

IIWM-Paper No. 18

Schriftenreihe des IIWM

Institut für Immobilienwirtschaft und -management

April 2026

Niklas Will
Verena Rock

Mensch vs. KI im Immobilien- Investment Management: Entscheidungsprozesse zur Bewertung von Investment- opportunitäten und Konzep- tion eines KI-Entscheidungs- agenten



IIWM Institut für
Immobilienwirtschaft
und -management
TH Aschaffenburg

ISSN (print) 2568-4140

ISSN (online) 2568-5872

www.iiwm.de

Zitierempfehlung: **Will, Niklas; Rock, Verena** (2026): Mensch vs. KI im Immobilien-Investment Management: Entscheidungsprozesse zur Bewertung von Investmentopportunitäten und Konzeption eines KI-Entscheidungsagenten. In: IIWM (Hrsg.): IIWM-Paper No. 18, Aschaffenburg: IIWM Institut für Immobilienwirtschaft und -management, April 2026.

Impressum

ISSN (print) 2568-4140
ISSN (online) 2568-5872

Herausgeber:

IIWM Institut für Immobilienwirtschaft und -management

Technische Hochschule Aschaffenburg

Vertreten durch die Institutsleitung:

Prof. Dr. Verena Rock MRICS

Prof. Dr.-Ing. Lars Bernhard Schöne

Würzburger Straße 45

63743 Aschaffenburg

Tel.: +49(0)6021-4206-745

Web: www.iiwm.de

E-Mail: institutsleitung@iiwm.de

Redaktion:

Niklas Will

c/o IIWM Institut für Immobilienwirtschaft und -management

Technische Hochschule Aschaffenburg

Würzburger Straße 45

63743 Aschaffenburg

Mensch vs. KI im Immobilien-Investment Management: Entscheidungsprozesse zur Bewertung von Investmentopportunitäten und Konzeption eines KI-Entscheidungsagenten

Human vs. AI in Real Estate Investment: Decision Processes in the Evaluation of Investment Opportunities and Design of an AI-Based Decision Agent

Niklas Will

Verena Rock

Zusammenfassung

Das Thema KI-gestützter Entscheidungsunterstützung im Immobilien-Investmentmanagement wird derzeit vor allem im Kontext von Effizienz, Standardisierung und Automatisierung diskutiert. Die vorliegende Arbeit analysiert daher Unterschiede in Argumentationslogik, Gewichtung zentraler Entscheidungsfaktoren und dem Auftreten heuristischer Muster zwischen menschlichen Entscheidungsgruppen und Large Language Models (LLMs).

Die Arbeit untersucht drei Forschungsfragen: (1) Wie unterscheiden sich Human- und KI-Bewertungen von Investmentopportunitäten in Bezug auf Entscheidungsdeterminanten und Heuristiken? (2) Welche Heuristiken zeigen sich bei menschlichen Entscheidergruppen und inwiefern repliziert oder variiert KI diese Muster? (3) Welche Anforderungen lassen sich daraus für einen KI-basierten Entscheidungsagenten zur Prozessunterstützung ableiten?

Zur Einordnung des Forschungsstandes wurde zunächst eine strukturierte Literaturanalyse durchgeführt; die Forschungsfragen wurden anschließend primär empirisch beantwortet. Methodisch basiert die Arbeit auf einer standardisierten Online-Befragung im experimentellen Zwei-Gruppen-Design mit manipulierten Investment-Teasern zur Prüfung von Anchoring, Framing, Representativeness, Herding und Overconfidence. Ergänzend wurden acht LLMs unter identischer Teaser- und Fragebogenlogik getestet; Experteninterviews dienen der Kontextualisierung und der Ableitung praxisnaher Gestaltungsprinzipien.

Die Ergebnisse zeigen in der Humanbefragung heuristikkonsistente Muster: Preisanker verschieben Kaufpreisindikationen (Expert:innen reagieren konservativer als Studierende), Framing beeinflusst Ton- und Risikowahrnehmung, Representativeness prägt Trendannahmen, während Herding insgesamt selten bleibt und Überkonfidenz nur schwach ausgeprägt ist. In den KI-Tests treten modellabhängige Muster auf: Mehrere Modelle sind ankersensitiv, Framing wirkt durchgängig auf Wahrnehmungsebene und teils auch auf die Investitionsentscheidung; Representativeness zeigt ein vergleichsweise robustes Profil, Herding ist nicht zu beobachten, während Überkonfidenz in Form hoher Entscheidungssicherheit bei einzelnen Modellen erkennbar ist.

Für die Praxis wird ein KI-Entscheidungsagent konzipiert, der die Screening-Phase durch strukturierte Aufbereitung von Teaser-Informationen, konsistente Ableitung zentraler Parameter und transparente Begründungslogik unterstützt. Durch gezieltes Prompt Engineering zielt der Agent darauf ab, heuristikbedingte Verzerrungen zu reduzieren und robustere, konsistentere Bewertungen zu fördern. Im Human-

in-the-Loop-Ansatz verbleibt die finale Entscheidung beim Investmentteam; der Agent erhöht Vergleichbarkeit und Qualität der Entscheidungsgrundlagen im Deal-Screening. Insgesamt leistet die Arbeit einen Beitrag zur Einordnung von Human-vs.-KI-Entscheidungslogiken im Immobilien-Investmentmanagement und zeigt, wie empirische Befunde in ein praxistaugliches Agentenkonzept für das Deal-Screening überführt werden können.

Summary

The role of AI-supported decision support in real estate investment management is increasingly discussed in the context of efficiency, standardization, and automation. This study therefore examines differences in reasoning logic, the weighting of key decision factors, and the occurrence of heuristic patterns between human decision-makers and Large Language Models (LLMs).

The analysis addresses three research questions: (1) How do human and AI-based evaluations of real estate investment opportunities differ with respect to decision determinants and heuristics? (2) Which heuristics can be observed among human decision-makers, and to what extent do AI models replicate or vary these patterns? (3) What requirements can be derived for a KI-based decision agent designed to support investment screening processes?

To contextualize the research landscape, a structured literature review was conducted first; the research questions were subsequently addressed primarily through empirical analysis. Methodologically, the study is based on a standardized online survey using an experimental two-group design with manipulated investment teasers to test Anchoring, Framing, Representativeness, Herding, and Overconfidence. In addition, eight LLMs were tested using identical teaser and questionnaire logic. Expert interviews were conducted to contextualize the findings and derive practice-oriented design principles.

The results show heuristic-consistent patterns in the human sample: price anchors shift indicated purchase prices (experts react more conservatively than students), framing affects tone perception and risk perception, and representativeness shapes trend expectations, while herding behavior remains rare and overconfidence appears only weakly pronounced. In the AI tests, model-dependent patterns emerge: several models are sensitive to anchors, framing consistently affects perception and in some cases also the investment decision, and representativeness shows a comparatively robust profile. Herding behavior is not observed, while overconfidence appears in some models in the form of high decision confidence.

For practical application, the study develops a KI-based decision agent that supports the early screening phase by structuring teaser information, consistently deriving key decision parameters, and providing transparent reasoning for evaluations. Through targeted prompt engineering, the agent aims to reduce heuristic-related distortions and promote more robust and consistent assessments. Within a human-in-the-loop framework, the final investment decision remains with the investment team, while the agent improves the comparability of opportunities and the quality of decision foundations in deal screening.

Overall, the study contributes to a better understanding of human versus AI decision logics in real estate investment management and demonstrates how empirical findings can be translated into a practical agent-based concept to support deal screening processes.

Will, Rock

Vorwort

Auch Praktiker aus der Immobilienwirtschaft sind nicht gefeit vor kognitiven Verzerrungen (Heuristiken) im Rahmen der anfänglichen Prüfung von Investmentmöglichkeiten im Investmentmanagement. Diese Arbeit untersucht Unterschiede im Entscheidungsprozess zwischen menschlichen und KI-gestützten Ansätzen, insbesondere im Hinblick auf die Ankerheuristik, den Rahmungseffekt, das Herdenverhalten, die Überkonfidenz und die Repräsentativitätsheuristik. Methodisch folgt die Untersuchung einem komparativen, quantitativ-experimentellen Design mit gezielten Informationsmanipulationen und wird durch qualitative Auswertungen zur Prozess- und Begründungslogik ergänzt. Neben der Analyse, ob und in welcher Form Menschen sowie große Sprachmodelle heuristikähnliche Muster zeigen, leitet die Arbeit Gestaltungsprinzipien für eine praxisnahe KI-Agentenstruktur ab. Ergebnis ist die Konzeption und Pilotierung eines KI-basierten Entscheidungsagenten, der menschliche Entscheider im Immobilien-Investmentmanagement durch Standardisierung, strukturierte Rückfragen und nachvollziehbare Begründungen unterstützt und so potenzielle Verzerrungsrisiken reduziert.

Die Arbeit ist bewusst praxisnah angelegt. Der entwickelte KI-Agent ist so konzipiert, dass er für reale Anwendungsfälle im Investment Management geeignet ist. Zur Veranschaulichung der praktischen Umsetzung ist ergänzend eine Videodemonstration des entwickelten KI-Agenten verfügbar.¹ Bei Interesse an einem fachlichen Austausch zu möglichen Einsatzfeldern und Weiterentwicklungen kann gerne Kontakt aufgenommen werden.²

Aschaffenburg, im April 2026

Niklas Will

Verena Rock

¹ Link: https://drive.google.com/file/d/1-AFevGlunMsJuLryvLjoS_Kj3TCzotb3/view?usp=sharing

² Kontakt: niklas.m.will@web.de

Inhalt

| | |
|---|--------|
| Tabellenverzeichnis | XI |
| Abbildungsverzeichnis | XII |
| 1 Einleitung | - 1 - |
| 1.1 Problemstellung und Relevanz | - 1 - |
| 1.2 Zielsetzung der Arbeit | - 2 - |
| 2 Theoretischer Rahmen | - 2 - |
| 2.1 Heuristiken | - 2 - |
| 2.1.1 Ankerheuristik (Anchoring) | - 3 - |
| 2.1.2 Herdenverhalten (Herding) | - 4 - |
| 2.1.3 Rahmungseffekt (Framing Effekt) | - 5 - |
| 2.1.4 Repräsentativitätsheuristik (Representativeness) | - 7 - |
| 2.1.5 Überkonfidenz (Overconfidence) | - 8 - |
| 2.2 Zusammenfassung und Überleitung Investment Management | - 9 - |
| 3 Investment Management | - 9 - |
| 3.1 Abgrenzung innerhalb des Investment Managements | - 10 - |
| 3.2 Heuristiken im Investment Management | - 10 - |
| 3.3 Treiber für Heuristik-anfällige Entscheidungen | - 11 - |
| 3.3.1 Angespanntes Finanzierungsumfeld | - 11 - |
| 3.3.2 Änderungen in der Nachfragestruktur | - 12 - |
| 3.3.3 Zunehmende Bedeutung von Nachhaltigkeit | - 12 - |
| 3.3.4 Implikationen und Relevanz | - 13 - |
| 4 Künstliche Intelligenz | - 13 - |
| 4.1 Einordnung und Abgrenzung | - 13 - |
| 4.2 Generative Künstliche Intelligenz | - 14 - |
| 4.3 KI-Agenten und die Abgrenzung zu LLMs | - 14 - |
| 4.4 KI in der Immobilienwirtschaft | - 16 - |
| 4.5 Akademische Literatur zu KI in der Immobilienwirtschaft | - 17 - |
| 5 Methodik | - 17 - |
| 5.1 Branchenumfrage | - 18 - |
| 5.1.1 Erhebung der Daten | - 18 - |
| 5.1.2 Zwei Gruppen Design | - 19 - |
| 5.2 KI-Modelltests | - 21 - |

| | | |
|-------|---|--------|
| 5.3 | Experteninterviews | - 21 - |
| 6 | Ergebnisse der Branchenumfrage | - 22 - |
| 6.1 | Ergebnisse der Gruppe 1 | - 22 - |
| 6.1.1 | Struktur der Teilnehmenden | - 22 - |
| 6.1.2 | Teaser East Star: Investitionsentscheidung, Entscheidungssicherheit, Herdenverhalten, Kaufpreis & IRR | - 23 - |
| 6.1.3 | Teaser „Courtyard“: Investitionsentscheidung, Kaufpreis & IRR sowie Tonwirkung und Risikowahrnehmung | - 24 - |
| 6.1.4 | Teaser „Primeverse“: Investitionsentscheidung, Kaufpreis & IRR sowie Mietpreiserwartung und Risikowahrnehmung | - 26 - |
| 6.1.5 | Entscheidungsmodus, Rolle der Berufserfahrung, Risikoprofil | - 27 - |
| 6.1.6 | Zusammenfassende Darstellung | - 27 - |
| 6.2 | Ergebnisse der Gruppe 2 | - 28 - |
| 6.2.1 | Struktur der Teilnehmenden | - 28 - |
| 6.2.2 | Teaser „East Star“: Investitionsentscheidung, Entscheidungssicherheit, Herdenverhalten, Kaufpreis & IRR | - 29 - |
| 6.2.3 | Teaser „Courtyard“: Investitionsentscheidung, Kaufpreis & IRR sowie Tonwirkung und Risikowahrnehmung | - 30 - |
| 6.2.4 | Teaser „Primeverse“: Investitionsentscheidung, Kaufpreis & IRR sowie Mietpreiserwartung und Risikowahrnehmung | - 31 - |
| 6.2.5 | Entscheidungsmodus, Rolle der Berufserfahrung, Risikoprofil | - 32 - |
| 6.2.6 | Zusammenfassende Darstellung der zentralen Ergebnisse | - 33 - |
| 6.3 | Entscheidungsdeterminanten | - 33 - |
| 6.4 | Informationsadäquanz | - 34 - |
| 6.5 | Entscheidungsunterstützung | - 35 - |
| 6.6 | Entscheidungssicherheit | - 35 - |
| 7 | Zusammenfassende Darstellung der Humanbefragung | - 35 - |
| 8 | KI-Test | - 37 - |
| 8.1 | Standardisierung der KI-Instruktion (Prompt Design) | - 37 - |
| 8.2 | Auswahl der getesteten Modelle | - 37 - |
| 8.3 | Durchführung und Datenerfassung | - 38 - |
| 8.4 | Deskriptive Beschreibung der KI-Tests | - 38 - |
| 8.4.1 | Ankereffekt: Preis- und Renditebezug als primäres Muster | - 38 - |
| 8.4.2 | Rahmungseffekt: Entscheidungswechsel als dominante Reaktion, aber nicht universell | - 38 - |

| | | |
|-------|---|--------|
| 8.4.3 | Repräsentativitätsheuristik: eher robuste Entscheidungen, punktuell starke Ausschläge | - 39 - |
| 8.4.4 | Herdenverhalten: robustes Verhalten über alle Modelle hinweg | - 39 - |
| 8.4.5 | Überkonfidenz: insgesamt hohes Niveau bei moderater Streuung | - 39 - |
| 8.5 | Menschliche Heuristiken in KI-Modellen | - 39 - |
| 8.6 | Plausibilisierung der beobachteten Ergebnisse | - 41 - |
| 8.7 | Entscheidungsdeterminanten in KI-Modellen | - 41 - |
| 9 | Vergleich zwischen Menschen und KI | - 43 - |
| 9.1 | Unterschied in Bezug auf Heuristiken | - 43 - |
| 9.1.1 | Ankerheuristik (Anchoring) | - 43 - |
| 9.1.2 | Rahmungseffekt (Framing) | - 43 - |
| 9.1.3 | Repräsentativitätsheuristik (Representativeness) | - 43 - |
| 9.1.4 | Herdenverhalten (Herding) | - 44 - |
| 9.1.5 | Überkonfidenz (Overconfidence) | - 44 - |
| 9.2 | Entscheidungsdeterminanten | - 44 - |
| 10 | Gestaltungsempfehlung für einen KI-Agenten | - 44 - |
| 10.1 | Prozessdesign | - 45 - |
| 10.2 | Zusammenfassung Prozessdesign | - 48 - |
| 10.3 | Test des Investment Agenten mit den Teasern | - 49 - |
| 11 | Schlussbetrachtung | - 50 - |
| 11.1 | Fazit | - 51 - |
| 11.2 | Limitationen | - 53 - |
| 11.3 | Ausblick für Forschung & Praxis | - 53 - |
| | Literatur- und Quellenverzeichnis | IX |

Abkürzungen

| | |
|-----------|--|
| bps | Basis points (Basispunkte) |
| CSV | Comma-Separated Values |
| CapEx | Capital Expenditures (Investitionsausgaben) |
| DCF | Discounted Cash Flow |
| ESG | Environmental, Social, Governance |
| EZB | Europäische Zentralbank |
| Gen KI | Generative Künstliche Intelligenz |
| GIS | Geoinformationssystem |
| GPQA | Graduate-level Google-Proof Q&A |
| HITL | Human-in-the-Loop (Mensch-in-der-Schleife) |
| IRR | Internal Rate of Return (interner Zinsfuß) |
| KI | Künstliche Intelligenz |
| KPI | Key Performance Indicator (Leistungskennzahl) |
| LLM | Large Language Model (großes Sprachmodell) |
| MDN | Median |
| NDA | Non-Disclosure Agreement (Vertraulichkeitsvereinbarung) |
| NKM | Nettokaltmiete |
| NOI | Net Operating Income (Netto-Betriebsergebnis) |
| RAG | Retrieval-Augmented Generation |
| REIM | Real Estate Investment Management (Immobilien-Investment- management) |
| REITs | Real Estate Investment Trust(s) |
| SD | Standard Deviation (Standardabweichung) |
| SWE-BENCH | Software Engineering Benchmark |

| | |
|--------|--|
| WAULT | Weighted Average Unexpired Lease Term (gewichtete durchschnittliche Restlaufzeit der Mietverträge) |
| ReHeAT | Representativeness Heuristic Assessment Test |

Tabellenverzeichnis

| | |
|--|--------|
| Tabelle 1: Zuordnung potenzieller Heuristiken in der anfänglichen Opportunitätsprüfung im Investmentmanagement..... | - 11 - |
| Tabelle 2: Deskriptive Kennwerte zur Tonwirkung und Risikowahrnehmung für "Courtyard" in Gruppe 1 (negatives Framing)..... | - 25 - |
| Tabelle 3: Deskriptive Kennwerte zur Tonwirkung und Risikowahrnehmung für "Courtyard" in Gruppe 2 (positives Framing)..... | - 30 - |
| Tabelle 4: Heuristiken in der Opportunitätsprüfung: Evidenz und Anknüpfungspunkte des Entscheidungsagenten. | - 45 - |
| Tabelle 5: KI-Agent Outputs auf ausgewählte Fragen in Bezug auf die Investment Teaser.- | 50 - |

Abbildungsverzeichnis

| | |
|---|--------|
| Abbildung 1: Vereinfachte Prozessabbildung zur Anwendung von LLMs und KI-Agenten im Investment Management. | - 15 - |
| Abbildung 2: Schematische Darstellung des Zwei-Gruppen-Studiendesigns und Versuchsablaufes. | - 19 - |
| Abbildung 3: Boxplot der Kaufpreisnennungen für "East Star" in Gruppe 1 (hoher Preisanker). Expert:innen und Gesamtheit (aggregiert). | - 23 - |
| Abbildung 4: Anteil der erwarteten Fortsetzung sinkender Mietpreise getrennt nach Erfahrungsniveau. | - 26 - |
| Abbildung 5: Boxplot der Kaufpreisnennungen für "East Star" in Gruppe 2 (niedriger Preisanker). | - 29 - |
| Abbildung 6: Anteil der erwarteten Fortsetzung steigender Mietpreise getrennt nach Erfahrungsniveau. | - 32 - |
| Abbildung 7: Heatmap zur wahrgenommenen Relevanz (Mediane) zentraler Entscheidungskriterien nach Erfahrungsgruppe | - 34 - |
| Abbildung 8: Darstellung der heuristikbezogenen Muster in der Humanbefragung | - 36 - |
| Abbildung 9: Heuristiken in KI-Modellen. | - 40 - |
| Abbildung 10: Heatmap der von KI-Modellen bewerteten Relevanz zentraler Entscheidungskriterien nach Bedingung (G1/G2). | - 42 - |
| Abbildung 11: Konzeptioneller Prozessablauf eines KI-Agenten im Rahmen der anfänglichen Opportunitätsprüfung im Investment Management. | - 46 - |

Will, Rock

Genderhinweis

Aus Gründen der Lesbarkeit wird in dem vorliegenden Paper das generische Maskulinum verwendet. Die verwendeten Personenbezeichnungen beziehen sich auf alle Geschlechter.

1 Einleitung

Investitionsentscheidungen im Immobilieninvestmentmanagement erfolgen unter Bedingungen ausgeprägter Unsicherheit, weil Immobilienmärkte im Vergleich zu liquiden Finanzmärkten geringere Transparenz, stärkere regionale Heterogenität und Informationsasymmetrien aufweisen.³ Zugleich ist die Assetklasse durch Illiquidität und lange Kapitalbindung gekennzeichnet, was sich u.a. in typischen Halteperioden von rund 8 - 11 Jahren für institutionelle Immobilieninvestments widerspiegelt.⁴ Gerade in solchen Kontexten werden Entscheidungen häufig nicht rein „rational“ im Sinne vollumfänglicher Informationsverarbeitung getroffen, sondern unter Rückgriff auf mentale Abkürzungen (Heuristiken), die kognitive Komplexität reduzieren.⁵ Diese Heuristiken sind häufig nützlich, können jedoch zu Fehlentscheidungen führen.⁶ Für das Immobilieninvestmentmanagement ist das besonders relevant, weil zentrale Teilschritte, beispielsweise die frühe Selektion und Bewertung von Investmentopportunitäten stark von Interpretation, Gewichtung und Plausibilisierung unvollständiger Informationen abhängen.⁷

Parallel gewinnt KI als Werkzeug zur Strukturierung, Beschleunigung und Standardisierung von Entscheidungsprozessen in der Branche stark an Bedeutung.⁸ Eine Befragung von 32 Immobilienakteuren ergab, dass 91% der Teilnehmenden KI innerhalb der Exposéprüfung verwenden würden und 84% waren der Meinung, dass KI in dem Bereich der Exposéprüfung relevant ist.⁹

1.1 Problemstellung und Relevanz

In diesem Umfeld gewinnt KI zunehmend an Relevanz, Unternehmen etablieren Fahrpläne für die kurzfristige sowie langfristige Integration, um Produktivität und Effizienz zu erhöhen oder ganze Workflows neu zu entwickeln.¹⁰ Für eine effektive Nutzung der KI sowie der KI-Agenten benötigt es eine differenzierte Betrachtung der Chancen und Möglichkeiten. Denn aktuelle Forschung zeigt, dass Large Language Models (LLMs) Entscheidungsmuster aufweisen können, die Heuristiken ähneln und potenziell verzerrte Empfehlungen erzeugen.¹¹

Daraus ergeben sich zwei zentrale Herausforderungen für das Investment Management: Menschliche Entscheider sind anfällig für Heuristiken, und KI-gestützte Systeme sind es potenziell ebenfalls – sei es

³ Vgl. Singh et al. (2023), S. 2.

⁴ Vgl. Ang & Westerfield (2013), S. 3.

⁵ Vgl. Tversky & Kahnemann (1974), S. 1124.

⁶ Ebd., S. 1131.

⁷ Vgl. Singh et al. (2023), S. 2.

⁸ Vgl. Jones Lang LaSalle (2025), o.S.

⁹ Vgl. Weidling & Rock (2025), S. 41.

¹⁰ Vgl. Jones Lang LaSalle (2025), o.S.

¹¹ Vgl. Echterhoff et al. (2024), S. 1 – 2.

durch replizierte Verzerrungen oder neue, durch Mensch-KI-Interaktion entstehende Fehlerquellen. Praxis und Forschung müssen daher die Entscheidungsprozesse beider Akteure gezielt vergleichen und daraus verzerrungsreduzierende Prozessdesigns ableiten.¹²

1.2 Zielsetzung der Arbeit

Da ungeklärt ist, welche kognitiven Verzerrungen in der anfänglichen Opportunitätsprüfung in KI-Systemen vorherrschen und wie sie sich von menschlichen Entscheidungsmustern unterscheiden, vergleicht diese Arbeit die frühe Bewertung von Immobilieninvestments durch Menschen und KI – und entwickelt darauf aufbauend einen KI-Agenten zur verzerrungsreduzierten Opportunitätsprüfung. Daraus ergibt sich folgendes Forschungsgebiet:

Wie unterscheiden sich Entscheidungsprozesse in der Bewertung von Investmentopportunitäten zwischen menschlichen und KI-gestützten Ansätzen im Investment Management? Konzeption eines KI-basierten Entscheidungsagenten zur Prozessunterstützung.

Es werden folgende Forschungsfragen im Rahmen dieser Arbeit untersucht:

- I. Wie unterscheidet sich die Bewertung von Investmentopportunitäten zwischen menschlichen Entscheidungsgruppen aus der Immobilienwirtschaft und künstlichen Intelligenzen, in Bezug auf Heuristiken und Entscheidungsdeterminanten?
- II. Welchen der untersuchten Heuristiken unterliegen die menschlichen Entscheidungsgruppen? Repliziert KI diese Muster oder unterscheidet sie sich systematisch?
- III. Wie lässt sich ein Konzept für eine KI basierte Entscheidungsagenten entwickeln, um im Prozess zu unterstützen und potenzielle Verzerrungen zu reduzieren oder gänzlich zu vermeiden?

2 Theoretischer Rahmen

Im Folgenden wird das theoretische Fundament dieser Arbeit aufgebaut. Wie eingangs angedeutet, handeln Menschen – auch Praktiker – in der Regel nicht durchgängig rational und verwenden in einigen Fällen unbewusst für ihre Entscheidungen mentale Abkürzungen.

2.1 Heuristiken

Heuristiken, auch bekannt als Daumenregeln, sind mentale Abkürzungen welche Individuen in Situationen mit begrenztem Informationszugang oder in Phasen erhöhter Unsicherheit verwenden. Heuristiken

¹² Vgl. Singh et al. (2023), S. 1 – 2.

ermöglichen es, Probleme zu vereinfachen und eine effiziente, anscheinend zufriedenstellende Lösung zu finden.¹³ Es gibt eine Bandbreite an Heuristiken, welche zu irrationalen Entscheidungen führen können. Im Folgenden werden die in dieser Arbeit untersuchten Heuristiken vorgestellt. Diese wurden im Rahmen einer Literaturrecherche identifiziert und für die anfängliche Opportunitätsprüfung als besonders relevant identifiziert.

2.1.1 Ankerheuristik (Anchoring)

Die Ankerheuristik beschreibt die Tendenz von Menschen, bei numerischen Urteilen von einem Anfangswert (sogenannt „Anker“) auszugehen und sich anschließend nur unzureichend davon zu lösen. Klassischerweise wird ein anfangs sehr hoher oder sehr niedriger Ausgangswert präsentiert, von dem aus Individuen im Rahmen möglicher Preisverhandlungen schrittweise anpassen, wobei diese Anpassungen systematisch zu gering ausfallen und die Endurteile somit in Richtung des Ankers verzerrt sind.¹⁴ Im Kontext von Immobilieninvestmententscheidungen äußert sich diese Heuristik insbesondere darin, dass Käufer sich an historischen Kaufpreisen, früher erzielten Renditen, Kursniveaus, Marktindizes oder vorgegebenen Werten orientieren und ihre Erwartungen sowie Entscheidungen daran ausrichten.¹⁵

2.1.1.1 Ankerheuristik in der Immobilienliteratur

Zu den einflussreichsten Arbeiten zählt die Studie von Northcraft und Neale (1987), die in einem realitätsnahen Setting zeigen, dass sowohl Laien als auch Immobilienexperten stark durch den Angebotspreis eines Einfamilienhauses beeinflusst werden. Trotz identischer objektiver Informationen lagen Bewertungen, Verhandlungsziele und Zahlungsbereitschaften systematisch höher bzw. niedriger, wenn der Angebotspreis hoch bzw. niedrig angesetzt war.¹⁶ Black und Diaz (1996) untersuchen explizit die Rolle des Angebotspreises in Verhandlungen und zeigen, dass Verhandlungsergebnisse deutlich in Richtung des geforderten Preises verzerrt sind. Selbst wenn zusätzliche Informationen bereitgestellt werden, dominiert der Angebotspreis als Referenzpunkt, was belegt, dass Marktteilnehmer den Anker nicht ausreichend korrigieren.¹⁷

2.1.1.2 Relevanz der Ankerheuristik für den Immobilienmarkt

Immobilieninvestitionen sind kapitalintensiv und mit hohen Transaktionskosten verbunden. Fehlentscheidungen lassen sich schwer korrigieren und wirken sich langfristig auf das Rendite-Risiko Verhältnis aus.¹⁸ Gerade in weniger transparenten Märkten greifen Marktteilnehmer stärker auf Heuristiken

¹³ Vgl. Tversky & Kahnemann (1974), S. 1124 – 1131.

