

Dynamic Demonstrations Selection for Few-Shot Legal Argument Mining

Francesco Alfieri¹, Giulia Grundler², Francesca Galloni², Rūta Liepiņa²,
Francesca Lagioia^{2,3,*}, Andrea Galassi^{1,*} and Paolo Torroni¹

¹Department of Computer Science and Engineering, University of Bologna, Bologna, Italy

²CIRSFID AlmaAI, University of Bologna, Bologna, Italy

³Law Department, European University Institute, Fiesole, Italy

Abstract

The selection of demonstrations for few-shot learning plays a pivotal role in the performance of LLMs. In the legal domain, selecting these examples becomes especially critical, since they must be both informative and economical. We address legal argument mining by adopting dynamic selection strategies where a specific set of demonstrations is selected for each inference, and we compare them to static approaches where examples are chosen by experts or by the LLMs themselves. We experiment with 34 learning configurations over 3 different tasks, i.e., classification of argumentative components, type of premises, and argumentative schemes. We find that dynamic selection methods are better than static ones in all three tasks, suggesting that similarity is an important criterion in this domain.

Keywords

Argument Mining, Large Language Models, In-Context Learning

1. Introduction

The ability to automatically extract and classify arguments from legal decisions holds significant promise for improving legal research and education, judicial transparency, and AI tools that support legal decision-making. Argument Mining (AM) in this domain enables structured access to Courts' reasoning, a critical asset for legal practitioners (e.g., suggesting relevant arguments and counterarguments), scholars, and AI systems dealing with a wide range of legal tasks, including the retrieval and classification of legal arguments from large corpora, the summarization of judicial decisions, and the construction of domain-specific ontologies. Recently, Large Language Models (LLMs) have dramatically advanced the field of Natural Language Processing (NLP), offering state-of-the-art performance across a wide range of tasks. Their versatility and generalization capabilities make them attractive tools for analyzing complex texts. However, the application of LLMs to the legal domain has yielded mixed results [1, 2, 3]. While some studies report strong performance on certain tasks, others highlight the difficulty of transferring general-purpose architectures to legal argumentation, particularly when nuanced legal reasoning and domain-specific language and cultural differences across jurisdictions are involved [4, 5]. One of the challenges underlying this gap lies in the nature of legal language itself, which is formal, layered, and often ambiguous. Legal reasoning frequently involves implicit knowledge, domain-specific interpretive schemes, and multi-step argumentation patterns that are difficult to capture through pre-training alone. These characteristics suggest that LLMs require carefully designed prompting strategies, particularly in few-shot settings, to perform effectively on legal argument mining tasks.

Proceedings of the First Argument Mining and Empirical Legal Research Workshop (AMELR 2025), June 20, 2025, Chicago, United States

*Corresponding author.

✉ francesco.alfieri5@studio.unibo.it (F. Alfieri); giulia.grundler2@unibo.it (G. Grundler); francesca.galloni5@unibo.it (F. Galloni); ruta.liepina@unibo.it (R. Liepiņa); francesca.lagioia@unibo.it (F. Lagioia); a.galassi@unibo.it (A. Galassi); p.torroni@unibo.it (P. Torroni)

🆔 0009-0007-2293-5682 (F. Alfieri); 0000-0002-7255-9343 (G. Grundler); 0009-0001-3727-6356 (F. Galloni); 0000-0002-2417-3219 (R. Liepiņa); 0000-0001-7083-3487 (F. Lagioia); 0000-0001-9711-7042 (A. Galassi); 0000-0002-9253-8638 (P. Torroni)



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

In few-shot learning, the provided demonstrations significantly shape the LLMs’ performance. In the legal domain, selecting these examples becomes especially critical. To be informative, they must encapsulate key reasoning patterns and linguistic cues. At the same time, they must remain concise, due to the token limitations of current models. Moreover, the multi-label nature of some tasks further complicates this process, as each example may represent multiple dimensions of legal reasoning. Consider, for instance the case of argumentation schemes, which are not mutually exclusive, i.e., a single legal premise may be assigned multiple schemes (e.g., from rule, precedent, principle).

In this paper, we address the open question of how to best select few-shot examples for LLMs, in the context of legal argument mining. We rely on a pre-existing dataset, Demosthenes [6, 7], and focus on four strategies for selecting demonstrations: (i) static selection performed by legal experts, (ii) static selection performed by LLMs, (iii) dynamic selection based on semantic similarity (k-nearest neighbors), and (iv) a novel dynamic graph-based method, accounting also for diversity in the demonstrations. Our goal is to assess how these strategies affect the performance of LLMs across different sub-tasks, as well as to provide guidance for leveraging few-shot learning in legal NLP applications.

2. Related Works

LLMs for Argument Mining. Argument mining aims to automatically detect and classify argumentative text. It has gained increasing attention in recent years, particularly within the legal domain, where legal reasoning relies heavily on well-structured arguments, which are fundamental for the decision-making process. Since LLMs can leverage in-context learning and few-shot classification, they can be particularly useful for this context, where annotated data is scarce and costly to obtain. However, the application of LLMs for argument mining shows limitations. Ruiz-Dolz et al. [8] apply LLMs to the detection of argumentative fallacies, but did not surpass the performance of Transformer models. Pan et al. [9] obtain ambivalent results on similar tasks. As concerns the legal domain, similar conclusions are drawn by Al Zubaer et al. [10]. Chen et al. [11] evaluate several LLMs on argument detection and argument generation tasks, in both zero-shot and few-shot settings, with results that are promising yet not consistently superior to those obtained with pre-trained language models. In contrast, in the work by Gorur et al. [12], Llama2-70B and Mixtral-8x7B outperform the Transformer baseline in the relation prediction task on several datasets. Additionally, Cabessa et al. [13] fine-tune LLMs, quantized and non-quantized, that achieve state-of-the-art results in classifying argument components and their relations.

In-Context Learning. In recent years, the popularity of in-context learning (ICL) as a research topic has seen a significant rise. Few-shot inference, which consists of providing LLMs with demonstration examples as part of the instruction prompt, has been proven to be an effective technique to use LLMs [14]. Similar to zero-shot inference, it allows using a general-purpose architecture without the need for training or fine-tuning. Unlike zero-shot inference, it necessitates labeled data, but the required amount is several orders of magnitude less than that needed for fine-tuning. The performance of ICL varies both with the instruction formatting and the demonstration organization, which concerns several aspects such as demonstration selection, format, and order [15, 16]. In this work, we focus on the demonstration selection aspect. Demonstrations are often randomly selected from the dataset, as in Chen et al. [11]. Alternative approaches include adopting unsupervised approaches to preliminarily select a static set of demonstrations for labelling out of a larger set of unlabeled data [17, 18], selecting representative examples consulting experts [19], or relying on LLMs themselves [20]. Choosing examples similar to the query in the embedding space has been shown to greatly improve the performance of LLMs [21]. In particular, Liu et al. [22] introduce a k-NN-based demonstration selector. Wang et al. [23] identify a contrast between the need for the demonstrations to be similar to the test instance and the need for diversity between the examples. They propose a reinforcement learning approach to demonstration selection, which aims to maximize both relevance and diversity. However, this approach computes diversity only by considering the distribution of labels among the selected demonstrations. Ye et al. [24]

Table 1
Dataset composition.

