基于双系统推理框架的法律判决研究

尹圣迪, 白泽文, 林鸿飞, 杨亮†
大连理工大学，计算机科学与技术学院，大连，116024
{20201071390,dlutbzw}@mail.dlut.edu.cn,{hflin,liang}@dlut.edu.cn

摘要

法律判决预测是法律人工智能领域的一项重要任务。本文提出了一种基于外部知识的可解释性双系统推理框架，来解决现有方法在刑期预测任务中精度不高且可解释性不强的问题。该框架借鉴认知科学领域的双系统理论，利用大型语言模型的文本理解和生成能力，模拟人类法官处理案件时的决策过程，最终给出具有清晰推理路径的刑期预测结果。此外，通过构建一个高质量思考增强数据集和一个外部法条知识库，提升了模型的解释能力并且有效地抑制法条判断模型出现法条幻觉。实验结果表明，该框架显著提升了CAIL-small和CAIL-big数据集中刑期预测子任务上的精度和可解释性。

关键词： 双系统推理框架 ； 刑期预测 ； 可解释性

Research on Legal Judgment Based on Dual-System Reasoning Framework

Shengdi Yin, Zewen Bai, Hongfei Lin, Liang Yang*
School of Computer Science and Technology,Dalian University of Technology,Dalian,116024
{20201071390,dlutbzw}@mail.dlut.edu.cn,{hflin,liang}@dlut.edu.cn

Abstract

Legal judgment prediction is an important task in the field of legal artificial intelligence. This paper proposes an external knowledge-enhanced interpretable dual-system reasoning framework to address the issues of low accuracy and weak interpretability in existing methods for prison term prediction tasks. Drawing on the dual-process theory from cognitive science, the framework leverages the text comprehension and generation capabilities of large language models to simulate the decision-making process of human judges when handling cases, ultimately producing prison term predictions with clear reasoning paths. Additionally, by constructing a high-quality chain-of-thought-enhanced dataset and an external legal provision knowledge base, the framework enhances the model's explanatory capability while effectively suppressing legal hallucination in statute judgment models. Experimental results demonstrate that the framework significantly improves both accuracy and interpretability on the prison term prediction subtask in both the CAIL-small and CAIL-big datasets.

Keywords: Dual-System Reasoning Framework , Prison Term Prediction , Interpretability

†* 通讯作者

第二十四届中国计算语言学大会论文集，第500页-第512页，济南，中国，2025年8月11日至14日。

1 引言

法律人工智能是为了让人工智能在法律领域内协助人类进行各种任务的自动化决策，以降低人类在各种任务上决策的难度并提高人类在各种任务上的决策准确性。法律判决预测任务（LJP）是法律人工智能领域的关键任务之一，其目的是让模型通过对案件的基本描述来判断该案件所要用到的法律条文(applicable law articles) 和罪名判定(charges)，并且预测嫌疑人被判处的刑期(term of penalty)。法律判决预测领域数据集具有语言简明、逻辑严密的优点，但是其中大量的专业化属于使得模型甚至人类都非常难以理解，数据集中的一条样例如图1所示。随着深度学习技术，比如注意力机制（Attention Mechanisms）以及特别是图神经网络（GNNs）在自然语言处理领域的突破，LJP任务取得了显著进展 (Luo et al., 2017; Zhong et al., 2018a; Xu et al., 2020; Zhao et al., 2023)。

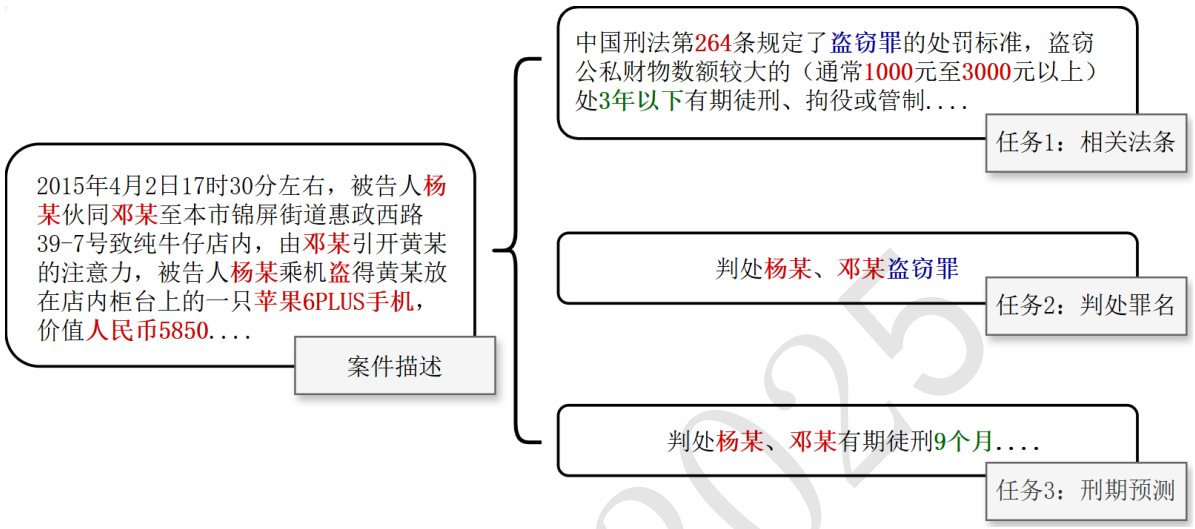

Figure 1: 法律判决预测任务图

现有研究通过捕捉案件描述中的实体与法条概念之间的语义关联在罪名预测子任务上已经达到了较高的准确度。但是在对精度要求更高的“刑期预测”子任务上现有方法仍然面临两个严峻的挑战。

1.现有研究很难理解案件情节与裁量情节（如初犯、累犯、立功、自首等）之间的语义关联，导致刑期预测的准确率仍有很大提升空间。

2.传统的判别式模型虽然能预测刑期，但是却无法生成其判决理由和思考过程，导致其严重缺乏可解释性。

在法律这个高度强调程序正义、判决说理和人权保障的领域，法官、律师、当事人都需要理解量刑预测的逻辑依据，然而判别式模型预测结果的不可解释性无法满足司法公开和透明的要求，是其应用于实际司法场景的主要障碍。

大语言模型（LLMs）强大的文本理解和生成能力为提升LJP任务的精度和可解释性带来了新的可能性。为了应对上述挑战，本文提出了**基于外部知识的可解释性双系统推理框架**(Interpretable Dual-System Reasoning Framework with External Knowledge,IDSRF)。训练**快速法条检索与校验系统**分析案情、查找法条的能力，来模拟司法人员凭借经验直觉筛选法条的过程。训练**可解释量刑推理系统**考量情节，给出具有推理路径的刑期预测结果的能力，来模拟司法人员权衡案情给出判决和解释的过程。

