WHITEPAPER · DELEGATIONSGRENZE

Eine Trefferquote ist keine Güte

Warum maschinelle Bewertungsstellen in zwei Richtungen irren, warum Sie nur eine der beiden sehen, und was vor dem Einsatz zu entscheiden ist

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

AUF EINEN BLICK

In vielen Unternehmen bereitet heute ein maschinelles Verfahren Entscheidungen vor. Es sichtet Bewerbungen, bewertet Anträge, filtert Meldungen, schlägt Lieferanten vor. Die Entscheidung selbst bleibt gewollt beim Menschen.

Nur wird sie dort meistens bestätigt und nicht getroffen.

Das ist kein Vorwurf an die Beteiligten. Es ist die Folge einer Architektur, die sich vor dem Einsatz festlegen lässt und danach kaum noch. Dieses Whitepaper beschreibt, woran das liegt, welche Zahl dabei fehlt und welche drei Festlegungen vor dem Start zu treffen sind.

Teil A richtet sich an die Geschäftsführung. Was an einer Übergabestelle wirklich passiert, warum die Herstellerangabe Ihre Frage nicht beantwortet und welche Entscheidungen vor dem Einsatz fallen müssen.

Teil B richtet sich an Fachbereich und Controlling. Die beiden Prüfschwellen, der Rückflusskanal, der Prüfvorlauf und die Frage, woher die Zahlen kommen.

Die vier Sätze, auf die es ankommt:

#	Kernsatz
1	Ein Verfahren irrt in zwei Richtungen, und beide haben ungleiche Folgen. Eine einzelne Trefferquote verbirgt, in welche der beiden Richtungen Sie irren
2	Sie beobachten den Ausgang nur für die Fälle, die Sie durchgelassen haben. Wer unter den Abgelehnten geeignet gewesen wäre, erscheint in keiner Beobachtung
3	Die Schwelle, bis zu der sich das Nachprüfen einer Ablehnung lohnt, hängt an genau der Größe, die Sie dadurch nie messen. Sie korrigiert sich nicht von selbst
4	Das wirksamste Gegenmittel ist ein Beschwerdeweg, den Sie nicht steuern. Er kostet Sie nichts, und wo er fehlt oder ungenutzt bleibt, hält der Irrtum

TEIL A

Für die Geschäftsführung

1. Bestätigen ist nicht entscheiden

Eine Übergabestelle sieht im Organigramm harmlos aus. Ein Verfahren bewertet, ein Mensch entscheidet. Die Zuständigkeit ist geklärt, die Aufsicht ist besetzt, der Haftungsfall ist zugeordnet.

In der Sache verschiebt sich dabei etwas, das kein Organigramm zeigt. Wer eine Empfehlung bestätigt, ohne sie zu prüfen, hat die Entscheidung nicht getroffen, sondern nachvollzogen. Die formale Zuständigkeit bleibt beim Menschen. Die tatsächliche liegt beim Verfahren, das die Empfehlung erzeugt.

Dafür gibt es seit langem einen Begriff. Die ökonomische Organisationstheorie unterscheidet formale und reale Autorität. Wer formal entscheidet, entscheidet nicht real, wenn die entscheidende Information bei einem anderen liegt und ihre Überprüfung Aufwand kostet.

Neu ist nicht diese Verschiebung. Neu ist, womit man es hier zu tun hat. Ein menschlicher Zuarbeiter kann befragt, widerlegt und verglichen werden, und er schweigt, wenn er nichts weiß. Ein maschinelles Verfahren schweigt nicht. Es liefert für jeden Fall eine Empfehlung, in gleicher Form und mit gleichem Nachdruck, und es hat keinen eigenen Vorteil, den es verfolgt.

Daraus folgt eine unbequeme Einsicht für die üblichen Abhilfen. Kontrollen gegen Eigeninteresse greifen nicht, weil es kein Eigeninteresse gibt. Die Prüfung unterbleibt nicht, weil jemand etwas anderes will, sondern weil sie sich nicht mehr zu lohnen scheint.

Die Prüfung verschwindet nicht durch Nachlässigkeit. Sie verschwindet, weil das Verfahren als so gut gilt, dass sie sich nicht mehr lohnt. Der Grund ist eine Rechnung, keine Haltung.

2. Zwei Richtungen, ungleiche Folgen

Jede Stelle einer Entscheidungskette kann auf zwei Arten irren. Stelle heißt hier jede Stufe der Kette, ob Verfahren oder Mensch. Sie kann einen Fall stoppen, der weitergehört hätte. Und sie kann einen Fall durchlassen, der zu stoppen gewesen wäre.

Das klingt nach einer Spitzfindigkeit und ist der Kern der Sache. Die beiden Richtungen haben ungleiche Folgen, weil der entgangene Ertrag und der eingetretene Schaden verschieden groß sind. Sie sind auch verschieden häufig, denn geeignete Fälle sind meist selten, etwa die passende Bewerbung unter vielen.

Deshalb beschreibt eine einzelne Zahl eine Stelle nicht. Eine Trefferquote, eine Genauigkeit, ein Qualitätswert sagt Ihnen, wie oft die Stelle richtig liegt. Die Richtung, in der sie falsch liegt, nennt sie nicht, obwohl Sie genau diese Angabe für eine Entscheidung brauchen.

Was daraus für die Gestaltung folgt. Mehrfachprüfung verschiebt die beiden Richtungen gegeneinander, sie verbessert die Stelle nicht. Lassen Sie zwei Prüfungen hintereinander laufen und verlangen Sie, dass beide zustimmen, so kommen weniger ungeeignete Fälle durch, aber mehr geeignete werden gestoppt. Lassen Sie es genügen, dass eine zustimmt, so ist es umgekehrt.

Wer mehrfach prüfen lässt, wählt damit die Fehlerrichtung, in die er irren will. Die Geschäftsführung trifft diese Entscheidung, sie ist keine Einstellung in einer Software.

3. Warum die Herstellerangabe Ihre Frage nicht beantwortet

Nehmen Sie an, der Anbieter gibt die beiden Fehlerraten seines Verfahrens vollständig und wahrheitsgemäß an. Sie wissen also genau, wie oft es einen guten Fall stoppt und wie oft es einen schlechten durchlässt.

Damit ist Ihre Frage noch nicht beantwortet. Ihre Frage lautet nicht, wie gut das Verfahren ist, sondern wie viele seiner Absagen zu Unrecht fallen. Das hängt davon ab, wie viele geeignete Fälle bei Ihnen überhaupt ankommen. Diese Mischung Ihrer eigenen Fälle kann der Anbieter nicht kennen.
Dasselbe Verfahren, vier Betriebe, verschiedene Fallmischung. Das Verfahren stoppt 35 von 100 geeigneten Fällen und lässt 15 von 100 ungeeigneten durch. In jedem Betrieb kommen 1.000 Fälle an.

