



Stardust
星 尘 数 据

产品手册

赋能人工智能，加速AI平民化



CHAPTER 1

星辰数据简介

Data is the new oil.

数据是AI时代的石油。

- 吴恩达

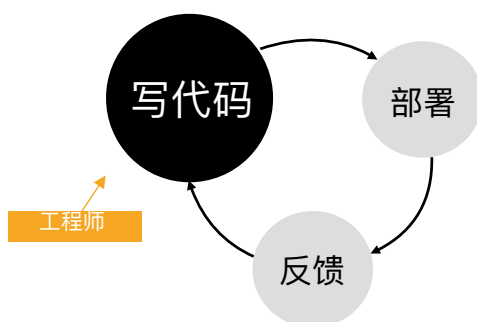
星辰数据是一家提供数据标注服务和AI数据中台系统的SaaS公司。我们为全球人工智能企业提供服务，为计算机视觉、语音识别、NLP任务提供了高质量的训练数据。我们专注于提供行业领先的AI数据中台系统，实现模型的生命周期中的数据生产、数据管理、模型训练、模型部署等环节。

数据作为AI算法的电能，如何在保证质量的前提下降低成本并大规模生产是行业的痛点。我们的数据处理平台Stardust提供AI数据全流程标注和质量管理，它提供了数据质量、成本、可扩展性的最佳解决方式。

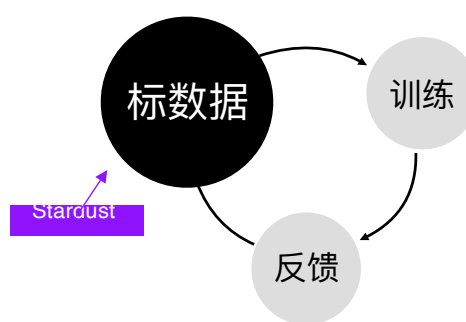
我们是一支世界级的团队，正在改变AI行业的运作方式。我们曾在硅谷、华尔街等地工作过，有着从创业公司、硅谷独角兽公司到上市公司从0到1的经验。星辰聚集了国内外人工智能科学家，数据处理数过亿，立志打造世界顶尖AI数据中台系统。

我们以赋能人工智能，加速AI平民化，让天下的企业都能做出高质量的算法为愿景，我们的使命是打造世界最领先的AI数据中台系统，简化算法团队研发和算法生命周期管理。

软件开发流程



AI开发流程





目录 CONTENTS

星辰数据简介 Brief Introduction

数据是AI应用的关键 Data is the New Oil

AI持续发展·数据服务成为趋势	With the Development of AI, Data Service Becomes a Trend	4
算法	Algorithm	4
数据	Data	4

问题 Problem

数据服务问题	Data Service Issues	5
--------	---------------------	---

数据平台介绍 Platform Description

数据质量是核心	Data Quality is the Core	6
自动化标注是关键	Automated Annotation is the Key	6
服务理念	Service Idea	7
服务流程	Service Process	7
六项数据服务质量承诺	Six Quality Commitments	7

产品优势 Product Advantages

质量管控	Quality Assurance	8
流程质量控制	Process Quality Control	8
系统质检	System Quality Inspection	9
机器辅助质检	Machine-Assisted Quality Inspection	9
数据采集质量保证系统	Data Collection Quality Assurance System	9

数据安全 Data Security

法律手段	Legal Means
权限控制	Permission Control
数据管理策略	Data Management Strategy
数据传输	Data Transmission
数据存储安全系统	Data Storage Security System

服务指标 Service Indicator

性价比	Cost Performance
可靠性	Reliability
稳定性	Stability

标注工具 Tools

点云工具	PointCloud Annotation
视频工具	Video Annotation
音频工具	Audio Annotation
图像工具	Image Annotation
文本工具	Text Annotation



服务场景 Use Cases

10	计算机视觉 Computer Vision	15
10	自动驾驶 Self Driving	16
10	语音/文本运用 Audio/Text	17
11	数据采集服务 Data Collection Service	18

产品亮点 Highlights

11	标注工具 Annotation Tools	20
11	算法辅助 Algorithmic Assistance	21
11	项目流程管理 Project Process Management	22
12	SaaS 自动化系统 SaaS Automation System	22
12	工作流管理系统 Workflow Management System	23
12	工作流统计系统 Work Statistics System	23
12	数据处理与管理 Data Processing and Management	23

无人车闭环解决方案 Self-driving Closed-loop Solution²⁴

12		
13	产品定价与定制化服务 Pricing and Customization	30

合作伙伴 Partners

13		
13		
14	成功案例 Cases	33



CHAPTER 2

数据是AI应用的关键

——AI 持续发展，数据服务成为趋势

算法

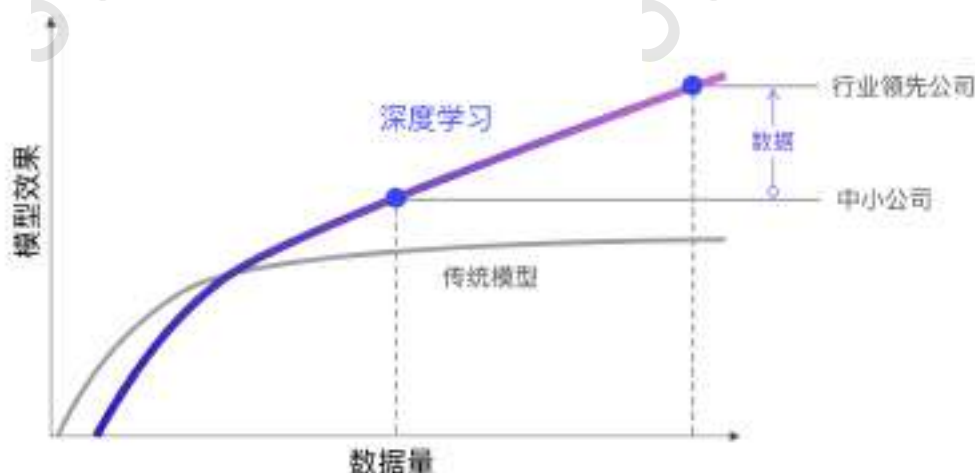
AI在技术领域不断突破，一方面云计算的快速发展，计算资源价格持续下降，解决了硬件和平台的问题。另一方面，几乎所有的算法都能够在GitHub上找到代码，Google AutoML等集合大量算法的工具出现，使得AI应用在框架及算法的不断开放下更容易实现。可见AI的竞争，理论上应该是人人平等、没有壁垒的。每一个需要AI技术的企业、机构，都应该能快速得到现成的、能解决实际问题的、所见即所得的AI技术（Off-the-shelf AI）。然而，现实情况却不尽相同，虽然算法理论不断丰富完善，能够应用在真实场景中的模型却极为稀少。

人工智能算法是一个开源的领域，没有竞争壁垒。数据是AI界最根本的竞争力，拥有了数据就像拥有了矿产。业界的一个共识是[大量数据+普通模型]往往会比[少量数据+高级模型]的准确度要高，竞争壁垒更强。小公司初期拥有的算法优势，会被后起的大公司用数据优势所消解。

