

Enhancing Legal Compliance and Regulation Analysis with Large Language Models

Shabnam Hassani 

University of Ottawa (Email: s.hassani@uottawa.ca)

Abstract—This research explores the application of Large Language Models (LLMs) for automating the extraction of requirement-related legal content in the food safety domain and checking legal compliance of regulatory artifacts. With Industry 4.0 revolutionizing the food industry and with the General Data Protection Regulation (GDPR) reshaping privacy policies and data processing agreements, there is a growing gap between regulatory analysis and recent technological advancements. This study aims to bridge this gap by leveraging LLMs, namely BERT and GPT models, to accurately classify legal provisions and automate compliance checks. Our findings demonstrate promising results, indicating LLMs’ significant potential to enhance legal compliance and regulatory analysis efficiency, notably by reducing manual workload and improving accuracy within reasonable time and financial constraints.

Index Terms—Legal Compliance, Legal Requirements, Large Language Models, GPT-4, GPT-3.5, BERT, GDPR, Food Safety, Industry 4.0, Internet of Things.

I. INTRODUCTION

The evolution of Industry 4.0 and regulations like GDPR demand innovative approaches to ensure that food safety software systems and regulatory artifacts, such as Data Processing Agreement (DPAs), comply with legal standards. Our review of the literature and existing methodologies reveals a gap in applying LLMs to domain-specific regulatory analysis. This study is propelled by the need to automate the interpretation and application of legal provisions in the food industry and data processing regulatory artifacts and is also motivated by industry partners. This research aims to develop an LLM-based methodology for classifying food safety and checking the compliance of regulatory provisions, seeking to outperform traditional methods in efficiency, accuracy, and cost.

My research is driven by two main objectives: to develop an automated methodology for classifying legal provisions pertaining to food safety, and to enhance the efficacy and interoperability of compliance checking of regulatory artifacts using LLMs. I propose that LLMs can significantly outperform existing methods in these tasks, thereby enhancing legal compliance and regulation analysis with improvements in terms of accuracy, time, and financial cost.

A. Research Questions

The main research questions (RQs) are:

RQ1: How accurate is the legal requirements classification approach? RQ1 examines alternative realizations of our classification approach using different BERT variants and GPT-3.5-turbo. To answer RQ1, we report accuracy using the standard classification metrics.

RQ2: How does the legal requirements classification approach fare against baselines? To assess whether our approach offers benefits over simpler solutions, RQ2 compares our approach against two baselines: one based on LSTM and the other based on keyword search. To answer RQ2, we report accuracy using the standard classification metrics.

RQ3: How do state-of-the-art LLMs, both open-source and closed-source, fare against one another in terms of accuracy for zero-shot learning and fine-tuning in regulatory compliance checking? This research question compares generative LLMs in terms of their ability to understand and apply regulatory requirements. To answer RQ3, we report accuracy using the standard classification metrics.

RQ4: Compared to traditional sentence-level analysis, how does incorporating paragraph context and the textual specification of compliance rules enhance the performance of compliance checking? This research question aims to assess the enhancement in accuracy of compliance checking brought about by the integration of paragraph-level context and rules, as opposed to traditional single-input sentence-level methods. To answer RQ4, we report accuracy using the standard classification metrics.

RQ5: What are the cost and time implications associated with our proposed approach? To be practical, our approach must generate predictions within a reasonable time. RQ5 provides a practical assessment of the approach’s resource efficiency, considering time and costs in terms of computing and financial resources.

Structure. Section II presents background. Section III compares the proposed method with related work. Section IV describes our proposed method. Section V provides current evaluation results. Section VI lays out the future research and the timeline.

II. BACKGROUND

In this section, the necessary legal and technical background for my approach is provided.

A. Food-safety Regulations

To ensure that the food supply is safe, countries around the world have established regulatory frameworks to govern the production, distribution, and consumption of food products. Our primary focus is on the Safe Food for Canadians Regulations (SFCR). SFCR consolidates several federal Canadian food regulations that apply, among other things, to the import, export, and inter-provincial trade of food, including

ingredients. In many cases, SFCR defers the elaboration of requirements for specific food products to Food-Specific Requirements and Guidance (FSRG) regulations. We use content from FSRG regulations both in our qualitative analysis and in our evaluation. In addition to SFCR and FSRG regulations, we consider selected parts of the food-safety regulations by the US Food and Drug Administration (FDA) to examine the extent to which our classification solution is transferable to jurisdictions outside Canada.

