OrderFlowAi

WISSENSCHAFTLICHER FORSCHUNGSBERICHT · ÖFFENTLICHE AUSGABE

Die Suche nach dem Edge

Wie man einen statistischen Vorteil an den Finanzmärkten beweist — und warum fast niemand es versucht

Ein Bericht über 471 Arbeitssitzungen, 54 selbst widerrufene Befunde und den Aufbau eines Beweisapparats, der ein Ergebnis **zertifizierbar** machen soll — bevor es überhaupt existiert. Vom ersten Alpha-Build bis zu **CORTEX-ONE**.

471	54	350	0
ARBEITSSITZUNGEN	SELBST WIDERRUFENE BEFUNDE	WERKZEUGE MIT SELBSTTEST	UNBELEGTE EDGE-BEHAUPTUNGEN

OrderFlowAi — OrderFlowAi Research
orderflowai.io/forschung

Stand: 23. September 2026 · Session 471
Öffentliche Ausgabe — Weitergabe erwünscht

INHALTSVERZEICHNIS

Was in diesem Bericht steht

Lesehinweis. Dieser Bericht behauptet **nicht**, dass ein Edge gefunden wurde. Er dokumentiert, wie systematisch danach gesucht wird — und warum genau diese Systematik der eigentliche Vermögenswert ist. Jede Zahl darin ist gemessen, jede widerlegte Zahl ist als widerlegt gekennzeichnet. Wer nur eine Seite liest, lese Seite 3. p = Zufallswahrscheinlichkeit (kleiner ist besser, Schwelle 0,05) · n = Stichprobengröße.

Kurzwege: Investoren · S. 3, 24–33, 36–43 | Techniker · S. 13–22 | Prüfer · S. 15–19, 34–35 | Neugierige · S. 4–14

Woher die Zahlen stammen

QUELLE	WAS DARAUS STAMMT
Projektarchiv 925 datierte Commits	Chronik, Meilensteine, Phasenzuordnung — jeder Eintrag datiert.
Widerrufsregister 54 Einträge, maschinell geprüft	Alle als falsch gekennzeichneten Aussagen samt Ursache und Kosten.
Auswertungswerkzeuge 350 Programme mit Selbsttest, 4.829 Einzeltests	Sämtliche statistischen Kennzahlen — p-Werte, Konfidenzintervalle, Zerfallskurven.
Handelsprotokolle **598 Trades / 57 Tage (Juni–September)**	**Alle Ergebniszahlen. Simulationskonto, vollständig aufgezeichnet.**

ZUSAMMENFASSUNG FÜR EILIGE

Wir haben den Edge nicht gefunden. Wir haben etwas Selteneres gebaut: die Maschine, die ihn beweisen könnte.

In der Trading-Software-Branche behaupten hunderte Anbieter einen „Vorteil". Kaum einer kann ihn belegen — weil dafür ein Apparat nötig ist, den fast niemand baut: Vorregistrierung, Placebo-Kontrollen, Nullhypothesen, ein Register der eigenen Irrtümer. **Diesen Apparat gibt es bei OrderFlowAi seit Juli 2026 vollständig.** Er ist das Ergebnis von zwölf Monaten Arbeit und — das ist der ungewöhnliche Teil — von 54 Befunden, die wir selbst gekippt haben, statt sie zu verkaufen.

DAS PRODUKT

Vier Software-Editionen laufen in Produktion, zahlende Kunden, offizielles NinjaTrader-Listing. **Der Umsatz hängt nicht am Edge** — er hängt an Visualisierung, Orderbuch-Analyse und Journal. Das ist bewusst so entkoppelt.

DIE FORSCHUNG

CORTEX — eine selbstlernende neuronale Engine mit 49 Merkmalen und fünf Vorhersageköpfen, seit dem 23. September mit frischem Lernstand. Letzter, mehrfach abgesicherter Befund: **kein nachweisbarer Vorteil**, global und in jedem einzelnen Marktregime. Das ist kein Rückschlag. Das ist ein Messergebnis.

DER VERMÖGENSWERT

Der Beweisapparat. **350 Analysewerkzeuge** mit eingebauten Selbsttests (4.829 Einzelprüfungen), eine vorregistrierte Prüfkette in neun Stufen und ein öffentlich geführtes Widerrufsregister. Wer damit einen Edge findet, kann ihn **zertifizieren lassen**.

Die fünf Kernaussagen

1. **Ehrlichkeit ist hier kein Wert, sondern eine Methode.**
   54-mal haben wir einen eigenen Befund widerrufen — darunter Ergebnisse, die bereits als Erfolg gemeldet waren. Jeder Widerruf steht mit Ursache und Kosten in einem maschinell geprüften Register. Ein System, das seine eigenen Fehler findet, ist die Voraussetzung für jede spätere Zertifizierung.

2. **Negative Ergebnisse sind teuer erkauftes Wissen.**
   Dass das Regime-Label keine Richtungsinformation trägt, dass der Microprice nur 90 Millisekunden vorausläuft, dass die Größen-Evidenz gegen ihr eigenes Placebo 0,2 Prozentpunkte gewinnt — jede dieser Erkenntnisse hat Wochen gekostet und spart jedem Nachfolger dieselben Wochen.

3. **Wir haben aufgehört, an Schwellen zu drehen.**
   Der teuerste Fehler dieser Branche ist, Parameter so lange zu justieren, bis die Rückschau gut aussieht. Seit Juli 2026 steht die Regel *vor* der Rechnung — vorregistriert, versioniert, im Commit datiert.

4. **Wir verfolgen Spuren bis zum Ende — auch wenn es negativ endet.**
   Die Juli-Spur blieb unentschieden, ihr Endpunkt wurde stillgelegt. Im September fielen drei Bilder gegen ihre Nulllinie; der Kandidat Flush-Absorption traf 65,6 %, verlor aber 5,17 $ je Trade und ist nach Regel verworfen. Tickgenaue Replays zeigen: die Einstiege tragen keine Kante — auch nicht umgekehrt. Was steht: eine Schatten-Sperre mit starkem, aber explorativem Signal (`p = 0,0001`) und eine gegen die Börse geprüfte Messbasis (97 %).
   Also: ehrlich negativ, nicht aufgegeben.

5. **Die Entwicklungsgeschwindigkeit wächst mit jedem Modell-Update.**
   Das Projekt wird von einem Gründer und einem KI-Agenten im „autonomen Mitgründer-Modus" gebaut. Was 2025 Tage brauchte, braucht heute Stunden — und die nächste Modellgeneration verschiebt die Grenze erneut. Details in Teil VI.

> **Der Investorensatz in einem Satz:** Sie investieren nicht in die Behauptung eines Edges — Sie investieren in das einzige Labor der Branche, das einen Edge *widerlegen* kann, und deshalb als einziges einen beweisen könnte.

GRUNDLAGEN

Was ein „Edge" eigentlich ist

Ein Edge ist kein Gefühl, keine Trefferquote und kein guter Monat. Ein Edge ist eine **messbare, wiederholbare Abweichung vom Zufall**, die auch dann bestehen bleibt, wenn man ernsthaft versucht, sie zu widerlegen.

Die vier Bedingungen

Damit man von einem Edge sprechen darf, müssen **alle vier** erfüllt sein:

- **A** **Ein Vergleichsmaßstab existiert.**
  Man muss wissen, was *Zufall* in genau dieser Situation liefern würde. Diese Zahl ist fast nie 50 %.
- **B** **Die Abweichung ist groß genug für die Stichprobe.**
  70 % Treffer aus 7 Trades sind bedeutungslos. Dieselbe Quote aus 400 Trades ist ein Befund.
- **C** **Sie hält außerhalb der Daten, mit denen sie gefunden wurde.**
  Alles andere ist Auswendiglernen.
- **D** **Sie überlebt einen ernsthaften Widerlegungsversuch.**
  Placebo-Kontrolle, vertauschte Reihenfolge, verschobene Zeitachse.

Warum die Nulllinie fast nie 50 % ist

Ein Beispiel aus unserem eigenen Widerrufsregister: Wir maßen bei einem Vorhersagekopf eine Trefferquote und verglichen sie mit 50 %. Ergebnis: „32 Punkte unter Zufall" — Alarm.

Tatsächlich lag die richtige Nulllinie je nach Tag zwischen **2 % und 20 %**, weil ein großer Teil der Ausgänge weder „auf" noch „ab" war, sondern seitwärts. Gegen die *richtige* Linie gerechnet lag der Kopf praktisch **genau auf Zufall**.

Widerrufen · S392

Der teuerste Satz der Branche

„Die Trefferquote liegt bei 75 %."
Diese Aussage ist wertlos, solange nicht dabeisteht: **75 % wovon**, **über wie viele unabhängige Beobachtungen**, **gegen welche Nulllinie** und **in welchem Zeitfenster gemessen**. Ohne diese vier Angaben ist es Marketing.

Der Unterschied zwischen Beschreiben und Vorhersagen

Der wichtigste und am häufigsten übersehene Unterschied. Ein Modell kann eine Marktbewegung perfekt *beschreiben* und trotzdem völlig wertlos sein — weil es die Bewegung erst beschreibt, wenn sie schon passiert.

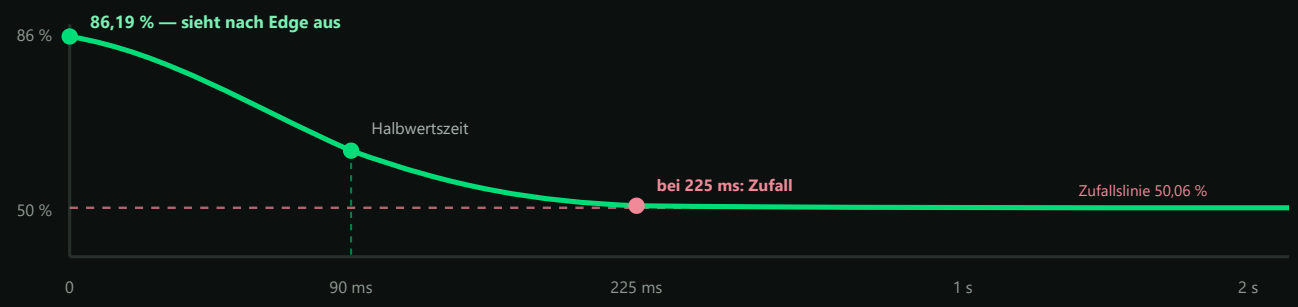


Abbildung 1: Die Zerfallskurve des Microprice-Signals, gemessen an 936.000 bereinigten Einzelabschlüssen (27.07.2026). Die 86 % Trefferquote sind echt — sie beschreiben aber, was gerade geschieht, und sagen nach 90 Millisekunden nichts mehr voraus. Ohne diese Kurve wäre daraus ein „Edge" geworden.

Regel, die daraus wurde: Eine Kennzahl ohne Zerfallskurve ist unvollständig — genauso wie eine Trefferquote ohne Nullhypothese. Beide sind seit Juli 2026 Pflichtspalten in jedem Auswertungswerkzeug.

FEHLERLEHRE

Die fünf Fallen — und wie wir in jede einzelne getappt sind

Diese Liste ist keine Theorie. Jede Falle hat uns Zeit gekostet, jede steht mit Datum und Kosten im Widerrufsregister. Wir zeigen sie, weil ein Prüfer genau danach fragen wird.

Falle 1 — Die Nulllinie raten statt rechnen

2× getappt

Man vergleicht ein Ergebnis mit 50 %, weil das intuitiv nach „Zufall" klingt. Die tatsächliche Zufallslinie hängt aber von der Verteilung der möglichen Ausgänge ab — bei uns lag sie zwischen 2 % und 20 %. **Folge:** Ein Fehlalarm („32 Punkte unter Zufall") und, schlimmer, ein Fortschrittsindikator, der ein Jahr lang Lernfortschritt meldete, wo in Wahrheit nur der Anteil der Seitwärtsausgänge sank.

Falle 2 — Die Abdeckungsquote ohne Gegenprobe

Kosten: ein Abend + ein Konzept

Eine Regel traf in **75,4 %** der Fälle zu. Das wurde als „erstes Werkzeug, das seine vorregistrierte Hürde genommen hat" gemeldet. Die Placebo-Kontrolle — dieselbe Regel angewandt auf einen **50.000 Abschlüsse entfernten**, garantiert unbeteiligten Datenpunkt — traf **75,2 %**. Informationsgehalt: **0,21 Prozentpunkte.** Die Regel maß nicht den Markt, sie maß ihre eigene Bauart.

Falle 3 — Log-Zeilen für Beobachtungen halten

3× getappt

Ein Ereignis erzeugt zehn Protokollzeilen. Zählt man Zeilen statt Ereignisse, wirkt eine Stichprobe zehnmal größer als sie ist — und jeder Signifikanztest wird falsch. Bei uns: **311 Zeilen = 13 Marktmomente**. Nach der Korrektur stieg der p-Wert von 0,005 auf 0,119; aus einem „Befund" wurde „kein Urteil". Dieselbe Falle in anderer Verkleidung: doppelt aufgezeichnete Abschlüsse, weil zwei Programminstanzen in dieselbe Datei schrieben (**44,5 % exakte Duplikate**).

Falle 4 — Zwei Zahlen mit verschiedenem Ausschnitt vergleichen

5× getappt

Zwei je für sich korrekte Messungen, die aber unterschiedliche Zeitfenster, Konten oder Abtastraster verwenden. Das Ergebnis sieht aus wie ein Widerspruch im Markt und ist ein Defekt in der Anzeige. Prominentester Fall: eine Delta-Kennzahl mit **+13.225** im einen Fenster und **−16.816** im anderen — kein Vorzeichenfehler, sondern zwei verschiedene Sessiondefinitionen.

Falle 5 — Sich gegen die eigene Rechnung verifizieren

zuletzt: 29.07.2026

Man prüft eine Umrechnung, indem man ihr Ergebnis mit einer Zahl vergleicht, die aus derselben Umrechnung stammt. Der Fehler bleibt unsichtbar. Bei uns: eine Zeitzonenverschiebung um zwei Stunden, die eine ganze Auswertung entwertete. Gefunden hat sie nicht die Ergebniskontrolle, sondern eine **fachliche** Frage: „Berührt der rekonstruierte Kursverlauf eines Vollverlusts überhaupt seinen Stop-Kurs?" Antwort damals: nur in 35 % der Fälle. Nach dem Fix: 100 %.

Was diese fünf Fallen gemeinsam haben: Keine davon ist ein Programmierfehler. Alle fünf sind *Messfehler* — der Code lief korrekt und maß die falsche Größe. Deshalb reicht Software-Qualitätssicherung nicht aus; es braucht einen zweiten, unabhängigen Prüfpfad. Genau den beschreibt Teil III.

SYSTEMATIK

Die Beweisleiter — sechs Stufen von „interessant" zu „zertifizierbar"

Jede Behauptung über einen Vorteil lässt sich auf dieser Leiter einordnen. Die Branche verkauft üblicherweise auf Stufe 1 und 2. Ein Zertifikat setzt Stufe 6 voraus.

Abbildung 2: Die Beweisleiter. Der entscheidende Sprung ist der von Stufe 4 auf 5 — er kostet nichts außer Disziplin und ist trotzdem der seltenste Schritt der Branche.

Wo wir stehen

Der **Apparat** steht auf Stufe 5 und ist auf Stufe 6 vorbereitet: Rohdaten werden archiviert, jedes Werkzeug prüft sich selbst, die Vorregistrierung ist versioniert und maschinell erzwungen. Das **Ergebnis** steht bei „kein Vorteil nachweisbar" — was ein sauberes Resultat auf Stufe 5 ist, nur eben ein negatives.

Warum das mehr wert ist als es klingt

Ein Apparat, der auf Stufe 5 ein *negatives* Ergebnis produziert, ist bewiesen funktionsfähig. Ein Apparat, der nur positive Ergebnisse produziert, ist verdächtig. Der erste ist ein Messgerät, der zweite ein Verkaufsprospekt.

PHASE 0 · AUGUST 2025 – FEBRUAR 2026

Die ersten Zeilen: ein Werkzeug, kein Orakel

Am **1. August 2025** läuft das erste interne Alpha. Es sagt nichts vorher. Es zeigt etwas an — und zwar etwas, das die meisten Trading-Programme gar nicht darstellen können: das **Orderbuch**, also die Liste aller offenen Kauf- und Verkaufsabsichten, live und in voller Tiefe.

Was in dieser Phase gebaut wurde

01.08.2025 · V0.01.1
Internes Alpha
Erste Darstellung von Markttiefe als Wärmebild. Reines Anzeigewerkzeug.

HERBST 2025
Anbindung an NinjaTrader 8
Die Handelsplattform liefert die Daten, die eigene Software rechnet und zeigt. Diese Trennung — **fremde Ausführung, eigene Analyse** — bleibt bis heute die Architektur.

01.01.2026 · V0.01.6
Spoofing-Erkennung
Erste *Interpretation* statt reiner Anzeige: Aufträge, die nur zum Schein im Buch stehen und vor Ausführung verschwinden, werden markiert.

01.02.2026 · V0.01.7
Sitzungsfilter & Export
Ab hier entstehen auswertbare Aufzeichnungen — die Grundlage für alles Spätere.

03.03.2026
4-Phasen Reversal-Scanner
Der erste Versuch, Wendepunkte zu klassifizieren. Regelbasiert, noch ohne Lernen.

Warum Level-2-Daten der Ausgangspunkt sind

Ein normaler Kurschart zeigt, *was* passiert ist. Das Orderbuch zeigt, *wer gerade bereit ist zu handeln* — und zu welchem Preis. Es ist die einzige Datenquelle im Privatkundenbereich, die Absicht vor Ausführung sichtbar macht.

Genau deshalb ist es auch die einzige Quelle, in der überhaupt ein Vorteil verborgen sein *könnte*. Wer nur Kerzencharts auswertet, wertet öffentlich verfügbare Vergangenheit aus.

Die Architekturentscheidung, die alles ermöglichte

Die Software wurde von Anfang an als **Beobachter** gebaut, nicht als Handelsplattform. Sie liest den Datenstrom, rechnet parallel und schreibt alles mit.

