COMPARISON OF LLM FEW-SHOT VS. SYNTHETIC DATA APPROACHES FOR LITHUANIAN EVENT EXTRACTION

Arūnas Čiukšys[1] and Rita Butkienė[2]

[1]Department of Information Systems, Kaunas University of Technology, Kaunas, Lithuania
[2]Department of Information Systems, Kaunas University of Technology, Kaunas, Lithuania

ABSTRACT

Automatic Event Extraction (EE) identifies events from unstructured text. For Lithuanian, a lack of annotated corpora limits the progress. This study compares two strategies: (1) ML models trained on synthetic data generated by LLMs and (2) few-shot prompting with advanced LLMs (Open AI GPT, Google Gemini). Results show that while synthetic data offers broad coverage, it suffers from lower precision. Few-shot approaches achieve higher precision but are recall-sensitive and require advanced prompt engineering. A hybrid approach combining both methods could optimize outcomes. These findings provide insights for developing scalable EE solutions that address the unique challenges of resource-scarce languages.

KEYWORDS

Event Extraction, Few-Shot Prompting, Synthetic Data, Open AI GPT, Google Gemini, Lithuanian Language, NLP, Comparative Analysis.

1. INTRODUCTION

For under-resourced languages such as Lithuanian, event extraction is particularly difficult due to the scarcity of annotated data and the complexities of the language's morphology. Analyzing how synthetic data generation and few-shot LLM prompting address these challenges provides insight into the strengths and limitations of each method, as well as their potential for building reliable EE solutions.

1.1. Motivation and Context

Event Extraction remains a critical subtask within the broader field [1] [2] of Natural Language Processing, aiming to identify and extract events—along with their triggers and participants—from unstructured text sources [3] [4]. Studies in [5] have further underlined its importance across various domains. In resource-rich languages like English, publicly available annotated corpora [6] [7] and model repositories [8] [9] have summoned rapid development and notable breakthroughs [10] [11]. Continued improvements in these resources [12] [13] have further advanced the field. However, for under-resourced languages such as Lithuanian [14], the scarcity of labeled training data continues to hinder the evolution of robust EE systems.

We introduced Approach I, a novel synthetic data generation for Lithuanian EE using OpenAI GPT-generated labeled examples [15] (unpublished). This work demonstrates how large language models could replace time-consuming and expensive manual annotations required [1], effectively augmenting the limited Lithuanian corpus. The methodology therein showed the feasibility of employing synthetic data to build machine learning models that attained decent accuracy, albeit with varying levels of precision and recall.

 DOI:10.5121/ijci.2025.140101

Building on this foundation, Approach II [16] explored few-shot prompting techniques for Lithuanian EE, leveraging the capabilities of two advanced LLMs—Open AI GPT and Google Gemini. We explored both layered and combined prompting strategies, illustrating how even minimal annotated examples could significantly improve performance. The study affirmed LLMs' strong potential for direct (few-shot) event extraction, particularly in contexts where annotated data is scarce. However, few-shot approaches have their own trade-offs, such as occasional overreliance on prompt design and a limited capacity to handle out-of-distribution text variations.

1.2. Problem Statement

Despite notable progress in Lithuanian EE, a persistent bottleneck remains: the lack of large, high-quality, manually annotated corpora. Annotating event data for Lithuanian is especially demanding, given the language's rich morphological system and relatively free word order. These linguistic properties compound the challenge of building reliable extraction pipelines.

Table 1. Pros and Cons of Different EE Approaches

LLM-Generated Synthetic Data		**Few-Shot LLM Approaches**	
Pros	Cons	Pros	Cons
Reduces annotation overhead, generates large volumes of labeled examples quickly, and provides diverse training data.	Susceptible to LLM "hallucinations" or inaccuracies, requires careful prompt engineering to maintain data quality, and may introduce artificial patterns not reflective of real-world use cases.	Minimizes the need for large, annotated datasets, adapts flexibly to new event types, and can yield high accuracy with carefully crafted prompts.	Prompt sensitivity can lead to inconsistent results, difficult to generalize across domains, and still limited by the inherent constraints of the underlying LLM.

As shown in Table 1, two competing strategies have arisen to address this data scarcity. The main objective of this article is thus to critically compare the performance, strengths, and limitations of these two approaches.