Split	Docs	Prem	Conc	Factual	Legal	Aut	Class	Itpr	Prec	Rule
Train	32	2,030	138	1,345	774	49	51	268	422	272
Test	8	345	22	231	132	4	5	29	82	51
Total	40	2,375	160	1,576	906	53	56	297	504	323

show that LLMs can benefit from example sets that exhibit both complementarity and relevance to a given test query. Here, the diversity is based on the same similarity metric used to assess the relevance, creating a trade-off between the two objectives. Other works also yield similar findings [25, 26]. In our work, we compare static and dynamic selection approaches and propose a novel graph-based method to leverage the advantages of both similarity and diversity of demonstrations.

3. Data

We rely on the Demosthenes corpus [6, 7], consisting of 40 decisions in English on Fiscal State Aids by the Court of Justice of the European Union (CJEU). We select this dataset for several reasons: (a) annotation was performed manually by legal experts; (b) CJEU rulings present a variety of legal reasoning (e.g., arguments based on statutes, legal principles and precedents); (c) decisions follow a relatively consistent structure; (d) all the selected cases belong to the same legal domain, Fiscal State Aid, which heavily depends on judicial interpretation. The argument structure is annotated at three hierarchical levels, that can be used for three classification tasks: (i) argument elements (*premises* and *conclusions*), (ii) type of premises (*legal* and *factual*), and (iii) argument schemes.

Argument Components and Types. Each document contains one or more argument chains, which are structured arguments supporting a final conclusion on a specific ground of appeal. It includes both supporting reasoning and any counterarguments considered by the Court (see B.1). Each argument is a set of connected inferences, where one or more premises lead to a conclusion. Each argument component that is part of an argumentative chain is a sentence labeled either as *premise* or *conclusion*. Conclusion of an inference can also serve as the premise for further inferences, and they are labeled as premises. Premises are categorized as either *factual* or *legal*. Factual premises describe real-world events or procedural aspects of the case, while legal premises refer to legal content such as laws, precedents, principles, or their interpretation. When a premise contains both legal and factual elements, it is labeled accordingly (see B.2)

Argument Schemes. Each legal premise is labeled with its argument scheme, based on the framework described in [27, 28] and adapted to CJEU reasoning, resulting in five schemes.¹ Since they are not mutually exclusive, a single legal premise may be assigned multiple schemes (see Appendix B.3).

The *Rule* (or established rule) scheme applies when a legislative rule applies to the case outcome, unless overridden by exceptions. It covers premises explicitly citing an EU norm as part of the legislative framework, excluding references to national laws or norms from the Court of First Instance, which cannot form the legal basis for a CJEU ruling.

The *Precedent* scheme (*Prec*) applies whenever the *ratio decidendi* of a past case applies to the current one unless a distinction is justified [29]. Premises in CJEU decisions citing the Court’s earlier rulings are annotated accordingly. Typical indicators include citations of prior judgments, along with standard expressions like “according to settled case-law”, “as is apparent from that case-law,” or “as the Court has consistently held.”

The *Authoritative* scheme (*Aut*) applies when a statement by an authority supports the case outcome,

¹The dataset includes a sixth scheme, *Princ*, which we do not consider in our work.

barring opposing reasons. It includes: (1) administrative authority, exercising command over others; (2) expert opinion, based on domain expertise; and (3) majority or common opinion [28, 30]. In our corpus, CJEU references to Advocate General opinions are annotated as authoritative inferences, as these serve as non-binding but influential sources of legal reasoning.

The *Classification* scheme (*Class*) applies when a concept is applicable to the case and used to support a classification, unless exceptions apply. Adapted from the Verbal Classification scheme in [31, 27], its acceptability depends on the acceptability of the classification and any applicable exceptions. Premises are marked under this scheme when they define a legal concept by specifying the conditions under which a fact, property, or entity qualifies as falling within it.

The *Interpretative* scheme (*Itpr*) applies when a meaning relevant to the decision of the case is ascribed to a legal source (e.g., legislation, precedent). The scheme includes different kinds of interpretative reasoning (e.g., literal, teleological, psychological, systematic interpretation).

Dataset Composition. In the original article [6], the evaluation was performed with 5-fold cross-validation, with manually created splits at the document level, in order to balance their composition. In our experiments, we use one of these folds as a test set, and the others as training sets (and validation sets if required by the models). Table 1 summarizes the composition of the dataset and the splits.

4. Method

We focus on few-shot inference for argument mining, addressing three classification tasks:

- **Argument Component Classification:** given an argumentative sentence, label it as premise (*prem*) or conclusion (*conc*)
- **Premise Type Classification:** given a premise, label it as factual (*F*), legal (*L*), or both.
- **Argument Scheme Classification:** given a legal premise, classify it as belonging to one or more argument schemes (*Aut*, *Class*, *Itpr*, *Prec*, *Rule*).

4.1. Demonstration Selection

We adopt two families of techniques for the few-shot inference: *static* selection and *dynamic* selection. In the former, the demonstrations are selected a priori, independently from the test instance, and the same examples are used for each inference. In the latter, specific examples from the training set are selected according to the test instance. Consequently, the content of the final prompt is different for each inference.

Static Expert-given Selection. In the selection strategy, we combine general and task-specific criteria. From the *general criteria* perspective, the selection is informed by the following principles: (i) clarity and completeness, favoring self-contained sentences with a clear argumentative function; (ii) length and readability, preferring concise yet informative formulations; (iii) linguistic patterns, prioritizing lexical indicators relevant for argument labeling; and (iv) diversity, to ensure variety across decisions while avoiding near-duplicates. Linguistic patterns and diversity were applied only when not in conflict with more relevant task-specific considerations. For argument component classification, we aim at reflecting the structure of argument chains, i.e., balancing initial and intermediate (supporting other) premises, as well as final conclusions. We also try to balance the factual or legal classes. For type classification, when selecting the legal premises we included one illustrative case per argument scheme. As regards factual ones, we focus on two main categories i.e., procedural aspects, reporting on actions or events in the course of litigation (e.g., filings, prior decisions) and disputed facts between parties. For argument scheme classification, we rely on the nature of the inferences and their content-specificity, as well as on the legal criteria used to assess a Fiscal State Aid (i.e., distortion of competition, economic advantage, selectivity, and state resources). Appendix B contains further details and relevant examples.

