**ICAIL 2025 Evaluation of the guiding principles**

# **Objectives:**

* High quality of evaluation while avoiding the reading of judgments
* 8 evaluation classes 🡪 may be reduced, if no meaningful results are expected
* If possible, transferability to international situations; for this purpose, evaluation as independent as possible of special features of the guidelines of the Federal Court of Justice?

# **Quality of the Guiding Principles?**

## What is important for a good (generated) guiding principle?

* Language
* Completeness
* Conciseness
* Correctness of content

## Guidelines from Beck:

* https://zitierportal.beck.de/documents/RedRL\_CHBeck\_Leits%C3%A4tze\_1\_8\_2021.pdf
* At least one guiding principle, in principle no more than six guiding principles; no more than six lines of 80 - 85 characters each.
* Reproduction of the main legal issues decided by the court, including the specific result of the decision. Each legal question in a separate guiding principle. Mere catchwords are not sufficient.
* Embedding the guiding principles in the overall context of the case law. Reference to previous case law in parentheses before the point.

# **Evaluation scale / Evaluation classes**

## Old classes:

|  |  |
| --- | --- |
| Class | Description |
| 1 | The guiding principles fulfil the minimum criterion of an intelligible result in German language |
| 2 | The guiding principles contain no error regarding the rules of German grammar |
| 3 | The guiding principles contain no error regarding the use of professional legal language |
| 4 | The guiding principles only refer to provisions if the original guiding principles include a reference, too |
| 5 | The guiding principles contain all aspects of the original guiding principles |
| 6 | The guiding principles contain no further information compared to the original guiding principle |
| 7 | The guiding principles reflect the core aspects of the original guiding principles |
| 8 | The guiding principles are superior to the original guiding principles |

## New classes:

|  |  |
| --- | --- |
| Classe | Description |
| 1 | Intelligible result (output control). |
| 2 | Correct use of German professional legal language (Language). |
| 3 | No additional, unnecessary information (Pertinence). |
| 4 | Inclusion of every aspect (Completeness) |
| 5 | Inclusion of the majority of aspects (Completeness – control) |
| 6 | No error in legal reasoning (Correctness) |
| 7 | Superior compared to the original (Subjective assessment) |

## Reasons for the amendment and guidelines:

|  |  |
| --- | --- |
| Classe | Description |
| 1 | Class 1 is retained: It can be assumed that the fine-tuned models will for the most part generate meaningful content. Nevertheless, this class is required to be able to mark incorrect output.  **Class 1 is fulfilled if a readable result has been generated. The output language is irrelevant.** |
| 2 | Classes 2 and 3 are combined, as there were hardly any differences during the evaluation.  **Class 2 is fulfilled if the legal terms are used correctly and a result is generated in German that is grammatically correct.** |
| 3 | The new class 3 includes the old class 6 and has been moved forward. What is important is the addition of “unnecessary” information, which is intended to make the assessment more flexible. Previously, strictly speaking, the addition of one word was enough to fail the class.  **Class 3 is fulfilled if the guiding principle does not contain any information that is unnecessary for the guiding principle. Individual, additional words or the additional citation of relevant regulations/judgments/etc. are generally harmless. Logically, there is room for judgment here, but the evaluation of three different lawyers will lead to a clear result by majority decision. In borderline cases, it is advisable to decide on the basis of first impressions whether information is considered “unnecessary”.**  **If particular patterns can be recognized in the unnecessary information, this can be briefly pointed out in a note. (Example: multiple repetition of statements). However, comments should be handled as restrictively as possible, as manual evaluation will be necessary. The comments are intended to ensure that important findings - which are not comprehensible through the evaluation classes - can be taken into account in the subsequent text of the contribution.** |
| 4 | Class 4 corresponds to the old class 5 and is intended to show whether the generated guiding principle is complete. Completeness is checked on the basis of a comparison with the original guiding principle.  **Class 4 is fulfilled if the generated guiding principle deals with every relevant legal aspect that the original guiding principle also deals with. It is irrelevant whether the legal issues are dealt with correctly or whether the legal statements are complete with regard to the individual aspect.**  **It is not necessary for the generated guiding principle to match the exact wording of the original guiding principle, as long as the underlying meaning remains unchanged. However, the exact wording would be required if it represents the main subject of the original guiding principle. An example would be the introduction of a legal concept with a proper name that is not specified in the generated guiding principle.** |
| 5 | Class 5 corresponds to the old Class 7 and is essentially a control class in order to better express the quality of the guiding principles. It is also only checked by means of a comparison (see Class 4).  **Class 5 is fulfilled if the generated guiding principle deals with at least the majority of the legal aspects that the original guiding principle also deals with. Class 5 is fulfilled if at least 3/4 of the legal aspects are dealt with. Logically, Class 5 is also fulfilled if Class 4 has already been fulfilled.** |
| 6 | Class 6 is new. The fine-tuned models may hallucinate or confuse the parties. In principle, the substantive correctness should only be determined on the basis of a comparison with the original guiding principles. Only if it is not possible to determine the accuracy of the content in this form should the judgment be used to determine whether the content is actually correct.  **Class 6 is fulfilled if the generated guiding principle reproduces the content of the original guiding principle without errors. Aspects that are contained in the generated guiding principle but not in the original guiding principle should not be examined, as these cannot be checked without looking at the judgment. Class 6 is also independent of the fulfillment of Classes 4 and 5. Only those aspects of the original guiding principle that are actually contained in the generated guiding principle are examined. However, if the generated guiding principle does not contain any aspect of the original guiding principle, class 6 is not fulfilled due to a lack of comparative subject matter.**  **If an obvious and serious error is identified in the additional information of the generated guiding principle, this can be pointed out in a note if it could be relevant for the subsequent discussion in the article.** |
| 7 | Class 7 remains unchanged in order to highlight particularly good guiding principles.  **Class 7 is fulfilled if the reviewer is of the opinion that the generated guiding principle is better than the original guiding principle. In this case, a brief explanation must be provided in the comments column, as a separate review will take place.** |