Mess-Vorlage

Den Nutzen von KI an einem Ablauf selbst messen

Mit dieser Vorlage messen Sie an einem einzelnen Arbeitsablauf, ob sich die Arbeit mit KI lohnt. Sie brauchen dafür eine Tabelle, einige Wochen Geduld und feste Regeln, die Sie vor dem Start aufschreiben.

Zur Vorlage gehört die Tabelle mess-vorlage.csv. Sie öffnet sich in Excel, Numbers, LibreOffice und Google Tabellen.

Sechs Grundsätze

1. **Die Regeln stehen vor den Daten.** Kennzahl, Zeiträume und Auswertungsregeln schreiben Sie auf, bevor Sie den ersten Wert ansehen. Danach ändern Sie sie nicht mehr.
2. **Gemessen wird Ende-zu-Ende.** Die Zeit läuft vom Beginn der Aufgabe bis zum geprüften, fertigen Ergebnis. Prüfen und Korrigieren gehören dazu.
3. **Systemdaten vor Selbsteinschätzung.** Geschätzte Zeitersparnis liegt nachweislich zu hoch: In einer Studie fühlten sich Entwickler 20 % schneller, waren aber 19 % langsamer (METR 2025). Zeitstempel aus Ihren Systemen sind die beste Quelle, ein Timer an festen Tagen die zweitbeste.
4. **Alle Fälle zählen.** Erfassen Sie jeden Fall im Messzeitraum, auch die ohne KI und die, bei denen es schlecht lief.
5. **Gemessen werden Abläufe, nicht Personen.** Sagen Sie das dem Team vorher. Gibt es einen Betriebsrat, beziehen Sie ihn vor dem Start ein.
6. **Die Qualität bleibt mindestens gleich.** Schnellere Arbeit zählt nur, wenn das Ergebnis nicht schlechter wird.

Schritt 1: Kennzahl und Regeln festlegen

Füllen Sie dieses Formular aus, bevor Sie Werte erheben oder auswerten. Datieren und unterschreiben Sie es.

Feld	Ihre Festlegung	Beispiel
Ablauf	[ ]	Produktbeschreibung aus Herstellerdaten schreiben
Start der Messung	[ ]	Aufgabe wird begonnen
Ende der Messung	[ ]	Text ist geprüft und im Shop veröffentlicht
Hauptkennzahl	[ ]	Median der Arbeitszeit pro Text in Minuten
Art der Zeit	[Durchlaufzeit / Arbeitszeit]	Arbeitszeit (Timer pausiert bei Unterbrechung)
Quelle	[Systemzeitstempel / Stichprobentage mit Timer]	Stichprobentage mit Timer
Baseline-Zeitraum (vorher)	[Daten]	drei feste Stichprobentage
Nachmess-Zeitraum (nachher)	[Daten, wird nicht verlängert]	drei feste Stichprobentage nach einigen Wochen Arbeit mit KI
Leitplanke	[ ]	Anteil der Fälle mit Nacharbeit steigt nicht
Ausgeschlossene Tage	[ ]	Feiertage, Aktionstage, Systemausfälle, in beiden Phasen
Stück pro Woche	[Wie zählen Sie das?]	Zahl veröffentlichter Texte laut Shopsystem
Stundensatz	[Euro pro Stunde]	interner Satz inklusive Nebenkosten
Gesparte Zeit geht in	[ ]	Überarbeitung älterer Produktseiten
Wechselregel	Die Kennzahl wird nur gewechselt, wenn die Datenquelle ausfällt. Der Wechsel wird begründet.	

Erstellt am: _____ Unterschrift: _____

Wenn sich die Aufgabe durch KI verändert: Die Aufgabe ist über Start, Ende und Umfang festgelegt. Neue Leistungen, die es vorher nicht gab, zählen als zusätzliches Ergebnis und nicht als Zeitersparnis.

Schritt 2: Die Baseline erheben

Die Baseline ist der Zustand vorher. Sie ist der Maßstab für alles Weitere.

Weg A: Zeitstempel aus Ihren Systemen (bevorzugt)

- Viele Systeme speichern ohnehin, wann etwas angelegt, bearbeitet und veröffentlicht wurde: Shop, Website-System, Ticketsystem, Postfach.

- Werten Sie diese Zeitstempel rückwirkend für einen Zeitraum im Normalbetrieb aus, zum Beispiel sechs Wochen.
- Vorteil: Niemand muss etwas tun, und die Werte sind frei von dem Effekt, dass Menschen anders arbeiten, wenn sie wissen, dass gemessen wird.
- Prüfen Sie vorher nur, ob die Zeitstempel vollständig und eindeutig sind. Sehen Sie sich die Werte erst an, wenn das Formular unterschrieben ist.

Weg B: Stichprobentage mit Timer

- Für Aufgaben ohne Zeitstempel legen Sie feste Tage fest. An diesen Tagen wird jeder Fall des Ablaufs mit einem Timer erfasst, etwa einer Zeiterfassungs-App oder einem einfachen Formular.
- Erfasst wird jeder Fall des Tages, nicht nur ausgewählte.
- Der Aufwand liegt bei wenigen Minuten pro Tag.

