WHITEPAPER · WIRKSAMKEIT VON KI

Eine Zeitersparnis ist kein Ertrag

Wie mittelständische Unternehmen den wirtschaftlichen Erfolg ihrer KI-Anwendungen nachprüfbar belegen, in sechs Nachweisstufen und fünf Nachweisgraden.

Ralf Platvoet · PPI – Platvoet Performance Intelligence · Stand: Oktober 2026

AUF EINEN BLICK

Die Werkzeuge funktionieren. In fast jedem Unternehmen, das KI einsetzt, bestätigen die Anwender eine Entlastung, und meistens haben sie recht. Trotzdem liegt in den seltensten Fällen eine Zahl vor, die eine Investitionsentscheidung tragen würde.

Das ist kein Messproblem. Es ist ein Zurechnungsproblem, und es entsteht lange vor der ersten Messung. Dieses Whitepaper beschreibt ein Verfahren, mit dem ein mittelständisches Unternehmen die Wirkung seiner KI-Anwendungen belegt, ohne ein neues IT-System zu beschaffen. Es benennt dabei auch die Grenzen des Nachweises, statt sie wegzurechnen.

Teil A richtet sich an die Geschäftsführung. Wohin das Geld fließt, was eine Freigabe tragen muss und welche drei Entscheidungen vor dem Start fallen müssen. Teil B richtet sich an Controlling und Projektleitung. Die sechs Nachweisstufen, die sechs Zurechnungsbedingungen, der Vergleich ohne Kontrollgruppe und eine Modellrechnung.

KERNSATZ 1

Eine Zeitersparnis ist kein Ertrag

Sie wird erst einer, wenn freigesetzte Kapazität abgebaut oder verkauft wird. Das ist eine Entscheidung und keine Eigenschaft der Technik.

KERNSATZ 2

Erst die Wirkung messen

Dann über ihre Verwendung entscheiden. Wer die Reihenfolge umkehrt, rechnet mit einer Wirkung, die auch negativ sein kann.

KERNSATZ 3

Die Basislinie ist nicht nachholbar

Wer sie überspringt, verliert den Nachweis dauerhaft und nicht nur vorübergehend. Sie kostet einen halben Tag.

KERNSATZ 4

Jede Zahl führt ihren Nachweisgrad mit

Was behauptet oder nur aus Annahmen gerechnet ist, erscheint nicht als Ergebnis. Fünf Grade genügen zur Einordnung.

TEIL A

Für die Geschäftsführung

Wohin die Ersparnis verschwindet, welche fünf Wege zum Geld führen, was eine Freigabe tragen muss und welche drei Entscheidungen vor dem Start fallen.

KAPITEL 1 · DIE LAGE

Wirkung wird gefordert, nicht nachgewiesen

Die Verbreitung von KI im Unternehmen ist belegt. Der Nutzen wird fast überall gesehen. Eine Zahl, die eine Investitionsentscheidung trägt, liegt fast nirgends vor.

Nach der DIHK-Digitalisierungsumfrage 2026, an der sich 4.686 Unternehmen aller Branchen und Größenklassen beteiligt haben, setzen 35 Prozent der Unternehmen KI ein. Weitere 34 Prozent planen den Einsatz innerhalb der nächsten drei Jahre. 31 Prozent haben weder das eine noch das andere vor.

Beim Produktivitätseinfluss bewerten 34 Prozent die Wirkung als hoch, 52 Prozent als moderat und 13 Prozent als gering. Rechnet man die ersten beiden zusammen, sehen 86 Prozent einen hohen oder moderaten Produktivitätseffekt. Die DIHK überschreibt das Kapitel mit dem Satz, KI sei als Produktivitätsfaktor in den Unternehmen angekommen.

35 %
der Unternehmen setzen KI ein, über alle Größenklassen

34 %
planen den Einsatz innerhalb der nächsten drei Jahre

86 %
sehen einen hohen oder moderaten Produktivitätseffekt (eigene Addition aus 34 und 52)

Das ist eine bemerkenswerte Konstellation. Die Wirkung wird fast überall gesehen. Eingesetzt wird KI von gut einem Drittel, ein weiteres Drittel plant es. Der Engpass ist also nicht die Überzeugung.

Er ist auch nicht die Technik. Die Anwendungsfelder mit der höchsten Verbreitung sind unspektakulär und erprobt. Die Erzeugung von Texten, Bildern und Code kommt auf 78 Prozent, Kundenansprache und Kundensupport auf 43 Prozent, Qualitätssicherung und Prozessüberwachung auf 38 Prozent. Das sind Felder, in denen die Werkzeuge seit Jahren tun, was sie sollen.

> Eine Einschränkung gehört dazu, und sie ist für dieses Whitepaper wichtig. Die Umfrage weist keinen eigenen Wert für den Mittelstand aus, und ihre Teilnehmerstruktur ist von kleinen Betrieben geprägt. 2.055 der 4.686 Unternehmen haben weniger als zehn Beschäftigte. Wer daraus eine mittelstandsspezifische Einsatzquote bildet, rechnet sie hoch. Nach der Regel dieses Whitepapers trägt eine solche Zahl Nachweisgrad 1 und erscheint deshalb hier nicht.

Der Engpass liegt dahinter. Er liegt bei der Frage, was der Einsatz gebracht hat.

Auf dem KI.Forum.NRW 2026 wurde das offen angesprochen. Nach den Beiträgen dort weist rund ein Drittel der Unternehmen überhaupt einen Ergebnisbeitrag ungleich null aus, und rund zwei Drittel der Piloten bleiben im

Pilotmodus. Diese Angaben stammen aus Mitschnitten einer Veranstaltung und nicht aus einer Erhebung. Sie taugen als Illustration und nicht als Beleg. Sie deckten sich allerdings mit dem, was in Projekten zu sehen ist.

Das Muster ist immer dasselbe. Alle fordern Wirkungsmessung. Niemand legt ein Verfahren vor.

> **Der wirtschaftliche Erfolg betrieblicher KI-Anwendungen ist nicht unklar, weil die Wirkung ausbleibt. Er ist unklar, weil kein Zurechnungsverfahren vorliegt.**

Das ist eine andere Diagnose als die üblichen, und sie führt zu anderen Maßnahmen. Wer glaubt, es fehle an Daten, kauft ein Werkzeug. Wer erkennt, dass es an Zurechnung fehlt, legt vor dem Start drei Dinge fest. Das kostet einen halben Tag und keine Lizenz.

KAPITEL 2 · DER MECHANISMUS

Warum die Zeitersparnis nicht in der Gewinn- und Verlustrechnung ankommt

Eine Sachbearbeitung braucht für einen Vorgang vorher 45 Minuten und nachher 32. Die Anwender bestätigen das, die Messung bestätigt es, und niemand bestreitet es. Wo sind die 13 Minuten?

