新闻热线：010—58884061
E-mail：kanaxi@126.com
科技日报
■责编 陈 莹
2018年7月9日 星期一

我们需要怎样的人工智能

『暗黑版』AI现身引忧虑——

实习记者 崔爽

近日，麻省理工学院媒体实验室出品了一个“暗黑版AI”，再次将人工智能的黑箱隐忧这个经久不衰的话题送上热门。据报道，实验室的三人团队联手创造了一个叫诺曼（Norman）的人工智能，与希区柯克经典电影《惊魂记》中的变态旅馆老板诺曼·贝兹同名。

名如其人。诺曼会以负面想法来理解它看到的图片。例如，一张在一般AI看来只是“树枝上的一群鸟”的普通图片，在诺曼眼中却是“一名男子触电致死”。

团队希望通过诺曼的表现提醒世人：用来教导或训练机器学习算法的数据，会对AI的行为造成显著影响。AI会成为什么样，有时人类可能束手无策。

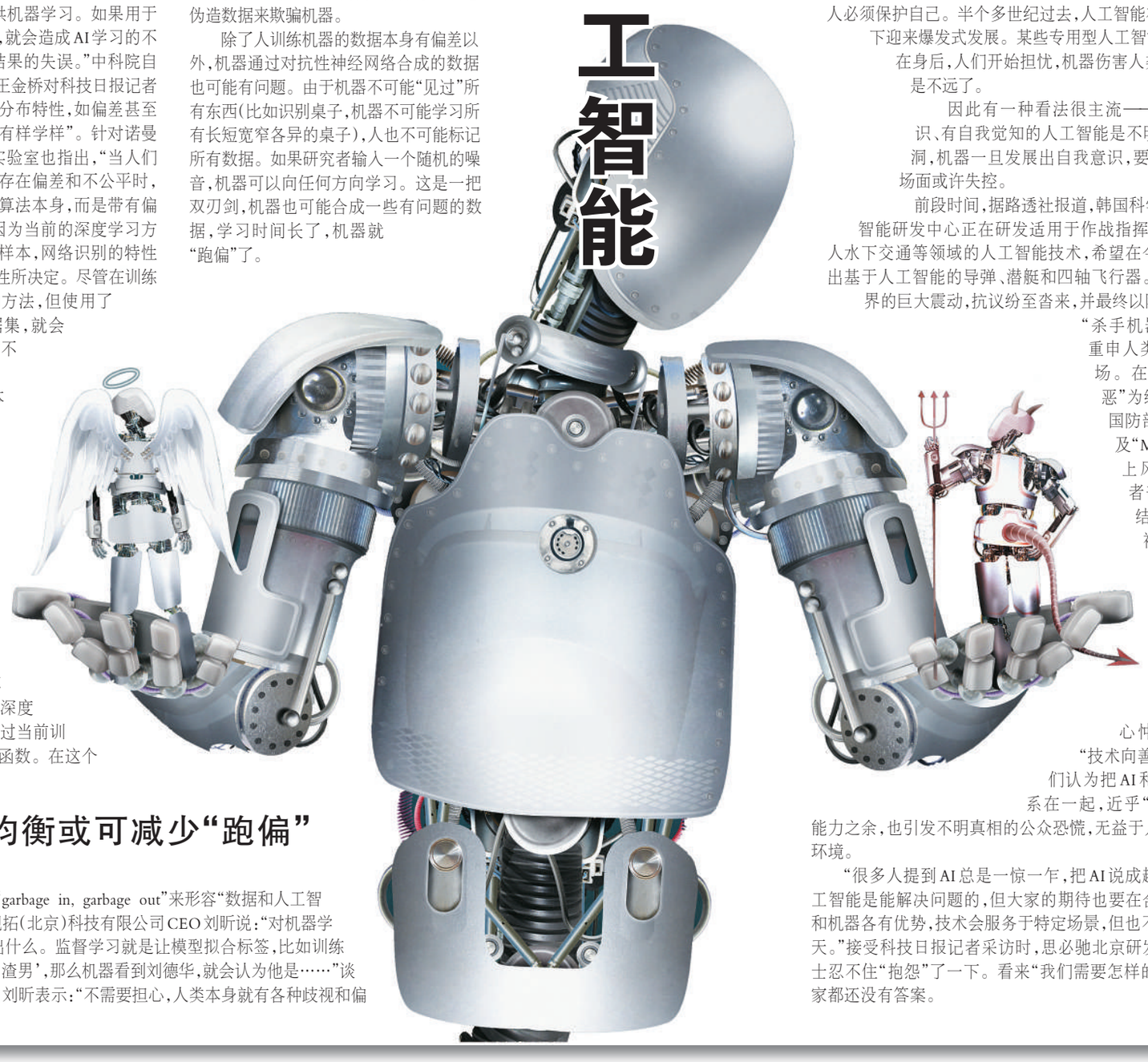
TA们的偏见就是人类的偏见

诺曼们从哪来？答案首先藏在数据里。

“人工智能识别出的结果不是凭空而来，是大量训练的结果。如果要训练AI某一方面的能力，比如下棋，就需要收集、清洗、标记大量数据供机器学习。如果用于训练的数据不够多，就会造成AI学习的不充分，导致其识别结果的失误。”中科院自动化研究所研究员王金桥对科技日报记者表示。数据本身的分布特性，如偏差甚至偏见，也会被机器“有样学样”。针对诺曼的表现，创造它的实验室也指出，“当人们谈论人工智能算法存在偏差和不公平时，罪魁祸首往往不是算法本身，而是带有偏差、偏见的数据。因为当前的深度学习方法依赖大量的训练样本，网络识别的特性是由样本本身的特性所决定。尽管在训练模型时使用同样的方法，但使用了错误或正确的数据集，就会在图像中看到非常不一样的东西”。

