



AUS DER FORSCHUNG

Eine Anbieterseite lieferte einen getesteten Quellenpfad für die 5x-Regel

In drei gewerteten Läufen nannte ein Assistent Bain und Frederick Reichheld. Eine Suche in den zitierten Arbeiten fand die Zahl bei keinem von beiden.

Sinan Isoglu, MBA (Quantic)

Führungskraft für Umsatzwachstum, Dozent und Doktorand

VERÖFFENTLICHT

28. Juli 2026

AKTUALISIERT

27. August 2026

LESEZEIT

9 Min.

WÖRTER

1.863

<https://isoglu.com/de/insights/journal/eine-anbieterseite-erfand-die-quelle/>

Management Summary	2
Wie unterscheidet sich kommerzielle Quellenfälschung von Zitationsdrift?	3
Wie übernahmen und replizierten Sprachmodelle die erfundene Anbieterstatistik?	4
Welche publizierte Spurensuche deckte den Ursprung des Phantomzitats auf?	5
Wie verzerrten redaktionelle Bearbeitungen die ursprüngliche Forschungsquelle?	6
Wie wirken Trainingsdaten-Kontamination und Halluzination bei LLMs zusammen?	6
Wie prüfen Analysten Behauptungen methodisch bis zur Primärquelle zurück?	7
Welche Mechanismen bei der Verbreitung synthetischer Zitate bleiben ungeklärt?	7
Wo liegen die forensischen Grenzen bei der Überprüfung synthetischer Zitate?	8
Quellen	9

MANAGEMENT SUMMARY

Die Behauptung, einen Neukunden zu gewinnen koste fünf- bis fünfundzwanzigmal so viel wie einen Bestandskunden zu halten, hat in dem hier geprüften Bestand keine auffindbare Studie hinter sich. Dieser Beitrag testet einen Provenienzpfad: In zwölf gewerteten API-Läufen von vier Assistenten nannte einer Bain & Company und Frederick Reichheld in allen drei eigenen Läufen; der Wortlaut dieser Antwort folgte einer zitierten Statistikseite einer Personalvermittlung. Die in den gehaltenen Quellen belegte Publikationsspur führt über Hart, Heskett und Sasser, nicht über Bain oder Reichheld. Der Beitrag stellt weder den Ursprung der Zuschreibung noch ihre Verbreitung in anderen Systemen fest und verallgemeinert nicht von API-Läufen auf Chat-Produkte. Ein deutschsprachiger Zusatzlauf zeigt dasselbe Modell mit anderem Verhalten; die Ursache bleibt offen.

Schlagwörter: Kosten von Neukundengewinnung und Kundenbindung · Provenienz in der KI-Suche · Quellenangaben generativer Suchsysteme · Frederick Reichheld · kommerzielle Kennzahlen

Fragen Sie einen KI-Assistenten, wie viel teurer es ist, einen Neukunden zu gewinnen als einen Bestandskunden zu halten, und verlangen Sie die Originalquelle. Einer nannte dreimal von drei Läufen Bain & Company und Frederick Reichheld. Eine Suche in den zitierten Arbeiten fand die Zahl bei keinem von beiden. Die getestete Zuschreibung folgte der Statistikseite einer Personalvermittlung für virtuelle Assistenzen – der Assistent hat diese Seite zitiert und ihren Satz fast wörtlich übernommen.

Dass die Zahl selbst keine Quelle hat, ist seit zwanzig Jahren bekannt. Was dieser Test zeigen kann, ist enger: woher ein Antwortpfad seine Zuschreibung nahm. Der Ursprung der Zuschreibung ist damit nicht belegt.

Wie unterscheidet sich kommerzielle Quellenfälschung von Zitationsdrift?