Der Nebeneffekt zeigte sich erst ein Jahr später: Weil **jede** Sitzung vollständig archiviert wurde, konnten 2026 Analysen auf Daten von 2025 gerechnet werden, die damals niemand geplant hatte. **Ohne dieses Archiv gäbe es keinen Forschungsbericht.**

Bestand nach Phase 0

1	0	~6 Mon.
PRODUKT	VORHERSAGEN	ARCHIV

Was diese Datenquelle von einem Kurschart unterscheidet

EBENE	WAS SIE ZEIGT	WARUM EIN VORTEIL DORT VERBORGEN SEIN KÖNNTE
Kerzenchart Standard, überall verfügbar	Abgeschlossene Kursbewegung in festen Zeitscheiben.	**Praktisch nicht.** Öffentlich, verzögert, von Millionen Marktteilnehmern gleichzeitig ausgewertet.
Abschlussband verbreitet	Jeder Handel mit Preis, Menge, Zeitstempel.	**Begrenzt.** Zeigt Ausführung, nicht Absicht — wer aggressiv war, muss geschätzt werden.
Orderbuch (Level 2) **Grundlage dieses Projekts**	**Alle offenen Kauf- und Verkaufsabsichten in voller Tiefe, laufend aktualisiert.**	**Ja, potenziell. Die einzige für Privatanwender zugängliche Ebene, auf der Absicht vor Ausführung sichtbar wird — inklusive vorgetäuschter Absicht.**

Rückblickend die wichtigste Entscheidung dieser Phase: alles mitzuschreiben, obwohl es zu diesem Zeitpunkt keinen Verwendungszweck gab. Datenarchive kann man nicht rückwirkend anlegen — und der Wert eines Marktdatensatzes wächst mit jedem Tag, den man ihn nicht angelegt hat.

PHASE 1 · MÄRZ 2026

Vom Anzeigen zum Vorhersagen — und die erste Versuchung

Im März 2026 wird aus dem Anzeigewerkzeug ein lernendes System. In vier Wochen entstehen die Bausteine, die bis heute tragen — und gleichzeitig entsteht der Fehler, der später ein Vierteljahr Korrektur kosten wird.

11.03.2026
Überarbeitung der KI/ML-Engine
Erste ernsthafte Modellarchitektur. Ziel: aus dem Orderbuchzustand die nächste Kursrichtung schätzen.

11.03.2026 · V0.02.3
Multi-Timeframe-Trend-Engine
Mehrere Zeitebenen gleichzeitig — mehr Kontext für dasselbe Modell.

13.03.2026
Marktgeschwindigkeit als Eingangsgröße
Wie viele Abschlüsse pro Sekunde? Ein einfacher, aber wirkungsvoller Kontextwert — derselbe Sensor, dessen stiller Ausfall im Juli 2026 monatelang unentdeckt bleiben wird (Seite 13).

17.03.2026
Institutional-Edition & GEX-Engine
Optionsbezogene Positionierungsdaten als zusätzliche Informationsebene.

26.03.2026
„Prometheus" Neural Engine
Die erste eigene neuronale Engine bekommt einen Namen. Zwei Tage später wird sie in **CORTEX** umbenannt — der Name, der bleibt.

Die erste Versuchung: 70–85 %

In dieser Phase entstehen Zahlen, die später als Erinnerung weiterleben: Trefferquoten von 70 bis 85 % über einzelne Handelstage. Sie waren nicht erfunden — sie waren nur nicht das, wofür man sie hielt.

Die Nachprüfung im Juli 2026 ergab: Die Belege bestanden aus **4 bis 9 Trades pro Tag**, gemessen in einem überwachten Zeitfenster rund um die New Yorker Eröffnung. Die Formulierung „70–85 %" taucht im Archiv mehrfach als **Ziel** auf, nicht als Messung. Eine Zeile lautet wörtlich: `10 trades 80 % WR +25 $` — vier von fünf Trades gewonnen, Ertrag: 25 Dollar.

Wie eine Erinnerung zur Kennzahl wird

Niemand hat gelogen. Es passiert schleichend:
1. Ein guter Tag wird notiert.
2. Die Notiz wird zitiert.
3. Das Zitat verliert die Stichprobengröße.
4. Aus „an einem Tag, 7 Trades" wird „die Trefferquote".

Gegenmittel, seit Juli 2026 in Kraft: Jede Quote wird nur zusammen mit Stichprobengröße, Nulllinie und Zeitfenster geschrieben. Werkzeuge, die das nicht ausgeben, gelten als defekt.

Der eigentliche Fortschritt dieser Phase war nicht das Modell, sondern die Infrastruktur: ab jetzt existierte ein durchgehender Pfad von der Handelsplattform über die Analyse bis in ein auswertbares Protokoll. Diese Kette ist die Voraussetzung dafür, dass ein Ergebnis später überhaupt *nachvollziehbar* sein kann.

PHASE 2 · 28. MÄRZ – 17. APRIL 2026

CORTEX wird geboren

Am 28. März 2026 bekommt die neuronale Engine ihren endgültigen Namen. Am 17. April geht sie als **CORTEX v0.3.0** an Kunden. Zwischen beiden Daten liegen drei Wochen, in denen aus einem Modell ein Produkt wird.

Was CORTEX technisch ist

Ein rekurrentes neuronales Netz (LSTM), das aus einer Folge von Momentaufnahmen des Marktes lernt. Jede Momentaufnahme besteht aus **41 Merkmalen** — Orderbuch-Ungleichgewichte, Absorptionsereignisse, Handelsgeschwindigkeit, Abstand zum Volumenschwerpunkt, Optionsdruck und weitere.

Das Modell läuft **lokal auf dem Rechner des Nutzers**, lernt aus dessen eigenen Marktsitzungen weiter und schickt keine Daten fort. Das ist ungewöhnlich — und es ist einer der Gründe, warum die Forschung überhaupt sauber möglich ist: Jede Instanz ist ein eigenständiges Experiment.

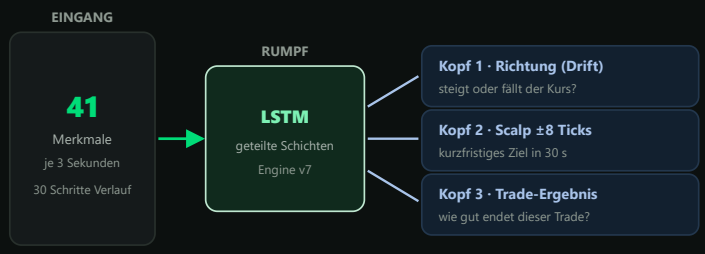


Abbildung 3: CORTEX-Architektur. Ein gemeinsamer Rumpf, drei spezialisierte Ausgänge. Der dritte Kopf wurde im Juli 2026 als Erster vollständig widerlegt (Seite 17).

Die Zahl, die alles kostet

```
NUM_FEATURES = 41
ENGINE_VERSION = 7
```

Ändert man die Anzahl oder Bedeutung der Merkmale, passt das gelernte Modell nicht mehr — **es muss vollständig gelöscht werden**. Der Fachbegriff im Projekt: ein *Wipe*.

Zwischen März und Juli 2026 geschah das mehrfach. Jedes Mal war das Modell wieder „null Sitzungen alt". **Deshalb wurde CORTEX nie besser** — nicht weil es nicht lernen konnte, sondern weil es nie lange genug durfte.

Die Erkenntnis daraus

Ein lernendes System braucht Stationarität. Man kann nicht gleichzeitig testen und lernen: Jeder Umbau am Eingang setzt den Lernstand zurück, jede Lernphase verbietet Umbauten.

Formuliert am 20. Juli 2026, nach vier Monaten, in denen beides gleichzeitig versucht wurde.

Woraus die 41 Merkmale bestehen

Buchstruktur
Ungleichgewicht zwischen Kauf- und Verkaufsseite, Tiefenverteilung, Wände, vorgetäuschte Aufträge

Fluss
Handelsgeschwindigkeit, kumulierte Differenz, Absorptionsereignisse, große Einzelabschlüsse

Lage
Abstand zum Volumenschwerpunkt, zur Eröffnung, zu Vortagesmarken, Tagesphase

Kontext
Optionsbezogener Positionierungsdruck, Marktphase, Sitzungsabschnitt

Jedes Merkmal wird alle drei Sekunden erhoben; das Modell sieht jeweils die letzten **30 Schritte**, also etwa eineinhalb Minuten Vorgeschichte. Die Auswahl war fachlich begründet — welche davon tatsächlich Information trugen, wurde erst im Juli 2026 gemessen (Ergebnis: der Zustandsvektor als Ganzes ist überangepasst, siehe Seite 17).

17. April 2026 — CORTEX-Launch v0.3.0. Pro-, Trial- und Starter-Edition gehen mit der neuronalen Engine live. Zu diesem Zeitpunkt gibt es **keine** Messung, die einen Vorteil belegt — es gibt eine funktionierende Engine und die begründete Hoffnung, dass sich einer zeigt. Der Unterschied zwischen beidem wird erst drei Monate später schmerzhaft klar.

PHASE 3 · APRIL – JUNI 2026

Die Euphorie — und die ersten Zahlen, die widersprachen

Zwei Monate lang läuft alles gleichzeitig: Produktausbau, Marketing, Kundengewinnung, Strategieentwicklung. Es ist die produktivste und zugleich unwissenschaftlichste Phase des Projekts.

Was gelang

05.05.2026
„NY Open Daily" — 11 Gewinntage, 1 Verlusttag
Eine öffentlich dokumentierte Serie. Tag 12 wird zum ersten Verlusttag, die Serie endet mit −700 $.

07.05.2026
CORTEX-OEM v0.5.0 — autonome Order-Engine
Die KI übernimmt Einstieg, Ausstieg und Positionsverwaltung. 26 Trades im Simulationskonto, +600 $, 69 % Trefferquote.

19.05.2026
Live 75 % Trefferquote
Wieder ein sehr guter Tag. Wieder eine kleine Stichprobe.

27.06.2026
Offizielles NinjaTrader-Listing
Aufnahme ins Vendor-Ökosystem — externe Bestätigung der Produktreife.

Was gleichzeitig sichtbar wurde

15.05.2026
11 Trades, 3 Gewinne, 8 Verluste
27,3 % Trefferquote, −747,82 $. Ein einzelner Tag — aber der erste, der in dieselbe Richtung zeigt wie die spätere Gesamtauswertung.

30.05.2026
Strategie-Auswertung: NO-GO
Die eigene Trade-Analyse verweigert die Freigabe der neuen Version.

29.06.2026
Ein Sicherheitsmechanismus sperrt den besten Tag
Zwei frühe Verluste lösen eine Tagessperre aus; der anschließende 22-Punkte-Aufwärtstrend wird komplett verpasst. **Ein Schutz, der mehr kostet als er schützt.**

Das Muster hinter allen drei Fällen

Gute Tage wurden dokumentiert und geteilt, schlechte Tage wurden analysiert und *behoben*. Das klingt vernünftig — führt aber dazu, dass die Gesamtbilanz nie entsteht. Sie wurde erst am **29. Juli 2026** gerechnet, auf Nachfrage, über alle **422 Trades aus 23 Tagen**.

Die Gesamtbilanz, die zwei Monate zu spät kam

KENNZAHL	GEMESSEN	BEDEUTUNG
Trefferquote gesamt	42,7 %	Break-even läge bei 49,0 % (Chance-Risiko-Verhältnis 1,04)
Ergebnis je Trade	−23,53 $	über 422 Trades
Kumuliert	−9.930,88 $	Simulationskonto, 23 Handelstage
Trades mit erreichtem ersten Kursziel	100 % Gewinn	n = 100 · durchschnittlich +320 $
Trades ohne erreichtes erstes Kursziel	25 % Gewinn	n = 322 · durchschnittlich −130 $
Verlierer, die binnen 30 Sekunden tot sind	40 %	**Es ist kein Richtungsproblem. Es ist ein Timing-Problem.**

Diese Tabelle ist der Wendepunkt des gesamten Projekts. Sie sagt: Wenn ein Trade die erste halbe Minute überlebt und sein erstes Ziel erreicht, gewinnt er *immer*. Damit verschiebt sich die Forschungsfrage von „In welche Richtung geht der Markt?" zu **„Ist dieser Moment ein guter Einstiegszeitpunkt?"** — eine Frage, die deutlich kleiner, konkreter und beweisbarer ist.

PHASE 4 · 6. – 20. JULI 2026

Der Bruch: vierzehn Tage, in denen fast jede Hoffnung fiel

Im Juli 2026 wird zum ersten Mal **ausreichend stark getestet** — mit genug Daten, mit korrekten Nulllinien, mit kreuzvalidierten Verfahren. Das Ergebnis ist eindeutig und unangenehm. Es ist zugleich der Moment, in dem aus einem Softwareprojekt ein Forschungsprojekt wird.

06.07.2026 · SITZUNG 337

CORTEX v5 — Verdikt: KEIN EDGE

Die erste vollständige Auswertung der neuronalen Engine gegen einen korrekten Maßstab. Kein nachweisbarer Vorteil.

07.07.2026 · SITZUNG 340

Offline-Lernbarkeitsstudie: kein lernbarer Vorteil — und unterpowert

Bemerkenswert ist der zweite Teil: Die Studie stellte **selbst fest**, dass ihre Datenmenge für ein sicheres Urteil noch nicht reichte. Ein Werkzeug, das die Grenzen seiner eigenen Aussagekraft meldet, ist der Anfang von seriöser Messung.

15.07.2026 · SITZUNG 358

Erstes ausreichend starkes Verdikt: KEIN EDGE (global)

Diesmal mit voller statistischer Aussagekraft. Kennzahl: Matthews-Korrelation `MCC ≈ 0,02` — praktisch Zufall.

15.07.2026 · SITZUNG 359

Kein Vorteil in *keinem* Marktregime

Die Rettungshypothese lautete: „Der Vorteil existiert, aber nur in bestimmten Marktphasen — im Mischtopf mittelt er sich weg." Geprüft über **26.940 Beobachtungen**: Seitwärtsmarkt `p = 0,29` · Aufwärtstrend `p = 0,30` · Abwärtstrend `p = 0,80`. **Falsifiziert.**

16.07.2026 · SITZUNG 362

Tick-genaue Prüfung: NICHT LERNBAR

Die letzte technische Ausrede — „die Daten sind zu grob aufgelöst" — fällt. Auch mit exakter Auflösung ist das Ziel aus diesen Merkmalen nicht lernbar.

20.07.2026 · SITZUNG 368

Das Regime-Label trägt keine Richtungsinformation

15 von 16 geprüften Zellen liegen unter der Grundlinie, in etwa **19 %** der Fälle zeigt das Label sogar in die falsche Richtung. Damit endet auch das Nachjustieren von Schwellenwerten: **Man kann ein uninformatives Signal nicht in Form tunen.**

20.07.2026 · SITZUNG 369

Strategiewechsel — Produkt und Forschung werden entkoppelt

Die zentrale unternehmerische Konsequenz. Das Produkt verkauft ab sofort das, was es nachweislich kann: Visualisierung, Orderbuchanalyse, Journal, Wärmebild. Der Edge-Nachweis läuft als **getrennter Forschungszweig** weiter — ohne Verkaufsdruck, ohne Terminzwang, ohne Versuchung.

Was in diesen 14 Tagen fiel

- Der globale Vorteil der neuronalen Engine
- Der regimebedingte Vorteil
- Die Lernbarkeit aus dem 41-Merkmale-Zustand
- Die Richtungsinformation des Regime-Labels
- Die Hoffnung, dass feinere Daten es lösen
- Vier Monate Schwellenwert-Kalibrierung

Was in diesen 14 Tagen entstand

- Ein Messapparat, der **Nein** sagen kann
- Die Trennung von Produkt und Forschung
- Das Prinzip: kein Urteil vor der Gegenprüfung
- Das Widerrufsregister als Pflichtinstanz
- Die Einsicht, dass Lernsysteme Stationarität brauchen
- Eine Forschungsfrage, die klein genug ist zum Beweisen

Warum dieser Abschnitt in einem Investorenbericht steht: Weil ein Team, das seine eigene Kernhypothese in vierzehn Tagen sechsmal widerlegt und daraufhin das Geschäftsmodell umbaut, genau das Verhalten zeigt, das man von einem Forschungsunternehmen erwarten muss. Die Alternative — weitersuchen, bis irgendeine Zahl passt — hätte ein Jahr länger gedauert und wäre bei der ersten externen Prüfung zerfallen.

PHASE 5 · 21. JULI 2026

CORTEX-ONE — ein Gehirn statt vieler Meinungen

Nach dem Bruch folgt der Neuaufbau. Die Diagnose lautete nicht „das Modell ist zu schwach", sondern: **es gab zu viele voneinander abweichende Wahrheiten im System.**

Das Problem: verteilte Wahrheit

Bis Juli 2026 berechneten mehrere Programmteile *dasselbe* Signal jeweils selbst — das Hauptfenster, das Wärmebild-Fenster, die Handelsstrategie. Weil jeder Teil leicht andere Zeitfenster oder Zwischenspeicher nutzte, konnten drei Anzeigen für dieselbe Sekunde drei verschiedene Richtungen zeigen.

Für eine Anzeige ist das ein Schönheitsfehler. Für eine **Messung** ist es tödlich: Man weiß nicht mehr, welche Zahl das Modell tatsächlich gesehen hat.

Die Lösung: ein Signal-Bus

Ein Produzent erzeugt das Signal, alle anderen konsumieren es identisch. Zusätzlich zwei feste, klar getrennte Zeithorizonte statt beliebig vieler Varianten:

SCALP
±8 Ticks binnen 30 Sekunden — kurzfristiges Ziel

SWING
größere Bewegung, längerer Horizont

Das eingefrorene Fundament

Am 21. Juli 2026 wird Version `v0.4.80` als **Lern-Grundlage festgenagelt** — mit einer Markierung im Versionsverlauf, die sich nicht mehr verschieben lässt.

Ab diesem Punkt gilt: **keine Änderung am Merkmalsvektor.** Kein neues Merkmal, keine neue Engine-Version, kein Löschen des Lernstands. Das Modell darf endlich über Wochen altern.

41 Merkmale · Engine v7 · eingefroren

Das vorregistrierte Freigabeprotokoll

Am selben Tag entsteht das **CORTEX Gating-Protokoll**. Es legt schriftlich fest, *bevor* irgendein Ergebnis vorliegt:
· ab wann ein Signal als tragfähig gilt
· welche statistische Hürde gilt (untere Grenze des Konfidenzintervalls, nicht der Punktwert)
· über wie viele Sitzungen es halten muss
· **drei Abbruchkriterien**, bei denen die Suche eingestellt wird

Und ausdrücklich, was **verboten** ist: Schwellen ändern, nachdem man ein Ergebnis gesehen hat.

Abbildung 4: CORTEX-ONE. Der Umbau brachte keine bessere Vorhersage — er brachte **Nachvollziehbarkeit**. Ohne sie ist keine Zertifizierung denkbar.

Warum das die eigentliche Voraussetzung für ein Zertifikat ist: Ein Prüfer wird fragen „Welchen Wert hat das Modell in diesem Moment gesehen?". Vor CORTEX-ONE gab es auf diese Frage drei Antworten. Seit CORTEX-ONE gibt es eine — und sie ist im Protokoll nachlesbar.

PHASE 6 · 28. – 30. JULI 2026

Vom Spiegel zum Messgerät

Die letzte und bislang wichtigste Diagnose: **CORTEX war ein Spiegel.** Es wurde aus denselben Rohgrößen gespeist wie die Handelsstrategie — und bestätigte deshalb zwangsläufig, was die Strategie ohnehin dachte. Am 28. Juli stimmten beide bei **33 von 33** Trades überein.

