

谁能训练人工智能

本报记者 余颖

最近,人社部公布了一批新职业,其中有个新职业叫作“人工智能训练师”。听起来高大上的人工智能还要人来训练?谁能训练这些看不见、摸不着的人工智能?“人工智能训练师”究竟做些啥?关键是,挣得多不多?

带着这些问题,经济日报记者近日走访了百度(山西)人工智能基础数据产业基地。这里是目前国内从业人员与产值规模最大的单体数据标注基地,入驻企业35家,AI数据标注师从业人员超过2300人。

“我是数据标注员,中专学历,每天工作8小时,大概能挣300多元。”面对记者的问题,山西麟诺网络科技有限公司的李宇龙自豪地给出了答案。这个收入对在太原生活的年轻人来说算高水平了——数据显示,2019年,太原市城镇居民人均可支配收入36362元。不过,李宇龙说自己不算最高的,“我有的同事一天能挣1000多元”。

李宇龙学历不高,原本与人工智能没什么关系。2018年,山西麟诺网络科技有限公司成立,李宇龙经朋友介绍来到这家公司。

“当时我也不知道数据标注员是做啥的,更没听说过人工智能。老板就交给我一套规则,让我按照规则在电脑上认车道线。”那时20岁出头的李宇龙正是肯学的时候,短时间内就熟悉了规则,还总结出标注车道线的新规律。

作为这项新职业的老员工,李宇龙已转型培训师,除了日常工作,还承担起公司新员工的培训教学,“有的同事会转型做项目或者内容审核员,还是有上升空间的”。

“数据标注技术门槛低,招工人群范围广泛,对于促进就业作用显著。”山西麟诺网络科技有限公司负责人李应维告诉记者,目前公司招聘的员工以大专生为主;标注规则由百度与相关领域专家共同制定,并形成可操作软件平台提供给企业,对员工简单培训就能上手。以人脸为例,目前他们能实现约150个特征点的标注。大致标注规则是数据标注师看到双眼皮就标记双眼皮,看到狮子鼻就标记狮子鼻,让机器自己理解与学习。标注特征点越多,AI就越能精确识别人脸。

看似简单的工作,背后是庞大数据库做支持。算法、算力与数据是人工智能发展的三大要素,其中数据是人工智能发展的燃料。每天,人类社会产生的海量数据必须经过清洗与标注,换成人工智能熟悉的语言才有价值。因此,人工智能训练师主要任务就是数据采集和标注,特别是数据标注。如果说人工智能是个孩子,数据标注师就是带领这个孩子认识世界的启蒙老师。有了足够多、足够好的数据,AI才能学会像人一样去感知、思考和决策,更好地为人类服务。

人工智能产业发展之初,数据采集与标注需要人力完成的部分比较多,现在数据平台已有了持续学习能力,大部分采集与标注工作人工智能自己就可完成,实现预标注与自动标注。因此,目前数据标注工作以机器辅助与人工标注结合为主。除了眼底检查片、肺部X光片等技术含量较高有专业要求,大多数员工均能按照规则,在平台上完成语音、人脸、道路的标注操作,只要能熟练操作电脑即可。

郭梅是新职业的受益者。34岁的她原本在山西一家煤矿做监控员,为照顾到太原上学的孩子,她到学校附近的数据标注基地求职。经过公司岗前培训与团队帮助,她从起初每天标注两三百张图,提升到每天能完成1300多张。

数据标注员的工作随着人工智能发展需求不断调整。前几年人工智能发展伊始,语音识别、人脸、图像等识别需求大;近几年无人驾驶驶入发展快车道,郭梅与李宇龙的主要工作就改成无人车、地图、3D点云等数据标注,天天看的是车道线、行人、障碍物。

