[klara]case

BENCHMARK · AUSGABE 2026-11

# Wie oft nennt eine KI ein Aktenzeichen, das nicht in unserem Bestand steht?

Wir haben vier KI-Modelle nach der einschlägigen Rechtsprechung gefragt: zehn Fragen aus sechs Rechtsgebieten. Jede Frage lief dreimal: ohne Werkzeuge, mit Websuche und mit klaracase. Danach haben wir jedes genannte Aktenzeichen mit unserem Bestand abgeglichen.

| 4 | 6 | 3 | 902 |
|---|---|---|---|
| KI-Modelle | Rechtsgebiete | Varianten je Frage | Genannte Aktenzeichen |

| | |
|---|---|
| AUSGABE | 2026-11 |
| DATUM | Veröffentlicht am 3. November 2026 |
| FRAGENSATZ | Fragensatz 1 |
| MODELLE | Getestete Modelle, gemessen am 3. November 2026: GPT-6.1 Terra · GPT-6 Astra · Claude Opus 5.5 · Claude Sonnet 5.5 |

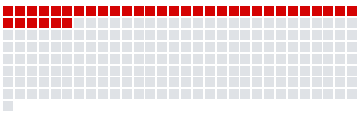
01 · ERGEBNIS

## Wie oft nennt eine KI ein Aktenzeichen, das nicht in unserem Bestand steht?

■ Ohne Werkzeuge

**36** von 241

genannten Aktenzeichen stehen nicht in unserem Bestand



Anteil: 15 % (36 von 241)
Falsches Datum: 5 von 241

◆ Mit Websuche

**27** von 306

genannten Aktenzeichen stehen nicht in unserem Bestand

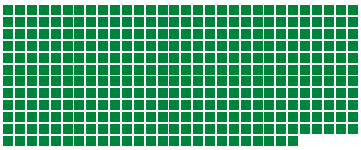


Anteil: 9 % (27 von 306)
Falsches Datum: 7 von 306

● Mit klaracase

**355** von 355

genannten Aktenzeichen stammen aus einem Werkzeugergebnis (nachvollziehbar)



Anteil: 100 % (355 von 355)
Nicht in unserem Bestand: 0 von 355
Falsches Datum: 1 von 355

02 · RECHTSGEBIETE

## Ergebnisse nach Rechtsgebiet

Je Rechtsgebiet und Variante: Aktenzeichen, die nicht in unserem Bestand stehen, von allen genannten.

| RECHTSGEBIET | ■ Ohne Werkzeuge | ◆ Mit Websuche | ● Mit klaracase |
|---|---|---|---|
| Strafrecht | **5 von 16** | **8 von 85** | 0 von 92 |
| Wirtschaftsstrafrecht | **6 von 59** | **1 von 47** | 0 von 59 |
| Mietrecht | **4 von 77** | 0 von 68 | 0 von 85 |
| Arbeitsrecht | **8 von 50** | **9 von 53** | 0 von 64 |
| Steuerrecht | **5 von 28** | **1 von 32** | 0 von 32 |
| Sozialrecht | **8 von 11** | **8 von 21** | 0 von 23 |

03 · JE MODELL

# Ergebnisse nach Modell

Alle sechs Rechtsgebiete zusammen, je Modell und Variante. Die Namen sind die exakt geprüften Modellversionen, gemessen am 3. November 2026.