Der Spiegel-Befund

Zwei Systeme, die dieselben Eingangsdaten verwenden, sind keine zwei Meinungen. Ihre Übereinstimmung beweist nichts.

Verschärfend: Die **Konfidenz** des Modells war *anti-prädiktiv*. Gewinner-Trades hatten im Mittel 53,2 % Modellvertrauen, Verlierer 57,4 %. Bei mindestens 70 % Vertrauen: **0 % Gewinnquote** (n = 4).

Ein besseres Modell hätte das nicht gelöst. Es brauchte einen zweiten, *unabhängigen* Messkanal.

Die Antwort: ein zweiter Kanal

Gewählt wurden zwei Größen, die **ohne** die fehleranfällige Zuordnung „war das ein Käufer oder ein Verkäufer?" auskommen und den defekten Pfad damit *umgehen* statt ihn zu reparieren:

OFI — Order-Flow-Imbalance: misst den *Fluss* im Orderbuch statt seines Zustands.
Microprice — der um die Warteschlangen gewichtete faire Preis.

Und das Ziel wurde bewusst verkleinert: **kein Richtungsmodell, sondern ein Einstiegs-Gate.** Nicht „wohin geht der Markt", sondern „ist jetzt ein guter Moment".

Drei Befunde aus drei Tagen

1. **Der Größen-Weg trägt keine Information (28./29.07.)**
   Eine Regel zur Erkennung aggressiver Marktteilnehmer traf in 75,4 % der Fälle. Gegen ihr eigenes Placebo: 75,2 %. Zusätzlich zeigte die Verteilungsanalyse, dass 87,7 % aller Abschlüsse nur einen einzigen Kontrakt umfassen — und die Regel bei den großen Abschlüssen (ab 25 Kontrakten) eine Trefferquote von **0,0 %** hat. Sie war blind für genau das, was sie finden sollte. widerrufen

2. **Unser Messgerät zeichnete doppelt auf (29.07.)**
   Zwei Programminstanzen schrieben in dieselbe Datei. **44,5 %** aller Zeilen waren exakte Duplikate. Jede abschlussgewichtete Statistik davor war doppelt gewichtet. Gefunden nur, weil seit Kurzem jede Zeile eine Instanzkennung trägt. Aufzeichnung repariert

3. **Ein toter Sensor blockierte 1.228 von 1.228 Entscheidungen (29.07.)**
   Der Marktgeschwindigkeits-Wert wurde in einem Fenster abgefragt, in dem die zugehörige Variable gar nicht existiert. Statt einen Fehler zu melden, füllte der Code den Ausfall mit dem Ersatzwert `"CALM"` auf — „ruhiger Markt". Die Strategie sah monatelang einen scheintoten Markt und handelte deshalb nicht.
   Daraus wurde eine Leitregel im Code:
   Gar nicht senden ist besser als einen Ersatzwert senden. Ein fehlender Wert sieht aus wie ein Defekt — ein getarnter Ersatzwert sieht aus wie eine Messung. behoben

Der rote Faden dieser drei Tage: Alle drei Befunde betrafen nicht den Markt, sondern das **Messinstrument**. Das ist kein Zufall — es ist die normale Reihenfolge in jeder empirischen Wissenschaft. Bevor man Aussagen über die Welt macht, muss man beweisen, dass das Gerät funktioniert. Die Branche überspringt diesen Schritt fast immer.

ÜBERBLICK

Zwölf Monate auf einen Blick

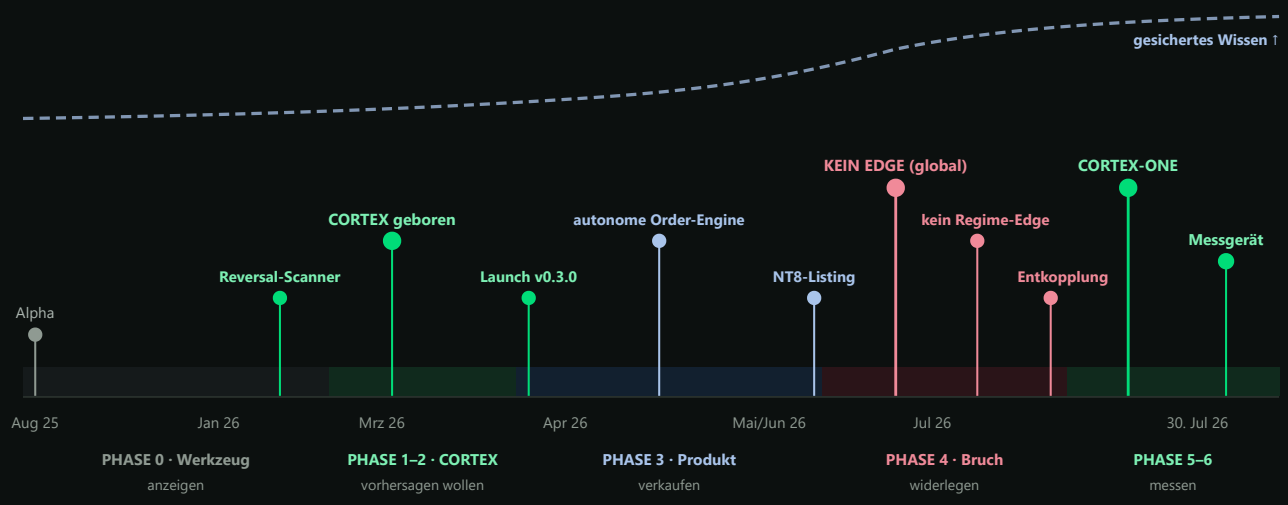


Abbildung 5: Zeitachse. Die gestrichelte Linie zeigt, wann das *gesicherte* Wissen am stärksten wuchs — nämlich genau dann, als die Ergebnisse am negativsten waren.

894	~207 500	808	4
PROTOKOLLIERTE ARBEITSSCHRITTE	ZEILEN CODE (JAVASCRIPT + C#)	WISSENSEINTRÄGE IN DER PROJEKTBASIS	PRODUKT-EDITIONEN IN PRODUKTION

Was jede Phase beigetragen hat

PHASE	LEITFRAGE	BLEIBENDER BEITRAG
0 · Werkzeug	Was passiert im Orderbuch?	Das Datenarchiv. Ohne die Entscheidung, ab Tag 1 alles mitzuschreiben, gäbe es heute keinen Datensatz — und Datensätze lassen sich nicht rückwirkend anlegen.
1–2 · CORTEX	Lässt es sich vorhersagen?	Die lernende Engine samt lokaler Ausführung beim Nutzer. Jede Installation ist ein eigenständiges Experiment, ohne dass Daten das Haus verlassen.
3 · Produkt	Will das jemand kaufen?	Vier Editionen in Produktion, zahlende Kunden, offizielles Ökosystem-Listing — der wirtschaftliche Unterbau, der die Forschung trägt.
4 · Bruch	Stimmt das überhaupt?	Sechs Falsifikationen in vierzehn Tagen und die Entkopplung von Produkt und Forschung. **Gemessen an gesichertem Wissen die wertvollste Phase.**
5–6 · Messgerät	**Wie beweist man es?**	**Der vollständige Beweisapparat: Vorregistrierung, Placebo-Kontrolle, Widerrufsregister, selbstprüfende Werkzeuge, Zertifikatsgenerierung.**

Beobachtung zum Nachdenken: Von August 2025 bis Juni 2026 wuchs das *Produkt* schnell und das *gesicherte Wissen* langsam. Ab Juli 2026 kehrte sich das um. Beide Phasen waren nötig — aber nur die zweite erzeugt etwas, das man zertifizieren lassen kann.

DER APPARAT · INSTRUMENT 1

Das Widerrufsregister — 54 Befunde, die wir selbst gekippt haben

Die ungewöhnlichste Einrichtung dieses Projekts ist eine Datei, in der jede Aussage steht, die sich als **falsch** erwiesen hat. Nicht als Fußnote — als eigenes, maschinell überwachtes Register mit vier Pflichtfeldern je Eintrag.

Warum es entstand

Im Juli 2026 kostete ein bereits widerlegter Befund einen kompletten Arbeitstag. Die Korrektur stand in der maßgeblichen Datei — aber eine automatisch geladene Kurzfassung trug die falsche Aussage weiter. Ein Versionsvergleich hätte das nie gefunden: **Beide Dateien waren „aktuell".** Nur ein Register widerrufener *Aussagen* findet so etwas.

Seitdem gilt: Wird ein Befund widerlegt, kommt er **im selben Arbeitsschritt** ins Register — nicht später, nicht „beim nächsten Aufräumen".

Die vier Pflichtfelder

FELD	INHALT
`pattern`	Ein Suchmuster, das die **falsche** Aussage im gesamten Projekt findet — Notizen, Code-Kommentare, Berichte.
`truth`	Was stattdessen gilt, in einem Satz.
`retracted`	Wann und wodurch widerrufen.
`cost`	**Was der Irrtum gekostet hat** — damit die Lehre nicht abstrakt bleibt.

Ein Prüfprogramm läuft bei jedem Projektstart über alle Dateien. Taucht ein Muster irgendwo noch als Behauptung auf, schlägt es an. **Die Vergangenheit kann sich nicht zurückschleichen.**

Die ersten 22 Einträge

22

WIDERRUFENE BEFUNDE

im selben Arbeitsschritt gefangen	**8**
durch eine Gegenprobe gefallen	**6**
vom Gründer korrigiert	**4**
erst Tage später gefunden	**4**

Teuerster Einzelirrtum: ein voller Arbeitstag — auf einer falschen Prämisse wurde eine komplette Erklärungskette gebaut *und* ein Umbau entworfen, der eine funktionierende Komponente ersetzt hätte.

Was ein Prüfer daraus liest

Ein Unternehmen, das ein Fehlerregister führt, macht nicht mehr Fehler als andere — es **findet** mehr. Für eine Zertifizierungsstelle ist genau das der entscheidende Unterschied zwischen einem Messlabor und einer Marketingabteilung.

Die unbequeme Zahl: Von den ersten 22 widerrufenen Befunden waren **fünf bereits an den Gründer gemeldet**, bevor die Gegenprüfung lief — darunter ein „erster echter Nachweis außerhalb der Trainingsdaten" und ein „erstes Werkzeug, das seine vorregistrierte Hürde genommen hat". Daraus wurde die härteste Regel des Projekts: *Kein Urteil vor der Gegenprüfung — und ein positives Resultat in einer Kette von Negativbefunden verdient mehr Misstrauen, nicht weniger.*

FALLSTUDIEN

Vier Falsifikationen im Detail (1 von 2)

Ausgewählt nicht nach Dramatik, sondern nach Lehrwert: Jede zeigt eine andere Art, wie ein scheinbarer Vorteil entsteht, wo keiner ist.

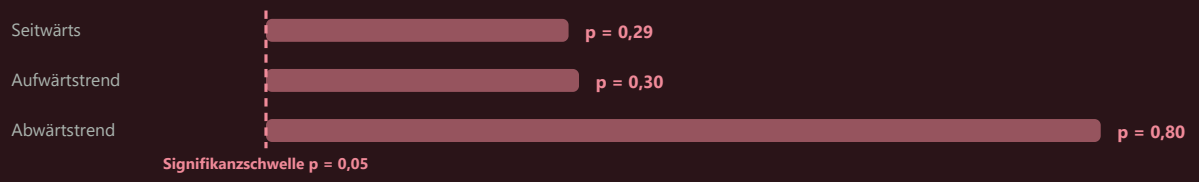
Fall 1 — „Der Vorteil ist regimebedingt"

falsifiziert · 15.07.2026

Die Hypothese. Das Modell zeigt global keinen Vorteil — aber vielleicht funktioniert es in bestimmten Marktphasen sehr gut und in anderen sehr schlecht, sodass sich beides im Gesamttopf wegmittelt. Das ist eine gute, plausible und in der Literatur verbreitete Idee.

Die Prüfung. Alle Beobachtungen wurden nach Marktphase getrennt ausgewertet — gegen eine *Grundwahrheit*, also den tatsächlich eingetretenen Kursverlauf, nicht gegen die Einschätzung des eigenen Systems. Stichprobe: **26.940 Beobachtungen**.

Das Ergebnis.



Je weiter der Balken nach rechts reicht, desto besser passt das Ergebnis zur Annahme „reiner Zufall". Alle drei Marktphasen liegen weit jenseits der Schwelle.

Die Konsequenz. Ein besseres Regime-Label holt keinen Einstiegsvorteil zurück. Damit endeten mehrere Wochen Arbeit an einem „Regime-Router" — und zugleich die Versuchung, an dessen Schwellenwerten zu drehen.

Fall 2 — „Das Modell lernt, die Kurve steigt"

widerrufen · 28.07.2026

Die Behauptung. Eine Lernkurve zeigte einen Anstieg der Trefferquote von **0,8 % auf 18,1 %** über mehrere Sitzungen. Diese Zahl stand seit Juli 2026 im Kopftext des Auswertungswerkzeugs und wurde bei jedem Lauf ausgegeben. Sie war der **einzige Fortschrittsindikator** der eingefrorenen Lernphase.

Der Fehler. Im selben Zeitraum fiel der Anteil der Seitwärtsausgänge von **96,1 % auf 60,0 %**. Damit stieg die theoretisch erreichbare Obergrenze mit — *unabhängig vom Modell*. Rechnet man gegen die tagesabhängige, korrekt bestimmte Zufallslinie, sind **4 von 5 Sitzungen negativ**.

Die Lehre. Ein Fortschrittsindikator, der nicht gegen eine *mitwandernde* Nulllinie rechnet, meldet Marktveränderung als Lernerfolg. Er hat ein Jahr lang genau das getan.

Was Fall 1 kostete

Mehrere Sitzungen Arbeit an einem Regime-Router, den die Daten nicht trugen — plus die Versuchung, an dessen Schwellen zu drehen.

Was Fall 2 kostete

Ein Jahr lang ein Fortschrittsindikator, der Marktveränderung als Lernerfolg meldete — der einzige, den die Lernphase hatte.

Was beide brachten

Die Pflicht, jede Nulllinie aus den Rohdaten zu rechnen statt sie anzunehmen. Heute in jedem Werkzeug erzwungen.

Gemeinsamer Nenner der beiden Fälle: In beiden war die Messung technisch korrekt und der Code fehlerfrei. Falsch war jeweils der **Bezugsrahmen** — wogegen verglichen wurde. Deshalb genügt es nicht, Software zu testen; man muss die *Frage* testen.

FALLSTUDIEN

Vier Falsifikationen im Detail (2 von 2)

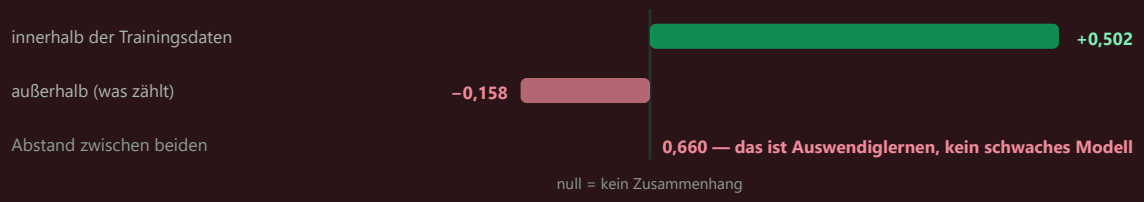
Fall 3 — Der dritte Vorhersagekopf, nach acht Monaten beantwortet

negativ · 29.07.2026

Die offene Frage. Seit November 2025 stand eine Freigabebedingung im Raum: Der dritte Kopf des Modells — der aus dem Marktzustand beim Einstieg das *Ergebnis* eines Trades schätzen soll — dürfe erst scharf geschaltet werden, wenn seine Vorhersagekraft **außerhalb der Trainingsdaten** nachweislich über null liegt.

Warum es acht Monate dauerte. Nicht weil die Rechnung schwer war. Sondern weil es keine *allgemeine* Prüfmethode gab — jede Frage dieser Art wurde einzeln, von Hand und jedes Mal etwas anders beantwortet. Erst als daraus eine wiederverwendbare Bibliothek wurde, war die Frage in Minuten geklärt.

Das Ergebnis. Gemessen auf **295 verwertbaren Trades**:



Rangkorrelation zwischen Modellschätzung und tatsächlichem Trade-Ergebnis. **p = 0,96 · 0 von 10 Datenschnitten positiv.**

Die zwei Nebenbefunde. Dieselbe Prüfung für die vereinfachte Frage „endet der Trade im Plus?" ergab `AUC 0,421`, für den ebenfalls vorgeschlagenen Einstiegs-Gate-Kopf `AUC 0,382` — beide unter der Zufallsgrenze von 0,5. **Positiv:** Die technische Voraussetzung — dass der Datenspeicher über Programmneustarts hinweg hält — ist damit im selben Zug bewiesen.

Fall 4 — „Die Beziehung hat sich nur gedreht"

vorregistriert widerlegt · 29.07.2026

Die letzte Ausrede. Wenn ein Modell außerhalb der Trainingsdaten *negativ* abschneidet, gibt es zwei mögliche Erklärungen. Die schmeichelhafte: Der Markt hat sich verändert, die Beziehung existiert, sie hat nur das Vorzeichen gewechselt. Die langweilige: Es gab nie eine Beziehung, das Modell hat die Trainingsdaten auswendig gelernt.

Der entscheidende Kunstgriff. Die Entscheidungsregel wurde *vor* der Rechnung schriftlich festgelegt — und zwar mit der **langweiligen Erklärung als Voreinstellung**. Nur ein klar definiertes Muster hätte die schmeichelhafte belegen können.

Das Ergebnis. Alle drei Zielgrößen liegen **vorwärts wie rückwärts** unter ihrer eigenen Nulllinie: `0,382 / 0,337` · `−0,158 / −0,151` · `0,421 / 0,391`. Hätte sich eine Beziehung gedreht, müsste mindestens eine Richtung positiv sein. **Sie war nie da.**

Zur Einordnung des Zeitraums: Die Frage aus Fall 3 stand seit November 2025 offen und blieb acht Monate unbeantwortet — nicht wegen fehlender Daten, sondern weil jede Prüfung dieser Art einzeln von Hand gebaut wurde. Erst als daraus eine **wiederverwendbare Prüfbibliothek** mit 23 eigenen Selbsttests wurde, war die Antwort eine Sache von Minuten. Dieselbe Bibliothek beantwortete anschließend die Fälle 3 und 4 sowie zwei weitere Fragen am selben Abend. **Werkzeugbau schlägt Einzelfallarbeit — und zwar um Größenordnungen.**

Warum Fall 4 der wertvollste ist: Er ist der Beweis, dass die Vorregistrierung funktioniert. Ohne sie hätte man nach dem negativen Ergebnis nach einer Erklärung gesucht und mit hoher Wahrscheinlichkeit die schmeichelhafte gefunden — es gibt immer einen Datenschnitt, der sie stützt. **Die Regel stand vorher. Also gab es nichts zu suchen.**

DER APPARAT · INSTRUMENT 2

Vorregistrierung: Die Regel steht vor der Rechnung

Das Verfahren stammt aus der klinischen Forschung. Dort ist es Pflicht, seit sich zeigte, dass Studien mit unerwünschtem Ergebnis systematisch nicht veröffentlicht wurden. In der Finanzsoftware ist es praktisch unbekannt — und genau deshalb der größte Hebel.

