

# Can Small Language Models Handle Vietnamese Legal Reasoning? Insights from Multi-Task Evaluation

Long S. T. Nguyen, Hung C. Luu, Quynh T. N. Vo, Hy N. G. La, Hoai M. Tran,  
Anh T. D. Dinh, Tuan H. Nguyen, Tri N. Ho, Tho T. Quan\*

URA Research Group, Ho Chi Minh City University of Technology, VNU-HCM

\*Correspondence: qttho@hcmut.edu.vn

## Abstract

Small Language Models (SLMs) offer a practical but underexplored avenue for advancing Vietnamese legal Natural Language Processing (NLP) in resource-constrained environments. Vietnamese legal NLP is particularly challenging due to scarce annotated datasets, linguistic complexity, and the structural difficulty of statutory texts. To address these issues, we investigate two complementary adaptation strategies: continual pretraining on statutory corpora to improve domain retention, and instruction fine-tuning on task-specific datasets to enhance reasoning. Building on resources from the VLSP 2025 Challenge, we construct multi-task, instruction-style benchmarks spanning Multiple-Choice Question-Answering (MCQ), Natural Language Inference (NLI, including citation inference), and free-text Question-Answering (QA). Using the Qwen3-4B backbone, we systematically evaluate whether models under 4B parameters can achieve competitive performance on complex legal reasoning tasks. Our results highlight a clear trade-off: fine-tuning delivers substantial gains on discriminative tasks such as NLI and MCQ, whereas statute-grounded free-text generation remains a persistent challenge, particularly in long and compositional contexts.

## 1 Introduction

The rapid progress of *Large Language Models* (LLMs) such as ChatGPT (OpenAI et al., 2024), Gemini (Team et al., 2025), DeepSeek (DeepSeek-AI et al., 2025), and Qwen (Yang et al., 2025) has transformed *Natural Language Processing* (NLP). These models have expanded the scope of intelligent applications, including the legal domain, where accurate analysis of complex documents is critical (Le et al., 2025b; Qin and Sun, 2024; Ariai et al., 2025). Yet applying general-purpose LLMs to law remains difficult, as this domain demands specialized expertise, contextual reasoning,

and strict factual precision (To et al., 2024). Legal NLP systems are increasingly used for tasks such as summarization, *Question Answering* (QA), and citation analysis, which can streamline workflows, enhance access to justice, and support decision-making (Quevedo, Ernesto and Cerny, Tomas and Rodriguez, Alejandro and Rivas, Pablo and Yero, Jorge and Sooksatra, Korn and Zhakubayev, Alibek and Taibi, Davide, 2024). Substantial advances have been made in resource-rich languages through specialized models such as Legal-BERT (Chalkidis et al., 2020), KL3M (Bommarito II et al., 2025), and Legal-LM (Shi et al., 2024) for English, ChatLaw (Cui et al., 2023) and InternLM-Law (Fei et al., 2025) for Chinese, and COLIEE-related systems for Japanese (Nguyen et al., 2024a). By contrast, foundational research on Vietnamese legal text processing remains limited (Nguyen et al., 2025b; Saito et al., 2024; Wang et al., 2024).

This gap stems from several factors: the scarcity of annotated datasets, the linguistic complexity of Vietnamese (e.g., tonal variations and compound expressions), the structural difficulty of legal texts, and the lack of resources for large-scale training (To et al., 2024; Nguyen et al., 2025b,c, 2023). Recent initiatives such as VLQA (Nguyen et al., 2025b) (3,129 expert-annotated legal questions) and ViGPTQA (Nguyen et al., 2023) represent important progress, but they also highlight the need for more robust benchmarks and models tailored to Vietnamese legal reasoning (Bach et al., 2017; Ha et al., 2024; Nguyen et al., 2025a).

Meanwhile, many institutions in developing countries face strict computational constraints (Le et al., 2025a), making large-scale models impractical. This motivates a central research question: *Can small-to-medium language models achieve competitive performance on complex Vietnamese legal reasoning tasks under resource limitations?* With sizes of  $\leq 4B$  parameters, *Small Language Models* (SLMs) are computationally efficient, deploy-

able on edge devices, and accessible to diverse research teams (Ling et al., 2024). Prior work shows that domain-specific adaptation can yield substantial improvements, with specialized legal models achieving up to 25% accuracy gains on classification tasks (Li et al., 2025; Siino, 2025).

In this work, we extend the legal corpora released through the VLSP 2025 Challenge on Vietnamese Legal SLMs<sup>1</sup>, transforming them into task-specific datasets suitable for instruction fine-tuning and multi-task evaluation. Our key contributions are as follows.

- We construct multi-task datasets for free-text QA, *Natural Language Inference* (NLI), and *Multiple-Choice QA* (MCQ), extending the VLSP corpora with higher coverage, diversity, and quality.
- We fine-tune pretrained small models ( $\leq 4B$ ) with domain-specific strategies tailored to Vietnamese law, enabling effective adaptation to complex reasoning tasks under computational constraints.
- We benchmark these models against official VLSP datasets, providing a systematic analysis of their strengths and limitations. Notably, our findings reveal a trade-off: fine-tuning markedly improves discriminative reasoning tasks, while statute-grounded free-text generation remains a persistent challenge.

## 2 Related Work

### 2.1 Vietnamese Legal NLP Techniques

*Retrieval-Augmented Generation* (RAG) has become a central strategy for improving factual accuracy in QA systems by grounding outputs in external knowledge. In the Vietnamese legal domain, several studies adapted this paradigm with domain-specific innovations. (Ba et al., 2024) proposed a framework that refines preprocessing, improves retrieval fusion between lexical and dense methods, and incorporates active retrieval for reranking, yielding more accurate QA. Other works developed legal document query systems integrating LLMs with RAG to address specialized retrieval challenges (Ngo et al., 2025; Le et al., 2025a). (Ha et al., 2024) showed that fine-tuning compact LLMs on curated legal corpora can mitigate hallucination while remaining effective un-

der limited resources. Dataset construction has also been essential: ViBidLQA (Nguyen et al., 2025c), focused on procurement law, supports both *Machine Reading Comprehension* (MRC) and *Answer Generation* (AG), while multilingual retrieval corpora (Nguyen et al., 2025a) extend cross-lingual QA to low-resource settings. Earlier systems such as vLawyer (Duong and Ho, 2014), which adapted similarity-based models and Lucene, achieved around 70% precision and provided a baseline for later neural approaches. Collectively, these studies highlight the centrality of retrieval for Vietnamese legal QA, while also indicating retrieval alone cannot fully capture complex reasoning.

