CHRISTOPH ENGEL
JOHANNES KRUSE

PROFESSOR GPT: HAVING A LARGE LANGUAGE MODEL WRITE A COMMENTARY ON FREEDOM OF ASSEMBLY

Christoph Engel   Johannes Kruse

Professor GPT
Having a Large Language Model Write a Commentary on Freedom of Assembly

I.   Task

Superintelligence is on the horizon (Bostrom 2017, Ashenbrenner 2024). The first research papers are out that have been completely designed, researched and written by large language models (Sakana 2024). Compared with this prospect, the achievement reported in this paper is modest. All we have GPT programmed to do is summarization. Yet for the legal community, this modest step is a big one.

In our project, GPT is not just summarizing a single text. It is writing a structured summary of the complete jurisprudence by the European Court of Human Rights on one of the fundamental freedoms protected by the European Convention on Human Rights. GPT has written a commentary on freedom of assembly, as protected by Art 11 EHCR. The text is written in the continental European tradition of a commentary. The output is organized the same way as much of European legal scholarship works - except that the author is taken out of the equation. As we demonstrate, the output looks exactly as this rich class of legal scholarship. Actually it even outperforms the competition. The GPT commentary is more encompassing, and also more functional, than competing texts written by human jurists. The result of the exercise is available here:

http://professor-gpt.coll.mpg.de/html/overview.html

As we will discuss, commentaries written by law professors may provide the legal community with additional services that are not part of the present project. But the core of every commentary is the summarization of legal thinking, organized along the elements of doctrine that have emerged from its application. And the core of the summarization part is the summarization of jurisprudence. This is not only the most laborious part. It is also why commentaries are so widely used in legal practice. Typically practitioners do not have time to, themselves, read through a rich and multifaceted body of legal materials. Commentaries are so popular precisely because they give practitioners easy and reliable access to the state of the art. It is this function that a commentary written by a large language model is able to fulfil.

As we show in this paper, the quality of the summarization performed by the large language model is on par with summaries written by trained lawyers. One more task can be delegated to machines. This delegation is not only appealing since law professors are not exactly cheap. With the help of language models, the task can be profoundly facilitated. As long as most of the legal literature is behind pay walls and copyrighted, it may be safer not to have the large language model also summarize the state of academic thinking on a given statutory provision. Experienced academics may see connections to analogous literatures that, at least for the

time being, would be harder for a language model to find. Academics may also amalgamate summarization with suggestions for legal evolution. These additional functions of commentaries invite a division of labor. The large language model does the tedious part, i.e. the summarization. The professor focuses on the finish.

The large language model does not shy away from the sheer amount of input. On the database of the European Court of human rights, 1198 cases are posted that discuss Art. 11 ECHR. Very assiduous legal scholars might read and organize them all. But practically, nobody has done so. By contrast, for a large language model, being comprehensive is no serious challenge.

We have chosen this provision as there is competition. Since we have started this project in the spirit of a proof of concept, we wanted the possibility to comparatively assess quality. Yet knowing that a large language model can write a very usable commentary is even more important for areas of law that have not been structured in this way. There are multiple reasons for the absence of a commentary. Probably the practically most important reason is what in data analysis would be called the long tail. Provisions that are of high practical importance have long been covered by commentaries, often by multiple competing commentaries. But many provisions are only relevant for a limited class of cases. For such matters, often no commentary is available. A second use case is motivated by different professional traditions. Commentaries are standard in the German speaking countries. There are some examples in other language families. But in many jurisdictions, including the UK and the US, the classic German commentary is not available for any statutory or jurisprudential rule. Given our proof of concept, it might be worth trying whether the legal community in these countries finds it appealing to have the jurisprudence organized in this alternative way. In addition, the prototype presented here can also be used (with minor modifications) for legal literature formats that play a significant role in common law jurisdictions. First and foremost, "annotations" to constitutions and statutes (such as the United States Code Annotated and the United States Code Service), provide not only the actual texts of the statute but also a summary of the cases that interpret the statute (Reimann 2020).

The most attractive feature of a commentary written by a large language model is timing. In areas of law that are both very active fields of jurisprudence and of high practical importance, commentaries may be updated every year. But especially if the field is active, a lot may happen during a year. And in many other fields, practitioners must wait much longer than a year before the next edition of the commentary becomes available. By contrast a large language model can be programmed such that the commentary is updated at very short intervals.

The primary interest in commentaries certainly stems from legal practice. But the structured analysis of jurisprudence is also useful for academic research. It is a straightforward next step to translate the classification that is necessary for writing the commentary into code. This provides academic research of judicial policy making with high-quality, fine grained sets of features that did not exist before.

The remainder of the paper as organized as follows: in Section II, we introduce the topic of our commentary, i.e. the guarantee of freedom of assembly in the European Convention on Human Rights. Section III explains the capabilities of large language models that make them a promising tool for the task. Section IV compares our task to other legal tasks for which LLMs

have already been used. Section V explains in detail how we have proceeded. Section VI assesses the performance of the commentary that GPT has written. Session VII concludes with discussion.

II.   Freedom of Assembly, as Protected by the European Convention on Human Rights

It is standard in computer science to assess the performance of a tool with the help of benchmarks (for an overview see Chang, Wang et al. 2024). Such benchmarks have also been developed for legal applications (Guha, Nyarko et al. 2024, Katz, Bommarito et al. 2024). Benchmarks allow to gauge the confidence a user may have in a tool. Our task is the structured summarization of an entire body of jurisprudence. This task does not lend itself to an easy quantitative assessment. We have therefore opted for a qualitative approach. We have selected an area of law where (a) there is sufficiently rich jurisprudence to make summarization meaningful and (b) summarisations, in the form of a commentary, are available that have been written by professional lawyers. Now commentaries are not a standard tool in legal practice in either the US or the UK. On the other hand, the lingua franca of the international academic community is English. This has led us to an international instrument that produces its output (at least predominantly) in English, while being applicable and of practical relevance in the German law speaking jurisdictions. The latter feature is responsible for the availability of human written commentaries in English language. The Germanic tradition has contaminated this international jurisdiction.

Specifically, we are able to compare the commentary written by GPT with the following two types of competitors: The first type consists of technical-functional competitors that do not address Article 11 of the European Convention on Human Rights (ECHR) but offer tools for summarizing court decisions. In this context, the offerings of major commercial providers are particularly noteworthy. Thomson Reuters (with Westlaw AI-Assisted Research and Ask Practical Law AI) and LexisNexis (with Lexis+AI) both provide platforms featuring LLM-supported tools capable of summarizing court decisions, among other functionalities (Nexis 2023, Reuters 2023). In Germany, for example, Wolters Kluwer offers GPT-based summaries of court decisions (Kluwer 2024).

On the other hand, content-related competitors—those that also offer a structured summary of the case law on Article 11 of the European Convention on Human Rights—provide an ideal starting point for a comparative evaluation. The Guide on Article 11 ECHR (Rights 2024) serves as a gold standard, representing an exceptionally expert human preparation of case law. The Guide is prepared by the Registry of the European Court of Human Rights (ECHR) rather than by the judges themselves. In reality, however, the Registry should be equally, if not more, informed about the Court's case law. This is because the Registry is responsible for providing the legal and administrative services required by the Court (see Rule 18 of the Rules of Court, 28 March 2024). Additionally, a unit within the Registry, Jurisconsult, is tasked with ensuring the quality and consistency of the Court's case law (see Rule 18B).