Das Problem, das es löst

Bei jeder Auswertung gibt es Dutzende legitime Entscheidungen: Welches Zeitfenster? Welche Mindeststichprobe? Welche Kennzahl? Welche Ausreißer bleiben drin?

Trifft man diese Entscheidungen, *nachdem* man die Daten gesehen hat, findet man fast immer eine Kombination, die gut aussieht — ohne jede böse Absicht. Der Fachbegriff dafür ist der **„Garten der sich verzweigenden Pfade"**.

Vorregistrierung schließt diesen Garten: Alle Entscheidungen werden schriftlich fixiert, versioniert und mit Datum abgelegt, **bevor** die Auswertung läuft.

Wie es hier umgesetzt ist

1. **Eine Datei mit Versionsnummer** enthält alle Vorgaben — Kennzahlen, Mindeststichproben, Hürden, Abbruchkriterien.
2. **Ein Prüfprogramm erzwingt sie.** Weicht ein Auswertungswerkzeug von der Vorgabe ab, läuft es nicht.
3. **Genau ein bestätigender Endpunkt.** Alle anderen Auswertungen sind ausdrücklich *erkundend* — sie dürfen Hypothesen erzeugen, aber keine belegen.
4. **Mindeststichproben in Handelstagen,** nicht in Beobachtungen. Ein einzelner Tag liefert tausende korrelierte Datenpunkte — aber nur *eine* unabhängige Beobachtung.

Was die Vorregistrierung sofort erzwang

Bei ihrer Einführung am 29. Juli 2026 fielen im selben Moment **drei Mängel** in einem bereits fertigen Auswertungswerkzeug auf:

· die Mindeststichprobe war mit 30 zu klein angesetzt → auf **150** erhöht
· es fehlte ein festgelegter Hauptzeithorizont → ergänzt
· die Kennzahl war ein Differenzwert statt eines normierten Skill-Scores → ersetzt

Zusätzlich wurde eine ganze Prüfstufe **gesperrt**, weil ihre statistische Aussagekraft nur bei 28 % lag — sie hätte ein echtes Ergebnis mit hoher Wahrscheinlichkeit übersehen und wäre trotzdem als „geprüft" gezählt worden.

Die härteste Selbstbeschränkung

Aus dem Freigabeprotokoll, wörtlich sinngemäß:

Verboten ist, Schwellen zu ändern, nachdem ein Ergebnis gesehen wurde, damit ein Kandidat qualifiziert.

Verboten ist, auf einem Punktschätzer freizugeben — immer auf der unteren Grenze des Konfidenzintervalls.

Verboten ist, die Schattenphase zu überspringen und direkt scharf zu schalten.

Jede dieser drei Regeln entstand aus einem konkreten, dokumentierten eigenen Fehler.

Für Investoren übersetzt: Vorregistrierung kostet nichts außer Disziplin und ist der Unterschied zwischen einem Ergebnis, das einer Due-Diligence standhält, und einem, das dabei zerfällt. Sie ist außerdem der Grund, warum ein späteres positives Ergebnis dieses Projekts **überhaupt zertifizierbar** wäre — ein nachträglich gefundener Vorteil ist es nie.

DER APPARAT · INSTRUMENT 3

Die Placebo-Kontrolle — der Test, der uns am meisten gekostet hat

In der Medizin bekommt eine Kontrollgruppe ein wirkungsloses Scheinmedikament. Wirkt das echte Mittel nicht deutlich besser, war die beobachtete Besserung kein Beleg. Dasselbe Prinzip lässt sich auf Marktdaten übertragen — und fast niemand tut es.

Der Fall, der die Regel erzwang

Am Abend des 28. Juli 2026 wurde ein Werkzeug gemeldet als *„das erste, das seine vorregistrierte Hürde genommen hat"*. Es sollte erkennen, ob hinter einem Abschluss ein aggressiver Käufer oder Verkäufer stand — die Grundlage praktisch jeder Orderfluss-Analyse.

Trefferquote: **75,4 %**. Gegen eine naive Vergleichsgröße von 8,9 % war das Faktor 8,5. Es sah nach dem ersten echten Durchbruch aus.

Dann lief das Placebo. Dieselbe Regel, angewandt auf einen Datenpunkt **50.000 Abschlüsse entfernt** — also garantiert ohne jeden Zusammenhang.

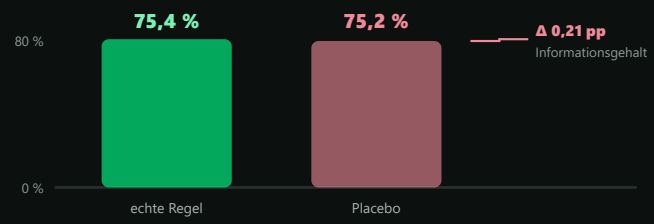


Abbildung 6: Die Regel maß nicht den Markt — sie maß ihre eigene Bauart.

Warum die Regel fast immer traf

Die Verteilungsanalyse erklärte es sofort:

· **87,7 %** aller Abschlüsse umfassen nur **einen einzigen Kontrakt**.
· Bei einer Toleranz von ±1 lautete die Bedingung damit faktisch „irgendeine Veränderung zwischen 0 und 2" — und die trifft fast immer zu.
· Bei den *großen* Abschlüssen ab 25 Kontrakten lag die Trefferquote bei **0,0 %** (n = 314).

Die Regel war blind für genau das, was sie finden sollte.

Der zweite Fund derselben Nacht

Dieselbe Datei, dieselben Daten, nur die Reihenfolge der Prüfschritte vertauscht:

Variante A: **+136.155**
Variante B: **−133.141**

Das Vorzeichen des Gesamtergebnisses kippt allein durch die Prüfreihenfolge. Das ist kein Marktbefund, das ist ein Konstruktionsfehler — sichtbar nur, weil jemand die Reihenfolge testweise umdrehte.

Der Vergleichsmaßstab

Eine simple, seit Jahrzehnten bekannte Faustregel erreicht auf denselben Kontrakten in einer begutachteten Veröffentlichung **86,43 %**.

Die eigenen 75,4 % lagen also **unter dem trivialen Vergleichswert** — was ohne Placebo-Kontrolle niemandem aufgefallen wäre, weil man gegen eine *selbstgewählte* Vergleichsgröße gerechnet hatte.

Die Regel, die daraus wurde: Placebo-Kontrolle, Größenverteilung und Duplikatquote sind seit dem 29. Juli 2026 **Pflichtzeilen** in jedem Klassifikations-Werkzeug. Eine Abdeckungsquote ohne Nullhypothese ist keine Evidenz — sie ist eine Zahl.

DER APPARAT · INSTRUMENT 4

Die Zerfallskurve — trennt Beschreiben von Vorhersagen

Die vielleicht eleganteste Prüfung des ganzen Apparats, weil sie eine Frage beantwortet, die sonst nie gestellt wird: **Wie lange gilt ein Signal eigentlich?**

Das Prinzip in einem Satz

Man misst dieselbe Trefferquote nicht nur „jetzt", sondern mit wachsendem zeitlichen Abstand — nach 5 Ereignissen, nach 50, nach einer Sekunde. Fällt sie sofort auf Zufallsniveau, war das Signal eine **Beschreibung** der Gegenwart, keine Vorhersage der Zukunft.

Zwei Messungen, zwei Konsequenzen

Microprice — Halbwertszeit 90 Millisekunden

Gemessen an **936.000 bereinigten Abschlüssen**. Trefferquote unmittelbar: **86,19 %** gegen eine Mehrheitsklasse von 50,06 %. Nach nur fünf Ereignissen (etwa 225 Millisekunden): **Zufall**.

Konsequenz: Kein zweiter Messkanal. Die Prüfstufe wurde geschlossen und ausdrücklich als „nicht wiederholen" markiert.

Berührungs-Ungleichgewicht — tot nach 1 Sekunde

Die Information sank von einem Skill-Wert 1,05 auf 1,00 bei einer Sekunde Verzögerung.

Konsequenz — und das ist der wertvolle Teil: Der bisherige Datenweg lieferte diesen Wert einmal pro Sekunde mit bis zu zwei Sekunden Alter. Er war beim Einstieg also *überwiegend bereits tot*. Daraufhin wurde ein eigener, direkter Datenkanal gebaut, der alle **200 Millisekunden** schreibt — und die Alterstoleranz von 2.000 auf **500 Millisekunden** gesenkt.

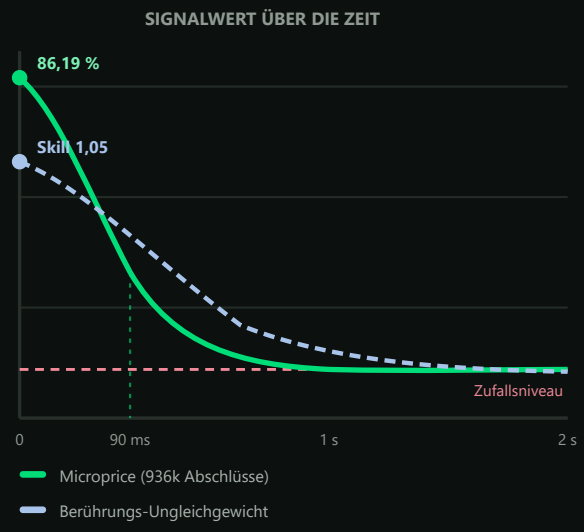


Abbildung 7: Beide Signale sind *real* — und beide sind vor dem Einstiegszeitpunkt bereits wertlos, wenn der Datenweg zu langsam ist.

Warum das eine Ingenieursfrage ist

Die Zerfallskurve verschiebt das Problem von der Statistik in die Technik. Wenn ein Signal 90 Millisekunden lebt, entscheidet nicht das Modell über den Erfolg, sondern die **Latenz der Datenkette**.

Das ist eine gute Nachricht: Latenz ist lösbar. Ein nicht existierender Zusammenhang ist es nicht.

Seit Juli 2026 Pflicht: Jede gemeldete Kennzahl trägt eine Zerfallsspalte. Ohne sie gilt der Befund als unvollständig — im selben Rang wie eine Trefferquote ohne Nullhypothese.

DER APPARAT · INSTRUMENT 5

Der Geltungsbereich — wem gehört diese Zahl eigentlich?

Die häufigste Fehlerklasse dieses Projekts, fünfmal in verschiedenen Verkleidungen aufgetreten: zwei je für sich korrekte Zahlen, die **verschiedene Ausschnitte** derselben Wirklichkeit beschreiben — und deshalb nicht verglichen werden dürfen.

ERSCHEINUNGSFORM	WAS VERGLICHEN WURDE	FOLGE
Zwei Sitzungsfenster	Eine Kennzahl war an 18:00 New Yorker Zeit verankert, die zweite an 9:30. Ergebnis: `+13.225` gegen `−16.816`.	Sah aus wie ein Vorzeichenfehler in der Handelslogik. War eine Anzeigefrage.
Zwei Abtastraster	Eine Berechnung lief alle 5 Sekunden, ihre Protokollzeile wurde nur alle 30 Sekunden geschrieben — also jede sechste.	63 % der Zeitfenster enthielten einen Zustand, den das Protokoll nie zeigte.
Zwei Konten	Eine Tagesbilanz mischte zwei Handelskonten. Die Zahl steuerte eine *Abschaltregel*.	Ein manueller Trade auf dem zweiten Konto beendete den automatischen Handelstag.
Zwei Zeitzonen	Zeitstempel trugen bereits lokale Uhrzeit, wurden aber als Weltzeit gelesen.	Zwei Stunden Versatz. Eine komplette Auswertung wertlos.
Zwei Definitionen von „verwertbar"	Zwei Prüfverträge hießen beide „zitierfähig" und maßen Verschiedenes: einer meldete 4 von 9 Tagen, der andere 1 von 9.	Die Schnittmenge war **leer** — beide Werte waren korrekt, gemeinsam unbrauchbar.

Die Gegenmaßnahme: ein Geltungsbereichs-Wächter

Ein eigenes Prüfmodul stellt vor jeder Auswertung sicher, dass die verwendeten Kategorien die **gesamte Grundgesamtheit ausschöpfen**. Klassisches Beispiel aus dem eigenen Betrieb: `2.529 + 1.126 ≠ 9.906` — es fehlten über 6.000 Fälle, die stillschweigend in keine Kategorie fielen.

Der Wächter hat **zehn eigene Selbsttests**, und seine Gegenprobe ist genau dieser reale Fall: Er muss ihn finden, sonst gilt er als defekt.

Die Leitfrage, die daraus wurde

Vor jedem Satz der Form *„diese Zahl steuert etwas"* steht seither die Frage:

Wem gehört diese Zahl?

Welches Konto, welches Zeitfenster, welche Instanz, welche Zeitzone, welche Definition? Erst wenn alle fünf beantwortet sind, darf die Zahl eine Entscheidung auslösen.

Und: Auch der Wächter selbst fiel einmal auf diese Falle herein — er meldete einen nach Handelsschluss ruhenden Schreibprozess als Defekt. Korrigiert mit Gegenproben für *beide* Fensterränder.

Warum das für eine Zertifizierung zentral ist: Ein Prüfer wird nicht fragen „ist die Zahl richtig?", sondern „worauf bezieht sie sich?". Ein System, das diese Frage für jede Kennzahl automatisch beantwortet, ist prüfbar. Eines, das es nicht tut, ist es nicht — unabhängig davon, wie gut sein Ergebnis aussieht.

DER APPARAT · INSTRUMENT 6

Werkzeuge, die sich selbst prüfen — und Wächter, die bewiesen wachen

Ein Prüfprogramm, das immer grün läuft, ist wertlos. Man erkennt den Unterschied nicht am Ergebnis, sondern nur daran, ob es **überhaupt noch rot werden kann**. Deshalb hat hier jedes Werkzeug zwei eingebaute Teile: einen Selbsttest und eine Gegenprobe.

170	50	22	17
ANALYSE-WERKZEUGE IM BESTAND	MIT EINGEBAUTEM SELBSTTEST	SCHRITTE IM ABENDLAUF	DAVON MIT FACHLICHER ABNAHME

Der Unterschied zwischen technischer und fachlicher Abnahme

Technische Abnahme — „lief es?"

Das Programm ist ohne Fehler durchgelaufen, hat eine Datei geschrieben, der Rückgabewert war null. **Das sagt nichts über die Richtigkeit des Ergebnisses.**

Ein Werkzeug kann fehlerfrei durchlaufen und dabei die falsche Größe messen — genau das ist bei allen fünf Fallen auf Seite 5 passiert.

Fachliche Abnahme — „stimmt das Ergebnis?"

Eine inhaltliche Frage, die nur mit korrekten Daten beantwortbar ist. Das Musterbeispiel aus dem eigenen Betrieb:

„Berührt der rekonstruierte Kursverlauf eines Vollverlusts überhaupt seinen Stop-Kurs?"

Antwort beim ersten Lauf: **35 %**. Nach der Reparatur des Datenpfads: **100 %**. **Und das Endergebnis kippte dabei ins Gegenteil.**

Der Fall, der zeigt, warum das nötig ist

Eine Auswertung zur Frage „wäre ein früherer Ausstieg besser gewesen?" meldete zunächst **+6,96 R** Verbesserung bei einer Schwelle von 2 Ticks. Die fachliche Abnahme deckte auf, dass der zugrunde liegende Kursverlauf zu grob war — er bestand aus etwa 20 Punkten je Trade statt aus dem tatsächlichen Verlauf.

Aus dem Rohdatenstrom neu geschnitten (**Median 1.447 Punkte je Trade**) kippte das Ergebnis: Die 2-Tick-Schwelle *verliert* nun 2,67 R, bester Wert ist eine 8-Tick-Schwelle mit +1,93 R. Auch das ist noch kein Urteil — die Stichprobe umfasst 30 Trades über 2 Tage. **Aber der grobe Pfad zeigte nachweislich in die falsche Richtung.**

Die vier Formen von „dokumentiert, aber nicht durchgesetzt"

1 · Regel ohne Wächter
Steht in der Doku, prüft niemand.

2 · Pfad, der schweigt
Tut bei Störung *nichts* — statt zu melden.

3 · Wächter, der grün läuft
Läuft, prüft aber seine Aussage nicht mehr.

4 · Wächter ohne Selbsttest
Kann nicht beweisen, dass er rot werden kann.

Die daraus abgeleitete Arbeitsregel: Wer eine Falle aufschreibt, legt **im selben Arbeitsschritt** fest, welcher automatische Test sie künftig findet. Sonst ist die Dokumentation nur eine Notiz darüber, wie man denselben Fehler beim nächsten Mal wieder machen wird.

FORSCHUNGSSTAND

Was heute gesichert ist — die Bilanz nach zwölf Monaten

Diese Seite ist die ehrlichste des Berichts. Sie trennt sauber zwischen dem, was **gemessen** ist, dem, was **widerlegt** ist, und dem, was **offen** ist.

Gesichert negativ — mit ausreichender statistischer Aussagekraft geprüft

HYPOTHESE	PRÜFUNG	BEFUND
Die neuronale Engine hat global einen Richtungsvorteil	3× powered	**Nein.** Matthews-Korrelation ≈ 0,02 — praktisch Zufall
Der Vorteil existiert in bestimmten Marktphasen	n = 26.940	**Nein.** p = 0,29 / 0,30 / 0,80
Das Regime-Label trägt Richtungsinformation	v1 *und* v2	**Nein.** 15/16 Zellen unter Grundlinie, ~19 % falsches Vorzeichen
Der 41-Merkmale-Zustand ist überhaupt lernbar	tick-genau	**Nein.** Rangkorrelation außerhalb −0,158, 0/10 Schnitte positiv
Der Microprice ist ein zweiter, unabhängiger Messkanal	936k Abschlüsse	**Nein.** Halbwertszeit 90 ms — beschreibend, nicht vorhersagend
Die Größen-Evidenz identifiziert Aggressoren	Placebo	**Nein.** 75,4 % gegen 75,2 % Placebo = 0,21 pp Information
Die Modellkonfidenz zeigt gute Trades an	33 Trades	**Umgekehrt.** Gewinner 53,2 %, Verlierer 57,4 % Vertrauen
Der KI-Positionsmanager schlägt die einfachen Regeln	Schattenlauf	**Nein.** 48 % gegen 64 % — bleibt abgeschaltet

Gesichert positiv

Der Timing-Befund. Über 422 Trades aus 23 Handelstagen:
· Trade erreicht sein erstes Kursziel → **100 % Gewinn** (n = 100)
· Trade erreicht es nicht → **25 % Gewinn** (n = 322)
· **40 % aller Verlierer sind nach 30 Sekunden tot**

Dieser Befund ist die tragfähigste Erkenntnis des Projekts, weil er die Forschungsfrage **verkleinert**: nicht mehr „wohin geht der Markt", sondern „ist dieser Moment tragfähig".

