

BALTIC JOURNAL OF LAW & POLITICS

A Journal of Vytautas Magnus University

VOLUME 18, ISSUE 2 (2025)

ISSN 2029-0454

Cit.: *Baltic Journal of Law & Politics* 18:2 (2025): 87–116

<https://reference-global.com/journal/BJLP>

DOI: 10.2478/bjlp-2025-0016

THE APPLICATION OF LARGE LANGUAGE MODELS IN ENFORCING PROHIBITIONS AGAINST HATE SPEECH IN LITHUANIAN

Milita Songailaitė

Doctoral student

Vytautas Magnus University, Faculty of Informatics, Lithuania

Contact information

Address: Universiteto str. 10, 53361 Akademija, Kauno r., Lithuania

E-mail address: milita.songailaite@vdu.lt

CRIS: <https://hdl.handle.net/20.500.12259/155037>

Aušrinė Pasvenskienė

Assoc. prof.

Vytautas Magnus University, Faculty of Law, Lithuania

Contact information

Address: Jonavos str. 66, 44191 Kaunas, Lithuania

E-mail address: ausrine.pasvenskiene@vdu.lt

CRIS: <https://hdl.handle.net/20.500.12259/153806>

Paulius Astromskis

Assoc. prof.

Vytautas Magnus University, Faculty of Law, Lithuania

Contact information

Address: Jonavos str. 66, 44191 Kaunas, Lithuania

E-mail address: paulius.astromskis@vdu.lt

CRIS: <https://hdl.handle.net/20.500.12259/153923>

Received: November 3, 2025; reviews 2; accepted: December 17, 2025.

SUMMARY

Large language models (LLMs) are increasingly used to detect and explain hate speech, yet their potential role within evidentiary processes remains underexplored. This article examines whether and how LLMs can support the enforcement of legal prohibitions against hate speech in Lithuania, focusing on the post-detection phase of legal analysis and case preparation. It first outlines the mechanics and evolutionary trends of LLMs, the specific risks associated with their deployment in a low-resource language such as Lithuanian, and the applicable substantive and procedural standards governing hate speech. Building on this framework, the article proposes and experimentally validates a conceptual “moot court” model in which distinct LLM agents assume the roles of plaintiff, defendant, and judge to generate structured legal arguments and reasoned decisions in Lithuanian hate speech cases. The findings indicate that, under carefully engineered constraints, LLMs can reliably distinguish criminal hate speech from lawful expression, reduce the cognitive burden on legal actors, and triangulate human and automated assessments, while persistent risks of hallucinations and opacity preclude their use as stand-alone evidence and instead support a complementary, assistive role in judicial practice.

KEYWORDS

Large language model; hate speech; evidence.

INTRODUCTION

The exponential growth of platforms for user-generated content has made hate speech one of the most significant and complex societal issues, with a multi-layered nature and the practical need to detect it automatically.¹ At the same time, the exponential growth of training dataset size and computational power has fuelled a myriad of large language models (LLMs), competing for dominance in generative AI, marked by intelligence, price, and speed. This evolution has opened new streams of research and development in the field of hate speech. Some authors analyse the efficiency of LLMs in hate speech detection using different prompting strategies², compare these models with human annotators³, or evaluate the quality of generated explanations in terms of linguistic fluency, informativeness, persuasiveness, and logical soundness.⁴

Within the Lithuanian hate speech context, researchers tend to focus on the legal evaluation and procedural issues of hate speech offences⁵, an explanation of this phe-

¹ Jūratė Ruzaitė, *Neapykantos kalba: Monografija (Hate speech: Monograph)* (Kaunas: Vytauto Didžiojo universitetas, 2024), 15 // <https://doi.org/10.7220/9786094676109>

² José Antonio García-Díaz, Ronghao Pan, Rafael Valencia-García, “Leveraging Zero and Few-Shot Learning for Enhanced Model Generality in Hate Speech Detection in Spanish and English,” *Mathematics* 11, No. 24, 5004 (December 2023): 16 // <https://doi.org/10.3390/math11245004>; Priya Gupta, Sachin Gupta, “Hate Speech Detection using OpenAI and GPT-3,” *International Journal of Emerging Technology and Advanced Engineering* Vol. 12, Issue 5 (May 2022): 136 // doi: 10.46338/ijetae0522_15; Ronghao Pan, José Antonio García-Díaz, Rafael Valencia-García, “Comparing fine-tuning, zero and few-shot strategies with large language models in hate speech detection in English,” *Computer Modeling in Engineering & Sciences* 140(3) (July 2024): 2850 // <https://doi.org/10.32604/cmes.2024.049631>

³ Fan Huang, Haewoon Kwak, Jisun An, “Is ChatGPT better than human annotators? Potential and limitations of ChatGPT in explaining implicit hate speech,” *Companion Proceedings of the ACM Web Conference* (March 2023) // <https://doi.org/10.48550/arXiv.2302.07736>

⁴ Han Wang et al., “Evaluating GPT-3 generated explanations for hateful content moderation,” *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence AI for Good (IJCAI-23)* (2023) // <https://doi.org/10.24963/ijcai.2023/694>

⁵ Dovilė Murauskienė, Raimundas Jurka, Jolanta Zajančauskienė, “The instigation of hatred: Questions of legal evaluation and procedural issues,” *Entrepreneurship and Sustainability Issues* 8(2) (2020) // [https://doi.org/10.9770/jesi.2020.8.2\(54\)](https://doi.org/10.9770/jesi.2020.8.2(54)); Mindaugas Bilius, Žaneta Navickienė, Vilius Velička, “Hate crimes: Evaluation of Lithuanian courts’ decisions in the light of the practice of the European Court of Human Rights,” *Baltic Journal of Law & Politics* 14(1) (October 2021) // <https://doi.org/10.2478/bjlp-2021-0002>;

nomenon from the linguistic perspective⁶, or the concept and power of counter-speech⁷, among others. Moreover, numerous projects and initiatives aim to develop a functioning prototype to detect hate speech⁸ or a multi-layered model of criteria for determining it⁹, or strengthen the resilience to hate speech through awareness-raising and capacity-building campaigns.¹⁰ Few emergent research initiatives aim to compare the performance of different LLMs in Lithuanian language hate speech detection tasks.¹¹ However, despite prior research and development projects, hate speech incidents in Lithuania remain a problem, especially among vulnerable communities.¹²

To counter hate speech and hate crimes, the European Parliament's commissioned study revealed, amongst other things, that besides soft measures that build social resilience against hate speech, enforcement of legal prohibitions should be more consequent on the level of investigation, prosecution, and the judiciary.¹³ Within this context, LLMs offer the potential to significantly reduce the cognitive burden and costs associated with enforcing legal prohibitions by automating hate speech detection, legal assessment, and the drafting of claims, defenses, or even judicial decisions. Accordingly, this research aims to propose a conceptual model for the application of LLMs in the post-detection stages of countering hate speech – specifically, performing legal analysis and initiating enforcement procedures.

To achieve this aim, the following research questions should be answered. First, what are the fundamental mechanics and evolutionary trends of LLMs, and what is the current state of research regarding their application to Lithuanian hate speech detection? Second, what are the inherent risks of using LLMs, including the challenges affected by the low-resource nature of the Lithuanian language? Third, what are the substantive legal standards and judicial practices governing hate speech in Lithuania, and what evidentiary challenges restrict the integration of LLM-generated outputs into the criminal process? Fourth, what conceptual model could be designed to apply LLMs in the post-detection phase of hate speech enforcement to mitigate cognitive burden and ensure legal validity? Finally, what does the experimental validation reveal about the proposed model in distinguishing criminal hate speech from lawful expression? These research questions determine the structure of this paper, with each section dedicated to exploring a specific question in the order presented.

Viltė Aleknaitė, "The qualification problem of incitement against any nation, race, ethnic, religious or other group of people in Lithuanian criminal law," *Teisės apžvalga / Law Review* 28 (2023).

⁶ Jūratė Ruzaitė, "Neapykantos kalba: medijų raštingumas iš lingvistinės perspektyvos": in: Aukšė Balčytienė (ed.), *Artimas tolimes medijų pasaulis: virsmas, ribos ir būdai tai suprasti* (Kaunas: Vytauto Didžiojo universitetas, 2024) // <https://doi.org/10.7220/9786094676253>

⁷ Jūratė Ruzaitė, *supra* note 1.

⁸ European Commission, *Project Be Hate-Free: Building Hate-Free Communities in Lithuania (Be Hate-Free)* // <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/org-details/925243619/project/963662/program/31076817/details>

⁹ Vytautas Magnus University, *Features of hate speech and criteria for its identification: A linguistic and legal analysis* // <https://www.vdu.lt/cris/entities/project/5eae72bf-e195-487a-b908-c9a835d15535>

¹⁰ Vidaus reikalų ministerija, *Ataskaita apie neapykantos nusikaltimų ir neapykantos kalbos situaciją Lietuvoje 2020–2021 m.*, (2022) // https://vrm.lrv.lt/uploads/vrm/documents/files/LT_versija/Viesasis_saugumas/2020_2021%20ataskaita%20d%C4%97l%20neapykantos%20kalbos%20ir%20neapykantos%20nusikaltim%C5%B3.pdf

¹¹ Eglė Kankevičiūtė, Milita Songailaitė, Justina Mandravickaitė, "Neapykantos kalbos atpažinimas lietuviškuose komentaruose panaudojant dirbtinį intelektą," *Vilniaus universiteto atvirosios serijos* (May 2023) // <https://doi.org/10.15388/LMITT.2023.3>; Justina Mandravickaitė et al., "Exploring hate speech detection models for Lithuanian language," *Proceedings of the 9th Workshop on Online Abuse and Harms (WOAH)* (2025).

¹² Katažyna Mikša, Monika Guliakaitė, *Neapykantos nusikaltimai ir neapykantos kalba: situacijos Lietuvoje apžvalga*, (2020) // <https://lt.efhr.eu/wp-content/uploads/2020/12/NEAPYKANTOS-NUSIKALTIMAI-IR-NEAPYKANTOS-KALBA-SITUACIJOS-LIETUVOJE-AP%C5%BDVALGA.pdf>

¹³ European Parliament, *Hate speech and hate crime in the EU and the evaluation of online content regulation approaches*, (2020) // <https://data.europa.eu/doi/10.2861/28047>

We expect that the outcomes of this research will contribute to knowledge, wisdom, and understanding of the ways LLMs can be used to counter hate speech. Such enhancement of mechanisms for enforcing the law may also strengthen societal resilience to the phenomenon. Notably, while the specific focus herein is limited to hate speech, inspired by the works of *Alma Mater* colleagues on automated detection of it, the architecture of the proposed model is universal. It may be used in various other fields of law, from business disputes to the protection of personal rights. Consequently, we expect that the outcomes of this research will be used to develop new practices not only in the field of hate speech but also across the broader legal landscape.

1. LLM MECHANICS, TRENDS, AND RELATED WORK ON LITHUANIAN HATE SPEECH DETECTION

LLMs are probabilistic sequence-to-sequence models that assign a probability to each possible next token (a sub-word unit) given the tokens they have seen so far and then generate text by repeatedly sampling from those probabilities. They do not “think” in the human sense; instead, they compute likelihoods conditioned on context using parameters learned from vast text corpora. Recent LLMs are built using the Transformer architecture¹⁴, whose self-attention mechanism allows the model to weigh linkages between all tokens in a sequence at once. This is crucial for capturing long-range dependencies in legal prose and other complex texts.

As the field has matured, the focus has shifted from the first wave of pretrained “just scale it” models to alignment and instruction following. After pretraining as next-token predictors, models are modified to reflect human preferences, first through reinforcement learning from human feedback (RLHF)¹⁵, preference optimization (e.g., Direct Preference Optimization)¹⁶ and synthetic instruction generation. Still, these actions do not result in understanding. Instead, they reshape the learned conditional distribution, so the model’s most likely continuations more closely fit what people want to read, reducing toxic or unhelpful outputs while preserving the statistical core.

