

DOCTORAL THESIS



UCAM

UNIVERSIDAD CATÓLICA
DE MURCIA

INTERNATIONAL DOCTORAL SCHOOL

Programa de Doctorado en Ciencias Sociales

An Analysis and Improvement of Robo-Advisory

Author:

Mahan Tahvildari, MSc MBA

Supervisors:

Professor Dr Laura Nieto Torrejón

Professor Dr Joachim Rojahn, CFA

Murcia, September 2025

DOCTORAL THESIS



UCAM

UNIVERSIDAD CATÓLICA
DE MURCIA

INTERNATIONAL DOCTORAL SCHOOL

Programa de Doctorado en Ciencias Sociales

An Analysis and Improvement of Robo-Advisory

Author:

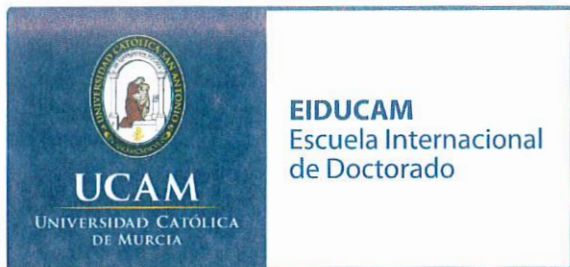
Mahan Tahvildari, MSc MBA

Supervisors:

Professor Dr Laura Nieto Torrejón

Professor Dr Joachim Rojahn, CFA

Murcia, September 2025

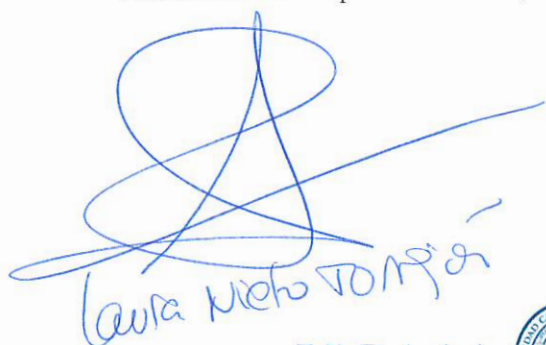


AUTORIZACIÓN DEL DIRECTOR DE LA TESIS
PARA SU PRESENTACIÓN/ AUTHORISATION OF THE THESIS
SUPERVISOR
FOR ITS PRESENTATION

El Dr. D. Laura Nieto Torrejón y el Dr. D. Joachim Rojahn como Directores(1) de la Tesis Doctoral titulada “An Analysis and Improvement of Robo-Advisory” realizada por D. Mahan Tahvildari en el Programa de Doctorado en Ciencias Sociales, autoriza su presentación a trámite dado que reúne las condiciones necesarias para su defensa./ Dr. D. Laura Nieto Torrejón and Dr. D. Joachim Rojahn as Directors(1) of the Doctoral Thesis entitled ‘An Analysis and Improvement of Robo-Advisory’ carried out by D. Mahan Tahvildari in the Doctoral Programme in Social Sciences, authorises its submission for processing as it meets the necessary conditions for its defence.

Lo que firmo, para dar cumplimiento al, Real Decreto 99/2011, de 28 de enero, modificado por el R.D. 576/2023, de 4 de julio, por el que se regulan las enseñanzas oficiales de doctorado/ Which I sign, in compliance with Royal Decree 99/2011, of 28 January, amended by R.D. 576/2023, of 4 July, which regulates official doctoral studies.

En Murcia a 12 de Septiembre de 2025/ In Murcia on 12 of September of 2025



UCAM  **EIDUCAM**
Escuela Internacional
de Doctorado

[Escriba aquí]

(1) Si la Tesis está dirigida por más de un director tienen que constar y firmar ambos/ (1) If the Thesis is supervised by more than one director, both of them have to be stated and signed

ABSTRACT

This dissertation examines and seeks to improve robo-advisory (RA) through an integrated analysis of German providers, a dynamic formal model of the workflow, an operational large-language-model (LLM) prototype, and a benchmark-consistent performance study. It tackles opacity in profiling and portfolios, the absence of a unified dynamic formalism, the lack of a scalable, auditable LLM-driven RA, and limited risk-matched evidence on realised outcomes. Six questions frame the work: RQ1–RQ2 on questionnaires, portfolios and strategies; RQ3 on a mathematical framework; RQ4 on a ChatGPT-4o-based RA; and RQ5–RQ6 on risk-adjusted performance overall and during COVID-19.

Methodologically, a July-2024 census of 45 German RAs is used to construct a profiling taxonomy validated with text analytics features and clustering. Holdings-based mapping is combined with structured analysis of strategy disclosures. A pragmatic two-layer catalogue—eleven equity cores with horizon/environmental, social, and governance (ESG)/knowledge overlays—is proposed. The formal contribution treats profiling, allocation and event-driven rebalancing as a unified control problem. A ChatGPT-4o prototype, built around a 32-item questionnaire and an exchange-traded fund (ETF) universe, generates 100 allocations; strict validation and regression relate questionnaire items to equity share. The performance study assembles monthly indices, assigns portfolios to equity buckets, and compares them with investable benchmarks using standard performance ratios.

Findings show convergence on eight profiling themes but uneven depth: risk, horizons, experience, and finances are widely covered, yet psychometric risk and knowledge checks are often sparse; ESG preferences have grown. Recommended portfolios are dominated by global equity/bond ETFs, with equity share the main risk dial; menus remain narrow relative to questionnaire breadth; mean–variance thinking dominates, and artificial intelligence (AI) use is nascent. The ChatGPT-4o-based prototype produces mixes in which equity exposure is chiefly explained by risk-attitude items and liquidity buffers; validation layers are necessary to prevent accounting/eligibility errors. On realised outcomes, RAs underperform matched benchmarks on average across buckets over 2017–2022 and during COVID-19, with

higher beta/correlation, deeper drawdowns and slower recoveries explaining efficiency gaps; a minority outperform in specific buckets.

The thesis recommends a supervisory profiling taxonomy grounded in validated psychometrics, machine-readable reporting and explicit mapping from client risk grades to product risk labels; event-driven rebalancing with beta/correlation control; and transparent, bucket-matched performance disclosure. Providers should adopt two-layer menus or dynamic instantiation and deploy explainable, auditable LLM pipelines with guardrails and human oversight. For investors, clearer mappings from responses to risk labels and risk-adjusted, benchmarked reporting are essential. Collectively, the programme offers a practical blueprint to make RA more transparent, accountable and performant.

RESUMEN

Esta tesis examina y busca mejorar el asesoramiento robótico (RA) mediante un análisis integrado de los proveedores alemanes, un modelo formal dinámico del flujo de trabajo, un prototipo operativo de modelo de lenguaje grande (LLM) y un estudio de rendimiento coherente con los parámetros de referencia. Aborda la opacidad en la elaboración de perfiles y carteras, la ausencia de un formalismo dinámico unificado, la falta de un RA escalable y auditable impulsado por LLM, y la evidencia limitada de resultados realizados que se ajusten al riesgo. Seis preguntas enmarcan el trabajo: RQ1-RQ2 sobre cuestionarios, carteras y estrategias; RQ3 sobre un marco matemático; RQ4 sobre un RA basado en ChatGPT-4o; y RQ5-RQ6 sobre el rendimiento ajustado al riesgo en general y durante la COVID-19.

Desde el punto de vista metodológico, se utiliza un censo de julio de 2024 de 45 RA alemanas para construir una taxonomía de perfiles validada con funciones de análisis de texto y agrupación. El mapeo basado en carteras se combina con un análisis estructurado de la divulgación de estrategias. Se propone un catálogo pragmático de dos capas: once núcleos de renta variable con superposiciones de horizonte/medioambiental, social y de gobernanza (ESG)/conocimiento. La contribución formal trata la elaboración de perfiles, la asignación y el reequilibrio impulsado por eventos como un problema de control unificado. Un prototipo ChatGPT-4o, construido en torno a un cuestionario de 32 preguntas y un universo de fondos cotizados en bolsa (ETF), genera 100 asignaciones; la validación estricta y la regresión relacionan las preguntas del cuestionario con la participación en acciones. El estudio de rendimiento recopila índices mensuales, asigna carteras a categorías de acciones y las compara con índices de referencia invertibles utilizando ratios de rendimiento estándar.

Los resultados muestran una convergencia en ocho temas de perfil, pero con una profundidad desigual: el riesgo, los horizontes, la experiencia y las finanzas se tratan ampliamente, pero las comprobaciones psicométricas de riesgo y conocimientos suelen ser escasas; las preferencias ESG han aumentado. Las carteras recomendadas están dominadas por ETF de renta variable/bonos globales, siendo la participación en renta variable el principal indicador de riesgo; los menús siguen siendo limitados en relación con la amplitud del cuestionario; predomina el pensamiento de media-varianza y el uso de la inteligencia artificial (IA) es

incipiente. El prototipo basado en ChatGPT-4o produce combinaciones en las que la exposición a la renta variable se explica principalmente por elementos de actitud al riesgo y reservas de liquidez; son necesarias capas de validación para evitar errores contables o de elegibilidad. En cuanto a los resultados obtenidos, los RA obtienen un rendimiento inferior al de los índices de referencia correspondientes en promedio en todos los segmentos durante 2017-2022 y durante la COVID-19, con un beta/correlación más alto, caídas más profundas y recuperaciones más lentas que explican las diferencias de eficiencia; una minoría obtiene un rendimiento superior en segmentos específicos.

La tesis recomienda una taxonomía de perfiles de supervisión basada en psicometría validada, informes legibles por máquina y una correspondencia explícita entre las calificaciones de riesgo de los clientes y las etiquetas de riesgo de los productos; un reequilibrio impulsado por eventos con control de beta/correlación; y una divulgación transparente del rendimiento por categorías. Los proveedores deben adoptar menús de dos capas o instancias dinámicas y desplegar canales LLM explicables y auditables con barreras de protección y supervisión humana. Para los inversores, es esencial disponer de correspondencias más claras entre las respuestas y las etiquetas de riesgo, así como de informes ajustados al riesgo y comparables con índices de referencia. En conjunto, el programa ofrece un plan práctico para que la RA sea más transparente, responsable y eficaz.

KEY WORDS

Robo-Advisory, Robo-Advisor, Financial Advice, Investor Profiling, Portfolio Recommendation, ChatGPT, Artificial Intelligence, Performance Analysis, COVID-19

PALABRAS CLAVES

Asesoramiento robótico, asesor robótico, asesoramiento financiero, perfil de inversor, recomendación de cartera, ChatGPT, inteligencia artificial, análisis de rendimiento, COVID-19.

ACKNOWLEDGEMENTS

I would like to thank everyone who contributed significantly to the completion of my doctoral thesis. Their consistent support and guidance have been invaluable to my academic progress. First and foremost, I would like to express my deepest gratitude to my esteemed doctoral supervisor, Professor Dr Joachim Rojahn. His unwavering dedication, reliability and insightful feedback and guidance have been instrumental in the successful completion of this dissertation. Furthermore, his exemplary guidance and support have contributed to my professional development in the field of economic research. I would also like to thank Professor Dr Laura Nieto Torrejón for her assistance and support throughout my doctoral studies, including all activities, including this dissertation.

I would also like to express my sincere gratitude to the Universidad Católica de Murcia and FOM University of Applied Sciences for the opportunity to write this doctoral thesis. Additionally, I would like to acknowledge the continuous supervision, administration, and support provided by Professor Mercedes Carmona Martínez, Dr. Vinny Flaviana Hyunanda, and Ms. Claudia Schmitz throughout the research process.

Beyond the academic sphere, I would like to thank my family and friends for their unwavering encouragement and patience throughout this journey. In particular, I would like to voice my deepest appreciation to my parents for their unrelenting support, encouragement and patience as I have pursued my academic career.

Mainz, September 10, 2025



Mahan Tahvildari, MSc MBA

DEDICATION

I dedicate this work to my parents, whose patience, sacrifices, and constant faith in me have been my greatest inspiration.

“It is not knowledge, but the act of learning, not possession
but the act of getting there, which grants the greatest
enjoyment.” Carl Friedrich Gauß (1777–1855)

GENERAL TABLE OF CONTENTS

ABSTRACT	5
I - INTRODUCTION	32
1.1. Research Motivation and Relevance	32
1.2. Research Objectives and Questions	35
1.3. Structure of the Dissertation.....	43
II - Theoretical Framework and Literature Review	48
2.1. Definition and Emergence of Robo-Advisory.....	48
2.2. Evolution and Generations of Robo-Advisory	53
2.3. Investor Profiling and Portfolio Recommendations	54
2.4. Performance of Robo-Advisors	74
III - Empirical Analysis of Investor Profiling and Portfolio Strategies by Robo-Advisors in Germany	85
3.1. Data and Methods.....	85
3.1.1. Market Research and Selection of Robo-Advisors	86
3.1.2. Collection of Questionnaires and Portfolio Data	86
3.1.3. Term Frequency–Inverse Document Frequency (TF-IDF) and Cluster Analysis for Categorisation of Questions	88
3.1.4. Modelling of the Building Blocks within the Robo-Advisory Process...	92
3.1.5. Development of a Custom GPT-Based Robo-Advisor	93
3.2. Results.....	98
3.2.1. Questionnaires for Investor Profiling of Robo-Advisors	99
3.2.2. Portfolio Recommendations and Strategies of Robo-Advisors.....	107
3.2.3. TF-IDF and Cluster Analysis of Robo-Advisors' Profiling Questionnaires.....	115

3.2.4.	Mathematical Model of the Robo-Advisory Process	127
3.2.5.	ChatGPT-4o-Based Robo-Advisor.....	134
3.3.	Discussion	146
3.3.1.	Scope, Categories and Typical Items of Investor Profiling	146
3.3.2.	Portfolio Menus and Strategy Patterns.....	156
3.3.3.	A Dynamic Formal Framework for Robo-Advice.....	163
3.3.4.	GPT-Based Prototype and Empirical Patterns.....	168
IV -	Empirical Performance Analysis of Robo-Advisors in Germany.....	176
4.1.	Data and Methods.....	176
4.1.1.	Selection and Data Collection of the Sample	177
4.1.2.	Benchmark Definitions and Assignments to Benchmark Groups	180
4.1.3.	Definitions of Performance Measures.....	184
4.1.4.	Mean–Variance Analysis Method	186
4.1.5.	Correlation Analysis Method.....	187
4.2.	Results for Calculated Performance Measures	188
4.2.1.	Total Period (31/01/2017–28/02/2022).....	188
4.2.2.	COVID-19 Period (31/12/2019–28/02/2022)	193
4.3.	Mean–Variance Analysis with Sharpe Isolines, Capital Market Line and Efficient Frontier	199
4.3.1.	Total Period (31/01/2017–28/02/2022).....	199
4.3.2.	COVID-19 Period (31/12/2019–28/02/2022)	206
4.4.	Correlation Results Between CAPM Key Figures	214
4.4.1.	Total Period (31/01/2017–28/02/2022).....	214
4.4.2.	COVID-19 Period (31/12/2019–28/02/2022)	217
4.5.	Discussion	221
V -	CONCLUSIONS	233
5.1.	Summary and Key Findings	236

5.1.1.	Robo-Advisory Questions in the Profiling Process.....	236
5.1.2.	Portfolios, Strategies and Asset Allocations of Robo-Advisors	238
5.1.3.	Dynamic Mathematical Modelling of the Robo-Advisory Process	240
5.1.4.	Implementation and Analysis of a GPT-Based Robo-Advisor.....	242
5.1.5.	Performance of German Robo-Advisors (31/01/2017–28/02/2022).....	246
5.1.6.	Performance of German Robo-Advisors over the COVID-19 period (31/12/2019–28/02/2022)	247
5.2.	Practical Implications and Recommendations for Action.....	248
5.2.1.	For Regulators and Standard-Setters	248
5.2.2.	For Robo-Advisory Clients and Investors.....	251
5.2.3.	For Robo-Advisory Providers	252
5.3.	Limitations	255
5.3.1.	Chapter III: Profiling and Portfolios—Data Scope, Disclosure Bias, and Interpretation	255
5.3.2.	Chapter IV: Performance—Sampling Frame, Benchmark Design, and Inference	259
5.4.	Outlook and Implications for Research	261
VI -	BIBLIOGRAPHICAL REFERENCES	269
VII -	APPENDIXES	301

ACRONYMS AND ABBREVIATIONS

Abs	Absolute value (function)
AI	Artificial intelligence
AIC	Akaike information criterion
AIRAL	AI Robo Advice Licence
AML	Anti-money laundering
AR	Annualised return
AUM	Assets under management
AVol	Annualised volatility
AVR	Annualised return-to-volatility risk
B2B2C	Business-to-business-to-consumer
BIC	Bayesian information criterion
CAPM	Capital asset pricing model
CCPS-5	Compact core profiling scheme
CHI	Calinski–Harabasz index
CML	Capital market line
Cov	Covariance
COVID-19	Coronavirus disease 2019
CPS-8	Core profiling scheme
CR	Cumulative return
DJIA	Dow Jones Industrial Average
DM	Developed markets
DOSPRT	Domain-specific risk-taking scale
EM	Emerging markets
ER	Excess return
ESG	Environmental, social, and governance

