AUGH UG (haftungsbeschränkt) i. G.

Angewandte Forschung — Quantitative Finanztechnologie

Der gläserne Quant

Ein Free-Stack-Multi-Agenten-Framework für disziplinierte, kostenbewusste Anlageentscheidungen von Privatanlegern

Design-Science-Entwicklung und ehrliche 29-Jahres-Validierung eines regelbasierten KI-Investmentagenten (1997–2026)

Wissenschaftliche Projektarbeit

Whitepaper / Forschungsbericht

Verfasser: Louis Lazay
Mitwirkung (Mensch–KI-Prozess): Sander Behrens
KI-Co-Pilot: Claude (Anthropic)
Stand: 14. Juni 2026
Version: 1.0

Böblingen, 14. Juni 2026

Inhaltsverzeichnis

Abstract . . . III
Abstract (English) . . . III
Abbildungsverzeichnis . . . V
Tabellenverzeichnis . . . V
Abkürzungsverzeichnis . . . V
1. Einleitung . . . 1
1.1 Problemstellung . . . 1
1.2 Forschungslücke und Forschungsfrage . . . 2
1.3 Scope und Aufbau der Arbeit . . . 3
2. Theoretische Grundlagen . . . 3
2.1 Markteffizienz, ihre Grenzen und der dünne Edge . . . 3
2.2 Quantitative Anlage-Perspektiven und ihre Eignung . . . 4
2.3 Agentische KI und Human-AI-Collaboration im Anlagekontext . . . 6
2.4 Behavioral Finance als größter Hebel . . . 7
2.5 Forschungslücke: governance-gesicherte, kostenbewusste Multi-Lens-Orchestrierung . . . 7
3. Methodik . . . 8
3.1 Design Science Research . . . 8
3.2 Methodische Triangulation . . . 9
3.3 Design-Anforderungen (DR1–DR8) . . . 10
4. Das OMLA-Framework (Kernbeitrag) . . . 12
4.1 Architekturüberblick . . . 13
4.2 Hybridansatz: Die 13 Analyse-Perspektiven (Lenses) . . . 14
4.2.1 Abgrenzung zu bestehenden Frameworks . . . 15
4.3 Multi-Lens-Konsens-Dissens (das zentrale Governance-Konstrukt) . . . 16
4.4 Graceful Degradation (L1: Datenreifegrad) . . . 17
4.5 Ablations-Governance & Kosten-Veto (L2) . . . 18
4.6 Overconfidence-Friction & Mensch-im-Orderpfad (L3/L6) . . . 19
4.7 Phasenübergreifender Tageszyklus . . . 19
5. Prototyp und Validierung . . . 20
5.1 Technische Architektur . . . 20
5.2 Implementierung des Kerns (V6.1 Mean-Reversion) . . . 21
5.3 Datensubstrat & Reproduzierbarkeit . . . 22
5.4 Ausgeführter Validierungslauf: 29-Jahres-Backtest & Benchmark . . . 22
5.5 Kosten- und Slippage-Sensitivität (Kernbeitrag zur Ehrlichkeit) . . . 24
5.6 Ablations-Validierung: Welche Perspektive trägt? . . . 25
5.7 Evaluation gegen die Design-Anforderungen . . . 26
6. Diskussion . . . 27
6.1 Beantwortung der Forschungsfragen . . . 27
6.2 Wissenschaftliche Einordnung . . . 29
6.3 Implikationen . . . 29
6.4 Wirtschaftlichkeitsanalyse . . . 30
6.5 Boundary Conditions . . . 31
6.6 Limitationen . . . 31

7. Fazit und Ausblick 32
7.1 Zusammenfassung 32
7.2 Ausblick 32
Literaturverzeichnis 33
Eidesstattliche Erklärung 34

Abstract

Privatanleger stehen vor einem strukturellen Dilemma: Etablierte quantitative Anlagemethoden setzen kostenpflichtige Datenfeeds, dedizierte Research-Teams und kontinuierliche Marktbeobachtung voraus — Ressourcen, über die eine Einzelperson nicht verfügt. Gleichzeitig versprechen agentische KI-Systeme neue Formen der autonomen Analyse und Entscheidungsunterstützung, drohen jedoch, undurchsichtige „Black-Box"-Empfehlungen ohne nachprüfbaren Edge zu erzeugen. Diese Arbeit adressiert die Forschungslücke, dass **kein publiziertes Framework agentische KI-Fähigkeiten systematisch, kostenbewusst und governance-gesichert auf den täglichen Anlageentscheidungsprozess eines einzelnen Privatanlegers abbildet** — bei zugleich radikaler methodischer Ehrlichkeit. Mit dem **Orchestrated Multi-Lens Advisory (OMLA, „Der gläserne Quant")** wird ein neuartiges Framework vorgestellt, das dreizehn unabhängige Analyse-Perspektiven (u.,a. Mean-Reversion, Regime-Trend, Makro-Liquidität, Credit-Spreads, Behavioral, Microstructure) zu einer governance-gesicherten Tagesempfehlung orchestriert. Es führt mehrere eigene Konzepte ein: eine **Multi-Lens-Konsens-Dissens-Mechanik** (Perspektiven-Widerspruch als Risikosignal), eine **Ablations-Governance** (jede Perspektive muss ihren Edge out-of-sample beweisen oder wird abgeschaltet — „Feature-Removal als erstklassiger Design-Zug"), ein **Kosten-Veto** (Free-Stack-Prinzip: kein bezahltes Datum) sowie ein **Overconfidence-Friction-Protokoll** (Anti-FOMO-Stopp bei verdächtigem Signal-Konsens). Methodisch folgt die Arbeit dem Design-Science-Research-Paradigma nach Hevner et al. (2004) und Peffers et al. (2007). Die Validierung erfolgt über einen 29-Jahres-Backtest (1997–2026, 14.646 Trades) sowie eine **kosten- und benchmark-ehrliche Re-Analyse** (Buy-&-Hold-SPY-Vergleich, Slippage-Sensitivität, Walk-Forward-IS/OOS, Layer-/Overlay-Ablation). Kernbefund mit Mehrwert: Das Artefakt ist **kein Rendite-Motor** (CAGR 7,1,% vs. SPY 9,9,%), sondern eine **risikoadjustiert klar überlegene Krisen-Hedge-Maschine** (Sharpe 1,25 vs. 0,58; maximaler Drawdown −8,6,% vs. −55,2,%; Beta 0,18; Alpha +5,1,% p.,a.), deren dünner Edge die Benchmark risikoadjustiert nur bis ca. 15 Basispunkte Round-Trip schlägt (Breakeven ~29–40 bps). Im Sinne der Knowledge Contribution Matrix nach Gregor und Hevner (2013) positioniert sich der Beitrag als **Exaptation** — die Anwendung reifer Lösungsartefakte (klassische Quant-Methoden, Medallion-Datenarchitektur, Multi-Agent-Patterns) auf einen neuartigen Problemraum (governance-gesicherte, kostenbewusste, beratende KI-Anlageunterstützung für eine Einzelperson) — mit *Improvement*-Beitrag auf Konstrukt-Ebene.

Schlüsselwörter: Agentic AI, KI-Investment, Quantitative Finance, Algorithmischer Handel, Multi-Agent-Systeme, Mean Reversion, Backtesting, Risk Management, Privatanleger, Design Science Research, Free-Stack, Human-AI-Collaboration

Abstract (English)

Retail investors face a structural dilemma: established quantitative methods presuppose paid data feeds, dedicated research teams and continuous monitoring. Agentic AI promises autonomous analysis but risks opaque "black-box" recommendations with no verifiable edge. This work addresses the gap that *no published framework systematically maps agentic AI capabilities onto a single retail investor's daily decision process under a hard cost constraint and strict human-in-the-loop governance, while maintaining radical methodological honesty.* This paper presents **OMLA — Orchestrated Multi-Lens Advisory ("the glass-box quant")**, a free-stack multi-agent framework that orchestrates thirteen independent analytical lenses into one governed daily recommendation, contributing four novel constructs: Multi-Lens Consensus-Dissent, Ablation Governance (a layer must prove out-of-sample edge or be switched off), a Cost Veto (no paid data), and an Overconfidence-Friction protocol. Following Design Science Research, validation rests on a 29-year backtest (14,646 trades) and a cost- and benchmark-honest re-analysis (including walk-forward IS/OOS testing and layer ablation). Key finding: the artifact is not a return engine (CAGR 7.1% vs. SPY 9.9%) but a risk-adjusted, crisis-hedge engine (Sharpe 1.25 vs. 0.58; max drawdown −8.6% vs. −55.2%) whose thin edge beats the benchmark risk-

adjusted only up to ~15 bps round-trip (break-even ~29–40 bps).

Abbildungsverzeichnis

Nr.	Titel	Kapitel
Abbildung 1	OMLA-Architektur — Pipeline-Phasen × Cross-Cutting-Layer	4.1
Abbildung 2	Multi-Lens-Konsens-Dissens — Governance-Fluss je Tages-Gate	4.3
Abbildung 3	Equity-Kurve & Drawdown Strategie vs. SPY Buy-&-Hold (1997–2026)	5.4
Abbildung 4	Kosten-/Slippage-Sensitivität — CAGR & Sharpe über Round-Trip-bps	5.5

Tabellenverzeichnis

Nr.	Titel	Kapitel
Tabelle 1	Bewertungsmatrix kandidierender Anlage-Perspektiven	2.2
Tabelle 2	DSRM-Mapping nach Peffers et al. (2007)	3.1
Tabelle 3	Methodische Triangulationsmatrix — Methoden × Kernbehauptungen	3.2
Tabelle 4	Design-Anforderungen (DR1–DR8) mit Kernel-Theorie & Falsifikationskriterium	3.3
Tabelle 4a	OMLA — Cross-Cutting-Layer L1–L7 mit DR-Zuordnung	4.1
Tabelle 5	Die 13 Analyse-Perspektiven (Lenses) mit Rolle & Reifegrad	4.2
Tabelle 6	Abgrenzung OMLA — gemeinsame Merkmale & Alleinstellungsmerkmale	4.2
Tabelle 7	Graceful Degradation nach Datenreifegrad	4.4
Tabelle 8	Performance-Kennzahlen Strategie vs. SPY Buy-&-Hold (1997–2026)	5.4
Tabelle 8a	Technologie-Stack des Prototyps	5.1
Tabelle 9	Risiko-/Relativkennzahlen (Alpha, Beta, IR, Korrelation)	5.4
Tabelle 10	Walk-Forward-Robustheit (IS/OOS-Degradation)	5.4
Tabelle 11	Kosten-Sensitivität — Break-even & risk-adjusted vs. SPY	5.5
Tabelle 12	Layer-/Overlay-Ablation — Beitrag je Perspektive	5.6
Tabelle 13	Evaluation gegen Design-Anforderungen (Status/Evidenz/Falsifikation)	5.7
Tabelle 14	Wirtschaftlichkeitsanalyse — Betriebskosten vs. Robo-Advisor/Fonds	6.4

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AMH	Adaptive Markets Hypothesis (Lo, 2004)
ATR	Average True Range
bps	Basispunkte (1 bps = 0,01,%)
CAGR	Compound Annual Growth Rate
DR	Design Requirement (Design-Anforderung)
DSR	Design Science Research
DSRM	Design Science Research Methodology
EMH	Efficient Market Hypothesis (Fama, 1970)
IR	Information Ratio
LLM	Large Language Model
MaxDD	Maximaler Drawdown
MAS	Multi-Agent-System
MCP	Model Context Protocol
MR	Mean Reversion
OMLA	Orchestrated Multi-Lens Advisory (das Framework)
OOS / IS	Out-of-Sample / In-Sample
RSI	Relative Strength Index
WR	Win Rate (Trefferquote)

1. Einleitung

1.1 Problemstellung

Quantitative, regelbasierte Anlagestrategien haben den professionellen Kapitalmarkt seit den 1990er-Jahren fundamental verändert. Ihr Zugang blieb jedoch strukturell asymmetrisch verteilt: Die etablierten Methoden setzen typischerweise kostenpflichtige Datenfeeds (Bloomberg, Polygon, Unusual Whales), dedizierte Research-Teams und eine kontinuierliche Marktbeobachtung voraus — Ressourcen, über die ein einzelner Privatanleger strukturell nicht verfügt. Diese **Free-Stack-Asymmetrie** ist nicht graduell, sondern qualitativ: Der professionelle Apparat operiert auf einem Informations- und Infrastruktur-Niveau, das einer Einzelperson per Definition verschlossen bleibt. Wer als Privatanleger dennoch systematisch handeln will, steht damit vor einem strukturellen Dilemma — entweder verzichtet er auf den methodischen Anspruch, oder er versucht, ihn unter Bedingungen zu erfüllen, die er nicht erfüllen kann.

Die zentrale These dieser Arbeit ist jedoch, dass die größte Renditebremse des Privatanlegers gar nicht in dieser Infrastruktur-Asymmetrie liegt, sondern in seinem eigenen Verhalten. Empirisch ist die dominante Verlustquelle nicht fehlendes *Alpha*, sondern **behaviorale Leckage**: Übermut, Verlustaversion, prozyklisches Hin-und-Her-Handeln und der *Disposition-Effekt* kosten den durchschnittlichen Privatanleger systematisch Rendite. Barber und Odean (2000) zeigen in ihrer klassischen Untersuchung von über 66.000 Privatdepots, dass die aktivsten Händler ihre Bruttorendite durch Transaktionskosten und Fehltiming derart erodieren, dass ihre Nettorendite deutlich hinter einer passiven Marktteilnahme zurückbleibt — der Befund verdichtet sich in ihrem Diktum *„trading is hazardous to your wealth"*. Diese verhaltensökonomisch begründete Lücke zwischen der Rendite eines Anlageinstruments (*investment return*) und der tatsächlich vom Anleger realisierten Rendite (*investor return*) wird in der Praktiker-Literatur als *behavior gap* geführt und über Methoden der DALBAR-QAIB-Logik regelmäßig auf mehrere Prozentpunkte pro Jahr beziffert (vgl. Barber & Odean, 2000; Kahneman & Tversky, 1979). Die hier vorgestellte Code- und Designanalyse verortet die so verlorene Rendite in einer Größenordnung von rund 3–8 % pro Jahr — ein Hebel, der die meisten realistisch erreichbaren *Alpha*-Quellen übersteigt. **Befund.** Wenn die größte Verlustquelle die Disziplin ist und nicht die Information, dann liegt der primäre Adressat eines Anleger-Werkzeugs nicht im Generieren neuer Signale, sondern in der Härtung des Entscheidungsprozesses gegen das eigene Verhalten — eine Verschiebung des Designziels von *Optimierung* hin zu *Bias-Hygiene*.

Mit der Reife agentischer KI-Systeme — autonome Planung, Werkzeugnutzung, mehrstufiges *Reasoning* — entsteht erstmals die realistische Möglichkeit, einen disziplinierten quantitativen Entscheidungsprozess auch für eine Einzelperson zu betreiben. Dieser Verheißung steht jedoch eine konkrete Gefahr gegenüber: Ein KI-System, das undurchsichtige Kauf- bzw. Verkaufsempfehlungen ohne nachprüfbaren, out-of-sample bestätigten *Edge* erzeugt, ist nicht neutral, sondern schädlich — es verleiht behavioraler Spekulation einen szientistischen Anstrich und kann so die behaviorale Leckage sogar verstärken, statt sie zu dämpfen. Diese **Black-Box-Gefahr** agentischer KI ist im Anlagekontext besonders virulent, weil Finanzdaten zirkelbezüglich und halluzinationsanfällig sind und ein plausibel formulierter, aber substanzloser Ratschlag vom Anwender kaum von einem fundierten zu unterscheiden ist. Die Branchenrealität bestätigt diese Zurückhaltung: Branchenbeobachtungen zufolge setzt unter den führenden quantitativen Fonds praktisch nur Bridgewaters *AIA Labs Large Language Models* im eigentlichen Entscheidungspfad ein; der dominante produktive Nutzen von LLMs liegt bislang in der *Research-Geschwindigkeit* („operational alpha"), nicht in der autonomen Order-Platzierung (vgl. Branchenanalysen, 2025). Das hier präsentierte Artefakt zieht aus dieser Beobachtung eine harte Designkonsequenz — das LLM verbleibt strikt *advisory-only* und niemals im Orderpfad (vgl. Kap. 4.6, DR7) —, die im weiteren Verlauf als eigenständiges Governance-Konstrukt entwickelt wird.

1.2 Forschungslücke und Forschungsfrage

Die Literatur adressiert entweder **generische quantitative Anlagestrategien** (Connors & Alvarez, 2009; López de Prado, 2018) **oder generische agentische KI-Architekturen** — jedoch nicht deren systematische, governance-gesicherte und **kostenbewusste** Verschränkung für den Anwendungsfall eines einzelnen Privatanlegers. Die Quant-Literatur setzt institutionelle Ressourcen und professionelle Datenfeeds implizit voraus und blendet die behaviorale wie die Kostendimension der Einzelperson weitgehend aus; die Agentic-AI-Literatur wiederum behandelt Architektur, Orchestrierung und *Tool*-Nutzung weithin domänenfrei und ohne harte ökonomische Randbedingung. Der Schnittpunkt — eine beratende, kostenbewusste und nachprüfbar fundierte KI-Anlageunterstützung für eine Einzelperson — bleibt damit von keiner der beiden Literaturen abgedeckt. Daraus folgt die Forschungslücke:

> **Es existiert kein publiziertes Framework, das agentische KI-Fähigkeiten systematisch auf den täglichen Anlageentscheidungsprozess eines einzelnen Privatanlegers abbildet — unter einer harten Kosten-Randbedingung (Free-Stack), strikter Mensch-im-Orderpfad-Governance und dem Anspruch radikaler methodischer Ehrlichkeit (out-of-sample-Nachweis statt Backtest-Hopium).**

Aus dieser Lücke leitet sich die **zentrale Forschungsfrage** ab:

> Inwiefern kann ein orchestriertes Multi-Perspektiven-Agentensystem den täglichen Anlageentscheidungsprozess eines Privatanlegers so unterstützen, dass ein nachprüfbarer, kosten- und risikobewusster Mehrwert gegenüber einer passiven Benchmark entsteht — ohne die Entscheidungsautonomie an die KI abzugeben?

Die Forschungsfrage wird durch drei **Unterfragen** operationalisiert, die jeweils einer Analyseebene des Artefakts entsprechen:

1. **Dekomposition & Aggregation.** Wie lassen sich heterogene Anlage-Perspektiven (technisch, makroökonomisch, behavioral, mikrostrukturell) in diskrete, einzeln falsifizierbare Analyse-Agenten („Lenses") dekonstruieren und zu einer kohärenten Tagesempfehlung aggregieren?
2. **Governance & Mensch-KI-Grenze.** Welche Governance-Mechanismen verhindern überoptimistische Signale, künstlichen Konsens und behaviorale Leckage — und wo verläuft die Mensch-KI-Grenze?
3. **Belastbarkeit unter Realbedingungen.** Wie verändert sich die belastbare Aussagekraft des Systems mit dem Datenreifegrad und unter realistischen Transaktionskosten?

Aus der Beantwortung dieser Fragen ergeben sich die folgenden **Beiträge dieser Arbeit**:

- Ein neuartiges, zweidimensionales Framework — *Orchestrated Multi-Lens Advisory* (OMLA, „Der gläserne Quant") — mit vier eigenen Konstrukten: einer **Multi-Lens-Konsens-Dissens-Mechanik** (Perspektiven-Widerspruch als Risiko- bzw. Chancensignal), einer **Ablations-Governance** (*Feature-Removal* als erstklassiger Design-Zug — jede Perspektive muss ihren *Edge* out-of-sample beweisen oder wird abgeschaltet), einem **Kosten-Veto** (Free-Stack-Prinzip: kein bezahltes Datum) und einem **Overconfidence-Friction-Protokoll** (Anti-FOMO-Stopp mit 60-Tage-Live-Stop gegen *Confirmation-Bias*).
- Eine **kosten- und benchmark-ehrliche** Re-Validierung eines 29-Jahres-Backtests (1997–2026, 14.646 Trades), die den verbreiteten Fehler absoluter — statt risiko- und benchmark-relativer — Performance-Berichterstattung korrigiert und den *Edge* explizit gegen eine realistische Transaktionskosten-Schwelle prüft.
- Eine offen dokumentierte **Negativ-Ergebnis-Kultur** (abgeschaltete Perspektiven, die offen ausgewiesene Break-even-Kostenschwelle) als methodischer Beitrag gegen *Backtest-Overfitting* (vgl. López de Prado, 2018).

