2022年11月29日 星期二
责编：王维萍 校对：黄凤 电话：0791-86821890
经济晚报

一个数据产品的交易历程

北数所所在的国际大数据交易产业园区外景

数据，作为一种新型生产要素，已快速融入生产、分配、流通、消费等各个环节。

数据交易是构建数据要素市场的关键一环。今年2月至7月，通过北京国际大数据交易所（以下简称“北数所”），北京海天瑞声科技股份有限公司和禾多科技（北京）有限公司完成了一笔人工智能算法训练数据产品交易。一个数据产品从采集、处理到交易、应用的过程是怎样的？如何探索建立合规高效的数据要素流通和交易制度？记者近日追踪了这次数据产品交易的全程，一探究竟。


在北京经济技术开发区，一辆“主驾无人、副驾配备安全员”的无人驾驶车正在行驶中

数据采集——
数据越真越全越精，越能提升人工智能“聪明”程度

打左转向灯起步、遇到过路行人减速绕行……在北京市石景山区的首钢园自动驾驶服务示范区内，一辆辆自动驾驶汽车可以精准识别路况，做到安全起步、行驶、落客。

“只要在手机应用程序上下单，车辆就自动开到眼前来；点击小程序的‘开始行程’按钮，车辆就启动了。”北京市朝阳区居民王女士对自动驾驶技术既赞叹不已，也十分好奇，“这左拐右拐、上坡下坡的，它是怎么判断的呢？”

“自动驾驶的实现，是基于人工智能技术、先进传感器、高精地图等进行的技术‘大综合’。人工智能技术就相当于自动驾驶系统的‘大脑’。”海天瑞声是一家人工智能数据资源和服务提供商，公司副总经理李科告诉记者，为了使这个“大脑”更“聪明”，就需要运用各类数据来训练人工智能算法。“人工智能算法做出判断大致要经历‘接收数据’‘总结规律’‘形成判断’3个环节，数据样本类型越全、精度越高、针对性越强，算法就会越聪明，自动驾驶系统的智能化水平就会越高。”

这次数据产品交易中，自动驾驶解决方案提供商禾多科技公司需要自己采集真实场景的原始数据，这些数据由海天瑞声进行专业处理后，形成人工智能算法训练数据，用于自动驾驶系统研发。

如何保证数据“原材料”的高质量？有效采集至关重要。

“数据采集要尽可能接近真实路况。”禾多科技副总裁戴震介绍，在近期的一次数据采集中，工程师驾车从北京市顺义区出发，途经望京区域、机场高速和4个停车场，行驶路程100多公里，现场采集到了道路状态、交通信号和标识、车辆和行人目标以及天气环境等信息。

“多位专业工程师驾驶数据采集车，车上安装了雷达、摄像头和传感器用以收集数据。采集到的数据经过合规处理，会被记录在车载硬盘内，之后通过网络闭环上传至数据处理系统，为下一步的筛选、标注做好准备。”戴震说。

据介绍，海天瑞声与禾多科技今年完成交易的数据产品，其中许多涉及停车场景。“为人工智能算法提供的训练数据，针对性越强，越有助于提升其在特定方面的智能化水平。”戴震说，有时根据客户的需求，为了提升场景的针对性，团队还会专门设置一些具体的情境。

数据处理——
由专业团队协作完成，创造规模可观的就业岗位

采集原始数据只是第一步，接下来需要技术人员对数据进行处理，让人工智能算法可以“读懂”这些数据。

处理数据的办法主要是进行数据标注。“虽然我们可以在原始视频上看出哪里是车道线、哪里是停车位，但如果不加以标注，人工智能算法是无法读懂这些数据的。”李科说，数据标注的基本原理是将原始视频数据分为若干帧，由技术人员运用公司自研的智能化数据处理平台及相关标注工具在每一帧上标注出相应内容，“例如，标出汽车的位置在哪里，某个交通标志是什么意思，等等。”

在海天瑞声公司总部，计算机视觉业务部高级项目经理秦子雄向记者现场演示了数据标注的步骤：

“我们使用这个矩形框将这辆汽车框起来，算法后期就会读‘明白’。”

如何精确定位这辆汽车？

“那就要使用接地线这个辅助工具，先确定几个汽车轮廓上的关键点，再画出数条接地线垂直于地面，这样就可以确定汽车轮廓投影在地面上的具体位置。”

