

What Are the Facts? Automated Extraction of Court-Established Facts from Criminal-Court Opinions

Klára Bendová², Tomáš Knap¹, Jan Černý^{1,2}, Vojtěch Pour¹, Jaromir Savelka³,
Ivana Kvapilíková² and Jakub Drápal¹

¹Faculty of Law, Charles University, Prague, Czechia

²Faculty of Mathematics and Physics, Charles University, Prague, Czechia

³School of Computer Science, Carnegie Mellon University, Pittsburgh PA, USA

Abstract

Criminal justice administrative data contain only a limited amount of information about the committed offense. There is, however, an unused source of extensive information in continental European courts' decisions: Descriptions of criminal behaviors in verdicts by which offenders are found guilty. In this paper, we study the feasibility of extraction of these descriptions from publicly available court decisions from Slovakia. We use two different approaches for retrieval: regular expressions and large language models (LLM). Our baseline was a simple method employing regular expressions to identify typical words occurring before the beginning of the description and after. The advanced regular expressions approach further focused on letter-spacing and its normalization (insertion of spaces between individual letters), typical for delineation of the description. The LLM approach involved prompting the Gemini Flash 2.0 model to extract the descriptions using a predefined set of instructions. Although the baseline identified descriptions in only 40.5% of the verdicts of our test set, both methods significantly outperformed it, achieving 97% with advanced regular expressions and 98.75% with LLM and a combination of both 99.5%. Evaluation by law students showed that both advanced methods matched human annotations in about 90% of cases, compared to just 34.5% for the baseline. LLMs fully matched human-labeled descriptions in 91.75% of instances and a combination of advanced regular expression with LLM brings 92% match.

Keywords

argument mining, court decisions, criminal behavior, NLP, LLM

1. Introduction

Empirical quantitative legal research is often hampered by insufficient detail in data, particularly in criminal law, where information about criminal behavior is frequently incomplete or lacking. Administrative datasets typically record only the relevant section of the penal code, offering a general definition of the offense. However, they usually provide little information beyond this classification, making it difficult for researchers to discern the specific behaviors involved and to understand how variations in behavior shape state responses.

More detailed information about criminal behavior is recorded in textual form. Criminal verdicts in most continental European countries contain a description of the criminal behavior: An authoritative description of what behavior an offender is found guilty of and for which a sentence is imposed. These descriptions thus provide crucial and otherwise unavailable pieces of information about behavior. Criminal verdicts are available online in several European and non-European countries (e. g., Slovakia, Estonia, Moldova and China), while in others they are available to researchers (e. g., Finland).

In this paper, we show the difficulties of extracting these descriptions of criminal behaviors from court verdicts. After discussing the related work and the data we employed, we present two different extraction methods and their specifics. Then, we present the outcomes of individual methods (and their combinations) and their reliability. We conclude by presenting how we worked with difficult cases and the future use of LLMs in perfecting this task.

Proceedings of the Seventh International Workshop on Automated Semantic Analysis of Information in Legal Text (ASAIL 2025), 16 June, 2025, Chicago, USA.

*Corresponding author: Tomáš Knap, Faculty of Law, Charles University, Prague, Czechia. Email: tomas.knap@prf.cuni.cz



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Related Work

Extracting various semantic and/or functional elements from court opinions has been established as a key task in legal text processing. This is because these loosely structured and sometimes noisy documents contain enormous amounts of useful knowledge that can potentially be utilized in many different applications. Prior research can be distinguished into two categories. First, the task may be defined as labeling small textual units, often sentences, according to some predefined type system (e. g., rhetorical roles such as evidence, reasoning, and conclusion). This approach has been applied to administrative decisions from the U.S. [1, 2], multi-domain court decisions from India [3], international arbitration decisions [4], multi-{domain,country} adjudicatory decisions [5], or opinions of the European Court of Human Rights [6]. Identifying a section that states an outcome of the case has also received considerable attention separately [7, 8, 9, 10]. In this work, we focus on detecting one specific sentence in each opinion—the one that authoritatively states the facts of the case.

Alternatively, the task could be to segment the text into a small number of contiguous parts, typically comprising multiple paragraphs. Different variations of this task were applied to several legal domains from countries such as Canada [11], the Czech Republic [12], France [13], and the U.S. [14]. One approach to segmentation has focused on automatically identifying the rhetorical roles of sentences [11, 15, 16]. In [11] the authors employed linguistic markers to segment Canadian decisions into four units: Introduction, Context, Juridical Analysis, and Conclusion. A similar scheme was proposed in [12], including some additional types such as Dissent, Footnotes, or Party Claims. In [17], the authors identify typical language structures that are used in various types of Premises or Conclusions. These are then expressed in the form of a Context Free Grammar for parsing legal arguments. In [18, 19], conditional random fields (CRF) were applied to segment legal documents into seven labeled components, with each label representing a corresponding rhetorical role.

