工智能风险精准识别机制

风险识别贯穿自身技术层与衍生应用层，是人工智能风险治理的首要步骤。人工智能技术演进和广泛应用带来了多元动态风险，精准化、系统性的风险识别成为治理的关键抓手。风险识别是发现和描述人工智能系统生命周期中可能造成负面影响的不确定性和潜在威胁的过程，旨在为风险评估和风险应对奠定基础。具体而言，人工智能的风险识别涵盖自身技术层和衍生应用层两个层面。

自身技术层风险识别，旨在确保人工智能系统的安全、可靠、可控，覆盖数据、算法系统、软件和基础设施四个要素。**数据风险识别应关注数据源的合法性与质量，**如采用数据脱敏和匿名化处理技术减少敏感信息泄露风险；利用实时数据质量监测和偏差检测工具，及时

识别不完整或被“投毒”的数据集；以及采用机器学习算法进行数据异常检测，快速发现潜在违规或有害输入。**算法模型风险识别聚焦于鲁棒性和透明性评估，**如通过对抗性训练和攻击模拟测试模型在扰动输入下的稳定性；结合可视化工具和可解释性算法分析模型决策逻辑，减少不可解释性导致的风险；以及对模型核心参数和权重进行加密保护，防范逆向攻击和知识产权窃取。**软件风险识别覆盖供应链全生命周期管理，**如利用静态和动态分析工具识别潜在漏洞与后门隐患；采用自动化兼容性测试发现系统集成中的冲突问题；以及持续监控跟踪软件更新和补丁应用，防止未修复的漏洞被恶意利用。**基础设施风险识别侧重于关键硬件与算力资源的保护，**如利用资源使用监控技术来识别恶意占用或滥用行为；应用硬件指纹和加密认证技术来确保计算资源的安全性；采取多级应急响应计划在硬件受损或遭受攻击时提供快速恢复的保障。**衍生应用层风险识别**，企业应积极研判各层次风险，部署基于行为模式的异常检测系统，通过跨阶段审查工具全面跟踪技术开发与应用过程。

此外，风险识别需要通过动态监测反馈，为优化治理框架提供支持。组织应建立持续的检测监测能力，精准识别人工智能各维度风险。**一是开展红队测试，提升安全能力。**通过模拟真实攻击者的策略和技术，红队对人工智能系统进行多层次的攻击模拟，以识别潜在安全漏洞和弱点。红队测试可涵盖软硬件基础设施、数据安全、模型安全方面的测试能力，对源代码、第三方库、开源框架等进行静态分析，发现代码中包含的漏洞、后门；对模型进行输入输出层面的仿真攻击测

试，使用模型诱导、常见的模型后门触发器或对抗样本检测模型表现。**二是设置蓝队参与系统防护，检验防御能力。**根据红队的模拟攻击情况，蓝队实施相应的防护工作，提升持续识别风险的能力。利用工具和人工审核相结合的方式，识别训练数据中包含的恶意数据、投毒数据。**三是建立持续监控能力。**由于当前的攻击方法层出不穷，很多攻击难以通过测试的方式发现。建立全网监测能力，持续发现最新的攻击方法，有助于提升潜在风险的识别能力；持续收集和分析用户的反馈意见，以便及时调整安全策略和防御措施。**四是建立跨领域合作。**人工智能技术涉及多个领域，特别是人工智能技术在垂直领域的应用需要结合行业知识识别特有风险。在理解行业特点的基础上，形成有针对性的解决方案。此外，还需要定期进行法规遵从和伦理及合规审查，确保系统符合相关法律法规和伦理标准。

（四）健全人工智能风险科学评估体系

风险评估是对人工智能风险进行挖掘分析和精准把脉的重要阶段。风险评估旨在量化风险严重程度，明确它们发生后可能对个人、组织和全社会可能带来的危害，以及这些危害实际发生的可能性。风险评价旨在明确个人、组织和全社会对风险可接受度，划定风险优先级。**人工智能风险评估在风险治理中起着桥梁作用**，直接连接风险识别与风险纾解两个重要步骤。它基于识别阶段发现的潜在风险，衡量具体的风险等级，从而为组织提供应对依据。有效的风险评估可以帮助组织合理分配有限资源，专注处理最关键的风险问题，避免资源的分散与浪费。

