



Künstliche oder Menschliche Intelligenz beim Investieren – was sagt die Wissenschaft?

Prof. Dr. Thorsten Hens, Universität Zürich

Künstliche oder Menschliche Intelligenz beim Investieren – was sagt die Wissenschaft?

- Alltagsbeobachtungen
- Trading Experimente
- Schlussfolgerungen

Alltagsbeobachtungen – was sagt die Wissenschaft?

Meine Oma trank jeden Tag Brennnessel Tee
und ist über 90 Jahre alt geworden.
--wegen oder trotz des Brennnessel Tees?



Alltagsbeobachtungen – was sagt die Wissenschaft?

Helmut Schmid hat jeden Tag mehrere Packungen Zigaretten geraucht
und ist über 90 Jahre alt geworden.

-- wegen oder trotz des Rauchens?



Alltagsbeobachtungen – was sagt die Wissenschaft?

Wenn es viele Leute gibt, die Brennnessel Tee trinken und nicht über 90 werden,
Ist die Aussage «Mit Brennnessel Tee wird man uralt» widerlegt.

→ Dies ist nicht der Fall.

Wenn es viele Leute gibt, die stark rauchen und nicht über 90 werden,
Ist die Aussage «Mit Rauchen wird man uralt» widerlegt.

→ Dies ist der Fall.

→ Es kommt auf den richtigen Vergleich an!

Künstliche oder Menschliche Intelligenz beim Investieren – was sagt die Wissenschaft?

- Alltagsbeobachtungen
- **Trading Experimente**
- Schlussfolgerungen

Trading Experimente

Eine KI hat innerhalb kurzer Zeit höhere Renditen als ein Mensch gemacht.

- Was sagt die Wissenschaft?
- Lag es an der Intelligenz der KI oder an der Dummheit des Menschen?
- Oder lag es vielleicht an dem Markt in dieser kurzen Zeit?

Investor Bench

INVESTORBENCH: A Benchmark for Financial Decision-Making Tasks with LLM-based Agent

Haohang Li^{1,*}, Yupeng Cao^{1,*}, Yangyang Yu^{1,*}, Shashidhar Reddy Javaji¹, Zhiyang Deng¹, Yueru He², Yuechen Jiang¹, Zining Zhu¹, Koduvayur Subbalakshmi¹, Guojun Xiong³, Jimin Huang⁴, Lingfei Qian⁴, Xueqing Peng⁴, Qianqian Xie^{4,†}, Jordan W. Suchow¹

¹Stevens Institute of Technology ²Columbia University ³Harvard University ⁴The Fin AI

*These authors contributed equally † Corresponding author: qianqian.xie@yale.edu

- Trading experiment mit 14 LLMs
 - Equities: Oct 20 to May 22
- Cryptos: April 23 to November 23
 - ETFs: Jan 20 to September 20