我们认识到**可解释量刑推理系统**必须在准确的法条内容基础上才能进行可信的案件推理，然而以LLMs为核心的**快速法条检索与校验系统**会不可避免的产生“法条幻觉” (Law Article Hallucination) (Huang et al., 2025)，严重影响预测的准确性和可靠性。为了减少幻觉所带来的影响，我们引入了一个精心构建的外部法条知识库对模型输出的法条进行校对，保证推理依据的法条绝对可靠。本文的主要贡献如下：

基金项目：国家重点研发计划（2024YFA1012700）

（1）提出了基于外部知识的可解释性双系统推理框架解决司法量刑缺乏可解释性的问题，提高了LJP任务中法律条文判断以及量刑预测两个子任务的精确度。

（2）快速法条检索与校验系统可以准确的识别法条并输出法条内容，可解释量刑推理系统可以给出具有推理路径的刑期预测结果。

（3）构建了一个经过人工校对的小规模、高质量的思考增强数据集，以及一个对法条概念进行精炼描述的知识库，有效的抑制模型出现法条幻觉。

2 相关工作

2.1 法律判决预测

法律判决预测（LJP）是法学领域和人工智能领域交叉而来得到的关键研究问题，专注于用自然语言处理（NLP）技术 (Minaee et al., 2021)来理解并处理法律文本，让模型得到较为准确的预测结果为人类的决策提供帮助。早期LJP任务主要集中在构建预定义的规则以及法律知识库，通过手动编码的形式结合预定义的规则来预测法律案件 (McCarty, 2013)，但是高昂的人工成本限制了这种方法在大量且复杂的法律文本上的能力。不久之后，人们开始尝试用支持向量机和Naive Bayes来预测法院判决 (Katz et al., 2017)，这种方法显著的提高了模型的预测性能，但是却无法得到合理的特征表示。

随着深度学习的出现，研究人员开始尝试将神经网络引入LJP任务中。Luo et al. (2017)通过捕捉案件描述和相关法条之间的关系从而提升了罪名预测子任务的准确率。Chen et al. (2019b)通过采用基于门控的模型，这有效减轻了传统RNN在处理长文本序列时常见的梯度消失与爆炸问题，进而提升了刑期预测的准确性。紧接着预训练语言模型的出现大大增强了模型对复杂文本的理解和表示能力，很多强大的预训练语言模型例如BERT (Devlin et al., 2019)、RoBERTa (Liu et al., 2019)也纷纷被研究人员用在LJP领域中并开发了专门针对法律文本的变体，例如Lawformer (Xiao et al., 2021)和Legal-BERT (Chalkidis et al., 2020)这些方法大幅提高了模型对文本的理解能力，进而提高了模型在LJP任务上的准确率。

随着图神经网络的发展 (Kipf and Welling, 2016; Hamilton et al., 2017; Bonner et al., 2019)，这类图结构模型能够通过聚合节点及其邻域特征来融合更多信息，从而展现出优异的图结构数据处理能力。由于法律文本具有很强的实体关联信息符合图结构的特征，所以最新研究将图神经网络引入LJP任务来增强模型对法律案件的理解。Zhong et al. (2018a)引入了有向无环图建立子任务之间的拓扑关系。Guo et al. (2024)首次提出了一种单层的异构图结构来捕捉文本中复杂的语义关系，Zhao et al. (2023)提出了多层异构图结构，并且将五种文本级异构图进行融合，得到可以承载文本关联、语义信息等更多信息的图，进一步完善了图架构。Meng et al. (2025)在LA-MGFM的基础上保留了三层图结构并且引入Lawformer模型来提取道义特征，得到了目前效果最好的图结构模型。

2.2 双系统理论在人工智能领域的应用

“快慢思考” (Kahneman, 2011)和“双系统理论” (Evans and Stanovich, 2013)是认知学领域中两个重要的理论。“快慢思考”指出了人类的思考过程是由快速的直觉和缓慢的理性共同驱动，“双系统理论”强调了两种认知过程的必要性。现有研究已有将双系统理论整合到机器学习中，Mittal et al. (2017)将向量空间模型概念化为快速思维（系统1），将知识图中的推理概念化为慢速思维（系统2），提出了一种用于搜索任务的混合查询处理引擎，Chen et al. (2019a)提出了一种端到端框架，其中包括一个表示快速思维的生成解码器（系统1）和一个表示慢速思维的推理模块（系统2），用于解决复杂任务。

基于上述工作，目前的图结构模型在LJP任务中的罪名预测子任务上已经达到了很高的准确率（98.44%）。但是基于图结构的方法在刑期预测任务上依然存在局限性，目前工作专注于让模型直接输出刑期的预测结果而忽略了预测结果的来源，导致其精度不高且严重缺乏可解释性。若想提升模型在刑期预测任务上的可解释性，其推理过程应做到与人类的认知过程严格对齐。所以我们提出了**基于外部知识的可解释性双系统推理框架**来模拟人类司法人员在执行量刑预测任务时的认知过程，解决了在刑期预测任务中量刑预测精度不高且可解释性不强的问题。

第二十四届中国计算语言学大会论文集，第500页-第512页，济南，中国，2025年8月11日至14日。

3 基于外部知识的可解释性双系统推理框架

在本节中，为了解决法律判决预测（LJP）任务中刑期预测部分的精度不高且可解释性不强的问题，我们提出了**基于外部知识的可解释性双系统推理框架**。该框架借鉴了认知科学中的双系统理论概念，模拟人类快速直觉判断（法条判断）与缓慢审慎推理（刑期预测）相结合的决策过程，旨在基于LLMs的文本理解和生成能力使其可以像司法人员一样思考，实现既准确又可解释的法律判决预测。图2展示了整个框架的框图，其主要由两个核心系统组成：系统1：快速法条检索与校验系统，用来根据案件描述判断所用法条并生成其内容，并辅以一个精心构建的外部法条知识库来校对系统1的输出，降低“法条幻觉”带来的影响；系统2：可解释量刑推理系统，用来根据案件描述和法条内容预测刑期并生成合理的解释。

