

Benchmark-Bogen: 30 Fragen, vier Kriterien, eine Schwelle

Schwelle vor dem ersten Kandidaten festlegen, jede Antwort gegen den Beleg bewerten, Bogen je Quartal aufheben.

FACHBEREICH	DATUM	AUSGEFÜLLT VON
-------------	-------	----------------

01 FRAGE (JE FRAGE EINE ZEILE, 30 ZEILEN AUF DER RÜCKSEITE) ☐ Nr. · Fachbereich · Fragetext ☐ Dokumentierte richtige Antwort ☐ Beleg: Dokument · Version · Abschnitt ☐ Typ: Fakt () · Mehrquellen () · Lücke (keine Antwort vorhanden) ()	02 BESTEHENSSCHWELLE (VOR DEM TEST, NICHT MEHR ÄNDERN) ☐ Mindestens ___ von 30 korrekt ☐ Quelle bei jeder Antwort () Pflicht ☐ Erfundene Angaben erlaubt: 0 ☐ Lückenfragen: ___ von 5 ehrlich als Lücke benannt ☐ Maximale Antwortzeit ___ s · maximale Kosten ___ Euro je Antwort
03 TESTBEDINGUNGEN ☐ Datenbasis · Stand vom ☐ Prompt-Version · Anbindung (RAG, MCP, Upload) ☐ Kandidat A · Kandidat B · Kandidat C ☐ Bewerter 1 · Bewerter 2 (Zweitbewertung)	04 BEWERTUNG JE KANDIDAT (SUMME ÜBER 30 FRAGEN) ☐ Korrekt: A ___ · B ___ · C ___ ☐ Mit Quelle: A ___ · B ___ · C ___ ☐ Lücke ehrlich benannt (von 5): A ___ · B ___ · C ___ ☐ Erfundene Angaben: A ___ · B ___ · C ___ ☐ Durchschnittliche Zeit / Kosten: A ___ · B ___ · C ___
05 ENTSCHEIDUNG ☐ Bestanden: A () · B () · C () ☐ Wenn keiner bestanden: fehlende oder veraltete Belege (Liste) ☐ Nächster Schritt · Verantwortlich · Termin ☐ Nächste Wiederholung am	

REIHENFOLGE

1 Fragen sammeln	2 Schwelle festlegen	3 Kandidaten messen	4 Entscheiden	5 Quartal: wiederholen
------------------	----------------------	---------------------	---------------	------------------------



TIPPS VOM BYTE

- Dreißig echte Fragen mit bekannten Antworten schlagen jeden Benchmark. Sammeln Sie sie im Service und im Vertrieb.
- Ändern Sie zwischen zwei Tests nur eine Sache: Modell oder Retrieval. Sonst wissen Sie nicht, was gewirkt hat.
- Zählen Sie auch die Zeit bis zur Antwort und die Kosten je Anfrage. Richtig, aber langsam und teuer, verliert.

GEGENZEICHNUNG

Fachbereich	IT	Geschäftsführung
-------------	----	------------------

Regel: Die Schwelle aus Block 02 wird vor dem ersten Kandidaten unterschrieben und während des Tests nicht verändert.