

WakeUP!

HUMAN BIAss

DAS MODELL GEGEN
DEN BIAS-KREISLAUF

DIE KI ALS SPIEGEL

WARUM SIE UNS
UNSERE MUSTER ZEIGT

12 INTERVENTIONEN

WAS DU AB MONTAG
TUN KANNST

DIE JAGD IN 12 WOCHEN

SO ZIEHT DAS RUDEL
MIT KI LOS

DIE POPPER-FRAGE

WANN GIBST DU ZU,
DASS DU FALSCH
LIEGST?

SPEERE WÄRMEN NICHT

DAS RUDEL SCHON.

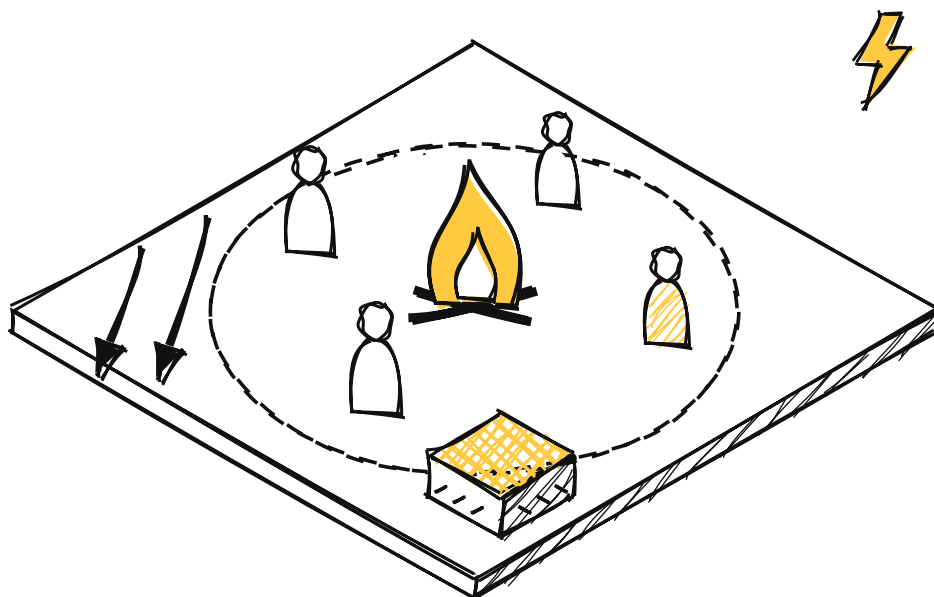
COVERSTORY

JÜRGEN MENHART

DER TRANSFORMANIST ÜBER DEN BIAS-KREISLAUF
IN MENSCH, ORGANISATION UND KI

WakeUP! Ausgabe 01.

Herbst 2026 · Das Magazin des Transformanisten · erscheint zweimal im Jahr



Menschen, die sind. Organisation und Technologie, die wirken.

Impressum gemäß § 24 Mediengesetz. **Medieninhaber, Herausgeber und Hersteller:** Jürgen Menhart, Weblingerstraße 75B, 8054 Graz, Österreich. **Verlags- und Herstellungsort:** Graz.

Offenlegung gemäß § 25 Mediengesetz: Medieninhaber ist Jürgen Menhart, Einzelunternehmen mit Sitz in Graz. Unternehmensgegenstand ist die Unternehmensberatung. Das WakeUP! Magazin erscheint zweimal im Jahr. Die grundlegende Richtung dieses Mediums ist Information und Weiterbildung zu Mensch, KI und Organisation in Unternehmen.

Erstellung: Teile dieses Dokuments wurden mit Unterstützung künstlicher Intelligenz erstellt. Inhalt, Auswahl und Freigabe verantwortet der Medieninhaber.

Bildnachweis: © 2026 Jürgen Menhart. Porträtfoto: Thomas Sattler, Bad Aussee, Steiermark. Veranstaltungsfotos IFM: Bayern Innovativ GmbH. Veranstaltungsfotos ILS365: Kevin Legenstein.

In dieser Ausgabe.

AUFTAKT

Vorwort	4
Mammuts warten nicht. Das Rudel entscheidet.	
Das Wichtigste auf einer Seite	5
Fünf Kernsätze · T.R.U.S.T. Hero Check	
Auf der Bühne	6
IFM Memmingen und ILS365 Graz · Keynotes	
Stimmen	7
Was bleibt, wenn der Applaus vorbei ist	

COVERSTORY · HUMAN BIASS

Gewissheit ist teurer als Unwissen	8
Der persönliche Einstieg	
HB Kernthese und Modell	10
Der Bias-Kreislauf und die vier Züge	
12 Die Jagd in zwölf Wochen	13
Wie das Rudel mit KI loszieht · dein 20-Minuten-Talk	

TIEFGANG

I Unser Gehirn ist auf Tempo gebaut	15
Heuristiken, System 1 und 2, begrenzte Rationalität	
II Der Teufelskreis aus Default, Filter und Wahrheit	19
Voreinstellung, Filter, Versteinerung	
III Organisationen sind Verstärker	26
Groupthink, HiPPO, Sunk Cost, Lagerfeuer	
IV Die Maschine lernt unser Gestern	32
Daten-Bias, Feedback Loops, Sprachmodelle	
V Der Trigger und die Struktur	37
Hybrid Intelligence, Prompts, Strukturen	
VI Mensch, Organisation und KI: Wer prüft wen?	46
Selbstreferenz, kalibriertes Vertrauen, EU AI Act	

PRAXIS

VII Das Lagerfeuer zuerst	49
Zwölf Interventionen, Praxisfall aus dem Mittelstand	

SERVICE

+ Anhang: Vertiefungen	55
Bias-Landkarte, Fallstudien, Fairness-Werkzeuge, Groupthink	
Literatur	59
Über den Autor	62
Keynotes und Bücher	

Drei Lesewege durch diese Ausgabe. In fünf Minuten liest du das Vorwort und das Wichtigste auf einer Seite. In zwanzig Minuten kommt die Coverstory mit dem Modell Human BIAss und der Jagd in zwölf Wochen dazu. Wer tiefer gehen will, findet im Tiefgang die Forschung hinter jedem Schritt und in der Praxis zwölf Interventionen für Montag. Das WakeUP! Magazin erscheint zweimal im Jahr, Ausgabe 02 im Frühjahr 2027.

Mammuts warten nicht. Das Rudel entscheidet.

In 680+ Projekten und Transformationen habe ich eines gelernt: Kein Unternehmen scheitert daran, dass ihm das Werkzeug fehlt. Es scheitert, wenn das Rudel nicht mehr gemeinsam jagt.

Die Mammutjagd war nie ein Spaziergang. Das Gelände hat sich aber verändert. Lange haben wir unsere Welt mit VUCA beschrieben, also als volatil, unsicher, komplex und mehrdeutig. Diese Beschreibung reicht heute nicht mehr aus. Der Zukunftsforscher Jamais Cascio spricht von einer BANI-Welt, die brüchig, beängstigend, nichtlinear und unbegreiflich ist. Brüchig bedeutet, dass Systeme, die stabil wirkten, über Nacht reißen können. Beängstigend bedeutet, dass diese Unsicherheit bei den Menschen ankommt.

Ich sehe das jeden Tag. Branchen, die über Jahrzehnte den Takt vorgegeben haben, verlieren den Boden unter den Füßen. Andere gehen durch die Decke, weil sie früher hingeschaut und schneller gehandelt haben. Dazwischen stehen Führungskräfte, die spüren, dass die Technologie sie überholt. Jede Woche erscheint ein neues Modell mit einem neuen Versprechen. Viele kaufen deshalb Speere, weil sich das Kaufen wie Handeln anfühlt.

Speere wärmen nicht. Ein verunsichertes Rudel schärft seine Speere und bleibt trotzdem in der Höhle. Es hält an dem fest, was es kennt, weil das Bekannte Sicherheit verspricht. Genau dort beginnt der Kreislauf, den dieses Heft beschreibt. Eine Annahme wird bestätigt, danach wiederholt und gilt irgendwann als Wahrheit. Aus dem Satz „Das geht bei uns nicht“ wird eine Höhlenwand, an der jede neue Idee abprallt.

Die Mammutjagd gelingt deshalb nicht mit dem schärfsten Speer, sondern am Lagerfeuer. Dort sitzt das Rudel zusammen, bevor es losgeht. Dort darf jemand sagen, dass die Spur in die falsche Richtung führt, ohne dass es ihn etwas kostet. Dort wird aus Verunsicherung wieder Mut, weil niemand allein vor dem Mammut steht.

Die KI ist in diesem Bild kein Gegner. Sie ist ein guter Speer, wenn das Rudel weiß, wohin es jagt. Zugleich ist sie ein Spiegel, der uns zeigt, welche Muster wir ihr beigebracht haben. Wer in diesen Spiegel schaut, erkennt seine eigenen Abkürzungen im Denken.

Dieses Heft ist meine Einladung, die eigenen Gewissheiten am Feuer auf den Tisch zu legen. Du findest darin die Mechanik des Bias-Kreislaufs, mein Modell Human BIAss und zwölf Interventionen, mit denen du am Montag beginnen kannst. Nimm dir das heraus, was dein Rudel jetzt braucht.

„Am Anfang benötigt es Begeisterung und Emotion, für Erfolg, Fokus und Disziplin!“

Jürgen Menhart · Der Transformant · Graz, im Oktober 2026

Das Wichtigste auf einer Seite.

- 1. Bias ist ein Kreislauf.** Eine Annahme wird bestätigt, danach wiederholt und gilt am Ende als Wahrheit, die zur neuen Voreinstellung wird. Aus sich heraus ist dieser Kreislauf gegen Fakten immun.
- 2. Organisationen verstärken ihn.** Groupthink, der HiPPO-Effekt und versunkene Kosten machen aus individuellen Denkabkürzungen strategische Fehler, an denen intern niemand zweifelt.
- 3. KI skaliert ihn.** Modelle lernen unsere Entscheidungen von gestern, Sprachmodelle neigen zur Zustimmung, während wir ihren Vorschlägen folgen. Auf diese Weise entstehen die Daten von morgen.
- 4. Der Kreislauf bricht durch Trigger und Struktur.** Ein Trigger allein reicht nicht aus. Es braucht zusätzlich einen Ort, an dem Widerspruch nichts kostet, also das Lagerfeuer der psychologischen Sicherheit.
- 5. Human BIAss ist die Antwort.** Das Modell verlangt Menschen, die in der Sache unbequem und im Ton wertschätzend sind. Es besteht aus vier Zügen: hinterfragen, Bias unterbrechen, neue Wege gehen sowie mit KI in einem klaren Rahmen ins Go kommen. Zwölf Interventionen in Teil VII machen das ab Montag umsetzbar.



T.R.U.S.T. HERO CHECK

Wie viel Lagerfeuer steckt in deinem Rudel? Der Check umfasst 28 Fragen, dauert fünf Minuten und ist kostenlos. Er ist kein Test, sondern ein Spiegel.

juergen-menhart.com/trust-hero-check/

Das Rudel am Lagerfeuer.



Oben und Mitte rechts: Keynote beim Internationalen Forum Mechatronik in Memmingen, September 2026, mit dem Lagerfeuer als Bild für das Rudel. **Mitte links und unten rechts:** Auftakt-Impuls beim ILS365 Main Event in Graz, Oktober 2026. **Unten links:** Hand aufs Herz in Memmingen. Fotos: Bayern Innovativ GmbH (IFM), Kevin Legenstein (ILS365).

KEYNOTE FÜR DEIN EVENT

» Speere wärmen nicht. Das Rudel schon.

Drei Keynotes, ein Gedanke: Menschen, Technologie und Organisation wirken nur gemeinsam. Anfragen unter juergen-menhart.com/#contact.

Was bleibt, wenn der Applaus vorbei ist.

ILS365 MAIN EVENT, GRAZ

„Ein 18 Minuten Impuls über genau das, was kein Roboter kann: ein Rudel zusammenhalten.“

„Daten, Roboter und KI sind die schärfsten Speere unserer Zeit. Technologie gibt uns dabei unglaubliche Möglichkeiten, aber nachhaltiger Mehrwert und Wärme entstehen dort, wo Menschen sich verbinden, Vertrauen schaffen und Technologie gemeinsam sinnvoll einsetzen.“

Kajetan Bergles, Managing Partner und ILS Projectlead

„Für mich ist Jürgen wirklich jemand der ganz klar nach vorne blickt, immer den Menschen im Fokus hat und versucht alle zu bündeln und – wie er so schön sagt – am Lagerfeuer zu vereinen.“

Christian Klimaschka, VP Technology, Gustav Hensel GmbH & Co. KG

„Jürgen has this rare combination of deep technical expertise in technology/product development and an empathetic leadership style.“

Charles Kammerer, Group CFO, Schneider Transport

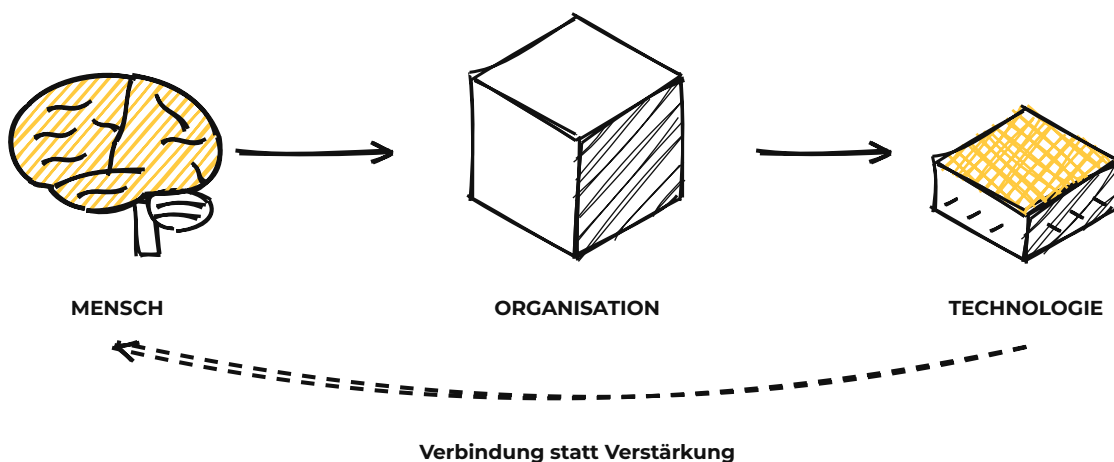


Internationales Forum Mechatronik, Memmingen, September 2026. Foto: Bayern Innovativ GmbH

Gewissheit ist teurer als Unwissen.

Transformationen scheitern selten an der Technik, sondern an Sätzen, die niemand mehr hinterfragt, etwa „Das haben wir immer so gemacht“, „Die Zahl steht seit dem Kick-off“ oder „Der Vorstand will das so“.

Jeder dieser Sätze war einmal eine Annahme, die erst wiederholt und dann bestätigt wurde. Am Ende galt er als Wahrheit, weil niemand mehr nachgefragt hat. Diesen Kreislauf beschreibt die erste Ausgabe des WakeUP! Magazins. Er beginnt in unserem Gehirn, das auf Tempo gebaut ist und nicht auf Vollständigkeit. In Organisationen wird er verstärkt, weil Hierarchie, Konsens und versunkene Investitionen den Zweifel teuer machen. Heute läuft er zusätzlich in Maschinen weiter. Wir trainieren KI auf unseren Entscheidungen von gestern. Weil ihre Vorschläge sachlich klingen, folgen wir ihnen, wodurch die Daten von morgen entstehen.



Der systemische Blick verändert dabei die Frage. Wer sich irrt, ist zweitrangig, entscheidend ist, welches System den Irrtum stabil hält. Mensch, Technologie und Organisation sind keine drei getrennten Baustellen, sondern ein gemeinsames System. Wer nur die Technologie verbessert, beschleunigt lediglich das Tempo, mit dem das System seine Fehler wiederholt. Speere wärmen nicht, auch die schärfsten nicht. Wärme entsteht am Lagerfeuer, also an einem Ort, an dem jemand „Ich glaube, wir irren uns“ sagen kann, ohne seinen Platz im Rudel zu riskieren.

Eine gelungene Verbindung entsteht, wenn Menschen ihre eigenen Denkabkürzungen kennen. Organisationen bauen dafür Widerspruch ein, statt ihn zu bestrafen. Technologie arbeitet dabei als Gegenspieler, statt als Ja-Sager aufzutreten. Diese Reihenfolge entspricht den drei Schritten aus WakeUP!: People fit, Project fit, Strategy fit.

WAKEUP!-MOMENT

„Wer bist du, wenn du nicht funktionieren musst?“ Sein statt Funktion.

Das ist die Frage aus der Keynote „Speere wärmen nicht. Das Rudel schon.“ mit ihrer Antwort. Sie trifft den Kern dieser Ausgabe, denn Bias schützt unsere Funktion. Wer ihn unterbrechen will, braucht den Mut, für einen Moment nicht zu funktionieren.



Speere wärmen nicht. Das Rudel schon. Auftakt-Impuls beim ILS365 Main Event in der Helmut-List-Halle Graz, Oktober 2026.
Foto: Kevin Legenstein

Jürgen Menhart · Graz, im Oktober 2026

Ein Kreislauf auf drei Ebenen erzeugt **scheinbare Wahrheit.**

Denkfehler sind keine Ausrutscher einzelner Menschen. Sie greifen ineinander und bilden einen geschlossenen Kreislauf, der sich selbst verstärkt, im Kopf ebenso wie im Team und in der Maschine. Mit jeder Ebene wird er schneller und schwerer zu erkennen.

PHASE	MECHANISMEN	WAS IM KREISLAUF PASSIERT
1 • Voreinstellung	Default, Status quo, Anker, Framing	Was zuerst da ist, wird zum Maßstab für alles, was danach kommt.
2 • Filter	Confirmation Bias, Motivated Reasoning, Dissonanzvermeidung	Was passt, kommt durch, während Störendes kleingeredet wird.
3 • Versteinerung	Wiederholung, Vertrautheit, sich selbst erfüllende Erwartungen	Das Gefilterte fühlt sich wie Wissen an, wird zu Stein und bildet den nächsten Default.