“还有一些语音识别项目,比如我们接了一些方言语音数据标注项目,这就需要招聘当地人完成了。”李应维说。

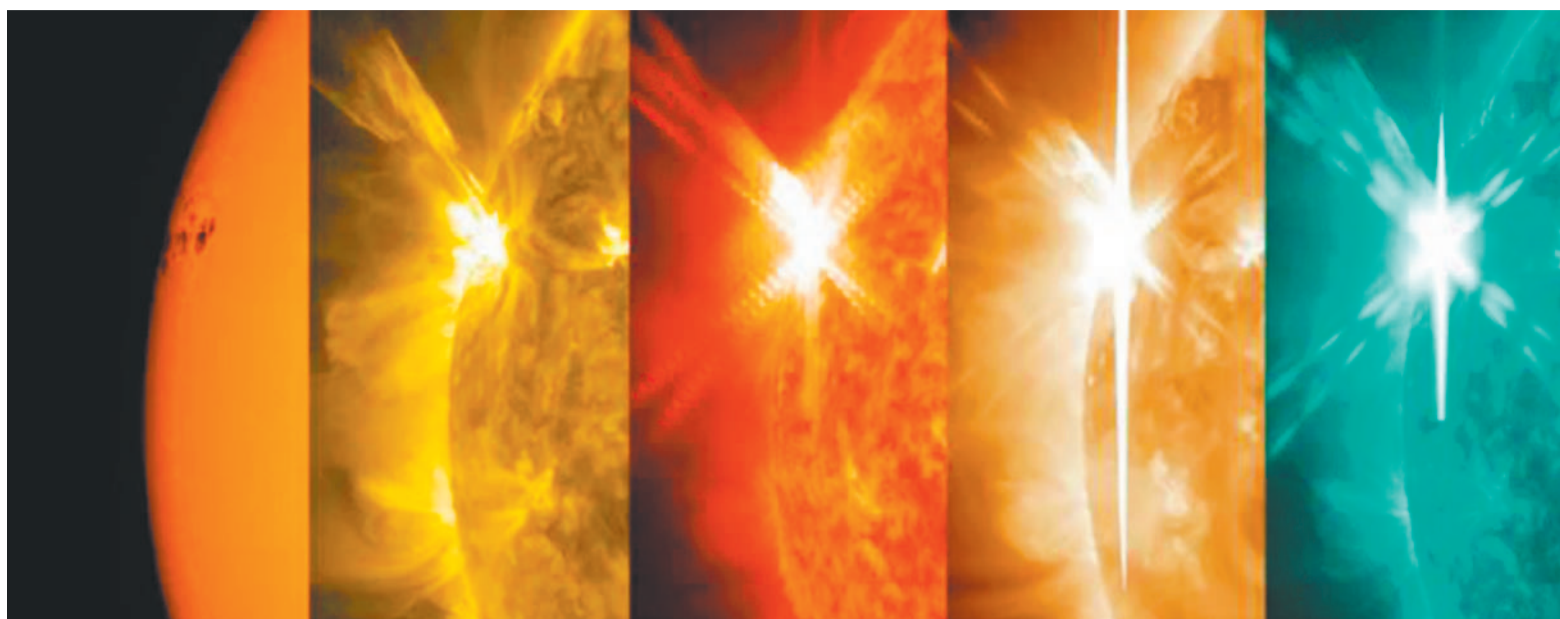
有了足够技术基础,数据标注员还能教会人工智能新任务。疫情期间,戴口罩的人脸识别需求激增。“根据客户需求,山西数据标注基地采集了大量戴口罩人脸照片,由数据标注师对眉毛、眼镜、颧骨等人脸关键点精准标注,训练人工智能完成了戴口罩人脸图像识别。”百度(山西)人工智能基础数据产业基地负责人尉亦说:“人们能在不戴口罩情况下实现精确体温测量,或是通过人脸闸机,背后就是这些数据标注员在训练人工智能。”

艾瑞咨询发布的《中国人工智能基础数据服务行业白皮书》预测,随着全球人工智能产业、物联网、5G的爆发式增长,到2025年,全球每年产生的数据量将从2016年的16.12B猛增至2025年的1632B,其中80%至90%是非结构化数据,需要经过清洗与标注才能被唤醒价值。在我国,每年需要标注的语音数据超过200万小时,图片则有数亿张。因此,人工智能数据标注师的岗位需求将持续增长。尉亦透露,今年基地新招了500多人,未来5年计划培养5万名AI数据标注师,并引入更多AI合作伙伴。李应维也预计,到明年自家企业用工将翻一番,从160多人扩展到300人左右。未来,数据标注将成为创造大量就业需求的新兴职业领域。



百度(山西)人工智能基础数据产业基地的AI数据标注师们正在工作。余颖摄

又一个火星发射窗口来临,阿联酋、中国与美国的3艘火星探测器将相继启程,去往这个我们探测了60年的红色星球。短期登陆火星何时能够照进现实,尚无明确时间表,但作为太阳系中与地球环境最接近的行星,在火星建立新的栖息地已是不少人的憧憬与幻想——移民火星,你想去吗?



