项目编号:

**暨南大学大学生创新创业训练计划**

**项目申报书**

项目名称 法芯智能—智能AI驱动的高效法律服务平台

项目类型  创新训练项目

项目类别  人工智能+法律

项目负责人  陈炜昊

学 号  2023102145

学院专业  网络空间安全学院/网络空间安全

联系电话  15976639950

E-mail  2426923853@qq.com

指导教师  古天龙

是否内外招学生联队 ☑是 □否

暨南大学教务处制

二О二四年十二月

**填 写 须 知**

一、项目类型分类说明：

1.创新训练项目是本科生个人或团队，在导师指导下，自主完成创新性实验方法的设计、研究条件准备和项目实施、研究报告报告撰写、成果（学术）交流等工作。

2.创业训练项目是本科生团队，在导师指导下，团队中每个学生在项目实施过程中承担一个或多个具体的角色，通过编制商业计划书、开展可行性研究、模拟企业运行、参加企业实践、撰写创业报告等工作。

3.创业实践项目是学生团队，在学校导师和企业导师共同指导下，采用前期创新训练项目（或创新性实验）的成果，提出一项具有市场前景的创新性产品或者服务，以此为基础开展创业实践活动。

二、项目类别分为一般项目和重点支持领域项目两类。其中，重点支持领域项目下设两个研究方向，分别为“基础学科、新工科、新医科、新农科、新文科”方向和 “人工智能+”方向。