ESMA	European Securities and Markets Authority
ETF	Exchange-traded fund
ETR	Excess Treynor ratio
EURIBOR	Euro Interbank Offered Rate
FHBSS	Filtered hierarchical best-subset selection
FINRA	Financial Industry Regulatory Authority
Fintech	Financial technology
GenAI	Generative artificial intelligence
GL-RTS	Guidelines on risk tolerance and risk-taking
GPT	Generative pre-trained transformer
HCSE	Heteroskedasticity-consistent standard error
HITL	Human-in-the-loop
HJB	Hamilton–Jacobi–Bellman
HMM	Hidden Markov model
IDF	Inverse document frequency
KPI	Key performance indicator
KYC	Know your customer
LLM	Large language model
LOWESS	Locally weighted scatterplot smoothing
LSTM	Long short-term memory
MDD	Maximum drawdown
MiFID II	Markets in Financial Instruments Directive II
MPT	Modern portfolio theory
mSR	Modified Sharpe ratio
MVO	Mean–variance optimisation
NASDAQ	National Association of Securities Dealers Automated Quotations
NAV	Net asset value
NFC	Near-field communication

OLS	Ordinary least squares
QR (code)	Quick response (code)
RA	Robo-advisor / Robo-advisory
RAPM	Risk-adjusted performance measure
RAs	Robo-advisors
RDR	Return-to-drawdown risk
Regtech	Regulatory technology
RMSE	Root mean square errors
RPS	Risk propensity scale
RSS	Root sum of squared distances
RT	Recovery time
S&P 500	Standard & Poor's 500 index
SD	Standard deviation
SFDR	Sustainable Finance Disclosure Regulation
SPTRS	Supervisory profiling taxonomy and reporting standard
SR	Sharpe ratio
SSD	Sum of squared distances
SWOT	strengths, weaknesses, opportunities, and threats
TF	Term frequency
TF-IDF	Term frequency-inverse document frequency
TIPS	Treasury Inflation-Protected Securities
VaR	Value-at-risk
VIF	Variance inflation factor
Wealthtech	Wealth technology
XAI	Explainable artificial intelligence
XGBoost	eXtreme gradient boosting

LIST OF FIGURES, TABLES AND APPENDIXES

LIST OF FIGURES

Figure 1. Typical Workflow Process of Robo-Advisors.....	51
Figure 2. Robo-Advisory Evolution Stages	54
Figure 3. Distribution of Portfolios by Target Equity Exposure.....	111
Figure 4. Merging Procedure of Question Categories ($K = 8-2$).....	116
Figure 5. Elbow Curve of Within-Category SSE by Number of Categories (K)...	118
Figure 6. Distribution of GPT-Recommended Portfolios by Equity Share	136
Figure 7. Mean–Variance Diagram of Portfolios (Total Period)	200
Figure 8. Mean–Variance Diagram with Sharpe Isolines (Total Period).....	201
Figure 9. Distribution of Sharpe Ratios of Portfolios (Total Period).....	202
Figure 10. Distances to CML per Benchmark Group (Total Period).....	204
Figure 11. Diagram of Distances to the CML by Robo-Advisor (Total Period) ...	206
Figure 12. Mean–Variance Diagram of Portfolios (COVID-19 Period)	207
Figure 13. Mean–Variance Diagram with Sharpe Isolines (COVID-19 Period) ...	208
Figure 14. Distribution of Sharpe Ratios of Portfolios (COVID-19 Period)	209
Figure 15. Distances to CML per Benchmark Group (COVID-19 Period)	211
Figure 16. Diagram of Distances to the CML by Robo-Advisor (COVID-19 Period)	213
Figure 17. Jensen’s α vs. ρ of CAPM with LOWESS (Total Period).....	216
Figure 18. Jensen’s α vs. β of CAPM with LOWESS (Total Period).....	217
Figure 19. Jensen’s α vs. ρ of CAPM with LOWESS (COVID-19 Period)	219
Figure 20. Jensen’s α vs. β of CAPM with LOWESS (COVID-19 Period)	220

LIST OF TABLES

Table 1. Sample Overview of Selected Robo-Advisors in Germany	99
Table 2. Representative Questions by Content Category	103
Table 3. Descriptive Statistics of Question Counts by Content Category	105
Table 4. Robo-Advisors by Question Count Intervals per Category	106
Table 5. Descriptive Statistics of Portfolios per Quartile	108
Table 6. Descriptive Statistics of Portfolios by Target Equity Exposure	110
Table 7. Portfolio Strategies Disclosed by Robo-Advisors	112
Table 8. Summary Metrics from TF-IDF Categorisation Variants ($K = 8-2$)	119
Table 9. Centroid Norms, Cardinality, and WCSSE per Category ($K = 8-2$)	120
Table 10. Top Five Typical Questions per Category ($K = 8$)	122
Table 11. Top Five Typical Questions per Category ($K = 4$)	125
Table 12. Descriptive Statistics of Recommended Portfolio Composition	135
Table 13. Descriptive Statistics of Equity Share by Risk Profile	137
Table 14. Independent Variables for Regressions Models of Equity Exposure	138
Table 15. Bilinear OLS Models Explaining Portfolio Equity Exposure	139
Table 16. Trilinear OLS Models Explaining Portfolio Equity Exposure	142
Table 17. 4-Linear OLS Models Explaining Portfolio Equity Exposure	144
Table 18. Key Findings from Robo-Advisors' Profiling Questionnaires	154
Table 19. Key Findings from Robo-Advisors' Portfolios and Strategies	162
Table 20. Key Findings from the Robo-Advisory Model	167
Table 21. Key Findings from the GPT-Based Robo-Advisor	173
Table 22. Sample of Robo-Advisors and Portfolios for Performance Analysis	179
Table 23. Benchmark Portfolio Definitions	181
Table 24. Definition of Benchmark Groups with Portfolios	183
Table 25. Descriptive Statistics of Performance Measures (Total Period)	189
Table 26. Robo-Advisors Outperforming Benchmarks (Total Period)	191
Table 27. Portfolios Outperforming Benchmarks (Total Period)	192

Table 28. Portfolios Outperforming per Benchmark (Total Period)	193
Table 29. Descriptive Statistics of Performance Measures (COVID-19 Period) ...	194
Table 30. Robo-Advisors Outperforming Benchmarks (COVID-19 Period).....	196
Table 31. Portfolios per RA Outperforming Benchmarks (COVID-19 Period)	197
Table 32. Portfolios Outperforming per Benchmark (COVID-19 Period).....	198
Table 33. Descriptive Statistics of Sharpe Ratios (Total Period)	203
Table 34. Descriptive Statistics of the Distances to the CML (Total Period)	204
Table 35. Distance Metrics to the CML by Robo-Advisor (Total Period).....	205
Table 36. Descriptive Statistics of Sharpe Ratios (COVID-19 Period).....	210
Table 37. Descriptive Statistics of the Distances to the CML (COVID-19 Period)	211
Table 38. Distance Metrics to the CML by Robo-Advisor (COVID-19 Period)	212
Table 39. Correlation Tests Between CAPM Metrics (Total Period)	214
Table 40. Bootstrapped Spearman's ρ of the CAPM Metrics (Total Period).....	215
Table 41. Correlation Tests Between CAPM Metrics (COVID-19 Period)	218
Table 42. Bootstrapped Spearman's ρ of the CAPM Metrics (COVID-19 Period)	218
Table 43. Key Findings from the Performance Analysis	228

APPENDIXES

APPENDIX 1. Proof of the TSSE–Global Mean Norm Identity.....	301
APPENDIX 2. Blacklist of Tokens for TF-IDF Preprocessing	301
APPENDIX 3. Python Script for TF-IDF Question Analysis.....	301
APPENDIX 4. Top Five Typical Questions per Category (K = 5).....	310
APPENDIX 5. Questionnaire for GPT-Based Robo-Advisor	311
APPENDIX 6. Investment Universe for the GPT-Based Robo-Advisor	314
APPENDIX 7. Logos of 45 Robo-Advisors in Germany	314
APPENDIX 8. Portfolios by Equity Share of the GPT-Based Robo-Advisor	315
APPENDIX 9. Simple OLS Regression Results of the GPT-Based Robo-Advisor	316
APPENDIX 10. Proof of the Distance from the Capital Market Line	317
APPENDIX 11. Profiling Sample for the GPT-Based Robo-Advisor	318

I – INTRODUCTION

I - INTRODUCTION

1.1. RESEARCH MOTIVATION AND RELEVANCE

Traditional investment advice is based on personal contact and the trust that develops between the client/investor and the advisor, a model that is time-consuming and costly due to personnel costs, individual advice, and the numerous financial products available, and therefore appeals primarily to wealthy clients. Research raises concerns about this effectiveness and value of financial advisors, citing high fees, misaligned incentives, and advisors' own investment errors (Allie et al., 2016; Foerster et al., 2017; Hackethal et al., 2012; Hoechle et al., 2017, 2018; Linnainmaa et al., 2021; Marsden et al., 2011; Montmarquette, 2015; West, 2012). The digital transformation in wealth management has led to the emergence of robo-advisors (RAs) as a new type of provider of digital and automated asset management for retail investors (Fisch et al., 2019; Iannarone, 2018; Jung, Dorner, Glaser, et al., 2018; Jung, Dorner, Weinhardt, et al., 2018; Jung et al., 2019; Sironi, 2016). These RAs are spreading and growing worldwide and are causing significant disruption to the entire wealth management sector. The total assets under management (AUM) by RAs worldwide is estimated at USD 2.06 trillion for 2025, of which USD 1.67 trillion is in the United States (Statista, 2025).

The robo-advisory (RA) process defined and described in Section 2.1 consists of two subprocesses: profiling, which involves an online questionnaire that uses an algorithm to assign a risk profile, and asset allocation, which recommends a portfolio tailored to the client's goals and preferences and then monitors and adjusts this portfolio (Iannarone, 2018; Jung, Dorner, Glaser, et al., 2018; Jung, Dorner, Weinhardt, et al., 2018; Jung et al., 2019; Sironi, 2016). While early studies praised the automation and low fees of RAs compared to human advisors, critics mentioned the generic nature of the profiling phase, as many platforms ask only a few questions and thus steer investors towards highly simplified, conservative portfolios with limited risk differentiation (Arenas-Parra et al., 2024; Blaschke & Kriebel, 2021; Brenner & Meyll, 2020; Faloon & Scherer, 2017; Scherer & Lehner, 2023; So, 2021; Tertilt & Scholz, 2018; Uhl & Rohner, 2018; Zhang et al., 2021).

Although robo-advisory is cost-effective, adoption and trust remain inconsistent, while modern portfolio theory (MPT) remains dominant and artificial intelligence (AI) engines are barely spreading (Beketov et al., 2018; D’Acunto & Rossi, 2021, 2023; Fisch et al., 2019; Jung et al., 2019; Kaya & von Martens, 2020; Tsai & Chen, 2022). Ad hoc scores with weak psychometrics and little balance sheet information from investors dominate profiling. Demand for depth is segmented, and regional typologies diverge. There is also a lack of a market-wide, validated question taxonomy (Blaschke & Kriebel, 2021; Faloon & Scherer, 2017; Scherer, 2017; Scherer & Lehner, 2022; So, 2021; Tertilt & Scholz, 2018). Currently, there is limited up-to-date information regarding the broad and fragmented RA market in Germany, the profiling process, and the specific question categories and typical questions posed to investors (Faloon & Scherer, 2017; Kordela, 2018; Oehler & Horn, 2024; Scherer, 2017; Scherer & Lehner, 2022).

Furthermore, the number of portfolios created, their composition and the assets they contain remain unclear, as does the evidence-based composition of holdings—including sustainable constituents, real estate, cryptocurrencies—and the translation of question depth into portfolio breadth (Faloon & Scherer, 2017; Kordela, 2018; Oehler & Horn, 2024; Scherer, 2017; Scherer & Lehner, 2022, 2025). There are no up-to-date findings on the application of specific portfolio strategies and techniques used by German RAs, such as the use of AI in portfolio management (Beketov et al., 2018; Jung, Dorner, Glaser, et al., 2018; Jung, Dorner, Weinhardt, et al., 2018; Jung et al., 2019).

While certain publications have previously explored the integration of AI within the domain of financial advisory, a concrete, scalable implementation of an large language model (LLM)-driven robo-advisor remains underdeveloped, and there is still limited knowledge regarding the implementation of comparisons of equity ratios/risk levels based on investor responses (Ali & Aysan, 2023; Ko & Lee, 2024; Lu et al., 2023; Oehler & Horn, 2024; Pelster & Val, 2024; Romanko et al., 2023; Tahvildari, 2025d). Existing works articulate what is required to be solved, but not how to build and test it at scale. Lo and Ross (2024) set a research agenda—domain specificity, individual tailoring, trustworthiness, and regulatory conformity—but stop short of an operational system. Oehler and Horn (2024) benchmarked an early version of the generative pre-trained transformer (GPT) tool “ChatGPT-3.x”,

developed by OpenAI (2025), on only three stylised profiles, demonstrating potential but not generalisability. While they used early versions of ChatGPT, newer, rapidly evolving generative artificial intelligence (GenAI) models offer an increasing number of potential applications that can be explored (Artificial Analysis Inc., 2025; Choi & Chang, 2025; Martínez-Cruz et al., 2025; OpenAI, 2025; Restrepo Echavarría, 2025; Zhao et al., 2025). Fieberg et al. (2025) expanded the scope to encompass more LLMs and profiles, concluding that LLMs have the capacity to generate recommendations with competitive historical performance. However, they did not implement a market-realistic pipeline that ingests actual RA questionnaires and maps answers to an investable financial instrument universe with auditable guardrails. It is crucial to emphasise that explainable AI (XAI) and regulatory-aligned justification are of paramount importance in practice. As specified by Vilone et al. (2024), the required explanations to be produced by RA systems are specified, but these requirements are not coupled to a full LLM portfolio constructor. The development of an LLM-based RA with, for instance, OpenAI's ChatGPT should also serve to further develop RAs (OpenAI, 2025), using state-of-the-art technology such as LLMs to ensure their competitiveness. This is particularly pertinent in the context of the question as to what added value traditional RAs with static profiling and fixed, pre-built portfolios have in the age of artificial intelligence.

Additionally, there is a gap between the current understanding of the business-operational robo-advisory process (Beketov et al., 2018; Faloon & Scherer, 2017; Jung, Dorner, Glaser, et al., 2018; Jung, Dorner, Weinhardt, et al., 2018; Jung et al., 2019; Scherer & Lehner, 2022) and financial mathematical methods for solving portfolio allocation and optimisation problems (Bourgeron et al., 2018; Cui et al., 2022; Khosravi, 2024; Monoyios, 2010, 2018; Perrin & Roncalli, 2020; Tahvildari, 2021). Important business-operational process components of the RA process include, for example, investor profiling with assessment based on questionnaires, portfolio recommendation with specific asset allocation, and other ongoing sub-processes during the investment period, such as the application of trading strategies, rebalancing, monitoring, and dynamic changes to the investor profile due to changes in preferences, which in turn can lead to new portfolio adjustments. Rigorous modelling of the entire process and formalisation of the optimisation process bridges this knowledge gap.

The evidence regarding RA performance is mixed. Helms et al. (2022) and Puhle (2019) found fee-adjusted underperformance, whereas Tao et al. (2021) reported average outperformance in the US, and Lee (2021) documented wide dispersion, with few positive alphas and still average underperformance. Limitations across studies, such as short time periods, simulated or secondary data, opaque portfolio compositions, cross-regional heterogeneity, and benchmarks that are not matched to equity exposure buckets, constrain inference (Helms et al., 2022; Lee, 2021; Puhle, 2019; Tao et al., 2021). The behaviour of German RAs during crises such as the market crash caused by COVID-19 measures is underexplored. While potential improvement levers emerge, they lack causal, risk-class-specific evidence in Germany regarding fees, rebalancing and transparency (Helms et al., 2022); implementation and factor tilts (Puhle, 2019); disciplined automation (Tao et al., 2021); and distribution channels and operations (Lee, 2021). Hence, a gap remains in understanding the of German RAs across different equity risk classes and over an extended time frame, particularly during drawdown phases like the COVID-19 crisis.

1.2. RESEARCH OBJECTIVES AND QUESTIONS

The dissertation comprises eight interconnected objectives that collectively map, explain, and evaluate the practice of robo-advisory in Germany. This goal is achieved by combining two empirical programmes: one focusing on investor profiling, portfolio recommendations and strategies (including formal modelling of the RA process and a ChatGPT-based RA prototype implementation), and the other focusing on realised performance. Chapter III aims to address and answer the following five central research questions:

- RQ1: What and how many questions and content categories are asked by the robo-advisors active in Germany as part of investor profiling, and what differences, similarities, and typical questions can be gained from this?
- RQ2: What portfolio strategies and specific compositions of asset classes and weightings are proposed by the robo-advisors active in Germany based on the results of investor profiling, and what are the differences, similarities, and typical portfolios that can be derived from this?

RQ3: How can the robo-advisory process with all its building blocks be formalised and dynamically extended into a coherent mathematical framework that generalises prior static formulations and equips providers with a foundation for subsequent algorithmic and operational development?

RQ4: How can a GPT-based robo-advisor be implemented using a suitable questionnaire and investment universe, and what empirical patterns emerge in an investor deployment regarding portfolio composition and the relationships between profiling variables (risk-preference items and household balance-sheet indicators) and equity exposure as a risk proxy?

RQ1 aims to identify the questions that German RAs ask investors and how these are organised. To this end, Subsection 3.2.1 conducts a market census of 45 active providers, harvesting the full onboarding questionnaires and coding 687 distinct questions into a data-driven taxonomy. The central objective here is to find out which questionnaire structure, most important representative questions and content categories are queried in the RA profiling process and what conclusions can be drawn from this. Moreover, the objective is to establish a comparable language for profiling across firms, test whether recent critiques that German RAs neglected household balance sheets and rigorous risk psychometrics still hold, and expand previous market analyses of RA questions using a more up-to-date and broader data basis of the German RA market (Faloon & Scherer, 2017; Kordela, 2018; Oehler & Horn, 2024; Scherer, 2017; Scherer & Lehner, 2022).