B. General Data Protection Regulation (GDPR)

GDPR is a privacy and data protection law enacted by the European Union to regulate the processing of personal data and enhance individuals' control over their information. We use Data Processing Agreements (DPAs), as defined by GDPR, to illustrate ideas and conduct evaluations. DPAs are binding agreements that specify the responsibilities and rights of controllers and processors. Most of the content in DPAs is software-related, making them of direct interest to the RE community [1]. For GDPR compliance, DPAs must follow GDPR Article 28.

III. RELATED WORK

Significant progress has been made in compliance checking of regulatory artifacts. Below we discuss limitations in existing research for compliance checking fall under two categories.

For our classification of requirements-related legal content from food safety regulations, we consider three prior lines of work: (1) extracting metadata from legal texts, (2) applying LLMs to natural-language requirements, and (3) monitoring food safety through IoT. To our knowledge, no prior work exists on automated classification of requirements-related provisions in food-safety regulations.

A. Related Work on Compliance Checking of Legal Content

1) *Reliance on sentences as units of analysis*: Existing approaches mostly promote a sentence-by-sentence analysis of legal texts [1]–[4]. We have observed three key issues caused by this decision that can significantly affect accuracy. First, interpreting sentences often requires contextual understanding beyond the immediate sentence structure, as sentences can be related to previously mentioned categories or definitions. Second, it is common for a legal concept to be defined, with subsequent sentences drawing upon the definition. Third, legal texts are frequently interwoven with cross-references [5]. This practice adds complexity by requiring an understanding of the broader content across a collection of legal documents.

2) *Automation strategies lack justification for decisions and are either coarse-grained or entail significant manual effort to build*: Automation for legal texts employs one of the following three strategies or a combination thereof. We argue that these strategies not only fail to offer justification and rationale for decisions – a capability now achievable with LLMs – but they either are too coarse-grained, potentially compromising accuracy, or require considerable manual effort, making compliance automation costly to implement.

Strategy 1: Extracting metadata or semantic frames and then executing predefined rules over the metadata/frames to check compliance such as [1]–[3], [6], [7].

Strategy 2: Projecting regulatory texts into an embedding space and then measuring the similarity between these texts against (textual) compliance rules or a labelled group of regulatory provisions. An example of research employing this strategy is that of Amaral et al. [1], [2], who create sentence embeddings and utilize semantic similarity for completeness checking of privacy policies and DPAs.

Strategy 3: Using encoder-only single-input transformer models, such as BERT [8], to classify which compliance rules are satisfied by each legal text. An example of work applying this strategy is that of Ilyas et al. [4], who use BERT to construct a multi-label classification pipeline for identifying sentences in DPAs that meet specific compliance rules.

B. Related Work on the Classification of Requirements-related Legal Content

1) *Metadata Extraction from Legal Texts*: Legal documents are typically lengthy and complex, making it difficult for individuals to locate relevant information quickly and accurately. This has led to the development of automated approaches for information extraction from legal texts such as [9]–[12]. These works, despite being amongst the first to focus on information extraction from legal texts, do not use ML techniques and are limited to privacy policies.

Studies such as [1]–[3] differ from ours in both domain and the utilization of feature-based learning as opposed to LLM-based approaches.

Some studies link privacy policies to software code [13]–[15]. In contrast, our approach focuses on food-safety regulations. We use BERT variants and GPT for classification and information extraction.

Zeni et al. [16], [17] and Sleimi et al. [18] extract semantic metadata from legal texts using NLP and semantic-web techniques, but their focus and domain differ from ours. Abualhaija et al. [19] employ LLMs for question answering over GDPR; they have different analytical goals and inputs from ours.