三、申报书请按顺序逐项填写，填写内容必须实事求是，表达明确严谨。空缺项要填“无”。

四、填写时可以改变字体大小等，但要确保表格的样式没有被改变；填写完后用A4纸张双面打印，不得随意涂改。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 项目名称 | | 法芯智能—智能AI驱动的高效法律服务平台 | | | | | | |
| 项目类型 | | 创新训练项目 | | | 是否内外招学生联队 | | ☑是 □否 | |
| 项目类别 | |  | | 研究方向 |  | | 项目研究周期 |  |
| 负责人情况 | 姓 名 | | 陈炜昊 | 性 别 | 男 | 平均绩点 | 3.11 | |
| 学 院 | | 网络空间安全学院 | 专 业 | 网络空间安全 | 生源 | ☑内招 □外招 | |
| 联系电话 | | 15976639950 | QQ号 | 2426923853 | Email | 2426923853@qq.com | |
| 项目  组成员  （不含负责人） | 姓 名 | | 性别 | 学号 | 学院 | 专 业 | 生源情况 | |
| 苏恒毅 | | 男 | 2023102696 | 网络空间安全学院 | 网络空间安全 | ☑内招 □外招 | |
| 罗尹诞 | | 女 | 2022104514 | 法学院/知识产权学院 | 法学 | ☑内招 □外招 | |
| 林歆婕 | | 女 | 2022104603 | 法学院/知识产权学院 | 法学 | ☑内招 □外招 | |
| 李东骏 | | 男 | 2024151951 | 法学院/知识产权学院 | 法学 | □内招 ☑外招 | |
| 指导老师情况  (第一导师) | 姓名 | | 古天龙 | 学科  专业 | 网络空间安全 | 所在单位 | 网络空间安全学院 | |
| 职称/职务 | | 教授、博士生导师 | 学历  学位 | 博士 | 联系电话 |  | |
| 校外导师情况（创业实践项目必填） | 姓名 | |  | 所在单位 | |  | | |
| 职称/职务 | |  | 学历  学位 |  | 联系电话 |  | |
| **一、项目简介** （一）项目简介：AI赋能的湾区法治协同引擎 近年来，中国大力推进法治社会建设，法律与我们每个公民的日常生活紧密关联，公民的法律意识也显著增强，然而，具备法律意识并不意味着可以在面对法律问题时能有效应对，普通民众常常因为信息检索能力不足、法律法规庞杂而难以精准获取有效指导，尤其是在法律纠纷中，当事人与律师之间的信息差更易导致沟通成本激增。无独有偶，尽管有着一定法律基础的法学学生也有着自身的法律痛点，在法律文献检索与实务研究中也面临资源分散、效率低下等问题。而一旦涉及不同法域等稍加复杂的法律问题时无疑是雪上加霜，以粤港澳大湾区为例，其作为国家战略发展区域，法律体系融合了内地与港澳的不同法系，是典型的地方法规复杂、跨区域法律适用难题。  与此同时，现代社会在精神文明高度建设的同时，数字科技也在充分发展。法律服务的智能化转型势在必行，与此同时，全球AI技术也正在加速向垂直领域渗透，法律科技（LegalTech）市场规模年均增长率超25%，AI驱动的法律咨询、文书生成、案件预测等应用已成为国际趋势。我国《新一代人工智能发展规划》明确提出“推动人工智能在法律服务等民生领域的深度应用”，为法律和AI 融合提供了政策支撑和技术落地的战略机遇。  法芯智能，一款由暨南大学法学院/知识产权学院、信息科学技术学院以及网络空间安全学院的师生团队组成核心管理团队，基于知识图谱与人工筛选构建的高质量法律数据集语言大模型，也是国内首个聚焦粤港澳大湾区的法律AI 服务平台，产品以“精准高效、大模型、本土化”为核心架构，面向普通民众提供法律咨询和科普服务，面向高校学生打造专业文献检索和法规背景查询工具。通过AI技术赋能法律场景，构建“法律知识库+智能交互”双引擎，致力于打造“全链条、多维度、深融合”的智能法律服务体系，突破传统法律服务的三大瓶颈：  1.破解湾区法律协同难题，构建“三法域”智能适配系统   1. 特色一：大湾区法规动态融合引擎。整合内地部门法、香港普通法、澳门大陆法系等差异化法律资源，基于知识图谱构建跨法域关联规则，实现“一案三地”法律条款智能对比与适用建议。针对跨境纠纷，系统可自动匹配相关地域法条，解析管辖权差异与诉讼策略。 2. 特色二：地方政策实时追踪与预警。接入广东省及湾区政务数据接口，动态更新自贸区条例、跨境投资指引等区域性政策，为企业和个人提供“政策合规性扫描“功能，降低制度性交易成本。   2.AI赋能法律服务全场景，打造“预-诊-解”闭环生态   1. 创新点一：证据链驱动的“判决预测2.0”。突破传统法律AI仅依赖案情描述的局限，首创“多模态证据智能分析系统”：用户可上传合同文本、录音录像、电子存证等材料，AI通过OCR、语义分割、时间轴还原等技术，自动构建证据效力图谱，结合最高人民法院指导性案例库，预测案件胜诉率与赔偿区间。 2. 创新点二：学术研究与实务应用双轮驱动。面对高校法学生推出“文献智析”功能，基于框架定制法律文献分析模型，自动提取裁判文书中的法律争议焦点、学术论文中的理论演进脉络，并生成可视化知识网络，大幅提升研究效率。   3.产学研用深度融合，构筑法律科技生态圈  本项目以“数字赋能司法，数据驱动公平”为核心理念，通过产、学、研、用四位一体协作，构建覆盖法律科技全链条的创新生态体系。