¹⁴ Ebd., S. 1128.

¹⁵ Vgl. Zain ul Abdin et al. (2017), S. 677f.

¹⁶ Vgl. Northcraft & Neale (1987), S. 84f.

¹⁷ Vgl. Black & Diaz (1996), S. 287 – 297.

¹⁸ Vgl. Singh et al. (2023), S. 2 – 4.

zurück, weil objektive Marktinformationen fehlen oder schwer zugänglich sind.¹⁹ Makler, Gutachter und Fondsmanager können selbst Anker setzen (z.B.: durch Listing-Preis, Gutachten oder Research-Berichte) und werden gleichzeitig von diesen Anker beeinflusst. Das macht die Ankerheuristik zu einem strukturellen Phänomen entlang der Wertschöpfungskette von Immobilien Investitionen.

2.1.1.3 Ankereffekt in KI-Systemen

Neuere Studien dehnen diese Fragestellung explizit auf KI-Systeme aus: Chen et al. (2025) untersuchen anhand eines Operations-Management-Szenarios, ob das KI Model ChatGPT bei Bestellmengenentscheidungen ähnlich wie menschliche Manager auf gegebene Referenzwerte (z.B.: Durchschnittsnachfrage) ankernd reagiert. Sie finden, dass die KI, trotz überdurchschnittlicher Rechenleistung, an menschliche Ankereffekte erinnernde Verzerrungen zeigt.²⁰ Nguyen zeigt, dass LLMs die in ihren Trainingsdaten enthaltenen Referenzwerte als Anker übernehmen und bei nachfolgenden quantitativen Einschätzungen systematisch unzureichend davon abweichen; ein klassischer Ankereffekt. Der Autor analysiert in randomisierten Experimenten, inwieweit große Sprachmodelle (u.a. GPT-4, Claude 2 und Gemini Pro) bei numerischen Prognosen durch vorab genannte hohe bzw. niedrige Werte beeinflusst werden. Die Ergebnisse zeigen klare Ankereffekte über alle getesteten Modelle hinweg.²¹ Darüber hinaus zeigen neuere Arbeiten zu Ankereffekten in LLMs, dass KI-Modelle auf in der Aufgabenstellung versteckte numerische Hinweise hochsensibel reagieren und ihre Antworten signifikant in Richtung dieser Anker verschoben werden.²²

2.1.2 Herdenverhalten (Herding)

Unter dem Begriff Herdenverhalten versteht man im finanzwirtschaftlichen Kontext die Tendenz von Investoren, die Entscheidungen anderer Marktteilnehmer zu imitieren, anstatt ausschließlich auf eigene Informationen oder Fundamentaldaten zu vertrauen.²³

2.1.2.1 Herdenverhalten in der Immobilienliteratur

In der Literatur wird verstärkt darauf hingewiesen, dass es bei der Ausprägung dieser Heuristik bereits einen relevanten Unterschied gibt zwischen Expert:innen (professionellen Investoren aus der Branche) und privaten Haushalten.

Für die Immobilienwirtschaft zeigt die systematische Literaturübersicht von Singh et al. (2023), dass das Herdenverhalten eine zentrale Heuristik ist und in unterschiedlichen Bereichen (u.a. REITs, Wohn-

¹⁹ Vgl. Zain-ul-Abdin et al. (2019), S. 2f.

²⁰ Vgl. Chen et al. (2025), S. 359 – 361.

²¹ Vgl. Nguyen (2024), S. 1f.

²² Vgl. Lou & Sun (2024), S. 5f.

²³ Vgl. Singh et al. (2023), S. 3.

und Gewerbeimmobilien) empirisch nachgewiesen wurde; auch bei professionellen Investoren wie Fondsmanager. Die Autoren empfehlen weitere Forschung aufgrund begrenzter Daten.²⁴

Das Herdenverhalten ist zudem stark kontextabhängig und variiert mit Marktregimen. Zahlreiche Studien, auf die sowohl Singh et al. (2023) als auch Lowies et al. (2016) verweisen, zeigen, dass Herdenverhalten häufig in Phasen hoher Unsicherheit, hoher Volatilität oder Krisen zunimmt.^{25, 26} Lowies et al. (2016) fanden kein statistisch signifikantes Herdenverhalten bei südafrikanischen Property Fondsmanagern (n=17). Irrationales Herdenverhalten wurde vollständig abgelehnt (0%), rationales Herding tritt bei besser informierten Konkurrenten oder Informationsunsicherheit auf. Irrationales Herdenverhalten beschreibt das „blinde Mitlaufen“ der Mehrheit ohne tragfähige Informationsgrundlage, während rationales Herdenverhalten ein begründetes Folgen darstellt, weil das Verhalten besser informierter Marktteilnehmer als Signal genutzt wird.²⁷

2.1.2.2 Relevanz von Herdenverhalten für den Immobilienmarkt

Herding ist ein soziales Phänomen: Menschen orientieren sich an anderen Menschen, wenn Informationen unsicher, teuer oder schwer interpretierbar sind, wie typischerweise in Immobilienmärkten mit illiquiden Assets, intransparenten Daten und langen Entwicklungszyklen.^{28, 29}

2.1.2.3 Herdenverhalten in KI-Systemen

LLM-basierte Systeme wie ChatGPT können einerseits **rationaler** entscheiden als Menschen, da sie keine Emotionen haben und in vielen Settings stärker auf konsistente Informationsverarbeitung und Berechnungen zurückgreifen. Experimente von Lee und Hansen (2025) zeigen z.B., dass LLM-Agenten in Experimenten zum Herdenverhalten weniger der Herde folgen und eher auf „private Informationen“ achten als menschliche Finanzexperten.³⁰

2.1.3 Rahmungseffekt (Framing Effekt)

Der Rahmungseffekt beschreibt, dass inhaltlich identische Informationen je nach Darstellung (z.B.: 75% Vermietungsquote versus 25% Leerstandsrate) zu unterschiedlichen Entscheidungen führen können.³¹ Im Kontext der Investment Opportunitäten bedeutet das, dass nicht nur der Inhalt der Informationen, sondern auch die Präsentation dieser Informationen einen Einfluss auf Investitionsentscheidungen haben kann.

²⁴ Ebd., S. 10.

²⁵ Ebd.

²⁶ Vgl. Lowies et al. (2016), S. 60 – 62.

²⁷ Ebd.

²⁸ Vgl. Singh et al. (2023), S. 10f.

²⁹ Vgl. Lowies et al. (2016), S. 54

³⁰ Vgl. Hansen & Lee (2025), S. 2f.

³¹ Vgl. Tversky & Kahnemann (1981), S. 453.

2.1.3.1 Rahmungseffekt in der Immobilienliteratur

Die spezifische Untersuchung von Framing-Effekten im Immobilienkontext ist, im Vergleich zu Heuristiken wie dem Ankereffekt oder dem Herdenverhalten, relativ jung und weniger umfangreich, gewinnt aber deutlich an Bedeutung.³² Jin und Gallimore (2010) untersuchen experimentell, wie die Darstellung identischer Marktinformationen die Wahrnehmung des Immobilienmarkts verändert. In ihrem Experiment erhalten Teilnehmende einen kommerziellen Marktbericht, dessen Daten in verschiedenen Rahmungen präsentiert werden (z.B.: stärker negativ bzw. positiv betonte Textpassagen und Diagramme). Sie zeigen, dass die Art der Informationsaufbereitung zu Unterschieden in der Einschätzung von Marktrisiken und Preisentwicklung führen kann. Die Ergebnisse belegen, dass Immobilieninvestoren nicht nur auf fundamentale Daten reagieren, sondern stark auf Narrative und Frames in Marktberichten, Research-Reports oder Präsentationen.³³

2.1.3.2 Relevanz des Rahmungseffektes in der Immobilienwirtschaft

Im Gegensatz zu stärker erforschten Heuristiken wie dem Ankereffekt oder dem Herdenverhalten ist die Untersuchung des Rahmungseffektes im Immobilienkontext noch relativ jung; gewinnt jedoch rapide an Bedeutung.³⁴ Bereits erwähnte Publikationen verdeutlichen, dass es für den Rahmungseffekt auch praktische Implikationen gibt und Entscheidungen je nach Darstellung zu unterschiedlichen Entscheidungen führen können. Darüber hinaus kann sogar die Risiko- und Preiseinschätzung beeinflussen werden obwohl objektiv betrachtet die gleichen Informationen präsentiert werden.³⁵

2.1.3.3 Rahmungseffekt in KI-Systemen

Zhang (2025) findet in einer systematischen Studie, dass verschiedene LLMs nicht auf das Framing reagieren und rationale Entscheidungen treffen. Die Modelle verhalten sich grundsätzlich risikoavers und unabhängig von Rahmungen.³⁶ Xiao und Wang fanden heraus, dass sich GPT-Modelle in *Life-and-Death*-Entscheidungen systematisch von Menschen unterscheiden, insbesondere im Hinblick auf den Framing-Effekt: Während Menschen bei positivem Framing eher risikoavers und bei negativem Framing eher risikofreudig entscheiden, produziert GPT-3.5 häufig invertierte oder instabile Rahmungseffekte und GPT-4 weitgehend kontextunabhängig risikoaverse Antworten, sodass zentrale charakteristische Muster menschlicher Entscheidungsanomalien von beiden Modellen nicht abgeleitet werden können.³⁷ Die Arbeit von Germani und Spitale zeigt, dass große Sprachmodelle bei der Bewertung politisch kontroverser Aussagen systematischen Verzerrungen durch Quellenabhängigkeit unterliegen. Ohne Quellen-

³² Vgl. Singh et al. (2023), S. 12.

³³ Vgl. Jin & Gallimore (2010), o.S.

³⁴ Vgl. Singh et al. (2023), S.12.

³⁵ Vgl. Jin & Gallimore (2010), o.S.

³⁶ Vgl. Zhang et al. (2025), S. 360.

³⁷ Vgl. Xiao & Wang (2025), S. 3f.

Will, Rock

Angabe stimmen die Modelle in ihren Urteilen stark überein, doch sobald identitätsbezogene Informationen zur vermeintlichen Quelle bereitgestellt werden (z.B. „Person aus China“), verändern sich die Zustimmungswerte plötzlich, insbesondere bei geopolitischen Themen. Besonders ausgeprägt ist eine Verzerrung gegenüber Aussagen, die einer Person aus China zugeschrieben werden. Die Autoren folgern, dass LLM-basierte Bewertungssysteme nicht nur inhalts-, sondern auch quellensensitiv sind und dadurch potenziell diskriminierende oder politisch verzerrte Entscheidungen reproduzieren können.³⁸

2.1.4 Repräsentativitätsheuristik (Representativeness)

Die Repräsentativitätsheuristik beschreibt die Tendenz von Menschen, Wahrscheinlichkeiten danach zu beurteilen, wie sehr ein Ereignis einem „typischen“ Fall ähnelt, statt nach statistischen Grundlagen zu entscheiden.^{39, 40} Aufbauend auf der klassischen Literatur zur Repräsentativitätsheuristik von Tversky und Kahnemann, entwickelt jüngere Forschung sechs unterscheidbare Subtypen: Base Rate Fallacy, Conjunction Fallacy, Disjunction Fallacy, Insensitivity to Sample Size, Misconceptions of Chance und Regression Fallacy.⁴¹

2.1.4.1 Repräsentativitätsheuristik in der Immobilien- und Finanzliteratur

In der allgemeinen Finanzmarktforschung wird Repräsentativität häufig mit Trendfolgen, „Hot-Stock“-Jagd in Verbindung gebracht, bei der Investoren die zukünftige Performance einer Aktie an der historischen Performance beurteilen und somit Schlüsse daraus ziehen, wie sich ein Aktienpreis in der Zukunft entwickeln wird.^{42, 43}

2.1.4.2 Effekt der Repräsentativitätsheuristik auf die Immobilienwirtschaft

Die Literaturübersicht von Singh et al. (2023) benennt die Repräsentativitätsheuristik explizit als eine vergleichsweise untererforschte Heuristik, hebt aber hervor, dass vorhandene Studien sie als dominanten Einflussfaktor für Immobilienentscheidungen identifizieren.⁴⁴

2.1.4.3 Repräsentativitätsheuristik in KI-Systemen

Wang et al. (2024) bauen mit „ReHeAT“ (Representativeness Heuristic Assessment Test) einen speziellen Testdatensatz für Repräsentativität und zeigen, dass mehrere LLMs in unterschiedlichen Varianten des Linda-Problems⁴⁵ typische Repräsentativitätsfehler machen, obwohl sie das korrekte statistische

³⁸ Vgl. Germani & Spitale (2025), S. 3f.

³⁹ Vgl. Tversky & Kahnemann (1974), S. 1124f.

⁴⁰ Vgl. Cascao et al. (2023), S. 99.

⁴¹ Vgl. Wang et al. (2024), S. 3 – 5.

⁴² Vgl. Abdin et al. (2017), S. 677.

⁴³ Vgl. Singh (2023), S. 2.

⁴⁴ Ebd., S. 11.

⁴⁵ Bei dem sog. Linda Problem erhalten LLM's eine Beschreibung einer Bankangestellten (Linda) und sollen daraufhin eine Frage beantworten, ob Linda zudem auch feministisch aktiv ist.

Wissen besitzen. Ein verbesserter Prompt bei, der auf die Heuristik hingewiesen wird, reduziert die Fehler deutlich.⁴⁶

2.1.5 Überkonfidenz (Overconfidence)

Überkonfidenz ist eine kognitive Verzerrung, bei der Investoren ihre Fähigkeiten, ihr Wissen und ihre Prognosen überschätzen sowie zukünftige Unsicherheiten unterschätzen. Überkonfidente Investoren neigen dazu, die Eintrittswahrscheinlichkeit positiver Ereignisse zu hoch einzuschätzen, sich selbst für überdurchschnittlich kompetent zu halten und zu glauben, den Markt schlagen zu können.^{47,48}

2.1.5.1 Überkonfidenz in der Immobilienliteratur

Singh et al. (2023) zeigen in ihrer Arbeit, dass Überkonfidenz zu einem häufig untersuchten Phänomen zählt, allerdings die Meinungen hinsichtlich der Anwendbarkeit auf die Immobilienbranche auseinander gehen. Während einige einen direkten Zusammenhang feststellen können, argumentieren andere, dass dies nicht nachvollziehbar ist.⁴⁹ Bao und Li (2016) argumentieren, dass Überkonfidenz ein zentraler Treiber von Marktineffizienzen in Immobilienmärkten ist und diskutieren zugleich die methodischen Schwierigkeiten, diese Heuristik zu messen.⁵⁰

2.1.5.2 Relevanz von Überkonfidenz in der Immobilienwirtschaft

Eichholtz/Yönder (2014) untersuchen US-REITs und zeigen, dass REITs mit überkonfidenten CEOs mehr Assets kaufen und weniger verkaufen als Vergleichsunternehmen. Zudem weisen diese REITs schwächere Performance-Kennzahlen auf.⁵¹ Singh et al. (2023) referenzieren diese Studie als erstes systematisches Evidenzstück, dass Überkonfidenz von Top-Managern direkt in realen Immobilienportfolios zu Überinvestition und Halteverhalten führt.⁵²

2.1.5.3 Überkonfidenz in KI-Systemen

LLMs reproduzieren die menschliche Verzerrung der Überkonfidenz systematisch: Ihr Selbstvertrauen übersteigt die tatsächliche Genauigkeit um 30 - 60%, was sie für risikobehaftete Anwendungen wie Immobilienprognosen unzuverlässig macht.⁵³ Chen et al. (2025) zeigen in umfassenden Tests, dass GPT-4 signifikant stärker selbstbewusst ist als GPT-3.5. Speziell im Überkonfidenz-Test überschätzen beide Modelle ihre Performance, wobei GPT-4 konsequent höhere Selbstsicherheit angibt.⁵⁴ Tian et al. (2025)

⁴⁶ Vgl. Wang (2024), S. 2.

⁴⁷ Vgl. Zain ul Abdin et al. (2017), S. 676.

⁴⁸ Vgl. Singh et al. (2023), S. 11.

⁴⁹ Ebd., S. 11f.

⁵⁰ Vgl. Bao & Li (2016), S. 1f.

⁵¹ Vgl. Eichholtz & Yönder (2014), S. 1.

⁵² Vgl. Singh et al. (2023), S. 11.

⁵³ Vgl. Tian et al. (2025), S. 2f.

⁵⁴ Vgl. Chen et al. 2025, Tabelle 1.

quantifizieren dies im „LLM-as-a-Judge“-Setting: LLMs erreichen Brier Scores⁵⁵ von 0.20 - 0.52 (menschlich: 0.15), reagieren jedoch hervorragend auf Prompt-Optimierung.⁵⁶

2.2 Zusammenfassung und Überleitung Investment Management

Die vorangegangene Analyse zeigt, dass zentrale Verhaltensmuster wie der Ankereffekt, das Herdenverhalten, der Rahmungseffekt, die Repräsentativitätsheuristik oder die Überkonfidenz nicht nur menschliche Entscheidungen, sondern zunehmend auch KI-basierte Systeme prägen und damit jene Werkzeuge, die in der Immobilienwirtschaft zunehmend an Bedeutung gewinnen. Gerade weil Immobilieninvestitionen durch hohe Kapitalbindung, Illiquidität und Informationsasymmetrien gekennzeichnet sind, können solche Heuristiken langfristig Risiko- und Renditeprofile ganzer Portfolios verzerren. Auf dieser Grundlage richtet sich der weitere Fokus der Arbeit auf die anfängliche Opportunitätsprüfung im Investment Management in der Immobilienwirtschaft und untersucht, wie sich diese Entscheidungsmechanismen zwischen Menschen und künstlichen Intelligenzen unterscheiden.

3 Investment Management

Immobilien Investment Management bezeichnet die strategische und operative Investition in Immobilienkapitalanlagen entlang des gesamten Investitionsprozesses.⁵⁷ Es umfasst insbesondere die Identifikation, das Sourcing, die Analyse, Strukturierung und Umsetzung von Immobilien Investitionen mit dem Ziel einem vorgegebenen Ankaufsprofil und Renditeerwartungen gerecht zu werden.⁵⁸ Das Investment Management fungiert als Gesicht des Investors im Transaktionsprozess und ist Ansprechpartner für Investitionsoportunitäten, unabhängig davon, ob der Investor ein Family Office, ein Fonds oder eine andere Einheit ist. Investmentopportunitäten – Angebote einzelner Objekte oder Portfolios durch Makler, Verkäufer oder eigene Marktansprache – bilden den zentralen Eingangspunkt des Investment Managements.⁵⁹ Nach Eingang prüft ein Mitarbeiter die Opportunität; bei Interesse folgen NDA-Unterzeichnung sowie die Übermittlung des Investment Memorandums und weiterer Objektinformationen. Diese Arbeit fokussiert die anfängliche Opportunitätsprüfung, in der das Investment Management anhand definierter Rahmenparameter (Lage, Risikoklasse, Transaktionsstruktur etc.) eine Vorauswahl trifft. Geeignete Objekte werden anschließend dem Investmentkomitee zur Entscheidung über die weitere Verfolgung vorgestellt.⁶⁰

⁵⁵ Der Brier Score misst die mittlere quadratische Abweichung zwischen vorhergesagten Wert und tatsächlichen Wert (0=perfekt, 1=unbrauchbar) (Vgl. Tian et al. (2025), S. 6.).

⁵⁶ Vgl. Tian et al. (2025), S. 5f.

⁵⁷ Vgl. gif (2011), o.S.

⁵⁸ Vgl. Experte A (2025), Interview.

⁵⁹ Ebd.

⁶⁰ Ebd.

3.1 Abgrenzung innerhalb des Investment Managements

Investmentopportunitäten werden über Teaser, Exposés und Investment Memoranden vermarktet. Der Teaser dient der Interessensweckung und enthält nur anonymisierte Kerndaten; das Exposé konkretisiert die Opportunität mit Lage, Kennzahlen und Unterlagen. Das Investment Memorandum wird erst nach NDA bereitgestellt und bildet als umfassendes, vertrauliches Dokument die Grundlage für Finanzmodellierung und finale Investmententscheidung.⁶¹ Gerade in den frühen Phasen, in denen eine große Zahl potenzieller Opportunitäten bei begrenzter Informationsgrundlage zu beurteilen ist, greifen Entscheidungsträger verstärkt auf vereinfachende Entscheidungsregeln und Erfahrungswerte zurück.^{62, 63} Welche Opportunitäten in diesem Screening-Prozess weiterverfolgt und welche verworfen werden, prägt maßgeblich das spätere Rendite-Risiko-Profil des Immobilienportfolios.⁶⁴

3.2 Heuristiken im Investment Management

Die untersuchten Heuristiken wurden nicht beliebig gewählt, sondern anhand des praktischen Anwendungskontexts im Investment Management abgeleitet. Tabelle 1 ordnet sie den jeweiligen Schritten der Opportunitätsprüfung zu und schlägt so die Brücke zwischen theoretischem Konstrukt und realem Entscheidungsrahmen – als konzeptionelle Übersicht auf Basis von Kapitel 2.

⁶¹ Vgl. Experte A (2025), Interview.

⁶² Vgl. Gallimore et al. (2000), S. 602f.

⁶³ Vgl. Lowies et al. (2016), S. 54 – 55.

⁶⁴ Vgl. Singh et al. (2023), S. 2 – 3.

| | Schritt in der Opportunitätsprüfung | Typischer Informationsreiz im Teaser/Exposé | Heuristik | Erwartetes Verzerrungsmuster |
|---|--|---|-----------------------------|--|
| 1 | Ersteindruck & "Investment Story" | "Investment Story", Ton des Textes & Darstellung der Fakten | Rahmungseffekt | Identische Fakten führen je nach Darstellung zu anderer Risiko-/Attraktivitätswahrnehmung. |
| 2 | Erste Preis- & Renditeplausibilisierung | Kaufpreisvorstellung / Angebotspreis | Ankereffekt | Numerischer Startwert dient als Anker; spätere Anpassungen lösen sich unzureichend davon. |
| 3 | Narrative aus dem Markt / Cash Flow | Trendgrafiken (z.B.: zu den Mieten) | Repräsentativitätsheuristik | Muster aus der Vergangenheit werden als repräsentativ für die Zukunft interpretiert: Trendfortschreibung. |
| 4 | Informationen aus dem Markt / von Marktteilnehmern | Hinweis auf andere Investoren/Mehrheitsmeinung/ Interesse | Herdenverhalten | Orientierung an Mehrheit bei Unsicherheit: Anpassung an soziale Information statt rationale Bewertung. |
| 5 | Subjektive Entscheidungssicherheit & Kompetenzgefühl | Selbsteinschätzung „Wie sicher bin ich?“ / „Wie kompetent?“ | Überkonfidenz | Überschätzung eigener Fähigkeiten/Wissen/Prognosen bei gleichzeitiger Unterschätzung künftiger Unsicherheiten. |

Tabelle 1: Zuordnung potenzieller Heuristiken in der anfänglichen Opportunitätsprüfung im Investmentmanagement.⁶⁵

Tabelle 1 verdeutlicht, dass sich Heuristiken in der Opportunitätsprüfung nicht isoliert, sondern entlang typischer Informationsreize aus Teasern/Exposés entwickeln. Frühe Rendite- und Preisnennungen begünstigen Ankereffekte. Narrative Markt- oder Asset-Stories können Urteile durch gewisse Rahmungen verstärken, indem sie Fakten in einer gewissen Art und Weise darstellen (eher positiv/eher negativ). Hinweise auf „andere Investoren“ oder vermeintliches Wettbewerbsinteresse fördern Herdenverhalten, während die Kombination aus Zeitdruck und unvollständigen Daten die Entscheidungssicherheit erhöhen kann und damit Überkonfidenz begünstigen kann.

3.3 Treiber für Heuristik-anfällige Entscheidungen

Die zuvor dargestellte Tabelle 1 ordnet die untersuchten Heuristiken den Entscheidungsschritten in der anfänglichen Opportunitätsprüfung zu. Der folgende Abschnitt erklärt, welche Unsicherheiten und Komplexitäten es im Bereich des Immobilien Investment Prozesses zu beachten gilt und wie diese erhöhte Unsicherheit und Komplexität die Heuristik anfälligen Entscheidungssituationen zusätzlich verstärken kann.⁶⁶

3.3.1 Angespantes Finanzierungsumfeld

Während COVID-19 ab 2020 einen exogenen Schock auslöste, gilt 2022 als eigentliche Trendwende: Das Zusammentreffen geopolitischer Spannungen, Energiepreisschocks und stark steigender Inflation

⁶⁵ Eigene Darstellung. Die theoretische Herleitung der fünf Heuristiken sowie die hier verwendeten Wirkannahmen zu (i) unzureichender Anpassung bei numerischen Startwerten, (ii) darstellungsabhängigen Urteilen bei inhaltlich identischen Informationen, (iii) Muster-/Trendfortschreibung, (iv) Orientierung an Mehrheits- bzw. Fremdentscheidungen unter Unsicherheit und (v) Überschätzung eigener Fähigkeiten bei gleichzeitiger Unterschätzung von Unsicherheit vgl. Kap. 2.1.1 - 2.1.5.

⁶⁶ Vgl. Gigerenzer & Gaissmaier (2011), S. 451.

veränderte Bewertungs- und Finanzierungsbedingungen sowie die Risikowahrnehmung im Transaktionsgeschehen nachhaltig.^{67, 68} Parallel dazu setzte seit Juli 2022 eine geldpolitische Normalisierung mit deutlichen Zinsschritten ein; die Deutsche Bundesbank dokumentiert u.a. Zinserhöhungen um jeweils 75 Basispunkte in September und Oktober 2022.⁶⁹ Die Bank-Lending-Survey der EZB berichtet für Q1 2023 eine markante Verschärfung von Kreditstandards sowie eine ausgeprägte Abschwächung der Kreditnachfrage, wobei das Zinsniveau als zentraler Treiber benannt wird.⁷⁰ Dies erschwert es, rentable Immobilienopportunitäten zu identifizieren, vor allem da durch das angespanntere Finanzierungsumfeld, der Fremdkapitalhebel deutlich geringer ausfällt.⁷¹

3.3.2 Änderungen in der Nachfragestruktur

Auf der Immobilienmarktebene verweist DZ HYP auf strukturelle Veränderungen, welche die Anforderungen der Branche anpasste. Hierzu zählen vor allem veränderte Flächenbedarfe, durch Home-Office Regelungen, aber auch erhöhte Anforderungen an die E-Commerce Branche.⁷² Auch die Attraktivität von Nutzungsarten hat sich grundlegend gewandelt. Während Büroimmobilien vor der COVID-19 Pandemie noch zu einer attraktiven Nutzungsart zählten, rangieren mittlerweile vor allem Rechenzentren und Logistikimmobilien auf den vorderen Plätzen.^{73, 74} In der Folge wird die Auswahl wirtschaftlich tragfähiger Immobilienakquisitionen anspruchsvoller, was sich auch in deutlich gesunkenen Transaktionsvolumina seit 2022 widerspiegelt.⁷⁵

Vor diesem Hintergrund sind Investment Manager stärker gefordert, Markt- und Strukturentwicklungen laufend zu verfolgen und Entscheidungen konsistent in einem angespannten, unsicheren Umfeld zu treffen.⁷⁶

3.3.3 Zunehmende Bedeutung von Nachhaltigkeit

ESG⁷⁷ ist im Real-Estate-Investment-Management inzwischen nicht mehr als Zusatzanforderung zu verstehen, sondern als Erwartung institutioneller Investoren und damit als relevante Rahmenbedingung der Opportunitätsbewertung zu berücksichtigen.⁷⁸ Der Druck auf die Akteure in der Immobilienwirtschaft hat sich durch die EU Taxonomie Verordnung sowie die Offenlegungsverordnung entsprechend erhöht

⁶⁷ Vgl. Europäische Zentralbank (2023 a), S. 4 – 13.

⁶⁸ Vgl. Hüther (2023), S. 63.

⁶⁹ Vgl. Deutsche Bundesbank (2022), S. 25.

⁷⁰ Vgl. Europäische Zentralbank (2023 b), S. 3 – 4.

⁷¹ Vgl. DZ HYP AG (2023), S. 13

⁷² Ebd., S. 2 – 5.

⁷³ Vgl. PWC (2023), S. 8f.

⁷⁴ Vgl. Deutsche Hypo & Bulwiengesa AG (2025), o.S.