3. Data

Slovak courts' verdicts are, in many respects, typical of the continental Germanic legal culture. They are clearly divided into two parts: the dispositive part, which presents the decisions that were taken, and the reasoning. The dispositive part (i) identifies the case, the court, the verdict, and the offender; (ii) announces whether the offender was found guilty or not; (iii) describes the criminal behavior for which the offender was found guilty (or innocent); (iv) subsumes the behavior within a legal definition of an offense; and (v) pronounces the consequences (sentences and reparation of damages). An example of this part is in Appendix B. Some verdicts further contain reasoning—an explanation of why the offender was found guilty, why a specific sentence was imposed, and a description of the proceedings. Reasoning is often infrequent and of low quality [20].

Slovakia allows anyone to download .json files with all court decisions and to use an API to search for specific decisions. While such convenient accessibility is very beneficial, researchers should be concerned about possible flaws in the published data, especially missing data and the possible resulting limited representativeness. If an administrative dataset containing secondary data about court decisions made in a country is available, it helps researchers identify both how many court verdicts are missing and whether there is a pattern among the missing values. Slovakia has an administrative dataset of high quality containing secondary data describing all criminal court verdicts from 2018 to 2022 (hereafter referred to as the "administrative dataset") [21]. There were 126,795 verdicts decided during this period, according to this dataset.

We first downloaded all court verdicts and linked the verdicts with the administrative dataset using a unique court docket number and court name. We were able to link 77.64% of cases. Then we employed the court docket numbers from the administrative dataset to retrieve the missing verdicts via the API, which allowed us to link an additional 12.42% of cases. Overall, we were unable to link 9.94% of cases that should exist; yet, they are (i) not present in the .json or API, (ii) present in .json but matching fails due to a corrupt .json file, or (iii) present in the API but matching fails due to repeated requests and

website limitations. Successful matching, to some extent, depended on which court decided the case. While 40 out of 54 district courts had a success rate higher than 90%, other courts had a success rate above 60%, with the exception of one that had a success rate of 12%. These differences present the need to work with administrative datasets to ensure that full-text court verdicts are representative and to determine their limitations. This provides us with 112,864 court verdicts, within which we attempted to identify descriptions of criminal behavior.

Table 1

Coverage of Slovak criminal-court verdicts (2018–2022)

Download / linkage step	n cases	% of 126 795
JSON dump (direct match)	98,500	77.64 %
API re-download (matched via docket number)	15,742	12.42 %
<i>Linked total</i>	112 864	88.06 %
Not linked (missing / corrupt / API errors)	13 931	11.0 %

We chose two test sets of verdicts to annotate the data. The first general stratified sample contains 400 judgments, in which we controlled for years and the representation of different courts during sampling. These were then annotated by trained law students (2 groups of 200 statements), with each statement being annotated by two persons (an agreement of 97% in both groups; disagreements were resolved by one of the authors). This dataset is intended to evaluate the overall performance of the individual methods. Since we expected high performance on the general dataset, we also created an additional set of 200 judgments to evaluate more challenging cases. This supplementary set includes a mix of judgments in which the letter-spacing extraction yielded only a single candidate expression—indicating a higher likelihood that the judgment lacks a factual sentence—and judgments in which the rule-based extraction method failed to identify any relevant sentence at all. The data were annotated under the same conditions (95.5% agreement; disagreements were resolved by one of the authors).

Table 2

Annotated evaluation sets

Set	n cases	Sampling	Agree.	Purpose
General stratified	400	year × court strata	97%	overall performance
Challenging cases	200	letter-spacing = 1 ∨ regex fail	95.5%	stress-test edge cases

4. Experimental Design

4.1. Structure of Court Decision

Our analysis leveraged two fundamental formal features observable across all judgments: A consistent structure (described above) and the systematic use of a specific typographic convention throughout the dataset, specifically letter-spacing. The letters of a word are spaced out (e. g., L I K E T H I S). Letter-spacing is generally applied to single words or short phrases and is typically placed between paragraphs. Its function is to introduce a new section of the judgment. As a result, letter-spacing tends to appear with a limited set of expressions used across decisions, as we can see in Appendix B. A key advantage of letter-spacing is its robustness during format conversion: even when judgments are somewhat messily transformed from PDF to .json, letter-spacing is usually preserved—unlike paragraph-ending characters, which often suffer from noisy encoding. However, letter-spacing has a domain-specific nature; while common in legal documents, it is rarely encountered in other textual domains.

4.2. Baseline

As we describe in Section 4.1, both the overall structure and the letter-spacing expression between paragraphs appear to be consistent at first glance. This observation motivated us to extract factual

statements using regular expressions due to their simplicity, accuracy, and ease of use. In the baseline approach, we chose not to introduce any complexity into the regex patterns. We manually listed starting and ending phrases as fixed patterns (examples of phrases are in Table 3) and used them for extraction. The baseline achieves a fact sentence extraction success rate of 31.69%.

Table 3
Typical starting and ending expressions

Starting Expression	Ending Expression
is found guilty that	therefore/thus
is found guilty	
they are guilty that	
is acknowledged as guilty that	
is acknowledged guilty that	

4.3. Advanced Regular Expressions

Building on observations of simple regular expressions, we developed more flexible patterns to eliminate variations in the input text. These flexible patterns focus on preventing unwanted irregularities—such as extra spaces, unexpected line breaks, and other conversion issues. Adding optional whitespace matching between characters in key phrases significantly improved resiliency to formatting inconsistencies.¹ This improved both the success rate and the quality of the extracted fact sentences, as shown in Table 5.

