

科创前线

数据“采矿人”的兴起和物理AI“淘金”的降本博弈

中经记者 杨让晨 上海报道

刚到上海一家数据采集工厂实习的王平(化名)是一名具身智能机器人数据采集员(以下简称“数采员”),工作内容是戴着数据采集设备为人形机器人采集数据。

和王平一样,张琪(化名)也在这家数据工厂实习,但她的工作是数据审核,她告诉《中国经营报》记者,尽管刚实习不久,但每天看着各种数据,“感觉很新奇”。

除实习生外,不少“宝妈”和退休人员也在通过数据采集赚钱。

觅蜂科技董事长、CEO姚卯青在日前举行的该公司第2万套MEgo系列设备下线仪式上表示,目前公司的数据采集已经不局限于数据工厂,而是推广到了社区。“不少‘宝妈’、退休人员在为公司生产数据。”

这背后是具身智能行业对数据的渴求,并且对数据的质量要求不断提高。姚卯青告诉记者,进入今年下半年,行业对数据的需求出现收敛,60FPS、1080P、双目、深度的需求越来越成为一种共识。

王平、张琪以及“宝妈”和退休人员不知道的是,他们所干的工作,最近被国家认定了新的职业和工种。

9月9日,人力资源和社会保障部发布第八批新职业,包括11个新职业和23个新工种,新工种中新增“具身智能机器人应用技术员”职业,下设“具身智能机器人数据采集员”“具身智能机器人培训师”2个工种。

“打工人的新选择

实际上,数据采集已经不只为数采工厂内进行。有不少“宝妈”、退休人员都进入了数据采集行业,一边赚钱一边提供数据。

“数据采集的流程是先接任务,之后准备采集需要的物料,随后开始采集。”王平告诉记者,自己工作的一天中大概有11个小时是在数据采集工厂度过的。每天采集完成后会上报采集量,采集的动作与录制场景随任务的要求而变化,并且采集时的动作不得超出摄像头录制区域,没有绝对意义上重复性高的动作,“基本都要做”。

从开始实习到现在,王平在数据工厂内的各种采集场景都工作过。“除超市搬运外,还有奶茶店、打印店等场景。”王平进一步告诉记者,自己虽然每天在数据采集工厂待的时间长,但是数据采集的效率并不高。

“现在的实习工作是计时的,每天工作的任务由组长下发,每天的目标采集时长是5个小时。”王平告诉记者,尽管目标采集时长看似只有5个小时,但是采集员在采集过程中会感到疲劳,连续采集半小时后就需要休息一段时间,同时在

数据采集过程中还会遇到网络连接失败的情况。

和王平不同,张琪的实习工作看上去显得稍微轻松一些。“工作的时间是每天早上九点到晚上七点,工作的内容是审核采集来的数据是否可用。这些数据除了数据工厂采集的以外,还有很多来自家庭、工厂等真实场景,由数采员通过佩戴无本体设备采集而来。”

除新奇和有趣外,不错的薪资水平也是吸引像王平、张琪这样的学生选择数采员实习的原因之一。王平告诉记者,自己每月的实习工资在4000~5000元。这比当下许多传统快消企业的实习月薪高出不少。

实际上,作为一个新职业,数据采集在一定程度上带动了就业。上海市的一家数据工厂就告诉记者,公司除了实习生以外,也有不少正式员工。

记者注意到,目前招聘平台

推倒数据高墙

随着数据采集设备在硬件标定和无线传输上的自动化程度提升,即使是非专业人员佩戴,也能稳定输出符合行业标准的高质量数据。

一线数采员的日常与薪资变化,正是整个具身智能数据市场爆发式增长的缩影。

据亿欧智库发布的《中国具身智能数据采集与数据产业发展展望》,在2024年,全球具身智能数据集市场规模大约为7.37亿美元,预计2031年达到70.14亿美元;到2031年,中国具身智能数据集市场占全球市场的比重将达到约50%。

实际上,尽管潜力十足,但目前对于数据采集而言,面临着成本、标准化、效率、复杂度、异构等核心挑战。

谢思远告诉记者,和大语言模型可以低成本获取海量互联网文本不同,具身智能需要的是真实物理世界交互数据。这类数据无法从公开网络直接抓取,采集要动用硬件、人力、真机环境,获取成本极高。