2) *LLMs and Natural-language (NL) Requirements*: Several studies use LLMs, particularly BERT, for analyzing NL requirements, with tasks including classifying non-functional requirements [20]–[22], classifying and summarizing contractual obligations [23], [24], detecting requirements smells [25], classifying security requirements [26], classifying user feedback [27], classifying requirements dependencies [28], detecting causality [29], classifying and clustering coreferences [30], [31], transforming NL requirements into formal specifications [32], predicting issue links [33], classifying issue sentences [34], identifying similar requirement [35], checking completeness [36], and generating elicitation-interview scripts [37]. These studies showcase the versatility of LLMs in requirements-analysis tasks. We have benefited from the best practices in the above-cited strands of work; however, our analytical objectives set us apart.

3) *Food-monitoring Systems*: Food monitoring is a crucial area of research for improving food safety and enhancing supply-chain efficiency. Bouzembrak et al. [38] have conducted a systematic literature review to examine the potential of IoT for food safety and quality monitoring, as well as food traceability and supply chain. They observe that most existing IoT research aimed at these objectives focuses on measurement solutions using sensors and communication technologies. These solutions are technology-driven and not directly linked or traceable to regulations. Our qualitative study highlights, from a regulatory perspective, the importance of all the core measurement types identified in the literature. The close alignment between the measurements implied by food-safety regulations and the existing measurement technologies confirms the validity of our interpretation of systems/software relevance in the context of food safety.

IV. PROPOSED METHOD

We conduct a comprehensive qualitative analysis of food safety regulations followed by utilizing LLMs, specifically BERT and GPT-3.5. We also conduct an analysis of compliance checking of regulatory artifacts with regulations followed by utilizing LLMs specifically GPT-3.5, GPT-4, Mixtral, and BERT. The methodology is further described in detail.

A. Automated Classification of Legal Provisions

Figure 1 presents an overview of our approach for the automated classification of requirements-related legal content, specifically instantiated for requirements-related concepts in food-safety regulations. The approach takes as input the textual content of the regulations being examined and produces as output labels for each provision in the input. The approach treats each sentence as one provision. Stated otherwise, our *unit of analysis* is a sentence since larger units like paragraphs and sections may not be precise enough for this task. Our approach has four steps: (1) Pre-processing, (2) LLM-based Classification, (3) Keyword-based Classification, and (4) Label Prediction (Fig 1).

Step 1) Pre-processing. In this step, we split the input into sentences. Legal documents frequently use lists. To ensure preservation of context during sentence splitting, we follow Sleimi et al. [18] and add the header in each list, known as a list-item prefix, to each itemized/enumerated item.

Step 2) LLM-based Classification. Using an LLM fine-tuned on labelled data from our qualitative study, we classify

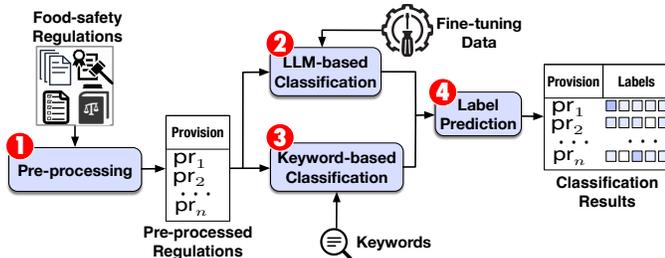


Fig. 1: Overview of our automated classification approach.

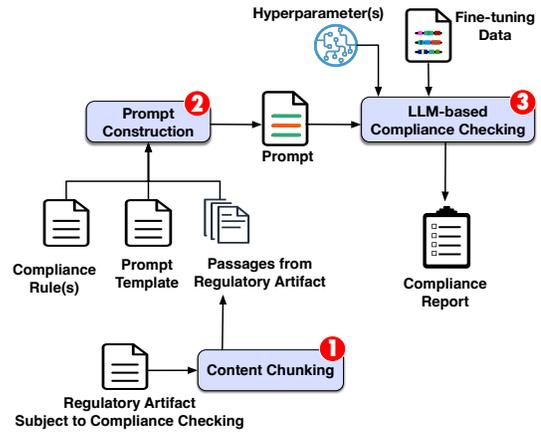


Fig. 2: Overview of our automated compliance checking approach.

provisions with occurrences of concepts from our conceptual model. In our evaluation, we examine BERT and GPT as alternative LLMs. For GPT, the pre-processed regulations from Step 1 are embedded into prompts.

