REVENUE OPERATIONS & KI

Was KI in Revenue Operations wirklich verändert

Zwei Feldexperimente und ein gestaffelter Rollout. Zwei Schäden sieht kein Mittelwert; den dritten verbucht er als Verbesserung.

Sinan Isoglu, MBA (Quantic)
Führungskraft für Umsatzwachstum, Dozent und Doktorand

VERÖFFENTLICHT	AKTUALISIERT	LESEZEIT	WÖRTER
27. Juli 2026	28. Juli 2026	16 Min.	3.345

https://isoglu.com/de/journal/was-ki-wirklich-veraendert/

INHALT

MANAGEMENT SUMMARY

Drei Studien haben gemessen, was KI mit kommerzieller Arbeit macht, und zwei davon fanden einen Schaden, den ein Mittelwert nicht auflöst, während die dritte einen fand, den er als Verbesserung verbucht: die erfahrensten Kräfte verlieren an Qualität, während der Durchschnitt steigt; eine Aufgabengrenze, die niemand beim Überschreiten spürt; und Kundschaft, die abbricht, sobald ein Bot offengelegt wird. Keine der drei handelt von Revenue Operations, und keine ihrer Zahlen gehört in eine Pipeline-Diskussion. Übertragbar ist das Instrumentierungsproblem. Der Beitrag beschreibt vier Schnitte — einen durch Daten, die in den meisten Systemen längst liegen, und drei, die neue Daten verlangen — und warum in Deutschland vor der Messfrage eine zweite steht: ob so ausgewertet werden darf.

Schlagwörter: Revenue Operations · Künstliche Intelligenz im Vertrieb · Produktivitätsmessung · Feldexperimente · Mitbestimmung

Die thematisch nächstliegende begutachtete Arbeit zu generativer KI im B2B-Vertrieb misst Vertriebsleistung, indem sie Vertriebsmitarbeitende nach sich selbst fragt. Sie findet einen positiven Effekt. Das ist ehrliche, saubere Adoptionsforschung. Belegt ist damit, dass die Beteiligten glauben, KI habe ihr kommerzielles Ergebnis verbessert — eine andere Frage als die, ob sie es getan hat.

Die drei Arbeiten, die tatsächlich gemessen haben, was KI mit kommerzieller Arbeit macht, sind außerhalb der Vertriebsliteratur erschienen, werden in den beiden hier behandelten Vertriebsarbeiten nicht zitiert, und zwei davon fanden einen Schaden, den ein Mittelwert nicht auflöst, während die dritte einen fand, den er als Verbesserung verbucht: am Rand einer Leistungsverteilung, an einer Aufgabengrenze oder auf der Kundenseite des Gesprächs. Darum geht es hier: Das Instrument, mit dem die meisten Rollouts diese Frage beantworten, kann die drei Arten, wie sie schadet, nicht auflösen.

Die Arbeit, die Sie am ehesten bekommen

Es ist die Studie von Rodriguez, Deeter-Schmelz und Krush, im vergangenen Jahr im Journal of Business & Industrial Marketing erschienen. Sie ist die nächstliegende begutachtete Antwort auf die Frage, und Nähe zur Frage ist das, wonach eine Abrufschicht sortiert. Generative KI, B2B-Vertrieb, ein empirisches Modell, partielle kleinste Quadrate. Wie oft sie tatsächlich zurückkommt, habe ich nicht gemessen. Ihr Befund: GenAI habe „a positive impact on the effectiveness of the sales process, administrative efficiency and sales performance."

Sehen Sie sich an, woraus diese drei Ergebnisgrößen bestehen. Das Abstract beschreibt den gesamten Apparat: Tiefeninterviews und Rückmeldungen aus der Führung, um einen Fragebogen zu entwickeln, eine Pilotbefragung, dann „an online survey distributed to a larger sample", ausgewertet mit partiellen kleinsten Quadraten. Jede einzelne Stufe davon ist ein Instrument, um Auskünfte einzusammeln. Der Limitationsabschnitt ergänzt, was daraus folgt — die Studie „focuses on a single health-care company" und „relies on self-reported data from sales professionals".

Als kommerzielle Führungskraft gelesen und nicht als Methodikerin: Was hier belegt wird, ist ein Zusammenhang innerhalb eines Unternehmens zwischen der Angabe, wie stark GenAI genutzt wird, und der Angabe, wie gut der Vertrieb läuft. Das ist ein realer Befund über eine reale Sache. Es ist keine Messung von Ergebnis.