### 2.2 Language Models for Vietnamese Law

Motivated by the limitations of retrieval-centric approaches, recent work has turned toward adapting language models directly for Vietnamese law. Studies such as (Nguyen et al., 2024b) proposed pipelines that preprocess legal corpora before fine-tuning LLMs, leading to better retention of statutory knowledge and more consistent responses. ViGPTQA (Nguyen et al., 2023) advanced this direction by introducing one of the first instruction-tuned Vietnamese LLMs, accompanied by a benchmark dataset combining AI- and human-generated questions, which demonstrated competitive performance and underscored the value of dedicated Vietnamese models. In parallel, VLQA (Nguyen et al., 2025b) added over 3,000 expert-annotated legal questions, while other efforts on retrieval and classification (Bach et al., 2017; Ha et al., 2024) confirmed the benefits of domain-specific adaptation. Despite these advances, most contributions remain focused on single-task QA, leaving open questions about how Vietnamese legal LMs can be extended to broader reasoning tasks beyond QA.

### 2.3 Research Gaps and Positioning

Overall, prior research has advanced Vietnamese legal NLP by improving retrieval pipelines, developing domain-specific datasets, and introducing initial instruction-tuned LLMs. However, three critical gaps persist. First, existing work largely concentrates on retrieval optimization or single-task QA, while reasoning tasks such as NLI and MCQ remain underexplored. Second, there has been limited systematic investigation into SLMs under parameter constraints ( $\leq 4B$ ), despite their advantages in accessibility and deployment for

<sup>1</sup><https://vlsp.org.vn/vlsp2025/eval/legalSLM>

resource-limited institutions. Third, comprehensive multi-task evaluations that jointly assess QA, NLI, and MCQ to probe reasoning capabilities are still missing. Our work addresses these gaps by constructing multi-task legal datasets from statutory corpora and systematically fine-tuning compact models to assess their effectiveness in diverse Vietnamese legal reasoning tasks under realistic constraints, thereby providing new insights into the feasibility of SLMs for legal NLP.

### 3 Methodology

Our methodology consists of a two-stage pipeline: (i) *continual pretraining* to inject domain-specific knowledge into pretrained small language models, and (ii) *multi-task instruction fine-tuning* to adapt the models to diverse legal reasoning tasks. Let  $\mathcal{M}_\theta$  denote a pretrained small language model with parameters  $\theta$ , originally trained on general-domain corpora, and let  $\mathcal{D}_{\text{law}}$  represent the Vietnamese legal corpus used for specialization.

#### 3.1 Continual Pretraining

To align  $\mathcal{M}_\theta$  with the linguistic and semantic characteristics of Vietnamese statutory texts, we perform continual pretraining on  $\mathcal{D}_{\text{law}}$  using the *Masked Language Modeling* (MLM) objective. The MLM loss is calculated as Equation 1, where  $\mathcal{M}(x)$  denotes the set of masked positions in input sequence  $x$ .

$$\mathcal{L}_{\text{MLM}} = -\mathbb{E}_{x \sim \mathcal{D}_{\text{law}}} \sum_{i \in \mathcal{M}(x)} \log P_\theta(x_i | x_{\setminus i}). \quad (1)$$

This stage allows the model to internalize legal terminology, complex syntactic constructions, and domain-specific discourse patterns, thereby providing a robust initialization for downstream fine-tuning.

#### 3.2 Multi-task Instruction Fine-tuning

Following domain adaptation, we fine-tune  $\mathcal{M}_\theta$  on three supervised tasks that capture complementary aspects of legal reasoning:

**Free-text Legal QA (QA).** Given a legal context  $c$  and question  $q$ , the model generates an answer  $a$ . The generation objective is defined in Equation 2.

$$\mathcal{L}_{\text{QA}} = -\log P_\theta(a | c, q). \quad (2)$$

This task equips the model with the ability to produce accurate, coherent, and statute-grounded answers in open-form legal consultation.

**Legal Citation Usefulness (NLI).** We formalize citation usefulness as a binary variant of NLI. Given a premise  $p$  (legal citation) and a hypothesis  $h$  (legal question), the model predicts whether  $p$  provides evidence useful for answering  $h$ . The classification loss is expressed in Equation 3.

$$\mathcal{L}_{\text{NLI}} = - \sum_{y \in \{\text{useful}, \text{not-useful}\}} \mathbf{1}[y] \cdot \log P_\theta(y | p, h). \quad (3)$$

This task enforces logical alignment between statutory provisions and legal queries, thereby strengthening reasoning consistency.

**Multiple-Choice Legal QA (MCQ).** Given a legal question  $q$  and candidate answers  $\{a_1, \dots, a_k\}$ , the model selects the correct  $a^*$ . The formulation is given in Equation 4:

$$\mathcal{L}_{\text{MCQ}} = -\log P_\theta(a^* | q, \{a_i\}_{i=1}^k). \quad (4)$$

This task directly evaluates the model’s capacity to recall statutory knowledge and perform discriminative reasoning under constrained options.

All tasks are reformulated into instruction–response pairs and unified into a multi-task dataset  $\mathcal{D}_{\text{multi}} = \mathcal{D}_{\text{QA}} \cup \mathcal{D}_{\text{NLI}} \cup \mathcal{D}_{\text{MCQ}}$ . Training alternates mini-batches from different tasks, and the global objective is defined in Equation 5, where  $\lambda_{\text{QA}}, \lambda_{\text{NLI}}, \lambda_{\text{MCQ}}$  are task-specific weights:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{QA}} \mathcal{L}_{\text{QA}} + \lambda_{\text{NLI}} \mathcal{L}_{\text{NLI}} + \lambda_{\text{MCQ}} \mathcal{L}_{\text{MCQ}}. \quad (5)$$

Finally, the overall training can be viewed as a joint empirical risk minimization problem across heterogeneous legal reasoning tasks, formalized in Equation 6:

$$\min_{\theta} \mathbb{E}_{(x,y) \sim \mathcal{D}_{\text{multi}}} [\mathcal{L}_{\text{task}(x)}(\theta; x, y)]. \quad (6)$$

By combining continual pretraining and multi-task fine-tuning, our methodology equips compact legal language models with statutory knowledge, logical inference ability, and robust QA skills, while maintaining computational efficiency for deployment in resource-constrained environments.

## 4 Proposed Datasets

### 4.1 Initial Dataset Overview

The VLSP 2025 Challenge on Vietnamese Legal Language Models provides a comprehensive set of corpora supporting both pretraining and downstream evaluation. The resources are organized into two components.

**Pretraining Corpus.** This corpus includes: (i) a large-scale statutory collection of 96,770 documents encompassing constitutional provisions, civil and criminal codes, administrative regulations, and specialized statutes that form the backbone of Vietnam’s legal system; and (ii) 47,967 legal news articles and commentary pieces that capture contemporary discourse, judicial interpretations, and real-world applications of law. Together, these resources ensure balanced coverage of formal statutory language and practical contextual usage.