另外是算法本身的影响。“这可能是无法完全避免的，由深度学习算法本身的缺陷决定，它存在内在对抗性。”王金桥表示，目前最流行的神经网络不同于人脑的生物计算，模型由数据驱动，和人类的认知不具有一致性。基于深度学习的框架，必须通过当前训练数据拟合到目标函数。在这个框架之下，如果机器要识别狗，它会通过狗的眼睛、鼻子、耳朵等局部特征进行可视化识别，而这些可视化特征却能给想利用深度学习漏洞的人机会，后者可以通过伪造数据来欺骗机器。

除了人训练机器的数据本身有偏差以外，机器通过对抗性神经网络合成的数据也可能有问题。由于机器不可能“见过”所有东西（比如识别桌子，机器不可能学习所有长短宽窄各异的桌子），人也不可能标记所有数据。如果研究者输入一个随机的噪音，机器可以向任何方向学习。这是一把双刃剑，机器也可能合成一些有问题的数据，学习时间长了，机器就“跑偏”了。



数据的均衡或可减少“跑偏”

不少科学家以“garbage in, garbage out”来形容“数据和人工智能的关系”。中科视拓（北京）科技有限公司CEO刘昕说：“对机器学习而言，进什么就出什么。监督学习就是让模型拟合标签，比如训练者把帅哥都标记成‘渣男’，那么机器看到刘德华，就会认为他是……”谈到诺曼引发的讨论，刘昕表示：“不需要担心，人类本身就有各种歧视和偏见，用人类生产的数据训练AI，再批判AI不够正直良善，这么说有点危言耸听。”

偏见、刻板印象、歧视都是人类社会的痼疾，有些流于表面，有些深入社会肌理，无法轻易剥离。在这样的语境中产生的数据，携带着大量复杂、难以界定、泥沙俱下的观点。如果研究者没有意识到或着手处理这一问题，机器学习的偏见几乎无解。真正的“公正算法”或许是不存在的。

据有关媒体报道，在谷歌研究自然语言处理的科学家 Daphne Luong 表示，正确地校准标签对机器学习来说非常关键，有些数据集其实并不平衡，像维基百科上的数据，“他”（He）出现的次数远比“她”（She）要多。