Beide Grenzen stehen im eigenen Abstract, was mehr ist, als die meisten Arbeiten tun. Die Arbeit beantwortet ihre eigene Frage gut. Ihr erklärter Zweck ist ein Moderationsmodell — ob Unterstützung durch die Geschäftsleitung den Weg von technologischer Selbstwirksamkeit zur GenAI-Nutzung moderiert —, und dafür ist eine Wahrnehmungsmessung das richtige Instrument. Das Problem entsteht, wenn man ihr eine andere Frage stellt: ob KI gewirkt hat, beantwortet von einer Studie, die wissen wollte, ob die Beteiligten das glauben. Diese Unterscheidung trägt die Abrufschicht nicht weiter — ein eigener Beitrag prüft, was diese Schicht zurückgibt, wenn man sie direkt nach der Herkunft fragt; hier gilt die Frage als geklärt.

Die Arbeitshypothese des Feldes zeigt in die andere Richtung

Es gibt eine zweite aktuelle Arbeit, die lesenswert ist und das Problem anders rahmt als alles, was danach kommt. Pia Hautamäki und Minna Heikinheimo haben aus Interviews mit 32 Führungskräften der obersten Ebene in B2B-Vertriebsorganisationen einen Rahmen entwickelt, im Grounded-Theory-Verfahren, erschienen im Maiheft 2025 des Journal of Business Research. Ihr Ausgangspunkt ist, dass „studies have demonstrated that artificial intelligence (AI) can enhance sales efficiency“ im B2B-Kontext — und dass das eigentliche Problem die Lücke danach ist: „despite the wide accessibility of AI, its adoption in B2B sales remains limited.“

Ihre Frage ist deshalb, was eine Organisation überhaupt in die Lage versetzt, KI zu nutzen. Ihre Antwort ist Führungsfähigkeit: „data-based human capital“, das soziale Kapital einer Kultur des Wissensteilens, und „a transformative AI-positive mindset“.

Halten Sie das beim Weiterlesen fest, denn es beschreibt zutreffend, wo die Aufmerksamkeit des Feldes liegt: auf Einführung, und auf der Fähigkeit, die Einführung gelingen lässt. Einführung ist eine andere Frage als Wirkung, und drei bestimmte Schäden liegen außerhalb dessen, was ein Adoptionsrahmen zu sehen gebaut ist.

Die belastbaren Befunde stehen woanders

Die drei Arbeiten, die tatsächlich gemessen haben, was KI mit kommerzieller Arbeit macht, stehen außerhalb der Vertriebsliteratur, und keine der beiden Vertriebsarbeiten oben zitiert eine von ihnen. Rodriguez u. a. führen 113 Quellen, Hautamäki und Heikinheimo 81; der einzige Luo in beiden Listen ist eine andere Arbeit von ihm, und Brynjolfsson und Dell’Acqua stehen in keiner.

Die Besten können schlechter werden. Erik Brynjolfsson, Danielle Li und Lindsey Raymond haben die gestaffelte Einführung eines generativen KI-Assistenten über 5.172 Servicekräfte hinweg untersucht. Die Ergebnisse erschienen im vergangenen Jahr im Quarterly Journal of Economics. Zitiert wird immer der Durchschnitt: 15 Prozent mehr gelöste Vorgänge pro Stunde. Die Zahl steht im Abstract nicht ohne Einschränkung: Derselbe Satz endet mit „with substantial heterogeneity across workers“, und kurz darauf steht, worin diese Heterogenität besteht: „Less experienced and lower-skilled workers improve both the speed and quality of their output, while the most experienced and highest-skilled workers see small gains in speed and small declines in quality.“ Die Autoren sagen es deutlich; verloren geht es erst danach.

Als kommerzielle Führungskraft gelesen und nicht als Ökonom: Am oberen Ende der Verteilung stand ein Tausch, kleine Zuwächse an Tempo gegen kleine Einbußen an Qualität. Und es sind Servicekräfte in einem Unternehmen, keine Menschen mit eigenem Kundenstamm — übertragbar ist die Position in der Verteilung, nicht der Beruf. Ein Mittelwert von 15 Prozent ist vollständig damit vereinbar, dass Ihre besten Verkäuferinnen und Verkäufer etwas schlechter werden. Und der Mittelwert ist das Einzige, was die meisten Rollouts berichten.