⁷⁵ Vgl. PWC (2024), S. 1.

⁷⁶ Vgl. DZ HYP AG (2023), S. 2 – 3.

⁷⁷ ESG (Environmental, Social, Governance) bezeichnet einen Kriterienrahmen zur Bewertung der Nachhaltigkeitsqualität von Immobilien entlang ökologischer (E), sozialer (S) und unternehmensführungsbezogener (G) Aspekte. (Vgl. ZIA Zentraler Immobilien Ausschuss e.V. (2023), S. 9 – 12.)

⁷⁸ Vgl. ZIA Zentraler Immobilien Ausschuss e.V. (2023), S. 5.

und soll Immobilienakteure dazu bringen, ihr Kapital verstärkt nachhaltig zu investieren.⁷⁹ Nachhaltigkeitszertifikate reduzieren den Analyseaufwand, ersetzen aber keine inhaltliche Prüfung: Zertifikate verschiedener Stellen sind nicht direkt vergleichbar und bilden tatsächliche Betriebsperformance sowie Taxonomie- oder Dekarbonisierungskonformität oft nur unzureichend ab.⁸⁰

3.3.4 Implikationen und Relevanz

Seit der Zinswende 2022 prägen erhöhte Unsicherheit und restriktivere Finanzierungsbedingungen das Marktumfeld. Investmententscheidungen erfordern daher eine fundiertere Prüfung von Cashflow-Robustheit, Finanzierungsstruktur und Exit-Szenarien. Zugleich erhöht die veränderte Beliebtheit einzelner Nutzungsarten, die Komplexität, passende Opportunitäten zu identifizieren. Gleichzeitig bleiben Informationslage und Datenqualität in frühen Phasen naturgemäß begrenzt. Vor diesem Hintergrund und angesichts fortbestehender Unsicherheiten hinsichtlich Geldpolitik und makroökonomischem Umfeld müssen Investitionsentscheidungen bewusster und risikosensitiver getroffen werden.

4 Künstliche Intelligenz

Vor dem Hintergrund eines Investmentumfelds, in dem die Opportunitätsprüfung durch geopolitische Spannungen, veränderte Finanzierungsbedingungen und strukturelle Veränderungen der Branche zunehmend komplexer wird, gewinnt die Perspektive der begrenzten Rationalität an Bedeutung: Entscheidungen werden unter einem gewissen Grad an Unsicherheit getroffen, wodurch vereinfachende Auswahlmechanismen herangezogen werden.^{81,82} Künstliche Intelligenz kann dabei als Ansatz zur Prozessunterstützung im Investment Management dienen, weil KI-Systeme potenziell dabei helfen können, textbasierte Informationen aufzubereiten und Analyse- und Bewertungslogiken stärker zu standardisieren. Auf der anderen Seite ist die Annahme eines per se rationaleren KI-Urteils kritisch zu prüfen.^{83,84} Es ist zu berücksichtigen, dass LLMs trotz großer Rechenleistung weiterhin unerwünschte Verhaltensweisen zeigen können und ihre Ergebnisse sensitiv gegenüber Prompt-Variationen sind, sodass die Qualität der Antworten nicht nur modell-, sondern auch eingabe- und kontextabhängig ist.⁸⁵

4.1 Einordnung und Abgrenzung

Im Rahmen dieser Arbeit bezeichnet „KI-basierter Entscheidungsagent zur Prozessunterstützung“ keine vollständige Automatisierung der Investitionsentscheidung, sondern ein konzeptionell unterstützendes System in der anfänglichen Prüfung von Investmentopportunitäten. Vor allem in dieser Phase treffen

⁷⁹ Ebd. S. 6 – 7.

⁸⁰ Vgl. Kitel (2022), S. 2 – 3.

⁸¹ Vgl. Simon (1955), S. 99 – 100.

⁸² Vgl. Kahneman (2003), S. 1449 – 1450.

⁸³ Vgl. Kahneman (2003), S. 1449 – 1450.

⁸⁴ Vgl. Ouyang et al. (2022), S. 1 – 2.

⁸⁵ Ebd.

Investmentteams vorläufige Go/No-Go Entscheidungen, oft unter Zeitdruck und einer unvollständigen Informationslage.⁸⁶ Damit ist das Screening besonders anfällig für Verzerrungen, die aus heuristischen Vereinfachungen und aus der Art der Informationsdarstellung in Teasern und Exposés entstehen. Ziel dieses Kapitels ist es, KI nicht technisch im Detail zu erklären, sondern einen Überblick über die in dieser Arbeit verwendeten Technologien (generative KI, KI-Agenten) zu geben, um daraus im späteren Verlauf der Arbeit konkrete Anforderungen an ein praxistaugliches Prozessdesign abzuleiten.⁸⁷ Darüber hinaus wird der aktuelle Status quo des KI-Einsatzes in der Immobilienwirtschaft beleuchtet.

4.2 Generative Künstliche Intelligenz

Generative Künstliche Intelligenz (Gen KI) bezeichnet ein Teilgebiet der Künstlichen Intelligenz, das sich auf Verfahren stützt, die aus vorhandenen Daten statistische Muster und Strukturen lernen, um daraus neue Inhalte wie Texte, Bilder, Audio, Video oder Programmcode zu erzeugen.⁸⁸ Generative KI – insbesondere große Sprachmodelle wie ChatGPT – bildet die technologische Grundlage dieser Arbeit. Sie wird sowohl als eigenständiges LLM als auch als Kernkomponente eines KI-basierten Entscheidungsagenten betrachtet.^{89,90}

4.3 KI-Agenten und die Abgrenzung zu LLMs

KI-Agenten werden schon seit den 90er Jahren als potenzieller Ansatz zur Realisierung künstlicher allgemeiner Intelligenz diskutiert, da sie Aufgaben durch eigenständige Planung und Handlung bewältigen sollen.⁹¹ Bei der Nutzung großer Sprachmodelle (LLM) bleibt der Ablauf, wie Abbildung 1 veranschaulicht, weitgehend menschenzentriert: Unterlagen werden manuell gesichtet, per Prompt verarbeitet und anschließend dennoch subjektiv interpretiert. Dies führt zu Wartezeiten und inkonsistenter Ergebnisqualität, weil die Prompt Formulierung und die Bewertung der Ergebnisse zwischen Personen variieren können. Genau diese Inkonsistenz in der Prompt-Formulierung und möglicherweise daraus resultierende Heuristiken werden in der aktuellen Literatur als Verbesserung zu reinen LLM-Anwendungen diskutiert.⁹² Während klassische LLM-Anwendungen primär auf einzelne Eingabe-Ausgabe-Interaktionen beschränkt sind, etablieren sich zunehmend Agentensysteme, die im Rahmen von Workflows Prozesse von Anfang bis Ende ausführen können. OpenAI beschreibt Agenten explizit als Systeme, die Aufgaben im Auftrag der Nutzer:innen mit einem hohen Maß an Unabhängigkeit erledigen.⁹³ Der KI-Agent über-

⁸⁶ Vgl. Tversky & Kahnemann (1974), S. 1124 – 1131.

⁸⁷ Vgl. Analge 10 für eine technische Erläuterung.

⁸⁸ Vgl. AWS (o.D.), o.S.

⁸⁹ Vgl. D'Onofrio (2024), S. 333f.

⁹⁰ Vgl. Schrodgers (2025), o.S.

⁹¹ Vgl. Wang et al. (2024), S. 1.

⁹² Vgl. Luo et al. (2025), S. 1f.

⁹³ Vgl. OpenAi (o.D.), o.S.

nimmt demgegenüber die automatisierte Prüfung entlang vordefinierter Parameter und liefert ein konsistent strukturiertes Ergebnis, wodurch Medienbrüche reduziert und die Vergleichbarkeit zwischen Opportunitäten erhöht wird.

Nachfolgendes Prozessdiagramm (Abbildung 1) verdeutlicht den zentralen Unterschied zwischen klassischen, LLM-basierten Lösungen und KI-Agenten im Sinne dieser Arbeit: Während große Sprachmodelle in der Praxis überwiegend als reaktives Werkzeug eingesetzt werden, fungiert ein KI-Agent als autonomer Entscheidungsakteur, der einen vollständigen vorher definierten Workflow eigenständig ausführt.⁹⁴ Die Abbildung 1 kontrastiert die praktische Nutzung großer Sprachmodelle mit einem agentischen Ansatz im Screening von Investmentopportunitäten. Im LLM-basierten Ablauf verbleiben wesentliche Prozessschritte beim Menschen (Sichtung, Überführung in eine Anwendung, Prompt-Formulierung und Interpretation), wodurch Zeitverzug sowie Ergebnisvarianz entstehen können.

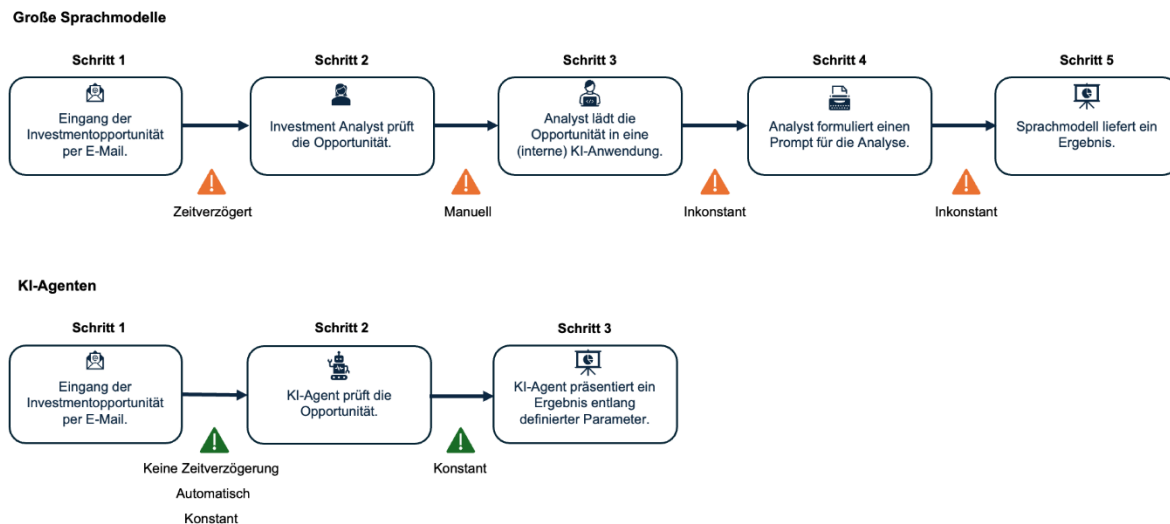


Abbildung 1: Vereinfachte Prozessabbildung zur Anwendung von LLMs und KI-Agenten im Investment Management.⁹⁵

Während Abbildung 1 eine vereinfachte Single-Agent-Struktur darstellt, werden in Forschung und Praxis auch Multi-Agenten-Systeme diskutiert. Masterman et al. (2024) differenzieren Multi-Agenten-Architekturen entlang eines Spektrums in vertikale und horizontale Organisationsformen. In vertikalen Architekturen übernimmt ein Leitagent die Koordination, indem er Teilaufgaben strukturiert und an unterstützende Agenten delegiert, die ihre Ergebnisse an ihn zurückspielen; die Interaktion kann dabei je nach Ausgestaltung ausschließlich über den Leitagenten oder innerhalb einer geteilten Konversation erfolgen. Horizontale Architekturen sind demgegenüber gleichrangig organisiert: Alle Agenten diskutieren die Aufgabe in einem gemeinsamen Kommunikationskanal, sehen die Nachrichten der anderen und

⁹⁴ Vgl. Open AI (o.D.), o.S.

⁹⁵ Eigene Darstellung in Anlehnung an Luo et al. (2025), S. 1 – 8.

können Aufgaben bzw. Tool-Aufrufe auch eigeninitiativ übernehmen, ohne durch einen Leitagenten zugewiesen zu werden.⁹⁶

4.4 KI in der Immobilienwirtschaft

KI hat sich in den vergangenen Jahren von einer Technologie mit eher abstrakter Bedeutung und wenigen Anwendungsfällen zu einem allgemein anerkannten Werkzeug entwickelt. Auch in der Immobilienwirtschaft wird das Thema stark diskutiert.⁹⁷ Während einige Sektoren, wie beispielsweise Finanzdienstleistungen oder Online-Handel bereits früh KI-basierte Anwendungen skaliert haben, gilt die Immobilienbranche tendenziell als technologischer Nachzügler. Das Potenzial wiederum erweist sich als recht hoch, da Entscheidungen stark datenbasiert sind und damit ein hohes Potenzial für KI-Einsatz besteht.⁹⁸ Das Vereinsmagazin *gif im Fokus* betont in seiner Ausgabe „Mensch Immobilie!“, dass sich die Rolle des Menschen in der Branche vor dem Hintergrund technologischer Entwicklungen, vor allem KI, grundlegend wandelt und Immobilienunternehmen ihre Arbeits- und Entscheidungsprozesse neu ausrichten müssen.⁹⁹ Konkrete Anwendungsfelder für KI lassen sich bereits heute in zentralen Funktionsbereichen der Immobilienwirtschaft identifizieren. Besonders weit fortgeschritten ist der Einsatz im Bereich der Immobilienbewertung sowie der Markt- und Standortanalyse. Krämer zeigt im Rahmen des gif-Immobilien-Forschungspreises, dass traditionelle Bewertungsprozesse jedoch nach wie vor stark von manuellen Arbeitsschritten, subjektiven Einschätzungen einzelner Gutachter und damit einhergehenden Verzerrungen geprägt sind.¹⁰⁰

Laut der Trendstudie von Drees & Sommer in Kooperation mit dem IIWM Institut der Technischen Hochschule liegt der digitale Reifegrad für die Immobilienbranche im Jahr 2025 bei einer 3,41 (auf einer Skala von 0 bis 5) und hat sich somit seit dem Jahr 2024 (3,37) um 1,2 % erhöht. Der Wille zur Transformation scheint vorhanden zu sein mit einem Ergebnis von 4,0 bei Transformation und Kultur. Die Aspekte Strategie (3,0) und Digitalisierung (3,22) hinken allerdings hinterher.¹⁰¹ Gleichzeitig erwarten Branchenanalysen, dass sich der Reifegrad dynamisch erhöht: CBRE prognostiziert für den weltweiten Property Technology Markt, getrieben durch KI-Anwendungen, ein Wachstum um rund 70 % auf 32,2 Mrd. US-Dollar bis 2030.¹⁰² Auch McKinsey argumentiert, dass insbesondere generative KI die Art, wie in Immobilien investiert, diese bewirtschaftet und vermarktet werden, grundlegend verändern kann. Hierzu zählen beispielsweise schnellere Deal-Analysen, intelligentere Asset-Selektion oder KI-Unterstützung in Verhandlungen.¹⁰³

⁹⁶ Vgl. Masterman (2024), S. 2 – 3.

⁹⁷ Vgl. Seagraves (2023), S. 190f.

⁹⁸ Vgl. Ochs (2022), S. 10 – 12.

⁹⁹ Vgl. Holzmann & Rock (2025), S. 5 – 10.

¹⁰⁰ Vgl. Krämer (2025), S. 54 – 55.

¹⁰¹ Vgl. Richter et al. (2024), S. 14.

¹⁰² Vgl. CBRE (2025), S. 1 – 3.

¹⁰³ Vgl. Fitzpatrick et al. (2023), o.S.

4.5 Akademische Literatur zu KI in der Immobilienwirtschaft

Aus wissenschaftlicher Perspektive liegt der bisherige KI-Forschungsfokus im Immobilienbereich stark auf Bewertung und Preisprognosen. Zekri analysierte 600 englischsprachige Publikationen aus dem Zeitraum 1988–2024 und zeigt, dass der Großteil der Arbeiten auf Preisprognosen, Bewertung und Marktanalysen entfällt, während Bereiche wie Facility Management, Smart Buildings oder risikoorientierte Entscheidungsmodelle deutlich unterrepräsentiert sind.¹⁰⁴ Auch Lalum (2025) argumentiert, dass die Recherche sich im Immobiliensektor zwar stark ausgeweitet habe aber nach wie vor fragmentiert ist mit einem Fokus auf der Immobilienbewertung.¹⁰⁵

Segura de la Cal et al. (2025) zeigen anhand von Fallstudien, dass KI in der Praxis vor allem zur Automatisierung von Asset-Monitoring, Marketing sowie Nachfrage- und Preisprognosen eingesetzt wird; Anwendungsfelder in Risikobewertung und Investitionsanalyse werden zwar diskutiert, sind aber noch kaum flächendeckend implementiert.¹⁰⁶

Auch Branchenstudien teilen das Bild von hohen Erwartungen aber einer begrenzten Reife. Die ZIA/EY-Digitalisierungsstudie 2024 bezeichnet KI explizit als „Game Changer“ für die Immobilienwirtschaft.¹⁰⁷ McKinsey sowie CBRE kommen zu dem Schluss, dass KI in der Immobilienwirtschaft zwar zunehmend für operative Effizienz (z.B.: im Property Management, Flächen- und Energieoptimierung) eingesetzt wird, aber die strategische Nutzung für Kapitalallokations- und Investitionsentscheidungen noch deutlich hinter dem technisch möglichen Potenzial zurückbleibt. McKinsey geht davon aus, dass generative KI im Bereich des Investment Managements hohes Potential hat, da es den operativen Aufwand reduziert und eine große Menge an Daten verarbeiten kann.¹⁰⁸

5 Methodik

Die Arbeit folgt einem quantitativ-experimentellen, komparativen Design mit ergänzenden Experteninterviews. Im quantitativen Teil werden (i) die Ergebnisse einer Humanbefragung (Zwei-Gruppen-Design) deskriptiv auf Heuristiken ausgewertet und (ii) mehrere LLMs unter identischen Instruktionen und identischem Input auf dieselben Heuristiken getestet. Der Vergleich erfolgt nicht inferenzstatistisch, sondern als heuristikbasiertes Pattern Matching – d.h. theoriegeleitete Erwartungsmuster werden den beobachteten Antwortmustern gegenübergestellt.¹⁰⁹

¹⁰⁴ Vgl. Zekri (2025), S. 8.

¹⁰⁵ Vgl. Lalum (2025), S. 2f.

¹⁰⁶ Vgl. Segura de la Cal et al. (2025), S. 5.

¹⁰⁷ Vgl. Ernst & Young Real Estate GmbH (2024), S. 2f.

¹⁰⁸ Vgl. Fitzpatrick et al. (2023), o.S.

¹⁰⁹ Vgl. Trochim (1989), S. 355 – 366.

Die Methodik gliedert sich in die folgenden Schritte:

- Ableitung des Forschungsbedarfs auf Basis der Literatur und praktischer Problemstellung.
- Konzeption und Durchführung einer experimentell angelegten Humanbefragung im Zwei-Gruppen-Design mit gezielt manipulierten Entscheidungsstimuli (in Form von Teasern) zur Analyse Heuristik-konsistenter Richtungsmuster bei menschlichen Entscheidern.¹¹⁰
- Durchführung identischer Teaser- und Fragebogenlogik über mehrere KI-Modelle unter konstanter Eingabelogik, um modellbezogene Antwortmuster mit Ergebnissen der Humanbefragung vergleichen zu können.
- Ergänzende Experteninterviews zur Kontextualisierung der Befunde sowie zur Ableitung von Gestaltungsprinzipien für einen KI-basierten Entscheidungsagenten.
- Konzeption eines KI-basierten Entscheidungsagenten.

Damit kombiniert die Arbeit Elemente aus der Verhaltensökonomie (Heuristiken) mit einem vergleichenden KI-Benchmarking. Ergänzend erfolgt eine qualitative Kontextualisierung durch Experteninterviews sowie die Konzeption eines KI-Agenten. Diese Vorgehensweise gründet auf dem Konzept des *systematic combining*, bei dem theoretische Konzepte und Empirie kontinuierlich weiterentwickelt werden.¹¹¹ Aufgrund der begrenzten Umfrageteilnehmenden im Versuchsdurchlauf wird auf inferenzstatistische Hypothesentests verzichtet und es erfolgt eine deskriptive Auswertung entlang der Heuristiken aus dem theoretischen Konstrukt dieser Arbeit.¹¹²

5.1 Branchenumfrage

Zur empirischen Adressierung der Forschungsfragen wurde eine standardisierte Online-Umfrage im Zwei-Gruppen-Design konzipiert. Drei realitätsnahe, fiktive Investment-Teaser (Frankfurt am Main) wurden den Teilnehmenden in identischer Reihenfolge vorgelegt. Die Teaser wurden so konstruiert, dass zentrale Kennzahlen (z.B. JNKM, WALT, Leerstand) nur begrenzt variieren, um die Vergleichbarkeit zu erhöhen und beobachtete Unterschiede primär auf die implementierten Heuristik-Manipulationen zurückführen zu können.

5.1.1 Erhebung der Daten

Die Umfrage wurde mit Typeform vom 17.11. bis 31.12.2025 durchgeführt. Die Rekrutierung erfolgte über LinkedIn, Mailings (u.a. GIF-Newsletter, firmeninterne Verteiler, Direktansprache) sowie im Rahmen

¹¹⁰ Vgl. Huijbregts et al. (2021), S. 8 – 11.

¹¹¹ Vgl. Dubois & Gadde (2002), S. 555f.

¹¹² Vgl. Kapitel 2.

eines Mastermoduls an der TH Aschaffenburg. Da die Rekrutierung über offene Kanäle erfolgte, ist die Stichprobe nicht probabilistisch – die Ergebnisse sind daher als explorative Befunde zu interpretieren.

5.1.2 Zwei Gruppen Design

Um zu testen ob Heuristiken und Verzerrungen in der Immobilienwirtschaft erkennbar sind, wurde ein Zwei-Gruppen-Design (*between-subject*) gewählt, siehe auch Abbildung 2. Jeder der beiden Gruppen wird der identische Fragebogen präsentiert, allerdings mit anderen (kontrolliert manipulierten) Teasern. Hierbei wissen die Teilnehmenden nicht, dass es zwei Gruppen gibt, in welcher Gruppe sie sich befinden und haben ebenfalls keine Hintergrundinformationen über die getesteten Heuristiken.¹¹³

Das Umfragedesign gliedert sich in drei Teile. In Teil 1 werden allgemeine Informationen zur Person erfragt (Frage 1 – 9), darunter die Berufserfahrung, das Tätigkeitsfeld in der Branche sowie der geografische Fokus und die Risikobereitschaft der Person. In Teil 2 wurde den Teilnehmenden die Teaser über einen Link in folgender Reihenfolge zur Verfügung gestellt: East Star (Ankerheuristik), Courtyard (Rahmungseffekt) und Primeverse (Repräsentativitätsheuristik). Bei East Star wird Herdenverhalten durch Präsentation einer Mehrheitsinformation getestet. Überkonfidenz wird anhand der Entscheidungssicherheit bei Primeverse sowie einer abschließenden hypothetischen Frage erfasst, wie sicher sich die Teilnehmenden wären, einen Investorenwettbewerb zu gewinnen.

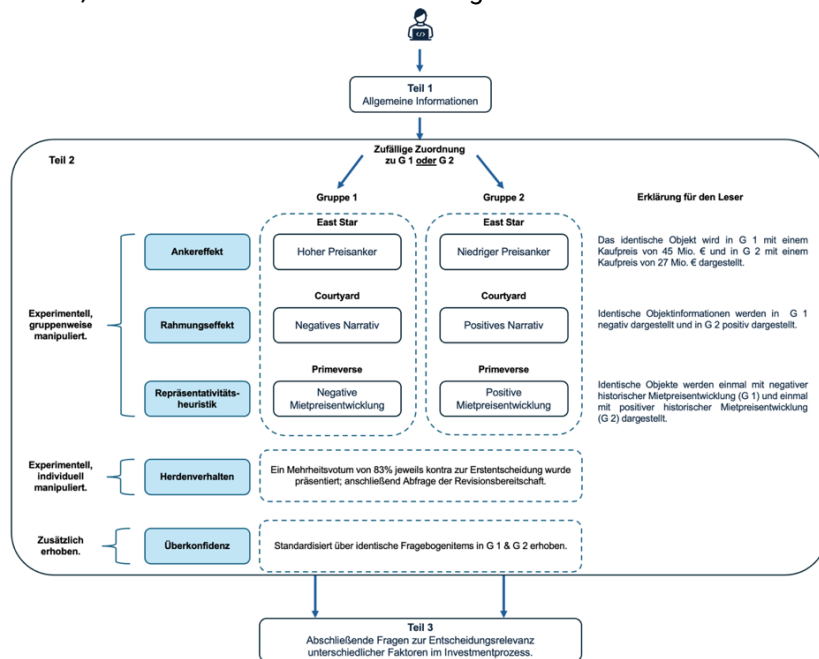


Abbildung 2: Schematische Darstellung des Zwei-Gruppen-Studiendesigns und Versuchsablaufes.¹¹⁴

¹¹³ In Anlehnung an das beschriebene *between subject design* aus Charness et al. (2011), S. 1f.

¹¹⁴ Eigene Darstellung.

Darüber hinaus wurden in Teil 3 Informationen zur Entscheidungsrelevanz unterschiedlicher Faktoren für eine Immobilieninvestition erfragt. Die Fragen 31 – 45 erfassen die subjektiv wahrgenommene Bedeutung einzelner Entscheidungsfaktoren (Mikrolage, Mieterstruktur, Entwicklungspotenzial, ESG, Pricing). Diese abgefragten Entscheidungsdeterminanten wurden abgeleitet aus dem Experteninterview mit Experten aus dem Investmentmanagement. Hier beschreibt er diese Kriterien als relevant in der frühen Opportunitätsprüfung.¹¹⁵ Darüber hinaus wird die genutzte Entscheidungsbasis (Daten vs. Bauchgefühl), die Rolle externer Informationsquellen sowie der eigenen Erfahrung erfasst. Diese Ergebnisse wurden gruppenübergreifend aggregiert ausgewertet.

Die Vorgehensweise, die erwähnten Heuristiken aus Teil 2 zu testen, wird im folgenden Teil erläutert.

5.1.2.1 Ankerheuristik (Anchoring), Herdenverhalten (Herding)

Zur Messung von Ankereffekten wird Teaser 1 (East Star) mit einem hohen Anker (45 Mio. €, 5.221 €/m², 2,85 % Rendite) und Teaser 2 mit einem niedrigen Anker (27 Mio. €, 3.133 €/m², 4,70 % Rendite) präsentiert – bei sonst identischen Objektinformationen. Die unabhängige Variable ist das Niveau des gesetzten Preisankers (hoch vs. niedrig), während alle übrigen objektbezogenen Informationen experimentell kontrolliert werden.

In den Fragen 10–15 werden Investitionsentscheidung und marktgerechte Kaufpreisschätzung erhoben. Die zwischen den Gruppen variierende Kaufpreisnennung soll Ankereffekte sichtbar machen. Anschließend wird – abhängig von der Erstentscheidung – eine aggregierte Mehrheitsmeinung („83 % investieren“ bzw. „83 % investieren nicht“) eingeblendet, um Herdenverhalten zu testen.¹¹⁶

5.1.2.2 Rahmungseffekt (Framing Effekt)

Mit dem Objekt Courtyard wird der Rahmungseffekt untersucht: In Gruppe 1 werden Leerstand, Mieterkonzentration, CapEx und Lage negativ als Risikofaktoren formuliert, in Gruppe 2 dieselben Faktoren positiv als stabile Cashflows, ausgewogener Mietermix, werterhöhende ESG-Investitionen und gute Innenstadtlage dargestellt.

In den Fragen 18–23 werden neben der Investitionsentscheidung eine Kaufpreis- und IRR-Schätzung sowie die wahrgenommene Ton- und Risikowirkung auf einer Likert-Skala erhoben. Die sprachliche Ausgestaltung des Teasers (chancen- vs. risikoorientiert) operationalisiert dabei den Framing-Effekt.¹¹⁷

¹¹⁵ Vgl. Experte A (2025), Interview.

¹¹⁶ In Anlehnung an Lowies et al. (2017), S. 60 – 61.

¹¹⁷ Vgl. Jin & Gallimore (2010), o.S.

5.1.2.3 Repräsentativitätsheuristik (Representativeness)

Beim Objekt Primeverse wird die Repräsentativitätsheuristik über die dargestellte historische Mietentwicklung manipuliert – positiv in Gruppe 1, negativ in Gruppe 2. Geprüft wird, ob Teilnehmende die Vergangenheitsentwicklung auf die Zukunft extrapolieren und das Objekt entsprechend höher oder niedriger bewerten.¹¹⁸ In den Fragen 24–30 werden zudem Kaufpreis und IRR-Einschätzung erhoben sowie ob die Teilnehmenden erwarten, dass sich die Mietpreisentwicklung fortschreiben wird.