Anteil geeigneter Fälle	Absagen insgesamt	davon zu Unrecht	Anteil falscher Absagen
2 von 100	840	7	0,8 Prozent
10 von 100	800	35	4,4 Prozent
20 von 100	750	70	9,3 Prozent
60 von 100	550	210	38,2 Prozent

Modellrechnung mit offengelegten Beispielwerten, 1.000 Fälle je Betrieb. Dasselbe Verfahren in jeder Zeile.
Eine Nachprüfung der Absagen lohnt, solange sie weniger kostet als dieser Anteil vom Ertrag eines richtig bearbeiteten Falls. Bei 0,8 Prozent lohnt sie fast nie, bei 38,2 Prozent fast immer. Die Schwelle streut hier um den Faktor 46, und das Verfahren ist in jeder Zeile dasselbe.
Die Größe, die über die Nachprüfung entscheidet, ist keine Eigenschaft des Verfahrens. Sie ergibt sich erst aus Verfahren und eigenem Zulauf zusammen. Herstellerangaben und Validierungsstudien liefern die eine Hälfte und können die andere nicht liefern.
Und jetzt kommt der unangenehme Teil. Die fehlende Hälfte ist genau die Größe, die Sie unter den üblichen Bedingungen nicht messen können.

4. Der Zustand, in dem nichts auffällt

Sie beobachten den Ausgang eines Falles nur, wenn Sie ihn durchgelassen haben. Was aus einem abgelehnten Fall geworden wäre, entsteht als Beobachtung nicht. Die Bewerberin, die Sie aussortiert haben, bewährt sich nicht bei Ihnen. Der Kredit, den Sie abgelehnt haben, wird nicht zurückgezahlt.
Daraus folgt eine Unsymmetrie, die alles Weitere trägt. Was unter den Durchgelassenen schiefgeht, ist messbar. Was unter den Abgelehnten richtig gewesen wäre, ist es nicht.

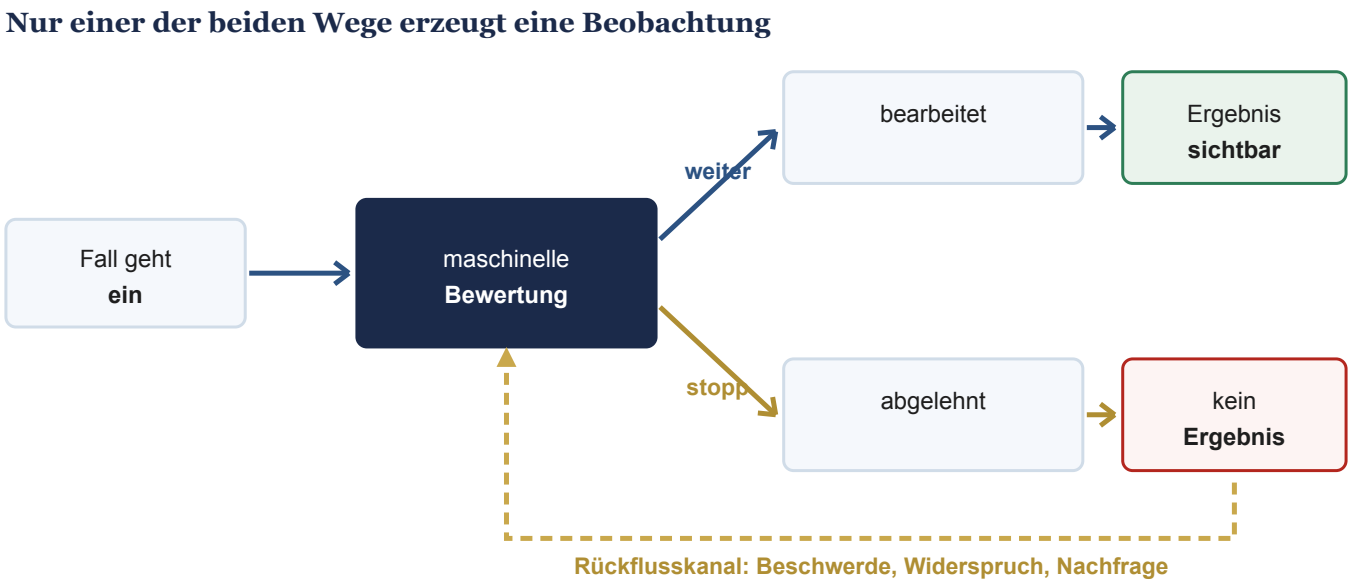


Abbildung 1. Die Übergabestelle und der zensierte Rückkanal. Was weitergereicht wird, erzeugt ein Ergebnis und damit eine Beobachtung. Was gestoppt wird, erzeugt keines. Der gestrichelte Weg ist der einzige, über den eine Ablehnung ohne Zutun der Organisation wieder aufgeht.

Für Ihre Zahlen heißt das: Die eine Fehlerrichtung lernen Sie mit der Zeit, die andere nie. Und die Schwelle, bis zu der sich das Nachprüfen einer Ablehnung lohnt, hängt an genau der Richtung, die Sie nicht lernen.

Daraus entsteht eine Lage, die sich selbst stabilisiert. Wer sein Verfahren für gut hält, prüft dessen Ablehnungen nicht. Wer sie nicht prüft, erfährt nichts über das Verfahren. Wer nichts darüber erfährt, hält es weiter für gut. Ein Denkfehler der Beteiligten liegt nicht vor. Die Lage folgt daraus, dass nur die Prüfung Daten erzeugt.

Besonders unangenehm ist die Richtung dieser Falle. Sie hält nur die zu gute Meinung fest, nicht die zu schlechte. Wer sein Verfahren für schlechter hält, als es ist, prüft, lernt und hört dann auf. Wer es für besser hält, prüft nicht und bleibt dabei.

Eine Organisation in dieser Lage kann nicht feststellen, ob sie darin ist. Der Unterschied zwischen einer richtigen und einer zu günstigen Einschätzung verlangt genau die Größe, die sie nicht hat.

Die unbequeme Aussage, auf die dieses Whitepaper hinausläuft, steht damit. Alles Weitere beschreibt, was sich dagegen tun lässt.

5. Wo das auftritt

Nicht jede maschinell vorbereitete Entscheidung ist betroffen. Es gibt einen einfachen Test, und er hat nichts mit der Technik zu tun.

Die Frage lautet: Geht eine Ablehnung ohne Ihr Zutun wieder auf, und zwar oft genug, dass Sie es in Ihren Zahlen sehen? Gemeint ist, ob ein Abgelehnter widersprechen kann und ob er es oft genug tut, dass Sie aus den Ergebnissen etwas lernen.