**Evaluation Sets.** For downstream validation, the challenge defines three supervised tasks that reflect complementary aspects of legal reasoning.

- *Legal Citation Usefulness.* This dataset contains 150 records, each consisting of a statutory excerpt, a specific legal inquiry, a meta-question of citation usefulness, binary choices (*Có/Yes, Không/No*), and the correct label. The task requires models not only to retrieve relevant statutes but also to assess their applicability. For example, provisions in the Law on Enterprises 2020 can address shareholder obligations but cannot resolve procedures for driver’s license applications. This benchmark therefore emphasizes filtering irrelevant evidence and reasoning about the sufficiency of legal citations.

- *Multiple-Choice Legal QA.* This dataset includes 146 carefully curated questions, each paired with four candidate answers (one correct option and three legally plausible distractors). The questions span diverse areas such as taxation, labor and social insurance, and administrative regulations. Since distractors are intentionally close to correct interpretations, the task provides an objective and discriminative benchmark for evaluating statutory knowledge and reasoning precision.

- *Free-Text Legal QA.* This dataset captures higher-level reasoning through case-based questions requiring structured narrative answers. Each record contains a question and a free-text response that typically follows a deductive reasoning pattern: (i) a general premise derived from statutes or decrees, (ii) a specific premise describing the factual scenario, and (iii) a conclusion applying the law to the facts. Covering domains such as land law, labor, social insurance, and healthcare, this dataset evaluates models’ ability to generate coherent, statute-grounded explanations, closely resembling the workflow of legal practitioners.

Collectively, these datasets form a robust foundation for our methodology. They provide complementary signals: factual grounding from citation usefulness, discriminative reasoning from multiple-choice QA, and generative legal analysis from free-text QA. This comprehensive benchmark allows us to rigorously evaluate the reasoning capacity of small legal language models under realistic constraints.

## 4.2 Data Cleaning and Filtering

Before fine-tuning, we implemented a two-stage pipeline to guarantee both structural consistency and legal validity of the pretraining corpus. The raw dataset, provided in heterogeneous HTML format, contained formatting noise as well as statutes that had been repealed or superseded. We therefore addressed these two issues systematically.

**HTML Cleaning.** We applied the `html2text`<sup>2</sup> converter to strip away markup, hyperlinks, and other formatting artifacts while preserving semantic content. For each record, we extracted the legal text (`doc_content`) and retained essential metadata fields (`DocIdentity`, `DocName`, `Id`, `IssueDate`, `OrganName`). The cleaned corpus was stored in JSONL format, with one normalized document per line, thereby facilitating consistent downstream processing.

**Filtering Outdated Statutes.** To avoid training on obsolete provisions, we developed an automated pipeline to detect and remove repealed laws:

- *Abrogation clause detection.* Each document was scanned for the section “*Điều i. Hiệu lực thi hành*” (“*Article i. Effectiveness*”), which typically lists statutes rendered invalid.

- *Name normalization and similarity matching.* We normalized document names by removing numerical identifiers and computed semantic similarity with Qwen3-Embedding-0.6B<sup>3</sup>. A similarity threshold of 0.6, chosen empirically, was used to distinguish direct predecessors (removed) from omnibus amendments (retained).

- *Iterative refinement.* All invalidated documents were logged under a new `Deprecated` field. Only statutes explicitly abrogated and exceeding the similarity threshold were removed.

This procedure eliminated 15,516 out of 95,468 documents ( $\approx 16.2\%$ ), yielding a curated corpus

<sup>2</sup><https://pypi.org/project/html2text/>

<sup>3</sup><https://huggingface.co/Qwen/Qwen3-Embedding-0.6B>

Table 1: Summary of constructed datasets, each aligned with a specific task type and supervision signal (examples in Appendix A).

Dataset	Task	Source(s)	Reasoning	Label/Output
qa_reasoning	Free-text QA	Statutes	Yes	Generated answer (+ reasoning trace)
qa_non_reasoning	Free-text QA	Statutes	No	Generated answer (final only)
nli_reasoning	Citation Usefulness	Statutes + News	Yes	{Yes/No} + explanation
nli_non_reasoning	Citation Usefulness	Statutes + News	No	{Yes/No} (label only)
mcq_reasoning	MCQ	Statutes + News	Yes	Option (A/B/C/D) + analysis
mcq_non_reasoning	MCQ	Statutes + News	No	Option (A/B/C/D)
legal_document_detection	Document Routing	Synthetic Stage I	N/A	Document name / identifier
legal_document_revision	Content Recall	Synthetic Stage I + Statutes	N/A	Statutory text span

Table 2: Dataset sizes and length statistics (measured in words).

Dataset	#Samples	Min Length	Max Length	Mean Length
qa_reasoning	29,992	193	906	402.07
qa_non_reasoning	29,992	144	735	278.97
nli_reasoning	30,123	189	94,659	832.60
nli_non_reasoning	30,123	95	94,508	728.83
mcq_reasoning	29,835	89	403	198.10
mcq_non_reasoning	29,835	42	189	81.23
legal_document_detection	30,000	67	348	126.08
legal_document_revision	30,000	96	921	304.92

of effective and up-to-date statutes.

**Outcome.** The resulting dataset is legally current, structurally normalized, and free of obsolete laws. By combining strict removal of repealed statutes with careful preservation of partially amended ones, our pipeline ensures both comprehensive coverage and reduced training noise. This guarantees that downstream models are trained on a corpus aligned with the effective Vietnamese legal framework.

### 4.3 Dataset Construction

We construct task-ready corpora from two primary sources—the curated Vietnamese statutory corpus and Vietnamese legal news articles—through a two-stage pipeline: (i) *Synthetic Dataset Generation*, where instruction-style prompting with Google Gemini systematically produces annotated samples, and (ii) *Secondary Dataset Transformation*, where the synthetic outputs are reorganized into additional task formats without further LLM generation.

#### Stage I: Synthetic Dataset Generation.

(a) *Legal Citation Usefulness.* A binary classification task using statutes and legal news. Each instance pairs a legal question with a statutory ex-

cerpt and a label indicating whether the excerpt provides a sufficient answer.

(b) *Multiple-Choice Legal QA.* Four-option questions generated from statutes and news, comprising one correct answer and three legally plausible distractors designed to require nuanced reasoning.

(c) *Free-Text Legal QA.* Case-based generative questions derived exclusively from statutes, producing narrative answers that follow structured legal reasoning patterns.

#### Stage II: Secondary Dataset Transformation.

(d) *Legal Document Detection.* Input: a legal question; Output: the most relevant document name or identifier, supporting domain classification and query routing.

(e) *Legal Document Revision.* Input: a document name (or provision reference); Output: faithful statutory content, testing models’ ability to accurately reproduce exact legal provisions.