1.3 Scope und Aufbau der Arbeit

Scope. Diese Arbeit präsentiert und validiert ein *designtes Artefakt* — das OMLA-Framework als größere Master-Vision mit dreizehn Analyse-Perspektiven — sowie dessen lauffähige Phase-A-Instanziierung, einen regelbasierten *Mean-Reversion*-Kern (V6.1-MR) auf US-Large-Caps. Sie validiert die Eigenschaften dieses Artefakts über einen 29-Jahres-Backtest (1997–2026, 14.646 Trades) sowie über konvergente Robustheits- und Kostenanalysen (Walk-Forward-IS/OOS, Slippage-Sensitivität, Layer-/Overlay-Ablation); sie ist **keine Anlageberatung**. Die Arbeit versteht sich zudem als eigenständiges Whitepaper im Mensch-Mensch-Prozess (Louis Lazay & Sander Behrens) mit Claude (Anthropic) als KI-Co-Pilot und ist **nicht** an eine prüfungsrechtliche Hochschulbindung gekoppelt.

Nicht-Scope. Explizit **nicht** Bestandteil dieser Arbeit sind: eine **Anlageberatung** im rechtlichen Sinne (die Ausführungen erfüllen keinen Beratungstatbestand), steuer- und regulatorische Aspekte, die **vollautomatische Order-Ausführung** (per Design ausgeschlossen — das LLM verbleibt *advisory-only*, niemals im Orderpfad, ausgeführt wird ausschließlich manuell über Trade Republic) sowie die **vollständige Implementierung aller dreizehn Perspektiven**. Mehrere Perspektiven sind im Framework spezifiziert, aber als Folgearbeit deklariert; live validiert ist allein der *Mean-Reversion*-Kern als Phase-A-Instanz (vgl. Kap. 4.2, Kap. 7.2). Diese Scope-Setzung entspricht der Design-Science-Anforderung, ein *nascent* Artefakt deskriptiv zu evaluieren, bevor experimentelle Validierung sinnvoll möglich wird (Hevner et al., 2004, Guideline 3).

Aufbau. Kapitel 2 fundiert die theoretischen Grundlagen und Kernel-Theorien — von der Effizienzmarkthypothese und ihren Grenzen (Fama, 1970; Grossman & Stiglitz, 1980) über die *Adaptive Markets Hypothesis* (Lo, 2004) bis zur *Behavioral Finance* als größtem Hebel (Kahneman & Tversky, 1979; Barber & Odean, 2000; Simon, 1956). Kapitel 3 entwickelt die Methodik (Design Science Research nach Hevner et al., 2004, und Peffers et al., 2007; methodische Triangulation; die Design-Anforderungen DR1–DR8 mit Kernel-Theorie und Falsifikationskriterium). Kapitel 4 stellt das OMLA-Framework als Kernbeitrag vor. Kapitel 5 dokumentiert Prototyp und Validierung mit echten Zahlen aus dem 29-Jahres-Backtest, der Kosten-Sensitivitätsanalyse, der Walk-Forward-Robustheit und der Layer-/Overlay-Ablation. Kapitel 6 diskutiert die Ergebnisse, ordnet sie wirtschaftlich und wissenschaftlich ein (Exaptation auf Framework-Ebene, *Improvement* auf Konstrukt-Ebene nach Gregor & Hevner, 2013) und benennt die Grenzen. Kapitel 7 schließt mit Fazit und Ausblick, mit klarer Trennung zwischen den Beiträgen dieser Arbeit und der Folgeforschung.

2. Theoretische Grundlagen

Dieses Kapitel fundiert das in Kapitel 4 entwickelte Artefakt theoretisch und benennt die Kernel-Theorien, an die jede zentrale Design-Entscheidung gebunden ist (vgl. Tabelle 4, Kap. 3.3). Die Argumentation verläuft von der ökonomischen Möglichkeitsbedingung eines Edges (2.1) über die Auswahl der Analyse-Perspektiven (2.2) und die Rolle agentischer KI (2.3) bis zum behavioralen Kernhebel (2.4) und mündet in die Synthese der Forschungslücke (2.5). Leitend ist dabei eine durchgängig demarkierende Haltung: Die Theorie begründet, *warum* das Artefakt so gebaut ist, sie liefert jedoch ausdrücklich keinen Wirksamkeitsnachweis — dieser ist Gegenstand der ehrlichen Validierung in Kapitel 5.

2.1 Markteffizienz, ihre Grenzen und der dünne Edge

Den theoretischen Ausgangspunkt bildet die *Efficient Market Hypothesis* (EMH) nach Fama (1970). In ihrer halbstrengen Form besagt sie, dass in den Marktpreisen sämtliche öffentlich verfügbaren Informationen bereits eingepreist sind, sodass auf Basis solcher Informationen kein systematischer Überrenditen-Vorteil erzielbar ist. Für ein Artefakt, das ausschließlich auf frei verfügbaren, öffentlichen Datenquellen operiert (*Free-Stack*-Prinzip, vgl. DR2), ist dies eine maximal harte Randbedingung:

Die EMH bestreitet die Existenz genau jenes Edges, den das System beansprucht. Eine Arbeit, die diese Position ignorierte, wäre szientistisch — entsprechend wird die EMH hier nicht als widerlegter Strohmann, sondern als ernstzunehmende Nullhypothese behandelt.

Das Grossman-Stiglitz-Paradoxon. Den entscheidenden theoretischen Spielraum eröffnet die Arbeit von Grossman und Stiglitz (1980). Sie zeigen, dass vollständig informationseffiziente Märkte logisch unmöglich sind: Wäre der gesamte private Informationsgehalt bereits kostenlos im Preis enthalten, bestünde kein ökonomischer Anreiz mehr, überhaupt Informationen zu beschaffen oder zu verarbeiten — wodurch der Mechanismus, der die Preise effizient macht, kollabierte. Märkte müssen daher *gerade so* ineffizient bleiben, dass die Informationsbeschaffung ihre Kosten deckt. Daraus folgt unmittelbar die zentrale Eigenschaft des real existierenden Edges: Er ist *dünn* und *kostensensitiv*. Ein nachweisbarer Vorteil kann existieren — er ist jedoch theoretisch genau auf die Größenordnung der Informations- und Transaktionskosten begrenzt, die ihn aufrechterhalten. **Theoretische Verankerung.** Das Grossman-Stiglitz-Paradoxon ist damit die unmittelbare Begründung des *Kosten-Vetos* (DR2): Ein Edge, der unter realistischer Transaktionsfriktion verschwindet, ist kein Marktversagen, sondern der theoretisch erwartete Normalfall. Die in Kapitel 5.5 berichtete Break-even-Kostenschwelle von rund 29 bis 40 Basispunkten (*Round-Trip*) ist insofern keine Überraschung, sondern die empirische Bestätigung eines theoretisch vorhergesagten dünnen Edges — vgl. Kap. 5.5.

Adaptive Markets Hypothesis. Die *Adaptive Markets Hypothesis* (AMH) nach Lo (2004) versöhnt die EMH mit der empirischen Beobachtung persistenter Anomalien, indem sie Markteffizienz evolutionär statt statisch fasst. Marktteilnehmer handeln unter *Bounded Rationality* (vgl. Simon, 1956; Abschn. 2.4), Strategien konkurrieren um begrenzte Ressourcen, und ein einmal vorhandener Edge wird durch Arbitrage und Adaption erodiert — bis veränderte Marktregime ihn erneut öffnen. Edge ist demnach *regime- und zeitabhängig*, nicht permanent. **Befund-Verankerung.** Die AMH liefert die theoretische Erklärung für den zentralen Empiriebefund dieser Arbeit: Die instanziierte *Mean-Reversion*-Strategie ist kein Allwetter-Rendite-Motor, sondern ein *Krisen-Hedge*. Die regime-konditionierte Attribution des Backtests (vgl. Kap. 5.4) zeigt eine systematische Unterperformance gegenüber SPY in Aufwärtsjahren (durchschnittlich rund −13 Prozentpunkte) bei gleichzeitig deutlicher Outperformance in Bärenmarkt-Jahren (rund +21,9 Prozentpunkte; im Krisenjahr 2008 −3,0 % gegenüber SPY −36,8 %). Dieses asymmetrische, konvex-defensive Profil — niedriges Beta (0,180), ein Bruchteil des Drawdowns der Benchmark (vgl. Tabelle 8) — ist exakt das, was die AMH für eine *Mean-Reversion*-Strategie vorhersagt: Ihr Edge ist an Regime erhöhter Volatilität und Angst gebunden, in denen kurzfristige Überreaktionen am stärksten ausgeprägt sind, und verschwindet in ruhigen Aufwärtsmärkten, in denen *Momentum* dominiert. Die theoretische Verankerung des Risiko-vor-Rendite-Prinzips (DR3) liegt damit in der AMH.

Epistemischer Status. Die EMH-Grenzen begründen die Möglichkeit eines Edges; sie beweisen ihn nicht. Ob der Edge im konkreten Artefakt vorliegt und ob er die Kostenschwelle überlebt, ist eine empirische Frage, die ausschließlich die *Out-of-Sample*-Validierung (Kap. 5.4–5.5) beantworten kann.

2.2 Quantitative Anlage-Perspektiven und ihre Eignung

Die *Master-Vision* des Artefakts versteht den Markt nicht durch eine einzelne Methode, sondern durch dreizehn unabhängige Analyse-Perspektiven (*Lenses*; vgl. Tabelle 5, Kap. 4.2). Welche dieser Perspektiven unter den Randbedingungen eines einzelnen Privatanlegers — *Free-Stack*-Daten, realistische Transaktionsfriktion, fehlendes Research-Team — überhaupt tragfähig sind, ist keine Selbstverständlichkeit, sondern erfordert eine begründete Vorauswahl. Dieser Abschnitt stellt die kandidierenden Perspektiven-Familien vor und bewertet sie argumentativ.

Mean Reversion bildet den live instanziierten Kern (Phase-A-Instanz, V6.1-MR). Die methodische Grundlage liefern Connors und Alvarez (2009) mit kurzfristigen, antizyklischen Setups (RSI-2, *ConnorsRSI*, TPS): Gekauft wird die kurzfristige Überreaktion nach unten innerhalb eines intakten langfristigen Aufwärtstrends. Die zugrunde liegenden Indikatoren — der *Relative Strength Index* (RSI) und die

Average True Range (ATR) — gehen auf Wilder (1978) zurück; der zusammengesetzte *ConnorsRSI* ist bei Connors, Alvarez und Radtke (2012) spezifiziert. Die Eignung ist hoch — die benötigten Eingangsdaten (Tagespreise) sind über *yfinance* und Stooq vollständig frei verfügbar, die Logik ist transparent und reproduzierbar, und der antizyklische Charakter ist behavioral robust (vgl. Abschn. 2.4). Die Kehrseite ist eine hohe Handelsfrequenz und damit Kostensensitivität (jährlicher Notional-Umschlag 18,6× der *Equity*; vgl. Kap. 5.5).

Trend / Momentum (CTA-Stil) ist die theoretische Gegen-Perspektive zur *Mean Reversion* und in der AMH-Logik komplementär: Wo *Mean Reversion* in volatilen Regimen trägt, dominiert *Momentum* in ruhigen Aufwärtsmärkten. Datenseitig ist die Perspektive ebenfalls *Free-Stack*-fähig (reine Preisreihen) und gilt out-of-sample als vergleichsweise robust; ihre Aktivierung als zweites Buch (*Dual-Book*) ist im Ausblick als Allwetter-Erweiterung deklariert (vgl. Kap. 7.2), aber nicht Bestandteil der live validierten Instanz.

Faktoren (Fama-French) — Size, Value, Profitability — sind empirisch breit belegt, operieren jedoch auf längeren Halteperioden und setzen Fundamentaldaten voraus, die im *Free-Stack* nur lückenhaft und nicht *point-in-time*-sauber vorliegen (verfügbar sind lediglich *yfinance*-Earnings-Events, keine vollständigen Bilanzreihen). Für den täglichen, kurzfristigen Entscheidungsprozess eines Einzelanlegers ist die Umsetzbarkeit daher eingeschränkt.

Makro-Liquidität und Credit-Spreads sind über die FRED-Schnittstelle hervorragend und kostenfrei abgedeckt (Credit-Spreads HY/IG/AAA/BBB/CCC-OAS ab 1996, Zinsstruktur, Liquiditätsaggregate). Ihre primäre Eignung liegt jedoch nicht in der Titelselektion, sondern als *Regime-* und *Veto-Schicht* — sie informieren das Marktregime-Gate und intermarket-Vetos, nicht das Einzelsignal.

Behavioral / Sentiment (AAII-, NAAIM-Daten, GDELT-Nachrichten-Sentiment) ist *Free-Stack*-verfügbar, leidet jedoch unter geringer Signaldichte und Datenlücken (die historische CBOE-Put/Call-Ratio ist seit 2024 kostenpflichtig und liegt nur als selbst-berechneter Einzelwert vor). Die *Out-of-Sample*-Robustheit der abgeleiteten Scoring-Signale erwies sich als schwach — ein Befund, der unten und in Kapitel 5.6 als Ablationsergebnis dokumentiert ist.

Microstructure (Spreads, Order-Flow, Intraday-Mikrostruktur) ist für die Krisen-Robustheits-Frage hochrelevant, datenseitig im *Free-Stack* jedoch praktisch nicht tragfähig: Belastbare Mikrostrukturdaten sind kostenpflichtig, und die Intraday-Schicht des Datensubstrats ist bewusst leer (deferiert). Für einen Privatanleger mit Tages-Entscheidungsrhythmus ist die Umsetzbarkeit gering.

Tabelle 1 — Bewertungsmatrix kandidierender Anlage-Perspektiven fasst diese Einschätzungen entlang vier privatanleger-relevanter Kriterien zusammen (Skala −− / − / o / + / ++):

Perspektive	Free-Stack-Datenverfügbarkeit	OOS-Robustheit	Kostensensitivität (günstig = +)	Privatanleger-Umsetzbarkeit	Verbleib im Artefakt
Mean Reversion (Connors & Alvarez, 2009)	++	+	−	++	**live (Phase-A-Kern)**
Trend / Momentum (CTA)	++	+	o	+	spezifiziert (Dual-Book, Kap. 7.2)
Faktoren (Fama-French)	−	+	+	−	nicht priori-siert
Makro-Liquidität	++	o	++	o	live (als Regime-/Veto-Schicht)

Perspektive	Free-Stack-Datenverfügbarkeit	OOS-Robustheit	Kostensensitivität (günstig = +)	Privatanleger-Umsetzbarkeit	Verbleib im Artefakt
Credit-Spreads	++	o	++	o	live (als Regime-/Veto-Schicht)
Behavioral / Sentiment	+	−	−	o	abgeschaltet (DISABLED, Kap. 5.6)
Microstructure	−−	o	−	−−	nicht tragfähig (Free-Stack)

Tabelle 1: Bewertungsmatrix kandidierender Anlage-Perspektiven (Skala −− kaum gegeben bis ++ voll gegeben).

Methodischer Disclaimer. Diese Matrix ist ausdrücklich als **argumentative Vorauswahl mit transparenten Kriterien zu lesen, nicht als Messung.** Die Bewertungen stellen eine literatur- und implementierungsgestützte Einzelbewertung des Autors dar; sie ersetzen weder eine empirische Frameworkbewertung noch eine systematische Inter-Rater-Validierung. Sie dienen allein der nachvollziehbaren Begründung, *warum* gerade diese Perspektiven in den Hybridansatz eingehen und welche zunächst zurückgestellt werden (vgl. Kap. 4.2). Die finale Selektion wird nicht durch die Matrix entschieden, sondern durch die *empirische Ablation* out-of-sample (vgl. Kap. 5.6): Die Behavioral-/Sentiment-Perspektive etwa erscheint hier nur leicht abgewertet, wurde jedoch nach negativem *Delta*-Erwartungswert out-of-sample hart abgeschaltet (`DISABLED_LAYERS`) — die argumentative Vorauswahl ist mithin der schwächere, die Ablation der stärkere Filter. Eine intersubjektive Validierung der Scores ist nicht Bestandteil dieser Arbeit.

2.3 Agentische KI und Human-AI-Collaboration im Anlagekontext

Agentische KI-Systeme zeichnen sich gegenüber klassischer generativer KI durch Autonomie, Zielorientierung, Werkzeugnutzung und mehrstufiges *Reasoning* aus (vgl. Wang et al., 2024). Im Anlagekontext ist die zentrale Design-Frage nicht, *ob* ein *Large Language Model* (LLM) eingebunden wird, sondern *an welcher Stelle des Entscheidungspfads*. Hierfür ist eine scharfe Unterscheidung zwischen zwei Einsatzmodi konstitutiv.

LLM als *operational alpha* versus LLM im Orderpfad. Der produktiv tragfähige Nutzen von LLMs liegt in der *Research-Geschwindigkeit* — der Beschleunigung von Hypothesengenerierung, Code-Erstellung, Dokumentensichtung und Plausibilisierung (*operational alpha*). Davon strikt zu trennen ist die autonome Order-Platzierung, bei der ein LLM Kauf- oder Verkaufsentscheidungen unmittelbar in Aufträge übersetzt. Die Branchenrealität stützt diese Trennung: Unter den führenden quantitativen Akteuren setzt praktisch nur ein einzelner Fonds LLMs im eigentlichen Entscheidungspfad ein, während der dominante Nutzen branchenweit in der Research-Augmentierung verbleibt (vgl. Kap. 1).

Halluzination und Zirkelbezug bei Finanzdaten. Der Ausschluss des LLM aus dem Orderpfad ist nicht vorsichtsbedingt, sondern theoretisch begründet. Erstens neigen LLMs zu *Halluzination* — der konfidenten Erzeugung faktisch falscher Aussagen —, was bei numerischen Finanzdaten (Kurse, Kennzahlen, Schwellen) unmittelbar handlungsrelevante Fehler erzeugt. Zweitens droht ein *Zirkelbezug*: Da Marktkommentar, Nachrichten und Modell-Output denselben öffentlichen Informationsraum teilen, kann ein LLM bereits eingepreiste Narrative als vermeintlich neue Evidenz reproduzieren und so einen Edge vortäuschen, der ökonomisch nicht existiert (vgl. das Grossman-Stiglitz-Paradoxon, Abschn. 2.1). Drittens erschwert die probabilistische Natur des LLM-Outputs die Reproduzierbarkeit einer einzelnen

Tagesentscheidung (vgl. DR6). **Begründung der Advisory-only-Governance.** Aus diesen drei Risiken folgt die *Advisory-only*-Festlegung (DR7): Das deterministische, regelbasierte Kernsystem trifft jede scoring- und sizing-relevante Entscheidung; das LLM unterstützt Briefing, Erläuterung und Forensik, **platziert oder triggert jedoch nie eine Order** — der Mensch bleibt im Orderpfad. Diese Architektur folgt den Strukturmodellen der Human-AI-Collaboration (vgl. Shrestha et al., 2019, zu hybriden Sequenz- versus Delegationsmodellen): Statt vollständiger Delegation an die KI wird eine hybride Sequenz gewählt, in der KI-Vorschlag und menschliche Gate-Entscheidung getrennt bleiben.

