让AI恶意攻击打在“棉花”上

■本报记者 王之康

人们普遍相信耳听为虚、眼见为实，但在当前的信息时代，耳听却未必为虚，眼见也未必为实。基于深度学习的人工智能（AI）也可能会“造假”，甚至作恶。

好评“灌水”、声音模仿等，都是人工智能造假搞的鬼，甚至最近网络上流传的一段有关美国前任总统奥巴马和现任总统特朗普的视频，也是它在作祟：视频中，两人不仅话语一致，连说话节奏、面部肌肉动作都一致，而谈吐风格却仍旧保持各自特点。

“随着深度学习应用越来越广泛，越来越多的人工智能安全问题也开始暴露出来。”南京理工大学计算机学院教授李千目告诉《中国科学报》，深度学习框架中的软件实现漏洞、对抗机器学习的恶意样本生成、训练数据的污染等都可能导致人工智能所驱动的识别系统出现混乱，形成漏判、误判，甚至导致系统崩溃或被劫持，并可以使智能设备变成僵尸攻击工具。

对于人工智能潜在的造假“作恶”，由李千目带领的研究团队所开展的“面向人工智能对抗性恶意样本的监测技术”，可以有效提高在对抗环境中对恶意软件监测的可靠性和安全性，让智能攻击仿佛打在“棉花”上一样绵软无力。

这项研究成果已在人工智能国际学术会议AAAI2019上发表，算法及其实验获得大会“挑战问题赢家”奖，这也是中国信息安全学者首次获得该奖项。

李千目

对于人工智能潜在的造假“作恶”，由李千目带领的研究团队所开展的“面向人工智能对抗性恶意样本的监测技术”，可以有效提高在对抗环境中对恶意软件监测的可靠性和安全性，让智能攻击仿佛打在“棉花”上一样绵软无力。

微扰动致AI“将金字塔认作骆驼”

人工智能之所以能够造假甚至作恶，在李千目看来，实际上是有多种技术和理论支撑的，“人工智能系统的攻击技术主要包含对抗性输入、数据中毒攻击及模型窃取技术三个方面”。

其中，对抗性输入攻击是一种专门设计的输入，确保被误分类以躲避检测。当前，这一手段已被大量使用在专门用来躲避防病毒程序的恶意文档、试图逃避垃圾邮件过滤器的电子邮件等多种场景。

数据中毒攻击涉及向分类器输入对抗性训练数据，最常见的攻击类型是模型偏斜，攻击者以这种方式污染训练数据，使得分类器在归类好数据和坏数据时向自己的偏好倾斜。

模型窃取攻击则是通过黑盒来探测/窃取（即复制）模型或恢复训练数据身份，比如，可以用来窃取某股市预测模型或者某垃圾邮件过滤模型。

李千目指出，对抗攻击的理论基础是神经网络的两个“bug”：一是高维神经网络的神经元并不是代表着某一个特征，而是所有特征混杂在所有神经元中；二是在原样本点上加上一些针对性的、不易察觉的扰动，从而导致神经网络的分类错误。

“其中，第二个就是对抗攻击的理论基础。”李千目说，后来伊恩·古德菲尔等人在研究中提出，原因并非深层神经网络的高度非线性和过拟合，即使是线性模型也存在对抗样本，“鉴于此，我觉得可以粗浅地认为，对抗攻击之所以能够成功，原因是误差放大效应。”

比如，匈牙利裔美国计算机科学家马里奥·塞格德等人通过微量扰动，成功地使人工智能将校车和孔雀识别为鸵鸟，将金字塔识别为骆驼，从而首次证明可以通过对图像添加微量的人类察觉不到的扰动来误导神经网络作出误分类。

“虽然人类视觉系统很难察觉到这些干扰，但是它们对深度学习模式的影响却是灾难性的。”李千目说。

以分类器研究为切入点

针对典型智能算法训练过程中存在的数据来源未知和算法参数被污染的安全风险，李千目团队开始研究对抗性攻击样本生成模型，并设计相应的对抗性样本算法，来实现对抗性攻击样本生成。