A further reason for selecting freedom of assembly as protected by the European Convention on Human Rights is a parallel exercise by the two authors of this paper. In a companion project,

we have programmed GPT to write a commentary on freedom of assembly as protected by article 8 Basic Law, i.e. by the German constitution (see the companion paper Engel and Kruse 2024). This makes it possible to compare the performance of GPT across both jurisdictions. We in particular are in a position to identify additional challenges present in the jurisprudence of the European Court of Human Rights.

While the main reason for selecting the application is thus pragmatic, freedom of assembly is an academically interesting and practically relevant topic in its own right. The guarantee reads:

> 1. Everyone has the right to freedom of peaceful assembly and to freedom of association with others, including the right to form and to join trade unions for the protection of his interests.
> 2. No restrictions shall be placed on the exercise of these rights other than such as are prescribed by law and are necessary in a democratic society in the interests of national security or public safety, for the prevention of disorder or crime, for the protection of health or morals or for the protection of the rights and freedoms of others. This Article shall not prevent the imposition of lawful restrictions on the exercise of these rights by members of the armed forces, of the police or of the administration of the State.

Freedom of assembly is a fundamental right in a democratic society and, like the right to freedom of expression, is one of the foundations of such a society (Salát 2015, Butler 2016, Rights 2024). This right has been instrumental in nearly every major social movement throughout history (Inazu 2010) and remains vital even in the information and internet age (Lewis 2006). Recently, the right to assemble has been central to discussions on significant social and global political conflicts, including the COVID-19 pandemic, the Black Lives Matter movement, and pro-Palestine protests. The pandemic, in particular, highlighted the challenging balance between the right to assemble and other protected interests, such as public health (Kruse and Langner 2021).

III. The Power of Large Language Models

The human mind is a black box, and so are large language models. Precisely what makes large language models so powerful also makes them opaque. Large language models no longer require fully determined if-then relations. They can handle the characteristic open texture of legal decision-making (Bix 1991, Schauer 2013). They do not shy away from ambiguity (Ellsberg 1961, Edelman 1992, Etner, Jeleva et al. 2012). They strive at making sense of the available input as best they can (Weick 1995, Turner, Allen et al. 2023). Still at a rather high level, it can be described how large language models work.

Language models make predictions. More precisely: they complete sentences that have been started. Given the text they have received so far: what is the most likely continuation? The user therefore controls language models with the input they provide. These are referred to as prompts. A prompt need not consist of a single sentence. The most advanced language models can process very long texts, even an entire book. Experience has shown that it is not only important to tell the language model as precisely as possible what it should do. A whole art

of particularly skillful ways of asking the computer questions has developed, prompt engineering (Sahoo, Singh et al. 2024), including applications to law (Choi 2023).

Language models use machine learning. Machine learning organizes large amounts of data. New observations are either classified (top down) in decision trees or they are assigned (bottom up) to other data points that are as closely related as possible (for an excellent introduction see James, Witten et al. 2022). Neural networks are particularly sophisticated instruments for this task. Not only can they process a large number of dimensions, they can also place these dimensions in complex relationships to one another. They can also have an architecture that allows preliminary assignments to be checked and gradually refined (for background see Goodfellow, Bengio et al. 2016). Transformers do not just translate inputs (e.g. natural language) into long chains of probabilities; computers can deal with such chains much more effectively. Rather, they provide the neural network with an attention mechanism (Vaswani, Shazeer et al. 2017). They use rich training data sets in this translation process. In this way, the local classification task is embedded in the "knowledge" that the architecture has previously acquired (Lin, Wang et al. 2022). Language models build on all these elements and add a generative component. The output no longer merely consists of an assignment of a data point to a class. Rather, the model can write texts (or generate images, or output sounds) (Chang, Wang et al. 2024).

The size of a language model refers to both the number of its parameters and the size of its training corpus. Large language models (LLMs) are models that contain billions of parameters and are trained with huge corpora that can be as large as the complete Internet.[1] The training data also includes legal information. However, legal texts regularly make up only a fraction of the data (Colombo, Pires et al. 2024). Before we present a prototype of the commentary without an author and report on our experience with GPT as an annotator, we first outline the current state of research.

IV.   Large Language Models in Legal Practice, and in Legal Research

To the best of our knowledge, language models in general and GPT in particular have not yet been used to write a legal commentary – except for our own companion project on the parallel provision in the German constitution (Engel and Kruse 2024). This is probably due to the fact that, although commentaries exist as a legal literature genre in many jurisdictions, they are not as important as in the German speaking countries (Kästle-Lamparter, Jansen et al. 2020). Nevertheless, there is a related genre, the annotated statutes, which is very similar to the commentary we present (Reimann 2020). Despite this similarity, there appear to have been no attempts to create such works using large language models. That said, the application of LLMs in law remains a very active area of research.

Language models have already been used for a variety of legal tasks (Kapoor, Henderson et al. 2024): from legal education (Choi, Hickman et al. 2021, Choi and Schwarcz 2023) to empirical legal research (Drápal, Westermann et al. 2023, Livermore, Herron et al. 2024) and legal

---

[1] Current frontier models are said to use 1.8 trillion (GPT-4o), 405 billion (Llama 3.1 405B) and 176 billion parameters (Mixtral 8x22B); Claude 3.5 Sonnet Opus has not disclosed the number of parameters.

practice (Rodgers, Armour et al. 2023, Bilgin and Licato 2024, Trozze, Davies et al. 2024). The use of LLMs has already achieved considerable success. For example, GPT-4 was able to answer questions on the US Bar Exam with an average accuracy rate that would have been sufficient to pass in all states (Katz, Bommarito et al. 2024) or showed high accuracy in extracting legal information from Employment Tribunal judgments (de Faria, Xie et al. 2024). The subsumption skills (statutory reasoning) of LLMs have also been examined (Trozze, Davies et al. 2024, Zou, Zhang et al. 2024). Even simple subsumption tasks caused difficulties though for the (now outdated) GPT-3 model: answers were only correct in around 4 out of 5 cases (Blair-Stanek, Holzenberger et al. 2023).

A commentary summarizes court decisions and literature, i.e. texts. Summaries of texts have also been aided by language models in other disciplines and areas of life, like news (Goyal, Li et al. 2022), research (Goyal, Li et al. 2022) or biomedical findings (Han and Choi 2024). Language models can summarize texts in two different ways: they either extract or abstract. The language model either attempts to identify individual sentences of particular importance. Or the model attempts to understand the entire text in order to then create an (original) summary (Shukla, Bhattacharya et al. 2022). LLMs are now able to summarize texts as accurately as human readers; this also applies to legal texts (Shukla, Bhattacharya et al. 2022, Haddaya, Macmillan et al. 2024).