To further mitigate the limitations of pure statistical prediction, such as factual drift, and add authenticity, systems increasingly pair generation with retrieval (RAG)¹⁷ and, in many workflows, explicit tool use or agentic planning, which invokes search, calculators, or databases mid-generation.¹⁸ In parallel, production systems pursue longer contexts and faster inference using linear and IO-aware attention kernels¹⁹, extended rotary embeddings²⁰, and speculative decoding that reduces latency without altering

¹⁴ Ashish Vaswani et al., “Attention Is All You Need,” *Proceedings of the 31st Conference on Neural Information Processing Systems* (2017): 2–3 // https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

¹⁵ Long Ouyang et al., “Training language models to follow instructions with human feedback,” *arXiv* (2022): 2 // <https://arxiv.org/abs/2203.02155>

¹⁶ Rafael Rafailov et al., “Direct Preference Optimization: Your Language Model is Secretly a Reward Model,” *arXiv* (2023) // <https://arxiv.org/abs/2305.18290v3>

¹⁷ Partick Lewis et al., “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” *Proceedings of the 34th Conference on Neural Information Processing Systems* (2020) // <https://proceedings.neurips.cc/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf>

¹⁸ Timo Schick et al., “Toolformer: Language Models Can Teach Themselves to Use Tools,” *arXiv* (2022) // <https://arxiv.org/abs/2302.04761>

¹⁹ Tri Dao et al., “FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness,” *arXiv* (2022) // <https://arxiv.org/abs/2205.14135>

²⁰ Shouyuan Chen, Sherman Wong, Liangjian Chen, Yuandong Tian, “Extending Context Window of Large Language Models via Positional Interpolation,” *arXiv* (2023) // <https://arxiv.org/abs/2306.15595>

outputs.²¹ At the frontier, mixture-of-experts simplicity enables scalability without proportional computation, whereas multimodal versions accept pictures alongside text but decode one token at a time using probability. None of these change the basic mechanics, an LLM remains an autoregressive estimator that selects the next token by its estimated likelihood given the context, even when it is consulting a database, sketching its reasoning, or reading a long brief.

The current model landscape reflects these divergent priorities between proprietary and open-weight systems. Today's most popular closed-source LLMs include models such as GPT, Claude, Gemini, Grok, and many other versions. All of these models share the Transformer design backbone, but differ in modalities, context lengths, and alignment methodologies. OpenAI's GPT model family has chosen a native multimodal and low-latency approach (e.g., GPT-4o and its Realtime API), enabling real-time search, multimodal voice and vision integration, and a variety of interactive use cases. One of Anthropic's most notable characteristics is its safety alignment via Constitutional AI²², which makes it one of the most ethical general-purpose LLMs. Claude models provide bigger context sizes than other models (e.g., Claude 3.5 Sonnet with ~200K tokens), making them ideal for extended legal documents and cautious help. Google's Gemini models prioritize specific ultra-long context and integration with enterprise applications (e.g., Google Docs), which means that in certain cases, you can have more access to a variety of programs, significantly altering document-based workflows. Finally, xAI's Grok program positions itself as a "reasoning-first" model, which dramatically increases the model's ability to do step-by-step problem solving tasks and aids in more accurate information recall (fewer hallucinations). However, Grok models have a more limited enterprise toolchain, making real-time data search more difficult.

The open-weight model community primarily relies on Meta's Llama line (with Llama 4 serving as the current lead) and Mistral's series. Both target users who prefer or require self-hosted deployments, putting cost control, data residency, and privacy entirely on the operator's shoulders. These models may stay behind closed-source systems in some specialized dimensions, particularly fully integrated multimodality, extensive business capabilities, and consistent multilingual coverage, but the gap is closing as releases mature and third-party ecosystems grow. DeepSeek's R1 "reasoning" versions²³ and Alibaba's Qwen family push the open-weight boundaries by focusing on step-by-step problem solving, robust long-context options, and competitive multilingual or vision-language capabilities. In practice, closed models continue to lead in end-to-end user experience and various platforms support, whereas open-weight families are increasingly winning in transparency, customization, and total cost of ownership, providing practitioners with a meaningful choice between maximum control and maximum convenience.

While these generative giants advance the frontiers of reasoning and text production, the specific task of detecting hate speech in morphologically complex languages like Lithuanian has historically relied on specialized, smaller architectures. The first systematic application of deep learning to Lithuanian hate-speech classification was

²¹ Yaniv Leviathan, Matan Kalman, Yossi Matias, "Fast Inference from Transformers via Speculative Decoding," *arXiv* (2022) // <https://arxiv.org/abs/2211.17192>

²² Yuntao Bai, "Constitutional AI: Harmlessness from AI Feedback," *arXiv* (2022) // <https://arxiv.org/abs/2212.08073>

²³ DeepSeek-AI, Daya Guo et al., "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning," *arXiv* (2025) // <https://arxiv.org/abs/2501.12948>

presented by Kankevičiūtė, Songailaitė and Mandravickaitė²⁴. Their study compared three classical transformer-based NLP models: Multilingual BERT, LitLat BERT, and Electra (fine-tuned on a 25k-comment corpus annotated into neutral, offensive, and hate categories). This work demonstrated that specialized medium-scale transformers (LitLat BERT) outperform both multilingual and more efficiency-focused architectures such as Electra, achieving an F1 score of 0.72. The advantage of transformer encoders in this setting stems from their bidirectional attention mechanism, which captures morphological variation and contextual cues essential for Lithuanian, a morphologically rich language. Electra's efficiency-oriented discriminator training proved less effective for nuanced category boundaries, suggesting that pretraining objective and linguistic coverage matter more than computational efficiency in this task.

Mandravickaitė et al.²⁵ expanded this line of research substantially by comparing three families of models on a larger annotated corpus (27k comments): (1) traditional ML (SVM, Random Forest), (2) classical deep learning (CNN, LSTM/BiLSTM), and (3) modern transformers and LLMs (Multilingual BERT, LitLat BERT, Electra, RWKV, LT-Llama-2, Gemma-2, ChatGPT). Their results reinforce the earlier finding that encoder-style transformers, especially LitLat BERT and Multilingual BERT, remain the most reliable for Lithuanian hate-speech detection, achieving weighted F1-scores above 0.76 on imbalanced data and above 0.84 after resampling and augmentation. In contrast, general LLMs used in zero-shot or few-shot mode performed inconsistently, and even large proprietary models underperformed smaller fine-tuned encoder models unless subjected to explicit domain adaptation. Fine-tuned LLMs (e.g., LT-Llama-2-7B) improved notably, but still required substantial prompt engineering and compute, while offering no clear advantage over classical fine-tuned BERT-family models in accuracy. Overall, this demonstrates that large-scale models without additional training cannot outperform specialized models; however, in scenarios where training data or resources are limited, they remain a viable and often quite effective alternative.

This dichotomy presents a specific challenge. While specialized smaller models are sufficient for detection, the enforcement phase – which involves legal analysis, reasoning, and document drafting – requires the generative capabilities of large-scale LLMs. However, utilizing these large models for legal tasks introduces risks that do not exist in simple classification, particularly the generation of plausible but incorrect information, also known as hallucinations. The issues related to the generative capabilities of LLMs are also exacerbated by the linguistic complexities of the Lithuanian language, which are discussed further.

2. HALLUCINATIONS, PROMPT ENGINEERING, AND THE *LINGUA EXOTICA LITUANA*

While LLMs represent a significant advancement, potentially ushering in a new era of legal productivity, they also exhibit limitations such as inaccurate information, misinterpreting facts, algorithmic discrimination and bias, and failing to adhere to ethical guidelines, amongst others.²⁶ The most critical problem for LLMs is hallucinations—mistakes

²⁴ Eglė Kankevičiūtė, Milita Songailaitė, Justina Mandravickaitė, *supra* note 11.

²⁵ Justina Mandravickaitė et al., *supra* note 11.

²⁶ Daniel Martin Katz, Michael James Bommarito, Shang Gao, Pablo Arredondo, "GPT-4 passes the bar exam," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 382(2243) (2024): 13 // <https://doi.org/10.1098/rsta.2023.0254>; Wang Zheng, "Research on generative artificial intelligence legal profession substitution," *Journal of Modern Law Research* 4(3) (2023): 38

in the generated text that are semantically or syntactically plausible but are, in fact, incorrect or nonsensical.²⁷ Hallucinations erode the trust of end users, prevent adoption in diverse fields, and raise safety and privacy concerns.²⁸ As Cossion observes, hallucinations come in many shapes, all of them caused by the data, model, or prompt factors²⁹, but the underlying reason is that LLMs sometimes guess when uncertain, producing plausible yet incorrect statements instead of admitting uncertainty.³⁰

Many studies have been presented on measuring and mitigating hallucinated texts. However, encouraging truthfulness through self-checking, robust instruction tuning, knowledge injection, self-reflection mechanisms, contrastive learning, or other supervision or reinforcement methods have been only partially successful and not without limitations.³¹ Automatic detection of hallucinations and metrics used to measure the quality of writing are still not well established and have not been widely validated, leaving human validation of hallucinations as the gold standard.³² That is, despite increased scientific attention, detecting and mitigating hallucinations remains a novel, challenging, and unresolved problem, requiring careful monitoring, ethical guidelines, and continuous refinement of their algorithms, training data, and prompts.³³

Indeed, LLMs' effectiveness significantly depends on the quality of prompts – an input text string that contains the instruction to these models.³⁴ LLMs have demonstrated extreme sensitivity even to subtle changes in prompt formatting, and therefore, choices in prompt design are critical in effectively using any modern pre-trained LLM.³⁵ Accordingly, one of the essential techniques to achieve excellent accuracy performance of LLMs is the optimisation of these inputs to the model, which is also described as

-
- // doi: 10.37420/j.mlr.2023.017; Sadie O'Connor, "Generative AI," *Georgetown Law Technology Review* 8(2) (2024): 401–403 // <https://georgetownlawtechreview.org/generative-ai/GLTR-05-2024/>; Francesco Contini, "Unboxing generative AI for the legal professions: Functions, impacts, and governance," *International Journal for Court Administration* 15(2) (2024) // <https://iacajournal.org/articles/10.36745/ijca.604>
- ²⁷ Craig S. Smith, "Hallucinations could blunt ChatGPT's success – IEEE spectrum," (2023) // <https://spectrum.ieee.org/ai-hallucination>
- ²⁸ George Chrysostomou, Zhixue Zhao, Miles Williams, Nikolaos Aletras, "Investigating hallucinations in pruned large language models for abstractive summarization," *Transactions of the Association for Computational Linguistics* 12 (2024) // https://doi.org/10.1162/tacl_a_00695; Ziwei Ji et al., "Survey of hallucination in natural language generation," *ACM Computing Surveys* 55(12) (2023) // <https://doi.org/10.1145/3571730>; Sebastian Farquhar, Jannik Kossen, Lorenz Kuhn and Yarin Gal, "Detecting hallucinations in large language models using semantic entropy," *Nature* 630 (2024) // <https://doi.org/10.1038/s41586-024-07421-0>
- ²⁹ Manuel Cossion, "A comprehensive taxonomy of hallucinations in Large Language Models," *arXiv* (2025) // <https://arxiv.org/abs/2508.01781v1>
- ³⁰ Adam Tauman Kalai et al., "Why Language Models Hallucinate," *arXiv* (2025) // <https://arxiv.org/abs/2509.04664>
- ³¹ Sebastian Farquhar, *supra* note 28; Muhammad Mudassar Yamin, Ehtesham Hashmi, Mohib Ullah, Basel Katt, "Applications of LLMs for generating cyber security exercise scenarios," *IEEE Access* (2024) // <https://doi.org/10.21203/rs.3.rs-3970015/v1>; Matthew Dahl, Varun Magesh, Mirac Suzgun, Daniel E Ho, "Large legal fictions: Profiling legal hallucinations in large language models," *Journal of Legal Analysis* 16(1) (2024) // <https://doi.org/10.1093/jla/laae003>
- ³² Sherry Yan et al. "Prompt engineering on leveraging large language models in generating response to InBasket messages," *Journal of the American Medical Informatics Association* 31(10) (2024): 2263–2270 // <https://doi.org/10.1093/jamia/ocae172>; Ziwei Ji, *supra* note 28.
- ³³ Mudassar Yamin, *supra* note 31; Rajvardhan Patil, Thomas F. Heston, Vijay Bhuse, "Prompt engineering in healthcare," *Electronics* 13(15), 2961 (2024) // <https://doi.org/10.3390/electronics13152961>; Alessia McGowan et al., "ChatGPT and Bard exhibit spontaneous citation fabrication during psychiatry literature search," *Psychiatry Research* 326, 115334 (2023) // <https://doi.org/10.1016/j.psychres.2023.115334>
- ³⁴ Aras Bozkurt, Ramesh C. Sharma, "Generative AI and prompt engineering: The art of whispering to let the genie out of the algorithmic world," *Asian Journal of Distance Education* 18 (2) (2023); Aras Bozkurt, "Tell Me Your Prompts and I Will Make Them True: The Alchemy of Prompt Engineering and Generative AI," *Open Praxis* 16(2) (2024) // doi: <https://doi.org/10.55982/openpraxis.16.2.661>
- ³⁵ Melanie Sclar, Yejin Choi, Yulia Tsvetkov and Alane Suhr, "Quantifying Language Models' Sensitivity to Spurious Features in Prompt Design or: How I learned to start worrying about prompt formatting," *arXiv* (2024) // <https://doi.org/10.48550/arXiv.2310.11324>; Albert Webson, Ellie Pavlick, "Do prompt-based models really understand the meaning of their prompts?" *arXiv* (2022) // <https://arxiv.org/pdf/2109.01247>

"prompt engineering"³⁶ or "promptology".³⁷ It is an emerging field of scientific inquiry into the principles, strategies, and techniques for creating, structuring, fine-tuning, and then integrating prompts that maximise interactions' utility, efficiency, and safety with generative AIs.³⁸ Prompt engineering, at its core, involves understanding what methods are most effective in prompting LLMs and strategic crafting of inputs to elicit desired outputs without modifying the model itself.³⁹