EBENE	VERSTÄRKER	TYPISCHES ERGEBNIS
Mensch	Begrenzte Aufmerksamkeit, Bedürfnis nach einer stimmigen Geschichte	Eine Überzeugung, die sich wie Wissen anfühlt
Organisation	Konsensdruck, Hierarchie, versunkene Kosten, Prozesse	Eine Strategie, die niemand mehr hinterfragt
KI	Trainingsdaten, Rückkopplung, Skalierung, Vertrauen in Maschinen	Ein Status quo, der objektiv aussieht

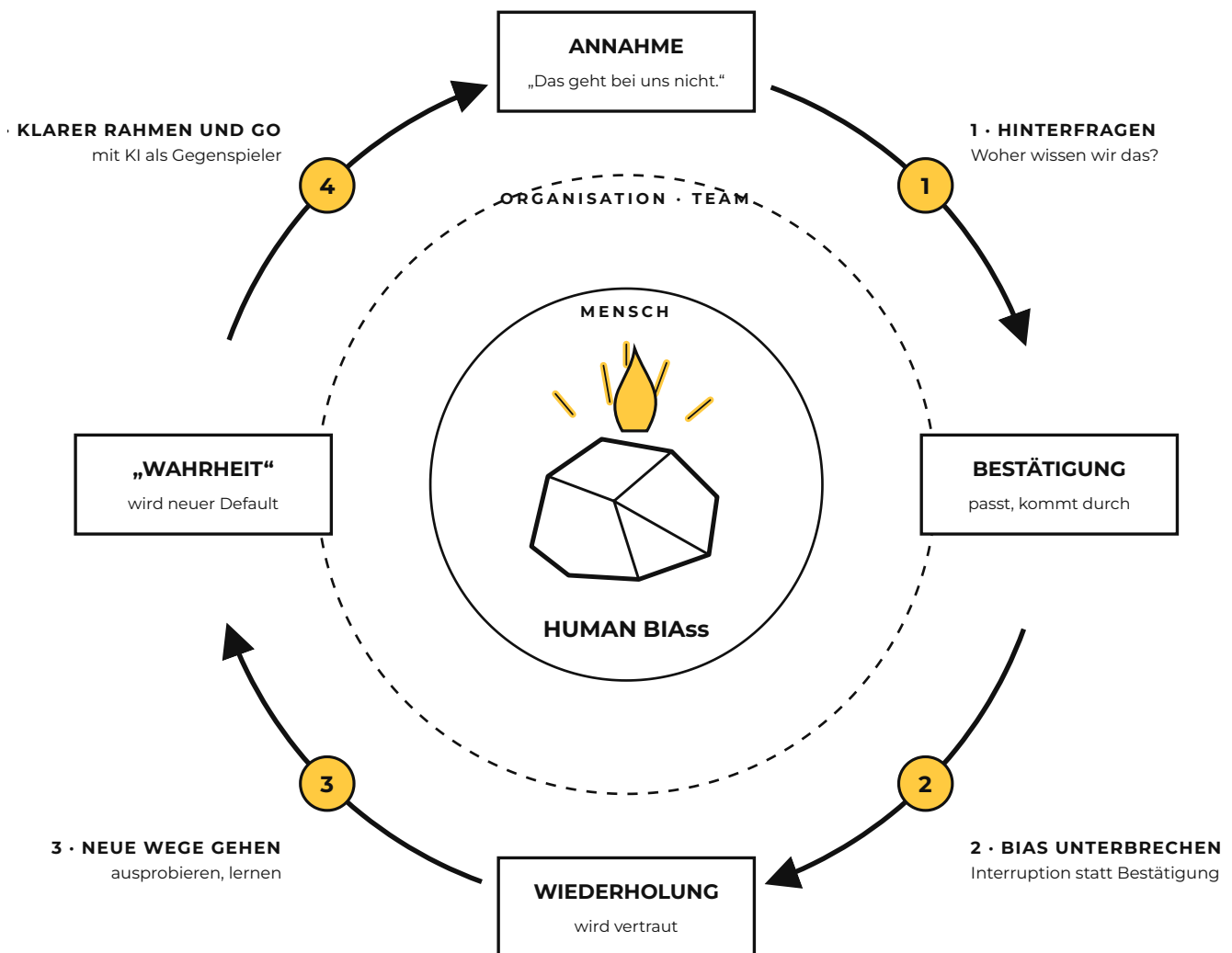
WAKEUP!-MOMENT

Aus sich heraus ist der Kreislauf gegen Fakten immun. Er braucht einen Trigger sowie eine Struktur, die diesen Trigger hörbar macht.

HB Das Modell: Human BIAss

Das Modell geht auf eine Keynote aus dem Jahr 2024 zurück, in der der „Circle of BIAss“ entstanden ist. Die Geschichte beginnt in der Höhle beim Rudel, das gemeinsam auf Mammutjagd geht. Seine Mitglieder haben überlebt, weil sie sich gemerkt haben, was gefährlich ist. Auf dem Flipchart stand dazu der Satz „Das geht bei uns nicht“, um den ein Kreis gezogen wurde.

Dieser Kreis beschreibt den Bias-Kreislauf: Eine Annahme wird bestätigt, danach wiederholt und gilt irgendwann als Wahrheit, die uns schützt, wie es früher die Höhle tat. Der Kreislauf wirkt im einzelnen Menschen, in Teams und Unternehmen sowie zunehmend in der KI, die ihn von uns gelernt hat und ihn uns nun wie einen Spiegel vorhält.



Innen steht der Mensch, um ihn liegen Organisation und Team, außen läuft der Kreislauf bis in die KI.

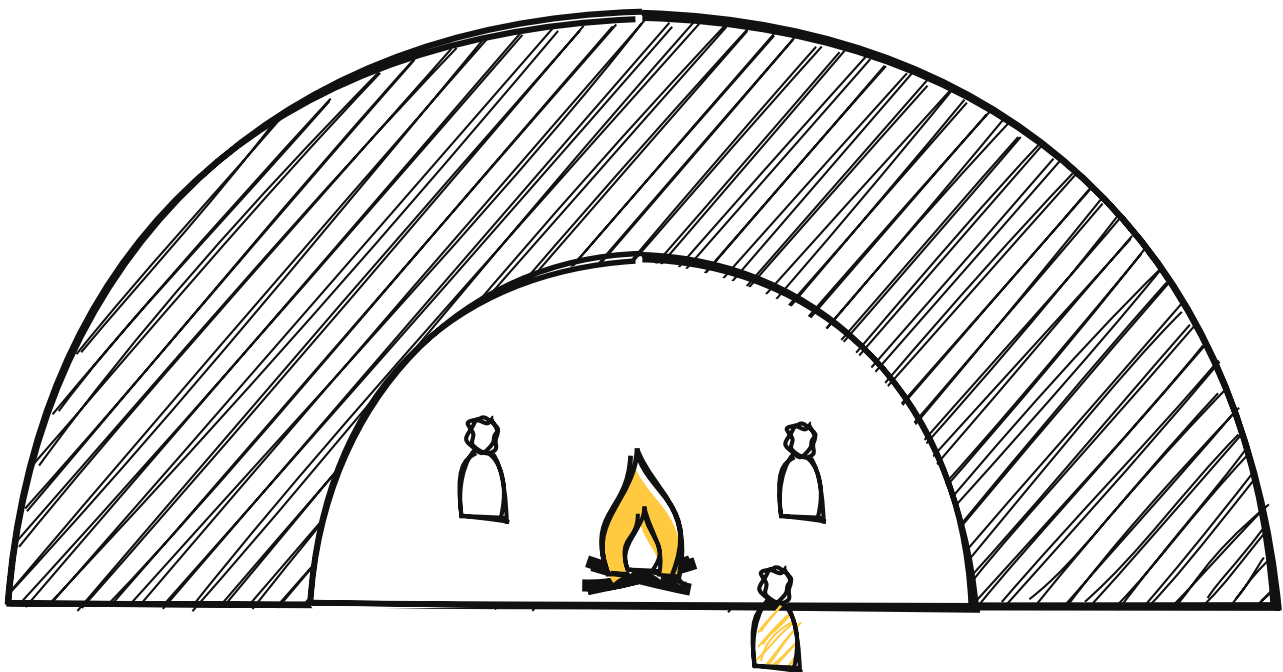
Human BIAss · Der Bias-Kreislauf und die vier Züge, die ihn unterbrechen. Jeder Zug setzt an einem Übergang an. In der Mitte schlägt der Feuerstein Funken: Aus dem Stein, zu dem eine Annahme geworden ist, entsteht durch Reibung das Feuer für Neues.

Human BIAss ist eine bewusste Überreizung. Im Wort steckt Bias, im Englischen steckt darin aber auch „ass“, also der Mensch, der nervt, weil er im Meeting die Frage stellt, die alle anderen vermeiden. Der Kreislauf bricht nicht durch freundliche Zustimmung, sondern durch jemanden, der bereit ist, für einen Moment unbequem zu sein, unbequem in der Sache und wertschätzend im Ton. Das Bild dafür ist der Feuerstein, der nur dann Funken schlägt, wenn er auf Widerstand trifft.

Wer in der Transformation diese Rolle übernehmen will, spielt vier Züge. Beim Hinterfragen lautet die Frage: Woher wissen wir das eigentlich? Beim Unterbrechen setzen wir bewusst dort an, wo sonst Bestätigung folgen würde. Neue Wege zu gehen bedeutet, auszuprobieren, sich auszutauschen, zu lernen und auch zu scheitern. Der vierte Zug heißt „Klarer Rahmen und Go“, weil wir KI in einem klaren Rahmen als Gegenspieler nutzen und dann ins Handeln kommen. Teil VII übersetzt diese vier Züge in zwölf konkrete Interventionen.

Systemisch gelesen ist Bias damit eine Eigenschaft des Systems und nicht nur der Menschen darin. Deshalb folgt diese Ausgabe der Logik von WakeUP! mit drei Schritten in fester Reihenfolge. People fit steht für Menschen, die Widerspruch wagen. Project fit steht für Strukturen, die Widerspruch einfordern. Strategy fit steht für Technologie, die Widerspruch liefert, statt Bestätigung zu verstärken. Wer mit dem dritten Schritt beginnt, kauft Speere und wundert sich, dass niemand warm wird.

„Der Kreislauf bricht nicht durch freundliche Zustimmung, sondern durch jemanden, der bereit ist, für einen Moment unbequem zu sein.“



12

EINBLICK · DAS WIE

Die Jagd in zwölf Wochen.

	Vorb.	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12
Phase	Feuer	Ankommen Kickoff	Gemeinsam jagen 2 bis 3 Use Cases je Person							Beute teilen Ergebnisse & Rollout			
Vor Ort am Lagerfeuer		Tag 1 Kickoff			Tag 2				Tag 3				Tag 4 Markt
Workshops remote		Intro-Workshops 2 × 60 min			Follow-up 1 h				Follow-up 1 h				Ergebnis 1 h
Weekly Sync 30 Minuten				Sync		Sync	Sync	Sync		Sync	Sync	Sync	
Rudel 4 h pro Woche		Mit KI arbeiten · im Rudel austauschen · reflektieren Keine Vorkenntnisse nötig, Scheitern ist erlaubt											
Rahmen Führung & IT	Mandat Plattform												

vor Ort am Lagerfeuer remote und im Alltag

Ein aktuelles Programm aus dem Mittelstand zeigt, wie eine Reise mit KI gelingt, bei der die Menschen vorangehen. Eine kleine Gruppe von KI-Champions aus mehreren Abteilungen zieht gemeinsam los, vom Lagerfeuer auf die Mammutjagd.

Wie das Rudel mit KI auf Mammutjagd geht.

Vorkenntnisse braucht dafür niemand, wohl aber die Bereitschaft, Gewohntes zu hinterfragen. Die Menschen gehen voran, die Technologie zieht mit.

01 · VORBEREITUNG

Das Feuer entzünden

Die Geschäftsführung gibt ein klares Mandat und nominiert das Rudel. Der Rahmen für Datenschutz und Plattform steht, bevor die Jagd beginnt.

02 · WOCHE 1 BIS 2

Am Lagerfeuer ankommen

Der Kickoff findet vor Ort statt, mit Impuls und ersten Workshops. Das Rudel lernt das Werkzeug kennen und erlebt, dass Fragen ausdrücklich erwünscht sind.

03 · WOCHE 3 BIS 8

Gemeinsam jagen

Jede Person arbeitet an zwei bis drei Use Cases aus dem eigenen Alltag. Ein kurzer Sync pro Woche hält das Rudel zusammen, Scheitern gehört dazu.

04 · WOCHE 9 BIS 12

Die Beute teilen

Die Ergebnisse werden dokumentiert, bewertet und auf einem Marktplatz vorgestellt. Daraus entsteht die Grundlage für den Rollout in die ganze Organisation.

4 Stunden

pro Woche: mit der KI arbeiten, im Rudel austauschen, reflektieren

4 Tage

vor Ort am Lagerfeuer, vom Kickoff bis zum Marktplatz

30 Minuten

Sync jede Woche, remote und verbindlich

HUMAN BIASS IM PROGRAMM

ZUG	WAS IM PROGRAMM PASSIERT
Hinterfragen	Prozesse kommen auf den Prüfstand, bevor KI sie beschleunigt.
Bias unterbrechen	Der Austausch im Rudel ersetzt die Bestätigung im eigenen Kopf.
Neue Wege gehen	Use Cases werden ausprobiert, verworfen und verbessert.
Klarer Rahmen und Go	Mandat, Datenschutz und Zeitbudget stehen fest, bevor es losgeht.

LUST AUF MEHR?

» Dein 20-Minuten-Talk mit Jürgen

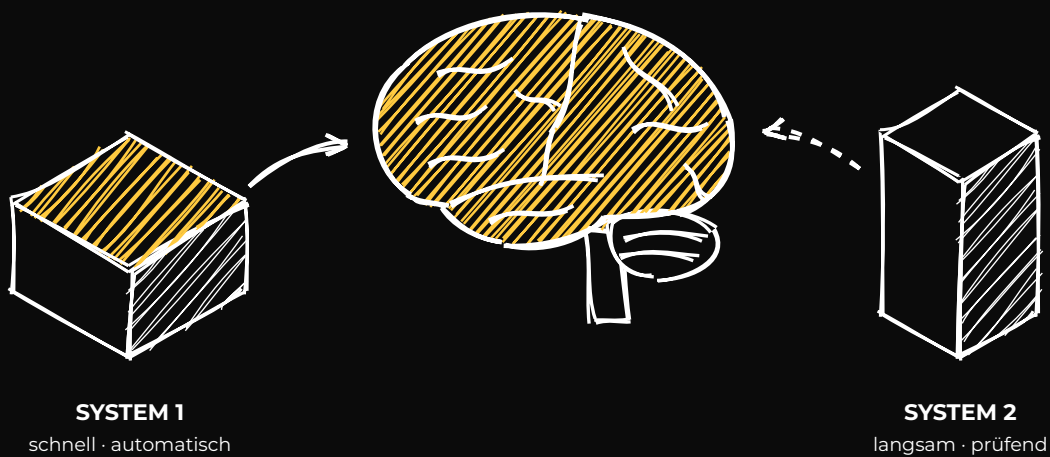
Wir schauen gemeinsam, wie die Jagd in deinem Unternehmen aussehen kann. Unverbindlich, persönlich und in Ruhe.



juergen-menhart.com/#contact

Der Fall ist anonymisiert und beschreibt das Vorgehen eines laufenden Programms.

Unser Gehirn ist auf **Tempo** gebaut und nicht auf Wahrheit.



Bevor wir über Organisationen und Maschinen sprechen, schauen wir auf das Organ, das all unsere Entscheidungen vorbereitet, lange bevor wir glauben, selbst zu entscheiden.

1.1 Schnell schlägt genau

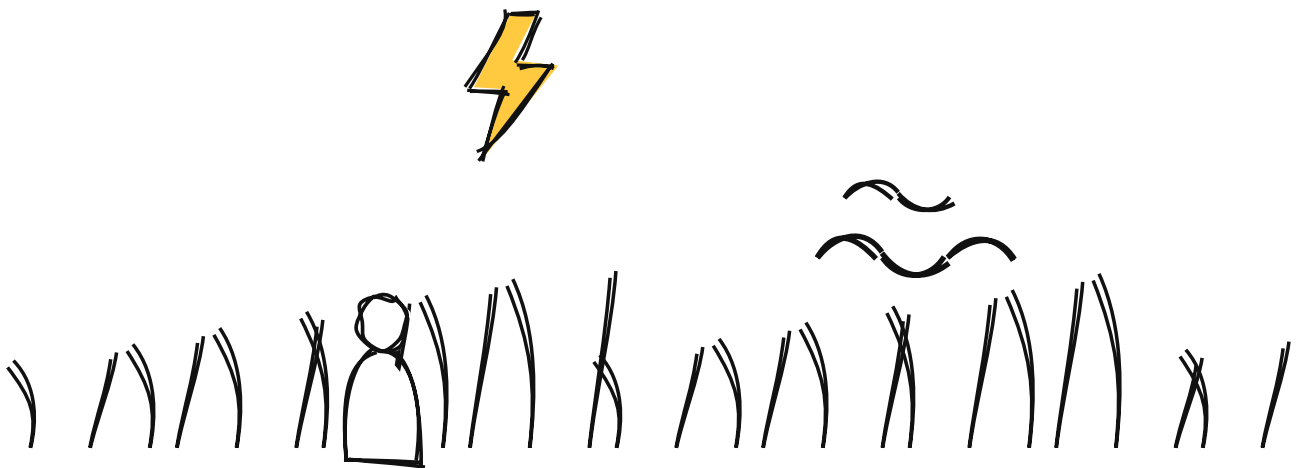
Wenn es im Gras raschelt, bleibt keine Zeit für eine Analyse, denn wer zuerst Daten sammelt, ist verloren. Die Evolution hat deshalb nicht das genaueste Denken belohnt, sondern ein Denken, das zum richtigen Zeitpunkt genau genug ist.

So entstehen Heuristiken, also Faustregeln, die mit wenig Information schnell brauchbare Ergebnisse liefern. Gerd Gigerenzer hat gezeigt, dass solche Regeln in vielen realen Umwelten erstaunlich gut funktionieren und teils sogar besser abschneiden als komplexe Modelle (Gigerenzer & Gaissmaier, 2011). Heuristiken sind damit ein Werkzeug, das zu einer bestimmten Umwelt passt.

Die Error Management Theory liefert eine zweite Logik (Haselton & Nettle, 2006): Wenn zwei mögliche Fehler unterschiedlich teuer sind, lohnt es sich, systematisch in Richtung des billigeren Fehlers zu irren. Wer ein Rascheln für einen Feind hält, verliert wenig, wer dagegen einen Feind für ein Rascheln hält, verliert alles.

Der knappe Faktor ist dabei Zeit und Aufmerksamkeit, kaum aber Energie. Das Gehirn verbraucht zwar rund ein Fünftel unseres Energieumsatzes, dieser Bedarf entsteht jedoch vor allem durch die ständige Grundaktivität. Bei Denkaufgaben steigt die Durchblutung nur lokal, meist um 5 % oder weniger, während sich für das gesamte Gehirn keine Veränderung messen lässt (Raichle & Gusnard, 2002). Auch die populäre Vorstellung, Nachdenken leere die Willenskraft wie einen Akku, hat einer Replikation mit 23 Laboren nicht standgehalten (Hagger et al., 2016).

Problematisch wird es dort, wo sich die Umwelt ändert, die Faustregel aber bleibt. Wir leben heute mit tausenden Signalen pro Tag, mit Feeds, Dashboards, Mails, Chats und Quartalszahlen. Unsere Abkürzungen laufen weiter, führen aber immer öfter an die falschen Orte.



1.2 System 1 und System 2

Daniel Kahneman hat in „Thinking, Fast and Slow“ (2011) Jahrzehnte der Forschung in ein Bild gefasst, das er selbst als Arbeitsmetapher verstand. System 1 arbeitet schnell, automatisch und mühelos, indem es Gesichter erkennt, Sätze vervollständigt und Gefahr spürt. System 2 arbeitet langsam, regelgeleitet und

anstrengend, indem es rechnet, prüft und vergleicht. System 1 liefert ununterbrochen Eindrücke, die System 2 meist ungeprüft übernimmt. Kahneman nennt System 2 deshalb „faul“, weil es nur anspringt, wenn etwas offensichtlich nicht passt oder wenn wir es bewusst einschalten.

WAKEUP!-MOMENT

„What you see is all there is.“

Mit dem Kürzel WYSIATI beschreibt Kahneman den Startpunkt des Kreislaufs. System 1 baut aus dem, was gerade verfügbar ist, die stimmigste Geschichte, ohne zu vermissen, was fehlt. Unsere Sicherheit hängt deshalb an der Stimmigkeit der Geschichte und nicht an der Qualität der Belege.

Das Fundament legten Kahneman und Amos Tversky 1974 mit „Judgment under Uncertainty: Heuristics and Biases“ in Science. Drei Heuristiken tragen bis heute. Bei der Repräsentativität urteilen wir nach Ähnlichkeit statt nach Wahrscheinlichkeit. Bei der Verfügbarkeit urteilen wir danach, wie leicht uns Beispiele einfallen. Bei der Verankerung entfernen wir uns zu wenig von einem Startwert. Diese Heuristiken sind meist nützlich, führen aber zu Fehlern, die systematisch und vorhersagbar sind.