Commentaries are popular not least because they relieve the legal practitioner of work. Practitioners do not need to evaluate the wealth of potentially relevant material themselves. They can limit themselves to looking up the references that the commentary flags as directly relevant. However, this form of division of labor between author and reader presupposes that the reader can rely on the commentary. Even commentaries written by humans occasionally make mistakes. However, these errors are not systematic. The first language models, on the other hand, made serious and systematic errors. This problem has diminished with each development step. However, it is not yet completely under control. This is why we also report on the literature on such errors.

Hallucinations have attracted particular attention (Mik 2023). If the language model hallucinates, it generates results that are not at all based on the input given to them. In the present context, this could mean that the model "references" a non-existent ruling, or one that discusses a different human right. A study has found that ChatGPT-4 answered legal questions incorrectly in 58% of cases (Dahl, Magesh et al. 2024). Deroy et al. investigated the extent to which GPT-3.5 Turbo is suitable for summarizing court decisions. They identified multiple hallucinations and came to the conclusion that language models are not yet capable of providing fully automated summaries of legal texts (Deroy, Gosh et al. 2023).

Such findings should, however, be treated with caution. The hallucination rate depends to a large extent on the specific circumstances of use: which LLM is used in which way; what exactly does the task look like; what information is made available to the LLM? A major cause of hallucinations is a lack of knowledge. Language models do everything they can to answer a query. If they do not have access to the knowledge they need, they are liable to filling the gap with non-existent, but probable information. This is why hallucinations occur more frequently when language models are asked open questions (Magesh, Surani et al. 2024).

Prompts also have a significant influence on the quality of the results. They can be used to provide the LLM with the necessary contextual knowledge, or to point out potential misunderstandings. The instruction to divide complex tasks into smaller tasks and then work through them step by step can also lead to substantial improvements. Kojima et al. were able to show that it is sufficient to add the following instruction at the end of the prompt: "Let's think step by step" (Kojima, Gu et al. 2022). It can also help to make the LLM understand the task with the help of an example (few shot prompt, (Brown 2020, Xie, Steffek et al. 2024)).

V.   Programming GPT to Write a Commentary

1.   Jurisprudence of the European Court of Human Rights

The first step of the process on which we report in this paper does not require a large language model. Happily the European Court of Human Rights is very transparent. On its website, it posts the complete text of 68,956 decisions. This is a large number, but still only a fraction of the cases that have been submitted: in the 10 years from 2014 to 2023, 467,300 cases have been allocated to one of the decision-making bodies set up by the Convention.[2] Were this a paper intending to causally analyze the jurisprudence of the court, we would have to worry about selection. Yet for our purposes, this limitation is mild. It would only matter if the court had kept rulings confidential that are of high importance for predicting the future decision of analogous cases. Theoretically, we cannot exclude this possibility. But it is a very unlikely concern. Informing the general public, and governments for that matter, about the development of its jurisprudence is the most important policy lever of the court. This is reflected in a very elaborate and outspoken policy of promulgating such developments.[3]

In principle, as the first step of the process, we might have downloaded all rulings that the court has posted, and would then have filtered them for freedom of assembly. We did not have to do that as the court maintains a very well-organized database.[4] This has allowed us to directly filter cases that the database highlights for discussing Art. 11 ECHR.[5] This gave us a wider set of 1198 cases.

Downloading these cases was a bit of a challenge, as the database of the court is constructed as a dynamic website.[6] We could therefore not directly target the .html code with the `beautiful soup` package in `python`, and had to mimic browsing to the dynamic site for each case, with the help of the `selenium` package. On each site, we had the program click

---

[2] https://www.echr.coe.int/documents/d/echr/stats-analysis-2023-eng?download=true, p. 6. For the early years of the Convention bodies, see https://www.echr.coe.int/documents/d/echr/survey_19591998_bil.

[3] See most notably the definition of „key cases", and their distribution in the official reporter, https://www.echr.coe.int/en/selection-of-key-cases, and the establishment of a separate knowledge-sharing institution, https://www.echr.coe.int/knowledge-sharing.

[4]
https://hudoc.echr.coe.int/eng#{%22languageisocode%22:[%22ENG%22],%22documentcollectionid2%22:[%22JUDGMENTS%22,%22DECISIONS%22]}.

[5] The exact filtering steps are documented in the ReadMe document.

[6] For background see https://en.wikipedia.org/wiki/Dynamic_web_page.

the .pdf button, and download the file linked to it.[7] Next we have extracted the raw text from each file, using the `PyPDF2` package. After a series of data cleaning steps[8] we had our set of 1198 raw data files.

The programmers of the court's website have erred on the cautious side. Quite a number of the raw data files are actually not discussing freedom of assembly. The main reason is the construction of Art. 11 ECHR. The provision also covers freedom of association (406 cases). Other files on closer inspection do not discuss either human right (101 cases). This leaves us with an actual set of 691 cases.

2. Classifying individual rulings

For the analysis we have used GPT-4o. This model has an impressive context window of 128,000 tokens. Still the totality of the 691 cases uses 28 MB, too much even for a frontier model.[9] Additionally, performance decreases when processing inputs that approach the token limit. A recent study found that large language models (LLMs) exhibit a "lost in the middle" effect when handling extensive amounts of text, similar to the serial position effect observed in humans (Feigenbaum and Simon 1962, Murdock Jr 1962). When large volumes of information are present, LLM performance significantly drops if relevant data is situated in the middle of a document rather than at the beginning or end (Liu, Lin et al. 2024). This poses particular challenges when processing a large set of court decisions. We therefore proceed iteratively. We first analyze each ruling individually, and only in a second step aggregate over the summaries that GPT has written about each individual ruling.

The ultimate goal of this first step is the multidimensional classification of each individual case. As our process starts with individual rulings, we cannot have the large language model infer doctrine from the complete body of jurisprudence. Rather we inform it with the help of the system prompt. One may of course consider this a limitation of the approach. But this limitation is in no way different from the approach of human commentators. They would also not start from scratch, but would build on the established doctrine. Moreover, in the case of Art. 11 ECHR the mapping between the wording of the provision and the structure of the established doctrine a very close. Hence merely by reading the relevant provision, one would already come close to an understanding of the relevant doctrine. This implies that our upfront intervention is indeed very mild. Moreover as we will show, GPT does not stop at the established doctrine, and proposes multiple refinements.

The system prompt adds structure to the process. We inform the large language model about the typical content of a ruling, and ask it to focus on the opinion of the court, using the statements of the parties only to the extent that they help better understand the court's decision. In the spirit of chain of thought prompting (Wei, Wang et al. 2022), we ask the language model to first characterize in natural language whether, and if so how, the ruling addresses each individual element of the established doctrine. If, for the element of doctrine

---

[7] Some old cases where on static websites, so that we had to split the process, after discriminating between static and dynamic websites. For detail see the ReadMe document, and the associated code.

[8] Documented in the ReadMe document.

[9] A rough estimate is 6 characters per token, which would result in a total of 4,666,667 tokens.

in question, the response is positive, we further ask the language model to note the paragraph or paragraphs within which the court discusses this element of doctrine. Finally, to facilitate the next steps of the process, we ask GPT to respond in JSON format whether the ruling discusses each individual element of doctrine, "Yes" or "No".