组织通过风险分析量化风险，明确潜在危害以及危害实际发生的可能性。这一过程能为识别到的每个风险附加一个潜在损失值，从而为后续的风险评价和风险应对提供量化依据。具体而言，**首先，明确分析框架，**增强评估结果的可信度和实效性。合理有效的分析框架应当明确列出需要量化的具体项目。例如，潜在危害的量化应当包含可预期的各类损失，包括已知风险可能导致的直接财务损失、对企业声誉的负面影响、法律争端导致的民事赔偿及刑事惩罚，以及对社会秩序与伦理的潜在冲击。**其次，广泛获取分析所需信息。**充足信息是有效分析的基础。因此，分析流程应当与风险识别紧密衔接，确保识别获取的信息完整纳入分析范围。分析过程中，组织应当开展调研，充分了解风险所处的法律、经济、社会和伦理环境，从个人、组织、社会等多维度全面理解潜在危害。**最后，选用有效的分析工具。**在危害程度量化方面，通常采用定量分析方法，如概率风险评估、影响图分析等，计算风险发生的可能性及潜在的经济、声誉等损害，以确定风险的直接危害程度。在影响范围分析方面，可通过分析风险事件可能波及的地域范围、受众群体及其社会经济特征，评估风险对国家经济安全、社会稳定、公众信任度、个人权益等方面的影响深度和广度。此外，在评估过程中，还可通过建立仿真模拟和风险评估模型的方式提升风险评估的准确性。如利用社交网络模拟模型，研判特定风险事件发生后，事件影响扩散路径、范围等，评估其后果。

组织通过风险评价，考虑风险可接受度，划定风险优先级。风险是任何系统的必然组成部分。彻底去除全部风险在价值上不必追求，

在技术上不可实现。基于此，组织应当在风险分析的基础上，为不同风险进行排序，并制定相应的风险应对策略。**首先，设计开放的风险可接受度评价标准。**风险可接受度不仅与现有风险应对的技术水平、组织能力和资源储备高度相关，而且受到具体应用场景的限制。。法律和监管要求、行业标准、社会伦理和经济发展水平等都可能影响风险可接受度。因此，组织应当采取有效方法，灵活调整可接受度评价标准。**其次，基于风险严重程度和可接受度的权衡，评价风险优先级。**对于严重程度远高于可接受度的风险种类，可以划定为“高优先级”风险。对于严重程度相当于和远低于可接受度的风险，则可以划定为“中优先级”和“低优先级”风险。**最后，动态调整优先级评价结论。**随着经济社会条件的变化，人工智能风险的影响程度和可接受程度处于持续的动态变化之中。例如，原本被认为是“低优先级”的风险，可能会随着情境变化加重，转变为“高优先级”风险。因此，组织应建立有效机制，定期检查和更新现有评价结论。

（五）完善人工智能风险敏捷应对方案

风险应对是验证人工智能风险治理有效性的关键一环。在识别和评估相关风险后，采取针对性措施来降低、缓解或管理这些风险，从而确保人工智能系统安全部署和应用。根据风险评估结果，实施风险应对具体措施来响应或消除风险，其实施效果将最终决定人工智能系统是否能安全部署、在何种条件下安全部署。

组织应具备涵盖计划、策略和技术手段三方面的综合风险应对工具。考虑到资源有限的现实，组织需要根据风险评估结果和优先级，

合理分配应对资源。**一是根据风险优先级制定详细的应对计划。**对于“高优先级”风险，应投入更多的资源，成立专门的团队并设立专项预算，确保及时响应；而对于“中优先级”和“低优先级”风险，则可以采取较为缓和的响应和监控策略，以保证资源的最优利用。**二是根据“成本－收益”分析选择合理的风险应对策略。**如果“高优先级”风险对特定应用场景的影响较大且短期内难以缓解，组织应优先考虑风险规避，推迟或取消人工智能系统的部署。若通过低成本措施能够有效降低风险发生的概率或减少其潜在损害，组织则应选择缓解风险。对于“低优先级”风险，组织可以选择接受风险，并持续进行监控。对于可能导致较大经济损失但社会影响较小的风险，则可以通过购买保险等手段进行风险分担。**三是掌握并选择适当的技术手段应对风险。**在**研发阶段**，可对训练数据采取不断地清洗、标注，剔除与研发或应用预期不符的数据，提升训练数据的质量和有效性。在训练阶段，使用隐私计算技术，建立安全可信环境，防止训练数据被未经许可的个体或组织访问。在**测试阶段**，可采用引入安全护栏的方式对输入模型和模型输出的内容进行审核、过滤，防范虚假、有害信息；还可通过模型微调等方式，动态提升模型的防护能力。在**部署和应用阶段**，可采用安全基准测试的方式，引入外部测试能力，对模型的安全能力进行持续的监测。