1979 folgte die Prospect Theory in Econometrica. Sie zeigt, dass wir Ergebnisse relativ zu einem Referenzpunkt bewerten, wobei Verluste schwerer wiegen als gleich große Gewinne. Dieser Verlustaversion begegnen wir in dieser Ausgabe immer wieder, beim Status quo ebenso wie beim Framing oder bei versunkenen Kosten. Wer den Referenzpunkt setzt, bestimmt, was als Verlust erlebt wird.

Dazu gehört eine Abgrenzung. Bias ist eine systematische Abweichung in eine Richtung, während Noise die ungewollte Streuung von Urteilen bezeichnet, die eigentlich gleich ausfallen sollten (Kahneman, Sibony & Sunstein, 2021). Algorithmen reduzieren Noise oft deutlich, Bias können sie dagegen einfrieren.

1.3 Gut genug und stimmig

Herbert Simon hat bereits 1955 das Bild des perfekt rationalen Entscheiders verabschiedet. Menschen verfügen über begrenzte Information, begrenzte Rechenkapazität und begrenzte Zeit. Sie optimieren deshalb nicht, sondern suchen eine Lösung, die gut genug ist. Simon nannte das Satisficing.

Hinzu kommt ein Motiv, denn unsere Überzeugungen sollen zusammenpassen. Widersprüche zwischen dem, was wir glauben, was wir tun und was wir sehen, erzeugen Spannung, weil wir Kohärenz herstellen wollen. Das macht uns handlungsfähig, da ein Team, das an seinen Plan glaubt, entschlossener arbeitet als eines, das ihn täglich zerlegt.

Dieselbe Kraft macht uns aber auch blind. Sobald eine Überzeugung Teil unserer Identität geworden ist, erleben wir jede Gegeninformation als Angriff auf unser Selbstbild. Dann arbeitet das Gehirn an der Verteidigung statt an der Wahrheit. Genau hier schließt sich der Ring.

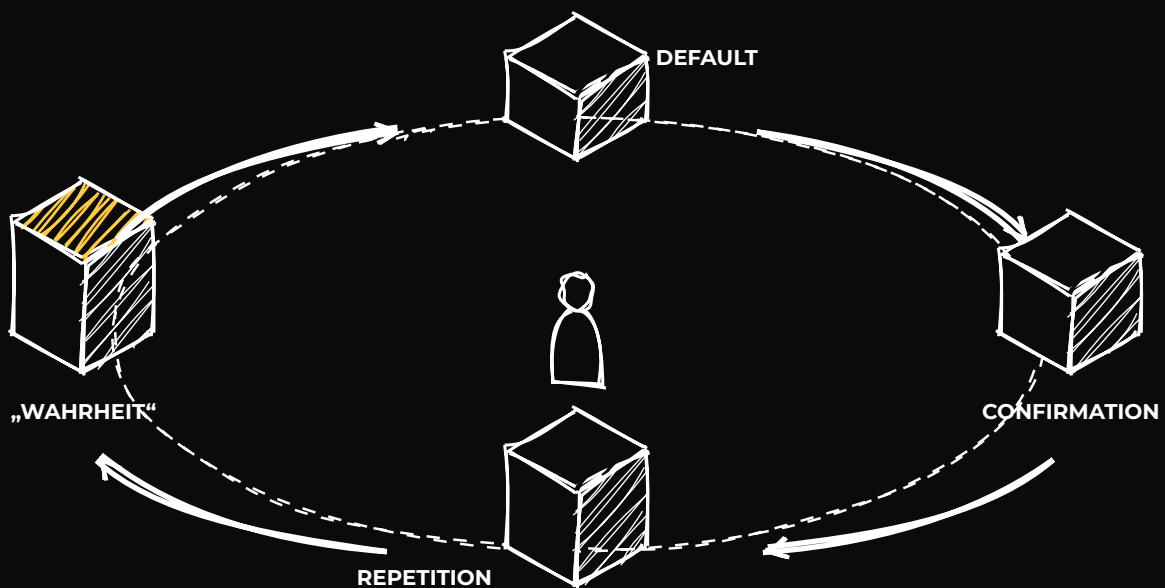
WAKEUP!-MOMENT

Welche Annahme in deinem Bereich ist so selbstverständlich geworden, dass seit über einem Jahr niemand mehr nach ihrer Herkunft gefragt hat?



TEIL II · MENSCHLICHE EBENE

Der Teufelskreis aus **Default, Filter** und **Wahrheit**.



Der Kreislauf besteht aus drei Phasen, von denen keine allein dramatisch ist. Zusammen bauen sie jedoch eine Überzeugung, die sich anfühlt wie Wissen.

2.1 Phase 1: Die Voreinstellung

■ Default und Status quo

Wir bleiben überproportional oft bei der Option, die bereits gesetzt ist. Samuelson und Zeckhauser (1988) haben diesen Status-quo-Bias in Experimenten sowie in echten Versicherungs- und Pensionsentscheidungen nachgewiesen. Drei Kräfte wirken dabei zusammen. Erstens spart die Voreinstellung Aufwand. Zweitens wirkt sie wie eine stille Empfehlung, hinter der sich jemand etwas gedacht haben wird. Drittens macht sie den Status quo zum Referenzpunkt, sodass wir jede Abweichung als möglichen Verlust erleben.

Johnson und Goldstein (2003) zeigen das in Science an der Organspende, denn Länder mit Widerspruchslösung haben dramatisch höhere Zustimmungsraten als Länder mit Zustimmungslösung. Die Menschen unterscheiden sich dabei kaum, die Voreinstellungen dagegen sehr. Thaler und Sunstein (2008) haben daraus das Konzept der Entscheidungsarchitektur entwickelt: Wer den Default gestaltet, gestaltet die Entscheidung.

Was hilft: Bei jeder wiederkehrenden Entscheidung lohnt sich die Frage, wer die Standardoption festgelegt hat, wann das geschah und welches Ziel damals dahinterstand.

■ Der Anker

Die erste Zahl, die wir hören, zieht alle weiteren Schätzungen in ihre Richtung, selbst wenn sie offensichtlich zufällig ist. Tversky und Kahneman (1974) ließen Versuchspersonen ein manipuliertes Glücksrad drehen, das bei 10 oder bei 65 stehen blieb. Danach sollten sie den Anteil afrikanischer Staaten in den Vereinten Nationen schätzen, wobei die Schätzungen nach dem hohen Anker deutlich höher lagen.

Fachwissen schützt davor nicht. Englich, Mussweiler und Strack (2006) zeigten, dass erfahrene Juristinnen und Juristen in einem fiktiven Fall höhere Strafen vorschlugen, wenn sie zuvor eine höhere Zahl gewürfelt hatten. Mussweiler und Strack (1999) erklären das mit selektiver Zugänglichkeit: Wenn wir zuerst prüfen, ob der Anker stimmen könnte, aktivieren wir dabei bevorzugt passende Information.

In Projekten heißen die Anker erste Budgetzahl, erstes Zieldatum oder erste Aufwandsschätzung. In Lenkungsausschüssen gilt eine Zahl aus der ersten Präsentation manchmal noch Jahre später als Maßstab, obwohl niemand mehr weiß, woher sie stammt.

Was hilft: Unabhängige Schätzungen werden eingeholt, bevor eine Zahl im Raum steht. Außerdem lohnt es sich, die Herkunft jeder Zahl zu dokumentieren, die als Maßstab dient.

■ Der Rahmen

Logisch gleiche Information führt zu unterschiedlichen Entscheidungen, je nachdem, wie sie verpackt ist. Im „Asian Disease Problem“ (Tversky & Kahneman, 1981) wählte die Mehrheit die sichere Option, solange von geretteten Leben die Rede war. Bei identischen Zahlen, die als Todesfälle formuliert waren, entschied sich die Mehrheit dagegen für die riskante Option. Im Gewinnrahmen handeln wir vorsichtig, während wir im Verlustrahmen ins Risiko gehen.

Bei KI-Einführungen lässt sich das unmittelbar beobachten. Der Satz „KI ersetzt Tätigkeiten“ aktiviert Verlust, während der Satz „KI nimmt uns Blindleistung ab“ Gewinn aktiviert. Beide Sätze können denselben Sachverhalt beschreiben, sie erzeugen aber völlig unterschiedliche Teams.

Was hilft: Wichtige Vorlagen werden in beiden Rahmen formuliert. Kippt dabei die Präferenz, entscheidet der Rahmen und nicht die Sache.

2.2 Phase 2: Der Filter

Confirmation Bias

Wir suchen, lesen und erinnern bevorzugt das, was unsere Annahme bestätigt. Raymond Nickerson (1998) bezeichnet das als eines der am besten belegten Phänomene der Psychologie.

Peter Wason (1960) gab Versuchspersonen die Zahlenfolge 2, 4, 6 und bat sie, die Regel dahinter zu finden, indem sie eigene Folgen testeten. Die Regel lautete schlicht „aufsteigende Zahlen“. Die meisten bildeten jedoch eine engere Hypothese wie „immer plus zwei“ und testeten ausschließlich Folgen, die dazu passten. So erhielten sie eine Bestätigung nach der anderen, wurden immer sicherer und fanden die richtige Regel trotzdem nicht.

Klayman und Ha (1987) zeigen, dass dahinter oft eine positive Teststrategie steckt, bei der wir Fälle testen, die wir für Treffer halten. Das ist häufig sinnvoll, führt in manchen Aufgaben aber zuverlässig in die Irre.

Lord, Ross und Lepper (1979) legten Befürwortern und Gegnern der Todesstrafe dieselben gemischten Studien vor. Beide Seiten hielten jene Studien für besser, die ihre eigene Position stützten. Diese verzerrte Bewertung ist vielfach bestätigt. Dass Menschen durch gemischte Evidenz extremer werden, zeigen spätere Studien mit echten Vorher-nachher-Messungen dagegen kaum (Kuhn & Lao, 1996; Guess & Coppock, 2020). Bei Gegenargumenten bewegen wir uns meist leicht in Richtung der Evidenz, allerdings langsamer, als wir sollten.

Was hilft: Die Technik „Consider the opposite“ wirkt nachweislich. Lord, Lepper und Preston (1984) zeigten, dass die Frage, was wir bei gegenteiliger Evidenz gedacht hätten, die verzerrte Bewertung deutlich senkt, während ein Appell an Objektivität kaum half.

Motivated Reasoning

Ziva Kunda (1990) hat gezeigt, dass wir nicht einfach glauben, was wir wollen. Wir bauen vielmehr Begründungen, die zum gewünschten Ergebnis führen. Diese Begründungen halten wir anschließend für sachlich. Gewünschte Schlüsse akzeptieren wir mit niedrigeren Beweisstandards als unerwünschte, sodass das Glaubenwollen vor dem Wissenwollen kommt.

Der Grund liegt im System, denn Überzeugungen sichern Selbstwert, Zugehörigkeit und Handlungsfähigkeit. Wer seine Position verliert, verliert unter Umständen auch seinen Platz im Rudel. Kahan und Kollegen (2017) fanden Hinweise darauf, dass gute Rechenfähigkeit bei politisch aufgeladenen Themen eher zur Verteidigung als zur Erkenntnis genutzt wird.

Was hilft: Identität und Position werden getrennt. Wer eine Gegenposition vertreten darf, ohne dass es seine eigene sein muss, denkt freier.

Kognitive Dissonanz

Leon Festinger beschrieb 1957 in „A Theory of Cognitive Dissonance“ den Zustand, in dem zwei Gedanken einander widersprechen. Diese Spannung fühlt sich unangenehm an, weshalb wir sie loswerden wollen. Wir können dazu unsere Überzeugung ändern, wir können aber auch die störende Information abwerten, gezielt bestätigende Information suchen oder den Widerspruch für unwichtig erklären. Dabei wählen wir, was am wenigsten kostet. Je mehr Handlungen, Investitionen und öffentliche Bekenntnisse an einer Überzeugung hängen, desto billiger wird es, die Realität umzudeuten.

Den klassischen Beleg liefert ein Laborexperiment. Festinger und Carlsmith (1959) ließen Versuchspersonen eine stumpfe Aufgabe erledigen und anderen danach erzählen, sie sei spannend gewesen. Wer dafür nur einen Dollar erhielt, bewertete die Aufgabe anschließend positiver als jemand, der zwanzig Dollar bekam. Ohne guten äußeren Grund für die Lüge passten die Menschen ihre Überzeugung an.

Was hilft: Die Kosten des Umdenkens müssen sinken. Wer einen Irrtum eingestehen kann, ohne Gesicht oder Platz zu verlieren, wählt häufiger den ehrlichen Weg.

2.3 Phase 3: Die Versteinerung

In der dritten Phase wird das Gefilterte zu Stein. Was oft genug wiederholt wird, fühlt sich irgendwann an, als wäre es in die Höhlenwand gemeißelt.

■ Wiederholung wird Wahrheit

Hasher, Goldstein und Toppino (1977) zeigten, dass wir Aussagen, die wir schon einmal gehört haben, eher für wahr halten. Robert Zajonc (1968) hatte bereits zuvor nachgewiesen, dass bloße Wiederholung Dinge sympathischer macht. Unser Gehirn verwechselt dabei Leichtigkeit mit Wahrheit. Was bekannt ist, lässt sich flüssiger verarbeiten. Flüssig Verarbeitetes fühlt sich vertraut an, während sich Vertrautes wiederum richtig anfühlt.

Wissen schützt davor nur begrenzt. Fazio und Kolleginnen (2015) fanden den Effekt selbst bei Aussagen, die dem Vorwissen der Versuchspersonen widersprachen. Pennycook, Cannon und Rand (2018) zeigten, dass schon eine einzige vorherige Begegnung die geglaubte Richtigkeit einer Falschmeldung erhöht.

In Organisationen geschieht das über Folienstapel, Leitsätze und immer gleiche Erfolgsgeschichten. Was in jeder Präsentation steht, muss irgendwann nicht mehr belegt werden.

Was hilft: Die Herkunft wiederkehrender Aussagen wird geprüft. Bei Korrekturen wiederholen wir die richtige Information und nicht die falsche, denn wer einen Mythos zehnmals zitiert, um ihn zu widerlegen, trainiert ihn.

■ Der Backfire-Effekt als Mythos der Wiederholung

Nyhan und Reifler (2010) berichteten, dass Korrekturen falscher politischer Aussagen bei manchen Menschen die Fehlwahrnehmung sogar verstärkten. Daraus wurde die verbreitete Annahme, Fakten gingen bei Überzeugten grundsätzlich nach hinten los. Die spätere Forschung zeichnet ein anderes Bild. Wood und Porter (2019) testeten Korrekturen zu Dutzenden umstrittenen Aussagen in großen Experimenten und fanden kaum echte Backfire-Effekte. Im Durchschnitt bewegten Korrekturen die Menschen in Richtung der Fakten, was Nyhan später selbst bestätigt hat. Swire-Thompson, DeGutis und Lazer (2020) kommen in ihrer Übersichtsarbeit zum gleichen Schluss.

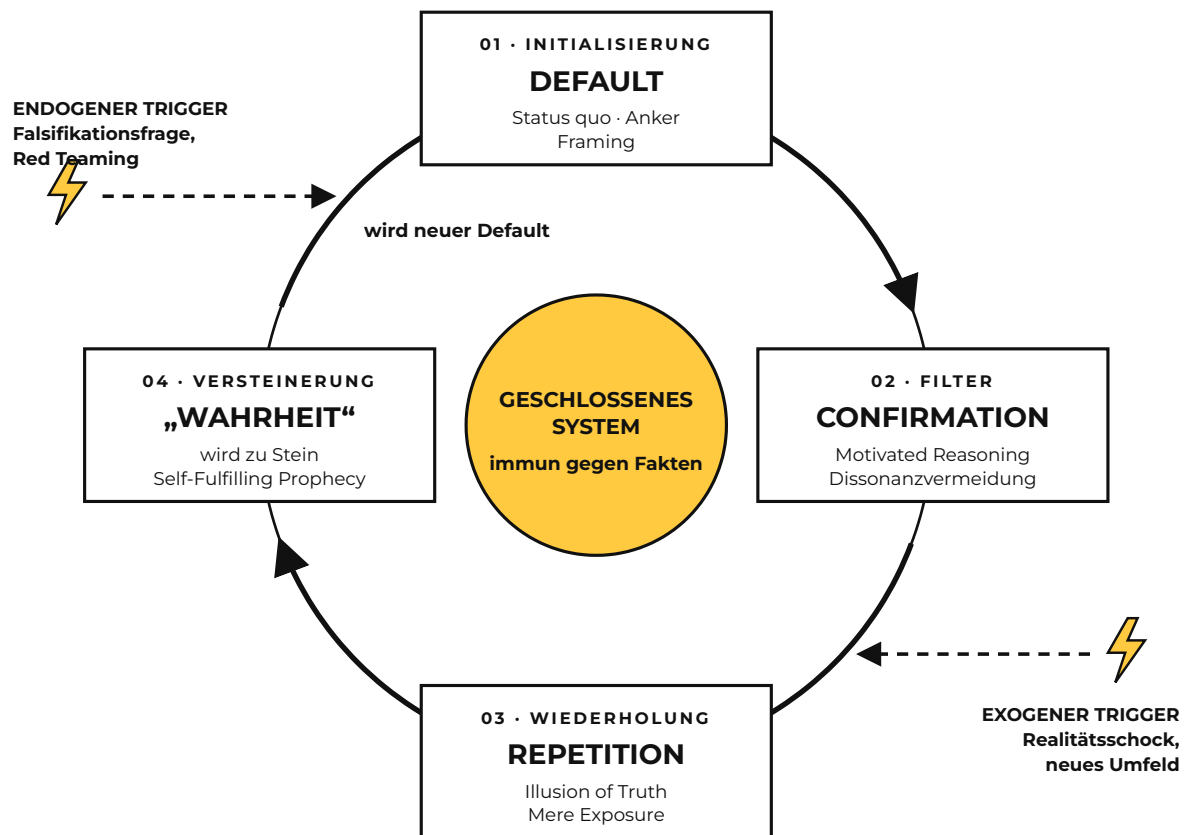
Der Kreislauf hält, weil Fakten gar nicht erst ankommen. Unsere Aufgabe liegt deshalb weniger im Widerlegen als im Öffnen des Filters.

■ **Die sich selbst erfüllende Prophezeiung**

Robert K. Merton (1948) beschrieb eine zunächst falsche Annahme, die ein Verhalten auslöst, durch das sie wahr wird. Sein Beispiel ist eine solvente Bank, über die ein Gerücht kursiert, sodass die Kunden ihr Geld abheben, bis die Bank tatsächlich zahlungsunfähig ist. Rosenthal und Jacobson (1968) berichteten vom Pygmalion-Effekt in Schulklassen. Spätere Analysen bestätigen den Mechanismus, finden aber meist kleinere Effekte, als oft erzählt wird (Jussim & Harber, 2005).

In Organisationen genügt dafür ein Etikett. Wer einem Team das Etikett „schwierig“ anheftet, gibt ihm weniger Spielraum, weniger Information und weniger Vertrauen. Das Team reagiert entsprechend, womit sich das Etikett bestätigt.

Was hilft: Etiketten werden wie Hypothesen behandelt. Dazu gehört die Frage, welches eigene Verhalten die beobachtete Wirkung miterzeugt.



Ansatzpunkte: zwischen Filter und Verstärkung (bevor Wiederholung versteinert)
und an der Übergabe von „Wahrheit“ zu neuem Default (bevor sie zur Voreinstellung wird)

Grafik 1 · Der geschlossene Bias-Kreislauf. Default, Confirmation, Repetition und „Wahrheit“ bilden einen Ring, in dem die Versteinerung den nächsten Default erzeugt. Exogene Trigger setzen idealerweise zwischen Filter und Wiederholung an, endogene Trigger an der Übergabe der „Wahrheit“ zur neuen Voreinstellung.