By optimising text-based instructions, questions, and/or contextual information, prompt engineers can guide the system to generate more accurate, relevant, and valuable outputs from LLMs.⁴⁰ To achieve this, prompt engineers must understand the AI model's capabilities and limitations and consider the model's training data, capacity to handle different tasks, and the desired output format. Clear, specific, and context-appropriate prompts are essential for eliciting high-quality responses from generative AI models. Ambiguous or overly broad prompts, lacking sufficient context, can lead to irrelevant, incomplete or nonsensical outputs.⁴¹ This makes the study of promptology not merely an academic interest but a vital component of practical AI usage.⁴² Yet, prompt engineering is an experimental discipline with only a handful of successful prompting methodologies⁴³ that have emerged recently and are continuously being improved.⁴⁴

Besides the quality of prompts, the amount and quality of training data generally improve an LLM's language abilities.⁴⁵ Lithuanian is *Lingua Exotica* compared with the high-resource languages. It is one of the low-resource languages, spoken by ~3 million native speakers and an estimate of 56 billion words indexed by Google⁴⁶, compared with ~400 million native English speakers⁴⁷ and ~30% of all web content being in English.⁴⁸ Developing LLMs for low-resource languages involves significant technical and economic challenges, linked to the limited availability of high-quality data, lower

³⁶ Brydon T Wang, "Prompts and Large Language Models: A New Tool for Drafting, Reviewing and Interpreting Contracts?" *Law, Technology and Humans* 6(2) (2024) // <https://doi.org/10.5204/lthj.3483>; Minghao Shao, Abdul Basit, Ramesh Karri, Muhammad Shafique, "Survey of Different Large Language Model Architectures: Trends, Benchmarks, and Challenges," *IEEE Access* 12 (2024) // <http://dx.doi.org/10.1109/ACCESS.2024.3482107>

³⁷ Phillip Olla, et al., "Promptology: Enhancing Human-AI Interaction in Large Language Models," *Information* 15(10) 634 (2024) // <https://doi.org/10.3390/info15100634>

³⁸ *Ibid.*; Aras Bozkurt, Ramesh C. Sharma, *supra* note 34; Jules White et al., "A prompt pattern catalog to enhance prompt engineering with ChatGPT," *arXiv* (2023) // <https://doi.org/10.48550/arXiv.2302.11382>; Juan David Velásquez-Henao, Carlos Jaime Franco-Cardona and Lorena Cadavid-Higuaita, "Prompt Engineering: a methodology for optimizing interactions with AI-Language Models in the field of engineering," *DYNA* 90 (230) (2023) // <https://doi.org/10.15446/dyna.v90n230.111700>

³⁹ Yoshija Walter, "Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education," *International Journal of Education Technology in Higher Education* 21, 15 (2024) // <https://doi.org/10.1186/s41239-024-00448-3>; Fouad Trad, Ali Chehab, "Prompt Engineering or Fine-Tuning? A Case Study on Phishing Detection with Large Language Models," *Machine Learning and Knowledge Extraction* 6(1) (2024) // <https://doi.org/10.3390/make6010018>

⁴⁰ Rajvardhan Patil, *supra* note 33.

⁴¹ *Ibid.*; Aras Bozkurt, Ramesh C. Sharma Sharma, *supra* note 34.

⁴² Rajvardhan Patil, *supra* note 33.

⁴³ See for e. g. Minghao Shao, *supra* note 36, Yoshija Walter, *supra* note 39; Liu Pengfei et al., "Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing," *arXiv* (2021) // <https://arxiv.org/pdf/2107.13586> for an eloquent summary of the most relevant prompting methods

⁴⁴ Yoshija Walter, *supra* note 39.

⁴⁵ Jared Kaplan et al., "Scaling laws for neural language models," *arXiv* (2020): 4 // <https://arxiv.org/abs/2001.08361>; Jordan Hoffmann et al., "Training Compute-Optimal Large Language Models," *arXiv* (2022) // <https://doi.org/10.48550/arXiv.2203.15556>

⁴⁶ Virginijus Dadurkevičius, Andrius Utkas, "Estimating the Amount of Lithuanian Text Indexed by Global Search Engines," *Baltic Journal of Modern Computing* 10 (2022) // doi:10.22364/bjmc.2022.10.3.06

⁴⁷ Babel, *How Many People Speak English, And Where Is It Spoken?* // <https://www.babel.com/en/magazine/how-many-people-speak-english-and-where-is-it-spoken?>

⁴⁸ Daniel Pimienta, "Is it true that more than Half of Web Contents are in English? Not if Multilingualism is Paid Due Attention!" *Forum for Linguistic Studies* 6(5) (2024) // doi:10.30564/fls.v6i5.7144

demand, fewer applications for such models, and the lack of financial incentives or rewards, making investment in these languages less attractive compared to more widely spoken ones.⁴⁹ As a result, most multilingual LLMs are trained on corpora where English is the overwhelmingly predominant language, leaving all other languages severely underrepresented.⁵⁰ This leads to the consistently observable performance gaps on tasks in low-resource languages.⁵¹ Moreover, Lithuanian is the most archaic Indo-European language still spoken⁵², coloured with many dialects⁵³, which makes it not only a low-resource, but also a high-complexity language.

To handle this underrepresentation, attempts are being made to balance the lack of quantity with better quality of specialized models and text corpora. The release of the first open Lithuanian-specific LLM, built on the Llama 2 model and additionally trained with the Lithuanian partition of the CulturaX dataset, is said to represent a major milestone in Lithuanian LLM development.⁵⁴ Also, the development of the Lithuanian hate speech corpus⁵⁵ and methods for its detection.⁵⁶ There are promising research initiatives intended to support automated hate-speech detection, despite the shortcomings of a *Lingua Exotica Lituāna*. Be that as it may, the Lithuanian language presents an additional layer of complexity, thus forming an interesting field of scientific inquiry.

3. LITHUANIAN LAW ON HATE SPEECH AND THE EVIDENTIARY CHALLENGES OF LLM INTEGRATION

The Criminal Code of the Republic of Lithuania establishes criminal liability for acts committed for the purpose of expressing hatred, prejudice, or bias on racial, nationalist, xenophobic, homophobic, religious, or other discriminatory grounds. Lithuanian law does not expressly employ the term “hate speech” to define this phenomenon. Nevertheless, on 26 July 2023, the new version of the Methodological Recommendations on the Pre-Trial Investigations concerning Hate Crimes and Hate Speech, approved by the Prosecutor General, entered into force, which set out the concepts of hate crimes and hate speech, the criteria for their distinction and specific features, as well as other terms relevant for the identification and investigation of such offences, the understanding of vulnerable groups, and the principal rules for conducting and organizing pre-trial investigations, international cooperation, and the particularities of the legal qualification of criminal acts. According to these recommendations, drawing upon both doctrinal and practical interpretation, all criminal acts committed against a person, society, or property—where the motive arises from the offender’s preconceived, prejudice-based, or

⁴⁹ Dominykas Pleševičius, *Large Language Models for Lithuanian Language* (Master’s thesis, VU, 2025).

⁵⁰ Yuemei Xu et al., “A survey multilingual large language models: corpora, alignment, and bias,” *Front. Comput. Sci.* 19, 1911362 (2025) // <https://doi.org/10.1007/s11704-024-40579-4>

⁵¹ Dominykas Pleševičius, *supra* note 49; Yevhen Kostyuk et al., “Towards Multilingual LLM Evaluation for Baltic and Nordic languages: A study on Lithuanian History,” *arXiv* (2025) // <https://doi.org/10.48550/arXiv.2501.09154>

⁵² The Encyclopedia Britannica, *Lithuanian language* // <https://www.britannica.com/topic/Lithuanian-language>

⁵³ Danguolė Mikulėnienė, Agnė Čepaitienė, “Lithuanian Dialect Classifications,” *Dialectologia*, Special issue 11 (2023) // doi: 10.1344/DIALECTOLOGIA2023.2023.8

⁵⁴ Dominykas Pleševičius, *supra* note 49; Yevhen Kostyuk et al., *supra* note 51.

⁵⁵ Rita Butkienė et al., “Lithuanian Hate Speech Corpus v.1,” *CLARIN-LT digital library in the Republic of Lithuania* (2025) // <http://hdl.handle.net/20.500.11821/69>; Eglė Kankevičiūtė, Milita Songailaitė, Justina Mandravickaitė, *supra* note 11.

⁵⁶ See for e. g. Eglė Kankevičiūtė, Milita Songailaitė, Justina Mandravickaitė, *supra* note 11; Justina Mandravickaitė et al., *supra* note 11; Jūratė Ruzaitė, “In search of hate speech in Lithuanian public discourse: A corpus-assisted analysis of online comments,” *Lodz Papers in Pragmatics* 14(1) // <https://doi.org/10.1515/lpp-2018-0005>

stereotypical negative attitudes toward an individual or a group of persons on grounds such as race, ethnicity, nationality, citizenship, religion, sex, sexual orientation, age, social status, disability, beliefs, or views—are classified as hate crimes and misdemeanours (hereinafter – hate crimes) and acts of hate speech.⁵⁷ Hate crimes are defined as criminal acts motivated by hatred, bias, or prejudice against a group of persons distinguished by age, sex, sexual orientation, disability, race, nationality, language, origin, social status, religion, beliefs, or opinions.⁵⁸ Hate speech, or incitement to hatred, is defined as the public dissemination—verbally, in writing, or by any other means—of information, including ideas, opinions, or knowingly false statements of fact, that mocks, denigrates, incites hatred, or encourages discrimination, violence, or physical assault against a group of persons or an individual belonging to such a group on the basis of age, sex, sexual orientation, disability, race, nationality, language, origin, social status, religion, beliefs, or opinions.⁵⁹ Acts constituting hate speech are criminalized under Article 170(2) and (3) and Article 170² of the Criminal Code.

Article 170 of the Lithuanian Criminal Code⁶⁰ makes a criminal offence to publicly ridicule, express contempt for, incite hatred toward, or incite discrimination against a group of people or an individual based on certain protected characteristics, including age, sex, sexual orientation, disability, race, nationality, language, descent, social status, religion, convictions, or views. The Supreme Court of Lithuania, in its practice, has provided the following interpretation of the terms contained in Article 170(2) of the Criminal Code: “Mockery refers to the degradation of honour and dignity, or the portrayal of a person or group of persons as an object of ridicule; denigration refers to a highly negative and disrespectful statement about a person or a group of persons, rejoicing at their misfortunes, or at crimes committed against them, and similar actions; incitement of hatred refers to the attribution to a person or group of persons of qualities or actions which are untrue and negatively characterize them in the eyes of society, thereby causing hostility or provoking intolerance towards that person or group; incitement to discriminate refers to a direct call or an indirect encouragement to restrict the rights and freedoms of a person or group of persons in comparison with other persons or groups.”⁶¹ However, it is noted that this interpretation of the Supreme Court of Lithuania is of a general nature, and it is rather difficult to apply it in practical work.⁶²

Given these interpretive challenges, the practical application⁶³ of this article in practice typically involves content analysis of the European Union policy recommendations, such as European Commission against Racism and Intolerance (ECRI) 2015-12-08 General Policy Recommendation No 15 on “Combating Hate Speech”⁶⁴ or Council of

⁵⁷ Office of the Prosecutor General of the Republic of Lithuania, *Recommendations on the Methodological Recommendations on the Pre-Trial Investigations concerning Hate Crimes and Hate Speech* (2023) // https://www.prokuraturos.lt/data/public/uploads/2023/09/20230726_neapykantos_nusikaltimai_rekomendacijos.pdf

⁵⁸ *Ibid.*

⁵⁹ *Ibid.*

⁶⁰ Article 170 is the most direct prohibition to promote hatred, although different forms of hatred are criminalized in Article 169 (discrimination), Article 1701 (organization of group, targeted to discriminate or hate), Article 1702 (Public endorsement or denial of certain war crimes), and others.

⁶¹ *Supreme Court of Lithuania*, ruling of 13 March 2018 in criminal case No. 2K-91-976/2018.

⁶² Gintarė Herasimenkienė, Miglė Keturkienė, “Neapykantos kalbos vertinimo gairės”: 258–266; in: Š. Zachar, J. Meteňko, M. Meteňková (eds.), *Zborník príspevkov: 17. medzinárodný kongres kriminalistika a forenzná veda: veda, vzdelávanie, prax, 16–17 September 2021*, Bratislava, Slovenská Republika (Akadémia Policajného zboru v Bratislave, 2021).

⁶³ See for e. g. the Supreme Court of Lithuania case No. 2K-91-976/2018 of 13 March 2018; the Supreme Court of Lithuania case No. 2K-58-489/2024 of 1 March 2024; the Constitutional Court of the Republic of Lithuania conclusion No. KT37-12/2024 in case No. 20/2023 of 25 April 2024 on the actions of Seimas member Remigijus Žemaitaitis.

