一、新一代人工智能治理原则

1、《治理原则》旨在更好协调人工智能发展与治理的关系，确保人工智能安全可控可靠，推动经济、社会及生态可持续发展，共建人类命运共同体。《治理原则》突出了发展负责任的人工智能这一主题，强调了和谐友好、公平公正、包容共享、尊重隐私、安全可控、共担责任、开放协作、敏捷治理等八条原则。

2、新一代人工智能治理原则——发展负责任的人工智能

3、全球人工智能发展进入新阶段，呈现出跨界融合、人机协同、群智开放等新特征，正在深刻改变人类社会生活、改变世界。为促进新一代人工智能健康发展，更好协调发展与治理的关系，确保人工智能安全可靠可控，推动经济、社会及生态可持续发展，共建人类命运共同体，人工智能发展相关各方应遵循以下原则：

4、一、和谐友好。人工智能发展应以增进人类共同福祉为目标；应符合人类的价值观和伦理道德，促进人机和谐，服务人类文明进步；应以保障社会安全、尊重人类权益为前提，避免误用，禁止滥用、恶用。

5、二、公平公正。人工智能发展应促进公平公正，保障利益相关者的权益，促进机会均等。通过持续提高技术水平、改善管理方式，在数据获取、算法设计、技术开发、产品研发和应用过程中消除偏见和歧视。

6、三、包容共享。人工智能应促进绿色发展，符合环境友好、资源节约的要求；应促进协调发展，推动各行各业转型升级，缩小区域差距；应促进包容发展，加强人工智能教育及科普，提升弱势群体适应性，努力消除数字鸿沟；应促进共享发展，避免数据与平台垄断，鼓励开放有序竞争。

7、四、尊重隐私。人工智能发展应尊重和保护个人隐私，充分保障个人的知情权和选择权。在个人信息的收集、存储、处理、使用等各环节应设置边界，建立规范。完善个人数据授权撤销机制，反对任何窃取、篡改、泄露和其他非法收集利用个人信息的行为。

8、五、安全可控。人工智能系统应不断提升透明性、可解释性、可靠性、可控性，逐步实现可审核、可监督、可追溯、可信赖。高度关注人工智能系统的安全，提高人工智能鲁棒性及抗干扰性，形成人工智能安全评估和管控能力。

9、六、共担责任。人工智能研发者、使用者及其他相关方应具有高度的社会责任感和自律意识，严格遵守法律法规、伦理道德和标准规范。建立人工智能问责机制，明确研发者、使用者和受用者等的责任。人工智能应用过程中应确保人类知情权

，告知可能产生的风险和影响。防范利用人工智能进行非法活动。

10、七、开放协作。鼓励跨学科、跨领域、跨地区、跨国界的交流合作，推动国际组织、政府部门、科研机构、教育机构、企业、社会组织、公众在人工智能发展与治理中的协调互动。开展国际对话与合作，在充分尊重各国人工智能治理原则和实践的前提下，推动形成具有广泛共识的国际人工智能治理框架和标准规范。

11、八、敏捷治理。尊重人工智能发展规律，在推动人工智能创新发展、有序发展的同时，及时发现和解决可能引发的风险。不断提升智能化技术手段，优化管理机制，完善治理体系，推动治理原则贯穿人工智能产品和服务的全生命周期。对未来更高级人工智能的潜在风险持续开展研究和预判，确保人工智能始终朝着有利于人类的方向发展。