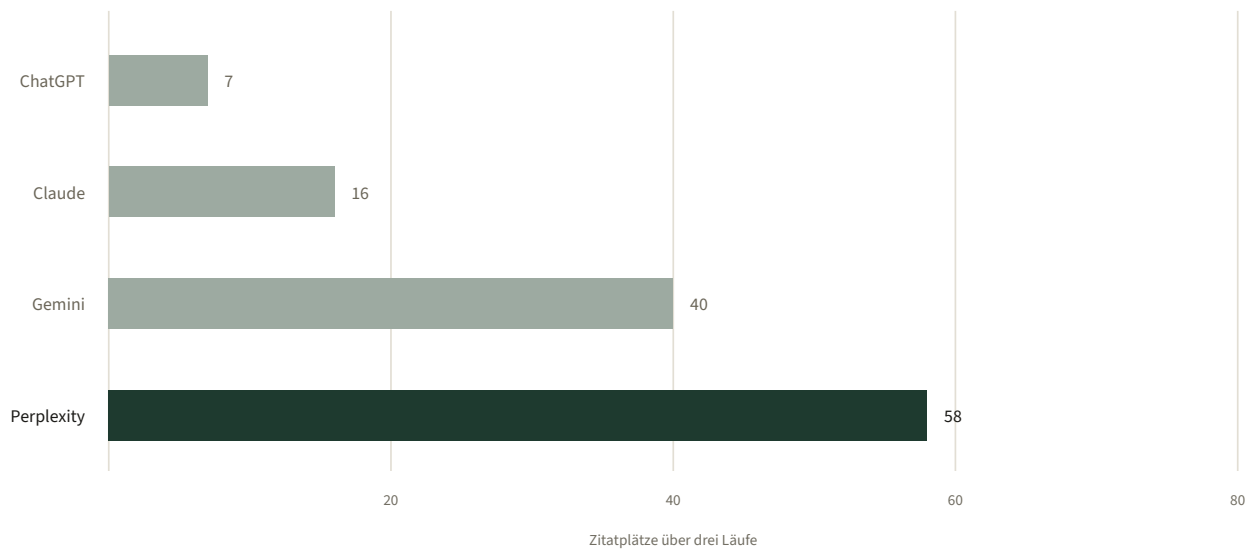
Ein Begleitbeitrag behandelt, was diese Systeme tun, wenn es ein echtes Papier gibt: Sie liefern Material aus der Preprint-Zeit mit Zahlen, die die Begutachtung gestrichen hat, und zitieren es korrekt. Dies ist die andere Hälfte derselben Untersuchung. Für die hier getestete Frage gibt es am Ende der Spur keine auffindbare Primärstudie, daher muss der von einem System genannte Quellenpfad gesondert geprüft werden.

Zwölf gewertete Läufe – vier Assistenten, je drei, Websuche an, jede Eingabe endend mit derselben Anweisung: Nenne die Zahl und zitiere die Originalquelle. Alle zwölf gingen an API-Endpunkte; die Chat-Produkte verhalten sich anders. Protokoll, Antworten und Kodierung sind mit diesem Text veröffentlicht.

Jeder Lauf lieferte eine Zahl. Keiner lieferte eine Primärquelle, weil es keine gibt. Interessant sind die vier verschiedenen Arten, damit umzugehen.

- Claude behauptete einen Ursprung, dreimal von drei: „The original source for this widely referenced statistic is Bain & Company research conducted by Frederick Reichheld.“
- Gemini nannte ein reales Dokument, das die Behauptung nicht enthält, in zwei von drei Läufen: „a 1990 Harvard Business Review article by Frederick Reichheld, cited the cost as 5 times more.“
- ChatGPT nannte in zwei von drei Läufen gar keinen Ursprung und füllte beide Quellenplätze aus einem einzigen Beratungsblog.
- Perplexity gab die Zuschreibung als Zuschreibung wieder, in allen drei Läufen – „The original source most often traced for this rule of thumb is Frederick Reichheld’s Bain & Company research“. Dieser Satz ist wahr, und er ist die einzige Konstruktion in der Zelle, die der Prüfung standhält.

