

Das erste KI-Pilotprojekt

Ein Testlauf mit Hypothese, Baseline und einer Entscheidung am Ende

KI-Projekte scheitern selten an der Technologie, sondern an einem unscharfen Start: zu viele Ziele, zu wenig Struktur, keine Messbarkeit. Ein Pilot ist kein Techniktest, sondern ein **strukturierter Versuchsaufbau** mit echten Anwendern, echten Daten und einem Vergleichswert.

DIE FÜNF SCHRITTE

1 HYPOTHESE

Was genau willst du herausfinden? Ein Satz, der prüfbar ist. Der Startschuss fällt nicht bei der Toolauswahl.

„Eine KI klassifiziert eingehende Servicemails. Dadurch sinkt die Bearbeitungszeit je Ticket.“

2 RAHMEN

Zeitraum (vier bis acht Wochen), Mindestfallzahl, Testgruppe, Rollen, Erfolgskriterien und die Baseline: der gemessene Ist-Zustand vor dem Start.

3 BASIS

Datenqualität und Zugänglichkeit klären, Datenschutz und IT-Sicherheit von Anfang an einbinden, interne Freigaben holen. Die beste KI scheitert an schlechten Daten.

4 TESTLAUF

Eng kontrolliert starten, Feedback standardisiert erheben statt frei sammeln, laufend nachschärfen. So entstehen Daten statt Einzelmeinungen.

5 AUSWERTUNG

Gegen die Baseline messen, in vier Dimensionen: Qualität, Zeit und Aufwand, Kosten-Nutzen-Verhältnis, Risiken.

WORAN DU AM ENDE MISST

Dimension	Was du erhebst
Qualität	wie oft lag die KI richtig, und welche Fehler traten auf?
Zeit und Aufwand	veränderte Bearbeitungszeit, plus die neue Kontrollzeit
Kosten und Nutzen	eingesparte Aufwände gegen Lizenz- und Einführungskosten
Risiken	kritische Ausreißer, Datenschutz, Haftungsfragen

Für einen Piloten genügt eine grobe Relation. Ein vollständiges ROI-Modell braucht es noch nicht.

DREI MÖGLICHE AUSGÄNGE

SKALIEREN

Qualitätsziele erreicht, Zeitersparnis belegt, Risiken beherrschbar.

ANPASSEN

Vielversprechend, aber noch zu instabil. Scope schärfen, Daten verbessern, erneut testen.

STOPPEN

Hypothese nicht bestätigt oder Aufwand nicht gerechtfertigt. Kein Scheitern, sondern Klarheit mit geringem Einsatz und eine vermiedene Fehlinvestition.

EIN BEWERTUNGS-PROMPT FÜR DIE TESTPHASE

	Prompt
je Ergebnis	„Bewerte das KI-Ergebnis fachlich entlang dieser Kriterien: 1. Vollständigkeit, 2. Richtigkeit, 3. Verständlichkeit. Vergib pro Kriterium eine Schulnote von 1 bis 5 und ergänze maximal zwei kurze Hinweise zur Verbesserung.“

Standardisiertes Feedback verhindert, dass am Ende nur anekdotische Einzelmeinungen stehen.

MERKSÄTZE

Je schärfer der Scope, desto eindeutiger das Ergebnis.

Ohne Baseline bleibt der Test ein Stimmungsbild.

Ein Stopp ist kein Scheitern, sondern gesparte Fehlinvestition.