Design

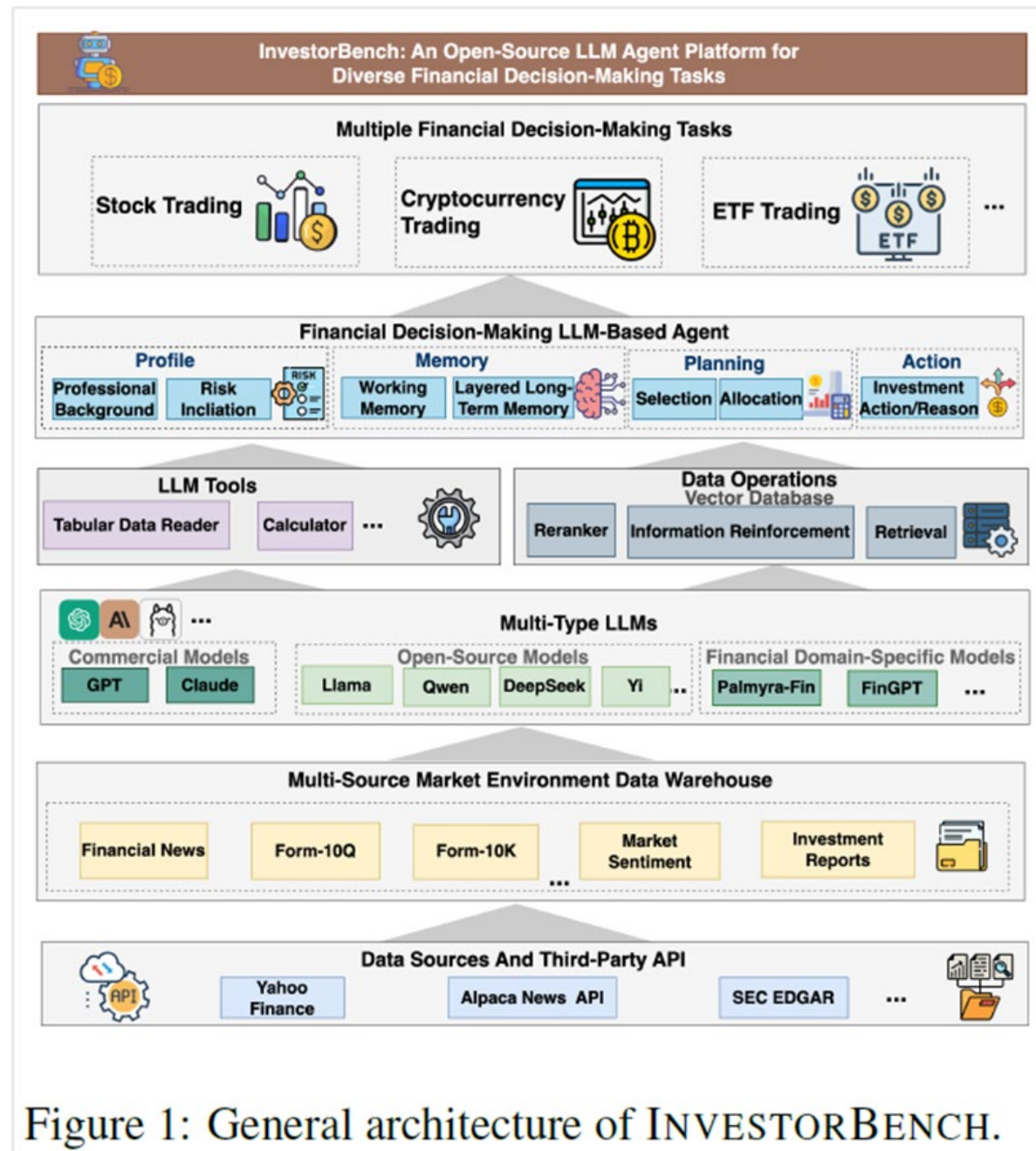