| VARIANTE | GENANNT | BELEGT | FALSCHES DATUM | FALSCHES GERICHT | NICHT IN UNSEREM BESTAND | ANDERSWO GEFUNDEN | OFFEN | NACHVOLLZIEHBAR |
|---|---|---|---|---|---|---|---|---|
| **GPT-6.1 Terra** | | | | | | | | |
| ■ Ohne Werkzeuge | 23 | 21 | 0 | 0 | **2** | 1 | 1 | – |
| ◆ Mit Websuche | 39 | 36 | 2 | 0 | **1** | 1 | 0 | – |
| ● Mit klaracase | 60 | 60 | 0 | 0 | 0 | 0 | 0 | 60 |
| **GPT-6 Astra** | | | | | | | | |
| ■ Ohne Werkzeuge | 46 | 43 | 1 | 0 | **2** | 0 | 2 | – |
| ◆ Mit Websuche | 62 | 61 | 0 | 0 | **1** | 0 | 1 | – |
| ● Mit klaracase | 65 | 64 | 1 | 0 | 0 | 0 | 0 | 65 |
| **Claude Opus 5.5** | | | | | | | | |
| ■ Ohne Werkzeuge | 69 | 59 | 0 | 0 | **10** | 7 | 3 | – |
| ◆ Mit Websuche | 97 | 78 | 2 | 0 | **17** | 5 | 12 | – |
| ● Mit klaracase | 123 | 123 | 0 | 0 | 0 | 0 | 0 | 123 |
| **Claude Sonnet 5.5** | | | | | | | | |
| ■ Ohne Werkzeuge | 103 | 77 | 4 | 0 | **22** | 10 | 12 | – |
| ◆ Mit Websuche | 108 | 97 | 3 | 0 | **8** | 4 | 4 | – |
| ● Mit klaracase | 107 | 107 | 0 | 0 | 0 | 0 | 0 | 107 |

Die Spalten „Belegt“, „Falsches Datum“, „Falsches Gericht“ und „Nicht in unserem Bestand“ ergeben zusammen die Zahl unter „Genannt“. „Belegt“ heißt: gleiches Gericht und gleiches Datum wie in unserem Bestand. „Anderswo gefunden“ und „Offen“ teilen die Spalte „Nicht in unserem Bestand“ auf. „Nachvollziehbar“ zählt die genannten Aktenzeichen, die aus einem Werkzeugergebnis stammen.

# So haben wir gemessen

### ■ Ohne Werkzeuge

1 Frage
2 KI-Modell
3 Antwort mit Aktenzeichen

Das Modell beantwortet die Frage allein. Keine Suche, keine Werkzeuge, kein Internet.

ABGEGLICHEN MIT
Unser Bestand an Gerichtsentscheidungen am Tag der Messung

### ◆ Mit Websuche

1 Frage
**2 KI-Modell mit Websuche**
3 Antwort mit Aktenzeichen

Das Modell darf im Internet suchen, bevor es antwortet.

ABGEGLICHEN MIT
Unser Bestand an Gerichtsentscheidungen am Tag der Messung

### ● Mit klaracase

1 Frage
**2 KI-Modell mit Zugriff auf unsere Datenbank**
3 Antwort mit Aktenzeichen

Das Modell liest unsere Datenbank über Werkzeuge. Wir zählen, wie viele genannte Aktenzeichen aus einem Werkzeugergebnis stammen. Das nennen wir nachvollziehbar.

ABGEGLICHEN MIT
Unser Bestand am Tag der Messung. Außerdem: Stammt das Aktenzeichen aus einem Werkzeugergebnis?

## Was zählt als eins?

Ein genanntes Aktenzeichen in einer Antwort.

## Wie groß ist die Stichprobe?

Vier Modelle, zehn Fragen in sechs Rechtsgebieten, jede Frage in drei Varianten. Genannt wurden 241 Aktenzeichen ohne Werkzeuge, 306 mit Websuche und 355 mit klaracase. Die Zahlen sind nicht gleich groß, weil jedes Modell selbst entscheidet, wie viele Entscheidungen es nennt. Darum steht neben jeder Zahl ihre eigene Gesamtzahl.

## Wie prüfen wir ein Aktenzeichen?

Wir zählen jedes genannte Aktenzeichen und gleichen es mit unserem Bestand vom Tag der Messung ab. Es ist belegt (gleiches Gericht, gleiches Datum), hat ein falsches Datum, hat ein falsches Gericht oder steht nicht in unserem Bestand.