Drei weitere Mechanismen füttern den Kreislauf zusätzlich. Die Verfügbarkeitsheuristik lässt einen spektakulären Einzelfall schwerer wiegen als eine stille Statistik. Der Bandwagon-Effekt macht Wiederholung sozial, weil plausibel wirkt, was viele

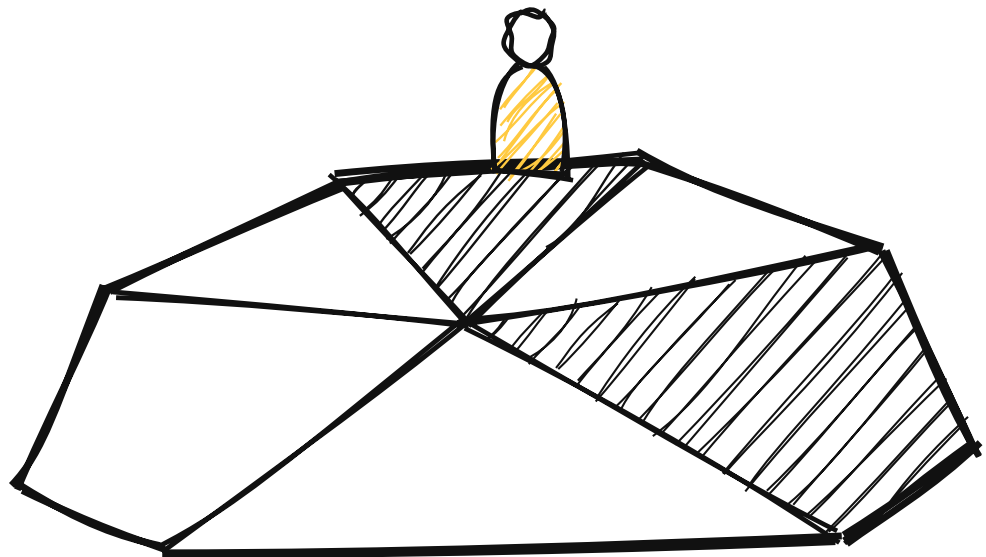
sagen. Der Survivorship Bias führt dazu, dass wir aus sichtbaren Erfolgen lernen, während wir die unsichtbaren Fehlschläge übersehen. Alle drei begegnen uns in Teil III wieder.

Unter welchen konkreten Bedingungen wärest du bereit zuzugeben, dass die aktuelle Annahme falsch ist?

Karl Popper hat in der „Logik der Forschung“ (1934) die Falsifizierbarkeit zum Kriterium gemacht. Eine Theorie ist demnach nur dann wissenschaftlich, wenn sie benennt, welche Beobachtung sie widerlegen würde. Hat im Strategiegelgespräch niemand eine Antwort auf diese Frage, ist die Annahme keine Hypothese mehr, sondern ein Glaubenssatz.

„Was in jeder Präsentation steht, muss irgendwann nicht mehr belegt werden.“

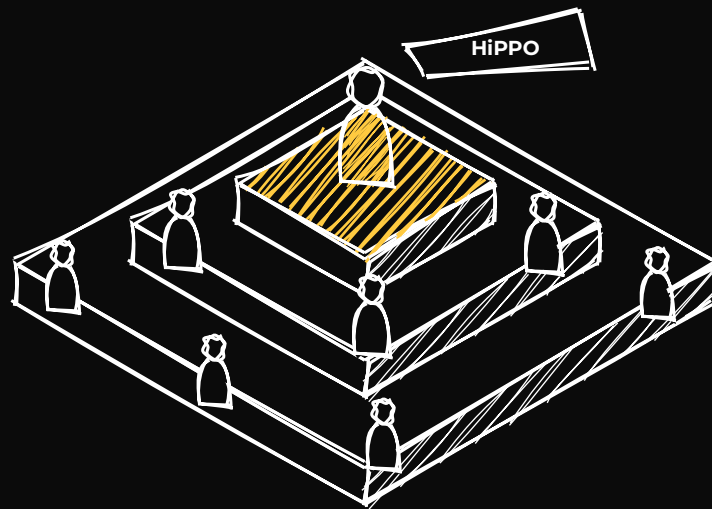
„Das geht bei uns nicht.“





TEIL III · KOLLEKTIVE VERZERRUNGEN

Organisationen sind **Verstärker.**



Was im einzelnen Kopf eine Neigung ist, wird im Team zur Norm, in der Hierarchie zur Weisung und im Prozess zur Voreinstellung. Dabei gilt seit der Höhle, dass man Mammuts gemeinsam jagt.

3.1 In Stein gemeißelte Annahmen

Jede Organisation speichert die Entscheidungen von gestern in Strukturen wie Prozessen, Formularen, Kennzahlen, Genehmigungswegen und Bewertungsrastern. Diese Strukturen sind in Stein gemeißelte Annahmen, die damals meist sinnvoll waren, ihre Begründung aber überleben. Damit wird der Default kollektiv, denn niemand muss eine Annahme aktiv vertreten, solange sie im Prozess steckt.

Eine Kennzahl, die seit Jahren im Monatsbericht steht, steuert Aufmerksamkeit, Prioritäten und Boni, auch wenn ihr Zweck längst weggefallen ist. Ein Bewertungsraster, das bestimmte Laufbahnen bevorzugt, filtert Jahr für Jahr dieselben Profile heraus.

Aus der Elektrotechnik kennen wir dafür den Begriff Blindleistung. Gemeint ist Energie, die im Netz pendelt, ohne Arbeit zu verrichten. In Organisationen entspricht ihr Aufwand ohne Wertschöpfung, etwa Berichte, die niemand liest, Abstimmungsschleifen, die niemand braucht, oder Kontrollen für Risiken, die es nicht mehr gibt. Blindleistung ist oft nichts anderes als institutionalisierter Status quo, weshalb die Gegenformel lautet: Wirkung statt Blindleistung.



Die KI-Illusion: Warum euer Rudel vor lauter Speeren die Jagd vergisst. Keynote beim Internationalen Forum Mechatronik in Memmingen, September 2026. Foto: Bayern Innovativ GmbH

WAKEUP!-MOMENT

Wer würde es merken, wenn dieser Bericht, dieses Meeting oder diese Freigabe morgen wegfällt?

Die ehrliche Antwort lautet erstaunlich oft, dass es niemand merken würde. Fragt man nach dem Warum, kommt selten ein Argument, sondern meist eine Geschichte über jemanden, der das früher einmal wollte.

3.2 Wenn das Rudel gemeinsam schweigt

■ Groupthink

Irving L. Janis beschrieb in „Groupthink: Psychological Studies of Policy Decisions and Fiascoes“ (1982) eine Denkweise eng verbundener Gruppen, in der das Streben nach Einmütigkeit die realistische Prüfung von Alternativen verdrängt. Seine Fallstudien reichen von der Invasion in der Schweinebucht 1961 bis zur Eskalation in Vietnam. Als Vorbedingungen nannte Janis hohe Kohäsion, die Abschottung von Außenmeinungen, direktive Führung, Zeitdruck und fehlende Verfahren für eine kritische Prüfung. Daraus entstehen typische Symptome wie das Gefühl der Unverwundbarkeit, wegerklärte Warnsignale, Selbstzensur, scheinbare Einstimmigkeit sowie „Meinungswächter“, die Störendes von der Gruppe fernhalten.

Groupthink ist Confirmation Bias im Plural. Jedes Mitglied filtert, während die Gruppe den Filter belohnt. Am Ende wirkt das Ergebnis besonders sicher, weil scheinbar alle unabhängig zum selben Schluss gekommen sind, obwohl in Wahrheit niemand unabhängig geurteilt hat.

Zur Einordnung gehört, dass das Modell zwar einflussreich, empirisch aber nur teilweise bestätigt ist. Nicht jede eng verbundene Gruppe entscheidet schlecht. Gefährlich wird Zusammenhalt erst, wenn er auf Konformitätsdruck und eine fehlende Widerspruchskultur trifft. Genau hier trennt sich ein Rudel, das zusammenhält, von einem Rudel, das gemeinsam schweigt.

Was hilft: Janis empfiehlt, dass die Führung ihre Meinung zunächst zurückhält, dass eine Person ausdrücklich kritisch prüft und dass externe Fachleute die bestehende Sicht angreifen.

■ Mitläufer und Social Proof

Solomon Asch (1956) ließ Versuchspersonen Linienlängen vergleichen. Obwohl die Aufgabe eindeutig war, schloss sich ein erheblicher Teil dem offensichtlich falschen Urteil einer Mehrheit an. Robert Cialdini (1984) beschrieb Social Proof als einen der stärksten Hebel sozialer Beeinflussung, weil wir uns unter Unsicherheit daran orientieren, was andere tun.

Sätze wie „Alle im Vorstand handeln so“ oder „Alle Mitbewerber haben jetzt eine KI-Strategie“ erklären Investitionswellen, Methodentrends und Beratungsmoden. Sie erklären auch, warum KI-Projekte starten, bevor jemand ihren konkreten Nutzen benannt hat.

WAKEUP!-MOMENT

Sobald in Aschs Experimenten eine einzige Person von der Mehrheit abwich, brach die Konformität stark ein.

Schon der erste Widerspruch verändert den Raum. Wir können ihn legitimieren, indem wir ihn als Rolle, als Regel oder als Teil der Kultur verankern.

■ Wir und die anderen

Henri Tajfel und Kollegen (1971) zeigten, dass Menschen ihre eigene Gruppe selbst dann bevorzugen, wenn die Zugehörigkeit auf einem völlig bedeutungslosen Merkmal beruht. Die Theorie der sozialen Identität (Tajfel & Turner, 1979) erklärt das mit dem Bedürfnis, Selbstwert über Zugehörigkeit zu gewinnen. In Organisationen entstehen daraus Silos, in denen die Entwicklung der Entwicklung glaubt und der Vertrieb dem Vertrieb. Bei Transformationen kommt die Grenze zwischen „den Projektleuten“ und „dem Tagesgeschäft“ hinzu.

Die belastbarste Antwort liefert die Kontaktforschung. Pettigrew und Tropp (2006) haben 515 Studien in einer Meta-Analyse zusammengeführt und gezeigt, dass Kontakt zwischen Gruppen Vorurteile abbaut. Am stärksten wirkt er unter den Bedingungen, die Gordon Allport bereits 1954 formuliert hat, nämlich bei gleichem Status, bei Zusammenarbeit, bei Unterstützung durch die Führung und bei gemeinsamen Zielen.

Mammuts jagt man also gemeinsam, allerdings unter einer wichtigen Bedingung. Ein Mammut, das eine einzelne Abteilung allein erlegen kann, verstärkt das Silo. Das Ziel muss deshalb so groß sein, dass es nur gemeinsam erreichbar ist.

3.3 Strukturelle Management-Fallen

■ „Wir haben schon zu viel investiert“

Rational betrachtet sollte bereits ausgegebenes Geld für die nächste Entscheidung keine Rolle spielen. Arkes und Blumer (1985) zeigten jedoch, dass es trotzdem zählt. Barry Staw (1976) beschrieb das eskalierende Commitment, bei dem Menschen, die persönlich für eine Fehlinvestition verantwortlich waren, nach schlechten Nachrichten weiter investierten. Verlustaversion, Selbstrechtfertigung und Dissonanzvermeidung greifen dabei ineinander.

Ein Abbruch würde den Verlust realisieren, die frühere Entscheidung als Fehler markieren und eine öffentliche Korrektur verlangen, während das Weitermachen all das in die Zukunft verschiebt. Das Überschallflugzeug Concorde hat dem Effekt sogar einen zweiten Namen gegeben.

Was hilft: Abbruchkriterien werden vor dem Start festgelegt. Über die Fortführung entscheiden Menschen, die den Start nicht verantwortet haben. Der Abbruch eines gescheiterten Vorhabens gilt dabei als Führungsleistung.

■ Die Überlebenden und der Blick zurück

Im Zweiten Weltkrieg analysierte der Statistiker Abraham Wald die Einschüsse an zurückgekehrten Bombern. Die Intuition empfahl, die Flugzeuge dort zu verstärken, wo die meisten Löcher waren. Wald drehte die Logik um, denn die Flugzeuge mit Treffern an den anderen Stellen waren gar nicht zurückgekommen. Auch in der Strategiearbeit lernen wir aus Erfolgsgeschichten, während jene, die mit derselben Strategie gescheitert sind, weder Fallstudien schreiben noch Keynotes halten.

Baruch Fischhoff (1975) zeigte den Rückschaufehler, bei dem Menschen, die den Ausgang kennen, glauben, sie hätten ihn vorhersehen können. Erfolge werden dann der Strategie zugeschrieben, auch wenn Glück entschieden hat. Misserfolge werden Menschen angelastet, die angeblich hätten wissen müssen, was niemand wissen konnte. Das zerstört die Bereitschaft, Risiken offen anzusprechen.

Was hilft: Entscheidungen werden im Moment der Entscheidung mit ihren Annahmen, Alternativen und Unsicherheiten dokumentiert. So bewerten wir später die Qualität der Entscheidung und nicht nur ihr Ergebnis.

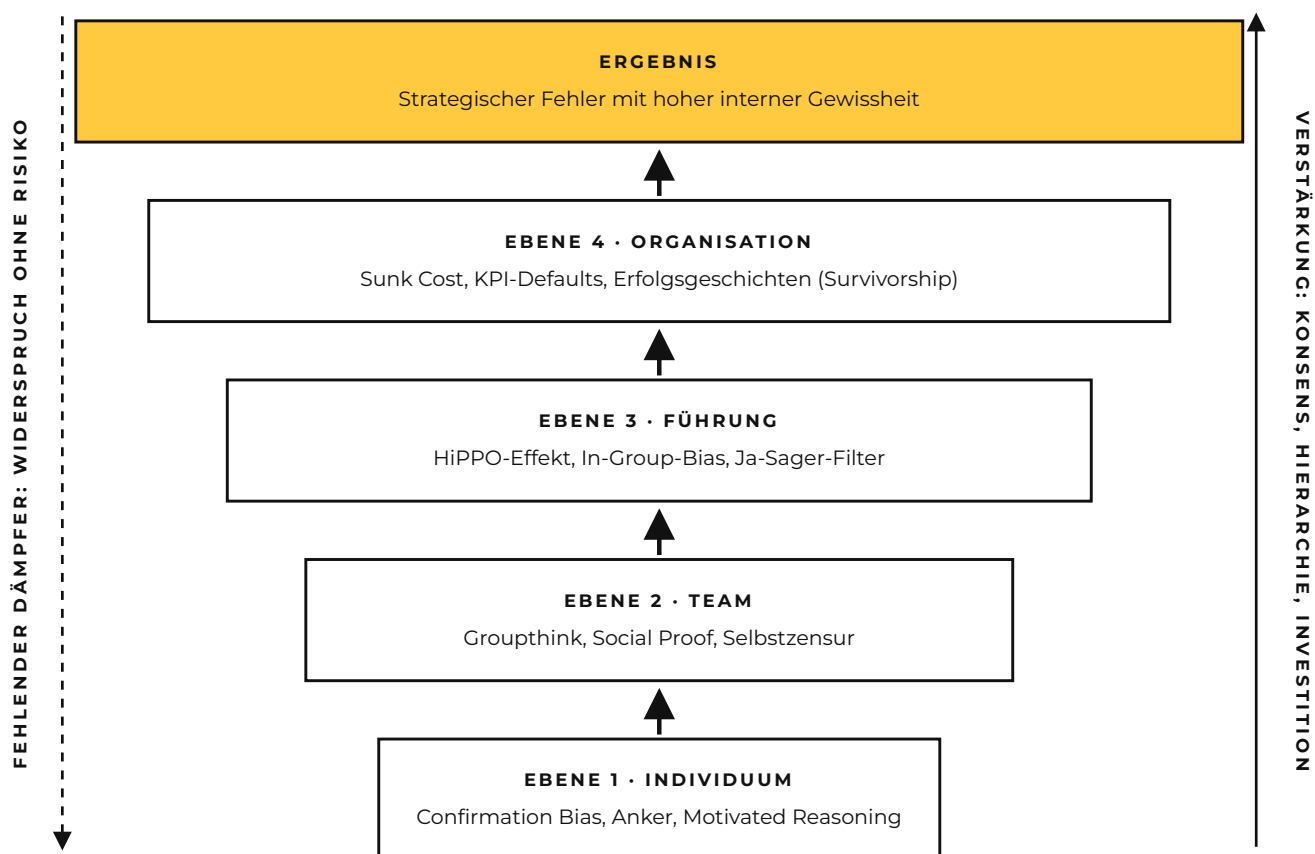
■ Der HiPPO im Raum

HiPPO steht für „Highest Paid Person’s Opinion“. Der Begriff, der dem Analytik-Experten Avinash Kaushik zugeschrieben wird, beschreibt Entscheidungen, die nach der Meinung der ranghöchsten Person fallen statt nach Evidenz. Der HiPPO bündelt drei Mechanismen aus Teil II. Seine Meinung wirkt als Anker, weil sie oft zuerst geäußert wird. Sie wirkt als Default, weil Abweichung ein Risiko bedeutet. Sie wirkt als Social Proof, weil Status als Kompetenz gelesen wird.

Kohavi und Kollegen beschreiben in „Trustworthy Online Controlled Experiments“ (2020), dass in kontrollierten Experimenten großer Digitalunternehmen nur ein Teil der Ideen, die intern als sicher galten, die erwartete Wirkung zeigte. Ohne Messung hätte niemand den Irrtum bemerkt.

Was hilft: Die ranghöchste Person spricht zuletzt. Schätzungen werden vorher schriftlich abgegeben. Wo immer es möglich ist, entscheidet ein kleiner Test statt einer großen Meinung.

Jede Ebene übernimmt die Verzerrung der darunterliegenden und fügt ihre eigene hinzu



Grafik 2 · Der organisationale Verstärkungsteppich. Individueller Confirmation Bias wird im Team durch Groupthink, in der Führung durch den HiPPO-Effekt und in der Organisation durch versunkene Kosten und Prozess-Defaults verstärkt. Es fehlt ein Dämpfer, also Widerspruch, der niemanden etwas kostet.

Organisationen scheitern selten an fehlendem Wissen, sondern daran, dass dieses Wissen den Raum nicht erreicht, in dem entschieden wird.

An dieser Stelle kommt ein zweiter Forschungsstrang ins Spiel. Amy Edmondson (1999) beschreibt psychologische Sicherheit als die gemeinsame Überzeugung im Team, dass zwischenmenschliche Risiken gefahrlos eingegangen werden können. Teams mit hoher psychologischer Sicherheit melden mehr Fehler, weil sie Fehler zeigen, nicht weil sie mehr davon machen. Googles Projekt „Aristotle“ kam 2015 zu dem Ergebnis, dass psychologische Sicherheit unter den untersuchten Faktoren der wichtigste für wirksame Teams war.

Psychologische Sicherheit ist damit der fehlende Dämpfer im Verstärkungsteppich, sie ist das Lagerfeuer. Ohne sie bleiben alle Methoden aus Teil V bloße Rituale, mit ihr werden sie zu wirksamen Werkzeugen.

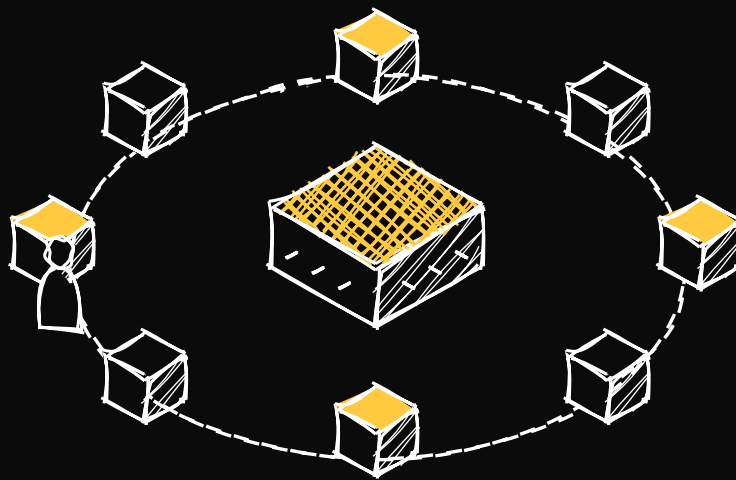
„Wo immer es möglich ist, entscheidet ein kleiner Test statt einer großen Meinung.“



IV

TEIL IV · KI UND ALGORITHMEN

Die Maschine lernt **unser** Gestern.