Static Self-Selection. For each task and class, we provide the model with the definitions of the task and classes, and query the LLM to choose the best examples. The examples are chosen among a subset of the training set containing at most 50 instances, including the examples selected by the experts, and training instances sampled randomly.

Dynamic k-NN Selection. In-context learning leads to better results when the examples used for the prompt are similar to the entity to be classified [32]. Hence, the first dynamic method that we use for demonstration selection is based on k-NN. First, we compute sentence embeddings for each available instance in our training set.² At inference time, given a test query, we first create its embedding. Then, we compute its dissimilarity from the training set embeddings. Finally, we select the most similar demonstrations and include them in the prompt. We have chosen to use the Euclidean distance as a measure of dissimilarity, but our methods are also compatible with other dissimilarities.

Dynamic Graph-based Selection. Another key factor that has been observed to improve the performances of in-context learning is the diversity among the retrieved demonstrations, both in terms of associated labels and the respective semantic contents [23, 24]. Therefore, we aim to improve the semantic diversity of the examples, while preserving the similarity to a given test query. We propose a novel graph-based method. First, we compute the embedding of each instance in our training set. Then, we instantiate a large graph, where each node represents an instance, and edges connect instances that are close to each other. Specifically, two instances are connected by an edge if the distance between their embeddings is smaller than a threshold hyperparameter r_a . At inference time, given an input query, we compute a subgraph that contains only nodes that are similar to the query. Specifically, we remove all nodes that represent instances whose distance from the query is greater than a threshold hyperparameter r_b . Last, to improve semantic diversity, we partition the resulting subgraph into dense communities and select a single node from each community. More in detail, we partition the subgraph via the Louvain method [33], which detects well-separated groups of examples that are similar to each other, optimizing modularity. The nodes in each community are chosen according to the PageRank centrality score, assigning larger scores to nodes that are connected to other important nodes.³ By construction, the retrieved demonstrations have the two desired properties: (i) they are *similar* to the query, thanks to the subgraph extraction, and (ii) they represent *diverse* concepts, as identified by the partitioning algorithm.

Balanced Dynamic Selection. As an attempt to further increase the diversity between demonstrations, we also experiment with a variant of the two dynamic approaches. In such a variant, the methods are constrained to include an approximately equal number of demonstrations for each class. For example, for the component classification task, the prompts always contain approximately the same number of premises and conclusions.

4.2. Models and Hyperparameters

As LLMs, we consider **Llama** and **Gemini**. Llama 3.1 [36] is a collection of multilingual generative LLMs, pretrained and instruction tuned, developed by Meta. Gemini 2.0 Flash [37] is the latest generative LLM accessible through Google’s API, with a one-million-token context window, meaning that it can process an input of 700k words.⁴

We also experiment with four other models and inference paradigms as references. We consider a **LinearSVC** classifier with TF-IDF features that we train from scratch and we fine-tune a **LEGAL-BERT** [38] model.⁵ For both classifiers, we train/fine-tune a separate model for each task. Additionally,

²We use the embedding methods associated with the LLM we use for inference.

³For undirected graphs, the PageRank distribution is statistically close but not equal to the degree distribution [34, 35].

⁴We use the following implementation of the models from HuggingFace: nlpaueb/legal-bert-base-uncased, meta-llama/Meta-Llama-3.1-8B-Instruct. For Gemini we use model gemini-2.0-flash-001.

⁵We fine-tune the whole model.

we also use the two LLMs in a zero-shot setting. Examples of prompts can be found in Appendix A.

For the Dynamic Graph method, given the wide hyperparameter space and the lack of previous research on the matter, we rely on a preliminary study to set the values for hyperparameters r_a and r_b , using different values for each task and number of shots. In particular, we define two configurations, named $c1$ and $c2$. In $c1$, r_a is chosen so that each node is connected, on average, to 1% of the other nodes. In $c2$, r_a is based on the value r_b : r_a is chosen so that each node is connected, on average, to half of the nodes inside the r_b range. The value of r_b is chosen so that, when performing inference over the training set, it yields a subgraph with average size \sqrt{kN} . N is the total number of examples in the training set and k is the number of demonstrations to be included in the prompt. This way, both subgraph extraction and node selection yield approximately the same reduction in the number of nodes.

5. Results and Discussion

As shown in Table 2, the best result is always obtained by LinearSVC or LEGAL-BERT, suggesting that specific training or fine-tuning is still crucial. As concerns the results of the LLMs, overall, dynamic k-NN is the best configuration in all three tasks, suggesting that similarity is an important criterion in this domain.

For the first task, the balanced versions seem particularly beneficial, especially with the Llama model. For each model and number of shots, at least one configuration of dynamic graph reaches equal or comparable results to k-NN, which are also very similar to the results obtained with the static expert prompts. However, the models are clearly underperforming compared to LEGAL-BERT, in particular on the *conc* class. In the second and third tasks, the results do not seem impacted as much by the balancing of classes. At least one configuration of dynamic graph performs similarly to k-NN in three out of four configurations. For the second task, the static expert selection yields results that can be considered comparable to the dynamic approaches. In contrast, for the third task, they perform consistently worse. For all three tasks, no configurations of dynamic graph between $c1$ and $c2$ is clearly superior to the other, as performance differences are consistently small and vary depending on the model and/or task.

As concerns the self-selected prompts, with Gemini they reach comparable results to the expert-selected ones. Notably, 14 of the demonstrations were chosen both by the model and by the experts, 11 of which for the Argument Scheme task. It was not possible to make the same comparison with Llama, because the model hallucinated and consistently generated new sentences, instead of choosing among the provided ones.

As for the relationship between these results and the available data, LinearSVC and LEGAL-BERT are the best-performing techniques, but they require the use of the entire training set. Dynamic few-shot techniques work well and use only 5 or 10 demonstrations in the prompts, but all the training instances are considered for selection for each inference. Static few-shot methods perform worse, but they always use the same demonstrations for each inference, even if they are selected from the whole training set or a subsample of it. It is therefore reasonable to hypothesize that reducing the amount of labeled data would have a greater negative impact on the best-performing methods. Nonetheless, we want to remark that for zero-shot and few-shot inference approaches we use the same general-purpose LLM for all the tasks. Instead, for LinearSVC and LEGAL-BERT we train or fine-tune a different model for each task.