## Wie oft haben wir jede Frage gemessen?

Einmal je Modell und Variante. Wir bilden keinen Mittelwert und behaupten keine Wiederholungen.

FANGFRAGEN

**20** von 24

Zwei weitere Fragen nennen Paragrafen, die es nicht gibt. Sie zählen nicht zu den zehn Fragen. Darauf kamen 24 Antworten. Eine automatische Prüfung hat bei 20 davon erkannt, dass die Antwort den Paragrafen als nicht vorhanden bezeichnet. Die übrigen 4 muss noch eine Person prüfen.

### Warum die Fragen nicht öffentlich sind

Wir veröffentlichen den Wortlaut der Fragen nicht. Veröffentlichte Fragen gelangen in die Trainingsdaten der Modelle. Dann kennt ein Modell die Fragen schon, und der Test verliert seinen Wert.

# Was dieser Benchmark nicht zeigt

### Was „nicht in unserem Bestand“ heißt

Es heißt nur: Wir haben das Aktenzeichen in unserem Bestand nicht gefunden. Die Entscheidung kann es trotzdem geben. Diese Aktenzeichen haben wir auf dejure.org nachgeschlagen. Manche gibt es dort („anderswo gefunden“). Der Rest ist noch offen.

| VARIANTE | NICHT IN UNSEREM BESTAND | ANDERSWO GEFUNDEN | OFFEN |
|---|---|---|---|
| Ohne Werkzeuge | 36 | 18 | 18 |
| Mit Websuche | 27 | 10 | 17 |
| Mit klaracase | 0 | 0 | 0 |

„Offen“ heißt: Eine Person hat diese Aktenzeichen noch nicht geprüft. Wir behaupten dazu nichts.

01 Wir haben nicht gemessen, ob eine Antwort rechtlich richtig ist. Niemand hat den juristischen Inhalt der Antworten geprüft. Wir zählen nur genannte Aktenzeichen.

02 klaracase hat das Werkzeug gebaut und die Messung selbst bewertet. Wir sind Anbieter und Prüfer zugleich.

03 Das falsche Datum zählen wir je Aktenzeichen aus dem Antworttext. Ein automatischer Leser ordnet die Daten zu. Er kann sich irren.

04 Die Zahlen gelten für die vier geprüften Modellversionen am Tag der Messung. Es waren zehn Fragen in sechs Rechtsgebieten. Über spätere Versionen und andere Fragen sagen die Zahlen nichts.

05 Alle Zahlen stammen aus einer festgeschriebenen Messung vom 3. November 2026. Sie ändern sich nicht, wenn unser Bestand wächst.

06 · AUSGABEN

## Alle Ausgaben

Eine neue Ausgabe erscheint etwa alle drei Monate. Die Zahlen sind nur zwischen Ausgaben mit demselben Fragensatz vergleichbar.

| AUSGABE | DATUM | FRAGENSATZ | MODELLE | ■ Ohne Werkzeuge NICHT IN UNSEREM BESTAND | ◆ Mit Websuche NICHT IN UNSEREM BESTAND | ● Mit klaracase NACHVOLLZIEHBAR |
|---|---|---|---|---|---|---|
| 2026-11 | 3. November 2026 | 1 | 4 | 36 von 241 | 27 von 306 | 355 von 355 |

07 · KONTAKT

## Sprechen Sie mit uns

### Forschung und Presse

Sie haben Fragen zur Methode oder zu den Daten dieses Benchmarks? Schreiben Sie uns. Forschenden stellen wir die Messdaten auf Anfrage zur Verfügung.

pr@klaracase.de

### Kanzleien und Unternehmen

Sie möchten wissen, ob klaracase zu Ihrer Arbeit passt? Schreiben Sie uns. Wir vereinbaren dann einen Termin für ein Gespräch.

info@klaracase.de

### Diese Ausgabe im Netz

klaracase.de/benchmark/edition/2026-11

### Selbst ausprobieren

Legen Sie ein kostenloses Konto an und testen Sie klaracase noch heute.

klaracase.de/register