## 当前人工智能存在哪些安全风险？如何应对？

 AI系统的安全问题主要包括模型安全、数据安全以及代码安全。

1. 在数据爆发的年代，90%的数据是近两年内生成的，据估计2020年，全球平均每秒种产生1.7MB的数据
2. 对抗性机器学习（Adversarial Machine Learning）

随着AI 技术的发展，

1. 模型缺陷与后门（Model Defects and Malicious Models）

近年来，各种结构复杂的二进制代码安全缺陷日益繁多。例如，后门、竞争条件及鉴权绕过等逻辑错误类型的代码安全缺陷。它们不仅存在于二进制单线程程序，而且分布于各种二进制并行程序。而符号执行、污点分析及模糊测试等现有方法难以发现结构复杂的代码安全缺陷，也不能有效地应用于二进制并行程序，使软件系统面临安全性风险

1. 数据污染攻击（Data Processing Vulnerabilities）

内存数据被污染往往是程序漏洞被利用的本质所在，从功能角度把内存数据划分为控制相关和非控制相关，由此引出控制流劫持攻击和非控制数据攻击。两者危害程度相当，前者因利用成本较低而成为主流，但随着控制流劫持防御方法的不断完善，非控制数据攻击逐渐被重视。研究者先后在顶级会议上提出了数据导向攻击得自动化利用框架Data-oriented Exploits （DOE）以及图灵完备性地证明Data-oriented Programming （DOP），使得非控制数据攻击成为热点。

1. 模型数据的推测窃取（Model Inference Attack）
2. 安全隐私问题

在人工智能的保护伞下，诸如机器学习和大数据等问题，都很容易触及到安全及隐私问题。有时基础设施扮演着很重要的角色。与隐私问题有关的安全需求，如将银行帐户和健康信息保密，将会对研究的安全性有更大的要求。2018年，有关安全和隐私的问题将得到解决，这一年，也是人工智能可能出现新的发展的一年。

人工智能的伦理问题也将成为2018年的主要问题，需要解决的伦理和道德问题包括人工智能对人类有哪些好处和坏处。人们也对机器人取代人类的可能性感到担忧，比如护士、治疗师或警察，另一个需要处理的问题是自主武器。

## 人工智能技术在网络安全中的应用场景有哪些？（举例企业的具体产品、技术或实践）

对于蓝盾来说，我们人工智能技术应用在网络安全的主要有以下几个安全产品中：

1. 边界安全，部署AI引擎于第三代FW,病毒文件检测。
2. 态势感知，基于机器学习多端点感知技术的用户行为分析，以及流量分析。
3. 蓝盾卫士，伪基站的防御，APK病毒文件检测。
4. 蓝盾云，AI云端技术（机器学习即服务）。
5. 云检测与云防线，AI图像与视频识别技术与文本分析来处理敏感信息

## 如何看待网络安全中人工智能的未来发展趋势？

1. **生成式对抗性网络**

生成式对抗网络GAN（Generative adversarial networks）目前已经成为人工智能学界一个热门的研究方向，GAN的基本思想源自博弈论，由一个生成器和一个判别器构成，通过对抗学习的方式来训练，目的是估测数据样本的潜在分布并生成新的数据样本。在图像和视觉计算、语音和语言处理、信息安全、棋类比赛等领域，GAN正在被广泛研究，具有巨大的应用前景。

无论是无监督学习、半监督学习，GAN给我们提供了一个处理问题的崭新思路，就是把博弈论引入到机器学习过程中来。可以预见，GAN本身的算法以及看问题的角度，必将对未来设计算法、以及解决实际问题产生深远的影响。

1. 技术孤岛的统一

**许多不同的框架在做同样的事情，如今，每一家从事机器学习的大公司都拥有自己的框架和其他开源解决方案。在单独的AI应用中，我们希望使用不同的框架，如 Cafle2, PyTorch和一些推荐系统的Tensorkow/Keras。合并它们需要大量的开发时间，它分散了数据科学家和软件开发人员从事更重要任务的注意力。解决方案必须是一种独特的神经网络格式，可以很容易的从任何框架中获得，然后开发人员能够轻松地部署，科学家可以轻松地使用，在这里我们推荐使用****ONNX标准。ONNX 的全称为“Open Neural Network Exchange”，即“开放的神经网络切换”。顾名思义，该项目的目的是让不同的神经网络开发框架做到互通互用。目前，Microsoft Cognitive Toolkit，PyTorch 和 Caffe2 已宣布支持 ONNX**