Es gibt genau drei Möglichkeiten. Die freigesetzte Zeit wird abgebaut, dann sinken die Personalkosten. Sie wird verkauft, dann steigt die Menge bei gleicher Besetzung und es entsteht Deckungsbeitrag. Oder sie verteilt sich im Arbeitstag, dann ändert sich in der Gewinn- und Verlustrechnung nichts.

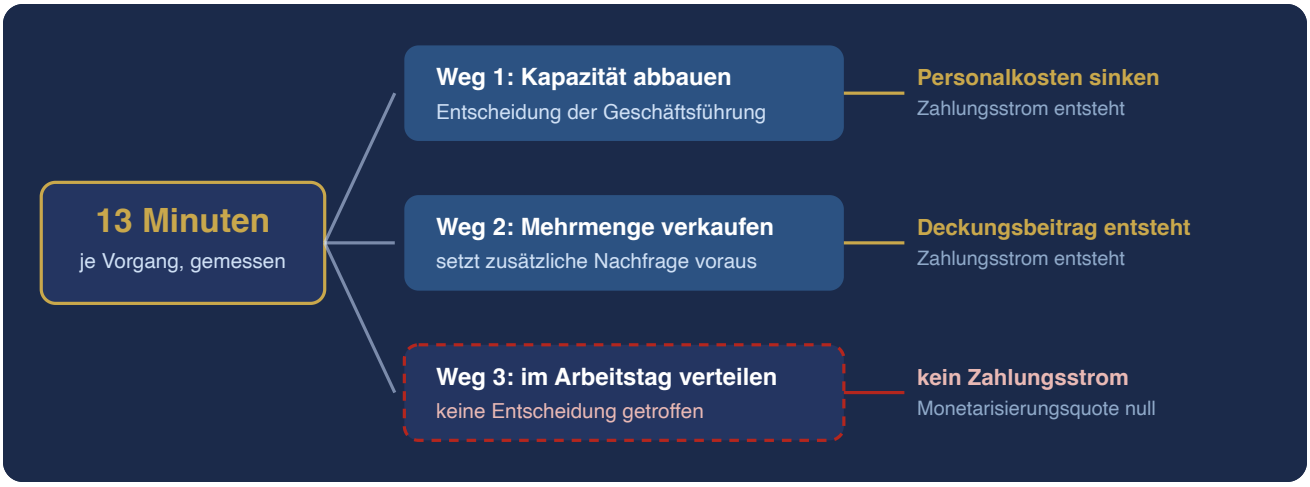


Abbildung 1: Drei Wege für eine gemessene Zeitersparnis. Nur zwei führen zu einem Zahlungsstrom, und beide verlangen eine Entscheidung. Der dritte Weg ist im Mittelstand der Regelfall.

Der dritte Fall ist der Regelfall. Und er ist kein Versäumnis der Mitarbeiter, sondern eine Folge der Struktur mittelständischer Unternehmen. Teams sind klein, Stellen sind schlecht teilbar, und eine Ersparnis von zwanzig Prozent an einer Halbtagsstelle ist nicht abbaubar. Damit verschiebt sich der ganze Fall zwingend auf die zweite Möglichkeit, also auf die Frage, ob zusätzliche Nachfrage vorliegt.

Daraus folgt eine Größe, die in kaum einem Business Case auftaucht. Die Monetarisierungsquote. Sie gibt an, welcher Anteil der technisch erzielten Wirkung tatsächlich zu einem Zahlungsstrom wird.

> **Freigesetzte Kapazität ist ein Angebot. Zum Ertrag wird sie erst durch eine Entscheidung, und die trifft nicht das Werkzeug.**

Die Größe ist unangenehm, weil sie den Blick von der Technik auf die Führung lenkt. Sie ist aber die Größe, an der das Ergebnis hängt, sobald die Wirkung gemessen ist. Eine Anwendung mit dreißig Prozent Zeitgewinn und einer Monetarisierungsquote von null erzeugt keinen Euro. Eine Anwendung mit acht Prozent Zeitgewinn, deren Gewinn als zusätzliche Aufträge abgewickelt wird, erzeugt Deckungsbeitrag. Der Zeitgewinn allein sagt darüber nichts.

Die Wirkung selbst kann dabei auch negativ ausfallen. Ein Feldversuch mit erfahrenen Entwicklern fand 2025 längere Bearbeitungszeiten mit KI-Unterstützung als ohne. Deshalb steht die Messung vor der Entscheidung und nicht umgekehrt.

In der Praxis fehlt an dieser Stelle meist nicht die Rechnung, sondern der Vermerk. Niemand hat entschieden und niemand hat aufgeschrieben, wofür ein Gewinn verwendet werden soll. Später lässt sich das nicht nachholen, weil dann niemand mehr unterscheiden kann, ob die Besetzung aus anderen Gründen gleich geblieben ist.

KAPITEL 3 · ORDNUNG

Fünf Wege zum Geld, und jeder bricht an einer anderen Stelle

Die üblichen Übersichten ordnen KI-Anwendungen nach dem Verfahren, also nach Sprachmodell, Klassifikation, Bildverarbeitung. Für die Bewertung des wirtschaftlichen Erfolgs ist diese Ordnung unbrauchbar.

Zwei Anwendungen derselben Technik führen über völlig verschiedene Pfade zu Geld, und sie scheitern an verschiedenen Stellen. Nützlich ist eine Ordnung nach der Wirkungsart. Sie beantwortet die Frage, über welchen Pfad aus der technischen Wirkung ein Zahlungsstrom wird. Fünf Typen reichen für den betrieblichen Bestand.

Typ	Wirkungsart	Monetarisierungspfad	Dominante Bruchstelle
T1	Kapazität	Abbau oder Verkauf freigesetzter Zeit	die Monetarisierung
T2	Vermeidung	entfallene Schadens- und Stillstandskosten	die Bewertung der vermiedenen Folge
T3	Qualität	weniger Ausschuss, Nacharbeit, Gewährleistung	die Basislinie
T4	Entscheidung	bessere Mittelverwendung	die Zurechnung überhaupt
T5	Erlös	Abschlussquote, Preis, Deckungsbeitrag	Gegenläufiges und die Marktseite

T1 Kapazitätswirkung

Ein Vorgang wird in weniger Zeit erledigt. Dazu gehören Textentwurf, Recherche, Angebotsvorbereitung, Zusammenfassung, Übersetzung und Codeunterstützung. Der Wirkungsgrad am Arbeitsplatz ist meist erhebbar, die Umwandlung in Geld ist es selten.