Wo das messbar geschieht, erledigt sich das Problem weitgehend von selbst, und Kapitel 10 zeigt mit welchen Größenordnungen. Wo es ausbleibt, hält der Irrtum, und niemand merkt es.

In Ihrem Unternehmen	Geht eine Ablehnung messbar ohne Sie wieder auf?
Bewerbungsabsage	Nein, jedenfalls nicht messbar. Die Abgelehnte kann eine menschliche Prüfung verlangen und eine Erläuterung, nur tut es kaum jemand, und erhoben wird es nicht
Kreditablehnung	Nein, jedenfalls nicht messbar. Bei einer ausschließlich automatisierten Entscheidung besteht ein Anfechtungsrecht, und ob Ihre Entscheidung als ausschließlich automatisiert gilt, hängt genau an der Frage dieses Whitepapers. Eine Quote der Inanspruchnahme ist nicht veröffentlicht
Lieferantenablehnung	Nein. Das Anfechtungsrecht der DSGVO schützt natürliche Personen und greift für einen Lieferanten als Unternehmen in der Regel nicht. Ein Beschwerdeweg ist selten vorgesehen, eine Quote nicht veröffentlicht
Vergabe unterhalb der Schwellenwerte	Nein. Ein Nachprüfungsverfahren gibt es erst oberhalb
Reklamation, Gewährleistungsfall	Ja. Der Betroffene hat ein Interesse und einen Weg
Antrag mit gesetzlichem Widerspruchsrecht	Ja, und gemessen. Gegen sieben bis acht von hundert Bescheiden wird Widerspruch eingelegt
Gemeldeter Inhalt auf einer großen Plattform	Ja, und gemessen. Das Beschwerdeverfahren ist seit 2024 vorgeschrieben, gegen fünf bis elf von hundert Entfernungen wird es genutzt

Ein Wort zur Rechtslage, denn sie verschiebt diesen Test nicht. Für die Bewerbungs- und die Kreditablehnung bestehen Rechte, nämlich das Eingreifen einer Person zu erwirken, den eigenen Standpunkt darzulegen, die Entscheidung anzufechten, und bei Hochrisikosystemen eine Erläuterung zu verlangen. An Ihren Zahlen ändert das nichts, solange niemand davon Gebrauch macht. Was im Betrieb wirkt, ist nicht das Recht, sondern die Quote, also der Anteil der Ablehnungen, der tatsächlich wieder aufgeht. Für diese Quote gibt es in den ersten vier Zeilen keine veröffentlichte Zahl.

Die ersten vier Zeilen sind die teuren, und sie haben eine Gemeinsamkeit, die über das Einzelne hinausgeht. Es sind genau die Entscheidungen, bei denen der Abgelehnte den Betrieb verlässt und nicht wiederkommt. Er beschwert sich nicht, weil er von einem Recht meist nichts weiß und davon nichts hat, und er taucht in keiner Statistik auf, weil kein Vorgang entsteht.

Eine zweite Unterscheidung lohnt sich. Betroffen sind vor allem Stellen, an denen der gesuchte Fall selten ist. Je seltener er ist, desto früher bricht die Nachprüfung von Ablehnungen weg, und desto später fällt es auf. Kapitel 7 rechnet das aus. Für die Einordnung genügt die Faustregel: Wo Sie viel ablehnen und wenig annehmen, ist die Gefahr am größten.

6. Die drei Entscheidungen vor dem Einsatz

Drei Festlegungen entscheiden über den Ertrag, und alle drei sind vor dem Start zu treffen. Danach sind sie teuer. Erstens, die Fehlerrichtung. Legen Sie fest, in welche Richtung die Kette irren soll, und begründen Sie es mit Ertrag und Schaden Ihres Falles und nicht mit der Güte des Werkzeugs. Aus dieser Festlegung folgt die Verschaltung, also ob mehrere Prüfungen zustimmen müssen oder ob eine genügt.

Zweitens, der Rückflusskanal. Legen Sie fest, auf welchem Weg eine Ablehnung wieder aufgehen kann, ohne dass Sie es entscheiden. Kapitel 10 zeigt, dass dies das wirksamste bekannte Mittel ist und dass es Sie nichts kostet. Drittens, der Prüfvorlauf. Legen Sie fest, wie viele Ablehnungen am Anfang nachgeprüft werden, auch wenn sich das im Einzelfall nicht lohnt. Kapitel 11 nennt die Größenordnung.

Was diese drei Festlegungen gemeinsam haben, ist der Zeitpunkt. Sie kosten vor dem Start wenig und sind danach kaum nachzuholen, weil die Beobachtungen, die Sie bräuchten, in der Zwischenzeit nicht entstanden sind.

TEIL B

Für Fachbereich und Controlling

7. Die zwei Prüfschwellen, und warum nur eine sich selbst korrigiert

An einer Übergabestelle gibt es nicht eine Prüfentscheidung, sondern zwei. Die Stelle empfiehlt entweder weiter oder stopp, und der Wert einer Prüfung ist in beiden Fällen ein anderer.

Empfehlung	Was ohne Prüfung geschieht	Was eine Prüfung wert ist
weiter	Der Fall wird bearbeitet	Der vermiedene Schaden einer falschen Bearbeitung
stopp	Der Fall wird abgelehnt	Der entgangene Ertrag einer zu Unrecht abgelehnten Bearbeitung

Daraus folgen zwei Schwellen mit verschiedenen Einsätzen. Geprüft wird jeweils dann, wenn die Prüfkosten unter dem Wert der Prüfung liegen. Je genauer die Stelle gilt, gleich in welcher der beiden Richtungen, desto weniger ist eine Prüfung an beiden Schwellen wert. Eine Stelle, der mehr zugetraut wird, wird in beide Richtungen weniger geprüft.

Die beiden Schwellen verschwinden aber nicht gleich schnell. Ist der geeignete Fall selten, so bricht die Schwelle für das Prüfen von Ablehnungen zuerst ein. Dann hört ein Betrieb auf, Ablehnungen nachzusehen, während er Freigaben noch prüft.

Und hier greifen zwei Mechanismen ineinander, die man leicht getrennt betrachtet:

Die Schwelle, die zuerst einbricht, ist dieselbe, die sich nie von selbst korrigiert. Die andere korrigiert sich.

Der Grund ist der aus Kapitel 4. Die Schwelle für Freigaben hängt an Größen, die aus jedem bearbeiteten Fall gelernt werden. Die Schwelle für Ablehnungen hängt allein an der Größe, über die keine Beobachtung entsteht, solange keine Ablehnung zurückkommt.