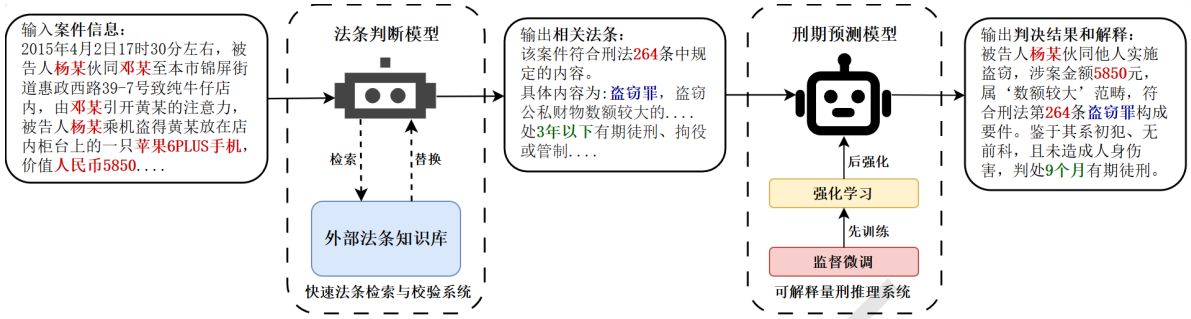

Figure 2: 基于外部知识的可解释性双系统推理框架图

3.1 快速法条检索与校验系统

为了模仿人类法官凭借经验快速判断法条的过程，我们首先使用了一个**快速法条检索与校验系统**，该系统的主要目标是快速、准确地识别与输入案件描述相关的法律条文并且为后续的可解释量刑推理系统提供准确的法律依据。为此，我们选用了有代表性的高效轻量级大型语言模型Qwen2.5-3b-Instruct (Yang et al., 2024)模型作为基线，通过在法条判断和法条内容概述两个特定的任务上进行了微调，训练出了一个法条判断模型，使其具备以下能力：

1. 法条序号识别:根据输入的案件基本事实描述，模型能够快速地分析案情并判断出最可能适用的法律条文的序号。

2. 初步内容生成:基于其在法条内容概括任务上学习到的知识，模型会尝试初步概括和生成它认为与识别出的法条序号相对应的核心内容。

我们发现法条判断模型由于本身参数量较小并且需要同时学习两个任务，导致其自行生成的法条内容描述特别是在处理细节或区分相似法条时不可避免的出现“法条幻觉”如图3所示，即生成的内容与真实的法律规定存在偏差。这种不准确的法条内容如果直接传递给后续的量刑推理系统，将会严重影响最终判决的准确性和可解释性。为了解决这一关键问题，我们人为的构建了一个可以用来校验模型的外部知识库，该知识库包含了我们为数据集中涉及的每个法条序号精心编写的描述，允许法条判断模型在输出预测的法条编号后从知识库中选择相对应的法律条文来对生成的内容进行校对，确保了知识的准确性和权威性。**快速法条检索与校验系统**的工作流程被设计为：

1.识别序号与初步生成：Qwen2.5-3B-Instruct模型首先处理案件描述，输出其判断的最相关法条序号，以及初步生成的对应法条内容。

2.法条检索与替换：在模型识别出法条序号后，系统会从外部法条知识库中检索对应的权威法条内容描述。并且将两者比对校验，如果出现显著不一致的情况则进行替换，将识别出的法条序号和最后经过校验的法条内容作为输出传递给**可解释量刑推理系统**，以此来保证法条内容的绝对准确性。

3.2 可解释量刑推理系统

为了模拟经验丰富的司法人员进行审慎量刑的思考过程，给出准确的刑期预测以及可理解的推理路径，在**快速法条检索与校验系统**输出了与案件相关的准确法条内容后，我们引入了系统2：**可解释量刑推理系统**。该系统是整个框架的核心推理引擎，旨在让刑期预测模型能够具

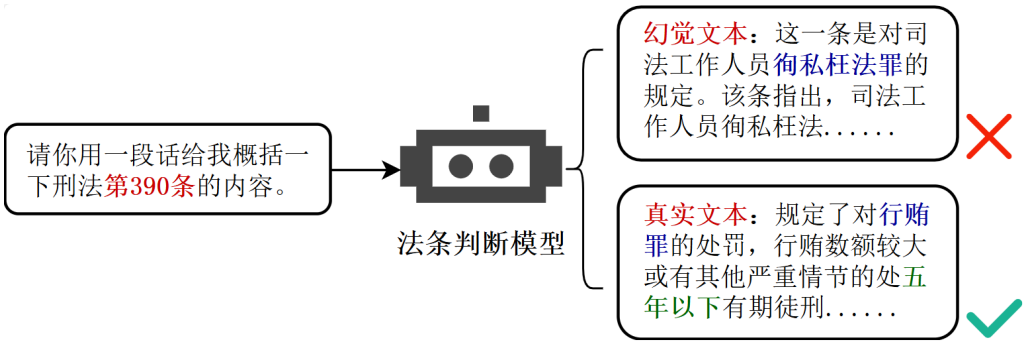

Figure 3: 法条幻觉示意图

有分析案情、援引法条和考量情节的能力。我们选用Qwen2.5-7B-Instruct (Yang et al., 2024)作为系统2的核心模型，因为它既具有一定的语言理解能力又不会因为过于庞大的参数导致成本过高并且效率低下。

为了使Qwen2.5-7B-Instruct具备类似司法人员的复杂推理能力，我们设计了一个分阶段的训练策略，首先让其学会分析案件以及根据判决结果给出推理过程的能力，然后利用全数据进行大量的监督微调来深化其预测判决结果的能力，最后利用强化学习来进一步提升其预测结果的精确性以及推理过程的规范性。这种从“学习范式”到“实践泛化”，再到“精炼优化”的训练流程，在一定程度上也契合了人类学习复杂技能的自然进阶过程：先通过高质量的指导建立认知框架，然后通过大量练习巩固应用，最终通过针对性反馈实现能力的精进。

3.2.1 思考过程的学习与模仿

我们首先认识到直接让刑期预测模型从案件描述, 法条内容, 刑期的简单映射中学习复杂的量刑逻辑和可解释的推理过程是极其困难的。所以我们构建了一个小规模、高质量的思考增强数据集（High-Quality Few-Shot Reasoning Dataset,HQFR）来引导刑期预测模型学会思考。首先我们将模型会遇到的法律条文做出统计并且从每一个法律条文对应的若干案件中选出最具有代表性的25条数据，条件为字数在450至500之间且具有清晰地犯罪情节描述。对于每一个选定的样本，我们整合了其原始的案件描述、从外部知识库中提取的对应且经过验证的法条内容概述，以及其基准刑期标签。我们选择使用deepseek-r1 (Guo et al., 2025)模型来作为“教师模型”，将这些整合后的信息输入给“教师模型”，并且引导“教师模型”模拟经验丰富的司法人员进行判决推理，生成详尽的、结构化的思考过程。该生成的思考过程显式地包含了对案件事实的分析、关键量刑情节（包括从重、从轻及酌定情节）的识别与评估、相关法条在当前案情下的具体适用性解释、对各项影响因素的权衡考量，形成最终推导出刑期结果的完整逻辑链条。为确保生成数据的严谨性和有效性，所有由“教师模型”产生的思考过程均经过了严格的人工审查与多轮校对，旨在修正任何潜在的逻辑谬误、事实偏差或表述歧义，从而构建出一个逻辑一致的高质量的思考增强数据集。