1. **混合学习模型**

**不同类型的深度神经网络（比如GAN或DRL）已经在性能和广泛应用方面展现出了大好前景，适用于许多不同类型的数据。然而，深度学习模型无法像贝叶斯方法或概率方法那样为不确定性建模。混合学习模型结合了这两种方法，集每种方法的优点于一身。混合模型的几个例子是贝叶斯深度学习、贝叶斯GAN和贝叶斯条件GAN。**

**有了混合学习模型，就可以扩大可处理的商业问题的种类，包括具有不确定性的深度学习。这可以帮助我们提升模型的性能和可解释性，进而会促使得到更广泛的采用。预计会看到更多的深度学习方法获得贝叶斯方法，而概率编程语言开始会结合深度学习。深度学习是机器学习中一种基于对数据进行表征学习的算法。**

1. **深度强化学习**

**这种神经网络的学习方式是通过观察、行动和奖励，与周围环境进行交互。深度强化学习（DRL）已用于学习游戏策略，比如Atari和Go，包括击败人类冠军的著名的AlphaGo软件**

**深度强化学习是所有学习技术中通用性最强的，所以它可以用于大多数商业应用。与其他技术相比，它训练模型所需要的数据较少。而尤其抢眼的是，它可以通过模拟来加以训练，因而完全不需要标记数据。鉴于这些优势，预计明年会出现更多的商业应用结合深度强化学习和基于代理的模拟。**

1. 自动化机器学习（**AutoML**）

**开发机器学习模型需要一个耗时、专家驱动的工作流程，这个流程包括数据准备、特征选择、模型或技术选择、训练以及调优等。AutoML使用许多不同的统计和深度学习技术，旨在使这个工作流程实现自动化。**

**AutoML是AI工具大众化的一部分，让商业用户能够在编程方面没有扎实背景的情况下开发机器学习模型。它也有望缩短数据科学家用来创建模型的时间。预计会看到更多的商业AutoML软件包、更庞大的机器学习平台里面整合AutoML。**

1. **语音交互入口AI**

对于许多人工智能领域研究者来说，语义学习是下一个主要目标。随着技术的进步，图像识别和语音识别都取得了重大突破，而这将有利于计算机更加高效地理解和生成语言，人工智能时代，服务的入口主要是自然语言的语音交互，自然语言处理会是人机交互的主要模式。

人工智能会全面包围我们生活的方方面面，近几年来，许多商业巨头已经开发出自己的聊天机器人工具，比如：微软的 “小冰” ，Facebook旗下的Messenger和还有Google Assistant等人工智能语音助手，这意味着，人工智能入口的竞争会更加惨烈，而越早参与，获得越多的用户，胜利的几率就越大。

1. **可解释的AI**

如今使用的机器学习算法种类繁多，这些算法可以在各种不同的应用中感知、思考和行动。然而其中许多算法被认为是“黑盒子”，它们如何得出结果方面几乎不为人知。当下可解释的AI蔚然成风，旨在开发这种类型的机器学习技术：在保持预测准确性的同时，生成更易解释的模型。

可解释、可证明、透明化的AI对于为技术赋予可信任性至关重要，并促进机器学习技术得到更广泛的采用。在开始大规模部署AI之前，企业会采用可解释的AI，作为一个要求或最佳实践，而各国政府在将来可能会将可解释的AI列作一项监管要求。

1. **GPU助力AI**

GPU，也就是我们所说的图形处理器，一直都是AI应用的主导硬件处理器，在图像语音识别、无人驾驶等人工智能领域，GPU正迅速扩大市场占比。

尽管有不少IT巨头和创新企业在尝试定制新的硬件架构，力图在性能、成本和功耗上与GPU竞争，很多企业建立了产业生态，可以肯定的是，GPU仍然会在相当长的一段时间内得到更广泛的应用。很多人都认为GPU的主导地位会被非GPU方案终结，例如FPGA或ASIC，但是非GPU方案都将面对两个艰巨的挑战：第一，能否实现量产，第二，整个产业生态链是否完整。目前，能够实现大规模商用、取代GPU地位的硬件尚未出现。