FOUNDATION ASCENT®

by prodct · UNIT [ELEMENTS]

WHITEPAPER

PRODUKTIONSREIFE IM HANDEL

Sind Ihre Daten bereit für KI?

Eine ehrliche Antwort in drei Wochen.

Fast jedes Handelsunternehmen hat KI-Projekte gestartet. Die wenigsten liefern im Tagesgeschäft. Dieses Whitepaper zeigt, woran das liegt. Und wie Sie herausfinden, wo es bei Ihnen hakt und was zu tun ist.

HERAUSGEBER prodct GmbH · Unit [Elements]

STAND Juli 2026 · Version 1.1 · frei teilbar

LESEDAUER ca. 15 Minuten

Warum so viel KI-Aktivität *so wenig bewirkt*

Kaum ein Thema beschäftigt den Handel so sehr wie Künstliche Intelligenz. Chatbots im Service, Prognosen im Einkauf, Agenten in der Disposition: Fast jedes Haus hat Projekte gestartet. Trotzdem berichten nur wenige von KI, die im Alltag zuverlässig arbeitet und sich rechnet. Die Zahlen zeigen beides zugleich. Viel Aktivität, wenig Wirkung.

89,1 %

der Handelsunternehmen nennen hohe Datenqualität als relevanten Erfolgsfaktor für KI. Sie ist der Spitzenwert der Erhebung.

HDE-KI-STUDIE 2025

42 %

der Unternehmen brechen die Mehrheit ihrer KI-Initiativen vor der Produktion ab. Das sind mehr als doppelt so viele wie im Vorjahr.

S&P GLOBAL MARKET INTELLIGENCE, VOICE OF THE ENTERPRISE 2025

> 40 %

der agentischen KI-Projekte werden bis Ende 2027 wieder eingestellt. Die Gründe: unklarer Nutzen und fehlende Kontrolle.

GARTNER · PROGNOSE JUNI 2025

Bemerkenswert ist, *woran* diese Vorhaben scheitern. Es ist fast nie das Modell. Die RAND Corporation nennt als häufigste Ursachen unklare Ziele und fehlende Daten. Nicht die Technologie. Der Cisco AI Readiness Index misst die Datendimension regelmäßig als eine der schwächsten. Und Gartner warnt, dass fehlende KI-taugliche Daten Projekte grundsätzlich gefährden. Drei unabhängige Quellen. Ein Befund:

■ **Der Engpass liegt in der Datengrundlage.** Es geht um Qualität, Zugänglichkeit und Verantwortung. Wer das fünfte Pilot-Projekt startet, ohne diese Grundlage zu kennen, wiederholt das Experiment. Mit demselben Ausgang. Die Frage ist nicht, ob Ihre KI gut genug ist. Die Frage ist, ob Ihre Daten gut genug für die KI sind.

Für Entscheider heißt das: Reihenfolge ändern. Erst messen, ob die Daten für den gewünschten Anwendungsfall ausreichen. Dann investieren. Und zwar dort, wo die Messung den Engpass zeigt. Wie so eine Messung aussieht, zeigt der Rest dieses Papiers.

Alle Werte mit Quelle und Jahr. Die Quellen sind am Ende verlinkt. Prognosen sind als solche gekennzeichnet. Wir zitieren nur belastbare Erhebungen und verzichten auf populäre, aber umstrittene Kennzahlen.

Warum der Pilot überzeugt *und der Betrieb enttäuscht*

Warum überzeugt der Pilot und scheitert der Betrieb? Ein Pilot ist ein Experiment unter Laborbedingungen. Jemand hat die Daten ausgewählt, bereinigt und einmal exportiert. Der Produktivbetrieb kennt diese Gnade nicht. Er läuft auf der echten Systemlandschaft. Mit allem, was sich dort über Jahre angesammelt hat:

DER PILOT läuft auf …

… einem kuratierten Datenauszug,
… von Hand bereinigt und einmalig exportiert,
… in einer kontrollierten Testumgebung,
… mit einem Menschen, der jede Ausgabe prüft.