2.4 Behavioral Finance als größter Hebel

Die zentrale These dieser Arbeit zur Privatanleger-Domäne lautet, dass der größte realisierbare Hebel nicht in zusätzlichem Alpha, sondern in der Vermeidung *behavioraler Leckage* liegt. Drei Theorien fundieren diese Position.

Prospect Theory. Kahneman und Tversky (1979) zeigen, dass Entscheidungen unter Risiko nicht dem Erwartungsnutzen folgen, sondern relativ zu einem Referenzpunkt bewertet werden, wobei Verluste stärker gewichtet werden als gleich große Gewinne (*Verlustaversion*). Für den Anlageprozess bedeutet dies eine systematische Verzerrung: Anleger realisieren Gewinne zu früh und halten Verluste zu lange. **Theoretische Verankerung.** Die *Prospect Theory* begründet die *Bias-Hygiene* und Regel-Disziplin des Artefakts (DR4): Feste, vorab definierte Exit-Regeln — ATR-basierte Stops, SMA5-Exit, ein harter T+3-Zeit-Exit — entziehen dem Anleger die diskretionäre Gelegenheit zur verlustaversen Verzerrung. Die Regel ersetzt nicht das bessere Urteil, sondern die schwächste Stelle des menschlichen Entscheiders im Affekt.

Disposition-Effekt. Den Effekt selbst — die Neigung, Gewinner zu früh zu realisieren und Verlierer zu lange zu halten — haben Shefrin und Statman (1985) theoretisch hergeleitet und benannt. Barber und Odean (2000) liefern die empirische Brücke: In einer großangelegten Analyse von Privatanleger-Konten zeigen sie, dass aktiv handelnde Anleger ihre Benchmark systematisch unterperformen — getrieben durch Übermut (*Overconfidence*), Hin-und-Her-Handeln und den *Disposition-Effekt*. Die Underperformance ist mithin nicht primär ein Selektions-, sondern ein Verhaltensproblem. **Verankerung.** Hieraus folgt das *Overconfidence-Friction*-Protokoll (Anti-FOMO): Bei verdächtig einhelligem Signal-Konsens wird bewusst Reibung erzeugt statt Bestätigung, und ein nicht verhandelbarer 60-Tage-Live-Stop institutionalisiert die Anti-*Confirmation-Bias*-Disziplin (vgl. Kap. 4.6). Konsistent damit dekonstruiert das System Perspektiven in unabhängige *Lenses*, deren Widerspruch (*Dissens*) als Risikosignal gewertet wird, statt zu einem trügerischen Konsens gemittelt zu werden (vgl. Kap. 4.3).

Satisficing. Simon (1956) zeigt, dass rationale Akteure unter kognitiven und informationellen Beschränkungen (*Bounded Rationality*) nicht nach dem Optimum, sondern nach einem hinreichend guten Ergebnis streben (*Satisficing*). **Verankerung.** Dies begründet zweierlei: Erstens die *beratende, nicht optimierende* Grundhaltung des Artefakts — es sucht eine disziplinierte, robuste Entscheidung, keine theoretisch optimale; zweitens die *Graceful Degradation* (DR5), bei der das System unter unvollständiger oder veralteter Datenlage defensiv in den NO-TRADE-Modus zurückfällt, statt auf unzureichender Basis zu ranken. Satisficing ist damit nicht Schwäche, sondern die rationale Antwort auf reale Beschränkung.

2.5 Forschungslücke: governance-gesicherte, kostenbewusste Multi-Lens-Orchestrierung

Die vorangehenden Abschnitte legen die Bausteine offen — und zugleich, dass sie in der Literatur unverbunden bleiben. Die Quant-Literatur (Connors & Alvarez, 2009; López de Prado, 2018) adressiert generische Anlagestrategien und Validierungsrigorosität, jedoch ohne agentische Orchestrierung und ohne die harte Kosten-Randbedingung eines Einzelanlegers. Die Literatur zu agentischer KI behandelt generische Multi-Agent-Architekturen, ohne sie governance-gesichert auf den täglichen Anlageentscheidungs-

prozess abzubilden. Die Behavioral-Finance-Forschung (Kahneman & Tversky, 1979; Barber & Odean, 2000) diagnostiziert die behaviorale Leckage präzise, operationalisiert sie aber nicht als architekturelle Governance-Schicht.

Synthese. Am Schnittpunkt dieser drei Stränge klafft die Lücke: **Es existiert kein publiziertes Framework, das heterogene quantitative Anlage-Perspektiven agentisch zu einer governance-gesicherten Tagesempfehlung orchestriert — unter einer harten *Free-Stack*-Kosten-Randbedingung, strikter Mensch-im-Orderpfad-Governance und dem Anspruch radikaler methodischer Ehrlichkeit (*Out-of-Sample*-Nachweis statt Backtest-Hopium).** Genau diese Verschränkung — und nicht eine ihrer Komponenten — ist der Beitrag der vorliegenden Arbeit. Im Sinne der Knowledge Contribution Matrix nach Gregor und Hevner (2013) positioniert sie sich als *Exaptation* auf Framework-Ebene (Übertragung reifer Lösungsartefakte in einen neuartigen Problemraum) und als *Improvement* auf Konstrukt-Ebene (Multi-Lens-Konsens-Dissens, Ablations-Governance, Kosten-Veto, Overconfidence-Friction). Die methodische Operationalisierung dieser Positionierung leistet Kapitel 3.

3. Methodik

Die Arbeit folgt dem Design-Science-Research-Paradigma (DSR) nach Hevner et al. (2004) und implementiert den darauf aufbauenden *Design Science Research Methodology Process* nach Peffers et al. (2007). Dieses Vorgehen ist für den vorliegenden Gegenstand angemessen, weil nicht ein naturwissenschaftliches Phänomen *erklärt*, sondern ein *Artefakt* — das OMLA-Framework und seine lauffähige Phase-A-Instanziierung — entworfen, demonstriert und ehrlich evaluiert wird. Der Autor betont an dieser Stelle ausdrücklich den epistemischen Status der gesamten Arbeit: Sie validiert die Eigenschaften des Artefakts über einen 29-Jahres-Backtest und konvergente Robustheits-/Kostenanalysen und ist **keine** Anlageberatung (vgl. Kap. 1.3). Die folgenden Abschnitte entwickeln das DSR-Mapping (3.1), die konvergente methodische Triangulation (3.2) sowie die acht falsifizierbaren Design-Anforderungen (3.3), die als Bewertungsraster für die Evaluation in Kapitel 5 dienen.

3.1 Design Science Research

Die sechs DSRM-Schritte nach Peffers et al. (2007) wurden für dieses Projekt wie in Tabelle 2 dargestellt operationalisiert. Bemerkenswert ist, dass das Artefakt nicht nachträglich auf einen Forschungsrahmen aufgesetzt, sondern über einen dokumentierten, mehrmonatigen Iterationsverlauf (Versionsstand V6.1 bis V7.21, April bis Juni 2026) entwickelt wurde — die DSRM-Schritte 3 bis 5 sind insofern nicht idealisiert, sondern aus einem real existierenden, im Tagesbetrieb laufenden System rekonstruiert.

DSRM-Schritt (Peffers et al., 2007)	Operationalisierung in dieser Arbeit
1. Problemidentifikation & Motivation	Forschungslücke: Kein publiziertes Framework bildet agentische KI kosten- und governance-gesichert auf den täglichen Anlageentscheidungsprozess eines einzelnen Privatanlegers ab; zugleich behaviorale Leckage als größte Renditebremse (Kap. 1.1–1.2).
2. Lösungsziele definieren	Acht falsifizierbare Design-Anforderungen DR1–DR8, je an eine Kernel-Theorie und ein Falsifikationskriterium gebunden (Kap. 3.3, Tabelle 4).
3. Design & Entwicklung	OMLA-Framework: 13 Analyse-Perspektiven × sieben Cross-Cutting-Layer mit den vier Eigen-Konstrukten Multi-Lens-Konsens-Dissens, Ablations-Governance, Kosten-Veto und Overconfidence-Friction (Kap. 4); live: V6.1-Mean-Reversion-Kern als Phase-A-Instanz.

DSRM-Schritt (Peffers et al., 2007)	Operationalisierung in dieser Arbeit
4. Demonstration	Ausgeführter 29-Jahres-Backtest (1997-01-02 bis 2026-04-15, 14.646 Trades, US-Large-Caps) sowie ein laufender, advisory-only Tageszyklus (Briefing → Empfehlung → manuelle Ausführung → Tracking) am Trade-Republic-Konto (Kap. 5.1–5.4).
5. Evaluation	Dreifache konvergente Triangulation (Kap. 3.2): kosten-/benchmark-ehrliche Re-Analyse gegen SPY Buy-&-Hold, Walk-Forward-IS/OOS-Test und Layer-/Overlay-Ablation — bewertet gegen DR1–DR8 (Tabelle 13).
6. Kommunikation	Dieses Whitepaper bzw. diese wissenschaftliche Projektarbeit.

Tabelle 2: DSRM-Mapping nach Peffers et al. (2007)

Epistemischer Status. Tabelle 2 dokumentiert die *prozessuale Vollständigkeit* des DSR-Vorgehens (alle sechs Schritte sind belegt instanziiert); sie behauptet ausdrücklich nicht, dass Schritt 5 einen externen Wirksamkeitsnachweis erbringt — die Evaluation ist, dem nascent-Charakter des Artefakts entsprechend (Hevner et al., 2004, Guideline 3), überwiegend deskriptiv und am Einzelfall orientiert.

Artefakttypisierung. Nach der Klassifikation von Hevner et al. (2004) vereint das primäre Artefakt zwei Typen: eine **Method** — der orchestrierte Multi-Lens-Bewertungs- und Governance-Prozess (Konsens-Dissens-Aggregation, Ablations-Governance, Kosten-Veto, Overconfidence-Friction) — und eine **Instantiation** — der lauffähige, im Tagesbetrieb befindliche V6.1-MR-Kern mit reproduzierbarer Pipeline (SHA256-Write-Once-Manifeste, Free-Stack-Datensubstrat). Auf Modell-Ebene tritt ergänzend die zweidimensionale Architektur (Pipeline-Phasen × Cross-Cutting-Layer, Abbildung 1) hinzu; der Schwerpunkt des Beitrags liegt jedoch auf *Method* und *Instantiation*.

Beitrags-Positionierung nach Gregor & Hevner (2013). In der *Knowledge Contribution Matrix* positioniert sich der Beitrag zweistufig. **Auf Framework-Ebene** handelt es sich um eine *Exaptation*: reife Lösungsartefakte — klassische Quant-Methoden (Mean Reversion nach Connors & Alvarez, 2009; Positionssizing nach Kelly, 1956, und Vince, 2007; methodische Härtung nach López de Prado, 2018), eine Medallion-artige Datenarchitektur sowie Multi-Agent-Patterns — werden auf einen neuartigen Problemraum übertragen: die governance-gesicherte, strikt kostenbewusste, *beratende* KI-Anlageunterstützung für eine Einzelperson. **Auf Konstrukt-Ebene** liefern die vier Eigen-Konstrukte (Multi-Lens-Konsens-Dissens; Ablations-Governance als Feature-Removal-Design-Zug; Kosten-Veto; Overconfidence-Friction) *Improvement*-Beiträge — neuartige Lösungsmechanismen für ein bekanntes Problem (disziplinierte Signal-Aggregation unter Overfitting- und Bias-Risiko). Diese Doppelpositionierung ist mit der Matrix konsistent, in der ein Artefakt auf verschiedenen Abstraktionsebenen unterschiedliche Quadranten besetzen kann (vgl. Gregor & Hevner, 2013, S. 345). Gregor und Hevner (2013) heben zudem hervor, dass Exaptation-Beiträge bei hohem praktischen Nutzen häufig unterschätzt werden — ein Vorbehalt, den die hier vorgelegte Kombination aus Übertragung etablierter Bausteine und neuartigen Verbindungs-Konstrukten adressiert.

3.2 Methodische Triangulation

Die Validierung folgt einem **konvergenten**, nicht **additiven** Triangulationsverständnis (vgl. Denzin, 1978): Es geht nicht darum, möglichst viele Methoden zu summieren, sondern darum, jede evaluative Kernbehauptung durch genau eine *primäre* und mehrere *unterstützende* Methoden zu stützen, sodass mehrere unabhängige Zugänge auf dasselbe Erkenntnisobjekt konvergieren. Drei Methoden kommen zum Einsatz: (1) die theoriegeleitete **Literaturanalyse** (Kernel-Theorien und Kandidaten-Perspektiven, Kap. 2); (2) der ausgeführte **29-Jahres-Backtest** (1997–2026, 14.646 Trades); und (3) eine eigens für diese Arbeit erstellte **kosten- und benchmark-ehrliche Re-Analyse** (SPY-Buy-&-Hold-Vergleich, Walk-

Forward-IS/OOS, Slippage-Sensitivität, Layer-/Overlay-Ablation). Tabelle 3 macht die Konvergenzlogik je Kernbehauptung explizit.

Methode	Edge existiert & ist kein Overfit	Edge überlebt realistische Kosten	Risiko-vor-Rendite / Krisen-Robustheit
Literaturanalyse (Kernel-Theorien)	unterstützend (EMH-Grenzen, AMH)	unterstützend (Grossman-Stiglitz)	unterstützend (Lo, 2004)
29-Jahres-Backtest (14.646 Trades)	**primär** (Walk-Forward IS/OOS)	unterstützend (Brutto-Erwartungswert, Umschlag)	**primär** (Sharpe/MaxDD-Spread)
Kosten-/Benchmark-Re-Analyse	unterstützend (Ablation: keine Komponente schlägt Baseline)	**primär** (Break-even-Kurve, Slippage-Sensitivität)	unterstützend (Beta, Alpha vs. SPY)

Tabelle 3: Methodische Triangulationsmatrix — Methoden × evaluative Kernbehauptungen (Konvergenzdesign)

Epistemischer Status. Tabelle 3 zeigt, dass jede der drei Kernbehauptungen von mindestens einer primären und einer unterstützenden Methode getragen wird, ohne dass eine einzelne Methode für mehrere primäre Behauptungen verantwortlich ist. Die Validierung ruht damit auf drei konvergenten, einander wechselseitig demarkierenden Methoden: (1) der theoriegeleiteten Literaturanalyse, (2) dem 29-Jahres-Backtest und (3) der kosten- und benchmark-ehrlichen Re-Analyse inklusive Walk-Forward und Ablation. Die methodische Stärke der Anlage liegt in dieser wechselseitigen Demarkation: Wo der Backtest zu optimistisch sein könnte — dokumentierter Same-Day-Close-Fill, Brutto-Kosten —, korrigiert die kostenehrliche Re-Analyse; wo die Re-Analyse die Robustheit prüft, erdet der Walk-Forward-IS/OOS-Test sie an out-of-sample nicht zur Parameterwahl verwendeten Daten; und wo die Literatur den theoretischen Anspruch formuliert, prüft der Backtest dessen empirische Tragfähigkeit out-of-sample.

3.3 Design-Anforderungen (DR1–DR8)

Vor der Artefakt-Konstruktion wurden acht Design-Anforderungen (DR) definiert, die zugleich als Bewertungskriterien für die Evaluation in Kapitel 5 dienen (vgl. vom Brocke et al., 2020). Jede Anforderung ist an eine **Kernel-Theorie** und ein **Falsifikationskriterium** gebunden (Tabelle 4) — diese Bindung macht das Artefakt *prüfbar* statt nur plausibel: Sie benennt für jede Anforderung ex ante die Bedingung, unter der sie als *gescheitert* gälte. Die anschließenden Begründungen folgen durchgängig der Kette Kernel-Theorie → Anforderung → Falsifikationskriterium.

DR	Anforderung	Kernel-Theorie	Falsifikationskriterium
DR1	Empirischer, out-of-sample bestätigter Edge	EMH-Grenzen (Fama, 1970; Grossman & Stiglitz, 1980)	OOS-Sharpe < 0,5 × IS-Sharpe (Overfit)
DR2	Kosten-Robustheit unter realistischer Friktion	Grossman & Stiglitz (1980); Marktmikrostruktur	Edge verschwindet unter realistischer Retail-Friktion (Break-even < ~15 bps)
DR3	Risiko-vor-Rendite / Krisen-Robustheit	Adaptive Markets Hypothesis (Lo, 2004)	kein positiver risikoadjustierter Spread (Sharpe/MaxDD) vs. Benchmark
DR4	Bias-Hygiene & Regel-Disziplin	Prospect Theory (Kahneman & Tversky, 1979)	diskretionäre Overrides schlagen die Regeln systematisch
DR5	Graceful Degradation & Daten-Integrität	Bounded Rationality / Satisficing (Simon, 1956)	System rankt auf veralteten/fehlenden Daten statt NO-TRADE
DR6	Reproduzierbarkeit & Provenienz	Wissenschaftstheorie (Reproduzierbarkeit)	eine Tagesentscheidung ist nicht reproduzierbar
DR7	Mensch-KI-Governance (advisory-only)	Human-AI-Collaboration; KI-Sicherheit	LLM platziert/triggert eine Order

DR	Anforderung	Kernel-Theorie	Falsifikationskriterium
DR8	Falsifikation-First / Ablations-Governance	Popper'sche Falsifikation; López de Prado (2018)	nachweislich edge-lose Komponenten werden beibehalten

Tabelle 4: Design-Anforderungen (DR1–DR8) mit Kernel-Theorie und Falsifikationskriterium

DR1 — Empirischer, out-of-sample bestätigter Edge. Die EMH (Fama, 1970) und ihre Präzisierung im Grossman-Stiglitz-Paradoxon (1980) implizieren, dass in real existierenden Märkten allenfalls ein *dünner* Edge erzielbar ist und dass im Backtest ausgewiesene Überrenditen unter Generalverdacht des Overfittings stehen. Daraus folgt die Anforderung, dass jeder behauptete Edge nicht in-sample geschätzt, sondern an unabhängigen, nicht zur Parameterwahl verwendeten Daten *out-of-sample* bestätigt sein muss. Das Falsifikationskriterium ist operational scharf gefasst: Sinkt das Verhältnis OOS-Sharpe zu IS-Sharpe unter 0,5, gilt der Edge als überangepasst und die Anforderung als verletzt — ein Schwellenwert, der die Strategie im ausgeführten Walk-Forward-Test (IS 1997–2012, OOS 2013–2026; Degradationsverhältnis ~1,04 bzw. ~0,96) bestehen muss, um als nicht-überangepasst gelten zu dürfen (vgl. Kap. 5.4).

DR2 — Kosten-Robustheit unter realistischer Friktion. Aus dem Grossman-Stiglitz-Argument (1980) und der Marktmikrostruktur folgt unmittelbar, dass ein dünner Edge *kostensensitiv* ist: Transaktionskosten, Spreads und Slippage konkurrieren direkt mit der erwarteten Überrendite je Trade. Bei einem Brutto-Erwartungswert von nur +0,401 % pro Trade und einem jährlichen Notional-Umschlag von 18,6× der Equity ist diese Sensitivität für die vorliegende Strategie strukturell hoch. Die Anforderung lautet daher, dass der Edge auch unter realistischer Retail-Friktion risikoadjustiert positiv bleiben muss. Das Falsifikationskriterium bindet dies an die Break-even-Kurve: Läge die kritische Round-Trip-Kostenschwelle, ab der die Strategie SPY risikoadjustiert nicht mehr schlägt, unterhalb von etwa 15 Basispunkten, gälte der Edge als praktisch nicht realisierbar und die Anforderung als verletzt (vgl. Kosten-Sensitivität, Kap. 5.5).

