全球智能制造研究专家杜如虚:

韧性智能制造把握三大关键点

当今世界面临前所未有的挑战,全球竞争空前激烈,未知的危机接踵而来。8月26日,全球智能制造研究专家杜如虚近日在报告会上围绕“智能制造产业创新、布局与应变”作了精彩演讲。杜如虚指出,“人类社会本身是具有韧性的,不会被这样或者那样的事件轻易地打垮”。关于人工智能如何获得韧性,杜如虚认为,需要抓住创新、布局、应变三大关键点。

创新是社会发展的引擎,目前社会创新速度在不断加快,而创新包括产品设计和制造技术创新,杜如虚强调,三分之二的创新跟制造有关系,此外还有供应链与市场的创新。

企业和社会在发展中想要持续保持强大的“韧性”,布局也是一大关键点。杜如虚表示,所谓布局包含“硬性”和“柔性”两大方面。硬性是强调严格的质量控制。在质量管理过程中,从头到尾都必须考究,每一个环节都一定要做好。杜院士对比了中国质量和日本质量指出,我们的质量管理系统相对来说不够健全,“不是说管理系统不够健全,而是管理理念有待提高”。他举例说,“丰田车的质量管理不是只看最后产品的质量,而是看产品加工中每一段的质量,它们有一个全程质量控制。

与此同时,在打造好硬件布局的基础上,还要兼顾“柔性”布局。在谈到如何让系统布局的时候更有柔性时,杜如虚分享了团队的研究经验,“尝试把应变机制植入制造系统”,并对比了新的方法与旧方法。传统的方法是多台机床并行或堆栈,这样会导致成本增加、材料浪费。而新的方法则可通过重构机床、可移动堆栈或动态规划。

杜如虚强调,保持韧性的第三个关键点是应变,这里的应变主要是工业人工智能。目前工业人工智能会面临非常多的挑战,特别是数据安全、数据孤岛、故障数据很少,许多故障的数据无法获得等问题,而解决这些问题的关键就是人工智能(AI),智能生产过程有三个步骤,首先是信号采集与处理;接着是建模预测,这需要AI;最后是控制决策,也需要AI。

杜如虚指出,智能制造系统的构建包含了三层:第一层各种装备集成、数据采集,第二层数字孪生,“最重要的是第三层,发掘深层问题”。他强调:“怎么样监控诊断,发现其中问题的根源,然后防患未然。最后还要优化系统,使得系统能够不断地提高,这才是智能制造的关键所在,就是韧性。”

人工智能的发展,包括硬件和软件。但算法是一个核心。杜如虚指出,人工智能算法的目的是从观测数据中学到尽可能多的知识。“如果人工智能是一个大饼,其中的机器学习占80%。机器学习的目的是从数据中学到尽可能多的东西。”

最近的一些算法包括谷歌的“自我专注学习”,对比学习,变分自编码器,及对抗神经元网络。杜如虚着重介绍了对抗神经元网络,并分享了他的团队基于对抗神经元网络开发的新算法——自增强学习。杜院士说:“如果我们有足够的数据,可以不用对抗的方法,直接用深度学习就可以了。但在实际应用中,总会遇到数据孤岛的问题。因此,要用对抗神经元网络,要用我们的自增强学习方法。”

袁斯茹

中国科学院自动化研究所研究员李兵:

人工智能是把“双刃剑”

新一代人工智能正在全球范围内蓬勃兴起,为经济社会发展注入新动能,但人工智能技术如果运用不当,也会给公共安全、道德伦理、社会治理等带来威胁。9月5日,人民中科董事长,中国科学院自动化研究所模式识别国家重点实验室研究员,知名青年AI专家李兵表示:“人工智能技术的快速发展,在一定程度上为网络安全的发展提供了很大的支撑,同时也是一把双刃剑,也会对网络安全带来新的隐患。”

人工智能发展 带来一定安全风险

人工智能技术存在算法黑箱、技术滥用、侵犯隐私等安全问题,随着人工智能与实体经济深度融合,这些风险将会进一步叠加放大,给公共安全、道德伦理、社会治理等带来挑战。李兵介绍,目前人工智能安全风险主要集中在三个方面:

一是技术内生风险。深度学习作为运用最为广泛的人工智能技术,存在“算法黑箱”“不可解释”等缺点,难以对其实施检测检查,一旦算法模型中存在漏洞,其产生的安全风险问题可能对社会安全产生重大威胁。华盛顿大学等高校的研究人员在一篇论文中披露,自动驾驶汽车并不能“真正理解”指示牌上的图案,而是通过一些特征来进行识别,人们只需要在交通标志上用贴纸进行相应处理,就可以让系统错误地将停车标志识别为限速标志。

二是技术滥用风险。技术发展是把双刃剑,随着人工智能应用日益深入,出现了利用计算机视觉、智能语音等技术实现“换脸”“换音”的情形,不但侵犯个人隐私,还可能被用于实施诈骗,产生严重后果。2019年3月,媒体报道,有犯罪分子使用深度伪造技术成功模仿了英国某能源公司在德国母公司CEO的声音,诈骗了220000欧元,造成严重的财产损失。

三是数据隐私风险。人工智能技术训练需要依靠大量数据,但当前数据的获取、使用的监督管理尚不完善,存在隐私泄露、数据滥用等风险,同时数据作为重要资产,跨境流动时处理不当还可能对国家安全产生威胁。2020年2月,服务于600多家执法机构及安保公司的美国人脸识别创业公司Clearview AI称其客户面部信息数据库被盗,据报道,Clearview AI从网络社交媒体上抓取了超过30亿张照片,这些数据在采集时并未明确获得用户的同意。

人工智能技术是把双刃剑

作为国内资深内容安全领域专家、人民中科内容安全公司创始人,李兵认为“网络安全一直是个永恒的主题,有新的技术引入就一定会带来新的安全隐患,随着人工智能技术的快速发展,在一定程度上为网络安全的发展提供了很大的支撑。但它也是一把双刃剑,也会对网络安全带来新的隐患。比如虚假信息伪造、深度模型本身的安全隐患等。”

目前各城市已在数字化、智能化的道路上获益,实现管理效能提升、运营成本降低,为智能城市发展注入“数字动力”。当前世界上已有2亿个摄像头监视着我们的一举一动,城市中随时有数字化摄像设备,观察和监视着个体的行为,并识别嫌疑人员的相关行为,提高了公共空间的安全性。但不利的一面是容易陷入到监控视频的海洋中,寻找准确的图像犹如大海捞针。

而面部识别技术为该问题提供了解决方案。李兵指出,面部识别技术目前比较主流的是基于深度学习、深度神经网络的方法,该方法是一个端到端的黑盒,只需要关注输入大量的训练样本,输出相应结果的过程,不需要关注模型本身,技术性能相对较好,能将人脸识别的性能大幅度提高。但同时也存在一个缺点,需要大量的大数据样本进行模型训练,同时在训练的过程中需要大量的计算资源,例如GPU、CPU等相应的硬件支持。

该技术解决了依赖高清图像才能识别的问题,但面部识别技术同样会遇到一些问题,如通过一些服饰伪装等手段,系统就无法正常完成人脸识别。如今智能技术与真实世界的融合效果更加自然,甚至出现了以假乱真的换脸技术,也就是将A人的脸换到B人的脸上,让人难以分辨真假。李兵表示:“随着AI技术的发展,AI换脸过程中是保持B人脸部基本特征情况下,将A脸的生物特征进行深度嵌入,视觉是完全没有办法视察出来的。人脸合成和人脸交换会带来巨大的安全隐患,造成人身份信息的混乱及虚假信息的传播,这是目前面临的非常大的问题。”

强化AI内容安全技术研发

人民中科作为中国科学院自动化研究所技术成果转化平台,以先进的算法研发能力和算法优化迭代能力为核心点,聚焦于内容安全领域的产业应用及产品开发,为内容安全产业提供新方案,构建有责任、有价值的人工智能生态。

不久前,人民中科与中科院自动化所国家模式识别实验室的研究团队提出了一种基于身份空间约束的伪造人脸检测新方法。团队从实际应用出发,通过大量的科学观察和实验发现,公众人物或特定的人物在实际的人脸交换检测任务中,他们的身份总是已知的,或者每个人至少有一个真实的面部图像。既然卷积神经网络单凭待测图像进行分类的泛化性能不佳,而参考人脸图像又包含了相应身份人物的先验信息,这些信息利用起来可为伪造人脸图像鉴别模型提供重要判定依据。该思想对实际应用较为简单、合理,有助于克服泛化问题。

李兵认为:“目前我们在安全领域做了关于对抗人脸交换,进行人脸鉴伪的AI技术工作,即对伪造人脸进行鉴别,从而保护公民的个人隐私。目前我们提出了在人的生物特征区域,在不同尺度上细分地找到对应的原始人脸和被交换人脸之间的不兼容点或伪造点,来判断脸部图像是否是伪造合成的。” **中科**