5.2 KI-Modelltests

Ergänzend zur Humanbefragung werden acht LLMs unter identischen Bedingungen, mit demselben Teaser, derselben Fragebogenlogik und einem einheitlichen Prompt, getestet. Modellname, Version, Datum und Prompt-Version werden dokumentiert, um Vergleichbarkeit sicherzustellen. Die Auswertung erfolgt heuristikbasiert per Pattern Matching, wobei die Modellantworten den theoretischen Erwartungsmustern sowie den deskriptiven Humanbefunden gegenübergestellt werden.^{119, 120}

5.3 Experteninterviews

Die Arbeit kombiniert Experteninterviews als qualitative Komponente mit dem beschriebenen quantitativen und experimentellen Ansatz um die Untersuchung zu operationalisieren, Befunde zu erklären und Gestaltungsprinzipien für einen KI-Agenten ableiten zu können.¹²¹ Zwei Experteninterviews flankieren die quantitative Erhebung. Das erste Interview (ein Experte aus dem Investment Management, nachfolgend Interview A) diene als explorative Vorphase zur Präzisierung praxisrelevanter Screening-Kriterien und Prozesslogiken, die in die Teaser- und Fragebogenlogik einfließen. Er verfügt über langjährige Erfahrung im institutionellen Investment Management eines international tätigen Immobilienunternehmens. Das zweite Interview (ein Experte für die Interaktion zwischen Menschen und KI, nachfolgend Interview B) wurde nach Auswertung der Humanbefragung und LLM-Durchläufe geführt, um erkannte Verzerrungsmuster zu plausibilisieren und Designentscheidungen für die Agentenarchitektur abzuleiten. Er ist im Bereich Künstliche Intelligenz und Agentensysteme tätig und beschäftigt sich insbesondere mit Mensch-KI-Interaktion. Beide Interviews folgten einem vorbereiteten Leitfaden, der jedoch bewusst Flexibilität für den Gesprächsverlauf ließ.¹²²

Aus Gründen der Vertraulichkeit und auf Wunsch der Interviewpartner erfolgt die Darstellung der Experteninterviews in anonymisierter Form.

¹¹⁸ Vgl. Abdin et al. (2017), S. 677.

¹¹⁹ Vgl. Liang et al. (2023), S. 1 – 2, S. 160 – 161.

¹²⁰ Vgl. Trochim (1989), S. 355 – 366.

¹²¹ Vgl. Dubois & Gadde (2002), S. 557.

¹²² Vgl. Kühl (2009), S. 36ff.

6 Ergebnisse der Branchenumfrage

Im folgenden Kapitel werden die Ergebnisse dieser Branchenumfrage erläutert. Zu Beginn wird die Gruppe 1 analysiert, daraufhin folgt die Gruppe 2. Im Anschluss werden gruppenübergreifende Daten präsentiert, darunter Entscheidungsdeterminanten und die Adäquanz der präsentierten Informationen aus den Teasern. Die gesammelten Daten zeigen einen Querschnittsatz, da die Beobachtungen (menschliche Befragung) jeweils zu einem Erhebungszeitpunkt vorliegen und keine zeitliche Dynamik abbilden.¹²³ Zur strukturierten Ergebnisdarstellung werden die Antworten in Microsoft Excel mittels deskriptiver Statistik zusammengefasst, um Unterschiede zwischen den experimentellen Bedingungen (manipulierte vs. nicht manipulierte Teaser).¹²⁴ Die Datenverarbeitung erfolgte mit Microsoft Excel, ergänzend wurde ChatGPT zur Plausibilisierung von Auswertungslogiken herangezogen; Interpretationen und Darstellungen wurden eigenständig erstellt und manuell gegengeprüft. Für dichotome Variablen – insbesondere Investitionsentscheidung und Entscheidungsänderungen im Herdenverhalten-Test – werden absolute und relative Häufigkeiten berichtet, um Gruppenvergleiche zu vereinfachen.¹²⁵ Bewertungs- und Einschätzungsfragen werden als Häufigkeitsverteilungen sowie über geeignete Lage- und Streuungsmaße zusammengefasst. Für vorhandene intervallskalierte Angaben werden ergänzend arithmetisches Mittel und Standardabweichung ausgewiesen.¹²⁶ Da Verteilungen in entscheidungsbezogenen Einschätzungen durch Ausreißer oder Schiefe beeinflusst sein können, wird der Median (MDN) als robuster Lageparameter ergänzend berichtet, weil er gegenüber Extremwerten unempfindlicher ist.¹²⁷

6.1 Ergebnisse der Gruppe 1

In der ersten Gruppe wurde die Umfrage insgesamt 117-mal angesehen, 83-mal aktiv gestartet und 20-mal abgeschlossen. Die entspricht einer Abschlussrate von 24%. Es lagen 20 auswertbare Datensätze vor (Studierende: $n = 11$; Experten: $n = 9$). Bei den Kaufpreisangaben wurden offenkundige Platzhalterwerte (Werte $<5,0$ Mio. €, $>75,0$ Mio. €) in der Auswertung exkludiert und als fehlend interpretiert. Dadurch reduzierte sich die Anzahl gültiger Fälle für die Kaufpreisvariablen auf $n = 16$ (East Star) bzw. $n = 18$ (Courtyard); für Primeverse lagen $n = 19$ gültige Werte vor.

6.1.1 Struktur der Teilnehmenden

Die Stichprobe war überwiegend männlich (70,0 %, $n=14$). Unter den 20 Teilnehmenden entfielen 55,0 % auf Studierende ($n=11$); die Expertenrollen verteilten sich auf Gutachter, Makler und Finanzierung/Bank (je 10,0 %) sowie Investment Manager, Asset Manager und Sonstiges (je 5,0 %). Die Berufserfahrung

¹²³ Vgl. Wooldridge (2009), S. 5.

¹²⁴ Vgl. Bourier (2022), S. 37 – 39.

¹²⁵ Ebd., S. 38 – 39.

¹²⁶ Ebd., S. 78 sowie S. 96.

¹²⁷ Ebd., S. 74.

konzentrierte sich auf 0–5 Jahre (70,0 %), erfahrenere Teilnehmende waren selten. Bei der Deal-Erfahrung dominierten <10 Deals (65,0 %), während >149 Deals mit 20,0 % überproportional vertreten waren. Asset-seitig überwogen Büro und Wohnen (je 35,0 %), regional Frankfurt am Main (40,0 %), DACH und übriges Deutschland (je 20,0 %).

6.1.2 Teaser East Star: Investitionsentscheidung, Entscheidungssicherheit, Herdenverhalten, Kaufpreis & IRR

Auf Basis des Teasers „East Star“ (mit dem hohen Kaufpreisanker von 45,0 Mio. €) entschieden sich 55,0% der 20 Teilnehmenden für ein Investment (Ja: n = 11; 55,0%) und 45,0% dagegen (Nein: n = 9; 45,0%). Die Entscheidungssicherheit (0 = sehr unsicher, 10 = sehr sicher) lag insgesamt über dem Mittelpunkt und damit leicht erhöht bei M = 6,75 (SD = 1,97; MDN = 7,00; Min = 1,00; Max = 10,00; n = 20).

Die Einschätzung des Kaufpreises wies (nach Bereinigung von Platzhalterwerten) n = 16 gültige Angaben auf. Der geschätzte Kaufpreis (in Mio. €) lag bei M = 34,6 (SD = 5,8; MDN = 35,00; Min = 20,0; Max = 43,00). Wie Abbildung 3 zeigt, liegen die zentralen Kaufpreisschätzungen in allen Teilstichproben deutlich oberhalb des marktüblichen Kaufpreises (27,27 Mio. €) und näher am vorgegebenen Preisanker (45,00 Mio. €). Die Boxplots verdeutlichen zudem eine linkssteile Verteilung (negative Schiefe) mit wenigen niedrigen Ausreißern (Schiefe -1,183). Entsprechend ist der Median als robustes Lagemaß besonders aussagekräftig (MDN = 35,00 Mio. €).

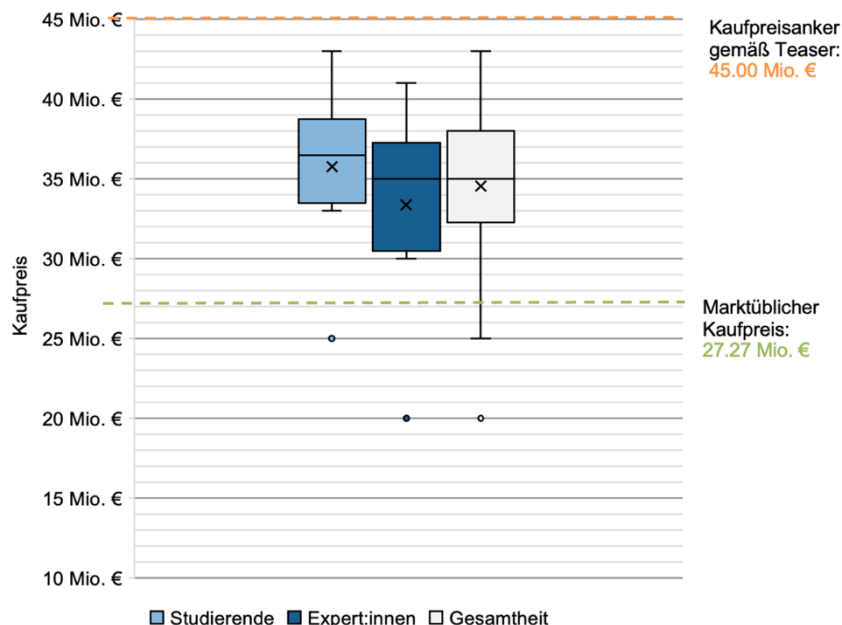


Abbildung 3: Boxplot der Kaufpreisnennungen für "East Star" in Gruppe 1 (hoher Preisanker). Expert:innen und Gesamtheit (aggregiert).¹²⁸ Orange Linie = Preisanker aus dem Teaser (45,00 Mio. €), grün gestrichelte Linie = marktüblicher Kaufpreis (27,27 Mio. €). Horizontale Linie in der Box = Median, x = Mittelwert, Punkte = Ausreißer.

¹²⁸ Eigene Darstellung.

Eine Differenzierung nach Erfahrungsstufe konkretisiert dieses Muster: Bei den Studierenden ($n = 8$) lagen die geschätzten Kaufpreise im Median bei rund 36,5 Mio. €. Der arithmetische Mittelwert betrug $M = 35,8$ Mio. € bei einer Standardabweichung von $SD = 5,31$ Mio. €, was auf eine moderate Streuung bei der Preisnennungen hinweist. Der gegenüber Ausreisern robustere Median lag mit $MDN = 36,5$ Mio. € leicht oberhalb des Mittelwerts. In Verbindung mit der negativen Schiefe von $-1,050$ deutet dies auf eine Häufung höherer Kaufpreisangaben bei gleichzeitig wenigen, vergleichsweise niedrigen Nennungen hin, die den Mittelwert nach unten beeinflussen. Bei den Expert:innen ($n = 8$) lag der geschätzte Kaufpreis im Mittel bei $M = 33,4$ Mio. € ($SD = 6,35$ Mio. €). Der Median betrug $MDN = 35,0$ Mio. € und lag deutlich oberhalb des Mittelwerts. In Verbindung mit der negativen Schiefe von $-1,398$ weist dies auf eine Konzentration höherer Kaufpreisnennungen bei wenigen, deutlich niedrigeren Angaben hin, die den Mittelwert nach unten ziehen.

Die erwartete Internal Rate of Return (IRR) wurde als Prozentwert angegeben und lag im Mittel bei $M = 4,77\%$ ($SD = 2,46$; $MDN = 5,00\%$; $Min = 0,13\%$; $Max = 10,00\%$; $n = 20$). Die Verteilung war hierbei annähernd symmetrisch ($Skew = -0,11$).

Im Anschluss an die East Star Befragung wurde eine Mehrheitsinformation präsentiert, um auf mögliches Herdenverhalten zu testen („83% investieren (nicht)“) und abgefragt, ob die eigene Aussage angepasst würde. Die Antworten waren überwiegend stabil: Nein: $n = 18$; 90,0%, Ja: $n = 1$; 5,0%, Unsicher: $n = 1$; 5,0% ($n = 20$). Rein deskriptiv zeigte sich die einzige Anpassungsnennung in der Studierendengruppe (Studierende: Ja $n = 1$; Unsicher $n = 1$; Experten: ausschließlich Nein $n = 9$).

6.1.3 Teaser „Courtyard“: Investitionsentscheidung, Kaufpreis & IRR sowie Tonwirkung und Risikowahrnehmung

In Gruppe 1 für „Courtyard“ (negative Rahmung) entschied sich die Mehrheit der Teilnehmenden für eine Investition: Ja: $n = 14$; 70,0%, Nein: $n = 6$; 30,0% ($n = 20$). Die Tonwirkung des Textes wurde auf einer Skala von 0 (= negativ) bis 10 (= positiv; 5 = neutral) insgesamt eher negativ eingeschätzt ($M = 2,95$; $SD = 2,04$; $MDN = 2,50$; $Min = 0,00$; $Max = 6,00$; $n = 20$). Der Median von 2,50 deutet auf eine negative Wahrnehmung des Textes unter den Teilnehmenden hin. Die Risikowahrnehmung (0 = sehr geringes Risiko, 5 = mittleres Risiko, 10 = sehr hohes Risiko) lag im mittleren bis erhöhten Bereich bei $M = 6,05$ ($SD = 1,76$; $MDN = 6,00$; $Min = 3,00$; $Max = 10,00$; $n = 20$).

Für den Kaufpreis (bereinigt um Platzhalterwerte) lagen $n = 18$ gültige Werte vor und wurde im Durchschnitt mit 26,58 Mio. € ($SD = 8,77$; $MDN = 26,00$; $Min = 10,00$; $Max = 50,00$) angegeben. Die erwartete IRR – ebenfalls bereinigt – lag bei $M = 6,20\%$ ($SD = 1,74$; $MDN = 6,00\%$; $Min = 3,5\%$; $Max = 10,00\%$; $n = 18$).

Die nach Investitionsentscheidung und Erfahrungsstufe aufgeschlüsselten Kennwerte zur Tonwirkung und Risikowahrnehmung sind in Tabelle 2 zusammengefasst. Zunächst zeigt die Aufteilung nach Investitionsentscheidung ein konsistentes Muster hinsichtlich der Risikowahrnehmung (vgl. Tabelle 2): Teilnehmende, die ein Investment ablehnten (n = 6), berichteten eine deutlich höhere Risikowahrnehmung als Investitionsbefürworter:innen (n = 14). Während die Ja-Gruppe im Mittel eine Risikowahrnehmung von M = 5,57 (MDN = 5,50; SD = 1,45) angab, lag die Nein-Gruppe bei M = 7,17 (MDN = 7,00; SD = 2,04). Dieses Ergebnis spricht deskriptiv dafür, dass eine erhöhte Risikoeinschätzung mit einer ablehnenden Investitionsentscheidung einhergeht. Demgegenüber unterscheiden sich die Tonbewertungen zwischen Ja und Nein nur begrenzt: Zwar fällt der Mittelwert in der Ja-Gruppe höher aus (M = 3,14; SD = 1,99) als in der Nein-Gruppe (M = 2,50; SD = 2,26), jedoch ist der Median in beiden Gruppen identisch (MDN = 2,50). Damit zeigt sich, dass die Teilnehmenden eine ähnliche Tonwahrnehmung haben aber dennoch in der Risikoeinschätzung bedeutend abweichen.

| | | Tonwirkung (0 = negativ, 10 = positiv) | | | Risikowahrnehmung (0 = sehr gering, 10 = sehr hoch) | | |
|--------------------------|-----------------------|---|------|------|--|------|------|
| | | M | MDN | SD | M | MDN | SD |
| Investitionsentscheidung | Ja n = 14 | 3,14 | 2,50 | 1,99 | 5,57 | 5,50 | 1,45 |
| | Nein n = 6 | 2,50 | 2,50 | 2,26 | 7,17 | 7,00 | 2,04 |
| Erfahrungsstufe | Studierende n = 11 | 3,00 | 3,00 | 2,32 | 5,36 | 5,00 | 0,81 |
| | Expert:innen n = 9 | 2,89 | 2,00 | 1,76 | 6,89 | 7,00 | 2,26 |

Hinweis: Basis N = 20 gültige Angaben; Teilgruppenbildung nach Filterkriterium.

Tabelle 2: Deskriptive Kennwerte zur Tonwirkung und Risikowahrnehmung für "Courtyard" in Gruppe 1 (negatives Framing).¹²⁹

Auch die Differenzierung nach Erfahrungsstufe legt eine systematische Heterogenität insbesondere bei der Risikowahrnehmung nahe. Studierende (n = 11) bewerteten das Risiko im Mittel niedriger (M = 5,36; MDN = 5,00; SD = 0,81) als Expert:innen (n = 9), die eine höhere Risikowahrnehmung berichteten (M = 6,89; MDN = 7,00; SD = 2,26). Dies kann deskriptiv als Hinweis auf eine unterschiedlich ausgeprägte Risikokalibrierung in Abhängigkeit vom Erfahrungsniveau interpretiert werden. Die Tonwirkung wurde in beiden Gruppen überwiegend negativ eingeschätzt, wobei Studierende tendenziell den Text leicht positiver wahrnahmen (M = 3,00; MDN = 3,00; SD = 2,32) als Expert:innen (M = 2,89; MDN = 2,00; SD = 1,76). Insgesamt zeigt die Analyse, dass Unterschiede in der Risikowahrnehmung deutlicher ausfallen als Unterschiede in der Tonwahrnehmung und zudem ein relevanter Unterschied nach Erfahrungsstufe festgestellt werden kann (vgl. Tabelle 2).

¹²⁹ Eigene Darstellung.

6.1.4 Teaser „Primeverse“: Investitionsentscheidung, Kaufpreis & IRR sowie Mietpreiserwartung und Risikowahrnehmung

Für „Primeverse“ lag die Investitionszustimmung am höchsten: Ja: n = 16; 80,0%, Nein: n = 4; 20,0% (n = 20). Die Einschätzung, ob sich die historisch negative Entwicklung der Mietpreise auf Objektebene in den nächsten Jahren fortsetzen werde, war exakt ausgeglichen (Ja: n = 10; 50,0%; Nein: n = 10; 50,0%). Eine Differenzierung nach Erfahrungsstufe zeigt jedoch ein gegenläufiges Muster (vgl. Abbildung 4): Studierende erwarteten häufiger eine Fortsetzung der negativen Mietpreisentwicklung (Ja: n = 7; 63,6%), während Expert:innen überwiegend keine Fortsetzung annahmen (Nein: n = 6; 66,7%).

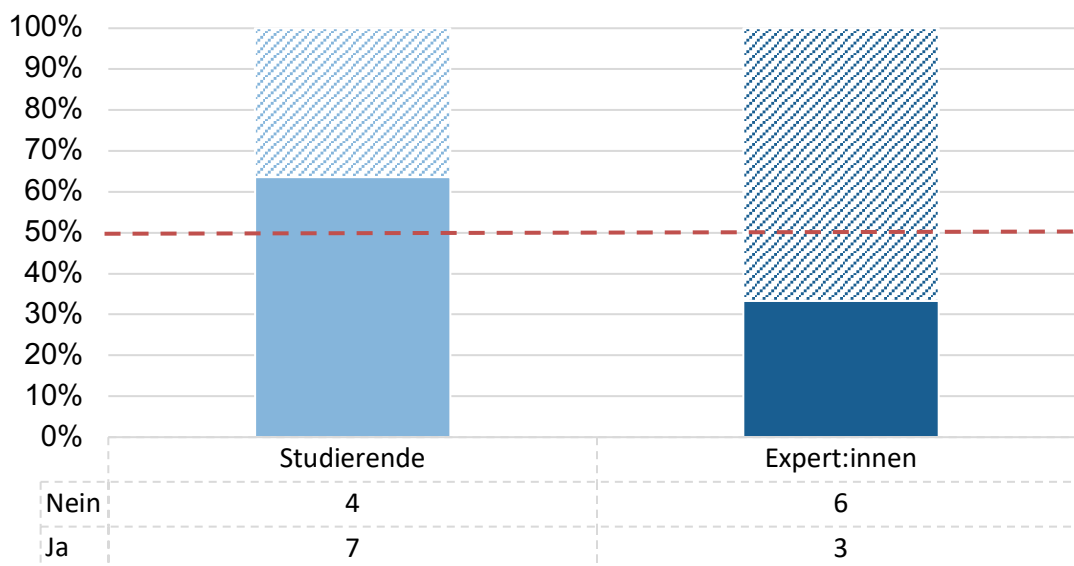


Abbildung 4: Anteil der erwarteten Fortsetzung sinkender Mietpreise getrennt nach Erfahrungsniveau. Die gestrichelte Linie markiert die Entscheidung der Gesamtheit: 50% „Ja“, 50% „Nein“.¹³⁰

Auch die Investitionsentscheidung in beiden Antwortgruppen fiel identisch aus: Sowohl bei Teilnehmenden, die eine Fortsetzung des Mietrückgangs erwarteten, als auch bei jenen, die dies verneinten, entschieden sich jeweils 80,0% für eine Investition (Ja: n = 8; Nein: n = 2; jeweils n = 10).

Trotz gleicher Investitionsquote zeigten sich deskriptive Unterschiede in den Bewertungskomponenten. Teilnehmende mit Erwartung einer Trendfortsetzung setzten im Mittel einen geringfügig niedrigeren Kaufpreis an (M = 30,45 Mio. €, SD = 5,65 Mio. €, MDN = 29,50 Mio. €, Min = 25,00 Mio. €, Max = 44,50 Mio. €, n = 10) als Teilnehmende ohne Trendfortsetzungserwartung (M = 31,61 Mio. €, SD = 7,73 Mio. €, MDN = 30,00 Mio. €, Min = 21,00 Mio. €, Max = 50,00 Mio. €, n = 9; eine Angabe fehlte). Gleichzeitig lagen die erwarteten Renditen bei Trendfortsetzung höher (IRR: M = 5,95%, SD = 2,22, MDN = 5,75, Min = 3,00, Max = 10,00, n = 10) als bei fehlender Trendfortsetzungserwartung (IRR: M = 4,86%, SD = 2,34, MDN =

¹³⁰ Eigene Darstellung.

Will, Rock

4,88, Min = 0,11, Max = 7,50, n = 10). Die Risikowahrnehmung war in der Trendfortsetzungsgruppe ebenfalls leicht erhöht (M = 5,30, SD = 1,89, MDN = 5,00, Min = 3, Max = 8, n = 10) gegenüber der Gruppe ohne Trendfortsetzung (M = 4,80, SD = 2,30, MDN = 5,00, Min = 0, Max = 8, n = 10).

6.1.5 Entscheidungsmodus, Rolle der Berufserfahrung, Risikoprofil

In Teil drei der Umfrage wurde der Entscheidungsmodus auf einer Skala von 0 (= Intuition) bis 10 (= Analyse der vorliegenden Daten) erhoben und lag in Gruppe 1 insgesamt bei M = 4,60 (SD = 3,49; MDN = 4,00; Min = 0,00; Max = 10,00; n = 20). Rein deskriptiv zeigte sich hierbei ein deutlicher Gruppenunterschied: Die Expertengruppe befand sich stärker auf der Analyse-Seite (M = 7,11; SD = 2,20; MDN = 8,00; n = 9), Studierende eher auf der Intuitions-Seite (M = 2,55; SD = 2,94; MDN = 2,00; n = 11). Die explizit abgefragte Grundlage der Entscheidungen wurde überwiegend als Bauchgefühl angegeben (n = 11; 55,0%), gefolgt von Zahlen/Daten/Fakten (n = 5; 25,0%) und Erfahrungswerten (n = 4; 20,0%). Deskriptiv wurde bei Expert:innen häufiger „Zahlen/Daten/Fakten“ genannt (n = 4; 44,4%), während Studierende überwiegend „Bauchgefühl“ auswählten (n = 8; 72,7%).

Die persönlich eingeschätzte Bedeutung der Berufserfahrung für die Entscheidung (0 = kein Einfluss, 10 = sehr hoher Einfluss) lag im Mittel bei M = 5,15 (SD = 3,72; MDN = 5,50; Min = 0,00; Max = 10,00; n = 20). Die Standardabweichung von 4 veranschaulicht, dass die Antworten stark verteilt waren. Die persönliche Risikobereitschaft lag insgesamt im mittleren Bereich (0 = risikoavers, 10 = risikofreudig) mit M = 5,40 (SD = 2,37; MDN = 6,00; Min = 0,00; Max = 9,00; n = 20). Die Verteilung zeigte eine moderate Linkssteilheit (Schiefe = -0,82), was auf eine tendenzielle Häufung höherer Werte hindeutet. Die Risikobereitschaft im Beruf fiel sehr ähnlich aus (M = 5,60; SD = 2,33; MDN = 6,00; Min = 1,00; Max = 10,00; n = 20). Rein deskriptiv lagen die Mittelwerte in der Expertengruppe etwas höher als bei Studierenden (persönlich: Expert:innen M = 5,89 vs. Studierende M = 5,00; beruflich: Expert:innen M = 6,33 vs. Studierende M = 5,00), ohne dass daraus inferenzielle Schlüsse abgeleitet werden.

6.1.6 Zusammenfassende Darstellung

Gruppe 1 umfasst 20 auswertbare Datensätze (Studierende: n=11, Experten: n=9) bei einer Abschlussrate von 24 %. Nach Ausschluss von Platzhalterwerten variieren die gültigen Fallzahlen je Objekt (East Star: n=16, Courtyard: n=18, Primeverse: n=19).

Bei East Star fiel die Investitionsentscheidung nahezu ausgeglichen aus (55 % Ja). Der Median der Kaufpreisschätzungen lag mit 35,00 Mio. € deutlich über dem Referenzwert (27,27 Mio. €), aber merklich unter dem gesetzten Anker (45,0 Mio. €). Die erwartete IRR betrug im Mittel 4,77 % bei hoher Streuung. 90 % der Teilnehmenden änderten ihre Entscheidung trotz Mehrheitsinformation nicht.

Bei Courtyard investierten 70 % der Teilnehmenden. Die Tonwirkung wurde im Mittel negativ eingeschätzt, die Risikowahrnehmung lag im mittleren bis erhöhten Bereich. Ablehner zeigten deskriptiv höhere Risikowahrnehmungswerte als Befürworter.

Bei Primeverse war die Zustimmung am höchsten (80 % Ja). Die Einschätzung zur Fortsetzung sinkender Mieten war ausgeglichen (50/50) und unterschied sich nach Erfahrungsniveau, ohne jedoch die Investitionsquote zu beeinflussen.

Übergreifend ist bemerkenswert, dass die Investitionsquoten trotz vergleichbarer Kennzahlen zwischen den Objekten deutlich variierten (55 % / 70 % / 80 %), was auf den Einfluss nicht-quantitativer Wahrnehmungskomponenten hindeutet. Experten berichteten zudem häufiger ein daten- und analyseorientiertes Vorgehen, Studierende eher intuitive Entscheidungen. Gruppe 1 liefert damit ein konsistentes deskriptives Ausgangsbild für den Vergleich mit Gruppe 2.

6.2 Ergebnisse der Gruppe 2

In der zweiten Gruppe wurde die Umfrage insgesamt 83-mal angesehen, 74-mal aktiv gestartet und 18-mal abgeschlossen. Dies entspricht einer Abschlussrate – ähnlich zu der Gruppe 1 – von 24%. 18 auswertbare Datensätze lagen vor, bestehend aus Experten (n = 12) und Studierenden (n = 6). Bei den Kaufpreisangaben wurden offenkundige Platzhalterwerte (Werte <5,0 Mio. €, >75,0 Mio. €) als fehlend interpretiert.

6.2.1 Struktur der Teilnehmenden

Die Stichprobe war überwiegend männlich (72,2 %, n=13). Unter den 18 gültigen Rollenangaben entfielen 33,3 % auf Studierende (n=6); die Expertenrollen verteilten sich auf Asset Manager (16,7 %), Investment Manager, Fund Manager und Makler (je 11,1 %) sowie Gutachter und Sonstiges. Die Berufserfahrung konzentrierte sich stark auf 0–5 Jahre (77,8 %), erfahrenere Teilnehmende (15–30 Jahre) waren selten. Bei der Deal-Erfahrung dominierten <10 Deals (44,4 %), gefolgt von 10–49 Deals (33,3 %). Asset-seitig überwogen Büro (38,9 %) und Wohnen (27,8 %), regional Frankfurt am Main (44,4 %) und DACH (22,2 %).