Ebenfalls gesichert — technisch

· Der Lernspeicher **hält über Programmneustarts** hinweg (Stufe 2 des dritten Kopfes belegt)
· Die Aufzeichnung ist seit dem 29.07. **duplikatfrei** (Schreiberwahl je Prozess)
· Die Datenkette ist auf **200 ms** beschleunigt und die Alterstoleranz gemessen begründet
· Die installierte Software wird gegen ihre Prüfsumme auf der Festplatte verifiziert — die Projektregel „der Nutzer verwendet die installierte Fassung" wird seit dem 29.07. erstmals **gemessen**, nicht angenommen

Die zusammenfassende Aussage, so klar wie möglich: Aus dem heute erfassten Marktzustand lässt sich die künftige Kursrichtung mit den bisher geprüften Verfahren **nicht** überzufällig vorhersagen. Was sich abzeichnet, ist etwas anderes und Kleineres — dass sich *Einstiegsmomente* nach ihrer Überlebenswahrscheinlichkeit unterscheiden lassen. Das ist die aktuelle Spur.

AKTUELLE SPUR

Der 29. Juli 2026 — eine Spur, ausdrücklich kein Urteil

Am Tag vor Abfassung dieses Berichts lief zum ersten Mal die auf CORTEX umgebaute Strategie mit repariertem Marktgeschwindigkeits-Sensor. Das Ergebnis ist auffällig genug, um es zu berichten — und schwach genug, dass wir es **nicht** als Befund bezeichnen.

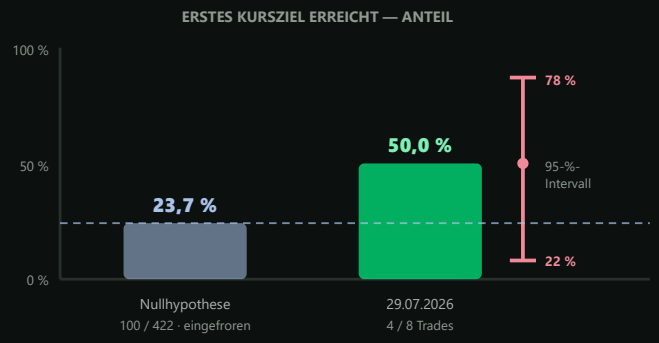


Abbildung 8: Das Konfidenzintervall des neuen Werts **schließt die Nullhypothese noch ein**. Genau deshalb ist dies eine Spur und kein Ergebnis.

Der Mechanismus reproduziert sich

Bemerkenswert ist nicht die Quote, sondern die **Struktur** des Tages — sie wiederholt exakt den Befund aus 422 Vergleichstrades:

AUSGANG	ANZAHL	ERGEBNIS
Erstes Kursziel erreicht	4	**4 von 4 gewonnen** — alle mit gesichertem Einstand
Nicht erreicht	4	**1 Kratzer, 3 Vollverluste** nach 18 / 26 / 26 Sekunden
Tagesergebnis	**8**	**+522,64 $ · 62,5 % Trefferquote**

Nicht die Richtung wurde besser — die Einstiege überleben die erste halbe Minute häufiger.

Was den Tag überhaupt möglich machte

Kein neues Modell, keine neue Regel — die Reparatur eines seit Monaten toten Messfühlers. Der Marktgeschwindigkeits-Wert wurde in einem Programmfenster abgefragt, in dem er nicht existiert, und stillschweigend mit dem Ersatzwert „ruhiger Markt" aufgefüllt.

Folge: Die Sperre *Markt zu ruhig* blockierte **1.228 von 1.228** buchbestätigten Entscheidungen. Nach der Reparatur fiel sie auf **null**.

Die Lehre steht seither als Leitregel im Code: Gar nichts zu senden ist besser, als einen Ersatzwert zu senden. Ein fehlender Wert sieht aus wie ein Defekt — ein getarnter Ersatzwert sieht aus wie eine Messung.

Warum wir das nicht feiern

1. Zu klein. Einseitig gerechnet: P(≥4 Ziele erreicht) = `0,097`. Für die Gesamt-Trefferquote sogar `0,217` — die trägt gar nichts.

2. Der Tag zählt formal nicht. Am 29.07. wurde die Handelslogik siebenmal neu geladen, es liefen **fünf verschiedene Programmstände**. Ein Tag mit mehr als einem Stand ist per Vorregistrierung ausgeschlossen — auch wenn das Ergebnis gefällt.

3. Ein Begrenzer hat gegriffen. Das Tageslimit von 8 Trades wurde erreicht; ab jetzt ist es der Engpass, nicht die Signalqualität.

Das Hauptbuch, das jeden Dreh sichtbar macht

Seit dem 29.07. läuft ein fortlaufendes Register, das in **jeder Zeile** die eingefrorene Nullhypothese mitschreibt (100 / 422). Ändert sie jemand später, ist die Änderung im Versionsverlauf sichtbar.

Zwei eingebaute Kontrollen:
· **Placebo** — Nullhypothese gegen sich selbst: `p = 0,519` ✓
· **Gegenrichtung** — 50 % bei n = 120: $p = 3{,}8 \cdot 10^{-10}$ ✓

Der Prüfer kann also beides: nicht zu früh anschlagen und ein echtes Signal erkennen.

Die Zielgröße hat gewechselt

Nicht mehr die Trefferquote, sondern die **Zielerreichungsquote**. Der Grund ist rein statistisch: Von 23,7 % auf 50 % braucht es rund **40 Trades** für ein belastbares Urteil — etwa **fünf Handelstage** statt zwanzig.

DIE OFFENE KETTE

Neun Stufen — die Kaskade ist gefallen, der Endpunkt ist gewandert

Der Forschungsplan vom Juli bestand aus neun aufeinander aufbauenden Prüfstufen mit genau **einem** bestätigenden Endpunkt. Am 1. September fiel Stufe 5 — und mit ihr, wie in der Vorregistrierung wörtlich festgelegt, alles darüber — kein Bruch des Plans, sondern sein vorgesehener Ausgang.

STUFE	FRAGE	STATUS	BEFUND · STAND 23. SEPTEMBER
0	Ist die Aufzeichnung vollständig und eindeutig?	erledigt	Lücken-Scan je Tag, seit dem 24. August in der Beweiskette
1	Stimmen Buch, Abschlüsse und Protokoll überein?	erledigt	Abnahme 10/10, Ladder-Abdeckung 98,90 %
2	Lässt sich der Aggressor sauber bestimmen?	verworfen	Placebo-Kontrolle: 0,21 pp Information
3	Trägt der Microprice Vorhersagekraft?	nein	Halbwertszeit 90 ms — ausdrücklich nicht wiederholen
4	Misst der Orderfluss (OFI) die Bewegung?	trägt	R^2 im Median **0,893** (42 Läufe) — er *beschreibt* die gleichzeitige Bewegung fast vollständig
5	Sagt er die nächste Bewegung voraus?	gefallen	**1. 9.:** Skill 0,995 gegen Hürde 1,25, 19 Bandtage, alle vier Horizonte — Kapitel 24
6a	Wirken Sweeps als Auslöser?	stillgelegt	Revision 13 — ohne Stufe 5 gegenstandslos
6b	Wirkt Absorption an Wänden?	stillgelegt	Revision 13; die Wand-Frage lebt im Autobahn-Test weiter
7	Sind Stop-Jagden erkennbar?	stillgelegt	Revision 13; der Kursziel-Abgleich bleibt als Werkzeug erhalten
8	**Verbessert das OFI-Einstiegs-Gate das Ergebnis?**	ersetzt	**Neuer Endpunkt „Bracket-MinP" (5. 9.) — der Blick steht aus**

Warum weiterhin genau ein Endpunkt bestätigend ist

Der Endpunkt ist gewandert, nicht vermehrt. Die neue Vorregistrierung fragt: *widerspricht der vierte CORTEX-Kopf bei verlorenen Einstiegen häufiger als bei gewonnenen?* — 10 Handelstage, **zweiseitig** (ein Ergebnis unter null belegt ein *schädliches* Tor und ist derselbe Befund wert), mit Pflicht-Gegenprobe ohne den ersten Tag.

Ungewöhnlich daran: das Feld **Vorsichtung**. Tag 1 war vor der Unterschrift bekannt (−4,8 Prozentpunkte, *gegen* die Hypothese) und steht offen im Dokument, mitgezählt im Fingerabdruck. Die nächste Vorregistrierung (tote Zonen) liegt als Entwurf bereit — ihr Anspruchsrecht wird erst **nach** diesem Urteil freigegeben.

Die Bremse — und ihr bewusster Bruch

Seit dem 30. Juli gilt: keine Neukompilierung der Handelslogik. Zwischen dem 1. und dem 4. September wurde sie **viermal gebrochen** (v0.9.5 bis v0.9.9) — jedes Mal auf Anweisung des Gründers, jedes Mal an drei Stellen dokumentiert: im Vorregistrierungs-Feld, im Projektstand, im Quelltext am Tor.

Der Zähler homogener Tage hat seither **nie gezählt: 0 von 20** — der Quelltext wechselte am 12., 15. und 22. September, jedes Mal dokumentiert. Ein bindender Anker ist er nicht mehr.

Zum Stand der Dinge: Der Apparat steht, die Regeln stehen, die Werkzeuge prüfen sich selbst. Was läuft, ist **Datensammlung** — seit Release 0.5.30 (23. 9.) erneut von null.

ZEITPLAN

Der Weg zum Urteil — was wann entschieden wird

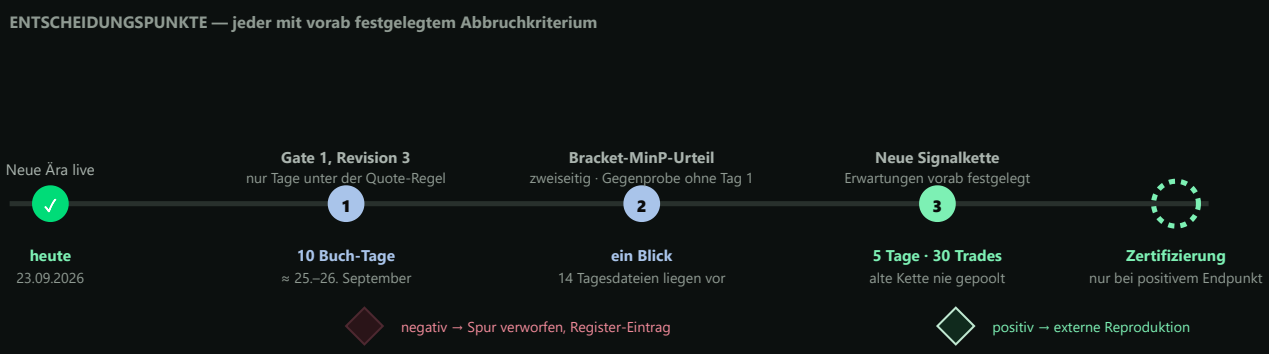


Abbildung 9: Jeder Meilenstein hat zwei Ausgänge, und beide sind vorab definiert. Ein Zeitplan ohne definiertes Scheitern ist ein Wunschzettel.

Was passiert, wenn es negativ ausgeht

Der Befund kommt ins Widerrufsregister, die Spur wird geschlossen, der nächste Kandidat aus der Liste rückt nach. **Das Produkt ist davon unberührt** — es verkauft seit Juli 2026 ausdrücklich keine Edge-Behauptung.

Kosten eines negativen Ausgangs: einige Wochen. Kosten eines *ungeprüft positiven* Ausgangs: die Glaubwürdigkeit des gesamten Unternehmens.

Was passiert, wenn es positiv ausgeht

Dann liegt zum ersten Mal ein Ergebnis vor, das **alle sechs Stufen der Beweisleiter** erfüllt: vorregistriert, placebo-kontrolliert, außerhalb der Trainingsdaten geprüft, gegen eine Nullverteilung getestet, unter eingefrorenen Bedingungen erhoben und aus archivierten Rohdaten **von Dritten nachrechenbar**.

Ab diesem Punkt beginnt der Zertifizierungsprozess — beschrieben in Teil V.

Warum der Zeitplan in Handelstagen und nicht in Wochen gerechnet ist: Ein Kalenderdatum lässt sich einhalten, indem man die Anforderungen senkt. Ein Zähler homogener Handelstage lässt sich nur durch Warten erfüllen — **und er springt bei jeder Änderung auf null zurück.** Das ist die einzige Form von Terminplanung, die gegen die eigene Ungeduld immun ist.

FÜNF TAGE IM SEPTEMBER

Drei Nullbefunde in fünf Tagen — und was sie verhindert haben

Zwischen dem 1. und dem 5. September fielen drei Bilder, die sich richtig anfühlten, gegen ihre uhrzeitgematchte Nulllinie. Keines war ein Rückschlag: **jedes hat einen Umbau verhindert**, der Wochen gekostet und Daten entwertet hätte. Alle drei waren vorregistriert; alle drei stehen im Widerrufsregister.

BILD	GRUNDMENGE	MESSUNG GEGEN KONTROLLE	VERHINDERT
Orderfluss-Richtung (Stufe 5) OFI-Vorzeichen → nächste Mid-Bewegung	19 Bandtage 77.972 Fenster (5 s)	Treffer **49,76 %**, Skill **0,995** (KI 0,986–1,004) gegen Hürde 1,25 — auf allen vier Horizonten, Bootstrap über Tage	das geplante OFI-Einstiegs-Gate; laut Vorregistrierung fällt damit die Kaskade 5–8
Liquiditätsmarker (LP) „zeigt fast immer am Wendepunkt"	693 Marker 19 Tage	≥ 4 Ticks in 60 s: **50,6 %** gegen 49,9 % Kontrolle (95. Perzentil 65,0 %) — alle vier Zellen fallen	ein 42. Merkmal für CORTEX — und damit die Löschung von 27 Sitzungen Lernstand
Trend-Fortsetzung das „Autobahn"-Modell, elf Zellen	19 Tage 436.088 RTH-Sekunden	Weiterfahrt nach Tankstopp **46,5 %** gegen 58,8 % — **0 von 19 Tagen** besser; Wände halten (42,9 % gegen 49,5 %)	ein Phasen-Modul in der Strategie; stattdessen Fade-Vorregistrierung, Stufe A läuft
Ausstieg auf Gegensignal **„bei Gegensignal sofort raus"**	**61 Trades 16 Tage**	**jede Regel schlägt das Ist (+1.067 bis +3.582 $) — und verliert gegen Zufallszeitpunkte (−296 bis −2.811 $)**	**einen Exit-Regler, der nur gut aussah, weil 79 % der Trades am Stop enden**

Was trägt: Stufe 4 — und eine Signatur

Dasselbe Instrument, das die Richtung *nicht* vorhersagt, misst die *gleichzeitige* Bewegung fast vollständig: R^2 im Median **0,893** über 42 Läufe. Die Vorregistrierung nennt die Verwechslung von Beschreiben und Vorhersagen „den häufigsten und teuersten Fehler in diesem Literaturfeld" — Kapitel 17 hat sie zum Prüfkriterium gemacht.

Alle drei Nullbefunde tragen dieselbe Signatur wie die Timing-Achse aus Kapitel 26: **Fluss ⇒ Rückkehr** in 30 bis 600 Sekunden, mit Vorlauf. Das ist keine Bestätigung — es ist der Grund, warum die Fade-Vorregistrierung unterschrieben wurde, *bevor* jemand an einer Schwelle drehte.

Der vierte Kopf — zweimal falsch bewertet, an einem Abend

CORTEX bekam Anfang September einen vierten Vorhersagekopf („erreicht der Trade 8 Ticks, bevor er 10 Ticks verliert?"). Seine erste Kennzahl las sich wie „doppelt so gut wie die Basisrate": Genauigkeit 0,631 neben Positiv-Rate 0,305. Die richtige Vergleichslinie für Genauigkeit ist **0,695** — der Kopf lag **darunter**. Der Brier-Wert 0,448 sah gegen 0,21 schlecht aus; seine naive Linie ist 0,408.

Beide Irrtümer stehen im Register. Konsequenz: Kennzahlen werden seit App 0.5.9 **mit ihrer Vergleichslinie** geschrieben, nicht daneben. Und bevor über einen Kopf geurteilt wird, gilt die Frage: **wer liest ihn überhaupt?** In der Strategie damals: niemand.

Sechs Fälle einer Klasse an zwei Tagen: ein Leser verwarf 2.941 auswertbare Zeilen, weil der Strom „UP/DN" schrieb und er „LONG/SHORT" erwartete — Ausgang grün. Eine Brücke setzte ein neues Feld je Nachricht auf null zurück (ein Handelstag, −448 $). Eine Spalte wurde seit Wochen nie befüllt (0 von 27.979 Zeilen). In allen sechs Fällen meldete etwas „nichts", wo „passt nicht zusammen" richtig gewesen wäre. Seit dem 5. September prüft ein Wächter das Werteinventar jedes Datenstroms gegen die Werte, auf die seine Leser verzweigen (Abendlauf, Schritt 7) — Positivkontrolle am echten Vorher-Stand gefahren.

DIE MESSGRENZE

Das Delta-Mikroskop — 56,7 % des Volumens haben kein sicheres Vorzeichen

Jede Delta-Anzeige am Markt behauptet zu wissen, wer bei einer Transaktion der Aggressor war — Käufer oder Verkäufer. Im August haben wir gemessen, **wie oft zwei gleichermaßen legitime Klassifikationsregeln auf denselben Rohdaten uneinig sind.** Das Ergebnis verändert, was ein „genaues Delta" überhaupt bedeuten kann.

Der Befund

Zwei Regeln, ein Rohdatenstrom, 14 Tage

Jede Transaktion aus 14 vollständig aufgezeichneten Tagen wurde **doppelt** klassifiziert: einmal nach der Tick-Regel (Vergleich zum letzten Preis), einmal nach der Quote-Regel (Lage zum Geld-/Briefkurs). Auf **56,7 % des gehandelten Volumens** widersprechen sich die beiden — das Vorzeichen dieser Anteile hängt an der *Wahl der Regel*, nicht an den Daten.

