

Post-mortem: Vier Features, die in die eigene Zukunft schauen

Vier Kontext-Features unserer Episoden-Basis enthielten Information, die zum Einstiegszeitpunkt noch nicht existierte. Betroffen waren 60.129 von 123.193 Episoden. Der Prognose-Spread eines Vorhersagemodells fiel nach dem Fix von +0,726 R auf +0,191 R. Wie wir das Leck gefunden, bewiesen und geschlossen haben — und warum die Modelle, die darauf trainiert wurden, nicht zu retten sind.

Dokument-Metadaten

Feld	Wert
Herausgeber	KruegerAlgorithms
Titel	Post-mortem: Vier Features, die in die eigene Zukunft schauen
Erscheinungsdatum	27. Juli 2026
Version	v1.0
Verdikt	Referenz
Datengrundlage	Datenbasis: 123.193 Episoden, 10 Märkte, 23 Setups, 05.01.2015 – 05.06.2026 . Exit: Trailing-Stop BE 0,5 / TS 1,0 / Schritt 0,5, netto nach Spread und Slippage. Gegenstand sind nicht die Trade-Ergebnisse, sondern die Kontext-Features unserer Episoden-Basis — die Spalten, die ein Modell zum Zeitpunkt des Einstiegs sehen darf. Keine Handelsempfehlung.
Hinweise	Korrektur vom 29. September 2026; Neurechnung vom 5. Oktober 2026. Wortlaut in der Zusammenfassung.
Dokumenttyp	Statistische Auswertung (keine Trading-Empfehlung)

Rein statistische Auswertung historischer Marktdaten. Keine Trading-Empfehlung, keine Finanzberatung. Historische Performance ist keine Garantie für die Zukunft. Trading mit Hebelprodukten kann zum Totalverlust führen.

Zusammenfassung

Korrektur vom 29. September 2026: Abschnitt 4 sagt, alle bisherigen Setup-Verdikte und die Persistenz-Studie blieben gültig, weil kein Trade-Ergebnis berührt wurde. Das hier beschriebene Leck hat tatsächlich kein Trade-Ergebnis berührt, aber die Trade-Ergebnisse selbst stammen aus unserer älteren Backtest-Kette, die auch auf einem Zufallsmarkt ohne Richtung positive R bucht. Diese Verdikte sind deshalb nicht belegt. Details: Der Zufallsmarkt als Lineal.

Neurechnung vom 5. Oktober 2026: Die Setup-Verdikte sind inzwischen neu gemessen: Keine der 25 Kombinationen aus Ausbruchsregel und Markt hat einen Vorteil gegenüber dem Zufallsmarkt. Alle Zahlen: Unsere eigenen Ausbruchsregeln, neu gemessen.

Vier Kontext-Features unserer Episoden-Basis enthielten Information, die zum Einstiegszeitpunkt noch nicht existierte. Betroffen waren 60.129 von 123.193 Episoden. Der Prognose-Spread eines Vorhersagemodells fiel nach dem Fix von +0,726 R auf +0,191 R. Wie wir das Leck gefunden, bewiesen und geschlossen haben — und warum die Modelle, die darauf trainiert wurden, nicht zu retten sind.

Lookahead-Bias ist der Fehler, vor dem jeder warnt und den trotzdem jeder irgendwann baut. Nicht, weil man den Schlusskurs von morgen einliest — so plump ist es nie. Sondern weil eine Kontextgröße, die *für den Tag* definiert ist, an einer Stelle abgefragt wird, an der der Tag noch gar nicht so weit ist.

Genau das ist uns passiert. Vier Spalten unserer Episoden-Basis waren ein Blick in die eigene Zukunft. Dieses Dokument beschreibt, wie das Leck entstanden ist, woran wir es erkannt haben, wie wir es bewiesen und geschlossen haben — und was davon nicht mehr zu reparieren war. Wir schreiben es auf, weil die Fehlerklasse allgemein ist und die Diagnose-Signatur sich auf jeden Backtest übertragen lässt.

Diese Auswertung enthält keine Trading-Empfehlung.

Inhaltsverzeichnis

Zusammenfassung	2
Inhaltsverzeichnis	3
Abbildungen	3
1. Der Fehler	4
2. Woran wir es bemerkt haben	4
3. Der Beweis	5
4. Fix und Verifikation	6
5. Was nicht zu retten war	6
6. Was daraus folgt	6
7. Grenzen	6

Abbildungen

Abb. 1: Was ist zum Entry bekannt? Zwei Feature-Paare wurden gelesen, bevor der Wert existierte: die Opening-Range-Grenzen stehen erst 30 Minuten nach dem Op...