6.2.2 Teaser „East Star“: Investitionsentscheidung, Entscheidungssicherheit, Herdenverhalten, Kaufpreis & IRR

Auf Basis des Teasers „East Star“ (Gruppe 2; niedriger Kaufpreisanker) entschieden sich 35,0% der Teilnehmenden für ein Investment (Ja: $n = 7$; 35,0%) und 55,0% dagegen (Nein: $n = 11$; 55,0%). Für $n = 18$ gültige Entscheidungen ergeben sich damit gültige Anteile von 38,9% (Ja) und 61,1% (Nein). Die Entscheidungssicherheit (0 = sehr unsicher, 10 = sehr sicher) lag bei den gültigen Angaben insgesamt im mittleren bis höheren Bereich ($M = 6,33$; $SD = 2,66$; $MDN = 7,50$; $Min = 2,00$; $Max = 10,00$; $n = 18$). Die Einschätzung des Kaufpreises wies nach Bereinigung der Ausreißer $n = 16$ gültige Nennungen auf.

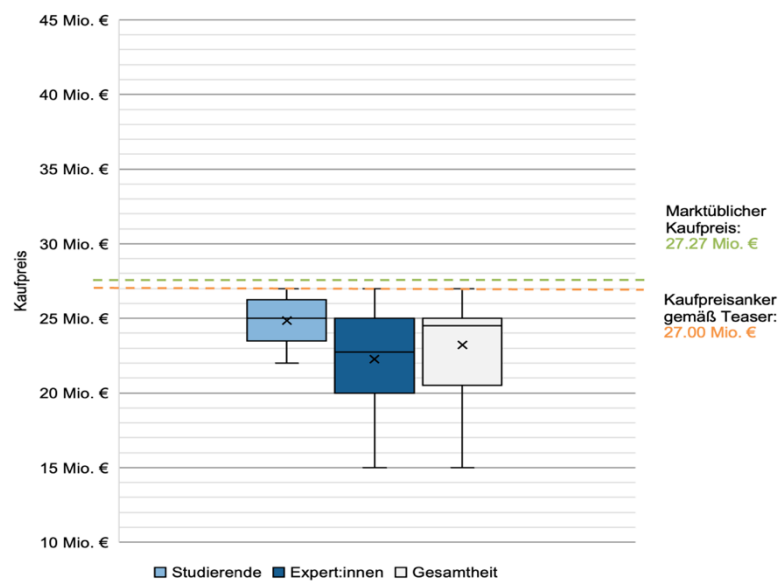


Abbildung 5: Boxplot der Kaufpreisenennungen für "East Star" in Gruppe 2 (niedriger Preisanker).¹³¹ Orange Linie = Preisanker aus dem Teaser (27,00 Mio. €), grün gestrichelte Linie = marktüblicher Kaufpreis (27,27 Mio. €). Horizontale Linie in der Box = Median, x = Mittelwert, Punkte = Ausreißer.

Wie Abbildung 5 zeigt, liegen die zentralen Kaufpreiseinschätzungen in Gruppe 2 unterhalb des gesetzten Ankers sowie des marktüblichen Kaufpreises. Der geschätzte Kaufpreis (in Mio. €) lag bei $M = 23,22$ ($SD = 3,20$; $MDN = 24,50$; $Min = 15,00$; $Max = 27,00$; $n = 16$).

Die Boxplots deuten zudem auf eine linkssteile Verteilung hin ($Schiefe = -1,162$), mit wenigen deutlich niedrigeren Nennungen. Der Median dient als robusteres Lagemaß für die typische Kaufpreiseinschätzung ($MDN = 24,50$ Mio. €).

Eine Differenzierung nach Erfahrungsstufe konkretisiert dieses Muster: Bei den Studierenden ($n = 6$) lagen die geschätzten Kaufpreise im Median bei etwa 25,0 Mio. €. Der arithmetische Mittelwert betrug $M = 24,83$ Mio. € bei einer Standardabweichung von $SD = 1,72$ Mio. €, was auf eine eher geringe bis moderate Streuung der Preisnennungen hindeutet. In Verbindung mit einer negativen Schiefe von $-0,678$

¹³¹ Eigene Darstellung.

spricht dies für eine tendenzielle Häufung höherer Nennungen bei einzelnen niedrigeren Werten, die den Mittelwert leicht nach unten beeinflussen. Bei den Expert:innen (n = 10) lag der geschätzte Kaufpreis im Mittel bei M = 22,25 Mio. € (SD = 3,55 Mio. €). Der Median betrug MDN = 22,75 Mio. € und lag damit leicht oberhalb des Mittelwerts. Abbildung 5 veranschaulicht die Unterschiede in der Kaufpreisenennung. Die ebenfalls negative Schiefe (Schiefe = -0,759) deutet auf eine asymmetrische Verteilung mit wenigen niedrigen Nennungen hin.

6.2.3 Teaser „Courtyard“: Investitionsentscheidung, Kaufpreis & IRR sowie Tonwirkung und Risikowahrnehmung

In Gruppe 2 fiel für „Courtyard“ (positives Framing) die Investitionsentscheidung häufiger positiv aus: Ja: n = 10; 55,6%, Nein: n = 8; 44,4% (n = 18). Die Tonwirkung des Textes wurde auf einer Skala von 0 (= negativ) bis 10 (= positiv; 5 = neutral) insgesamt eher positiv eingeschätzt (M = 5,78; SD = 1,96; MDN = 5,50; Min = 2,0; Max = 9,00; n = 18). Der Median von 5,5 deutet auf eine leicht positive Wahrnehmung des Textes unter den Teilnehmenden hin. Die Risikowahrnehmung (0 = sehr geringes Risiko, 5 = mittleres Risiko, 10 = sehr hohes Risiko) lag im mittleren Bereich bei M = 5,11 (SD = 1,75; MDN = 5,00; Min = 3,00; Max = 8,00; n = 18). Die nach Investitionsentscheidung und Erfahrungsstufe aufgeschlüsselten Kennwerte zur Tonwirkung und Risikowahrnehmung sind in Tabelle 3 zusammengefasst.

Nach Bereinigung der Kaufpreisdaten verbleiben 16 gültige Eingabe, welche mit einer Schiefe von 2,36 leicht schief verteilt sind. Der Mittelwert mit M = 26,46 Mio. € lag (SD = 5,68), betrug der Median nur MDN = 26,00 Mio. € (Min = 19,00; Max = 45,00; n = 16). Die erwartete IRR lag bei M = 7,39% (SD = 2,04; MDN = 7,75%; Min = 4,00%; Max = 11,00%; n = 18) mit unauffälliger Verteilung (Skew = -0,06).

| | | Tonwirkung (0 = negativ, 10 = positiv) | | | Risikowahrnehmung (0 = sehr gering, 10 = sehr hoch) | | | |
|-------------------------------|--------------|---|------|------|--|------|------|------|
| | | M | MDN | SD | M | MDN | SD | |
| Investitions- entscheidung | Ja | n = 10 | 5,80 | 6,00 | 2,57 | 4,30 | 4,00 | 1,34 |
| | Nein | n = 8 | 5,75 | 5,50 | 0,89 | 6,13 | 6,50 | 1,73 |
| Erfahrungs- stufe | Studierende | n = 6 | 4,50 | 4,50 | 2,59 | 5,33 | 5,50 | 1,75 |
| | Expert:innen | n = 12 | 6,42 | 6,50 | 1,24 | 5,00 | 5,00 | 1,81 |

Hinweis: Basis N = 18 gültige Angaben; Teilgruppenbildung nach Filterkriterium.

Tabelle 3: Deskriptive Kennwerte zur Tonwirkung und Risikowahrnehmung für "Courtyard" in Gruppe 2 (positives Framing)¹³²

¹³² Eigene Darstellung.

Die Aufteilung nach Investitionsentscheidung zeigt dabei ein klares Muster in der Risikowahrnehmung (vgl. Tabelle 3): Teilnehmende, die ein Investment ablehnten (Nein; $n = 8$), berichteten eine deutlich höhere Risikowahrnehmung als Investitionsbefürworter:innen (Ja; $n = 10$). Während die Ja-Gruppe im Mittel eine Risikowahrnehmung von $M = 4,30$ ($MDN = 4,00$; $SD = 1,34$) angab, lag die Nein-Gruppe bei $M = 6,13$ ($MDN = 6,50$; $SD = 1,73$). Demgegenüber unterscheiden sich die Tonbewertungen zwischen Ja und Nein nur gering: Die Tonwirkung liegt in beiden Gruppen im neutral bis leicht positiven Bereich (Ja: $M = 5,80$; $MDN = 6,00$; $SD = 2,57$; Nein: $M = 5,75$; $MDN = 5,50$; $SD = 0,89$), sodass sich die Ablehnung eines Investments in dieser Bedingung deskriptiv primär über eine erhöhte Risikowahrnehmung, nicht jedoch über eine deutlich negativere Tonwahrnehmung erklären lässt.

Auch die Differenzierung nach Erfahrungsstufe weist auf Unterschiede in der Tonwahrnehmung hin. Studierende ($n = 6$) bewerteten die Tonwirkung im Mittel niedriger ($M = 4,50$; $MDN = 4,50$; $SD = 2,59$) als Expert:innen ($n = 12$), die den Ton deutlich positiver einschätzten ($M = 6,42$; $MDN = 6,50$; $SD = 1,24$). Die Risikowahrnehmung fällt hingegen in beiden Erfahrungsgruppen vergleichbar aus (Studierende: $M = 5,33$; $MDN = 5,50$; $SD = 1,75$; Expert:innen: $M = 5,00$; $MDN = 5,00$; $SD = 1,81$). Insgesamt deuten die Teilgruppenanalysen damit darauf hin, dass sich innerhalb der positiv gerahmten Bedingung Unterschiede in der Investitionsentscheidung vor allem in der Risikowahrnehmung widerspiegeln, während Unterschiede nach Erfahrungsstufe stärker die Tonwahrnehmung betreffen.

6.2.4 Teaser „Primeverse“: Investitionsentscheidung, Kaufpreis & IRR sowie Mietpreiserwartung und Risikowahrnehmung

Bei dem Objekt „Primeverse“ (historisch positive Mietpreisentwicklung im Objekt) stimmten 55,6% einer Investition zu (Ja: $n = 10$), 44,4% lehnten ab (Nein: $n = 8$). Die Erwartung, dass die Mietpreisentwicklung fortgesetzt wird, wurde mehrheitlich bejaht (Ja: $n = 11$; 61,1%; Nein: $n = 7$; 38,9%; $n = 18$). Abbildung 6 veranschaulicht das deskriptive Muster: 50,0% der Studierenden stimmen zu, dass sich die Mietpreisentwicklung so fortsetzen wird, während es bei den Expert:innen 66,7% sind. Die Risikowahrnehmung lag bei „Primeverse“ näher am Skalenmittelpunkt ($M = 5,11$; $SD = 2,05$; $MDN = 5,50$; $Min = 0,00$; $Max = 8,00$; $n = 18$). In Gruppe 2 (positive Mietpreisentwicklung) erwartete die Mehrheit der Teilnehmenden eine Fortsetzung der Mietentwicklung (Ja: $n = 11$; 61,1%), während 38,9% von keiner Fortsetzung ausgingen (Nein: $n = 7$). Die Investitionsentscheidung fiel insgesamt in 55,6% der Fälle positiv aus (Ja: $n = 10$; Nein: $n = 8$; $n = 18$). Deskriptiv zeigte sich hierbei kein relevanter Unterschied in der Investitionsquote zwischen Teilnehmenden mit Trendfortsetzungserwartung (Invest Ja: 6/11; 54,5%) und Teilnehmenden ohne Trendfortsetzungserwartung (Invest Ja: 4/7; 57,1%).

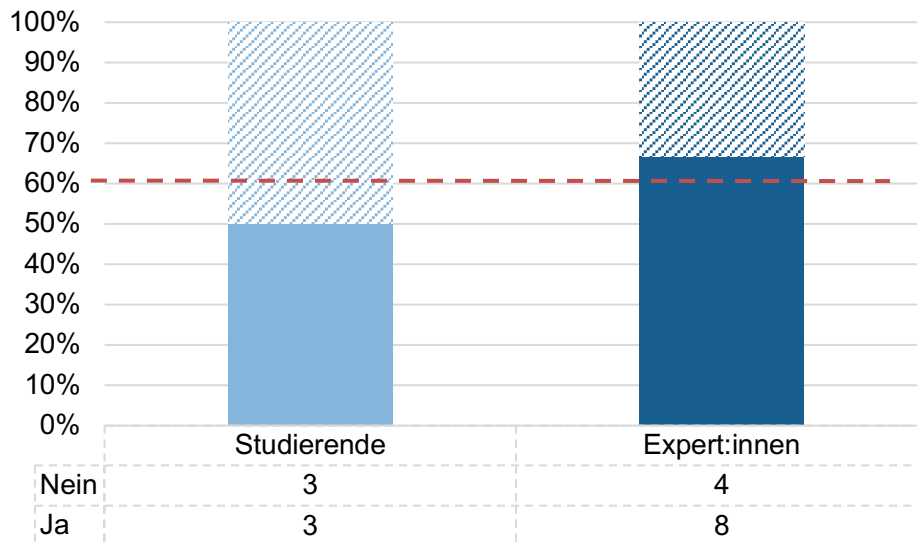


Abbildung 6: Anteil der erwarteten Fortsetzung steigender Mietpreise getrennt nach Erfahrungsniveau. Die gestrichelte Linie markiert den Gesamtanteil „Ja“ = 61,1%.¹³³

Unterschiede zeigten sich eher in den Bewertungskomponenten: In der Trendfortsetzungsgruppe lag der durchschnittlich angesetzte Kaufpreis höher ($M = 31,27$ Mio. €) als in der Vergleichsgruppe ($M = 27,64$ Mio. €), wobei dieser Unterschied maßgeblich durch einen Ausreißer ($Max = 65,00$ Mio. €) beeinflusst wird und die Mediane nahezu identisch ausfielen ($28,50$ vs. $28,00$ Mio. €). Die erwartete IRR war bei Trendfortsetzung leicht erhöht ($M = 7,67\%$ vs. $7,15\%$). Die Objektwahrnehmung fiel hingegen in der Gruppe ohne Trendfortsetzungserwartung höher aus ($MDN = 7,00$ vs. $5,00$). Insgesamt deutet dies darauf hin, dass die Trendannahme in dieser Manipulation weniger die binäre Investitionsentscheidung beeinflusst, sondern eher mit der Kalibrierung einzelner Bewertungsparameter zusammenhängt; aufgrund der Stichprobengröße erfolgt die Interpretation rein deskriptiv.

6.2.5 Entscheidungsmodus, Rolle der Berufserfahrung, Risikoprofil

In Teil 3 der Umfrage wurden wieder Informationen zur Entscheidungsgrundlage erfasst. Bei insgesamt 18 gültigen Antworten, verteilen sich die Antworten wie folgt: Bauchgefühl ($n = 8$; $44,4\%$), Zahlen/Daten/Fakten ($n = 5$; $27,8\%$) und Erfahrungswerte ($n = 5$; $27,8\%$). Der Entscheidungsmodus ($0 =$ Intuition, $10 =$ Analyse) lag bei $M = 4,28$ ($SD = 3,10$; $MDN = 4,50$; $Min = 0$; $Max = 9$; $n = 18$) mit einer Standardabweichung von 3, was auf eine relative Streuung hindeutet und somit eine heterogene Herangehensweise in der Stichprobe nahelegt. Der berichtete Einfluss der Berufserfahrung wurde als eher hoch eingeschätzt und lag im Durchschnitt bei $M = 5,94$ ($SD = 3,57$; $MDN = 6,50$; $Min = 0$; $Max = 10$; $n = 18$) mit einer Standardabweichung von 3,6, was ebenfalls eine Streuung unter den Teilnehmenden impliziert.

¹³³ Eigene Darstellung.

Will, Rock

Die Risikobereitschaft im Privaten lag im oberen Mittelfeld bei $M = 6,22$ ($SD = 2,07$; $MDN = 6,50$; $Min = 2$; $Max = 10$; $n = 18$). Die Risikobereitschaft im Beruf lag bei $M = 5,50$ ($SD = 2,12$; $MDN = 6,00$; $Min = 2$; $Max = 10$; $n = 18$). Rein deskriptiv zeigten Experten höhere Mittelwerte und damit eine höhere Risikobereitschaft als Studierende (persönlich: Experten $M = 7,25$ vs. Studierende $M = 4,17$; beruflich: Experten $M = 6,00$ vs. Studierende $M = 4,50$).

6.2.6 Zusammenfassende Darstellung der zentralen Ergebnisse

Gruppe 2 umfasst 18 auswertbare Datensätze bei einer Abschlussrate von 24 %.

Bei East Star (niedriger Anker) lag die Entscheidungssicherheit im mittleren bis höheren Bereich ($MDN=7,50$). Die Kaufpreisschätzungen konzentrierten sich eng um den Median ($MDN=24,50$ Mio. €, $n=16$) und lagen damit nahe am gesetzten Anker.

Bei Courtyard (positives Framing) wurde die Tonwirkung neutral bis leicht positiv bewertet ($MDN=5,50$), die Risikowahrnehmung lag im mittleren Bereich ($MDN=5,00$). Analog zu Gruppe 1 zeigte die Risikowahrnehmung eine stärkere Differenzierung nach Investitionsentscheidung als die Tonbewertung.

Bei Primeverse stimmten 55,6 % einer Investition zu. Die Fortsetzung der positiven Mietpreisentwicklung wurde mehrheitlich erwartet (61,1 % Ja), ohne jedoch die Investitionsquote zwischen den Antwortgruppen wesentlich zu beeinflussen.

Übergreifend ist trotz vergleichbarer Objektkennzahlen eine relevante Streuung der Investitionsentscheidungen festzustellen: East Star 38,9 % Ja, Courtyard und Primeverse je 55,6 % Ja. Auch Gruppe 2 bestätigt damit, dass Entscheidungen nicht allein entlang objektiver Kennzahlen verlaufen.

6.3 Entscheidungsdeterminanten

In beiden Umfragegruppen wurde die Relevanz zentraler Entscheidungskriterien auf einer Likert Skala (0 – 10 Skala) insgesamt hoch eingeschätzt (vgl. Abbildung 7). Hierbei zeichnen sich vor allem Pricing und Entwicklungspotenzial als relevante Entscheidungskriterien ab. In Gruppe 1 lagen die höchsten Werte bei Pricing ($M = 8,45$; $MDN = 9,00$) und Entwicklungspotenzial ($M = 8,20$; $MDN = 8,50$), während in Gruppe 2 Pricing nochmals leicht höher ausfiel ($M = 8,78$; $MDN = 9,00$) und Entwicklungspotenzial ebenfalls hoch bewertet wurde ($M = 7,83$; $MDN = 8,50$). Die Wichtigkeit der Mikrolage wurde in Gruppe 1 stärker gewichtet ($M = 7,95$; $MDN = 8,00$) als in Gruppe 2 ($M = 7,00$; $MDN = 7,00$). Die Mieterstruktur zeigte in beiden Gruppen hohe Relevanz, mit tendenziell höheren Mittelwerten in Gruppe 2 ($M = 7,56$) gegenüber Gruppe 1 ($M = 7,10$). ESG-Faktoren wurde in beiden Gruppen moderat bewertet (Gruppe 1: M

= 5,40; Gruppe 2: M = 6,17) und zeigte somit unter den zur Verfügung stehenden Parametern die geringste Relevanz.

| | | Mikrolage | Mieterstruktur | Entwicklungspotential | ESG Faktoren | Pricing |
|------------------------|---------------|-----------|----------------|-----------------------|--------------|---------|
| Studierende n = 17 | Median | 8,0 | 8,0 | 9,0 | 6,0 | 8,0 |
| | SD | 1,8 | 1,6 | 1,7 | 2,2 | 1,9 |
| Expert:innen n = 21 | Median | 7,0 | 7,0 | 8,0 | 5,0 | 9,0 |
| | SD | 1,8 | 2,3 | 1,9 | 2,4 | 0,7 |

Hinweis Zellen zeigen Medianwerte (0–10; 0 = gar nicht relevant, 10 = sehr relevant).
Farbintensität entspricht der Höhe des Medians.
Standardabweichungen (SD) sind hellgrau hinterlegt.

| Farbskala | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------|---|---|---|---|---|---|---|---|---|---|----|
|-----------|---|---|---|---|---|---|---|---|---|---|----|

Abbildung 7: Heatmap zur wahrgenommenen Relevanz (Mediane) zentraler Entscheidungskriterien nach Erfahrungsgruppe¹³⁴

Wie Abbildung 7 zeigt, unterscheiden sich die Entscheidungsdeterminanten deskriptiv nach Erfahrungsniveau. Studierende (n=17) priorisierten Entwicklungspotenzial am höchsten (MDN=9), Experten (n=21) hingegen Pricing (MDN=9, SD=0,7 – mit der geringsten Streuung aller Kriterien). Mikrolage und Mieterstruktur wurden von Studierenden höher bewertet (je MDN=8) als von Experten (je MDN=7), wobei die Mieterstruktur in der Expertengruppe heterogener eingeschätzt wurde (SD=2,3). ESG-Faktoren wiesen in beiden Gruppen die niedrigsten Medianwerte (MDN=6 bzw. 5) und höchsten Streuungen auf, was auf uneinheitliche Bedeutungszuschreibungen hindeutet.

Die anschließende Mehrfachauswahl bestätigt dieses Muster: In Gruppe 1 wurde Entwicklungspotenzial am häufigsten als wichtigstes Kriterium genannt (75,0 %), in Gruppe 2 dominierte Pricing (77,8 %). Beide Gruppen zeigen ein konsistentes Präferenzmuster zugunsten wirtschaftlicher Kernparameter bei nachrangiger Gewichtung von ESG.

6.4 Informationsadäquanz

Über beide Gruppen hinweg zeigt sich ein konsistentes Bild einer niedrigen bis mittleren Informationsadäquanz der Teaser. In Gruppe 1 lag die Informationsausreichendheit bei M = 4,35 (SD = 2,11; MDN = 4,00; Min = 1,00; Max = 10,00; n = 20) und in Gruppe 2 bei M = 4,89 (SD = 2,27; MDN = 5,00; Min = 0; Max = 8; n = 18). Die Freitextangaben zu fehlenden Informationen (Gruppe 1: n = 10; Gruppe 2: n = 8) verweisen in beiden Gruppen überwiegend auf Miet-/Mietvertrags- bzw. Marktmietinformationen sowie Mieter-/Bonitätsinformationen als zentrale Lücken. Ergänzend wurden in Gruppe 1 CapEx/Instandhal-

¹³⁴ Eigene Darstellung.

Will, Rock

tung/Zustand sowie vereinzelt Leerstands- und Objektdetails genannt. In Gruppe 2 traten darüber hinaus vereinzelt Hinweise auf Bewertungs-/Finanzierungskennzahlen, ESG, Business-Case-Aspekte, Marktdaten und weitere Objektdetails auf.

6.5 Entscheidungsunterstützung

Die Relevanz externer Quellen wurde in beiden Gruppen als gering eingeschätzt (Gruppe 1: MDN=2,50; Gruppe 2: MDN=2,50). Entsprechend dominierten einfache oder keine Hilfsmittel: In Gruppe 1 nutzten je 50,0 % Taschenrechner/Excel bzw. keine Hilfsmittel, in Gruppe 2 überwog „keine Hilfsmittel“ (66,7 %). Weitere Tools wie Marktberichte, Online-Suche oder interne Datenbanken wurden selten eingesetzt. Die Gesamtzahl genutzter Hilfsmittel war in beiden Gruppen niedrig und rechtsschief verteilt (Gruppe 1: MDN=0,50; Gruppe 2: MDN=0). Deskriptiv zeigte sich in Gruppe 1 eine höhere Tool-Nutzung bei Experten (MDN=1) gegenüber Studierenden (MDN=0).

6.6 Entscheidungssicherheit

Die Entscheidungssicherheit liegt über beide Gruppen hinweg im mittleren bis leicht erhöhten Bereich. Bei East Star zeigt sich gruppenübergreifend ein Wert von $M=6,55$ (MDN=7,00, $n=38$), bei der globalen Sicherheit über alle Teaser ein leicht niedrigerer Wert von $M=5,92$ (MDN=7,00). Im hypothetischen Investorenwettbewerb fällt die Zuversicht deutlich geringer aus ($M=4,55$, MDN=5,00, $n=38$). Insgesamt zeigt sich: Während Teilnehmende ihre eigene Investitionsentscheidung moderat sicher treffen, ist die Zuversicht im kompetitiven Bietkontext spürbar gedämpft.

7 Zusammenfassende Darstellung der Humanbefragung

Über beide Gruppen hinweg zeigt sich ein konsistentes Bild, dass die frühe Opportunitätsprüfung trotz weitgehend vergleichbarer Objektkennzahlen eine relevante Varianz in Investitionsentscheidungen und Bewertungskomponenten aufweist. Entlang der untersuchten Heuristiken lassen sich folgende Muster erkennen:

- (1) **Ankereffekt:** Die Preisnennungen verschieben sich deutlich in Richtung des gesetzten Ankers. Bei hohem Anker (Gruppe 1) liegen die zentralen Preisindikationen messbar höher als bei niedrigem Anker (Gruppe 2). Studierende tendieren in beiden Gruppen zu höheren Preisnennungen als Expert:innen, während Expert:innen insgesamt konservativer auf den Anker reagieren.
- (2) **Rahmungseffekt:** Die Tonwahrnehmung folgt der Manipulation (Gruppe 1 deutlich negativer, Gruppe 2 neutral bis leicht positiv). Gleichzeitig differenziert die Risikowahrnehmung konsistent nach Investitionsentscheidung (Ablehnung führt zu höheren Risikowerten), während sich die Investitionsquote nicht eindeutig spiegelbildlich zur Tonalität verschiebt. Nach Erfahrungsniveau zeigen sich Unterschiede in der Risikokalibrierung und Tonwahrnehmung. Expert:innen

schätzen das Risiko deutlich höher ein als Studierende und unter positivem Framing bewerten Expert:innen den Textton deutlich positiver als Studierende.

- (3) **Repräsentativitätsheuristik:** Die Manipulation der Mieten beeinflusst die Trendannahmen (bei positiver Historie wird häufiger eine Fortsetzung erwartet). Die Investitionsquote bleibt jedoch weitgehend ähnlich zwischen Trend- und Nicht-Trend-Antwortgruppen; Unterschiede zeigen sich in der Kaufpreisnennung. Expert:innen interpretieren negative Trends tendenziell zurückhaltender als Studierende.
- (4) **Herdverhalten:** Die eingeblendete Mehrheitsmeinung führt nur sehr selten zu einer Entscheidungsanpassung. Sofern eine Anpassung genannt wird, tritt diese eher bei Studierenden auf.
- (5) **Überkonfidenz:** Es zeigen sich keine starken Hinweise auf ausgeprägte Überkonfidenz: Die Teilnehmenden berichten zwar eine moderate Entscheidungssicherheit, schätzen ihre Erfolgchancen im Investorenwettbewerb jedoch deutlich niedriger ein. Das spricht für eine realistische Kalibrierung (Bewusstsein für Unsicherheit) als für Selbstüberschätzung im Screening-Setting.

Insgesamt zeigt sich damit ein differenziertes, aber in sich konsistentes Bild heuristikbezogener Entscheidungsmuster in der Humanbefragung. Die nachfolgende Abbildung 8 bündelt diese Befunde in verdichteter Form und illustriert die Ausprägung der untersuchten Heuristiken getrennt nach Gruppe 1 und Gruppe 2 als Grundlage für den folgenden Vergleich mit den KI-Modellen.

| | | Menschliche Heuristiken | | | | |
|------------|----------|-------------------------|---------------------------------|-------------------------------|---------------------|---------------|
| | | Ankereffekt | Rahmungseffekt | Repräsentativitätsheuristik | Herdverhalten | Überkonfidenz |
| Der Mensch | Gruppe 1 | Δ € +7.73m | Ton: MDN 2,5 Risiko: MDN 6,0 | Trendfortschreibung 50,0 % | Anpassung: 5,0 % | Δ M = +1,85 |
| | Gruppe 2 | Δ € - 2.77m | Ton: MDN 5,5 Risiko: MDN 5,0 | Trendfortschreibung 61,1 % | Anpassung: 5,6 % | Δ M = +2,16 |

Legende:

| | | | | | |
|--|-------|---------|--------|----------------|--------|
| | stark | moderat | robust | schwach invers | invers |
|--|-------|---------|--------|----------------|--------|

Ankereffekt 1. Median-Kaufpreis minus Marktwert (27,27 Mio. €).
 Rahmungseffekt 2. Medianwerte von Tonwirkung und Risikowahrnehmung.
 Repräsentativitätsheuristik 3. Anteil der Personen die eine Fortschreibung des Mietpreistrends erwarten.
 Herdverhalten 4. Anteil der Entscheidungsanpassungen nach Mehrheitsinformation.
 Überkonfidenz 5. Differenz Frage 30 (Entscheidungssicherheit East Star) zu Frage 45 (Entscheidungssicherheit Investorenwettbewerb).