T2 Vermeidungswirkung

Ein Schaden tritt nicht ein. Dazu gehören Zustandsüberwachung, vorausschauende Instandhaltung, Absatzprognose und Anomalieerkennung. Der Ertrag ist eine Differenz zu einem Zustand, der nicht eingetreten ist. Ist die Ausfallfolge je Anlage nicht quantifiziert, fehlt die Bezugsgröße. Für die vorausschauende Instandhaltung zeigt unser Working Paper 02/2026, dass die nötigen Daten in aller Regel vorliegen und nur nie zu einer Zahl je Anlage verdichtet worden sind.

T3 Qualitätswirkung

Ein Fehler wird früher oder häufiger erkannt. Dazu gehören optische Prüfung, Qualitätsvorhersage, Eingangsprüfung und Plausibilisierung im Rechnungswesen. Hier liegt eine Falle, die regelmäßig zuschnappt. Unentdeckte Fehler stehen definitionsgemäß nicht in der Statistik. Der Erfolg zeigt sich deshalb zuerst als steigende erfasste Fehlerzahl. Wer das nicht vorher festlegt, deutet den Erfolg als Misserfolg.

T4 Entscheidungswirkung

Eine Entscheidung wird besser getroffen, ohne dass sich eine Menge oder eine Zeit ändert. Dazu gehören Betriebslagebild, Priorisierung, Investitionsauswahl und Frühwarnung. Der Ertrag besteht in der Differenz zu einer Wahl, die nicht getroffen wurde, und die ist nicht beobachtbar. T4 ist der Grenzfall des Verfahrens. Kapitel 10 behandelt ihn als solchen.

T5 Erlöswirkung

Mehr oder besser bezahlte Abschlüsse. Dazu gehören Angebotsqualität, Priorisierung im Vertrieb, Preisdifferenzierung und Empfehlung. Konjunktur, Wettbewerb und gleichzeitige Vertriebsmaßnahmen wirken auf dieselbe Größe. Bei zwanzig Angeboten im Quartal ist ein Effekt auf die Abschlussquote nicht von Zufall zu trennen.

> Die Einordnung ist keine Formalie. Sie legt fest, welche Größe zu messen ist und welche Bedingung am schwersten zu erfüllen sein wird. Eine Anwendung kann zwei Typen bedienen, etwa eine Prüfanwendung, die zugleich Prüfzeit spart. Dann werden die Wirkungen getrennt geführt und am Ende einmal zusammengeführt, damit nicht doppelt gezählt wird.

KAPITEL 4 · DER MASSSTAB

Was eine Freigabe tragen muss

Vor jeder Investitionsentscheidung steht eine Zahl. Entscheidend ist nicht ihre Höhe, sondern wie sie entstanden ist. Dafür reichen fünf Grade.

Grad	Bezeichnung	Bedingung	Zulässige Verwendung
0	behauptet	keine Messung	**keine. Nicht als Ergebnis ausweisen**
1	plausibilisiert	aus Annahmen gerechnet, ohne Messung am eigenen Fall	Bewertung vor dem Start, ausdrücklich als Annahme gekennzeichnet
2	gemessen	Basislinie, Bezugsgröße, Gegenläufiges und Gleichzeitiges erfasst, kein Vergleichsfall	interne Steuerung, kein Wirkungsnachweis
3	belegt	zusätzlich ein Vergleich über eine der Ersatzformen aus Kapitel 8	Wirkungsnachweis, Investitionsentscheidung
4	nachgewiesen	zusätzlich eine echte Vergleichsgruppe	Wirkungsnachweis, Veröffentlichung

Zwei Regeln machen daraus ein Werkzeug.

Erstens führt jede Ertragsaussage ihren Grad mit sich. Grad 0 und Grad 1 erscheinen nicht als Ergebnis. Grad 1 ist die legitime Form der Bewertung vor dem Start, und sie wird dort nicht verschwiegen, sondern benannt. Ein Business Case, der vor dem Projektstart Grad 1 trägt, ist in Ordnung. Ein Erfolgsbericht nach einem Jahr, der noch Grad 1 trägt, ist keiner.

Zweitens ist Grad 3 die realistische Obergrenze. Grad 4 verlangt zwei vergleichbare Einheiten, die sich im Mittelstand selten finden. Ein Verfahren, das Grad 4 fordert, wird nicht angewandt und ist damit wirkungslos. Grad 3 ist das Ziel.

> **Die Frage an jede Erfolgsmeldung lautet nicht, wie hoch der Nutzen ist. Sie lautet, welchen Grad die Aussage trägt.**

Diese zwei Regeln kosten nichts und wirken sofort. Sie sortieren jeden Anbieterprospekt in wenigen Minuten. Eine Angabe ohne Basislinie und ohne Vergleichsfall trägt Grad 0 oder Grad 1, unabhängig davon, wie viele Nachkommastellen sie hat.

KAPITEL 5 · VOR DEM START

Die drei Entscheidungen, die nicht nachholbar sind

Von den sechs Zurechnungsbedingungen, die Kapitel 7 im Einzelnen beschreibt, lassen sich vier nachholen. Zwei nicht. Dazu kommt eine Festlegung, die nichts kostet, wenn man sie früh trifft, und später unmöglich ist.

Erste Entscheidung: die Basislinie

Vor dem Einsatz wird gemessen, wie lange ein Vorgang dauert, wie viele es sind und wie hoch die Fehlerquote ist. Eine Stichprobe genügt, flächige Erhebung ist nicht nötig. Entscheidend ist der Zeitpunkt. Nach dem Start lässt

sich eine Basislinie nicht rekonstruieren, weil die Beteiligten den neuen Weg dann schon kennen. Bei Qualitätsanwendungen gehört zur Basislinie zusätzlich der Erkennungsgrad, also welcher Anteil der Fehler vorher überhaupt auffiel.

Zweite Entscheidung: der Verwendungsvermerk

Wer entscheidet, dass freigesetzte Kapazität abgebaut oder verkauft wird, und wann? Ein Satz im Entscheidungsvermerk reicht. Fehlt er, steht in der Monetarisierungsspalte später null, und zwar zu Recht.

Dritte Entscheidung: die Vergleichsanlage

Die Einführung findet ohnehin nacheinander statt, in Teams, an Standorten oder nach Vorgangsarten. Wer diese Reihenfolge vorher festhält und die später startende Einheit als Vergleich behandelt, hat einen Vergleichsfall ohne Verzicht und ohne Zusatzkosten. Wer sie nicht festhält, hat am Ende zwei Gruppen, die sich in zu vielen Punkten unterscheiden.

Entscheidung	Aufwand	Nachholbar
Basislinie messen	ein halber Tag mit Controlling und Fachbereich	**nein**
Verwendung festlegen	ein Satz im Entscheidungsvermerk	**nein**
Reihenfolge der Einführung festhalten	eine Zeile im Projektplan	**nein**

Alle drei sind Entscheidungen und keine Investitionen. Keine davon hängt an einem System, an einer Lizenz oder an einem Berater. Sie hängen daran, dass jemand vor dem Start eine halbe Stunde darüber nachdenkt, woran der Erfolg später erkennbar sein soll.

