新闻热线：010—58884102
E—mail：zdd@stdaily.com
■责编 翟冬冬
科技日报 2020年5月6日 星期三


视觉中国供图

数据投毒致人工智能失控
AI杀毒软件市场尚为一片蓝海

实习记者 **代小佩**

一辆正常行驶的自动驾驶汽车，突然驶入了逆行车道；胸前贴一款特殊贴纸，犹如披上隐形斗篷，在监控系统中成功遁形；戴上一幅特制眼镜，轻松骗过人脸识别系统后，用别人的手机也可实现刷脸解锁或刷脸支付……

小心，这可能是遇上了难缠的AI病毒！

近日，清华大学人工智能研究院孵化企业推出了针对人工智能算法模型本身安全的RealSafe安全平台，据介绍，该平台可快速缓解对抗样本的攻击威胁。

人工智能感染的是什么病毒？其安全问题有哪些特点？人工智能时代，杀毒软件如何修炼才能化作身怀绝技的病毒猎手？

是敌又是友 对抗样本戴着双重面具

RealSafe人工智能安全平台，是针对AI在极端和对抗环境下的算法安全性检测与加固的工具平台，包括模型安全测评、防御解决方案两大功能模块。平台内置AI对抗攻防算法，提供从安全测评到防御加固整体解决方案。

北京理工大学计算机网络及对抗技术研究所所长闫怀志接受科技日报记者采访时表示，上述平台目前侧重于模型和算法安全性检测与加固，可以说是人工智能算法的病毒查杀工具。

闫怀志说，针对人工智能系统实施对抗样本攻击的这类恶意代码，常被称为“AI病毒”。对抗样本是指在数据集中通过故意添加细微的干扰所形成的输入样本，会导致模型以高置信度给出一个错误的输出。

“其实在实验室中，使用对抗样本可以检测许多训练学习类人工智能方法的分类有效性，也可以利用对抗样本来进行对抗训练，以提升人工智能系统的分类有效性。”闫怀志告诉科技日报记者。也就是说，对抗样本可以看成是训练人工智能的一种手段。

“但是在现实世界，攻击者可以利用对抗样本来实施针对AI系统的攻击和恶意侵扰，从而演变成令人头疼的‘AI病毒’。”闫怀志表示，对抗样本攻击可逃避检测，例如在生物特征识别应用场景中，对抗样本攻击可欺骗基于人工智能技术的身份鉴别、活体检测系统。2019年4月，比利时鲁汶大学研究人员发现，借助一张设计的打印图案就可以避开人工智能视频监控系统。

在现实世界中，很多AI系统在对抗样本攻击面前不堪一击。闫怀志介绍，一方面，这是由于AI系统重应用、轻安全的现象普遍存在，很多AI系统根本没有考虑对抗样本攻击问题；另一方面，虽然有些AI系统经过了对抗训练，但由于对抗样本不完备、AI算法欠成熟等诸多缺陷，在对抗样本恶意攻击面前，也毫无招架之力。

对训练数据投毒 与传统网络攻击存在明显不同

360公司董事长兼CEO周鸿祎曾表示，人工智能是大数据训练出来的，训练的数据可以被污染，也叫“数据投毒”——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。

中国信息通信研究院安全研究所发布的《人工智能数据安全白皮书（2019年）》（以下简称白皮书）也提到了这一点。白皮书指出，人工智能自身面临的数据安全风险包括：训练数据污染导致人工智能决策错误；运行阶段的数据异常导致智能系统运行错误（如对抗样本攻击）；模型窃取攻击对算法模型的数据进行逆向还原等。

值得警惕的是，随着人工智能与实体经济深度融合，医疗、交通、金融等行业对于数据集建设的迫切需求，使得在训练样本环节发动网络攻击成为最直接有效的方法，潜在危害巨大。比如在军事领域，通过信息伪装的方式可诱导自主性武器启动或攻击，带来毁灭性风险。

白皮书还提到，人工智能算法模型主要反映的是数据关联性和其特征统计，没有真正获取数据之间的因果关系。所以，针对算法模型这一缺陷，对抗样本通过对数据输入样例，添加难以察觉的扰动，使算法模型输出错误结果。

如此一来，发生文章开头所谈到的一类事故就不足为奇了。

此外，模型窃取攻击也值得注意。由于算法模型在部署应用中需要将公共访问接口发布给用户使用，攻击者就可以通过公共访问接口对算法模型进行黑盒访问，并且在没有算法模型任何先验知识（训练数据、模型参数等）的情况下，构造出与目标模型相似度非常高的模型，实现对算法模型的窃取。