The methods combine qualitative content analysis, descriptive statistics and quantitative text analytics (term frequency—inverse document frequency [TF-IDF] feature construction over the full question corpus), as well as clustering techniques, which are evaluated using elbow plots and various distance metrics, including within-cluster errors and the Calinski–Harabasz index (CHI) (see also Subsection 3.2.3). This validates the chosen categories and reveals themes such as risk, experience and financial situation. The objective is not only to achieve descriptive harmonisation but also to motivate the replacement of ad hoc risk items with validated psychometric scales (e.g., GL-RTS, DOSPERT, RPS) and to align profiling with EU and MiFID II suitability rules (ESMA, 2009, 2016, 2023; European Union, 2014; Grable & Lytton, 1999; Weber et al., 2002; Zhang et al., 2019).

RQ2 investigates the portfolios recommended by RAs following profiling and whether the portfolios' breadth and construction mirror the questionnaire's depth. The objective is to characterise asset classes, weights, and the granularity of risk bands across 405 live strategies. The goal of this study is to ascertain the precise portfolio compositions of the portfolios recommended by the 45 RAs active on the German market. This will be achieved through comprehensive holdings analysis, with the aim of identifying investment universes and examining newer instruments such as environmental, social, and governance (ESG) instruments and alternative investments such as cryptocurrencies that have not been identified or considered in previous studies.

Furthermore, the explicit equity proportions and the relevant risk drivers of the portfolios must be identified and grouped into different risk classes. Identifying explicit equity shares and risk drivers allows clearer assignment of matching benchmarks and reduces misclassification that has hampered prior comparisons. This is a novel element of the study, as earlier studies did not address this issue or only addressed it to a diluted extent due to the mixing of asset classes (Faloon & Scherer, 2017; Kordela, 2018; Oehler & Horn, 2024; Scherer, 2017; Scherer & Lehner, 2022, 2025).

Subsection 3.2.2 thus employs a holdings-based approach to map each portfolio, taking into account equity exposure bins. This method is consistent with the market's own risk labelling and the risk-bucket logic later used for performance benchmarking in Chapter IV. The method is both descriptive and disciplined in nature. This granular, holdings-based mapping enhances market transparency, enables regulators and investors to evaluate suitability and implementation quality more effectively, and gives providers actionable insights to improve product design and execution. Portfolio constituents are grouped by asset class (equities, bonds, commodities, real estate, crypto), ESG variants are flagged, and the distribution of equity shares across providers is measured. Furthermore, the objective extends to explaining an observed tension: abundant questionnaire dimensions but relatively few model portfolios. To address this, Subsection 3.3 proposes a two-layer catalogue comprising 11 "core" equity bands overlaid with horizon, ESG, and knowledge/frequency variants. This is proposed as a pragmatic lower bound that honours profiling without proliferating static, pre-defined

portfolios. Practically, the catalogue reduces operational complexity for providers while preserving meaningful personalisation; academically, it creates a parsimonious framework for subsequent causal research that links disclosed strategies and implementation choices to realised performance.

The analysis of sustainability preferences, which were not examined in the aforementioned studies on RAs, will also be pursued in the analysis of RA questions and the composition analysis. The aim is to examine whether and how sustainability preferences are reflected in portfolio compositions. A detailed analysis is essential to protect investors from confusion and misallocation and to ensure compliance with Sustainable Finance Disclosure Regulation (SFDR) obligations to integrate sustainability risks and provide product-specific disclosures (Regulation (EU) 2019/2088) (European Union, 2019). The relevance of the analysis for the implementation of sustainability requirements in RA has also already been highlighted in academic literature (Barile et al., 2025; Chen et al., 2025; Faradynawati & Söderberg, 2022; Shan et al., 2022).

Furthermore, RQ2 targets the strategic “how” behind those portfolios by documenting which construction/management approaches providers disclose and with what prevalence. The objective is to trace the field’s method mix, from modern portfolio theory (MPT) to hybrid human oversight and emerging AI. Subsection 3.2.2 implements a structured content analysis of public disclosures, coding for mentions of MPT (mean–variance), buy-and-hold, hedging/tactical overlays, Black–Litterman, value-at-risk (VaR), factor investing (smart beta), core–satellite, risk parity, hybrid governance, and AI techniques. The findings should demonstrate the current spectrum, which has thus far concentrated on MPT, but novel active overlays such as ESG switches (impact investing), active algorithmic trading, and hybrid and AI techniques should also be compared with the results of previous studies (Beketov et al., 2018; Jung, Dorner, Glaser, et al., 2018; Jung, Dorner, Weinhardt, et al., 2018; Jung et al., 2019).

Analysing robo-advisors’ portfolio compositions and methods is crucial due to the rapid growth of the RA market and its current opacity. Systematic mapping of assets, allocation strategies, and algorithms can reveal research gaps, inform regulation, and improve the design of client-facing profiling and allocation processes (Arenas-Parra et al., 2024; Helms et al., 2022). The correlation between methodological sophistication and larger AUM indicates a discernible discrepancy

between deployed approaches and state-of-the-art research (Beketov et al., 2018). Given that trust, supervisory control, information quality, and regulatory assurances exert a significant influence on user adoption, it is imperative to comprehend and disclose the composition of concrete portfolios, equity exposures (risk levels), and strategies employed by RAs to ensure the development of trustworthy, scalable products and facilitate actionable vendor and policy responses (Cao et al., 2025; Cheng et al., 2019; Nourallah et al., 2023). Although RQ2 does not itself link disclosed methods to realised performance, documenting the field's method mix establishes the empirical foundation for subsequent tests that will determine which allocation and implementation choices improve outcomes, thereby deepening understanding of the asset-allocation and management subprocess that ultimately affects investor welfare and systemic risk.

RQ3 has a design objective: to formalise the end-to-end RA workflow as a coherent mathematical framework that generalises static formulations. Accordingly, in Subsection 3.2.4, the objective is to build an interdisciplinary bridge between the business-operational robo-advisory sub-processes (profiling, asset allocation and management, etc.) researched to date in previous studies (Beketov et al., 2018; Faloon & Scherer, 2017; Jung, Dorner, Glaser, et al., 2018; Jung, Dorner, Weinhardt, et al., 2018; Jung et al., 2019; Scherer & Lehner, 2022) and the financial mathematical strategies of portfolio allocation and stochastic-analytical optimisation problems from other studies (Bourgeron et al., 2018; Cui et al., 2022; Khosravi, 2024; Monoyios, 2010, 2018; Perrin & Roncalli, 2020; Tahvildari, 2021). Furthermore, the results previously obtained from this thesis by researching the profiling process and portfolios, strategies, and the management of the 45 RAs in Germany are to be considered, with integrations of dynamic processes (such as market and customer preferences) being enabled to allow for more flexible modelling.

RQ4 seeks to design and implement a GPT-based RA prototype, encompassing a profiling questionnaire and an exchange-traded fund (ETF) universe (see Subsection 3.2.5). The objective is to generate personalised recommendations based on a large (e.g., 100) random sample of investor profiles. Despite advances in AI for financial advice, the field still lacks a concrete, scalable LLM-driven RA implementation that operationalises market-realistic

questionnaires, systematically maps investor responses to equity-ratio/risk-level decisions within investable universes, and provides legally explainable outputs—leaving mostly conceptual roadmaps and limited-scope demonstrations rather than audited pipelines (Ali & Aysan, 2023; Bose & Srinivasan, 2024; Fieberg et al., 2025; Ko & Lee, 2024; Lo & Ross, 2024; Lu et al., 2023; Oehler & Horn, 2024; Pelster & Val, 2024; Romanko et al., 2023; Tahvildari, 2025c). A GPT-based RA tested across a large, diverse set of investor profiles would address this gap by delivering adaptive, explainable, compliance-aware portfolio construction at scale and clarifying the incremental value of LLM-enabled advice relative to static, pre-built portfolios, thereby strengthening provider competitiveness in the AI era.

Hence, RQ4 aims to describe portfolio composition (equity/bond shares, use of ESG and alternatives) and quantify how profiling variables relate to equity exposure as a risk proxy. Methods include strict output validation, composition analysis of assets, and multilinear regression of equity share on the questionnaire's items to explore the RA's asset allocation based on the profiles. The intention is also to provide a blueprint for RAs that do not currently use AI, particularly LLMs, but can integrate this technology into their RA process.

Chapter IV addresses the following three key research questions:

RQ5: How do RAs in Germany perform in terms of their (risk-adjusted) performance compared to benchmarks between January 2017 and February 2022?

RQ6: How do RAs in Germany perform in terms of their (risk-adjusted) performance compared to benchmarks in times of crisis, taking the period between the outbreak of the COVID-19 pandemic in December 2019 and the beginning of the war in Ukraine in February 2022?

RQ5 and RQ6 aim to address the gap in data-based performance differences between a broader sample of RAs with many portfolios grouped according to numerous equity exposure weightings to capture different risk levels and compare them with each other and against matching benchmarks. The reason for this is that previous studies had methodological data-based limitations such as smaller samples, significantly shorter time series, simulated rather than realised data, heterogeneous groups from different regions, and only certain equity exposure groups (certain risk levels) (Helms et al., 2022; Lee, 2021; Puhle, 2019; Tao et al.,

2021). This is also due to the fact that the data available on RAs is limited due to their relatively recent history, and in addition, numerous RAs on the market are not transparent and do not publish public price/performance data for their portfolios. In addition, the gap in performance for German RAs for different equity risk classes, RAs and a longer time period, including drawdown phases such as during the COVID-19 crisis, is addressed by RQ5 and RQ6. RQ6 aims to specifically examine the benchmarked performance of RAs in market phases characterised by crises. Therefore, the present study will encompass both the 2020 stock market crash, precipitated by the outbreak of COVID-19, and the subsequent recovery phase to measure its impact on RA performance (Huber et al., 2021; Mazur et al., 2021; Song et al., 2022). The onset of the war in Ukraine in 2022 signifies the commencement of another bear market phase and a structural break, accompanied by sharp central bank interventions, which is employed as the termination of the investigation interval in the context of the COVID-19 period analysis (Ahmed et al., 2022; Bounou & Yatié, 2022; Federal Reserve Bank, 2022).

Additionally, RQ5 and RQ6 address the performance gap of German RAs across various equity risk classes and over an extended time period, including drawdown phases like those experienced during the COVID-19 crisis. The method builds a transparent monthly dataset of 62 portfolios from 12 RA providers (later expanding to 86/16 for COVID-19), assigns each to one of ten equity-exposure benchmark groups, and constructs passive, low-cost reference portfolios from EUR-hedged MSCI ACWI equity and a EUR-hedged global bond sleeve with a 3% threshold rebalancing rule (see Subsections 4.1.1–4.1.2). Performance measures include annualised return/volatility, maximum drawdown, return-to-drawdown and return-to-volatility ratios, the modified Sharpe ratio with small-sample correction, and Capital Asset Pricing Model (CAPM) alpha, beta and Treynor (see Subsection 4.1.3).

Additionally, more detailed mean–variance analyses according to dominance rules, comparisons of Sharpe isolines and distances to the capital market line, and efficient frontiers will be examined (see Section 4.3). Correlation analyses between CAPM ratios aim to provide additional information about the relationships between risk-adjusted performance ratios within the model (see Section 4.4). RQ6 narrows to the COVID-19 shock (31/12/2019–28/02/2022), holding methods

constant to isolate crisis behaviour and recovery dynamics (Huber et al., 2021; Mazur et al., 2021).

Against this backdrop, RQ5 and RQ6 provide a neutral, benchmark-consistent evaluation framework that stakeholders can use to assess the efficacy of RA designs in translating profiling inputs into portfolios that genuinely enhance clients' risk-return positions within clearly defined equity-exposure buckets. For regulators and standard-setters, such evidence can inform the establishment of proportional disclosure rules and supervisory expectations around the selection of benchmarks, the publication of drawdown/recovery metrics, and the use of risk-adjusted performance measures in investor communications. These are areas where prior work has highlighted methodological heterogeneity and comparability problems (Helms et al., 2022; Lee, 2021; Puhle, 2019; Tao et al., 2021). For clients and their advisors, the same framework under RQ5 and RQ6 enables due-diligence practices that focus on matched-bucket benchmarking, crisis-phase behaviour, and preference-consistent risk management rather than headline point returns. It also clarifies when a low-cost benchmark is a more suitable comparator if a provider largely tracks market risk. For RA providers, the analyses supply design-relevant diagnostics—how rebalancing rules, beta/correlation control, and portfolio architecture manifest in drawdown depth, recovery speed, and risk-adjusted ratios across buckets and through stress—thereby creating a disciplined basis for model audit, fee policy, and product positioning without prescribing any one “house view”.

In terms of research, RQ5 and RQ6 extend the extant empirical literature by offering realised, long-horizon, bucketed performance for a single jurisdiction, which can be replicated cross-country, linked to disclosed strategy taxonomies (Chapter III), and decomposed into drivers (fees, rebalancing cadence, beta management) using future transaction-level data. Furthermore, they facilitate the exploration of stress-testing protocols and comparative studies of calendar- versus threshold-based rebalancing, crisis-conditioned alpha generation, and the interaction between portfolio design choices and regulatory disclosure regimes. This contributes to the development of a cumulative evidence base for robo-advisory performance analysis.

The results of the performance analysis in Chapter IV show that, on average, the RAs largely underperform their benchmarks in the corresponding risk classes

(equity buckets) on a risk-adjusted basis, both over the entire period of the analysis and during the crisis phase, using the example of the phase since COVID-19. It is evident that none of the RAs examined manage to dominate in all benchmark groups (equity exposure groups) according to any risk-adjusted performance metric. Conversely, in the majority of cases concerning performance metrics, with a few exceptions, the benchmark demonstrates superiority across various equity groups. However, a minority of RAs have managed to outperform their benchmark with at least some of their portfolios or to limit risks through lower drawdowns. Furthermore, the performance of RAs differs significantly both between each other and within the portfolios of a single RA. Further details on the findings can be found in the discussion in Section 4.5.

1.3. STRUCTURE OF THE DISSERTATION

The rest of the thesis is structured as follows: Chapter II establishes the conceptual foundations and positions the study within the existing subject of study. Section 2.1 defines robo-advisory, traces its emergence, and delineates the core workflow into profiling and asset-allocation subprocesses, introducing the building blocks and their interactions. Section 2.2 then maps the field's evolution across four generations, from questionnaire-only recommenders to fully automated, AI-enabled platforms, highlighting hybrid human-machine variants, oversight, and rebalancing logics. Section 2.3 reviews studies on investor profiling and portfolio recommendations, covering questionnaires, portfolio composition, and disclosed construction methods. Section 2.4 synthesises empirical work on RA performance across geographies and metrics, foregrounding methodological constraints and benchmarking issues.

Chapter III undertakes an empirical assessment of investor profiling and portfolio strategies across RAs operating in Germany. It combines market mapping, qualitative content analysis, text analytics, and a mathematical modelling framework with an implementation using a custom GPT-based RA. The chapter is organised to move from data construction to inference and interpretation, and it signposts where subsequent material extends, validates, or stress-tests the findings (see, for example, Subsections 3.2.3, 3.2.4, or 3.2.5).

Section 3.1 sets out the evidential base and methodological choices in five steps. Subsection 3.1.1 outlines the approach for analysing the RA market and the data collection method for the questionnaires, portfolios, and strategies. Subsection 3.1.2 specifies the inductive qualitative content analysis method applied to derive question categories, select representative questions, and compile descriptive statistics, and the extraction of portfolio data with subsequent composition analysis methods. Subsection 3.1.3 then introduces the term frequency–inverse document frequency (TF-IDF) representation and clustering analysis machinery employed to validate and triangulate the previous analysis of the questionnaires. Subsection 3.1.4 describes the approach for the development of a formal model of the RA process that integrates profiling uncertainty, state-dependent allocation, and event-driven rebalancing as a tractable stochastic control problem. Subsection 3.1.5 documents the construction of a GPT-4o-based prototype, the design of a 32-item questionnaire and investment universe, based on the previous subsections, and the generation of a 100-investor sample used for downstream analysis.

Section 3.2 presents the empirical findings in five corresponding parts. Subsection 3.2.1 reports the results of the questionnaires' analyses. Subsection 3.2.2 reports the analysis results of the portfolio compositions and strategies. Subsection 3.2.3 presents the results of the machine learning methods (TF-IDF, cluster analysis) applied to the sample of questions. Subsection 3.2.4 sets out the mathematical RA model's components (profiling map, allocation operator, and rebalancing rule) and clarifies how profiling uncertainty propagates to allocation decisions. Subsection 3.2.5 evaluates the GPT-based RA's portfolio recommendations, including allocation patterns across asset classes and the mapping from question responses to equity exposures.

Section 3.3 discusses and interprets the results in light of previous literature, with the findings and discussion sections divided into subsections.

Chapter IV examines the empirical performance of RAs operating in Germany and assesses whether their portfolio strategies deliver superior, risk-adjusted outcomes relative to transparent, investable benchmarks matched by equity exposure. Building on the market mapping in Section 3.2, the analysis uses monthly data and adopts both a full-sample view (January 2017–February 2022) and a crisis subsample centred on the COVID-19 shock (December 2019–February 2022) to probe robustness across regimes. It follows a transparent, replicable

structure that first specifies data and measurement, then applies complementary analytical lenses—performance metrics, mean–variance geometry, and CAPM-based correlation analysis—before synthesising findings and limitations.