The regular expressions themselves were automatically extracted directly from the court verdicts. We focused only on preserving the order of the letter-spaced expressions from the rulings in the document. We then grouped expressions by their relative position and annotated those that function as openers or closers of factual sentences. This yielded a set of starting ($N = 40$) and ending ($N = 2$) points to identify fact sentences.

Table 4
Comparison of simple regex and automatic approach extraction

Method	Extracted	Extracted (%)	Total
Simple Regex	35,767	31.69 %	112,864
Automatic approach	109,292	96.80 %	112,864

4.4. LLMs: Standalone and Building on Previous Methods

During the initial phases of our research, relying solely on a rule-based approach for extracting factual statements proved insufficient. We have made a natural progression to using more sophisticated approaches. The main challenge with defining a narrow scope of rules, especially in a language where the individual words are typically more varied, is that one can hardly create a complete list of all the possible words and their combinations. LLMs are an ideal solution for extending a narrow list of rules to a dataset that presents uncertainties about the exact wording in each individual case.

Utilizing LLMs involves two main aspects: selecting an appropriate model and providing it with the necessary context, a process known as prompting. Our model selection was guided by two primary criteria: its established suitability for relevant applications such as parsing and extracting from large texts, and overall cost. Based on this evaluation, Gemini Flash 2.0² was identified as the most suitable option. Our temperature was 0.0, which helps reduce hallucinations in extracting factual sentences from judgments and supports better replicability of our research. Our initial experiments revealed that generic descriptions of factual statements were insufficient for reliable extraction using LLMs. Achieving success requires a carefully constructed prompt specifically designed for this task. We developed a specialized prompt that combines concrete examples of factual statements, a clear definition of the

¹<https://github.com/vanickovak/Factual-sentence-extraction.git>

²<https://blog.google/technology/google-deepmind/google-gemini-ai-update-december-2024>

expected output, and, importantly, explicit indicators of where these statements typically begin and end (Appendix A).

The critical breakthrough in our approach occurred when we followed the phrases used by previous methods. We incorporated specific textual markers commonly found at the beginning or end of factual statements, as shown in Table 3. This structural guidance transformed the task from pure content understanding to a more focused pattern recognition and extraction challenge. The model was directed to identify text between these markers and extract only the factual components while omitting legal evaluations. The results were returned in .json format for further processing.

A single Slovak judgment contains on average 4,083 characters (approx. 1,020 tokens), to which we prepend a 10,497-character prompt (approx. 2,624 tokens). The Gemini Flash 2.0 output averages 1 257 characters (approx. 314 tokens). At Google’s July 2025 pricing (USD 0.10 / M input tokens, 0.40 / M output tokens), this corresponds to 0.00049 USD per judgment (approx. 0.49 USD for 1,000 decisions).

Despite recent improvements in LLM accuracy, we observed occasional instances of text hallucination, particularly with special characters or uncommon legal terminology. To address this issue, we implemented a post-processing step using a function that aligns the model’s output with the original text. This verification method effectively eliminated hallucinations, ensuring fidelity to the source documents.

4.5. Combination Approach

Our experimental design includes a combined methodology that integrates the strengths of both advanced regular expressions and LLM approaches. This hybrid pipeline operates sequentially:

1) Advanced regular expressions first attempt to extract factual statements from court decisions. 2) For cases where the regular expression approach fails to identify any factual sentences, we apply the LLM-based extraction.

This combination strategy maximizes efficiency while addressing the limitations of each individual method. The regular expressions provide fast processing for standard document structures, while the LLM handles more complex linguistic variations and non-standard document formats.

5. Results

To empirically assess our approach, we tested the ability of the method to extract specific factual sentences against a set of 400 reference statements.

First, we assess the method’s ability to identify a factual statement in the court decision. While the baseline identified a description of criminal behavior in only 40.5% of verdicts, both advanced regular expressions and LLMs performed significantly better, achieving success rates of 97% and 98.75%, respectively. If we combine both approaches, specifically by employing LLMs only on cases where advanced regular expressions fail to extract any factual sentences, we achieve an improved overall accuracy of 99.5%. The reason the LLM did not achieve the extraction of 100% factual statements is that five cases could not be processed due to exceeding the model token limits or triggering content safety flags within the API, likely related to sensitive topics such as the narcotics mentioned in the cases. Feeding the texts to the model in pieces and employing prompt engineering to bypass the safety mechanism would likely further increase performance.

Table 5
Method performance on test data

Method	Extraction Rate	Quality $\geq 95\%$
Baseline	40.5%	34.5%
Advanced Regex	97.0%	89.5%
LLM	98.75%	91.75%
Combined	99.5%	92.0%

We further focused on the quality of the extracted sentences. Extraction quality was measured at the character level, ignoring diacritics. The results are displayed in Table 5 by the quality of the match. When evaluating based on exact match, the performance of the LLM is undeniable. However, if we allow for minor character-level variations of up to 5%, the results of advanced regex and LLMs are surprisingly comparable—89.5% for the advanced regex and 91.75% for the LLM. In contrast, the baseline approach achieved this level of approximate matching in only 34.5% of cases. By combining both approaches—specifically applying the LLM only to cases where advanced regex fails to extract accurate factual sentences—we achieve 92% accuracy on the test data.