Dieselbe Studie zeigt auch, woher der Gewinn kam. Er war am größten „for moderately rare problems, where human agents have less baseline experience but the system still has adequate training data“. Die häufigen Fälle laufen ohnehin gut, und für die wirklich neuen hat das System nichts. Der Gewinn lag im Band dazwischen, während die meisten Rollouts auf Menge zielen.

Im selben Abstract steht ein vierter Befund, und er läuft dem kundenseitigen Schaden zuwider, um den es weiter unten geht: „customers are more polite and less likely to ask to speak to a manager“. Wo der Assistent hinter einer menschlichen Servicekraft saß, wurde die Kundenseite dieses Rollouts besser.

Es gibt eine Grenze, und wer sie überschreitet, merkt es nicht. Fabrizio Dell'Acqua und Kolleginnen und Kollegen haben ein präregistriertes Experiment mit 758 Wissensarbeitenden bei der Boston Consulting Group über achtzehn realitätsnahe Aufgaben durchgeführt, die sämtlich innerhalb dessen liegen, was sie die zerklüftete technologische Grenze nennen, dazu eine neunzehnte: „a complex managerial task selected to be outside the frontier“. Erschienen ist die Arbeit in diesem Jahr in Organization Science. Bei den achtzehn bearbeiteten die Teilnehmenden mit GPT-4 12,2 Prozent mehr Aufgaben, 25,1 Prozent schneller, in besserer Qualität. Bei der neunzehnten war die Wahrscheinlichkeit einer richtigen Lösung um 19 Prozent geringer.

Das entscheidende Wort ist zerklüftet. In ihrer Beschreibung verbessert KI-Unterstützung „performance for some tasks but worsens it for others, even within the same knowledge workflow and with a seemingly similar level of difficulty“. Zwei Aufgaben, die der bearbeitenden Person gleich schwer vorkommen, können auf verschiedenen Seiten liegen. Auf welcher Seite eine Ihrer Aufgaben liegt, zeigt sich erst am Ergebnis — Schnitt 3 weiter unten.

Ein Teil des Verlusts hat mit Leistungsfähigkeit nichts zu tun. Xueming Luo, Siliang Tong, Zheng Fang und Zhe Qu haben mehr als 6.200 Kundinnen und Kunden zufällig auf Chatbots und menschliche Kräfte verteilt, bei stark strukturierten Outbound-Verkaufsanrufen. Die Studie steht in Marketing Science. Ohne Offenlegung waren die Bots „as effective as proficient workers“. Wurde die Maschine vor dem Gespräch offengelegt, fielen die Abschlussquoten um mehr als 79,7 Prozent. Der Mechanismus, auf den die Autoren hinauswollen, ist nicht Leistung: Der Effekt „seems to be driven by a subjective human perception against machines“. Die Kundschaft „are curt and purchase less because they perceive the disclosed bot as less knowledgeable and less empathetic“ — und die Autoren halten ausdrücklich fest, dass das „despite the objective competence of AI chatbots“ geschieht. Die Grenze ihres Befundes halten sie im Satz darauf fest: „such negative impact can be mitigated by a late disclosure timing strategy and customer prior AI experience.“

Eine Durchsatzkennzahl kann eine Kundin, die innerlich abgebrochen hat, nicht sehen. Sie verbucht ein kürzeres Gespräch — was die Studie ebenfalls fand: Die Offenlegung „substantially decreases call length“.

Tabelle 1 **Was diese Arbeiten tatsächlich gemessen haben**

Arbeit	Design	Untersuchte	Feld	Wo ein Schaden sichtbar würde
Rodriguez u. a. (2025)	Befragung, partielle kleinste Quadrate	Vertriebsmitarbeitende, ein Gesundheitsunternehmen	B2B-Vertrieb	Nicht gemessen — das Ergebnis ist eine Auskunft
Hautamäki und Heikinheimo (2025)	Grounded Theory, Interviews	32 Führungskräfte der obersten Ebene	B2B-Vertriebsorganisationen	Nicht gemessen — das Ergebnis ist Führungsfähigkeit
Brynjolfsson u. a. (2025)	Gestaffelte Einführung, nicht randomisiert	5.172 Servicekräfte	Kundenservice nach dem Kauf, ein Unternehmen	Am oberen Ende der Leistungsverteilung
Dell'Acqua u. a. (2026)	Präregistriertes randomisiertes Experiment	758 Wissensarbeitende	18 Aufgaben innerhalb der Grenze, 1 außerhalb	Auf einer Seite einer Aufgabengrenze
Luo u. a. (2019)	Randomisiertes Feldexperiment	Mehr als 6.200 Kundinnen und Kunden	Stark strukturierte Outbound-Anrufe	Auf der Kundenseite des Gesprächs