> **Nicht die Messbarkeit fehlt im Mittelstand. Es fehlt die Entscheidung, vorher zu messen und vorher festzulegen, wofür ein Gewinn verwendet wird.**

TEIL B

Für Controlling und Projektleitung

Sechs Nachweisstufen, sechs Zurechnungsbedingungen, vier Ersatzformen für den fehlenden Vergleichsfall, die Erhebungstabelle und eine Modellrechnung.

KAPITEL 6 · DAS GERÜST

Sechs Nachweisstufen

Eine Wirkung läuft durch mehrere Stationen, bis sie im Ergebnis ankommt. An jeder Station lässt sich prüfen, ob sie dort angekommen ist.

Diese Stationen heißen hier Nachweisstufen, weil sie Messgrößen tragen und keine Verlustmechanismen. Die Stufen sind keine Reifegrade. Ein Unternehmen durchläuft sie nicht nacheinander, es belegt sie nacheinander. Die Reihenfolge ist eine Belegreihenfolge, und sie sagt nichts über die Qualität der Organisation.

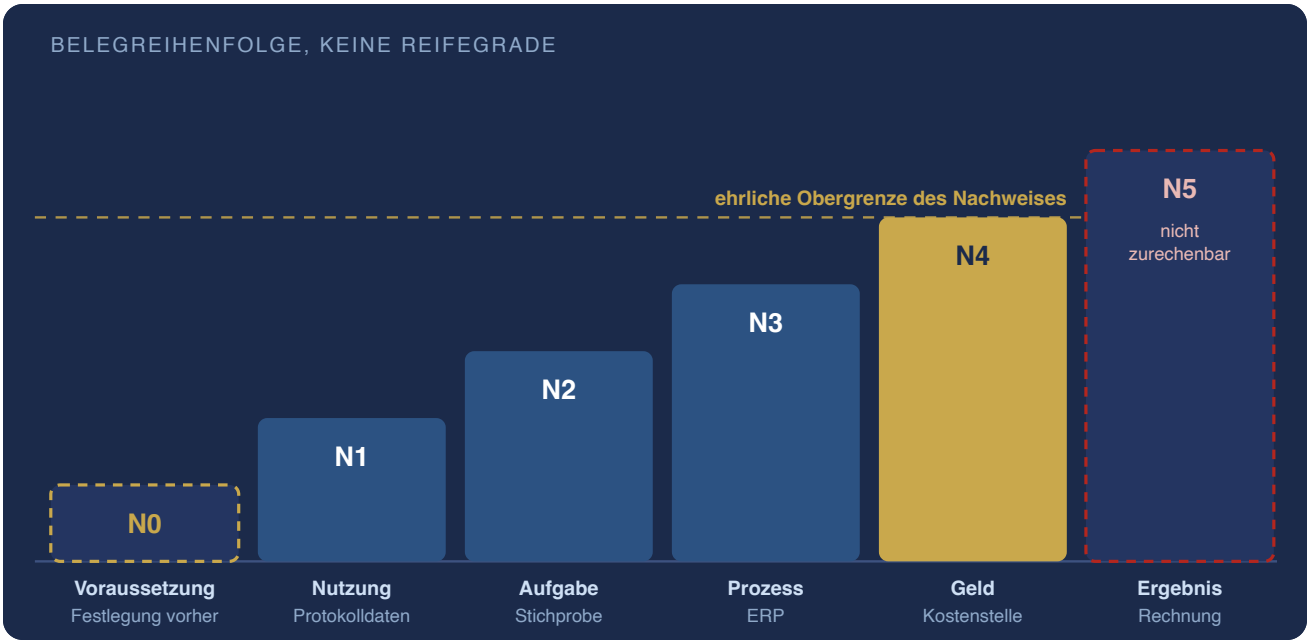


Abbildung 2: Die sechs Nachweisstufen. N0 hängt an einer Festlegung vor dem Start und nicht an einem System. N4 ist die ehrliche Obergrenze. N5 wird gerechnet und nicht gemessen, und das Ergebnis trägt den Nachweisgrad von N4.

N0 Voraussetzung

Was vor dem Einsatz gegeben sein muss, damit überhaupt etwas messbar wird. Dazu gehören die Verfügbarkeit der benötigten Daten, eine Beschreibung des betroffenen Vorgangs, die Zahl qualifizierter Beschäftigter und eine Vorperiodenmessung. N0 ist die einzige Stufe, die vor dem Start abgeschlossen sein muss. Wird sie übersprungen, ist Grad 2 die Obergrenze für alles Folgende, weil die Basislinie fehlt.

N1 Nutzung

Ob das Werkzeug benutzt wird, und in welchem Umfang. Größen sind der Abdeckungsgrad, also der Anteil der betroffenen Vorgänge, die tatsächlich über das Werkzeug laufen, die Nutzungsfrequenz, der Anteil aktiver an berechtigten Nutzern und der Anteil der Vorgänge mit Rückfall auf den alten Weg.

Diese Stufe ist die am besten erhebbare, weil die Protokolldaten des Werkzeugs sie hergeben. Das setzt voraus, dass die Auswertbarkeit bei der Beschaffung mitvereinbart wurde. Ist sie es nicht, liegt ein Beschaffungsfehler vor, und er gehört in die Anbieterauswahl des nächsten Projekts.

N2 Aufgabenwirkung

Ob der einzelne Vorgang schneller oder besser wird. Größen sind die Bearbeitungszeit je Vorgang, die Qualität je Vorgang, der Prüf- und Korrekturaufwand und der Anteil der Ausgaben, die unverändert übernommen werden.

Der Prüf- und Korrekturaufwand ist die Größe, die am häufigsten fehlt. Ohne sie wird eine Verlagerung als Ersparnis ausgewiesen. Wer zehn Minuten beim Schreiben gewinnt und sechs beim Prüfen verliert, hat vier gewonnen und nicht zehn. Die Trennung beider Zeiten ist deshalb zwingend und nicht optional. Erhebbar ist diese Stufe über eine Stichprobe, nicht über flächige Zeitmessung. Vorher und nachher dieselbe Vorgangsart, dieselbe Zählweise, dieselbe Person, die zählt.

N3 Prozesswirkung

Ob der Prozess als Ganzes schneller, fehlerärmer oder leistungsfähiger wird. Größen sind die Durchlaufzeit des Gesamtvorgangs, die Fehler- und Nacharbeitsquote, die Zahl bearbeiteter Vorgänge je Kopf und die ausgewiesene freie Kapazität.