Table 6
Detailed extraction quality breakdown

Method	Perfect 100%	Good 80-95%	Failed <50%
Baseline	2.0%	0.8%	4.5%
Adv. Regex	9.0%	2.8%	2.8%
LLM	91.8%	5.3%	1.3%
Combined	11.5%	2.8%	2.8%

We then focused on problematic cases, i.e., verdicts where no factual statement was extracted or where the extracted statement shared less than 50% character overlap with the manually annotated version. Such cases typically result from non-standard typographic formatting of the verdict or the use of less frequent terms in letter-spaced expressions. In some instances, the extraction failed because the factual statement was entirely missing from the verdict, which is usually due to their procedural nature (e. g., acquittal judgments). We prepared a dataset of 200 such verdicts exhibiting these deviations and had it annotated by two annotators under the same conditions as the main test dataset. The same LLM prompt was then applied to this data.

Table 7
LLM performance on difficult cases ($N = 200$)

Outcome	Share
Perfect match (100%)	81.5 %
High quality (95–100%)	2.5 %
Good quality (80–95%)	3.5 %
Fair quality (50–80%)	0.5 %
Failed extraction	12.0 %
Total high quality ($\geq 95\%$)	84.0 %

As shown in Table 7, the LLM was able to perform high-quality factual statement extraction in 84% of cases within this challenging dataset, and it failed to identify a factual statement in 12% of cases, resulting in 0% similarity. Investigation confirmed that, in nearly all these instances, the LLM correctly followed its strict prompt instructions. The failures occurred because the factual sentences in this challenging subset used grammatical structures or formats not anticipated by the specific rules in the prompt. The LLM, therefore, acted as instructed by reporting no match rather than misinterpreting the content. This highlights the need to update the prompt, generalizing the description of a factual sentence based on the wider variety of formats observed.

To analyze hallucinations, we compared 100 LLM-generated factual statements with the corresponding spans in the source judgments. Character-level similarity averaged 0.99; only one case (1 %) fell below 90 %.

To avoid leaking hallucinated text into downstream data, we post-process every model output: we locate the predicted span in the source document and replace the generated string with that exact substring. The factual sentences stored in our dataset are therefore *verbatim* excerpts from the original judgments; hallucinations appear only in cases where no sufficiently similar span can be aligned.

Table 8
Hallucination breakdown ($N = 100$)

Similarity band	Cases	Share
100 %	26	26.0 %
95–99 %	71	71.0 %
90–94 %	2	2.0 %
80–89 %	1	1.0 %
<80 %	0	0.0 %

6. Discussion: Integrating all Methods

We did not train a supervised model due to the high cost of large-scale annotations. Our strategy was to first explore few-shot prompting of an LLM, measure its performance against manually annotated data, and consider supervised training only if the LLM fell short. A small experiment with training a model on the annotated test set could be interesting; however, it would leave us without a clean held-out set for evaluation, making it difficult to fairly compare the approaches.

Our goal was to develop a scalable solution that could be applied across a wide range of court decisions without the need for extensive labeled training data. To that end, we combined regular expressions with few-shot prompting of an LLM. This approach allows us to inject a small number of annotated examples directly into the prompt, guiding the model’s output in a flexible and easily adjustable way. By avoiding model training and relying instead on prompt-level supervision and rule-based heuristics, we were able to build a system that is applicable in low-resource legal settings and avoids the high cost of large-scale manual annotations.

The LLM-based approach introduces computational demands that must be balanced against extraction quality. While more resource-intensive than simpler methods, the accuracy benefits justify their application, particularly for complex or non-standard documents. For large-scale processing of thousands of verdicts, we recommend a staged approach in which computationally expensive LLM processing is reserved for documents where simpler methods yield low-confidence results.

It is worth noting that a recurring pattern observed in cases with imperfect matches was the LLMs’ occasional difficulty in precisely delineating the boundaries of the target factual statement within the broader text. Specifically, mismatches often arose not from extracting incorrect information; rather, they stemmed from the model including extraneous text immediately following the intended end of the factual statement or, in some other cases, the LLM finished the factual sentence before it was instructed to do so.

These detailed results affirm the LLMs’ precision, particularly their high success rate, but also highlight challenges with edge cases and computational intensity, which require careful consideration for scalability. Future enhancements include employing models with larger context windows to handle longer texts without truncation and refining prompts for greater task specificity to further improve accuracy and relevance, particularly for the less successful cases.

Finally, transferring the methods to other jurisdictions also remains a challenge. Each new legal system would require its own regex inventory, reflecting local idioms and citation habits. The LLM approach proved more robust in our pilots, but our sample is too narrow to draw general conclusions about out-of-domain performance without retraining and expert validation. We therefore refrain from reporting cross-jurisdiction metrics and leave systematic evaluation on foreign corpora for future work.