据了解,数据采集面临着高昂

的成本。以仿真数据为例,具身智能行业的仿真数据虽然边际生成成本低,但存在较高的初始搭建门槛与前期投入成本。姚卯青告诉记者,仿真数据需要购买数字资产,通过手绘、建模以及运用GPU算力渲染。

记者注意到,使用不同的采集方式,也会对数据采集的成本带来影响。当前行业内主要有本体真机遥操作、无本体EGO第一人称数据、便携UMI以及仿真数据四种采集方式。其中,有本体真机遥操作采集数据质量最高,但是成本最大,效率也较低;而无本体EGO数据采集依靠头戴式或颈挂式设备,成本较低,规模化潜力最大。

在谢思远看来,当前行业内数据采集从真机采集到无本体采集的变化,对采集成本而言只是“两害取其轻”,“事实上,数据采集最终仍然离不开真实真机数据



上海市一家数据采集工厂中,一位数据采集员正在做数据采集。

杨让晨/摄影

上,不少具身智能企业正在招聘数据采集相关正式岗位。以一家企业的具身智能数据采集岗为例,该岗位需要负责EGO centric(第一人称)数据采集和遥控操作,确保高质量数据上传与设备管理等职责,薪资在8000元至1.2万元,同时缴纳五险一金以及保证周末双休。

不过,沂景资本董事总经理谢

思远告诉记者,如果无本体采集依然局限在数据工厂的流水线模式中,数采员对采集过程中枯燥重复动作产生的疲劳会影响数据采集的质量和成本。从当前来看,数采企业并没有很好的解决办法,并且多次重复数采,对于具身泛化本身也没有太大的科学意义。

实际上,数据采集已经不只是

BPO公司,而是真正发动各行各业的人帮公司采集。”姚卯青告诉记者,截至目前,该公司已借助Mego系列产品实现了百万小时有效数据采集。

记者了解到,目前部分经过去重清洗的高质量具身智能数据,已逐步实现上架交易。不过,数据采集后如何变现,存在不同的路径。

“由于数据并非有形的物理实体,商业模式存在特殊性。”姚卯青告诉记者,目前数据的商业化有两种模式,一是最终使用数据方出售使用权,但产权属于数据采集方,同时使用方也不能对数据进行二次商业化;二是数据产权归属买断的用户,可以使用、再次出售都是可以自由支配的。对应的数据提供商要把数据从服务器上移除。“如何确保在过程中数据产权版权的正确使用、转移,这确实相比实物来讲更难一些。”姚

卯青说。

谢思远也告诉记者,数据能够资产化,并且对于后训练有价值。“经过清洗加工后的数据集,对模型后训练具备明确价值。隐私问题可能比较好解决,机器自动打码和隐私计算,这些在以无人驾驶为代表的领域已经足够成熟,他们可以为具身智能提供可用的技术工具箱。”不过,谢思远也强调,技术不等于彻底解决全部风险。

就整个数据行业的商业化而言,业内人士认为,数据产业可能比模型更快迎来产业化。“在未来两年,数据行业会变得越来越成熟、越来越规范,在技术侧随着众多技术方案的出现,以及中国有完备的从数据到模型的生态闭环,能在实际使用过程中得到越来越多的反馈,逐步收敛和沉淀出一套非常标准化的数据从格式到质量标准的定义。”姚卯青表示。

三大封测厂上半年利润增速与毛利率分化 先进封装卡位影响未来格局

中经记者 李玉洋 上海报道

在摩尔定律趋近物理极限叠加国内先进晶圆制程外部约束的背景下,封装从传统后道“保护+引脚引出”的配套工序,升级为“靠异质集成、芯粒并接延续芯片性能爬坡”的核心路径。

国内龙头企业业绩普增

今年上半年,长电科技收入按市场应用领域划分来看,通讯电子占比27.4%、消费电子占比23.2%、运算电子占比30.1%、汽车电子占比11.1%、工业及医疗电子占比8.2%。其中,运算电子营收同比增长40.4%,汽车电子营收同比增长25.0%。

而长电科技在今年7月披露的一则调研纪要中也提到,公司已在去年先于行业推动业务结构调整及针对传统通讯业务的取舍,通讯及消费电子收入占比已较历史周期显著收敛,且内部结构主要面向当前需求韧性较强的高端市场。