Die Uneinigkeit ist kein Rauschen — sie hat eine Form

Sortiert nach dem Alter der letzten Quote fällt die Widerspruchsquote **monoton von 51,6 % auf 3,2 %**. Je frischer die Quote, desto einiger die Regeln. Die Unbestimmtheit ist damit **messbar, erklärbar und je Transaktion bezifferbar** — sie ist eine Eigenschaft des Datenstroms, kein Fehler eines Anbieters.

Der unbequeme Nebenbefund

Bei der Messung fiel auf: **unsere eigene Software führte zwei Delta-Pfade mit zwei verschiedenen Regeln** — Chart und Heatmap konnten demselben Moment verschiedene Vorzeichen geben, und niemand wusste es. Der Befund steht im Widerrufsregister; die Vereinheitlichung ist Teil des nächsten Auslieferungsstands.

Die Konsequenz

„Das genaueste Delta" ist eine unbelegbare Aussage

Wenn über die Hälfte des Volumens je nach Regel das Vorzeichen wechselt, kann **kein** Anbieter beanspruchen, „das richtige" Delta zu zeigen — es gibt auf diesen Daten kein beobachtbares Richtig. Belegbar ist etwas anderes: ein Delta, das **seine eigene Fehlerbreite ausweist** — das zu jedem Wert sagt, welcher Anteil davon regelfest ist und welcher an einer Konvention hängt.

Genau das ist die Produkteigenschaft, die aus dieser Messung folgt: **nicht „am genauesten", sondern „mit ausgewiesener Unbestimmtheit"** — eine Eigenschaft, die heute kein Wettbewerber anbietet und die ohne diese Messung niemand anbieten kann.

Und das benannte Produktrisiko

Kunden vergleichen unsere Anzeige live mit Fremdanbietern, die stillschweigend die *andere* Regel fahren. An Tagen mit vielen alten Quotes können beide Anzeigen gegenläufig wirken — **beide konsistent mit den Rohdaten**. Ohne ausgewiesene Spanne sieht das wie ein Defekt aus. Mit ihr ist es ein Messwert.

Die Zahl, die bleibt: Kein Anbieter kann die 56,7 % kleiner machen — die Rohdaten geben nicht mehr her. Man kann sie nur **ausweisen oder verschweigen**. Wir haben uns für das Ausweisen entschieden, weil es die einzige überprüfbare Position ist.

DIE ZWEITE SUCHRUNDE

Vier leere Leser-Klassen — und eine Achse, die trägt

Das „NICHT_LERNBAR" aus Gate 1 galt ausdrücklich nur für einen **linearen, punktuellen Leser**. Die naheliegende Frage: Sehen stärkere Modellklassen auf denselben Daten mehr? Ende August wurde sie systematisch beantwortet — mit einem doppelten Ergebnis.

Die Antwort der stärkeren Leser: leer

1. **Vier Leser-Klassen**
   — vom Baum-Ensemble bis zum sequenziellen Netz — auf dem identischen Datenraum (12 Merkmale, 8 Schritte, 600-s-Horizont):
   keine liefert ein replizierbares Signal.

2. **Der eine „LERNBAR"-Treffer widerlegte sich selbst.**
   Ein rekurrentes Netz meldete zunächst Lernbarkeit — die vorregistrierte Replikation maß dann
   AUC 0,504
   : eine Münze. Ohne Replikationspflicht wäre das als Durchbruch gefeiert worden.

3. **Gate 1 steht damit breiter da als je:**
   NICHT_LERNBAR bei
   effN 2.279
   auf
   22 von 22
   zitierfähigen Aufzeichnungstagen (Urteil vom 27. August, quittiert am 28.). Die obere Schranke aus Kapitel 23 bleibt unverändert gültig.

Was das Urteil jetzt abdeckt — und was nicht

Abgedeckt: lineare *und* die vier geprüften stärkeren Klassen auf dem punktuellen Merkmalsraum. **Nicht abgedeckt:** andere Merkmalsräume und andere Zeitauflösungen — genau dort suchte die Wochenendstudie des nächsten Kapitels weiter.

Der Nebenbefund mit Substanz

Die Timing-Achse: Fluss ⇒ Rückkehr im Mittelfenster

Quer zu den Leser-Klassen zeigte sich eine gerichtete Struktur: **starker aggressiver Fluss sagt im Fenster 30–600 Sekunden eine Gegenbewegung voraus** (Mean-Reversion), nicht die Fortsetzung. Der Effekt überlebt die Bounce-Kontrolle, den Zirkular-Shift und den Tages-Split (**9 von 19 Tagen** einzeln signifikant, kein Tag gegenläufig) — und die eigenen aufgezeichneten Trades sind mit ihm konvergent (Korrelation −0,27, n = 48).

Status: Kandidat — mit Datum

Der Effekt ist **ohne Handelskosten** gemessen und noch nicht repliziert. Er trägt deshalb den Status **HYPOTHESEN-KANDIDAT** mit vorregistriertem Replikationsfenster um den **9. September**. Erst wenn er dort erneut erscheint, wird aus der Achse eine Vorlage für den konfirmatorischen Pfad — vorher ist sie eine Beobachtung.

Der Unterschied zwischen „Fund" und „Kandidat" ist ein Datum: das der vorregistrierten Replikation. Alles, was dieses Datum nicht abwartet, ist eine Geschichte über die Vergangenheit — davon hatte dieses Projekt genug.

Warum die Null der Leser-Klassen wertvoll ist: Sie schließt die billige Erklärung aus, wir hätten nur das falsche Modell benutzt. Wenn im nächsten Kapitel ein Kandidat gefunden wird, dann **nicht**, weil ein stärkeres Modell mehr hineinlesen konnte — sondern weil ein *anderer Datenraum* geöffnet wurde.

DIE WOCHENENDSTUDIE

Fünfzehn Millionen Konfigurationen — eine Fata Morgana und ein Kandidat

Am 29. August liefen zwei erschöpfende Suchen über den gesamten aufgezeichneten Datenbestand. Die erste bewies, dass unser Prüftrichter einen echten Vorteil **durchlassen würde**. Die zweite fand einen Kandidaten, der jede Erstprüfung besteht — und entlarvte unterwegs einen zweiten, der keine bestand.

Lauf 1 — das Buch im Sekundenraster

631.800 Konfigurationen, null Finalisten

Sechs Signalfamilien × feines Exit-Raster, erschöpfend statt gestichprobt. Disziplin vorab deklariert: chronologischer Split 13/4/4, **Testblock nur einmal berührbar**, Plateau-Pflicht, 10.000 Permutationen. Ergebnis: **kein Buch-Merkmal trägt in diesem Raum etwas über die reine Preisbewegung hinaus.**

Die Positiv-Kontrolle — der Beweis, dass die Null zählt

In 21 synthetische Tage wurde ein bekannter Vorteil eingepflanzt und durch denselben Trichter geschickt: **582 → 60 → 49 → 20 Finalisten, 20 von 20 bestehen den Einmal-Test** (getrennter Ledger, danach entfernt). Der Trichter lässt echte Edges durch. **Die Null von Lauf 1 ist damit eine Aussage über die Daten — nicht über das Werkzeug.**

Die Fata Morgana

Ein Cluster („Quote-Rückzug") sah in der Validierung glänzend aus: **+51 bis +64 $ je Trade**. Der einmalige Testblock zeigte **−10 bis −78 $** — nur 1 von 12 Exit-Varianten positiv, eine Messerschneide. Ohne die Einmal-Disziplin wäre genau das als Fund gefeiert worden. **Das ist der Apparat bei der Arbeit.**

Lauf 2 — die Rohdaten im Viertelsekundenraster

14,9 Millionen Konfigurationen auf 103 Millionen Ereignissen

250-ms-Raster direkt aus den Quote-Rohbändern (21 Sessions — darunter drei Tage, die der App-Strom verloren hatte und die Aufzeichnung des Indikators lückenlos trug), konservative Limit-Ausführung, 16 Kontext-Tore, Konfluenz-Paare, iterative Verfeinerung — **81,7 Minuten Rechenzeit**, Deklaration vor dem Lauf.

Der Überlebende: „Flush-Absorption"

Preis stürzt ≥ 12 Ticks in 10 s, **während** der 2-Sekunden-Orderfluss starkes aggressives *Kaufen* zeigt → Einstieg long per Limit in den Sturz. Training **+17…23 $** je Trade (n ≈ 111) · Validierung **+15…19 $** · einmaliger Test **+27…43 $ bei 75 % Trefferquote — alle 7 Exit-Varianten positiv** (ein Plateau, keine Spitze), p ≤ 0,003 bei 10.000 Permutationen. Mikrostrukturell kohärent: der klassische Absorptions-Abpraller — und die Konfluenz-Form der Timing-Achse aus Kapitel 26.

Die Ehrlichkeit dazu: Der Testblock trug nur **n = 16** Trades (Wilson-Untergrenze ~51 %). Deshalb lautet der Status nicht „Edge", sondern **HYPOTHESEN-KANDIDAT mit bestandener Erstprüfung** — eingefroren (jede Änderung = neuer Kandidat, Zähler null) und seit dem 29. August im **täglichen Forward-Test** auf jeder neuen Session. Schwelle: **n ≥ 50 und Wilson-Untergrenze ≥ 55 %. Nachtrag 23. 9.:** Der Forward-Test hat gegen den Kandidaten entschieden. Bei n = 90 trifft er 65,6 % (Wilson 55,3 %) — und verliert **5,17 $ je Trade**. Seit dem 18. 9. verworfen. Am 14. 9. war die Schwelle kurz erreicht (n = 52); die Regel kennt kein Anhalten beim ersten Überschreiten. Lehre: eine Trefferquote ohne Erwartungswert ist keine Kante.

ZEHN TAGE IM SEPTEMBER · I

Die Replay-Maschine — hätte die Strategie andersherum gewonnen?

Am 9. und 10. September wurden alle **105 nachspielbaren Strategie-Trades** seit dem 29. Juli tickgenau auf dem eigenen Tape nachgespielt — mit echtem Preispfad, Stop, Ziel und Kosten (0,64 Ticks je Kontrakt). Vier Vorregistrierungen, ein Blick je Variante. Die Frage dahinter war unbequem: **liegt der Fehler im Management — oder im Einstieg selbst?**

FRAGE	VARIANTEN	NETTO-TICKS JE KONTRAKT	URTEIL
Kalibrierung Original, unverändert	das tatsächliche Regelwerk	**−1,76** · Trefferquote 35,2 %	Referenz
Umkehr der Richtung „andersherum handeln"	jeder Einstieg gespiegelt, eigener Stop	**−1,73** mit 1 Tick Schlupf (−1,04 ohne) · 39,0 % gegen Break-even 50,4 %	trägt nicht
Struktur Zeit-Stop, T1, Break-even-Lock	ohne Zeit-Stop · T1 = Stop · ohne frühen Lock · alle drei	**−1,36 bis −1,72** — alle vier verlieren	trägt nicht
Skala ist 8/10 Ticks zu klein?	16/12 und 24/12 Ticks, je mit und ohne Lock	**−2,01 · −1,56 · −1,39 · −0,89** — alle unter Break-even; Merkmale AUC ≈ 0,50	trägt nicht
Mitwachsende Brackets **in schnellen Märkten**	**Faktor 1,5 und 2,0 im Vollgas**	**−1,86 · −1,87 gegen −1,64; im Vollgas selbst schlechter (−1,46 gegen −0,11)**	trägt nicht

Was daraus folgt

Die Einstiege selbst tragen keine Kante. Bei einem Chance-Risiko-Verhältnis nahe 1 ist ein Einstieg ohne Information ein Münzwurf minus Kosten — und genau so sehen alle Varianten aus. Kein Management, keine Skala und keine Richtung rettet ihn.

Konsequenz: Die Strategie wurde **nicht** gedreht, die Brackets wurden **nicht** vergrößert. Die Arbeit verlagerte sich auf die Frage, *wer* die Richtung bestimmt — und ob die Messgröße darunter überhaupt stimmt (Kapitel 29 und 30).

Die Falle, die der Apparat abgefangen hat

Die schnelle Näherung „Vorzeichen der Ergebnisse drehen" ergab **+0,95 Ticks** und 57,3 % Treffer — eine Umkehr-Strategie hätte sich wie ein Fund angefühlt. Aber der gespiegelte Trade hat seinen eigenen Stop, und der liegt genau dort, wohin der Preis vorher lief. Auf dem echten Preispfad blieb davon nichts. Widerrufen am 10. September — bevor eine Zeile Code geändert war.

Warum ein Replay nur töten, nicht stützen kann: Die Varianten wurden aus den Ausstiegsgründen des Originals abgeleitet — also aus dem Ergebnis. Die Vorregistrierung nennt den Lauf deshalb ausdrücklich *deskriptiv*: Ein positives Resultat wäre nur ein Anlass für einen Forward-Test derselben Variante gewesen, kein Beleg. Ein negatives Resultat für *alle* Varianten dagegen sagt genau eine Sache — und die ist eindeutig.

ZEHN TAGE IM SEPTEMBER · II

Die Börse als Schiedsrichter — 97,5 % statt 69,9 %

Seit dem 10. September liegen **echte Orderdaten der CME** vor — jede einzelne Order mit ihrer Seite. Zum ersten Mal ließ sich prüfen, ob unsere Regel, die jedem Abschluss einen Aggressor zuweist, stimmt: nicht gegen sich selbst, sondern **gegen die Börse**. Jede Größe, die aus dem Delta abgeleitet ist — Absorption, Erschöpfung, CVD, Divergenz, Orderfluss — hängt daran.

FRAGE	GRUNDMENGE	ERGEBNIS GEGEN DIE BÖRSENWAHRHEIT	FOLGE
Welche Regel trifft den Aggressor?	7 Tage 94–96 % Abdeckung	Quote-Regel **97,5 %** des Volumens richtig, Tick-Regel nur **69,9 %** — Tages-Delta mit falschem Vorzeichen an 4 von 7 Tagen	Produktentscheidung: Delta auf die Quote-Regel
Gesperrte Notierung Bid = Ask	7 Tage	Nachbar-Notierung trifft 40,2 % bzw. 14,3 % — *unter* Zufall; erklärt 104 % der Restabweichung	„Seite nicht bestimmbar" statt raten
Absorption und Erschöpfung echt oder Regel-Artefakt?	7 Tage	Vorzeichen nur zu 74,1 % einig; die Tick-Fassung meldet Erschöpfung in 41–49 % der Sekunden, die Wahrheit in 34 %	wesentlich Artefakt
Trägt die *wahre* Erschöpfung?	3 Tage 7.723 Ereignisse	AUC 0,510 gegen 0,510 — die korrekte Regel macht sie wahr, aber nicht nützlich	kein Signal verloren
Große Prints als KI-Merkmal? Seite der ≥ 50er an der Einstiegsstufe	30 Tage	AUC 0,503; über dem Null-Band an 12 von 30 Tagen; das Vorzeichen wechselt von Tag zu Tag	bleibt Anzeige, kein Merkmal
Bestätigungstag nach dem Einbau	**11. 9. 780.123 Fills**	**Produkt = Nachrechnung an 100 % der Fills · 97,0 % richtig · je Minute r 0,985**	**ausgeliefert: Indikator, App, KI**

Warum das der wichtigste Befund des Monats ist

Fast ein Drittel des Volumens stand auf der falschen Seite. Jede Studie der Tick-Ära hat auf dieser Grundlage gerechnet — dass viele von ihnen leer blieben, kann teils genau daran liegen. Das beweist nichts in die andere Richtung, aber es erzwingt eine Regel: **alte und neue Tage werden nie gepoolt.** Gate 1 startete deshalb am 12. September neu, nur auf Tagen unter der Quote-Regel.

Der Irrtum davor — zirkulär gemessen

Im Juli stand im Projekt, die Tick-Regel schlage die Quote-Varianten. Die Zahl dahinter (r = 0,965) maß die Übereinstimmung der Tick-Regel mit den *eigenen* Indikatorfeldern — ein Messgerät, das sich selbst abliest. Gegen die Börse: Quote r = 0,983, Tick r = 0,677. Widerrufen am 10. September, im selben Arbeitsschritt.

Die Konsequenz mit Preis: Die neue Regel verändert die Bedeutung der Buch-Merkmale, auf denen CORTEX lernt. Ein Modell, das auf zwei Definitionen lernt, lernt deren Unterschied. Deshalb wurde der Lernstand neu gestartet — beim Gründer am 11. September, mit Release 0.5.30 am 23. September für alle.

DIE NEUE ÄRA

ONE CORTEX entscheidet — und vorher wurde gemessen, was ihn bremst

Zwischen dem 14. und dem 22. September wurde die Signalkette umgebaut: Die Richtung kommt **allein von CORTEX**, ein CORTEX-Gegensignal kann den Trade vor dem ersten Ziel beenden, die strategie-eigenen Sperren protokollieren nur noch. Am 23. September ging die neue Ära als **Release 0.5.30** an alle Kunden — mit einem Lernstand, der bei null beginnt.

MESSUNG	ERGEBNIS	KONSEQUENZ
Merkmals-Inventur 22. 9., vor dem Neustart	**6 von 49** Eingängen praktisch tot (Schwellen über der realen Verteilung); kein Merkmal trennt aufwärts von abwärts für sich allein (AUC 0,47–0,51)	vorregistriert, nicht still korrigiert
Signalkette Buch → Signal → Order	drei Drosseln (3.000 / 1.500 / 2.000 ms) ⇒ **3.290 ms**; danach **≈ 208 ms** — aus Taktzeiten gerechnet, nicht Ende-zu-Ende gemessen	Drosseln entfernt; Messpunkt am Einstieg fehlt noch
Rechenzeit	45 ms am kalten Modell — aber **348 ms** im Handel, weil das Training im Takt lief	Training im eigenen Prozess: warm ≈ 200 ms, Ziel < 100 ms verfehlt
Einstieg	CORTEX ≥ 55 % · Buch ≤ 200 ms alt, Schieflage ≥ 10 % gleichgerichtet · Target-Odds-Tor erst ab 3.000 Labels	übrige Sperren nur im Schatten
CORTEX-Exit	Gegensignal ≥ 55 % für ≥ 300 ms vor T1 ⇒ glattstellen; erster Tag: 1 von 4 Trades	Urteil ab 5 Tagen und 30 Trades
HUD-Ereignisse **vier Vorregistrierungen**	**Umkehrsignal 57,0 % (Untergrenze 46,8 %) gegen 48,2 % · Übernahme 58,5 % (n = 41) · Absorptions-Preis hält 48,6 % gegen 46,5 % · stärkste Ereignisse 45,7 % gegen 50,1 %**	**alle „trägt nicht" — bleiben Anzeige, mit ehrlicher Legende**

Der Bruch — dreifach dokumentiert

Der Umbau geschah bewusst *vor* dem Urteil seiner Vorregistrierung — auf Anweisung des Gründers, festgehalten in Vorregistrierung, Projektstand und Quelltext. Die Erwartungen standen vorher fest: mehr als 6,6 Einstiege je Tag, Exit-Anteil 10–30 %. **Nicht gebaut** wurde, was vorab zu schwach gemessen war: eine Bracket-Weite aus dem Bracket-Kopf (AUC 0,54 über 28.986 Proben).