数据

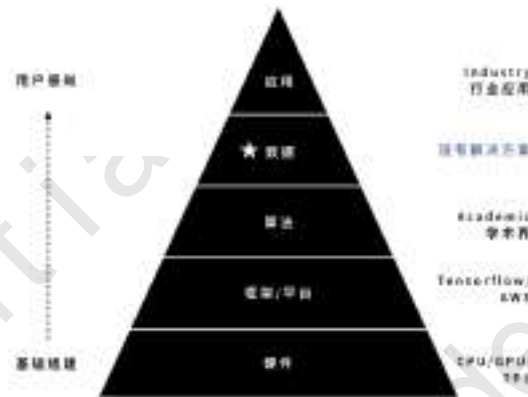
现实中供AI训练应用的数据没有跟上理论和行业需求的快速发展，从而限制了模型的大量落地应用，前期的数据与算法之争逐渐偏向于对数据的争夺，数据成为现阶段AI应用的关键。

目前AI训练的数据主要以人工标注为主，导致了AI应用的成本较高。随着深度学习成为AI行业的主流算法，数据的需求量也极大地增加。训练数据量的增加，对模型的效果永远具有正向作用。





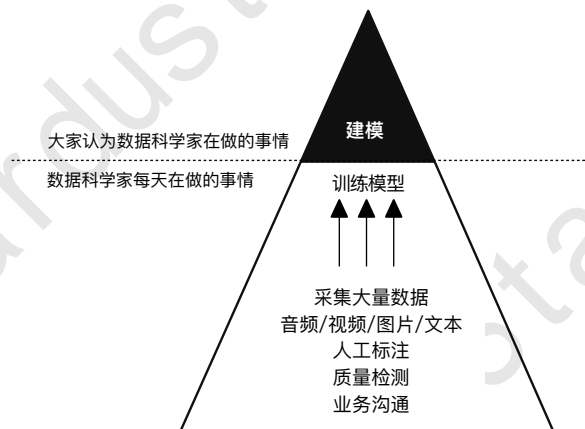
同时，随着AI进入更多、更细分的行业，虽然AI计算平台、框架、算法持续完善、更加易用，但供AI训练的数据却分散在各场景中。另外互联网的快速发展虽然积累了大量数据，但缺少供AI场景落地的数据，因此数据采集等数据化服务成为AI行业的机会。专业AI应用数据服务成为当前的行业发展瓶颈。



人工智能从产业角度

CHAPTER 03

问题



李飞飞的ImageNet的出现，直接改变了AI的发展轨迹，历史向深度学习抛出了橄榄枝。虽然我们目前仍然不能确切的知道，深度神经网络是否是最有效的高维数据空间搜索和求解方式；迁移学习（Transfer Learning）、元学习（Meta Learning）、深度强化学习（Deep Reinforcement Learning）、胶囊网络（Capsule Network）等新型的网络架构，也正在逐步走进我们的视线，训练数据在主流的深度学习训练中，仍然起到了决定性的作用。



而当前大规模训练数据的缺失，限制了人工智能的快速发展。究其原因有以下几点：

- 1.数据生产服务没有跟上AI理论和行业需求的快速发展
- 2.少数公司垄断了大量数据，形成了数据霸权
- 3.训练数据效果无法评价，市场交易阻力过大
- 4.缺乏安全有效的交易机制，保证AI交易的安全

数据服务问题：

- 1.数据的生产流程混乱，难以管理
- 2.数据经过多道贩卖，中间商费用过高
- 3.数据质量无法有效测量和保证

星云数据及其平台的出现，目的是要解决数据服务的问题。

CHAPTER 04

数据平台介绍

Stardust平台运用“人机结合”（Human-in-the-loop）的智能辅助工具，实现人力采集、快速高效标注、质量监控等工作。我们通过“准入考核”、“动态测量”、“动态发题”和“人机结合”等标注方法确保了标注的全流程管理。智能化辅助标注工具，实现了快速标注及有效标注。全流程标注平台实现了数据标注的交互审核与验证功能，使得标注的数据质量更高，更适合人工智能应用场景的开发。相比于普通数据标注和检查方式，该功能不仅将数据标注精度提高到 99%以上，还能轻松节省 1/3 的时间成本，提供的数据及模型服务更适合人工智能场景的开发。目前公司已有涵盖各种应用场景的数据标注服务。

数据质量是核心

监督学习下的深度学习算法训练十分依赖于标注数据，数据集质量的高低将直接决定算法模型的效果。Stardust平台的质量控制系统，支持多重指标定义、标注对象级别的质检和错误记录，能以多重方式进行抽检和统计、回溯历史记录，确保标注对象级别的数据质量。

自动化标注是关键

Stardust的SaaS自动化，支持整个标注、质检、上岗、系统对接的自动化。我们自研的需求标准化工具、培训考核上岗自动化工具、质检自动化工具以及AI辅助标注自动化工具为行业提供了一套完整的解决方案，通过事件触发脚本，极大减短了客户从提出需求到项目完结的时间。其中3D框AI辅助标注有效将行业平均1.5框/分钟的标注速度提升到20框/分钟，在AI辅助下的语音分割则提升了10倍速率。



服务理念

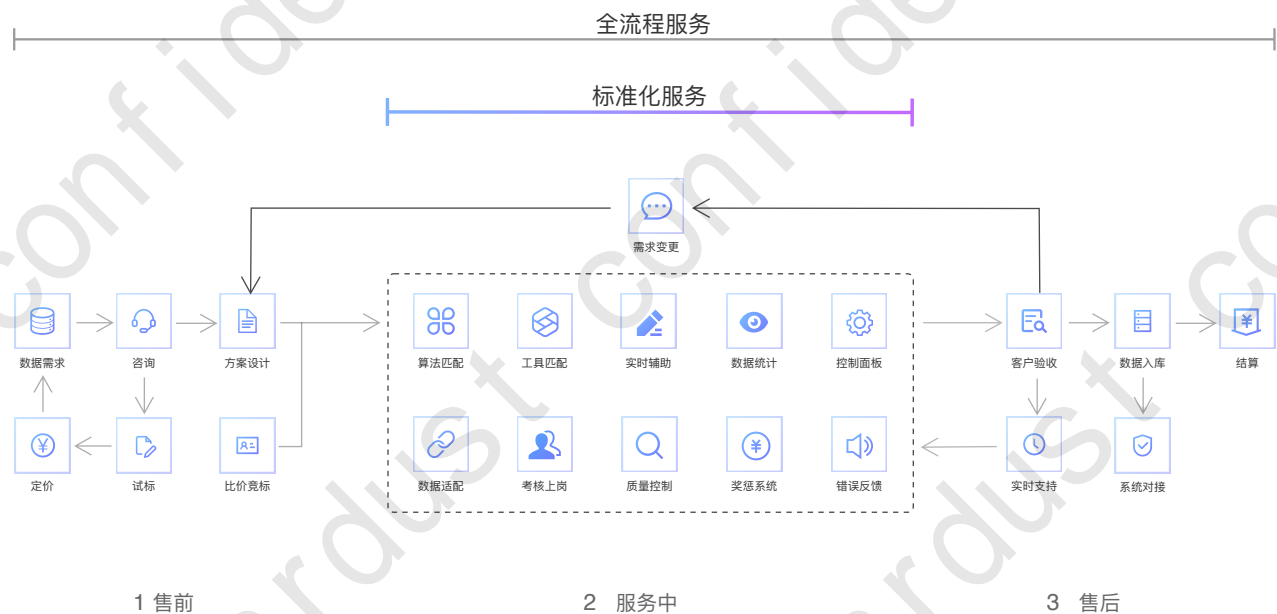
以提供低成本、高质量的数据服务为目标，我们的服务覆盖各类人工智能应用场景。基于目前AI行业的数据需求，星尘数据依托智能化的标注工具，安全的数据管理平台，实现了对场景化数据的收集、标注、管理及分发等工作，致力于打造人工智能时代的数据工厂，为各种场景提供可直接应用的AI训练数据。

