SHOULD BE REPLACED ON REQUIRED TITLE PAGE

Instruction

1. Open needed docx template (folder "title"/<your department or bach if bachelor student>.docx).

2. Put Thesis topic, supervisor's and your name in appropriate places on both English and Russian languages.

3. Put current year (last row).

4. Convert it to "title.pdf," replace the existing one in the root folder.

Contents

List of Tables

List of Figures

Abstract

My abstract starts from here.

Chapter 1

Introduction

1.1 Early Approaches to Automation in the Legal Sphere

...

Chapter 2

Literature Review

2.1 Early Approaches to Automation in the Legal Sphere

The application of Artificial Intelligence in the legal domain began in 1958 with Lucien Mehl's seminal article [1], which explored AI's potential in law and decision-making processes. Mehl proposed two primary approaches for legal AI systems:

- *Information machine* - "the machine for finding precedent"
- *Consultation machine* - "the judgement machine"

The author argued that both approaches required transforming natural language in legal documents into a strictly formalized, machine-readable format. This task was considered a preliminary step toward formalizing natural language as a whole. While Mehl proposed a framework for legal text formalization, this remains an unsolved challenge today.[1]

[1] Modern projects like [2] continue to explore machine-readable formats for legislative texts.

Early probabilistic approaches to natural language processing emerged in 1952 [3], but were deemed impractical due to computational constraints. Only in 1990 did IBM introduce efficient, computationally feasible methods [4]. A 2003 review [5] confirmed that case-based and rule-based reasoning systems remained the most prevalent AI applications in law at that time.

2.2 Large Language Models in Law

Advances in computational power enabled opportunity to utilize complex natural language processing models. The field transformed with Word2Vec's semantically meaningful word vectors (2013) [6], followed by Google's Transformer architecture (2017) [7]. OpenAI's 2020 paper [8] introduced few-shot learning through prompting, leading to powerful, accessible AI assistants applicable to legal domains.

A 2024 Texas survey [9] revealed that over 50% of respondents would consult AI assistants for tenancy, tax, and traffic legal matters.[2] Similarly, LexisNexis [10] reported 35% of lawyers using AI assistants monthly.

Specialized legal LLMs like LawGPT (2023) [11], LawyerLLaMA (2023) [12], and ChatLaw (2023) [13] have emerged.

[2]However, only ≈30% trusted AI for divorce, juvenile, or civil dispute cases.

2.3 Limitations of Language Models in the Legal Domain

Despite the active use of Large Language Models (LLMs) by legal professionals, their application to a broad range of legal tasks remains constrained by several factors that are particularly critical in the legal field:

2.3.1 Continuous Changes in Legal Frameworks

Legal systems undergo constant modifications. A comprehensive study on parliamentary productivity [14] reports that between 1990–2003, national parliaments passed between 698 laws (≈49 annually, United Kingdom) and 3,346 laws (≈239 annually, United States). This issue is particularly acute in Russia, where the State Duma adopted 653 laws in 2022 alone.[3] The described statistics only account for parliamentary legislation. When including regulatory acts issued by government agencies, the volume increases to levels where continuous LLM retraining becomes impractical. Current research on purely fine-tuned Law-LLMs, including LawGPT [11] and Lawyer LLaMa [12], fails to adequately address this legislative volatility due to the novelty of the domain.

Legal Changes as Dataset Shifts

Following the dataset shift framework [15], we model legislative changes as following:

Let $P_1(X)$ represent the probability distribution of legal provisions before changes, and $P_2(X)$ after changes. Then:

[3] https://tass.ru/politika/16661451

- $P_1(Y)$: Distribution of correct legal interpretations pre-change
- $P_2(Y)$: Distribution post-change

Where $P_1(X) \neq P_2(X) \Rightarrow P_1(Y) \neq P_2(Y)$ [4]. This yields three shift types:

1. **New legislation**:

$$P_1(Y \mid X) = P_2(Y \mid X) \quad \text{(covariate shift)}$$

2. **Amended legislation**:

$$P_1(Y \mid X) \neq P_2(Y \mid X) \quad \text{(concept shift)}$$

3. **Repealed legislation**:

$$P_1(Y) \neq P_2(Y) \quad \text{with} \quad P(X \mid Y) \quad \text{constant (prior probability shift)}$$

2.3.2 Temporal Shift

Recent research by Chenghao Zhu et al. [16] demonstrates that language models exhibit two critical limitations:

- Inability to process information absent from their training data
- Systematic preference for chronologically earlier information (termed "nostalgia bias")

[4] If this implication doesn't hold, model outputs remain valid without retraining

The study reveals that nostalgia bias intensifies with improved model performance metrics. Consequently, even with continuous retraining, LLMs maintain significant risk of utilizing outdated information - a particularly critical limitation for the legal domain given its constant evolution.