Abb. 2: Die Hälfte aller Episoden trug ein Feature aus der eigenen Zukunft

1. Der Fehler

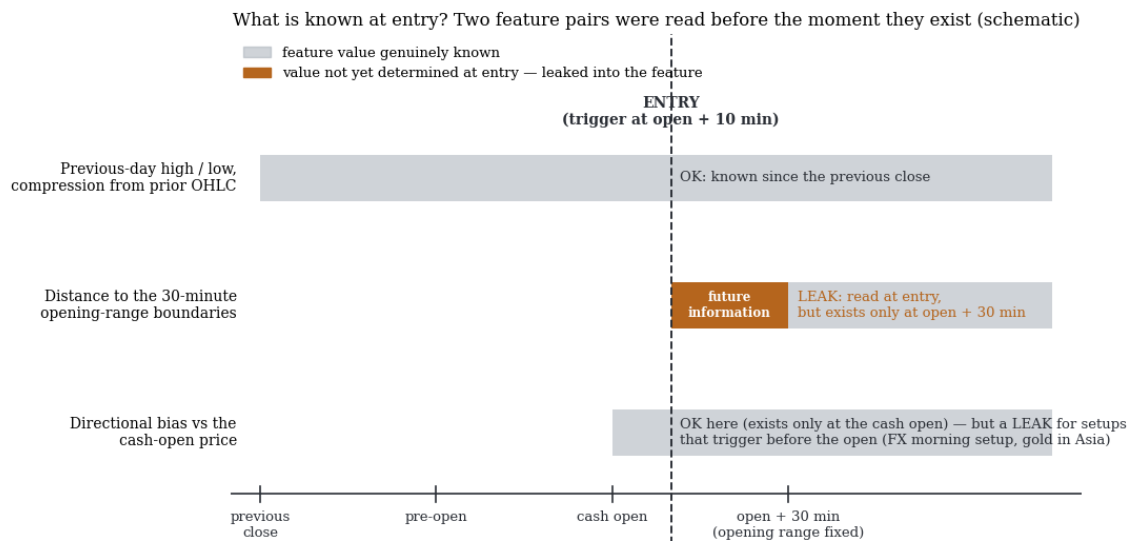


Abb. 1: Was ist zum Entry bekannt? Zwei Feature-Paare wurden gelesen, bevor der Wert existierte: die Opening-Range-Grenzen stehen erst 30 Minuten nach dem Open fest.

Jede Episode in unserer Basis trägt neben dem Trade-Ergebnis eine Reihe statischer Kontext-Features, die den Tag beschreiben: Vortages-Hoch und -Tief, eine Kompressionskennzahl aus dem Vortages-OHLC, die Distanz des Einstiegs zu den Grenzen der 30-Minuten-Opening-Range und eine Richtungs-Tendenz relativ zum Cash-Open-Preis.

Die Features wurden aus einer Tages-Zeile geholt — **ohne Prüfung, ob der jeweilige Wert zum Trigger-Zeitpunkt schon bekannt war**. Für Vortages-Werte ist das harmlos, die sind seit dem Vorabend fix. Für zwei Feature-Paare war es ein Leck:

Feature	definiert über	unbekannt für	betroffen
Distanz zur oberen / unteren Opening-Range-Grenze	30-Minuten-Opening-Range	Trigger innerhalb der ersten 30 Minuten	60.129 Episoden (48,8 %)
Richtungs-Tendenz (absolut und richtungsbezogen)	Cash-Open-Preis	Trigger vor dem Cash-Open	8.479 Episoden (ein FX-Vormittags-Setup 93,5 %, ein Gold-Setup in der Asien-Session 100 %)

Ein Setup, das nach zehn Minuten auslöst, bekam die Opening-Range-Grenzen mitgeliefert, die erst zwanzig Minuten später feststehen. Ein Pre-Open-Setup bekam den Open-Preis, der noch nicht existierte. Für die Hälfte aller Episoden enthielt der Zustand damit Information über den Kursverlauf *nach* dem Einstieg.

Dieselbe Fehlerklasse hatten wir schon einmal gefixt: Am 11.06.2026 hatten wir für die Features aus den ersten 5 und 15 Minuten eine Zeitprüfung eingebaut. Der Fix war richtig — aber er deckte nur die beiden Spalten ab, für die wir damals hingeschaut hatten. Die anderen vier blieben ungeprüft.

2. Woran wir es bemerkt haben

Das Leck fiel nicht beim Bau auf, sondern als Nebenprodukt unserer Studie zur Edge-Persistenz. Dort lieferte ein einfaches Nachbarschafts-Modell (kNN, walk-forward) auf den Kontext-Features eine Trennschärfe, die zu gut war, um wahr zu sein.