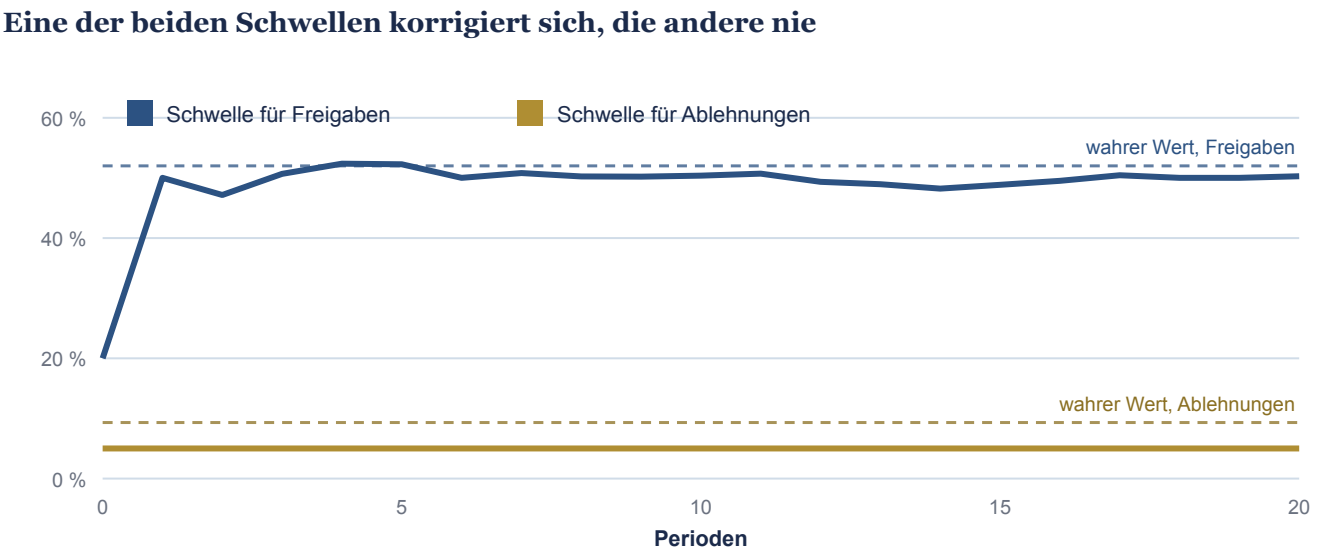


Abbildung 2. Die beiden Schwellen über zwanzig Perioden. Die Schwelle für Freigaben startet zu niedrig und läuft binnen weniger Perioden auf den wahren Wert. Die Schwelle für Ablehnungen bleibt genau dort stehen, wo sie begonnen hat, und zwar für immer. Modellrechnung, Beispielwerte offengelegt.

Die Abbildung zeigt beides an einer Modellrechnung. Beide Schwellen starten zu niedrig, weil die Stelle zu günstig eingeschätzt wird. Die obere ist nach vier Perioden am richtigen Wert. Die untere steht nach zwanzig Perioden noch genau dort, wo sie begonnen hat. Gerechnet ist das für den Fall, dass keine Ablehnung ohne Ihr Zutun wieder aufgeht, und unter dieser Bedingung stünde sie dort auch nach zwanzig Jahren. Wie schnell sich das ändert, sobald es einen Rückflusskanal gibt, steht in Kapitel 10.

Eine Nebenbemerkung, die in der Praxis oft übersehen wird. Die Prüfung einer Freigabe hat keinen Informationswert, denn der Ausgang wird ohnehin sichtbar. Sie vermeidet einen Schaden, sie lehrt nichts. Der einzige Kanal, über den Sie etwas über Ihre Ablehnungen erfahren, ist die Prüfung einer Ablehnung. Genau dieser Kanal wird als erster geschlossen.

8. Wie Sie Ihre Fallmischung schätzen

Die ganze Rechnung hängt an einer Größe, die Ihnen niemand liefert, nämlich am Anteil der wirklich geeigneten Fälle in dem, was bei Ihnen eingeht. Die Frage ist also berechtigt, woher Sie ihn nehmen sollen, wenn Sie die Ablehnungen gerade nicht beobachten.

Vier Wege gibt es, verschieden gut. Keiner ist exakt, und exakt muss auch keiner sein.

Erstens, die bewährten Annahmen als Untergrenze. Der Anteil der Angenommenen, die sich später bewährt haben, ist eine untere Schranke für den Anteil der Geeigneten, denn zu ihnen kommen nur noch die zu Unrecht Abgelehnten hinzu. Die reine Annahmequote taugt dafür nicht, weil unter den Angenommenen auch Ungeeignete sind, die das Verfahren durchgelassen hat. Die Schranke ist grob und trotzdem brauchbar, denn sie grenzt nach unten ein.

Zweitens, die Stelle vor der Stelle. Wenn vor der maschinellen Bewertung eine Vorauswahl steht, dann kennen Sie deren Durchlassquote. Je schärfer vorgefiltert wird, desto höher ist der Anteil der Geeigneten in dem, was

ankommt. Eine Stelle am Ende einer scharfen Kette sieht einen ganz anderen Fallstrom als dieselbe Stelle am Anfang.
Drittens, die Veränderung über die Zeit. Sie müssen den Anteil nicht kennen, um zu merken, dass er sich bewegt. Ein Einbruch der Bewerberzahl bei gleichbleibender Annahmequote, ein neuer Kanal, eine veränderte Ausschreibung: Jede dieser Änderungen verschiebt die Mischung, und damit verschiebt sich die Schwelle, bis zu der sich eine Nachprüfung lohnt. Die Schwelle ist keine Konstante, die einmal gesetzt wird.
Viertens der einzige harte Weg, eine Stichprobe von Ablehnungen. Nehmen Sie eine kleine, zufällig gezogene Zahl abgelehnter Fälle und lassen Sie sie unabhängig zweitbeurteilen. Sie ist aufwendig und einmalig, liefert aber als einzige Beobachtung die fehlende Hälfte. Wie groß die Stichprobe sein muss, ist dabei eine Frage der Genauigkeit und nicht dieselbe Frage wie die Länge des Prüfvorlaufs aus Kapitel 11. Bei einem erwarteten Anteil um zehn Prozent liefern 75 Fälle etwa sieben bis acht Treffer, und der Schätzfehler liegt bei rund dreieinhalb Prozentpunkten. Für die Einordnung genügt das, für eine Schwelle auf den Euro nicht.
Die Fallmischung ist keine Zahl, die Sie beschaffen. Sie ist eine Annahme, die Sie treffen, aufschreiben und bei jeder Änderung des Zugangs überprüfen. Wer sie nicht aufschreibt, hat sie trotzdem getroffen, nur unbemerkt.
Eine Warnung zum Schluss. Die Verzerrung Ihrer eigenen Schätzung hat eine Richtung, die vom Arbeitspunkt des Verfahrens abhängt und deshalb nicht allgemein anzugeben ist. Das Modell sagt nur, in welche Richtung eine zu niedrig angesetzte Mischung wirkt, nämlich hin zu zu wenig Nachprüfung. Dass Betriebe ihre Mischung systematisch zu niedrig ansetzen, behauptet es nicht.