2.3.3 The Problem of Catastrophic Forgetting

The phenomenon of Catastrophic Forgetting, first defined by Anthony Robins in [17] as "the loss or disruption of previously learned information when new information is learned," remains relevant for LLMs. Recent studies demonstrate that both LLMs [18] and Multimodal LLMs [19] exhibit tendencies toward catastrophic forgetting.

The research reveals a linear relationship where:

- Improvement in loss function values for new tasks during fine-tuning
- Corresponding degradation in loss function values for original tasks

The exact coefficient of performance degradation on general tasks varies depending on specific model architecture, training data characteristics and task parameters. Nevertheless, the cited works investigate various methods to decrease this forgetting effect during model fine-tuning.

2.3.4 Hallucinations

Despite LLMs' strong performance in general domains, legal applications face accuracy challenges. Stanford research (2024) [20] shows fine-tuned legal LLMs still hallucinate in 17-33% of cases. OpenAI's theoretical analysis [21] confirms all language models retain measurable hallucination probabilities regardless of size or training. Consequently, even a theoretical model that is retrained daily cannot guarantee response accuracy in legal applications.

2.3.5 Cost of Fine-Tuning

While being significantly cheaper than training a language model from scratch in terms of computational time, resources, and financial costs, fine-tuning large language models still leads to substantial expenses.

In their 2024 study, Xia et al.[22] examine the fine-tuning costs of the Mixtral model [23] on the GSM8K dataset [24], which contains 8.5 thousand grade-school level math problems. Their estimated fine-tuning costs range from $17 to $32, depending on hardware configurations, excluding equipment acquisition costs and labor costs for specialists conducting the fine-tuning.

While a comprehensive cost assessment for maintaining model relevance falls beyond this work's scope, we note that given the aforementioned need for daily fine-tuning, these expenses may reach prohibitive levels for practical deployment.

2.4 RAG Approach

Retrieval-Augmented Generation (RAG), introduced in 2020 [25], combines LLMs with reliable information retrieval systems (see Section 2). [20] demonstrates RAG can reduce legal QA hallucinations to near-zero with proper context retrieval.

While RAG-based legal QA systems are rapidly developing (detailed in Section 2), existing research focuses on Western (primarily US and Australian) legal systems. Though Russian applications exist (e.g., [26]), none address legal domains. Our work applies and evaluates RAG in Russian legal contexts, providing foundational insights for future research. The idea of RAG is novel and simple at the same time: it combines the power of LLMs with the reliability of information retrieval systems. Its architecture consists of two main componentents: a retriever and a generator. The retriever is responsible for searching and retrieving relevant documents from a large corpus, while the generator uses these documents to produce coherent and contextually relevant responses. The retriever can be based on various techniques, such as dense retrieval, sparse retrieval, or a combination of both. The generator is typically a pre-trained language model, such as BART or T5, which is fine-tuned on the task of generating text based on the retrieved documents.

2.4.1 Diversity and Hierarchy of Legal Systems

Legal systems vary significantly across jurisdictions, often featuring multiple hierarchical levels (national, regional, municipal), where the set of applicable legal acts depends on specific territories. This challenge is particularly acute in Russia due to its federal state structure.

According to data from the Ministry of Justice of the Russian Federation [5], as of January 1, 2021, there were 20,184 distinct municipal entities in Russia. Consequently, achieving comprehensive coverage of only publicly available Russian legal framework would require maintaining at least 20,184 unique continuously-retrained models.

[5] https://minjust.gov.ru/uploaded/files/monitoring-msu-202115.docx

2.5 Juridical RAG

Chapter 3

Methodology

Chapter 4

Implementation

...

Chapter 5

Evaluation and Discussion

...

Chapter 6

Conclusion

. . .

Chapter 7

template data

The body of the text and abstract must be double-spaced except for footnotes or long quotations. Fonts such as Times Roman, Bookman, New Century Schoolbook, Garamond, Palatine, and Courier are acceptable and commonly found on most computers. The same type must be used throughout the body of the text. The font size must be 10 point or larger and footnotes[1] must be two sizes smaller than the text[2] but no smaller than eight points. Chapter, section, or other headings should be of a consistent font and size throughout the ETD, as should labels for illustrations, charts, and figures.

Referencing other chapters 2, 7, 4, 5 and 6

TABLE 7.1
Simulation Parameters

A	B
Parameter	**Value**
Number of vehicles	$\mid \mathcal{V} \mid$

[1]This is a footnote.
[2]This is another footnote.