Wie viele Werte Sie brauchen: Planen Sie mindestens 15 Werte vorher und 15 Werte nachher ein. Mit weniger Werten können Sie nur Einzelfälle beschreiben.

Wenn Sie Vorjahresdaten haben: Werten Sie dieselben Kalenderwochen des Vorjahres aus. So sehen Sie, wie stark sich die Kennzahl allein durch die Jahreszeit bewegt.

Schritt 3: Der Messplan

Phase	Zeitraum	Quelle	Wer erfasst	Aufwand
Vorher (Baseline)	[Daten]	[System / Stichprobentage]	[Rolle]	[ ]
Einführung der KI-Arbeitsweise	[Daten]	–	–	–
Nachher (Nachmessung)	[Daten]	[gleiche Quelle wie vorher]	[Rolle]	[ ]
Optional: Haltbarkeit	[einige Wochen später]	[gleiche Quelle]	[Rolle]	[ ]

- Messen Sie vorher und nachher mit derselben Quelle und denselben Regeln.
- Messen Sie nachher nicht in der ersten Woche. Am Anfang ist alles neu, die Werte schwanken stark.
- Der Nachmess-Zeitraum steht fest. Verlängern Sie ihn nicht, auch wenn er wenige Werte liefert.
- Eine zweite Nachmessung einige Wochen später zeigt, ob der Effekt hält.

Die Erfassungstabelle

Tragen Sie jeden Fall in eine Zeile ein. Die Datei mess-vorlage.csv hat diese Spalten:

Spalte	Inhalt
Datum	Tag des Falls, TT.MM.JJJJ
Fall-ID	eindeutige Kennung, z. B. Auftrags- oder Ticketnummer, keine Personennamen
Phase_vorher_nachher	„vorher“ oder „nachher“
Start	Uhrzeit Beginn, hh:mm
Ende	Uhrzeit Ende, hh:mm
Dauer_min	Dauer in Minuten. Bei Durchlaufzeit Ende minus Start, bei Arbeitszeit der Wert des Timers
Nacharbeit_ja_nein	„ja“, wenn das Ergebnis nach der Freigabe noch korrigiert werden musste
Bemerkung	z. B. „ohne KI“, „Sonderfall“, „Systemausfall“

Die Datei trennt die Spalten mit Semikolon und ist für deutsche Tabellenprogramme eingerichtet. Löschen Sie die drei Beispielzeilen vor dem Gebrauch.

Schritt 4: Auswerten

Was Sie berechnen

Für jede Phase getrennt:

- **Median:** der mittlere Wert, wenn Sie alle Werte der Größe nach sortieren. Bei einer geraden Zahl von Werten ist es der Durchschnitt der beiden mittleren. Der Median wird von Ausreißern kaum verzerrt.
- **Spanne:** kleinster und größter Wert.
- **Anzahl der Werte (n).**

Sehen Sie sich zusätzlich alle Einzelwerte an, am besten als Punktdiagramm. Verzichten Sie auf Durchschnitte über verschiedene Aufgaben hinweg.

Formeln in Excel (Dauer in Spalte F, Werte vorher in F2:F19, nachher in F20:F35):

- Median: =MEDIAN(F2:F19)
- Anzahl: =ANZAHL(F2:F19)
- Nachher-Werte unter dem Baseline-Median (Baseline-Median in Zelle H2): =ZÄHLENWENN(F20:F35;"<"&H2)
- Oberes Quartil (für die vorsichtige Rechnung): =QUARTILE.INKL(F20:F35;3)

In anderen Tabellenprogrammen heißen die Funktionen ähnlich.

Wie Sie das Ergebnis einstufen

Einstufung	Bedingung
Deutlich	Mindestens 15 Werte vorher **und** mindestens 15 Werte nachher, **und** mindestens drei Viertel der Nachher-Werte liegen unter dem Baseline-Median. Zusätzlich darf der Vorjahresvergleich keine gleich große Veränderung ohne die neue Arbeitsweise zeigen.
Tendenz	Der Median ist gesunken, aber die Bedingungen für „deutlich“ sind nicht erfüllt.
Nicht belegt	alles andere

Seltene Aufgaben mit weniger als 15 Fällen je Phase beschreiben Sie nur als Einzelfälle, ohne Prozentzahl.

Leitplanke und Qualität prüfen

- Vergleichen Sie den Anteil der Fälle mit Nacharbeit vorher und nachher. Steigt er, ist die Zeitersparnis zum Teil eine Verschiebung in die Korrektur.
- Für die Qualität legen Sie alte und neue Ergebnisse gemischt und ohne Datum zwei Personen vor. Beide bewerten, ohne zu wissen, was mit KI entstand.

Schritt 5: Den Wert berechnen

> Zeitwert pro Woche = (Baseline-Median – Nachher-Median)
> × Stück pro Woche × Stundensatz

Die Mediane sind in Minuten. Teilen Sie das Ergebnis der Klammer mal Stückzahl durch 60, um Stunden zu erhalten.