服务流程

我们的服务包括「标准化流程」和「全流程服务」：

相比标准化流程，全流程服务支持自定义数据质量、支持加急处理、支持更多数据类型。

在服务流程上，全流程服务提供全面的售前服务以及个性化的售后支持。此外，全流程服务可以满足客户验收时产生的需求变更的需要。



六项数据服务质量指标

可靠性：随时处于可服务状态，保证时效。

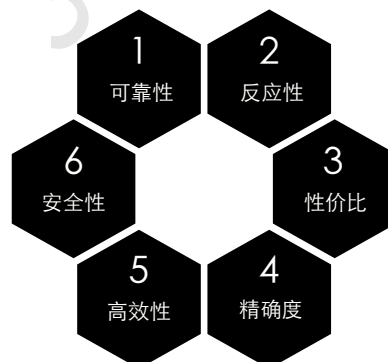
反应性：及时给顾客答复，提供服务的主动性。

性价比：以最优性价比，提供高质量数据服务。

精确度：确保最终交付给客户数据的高质量。

高效性：高效率服务，缩短交付时间。

安全性：数据安全保障，不被泄露。





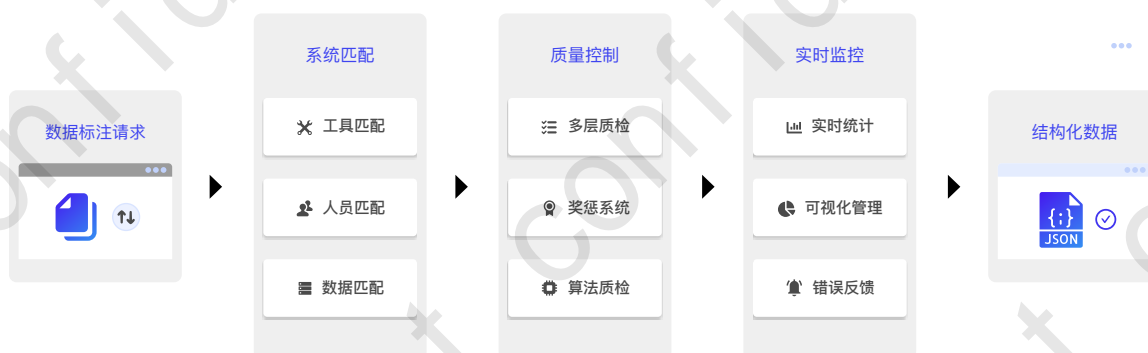
CHAPTER 05

产品优势

质量管控

针对数据标注与数据采集采取不同的质量保证系统。

对于标注系统，Stardust平台通过多层质检，以及在奖惩系统和算法质检采取不同的严密措施来进行数据标注的质量管控。



我们通过多重机制保证质量，最高可达99.9%的精确度

流程质量控制

结合谷歌等行业各种标注流程质检算法，设计一套结合标注、审核、抽样、去重等流程严格把关的系统。针对日益复杂的标注需求，将复合标注需求自动拆解成为单一标注单元。通过全过程中的层层严密的质量监控，保证数据的高质量交付。

数据经过预处理后，进入标注流程时被智能分配到人工标注或算法标注，被标注后进入质检流程，同样被智能分配到人工质检、系统质检或算法质检进行质量控制，最后再由星尘优质的QC质检团队进行质量检测。

此外，在标注流程中，以做审分离式标注的形式对标注数据进行交叉审核，被标注数据在用户自定义的标注池、检查池、抽查池中来回流转，并通过在不同池子中配置不同职能的标注员进行交叉审核的机制来保证最终进入完成池时的高精确度。



系统质检

根据数据标注的类型分配在此类型答题准确度和效率高的标注员。此外，在新项目开始前，对筛选出的标注员进行测试。若测试结果不达标，则无法参与对应项目的标注或查验。

我们平台支持实时正确率估算，且在标注人员标注过程中通过埋点的方式动态地发放一些有标准答案的真题，以不定时对标注员进行准确度监测、定义对应能力标签、奖励惩罚等的机制来保障标注员答题的效率和准确性。

机器辅助质检

作为一套独立系统，算法质检系统持续对Stardust标注系统中的新提交题进行判断。对于每一个项目，算法质检系统都有对应的算法进行质检。这套逻辑是和Stardust系统解耦的，是完全不透明的信息。甚至连Stardust都不知道检查时由人完成还是机器完成的。

Stardust系统创建新的项目A，并广播；算法质检系统接收到广播并判断是否能够质检；如果具有质检算法匹配该项目（如Algo A），则开启一个实时质检服务；算法质检系统对A项目的检查池进行轮询，发现有未检查的题目就发送到相应的质检算法；算法返回判断结果：若通过，则直接调用接口通过；若不通过，则调用接口驳回，并带有错误标签。

算法质检系统的目的是保障数据服务质量的下限，即尽量避免由于标注员的疏忽和漏题导致的显而易见的错误。

数据采集质量保证系统

Stardust平台通过严格的质量管理系统和相似度检测来保障所采集数据的高质量和真实性。例如人脸数据采集，我们采取三重检测算法对图片进行审核。首先审核该用户递交的数据是否为同一人脸，其次检查上传人脸是否和已上传的图库重复或明星脸相似度过高，最后检查图片是否在全局图片库存在相似性。

CHAPTER 06

数据安全

星尘数据安全方案结合相应的法律措施、系统措施、数据安全措施和安防措施来保障用户的数据安全。

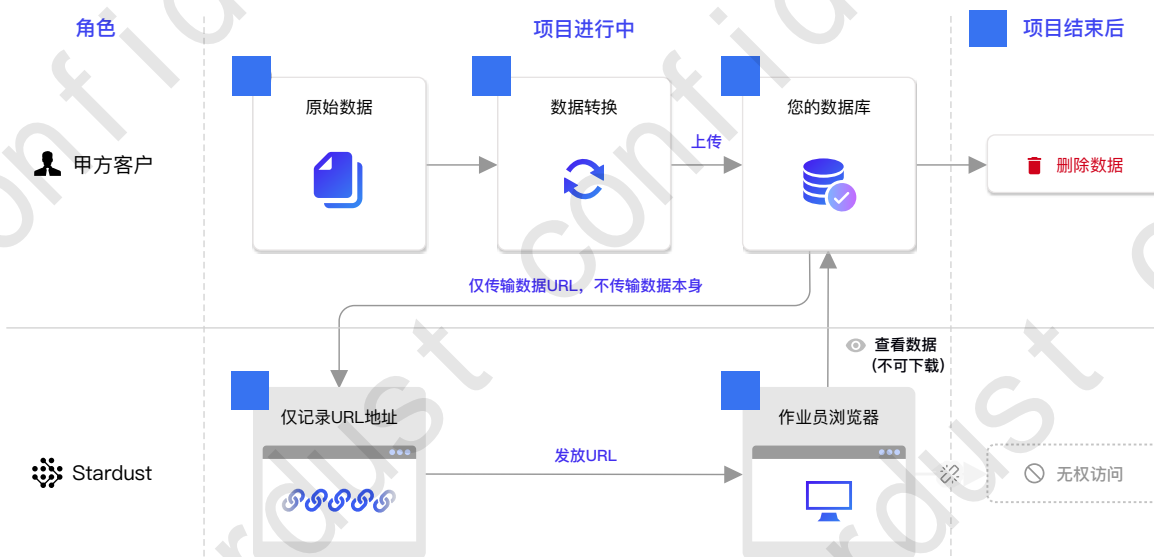


法律手段

星尘数据通过一定的法律手段来保障用户的数字版权，杜绝二次售卖与恶意泄露，例如与用户签订保密合同以及一些合法性隐私保护与授权协议；制定严格数据安全规范并进行周期培训，加强员工的保密意识。

