

KI-Tool Scorecard

Auswahl & Bewertung

Nutzen, Datenschutz, Integration, Kosten, Governance – strukturiert entscheiden statt impulsiv kaufen. Diese Scorecard gibt Ihnen ein erprobtes Framework, um KI-Tools objektiv zu vergleichen und fundierte Entscheidungen zu treffen.



Ziel & Definition of Done

„Tool-Entscheidung statt Hype-Kauf.“

Der Markt für KI-Tools wächst exponentiell. Wöchentlich erscheinen neue Lösungen, die mit beeindruckenden Demos und großen Versprechen um die Aufmerksamkeit von Entscheidern buhlen. Doch ohne strukturierten Bewertungsprozess landen Unternehmen schnell bei Werkzeugen, die nicht zur eigenen Infrastruktur passen, Datenschutzerfordernungen verletzen oder nach der Testphase ungenutzt bleiben. Diese Scorecard schafft Abhilfe: Sie ersetzt Bauchgefühl durch Methodik und stellt sicher, dass jede Tool-Entscheidung nachvollziehbar, vergleichbar und dokumentiert ist.

Die **Definition of Done** definiert klare Erfolgskriterien für Ihren Auswahlprozess. Erst wenn alle drei Bedingungen erfüllt sind, gilt die Evaluation als abgeschlossen:

1

Anforderungen klar definiert

Use Case, Muss-Kriterien, Datenarten und Integrationsanforderungen sind schriftlich festgehalten. Alle Stakeholder haben die Anforderungen reviewed und freigegeben. Ohne diese Grundlage ist jeder Vergleich wertlos.

2

2-5 Tools vergleichbar bewertet

Jedes Tool wurde anhand der acht Kriterien dieser Scorecard mit gewichteten Scores bewertet. Die Ergebnisse sind in einer einheitlichen Matrix dokumentiert und ermöglichen einen direkten, objektiven Vergleich.

3

Pilotplan + Entscheidung dokumentiert

Ein konkreter Pilotplan mit Zeitrahmen, KPIs und Verantwortlichkeiten liegt vor. Die finale Entscheidung ist mit Begründung protokolliert – nachvollziehbar für Geschäftsführung, IT und Datenschutz gleichermaßen.



Praxistipp: Starten Sie den Prozess immer mit der Anforderungsdefinition (Seite 5), bevor Sie Tools recherchieren. So vermeiden Sie den häufigsten Fehler: eine Lösung zu suchen, bevor das Problem klar ist.

Bewertungsmodell – 8 Kriterien

Das Herzstück der Scorecard sind acht sorgfältig ausgewählte Bewertungskriterien. Sie decken alle relevanten Dimensionen ab – von der unmittelbaren Nutzenstiftung bis zur langfristigen Zukunftsfähigkeit. Jedes Kriterium wird individuell gewichtet, sodass Sie die Scorecard exakt an die Prioritäten Ihres Unternehmens anpassen können. Die Kriterien basieren auf realen Evaluationsprozessen aus mittelständischen und großen Unternehmen und adressieren gezielt die Bereiche, in denen KI-Tool-Entscheidungen am häufigsten scheitern.



1 – Nutzen / Impact

Messbarer Mehrwert für den definierten Use Case



2 – Daten / Datenschutz

DSGVO-Konformität, Datenverarbeitung und Speicherort



3 – Integration / API

Anbindung an bestehende Systeme und Schnittstellen



4 – Kosten / TCO

Gesamtkosten inkl. Lizenz, Setup, Betrieb und Schulung



5 – Bedienbarkeit

Intuitive Nutzung, Onboarding-Aufwand und UX-Qualität



6 – Governance / Logging

Nachvollziehbarkeit, Audit-Trails und Compliance-Features



7 – Vendor-Risiko

Anbieter-Stabilität, Lock-in-Gefahr und Exit-Strategie



8 – Skalierung / Zukunft


Wachstumsfähigkeit, Roadmap und technologische Weiterentwicklung

Bewerten Sie jedes Kriterium auf einer Skala von **0 (nicht erfüllt)** bis **5 (vollständig erfüllt)**. Durch die individuelle Gewichtung (0–3) fließen geschäftskritische Kriterien stärker in das Gesamtergebnis ein als Nice-to-have-Aspekte.