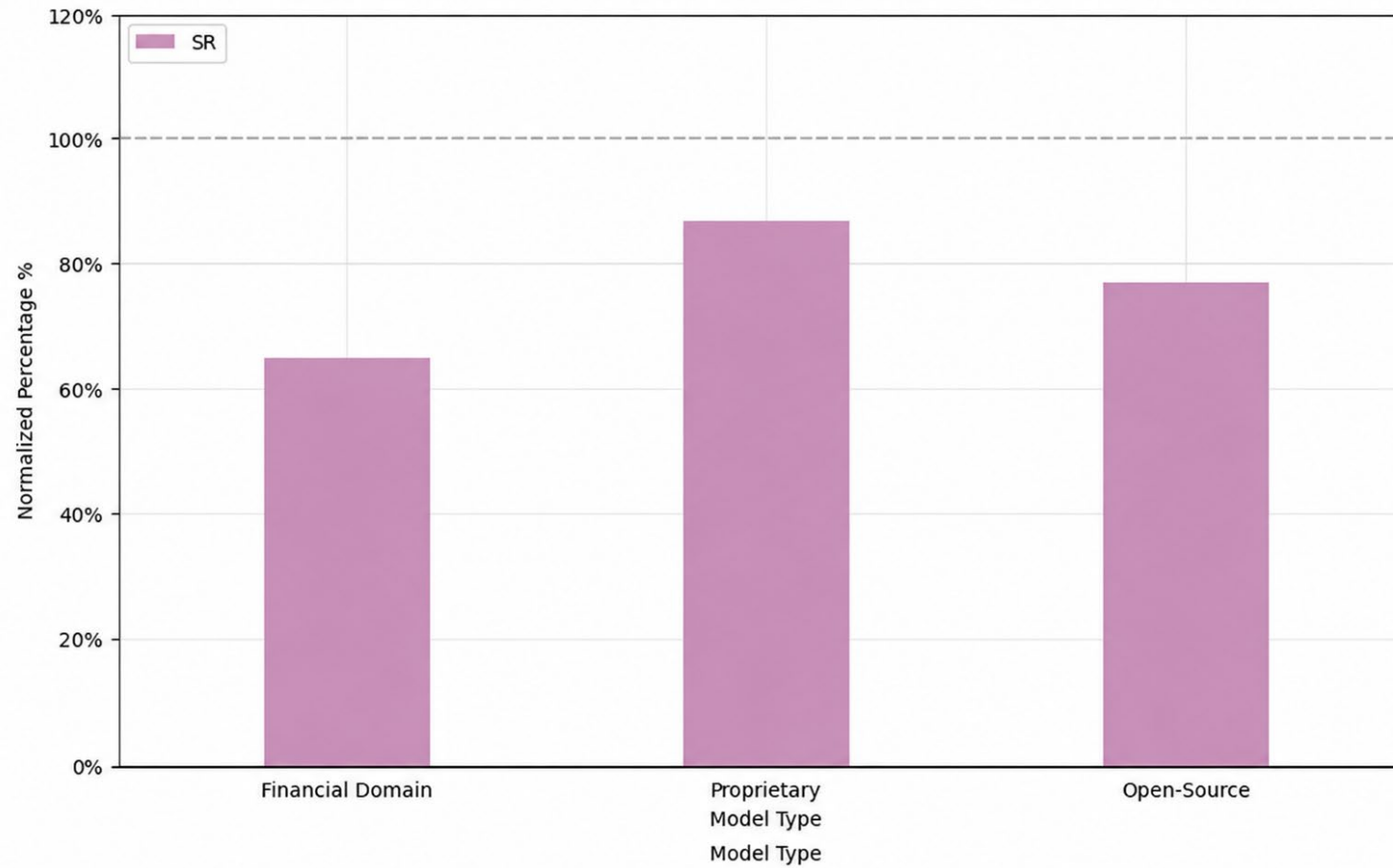
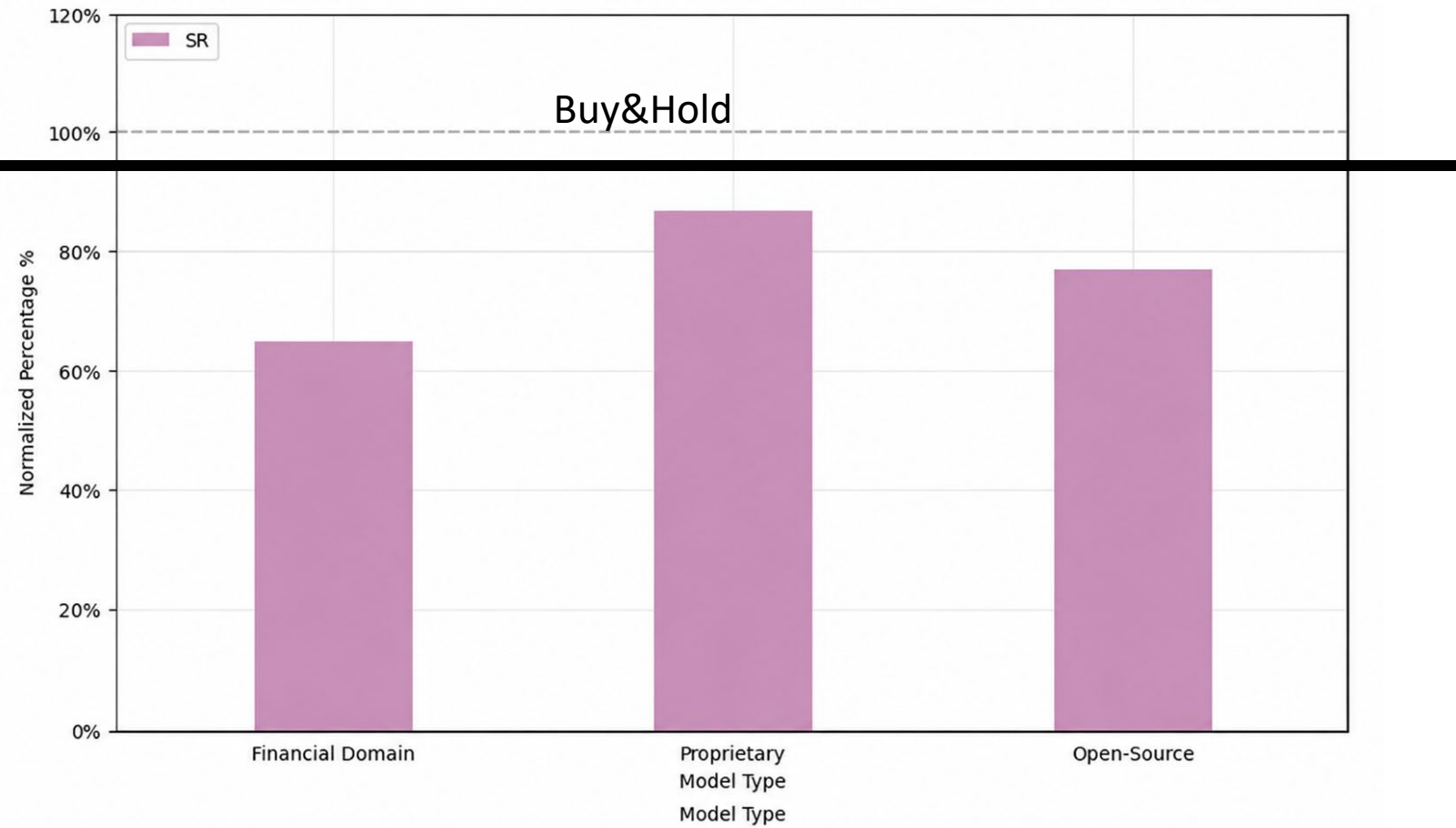