As shown in Table 1, the eight datasets collectively span classification, multiple-choice, free-text generation, and document-centric tasks. Table 2 further summarizes dataset sizes and length statistics, confirming their large-scale coverage and diversity.

**Rationale.** Splitting QA, NLI, and MCQ into

reasoning vs. non-reasoning variants yields complementary supervision: the reasoning sets emphasize stepwise legal analysis, while the non-reasoning sets emphasize concise finalization. The two document-centric datasets introduce additional signals for query routing and faithful statutory recall. Together, these design choices provide a comprehensive training corpus that strengthens both reasoning depth and practical reliability of Vietnamese legal language models.

## 5 Experimentations

We investigate whether compact Vietnamese legal LMs benefit more from *continual pretraining* on statutes or from direct *instruction finetuning* on downstream tasks.

### 5.1 Datasets

We evaluate on the VLSP 2025 Legal Benchmark, which provides standardized test splits for three tasks: (i) *MCQ* with 146 questions, each containing one correct answer and three legally plausible distractors; (ii) *NLI* with 150 (*legal excerpt, question*) pairs annotated as {*Có/Yes, Không/No*}; and (iii) *QA* with 120 case-based free-form questions requiring statute-grounded narrative responses. These datasets jointly probe factual recall, discriminative judgment, and generative legal reasoning.

### 5.2 Evaluation Metrics

For classification tasks (MCQ and NLI), we use *accuracy*, defined in Equation 7.

$$\text{Acc} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}[\hat{y}_i = y_i], \quad (7)$$

where  $N$  is the dataset size,  $y_i$  the gold label, and  $\hat{y}_i$  the model prediction. For free-form QA, we adopt *Exact Match* (EM) with light normalization  $\nu(\cdot)$  (lowercasing, punctuation and whitespace removal), as shown in Equation 8.

$$\text{EM} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}[\nu(\hat{a}_i) = \nu(a_i)], \quad (8)$$

which measures the proportion of predictions that exactly match the reference answers after normalization.

### 5.3 Experimental Setup

**Backbone.** All experiments are conducted with the qwen3-4b backbone (Yang et al., 2025).

Qwen3 is the latest generation of the Qwen family, spanning 0.6B to 235B parameters across both dense and *Mixture-of-Experts* (MoE) architectures. A key feature is its unified framework for *thinking mode* (multi-step reasoning) and *non-thinking mode* (fast, context-driven responses), complemented by a *thinking budget* mechanism that adaptively allocates computation during inference. Compared to Qwen2.5, Qwen3 extends multilingual coverage from 29 to 119 languages while preserving strong efficiency in smaller variants. These characteristics make the 4B model an appropriate backbone for Vietnamese legal NLP under resource-constrained settings.

**Adaptation strategies.** We consider two strategies: (i) *Continual Pretraining*, which further pre-trains models on approximately 145k curated legal documents to inject statutory knowledge; (ii) *Finetuning*, which performs instruction tuning on single- or multi-task mixtures of 12k–30k samples, with explicit *reasoning* and *non-reasoning* variants.

**Training configuration.** Unless otherwise specified, training uses the AdamW (Loshchilov and Hutter, 2019) optimizer with learning rate  $2 \times 10^{-5}$ , batch size 64, and early stopping based on validation accuracy.

**Model naming.** Models follow the convention qwen3-4b-{tasks}-{reasoning flag}-{size}.

### 5.4 Results and Findings

Table 3 summarizes performance. Best scores per column are highlighted in **bold**.

**MCQ.** Finetuning consistently achieves the strongest MCQ performance. The non-reasoning variant reaches the highest accuracy of 0.9110, showing that concise, outcome-focused supervision is more effective than explicit reasoning traces for compact models.

**NLI.** Multi-task finetuning enhances NLI. Joint training on MCQ and NLI attains 0.9200, surpassing NLI-only training (0.9067). This suggests that option discrimination and citation usefulness provide complementary supervision signals.

**QA.** QA remains the most challenging task. Only CP models were evaluated, both scoring below 0.20 EM. This indicates that statute-grounded free-text generation requires either substantially larger QA corpora or retrieval-augmented finetuning. We did not evaluate finetuning-only models on QA in this study.

Table 3: Performance on VLSP 2025 legal tasks. “–” indicates models not trained or evaluated on QA.

Model	MCQ	NLI	QA (EM)
<i>Continual Pretraining</i>			
qwen3-4b-legal-pretrain-thinking	0.7808	0.9133	0.1806
qwen3-4b-legal-pretrain-non_thinking	0.8630	0.9000	0.1736
<i>Finetuning</i>			
qwen3-4b-mcq-reasoning-30k	0.8630	0.7667	–
qwen3-4b-mcq-non_reasoning-30k	<b>0.9110</b>	0.7400	–
qwen3-4b-nli-non_reasoning-30k	0.8836	0.9067	–
qwen3-4b-qa-non_reasoning-30k	0.8493	0.8933	–
qwen3-4b-legal_detection-30k	0.7877	0.8200	–
qwen3-4b-mcq-nli-qa-detection-revision-15k	0.8356	0.8667	–
qwen3-4b-mcq-nli-qa-detection-revision-13k	0.8630	0.9133	–
qwen3-4b-vietnamese_legal_documents_revision-30k	0.8356	0.7600	–
qwen3-4b-mcq-nli-detection-15k	0.8356	<b>0.9200</b>	–
qwen3-4b-mcq-nli-qa-detection-15k	0.7877	0.6933	–
qwen3-4b-mcq-nli-12k	0.8699	<b>0.9200</b>	–

**Progression of Experiments.** The results follow an iterative trajectory. (i) Baseline benchmarking with CP models (thinking and non-thinking modes) yielded competitive MCQ and NLI but exposed QA weaknesses, often due to multi-document reasoning beyond our synthetic pipeline. (ii) Single-task finetuning revealed that non-thinking supervision was consistently stronger, with NLI-focused training delivering the best transfer effects. (iii) Multi-task finetuning showed that MCQ+NLI training achieved the highest overall accuracy (0.9200 for NLI), while adding QA degraded results and auxiliary tasks (detection, revision) provided only limited gains (e.g., NLI 0.8667).

**Summary.** For models under 4B parameters, targeted finetuning provides the largest gains on discriminative tasks (MCQ, NLI). In contrast, statute-grounded free-text QA remains a bottleneck, motivating future work on hybrid strategies that combine multi-task training with retrieval-augmented generation and larger QA datasets.