DR3 — Risiko-vor-Rendite / Krisen-Robustheit. Die *Adaptive Markets Hypothesis* (Lo, 2004) verortet jeden Edge als regime- und zeitabhängig und legt nahe, dass eine Mean-Reversion-Strategie nicht als Allwetter-Rendite-Motor, sondern als regimespezifischer — hier defensiver, konvexer — Krisen-Hedge zu verstehen ist. Daraus folgt die Anforderung, das Artefakt primär an risikoadjustierten und Drawdown-Kennzahlen statt an der absoluten Rendite zu messen. Das Falsifikationskriterium verlangt einen positiven risikoadjustierten Spread gegenüber der passiven Benchmark: Erzielte die Strategie keinen überlegenen Sharpe-/MaxDD-Spread gegenüber SPY Buy-&-Hold, verfehlte sie ihren eigentlichen Designzweck. Die Strategie muss diesen Test über den Spread Sharpe 1,249 vs. 0,581 und MaxDD −8,62 % vs. −55,19 % bestehen (vgl. Kap. 5.4).

DR4 — Bias-Hygiene & Regel-Disziplin. Die *Prospect Theory* (Kahneman & Tversky, 1979) und der Dispositionseffekt (Barber & Odean, 2000) identifizieren behaviorale Leckage — Verlustaversion, prozyklisches Hin-und-Her-Handeln, übermütige Overrides — als die empirisch größte Renditebremse des Privatanlegers. Daraus folgt die Anforderung, dass das System Regel-Disziplin erzwingt und diskretionäre Eingriffe systematisch eindämmt (Hard-Exit-T+3, daily-loss-lock, Overconfidence-Friction gegen verdächtigen Konsens). Das Falsifikationskriterium ist der direkte empirische Test: Schlügen diskretionäre Overrides die regelbasierten Entscheidungen *systematisch*, wäre die Disziplin-Hypothese widerlegt und die Anforderung verfehlt; das kohorten-partitionierte Journal (Trennung systematisch/diskretionär) ist genau die Datenbasis, die diese Prüfung erlaubt.

DR5 — Graceful Degradation & Daten-Integrität. Simons Konzept der *Bounded Rationality* bzw. des *Satisficing* (Simon, 1956) begründet, dass ein Agent unter Unsicherheit und unvollständiger Information nicht Optimalität, sondern hinreichend gute *und sichere* Ergebnisse anstreben sollte. Übertragen auf einen Free-Stack-Datenfluss mit unzuverlässigen Quellen folgt die Anforderung, bei fehlenden oder veralteten Daten kontrolliert in einen Defensiv-Modus (NO-TRADE) zu degradieren, statt auf korruptem Substrat zu ranken. Das System operationalisiert dies über ENV-Guard, Canary-Probe (SPY/AAPL/MSFT),

einen COVERAGE-FLOOR (Default 50 % des Universums) und eine „Quellen-Quittung", die bei $\geq$2 fehlenden Pflichtquellen NO-TRADE erzwingt. Das Falsifikationskriterium ist verletzt, sobald das System auf nachweislich veralteten oder fehlenden Daten dennoch eine Empfehlung rankt, statt auf NO-TRADE umzuschalten — ein Fehlerbild, das ein realer certifi-Bruch am 2026-06-08 motiviert hat.

DR6 — Reproduzierbarkeit & Provenienz. Die wissenschaftstheoretische Grundforderung der Reproduzierbarkeit verlangt, dass eine Entscheidung anhand ihrer Eingangsdaten exakt nachvollzogen werden kann. Daraus folgt die Anforderung, jede Tagesentscheidung deterministisch an ein eingefrorenes, versioniertes Daten-Artefakt zu binden. Das System realisiert dies über SHA256-Write-Once-Manifeste (eine Tagesempfehlung referenziert genau das Morgen-Artefakt, gegen das Briefing und Tracking gebaut wurden) sowie über einen festen Zufalls-Seed (`default_rng(42)`) in den Bootstrap-Analysen. Das Falsifikationskriterium ist erfüllt — und die Anforderung verletzt —, sobald sich eine einzelne Tagesentscheidung nicht aus ihren persistierten Eingaben reproduzieren lässt. *Caveat:* Die Reproduzierbarkeit ist für die Tagesartefakte und die kanonischen Backtest-Ergebnisse gegeben; der historische Daten-Lake liegt derzeit als eingefrorener Snapshot (Stand ~2026-04-15) und nicht als vollständig re-baubarer Ingest mit gepinnten Versionen vor (vgl. Limitationen, Kap. 6.6).

DR7 — Mensch-KI-Governance (advisory-only). Aus der Human-AI-Collaboration-Forschung und elementaren KI-Sicherheitsüberlegungen folgt, dass ein potenziell halluzinationsanfälliges LLM in einem finanziell konsequenten Entscheidungspfad niemals exekutiv handeln darf. Daraus folgt die härteste Governance-Anforderung des Frameworks: Das LLM bleibt strikt beratend; jede Order wird vom Menschen manuell im Broker platziert. Diese Grenze ist im System mehrfach abgesichert — die einzige schreibende MCP-Funktion (`run_quick_trade`) ist hinter `confirm=True` hart gegated, das Dashboard bindet ausschließlich an 127.0.0.1, und es existiert keinerlei Broker-API (Trade Republic bietet keine). Das Falsifikationskriterium ist denkbar einfach und absolut: Platziert oder triggert das LLM auch nur eine einzige Order, ist die Anforderung verletzt — „kein LLM jemals im Orderpfad" ist das ausgewiesene Anti-Pattern Nr. 1.

DR8 — Falsifikation-First / Ablations-Governance. Die Popper'sche Falsifikationslogik und die methodische Härtung gegen Backtest-Overfitting nach López de Prado (2018) begründen, dass *Feature-Removal* ein erstklassiger Design-Zug ist: Jede Komponente muss ihren Edge out-of-sample beweisen oder wird abgeschaltet. Daraus folgt die Anforderung, dass nachweislich edge-lose Bestandteile nicht aus Trägheit beibehalten, sondern aktiv deaktiviert werden. Das System operationalisiert dies über `DISABLED_LAYERS` — vier Live-Scoring-Layer (News, Options-Flow, Insider, Market-PCR) wurden nach nicht-positiver OOS-Delta-Erwartung zum 2026-06-01 empirisch abgeschaltet (reversibel, aber inaktiv), und sämtliche Makro-Overlays sowie Circuit-Breaker wurden gegen die Baseline getestet und als performance-mindernd publiziert. Das Falsifikationskriterium ist verletzt, sobald eine als edge-los gemessene Komponente dennoch im Live-Scoring belassen wird. *Caveat:* Die Ablation der vier Layer ruht bislang auf einem einzelnen, dünn besetzten Panel (Tier-A-only, ~68 statt der im projekteigenen Runbook geforderten ~250 Zeilen/Tag); die vollständige H2-Re-Validierung ist als Folgearbeit deklariert (vgl. Kap. 6.6).

4. Das OMLA-Framework (Kernbeitrag)

Das *Orchestrated Multi-Lens Advisory* (OMLA, „Der gläserne Quant") ist ein hybrides Multi-Agenten-Framework, das etablierte quantitative Anlage-Perspektiven mit agentischer KI-Orchestrierung unter einer harten Kosten-Randbedingung verschränkt. Dieses Kapitel stellt das Artefakt als Beitrag vor; die Trennung zwischen der designten *Master-Vision* (dreizehn Perspektiven) und der lauffähigen *Phase-A-Instanziierung* (ein regelbasierter Mean-Reversion-Kern auf US-Large-Caps) wird durchgängig markiert. Das Kapitel beschreibt damit, was das System *ist* und *sein soll* — nicht, was es empirisch *leistet*; die Validierung seiner Eigenschaften ist Gegenstand von Kapitel 5. Jede Design-Entscheidung wird an eine benannte Kernel-Theorie und an die in Kapitel 3.3 eingeführten Design-Anforderungen (DR1–DR8)

zurückgebunden.

4.1 Architekturüberblick

OMLA ist zweidimensional aufgebaut (Abbildung 1). Die **horizontale Dimension** beschreibt den deterministischen Tages-Pipeline-Fluss von der Datenaufnahme bis zur Lern-Rückkopplung; die **vertikale Dimension** beschreibt sieben Querschnitts-Layer (*Cross-Cutting Layers* L1–L7), die das Verhalten in jeder Pipeline-Phase prägen. Die Pipeline-Phasen folgen der Sequenz INGEST → ANALYZE → SCORE → GATE/REVIEW → LOG/LEARN mit einer Lern-Rückkopplung von der letzten zur ersten Phase. In **INGEST** lädt das System ausschließlich kostenfreie Daten (Free-Stack: *yfinance*, Stooq, FRED, GDELT, SEC EDGAR, openinsider.com, capitoltrades.com) und prüft deren Integrität; in **ANALYZE** berechnen die Perspektiven ihre jeweiligen technischen, makroökonomischen und behavioralen Kennzahlen; in **SCORE** werden diese zu einem additiven, eng begrenzten Kompositscore zusammengeführt; in **GATE/REVIEW** entscheidet ein Regime-Gate gemeinsam mit einer Reihe von Governance-Prüfungen über eine Tagesempfehlung (0–5 Kandidaten oder NO-TRADE), die ein Mensch manuell ausführt; in **LOG/LEARN** werden Empfehlung, Ausführung und Ergebnis unveränderlich journalisiert und neue Ideen im *Shadow*-Modus für spätere Falsifikation protokolliert.

Abbildung 1: OMLA-Architektur — Pipeline-Phasen × Cross-Cutting-Layer.

Die sieben Layer wirken orthogonal zu dieser Sequenz. Jeder Layer ist genau einer primären Design-Anforderung zugeordnet, was das Artefakt prüfbar statt nur plausibel macht (vgl. Kap. 3.3).

Layer	Name	Funktion	Design-Anf.
L1	Daten-Reifegrad	Graceful Degradation; bei fehlenden/veralteten Quellen Defensiv-Modus → NO-TRADE statt Müll-Ranking	DR5
L2	Signal-Governance	Ablations-Governance: edge-lose Perspektiven werden abgeschaltet; Kosten-Veto (kein bezahltes Datum)	DR2, DR8
L3	Mensch-in-the-Loop	advisory-only Gate; das LLM platziert nie eine Order	DR7
L4	Risiko-Adaptivität	Regime-abhängiges *Sizing* (Kelly/Optimal-f, Vol-Targeting, Recovery-Faktor, DD-Lock)	DR3
L5	Lernen / Shadow-Logging	Shadow-first-Disziplin; Kandidaten werden vor jedem Go-Live forward-evaluiert	DR1, DR8
L6	Bias-/Verhaltens-Hygiene	Overconfidence-Friction, Pflicht-Gegenargument, Size-Freeze, 60-Tage-Live-Stop	DR4
L7	Integration / Reproduzierbarkeit	Write-Once-Artefakte (SHA256-Manifest), Provenienz, Quellen-Quittung	DR6

Tabelle 4a: OMLA-Architektur — Cross-Cutting-Layer L1–L7 mit DR-Zuordnung

Epistemischer Status. Die Tabelle dokumentiert die *intendierte* Zuordnung von Querschnittsfunktion zu Anforderung; sie behauptet nicht, dass jeder Layer in der lauffähigen Instanz vollständig realisiert ist. Mehrere Layer sind in der Phase-A-Instanz nur partiell aktiv — so ist die Risiko-Adaptivität (L4)

jenseits des Regime-Gates und der Stop-Logik bewusst auf den nicht-eskalierenden Zustand eingefroren, solange die Out-of-Sample-Edge-Re-Validierung (vgl. Kap. 5.6) offen ist. Diese Lücke zwischen Vision und Live-Betrieb wird in Kapitel 6.6 als Limitation geführt.

4.2 Hybridansatz: Die 13 Analyse-Perspektiven (Lenses)

Die *Master-Vision* dekonstruiert den Anlageentscheidungsprozess in dreizehn unabhängige Analyse-Perspektiven („*Lenses*"), die jeweils wie ein eigenständiger *Senior-Researcher* eine spezifische Leitfrage an dieselben Daten stellen. Das methodische Prinzip ist dabei nicht additive Meinungsmittelung, sondern *Risiko-Aggregation*: Jede Perspektive bringt ein eigenes Risikokonzept mit, und kein einzelnes Signal regiert die Entscheidung allein. Die theoretische Verankerung dieser Pluralität liefert die *Adaptive Markets Hypothesis* (Lo, 2004), nach der ein Edge regime- und zeitabhängig ist — eine einzelne Perspektive kann daher nur regimespezifisch tragen, was die Existenz mehrerer, teils gegenläufiger Linsen rechtfertigt.

Tabelle 5 listet die dreizehn Perspektiven mit ihrer Rolle, ihrer Free-Stack-Datenquelle und ihrem **Reifegrad** auf. Der Reifegrad ist dabei ehrlich aus dem code-verifizierten Systembefund abgeleitet und unterscheidet drei Zustände: *live* (im produktiven Scoring wirksam), *shadow* (berechnet und protokolliert, aber ohne Einfluss auf die Entscheidung) und *spezifiziert* (konzeptionell ausgearbeitet, aber nicht implementiert oder als Folgearbeit deklariert). Mehrere ursprünglich als Scoring-Layer vorgesehene Perspektiven sind nach negativem Out-of-Sample-Erwartungswert per `DISABLED_LAYERS` abgeschaltet (vgl. Kap. 4.5 und 5.6) und damit auf *shadow* zurückgestuft.

#	Perspektive	Rolle / Leitfrage	Free-Stack-Datenquelle	Reifegrad
P1	Multi-Lens-Audit (Hedgefonds-Manager)	Wie risikoadjustiert ist die Rendite wirklich? (Sharpe, Sortino, Walk-Forward, TCA)	Eigenes Journal (SQLite), Backtest-Korpus	*live*
P2	Bias-Hygiene (Behavioral-Ökonom)	Welche Biases zerstören den Edge?	Eigenes Bias-Journal, Regelwerk	*live*
P3	Gamma-Overlay (Options-Dealer)	Wo sitzt das Gamma, wohin hedgt der Markt?	CBOE-Optionschain (paywalled), *yfinance* `option_chain` (Fallback)	*spezifiziert*
P4	Makro-Liquidität (Zentralbank-Watcher)	Wie füllt/entzieht Liquidität dem Markt?	FRED (NFCI, STLFSI4, Netliq-Proxies)	*shadow*
P5	Credit-Spreads (Fixed-Income-Analyst)	Was sagt der Credit-Markt vor dem Equity-Markt?	FRED (HY-OAS, Kurven-Slope, 1997 ff.)	*shadow*
P6	Vol-Skew (Volatilitäts-Stratege)	Wann sind Hedges wirklich billig?	*yfinance* (^VIX 1990 ff., ^MOVE), CBOE	*spezifiziert*
P7	Regime-Trend (systematischer CTA)	In welchem Marktregime befinden wir uns?	*yfinance* (VIX, ADX, SPY-Trend, 10Y, DXY, Öl, Gold)	*live*
P8	Forensik (Bilanz-Analyst)	Welche Namen sind rechnerisch unsauber?	*yfinance* Fundamentaldaten (Beneish-M-Score; Beneish, 1999)	*spezifiziert*
P9	ML-Research (López de Prado)	Ist der Backtest methodisch sauber?	Eigener Backtest-Lauf, HMM-Regime (Shadow)	*shadow*
P10	Microstructure (Dark-Pool/Tape)	Was tun Institutionelle intraday?	*yfinance* `option_chain`, Options-Flow-Proxy	*shadow*
P11	Event-Driven (Merger-Arb/PEAD)	Welche Kalender-Edges existieren?	*yfinance* Earnings-History, GDELT	*shadow*
P12	Risk-Parity (Makro-Allokator)	Wie ist das Risikobudget pro Sleeve verteilt?	Eigenes Journal, Kelly-/Optimal-f-Statistik	*shadow*
P13	Behavioral / Retail-Contrarian (Sentiment)	Was tut die Masse, wo liegt Contrarian-Edge?	GDELT, Reddit-Trending (Shadow), Market-PCR	*shadow*

Tabelle 5: Die 13 Analyse-Perspektiven (Lenses) mit Rolle, Datenquelle und Reifegrad

Befund. Tabelle 5 zeigt die zentrale Ehrlichkeits-Asymmetrie des Artefakts: Von dreizehn Perspektiven sind in der Phase-A-Instanz lediglich vier *live* wirksam (Multi-Lens-Audit, Bias-Hygiene, Regime-Trend sowie — als positiv getestete *boost*-Overlays — Sektor-Relative-Strength und Multi-Timeframe, die dem Regime-Trend zugeordnet sind); der live tragende Score reduziert sich code-verifiziert auf einen technischen Mean-Reversion-Kern (Triple-RSI + Connors-Quality) plus zwei *boost*-Anpassungen. Vier ursprünglich aktive Scoring-Layer — *News*, *Options-Flow*, *Insider* und *Market-PCR* — sind seit dem 01.06.2026 per `DISABLED_LAYERS = {news, options_flow, insider, market_pcr}` abgeschaltet, da ihr Δ-Erwartungswert in den Out-of-Sample-Splits nie positiv war. Die Tabelle zeigt damit *nicht*, dass dreizehn Agenten produktiv kooperieren; sie zeigt einen designten Möglichkeitsraum, dessen live tragende Teilmenge bewusst klein und falsifikationsgetrieben gehalten ist. Diese Demarkation ist nicht Schwäche, sondern Beitrag (DR8): Der publizierte Negativbefund über das Nicht-Tragen mehrerer Perspektiven ist Teil der methodischen Aussage (vgl. Kap. 5.6).

4.2.1 Abgrenzung zu bestehenden Frameworks

Die einzelnen Methoden, aus denen OMLA schöpft — Connors-artige Mean-Reversion (Connors & Alvarez, 2009), Walk-Forward-Validierung und Triple-Barrier-Logik (López de Prado, 2018), Kelly-/Optimal-f-*Sizing* (Kelly, 1956; Vince, 2007), Medallion-artige Datenarchitektur — sind reife Bausteine des klassischen Quant-Stacks. Ein direkter Feature-Vergleich auf diesen Bausteinen wäre wenig aussagekräftig; OMLAs Beitrag liegt nicht in einer neuen Einzelmethode, sondern in deren governance-gesicherter, kostenbewusster Orchestrierung für eine Einzelperson. Die Abgrenzung erfolgt daher zweistufig.

Gemeinsame Merkmale (was OMLA mit dem klassischen Quant-Stack teilt):

Merkmal	Klassischer Quant-Stack	Robo-Advisor	**OMLA**
Regelbasierte Signal-Erzeugung	Ja	Teilweise	Ja
Backtest-/OOS-Validierung	Ja	Selten	Ja (Walk-Forward)
Risiko-vor-Rendite-Orientierung	Ja	Ja	Ja (Krisen-Hedge)
Positionsgrößen-Steuerung	Ja (Kelly/Vol)	Ja	Ja (Kelly/Optimal-f)

OMLA-spezifische Alleinstellungsmerkmale (Konstrukte, die durch die agentische, kostenbewusste Architektur ermöglicht werden und im klassischen Stack so nicht existieren):

1. **Multi-Lens-Konsens-Dissens** — Perspektiven-Widerspruch wird nicht weggemittelt, sondern als Risiko-/Chancensignal behandelt (vgl. Kap. 4.3).
2. **Ablations-Governance** — *Feature-Removal* ist ein erstklassiger Design-Zug; jede Perspektive muss ihren Edge out-of-sample beweisen oder wird abgeschaltet (vgl. Kap. 4.5).
3. **Kosten-Veto** — eine harte Free-Stack-Randbedingung: kein bezahltes Datum, auch um den Preis abgeschalteter Perspektiven (vgl. Kap. 4.5).
4. **Overconfidence-Friction** — proaktive Anti-FOMO-Reibung bei verdächtigem Konsens, plus ein nicht-verhandelbarer 60-Tage-Live-Stop gegen Confirmation-Bias (vgl. Kap. 4.6).

Tabelle 6: Abgrenzung OMLA — gemeinsame Merkmale und Alleinstellungsmerkmale

Epistemischer Status. Tabelle 6 stützt die DSR-Positionierung des Beitrags als *Exaptation* auf Framework-Ebene und *Improvement* auf Konstrukt-Ebene nach Gregor und Hevner (2013): Die reifen

Bausteine werden auf einen neuartigen Problemraum (governance-gesicherte, kostenbewusste, beratende KI-Anlageunterstützung für eine Einzelperson) übertragen, während die vier benannten Konstrukte einen eigenständigen Konstrukt-Beitrag darstellen. Die Tabelle behauptet keine Überlegenheit gegenüber den Vergleichssystemen — sie grenzt ab, sie misst nicht.