Daten · Modell · Empfehlung · Handeln · neue Daten

„Wir kaufen Werkzeuge, die schneller lernen als die Organisationen, die sie einsetzen. Deshalb scheitert Transformation selten an der Technik. Sie scheitert an den Menschen, die dabei vergessen wurden.“ Drohne, Roboter und KI sind die schärfsten Speere unserer Zeit.

4.1 Wie KI unsere Biases lernt

Maschinelles Lernen sucht Muster, ohne zu unterscheiden, ob ein Muster eine Ursache abbildet oder eine historische Ungleichheit. Harini Suresh und John Gutttag (2021)

haben systematisch beschrieben, an welchen Stellen eines Machine-Learning-Lebenszyklus Verzerrungen entstehen. Drei davon sind für Führungskräfte zentral.

ART	WAS PASSIERT
Historical Bias	Auch perfekte Daten bilden eine ungleiche Welt ab. Ein Modell, das aus vergangenen Beförderungen lernt, lernt, wer befördert wurde, nicht aber, wer gute Arbeit geleistet hat. So wird die Vergangenheit zum Default der Zukunft.
Representation Bias	Wenn Gruppen in den Daten unterrepräsentiert sind, funktioniert das Modell für sie schlechter. Die Studie Gender Shades hat das eindrücklich gezeigt.
Measurement Bias	Wenn das eigentliche Ziel nicht messbar ist, wird eine Ersatzgröße gemessen. Hängt diese für verschiedene Gruppen unterschiedlich mit dem Ziel zusammen, entsteht ein Bias, der im Code nicht zu finden ist.

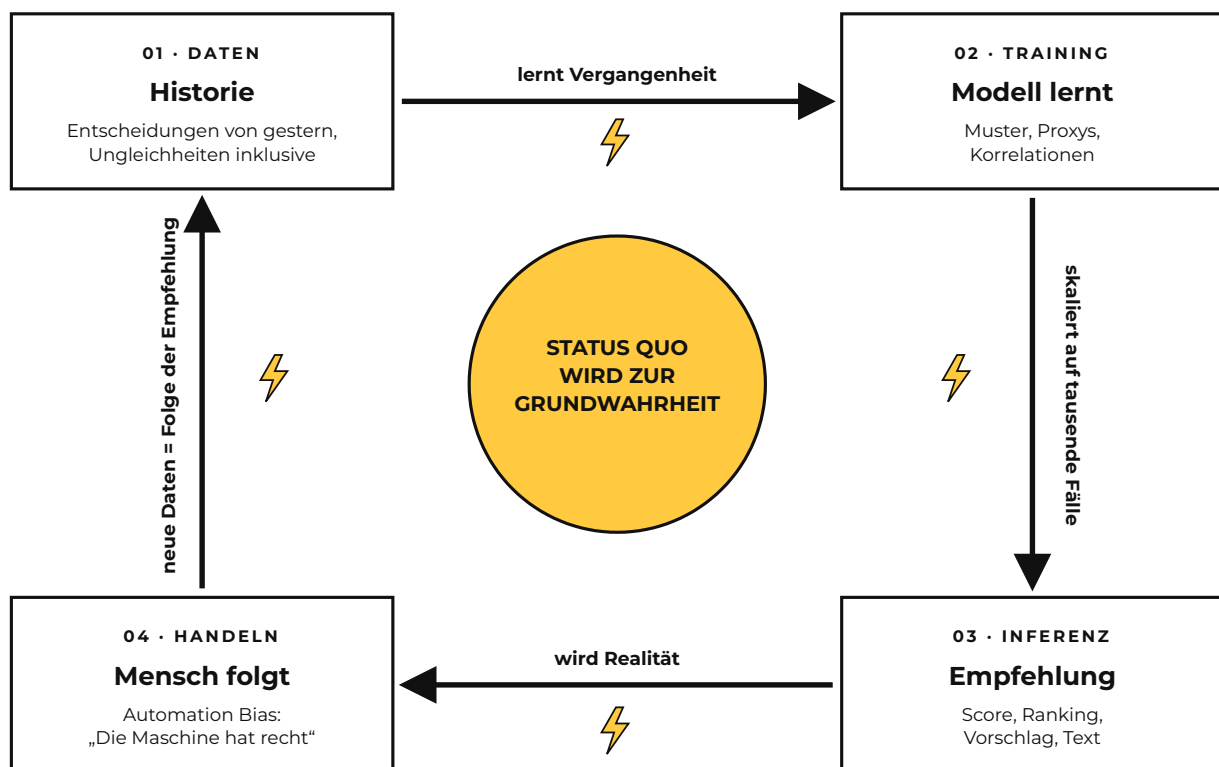
4.2 Der technische Verstärker

Feedback Loops

Ein Modell gibt eine Empfehlung, nach der Menschen handeln. Dieses Handeln erzeugt neue Daten, die in das nächste Training fließen. Dadurch sieht das Modell vor allem die Welt, die es selbst mitgestaltet hat. Was es nie vorgeschlagen hat, wird nie beobachtet und kann deshalb auch nie als gut erkannt werden.

Ensign und Kollegen (2018) haben das für Predictive Policing mathematisch als „Runaway Feedback Loop“ beschrieben. Streifen fahren dorthin, wo das Modell viel Kriminalität erwartet, wodurch dort mehr registriert wird, was wiederum das Modell bestätigt. Andere Gebiete, in denen ebenso viel geschieht, bleiben unterbeobachtet. Das System läuft so auf seine Ausgangsannahme zu, unabhängig davon, wie die Wirklichkeit aussieht.

Was hilft: Unabhängige Kontrollstichproben ohne Modelleinfluss decken die Verschiebung auf. Gezielte Exploration ergänzt die reine Ausnutzung bekannter Muster. Zudem prüfen wir regelmäßig, ob das Modell seine eigene Datenbasis verschoben hat.



Gold markiert: Unterbrechungspunkte. Datenaudit vor dem Training, Fairness-Metriken am Modell, Erklärung mit Einspruchsrecht an der Empfehlung, Kontrollstichproben ohne Modelleinfluss beim Handeln.

Grafik 3 · Der algorithmische Feedback-Loop. Historische Daten trainieren ein Modell, dessen Empfehlungen menschliches Handeln steuern. Dieses Handeln erzeugt die Daten für das nächste Training. Ohne Unterbrechung wirkt der Status quo mit jeder Runde objektiver, während er immer weniger überprüfbar wird.

Filterblasen

Empfehlungssysteme optimieren meist auf Interaktion und zeigen deshalb, was geklickt, angesehen oder geteilt wird. Eli Pariser (2011) nannte das Ergebnis Filterblase. Technisch betrachtet handelt es sich um skalierte Wiederholung, weil wir mehr von dem sehen, was wir bereits kennen. Deshalb halten wir es zunehmend für wahr und repräsentativ.

Bakshy, Messing und Adamic (2015) fanden allerdings heraus, dass die eigene Auswahl der Nutzer die Begegnung mit abweichenden Inhalten stärker einschränkte als der Algorithmus. Das entlastet die Technik nicht, zeigt aber, dass Maschine und menschlicher Confirmation Bias als Team arbeiten.

Automation Bias

Parasuraman und Manzey (2010) beschreiben die Tendenz, automatisierten Empfehlungen zu folgen und dabei eigene Prüfungen oder widersprechende Hinweise zu vernachlässigen. Bei einem Kommissionsfehler wird eine falsche Empfehlung umgesetzt, bei einem Omissionsfehler wird ein Problem übersehen, weil das System nicht gewarnt hat.

Automation Bias ist System 1 in Reinform, denn die Empfehlung liegt bereits vor, wirkt sachlich und spart Mühe. Sie ist damit Default, Anker und Autorität zugleich. Hinzu kommt, dass sich der Mensch weniger verantwortlich fühlt, wenn die

Maschine entschieden hat. Bei generativer KI verschärft sich das, weil ein Sprachmodell flüssigen, strukturierten und selbstbewussten Text liefert. Genau diese Flüssigkeit haben wir in Teil II als Wahrheitssignal kennengelernt.

WAKEUP!-MOMENT

Ein gut formulierter Irrtum wirkt glaubwürdiger als eine holprig formulierte Wahrheit.

4.3 Sprachmodelle zwischen Halluzination, Gefälligkeit und Selbstbezug

Halluzinationen sind Aussagen, die plausibel klingen und dennoch falsch sind, etwa erfundene Quellen, falsche Zahlen oder nicht existierende Urteile. Sie entstehen, weil Sprachmodelle wahrscheinliche Fortsetzungen erzeugen statt geprüfter Fakten. Dabei sind sie nicht zufällig verteilt, sondern folgen dem, was in den Trainingsdaten häufig vorkam, sodass das Modell bevorzugt das Erwartbare erfindet.

Gefälligkeit, im Englischen Sycophancy, wiegt für unser Thema noch schwerer. Sharma und Kollegen (2023) zeigten, dass Sprachmodelle, die mit menschlichem Feedback trainiert wurden, dazu neigen, der Meinung der fragenden Person zuzustimmen. Steckt in der Frage eine Annahme, bestätigt das Modell sie tendenziell.

Daraus entsteht eine gefährliche Kopplung. Ein Mensch mit Confirmation Bias stellt eine Frage, in der seine Annahme bereits steckt, woraufhin ein gefälliges Modell sie in sachlichem Ton bestätigt. Der Mensch erlebt flüssige Sprache und externe Bestätigung zugleich, bevor er die Aussage im Team wiederholt. Der Kreislauf aus Teil II läuft damit mit maschineller Unterstützung weiter, rund um die Uhr und in jedem Büro.

Shumailov und Kollegen (2024) zeigten in Nature, dass Modelle, die wiederholt auf Daten anderer Modelle trainiert werden, seltene Muster verlieren und verarmen. Sie nennen das „Model Collapse“, wobei der Rand der Verteilung zuerst verschwindet. Für eine Gesellschaft, die immer mehr Texte mit KI schreibt, bedeutet das, dass das Ungewöhnliche statistisch unsichtbar wird.

Alignment bezeichnet die Aufgabe, KI so zu gestalten, dass sie tut, was Menschen tatsächlich beabsichtigen. Gefälligkeit ist ein Alignment-Problem, weil das Modell auf Zustimmung optimiert statt auf Hilfe. Die praktische Konsequenz ist einfach: Wer KI zur Entscheidungsunterstützung nutzt, fordert sie ausdrücklich zum Widerspruch auf, weil sie das von selbst oft nicht tut.

4.4 Zwei Fallstudien, die jede Führungskraft kennen sollte

Gender Shades

Joy Buolamwini und Timnit Gebru prüften 2018 in „Gender Shades“, erschienen in den Proceedings of Machine Learning Research, drei kommerzielle Systeme zur Geschlechtsklassifikation. Bei Männern mit heller Haut lag die Fehlerquote bei höchstens 0,8 %, bei Frauen mit dunkler Haut dagegen bei bis zu 34,7 %. Ursache waren vor allem Daten, in denen helle Gesichter stark überrepräsentiert waren. Sichtbar wurden die Fehler erst an den Schnittstellen mehrerer Merkmale. Mehrere Anbieter haben ihre Systeme danach messbar verbessert.

Der falsche Proxy im Gesundheitswesen

Ziad Obermeyer, Brian Powers, Christine Vogeli und Sendhil Mullainathan untersuchten 2019 in Science einen in den USA weit verbreiteten Algorithmus. Er sollte Patientinnen und Patienten mit hohem Versorgungsbedarf für zusätzliche Betreuung identifizieren und nutzte dafür die erwarteten Gesundheitskosten als Zielgröße. Weil für Schwarze Patientinnen und Patienten bei gleichem Gesundheitszustand im Durchschnitt weniger Geld ausgegeben wurde, unterschätzte das Modell ihren Bedarf systematisch. Eine Korrektur hätte ihren Anteil an der zusätzlichen Betreuung von 17,7 auf 46,5 % erhöht.

WAKEUP!-MOMENT

Das Modell nutzte die Hautfarbe gar nicht und war technisch gut. Der Fehler lag in einer Annahme, die lange vor dem Training getroffen wurde, nämlich in der Wahl der Zielgröße.

Weitere Fälle aus Recruiting, Strafjustiz, Polizeiarbeit und Kreditvergabe sind im Anhang in Vertiefung B ausführlich beschrieben. Sie zeigen ein gemeinsames Muster: Fairness ist eine Wertentscheidung, die Menschen treffen müssen.

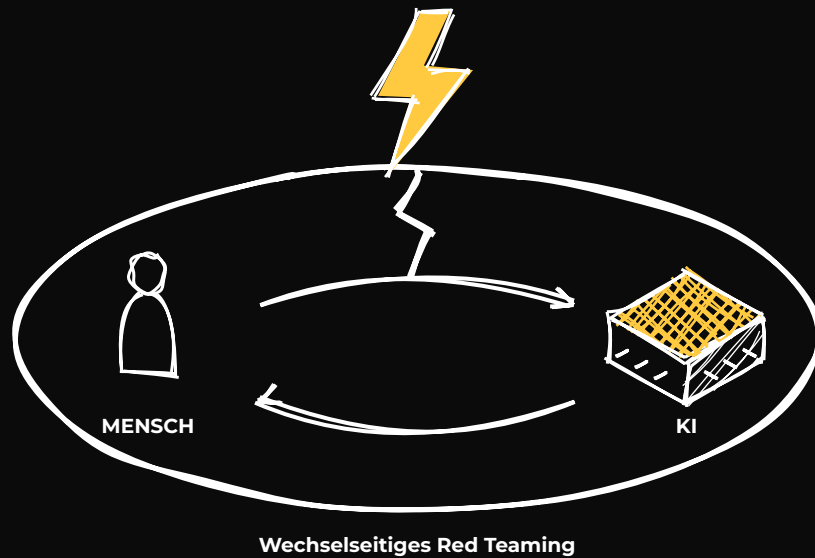
Kein Regler im Code nimmt ihnen diese Entscheidung ab. Cathy O’Neil hat das in „Weapons of Math Destruction“ (2016) zugespitzt, indem sie Modelle als in Mathematik eingebettete Meinungen beschreibt.

Ein Algorithmus wirkt objektiv, weil wir die Annahmen nicht mehr sehen, die in ihm stecken.

V

TEIL V · DEBIASING

Der Trigger und die Struktur, die ihn hörbar macht.



Ein geschlossenes System öffnet sich nicht durch Appelle. Es braucht einen Moment, in dem sich die Realität nicht mehr wegerklären lässt, sowie einen Raum, in dem dieser Moment ausgesprochen werden darf.

5.1 Die Anatomie des Triggers

Ein Trigger macht das Festhalten teurer als das Umdenken. Nach der Dissonanztheorie wählen wir immer die billigste Form, eine

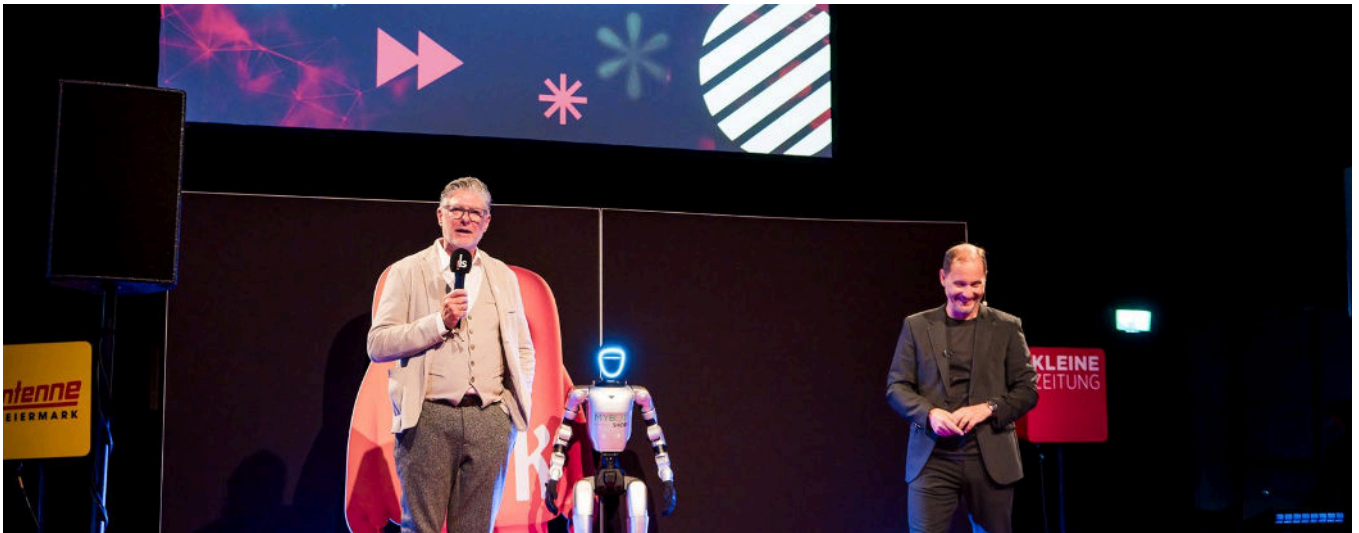
Spannung loszuwerden. Ein Trigger wirkt deshalb nur, wenn das Wegerklären mehr kostet als das Umdenken.

ART	BEISPIELE	WANN ER WIRKT
Exogen, abrupt	Marktschock, Kundenverlust, gescheiterter Launch, Regulierung	Er ist unübersehbar, kommt aber oft spät und teuer.
Exogen, sozial	Neue Menschen im Team, Blick von außen, Kundengespräch ohne Filter	Er wirkt, wenn die neue Stimme Gewicht bekommt und nicht als Außenseiter abgewertet wird.
Endogen	Popper-Frage, Pre-Mortem, Red Teaming, ein persönlicher Moment der Klarheit	Er wirkt früh und kostet wenig, braucht aber Disziplin und Struktur.

Exogene Trigger kommen ohnehin, offen ist nur, wie spät sie kommen. Ein Unternehmen, das erst durch den Verlust eines Großkunden erfährt, dass seine Qualitätsannahme falsch war, hat den Trigger zwar bekommen, allerdings zum teuersten möglichen Zeitpunkt. Endogene Trigger sind deshalb die eigentliche Führungsaufgabe, weil sie die Erschütterung in einen Moment holen, in dem sie noch wenig kostet.

Lewandowsky und Kollegen (2012) zeigen in ihrer Übersicht zur Korrektur von Fehlinformation, dass Menschen eine widerlegte Annahme oft weiter nutzen, wenn ihnen keine alternative Erklärung angeboten wird. Eine Lücke im Weltbild fühlt sich unangenehmer an als eine falsche Erklärung. Reißt der Trigger die alte Geschichte ein, ohne dass eine neue bereitsteht, bauen wir die alte wieder auf.

Wirksam wird der Bruch deshalb unter drei Bedingungen. Er muss unübersehbar sein, also so konkret, dass er sich nicht als Ausnahme abtun lässt. Er muss gesichtswahrend sein, damit das Umdenken weder Status noch den Platz im Rudel kostet, woran die meisten Organisationen scheitern. Schließlich muss er anschlussfähig sein, indem er eine neue Erklärung sowie einen nächsten Schritt anbietet.