Section 4.1 defines the evidential base and methods in five subsections. Subsection 4.1.1 sets out the sample selection, data collection, and time framing. Subsection 4.1.2 defines the benchmark architecture and the assignment of RA portfolios to benchmark groups. Subsection 4.1.3 defines the performance measures applied for comparison. Subsection 4.1.4 sets out a mean–variance approach in the (σ, μ) -plane. Subsection 4.1.5 details a correlation study linking CAPM diagnostics to tracking behaviours, together with locally weighted scatterplot smoothing (LOWESS).

Section 4.2 reports descriptive statistics and comparisons of all performance measures within each benchmark group for both periods. Section 4.3 presents the mean–variance evidence, identifies tangency portfolios, and compares distances to the CML across groups and providers. Section 4.4 reports the correlation findings, discussing the link between systematic exposure, tracking intensity, and risk-adjusted outperformance. The chapter concludes with a discussion in Subsection 4.5 that synthesises evidence across methods and periods.

Chapter V consolidates the contributions of the thesis. Section 5.1 summarises the key findings from Chapters III and IV, organised by research questions defined in Subsection 1.2. Section 5.2 translates the findings into actionable implications for the stakeholders, regulators and standard-setters (Subsection 5.2.1), RA clients and investors (Subsection 5.2.2), and RA providers (Subsection 5.2.3). Section 5.3 outlines methodological and data limitations relating to scope, measurement, and modelling. Section 5.4 proposes future research directions.

II – THEORETICAL FRAMEWORK AND LITERATURE REVIEW

II - THEORETICAL FRAMEWORK AND LITERATURE REVIEW

2.1. DEFINITION AND EMERGENCE OF ROBO-ADVISORY

In this section, robo-advisory (RA) is defined as the subject of research, the technology is described, and the background and origins are explained. The ongoing digital transformation of financial services, driven by new technological possibilities such as smartphones, the spread of the internet, algorithms, computing capacity and speed, and changes in consumer behaviour, has enabled innovative financial technology start-ups, known as “FinTechs” (fintechs), to develop business models that disrupt established companies and force established financial institutions to digitise their processes, products, and services (Baek, 2021; Dorfleitner & Hornuf, 2019; Eickhoff et al., 2017; Go et al., 2020; Gomber et al., 2017; Haddad & Hornuf, 2019; Lee & Shin, 2018; Puschmann, 2017; Schueffel, 2016). A particularly disruptive innovation in wealth management is the provision of automated investment advice and portfolio management by robo-advisors (RAs), which are the subject of this thesis (Fisch et al., 2019; Iannarone, 2018; Jung, Dorner, Glaser, et al., 2018; Jung, Dorner, Weinhardt, et al., 2018; Jung et al., 2019; Sironi, 2016). The first scientific articles on RAs appeared in the mid-2010s (Arwas & Soleil, 2016; Bradbury, 2014; Cocca & Teodoro, 2016; Darren, 2015; Kinder, 2015; Kitces, 2015; Matt, 2015; Paracampo, 2022; Sironi, 2016).

Robo-advisory is one major development within the automation and digitisation of wealth management, known as “WealthTech” (wealthtech), which also includes other trends such as social trading and crowdfund investing (Dziawgo, 2021; Jokanovic, 2024; Kwok, 2022; Shahbaz, 2018). Social trading platforms enable users to mirror the trades of experienced investors in real time (Apesteguia et al., 2020; Kern, 2017; Röder & Walter, 2019; Tahvildari, 2025b). In contrast to robo-advisory, social trading employs human asset managers, referred to as “social traders”, to manage assets. Social trading allows social traders to create a virtual portfolio, and based on this, issuers can issue a tradable instrument (e.g., a certificate) on the stock exchange so that retail investors can buy it, which participates 1:1 in the performance of the virtual portfolio managed by the social trader (Horn et al., 2024; Kern, 2017; Oehler et al., 2016; Oehler & Schneider, 2023).

They may serve as alternative instruments for markets where there are few or no tradable, liquid, exchange-traded products.

Crowdinvesting platforms are digital marketplaces that match many small funders (the “crowd”) with borrowers or issuers without traditional intermediaries (Belleflamme et al., 2015; Moritz & Block, 2016). Some platforms provide portfolio builders that automatically invest based on investor preferences, which intersects with robo-advisory services, but using crowdfunding projects instead of ETFs as instruments (Kleverlaan et al., 2021; Republic Europe, 2025; Sharma & Lertnuwat, 2016). However, social trading and crowd investing platforms are excluded from the scope of this study. There are, though, providers on the market such as Revolut, which, as so-called super apps, combine these different fintech services with non-financial services (Agarwal & Ann, 2024; Fortagne et al., 2023; Hasselwander, 2024; Polasik et al., 2022; Revolut Technologies Inc., 2025).

The name “robo-advisor” is made up of the two terms “robo” (for robotics) and “advisor” (for advisory). The former term was given to an automated process that does not require the intervention of a human, using mathematical algorithms to assist investment decisions (Moulliet et al., 2016, p. 2), and the latter term is used to describe financial investment advisory (Sironi, 2016). The two terms put together refer to a portfolio management solution that plans to invest client assets using an automated client advisory. Several different definitions for RAs can be found in the literature. As Torno et al. (2021, p. 3) note, RAs are usually referred to as “robo-advisors”, “robo-advice”, or “robo-advising”, but are also broadly termed “digital investment management systems” (Coşkun, 2022; Rühr et al., 2019), “automated financial advisors” (Abraham et al., 2019; Altrock et al., 2023; Reher & Sun, 2019; Severino & Thierry, 2022; Tao et al., 2021), or articles that focus on portfolio allocation and rebalancing refer to them as “robo-investors” (D’Hondt et al., 2020; Schouten, 2018). Fisch et al. (2019) define robo-advisors as “automated online services that use computer algorithms to provide financial advice and manage customers’ investment portfolios” (p. 13). RAs provide automated, algorithm-driven financial investment services, combining the advantages of low fees from passive portfolio management and the adaptability of active asset allocation methods (Abraham et al., 2019; Beketov et al., 2018; Phoon & Koh, 2018; Renaux et al., 2019, p. 5). Beketov et al. (2018) define RA as “an automated investment

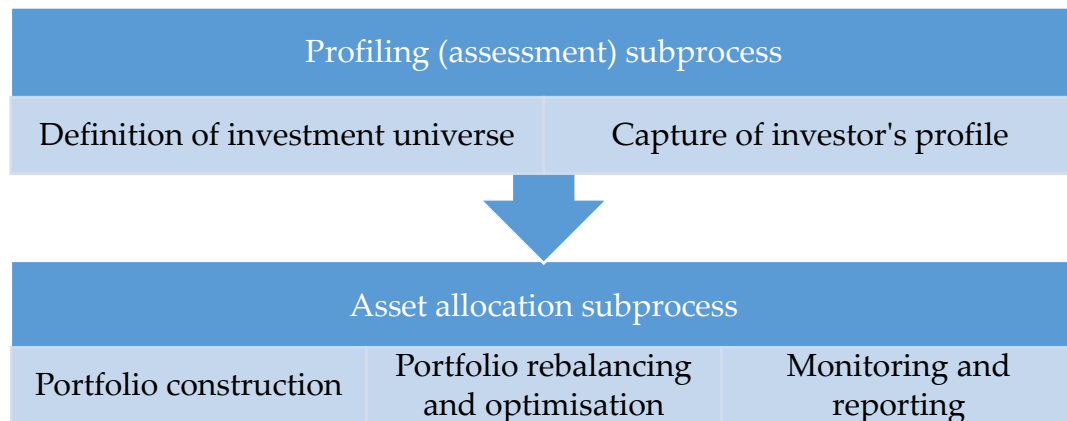
platform that uses quantitative algorithms to manage investors' portfolios and is accessible to clients online" (p. 364). Sironi (2016, p. 21) defines RAs as digital businesses with five main characteristics: they rely on automation, favour passive investing, offer portfolio management and tax optimisation, focus on customers and their individual goals and behaviours, and are single-minded companies. RAs aim to transform the traditional human-to-human advisory process into a digital human-to-computer process and are becoming increasingly popular for individuals to invest in and manage portfolios of assets (Jung, Dörner, Glaser, et al., 2018). Although there are slightly different terms and definitions for RAs, they all describe the same technological development. In this thesis, the abbreviation "RA" is used interchangeably for "robo-advisor" and "robo-advisory" depending on the context.

The typical workflow process of an RA has five building blocks, which are divided into two subprocesses. The first two blocks are part of the profiling (assessment) subprocess defining the investment universe and identifying the investor's profile (Beketov et al., 2018; Renaux et al., 2019, pp. 13–16; Tahvildari, 2025c). The next three blocks are portfolio construction, portfolio optimisation and rebalancing, and lastly monitoring and reporting. They belong to the asset allocation subprocess. These two subprocesses complement each other in this order, whereby the first subprocess in this thesis is referred to as the investor profiling process. The aim is to use the information about the investor, which is usually provided to the RA in a questionnaire, to create as accurate a picture as possible of the investor's profile, including preferences such as risk, amount, and maturity of the investment, such that this data can be used as a basis for portfolio recommendations and composition in the next subprocess.

The initial onboarding and assessment phase constitutes a relatively brief component within the broader RA procedure. Conversely, the subsequent subprocess of asset allocation and management typically encompasses a considerably more prolonged duration, particularly in instances where the RA not only proposes a portfolio but also undertakes its management, rebalancing, and continuous reporting in an ongoing manner. It is crucial to acknowledge that not all RAs are homogeneous in their functionality and scope and that the extent to which they cover the fundamental components may vary. First-generation RAs concentrated exclusively on the profiling phase, offering portfolio

recommendations, as will be examined in Section 2.2, which outlines the evolutionary stages of RA. Figure 1 illustrates the described RA process, including the building blocks.

Figure 1. Typical Workflow Process of Robo-Advisors



Note. Author's elaboration: This diagram shows the two main subprocesses of investor profiling and asset allocation in the entire robo-advisory process with the various building blocks.

Although some authors propose slightly different divisions of the RA process into blocks or subprocesses, these variations can ultimately be subsumed under the two fundamental components of profiling and asset allocation (Cocca, 2016; Jung, Dorner, Glaser, et al., 2018; Jung, Dorner, Weinhardt, et al., 2018; Jung et al., 2019; Rühr et al., 2019).

The subprocesses described for creating the investor profile and the portfolio are not the termination of the entire RA process but merely the inception, as this is usually followed by a longer-term procedure for the ongoing management of the portfolio, including rebalancing. The assignment of these building blocks to the asset allocation subprocess in Figure 1 derives from the existing literature, as, strictly speaking, they relate to the portfolio structure. Subsection 3.2.4 expands these concepts further by developing a mathematical model of the dynamic RA process. This model integrates the investor's behavioural patterns and alterations to their profile, such as adjustments to their risk appetite, assets, or other parameters, along with market dynamics. In light of this, a re-evaluation of the subprocesses is undertaken, culminating in an updated portfolio that aligns with

the revised investor profile. These are sophisticated features that are particularly relevant for fourth-generation RAs, which are outlined in the following Section 2.2.

In 1996 the Nobel laureate William F. Sharpe co-founded Financial Engines in Palo Alto to translate contemporary asset-pricing theory into an automated advisory tool for members of defined-contribution pension schemes. The firm's modelling framework simulated thousands of economic paths and applied mean-variance optimisation to recommend portfolio mixes that matched each investor's risk tolerance, horizon, and plan menu (Caramanis, 2025; Jones & Sharpe, 2010; Schulz, 2007; Spedding, 2002). The "Online Advice" platform, introduced in 1998, delivered these allocations—and subsequent rebalancing instructions—directly to 401(k) participants via the web with minimal human intervention. Trade press and later academic studies therefore identify Financial Engines as the first operational "robo-advisor" in personal finance (Arnott et al., 2023; StreetAuthority, 2016). Its debut coincided with two structural shifts: the migration of retirement risk from employers to employees following the rise of defined-contribution plans and the rapid diffusion of internet connectivity, both of which generated demand for scalable, low-cost advice (Fisch et al., 2019). Subsequent evaluations demonstrate that Financial Engines established the technical feasibility of web-based portfolio optimisation and revealed a willingness among retail investors to engage with algorithmic guidance, thereby providing a conceptual and technological template later adapted by direct-to-consumer platforms (Andrus, 2015; Arnott et al., 2023; Fisch et al., 2019).

The first proper RAs, as understood today, emerged in the US in the aftermath of the 2008–2009 financial crisis, marked by the market entries of Betterment in New York (2010) and Wealthfront in Palo Alto (2011) (D'Acunto & Rossi, 2023; Ferri, 2014; Gupta & Tham, 2019, p. 329; Renaux et al., 2019, p. 4; D. Wang & Nguyen, 2015). The RA landscape expanded rapidly in the 2010s, thanks in part to these pioneers' early success. This expansion was marked by the launch of a wide range of products and services by both new entrepreneurs and traditional financial institutions, with the latter becoming increasingly diverse in their offerings. A Google search for the term "robo-advisor" on 14 July 2025 yielded approximately 38 million results (Google, 2025). The AUM by RAs worldwide is estimated at USD 2.06 trillion for 2025, of which USD 1.67 trillion is in the United States (Statista, 2025). According to forecasts, AUM will exceed the USD 2 trillion

mark in 2025 and grow by an average of 3.66% per year to around USD 2.38 trillion by 2029, serving well over 34 million customers (Statista, 2025). This rapid proliferation has given rise to low-cost, ETF-centric platforms that are expanding access to investments, streamlining decision-making, and forcing established advisors to rethink their business models (Tiberius et al., 2022). Trends such as AI and big data analytics can accelerate the automation of RA services and help capitalise on returns (D'Acunto & Rossi, 2023; Tahvildari, 2025d).

2.2. EVOLUTION AND GENERATIONS OF ROBO-ADVISORY

Having defined and described what RA is, a closer look will now be taken at the evolutionary stages. Moulliet et al. (2016) identify four stages in the evolution of RAs, differentiated by the level of service they provide. The first generation of RAs supports customer assessment (customer profiling), which is conducted via a survey aimed at evaluating the investor's risk and return preferences and utility. The profiling process varies between different RAs. Following profiling, the RA typically proposes an investment product or portfolio consisting of various instruments such as ETFs, bonds, stocks, or commodities (Tertilt & Scholz, 2018). The RA's service ends at this point, as there is no bank or broker interface to manage the execution, meaning clients must buy and manage their portfolio independently using their own accounts. Some RAs combine with traditional human advisory services to become hybrid (Fisch et al., 2019).

The second generation of RAs goes further, offering appropriate accounts with dedicated fund management, as well as ongoing advice, managed adjustments, and rebalancing (Moulliet et al., 2016). This service includes setting up an investment account and executing orders. Questionnaires are used to allocate clients to predefined portfolios managed by real asset managers who oversee the algorithms and rule sets. Moreover, portfolio monitoring is a process that provides information regarding key portfolio figures. Such information in turn enables investors to assess the performance of their portfolios.

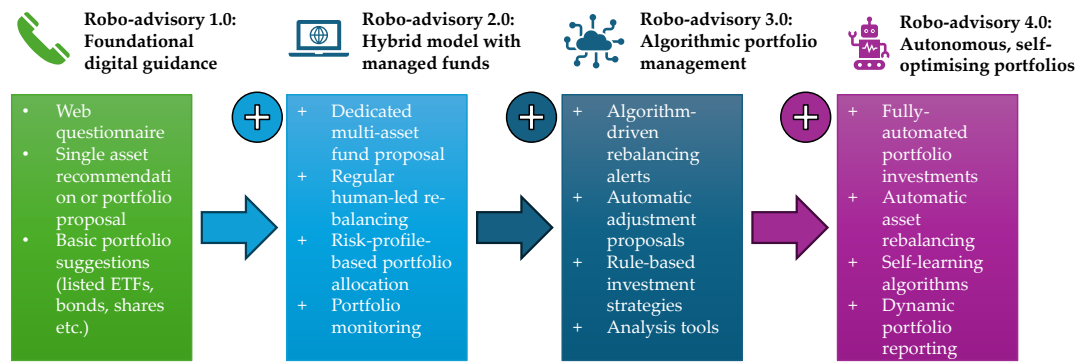
The third generation represents an innovative step forward with the addition of algorithmic-driven rebalancing alerts and automatic adjustment proposals based on dynamic market developments. There are rule-based investment strategies and additional analysis tools that provide additional dynamic performance results and

insights based on portfolio data. Some RAs offer their clients more customisation options for portfolio adjustments and analytical tools. Notwithstanding, it is crucial to emphasise that the rebalancing and investment algorithm is not yet fully automated; manual reallocations by the customer or a human investment advisor remain necessary.

The fourth generation provides fully automated portfolio investment with automatic rebalancing and dynamic reporting, thereby negating the need for customers to intervene in the portfolio after initialisation. This can even extend to the utilisation of self-learning algorithms or AI techniques based on market factors and other input factors, such as behavioural and profile changes and changing customer needs.

Many RAs currently available on the market are classified as third or fourth generation, with a growing number of fourth-generation models, particularly those employing AI in robo-advisory (Abbas, 2024; Arenas-Parra et al., 2024; Beketov et al., 2018; Dias, 2025; Roy et al., 2024; Tahvildari, 2025c). This leads to the categorisation of these generations, as depicted in Figure 2.

Figure 2. Robo-Advisory Evolution Stages



Note. Author's elaboration based on Moulliet et al. (2016, p. 4).