Figure 1: General architecture of INVESTORBENCH.



Investor Bench: Resultat SR



Investor Bench: Resultat SR



Trading Characteristics of ChatGPT



Norwegian School of Economics
Bergen, Fall 2025

Trading Characteristics of ChatGPT

A Study on ChatGPT-5 Thinking in Day-to-Day Trading

Signe Fu Yu Halvorsen Ryen & Tina Nguyen

Supervisor: Thorsten Hens

Master's Thesis, MSc in Economics and Business Administration,
Major in Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Design

– Starting Prompt

- Asset universe (10-15 Stocks from Europe and US)
- Budget (1 Mio Euros)
- 4 Risk and Return Targets 4 portfolios each (V10 A bis R11 B)

– Daily Prompt

- Information on your stock?
- Reallocation?

– Correction

- Budget, Risk and Return target

Anlageentscheide der künstlichen Intelligenz: Ergebnisse

– Operational Mistakes...

Portfolio	Number of recommendations outside the allowed range
V10-A	-
V15-A	-
V20-A	2
R11-A	1
V10-B	2
V15-B	1
V20-B	-
R11-B	-

Table 5: Frequency of recommendations outside the allowed portfolio size.

– ... treten gelegentlich auf.

Anlageentscheide der künstlichen Intelligenz: Ergebnisse

– Violations of Budget Constraint...

Portfolio	Number of cash-insufficient trade proposals	Number of output portfolios with negative cash	Total violations of the budget constraint
V10-A	1	1	2
V15-A	2	3	5
V20-A	1	-	1
R11-A	1	-	1
V10-B	2	2	4
V15-B	2	2	4
V20-B	5	-	5
R11-B	-	-	-

Table 6: Number of budget constraint violations across portfolios.