6. Conclusion

In this work, we explore the open challenge of selecting effective few-shot examples for LLMs for legal argument mining. We compare several selection strategies and find out that the best approach is dynamic selection based on k-NN, which indicates the importance of semantic similarity in this context. This approach consistently matches or outperforms the expert selection, suggesting that comparable or improved performance can be achieved without the cost and effort associated with manual expert curation. We notice that at least one configuration of the dynamic graph method performs similarly to k-NN, encouraging further exploration of the large hyperparameter space of this method. Our baselines,

Table 2

Experimental results for the three tasks. For each task, we report the F1 scores of each class and their macro average. For each macro average, we highlight the best score overall in bold and underline the best score reached by LLMs.

Model and Configuration	Arg Component			Premise Type			Argument Scheme					
	prem	conc	Avg	F	L	Avg	Aut	Class	Itpr	Prec	Rule	Avg
<i>Other Models</i>												
LinearSVC	0.99	0.81	0.90	0.86	0.78	0.82	0.86	1.00	0.48	0.93	0.90	0.83
LEGAL-BERT	0.99	0.82	0.91	0.91	0.88	0.89	1.00	0.00	0.81	0.98	0.97	0.75
Llama 0-shot	0.96	0.29	0.63	0.80	0.57	0.68	0.89	0.00	0.23	0.86	0.62	0.52
Gemini 0-shot	0.78	0.26	0.52	0.74	0.61	0.68	0.89	0.26	0.06	0.95	0.91	0.61
<i>Llama 5-shot</i>												
Static Expert	0.91	0.44	0.67	0.80	0.60	0.70	0.67	0.00	0.22	0.85	0.68	0.48
Dynamic k-NN	0.16	0.12	0.14	0.80	0.60	0.70	0.89	0.36	0.37	0.80	0.63	0.61
Dynamic k-NN bal.	0.93	0.46	0.69	0.81	0.58	0.69	0.89	0.24	0.27	0.88	0.65	0.59
Dynamic graph c1	0.07	0.11	0.09	0.80	0.64	0.72	0.62	0.21	0.24	0.61	0.59	0.45
Dynamic graph c2	0.09	0.09	0.09	0.80	0.64	0.72	0.67	0.15	0.25	0.74	0.62	0.49
Dynamic graph c1 bal.	0.90	0.34	0.62	0.81	0.61	0.71	1.00	0.28	0.24	0.88	0.65	0.61
Dynamic graph c2 bal.	0.92	0.44	0.68	0.81	0.59	0.70	1.00	0.26	0.28	0.84	0.66	0.61
<i>Llama 10-shot</i>												
Static Expert	0.95	0.53	<u>0.74</u>	0.80	0.60	0.70	0.89	0.19	0.21	0.80	0.67	0.55
Dynamic k-NN	0.16	0.09	0.13	0.81	0.62	0.72	0.80	0.44	0.34	0.79	0.65	0.61
Dynamic k-NN bal.	0.95	0.54	<u>0.74</u>	0.81	0.59	0.70	1.00	0.29	0.25	0.86	0.64	0.61
Dynamic graph c1	0.09	0.10	0.10	0.81	0.61	0.71	0.73	0.29	0.30	0.82	0.58	0.54
Dynamic graph c2	0.11	0.09	0.10	0.81	0.58	0.70	0.73	0.22	0.24	0.70	0.62	0.50
Dynamic graph c1 bal.	0.94	0.51	0.73	0.81	0.60	0.70	1.00	0.30	0.24	0.84	0.69	0.61
Dynamic graph c2 bal.	0.94	0.49	0.72	0.81	0.56	0.68	1.00	0.30	0.15	0.83	0.67	0.59
<i>Gemini 5-shot</i>												
Static Expert	0.88	0.37	0.62	0.80	0.71	0.76	1.00	0.32	0.47	0.93	0.69	0.68
Static Self	0.86	0.34	0.60	0.82	0.77	0.80	1.00	0.36	0.42	0.94	0.70	0.68
Dynamic k-NN	0.85	0.33	0.59	0.91	0.74	0.82	1.00	0.55	0.57	0.95	0.95	<u>0.80</u>
Dynamic k-NN bal.	0.86	0.37	0.63	0.86	0.71	0.78	1.00	0.43	0.51	0.96	0.94	0.77
Dynamic graph c1	0.82	0.29	0.56	0.80	0.63	0.72	1.00	0.33	0.41	0.95	0.91	0.72
Dynamic graph c2	0.84	0.32	0.58	0.77	0.70	0.74	1.00	0.45	0.45	0.94	0.87	0.74
Dynamic graph c1 bal.	0.88	0.37	0.63	0.84	0.63	0.74	1.00	0.31	0.30	0.94	0.88	0.69
Dynamic graph c2 bal.	0.88	0.37	0.63	0.87	0.69	0.78	1.00	0.38	0.31	0.96	0.91	0.71
<i>Gemini 10-shot</i>												
Static Expert	0.92	0.46	0.69	0.84	0.77	0.81	1.00	0.20	0.48	0.97	0.87	0.71
Static Self	0.91	0.43	0.67	0.81	0.78	0.80	1.00	0.40	0.55	0.95	0.84	0.75
Dynamic k-NN	0.91	0.40	0.65	0.90	0.79	<u>0.84</u>	1.00	0.59	0.52	0.94	0.95	<u>0.80</u>
Dynamic k-NN bal.	0.91	0.43	0.67	0.86	0.77	0.82	1.00	0.53	0.58	0.95	0.94	<u>0.80</u>
Dynamic graph c1	0.90	0.40	0.65	0.81	0.77	0.79	1.00	0.53	0.56	0.96	0.90	0.79
Dynamic graph c2	0.91	0.42	0.67	0.84	0.83	0.83	1.00	0.40	0.51	0.94	0.86	0.74
Dynamic graph c1 bal.	0.91	0.44	0.68	0.90	0.75	0.82	1.00	0.36	0.54	0.97	0.80	0.74
Dynamic graph c2 bal.	0.90	0.40	0.65	0.84	0.73	0.78	1.00	0.32	0.54	0.97	0.87	0.74

LinearSVC and LEGAL-BERT, obtain the best results in all three tasks, suggesting that training or fine-tuning is still crucial in this domain.

In future work, we want to investigate how reducing the amount of available data impacts performances. We also want to evaluate the robustness of our approaches, considering the impact of prompt format and demonstration order.