Eigene Zusammenstellung der fünf zitierten Arbeiten, jeweils nach dem veröffentlichten Abstract. Die Designs sind so benannt, wie die Autorinnen und Autoren sie benennen. Ergebniszahlen sind hier bewusst nicht wiedergegeben; das Argument des Beitrags ist, dass diese Zahlen nicht wandern sollten.

Was nicht übertragbar ist

Keine dieser drei Arbeiten handelt von Revenue Operations. Brynjolfsson ist Kundenservice nach dem Kauf in einem Unternehmen, und eine gestaffelte Einführung, keine randomisierte. Dell'Acqua sind Wissensarbeitende an konstruierten Aufgaben. Luo ist Feldforschung aus der Zeit vor den großen Sprachmodellen, 2019 veröffentlicht, an stark strukturierten Outbound-Anrufen, was über einen Agenten der GPT-Klasse nichts aussagt. Niemand sollte 15, 19 oder 79,7 Prozent in ein Pipeline-Gespräch tragen. Ich werde es nicht tun.

Übertragbar ist die Form. Drei Methoden, drei Felder, drei verschiedene Mechanismen. In jedem liegt der Schaden dort, wo ein Mittelwert nicht hinreicht: am Rand einer Leistungsverteilung, an einer Aufgabengrenze oder auf der Kundenseite des Gesprächs. Das ist eine Aussage über Instrumentierung, und Instrumentierung überträgt sich.

Es lohnt zu sagen, wie sich das zu Daron Acemoglu und Pascual Restrepo verhält. Ihr Argument ist, dass der technologische Wandel der jüngeren Zeit auf Automatisierung ausgerichtet war statt auf das Schaffen neuer Tätigkeiten, und dass die Folgen „stagnating labour demand, declining labour share in national income, rising inequality and lowering productivity growth" gewesen seien. Das ist dieselbe Sorge auf einer anderen Flughöhe. Ihr Rahmen setzt bei der einzelnen Tätigkeit an — KI „can automate tasks previously performed by labour or create new tasks" —, ihre Ergebnisgröße ist das Volkseinkom-

men. Hier ist es ein einzelnes Unternehmen, und die Ergebnisgröße ist das, was am Montag auf einem Dashboard steht. Beides kann zutreffen. Nur eines davon sagt Ihnen, was Sie instrumentieren müssen.

Die vier Schnitte

Wenn Ihr Rollout einen moderaten durchschnittlichen Zuwachs gebracht hat, ist der Durchschnitt das Uninteressanteste, was Sie jetzt wissen. Vier Schnitte. Der erste ist ein Neuschnitt vorhandener Daten, sofern der Rollout eine Gruppe ohne Werkzeug gelassen hat. Die anderen drei brauchen etwas, das Sie noch nicht haben: eine Häufigkeitsverteilung über Fallarten und ein bewertetes Qualitätsmaß.