Step 3) Keyword-based Classification. Our qualitative study did not yield a sufficiently large number of examples for a few concepts: *Colour*, *Firmness*, *Pathogen*, and *Water Content*. As previous studies on automated classification of legal texts point out [2], [18], [39], some concepts may turn out to be too scarce for learning-based techniques, thereby warranting the use of keywords for classification. In the third step of our approach, we conduct a keyword lookup for the four scarce concepts mentioned above. The keywords associated with each concept were derived during our qualitative study. If a provision P contains one or more of the keywords associated with a concept M , we label P with M .

Step 4) Label Prediction. This step combines the labels computed by the LLM-based classifiers (Step 2) and those computed by the keyword-based classifiers (Step 3) to produce the final label recommendations for each provision.

B. Automated Compliance Checking

Figure 2 presents an overview of our approach for automated compliance checking, which systematically employs LLMs to assess regulatory artifacts against specific regulations. The approach consists of three steps. The first step (1) is to create passages from the regulatory artifact that the user wants to verify for compliance, e.g., a DPA. The second step (2) is to generate a prompt for LLMs. This step takes as input the passages generated in the first step, along with the compliance rules to be verified (e.g., from the GDPR), and a configurable prompt template. The third step (3) is to present the prompt generated in the second step to LLMs to draw inferences about compliance and non-compliance. The output of the approach is a compliance report, accompanied by an explanation and justification for each determination.

Step 1) Content Chunking. The regulatory artifact subject to compliance checking is fed to the pipeline. The pipeline segments the regulatory artifact into manageable content chunks,

primarily paragraphs (Fig. 2, Step ①). This process ensures that the chunks, once incorporated into the overall prompt, will fit within the token limit of the LLM. If a full paragraph were to exceed the token limit, it would need to be either truncated or summarized. However, in our investigation of LLMs, we did not encounter such situations due to their reasonably large token limit. The output of this step is a set of passages each within the LLM’s token limit, ready for prompt construction.

Step 2) Prompt Construction. Our approach constructs tailored prompts to guide the LLMs in eliciting rule-specific responses, based on the input compliance rules. Prompts, designated to extract both the rule applicability and the LLMs’ explanation and justification, consist of three parts: *Prompt Template*, *Compliance Rules*, and *Passages from Regulatory Artifact* (Fig. 2, Step ②).

The prompt is presented in a structured chat format for the LLMs, specifying roles for system instructions, user inquiries, and the assistant’s responses, following best practices where applicable [40], [41]. The system role provides instructions for the model to follow (detailed in the Prompt Template). The user role presents input that model should respond to (Passages from Regulatory Artifact and Compliance Rule(s)). The assistant role represents the model’s response to the user’s input (Compliance Report). *Prompt Template: You are a legal expert trained to identify applicable {Compliance Rules} based on a given {text} within its specific {context}. When provided with the {text} and its {context}, your response should only include the rule identifier (e.g., 'R5') if applicable. If there is no direct connection to any Compliance Rule within the context provided, respond with 'R99'. Follow this format strictly. Then, provide your rationale for the decision.*

Step 3) LLM-based Compliance Checking. Once tailored prompts are constructed, we employ zero-shot learning, where the model generates responses without prior training on similar tasks, based solely on the instructions provided in the prompt. In the future, this process may be enhanced with (optional) fine-tuning to further refine the models’ responses to the specific language of DPAs. Subsequently, the generated responses are analyzed to determine the compliance of the DPA text with the GDPR, identifying areas of compliance and non-compliance in a comprehensive report (Fig. 2, Step ③).

V. CURRENT EMPIRICAL RESULTS

We conducted (1) a qualitative study that characterizes food-safety concepts in regulatory provisions impacting modern software-intensive food-safety systems, and (2) an LLM-based approach that classifies the provisions of food-safety regulations based on their relevance to systems and software requirements. We conducted an extensive evaluation of our approach by instantiating it with both BERT and GPT-3.5.