## 6 Conclusion

We presented the first systematic evaluation of SLMs for Vietnamese legal reasoning under realistic resource constraints, combining continual pretraining on statutory corpora with multi-task instruction fine-tuning across MCQ, NLI, and free-text QA. Our results show that compact models can deliver competitive performance on discriminative

reasoning tasks, while statute-grounded free-text QA remains a persistent challenge in long and compositional contexts. These findings demonstrate that domain-specific fine-tuning is the most effective strategy for sub-4B models on classification tasks, whereas reliable generation requires complementary mechanisms. In practice, SLMs provide a viable solution for institutions with limited compute to deploy accurate, low-latency tools for legal reasoning. However, generative assistants should integrate retrieval-augmented methods, richer QA corpora, and balanced curricula to ensure factuality and compliance. Our study is limited by the scope of generative fine-tuning and the reliance on automatic metrics, which under-represent explanation quality and safety. Future work will expand human evaluation of faithfulness, explore parameter-efficient tuning and thinking-budget control, and develop hybrid strategies that combine retrieval-augmented generation with multi-task learning. Overall, SLMs offer a feasible and effective path for advancing Vietnamese legal NLP on discriminative reasoning, while closing the gap in statute-grounded generation remains a central challenge for building trustworthy legal assistants.

## References

Farid Ariai, Joel Mackenzie, and Gianluca Demartini. 2025. [Natural Language Processing for the Legal](#)

**Domain: A Survey of Tasks, Datasets, Models, and Challenges.**

- Thiem Nguyen Ba, Vinh Doan The, Tung Pham Quang, and Toan Tran Van. 2024. Vietnamese Legal Information Retrieval in Question-Answering System.
- Ngo Xuan Bach, Le Thi Ngoc Cham, Tran Ha Ngoc Thien, and Tu Minh Phuong. 2017. [Question analysis for Vietnamese legal question answering](#). In *2017 9th International Conference on Knowledge and Systems Engineering (KSE)*, pages 154–159.
- Michael J Bommarito II, Jillian Bommarito, and Daniel Martin Katz. 2025. The KL3M Data Project: Copyright-Clean Training Resources for Large Language Models.
- Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. [LEGAL-BERT: The Muppets straight out of Law School](#). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2898–2904, Online. Association for Computational Linguistics.
- Jiaxi Cui, Munan Ning, Zongjian Li, Bohua Chen, Yang Yan, Hao Li, Bin Ling, Yonghong Tian, and Li Yuan. 2023. [ChatLaw: Open-Source Legal Large Language Model with Integrated External Knowledge Bases](#).
- DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, and 181 others. 2025. [DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning](#).
- Huu-Thanh Duong and Bao-Quoc Ho. 2014. A Vietnamese Question Answering System in Vietnam’s Legal Documents. In *Computer Information Systems and Industrial Management*, pages 186–197, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Zhiwei Fei, Songyang Zhang, Xiaoyu Shen, Dawei Zhu, Xiao Wang, Jidong Ge, and Vincent Ng. 2025. [InternLM-Law: An Open-Sourced Chinese Legal Large Language Model](#). In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 9376–9392, Abu Dhabi, UAE. Association for Computational Linguistics.
- Nguyen Thu Ha, Truong-Phuc Nguyen, Khang T. Trung, Huu-Loi Le, Le Thi Viet Huong, Chi Thanh Nguyen, and Minh-Tien Nguyen. 2024. [Vietnamese Legal Question Answering: An Experimental Study](#). In *2024 16th International Conference on Knowledge and System Engineering (KSE)*, pages 440–446.
- Huong Le, Ngoc Luu, Thanh Nguyen, Tuan Dao, and Sang Dinh. 2025a. [Optimizing Answer Generator in Vietnamese Legal Question Answering Systems Using Language Models](#). *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 24(6).
- Van-Hoang Le, Duc-Vu Nguyen, Kiet Van Nguyen, and Ngan Luu-Thuy Nguyen. 2025b. [Optimizing Legal Document Retrieval in Vietnamese with Semi-Hard Negative Mining](#).
- Haitao Li, Qingyao Ai, Jia Chen, Qian Dong, Zhi-jing Wu, and Yiqun Liu. 2025. [BLADE: Enhancing Black-Box Large Language Models with Small Domain-Specific Models](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(23):24422–24430.
- Chen Ling, Xujiang Zhao, Jiaying Lu, Chengyuan Deng, Can Zheng, Junxiang Wang, Tanmoy Chowdhury, Yun Li, Hejie Cui, Xuchao Zhang, Tianjiao Zhao, Amit Panalkar, Dhagash Mehta, Stefano Pasquali, Wei Cheng, Haoyu Wang, Yanchi Liu, Zhengzhang Chen, Haifeng Chen, and 5 others. 2024. [Domain Specialization as the Key to Make Large Language Models Disruptive: A Comprehensive Survey](#).
- Ilya Loshchilov and Frank Hutter. 2019. [Decoupled Weight Decay Regularization](#). In *International Conference on Learning Representations*.
- Anh Tuan Ngo, Trung Tung Doan, Tung Thanh Ngo, and Viet Hoang Nguyen. 2025. [Legal Documents Query Application for Vietnamese Law Using LLM and RAG Techniques](#). In *Proceedings of the 2025 10th International Conference on Intelligent Information Technology, ICIIT ’25*, page 152–157, New York, NY, USA. Association for Computing Machinery.
- Chau Nguyen, Phuong Nguyen, Thanh Tran, Dat Nguyen, An Trieu, Tin Pham, Anh Dang, and Le-Minh Nguyen. 2024a. [Captain at coliee 2023: Efficient methods for legal information retrieval and entailment tasks](#).
- Long S. T. Nguyen, Truong P. Hua, Thanh M. Nguyen, Toan Q. Pham, Nam K. Ngo, An X. Nguyen, Nghi D. M. Pham, Nghia H. Nguyen, and Tho T. Quan. 2025a. [A Benchmark Dataset and Evaluation Framework for Vietnamese Large Language Models in Customer Support](#).
- Minh Thuan Nguyen, Khanh Tung Tran, Nhu Van Nguyen, and Xuan-Son Vu. 2023. [ViGPTQA - State-of-the-Art LLMs for Vietnamese Question Answering: System Overview, Core Models Training, and Evaluations](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 754–764, Singapore. Association for Computational Linguistics.
- Nguyen P. Nguyen, Thang V.Q. Le, Anh-Cuong Le, Viet-Ha Nguyen, and Viet-Cuong Nguyen. 2024b. [Adapting large language models to vietnamese law: Pretrained llm refinement vs retrieval augmented generation](#). In *2024 16th International Conference on Knowledge and System Engineering (KSE)*, pages 01–06.
- Tan-Minh Nguyen, Hoang-Trung Nguyen, Trong-Khoi Dao, Xuan-Hieu Phan, Ha-Thanh Nguyen, and Thi-Hai-Yen Vuong. 2025b. [VLQA: The First Com-](#)



prehensive, Large, and High-Quality Vietnamese Dataset for Legal Question Answering.

