

哈尔滨工业大学（深圳）

大一年度项目立项报告

项目名称：基于大模型的智能法律维权助手

项目负责人：艾烨

学 号：2024311176

联系电话：18988703523

电子邮箱：1486100749@qq.com

学 院：基础学部

指导教师：白雪峰

职 称：助理教授

联系电话：15754604524

电子邮箱：baixuefeng@hit.edu.cn

学 院：计算机科学与技术学院

填表日期：2024 年 10 月 26 日

一、项目团队成员

姓名	性别	所在学院	学号	联系电话	本人签字
艾烨	男	基础学部	2024311176	18988703523	艾烨
曾龙安	男	基础学部	2024311389	15828509957	曾龙安
白亦恒	男	基础学部	2024311111	15220134648	白亦恒
耿亦桐	男	基础学部	2024311305	13160877819	耿亦桐

二、立项报告

基于大模型的智能法律维权助手

（一）立项背景

在当今社会，随着法治建设的深入和公众法律意识的提高，法律维权的需求日益增加。然而，普通消费者和劳动者在面对侵权行为时，往往感到无所适从，缺乏有效的法律知识和维权手段。特别是在信息化迅速发展的背景下，尽管法律信息的获取变得更加便利，但许多人在复杂的法律环境中依然难以找到所需的支持和帮助。这种现象亟需引起关注，并促使我们探索基于人工智能技术的法律智能助手解决方案。

首先，传统的法律服务模式在资源和服务能力方面存在一定的局限性。当前，法律咨询主要依赖于专业律师的面对面服务，而这对于普通民众而言，常常需要付出高昂的费用和时间成本。尤其在一些偏远地区，专业法律服务的获取难度更大，许多人在遇到法律问题时往往缺乏及时的法律援助。此外，现有的法律咨询平台虽然能够提供一些在线服务，但通常只限于简单的问题解答，未能覆盖到多样化的维权场景和复杂的法律问题。因此，建立一个智能化的法律维权助手，以便普通用户能够在任何时间、任何地点都能获得高效、精准的法律咨询和维权建议，显得尤为重要。

其次，法律条文和相关规定的复杂性也是普通民众面临的一大挑战。法律语言通常晦涩难懂，普通用户很难准确理解其中的含义。在实际维权过程中，用户往往需要在繁杂的法律条款中寻找适用于自身情况的依据，这一过程不仅耗时，而且容易导致错误理解和应用。由于缺乏专业的法律知识，用户在处理侵权事件时，往往会错失合理的维权机会，甚至在未能正确运用法律条款的情况下，可能会面临更加复杂的法律后果。因此，利用大语言模型的强大语义理解能力，为用户提供简洁明了的法律条文解读和建议，是提升公众法律维权能力的关键。在技术层面，近年来大型语言模型（LLM）如 GPT-3、GPT-4 等的兴起，极大地推动了自然语言处理（NLP）技术的进步。这些模型在语言理解、生成和问答等任务中展现出了卓越的性能，使得机器能够理解和生成自然语言，具有非常强大的应用潜力。特别是在法律领域，随着法律数据的大规模积累和处理技术的不断进步，智能法律助手的构建成为可能。通过对大模型的微调和领域适应，可以实现针对特定法律问题的智能问答、法律条文匹配以及维权路径推荐等功能，从而在更大程度上帮助用户解决实际问题。

另外，社会公众对法律服务的期望也在不断提升。随着社会信息化程度的加深，人们在寻求法律帮助时，越来越倾向于使用便捷的在线服务，而不再满足于传统的线下咨询。因此，如何利用现代科技手段，提高法律服务的可获取性和普及率，成为法律行业转型的重要方向。智能法律助手不仅能够帮助用户快速找到所需的信息，还能够提供个性化的服务，针对用户的具体情况提出针对性的法律建议。这一智能化、个性化的服务模式，将极大提升法律服务的效率和效果，推动法律服务行业的数字化转型。