闫怀志在采访中表示，AI安全更突出功能安全问题（safety），这通常是指人工智能系统被恶意数据（比如对抗样本数据）所欺骗，从而导致AI输出与预期不符乃至产生危害性的结果。“AI功能安全问题与传统的网络安全强调的保密性、完整性、可用性等信息安全问题（security），存在本质不同。”

预防“中毒”困难重重 AI技术也可构筑网络安全利器

闫怀志表示，目前种种原因导致了预防人工智能“中毒”困难重重，原因具体表现在三个方面。

一是很多AI研发者和用户并没有意识到AI病毒的巨大风险和危害，重视并解决AI病毒问题根本无从谈起；二是由于AI正处于高速发展阶段，很多AI研发者和生产商“萝卜快了不洗泥”，根本无暇顾及安全问题，导致带有先天安全缺陷的AI系统大量涌入应用市场；三是部分AI研发者和供应商虽然意识到了AI病毒问题，但由于技术能力不足，针对该问题并无有效的解决办法。

“当然，网络安全本来就是一个高度对抗、动态发展的领域，这也给杀毒软件领域开辟了一个蓝海市场，AI杀毒行业面临着重大的发展机遇。”闫怀志强调，杀毒软件行业首先应该具有防范AI病毒的意识，然后在软件技术和算法安全方面重视信息安全和功能安全问题。

“以现实需求为牵引，以高新技术来推动，有可能将AI病毒查杀这个严峻挑战转变为杀毒软件行业发展的重大契机。”闫怀志强调，AI技术既会带来网络安全问题，也可以赋能网络安全。

一方面，人工智能的广泛应用带来了许多安全风险。由技术性缺陷导致的AI算法安全风险，包括可导致AI系统被攻击者控制的信息安全问题；也可导致AI系统输出结果被攻击者任意控制的功能安全问题。

但另一方面，人工智能技术也可以成为构筑网络空间安全的利器，这主要体现在主动防御、威胁分析、策略生成、态势感知、攻防对抗等诸多方面。“包括采用人工神经网络技术来检测入侵行为、蠕虫病毒等安全风险源；采用专家系统技术进行安全规划、安全运行中心管理等；此外，人工智能方法还有助于网络空间安全环境的治理，比如打击网络诈骗。”闫怀志说。

中国信息通信研究院安全研究所的专家称，为有效管控人工智能安全风险并积极促进人工智能技术在安全领域应用，可从法规政策、标准规范、技术手段、安全评估、人才队伍、可控生态等方面构建人工智能安全管理体系。

靠“脑补” AI将卫星“废片”变成高分辨率地图资源

本报记者 **张佳星**

卫星虽被形象地称为“天眼”，事实上卫星数据的应用，却难以做到“尽收尽用”。

中国工程院院士杨小牛就曾经表示，卫星在天上飞来飞去，效能其实并不高，每天只有几十分钟时段内采集到的数据是地面需要的。

那些大量的被认为是无效数据的“废片”，有没有可能被利用起来？

近日，清华大学理学院院长、地球系统科学系主任宫鹏介绍，在高性能云计算的支持下，通过数据建模、人工智能算法等手段，清华大学地球系统科学系制作完成了首套中国30米逐日无缝遥感观测数据集，以及逐季节土地覆盖和逐年土地利用的数据集，让“拼图无解”的卫星“废片”能够成为高分辨率的地图资源。

人工智能深度“补片”

“传统的对地卫星观测，拍下来的照片是不同时间采集的，拼在一起并不完整，使用门槛很高。”宫鹏解释，卫星直接获得的图片不能拿来直接用，因为卫星图片不是自然连续的，很可能像100块的拼图，有时候是少了50块的效果，但也有可能同样的拼图来了好几块。

不止如此，卫星轨道的偏差还会造成同一地方不同时间拍摄的图片难以重叠，云彩的遮挡、雾气不均匀的散射都会导致大量的卫星遥感数据难以反映地表的真实情况，而成为难用的“废片”。

基于此前完成10米分辨率全球地表覆盖制图数据处理过程中积累的经验，清华大学地球系统科学系团队自主研发了时空数据融合重建的技术。

“我们构建了人工智能需要的知识库，其中包括世界首套全球全季节普适样本库和相关领域知识。库中分为训练样本库和完全独立的验证样本库。”清华大学博士生刘涵介绍，团队设计了一套适应遥感大数据的深度遥感特征学习和分类模型，利用机器学习和数据建模对人工智能系统进行了训练，使其能够“理解”或者说“推断”出缺失的图块，进而补缺。