⁶⁴ Council of Europe, *European Commission against Racism and Intolerance (ECRI) General Policy Recommendation No. 15 on combating hate speech of 8 December 2015*.

Europe Committee of Ministers 2022-05-20 Recommendation CM/Rec(2022)16 on combating hate speech⁶⁵, and analysis of the seminal ECtHR cases on hate speech, such as *Perinçek v. Switzerland* (2015)⁶⁶ or *Sanchez v. France* (2023)⁶⁷, among others. Despite the existence of various guidelines, criteria, and the case law of the ECtHR, a certain inconsistency has been observed in Lithuania in the assessment of hate speech cases.⁶⁸ This inconsistency was also noted by the ECtHR in its 2020 judgment in *Beizaras and Levickas v. Lithuania*.⁶⁹ The ECtHR criticized the criteria of “dangerousness” and “systematic nature” formulated in Lithuanian judicial practice, emphasizing that even a single comment inciting hatred may suffice for law enforcement authorities to take action, and that incitement to hatred does not necessarily have to involve calls for violence or other criminal conduct.⁷⁰ Nevertheless, the most common grounds for refusing to initiate pre-trial investigations continue to be the alleged lack of systematic conduct or insufficient dangerousness.⁷⁸ Recently, however, the situation has begun to show signs of change.

This shift was notably visible on 1 March 2024, the Supreme Court of Lithuania examined a criminal case in which an individual was convicted for publicly denigrating a member of the Seimas (Parliament of the Republic of Lithuania) on the basis of his sexual orientation by posting an insulting, derogatory, and discriminatory comment on the social network *Facebook*, in violation of Article 170(2) of the Criminal Code of the Republic of Lithuania.⁷¹ Taking into account ECtHR jurisprudence, the Supreme Court observed that, in the Court’s practice, comments posted on *Facebook* are considered as potentially harmful as comments made on online news portals, and it stressed that even a single hate-inciting comment must be treated with seriousness. The panel of judges held that such an assessment had indeed been carried out in the present case, leading to the defendant’s conviction for denigrating the victim—and, by extension, the relevant group of persons—on the grounds of sexual orientation, under Article 170(2) of the Criminal Code.⁷²

Consequently, while hate speech cases require deep contextual, historical, and dynamic analysis of all attributes, the typical starting point of the reasoning is a linguistic analysis of the speech. Where necessary, prosecutors, defence attorneys, or courts ask linguistic experts to assist in correctly interpreting the language, its overall content, the meaning of individual words in linguistic terms, and in distinguishing the dissemination of information from the expression of opinion, among other aspects.⁷³ Of course, the determination of whether a particular comment meets the elements of a criminal offence is made by the court, based on their own internal conviction, formed through a comprehensive and impartial assessment of all case circumstances, following

⁶⁵ Council of Europe, *Recommendation CM/Rec(2022)16 of the Committee of Ministers to member States on combating hate speech of 20 May 2022*.

⁶⁶ *Perinçek v. Switzerland*, European Court of Human Rights (Grand Chamber) (2015, Application No. 27510/08).

⁶⁷ *Sanchez v. France*, European Court of Human Rights (Grand Chamber) (2023, Application No. 45581/15).

⁶⁸ Monika Guliakaitė-Danisevičienė, *Neapykantos kalba Lietuvoje: situacijos ir atsisakymų pradėti ikiteisminius tyrimus apžvalga*, (2023) // <https://lt.efhr.eu/wp-content/uploads/2023/03/NEAPYKAN-TOS-KALBA-LIETUVOJE.pdf>

⁶⁹ *Beizaras and Levickas v. Lithuania*, European Court of Human Rights (Second Section) (2020, Application No. 41288/15).

⁷⁰ *Ibid.*

⁷¹ Lithuanian Supreme Court, 2K-58-489/2024.

⁷² *Ibid.*

⁷³ See for e. g. the Supreme Court of Lithuania case No. 2K-91-976/2018 of 13 March 2018; the Supreme Court of Lithuania case No. 2K-58-489/2024 of 1 March 2024; the Constitutional Court of the Republic of Lithuania conclusion No. KT37-12/2024 in case No. 20/2023 of 25 April 2024 on the actions of Seimas member Remigijus Žemaitaitis.

the requirements established by law and refraining from any pre-assigned value to any particular evidence.

This requirement for nuanced human analysis is not merely procedural but constitutional. Freedom of Expression is one of the most protected human rights by the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, the European Convention on Human Rights, the Charter of Fundamental Rights of the European Union, and the constitutions of the member states. Accordingly, any limitation of expression is a complex phenomenon involving not only linguistic interpretation but also the balance of dignity, liberty, and equality.⁷⁴ Therefore, each case of alleged hate speech should be subjected to a careful individual analysis, which makes it difficult to quantify the criteria or fully automate the processes. Moreover, quantification or automation may be misused to suppress dissenting voices by directing the system at individuals or groups not dedicated to expressing hatred, thus oppressing the freedom to speak openly and express unpalatable or unpopular views.⁷⁵

It is within this high-stakes courtroom context that the introduction of Large Language Models creates evidentiary friction. There is still an unanswered confusion on what legal power should be assigned to an LLM-generated outcome.⁷⁶ Is it to be considered a witness providing expert or percipient testimony, or should it be categorized as a statement by a litigant?⁷⁷ Generative AI, as a form of “intelligent machine”, on one hand, has the potential to provide new sources of information and, thus, a chance for more accurate fact-finding in criminal trials.⁷⁸ On the other hand, unlike traditional forms of evidence, generative AI outputs are often opaque, difficult to verify, and not subject to human cross-examination, which presents unique challenges to established evidentiary standards and procedural safeguards.⁷⁹

There are compelling legal grounds to exclude LLM-generated outcomes from the list of possible evidence and treat them as a mere statement by a litigant at most. Firstly, there have been no certification processes to support the reliability of the data generated for use in criminal trials because the products were not designed to accurately record specific data for this purpose, but rather to meet broader consumer demand. Secondly, the data is often owned by private corporations and stored in cloud-based systems, where it may be encrypted and shielded by manufacturers claiming trade secret privileges.⁸⁰ Thirdly, unlike human witnesses or physical evidence, AI outputs do not testify or possess an independent provenance that can be examined, which risks allowing evidence that cannot be effectively challenged or tested, undermining the adversarial process and clashing with the legal system’s foundational requirements of a fair trial.⁸¹

Consequently, while current legal standards rightly exclude LLMs from serving as independent witnesses or primary evidence due to these reliability and transparency

⁷⁴ Iginio Gagliardone, Danit Gal, Thiago Alves, Gabriela Martinez, “Countering online hate speech,” *UNESCO Publishing* (2015): 10 // <https://unesdoc.unesco.org/ark:/48223/pf0000233231>

⁷⁵ Haji Mohammad Saleem, Kelly P Dillon, Susan Benesch, Derek Ruths, “A web of hate: Tackling hateful speech in online social spaces,” *arXiv* (2017) // <https://arxiv.org/abs/1709.10159>

⁷⁶ Sabine Gless, “AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials,” *Georgetown University Law Center* 51 (2020) // doi: 10.5451/UNIBAS-EP76669; Hua Zhang, “The Evidentiary Value of AI-Generated Data: A Framework for Reliability and Admissibility,” *Education and Social Work* 2(2) (2025) // <https://doi.org/10.63313/ESW.9074>

⁷⁷ Sabine Gless, *supra* note 76.

⁷⁸ *Ibid.*

⁷⁹ Hua Zhang, *supra* note 76.

⁸⁰ *Ibid.*

⁸¹ *Ibid.*

issues, their utility lies elsewhere in the legal process. By shifting the focus from evidentiary proof to procedural assistance, the technology can be leveraged without violating the principles of a fair trial. In this regard, LLMs offer the potential to significantly reduce the cognitive burden and costs associated with enforcing the law. Notably, this utility extends beyond hate speech detection and includes automating the drafting of procedural documents. This provides an opportunity for legal professionals to streamline their workflows while retaining the final human oversight.

4. A CONCEPTUAL MODEL FOR LLM-ASSISTED HATE SPEECH ENFORCEMENT

Having established the opportunities and prerequisites for legal assistance in the post-detection phase, Figure 1 provides a conceptual model for an LLM-assisted workflow. This model is constructed as a Human-in-the-Loop system, aimed at bridging the transition from the automated detection of hate speech to the decision to initiate legal procedures. Such architecture ensures that while leveraging the stochastic power of LLMs to reduce the cognitive burden and costs of legal analysis, the ultimate decision-making authority and accountability remain strictly with the human operator.

The architectural logic is organized into three distinct operational domains, represented in the scheme as vertical pillars: the User, the LLM Assistant, and the Toolbox.

a) **The User.** The first pillar represents the human being, who, for personal or professional reasons, is interested in testing the specific act of speech. In this model the User is not merely a passive recipient of AI outputs but the active driver of the workflow. To address the evidentiary challenges discussed in the previous section, the User retains exclusive agency at every critical decision node – initiating the review, validating the preliminary opinion, and authorizing the final procedural action. The primary function of the User in this architecture is to exercise cognitive control and validate the AI's reasoning against their own internal conviction. This design effectively shifts the User's role from high-volume data processing to high-level quality assurance.

b) **The LLM Assistant.** The central pillar acts as the generative engine of the system. Unlike standard chatbot interfaces that provide a single, linear response, this component is designed to operate dynamically through distinct legal personas. By adopting the roles of Plaintiff, Defendant, and Judge, the LLM Assistant moves beyond simple text generation to perform adversarial reasoning – debating the merits of the case and presenting thus debate and conclusion to the User for an informed decision to proceed with legal action. Such multi-agent design aims to mitigate the risks of hallucination and algorithmic bias. It forces LLMs to challenge its own outputs through an adversarial debate, much alike in a typical court litigation.

c) **The Toolbox.** The third pillar serves as the grounding mechanism, effectively implementing a RAG framework. This section aims to provide a curated repository containing the relevant laws, Supreme Court case law, ECtHR guidelines, specific drafting templates, and ethical risk warnings. This component should ensure that every argument or opinion generated by LLM assistant is empirically anchored in valid material and procedural norms and best drafting practices.

That said, the workflow begins at the **Hate Speech Detection** phase, where raw textual input ("Speech") is filtered and probable actionable instances of illegal content are identified. Unlike the subsequent phases, which rely on generative reasoning, here the pre-trained AI model is utilised. As demonstrated by the earlier discussed works on

automated detection of hate speech, large generative models often struggle with the nuanced classification required for low-resource languages. Therefore, the Toolbox at this stage employs a specialized discriminative architecture (such as the encoder-based LitLatBERT) which has proven to offer superior performance in binary hate speech classification compared to general-purpose LLMs.

However, even specialized models are not perfect. Therefore this process imposes a mandatory user verification step ("Agree with AI?"), which serves as a Human-in-the-Loop checkpoint. This design explicitly accounts for the linguistic complexities identified by J. Ruzaitė⁸², specifically that Lithuanian hate speech is highly context-dependent and often indirect rather than explicit, relying on local knowledge and pragmatic inference. A purely statistical model might fail to detect culturally coded stereotypes or indirect aggression. To mitigate this risk, the workflow includes a decisive override mechanism. As illustrated in the scheme, if the AI classifies the text as safe ("No"), the User retains the authority to reject this assessment and proceed to the next stage regardless – first, with the neutral **Preliminary Opinion**, and then with the full Moot Court Simulation, as described below. This ensures that subtle, context-heavy violations do not escape legal scrutiny. Conversely, if the User agrees that the content is benign, the process terminates ("Drop"), preserving resources. Only when a consensus is reached, or the User exercises their override prerogative, does the system transition to the generative Moot Court phase.

The **Moot Court Simulation phase** transitions the system from simple classification to complex legal reasoning. Rather than asking the model for a single, binary determination, this phase employs an adversarial prompting approach, which is designed to replicate the nature of the adversarial legal system, ensuring that any conclusion reached is the result of a rigorous debate rather than a stochastic guess. The process is orchestrated through three distinct generative steps, mimicking the roles of the primary legal actors:

a) **The Plaintiff.** Triggered by a positive detection (or a user override), the first instance of the model is prompted to construct the strongest possible argument for prosecution. Drawing upon the legal definitions stored in the Toolbox it drafts an accusatory rationale, highlighting the specific linguistic markers that constitute the offence.

b) **The Defendant.** Immediately following this, a second instance in a distinct context window (or another LLM model) is tasked with the opposing role. It is prompted to dismantle the Plaintiff's arguments, targeting the common grounds for dismissal, such as the protection of Freedom of Speech, the lack of "systematic nature", or insufficient "dangerousness".

c) **The Judge.** The final step involves a synthesis of these conflicting viewpoints. Acting in the role of an impartial arbiter, this agent weighs the arguments of the Plaintiff against the counter-arguments of the Defendant and generates a final opinion in this Moot Court Simulation.