This intermediate step already yields an interesting observation: the discussion of the elements of doctrine is very unevenly distributed (Figure 1). The large majority of the rulings spell out whether the governmental act or omission against which the complaint is directed falls into the substantive and personal scope of freedom of assembly, and whether the act interferes with this human right. A not much smaller number of rulings also applies the "necessary in a democratic society" test. By contrast, discussions of the various aims that the provision considers legitimate are rarer. Interestingly, the most frequently invoked justification is prevention of disorder, not the protection of the rights of others, let alone the remaining legitimate aims. Finally it is remarkable how frequently the court explicitly discusses whether the applicant deserves some form of just satisfaction.
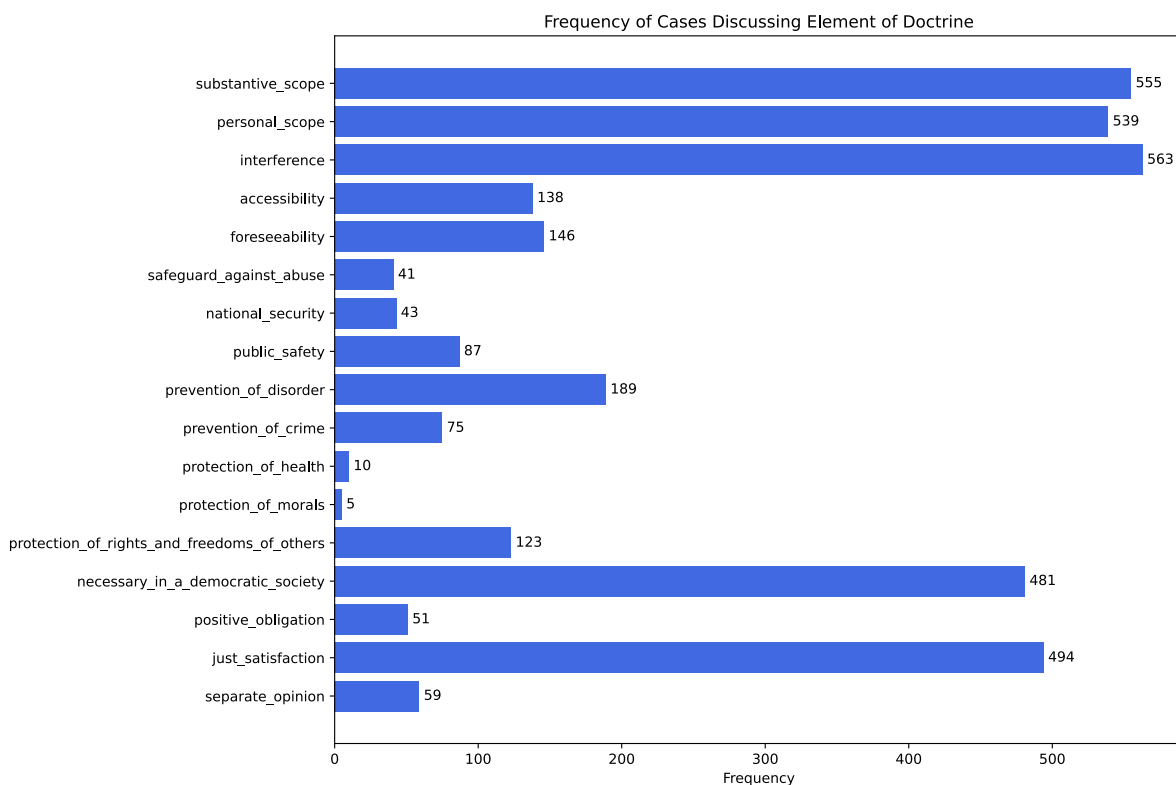

Figure 1
Intensity of doctrinal discussion

3. Summarization across rulings

The summarization across rulings starts with creating a folder, separately for each element of doctrine, with those paragraphs of all rulings that GPT has identified as discussing the element in question. As Figure 1 suggests, just combining all these relevant paragraphs into one single

file and asking GPT to summarize this file is not feasible. There is just too much information. Even if, technically, the file remains within the token window, results do not look convincing. GPT is overwhelmed with the sheer amount of text. This is why we handle the input data by batches of 100 rulings.[10]

While this split makes the process manageable, it creates a new challenge. When summarizing the next batch, GPT does not remember how it had summarized earlier batches. This is why there is pronounced heterogeneity both in style, but also in substance, across per batch summaries. Actually, this challenge is also an advantage. GPT gets multiple opportunities to structure the jurisprudence regarding the element of doctrine in question. We react the following way: we first ask GPT to only extract the subtopics it finds in the different per batch summaries, and to bring them into a coherent order. We then have GPT revisit each batch of input data in the light of this more coherent list of subtopics. This gives us summaries that are reasonably coherent across batches.

We concatenate these revised per batch summaries, separately for each element of doctrine, and for each subtopic within the element of doctrine that GPT has identified. To these raw summaries, we apply a number of data cleaning steps. In earlier steps of our process, we have singled out paragraphs that truly discuss the element of doctrine in question. Yet this step does not filter out text that conjointly discusses more than one element of doctrine. Despite the fact that we always have reminded GPT of the structure of doctrine, it repeatedly has summarized everything that has been said, not just what has been said about the element of doctrine in question. In a first cleaning step, we ask GPT to remove text from the summaries that belongs to other elements of doctrine. A further challenge results from the tendency of GPT to be redundant. In a second cleaning step, we ask it to also remove repetitive statements. Finally in a third pass, we ask GPT to double check whether summaries of one element of doctrine still cover material that belongs to other elements of doctrine, and to remove such text as well.

The complete code is available for scrutiny. In the Appendix, we illustrate the approach with the master for writing the system prompts for the final summarization step.

4. Presentation

We present the result on a set of websites. The commentary is structured hierarchically. At the top level, it is organized by the elements of doctrine that result from the wording of Art. 11 ECHR. At the next level below, the subtopics are listed that GPT itself has formulated. On the page for each subtopic, there is a general definition, and a list of applications. Each statement comes with references to the paragraphs from which GPT has taken the statement in question. Each reference is clickable and leads to the wording of the respective paragraph. We also provide a list of cases, complete with the case number in the system of the court, and a link to the full text of the ruling.

---

[10] Specifically, we split the input file at each 100 case numbers. Effectively therefore each batch contains less than extracts from 100 rulings, as no element of doctrine is discussed in each and every ruling.

VI.   Comparative performance

It is interesting to compare how our commentary and summarization pipeline performs against the competitors mentioned above.

Starting with technical-functional competitors, Wolters Kluwer only provides summaries of recent higher court decisions, preventing users from summarizing other cases of interest. In contrast, Thomson Reuters and LexisNexis offer this functionality. However, they limit users to summarizing individual decisions and then prompt them with follow-up questions (Beberness 2024). There are no options to specify detailed criteria for summaries, restricting personalization. Additionally, these tools cannot analyze and organize multiple decisions simultaneously. Furthermore, these commercial solutions are susceptible to hallucinations, despite claims of being "hallucination-free" (Nexis 2023). A study found that LexisNexis and Thomson Reuters' AI tools hallucinate 17% to 33% of the time (Magesh, Surani et al. 2024).