Toan Ngoc Nguyen, Nam Le Hai, Nguyen Doan Hieu, Dai An Nguyen, Linh Ngo Van, Thien Huu Nguyen, and Sang Dinh. 2025c. [Improving Vietnamese-English Cross-Lingual Retrieval for Legal and General Domains](#). In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pages 142–153, Albuquerque, New Mexico. Association for Computational Linguistics.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, and 262 others. 2024. [GPT-4 Technical Report](#).

Weicong Qin and Zhongxiang Sun. 2024. [Exploring the Nexus of Large Language Models and Legal Systems: A Short Survey](#).

Quevedo, Ernesto and Cerny, Tomas and Rodriguez, Alejandro and Rivas, Pablo and Yero, Jorge and Sooksatra, Korn and Zhakubayev, Alibek and Taibi, Davide. 2024. [Legal natural language processing from 2015 to 2022: A comprehensive systematic mapping study of advances and applications](#). *IEEE Access*, 12:145286–145317.

Koshiro Saito, Sakae Mizuki, Masanari Ohi, Taishi Nakamura, Taihei Shiotani, Koki Maeda, Youmi Ma, Kakeru Hattori, Kazuki Fujii, Takumi Okamoto, Shigeki Ishida, Hiroya Takamura, Rio Yokota, and Naoaki Okazaki. 2024. [Why We Build Local Large Language Models: An Observational Analysis from 35 Japanese and Multilingual LLMs](#).

Juanming Shi, Qinglang Guo, Yong Liao, Yuxing Wang, Shijia Chen, and Shenglin Liang. 2024. [Legal-LM: Knowledge Graph Enhanced Large Language Models for Law Consulting](#). In *Advanced Intelligent Computing Technology and Applications: 20th International Conference, ICIC 2024, Tianjin, China, August 5–8, 2024, Proceedings, Part IV*, page 175–186, Berlin, Heidelberg. Springer-Verlag.

Marco Siino. 2025. [Exploring the use of LLMs in the Italian legal domain: A survey on recent applications](#). *Computer Law & Security Review*, 58:106164.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M. Dai, Anja Hauth, Katie Millican, David Silver, Melvin Johnson, Ioannis Antonoglou, Julian Schrittwieser, Amelia Glaese, Jilin Chen, Emily Pitler, Timothy Lillicrap, Angeliki Lazaridou, and 1332 others. 2025. [Gemini: A Family of Highly Capable Multimodal Models](#).

Long Truong To, Hung Tuan Le, Dat Van-Thanh Nguyen, Manh Trong Nguyen, Tri Thien Nguyen,

Tin Van Huynh, and Kiet Van Nguyen. 2024. [Evaluating Large Language Model Capability in Vietnamese Fact-Checking Data Generation](#).

Yizhen Wang, Xueying Shen, Zixian Huang, Lihui Niu, and Shiyan Ou. 2024. [cLegal-QA: a Chinese legal question answering with natural language generation methods](#). *Complex & Intelligent Systems*, 11(1):77.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025. Qwen3 technical report.

## A Dataset Illustrations

To illustrate the supervision signals and instruction formats introduced in Section 4, this appendix presents representative examples of five dataset families. Datasets with the suffix `_reasoning` include an explicit chain-of-thought trace, delimited by `<think>...</think>` and displayed in a grey box, while their `_non_reasoning` counterparts follow the same input format but omit this trace, returning only the final answer. Samples are provided for free-text QA (`qa`, Table 4), NLI (`nli`, Table 5), MCQ (`mcq`, Table 6), Legal Document Detection (`legal_detection`, Table 7), and Legal Document Revision (`legal_revision`, Table 8). Together, these examples showcase the breadth of tasks in our benchmark—from binary classification and multiple-choice discrimination to statute-grounded free-text generation—and highlight how contrasting reasoning versus non-reasoning settings enables a systematic evaluation of different dimensions of legal reasoning.

Table 4: Overview of the qa\_reasoning dataset. The qa\_non\_reasoning dataset follows the same format but omits the grey box (no reasoning trace).

---

**Vietnamese:**

```
<|im_start|>user
### Instruction
Bạn là một luật sư nhiều năm kinh nghiệm, nắm vững Pháp luật Việt Nam. Nhiệm vụ của bạn là sử dụng kiến thức luật của bản thân để trả lời câu hỏi cho tình huống đặt ra.
### Response Format
Hãy trả về theo format sau:
'Tiền đề lớn: ... Tiền đề nhỏ: ... Kết luận: ...'
### Giải thích các trường thông tin
**Tiền đề lớn**: Quy định nào liên quan? Nội dung của quy định đó như thế nào?
**Tiền đề nhỏ**: Áp dụng vào trường hợp trong câu hỏi
**Kết luận**: Kết luận cuối cùng theo yêu cầu của câu hỏi
### Question
Giả sử Hệ thống E-Office của Bộ Giáo dục và Đào tạo gặp sự cố, đơn vị nào có trách nhiệm thông báo cho các đơn vị khác về sự cố này và kết quả khắc phục?
<|im_end|>
<|im_start|>assistant
```

```
<think>
Bước 1: Vấn đề của câu hỏi là xác định đơn vị chịu trách nhiệm thông báo về sự cố và khắc phục của Hệ thống E-Office.
Bước 2: Điều 17 của Quyết định 388/QĐ-BGDĐT quy định rõ trách nhiệm của Cục Công nghệ thông tin, bao gồm việc thông báo về sự cố và kết quả khắc phục.
Bước 3: Áp dụng quy định vào tình huống, Cục Công nghệ thông tin là đơn vị chịu trách nhiệm thông báo.
Bước 4: Kết luận, Cục Công nghệ thông tin có trách nhiệm thông báo cho các đơn vị khác về sự cố và kết quả khắc phục.
</think>
```

Tiền đề lớn: Theo Điều 17 của Quyết định 388/QĐ-BGDĐT, Cục Công nghệ thông tin có trách nhiệm duy trì Hệ thống E-Office hoạt động thường xuyên, thông suốt; thông báo kịp thời các đơn vị biết khi Hệ thống E-Office gặp sự cố và kết quả khắc phục sự cố.

Tiền đề nhỏ: Hệ thống E-Office gặp sự cố.