Following the introduction, the article compares synthetic data generation (Approach I) and few-shot prompting with advanced LLMs (Approach II), focusing on methodologies, experimental results, and key metrics such as accuracy, precision, recall, and F1-score (Section 2 and 3). Section 4 synthesizes these findings, highlighting implications for Lithuanian event extraction and potential synergies between the methods. Practical applications, trade-offs, and hybrid strategies are discussed in Section 5, while Section 6 concludes with recommendations for advancing research in under-resourced languages.

2. BACKGROUND AND RELATED WORK

Event extraction is a pivotal NLP subtask that targets the identification of events—along with their triggers and participants—in unstructured text. While abundant annotated corpora and model repositories have spurred progress in resource-rich languages, under-resourced languages such as Lithuanian continue to face challenges stemming from limited labeled data and complex morphological features.

2.1. Overview of Event Extraction

Event Extraction [17] refers to the process of identifying event mentions in text and extracting corresponding triggers, arguments, and other relevant attributes. Standard benchmarks in this domain

include the Message Understanding Conference (MUC) series [18]—an early effort that shaped fundamental EE conventions—and the Automatic Content Extraction (ACE) [17] guidelines, which further refined the definitions of events, their triggers, and participant roles.

In resource-rich languages such as English, extensive annotated corpora have driven rapid progress in EE. However, less-resourced languages like Lithuanian lag due to limited labeled data and the complexity of their morphological and syntactic structures. Recently, large language models [19] [20], which are pre-trained on vast multilingual corpora [21] [22], have shown promise in bridging this resource gap. LLMs can be prompted, even with few or no examples [23], to perform tasks ranging from text generation to event detection. For Lithuanian, whose rich inflectional morphology and free word order complicate traditional rule-based approaches, LLM-driven methods open new avenues for scalable EE systems with minimal reliance on manually annotated resources.

2.2. Approach I - Synthetic Data Generation for EE

While few-shot methods minimize the necessity for annotations, Approach I took a complementary path by using LLMs to generate synthetic Lithuanian EE data at scale. Specifically, Open AI GPT was instructed to produce thousands of labeled sentences, each containing either a Contact. Meet or Contact. Phone-Write event—or no event at all. This Synthetic Data Generation (SDG) process was refined via Output Limitation Rules (OLR), which control the diversity of prompts and limit repetitive patterns in the generated text. The resulting datasets—1,290 sentences in the Small Dataset (SD) and over 54,000 in the Large Dataset (LD)—were used to train ML models (specifically the Sdca Maximum Entropy algorithm).

The advantage of this approach lies in rapidly creating sizable corpora without costly human annotation. Trained on the larger synthetic set, the EE model achieved substantially higher accuracy than the one trained on the smaller dataset. However, this approach also highlights inherent risks. Synthetic text may contain “hallucinations” or inaccuracies introduced by the LLM. These artifacts can degrade precision if the model learns patterns unrepresentative of real-world Lithuanian usage. Maintaining data quality and relevance is challenging and requires careful prompt design, thorough filtering processes, and continuous refinement.

2.3. Approach II - Few-Shot Prompting in Event Extraction

Approach II explored how few-shot prompting could mitigate the lack of Lithuanian EE datasets[24] by leveraging the strengths of modern LLMs—specifically, Open AI GPT and Google Gemini. In a few-shot setup, the model is provided with only a handful of annotated examples illustrating the task. By using these examples, the LLMs are guided to classify or annotate Lithuanian sentences with event types such as Contact. Meet and Contact. Phone-Write.

Two innovations were introduced to enhance model performance. First, a Layered Prompting Approach (LPA) that incrementally refines the model’s outputs through multiple prompt stages, ensuring that potential uncertainties in the LLM’s initial classification can be revisited with clearer definitions or additional context. Second, a Combined Approach (CA) that harnesses outputs from two different LLMs and either merges them (OR) or requires both models to agree (AND). These methods proved effective for Lithuanian, showing that carefully designed few-shot strategies could yield results competitive with more data-intensive solutions.

2.4. Comparison to Other NLP Efforts

Beyond Lithuanian, similar strategies have been adopted for other resourced and under-resourced languages[19] [25] [26] [21]. For instance, synthetic data generation has been explored for Arabic sentiment analysis, Indonesian text classification, and Chinese event extraction. In each case, LLM-

driven approaches—be they few-shot or synthetic-data-based—address the core bottleneck of insufficient annotated corpora. However, cross-lingual transfer learning has also been proposed, where knowledge from a resource-rich language model is adapted to a low-resource language through parallel data or translations. For Lithuanian, such transfer solutions remain largely unexplored or limited by the mismatch in linguistic structures.