This three-part simulation serves as a measure to mitigate the risks of LLM hallucinations and confirmation bias. By forcing the LLM assistant to generate counter-arguments, the model engages in a form of self-critical reasoning. Assumedly, if the Defendant agent can easily refute the Plaintiff's claims using valid case law from the Toolbox, the Judge agent is statistically less likely to recommend proceeding with a

⁸² Jūratė Ruzaitė, *supra* note 56.

weak or hallucinated case, thereby filtering out false positives before they ever reach the drafting stage.

It shall be noted that the value of this simulation extends beyond the simulated Judge's conclusion. In this process, the User is presented with the full spectrum of possible arguments for and against the qualification of hate speech. This empowers the human agent to make an informed decision on whether the probability of a successful conviction justifies initiating costly and time-consuming legal proceedings. Consequently, this simulation should effectively filter out weak cases at the earliest possible stage.

The final phase of this workflow – **Claim Drafting** – is triggered only if the User, having reviewed the adversarial arguments and the impartial opinion, affirmatively decides to proceed with a legal action. At this stage, the system shifts its function from analytical reasoning to procedural formalization. The LLM is tasked with converting the substantiated arguments of the Plaintiff persona into a formal legal document suitable for submission to pre-trial investigation authorities or the court.

This step is designed to directly address the evidentiary and procedural challenges outlined earlier. Since LLM-generated text cannot serve as independent evidence, the drafting process is not an exercise in creative writing. Instead, it is rigidly constrained by the Toolbox, which injects specific rules of procedure, standardized templates, and ethical disclaimers into the context window. By grounding the output in approved legal forms, the model ensures that the draft claim complies with the formal requirements of Lithuanian law, thereby transforming raw linguistic analysis into a valid procedural instrument.

The workflow concludes with the **Decision on (In)action**, the ultimate assertion of human agency. The model presents the finalized draft to the User for review. The AI possesses no autonomy to initiate legal proceedings; it cannot file a document or contact authorities. The User must explicitly validate the draft and make the final physical choice to file the Claim or, upon further reflection, to Drop the matter. This hard-stop mechanism ensures that while the AI acts as a high-efficiency accelerator for detection and drafting, the moral and legal responsibility for the accusation remains exclusively with the human operator. Furthermore, the resolution of each case completes a recursive feedback loop by reintegrating relevant workflow data into the system's repositories, thereby continuously enriching the underlying corpora and refining the predictive capabilities of the AI models for future cycles.

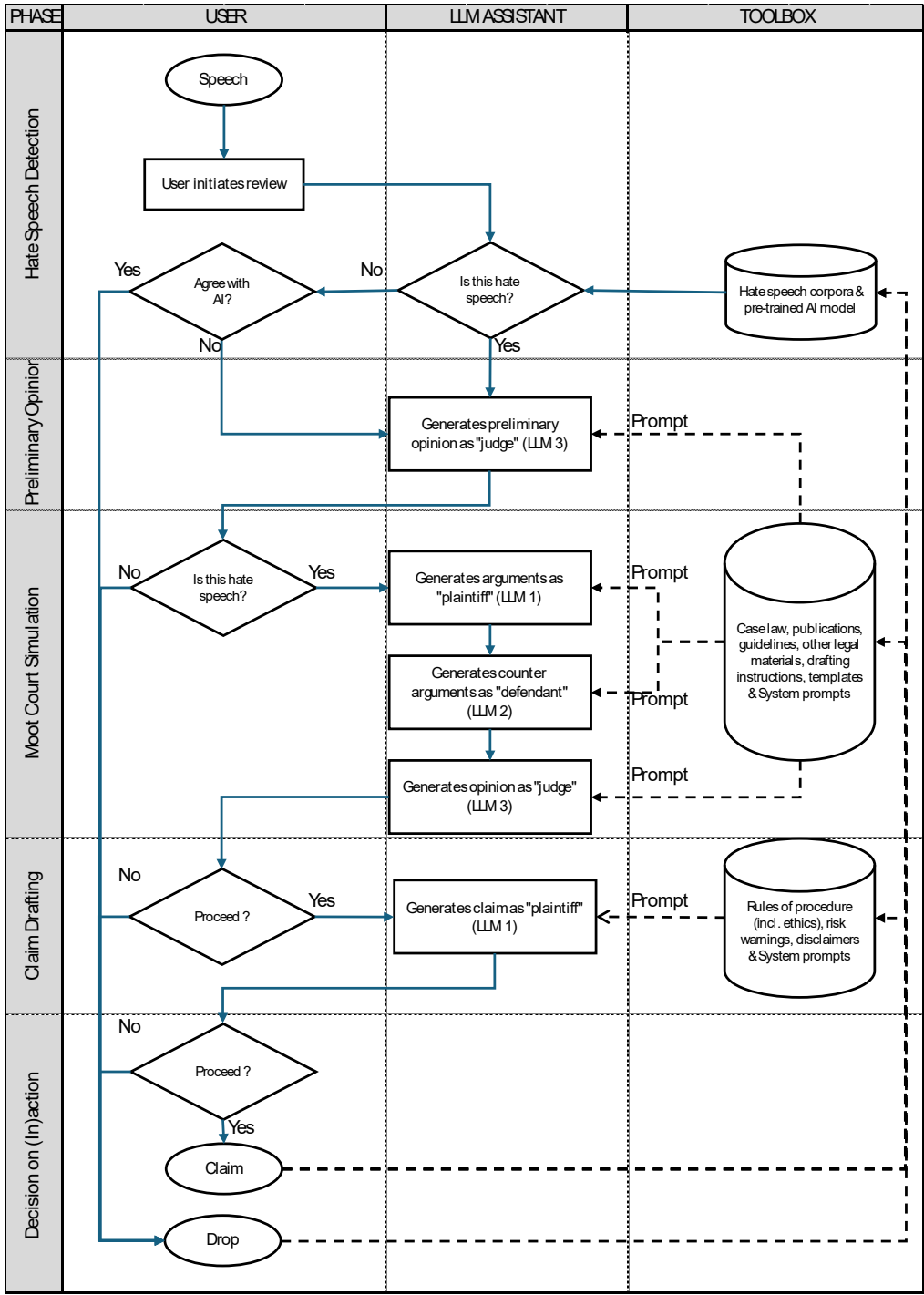


Fig. 1 Conceptual Model for LLM-Assisted Hate Speech Enforcement

5. EXPERIMENTAL VALIDATION OF THE CONCEPTUAL MODEL

To test the feasibility of the proposed Conceptual Model for LLM-Assisted Hate Speech Enforcement, a limited qualitative experiment for a Moot Court Simulation was conducted.

The objective of the experiment was not to measure statistical accuracy, but to evaluate the quality of legal reasoning and the effectiveness of the adversarial prompting architecture in a real-world Lithuanian context.

Model selection was guided by Artificial Analysis Intelligence Index, which combines a comprehensive suite of evaluation datasets to assess language model capabilities, including reasoning and knowledge⁸³. The leading model in this benchmark – Gemini 3 Pro – was selected for the experiments.

The dataset combined a case of a well-documented violation and a set of 12 less-documented speech samples. The former case involved five antisemitic statements⁸⁴ made by Lithuanian parliament member Remigijus Žemaitaitis. These statements have been subject to extensive media coverage, scholarly analysis⁸⁵ and most importantly, qualified as a breach of the parliamentary oath by the Lithuanian Constitutional Court.⁸⁶ This well-documented case provides a validated benchmark where the adversarial reasoning of parties is known and legal classification of hate speech is already established, allowing us to test the working of the abovementioned models. The latter part aimed to test the selected models on more varied and less covered speech samples, manually selected from the hate speech corpora developed by Kankevičiūtė et al. (2023)⁸⁷. A set of 12 samples was specifically chosen to represent four distinct scenarios (3 samples per scenario)⁸⁸, grouped by the classification of human and specifically trained AI model:

Table 1. The matrix of selected speech samples

	Neutral (AI)	Hate Speech (AI)
Neutral (Human)	False-False (FF) samples deemed not to be hate speech by both humans and AI model	False-True (FT) samples deemed not to be hate speech by humans, but flagged as such by AI model
Hate Speech (Human)	True-False (TF) samples deemed to be hate speech by humans, but flagged as not by AI model	True-True (TT) samples deemed to be hate speech by both humans and AI model

This diversified corpus allowed us not only to test the performance of a selected LLM on less-covered speech samples but also to compare results with a variety of automated detection outcomes, especially in cases of disagreement and related ambiguity (labelled as TF and FT in a sample).

A set of system prompts was engineered using a selected model (Gemini 3 Pro) for each role (Judge, Plaintiff, and Defendant) in the Lithuanian language. To ensure deterministic outputs, the prompt design moved beyond simple instructional queries, employing a combination of prompt engineering techniques:

- 1) To prevent “instruction drift” and clearly separate the model’s role from the input data, the prompts utilized XML tagging (e.g., <role>, <objective>, <workflow>, <constraints>). This syntax allows the model to process complex instructions as a structured algorithm rather than unstructured text.
- 2) Each instance was primed with a specific high-context persona (e.g., “Elite Prosecutor specializing in Article 170,” “Defense Attorney specializing in Freedom of Speech”). This technique guides the model’s to the specific legal vocabulary, tone, and argumentative strategies typical for that role.

⁸³ Artificial Analysis, *Artificial Analysis Intelligence Index* // <https://artificialanalysis.ai/models> (accessed 2025-12-13)

⁸⁴ See annex 1.

⁸⁵ Jūratė Ruzaitė, *supra* note 1.

⁸⁶ Constitutional Court of the Republic of Lithuania conclusion No. KT37-12/2024, *supra* note 63.

⁸⁷ Eglė Kankevičiūtė, Milita Songailaitė, Justina Mandravickaitė, *supra* note 11.

⁸⁸ See annex 2.

- 3) To prevent the model from jumping to a stochastic conclusion, the <workflow> section was included to enforce a step-by-step reasoning process. The model was instructed to perform a linguistic analysis first, followed by an objective legal qualification, an assessment of intent, and finally, a conclusion.
- 4) The simulation was designed as a sequential dependency chain. The Defendant prompt was engineered to receive the Plaintiff's generated memorandum as input context (<input_data>), with specific instructions to rebut those arguments. Accordingly, the Judge prompt was engineered to receive both the Plaintiff's and the Defendant's arguments, with the specific instructions to provide an impartial decision.
- 5) To address the Lithuanian language challenges, the prompts included specific constraints to detect implicit meanings (irony, dog-whistles) and strictly forbade the fabrication of case law.

The results of the experiments, summarized in *Table 2*, confirmed the structural integrity of the Moot Court Simulation architecture, showing no hardship related to the Lithuanian language. Throughout the simulation, the selected model demonstrated consistency in maintaining distinct agentic personas, effectively replicating the adversarial dynamics of a courtroom. The Plaintiff agent consistently constructed prosecutorial arguments, identifying linguistic markers of hate speech. Conversely, the Defendant agent consistently applied defense doctrines, prioritizing the Ultima Ratio principle and Freedom of Expression standards. Crucially, the Judge persona proved capable of independent evaluation, refusing to merely average the inputs.

Table 2. Summary of Moot Court Simulation experiments

Speech	Plaintiff	Defendant	Judge	Human Expert v. Judge
Žemaitaitis	Hate speech	Not hate speech	Hate speech	Aligns with the court ruling
FF1	Hate speech	Not hate speech	Not hate speech	Aligns with human annotator
FF2	Hate speech	Not hate speech	Not hate speech	Aligns with human annotator
FF3	Hate speech	Not hate speech	Not hate speech	Aligns with human annotator
TF1	Hate speech	Not hate speech	Not hate speech	Different from a human annotator
TF2	Hate speech	Not hate speech	Not hate speech	Different from a human annotator
TF3	Hate speech	Not hate speech	Not hate speech	Different from a human annotator
FT1	Hate speech	Not hate speech	Not hate speech	Aligns with human annotator
FT2	Hate speech	Not hate speech	Not hate speech	Aligns with human annotator
FT3	Hate speech	Not hate speech	Not hate speech	Aligns with human annotator
TT1	Hate speech	Not hate speech	Not hate speech	Different from a human annotator
TT2	Hate speech	Not hate speech	Not hate speech	Different from a human annotator
TT3	Hate speech	Not hate speech	Not hate speech	Different from a human annotator

This capacity for independent adjudication was most clearly evidenced in the calibration against the well-documented violation case (Remigijus Žemaitaitis). In this instance, where the ground truth was established by the Constitutional Court of Lithuania, the LLM Judge correctly identified all five statements as actionable hate speech. This

confirms that the model is capable of identifying and prosecuting hate speech when the legal threshold is clearly met.

However, a distinct pattern emerged in the comparative set of 12 less-documented samples. In 6 of these cases (Group FT and TF), the LLM Judge diverged from the Plaintiff's recommendation and ruled in favor of the Defendant.