Hier greift die Durchsatzkopplung. Der Gewinn aus N2 schlägt nur in dem Maß durch, in dem der beschleunigte Schritt am Gesamtvorgang beteiligt ist und kein anderer Schritt den Engpass bildet. Eine um dreißig Prozent schnellere Angebotserstellung verkürzt die Durchlaufzeit nicht um dreißig Prozent, wenn danach vier Tage auf eine Freigabe gewartet wird. Freie Kapazität ist auf dieser Stufe nur dann eine Größe, wenn ihr eine Verwendungsangabe beiliegt. Ohne sie ist sie eine Behauptung.

N4 Monetäre Wirkung

Ob aus der Prozesswirkung ein Zahlungsstrom wird. Auf der Habenseite stehen Personalkosten der betroffenen Stellen, Sachkosten des Vorgangs, Deckungsbeitrag zusätzlich abgewickelter Vorgänge, entfallene Schadens- und Ausschusskosten und gebundenes Kapital.

Auf der Sollseite steht die Kostenseite der Anwendung, und sie steht vollständig. Lizenz, Einführung, Anpassung, Prüf- und Nacharbeitsaufwand, Betrieb und Kompetenzaufbau. Diese Positionen gehören auf eine eigene Kostenstelle, sonst sind sie am Jahresende nicht mehr auffindbar.

N4 ist die Stufe, an der die Entscheidung über die Verwendung sichtbar wird oder eben nicht. Bleibt die Personalkostenzeile unverändert und ist keine Mehrleistung verkauft, ist die Monetarisierungsquote null, unabhängig davon, wie gut N2 und N3 ausfallen. Die Zuordnung zur Anwendung ist auf dieser Stufe eine Rechnung und keine Ablesung. Sie ist als solche offenzulegen.

N5 Unternehmenswirkung

Ob sich das Unternehmensergebnis ändert. Größen sind Periodenergebnis, operativer Cashflow und Kapitalwert des Vorhabens. Bei einem einzelnen Vorhaben im Mittelstand ist diese Stufe in aller Regel nicht zurechenbar. Der Effekt liegt unter der Schwankungsbreite des Ergebnisses. Das ist kein Mangel des Verfahrens, sondern ein Befund, und Kapitel 10 behandelt ihn. Die ehrliche Obergrenze des Nachweises liegt bei N4.

> Außerhalb der Stufen stehen aufgebaute Fähigkeit und gewonnene Handlungsoption. Beide sind wirtschaftlich bedeutsam und werden hier nicht in Geld umgerechnet. Eine Umrechnung trüge Grad 1 und dürfte nach der Regel aus Kapitel 4 nicht als Ergebnis erscheinen. Beides wird benannt und nicht bewertet.

KAPITEL 7 · ZURECHNUNG

Sechs Bedingungen, damit die Zahl trägt

Eine gemessene Veränderung ist noch keine zugerechnete Wirkung. Dazwischen liegen sechs Bedingungen. Jede ist eine Frage, die mit Ja oder Nein beantwortet wird.

#	Bedingung	Prüffrage	Erfüllbar im Mittelstand
Z1	Basislinie	Liegen Vorperiodenwerte vor, die **vor** dem Start festgelegt wurden?	ja, wenn früh geplant. Nachträglich nicht herstellbar
Z2	Bezugsgröße	Ist das Mengengerüst konstant oder korrigiert, so dass Wachstum nicht als Wirkung erscheint?	ja
Z3	Gegenläufiges	Sind Lizenz, Einführung, Anpassung, Prüfaufwand, Nacharbeit, Betreuung und Kompetenzaufbau erfasst?	ja, verlangt eine eigene Kostenstelle
Z4	Gleichzeitiges	Sind andere Maßnahmen im selben Zeitraum benannt, die auf dieselbe Größe wirken?	ja, verlangt Disziplin, keine Technik
Z5	Vergleichsfall	Liegt eine Vergleichsgrundlage vor, echt oder als Ersatzform nach Kapitel 8?	Ersatzform ja, echte Kontrollgruppe meist nein
Z6	Verwendungsentscheidung	Ist dokumentiert, wer entschieden hat, dass freigesetzte Kapazität abgebaut oder verkauft wird?	ja, wird aber fast nie getan

Z1 und Z6 müssen vor dem Start erfüllt sein. Alle anderen lassen sich nachholen. Deshalb beginnt die Arbeit an der Wirksamkeit bei der Bewertung und nicht bei der Messung.

Z2 wird am häufigsten übersehen. Wächst die Auftragszahl um zwölf Prozent, während die Anwendung läuft, steigt die Zahl der bearbeiteten Vorgänge je Kopf auch ohne jede Wirkung. Wer durch die Menge teilt, hat das Problem gelöst. Wer es nicht tut, verbucht Wachstum als Produktivität.

Z3 entscheidet darüber, ob die Rechnung ehrlich ist. Die Kostenseite einer KI-Anwendung besteht nicht aus der Lizenz. Sie besteht aus der Lizenz, der Einführung, der Anpassung, der laufenden Prüfung der Ausgaben, der Nacharbeit, der Betreuung der Nutzer und dem Kompetenzaufbau. Vier dieser sieben Positionen sind Arbeitszeit und erscheinen in keiner Rechnung, solange niemand sie bucht.

Z6 wird nach unserer Projekterfahrung am seltensten erfüllt. Die Bedingung verlangt keine Technik und keinen Aufwand. Sie verlangt einen Satz im Protokoll.

KAPITEL 8 · DER VERGLEICHSFALL

Vergleich ohne Kontrollgruppe

Eine echte Kontrollgruppe setzt zwei vergleichbare Einheiten voraus, von denen eine ohne das Werkzeug weiterarbeitet. Im Mittelstand gibt es sie selten. Vier Ersatzformen bleiben, und drei davon kosten nichts.

#	Form	Vorgehen	Aussagekraft	Grenze
E1	Gestaffelte Einführung	Einführung nacheinander in Teams, Standorten oder Vorgangsarten	hoch	verlangt zwei vergleichbare Einheiten und eine vertretbare Verzögerung
E2	Interne Vergleichsgruppe	Ein Teil der Vorgänge läuft weiter auf dem alten Weg	hoch	Selbstselektion der Nutzer, schwer durchzuhalten
E3	Unterbrechungstest	Das Werkzeug wird für einen begrenzten Zeitraum abgeschaltet	mittel bis hoch	organisatorisch unbeliebt, Gewöhnung wirkt nach
E4	Vorher und Nachher mit Mengenkorrektur	Perioden vergleichen, korrigiert um Mengengerüst und benanntes Gleichzeitiges	mittel	trägt nur, wenn Z2 und Z4 sauber erfüllt sind

E1 ist die erste Wahl, und zwar aus einem einfachen Grund. Die Einführung findet ohnehin nacheinander statt. Sie muss nur als Vergleich angelegt werden. Das kostet nichts außer der Festlegung vorher.