权限控制

Stardust平台通过权限管理（如数据权限以及访问限制），通过白名单和访问控制、动态水印以及让数据具有不可复制性，使得黑客无法得知下载链接、无法伪造IP地址、无法批量下载，有效防止数据泄露；此外，平台设置文件服务器的访问频率限制，同时记录所有用户的访问地址和时间，设立高频预警和禁止IP访问的规则也可以防止用户窃取大批量数据。



数据管理策略

Stardust平台在数据传输时通过一定的数据适配手段并使用安全的TLS与数据隔离手段来保障数据的安全始终牢牢攥在用户的手中。

数据传输

在数据输入与导出Stardust平台时通过统一的数据结构对客户的原始数据进行数据转换，并使用传输层安全性协议（TSL）对数据进行加密传输，保证两个数据库间通信的保密性与可靠性，保障传输的信息不会被窃取。



数据储存安全系统

Stardust平台的数据管理策略中首要的就是数据是和标注平台分离开的。在客户的原始数据经由转换之后，仅上传到甲方的数据库中；此外，数据管理策略中URL保密机制也尤为重要，甲方数据库仅向Stardust平台传输URL，而不传输数据本身，此过程大大削减了黑客大规模爬取数据的风险。

除此之外，Stardust平台在数据安全方面提供定制化服务，客户可以定制更加严格的数据加密以及数据访问限制策略为数据安全上多重保险。

CHAPTER 07

服务指标

50%

性价比

99.9%

可靠性

100M

稳定性

专业、快速、优质的数据服务，为您省去数据收集与标注的繁琐过程，助力您的人工智能模型；包含菜单化服务、透明化行业服务标准的标准化服务，为您避开行业中坑蒙拐骗的行为。

性价比

国际顶尖数据科学家结合“人机结合”（Human-in-the-loop）智能辅助工具打造的标注系统确保数据服务更加高效；严格的质量管控保证系统保障了最终提交给客户的标注结果具有更高准确度；更低廉的价格为客户保障了数据服务的最大性价比。

可靠性

我们的平台支持实时精确度估算和AI辅助检查，凭借丰富的项目管理经验和灵活的定制化工具，获得数百家行业领先公司的信赖。

稳定性

我们在全球拥有丰富的人力资源，星尘数据可支持每天100M + API请求。



CHAPTER 08

标注工具

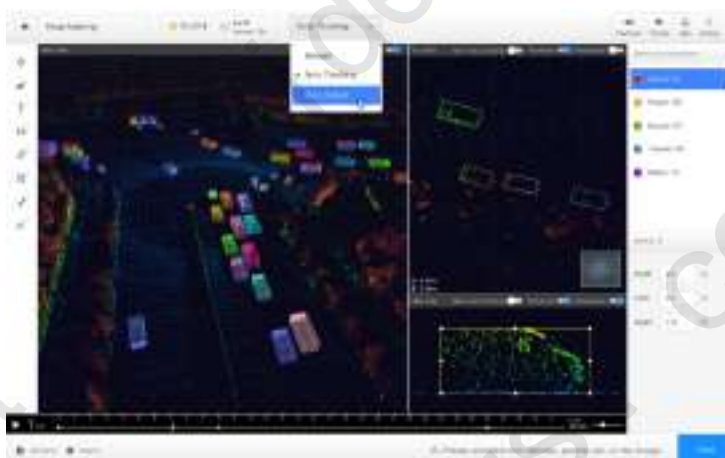
平台支持大规模数据采集、标注工作，适用于各种复杂场景数据标注的需求，标注数据类型包括点云、视频、音频、图像和文本。

点云工具

行业里点云工具存在开发、维护成本高，工具功能不完善，工具标注效率低，标注质量低等痛点。

我们的工具具有五大视图精准标注、三大模式辅助标注、连续帧图自动标注等优势。

点云工具支持对点云中的车辆、行人进行标注或者对物体进行语义分割。



视频工具

视频工具支持对视频中出现的物体、文字、语音、情景等进行标注，包含2D框、多边形、直线、曲线、点、立体框标注工具，支持2D拉框、3D拉框、语义分割、标签标注、关键帧等标注功能，支持子任务标注。





音频工具



音频工具支持对全球主要语言和语料、包括方言、特殊情景语音进行识别标注。标注工具具有音频提取、分割、分类、识别、转写能力，支持多层任务嵌套，可自定义选项，支持多媒体信息，支持正则辅助验证。

图像工具

图像工具支持多选、复制、旋转、放大/缩小、显示/隐藏、关键帧标注、支持自动补间、共用边切割、分割、历史标注结果回看、关键帧标注、自动贴边、自动映射、预标注能力、预加载、多媒体信息等功能。



图像工具

文本工具



文本工具

文本工具支持情感分析、知识库、关键词提取、文字翻译、搜索引擎、文本选择、文本对象、实体提取、文本摘要等标注等。

可对文本进行分词、槽位、句法、情感、事件要素、语种互译等不同层级标注。提供正则辅助验证、自动补全，文本算法预标注、自定义 API 接入、自定义辅助算法，支持实体识别、关系提取、摘要、自动分类等算法。



CHAPTER 09

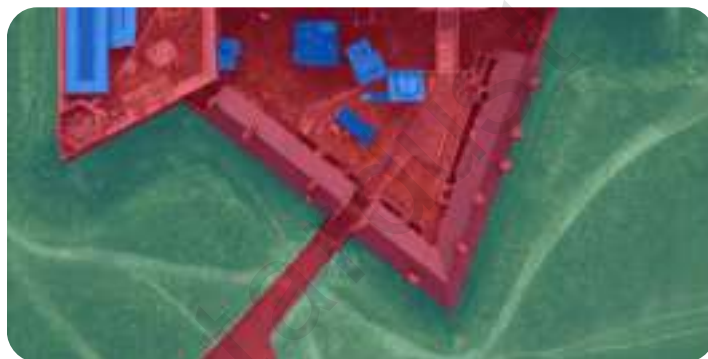
服务场景

01. <u>无人车</u>	02. <u>安防</u>	03. <u>消防场景</u>	04. <u>客服</u>	05. <u>金融</u>	06. <u>医疗</u>	07. <u>电商</u>
车辆标注 车道线标注 高清地图 LIDAR标注 2D分割 3D分割	人脸标注 人体标注 视频标注 语音识别 行人跟踪	行为分析 视频分析 物品分类 内容审核 用户画像 商超收集	文字翻译 语义关联 词性标注 语音转录 语音采集	知识图谱 证件识别 OCR 语义分割 手写识别	病例转录 人体标框 3D画框 骨骼点标记	搜索完善 情绪分析 商品推荐 语音采集