The divergence between the human annotators (who flagged these samples as Hate Speech) and the LLM Judge (who dismissed them) provides an interesting insight into the outcome. Human annotators most likely relied on a semantic definition of hate speech (flagging content that felt hostile, rude, or offensive), while the LLM Judge applied a strict legal approach, distinguishing moral and legal wrongs. In all cases where the LLM overruled the human label, the LLM Judge also labeled language using terms such as "vulgar" (*vulgarus*), "unethical" (*neetiškas*), "primitive" (*primityvus*), and "morally condemnable" (*moraliniu požiūriu smerktinas*). However, by consistently citing the *Ultima Ratio* principle, LLM Judge refused to equate awful language with criminal wrong, therefore effectively filtering out "awful but lawful" cases.

Accordingly, the ultimate value of this three-part simulation extends beyond the specific verdict rendered by the AI Judge. The generation of well-developed, competing arguments provides the User with a certain legal roadmap. This transparency allows the human operator to step back from the immediate emotional reaction of being offended and objectively evaluate the litigation risks. By exposing the specific legal risks before any official documents are filed, the system empowers the User to make a prudent, informed decision. This acts as a vital filter, preventing the initiation of legal actions that, while emotionally justified, are legally weak and likely to result in dismissal, thereby saving significant time, financial resources, and the frustration of an unsatisfactory legal outcome.

CONCLUDING REMARKS

The experiment with the Moot Court Simulation demonstrated the viability of an LLM-centric resolution of the complexities inherent in the legal classification of hate speech and the process of enforcement of legal prohibitions of it. When embedded within a Plaintiff-Defendant-Judge adversarial triangulation, a state-of-art LLM is capable of effectively bridging the automated detection with the demands of legal enforcement. Thus, the specific utility of the proposed model is not merely processing velocity, but rather the capability to enforce the legal principles of proportionality, dangerousness, and intent with a consistency that arguably surpasses lay human observation or automated detection solutions.

Accordingly, the practical implications for the litigation attorneys or their clients, and the judiciary, are distinct yet complementary. For the former, the proposed system may serve as a pre-litigation filter, assisting in assessing litigation risk and ethically compelling counsel to advise against weak claims or defenses. For both, the system may function as a drafting auxiliary. Of course, neither party should rely on LLM. Still, the data-driven structuration of arguments could be used to ensure all legal angles are evaluated and integrated into the reasoning of the claim, rebuttal, or the final judgment.

Concerning the selection of specific AI architectures, reliance should not be placed on a single proprietary model. Of course, the state-of-the-art LLMs, capable of maintaining complex personas over extended context windows, should be utilized. However,

trust must be anchored in the standardization of the system prompts and the libraries (the Toolbox) used to ground the model in the local legal frameworks and practices. Ultimately, the human element remains non-negotiable. Only through a judicious and ethical approach, legal professionals can responsibly leverage GenAI, maximizing its transformative potential while mitigating potential risks.⁸⁹

The Guidelines developed by the European Bars Federation New Technologies Commission emphasize the imperative that European lawyers recognize and comply with the ethical and regulatory frameworks governing AI use, including the evolving EU Artificial Intelligence Act and relevant national regulations set by Bar Associations, regardless of whether such rules are legally binding.⁹⁰ Legal professionals are expected to remain vigilant, staying informed about regulatory updates and adapting practices accordingly. Generative AI should only enhance legal expertise, streamlining processes without replacing critical legal analysis or professional judgment. While GenAI can assist in drafting documents, lawyers must avoid overreliance, always verify AI outputs independently, and uphold the value of human decision-making.⁹¹

Moreover, the Ethical Charter on the use of AI in judicial systems, adopted in 2018 by the European Commission for the Efficiency of Justice (CEPEJ), also identified the risk of the so called “*effet moutonnier*” (sheep effect), wherein the judge may avoid responsibility and simply follow the algorithm’s advice.⁹² A lazy judge may tend to be satisfied with a solution proposed by AI without looking any further, thus drying up the evolution of case-law and failing to pay attention to the details of the concrete case.⁹³ The danger of a judge being captured by the algorithm, seeking the veneer of “scientificity” and “neutrality” which surrounds algorithmic evaluation⁹⁴, cannot be underestimated. It is therefore essential that a judge must maintain full autonomy of judgment and not base their decision exclusively on the indications derived from AI.⁹⁵ That is, judges must remain the final decision-makers. AI may provide support but cannot replace human legal judgment or autonomy. Judges must use AI as a tool, maintaining informed, user-level control over choices made by the system.⁹⁶

In conclusion, if it is accepted that legal practitioners must be cognizant of the evolution of information technologies, it follows that they must also be capable of a comprehensive and impartial examination of algorithmic prompts and outcomes.

⁸⁹ Natalie Pierce, Stephanie Goutos, “Why lawyers must responsibly embrace generative AI,” *Berkeley Business Law Journal* 21(2) (2024) // <https://doi.org/10.2139/ssrn.4477704>

⁹⁰ The European Bars Federation New Technologies Commission, *European lawyers in the era of ChatGPT. Guidelines on how lawyers should take advantage of the opportunities offered by large language models and gene*, (2023) // <https://www.fbe.org/nt-commission-guidelines-on-generative-ai/>

⁹¹ European Commission, *The General-Purpose AI Code of Practice* // <https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai>

⁹² JuLIA (101046631) Justice, fundamental rIghts and Artificial intelligence, Handbook *Artificial Intelligence, Judicial Decision-Making and Fundamental Rights*, (2024) // https://ssm-italia.eu/wp-content/uploads/2025/02/JuLIA_handbook-Justice_final.pdf. Council of Europe, European Commission for the Efficiency of Justice (CEPEJ), *European Ethical Charter on the use of Artificial Intelligence in Judicial Systems and their environment*, (2018) // <https://www.europarl.europa.eu/cmsdata/196205/COUNCIL%20OF%20EUROPE%20-%20European%20Ethical%20Charter%20on%20the%20use%20of%20AI%20in%20judicial%20systems.pdf>

⁹³ Maria Giuliana Civinini, “The perks of using AI tools in the justice domain,” JuLIA Handbook *Artificial Intelligence, Judicial Decision-Making and Fundamental Rights*, (2024) // https://ssm-italia.eu/wp-content/uploads/2025/02/JuLIA_handbook-Justice_final.pdf

⁹⁴ JuLIA, *supra* note 92.

⁹⁵ *Ibid.*

⁹⁶ NCSC, *AI & the courts: Judicial and legal ethics issues* // <https://www.ncsc.org/resources-courts/ai-courts-judicial-and-legal-ethics-issues>; Council of Europe, European Commission for the Efficiency of Justice (CEPEJ), *Reflections of the AIAB on the use of Artificial Intelligence in Judicial Systems*, (2025) // <https://rm.coe.int/cepej-aiab-2025-1rev5-en-reflections-of-the-aiab-on-the-use-of-artific/1680b42a56>; Council of Europe, *supra* note 92.

Furthermore, if legal education is to remain relevant, the curriculum must inevitably be expanded to include prompt engineering and the critical evaluation of adversarial AI outputs, thereby ensuring the requisite technological literacy for modern legal education and practice.

BIBLIOGRAPHY

1. Aleknaitė, Viltė. "The qualification problem of incitement against any nation, race, ethnic, religious or other group of people in Lithuanian criminal law." *Teisės apžvalga / Law Review* 28 (2023): 131–156.
2. Artificial Analysis. *Artificial Analysis Intelligence Index* // <https://artificialanalysis.ai/models> (accessed 2025-12-13)
3. Babbel. *How Many People Speak English, And Where is it Spoken?* // <https://www.babbel.com/en/magazine/how-many-people-speak-english-and-where-is-it-spoken?>
4. Bai, Yuntao. "Constitutional AI: Harmlessness from AI Feedback." *arXiv* (2022): 1–34 // <https://arxiv.org/abs/2212.08073>
5. Bilius, Mindaugas, Žaneta Navickienė, Vilius Velička. "Hate crimes: Evaluation of Lithuanian courts' decisions in the light of the practice of the European Court of Human Rights." *Baltic Journal of Law & Politics* 14(1) (October 2021): 22–47 // <https://doi.org/10.2478/bjlp-2021-0002>
6. Bozkurt, Aras, Ramesh C. Sharma. "Generative AI and prompt engineering: The art of whispering to let the genie out of the algorithmic world." *Asian Journal of Distance Education* 18(2) (2023).
7. Bozkurt, Aras. "Tell Me Your Prompts and I Will Make Them True: The Alchemy of Prompt Engineering and Generative AI." *Open Praxis* 16(2) (2024): 111–118 // doi: <https://doi.org/10.55982/openpraxis.16.2.661>
8. Butkienė, Rita et al. "Lithuanian Hate Speech Corpus v.1." *CLARIN-LT digital library in the Republic of Lithuania* (2025) // <http://hdl.handle.net/20.500.11821/69>
9. Chen, Shouyuan, Sherman Wong, Liangjian Chen, Yuandong Tian. "Extending Context Window of Large Language Models via Positional Interpolation." *arXiv* (2023): 1–18 // <https://arxiv.org/abs/2306.15595>
10. Chrysostomou, George, Zhixue Zhao, Miles Williams, Nikolaos Aletras. "Investigating hallucinations in pruned large language models for abstractive summarization." *Transactions of the Association for Computational Linguistics* 12 (2024): 1163–1181 // https://doi.org/10.1162/tacl_a_00695
11. Civinini, Maria Giuliana. "The perks of using AI tools in the justice domain." *JuLIA Handbook Artificial Intelligence, Judicial Decision-Making and Fundamental Rights* (2024): 20–37 // https://ssm-italia.eu/wp-content/uploads/2025/02/JuLIA_handbook-Justice_final.pdf
12. Contini, Francesco. "Unboxing generative AI for the legal professions: Functions, impacts, and governance." *International Journal for Court Administration* 15(2) (2024): 1–22 // <https://iacajournal.org/articles/10.36745/ijca.604>
13. Cossion, Manuel. "A comprehensive taxonomy of hallucinations in Large Language Models." *arXiv* (2025) // <https://arxiv.org/abs/2508.01781v1>
14. Council of Europe. *European Commission against Racism and Intolerance (ECRI) General Policy Recommendation No. 15 on combating hate speech of 8 December 2015*.

15. Council of Europe. European Commission for the Efficiency of Justice (CEPEJ). *European Ethical Charter on the use of Artificial Intelligence in Judicial Systems and their environment*. (2018) // <https://www.europarl.europa.eu/cmsdata/196205/COUNCIL%20OF%20EUROPE%20-%20European%20Ethical%20Charter%20on%20the%20use%20of%20AI%20in%20judicial%20systems.pdf>
16. Council of Europe. European Commission for the Efficiency of Justice (CEPEJ). *Reflections of the AIAB on the use of Artificial Intelligence in Judicial Systems*. (2025) // <https://rm.coe.int/cepej-aiab-2025-1rev5-en-reflections-of-the-aiab-on-the-use-of-artific/1680b42a56>
17. Council of Europe. *Recommendation CM/Rec(2022)16 of the Committee of Ministers to member States on combating hate speech of 20 May 2022*.
18. Dadurkevičius, Virginijus, Andrius Utkas. "Estimating the Amount of Lithuanian Text Indexed by Global Search Engines." *Baltic Journal of Modern Computing* 10 (2022) // doi: 10.22364/bjmc.2022.10.3.06
19. Dahl, Matthew, Varun Magesh, Mirac Suzgun, Daniel E Ho. "Large legal fictions: Profiling legal hallucinations in large language models." *Journal of Legal Analysis* 16(1) (2024): 64–93 // <https://doi.org/10.1093/jla/laae003>
20. Dao, Tri et al. "FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness." *arXiv* (2022): 1–34 // <https://arxiv.org/abs/2205.14135>
21. DeepSeek-AI, Daya Guo, et al. "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning." *arXiv* (2025): 1–22 // <https://arxiv.org/abs/2501.12948>
22. European Commission. *Project Be Hate-Free: Building Hate-Free Communities in Lithuania (Be Hate-Free)* // <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/org-details/925243619/project/963662/program/31076817/details>
23. European Commission. *The General-Purpose AI Code of Practice* // <https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai>
24. European Parliament. *Hate speech and hate crime in the EU and the evaluation of online content regulation approaches*. (2020) // <https://data.europa.eu/doi/10.2861/28047>
25. Farquhar, Sebastian, Jannik Kossen, Lorenz Kuhn, Yarin Gal. "Detecting hallucinations in large language models using semantic entropy." *Nature* 630 (2024): 625–630 // <https://doi.org/10.1038/s41586-024-07421-0>
26. Gagliardone, Iginio, Danit Gal, Thiago Alves, Gabriela Martinez. "Countering online hate speech." *UNESCO Publishing* (2015): 1–71 // <https://unesdoc.unesco.org/ark:/48223/pf0000233231>
27. García-Díaz, José Antonio, Ronghao Pan, Rafael Valencia-García. "Leveraging Zero and Few-Shot Learning for Enhanced Model Generality in Hate Speech Detection in Spanish and English." *Mathematics* 11, No. 24, 5004 (December 2023): 1–19 // <https://doi.org/10.3390/math11245004>.
28. Gless, Sabine. "AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials." *Georgetown University Law Center* 51 (2020): 195–253 // doi: 10.5451/UNIBAS-EP76669
29. Guliakaitė-Danisevičienė, Monika. *Neapykantos kalba Lietuvoje: situacijos ir atsisakymų pradėti ikiteisminius tyrimus apžvalga*. (2023) // <https://lt.efhr.eu/wp-content/uploads/2023/03/NEAPYKANTOS-KALBA-LIETUVOJE.pdf>