Was der Neustart kostet — und warum er richtig ist

Mit 0.5.30 beginnt jedes Modell bei null — beim Kunden wie beim Gründer. Kein vortrainiertes Modell, keine fremden Trades. Der Preis ist die bisherige Lerngeschichte. Der Gewinn: ein Modell, das auf **einer einzigen, gegen die Börse geprüften Definition** lernt — und ein Ergebnis, das dann eindeutig CORTEX gehört, nicht einem Regelwerk drumherum.

Zwei Befunde, die ein Kunde hätte sehen können: Der Spoof-Melder stand in praktisch jeder Sekunde an — er maß eine Richtung, keinen Anteil; das Spoof-Sperrtor ging in den Schatten. Und der Wand-Tracker zählte Stornos als Absorption. Beides fiel beim Anlegen von Vorregistrierungen auf, bevor die eigentliche Frage gerechnet war.

ZERTIFIZIERUNG

Warum ein Zertifikat möglich ist — und wer es ausstellen könnte

Die Frage lautet nicht „gibt es eine Behörde, die Handelsvorteile zertifiziert?" — die gibt es nicht. Die Frage lautet: **Welche prüfbaren Eigenschaften muss ein Ergebnis haben, damit etablierte Prüfinstanzen es bestätigen können?** Darauf gibt es klare Antworten.

Vier reale Wege, jeder mit klarem Geltungsbereich

WEG	WAS BESTÄTIGT WIRD	VORAUSSETZUNG — UND OB SIE HEUTE ERFÜLLT IST
Wirtschaftsprüfer Vereinbarte Prüfungshandlungen (ISAE 3000 / AT-C 105)	Dass ein definiertes Rechenverfahren auf definierten Rohdaten **exakt das ausgewiesene Ergebnis** liefert.	**erfüllbar.** Rohdaten sind archiviert, jedes Werkzeug ist quelloffen im Projekt, jedes Ergebnis trägt einen SHA-256-Fingerabdruck über Ergebnis *und* Datenherkunft. Ein Prüfer rechnet nach — stimmt der Fingerabdruck, ist nichts nachträglich verändert worden.
Akademische Begutachtung Peer Review / Preprint	Dass die **Methodik** dem Stand der Forschung entspricht und die Schlussfolgerung trägt.	**erfüllbar.** Vorregistrierung, Permutations-Nullverteilungen, kreuzvalidierte Robustheit und Placebo-Kontrollen sind genau die Elemente, nach denen begutachtet wird. Das negative Ergebnis ist dabei kein Hindernis — **publizierte Negativbefunde sind in diesem Feld selten und gefragt.**
ISO/IEC 42001 KI-Managementsystem	Dass der **Entwicklungs- und Prüfprozess** für das KI-System nachvollziehbar, dokumentiert und überwacht ist.	**weitgehend vorbereitet.** Änderungsverfolgung, Risikoregister, Freigabekriterien und ein Verfahren für widerlegte Befunde existieren bereits. Zu ergänzen wären formale Rollen- und Verantwortungsdefinitionen.
Unabhängige Nachrechnung Dritte Partei, Rohdaten + Protokoll	Dass ein **fremdes Team** aus denselben Rohdaten dasselbe Ergebnis erhält.	**erfüllbar.** Dies ist die stärkste Form und der eigentliche Zielzustand (Stufe 6 der Beweisleiter). Sie setzt voraus, dass die Rohdaten weitergegeben werden können — was bei selbst aufgezeichneten Marktdaten der Fall ist.

Was ausdrücklich *nicht* zertifizierbar ist

Künftige Erträge. Keine Instanz der Welt bestätigt, dass ein Verfahren morgen funktioniert. Wer das anbietet, ist unseriös.

Zertifizierbar ist immer nur: *Dieses Verfahren hat auf diesen Daten, unter diesen vorab festgelegten Bedingungen, dieses Ergebnis erzielt, und es wurde korrekt gerechnet.*

Genau das ist aber der entscheidende Unterschied zu allem, was der Markt heute bietet.

Die drei Eigenschaften, die alles ermöglichen

1 · Manipulationssicher. Jedes Ergebnisdokument trägt einen kryptografischen Fingerabdruck über Ergebnis und Datenherkunft. Neu berechnen ergibt denselben Hash — oder deckt eine Änderung auf.

2 · Reproduzierbar. Rohdaten, Werkzeug und Parameter sind archiviert und versioniert.

3 · Vorregistriert. Die Kriterien standen fest, bevor die Daten gesehen wurden — nachweisbar über datierte Einträge im Versionsverlauf.

Der Kern in einem Satz: Zertifizierbarkeit entsteht nicht durch ein gutes Ergebnis, sondern durch einen prüfbaren **Weg** zum Ergebnis. Dieser Weg existiert bei OrderFlowAi heute vollständig — er wartet nur darauf, dass am Ende etwas Positives steht.

ZERTIFIZIERUNG

Das Dossier — was heute bereits existiert

Ein Zertifizierungsdossier lässt sich nicht rückwirkend erstellen; es muss **mitlaufen**. Deshalb wurde die Beweisführung von Anfang an als Maschine gebaut, nicht als Dokument. Diese Seite listet auf, was ein Prüfer heute vorfinden würde.

1 · Das Edge-Validierungs-Zertifikat

Ein Programm erzeugt aus der jeweils aktuellen Studie ein vollständiges Dokument — als Text, als gestaltetes HTML und als PDF. Es enthält:

· **Datenherkunft** — welche Handelstage, wie viele, je Marktphase · **Merkmalsgesundheit** — welche Eingangsgrößen überhaupt Information trugen · **Das vorregistrierte Verdikt** nach festgelegtem Paragrafen · **Kreuzvalidierte Robustheitsverteilung** (CPCV) · **p-Wert gegen die Nullverteilung** · **Leckage-Prüfung** über einen Parameterbereich · **Geltungsbereichs-Vorbehalte** — was die Aussage *nicht* abdeckt · **SHA-256-Fingerabdruck** über Ergebnis und Herkunft

2 · Das Strategie-Änderungs-Zertifikat

Das Gegenstück für Regeländerungen: Erwartungswert vorher gegen nachher, Anzahl Trades, Bootstrap-Signifikanz der Verbesserung, vollständige Aufstellung aller unterdrückten Signale samt näherungsweiser Gegenrechnung — **und dieselben Vorbehalte und denselben Fingerabdruck.**

3 · Das Datenintegritäts-Zertifikat mit Historie

Eine tägliche Ampel über sechs unabhängige Qualitätswächter — mit fortgeschriebener **Historie**. Der Grund dafür ist der wichtigste Teil: Wenn im August eine Auswertung seltsam aussieht, lautet die Frage „wie war die Datenqualität an den Tagen, die hineinflossen?". Ohne datierte Aufzeichnung ist das im Nachhinein nicht mehr rekonstruierbar.

4 · Die laufenden Hauptbücher

Fortgeschriebene, nur anfügbare Register — jedes mit eingefrorener Nullhypothese in **jeder** Zeile:

· **Zielerreichungs-Register** — die aktuelle Spur
· **Kopf-Register** — jeder Vorhersagekopf mit seinem Urteil
· **Lern-Register** — Modellalter und Leistung je Sitzung
· **Sperren-Register** — jede Filterregel mit ihrer gemessenen Wirkung
· **Aufzeichnungs-Register** — Vollständigkeit je Handelstag
· **Buchabdeckungs-Register** — verwertbare Tage

5 · Das Widerrufsregister

54 Einträge, maschinell überwacht. Für einen Prüfer ist das **der wertvollste Teil des gesamten Dossiers** — es ist der Nachweis, dass das System Fehler nicht nur machen, sondern auch finden und zurücknehmen kann.

6 · Das Abendprotokoll

22 Schritte, jeden Handelstag, automatisiert. Von der Fingerabdruck-Prüfung der installierten Software über die Aufzeichnungsvollständigkeit bis zu den Registern und der nächtlichen Studie.

17 der 22 Schritte haben eine fachliche Abnahme — sie prüfen nicht nur, *ob* etwas lief, sondern *ob das Ergebnis stimmt*. Die fehlenden fünf sind benannt und bewusst offen ausgewiesen, statt weggeeicht zu werden.

Der Unterschied zu einem üblichen Backtest-Bericht: Ein Backtest-Bericht ist eine Momentaufnahme, die jemand erstellt hat. Dieses Dossier ist ein laufender Prozess, den niemand anhalten kann, ohne dass es auffällt. **Man kann es nicht schönen — man kann es nur abschalten, und das wäre sichtbar.**

ZERTIFIZIERUNG

Der Erstheitsanspruch — was genau wäre „das erste Mal"?

Ein Anspruch auf Erstmaligkeit muss präzise formuliert sein, sonst hält er keiner Prüfung stand. Diese Seite formuliert ihn so eng, dass er verteidigbar ist — und benennt zugleich, was er **nicht** behauptet.

Die verteidigbare Formulierung

„Die erste am Markt erhältliche Orderfluss-Analysesoftware für Privatanwender, deren Vorhersagekomponente einem vorregistrierten, placebo-kontrollierten und von unabhängiger Seite nachrechenbaren Prüfverfahren unterzogen wurde — mit veröffentlichtem Ergebnis, unabhängig davon, ob dieses positiv oder negativ ausfiel."

Was dieser Anspruch beinhaltet

- **Vorregistriert** — Kriterien datiert vor der Auswertung
- **Placebo-kontrolliert** — mit ausgewiesener Kontrollgruppe
- **Nachrechenbar** — Rohdaten + Werkzeug + Fingerabdruck
- **Veröffentlicht** — auch bei negativem Ausgang
- **Am Markt erhältlich** — kein Forschungsprototyp
- **Für Privatanwender** — nicht institutionell abgeschirmt

Was er ausdrücklich nicht behauptet

- **Nicht**: „die erste KI im Trading" — das wäre falsch
- **Nicht**: „die erste mit einem Vorteil" — unbewiesen
- **Nicht**: „die genaueste" — nicht vergleichend gemessen
- **Nicht**: Aussagen über künftige Erträge
- **Nicht**: dass institutionelle Häuser das nicht intern tun — sie tun es vermutlich, nur eben **nicht öffentlich prüfbar**

Warum die Enge des Anspruchs seine Stärke ist

Ein weiter Anspruch („die beste KI der Welt") ist wertlos, weil er unprüfbar ist und bei der ersten kritischen Nachfrage zerfällt. Ein enger Anspruch ist verteidigbar — und deshalb **zitierfähig**: in Fachartikeln, in Vertriebsgesprächen, gegenüber Regulierern, in einer Due-Diligence.

Hinzu kommt: Der enge Anspruch ist bereits **zur Hälfte eingelöst**. Der Apparat existiert, die Prüfungen laufen, die Ergebnisse werden dokumentiert — auch die negativen. Was fehlt, ist nur der Abschluss des laufenden Prüfzyklus.

Der historische Vergleich

In der Medizin führte die Einführung vorregistrierter Studien nicht dazu, dass bessere Medikamente entstanden. Sie führte dazu, dass man **zum ersten Mal wusste, welche wirken** — und der Markt bereinigte sich innerhalb weniger Jahre.

Die Finanzsoftware für Privatanwender hat diesen Schritt noch vor sich. Wer ihn als Erster geht, definiert den Maßstab, an dem alle anderen anschließend gemessen werden.

Belegbar heute

Vorregistrierung datiert im Versionsverlauf · Placebo-Kontrollen als Pflichtzeile · Fingerabdruck je Ergebnis · 22 veröffentlichte Widerrufe

Belegbar nach dem bestätigenden Endpunkt

Der bestätigende Endpunkt unter eingefrorenen Bedingungen — positiv oder negativ, in beiden Fällen zitierfähig

Belegbar nach externer Prüfung

Nachrechnung durch eine dritte Partei aus denselben Rohdaten — Stufe 6 der Beweisleiter

Die strategische Pointe: Selbst wenn die aktuelle Spur negativ endet, bleibt der Erstheitsanspruch bestehen — er hängt am **Verfahren**, nicht am Ergebnis. Damit ist er der einzige Teil dieses Projekts, der nicht scheitern kann.

DIE EVOLUTIONSSTUFE

Der autonome Mitgründer — was passiert, wenn die KI mitdenkt statt nur auszuführen

Dieses Projekt wird von **einer Person** und **einem KI-Agenten** gebaut. Das allein ist 2026 nicht mehr bemerkenswert. Bemerkenswert ist die Betriebsart: Der Agent ist nicht als Werkzeug eingerichtet, sondern als **Mitgründer mit Pflicht zur Eigeninitiative**.

Die vier Regeln des Modus

1. **Kontext-Routing.**
   Je nach bearbeitetem Bereich lädt der Agent selbstständig das passende Fachwissen — Marktmikrostruktur und Quantitative Methoden für die Analyse-Engine, Web-Architektur und Konversionsoptimierung für die Vermarktung.
   Ohne Aufforderung.

2. **Architekt statt Ausführer.**
   Der Agent zieht aus dem Projektstand eigene Schlüsse und sucht permanent nach dem statistischen oder technischen Vorteil — auch dann, wenn die gestellte Aufgabe eine andere war.

3. **Pflicht-Block am Ende jeder Antwort.**
   Jede substanzielle Arbeitseinheit endet mit einem festgelegten Abschnitt `[Autonome Projekt-Evolution]`, der zwei Dinge enthalten *muss*: einen konkreten Verbesserungsimpuls für den gerade bearbeiteten Bereich und einen
   Feature-Vorschlag
   Richtung Fernziel.

4. **Kein Code ohne Freigabe.**
   Vorschläge werden erst nach ausdrücklichem „Go" umgesetzt. Der Mensch behält die Entscheidung, die KI übernimmt das Weiterdenken.

Warum Regel 3 der eigentliche Hebel ist

Ein Agent, der nur Aufträge abarbeitet, denkt nie über den Auftrag hinaus. Der erzwungene Abschlussblock kehrt das um: **Am Ende jeder Arbeitseinheit muss ein Schritt weiter gedacht worden sein.** Über 395 Sitzungen ergibt das mehrere hundert unaufgeforderte Verbesserungsvorschläge — von denen ein erheblicher Teil in genau die Werkzeuge floss, die in Teil III beschrieben sind.

Das Widerrufsregister, die Placebo-Pflichtzeile, der Geltungsbereichs-Wächter und die Abnahme-Inventur entstanden **alle** aus diesem Block — keiner davon war ein Auftrag.

Das persistente Gedächtnis

Der Agent verfügt über eine dateibasierte Wissensbasis mit heute **808 Einträgen** — Nutzerpräferenzen, Projektentscheidungen, technische Fallen, widerlegte Annahmen, jeweils mit Begründung und Querverweisen.

Der Effekt: **Ein Fehler wird höchstens zweimal gemacht.** Beim zweiten Mal entsteht ein Eintrag, und beim dritten Mal schlägt ein automatischer Prüfer an.

Genau darauf beruht auch die Selbstkorrektur-Fähigkeit, die dieser Bericht durchgehend dokumentiert.

Die Selbstbindung, die daraus wuchs

Bemerkenswert ist die Richtung der Entwicklung. Der Agent hat sich über die Sitzungen hinweg selbst **immer strengere Regeln** auferlegt:

· kein Urteil vor der Gegenprüfung
· jede Quote nur mit Nullhypothese
· jede Kennzahl nur mit Zerfallskurve
· jeder Wächter nur mit Selbsttest
· jede Falle sofort mit automatischem Prüfer

Keine dieser Regeln wurde vorgegeben. Alle entstanden aus der Analyse eigener Fehler — und alle machen die Arbeit langsamer und das Ergebnis belastbarer.

Die ehrliche Einschränkung

Derselbe Agent hat auch die 54 widerrufenen Befunde produziert. Die Betriebsart erzeugt **mehr** Hypothesen — richtige wie falsche. Ihr Wert liegt nicht in höherer Treffsicherheit, sondern darin, dass die Prüfmaschinerie mitwächst.

DIE EVOLUTIONSSTUFE

Der Verstärkungseffekt — gemessen, nicht behauptet

Die These lautet: Weil das Projekt von einem KI-Agenten mitentwickelt wird, wächst seine Entwicklungsgeschwindigkeit und -qualität mit jeder Modellgeneration. Diese These ist prüfbar — und wir haben sie am eigenen Versionsverlauf gemessen.

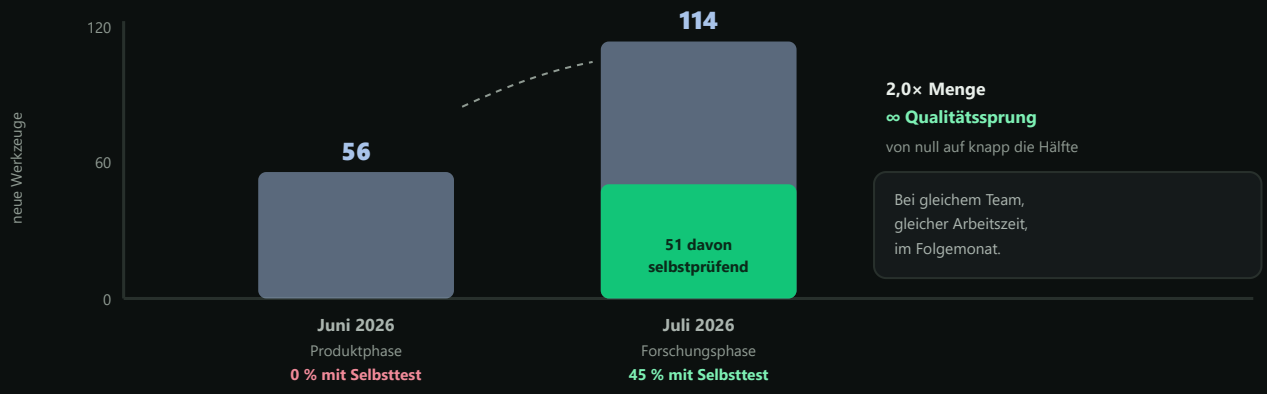


Abbildung 10: Neu entstandene Analysewerkzeuge je Monat, aus dem Versionsverlauf gezählt. Der grüne Anteil sind Werkzeuge mit eingebautem Selbsttest.

Die ehrliche Einordnung

Diese Zahlen belegen **nicht** allein einen Modelleffekt. Im Juli fiel zusätzlich der Methodenwandel: Nach dem Bruch (Seite 11) wurde von Produktbau auf Forschung umgestellt, und Forschung braucht naturgemäß mehr Messwerkzeuge.