我们使用这个带有高质量、经校对思考过程的小规模数据集，对Qwen2.5-7B-Instruct进行监督微调。目标是让Qwen2.5-7B-Instruct学会并模仿“教师模型”所展示的结构化、逻辑化的思考方式。通过学习这些带有明确推理步骤的样本，Qwen2.5-7B-Instruct能够初步掌握如何将案件事实、法条内容与量刑情节联系起来，并以连贯的语言表达出来。最终证明这个小规模高质量的数据集对提升刑期预测模型的解释能力是有效的。

3.2.2 全数据监督微调与解释生成

刑期预测模型在小规模高质量的数据集上预训练之后，初步掌握了“教师模型”结构化的司法推理范式。为了泛化刑期预测模型学到的推理能力到整个训练数据集覆盖的更广泛、更多样的案例场景中，我们用了全数据对刑期预测模型进行进一步的训练。训练实例的输入由原始的案件描述和经由系统1校验并提供的准确法条内容构成。模型的训练目标则是一个复合输出，它不仅包括对最终刑期精确到月份的预测，还要求模型同时生成其得出该刑期预测所依据的内部思考过程，该过程即被视为模型预测的伴随解释。

通过这种让刑期预测模型给出预测结果的同时强制其利用所学逻辑加入思考的方式，让刑

期预测模型更加理解了案件事实、法条规定、量刑情节与刑期结果之间复杂映射，强化了其独立生成连贯、逻辑化思考过程的能力，进而也提升了刑期预测的准确性。这一阶段的目标是确保模型在面对未曾见过的案例时，既能做出合理的预测，又能提供透明的、可追溯的解释依据。

3.2.3 组间相对策略优化精炼

当刑期预测模型经过上述两个步骤的训练之后，已经可以基本掌握了法官最基本的推理模式并且泛化至全量数据。但是仅依赖监督微调难以完全满足现实司法实践的高标准，训练后的模型有时生成的解释中只包含刑期预测的大致区间或者模棱两可的输出而没有准确的判决答案，有时其预测的刑期虽然在宏观上接近基准答案，但在具体数值的精确度上尚有提升空间。考虑到司法判决对于预测精确性和解释规范性与合理性的严格要求，为了弥补这些细微差距并进一步精炼模型的性能，我们引入了组间相对策略优化(Group Relative Policy Optimization,GRPO) (Shao et al., 2024)让经过训练的模型在高质量思考数据集上进行强化学习，作为训练流程的最后关键阶段，来进一步提升模型预测的精确性和解释的合理性。GRPO作为一种先进的强化学习技术，在近端策略优化（PPO）的基础上进行改进。不需要训练一个较大的价值模型，转而使用同一问题下多个输出的平均奖励作为基线来计算输出中每个token的相对优势，通过这种方式显著降低了训练所需的资源。GRPO的设计初衷为了解决数学推理的复杂性与结构化特性，这与刑期预测任务在提升预测结果的精确性的同时，保证思考过程合理性和规范性的目标高度一致。

为了利用GRPO有效精炼模型的刑期预测精确度和思考过程的质量，我们设计了一个多维度、结构化的奖励函数评估模型针对给定案件生成的预测刑期和思考过程输出。该奖励函数旨在量化输出结果在多个关键维度上的表现，奖励总得分为3分，其中刑期预测的准确度被赋予更高的权重（最高2分）以强调量刑结果的精确性，同时思考过程的质量也占有重要比重（最高1分）以保证解释的规范性与合理性。

总奖励$Reward$是由两部分组成：准确率评分R_{acc}和质量评分$R_{quality}$，公式如下：

$$Reward(y, y_{gt}) = R_{acc}(v_s, v_{s,gt}) + R_{quality}(y, T_{r,gt}) \tag{1}$$

其中$y_{gt} = (T_{r,gt}, v_{s,gt})$是标准答案，$y = (T_r, v_s)$是刑期预测模型输出的答案，其中$T_r$代表生成的思考过程文本，$v_s$代表预测的刑期数值。$R_{acc}$是刑期**预测准确度奖励**，用来评估预测刑期$v_s$与标准答案刑期$v_{s,gt}$的接近程度。考虑到案件中可能会存在无罪释放的样例，也就是刑期为0，所以我们采用了相对误差来量化两者之间的误差，具体公式如下：

$$R_{acc}(v_s, v_{s,gt}) = \begin{cases} 2.0 & \text{if } v_{s,gt} = 0 \wedge v_s = 0 \\ 0.0 & \text{if } v_{s,gt} = 0 \wedge v_s \neq 0 \\ 0.0 & \text{if } v_{s,gt} \neq 0 \wedge v_s = 0 \\ \max\left(0.0, 2.0 - 2.0 \cdot \frac{|v_s - v_{s,gt}|}{v_{s,gt}}\right) & \text{if } v_{s,gt} \neq 0 \wedge v_s \text{ is valid} \\ 0.0 & \text{otherwise (e.g., } v_s \text{ is invalid)} \end{cases} \tag{2}$$

当标准答案与模型预测都有意义且均不为0的时候，利用相对误差来定义奖励函数，以训练刑期预测模型的预测结果逼近标准答案的能力。当标准答案不为0而刑期预测模型的预测结果为0则会造成犯罪分子逍遥法外，当标准答案为0而刑期预测模型的预测结果不为0时则会出现“冤枉好人”的情况。由于司法判决的严肃性与公平性，这两种情况需要刑期预测模型坚决杜绝，所以以上两种情况的刑期预测准确度奖励R_{acc}均为0，保证了刑期预测模型“不会放过一个坏人，也不会冤枉一个好人”。除了设计准确度奖励来确保模型的精确度，我们还设计了思考过程质量奖励$R_{quality}$，该奖励综合评估生成输出y的格式、长度和内容，由三部分加权求和，公式如下：

$$R_{quality}(y, T_{r,gt}) = \min(1.0, w_f \cdot \mathbb{I}_{format}(y) + w_l \cdot R_{length}(T_r) + w_c \cdot R_{content}(T_r, T_{r,gt})) \tag{3}$$