7. Conclusion

In this paper, we presented a method for collecting published court decisions and extracting factual sentences—coherent descriptions of criminal conduct—with strong potential for further research. In the extraction process, we identified the consistent structure of court decisions and a legal typographic convention known as “letter-spacing” as key features. We leveraged these observations in three extraction approaches: an automated search for the start and end-markers of factual sentences (baseline),

an advanced regular expression-based script, and, surprisingly effectively, extraction using a LLM. The baseline approach lacked the complexity to successfully extract factual sentences from verdicts. In contrast, the advanced regular expressions identified descriptions in 97% of verdicts and Gemini Flash 2.0 extracted them in 98.75% of the test data. The combination of both methods extracted descriptions in 99.5% of cases. Manual annotation revealed that 91.75% of descriptions retrieved by the LLM and 89.5% of those retrieved by regular expressions match the descriptions identified by human annotators. The combination of advanced regular expressions and LLM achieved 92% accuracy. In future work, we aim to expand the dataset with court decisions from additional countries, providing empirical data for comparative research in criminal law.

Acknowledgments

This study was funded by the Czech Grant Foundation (grant number 25-16848M entitled "Just Sentences: Analyzing and Enhancing Proportionality and Consistency Using Typical Crimes"). The authors have no competing interests.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

- [1] V. R. Walker, K. Pillaipakkamnatt, A. M. Davidson, M. Linares, D. J. Pesce, Automatic classification of rhetorical roles for sentences: Comparing rule-based scripts with machine learning., in: *ASAIL@ ICAIL 2019*, 2019.
- [2] L. Zhong, Z. Zhong, Z. Zhao, S. Wang, K. D. Ashley, M. Grabmair, Automatic summarization of legal decisions using iterative masking of predictive sentences, in: *ICAIL 2019*, 2019, pp. 163–172.
- [3] P. Bhattacharya, S. Paul, K. Ghosh, S. Ghosh, A. Wyner, Identification of rhetorical roles of sentences in indian legal judgments, in: *JURIX 2019*, volume 322, IOS Press, 2019, p. 3.
- [4] K. Branting, B. Weiss, B. Brown, C. Pfeifer, A. Chakraborty, L. Ferro, M. Pfaff, A. Yeh, Semi-supervised methods for explainable legal prediction, in: *ICAIL 2019*, 2019, pp. 22–31.
- [5] J. Šavelka, H. Westermann, K. Benyekhlef, Cross-domain generalization and knowledge transfer in transformers trained on legal data, in: *ASAIL@ JURIX 2020*, 2020.
- [6] I. Habernal, D. Faber, N. Recchia, S. Bretthauer, I. Gurevych, I. Spiecker genannt Döhmann, C. Burchard, Mining legal arguments in court decisions, *Artificial Intelligence and Law* 32 (2024) 1–38.
- [7] H. Xu, J. Šavelka, K. D. Ashley, Using argument mining for legal text summarization, in: *JURIX 2020*, volume 334, IOS Press, 2020.
- [8] A. Petrova, J. Armour, T. Lukasiewicz, Extracting outcomes from appellate decisions in us state courts, in: *JURIX 2020*, 2020, p. 133.
- [9] M. Grabmair, K. D. Ashley, R. Chen, P. Sureshkumar, C. Wang, E. Nyberg, V. R. Walker, Introducing luima: an experiment in legal conceptual retrieval of vaccine injury decisions using a luima type system and tools, in: *Proceedings of the 15th international conference on artificial intelligence and law*, 2015, pp. 69–78.
- [10] A. Bansal, Z. Bu, B. Mishra, S. Wang, K. Ashley, M. Grabmair, Document ranking with citation information and oversampling sentence classification in the luima framework, in: *Legal Knowledge and Information Systems*, IOS Press, 2016, pp. 33–42.
- [11] A. Farzindar, G. Lapalme, Letsum, an automatic text summarization system in law field, in: *Proceedings of JURIX 2004*, 2004.
- [12] J. Harašta, J. Šavelka, F. Kasl, J. Míšek, et al., Automatic segmentation of czech court decisions into multi-paragraph parts, *Jusletter IT* (2019).

- [13] P. Boniol, G. Panagopoulos, C. Xypolopoulos, R. El Hamdani, D. R. Amariles, M. Vazirgiannis, Performance in the courtroom: Automated processing and visualization of appeal court decisions in france, in: Proceedings of the Natural Legal Language Processing Workshop 2020, 2020.
- [14] J. Savelka, K. D. Ashley, Segmenting us court decisions into functional and issue specific parts., in: JURIX 2018, 2018, pp. 111–120.
- [15] B. Hachey, C. Grover, Extractive summarisation of legal texts, *Artificial Intelligence and Law* 14 (2006) 305–345.
- [16] M.-F. Moens, Summarizing court decisions, *Information processing & management* 43 (2007) 1748–1764.
- [17] A. Wyner, R. Mochales-Palau, M.-F. Moens, D. Milward, Approaches to text mining arguments from legal cases, Springer, 2010.
- [18] M. Saravanan, B. Ravindran, S. Raman, Improving legal document summarization using graphical models, *Frontiers in Artificial Intelligence and Applications* 152 (2006) 51.
- [19] M. Saravanan, B. Ravindran, Identification of rhetorical roles for segmentation and summarization of a legal judgment, *Artificial Intelligence and Law* 18 (2010) 45–76.
- [20] J. Drápal, A. Bolocan-Holban, J. Ginter, M. Plesničar, K. Tomšů, M. Vidaicu, Sentence justification in first-level courts in post-communist europe, in: *Justifying Punishment*, Routledge, 2024, pp. 39–62.
- [21] M. Terkanič, D. Gálisová, Revízia trestných štatistických listov a výkazov na slovensku: Desať vecí, ktoré by sme najradšej spravili inak, *Česká kriminologie* 6 (2021) 1–11.