Acknowledgments

This work was partially supported by the following projects: CompuLaw - Computable Law - funded by the ERC under the Horizon 2020 (Grant Agreement N. 833647); PRIN2022 PRIMA - PRivacy Infringements Machine-Advice (Ref. Prot. n.: 20224TPEYC - CUP J53D23005130001); PRIN2022 EQUAL - EQUitableALgorithms (Ref. Prot n. 2022KFLF3E_001 - CUP J53D23005560001); CLAUDETTE IV, funded by the EUI Research Council for funding; “FAIR - Future Artificial Intelligence Research” – Spoke 8 “Pervasive AI”, under the European Commission’s NextGeneration EU programme, PNRR – M4C2 – Investimento 1.3, Partenariato Esteso (PE00000013).

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: Grammar and spelling check. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication’s content.

References

- [1] A. Blair-Stanek, N. Holzenberger, B. V. Durme, Can GPT-3 perform statutory reasoning?, in: ICAIL, ACM, 2023, pp. 22–31.
- [2] E. Martínez, Re-evaluating gpt-4’s bar exam performance, *Artificial Intelligence and Law* (2024).
- [3] M. Panarelli, A. Galassi, F. Lagioia, R. Liepiņa, M. Lippi, P. Palka, G. Sartor, Is it worth using llms for unfair clause detection in terms of service?, in: ICAIL, ACM, 2025.
- [4] I. Habernal, D. Faber, N. Recchia, S. Bretthauer, I. Gurevych, I. S. genannt Döhmann, C. Burchard, Mining legal arguments in court decisions, *Artif. Intell. Law* 32 (2024) 1–38.
- [5] P. Poudyal, J. Savelka, A. Ieven, M. F. Moens, T. Goncalves, P. Quaresma, ECHR: Legal corpus for argument mining, in: ArgMining@COLING, Association for Computational Linguistics, Online, 2020, pp. 67–75. URL: <https://aclanthology.org/2020.argmining-1.8>.
- [6] G. Grundler, P. Santin, A. Galassi, F. Galli, F. Godano, F. Lagioia, E. Palmieri, F. Ruggeri, G. Sartor, P. Torroni, Detecting arguments in CJEU decisions on fiscal state aid, in: G. Lapesa, J. Schneider, Y. Jo, S. Saha (Eds.), *Proceedings of the 9th Workshop on Argument Mining, ArgMining@COLING 2022*, Online and in Gyeongju, Republic of Korea, October 12 - 17, 2022, International Conference on Computational Linguistics, 2022, pp. 143–157. URL: <https://aclanthology.org/2022.argmining-1.14>.
- [7] P. Santin, G. Grundler, A. Galassi, F. Galli, F. Lagioia, E. Palmieri, F. Ruggeri, G. Sartor, P. Torroni, Argumentation structure prediction in CJEU decisions on fiscal state aid, in: ICAIL, ACM, 2023, pp. 247–256.
- [8] R. Ruiz-Dolz, J. Lawrence, Detecting argumentative fallacies in the wild: Problems and limitations of large language models, in: M. Alshomary, C.-C. Chen, S. Muresan, J. Park, J. Romberg (Eds.), *Proceedings of the 10th Workshop on Argument Mining, Association for Computational Linguistics, Singapore, 2023*, pp. 1–10. URL: <https://aclanthology.org/2023.argmining-1.1/>. doi:10.18653/v1/2023.argmining-1.1.
- [9] F. Pan, X. Wu, Z. Li, A. T. Luu, Are llms good zero-shot fallacy classifiers?, in: EMNLP, Association for Computational Linguistics, 2024, pp. 14338–14364.
- [10] A. Al Zubaer, M. Granitzer, J. Mitrović, Performance analysis of large language models in the domain of legal argument mining, *Frontiers in Artificial Intelligence Volume 6 - 2023* (2023). URL: <https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2023.1278796>. doi:10.3389/frai.2023.1278796.
- [11] G. Chen, L. Cheng, A. T. Luu, L. Bing, Exploring the potential of large language models in computational argumentation, in: L.-W. Ku, A. Martins, V. Srikumar (Eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long*

- Papers), Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 2309–2330. URL: <https://aclanthology.org/2024.acl-long.126/>. doi:10.18653/v1/2024.acl-long.126.
- [12] D. Gorur, A. Rago, F. Toni, Can large language models perform relation-based argument mining?, in: O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. D. Eugenio, S. Schockaert (Eds.), Proceedings of the 31st International Conference on Computational Linguistics, Association for Computational Linguistics, Abu Dhabi, UAE, 2025, pp. 8518–8534. URL: <https://aclanthology.org/2025.coling-main.569/>.
 - [13] J. Cabessa, H. Hernault, U. Mushtaq, Argument mining with fine-tuned large language models, in: O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. D. Eugenio, S. Schockaert (Eds.), Proceedings of the 31st International Conference on Computational Linguistics, Association for Computational Linguistics, Abu Dhabi, UAE, 2025, pp. 6624–6635. URL: <https://aclanthology.org/2025.coling-main.442/>.
 - [14] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language models are few-shot learners, in: H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin (Eds.), Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6–12, 2020, virtual, 2020. URL: <https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfc4967418bfb8ac142f64a-Abstract.html>.
 - [15] Z. Zhao, E. Wallace, S. Feng, D. Klein, S. Singh, Calibrate before use: Improving few-shot performance of language models, in: ICML, volume 139 of *Proceedings of Machine Learning Research*, PMLR, 2021, pp. 12697–12706.
 - [16] Y. Lu, M. Bartolo, A. Moore, S. Riedel, P. Stenetorp, Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity, in: ACL (1), Association for Computational Linguistics, 2022, pp. 8086–8098.
 - [17] E. Chang, X. Shen, H.-S. Yeh, V. Demberg, On training instance selection for few-shot neural text generation, in: C. Zong, F. Xia, W. Li, R. Navigli (Eds.), Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Association for Computational Linguistics, Online, 2021, pp. 8–13. URL: <https://aclanthology.org/2021.acl-short.2/>. doi:10.18653/v1/2021.acl-short.2.
 - [18] H. Su, J. Kasai, C. H. Wu, W. Shi, T. Wang, J. Xin, R. Zhang, M. Ostendorf, L. Zettlemoyer, N. A. Smith, T. Yu, Selective annotation makes language models better few-shot learners, in: ICLR, OpenReview.net, 2023.
 - [19] G. Grundler, A. Galassi, P. Santin, A. Fidelangeli, F. Galli, E. Palmieri, F. Lagioia, G. Sartor, P. Torroni, AMELIA - argument mining evaluation on legal documents in italian: A CALAMITA challenge, in: CLiC-it, volume 3878 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024.
 - [20] C. Qin, A. Zhang, C. Chen, A. Dagar, W. Ye, In-context learning with iterative demonstration selection, in: EMNLP (Findings), Association for Computational Linguistics, 2024, pp. 7441–7455.
 - [21] J. D’Abramo, A. Zugarini, P. Torroni, Dynamic few-shot learning for knowledge graph question answering, CoRR abs/2407.01409 (2024).
 - [22] J. Liu, D. Shen, Y. Zhang, B. Dolan, L. Carin, W. Chen, What makes good in-context examples for gpt-3?, in: E. Agirre, M. Apidianaki, I. Vulic (Eds.), Proceedings of Deep Learning Inside Out: The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures, DeeLIO@ACL 2022, Dublin, Ireland and Online, May 27, 2022, Association for Computational Linguistics, 2022, pp. 100–114. URL: <https://doi.org/10.18653/v1/2022.deelio-1.10>. doi:10.18653/v1/2022.DEELIO-1.10.
 - [23] X. Wang, J. Wu, Y. Yuan, M. Li, D. Cai, W. Jia, Demonstration selection for in-context learning via reinforcement learning, in: ICML, 2025.
 - [24] X. Ye, S. Iyer, A. Celikyilmaz, V. Stoyanov, G. Durrett, R. Pasunuru, Complementary explanations for effective in-context learning, in: A. Rogers, J. L. Boyd-Graber, N. Okazaki (Eds.), Findings