Ergebnis: Die Demo überzeugt.

DER BETRIEB läuft auf …

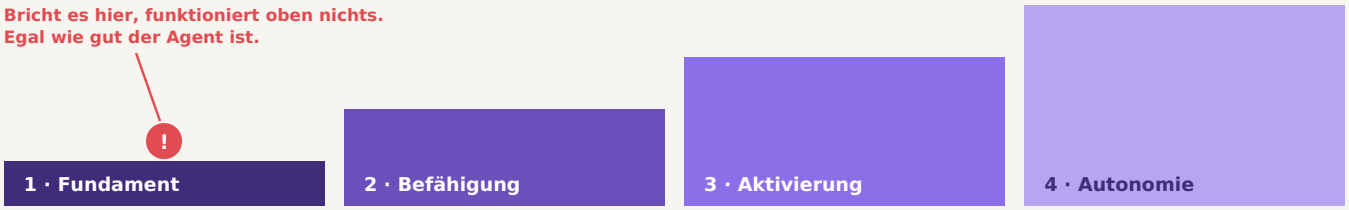
… der echten Systemlandschaft (ERP, PIM, POS, WMS),
… mit mehreren Wahrheiten für dieselbe Kennzahl,
… mit Pipelines, Berechtigungen, Latenzen,
… in Masse und ohne Rückfrage.

Ergebnis: Die Fehler skalieren.

Ein Mensch kommt mit fehlerhaften Stammdaten zurecht. Er erkennt den Fehler, fragt nach oder korrigiert still. Ein KI-System tut das nicht. Es verarbeitet, was da ist. In Sekunden, in Masse, ohne Rückfrage. **KI löst Datenprobleme nicht. Sie skaliert sie.** Ein Beispiel: Fehlt am Artikel die Lieferzeit, rät der Agent bei jeder Bestellung. Nur eben automatisiert und tausendfach.

Die schwächste Schicht begrenzt das Ganze

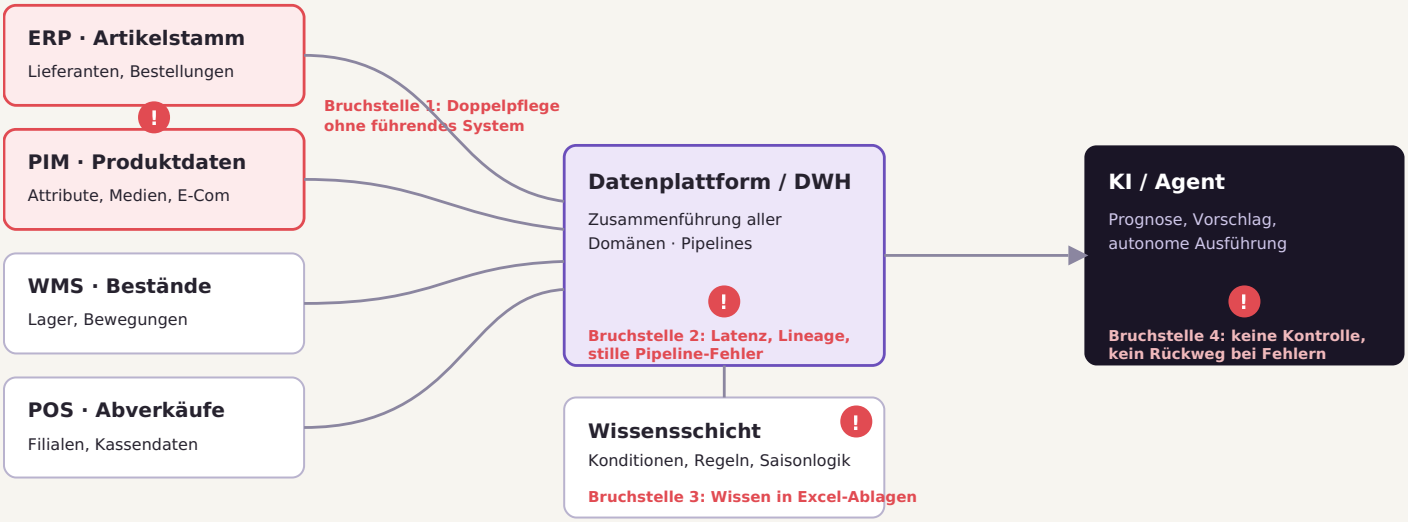
Jeder KI-Anwendungsfall ruht auf vier Stufen. Das **Fundament**: saubere, verantwortete Daten. Die **Befähigung**: Daten werden über Pipelines und Schnittstellen nutzbar. Die **Aktivierung**: Unternehmenswissen wird für KI abrufbar. Die **Autonomie**: Agenten führen Aufgaben selbstständig aus. Man kann Stufen überspringen. Aber nicht folgenlos. Die schwächste benötigte Schicht begrenzt, wie zuverlässig das Ganze werden kann. Kein Agenten-Tuning hebt diese Grenze. Nur das Schließen der Lücke tut es.