此外，社会对法律公正和公平的期待也在增加。法律的适用和执行不仅要依赖于法律条文的严格规定，更要体现在法律服务的可及性和有效性上。通过智能法律助手的推广与应用，可以有效缩小法律服务的供给和需求之间的差距，让更多人能够平等地享有法律保护和维权支持。这不仅有助于提高公众对法律的信任和依赖，还能在一定程度上促进社会的和谐与稳定。综上所述，基于大模型的智能法律维权助手项目的立项背景，正是基于当前社会对法律服务需求的迫切性和传统法律服务模式的局限性，结合了技术发展趋势和公众期待。这一项目不仅有助于提升普通民众的法律维权能力，促进社会公平正义，也将推动法律行业的智能化进程，实现更高效的法律服务体系。通过整合先进的自然语言处理技术与法律知识，项目目标是构建一个普惠型的法律智能助手，让每一个需要帮助的人都能够通过简单的方式获得有效的法律支持。

（二）项目研究内容及实施方案

研究内容

（1）智能法律条文匹配与查询

本项目的核心功能是帮助用户在日常生活中遇到的复杂且多变的侵权问题中找到相应的法律依据。该功能将通过微调的法律大模型，在法律法规文本的基础上进行语义理解和检索，确保模型可以对用户描述的具体情境（例如消费纠纷、劳资纠纷等）进行深度解析，并能够识别出相应的法律条款。不同于传统的法律检索方式，智能助手将根据用户的具体表达方式进行条文的自动化匹配，即便用户使用非正式语言描述侵权情况，模型也能通过语义匹配找到相关的法律条款。同时，通过内置的法条解释模块，助手会对法律术语进行简明解释，使得用户能够轻松理解法律条文的具体含义。

（2）多维度维权路径生成

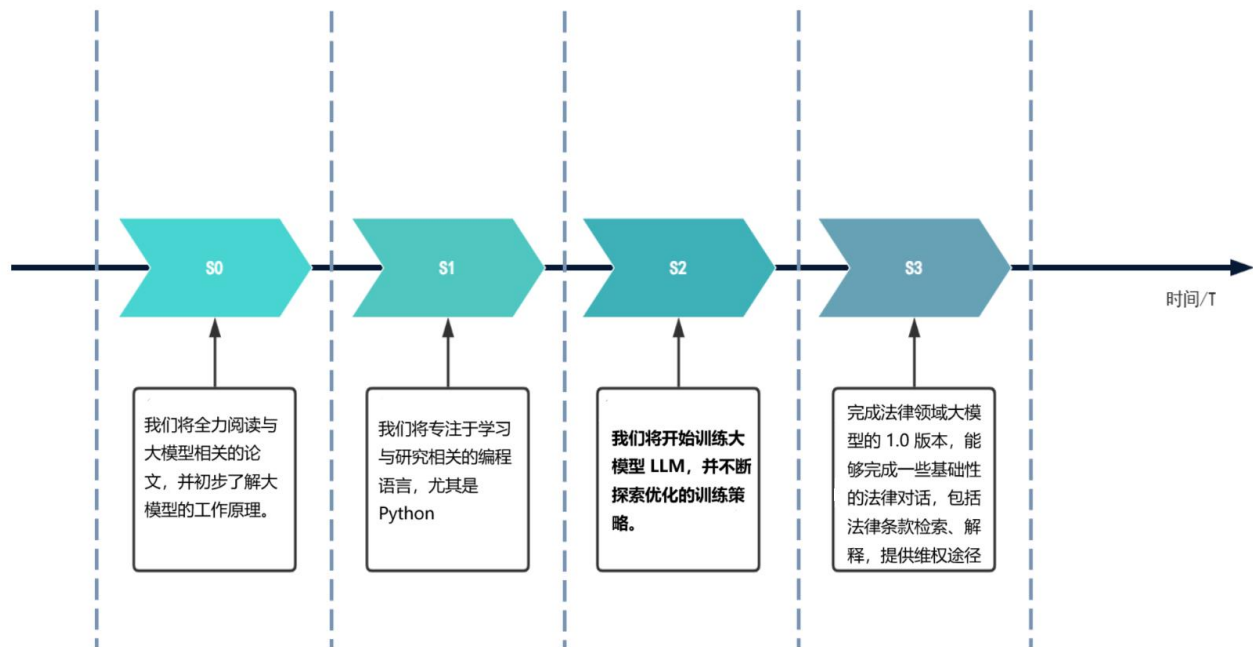
智能法律维权助手将针对不同类型的侵权情境，生成相应的维权路径和指导信息。例如，在消费者遇到产品质量问题或服务纠纷时，助手会根据相关法律法规，生成投诉、调解、法律咨询等多维度的解决建议；在工人遇到劳动权益纠纷时，助手会根据劳动法和相关政策，提供投诉劳动监察部门、申请劳动仲裁、寻求法律援助等具体路径。项目中将深入分析消费权益保护、劳动权益保障等法律的实际应用场景，并通过多层次的知识链路设定，在助手系统中实现法律条文和实际维权路径的智能链接。此外，维权路径还将集成区域性建议，结合用户所在的地理位置或场景细分，推荐最直接、可行的维权路径。例如，当用户位于特定城市或地区时，系统会智能调取当地相关部门的联系方式或程序要求，以便用户高效获取实际支持。