此外，组织还需持续迭代风险应对措施，共享最佳实践提升风险治理能力。首先，定期衡量风险应对的效果，加以改进。在实施风险缓解措施后，组织应评估剩余的风险，包括衡量已采取措施后仍未完

全控制的风险，并与企业设定的风险可接受度进行对比，决定人工智能系统是否达到特定场景的部署条件。如果发现风险缓解措施未达到预期效果，需及时改进或更新相应技术手段。**其次，积极分享风险应对最佳实践以促进业界整体安全水位提升。**组织应当与行业内外的相关方分享其在风险应对过程中积累的经验，帮助其他组织提升风险治理能力，进而增强行业整体的风险应对能力，共同打造人工智能的安全环境。

五、展望建议

人工智能技术以其革命性影响迅速融入全球经济与社会，但同时也带来了技术风险与治理挑战。本报告系统梳理了人工智能风险的认知转型、治理实践和应对框架，深入剖析了人工智能风险的多维特征与治理需求。在此基础上，报告总结各方努力，提出以下四方面的具体建议，旨在构建一个多元、协同、系统的治理体系，从而推动人工智能技术的安全发展与风险治理，为人工智能的可持续发展提供有力支撑。

（一）促进全球合作，共筑人工智能治理新格局

中国充分发挥引领作用，推动国际规则与合作机制的创新与完善。一是参与并主导多边治理框架，通过联合国、二十国集团（G20）、金砖国家及“一带一路”等平台加强沟通与协作，推动《全球人工智能治理倡议》和《人工智能能力建设普惠计划》迈向更为具体的实际行动。**二是**关注全球人工智能安全峰会，参与治理框架的互操作性与协作性，积极代表新兴经济体发声，推动形成普惠性国际治理规则。

三是建立人工智能风险监测与信息共享平台，探索国际合作模式，通过整合全球风险数据、开发协同监测工具、组织联合演练等手段，提升跨境风险的动态管理与协同应对能力，为全球治理网络注入中国智慧。

（二）优化治理体系，夯实人工智能安全发展基础

政府进一步优化治理体系，提升政策执行力和跨领域协同能力。一是不断完善我国人工智能治理体系，鼓励地方开展先行先试，创新探索人工智能治理规则，为国家层面的人工智能治理提供实践指导和参考借鉴。**二是**制定技术复杂性和应用场景结合的风险分级标准，细化从内生安全风险到衍生安全风险的具体分类，明确配套治理规则，实现精准施策。**三是**构建覆盖全国的人工智能风险治理数据库，整合行业数据资源，支持实时监测、动态评估和风险预警；同时引入智能化监管工具，以全面提升风险治理的效率和精准性。

（三）发挥桥梁作用，推进行业治理标准化与动态化

产业组织推动技术研发与政策制定深度融合，助推行业治理标准化与动态化。一是搭建双向交流平台，增进技术应用与政策制定的协作机制，以产业智慧增强政策法规的前瞻性与科学性，促进政策和法律积极适应并有效引导人工智能产业实践。**二是**推动人工智能风险认知评估动态滑尺的应用，将动态风险识别和适配机制融入企业、政府和社会的治理实践，增强对技术风险的敏感性和应对的灵活性。**三是**结合人工智能风险治理落地框架，推动行业统一的治理标准和操作指南，构建可复制、可推广的治理实践模式，确保人工智能技术发展与

风险治理的动态平衡。

（四）强化内外协同，助力技术创新与责任共担

组织应健全内外部风险治理机制，实现技术创新与社会责任统一。一是加强内部治理机制建设，明确责任分工，设立伦理委员会或技术监督小组，制定覆盖设计、开发、测试和部署全流程的治理规则，确保合规与透明。**二是**定期开展多维度风险评估与测试，聚焦算法可释、隐私保护与安全漏洞，通过技术工具实现自动化评估，并及时采取防护措施，促进安全能力提升。**三是**积极参与外部协同合作，推进行业标准制定，与其他企业联合开发解决方案，共同应对深度伪造、算法歧视等系统性挑战，为行业可持续发展贡献力量。

中国信息通信研究院 人工智能研究所

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62302914

传真：010-62304980