4.3 Multi-Lens-Konsens-Dissens (das zentrale Governance-Konstrukt)

Das konzeptionell neuartigste Element von OMLA ist die Behandlung von Widersprüchen zwischen den Perspektiven. In klassischen Multi-Agenten-Systemen gilt Dissens als Problem, das durch Konsensverfahren oder erzwungene Mittelung aufzulösen sei. OMLA invertiert diese Perspektive:

> **Wenn unabhängige Perspektiven sich widersprechen, ist dies kein Rauschen — es ist ein Risiko- bzw. Chancensignal.**

Die methodische Begründung liegt im Leitprinzip der Master-Vision: Wo zwei Perspektiven dasselbe Problem signalisieren, gilt der Befund als härter validiert; wo sie sich widersprechen, wird die Spannung dokumentiert und in der Portfolio-Governance aufgelöst, nicht wegargumentiert. Theoretisch stützt sich dies auf das Grossman-Stiglitz-Paradoxon (1980) und die *Adaptive Markets Hypothesis* (Lo, 2004): Wenn der real existierende Edge dünn und regimeabhängig ist, dann ist die seltene Übereinstimmung gegenläufiger Linsen ein verlässlicheres Signal als der Konsens gleichgerichteter — und gleichzeitig ist ein zu glatter, allseitiger Konsens ein Verdachtsmoment für ein überoptimistisches oder überangepasstes Signal (Anschluss an die Overconfidence-Friction, Kap. 4.6).

Instanziierung. In der Phase-A-Instanz manifestiert sich das Konstrukt nicht als freie Meinungsabstimmung, sondern als additive Score-Komposition unter einem Regime-Gate. Der technische Mean-Reversion-Pre-Score (code-verifiziert `pre = 0,667 · triple_rsi + 0,333 · connors_scaled`) wird durch eng begrenzte, einzeln auditierbare *adjust*-Overlays moduliert (Sektor-RS $\pm$0,30, Multi-Timeframe $\pm$0,30), wobei die Summe der Anpassungen hart auf $\pm$1,0 geklemmt ist. Über diesem Score entscheidet ein regelbasiertes Regime-Gate (sieben Regime, Größenmultiplikator 0,3–1,2, erlaubte Tiers, Intermarket-Vetos für WTI, 10Y, DXY und VIX). Manifestiert sich Dissens — etwa wenn das Regime „Bullish-Confirmed" einem Mean-Reversion-Entry widerspricht (historisch negativer Edge) —, reduziert das System die Positionsgröße oder wählt NO-TRADE, statt eine erzwungene Mittelung zu handeln (DR3, DR4). Eine Pflicht-Reibung verlangt zudem, dass jede Empfehlung mindestens ein widersprechendes Gegenargument explizit ausweist (Anti-Confirmation-Bias).

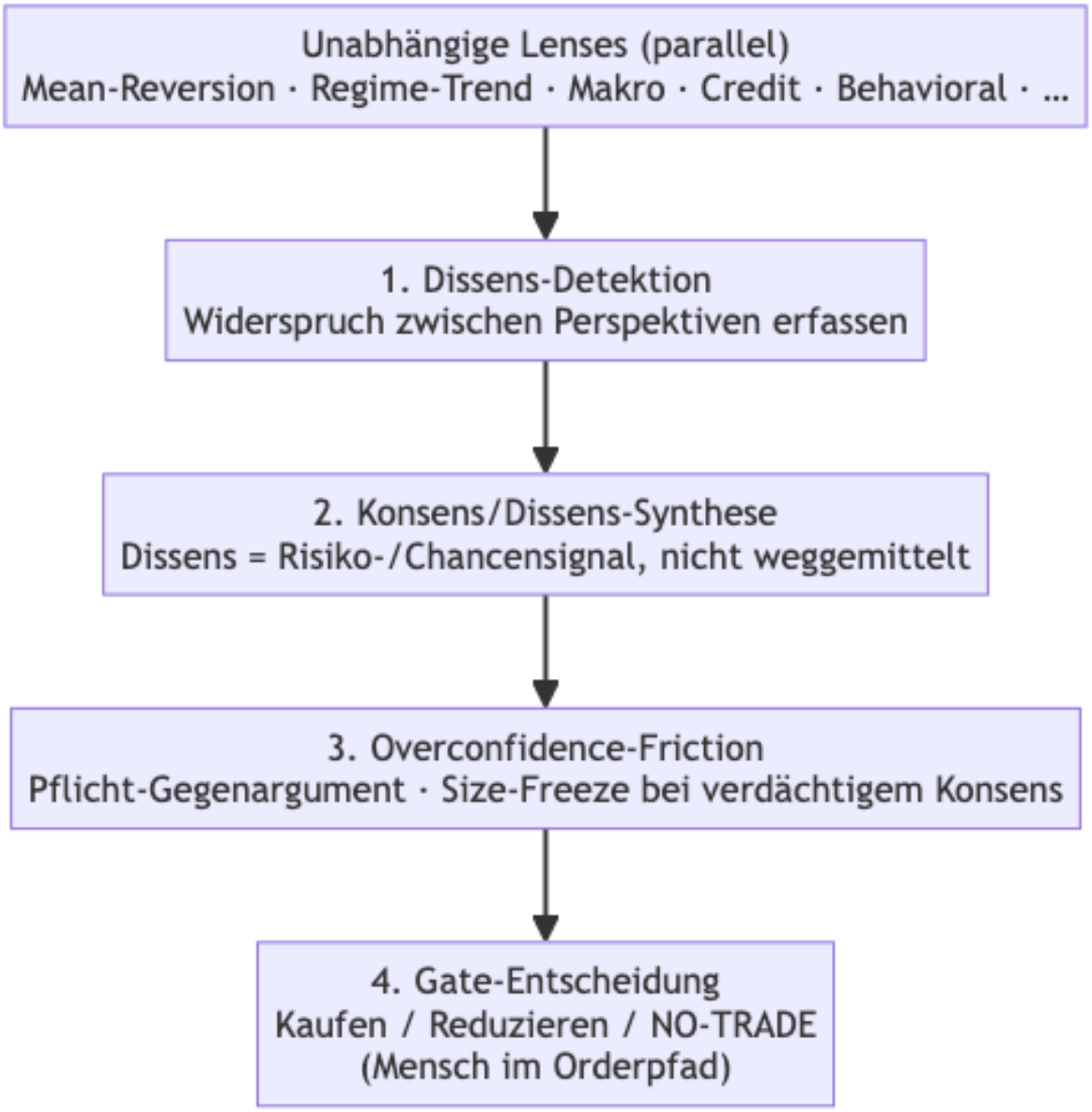


Abbildung 2: Multi-Lens-Konsens-Dissens — Governance-Fluss je Tages-Gate.

Epistemischer Status. Die hier beschriebene Mechanik ist konzeptionell vollständig und in der Phase-A-Instanz für den Mean-Reversion-Kern realisiert; die volle Multi-Lens-Abstimmung über alle dreizehn Perspektiven ist jedoch designt, nicht live (vgl. Tabelle 5). Der Existenzbeweis betrifft das *Prinzip* (Dissens senkt Size oder erzwingt NO-TRADE), nicht dessen dreizehnstimmige Vollform.

4.4 Graceful Degradation (L1: Datenreifegrad)

Ein unbeaufsichtigt laufendes Privatanleger-System steht vor einem fundamentalen Integritätsproblem: Eine fehlende oder veraltete Quelle darf nicht dazu führen, dass auf Müll-Daten gehandelt wird. OMLA löst dies über eine *Graceful Degradation*, deren theoretische Fundierung Simons *Satisficing*-Konzept (Simon, 1956) bildet — unter Unsicherheit ist der defensive, hinreichend gute Zustand dem scheinpräzisen, aber fragilen vorzuziehen. Die Instanziierung erfolgt über `tools/data_health.py` (Quellen-Erreichbarkeit und Pool-Abdeckung), eine `ENV-GUARD` mit *certifi*-Selbsttest und einer Canary-Probe über SPY/AAPL/MSFT sowie einen `COVERAGE_FLOOR` (Standard: 50 % des Universums). Fällt die Abdeckung unter den Schwellenwert, schreibt das System bewusst *kein* Scan-Artefakt, sondern eine

`scan_abort`-Quittung — eine fehlende Empfehlung (sicher → NO-TRADE) ist einem fehlerhaften Ranking stets vorzuziehen.

Stufe	Bezeichnung	Auslöser	System-Verhalten
1	Vollbetrieb	Alle Pflichtquellen frisch (z. B. 6/7 OK), Abdeckung $\geq$ Floor	Normales Scoring & Ranking, reguläre Empfehlung
2	Eingeschränkt	Einzelne nicht-kritische Quelle fehlt/abgeschaltet (z. B. `options_flow` DISABLED)	Scoring auf verbleibender Basis; betroffener Layer auf 0 gezwungen; Hinweis in Quellen-Quittung
3	Defensiv-Modus	≥ 2 fehlende Pflichtquellen / Abdeckung < Floor	NO-TRADE (Defensiv-Modus); Quellen-Quittung tallyt Fehlversuche
4	Abbruch / Quittung	Umgebung defekt (Canary tot, CA-Bundle, Rate-Limit-Abort)	Kein Artefakt; `scan_abort` mit Fix-Hinweis; Exit-Code 4/5; fail-loud statt fail-silent

Tabelle 7: Graceful Degradation nach Datenreifegrad

Befund. Tabelle 7 dokumentiert die ehrliche Beobachtung, dass die meisten Tage *nicht* im Vollbetrieb laufen: Der jüngste `data_health`-Stand (12.06.2026) meldet 6/7 Quellen OK bei abgeschaltetem `options_flow` (Stufe 2), und jedes Tagesdokument schließt mit einer Quellen-Quittung, die fehlende Pflichtquellen zählt und bei ≥ 2 Ausfällen in den Defensiv-Modus degradiert. Die Tabelle zeigt damit, dass NO-TRADE der Normalfall-Schutz und nicht die Ausnahme ist (DR5); sie zeigt *nicht*, dass das System auf Stufe 4 jemals fehlerfrei reagiert hat — die Resilienz-Logik ist durch dokumentierte Live-Inzidente (etwa der `COVERAGE_FLOOR`-Abbruch durch Yahoo-Rate-Limits) iterativ gehärtet, aber nicht formal verifiziert.

4.5 Ablations-Governance & Kosten-Veto (L2)

Die **Ablations-Governance** erhebt das Entfernen von Funktionalität zu einem erstklassigen Design-Zug. Die theoretische Verankerung bildet die Popper'sche Falsifikation in der Lesart von López de Prado (2018): Eine Komponente verbleibt nur dann im System, wenn sie ihren Edge out-of-sample beweist; andernfalls wird sie abgeschaltet. Operativ ist dies über die Menge `DISABLED_LAYERS` implementiert. Nach einer H2-Ablation (01.06.2026) sind vier Scoring-Layer — `news`, `options_flow`, `insider` und `market_pcr` — hart deaktiviert, da ihr Δ-Erwartungswert in den Out-of-Sample-Splits nie positiv war; die Deaktivierung erfolgt durch inline-Nullsetzen der jeweiligen *adjust*-Anpassung und ist durch Entfernen des Namens aus der Menge reversibel. Bezeichnend ist, dass die abgeschalteten Layer weiterhin als informationelle Artefakte berechnet und in den nicht-blockierenden Pipeline-Schwanz verschoben werden — *Feature-Retirement* ohne Löschung. Auch flankierende Befunde stützen das Prinzip: Sämtliche getesteten Makro-Overlays und Circuit-Breaker kosteten in der Ablation Performance, und das Congress-Smart-Money-Signal wurde nach gemessenen lediglich +0,2 Prozentpunkten Trefferquote-Uplift auf „nur informationell" zurückgestuft (vgl. Kap. 5.6).

Das **Kosten-Veto** ist die zweite harte Randbedingung dieses Layers: kein bezahltes Datum, ausnahmslos. Die theoretische Begründung liefert das Grossman-Stiglitz-Paradoxon (1980) — in einem nahezu effizienten Markt ist der real verbleibende Edge so dünn, dass die Kosten der Informationsbeschaffung ihn aufzehren können; ein Privatanleger ohne Skaleneffekte muss daher Datenkosten strukturell vermeiden. Das Veto manifestiert sich code-verifiziert in der Free-Stack-Architektur (E8-Cost-Veto, Ersatz von Quiver, Unusual Whales und CBOE-Historie durch SEC-EDGAR-, openinsider-, capitoltrades- und FRED-Scrapes) und akzeptiert bewusst Abdeckungs- und Reifegrad-Lücken als Preis: So ist die

historische CBOE-Put-Call-Ratio seit 2024 paywalled und wird ehrlich als Gap geführt, statt sie zu kaufen.

Für die Praxis: Ablations-Governance und Kosten-Veto greifen ineinander — gerade weil keine bezahlten Daten den Edge stützen dürfen, muss jede Perspektive ihren Beitrag auf Free-Stack-Daten beweisen, und mehrere bestehen diesen Test nicht. Das ist der designte Mechanismus gegen *Backtest-Overfitting* und gegen den szientistischen Anstrich, vor dem Kapitel 1.1 warnt.

4.6 Overconfidence-Friction & Mensch-im-Orderpfad (L3/L6)

Die größte Renditebremse des Privatanlegers ist nicht fehlendes Alpha, sondern behaviorale Leckage (Barber & Odean, 2000; Kahneman & Tversky, 1979). OMLA adressiert dies über zwei verschränkte Layer. Die **Overconfidence-Friction** (L6) ist eine proaktive Reibung gegen überoptimistisches Handeln: Nach drei aufeinanderfolgenden Gewinntrades friert ein Auto-Size-Cap die Positionsgröße für 48 Stunden auf den Standard ein; jede Empfehlung muss ein widersprechendes Gegenargument tragen (Anti-Confirmation); Stops sind ausschließlich harte Broker-Stops relativ zum aktuellen ATR, nie mentale Anker am Einstandskurs (Anti-Anchoring, Anti-Loss-Aversion); Nachkäufe in Verlustpositionen sind regelseitig untersagt (Anti-Sunk-Cost). Theoretisch ist dies in der *Prospect Theory* (Kahneman & Tversky, 1979) und in Simons *Satisficing* (Simon, 1956) verankert — das System strebt regeldiszipliniert nach „gut genug" statt nach diskretionärer Optimalität (DR4). Als härteste Anti-FOMO-Maßnahme setzt der Autor einen nicht-verhandelbaren **60-Tage-Live-Stop** (Stichtag 04.07.2026): Bestätigt der Live-Betrieb den Edge bis dahin nicht, wird ein Pivot erzwungen — eine bewusste Konstruktion gegen den eigenen Confirmation-Bias.

Der **Mensch-im-Orderpfad** (L3) ist die strikte Governance-Grenze: Das LLM und die Pipeline sind *advisory-only* — sie erzeugen disziplinierte, auditierbare Markdown-Empfehlungen, platzieren aber **nie** eine Order. Die Architektur erzwingt dies auch technisch: Der Broker Trade Republic besitzt keine Order-API, sodass jede Ausführung physisch und manuell durch den Menschen erfolgt; selbst der einzige schreibende MCP-Endpunkt (`run_quick_trade`) ist hart hinter `confirm=True` gegated und dient ausschließlich der nachträglichen Journalisierung. Theoretisch folgt dies dem Human-AI-Collaboration-Prinzip und Erwägungen der KI-Sicherheit: Ein KI-System, das undurchsichtige Order ohne nachprüfbaren Edge auslöste, wäre schlimmer als nützlich (vgl. Kap. 1.1). Die Falsifikationsbedingung von DR7 ist scharf — sie gilt als verletzt, sobald das LLM eine Order platziert oder triggert.

Epistemischer Status. Diese Layer beschreiben Governance-*Regeln*. Die designte Verschränkung von Kernel-Theorie und durchgesetzter Regel ist der Beitrag; eine gemessene Verhaltenswirkung über alle dreizehn Perspektiven ist designt, nicht über die Phase-A-Instanz hinaus instanziiert.

4.7 Phasenübergreifender Tageszyklus

Der OMLA-Tageszyklus integriert die Pipeline-Phasen (Kap. 4.1) und die Querschnitts-Layer zu einem deterministischen, *launchd*-getriebenen Ablauf um zwei tägliche Fixpunkte (Briefing am Morgen, Tracking am Abend) plus wöchentliche Rituale. Code-verifiziert laufen vier Jobs unter `com.louis.dailyinvest`: ein News-Job (06:30 CET), die deterministische Daten-Spine `run_morning_pipeline.py` (08:00 CET, neun fail-soft Schritte in rate-budget-optimierter Reihenfolge), ein Katalysator-Job (Sonntag 20:00 CET) und ein Gatekeeper (Montag 09:45 CET, Selbstheilung fehlender Shadow-/Tripwire-Logs). Aus Anwendersicht erzeugt das System drei advisory-only Markdown-Dokumente pro Handelstag:

- **06:30 CET — Briefing (INGEST/ANALYZE):** Globale Markt-Synthese (Asien-Schluss, EU-Vorbörse, US-Futures, Intermarket) mit Regime-Klassifikation, Intermarket-Veto-Tabelle und Quellen-Quittung; degradiert bei ≥ 2 fehlenden Pflichtquellen sichtbar in den Defensiv-Modus.

- **09:00 CET — Empfehlung (SCORE/GATE):** Entscheidung zuerst (NO-TRADE oder 0–5 Kandidaten mit Entry/SL/TP/R:R/Hard-Exit-Tag), nach einem 15-Punkte-Selbstcheck, mit aktiven Vetos, übersprungenen Kandidaten, Compliance-Erinnerung (YTD), Quellen-Quittung und einem exakten Journal-Persistierungsblock.
- **Manuelle Ausführung (L3):** Der Mensch platziert jede Order physisch in Trade Republic; das System bleibt außerhalb des Orderpfads.
- **20:00 CET — Tracking (LOG/LEARN):** Compliance-Status offener Positionen (dreifach gegen die kanonische Datenbank verifiziert), geschlossene Trades, Performance-Bilanz, Regime-Gate-Abgleich, Kennzahlen-Refresh (Sharpe/Sortino/MaxDD/Calmar) und ein Compliance-Score.

Die **Lern-Rückkopplung** schließt den Zyklus: Nahezu jede neue Idee — alternative Strategien, HMM-Regime, Optimal-f-*Sizing*, Reddit-Sentiment, Congress-Signale — wird im *Shadow*-Modus forward-protokolliert und vor jedem Go-Live empirisch evaluiert (L5). Genau diese Shadow-Disziplin ist die Quelle der in Kapitel 4.5 beschriebenen Ablations-Entscheidungen: Die Abschaltung der vier Scoring-Layer und die Zurückstufung des Congress-Signals sind datengetriebene Konklusionen dieses Loops, keine Annahmen. Reproduzierbarkeit sichert L7 durch Write-Once-Artefakte mit SHA256-Manifest, sodass jede Tagesentscheidung nachvollziehbar bleibt (DR6).

Epistemischer Status. Der beschriebene Zyklus ist die produktiv laufende Phase-A-Instanz; die Kadenz- und Quellen-Quittungs-Disziplin sind die definierenden Produkteigenschaften. Der Zyklus belegt die *Lauffähigkeit* und *Reproduzierbarkeit* des Artefakts; die Eigenschaftsvalidierung selbst ruht auf dem 29-Jahres-Backtest und den konvergenten Robustheits-/Kostenanalysen (vgl. Kap. 5.4–5.6).