（3）基于法律三段论的法律推理能力增强

为了提高系统的智能性和法律推理的准确性，智能助手将利用法律三段论的逻辑推理方法。在法律应用中，法律三段论包含“法律条文-案件事实-判决结果”的推理过程。这种推理过程是构建系统推理能力的基础，有助于让系统能够模拟法律专家的思维方式进行分析推理。例如，系统能够基于用户陈述的侵权事实（如产品质量不合格、劳动报酬被拖欠等），结合适

用的法律条文，得出具体的维权建议。通过三段论法则，系统不仅能提高推理准确性，还能为用户提供更加清晰、完整的推理路径，从而提升用户的理解度和信任度。三段论推理能力的应用有助于系统在处理复杂情境时仍然保持法律逻辑的严谨性，使得用户对维权建议的合法性和专业性有更高的认可。

实施方案



(1) 数据准备与法律语料库构建

系统的数据基础是法律条文和案例的语料库。首先，项目团队将收集涵盖主要法律领域的法律文本，包括民法典、消费者权益保护法、劳动法、合同法等文本资源。这些法律条文和案例构成了系统进行语义理解和匹配的知识基础。数据准备阶段还将包括文本清洗和去重，以确保数据的质量和一致性。此外，将构建覆盖全国主要地区的法律部门和维权机构联系信息数据库，并定期更新，以确保用户在系统中获取的联系信息准确、有效。所有文本将经过向量化处理，便于模型在推理和检索时使用。系统还将整合一部分法律服务平台的公共数据和用户维权案例文本，以增强智能模型对用户语言特征和具体情境的适应性。数据准备的最终目标是构建一个既涵盖常用法律条款、也能进行多场景推理的法律知识库，作为模型智能推理和决策的核心支持。

(2) 知识检索增强与生成模块的构建

本项目将借助检索增强生成（RAG）技术来优化用户交互的准确性和法律知识输出的完整性。RAG 模块将使用一套基于法条和案例的检索系统，结合深度神经网络的语言生成能力，帮助模型在回答法律问题时提供先验的知识背景。在实际应用中，检索增强生成的工作流程包括：首先，用户提出问题后，系统会在知识库中检索相关的法律条文和案例段落，然后在生成回答时自动整合这些检索结果，以增强回答的可信度和严谨性。为了支持检索增强生成的运行，项目团队将构建一个专门的法律知识库，将收录包括宪法、民法典、刑法、劳动法等常见法条，确保系统在回答时能提供有据可依的法律依据。此模块将定期更新，以适应法律法规的变化，提高生成内容的时效性和准确性。