其中，$w_f = 0.2$, $w_l = 0.2$, $w_c = 0.6$ 是各部分的权重。$\mathbb{I}_{format}(y)$ 是**格式规范性奖励**，判断输出y 是否包含所有必需的结构标签(e.g., `<reasoning>`, `</reasoning>`, `<sentence>`,

</sentence>)，强制要求刑期预测模型分别将思考过程输出到<reasoning>, </reasoning>中，预测刑期输出到<sentence>, </sentence>中，以此来确保模型的输出既有思考过程又有明确的刑期预测输出，$\mathbb{I}_{format}(y)$ 公式为：

$$\mathbb{I}_{format}(y) = \begin{cases} 1 & \text{if all required tags present in } y \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

为了提升刑期预测模型输出解释的可读性、简洁性并模拟真实法律文书的风格，我们还设计了**长度控制奖励** R_{length}限制思考过程的输出长度。过长的推理文本容易包含冗余信息、重复论证或旁枝末节，不仅会模糊核心逻辑，降低用户的阅读效率，也与法律文书追求精确、必要、简明的特性相悖。通过设定长度上限，我们鼓励模型生成更精炼、更聚焦于关键量刑步骤和核心法律要素的解释。这不仅使得推理过程更易于快速理解和把握，也引导模型学习并贴近真实判决书中针对特定论点进行简洁、有力阐述的专业写作风格，从而提高生成解释的实用性和专业度。R_{length}对思考过程文本T_r的长度$L(T_r)$进行评估，鼓励其最大长度不超过L_{max}。R_{length}公式为：

$$R_{length}(T_r) = \begin{cases} 1 & \text{if } L(T_r) \leq L_{max} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

为了确保刑期预测模型生成的思考过程贴近理想的、符合法律规范的推理路径，我们还设计了**内容相似度奖励**。通过对模型生成的解释文本与构建的高质量思考增强数据集中的解释文本使用Levenshtein Ratio方法进行相似度比较，来度量其思考的过程与标准的思考过程的相似度。其中Levenshtein Ratio方法是基于Levenshtein Distance（莱文斯坦距离）(Levenshtein and others, 1966)计算出的一个字符串相似度度量，作用是衡量的是两个字符串之间的相似程度，取值范围通常在[0, 1]之间，值越接近1表示两个字符串越相似。该奖励项能够直接引导模型学习并再现正确的法律逻辑、关键量刑要素的考量顺序以及恰当的法律术语运用，例如初犯、从犯、主动自首、取得谅解等。$R_{content}(T_r, T_{r,gt})$公式为：

$$R_{content}(T_r, T_{r,gt}) = \begin{cases} Sim_{Lev}(T_r, T_{r,gt}) & \text{if } T_{r,gt} \text{ is available and valid} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

在每个GRPO训练步骤中，对于一个输入的案件描述和对应的准确法条内容，我们让当前的策略模型生成一组候选的输出$y = (T_r, v_s)$。随后，我们利用上述设计的奖励函数对每个候选输出进行打分，得到一个包含若干个总奖励值的向量$\mathbf{r} = (r_1, r_2, \ldots, r_n)$。接下来，GRPO算法利用这些奖励分数计算每个候选输出中每个token的相对优势$(\hat{A}_{i,t})$。这些相对优势信号随后被用于更新策略模型的参数，依据GRPO的目标函数，并结合KL散度正则化。通过迭代这个过程，模型逐渐学习调整其生成策略，尤其侧重于产生能够获得更高刑期准确度得分的输出，同时也兼顾生成格式规范、长度适中、内容合理的判决解释。

4 实验结果与分析

4.1 数据集和基线模型

为了验证本文提出的方法的有效性，我们做了大量的实验来验证上述方法的有效性，在实验中我们使用的数据集是中国人工智能与法律挑战赛公开数据集CAIL-2018 (Xiao et al., 2018; Zhong et al., 2018b)。CAIL-2018数据集是这是一个专门为中国人工智能和法律挑战赛设计的大型法律数据集，数据集包含大量的各种类型的中国刑事案件，案件的主要来源是裁判文书网等大型公开的裁判文书开源网站保证了数据集内容的规范性和可信性，数据集中每个案件都有详细的标签包括案件事实、相关法条、判决刑期等。

CAIL-2018数据集的数据总量超过260万，为了减少训练成本在LADAN (Xu et al., 2020)任务中又将CAIL-2018进行了进一步处理将样本中的刑期预测按照表划分成了11个互不重叠的区间。除此之外还过滤掉数据集中案件描述过于简短且不足十个有意义单词的样本，因为它们缺少足够的上下文信息，并且筛选出了同时涉及多个法条或者多项罪名的案件样本以及总涉及

法律条文不足100个相应案例样本的极少数样本。最终转化成了CAIL-big（来自the first stage dataset）包含超过170万条数据和CAIL-small（来自the exercise stage dataset）包含128,368条数据。为了有效的验证我们方法的有效性，并确保对比实验的一致性和公平性，所以我们选择在CAIL-small和CAIL-big两个数据集上进行训练和测试，数据集详细的统计数据如表1所示，更多数据集信息可以参考官方发布的论文信息。

	训练集规模	测试集规模	法律条文	刑期区间	案例平均长度	解释平均长度
CAIL-small	101,619	26,749	103	11	408	–
CAIL-big	1,587,979	185,120	118	11	435	–
HQFR	3,660	1,216	103	11	462	154

Table 1: 数据集统计，包括CAIL-small、CAIL-big和构建的HQFR。

CAIL-2018的两个数据集也同样存在局限性，一个潜在的问题是不同类型的案件之间分布严重不平衡。例如，其中涉及**法条264（盗窃罪）**、**法条133（危险驾驶罪）**、**法条293（寻衅滋事罪）**以及**法条266（诈骗罪）**，与这四条罪名相对应的案例样本共占总样本的**60%**。而像**法条268（聚众哄抢罪）**，在数据集中仅有105条样本与之对应，仅占总样本的**0.08%**。这种情况可能会在模型训练过程中引入偏见，从而降低其在极小样本量类别上的刑期预测精确性以及解释的合理性。

为了应对数据不均衡所带来的挑战，在我们构建的小规模、高质量的思考过程增强数据集中，针对CAIL-small数据集中出现的所有法条均选取了25条案件情节描述较为详细的案件，并对这些案件的思考过程进行了详细的生成与增强。这种方法不仅旨在显著增强量刑预测模型在极小样本量法条类别上的刑期预测精度和可解释性，也在一定程度上提高模型在整体样本上的预测性能。为了有效地评估我们系统在LJP任务上的性能，我们选择了一些在领域内被广泛认可且具有代表性的基线模型进行比较。以下是所选的基线模型：