30. Gupta, Priya, Sachin Gupta. "Hate Speech Detection using OpenAI and GPT-3." *International Journal of Emerging Technology and Advanced Engineering*, Vol. 12, Issue 5 (May 2022): 132–138 // doi: 10.46338/ijetae0522_15
31. Hoffmann, Jordan, et al. "Training Compute-Optimal Large Language Models." *arXiv* (2022): 1–36 // <https://doi.org/10.48550/arXiv.2203.15556>
32. Herasimenkienė, Gintarė, Miglė Keturkienė. (2021). "Neapykantos kalbos vertinimo gairės": 258–266. In: Š. Zachar, J. Meteňko, M. Meteňková (eds.), *Zborník príspevkov: 17. medzinárodný kongres kriminalistika a forenzné vedy: veda, vzdelávanie, prax, 16–17 September 2021, Bratislava, Slovenská Republika*. Akadémia Policajného zboru v Bratislave.
33. Huang, Fan, Haewoon Kwak, Jisun An. "Is ChatGPT better than human annotators? Potential and limitations of ChatGPT in explaining implicit hate speech." *Companion Proceedings of the ACM Web Conference* (March 2023): 294–297 // <https://doi.org/10.48550/arXiv.2302.07736>
34. Ji, Ziwei, et al. "Survey of hallucination in natural language generation." *ACM Computing Surveys* 55(12) (2023): 1–38 // <https://doi.org/10.1145/3571730>
35. JuLIA (101046631) Justice, fundamental rights and Artificial intelligence. Handbook *Artificial Intelligence, Judicial Decision-Making and Fundamental Rights*. (2024) // https://ssm-italia.eu/wp-content/uploads/2025/02/JuLIA_handbook-Justice_final.pdf
36. Kalai, Adam Tauman, et al. "Why Language Models Hallucinate." *arXiv* (2025) // <https://arxiv.org/abs/2509.04664>
37. Kankevičiūtė, Eglė, Milita Songailaitė, Justina Mandravickaitė. "Neapykantos kalbos atpažinimas lietuviškuose komentaruose panaudojant dirbtinį intelektą." *Vilniaus universiteto atvirosios serijos* (May 2023): 26–34 // <https://doi.org/10.15388/LMITT.2023.3>
38. Kaplan, Jared, et al. "Scaling laws for neural language models." *arXiv* (2020): 1–30 // <https://arxiv.org/abs/2001.08361>
39. Katz, Daniel Martin, Michael James Bommarito, Shang Gao, Pablo Arredondo. "GPT-4 passes the bar exam." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 382(2243) (2024): 1–17 // <https://doi.org/10.1098/rsta.2023.0254>
40. Kostiuk, Yevhen, et al. "Towards Multilingual LLM Evaluation for Baltic and Nordic languages: A study on Lithuanian History." *arXiv* (2025): 1–11 // <https://doi.org/10.48550/arXiv.2501.09154>
41. Leviathan, Yaniv, Matan Kalman, Yossi Matias. "Fast Inference from Transformers via Speculative Decoding." *arXiv* (2022): 1–13 // <https://arxiv.org/abs/2211.17192>
42. Lewis, Partick, et al. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." *Proceedings of the 34th Conference on Neural Information Processing Systems* (2020): 1–16 // <https://proceedings.neurips.cc/paper/2020/file/6b-493230205f780e1bc26945df7481e5-Paper.pdf>
43. Mandravickaitė, Justina, et al. "Exploring hate speech detection models for Lithuanian language." *Proceedings of the 9th Workshop on Online Abuse and Harms (WOAH)* (2025): 206–218.
44. McGowan, Alessia, et al. "ChatGPT and Bard exhibit spontaneous citation fabrication during psychiatry literature search." *Psychiatry Research* 326, 115334 (2023): // <https://doi.org/10.1016/j.psychres.2023.115334>

45. Mikša, Katažyna, Monika Guliakaitė. "Neapykantos nusikaltimai ir neapykantos kalba: situacijos Lietuvoje apžvalga." (2020) // <https://lt.efhr.eu/wp-content/uploads/2020/12/NEAPYKANTOS-NUSIKALTIMAI-IR-NEAPYKANTOS-KALBA-SITUACIJOS-LIETUVOJE-AP%C5%BDVALGA.pdf>
46. Mikulėnienė, Danguolė, Agnė Čepaitienė. "Lithuanian Dialect Classifications." *Dialectologia* Special issue 11 (2023): 207–235 // doi: 10.1344/DIALECTOLOGIA2023.2023.8
47. Murauskienė, Dovilė, Raimundas Jurka, Jolanta Zajančauskienė. "The instigation of hatred: Questions of legal evaluation and procedural issues." *Entrepreneurship and Sustainability Issues* 8(2) (2020): 896–913 // [https://doi.org/10.9770/jesi.2020.8.2\(54\)](https://doi.org/10.9770/jesi.2020.8.2(54))
48. NCSC. *AI & the courts: Judicial and legal ethics issues* // <https://www.ncsc.org/resources-courts/ai-courts-judicial-and-legal-ethics-issues>
49. O'Connor, Sadie. "Generative AI." *Georgetown Law Technology Review* 8(2) (2024): 394–404 // <https://georgetownlawtechreview.org/generative-ai/GLTR-05-2024/>
50. Olla, Phillip, et al. "Promptology: Enhancing Human–AI Interaction in Large Language Models." *Information* 15(10), 634 (2024) // <https://doi.org/10.3390/info15100634>
51. Ouyang, Long et al. "Training language models to follow instructions with human feedback." *arXiv* (2022): 1–68 // <https://arxiv.org/abs/2203.02155>
52. Pan, Ronghao, José Antonio García-Díaz, Rafael Valencia-García. "Comparing fine-tuning, zero and few-shot strategies with large language models in hate speech detection in English." *Computer Modeling in Engineering & Sciences* 140(3) (July 2024): 2849–2686 // <https://doi.org/10.32604/cmes.2024.049631>
53. Pan, Ronghao, José Antonio García-Díaz, Rafael Valencia-García. "Leveraging Zero and Few-Shot Learning for Enhanced Model Generality in Hate Speech Detection in Spanish and English." *Mathematics* 11, No. 24, 5004 (December 2023): 16 // <https://doi.org/10.3390/math11245004>
54. Patil, Rajvardhan, Thomas F. Heston, Vijay Bhuse. "Prompt engineering in healthcare." *Electronics* 13(15), 2961 (2024): 1–26 // <https://doi.org/10.3390/electronics13152961>
55. Pengfei, Liu et al. "Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing." *arXiv* (2021): 1–46 // <https://arxiv.org/pdf/2107.13586>
56. Pierce, Natalie, Stephanie Goutos. "Why lawyers must responsibly embrace generative AI." *Berkeley Business Law Journal* 21(2) (2024): 1–51 // <https://doi.org/10.2139/ssrn.4477704>
57. Pimienta, Daniel. "Is it true that more than Half of Web Contents are in English? Not if Multilingualism is Paid Due Attention!" *Forum for Linguistic Studies* 6(5) (2024): 201–212 // doi: 10.30564/fls.v6i5.7144
58. Pleševičius, Dominykas. *Large Language Models for Lithuanian Language*. Master's thesis. VU, 2025.
59. Rafailov, Rafael, et al. "Direct Preference Optimization: Your Language Model is Secretly a Reward Model." *arXiv* (2023) // <https://arxiv.org/abs/2305.18290v3>
60. Ruzaitė, Jūratė. "In search of hate speech in Lithuanian public discourse: A corpus-assisted analysis of online comments." *Lodz Papers in Pragmatics* 14(1): 93–116 // <https://doi.org/10.1515/lpp-2018-0005>

61. Ruzaitė, Jūratė. *Neapykantos kalba: Monografija (Hatespeech: Monograph)*. Kaunas: Vytauto Didžiojo universitetas, 2024 // <https://doi.org/10.7220/9786094676109>
62. Ruzaitė, Jūratė. "Neapykantos kalba: medijų raštingumas iš lingvistinės perspektyvos": 166–188. In: Auksė Balčytienė, ed. *Artimas tolimes medijų pasaulis: virsmas, ribos ir būdai tai suprasti*. Kaunas: Vytauto Didžiojo universitetas, 2024 // <https://doi.org/10.7220/9786094676253>
63. Saleem, Haji Mohammad, Kelly P Dillon, Susan Benesch, Derek Ruths. "A web of hate: Tackling hateful speech in online social spaces." *arXiv* (2017) // <https://arxiv.org/abs/1709.10159>
64. Schick, Timo, et al. "Toolformer: Language Models Can Teach Themselves to Use Tools." *arXiv* (2023): 1–17 // <https://arxiv.org/abs/2302.04761>
65. Sclar, Melanie, Yejin Choi, Yulia Tsvetkov, Alane Suhr. "Quantifying Language Models' Sensitivity to Spurious Features in Prompt Design or: How I learned to start worrying about prompt formatting." *arXiv* (2024): 1–29 // <https://doi.org/10.48550/arXiv.2310.11324>
66. Shao, Minghao, Abdul Basit, Ramesh Karri, Muhammad Shafique. "Survey of Different Large Language Model Architectures: Trends, Benchmarks, and Challenges." *IEEE Access* 12 (2024): 188664–188706 // <http://dx.doi.org/10.1109/ACCESS.2024.3482107>
67. Smith, Craig S. *Hallucinations could blunt ChatGPT's success – IEEE spectrum*. (2023) // <https://spectrum.ieee.org/ai-hallucination>
68. The Encyclopedia Britannica. *Lithuanian language*. // <https://www.britannica.com/topic/Lithuanian-language>
69. The European Bars Federation New Technologies Commission. *European lawyers in the era of ChatGPT. Guidelines on how lawyers should take advantage of the opportunities offered by large language models and gene*. (2023) // <https://www.fbe.org/nt-commission-guidelines-on-generative-ai/>
70. Trad, Fouad, Ali Chehab. "Prompt Engineering or Fine-Tuning? A Case Study on Phishing Detection with Large Language Models." *Machine Learning and Knowledge Extraction* 6(1) (2024): 367–384 // <https://doi.org/10.3390/make6010018>
71. Vaswani, Ashish, et al. "Attention Is All You Need." *Proceedings of the 31st Conference on Neural Information Processing Systems* (2017): 1–11 // https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c-1c4a845aa-Paper.pdf
72. Velásquez-Henao, Juan David, Carlos Jaime Franco-Cardona, Lorena Cadavid-Higuera. "Prompt Engineering: a methodology for optimizing interactions with AI-Language Models in the field of engineering." *DYNA* 90 (230) (2023): 9–17 // <https://doi.org/10.15446/dyna.v90n230.111700>
73. Vidaus reikalų ministerija. *Ataskaita apie neapykantos nusikaltimų ir neapykantos kalbos situaciją Lietuvoje 2020–2021 m.* (2022) // https://vrm.lrv.lt/uploads/vrm/documents/files/LT_versija/Viesasis_saugumas/2020_2021%20ataskaita%20d%C4%97l%20neapykantos%20kalbos%20ir%20neapykantos%20nusikaltim%C5%B3.pdf
74. Vytautas Magnus University. *Features of hate speech and criteria for its identification: A linguistic and legal analysis*. // <https://www.vdu.lt/cris/entities/project/5eae72bf-e195-487a-b908-c9a835d15535>

75. Wang, Brydon T. "Prompts and Large Language Models: A New Tool for Drafting, Reviewing and Interpreting Contracts?" *Law, Technology and Humans* 6(2) (2024): 88–106 // <https://doi.org/10.5204/lthj.3483>
76. Wang, Han, et al. "Evaluating GPT-3 generated explanations for hateful content moderation." *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence AI for Good (IJCAI-23)* (2023): 6255–6263 // <https://doi.org/10.24963/ijcai.2023/694>
77. Walter, Yoshija. "Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education." *International Journal of Education Technology in Higher Education* 21, 15 (2024): 1–29 // <https://doi.org/10.1186/s41239-024-00448-3>
78. Webson, Albert, Ellie Pavlizk. "Do prompt-based models really understand the meaning of their prompts?" *arXiv* (2022) // <https://arxiv.org/pdf/2109.01247>
79. White, Jules, et al. "A prompt pattern catalog to enhance prompt engineering with ChatGPT." *arXiv* (2023): 1–19 // <https://doi.org/10.48550/arXiv.2302.11382>
80. Xu, Yuemei, et al. "A survey multilingual large language models: corpora, alignment, and bias." *Front. Comput. Sci.* 19, 1911362 (2025) // <https://doi.org/10.1007/s11704-024-40579-4>
81. Yamin, Muhammad Mudassar, Ehtesham Hashmi, Mohib Ullah, Basel Katt. "Applications of LLMs for generating cyber security exercise scenarios." *IEEE Access* (2024): 1–17 // <https://doi.org/10.21203/rs.3.rs-3970015/v1>
82. Zhang, Hua. "The Evidentiary Value of AI-Generated Data: A Framework for Reliability and Admissibility." *Education and Social Work* 2(2) (2025): 34–41 // <https://doi.org/10.63313/ESW.9074>
83. Zheng, Wang. "Research on generative artificial intelligence legal profession substitution." *Journal of Modern Law Research* 4(3) (2023): 32–40 // doi: 10.37420/j.mlr.2023.017.