Appendix A: Prompt

Task: Extracting factual statements from text

Your task is to extract the main factual statement ONLY from the provided text.

A factual statement in a criminal judgment is that part of the judgment which precisely and specifically describes the act for which the accused is convicted. It is a detailed description of what happened - when, where, how, and with what consequence the criminal offense was committed.

The factual statement must contain all the legal elements of the criminal offense for which the accused is convicted, both the objective aspect (action, consequence, causal link) and the subjective aspect (intent or negligence).

It is an important part of the operative part of the judgment and must be formulated in such a way that it is completely clear for which specific act the accused is convicted, and that this act is unmistakable from any other. This is key for the principle of "ne bis in idem" (not twice in the same matter), which prevents someone from being tried twice for the same act.

How else can we recognize the factual statement?

It is located after the phrase "sa uznáva vinnou, že..." (is found guilty that...) - which is a typical introduction to the factual statement in Slovak criminal judgments.

It contains a detailed and specific description of the committed act:

Who: "The accused P. Z. ... as a parent obliged to continuously care for the upbringing..."

What: "inconsistently approached the fulfillment of her duties... and created conditions for minors... for the emergence of undesirable habits in the form of long-term truancy"

When: "in the school year 2017/2018... in the period from 01. 09. 2017 to 30. 06. 2018"

How: detailed description of the conduct including specific numbers (93 and 194 unexcused hours)

Consequence: "through negligence, exposed persons younger than eighteen years to the danger of neglect by allowing them to lead an idle life"

It contains all the legal elements of the constituent facts of the criminal offense of endangering the moral upbringing of youth according to § 211 par. 1 letter b) of the Criminal Code.

It ends with a comma, followed by the legal qualification: "thereby committing the offense of endangering the moral upbringing of youth according to § 211 par. 1 letter b) of the Criminal Code."

It is located in the operative part of the judgment (before the reasoning).

It unequivocally and unmistakably describes the act for which the accused is convicted.

THE FACTUAL STATEMENT IS USUALLY FOUND AFTER THESE AND SIMILAR TERMS:

sa uznáva za vinného, že/uznáva sa za vinného/sú vinní, že/uznáva sa vinným, že/sa uznáva vinným, že/skutkovom základe, že (all can be either spaced out or not)

THE FACTUAL STATEMENT ALWAYS IN 100% OF CASES ENDS WITH THIS AND SIMILAR TERM

t e d a (t h u s / t h e r e f o r e)

If the text does NOT meet these conditions, it means it does not have a factual statement

If you do not find a factual statement in the text, use this structure:

```
{
  "skutkova_veta": null,
  "no_factual_statement_reason": "Explanation why the text does not contain a factual statement."
}
```

Example:

Provided text: "Obvinený Ján Novák dňa 12.5.2022 o 15:30 hod. na ulici Dlhá v Bratislave fyzicky napadol poškodeného Petra Svobodu, čím sa dopustil trestného činu ublíženia na zdraví podľa § 156 trestného zákona." (Accused Ján Novák on 12.5.2022 at 15:30 hrs. on Dlhá street in Bratislava physically assaulted the victim Peter Svoboda, thereby committing the criminal offense of bodily harm according to § 156 of the Criminal Code.)

Expected AI output:

```
{
  "skutkova_veta": "Ján Novák dňa 12.5.2022 o 15:30 hod. na ulici Dlhá v Bratislave fyzicky napadol Petra Svobodu.",
  "no_factual_statement_reason": null
}
```

Expected AI output:

```
{
  "skutkova_veta": null,
```

```
"no_factual_statement_reason": "The text only contains
procedural information about the termination of proceedings, it
does not contain a description of the act."
}
```

HOW TO WORK WITH TEXT THAT HAS A FACTUAL STATEMENT?

Examples of factual statements:

1.

on 29.12.2019 at around 17.05 h in Bratislava in OC Centrál in the DM Drogerie store gradually took from the shelves 1 pc Denim VPH original 100 ml valued at 12.58 Eur, 3 pcs Gillette VPH Artic ice 100 ml valued at 48.93 Erur, 1 pc Denim VPH original 100 ml valued at 12.58 Eur, 4 pcs Gillette M Fusion power syst. valued at 89.94 Eur, 2 pcs Gillette M Fusion 1+1NH valued at 35.97 Eur, 2 pcs Gillette M Fusion proshield syst. 1+4 NH valued at 49.98 Eur, 1 pc Denim VPH black 100 ml valued at 56.61Eur, 3 pcs Gillette Mach 3 syst. 1+1 NH valued at 49.95, Eur, immediately placed the mentioned goods into a plastic bag and subsequently passed through the checkout zone without paying, thereby causing damage to the company DM drogerie markt, s.r.o., Bratislava, Na pántoch 18, IČO: 31 393 781, in the amount of 356.54 Eur,

2.