从不同波段观察到的一次耀斑爆发现象。(NANS)

移民火星,这一出现于科幻作品中的桥段,如今距离照进现实似乎越来越接近。然而,如果人类真要在火星生存,这些新移民的手机天气APP中需要提供两类天气信息:一类与地球日常天气一样,是火星大气变化造成的风暴、沙尘等现象;另一类则是与宇宙线辐射相关的空间天气信息。无论是乘坐飞船往返火星与地球之间,还是在火星表面生活,辐射问题都不可忽视。但与人们日常对辐射诸多不合理的担忧与恐惧不同,辐射是登陆火星的确切威胁,必须直视。

来自太阳的威胁

1972年,美国阿波罗登月计划接近尾声,人们已经习惯了宇航员们乘坐飞船来往穿梭于地球与月球之间,似乎只是出了一趟远门。

然而,若干年后,当空间气象学家们对太阳活动有了更深入认识,他们惊恐地发现,原来阿波罗16号宇航员曾与足以夺取生命的灾难擦肩而过——1972年8月2日,就在阿波罗16号返回地球后几个月,阿波罗17号准备发射时,太阳爆发了一次巨大耀斑。据科学家估计,耀斑爆发时,如果宇航员刚好在月球表面活动,将会在短时间内受到4西弗剂量的辐射,超过一般认为的致死剂量3西弗。西弗的辐射剂量有多大?大概相当于人体经受4万次胸透后所累积的辐射。

这一辐射的元凶太阳耀斑,是太阳表面的一种爆发现象,能导致磁场中储存的能量被快速释放。一次大耀斑释放的能量,与成百上千颗氢弹同时爆炸释放的能量相当,可以将太阳日冕中的粒子加速到接近光速向太空奔去。仅需几十分钟,这些高能粒子就能到达地球与火星附近。也就是说,当地球附近探测器发现耀斑爆发并给出警告后,留给宇航员做出反应并采取防护措施的时间很短。

太阳耀斑是一种偶发性爆发事件。当太阳表面的太阳黑子数量较多时,爆发耀斑的可能性更大。大型黑子群是孕育强耀斑的土壤,这些黑子群一般有着复杂日冕磁场结构,在太阳内部对流作用下,磁场不断剪切、扭缠,如同越绷越紧的发条不断积蓄能量,并在瞬间释放。消亡之前,这些黑子群会反复将日冕磁场的“发条”拧紧,产生多个耀斑爆发。

大耀斑爆发的同时,一般会伴随着另一种太阳爆发——日冕物质抛射。如果将耀斑看作雷电,那么日冕物质抛射就是一阵暴风雨。从太阳出发,日冕物质抛射最快10多个小时即可到达地球,20多个小时就能到达火星。日冕物质抛射前的激波也会产生高能粒子流,有时还比耀斑更剧烈。不过有个好消息,一旦我们监测到太阳附近日冕物质抛射,宇航员的避险时间比耀斑爆发要长一些。

虽然某次特定太阳耀斑与日冕物质抛射威力十分惊人,但是这些太阳爆发现象并不是持续出现的。太阳爆发会出现一定间歇期,在太阳活动极小年份,这种间歇期还可能比较长。

你以为航天员可以松了一口气了?并没有,因为宇宙中的辐射威胁可不止来自于太阳。

防不胜防的银河宇宙线

在更加遥远的太空中,诸如超新星爆发等现象更为剧烈。这些现象产生的高能粒子流可以侵入太阳系之内,到达地球与火星附近,威胁航天员安全。与间歇发生的太阳耀斑不同,由于银河宇宙线来源遍布全宇宙,它几乎持续存在。源自太阳的行星际磁场对银河宇宙线具有一定屏蔽作用,在黑子比较多的太阳活动高年,行星际磁场的磁通量有所增加,银河宇宙线强度会变得较低。然而,当太阳活动降低时,银河宇宙线强度会随之增强。此外,银河宇宙线中存在一些原子量比较大的重元素高能粒子,甚至能穿透目前载人飞船保护层,或轰击宇宙飞船外壳材料产生二次辐射,更加防不胜防。