在半年报中,公司表示受益于人工智能相关的基础设施及端侧领域需求快速增长,国产半导体替代趋势明显,同时,也受益于客户订单持续增长、产能利用率高位运行等因素。

另一封测大厂通富微电2026年上半年实现营收160.41亿元,同比增长23.03%;归属于上市公司股东的净利润17.17亿元,同比增长316.77%;扣除非经常性损益后的净利润7.47亿元,同比增长77.78%。其中,2026年第二季度,公司实现

日前,中国大陆地区两大半导体封测厂长电科技(600584.SH)和通富微电(002156.SZ),以及中国大陆地区唯一实现2.5D异质集成大规模量产的盛合晶微(688820.SH)均发布了2026年半年度业绩报告,整体受益于AI算力与存储市场景气上行,业绩增

长,但呈现出不同的结构性特征。

财报显示,2026年上半年,长电科技实现营业收入195.27亿元,同比增长4.96%;归母净利润8.45亿元,同比增长79.41%;扣非净利润8.08亿元,同比增长84.73%。其中,第二季度实现净利润约5.54亿

元,环比增长约90%。

《中国经营报》记者注意到,真正有信息量的地方,并非5%左右的收入增长,而是利润增速远远甩开营收。这背后反映的是公司业务结构的变化,收入结构正进一步向汽车电子、高性能计算等方向倾斜。

长电科技、通富微电、华天科技三大龙头集体扩产,甬矽电子、盛合晶微、深科技等封测厂商已纷纷启动扩产计划。

大客户业务方面,通富微电表示,公司大客户AMD 2026年上半年营收和盈利均创下历史新高,其中数据中心业务营收同比增长一倍以上。据了解,该公司与AMD保持“合资+合作”模式,是AMD最主要的封测供应商,占其相关产品的80%以上。

相较以上两大传统龙头,主攻先进封装的盛合晶微上半年营收34.07亿元,同比增长7.20%;归母净

国内企业找到突破口

记者注意到,目前半导体封测行业已跳出传统封测的同质化竞争怪圈,拿到与晶圆厂、芯片设计公司平等协同的话语权,也能承接国产算力芯片、AI芯片的自主化封装需求,将成为国内半导体产业链绕开单点制程卡脖子的关键突破口。

根据世界半导体贸易统计组织(WSTS)的统计,今年上半年以AI服务器与存储芯片为代表的算力相关需求已成为不可逆的全局增长引擎。那么,本轮AI带动的先进封装高景气,是短期供需错配,还是长期结构性增长的开端?

对此,研究机构TrendForce集邦咨询分析师乔安表示,因为AI HPC对芯片算力需求爆发式增长,先进封装将迎来长期结构性的增长机遇,主要包含以下动能:“一是突破摩尔定律物理极限:传统单芯片依靠晶体管微缩来提升效能的方式,已面临物理极限与成本效益瓶颈。IC开始转向Chiplet架构,将不同功能及不同制程的小芯片进行异质整合,而先进封装正是实现高带宽、低延迟异质整合的核心关键技术。二是突破存储墙的限制:AI模型不论Training还是Inference皆极度依赖HBM,而HBM必须通过先进封装才能与GPU/xPU紧密串联。只要AI

对海量数据传输与算力的需求不减,先进封装的需求就会持续增长。”

乔安指出,封装技术已从过去低附加价值的“后段加工”,升格为决定芯片最终效能的前段延伸核心。“各大晶圆代工厂与封测厂都大幅增加先进封装资本支出,显现其作为未来数十年半导体产业主要成长引擎的长期趋势。”乔安说。

在这一长期趋势下,国内头部封测厂商正在加大资本投入扩产。长电科技、通富微电、华天科技三大龙头集体扩产,甬矽电子、盛合晶微、深科技等封测厂商已纷纷启动扩产计划。

6月,长电科技宣布拟在上海临港建设高端先进封测工厂,总投资78亿元;通富微电则通过定增进一步扩充“弹药库”,4个生产性项目分别投向存储芯片、汽车等新兴应用、晶圆级封测以及高性能计算和通信;6月底,盛合晶微上海临港东盛合芯3维集成芯片制造一期项目正式开工,总投资约100亿元,建设内容直指3DIC规模量产能力。

9月3日晚间,长电科技披露2026年度向特定对象发行A股股票预案,拟募集资金总额不超过65亿元,用于高性能计算高端先进封装平台扩产及其他4个项目及补充流动资金。



长电科技、通富微电、华天科技三大龙头集体扩产,甬矽电子、盛合晶微、深科技等封测厂商已纷纷启动扩产计划。

视觉中国/图