9. Modellrechnung: zwei Betriebe, dasselbe Verfahren

Die folgende Rechnung zeigt, worauf es hinausläuft. Sie bleibt eine Modellrechnung mit offengelegten Beispielwerten. Sie ersetzt keine betriebsindividuelle Kalkulation.
Die Lage. Zwei Betriebe setzen dasselbe Verfahren zur Vorauswahl von Bewerbungen ein. Es stoppt 35 von 100 geeigneten Bewerbungen und lässt 15 von 100 ungeeigneten durch. Beide Betriebe kennen diese beiden Zahlen, der Anbieter hat sie offengelegt. Eine richtige Einstellung ist dem Unternehmen 20.000 Euro wert. Eine strukturierte Zweitdurchsicht einer Absage kostet 1.200 Euro.
Der Unterschied liegt allein im Zulauf. Bei Betrieb A sind 5 von 100 Bewerbungen wirklich geeignet, bei Betrieb B sind es 25 von 100.

	Betrieb A	Betrieb B
geeignete Bewerbungen im Zulauf	5 Prozent	25 Prozent
Absagen je 1.000 Bewerbungen	825	725
davon zu Unrecht abgesagt	2,1 Prozent	12,1 Prozent
Zweitdurchsicht lohnt, bis sie so viel kostet wie	424 Euro	2.414 Euro
Entscheidung bei 1.200 Euro Prüfkosten	nicht prüfen	**prüfen**

Modellrechnung, Beispielwerte offengelegt.
Jetzt der Punkt. Beide Betriebe orientieren sich an derselben Zahl, etwa an einem Erfahrungswert aus der Branche oder an dem, was der Anbieter aus anderen Einführungen berichtet. Diese Zahl liegt bei rund 2 Prozent zu Unrecht

Abgesagten, also bei dem Wert von Betrieb A. Gemeint ist damit nicht die Fallmischung aus Kapitel 3, sondern der Anteil unter den Absagen.
Beide verzichten daraufhin auf die Zweitdurchsicht, denn 1.200 Euro liegen über der Schwelle von 424 Euro.
Für Betrieb A ist das richtig. Seine wahre Schwelle liegt tatsächlich bei 424 Euro, die Prüfung lohnt nicht, und der Verzicht kostet ihn nichts.
Für Betrieb B ist es falsch. Seine Durchsicht kostet 1.200 Euro und wäre 2.414 Euro wert. Jede Absage, die er nicht durchsieht, ist eine Entscheidung gegen eine Prüfung, die sich gelohnt hätte.
Zwei Betriebe, dasselbe Verfahren, dieselbe Annahme, dieselbe Entscheidung. Für einen ist sie richtig und für den anderen falsch, und keiner von beiden kann das an seinen eigenen Daten erkennen.
Was in Betrieb B anschließend geschieht, ist das eigentlich Bemerkenswerte. Nichts. Die eingestellten Bewerber bewähren sich wie erwartet, die Quote der Fehlbesetzungen bleibt unauffällig, die Kennzahlen des Verfahrens sehen gut aus. Die 12,1 Prozent zu Unrecht Abgesagten erscheinen in keiner Auswertung, weil sie nie eingestellt wurden.
Eine wichtige Grenze dieser Rechnung ist zu nennen. Das Modell unterstellt, dass jeder zu Unrecht abgesagte Bewerber bei Einstellung den vollen Ertrag erbracht hätte. In der Praxis ist die Zahl der Stellen begrenzt, ein übersehener Bewerber wird also teilweise durch einen anderen ersetzt. Der Gesamtschaden ist deshalb kleiner als die einfache Hochrechnung, und dieses Whitepaper rechnet ihn aus diesem Grund nicht hoch. Was die Rechnung trägt, ist der Vergleich der beiden Schwellen und nicht eine Schadenssumme.

10. Der Rückflusskanal als wirksamstes Mittel

Bis hierher klingt die Lage ausweglos. Sie ist es nicht, und das Gegenmittel ist weniger aufwendig als erwartet.
Entscheidend ist, ob ein abgelehnter Fall ohne Ihr Zutun wieder aufgehen kann. Ein Widerspruch, eine Beschwerde, eine Nachprüfung auf Verlangen des Betroffenen: Jeder solche Vorgang erzeugt genau die Beobachtung, die Ihnen sonst fehlt, und er kostet Sie keine Prüfkosten, weil ihn ein anderer auslöst.
Wie stark das wirkt, lässt sich an Feldern ablesen, die solche Zahlen veröffentlichen.

Feld	Anfechtungen je 100 Entscheidungen	Anteil der Anfechtungen mit Erfolg
Pflegebegutachtung, 2022	7,3 Prozent	29,6 Prozent
Rentenbescheide, 2024	8,0 Prozent	27,1 Prozent
Bürgergeld, 2025	nicht ausgewiesen	30,9 Prozent
Inhaltsmoderation, Anbieter 1, 2025	11,1 Prozent	44,2 Prozent
Inhaltsmoderation, Anbieter 2, 2024	5,2 Prozent	32,3 Prozent

Amtliche Meldungen und Pflichtberichte, Quellen im Anhang. Die erste Spalte bezieht sich auf alle Bescheide beziehungsweise alle Entfernungen und nicht allein auf die Ablehnungen; als Anteil der Ablehnungen wäre sie höher. Die Erfolgsquote bezieht sich auf die Anfechtungen und ist nach oben verzerrt, weil anficht, wer sich im Recht fühlt.
Zwei Dinge sind daran bemerkenswert. Erstens liegen die Quoten über fünf Felder, zwei Gattungen und mehrere Jahre im selben Bereich. Zweitens genügt dieser Bereich vollständig. Nach dem zugrunde liegenden Modell hält

der beschriebene Zustand in diesen Feldern nicht Jahre, sondern Tage, in der Inhaltsmoderation sogar nur Minuten, weil die Fallzahlen dort so groß sind.
Ein Beschwerdeweg mit wenigen Prozent Inanspruchnahme erledigt das Problem. Er muss nicht funktionieren wie ein Kontrollsystem. Er muss nur existieren und darf nicht von Ihnen abhängen.
Die Umkehrung ist der eigentliche Befund dieses Whitepapers. Zu erwarten ist der Zustand dort, wo kein solcher Weg gemessen begangen wird. Das sind die Felder, in denen eine Ablehnung praktisch endgültig bleibt, gleich ob ein Recht darauf besteht:
die Absage auf eine Bewerbung,
die Ablehnung eines Kredits,
die Nichtberücksichtigung bei einer Vergabe unterhalb der Schwellenwerte.
Und für keines dieser Felder gibt es eine Zahl. Zufall oder Versäumnis der Statistik ist das nicht. Wo kein Vorgang entsteht, entsteht keine Erhebung. Die Größe ist genau dort nicht belegbar, wo sie wirkt.
Was das praktisch heißt. Ein freiwilliger Rückkanal, der nichts kostet, ist in diesen Feldern das wirksamste, was Sie tun können. Eine Begründung auf Nachfrage, eine zweite Durchsicht auf Wunsch, eine offene Rückfragemöglichkeit mit ausgewertetem Ergebnis. Entscheidend ist nicht die Form, sondern dass der Anstoß nicht von Ihnen kommt.