5. Prototyp und Validierung

> Alle Performance-Zahlen dieses Kapitels stammen aus dem kanonischen Engine-Lauf `backtest/results_v1/` (2026-04-15) sowie den dezidierten Robustheits-Modulen (`results_walkforward/`, `results_overlay/`, `results_circuit_breakers/`, `results_survivorship/`) und der eigens für diese Arbeit erstellten kosten-/benchmark-ehrlichen Re-Analyse (`scientific_addendum.py`). Die Methodik wird je Tabelle offengelegt; jede starke Tabelle trägt unmittelbar einen Satz zu ihrem epistemischen Status. Dieses Kapitel validiert die Eigenschaften des Artefakts über einen 29-Jahres-Backtest und konvergente Robustheits-/Kostenanalysen (Walk-Forward, Ablation, Kosten-Sensitivität); es ist **keine Anlageberatung** (vgl. Kap. 1.3).

5.1 Technische Architektur

Die lauffähige Phase-A-Instanziierung von OMLA — der V6.1-Mean-Reversion-Kern — ist als deterministische, vollständig lokale Python-Pipeline realisiert, die das Kosten-Veto (DR2) auf Architektur-Ebene durchsetzt: Kein Bestandteil des produktiven Pfades konsumiert ein bezahltes Datum. Der Entwurf folgt damit unmittelbar der Grossman-Stiglitz-Logik (1980), wonach der real existierende Edge eines Privatanlegers *dünn* und *kostensensitiv* ist und jede fixe oder marginale Datengebühr ihn potenziell aufzehrt (vgl. Kap. 5.5). Tabelle 8a dokumentiert den Stack und ordnet jede Komponente der durch sie adressierten Design-Anforderung zu.

Schicht	Komponente (Free-Stack)	Funktion	DR-Bezug
Sprache/Numerik	Python, *pandas*, *numpy*, *pyarrow*	deterministische Indikatorberechnung (Wilder-RSI, ATR, SMA), Backtest-Engines	DR6

Schicht	Komponente (Free-Stack)	Funktion	DR-Bezug
Marktdaten	*yfinance* (primär) + *Stooq*-Fallback via `data_source_orchestrator`	OHLCV-Preise, Optionsketten, Earnings — kostenfrei, ohne API-Schlüssel	DR2
Makro/Sentiment	FRED (41 Serien), AAII/NAAIM, SEC EDGAR Form-4	Credit-Spreads, Liquidität, Vol, Insider — kostenfrei	DR2, DR5
Journal/State	*SQLite* (WAL-Modus), `trade_journal.sqlite`	unveränderliches, kohorten-partitioniertes Trade-Journal, Setup-/Kohorten-Statistik	DR6
Datensubstrat	Apache *Parquet*-Datenlake (212 Serien, ~4,06 Mio. Zeilen, 152,6 MB), *DuckDB*/*Polars*	30-Jahres-Historie, Walk-Forward-Substrat; SHA256-Provenienz je Serie	DR6
Daten-Health	`data_health.py` (v1.1), `dq_check.py`	Quellen-Erreichbarkeit, Coverage-Floor, Canary-Probe (SPY/AAPL/MSFT) → NO-TRADE bei Degradation	DR5
Agenten-Schnittstelle	*FastMCP*-Server	exponiert Indikatoren/Empfehlungen an den KI-Co-Piloten (advisory-only, write-gated)	DR7
Scheduling	macOS *launchd*-Cron (`install_launchd_crons.sh`)	zwei Tagesfixpunkte (08:00/20:00 CET), idempotenter Catch-up	DR5, DR6
Cockpit	*Flask*/SSE-Dashboard, *Chart.js*, PWA (Port 5757)	Zustandsvisualisierung, Risk-Endpoint	—

Tabelle 8a: Technologie-Stack des Prototyps (Phase-A-Instanziierung) mit DR-Zuordnung.

Epistemischer Status. Die Tabelle belegt die technische Realisierbarkeit und die durchgängige Free-Stack-Disziplin des Artefakts; sie ist kein Performance-Nachweis. Bemerkenswert ist, dass die operationale Resilienz (dreischichtige Cron-Ausfall-Abwehr, `certifi`-Selbsttest, Coverage-Floor mit Abbruch) relativ zu einem Privatdepot überdimensioniert erscheint — ein bewusster Reflex auf drei reale macOS-Sleep-Ausfälle im Juni 2026 und auf die Erkenntnis, dass die größte Gefahr nicht ein fehlendes Signal, sondern ein *stilles* Ranking auf degradierten Daten ist (DR5).

5.2 Implementierung des Kerns (V6.1 Mean-Reversion)

Der MR-Kern instanziiert eine Connors-artige *Mean-Reversion*-Strategie (Connors & Alvarez, 2009) auf US-Large-Caps. Die kanonische Backtest-Baseline (`backtest_engine.py`) ist bewusst minimal gehalten — vier Entry-Bedingungen, drei Exit-Pfade —, da die Ablations-Validierung (vgl. Kap. 5.6) zeigt, dass jede zusätzliche Komplexität den Edge eher kostet als mehrt:

- **Entry.** Close > SMA200 (Trendfilter), 2-Perioden-RSI < 25 (Überverkauft-Signal), Tagesrendite ret_1d > −5 % (Crash-Schutz). Ranking der Kandidaten nach aufsteigendem RSI-2 (am stärksten

überverkauft zuerst).

- **Exit.** Close > SMA5 (*Mean-Reversion*-Ziel) **oder** Intraday-Tief ≤ ATR-Stop (2,5 × ATR-10) **oder** Zeit-Stop nach maximaler Haltedauer. Im Live-System verschärft ein **Hard-Exit-T+3** den Zeit-Stop (verifiziert in `exit_engine.py`: `HARD_EXIT_MAX_HOLD = 3`, `HARD_EXIT_MIN_RET_PCT = 1.0` — Glattstellung am nächsten Open, falls ein Trade nach drei Handelstagen nicht ≥ +1 % liegt).
- **Sizing & Governance.** 1 % Risiko pro Trade (Dollar-Risiko = Kapital × 1 % / [Entry − Stop]), maximal 10 Positionen, 1 € Kommission pro Trade. Im Live-Pfad ergänzt durch Cohort-Kelly-Sizing (half-/quarter-Kelly, harte 20-%-Kappe; Kelly, 1956; Vince, 2007) und ein Regime-Gate (`regime_gate.py`), das die Positionsgröße über einen Multiplikator 0,3–1,2 an das Marktregime koppelt.

Live-vs-Ideal-Divergenz (DR8). Das Live-System ist gegenüber der dokumentierten Master-Vision **ehrlich abgespeckt**: Mit Stand 2026-06-01 sind vier Scoring-Layer (`news`, `options_flow`, `insider`, `market_pcr`) per `DISABLED_LAYERS` hart deaktiviert (`daily_scanner.py`, Zeile 56), weil ihre Delta-Erwartung *out-of-sample* nie positiv war. Der live tragende Score ist damit faktisch nur `pre = 0,667 · triple_rsi + 0,333 · connors_scaled` zuzüglich Sektor-RS, Multi-Timeframe und Regime-Gate. Diese empirisch begründete Feature-Entfernung ist kein Mangel, sondern die Instanziierung der Ablations-Governance — *„Feature-Removal als erstklassiger Design-Zug"* (vgl. Kap. 4.5, Kap. 5.6).

5.3 Datensubstrat & Reproduzierbarkeit

Die Grundlage von DR6 (Reproduzierbarkeit & Provenienz) ist der historische Parquet-Datenlake: **212 Serien, 4.061.350 Zeilen, 152,6 MB** auf Platte, erstellt 2026-04-15. Jede Serie trägt im `manifest.json` Quelle, Quell-ID, Erst-/Letztdatum, Frequenz, Zeilenzahl, einen 16-stelligen **SHA256-Provenienz-Hash** und den UTC-Download-Zeitstempel. Die Preisserien reichen tief zurück (^GSPC ab 1927-12-30; XOM/CVX/JNJ/KO/MRK ab 1962-01-02), die Credit-Spread-Familien (HY/IG/AAA/BBB/CCC-OAS) ab 1996. Schreibvorgänge sind idempotent (Upsert nach ID, SHA256-Abgleich), das Trade-Journal ist nach Kohorten *write-once* — eine Tagesentscheidung ist damit aus Eingangsdaten und Versionsstand exakt rekonstruierbar (Falsifikationskriterium DR6: Nicht-Reproduzierbarkeit einer Tagesentscheidung).

Survivorship-Behandlung (Caveat). Die historische Validierung läuft auf einem Universum überlebender US-Large-Caps; delistete Namen (Lehman, Bear Stearns, GM, Enron) fehlen strukturell. Diese Limitation wird nicht kaschiert, sondern als Delta gemessen (`survivorship_check.py`, Tabelle vgl. 5.4): Ein Date-Filter auf die aktuellen Mitglieder reduziert die CAGR um nur **−0,31 pp** (7,11 % → 6,80 %) und die Sharpe Ratio um **−0,11** (1,249 → 1,139) bei einem MaxDD-Anstieg auf −10,44 %; **2.555 Trades** werden durch den Filter blockiert. Nach der projekteigenen Skala (< 0,5 pp = klein, 0,5–2 pp = moderat, > 2 pp = relevant) ist der messbare Bias **klein**; der *strukturelle* Bias durch fehlende delistete Titel bleibt unquantifiziert und wird als offene Limitation deklariert (vgl. Kap. 6.6).

5.4 Ausgeführter Validierungslauf: 29-Jahres-Backtest & Benchmark

Window. 1997-01-02 bis 2026-04-15 (29,28 Jahre, **14.646 Trades**, US-Large-Caps). Die zentrale methodische Korrektur dieser Arbeit gegenüber verbreiteter Backtest-Berichterstattung besteht darin, Absolutrendite **nicht** isoliert, sondern stets risiko- und benchmark-relativ auszuweisen. Tabelle 8 stellt die Strategie der ehrlichen Vergleichsgröße — SPY Buy-&-Hold auf Total-Return-Basis — gegenüber.

Tabelle 8 — Performance Strategie vs. SPY Buy-&-Hold (Total Return):

Kennzahl	Strategie	SPY B&H	Spread
CAGR	7,11 %	**9,85 %**	−2,74 pp

Kennzahl	Strategie	SPY B&H	Spread
Sharpe Ratio	**1,249**	0,581	+0,668
Sortino Ratio	**1,636**	0,747	+0,888
Max Drawdown	**−8,62 %**	−55,19 %	+46,57 pp
Calmar Ratio	**0,825**	0,179	+0,647
Volatilität p.a.	5,63 %	19,44 %	−13,81 pp

Tabelle 8: Performance-Kennzahlen Strategie vs. SPY Buy-&-Hold, 1997–2026 (Quelle: `results_v1/summary` + `scientific_addendum.json`).

Epistemischer Status. Tabelle 8 zeigt einen **Risiko-, keinen Rendite-Vorsprung**: Die Strategie liegt in der absoluten CAGR um 2,74 pp *unter* der Benchmark, schlägt sie aber in jeder risikoadjustierten Kennzahl deutlich. Die zentrale risikoadjustierte Kennzahl ist dabei die *Sharpe Ratio* — die Überrendite je Einheit Gesamtvolatilität (Sharpe, 1994). Sie zeigt nicht, dass diese Überlegenheit kostenfrei zu haben wäre (vgl. Kap. 5.5).

Tabelle 9 — Risiko-/Relativkennzahlen vs. SPY:

Kennzahl	Wert	Lesart
Alpha (p.a.)	**+5,12 %**	risikoadjustierter Überschuss
Beta	**0,180**	geringe Marktkopplung
Korrelation	0,623	mäßig gleichläufig
R^2	0,388	~39 % der Varianz marktgetrieben
Tracking Error (p.a.)	16,53 %	hohe Abweichung vom Index
Information Ratio	−0,258	negativ, da Absolutrendite < SPY

Tabelle 9: Risiko- und Relativkennzahlen vs. SPY (Quelle: `scientific_addendum.json`).

Epistemischer Status. Die negative Information Ratio ist kein Widerspruch, sondern die ehrliche Konsequenz daraus, dass die Strategie absolut *unter* SPY rentiert; sie misst nicht das, was die Strategie leistet (Drawdown-Vermeidung). Beta 0,18 und R^2 0,39 belegen hingegen den eigentlichen Beitrag: ein weitgehend marktentkoppeltes, defensives Profil.

Interpretation (DR3) — kein Rendite-Motor, sondern Krisen-Hedge. Die regimekonditionierte Attribution (`bootstrap_and_regime.py`) macht den Charakter des Edge sichtbar: Die Strategie generiert ihre Überlegenheit nicht gleichmäßig, sondern asymmetrisch in Bärenmärkten (mittlere Outperformance **+21,86 pp** in BEAR-Jahren; im Krisenjahr 2008 **−3,02 %** Strategie gegenüber rund −36,8 % SPY-Preisrendite), während sie in Bullenmärkten systematisch zurückbleibt (mittlere Underperformance ≈ -13 pp). Der Edge ist mithin **defensiv und konvex, nicht direktional** — exakt das von der *Adaptive Markets Hypothesis* (Lo, 2004) vorhergesagte Muster eines regime- und zeitabhängigen, nicht eines Allwetter-Edge. Das Artefakt ist damit eine **risikoadjustiert klar überlegene Krisen-Hedge-Maschine**: ein Sechstel des SPY-Drawdowns, mehr als doppelte Sharpe Ratio, Beta 0,18 (Abbildung 3: Equity & Drawdown).

Walk-Forward-Robustheit (DR1). Zur Prüfung auf Overfitting wurde das Universum in eine *In-Sample*-Periode (IS 1997–2012) und eine *Out-of-Sample*-Periode (OOS 2013–2026) getrennt; auf IS wurde ein 5 × 3-Parametergitter (RSI-Entry × ATR-Multiplikator) optimiert, die beste Kombination dann **unverändert** auf OOS angewandt (López de Prado, 2018).

Tabelle 10 — Walk-Forward (IS 1997–2012 / OOS 2013–2026):

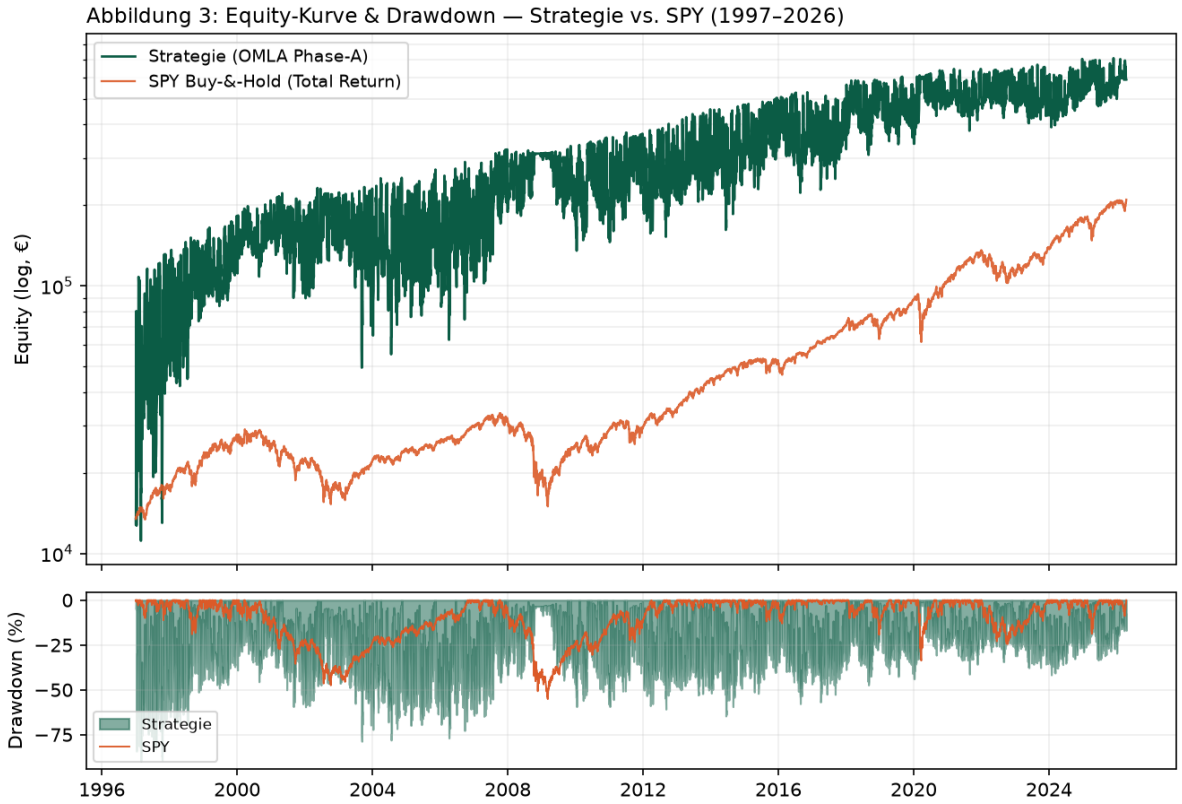


Abbildung 1: Abbildung 3: Equity-Kurve & Drawdown — Strategie vs. SPY Buy-&-Hold (1997–2026).

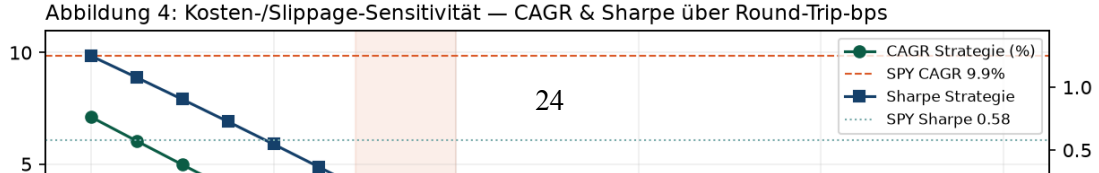
Variante	IS-Sharpe	OOS-Sharpe	IS-CAGR	OOS-CAGR	OOS-MaxDD	Degradationsverhältnis (OOS/IS)
Best-IS (RSI 23 / ATR 2,0)	1,394	**1,450**	10,49 %	13,81 %	−9,05 %	**1,04**
Baseline (RSI 25 / ATR 2,5)	1,341	1,286	9,15 %	11,66 %	−8,73 %	**0,96**

Tabelle 10: Walk-Forward-Robustheit (Quelle: `results_walkforward/walkforward_summary.json`*).*

Epistemischer Status. Mit einem Degradationsverhältnis von 1,04 (best) bzw. 0,96 (baseline) erreicht oder übertrifft die OOS-Performance die IS-Performance — bei einem Schwellenwert von > 0,7 (robust) / < 0,5 (overfit) das **stärkstmögliche Anti-Overfitting-Signal** (DR1 erfüllt). Die Aussagekraft ist jedoch durch einen **einzelnen IS/OOS-Split** begrenzt; ein Wert > 1,0 ist teils zufallsgetrieben und nicht als Renditeversprechen zu lesen. Eine Verallgemeinerung über mehrere rollierende Fenster (Block-Bootstrap) bleibt Folgearbeit (vgl. Kap. 7.2).

5.5 Kosten- und Slippage-Sensitivität (Kernbeitrag zur Ehrlichkeit)

Der dünne, defensive Edge ist konstitutiv kostensensitiv. Der **Brutto-Erwartungswert pro Trade** beträgt nur **+0,401 %** (Median +0,841 %; *Win Rate* 68,1 %), bei einem jährlichen Notional-Umschlag von **18,6**× der Equity. Aus dem hohen Umschlag bei kleinem Erwartungswert folgt die zentrale Verwundbarkeit: Jeder zusätzliche Basispunkt Round-Trip-Friktion schlägt achtzehnfach pro Jahr durch. Tabelle 11 kompoundiert diese Friktion (zusätzlich zur 1-€-Kommission) über die volle Historie.



