## 当前人工智能存在哪些安全风险？如何应对？

伴随着人工智能的发展，安全问题也无处不在。从手机扫码支付，智能穿戴设备，以及到人脸识别，自动驾驶等等人工智能技术深刻改变每个人的生活与工作，但这些都存在一定的安全风险，很容易触及到安全及隐私问题，黑客便可以获取用户信息，利用系统漏洞直接获取控制权限为所欲为。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。

对于安全性考虑不够，没有防攻技术支撑，是人工智能存在安全风险的主要原因。

例如 （1）人脸识别，技术隐患主要在活体认证部分，无特殊硬件支持的传统手机端的支付存在安全隐患，新的苹果手机基于深度摄像头，在活体认证部分极大提高，但依然利用假面具对其破解的报道见诸报端。

1. 自动驾驶，人工智能技术带来的潜在的安全风险也是多方面：传感器风险，不法分子可能通过干扰器让传感器失效，或者误导传感器，比如误导摄像头将车跑出路面。代码风险，人工智能算法的代码设计不周，可能让黑客有机会篡改实时代码，让结果出现偏差。产品风险，黑客可能远程挟持硬件或植入恶意代码。例如进行计算机深度学习时，一个’禁止停车’标志，可能被计算机解读为’限速’。

基于以上人工智能存在安全风险的例子，给计算机系统杀病毒、建防火墙，都是被动的，不解决根本问题，也是过时的。我们应该主动进行安全防护，例如从应用层，技术层，以及基础层等层面上做好安全性考虑，防攻技术支撑。

## 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的二人零和博弈，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. **数据众包**

所有的人工智能公司都追求巨大的数据库，以实现他们对人工智能的雄心壮志。这些公司将开始通过众包方式获取大量数据。企业已经找到了一种方法来评估众包数据的质量和真实性，不仅给企业提供了便利，还能反馈信息给消费者。

目前，谷歌通过众包的方式，获得了大量的图片，并构建了成像算法。该公司还利用众包来帮助改善服务，比如翻译、转录、手写识别和地图应用。而亚马逊还利用众包的技术改善了Alexa的1.5万项现有技能。

1. **所有神经网络的独特格式**

**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，例如计算机视觉的Cafle2, NLP的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，在这里我们遇到了ONNX标准：**

1. **AutoML更换管道**

**设计一个神经网络框架是一个非常痛苦的任务——虽然有时通过叠加卷积层能获得相当好的结果，但是大多数时候需要非常小心的使用直觉和超参数搜索方法设计宽度、深度和超参数，如随机搜索或者贝叶斯优化。特别是当你不在计算机视觉研究时，这意味着没法在ImageNet上完成一些DenseNet模型的训练，但是可以使用一些3D数据分类或多变量时间序列应用。**

1. **Zoos大爆发**

**深度学习中在计算机视觉任务和自然语言处理任务中将预训练的模型作为新模型的起点是一种常用的方法，通常这些预训练的模型在开发神经网络的时候已经消耗了巨大的时间资源和计算资源，迁移学习可以将已习得的强大技能迁移到相关的的问题上。**

**很多人开始用Cafle Zoo的模型进行迁移学习或特性提取，最近发现它就像大型计算机视觉管道的一部分一样。这意味着实际上没有必要训练自己的网络，例如，对于ImageNet对象的识别或局部识别，这些基本的东西可以下载并插入到您的系统中。使模型能够更快的训练使用到对应的场景中**

1. **加速转向对话式交互**

对于许多人工智能领域研究者来说，语义学习是他们的下一个主要目标。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，谁能让机器更懂人类的语言，谁就能有可能取胜。人工智能会全面包围我们生活的方方面面，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. **在智能芯片方面，GPU将进一步得到更广泛的应用**

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，但是由于英伟达公司投入巨资成功建立了产业生态，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。

1. **智能堆栈**

“AI堆栈”更深层次，在每一个层面都有自己的发展和研究，而不只是由机器学习算法和你最喜欢的框架构成。

阿里云云栖社区认为人工智能开发行业已经足够成熟，它需要有更多不同的专家。在团队中只有一名数据科学家是远远不够的——你需要不同的人来进行硬件优化、神经网络研究、人工智能编译器、解决方案优化、生产实现。以上人员必须是不同的团队领导，软件架构师（必须分别为每个问题设计上面的堆栈）和管理者。我曾经提到过这个概念，它提供了一些愿景，在未来人工智能领域的技术专家能够成长成什么样子(对于那些想成为人工智能或技术领域的软件架构师的人来说，你需要知道该研究什么)。