A	B
Number of RSUs	$\mid \mathcal{U} \mid$
RSU coverage radius	150 m
V2V communication radius	30 m
Smart vehicle antenna height	1.5 m
RSU antenna height	25 m
Smart vehicle maximum speed	v_{max} m/s
Smart vehicle minimum speed	v_{min} m/s
Common smart vehicle cache capacities	$[50, 100, 150, 200, 250]$ mb
Common RSU cache capacities	$[5000, 1000, 1500, 2000, 2500]$ mb
Common backhaul rates	$[75, 100, 150]$ mb/s

...

7.1 NEW SECTION

7.1.1 Creating a Subsection

Creating a Subsubsection

Creating a Subsubsection

Creating a Subsubsection

This is a heading level below subsubsection And this is a quote:

> quote

Fig. 7.1. One kernel at x_s (*dotted kernel*) or two kernels at x_i and x_j (*left and right*) lead to the same summed estimate at x_s. This shows a figure consisting of different types of lines. Elements of the figure described in the caption should be set in italics, in parentheses, as shown in this sample caption.

This is a table:

TABLE 7.2
This Is a Table Example

A	B	C
a1	b1	c1
a2	b2	c2
a3	b3	c3
a4	b4	c4

The package “upgreek” allows us to use non-italicized lower-case greek letters. See for yourself: β, **β**, β, $\boldsymbol{\beta}$. Next is a numbered equation:

$$\|\boldsymbol{X}\|_{2,1} = \underbrace{\sum_{j=1}^{n} f_j(\boldsymbol{X})}_{\text{convex}} = \sum_{j=1}^{n} \|\boldsymbol{X}_{.,j}\|_2 \tag{7.1}$$

The reference to equation (7.1) is clickable.

7.2 Theorems, Corollaries, Lemmas, Proofs, Remarks, Definitions,and Examples

Theorem 1. *Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.*

Proof. I'm a (very short) proof. □

Lemma 1. *I'm a lemma.*

Corollary 1. *I include a reference to Thm. 1.*

Proposition 1. *I'm a proposition.*

Remark. I'm a remark.

Definition 1. I'm a definition. I'm a definition. I'm a definition. I'm a definition. I'm a definition. I'm a definition. I'm a definition. I'm a definition. I'm a definition. I'm a definition. I'm a definition.

Example. I'm an example.

7.3 Section with linebreaks in the name

Bibliography cited

[1] D. L. Mehl, “Automation in the legal world,” *Mechanisation of Thought Processes*, 1958.

[2] D. Merigoux, N. Chataing, and J. Protzenko, “Catala: A Programming Language for the Law,” *Proceedings of the ACM on Programming Languages*, vol. 5, no. ICFP, 77:1, 2021. DOI: 10.1145/3473582. Accessed: Mar. 30, 2025.

[3] W. Weaver, “Translation,” in *Proceedings of the Conference on Mechanical Translation*, Massachusetts Institute of Technology, 1952. Accessed: Mar. 30, 2025.

[4] P. F. Brown et al., “A Statistical Approach to Machine Translation,” *Computational Linguistics*, vol. 16, no. 2, pp. 79–85, 1990. Accessed: Mar. 30, 2025.

[5] E. L. Rissland, K. D. Ashley, and R. Loui, “AI and Law: A fruitful synergy,” *Artificial Intelligence*, vol. 150, no. 1-2, pp. 1–15, Nov. 2003, ISSN: 00043702. DOI: 10.1016/S0004-3702(03)00122-X. Accessed: Mar. 30, 2025.

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean, *Efficient Estimation of Word Representations in Vector Space*, Sep. 2013. DOI: 10.48550/arXiv.1301.3781. arXiv: 1301.3781 [cs]. Accessed: Mar. 30, 2025.

[7] A. Vaswani et al., "Attention is All you Need," in *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., 2017. Accessed: Mar. 30, 2025.

[8] T. Brown et al., "Language Models are Few-Shot Learners," in *Advances in Neural Information Processing Systems*, vol. 33, Curran Associates, Inc., 2020, pp. 1877–1901. Accessed: Mar. 30, 2025.

[9] T. Seabrooke et al., "A Survey of Lay People's Willingness to Generate Legal Advice using Large Language Models (LLMs)," in *Proceedings of the Second International Symposium on Trustworthy Autonomous Systems*, Austin TX USA: ACM, Sep. 2024, pp. 1–5, ISBN: 979-8-4007-0989-0. DOI: 10.1145/3686038.3686043. Accessed: Mar. 30, 2025.