Abbildung 8: Darstellung der heuristikbezogenen Muster in der Humanbefragung¹³⁵

Auf Basis dieses menschlichen Ausgangsbildes wird im nächsten Kapitel geprüft, ob KI-Modelle unter identischer Teaser- und Fragebogenlogik vergleichbare heuristik-konsistente Muster replizieren oder abweichende Entscheidungsprofile zeigen.

¹³⁵ Eigene Darstellung.

8 KI-Test

Zur Beantwortung der Forschungsfrage wurden acht LLMs in einem standardisierten Testablauf mit identischen Teasern und Fragebogenlogik getestet – je Modell ein Durchlauf pro Experimentalgruppe. Die Entscheidung gegen Mehrfachläufe folgt dem Fokus der Arbeit: Der Vergleich menschlicher und KI-gestützter Entscheidungsprozesse steht im Vordergrund, nicht eine modellinterne Robustheitsanalyse. Um Abhängigkeiten von einzelnen Systemen zu reduzieren, wurden acht etablierte Modelle einbezogen. Die Ergebnisse sind als Momentaufnahme (Stand: Dezember 2025) zu interpretieren.

8.1 Standardisierung der KI-Instruktion (Prompt Design)

Da LLM-Antworten sensibel gegenüber Prompt-Variationen sein können, wurde für alle Tests – in Anlehnung an Liang (2023) et al. – ein möglichst natürlicher und über alle Modelle hinweg konsistenter Prompt verwendet.¹³⁶ Der in der Humanbefragung enthaltene Disclaimer wurde für die KI-Durchläufe entfernt, um unbeabsichtigte Hinweise auf das Studiendesign zu vermeiden. Alle Modelle erhielten folgenden einheitlichen Prompt:

„Versetze dich in die Rolle eines Investment Analysten im Bereich Real Estate. Ich werde dir nun drei Investment Teaser zur Verfügung stellen und möchte, dass du diese in der Reihenfolge wie sie in der Umfrage erscheinen untersuchst und die entsprechenden Fragen streng nach dem Fragebogen beantwortest.“

Die rollenbasierte Instruktion („Investment Analyst“) gibt dem Modell eine konsistente fachliche Perspektive vor und reduziert Interpretationsspielräume. Auf zusätzlichen Kontext wurde bewusst verzichtet, da LLMs durch ergänzende Prompt-Informationen stark beeinflusst werden können.¹³⁷

8.2 Auswahl der getesteten Modelle

Die Modellauswahl basiert auf drei anerkannten Benchmarks (abgerufen am 08.12.2025): GPQA¹³⁸ testet wissenschaftliches Expertenwissen, SWE-Bench¹³⁹ bewertet Softwareengineering-Kompetenz anhand realer GitHub-Aufgaben, und AIME¹⁴⁰ dient als Indikator für fortgeschrittenes mathematisches Reasoning. Auf dieser Grundlage wurden ChatGPT 5.2, Gemini 3 Pro, Grok 4, KIMI K2 Thinking, DeepSeek V3, Mistral Le Chat, Claude Sonnet 4.5 sowie Perplexity in den Vergleich aufgenommen.

¹³⁶ Vgl. Liang et al. (2023), S. 2, S. 161.

¹³⁷ Vgl. Salinas & Morstatter (2024), S. 1

¹³⁸ Vgl. Rein et al. (2023), S. 1.

¹³⁹ Vgl. Jimenez et al. (2024), S. 1.

¹⁴⁰ Vgl. Yue & Klein (2025), S. 11.

8.3 Durchführung und Datenerfassung

Die KI-Tests wurden im Zeitraum vom 08.12.2025 und 21.12.2025 durchgeführt, indem jedes Modell den identischen Prompt erhielt (vgl. Kap. 8.1) und anschließend die drei Investment-Teaser in der im Fragebogen vorgegebenen Reihenfolge bearbeitet hat. Der Zugriff auf die Modelle erfolgte über den Webbrowser, wodurch eine Kontrolle technischer Parameter nicht möglich war. Teaser aus Gruppe 1 und 2 wurden jeweils in separaten Chats eingegeben, um eine modellseitige Erkennung der Manipulationen zu verhindern. Die Antworten wurden als CSV exportiert und in Microsoft Excel weiterverarbeitet.

8.4 Deskriptive Beschreibung der KI-Tests

Die folgenden Ergebnisse stellen die KI-Modellläufe im Zwei-Gruppen-Design deskriptiv dar, mit Fokus auf Unterschiede entlang der untersuchten Heuristiken. Da je Modell ein Durchlauf pro Gruppe vorliegt, handelt es sich ausdrücklich um eine nicht-inferenzstatistische Bestandsaufnahme – die Befunde sind als modellspezifische Beobachtungen (Stand: Dezember 2025) zu verstehen, nicht als stabile, generalisierbare Modelleigenschaften.

8.4.1 Ankereffekt: Preis- und Renditebezug als primäres Muster

Beim Ankereffekt zeigt sich auf der Preisebene ein konstantes Muster: In 7 von 8 Modellen liegt der vorgeschlagene Kaufpreis in Gruppe 1 über dem in Gruppe 2, mit Differenzen von -1,0 bis +15,0 Mio. €. Besonders ausgeprägt sind die Abstände bei Perplexity Pro (+15,0 Mio. €), Mistral Le Chat (+14,5 Mio. €), Claude Sonnet 4.5 (+13,5 Mio. €) und DeepSeek V3 (+12,5 Mio. €). KIMI K2 (+1,2 Mio. €) und Grok 4.1 (+3,5 Mio. €) zeigen geringere Differenzen. ChatGPT 5.2 Thinking bildet mit -1,0 Mio. € die einzige Ausnahme und deutet auf eine inverse Reaktion hin.

Auf der Renditeebene sind die Muster heterogener: In 6 von 8 Fällen liegt die IRR in Gruppe 1 unter der in Gruppe 2; Claude Sonnet 4.5 und ChatGPT 5.2 Thinking weichen positiv ab. Dies deutet darauf hin, dass einige Modelle die Preisinformation in die Renditelogik „durchreichen“, während andere Renditeannahmen stärker unabhängig variieren.

8.4.2 Rahmungseffekt: Entscheidungswechsel als dominante Reaktion, aber nicht universell

Beim Rahmungseffekt wechselten 5 von 8 Modellen ihre Investitionsentscheidung von Nein zu Ja (DeepSeek V3, Gemini Flash 2.5, KIMI K2, Mistral Le Chat, Perplexity Pro), was auf eine Wirkung des Framings hindeutet. Grok 4.1 und Claude Sonnet 4.5 zeigen stabile Entscheidungen, obwohl beide eine deutliche Verschiebung der Risikowahrnehmung aufweisen ($\Delta \text{Risiko} \geq 2$) – diese Modelle scheinen Entscheidungen an einer stabileren Schwelle auszurichten. ChatGPT 5.2 Thinking bildet mit einem inversen

Will, Rock

Effekt („Ja → Nein“) die Ausnahme. Insgesamt zeigt sich: Framing wirkt entweder als Entscheidungshebel (Nein → Ja) oder als Bewertungsverschiebung bei stabiler Entscheidung.

8.4.3 Repräsentativitätsheuristik: eher robuste Entscheidungen, punktuell starke Ausschläge

Bei der Repräsentativitätsheuristik zeigt sich ein robusteres Bild als beim Framing: In 5 von 8 Fällen bleibt die Entscheidung stabil („Ja → Ja“), 2 Modelle wechseln von Nein zu Ja (DeepSeek V3, KIMI K2). Vier Modelle (u.a. ChatGPT 5.2 Thinking, Gemini Flash 2.5, Mistral Le Chat, Perplexity Pro) zeigen bei stabiler Entscheidung keine relevanten IRR-Verschiebungen und werden als robust eingestuft.

Auf IRR-Ebene zeigen Claude Sonnet 4.5 und KIMI K2 mit über 270 Basispunkten Unterschied die stärksten Verschiebungen, begleitet von veränderten Aussagen zur Trendfortschreibung. DeepSeek V3 und Grok 4.1 weisen moderate Anpassungen von 100 Basispunkten auf. Mistral Le Chat ist vollständig robust – ohne Anpassung der Mietpreisentwicklungseinschätzung und identischer IRR.

8.4.4 Herdenverhalten: robustes Verhalten über alle Modelle hinweg

Modellübergreifend zeigt sich, dass keines der getesteten Modelle bei Einblendung einer Mehrheitsinformation die Entscheidung anpasst.

8.4.5 Überkonfidenz: insgesamt hohes Niveau bei moderater Streuung

Modellübergreifend liegt die geäußerte Entscheidungssicherheit auf einem relativ hohen Niveau (Durchschnitt von ca. 7,09 auf einer Skala von 0 – 10), mit einer Spannweite von 6,00 (ChatGPT 5.2 Thinking) bis 7,75 (Grok 4.1). Auffällig ist dabei, dass hohe Entscheidungssicherheit nicht zwingend mit geringer Sensitivität gegenüber Heuristiken einhergeht: So zeigt Grok 4.1 hohe Entscheidungssicherheit bei gleichzeitig moderater Anfälligkeit für den Rahmungseffekt oder die Repräsentativitätsheuristik. Umgekehrt weist ChatGPT 5.2 Thinking die niedrigste durchschnittliche Entscheidungssicherheit auf, zeigt jedoch inverses Verhalten in Bezug auf den Rahmungseffekt und den Ankereffekt. Es zeigt sich also, dass die ausgedrückte Entscheidungssicherheit nicht als verlässlicher Indikator für die Robustheit dient.

8.5 Menschliche Heuristiken in KI-Modellen

Die Auswertung der LLM-Antworten im zweigruppigen Experimentaldesign zeigt, dass heuristikähnliche Muster in großen Sprachmodellen vorzufinden sind und sich diese je nach Modell unterscheiden können. Abbildung 8 veranschaulicht die dargestellten Muster mit einer entsprechenden farblichen Darstellung. Der Ankereffekt tritt bei mehreren Modellen auf: Perplexity Pro, Mistral Le Chat, Claude Sonnet 4.5 und DeepSeek V3 setzen in Gruppe 1 deutlich höhere Kaufpreise an, was auf eine starke Verarbeitung

numerischer Anker als Referenzpunkt hindeutet. ChatGPT 5.2 Thinking zeigt hingegen eine gegenläufige Differenz und damit vergleichsweise höhere Robustheit.

Beim Rahmungseffekt ist die Manipulation auf Wahrnehmungsebene bei allen Modellen nachweisbar: Ton wird in Gruppe 2 positiver, Risiko geringer bewertet. Entscheidungsrelevant wird das Framing bei DeepSeek V3, Gemini Flash 2.5, KIMI K2, Mistral Le Chat und Perplexity Pro, die von Ablehnung zu Zustimmung wechseln und zugleich höhere Kaufpreise ansetzen. Claude Sonnet 4.5 und Grok 4.1 bleiben entscheidungsstabil, vollziehen die Risikoverschiebung jedoch nach – was auf partielle Robustheit auf Entscheidungsebene hindeutet. ChatGPT 5.2 Thinking zeigt erneut ein inverses Muster.

| | Menschliche Heuristiken | | | | | |
|------------------------------------|-------------------------|----------------|-----------------------------|-------------------------|---------------|--------|
| | Ankereffekt | Rahmungseffekt | Repräsentativitätsheuristik | Herdenverhalten | Überkonfidenz | |
| Sprachmodell alphabetisch sortiert | Claude Sonnet 4.5 | Δ € +13.5m | Nein → Nein Δ R 2 | Nein n/a Δ 275 bps | Nein | Ø 6,50 |
| | DeepSeek V3 | Δ € +12.5m | Nein → Ja Δ R 3 | Nein Ja Δ -100 bps | Nein | Ø 7,50 |
| | Google Gemini Flash 2.5 | Δ € +4.0m | Nein → Ja Δ R 2 | Ja Ja Δ -30 bps | Nein | Ø 7,25 |
| | Grok 4.1 | Δ € +3.5m | Ja → Ja Δ R 4 | Ja Ja Δ 100 bps | Nein | Ø 7,75 |
| | KIMI K2 | Δ € +1.2m | Nein → Ja Δ R 3 | Nein Ja Δ 270 bps | Nein | Ø 7,50 |
| | Mistral Le Chat | Δ € +14.5m | Nein → Ja Δ R 4 | Ja Ja Δ 0 bps | Nein | Ø 7,50 |
| | Open AI Chat GPT 5.2 | Δ € -1.0m | Ja → Nein Δ R 1 | Ja Ja Δ 50 bps | Nein | Ø 6,00 |
| | Perplexity Pro | Δ € +15.0m | Nein → Ja Δ R 1 | Ja Ja Δ -50 bps | Nein | Ø 6,75 |

| | | | | | |
|-----------------|--------------|----------------|---------------|-----------------------|---------------|
| Legende: | stark | moderat | robust | schwach invers | invers |
|-----------------|--------------|----------------|---------------|-----------------------|---------------|

| | |
|-----------------------------|--|
| Ankereffekt | 1. Preisdifferenz Frage 16 (G1 - G2) (= Δ €) |
| Rahmungseffekt | 2.1. Entscheidungsänderung Frage 19 (G1 zu G2) |
| Rahmungseffekt | 2.2. Differenz Risikowahrnehmung Frage 23 (G1 zu G2) (= Δ R) |
| Repräsentativitätsheuristik | 3.1. Trendförschreibung Mietpreisentwicklung Frage 28 (G1 zu G2) |
| Repräsentativitätsheuristik | 3.2. Differenz in der IRR Frage 27 (G1 zu G2) (= Δ ... bps) |
| Herdenverhalten | 4.1. Entscheidungsänderung bei Mehrheitsanzeige Frage 14 (G1 zu G2) |
| Überkonfidenz | 5.1. Durchschnittswert Frage 30 G1, Frage 30 G2, Frage 45 G1, Frage 45 G2 (=Ø) |

Abbildung 9: Heuristiken in KI-Modellen.¹⁴¹

Bei der Repräsentativitätsheuristik zeigt sich ein heterogeneres Bild: Nur KIMI K2 und DeepSeek V3 passen die explizite Trenderwartung in der intendierten Richtung an. Bei weiteren Modellen bleibt die Erwartungsfrage stabil, während sich dennoch implizite Anpassungen über IRR-Annahmen beobachten lassen. Insgesamt erscheint dieser Effekt weniger konsistent als Anker- und Rahmungseffekt.

Herdenverhalten zeigt sich bei keinem Modell – die Frage zur Meinungsanpassung bei abweichender Mehrheit wird durchgehend verneint. Die Entscheidungssicherheit liegt modellübergreifend im oberen Bereich (6,00–7,75), mit ChatGPT 5.2 Thinking als niedrigstem und Grok 4.1 als höchstem Wert.

¹⁴¹ Eigene Darstellung.

8.6 Plausibilisierung der beobachteten Ergebnisse

Die Unterschiede in den Antworten der KI beweisen aufgrund der geringen Durchlaufzahl noch keine menschlichen Heuristiken, veranschaulichen allerdings Heuristik ähnliche Muster in LLMs. Um zu eruieren, was Gründe für dieses Verhalten sein könnte, wurde ein Experte für Mensch-KI-Interaktion konsultiert.

Als zentraler Mechanismus für Ausgabereaktionen von KI-Modellen ist die Trainingsdatenbasis relevant: LLMs werden auf sehr großen Mengen überwiegend menschlich erzeugter Daten trainiert; darin sind menschliche Argumentations- und Bewertungsmuster implizit enthalten, sodass verzerrungsähnliche Muster im Output grundsätzlich auftreten können.¹⁴²

Zweitens hebt der Experte zur Mensch-KI Interaktion eine Rahmungssensitivität hervor: Modelle tendieren häufig dazu, die im Kontext angelegte Rahmung zu „bestätigen“; Neutralität der Formulierungen ist daher für belastbare Bewertungen zentral.¹⁴³ Bei der Durchführung der KI-Tests wurde ein recht kurzer, neutraler Prompt verwendet, um die KI nicht durch die Prompt Formulierung zu beeinflussen. Daraus lässt sich schließen, dass die KI-Modelle allerdings auf etwaige Rahmungen im Rahmen der Teaser sensitiv reagiert haben könnten.

Drittens betont der Experte zur Mensch-KI Interaktion im Interview, dass Teaser und Exposé s zwar regelmäßig die zentralen Objekt- und Transaktionsinformationen für ein erstes Screening enthalten, für eine belastbare Investmententscheidung jedoch häufig entscheidungsrelevante Kontext- und Meta-Informationen nicht in ausreichender Tiefe abbilden.¹⁴⁴ Insbesondere hebt er die Bedeutung der Mikrolage hervor: In urbanen Märkten könne bereits geringe Unterschiede in der Lage Auswirkungen auf Vermietbarkeit, Mietniveau, Nachfrage und Risiko haben, während Teaser diese Feinräumigkeit typischerweise nur aggregiert oder narrativ beschreiben.¹⁴⁵

8.7 Entscheidungsdeterminanten in KI-Modellen

Wie Abbildung 9 zeigt, werden alle zentralen Entscheidungskriterien von den KI-Modellen hoch eingestuft. Pricing und Mikrolage dominieren: Pricing erreicht in Gruppe 1 $M=9,75$ und in Gruppe 2 $M=9,50$, die Mikrolage wird in beiden Gruppen identisch bewertet ($M=9,25$). Entwicklungspotenzial und Mieterstruktur liegen ebenfalls auf hohem Niveau ($M=8,25-8,50$). Den klarsten gruppenspezifischen Unterschied zeigen ESG-Faktoren, die in Gruppe 2 deutlich höher gewichtet werden (G1: $M=6,25$ vs. G2: $M=7,13$). In der zusammenfassenden Darstellung der Heatmap (vgl. Abbildung 9) wird sichtbar, dass die Rangfolge der Kernparameter über die Modelle hinweg bemerkenswert konsistent ist, während die

¹⁴² Vgl. Experte B (2026), Interview.

¹⁴³ Ebd.

¹⁴⁴ Ebd.

¹⁴⁵ Ebd.

Sensitivität gegenüber der Bedingung ($\Delta = G1 - G2$) primär bei ESG und punktuell bei Pricing bzw. Entwicklungspotenzial auftritt. Für Mikrolage und Mieterstruktur sind keine Differenzen zwischen G1 und G2 erkennbar. Demgegenüber zeigen mehrere Modelle für ESG negative Differenzen (höhere Gewichtung in G2), mit den größten Ausschlägen von $\Delta = -2$ (u.a. bei Gemini Flash 2.5 und KIMI K2). In Bezug auf das Pricing ist das Muster insgesamt stabil (mehrheitlich Werte von 9 bis 10), jedoch mit einzelnen stärkeren Abweichungen (z.B.: DeepSeek V3: G1 = 10 vs. G2 = 8; $\Delta = +2$). Insgesamt deutet die Matrix damit auf ein KI-seitig stark wirtschaftlich geprägtes Präferenzmuster (Pricing/Mikrolage/Entwicklung/Mieterstruktur) hin, bei gleichzeitig nachrangiger, aber bedingungsabhängig variierender Bedeutung von ESG.

| | | G1 | G2 | G1 | G2 | G1 | G2 | G1 | G2 | G1 | G2 | |
|-------------------------------------|-------------------------|-----------|------|----------------|------|-----------------------|------|--------------|------|---------|------|----|
| | | Mikrolage | | Mieterstruktur | | Entwicklungspotential | | ESG Faktoren | | Pricing | | |
| Sprachmodelle alphabetisch sortiert | Claude Sonnet 4.5 | Score | 9 | 9 | 8 | 8 | 9 | 9 | 6 | 7 | 10 | 10 |
| | | Δ | 0,0 | | 0,0 | | 0,0 | | -1,0 | | 0,0 | |
| | Deep Seek V3 | Score | 8 | 9 | 9 | 8 | 9 | 9 | 6 | 7 | 10 | 8 |
| | | Δ | -1,0 | | 1,0 | | 0,0 | | -1,0 | | 2,0 | |
| | Google Gemini Flash 2.5 | Score | 10 | 9 | 8 | 8 | 9 | 10 | 7 | 9 | 10 | 10 |
| | | Δ | 1,0 | | 0,0 | | -1,0 | | -2,0 | | 0,0 | |
| | Grok 4.1 | Score | 10 | 10 | 8 | 9 | 9 | 8 | 7 | 7 | 10 | 10 |
| | | Δ | 0,0 | | -1,0 | | 1,0 | | 0,0 | | 0,0 | |
| | KIMI K2 | Score | 10 | 10 | 9 | 9 | 7 | 8 | 5 | 7 | 10 | 10 |
| | | Δ | 0,0 | | 0,0 | | -1,0 | | -2,0 | | 0,0 | |
| | Mistral Le Chat | Score | 9 | 9 | 8 | 8 | 8 | 9 | 7 | 7 | 9 | 8 |
| | | Δ | 0,0 | | 0,0 | | -1,0 | | 0,0 | | 1,0 | |
| | Open AI Chat GPT 5.2 | Score | 9 | 9 | 8 | 8 | 8 | 7 | 6 | 6 | 10 | 10 |
| | | Δ | 0,0 | | 0,0 | | 1,0 | | 0,0 | | 0,0 | |
| | Perplexity Pro | Score | 9 | 9 | 8 | 8 | 7 | 8 | 6 | 7 | 9 | 10 |
| | | Δ | 0,0 | | 0,0 | | -1,0 | | -1,0 | | -1,0 | |
| Modellensemble, n=8 | Median | 9,0 | 9,0 | 8,0 | 8,0 | 8,5 | 8,5 | 6,0 | 7,0 | 10,0 | 10,0 | |
| | Δ | 0,0 | | 0,0 | | 0,0 | | -1,0 | | 0,0 | | |

Hinweis Zellen zeigen je KI-Modell und Bedingung (G1/G2) den generierten Wichtigkeits-Score (0-10; 0 = gar nicht relevant, 10 = sehr relevant). Die Zeile Δ gibt die Differenz zwischen den Bedingungen an ($\Delta = G1 - G2$).



Abbildung 10: Heatmap der von KI-Modellen bewerteten Relevanz zentraler Entscheidungskriterien nach Bedingung (G1/G2).¹⁴⁶

¹⁴⁶ Eigene Darstellung.

9 Vergleich zwischen Menschen und KI

Die vorangegangenen Kapitel 6 und Kapitel 7 haben die Ergebnisse der Humanbefragung sowie des KI-Modelltests (Kapitel 8) betrachtet. Nachfolgend werden die Befunde deskriptiv zusammengeführt. Der Vergleich ist – entsprechend der Methodik – als heuristikbasierter Musterabgleich (Pattern Matching) zu verstehen und nicht als inferenzstatistische Prüfung.

9.1 Unterschied in Bezug auf Heuristiken

Nachfolgend wird zuerst ein Vergleich zwischen den untersuchten Heuristiken vollzogen und im Anschluss erfolgt die differenzierte Betrachtung nach Entscheidungsdeterminanten.

9.1.1 Ankerheuristik (Anchoring)

In der Humanbefragung zeigt sich ein anker-konsistentes Richtungsmuster: Bei hohem Preisanker liegen die zentralen Kaufpreisnennungen deutlich höher als bei niedrigem Preisanker; zudem reagieren Studierende in beiden Bedingungen tendenziell stärker (höhere Preisnennungen) als Expert:innen, die insgesamt konservativer adjustieren. In den KI-Tests zeigt sich ebenfalls bei der Mehrheit der Modelle eine Preisverschiebung in Richtung des gesetzten Ankers (in 7 von 8 Fällen liegt der Kaufpreis in Gruppe 1 über Gruppe 2), wobei die Abstände teils substantiell unterschiedlich ausfallen; gleichzeitig existiert mit ChatGPT 5.2 Thinking eine gegenläufige Einzelbeobachtung (inverse Ausprägung).

9.1.2 Rahmungseffekt (Framing)

Beim Menschen folgt die Tonwahrnehmung der experimentellen Manipulation deutlich, und höhere Risikowahrnehmung geht konsistent mit Ablehnung einher. Die Investitionsbereitschaft verschiebt sich jedoch nicht spiegelbildlich zur Tonalität; Experten zeigen unter negativem Framing höhere Risikobewertungen und unter positivem Framing deutlich positivere Tonbewertungen. Bei den KI-Modellen ist Rahmungssensitivität auf Wahrnehmungsebene durchgängig sichtbar: Alle Modelle bewerten die positiv formulierte Variante positiver und risikoärmer. Auf Entscheidungsebene ist Framing bei der Mehrheit handlungsrelevant (überwiegend Wechsel von „Nein“ zu „Ja“), während einzelne Modelle trotz Wahrnehmungsverschiebung entscheidungsstabil bleiben. ChatGPT 5.2 Thinking zeigt erneut ein inverses Muster.

9.1.3 Repräsentativitätsheuristik (Representativeness)

In der Humanbefragung beeinflusst die historische Mietentwicklung die Trenderwartungen (positive Historie → häufiger Fortsetzungsprognose), während die Investitionsquote weitgehend stabil bleibt. Unterschiede zeigen sich eher in der Kalibrierung einzelner Bewertungsparameter. Bemerkenswert ist,

dass Studierende im negativen Setting häufiger eine Trendfortschreibung erwarten, Experten im positiven Setting. In den KI-Tests zeigt sich ein robusteres Bild als beim Framing: Entscheidungen bleiben mehrheitlich stabil; nur einzelne Modelle passen die explizite Trenderwartung in der intendierten Richtung an, während bei anderen implizite Verschiebungen über Renditeannahmen auftreten.

9.1.4 Herdenverhalten (Herding)

In der Humanbefragung führt die eingeblendete Mehrheitsinformation nur sehr selten zu einer Entscheidungsanpassung. Die beobachteten Anpassungen treten ausschließlich bei den Studierenden auf. In den KI-Tests passt keines der Modelle seine Entscheidung aufgrund der Mehrheitsinformation an.

9.1.5 Überkonfidenz (Overconfidence)

In der Humanbefragung deutet die Kombination aus mittlerer Entscheidungssicherheit und gedämpfter Wettbewerbszuversicht auf eine eher realistische Kalibrierung hin. KI-Modelle zeigen demgegenüber modellübergreifend höhere geäußerte Sicherheit – wobei hohe Entscheidungssicherheit kein belastbarer Robustheitsindikator ist, da einzelne Modelle trotz hoher Sicherheit sensitiv gegenüber Heuristiken reagieren.

9.2 Entscheidungsdeterminanten

Sowohl Menschen als auch KI-Modelle priorisieren wirtschaftlich geprägte Kriterien: Pricing, Entwicklungspotenzial, Mikrolage und Mieterstruktur dominieren in beiden Gruppen, während ESG-Faktoren die geringste Relevanz erhalten. Menschen und KI teilen damit eine ähnliche ökonomische Grundlogik der Opportunitätsbewertung.