Approach II few-shot paradigm and Approach I synthetic data approach both reflect broader trends in leveraging large language models for low-resource NLP. Each method offers distinct advantages—rapid adaptation with minimal data versus large-scale, systematically generated corpora—and faces challenges in prompt engineering, data fidelity, and model calibration. These two complementary strategies form the basis of the comparative discussion that follows.

3. METHODOLOGIES COMPARISON

An overview of the two methodologies examined for Lithuanian event extraction.

Table 2. Methodologies Comparison Table

Synthetic Data Approach (I)		**Few-Shot LLM Approach (II)**	
Key Components	**Scope and benchmarking**	**Key Components**	**Scope and benchmarking**
Prompt Engineering Dataset Creation ML Model Creation	Accuracy, precision, recall, and F1-score Contact. Meet and Contact. Phone-Write event types	Prompt Engineering	Accuracy, precision, recall, and F1-score Contact. Meet and Contact. Phone-Write event types

3.1. Recap of Methodologies

Approach I leverages Open AI GPT to generate large-scale synthetic datasets for Lithuanian event extraction, using structured prompt engineering and Output Limitation Rules to ensure both diversity and relevance in the generated corpus.

3.1.1. Synthetic Data Approach

Approach I uses Open AI GPT to generate synthetic Lithuanian event extraction datasets in two steps: (1) Prompt Engineering (defining event types, examples, and Output Limitation Rules to ensure diverse, relevant outputs) and (2) Dataset Creation (producing small and large synthetic datasets for training). Models trained with the Sdca Maximum Entropy algorithm.

3.1.2. Few-Shot LLM Approach

Approach II applied few-shot prompting to Open AI GPT and Google Gemini with two methods: a Layered Prompting Approach (LPA)—a three-step iterative process (IP, SP, SSP) to refine outputs—and a Combined Approach (CA) merging results (AND/OR) to validate or expand identified events. Both were benchmarked on the same gold corpus from Approach I, showing high accuracy but underscoring trade-offs in recall and precision.

3.2. Common Ground for Comparison

Both methodologies (Table 2) focused on two event types—Contact. Meet and Contact. Phone-Write. These event types were evaluated using a shared gold-standard corpus of manually annotated Lithuanian texts. This consistency provides a robust basis for comparing the performance of synthetic

data models and few-shot LLM approaches. The studies employed similar metrics (accuracy, precision, recall, and F1-score) to evaluate model performance.

4. EXPERIMENTS AND RESULTS COMPARISON

The metrics of interest include accuracy, precision, recall, and F1-score. We first briefly restate the main experimental results of each Approach and then provide an integrated comparison.

4.1. Summary of Approach I Results (Synthetic Data)

Table 3. Synthetic Data Approach Results

	Small Dataset	**Large Dataset**
Accuracy	57.97%	86.87%
Precision	15.45%	35.47%
Recall	82.45%	56.47%
F1-score	26.02$	43.57%

As shown in Table 3, the SD-trained model achieved high recall (82.35%) but low precision (15.45%), resulting in an F1 of 26.02%. The LD-trained model improved accuracy (86.87%) and F1 (43.57%) with more balanced precision and recall, though recall dropped to 56.47%. Increasing dataset size enhanced performance but introduced a recall-precision trade-off, highlighting the need for better prompt engineering and more diverse synthetic data.

4.2. Summary of Results from Approach II (Few-Shot LLM)

Table 4. Few-shot LLM Results

LLM, Methodology	**Accuracy**	**Precision**	**Recall**	**F1-Score**
GPT-mini	92,92%	74,76%	30,2%	43,02%
GPT-mini LPA	92,61%	63,64%	41,18%	50%
GPT-4o	93,17%	63,32%	56,86%	59,92%
GPT-4o LPA	92.82%	62.29%	49.41%	55.26%
Gemini 1.5 Flash	85,6%	35,79%	76,08%	48,68%
Gemini 1.5 Flash LPA	83,07%	31,72%	76,86%	44,91%
Gemini Pro 1.5	92,29%	57,69%	52,94%	55,21%
Gemini Pro 1.5 LPA	92,64%	60%	54,12%	56,91%
CA (GPT-4o + Gemini Pro 1.5) And	93,03%	67,7%	42,75%	52,41%
CA (GPT-4o + Gemini Pro 1.5) OR	92,43%	56,62%	67,06%	61,4%

As shown in the Table 4, in the second study, four LLMs (Google Gemini 1.5 Pro, Gemini 1.5 Flash, Open AI GPT-4o, GPT-4o mini) were evaluated using few-shot, layered, and combined approaches (CA). Layered prompting showed mixed results, with Gemini 1.5 Pro improving precision and recall, while GPT-4o saw minor changes. The CA OR strategy achieved higher recall (67.06%) and an F1 of 61.40%, while CA AND prioritized precision (67.70%) at the cost of recall (42.75%). Gemini 1.5

Flash excelled in recall but struggled with precision, while GPT-4o mini favored precision over recall. Few-shot prompting delivered strong accuracies and balanced performance, with CA OR emerging as a top strategy for recall improvement.