- **Erwartet:** Rechnung mit dem Nachher-Median.
- **Vorsichtig:** dieselbe Rechnung mit dem oberen Quartil der Nachher-Werte statt dem Median.
- **Stück pro Woche:** die tatsächliche Zahl, nicht eine geplante.
- **Euro-Wert:** Zählen Sie nur die Zeit, die nachweislich in das vereinbarte Ziel geflossen ist („Gesparte Zeit geht in“). Den Rest nennen Sie „freigewordene Zeit“, und zwar nur in Stunden.
- **Jahreswert:** eine Hochrechnung unter der Annahme, dass das Niveau hält. Beschriften Sie ihn genau so.
- **Kosten:** Lizenzen, Einrichtung und die Zeit des Teams für Einführung und Messung. Stellen Sie sie dem Wert gegenüber.

Durchgerechnetes Beispiel

> **BEISPIEL MIT ERFUNDENEN ZAHLEN.** Alle Werte in diesem Abschnitt sind ausgedacht und dienen nur dazu, die Rechnung zu zeigen. Sie sind kein Messergebnis und keine Erwartung für Ihr Team.

Ablauf: Produktbeschreibung aus Herstellerdaten schreiben, bis der Text geprüft im Shop steht. Gemessen wird die Arbeitszeit an Stichprobentagen mit Timer.

Erhobene Werte (Minuten pro Text, sortiert):

Phase	Werte	n	Median	Spanne
Vorher	22, 24, 25, 26, 27, 27, 28, 29, 29, 30, 30, 31, 32, 33, 34, 35, 37, 40	18	29,5	22–40
Nachher	15, 16, 17, 18, 18, 19, 19, 20, 21, 21, 22, 23, 24, 26, 30, 31	16	20,5	15–31

Einstufung:

- Vorher 18 Werte, nachher 16 Werte: beide mindestens 15.
- Nachher-Werte unter dem Baseline-Median von 29,5 Minuten: 14 von 16. Drei Viertel wären 12.
- Vorjahresdaten liegen im Beispiel nicht vor. Das wird im Ergebnis genannt.
- **Ergebnis: deutlich.**

Leitplanke: Nacharbeit vorher bei 3 von 18 Texten, nachher bei 2 von 16. Der Anteil ist nicht gestiegen.

Wertrechnung (Beispielannahmen: 25 Texte pro Woche laut Shopsystem, Stundensatz 45 €):

	Rechnung	Stunden pro Woche	Zeitwert pro Woche
Erwartet	(29,5 – 20,5) × 25 = 225 Minuten	3,75 h	168,75 €
Vorsichtig (oberes Quartil 23,25)	(29,5 – 23,25) × 25 = 156,25 Minuten	2,60 h	117,19 €

Euro-Wert: Vereinbart war, dass die gesparte Zeit in die Überarbeitung älterer Produktseiten geht. Im Beispiel sind dort nachweislich 2 Stunden pro Woche eingeflossen. Euro-Wert: 2 h × 45 € = **90 € pro Woche**. Die übrigen 1,75 Stunden (erwartete Rechnung) werden als freigewordene Zeit genannt, ohne Euro-Betrag.

Hochrechnung aufs Jahr (Annahme: 44 Arbeitswochen, Niveau hält): 90 € × 44 = 3.960 €. Beschriftung: „Hochrechnung, falls das Niveau der Nachmessung hält“.

Kosten im Jahr (Beispielannahmen): Lizenzen für 2 Personen à 25 € im Monat = 600 €. Einrichtung und Messung 9 Stunden × 45 € = 405 €. Zusammen 1.005 €.

So klingt das Ergebnis:

> „Im Nachmess-Zeitraum dauerte eine Produktbeschreibung im Median 20,5 statt 29,5 Minuten (vorher 18 Texte, Spanne 22–40; nachher 16 Texte, Spanne 15–31). Einstufung: deutlich. Der Anteil an Nacharbeit ist nicht gestiegen. Vorjahresdaten lagen nicht vor, ein Einfluss der Jahreszeit ist daher nicht ausgeschlossen.“

Vermeiden Sie Aussagen wie „60 % produktiver“, Hochrechnungen über den gemessenen Ablauf hinaus und Herstellerzahlen als eigene Ergebnisse.

Was diese Messung nicht leisten kann

- Sie zeigt die Wirkung der neuen Arbeitsweise insgesamt, nicht einen reinen „KI-Effekt“. Neue Abläufe und Vorlagen wirken mit.
- Bei seltenen Aufgaben erreichen Sie in wenigen Wochen kaum genug Werte für „deutlich“.
- Ob die Wirkung über Monate hält, zeigt erst eine spätere zweite Nachmessung.

Wir bieten Praxistage an, nach denen Teams eine solche Messung mit festen Regeln für ihren wichtigsten Ablauf starten. Mehr dazu unter https://lakeworkai.com (https://lakeworkai.com).