Scorecard Template

Die folgende Tabelle ist das zentrale Arbeitsinstrument Ihrer KI-Tool-Evaluation. Tragen Sie für jedes zu bewertende Tool die Scores (0–5) ein und multiplizieren Sie diese mit der individuellen Gewichtung (0–3). Das Tool mit dem höchsten Gesamtscore ist Ihr Top-Kandidat – vorausgesetzt, alle Muss-Kriterien sind erfüllt. Passen Sie die Gewichtungen an Ihre spezifische Unternehmenssituation an: Ein stark reguliertes Unternehmen wird Datenschutz und Governance höher gewichten, ein Startup möglicherweise Kosten und Bedienbarkeit.

Kriterium	Gewicht (0–3)	Tool A Score (0–5)	Tool B Score (0–5)	Tool C Score (0–5)	Notizen
1 – Nutzen / Impact					
2 – Daten / Datenschutz					
3 – Integration / API					
4 – Kosten / TCO					
5 – Bedienbarkeit					
6 – Governance / Logging					
7 – Vendor-Risiko					
8 – Skalierung / Zukunft					
Gesamt (Gewicht * Score)	Σ	Σ	Σ	Σ	

 **Hinweis zur Gewichtung:** Gewicht 3 = geschäftskritisch, 2 = wichtig, 1 = wünschenswert, 0 = nicht relevant. Wir empfehlen, maximal 3 Kriterien mit Gewicht 3 zu versehen, um eine echte Priorisierung zu erzwingen.

Anforderungen definieren

Bevor Sie überhaupt beginnen, Tools zu recherchieren, müssen die Anforderungen klar und schriftlich fixiert sein. Dieses Formular zwingt alle Beteiligten, die entscheidenden Fragen zu beantworten. Füllen Sie es gemeinsam mit den Fachabteilungen aus – idealerweise in einem Workshop mit IT, Datenschutz und den zukünftigen Nutzern. Nur so stellen Sie sicher, dass keine blinden Flecken entstehen und alle Perspektiven berücksichtigt werden.

Use Case

Beschreiben Sie den konkreten Anwendungsfall in 2-3 Sätzen. Was soll das Tool leisten? Welches Problem wird gelöst? Wer sind die primären Nutzer?

Muss-Kriterien (3 Bulletpoints)

- _____
- _____
- _____

Kann-Kriterien (3 Bulletpoints)

- _____
- _____
- _____

Datenarten

Welche Daten werden verarbeitet?
Personenbezogene Daten,
Geschäftsgeheimnisse, öffentliche
Informationen?

Integrationen

Welche bestehenden Systeme müssen angebunden werden?

- ERP-System: _____
- CRM-System: _____
- Shop-System: _____
- Ticket-System: _____
- Sonstige: _____

Budgetrahmen

Monatlich / jährlich, inkl. Setup-Kosten:

Go-Live Ziel


Wann soll das Tool produktiv sein?

Tipp: Unterscheiden Sie klar zwischen Muss- und Kann-Kriterien. Ein Muss-Kriterium ist ein K.O.-Kriterium – wird es nicht erfüllt, scheidet das Tool unabhängig vom Gesamtscore aus.

Datenschutz & Governance Kurzcheck

KI-Tools verarbeiten häufig sensible Unternehmensdaten – von Kundenkommunikation über interne Strategiedokumente bis hin zu personenbezogenen Daten. Dieser Kurzcheck hilft Ihnen, die wichtigsten Datenschutz- und Governance-Fragen systematisch abzuclarbeitern, bevor ein Tool in die engere Auswahl kommt. Jede Frage, die mit „Nein“ oder „Unklar“ beantwortet wird, erfordert eine Risikobewertung und ggf. eine Klärung mit dem Anbieter. Beziehen Sie unbedingt Ihren Datenschutzbeauftragten in diesen Schritt ein.