4.3. Direct Comparison: Synthetic-Data-Trained ML vs. Few-Shot LLM

Table 5. Direct Results Comparison Table

Approach	Accuracy	Precision	Recall	F1
Synthetic SD Model	57.97%	15.45%	82.35%	26.02%
Synthetic LD Model	86.87%	35.47%	56.47%	43.57%
Few-Shot LLM (best single) GPT-4o	93.17%	63.32%	56.86%	59.92%
Few-Shot LLM (best combined) CA OR	92.43%	56.62%	67.06%	61.40%

As shown in Table 5, the LD-based model (86.87% accuracy) already surpassed the SD model (57.97%), but it still lags behind GPT-4o (93.17%) and CA OR (92.43%).

Few-shot LLM systems deliver substantially higher precision than the synthetic-data-only model. Even the Gemini-based LLM with relatively low precision outperforms the SD model's 15.45%. The LD model narrows the gap somewhat, but still does not reach the 50–60% precision range commonly seen with GPT-4o or Gemini Pro.

The best synthetic-data F1 is 43.57% (LD model), while the top few-shot LLM approaches reach 59–61%. This suggests that the immediate "off-the-shelf" LLM usage for Lithuanian event extraction can yield better precision-recall balance than an ML model trained strictly on synthetic data—unless the synthetic data is drastically increased in volume or improved in quality.

In short, few-shot LLM strategies (especially when combining multiple models) consistently outperform the synthetic ML approach on accuracy, precision, recall, and F1. However, synthetic datasets still present a flexible, lower-cost pipeline for specialized tasks or domains where immediate access to LLMs or computational APIs is constrained.

5. DISCUSSION

By examining dataset size, morphological complexity, and the recall-precision balance, we uncover the key factors that shape performance in resource-scarce environments.

5.1. Interpretation of Comparative Findings

The comparative evaluation of a synthetic-data-based ML approach (Approach I) and a few-shot LLM approach (Approach II) highlights several dimensions influencing performance—most notably accuracy, precision, recall, and F1-score. One of the most striking observations is that larger synthetic datasets (e.g., LD with over 50,000 sentences) can boost accuracy and yield a balanced F1-score, yet still suffer from modest precision. Conversely, the few-shot LLM methods often exhibit higher precision (especially when using top-tier models like GPT-4o), at times sacrificing recall.

5.1.1. Dataset Size and Diversity

A primary reason why the synthetic-data model (particularly the LD model) can deliver strong accuracy is the sheer volume of training examples. Even though these are artificially generated, the model benefits from the extensive coverage of event structures—especially if the prompts are well engineered and ensure variety (via output limitation rules). However, the synthetic data may introduce certain repetitive or "hallucinatory" elements, affecting precision and occasionally leading to over fitted triggers or context mismatches.

5.1.2. Morphological Complexity of Lithuanian

Lithuanian's rich morphology and free word order pose challenges for both paradigms. The synthetic model benefits from repeated exposure to morphological variants in the data. However, if the synthetic examples lack truly representative morphological inflections, the model might generalize poorly, thus elevating false positives (and lowering precision). On the LLM side, few-shot approaches rely on robust internal multilingual capabilities; yet, these may still yield inconsistent handling of morphological nuances if the prompts are insufficiently detailed or if the model underestimates the complexities in syntactic structure.

5.1.3. Recall-Precision Trade-offs

In practice, seeking high recall (capturing most actual events) often increases false positives, while demanding high precision risks missing some valid events. The LD-trained synthetic model tended to miss fewer events overall than the smaller dataset model, but it still produced non-trivial false positives. In contrast, certain LLM runs, particularly those with layered prompting (Approach 2), tightened precision while occasionally failing to detect subtler event mentions.

Overall, the synthetic-data approach can yield robust general coverage but may require substantial prompt-engineering refinements to improve precision. By contrast, few-shot LLM methods often start with better precision but need additional prompting or combined-model strategies to attain strong recall.

