



模型，扩大参保人群范围，让“带病体”保险的商业模式更可持续。

数据与 AI 双轮驱动

商保协同创新之外，上海临转院在数据挖掘和人工智能方面还有诸多亮点成绩。

高质量多模态语料知识库建设，以“服务产业需求”为核心，聚焦关键疾病领域的标准化。段琦介绍，语料知识库首先围绕申康重点推进的肿瘤综合诊治（COC）、心脑血管疾病、重症、儿科四大方向布局，实现常见高发重大疾病的全面覆盖；具体建设逻辑是“先定标准，再扩范围”——先与各疾病领域的牵头医院合作，由牵头医院联合全国知名专家共同制定数据标准，明确关键核心字段、模态及内容的覆盖范围，确保最终语料数据的质量。

针对不同行业的差异化需求，语料中心提供“定制化数据服务”：对商业保险，提供大队列人群的疾病发生率、治疗费用等统计数据，无需标注，仅需核心字段以控制成本；对医疗 AI 企业，联合医院组织高水平医生对数据进行标注、加工，如提供医学影像、病理图像的专家标注，满足模型训练需求；对生物医药企业，则提供基因测序、蛋白质组学、代谢组学等数据，即使是几十人的小规模队列，例如具备某些共性的肿瘤患者的多模态多组学数据，只要内容和维度全面准确，也能够支撑诸如靶点发现、药物设计等行业研发需求。

目前，该项目已获得上海市经信委“高质量产业发展专项”的支持，已经全面梳理了 1500 万余篇国内外医疗文献数据，完成对美国、中国、日本、欧盟、澳新等地共近百万项临床试验数据的解析，率先实现 8 个肿瘤疾病方向的文献及临床试验数据筛选建设。未来，临转院将联合更多医院扩大语料库规模，打造全国顶尖甚至全球领先的医疗数据体系。

在 AI 赋能临床科研领域，上海临转院的代表性成果是“杏林智研 X-Scholar 临床研究大模型”。段琦介绍，当前医疗领域的知识更新呈爆发式增长，每天新增数千篇临床医学文献、临床研究方案及专家指南，医生难以全部研读，更难快速消化吸收转化为临床诊疗的经验。“杏林智研”正是为解决这一痛点而生。

该大模型的核心功能分为两部分：一是“精准知识推送”，通过微信小程序向注册医生个性化推送感兴趣临床知识内容。根据医生关注的领域，每日筛选新增的文献、指南，确保医生第一时间掌握前沿动态。二是“科研与诊疗辅助”，以疾病为单位梳理知识体系，帮助医生快速设计临床实验方案，撰写高水平研究报告。

更重要的是，它正逐步向临床诊疗延伸：当医生遇到疑难患者时，大模型可梳理相关文献中记载的类似病例及诊疗方法，为医生提供参考。段琦表示，“杏林智研”的目标是成为医生的“智能知识助手”，打通“知识获取 - 科研转化 - 临床应用”的链路，提升科研与诊疗效率。

上海临转院的“医疗大数据与人工智能概念验证平台”，旨在解决医疗 AI “需求与技术脱节”的问题。段琦说：医院临床工作存在大量痛点，如疾病诊断任务繁重、智能辅助工具功能有限，各类管理工作耗时等问题，但缺乏技术团队对接；高校、企业有强大的 AI 技术，却找不到合适的临床场景。“概念验证平台”为二者搭建了对接的桥梁。

它的运作模式是：医疗专家提出临床问题，并提供少量数据用于算法验证；高校、企业等技术方提交解决方案，双方双向选择后，在 3-6 个月内快速研发；研发成果后期在多家合作医院验证，如 A 医院提出的问题，需在 B、C 医院测试普适性，避免“仅适用于单家医院，不具备普适泛化性”的局限。整个过程中，医疗专家与技术导师共同指导，成功后形成“利益共同体”，共同孵化新公司或推动技术转化。这一模式在全国范围内具有创新意义。■