Nr.	Prüffrage	Ja	Nein	Unklar	Notiz
1	Ist ein Auftragsverarbeitungsvertrag (AVV) verfügbar?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
2	Ist EU-Hosting (DSGVO-konform) möglich?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
3	Werden Prompts/Inputs vom Anbieter gespeichert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
4	Werden Daten für Modelltraining verwendet?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
5	Sind Logs exportierbar (Audit-Trail)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
6	Sind Rollen und Rechte granular steuerbar?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
7	Gibt es eine Datenlöschfunktion (Right to Erasure)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
8	Ist SSO / SAML-Integration möglich?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
9	Liegt eine SOC 2 / ISO 27001 Zertifizierung vor?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
10	Gibt es eine transparente Subauftragnehmer-Liste?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

 **Bewertungsregel:** Fragen 1–3 sind K.O.-Kriterien. Wird eine davon mit „Nein“ beantwortet und ist keine Abhilfe möglich, scheidet das Tool für den Einsatz mit personenbezogenen oder vertraulichen Daten aus. Dokumentieren Sie jede Klärung schriftlich.

Pilot-Plan Template (2 Wochen)

Ein strukturierter Pilot ist der entscheidende Schritt zwischen theoretischer Bewertung und fundierter Entscheidung. Planen Sie zwei Wochen ein, um das favorisierte Tool unter realen Bedingungen zu testen. Der folgende Pilotplan gliedert sich in vier Phasen und stellt sicher, dass am Ende belastbare Ergebnisse vorliegen – nicht nur subjektive Eindrücke. Definieren Sie vorab klare KPIs und Erfolgskriterien, damit die Entscheidung am Ende datenbasiert erfolgt.



Nutzen Sie die folgende Tabelle, um jede Phase konkret auszuplanen:

Tag / Phase	Ziel	Testfälle	KPI	Verantwortlich	Ergebnis
Tag 1-2: Setup	Zugang, Konfiguration, Testdaten	Account einrichten, API-Key generieren, Testdaten laden	Setup abgeschlossen Ja/Nein		
Tag 3-5: Test I	Kernfunktionen prüfen	3-5 typische Use Cases durchspielen	Qualität Output (1-5), Geschwindigkeit		
Tag 6-8: Test II	Edge Cases, Integration	Grenzfälle testen, API- Anbindung prüfen	Fehlerrate, Integrationsaufw and (h)		
Tag 9-10: Review	Ergebnisse auswerten	Feedback aller Tester sammeln, KPIs auswerten	Nutzerzufrieden heit (NPS)		
Tag 11-14: Entscheidung	Go/No-Go	Ergebnisse präsentieren, Entscheidung dokumentieren	Entscheidung dokumentiert Ja/Nein		

Praxisbeispiel: Zwei Tools bewertet

Um die Anwendung der Scorecard zu veranschaulichen, zeigen wir eine ausgefüllte Bewertung für zwei anonymisierte KI-Tools. Der Use Case: ein mittelständisches Unternehmen sucht ein KI-gestütztes Tool zur automatisierten Kundenkommunikation. Die Gewichtungen spiegeln die Prioritäten wider: Datenschutz und Integration haben höchste Priorität, Bedienbarkeit ist wichtig, Vendor-Risiko nachrangig.