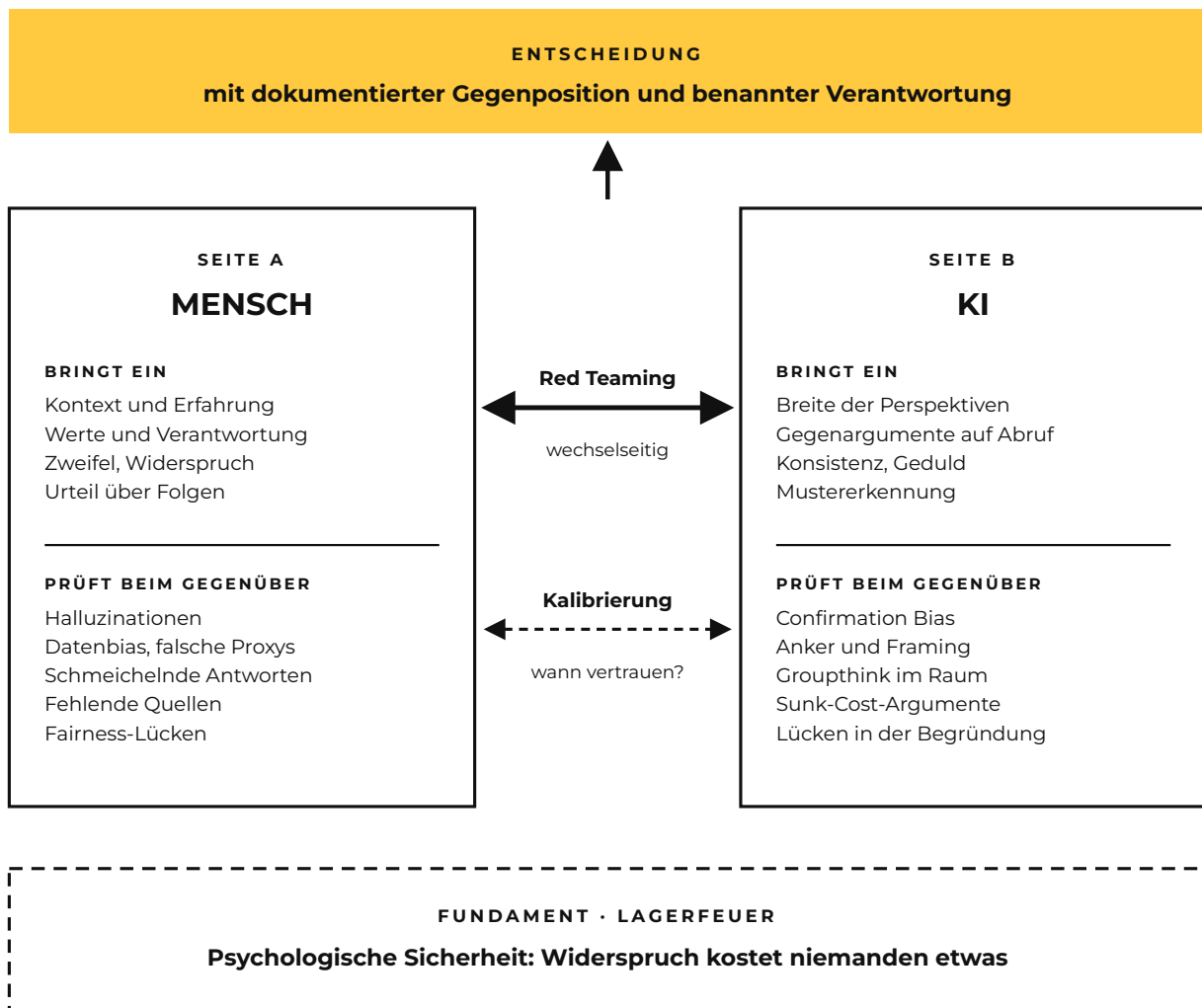
Mensch und Roboter gemeinsam auf der Bühne. ILS365 Main Event in Graz, Oktober 2026. Foto: Kevin Legenstein

5.2 Hybrid Intelligence mit KI als Gegenspieler

Dellermann und Kollegen (2019) beschreiben Hybrid Intelligence als Kombination menschlicher und maschineller Intelligenz, bei der beide Seiten voneinander lernen und gemeinsam bessere Ergebnisse erzielen als jede für sich. Für unser Thema ist eine Form besonders wertvoll, nämlich das wechselseitige Red Teaming.

Die Logik dahinter ist einfach. Menschen neigen zum Confirmation Bias, während Sprachmodelle zur Gefälligkeit neigen. Gemeinsam ergeben sie eine Bestätigungsmaschine. Setzen wir das Modell jedoch ausdrücklich in die Rolle des Gegenspielers, dreht sich die Dynamik. Die Maschine hat keine Angst um ihren Platz im Rudel und kann die Gegenposition deshalb in voller Stärke formulieren, ohne dass jemand im Raum sein Gesicht riskiert.

Charlan Nemeth und Kolleginnen (2001) liefern dazu eine wichtige Einschränkung. Ein nur gespielter Advocatus Diaboli regt deutlich weniger eigenständiges Denken an als echter Widerspruch, weil Teams merken, wenn Kritik ein Ritual ist. Die Gegenargumente der KI sind inhaltlich erzeugt und nicht gespielt. Ob sie gehört werden, entscheidet dennoch der Raum. Eine KI als Red Teamer in einem Team ohne psychologische Sicherheit erzeugt lediglich ein Protokoll, das niemand liest.



Grafik 4 · Das Hybrid-Intelligence-Modell. Mensch und KI prüfen sich gegenseitig. Die KI prüft den Menschen auf Bestätigungsfehler, Anker und Gruppendruck, während der Mensch die KI auf Halluzinationen, Datenbias und Gefälligkeit prüft. Das Fundament bildet psychologische Sicherheit, ohne die jede Gegenposition folgenlos bleibt.

WAKEUP!-MOMENT · PROMPT FÜR DEINE NÄCHSTE ENTSCHEIDUNGSVORLAGE

Du bist der kritische Gegenspieler in einem Entscheidungsgremium. Deine Aufgabe ist es nicht, mir zuzustimmen.

Hier ist unsere Vorlage: [Vorlage einfügen].

1. Formuliere die drei stärksten Argumente gegen diese Entscheidung so überzeugend, wie ein kluger Gegner sie formulieren würde.
2. Benenne die Annahmen, die nicht belegt sind, sowie jede Zahl, die wie ein Anker mit unklarer Herkunft wirkt.
3. Beschreibe, welche Beobachtung in den nächsten 90 Tagen zeigen würde, dass wir falsch liegen.
4. Wenn du mir in Teilen zustimmst, sag es erst nach den Punkten 1 bis 3.

WAKEUP!-MOMENT · PROMPT FÜR EIN PRE-MORTEM

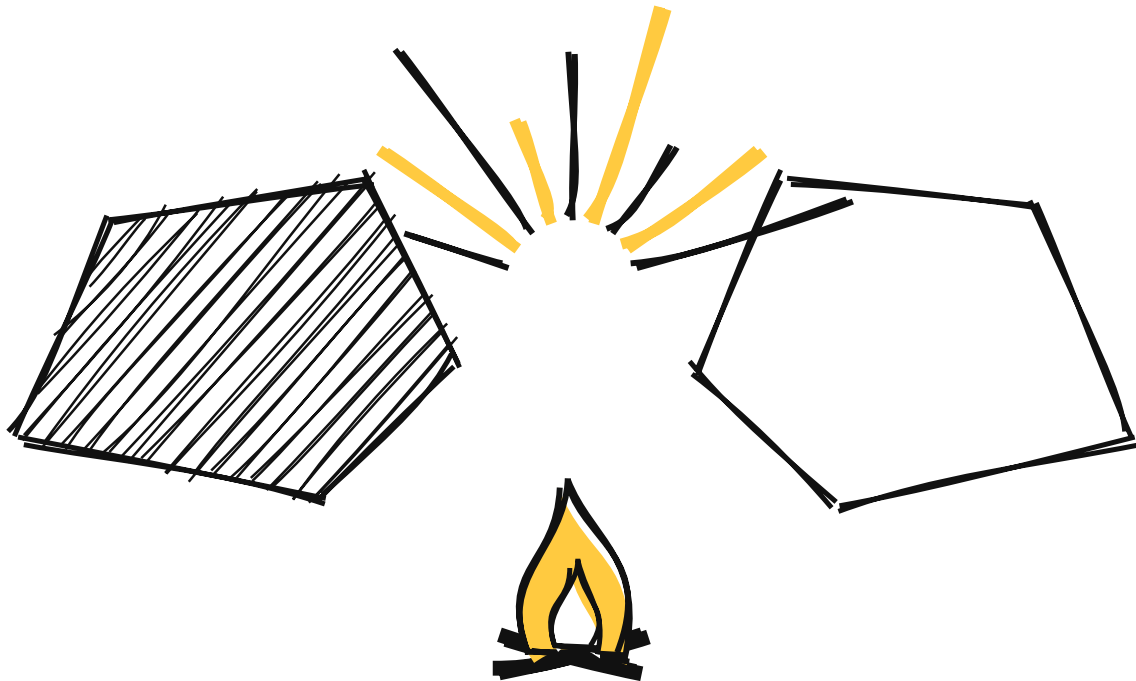
Stell dir vor, wir schreiben den [Datum in 12 Monaten] und unser Vorhaben [kurz beschreiben] ist klar gescheitert.

Schreib die nüchterne Ursachenanalyse für den Lenkungsausschuss.

Nenne mindestens fünf plausible Ursachen, von denen mindestens zwei mit Menschen, Kultur oder Zusammenarbeit zu tun haben und nicht mit Technik.

Ordne jeder Ursache ein frühes Warnsignal zu, das wir heute schon beobachten könnten.

„Eine KI als Red Teamer in einem Team ohne psychologische Sicherheit erzeugt lediglich ein Protokoll, das niemand liest.“

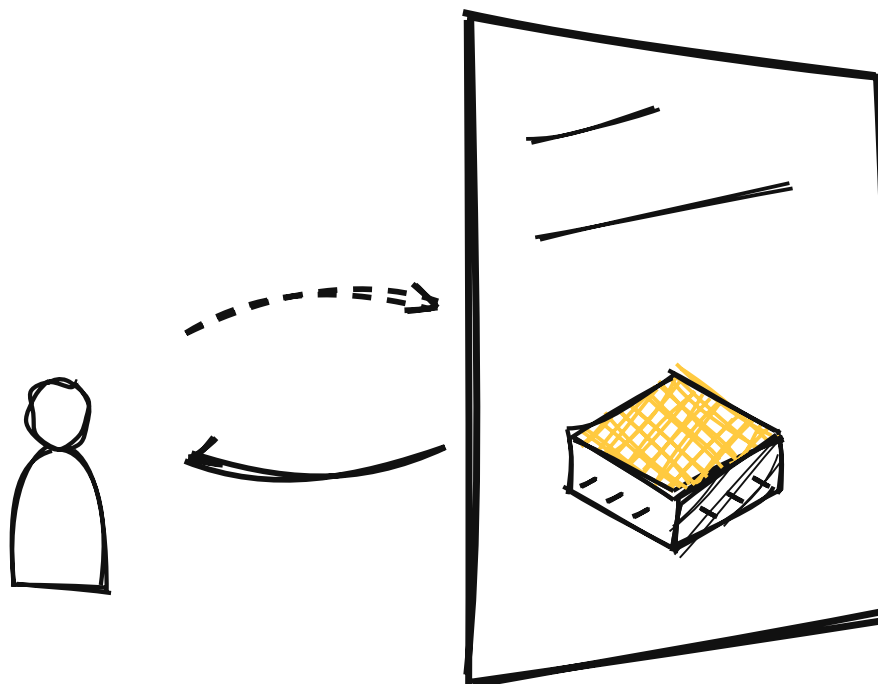


5.3 Erklärbare KI gegen Automation Bias

Erklärbare KI soll zeigen, warum ein System eine bestimmte Empfehlung gibt. Die naheliegende Hoffnung lautet, dass wir Fehler erkennen, sobald wir die Begründung sehen. So einfach ist es jedoch nicht. Bansal und Kollegen (2021) fanden heraus, dass Erklärungen die Akzeptanz von KI-Empfehlungen erhöhten. Das galt unabhängig davon, ob die Empfehlung richtig oder falsch war. Eine Begründung lässt eine Aussage stimmiger wirken, wobei sich Stimmiges wahr anfühlt, ein Muster, das wir bereits kennen.

Wirksamer waren in Studien von Buçinca, Malaya und Gajos (2021) sogenannte kognitive Zwangspunkte, bei denen Menschen zuerst ein eigenes Urteil bilden, bevor sie die KI-Empfehlung sehen. Das senkte übermäßiges Vertrauen deutlich, wurde aber als anstrengender erlebt. Darin liegt der Kern, denn Debiasing kostet Aufmerksamkeit und Zeit. Wer es gratis haben will, bekommt es nicht.

MASSNAHME	WIRKUNG AUF AUTOMATION BIAS
Erklärung neben der Empfehlung	Sie kann das Vertrauen ungeprüft erhöhen und reicht allein nicht aus.
Eigenes Urteil vor der Empfehlung	Es senkt die Ankerwirkung der Maschine deutlich.
Unsicherheit sichtbar machen	Das hilft, wenn die Konfidenzangaben gut kalibriert sind.
Begründungspflicht bei Übernahme	Sie schaltet System 2 ein, solange sie nicht zum Formular zum Abhaken wird.
Quellenpflicht bei generativer KI	Sie macht Halluzinationen prüfbar, sofern die Quellen auch geöffnet werden.



5.4 Debiasing als Struktur

P re-Mortem. Gary Klein (2007) hat das Pre-Mortem als Gegenstück zur Nachbetrachtung beschrieben. Vor dem Start stellt sich das Team vor, das Vorhaben sei bereits gescheitert, woraufhin jede Person allein die Gründe aufschreibt. Die Methode beruht auf Befunden zur „prospective hindsight“ (Mitchell, Russo & Pennington, 1989), wonach wir mehr und konkretere Erklärungen finden, wenn wir uns ein Ereignis als bereits eingetreten vorstellen. Zweifel wird so zur Aufgabe statt zum Verrat.

Erst schreiben, dann reden. Jede Person hält ihre Einschätzung schriftlich fest, bevor im Raum gesprochen wird, wobei die ranghöchste Person zuletzt spricht. Kahneman, Sibony und Sunstein (2021) empfehlen diese Unabhängigkeit der Urteile als Kernprinzip guter Entscheidungshygiene.

Blind Review. Goldin und Rouse (2000) berichteten, dass Vorspiele hinter einem Vorhang die Chancen von Frauen in US-Orchestern erhöhten. Die statistische Aussagekraft der Studie ist umstritten, das Prinzip bleibt jedoch stark: Was für eine Entscheidung keine Rolle spielen soll und nicht sichtbar ist, kann auch nicht wirken. Übertragen lässt sich das auf Bewerbungen, Projektanträge, Ideenwettbewerbe und die Lieferantenauswahl.

Vielfalt im Gremium. Samuel Sommers (2006) zeigte in Simulationen von Geschworenenberatungen, dass gemischte Gruppen länger berieten, mehr Fakten diskutierten und weniger sachliche Fehler machten. Auch die Mitglieder der Mehrheitsgruppe arbeiteten dabei gründlicher. Ohne psychologische Sicherheit erzeugt Vielfalt allerdings eher stille Anwesenheit als hörbare Perspektive.

Abbruchkriterien und Entscheidungsjournal. Abbruchkriterien, die vor dem Start feststehen, sowie ein Journal, das Annahmen, Alternativen und Unsicherheiten zum Zeitpunkt der Entscheidung festhält, machen Poppers Gedanken alltagstauglich.

5.5 Algorithmen prüfen

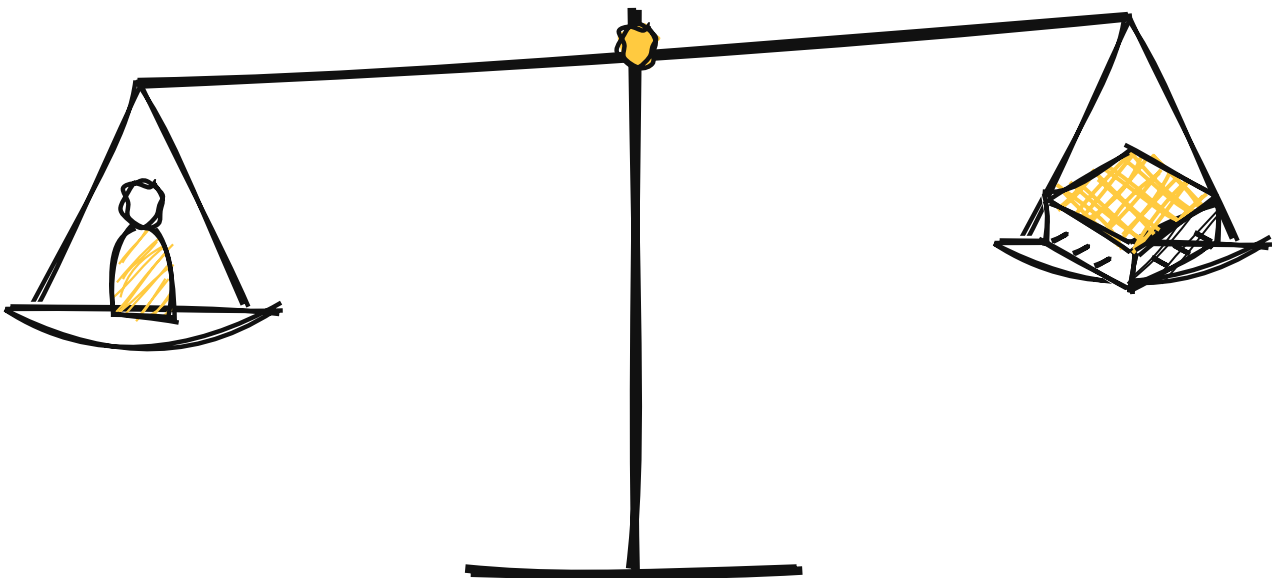
Für KI-Systeme gibt es inzwischen einen erprobten Werkzeugkasten. Datasheets for Datasets dokumentieren Herkunft, Zusammensetzung und Lücken von Daten. Model Cards beschreiben, wofür ein Modell gedacht ist und wie es für verschiedene Gruppen abschneidet. Fairness-Metriken machen Unterschiede messbar, disaggregierte Evaluation prüft Untergruppen statt Durchschnittswerte, laufendes Monitoring erkennt, ob ein Modell seine eigene Datenbasis verschiebt. Details dazu stehen im Anhang in Vertiefung C.

Eine Grenze bleibt allerdings bestehen. Weil sich nicht alle Fairness-Kriterien gleichzeitig erfüllen lassen, kann keine Metrik entscheiden, welche Fairness ein System haben soll.

WAKEUP!-MOMENT

Ein Audit macht Bias sichtbar, doch ob er akzeptabel ist, entscheidet kein Algorithmus.

Das entscheiden Menschen, die Verantwortung tragen und die Betroffenen kennen.



5.6 Wenn Lehrbücher im Kreis laufen

Der Kreislauf macht auch vor der jahrzehntelang zitiert, gelten heute aber als Wissenschaft nicht halt. Zwei widerlegt oder als Beleg unbrauchbar. Lehrbuchgeschichten wurden

LEHRBUCHGESCHICHTE	WAS LANGE ERZÄHLT WURDE	WAS DIE PRÜFUNG ZEIGT
Robbers Cave (Sherif, 1954)	Konkurrenz erzeugt spontan Feindschaft, die gemeinsame Ziele wieder auflösen.	Die Archivauswertung von Perry (2018) zeigt, dass ein erster Versuch scheiterte und unveröffentlicht blieb. Konflikt und Kooperationskrisen wurden von Mitarbeitern geschürt.
Denken kostet Energie (Glukose- und Ego- Depletion-Modell)	Nachdenken verbraucht spürbar Energie, weshalb das Gehirn mit Abkürzungen spart.	Der Hirnstoffwechsel ändert sich bei Denkaufgaben kaum (Raichle & Gusnard, 2002). Ego Depletion zeigte in 23 Laboren keinen Effekt (Hagger et al., 2016).