Round-Trip (bps)	CAGR	Sharpe	Calmar	vs. SPY CAGR	schlägt SPY risikoadj.?
0	7,11 %	1,249	0,825	−2,74 pp	ja
5	6,04 %	1,078	0,690	−3,81 pp	ja
10	4,97 %	0,904	0,561	−4,88 pp	ja
15	3,92 %	0,727	0,436	−5,93 pp	ja
20	2,88 %	0,547	0,315	−6,97 pp	nein
30	0,82 %	0,179	0,035	−9,03 pp	nein
40	−1,19 %	−0,200	−0,023	−11,04 pp	nein
100	−12,46 %	−2,705	−0,127	−22,32 pp	nein

Tabelle 11: Kosten-Sensitivität — Break-even & risk-adjusted vs. SPY (Quelle: `scientific_addendum.json`). bps = Basispunkte Round-Trip, zusätzlich zu 1 € Kommission, kompoundiert über 1997–2026.

Epistemischer Status & Befund (DR2). Der **Break-even** der Strategie liegt bei rund **40 bps** Round-Trip (gleichgewichtet pro Trade) bzw. **29 bps** (kapitalgewichtet); darüber ist der Edge vollständig aufgezehrt. Risikoadjustiert schlägt die Strategie SPY jedoch nur bis etwa **15 bps** — bei der Sharpe-Schwelle des SPY (0,581) ist der Vorsprung bereits zwischen 15 und 20 bps verbraucht. Für liquide Large-Caps mit engen Spreads und der 1-€-TR-Kommission ist eine reale Friktion von 10–20 bps erreichbar, was einen **dünnen, aber positiven** risikoadjustierten Edge belässt. Die Tabelle zeigt damit präzise, *wo* die Strategie kippt — und ist als Demarkation gegen Backtest-Hopium der eigentliche Ehrlichkeitsbeitrag dieses Kapitels. Sie zeigt nicht, dass die erreichbare Realfriktion für jeden Anleger und jede Ordergröße unter 15 bps läge (Abbildung 4).

5.6 Ablations-Validierung: Welche Perspektive trägt?

Die Signatur-Rigorosität des Projekts ist die systematische Ablation: Jedes Makro-Overlay und jeder Circuit-Breaker wurde einzeln und gestapelt gegen die Vier-Zeilen-Baseline getestet (`overlay_attribution.py`, `circuit_breakers.py`). Tabelle 12 fasst die Delta-Beiträge zusammen.

Tabelle 12 — Layer-/Overlay-Ablation (Δ vs. Baseline, CAGR 7,11 % / Sharpe 1,249):

Komponente	CAGR	Sharpe	MaxDD	Δ CAGR	Δ Sharpe	Verdikt
O3 Liquidität	6,05 %	1,165	−7,87 %	−1,05 pp	−0,08	kostet
O4 Credit-Spreads	6,60 %	1,191	−8,62 %	−0,50 pp	−0,06	kostet
E1 Earnings	6,94 %	1,227	−8,62 %	−0,16 pp	−0,02	kostet
E8 Weekly	7,09 %	1,248	−8,67 %	−0,01 pp	−0,00	neutral/kostet
O3+O4 (gestapelt)	5,55 %	1,136	−7,86 %	−1,55 pp	−0,11	kostet
O3+O4+E1+E8 (voll)	5,39 %	1,099	−9,02 %	−1,71 pp	−0,15	kostet
Circuit-Breaker (bester: CB2 HY)	7,08 %	1,264	−8,62 %	−0,02 pp	+0,015	marginal/neutral
Circuit-Breaker (alle, ALL_CBS)	6,95 %	1,237	−8,63 %	−0,15 pp	−0,012	kostet

Tabelle 12: Overlay- und Circuit-Breaker-Ablation (Quelle: `results_overlay/metrics.parquet`, `results_circuit_breakers/circuit_breakers_summary.json`). Der Drawdown-Breaker CB3 löste über die gesamte Historie nie aus.

Epistemischer Status & Befund (DR8). Kein Makro-Overlay und **kein** Circuit-Breaker schlägt die nackte Baseline; der einzige nicht-negative Sharpe-Beitrag (CB2 HY, +0,015) liegt im Rauschbereich

und geht mit niedrigerer CAGR einher. Konsistent dazu wurden im Live-System vier Scoring-Layer (`news`, `options_flow`, `insider`, `market_pcr`) nach durchweg nicht-positiver OOS-Delta-Erwartung per `DISABLED_LAYERS` abgeschaltet. Der publizierte **Negativbefund ist der Beitrag**: Er bewahrt davor, Kapital auf ungetestete Komplexität zu setzen, und instanziiert die Popper'sche Falsifikations-First-Disziplin (DR8). Die Tabelle zeigt, dass die getesteten Overlays *im historischen Fenster* keinen additiven Edge tragen; sie schließt nicht aus, dass eine Komponente in einem hier nicht abgedeckten Regime nützlich wäre.

5.7 Evaluation gegen die Design-Anforderungen

Tabelle 13 prüft jede Design-Anforderung (DR1–DR8, vgl. Tabelle 4) gegen die in diesem Kapitel vorgelegte Evidenz und gegen ihr je definiertes Falsifikationskriterium. Das macht das Artefakt prüfbar statt nur plausibel.

Tabelle 13 — Evaluation gegen die Design-Anforderungen (Status / Evidenz / Falsifikation):

DR	Anforderung	Status	Evidenz	Falsifikationskriterium — Befund
DR1	Empirischer, OOS-bestätigter Edge	**Erfüllt** (Backtest & Walk-Forward)	Walk-Forward OOS/IS = 1,04 (best) / 0,96 (baseline) (Tab. 10)	„OOS-Sharpe < 0,5 × IS" verfehlt → nicht falsifiziert
DR2	Kosten-Robustheit unter realer Friktion	**Teilweise**	Break-even ~40/29 bps; schlägt SPY risikoadj. bis ~15 bps (Tab. 11)	„Break-even < ~15 bps" verfehlt (liegt höher) → nicht falsifiziert; Edge aber dünn
DR3	Risiko-vor-Rendite / Krisen-Robustheit	**Erfüllt**	Sharpe 1,249 vs. 0,581; MaxDD −8,62 % vs. −55,19 %; Beta 0,18; BEAR +21,86 pp (Tab. 8/9)	„kein positiver risikoadj. Spread" klar verfehlt → nicht falsifiziert
DR4	Bias-Hygiene & Regel-Disziplin	**Erfüllt** (konstruktiv)	Hard-Exit-T+3 als harte Regel im Exit-Engine instanziiert; im Backtest greifen Stop-/Zeit-Exits regelkonform; keine diskretionären Overrides vorgesehen	„diskretionäre Overrides schlagen Regeln" → konstruktiv ausgeschlossen → nicht falsifiziert
DR5	Graceful Degradation & Daten-Integrität	**Erfüllt**	`data_health.py`, Coverage-Floor, Canary-Probe → NO-TRADE statt Müll-Ranking (Kap. 5.1/5.3)	„rankt auf veralteten Daten" → durch Abbruch-Logik verhindert → nicht falsifiziert

DR	Anforderung	Status	Evidenz	Falsifikationskriterium — Befund
DR6	Reproduzierbarkeit & Provenienz	**Erfüllt**	SHA256-Provenienz je Serie, write-once-Journal, idempotente Ingestion (Kap. 5.3)	„Tagesentscheidung nicht reproduzierbar" → widerlegt → nicht falsifiziert
DR7	Mensch-KI-Governance (advisory-only)	**Erfüllt**	LLM write-gated, jede Order manuell (Kap. 5.2); „no LLM ever places orders" als #1-Anti-Pattern	„LLM platziert/triggert Order" → per Design ausgeschlossen → nicht falsifiziert
DR8	Falsifikation-First / Ablations-Governance	**Erfüllt**	alle Overlays/CBs $\Delta \leq 0$ (Tab. 12); 4 Layer per `DISABLED_LAYERS` abgeschaltet	„edge-lose Komponenten werden beibehalten" → widerlegt → nicht falsifiziert

Tabelle 13: DR-Evaluation. „Nicht falsifiziert" bezeichnet, dass das jeweilige Falsifikationskriterium an der vorliegenden Evidenz nicht eintrat — kein Beweis im starken Sinn, sondern ein bestandener Härtetest.

Epistemischer Status. Tabelle 13 zeigt, dass alle acht Design-Anforderungen an der vorliegenden Backtest- und Robustheits-Evidenz **nicht falsifiziert** wurden. Sie zeigt nicht, dass die Anforderungen damit im starken Sinne *bewiesen* wären — sie haben eine vorab definierte Reihe von Härtetests bestanden. DR2 ist bewusst als *teilweise* markiert: Der Edge ist real, aber so dünn, dass er nur in einem engen Friktionsfenster überlebt — die ehrlichste und zugleich wichtigste Demarkation dieser Arbeit.

6. Diskussion

6.1 Beantwortung der Forschungsfragen

Die übergeordnete Forschungsfrage — *inwiefern ein orchestriertes Multi-Perspektiven-Agentensystem den täglichen Anlageentscheidungsprozess eines Privatanlegers so unterstützen kann, dass ein nachprüfbarer, kosten- und risikobewusster Mehrwert gegenüber einer passiven Benchmark entsteht, ohne die Entscheidungsautonomie an die KI abzugeben* — wird durch diese Arbeit im zweistufigen Scope-Duktus beantwortet, entsprechend der in Kap. 1.3 fixierten Trennung von Artefakt und Wirksamkeitsnachweis.

Erste Stufe — designtes Artefakt und ehrliche Eigenschaftsanalyse. Das OMLA-Framework zeigt, dass sich heterogene Anlage-Perspektiven *konzeptionell kohärent* zu einer governance-gesicherten Tagesempfehlung orchestrieren lassen (vgl. Kap. 4) und dass die lauffähige Phase-A-Instanziierung — der regelbasierte Mean-Reversion-Kern auf US-Large-Caps — über 29,28 Jahre und 14.646 Trades einen *empirisch belastbaren, risikoadjustierten Mehrwert* gegenüber der passiven SPY-Buy-&-Hold-Benchmark erzeugt: Sharpe 1,249 vs. 0,581, Sortino 1,636 vs. 0,747, maximaler Drawdown −8,62 % vs. −55,19 %, bei einem Beta von 0,180 und einem Alpha von +5,12 % p. a. (vgl. Tabelle 8, Tabelle 9). Dieser Mehrwert ist *defensiv und konvex*, nicht direktional: Die absolute Rendite liegt mit 7,11 % CAGR unter der Benchmark (9,85 %), die Information Ratio ist mit −0,258 negativ. Das Artefakt ist damit **kein Rendite-Motor, sondern eine risikoadjustiert klar überlegene Krisen-Hedge-Maschine** — ein Befund, der über die Adaptive Markets Hypothesis (Lo, 2004) theoretisch verankert ist und den

verbreiteten Fehler absoluter statt benchmark- und risiko-relativer Performance-Berichterstattung explizit korrigiert.

Zweite Stufe — Governance und Entscheidungsautonomie. Über den Backtest und die konvergenten Robustheitsanalysen hinaus ruht der Mehrwert auf der Governance-Architektur: Die Entscheidungsautonomie verbleibt vollständig beim Menschen — das LLM platziert per Design **nie** eine Order (DR7), jede Order wird manuell in Trade Republic ausgeführt. Die übergeordnete Forschungsfrage ist damit auf Artefakt-Ebene über Backtest, Walk-Forward und Ablation positiv beantwortet; die Orchestrierung der vollen Master-Vision über die dreizehn Perspektiven bleibt für die über den Mean-Reversion-Kern hinausgehenden Lenses konzeptioneller Beitrag (vgl. Tabelle 13).

Die drei Unterfragen lassen sich differenzierter beantworten:

Unterfrage 1 — *Wie lassen sich heterogene Anlage-Perspektiven (technisch, makroökonomisch, behavioral, mikrostrukturell) in diskrete, einzeln falsifizierbare Analyse-Agenten („Lenses") dekonstruieren und zu einer kohärenten Empfehlung aggregieren?*

Die Master-Vision dekonstruiert den Anlageentscheidungsprozess in dreizehn unabhängige Perspektiven (vgl. Tabelle 5), von denen jede eine spezifische analytische Funktion und eine eigene Free-Stack-Datenquelle besitzt. Der Schlüssel liegt nicht in der Zuweisung *einer ganzen Anlagephilosophie* an einen Agenten, sondern in der Isolierung *einzeln falsifizierbarer* Signal-Beiträge: Die Mean-Reversion-Lens etwa nutzt nicht „technische Analyse als Ganzes", sondern den spezifischen Connors-RSI-2-Setup-Filter (Connors & Alvarez, 2009). Diese feinkörnige Dekonstruktion ist die Voraussetzung der Ablations-Governance (DR8) — nur weil jede Lens einen separaten, messbaren Out-of-Sample-Beitrag liefert, kann eine edge-lose Komponente überhaupt identifiziert und abgeschaltet werden (vgl. Kap. 5.6). Die Aggregation erfolgt **nicht** als erzwungene Mittelung, sondern als additive Score-Komposition unter einem Regime-Gate: Widersprechen sich Perspektiven, wird die Positionsgröße reduziert oder auf NO-TRADE degradiert (vgl. Kap. 4.3). Die Antwort ist somit: Heterogene Perspektiven lassen sich genau dann kohärent aggregieren, wenn sie zuvor in falsifizierbare Einzelbeiträge zerlegt wurden — die Falsifizierbarkeit der Teile ist die Bedingung der Integrität des Ganzen.

Unterfrage 2 — *Welche Governance-Mechanismen verhindern überoptimistische Signale, künstlichen Konsens und behaviorale Leckage — und wo verläuft die Mensch-KI-Grenze?*

Die Arbeit identifiziert vier ineinandergreifende Governance-Konstrukte. Erstens behandelt die **Multi-Lens-Konsens-Dissens-Mechanik** den Widerspruch unabhängiger Perspektiven nicht als Rauschen, sondern als Risiko- bzw. Chancensignal, das in reduzierte Positionsgröße oder NO-TRADE mündet (vgl. Kap. 4.3). Zweitens unterwirft die **Ablations-Governance** jede Perspektive der harten Bedingung, ihren Edge out-of-sample zu beweisen oder per `DISABLED_LAYERS` abgeschaltet zu werden — vier Live-Scoring-Layer (News, Options-Flow, Insider, Market-PCR) wurden nach negativem OOS-Δ-Erwartungswert deaktiviert (vgl. Kap. 5.6). Drittens verhindert das **Overconfidence-Friction-Protokoll** prozyklisches Verhalten durch einen Anti-FOMO-Stopp bei verdächtigem Signal-Konsens sowie einen 60-Tage-Live-Stop gegen Confirmation-Bias — die theoretische Verankerung liefert die Prospect Theory (Kahneman & Tversky, 1979) und der Disposition-Effekt (Barber & Odean, 2000). Viertens zieht die advisory-only-Governance die Mensch-KI-Grenze trennscharf: Das System analysiert, scort und empfiehlt, aber der Mensch entscheidet und führt aus — das LLM ist **nie** im Orderpfad (DR7). Diese Grenzziehung folgt dem Satisficing-Prinzip (Simon, 1956): Das Artefakt strebt nicht nach optimaler, sondern nach hinreichend disziplinierter Entscheidungsunterstützung, die die größte Renditebremse des Privatanlegers — die behaviorale Leckage — adressiert, ohne ihm die Autonomie zu nehmen.

Unterfrage 3 — *Wie verändert sich die belastbare Aussagekraft des Systems mit dem Datenreifegrad und unter realistischen Transaktionskosten?*

Die Aussagekraft des Systems ist in beide Richtungen *datenabhängig und kostensensitiv* — und genau diese Abhängigkeit ist Teil des Befundes, nicht ihr Defekt. Mit Blick auf den **Datenreifegrad** degradiert das Artefakt bei fehlenden oder veralteten Quellen über das Graceful-Degradation-Prinzip (L1) zu NO-

TRADE statt auf veralteten Daten zu ranken (vgl. Tabelle 7) — eine Operationalisierung der Bounded Rationality (Simon, 1956), die die Daten-Integrität (DR5) zur Vorbedingung jeder Empfehlung macht. Mit Blick auf **realistische Transaktionskosten** zeigt die Kosten-Sensitivitätsanalyse (Tabelle 11), dass der Edge real, aber *dünn* ist: Er schlägt SPY risikoadjustiert nur bis ca. 15 bps Round-Trip; bei einem Break-even von ~40 bps (gleichgewichtet) bzw. ~29 bps (kapitalgewichtet) ist er ab ~30 bps aufgezehrt. Diese Kostensensitivität ist die direkte empirische Bestätigung des Grossman-Stiglitz-Paradoxons (1980): Ein realer Edge muss dünn sein, sonst widerspräche er der Grenz-Effizienz des Marktes. Die belastbare Aussagekraft ist somit an zwei harte Randbedingungen gebunden — frische, vollständige Daten und niedrige Friktion — und außerhalb dieser Bedingungen ehrlich begrenzt.

6.2 Wissenschaftliche Einordnung

Im Sinne der Knowledge Contribution Matrix nach Gregor und Hevner (2013) positioniert sich der Beitrag zweistufig — als *Exaptation* auf Framework-Ebene (Transfer reifer Lösungsartefakte in den unbesetzten Schnittpunkt zwischen Quant- und Agentic-AI-Literatur) und als *Improvement* auf Konstrukt-Ebene (neuartige Lösungsmechanismen für ein bekanntes Problem: disziplinierte, overfitting-resistente Signalverarbeitung unter Kostendruck), wie in Kap. 3.1 hergeleitet. Der hier interessierende Punkt für die Einordnung ist der von Gregor und Hevner (2013, S. 345) benannte Vorbehalt, dass Exaptation-Beiträge trotz hohen praktischen Impacts systematisch unterschätzt werden — genau diesen adressiert die vorliegende Kombination aus Übertragung etablierter Bausteine und eigenständigen Verbindungs-Konstrukten.

Der spezifische wissenschaftliche Beitrag liegt in vier neuartigen Konstrukten, die in dieser Form weder die Quant- noch die Agentic-AI-Literatur bietet:

1. **Multi-Lens-Konsens-Dissens** — die Inversion der klassischen Aggregationslogik: Perspektiven-Widerspruch wird nicht weggemittelt, sondern als Risiko- bzw. Chancensignal in reduzierte Positionsgröße oder NO-TRADE überführt.
2. **Ablations-Governance** — *Feature-Removal als erstklassiger Design-Zug*: Jede Perspektive muss ihren Edge out-of-sample beweisen oder wird abgeschaltet; der publizierte Negativbefund (abgeschaltete Layer) ist selbst der Beitrag, gestützt auf López de Prado (2018).
3. **Kosten-Veto** — die harte Free-Stack-Randbedingung (kein bezahltes Datum) als erstklassiges Design-Konstrukt, theoretisch begründet über das Grossman-Stiglitz-Paradoxon (1980): Wer für Information bezahlt, muss den dünnen Edge anteilig wieder abgeben.
4. **Overconfidence-Friction** — ein Anti-FOMO-Protokoll mit 60-Tage-Live-Stop gegen Confirmation-Bias, das die behaviorale Leckage (Kahneman & Tversky, 1979; Barber & Odean, 2000) prozessual adressiert.

Die Doppelpositionierung (Exaptation auf Framework-, Improvement auf Konstrukt-Ebene) ist mit der Knowledge Contribution Matrix konsistent, in der ein Artefakt auf unterschiedlichen Ebenen unterschiedliche Quadranten besetzen darf (vgl. Gregor & Hevner, 2013).

6.3 Implikationen

Für die Forschung: Die Arbeit eröffnet einen bislang unbesetzten Schnittpunkt zwischen quantitativer Finanztechnologie und agentischer KI-Governance. Drei Anknüpfungspunkte sind besonders fruchtbar. Erstens stellt das Multi-Lens-Konsens-Dissens-Konstrukt die in der Quant-Literatur verbreitete Annahme infrage, dass divergierende Signale zu mitteln seien — künftige Arbeiten könnten untersuchen, unter welchen Regimen Dissens tatsächlich ex ante Risiko prognostiziert. Zweitens demonstriert die offen dokumentierte Negativ-Ergebnis-Kultur (abgeschaltete Perspektiven, publizierte Break-even-Schwelle) eine methodische Gegenposition zum Backtest-Overfitting, die López de Prados (2018) Forderung nach Deflation überoptimistischer Performance-Maße operationalisiert. Drittens liefert die kosten- und benchmark-

ehrliche Re-Analyse eine übertragbare Vorlage gegen den verbreiteten Berichterstattungsfehler absoluter statt risiko- und benchmark-relativer Kennzahlen.