1. at an unspecified time at the end of August 2010 in the village of Y. the accused K. entered without consent and knowledge of the owner into the yard of family house no. XXX and from an unlocked shed took a motor mower brand Jičín, red color, which he pushed out in front of the gate, where all three accused together loaded it into a motor vehicle brand Q. U. and drove away, thereby causing damage to the owner of the mower Q. D., born XX.XX.XXXX in the amount of 179.20 €
2. at an unspecified time, in the period from 21:30 hrs. on 31.08.2010 to 07:00 hrs. on 01.09.2010, in S. Y., city district Y. all three accused came to the grocery store X. N., with a brought crowbar removed two padlocks on the iron cage located at the entrance to the store, from where they took 14 pcs

propane-butane cylinders with gas filling weighing 10 kg, loaded them into a motor vehicle brand Ford

Transit and drove away, thereby causing damage to the owner of the stolen cylinders, company C. M. B. H. O. K. Z. G. Z..G.,

VAT ID: XX XXX XXX damage in the amount of 370.44 € and to the owner of the damaged locks and stolen

gas, cooperative X. N. U., G. S., VAT ID: XX XXX XXX damage in the amount of 202.80 €

3. at an unspecified time, in the period from 17:00 hrs. on 31.08.2010 to 10:00 hrs. on 02.09.2010 in

the village of G. S. all three accused came to the grocery store, with a brought crowbar removed the padlocks

on the iron cage located at the entrance to the store, from where they took 10 pcs propane-butane

cylinders with gas filling weighing 10 kg, loaded them into a Ford Transit motor vehicle and drove away, thereby

causing damage to the owner of the stolen cylinders, company C. Z..G., VAT ID: XX XXX XXX, damage in the amount of 264.60

€ and to the owner of the damaged locks and stolen gas, I.. M. S., born XX.XX.XXXX damage in the amount of

142.80 €

4. at an unspecified time in the period from the beginning of December 2010 to 13.02.2011 in S. Y.

on I.. R. street the accused K. broke the entrance door lock area on garage no. 1429, entered inside

and inside the garage broke the locked trunk door of the vehicle parked there, brand X.

D., Lic. No. S.

-XXXAJ, from which he took a four-wheeled tractor mower brand Rider F XX, serial no. XXXXX

and from the garage took 8 pcs winter tires, a men's mountain bike, blue color, brand Author and wooden

sleds, thereby causing damage to the owner of the garage and stolen items, N.. K. R., born XX.XX.XXXX damage in the amount of

1,921.01 €,

whereas the accused K. acted this way despite the fact that by the penal order of the District Court M. no. XT/XX/XXXX

dated 29.05.2010, legally effective on 29.05.2010, he was found guilty of committing the offense "Theft" under § 212 par. 2 letter a), par. 3 letter a) of the Criminal Code,

3.

that

on 29. 3. 2013 at 10.45 hrs. drove a personal motor vehicle brand Renault Scénic, reg. no. LM-040 CF, in

Liptovský Mikuláš along Demänovská cesta in the direction from OD Kaufland towards Palúčanská street and near the

Elementary School Demänovská cesta was stopped by a patrol of the Regional Directorate of the Police Force Žilina, rapid response unit

PZ Žilina and performed this activity despite the fact that by the decision on the offense of the OR PZ, District

Traffic Inspectorate Liptovský Mikuláš under no. ORPZ-LM-ODI2-P-364/2011 dated 26. 8. 2011, which

became legally effective on 26. 8. 2011, he was imposed a ban on driving motor vehicles for a period of

36 months from the legal effectivity of the decision,

4.

in the period for the month of March 2018, December 2018, January 2019 until

21.08.2019 inclusive as the father of minor

son G. Č., born XXXX, he fails to fulfill his maintenance obligation in Y. and in other places where he stays,

although this obligation arises from the Family Act and was determined for him by the judgment of the District Court Skalica file ref.

8P/155/2017 dated 28.02.2018, legally effective on 07.03.2018, by which he was entrusted to the custody of the mother

D. O., residing at Y., M. XX and the father was ordered to contribute to the maintenance of the minor son G. Č. in the amount of 150,

€ by the 15th day of each month in advance into the hands of the mother, thereby causing for the specified period a debt on maintenance in the amount of 1,200,

- € into the hands of the mother D. O., residing at Y., M. XX,

5.