Kết luận: Cục Công nghệ thông tin có trách nhiệm thông báo cho các đơn vị khác về sự cố và kết quả khắc phục.

```
<|im_end|>
```

---

**English:**

```
<|im_start|>user
### Instruction
You are an experienced lawyer, well-versed in Vietnamese law. Your task is to use your legal expertise to answer the given case.
### Response Format
Please return the answer in the following format:
'Major premise: ... Minor premise: ... Conclusion: ...'
### Explanation of the fields
**Major premise**: Which legal regulation is relevant? What does it stipulate?
**Minor premise**: Application of the regulation to the given case.
**Conclusion**: The final conclusion that directly addresses the question.
### Question
Suppose the E-Office system of the Ministry of Education and Training encounters a failure. Which unit is responsible for informing other departments about the incident and the resolution?
<|im_end|>
<|im_start|>assistant
```

```
<think>
Step 1: The question requires identifying the unit responsible for reporting incidents and their resolution in the E-Office system.
Step 2: Article 17 of Decision 388/QĐ-BGDĐT explicitly assigns this responsibility to the Department of Information Technology, including notifying other departments of the incident and its resolution.
Step 3: Applying this regulation to the case, the Department of Information Technology is the responsible unit.
Step 4: Conclusion: The Department of Information Technology is responsible for informing other departments about the incident and the resolution.
</think>
```

Major premise: According to Article 17 of Decision 388/QĐ-BGDĐT, the Department of Information Technology is responsible for ensuring the E-Office system operates continuously and smoothly, and for promptly notifying other units in the event of a failure together with its resolution.

Minor premise: The E-Office system encounters a failure.

Conclusion: The Department of Information Technology is responsible for informing other units about the incident and its resolution.

```
<|im_end|>
```

---

Table 5: Overview of the nli\_reasoning dataset. The nli\_non\_reasoning dataset follows the same format but omits the grey box (no reasoning trace).

---

**Vietnamese:**  
<|im\_start|>user  
### Yêu cầu  
Điều luật được cung cấp có thể dùng để trả lời câu hỏi trên hay không? Hãy trả về "Có" hoặc "Không"  
### Legal Document  
Quyết định 1101/QĐ-UBND của Ủy ban nhân dân tỉnh Thanh Hóa về việc công bố thủ tục hành chính được sửa đổi, bổ sung trong lĩnh vực Đất đai thuộc thẩm quyền giải quyết của Văn phòng Đăng ký đất đai trực thuộc Sở Tài nguyên và Môi trường  
Điều 3 Chánh Văn phòng UBND tỉnh, Giám đốc Sở Tài nguyên và Môi trường; Giám đốc Văn phòng Đăng ký đất đai Thanh Hoá; Cục Thuế tỉnh; Chủ tịch UBND các huyện, thị xã, thành phố; Chủ tịch UBND các xã, phường, thị trấn và Thủ trưởng các cơ quan, tổ chức, cá nhân có liên quan chịu trách nhiệm thi hành Quyết định này./.  
Nơi nhận: - Như Điều 3 Quyết định; - Cục Kiểm soát TTTHC-VPCP (bản điện tử); - Chủ tịch UBND tỉnh (để báo cáo); - Trung tâm Phục vụ hành chính công tỉnh; - Cổng thông tin điện tử tỉnh (để đăng tải); - Lưu: VT, KSTTHCNC. | KT. CHỦ TỊCH PHÓ CHỦ TỊCH Nguyễn Văn Thi  
### Question  
Theo Quyết định 1101/QĐ-UBND của Ủy ban nhân dân tỉnh Thanh Hóa, những cơ quan, tổ chức, cá nhân nào chịu trách nhiệm thi hành quyết định này?  
<|im\_end|>  
<|im\_start|>assistant  
<think>  
Bước 1: Câu hỏi yêu cầu xác định các đối tượng chịu trách nhiệm thi hành Quyết định 1101/QĐ-UBND.  
Bước 2: Điều 3 của Quyết định 1101/QĐ-UBND liệt kê rõ ràng các cơ quan, tổ chức, cá nhân chịu trách nhiệm thi hành, bao gồm Chánh Văn phòng UBND tỉnh, Giám đốc Sở Tài nguyên và Môi trường, Giám đốc Văn phòng Đăng ký đất đai Thanh Hoá, Cục Thuế tỉnh, Chủ tịch UBND các huyện, thị xã, thành phố, Chủ tịch UBND các xã, phường, thị trấn và Thủ trưởng các cơ quan, tổ chức, cá nhân có liên quan.  
Bước 3: Kết luận: Câu hỏi có thể được trả lời dựa trên thông tin được cung cấp trong văn bản pháp luật.  
</think>  
Có  
<|im\_end|>

---

**English:**  
<|im\_start|>user  
### Task  
Can the provided legal document be used to answer the above question? Please return "Yes" or "No".  
### Legal Document  
Decision 1101/QĐ-UBND of the People's Committee of Thanh Hóa Province on the promulgation of amended and supplemented administrative procedures in the field of Land Management under the jurisdiction of the Land Registration Office, Department of Natural Resources and Environment.  
Article 3: The Chief of the Office of the Provincial People's Committee, the Director of the Department of Natural Resources and Environment, the Director of the Thanh Hóa Land Registration Office, the Provincial Tax Department, the Chairpersons of district, town, and city People's Committees, the Chairpersons of commune, ward, and township People's Committees, and the heads of relevant agencies, organizations, and individuals shall be responsible for implementing this Decision.  
Recipients: - As stated in Article 3; - Administrative Procedure Control Agency, Government Office (electronic copy); - Chairman of the Provincial People's Committee (for reporting); - Provincial Public Administration Service Center; - Provincial e-portal (for publication); - Filed at: VT, KSTTHCNC. | On behalf of the CHAIRMAN, VICE CHAIRMAN Nguyễn Văn Thi  
### Question  
According to Decision 1101/QĐ-UBND of the People's Committee of Thanh Hóa Province, which agencies, organizations, and individuals are responsible for implementing this decision?  
<|im\_end|>  
<|im\_start|>assistant  
<think>  
Step 1: The question requires identifying the entities responsible for implementing Decision 1101/QĐ-UBND.  
Step 2: Article 3 of Decision 1101/QĐ-UBND clearly lists the responsible entities, including the Chief of the Office of the Provincial People's Committee, the Director of the Department of Natural Resources and Environment, the Director of the Thanh Hóa Land Registration Office, the Provincial Tax Department, the Chairpersons of district, town, and city People's Committees, the Chairpersons of commune, ward, and township People's Committees, and the heads of relevant agencies, organizations, and individuals.  
Step 3: Conclusion: The question can be answered based on the provided legal document.  
</think>  
Yes  
<|im\_end|>

---

Table 6: Overview of the mcq\_reasoning dataset. The mcq\_non\_reasoning dataset follows the same format but omits the grey box (no reasoning trace).