计算机视觉

语义分割

视频工具支持对视频中出现的物体、文字、语音、情景等进行标注。



立方体

2D图像中的立方体标注，为您的机器视觉模型提供空间感知能力。





多边形

对于不规则形状的目标物体标注，助力您的机器视觉模型。



轮廓标注

高精度物体轮廓标注，助力您的机器视觉模型。

关键点

对象物体的关键点标注，助力您的机器视觉模型。



图像分类

将对象物体分类，助力您的机器视觉模型。



自动驾驶

点云标注

可视化3D激光雷达点云标注工具助力您的自动驾驶模型训练。支持物体的位置、大小、朝向等属性，以及多传感器融合及连续帧对象追踪。



点云语义分割

对3D激光雷达点云进行语义分割，支持对栏杆、行驶区域、道路分隔线等多种区域进行单点级别的高精度分割。

2D矩形

2D框标注助力您的物体识别模型。



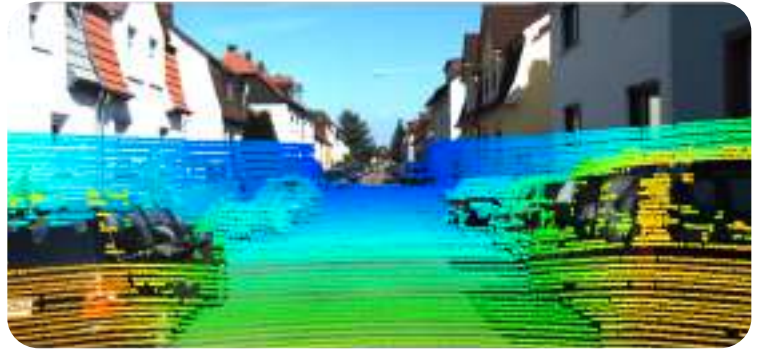
图像语义分割

语义分割可以将图像进行像素级的分类，高精度对象分割将助力您的机器视觉模型获得更完整的场景知识。

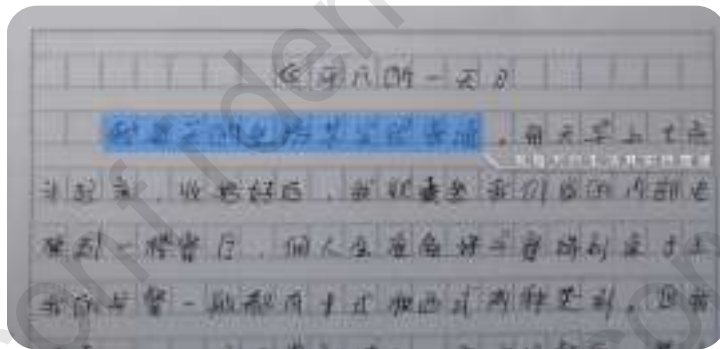


融合标注

为自动驾驶技术更快注入数据燃料，以帮助汽车制造商和自动驾驶汽车公司加速感知算法的开发。



语音 / 文本应用



文字识别

支持印刷体、手写体及定制化场景的图片文字识别服务。

语音标注

语音转写，可将音频数据分段并转换成文本数据，加速您的信息处理、数据挖掘和机器学习模型。



文本实体

识别文本实体，提取实体关系，助力自然语言处理。





数据采集服务

采集内容

- ① 视频、人像采集：非敏感区域视频采集、特殊场景视频采集，不同光照、角度、背景的人脸、人体、手势、面部表情等；
- ② 自动驾驶：支持对道路和图片进行信息采集；
- ③ 常规类采集：支持车辆、服饰、食品、动物、植物、商标、场景、手写体、印刷文字等信息采集；
- ④ 语音采集：支持对普通话，各地方言，少数民族语言，各式英语、亚洲语系、欧洲语系、美洲语系等语言进行采集。

设备&条件

- ① 具备特殊设备采集能力：手机、红外相机、3D深度相机、双目视觉、高保真设备等；
- ② 支持特殊场景信息采集：车载、安防、无人商超。

采集范围

- ① 覆盖国内30个省份，具备汉族，少数民族采集能力；
- ② 覆盖全球18个国家，具备白人、黑人、印第安人、中东人、中亚人、南亚人、东南亚人种采集能力。



自动驾驶采集

- 采集方式：前视、环视；
- 采集范围：高速公路、市区内道路、测试场地等均可；
- 车辆类型：货运、客运；
- 采集能力：北上广深等国内大部分城市均可；
- 采集设备：摄像机、高度相机、双目摄像头、3D激光雷达等；
- 采集类型：图像、视频。





CHAPTER 10

产品亮点

标注工具

平台支持大规模数据采集、标注工作，适用于各种复杂场景数据标注的需求，标注数据类型包括点云、视频、音频、图像、文本和数据采集。



标注通用

支持项目灵活配置、模板导入、父子任务、AI算法辅助等配置能力



标注界面

标注工具经过长期打磨，具有30多项功能，支持各种标注流程。例如：标尺、辅助线、高亮、筛选、边界、附加信息、挂起、消息



结果面板

简洁的结果显示方式，让标注流程和检验流程非常高效。包括高亮、隐藏、锁定、预览、属性弹出等功能



点

点标注工具，支持序号、点组、颜色等工作



线

线标注工具，支持曲线、增删点、形状调整等



2D框

2D Box工具，支持大小数量限制、旋转、形状调整、预设大小等功能



多边形

多边形标注工具，支持分割、调整形状、增删点、笔刷、橡皮擦等功能



立方体

支持立方体标注工具，支持前后面、灭点、旋转等功能



3D标注

支持3D标注的各种场景，例如3DBox、3D语义分割、融合标注、连续帧标注



选择

选择属性工具，支持默认选项、多选、搜索、API接入等功能



语音/音频

语音标注支持音频段分割、分类、转译等功能



文本输入

文本输入工具，支持API接入、知识图谱、数据校验等功能



NLP

NLP标注工具，支持实体标注和实体关系等功能



数据采集工具

数据采集、收集工具，支持表单、语音、文本、图像等采集方式。

算法辅助



算法辅助标注

辅助标注算法，包括3D、CV、NLP、ASR等标注工具，支持10余种场景，极大提高标注效率



算法辅助质检

算法辅助质检模块，支持多种质检场景，可自定义API接入

采用人机协同（Human-in-the-loop）的智能标注工具，借助深度学习和人群协作，提高标注的速度和质量。通过对人工智能数据管理的研究，借助智能算法，将标注过程中人工劳动的重复率降到最低。星尘数据开发能够处理大量数据标注的AI辅助标注平台，（基于Mask R-CNN模型实现对标注图片中每个物体的边界框（boundary boxes）和分割掩膜，排序、过滤Anchor，通过最后生成的识别框，分别以虚线、实线来表示，这些 mask 将被调整并最终放到原图的正确位置）。在标注时，能够通过智能识别功能辅助画框与分割工具，标注者只需要简单的对需要标注的数据进行选点，标注工具会自动的完成所有的工作，提升了50%–80%的数据标注效率。结合各类标注场景，应用深度学习算法，开发适应各种场景的标注工具。



平台支持多种不同类型的数据采集与标注，提供多样化标注工具与灵活配置项目的通用界面。流畅的标注流程与高效的项目配置机制不仅可覆盖各种复杂场景的数据采集与标注，更能保障高效率高质量地实现客户需求。

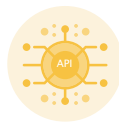
项目流程管理

除了便捷的“人机结合”标注工具，Stardust平台构建了成熟的项目管理系统，避免了许多供应商存在的项目生产过程混乱的问题。项目管理系统包括项目管理、用户管理、任务流管理、质量控制、工作统计、SaaS自动化和数据处理与管理机制，通过SaaS自动化实现了数据服务请求流程的自动与便利。