– ... gibt es häufiger mal.

Anlageentscheide der künstlichen Intelligenz: Ergebnisse

– Compliance with Vola and Target Returns...

Portfolio	Annualized Volatility	Annualized Return	Compliance
V10-A	12.62 %	19.40 %	X
V15-A	13.58 %	22.92 %	✓
V20-A	12.51 %	21.20 %	X
R11-A	10.60 %	10.30 %	X
V10-B	7.95 %	8.63 %	X
V15-B	12.89 %	23.88 %	X
V20-B	12.35 %	18.76 %	X
R11-B	9.98 %	17.93 %	✓

Table 7: Historical volatility and return of day-0 portfolio weights. The annualized return refers to the gross return, based on adjusted close prices.

– ... war sehr schwierig für ChatGPT

Anlageentscheide der künstlichen Intelligenz: Ergebnisse

- Nestle S.A.
- LVMH
- Roche
- Tota I
- Siemens AG
- Iberdrola S.A.
- Novartis AG
- ASML
- Fresenius SE
- National Grid
- Vinci
- Novo Nordisk
- Shell plc
- Unilever PLC
- AstraZeneca
- Danone SA
- Allianz SE
- Airbus SE
- Beiersdorf
- Henkel Vz
- ProcterGamble
- Vodafone
- Deutsche
- Telekom
- Enel SpA
- L'Orea I S.A.
- SAP SE
- Apple Inc.
- Microsoft
- Corp.
- NVIDIA Corp.
- ENGIE
- Koninklijke
- Ahold Delhaiz
- Symrise
- Schneider
- Electric SE
- Orange SA
- Sanofi S.A.
- E.ON SE
- Adyen N.V.
- Alphabet
- Amazon
- Meta
- Munich Re
- Coca-Cola

– Nur 49 von 3500 möglichen Aktien wurden ausgewählt...

Anlageentscheide der künstlichen Intelligenz: Ergebnisse

– Attention based stock selection (Google SVI)...



Figure 4: Distribution of z-scores of the 49 stocks.

– ... eher der Fall.

Anlageentscheide der künstlichen Intelligenz: Ergebnisse

Portfolio	SR	Portfolio	SR
– VIO-A (GPT)	0.918	– VIO-B (GPT)	0.867
– VIO-A (buy&hold)	0.88	– VIO-B (buy&hold)	1.193
– V15-A (GPT)	0.854	– V15-B (GPT)	0.918
– V15-A (buy&hold)	0.855	– V15-B (buy&hold)	1.160
– V20-A (GPT)	0.11	– V20-B (GPT)	0.470
– V20-A (buy&hold)	0.880	– V20-B (buy&hold)	0.923
– RI 1-A (GPT)	1.381	– RI 1-B (GPT)	0.516
– RI 1-A (buy&hold)	1.28	– RI 1-B (buy&hold)	1.363

→ Die initiale Aktienausswahl war sehr gut. → Das Trading nicht sehr hilfreich!

KI hat weniger Behavioral Biases als der Mensch

Average and	Max	Behavior
	Human	ChatGPT
Activity Level	-1% (81%)	3% (5.9%)
Adaptive Behavior	0 % (63%)	-0.7% (-22.5%)
Disposition Effect	13% (100%)	-3.2% (-10%)
Emotional Flexibility	17% (100%)	3.9% (40%)
Greediness	-2% (100%)	0% (15%)
No Regret	-6% (100%)	17.2% (76%)
Temperament	3% (90%)	-5.2% (-48%)
Turnover	1.3% (100%)	5.2% (8.6%)

Künstliche oder Menschliche Intelligenz beim Investieren – was sagt die Wissenschaft?

- Alltagsbeobachtungen
- Trading Experimente
- **Schlussfolgerungen**

KI oder MI – was sagt die Wissenschaft?

- Die Wissenschaft macht korrekte Aussagen und lässt sich nicht blenden
- KI macht nicht weniger sondern andere Fehler als der Mensch
- Insgesamt erzielte die KI keine höhere Performance als der Mensch