E3 wird unterschätzt. Ein Werkzeug für zwei Wochen abzuschalten klingt nach Rückschritt. Es ist aber der schnellste Weg zu einer belastbaren Aussage, wenn die Basislinie fehlt. Und es beantwortet eine Frage, die sonst offen bleibt, nämlich ob die Beteiligten das Werkzeug vermissen würden.

E4 ist die schwächste Form und die am häufigsten verwendete. Sie trägt, wenn das Mengengerüst korrigiert ist und die gleichzeitigen Maßnahmen benannt sind. Ohne beides trägt sie nicht.

KAPITEL 9 · ERHEBUNG

Welche Größe woher kommt

Die Frage, die über die Umsetzbarkeit entscheidet, lautet nicht, welche Größe ideal wäre. Sie lautet, welche Größe ein mittelständisches Unternehmen ohne neues IT-System erheben kann.

Dafür genügen acht Orte im Unternehmen.

Kürzel	Datenquelle
D1	Protokolldaten des Werkzeugs
D2	Stichprobenmessung am Vorgang, von Hand
D3	ERP, Warenwirtschaft, Auftragsverwaltung
D4	Instandhaltungs- und Stördokumentation, Leitstand
D5	Qualitäts-, Reklamations- und Gewährleistungsdokumentation
D6	CRM, Angebots- und Kontaktverwaltung
D7	Buchhaltung und Kostenstellenrechnung
D8	dokumentierte Festlegung, also Beschaffungsvereinbarung, Entscheidungsvermerk, Protokoll

D1 bis D7 sind Systeme oder Messungen, die im Mittelstand typischerweise vorliegen. D8 ist kein System, sondern eine Disziplin. Dieser Unterschied trägt den wichtigsten Befund der folgenden Tabelle.

Die Tabelle nennt je Stufe und Typ eine Leitgröße und ihre Quelle. Ein Vermerk Festl. bedeutet, dass die Größe erhebbar ist, aber nur bei Festlegung vor dem Start.

Stufe	T1 Kapazität	T2 Vermeidung	T3 Qualität	T4 Entscheidung	T5 Erlös
N0	Zeit und Menge je Vorgangsart, Verwendungsvermerk (D2, D3, D8) *Festl.*	Ereignisrate und bewertete Folge je Anlage (D4, D7) *Festl.*	Fehlerquote mit Fehlerdefinition und Erkennungsgrad (D5, D8) *Festl.*	Entscheidungsliste der Vorperiode mit Grundlage (D8) *Festl.*	Abschlussquote, Preisniveau, Liste gleichzeitiger Maßnahmen (D6, D7, D8) *Festl.*
N1	Abdeckungsgrad, Rückfallquote (D1, D3) ja	Abdeckung des Anlagenbestands, Quittierungsquote (D1, D4) ja	Anteil geprüfter an prüfpflichtigen Stücken (D1, D5) ja	Abrufhäufigkeit, Heranziehung im Vermerk (D1, D8) *Festl.*	Anteil gestützter Angebote und Kontakte (D1, D6) ja
N2	Bearbeitungszeit, davon getrennt Prüf- und Korrekturzeit (D2) ja	Trefferquote, Fehlalarmquote, Vorwarnzeit (D1, D4) ja	Erkennungs- und Falschausschussquote am Prüfsatz (D2, D5) ja	Zeit bis zur Entscheidung, Zahl geprüfter Alternativen (D8) *Festl.*	Angebotsdurchlaufzeit, Antwortquote (D2, D6) ja
N3	Durchlaufzeit, Vorgänge je Kopf, ausgewiesene freie Kapazität (D3, D8) ja	ungeplante Stillstandszeit, Verhältnis geplant zu ungeplant (D3, D4) ja	Ausschuss-, Nacharbeits- und Durchrutschquote (D3, D5) ja	Mittelanteil auf vorrangige Vorhaben, Reaktionszeit (D7, D8) *Festl.*	Abschlussquote, Auftragseingang je Vertriebskopf (D3, D6) ja

Stufe	T1 Kapazität	T2 Vermeidung	T3 Qualität	T4 Entscheidung	T5 Erlös
N4	Personalkosten der Kostenstelle, Deckungsbeitrag der Mehrmenge, Monetarisierungsquote (D7, D8) ja	bewertete entfallene Ausfallfolge (D4, D7) *Festl.*	entfallene Fehlerfolgekosten gegen Kosten des Falschausschusses (D5, D7) ja	**nicht zurechenbar, Ersatz nur als Bandbreite mit Grad 1**	Deckungsbeitrag der Mehrabschlüsse gegen Preiszugeständnisse (D6, D7) *Festl.*
N5	**nicht zurechenbar**	**nicht zurechenbar**	**nicht zurechenbar**	**nicht zurechenbar**	**nicht zurechenbar, höchstes Risiko einer Scheinzurechnung**

Drei Befunde stehen in dieser Tabelle.

Die Zeile N0 trägt nirgends ein Ja. Keine der fünf Zellen ist ohne Festlegung vor dem Start erhebbar. Und keine hängt an einem fehlenden System. Sie hängen an einer Entscheidung.

Die Obergrenze des Nachweises hängt vom Typ ab. T1 bis T3 und T5 erreichen N4. T4 erreicht N3, weil der Vergleichsfall einer besseren Entscheidung nicht beobachtbar ist.

Erhebbarkeit und Nachweisgrad sind zwei verschiedene Fragen. Die Tabelle sagt, ob eine Größe zu bekommen ist. Sie sagt nicht, welchen Grad die daraus gebildete Aussage trägt. Eine erhebbare Größe auf N3 ohne Vergleichsfall ergibt Grad 2 und damit keinen Wirkungsnachweis.

KAPITEL 10 · GRENZEN

Grenzen, die benannt werden statt weggerechnet

Ein Verfahren, das für jede Frage eine Zahl liefert, ist unseriös. Vier Stellen bleiben offen, und sie bleiben offen aus strukturellen Gründen.

Die Unternehmenswirkung ist bei einem einzelnen Vorhaben nicht zurechenbar

Der Beitrag einer Anwendung zum Periodenergebnis liegt unter der Schwankungsbreite dieses Ergebnisses. Er wird nicht gemessen, sondern aus den Größen der Stufe N4 gerechnet, und das Ergebnis trägt den Grad von N4 und keinen höheren. Wer eine Ergebnisveränderung einer Anwendung zuschreibt, rechnet nicht, er behauptet.

Die Entscheidungswirkung ist der Grenzfall

Bei T4 ist der Vergleichsfall die Entscheidung, die nicht getroffen wurde. Als Ersatz bleibt eine Bandbreite gegenüber der nächstbesten, zum Zeitpunkt dokumentierten Alternative. Sie trägt Grad 1 und erscheint deshalb nicht als Ergebnis. Diese Grenze lässt sich nicht durch mehr Daten verschieben.