（3）多轮对话与自适应优化机制

项目的多轮对话模块将支持系统与用户在多轮交互中的连续性和动态调整能力。通过用户历史问题和对话上下文的记忆，系统可以在每次新对话中自动调整答案的结构和内容。此外，自适应优化机制将根据用户反馈数据进行实时学习。例如，若用户表示当前建议无法满足需求，系统会自动对之前的回答内容进行优化，并为用户提供替代路径。为了实现该功能，系统将采用基于深度学习的上下文记忆网络，使系统在识别用户意图时能够结合对话历史，以生成符合用户具体需求的建议。优化机制还将通过反馈回路对对话内容进行持续改进，例如在用户完成投诉后，对进一步的申诉流程给出个性化提醒，确保用户的维权进展顺畅。多轮对话机制的目标在于让用户获得更自然、连贯的体验，提高用户的满意度与使用粘性。

（三）进度安排

- **10.24——11.4** 在这段时间内，我们将全力阅读与大模型相关的论文，并初步了解大模型的工作原理。通过深入研究先前的研究成果，我将对大模型的架构、训练方法和应用领域有更清晰的认识。在此期间，我还将撰写立项报告并制作开题答辩 PPT，以准备开题答辩。这将使我们能够系统地介绍我们的研究计划，并提供对研究目标和方法的详细说明，以便获得导师和评审专家的认可。

- **11.5——3.1** 在接下来的时间里，我们将专注于学习与研究相关的编程语言，尤其是 Python。Python 是一种功能强大且广泛应用于人工智能领域的编程语言，对于训练和实现大型语言模型 LLM 至关重要。通过系统学习 Python，我们将掌握语言的基本语法、数据结构和常用库，为后续的研究工作打下坚实的基础。同时，我们还将深入了解训练 LLM 的代码实现，包括模型构建、数据预处理、训练过程和模型评估等方面的内容。这将使我们能够理解和运用现有的训练代码，为后续的研究工作提供支持。

- **3.1——中期** 在导师提供的算力支持下，我们将开始训练大模型 LLM，并不断探索优化的训练策略。通过使用强大的算力资源，我们将能够在更大规模和更复杂的数据集上进行训练，以获得更准确、更灵活的模型。同时，我们还将积极探索不同的训练策略，如预训练和微调方法，以进一步提升模型的性能和泛化能力。通过这一阶段的努力，我们将初步形成一个能够进行实际应用的法律对话机器人，为法律咨询、维权方面提供智能化的支持和解决方案。

（四）中期及结题预期目标

- **中期目标：**完成法律领域大模型的 1.0 版本，能够完成一些基础性的法律对话，包括法律条款检索、解释，提供维权途径等。

- **结题目标：**完成法律领域大模型的 2.0 版本，能够完成更多法律对话任务（包括复杂情景下的法律问答、维权回复生成等），并在准确度和可靠度上能与其他在法律领域表现出色的大模型旗鼓相当。

（五）经费使用计划

- 购买相关课程的教材
- 租用 GPU 算力
- 差旅、会议费

（六）主要参考文献

- [1] 张钦彤,王昱超,王鹤羲,王俊鑫,陈 海. 大语言模型微调技术的研究综述.计算机工程与应用.2024,60 (17)
- [2] 魏 斌. 法律大语言模型的司法应用及其规范.东方法学. 2024.5(57)
- [3] 沈晨晨,岳盛斌,刘书隽,周宇轩,王思远,陈伟,萧尧,李秉轩,宋鋆,沈晓宇,黄萱菁,魏忠钰.面向法律领域的大模型微调与应用.大数据. 2024,57(12)
- [4] 周可婧. 自然语言处理模型辅助法律检索的限度及应对——以 ChatGPT 为例. 浙江万里学院学报. 2023,36(53)
- [5] 王彬,董焱焱. 基于大语言模型的司法人工智能.山东社会科学.2024,7(176)

三、指导教师意见

签 名：

年 月 日

四、评审专家组意见

批准经费：元 组长签名：

年 月 日