“我们尝试增强深度学习模型，主要利用人工智能分类器对恶意软件进行分类，以恶意软件为输入样本，分为训练数据集和测试数据集两部分。”他介绍说，在训练阶段，训练多个人工智能分类器的集合，在每个分类器上都将所提出的原则系统化地加以运用；在测试阶段，将样本输入至每个分类器，最后根据所有分类器的投票结果确定样本是否为对抗性恶意软件。

在团队成员李德强博士看来，这项技术的关键是在训练阶段，同时也是研究的难点和创新点所在，“我们提出了一些设计和规避的原则，用于恶意软件分类的设置中，增强智能分类器对抗性逃避攻击的可靠性和安全性，实现了部分种类恶意数据样本的高效快速检测以及样本数据防污染、防篡改等”。

“常规的人工智能完全是用计算机去识别，如果碰到恶意攻击，即把一些偏差样本不断地加入其中，慢慢的，人工智能就会认为这件事情是真的，从而导致判断出现偏差。”中国科学院上海技术物理研究所杭州院副院长、研究员陈峰磊说，比如，1000个样本中有990个是对的，那么人工智能就会识别这些是对的；如果其中再加入10000个偏差样本，人工智能就会认为这10000个偏差样本是对的。

因此，他认为，该研究最大的创新点在于人工智能与人重新结合起来，也就是人必须干预人工智能，特别是当样本具有明显倾向性的时候，人就要介入其中进行判断。有些倾向性是正常的，那么就是正常的加权；有些倾向性是错误的，就要把这个加权降下来。“同时，这也是该研究非常重要的价值之一，就是人工智能不能离开人，否则迟早会被带偏。”

阶段性成果亦有用武之地

李千目介绍，该领域的研究在国内外都处于起步阶段，相信在两年内会有突破性成果出现，“就目前来说，我们的研究处于国际同类研究的先进水平”。

虽然处于起步阶段，但他们的研究成果却有着广泛的应用前景。

比如，可用于规范工业互联网、智能无人系统等领域的顶层安全设计，通过统一智能系统的安全体系架构，增强无人系统等智能系统的安全互联互通互操作能力，提升智能系统的信息防御能力；也可用于指导智能平台、工业互联网的产品安全研制，通过通用化、标准化、组件化，使得各种安全功能构件可重用、可替换，大量减少采购费用等。

“不过，这项研究和其他研究不同，即便是阶段性成果，也可以在领域里面进行应用。”李千目说，因为恶意监测就像医生治病一样，一个医生不可能治愈所有疾病，但拥有一种有效治疗手段就可以将其用于临床。“目前，我们正牵头制定某项工业互联网安全的标准，其中就涉及部分研究成果，同时我们也在某示范项目中推动了该成果的应用。”

不过他也表示，这类研究的成果由于需要定制、部署和依照对象不同而进行适配的工作非常多，所以短期来看实现产业化比较难；或者说，作为一项产品，需要它的用户比较少。“但无论是军用还是民用，这类研究都非常有价值，将助力网络强国、智造强国建设。”

相关论文信息：

DOI:arxiv.org/abs/1812.08108

异言堂

5G时代的安全问题将是『核心』问题

■田溯宁

在过去一段时间，没有什么话题比5G更热。考虑到5G跟云计算联系密切，可以说，我们进入了一个云网一体化的时代。我关心的是，云网一体化会给网络安全带来哪些挑战和机会？

30年前，互联网开启了计算信息技术领域波澜壮阔的历史。但是直到今天，网络的密切连接才刚刚开始。过去互联网连接了几十亿人，随着5G在2020年前后的部署，未来互联网可能要连接7万亿个设备、数据和商业流程。因此，当我们谈论5G，一个核心的命题就是，5G将实现从人的连接到物的连接、从知识连接到商业流程的连接，开启产业互联网元年。

如果说3G、4G是移动互联网的基础设施、带来了移动互联网应用，那么5G正在成为产业互联网的基础设施，其特征就是云网一体化。云可以不断积累数据，连接则不断进化出智能。