也就是说,源自太阳耀斑的太阳宇宙线与源于遥远天体的银河宇宙线此消彼长,共同构成极寒与接近真空以外,太空中又一个恶劣环境来源。虽然,一次产生致命辐射剂量的太阳爆发相当罕见,但在银河宇宙线与规模相对小的太阳爆发共同作用下,长时间执行深空任务的宇航员可能会累积过量辐射。

幸运的是,生活在地球上的人们无需对这些来自太空中的高能粒子担心。这是因为大自然给予了地球磁场和浓厚的大气层,成为抵御宇宙线的天然屏障。一旦我们远离地面,宇宙线辐射强度会随之升高。例如,民航飞机飞行在万米高空失去一部分大气的保护,乘客与机组人员所受辐射剂量就比地面要高。在低地球轨道中生活的宇航员,则会承受比飞机乘客更大的辐射风险。至于飞往月球的宇航员,风险将更大。

一旦去往火星,必须脱离地球磁层保护,进入深空飞行。此时,银河宇宙线与太阳宇宙线辐射均将更强,就像开着敞篷跑车去火星,缺乏相应防护,恐怕凶多吉少。

还有很长的路要走

2012年发射的好奇号火星车上,搭载了一台名为“辐射评估探测器”(RAD)的仪器,用来测量在飞往火星途中与降落在火星表面后,好奇号火星车所接收的辐射剂量。在飞往火星与在火星着陆过程中,好奇号实际是“火星科学实验室”号飞船的乘客,而这艘飞船对宇宙线的防护水平与美

国国家航空航天局(NASA)未来执行深空探测任务的载人飞船基本相当,可以估计未来宇航员可能受到的辐射。

RAD探测结果显示,在来往火星共约360天的路程中,一位宇航员可能受到0.662西弗辐射,相当于NASA对宇航员在职业生涯中总辐射限制的62%。此外,就算到了火星,宇宙线问题仍继续伴随着人们。火星没有较强的内禀磁场,无法像地球那样拥有自己的磁层。在太阳风吹拂下,火星大气逐渐被侵蚀,变得稀薄,密度仅相当于地球大气密度的1%左右。因此,火星大气无法为人类提供足够保护。

根据“火星奥德赛”号的探测数据,火星表面平均辐射水平约为国际空间站内宇航员所受辐射水平的2.5倍,是地球表面日常辐射水平近13倍。这样的辐射水平短时间内不会对人造成不可逆伤害,但长时间累积会带来癌症等各种健康风险。美国内华达大学研究者在2017年发表论文显示,通过更优化的伤害模型评估,火星辐射带来的健康风险比此前预计要高一倍。在火星生活,更容易罹患癌症、中枢神经系统疾病、白内障、循环系统疾病与急性辐射综合征。为了健康生活,大部分时间内,火星移民们可能不得不待在地下庇护所中。

目前,我们对飞往火星与火星表面辐射状况的认识仍比较有限,新数据在不断刷新人类认知。2017年,科学家发现MAVEN探测器在一次强烈太阳爆发事件后测量到的辐射水平,比好奇号火星车之前所测到的最高水平要高一倍。太阳活动存在11年周期,每个太阳活动周强度又有所不同。距离我们最近的23与24太阳周强度均比之前几周强度低不少,而在强度更高太阳活动周中出现的几十年一遇,甚至百年或千年一遇的超强太阳风暴,火星移民们要怎样面对?人类尚无法回答。

实际上,辐射问题仅仅是移民火星之路上要解决的诸多问题之一。在火星安营扎寨后,火星移民们如何自给自足获取水、食物与能源,扩大火星建筑的材料能否从火星表面提炼,火星移民的经济与社会秩序如何形成,是否存在与地球建立贸易往来的可能?这些问题需要科学家们进一步探索,需要新的探测数据一点点解答。

目前,对于火星辐射问题,空间气象学家们能做的是,进一步认识太阳活动规律,发展出更高效、更准确的预报模型,从而更加准确估计宇宙线强度,在太阳爆发事件发生之前及时发出预警。航天工程师们则致力于研制出更轻巧、屏蔽效能更高的防护材料,给飞船与火星建筑提供更强防护性能。相信当火星移民真正实现的那一天,去往火星的旅程一定舒适而安全。