11. Der Prüfvorlauf, die Zahl für den Anfang

Die Schwellenregel aus Kapitel 7 wägt den Nutzen einer Prüfung für den einzelnen Fall gegen ihre Kosten. Sie zählt nicht mit, dass eine Prüfung auch etwas lehrt.
Wer das mitzählt, prüft am Anfang mehr, als sich im Einzelfall lohnt. Nicht dauerhaft, denn die Information ist nach wenigen Durchgängen gekauft und würde danach weiter bezahlt. Die praktikable Form ist deshalb ein Vorlauf. Prüfen Sie die ersten Ablehnungen nach, auch wenn Sie sie für richtig halten und sich jede einzelne Prüfung nicht rechnet. Danach gilt wieder die Schwellenregel.
Wie lang dieser Vorlauf ist, lässt sich ausrechnen. In der Modellrechnung des zugrunde liegenden Working Papers beträgt er eine Periode. Bei dem dort gerechneten Zulauf von 100 Fällen je Periode sind das 75 abgelehnte Fälle. Legen Sie ihn deshalb als Anzahl abgelehnter Fälle fest, nicht als Zeitraum. Wie lange er dauert, hängt an Ihrem Zulauf.
Drei Lagen, in denen der Vorlauf null ist, jede mit eigenem Grund:

Lage	Warum kein Vorlauf nötig ist
Es gibt einen dichten Rückflusskanal	Die Aufdeckung kommt dann ohnehin, und die Prüfung kauft nur Zeit. Wenige Prozent Inanspruchnahme beenden den Zustand, ersetzen den Vorlauf aber nicht; in der Modellrechnung ist der Vorlauf auch bei zehn Prozent Rückfluss noch positiv
Die Einschätzung sitzt sehr fest	Was Sie erfahren, ändert Ihr Verhalten innerhalb des Horizonts nicht mehr
Die Prüfung ist sehr teuer	Auch mit richtiger Einschätzung würde nicht geprüft

Die zweite Zeile ist die unangenehme. Dort, wo eine Einschätzung am längsten hält, lohnt es am wenigsten, sie zu erschüttern. Eine Paradoxie liegt nicht vor, es wirkt dieselbe Größe in zwei Rollen.

Was ausdrücklich nicht empfohlen wird. Ein Anteil von Fällen, die zufällig gegen die Empfehlung behandelt werden, erzeugt dieselbe Information. In den Rechnungen des Working Papers lohnt er sich fast nirgends, und dort, wo er positiv wird, ist das Motiv ein anderes als Lernen. Dazu kommt ein Einwand aus der Fachliteratur, der schwerer wiegt als jede Rechnung. Wo Entscheidungen folgenreich sind, ist es nicht vertretbar, einen Teil davon zu würfeln, nur um die eigene Datenlage zu verbessern. Der Prüfvorlauf hat dieses Problem nicht, denn er prüft und würfelt nicht.

12. Das Handlungsfenster, und warum mehr Prüfung nicht mehr Sicherheit ist

Jede Prüfung verbraucht Zeit, und die Zeit ist in den meisten Fällen nicht Ihre. Eine Bewerbung ist nach zwei Wochen vergeben, ein Angebot läuft ab, ein Vorgang ist fristgebunden.

Daraus folgt eine Obergrenze, die keine Qualitätsfrage ist. Jede zusätzliche Prüfung verbraucht einen Teil des Fensters. Ist das Fenster kürzer, als die Kette braucht, geht sie nicht mehr. Das Fenster begrenzt nicht die Güte Ihrer Kette, sondern ihre Länge.

Zwei Beobachtungen aus der Modellrechnung sind dabei nicht offensichtlich:

Erstens wirkt das Fenster zuerst auf die Mehrfachprüfung und erst dann auf die Frage, wer prüft. Wird die Frist knapper, fällt zunächst die Mehrfachprüfung weg, danach ändert sich, ob Mensch oder Verfahren die Stufe übernimmt.

Zweitens tritt die menschliche Besetzung gerade dann in die Kette, wenn die Zeit knapp wird. Das klingt verkehrt, weil die menschliche Stelle die langsamere ist. Der Grund liegt in der Reihenfolge. Eine scharf stoppende Stelle vorn senkt die Zahl der Fälle, die die aufwendige Stelle dahinter überhaupt erreichen, und kauft damit Zeit.

Für die Praxis folgt daraus eine Prüffrage an jedes Vorhaben. Wie lang ist das Fenster, und wie viel davon verbraucht die Kette heute? Wer diese Zahl nicht hat, kann über Mehrfachprüfung nicht entscheiden.

13. Was tun, wenn die Stelle schon läuft

Die drei Festlegungen aus Kapitel 6 gehören vor den Start. Die meisten Leser haben diesen Punkt überschritten, und die Frage lautet dann, was von außen noch geht.

Die gute Nachricht zuerst. Zwei der drei Festlegungen sind nachholbar, und die dritte ist es teilweise.

Der Rückflusskanal ist jederzeit nachrüstbar. Er verlangt keine Software und keine Datenhistorie. Eine Rückfragemöglichkeit mit ausgewertetem Ergebnis lässt sich in der nächsten Woche einrichten, und sie wirkt von da an. Was vorher geschehen ist, erfahren Sie dadurch nicht, aber das Weiterlaufen hört auf.

Der Prüfvorlauf ist jederzeit nachholbar, unter anderem Namen. Eine einmalige Stichprobe aus den Ablehnungen der letzten Monate, unabhängig zweitbeurteilt, liefert dieselbe Art von Beobachtung wie ein Vorlauf am Anfang. Sie ist sogar leichter zu organisieren, weil die Fälle schon vorliegen. Über die Größe der Stichprobe entscheidet die gewünschte Genauigkeit, siehe Kapitel 8.

Die Fehlerrichtung ist nur teilweise nachholbar. Die Verschaltung lässt sich ändern, die bereits getroffenen Entscheidungen nicht. Was Sie in den vergangenen zwei Jahren abgelehnt haben, bleibt abgelehnt, und die betroffenen Fälle kommen nicht zurück.