1. Schneiden Sie den Effekt nach Erfahrung — und gegen eine Vergleichsgruppe. Brynjolfsson, Li und Raymond schneiden nach Erfahrung und Können, und sie können den Effekt überhaupt nur zeigen, weil die noch nicht umgestellten Servicekräfte im selben Zeitraum mitgemessen werden. Beides muss mitwandern. Gruppieren Sie Ihre Vertriebsmitarbeitenden nach etwas, das vor dem Rollout feststand und selbst keine Ergebnisgröße ist — Betriebszugehörigkeit, Einarbeitungsstand, eine Kompetenzeinschätzung von vorher — und vergleichen Sie die Veränderung je Gruppe mit Personen, die das Werkzeug noch nicht haben. Fällt der Zuwachs der erfahrensten Gruppe im umgestellten Arm hinter den der unerfahrensten zurück, und zeigt sich derselbe Abstand in der Vergleichsgruppe nicht, haben Sie das Brynjolfsson-Muster. Erst die Vergleichsgruppe macht diesen Satz belastbar: Ordnen Sie Menschen nach einer verrauschten Ausgangsmessung und rechnen Sie die Veränderung nach genau dieser Ordnung, fällt die Spitzengruppe auch ganz ohne Maßnahme — das ist Regression zur Mitte, und bei der Zuverlässigkeit von Quartalsleistungen im Vertrieb ist sie größer als jeder KI-Effekt, den Sie suchen. Dezile lohnen sich erst ab einigen hundert Vertriebsmitarbeitenden; bei vierzig schneiden Sie in Hälften und rechnen Sie nur mit großen Unterschieden.
2. Schneiden Sie danach, wie selten eine Fallart ist. Sortieren Sie Fälle, Deals oder Tickets danach, wie oft der Typ vorkommt. Die Vorhersage ist ein mittleres Band: wenig Wirkung beim Routinefall, wenig beim wirklich neuen, das meiste dazwischen. Sie hängt an einer Voraussetzung, die Brynjolfssons Assistent erfüllte — „the system still has adequate training data“. Er war auf der Lösungshistorie des Unternehmens trainiert, selten im Unternehmen hieß also dünn im Modell. Ein Modell von der Stange hat dieses Gefälle nicht: Eine öffentliche Ausschreibung, die in Ihrem Bestand selten ist, kann in seinen Trainingsdaten häufig sein, und die Sonderfälle Ihres Produkts sind in beidem selten. Prüfen Sie, ob die beiden Seltenheiten zusammenfallen, bevor Sie ein mittleres Band als denselben Befund lesen. Liegt Ihr Gewinn bei den häufigen Fällen, messen Sie eingesparte Zeit an Arbeit, die schon vorher billig war.
3. Finden Sie Ihre Grenze, indem Sie Stichproben der Arbeitsergebnisse bewerten lassen. Ziehen Sie eine geschichtete Stichprobe KI-gestützter Arbeitsergebnisse über Aufgabentypen hinweg und lassen Sie sie anhand eines schriftlichen Rasters von jemandem bewerten, der nicht beteiligt war. Die Grenze ist zerklüftet. Welche Seite eine Aufgabe trifft, lässt sich nur am Ergebnis prüfen, nicht an der Zuversicht dessen, der es erzeugt hat. Lassen Sie eine Stichprobe ohne Werkzeug erstellter Arbeit durch dieselbe Bewertung laufen: Ohne diese zweite Stichprobe haben Sie ein Qualitätsniveau, keinen Effekt. Und gesucht ist ein Ergebnis, das falsch ist und richtig aussieht — die bewertende Person muss die Arbeit nachrechnen können, nicht nur lesen.
4. Messen Sie Qualität und Durchsatz zusammen oder gar nicht. Zwei der drei Befunde oben sind für eine allein gelesene Durchsatzkennzahl unsichtbar. Der dritte ist schlimmer als unsichtbar: Die Offenlegung verkürzte die Gespräche, der Durchsatz verbuchte den Schaden also als Verbesserung.

Lösungsquote gegen Qualitätsbewertung. Vereinbarte Termine gegen eine bewertete Stichprobe, wie gut sie qualifiziert waren. Verschickte Angebote gegen die Quote derer, die zur Nacharbeit zurückkommen. Eine eindimensionale Kennzahl kann keinen Tausch abbilden, und jedes Ergebnis oben ist ein Tausch.

Dafür braucht es keine neue Software. Drei der vier brauchen Daten, die Sie noch nicht haben, und zusammengenommen sind das zwei Dinge: eine Häufigkeitstaxonomie über Fallarten und ein Qualitätsmaß mit Raster und bewertender Person dahinter. Beides fehlt nicht aus Versehen: Erfasst wird, was ein System besitzt — aus demselben Grund hat auch die Spanne zwischen der Entstehung eines Leads und der ersten menschlichen Reaktion nirgends ein Feld. Der Rest sind dieselben Zahlen, anders geschnitten, eine Vergleichsgruppe, die vor dem Rollout festgelegt wird statt danach – eine der drei Fragen, die jede Zahl vor einer Entscheidung bestehen sollte –, und die Bereitschaft, an den Rand zu sehen statt in die Mitte. Darunter liegt eine zeitliche Grenze für alle vier: Bei einem Vertriebszyklus von drei bis neun Monaten liegen für einen nach neunzig Tagen ausgewerteten Rollout noch keine abgeschlossenen Geschäfte aus der Zeit danach vor. Schneiden Sie nach Stufen, die innerhalb des Fensters fallen, oder warten Sie auf die Kohorte.