We evaluated our models using standard classification metrics, including *Accuracy*, *Precision*, *Recall*, and *F-score*, to classify legal provisions effectively. This evaluation facilitated a comprehensive comparison of LLM performances, identifying strengths and highlighting areas for enhancement. To account for variability, we conducted multiple experiments and

reported the average results. Boxplots were generated to visually represent the performance outcomes of these experiments. Our results indicate that accuracy is largely consistent across BERT (*Precision* of 87%, *Recall* of 86%, and *F-score* of 87%) and GPT-3.5 (*Precision* of 89%, *Recall* of 83%, and *F-score* of 86%), with BERT achieving slightly better overall results.

For our compliance checking of DPAs with GDPR, our early experiments with *gpt-3.5-turbo-0125* [42], *Mixtral-8x7B-Instruct-v0.1* [43], and *gpt-4-0125-preview* [42] show promising improvements when transitioning from sentence-level to paragraph-level passages. On average, the improvements in *Accuracy* obtained through using these three models are in the ranges of 33% (from 30% and 63%), 35% (from 33% and 69%), and 40% (from 41% and 81%), respectively.

VI. FUTURE RESEARCH

My research aims to make a case for the need to reconsider current practices in legal compliance automation in light of recent advances in AI. Specifically, I argue that the enhanced capacity of modern LLMs to handle context is likely to induce a major shift in our treatment of textual legal artifacts. This shift will involve transitioning from analyzing smaller contexts, such as individual sentences and phrases, to considering larger volumes of content, such as paragraphs and beyond, as context. I posit that the larger context will be able to provide the prerequisite knowledge, including cross-referenced legal materials, to create a self-contained basis for accurate automated decision-making regarding compliance.

In the domain of food-safety regulations, my focus will extend to exploring concepts beyond North America, with a keen interest in the regulations of countries like India and the UK, which, while operating under English law, have distinct regulations.

In the domain of DPAs, my future work will focus on four main aspects: (1) enriching the DPA dataset [1] with paragraph-level annotations; (2) conducting comprehensive empirical evaluations to validate the effectiveness of paragraph-level context in increasing LLM accuracy; (3) benchmarking against prior BERT-based approaches, e.g., [4], to showcase benefits; and (4) involving legal experts to critically assess LLM outputs, especially concerning their explanations and justifications. A pertinent question arises on evaluating legal experts’ verification of these justifications.

I foresee dedicating at least eight months to these mentioned activities, paralleling these efforts with the composition of a research paper for submission to RE 2025.

In addition, and time permitting, I plan to examine techniques like Retrieval Augmented Generation [44] to facilitate question answering over domains such as DPAs and Privacy Policies. A preparatory period of four months will be reserved for this initiative, post-investigation and prior to full-scale implementation. As I progress, I trust that the constructive insights garnered from our ongoing research, coupled with the sufficient time frame, will empower me to not only expand upon my thesis but also to finalize it punctually by the culmination of my fourth academic year.