Der erste Verdacht kam trotzdem nicht sofort. Die Richtungs-Tendenz allein sah unauffällig aus: Korrelation mit dem Trade-Ergebnis $+0,011$. Erst die richtungsbezogene Variante — Tendenz mal Trade-Richtung — zeigte $+0,173$. Der Effekt mittelt sich über Long und Short weg, weil ein positiver Bias Longs hilft und Shorts schadet. **Ein richtungsneutrales Feature einzeln zu prüfen reicht nicht.** Man muss es mit der Trade-Richtung multiplizieren, sonst versteckt sich das Leck in der Symmetrie.

3. Der Beweis

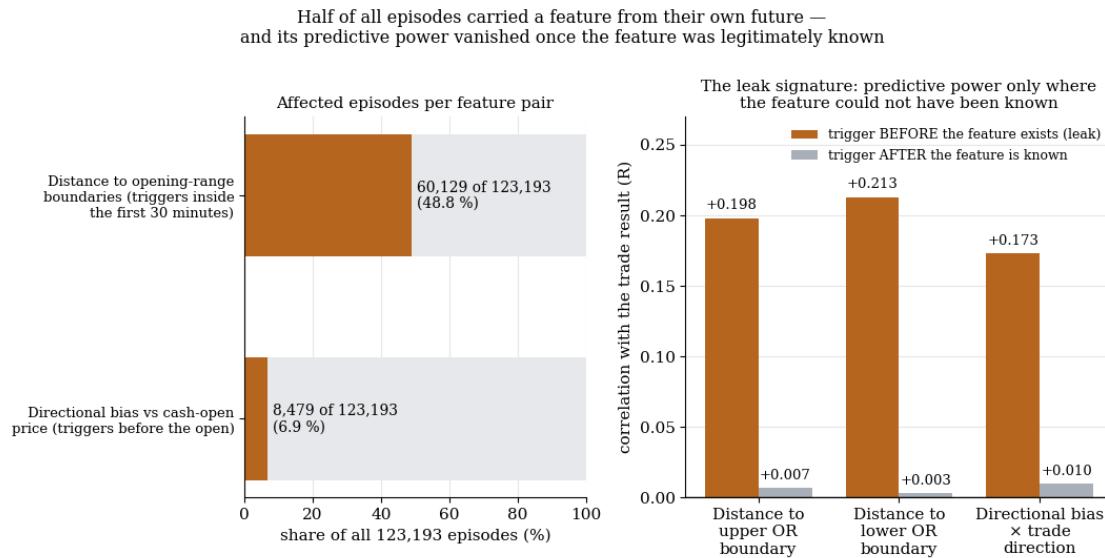


Abb. 2: Die Hälfte aller Episoden trug ein Feature aus der eigenen Zukunft. Die Leck-Signatur: Prognosekraft nur dort, wo das Feature nicht bekannt sein konnte.

Die Diagnose, die den Bug zweifelsfrei nachweist, ist einfach: Man teilt die Episoden danach, ob der Trigger *vor* oder *nach* dem Zeitpunkt lag, an dem das Feature real bekannt wird, und misst die Korrelation zum Trade-Ergebnis getrennt.

Feature	Trigger vor dem Zeitpunkt	Trigger danach	nach dem Fix
Distanz obere OR-Grenze	+0,198	+0,007	konstant 0 vor OR-Ende
Distanz untere OR-Grenze	+0,213	+0,003	konstant 0 vor OR-Ende
Richtungs-Tendenz $\times$ Richtung	+0,173	+0,010	konstant 0 vor Open

Das ist die Signatur eines Lecks, nicht eines Edges. Ein echtes Feature würde *nach* dem Bekanntwerden mindestens so gut prognostizieren wie davor. Hier ist es umgekehrt: Sobald das Feature legitim verfügbar ist, ist seine Vorhersagekraft weg. Die $+0,2$ stammen ausschließlich aus Episoden, für die das Feature noch gar nicht existieren durfte.

Auf Modell-Ebene sah der Unterschied so aus: Der kNN-Terzil-Spread — Erwartungswert bestes minus schlechtestes Drittel, walk-forward, innerhalb der Zelle — fiel von **$+0,726$ R auf $+0,191$ R**. Drei Viertel der scheinbaren Prognosekraft waren Zukunft.

Das schärfste diagnostische Detail war die Kalibrierung. Vor dem Fix hat das Modell **untertrieben**: Wo es $+0,316$ R prognostizierte, kamen $+0,493$ R. Nach dem Fix übertreibt es, wie es sich gehört: prognostiziert $+0,241$ R, realisiert $+0,197$ R. Ein sauberer, verrauschter Prädiktor *muss* übertreiben — Regression zur Mitte. **Ein Prädiktor, der systematisch untertreibt, weiß etwas, das er nicht wissen dürfte.** Diese Regel ist der günstigste Lookahead-Test, den wir kennen, und sie braucht keinen Blick in den Code.