Was die Zahlen aber sehr wohl belegen: Die Umstellung von 0 % auf 45 % Selbsttestquote geschah *ohne* zusätzliche Personen, *ohne* mehr Zeit und *ohne* externe Vorgabe. Sie geschah, weil die Regel „jeder Wächter braucht einen Selbsttest" im Verlauf des Monats entstand — und dann konsequent angewandt wurde.

Warum daraus ein Zinseszins wird

Der Effekt ist nicht linear, weil jedes neue Werkzeug die **Fehlerentdeckungsrate** aller künftigen Arbeit erhöht. Ein Beispiel aus dem Juli:

Der Geltungsbereichs-Wächter fand einen Fehler → daraus entstand die Abnahme-Inventur → die fand fünf Schritte ohne fachliche Prüfung → einer davon deckte auf, dass die installierte Software nie gegen ihre Prüfsumme verifiziert wurde.

Ein Werkzeug erzeugte drei weitere — und deckte dabei eine Projektregel auf, die seit einem Jahr galt und nie gemessen worden war.

Was sich dadurch am Rollenbild ändert. Der Engpass eines Forschungsunternehmens war bisher die Zahl der Menschen, die gleichzeitig sorgfältig denken können. Verschiebt sich ein Teil davon in einen Agenten, der nicht ermüdet, nichts vergisst und aus jedem Fehler einen dauerhaften automatischen Prüfer macht, verschiebt sich der Engpass auf etwas anderes: **auf die Zahl der Handelstage, die vergehen müssen.** Genau dort steht dieses Projekt heute — es wartet nicht auf Entwicklungsarbeit, sondern auf den Kalender. Das ist ein guter Ort, um zu warten.

Die Prognose, klar als Prognose gekennzeichnet: Wenn jede Modellgeneration die Qualität autonomer Beiträge erhöht, verschiebt sich die Grenze dessen, was ein Ein-Personen-Unternehmen an Forschungstiefe leisten kann, weiter nach oben. Der bisher gemessene Verlauf ist mit dieser Erwartung vereinbar. **Er beweist sie nicht** — dafür braucht es mehr als zwei Monate. Aber genau diese Unterscheidung zwischen „vereinbar mit" und „bewiesen" ist der Grund, warum dieser Bericht überhaupt Gewicht hat.

WERT

Die Wertschöpfungskette — fünf Stufen, vier davon unabhängig vom Edge

Der entscheidende strategische Zug vom 20. Juli 2026 war die **Entkopplung**: Das Produkt verkauft, was nachweislich funktioniert. Die Forschung läuft getrennt. Dadurch hängt der Unternehmenswert **nicht** am Ausgang eines einzelnen Experiments.

Abbildung 11: Vier von fünf Stufen tragen unabhängig davon, ob der Edge-Nachweis gelingt. Stufe 5 ist kein Fundament, sondern ein Multiplikator.

STUFE	WAS VERKAUFT WIRD	WARUM ES TRÄGT	STATUS
1	**Softwarelizenzen** Pro · Trial · Institutional	Level-2-Visualisierung, Wärmebild, Fußabdruck, DOM-Analyse, Handelsjournal. Kunden kaufen **Sichtbarkeit**, nicht Prognose — und diese Sichtbarkeit funktioniert nachweislich. Offizielles NinjaTrader-Ökosystem-Listing als externe Bestätigung.	live
2	**Der Datensatz**	Vollständig aufgezeichnete Orderbuch- und Abschlussdaten mit Herkunftskennzeichnung je Zeile, durchgehend seit 2025. **Nicht nachträglich beschaffbar** — der Wert wächst mit jedem Tag, an dem ein Wettbewerber nicht aufzeichnet.	wächst
3	**Die Methodik**	Der Beweisapparat selbst — 170 Werkzeuge, Vorregistrierung, Zertifikatsgenerierung, Widerrufsregister. Übertragbar auf jede quantitative Fragestellung. Denkbar als Lizenz, als Beratungsleistung oder als eigenständiges Produkt.	Option
4	**Signal-Schnittstelle**	Maschinenlesbarer Zugang zur Analyse-Engine für institutionelle Abnehmer. Technisch vorbereitet; wirtschaftlich sinnvoll erst mit belastbarem Nachweis.	vorbereitet
5	**Der zertifizierte Edge**	**Bei positivem Endpunkt: eine belegte, extern nachrechenbare Aussage über einen statistischen Vorteil. Der Preis eines Werkzeugs, das nachweislich wirkt, ist ein anderer als der eines Werkzeugs, das schön aussieht.**	offen

Für die Bewertung entscheidend: Ein Investment in OrderFlowAi ist keine Wette auf Stufe 5. Es ist der Erwerb eines laufenden Softwaregeschäfts mit einem proprietären, nicht nachholbaren Datensatz und einem Forschungsapparat, der als eigenständiger Vermögenswert verwertbar ist — **mit einer kostenlosen Option auf Stufe 5 obendrauf.**

WERT

Risiken, ehrlich benannt

Ein Bericht, der 22 eigene Irrtümer dokumentiert und dann bei den Risiken schweigt, wäre unglaubwürdig. Diese Seite benennt, was schiefgehen kann — und was dagegen bereits getan ist.

RISIKO	SCHWERE	EINSCHÄTZUNG UND GEGENMASSNAHME
Der Edge existiert nicht	hoch	Das ist der wahrscheinlichste Einzelausgang — die bisherige Befundlage ist überwiegend negativ. **Abgefedert durch die Entkopplung:** Vier von fünf Wertschöpfungsstufen tragen unabhängig davon. Ein negativer Ausgang kostet Wochen, nicht das Unternehmen.
Schlüsselpersonen-abhängigkeit	hoch	Ein Gründer, ein Agent. Das ist der stärkste Beschleuniger und die größte Verwundbarkeit zugleich. **Teilweise abgefedert** durch die außergewöhnliche Dokumentationstiefe: 808 Wissenseinträge, jede Entscheidung mit Begründung, jede Falle mit automatischem Prüfer. Ein neues Team fände einen ungewöhnlich gut übergebbaren Zustand vor — **aber es ersetzt keine Person.**
Zeitbedarf der Prüfkette	mittel	Der bestätigende Endpunkt braucht 20 Handelstage *unter unveränderten Bedingungen*. Jede Änderung setzt den Zähler zurück. Das ist methodisch zwingend und praktisch unbequem — **und es ist genau die Disziplin, die den Nachweis später belastbar macht.**
Marktveränderung entwertet Befunde	mittel	Ein Vorteil, der heute existiert, kann morgen verschwinden. **Adressiert** durch den Stationaritätstest, der vorwärts *und* rückwärts prüft, und durch das Prinzip, Befunde mit ihrer Halbwertszeit auszuweisen.
Wettbewerber kopieren die Methodik	gering	Technisch möglich, praktisch unwahrscheinlich. Die Methodik ist unbequem: Sie verlangt, eigene Erfolgsmeldungen zurückzunehmen. Ein Anbieter mit Vertriebsdruck wird das nicht tun. **Der Wettbewerbsvorteil ist kulturell, nicht technisch** — und deshalb schwer kopierbar.
Regulatorische Anforderungen	gering	Die Software ist ein Analysewerkzeug ohne Anlageberatung und ohne Vermögensverwaltung. Risikohinweise sind auf allen kundensichtbaren Flächen umgesetzt. Der KI-Rechtsrahmen der EU ist berücksichtigt; das Managementsystem ist auf ISO/IEC 42001 vorbereitet.
Datenqualität am Messgerät	gering	Historisch die häufigste Fehlerquelle — drei der letzten fünf Widerrufe. **Heute am stärksten abgesichert:** Herkunftskennzeichnung je Datenzeile, Duplikatprüfung, Vollständigkeitsprüfung vor Handelsbeginn, tägliches Integritätszertifikat mit Historie.

Das größte Restrisiko in einem Satz: Es ist nicht, dass der Edge nicht existiert — darauf ist das Unternehmen vorbereitet. Es ist, dass jemand aus Ungeduld die Prüfkette abkürzt und ein ungeprüftes Ergebnis nach außen trägt. **Genau dagegen wurde der gesamte in Teil III beschriebene Apparat gebaut** — inklusive der Regel, dass eine Änderung an den Kriterien im Versionsverlauf sichtbar bleibt.

STANDORT

Wo wir heute stehen — in Zahlen, nicht in Absichten

Dieses Kapitel ist der Kompass. Es beantwortet drei Fragen: **Wo stehen wir? Was fehlt bis zum Urteil?** Und: **Worauf trifft die Strategie ihre Entscheidungen?** Alle Zahlen sind Messwerte vom 23. September 2026; wo ein älterer Messpunkt gilt, steht sein Datum dabei.

Die vier laufenden Beweisketten

① Der bestätigende Endpunkt — „Bracket-MinP"

Unterschrieben am 5. September, zweiseitig. 14 Tagesdateien liegen vor, die Hürde von 10 Tagen ist erreicht — **der eine vorregistrierte Blick steht noch aus**. Umfeld: die Kalibrierung vom 10. 9. zeigte, dass der Kopf nicht trennt (oberstes Dezil vorhergesagt 78 %, real 33 %); seit dem 23. 9. lernt er von null. Der Anker hat **nie gezählt** (0 von 20).

② Gate 1 — das Buch als zweiter Eingang, Revision 3

Neu gestartet am 12. September, **nur auf Tagen unter der Quote-Regel** (Kapitel 29). Stand: **8 von 10** Buch-Tagen, Urteil um den 25.–26. September. Die Vorgängerfassung lautete auf 61 Tagen der Tick-Ära **NICHT_LERNBAR** — sie wird nicht mehr gepoolt.

③ CORTEX — das lernende Modell

ONE CORTEX: 49 Merkmale, fünf Köpfe, seit Release 0.5.30 **allein** für die Richtung zuständig und seit dem 23. September mit frischem Lernstand. Das letzte belastbare Urteil über die alte Generation (29. Juli: **KEIN EDGE**, MCC 0,0201) bleibt die Vergleichslinie. Vor dem Neustart gemessen: 6 von 49 Eingängen praktisch tot (Kapitel 30).

④ Die Schatten-Sperren

Sie melden, greifen aber nicht ein. **DELTA-MISMATCH** ist stärker geworden: markierte Einstiege gewinnen **29 %** (23 von 80) gegen **50 %** unmarkierte, `p = 0,0001`, 27 Tage. Die Zahl ist **explorativ** — sie poolt mehrere Versionen und die Umstellung auf die Quote-Regel. Eine Schranke ist nicht registriert.

Der Apparat, der das absichert

Werkzeuge mit Selbsttest (348 grün)
350
Einzeltests
4.829
Vorregistrierungen
43
Widerrufene eigene Befunde
54
Arbeitssitzungen
471

Was seit dem letzten Berichtsstand (5. September) dazukam

Die Replay-Maschine (Kapitel 28): 105 Trades tickgenau nachgespielt — Umkehr, Struktur, Skala und mitwachsende Brackets verlieren alle. Die Einstiege tragen keine Kante.

Die Börse als Schiedsrichter (Kapitel 29): Die Quote-Regel trifft den Aggressor zu 97,5 %, die bisherige Tick-Regel zu 69,9 %. Seit dem 11. September ist das Produkt umgestellt.

Die neue Ära (Kapitel 30): CORTEX entscheidet allein, die Signalkette ist von 3,3 s auf rund 0,2 s verkürzt, der Lernstand beginnt für alle bei null. Und der letzte Kandidat aus der Wochenendstudie ist nach Regel verworfen (Kapitel 27).

Die unbequeme Zeile. Das *konfirmatorische* Hauptergebnis ist weiterhin **negativ**, und alle drei Forward-Kandidaten sind inzwischen verworfen. Neu ist nicht ein Fund, sondern die Messbasis: **zum ersten Mal ist das Delta gegen die Börse geprüft**.

DER WEG ZUM URTEIL

Was fehlt — und worauf die Strategie jetzt entscheidet

Das Programm kennt **zwei zulässige Ausgänge**, und beide sind ein Ergebnis: **A — ein nachgewiesener Vorteil. B — ein belastbares Nein.** Ein drittes Ergebnis („wir wissen es immer noch nicht") wäre das einzige Scheitern. Diese Seite sagt, was jeder der beiden Ausgänge noch braucht.

Ausgang A — der Vorteil ist nachgewiesen

1. **Der eine Blick auf „Bracket-MinP"**
   — die Tage liegen vor. Danach wird das Anspruchsrecht der nächsten Vorregistrierung frei; nie zwei offene Endpunkte zugleich.
2. **`DELTA-MISMATCH` bekommt eine registrierte Schranke**
   und wird auf neuen Tagen geprüft — nur unter der Quote-Regel, damit das starke Signal nicht aus zwei Messgrößen zusammengesetzt ist.
3. **Unabhängige Nachrechnung**
   durch eine dritte Partei auf den Rohdaten. Das Dossier dafür existiert: Vorregistrierung, Widerrufsregister, 4.829 Selbsttests, jede Zahl mit Messpunkt.

Ausgang B — das belastbare Nein

1. **Obere Schranke statt Nullbefund.**
   „Kein Edge gefunden" ist wertlos ohne die Angabe, welcher Edge noch *hineingepasst* hätte. Für die OFI-Richtung steht sie (Skill unter 1,016); für Gate 1 wird sie unter der Quote-Regel neu gerechnet.
2. **Der Geltungsbereich muss stehen.**
   Das Nein deckt inzwischen den punktuellen Merkmalsraum, die OFI-Richtung, die Trend-Fortsetzung — und seit September auch Richtung, Struktur und Skala der Einstiege (Replay).
3. **Die Messgröße muss stimmen.**
   Das Delta ist seit dem 11. September gegen die Börse verankert (97,0 %) — die größte einzelne Vorzeichen-Sicherung des Programms.

Worauf die Strategie jetzt tatsächlich entscheidet

Seit Strategie 0.9.14 (22. September) ist die Kette kurz:

CORTEX ≥ 55 % → Buch ≤ 200 ms alt und gleichgerichtet (Schieflage ≥ 10 %) → Target-Odds-Tor (erst ab 3.000 aufgelösten Labels) → Risiko-Tore: ruhiger Markt, 12–14 Uhr ET, 300 s Mindestabstand, Orderfluss gegen das Signal, Vollgas.
Ausstieg: CORTEX-Exit · Zeit-Stop (120 s vor T1, Runner 600 s) · 15:40 ET flach.

Am 4. September war das Nadelöhr noch das Regelwerk *nach* dem Modell — die Inventur zählte 120 Sperren je Einstieg. Jetzt entscheidet das Modell, und es beginnt bei null. Das ist die ehrlichere Versuchsanordnung: **ein Ergebnis gehört dann eindeutig CORTEX.**

Die Reihenfolge, in der gehärtet wird

1. Aufzeichnung sichern — ohne zitierfähige Tage ist jede Auswertung wertlos. **erledigt**
2. Die Messgröße gegen die Wahrheit prüfen — ein falsches Delta entwertet alles darüber. **Delta verankert**
3. Sperren einzeln nachweisen — jede mit eigener Nullhypothese. **im Schatten, Schranke fehlt**
4. Erst dann das Modell schärfen — ein Kopf, der auf ungesicherten Eingängen lernt, lernt das Rauschen. **Neustart 23. 9.**

Warum diese Reihenfolge nicht verhandelbar ist:
Jede Stufe misst die darunter mit. Wer das Modell vor den Eingängen schärft, optimiert auf einen Messfehler — und merkt es erst, wenn Monate an Daten unbrauchbar sind. Im September ist genau das sichtbar geworden: ein Drittel des Volumens stand auf der falschen Seite.

AUSBLICK

Was als Nächstes geschieht

Die nächsten vier Wochen — Stand 23. September

1. **Gate 1, Revision 3 — Urteil um den 25.–26. September**, auf 10 Buch-Tagen unter der Quote-Regel. Fällt ein Tag aus dem Buch-Vertrag, rückt das Urteil einen Tag nach — die Hürde nicht.
2. **Der eine Blick auf „Bracket-MinP"**, zweiseitig, mit Pflicht-Gegenprobe ohne Tag 1. Danach wird das Anspruchsrecht der nächsten Vorregistrierung frei — nie zwei offene Endpunkte zugleich.
3. **Die neue Signalkette misst sich selbst:** Urteil ab 5 Handelstagen und 30 Trades gegen die vorab festgelegten Erwartungen; alte und neue Kette werden nie gepoolt. Dazu der fehlende Messpunkt: Signalalter am Einstieg.
4. **Sperren und Eingänge zur Reife bringen.** `DELTA-MISMATCH` bekommt eine registrierte Schranke; die sechs toten Merkmals-Eingänge werden vorregistriert behoben — beim nächsten ohnehin nötigen Neustart.
5. **Bei positivem Endpunkt:** Aufnahme des Zertifizierungsprozesses — beginnend mit der unabhängigen Nachrechnung durch eine dritte Partei.

Das Fernziel, unverändert seit Beginn

Die weltweit beste Analyse-KI für Orderfluss — definiert nicht über Funktionsumfang, sondern über eine Eigenschaft, die heute kein Anbieter beansprucht:

dass jede Aussage, die sie trifft, überprüfbar ist.

Der Weg dorthin führt nicht über mehr Funktionen. Er führt über mehr bewiesene Aussagen — und über den Mut, die unbewiesenen als unbewiesen zu kennzeichnen.

Was dieser Bericht belegen sollte

Nicht, dass wir einen Vorteil gefunden haben. Sondern:

· dass wir wissen, **wie man ihn findet**
· dass wir wissen, **wie man sich dabei täuscht**
· dass wir es 54-mal **selbst bemerkt** haben
· und dass daraus ein Apparat entstand, der ein künftiges Ergebnis **zertifizierbar** macht

Das ist die Grundlage, auf der sich ein Unternehmen bauen lässt. Eine unbelegte Erfolgsmeldung ist es nicht.

Eine Bitte an jeden Leser, der dieses Projekt bewertet: Fragen Sie nicht nach der Trefferquote. Fragen Sie nach der Nullhypothese, der Stichprobengröße und danach, wann die Kriterien festgelegt wurden. Wer diese drei Fragen nicht beantworten kann, hat keinen Vorteil gemessen — er hat eine Zahl gefunden. **Wir haben diesen Bericht so geschrieben, dass alle drei Antworten auf jeder Seite stehen.**

Schlusssatz. Dreizehn Monate, 471 Arbeitssitzungen, 54 widerrufene Befunde, ein negatives konfirmatorisches Hauptergebnis, eine gegen die Börse geprüfte Messgröße, ein Neustart des Lernstands — und ein Messapparat, wie ihn dieser Markt bisher nicht kennt. Wir kratzen an der Oberfläche eines Datenraums, den wir als Einzige systematisch aufzeichnen. **Ob dort ein Vorteil verborgen ist, wissen wir nicht. Aber wir sind die Einzigen, die es beweisen könnten — und die Einzigen, die es zugeben würden, wenn nicht.**

OrderFlowAi

OrderFlowAi Research
orderflowai.io/forschung