LEGAL REFERENCES

1. *Beizaras and Levickas v. Lithuania*. European Court of Human Rights (Second Section) (2020, Application No. 41288/15).
2. Constitutional Court of the Republic of Lithuania conclusion No. KT37-12/2024 in case No. 20/2023 of 25 April 2024 on the actions of Seimas member Remigijus Žemaitaitis.
3. Criminal Procedure Code of the Republic of Lithuania, Official Gazette (2002, No. 37-1341).
4. Office of the Prosecutor General of the Republic of Lithuania. *Recommendations on the Methodological Recommendations on the Pre-Trial Investigations concerning Hate Crimes and Hate Speech*. (2023) // https://www.prokuraturos.lt/data/public/uploads/2023/09/20230726_neapykantos_nusikaltimai_rekomendacijos.pdf
5. *Perinçek v. Switzerland*. European Court of Human Rights (Grand Chamber) (2015, Application No. 27510/08).
6. *Sanchez v. France*. European Court of Human Rights (Grand Chamber) (2023, Application No. 45581/15).
7. Supreme Court of Lithuania case No. 2K-91-976/2018 of 13 March 2018.
8. Supreme Court of Lithuania case No. 2K-58-489/2024 of 1 March 2024.

Acknowledgements

In strict adherence to the principles of transparency and research integrity mandated by the European Research Area Forum, the authors explicitly disclose the utilisation of generative artificial intelligence technologies during the manuscript preparation process. Specifically, Perplexity was employed to facilitate the preliminary drafting of the abstract, whereas Gemini 3 PRO and Grammarly (v1.2.208.1781) were utilised for the purposes of linguistic refinement, translation, and structural optimization. It is imperative to emphasize that these tools functioned solely as auxiliary mechanisms; accordingly, the authors retain ultimate responsibility for the scientific accuracy, originality, and ethical integrity of the final output, ensuring full compliance with the normative standards established in the Living Guidelines on the Responsible Use of Generative AI in Research⁹⁷.

ANNEX 1: ANTISEMITIC STATEMENTS MADE BY LITHUANIAN PARLIAMENT MEMBER REMIGIJUS ŽEMAITAITIS

In Lithuanian (original)

„Pasirodo be Putino dar vieni gyvuliai atsirado Pasaulyje – IZRAELIS <...> Reikia tik konstatuoti, kad ši mokykla buvo pastatyta ES finansų dėka. Po tokių įvykių, nelieta nuostabos, kodėl gimsta tokie pasakymai: Lipo žydas kopėčiom ir nukrito netyčiom. Imkit, vaikai, pagaliuką ir užmuškite tą žyduką... Kas turi dar nutikti, kad Izraelis suprastų, kad tokios provokacijos ir tokie veiksmai tik dar labiau kelia pyktį ir kartu neapykantą žydams ir jų tautai“;

„Gal kitą kartą Izraelis numes bombą ant Palestiniečių ligoninės su prierašu su-sprogdinome, nes bijojome, kad jie mus puls <...> AŠ JUMS NORIU SUTEIKTI ŠANSO MIELI IZRAELIO ŽYDAI ATSIPRAŠYTI PALESTINOS IR ES, UŽ JŪSŲ ŠLYKŠČIUS VEIKSMELIUS SVETIMOJE ŠALYJE Ir pasikartosiu „Po tokių įvykių, nelieta nuostabos, kodėl gimsta tokie pasakymai: Lipo žydas kopėčiom ir nukrito netyčiom. Imkit, vaikai, pagaliuką ir užmuškite tą žyduką“;

„Vienas barbaras gyvena Rusijoje, dabar, pasirodo, Izraelis yra barbarai“, „Jei komunistuojanti ambasadorė labai jau atgailauja ir prašo atsiprašymo, tikiuosi, ta komunistuojanti ambasadorė viešai atsiprašys palestiniečių žmonių, palestiniečių vaikų už sugriautą mokyklą, kuri buvo pastatyta už mano pinigus ir už jūsų pinigus, nes tie pinigai buvo skirti iš Europos Sąjungos“;

„Šlykščiau mūsų Ministrė Pirmininkė I. Šimonytė ir negalėjo nepasakyti savaitgalį būdama Izraelyje! I. Šimonytė „Premjerė susitiko su Izraelio prezidentu: mūsų istoriniai saitai ir šalių draugystė yra tai, kuo didžiuojamės.“ Šlykštu yra tai, kad I. Šimonytė tokias kvailystes kalba, kai 1944 m. birželio 3 dieną Lietuvos ŽYDAI kartu su RUSAIS IŠŽUDĖ PIRČIUPŲ KAIMĄ IR TEN GYVENUSIUS ŽMONES <...> KIEK DAR ILGAI MŪSŲ POLITIKAI TOLIAU KLAUPSIS ŽYDAMS, KURIE ŽUDĖ MŪSŲ TAUTIEČIUS, PRISIDĖJO PRIE LIETUVIŲ SEKIMŲ, KANKINIMŲ IR MŪSŲ VALSTYBĖS NAIKINIMO <...> Buvo HOLOKAUSTAS ŽYDŲ, bet dar DIDESNIS HOLOKAUSTAS LIETUVIŲ BUVO LIETUVOJE!“;

⁹⁷ European Commission. Living Guidelines on the Responsible Use of Generative AI in Research, *ERA Forum Stakeholders' Document*, 2nd edn. (Brussels: Directorate-General for Research and Innovation, 2025) // https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf

„Mes, lietuvių tauta, niekada neturi pamiršti žydų ir rusų, kurie labai aktyviai prisidėjo prie mūsų TAUTOS NAIKINIMO!“

Translation to English

“It turns out that apart from Putin, yet more animals have appeared in the world – ISRAEL <...> It must only be stated that this school was built thanks to EU funds. After such events, there is no surprise why such sayings are born: A Jew climbed the ladder and accidentally fell. Take, children, a stick and beat that little Jew to death... What else has to happen for Israel to understand that such provocations and such actions only further raise anger and at the same time hatred toward Jews and their nation”

“Maybe next time Israel will drop a bomb on a Palestinian hospital with a note saying we blew it up because we were afraid they might attack us <...> I WANT TO GIVE YOU, DEAR ISRAELI JEWS, A CHANCE TO APOLOGIZE TO PALESTINE AND THE EU FOR YOUR DISGUSTING LITTLE ACTIONS IN A FOREIGN COUNTRY. And I will repeat again: ‘After such events, there is no surprise why such sayings are born: A Jew climbed the ladder and accidentally fell. Take, children, a stick and beat that little Jew to death’”

“One barbarian lives in Russia, now, it turns out, Israel are barbarians. If the communizing ambassador is so very repentant and begs for an apology, I hope that communizing ambassador will publicly apologize to the Palestinian people, to the Palestinian children, for the destroyed school, which was built with my money and with your money, because that money came from the European Union”

“More disgusting could not have been what our Prime Minister I. Šimonytė said over the weekend while in Israel! I. Šimonytė: ‘The Prime Minister met with the President of Israel: our historical ties and the friendship of our countries are something we are proud of.’ What is disgusting is that I. Šimonytė speaks such nonsense, when on June 3, 1944, Lithuanian JEWS together with RUSSIANS MASSACRED THE VILLAGE OF PIRČIUPIAI AND THE PEOPLE WHO LIVED THERE <...> HOW MUCH LONGER WILL OUR POLITICIANS CONTINUE TO KNEEL BEFORE THE JEWS, WHO KILLED OUR COUNTRYMEN, TOOK PART IN THE PERSECUTION AND TORTURE OF LITHUANIANS, AND THE DESTRUCTION OF OUR STATE <...> There was the HOLOCAUST of the JEWS, but an even GREATER HOLOCAUST OF LITHUANIANS TOOK PLACE IN LITHUANIA!”

“We, the Lithuanian nation, must never forget the Jews and Russians who very actively contributed to the DESTRUCTION OF OUR NATION!”

ANNEX 2: A SET OF 12 SAMPLES FROM KANKEVICIUTE ET AL. (2023) CORPORA

In Lithuanian (original)

	Neutral (AI)	Hate Speech (AI)
Neutral (Human)	FF1: „kas te rase apie lenkija kad jie is kitokio molio ir maziau serga taip ju 12 kartu daugiau bet ir mirusiu 14 kartu daugiau nei lietuvoje o tai 9 tukstanciais skirtumas ir t y dvigubai nei pas mus visu mirciu skaicius lenkija“ FF2: „nepritariu gp [galimybių pasas] jis negydo o tik platina legaliai virusą“ FF3: „gyvunu pasaulyje nera homoseksualiu santykiu kuo blogas šitas teiginys gamtoje viskas sutverta taip kad visi buna su priešingos lyties partneriais, nes yra prigintinis jausmas susilaukti palikuonių o kaip susilauks palikuonių tos pacios lyties gyvunai yra gamtos desniai neikit prieš gamta ir plius negali mokint kad 2 teciai arba 2 mamos pagimde vaika ir visi gyvena laimingai taip gamtoj nebuna“	FT1: „taip ir neaisku o su zoofilais ir kitais filais ka as zinau koku dar buna“. FT2: „lietuva lietuviams su dideliu tautiniu liudesių tegalima pasakyti kad kaip rodo istorija ir tautinio senojo tikejimo sugrizimo perspektyvos šiandienineje nuo sveiko proto išsilaisvinusioje ir išsivaikščiojancioje lietuvoje niekas negali parodyti kur pradingo seno tikejimo neprarade ar ji susigražine australijos aborigenai amerikos indenai ir kitos senu ir tikru tikejimu tautos“ FT3: „bartkus visiems nori irodyti kad asilas yra grynaveislis arklys gali buti kad to irodyti ir nepavyks“
Hate Speech (Human)	TF1: „moterys kalba daugiau nes jų liežuvis ilgesnis o mąstymas sisuresnis“ TF2: „tai aš butent ir emiau iš vokiecių reikia stipresnes rankos nei nacių“ TF3: „elektros srove reikia per ta koncertina paleisti ir ramu bus“	TT1: „laikas talibam bombe uzvest ir tuo viskas baigtusi tisina“ TT2: „liko zydu sako arabai skiepus butina kartoti ir kartoti kol neliks“ TT3: „boba už vairo avarija už kampo o kai dvi bobos susitinka tai ir kampo nereikia“

Translation to English

	Neutral (AI)	Hate Speech (AI)
Neutral (Human)	FF1: “who wrote here about Poland that they are from a different clay and get sick less, yes, there are 12 times more of them, but also 14 times more deaths than in Lithuania, and that is a difference of 9 thousand, which is double the number of all our deaths in Poland” FF2: “I do not support the gp [opportunity pass], it does not cure but only legally spreads the virus”	FT1: “it’s still not clear what to do with zoophiles and other -philes, what do I know what else there is”. FT2: “Lithuania for Lithuanians with great national sadness it can only be said that as history shows and the prospects for the return of the old national faith in today’s Lithuania, which has been liberated from common sense and is walking away, no one can show where the Australian Aborigines, American Indians and other nations of old and true faiths who have not lost their old faith or have regained it have disappeared”

Neutral (Human)	FF3: "in the animal world there are no homosexual relationships, what is wrong with this statement, in nature everything is created so that everyone is with partners of the opposite sex because there is an innate feeling to have offspring, and how will same-sex animals have offspring, there are laws of nature, do not go against nature and plus you cannot teach that 2 dads or 2 moms gave birth to a child and everyone lived happily ever after, it does not happen in nature"	FT3: "Bartkus wants to prove to everyone that a donkey is a purebred horse, it may be that he will not succeed in proving it"
Hate Speech (Human)	TF1: "women talk more because their tongue is longer and their thinking is narrower" TF2: "that's exactly what I took from the Germans, a stronger hand than the Nazis is needed" TF3: "an electric current needs to be run through that concertina wire and it will be quiet"	TT1: "it's time to drop a bomb on the Taliban and that would be the end of it, silence" TT2: "what's left are Jews, say the Arabs, the vaccines must be repeated and repeated until none are left" TT3: "a woman behind the wheel, an accident around the corner, and when two women meet, you don't even need a corner"