项目管理

项目管理系统，支持进度查看、人员管理、角色管理、错误统计、团队记录、操作项配置，极大提高标注效率



算法接入

算法接入管理系统，可以接入算法、脚本、API，可以管理大型任务的运行时

SaaS 自动化系统

SaaS自动化系统包括自动化报价、自动化考核上岗、自动化质量控制、数据验收和API接口文档。



SaaS自动化系统

SaaS自动化，支持整个标注、质检、上岗、系统对接的自动化，通过事件触发脚本，极大提高效率

① 自动化报价

报价自动化能够极大提高inbound客户需求的沟通效率，解决90%的沟通需求。通过交互式操作，能够让客户想清楚他们的需求细节，并直接获得他们想要的信息（项目需求、报价、工期）。

② 自动化考核上岗

客户创建题库项目后，平台自动创建考试项目，选取候选人进行上岗考试，最后考试记录保存于数据库中。如果通过，标注员才可以加入正式项目或团队。



③ 机器自动化辅助

包括预标注、实时辅助标注和算法自动辅助。通过这三个方式自动辅助标注员工作的流程，目的是提高标注准确率和效率。

④ 项目流程可视化

可随时进入项目管理页面查看实时项目进度；可导出项目Excel查看项目统计内容（统计当下整个项目完成情况）。

⑤ 数据验收

SaaS自动化系统中的数据验收主要特点为整个数据验收过程可视化和算法验收。

⑥ API接口文档

整个数据服务流程可以用API调用。

workflow 管理系统



任务流管理

领先的多复合工作流管理系统，支持标注、检查、抽查、验收、驳回、分发逻辑、高级流程逻辑、callback等功能



质量控制

质量控制系统，支持多重指标定义、标注对象级别的质检和错误记录，能以多重方式进行抽检和统计、回溯历史记录，确保输出合格的数据



用户管理

用户管理系统，支持角色配置、用户权限、人员管理、历史查看等功能

① 任务流管理

包括流转系统和分发逻辑、高级流程逻辑、callback等功能。确保题目根据不同的状态在不同的工作池子（标注池、检查池等）中按照项目配置的工作流程来回流转，以确保每道题都经过标注、审核、抽样、去重严格把关的流程，保证题目的正确率和完成效率。

② 质量控制

通过多层质检、系统质检和机器辅助质检等方式对数据服务过程进行质量控制。具体见产品优势-质量管控。

③ 用户管理

通过角色配置和人员管理为用户配置清晰的角色和职权。用户权限确定明确的界限范围，防止信息泄露；即时对反复进入某一敏感权限范围的ID进行相应的反应并处理。



工作统计系统

包括项目数据统计和工作量数据统计，支持20余种统计值，对每个人的工作量统计到标注对象层面，可以实时或分时段地对不同池子进行监控，可以自定义统计项和统计界面。



工作统计

工作统计系统，能够按多种层级、对象、视角统计，支持20余种统计值，并可以自定义统计项和统计界面

数据处理与管理

管理数据集、数据库、数据、访问权限、数据安全等方面的问题。以原始数据为主体，同时可以带着结构化的数据(optional)。



数据管理

数据管理模块，可以自定义接入外部数据库，可删选查看数据，并自动同步数据



结构化数据管理

标注结果可以映射入库、支持结构化查询检索、支持数据可视化



数据处理

简洁的结果显示方式，让标注流程和检验流程非常高效。包括高亮、隐藏、锁定、预览、属性弹出等功能



数据安全

通过水印、不可复制、数据权限、访问控制、数据加等方式保证数据安全

① 数据管理

数据管理模块，可以自定义接入外部数据库，可删选查看数据，并自动同步数据。

② 结构化数据管理

标注结果可以映射入库、支持结构化查询检索、支持数据可视化。

③ 数据处理

支持预处理脚本、入库脚本和数据流转通知。

④ 数据安全

Stardust数据安全方案结合相应的法律措施、系统措施、数据安全措施和安防措施来保障用户的数据安全，具体可见产品优势-数据安全。



CHAPTER 11

无人车闭环解决方案

星辰数据Stardust是业界领先的数据标注及AI数据中台解决方案供应商，我们基于多年在无人车领域数据采集和标注的经验，致力于打通无人车训练的闭环。

星辰协同业界生态合作伙伴，为无人车研发提供从改装车、数据采集、数据标注、仿真测试到真车测试的一站式解决方案。

我们来负责无人车开发中的标注与落地，让您的团队可以专注于最重要的地方——模型开发。

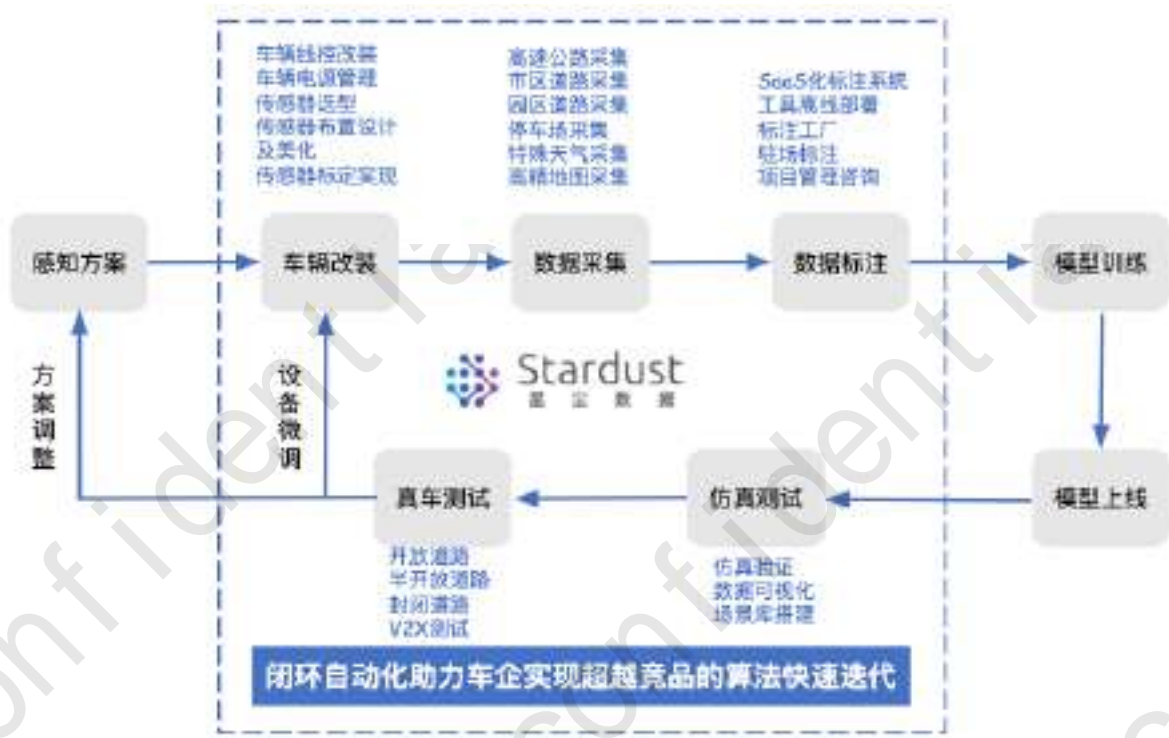


行业痛点

缺乏行业一站式解决方案，从改装、数据采集、数据清洗、数据蒸馏、数据标注、前置算法、仿真系统测试到测试场真车测试无法实现闭环，耗费大量的时间、人力成本：

1. 车辆电气化改装难度大、定制化需求多且必须满足安全性需求
2. 数据采集需求极大、且需要前置算法进行知识蒸馏以得到有效数据
3. 数据标注工具专业化程度高，标注类别多，对应的产品功能点繁杂，尤其是针对点云数据，需要优质的工具才能满足数据的标注需求
4. 人员管理难：安全员、车辆运营、政府关系、公安报备等都需要专业的人员
5. 封闭道路、V2X测试需要场地及物理基础设施，封闭测试区场地建设水平不同，场景搭建情况差异大，真正
8. 各地测试规程不一致，示范区协同存在壁垒；测试结果未实现完全互认，重复测试增加企业负担