Die Wissenschaft hat damit genau den Kreislauf durchlaufen, den diese Ausgabe beschreibt. Eine stimmige Geschichte wurde zur Voreinstellung, wobei Lehrbücher Kritik herausfilterten, weil die

Geschichte so gut zur Theorie passte. Jede neue Auflage wiederholte sie, sodass sie versteinerte, bis schließlich jemand in die Archive ging.

WAKEUP!-MOMENT · DREI LEHREN FÜR ORGANISATIONEN

Auch Profis laufen im Kreis. Gute Absicht und Intelligenz schützen nicht davor, Struktur dagegen schon.

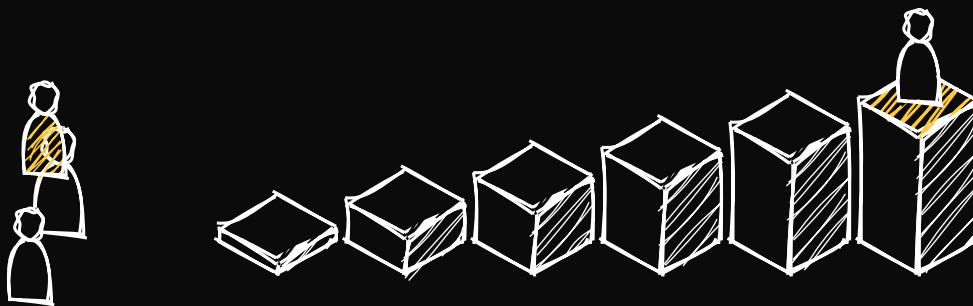
Der wirksamste Trigger ist die Quelle. Gina Perry hat nichts Neues erfunden, sondern im Archiv nachgeschaut. Die Frage „Woher wissen wir das eigentlich?“ verdient deshalb einen festen Platz in jedem Team.

Korrektur ist Stärke. Die Wissenschaft verfügt mit Replikation und Archivprüfung über eingebaute Trigger. Organisationen ohne solche Trigger laufen länger im Kreis, als ihnen guttut.

VI

TEIL VI · AUSBLICK

Mensch, Organisation und KI: **Wer prüft wen?**



Kompetenz · Governance · Kultur: Stufe für Stufe

Offen ist nicht, ob Maschinen verzerrt sind. Offen ist, ob wir eine Kultur bauen, in der Verzerrungen bemerkt und ausgesprochen werden, bevor sie zur Infrastruktur werden.

6.1 Das Risiko der Selbstreferenz

Bis vor wenigen Jahren war der Bias-Kreislauf an menschliches Tempo gebunden, denn eine Annahme brauchte Monate, um sich durch Sitzungen und Berichte zu verfestigen. Heute durchlaufen Empfehlungssysteme, Bewertungsmodelle und Sprachmodelle denselben Kreislauf in Stunden, noch dazu in tausenden Fällen gleichzeitig.

Denken wir an eine Gesellschaft, in der Modelle, die auf historischen Daten trainiert sind, zunehmend Entscheidungen vorbereiten, während Texte zunehmend von Modellen stammen, die auf Texten anderer Modelle trainiert wurden. Eine solche Gesellschaft läuft Gefahr, ihre Vergangenheit als Zukunft zu wiederholen. Model Collapse auf der technischen Ebene und Automation Bias auf der menschlichen Ebene verstärken sich dabei gegenseitig. Das Ungewöhnliche verschwindet zuerst, obwohl gerade das Ungewöhnliche die Quelle von Innovation und Korrektur ist.

6.2 Zwischen Bauchgefühl und Datenhörigkeit

Zwei Fehlhaltungen prägen viele Diskussionen über KI. Wer das Bauchgefühl zum Dogma macht, erlebt jede Datenanalyse als Angriff auf die eigene Erfahrung, was einem System 1 ohne Prüfung entspricht. Wer dagegen datenhörig ist, hält jede Zahl für objektiv, solange sie aus einem System kommt, was einem Automation Bias mit gutem Gewissen gleichkommt. Der dritte Weg heißt kalibriertes Vertrauen, bei dem wir einer Quelle so weit vertrauen, wie ihre nachgewiesene Treffsicherheit in genau diesem Kontext es rechtfertigt.

Das gilt für Menschen ebenso wie für Maschinen. Fachleute sind dort verlässlich, wo sie regelmäßig und schnell Feedback auf ihre Urteile bekommen. Darauf haben sich Kahneman und Klein (2009) als Vertreter zweier lange gegensätzlicher Forschungstraditionen geeinigt. Modelle sind dort verlässlich, wo ihre Trainingsdaten die Einsatzsituation gut abbilden. Außerhalb dieser Zonen werden beide zum Risiko.

Die wertvollste Fähigkeit im KI-Zeitalter besteht darin, die eigene Antwort ernsthaft zu prüfen.

6.3 Leitplanken durch den EU AI Act

Mit der KI-Verordnung, der Verordnung (EU) 2024/1689, hat die EU den ersten umfassenden Rechtsrahmen für KI geschaffen. Sie gilt seit dem 1. August 2024 und wird stufenweise anwendbar. Drei Artikel greifen direkt in die Mechanik dieser Ausgabe ein.

ARTIKEL	INHALT UND BEZUG ZUM KREISLAUF
Art. 10 · Datenqualität	Anbieter von Hochrisiko-Systemen müssen ihre Daten auf mögliche Verzerrungen prüfen und gegensteuern. Das ist die rechtliche Antwort auf Historical und Representation Bias.
Art. 14 · Menschliche Aufsicht	Wer ein Hochrisiko-System beaufsichtigt, muss sich der Tendenz bewusst sein, sich automatisch auf dessen Ausgaben zu verlassen. Der Gesetzestext nennt den Automation Bias dabei ausdrücklich.
Art. 4 · KI-Kompetenz	Seit dem 2. Februar 2025 müssen Anbieter und Betreiber für ausreichende KI-Kompetenz ihres Personals sorgen. Das Bewusstsein für Bias gehört zum Kern dieser Kompetenz.

Systeme im Personalwesen, etwa zur Auswahl von Bewerbungen, zählen zu den Hochrisiko-Anwendungen nach Anhang III. Mit dem „Digital Omnibus“ ist vorgesehen, diese Pflichten auf Dezember 2027 zu verschieben. Maßgeblich ist jeweils der aktuelle Stand des Verfahrens.

Regulierung setzt Mindeststandards, kann aber keine Kultur verordnen. Ein Unternehmen kann alle Dokumentationspflichten erfüllen und trotzdem im geschlossenen Kreislauf entscheiden. Wachsamkeit entsteht durch Menschen, die bemerken, dass etwas nicht stimmt, sofern sie einen Ort haben, an dem sie es aussprechen können.

VII

■ TEIL VII · TRANSFER

Das Lagerfeuer **zuerst**.



Wo Widerspruch nichts kostet, hat Wahrheit eine Chance.

Was das ab Montag bedeutet, folgt der Reihenfolge, die sich in Transformationen bewährt hat: zuerst die Menschen, dann die Projekte und danach die Strategie.

7.1 Warum die Reihenfolge zählt

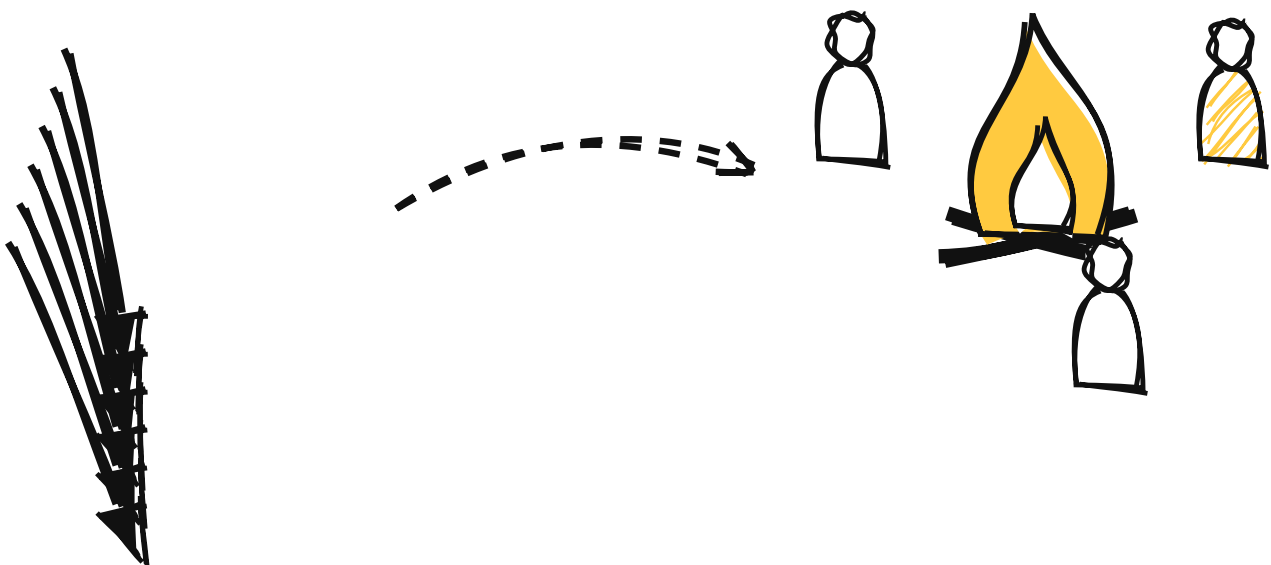
Viele Organisationen beginnen beim Debiasing mit der Technik, also mit Fairness-Tools, Audit-Software und Dashboards. Das ist verständlich, weil sich Technik kaufen und messen lässt, es ist aber trotzdem die

falsche Reihenfolge. Ein Audit, dessen Befund niemand auszusprechen wagt, verändert nichts. Ein Pre-Mortem ohne Vertrauen erzeugt Floskeln. Ein KI-Red-Teamer in einem Gremium mit starkem HiPPO liefert Argumente fürs Archiv.

WAKEUP!-MOMENT

**Technologie ohne Menschen scheitert. Projekte ohne Struktur verpuffen.
Strategie ohne Momentum bleibt Theorie.**

„Ein Audit, dessen Befund niemand auszusprechen wagt, verändert nichts.“



7.2 Zwölf Interventionen mit Aufwand und Wirkung

Jede Intervention ist einem der vier Züge von **neue Wege gehen** und **G für klaren Rahmen und Go**. Human **BIAss** zugeordnet. **H** steht für Hinterfragen, **U** für Bias unterbrechen, **N** für

SCHRITT	ZUG	INTERVENTION	AUFWAND	ERWARTETE WIRKUNG
People fit	H	Popper-Frage als feste Schlussfrage jeder Strategiesitzung	sehr gering	Macht Glaubenssätze sofort sichtbar
	U	Führung spricht zuletzt und benennt pro Quartal einen eigenen Irrtum	gering	Stärkster Hebel für psychologische Sicherheit
	H	„Consider the opposite“ im Team, wöchentlich fünf Minuten	gering	Trainiert System 2, wirkt über Wochen
	U	Rahmenprüfung: jede wichtige Vorlage als Gewinn und als Verlust formulieren	gering	Deckt Framing vor der Entscheidung auf
Project fit	N	Pre-Mortem vor jedem größeren Vorhaben	ein bis zwei Stunden	Hoch bei übersehenen Risiken
	U	Abbruchkriterien vorab, Fortführung durch Unbeteiligte	mittel	Bricht die Sunk-Cost-Dynamik
	U	Erst schreiben, dann reden: unabhängige Schätzungen	gering	Senkt Anker und HiPPO-Effekt deutlich
	H	Blindleistungs-Check: Wer würde es merken, wenn es wegfällt?	gering	Legt Defaults frei, schafft Zeit
Strategy fit	G	KI als Red Teamer für alle strategischen Vorlagen	Prompt-Standard und Schulung	Gegenposition ohne sozialen Preis, wirkt nur mit Lagerfeuer
	G	Eigenes Urteil vor der KI-Empfehlung als Arbeitsregel	Disziplin	Senkt Automation Bias spürbar
	N	Entscheidungsjournal mit Annahmen, Alternativen und Unsicherheit	mittel	Ehrliches Lernen, Schutz vor dem Rückschaufehler
	G	KI-Governance mit Datenblatt, Modellbeschreibung, Untergruppen-Prüfung und Monitoring	hoch	Erfüllt Pflichten, macht maschinellen Bias bearbeitbar

7.3 Aus der Praxis: Late Follower mit Turbo

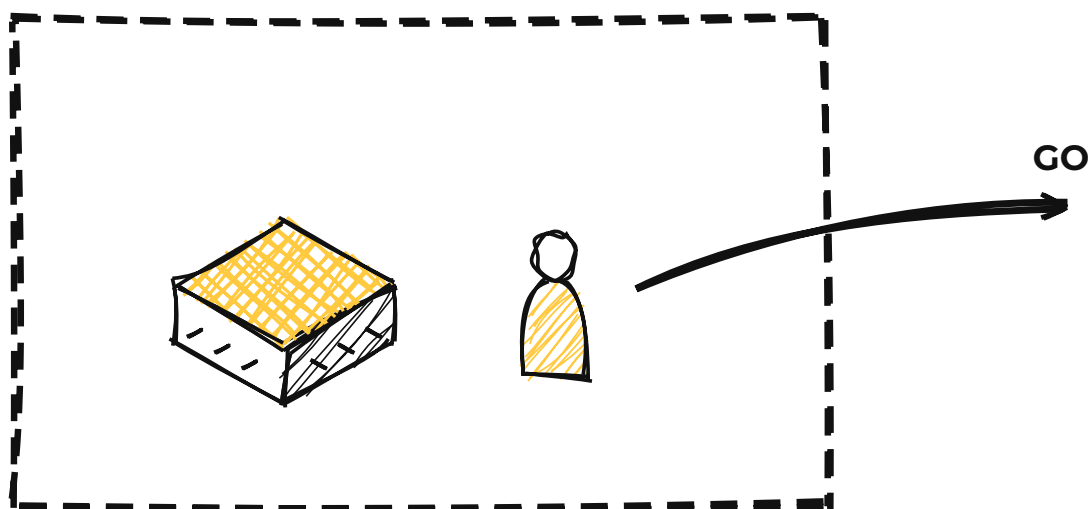
Wie Human BIAss im Mittelstand aussieht, zeigt ein aktueller Fall aus einem produzierenden Unternehmen. Führung und Mitarbeitende haben dort gemeinsam einen konsequenten Modernisierungskurs eingeschlagen.

Das Unternehmen versteht sich selbst als „Late Follower“, der KI nun mit hoher Geschwindigkeit einführt. Dahinter steht die Erfahrung, dass ein mittelständisches Unternehmen mit KI so schnell und so breit handeln kann wie ein Konzern.

ZUG	WAS IM FALL SICHTBAR WIRD
Hinterfragen	Der Glaubenssatz, KI sei etwas für Konzerne, kippt, weil kleine Unternehmen durch KI plötzlich wie Konzerne agieren können. Wer KI nutzt, muss dafür kein Coder sein.
Bias unterbrechen	Entschieden wurde in Tagen. Das ist das Gegenmodell zu jener Zögerlichkeit, die in Europa oft den Takt vorgibt.
Neue Wege gehen	Der Start erfolgte bei den „Low Hanging Fruits“, also bei dem Use Case, der jeden Tag nervt. Dazu gehört eine Fehlerkultur, in der Scheitern kein Scheitern ist.
Klarer Rahmen und Go	Eine DSGVO-konforme Plattform bildet den Rahmen, während die Führung KI vorlebt, statt sie zu delegieren. KI dient dabei als Sparring-Partner für bessere Entscheidungen.

Die Herausforderungen liegen ebenso offen auf dem Tisch. Ängste im Team lassen sich nur abbauen, wenn Führungskräfte KI selbst vorleben. Schulungsprogramme sind nach zwei Wochen überholt. Außerdem müssen die Generationen verbunden werden, vom Produktionsmitarbeiter bis zum Controller. Daraus folgt kein Zögern, sondern ein Go für mehr. Der Appell aus dem Unternehmen ist entsprechend klar: Seid mutig und geht den ersten Schritt.

Der Fall ist anonymisiert. Hinweis: Der Autor ist Creator des meinGPT Austria Hub, die im Unternehmen eingesetzte Plattform ist meinGPT.



7.4 Fünf Sätze für unser Lagerfeuer

K

ultur zeigt sich in Sätzen, die in einem Team selbstverständlich sind. Die folgenden fünf Sätze übersetzen Debiasing in

Alltagssprache. Wenn sie in deinem Team ohne Risiko gesagt werden können, ist der wichtigste Teil der Arbeit getan.

WAKEUP!-MOMENT

„Wie haben wir diese Zahl belastet?“

„Was müsste passieren, damit wir unsere Meinung ändern?“

„Ich glaube, ich habe mich geirrt.“

„Lass uns kurz so tun, als wäre es schon schiefgegangen.“

„Was würde jemand sagen, der nicht in diesem Raum sitzt?“

7.5 Was der Mensch beiträgt

K I wird viele Funktionen übernehmen, weil sie schneller zusammenfasst, breiter recherchiert und konsistenter bewertet. Was sie nicht übernimmt, ist das, was zwischen

Menschen entsteht, nämlich das Vertrauen, einen Zweifel auszusprechen, sowie die Verbundenheit, die einen Irrtum überlebt. Genau daraus ist der Bruch eines Bias-Kreislaufs gemacht.



Wer bist du, wenn du nicht funktionieren musst? Sein statt Funktion. Die Leitfrage im Saal beim Internationalen Forum Mechatronik in Memmingen. Foto: Bayern Innovativ GmbH

WAKEUP!-MOMENT

„Am Anfang benötigt es Begeisterung und Emotion, für Erfolg, Fokus und Disziplin!“

Für die Wahrheit braucht es zusätzlich einen Ort, an dem Widerspruch willkommen ist.

Speere wärmen nicht. Das Rudel schon. Entzünde das Lagerfeuer.



Wie viel Lagerfeuer steckt in deinem Rudel? Der T.R.U.S.T. Hero Check umfasst 28 Fragen, dauert fünf Minuten und ist kostenlos.

juergen-menhart.com/trust-hero-check/



■ ANHANG · VERTIEFUNGEN

Für alle, die es **genau wissen wollen.**

Vier Vertiefungen bündeln die Mechanismen, Fallstudien und Werkzeuge in kompakter Form, damit du sie nachschlagen, weitergeben und prüfen kannst.

A Die Bias-Landkarte

Die folgende Übersicht zeigt alle Mechanismen des Kreislaufs mit ihrem zentralen Befund, ihrer typischen Wirkung in Organisationen und dem wirksamsten Gegenmittel.