Diese eine Schicht nennen wir den **Engpass**. Ihn zu finden ist der Zweck des Produktionsreife-Audits. Denn er zeigt, wohin die nächste Investition gehört. Fast nie in mehr KI. Fast immer in eine bestimmte, benennbare Datenschicht.

Wo der Engpass *in Ihren Datenflüssen entsteht*

Im Handel entsteht eine KI-Entscheidung nie aus einem System. Ein Bestellvorschlag ist das Ende einer Kette. Stammdaten kommen aus ERP und PIM. Bestände aus dem Lager. Abverkäufe von der Kasse. Konditionen vom Lieferanten. Eine Datenplattform führt alles zusammen, die KI arbeitet mit dem Ergebnis. Jedes Glied dieser Kette kann der Engpass sein. Und die typischen Bruchstellen ähneln sich von Haus zu Haus:



Die vier Bruchstellen entsprechen den vier Stufen. Doppelpflege und fehlende Verantwortung sind Fundament-Probleme. Pipeline-Fehler und Latenzen gehören zur Befähigung. Ungepflegtes Wissen ist ein Aktivierungs-Problem. Fehlende Kontrolle ein Autonomie-Problem. **Ein Audit prüft genau diese Kette. Aber nur die Glieder, die Ihr Anwendungsfall wirklich braucht.** Das unterscheidet eine gezielte Prüfung von einem unternehmensweiten Datenprojekt.

Was eine ehrliche Prüfung *von einem Fragebogen unterscheidet*

Der Markt ist voll von kostenlosen Readiness-Checks per Fragebogen. Ihr Problem: Selbsteinschätzungen fallen zu gut aus. Jede Abteilung bewertet die eigene Arbeit. Ein Score ohne Belege ist deshalb eine Meinung mit Nachkommastelle. Eine Prüfung, auf die Sie eine Investitionsentscheidung stützen können, folgt drei Prinzipien:

PRINZIP 1

Belegt oder gedeckelt

Jedes Kriterium braucht ein prüfbares Artefakt. Einen Datenqualitäts-Report, eine Schnittstellen-Doku, ein Betriebs-Log. Ohne Beleg gibt es keine gute Note. Egal, was berichtet wird.

PRINZIP 2

Je Use Case, nicht pauschal

Nicht: Wie reif ist das Unternehmen? Sondern: Welche Datenschichten braucht dieser Anwendungsfall? Und sind sie dafür gut genug? Gemessen wird eine Matrix aus Stufen und Datendomänen (siehe unten).

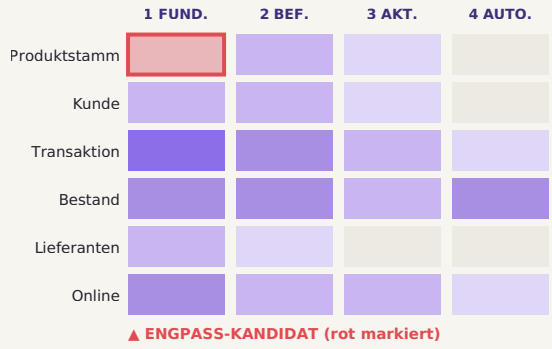
PRINZIP 3

Reproduzierbar

Kriterien, Punkteanker und Formeln sind offen dokumentiert. Zwei Prüfer müssen mit denselben Belegen zum selben Ergebnis kommen. Sonst ist es kein Standard, sondern Bauchgefühl.