2.3. INVESTOR PROFILING AND PORTFOLIO RECOMMENDATIONS

This section provides a literature review of studies concerning investor profiling and portfolio strategies with a focus on the questionnaires, portfolio compositions and strategies of RAs available on the market, regarding the research questions RQ1, RQ2, RQ3, and RQ4, defined in Section 1.2. Sources are selected based on a literature search using scientific meta-search engines, such as Google

Scholar and Crossref, and scientific platforms, such as ScienceDirect. Further relevant sources are identified using forward and backward citation searching based on those already found. This general literature search procedure has already been applied to the literature search for the entire dissertation, particularly for Sections 2.1 and 2.2. However, this section will refer specifically to research questions RQ1–RQ4, discussing relevant sources on investor profiling and portfolios of RAs. As mentioned in Section 2.1, the first scientific literature on RAs emerged in the mid-2010s, so the period following the 2009 financial crisis serves as a reference point for the literature search. In addition to keywords, titles and abstracts, studies relevant to the specific investor profiling process and portfolio recommendations of RAs will be examined in detail in terms of their questions and objectives, data and methods, and results and conclusions. Finally, the results will be presented and discussed below.

An early study by Scherer (2017) investigated which household characteristics drove participation in, and allocation to, risky financial assets, with direct implications for RAs. It developed a simple normative framework with decreasing relative risk aversion and fixed participation costs, yielding a wealth-dependent threshold for market entry and optimal risky weights. Using German HFCS panel data (3,565 households), the study estimated interpretable classification trees with repeated five-fold cross-validation. Empirically, net wealth and risk aversion emerged as the dominant predictors of financial market participation, followed by income; education and age further raised participation probabilities. Real estate holdings crowded out financial investment, while mortgage leverage and background risks dampened risky allocations; results were robust when equity and bond participation were analysed separately. Although the predictive lift over a naïve baseline was moderate, it was statistically meaningful. The principal implication is that RAs should adopt a holistic, balance-sheet-aware profiling approach that captures both ability (assets, liabilities) and willingness (risk attitudes), consistent with MiFID II suitability expectations.

Faloon and Scherer (2017) investigated the extent to which modern RAs adapt portfolio recommendations to the specific circumstances of individual investors. They began by presenting robo-advice as a response to two structural pressures: fee compression in a low-interest rate environment and the global shift

from defined-benefit to defined-contribution pension schemes. These changes leave households with modest wealth in need of low-cost, scalable guidance. Drawing on Sharpe's four principles of advice—diversification, economisation, contextualisation, and personalisation—the authors argued that RA platforms mainly excel at providing cheap, diversified beta (β) but neglect meaningful personalisation.

Using the generic questionnaire adopted by most providers (enquiring about factors such as age, income, savings, and risk tolerance), the authors outlined two translation mechanisms. The first was a legally driven scoring model that mapped ad hoc point totals to preset efficient frontier portfolios. The second, more rigorous method embedded the same inputs in a mean–variance household optimisation model that treats human capital as a shadow asset and retirement liabilities as future obligations. This yielded a four-fund separation of speculative, diversification, liability-hedging, and cash positions. In practice, however, most RAs defaulted to the simpler scoring approach, producing “canned” portfolios that ignore heterogeneity in career risk, property leverage, pension entitlements, or gender-specific longevity. The authors then scrutinised risk-profiling practices. They contended that standard multiple-choice questions lack psychometric validity and convert coarse answers into numeric risk-aversion parameters with questionable reliability. The authors pointed out that evidence from experience-sampling experiments indicates that interactive simulations of wealth trajectories tend to produce more stable and typically higher risk appetites. Yet this methodology was rarely implemented. Two partial attempts at customisation are examined: tax-optimised rebalancing and the concept of “mental accounts” based on goals. While tax overlays add genuine value, goalification may induce aggregate inefficiency by preventing the netting of opposing hedging and speculative demands across goals.

To truly individualise, the authors proposed extending the household balance sheet model to include diversified shadow assets, illiquid outside wealth, and multiple liability streams, each with its own covariance structure. Such granularity would generate investor-specific glide paths and distinguish fiduciary RAs from merely compliant ones. Until this evolution occurs, the authors concluded that most platforms would remain far simpler than genuinely intelligent, client-centric robots. Nonetheless, it is uncertain how many and which

RAs the authors assessed. Furthermore, the questions were vague and did not encompass all aspects and dimensions of true RAs, perhaps due to the study's dependence on a prior version of RAs.

Tertilt and Scholz (2018) studied how RAs elicited and translated the risk preferences of retail investors and whether algorithmic routines could rival human practice. They compiled a sample of thirteen proprietary, anonymised platforms—seven German, three US, and three UK services—which were selected because their web questionnaires could be accessed without credentials. A custom-made Selenium/Excel script generated up to 10,000 answer permutations per site (more than one million in total) and captured the resulting risk ratings, back-solving each platform's implied equity weight in the process. Five German bank questionnaires and the 25-item FinaMetrica psychometric survey served as conventional benchmarks, while the portfolio outputs were compared with a German retail bank dataset and the *Finanztest* consumer magazine. While bank advisors posed 19-38 questions, RAs presented only 4–27 questions, largely to reduce user drop-off. However, roughly 40% of these items had no bearing on the final risk score. As almost every platform applied fewer than 25 targeted risk-tolerance items, the reliability threshold (Cronbach's $\alpha \geq 0.8$) was not met. Many algorithms relied on only one to three answers, commonly age or a self-selected risk level, and several platforms allowed clients to override the computed profile entirely, thereby undermining the advisory function.

Equity-quota analysis revealed significant differences between countries. For the lowest risk band, German RAs recommended an average equity share of 19%, UK peers recommended 7%, and US peers recommended 34%. At the highest risk level, German services capped equities at around 74%, whereas both Anglo-Saxon cohorts approached 86%. German portfolios therefore appeared more conservative than the *Finanztest* templates at low and medium risk, yet slightly less conservative than the retail-bank benchmark at the aggressive end. Portfolio granularity varied from three to twenty-three preset allocations, but this range did not align with the narrowness of many risk-profiling routines. The authors concluded that, despite their advantages in terms of scale and fees, RAs in 2018 delivered only basic, sometimes regulation-driven, risk assessments. Without richer psychometrics, goal-based modelling, and adaptive AI, the technology could not meet fiduciary

standards. They suggested that future progress depended on engaging data collection methods, behavioural metrics, and bespoke, goal-orientated asset allocations. Tertilt and Scholz (2018) concluded, as Faloony and Scherer (2017) had already done, that until such advances are realised, robo-advice would offer convenience but only limited advisory value.

Beketov et al. (2018) catalogued the quantitative methods embedded in RAs and assessed their prevalence and associated assets under management (AuM). From a universe of 219 global providers (third/fourth-generation systems), they extracted method disclosures for 73 RAs and examined a representative sub-sample of 28 to map the typical workflow: ETF universe selection, investor profiling via online questionnaires, portfolio construction, monitoring/rebalancing, and reporting. MPT dominated as the headline framework ($\approx 40\%$), well ahead of “sample portfolios” and “constant weights”. A word-cloud analysis showed frequent references to MPT-adjacent tools, while a method–AUM comparison revealed that more sophisticated techniques, such as Black–Litterman and full-scale optimisation, attracted disproportionately higher AUM despite lower incidence.

Most RAs implemented what the authors termed a multidimensional improvement of MPT: alternative return models (Black–Litterman, Gordon growth, CAPM), risk measures (VaR), constraints, tax/trading frictions, correlation choices, simulations, and pragmatic rebalancing (typically threshold/event-based with daily checks; some calendar-based, with dividend reinvestment). The study also highlighted a gap between widely used methods and state-of-the-art approaches in the literature. Looking forward, the authors expected greater adoption of risk parity, full-scale scenario optimisation, and risk-parity variants with skewness—both for performance potential and marketing appeal. They further anticipated hybrid advisory models (automation plus human oversight), particularly for wealthier clients seeking manual overrides in atypical or stressful conditions.

Bourgeron et al. (2018) argued that RAs required fully automated, robust portfolio engines beyond plain mean–variance optimisation (MVO). They showed that MVO was essentially an “alpha optimiser” driven by hedging/arbitrage factors and was numerically unstable owing to small covariance eigenvalues; modest input changes could yield extreme, non-investable weights. The proposed

remedies were regularisation and sparsity. The authors formalised ridge/Tikhonov and lasso penalties, alone or as an elastic net, linking weight constraints to implicit covariance shrinkage and advocated spectral filtering/denoising to stabilise inversions. Because these penalties created non-standard programmes, they proposed the proximal and alternating direction method of multipliers (ADMM) solvers with closed-form projections for common constraints (budget, bounds, turnover, cardinality). For practice, they separated (i) expected-return formation (e.g., Black–Litterman from strategic benchmarks and view ratings), (ii) a tracking-error-based objective with time-varying tracking error targets, and (iii) calibration of penalty strength via (generalised) cross-validation. The result was a scalable, dynamic allocation framework that could rebalance hundreds of portfolios without human trial and error while keeping allocations stable, sparse, and transparent.

Jung et al. (2019) investigated the emergence of RA services within retail and private banking, evaluating their prospective influence on the German and wider wealth management sector. First, they retraced the historical evolution of financial advice. They observed that the 1950s model had been characterised by exclusivity and high fees, while discount broking had broadened participation in the 1970s. The internet then enabled self-directed online trading in the 1990s, and algorithmic trading paved the way for fully automated advisory platforms in the 2010s. Against this backdrop, the authors defined RAs as digital platforms that automate customer profiling and portfolio management through questionnaires, passive ETF baskets, mean–variance optimisation, and algorithmic rebalancing. This procedure lowers minimum investment requirements and management fees. The authors formulated the research question of how RAs reshape the advisory industry and which factors might support or hinder this process and answered it by conducting a structured strengths, weaknesses, opportunities, and threats (SWOT) analysis.

They noted the following strengths: reduced fees and entry thresholds; tax-loss harvesting facilities; an intuitive digital user experience; evidence-based portfolio construction; automated discipline that suppressed emotional trading; and 24/7 accessibility. They identified weaknesses as comprising hidden or variable cost layers, potential conflicts of interest with affiliated brokers or proprietary products, basic risk profiling via simplistic questionnaires, limited personalisation, the absence of human interaction, and possible breaches of fiduciary duty.

Regarding opportunities, the authors highlighted the ubiquity of digital services among Millennials, the potential to extend RA logic to complex, goal-based investing, the prospect of hybrid business-to-business-to-consumer (B2B2C) models in which traditional institutions integrate RA engines, and the possibility of standardising ancillary services, such as tax and legal advice. Threats stemmed from intense competition among stand-alone fintechs and incumbent banks launching proprietary RAs, users' distrust of impersonal algorithms, tightening regulatory scrutiny concerning suitability and disclosure, and uncertainty regarding RA performance in severe bear markets, a context that none had yet experienced. The study concluded that RAs had significant disruptive potential in democratising professional portfolio management and pressuring fee structures. However, their long-term success would depend on overcoming personalisation deficits, aligning incentives, gaining consumer trust, and navigating future regulations. The authors contended that RAs should not be viewed as a wholesale replacement for human advice but rather as complementary technologies whose future impact would depend on their integration with established advisory frameworks.

Fisch et al. (2019) explored the development and disruptive potential of RAs within the financial advisory industry, particularly in the context of retirement systems. The authors described the emergence of RAs as a response to several market inefficiencies, namely limited access to affordable advice, widespread financial illiteracy, and increasing investment complexity. It was demonstrated by early firms such as Wealthfront and Betterment that the feasibility of low-fee, scalable advisory models targeting underserved clients who previously lacked sufficient assets to engage human advisors was possible. Risk-based questionnaires, automated portfolio rebalancing, and features like tax-loss harvesting have demonstrated the ability of RAs to facilitate investment.

Despite these advantages, RAs had not yet seen widespread adoption. As of the late 2010s, the utilisation of these tools was minimal, with only 5% of US investors employing them, and awareness levels remained low (Jones, 2020). The typical RA client was found to be younger, more technologically literate, and to hold fewer investable assets. The barriers identified included a degree of scepticism concerning the quality of advice, a perception that it lacked personalisation, and an enduring preference for human interaction. The range of services provided by RAs

was found to be diverse, with many providing only fundamental portfolio management and limited capability to address intricate financial circumstances, such as estate or retirement planning. However, RAs had some clear benefits: much lower fees (usually 0.25–0.50% compared to 1–2% for traditional advisors), easy access with little or no minimum account requirements, and a lower chance of conflicts of interest caused by selling products. In contrast to the practices of many human advisors, RAs were less incentivised to promote high-fee investment products. However, when integrating RAs within larger financial institutions, the possibility of conflict persists. The study noted several industry trends, including the emergence of hybrid models combining RA automation with human support; increasing product diversification to target niche demographics (e.g., women, retirees, and ethical investors); and vertical integration, with large asset managers acquiring RA firms to distribute their own products. The utilisation of RA tools by conventional financial advisors has also witnessed a surge in popularity, thereby facilitating enhanced efficiency in the domains of portfolio construction and risk profiling. Fisch et al. (2019) advanced the argument that RAs exhibited clear advantages in terms of cost and accessibility, but their long-term impact was contingent on the enhancement of personalisation, client trust, and hybrid integration. The emotional aspect of human advice remained a key differentiator, suggesting that RA-human collaboration may define the future of financial advisory services.

Kaya and von Martens' (2020) report from Deutsche Bank Research analysed the emerging role of RAs in the German financial advisory landscape and their potential to promote passive investment strategies, particularly ETFs. The report indicated that, historically, German households had demonstrated a preference for savings accounts and actively managed retail funds, as opposed to direct capital market investments or low-cost passive alternatives such as ETFs. Despite their growing popularity on a global scale, ETFs accounted for a relatively modest proportion (EUR 30–35 billion) of German retail fund holdings in 2019, in comparison to EUR 633 billion in mutual funds. RAs have been gradually introducing changes to this traditional investment culture. The report projected that if current trends persisted, German RAs could manage up to EUR 25–35 billion by 2025. Current statistics demonstrate that the German market was valued at

approximately EUR 19 billion in 2024, thereby substantiating the forecast's precision at the lower end of the conservative range (Statista, 2025).

The analysis identified a prototype RA client as being middle-aged, male, and with an income that was three times higher than that of the average bank client. These individuals exhibited a propensity for autonomy and digital interaction, yet concurrently sustained engagement with conventional banking branches. While women and low-income groups were under-represented, the study anticipated a gradual expansion of RA participation due to demographic shifts, inheritance-driven wealth transfers, and increased digital literacy. The report attributed the slow uptake of ETFs in Germany to structural barriers, including the dominance of the universal banking model (where advisors prefer in-house actively managed funds for commission-based incentives) and the prevalence of pay-as-you-go pension schemes, which reduced household motivation to accumulate capital market wealth. Nevertheless, the report concluded that RAs could act as a catalyst for increasing ETF adoption, thereby bridging the "advice gap" for underserved client segments. The appeal of these investments stemmed from their cost-efficiency, simplicity, and alignment with evolving investor preferences, particularly among digital natives. Despite their digital orientation, RA clients still valued in-person services for complex transactions, suggesting that hybrid advisory models could be the way forward. The report emphasised the transformative capacity of RAs within Germany's investment ecosystem, particularly in the direction of retail investors towards passive strategies, the transformation of financial advisory structures, and the promotion of long-term capital market engagement.

Boreiko and Massarotti (2020) examined why portfolio recommendations differed across RAs and whether these services mapped investor risk profiles into asset allocations. Using 53 platforms (25 US, 28 German) and three stylised investor types (conservative, moderate, aggressive), they collected 159 recommended portfolios and modelled the equity share as the dependent variable. Ordinary least squares with Newey–West corrections tested four drivers: investor risk type, number of model portfolios, provider "expertise" (relative weight of equity, fixed income, and other assets in the investable universe), and domicile. They found substantial cross-firm variation, especially for conservative and moderate types. Risk profiles strongly predicted equity weights, indicating that RAs generally

respected suitability inputs. The breadth of model portfolios was not significant. By contrast, firm expertise mattered: equity (fixed-income) specialisation was associated with higher (lower) recommended equity shares. US providers recommended materially higher equity exposures than German peers. The authors called for clearer standards to reduce unwarranted dispersion.

D’Acunto and Rossi (2021) reviewed RA as a response to the limitations of traditional financial advice: high fees, capacity constraints, and the transmission of advisors’ biases. They proposed a four-dimensional taxonomy (personalisation, investor involvement, client discretion, and human interaction) for RAs with investor approval and RAs with fully automated execution. Surveying theoretical and empirical evidence, they documented that RAs typically improved diversification, reduced idiosyncratic risk and behavioural biases, and raised risk-adjusted performance, especially for previously undiversified or inexperienced investors; the benefits for already well-diversified users were limited. Long-horizon services emphasised ETF-based asset allocation, periodic rebalancing, and, in some cases, tax-loss harvesting or VaR-based risk controls. Beyond investments, emerging tools nudged spending–saving choices by aggregating accounts and delivering behaviourally informed prompts. They concluded with open issues: designing holistic life-cycle advice; addressing algorithmic aversion via hybrid models; distributional impacts and access; potential systemic effects of portfolio homogeneity; and fiduciary, privacy, and data-governance questions.

A further study by Blaschke and Kriebel (2021) investigated which RA clients were actively seeking in-depth risk-profiling (“guidance customers”) and how such investors could be identified ex-ante. The researchers analysed a set of 10,447 onboarding records from a prominent German RA, focusing on a subset of 5,176 single-account holders who had been legitimated and collected between mid-2019 and mid-2020. During the onboarding process, each client was required to select either a brief, self-selected risk level or to complete a more extensive questionnaire that elicited risk tolerance through the addition of further items. The acceptance of the long path was identified as the dependent variable, serving as a direct proxy for valuing advice. As a mere 12% of users elected to complete the extended questionnaire, the authors were compelled to balance the training data by employing the random oversampling examples algorithm and training a random-

forest classifier; a logistic-regression model was utilised as a benchmark. The ten-fold cross-validation procedure selected a number of features randomly selected for consideration at each split of five, and a 70/30 train-test split evaluated the out-of-sample performance. The random forest achieved an area under the receiver operating characteristic curve of 0.66, outperforming the logistic regression's 0.63. This demonstrates meaningful, albeit not perfect, predictability.