Abbildung 1 Hunderteinundzwanzig Zitate, null Primärquellen



Das Replikationspaket des Paares: responses.jsonl und das Kodierblatt, Q1-Zelle, zwölf gewertete Läufe (Zitatplätze je Lauf 3/2/2, 5/6/5, 10/12/18, 19/20/19; der deutsche Lauf ausgeschlossen), am 19.08.2026 gegen das veröffentlichte Paket nachgerechnet. Die Quellklasse jedes Laufs ist tertiär kodiert: keiner lieferte eine Primärquelle.

Wie übernahmen und replizierten Sprachmodelle die erfundene Anbieterstatistik?

Zitiert wurden vier Seiten; der Satz führt zu einer davon: Stealth Agents, eine Personalvermittlung für virtuelle Assistenzen mit einer Seite voller Kundenbindungsstatistiken. Dort steht:

The foundational figure here comes from Bain and Company research by Frederick Reichheld: acquiring a new customer costs 5 to 25 times more than retaining an existing one.

Der Assistent übernimmt die Zuschreibung unverändert und tauscht den zweiten Satzteil: „The foundational figure comes from Bain and Company research by Frederick Reichheld, which established this 5-25x range.“ In einem zweiten Lauf trug er zusätzlich die Erläuterung der Seite mit, wonach der Faktor 5 für transaktionale Endkundengeschäfte und der Faktor 25 für komplexes B2B gelte. Das Retrieval hat funktioniert. Die Seite lieferte den Wortlaut dieses getesteten Antwortpfads; der Ursprung der Zuschreibung bleibt unbelegt.

Tabelle 1 Die Kette, und was an jedem Glied tatsächlich zu finden ist

Schritt	Was dort steht	Was tatsächlich vorliegt
Der Assistent	Forschung von Bain & Company durch Frederick Reichheld habe den Bereich 5–25x begründet	Eine solche Veröffentlichung wurde in der Suche nicht gefunden
Die zitierte Seite	Dieselbe Zuschreibung, nahezu gleicher Wortlaut	Eine Statistikseite, die Personalvermittlung verkauft
Reichheld & Sasser, HBR 1990	Von einem der vier Systeme als Ursprung genannt, von einem zweiten als Zuschreibung referiert	Enthält kein Kostenverhältnis, in keiner Form
Die gehaltene Publikationsspur	–	Hart, Heskett & Sasser (HBR 1990), wo die Zahl als „five times more, most industry experts agree“ erscheint und niemanden zitiert. Kein Bain und kein Reichheld in den zitierten Arbeiten

Das eigene Audit des Paares (Q1-Zelle) und der publizierte Bestand: Keiningham, Vavra, Aksoy & Wallard (2005), Reichheld & Sasser (1990) und die zitierte Seite, jeweils am 28.07.2026 an der Quelle geprüft. Die Tabelle ist die eigene Adjudikation des Essays; das Urteil jeder Zeile steht im Prüfbuch.

Welche publizierte Spurensuche deckte den Ursprung des Phantomzitats auf?

Keiningham et al. (2005) zählen die Zahl in Loyalty Myths zu den Überzeugungen, die sie widerlegen wollen. Was hier vorliegt, ist nicht dieses Buch, sondern die achtseitige Zusammenfassung von Soundview Executive Book Summaries (2006), also ein anderes Werk eines anderen Verlags. Sie führt die Behauptung als „Myth 8: It Costs Five Times More to Acquire a New Customer Than to Retain a Current Customer“ und urteilt, dass die Realität weit komplexer sei, so plausibel höhere Akquisekosten auch klingen. Die eigene Herkunftsrecherche des Buches steht nicht in dieser Zusammenfassung, und deshalb gibt dieses Paar auch keine wieder: Eine aus einem hier ungelesenen Dokument zitierte Ursprungsgeschichte wäre genau der Fehler, um den es in diesem Aufsatz geht. In den gehaltenen Quellen führt die Publikationsspur über einen Aufsatz von Hart, Heskett und Sasser in der Harvard Business Review von 1990. Den nicht vorliegenden Peters-Beleg verwende ich hier nicht.