Gemessen wird eine Heatmap, keine Einzelzahl

Bewertet wird jede Kombination aus Datendomäne und Stufe. Die Domänen im Handel: Produktstammdaten, Kunde, Transaktion, Bestand, Lieferanten, Online-Verhalten. Jede Zelle bekommt einen Reifewert von 0 bis 100. Ein einzelner Gesamtwert würde täuschen, denn Reife ist in der Praxis ungleich verteilt. Die Heatmap zeigt, wo Ihre Daten stark sind. Und wo genau es klemmt:



Gelesen wird die Heatmap vom Anwendungsfall her. Ein Nachbestell-Agent braucht Produktstamm und Bestand. Die Online-Domäne braucht er nicht. Bewertet und geschlossen werden nur die Zellen, die der Use Case benötigt.

Die Skala hat fünf Reifebänder: Ad hoc, Initial, Etabliert, Gesteuert, Optimiert. Jedes Ergebnis wird als Band mit Konfidenz angegeben. Nie als scheinpräzise Einzelzahl.

Eine Prüfdimension wird oft vergessen: Kommen KI-Agenten überhaupt an die Daten heran? Agenten arbeiten über Schnittstellen, nicht über Excel-Exporte. Wir prüfen das deshalb technisch. Gibt es APIs? Sind sie dokumentiert? Sind die Datenmodelle maschinenlesbar? Diese Prüfung liest die Systeme aus. Interviews reichen dafür nicht.

Illustrative Heatmap. Die vollständige Bewertungslogik (Kriterienkataloge, Reifebänder 0–100, Soll-Ableitung, Formeln) ist im offenen Messstandard dokumentiert: prodct.group/foundation-ascent/methode

Was Sie am Ende *in der Hand haben*

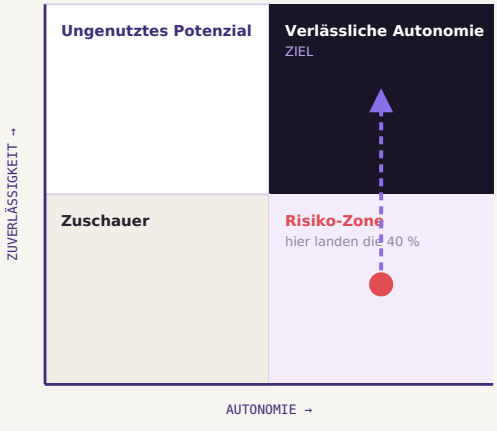
Am Ende eines Audits steht kein Foliensatz mit Empfehlungs-Prosa. Es steht ein Befund mit vier Ergebnissen. Jedes ist für sich nutzbar. Zusammen sind sie die Grundlage für Ihre nächste KI-Investition:

R1
Der Zuverlässigkeits-Wert
Wie zuverlässig kann Ihr Use Case heute werden? Eine belegte Zahl von 0 bis 100, als Band mit Konfidenz. Keine Meinung. Sie zeigt, ob der angestrebte Automatisierungsgrad heute realistisch ist.

R2
Der Engpass
Welche Datenschicht begrenzt das Ergebnis? Mit Belegen aus Ihren Systemen und dem Preis der Lücke. Der Engpass beantwortet die Frage: Wohin zuerst?

R3
Der Maßnahmenplan
Was ist zuerst zu tun? Was kostet es, was bringt es? Der Plan ist priorisiert und so beschrieben, dass Sie ihn auch ohne uns umsetzen können.

R4
Die Einordnung
Wo stehen Sie im Zielbild aus Zuverlässigkeit und Autonomie? Dazu eine ehrliche Empfehlung für den nächsten Schritt. Auch wenn sie „noch kein Agent“ lautet.