In contrast, our pipeline offers a comprehensive approach to court decisions. We extract both the core elements of individual judgments and entire strands of case law. The results are organized according to a predefined structure. Our system allows extensive personalization through prompts, enabling, for example, commentaries focused on specific periods, regions, or types of assembly concerns, such as environmental protection or LGBTQ rights.

The Guide on Article 11 ECHR prepared by the Registrar of the Council of Europe is our most serious content-related competitor. First, we compare the costs of preparing the commentary. Because information for the Guide on Article 11 ECHR is unavailable, we can only report the costs of Professor GPT. Since there were no project-specific expenses for the authors' work and standard IT usage, we estimate only the costs of using the OpenAI API, which remain below $100.

Our Commentary Without Author covers Article 11 ECHR and relevant case law in much greater detail (427,000 characters, excluding spaces) than the Guide on Article 11 ECHR (62,000 characters, excluding spaces).

Like the Guide on Article 11 ECHR, Professor GPT focuses solely on ECtHR case law and does not incorporate academic literature, unlike typical commentaries. This makes this comparison especially suitable. We analyze how many different ECtHR decisions are cited at least once, and how frequently each decision is cited. The first metric indicates the breadth of citations, while the second measures their depth. The comparison yields the following results:

Metric	Commentary Without Author	Guide on Article 11 ECHR
Total Number of Case Citations (Unweighted)	12254	267
Total Number of Case Citations (Normalized by Text Length, per 1,000 characters)	28,7	4,3
Number of Unique Cases Cited (Unweighted)	572	118
Number of Unique Cases Cited (Normalized by Text Length)	1,34	1,9

Table 1
Comparison of Citation Metrics Between 'Commentary Without Author' and 'Guide on Article 11 ECHR'

This demonstrates that Professor GPT not only engages with case law more thoroughly, but also provides more comprehensive coverage. Out of 691 decisions related to Art. 11 ECHR identified in our evaluation, Professor GPT cites nearly 83% (572 decisions). In comparison, the Guide references just over one-sixth (17%) of the relevant case law.

Furthermore, the precision of citations is noteworthy. Readers benefit more when they learn not only that the ECtHR has addressed a specific interpretative question in a particular decision, but also the exact paragraph where the statement appears. Accurate citations also simplify the verification of the commentary. Unlike traditional commentaries (cf. Engel and Kruse 2024), both the Commentary Without Author and the Guide perform exceptionally well in this regard. The Guide references specific paragraph numbers in 95% of citations, while the Commentary Without Author achieves a perfect 100%. Additionally, the ease of accessing the relevant paragraphs is a significant advantage of the Commentary Without Author, especially in legal applications where time is critical. Although the Guide allows users to click directly to the referenced decision, it does not link to specific paragraphs, requiring users to manually navigate to the relevant sections.

The innovative advantage of Professor GPT should also be highlighted. The Guide, produced by the Registry of the European Court of Human Rights, adheres strictly to established case law to make it more accessible to readers. Consequently, it cannot highlight potential new interpretative approaches or consider individual circumstances or questions. In contrast, Professor GPT can emphasize different aspects in its summaries and introduce new categories. This capability is evident through the inclusion of subtopics not found in the Guide, aligning with research on the creativity of large language models. While not fully creative in the human sense, these models can generate novel content (Franceschelli and Musolesi 2023, Bellemare-Pepin, Lespinasse et al. 2024, Orwig, Edenbaum et al. 2024). The innovative potential of Professor GPT can be managed within certain limits. For instance, one study showed that higher temperature settings in language models lead to slightly more novel outputs

(Peeperkorn, Kouwenhoven et al. 2024). Since both judges and lawyers have access to the Commentary Without Author, any innovations it introduces could eventually influence decision-making practices. Additionally, the extensive personalization options offered by Professor GPT can provide new insights into case law content.

Finally, a comparative analysis of the commentary on Art. 8 of the German Basic Law ("Grundgesetz" or "GG") reveals that its scope (approximately 34,000 characters, excluding spaces) is significantly smaller than that of Art. 11 ECHR. The difference in size (1:12.5) is only partly due to the inclusion of more decisions in the Commentary Without Author for Art. 11 ECHR (1:6.5). The use of different GPT models (GPT-4 Turbo and GPT-4o) may also have contributed to this difference.

Metric	Commentary Without Author on Art. 11 ECHR	Commentary Without Author on Art. 8 GG
Total Number of Case Citations (Unweighted)	12254	553
Total Number of Case Citations (Normalized by Text Length, per 1,000 characters)	28,7	16,3
Number of Unique Cases Cited (Unweighted)	572	88
Number of Unique Cases Cited (Normalized by Text Length)	1,34	2,6

Table 2
Comparison of Citation Metrics Between 'Commentary Without Author on Art. 11 ECHR ' and 'Commentary Without Author on Art. 8 GG '

Additionally, examining the two citation metrics further highlights clear advantages for the Commentary Without Author on Art. 11 ECHR.

VII.   Discussion

Judges and administrators wield political power. It has always been suspected that these public officials abuse sovereign powers for advancing personal benefit (Posner 1993, Schauer 1999), individual political convictions (Segal and Cover 1989, Segal, Epstein et al. 1995), or that the fall prey to subconscious bias (Rachlinski, Johnson et al. 2008, Rachlinski, Wistrich et al. 2015). In this perspective, delegating judicial decision-making to machines may be considered beneficial (cf. Joh 2007). Computerized decision-making is also easier to scale. This might make sovereign decision making more affordable, and might reduce barriers for access to justice (Lippman 2014, OECD 2019). But automated decision-making may not be sufficiently accurate (Grgić-Hlača, Engel et al. 2019), it may be perceived as unfair (Dietvorst, Simmons et al. 2015, Grgic-Hlaca, Redmiles et al. 2018), and bias may sneak in unnoticed (Engel, Linhardt et al. 2024). There are thus opportunities, but also risks. The European legislator clearly

considers the risks to be more concerning. In the new AI Act, judicial applications are considered "high-risk", and require very heavy-handed institutional protection.[11]

This is not the right context to discuss whether a structured summary of the jurisprudence of a court already qualifies for a high-risk application in the sense of the AI Act. At any rate, a tool that merely makes it easier for legal practitioners to predict how the court is likely to decide a new case is less concerning. Such a tool does not impinge on the ultimate decision made by the competent panel of the court. On the other hand, the tool makes it considerably easier to navigate the particularly rich jurisprudence of the European Court of Human Rights. This may not only help applicants, and governments on the defense side for that matter. Easy structured access to the complete body of jurisprudence may also help the court itself see how its jurisprudence develops, and whether some of the parallel decision-making panels move into different directions.