- of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023, Association for Computational Linguistics, 2023, pp. 4469–4484. URL: <https://doi.org/10.18653/v1/2023.findings-acl.273>. doi:10.18653/v1/2023.FINDINGS-ACL.273.
- [25] I. Levy, B. Bogin, J. Berant, Diverse demonstrations improve in-context compositional generalization, in: A. Rogers, J. L. Boyd-Graber, N. Okazaki (Eds.), Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023, Association for Computational Linguistics, 2023, pp. 1401–1422. URL: <https://doi.org/10.18653/v1/2023.acl-long.78>. doi:10.18653/v1/2023.ACL-LONG.78.
- [26] S. An, Z. Lin, Q. Fu, B. Chen, N. Zheng, J. Lou, D. Zhang, How do in-context examples affect compositional generalization?, in: A. Rogers, J. L. Boyd-Graber, N. Okazaki (Eds.), Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023, Association for Computational Linguistics, 2023, pp. 11027–11052. URL: <https://doi.org/10.18653/v1/2023.acl-long.618>. doi:10.18653/v1/2023.ACL-LONG.618.
- [27] D. Walton, C. Reed, F. Macagno, Argumentation Schemes, Cambridge University Press, 2008. URL: <http://www.cambridge.org/us/academic/subjects/philosophy/logic/argumentation-schemes>.
- [28] D. Walton, F. Macagno, G. Sartor, Statutory interpretation: Pragmatics and argumentation, Cambridge University Press, 2021. doi:10.1017/9781108554572.
- [29] K. Langenbucher, Argument by analogy in european law, The Cambridge Law Journal 57 (1998) 481–521. doi:10.1017/S0008197398003031.
- [30] D. Walton, M. Koszowy, Two kinds of arguments from authority in the ad verecundiam fallacy, in: Proceedings of the 8th Conference of the International Society for the Study of Argumentation, 2015, pp. 1483–1492. URL: <https://scholar.uwindsor.ca/crrarpub/17/>.
- [31] F. Macagno, D. Walton, Classifying the patterns of natural arguments, Philosophy & Rhetoric 48 (2015) 26–53. doi:10.2139/SSRN.2577387.
- [32] J. Liu, D. Shen, Y. Zhang, B. Dolan, L. Carin, W. Chen, What makes good in-context examples for GPT-3?, in: E. Agirre, M. Apidianaki, I. Vulić (Eds.), Proceedings of Deep Learning Inside Out (DeeLIO 2022): The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures, Association for Computational Linguistics, Dublin, Ireland and Online, 2022, pp. 100–114. URL: <https://aclanthology.org/2022.deelio-1.10/>. doi:10.18653/v1/2022.deelio-1.10.
- [33] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment 2008 (2008) P10008. URL: <http://dx.doi.org/10.1088/1742-5468/2008/10/P10008>. doi:10.1088/1742-5468/2008/10/p10008.
- [34] V. Grolmusz, A note on the pagerank of undirected graphs, Inf. Process. Lett. 115 (2015) 633–634. URL: <https://doi.org/10.1016/j.ipl.2015.02.015>. doi:10.1016/J.IPL.2015.02.015.
- [35] D. F. Gleich, Pagerank beyond the web, SIAM Rev. 57 (2015) 321–363. URL: <https://doi.org/10.1137/140976649>. doi:10.1137/140976649.
- [36] LlamaTeam, The llama 3 herd of models, 2024. URL: <https://arxiv.org/abs/2407.21783>. arXiv:2407.21783.
- [37] GeminiTeam, Gemini: A family of highly capable multimodal models, 2024. URL: <https://arxiv.org/abs/2312.11805>. arXiv:2312.11805.
- [38] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, I. Androutsopoulos, LEGAL-BERT: The muppets straight out of law school, in: Findings of the Association for Computational Linguistics: EMNLP 2020, Association for Computational Linguistics, Online, 2020, pp. 2898–2904. doi:10.18653/v1/2020.findings-emnlp.261.

Appendix

A. Definitions (Zero-Shot)

Argument component.

Classify the following argumentative text as premise `prem` or conclusion `conc`. A premise (prem) is a proposition that provides a reason or support for the argument. A conclusion (conc) is the statement that follows logically from the premise(s) and represents the final point being argued for. Only reply with `prem` or `conc`.

Premise Type.

Classify the following premise as factual `F`, legal `L` or both. Factual premises (F) describe factual situations and events, pertaining to the substance or the procedure of the case. Legal premises (L) specify the legal content (legal rules, precedents, interpretation of applicable laws and principles). The expected output is a list with all applicable labels. For example: [`F`, `L`]. Only reply with the list of labels.

Argument Scheme.

Classify the following legal premise as one or more of the following argumentative schemes: Rule, Prec, Class, Itpr, Aut. Rule: whether there is an explicit or implicit reference to an article of law or citation of the text of a certain article. Prec: whether there is a reference to a previous ruling of the Supreme Court or the Court of Justice of the European Union. Class: if there is a definition of a legal concept or its constituent elements. Itpr: if there is reference to one of the interpretative criteria contained in Article 12 of the prelegislations (literal, teleological, psychological, systematic) to the Civil Code. Aut: if there is a reference to an indication by an authority (e.g., an opinion of the Advocate General). The expected output is a list with all applicable labels. For example: [`Prec`, `Aut`, `Rule`]. Only reply with the list of labels.

B. Detailed Examples (Few-Shot)

B.1. Argument components

As noted in Section 3 and 4, by an argument, we mean a set of connected inferences. In the selection, we aimed at reflecting the structure of argument chains, i.e., initial and intermediate (supporting other) premises, as well as final conclusions.