5.2. Practical Implications

Selecting between a synthetic-data-trained model and a few-shot LLM depends on constraints such as computation, API access, and the need for high recall or precision. For offline or budget-limited scenarios, locally hosted models trained on synthetic data reduce manual annotation and avoid recurring API costs, though they require iterative curation to maintain precision. In contrast, few-shot LLMs offer faster iteration and broader coverage but incur ongoing fees. Balancing these factors, a hybrid approach—beginning with a modest dataset and using LLM prompting for fine-tuning or incremental annotation—may provide an effective middle ground.

6. CONCLUSIONS

A central takeaway is that quantitative comparison of both approaches (accuracy, precision, recall, and F1-score) confirm the ability to reach practical performance levels for Lithuanian EE with carefully designed strategies. Combining or layering these methods (as seen in the CA OR/AND and layered prompting approaches) often resulted in more balanced performance, highlighting a promising direction for future improvements.

6.1. Balancing Performance, Trade-offs, and Synergies

Few-shot LLMs consistently outperform the best synthetic-data-trained model (LD) by up to six percentage points in accuracy (e.g., 86.87% vs. over 92%) and 18 points in F1 (43.57% vs. 61.40%),

though each approach carries trade-offs: scaling synthetic data boosts precision but lowers recall, while certain LLM configurations optimize recall at the expense of precision. Layered and combined prompting (AND or OR) allows practitioners to toggle between high precision or high recall, while resource considerations favor synthetic models for offline use and LLMs for higher accuracy if cost is not prohibitive. Notably, synthetic data can fill lexical gaps that LLMs miss, and few-shot outputs can refine synthetic datasets by revealing underrepresented linguistic triggers.

6.2. Contributions

This comparative analysis clarifies best practices for under-resourced languages like Lithuanian, showing how synthetic data generation can bootstrap robust ML models—particularly as larger synthetic sets drive significant accuracy and F1 gains—and how few-shot LLMs can achieve strong precision and coverage with minimal labeled data. It also highlights how to weigh trade-offs in precision, recall, and cost when deciding between offline ML deployments or on-demand LLM usage.

6.3. Limitations and Future Directions

While this study advances Lithuanian event extraction, several limitations remain: coverage is restricted to two event types, the gold corpus is small and may not capture the language's full complexity, and prompt engineering can be further refined—especially for morphological variations. Although larger synthetic datasets improved performance, more sophisticated prompt-engineering could boost data diversity and reduce repetitive outputs. Future efforts could focus on broadening event coverage, enhancing gold-standard annotations, and improving methods for generating and refining synthetic data.

REFERENCES

[1] C. Lou, J. Gao, C. Yu, W. Wang, H. Zhao, W. Tu and R. Xu, "Translation-Based Implicit Annotation Projection for Zero-Shot Cross-Lingual Event Argument Extraction," in SIGIR '22: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022.

[2] Z. Sun, G. Pergola, B. Wallace and Y. He, "Leveraging ChatGPT in Pharmacovigilance Event Extraction: An Empirical Study," in Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 2: Short Papers), 2024.

[3] T. Wu, F. Shiri, J. Kang, G. Qi, G. Haffari and Y.-F. Li, "KC-GEE: knowledge-based conditioning for generative event extraction," World Wide Web (Springer), vol. 26, p. 3983–3999, 2023.

[4] C. Saetia, A. Thonglong, A. Thonglong, T. Amornchaiteera, T. Amornchaiteera, T. Chalothorn, T. Chalothorn, S. Taerungruang, S. Taerungruang, P. Buabthong and P. Buabthong, "Streamlining event extraction with a simplified annotation framework," Frontiers Artificial Intelligence, vol. 7, 2024.

[5] W. Liu, L. Zhou, D. Zeng, Y. Xiao, S. Cheng, C. Zhang, G. Lee, M. Zhang and W. Chen, "Beyond Single-Event Extraction: Towards Efficient Document-Level Multi-Event Argument Extraction," in Findings of the Association for Computational Linguistics ACL 2024, Bangkok, 2024.

[6] J. Pustejovsky, P. Hanks, R. Sauri, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day and L. Ferro, "The timebank corpus," Corpus linguistics, volume 2003, p. 40, 2003.

[7] Linguistic Data Consortium, "ACE 2005 Multilingual Training Corpus," 2006. [Online]. Available: https://catalog.ldc.upenn.edu/LDC2006T06.