[10] *Lawyers gear up for generative AI*, https://www.lexisnexis.co.uk/insights/lawyers-cross-into-the-new-era-of-generative-ai/. Accessed: Mar. 30, 2025.

[11] H.-T. Nguyen, *A Brief Report on LawGPT 1.0: A Virtual Legal Assistant Based on GPT-3*, Feb. 2023. DOI: 10.48550/arXiv.2302.05729. arXiv: 2302.05729 [cs]. Accessed: Mar. 30, 2025.

[12] Q. Huang et al., *Lawyer LLaMA Technical Report*, Oct. 2023. DOI: 10.48550/arXiv.2305.15062. arXiv: 2305.15062 [cs]. Accessed: Mar. 30, 2025.

[13] J. Cui et al., *Chatlaw: A Multi-Agent Collaborative Legal Assistant with Knowledge Graph Enhanced Mixture-of-Experts Large Language Model*,

May 2024. DOI: 10.48550/arXiv.2306.16092. arXiv: 2306.16092 [cs]. Accessed: Mar. 30, 2025.

[14] S. Brouard et al., "Legislative productivity in comparative perspective: An introduction to the comparative agendas project," in *ECPR Joint Sessions*, 2008.

[15] M. Kull and P. Flach, "Patterns of dataset shift,"

[16] C. ChenghaoZhu, N. Chen, Y. Gao, Y. Zhang, P. Tiwari, and B. Wang, "Is Your LLM Outdated? A Deep Look at Temporal Generalization," in *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, L. Chiruzzo, A. Ritter, and L. Wang, Eds., Albuquerque, New Mexico: Association for Computational Linguistics, Apr. 2025, pp. 7433–7457, ISBN: 979-8-89176-189-6. Accessed: May 11, 2025.

[17] A. ROBINS, "Catastrophic Forgetting, Rehearsal and Pseudorehearsal," *Connection Science*, vol. 7, no. 2, pp. 123–146, Jun. 1995, ISSN: 0954-0091. DOI: 10.1080/09540099550039318. Accessed: May 11, 2025.

[18] D. Kalajdzievski, *Scaling Laws for Forgetting When Fine-Tuning Large Language Models*, Jan. 2024. DOI: 10.48550/arXiv.2401.05605. arXiv: 2401.05605 [cs]. Accessed: May 11, 2025.

[19] Y. Zhai et al., "Investigating the Catastrophic Forgetting in Multimodal Large Language Model Fine-Tuning," in *Conference on Parsimony and Learning*, PMLR, Jan. 2024, pp. 202–227. Accessed: May 11, 2025.

[20] V. Magesh, F. Surani, M. Dahl, M. Suzgun, C. D. Manning, and D. E. Ho, *Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools*, May 2024. DOI: 10.48550/arXiv.2405.20362. arXiv: 2405.20362 [cs]. Accessed: Mar. 30, 2025.

[21] A. T. Kalai and S. S. Vempala, *Calibrated Language Models Must Hallucinate*, Mar. 2024. DOI: 10.48550/arXiv.2311.14648. arXiv: 2311.14648 [cs]. Accessed: Mar. 30, 2025.

[22] Y. Xia et al., *Understanding the Performance and Estimating the Cost of LLM Fine-Tuning*, Aug. 2024. DOI: 10.48550/arXiv.2408.04693. arXiv: 2408.04693 [cs]. Accessed: May 11, 2025.

[23] A. Q. Jiang et al., *Mixtral of Experts*, Jan. 2024. DOI: 10.48550/arXiv.2401.04088. arXiv: 2401.04088 [cs]. Accessed: May 11, 2025.

[24] K. Cobbe et al., *Training Verifiers to Solve Math Word Problems*, Nov. 2021. DOI: 10.48550/arXiv.2110.14168. arXiv: 2110.14168 [cs]. Accessed: May 11, 2025.

[25] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *Advances in Neural Information Processing Systems*, vol. 33, Curran Associates, Inc., 2020, pp. 9459–9474. Accessed: Mar. 21, 2025.

[26] A. G. Oleynik, I. O. Datyev, A. A. Zuenko, A. M. Fedorov, A. V. Shestakov, and I. G. Vishnyakov, "Using rag technology to design an intelligent information system for support exploratory search," *Transactions of the Kola Science Centre of RAS. Series: Engineering Sciences*, vol. 15, no. 3, pp. 5–27, 2024. DOI: 10.37614/2949.1215.2024.15.3.001. [Online]. Available:

https://rio.ksc.ru/data/documents/60_3_2024_15_IIMM/60_Trud_Teh_3_2024_15.pdf.

Appendix A

Extra Stuff

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Appendix B

Even More Extra Stuff

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.