CNN：(Chen, 2015)具有多层的卷积神经网络并且最后用softmax作为分类器，这也是CAIL-2018原始论文中提到的最好的基线模型。

FLA：(Luo et al., 2017)该方法将注意力机制融合进法律的基础信息中，从而可以让模型理解案件事实和法律条文之间的关系。

Few-Shot：(Hu et al., 2018)该法律判决预测模型尝试使用拓扑学习理论，其核心在于构建案件间的拓扑表征结构，并应用图神经网络技术捕捉案件事实描述与相应法律条款之间所蕴含的深层复杂关联，最终达成判决预测的目标。

BERT：(Devlin et al., 2019)一个基于双向transformer的经典预训练语言模型，该方法可以让模型更好的处理上下文信息，从而具有更好的广泛适用性和语义通用性。

TOPJUDGE：(Zhong et al., 2018a)该方法尝试将LJP任务中的多个子任务转化成一个拓扑关系来互相学习，并且抽象为一个有向无环图提出了一个拓扑多任务学习框架来提升整体效果。

LADAN：(Xu et al., 2020)将CAIL-2018拆分成big和small的关键工作，该方法引入了一种特殊的注意力机制来区分和处理具有相似法律条文的案件，增强了模型对相似案件的处理能力和对文本理解的深度，进而提高了模型的效果。

NeurJudge：(Yue et al., 2021)该方法运用了一种上下文感知机制，并且搭建了一个全面的神经网络框架来理解复杂的法律案件。

GCLA：(Dong et al., 2024)该方法构建了句子级的图结构，采用了图对比学习方法并且结合了数据增强技术。增强了模型捕捉关键特征和关系和区分相似案例的能力。

HD-LJP：(Zhang et al., 2024)该方法引入了多任务学习的策略，在多个法律相关任务之间构建了一种层次依赖关系，以此来同时处理多个法律相关的任务。

LA-MGFM：(Zhao et al., 2023)该方法引入了多层异构图结构并融合以此整合多个维度的语义信息，训练语义增强图神经网络，使模型可以捕捉复杂的法律关系和语义特征。

DPFSI：(Meng et al., 2025)该方法在LA-MGFM构建的多层异构图结构的基础上删去了部分影响较小的异构图，并引入了道义逻辑来处理法条中的特殊词汇，进而提升模型的整体效果，这也是目前为止最为先进的方法。

lawLLM-7b: (Yue et al., 2024)该方法提出了一个基于LLMs的智能法律系统，可以在原始模型上进行指令微调，使模型具有更强的推理能力。

lawLLM-13b: (Cheng et al., 2024)该方法提出一种将原始语料转化为阅读理解文本的简易方法，可以提升LLMs在法律领域中的表现。

4.2 实验设计

我们的系统整体是使用Python中的Pytorch框架实现，并且选择了上文提到的基线模型作为比较实验。其中CNN和BERT的结果分别来自LA-MGFM和HD-LJP，lawLLM-7b和lawLLM-13b两者结果均为参考相应框架进行下游训练所得出的结果，其余基线模型结果分别于其原始论文中发布的结果保持一致。

在“快思考”阶段，我们使用在全数据集上经过lora微调过的Qwen2.5-3b-Instruct模型作为**快速法条检索与校验系统**中的法条判断模型，在“慢思考”阶段，使用分别经过在小规模高质量数据集(HQFR)和全数据集两个阶段微调并且用GRPO精炼的Qwen2.5-7b-Instruct模型作为**可解释量刑推理系统**中刑期预测模型。具体实验参数请参考**附录A.实验参数设置**

模型评估阶段，因为数据集存在数据不均衡的问题，所以我们选用了 (Zhong et al., 2018a)中所提到的准确性（Acc）、宏观精确度（MP）、宏观召回率（MR）和宏观F1（F1）作为评估指标，更为科学的地评估我们提出的系统在法律判决预测任务中的性能。

4.3 实验结果与分析

针对我们提出的方法，在CAIL-small和CAIL-big两个数据集上开展了如下分析，具体结果表2和3所示。实验结果重点关注相关法条判断和刑期预测两个子任务，使用了四个评估指标：Acc、MP、MR和F1。从表2可以看出，在法条预测子任务上我们的快速法条检索与校验系统超过了当前最佳基线lawLLM-13b，1.06%（Acc），在刑期预测子任务上我们的可解释量刑推理系统超过了当前最佳基线lawLLM-13b，10.11%（Acc），9.12%（MP），7.03%（MR），8.55%（F1），实验结果说明我们的框架在两个子任务中的表现通常优于基线模型，证明了该框架的有效性。从表3中可以看出，我们提出的框架在数据量更大、种类更多的CAIL-big数据集上的效果也同样优于其他基线模型，证明了该框架具有良好的泛化能力。

方法	法条判断				刑期预测			
	Acc	MP	MR	F1	Acc	MP	MR	F1
CNN	78.61	75.86	74.60	73.59	35.20	32.96	29.09	29.68
FLA	77.72	75.21	74.12	72.78	36.32	30.81	28.22	27.83
Few-Shot	79.30	77.80	77.59	76.09	36.52	35.07	26.88	27.14
BERT	79.14	78.77	73.26	72.63	38.41	33.96	28.71	28.43
TOPJUDGE	79.79	79.52	75.39	73.33	36.05	34.54	32.49	29.19
LADAN	81.02	78.24	77.38	76.47	38.29	36.16	32.49	32.65
NeurJudge	79.81	78.25	79.59	77.55	36.66	34.85	33.82	34.13
GCLA	82.14	80.08	77.71	77.68	37.03	33.80	29.86	29.88
HD-LJP	81.47	79.63	78.26	77.42	42.46	40.20	36.67	37.07
LA-MGFM	84.95	83.91	83.32	82.93	43.01	41.94	40.06	41.04
DPFSI	87.83	86.72	**85.48**	85.33	43.19	41.65	39.76	40.75
lawLLM-7b	90.33	90.52	83.54	85.69	41.10	35.39	33.06	33.41
lawLLM-13b	90.53	**91.05**	83.99	**86.12**	44.17	45.89	44.22	44.27
DERF (Ours)	**91.59**	76.84	79.39	78.10	**54.28**	**55.01**	**51.25**	**52.82**

