

数据标注众包服务平台

修订历史记录

日期（年/月/日）	版本	原因与修改描述	作者
2022/3/16	1.0	初稿	舒康城

数据标注众包服务平台

一、引言

- 编写目的
- 阅读对象

二、项目综述

- 项目背景介绍
- 项目应用场景举例
 - 人工智能行业
 - 调研咨询行业
- 项目功能概述
 - 数据采集
 - 语音采集
 - 图片采集
 - 文本采集
 - 视频采集
 - 数据清洗
 - 数据分类
 - 相关性评分
 - 数据标注
- 项目目标
- 产品整体流程和逻辑关系
 - 分发流程图
 - 分发信息图
 - 监控流程图

一、引言

1、编写目的

- ①对项目需求进行描述与界定，以便达成一致认可
- ②供团队成员了解需求情况，同时作为追踪项目进度的依据

2、阅读对象

本小组的前端，后端人员

二、项目综述

1、项目背景介绍

众包（Crowdsourcing）指的是一个公司或机构把过去由员工执行的工作任务，以自由自愿的形式外包给非特定的（而且通常是大型的）大众网络的做法。



人工智能众包打造的面向人工智能行业所需的数据采集、数据清洗和数据标注等数据服务的智能众包平台。

人工智能众包优势

- 多样本：有足够的用户满足多样本的需求（不同年龄、各地方言）
- 高效：有足够的用户保障需求完成效率（秒级回收数据）
- 弹性：解决高峰低估的波动

2、项目应用场景举例

人工智能行业

- 智能家居的指定唤醒词语音收集、数字串朗读，可通过语音识别接口自动为语音质量打分，只交付合格的数据
- 智慧医疗领域：X光片、病历等数据采集，以及对病源部位进行标注
- 自动驾驶领域道路图片采集和标注：采集道路相关图片，对道路和行人等进行标注，供算法识别

