**ICAIL 2025 Bewertung der Leitsätze**

# **Ziele:**

* Hohe Qualität der Auswertung unter gleichzeitiger Vermeidung des Lesens von Urteilen
* 8 Bewertungsklassen 🡪 ggf. reduzieren, sofern keine sinnvollen Ergebnisse zu erwarten
* Nach Möglichkeit Übertragbarkeit auf internationale Sachverhalte; dazu Evaluation möglichst unabhängig von Besonderheiten der Leitsätze des BGH?

# **Qualität der Leitsätze?**

## Was ist wichtig für einen guten (generierten) Leitsatz?

* Sprache
* Vollständigkeit
* Prägnanz
* Inhaltliche Richtigkeit

## Vorgaben von Beck:

* https://zitierportal.beck.de/documents/RedRL\_CHBeck\_Leits%C3%A4tze\_1\_8\_2021.pdf
* Mindestens ein Leitsatz, grds. nicht mehr als sechs Leitsätze; nicht mehr als sechs Zeilen à 80 – 85 Zeichen.
* Wiedergabe der wesentlichen vom Gericht entschiedenen Rechtsfragen, einschließlich des konkreten Entscheidungsergebnisses. Jede Rechtsfrage in einem eigenen Leitsatz. Bloße Schlagworte reichen nicht.
* Einbettung der Leitsätze in den Gesamtkontext der Judikatur. Hinweis auf bisherige Rechtsprechung in Klammerzusatz vor dem Punkt.

# **Bewertungsmaßstab/Evaluationsklassen**

## Alte Klassen:

|  |  |
| --- | --- |
| Klasse | Beschreibung |
| 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 |

## Neue Klassen:

|  |  |
| --- | --- |
| Klasse | Beschreibung |
| 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) |