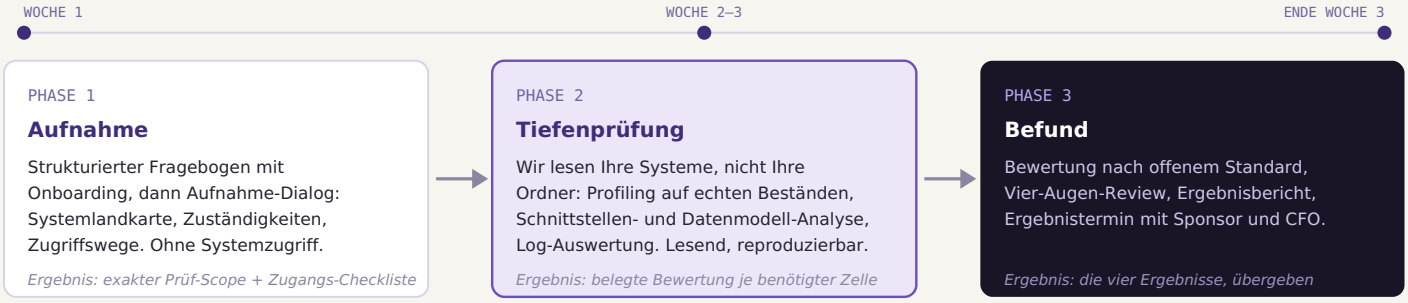
Der Weg führt nach oben, nicht nach rechts.

Rolle	Was das Audit Ihnen gibt
COO / Fachbereich	Die Antwort, warum der Prozess noch nicht zuverlässig läuft. Und einen Plan mit erwartbarem Effekt je Maßnahme. Statt eines weiteren Piloten.
CDO / CIO	Den Beleg, welche Datenschicht in Ordnung ist und welche bremst. Messbar statt Bauchgefühl. Dazu eine klare Investitionsreihenfolge als Budgetierungshilfe.
CFO / Geschäftsführung	Festpreis, fester Endtermin, ergebnisoffener Befund. Und die Sicherheit, kein Geld mehr in einen Use Case zu stecken, den eine andere Schicht begrenzt.

■ Der rote Punkt im Quadranten ist der häufigste Befund: hoher Autonomie-Anspruch auf schwachem Fundament. Das ist der direkte Weg in die Abbruch-Statistik. Niemand braucht Autonomie um ihrer selbst willen. Was zählt, sind Prozesse, die verlässlich laufen. Mit so viel Autonomie, wie das Fundament erlaubt. **Zuverlässigkeit vor Autonomie.**

Wie das Audit *in drei Wochen abläuft*

Ein Audit muss den Betrieb nicht lahmlegen. Der Ablauf hat deshalb zwei Teile. Erst nehmen wir strukturiert auf, ganz ohne Systemzugriff. Dann prüfen wir lesend in den Systemen, die Ihr Anwendungsfall braucht. So bleibt Ihr Aufwand klein und die Prüftiefe hoch:



Ihr Aufwand bleibt überschaubar. Ein halbtägiger Workshop. Ein Ansprechpartner mit zwei bis drei Stunden pro Woche. Lesezugänge nach Checkliste. Keine Vorarbeiten, kein Aufräumen vorab. Das Audit bewertet den Ist-Zustand, genau dafür ist es da. Fehlen Zugänge, pausiert die Frist transparent.

Das Audit ist ergebnisoffen und kennt drei Ausgänge

AUSGANG 1

Fundament schwach: Wir schließen die Engpass-Lücke. Im Vertical Slice Sprint, der den Use Case in den Produktivbetrieb führt. Nicht das ganze Fundament, nur das richtige.

AUSGANG 2

Fundament stark: Gute Nachricht. Kein Datenprojekt nötig. Sie steigen weiter oben ein, bei Wissensschicht oder Agent.