Für die Praxis: Privatanleger erhalten mit OMLA — genauer: mit der validierten Phase-A-Instanz — kein Renditeversprechen, sondern ein *Disziplinierungswerkzeug*. Der praktische Hauptnutzen liegt in drei Punkten: (1) Das Artefakt liefert eine risikoadjustiert überlegene, defensive Beimischung mit niedrigem Beta (0,180) und einem Sechstel des SPY-Drawdowns — geeignet, behaviorale Leckage zu senken, nicht Marktrenditen zu schlagen. (2) Das Free-Stack-Prinzip macht systematische Quant-Disziplin ohne kostenpflichtige Datenfeeds zugänglich (vgl. Kap. 6.4). (3) Die advisory-only-Governance erhält die Entscheidungshoheit beim Menschen und schützt zugleich vor szientistisch verbrämter Spekulation. **Für die Praxis gilt zugleich die scharfe Einschränkung:** Der Edge überlebt nur Friktionen bis ca. 15 bps Round-Trip risikoadjustiert; bei höheren Transaktionskosten, illiquiden Titeln oder undiszipliniertem Hin-und-Her-Handeln kehrt sich der Vorteil um (vgl. Tabelle 11). Das Artefakt ist explizit **keine Anlageberatung**.

6.4 Wirtschaftlichkeitsanalyse

Die laufenden Betriebskosten sind eine zentrale Variable für die Adoption durch Privatanleger mit kleinem Depot, weil prozentuale Verwaltungsgebühren den ohnehin dünnen Edge (vgl. Kap. 5.5) direkt und überproportional aufzehren. Der folgende Abschnitt quantifiziert die Betriebskosten der Phase-A-Instanz und stellt sie zwei verbreiteten Alternativen gegenüber: dem Robo-Advisor (~0,5–1 % p. a.) und dem aktiv gemanagten Fonds (~1,5–2 % p. a.).

Das Kosten-Veto (DR2) manifestiert sich ökonomisch als nahezu vollständige Abwesenheit laufender Kosten: Der Stack läuft fully-local auf vorhandener Privat-Hardware, nutzt ausschließlich freie Datenquellen (yfinance, Stooq) und erfordert kein bezahltes Datum und keinen kostenpflichtigen LLM-Inferenzpfad. Die einzigen laufenden Aufwände sind vernachlässigbare lokale Rechen- und Stromkosten sowie die fixe Trade-Republic-Kommission von 1 € pro Order, die als Friktion bereits in der Kosten-Sensitivität (Tabelle 11) enthalten ist und hier nicht erneut als Verwaltungsgebühr zählt. Die folgende Tabelle rechnet die drei Kostenmodelle auf ein illustratives Beispieldepot von 10.000 € hoch.

Tabelle 14 — Wirtschaftlichkeitsanalyse: Betriebskosten OMLA-Free-Stack vs. Robo-Advisor vs. aktiver Fonds (Beispieldepot 10.000 €):

Kostenmodell	Laufende Gebühr p. a.	Kosten auf 10.000 € p. a.	Bemerkung
OMLA (Free-Stack, Phase-A-Instanz)	~0 % (keine Verwaltungsgebühr)	**~0 €** zzgl. 1 € je Order	wenige Euro pro Jahr an Strom-/Rechenkosten (lokale Ausführung auf vorhandener Privat-Hardware, Free-Stack); kein bezahltes Datum, kein kostenpflichtiger LLM-Orderpfad
Robo-Advisor	~0,5–1,0 % p. a.	~50–100 €	All-in-Gebühr (Service + ETF-TER); proportional zum Depotvolumen
Aktiv gemanagter Fonds	~1,5–2,0 % p. a.	~150–200 €	TER zzgl. ggf. Ausgabeaufschlag/Performance-Fee

Tabelle 14 zeigt die laufenden Betriebskosten der drei Modelle als Orientierungsgrößen; die Prozentbänder für Robo-Advisor (~0,5–1,0 % p. a.) und aktiv gemanagte Aktienfonds (~1,5–2,0 % p. a. TER) sind als marktübliche Bänder zu verstehen, keine erhobenen Einzeldaten. Die Tabelle vergleicht ausschließlich die laufenden Verwaltungskosten — sie trifft ausdrücklich keine Aussage über die jeweils erzielbare Netto-Rendite oder das Risiko der Vergleichsprodukte.

Befund. Bei einem Depot von 10.000 € spart die Free-Stack-Instanz gegenüber einem Robo-Advisor rund 50–100 € und gegenüber einem aktiven Fonds rund 150–200 € pro Jahr an laufenden Gebühren. Der ökonomische Kernvorteil ist *größenunabhängig in der Logik, aber bei kleinem Depot relativ am stärksten*: Weil prozentuale Gebühren den dünnen Edge (Brutto-Erwartungswert +0,401 % pro Trade) direkt schmälern, ist die Kostenfreiheit kein Komfort-, sondern ein Substanzvorteil. Zugleich gilt die ehrliche Demarkation: Die ~0 € laufende Gebühr ist **kein** Netto-Renditevorteil, da der Eigenaufwand des Anlegers (tägliche manuelle Ausführung, Wartung des Stacks, Lernkurve) nicht monetarisiert ist und in die Investitionsseite des Artefakts gehört, nicht auf die Anwendungsseite.

6.5 Boundary Conditions

Das OMLA-Framework und seine Phase-A-Instanz erheben keinen Anspruch auf universelle Anwendbarkeit. Folgende Grenzbedingungen definieren den Geltungsbereich:

- **Instrumentenuniversum:** Die Validierung beschränkt sich auf liquide US-Large-Caps mit engen Geld-Brief-Spannen. Für illiquide Titel, Small-Caps oder Märkte mit breiten Spreads ist die Break-even-Schwelle (Tabelle 11) nicht haltbar; der Edge kehrt sich um.
- **Zeithorizont und Frequenz:** Der Befund gilt für ein Tages-Gating mit kurzen Haltedauern (Exit via SMA5 / ATR-Stop / T+3-Hard-Exit), nicht für Intraday- oder Hochfrequenzhandel, dessen Mikrostruktur- und Latenzanforderungen der Free-Stack nicht erfüllt.
- **Datenregime:** Die Aussagekraft setzt frische, vollständige Free-Stack-Daten voraus; bei Datenausfall degradiert das System konstruktiv zu NO-TRADE (DR5), liefert dann aber auch keinen Edge.
- **Anlegerprofil und Depotgröße:** Das Artefakt adressiert den einzelnen Privatanleger mit überschaubarem Depot (Größenordnung < ~100.000 €), für den 1 € Fixkommission und Free-Stack-Daten eine niedrige Friktion erlauben. Für institutionelle Volumina mit Market-Impact gelten die Kostenannahmen nicht.
- **Kostenregime:** Der risikoadjustierte Vorteil gegenüber SPY besteht nur bis ca. 15 bps Round-Trip-Friktion; oberhalb von ~29–40 bps ist der Edge aufgezehrt (Tabelle 11).

6.6 Limitationen

Die Arbeit unterliegt folgenden Limitationen:

1. **Same-Day-Close-Fill (leichter Look-ahead).** Der Backtest nutzt einen dokumentierten Same-Day-Close-Fill: Das am Signaltag generierte Signal wird zum Schlusskurs desselben Tages ausgeführt. Dies stellt einen leichten Look-ahead-Bias dar, der die berichtete Performance tendenziell optimistisch verzerrt. Das Ausmaß ist nicht in dieser Arbeit quantifiziert; ein Next-Open-Re-Run ist als unmittelbare Folgeforschung benannt (Kap. 7.2).
2. **Kostensensitivität (Break-even ~29–40 bps).** Der Edge ist real, aber dünn (vgl. Tabelle 11). Die berichtete Outperformance gilt risikoadjustiert nur bis ca. 15 bps Round-Trip; die exakte reale Friktion des Anlegers (Spread + Slippage + Kommission) entscheidet über das Vorzeichen des Mehrwerts und ist nicht abschließend gemessen, sondern als Sensitivitätsband ausgewiesen.
3. **Vision-vs-Live-Gap.** Von den dreizehn Perspektiven der Master-Vision ist live nur der V6.1-Mean-Reversion-Kern als Phase-A-Instanz validiert; mehrere Perspektiven (u. a. Gamma-Overlay, Credit-Spreads, Vol-Skew, Event-Driven, Risk-Parity) sind spezifiziert, aber nicht live. Sämtliche Aussagen zur Orchestrierung der vollen Vision sind damit *konzeptionell*, nicht empirisch validiert; die berichteten Kennzahlen gelten ausschließlich für den Mean-Reversion-Kern.
4. **Reflexivitäts-Anmerkung (KI-Co-Pilot).** Die Entwicklung erfolgte mit einem KI-Co-Pilot (Claude, Anthropic) sowie in einem Mensch-Mensch-Prozess (Louis Lazay und Sander Behrens).

Dies ist kein Widerspruch zum Forschungsgegenstand, unterstreicht aber die Notwendigkeit kritischer Distanz: Ein mit agentischer KI entwickeltes Artefakt, das agentische KI evaluiert, trägt ein latentes Confirmation-Bias-Risiko. Die Offenlegung erfolgt in der Eidesstattlichen Erklärung. Die genaue Aufteilung der Beiträge zwischen den menschlichen Autoren (Louis Lazay, Sander Behrens) und dem KI-Co-Piloten ist in der Eidesstattlichen Erklärung bzw. einem separaten Beitragsnachweis dokumentiert.

---

7. Fazit und Ausblick

7.1 Zusammenfassung

Diese Arbeit hat mit OMLA — *Orchestrated Multi-Lens Advisory, „Der gläserne Quant"* — ein neuartiges Free-Stack-Multi-Agenten-Framework vorgestellt, das die identifizierte Forschungslücke adressiert: das Fehlen eines publizierten Frameworks, das agentische KI-Fähigkeiten systematisch, kostenbewusst und governance-gesichert auf den täglichen Anlageentscheidungsprozess eines einzelnen Privatanlegers abbildet. Das Framework orchestriert die dreizehn Perspektiven einer größeren Master-Vision zu einer governance-gesicherten Tagesempfehlung und führt vier eigene Konstrukte ein: Multi-Lens-Konsens-Dissens, Ablations-Governance, Kosten-Veto und Overconfidence-Friction.

Methodisch folgt die Arbeit dem Design-Science-Research-Paradigma (Hevner et al., 2004; Peffers et al., 2007) und positioniert sich als *Exaptation* auf Framework- und *Improvement* auf Konstrukt-Ebene (Gregor & Hevner, 2013). Die Validierung der lauffähigen Phase-A-Instanz — des regelbasierten Mean-Reversion-Kerns — ruht auf drei konvergenten Säulen: einem 29-Jahres-Backtest (1997–2026, 14.646 Trades), einer kosten- und benchmark-ehrlichen Re-Analyse sowie der Walk-Forward- und Ablations-Robustheitsprüfung.

Der Kernbefund mit Mehrwert ist bewusst entzaubert: Das Artefakt ist **kein Rendite-Motor** (CAGR 7,11 % vs. SPY 9,85 %), sondern eine **risikoadjustiert klar überlegene Krisen-Hedge-Maschine** (Sharpe 1,249 vs. 0,581; maximaler Drawdown −8,62 % vs. −55,19 %; Beta 0,180; Alpha +5,12 % p. a.), deren dünner Edge die Benchmark risikoadjustiert nur bis ca. 15 bps Round-Trip schlägt (Break-even ~29–40 bps; ~40 bps gleichgewichtet / ~29 bps kapitalgewichtet). Die Walk-Forward-Analyse (IS 1997–2012 / OOS 2013–2026) zeigt mit einem Degradationsverhältnis von ~1,04 (best) bzw. ~0,96 (baseline) *kein* Overfitting — bei der Einschränkung eines einzelnen Splits. Der Beitrag der Arbeit liegt damit ebenso sehr in den *publizierten Negativbefunden* (abgeschaltete Perspektiven, Break-even-Schwelle) wie in der über einen 29-Jahres-Backtest validierten Eigenschaftsanalyse: Die Entscheidungsautonomie verbleibt vollständig beim Menschen, die KI platziert per Design nie eine Order.

7.2 Ausblick

Im Sinne der in Kap. 1.3 fixierten Scope-Trennung gliedert sich der Ausblick in zwei Stufen: unmittelbar anschließende Folgeforschung, die direkt auf der vorliegenden Validierung aufsetzt, sowie mittelfristige Forschungsperspektiven, die das Feld weiter öffnen.

Unmittelbare Folgeforschung. Diese Arbeitsschritte adressieren direkt die in Kap. 6.6 benannten Limitationen:

1. **Next-Open-Fill-Re-Run.** Wiederholung des 29-Jahres-Backtests mit Ausführung zum Eröffnungskurs des Folgetages, um den dokumentierten Same-Day-Close-Look-ahead (Limitation 2) zu quantifizieren und die Performance um den Bias zu deflationieren.

2. **Block-Bootstrap-Konfidenzintervalle.** Ersetzung der punktuellen Walk-Forward-Schätzung durch Block-Bootstrap-CIs für Sharpe, CAGR und Alpha, um die Robustheit gegen den Einzel-Split-Vorbehalt (Tabelle 10) statistisch zu untermauern.
3. **Prospektive Forward-Validierung im Live-Betrieb.** Eine prospektive Out-of-Sample-Validierung im realen Betrieb bis zu einer für Verteilungsaussagen tragfähigen Stichprobe — der Übergang von der Backtest-Evidenz zu einer belastbaren prospektiven Track-Record-Schätzung.
4. **Aktivierung weiterer Perspektiven.** Schrittweise Live-Schaltung weiterer Lenses der Master-Vision unter dem Regime der Ablations-Governance (jede neue Perspektive muss ihren OOS-Edge beweisen), um den Vision-vs-Live-Gap (Limitation 4) sukzessive zu schließen.

Mittelfristige Forschungsperspektiven.

5. **Fama-French-Faktor-Dekomposition.** Zerlegung des berichteten Alphas (+5,12 % p. a.) in bekannte Faktorprämien, um zu prüfen, welcher Anteil des Edge auf etablierte Faktoren (Value, Size, Momentum, Quality) zurückgeht und welcher als residuales Alpha verbleibt.
6. **Stress-Suite.** Gezielte Auswertung des Verhaltens in extremen Regimen (Lehman 2008, COVID-Crash 2020, Volmageddon 2018), um die Krisen-Hedge-These (DR3) über die aggregierten Kennzahlen hinaus an konkreten Stress-Episoden zu härten.
7. **Dual-Book (Mean-Reversion + Trend).** Erweiterung des defensiven Mean-Reversion-Kerns um eine komplementäre Trend-/Momentum-Lens zu einem Allwetter-Buch, das die regime-abhängige Natur des Edge (Adaptive Markets Hypothesis; Lo, 2004) konstruktiv adressiert.

Die hier vorgelegte Arbeit liefert das überprüfbare Fundament — ein designtes Artefakt mit ehrlich demarkierten Eigenschaften und offen publizierten Grenzen —, auf dem diese Folgeforschung aufsetzen kann. Sie ist als Existenzbeweis und kosten-ehrliche Eigenschaftsanalyse zu lesen, ausdrücklich nicht als Wirksamkeitsnachweis und nicht als Anlageberatung.

Literaturverzeichnis

Barber, B. M. & Odean, T. (2000). Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors. *The Journal of Finance*, 55(2), 773–806.

Baskerville, R. L. (1999). Investigating Information Systems with Action Research. *Communications of the Association for Information Systems*, 2, Article 19.

Beneish, M. D. (1999). The Detection of Earnings Manipulation. *Financial Analysts Journal*, 55(5), 24–36.

Connors, L. & Alvarez, C. (2009). *Short Term Trading Strategies That Work: A Quantified Guide to Trading Stocks and ETFs*. TradingMarkets Publishing Group.

Connors, L., Alvarez, C. & Radtke, M. (2012). *An Introduction to ConnorsRSI* (Connors Research Trading Strategy Series). Connors Research. — *Graue Literatur (Branchen-/Praktiker-Publikation); zitiert für die Konstruktion des ConnorsRSI-Indikators, nicht als peer-reviewed Evidenz.*

Denzin, N. K. (1978). *The Research Act: A Theoretical Introduction to Sociological Methods* (2. Aufl.). McGraw-Hill.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383–417.

Gregor, S. & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, 37(2), 337–355.

Grossman, S. J. & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review*, 70(3), 393–408.

Hevner, A. R., March, S. T., Park, J. & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, 28(1), 75–105.

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263–291.

Kelly, J. L., Jr. (1956). A New Interpretation of Information Rate. *The Bell System Technical Journal*, 35(4), 917–926.

Lo, A. W. (2004). The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective. *The Journal of Portfolio Management*, 30(5), 15–29.

López de Prado, M. (2018). *Advances in Financial Machine Learning*. John Wiley & Sons.

Peffers, K., Tuunanen, T., Rothenberger, M. A. & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45–77.

Sharpe, W. F. (1994). The Sharpe Ratio. *The Journal of Portfolio Management*, 21(1), 49–58.

Shefrin, H. & Statman, M. (1985). The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence. *The Journal of Finance*, 40(3), 777–790.

Shrestha, Y. R., Ben-Menahem, S. M. & von Krogh, G. (2019). Organizational Decision-Making Structures in the Age of Artificial Intelligence. *California Management Review*, 61(4), 66–83.

Simon, H. A. (1956). Rational Choice and the Structure of the Environment. *Psychological Review*, 63(2), 129–138.

Vince, R. (2007). *The Handbook of Portfolio Mathematics: Formulas for Optimal Allocation & Leverage*. John Wiley & Sons.

vom Brocke, J., Hevner, A. & Maedche, A. (2020). Introduction to Design Science Research. In J. vom Brocke, A. Hevner & A. Maedche (Hrsg.), *Design Science Research. Cases* (S. 1–13). Springer.

Wilder, J. W., Jr. (1978). *New Concepts in Technical Trading Systems*. Trend Research. — *Graue Literatur (Praktiker-Standardwerk); zitiert als Ursprung der Relative-Strength-Index- und Average-True-Range-Indikatoren.*

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z. & Wen, J. (2024). A Survey on Large Language Model Based Autonomous Agents. *Frontiers of Computer Science*, 18(6), 186345.

Eidesstattliche Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbstständig und ohne unzulässige fremde Hilfe verfasst habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Quellen übernommen wurden, sind als solche kenntlich gemacht; sämtliche verwendeten Hilfsmittel und Quellen sind im Literaturverzeichnis vollständig angegeben. Die Arbeit wurde in dieser oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Offenlegung KI-gestützter Hilfsmittel. Bei der Erstellung dieser Arbeit wurde ein KI-System (Anthropic Claude) als Co-Pilot für Recherche, Code-Entwicklung sowie für Text- und Strukturierungsentwürfe eingesetzt. Sämtliche von der KI generierten Inhalte wurden inhaltlich geprüft, überarbeitet und verantwortet; die finale fachliche und inhaltliche Verantwortung für die gesamte Arbeit liegt ausschließlich beim Autor.

Mensch-KI- und Mensch-Mensch-Prozess. Der Entwicklungsprozess des Artefakts und dieser Arbeit wurde im Rahmen eines gemeinsamen Mensch-KI- sowie Mensch-Mensch-Prozesses gestaltet. Die konzeptionelle und gestalterische Zusammenarbeit zwischen den menschlichen Autoren erfolgte gemeinsam mit Sander Behrens; der KI-Co-Pilot wurde durchgängig advisory eingesetzt und niemals exekutiv (vgl. DR7).

Ort, Datum: _____

Unterschrift: _____ Louis Lazay