MECHANISMUS	KERNBEFUND	WIRKUNG IN ORGANISATIONEN	GEGENMITTEL
Default / Status quo	Voreinstellungen bestimmen Entscheidungen massiv (Samuelson & Zeckhauser, 1988; Johnson & Goldstein, 2003)	Geerbte Prozesse, Kennzahlen und Raster laufen unhinterfragt weiter	Herkunft jeder Standardoption prüfen
Anker	Zufällige Startwerte verschieben Schätzungen, auch bei Fachleuten (Tversky & Kahneman, 1974; Englich et al., 2006)	Erste Budgetzahl, erstes Zieldatum werden zum Maßstab	Unabhängige Schätzungen vor der Diskussion
Framing	Gleiche Fakten, anderer Rahmen, andere Wahl (Tversky & Kahneman, 1981)	Veränderung als Verlust oder Gewinn kommuniziert	Vorlage in beiden Rahmen formulieren
Confirmation Bias	Bestätigendes wird gesucht und besser bewertet (Wason, 1960; Lord et al., 1979; Nickerson, 1998)	Business Cases sammeln Belege statt Gegenbelege	Consider the opposite (Lord et al., 1984)
Motivated Reasoning	Niedrigere Beweisstandards für Gewünschtes (Kunda, 1990)	Position wird Identität, Kritik wird persönlich	Rolle und Position trennen
Dissonanz	Überzeugung passt sich dem Verhalten an (Festinger, 1957; Festinger & Carlsmith, 1959)	Öffentliche Bekenntnisse machen Umkehr teuer	Kosten des Umdenkens senken
Illusion of Truth	Wiederholung erhöht geglaubte Wahrheit (Hasher et al., 1977; Fazio et al., 2015)	Leitsätze und Foliensätze werden unhinterfragte Fakten	Herkunft wiederkehrender Aussagen prüfen
Backfire	Tritt selten auf, Korrekturen wirken meist (Wood & Porter, 2019; Swire-Thompson et al., 2020)	Falsche Annahme, dass Fakten nichts bringen	Korrekturen mit Alternative anbieten
Self-Fulfilling Prophecy	Erwartung erzeugt Verhalten, das sie bestätigt (Merton, 1948; Jussim & Harber, 2005)	Etiketten für Teams und Personen	Etiketten als Hypothesen behandeln
Groupthink	Einmütigkeit verdrängt Prüfung (Janis, 1982)	Scheinbare Einstimmigkeit im Lenkungsreis	Kritischer Prüfer, Führung hält Meinung zurück
Social Proof	Mehrheit zieht mit, ein Abweichler bricht den Bann (Asch, 1956; Cialdini, 1984)	Methoden- und Investitionsmoden	Den ersten Widerspruch legitimieren
In-Group-Bias	Bevorzugung der eigenen Gruppe (Tajfel et al., 1971); Kontakt mit gemeinsamen Zielen hilft (Pettigrew & Tropp, 2006)	Silos, Projekt gegen Tagesgeschäft	Gemeinsames Mammut, gleicher Status

Sunk Cost	Weiterinvestieren nach Fehlschlag (Arkes & Blumer, 1985; Staw, 1976)	Großprojekte, die niemand stoppt	Abbruchkriterien vorab
Survivorship / Hindsight	Lernen nur von Überlebenden; Rückschau wirkt zwangsläufig (Fischhoff, 1975)	Falsche Erfolgsrezepte, Schuld statt Lernen	Entscheidungsjournal
HiPPO	Rang schlägt Evidenz (Kohavi et al., 2020)	Meinung der Spitze als Default	Führung spricht zuletzt, kleine Tests
Automation Bias	Übervertrauen in Maschinen (Parasuraman & Manzey, 2010)	KI-Vorschläge werden ungeprüft übernommen	Eigenes Urteil vor der Empfehlung

B Fallstudien im Detail

Recruiting. Reuters berichtete 2018, dass Amazon ein experimentelles Werkzeug zur Bewertung von Bewerbungen eingestellt hatte. Das System war auf Bewerbungen aus rund zehn Jahren trainiert, die überwiegend von Männern stammten. Es lernte, Lebensläufe mit Bezug zu Frauen abzuwerten. Nach Angaben des Unternehmens wurde es nie allein für Einstellungsentscheidungen genutzt. Der Fall ist Historical Bias in Reinform: Das Modell lernte korrekt, wer früher eingestellt wurde.

Strafjustiz. ProPublica analysierte 2016 das Risikobewertungssystem COMPAS. Schwarze Angeklagte, die später nicht rückfällig wurden, waren fast doppelt so häufig fälschlich als hohes Risiko eingestuft wie weiße. Der Hersteller hielt dagegen, das System sei für beide Gruppen gleich gut kalibriert. Kleinberg, Mullainathan und Raghavan (2016) sowie Chouldechova (2017) zeigten, dass beide Seiten rechnerisch recht haben konnten. Unterscheiden sich die Grundraten zwischen Gruppen, lassen sich gleiche Kalibrierung und gleiche Fehlerraten mathematisch nicht gleichzeitig erreichen. Welche Fairness gelten soll, ist deshalb eine Wertentscheidung.

Polizeiarbeit. Lum und Isaac (2016) simulierten ein Vorhersagemodell auf Basis historischer Festnahmedaten zu Drogendelikten in Oakland. Das Modell lenkte Einsätze immer wieder in Viertel, die bereits stark kontrolliert wurden, obwohl Gesundheitsdaten auf eine breitere Verteilung des Konsums hinwiesen. Ensign und Kollegen (2018) haben diese Dynamik als „Runaway Feedback Loop“ formalisiert.

Kreditvergabe. Bartlett, Morse, Stanton und Wallace (2022) untersuchten US-Hypothekendaten. Algorithmische Kreditgeber verlangten von Kreditnehmenden aus Minderheiten höhere Zinsen, die Diskriminierung fiel aber geringer aus als bei persönlicher Beratung. Algorithmen können Bias also verringern. Sie beseitigen ihn nicht automatisch, solange sie aus Daten lernen, in denen alte Muster stecken.

C Werkzeuge für faire Algorithmen

WERKZEUG	WELCHE FRAGE ES BEANTWORTET	GRENZE
Datasheets for Datasets (Geburu et al., 2021)	Woher kommen die Daten, wer fehlt, welche Lücken sind bekannt?	Nur so gut wie die Ehrlichkeit der Dokumentation
Model Cards (Mitchell et al., 2019)	Wofür ist das Modell gedacht, wofür nicht, wie schneidet es je Gruppe ab?	Beschreibt, korrigiert aber nicht
Demografische Parität	Erhalten alle Gruppen gleich häufig ein positives Ergebnis?	Ignoriert unterschiedliche Grundraten
Equalized Odds (Hardt et al., 2016)	Sind Fehlerraten in allen Gruppen gleich?	Nicht gleichzeitig mit Kalibrierung erfüllbar, wenn Grundraten differieren
Kalibrierung	Bedeutet ein Risikowert in allen Gruppen dasselbe?	Kann ungleiche Fehlerraten verdecken
Disaggregierte Evaluation	Wie gut ist das Modell für Untergruppen und deren Schnittstellen?	Braucht ausreichend Daten je Gruppe
Laufendes Monitoring	Verschiebt das Modell durch seinen Einsatz die eigene Datenbasis?	Erkennt Drift, entscheidet nicht über Konsequenzen

D Die acht Symptome von Groupthink nach Janis

SYMPTOM	WORAN MAN ES IM MEETING ERKENNT
Illusion der Unverwundbarkeit	Risiken werden belächelt, „uns passiert das nicht“
Kollektives Wegerklären	Warnsignale werden gemeinsam kleingeredet
Glaube an die eigene Moral	Ethische Folgen werden nicht mehr diskutiert
Stereotype über Außenstehende	Kritiker gelten als ahnungslos oder böswillig
Druck auf Abweichler	Wer zweifelt, gilt als illoyal
Selbstzensur	Bedenken werden gedacht, aber nicht ausgesprochen
Illusion der Einstimmigkeit	Schweigen wird als Zustimmung gezählt
Selbsternannte Meinungswächter	Einzelne halten unbequeme Information von der Gruppe fern

Literatur

- Allport, G. W. (1954). *The Nature of Prejudice*. Cambridge, MA: Addison-Wesley.
- Arkes, H. R. & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, 35(1), 124–140.
- Asch, S. E. (1956). Studies of independence and conformity. *Psychological Monographs*, 70(9).
- Bakshy, E., Messing, S. & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132.
- Bansal, G. et al. (2021). Does the whole exceed its parts? The effect of AI explanations on complementary team performance. *Proceedings of CHI 2021*.
- Bartlett, R., Morse, A., Stanton, R. & Wallace, N. (2022). Consumer-lending discrimination in the FinTech era. *Journal of Financial Economics*, 143(1), 30–56.
- Buçinca, Z., Malaya, M. B. & Gajos, K. Z. (2021). To trust or to think. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1).
- Buolamwini, J. & Gebru, T. (2018). Gender Shades. *Proceedings of Machine Learning Research*, 81, 77–91.
- Cascio, J. (2020). Facing the Age of Chaos. *Medium*.
- Chouldechova, A. (2017). Fair prediction with disparate impact. *Big Data*, 5(2), 153–163.
- Cialdini, R. B. (1984). *Influence: The Psychology of Persuasion*. New York: Morrow.
- Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, 10.10.2018.
- Dellermann, D., Ebel, P., Söllner, M. & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering*, 61(5), 637–643.
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350–383.
- Englich, B., Mussweiler, T. & Strack, F. (2006). Playing dice with criminal sentences. *Personality and Social Psychology Bulletin*, 32(2), 188–200.
- Ensign, D. et al. (2018). Runaway feedback loops in predictive policing. *Proceedings of Machine Learning Research*, 81, 160–171.
- Europäische Union (2024). Verordnung (EU) 2024/1689 (KI-Verordnung). *Amtsblatt der EU*, L, 12.7.2024.
- Fazio, L. K., Brashier, N. M., Payne, B. K. & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, 144(5), 993–1002.
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.
- Festinger, L. & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58(2), 203–210.
- Fischhoff, B. (1975). Hindsight is not equal to foresight. *Journal of Experimental Psychology: Human Perception and Performance*, 1(3), 288–299.
- Gebru, T. et al. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92.
- Gigerenzer, G. & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451–482.
- Goldin, C. & Rouse, C. (2000). Orchestrating impartiality. *American Economic Review*, 90(4), 715–741.
- Guess, A. & Coppock, A. (2020). Does counter-attitudinal information cause backlash? *British Journal of Political Science*, 50(4), 1497–1515.
- Hagger, M. S. et al. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, 11(4), 546–573.
- Hardt, M., Price, E. & Srebro, N. (2016). Equality of opportunity in supervised learning. *NeurIPS 29*.
- Hasher, L., Goldstein, D. & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16(1), 107–112.
- Haselton, M. G. & Nettle, D. (2006). The paranoid optimist. *Personality and Social Psychology Review*, 10(1), 47–66.
- Janis, I. L. (1982). *Groupthink* (2. Aufl.). Boston: Houghton Mifflin.

- Johnson, E. J. & Goldstein, D. (2003). Do defaults save lives? *Science*, 302(5649), 1338–1339.
- Jussim, L. & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies. *Personality and Social Psychology Review*, 9(2), 131–155.
- Kahan, D. M., Peters, E., Dawson, E. C. & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, 1(1), 54–86.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kahneman, D. & Klein, G. (2009). Conditions for intuitive expertise. *American Psychologist*, 64(6), 515–526.
- Kahneman, D., Sibony, O. & Sunstein, C. R. (2021). *Noise*. New York: Little, Brown Spark.
- Kahneman, D. & Tversky, A. (1979). Prospect theory. *Econometrica*, 47(2), 263–291.
- Klayman, J. & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94(2), 211–228.
- Klein, G. (2007). Performing a project premortem. *Harvard Business Review*, 85(9), 18–19.
- Kleinberg, J., Mullainathan, S. & Raghavan, M. (2016). Inherent trade-offs in the fair determination of risk scores. *arXiv:1609.05807*.
- Kohavi, R., Tang, D. & Xu, Y. (2020). *Trustworthy Online Controlled Experiments*. Cambridge University Press.
- Kuhn, D. & Lao, J. (1996). Effects of evidence on attitudes: Is polarization the norm? *Psychological Science*, 7(2), 115–120.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498.
- Lewandowsky, S. et al. (2012). Misinformation and its correction. *Psychological Science in the Public Interest*, 13(3), 106–131.
- Lord, C. G., Lepper, M. R. & Preston, E. (1984). Considering the opposite. *Journal of Personality and Social Psychology*, 47(6), 1231–1243.
- Lord, C. G., Ross, L. & Lepper, M. R. (1979). Biased assimilation and attitude polarization. *Journal of Personality and Social Psychology*, 37(11), 2098–2109.
- Lum, K. & Isaac, W. (2016). To predict and serve? *Significance*, 13(5), 14–19.
- Merton, R. K. (1948). The self-fulfilling prophecy. *The Antioch Review*, 8(2), 193–210.
- Mitchell, D. J., Russo, J. E. & Pennington, N. (1989). Back to the future. *Journal of Behavioral Decision Making*, 2(1), 25–38.
- Mitchell, M. et al. (2019). Model cards for model reporting. *Proceedings of FAT* 2019*, 220–229.
- Mussweiler, T. & Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm. *Journal of Experimental Social Psychology*, 35(2), 136–164.
- Nemeth, C., Brown, K. & Rogers, J. (2001). Devil's advocate versus authentic dissent. *European Journal of Social Psychology*, 31(6), 707–720.
- Nickerson, R. S. (1998). Confirmation bias. *Review of General Psychology*, 2(2), 175–220.
- Nyhan, B. & Reifler, J. (2010). When corrections fail. *Political Behavior*, 32(2), 303–330.
- Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453.
- O'Neil, C. (2016). *Weapons of Math Destruction*. New York: Crown.
- Parasuraman, R. & Manzey, D. H. (2010). Complacency and bias in human use of automation. *Human Factors*, 52(3), 381–410.
- Pariser, E. (2011). *The Filter Bubble*. New York: Penguin Press.
- Pennycook, G., Cannon, T. D. & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, 147(12), 1865–1880.
- Perry, G. (2018). *The Lost Boys: Inside Muzafer Sherif's Robbers Cave Experiment*. Melbourne: Scribe.
- Pettigrew, T. F. & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5), 751–783.
- Popper, K. (1935). *Logik der Forschung*. Wien: Springer (ausgeliefert 1934).
- Raichle, M. E. & Gusnard, D. A. (2002). Appraising the brain's energy budget. *PNAS*, 99(16), 10237–10239.
- Rosenthal, R. & Jacobson, L. (1968). *Pygmalion in the Classroom*. New York: Holt, Rinehart and Winston.
- Samuelson, W. & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1), 7–59.
- Sharma, M. et al. (2023). Towards understanding sycophancy in language models. *arXiv:2310.13548*.
- Shumailov, I. et al. (2024). AI models collapse when trained on recursively generated data. *Nature*, 631, 755–759.

- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1), 99–118.
- Sommers, S. R. (2006). On racial diversity and group decision making. *Journal of Personality and Social Psychology*, 90(4), 597–612.
- Staw, B. M. (1976). Knee-deep in the big muddy. *Organizational Behavior and Human Performance*, 16(1), 27–44.
- Suresh, H. & Guttag, J. (2021). A framework for understanding sources of harm throughout the machine learning life cycle. *Proceedings of EAAMO 2021*.
- Swire-Thompson, B., DeGutis, J. & Lazer, D. (2020). Searching for the backfire effect. *Journal of Applied Research in Memory and Cognition*, 9(3), 286–299.
- Tajfel, H., Billig, M. G., Bundy, R. P. & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, 1(2), 149–178.
- Tajfel, H. & Turner, J. C. (1979). An integrative theory of intergroup conflict. In Austin & Worchel (Hrsg.), *The Social Psychology of Intergroup Relations*. Monterey: Brooks/Cole.
- Thaler, R. H. & Sunstein, C. R. (2008). *Nudge*. New Haven: Yale University Press.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.
- Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12(3), 129–140.
- Wood, T. & Porter, E. (2019). The elusive backfire effect. *Political Behavior*, 41(1), 135–163.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9(2, Pt. 2), 1–27.

Jürgen Menhart. Der Transformanist.



Jürgen Menhart ist Keynote Speaker und Autor. Er verbindet Menschen sowie Mensch, Technologie und Organisation, für Wertschöpfung im Unternehmen und für die Menschen selbst.

Menschen, die sind. Organisation und Technologie, die wirken.

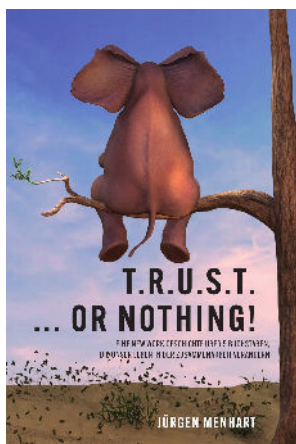
Nach 25 Jahren in Automatisierung und Industrie, vom Elektriker bis zum CEO, begleitet Jürgen Menhart heute Unternehmen in ihrer Transformation. Seine These lautet, dass der Engpass gelingender Veränderung selten in der Technologie liegt, sondern in Führung, Kultur und der Zusammenarbeit der Menschen.

Als Creator des meinGPT Austria Hub bringt er KI praxisnah in Organisationen und zeigt zugleich, warum Technologie allein keine wirksame Veränderung schafft. Er ist Autor von „WakeUP! Die 3 Schritte zur zukunftsfähigen Arbeitswelt“ und „T.R.U.S.T. ... or nothing!“. Der komplette Bucherlös von „T.R.U.S.T. ... or nothing!“ geht 1:1 an den Verein Steirer mit Herz.

KEYNOTE	WORUM ES GEHT
Speere wärmen nicht.	Die Keynote zeigt, warum High-Performance mehr braucht als perfekte Tools und warum psychologische Sicherheit die Basis jeder Veränderung ist.
Mammuts jagt man gemeinsam.	Die Keynote zeigt, warum Blindleistung der wahre Feind ist und wie aus Einzelkämpfern ein Team wird.
Die KI-Illusion.	Die Keynote beim Internationalen Forum Mechatronik 2026 zeigt, warum euer Rudel vor lauter Speeren die Jagd vergisst.

Der nächste Schritt am Lagerfeuer.

Dieses Heft zeigt, wie der Bias-Kreislauf entsteht und wo wir ihn unterbrechen. Die beiden Bücher führen weiter: das eine zum Vertrauen im Rudel, das andere zu den drei Schritten in eine zukunftsfähige Arbeitswelt.



T.R.U.S.T. ... OR NOTHING!

Vertrauen ist das einzige Betriebssystem, das im 21. Jahrhundert funktioniert. Eine wahre Geschichte über fünf Buchstaben, die Führung und Zusammenarbeit verändern.

DAS LAGERFEUER IST KEINE METAPHER.

Es ist ein Überlebenssystem. Die Fabel hinter der Keynote zur Mammutjagd zeigt, warum jedes Rudel alle fünf Charaktere braucht.

Der komplette Bucherlös geht 1:1 an den Verein Steirer mit Herz.

juergen-menhart.com/#buecher



WAKEUP! DIE 3 SCHRITTE

Der Weckruf für eine erstarrte Arbeitswelt. Drei Schritte in fester Reihenfolge führen Teams zurück ins Handeln: People fit, Project fit und Strategy fit.

DIE EISZEIT IST VORBEI.

Deine Teams wissen es vielleicht noch nicht. Dieses Buch führt sie vom Erstarren zurück ins Erwachen.

Mit einer 24-Wochen-Roadmap, mit der du sofort in die Umsetzung startest.

juergen-menhart.com/#buecher

WakeUP! Ausgabe 01, Oktober 2026. Angaben zum EU AI Act geben den Stand vom Oktober 2026 wieder. Die Fallstudien im Anhang beziehen sich auf veröffentlichte Forschung und Berichterstattung, der Praxisfall in Teil VII ist anonymisiert.

Jürgen Menhart · Keynote Speaker, Transformant und Autor · Original Human · Graz, Österreich · juergen-menhart.com · office@juergen-menhart.com

Speere wärmen nicht. Das Rudel schon.

„Wer bist du, wenn du nicht funktionieren musst?“
Sein statt Funktion.



Wo Widerspruch nichts kostet, hat Wahrheit eine Chance.



T.R.U.S.T. HERO CHECK

Wie viel Lagerfeuer steckt in deinem Rudel?

Der Check umfasst 28 Fragen, dauert fünf Minuten und ist kostenlos.

juergen-menhart.com/trust-hero-check/

Jürgen Menhart

Keynote Speaker, Transformant und Autor

ORIGINAL HUMAN

juergen-menhart.com · office@juergen-menhart.com