Most importantly, the main purpose for developing the tool, and for writing this paper, has not been providing a decision aid for cases dealing with freedom of assembly as protected by the European Convention on Human Rights. This specific application is predominantly meant as a proof of concept. As the example shows, a rather elaborate process is required. Technology is not yet at a point where a commentary on any chosen legal provision can be provided by just "pushing a button". Maybe, over the years to come an even simpler process will suffice. So many tasks that seemed outside the reach of language models have been implemented over the last two years that optimism is not speculation. But even if, ultimately, heavier intervention by legal experts remains necessary to guarantee sufficient quality and usability, this intervention requires way less human input than the writing of a traditional commentary. The human legal expert may turn into a supervisor of the machine, rather than having to do all the heavy lifting in person. This opens up in the opportunity for much wider coverage, and much more rapid updating. These are substantial benefits for the legal community.

---

[11] Appendix III.6 and III.8, Regulation 2024/1689, OJ 2024 L 1689.

References

Ashenbrenner, Leopold (2024). <u>Situational Awareness. The Decade Ahead</u>.

Beberness, Ethan (2024). Inside Lexis+Ai. Lexisnexis' Latest Research Tool.

Bellemare-Pepin, Antoine, François Lespinasse, Philipp Thölke, Yann Harel, Kory Mathewson, Jay A Olson, Yoshua Bengio and Karim Jerbi (2024). Divergent Creativity in Humans and Large Language Models. <u>arXiv preprint arXiv:2405.13012</u>.

Bilgin, Onur and John Licato (2024). Determining Legal Relevance with Llms Using Relevance Chain Prompting. <u>The International FLAIRS Conference Proceedings</u>. **37**.

Bix, Brian (1991). "Hla Hart and the" Open Texture" of Language." <u>Law and Philosophy</u> **10**: 51-72.

Blair-Stanek, Andrew, Nils Holzenberger and Benjamin Van Durme (2023). <u>Can Gpt-3 Perform Statutory Reasoning?</u> Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law.

Bostrom, Nick (2017). <u>Superintelligence</u>, Dunod.

Brown, Tom B (2020). "Language Models Are Few-Shot Learners." <u>arXiv preprint arXiv:2005.14165</u>.

Butler, Judith (2016). "We, the People": Thoughts on Freedom of Assembly. <u>What Is a People?</u> G. Didi-Huberman, S. Khiari, J. Rancière et al., Columbia University Press**:** 49-64.

Chang, Yupeng, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang and Yidong Wang (2024). "A Survey on Evaluation of Large Language Models." <u>ACM Transactions on Intelligent Systems and Technology</u> **15**(3): 1-45.

Choi, Jonathan H (2023). "How to Use Large Language Models for Empirical Legal Research." <u>Journal of Institutional and Theoretical Economics</u> **180**: 214-233.

Choi, Jonathan H, Kristin E Hickman, Amy B Monahan and Daniel Schwarcz (2021). "Chatgpt Goes to Law School." <u>Journal of Legal Education</u> **71**: 387-400.

Choi, Jonathan H and Daniel Schwarcz (2023). "Ai Assistance in Legal Analysis. An Empirical Study." <u>Journal of Legal Education</u> **73**: ***.

Colombo, Pierre, Telmo Pessoa Pires, Malik Boudiaf, Dominic Culver, Rui Melo, Caio Corro, Andre FT Martins, Fabrizio Esposito, Vera Lúcia Raposo and Sofia Morgado (2024). Saullm-7b: A Pioneering Large Language Model for Law. <u>arXiv preprint arXiv:2403.03883</u>.

Dahl, Matthew, Varun Magesh, Mirac Suzgun and Daniel E Ho (2024). "Large Legal Fictions. Profiling Legal Hallucinations in Large Language Models." <u>Journal of Legal Analysis</u> **16**: 64-93.

de Faria, Joana Ribeiro, Huiyuan Xie and Felix Steffek (2024). Automatic Information Extraction from Employment Tribunal Judgements Using Large Language Models. <u>arXiv preprint arXiv:2403.12936</u>.

Deroy, Aniket, Kripabandhu Gosh and Saptarshi Gosh (2023). How Ready Are Pre-Trained Abstractive Models and Llms for Legal Case Judgement Summarization?

Dietvorst, Berkeley J, Joseph P Simmons and Cade Massey (2015). "Algorithm Aversion. People Erroneously Avoid Algorithms after Seeing Them Err." <u>Journal of Experimental Psychology: General</u> **144**(1): 114-126.

Drápal, Jakub, Hannes Westermann and Jaromir Savelka (2023). <u>Using Large Language Models to Support Thematic Analysis in Empirical Legal Studies</u>. JURIX.

Edelman, Lauren B (1992). "Legal Ambiguity and Symbolic Structures. Organizational Mediation of Civil Rights Law." <u>American Journal of Sociology</u> **97**(6): 1531-1576.

Ellsberg, Daniel (1961). "Risk, Ambiguity, and the Savage Axioms." Quarterly Journal of Economics **75**: 643-669.

Engel, Christoph and Johannes Kruse (2024). "Kommentar Ohne Autor. Können Sprachmodelle Das Kommentieren Übernehmen?" Juristenzeitung *******: ***.

Engel, Christoph, Lorenz Linhardt and Marcel Schubert (2024). "Code Is Law: How Compas Affects the Way the Judiciary Handles the Risk of Recidivism." Artificial Intelligence and Law: 1-23.

Etner, Johanna, Meglena Jeleva and Jean-Marc Tallon (2012). "Decision Theory under Ambiguity." Journal of Economic Surveys **26**(2): 234-270.

Feigenbaum, Edward A and Herbert A Simon (1962). "A Theory of the Serial Position Effect." British Journal of Psychology **53**(3): 307-320.

Franceschelli, Giorgio and Mirco Musolesi (2023). On the Creativity of Large Language Models. arXiv preprint arXiv:2304.00008.

Goodfellow, Ian, Yoshua Bengio and Aaron Courville (2016). Deep Learning, MIT press.

Goyal, Tanya, Junyi Jessy Li and Greg Durrett (2022). News Summarization and Evaluation in the Era of Gpt-3. arXiv preprint arXiv:2209.12356.

Grgić-Hlača, Nina, Christoph Engel and Krishna P Gummadi (2019). Human Decision Making with Machine Assistance. An Experiment on Bailing and Jailing. Proceedings of the ACM on Human-Computer Interaction.

Grgic-Hlaca, Nina, Elissa M Redmiles, Krishna P Gummadi and Adrian Weller (2018). Human Perceptions of Fairness in Algorithmic Decision Making. A Case Study of Criminal Risk Prediction. Proceedings of the 2018 World Wide Web Conference.

Guha, Neel, Julian Nyarko, Daniel Ho, Christopher Ré, Adam Chilton, Alex Chohlas-Wood, Austin Peters, Brandon Waldon, Daniel Rockmore and Diego Zambrano (2024). "Legalbench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models." Advances in Neural Information Processing Systems **36**.

Haddaya, Mourad, Kyle Macmillan, Anup Malani, Hongyuan Mei and Chenhao Tan (2024). Comparing Summaries of Legal Opinions Generated by Llms and by Experts.

Han, Hyunkyung and Jaesik Choi (2024). "Optimal Path for Biomedical Text Summarization Using Pointer Gpt." arXiv preprint arXiv:2404.08654.

Inazu, John D (2010). "The Forgotten Freedom of Assembly." Tulane Law Review **84**: 565-612.