……

几番操作下来，经过各种线和框“勾勾画画”，一帧视频图像标注完成。

数据标注不是一项轻松的工作，需要专业的技术团队协作完成。“为了顺利完成这次与禾多科技的交易，我带领100多人的数据标注服务团队工作了近5个月，标注完成了十几万帧的原始视频数据。”秦子雄说，在这个过程中，需要通过培训帮助团队人员熟练掌握规范，还要依靠公司平台管理团队、追踪工作进度、交付最终成果，“数据标注是一个既有技术含量，也需要较多人力投入的工作，下一步公司将继续加大数据处理平台的研发力度，提升数据标注的智能化水平。”

从宏观层面上看，人工智能产业的快速发展催生了对数据标注服务的庞大需求。《2022人工智能基础数据服务产业发展白皮书》显示，2022年，我国人工智能基础数据服务市场规模将达47.8亿元，预计2025年这一数字将突破120亿元。目前，许多数据服务企业在中西部地区建立了数据标注基地，为当地创造出可观的高质量就业岗位。

数据交易——
建立数据流通信任机制，实现数据“上市有审核、采买有资质”

海天瑞声与禾多科技能顺利完成这次数据产品交易，离不开北数所的撮合与服务。

“在去年3月底北数所成立之初，我们就受邀加入了其牵头成立的北京国际数据交易联盟，并在去年9月至10月上线了几款数据产品。”李科说，数据交易所在国内还属于新生事物，海天瑞声作为首批“尝鲜”的企业之一，在与北数所的交流合作中，也在不断更新对数据交易模式的认知。

“过去，我们寻找客户主要靠广告推广、参与展会等方式，得一个客户一个客户地谈，属于‘点对点’的模式。”李科说，近一年多来，随着买家在北数所数据交易平台上相继出现，企业有条件从“点对点”过渡到“点对面”模式，依靠交易平台提供的撮合服务来获取客户。

北数所相关负责人郎佩佩介绍，这两家企业都是北数所的合作伙伴。了解到海天瑞声在数据领域的综合实力后，禾多科技决定与其开展合作。相关数据处理产品于今年2月至7月分两期交付完成，合同在北数所进行了备案。

除了撮合供需双方外，北数所还要对数据交易主体、数据来源、交易产品、数据用途等进行合规审核。郎佩佩说：“北数所要研判这些人工智能训练数据的来源是否合规，数据产品交付后的用途是否正当等。”

目前，北数所构建了由数据提供方、购买方、中介服务方和交易场所组成的北京国际数据交易联盟，合力打造数据要素市场体系。统计显示，北京国际数据交易联盟已吸纳大型商业银行、电信运营商、互联网企业、跨国机构等150多家机构或企业。“只有实现确权、流通和交易后，数据资源才会转变成可以量化的数字资产。”北京金控集团党委书记、董事长、北数所董事长范文仲表示，数据交易所要做的不仅是撮合交易，更应该建立一套技术、规则、机制、流程健全的数据流通信任机制，实现“上市有审核、采买有资质”的数据交易良性生态。

数据应用——
训练人工智能算法，赋能实体经济、提升用户体验

在地下车库，上海市长宁区居民沈先生体验了一把爱车的“记忆泊车”功能。

“开启‘记忆泊车’功能后，我驾车从地下车库的入口出发，先完整地进行了一遍泊车入库。这时车辆的自动驾驶系统已经‘记住’了泊车路线。待再次出发时，车辆便由系统自动操控，按照设定的路线从车库入口驶入车位。”沈先生说。

“记忆泊车”“跨层泊车”等高阶自动驾驶功能的实现，是人工智能算法通过训练不断“进化”的结果。“经过几个月的迭代升级，我们的人工智能算法在泊车等场景上的智能化水平有了较大提高。”戴震说，目前企业研发的自动泊车系统已经在广汽集团的量产车上得到应用，将为消费者带来更好的出行体验。

将采集到的原始数据进行筛选、标注，把处理完成的数据用于训练人工智能算法，最终赋能实体经济、提升用户体验。业内人士表示，数据流通的这一过程折射出近年来我国数字经济的蓬勃发展态势，也将促进各行业更好地应用数据要素。

“当前，我国数字经济发展成效显著，但适应数字经济发展的规则制度体系仍有待健全。”浙江大学国际联合商学院数字经济与金融创新研究中心联席主任盘和林说，下一步，应加快出台数据要素基础制度及配套政策，推进公共数据、企业数据、个人数据分类分级确权授权使用，构建数据产权、流通交易、收益分配、安全治理制度规则，统筹推进全国数据要素市场化配置改革。

培育数据要素市场逐步取得了成效。“有了这次成功交易，我们和海天瑞声将继续深化合作，未来双方有望达成更大量级的合作。”戴震说。

（据《人民日报》）