

妙数标注团队

TAGGING TEAM - TAGGING TEAM - TAGGING TEAM

为智能世界提供充足的
数据生产力



目录

CONTENTS



01

团队介绍



02

业务矩阵



03

应用场景



04

核心优势



05

客户体验



SECTION 01

团队介绍

| 团队简介

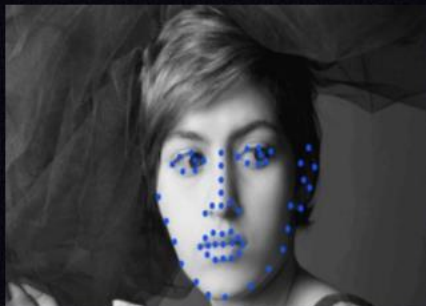
| 团队服务



团队介绍

INTRODUCE

妙数标注团队是一家为数字化服务而生的数据智能团队，团队以“为智能世界提供充足的数据生产力”为使命，提供图像、语音、文字、视频等内容的数据采集、数据清洗及标注服务，并可以根据客户的业务需求定制标注工具，生成高质量的训练数据。采用具有社交属性的运营模式，团队于2019年实现快速扩编，产出质量远超同期友商。



| 标注服务

团队可根据客户的定制化需求，提供多种类型的标注服务，以满足客户在模型识别领域的不同需要，如人物表情采集，OCR转写，大模型提示纠正等。



| 平台服务

适用于各种复杂场景数据采集需求，海量标注团队线下采集，为客户提供高质量源数据；提供高标准、高质量的数据清洗及数据标注服务，可以根据客户的业务需求定制标注工具，生成高质量的训练数据。

语音情绪判定

情感标注

样本筛选

人脸关键点标注

目标分割

SECTION 02

业务服务矩阵

| 计算机视觉

| 自然语言处理

| 语音识别

ASR语音转

OCR

文本分类

目标检测





业务矩阵

PRODUCTS



计算机视觉



自然语言处理



语言识别



关键点标注

人脸关键点、手势关键点、身体关键点、物体关键点.....