James, Gareth, Daniela Witten, Trevor Hastie and Robert Tibshirani (2022). An Introduction to Statistical Learning, Springer.

Joh, Elizabeth E (2007). "Discretionless Policing: Technology and the Fourth Amendment." Calif. L. Rev. **95**: 199.

Kapoor, Sayash, Peter Henderson and Arvind Narayanan (2024). Promises and Pitfalls of Artificial Intelligence for Legal Applications**:** 5.

Kästle-Lamparter, David, Niels Jansen and Reinhard Zimmermann (2020). Juristische Kommentare. Ein Internationaler Vergleich. Tübingen, Mohr (Siebeck).

Katz, Daniel Martin, Michael James Bommarito, Shang Gao and Pablo Arredondo (2024). "Gpt-4 Passes the Bar Exam." Philosophical Transactions of the Royal Society A **382**(2270): 20230254.

Kluwer, Wolters (2024). Effizientere Juristische Recherche Durch Generative Ki.

Kojima, Takeshi, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo and Yusuke Iwasawa (2022). "Large Language Models Are Zero-Shot Reasoners." Advances in neural information processing systems **35**: 22199-22213.

Kruse, Johannes and Christian Langner (2021). "Covid-19 Vor Gericht: Eine Quantitative Auswertung Der Verwaltungsgerichtlichen Judikatur." Neue Juristische Wochenschrift **74**: 3707-3712.

Lewis, Charlie (2006). The Right of Assembly and Freedom of Association in the Information Age. Human Rights in the Global Information Society. R. F. Jorgensen**:** 151-184.

Lin, Tianyang, Yuxin Wang, Xiangyang Liu and Xipeng Qiu (2022). "A Survey of Transformers." AI open **3**: 111-132.

Lippman, Jonathan (2014). "The Judiciary as the Leader of the Access-to-Justice Revolution." New York University Law Review **89**: 1569-1588.

Liu, Nelson F, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni and Percy Liang (2024). "Lost in the Middle: How Language Models Use Long Contexts." Transactions of the Association for Computational Linguistics **12**: 157-173.

Livermore, Michael A, Felix Herron and Daniel Rockmore (2024). "Language Model Interpretability and Empirical Legal Studies." Journal of Institutional and Theoretical Economics **180**: 244-276.

Magesh, Varun, Faiz Surani, Matthew Dahl, Mirac Suzgun, Christopher D Manning and Daniel E Ho (2024). Hallucination-Free? Assessing the Reliability of Leading Ai Legal Research Tools. arXiv preprint arXiv:2405.20362.

Mik, Eliza (2023). "Caveat Lector: Large Language Models in Legal Practice." Rutgers Business Law Journal **19**: 70-128.

Murdock Jr, Bennet B (1962). "The Serial Position Effect of Free Recall." Journal of Experimental Psychology **64**(5): 482-488.

Nexis, Lexis (2023). Lexisnexis Launches Lexis+ Ai, a Generative Ai Solution with Linked Hallucination-Free Legal Citations.

OECD (2019). Legal Needs Surveys and Access to Justice.

Orwig, William, Emma R Edenbaum, Joshua D Greene and Daniel L Schacter (2024). "The Language of Creativity: Evidence from Humans and Large Language Models." The Journal of Creative Behavior **58**(1): 128-136.

Peeperkorn, Max, Tom Kouwenhoven, Dan Brown and Anna Jordanous (2024). Is Temperature the Creativity Parameter of Large Language Models? arXiv preprint arXiv:2405.00492.

Posner, Richard A. (1993). "What Do Judges and Justices Maximize? (the Same Thing Everybody Else Does)." Supreme Court Economic Review **3**: 1-41.

Rachlinski, Jeffrey J, Sheri Lynn Johnson, Andrew J Wistrich and Chris Guthrie (2008). "Does Unconscious Racial Bias Affect Trial Judges." Notre Dame Law Review **84**: 1195-1246.

Rachlinski, Jeffrey J, Andrew J Wistrich and Chris Guthrie (2015). "Can Judges Make Reliable Numeric Judgments. Distorted Damages and Skewed Sentences." Indiana Law Journal **90**: 695-739.

Reimann, Mathias (2020). Legal "Commentaries" in the United States. Division of Labor. Juristische Kommentare. Ein Internationaler Vergleich. D. Kästle-Lamparter, N. Jansen and R. Zimmermann. Tübingen, Mohr Siebeck**:** 277-294.

Reuters, Thomson (2023). Introducing Ai-Assisted Research. Legal Research Meets Generative Ai.

Rights, European Court of Human (2024). Guide on Article 11 of the European Convention on Human Rights.

Rodgers, Ian, John Armour and Mari Sako (2023). "How Technology Is (or Is Not) Transforming Law Firms." Annual Review of Law and Social Science **19**(1): 299-317.

Sahoo, Pranab, Ayush Kumar Singh, Sriparna Saha, Vinija Jain, Samrat Mondal and Aman Chadha (2024). A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. arXiv preprint arXiv:2402.07927.

Sakana (2024). Adaptive Learning Rates for Transformers Via Q-Learning.

Salát, Orsolya (2015). The Right to Freedom of Assembly.

Schauer, Frederick (1999). "Incentives, Reputation, and the Inglorious Determinants of Judicial Behavior." University of Cincinnati Law Review **68**: 615-636.

Schauer, Frederick (2013). "On the Open Texture of Law." Grazer Philosophische Studien **87**: 197-215.

Segal, Jeffrey A and Albert D Cover (1989). "Ideological Values and the Votes of Us Supreme Court Justices." American Political Science Review **83**(2): 557-565.

Segal, Jeffrey A, Lee Epstein, Charles M Cameron and Harold J Spaeth (1995). "Ideological Values and the Votes of Us Supreme Court Justices Revisited." Journal of Politics **57**(3): 812-823.

Shukla, Abhay, Paheli Bhattacharya, Soham Poddar, Rajdeep Mukherjee, Kripabandhu Ghosh, Pawan Goyal and Saptarshi Ghosh (2022). Legal Case Document Summarization: Extractive and Abstractive Methods and Their Evaluation. arXiv preprint arXiv:2210.07544.

Trozze, Arianna, Toby Davies and Bennett Kleinberg (2024). "Large Language Models in Cryptocurrency Securities Cases: Can a Gpt Model Meaningfully Assist Lawyers?" Artificial Intelligence and Law: 1-47.

Turner, John R, Jeff Allen, Suliman Hawamdeh and Gujjula Mastanamma (2023). "The Multifaceted Sensemaking Theory: A Systematic Literature Review and Content Analysis on Sensemaking." Systems **11**(3): 145.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin (2017). Attention Is All You Need. Advances in Neural Information Processing Systems.

Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le and Denny Zhou (2022). "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models." Advances in neural information processing systems **35**: 24824-24837.

Weick, Karl E. (1995). Sensemaking in Organizations. Thousand Oaks, Sage Publications.

Xie, Huiyuan, Felix Steffek, Joana Ribeiro de Faria, Christine Carter and Jonathan Rutherford (2024). The Clc-Uket Dataset: Benchmarking Case Outcome Prediction for the Uk Employment Tribunal. arXiv preprint arXiv:2409.08098.