首先是产学研协同机制，打破边界，共享成果：一方面是产业资源整合，合作搭建法律垂直领域大模型训练平台，共享算力资源和联邦学习框架，努力提高模型训练效率；另一方面是学界智库支撑，联合法学研究所成立“跨境法律数字化研究中心”，力求填补法律科技领域标准空白；  然后是技术转化路径，从实验室到大环境的闭环验证：一方面建立“用户反馈-模型迭代-场景拓展”正循环，每日处理用户咨询产生的合规数据，通过区块链确权后形成国内首个法律AI训练数据集，与此同步进行场景化验证体系和商业化落地模式，前者与法院、律所等开展试点，实现案件要素自动提取，辅助法律关键词提取效率。  综上，本项目聚焦粤港澳大湾区法治建设前沿需求，立足三地法律体系衔接的时代课题，深度融合人工智能技术与法律专业智慧，打造湾区法律智能服务平台。依托多法域知识图谱与多语言交互技术，为湾区企业及居民提供精准高效的法律咨询、合同审查和风险预警服务，助力破解区域协同发展中的制度壁垒，以科技赋能法治融合，为营造国际化、法治化营商环境贡献智慧力量。 | | | | | | | | | |
| **二、立项依据**  （一）研究意义  本项目组成员市场调研过程中，发现目前法律AI市场上主要存在以下四大痛点：  1.法律解读准确性不足  复杂性与地域差异:法律条文体系庞杂，不同地区法规存在差异，AI可能因训练数据局限或算法缺陷导致解读错误。  上下文理解局限:用户问题的背景信息可能涉及多因素交织(如情感、文化等)，AI难以精准捕捉深层需求，易给出模糊或片面的建议。本产品的模态数据预处理管道与文档结构识别模块（LayoutLMv3），能对用户输入的内容进行更加准确的解析有助于精准理解用户需求；法律术语知识图谱（Neo4j）与领域检索增强生成（RAG）的引入有助于根据用户需求更准确地匹配对应的信息生成高质量的回答，避免片面建议。案例推荐则提供完全真实的历史案例，不仅能为用户提供参考，也能“借助用户的自行判断”，避免不准确的直接回答。  时效性挑战:法律更新频繁，若模型未能实时同步最新条款或判例，可能提供过时答案，增加用户风险。本产品联邦学习支持模块的引入使得大模型的训练数据“来源更广”且能“实时更新”，适应了法律条文不断更新的实际情况。  2.数据隐私与安全风险  用户咨询常涉及个人隐私、商业秘密等，若平台数据保护措施不足(如加密不严、第三方共享)，易引发信息泄露。我们采用严格的数据加密和访问控制措施，确保用户数据在传输和存储过程中的安全性。合规性校验模块的引入能对用户输入和模型输出进行审查，避免模型输出包含个人信息的敏感内容、越权建议或其他违规内容。  3.用户认知与表达偏差  普通用户常因法律知识匮乏，无法精准描述问题(如混淆“劳动仲裁”与"民事诉讼”)，导致Al误判需求方向。用户可能混淆法律术语，导致 AI 难以准确判断需求。为此，我们产品引入的多模态数据预处理管道和文档结构识别模块（LayoutLMv3）能够对用户输入的内容进行更准确的解析，帮助模型理解用户的真实意图。  4.个案指导不足  现有法律AI工具通常提供一般性的法律信息，缺乏对用户特定案例的深入分析，难以满足用户对自身案件结果预判的需求。我们的模型基于类案检索、深度学习模型和用户交互引导，当模型识别到用户对特定案例（如自身正在面临的法律问题）寻求指导时，系统会引导用户提供更多判案所需的详细信息。随后，模型利用数据库中的历史案例，模拟法院判决过程，提供对该个案的结果预判。具体来说，模型将从法律案例数据库中检索与用户提供的信息相似的历史案例，确保参考的案例具有相关性和代表性；设计了智能交互机制，引导用户提供必要的案件细节，以提高结果预判的准确性；并利用深度学习技术，分析和理解复杂的法律文本和案例细节，模拟法院的判决过程并给出个案结果预判，为用户提供参考。  综上，本项目聚焦粤港澳大湾区法治建设前沿需求，立足三地法律体系衔接的时代课题，深度融合人工智能技术与法律专业智慧，打造湾区法律智能服务平台。依托多法域知识图谱与多语言交互技术，为湾区企业及居民提供精准高效的法律咨询、合同审查和风险预警服务，助力破解区域协同发展中的制度壁垒，以科技赋能法治融合，为营造国际化、法治化营商环境贡献智慧力量：  1.推动法治社会建设：通过智能法律服务平台降低公众获取法律服务的门槛，提升公民法律问题解决效率，助力国家法治化进程。  2.解决大湾区法律协同难题：聚焦粤港澳大湾区三地法律体系差异，构建跨法域智能适配系统，促进区域法治协同与制度衔接。  3.技术创新与产业融合：结合AI技术与法律专业知识，探索法律科技（Legal Tech）的垂直领域应用，推动法律服务业智能化转型。  4.促进普惠法律教育：为高校学生提供法律文献智能检索工具，为普通民众提供法律科普服务，推动法律知识普及。   1. 国内外研究现状和发展趋势   1.1国外研究现状 美国在大模型技术研发和商业化应用方面处于全球领先地位。例如：OpenAI的GPT系列：通过微调技术，已被应用于法律合同审查、法律咨询等场景。例如，摩根士丹利利用GPT优化财富管理咨询流程，提升法律文本分析的效率；Google的PaLM：在司法判决预测领域取得突破，通过分析历史案例数据辅助法官评估证据链的完整性。 欧洲的研究机构（如DeepMind）探索将视觉-语言模型应用于法律文件的多模态解析，例如自动识别合同中的关键条款与风险标记。 2025年复旦大学联合国际学者发布的《大模型安全技术综述》指出，大语言模型在法律领域的攻击风险（如越狱攻击、提示注入）成为研究重点，需结合安全对齐技术提升鲁棒性。  