Table 2: CAIL-small上对比实验结果,对最好的结果加粗且次好结果用下划线标注。

从表2中可以看出我们的框架在法条预测子任务上的MP、MR和F1这三个指标得分较低，然而在表3中这三个指标有了明显的提高，说明此现象是由于法条判断模型参数量较小，在极少数法条对应案件数据上训练效果不佳导致宏指标偏低，随着数据集规模的提升，极少数法条对应案件数量的增加，效果也变得更好。在CAIL-small数据集上，刑期预测子任务中我们的宏精确率略微高于准确率，这是由于少量法条对应的样本往往刑期区间较大，所以分类的正确率反而会有所提高，然而在CAIL-big数据集上，随着数据规模的增大，模型在少量法条对应案件上效果有了提升，所以结果中宏精确率略微低于准确率。

方法	法条判断				刑期预测			
	Acc	MP	MR	F1	Acc	MP	MR	F1
CNN	95.79	82.79	75.15	76.62	55.41	45.23	38.73	39.96
FLA	93.22	72.81	64.27	66.57	57.66	43.01	38.89	41.63
Few-Shot	96.12	85.43	80.07	81.49	57.84	47.27	42.55	43.44
BERT	93.54	82.65	64.66	68.97	52.38	41.75	32.90	33.71
TOPJUDGE	95.81	84.41	74.36	76.67	57.29	47.35	42.61	44.03
LADAN	96.57	86.22	80.78	82.36	59.66	51.78	45.34	46.93
NeurJudge	96.19	84.75	78.88	80.56	57.67	51.72	45.77	46.96
GCLA	96.74	87.43	80.82	82.74	56.13	47.87	43.36	44.65
HD-LJP	96.81	88.68	82.08	84.19	60.35	52.66	50.04	50.17
LA-MGFM	97.98	88.97	87.21	87.95	63.06	54.29	52.68	53.56
DPFSI	**98.89**	91.64	**89.54**	90.67	64.06	53.98	52.84	53.76
lawLLM-7b	97.23	92.64	86.84	88.68	60.59	54.38	50.14	51.60
lawLLM-13b	97.39	92.55	87.82	89.65	62.95	56.26	53.61	54.48
DERF (Ours)	97.65	**93.43**	89.17	**90.85**	**65.32**	**57.44**	**54.98**	**55.79**

Table 3: CAIL-big上对比实验结果,对最好的结果加粗且次好结果用下划线标注。

4.4 消融实验

为了确定我们提出的双系统框架在刑期预测子任务上的提升是由某一特定方法还是所有方法的组合效应，我们也在CAIL-small数据集上进行了消融实验来分析。按照表4从上向下依次逆序的去除了我们在训练过程中所用的技术，实验结果如下：

方法	刑期预测			
	Acc	MP	MR	F1
DERF (Full)	**54.28**	**55.01**	**51.25**	**52.82**
W/O GRPO	51.22	52.88	49.17	50.71
W/O HQFR	46.55	47.18	42.39	43.37
W/O Law	39.43	46.52	37.85	34.10
only GRPO	37.76	36.60	34.53	35.62
only HQFR	36.26	34.53	37.30	35.60

Table 4: 消融实验结果，按照表格顺序从上至下依次逆序删除所用技术。

从表4中可以看到随着GRPO方法和在HQFR数据集上面预训练方法的去除，效果也逐渐下降，说明这两项技术可以提升模型在刑期预测任务上的准确率。当我们去除掉了由系统1给出的相关法条信息后准确率下降了**7.12%**,可见系统1在我们的双系统架构中也具有十分关键的作用。同时我们也测试了只用GRPO的方式学习而不加任何其他功能，效果有大幅下降，这说明模型在没有学到前提知识时就进行精炼结果并不理想。我们也测试了只用构建的小规模数据集来训练系统，并用全数据的测试集进行测试的结果依然有**36.26%**的准确率，与在全数据集上直接微调原始数据的结果(W/O Law)只相差了**3.17%**,可见我们的小规模数据集是有效的。

4.5 样例分析

为了进一步证明本文提出的双系统架构输出的可解释性，我们对刑期预测模型在小规模数据集上的输出结果进行了人工分析，分别选取包含“原告信息”和“涉案金额”的案例以及模型对其输出的具体解释进行展示，所选样例罪名也同样为生活中较为常见的“故意伤害罪”和“盗窃罪”。从表5中可以看出刑期预测模型在输出解释时兼顾到了案件要素、法律条文、量刑情节等多个方面，证明其具备了一定的解释能力。

5 结论与展望

针对目前方法在LJP任务中刑期预测子任务上的精度不高且可解释不强的问题，本文提出了**基于外部知识的可解释性双系统推理框架**，通过引入大模型的文本理解和生成能力来解决这一瓶颈。同时构建了一个小规模推理增强数据集HQFR来提高模型的推理能力和一个外部法条

解释1：“被告人王某因经济纠纷殴打黄某致轻伤二级，构成故意伤害罪。鉴于其系累犯，依法从重处罚；但家属已赔偿并获谅解，可酌情从轻。综合《刑法》第234条及累犯条款，判处8个月有期徒刑。”
解释2：“被告人杨某伙同他人实施盗窃，涉案金额5850元，构成盗窃罪，且属‘数额较大’范畴（1000-3000元以上），符合刑法第264条盗窃罪构成要件。鉴于其系初犯、无前科，且未造成人身伤害，判处9个月有期徒刑。”

Table 5: 模型输出的解释部分样例并将关键信息用不同颜色标注。

知识库来抑制“法条幻觉”。除此之外，为了验证我们方法的有效性我们还在CAIL-small数据集上进行了大量的实验，结果表明我们的方法显著提升了刑期预测子任务的精度和可解释性。

在未来工作中，我们将尝试使用一些预训练语言模型来尝试理解模型生成的推理过程，进一步量化测试模型生成推理的可解释性。此外，我们也将尝试将传统的基于图结构的方法与我们的双系统框架进行融合，新的方法也会在更多的数据集上进行测试。

参考文献

Stephen Bonner, Ibad Kureshi, John Brennan, Georgios Theodoropoulos, Andrew Stephen McGough, and Boguslaw Obara. 2019. Exploring the semantic content of unsupervised graph embeddings: An empirical study. *Data Science and Engineering*, 4:269–289.

Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. Legal-bert: The muppets straight out of law school. *arXiv preprint arXiv:2010.02559*.

Di Chen, Yiwei Bai, Wenting Zhao, Sebastian Ament, John M Gregoire, and Carla P Gomes. 2019a. Deep reasoning networks: Thinking fast and slow. *arXiv preprint arXiv:1906.00855*.