拳头服务



目标检测

依据需求框选标注出指定的目标对象

拳头服务



目标分割

对图像中存在的内容及位置，根据属性进行分割，支持单属性、多级属性。如人像分割、人头分割遮挡数据分类等。



2D/3D标注

依据需求对物体在平面或立体中的位置和大小进行标注，帮助自动驾驶模型增强视觉

拳头服务



样本筛选

依据需求可对图片集做属性归类



短视频、游戏打标签

依据需求可对短视频、游戏做属性归类，支持多级、自定义标签。



业务矩阵

PRODUCTS



计算机视觉



自然语言处理



语言识别



长、短文本多标签

按照规则对文本属性进行分类

拳头服务



OCR文字识别

对图像中的文字内容进行识别及标框，并以文本的形式进行转写



情感标注

标注文本中的情感倾向，分类积极和消极的文本，例如正面、负面或中性。



语义分类

对文本中的词或短语按照预定义的语义类别进行标注。





业务矩阵

PRODUCTS



计算机视觉



自然语言处理



语言识别



语音清洗

从语音数据中识别说话人的情绪状态；



语音情绪判定

对语音中的文字内容进行识别及标框，并以文本的形式进行转写



ASR语音转写

提取语音数据特征，转写为对应的文本或者命令

拳头服务





SECTION 03

应用场景

| 内容审核/内容标标签场景

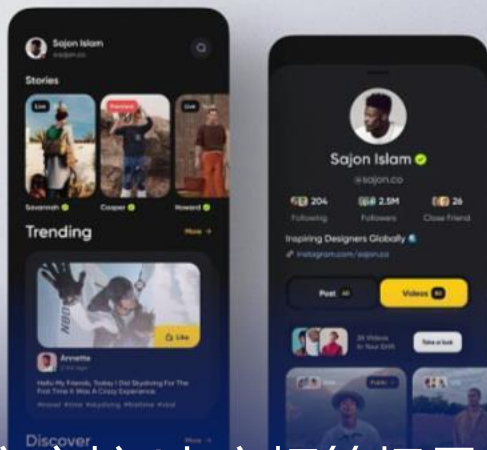
| 智能驾驶

| 大模型训练



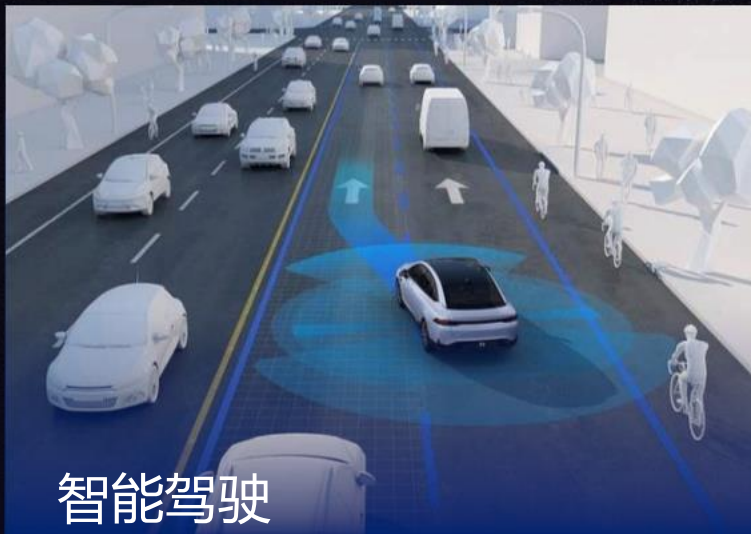
应用场景

SECTION TWO



内容审核/内容标签场景

通过专业的标注实现对问答评论、昵称签名、日记帖子、恶意推广、图文图集、音视频的内容合规合法性进行识别及个性化推荐。



智能驾驶

为训练自动驾驶领域的模型和算法提供了关键的数据支持，帮助实现车辆的感知、决策和控制能力。



大模型训练

用于训练通用大模型，让其具备对多种自然语言和图像数据进行分析和理解的能力。具体的任务需求会决定需要提供哪些类型的标注数据。

内容审核/内容标签场景

内容审核模型

- **图像/视频分类标注**: 对图像或视频进行分类, 标注其中是否包含某些敏感内容, 如暴力、色情、恶心等。
- **目标检测标注**: 对图像或视频中特定目标的位置进行标注, 以便识别和过滤敏感或违规物体。
- **文本分类标注**: 对文本进行分类, 标注其中是否包含违规、辱骂、恶意等内容。
- **实体识别标注**: 对文本中的实体进行标注, 标注关键信息以便识别和过滤敏感实体, 如人名、地址等。
- **关键词标注**: 对文本中的关键词进行标注, 以便识别和过滤特定的敏感词汇。

内容推荐模型

- **用户行为标注**: 对用户在平台上的行为进行标注, 包括浏览、点击、收藏、购买等, 以了解用户兴趣和喜好。
- **内容标签标注**: 对内容进行标签 (一般为多级)。
- **用户偏好标注**: 对用户的偏好进行标注, 如用户对不同类型内容的偏好程度。
- **相似性标注**: 对内容进行相似性度量的标注, 以便推荐与用户兴趣相似的内容。
- **热门度标注**: 对内容在平台上的热门程度进行标注, 以便在推荐中考虑热门度因素。