AUSGANG 3

Use Case lohnt sich nicht: Wir sagen es Ihnen, bevor Sie investieren. Und priorisieren gemeinsam den Fall, der es wert ist.

■ Alle Ergebnisse gehören Ihnen und sind ohne uns umsetzbar. Eine unbequeme Antwort nach drei Wochen Audit ist billiger als zwölf Monate Pilot ohne Produktion.

EINBLICK

Ein Blick in einen fertigen Befund *am Beispiel ACME Inc.*

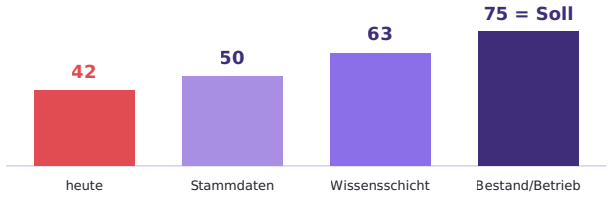
Wie konkret ein Befund wird, zeigt ein Beispiel. ACME Inc. ist fiktiv, aber vollständig durchgerechnet. Ein Multichannel-Händler mit rund 480 Mio. € Umsatz will die Nachbestellung automatisieren. Der Agent soll Bestellvorschläge selbst auslösen. Ein Disponent greift nur bei Ausnahmen ein. Das Audit findet den Engpass nicht im Agenten. Es findet ihn im Artikelstamm. Der wird in ERP und PIM parallel gepflegt, ohne führendes System. Die Folgen sind messbar: Dubletten, fehlende Lieferzeiten, fehlende Mindestbestellmengen. Genau die Felder, die eine Bestellentscheidung braucht.

DER BEFUND IN ZAHLEN (FIKTIV)

Benötigte Zelle	Soll	Ist	Lücke
Produktstamm · Stufe 1	75	42	**33 ◀ ENGPASS**
Bestand · Stufe 1	75	63	12
Bestand · Stufe 2	75	54	21
Bestand · Stufe 3	75	50	25
Bestand · Stufe 4	75	55	20

Zuverlässigkeits-Wert heute: **Band 37–47 von 100**. Zu wenig für den Zielgrad „autonom mit Ausnahme-Freigabe“ (Soll 75). Kein Agenten-Tuning hebt diesen Wert. Nur das Schließen der Stammdaten-Lücke.

DER EFFEKT DES MASSNAHMENPLANS (FIKTIV)



Der Maßnahmenplan hebt den Wert Schritt für Schritt. Jede Maßnahme hat einen Aufwand und einen erwarteten Effekt. Am Ende reicht die Datenbasis für das Ziel.

■ Der vollständige Beispiel-Report hat elf Seiten. Er enthält Management Summary, Systemlandschaft, Datenmodell mit gemessener Feldbefüllung, Engpass-Belege, Maßnahmenplan und Einordnung. Ein Dossier macht jede Zahl nachvollziehbar. Frei zum Download: **prodct.group/foundation-ascent/produktionsreife-audit**

ACME Inc. ist fiktiv. Die Zahlen zeigen die Methodik. Echte Audits sind vertraulich. In die Benchmark fließen nur anonymisierte, aggregierte Werte.

FAZIT

Woran Ihr KI-Vorhaben wirklich scheitert, *lässt sich herausfinden*

Drei Sätze fassen die Lage zusammen. Die meisten KI-Vorhaben im Handel scheitern nicht am Modell, sondern an der Datengrundlage. Welche Schicht einen Anwendungsfall begrenzt, lässt sich in drei Wochen messen. Mit Belegen und nach offenem Standard. Und wer den Engpass kennt, investiert anders: gezielt in eine benennbare Schicht statt breit in mehr KI.

Die Frage ist also nicht, ob Sie mit KI arbeiten werden. Die Frage ist, ob Ihre Daten schon so weit sind. Beides beginnt an derselben Stelle: mit einer ehrlichen Messung.

NÄCHSTER SCHRITT

Produktionsreife-Audit anfragen

Drei Wochen, Festpreis, ergebnisoffen. Antwort innerhalb von 24 Stunden. Von einer Person, nicht von einem Funnel.