1.2国内研究现状  国内已经有多个模型落地实现开发如得理法律大模型和通达海的法律大模型等，已取得技术突破与学术成果如下：  四阶段认知框架：上海算法研究院提出基于注意力头的KR-ICI-LR-EP四阶段推理机制，模拟人类法律专家的思维过程，提升模型的可解释性。  模型安全与伦理研究：复旦大学团队在《Safety at Scale》综述中强调，国内法律大模型需解决数据隐私泄露和伦理对齐问题，例如通过联邦学习技术保护敏感司法数据。  行业应用案例  政务领域：北京大模型应用于政协提案的智能分析，实现法律条款的快速匹配。  金融领域：平安科技利用法律大模型优化信贷合同的风险识别，降低违约率。  2发展趋势  2.1技术优化方向  知识增强与多模态融合 通过引入法律知识图谱和外部数据库（如裁判文书网），结合RAG技术提升回答的权威性。例如，Meta提出的KI-NLP框架通过动态检索增强法律推理能力。  模型轻量化与边缘计算 针对法律场景的高实时性需求，华为、腾讯等企业探索小型化模型（如TinyLLM），在本地设备部署法律问答系统，降低云端依赖。  安全与可信技术 防御越狱攻击和提示注入成为重点。例如，OpenAI开发的“宪法AI”框架通过多轮对齐机制约束模型输出，确保法律建议的合规性。  2.2挑战与瓶颈  数据与算力限制 高质量法律数据的标注成本高昂，且涉及隐私保护。例如，中国裁判文书网的公开数据仅覆盖部分案例，难以支撑全领域训练。  模型幻觉与逻辑缺陷 法律大模型在复杂推理任务中易产生错误结论。例如，某测试显示模型在劳动法领域的回答错误率达15%，主要源于训练数据的时序性偏差。  伦理与监管风险 生成内容的权责归属不明确，可能引发法律纠纷。欧盟已出台《人工智能法案》，要求法律AI需通过透明性认证。  （三）现有AI法律平台相关竞品分析   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **产品名称** | **得理法搜** | **法行宝** | **合同嗖嗖** | **万象法律大模型** | | **主要功能界面图** | c85d29cdca513de87d2f1debc6aa7f9 | ee8e746f54e13a13dc1d056865682ee | df1748592f7164a28b79c0b0e972c32 | 64dbb73a661f546b7df1710c9ea6d00 | | **产品功能类型** | 案件检索类 | 法律问答类 | 合同相关类 | 专业领域类 | | **商业模式** | 付费会员模式；定制化服务收费；与其他企业或机构建立合作 | 免费提供法律咨询服务，为其他相关行业如法律广告营销、法律培训课程推广等拓展盈利渠道 | 基础功能免费+增值服务收费；定制化服务收费 | 付费会员模式；项目制合作，收取解决方案费用 | | **技术模型** | 得理法律GPT大模型 | 以文心大模型为基础，结合知识图谱，运用自然语言处理技术 | 自然语言处理技术+深度学习模型+知识图谱技术 | 以通用大模型为基座，运用法律知识图谱进行知识增强 | | **优势** | 法律数据全面丰富，且实时更新；产品功能强大 | 免费提供基础服务；用户可通过手机或电脑、iosAPP使用，操作简便 | 能快速生成专业合同文本，高效便捷，专业准确 | 应用场景广泛；数据安全有保障 | | **劣势** | 不同地区、不同法律领域的特殊性和差异性较大，产品适配性面临挑战 | 处理复杂法律问题的能力不足；存在信息不准确或不完整的 | 特殊情况处理不足，同时，缺乏深度沟通 | 技术迭代成本和压力较大；复杂案件处理问题存在局限 |   **图1 专门功能类AI法律工具产品竞品分析图**  目前市面上确有繁多AI法律产品，市场竞争相对激烈，“法芯智能”产品在结合上述产品优势的基础上，且被赋予有其独特的产品特色，包括证据分析和判决预测等。 | | | | | | | | | |
| **三、项目方案**  （一）研究内容   1. 数据资源整合与知识图谱构建：收集粤港澳三地法律法规、案例数据库，建立动态更新的法律数据集。利用自然语言处理（NLP）技术构建多法域关联知识图谱。 2. AI模型开发与功能模块设计：开发多模态证据分析系统，支持合同文本、录音录像等材料的智能解析与证据链构建。训练法律大模型，实现法律文书自动生成、案件胜诉率预测等功能。 3. 平台功能实现与验证：设计用户交互界面，开发法律咨询、政策合规扫描、文献智能检索等核心功能。联合律所开展试点应用，验证系统在跨境纠纷、企业合规等场景中的实际效果。 4. 产学研合作与生态建设：与法学院、律所共建“跨境法律数字化研究中心”，制定法律科技领域标准。构建联邦学习框架，共享算力资源，优化模型训练效率。   （二）研究目标  1.构建多法域知识图谱：整合粤港澳三地法律资源，通过知识图谱技术实现跨法域法律条款智能对比与适用建议。  2.开发AI驱动的全场景法律服务平台：包括法律咨询、合同审查、风险预警、案件预测等功能模块，覆盖“预-诊-解”全流程服务闭环。  3.实现多语言交互与本土化适配：支持粤语、英语、普通话多语言法律问答，解决大湾区跨境服务语言障碍。  4.形成产学研用生态闭环：联合律所、高校、企业构建法律科技生态圈，推动技术成果转化与商业化落地。   1. 技术路线   1. 数据集构建与预处理  1.1多源法律数据采集  指由于法律体系差异，不同地域（如内地的不同省份、香港、澳门）在法源、司法程序等方面存在明显差异。这种现象称为多法域。