“就好像现在一些APP身份验证时，会有一个补图的步骤，经过训练的模型，也可以大规模分析现有的卫星图片，自动补图，且做到数据与真实情况相符合。”刘涵说。

通过训练，模型可完成高性能的推理，把不完整的“拼图”重建成时空一致的图像库，建立起这一深度遥感制图模型的“超能力”，完成各种不合格“废片”的补片工作，从而生成与真实情况相匹配的遥感观测数据集。例如通过人工智能技术可识别路面是沥青、土路还是水泥路面等地表覆盖类型。

计算上云避免巨大资源消耗

“地球系统科学使用和产生的数据是极其巨大的，例如气候模拟和预测会生成时间间隔在小时级、地面分辨率是3公里的气候数据，这些数据的数据量级往往相当于数百万集高清电影的量级。”宫鹏介绍，因此需要超强算力来完成。

如果为这些数据进行数据中心建设的话，需要三四百个机柜，占地成本和时间成本耗费巨大。对这些数据集中的数据进行AI处理时，如果不在云上进行而是搬运下载后再运算，那光是用来搬运的时间也可能需要几个月。

而通过云上高性能计算，则能够把算力部署在公共数据集周边，围绕数据进行计算。据介绍，亚马逊云服务（AWS）为此次项目的完成提供了10万核左右的云上高性能计算资源。

此外，AWS上还提供一整套完善的人工智能和机器学习的套件和服务以及自动化多层堆叠集成技术，可用于对模型结构和参数进行深度调优，并进行分布式高性能推理。

“现在中学生、小学生想拿数据做点什么，从里面拉几条曲线，或者把一个区域拿出来做一些探测、变化、趋势的分析，都已经变得非常容易。”宫鹏说，对于卫星公共数据的梳理、重建，让卫星遥感图的使用门槛大大降低，如果说之前只有专业用户能从数据中获得价值，那么以后更多的普通用户也能看懂和利用这些数据。

情报所

美国专利商标局裁定 人工智能不能被列为发明人

据外媒报道，近日美国专利商标局（USPTO）公布了一项裁定结果，表示人工智能不能被列为发明人。目前，只有自然人才有权利获得专利。

去年，两项专利——一种可变形的食品容器和一种应急手电筒给专利法规提出了一个疑问：专利发明人必须是人吗？

这两项发明是物理学家和人工智能研究者斯蒂芬·泰勒创造的人工智能系统DABUS的作品。现在，USPTO已经裁定，DABUS和其他任何人工智能都不能被列为专利申请中的发明人。

此案之前，美国专利法对机器是否可以作为发明人的规定并不明确。斯蒂芬·泰勒和部分专利法专家认为，由于泰勒在容器或手电筒方面没有任何专业知识，也没有帮助人工智能系统制造这些发明，因此他被列为发明人是不合适的。“如果我教我的博士生相关知识，而他们最终做了一个比我之前教导的复杂得多的决定，这并不能使我成为他们专利上的发明人，对于人工智能系统来说也是如此。”领导人工智能专利项目的法律专家、英国萨里大学的法律和健康科学教授瑞安·雅培表示。

日前，在英国，根据当地禁止非自然人作为发明人的专利法，人工智能系统DABUS的专利申请已被驳回。随着近日裁定结果的宣布，美国也明确了“只有自然人可以在专利申请中被列为发明人。”

（来源：cnBeta.COM）

图说智能

一批具有消毒防疫功能的地埋式垃圾桶日前在深圳启用。据介绍，地埋式垃圾桶的收集桶位于地面以下，清运时才露出地面，桶内垃圾满溢可自动报警，通知环卫人员清运。此外，该类地埋式垃圾桶还配备排水系统与灭火系统，并可在地下密闭空间进行自动消毒。

图为近日在深圳市福田区拍摄的地埋式垃圾桶。

新华社记者 **卢烨**摄

素有“渝西门户”之称的荣昌，已从重庆、成都中间地带的“产业塌陷区”变身为投资热土，成为重庆发展最快的地区之一，也见证着成渝从“背向发展”到“相向发展”的沧桑巨变。

图为近日在唯美陶瓷荣昌生产基地的自动化生产车间内，一台工业智能机器人在搬运成品瓷砖产品。 新华社记者 **刘潺**摄

河北省唐山国家高新技术产业开发区的高端装备制造业企业在做好疫情防控全面复工复产以来，开足马力赶制订单，满足客户需求。据介绍，目前该区高端装备制造企业达44家。

图为近日，工人在唐山国家高新技术产业开发区一家特种机器人生产车间工作。 新华社记者 **杨世尧**摄