Fähigkeit und Option werden nicht bewertet

Ein Unternehmen, das zwanzig Mitarbeiter im Umgang mit diesen Werkzeugen sicher gemacht hat, hat etwas Wertvolles erworben. Jede Umrechnung in Euro wäre eine Setzung. Beides wird benannt, getrennt vom Ertrag, und ohne Zahl.

Doppelzählung droht auf N4

Eine Prüfanwendung wirkt als Qualitätsanwendung und spart zugleich Prüfzeit. Beide Wirkungen laufen in dieselbe Kostenstelle. Sie werden bis N3 getrennt geführt und auf N4 einmal zusammengeführt. Die eingesparte Prüfzeit wird genau einem der beiden Typen zugerechnet, und die Zuordnung wird dokumentiert.

> **Eine Grenze, die im Bericht steht, kostet Glaubwürdigkeit nicht. Eine Grenze, die jemand später findet, kostet sie vollständig.**

KAPITEL 11 · MODELLRECHNUNG

Derselbe Fall, zwei Ergebnisse

> Die folgende Rechnung ist eine Modellrechnung mit Beispielwerten. Sie trägt Nachweisgrad 1. Alle Annahmen stehen offen, keine stammt aus einer Studie, und die Rechnung ersetzt keine betriebsindividuelle Kalkulation. Sie zeigt nur, wie weit zwei Ergebnisse auseinanderliegen, wenn sich eine einzige Entscheidung unterscheidet.

Angenommener Fall. Kaufmännische Sachbearbeitung, Angebotserstellung. 1.200 Vorgänge im Jahr. Bearbeitungszeit vorher 45 Minuten, nachher 32 Minuten. Zusätzlich fallen 4 Minuten je Vorgang für die Prüfung der Ausgaben an. Vollkostensatz 55 Euro je Stunde. Kosten der Anwendung 9.000 Euro im Jahr, einschließlich Lizenz, Einführung und Betreuung.

Schritt 1, die Aufgabenwirkung

45 minus 32 ergibt 13 Minuten. Davon gehen 4 Minuten Prüfzeit ab. Netto bleiben 9 Minuten je Vorgang. Wer die Prüfzeit nicht trennt, weist 13 aus und übertreibt um 44 Prozent.

Schritt 2, die Prozesswirkung

9 Minuten mal 1.200 Vorgänge ergibt 10.800 Minuten, also 180 Stunden im Jahr. Bezogen auf eine Vollzeitstelle sind das etwa elf Prozent der Jahresarbeitszeit.

Schritt 3, die monetäre Wirkung in zwei Fällen

	Fall A: keine Entscheidung	Fall B: Verwendung festgelegt
Freigesetzte Zeit	180 Stunden	180 Stunden
Verwendung	verteilt sich im Arbeitstag	180 Stunden gehen in zusätzliche Angebote
Monetarisierungsquote	0 Prozent	angenommen 60 Prozent
Bewerteter Nutzen	0 Euro	108 Stunden, als Deckungsbeitrag der Mehrmenge bewertet
Kosten der Anwendung	9.000 Euro	9.000 Euro
Ergebnis	**minus 9.000 Euro**	abhängig vom Deckungsbeitrag je Vorgang

In Fall A steht ein negatives Ergebnis, obwohl die Anwendung technisch einwandfrei funktioniert und die Anwender zufrieden sind. In Fall B steht ein Ergebnis, das vom Deckungsbeitrag der zusätzlichen Vorgänge abhängt und nicht vom Zeitgewinn. Die Rechnung bricht hier bewusst ab, weil der Deckungsbeitrag je Vorgang betriebsindividuell ist und jede eingesetzte Zahl Grad 0 tragen würde.

Der Unterschied zwischen beiden Fällen ist kein technischer. Er ist eine Zeile im Entscheidungsvermerk.

Eine Nebenrechnung macht den Punkt noch deutlicher. Hätte man in Fall A auf die Trennung der Prüfzeit verzichtet und zusätzlich mit dem vollen Zeitgewinn gerechnet, stünden 13 Minuten mal 1.200 Vorgänge mal 55 Euro im Bericht, also rund 14.300 Euro Nutzen gegen 9.000 Euro Kosten. Dieselbe Anwendung, dieselbe Periode, und ein Vorzeichenwechsel im Ergebnis. Beide Zahlen wären gerechnet. Nur eine wäre zugerechnet.

KAPITEL 12 · AUS DER PROJEKTPRAXIS

Die Einwände, die in jedem Projekt kommen

ZUM AUFWAND

„Dafür haben wir keine Kapazität. Wir sind froh, wenn das Werkzeug überhaupt läuft.“

Der Aufwand liegt niedriger als erwartet, weil er zum größten Teil vor dem Start anfällt. Eine Stichprobenmessung an einer Vorgangsart, ein Satz im Vermerk, eine Zeile im Projektplan. Zusammen ein halber Tag. Danach fallen die Größen aus Systemen an, die ohnehin laufen. Teuer ist nicht die Messung. Teuer ist ein zweites Projekt, das man beginnt, ohne das erste bewerten zu können.

„Unsere Leute sagen, es hilft. Reicht das nicht?“

Für die Entscheidung, das Werkzeug weiter zu betreiben, reicht es oft. Für eine Investitionsentscheidung reicht es nicht, und für die Frage, welche von vier Anwendungen ausgebaut wird, erst recht nicht. Zustimmung der Anwender ist ein Indiz für Nutzung und damit für Stufe N1. Sie sagt nichts über N3 und N4.

„Das klingt nach Wissenschaft. Wir brauchen etwas Praktisches.“

Praktisch ist die Checkliste im nächsten Kapitel. Die Begriffe dieses Whitepapers stammen aus einem Working Paper, das die Belege und das Rechenmodell dazu enthält. Für die Anwendung im Betrieb reichen zwölf Zeilen.

ZUR MESSUNG

„Wir haben keine Basislinie. Das Projekt läuft seit einem Jahr.“

Dann ist Grad 2 die Obergrenze, und das Verfahren sagt es offen. Zwei Wege bleiben. Entweder der Unterbrechungstest, also eine befristete Abschaltung mit Messung in beiden Zuständen. Oder die Basislinie wird für die nächste Anwendung sauber angelegt, und für die laufende begnügt man sich mit einer Messung ohne Wirkungsnachweis. Beides ist ehrlicher als eine nachträglich rekonstruierte Vorperiode.

„Unsere Anwendung lässt sich nicht gestaffelt einführen.“

Dann prüfen Sie die Staffelung nach Vorgangsart statt nach Team. Selten läuft ein Werkzeug von Tag eins über alle Vorgangsarten. Fast immer gibt es eine Reihenfolge, und fast immer ist sie nur nicht aufgeschrieben.