The permutation importance and partial-dependence plots indicated that self-reported financial knowledge was a significant predictor of prediction accuracy. A strong correlation was observed between higher literacy levels and a greater propensity to decline advice, while limited knowledge was found to be associated with an increased likelihood of opting for guidance. Several other significant variables were identified, including age, net income, investable assets, and the amount initially invested. Clients who were younger, less affluent, or planning smaller contributions displayed a greater need for risk elicitation. Contrary to the findings of some earlier studies, older investors demonstrated a lower demand for advice, suggesting that confidence (or questionnaire fatigue) may have increased with age. The findings indicated that females exhibited a modestly higher propensity to accept advice than males, thereby corroborating extant evidence on gendered financial overconfidence. The authors concluded that RAs could enhance user experience and conversion by dynamically tailoring onboarding flows. In this regard, customers predicted to require guidance should receive richer questionnaires and explanatory content, whereas self-confident segments could be routed to abbreviated forms. The implementation of such segmentation techniques was poised to enhance the perceived value of advisory services for users, while concurrently ensuring that the process did not impose an undue burden on other individuals. This approach underscored the efficacy of machine-learning algorithms in personalising digital wealth management operations on a large scale.

So (2021) explored how banks and investment houses assessed clients' risk preferences and derived a concise, theoretically grounded framework that RAs could apply in the Hong Kong market and beyond. The study collated 20 publicly available risk-assessment questionnaires (11 from Hong Kong and nine from Australia, Canada, the UK, and the USA) that collectively encompassed 180 individual items. Using inductive content analysis, So (2021) first classified almost

every item (approximately 97%) into 15 question types according to their manifest wording and response scales. A comparison of the type frequencies revealed regional differences in design: Hong Kong providers demonstrated a greater emphasis on financial capacity questions (percentage of income/net worth invested, employment, and liquidity checks), whereas overseas providers placed a stronger emphasis on behavioural reactions to losses and utilised data-visualised answer options with greater frequency. To make the instrument actionable, the 15 types were then collapsed into five higher-order risk factors that were mutually exclusive yet collectively exhaustive:

- i. Realistic investment objectives (goal/return expectations & time horizon, factual consistency checks).
- ii. Risk appetite (subjective willingness via literal & quantitative descriptions).
- iii. Investment knowledge & experience (practical and product-specific).
- iv. Behaviour under loss (intended actions during drawdowns, behavioural validation).
- v. Ability to bear risk (income-, wealth-, and employment-based capacity).

Reliability was ensured through independent recording by a research assistant and subsequent expert review, while validity was cross-checked against proprietary frameworks disclosed by Deloitte and Wealthfront. So (2021) advanced the argument that RAs should develop streamlined questionnaires that target only these five factors to avoid respondent fatigue and inconsistent answers. The additional implementation guidelines included the following:

- i. the maintenance of a low item count,
- ii. the formulation of questions in a straightforward manner,
- iii. the flagging of contradictory responses in real time,
- iv. the scheduling of periodic score updates to capture evolving sentiment, and
- v. the prevention of clients from “gaming” the system by requesting frequent, ad hoc risk-grade changes.

The framework promised more accurate, regulator-compliant profiling, stronger customer trust, and ultimately more sustainable, contact-free investment

services. The study's implementation of social-distancing measures reinforced this imperative.

Bräuer (2021) examined how an RA influences the mutual fund/ETF savings plan choices of households at a major German online bank. Using administrative data from 2003 to 2017 covering over 100,000 clients and employing a natural field experiment setting, he compared first-time and experienced savings plan users before and after adoption by combining matching with difference-in-differences analysis. Defaults were found to be powerful, with around 90% of chosen funds being tool defaults, which lowered total expense ratios and diversification losses. However, a prominently displayed 50/50 equity-bond default was found to depress equity exposure. The effects persisted but were heterogeneous: inexperienced users remained inert, while experienced users subsequently increased their equity holdings, retaining lower fees and improved diversification. Simulations showed considerable fee savings. However, for inexperienced users, the lower equity share could reduce long-term wealth despite cost savings. The study highlights the benefits and suitability risks of digital defaults.

The study by Tsai and Chen (2022) explored the diffusion of AI-driven big data RAs in stock market forecasting, using Rogers' innovation diffusion theory as its conceptual framework (Miller, 2015; Wonglimpiyarat & Yuberk, 2005). Despite the considerable academic effort expended on the development of AI-based prediction models, including time series, decision trees, and neural networks, this study addresses a significant gap in the literature, namely the lack of empirical understanding regarding their real-world adoption by institutional investors.

Employing an abductive reasoning approach, the authors triangulated insights from extant literature with data obtained via semi-structured interviews conducted with representatives of 21 fund companies. A review of the academic literature revealed that AI neural network models demonstrated prediction accuracies ranging from 52% to 97%. Nevertheless, these accuracies varied substantially across models and contexts, with few models consistently outperforming traditional methods. The unpredictability and lack of stability in the performance of these models emerged as a key concern. The interviews corroborated these findings. While 95% of respondents expressed awareness of big data RA models, none (with the exception of one individual) reported adopting them for active fund management. The primary reasons for non-adoption were the

models' perceived complexity, insufficient relative advantage, lack of observability, and unreliable forecast accuracy. In accordance with the principles of innovation diffusion theory, the majority of respondents remained in the "knowledge" stage and had not progressed to the subsequent stages of "persuasion" or "adoption". This finding suggests a failure in the diffusion process.

The study concluded that AI-based RAs, particularly those leveraging neural networks, had not yet achieved widespread practical uptake despite significant academic interest. This finding aligns with Fama's (1965) efficient market hypothesis, which posits that market prices already incorporate all available information, thereby rendering predictive models ineffective in delivering excess returns. Consequently, despite the theoretical novelty offered by neural network models, they have yet to translate into disruptive tools within the investment industry. From a practical perspective, the research highlighted a significant disconnect between academic innovation and institutional application. Theoretical models remained confined to research settings due to implementation barriers, including trade secrecy, cognitive load on decision-makers, and lack of observable success. Nevertheless, the study suggested that increasing familiarity and potential improvements in transparency and accuracy may pave the way for these models to become supplementary tools in technical analysis in the future. The authors called for further research incorporating broader influencing variables and emphasised the necessity for greater testability and demonstration effects to enhance diffusion potential.

D'Acunto & Rossi (2023) conducted a review of the emerging yet rapidly expanding body of literature on RA. They argued that this digital form of financial intermediation has the potential to transform households that are biased and ill-informed into decision-makers whose behaviour aligns with the assumptions of standard neoclassical models. The authors highlighted that RAs are different from regular human advisors because they use large amounts of data, provide clear expectations, and automate processes, which helps reduce financial, mental, and emotional costs.

They showed that RA mitigates information frictions. Income-aggregator apps compiled users' transaction-level data and generated balance-sheet dashboards. Exposure to these dashboards prompted Icelandic households to

reduce overdraft debt and curb frivolous spending, while Canadian users who received overspending alerts consistently reduced their spending. Furthermore, when US users were shown anonymised peer benchmarks, their spending converged towards the peer mean, with those who overspent reducing their consumption the most. However, the review also warned that the “ostrich effect”, whereby users avert their gaze from negative financial feedback, might limit the long-term impact of RA and exacerbate inequality. Moreover, RA reduces distorted beliefs. Goal-setting interfaces in an Italian app encouraged regular monitoring and increased monthly savings, particularly among impatient or low-income users. Experiments on debt repayment tools revealed that households were willing to pay more for RA than they gained financially, indicating an awareness of their limitations. In peer-to-peer lending, algorithmic recommendations corrected cultural and in-group biases, redirecting funds towards previously marginalised borrower groups. Furthermore, personalised forecasts of equity returns improved portfolio Sharpe ratios and eliminated trading biases, such as trend-chasing and the disposition effect. Lastly, the technology lowered transaction costs. Evidence from the US indicated that reducing minimum account sizes increased stock market participation among middle-class households, while French employees who received automatic rebalancing targets achieved higher risk-adjusted returns due to reduced inertia. Mobile wallets in India and Singapore increased purchasing power and stimulated sales growth for small firms, although the increased visibility of luxury goods occasionally led to excessive spending.

Despite these advantages, adoption remained uneven. Algorithmic aversion, driven by distrust, discomfort, or uncertainty, was strongest among older and less technically adept users. However, some segments displayed algorithmic appreciation. Provider incentives also mattered: banks might temper RA that prevented profitable consumer mistakes, whereas third-party platforms aligned fees with user outcomes. The review concluded that RA was more than just a FinTech novelty; it was also a potential policy instrument and a laboratory for testing economic theory. However, it also cautioned that long-term learning, dependence on algorithms, distributional effects, and regulatory design remained areas of ongoing research.

Scherer and Lehner (2022) conducted a study to ascertain whether German RAs genuinely provide personalised, economically sound portfolios, as frequently

promoted in their advertising campaigns. The authors of the study meticulously collected data on 243,000 portfolio recommendations, encompassing 16 RAs. These recommendations collectively constituted approximately 78% of assets under management in the German market during the period from 1 to 23 June 2021. For each recommendation, the full set of questionnaire answers provided by a prospect was captured, thereby producing a data-intensive cross-section that permitted direct linkage between client inputs and suggested equity weights. The majority of providers offered only a small number of pre-baked model portfolios, with 10 or fewer being offered in 12 out of 16 cases. This meant that advisors were able to map clients onto a very limited efficient set, rather than constructing portfolios “from first principles”. Extreme financial strategies, such as 100% bonds or equities, were seldom, if ever, available, presumably with a view to mitigating litigation risk and maintaining regulatory compliance.

Utilising both ordered-factor ordinary least squares (OLS) regressions and model-agnostic permutation importance, the study demonstrated that risk-aversion items accounted for the largest proportion of variance in recommended equity shares. The time horizon was the second most significant factor, while investor experience was only relevant for a small number of firms. The impact of wealth and income on allocations was observed in a limited number of cases, exhibiting only a marginal effect. This finding stands in contrast to models predicated on decreasing relative risk aversion. Objectives such as wealth preservation or retirement planning influenced outcomes at some providers, but the authors argued that these goal-based questions reflected behavioural nudges and regulatory check-boxes more than normative portfolio theory.

The overall explanatory power of three variables (typically risk aversion, horizon, and goal) averaged 82% in R^2 terms; additional questions added negligible value, implying that many questionnaire items served disclosure rather than optimisation purposes. Regression-tree analysis corroborated these findings and revealed non-linear interactions consistent with “trust-building” design choices: cautious investors with short investment horizons received low-equity portfolios, while long-horizon, goal-orientated investors who accepted volatility obtained higher equity weights. Consequently, Scherer and Lehner (2022) concluded that German robo-advice delivered standardised, regulator-friendly solutions that

aligned with clients' prior beliefs instead of Merton-type, balance-sheet-specific prescriptions. This strategy is hypothesised to have enhanced user trust and simplified compliance; however, it is theorised that it limited the economic value-add of RAs and undermined the original promise of low-cost, genuinely customised advice.

A study by Oehler and Horn (2024) examined whether the large-language-model (LLM) chatbot ChatGPT delivered more suitable one-off investment recommendations than commercial RAs. Three stylised retail-investor profiles were constructed, which differed primarily in terms of risk aversion (high, moderate, low) while sharing realistic characteristics (e.g., EUR 50,000 in investable funds, a long horizon, adequate liquidity reserves). In April 2023, a query was submitted to 17 internationally available RAs, with each profile analysed and the recommended portfolio allocations recorded. In addition, the subjects were presented with bespoke prompts that featured the same profiles. These prompts prompted the subjects to solicit advice from ChatGPT (GPT-3.5, free version). The authors derived a benchmark allocation for each profile by utilising established academic guidelines and 60-40 heuristics adjusted for risk tolerance to assess suitability. The subsequent step involved a comparison of each provider's stock-bond-(alternatives) split with the benchmark, with the results of this comparison being classified as either matches, partial matches, or failures. ChatGPT produced coherent, succinct recommendations:

- Risk-averse Profile A: 60% high-quality bond fund, 30% global equity index fund, 10% REIT/commodity mix.
- Balanced Profile B: 50% equity index, 30% bond fund, 20% broad REIT/commodity exposure.
- Risk-seeking Profile C: 70% equity index, 15% bond fund, 15% commodities/REITs.

These allocations exhibited a high degree of congruence with the literature-based benchmarks, and ChatGPT furnished straightforward justifications (diversification, inflation hedging, and rebalancing frequency). Subsequent follow-up prompts demonstrated that the recommendations provided remained consistent over time, despite minor fluctuations in percentage, which did not compromise their appropriateness. Among the 17 RAs, only three (approximately

18%) approximated the benchmark across all profiles; eight matched at most one profile; and three failed for all profiles. The latter failed typically by over- or underweighting equities relative to their stated risk attitudes. Robo-advisors required the completion of extensive questionnaires, whereas ChatGPT posed a limited number of clarifying questions and provided counsel in a substantially reduced timeframe. Oehler and Horn (2024) concluded that, as of early 2023, ChatGPT provided superior first-pass portfolio guidance for lump-sum investors compared with mainstream RAs. In terms of policy, the authors proposed the encouragement of retail investors to consult independent generative-AI chatbots as a low-cost “second opinion”, while traditional advisors could integrate such tools to streamline processes and upgrade recommendation quality. More broadly, generative AI appeared poised to become the next significant advance in accessible, high-quality financial advice, benefiting both providers (lower costs) and consumers (better fit, clearer explanations).

Kumar (2025) conducted a study on the evolving dynamics of the financial advisory market in the wake of the introduction of RAs. The objective of the study was to comprehend how the advent of financial technology, particularly RA services, has reconfigured the market’s structure and exerted an influence on the strategic conduct of firms offering such services or not. Applying a novel, hand-collected dataset, the author differentiated between incumbent financial advisory firms and new entrants offering RA technology. The research employed a dual approach, incorporating theoretical and empirical methodologies. Firstly, a basic extensive game model was constructed to propose a mechanism explaining the observed patterns in the market. The theoretical framework thus illustrated strategic interactions among firms and allowed the researcher to identify potential equilibria in response to technological adoption. The model’s implications were then subjected to empirical analysis, based on data collected from the evolving market.

The key finding of the study was that incumbent firms responded to the entry of RA competitors by adopting the same technology. This technological adoption by incumbents subsequently prompted a strategic reaction from the original RA firms, who responded by expanding their workforce of human financial advisors. Contrary to the widely held belief that automation would result in a reduction in

demand for human advisors, the study demonstrated that the presence of RAs had induced an overall expansion in demand for advisory services. Kumar (2025) concluded that the introduction of RAs did not merely result in the displacement of traditional services but rather precipitated a complementary evolution of the market. The increased demand for human advisors indicated that RA platforms had expanded the consumer base and enhanced overall market engagement. This dynamic response among market participants revealed a form of co-evolution, whereby technology adoption by one segment of the market encouraged further innovation and adaptation by others. The study made a novel contribution by analysing firm-level strategic behaviour in the wake of fintech adoption. It also emphasised the significance of comprehending these patterns for policymaking purposes, particularly in the context of formulating regulations that can balance innovation with market welfare.

Scherer and Lehner (2025) assessed how closely commercial robo-advice adhered to normative portfolio theory. They web-scraped portfolio recommendations for 151,200 investor profiles from a major US robo-advisor to test whether advice reflected Merton-style hedging demands and balance-sheet heterogeneity. They found that stated goals and investment horizons materially shifted recommended equity weights, yet hedging demands linked to human capital, liabilities, and systematic risk exposures were largely ignored. The provider mapped questionnaire responses to one of eleven portfolios approximating an efficient frontier, privileging simplicity and client comprehension over multi-factor optimality; in such settings, the market portfolio sat below the efficient set, and factor exposures were unsuitable. Augmenting with Financial Industry Regulatory Authority (FINRA) survey data, the authors showed that demographic characteristics helped explain robo-advice adoption. Overall, commercial implementations traded theoretical completeness for scalability and user-friendliness, motivating designs that better incorporate investor heterogeneity without sacrificing accessibility.

Taken together, prior work establishes robo-advisory as a low-cost, scalable channel that can broaden participation, yet real-world uptake, trust, and personalisation remain uneven and increasingly mediated by hybrid human-machine models (D'Acunto & Rossi, 2021, 2023; Fisch et al., 2019; Jung et al., 2019; Kaya & von Martens, 2020; Kumar, 2025). At the profiling stage, many providers

still rely on ad hoc scoring with limited psychometric validity, sparse use of validated scales, and only partial consideration of household balance sheets (Faloon & Scherer, 2017; Scherer, 2017; Scherer & Lehner, 2022; So, 2021; Tertilt & Scholz, 2018). Demand for deeper elicitation is heterogeneous and predictable in terms of literacy and demographics, suggesting that one-size questionnaires are inefficient (Blaschke & Kriebel, 2021).

On the portfolio side, recommendations frequently map many client inputs into a few model portfolios, producing wide cross-platform dispersion, especially for conservative and moderate types, and domicile effects in equity weights (Beketov et al., 2018; Boreiko & Massarotti, 2020; Scherer & Lehner, 2022). Defaults and presentation matter economically, e.g., equity-bond anchors that depress equity exposure while lowering fees (Bräuer, 2021). Normatively, commercial implementations trade theoretical fidelity for simplicity, often ignoring hedging demands from human capital and liabilities (Scherer & Lehner, 2025). Against this backdrop, comparative evidence that a general-purpose LLM can sometimes match or exceed first-pass suitability further underscores shortcomings in today's workflows (Oehler & Horn, 2024).

