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

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.

1.2 Large Language Models in Law

Advances in computational power enabled 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. Despite LLMs' strong performance in general domains, legal applications face accuracy challenges. Stanford research (2024) [14] shows fine-tuned legal LLMs still hallucinate in 17-33% of cases. OpenAI's theoretical analysis [15] confirms all language models retain measurable hallucination probabilities regardless of size or training.

Legal domains present unique challenges:

- **Continuous evolution**: Legislative acts constantly expand and change

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

- **Outdated information**: Documents may be deprecated, replaced, or amended
- **Jurisdictional variation**: Legal documents apply differently across regions, requiring localized fine-tuning

These factors heighten risks when using hallucination-prone AI for felony or misdemeanor cases, particularly when relying solely on "consultation machines" without verified document retrieval.

1.3 RAG Approach

Retrieval-Augmented Generation (RAG), introduced in 2020 [16], combines LLMs with reliable information retrieval systems (see Section 2). [14] 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., [17]), none address legal domains. Our work applies and evaluates RAG in Russian legal contexts, providing foundational insights for future research.

1.3.1 Creating a Subsection

Creating a Subsubsection

Creating a Subsubsection

Creating a Subsubsection

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

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.

Fig. 1.1. 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 1.1
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: $\upbeta$, $\boldsymbol{\upbeta}$, β, $\boldsymbol{\beta}$. Next is a numbered equation:

$$\|\boldsymbol{X}\|_{2,1} = \underbrace{\sum_{j=1}^{n} f_j(\boldsymbol{X})}_{\text{convex}} = \sum_{j=1}^{n} \|\boldsymbol{X}_{.,j}\|_2 \tag{1.1}$$

The reference to equation (1.1) is clickable.

1.4 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.

1.5 Section with linebreaks in the name

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.

This is the second paragraph. 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.

Chapter 2

Literature Review

2.1 Foundations of Retrieval-Augmented Generation

2.1.1 Core architecture of RAG-systems

As it was said in Chapter 1, in 2020 a Retrieval-Augmented approach was presented by Lewis et al. in [16]. 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.

Chapter 3

Methodology

Referencing other chapters 2, 3, 4, 5 and 6

TABLE 3.1
Simulation Parameters

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

A	B
Common backhaul rates	$[75, 100, 150]$ mb/s

. . .

Chapter 4

Implementation

. . .

Chapter 5

Evaluation and Discussion

. . .

Chapter 6

Conclusion

. . .

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] 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.

[15] 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.

[16] 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.

[17] 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.