Zou, Xinrui, Ming Zhang, Nathaniel Weir, Benjamin Van Durme and Nils Holzenberger (2024). Reframing Tax Law Entailment as Analogical Reasoning. arXiv preprint arXiv:2401.06715.

Master for Final System Prompts

Task: Summarization of Jurisprudence Regarding a Specific Element of Doctrine

The user prompt corresponding to this system prompt consists of a list of snippets from decisions taken by the European Court of Human Rights, and the former European Commission of Human Rights. All snippets have been selected (in a previous step, also by GPT-4o) as covering one specific element of the doctrine of freedom of assembly, as guaranteed by Art. 11 of the European Convention on Human Rights. The relevant element of doctrine is listed at the beginning of the prompt document. It is, however, possible that the snippets also cover further elements of doctrine. Please only summarize the jurisprudence regarding the one element of doctrine that is mentioned on top of the prompt document.

Let me illustrate this request with an example. Assume the following: the snippet says that a governmental act falls within the substantive scope of freedom of assembly, that the applicant is protected by the human right, that the act interferes with the freedom, and that it is sufficiently accessible. Further assume that I am asking you to summarize the jurisprudence regarding the question whether an act "interferes" with freedom of assembly. In that case please do NOT explain how the court interprets the substantive and the personal scope of freedom of assembly, and the requirement that interferences must be sufficiently accessible. Please ONLY use what the snippet says about interference.

Let me explain why: at a later point, I will also ask you to write the analogous summary for the remaining elements of the doctrine of freedom of assembly. Yet for the users of the commentary that I want to write with your help, it is important that they can zero in on the element of doctrine that is critical for their individual case.

Structure of the Doctrine of Freedom of Assembly

In the interest of facilitating your task, I am informing you about the elements of the doctrine of human rights, as developed by the European Court of Human Rights.

1. Scope

The Court defines the scope of the protection in two dimensions:

a) substantive scope

Which actions are protected by the human right?

b) person scope

Which persons, groups of persons or institutions are protected by the human right?

2. interference

Does the governmental act against which the complaint is directed interfere with freedom of assembly? The Court uses this element of doctrine in particular to decide whether the act sufficiently severely impinges on the human right.

3. prescribed by law

If the governmental act interferes with freedom of assembly, the Court investigates whether the interference is justified. According to Art. 11 (2) justification first requires that the act is "prescribed by law". In its jurisprudence, the Court has split this into the following requirements:

a) accessibility

It must be sufficiently easy for the individuals covered by the legal rule that empowers government to intervene into freedom of assembly to find out in which way this rule constrains their freedom.

b) foreseeability

It must also be sufficiently foreseeable in which ways government may constrain freedom of assembly.

c) safeguards against abuse

Finally the law must sufficiently protect the applicant against the risk that government abuses the powers with which it has been endowed.

4. legitimate aim

For an interference with freedom of assembly to be justified, government must pursue a legitimate aim. Art. 11 (2) comes with a list of aims that government may pursue. To be justified, the interference must be brought under one of these aims. The list comprises:

a) national security
b) public safety
c) prevention of disorder
d) prevention of crime
e) protection of health
f) protection of morals
g) protection of the rights and freedoms of others

5. necessary in a democratic society

It is not enough that government pursues a legitimate aim. The intervention must also be motivated by a pressing social need, it must be proportionate to the legitimate aim, and the reasons given by the government must be relevant and sufficient.

6. positive obligations

The European Convention on Human Rights does not only protect individuals from state intervention. States are also obliged to actively intervene, to protect those enjoying freedom of assembly against the risk that third parties impinge on the exercise of the right.

7. just satisfaction

If the Court comes to the conclusion that government has violated freedom of assembly, it may oblige government to provide the applicant with just satisfaction.

8. separate opinion

If a judge disagrees with the way how the majority of the panel has decided the case, she may add a separate opinion.

Character of the Snippets

Each snippet is taken from a ruling by the European Court of Human Rights, or by the former European Commission of Human Rights. Cases that discuss freedom of assembly are numbered consecutively. Each snippet is taken from a case, and from one or more paragraphs of the case. Each snippet starts with <case_number>_<paragraph_number>.

The Summarization Task

Please write a text that summarizes the snippets in the prompt. Please list for every statement from which snippet you have taken it. If more than one snippet supports the statement, please list them all. Please add these references to the end of the statement, in the format

        (<case_number>_<paragraph_number>)

If applicable enclose multiple references in these brackets.

Please do order your summary by substance matter, not by case_numbers.

Please structure your summary as follows:

a) general principles

How does the Court define the element of doctrine in question? Which purpose is it meant to serve, according to the jurisprudence of the court? Does the court discuss alternative interpretations?

b) applications

For which situations of life the element of doctrine has been fulfilled, and for which it has not? Please be a specific as possible. The more the ruling has been detailed, the more report. If possible, do not just write a long list of more or less relevant cases. Such a list is much less helpful for users than the structured report of detail. For users it is important a) to learn about the facets of jurisprudence, and b) via the references to quickly access cases that have been particularly close to their current case.

What I do not want

Let me use an example to explain what I do not want: when asked about "interference", in an earlier attempt, I have received the following response:

"The European Court of Human Rights (ECtHR) defines "interference" with the right to freedom of assembly as any action by public authorities that restricts or impedes the exercise of this right. This includes measures taken before, during, or after an assembly, such as bans, dispersals, arrests, and punitive measures (1_97, 2_103, 2_239, 3_93, 4_71, 5_56, 5_87, 6_81, 6_88, 6_89, 7_81, 7_82, 9_242, 9_243, 10_2, 10_38, 10_39, 10_40, 10_41, 11_94, 12_97, 12_98, 13_86, 14_88, 15_66, 16_88, 18_99, 19_73, 22_48, 22_49, 22_50, 22_125, 23_114, 24_108, 24_109, 25_106, ..."

This is not sufficiently helpful. The first sentence essentially repeats the doctrinal element. The second sentence lumps a multitude of potential interferences together. Users do not learn anything about the specifics of the cases that the court has considered to interfere with freedom of assembly. They do not learn about the conditions under which the court has considered the governmental act NOT to interfere with freedom of intervention.

Also the summary should be much more detailed. Several hundred snippets should not just be summarized by a single paragraph (even if the paragraph is a long one).

Provisional Structure

In a first attempt, GPT has given responses to these requests, separately per batches of 100 rulings. The results however differ quite strongly from one batch to the next. In this second round, I am trying to create a more coherent picture of the jurisprudence. I exploit that, by another call to GPT, I have received the list of topics, and some illustrations describing them,

that I am reproducing in the end of this prompt. Please try to organize the new summary of each batch of data along the lines of this list of topics.

To be clear: not every topic has to feature in every summary; it could be that the Court has been more engaged with certain topics at certain points in time. It is also not unlikely that the Court has used different language over time for essentially the same, or a closely related, concern. But do not force coherence if actually there is none. With this additional step, I just want to avoid that the historically contingent composition of some batches of data makes the summary across batches of data unnecessarily incoherent.

Here is the summary of topics that GPT, for this element of the doctrine, has identified in the first round: