



天堂声谷数据标注众包服务

杭州远传新业科技有限公司

天堂声谷使用低成本高效率的众包模式，将海量数据处理工作以任务形式，匹配给合适的会员，根据客户要求对图片、文本、语音、视频等数据进行采集及标注处理，输出客户可用的数据成果。帮助客户训练算法模型、开展机器学习，提高AI领域的竞争力。



数据采集

众包会员采集，提供各种场景下的源数据。

- 文本数据采集：广告、杂志、试卷教材等
- 图片数据采集：实物图片、人物图片、场景图片等
- 语音数据采集：普通话、方言、特殊情景语音等
- 问卷调查：市场调研、产品使用体验调研等



数据标注

众包会员进行数据处理，将源数据转化为可用数据。

- 数据分类清洗：评价内容分类、语义分割、有效语音筛选、语料整理等
- 数据内容提取：语音转写文本、图片文字提取、文本关键字提取等

U 数据众包服务流程

您的需求 ○ 天堂声谷数据众包服务 ○ 数据交付



需求确认

- 工具选择
- 规范确认



标准校验

- 样本数据预处理
- 预处理成果标准校验
- 评估工作效能



合同签订

- 约定交付期限
- 约定整体质量要求
- 客户支付预付款



团队组建

- 会员招募筛选
- 质培团队组建
- 会员业务培训



数据处理

- 进度管控
- 质量管控



交付结算

- 客户验收数据成果
- 确认有效数据量
- 客户尾款支付

01 客户提供平台

会员在天堂声谷众包平台领取任务后，跳转至客户提供的标注平台在线标注。

- 信息安全性高，原始数据及标注成果均在客户平台。
- 标注整体质量及进度可视化，客户可随时抽检。

02 私有化部署平台

支持在内网环境私有化部署标注平台。会员领取任务后通过平台客户端工作。

- 私有化部署平台可实现多场景预标注，提升标注效率。
- 批量提交源数据并获取处理数据成果，数据隐私性高。

03 线下标注软件

根据客户指定或我司提供的标注软件，会员进行线下标注工作。

- 语音标注工具：Praat
- 图片标注工具：Labelme

平台工具



承接人员

01 专属职场会员

远传在杭州和重庆基地已建立专属职场，招募专/兼职会员在指定职场数据处理。

- 自有人员全程进行管理，沟通及问题响应更便捷。
- 会员量相对有限，以初次尝试或要求较高的数据处理为主。

02 合作基地会员

具备固定场地和管理人的全国合作基地，招募具备经验的专/兼职会员数据处理。

- 不亚于专属职场的现场和管理人员配备，支持线下培训。
- 会员稳定，可以满足数据涉密性高或长期大批量的数据处理需求。

03 平台云端会员

有过往服务经验的会员，通过线上培训和试标后，云端形式远程数据处理。

- 丰富的会员群体，可以满足各类有地域环境，人群年龄等要求的数据处理。
- 可大批量承接隐私性及保密性要求不高的数据处理需求。

1 标准校验

- 熟悉标注过程，深入理解标注要求和成果要求，形成标准样例。
- 针对性梳理易错点和提高效率的小窍门，用于培训分享。

2 培训答疑

- 每项任务培训师通过现场或远程形式，对标注会员进行专项培训。
- 标注开始7天内，针对标注成果中出现较多的错误点进行巩固培训，强调要求。
- 每项任务设置交流群，会员标注过程中遇到的疑点，难点直接交流。

安排专人与客户对接，标注中的疑点和质检偏差问题与客户及时沟通明确，质检标准阶段性做校验，避免交付出现批量错误。

3 双重质检

- 对现场质检人员进行专项质检能力培训，现场质检人员对标注成果进行初检。
- 设置专项质检人员，对初检结果进行复检和比对。

会员标注合格率持续不达标，取消此任务标注资格，对应数据回退数据池由其他人重标。

发现质检人员初检准确率持续偏差较大，取消质检资格，已质检数据安排其他质检员重检。

U 数据信息安全

数据隔离

针对客户采用独立数据库多数据源存储方式，保障源数据、成果数据的独立性、安全性。

数据脱敏

对源数据敏中的敏感内容进行脱敏预处理，避免会员数据处理过程中出现敏感数据泄漏风险。

数据交付安全

私有化交付：通过客户端传输数据，客户可以批量提交源数据并获取成果数据，隐私性安全性更高。

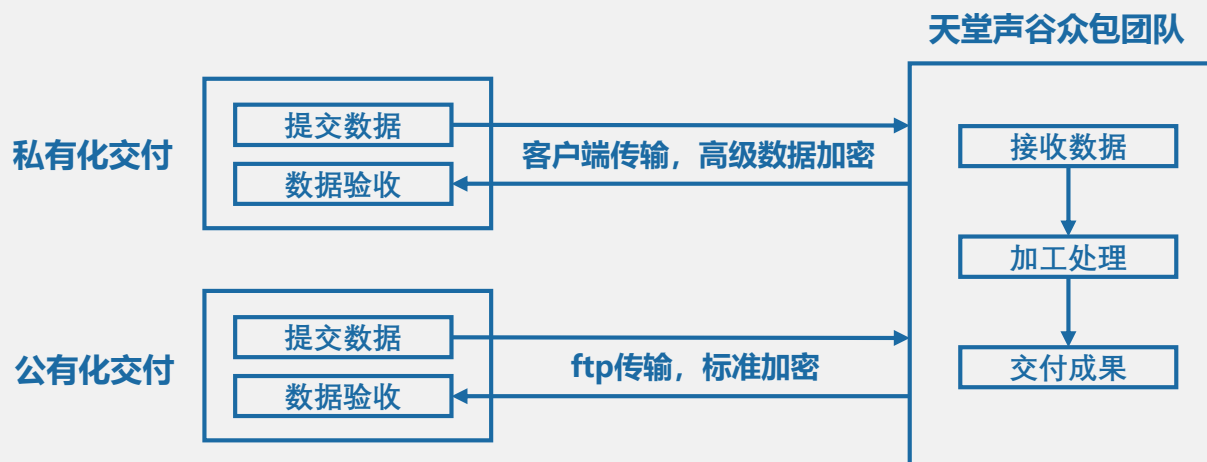
公有化交付：客户可通过ftp上传待处理数据，项目执行完毕后，下载结果数据，方便快捷。

会员管控

会员需完成实名认证，通过征信背调，签订保密协议后，才可以领取任务进行数据处理。同时设有黑名单机制，数据处理过程中出现违规行为的会员，将失去数据处理资格。

现场管控

职场区域实现摄像头共享监控，巡场人员对现场进行管理监督，禁止会员拍照，截屏。同时计算机均禁用USB口，以免数据泄露。



基于数据处理工作注重经验的特性，众包平台建立会员标签管理体系、技能培训认证体系，持续进行数据处理人才的培养及储备，保障业务匹配经验会员以及急单快速组队承接，完成各类行业及场景要求的数据处理工作。



数据处理会员选育流程



智能语音数据应用

语料采集，语音内容加工处理（性别，情感判断，语音文字转换等），为语音识别、语音合成等提供高质量语音数据。



智能交通数据应用

交通影像语义分割，车道线，障碍物、红绿灯等标识处理。为自动驾驶等智能交通技术提供精确的训练数据。



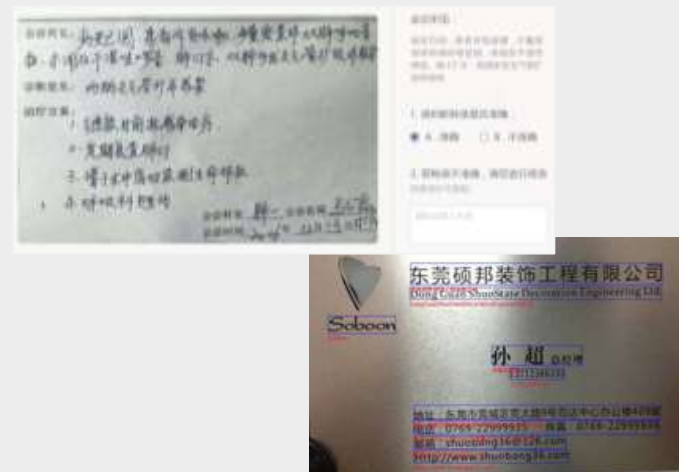
智能安防数据应用

人脸、人像面部识别，车辆识别等标注处理，为安防监控用的智能摄像头等设备的算法训练提供精准的数据。



OCR数据应用

名片，身份证，车牌，银行卡，通用印书体以及手写轨迹等多种垂类图片采集及内容识别提取，提升智能设备对多场景文字识别准确率。



OCR图片转写标注

累计招募培训3000余名会员，使用Labelme工具，近半年时间完成85万张图片处理，累计画框量达1235万，复检合格率98%。

带口音普通话采集

定向招募不同地域、年龄、性别的会员，在安静/噪音环境下按指定文本录制带口音普通话。累计采集语音8万余条。

粤语语音标注

定向推送任务给广州基地会员，使用客户提供的标注平台进行粤语语音标注，1个月完成100小时标注，质检合格率98%。

客服录音清洗分类

会员听取未接通录音，判断未接通的原因是关机，暂时无法接通还是停机，从而进行清洗分类，完善外呼系统。清洗准确率99%。

客服录音语音标注

为保障信息安全，会员驻场使用GoldWave软件对录音数据脱敏处理。脱敏后的数据匹配给平台有经验的会员使用Parrrt软件标注。1个月完成100小时有效录音标注。

行业客服知识库校准

招募有对应行业客服经验的会员，根据标准文本内容结合服务经验，对庞大的知识库内容进行比对和校准，校准准确率达98%。

小区道路异物采集标注

会员领取任务后，按指定要求线下拍摄小区道路照片。拍摄采集到的照片经过质检筛选后，会员继续使用Labelme对照片中道路，异物，障碍物进行标注，用于后期小区巡逻机器人算法训练。不到2个月即完成8万张有效标注。

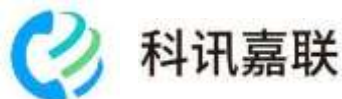
交通道路车道图片标注

通过百度数据平台，对图片中的车道线进行对应属性标注，为无人驾驶做优质的数据服务。已累计完成5万张标注。

道路交通标志图片标注

对道路交通标志进行标注，并按类目和清晰度分类。会员领取任务后，跳转至百度数据平台标注。2个月持续完成近20万张图片的标注处理，复检合格率达99%。

合作客户





我们的优势



处理速度高效

1万+标注经验会员，5万场外数据采集会员，构建数据处理人才库，具备百万数据处理能力。



承接价格优惠

成熟的众包分发模式，AI赋能工具实现预标注提升效率降低成本，收费低于竞品。



处理结果精准

专业项目人员全一对一跟进，完善的培训体系和审核机制，成果数据平均准确率达到98%。



数据信息安全

公有化私有化交付模式，敏感数据支持专人预处理，会员，现场，平台三重安全监管保障。



服务流程灵活

工具，会员按需多形式组合承接，支持急单快速组队，合作模式弹性灵活。



服务经验丰富

各类场景数据处理经验丰富，客户服务应用场景及AI服务机器人训练具有足够专业保障。



400-028-2200

Thank you
For your precious time and attention