调研咨询行业

通过匹配用户基本信息（性别、年龄、职业等）和其他用户行为（购买力、购买偏好等），获取用户对产品或市场的第一手资料，从而进行调研分析

3、项目功能概述

数据采集

语音采集

基于文本朗读或者基于场景自行采集语音

图片采集

人脸图片、手势图片、手写体图片以及各种场景的图片采集

文本采集

如询问天气、叫外卖、查询高考分数、查询新闻等场景的文本语料采集，产品支持用户必须一次性提交指定条数的语料，确保语料句式的丰富

视频采集

各种场景如人脸、手势的短视频采集

数据清洗

通过检查数据一致性，处理无效值和缺失值，保障数据可用性

数据分类

可支持对文本或图片的情感色彩、表述目的、语言内容做分类打标，主要应用场景包括评论内容中的垃圾广告监测、品牌的负面舆情监测

相关性评分

可支持多张图片的相似度判定、给定query的搜索结果准确情况评估，适合电商平台、搜索引擎产品了解算法盲区、优化搜索相关算法

数据标注

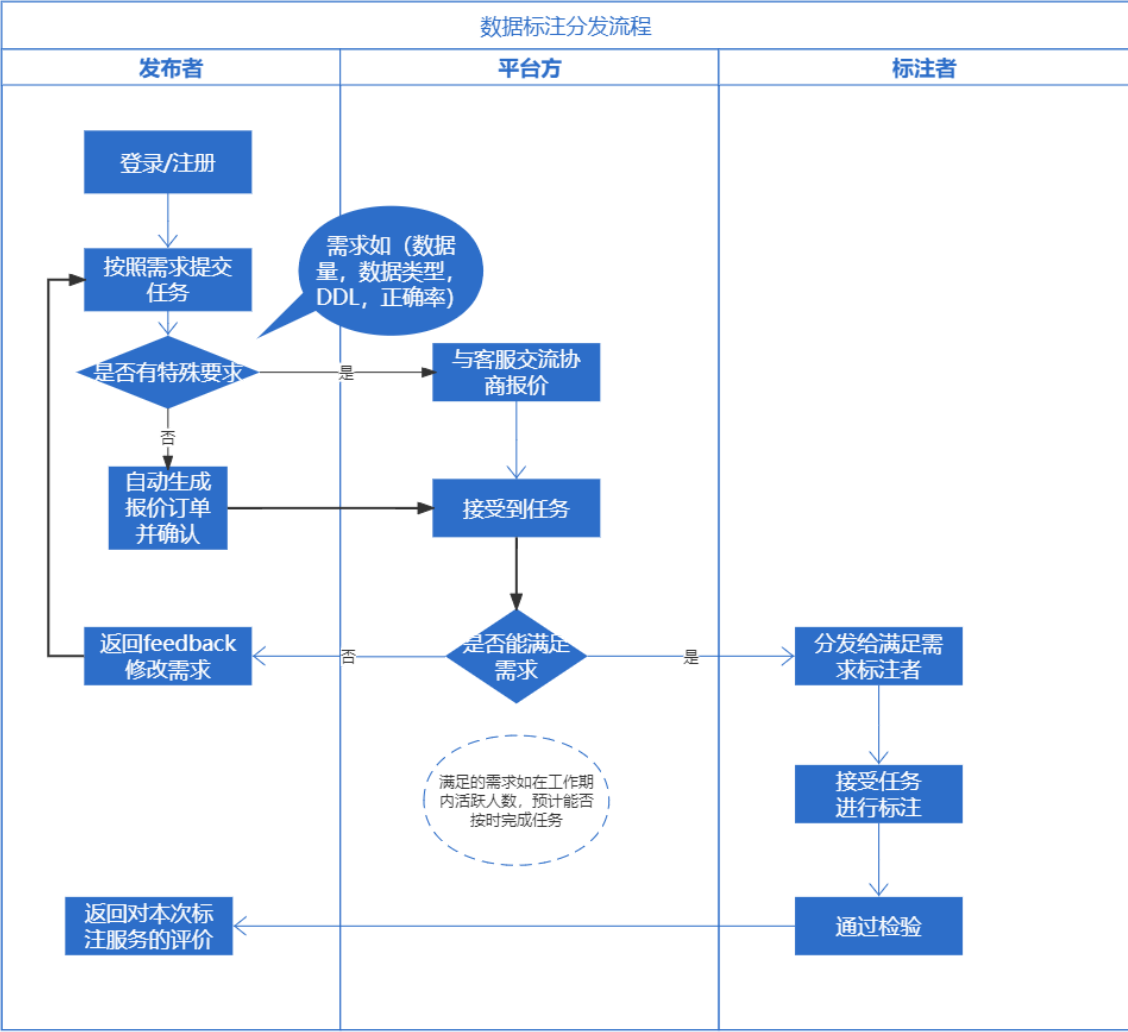
对数据进行标注，通过标注数据的审核机制把控标注数据的准确性，提升数据输出质量

4、项目目标

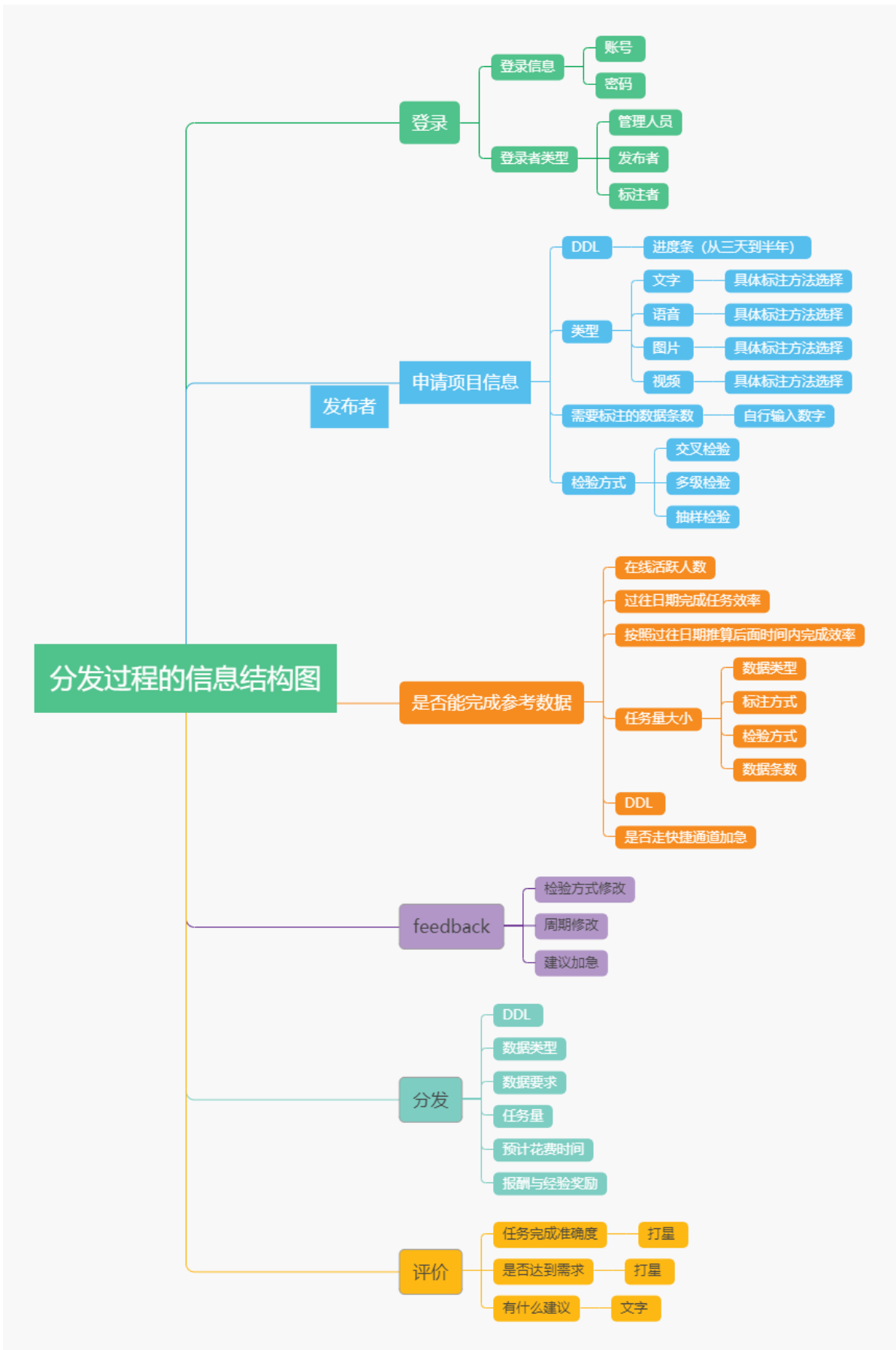
	发布者	管理员	标注者
功能	发布任务 查询已发布任务完成情况 对已完成任务评价	任务分发 任务提醒 人员管理 风险与合规监控 客服交流界面	标注者类型选择 标注者能力评级 标注者任务列表 标注者薪酬状态

5、产品整体流程和逻辑关系

分发流程图



分发信息图



监控流程图