二、ai人工智能真的存在很大风险吗

人工智能（AI）的出现确实为我们带来了很多便利和创新，但同时也存在一些风险和挑战。以下是一些可能存在的风险：

1.数据隐私和安全问题：AI需要大量的数据来学习和发展，但这些数据可能包含个人私密信息，如果泄露或被滥用，将会对我们的隐私和安全带来威胁。

2.就业岗位减少：随着AI技术的不断发展，可能会取代一些传统的工作岗位，导致失业率上升，社会不稳定。

3.伦理和道德问题：AI技术的应用范围越来越广泛，但有些应用可能会涉及到伦理和道德问题，例如自动化武器的使用、AI决策制定的公平性等。

4.对人类的控制问题：如果AI超过了人类的控制力，可能会对人类造成威胁，例如机器人的失控。

5.信息失真和偏差问题：AI的学习和判断是基于大量的数据，而这些数据可能会存在偏差和失真，导致AI的决策出现问题，影响公正和公平。

需要注意的是，AI风险并不是绝对的，而是取决于AI的设计、应用和管理方式。只有在充分认识到其风险的同时，采取相应的措施才能最大程度地发挥其利益，同时避免其风险。

三、人工智能风险有哪些

人工智能的风险包括但不限于以下几点：

1.数据隐私和安全问题：人工智能需要大量数据来训练模型，但数据的获取、存储和使用过程中存在隐私和安全风险。

2.偏见和歧视：人工智能算法可能存在偏见和歧视，这可能导致不公平的结果。

3.失业问题：人工智能的发展可能导致一些传统职业的消失，从而引起失业问题。

4.法律和伦理问题：人工智能的决策过程缺乏透明度，可能导致法律和伦理问题。

5.安全和稳定性问题：人工智能系统可能存在安全漏洞，遭受黑客攻击，或者出现不稳定的情况，导致损失。

6.人机关系失衡：过度依赖人工智能可能导致人与人之间的沟通减少，人际关系逐渐疏远。

7.无法应对复杂情况：尽管人工智能在许多方面表现出色，但在复杂情况下，它们可能无法像人类一样做出灵活的决策。

8.不可预测性：人工智能的决策过程往往缺乏透明度，结果可能不可预测，这可能会引发社会问题。

9.技术依赖：对人工智能技术的过度依赖可能导致技术失控，一旦出现故障，可能产生严重后果。

10.智慧反击：高级的人工智能系统可能学会如何利用人类的弱点，进行反击，这将引发严重的安全问题。

为了降低这些风险，我们需要制定相应的政策、法规和技术标准，以确保人工智能的发展和使用符合社会的价值观和道德标准。同时，我们也需要开展更多的研究，以更好地理解和解决这些问题。

四、人工智能应用不当会产生哪些风险

所谓的“数据投毒”指人工智能训练数据污染导致人工智能决策错误。通过在训练数据里加入伪装数据、恶意样本等，破坏数据的完整性，进而导致训练的算法模型决策出现偏差。

一方面逆向攻击可导致算法模型内部的数据泄露;

另一方面，人工智能技术可加强数据挖掘分析能力，加大隐私泄露风险。比如各类智能设备（如智能手环、智能音箱）和智能系统（如生物特征识别系统、智能医疗系统），人工智能设备和系统对个人信息采集更加直接与全面。人工智能应用采集的信息包括了人脸、指纹、声纹、虹膜、心跳、基因等，具有很强的个人属性。这些信息具有唯一性和不变性，一旦泄露或者滥用将产生严重后果。

运行阶段的数据异常可导致智能系统运行错误，同时模型窃取攻击可对算法模型的数据进行逆向还原。此外，开源学习框架存在安全风险，也可导致人工智能系统数据泄露。

图像识别、图像欺骗等会导致算法出问题，比如自动驾驶，谷歌也做了一些研究，如果模型文件被黑客控制恶意修改，并且给它学习，会产生完全不一样的结果;

算法设计或实施有误可产生与预期不符甚至伤害性结果;

算法潜藏偏见和歧视，导致决策结果可能存在不公;

算法黑箱导致人工智能决策不可解释，引发监督审查困境;

含有噪声或偏差的训练数据可影响算法模型准确性。

人工智能不可避免的会引入网络连接，网络本身的安全风险也会将AI带入风险的深坑;

人工智能技术本身也能够提升网络攻击的智能化水平，进而进行数据智能窃取;

人工智能可用来自动锁定目标，进行数据勒索攻击。人工智能技术通过对特征库学习自动查找系统漏洞和识别关键目标，提高攻击效率;

人工智能可自动生成大量虚假威胁情报，对分析系统实施攻击。人工智能通过使用机器学习、数据挖掘和自然语言处理等技术处理安全大数据，能自动生产威胁性情报，攻击者也可利用相关技术生成大量错误情报以混淆判断;

人工智能可自动识别图像验证码，窃取系统数据。图像验证码是一种防止机器人账户滥用网站或服务的常用验证措施，但人工智能通过学习可以让这一验证措施失效。

第三方组件问题也会存在问题，包括对文件、网络协议、各种外部输入协议的处理都会出问题。被黑客利用，带来的是灾难性的毁灭。

五、人工智能对企业发展与风险管理的影响

1、智能制造。人工智能引入到企业的生产制造系统，包括机器人、大数据、云制造等，能使企业实现智能制造，从而提高制造效率。

2、智能研发。运用人工智能可以辅助企业进行科学的研发决策，使企业把握研发方向，洞悉研发风险。

3、智能管理。包括智能决策、精准营销等等。