Eine Reihenfolge, die sich bewährt hat.

Schritt	Aufwand	Was er bringt
1. Rückfragemöglichkeit einrichten und auswerten	gering	Ab sofort entstehen Beobachtungen, und das Problem wächst nicht weiter
2. Stichprobe aus den Ablehnungen zweitbeurteilen	einmalig, überschaubar	Die fehlende Zahl, erstmals
3. Fallmischung aufschreiben und die Schwelle neu rechnen	ein halber Tag	Die Entscheidung, ob künftig geprüft wird
4. Verschaltung an die gewählte Fehlerrichtung anpassen	Projektarbeit	Die Gestaltung folgt der Entscheidung und nicht umgekehrt

Was an Schritt 2 erfahrungsgemäß scheitert, ist nicht der Aufwand. Es ist die Frage, wer die Zweitbeurteilung vornimmt. Sie darf nicht von derselben Stelle kommen, die ursprünglich entschieden hat, und sie darf das maschinelle Urteil nicht kennen. Sonst messen Sie die Übereinstimmung zweier Urteile und nicht die Güte des ersten.
Nichts an dieser Reihenfolge verlangt ein neues System. Der teuerste Posten ist die Zweitbeurteilung einer Stichprobe, und sie fällt einmal an.

14. Woher die Zahlen kommen

Dieses Whitepaper führt drei Gattungen von Zahlen, die streng zu unterscheiden sind.

Gattung	Beispiel	Was sie trägt
Amtliche Zahlen und Pflichtberichte	die Quoten in Kapitel 10	Sie sind am Primärdokument geprüft und im Anhang nachgewiesen
Modellrechnungen	der Faktor 46, der Prüfvorlauf von 75 Fällen	Sie arbeiten mit offengelegten Beispielwerten und ersetzen keine betriebsindividuelle Rechnung
Was fehlt	der Anteil der zu Unrecht Abgelehnten in den vier Feldern aus Kapitel 5	Es gibt ihn nicht, und der Grund ist derselbe, aus dem das Problem dort auftritt

Die dritte Zeile ist keine Verlegenheit, sondern ein Ergebnis. Die drei Größen, an denen alles hängt, nämlich der Anteil der zu Unrecht Abgelehnten, Ihre eigene Einschätzung dazu und die Dichte des Rückflusses, sind genau dort belegbar, wo das Problem nicht auftritt, und dort nicht belegbar, wo es auftritt. Belegbar sind sie, wo ein Verfahren abgelehnte Fälle gegen den Willen der ablehnenden Stelle wieder aufgehen lässt. Genau dieses Verfahren erledigt das Problem. Eine vierte Größe, nämlich wie fest Ihre Einschätzung sitzt, ist nirgends belegbar, denn sie ist eine Eigenschaft einer Meinung und keine der Welt.
Eine Konsequenz für Ihre eigene Berichterstattung ergibt sich daraus unmittelbar. Eine Kennzahl zur Güte Ihrer Übergabestelle, die allein auf den bearbeiteten Fällen beruht, misst eine Hälfte und schweigt über die andere. Sie ist nicht falsch, sie ist unvollständig, und die fehlende Hälfte ist die teure.
Ein Fremdbefund zum Schluss dieses Kapitels. Eine vorab angemeldete Zusammenschau von 370 Einzelergebnissen aus 106 Versuchen hat untersucht, wann die Verbindung von Mensch und Verfahren besser ist

als beide allein. Im Mittel ist sie es nicht, und bei Aufgaben, die eine Entscheidung zwischen Möglichkeiten verlangen, liegt sie im Minus. Eine vorab festgelegte Arbeitsteilung zwischen Mensch und Verfahren kam in dieser Zusammenschau in vier von 370 Einzelergebnissen vor, also in etwa einem Prozent.
Das ist keine Empfehlung, denn vier Messungen tragen keine. Es ist eine Angabe darüber, wie wenig untersucht ist, was hier vorgeschlagen wird.

15. Was im Bericht an die Geschäftsführung stehen muss

Die meisten Berichte über eine maschinell gestützte Stelle enthalten eine Trefferquote, eine Durchlaufzeit und eine Mengenangabe. Nach allem Vorstehenden fehlt darin die teure Hälfte.
Vier Angaben gehören hinein, und drei davon haben Sie bereits.
Erstens, die beiden Fehlerrichtungen getrennt. Statt einer Trefferquote zwei Zeilen: wie viele der bearbeiteten Fälle sich als Fehlgriff erwiesen haben, und wie viele der abgelehnten nachträglich doch bearbeitet wurden. Die zweite Zeile ist in den meisten Berichten leer. Dass sie leer ist, ist die Aussage, und sie gehört sichtbar in den Bericht und nicht weggelassen.
Zweitens, die unterstellte Fallmischung und wann sie zuletzt geprüft wurde. Eine Zahl, ein Datum, eine Begründung. Ändert sich der Zugang, ändert sich die Schwelle, und das fällt sonst niemandem auf.
Drittens, die Zahl der nachgeprüften Ablehnungen und ihr Ergebnis. Wenn Sie einen Prüfvorlauf gefahren haben, steht hier, was er ergeben hat. Wenn nicht, steht hier eine Null, und auch das ist eine Information.
Viertens, der Rückflusskanal und seine Inanspruchnahme. Wie viele Abgelehnte haben nachgefragt, und wie oft wurde die Entscheidung daraufhin geändert. Diese beiden Zahlen werden in den Feldern mit Beschwerdeweg veröffentlicht, und Sie können sie selbst erheben.

Angabe	Woher sie kommt	Was sie zeigt
Fehlgriffe unter den Bearbeiteten	Ihre bestehenden Daten	die eine Fehlerrichtung
nachträglich doch bearbeitete Ablehnungen	nur aus Nachprüfung oder Rückkanal	die andere Fehlerrichtung
unterstellte Fallmischung, mit Datum	Festlegung, siehe Kapitel 8	ob die Schwelle noch stimmt
Nachfragen und daraufhin geänderte Entscheidungen	der Rückkanal	ob Sie überhaupt etwas erfahren

Ein Bericht, der nur die bearbeiteten Fälle auswertet, misst eine Hälfte und schweigt über die andere. Er ist nicht falsch. Er ist unvollständig, und die fehlende Hälfte ist die, über die Sie sich sorgen sollten.