选用的数据范围包括：从内地各个省份、香港、澳门等多法域采集法律法规文本、判决文书、合同范本、学术论文与案例库。还包括法律文档扫描件、庭审音视频等非结构化数据。  1.2地域化标注与分类  标注（Annotation）是指在数据上添加人工或自动化标签的过程，以便在训练时让模型能够学习到更丰富、准确的特征信息。我们在数据清洗过程中，根据文档或案例的来源地、适用法域（如《民法典》内地适用、《基本法》港澳适用）添加地域标签；对于港澳跨境法律文件则进行“多区域”标记。  我们还将对数据进行分层次标注。分层次标注指的是对每份文档进行多级标注，包括以下内容。  法律条文标注：提取并标记具体法条编号及其生效时段；  案件类型标注：如民事案件、刑事案件、行政案件等；  地域特性标注：标明适用的地方法规或跨区域协议；  领域术语标注：如“港澳普通法原则”“大陆法系”“合同法争议焦点”等，便于模型针对性学习。  在使用上述方法进行标注后，同一文档上可能标记多个维度，如法域、条文、案件类型等。称为多级标签（Multi-Level Labeling）。  1.3数据清洗与脱敏  我们将对数据进行去重与去噪和隐私信息保护，剔除重复、格式错误、过时失效的法律文本，统一编码格式；对当事人姓名、身份证号码、企业财务数据等敏感信息进行模糊化或替换，避免在训练集或线上服务时泄露个人隐私。  1.4结构化处理  利用OCR（光学字符识别）与ASR（自动语音识别）对扫描件、音视频进行文字转录，结合文档结构识别（LayoutLMv3）解析文档的段落、表格、印章、手写批注等布局信息，转化为可被模型理解的统一文本格式。  2. 大模型训练与微调  2.1基础模型选型与初始化  选用目前最先进的高质量中文大语言模型DeepSeek-R1，导入公开领域的预训练权重。  2.2地域化与不同地区的适配训练  训练大模型进行多任务学习（Multi-Task Learning, MTL）。在统一模型结构中同时学习多种法律任务（法规检索、判例引用、合同审查等），并针对不同地域标签进行差异化训练，使模型具备“地域针对性”。在训练时，通过地域标签或地方法规特征向量，引导模型对来自内地、港澳等不同法域的数据进行差异化学习，以提升地域化回答准确度。训练过程如下。  (1)领域初步适配  将前三层layer层的维度冻结，仅对后若干层及新加入的法域标签嵌入层进行对比学习微调。在包含多地域标签的数据集上进行初步训练，让模型学会分辨不同法域文本特征。  (2)全模型微调  解冻模型全部参数，使用法律专属数据进行更深度的训练。加强对关键法律条文、案例推理、判决结果预测等任务的准确度。  2.3强化学习与用户反馈  基于RLHF（Reinforcement Learning from Human Feedback）框架，通过平台前端收集用户对模型回答的评价，形成奖励信号并用于优化模型策略。建立奖励评价机制，基于用户的打分或反馈训练的模型，用于度量生成结果的好坏，以此指导大模型在生成时趋向更优回答。  3. 多模态信息解析与融合  3.1文档结构识别  采用LayoutLMv3对扫描版判决书、带印章的合同等进行段落、表格、批注的结构化解析。  将文档结构信息与文本语义特征拼接或交叉注意力（Cross-Attention）机制相结合，为后续问答或检索提供更丰富的上下文信息。  3.2图像、语音与文本对齐  OCR用于提取法律文档中的文字信息；ASR则将音视频转化为文本。使用多模态嵌入（Multimodal Embedding）对图像、语音、文本特征进行统一向量化表示，实现跨模态关联分析。  4.生成式数据增强  4.1基于法律大模型构建「对抗样本生成器」，自动合成跨境贸易纠纷、数字资产侵权等稀缺案例，通过合成数据扩充训练集多样性  4.2开发「法域迁移学习」模块，利用对比学习技术对齐不同法域条文语义空间，突破普通法系与大陆法系的知识壁垒  5.法律认知智能引擎架构升级  5.1MoE-Transformer混合架构  采用「动态稀疏MoE架构」，为民事/刑事/跨境等不同领域配备专属专家模型，推理效率提升5倍；引入「神经符号系统」，将法律条文转化为可解释的逻辑规则树，实现黑箱模型与符号推理的协同决策  5.2多法系认知引擎  开发「双轨注意力机制」，对大陆法系条文解析采用因果注意力，普通法系判例分析采用时序注意力；构建「法律认知孪生系统」，通过对比学习同步训练内地版（民法典体系）与港澳版（基本法体系）双模型  6.联邦学习下的隐私计算  6.1构建「区块链联邦学习网络」，采用同态加密+零知识证明技术，实现粤港澳司法机构间数据安全流通  6.2设计「差分隐私数据湖」，对港澳敏感判例进行噪声注入与k-匿名化处理，满足《个人信息保护法》跨境合规要求  本项目构建多模态联邦学习数据工厂，通过对抗样本生成与法域迁移学习突破数据壁垒，采用动态稀疏MoE架构与神经符号系统打造双轨认知引擎（大陆法系因果注意力+普通法系时序注意力），实现跨法域精准推理。基于区块链联邦学习网络与差分隐私数据湖保障跨境数据安全，结合生成式RAG与法律决策溯源链，形成“多模态解析-认知增强生成-合规存证”全链路。通过边缘端轻量化部署与自进化知识图谱，持续提升多法系场景下的服务时效性与解释性，建立安全、高效、可信的跨境法律智能基础设施。  含前端.drawio  **图2 技术路线图**  （四）人员分工  “法芯智能”团队主要成员由6人组成，分别是：  **罗尹诞**，22级法学专业，暨南大学番禺校区学生工作办公室助理，有良好的组织、领导能力，在本项目中作为核心负责人，负责人员分工、统筹项目开展和推进工作，并把握项目总体方向，为团队发展提供资源渠道，确保项目高效运作。  **陈炜昊** ，23级网络空间安全专业，熟悉办公软件，Abobe Premiere Pro,Abobe photoshop等，熟悉C语言，python，java，mysql，HTML，css等编程语言。熟练掌握web攻防，渗透测试技能，对于信息打点，webshell上线，免杀绕过和权限维持等方面都有一定的建树。对通义千问等语言模型了解深入，能熟练掌握powerBI对数据进行分析可视化。参加第三届广东大学生网络安全攻防大赛，以及进行数据分类分级项目的研究，作为主要研究人员独立自主设计了一套模型可用于基于paddle框架的数据分类分级训练，并准备推出专利论文、参与昇腾AI发起的云图片模型计算组研究并获校内三等奖。  **苏恒毅**，23级网络空间安全专业，熟悉C、Python、Java等编程语言，已修读自然语言处理（NLP）、机器学习和计算机视觉相关课程，熟悉人工智能、大模型训练及深度学习技术领域，具备前端开发、API 设计、系统架构搭建等技术能力。英语水平突出，具备良好的学术英语能力、中外文献检索能力和信息分类整合能力，能流畅阅读大模型领域的学术英文文献、搜集并整合各种信息。校期间多次参加数模相关竞赛，最高获得国家级二等奖，具有扎实的理论与应用数学知识基础和团队协作沟通能力，通过沟通和协作确保团队中的非技术成员都能理解法律大模型的技术实现和细节。在团队中负责法律大模型的前端UI界面设计、数据处理流程构建以及整体架构设计，确保项目技术方案的可行性与高效性。  **林歆婕**，22级法学专业，擅长使用Xmind、EndNote等工具，已修读人工智能通识，以及国际私法等相关课程，法律素养高，文书撰写能力强。在本团队中主要负责调研市场需求，以及补充内地法律相关知识，对比三法域之间法律差异，为项目提供市场数据支持和法律决策依据，分析行业动态，进行法律服务和市场需求的对接，提升项目产品的法律专业度以赢得市场。  **李东骏**，24级法学专业，擅长法律检索、法律分析等，熟悉港澳与内地法律法规的差异，有良好的学习和研究能力。曾参加过大学生职业规划大赛并取得优异成绩，以及在班级担任学习委员。在团队中负责风控板块，以及提供港澳法律资源的补充，致力于提升团队项目应对风险的能力和增强项目平台粤港澳大湾区的特色。  本项目团队是内外招联合组队，在文献检索、法域差别等方面具有显著优势，充分结合内外招学生的各种优势，为团队项目注入特殊意义。首先是多元文化视角和跨法域适配能力，在构建“三法域智能适配系统”时，外招学生的本地经验可以辅助设计法律条款对比功能，确保平台在跨境纠纷中的实用性；其次是跨学科技术协同，本项目团队联合网络空间安全学院、法学院等不同学院学生，形成“技术+法律+语言”的复合型团队；再而是多语言支持和用户体验优化，外招学生的语言优势可以直接提升平台的多语言服务能力，覆盖多人群用户；最后是创新思维和国际化视野，不同教育背景的团队成员在思维交流中碰撞出新方案，外招学生也可以链接港澳高校等多样资源，助力项目在湾区的推广和商业化。内外招联合团队通过“文化适配力+技术复合性+资源多样性”，能够为项目提供从技术开发到市场落地的全方位支持，尤其是在粤港澳大湾区——多法域、多语言、多文化融合的场景中，形成项目团队的独特优势。 | | | | | | | | | |
| 四、项目特色与创新点 （一）技术特色：智能AI技术优势 1．前端优势  （1）简洁美观的UI界面与高效交互，支持多终端使用、多渠道接入与统一体验   * 方案：主要采用基于Java的现代化UI框架为用户提供简洁、友好且美观的界面；支持快捷注册以及“游客模式”访问；支持多终端同时登录与数据同步并确保功能和用户体验的一致性。 * 技术特征：使用JavaFX前端技术栈保证在不同分辨率和设备上都能良好显示，且与后端服务进行实时接口通讯；采用分级权限策略为注册用户和游客使用提供技术支持。  1. 前后端分离与高可扩展性  * 方案：前端与后端严格分离，主要分别基于Java技术栈和Python深度学习框架，二者通过API进行通信，部署于阿里云等云平台。 * 技术特征：参考当今Deepseek等大语言模型服务方式，前端通过HTTP/2或WebSocket协议与后端进行实时通信并采用标准化的RESTful或GraphQL接口，实现高吞吐与低延迟的请求处理。   2.后端优势  （1）多模态数据预处理管道与文档结构识别模块（LayoutLMv3）助力更好地理解用户需求   * 方案：整合OCR（光学字符识别）、ASR（自动语音识别）和文本数据构建混合数据引擎，处理和理解多种格式的法律文档；基于LayoutLMv3的文档结构识别模块，自动识别和理解法律文档的布局和结构。 * 技术特征：引入多模态对齐机制，如结合最高人民法院“法信”平台的法律知识体系编码（18万法律知识节点），增强对法律文档中图表、印章、手写批注等非结构化数据的解析能力；利用先进的文档布局识别技术提高文档内容抽取和信息理解的准确性。   （2）法律术语知识图谱（Neo4j）与领域检索增强生成（RAG）助力更好地提供准确回答   * 方案：集成“法信大纲”的18万法律知识体系编码，结合Neo4j构建动态更新的术语网络支持条文修订时效性追踪（如民法典更新标注）。在模型推理时，基于RAG思想实现领域特定的检索增强生成方法。 * 技术特征：通过知识图谱助力模型更好地理解和处理法律术语；模型推理时基于RAG思想从预先构建的向量索引数据库或关键词搜索中检索到相关法律条文或案例，再由生成模型进行融合回答。该过程引入“快慢系统”。 * 分层检索策略，快系统基于关键词匹配快速返回法条，慢系统通过向量检索+知识图谱关联展开深度分析。   （3）可解释性与合规性校验模块、案例推荐、联邦学习支持模块助力模型的可扩展性和回答的全面客观性   * 方案：在生成答案时，为模型配置一个可解释性模块，用于输出参考条文、判例来源或关键推理路径，并通过加入合规性校验识别是否涉及隐私或敏感信息；在回答中添加案例推荐机制，向用户提供类似或相关案例参考；引入联邦学习支持模块，支持多机构协作训练，保护数据隐私的同时确保模型的可扩展性，提高模型性能。   技术特征：借助可解释AI（XAI）方法，使模型输出可溯源的证据链（例如依据哪些条文、案例），同时结合合规性扫描模块判断回答中是否包含敏感信息或越权建议。在回答中添加案例推荐时，结合DPO技术（Direct Preference Optimization），根据用户反馈动态优化检索结果排序，优先推荐高采纳率的判例。通过联邦学习，使模型在不共享原始数据的情况下利用多个机构的数据进行训练，提高模型的泛化能力和可扩展性。  （二）产品法律特色：湾区基因驱动   1. 聚焦大湾区多法系融合场景   大湾区法规动态融合引擎。整合内地部门法、香港普通法、澳门大陆法系等差异化法律资源，基于知识图谱构建跨法域关联规则，实现“一案三地”法律条款智能对比与适用建议。针对跨境纠纷，系统可自动匹配相关地域法条，解析管辖权差异与诉讼策略。  多语言法律助手。支持粤语口语识别、繁体/简体中文及英文法律文档互译，适配大湾区多语言沟通场景，降低港澳用户使用门槛。  2. 响应大湾区政策与科技创新需求  服务国家战略。契合《粤港澳大湾区发展规划纲要》中“建设国际法律服务中心”目标，提供跨境争议在线调解、ODR（在线纠纷解决）模块，助力湾区法治化营商环境建设。  赋能重点产业。针对湾区跨境电商、跨境金融、科技创新企业高频法律需求（如数据跨境合规、知识产权跨境保护），开发定制化风险评估工具，生成符合三地法律的合规报告。  3. 技术创新结合本土实践  湾区数据训练模型。采用粤港澳司法案例、仲裁裁决等本土化数据训练AI，提升对湾区法律实务问题的理解精度，如“港澳居民内地购房政策解读”“深港跨境继承案例研判”“内地与港澳居民纠纷管辖权认定规则”，使之更具湾区特色。  政产学研协同生态。与湾区高校（如中山大学法学院）、律所（如金杜深圳分所）、仲裁机构（深圳国际仲裁院）合作共建法律语料库，确保工具专业性与权威性。   1. 文化适配与区域辐射力   粤港特色交互设计。界面提供粤语语音导航、繁体版切换，内置“大湾区法律常识问答”模块，以“港式调解文化”“广府商业习惯”等案例增强用户认同感，优化用户体验感。一带一路延展性。以湾区为起点，探索工具在东盟跨境法律服务中的应用，输出“中国版LegalTech解决方案”。 （三）“法芯智能”创新点 我们的产品与同类产品相比，UI界面采用极简设计，带来了极致的美学体验；界面视觉优化与高效动态交互提升用户粘性与使用效率。大幅提升非专业用户的友好度，降低使用门槛，扩大产品受众；灵活的账号体系满足不同用户的需求。  我们的产品前端与后端严格分离，前端主要基于Java技术栈，后端基于Python深度学习框架，二者通过API进行通信，部署于阿里云等云平台，具备可伸缩、高并发的性能优势。前端可独立迭代和优化，不影响后端大模型的更新；同时为第三方系统或不同语言的开发者提供统一API，方便整合或二次开发，扩大项目生态圈。  我们的大模型提供模态数据预处理管道与文档结构识别模块（LayoutLMv3），能对用户输入的内容进行更加准确的解析，精准理解用户需求，从而有针对性的提供准确回答；而法律术语知识图谱（Neo4j）与领域检索增强生成（RAG）的引入则有助于根据用户需求更准确地匹配对应的信息，并生成高质量的回答。  传统的大模型基于既有的训练数据，生成的回答在时间上具有滞后性。而我们产品对联邦学习支持模块的引入使得大模型的训练数据“来源更广”且能“实时更新”，适应了法律条文不断更新的实际情况。  DeepSeek-R1是由中国 AI 初创公司DeepSeek于2025年1月发布的开源模型，采用了强化学习技术展现了卓越的推理能力。我们的模型基于 DeepSeek-R1 进行了进一步的训练，以适应法律领域的专业需求。然而，原始模型并未专注于法律领域。因此，我们收集了大量法律文本、案例和法规，针对这些专业数据对DeepSeek-R1进行了微调。这一过程使模型能够更好地理解法律术语和逻辑，提供更准确、专业的法律咨询服务。 （四）法律大模型生态优势   **图3 竞品分析图**  与Deepseek 等AI法律服务平台相比的优势：  1.目前部分AI法律平台不能及时更新新颁布或修订的法律条文。如在新《公司法》已经生效的情况下，Deepseek仍呈现旧法条款，会导致系统依据过时条文进行案件分析和策略制定，给案件带来不利影响。本产品联邦学习支持模块的引入使得大模型的训练数据“来源更广”且能“实时更新”，适应了法律条文不断更新的实际情况。DeepSeek-R1是由中国 AI 初创公司Deepseek于2025年1月发布的开源模型，采用了强化学习技术展现了卓越的推理能力，我们的模型基于 DeepSeek-R1 进行了进一步的训练。然而，原始模型并未专注于法律领域。因此，我们收集了大量法律文本、案例和法规，针对这些专业数据对DeepSeek-R1进行了微调。