然而当这么多连接在一起的时候，安全就成了重要的课题。

我们认为人连接产生的安全问题小于物连接产生的安全问题。这是因为，无论是汽车飞机还是终端设备，一旦产生密切连接，人们主要的生产和生活流程也就跟它们密切连接在一起——人遇到安全问题时，能够有各种防护措施，“物”则不然。因此，5G将使安全问题变得非常困难和尖锐。

在此背景下，当我们向往5G新时代、憧憬中国在5G时代走在前面的时候，云网一体的安全也必须跟上时代。因此，在产业互联网语境下谈到信息技术（IT）、运营技术（OT）、通信技术（CT）的融合时，网络安全技术（ST）同样不容忽视。也就是说，5G云网一体化的时代，IT、OT、CT将会与ST相融合。如何融合？这需要共同探讨。

现在的企业很缺乏知识积累。一家企业想要拥有“智商”，必须把数据通过5G、通过万物互联收集起来，如此企业的智识才能沉淀——这将是产业互联网的一个特征。然而就像电力是工业革命的基础设施一样，产业互联网基础设施的一个前提条件，就是必须把安全技术深入到每一个环节。从事网络安全事业的人们会逐渐发现，ST将成为未来人类信息化、智能化进程中的核心技术之一。

（作者系宽带资本董事长、亚信集团董事长，本报记者赵广立根据其在2019 C3安全峰会上的演讲整理）

“人工智能开放科研教育平台”扩展共享资源

本报讯 5月10日，《中国科学报》记者从在京举行的“中国高校人工智能人才国际培养计划”2019国际人工智能专家论坛暨2019微软新一代人工智能开放科研教育平台合作论坛上获悉，经过一年的建设，新一代人工智能开放科研教育平台推进了“基础支撑”“实践案例”“打造‘金’课”等七个维度的核心工作，吸引了越来越多高校的参与，不断完善其教育合作体系。2019年，该平台将在平台、工具、数据、课程与实践等四个方面进一步增加开放共享资源，促进以该平台为核心的人工智能产教融合生态健康、可持续发展。

2018年5月22日，新一代人工智能开放科研教育平台在微软亚洲研究院正式对外发布，面向中国高校开放合作。平台首席顾问包括中国科学院院士、中国科学技术大学校长包信和，中国工程院院士、北京大学博雅讲席教授高文，中国工程院院士、浙江大学教授潘云鹤，美国工程院院士、微软全球执行副总裁沈向洋，中国工程院院士、西安交通大学教授郑南宁等院士专家。

微软亚洲研究院副院长周礼栋向《中国科学报》介绍，经过一年的建设，30所高校通过该平台与微软展开了技术交流、联合科研、课程共建、师资培训等合作。例如，在平台的合作框架下，由微软推出的国内首个针对人工智能深度学习领域、由多方共同开发的开源平台OpenPAI已帮助多所高校与科研机构建立了属于自己的人工智能基础支撑平台。其中，中国科学技术大学的“类脑智能技术及应用国家工程实验室”就是基于Open-PAI搭建的一个开放、共享的科研教育平台，并已在真实的科研、教育场景中进行应用。

中国科学技术大学校长助理、信息科学技术学院执行院长、类脑智能技术及应用国家工程实验室主任吴枫告诉记者：“中国科学技术大学基于微软OpenPAI所搭建的类脑智能云平台很好地支持了类脑实验室的科研工作，未来平台的应用范围将会继续扩展。”截至目前，OpenPAI提供的智能平台确保了中国科大类脑实验室每月49万GPU小时数的稳定运行，以及学校600多名师生的研究需求，让他们可以为多媒体、自然语言处理、基础理论研究等不同研究方向定制不同的范式。

周礼栋透露，2019年微软亚洲研究院将继续提升OpenPAI的用户体验、核心深度学习能力支持、调度系统的可扩展性以及系统稳定性，并与合作高校进一步以OpenPAI为基础进行深度学习算法和系统相关的研究，包括自动化机器学习算法、分布式深度学习、自动化网络压缩、自动深度神经网络搜索以及GPU调度算法等。 **（计红梅）**