Huajie Chen, Deng Cai, Wei Dai, Zehui Dai, and Yadong Ding. 2019b. Charge-based prison term prediction with deep gating network. *arXiv preprint arXiv:1908.11521*.

Yahui Chen. 2015. Convolutional neural network for sentence classification. Master's thesis, University of Waterloo.

Daixuan Cheng, Shaohan Huang, and Furu Wei. 2024. Adapting large language models to domains via reading comprehension.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186.

Yao Dong, Xinran Li, Jin Shi, Yongfeng Dong, and Chen Chen. 2024. Graph contrastive learning networks with augmentation for legal judgment prediction. *Artificial Intelligence and Law*, pages 1–24.

Jonathan St BT Evans and Keith E Stanovich. 2013. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3):223–241.

Yao Guo, Yanling Li, Fengpei Ge, Haiqing Yu, Sukun Wang, and Zhongyi Miao. 2024. Legal judgment prediction via fine-grained element graphs and external knowledge. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.

Zikun Hu, Xiang Li, Cunchao Tu, Zhiyuan Liu, and Maosong Sun. 2018. Few-shot charge prediction with discriminative legal attributes. In *Proceedings of the 27th international conference on computational linguistics*, pages 487–498.

Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, et al. 2025. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2):1–55.

Daniel Kahneman. 2011. *Thinking, fast and slow*. macmillan.

Daniel Martin Katz, Michael J Bommarito II, and Josh Blackman. 2017. A general approach for predicting the behavior of the supreme court of the united states. *PloS one*, 12(4):e0174698.

Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Vladimir I Levenshtein et al. 1966. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710. Soviet Union.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Bingfeng Luo, Yansong Feng, Jianbo Xu, Xiang Zhang, and Dongyan Zhao. 2017. Learning to predict charges for criminal cases with legal basis. *arXiv preprint arXiv:1707.09168*.

L Thome McCarty. 2013. Reflections on taxman: An experiment in artificial intelligence and legal reasoning. In *Scientific Models of Legal Reasoning*, pages 145–202. Routledge.

Chunyun Meng, Yuki Todo, Cheng Tang, Li Luan, and Zheng Tang. 2025. Dpfsi: A legal judgment prediction method based on deontic logic prompt and fusion of law article statistical information. *Expert Systems with Applications*, page 126722.

Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, and Jianfeng Gao. 2021. Deep learning–based text classification: a comprehensive review. *ACM computing surveys (CSUR)*, 54(3):1–40.

Sudip Mittal, Anupam Joshi, and Tim Finin. 2017. Thinking, fast and slow: Combining vector spaces and knowledge graphs. *arXiv preprint arXiv:1708.03310*.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Y Wu, et al. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.

Chaojun Xiao, Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Zhiyuan Liu, Maosong Sun, Yansong Feng, Xianpei Han, Zhen Hu, Heng Wang, et al. 2018. Cail2018: A large-scale legal dataset for judgment prediction. *arXiv preprint arXiv:1807.02478*.

Chaojun Xiao, Xueyu Hu, Zhiyuan Liu, Cunchao Tu, and Maosong Sun. 2021. Lawformer: A pre-trained language model for chinese legal long documents. *AI Open*, 2:79–84.

Nuo Xu, Pinghui Wang, Long Chen, Li Pan, Xiaoyan Wang, and Junzhou Zhao. 2020. Distinguish confusing law articles for legal judgment prediction. *arXiv preprint arXiv:2004.02557*.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, et al. 2024. Qwen2. 5 technical report. *arXiv preprint arXiv:2412.15115*.

Linan Yue, Qi Liu, Binbin Jin, Han Wu, Kai Zhang, Yanqing An, Mingyue Cheng, Biao Yin, and Dayong Wu. 2021. Neurjudge: A circumstance-aware neural framework for legal judgment prediction. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 973–982.

Shengbin Yue, Shujun Liu, Yuxuan Zhou, Chenchen Shen, Siyuan Wang, Yao Xiao, Bingxuan Li, Yun Song, Xiaoyu Shen, Wei Chen, et al. 2024. Lawllm: Intelligent legal system with legal reasoning and verifiable retrieval. In *International Conference on Database Systems for Advanced Applications*, pages 304–321. Springer.

Yunong Zhang, Xiao Wei, and Hang Yu. 2024. Hd-ljp: A hierarchical dependency-based legal judgment prediction framework for multi-task learning. *Knowledge-Based Systems*, 299:112033.

第二十四届中国计算语言学大会论文集，第500页-第512页，济南，中国，2025年8月11日至14日。

Qihui Zhao, Tianhan Gao, and Nan Guo. 2023. La-mgfm: A legal judgment prediction method via sememe-enhanced graph neural networks and multi-graph fusion mechanism. *Information Processing & Management*, 60(5):103455.

Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. 2018a. Legal judgment prediction via topological learning. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pages 3540–3549.

Haoxi Zhong, Chaojun Xiao, Zhipeng Guo, Cunchao Tu, Zhiyuan Liu, Maosong Sun, Yansong Feng, Xianpei Han, Zhen Hu, Heng Wang, et al. 2018b. Overview of cail2018: legal judgment prediction competition. *arXiv preprint arXiv:1810.05851*.

6 附录

A.实验参数设置

在“快思考”阶段我们在全数据集上进行训练，实验设置：*learningrate*=5e-5，*batchsize*=2，*epoch*=4，*maxlength*=1048 tokens。之后测试法条判断模型在法条序号预测任务上的准确率，并且从法条知识库中检索出相关法条与模型生成的法条概述用Levenshtein Ratio方法进行比对，当两者相似度低于0.7 时视为产生了法条幻觉并进行法条内容的替换。

在“慢思考”阶段我们将构建的小规模高质量的思考增强数据集(HQFR)按照3:1的比例分成训练集和测试集，在此基础上进行第一阶段的lora微调，实验设置：*learningrate*=5e-5，*batchsize*=1，*epoch*=5，*maxlength*=1048 tokens。在全数据监督微调阶段，为了防止二次微调之后的模型对首次微调的能力出现灾难性遗忘，我们降低二次微调的学习率，实验设置：*learningrate*=5e-6，*batchsize*=1，*epoch*=5，*maxlength*=1048 tokens。在GRPO精炼阶段，我们选择用之前提到的奖励函数对二次微调之后的模型进行精度和可解释性方面的进一步训练。具体的训练部分则是利用GRPO来训练自注意力机制中的键投影层（K_{proj}）和值投影层（V_{proj}），以及前馈网络（FFN）中的门控投影层（G_{proj}）、升维投影层（U_{proj}）和降维投影层（D_{proj}）。