闭环解决方案



- 车辆改装：车辆线控改装、车辆电源管理、传感器选型、传感器布置设计及美化、传感器标定实现
- 数据采集：高速公路采集、市区道路采集、园区道路采集、停车场采集、特殊天气采集、高精地图采集
- 数据标注：SaaS化标注系统、工具离线部署、标注工厂、驻场标注、项目管理咨询
- 仿真测试：仿真验证、数据可视化、场景库搭建
- 真车测试：开放道路、半开放道路、封闭道路、V2X测试



车辆改装

支持L2-L5级别自动驾驶的改装

- 车辆线控改装
- 车辆电源管理
- 传感器选型
- 传感器布置设计及美化
- 传感器标定实现



Velodyne激光雷达



Continental毫米波雷达



Leopard相机



Ublox导航模块



Xsens IMU



Bosch超声波

支持传感器（硬件安装及驱动开发标定）

- 激光雷达(Velodyne, Ibeo, 禾赛, RS)
- 毫米波雷达(Bosch, Continental, Delphi)
- 相机(Bosch, Mobileye, Leopard)
- 导航模块(Ublox, Novatel)
- IMU(Xsens, Aceinna)
- 超声波(Bosch, Yis)

定制化改装服务

- 车辆电源系统改制
- 车辆散热系统改制
- 车辆外饰美化
- 交互灯光改制





仿真测试

运行流程：

1. 导入自建的场景，或使用平台预建的典型场景；
2. 按类别或手动选取场景库文件，或选择待测模块的特定版本进行测试；
3. 场景测试会自动运行，结果会在网页或本地平台显示；
4. 具体的测试指标可以下载并详细分析。未达标的测试用例会反馈给研发部门进行进一步验证。

ID	名称	创建时间	更新时间	状态	操作
101	101	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
102	102	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
103	103	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
104	104	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
105	105	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
106	106	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
107	107	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
108	108	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
109	109	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
110	110	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
111	111	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
112	112	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
113	113	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
114	114	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
115	115	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
116	116	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
117	117	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
118	118	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
119	119	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除
120	120	2020-01-01 10:00:00	2020-01-01 10:00:00	成功	删除

结果列表



停车场避障

系统自有场景：

- 支持150多种典型工况的6W多个场景
 - 高速公路自主超车
 - 停车场避障
- 仍在持续扩建中

自建场景：

- 支持从路网（主路，匝道，隧道等）、交通参与物（车辆、行人等）、障碍物的交通行为（脚本控制，输入设备控制等）、硬件状态（传感器配置和故障录入等）等维度创建场景库的能力
- 充分兼容业内主流的OpenScenario格式



Python脚本编写



在线配置自动生成XML



随机交通流测试：

平台提供了随机交通流仿真的能力，用户可以通过设定地图/障碍物车行为方式/场景属性等参数生成长时间长距离的随机交通流仿真，并用与场景库一致的标准验证本车算法效果



可视化：

- 可以实现自动驾驶主流传感器，主流格式的数据可视化
- 支持感知、决策、定位算法中间结果的可视化
- 所支持的传感器原始数据：点云、移动障碍物、轮速计、相机数据
- 所支持的核心系统状态：规划路径、控制命令、融合定位、融合后车道线、障碍物融合

可视算法调试工具：

平台自带自动驾驶各算法调试工具箱，支持定位，点云层，障碍物属性，误差轨迹等的属性查看和可视化



多定位显示



定位属性、误差、车速



点云及影像



障碍物属性



真车测试

大多数的开放场景大多选择交通流量小的封闭测试场地周边城市道路，交通流量小、弯道、坡道、桥梁、隧道等覆盖不够典型，代表性有限。星尘为您提供全国最大的城市区域开放测试场地，并能够提供全套的配套运营服务。

- 场景类型：弯道、隧道、坡道、桥梁、十字路口、环岛、机非混行场景、水泥路等
- 道路类型：城市道路



开放测试场地



开放测试场地

封闭测试场地建设规划投入大，运营成本高，并且测试主体路试前期投入压力大。星尘为您提供场景类型全面涵盖的测试场地以及专业的运营人员：

- 场景类型：城市道路模拟区、高速道路模拟区、乡村/城郊道路模拟区、越野（无人战车）道路模拟区、涵盖弯道、十字交叉口、环岛场景等（部分建设中）
- 道路类型：高速、城市、乡村道路
- 安全员：50+名



安全员上岗培训



安全员上岗培训



CHAPTER 12

产品定价与定制化

典型任务指导价格

机器视觉 Computer Vision	语音转写 Audio Speech Recognition	OCR转写 Optical Character Recognition	3D点云 LIDAR AR VR	NLP Natural Language Processing
<ul style="list-style-type: none"> 2D框 ¥0.06/个 点 ¥0.02/个 多边形 ¥0.10/个 语义分割 ¥0.20/个 线 ¥0.10/条 立体框 ¥0.20/个 	<ul style="list-style-type: none"> 普通话 ¥3.00/分 中英混合 ¥5.00/分 方言 ¥6.00/分 小语种 ¥8.00/分 	<ul style="list-style-type: none"> 英文 ¥0.04/词 文章 ¥0.008/字 小语种 ¥0.25/行 公式 ¥0.10/条 手写体 ¥0.01/字 	<ul style="list-style-type: none"> 车辆 ¥0.16/框 行人 ¥0.20/框 其他障碍物 语义分割 ¥5.00/帧 	<ul style="list-style-type: none"> 实体提取 ¥0.10/条 判断 ¥0.20/个 分类 ¥0.10/个 排序 ¥0.10/个 问答 ¥0.20/个

企业化部署

标注系统部署包括：企业部署、SaaS平台和定制化开发。

企业级部署

支持企业级部署，包括云端部署和私有化部署，您都将获得最适合、最安全的服务。

```

stardust install
> downloading pre-requirements...
> pulling docker images...
> installing software...
> Done! Stardust is running on port: 80

stardust update
> checking for updates...
> Done! Stardust is running on port: 80