## Begründung der Änderung und Leitfaden:

|  |  |
| --- | --- |
| Klasse | Beschreibung |
| 1 | Klasse 1 wird beibehalten: Es ist davon auszugehen, dass die fine-tuned models größtenteils sinnvollen Inhalt generieren werden. Dennoch bedarf es dieser Klasse, um fehlerhafte Ausgaben markieren zu können.  **Klasse 1 ist erfüllt, wenn ein lesbares Ergebnis generiert wurde. Die Ausgabesprache ist dabei unerheblich.** |
| 2 | Klassen 2 und 3 werden zusammengefasst, da sich bei der Evaluation kaum Unterschiede ergeben haben.  **Klasse 2 ist erfüllt, sofern die juristischen Begriffe richtig verwendet werden und ein Ergebnis in deutscher Sprache generiert wurde, das grammatikalisch korrekt ist**. |
| 3 | Die neue Klasse 3 umfasst die alte Klasse 6 und wurde nach vorne verschoben. Wichtig ist der Zusatz der „unnötigen“ Informationen, der die Bewertung flexibler gestalten soll. Zuvor reichte – streng gesehen – bereits der Zusatz eines Wortes, um die Klasse nicht zu erfüllen.  **Klasse 3 ist erfüllt, wenn der Leitsatz keine Informationen enthält, die unnötig für den Leitsatz sind. Einzelne, zusätzliche Wörter oder das zusätzliche Zitieren von relevanten Vorschriften/Urteilen/etc. ist grundsätzlich unschädlich. Logischerweise besteht hier ein Beurteilungsspielraum, der jedoch durch die Auswertung von drei verschiedenen Juristen zu einem klaren Ergebnis per Mehrheitsentscheidung führen wird. Bei Grenzfällen empfiehlt es sich nach dem ersten Eindruck zu entscheiden, ob eine Information als „unnötig“ angesehen wird.**  **Sofern besondere Muster in den unnötigen Informationen zu erkennen sind, kann darauf per Anmerkung kurz hingewiesen werden. (Beispiel mehrfache Wiederholung der Ausführungen). Anmerkungen sollten jedoch möglichst restriktiv gehandhabt werden, da eine manuelle Auswertung erforderlich sein wird. Die Anmerkungen sind dafür gedacht, dass wichtige Erkenntnisse – die durch die Evaluationsklassen nicht nachvollziehbar sind – im anschließenden Text des Beitrags berücksichtigt werden können.** |
| 4 | Klasse 4 entspricht der alten Klasse 5 und soll aufzeigen, ob der generierte Leitsatz vollständig ist. Die Vollständigkeit wird anhand eines Vergleichs zum originalen Leitsatz überprüft.  **Klasse 4 ist erfüllt, wenn der generierte Leitsatz jeden relevanten rechtlichen Aspekt behandelt, den auch der originale Leitsatz behandelt. Dabei ist es unerheblich, ob die Rechtsfragen korrekt behandelt werden oder die rechtlichen Ausführungen hinsichtlich des einzelnen Aspekts vollständig sind.**  **Dabei ist es nicht erforderlich, dass der generierte Leitsatz den genauen Wortlaut des originalen Leitsatzes trifft, sofern der dahinterstehende Sinn unverändert bleibt. Der genaue Wortlaut wäre jedoch erforderlich, wenn dieser den Hauptgegenstand des originalen Leitsatzes darstellt. Beispiel wäre die Einführung einer Rechtsfigur mit einem Eigennamen, der im generierten Leitsatz nicht genau benannt wird.** |
| 5 | Klasse 5 entspricht der alten Klasse 7 und ist im Wesentlichen eine Kontrollklasse, um die Qualität der Leitsätze besser ausdrücken zu können. Prüfung erfolgt auch nur durch einen Vergleich (siehe Klasse 4).  **Klasse 5 ist erfüllt, wenn der generierte Leitsatz zumindest den Großteil der rechtlichen Aspekte behandelt, die auch der originale Leitsatz behandelt. Klasse 5 ist erfüllt, wenn mindestens 3/4 der rechtlichen Aspekte behandelt werden. Klasse 5 ist logischerweise auch dann erfüllt, wenn bereits Klasse 4 erfüllt wurde.** |
| 6 | Klasse 6 ist neu. Die fine-tuned models können ggf. halluzinieren oder die Parteien vertauschen. Die inhaltliche Richtigkeit soll grundsätzlich nur anhand eines Vergleichs mit dem originalen Leitsatz festgestellt werden. Nur wenn eine Feststellung der inhaltlichen Richtigkeit in dieser Form nicht möglich ist, sollte das Urteil herangezogen werden, um festzustellen, ob die inhaltliche Richtigkeit wirklich gegeben ist.  **Klasse 6 ist erfüllt, wenn der generierte Leitsatz die Aspekte des originalen Leitsatzes inhaltlich ohne Fehler wiedergibt. Nicht zu untersuchen sind die Aspekte, die im generierten Leitsatz, aber nicht im originalen Leitsatz enthalten sind, da diese nicht ohne einen Blick in das Urteil überprüft werden können. Klasse 6 ist zudem unabhängig von der Erfüllung der Klassen 4 und 5. Es werden nur die Aspekte des originalen Leitsatzes untersucht, die der generierte Leitsatz auch tatsächlich beinhaltet. Sofern der generierte Leitsatz keinen Aspekt des originalen Leitsatzes beinhaltet, ist Klasse 6 jedoch mangels Vergleichsgegenstand nicht erfüllt.**  **Sofern ein offensichtlicher und schwerwiegender Fehler in den zusätzlichen Informationen des generierten Leitsatzes erkannt wird, kann darauf per Anmerkung hingewiesen werden, wenn dies für die spätere Auseinandersetzung im Beitrag relevant sein könnte.** |
| 7 | Klasse 7 bleibt unverändert, um besonders gute Leitsätze hervorheben zu können.  **Klasse 7 ist erfüllt, wenn die Prüferin oder der Prüfer der Auffassung ist, dass der generierte Leitsatz besser als der originale Leitsatz ist. In diesem Fall ist eine kurze Begründung in der Spalte für Anmerkungen anzufertigen, da eine gesonderte Nachprüfung stattfinden wird.** |