Methodologically, MPT remains the dominant disclosed framework, with more advanced overlays comparatively scarce; when adopted, robust formulations and dynamic controls are recommended to mitigate instability and operational complexity (Beketov et al., 2018; Bourgeron et al., 2018; D'Acunto & Rossi, 2021). While AI is widely discussed, its diffusion into production portfolio engines has been limited by opacity and perceived fragility (Tsai & Chen, 2022), even as institutions converge on hybrid governance (Jung et al., 2019; Kumar, 2025).

While the existing literature portrays robo-advisory as a low-cost, scalable channel that can increase retail participation, it also documents significant shortcomings in personalisation, profiling rigour, and transparency. At the intake stage, questionnaires tend to be short and vary between providers, only partially capturing household balance sheet realities. Validated psychometrics are rarely used, and sustainability prompts are implemented inconsistently. This implies that neither the breadth of information elicited nor the typical wording of items can be characterised by a single market. Furthermore, there is a lack of current market analyses based on a broader, up-to-date inventory of RA questionnaires to review

further developments such as sustainability requirements. Therefore, a systematic census of items and a defensible taxonomy are required to establish what providers actually ask, how consistently they do so and which questions appear to influence later allocations (see RQ1).

Regarding portfolios, evidence suggests that many platforms consolidate diverse client inputs into a limited selection of model portfolios based on mean–variance logic, with significant cross-platform variation in equity weights, particularly at lower risk levels. Defaults and presentation have economically meaningful effects, advanced allocation overlays exist but are far from universal, and the practical treatment of rebalancing, beta/correlation control, and drawdown management varies considerably. These observations justify a comprehensive mapping of portfolio architectures and instrument mixes aligned to clear equity exposure buckets to identify differences, common patterns and 'typical' questions. Similarly, there is a lack of up-to-date market analyses on the status of portfolio construction with regard to asset classes and weightings and on the methods used by current RAs to study further developments (see RQ2).

Methodologically, the field still lacks a formalisation of the entire RA process, including profiling and portfolio construction subprocesses, that is dynamic and end-to-end. Current formulations are either absent or static or treat modules informally in isolation. This leaves a gap for a coherent mathematical framework that generalises prior approaches, makes explicit the control levers (e.g., dynamic rebalancing rules), and can be implemented at scale (see RQ3).

Finally, the rapid advancement of LLMs raises the practical question of how a GPT-based robo-advisor can be developed—using a fit-for-purpose questionnaire and an investable universe—to assess the feasibility, consistency, and user-level patterns during live deployment. Of particular interest is whether the relationships observed between profiling variables (e.g., risk-preference items and balance-sheet indicators) and equity exposure replicate or improve upon existing mappings within a compliant workflow (see RQ4).

2.4. PERFORMANCE OF ROBO-ADVISORS

This section outlines empirical studies dealing with the performance of portfolios proposed and managed by RAs. The literature review is conducted as in

the previous Section 2.3, but focuses on performance analyses of RAs with regard to research questions RQ5 and RQ6 from Section 1.2.

Helms et al. (2022) investigated whether RAs delivered portfolio designs and investment performance that justified their growing popularity in the US, UK, and Germany. Against the backdrop of limited academic literature on digital discretionary asset management, the study addressed two key research questions: (i) How do the qualitative design features of leading international RAs differ? And (ii) do these differences result in superior risk-adjusted performance when fees, rebalancing policies, and market conditions are considered?

Their mixed-methods design began with a taxonomy-based census of national markets. 15 RAs (five per country) were selected based on assets under management, disclosure quality, and purely algorithmic execution and were benchmarked qualitatively on regulation, minimum investment, fee schedule, risk-profiling architecture, admissible asset classes, risk metrics, portfolio-construction models, and rebalancing rules. Significant heterogeneity was evident: minimum ticket sizes ranged from EUR/GBP/USD 0 to EUR/GBP/USD 100,000, fee bundles were approximately 0–1% p.a., risk buckets ranged from 5 to 23, and portfolio theories ranged from classic mean–variance to prospect-theoretic overlays.

The quantitative phase evaluated realised portfolio structures for each risk bucket (low, medium, or high). For each underlying ETF, daily total returns were converted into monthly logarithmic returns. Then, 20-year stationary bootstrap simulations (with 20,000 paths each) generated four million synthetic months per instrument. The authors then rebuilt each RA's portfolios, incorporating disclosed fees and rebalancing triggers. Where parameters were opaque, conservative assumptions were imposed, most notably a 10% drift threshold. Standard performance diagnostics were then conducted: annual return, volatility, 95 percent Value-at-Risk (VaR), maximum drawdown (MDD), and three risk-adjusted ratios (Sharpe, VaR, and Sortino). Pairwise performance gaps were computed (400 million differences per comparison), and their statistical significance was assessed.

The results were stark. Across risk classes and metrics, only two RAs, Schwab Intelligent Portfolios (US) and Wealthify (UK), achieved double-digit percentages of significant outperformance relative to their peers. No provider dominated universally. Performance rankings remained remarkably stable across the three

risk-adjusted ratios, underscoring the robustness of the results to the risk measure chosen. The impact of country-specific factors was minimal: each jurisdiction hosted a single top-three performer, and regulatory regimes appeared to influence disclosure more than investment quality. Sensitivity tests revealed that disabling rebalancing increased both returns and risk yet rarely altered the order of the league tables. Substituting a uniform 10% drift rule had a negligible impact. Predictably, stripping out platform fees lifted all risk-adjusted ratios but changed peer rankings only marginally, suggesting that cost differentials were too small to outweigh strategic effects.

The authors concluded that design diversity, rather than geography or pricing, drove RAs outcomes. Most providers failed to deliver consistently superior fee-adjusted risk–return profiles, implying that investors should scrutinise portfolio construction and risk controls rather than assuming that digital sophistication guarantees performance. They added that regulators may need to tighten transparency standards so that algorithmic decision rules and historical results can be evaluated independently.

However, the study does not include a database with real price data based on real money deposits. As a result, the study is forced to make a number of assumptions during the analysis based on initial portfolio allocation. It is evident that real rebalancing, fees, and other effects are either inadequately considered. In practice, achieving complete replication is challenging, unless all trades and positions are copied (mirrored) at the same time. Consequently, for research purposes, it would be preferable to utilise the actual portfolio values to calculate performance, as simulations based on simplified assumptions and unknown information regarding the techniques used by RAs, e.g., how they specifically structure their rebalancing or how reallocations and other portfolio allocation changes take place, have limited significance and do not accurately reflect real performance.

Tao et al. (2021) set out to determine whether fully automated RAs in the US, emblematic of the push towards algorithmic finance in the fourth industrial revolution, deliver superior risk-adjusted returns compared to traditional mutual funds and passive benchmarks. Motivated by the lack of empirical evidence on RA performance and the costs and behavioural issues that affect conventional funds, the authors asked two related questions: First, did RAs outperform equity, fixed-

income, money-market, and hybrid funds once risk was controlled? Second, were any observed advantages robust to alternative asset-pricing models? To answer these questions, the study compiled a balanced dataset of weekly data from 2016 to 2019 comprising 100 US RA portfolios (selected for the public disclosure of net asset values (NAVs)), a matched sample of 100 of the largest equity funds, and 100 funds each from the fixed-income, money market, and hybrid categories, together with the S&P 500, Dow Jones Industrial Average (DJIA), and NASDAQ stock indices. Returns were value-weighted, and the 10-year Treasury yield was used as a proxy for the risk-free rate. Risk factors were drawn from the library by Fama and French (1992, 2004), and a momentum series was engineered according to the method of Liu and Zhang (2008). Performance was evaluated through three lenses. Correcting the classic statistic for skewness and kurtosis, the first lens was an adjusted Sharpe ratio. Secondly, a reward-to-VaR measure connected excess return to extreme downside risk, assessed at the 95% confidence level. Finally, Jensen's alpha (α) was estimated using the Capital Asset Pricing Model (CAPM), the Fama–French three-factor specification, and Carhart's four-factor momentum extension (Carhart, 1997; Jensen, 1968; Jensen et al., 1972; Sharpe, 1964). T-statistics were used to assess significance.

Over the four-year period, RAs achieved an average annual return of 18.5%, surpassing equity funds (17%), hybrid funds (8.2%), fixed-income funds (7.6%), money-market instruments (4.6%), and all three indices (10–13%). Their adjusted Sharpe ratios ranged from 0.81 to 2.18, earning them the top ranking every year, and their reward-to-VaR values ranged from 1.15 to 1.62, again outperforming comparators. Jensen's α remained positive and significant regardless of the model: approximately 0.9% under the CAPM, 1.0% with size and value controls, and 1.1% after adding momentum. By contrast, equity funds recorded modest or negative α once momentum was considered, and passive indices delivered only small positive intercepts. Tao et al. (2021) attributed the superiority of RAs to their low fees, systematic mean–variance optimisation, algorithmic rebalancing, and the absence of behavioural biases or agency conflicts. Sensitivity analysis emphasised that these advantages persisted after alternative risk adjustments, with only hybrid funds approaching RA performance in the most favourable year. In conclusion, the authors argued that RAs not only widened retail access to professional portfolio management but also provided demonstrably better risk-adjusted outcomes than

active and passive incumbents. They recommended that investors and regulators recognise the benefits to investment quality of automation, yet cautioned that broader data sets and evidence from stress periods (e.g., the COVID-19 pandemic) are still needed before definitive recommendations can be made. Chapter IV also addresses this gap in order to analyse the risk-adjusted performance of RAs during the COVID-19 pandemic. The time frame of the study by Tao et al. (2021), from 2016 to 2019, is relatively short, and it is also unclear which RAs the authors used for their analysis. However, the study refers to 100 RAs in the US, but it is likely that this refers to 100 RA portfolios, rather than providers, as the comparison is based on portfolios. This ambiguity arises from the uncertainty surrounding the nature of the entities in question. Specifically, it remains unclear whether these entities are RAs that manage funds automatically or whether there are conventional funds that are distributed by RAs.

Puhle (2019) conducted a study to examine whether RAs operating in Germany could achieve investment returns that surpassed market benchmarks and whether disparities in cross-provider performance could be attributed to variations in asset allocation strategies. The study utilised monthly net returns for a selection of medium-risk portfolios (comprising approximately 50% equities and 50% bonds) held with easyfolio, fintego, quirion, SutorBank, and Vaamo. The analysis encompassed the period from May 2015 to December 2018. The figures, drawn from secondary market data provider BrokerVergleich real-money accounts, were compared with a passive 50/50 benchmark composed of the MSCI World (EUR) and Bloomberg Barclays Global Aggregate (hedged EUR). The gross and “after-fee” benchmark series were constructed by deducting typical ETF and RA charges.

The performance of the investment strategy was assessed through the application of descriptive statistics, Sharpe’s risk-adjusted ratio, and Jensen’s α , the latter of which was derived from a single-factor CAPM with the MSCI World Index as the market and the one-month Euro Interbank Offered Rate (EURIBOR) as the risk-free rate. In order to ascertain effective asset mixes, Puhle (2019) applied Sharpe (1992) returns-based style analysis, regressing RA returns on five mutually exclusive factors—developed-market bonds, developed-market equities, emerging-market equities, Treasury Inflation-Protected Securities (TIPS), and commodities—under non-negativity and full-investment constraints. The resulting weights permitted the construction of replicating portfolios whose behaviour was

contrasted with the originals, and a “common style” metric gauged overlap among providers.

It was observed that none of the five RAs managed to outperform the benchmark, either before or after accounting for fees. The average monthly returns ranged from -0.06% (easyfolio) to 0.06% (Vaamo), in comparison to 0.17% for the gross benchmark and 0.06% for the net series. The Sharpe ratios ranged from -0.02 to 0.05, which is well below the benchmark of 0.20 after fees. The findings revealed that Jensen’s α were uniformly negative and statistically insignificant, indicating that the managers contributed no excess return once market exposure was controlled. However, a considerable degree of dispersion was observed among RAs: Vaamo exhibited a 5.2 percentage point cumulative advantage over easyfolio, while SutorBank demonstrated a maximum drawdown of 10.6%, in comparison to fintego’s 7.6%.

Style analysis explained between 88% and 95% of return variance, showing that all providers maintained roughly half-equity, half-bond profiles, though fintego showed an idiosyncratic tilt toward TIPS. Despite the high stylistic similarity between Quirion and SutorBank (over 92% common style), allocation differences could not account for the observed performance spread. This was evidenced by the far smaller divergences exhibited by mimicking portfolios with identical factor weights. The author attributed the residual variation to unmodelled exposures (such as size, value, or momentum factors), implementation quality, ETF selection, and rebalancing technique.

The study concluded that German RAs neither outperformed a simple passive alternative nor delivered homogeneous outcomes, even for investors with comparable risk appetites. Therefore, the selection of an underperforming provider has the potential to result in a substantial erosion of terminal wealth. Nevertheless, the study has significant limitations. The period under consideration is relatively brief, spanning a mere 3.5 years, and encompasses only five portfolios, with one portfolio per RA. Moreover, it is important to note that the sample exclusively comprises portfolios with a stock-to-bond ratio of 50:50. Hence, the findings are not indicative of the broader RA industry in Germany or of all portfolios with varying asset allocation structures and, consequently, disparate risk profiles with equity exposures ranging from 0 to 100%. This is a gap that is addressed in Chapter IV.

Lee (2021) provided the first comprehensive evaluation of South Korea's nascent RA sector five years after its 2016 debut. The study posed three interconnected questions: firstly, the size and diversity of the domestic market; secondly, the effectiveness of different RA types in managing assets; and thirdly, whether superior portfolio performance would result in client acquisition. The author has drawn on data from Korea Securities Computing Corporation (Koscom), supplemented by proprietary desk research, to map 31 active providers (namely banks, securities firms, and RA companies). The author then proceeded to classify the services of these providers as falling into the following categories: product recommendation, advisory, discretionary, or hybrid.

To conduct a performance assessment, Lee (2021) enlisted the services of 14 providers (19 algorithms) and conducted continuous observations over the period from October 2016 to March 2021. However, the analysis was restricted to domestically invested portfolios. The daily returns were subjected to regression analysis using the Fama–French three-factor model, with the aim of capturing market-timing and rebalancing skill. The heteroskedasticity of the data was addressed through the implementation of Newey-West corrections. The client-gathering abilities of 13 providers (22 algorithms) were analysed using monthly panel data from June 2019 to March 2021. The analysis employed random-effects models to relate subscriber counts to annualised returns, Sharpe ratios, volatility, and a set of time-variant design variables (provider category, distribution channel quality, minimum ticket fee, fee schema, service and asset type).

The market census revealed accelerated growth, with 380,000 accounts recorded by mid-2021, yet a mere KRW 1.8 trillion (approx. USD 12.6 billion) under management. Furthermore, the census revealed a notable degree of diversity, with 91% of assets held by bank-run fund-recommendation engines and a surge in advisory/discretionary start-ups following the retail-trading boom caused by the pandemic. The performance results indicated significant heterogeneity. Across the full range of investor risk profiles, merely nine algorithms exhibited statistically significant positive α , and none concomitantly attained abnormal excess returns, a beta near unity, and a positive rebalancing coefficient. The majority of betas ranged around 0.2, indicating minimal market tracking and, consequently, limited diversification or substantial cash reserves. In four out of five cases, the rebalancing terms were either insignificant or negative, indicating that automated timing had a

negligible or even deleterious effect on the outcome. In the context of financial analysis and investment strategy, the performance of product-recommendation and pure advisory models was found to be marginally superior to that of discretionary systems. Furthermore, it was observed that securities-firm offerings exhibited the strongest stock-selection effects. But it is important to note that banks showed a remarkable ability to mimic benchmark returns. The study found that increased subscriber numbers were associated with higher Sharpe ratios and lower volatility. However, the study observed that raw return levels did not contribute significantly. It is important to note that distribution variables dominated. Bank branding and mobile-app accessibility had large, significant positive effects, whereas securities houses, despite comparable risk-adjusted results, lagged badly in attracting clients. Furthermore, minimum investment thresholds and charging structures proved irrelevant once provider identity was controlled.

The indicated that Korean RA services exhibited significant heterogeneity, with consumer outcomes being contingent on the specific provider selected. Given the evident performance gaps, fragmented disclosure, and blurred regulatory boundaries between recommendation, advice, and information services, the study urged the implementation of a “same-activity, same-rule” framework, mandatory algorithm and performance reporting, conflict-of-interest audits, and infrastructure such as fractional-share trading to facilitate proper diversification for small investors. In the absence of such reforms, the rapid growth in subscribers may amplify the risks associated with mis-selling and erode confidence in digital wealth management.

The four studies discussed yield largely mixed results regarding the performance of RAs. Using reconstructed portfolios in the US, UK, and Germany, Helms et al. (2022) found that only a few providers achieved statistically robust outperformance. Fee-adjusted rankings were largely insensitive to rebalancing or fee assumptions, but most providers underperformed. In Germany, realised mid-risk portfolios in Puhle (2019) underperformed a simple 50/50 benchmark, exhibiting negative Jensen’s α and notable dispersion, which is plausibly driven by factor tilts and implementation quality. In contrast, Tao et al. (2021) found that, on average, US RA portfolios deliver higher risk-adjusted returns and positive alphas than mutual funds and indices. Lee (2021) documented significant heterogeneity in

South Korea, finding only a few positive alphas and weak timing effects, as well as subscriber growth that is more influenced by brand and mobile access than raw returns. They also note the widespread underperformance of RAs.

