

AI 数据服务

数据收集与数据注释



Lionbridge 提供高质量的数据收集和注释服务,为人工智能 (AI) 发展提供强力支持。我们的全球众包团队提供多样化、符合各地文化背景的数据集,助力 AI 模型在不同语言、地区及真实场景中实现精准表现。



使用原因及常见用户

为什么您需要使用优质数据的收集和注释服务？

AI 模型的性能取决于其训练数据的质量，而许多企业难以获取足够规模、质量和多样性的数据，因而无法开发准确、可靠的 AI 系统。对此，Lionbridge 着意在以下方面提供支持：

1

数据稀缺

大多数组织缺乏内部资源来收集所需的海量、多样化数据集 — 尤其是在面对非典型案例、代表性不足的地区或不常见的语言等情况时。

2

数据质量和一致性

标注质量差或不一致的数据会导致模型存在偏差、准确率低下以及部署失败。我们专业的注释流程和严格的质量保证 (QA) 机制，从项目初期即可确保数据的高精度。

3

多语言和多模态需求

从语音助手到计算机视觉，AI 系统必须能够在多种语言、格式和环境下有效运行。而我们大规模提供符合文化背景、多语言、多模态的数据集，可为其运行提供强力支持。

4

速度和可扩展性

组建内部数据团队不仅耗时长，成本也高。而我们能够通过快速启动、全球扩展和支持快速迭代周期，助力您的模型更快应用于生产。

5

融入人工介入 (HITL)

即使是最先进的模型也需要人工验证。我们的服务通过人工审核的训练数据实现闭环，能够持续提升模型性能。

哪些用户通常选择数据收集和注释服务？

各行业致力于构建或扩展 AI 应用的服务支持团队。常见用户包括：



机器学习 (ML) 和数据科学团队

此类用户需要大规模的优质数据集, 用以对模型进行训练、验证和微调。



AI 产品和工程团队

此类用户依赖于精确标注的数据, 以推出语音助手、计算机视觉工具和搜索算法等功能。



神经语言程序学 (NLP)、计算机视觉 (CV) 和多模态 AI 团队

此类用户需要经注释的专业化数据 (文本、音频、图像、视频) 以执行特定任务, 例如实体识别、情感分析、对象检测或转录。



医疗、金融、汽车和零售公司

此类用户利用 AI 执行诊断、欺诈检测、自动化、个性化等任务, 每项任务均依赖于特定领域的准确训练数据。



正在开发生成式 AI (GenAI) 的初创企业和大型企业

此类用户依赖于提示词数据、合成的输入内容、精细调优的训练数据集以及人工评估的输出结果, 以构建具有竞争力的 AI 应用程序。

我们的服务



数据注释



数据收集



模型评估

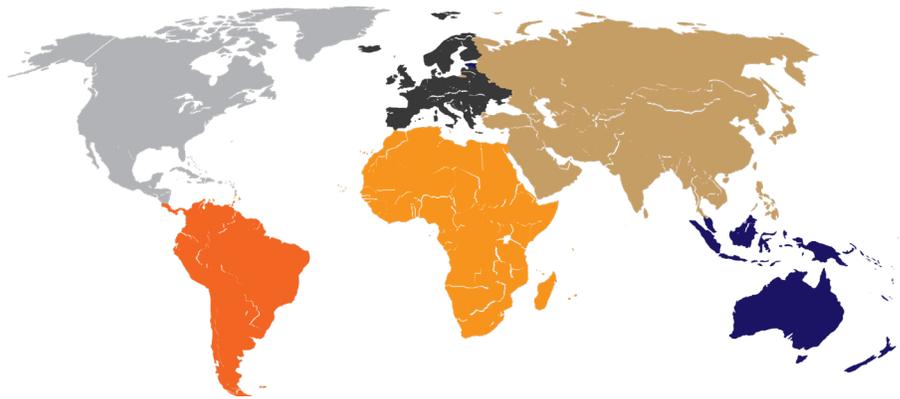
众包

我们的众包服务成就您的竞争优势

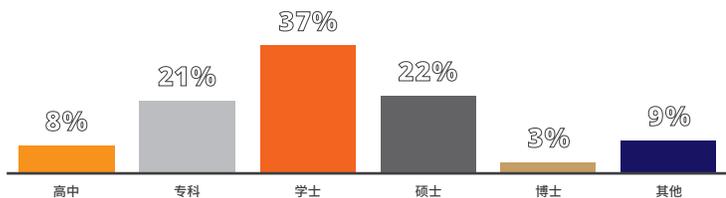
Lionbridge 的全球众包团队融合了普通用户与主题专家,为您提供兼具规模性与专业深度的数据支持。无论您是在训练 AI 模型,还是注释复杂的特定领域数据,我们都能为您的任务搜寻最合适的人员。依托专职项目经理和完善的 QA 工作流程,我们可确保交付高质量、符合文化背景的数据,紧密贴合真实应用场景,从而提升模型性能。

这需要一个拥有高度专业化、多元化专业知识与兴趣的全球社群。

- 北美:26%
- 欧洲:29.5%
- 亚洲:27%
- 非洲:6.5%
- 南美:8.5%
- 大洋洲/澳大利亚:2.5%



最高教育水平



性别



年龄



60万
全球社群成员



350+
语言



450+
地区

行业



自建还是采购?单打独斗的成本与正确行事的回报:真实的投资回报率案例。

自行组建内部数据运营团队可能会在人力、工具和时间上耗费数百万,且无法保证数据质量和规模。一家全球劳动力分析公司需要注释 30 种语言的劳动力市场数据,每种语言需要 1,200 小时的工作量 — 注释任务总计时长达 36,000 小时。

预计的内部方案:

- 语言成本:约 180 万美元
- 招聘周期:招聘 30 名语言专家并完成入职,耗时 3 至 4 个月
- 项目工期:由于项目启动和分阶段交付,预计耗时约 6 个月以上
- 训练数据就绪时间:总计约 8 个月以上,之后数据才可用于训练模型

Lionbridge 的方案:

- 按需调用的全球众包团队,包括具备资质的语言专家及双语注释人员
- Lionbridge 项目经理 (PM) 全程管理,配合多层次 QA 流程
- 项目数周内启动,而非数月
- 训练数据在 12 周内准备就绪,使模型开发进程加快数月
- 交付速度提升 40%,相比内部实施节省超 100 万美元

关键洞察

- 通过减少数据净化和返工所耗时间,模型部署速度提升 30-50%
- 得益于我们的全球贡献者网络和高效基础架构,与自行组建内部数据团队相比,成本降低 50-60%
- 投入生产速度加快 3 倍,避免了冗长的招聘、工具配置和培训周期
- 减轻内部负担,释放您的 ML 团队,使其专注于模型开发,而非繁琐的数据处理
- 我们让 AI 数据生产更具可扩展性和成本效益,强化结果导向。您将获得切实成果,而无需额外开销。