PRODCT.GROUP/FOUNDATION-ASCENT

DIE METHODE IM DETAIL

Offener Messstandard

Kriterien, Reifebänder und Formeln sind öffentlich dokumentiert. Nachvollziehbar statt Blackbox.

PRODCT.GROUP/FOUNDATION-ASCENT/METHODE

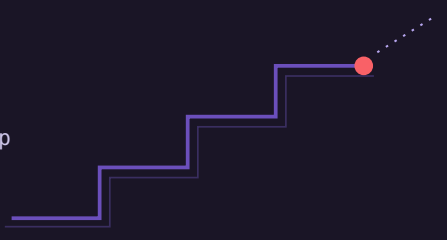
DER BEWEIS

Beispiel-Report ansehen

Der vollständige ACME-Befund als PDF. So sieht das Ergebnis aus, das Sie nach drei Wochen in der Hand haben.

FREI ZUM DOWNLOAD AUF DER AUDIT-SEITE

KONTAKT Jan Fischer · jan.fischer@prodct.group · WEB prodct.group



Die Quellen dieses Papiers, *verlinkt und eingeordnet*

Zahlen zum Scheitern von KI-Projekten kursieren viele. Sie unterscheiden sich methodisch erheblich. Wir zitieren nur Erhebungen mit nachvollziehbarer Herkunft und Methodik. Prognosen kennzeichnen wir als Prognosen. Auf populäre, aber umstrittene Kennzahlen verzichten wir.

Quelle	Befund & Einordnung
HDE · KI-Studie im Handel 2025 (mit Safaric Consulting)	89,1 % der Handelsunternehmen nennen hohe Datenqualität als relevanten oder sehr relevanten Erfolgsfaktor für KI, der höchste Wert der Erhebung. Die handelsspezifischste Quelle dieses Papiers.
S&P Global Market Intelligence · Voice of the Enterprise: AI & Machine Learning, Use Cases 2025	Der Anteil der Unternehmen, die die Mehrheit ihrer KI-Initiativen vor der Produktion abbrechen, stieg von 17 % auf 42 % (n = 1.006, Nordamerika und Europa). Die solideste Abbruch-Kennzahl am Markt.
Gartner · Prognose zu agentischen KI-Projekten, Juni 2025	Über 40 % der agentischen KI-Projekte werden bis Ende 2027 eingestellt. Die Gründe: steigende Kosten, unklarer Nutzen, fehlende Risikokontrollen. *Prognose, kein gemessenes Ergebnis.*
Gartner · „Lack of AI-ready data puts AI projects at risk", Februar 2025	63 % der Datenmanagement-Verantwortlichen haben keine passenden Datenpraktiken für KI oder sind unsicher, ob sie sie haben. Fehlende KI-taugliche Daten gefährden Projekte grundsätzlich.
RAND Corporation · The Root Causes of Failure for AI Projects, 2024	Interviews mit 65 erfahrenen Data Scientists und Engineers. Häufigste Ursachen: missverstandene Ziele und fehlende Daten. Nicht die Technologie. Qualitative Studie.
Cisco · AI Readiness Index 2024	Befragung von 7.985 Führungskräften in 30 Märkten. Nur 13 % der Unternehmen sind voll bereit. Die Datendimension gehört regelmäßig zu den schwächsten Säulen.
Informatica · CDO Insights 2025	Befragung von 600 Datenverantwortlichen. Datenthemen sind die häufigste Hürde auf dem Weg vom Piloten in die Produktion. Anbieternahe Quelle. Die Zahlen passen zu den unabhängigen Erhebungen.

■ **Was wir bewusst nicht zitieren:** die vielzitierte „95 %-Scheiternsquote" (MIT/NANDA 2025). Sie ist eng definiert, methodisch umstritten und wird meist falsch wiedergegeben. Wer mit Zahlen argumentiert, muss sich auch von schlechten trennen können. Das gilt für Studien wie für Stammdaten.