The following is an example of a starting (factual) premise:

“As a preliminary point, it must be recalled – as held earlier in paragraph 39 of the present judgment – that the Court of First Instance did not commit an error of law in finding that the aid covered by the contested decision fell within the scope of the prohibition laid down in Article 4(c) CS.” (Case A2008 Commission of the European Communities v Salzgitter)

Rationale: the statement is self-contained and of medium length. It refers to a legal rule, i.e., to a relevant legal aspect of the case, routinely characterizing premises rather than conclusions. It presents a typical linguistic indicator of starting premises introducing a new argument, i.e., “As a preliminary point”.

The following is an example of an intermediate (legal) premise, i.e., a conclusion of an inference serving as a premise for further inferences:

Also, it is clear from consistent case-law that Articles 4 CS and 67 CS concern two distinct areas, the first abolishing and prohibiting certain actions by Member States in the field which the ECSC Treaty places under Community jurisdiction, the second intended to prevent the distortion of competition which exercise of the residual powers of the Member States inevitably entails (see Banks, paragraph 88, and the case-law cited there) (A2008 Commission of the European Communities v Salzgitter AG)

Rationale: The sentence is self-contained and of medium length. It opens with “it is clear from consistent case-law”, a standard CJEU formulation to reaffirm settled interpretations, thereby invoking argument from precedent. It provides a legal interpretation distinguishing the scopes of Articles 4 CS and 67 CS – prohibited State actions versus competition distortion under residual powers– and cites established jurisprudence (Banks, paragraph 88) to reinforce its reasoning. Although it expresses a conclusion drawn from case-law, it does not close the argument but instead sets up a further inference, likely about legal classification or compatibility. As such, it functions as an intermediate legal premise within a broader chain of judicial reasoning.

The following is an example of a conclusion:

“Since the Court of First Instance interpreted Albany wrongly, and consequently failed to address the appellant’s argument relating to its competitive position in relation to other trade unions in the negotiation of collective agreements for seafarers, the order under appeal must accordingly be set aside on this point.” (Case A2009 3F v Commission of the European Communities)

Rationale: the statement is self-contained and of medium length. It is structurally coherent, linguistically explicit, and rhetorically decisive, functioning as the final step in a legal reasoning chain. It follows a premise–reason–conclusion pattern, including clear discourse markers: “Since” introduces the underlying justification (misinterpretation of Albany and failure to consider the appellant’s argument), while “accordingly” signals the conclusive remedial action. The conclusion is concrete, linking the legal misinterpretation to a specific outcome, i.e., “the order under appeal must accordingly be set aside on this point”, thereby reinforcing its classification as a conclusion.

B.2. Type of Premise

As pointed in Section 3 a distinction was made between factual and legal premises. Factual premises describe situations or events related to the substance or procedure of the case, while legal premises articulate legal content, including rules, precedents, and interpretations of applicable laws and principles. When a premise contains both legal and factual elements, it was classified accordingly.

The following is an example of a(n) (intermediate) legal premise:

“Also, it is clear from consistent case-law that Articles 4 CS and 67 CS concern two distinct areas, the first abolishing and prohibiting certain actions by Member States in the field which the ECSC Treaty places under Community jurisdiction, the second intended to prevent the distortion of competition which exercise of the residual powers of the Member States inevitably entails (see Banks, paragraph 88, and the case-law cited there).” (Case A2008 Commission of the European Communities v Salzgitter AG)

Rationale: the statement is self-contained and of medium length. It clearly functions as a legal premise, offering a legal interpretation grounded in consistent case-law, which clarifies the distinct purposes of Articles 4 CS and 67 CS. The reference to prior jurisprudence (Banks, paragraph 88) and treaty provisions signals a reasoning step that informs later conclusions. The phrase “it is clear from consistent case-law” is a standard CJEU formulation used to introduce or reaffirm established interpretations. By combining legal rule, precedent, and interpretation, the sentence reflects a typical argumentative pattern in the Court’s reasoning.

The following is an example of a (starting) legal premise:

“It must be held, first, that while Article 4(c) CS prohibits the granting of State aid to steel and coal undertakings, without drawing a distinction between individual aid or aid disbursed under a State aid scheme, Article 67 CS refers expressly to State aid only in respect of protective measures that the Commission may authorise, pursuant to the first indent of paragraph 2 of that article, in favour of coal or steel undertakings where they suffer competitive disadvantages because of general economic policy measures.” (Case A2008 Commission of the European Communities v Salzgitter AG)

Rationale: the statement is self-contained, of medium length, effectively illustrating a legal premise. It begins with the formal expression “It must be held, first, that...”, a conventional marker to introduce legal interpretation or establish foundational reasoning. The content provides a comparative analysis

of Article 4(c) CS and Article 67 CS, clarifying their respective scopes – general prohibition of State aid versus authorized exceptions. The reference to Treaty provisions and the interpretive nature of the analyses, coupled with the absence of remedial or conclusive language, supports its role as an early-stage premise in legal reasoning.

The following is an example of a factual premise:

“In paragraph 37 of the order under appeal, the Court of First Instance found that the appellant could not be regarded as individually concerned merely because the aid in question was passed on to its recipients by means of a reduction in the wage claims of the seafarers enjoying the exemption from income tax introduced by the fiscal measures at issue.” (Case A2009 3F v Commission of the European Communities)

Rationale: The statement is self-contained and of moderate length, serving as a clear example of a factual premise. It reports a specific finding by the Court of First Instance – “In paragraph 37 of the order under appeal, the Court of First Instance found” – a standard formulation for referencing procedural determinations. The content presents a factual assessment of how the aid was transferred and its effect on wage claims, relevant to the appellant’s legal standing. The reference to “individually concerned” and the absence of evaluative or conclusive language confirm its role in establishing the factual basis for subsequent legal reasoning, rather than stating a normative conclusion.

The following is an example of a factual premise:

“As regards in particular the members of the trade union, the Court of First Instance found that, since they appeared to be persons falling within the definition of a worker within the meaning of Article 39 EC, they were not themselves undertakings.” (Case A2009 3F v Commission of the European Communities)

Rationale: The premise is self-contained and of relatively short length. It reports a factual finding by the Court of First Instance on the classification of trade union members as workers rather than undertakings. The introductory statement “the Court of First Instance found that ...” frames it as a procedural observation grounded in case-specific context. It provides a factual basis relevant to assessing standing or eligibility, without presenting a normative conclusion, and serves to support subsequent legal reasoning.