16. Die Einwände, die kommen

„Unsere Trefferquote liegt bei 94 Prozent." Das kann stimmen und beantwortet die Frage nicht. Gefragt ist, wie sich die verbleibenden 6 Prozent auf die beiden Richtungen verteilen und welche der beiden in Ihrem Fall teurer ist. Diese Aufteilung steht in den wenigsten Berichten, und sie hängt zusätzlich an Ihrer Fallmischung.
„Wir prüfen ohnehin jede Ablehnung." Dann haben Sie das Problem nicht, und Sie haben Kosten, die Sie prüfen sollten. Die Frage ist nicht, ob geprüft wird, sondern ob die Prüfung dem Fall oder der Information dient. Für die Information genügt der Vorlauf aus Kapitel 11.

„Dafür haben wir keine Daten." Das ist der Befund dieses Whitepapers und nicht ein Einwand dagegen. Die Daten entstehen erst, wenn der Rückkanal besteht. Wer wartet, bis sie da sind, wartet dauerhaft.
„Das klingt nach einem Nischenproblem." Es betrifft jede Stelle, an der eine Ablehnung praktisch endgültig bleibt, weil niemand sie anfechtet. Für Ihr Unternehmen heißt das: jede Bewerbungsabsage, jede abgelehnte Finanzierung, jede nicht berücksichtigte Angebotsanfrage.
„Wir wollen keine Zahl, die später gegen uns spricht." Verständlich und teuer. Eine Fehlentscheidung, die niemand erhebt, tritt trotzdem ein. Sie erscheint nur später und in einem schlechteren Moment, meist dann, wenn jemand von außen danach fragt.
„Das klingt nach Theorie." Die Begriffe stammen aus einem Working Paper mit Modell, Rechenkern und Belegapparat. Für die Anwendung im Betrieb genügen die acht Zeilen im nächsten Kapitel.

17. Die Checkliste vor dem Einsatz

Acht Punkte, unabhängig vom Anbieter und vom Werkzeug.
Die beiden Fehlerrichtungen der Stelle sind getrennt benannt, nicht als eine Trefferquote.
Für beide Richtungen ist festgehalten, was ein Fehler kostet, einmal als entgangener Ertrag und einmal als eingetretener Schaden.
Die Mischung der eigenen Fälle ist geschätzt, mit Vermerk, worauf die Schätzung beruht.
Aus beidem ist abgeleitet, in welche Richtung die Kette irren soll, und die Verschaltung folgt dieser Festlegung.
Ein Rückflusskanal ist eingerichtet, den nicht Sie auslösen, und sein Ergebnis wird ausgewertet.
Ein Prüfvorlauf ist festgelegt, in abgelehnten Fällen und nicht in Wochen.
Die Länge des Handlungsfensters ist bekannt, und der Zeitverbrauch der Kette ist dagegen gerechnet.
Im Bericht an die Geschäftsführung steht, welche der beiden Fehlerrichtungen gemessen wird und welche nicht.
Die Technik ist die einfache Hälfte. Über das Ergebnis entscheiden drei Festlegungen vor dem Start und die Frage, ob eine Ablehnung ohne Ihr Zutun wieder aufgehen kann.

Die sieben Begriffe, in einem Satz

Begriff	Bedeutung in diesem Whitepaper
Stelle	Jede Stufe einer Entscheidungskette, ob Verfahren oder Mensch
Übergabestelle	Die Stelle, an der ein Verfahren bewertet und ein Mensch formal entscheidet
Fehlerrichtung	Eine Stelle kann einen guten Fall stoppen oder einen schlechten durchlassen. Beides sind Fehler, und sie kosten Verschiedenes
Fallmischung	Der Anteil der wirklich geeigneten Fälle in dem, was bei Ihnen eingeht. Die Größe, die der Anbieter nicht kennen kann
Rückflusskanal	Ein Weg, auf dem eine Ablehnung ohne Ihr Zutun wieder aufgeht, etwa eine Beschwerde oder ein Widerspruch
Prüfvorlauf	Die ersten Ablehnungen, die nachgeprüft werden, obwohl man sie für richtig hält, um die fehlende Zahl zu gewinnen
Handlungsfenster	Die Zeit, in der ein Fall entschieden sein muss. Sie begrenzt die Zahl der Prüfungen und damit die Länge der Kette

Quellen

Nr.	Quelle
1	Medizinischer Dienst Bund, Meldung vom 22.11.2023, Bezugsjahr 2022. Widerspruchsquote und Korrekturen bei Pflegegutachten. Am Primärdokument geprüft
2	Deutsche Rentenversicherung Bund, Faktencheck Rentenbescheide, Meldung vom 30.04.2025, Bezugsjahr 2024. Am Primärdokument geprüft
3	Bundesagentur für Arbeit, Widersprüche und Klagen im Bürgergeld, Meldung vom 13.01.2026, Bezugsjahr 2025. Am Primärdokument geprüft
4	TikTok, Transparenzbericht nach dem Digital Services Act, Berichtszeitraum Januar bis Juni 2025, Europäische Union. Am Primärdokument geprüft
5	Meta, Transparenzbericht nach dem Digital Services Act für Facebook, Berichtszeitraum Oktober bis Dezember 2024, Europäische Union. Am Primärdokument geprüft
6	Vaccaro, Almaatouq und Malone (2024), *When Combinations of Humans and AI Are Useful*, Nature Human Behaviour 8 (12), S. 2293 bis 2303
7	PPI Working Paper 09/2026, *Die Delegationsgrenze. Maschinelle Bewertung, reale Autorität und die Architektur betrieblicher Entscheidungsketten*. Quelle des Modells, der beiden Prüfschwellen, des Prüfvorlaufs und der Feldrechnungen. MPRA Paper Nr. 131392, https://mpra.ub.uni-muenchen.de/131392/

Zu den Quoten in Kapitel 10. Die Anteile der angefochtenen Ablehnungen beziehen sich auf alle Bescheide beziehungsweise auf alle Entfernungen. Die Erfolgsquoten beziehen sich auf die Anfechtungen. Beide Berichte der Anbieter weisen darauf hin, dass sich Beschwerden auf Maßnahmen eines früheren Zeitraums beziehen können; die Quotienten sind deshalb Größenordnungen und keine Raten. In den Primärquellen der amtlichen Zahlen stehen vier Unstimmigkeiten, die im Working Paper einzeln aufgeführt sind. Gerechnet ist durchweg mit den vorsichtigeren Werten.
Zu den Modellrechnungen. Der Faktor 46 in Kapitel 3, die Rechnungen in den Kapiteln 7 und 9 und der Prüfvorlauf von 75 abgelehnten Fällen in Kapitel 11 stammen aus dem Modell des Working Papers und arbeiten mit offengelegten Beispielwerten. Sie sind Modellaussagen und keine Messungen. Sie ersetzen keine betriebsindividuelle Rechnung, und die Größenordnung, nicht die Ziffer, ist die Aussage.

Ralf Platvoet · Diplom-Ökonom · Platvoet Performance Intelligence info@platvoet.org · platvoet.org