智能驾驶场景

以下是一些常见的样本标注服务，它们为训练自动驾驶领域的模型和算法提供了关键的数据支持，帮助实现车辆的感知、决策和控制能力。

- **图像/视频标注**：针对自动驾驶中使用的传感器数据，如摄像头和雷达，需要进行图像或视频标注。这可以包括对象检测、语义分割、实例分割、行人检测、车道线检测等任务。
- **地图和道路标注**：为了实现车辆在道路上的导航和路径规划，需要对道路、交通标志、交通信号灯、交叉口、车道线等地图要素进行标注。
- **行为和动作标注**：为了训练自动驾驶车辆的行为决策和规划能力，需要对不同驾驶行为和动作进行标注。这可以包括变道、超车、停车等常见的驾驶行为。
- **语音和语言标注**：如果需要进行语音识别或自然语言理解方面的训练，需要对语音数据或文本数据进行标注。包括语音识别、实体识别、情感分析等任务。
- **图像标注**：提供标注了图像中目标物体位置和类别的图像样本。例如，目标检测、图像分割等。
- **语音识别标注**：提供带有对应文本转录的语音数据标注。有助于训练语音识别模型。
- **模拟器和虚拟环境标注**：使用仿真环境进行自动驾驶能力的训练时，需要对虚拟场景、虚拟车辆、虚拟道路等进行标注。

垂直通用大模型训练场景

以下是一些常见的样本标注服务，它们为训练自动驾驶领域的模型和算法提供了关键的数据支持，帮助实现车辆的感知、决策和控制能力。

- **文本分类标注**：提供大量文本样本及其对应的预定义类别标签，用于训练模型进行文本分类和预测。
- **文本序列标注**：提供标注了每个词或字符的标签的文本样本，有助于模型对文本进行实体识别和词性标注。
- **语义角色标注**：提供标注了句子中每个词语的语义角色信息的样本数据，有助于模型理解句子中不同成分之间的语义关系。
- **句法依存标注**：标注句子中每个词与其在句中的依存关系，包括词与词之间的句法依存结构。有助于模型理解句子的语法结构。
- **图像标注**：提供标注了图像中目标物体位置和类别的图像样本。例如，目标检测、图像分割等。
- **语音识别标注**：提供带有对应文本转录的语音数据标注。有助于训练语音识别模型。
- **关系抽取标注**：标注文本中实体之间的关系，例如人物关系、地理关系等。这有助于模型理解和提取文本中的关系。
- **情感分析标注**：对文本进行情感分类，标注文本的情感倾向，如正面、负面或中性。这有助于模型进行情感分析。