王金桥也着重强调了“数据的均衡”。就算对人来说，不同人秉持着不同的价值观，但多听多看多受教育可以让人向良好的方向改进。机器学习也是如此。“训练机器时，要注重数据的均衡，给它更多可靠的数据。研究团队成员具有多元的学术背景（如吸纳更多社会学、心理学等领域学者加入）、性别、年龄、价值观，也可以帮助机器学习更加均衡，减少机器出现偏见、谬误甚至失控的可能。”王金桥说。

“机器学习的模型本身也要有一定的防攻击能力，从技术上防止本身结构设计的漏洞被攻击，研究者可以使用各种数据攻击机器，训练机器的反攻击能力。”王金桥说。

作恶还是向善，是人类的选择

1942年，阿西莫夫在短篇小说《环舞》中首次提出著名的机器人三定律：机器人不得伤害人类，或因不作为使人类受到伤害；除非违背第一定律，机器人必须服从人类的命令；除非违背第一及第二定律，机器人必须保护自己。半个多世纪过去，人工智能在大数据的加持下迎来爆发式发展。某些专用型人工智能把人类智能甩在身后，人们开始担忧，机器伤害人类的那一天是不是不远了。

因此有一种看法很主流——人类训练有意识、有自我觉知的人工智能是不明智的。开个脑洞，机器一旦发展出自我意识，要反向攻击人类，场面或许失控。

前段时间，据路透社报道，韩国科学技术院的人工智能研发中心正在研发适用于作战指挥、目标追踪和无人水下交通等领域的人工智能技术，希望在今年年底前研发出基于人工智能的导弹、潜艇和四轴飞行器。此事引发学术界的巨大震动，抗议纷至沓来，并最终以院长保证无意于“杀手机器人”的研发并重申人类尊严和伦理收场。在美国，以“不作恶”为纲的谷歌也因与国防部的合作协议涉及“Maven项目”被推上风口浪尖，反对者普遍认为，识别结果完全有可能被用于军事用途，比如说精准打击。谷歌最终表示终结协议。

相较于舆论环境的忧心忡忡，研究者对“技术向善”普遍乐观。他们认为把AI和“杀人机器”联系在一起，近乎“捧杀”，夸大AI能力之余，也引发不明真相的公众恐慌，无益于人工智能的发展环境。

“很多人提到AI总是一惊一乍，把AI说成超人。我相信人工智能是能解决问题的，但大家的期待也要在合理范围内。人和机器各有优势，技术会服务于特定场景，但也不需要把AI捧上天。”接受科技日报记者采访时，思必驰北京研发院院长初敏博士忍不住“抱怨”了一下。看来“我们需要怎样的AI”这一题，大家都还没有答案。

新鲜事

欢迎来到AI唤醒的世界

——智能应用落地场景新观察

实习记者 唐芳

人工智能现在有多牛？近日，在主题为“聚焦人工智能，畅享美好未来”的第二十二届中国国际软件博览会第四场全球软件产业发展高峰论坛上，科大讯飞、京东和阿里巴巴的人工智能技术相关负责人，分享了AI应用落地场景的最新成果。

AI速记员：识别22种方言才稀罕

软博会现场的环形屏幕上，演讲者的发言内容实时显示为字幕。科大讯飞副总裁李世鹏在演讲中表示，这正是科大讯飞的听见系统在做实时速记和翻译。

听见系统是一位既“老实”也“不老实”的AI速记员。记者在现场看到，它会将演讲者口语中出现的“嗯”“啊”“这个”“那个”等语助词如实照写，而不是过滤掉；偶尔会将同音的“攻击”写成“公积”，大致是没“听懂”演讲者的原意。

“科大讯飞输入法不仅可以实时识别普通话和英语，还能识别我国22种方言。”李世鹏介绍，通过该应用，科大讯飞俘获大批不会说普通话和英语的“银发”用户群。

他还表示，人工智能在教育领域有两个新的落地场景。学习英语口语时，学习者可能无法确定自己的发音是否标准，一款自动语言考试打分系统可以帮助解决这个烦恼。此外，学生们的作业和试卷，成为人工智能分析学生知识薄弱点和长处的“密卷”，老师们可以“对症下药”帮学生答疑解惑，进行有效的个性化教育。