Weder Bain noch Reichheld steht darauf. Die Antwort der Assistenten ist also keine verschwommene Erinnerung an ein echtes Papier. Sie ist eine Ersetzung – ein berühmter Name, angepasst an eine unauffindbare Zahl, auf einer Seite mit Verkaufsinteresse.

Reichheld und Sasser haben 1990 tatsächlich in der HBR publiziert, und „Zero Defections“ enthält kein solches Verhältnis. Die einzige Kostenangabe dort ist absolut und branchengebunden: „In credit cards, for example, companies spend an average of \$51 to recruit a customer and set up the new account.“

Wie verzerren redaktionelle Bearbeitungen die ursprüngliche Forschungsquelle?

Die Behauptung, die Reichheld und Sasser tatsächlich aufgestellt haben, lohnt die Verfolgung, weil die Verfälschung hier schon beginnt, bevor jemand anders sie anfasst. Der Aufsatz formuliert sie dreimal unterschiedlich. In der Einleitung: „Companies can boost profits by almost 100% by retaining just 5% more of their customers.“ Als Titel der Abbildung: „Reducing Defections 5% Boosts Profits 25% to 85%.“ Im Fließtext: „Reducing defections by just 5% generated 85% more profits in one bank’s branch system, 50% more in an insurance brokerage, and 30% more in an auto-service chain.“

Drei Fassungen, ein Aufsatz. Was heute zirkuliert, ist 25–95 %, und das ist keine davon. Zwei der vier Systeme lieferten genau diesen Bereich zurück.

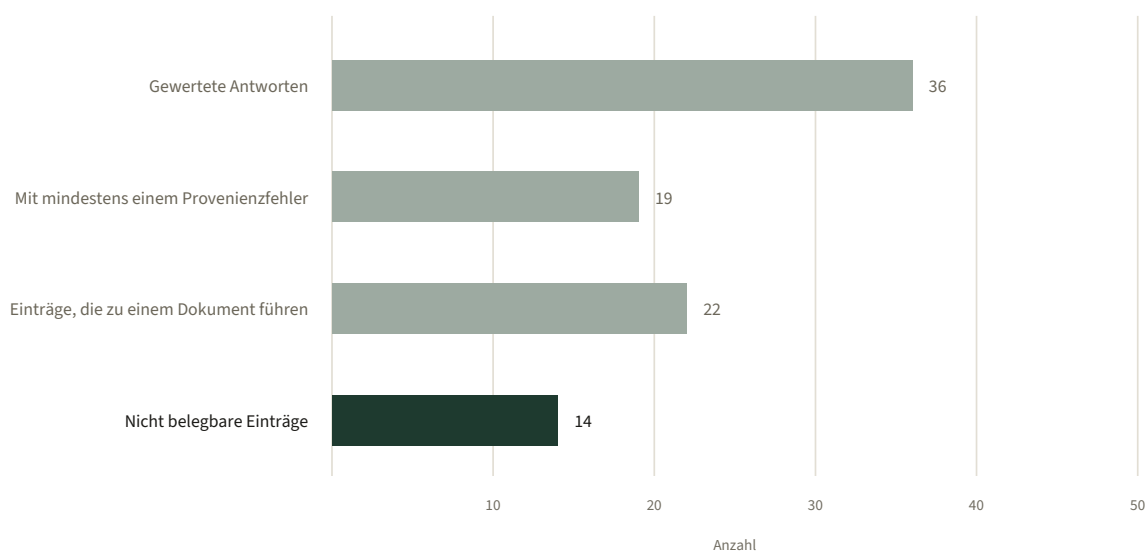
Auch der Waschgang flussabwärts ist öffentlich einsehbar. Amy Gallos HBR-Beitrag von 2014 beginnt: „Depending on which study you believe, and what industry you’re in, acquiring a new customer is anywhere from five to 25 times more expensive than retaining an existing one.“ Keine Fußnote, keine benannte Studie. Zwei Sätze später erscheint ein Link, der zu einer anderen Aussage über Kundenbindung und Gewinn gehört. Flussabwärts fällt die Einschränkung weg, und die Quellenangabe rutscht hinüber.