10 Gestaltungsempfehlung für einen KI-Agenten

Die nachfolgende Tabelle 4 bündelt die untersuchten Verzerrungsrisiken entlang der Opportunitätsprüfung und verknüpft diese mit konkreten Anknüpfungspunkten für die Agentenkonzeption, basierend auf Tabelle 1 (vgl. Kapitel 3). Es werden die dort bereits genannten typischen Informationsreizen aus Teasern sowie die entsprechend relevante Heuristiken aufgegriffen und aufgezeigt, wie diese durch eine Agentenstruktur adressiert werden können. Die Tabelle leitet für jeden Prozessschritt (vgl. Abbildung 10, Kap. 10.1) ab, inwiefern der Agent in der frühen Opportunitätsprüfung unterstützen kann und welche Werkzeuge hierfür eingesetzt werden sowie die Ausgaben welche als Entscheidungsgrundlage bereitgestellt werden. Damit bildet die Tabelle die Brücke von der theoretisch und empirisch hergeleiteten Problemstruktur zur Ausgestaltung des Prozessdesigns, in dem die beschriebenen Anknüpfungspunkte in einen praxistauglichen Agentenablauf überführt werden. Dabei wird KI nicht als autonomes Bewer-

Will, Rock

tungsmodell verstanden, sondern als unterstützendes System in der frühen Phase der Opportunitätsprüfung mit klaren Rollen- und Prozessvorgaben.¹⁴⁷ Das Kapitel und die Entwicklung des KI-Agenten stützt sich auf die durchgeführten Experteninterviews. Falls notwendig werden externe Quellen referenziert, um mögliche Lücken in den Aussagen zu komplementieren.

| Schritt in der Opportunitätsprüfung | Typischer Informationsreiz im Teaser/Exposé | Heuristik | Erwartetes Verzerrungsmuster | Evidenz im Datensatz* | Anknüpfungspunkt Agent (vgl. Schritte im konzeptionellen Prozessablauf) | Werkzeuge | Agenten Ausgabe |
|---|--|-----------------------------|--|-----------------------|---|------------------------------|---|
| 1 Ersteindruck & "Investment Story" | "Investment Story", Ton des Textes & Darstellung der Fakten | Rahmungseffekt | Identische Fakten führen je nach Darstellung zu anderer Risiko-/Attraktivitätswahrnehmung. | ✓ | Schritt 4 & 7: Nach Extraktion (Schritt 3) neutralisiert der Agent die Darstellung und trennt Fakten von Wertung. Wenn notwendig werden durch RAG externe Quellen hinzugezogen. | - RAG - HITL | - Möglichst neutrale Zusammenfassung - Quellenzuordnung wenn notwendig |
| 2 Erste Preis- & Renditeplausibilisierung | Kaufpreisvorstellung / Angebotspreis | Ankereffekt | Numerischer Startwert dient als Anker; spätere Anpassungen lösen sich unzureichend davon. | ✓ | Schritt 5 & 6: Der Agent berechnet Kennzahlen und erzeugt eine realistische Bandbreite; zusätzlich wird ein „Anker-Check“ durchgeführt. Rückfragen lösen ein Review-Gate aus. | - Rechner - RAG - HITL | - KPIs - Wertbandbreite - Darstellung des Ankereinflusses |
| 3 Narrative aus dem Markt / Cash Flow | Trendgrafiken (z.B.: zu den Mieten) | Repräsentativitätsheuristik | Muster aus der Vergangenheit werden als repräsentativ für die Zukunft interpretiert: Trendfortschreibung. | ✓ | Schritt 5 & 6: Der Agent ergänzt die Teaser-Narrative um Fakten und erstellt ein Mini-Szenario (Base/Downside). Trends werden mit neutralen Fakten aus dem Markt gegengeprüft. | - RAG - Rechner | - Szenarioberechnung - Risikotreiber |
| 4 Informationen aus dem Markt / von Marktteilnehmern | Hinweis auf andere Investoren/Mehrheitsmeinung/ Interesse | Herdverhalten | Orientierung an Mehrheit bei Unsicherheit: Anpassung an soziale Information statt rationale Bewertung. | X | Schritt 4 & 6: Der Agent detektiert Social-Proof-Signale und markiert diese als Kontextinformation. Er erzwingt eine unabhängige Begründung auf objektiveren Kriterien und fordert mindestens ein Gegenargument. | - HITL - RAG | - Unabhängige Begründung - Kriterienbasiert |
| 5 Subjektive Entscheidungssicherheit & Kompetenzgefühl | Selbst einschätzung „Wie sicher bin ich?“ / „Wie kompetent?“ | Überkonfidenz | Überschätzung eigener Fähigkeiten/Wissen/Prognosen bei gleichzeitiger Unterschätzung künftiger Unsicherheiten. | X | Schritt 4 & 6: Der Agent koppelt Entscheidungssicherheit an Datenqualität. Wenn zentrale Informationen fehlen, erzeugt er eine Missing-Info-Liste und formuliert gezielte Rückfragen; hohe Sicherheit ist nur mit Evidenz zulässig. | - HITL - RAG | - Entscheidungssicherheit - Liste fehlender Pflichtangaben |

Hinweis:

* in Bezug auf die Humanbefragung dieser Arbeit.
 ✓ = beobachtet, X = uneindeutig
 RAG = Retriever Augmented Generation
 HITL = Human in the Loop

Tabelle 4: Heuristiken in der Opportunitätsprüfung: Evidenz und Anknüpfungspunkte des Entscheidungsagenten.¹⁴⁸

10.1 Prozessdesign

Abbildung 10 visualisiert den konzeptionellen Aufbau eines KI-basierten Entscheidungsagenten zur Unterstützung der anfänglichen Opportunitätsprüfung. Der konzeptionell entwickelte Agent wurde auf der Plattform n8n nachgebaut und getestet. Zur Veranschaulichung der praktischen Umsetzung ist ergänzend eine Videodemonstration des entwickelten KI-Agenten verfügbar.¹⁴⁹

Hierfür wurde ergänzend zu Abbildung 10 eine vertikale Multi-Agentenstruktur gewählt, in der die Agenten hierarchisch in einer Ebene angeordnet sind und Hand in Hand arbeiten.¹⁵⁰ Der Designfokus liegt auf einer standardisierten, nachvollziehbaren und datenbasierten Vorprüfung, die menschliche Opportunitätsprüfung nicht ersetzt, sondern käuferseitig in der frühen Filterphase unterstützt.

¹⁴⁷ Vgl. Experte B (2026), Interview.

¹⁴⁸ Eigene Darstellung, in Anknüpfung an Kap. 3.2.

¹⁴⁹ Link: https://drive.google.com/file/d/1-AFevGlunMsJulRyvLjoS_Kj3TCzotb3/view?usp=sharing

¹⁵⁰ Vgl. Masterman et al. (2024), S. 2 – 3.

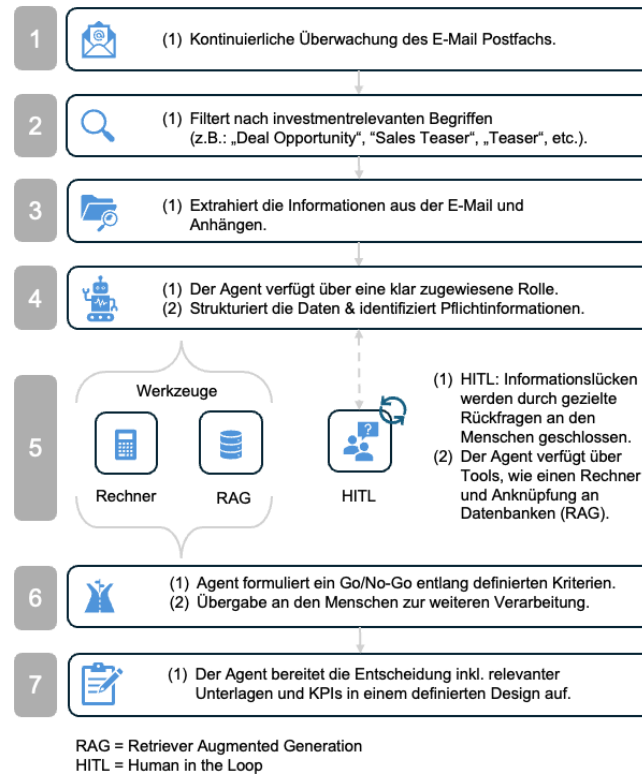


Abbildung 11: Konzeptioneller Prozessablauf eines KI-Agenten im Rahmen der anfänglichen Opportunitätsprüfung im Investment Management.¹⁵¹

(1) E-Mail-Trigger und Filterung investmentrelevanter Opportunitäten

Der Prozess startet mit der kontinuierlichen Überwachung des E-Mail-Postfachs als primärem Eingangskanal für Akquisitionsoportunitäten (sog. *E-Mail Trigger*).¹⁵² Zuerst werden eingehende Nachrichten (ungelesene Nachrichten) über einen Stichwortfilter (z.B.: „Teaser“, „Sales Teaser“, „Deal Opportunity“, etc.) als potenziell investmentrelevant klassifiziert. Die Filterlogik dient dabei als pragmatische Vorselektion, um die korrekten Dokumente und E-Mails zu identifizieren.

(2) Extraktion aus E-Mail-Inhalt und Anhängen

Sobald der Workflow eine Übereinstimmung mit einem der definierten Suchparameter hat, startet der nächste Schritt: die Extraktion der Informationen aus dem E-Mail Body und den Anhängen (sog. *Extract from file*). Der Investment Agent nutzt diese zuerst unstrukturierten Informationen und bringt sie nun in eine ordentliche, dem Agenten vorgegebene Datenstruktur. Hierfür wurde dem Agenten seine Rolle im Rahmen des *Prompt Engineering* vorgegeben.^{153, 154} Die zentrale Funktion des Agenten passiert im sogenannten *Brain* – also der LLM-Anbindung des Agenten – in diesem Fall Open AI Chat GPT 4.1 Mini.

¹⁵¹ Eigene Darstellung.

¹⁵² Vgl. Experte A (2025), Interview.

¹⁵³ Vgl. Wang et al. (2024), S. 15.

¹⁵⁴ Vgl. Experte B (2026), Interview.

Will, Rock

Im Kontext dieser Arbeit wurden die untersuchten Heuristiken explizit benannt und der Agent entsprechend instruiert, diese bewusst zu vermeiden. Der Experte zur Mensch-KI Interaktion verweist darauf, dass eine möglichst neutrale Rahmung des Agenten essenziell ist, damit dieser seine Aufgabe objektiv ausführen kann.¹⁵⁵

(3) Werkzeuge des Agenten

Um die Aufgaben effizienter ausführen zu können, werden dem Agenten unterschiedliche Werkzeuge (sog. *Tools*) bereitgestellt.¹⁵⁶ Abbildung 10 bündelt Rechner und Retriever Augmented Generation (RAG) als Werkzeuge. Die RAG-Anbindung adressiert die Notwendigkeit externer Datenquellen. Aufgrund begrenzter Datenverfügbarkeit wurde im Rahmen dieses n8n Workflows eine sog. *Serp API* angebunden, welche es dem Agenten erlaubt, die Google Suche zu verwenden, sollten beispielsweise Informationen fehlen. Der Experte aus dem Investmentmanagement verweist für die Nutzung externer Quellen auf Research-Reports, Sozioökonomika sowie Miet- und Kaufpreis-Comparables und interne Datenbanken zur Einordnung von Opportunitäten.¹⁵⁷ Der Experte zur Mensch-KI Interaktion nennt RAG als zentrale Technik, um fallbezogen passende Informationen aus Datenbanken abzurufen und als Kontext in die Modellverarbeitung einzuspeisen.¹⁵⁸

Der Rechner dient der Berechnung und Plausibilisierung von Kennzahlen. Dazu gehören beispielsweise Anfangsrenditen sowie die Cash-on-Cash und die IRR. Zur Plausibilisierung werden Marktmiete vs. Ist-Miete, Kaufpreis je Quadratmeter und Anfangsrendite herangezogen.¹⁵⁹ Diese zahlenbasierten Checks stützen zudem den Grundsatz, dass Ergebnisse objektivierbar und zahlenbasiert begründbar sein müssen und Kennzahlen besonders gut geeignet sind, um früh zu filtern.¹⁶⁰

(4) Strukturierung & Ausgabeformat

Für ein konstant strukturiertes Ausgabeformat wird dem Agenten ein Output Schema (mithilfe eines sog. *Structured Output Parser*) vorgegeben, welches er einhalten muss und die Basis für die weiterführende Analyse bildet. Dieses Schema orientiert sich an den in der Frühphase entscheidungsrelevanten Kriterien, welche von dem Experten hervorgehoben wurden: Lage, Risikoprofil und Objektqualität (u.a. Baujahr, technischer Zustand) sowie auf Vermietungsseite Nettokaltmiete sowie das Net Operating Income, Vermietungsstand, Restlaufzeiten (WALT) sowie der Mietermix im Objekt.¹⁶¹ Die Extraktion bildet damit die Grundlage für eine konsistente, vergleichbare Vorprüfung über heterogene Dokumentqualitäten hinweg. Diese Informationen werden im Rahmen des Workflows automatisiert in eine Google Sheets

¹⁵⁵ Vgl. Experte B (2026), Interview.

¹⁵⁶ Vgl. Masterman (2024), S. 2.

¹⁵⁷ Vgl. Experte A (2025), Interview.

¹⁵⁸ Vgl. Experte B (2026), Interview.

¹⁵⁹ Vgl. Experte A (2025), Interview.

¹⁶⁰ Ebd.

¹⁶¹ Ebd.

Liste eingetragen. Diese Ausgestaltung ist unmittelbar mit den genannten Potenzialen für KI-Unterstützung konsistent: Effizienzgewinne liegen insbesondere in der strukturierten Datenübernahme aus Opportunitäten in interne Modelle sowie in der standardisierten Erstellung von Entscheidungsvorlagen und Präsentationen für Team- und Komitee Prozesse.¹⁶² Identifiziert der Agent fehlende, entscheidungsrelevante Angaben aus der vorgegebenen Kriterienliste wird der Prozess nicht durch Annahmen fortgesetzt, sondern in eine Rückfrageschleife überführt. Der Experte zur Mensch-KI Interaktion empfiehlt explizit, Modelle dazu anzuhalten, Rückfragen zu stellen, um Informationslücken zu schließen, statt automatisch zu abstrahieren.¹⁶³

(5) Aufbereitung fehlender Informationen & Human in the Loop

Fehlende Informationen bzw. Rückfragen werden vom E-Mail-Agenten in eine strukturierte E-Mail formuliert, welche im Anschluss einem Menschen zur Freigabe vorgelegt wird. Dieser *Human-in-the-Loop* (HITL) Ansatz erfüllt eine doppelte Funktion: Er schließt Informationslücken durch gezielte Rückfragen und wirkt als Mechanismus gegen Fehlannahmen bzw. Übervertrauen in automatisierte Empfehlungen.¹⁶⁴ Der Mensch hat im Rahmen der Freigabe die Möglichkeit konstruktives Feedback zu hinterlassen, welches vom Revisions Agenten weiterverarbeitet werden würde oder die E-Mail freizugeben. Daraufhin sendet der E-Mail-Agent automatisiert die Mail mit den Rückfragen an den Absender der Eingangsnachricht.

(6) Go/NoGo Empfehlung entlang definierter Parameter

Ebenfalls formuliert der Agent basierend auf den vorliegenden Informationen eine Entscheidung, ob das Investment sinnvoll ist. In den Praxistests zeigt sich, dass dies aufgrund der begrenzten Informationen zur Opportunität in der anfänglichen Phase, durchgehend abgelehnt wurde (vgl. Tabelle 5). Diese Kriterien spiegeln die von dem Experten aus dem Investmentmanagement beschriebene Filterlogik der Frühphase wider.¹⁶⁵ Der Agent dient damit als standardisierende Entscheidungsunterstützung, während die finale Verantwortung im Prozess beim menschlichen Entscheider verbleibt.¹⁶⁶

10.2 Zusammenfassung Prozessdesign

Die in der Literatur identifizierten und in der Praxis teilweise replizierten Verzerrungsrisiken in der frühphasigen Opportunitätsprüfung werden mithilfe dieses konzeptionell entwickelten KI-Agenten adressiert. Heterogene Teaser-Informationen werden strukturiert, zentrale Kennzahlen werden extrahiert und plausibilisiert, Informationslücken werden erkannt und über eine HITL-Schleife gezielt an den Absender

¹⁶² Vgl. Experte A (2025), Interview.

¹⁶³ Vgl. Experte B (2026), Interview.

¹⁶⁴ Ebd.

¹⁶⁵ Vgl. Experte A (2025), Interview.

¹⁶⁶ Ebd.

zurückgespielt. Damit adressiert der Agent Schwachstellen menschlicher Erstbewertungen und der klassischen Bewertung mithilfe von LLM-Modellen durch klares Prompt Engineering, definierte Output-Formate und prüfbare Datenbezüge. Gleichzeitig wird durch die prototypische Instanziierung in n8n die praktische Umsetzbarkeit des konzeptionellen Entwurfes (Abbildung 10) demonstriert und die Brücke von der Theorie zur Anwendung geschlossen.

10.3 Test des Investment Agenten mit den Teasern

Zur Plausibilisierung und praxisnahen Prüfung der agentischen Outputs wurde die Pilotierung mit zwei identischen Prozessdurchläufen je Teaser-Variante (Gruppe 1/Gruppe 2) durchgeführt (vgl. Tabelle 5). Damit basieren beide Durchläufe auf denselben Stimuli wie die Humanbefragung und die KI-Tests. Die Durchläufe wurden mit unverändertem Workflow sowie gleichen Systemeinstellungen (inkl. identischer Promptlogik und Outputstruktur) ausgeführt. Dadurch lässt sich neben dem gruppenbezogenen Effekt (G1 vs. G2) zugleich der Effekt eines LLM-basierten Entscheidungsagenten in Bezug auf die untersuchten Heuristiken beobachten. Die Ergebnisse zeigen zunächst eine hohe Robustheit gegenüber dem Ankereffekt im relativen Sinn: Bei East Star variiert der Angebotskaufpreis zwischen G1 und G2 deutlich (45,0 Mio. € vs. 27,0 Mio. €; $\Delta = 18,0$ Mio. €), während der vom Agenten abgeleitete realistische Kaufpreis innerhalb des jeweiligen Durchlaufs nur gering differiert (Durchlauf 1: 17,34 Mio. € vs. 17,00 Mio. €; $\Delta = 0,34$ Mio. €; Durchlauf 2: 23,75 Mio. € vs. 24,30 Mio. €; $\Delta = 0,55$ Mio. €). Damit bleibt die gruppenbezogene Preisreaktion im Verhältnis zur Ankeränderung sehr klein, was als Indiz dafür gewertet werden kann, dass der Agent den Angebotskaufpreis nicht als dominanten Ankerpreis übernimmt, sondern die Bewertung rationaler, unabhängig vom gesetzten Anker erfolgt. Gleichzeitig werden absolute Niveauunterschiede in Bezug auf den genannten Kaufpreis zwischen den Durchläufen sichtbar (z.B.: East Star G1 (Lauf 1 \rightarrow Lauf 2): 17,336 \rightarrow 23,750 Mio. €, East Star G2 (Lauf 1 \rightarrow Lauf 2): 17,000 \rightarrow 24,300 Mio. €), obwohl die Eingabelogik und Einstellungen konstant sind.

Für Rahmungs- und Repräsentativitätssignale zeigen die beiden Durchläufe eine differenziertere Stabilität. Bei Courtyard bleibt der Angebotskaufpreis konstant (28,3 Mio. €), der realistische Kaufpreis verändert sich jedoch zwischen den Durchläufen (G1: 26,5 \rightarrow 31,13 Mio. €, G2: 24,5 \rightarrow 24,0 Mio. €); zudem kippt bei G1 die Einschätzung der Mietpreisentwicklung von „Ja“ (Durchlauf 1) zu „Nein“ (Durchlauf 2), während G2 konsistent „Nein“ liefert. Für Primeverse (Angebot konstant 32,0 Mio. €) wechselt die Einschätzung der Mietpreisentwicklung bei G1 von „Nein“ (Durchlauf 1) zu „Ja“ (Durchlauf 2), während G2 in beiden Durchläufen „Ja“ liefert. Darüber hinaus verschiebt sich der dem Agenten zufolge realistische Kaufpreis (G1: 27,0 \rightarrow 31,0 Mio. €, G2: 29,358 \rightarrow 28,8 Mio. €). Die Framing-Indikatoren „Risikowahrnehmung“ und „Tonwahrnehmung“ zeigen punktuelle Abweichungen, auch wenn sie sich insgesamt in einem engen Korridor bewegen (z.B.: Courtyard G1 Risiko 6 \rightarrow 7, Primeverse G1 Risiko 5 \rightarrow 6, Ton teils 7 \rightarrow 8 (konstant in Gruppe 2)). Insgesamt deutet dies darauf hin, dass der Agent gegenüber klaren numerischen Ankerreizen eine hohe Robustheit aufweist, während es bei text- bzw. narrativen Urteilen in

diesem Setting zu minimalen Abweichungen kommen kann, auch wenn diese sich in einem überschaubaren Rahmen halten.

Aus diesem Test ergeben sich zentrale Implikationen für das Prozessdesign: Erstens liefern die Ergebnisse indikative Evidenz, dass ein agentisches Setup Anker-Sensitivitäten reduzieren kann. Zweitens machen die beobachteten Niveauunterschiede zwischen den Durchläufen deutlich, dass für eine produktive Prozessunterstützung zusätzliche Stabilisierungsmechanismen sinnvoll sind, um die Reproduzierbarkeit insbesondere bei trend- und rahmungsnahen Einschätzungen zu erhöhen. Darüber hinaus illustrieren die Tests praxisnah, dass der Agent relevante Kennzahlen korrekt erkennt und in ein strukturiertes Datenformat integriert. Außerdem lehnt der Agent die Investitionsentscheidungen systematisch in allen Durchläufen ab, hinterfragt die dargestellten Informationen und schließt durch automatisierte Erstellung von Rückfragen die relevanten Informationslücken.

| Name der Opportunität | Baujahr | Fläche in m ² | NKM p.a. | Vermietungsstand | WAULT | Angebotskaufpreis | Realistischer Kaufpreis | Infos vollständig? | Mietpreisentwicklung | Risiko (0 - 10) | Ton (0 - 10) | Investitionsentscheidung | Sicherheitsgefühl |
|------------------------------|---------|--------------------------|-------------|------------------|-------|-------------------|-------------------------|--------------------|----------------------|-----------------|--------------|--------------------------|-------------------|
| East Star Gruppe 1 Lauf 1 | 1999 | 8.618 | 1.282.507 € | 73% | 6,5 | 45.000.000 € | 17.336.000 € | Nein | Ja | 6 | 7 | Nein | 8 |
| East Star Gruppe 1 Lauf 2 | 1999 | 8.618 | 1.282.507 € | 73% | 6,5 | 45.000.000 € | 23.750.000 € | Nein | Ja | 5 | 7 | Nein | 8 |
| East Star Gruppe 2 Lauf 1 | 1999 | 8.618 | 1.282.507 € | 73% | 6,5 | 27.000.000 € | 17.000.000 € | Nein | Ja | 6 | 7 | Nein | 7 |
| East Star Gruppe 2 Lauf 2 | 1999 | 8.618 | 1.282.507 € | 73% | 6,5 | 27.000.000 € | 24.300.000 € | Nein | Ja | 6 | 8 | Nein | 7 |
| Courtyard 46 Gruppe 1 Lauf 1 | k.A. | 8.906 | 1.324.200 € | 76% | 6,9 | 28.300.000 € | 26.500.000 € | Nein | Ja | 6 | 7 | Nein | 7 |
| Courtyard 46 Gruppe 1 Lauf 2 | k.A. | 8.906 | 1.324.200 € | 76% | 6,9 | 28.300.000 € | 31.130.000 € | Nein | Nein | 7 | 7 | Nein | 8 |
| Courtyard 46 Gruppe 2 Lauf 1 | k.A. | 8.906 | 1.324.200 € | 76% | 6,9 | 28.300.000 € | 24.500.000 € | Nein | Nein | 6 | 7 | Nein | 7 |
| Courtyard 46 Gruppe 2 Lauf 2 | k.A. | 8.906 | 1.324.200 € | 76% | 6,9 | 28.300.000 € | 24.000.000 € | Nein | Nein | 6 | 8 | Nein | 7 |
| Primeverse Gruppe 1 Lauf 1 | 2005 | 8.830 | 1.324.200 € | 75% | 6,9 | 32.000.000 € | 27.000.000 € | Nein | Nein | 5 | 7 | Nein | 7 |
| Primeverse Gruppe 1 Lauf 2 | 2005 | 8.830 | 1.324.200 € | 75% | 6,9 | 32.000.000 € | 31.000.000 € | Nein | Ja | 6 | 7 | Nein | 4 |
| Primeverse Gruppe 2 Lauf 1 | 2005 | 8.830 | 1.324.200 € | 75% | 6,9 | 32.000.000 € | 29.358.000 € | Nein | Ja | 5 | 8 | Nein | 6 |
| Primeverse Gruppe 2 Lauf 2 | 2005 | 8.830 | 1.324.200 € | 75% | 6,9 | 32.000.000 € | 28.800.000 € | Nein | Ja | 5 | 7 | Nein | 7 |

| |
|----------|
| Gruppe 1 |
| Gruppe 2 |

Tabelle 5: KI-Agent Outputs auf ausgewählte Fragen in Bezug auf die Investment Teaser.

11 Schlussbetrachtung

Diese Arbeit untersuchte, wie sich Entscheidungsprozesse in der frühen Bewertung von Immobilien-Investmentopportunitäten zwischen menschlichen Entscheidungsgruppen und ausgewählten LLMs in Bezug auf die ausgewählten Heuristiken und Entscheidungsdeterminanten unterscheiden. Basierend darauf wurde ein Konzept für einen KI-basierten Entscheidungsagenten zur Prozessunterstützung abgeleitet und praxisnah mithilfe von dem Workflow Automatisierungstool n8n aufgebaut.

11.1 Fazit

Unter Berücksichtigung der Forschungsfragen ergeben sich folgende Erkenntnisse.

Forschungsfrage I: Wie unterscheidet sich die Bewertung von Investmentopportunitäten zwischen menschlichen Entscheidungsgruppen aus der Immobilienwirtschaft und künstlichen Intelligenzen, in Bezug auf Heuristiken und Entscheidungsdeterminanten?

Auf Heuristikebene reagieren sowohl Menschen als auch LLMs auf die gesetzten Reize, mit unterschiedlichen Ausprägungen. Beim Ankereffekt zeigt sich in beiden Gruppen ein anker-konsistentes Muster: Höhere Preisanker führen zu höheren Kaufpreisnennungen. LLMs replizieren dieses Muster in 7 von 8 Fällen; ChatGPT 5.2 Thinking verhält sich invers.

Beim Rahmungseffekt folgt die menschliche Tonwahrnehmung der Manipulation, und Risikowahrnehmung korreliert mit der Investitionsentscheidung – ohne sich jedoch spiegelbildlich zur Tonalität zu verschieben. Bei KI-Modellen ist Rahmungssensitivität auf Wahrnehmungsebene durchgängig sichtbar; auf Entscheidungsebene ist Framing bei der Mehrheit handlungsrelevant, einzelne Modelle bleiben stabil.

Bei der Repräsentativitätsheuristik verändern Menschen vor allem Trendannahmen und Bewertungsparameter, während die Investitionsquote stabil bleibt. KI-Modelle zeigen ebenfalls überwiegend stabile Entscheidungen; Anpassungen treten eher punktuell über Renditeannahmen auf.

Herdenverhalten tritt in beiden Gruppen kaum auf: Menschen passen ihre Entscheidung selten an Mehrheitsinformationen an – und wenn, dann ausschließlich Studierende – während KI-Modelle modellübergreifend keine Anpassung zeigen.

Bei Überkonfidenz verhalten sich Menschen im hypothetischen Wettbewerbskontext zurückhaltend, trotz grundsätzlicher Entscheidungssicherheit. KI-Modelle berichten durchgängig hohe Sicherheitswerte.

Auf Ebene der Entscheidungsdeterminanten teilen Menschen und KI eine ähnliche ökonomische Grundlogik: Pricing, Entwicklungspotenzial, Mikrolage und Mieterstruktur dominieren, ESG wird am niedrigsten gewichtet. Die KI-Modelle zeigen dabei eine bemerkenswert konsistente Rangfolge, während sich in der Humanstichprobe erfahrungsabhängige Unterschiede zeigen – Experten priorisieren Pricing stärker und mit geringerer Streuung, Studierende eher Entwicklungspotenzial.

Forschungsfrage II: Welchen Heuristiken unterliegen die menschlichen Entscheidungsgruppen? Repliziert KI diese Muster oder unterscheidet sie sich systematisch?

Die Humanbefragung zeigt über die untersuchten Heuristiken hinweg konsistente Muster: (1) der Ankereffekt hat einen Effekt auf die Kaufpreisenennungen; (2) der Rahmungseffekt beeinflusst die Ton- und Risikowahrnehmung, ohne dass die Investitionsquote dadurch zwingend spiegelbildlich folgt; (3) die Repräsentativitätsheuristik beeinflusst Trendannahmen und Kalibrierung einzelner Bewertungsparameter stärker als die binäre Entscheidung; (4) Herdenverhalten ist nur schwach ausgeprägt; (5) Überkonfidenz ist ebenfalls schwach ausgeprägt beim Menschen und spiegelt sich eher in einer leicht hohen Entscheidungssicherheit wieder.

Auch die KI-Modelle zeigen heuristikähnliche Muster, jedoch mit Unterschieden in den Modellen: der Ankereffekt und der Rahmungseffekt tritt bei mehreren Modellen deutlich auf wohingegen die Modelle sich gegenüber der Repräsentativitätsheuristik insgesamt eher robust zeigen. Herdenverhalten wird grundsätzlich nicht beobachtet und im Rahmen der Überkonfidenz zeigt sich bei den KI-Modellen allgemein eine etwas stärkere Ausprägung. Hierbei wird ersichtlich, dass hohe Entscheidungssicherheit nicht automatisch mit geringer Heuristik-Sensitivität einhergeht.