---

**Vietnamese:**

<|im\_start|>user  
### Yêu cầu  
Hãy chọn đáp án đúng nhất cho câu hỏi sau. Hãy chỉ trả lời A/B/C/D  
### Question  
Yêu cầu nghiên cứu trong Quy hoạch xây dựng vùng tỉnh Gia Lai bao gồm điều gì?  
### Options  
A. Nghiên cứu tác động ngoại vùng và nội vùng  
B. Chỉ nghiên cứu tác động nội vùng  
C. Không cần nghiên cứu tác động  
D. Chỉ cần nghiên cứu về dân số  
<|im\_end|>  
<|im\_start|>assistant

<think>  
Câu hỏi này yêu cầu xác định yêu cầu nghiên cứu trong Quy hoạch xây dựng vùng tỉnh Gia Lai. Theo nội dung quyết định, yêu cầu nghiên cứu bao gồm cả tác động ngoại vùng và nội vùng, nhằm xác định các ngành, lĩnh vực có tác động đến vai trò và vị thế của tỉnh Gia Lai. Phân tích các lựa chọn: A đúng vì nó bao quát cả hai khía cạnh; B sai vì không chỉ nghiên cứu nội vùng; C sai vì nghiên cứu tác động là cần thiết; D không đầy đủ vì chỉ nghiên cứu về dân số là không đủ. Do đó, đáp án đúng là A.  
</think>

A  
<|im\_end|>

---

**English:**

<|im\_start|>user  
### Instruction  
Choose the best answer to the following question. Reply with A/B/C/D only.  
### Question  
What does the research requirement in Gia Lai Province's Regional Construction Planning include?  
### Options  
A. Study both external and internal impacts  
B. Study only internal impacts  
C. No need to study impacts  
D. Only study population aspects  
<|im\_end|>  
<|im\_start|>assistant

<think>  
This question asks about the research requirements in Gia Lai Province's Regional Construction Planning. According to the decision, the research must include both external and internal impacts, to identify sectors and fields affecting the province's role and position. Option A is correct since it covers both; B is wrong since it only covers internal impacts; C is wrong since impact studies are necessary; D is incomplete since focusing only on population is insufficient. Therefore, the correct answer is A.  
</think>

A  
<|im\_end|>

---

Table 7: Overview of the legal\_detection dataset.

---

**Vietnamese:**

<|im\_start|>user

### Instruction

Bạn là một luật sư nhiều năm kinh nghiệm, nắm vững Pháp luật Việt Nam, bạn có khả năng ghi nhớ toàn bộ văn bản pháp luật. Nhiệm vụ của bạn là xác định câu hỏi cần sử dụng quy định nào?

### Câu hỏi

Quyết định 1344/QĐ-UBND của Ủy ban nhân dân tỉnh Lai Châu có đề cập đến các biện pháp xử lý vi phạm hành chính trong lĩnh vực trồng trọt hay không?

<|im\_end|>

<|im\_start|>assistant

Quyết định 1344/QĐ-UBND của Ủy ban nhân dân tỉnh Lai Châu về việc phê duyệt quy trình nội bộ giải quyết thủ tục hành chính trong lĩnh vực trồng trọt thuộc phạm vi chức năng quản lý của Sở Nông nghiệp và Phát triển nông thôn tỉnh Lai Châu

<|im\_end|>

---

**English:**

<|im\_start|>user

### Instruction

You are an experienced lawyer, highly knowledgeable about Vietnamese law, with the ability to recall all legal documents. Your task is to determine which legal provision should be applied to the given question.

### Question

Does Decision 1344/QĐ-UBND of the People’s Committee of Lai Châu Province mention administrative violation handling measures in the field of cultivation?

<|im\_end|>

<|im\_start|>assistant

Decision 1344/QĐ-UBND of the People’s Committee of Lai Châu Province on approving the internal process for handling administrative procedures in the field of cultivation within the management scope of the Department of Agriculture and Rural Development of Lai Châu Province

<|im\_end|>

---

Table 8: Overview of the legal \_revision dataset. The dataset provides precise legal provisions as answers, with reasoning traces enclosed in a grey box.

---

**Vietnamese:**

<|im\_start|>user

### Instruction

Bạn là một luật sư nhiều năm kinh nghiệm, nắm vững Pháp luật Việt Nam, bạn có khả năng ghi nhớ toàn bộ văn bản pháp luật. Nhiệm vụ của bạn là trả lời chính xác quy định được hỏi.

### Question

Điều 27 - Xử lý đối với trường hợp hàng hoá, vật phẩm nhập khẩu quá thời hạn làm thủ tục hải quan hoặc buộc phải đưa ra khỏi Việt Nam hay tái xuất, Nghị định 16/CP của Chính phủ về việc quy định việc xử phạt vi phạm hành chính trong lĩnh vực quản lý Nhà nước về hải quan

<|im\_end|>

<|im\_start|>assistant

<think>

Bước 1: Câu hỏi yêu cầu trích dẫn chính xác điều luật quy định về xử lý hàng hoá nhập khẩu quá thời hạn.

Bước 2: Điều 27 của Nghị định 16/CP quy định rõ về hai trường hợp: (i) quá thời hạn làm thủ tục, và (ii) quá thời hạn tái xuất hoặc đưa hàng ra khỏi Việt Nam.

Bước 3: Vì vậy cần trả lời nguyên văn quy định.

</think>

1. Hàng hoá, vật phẩm nhập khẩu quá thời hạn 60 ngày kể từ sau thời hạn cuối cùng phải làm thủ tục hải quan theo quy định của pháp luật hoặc quá 30 ngày, kể từ ngày cơ quan Hải quan thông báo trên phương tiện thông tin đại chúng mà không có người đến làm thủ tục hải quan thì làm thủ tục sung công quỹ Nhà nước.

2. Mọi trường hợp hàng hoá theo quy định phải tái xuất hoặc buộc phải đưa ra khỏi Việt Nam, nếu quá thời hạn quy định trong quyết định xử phạt mà đương sự không thực hiện sẽ bị sung công quỹ.

<|im\_end|>

---

**English:**

<|im\_start|>user

### Instruction

You are an experienced lawyer, highly knowledgeable about Vietnamese law, with the ability to recall all legal documents. Your task is to provide the exact legal provision being asked about.

### Question

Article 27 – Handling of imported goods or items that exceed the time limit for customs procedures, or that must be removed from Vietnam or re-exported, Decree 16/CP of the Government on administrative sanctions in the field of state management of customs.

<|im\_end|>

<|im\_start|>assistant

<think>

Step 1: The question requires citing the exact provision regarding overdue imported goods.

Step 2: Article 27 of Decree 16/CP clearly specifies two cases: (i) exceeding the customs deadline, and (ii) exceeding the re-export/removal deadline.

Step 3: Therefore, the full legal text must be returned verbatim.

</think>

1. Imported goods or items that exceed the time limit of 60 days from the final deadline for completing customs procedures under the law, or exceed 30 days from the date the Customs authority announces publicly without anyone coming forward to complete customs procedures, shall be confiscated into the State Treasury.

2. In all cases where goods are required by law to be re-exported or removed from Vietnam, if the sanctioned party fails to comply within the prescribed time limit, the goods shall be confiscated into the State Treasury.

<|im\_end|>

---