4. Fix und Verifikation

Der Fix ist eine Zeitprüfung: Jedes Kontext-Feature wird auf null gesetzt, wenn der Trigger vor dem Zeitpunkt liegt, an dem der Wert real feststeht. Die Basis wurde komplett neu gebaut.

Drei Kontrollen mussten danach halten:

- **Alle vier Validierungsanker unverändert.** Die Referenzwerte, gegen die wir jede Neuberechnung der Basis prüfen, lagen exakt auf den alten Zahlen.
- **R-Werte über alle 123.193 Episoden bitgenau identisch** ($\max |\Delta R| = 0$). Der Fix ändert Features, nie Ergebnisse.
- **Der alte Zustand ist archiviert**, sodass jede Vorher-nachher-Behauptung nachrechenbar bleibt.

Die zweite Kontrolle ist die wichtige. Weil kein Trade-Ergebnis berührt wurde, bleiben alle bisherigen Setup-Verdikte und die Persistenz-Studie gültig. Sie beruhen auf Ergebnissen, nicht auf Features. Das Leck hat nur die *Erklärungen* verfälscht, nicht die *Messungen*.

5. Was nicht zu retten war

Vor dem Fix hatten wir mehrere Reinforcement-Learning-Modelle auf dieser Basis trainiert. Die Beobachtung, die ein solches Modell in jedem Schritt sieht, enthielt die Kontext-Features pauschal — und damit die vier Orakel-Spalten.

Diese Modelle sind kontaminiert und lassen sich nicht nachträglich bereinigen. Ein neuronales Netz, das im Training gelernt hat, auf eine Spalte zu reagieren, die Zukunft enthält, hat seine Gewichte um dieses Signal herum gebaut. Es gibt keinen Weg, das Signal aus den Gewichten zu subtrahieren. Ihre Evaluationszahlen sind zu optimistisch, und zwar am stärksten genau dort, wo das Leck am größten war: bei Setups, die früh in der Session auslösen. **Die einzige Konsequenz ist neu trainieren, und kein Ergebnis aus der alten Generation zählt.**

Das ist der teure Teil eines Lookahead-Bugs. Die Basis lässt sich in einer Stunde neu bauen. Alles, was auf der Basis gelernt wurde, muss weggeworfen werden.

6. Was daraus folgt

Drei Regeln, die wir seitdem als Pflicht führen:

Jedes Kontext-Feature trägt einen Zeitstempel, ab dem es bekannt ist. Nicht der Tag, sondern die Uhrzeit. Die Zeitprüfung sitzt beim Bau der Basis, nicht beim Modell — sonst vergisst sie jemand.

Richtungsneutrale Features werden mit der Trade-Richtung multipliziert geprüft. Sonst hebt sich das Leck über Long und Short auf und bleibt unsichtbar.

Kalibrierung ist ein Lookahead-Test. Ein Modell, dessen Prognosen zu klein ausfallen, ist verdächtig. Ein sauberes Modell übertreibt.

Dass wir dieselbe Fehlerklasse zweimal fixen mussten, ist die eigentliche Lehre. Der erste Fix war lokal — zwei Spalten, die gerade auffällig waren. Der zweite war strukturell — eine Zeitprüfung für jede Spalte, mit einem Suchskript, das die Fehlerklasse über alle Features abfragt. Nur der zweite ist ein Fix.

7. Grenzen

- **Das ist ein Post-mortem, kein Backtest.** Die Zahlen beschreiben die Größe eines Lecks, keinen Edge.

- **Die Korrelationen sind In-Sample-Diagnosen.** Sie belegen das Leck, sie messen nicht die verbleibende Prognosekraft der sauberen Features; die liegt mit +0,191 R im kNN-Terzil-Spread deutlich unter dem kontaminierten Wert und ist selbst noch nicht OOS-bestätigt.
- **Nicht getestet:** ob weitere, subtilere Lecks existieren, die keine der drei Signaturen (Vorher-nachher-Split, richtungsbezogene Korrelation, Untertreiben) auslösen. Das Suchskript deckt die Fehlerklasse „Tages-Wert vor seinem Zeitpunkt abgefragt“ ab, nicht jede denkbare Klasse.
- **Die kontaminierten RL-Modelle wurden nicht neu trainiert.** Alle Aussagen über ihre Optimismus-Größe sind qualitativ, bis saubere Vergleichsläufe existieren.

Rein statistische Auswertung historischer Marktdaten. Keine Trading-Empfehlung, keine Finanzberatung. Historische Performance ist keine Garantie für die Zukunft. Trading mit Hebelprodukten kann zum Totalverlust führen.