【作者单位:哈尔滨工业大学(深圳),本文授权转自中国科普博览,微信公众号:kepobolan,略有删改。】

想移民去火星?留心宇宙辐射

李会超



在飞往火星的过程中,火星车是“火星科学实验室”飞船的乘客。降落到火星后才会开始独自执行任务。(NANS)

循天神器 破题气象灾害风险管理

本报记者 郭静原

气象灾害风险管理室,实地探究一款破题神器——气象灾害风险管理系统。通过系统界面,点击彼时南方部分汛情严峻省份,相关灾情与灾害事件信息一目了然。

“气象灾害风险管理系统集重大灾害事件监测、影响评估、灾害风险预估等功能于一体,可时刻监视正在或将要发生的气象灾害事件。”国家气候中心气象灾害风险管理室主任王国复介绍,该系统支撑气象防灾减灾第一道防线的作用日益显现。

今年6月以来,我国江南、长江中下游与江淮地区先后入梅,天气气候形势复杂,灾害事件频发,南方地区出现大范围持续性强降雨,累计雨量大、极端性强,灾害影响大。“通过该系统,我们可以实时跟踪监测南方地区灾害发展情况,滚动发布针对灾害事件的影响评估产品与风险预估产品,为防灾减灾持续提供支撑服务。”王国复说。

由此,一份题为“2020年6月以来南方暴雨事件风险评估”的《灾害风险评估快报》应运而生。国家气候中心基于气象灾害风险管理系统提供的客观化、定量化灾害评估产品,得出南方暴雨灾害的致灾强度、影响范围

与持续时间等翔实数据与结论,从纵向和横向不同角度权威解读,对于提高公众对暴雨灾害事件的科学认知发挥了重要作用。

该系统另一个重要应用领域是风险预估。在国家气候中心发布的多份《重要气候信息》产品中,不但揭示了未来气候趋势预测,还对有可能发生的气象灾害及其影响作出风险预估,产品从单纯提供灾害性天气预警扩展到灾害性天气预警与灾害风险预估并举等多样化应用服务。

该系统“捕获”的灾害信息种类丰富,可提供暴雨、台风、干旱、高温、低温等主要气象灾害的实时滚动监测识别、影响评估与风险预估等服务。王国复告诉记者,系统集纳了1951年以来的2800多起区域性暴雨事件与近500起台风事件,以及历年高温、低温与干旱事件的详细致灾危险性因子、灾情信息和影响评估数据与产品。这为决策服务与专题服务提供了最全面、最权威信息,使得汛期服务对重大灾害的评估更加客观。

气象灾害风险管理系统是随着气象灾害风险管理业务不断发展而逐步建立起来的。2016年起,王国复就带领团队开发该系统,

时隔4年,经过不断改进,该系统已在国家级业务平台运行,科技成果转化效益显著。对于系统定位,王国复明确表示:“系统要解决的是灾害风险产品制作、业务集成、科技转化以及业务体系和学科发展的支撑等问题。”

要解决这些问题,需要具备强大功能与新技术不断注入。目前,该系统功能已覆盖气象灾害风险管理六大类业务,实现从实时灾害监测识别、影响评估,到风险预估与精细化区划等在线分析与产品制作功能。研发团队还建设了大数据应用、模式算法等4个中心系统,进一步增强系统功能。

“系统已在部分试点省份得到应用,2021年将在全国推广使用。”研发团队负责人李莹说,基于国家重大建设工程,团队正在研究系统的省级版本,推动省级气象灾害风险业务发展,从而实现国省两级协同发展。

循天之迹,鉴灾之事。随着防灾减灾从传统的灾后救助向关口前移与注重灾前预防转变,从应对单一灾种向应对综合减灾转变,从减少灾害损失向减轻灾害风险转变,气象灾害风险管理系统还将在防灾减灾中发挥越来越重要的作用。



国家气候中心工作人员正在操作气象灾害风险管理系统。郭静原摄

“受8月11日至13日强降雨影响,四川盆地西北部、甘肃南部、河北北部、北京等局地发生山洪、地质灾害、中小河流洪水气象风险高……”细心的公众会发现,现在的天气预报里,不仅有熟悉的天气预测内容,还增加了对气象灾害风险的预警。

当极端天气事件来袭,我们该如何监测识别与客观评估灾害风险发生的可能?带着这个问题,经济日报记者走进国家气候中心