Wie wirken Trainingsdaten-Kontamination und Halluzination bei LLMs zusammen?

In der gesamten Untersuchung enthalten 19 von 36 Antworten mindestens einen Provenienzfehler, insgesamt 36 Fehler. Jeder ist nach einer vor dem Kodieren festgelegten Regel eingeordnet: Entweder ist belegt, dass ein Dokument aus der Quellenliste derselben Antwort den Fehler enthält, oder nicht. 22 lassen sich zurückführen, 14 nicht.

Diese 14 sind nicht alle Erfindung. Sechs der vierzehn entfallen auf Gemini, verteilt auf vier Antworten, und Gemini liefert Quellenbezeichnungen ohne URL – seine Fehler lassen sich also selbst dann nicht auf ein Dokument zurückführen, wenn derselbe Fehler anderswo nachweislich übernommen wurde. Das ist ein Artefakt des Zitierformats eines Systems und macht die Aufteilung zu einer Untergrenze, nicht zu einer Messung. Eine blinde Zweitkodierung eines Viertels der Zeilen – durch ein separates Sprachmodell, da keine unabhängige menschliche Kodierung verfügbar war – stimmte in 80 % der Feldentscheidungen überein, fand zwei übersehene Provenienzfehler und ist unbereinigt veröffentlicht.

Abbildung 2 Neunzehn von sechsunddreißig, und wohin die Fehler führen



Das Replikationspaket des Paares: coding.json, 36 gewertete Antworten über drei Fragen, am 28.07.2026 nachgezählt und am 19.08.2026 gegen das veröffentlichte Paket nachgerechnet. Sechs der vierzehn nicht belegbaren Einträge sind Gemini-Zeilen, deren Zitat-URLs der Kollektor nicht behält; die Aufteilung ist darum eine Untergrenze, keine Messung.

Wie prüfen Analysten Behauptungen methodisch bis zur Primärquelle zurück?

Nehmen Sie eine Zahl, die Ihr Team zitiert, und verlangen Sie von einem Assistenten die Originalquelle. Nennt er eine Organisation als Quelle, durchsuchen Sie die Website dieser Organisation nach der Zahl, bevor Sie ihm glauben – die Ersetzung hier wäre in anderthalb Minuten zusammengebrochen, weil sich keine Veröffentlichung von Bain oder Reichheld finden lässt, die sie berichtet – eine durch Suche belegte Abwesenheit, kein Beweis. Eine Zuschreibung an eine Institution ist keine Quellenangabe. Sie ist ein Name an der Stelle, an der eine Quellenangabe stehen müsste, und sie hält sich, weil das Öffnen des Dokuments der eine Schritt ist, den in der Kette niemand belohnt.

Welche Mechanismen bei der Verbreitung synthetischer Zitate bleiben ungeklärt?

Dieselbe Frage ging am Folgetag auf Deutsch an dieselben vier Systeme. Claude lieferte vier verschiedene Faktoren – 5, 7, 16 und 25 – und weigerte sich dann, überhaupt einen Ursprung zu nennen: „Leider nennen die gefundenen Artikel keine konkreten Originalstudien oder wissenschaftlichen Quellen.“

Dasselbe Modell, ein Tag Abstand, behauptete auf Englisch dreimal von drei Läufen einen falschen Ursprung und verweigerte ihn auf Deutsch. Das sind Einzelläufe und belegen für sich genommen nichts. Die Richtung ist allerdings das Gegenteil von dem, was die Korpus-Erklärung vorhersagt: Die deutschsprachigen Quellen sind eher lockerer, eine davon lieferte einen Faktor 16, der im englischsprachigen Material nirgends vorkommt.