REFERENCES

- [1] O. Amaral, S. Abualhaija, and L. Briand, "ML-based compliance verification of data processing agreements against GDPR," in *Proceedings of 31st IEEE International Requirements Engineering Conference (RE)*, 2023.
- [2] O. Amaral, S. Abualhaija, D. Torre, M. Sabetzadeh, and L. C. Briand, "AI-enabled automation for completeness checking of privacy policies," *IEEE Transactions on Software Engineering*, vol. 48, no. 11, 2021.
- [3] O. Amaral, M. I. Azeem, S. Abualhaija, and L. C. Briand, "NLP-based automated compliance checking of data processing agreements against GDPR," *IEEE Transactions on Software Engineering*, 2023.
- [4] M. Ilyas Azeem and S. Abualhaija, "A multi-solution study on GDPR AI-enabled completeness checking of DPAs," *arXiv:2311.13881*, 2023.
- [5] N. Sannier, M. Adedjouma, M. Sabetzadeh, and L. Briand, "An automated framework for detection and resolution of cross references in legal texts," *Requirements Engineering*, vol. 22, 2017.
- [6] O. Amaral, S. Abualhaija, M. Sabetzadeh, and L. Briand, "A Model-based conceptualization of requirements for compliance checking of data processing against GDPR," in *Proceedings of 29th International Requirements Engineering Conference Workshops (REW)*, 2021.
- [7] A. Xiang, W. Pei, and C. Yue, "PolicyChecker: Analyzing the GDPR completeness of mobile apps' privacy policies," in *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, 2023.
- [8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of 17th American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2019.
- [9] T. D. Breaux, M. W. Vail, and A. I. Antón, "Towards regulatory compliance: Extracting rights and obligations to align requirements with regulations," in *Proceedings of 14th IEEE International Requirements Engineering Conference (RE)*, 2006.
- [10] T. Breaux and A. Antón, "Analyzing regulatory rules for privacy and security requirements," *IEEE Transactions on Software Engineering*, vol. 34, no. 1, 2008.
- [11] T. D. Breaux and D. G. Gordon, "Preserving traceability and encoding meaning in legal requirements extraction," in *Proceedings of 6th International Workshop on Requirements Engineering and Law (RELAW)*, 2013.
- [12] J. Bhatia and T. D. Breaux, "Semantic incompleteness in privacy policy goals," in *Proceedings of 26th IEEE International Requirements Engineering Conference (RE)*, 2018.
- [13] M. Fan, L. Yu, S. Chen, H. Zhou, X. Luo, S. Li, Y. Liu, J. Liu, and T. Liu, "An empirical evaluation of GDPR compliance violations in Android mHealth apps," in *Proceedings of 31st IEEE International Symposium on Software Reliability Engineering (ISSRE)*, 2020.
- [14] R. E. Hamdani, M. Mustapha, D. R. Amariles, A. Troussel, S. Meeüs, and K. Krasnashchok, "A combined rule-based and machine learning approach for automated GDPR compliance checking," in *Proceedings of 18th International Conference on Artificial Intelligence and Law (ICAIL)*, 2021.
- [15] F. Xie, Y. Zhang, C. Yan, S. Li, L. Bu, K. Chen, Z. Huang, and G. Bai, "Scrutinizing privacy policy compliance of virtual personal assistant apps," in *Proceedings of 37th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 2022.
- [16] N. Zeni, N. Kiyavitskaya, L. Mich, J. R. Cordy, and J. Mylopoulos, "GaiusT: supporting the extraction of rights and obligations for regulatory compliance," *Requirements Engineering*, vol. 20, 2015.
- [17] N. Zeni, E. A. Seid, P. Engiel, S. Ingolfo, and J. Mylopoulos, "Building large models of law with NómósT," in *Proceedings of 35th International Conference on Conceptual Modeling (ER)*, 2016.
- [18] A. Sleimi, N. Sannier, M. Sabetzadeh, L. Briand, and J. Dann, "Automated extraction of semantic legal metadata using natural language processing," in *Proceedings of 26th IEEE International Requirements Engineering Conference (RE)*, 2018.
- [19] S. Abualhaija, C. Arora, A. Sleimi, and L. C. Briand, "Automated question answering for improved understanding of compliance requirements: A multi-document study," in *Proceedings of 30th IEEE International Requirements Engineering Conference (RE)*, 2022.
- [20] T. Hey, J. Keim, A. Koziolk, and W. F. Tichy, "NoRBERT: Transfer learning for requirements classification," in *Proceedings of 28th IEEE International Requirements Engineering Conference (RE)*, 2020.
- [21] R. Chatterjee, A. Ahmed, P. R. Anish, B. Suman, P. Lawhatre, and S. Ghaisas, "A pipeline for automating labeling to prediction in classification of NFRs," in *Proceedings of 29th IEEE International Requirements Engineering Conference (RE)*, 2021.
- [22] W. Alhoshan, L. Zhao, A. Ferrari, and K. J. Letsholo, "A zero-shot learning approach to classifying requirements: A preliminary study," in *Proceeding of 28th International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ)*. Springer, 2022.