Abbildung 1 **Vier Schnitte durch Daten, die Sie schon haben**

DER SCHNITT	WAS DER MITTELWERT SAGT	WAS DER RAND SAGT	DECKT SICH DAS?
Erfahrung, Seltenheit, Aufgabentyp, Qualität.	Die berichtete Zahl.	Erfahrenste Gruppe, mittleres Band, jenseits der Grenze.	Wenn nicht, war der Mittelwert falsch.

Vier Zeilen, weil es vier Schnitte sind. Sie sind keine Alternativen — ein Rollout kann drei bestehen und am vierten scheitern. Schnitt 1 braucht eine Vergleichsgruppe; ohne sie füllt sich die dritte Spalte

Eigenes Arbeitsblatt.

In Deutschland ist die Messfrage nicht die erste Frage

Die vier Schnitte sind Vorschläge, mehr zu messen. Hier ist das nicht zuerst eine Methodenfrage.

§ 87 Abs. 1 Nr. 6 BetrVG gibt dem Betriebsrat ein Mitbestimmungsrecht bei „Einführung und Anwendung von technischen Einrichtungen, die dazu bestimmt sind, das Verhalten oder die Leistung der Arbeitnehmer zu überwachen“. Entscheidend ist dabei nicht, wozu der Arbeitgeber die Einrichtung bestimmt hat, sondern wozu sie objektiv geeignet ist. Die Absicht ist kein Maßstab, weil sie sich nicht prüfen lässt.

Wie weit das trägt, hat das Bundesarbeitsgericht am 16. Juli 2024 gezeigt (1 ABR 16/23). Der Leitsatz:

> **Ein Headset-System, das es den Vorgesetzten ermöglicht, die Kommunikation unter Arbeitnehmern mitzuhören, ist eine technische Einrichtung, die zur Überwachung der Arbeitnehmer iSd. § 87 Abs. 1 Nr. 6 BetrVG bestimmt ist. Seine Einführung und Nutzung unterliegt auch dann der betrieblichen Mitbestimmung, wenn die Gespräche nicht aufgezeichnet oder gespeichert werden.**

Der zweite Satz ist der teure. „Wir speichern nichts" ist keine Verteidigung.

Lesen Sie die vier Schnitte jetzt noch einmal. Schnitt 1 gruppiert namentlich bekannte Mitarbeitende nach Erfahrung und rechnet die Veränderung je Gruppe. Schnitt 3 zieht eine Stichprobe aus der Arbeit einzelner Personen und lässt sie fremdbewerten. Beides ist genau das, wovon Nummer 6 spricht: eine technische Einrichtung, mit der sich die Leistung von Beschäftigten auswerten lässt. Dass Sie damit einen Schaden suchen, statt Leistung zu kontrollieren, ändert an der objektiven Eignung nichts — und die Eignung ist der Maßstab.

Das ist kein Grund, es zu lassen. Es ist der Grund, warum es früher auf die Tagesordnung gehört: nicht als Auswertung, die man hinterher erklärt, sondern als Vorhaben, das man vorher beschreibt. Wer den Schaden im obersten Dezil sucht, braucht dafür ohnehin einen Zweck, den er formulieren kann. Die Mitbestimmung zwingt nur dazu, ihn vor der Auswertung zu formulieren statt danach.

Damit steht die Frage in Deutschland anders als im englischen Text. Dort lautet sie, ob Ihre Kennzahlen den Schaden sehen können. Hier lautet sie zuerst, ob Sie so messen dürfen — und diese Frage beantworten Sie nicht allein.