Forschungsfrage III: Wie lässt sich ein Konzept für einen KI-basierten Entscheidungsagenten entwickeln, um im Prozess zu unterstützen und potenzielle Verzerrungen zu reduzieren?

Der praktische Nutzen liegt nicht darin, dass KI besser entscheidet, sondern in einem Prozessdesign, das Standardisierung, Plausibilisierung und Human-in-the-Loop-Kontrollen kombiniert. Der entwickelte Agent ist als unterstützendes System für die anfängliche Opportunitätsprüfung konzipiert – mit strukturierter KPI-Extraktion, Informationslücken-Management und standardisierten Entscheidungsvorlagen.

Die Pilotierung zeigt eine hohe relative Robustheit gegenüber dem Ankereffekt: Trotz variierendem Anker bleibt die Differenz im abgeleiteten Kaufpreis klein ($\Delta=0,34$ bzw. 0,55 Mio. €). Gleichzeitig sind Niveauunterschiede zwischen den Durchläufen sichtbar, was darauf hindeutet, dass auch zweckspezifisch entwickelte KI-Systeme dynamische Antwortmuster zeigen.

Insgesamt unterscheiden sich menschliche und KI-gestützte Entscheidungsprozesse im frühen Screening weniger durch die zugrunde liegenden Kriterien als durch die Art und Robustheit der Reaktion auf gesetzte Reize. Der zentrale Beitrag dieser Arbeit liegt in der Überführung dieser Befunde in ein praxistaugliches Agentendesign, das KI nicht als autonome Blackbox positioniert, sondern als kontrollierten Prozessbaustein mit nachvollziehbaren Outputs und menschlicher Kontrollinstanz.

11.2 Limitationen

Die Ergebnisse sind vor folgenden Limitationen einzuordnen. Erstens ist die Humanbefragung aufgrund freiwilliger Teilnahme und begrenzter Fallzahlen (G1: n=20, G2: n=18) nicht repräsentativ – die Befunde sind als explorative Ergebnisse zu verstehen. Zweitens bestehen typische Online-Umfrage-Risiken hinsichtlich Aufmerksamkeit und Antwortqualität; trotz Bereinigung offensichtlicher Platzhalterwerte können einzelne unaufmerksame Antworten nicht ausgeschlossen werden. Drittens ist der KI-Test mit einem Durchlauf pro Modell und Gruppe methodisch als Momentaufnahme zu interpretieren: Ohne Mehrfachläufe und kontrollierte Parameter kann die Stabilität der Muster nicht statistisch abgesichert werden. Die Auswahl mehrerer Modelle erhöht zwar die Vergleichsbasis, ersetzt aber keine Replikation. Der Mensch-KI-Vergleich ist schließlich bewusst als heuristikbasiertes Pattern Matching konzipiert und erlaubt keine inferenzstatistischen Rückschlüsse.

11.3 Ausblick für Forschung & Praxis

Aus den Ergebnissen lassen sich mehrere Anschlusschritte ableiten. Erstens sollte eine robuste Evaluationsroutine für KI-gestützte Systeme in der Opportunitätsprüfung etabliert werden – insbesondere durch Mehrfachläufe pro Modell zur Absicherung der Ergebnisstabilität. Zweitens bietet sich eine Erweiterung in Richtung Realprozess an: Die Integration zusätzlicher Informationsquellen und die Einbettung in reale Teamprozesse würden es ermöglichen zu prüfen, ob beobachtete Heuristikmuster unter realistischeren Bedingungen abnehmen oder sich verlagern. Drittens könnte die Humanstudie mit größeren Stichproben skaliert werden, um stärkere Differenzierungen nach Rolle, Erfahrung und Assetklasse zu ermöglichen und den explorativen Charakter in Richtung inferenzfähiger Aussagen weiterzuentwickeln. Viertens bietet der konzipierte Agent einen klaren Pfad zur praktischen Implementierung: Ein Pilotprojekt in einem realen Unternehmen könnte Zeitersparnis, Bewertungskonsistenz und Nachvollziehbarkeit als Erfolgskriterien validieren – und die systematische Erfassung eingehender Opportunitäten könnte zugleich wertvolle Marktdaten generieren.

Literatur- und Quellenverzeichnis

- Ang, A., Papanikolaou, D. und Westerfield, M. (2013) Portfolio Choice with Illiquid Assets (NBER Working Paper). National Bureau of Economic Research. Verfügbar unter: https://www.nber.org/system/files/working_papers/w19436/w19436.pdf (Zugegriffen: 16. Januar 2026).
- Hüther, M. (2023) Volkswirtschaftliche Einordnung der Pandemieschäden. In: Arnold, R. u. a. (Hrsg.) Risiko im Wandel: Herausforderung für die Versicherungswirtschaft. Wiesbaden: Springer Gabler, S. 63-75. doi: 10.1007/978-3-658-37071-8_4.
- AWS (ohne Datum) Was ist generative KI? Verfügbar unter: <https://aws.amazon.com/what-is/generative-ai/> (Zugegriffen: 14. November 2025).
- Bao, H.X.H. und Li, S.H. (2016) Overconfidence and Real Estate Research: A Review of the Literature. University of Cambridge. Verfügbar unter: https://www.researchgate.net/publication/303386509_Overconfidence_and_Real_Estate_Research_A_Review_of_the_Literature (Zugegriffen: 28. November 2025).
- Black, R.T. und Diaz III, J. (1996) The Use of Information Versus Asking Price in the Real Property Negotiation Process. *Journal of Property Research*, 13(4), S. 287-297. Verfügbar unter: <https://doi.org/10.1080/095999196368808> (Zugegriffen: 21. Februar 2026).
- BNP Paribas Real Estate (2025 a): BNP Paribas Real Estate: Frankfurter Büromarkt mit bester Zwischenbilanz der letzten 20 Jahre. Büromarktbericht. Online im Internet: <https://www.realestate.bnpparibas.de/marktberichte/buromarkt/frankfurt-at-a-glance> (Zugegriffen: 06. Oktober 2025.)
- BNP Paribas Real Estate (2025 b): BNP Paribas Real Estate: Investmentmarkt Frankfurt. Online im Internet: <https://www.realestate.bnpparibas.de/marktberichte/investmentmarkt/frankfurt-at-a-glance> (Zugegriffen: 08. Oktober 2025.)
- Bourier, G. (2022) Beschreibende Statistik: Praxisorientierte Einführung - Mit Aufgaben und Lösungen. 14. Aufl., Wiesbaden: Springer Gabler. Verfügbar unter: <https://link.springer.com/book/10.1007/978-3-658-37021-3> (Zugegriffen: 21. Februar 2026).
- Cascao, A., Quelhas, A.P. und Cunha, A.P. (2023) Heuristics and cognitive biases in the housing investment market. Volume 16, Issue 5. Verfügbar unter: <https://doi.org/10.1108/IJHMA-05-2022-0073> (Zugegriffen: 28. November 2025).
- CBRE GmbH (2025) Business Insights | How AI is Advancing Decision-Making In Corporate Real Estate. Verfügbar unter: <https://www.cbre.com/insights/viewpoints/how-ai-is-advancing-decision-making-in-corporate-real-estate> (Zugegriffen: 16. November 2025)
- Charness, G., Gneezy, U. und Kuhn, M.A. (2012) Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1), S. 1-8. Verfügbar unter: <https://doi.org/10.1016/j.jebo.2011.08.009> (Zugegriffen: 21. Februar 2026).

- Chen, Y. u. a. (2025) A Manager and an AI Walk into a Bar: Does ChatGPT Make Biased Decisions Like We Do? Verfügbar unter: <https://pubsonline.informs.org/doi/10.1287/msom.2023.0279> (Zugegriffen: 17. Januar 2026).
- Collins, C., Dennehy, D., Conboy, K. und Mikalef, P. (2021) Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, 60, 102383. Verfügbar unter: <https://doi.org/10.1016/j.ijinfomgt.2021.102383> (Zugegriffen: 22. Februar 2026).
- D´Onofrio, S. (2024) Generative Künstliche Intelligenzen - die neue Ära der kreativen Maschinen. Verfügbar unter: <https://doi.org/10.1365/s40702-024-01069-0>.
- Deutsche Bundesbank (2022) Monetary policy and banking business. Verfügbar unter: <https://www.bundesbank.de/resource/blob/901692/af4fa9ac8f60e2b54ee-afc72d1461ee3/mL/2022-11-geldpolitik-data.pdf> (Zugegriffen: 18. Januar 2026).
- Dubois, A. und Gadde, L.-E. (2002) Systematic combining: an abductive approach to case research. *Journal of Business Research*. Verfügbar unter: <https://www.sciencedirect.com/science/article/pii/S0148296300001958?via=ihub> (Zugegriffen: 20. Januar 2026).
- DZ HYP AG (2023) Immobilienmarkt Deutschland 2023 / 2024. Verfügbar unter: https://www.pfandbrief.market/wp-content/uploads/2023/10/Immobilienmarkt_Deutschland_2023_DE_final.pdf (Zugegriffen: 15. Januar 2026).
- Echterhoff, J. u. a. (2024) Cognitive Bias in Decision-Making with LLMs. Verfügbar unter: <https://aclanthology.org/2024.findings-emnlp.739.pdf> (Zugegriffen: 16. Januar 2026).
- Eichholtz, P. und Yönder, E. (2014) CEO Overconfidence, REIT Investment Activity and Performance. Volume 43, Issue 1. Verfügbar unter: <https://doi.org/10.1111/1540-6229.12054> (Zugegriffen: 28. November 2025).
- Ernst & Young Real Estate GmbH und ZIA Zentraler Immobilien Ausschuss e.V. (2024) Digitalisierungsstudie 2024: „KI - ein ‚Game Changer‘ in der Immobilienwirtschaft?“ Verfügbar unter: <https://zia-deutschland.de/project/digitalisierungsstudie-2024-ki-ein-game-changer-in-der-immobilienwirtschaft/> (Zugegriffen: 10. Mai 2025).
- Europäische Zentralbank (2023 a) Annual Report 2022. Verfügbar unter: <https://www.ecb.europa.eu/pub/pdf/annrep/ecb.ar2022~8ae51d163b.en.pdf> (Zugegriffen: 16. Januar 2026).
- Europäische Zentralbank (2023 b) The euro area bank lending survey - Third quarter of 2023. Verfügbar unter: https://www.ecb.europa.eu/stats/ecb_surveys/bank_lending_survey/html/ecb.blsurvey2023q3~b960111b2d.en.html (Zugegriffen: 21. Februar 2026).
- Europäisches Parlament (2020) Was ist künstliche Intelligenz und wie wird sie genutzt? Verfügbar unter: <https://www.europarl.europa.eu/topics/de/article/20200827STO85804/was-ist-kunstliche-intelligenz-und-wie-wird-sie-genutzt> (Zugegriffen: 22. Februar 2026).

- Experte A (2025): Leitfadengestütztes Experteninterview im Bereich institutionelles Investment Management (international tätiges Immobilienunternehmen).
- Experte B (2026): Leitfadengestütztes Experteninterview zu Mensch-KI-Interaktion und agentenbasierten Systemen (Forschungsumfeld Künstliche Intelligenz).
- Fitzpatrick, M. u. a. (2023) Generative AI can change real estate, but the industry must change to reap the benefits. McKinsey. Verfügbar unter: <https://www.mckinsey.com/industries/real-estate/our-insights/generative-ai-can-change-real-estate-but-the-industry-must-change-to-reap-the-benefits> (Zugegriffen: 16. November 2025).
- Gallimore, P., Hansz, J.A. und Gray, A. (2000) Decision making in small property companies. *Journal of Property Investment & Finance*. Verfügbar unter: <http://dx.doi.org/10.1108/14635780010357569> (Zugegriffen: 17. Januar 2026).
- Germani, F. und Spitale, G. (2025) Source framing triggers systematic bias in large language models. Verfügbar unter: <https://www.science.org/doi/10.1126/sciadv.adz2924> (Zugegriffen: 25. November 2025).
- gif Gesellschaft für Immobilienwirtschaftliche Forschung e.V. (2011) Real Estate Investment Management. Verfügbar unter: <https://gif-ev.com/glossar-eintrag/real-estate-investment-management/> (Zugegriffen: 9. Dezember 2025).
- Gigerenzer, G. und Gaissmaier, W. (2011) Heuristic Decision Making. *Annual Review of Psychology*, 62, S. 451-482. Verfügbar unter: <https://doi.org/10.1146/annurev-psych-120709-145346> (Zugegriffen: 21. Februar 2026).
- Hansen, A.L. und Lee, S.J. (2025) Financial Stability Implications of Generative AI: Taming the Animal Spirits. Verfügbar unter: <https://doi.org/10.17016/FEDS.2025.090> (Zugegriffen: 24. November 2025).
- Holzmann, Dr.C. und Rock, Dr.V. (2025) „Mensch vs. Maschine. Wie der Mensch in einer KIgeprägten Arbeitswelt agiert“, gif im Fokus - Mensch Immobilie!, Januar, S. 5-10.
- Huijbregts, R., George, B. und Bekkers, V. (2021) Valuation tools and politicians' willingness to sell public real estate: a survey experiment. In: *Public Management Review*. Verfügbar unter: <https://doi.org/10.1080/14719037.2021.1874496> (Zugegriffen: 14. April 2026).
- Jimenez, C.E., Yang, J., Wettig, A., Yao, S., Pei, K., Press, O. und Narasimhan, K.R. (2024) SWE-bench: Can Language Models Resolve Real-world GitHub Issues? In: *Proceedings of the Twelfth International Conference on Learning Representations (ICLR)*. Verfügbar unter: <https://openreview.net/forum?id=VTF8yNQM66> (Zugegriffen: 21. Februar 2026).
- Jin, C. und Gallimore, P. (2010) The effects of information presentation on real estate market perceptions. Verfügbar unter: <https://doi.org/10.1080/09599916.2010.518404> (Zugegriffen: 24. November 2025).

- Kahneman, D. (2003) Maps of Bounded Rationality: Psychology for Behavioral Economics. *The American Economic Review*. Verfügbar unter: https://kahneman.scholar.princeton.edu/sites/g/files/toruqf3831/files/kahneman/files/maps_bounded_rationality_dk_2003.pdf (Zugegriffen: 20. Dezember 2025).
- Kitel, R. (2022) Ratings und Zertifikate bei Immobilien. ZIA Zentraler Immobilien Ausschuss e.V. Verfügbar unter: https://zia-deutschland.de/wp-content/uploads/2022/03/zia_csr_nl_2022_01_carbon_gap.pdf (Zugegriffen: 14. Januar 2026).
- Kleiber, W., Simon, J., Weyers, G. und Schröter, K. (2007) Verkehrswertermittlung von Grundstücken: Kommentar und Handbuch zur Ermittlung von Verkehrs-, Versicherungs- und Beleihungswerten unter Berücksichtigung von WertV und BelWertV. 5. Aufl., Köln: Bundesanzeiger Verlag.
- Krämer, Dr.B. (2025) „Künstliche Intelligenz in der Immobilienbewertung? Wie maschinelles Lernen und Erklärbare Künstliche Intelligenz den Bewertungsprozess transformieren“, gif im Fokus - Mensch Immobilie!, Januar, S. 54-55.
- Kühl, S., Strodtholz, P. und Taffertshofer, A. (2009) Handbuch Methoden der Organisationsforschung. Verfügbar unter: https://doi.org/10.1007/978-3-531-91570-8_3.
- Lalum, A., López del Rio, L.C. und Villamandos, N.C. (2025) Open Access Article Multi-Dimensional AI-Based Modeling of Real Estate Investment Risk: A Regulatory and Explainable Framework for Investment Decisions. Verfügbar unter: <https://doi.org/10.3390/math13213413> (Zugegriffen: 16. November 2025).
- LeCun, Y., Bengio, Y. und Hinton, G. (2015) Deep learning. *Nature*, 521(7553), S. 436-444. Verfügbar unter: <https://doi.org/10.1038/nature14539> (Zugegriffen: 22. Februar 2026).
- Liang, P. u. a. (2023) Holistic Evaluation of Language Models. arXiv Preprint (arXiv:2211.09110). Verfügbar unter: <https://doi.org/10.48550/arXiv.2211.09110> (Zugegriffen: 21. Februar 2026).
- Lou, J. und Sun, Y. (2024) Anchoring Bias in Large Language Models: An Experimental Study. arXiv Preprint (arXiv:2412.06593). Verfügbar unter: <https://doi.org/10.48550/arXiv.2412.06593> (Zugegriffen: 21. Februar 2026).
- Lowies, G.A., Hall, J.H. und Cloete, C.E. (2016) Heuristic-driven bias in property investment decision-making in South Africa. *Journal of Property Investment & Finance*, 34(1), S. 51-67. Verfügbar unter: <https://doi.org/10.1108/JPIF-08-2014-0055> (Zugegriffen: 21. Februar 2026).
- Luo, J. u. a. (2025) Large Language Model Agent: A Survey on Methodology, Applications and Challenges. Verfügbar unter: <https://arxiv.org/abs/2503.21460> (Zugegriffen: 10. Januar 2026).
- Masterman, T. u. a. (2024) THE LANDSCAPE OF EMERGING AI AGENT ARCHITECTURES FOR REASONING, PLANNING, AND TOOL CALLING: A SURVEY. Verfügbar unter: <https://www.rivista.ai/wp-content/uploads/2024/06/2404.11584v1.pdf> (Zugegriffen: 1. Februar 2026).
- Mitchell, T.M. (1997) Machine Learning. New York: McGraw-Hill. ISBN 0-07-042807-7. Verfügbar unter: <https://www.cs.cmu.edu/~tom/mlbook.html> (Zugegriffen: 22. Februar 2026).

- Moring, A. (2023) Künstliche Intelligenz und Intuition: Robuste und nachhaltige Entscheidungen in digitalen Arbeitswelten. Wiesbaden: Springer Gabler. Verfügbar unter: <https://doi.org/10.1007/978-3-658-42018-5> (Zugegriffen: 22. Februar 2026).
- Nguyen, J.K. (2024) Human bias in AI models? Anchoring effects and mitigation strategies in large language models. Verfügbar unter: <https://doi.org/10.1016/j.jbef.2024.100971>.
- Northcraft, G.B. und Neale, M.A. (1987) Experts, Amateurs, and Real Estate: An Anchoring-and-Adjustment Perspective on Property Pricing Decisions. *Organizational Behavior and Human Decision Processes*, 39(1), S. 84-97. Verfügbar unter: [https://doi.org/10.1016/0749-5978\(87\)90046-X](https://doi.org/10.1016/0749-5978(87)90046-X) (Zugegriffen: 21. Februar 2026).
- Ochs, M. (2022) The Impact of Artificial Intelligence on the Decision-Making Process of a Real Estate Company. Master's Thesis, Universidade Católica Portuguesa. Verfügbar unter: <https://ciencia.ucp.pt/en/studentTheses/the-impact-of-artificial-intelligence-on-the-decision-making-proc> (Zugegriffen: 21. Februar 2026).
- OpenAI (ohne Datum) A practical guide to building agents. Verfügbar unter: <https://openai.com/business/guides-and-resources/a-practical-guide-to-building-ai-agents/> (Zugegriffen: 21. Februar 2026).
- Ouyang, L. u. a. (2022) Training language models to follow instructions with human feedback. Open AI. Verfügbar unter: https://cdn.openai.com/papers/Training_language_models_to_follow_instructions_with_human_feedback.pdf (Zugegriffen: 10. Januar 2026).
- PWC (2023) Emerging Trends in Real Estate - In the eye of the storm. Verfügbar unter: <https://www.pwc.de/de/real-estate/emerging-trends-in-real-estate/2023/emerging-trends-in-real-estate-europe-2023.pdf> (Zugegriffen: 14. November 2025).
- PWC (2024) Kreditausfallrisiken bei Gewerbeimmobilien wachsen: PwC-Whitepaper zu Distressed Real Estate. PwC Deutschland. Verfügbar unter: <https://www.pwc.de/de/real-estate/pwcs-real-estate-institute/kreditausfallrisiken-bei-gewerbeimmobilien-wachsen.html> (Zugegriffen: 21. Februar 2026).
- Rein, D. u. a. (2023) GPQA: A Graduate-Level Google-Proof Q&A Benchmark. Cornell University. Verfügbar unter: <https://arxiv.org/abs/2311.12022> (Zugegriffen: 22. Dezember 2025).
- Richter, C., Rademacher, K., Heck, M., Rock, V. und Will, P. (2024) Transform to Succeed. Studie. Drees & Sommer SE in Zusammenarbeit mit der Technischen Hochschule Aschaffenburg (IIWM – Institut für Immobilienwirtschaft und -management), Stand: Februar 2024. Verfügbar unter: https://www.th-ab.de/fileadmin/th-ab-redaktion/dokumente/Forschung/Projekte/Studie_Transform_to_Succeed_Drees_Sommer_2024.pdf (Zugegriffen: 27. November 2026).
- Salinas, A. und Morstatter, F. (2024) The Butterfly Effect of Altering Prompts: How Small Changes and Jailbreaks Affect Large Language Model Performance. Cornell University. Verfügbar unter: <https://aclanthology.org/2024.findings-acl.275.pdf> (Zugegriffen: 22. Dezember 2025).

- SAP SE (a) (ohne Datum) Was ist generative KI? Beispiele, Anwendungsfälle. SAP. Verfügbar unter: <https://www.sap.com/germany/products/artificial-intelligence/what-is-generative-ai.html> (Zugegriffen: 22. Februar 2026).
- Schroders (2025) Schroders Capital erweitert KI-Toolkit mit virtuellem „Investment Committee Agent“. Verfügbar unter: <https://www.schroders.com/de-de/de/finanzberater/media-center/schroders-capital-erweitert-ki-toolkit-mit-virtuellem-investment-committee-agent/> (Zugegriffen: 17. Januar 2026).
- Seagraves, P. (2023) Real Estate Insights: Is the AI revolution a real estate boon or bane? *Journal of Property Investment & Finance*, 42(2), S. 190-199. Verfügbar unter: <https://doi.org/10.1108/JPIF-05-2023-0045> (Zugegriffen: 21. Februar 2026).
- Segura de la Cal, A., Martínez Raya, A. und Morales-Alonso, G. (2025) Mapping the role of Artificial Intelligence in real estate: A bibliometric and case study analysis. *Journal of Entrepreneurship, Management and Innovation*, 21(3), S. 5-23. Verfügbar unter: <https://doi.org/10.7341/20252131> (Zugegriffen: 21. Februar 2026).
- Simon, H.A. (1955) A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1), S. 99-118. Verfügbar unter: <https://doi.org/10.2307/1884852> (Zugegriffen: 21. Februar 2026).
- Singh, A. u. a. (2023) Behavioural biases in real estate investment: a literature review and future research agenda. Verfügbar unter: <https://www.nature.com/articles/s41599-023-02366-7> (Zugegriffen: 20. November 2025).
- Tian, Z. u. a. (2025) Overconfidence in LLM-as-a-Judge: Diagnosis and Confidence-Driven Solution. Verfügbar unter: <https://arxiv.org/html/2508.06225v2> (Zugegriffen: 17. Januar 2026).
- Trochim, W.M.K. (1989) Outcome Pattern Matching And Program Theory. Cornell University. Verfügbar unter: [https://doi.org/10.1016/0149-7189\(89\)90052-9](https://doi.org/10.1016/0149-7189(89)90052-9) (Zugegriffen: 12. Januar 2026).
- Tversky, A. und Kahneman, D. (1974) Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), S. 1124-1131. Verfügbar unter: <https://doi.org/10.1126/science.185.4157.1124> (Zugegriffen: 21. Februar 2026).
- Tversky, A. und Kahneman, D. (1981) The Framing of Decisions and the Psychology of Choice. *Science*, 211(4481), S. 453-458. Verfügbar unter: <https://doi.org/10.1126/science.7455683> (Zugegriffen: 21. Februar 2026).
- Wang, L. u. a. (2024) A survey on large language model based autonomous agents. Verfügbar unter: <https://doi.org/10.1007/s11704-024-40231-1> (Zugegriffen: 4. Dezember 2025).
- Weidling, J. und Rock, V. (2025) KI im Immobilienfondsmanagement. IIWM-Paper, No. 14. IIWM Institut für Immobilienwirtschaft und -management (TH Aschaffenburg). Verfügbar unter: https://www.iiwm.de/wp-content/uploads/2025/07/IIWM_2025_KI-im-Immobilienfondsmanagement_final.pdf (Zugegriffen: 21. Februar 2026).
- Wooldridge, J.M. (2009) Introductory Econometrics - A Modern Approach. Verfügbar unter: https://commecsinsitute.edu.pk/wp-content/uploads/2024/08/IntroductoryEconometrics_AModernApproach_FourthEdition_Jeffrey_Wooldridge.pdf (Zugegriffen: 15. Januar 2026).

- Xiao, F. und Wang, X.X. (2025) Evaluating the ability of large Language models to predict human social decisions. Verfügbar unter: <https://www.nature.com/articles/s41598-025-17188-7> (Zugegriffen: 25. November 2025).
- Yue, J. und Klein, D. (2025) Benchmarking LLMs on Advanced Mathematical Reasoning. Verfügbar unter: <https://www2.eecs.berkeley.edu/Pubs/TechRpts/2025/EECS-2025-121.pdf> (Zugegriffen: 22. Dezember 2025).
- Zain ul Abdin, S. u. a. (2017) The impact of heuristics on investment decision and performance: Exploring multiple mediation mechanisms. Verfügbar unter: <https://doi.org/10.1016/j.ribaf.2017.07.010> (Zugegriffen: 24. November 2025).
- Zekri, M. (2025) Data-driven literature analysis of artificial intelligence (AI) in real estate research: a bibliometric and machine learning-based approach. Property Management. Verfügbar unter: <https://doi.org/10.1108/PM-04-2025-0059> (Zugegriffen: 21. Februar 2026).
- ZIA Zentraler Immobilien Ausschuss e.V. (2023) ESG Leitfaden für Büroimmobilien. Verfügbar unter: https://zia-deutschland.de/wp-content/uploads/2023/11/zia_leitfaden_esg_buroimmobilien.pdf (Zugegriffen: 16. Januar 2026).

Kurzbiografie der Autoren

Niklas Will

erlangte im Sommer 2021 den Bachelorabschluss im Studiengang Business Administration Dual Degree an der EBS Universität für Wirtschaft und Recht.

Seine berufliche Laufbahn in der Immobilienwirtschaft begann er Anfang 2022 bei Cushman & Wakefield im Bereich Immobilienbewertung, zunächst als Analyst, später als Consultant mit Verantwortung für nationale und internationale Projekte.

Seit dem Wintersemester 2024 studiert er Immobilienmanagement im Master an der Technischen Hochschule Aschaffenburg. Parallel hierzu war er bei Arcadis im Account Management sowie bei AEW im Fund Management tätig.

Sein besonderes Interesse gilt den Anwendungspotenzialen Künstlicher Intelligenz im Investment- und Immobilienmanagement. Diese Schnittstelle zwischen Immobilienpraxis und technologischer Innovation war ein zentraler Impuls für die vorliegende Arbeit.

Verena Rock,

ist seit 2010 Professorin für Immobilieninvestment und seit 2020 Leiterin des Studiengangs Digitales Immobilienmanagement an der Technischen Hochschule Aschaffenburg. Dort leitet sie ebenfalls das Institut für Immobilienwirtschaft und -management (IIWM).

Zuvor war Rock Fondsmanagerin eines international investierten Immobilienspezialfonds bei Morgan Stanley Real Estate und Managerin International Real Estate Advisory bei Corpus Sireo.

Sie studierte Betriebswirtschaft mit den Schwerpunkten Immobilienökonomie und Marketing und promovierte bei Prof. Dr. Karl-Werner Schulte an der EBS Universität.

Seit 2023 ist sie Präsidentin der Gesellschaft für Immobilienwirtschaftliche Forschung (gif e.V.) und hält diverse Aufsichts- und Beiratsmandate in Unternehmen der Immobilienwirtschaft. Darüber hinaus ist sie Jurorin des immobilienmanager Awards in der Kategorie Digitalisierung und Mitherausgeberin der Standardwerke „Praxishandbuch Immobilienfondsmanagement“ und „Next Generation Real Estate“.



IIWM Institut für
Immobilienwirtschaft
und-management
TH Aschaffenburg

Technische Hochschule Aschaffenburg
Würzburger Straße 45
D-63743 Aschaffenburg