这一过程使模型能够更好地理解法律术语和逻辑，提供更准确、专业的法律咨询服务。  2.法芯智能是国内首个聚焦粤港澳大湾区的法律AI 服务平台，目前许多AI法律平台虽然在跨境案件中可以通过接入国外法律数据库的方式来解决跨境纠纷，但并未像本产品一样深入研究如何解决港澳与内地的跨境案件。如目前没有公开资料显示 DeepSeek 专门针对港澳与内地跨境案件有特定处理方式。针对跨境纠纷，本产品可自动匹配相关地域法条，解析管辖权差异与诉讼策略，高效解决粤港澳跨境案件。  3.现有的AI法律服务平台虽然配有法律专家对其数据建设、案情分析等方面提供建议，但服务结果大多是AI进行客观分析得出，并且在没有专业人士的辅助下，反馈的结果也很有可能与案件实况存在差异，因此表现出案件分析结果的不准确性。本产品在标准化服务中就可以提供人工专家在线案件预测结果审查，可以为平台服务及时提供专业的建议，提高产品服务质量，以及降低客户咨询成本。  4.Deepseek等AI法律服务平台会出现 “AI 幻觉”，无中生有地创造出看似合理但实际上并不存在的法条或案例。如生成的《离婚协议》里，财产分割条款引用了根本不存在的《民法典婚姻家庭编司法解释（五）》。除此之外，部分平台在检索非法律类但与法律业务相关的信息时也可能出错，如检索 “深圳 2023 年最低工资标准” 时，给出错误答案。本产品专门用于法律服务，客户所提出的疑问平台都会以专业严谨的法律术语来回答，并且不会出现引用不存在的法律甚至造法的情况。 | | | | | | | | | |
| 五、项目进度安排  图4 进度安排   |  |  |  | | --- | --- | --- | | 阶段名称 | 时间安排 | 主要任务 | | 项目筹备阶段 | 2024年10月-2025年2月 | 收集整理资料数据、调研国内外同类课题 | | 产品开发阶段 | 2025年4月-2025年6月 | 设计系统架构，建立产品初步模型且试运行和测试 | | 推广提升阶段 | 2025年7月-2025年9月 | 开展市场营销，根据用户反馈更新产品 | | 优化升级阶段 | 2025年10月-2025年12月 | 持续升级产品，优化模型性能和提高用户体验 | | 总结规划阶段 | 2026年1月-2026年4月 | 总结项目运行情况，评估综合效益，规划后续战略 | | | | | | | | | | |
| 六、经费预算及依据    图5 预算表   |  |  |  | | --- | --- | --- | | 类别 | 开支项目 | 预算 | | 文献资料费 | 用于支付资料费、文献检索费 | 1000 | | 调研实践类 | 调研交通费、材料打印费 | 100 | | 技术支持类 | 数据分析、计算、测试、文字挖掘、数据爬取费 | 5000 | | 模型算力类 | 服务器租借、模型算力、显存并发量 | 5000 | | 总计 |  | 11100 | | | | | | | | | | |
| **九、项目预期成果** （一）AI法律服务生态链规模化突破 1.技术底座升级：完成AI大模型3.0版本迭代，提高对话意图识别准确率和问题解决率，建立覆盖细分行业的法律知识库，形成行业领先的技术壁垒。  2.生态规模扩张：实现平台入驻商家跨越式增长，重点拓展粤港澳大湾区跨境贸易、科技创新等特色产业客户，市场占有率完成湾区法律科技服务领域提升。  3.协同效应显现：构建商家数据共享机制，通过客户服务数据回流持续优化模型，形成“企业需求洞察-智能服务升级-商业价值反哺”的闭环生态，降低客户服务成本。 （二）全栈式智能法律生态体系构建 1.认知智能升级：实现法律大模型4.0升级，对话意图识别准确率突破92%（司法考试基准），构建多模态法律推理框架，引入联邦学习实现跨法域知识融合  2.生态协同网络：搭建"政府-企业-律所"三维数据中台，建立跨境法律区块链存证系统，实现大湾区2000+政企单位智能合约部署，市场渗透率达35%  3.动态进化机制：打造法律认知飞轮系统，通过实时多源反馈（司法判例更新率99.2%响应）实现模型周级迭代，服务成本降低40%同时效能提升3倍 （三）商业价值与实践创新双提升 1.经济效益：增加年度平台服务收入，定制化服务客单价提升，实现生态合作伙伴收入占比达总营收的分成每年稳中有升，形成可持续盈利模式。  2.实践创新：结合当下流行的deepseek R1大模型进行预训练，采用bert掩码机制对高质量数据集进行模型微调，在原有参数量不变的情况下采用更多头的自注意力模块机制提高模型精度和效果，扩展deepseek在法律板块的效应。  3.技术溢出效应：孵化法律知识图谱构建、跨法域语义理解等核心技术专利，吸引粤港澳高校联合建立AI法律实验室  通过以上成果，项目将构建法律认知智能新范式，打造政企协同的跨境法治服务平台，形成技术突破-商业闭环-社会价值三位一体的法律科技新基建。 | | | | | | | | | |
| 八、申请人承诺  我保证上述填报内容的真实性。如果获得资助，我与本项目组成员将严格遵守有关规定，切实保障研究工作的进行，按时报送有关材料。  项目负责人（签名）：  时 间： | | | | | | | | | |
| 九、指导教师意见：  签 名：  时 间： | | | | | | | | | |
| 十、学生所在学院意见：    签字盖章：  时间： | | | | | | | | | |
| 十一、学校意见：    签字盖章：  时间： | | | | | | | | | |