Kriterium	Gewicht	Tool A (Score)	Tool B (Score)	Notizen
Nutzen / Impact	3	4 → 12	5 → 15	Tool B: bessere Qualität der Antworten
Daten / Datenschutz	3	5 → 15	3 → 9	Tool A: EU-Hosting, AVV vorhanden
Integration / API	3	4 → 12	2 → 6	Tool B: keine native CRM-Anbindung
Kosten / TCO	2	3 → 6	4 → 8	Tool B: günstiger pro Seat
Bedienbarkeit	2	3 → 6	5 → 10	Tool B: deutlich intuitivere UI
Governance / Logging	2	4 → 8	2 → 4	Tool A: vollständiger Audit-Trail
Vendor-Risiko	1	4 → 4	3 → 3	Beide stabil, Tool A größerer Anbieter
Skalierung / Zukunft	2	4 → 8	4 → 8	Beide mit guter Roadmap
Gesamt	18	71	63	Tool A gewinnt

Die 3 wichtigsten Learnings

Gewichtung entscheidet

Tool B hat in mehr Einzelkategorien bessere Scores – gewinnt aber nicht, weil es bei den hochgewichteten Kriterien (Datenschutz, Integration) schwächer ist. Die Gewichtung macht den Unterschied.

UX ≠ Eignung

Das intuitivere Tool ist nicht automatisch das bessere. Bedienbarkeit ist wichtig, aber ohne passende Integration und Datenschutz-Compliance wertlos im Enterprise-Kontext.

Dokumentation lohnt sich

Durch die strukturierte Bewertung konnte die Entscheidung gegenüber der Geschäftsführung in 10 Minuten begründet werden – ohne Diskussionen über „Gefühl“ oder „Demos“.

Der Prozess im Überblick

Von der ersten Idee bis zur finalen Entscheidung: So fügen sich alle Bausteine der Scorecard zu einem durchgängigen Evaluationsprozess zusammen. Jeder Schritt baut auf dem vorherigen auf und stellt sicher, dass nichts vergessen wird. Die typische Durchlaufzeit beträgt 3–4 Wochen – deutlich schneller und strukturierter als die üblichen Ad-hoc-Evaluierungen, die sich über Monate ziehen können.



Typischer Zeitrahmen

- Woche 1: Anforderungen + Marktrecherche
- Woche 2: Datenschutz-Check + Scorecard
- Woche 3–4: Pilotphase
- Ende Woche 4: Dokumentierte Entscheidung

Beteiligte Rollen

- **Fachbereich:** Use Case + Anforderungen
- **IT:** Integration + Sicherheit
- **Datenschutz:** DSGVO-Kurzcheck
- **Einkauf:** Kosten + Verträge
- **Management:** Go/No-Go-Entscheidung

Häufige Fehler bei der KI-Tool-Auswahl

Aus der Erfahrung mit dutzenden KI-Tool-Evaluationen in Unternehmen unterschiedlicher Branchen und Größen haben sich wiederkehrende Fehlerquellen herauskristallisiert. Diese zu kennen und aktiv zu vermeiden, ist mindestens genauso wichtig wie die strukturierte Bewertung selbst. Die folgenden sechs Fehler kosten Unternehmen regelmäßig Zeit, Geld und Vertrauen – und sind mit der richtigen Methodik vollständig vermeidbar.

1 Demo-Blindheit
Das Tool sieht in der Demo des Vertriebsteams großartig aus – aber im Alltag mit echten Daten und realen Edge Cases versagt es. Testen Sie immer mit eigenen Daten und Szenarien.

2 Datenschutz als Nachgedanke
Erst nach dem Kauf wird der Datenschutzbeauftragte einbezogen – und stellt fest, dass das Tool nicht DSGVO-konform eingesetzt werden kann. Der Kurzcheck gehört an den Anfang.

3 Fehlende Gewichtung
Alle Kriterien werden gleich behandelt. Das führt dazu, dass ein Tool mit perfekter UX aber mangelhafter Integration gewinnt – und später nicht produktiv eingesetzt werden kann.

4 Einzelperson entscheidet
Eine Person testet, eine Person entscheidet. Ohne Perspektiven aus IT, Fachbereich und Datenschutz werden kritische Schwächen übersehen.

5 Kein Pilotplan
Das Tool wird ohne definierte KPIs und Testfälle „mal ausprobiert“. Ohne Struktur ist das Ergebnis immer: „Fühlt sich ganz gut an“ – aber keine belastbare Entscheidungsgrundlage.