```

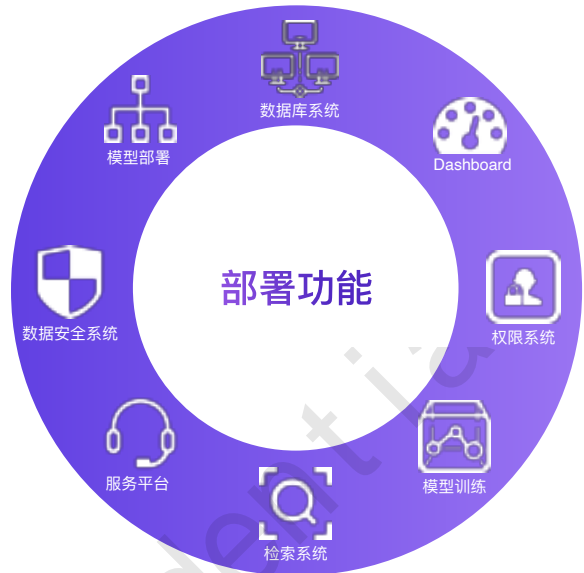


专属项目经理

我们为每个客户提供技术、开发和集成支持，并有专属的项目经理全程陪伴，软件部署的全流程都由星辰来帮您分担。

部署功能

除了标注工具、 workflow 管理，还包括模型部署、数据管理系统、Dashboard、权限管理、训练管理、结构化数据检索、数据安全、实时统计等功能。



产品定价

企业部署		
<h3>SaaS平台</h3> <p>¥ 1000 /人月</p> <ul style="list-style-type: none"> 使用在线平台操作 全量的标注工具 质量控制 人员管控 API接口，对接您的内部工具 数据验收平台 <p>选择</p>	<p>¥ 50万 元起</p> <ul style="list-style-type: none"> 包含SaaS平台的全部功能 部署到企业内部环境 一次性部署费用 20位标注员席位年费 免费升级和售后维护 <p>选择</p>	<h3>定制化开发</h3> <p>联系我们</p> <ul style="list-style-type: none"> 定制化的功能开发 定制化工具 定制化质量控制 定制化人员管控 定制化标注流程 <p>选择</p>



CHAPTER 13

合作伙伴





CHAPTER 14

成功案例

无人驾驶-禾多科技



HOLOMATIC
禾多科技



概述

为无人车获得更加清晰灵敏的眼睛

为了确保禾多的无人车跟踪和避障系统在各种场景和位置均能正常运行，他们的团队需要庞大而多样的数据集来训练模型。

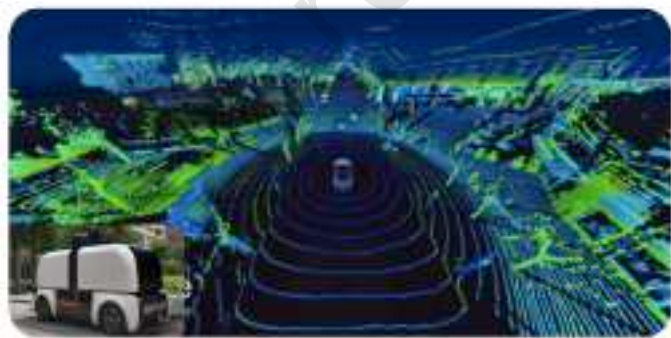


问题&解决方案

点云标注获取门槛高，价格高昂且有效数据少

问题：优质的点云训练数据需要成熟、完善的工具支持，而产品开发难度大，成本高；在训练数据需求突增的情况下很难迅速建立和管理一个专业的点云标注团队。

解决方案：星云平台支持点云连续帧标注，2D图像映射和语义分割；支持数据采集，同步上传以及数据清洗、格式转换。支持API，支持自动化；支持500人以上同时在线标注；准确度可达99.5%以上，可为客户节约50%以上成本；高效，再大量的数据也能在1-2个月内完成。



未来

更成熟的追踪与避障系统

更优越的全栈式无人配送解决方案和无人车常态化运营，为客户提供稳定、安全的城市区域配送服务。



客户赠言

韩总监
感知系统Leader
禾多科技

“星云的平台可以实现API化的数据验收、数据质量实时监测，输出高效、高质、精准且安全的数据；并且星云团队依靠在无人车领域的丰富标注经验可以给我们专业建议，这一点是难能可贵的。”



融媒体+人工智能



目标：打造国家级NLP/新媒体实验室

概述

为智能化算法训练提供优质语料资源，进而实现精准检索、关系挖掘、智能化应用等多种服务，实现大数据与人工智能相关技术在媒体行业的融合应用。



问题&解决方案

中文世界缺少高质量的NLP语料训练任务

问题：NLP自然语言标注主观性强，难度大，需要标注员具有较强的中文语言能力和较好的政治素养；项目前期，标注需求不够清晰，缺乏完善的标注规则体系；文章篇幅长，大标签数量22个，累积标签和选项达70-80个；文章类型多，场景多，包含9大场景，数据形式多，包含文字、图片和视频等。

解决方案：星尘自动考试系统，部署对应需求的考题设置标注门槛对标注员进行筛选，保证标注员的质量；对于不够完善的标注规则体系，项目配置的专属PM（项目经理），即时反馈标注过程遇到的标注问题并提供多个解决方案，推动优化完善标注规则体系；对于复杂的标注场景与需求，星尘系统自带的预标注与自动质检大大地提高了标注的效率与质量。



新闻即数据，数据即新闻

未来

新华社：“凭计算之力，求数据洞察，赋万物为媒，迎智能时代。”与AI结合的媒体大脑是未来媒体行业发展的方向。



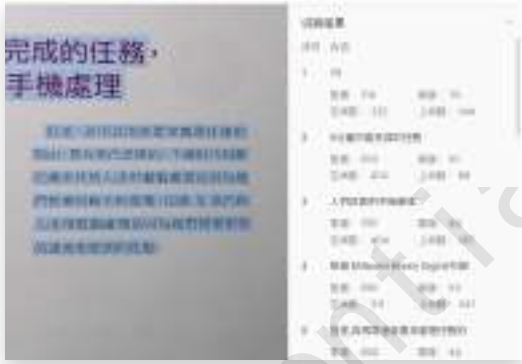
林老师
技术局主管

客户赠言 新华社

“星尘是我们在建设开拓融媒体国家重点实验室道路上的可靠合作伙伴，他们的系统具备高实用性、实践性，在标注层面用算法辅助大幅提高效率；项目人员的新闻敏感度和政治素养使得他们高质量地完成新闻稿件标注工作。”



OCR与ASR



概述

目标：让机器能说会写，训练更灵敏的OCR模型和ASR模型

为了实现这一目标，百度需要庞大而多样的数据集来训练他们的模型。他们的团队需要采集和标注稀缺小语种图片、语音以及手写体。



对于机器听说读写能力的提升，数据量是极大的瓶颈

问题&解决方案

问题：稀缺语言数据采集困难。由于该项目主要是对稀缺语言进行图片、语音以及手写体的采集，因此需要懂该稀缺语言的人在当地进行收集；此外，由于地域问题以及采集需求复杂，数据的采集与收集方式成为一个比较大的问题；最后，多形式与多种复杂场景的数据给质检带来比较大的难题。

解决方案：星云数据通过调用在全球的采集资源，寻找最适合的采集伙伴对稀缺语言进行采集；在短时间内为了适配采集和标注的需求开发更加适合的软件以满足采标的需求；通过多层质检和动态监控保障稀缺语言标注的质量。



未来

更精准的OCR系统和ASR系统

星云数据采集及标注系统可以帮助百度训练更精准的OCR系统和ASR系统，未来将能服务于更多的场景，在不同场景下提高效率。



客户赠言

范经理
语音研发主管
百度AI

“在过往合作中，星云能够支持我们定制化程度超高的数据需求，并依靠其丰富的海外资源，综合采集、标注、质检、输出等环节，构建专业的一站式数据服务解决方案。”



Stardust
星 尘 数 据

CONTACTS



地址

北京市朝阳区
尚都SOHO
北塔b座



电话/邮箱

T. +010-8646 2864
E. hello@stardust.ai



网址

<https://stardust.ai>