„Wenn wir die Prüfzeit mitrechnen, bleibt kaum etwas übrig.“

Das ist ein Ergebnis und kein Einwand. Es ist genau die Information, die eine Entscheidung braucht. Bei Anwendungen, deren Ausgaben vollständig geprüft werden müssen, liegt der Nettogewinn systematisch niedriger. Das spricht nicht gegen das Werkzeug. Es spricht gegen den Einsatz an einer Stelle, an der jede Ausgabe geprüft werden muss.

„Der Anbieter liefert ein Dashboard mit ROI-Anzeige.“

Prüfen Sie zwei Dinge. Woher kommt die Basislinie, und wo steht der Prüfaufwand? Fehlt eines von beiden, zeigt das Dashboard eine Hochrechnung aus Nutzungsdaten. Das ist Grad 1, verlässlich berechnet und als Ergebnis unbrauchbar.

ZUR VERWENDUNG DES GEWINNS

„Kapazitätsabbau ist bei uns nicht durchsetzbar.“

Dann ist der zweite Monetarisierungspfad der richtige, also zusätzliche Menge bei gleicher Besetzung. Das ist eine legitime Entscheidung und muss nur eine sein. Die Alternative besteht nicht darin, die Frage offen zu lassen. Sie besteht darin, sie zu beantworten und das Ergebnis zu akzeptieren, auch wenn es null lautet.

„Wir wollen keine Zahl, die am Ende gegen das Projekt spricht.“

Verständlich und teuer. Eine Zahl, die niemand erhebt, existiert trotzdem. Sie erscheint nur später und in einem schlechteren Moment, meist wenn das zweite Vorhaben zur Freigabe ansteht und niemand das erste belegen kann. Ein offen ausgewiesenes Ergebnis von null bei einer Anwendung schützt die Freigabe für die nächsten drei.

KAPITEL 13 · ZUSAMMENFASSUNG

Die Checkliste vor dem Start

Wer diese zwölf Punkte beantwortet, erreicht Nachweisgrad 3, und zwar unabhängig davon, welches Werkzeug und welcher Anbieter gewählt wird.

- [x] Die Anwendung ist einem der fünf Typen zugeordnet, und bei zwei Typen ist festgelegt, wie doppelt gezählte Wirkung behandelt wird.
- [x] Die dominante Bruchstelle dieses Typs ist benannt, und die zugehörige Zurechnungsbedingung ist bearbeitet.
- [x] Die Basislinie ist vor dem Start gemessen, mit Zeit, Menge und Qualität je Vorgangsart.
- [x] Bei Qualitätsanwendungen gehört der Erkennungsgrad der Basislinie dazu, damit ein Erkennungsgewinn später nicht als Verschlechterung erscheint.
- [x] Es ist dokumentiert, wer entschieden hat, wofür freigesetzte Kapazität verwendet wird.
- [x] Die Reihenfolge der Einführung steht im Projektplan und ist als Vergleich angelegt.
- [x] Die Auswertbarkeit der Protokolldaten ist im Beschaffungsvertrag vereinbart, nicht nachträglich erbeten.
- [x] Prüf- und Korrekturzeit werden getrennt von der Bearbeitungszeit gemessen.
- [x] Die Anwendung hat eine eigene Kostenstelle, auf der Lizenz, Einführung, Anpassung, Betreuung und Kompetenzaufbau zusammenlaufen.
- [x] Das Mengengerüst der Vergleichsperioden ist korrigiert, und gleichzeitige Maßnahmen auf dieselbe Größe sind benannt.
- [x] Jede Ertragsaussage im Bericht trägt ihren Nachweisgrad, und Grad 0 und 1 erscheinen nicht als Ergebnis.
- [x] Die Grenzen sind im Bericht benannt, insbesondere die fehlende Zurechenbarkeit auf Unternehmensebene.

> **Die Technik ist die einfache Hälfte. Über den Ertrag entscheiden drei Festlegungen, die vor dem Start getroffen werden.**

ANHANG

Quellen

Nr.	Quelle
1	DIHK, *Künstliche Intelligenz, Souveränität und Resilienz. DIHK-Digitalisierungsumfrage 2026*, Stand Januar 2026, Erhebungszeitraum 10. bis 28. November 2025, Beteiligung 4.686 Unternehmen. Alle Werte am Primärdokument geprüft
2	KI.Forum.NRW 2026, Mitschnitte der Veranstaltung. Im Text ausdrücklich als Praxisillustration gekennzeichnet, nicht als Beleg
3	PPI Working Paper 08/2026, *Der Ertragsnachweis. Bewertung und Nachprüfung des wirtschaftlichen Erfolgs betrieblicher KI-Anwendungen im mittelständischen Unternehmen*, in Vorbereitung. Quelle der Nachweisstufen, Zurechnungsbedingungen, Ersatzformen, Nachweisgrade und Anwendungstypen
4	PPI Working Paper 02/2026, *Die Wirkungskette in Predictive-Maintenance-Projekten*. Quelle des Befundes zur Bewertung der Ausfallfolge
5	PPI Working Paper 07/2026, *Die Wirkung von Betriebslagebildern*. Quelle der Begründung, warum keine aggregierte Wirkungskennzahl gebildet wird
6	PPI Working Paper 04/2026, *Lagebildreife*. Quelle der Abgrenzung zwischen Nachweisstufe und Reifegrad
7	Becker, J., Rush, N., Barnes, B. und Rein, D. (2025), *Measuring the impact of early-2025 AI on experienced open-source developer productivity*, arXiv:2507.09089, Vorabfassung. Quelle der Richtung des Befundes zur Bearbeitungszeit in Kapitel 2, es wird keine Zahl übernommen

> Die Rechnung in Kapitel 11 ist als Modellrechnung gekennzeichnet und arbeitet mit offengelegten Beispielwerten. Sie trägt Nachweisgrad 1 und ersetzt keine betriebsindividuelle Kalkulation. Kursierende Branchenzahlen zum Nutzen von KI und zu Stillstandskosten sind in diesem Whitepaper bewusst nicht verwendet, weil sie sich bei Prüfung der Primärquellen nicht auf den Mittelstand beziehen lassen.

RP

Ralf Platvoet

Diplom-Ökonom, unabhängiger Berater für Controlling, Strategie und KI-Einsatz im Mittelstand. Platvoet Performance Intelligence verbindet betriebswirtschaftliche Bewertung mit KI-gestützter Arbeitsweise, von der Bewertung einzelner Anwendungen bis zum Nachweisverfahren für ein ganzes Vorhabenportfolio.

info@platvoet.org · platvoet.org

© 2026 Platvoet Performance Intelligence · platvoet.org

Alle Angaben ohne Gewähr. Die Modellrechnung in Kapitel 11 arbeitet mit Beispielwerten und ersetzt keine betriebsindividuelle Kalkulation.

Impressum · Datenschutz