依图科技首席创新官吕昊展示基于“求索”的服务器板

依图科技推出云端AI处理器

本报讯 近日，人工智能创企“四小龙”之一依图科技在上海发布了其首款自研的云端视觉推理芯片“求索”（questcore），并在发布会现场实时演示集成了4颗该芯片的云端服务器对多达200路摄像头的1∶1人脸识别的智能计算支持。

“‘求索’芯片不是一个AI加速模块，而是一个完整的具有端到端能力的AI处理器。”依图科技联合创始人、CEO朱珑对“求索”如此定位。

基于现场演示，朱珑表态依图科技不是“为了造芯而造芯”，而是构建基于该芯片的软硬件一体化的行业解决方案。同时他还表示，现场成功的演示也宣告了该芯片已经可实现量产、投入商用。

性能方面，依图科技首席创新官吕昊介绍说，基于“求索”芯片打造的依图“原子服务器”（搭载4核芯片），提供的算力与8张英伟达P4卡服务器相当，而体积仅为后者的一半，功耗不到20%。在进行视频解析时，1台依图原子服务器，与8卡英伟达T4服务器（含双核英特尔x86 CPU）对比，单路视频解析功耗仅为后者的20%，与8卡英伟达P4服务器（含双核英特尔x86 CPU）相比，功耗约为后者的10%。

对于性能的获得，朱珑解释称这是由“算法即芯片”的理念达成的，“找对问题，找对场景，用对算法，并为此定制芯片，才有可能做到极致性价比。”

《中国科学报》了解到，该芯片系依图科技与上海熠知电子科技有限公司（以下称ThinkForce）合作研发，其中依图主要提供视觉算法，ThinkForce主要承担硬件研发。值得一提的是，ThinkForce是依图科技在2017年战略投资的AI芯片初创团队。天眼查数据显示，依图科技是其第二大股东，持股比例为28.92%。

此外，“求索”芯片基于ARM+Manycore架构研制，采用16nm制程工艺封装。 **（赵广立）**

速递

商汤连推11款产品 打响AI落地战

本报讯 5月15日，人工智能创企“四小龙”之一商汤科技围绕智慧城市、AI教育、AI医疗、智慧零售和增强现实（AR），在第四届商汤人工智能峰会上一口气发布了覆盖5类应用场景的11款AI产品，打响了商汤AI产品落地战。

11款产品分别是面向智慧城市领域的能力基础设施SenseFoundry方舟2.0、边缘AI能力中心SenseNebula星云、AI端节点产品SensePass、视觉管理平台SenseStudio，面向AI教育的初中AI教材《人工智能入门》、AI课程创作工具SenseStudy、教育机器人小车SenseRover Mini/Pro，面向AI医疗场景的SenseCare智慧诊疗平台，针对智慧零售场景的SenseGo智慧零售视觉平台，面向AR的SenseAR特效引擎以及SenseAR 2.0开发者平台。

“如果去年是商汤单点落地元年，今年则是全场景覆盖的重要一年。”商汤科技智慧城市事业群首席运营官高焱说。

谈及利用AI技术打破不同行业、不同领域、不同地域之间的界限，商汤科技创始人汤晓鸥表示“和朋友一起走才能成功”，表达了商汤科技希望与合作伙伴和客户紧密合作，共同推进AI的创新与落地。

商汤科技联合创始人、首席执行官徐立也表示，希望通过新产品和解决方案在各个垂直领域的渗透，与合作伙伴“共赢共生”。 **（赵鲁）**

2019 ABB杯 智能技术创新大赛启动

本报讯 近日，由ABB（中国）有限公司主办、中国自动化学会协办的“2019 ABB杯智能技术创新大赛”发布参赛题目、赛程安排及赛制规则，这标志着该项赛事启动。

本届大赛报名及作品递交的截止日期为2019年8月31日，入围选手将在今年10月举行的决赛答辩环节角逐一、二、三等奖。除了可赢取奖金之外，获奖选手还将有机会拿到ABB（中国）有限公司的招聘直通车“绿卡”。