[8] J.-D. Kim, T. Ohta, Y. Tateisi and J. I. Tsujii, "GENIA corpus—a semantically annotated corpus for bio-textmining," Bioinformatics, Volume 19, p. 180–182, 2003.

[9] R. Saurí and J. Pustejovsky, "FactBank: A Corpus Annotated with Event Factuality," Language Resources and Evaluation, Vol. 43, No. 3, pp. 227-268, 2009.

[10] S. Chen, A. Bies, K. Griffitt and S. S. Joe Ellis, "DEFT English Light and Rich ERE Annotation," Linguistic Data Consortium, 17 April 2023. [Online]. Available: https://catalog.ldc.upenn.edu/LDC2023T04. [Accessed 08 September 2024].

[11] X. Wang, Z. Wang, X. Han, W. Jiang, R. Han, Z. Liu, J. Li, P. Li, Y. Lin and J. Zhou, "MAVEN: A Massive General Domain Event Detection Dataset," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.

[12] T. O'Gorman, K. Wright-Bettner and M. Palmer, "Richer Event Description: Integrating event coreference with temporal, causal and bridging annotation," in Proceedings of the 2nd Workshop on Computing News Storylines (CNS 2016), 2016.
[13] A. Cybulska and P. Vossen, "Using a sledgehammer to crack a nut? Lexical diversity and event coreference resolution," in Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), 2014.
[14] G. A. Melnik-Leroy, J. Bernatavičienė, G. Korvel, G. Navickas, G. Tamulevičius and P. Treigys, "An Overview of Lithuanian Intonation: A Linguistic and Modelling Perspective," Informatica, vol. 33, no. 4, p. 795–832, 2022.
[15] A. Čiukšys and R. Butkienė, "Development of a Lithuanian Event Extraction Model Using a Training Dataset Generated by OpenAI GPT," unpublished manuscript, submitted to IEEE ACCESS, 09-2024, 2025.
[16] A. Čiukšys and R. Butkienė, "FEW-SHOT EVENT EXTRACTION IN LITHUANIAN WITH GOOGLE GEMINI AND OPENAI GPT," in CSITAI 2024, Dubai, 2024.
[17] Linguistic Data Consortium, ACE (Automatic Content Extraction) English Annotation Guidelines for Events, 2005.
[18] B. Sundheim and R. Grishman, "Message Understanding Conference - 6: A brief history," in Volume 1: The 16th International Conference on Computational Linguistics., 1996.
[19] J. Chen, P. Chen and X. Wu, "Generating Chinese Event Extraction Method Based on ChatGPT and Prompt Learning," Applied Sciences, 2023.
[20] Google, "Google Blog," Google, 2024. [Online]. Available: https://blog.google/technology/ai/google-gemini-next-generation-model-february-2024/.
[21] R. Chen, C. Qin, W. Jiang and D. Choi, "Is a Large Language Model a Good Annotator for Event Extraction?," in Proceedings of the AAAI Conference on Artificial Intelligence, 2024.
[22] OpenAI, "ChatGPT," 2022. [Online]. Available: https://openai.com/.
[23] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child and A. R. Dan, "Language Models are Few-Shot Learners," in Advances in Neural Information Processing Systems 33 (NeurIPS 2020), 2020.
[24] "GitHub, KTU_Lithuanian_Events_Dataset," 2024. [Online]. Available: https://github.com/device7/KTU_Lithuanian_Events_Dataset.
[25] C. Suhaeni and H.-S. Yong, "Mitigating Class Imbalance in Sentiment Analysis through GPT-3-Generated Synthetic Sentences," Applied Sciences, 2023.
[26] O. Litake, B. H. Park, J. L. Tully and R. A. Gabriel, "Constructing synthetic datasets with generative artificial intelligence to train large language models to classify acute renal failure from clinical notes," Journal of the American Medical Informatics Association, Volume 31, Issue 6, June 2024, p. 1404–1410, 2024.

AUTHORS

ARŪNAS ČIUKŠYS received the B.Sc. and M.Sc. degrees in computer science from the Kaunas University of Technology, in 2012 and 2016, respectively, where he is currently pursuing the Ph.D. degree with the Faculty of Informatics. His research interests include machine learning, AI, information system engineering, databases, software, and information extraction.

RITA BUTKIENĖ received the B.Sc., M.Sc., and Ph.D. degrees from the Kaunas University of Technology, in 1993, 1995, and 2002, respectively. She is currently a Professor at the Kaunas University of Technology. Her research interests encompass information system engineering, ontologies, semantic technologies, databases, distributed ledgers, information extraction, and information retrieval.