The following is an example of an (intermediate) legal-factual premise:

“As regards the Court of First Instance’s finding that the adoption of the Second and the Third Steel Aid Code led to ambiguity as to whether subsequent application of the ZRFG had to be notified as a ‘plan’ within the meaning of Article 6 of that third code, it must, first, be held that that article expressly provides that there is an obligation to inform the Commission of plans to grant aid to the steel industry under schemes on which it has already taken a decision under the EC Treaty.” (Case A2008 Commission of the European Communities v Salzgitter AG)

Rationale: The premise is self-contained and of medium length. It opens with “As regards the Court of First Instance’s finding ...”, a commonly used formulation to recall the history of a case, followed by a legal interpretation signaled by “it must, first, be held that...” This transition from factual observation to normative interpretation under Article 6 of the Third Steel Aid Code reflects the dual structure characteristic of legal-factual premises.

B.3. Argument schemes

The following is an example of an argument from *Precedent*:

“With regard to the requirement that competition should be distorted, it must be borne in mind in that regard that, in principle, aid intended to release an undertaking from costs which it would normally have to bear in its day-to-day management or normal activities distorts the conditions of competition (judgment of 30 April 2009, Commission v Italy and Wam, C-494/06 P, EU:C:2009:272, paragraph 54 and the case-law cited)” (R2016 Orange v European Commission)

Rationale: the statement is self-contained and of medium length, functioning as a legal premise grounded in precedent. It introduces the issue of whether competition is distorted through a general legal condition, and then reinforces the normative assertion with an explicit reference to prior case-law. The use of “it must be borne in mind” signals a restatement, while the citation of a specific judgment (Commission

v Italy and Wam, C-494/06 P) and the reference to the case-law cited clearly anchor the reasoning in established jurisprudence.

The following is an example of an argument from *Rule*:

“Article 92 of the Treaty states: Save as otherwise provided in this Treaty, any aid granted by a Member State or through State resources in any form whatsoever which distorts or threatens to distort competition by favouring certain undertakings or the production of certain goods shall, in so far as it affects trade between Member States, be incompatible with the common market.” (R1997 Tiercé Ladbroke SA v Commission of the European Communities) ”

Rationale: the statement is self-contained and of moderate length. It consists of a direct quotation from Article 92 of the Treaty, restating the general legal standard governing the compatibility of Fiscal State Aid with the common market. The use of this treaty provision as the basis for subsequent legal reasoning draws its normative force from the explicit content of a binding legal text. This type of premise typically establishes the applicable legal framework, which is then interpreted or applied to the facts of the case. By invoking the rule in its full textual form, the sentence serves as a foundational step in deductive legal argumentation.

The following is an example of an argument from interpretation:

“As was noted in paragraph 73 above, only selective advantages, and not advantages resulting from a general measure applicable without distinction to all economic operators, fall within the concept of State aid.” (A2011_European Commission (C-106_09 P) and Kingdom of Spain (C-107_09 P) v Government of Gibraltar and United Kingdom of Great Britain and Northern Ireland)

Rationale: The statement is self-contained and of short length. It begins with a referential phrase – “As was noted in paragraph 73 above” – which anchors the reasoning in an earlier step. The premise clarifies the meaning of a key legal concept, namely “State aid”, by distinguishing between selective advantages and general measures. This definitional refinement illustrates an argument from interpretation, where the legal effect of a legal provision depends on how its terms are construed in context. Rather than quoting a rule or invoking precedent directly, the sentence draws on internal interpretive reasoning to specify the scope and application of an existing legal standard. It thereby functions as a premise that shapes the analytical boundaries for assessing whether a measure qualifies as aid.

The following is an example of an argument from *Authority*:

Accordingly, first, as the Advocate General observed in point 86 of his Opinion, the judgment under appeal sets out clearly the reasons why the General Court rejected Orange’s claims. (R2016_Orange v European Commission)

Rationale: the statement is self-contained and of short length, functioning as a premise grounded in authoritative endorsement. It explicitly refers to the observations made by the Advocate General – “as the Advocate General observed ...” – which are used to reinforce the validity of the Court’s reasoning. Such opinions can be considered as authoritative sources of knowledge on which the Court relies, even though they are not legally binding. The argument does not offer new factual or legal reasoning, but rather strengthens the justification for the conclusion by referencing an external source regarded as knowledgeable and reliable.

The following is an example of an argument from *Classification*:

Any activity consisting in offering services on a given market, that is, services normally provided for remuneration, is an economic activity. (A2018_Scuola Elementare Maria Montessori Srl v European Commission)

Rationale: the statement is self-contained and of short length. It establishes a definitional boundary by linking a general category – economic activity – to a set of specific conditions: the offering of services on a given market for remuneration. The argument classifies a particular kind of conduct (service provision) within a legal category. Indeed, in argument from classification, the legal consequence depends on placing a fact pattern within a broader conceptual category, thus determining the applicability of legal regimes.

The following is an example of a multi-scheme argument from *Classification, Precedent* and *Rule*:

“First, it must be recalled that, according to the Court’s settled case-law, classification of a national measure as ‘State aid’, within the meaning of Article 107(1) TFEU, requires all the following conditions to be fulfilled ...” (Case A2016_European Commission v World Duty Free)

Rationale: the statement is self-contained and, despite its short length, exemplifies a composite legal argument drawing simultaneously on rule, precedent, and classification. The starting “it must be recalled that according to the Court’s settled case-law” signals reliance on established jurisprudence, invoking argument from precedent. The reference to Article 107(1) TFEU introduces a rule-based argument, linking legal outcomes to specific normative criteria. Simultaneously, the statement engages in classification, as it outlines the definitional conditions under which a national measure qualifies as State aid. This combination of doctrinal restatement, rule application, and conceptual framing gives the sentence a foundational and multi-layered argumentative function.

The following is an example of an argument from *Authority* and *Interpretation*:

“It follows, as the Advocate General observed, in essence, in point 62 of his Opinion, that recovery of such aid entails the restitution of the advantage procured by the aid for the recipient, not the restitution of any economic benefit the recipient may have enjoyed as a result of exploiting the advantage.” (Case A2016_European Commission v Aer Lingus Ltd and Ryanair Designated Activity Company)

Rationale: the statement is self-contained and, despite its short length, exemplifies a composite argument drawing on both authority and interpretation. It opens with “It follows”, introducing an inference, and cites the “Advocate General in point 62 of his Opinion”, invoking argument from authority. The substantive content offers a clarifying interpretation: recovery concerns the restitution of the advantage, not any resulting economic benefit. This reflects argument from interpretation, as it defines the legal scope of aid recovery. Together, the citation of authoritative opinion and the interpretive refinement of a legal effect give the sentence dual argumentative weight.