AI客服：带有情感的回答好温柔

“AI正在重构零售业的服务体验。”京东AI研究院副院长梅涛说，在未来的购物环境里，消费者可以在任意时间点、任意空间买到他想要的东西。

接着，他播放了几段视频进行演示。比如不用脱衣也能换装试衣魔镜，客户站在“镜子”前在线“试衣试妆”。好不好看？魔镜自带打分系统，通过这种方式，推荐客户适合的服饰、配饰和产品。另一段视频中，顾客走进一家智慧门店，随即“画像”被捕捉。梅涛介绍，店员走来时，已经知道你是不是会员、偏好哪类商品，从而提供精准个性化服务。

京东智能客服已服务上亿客户，“她们”的回复可以说是礼貌、周到和温柔。你很有可能已经与情感智能AI客服接洽过，还私下感叹这位“小姐姐”性格好好。“因为‘她’能感知用户情绪，回答问题时还带着情感和同理心。”梅涛说。

AI+万物：正走入我们的生活

人工智能带来大融合时代，一方面与物联网、混合现实、3D打印、机器人等技术结合；另一方面与城市、农业、工业、工业互联网、汽车制造、智能家居等垂直领域融合，在物理世界和实体经济中体现巨大价值。

人工智能能养猪？阿里巴巴副总裁李松介绍：“通过图像识别技术，可以实现对上百万头猪的百分之百行为监控和识别。”此外，他表示现在的人工智能结合遥感数据、气象数据，还可以大范围地解决农业种植业病虫害的问题。

“从城市大脑到工业大脑，再到我们去年年底跟首都机场合作的航空大脑，也都是产业AI融合的方式。”李松说，比如城市大脑可以做到在火车站或者公共场合辨别可能的犯罪嫌疑人。还可以20秒为单位自动判定高峰期车辆的刮蹭和碰瓷，让巡逻交警5分钟处理冲突。

在他看来，智能化产品将给中国制造业带来极大升级机会。比如汽车是未来十年最重要的物联网和智能化产品，大量应用大数据、智能算法推荐技术与移动互联网场景相联动。如果在“淘票票”上购买一张电影票，上车可以直接导航去电影院，同时推荐附近餐馆。最新一代的智能联网车上追加了人脸识别功能，驾驶座上坐的如果不是车主就会预警。

淘宝的天猫精灵智能音箱已经在北京一些养老社区应用，老人行动不便时“动动嘴”，可以启动房间里的窗帘、电视机，把空调开到恰当的温度。同样，这种应用将来或许可以在酒店或者家中使用，成为“懒人”福音。

智能硬件大家族中也有“清新派”。一款“自动跟随行李箱”让女士在旅行中变得自由自在，它实际上是将智能驾驶识别技术“下移”到行李箱中。“能直播的烤箱”自动以秒计时将烘焙过程晒到朋友圈。

毒瘾评估？用AI翻译“不说谎”的生理信号

好机友

本报记者 张盖伦

近日，在关于国际禁毒日的一篇报道中，出现了“AI戒毒瘾”的说法。但文章着墨不多，只说北京的强制性戒毒所已将人工智能运用于戒毒领域，大数据交互信息为提升戒毒人员的心理健康提供了科技支撑。

AI和戒毒怎么扯上关系了？其实，更准确来说，是用AI来进行毒瘾渴求度评估。为北京提供这项服务的，是成立不到两年的望里科技，他们希望用技术手段探索和学习人类的内心与大脑。人说出口的话可以经过修饰，但是脑电反应等生理信号却很忠实。他们要做的，是破译这些复杂信号构筑成的密码。

毒瘾评估，只是人工智能与心理认知结合的多项可能应用中的一种。

从生理信号中找规律

毒瘾评估，就是要了解成瘾者在想什么。要炼成“读心术”，可以从脑电等生理信号中的特异性标记中找答案。

传统的毒瘾评估主要靠询问。

2015年，江苏禁毒网上刊登了一篇文章，详尽描述了强制隔离诊断评估实践中的主要困难。比如戒毒人员矫治积极性不高，诊断评估体系标准化程度较低，各家评估的内容、方法和标准千差万别，而实际操作中许多测评指标更是难以客观量化。