- [23] A. Sainani, P. R. Anish, V. Joshi, and S. Ghaisas, "Extracting and classifying requirements from software engineering contracts," in *Proceedings of 28th IEEE International Requirements Engineering Conference (RE)*, 2020.
- [24] C. Jain, P. R. Anish, A. Singh, and S. Ghaisas, "A transformer-based approach for abstractive summarization of requirements from obligations in software engineering contracts," in *Proceedings of 31st IEEE International Requirements Engineering Conference (RE)*, 2023.
- [25] M. K. Habib, S. Wagner, and D. Graziotin, "Detecting requirements smells with deep learning: Experiences, challenges and future work," in *Proceedings of 29th IEEE International Requirements Engineering Conference Workshops (REW)*. IEEE, 2021.
- [26] V. Varenov and A. Gabdrachmanov, "Security requirements classification into groups using NLP transformers," in *Proceedings of 29th IEEE International Requirements Engineering Conference Workshops (REW)*. IEEE, 2021.
- [27] R. R. Mekala, A. Irfan, E. C. Groen, A. Porter, and M. Lindvall, "Classifying user requirements from online feedback in small dataset environments using deep learning," in *Proceedings of 29th IEEE International Requirements Engineering Conference (RE)*. IEEE, 2021.
- [28] G. Deshpande, B. Sheikhi, S. Chakka, D. L. Zotegouon, M. N. Masahati, and G. Ruhe, "Is BERT the new silver bullet? An empirical investigation of requirements dependency classification," in *Proceedings of 29th IEEE International Requirements Engineering Conference Workshops (REW)*. IEEE, 2021.
- [29] J. Fischbach, J. Frattini, A. Spaans, M. Kummeth, A. Vogelsang, D. Mendez, and M. Unterkalmsteiner, "Automatic detection of causality in requirement artifacts: The CiRa approach," in *Proceedings of 27th International Working Conference on Requirements Engineering: Foundation for software quality (REFSQ)*. Springer, 2021.
- [30] Y. Wang, L. Shi, M. Li, Q. Wang, and Y. Yang, "A deep context-wise method for coreference detection in natural language requirements," in *Proceedings of 28th IEEE International Requirements Engineering Conference (RE)*, 2020.
- [31] —, "Detecting coreferent entities in natural language requirements," *Requirements Engineering*, vol. 27, no. 3, 2022.
- [32] A. Nayak, H. P. Timmapathini, V. Murali, K. Ponnalagu, V. G. Venkoparao, and A. Post, "Req2Spec: Transforming software requirements into formal specifications using natural language processing," in *Proceedings of 28th International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ)*, 2022.
- [33] C. M. Luders, T. Pietz, and W. Maalej, "Automated detection of typed links in issue trackers," in *Proceedings of 30th IEEE International Requirements Engineering Conference (RE)*, 2022.
- [34] Ş. Mehder and F. B. Aydemir, "Classification of issue discussions in open source projects using deep language models," in *2022 IEEE 30th International Requirements Engineering Conference Workshops (REW)*, 2022.
- [35] M. Abbas, A. Ferrari, A. Shatnawi, E. Enoiu, M. Saadatmand, and D. Sundmark, "On the relationship between similar requirements and similar software," *Requirements Engineering*, vol. 28, no. 1, 2023.
- [36] D. Luitel, S. Hassani, and M. Sabetzadeh, "Using language models for enhancing the completeness of natural-language requirements," in *Proceedings of 29th International Working Conference on Requirement Engineering: Foundation for Software Quality (REFSQ)*. Springer, 2023.
- [37] B. Görner and F. B. Aydemir, "Generating requirements elicitation interview scripts with large language models," in *Proceedings of the 31st IEEE International Requirements Engineering Conference Workshops (REW)*, 2023.
- [38] Y. Bouzembrak, M. Klüche, A. Gavai, and H. J. Marvin, "Internet of Things in food safety: Literature review and a bibliometric analysis," *Trends in Food Science & Technology*, vol. 94, 2019.
- [39] D. Torre, S. Abualhaija, M. Sabetzadeh, L. Briand, K. Baetens, P. Goes, and S. Forastier, "An AI-assisted approach for checking the completeness of privacy policies against GDPR," in *Proceedings of 28th IEEE International Requirements Engineering Conference (RE)*, 2022.
- [40] OpenAI, "OpenAIChatData," <https://bit.ly/4bBRYGw>.
- [41] Huggingface, "Chatcompletion," <https://bit.ly/48cuHrV>.
- [42] OpenAI, "OpenAIModels," <https://bit.ly/495WdZk>.
- [43] Huggingface, "Mixtral," 2023, <https://bit.ly/42DgQtI>.
- [44] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, and H. Wang, "Retrieval-augmented generation for large language models: A survey," *arXiv:2312.10997*, 2023.