6 TCO unterschätzt
Nur die Lizenzkosten werden verglichen. Setup, Schulung, Integration, laufender Support und potenzielle Skalierungskosten werden ignoriert – und sprengen später das Budget.

Checkliste vor der Entscheidung

Bevor Sie die finale Entscheidung treffen und den Beschaffungsprozess starten, gehen Sie diese Checkliste durch. Sie fasst alle wesentlichen Schritte zusammen und stellt sicher, dass kein kritischer Aspekt übersehen wurde. Erst wenn alle Punkte abgehakt sind, haben Sie eine belastbare Entscheidungsgrundlage – für sich selbst, für Ihr Team und für die Geschäftsführung.

01

Anforderungsdokument vollständig

Use Case beschrieben, Muss-/Kann-Kriterien definiert, Datenarten identifiziert, Integrationen benannt, Budget und Go-Live-Ziel festgelegt.

02

Shortlist erstellt

2–5 Tools identifiziert, die grundsätzlich zum Use Case passen. Offensichtlich ungeeignete Kandidaten bereits aussortiert.

03

Datenschutz-Check durchgeführt

Alle 10 Prüffragen für jedes Tool beantwortet. K.O.-Kriterien geprüft. Datenschutzbeauftragter einbezogen und Ergebnis dokumentiert.

04

Scorecard ausgefüllt

Alle 8 Kriterien für jedes Tool bewertet. Gewichtungen festgelegt und mit Stakeholdern abgestimmt. Gesamtscores berechnet.

05

Pilot abgeschlossen

2-Wochen-Pilot mit dem Top-Kandidaten durchgeführt. KPIs gemessen, Feedback gesammelt, Ergebnisse ausgewertet.

06

Entscheidung dokumentiert

Go/No-Go-Entscheidung mit Begründung schriftlich festgehalten. Nächste Schritte definiert (Vertrag, Rollout-Plan, Schulungskonzept).

Ressourcen & Kontakt

Diese Scorecard ist Teil einer umfassenden Methodensammlung für den strukturierten Einsatz von KI in Unternehmen. Alle Templates, Checklisten und Frameworks sind darauf ausgelegt, Ihnen Zeit zu sparen und Entscheidungen auf eine solide Grundlage zu stellen. Wenn Sie tiefer einsteigen möchten, finden Sie im Buch „**KI mit Kaffee-Intelligenz**“ praxisnahe Anleitungen, Fallbeispiele und strategische Frameworks für den gesamten KI-Lebenszyklus – von der Ideenfindung über die Tool-Auswahl bis zur unternehmensweiten Skalierung.

Buch-Teaser

KI mit Kaffee-Intelligenz – Der pragmatische Leitfaden für Entscheider, die KI nicht nur einführen, sondern wirksam machen wollen. Keine Buzzwords, keine übertriebenen Versprechen – stattdessen erprobte Methoden, ehrliche Einschätzungen und sofort umsetzbare Templates.

Das Buch liefert Ihnen:


- Strategien für die KI-Einführung ohne Überforderung
- Bewertungsframeworks wie diese Scorecard
- Governance-Modelle für den verantwortungsvollen Einsatz
- Praxisberichte aus dem deutschsprachigen Mittelstand



Scannen Sie den QR-Code für weitere Templates, Updates und kostenlose Ressourcen.

Kontakt

Philipp Diekmann

 kaffee-intelligenz.de

Für Fragen, Feedback oder individuelle Beratung zur KI-Tool-Auswahl in Ihrem Unternehmen stehe ich Ihnen gerne zur Verfügung.

Weitere Ressourcen

● **KI-Strategie-Canvas**

Von der Vision zum konkreten Maßnahmenplan

● **Prompt-Engineering-Guide**

Bessere Ergebnisse durch strukturierte Prompts

● **KI-Governance-Framework**

Regeln, Rollen und Verantwortlichkeiten definieren