“戒毒人员未必对你讲真话，你不清楚他究竟处在怎样的状态，怎么进行诊断和矫正？”望里科技CEO李岱介绍，他们采用的方法，是采集戒毒人员那些没法撒谎的数据——脑电、心率和皮电信号，用生物指标辅助或替代主观询问。

具体来说，每名戒毒者的数据采集时间约为15分钟。他们会戴上虚拟现实头盔，进入一个有吸毒诱惑的场景，比如纸醉金迷的KTV，比如有同伴拿着吸食冰毒的器具在旁怂恿。

给烟瘾者一杯水，他可能内心毫无波动；但给烟瘾者一支烟，他大脑特定区域的活动可能会显著增高。你是否渴望，是否享受，程度如何，都可以从生理信号中一窥端倪。

听起来简单，但采集到的数据会非常多。“和心电图一样，你能看到各种各样的波形。”

Peter是望里科技在硅谷的技术人员。他告诉科技日报记者，靠人从这些数据中找规律很难，但机器学习擅长的，恰恰就是从大量数据中提取客观指标，找到规律。

据望里科技介绍，基于生理大数据的人工智能毒瘾渴求评估系统已在全国各省、市强制隔离戒毒所对冰毒戒毒人员进行了大规模的评估施测。目前已有1000多名甲基苯丙胺（冰毒）成瘾者和600多名正常对照组，以此为基础自建的机器学习分类模型已具有较好的预测效果，对戒毒人员毒瘾渴求状态有87%的判断准确率。

系统会为戒毒者的毒瘾渴求度打分。100分代表着极度渴求，这意味着，如果戒毒者暴露在毒品环境中，他有极高概率想吸。

这些分数，都是给强制戒毒所提供的参考性数据。

打开精神科领域的“黑匣子”

要做到准确评估，需要数据，但更需要的是具有信度和效度的数据，于是，它也就需要一个合理、严谨但又不复杂的实验设计。

李岱强调，他们是个跨学科的科研团队。懂心理认知，也懂人工智能，是团队的一大优势。将脑科学、神经科学和人工智能结合在一起，“背后是大量的研发和试错”。

除了做毒瘾渴求度评估，他们也在用类似的方法，做焦虑情绪评估与干预系统，做抑郁辅助诊断系统等。

“精神科领域像是个黑匣子。”李岱坦言。大部分生理疾病都有客观生物指标，但心理疾病则主要依赖主观量表。四川大学华西医院李静教授在之前召开的一次学术会议上，对将人工智能引入精神科领域给予了肯定：“人们常说‘邮差的腿、精神科大夫的嘴’，意思是大夫是靠嘴挣钱的。所以，我们也需要一些客观的生物学指标，这样一项人工智能技术是非常值得期待的。”

Peter说，人工智能进入心理疾病诊断领域是顺理成章，在治疗方面，它也能进一步发挥作用。

对精神科大夫来说，给病人开什么药，常常靠的是经验。不同医生有自己习惯性的药物组合，但同样的药物，放在患有同样疾病的不同人身上，疗效就是不一样。所以医生往往需要试药。患者服用某种药物一段时间之后，可能需要再度去往医院，由医生评估要不要换药，怎么换。但这个评估，同样也靠的是医生问、患者答，类似疾病的诊疗过程缺乏客观的、具有特异性敏感性的公认指标。

李岱说，行业从业者都知道传统诊断和治疗手段的局限性，因此，业内人士也渴望新技术。

Peter表示，用强化学习的方法，是可以让人工智能找到最优治疗方案的。

这个道理和训练机器下围棋、打游戏一样。在围棋中，机器的最终目标是取胜；在治疗方案的选择中，机器的最终目标是改善病人精神状态。经过训练，假以时日，机器能够帮忙，让最优治疗方案“浮出水面”。

“我们无意做替代者，只想给这个传统行业带来一些新东西，用技术手段辅助医生问诊。”李岱说，人工智能是对医生的解放，他们可以把更多精力放在治疗方案的拟定上。

（本版图片来源于视觉中国）

扫一扫
欢迎关注
AI瞭望站
微信公众号