04

SECTION 04

核心优势

| 服务优势

| 技术优势



核心优势

ADVANTAGES

PART1 服务优势

-MOUTH- -DAY-
6000 **1000**
月稳定在线 日稳定在线

高效率，高产出

月稳定6000名专业标注人员，日稳定在线400名专业标注人员



定制化，细追踪

专属客户经理，全程跟踪服务。同时根据不同客户提供定制工具，支持各类定制化需求



扩容快，质量高

采用具有社交属性的运营模式，实现快速扩编。同时具备完善的分组管理培训机制，实现快的同时，输出稳定的质量



核心优势

ADVANTAGES

PART2 技术优势

多重质检+复检机制保证数据标注质量，专注、专业、质量永远第一，绝不含糊

自动化质检

通过自动语言识别（ASR）、自然语言处理（NLP）等技术，质检规则模型灵活可配置，支持常见的内容关键词检测及复杂的质检场景，辅以人工抽检确保标注质量



自动化预标注结合人工标注

智能数据处理系统根据定制算法对数据包进行预标注，提升标注效率



高质量

人工质检审核

确保标注质量，不合格的数据打回重标。重标数据将再次质检审核过后，交付客户





核心优势

ADVANTAGES

PART3 丰富经验

团队拥有丰富的标注经验，在多个领域和行业完成了众多成功的标注项目，高质量地交付工作，满足各种复杂需求。

				交付时间	需求类型	任务名称	状态
交付时间	需求类型	任务名称	状态	2022/5/24	音频标注	音频指纹识别任务四期	已验收通过
2023/1/11	图片标注	人像分割	已验收通过	2022/5/27	音频下载	歌曲下载任务	已验收通过
2023/1/25	图片标注	手势关键点	已验收通过	2022/6/7	目标检测	真人衣服分割	已验收通过
2023/2/9	图片标注	视频人像分割	已验收通过	2022/6/10	样本筛选	皮肤筛选	已验收通过
2023/4/8	图片标注	速度测试分类	已验收通过	2022/6/17	音频标注	音乐重拍标注任务	已验收通过
2023/2/15	图文标注	图文多标签	已验收通过	2022/6/17	音频标注	英文ASR标注任务	已验收通过
2023/2/16	图片标注	人像分割三期	已验收通过	2022/6/24	音频标注	歌名修正任务二期	已验收通过
2023/3/9	图片标注	短视频app图标检测	已验收通过	2022/7/3	音频下载	歌曲下载任务三期	已验收通过
2023/3/24	视频收集	游戏视频收集	已验收通过	2022/7/7	图片标注	车道线标注	已验收通过
2023/3/30	图片标注	视频人像分割二期	已验收通过	2022/7/13	图片标注	SAM分割标注	已验收通过
				2022/5/24	文本标注	广告联系方式筛选	已验收通过
				2022/5/27	文本标注	广告语义筛选-第八期	已验收通过