Accused P. Z. born XX. XX. XXXX P. X. permanently residing at L.Á. S. XX, X. X. is found guilty that as a parent obliged to continuously care for the upbringing and comprehensive development of the child, she inconsistently approached the fulfillment of her duties according to the Family Act No. 36/2005 Coll. as amended and created conditions for the minors C. Z., born XX. XX. XXXX, permanently residing at L. S. XX, X. X. and Š. Z., born XX. XX. XXXX, permanently residing at L. S. XX, X. X., for the emergence of undesirable habits in the form of long-term truancy, when C. Z. in the school year 2017/2018 missed unexcusedly 93 hours from the teaching process at Š. Z. Š., Ď. XX, X. X., in the period from 01. 09. 2017 to 30. 06. 2018, and Š. Z. in the school year 2017/2018 missed unexcusedly 194 hours from the teaching process at Š. Z. Š., Ď.D. XX, X. X., in the period from 01. 09. 2017 to 30. 06. 2018, through negligence, exposed persons younger than eighteen years to the danger of neglect by allowing them to lead an idle life,

Expected AI output:

```
{
  "skutkova_veta": "ako rodič povinný sústavne sa starať o
výchovu a všestranný rozvoj dieťaťa nedôsledne pristupovala k
plneniu si povinností podľa Zákona o rodine č. 36/2005 Z. z. v
platnom znení a vytvorila maloletým C. Z., narodenému XX. XX.
XXXX, trvalo bytom L. S. XX, X. X. a Š. Z., narodenému XX. XX.
XXXX, trvalo bytom L. S. XX, X. X., podmienky pre vznik
nežiaducich návykov vo forme dlhodobého záškoláctva, keď C. Z.
v školskom roku 2017/2018 z vyučovacieho procesu na Š. Z. Š.,
Ď. XX, X. X., v období od 01. 09. 2017 do 30. 06. 2018 vymeškal
neospravedlnene 93 hodín a Š. Z. v školskom roku 2017/2018 z
vyučovacieho procesu na Š. Z. Š., Ď.D. XX, X. X., v období od
01. 09. 2017 do 30. 06. 2018 vymeškal neospravedlnene 194
hodín, vydala z nedbanlivosti osoby mladšie ako osemnásť rokov
nebezpečenstvu spustnutia tým, že im umožnila viesť záhaľčivý
život",
  "no_factual_statement_reason": null
}
```

Instructions:

1. Identify the main factual statement in the provided text
2. Extract the factual statement exactly as it is stated in the text
3. DO NOT invent any new facts or information that are not explicitly stated in the provided text.

4. NEVER include the text before and after the key passage sa u z n á v a z a v i n n é h o, ž e / u z n á v a s a z a v i n n é h o / s ú v i n n í, ž e / u z n á v a s a v i n n ý m, ž e / s a u z n á v a v i n n ý m, ž e / s k u t k o v o m z á k l a d e, ž e (all can be either spaced out or not)/teda...
5. Return the response in JSON format with the following structure:

```
{  
  "skutkova_veta": "Extracted factual statement from the  
provided text",  
  "no_factual_statement_reason": null  
}
```

Appendix B: Example Slovak Criminal Verdict

Súd:	Okresný súd Malacky
Spisová značka:	3T/52/2022
Identifikačné číslo súdneho spisu:	1622010469
Dátum vydania rozhodnutia:	02. 11. 2022
Meno a priezvisko sudcu, VSÚ:	Mgr. Adriana Šimková
ECLI:	ECLI:SK:OSMA:2022:1622010469.2

Case identification

ROZSUDOK V MENE SLOVENSKEJ REPUBLIKY

Okresný súd Malacky samosudkyňou Mgr. Adrianou Šimkovou v trestnej veci obžalovaného J. R., nar. XX.XX.XXXX, pre zločin ublíženia na zdraví podľa § 155 ods. 1 Tr. zák., štádiom: pokus podľa § 14 ods. 1 Tr. zák., na hlavnom pojednávaní konanom dňa 02.11.2022, takto

General information

rozhodol:

Obžalovaný
J. R., nar. XX.XX.XXXX v L., trvalé bydlisko: J. Y.,

sa uznáva za vinného, že

Starting point

obvinený dňa 20.06.2022 v čase približne o 19.00 h na priestranstve pešej zóny pri predajni Coop Jednota a reštaurácii Oliveto na ul. D. Skuteckého v Malackách po vzájomnej slovnej roztržke s poškodeným K. K. držiac v rukách lopatu s drevenou rukoväťou, s touto lopatou sa obvinený niekoľkokrát zahnal po poškodenom K. K. a následne poškodeného udrel kovovou časťou lopaty do oblasti ľavej časti hlavy, čím poškodenému K. K. spôsobil otras mozgu, tržnozmliaždenú ranu čela vľavo v dĺžke 5 cm a tržnozmliaždenú ranu až čiastočnú amputáciu ľavej ušnice, ktoré zranenie si vyžiadali liečenie v trvaní 20 dní, čím vykonal útok proti takej časti tela poškodeného K. K., ktorá je dôležitá pre život,

Fact sentence

teda

Ending point

v úmysle spáchať trestný čin dopustil sa konania, ktoré bezprostredne smerovalo k úmyselnému spôsobeniu ťažkej ujmy na zdraví,

Legal subsumption

čím spáchal

zločin: ublíženie na zdraví podľa § 155 odsek 1 Trestného zákona, štádiom: pokus podľa § 14 odsek 1 Trestného zákona.

Za to sa
odsudzuje

Podľa § 155 ods. 1 Trestného zákona, s použitím § 38 ods. 2, ods. 3 Trestného zákona, § 36 písm. písm. d), l) Trestného zákona a s použitím § 39 ods. 2 písm. d), ods. 4 Trestného zákona per analogiam na trest odňatia slobody v trvaní 3 (tri) roky.