Wo das nicht mehr trägt

Nehmen Sie an, Ihr Effekt ist tatsächlich gleichmäßig: derselbe Zuwachs bei den stärksten und den schwächsten Vertriebsmitarbeitenden, derselbe über alle Fallarten, nirgends eine Maschine, die Kundschaft sehen kann. Dann ist der Mittelwert eine völlig taugliche Größe und nichts von alledem trifft zu. Diesen Fall gibt es. Wie oft er zutrifft, ist unbekannt: Jede Arbeit in diesem Beitrag wurde ausgewählt, weil sie Heterogenität gefunden hat, aus dieser Auswahl lässt sich keine Grundrate ablesen. Die vier Schnitte sind der Weg herauszufinden, in welchem Fall Sie sich befinden.

Auch die deutsche Einordnung hat ihre Grenzen, und sie werden regelmäßig überdehnt. Das Recht aus § 87 Abs. 1 Nr. 6 steht dem Betriebsrat zu; wo keiner besteht, besteht auch keine Mitbestimmungspflicht. Der Absatz gilt zudem nur, „soweit eine gesetzliche oder tarifliche Regelung nicht besteht". Und eine Betriebsvereinbarung ist kein Instrument des § 87, sondern des § 77. Die verbreitete Kurzfassung, ein Vertriebsdashboard brauche in Deutschland immer eine Betriebsvereinbarung, ist an beiden Enden falsch.

Die ehrliche Zusammenfassung ist, dass „hat es gewirkt" beantwortet wurde, bevor es gemessen war. Die beiden thematisch nächstliegenden Arbeiten der Vertriebsliteratur bestehen beide aus dem, was Menschen über KI sagen, und nicht aus dem, was sich verändert hat, als sie sie benutzten. Die Arbeiten, die etwas gemessen haben, stehen in der Volkswirtschaftslehre, in der Marketingwissenschaft

und in der Organisationsforschung. Keine davon handelt von Ihnen. Jede von ihnen hat ein Ergebnis gemessen, statt nach einem zu fragen, und erst das hat den Schaden überhaupt auffindbar gemacht.

Acemoglu, D., & Restrepo, P. (2020). The wrong kind of AI? Artificial intelligence and the future of labour demand. Cambridge Journal of Regions, Economy and Society, 13(1), 25–35. https://doi.org/10.1093/cjres/rsz022

Betriebsverfassungsgesetz (BetrVG), § 87 Abs. 1 Nr. 6. Abgerufen am 27. Juli 2026. https://dejure.org/gesetze/BetrVG/87.html

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. The Quarterly Journal of Economics, 140(2), 889–942. https://doi.org/10.1093/qje/qjae044

Bundesarbeitsgericht. (16. Juli 2024). Beschluss des Ersten Senats – 1 ABR 16/23. https://www.bundesarbeitsgericht.de/entscheidung/1-abr-16-23/

Dell'Acqua, F., McFowland, E., III, Mollick, E. R., Lifshitz, H., Kellogg, K. C., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. Organization Science, 37(2), 403–423. https://doi.org/10.1287/orsc.2025.21838

Hautamäki, P., & Heikinheimo, M. (2025). Fully leveraging AI in B2B sales: Exploring sales managers' capabilities and organizational knowledge processes. Journal of Business Research, 194, 115396. https://doi.org/10.1016/j.jbusres.2025.115396

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. Marketing Science, 38(6), 937–947. https://doi.org/10.1287/mksc.2019.1192

Rodriguez, M., Deeter-Schmelz, D. R., & Krush, M. T. (2025). The impact of generative AI technology on B2B sales process and performance: An empirical study. Journal of Business & Industrial Marketing, 40(10), 2013–2027. https://doi.org/10.1108/JBIM-02-2025-0097

ENDE DES BEITRAGS

ZITIERWEISE

Isoglu, S. (2026, 27. Juli). Was KI in Revenue Operations wirklich verändert. Sinan Isoglu. https://isoglu.com/de/journal/was-ki-wirklich-veraendert/

WEITERLESEN

Revenue Operations & KI
Die eine Zahl, die ein kommerzielles Team teilen sollte.
https://isoglu.com/de/journal/die-eine-zahl-kommerzielles-team/

Revenue Operations & KI
Der Funnel-Engpass, den niemand misst.
https://isoglu.com/de/journal/der-funnel-engpass-den-niemand-misst/

Go-to-Market & Pricing
Wenn ein bewährtes Playbook auf einen neuen Markt trifft.
https://isoglu.com/de/journal/bewaehrtes-playbook-neuer-markt/

Sinan Isoglu, MBA (Quantic)
Führungskraft für Umsatzwachstum, Dozent und Doktorand