核心优势

ADVANTAGES



多级标签内容标注

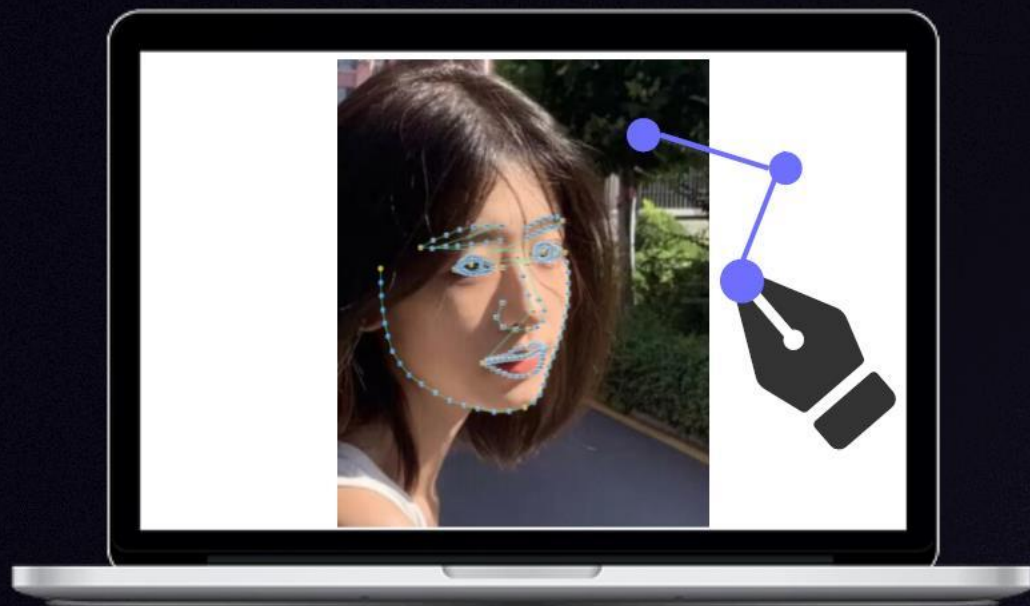


SAM分割标注



核心优势

ADVANTAGES



人脸关键点-222个点



文本标注

ASR音转文



SECTION 05

05

客户体验

| 案例分享

| 审核流程

| 业务流程



客户体验

CUSTOMER CASE

项目一：某互联网上市公司

CASE SHARING



项目痛点

任务需求量大，时间紧急

解决方案

我们团队拥有海量人才池，通过任务发布和可批量复制的培训流程，短时间内搭建起70人的专项项目组，快速消化任务

效果

对标其余专业标注团队，最早交付任务



客户体验

CUSTOMER CASE

项目二：某互联网上市公司

CASE SHARING



项目痛点

任务需要再人脸上标224个点，质量要求高，难度大，需要人员进行精细化操作

解决方案

团队拥有来自各行各业，专业程度高的项目成员，团队根据任务特征，精确定位目标人群，2天招募30余名美术专业成员，高效消化任务

效果

在确保时效的情况下，以100%的准确率完成任务



业务流程

PROCESS

01

需求评估

02

标注方案确认

03

试标注

04

质量确认

05

给出估价

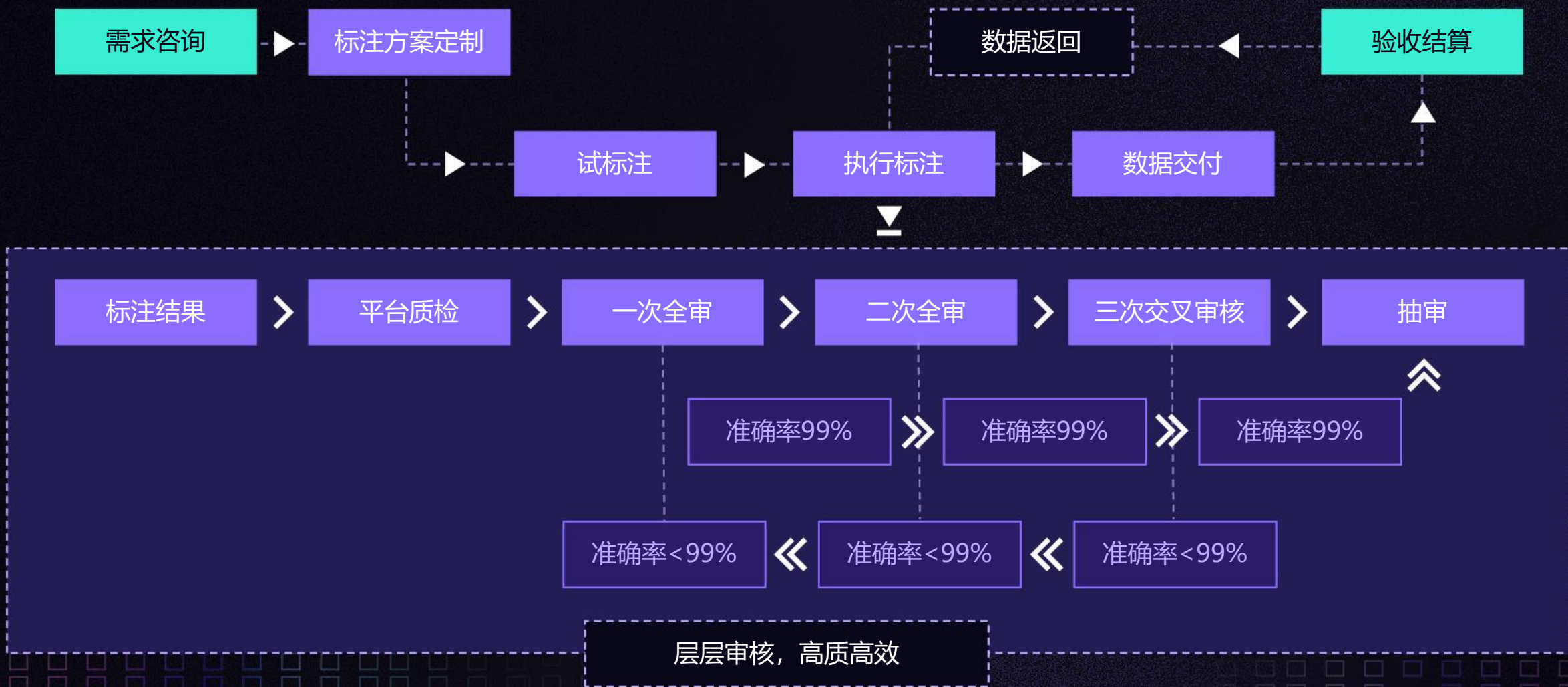
06

批量标注



审核流程

PROCESS



THANKS !

感谢大家观看！