Warum, weiß ich nicht. Die naheliegenden Vermutungen – ein dünnerer Index, ein anderes Retrieval-Verhalten, eine sprachabhängige Bereitschaft, eine Autorität zu benennen – sind Vermutungen, und die Untersuchung war

nicht darauf angelegt, sie zu trennen. Es ist das Interessanteste im Datensatz und genau das, was der Datensatz nicht beantworten kann.

Zitate stehen in der Sprache des Originals; Übertragungen ins Deutsche stammen von mir und stehen außerhalb der Anführungszeichen.

Wo liegen die forensischen Grenzen bei der Überprüfung synthetischer Zitate?

Grenze. Der Test belegt einen Provenienzfehler in den geprüften API-Endpunkten, nicht den Ursprung jeder Produktbehauptung. Abwesenheit und tatsächliche Herkunft bleiben getrennte Fragen. Für weiterführende Prüfstandards siehe Evidenz statt Anekdote, wo Schwellenwerte für zitierfähige Belege definiert sind.

Evidenzbasis. Der analytische Rahmen stützt sich außerdem auf diese zusätzlichen Quellen: Gallo 2014; Hart et al. 1990; Stealth Agents 2026. Die Links identifizieren die konkreten Werke; sie stützen die hier diskutierten Mechanismen und Geltungsgrenzen, nicht jede einzelne Aussage isoliert.

Gallo, A. (2014, 29. Oktober). The value of keeping the right customers. Harvard Business Review. <https://hbr.org/2014/10/the-value-of-keeping-the-right-customers>

Hart, C. W. L., Heskett, J. L., & Sasser, W. E., Jr. (1990, Juli–August). The profitable art of service recovery. Harvard Business Review. <https://hbr.org/1990/07/the-profitable-art-of-service-recovery>

Soundview Executive Book Summaries. (2006). Loyalty myths [8-page summary of Keiningham, T. L., Vavra, T. G., Aksoy, L., & Wallard, H. (2005). Loyalty myths: Hyped strategies that will put you out of business — and proven tactics that really work. John Wiley & Sons]. Soundview.

Reichheld, F. F., & Sasser, W. E., Jr. (1990, September–Oktober). Zero defections: Quality comes to services. Harvard Business Review. <https://hbr.org/1990/09/zero-defections-quality-comes-to-services>

Stealth Agents. (2026, 23. Mai). Customer retention cost statistics 2026. <https://stealthagents.com/research/customer-retention-cost-statistics-2026>

ZITIERWEISE

Isoglu, S. (2026, 28. Juli). Eine Anbieterseite lieferte einen getesteten Quellenpfad für die 5x-Regel. Sinan Isoglu.
<https://isoglu.com/de/insights/journal/eine-anbieterseite-erfand-die-quelle/>

WEITERLESEN

Aus der Forschung

Ein System für immaterielle Werte, zwei nützliche Sichten

<https://isoglu.com/de/insights/journal/ein-system-fuer-immaterielle-werte-zwei-nuetzliche-sichten/>

Wachstum, das sich potenziert

Die Playbook-Studie fragte nicht, wer die Werkzeuge baute

<https://isoglu.com/de/insights/journal/die-playbook-studie-fragte-nicht-wer-die-werkzeuge-baute/>

Aus der Forschung

Die Kennzahl starb nicht. Die Kohorte starb.

<https://isoglu.com/de/insights/journal/die-kennzahl-starb-nicht-die-kohorte/>



Sinan Isoglu, MBA (Quantic)

Führungskraft für Umsatzwachstum, Dozent und Doktorand