记者了解到，ABB杯智能技术创新大赛的前身是“ABB杯全国自动化系统工程师论文大赛与ABB大学生创新大赛”，前者由中国自动化学会与ABB于2005年联合创办，累计已有10910人参赛，提交有效论文4314篇；后者自2012年首次举办以来，已成功举办7届，累计吸引8000余名高校学生参与。

融合两大赛事后，ABB杯智能技术创新大赛将面向中国内地及港澳台地区公众及普通高校和高职院校学生开放报名通道，参赛选题包括智能制造、智慧园区和智能建筑三大类别，供选手从中选择题目、设计与创作。

同时，ABB还将在官网及微信号设置投票页面，评选“最佳人气团队奖”。

ABB集团是世界500强企业，作为电力和自动化技术领域巨头，其业务遍布全球100多个国家和地区。ABB在中国拥有研发、制造、销售和工程服务等全方位的业务，拥有44家本地企业，并有近2万名员工遍布于130余个城市。 **（赵广立）**

中科视拓发起 成立“AI++人工智能学院”

本报讯 近日，由中科视拓建立的“AI++人工智能学院”在上海举行揭牌仪式，中科视拓华东区负责人、上海分公司总经理方圆圆为与会代表作了《中国AI产业人才报告》。

报告介绍了视拓探索出来的AI高质量人才培训之路：AI人才通过中科视拓行业应用平台进行理论系统+行业开发实训培养+平台测试，可清楚了解人才的进阶程度，大大缩短人才培养周期，并依靠长期人才发展计划与创业扶持计划进行长足的发展。

据了解，AI++人工智能学院将通过线上与线下相结合，多行业应用案例课、阶梯课程、前沿讲座、实操实测、水平审核等系列设置，将理论、实践、人才评测相结合，为企业培养应用级AI技术人才。

会上，AI++人工智能学院与首期6家合作企业进行了签约仪式。此次战略签约打通了AI产、学、研、创，助力企事业AI升级全链条推进与学员职业发展及创业孵化。 **（田瑞颖）**

2019工业安全大会聚焦信息安全

本报讯 5月9日，由工业控制系统信息安全产业联盟主办的“2019工业安全大会”在京召开，130余位与会代表共同探讨了数字经济时代下工业安全的前沿技术和产业生态建设。

工业和信息化部信软司副司长王建伟在会上表示，加强工业信息安全防护、推动新一轮工业变革，已经成为当前重塑工业发展新优势、抢占竞争制高点的战略选择。在今后的工作中，将继续加大支持力度，围绕关键环节，完善工业信息安全规章制度，强化工业企业安全主体责任，筑牢工业信息安全技术防线，培育工业信息安全产业生态，推动工业控制系统信息安全产业健康发展和工业互联网安全保障体系的快速构建。

大会还发布了工业安全产业联盟首批38位智库专家名单。 **（张楠）**

王东升获 “SID David Sarnoff产业成就奖”

本报讯 美国时间5月13日，国际信息显示学会在美国圣何塞开幕的国际显示周（SID Display Week 2019）上评选出了全球信息显示领域杰出贡献者。京东方（BOE）创始人、董事长王东升因其对全球半导体显示产业发展做出的贡献获得了“SID David Sarnoff产业成就奖”。

据悉，David Sarnoff产业成就奖是针对在全球显示产业发挥了卓越领导力、产生了长远影响、被业界广泛认可的杰出贡献者而设立的奖项。

王东升提出的“显示产业生存定律”，也被业界称之为“王氏定律”：若保持价格不变，显示产品性能每36个月须提升一倍以上，这一周期正被缩短。该定律揭示了技术价值创造驱动对于提升企业价值的作用和影响，成为企业实现稳定盈利的生存法则。他还提出了半导体显示的概念，为全球业内所采用。 **（计红梅）**