While Helms et al. (2022) and Puhle (2019) found systematic underperformance of RAs in the US, UK, and Germany, Tao et al. (2021) found outperformance of RAs in the US, and Lee (2021) found partial outperformance of some RAs in South Korea. However, the authors note in their studies that there is considerable variation in performance between RAs. The issues are that the studies either use simulated data, secondary data, or data where it is not clear which RAs and portfolios are actually included. The time frames are too short, and the number of portfolios is small. This partly explains the contradictory results and weak significance. Another flaw is that the studies do not divide the entire spectrum of RA portfolios by asset allocation for benchmark comparisons in order to assign the correct benchmarks with the same asset type weighting (stock-to-bond ratios) for the performance analyses for each risk class or equity exposure bucket. For this reason, this study covers a significantly longer period and is based on a broader, more reliable, and more transparent database of real RA portfolio prices for performance analyses depending on equity exposure (see RQ5). It also includes a separate performance analysis for the COVID-19 downturn period when the markets crashed (see RQ6).

III – EMPIRICAL ANALYSIS OF INVESTOR PROFILING AND PORTFOLIO STRATEGIES BY ROBO- ADVISORS IN GERMANY

III - EMPIRICAL ANALYSIS OF INVESTOR PROFILING AND PORTFOLIO STRATEGIES BY ROBO-ADVISORS IN GERMANY

3.1. DATA AND METHODS

This section sets out the data and methods that underpin the analysis of robo-advisory profiling and portfolio design. It proceeds in five steps. Subsection 3.1.1 delineates the market research protocol and selection rules for German providers and documents how each RA is profiled repeatedly to capture the full breadth of questionnaire paths. Subsection 3.1.2 then compiles the raw materials—questionnaires and portfolio recommendations—into a structured corpus and metadata set covering asset allocations and equity exposures. While Tahvildari (2025c) focuses solely on analysing the questions and portfolio compositions, the analyses in this chapter delve much deeper into the questions in the RAs’ questionnaires, as well as the asset classes and weightings, and also include evaluations of the strategies and techniques applied by them (see Subsection 3.1.2).

Building on the obtained results and beyond Tahvildari (2025c), Subsection 3.1.3 applies TF-IDF with clustering to derive and validate question categories with typical questions, quantify term salience, and determine category granularity via standard internal validity criteria.

Furthermore, Subsection 3.1.4 develops a formal stochastic model of the RA process that integrates profiling uncertainty, allocation logic, market-state dynamics and event-driven rebalancing within a unified optimisation framework. The entire RA process is formally modelled, and the gap between previous understandings of the building blocks and the mathematical portfolio allocation strategies is closed.

Additionally, Subsection 3.1.5 operationalises a GPT-based RA prototype, generates a sample of 100 investor profiles, and analyses the resulting portfolio advice using standardised variables and filtered best-subset OLS for equity-share modelling. The model is developed and implemented based on the typical questions previously identified and is intended to serve as a blueprint for future LLM-based RAs.

3.1.1. Market Research and Selection of Robo-Advisors

The first phase was an in-depth market analysis of RAs operating in Germany, due to the high market relevance and the disruptive potential for asset management (Kordela, 2018; Statista, 2025). The German RA market has undergone substantial growth since the introduction of RAs such as Quirion in 2013. With an AUM of EUR 17.6 billion (approximately USD 19 billion) in 2024, it is the largest RA market in Europe and the second largest worldwide after the United States (approx. USD 1.459 billion). Consequently, the German market constitutes a suitable setting for the analysis of investor profiling and asset allocation processes.

The data collection included all RAs active in Germany in July 2024 to provide the broadest possible empirical basis for the analysis. It was based on the dataset compiled by Tahvildari (2025c), which was conducted in July 2024. However, RAs that existed prior to July 2024 but no longer operate as a distinct entity (e.g., due to liquidation, takeover, or closure) were excluded. If a provider marketed its services under several brand names, it was only counted once to avoid duplication. As a result, the data collection method tends to be biased towards survival: companies that have exited the market are systematically omitted. This selection process tends to lead to an upward bias in estimates of the typical quality and resilience of providers and to an underestimation of downside risks and heterogeneity, as platforms with poor performance or other problems are less likely to be represented in the sample (see limitations in Section 5.3.1).

Then, the profiling process for each RA was repeated manually numerous times with different combinations of answers to the questionnaires. All questions and possible answers displayed on the RA providers' websites and applications were recorded and documented. The data collection method, which consists of selecting RAs on the German market to then undergo the profiling processes and documenting the questions asked by RAs and the recommended portfolios, has already been used in previous studies such as Kordela (2018), Tertilt and Scholz (2018), Scherer and Lehner (2022), and Oehler and Horn (2024).

3.1.2. Collection of Questionnaires and Portfolio Data

Subsequently, the material was then subjected to inductive qualitative content analysis (Elo & Kyngäs, 2008; Mayring, 2015), whereby question categories

were defined from the extracted question and answer material, and the questions were assigned to these categories for each RA. Examples of question categories include risk preferences and financial situation. Additionally, the number of questions per category was recorded for each RA. Previous studies (Scherer & Lehner, 2022; So, 2021; Tertilt & Scholz, 2018) have used similar classifications, but only So (2021) has described an inductive procedure in detail.

The method applied here likewise employed an inductive approach to the content analysis of questions from questionnaires to develop categories, because no other studies have examined these RAs, and no prior theory exists to test. The topics covered by the questions are already apparent in the RA profiling process, as the questions usually have clear themes/headings that serve to categorise them, e.g., goals, investment amount or risk preference. These topics have already been used as a guide in the preceding studies (Scherer & Lehner, 2022; So, 2021; Tertilt & Scholz, 2018).

The next step involved repeatedly observing and comparing the questions from the questionnaires to identify patterns, followed by developing content categories through inductive reasoning. This reduced the number of specific observations to generalisations. After defining the question categories, the wording of the questions within each theme was reviewed to select prototypes that summarise the underlying content with minimal loss of information. Through this approach, redundancies caused by nearly identical questions from different RAs were removed. The resulting consolidated matrix is therefore representative at the category level, and descriptive statistics were then generated for the number of questions per category.

Thereafter, the investment proposals generated after profiling were evaluated. Portfolio data was documented along with the strategies and portfolio components. The composition analysis considered the entire asset allocation, documenting each asset class and weighting of equity exposures. This information was compiled from public websites and accompanying regulatory disclosures. Furthermore, the entire profiling process was performed multiple times for all 45 RAs identified to ensure comprehensive coverage of portfolio structures and underlying asset compositions. The portfolios were then examined using descriptive statistical methods.

3.1.3. Term Frequency-Inverse Document Frequency (TF-IDF) and Cluster Analysis for Categorisation of Questions

Subsequently, the text analytics method term frequency-inverse document frequency (TF-IDF) was applied with cluster analysis measures to evaluate the questions in the RAs' questionnaires and the previously derived question categories with the most important questions associated with them (Bafna et al., 2016; Christian et al., 2016; Khan et al., 2019). The clusters in this application were the question categories. TF-IDF is a weighting scheme that is widely used in information retrieval and text mining to evaluate the importance of a term (e.g., a word) within a specific document in a corpus (Al-Obaydy et al., 2022; Bafna et al., 2016; Dey & Das, 2023; Havrlant & Kreinovich, 2017; Khan et al., 2019; Liu et al., 2018).

In this application, the documents are the questions extracted from the RAs' questionnaires. The method is explained as follows. Let

$$Q = \{q_1, \dots, q_m\} \quad (1)$$

be the set (corpus) of $m \in \mathbb{N}$ questions extracted. The data set is a sheet with one question per row in a single column (long format). Then, denote by

$$V = \{v_1, v_2, \dots, v_n\} \quad (2)$$

the vocabulary, i.e., the set of all $n \in \mathbb{N}$ unique terms (N -grams). An N -gram consists of $N \in \mathbb{N}$ words separated by spaces. The goal is to represent each question q_i by a numeric vector

$$x_i = (w_{i,1}, w_{i,2}, \dots, w_{i,n}) \in \mathbb{R}^n, \quad (3)$$

where $w_{i,j}$ is the TF-IDF weight of the vocabulary term v_j in question q_i . After this, each question q_i is represented by a list of terms, which form candidates for V . For each question q_i and each vocabulary term $v_j \in V$, define the raw count:

$$\text{count}(v_j, q_i) = |\{u \in q_i \mid u = v_j\}|. \quad (4)$$

When normalising by the question's length, the term frequency (TF)

$$TF(v_j, q_i) = \frac{\text{count}(v_j, q_i)}{\sum_{u \in q_i} \text{count}(u, q_i)} \in [0,1] \quad (5)$$

is obtained. TF measures how often a term appears in a question (document) and is normally normalised by question (document) length to prevent bias towards longer texts. The numerator represents the number of times v_j appears in question

q_i , and the denominator represents the total number of token occurrences in q_i . The document frequency of term v_j is

$$n_j = |\{q \in Q | v_j \text{ appears in } q\}|. \quad (6)$$

It is the number of questions in which v_j occurs. The inverse document frequency (IDF) can be calculated as

$$IDF(v_j) = \ln\left(\frac{m}{n_j + 1}\right), \quad (7)$$

where $m = |Q|$ is the total number of questions. The “+1” in the denominator avoids division by zero if a term happens not to appear. IDF quantifies how unique a term is across all questions. Common terms ($n_j \approx m$) receive $IDF \approx \ln(1) = 0$. Rare, discriminative terms with small n_j receive larger IDF values. The TF-IDF weight for term v_j in question q_i is the product

$$w_{i,j} = TFIDF(v_j, q_i) = TF(v_j, q_i) \cdot IDF(v_j). \quad (8)$$

Thus, each question q_i is mapped to the TFIDF vector $x_i = (w_{i,1}, w_{i,2}, \dots, w_{i,n})$ from (3). Multiplying TF by IDF assigns higher weights to terms that are frequent in a given question but rare across the set of all questions, which effectively emphasises discriminative terms and reduces the importance of common stop words such as “the”, “is”, “and” etc. (Sarica & Luo, 2021). The weight vectors x_i deliver the TF-IDF $m \times n$ -matrix

$$\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} w_{1,1} & \cdots & w_{1,n} \\ \vdots & \ddots & \vdots \\ w_{m,1} & \cdots & w_{m,n} \end{pmatrix} \in \mathbb{R}^{m \times n}. \quad (9)$$

The representations (8) and (9) transform textual data into numerical feature vectors and matrices, respectively, enabling algorithms from cluster analysis to operate on content based on term importance and on the principle of feature similarity (Al-Obaydy et al., 2022; Jin & Han, 2017; R. Khan et al., 2019; Likas et al., 2003; Sinaga & Yang, 2020). Once, the TF-IDF matrix is derived, for any question category $C \subseteq Q$, the sum of TF-IDF weights for each term

$$S(v_j, C) = \sum_{q_i \in C} w_{i,j} \quad (10)$$

is the category-level term score. The normalised category-level term score is

$$\bar{S}(v_j, C) = \frac{1}{|C|} \sum_{q_i \in C} w_{i,j}. \quad (11)$$

Sorting all v_j by descending $S(v_j, C)$ or $\bar{S}(v_j, C)$ provides the most characteristic terms for the category C . For a $K \in \{2, 3, \dots, K_{max}\}$, let $\{C_1, \dots, C_K\}$ be the partition of K categories

$$C_k = \{i | q_i \text{ is assigned to question category } k\} \quad (12)$$

with $k = 1, \dots, K$. For each category $C_k \in \{C_1, \dots, C_K\}$ with $C_k \subseteq Q = \{q_1, \dots, q_m\}$, the centroid vector (mean vector) of category C_k is

$$\mu_k = \frac{1}{|C_k|} \sum_{q_i \in C_k} x_i, \quad (13)$$

implying its centroid quadratic norm

$$\|\mu_k\|^2 = \sum_{j=1}^m \mu_{k,j}^2 \quad (14)$$

where $\mu_{k,j} = \bar{S}(v_j, C_k)$ is the j th component of μ_k . For each question q_i , its squared Euclidian distance (quadratic norm) to each centroid is given by

$$\text{dist}(i, k)^2 = \|x_i - \mu_k\|^2 = \sum_{j=1}^n (w_{i,j} - \mu_{k,j})^2. \quad (15)$$

The within-category sum of squared errors (WCSSE) per category C_k is

$$WCSSE_k = \sum_{q_i \in C_k} \text{dist}(i, k)^2, \quad (16)$$

and the sum of WCSSE over all categories

$$WCSSE = \sum_{k=1}^K WCSSE_k. \quad (17)$$

The quantities $WCSSE_k$ and their total $WCSSE$ measure how tightly each question is grouped around its category centroid per category-level and overall, respectively. The total sum of squared errors (TSSE) of the weight vectors is

$$TSSE = \sum_{i=1}^m \|x_i - \bar{x}\|^2 \quad (18)$$

with the global mean of the weight vectors

$$\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i, \quad (19)$$

It satisfies the identity

$$\sum_{i=1}^m \|x_i\|^2 = TSSE + m\|\bar{x}\|^2, \quad (20)$$

the proof of which is given in APPENDIX 1. The between-category sum of squared errors (BCSSE) given by the difference $BCSSE = TSSE - WCSSE$ yields the Calinski–Harabasz index (CHI)

$$CHI = \frac{(m - K)BCSSE}{(K - 1)WCSSE}, \quad (21)$$

which is a metric used to evaluate the quality of clustering (Wang & Xu, 2019; Zhao & Fränti, 2014). As a ratio between the separation between categories (BCSS) and the compactness within a category (WCSSE), CHI represents an indicator for geometrically optimal granularity.

To ensure that the TF-IDF representation captures only those lexical items most informative, a sequence of carefully chosen parameter filters was applied. Firstly, the built-in English stop-word list from the open-source Python library *Scikit-Learn* (`stop_words = 'english'`) was invoked to eliminate common function words such as “and”, “the”, or “is”, which appear ubiquitously but convey little topical content (Bisong, 2019; Hao & Ho, 2019; Kramer, 2016; Moncecchi, 2013; Pedregosa et al., 2011). Secondly, a blacklist of tokens (terms) is used to clean up the raw data of the questions for the TF-IDF algorithm for analysing the questions, which can be found in APPENDIX 2.

Unigrams, bigrams, trigrams and quadrigrams (1 to 4-grams) were used to capture more content-specific terms. Higher dimensional N-grams frequently capture more precise semantics (for instance, “risk appetite” versus “risk” alone), thereby improving the clarity of subsequent category interpretation. Subsequently, the minimum number of questions required for inclusion was specified as two, i.e., any N-gram (e.g., “financial” or “investment horizon”) must occur in a minimum of two questions. Moreover, any term that appeared in more than 80% of the questions was eliminated, leading to the removal of expressions that were almost omnipresent and offered minimal discriminatory power. The implementation of these parameterisations resulted in a compact, higher-quality feature set, which was conducive to a robust and reproducible analysis of the questions and categories. The executable Python script for the TF-IDF algorithm can be found in APPENDIX 3.

The question categorisation in Subsection 3.1.2 was used and further analysed and evaluated by applying the TF-IDF algorithm. In addition, the categorisation was gradually reduced by merging categories to calculate the TF-IDF algorithm for different numbers of categories K . The key figures for the categories were then compared. To determine the optimal number of categories K^* , $WCSSE(K)$ was plotted on a vertical axis against each K in the candidate range $\{2, 3, \dots, K_{max}\}$ on a horizontal axis. The graph

$$\Gamma(K, WCSSE) = \{(K, WCSSE(K)) \in \mathbb{N} \times \mathbb{R}_{\geq 0} \mid K \in \{2, 3, \dots, K_{max}\}\} \quad (22)$$

is the “elbow curve”. The elbow curve was then inspected to identify the point at which the rate of decrease in $WCSSE(K^*)$ slows sharply (“elbow”). This method is called the “elbow method” (Bholowalia & Kumar, 2014; Khan et al., 2019; Syakur et al., 2018). For the geometric evaluation, the two discretisations of the second derivative were calculated, i.e. the backward difference

$$\Delta^2 WCSSE(K) = WCSSE(K) - 2WCSSE(K - 1) + WCSSE(K - 2) \quad (23)$$

and the central difference

$$\delta^2 WCSSE(K) = WCSSE(K + 1) - 2WCSSE(K) + WCSSE(K - 1). \quad (24)$$

The backward difference of first order is

$$\Delta WCSSE(K) = WCSSE(K) - WCSSE(K - 1). \quad (25)$$

3.1.4. Modelling of the Building Blocks within the Robo-Advisory Process

This step comprised the development of a formal mathematical (stochastic) model of the RA process based on the building blocks identified in Subsection 2.1. Studies like Faloon and Scherer (2017), Beketov et al. (2018) and Jung, Dorner, Glaser, et al. (2018) described the building blocks and their connections within the RA process qualitatively or visually but did not involve any formal mathematical modelling of the process. Bourgeron et al. (2018) and Perrin and Roncalli (2020) paved the way by showing that modern optimisation techniques, such as alternating-direction methods, coordinate descent, and proximal operators, could be used to solve high-dimensional portfolio allocation problems more effectively than classical quadratic programming. Building on these insights, a unified, end-to-end framework was developed that integrates investor profiling, asset

allocation, market state modelling, and dynamic rebalancing in a single mathematical structure.

Methodologically, the questionnaire responses were modelled as a stochastic mapping into a low-dimensional profile space and market factors were simulated via time-dependent processes. Event-driven optimisation was then applied, balancing tracking errors against transaction costs to determine each portfolio update. Through this approach, previous work by Bourgeron et al. (2018) and Perrin and Roncalli (2020) was extended by introducing profiling uncertainty directly into the control problem and formulating the allocation step as a stochastic optimal control problem rather than a static optimisation problem. Formalising the RA process, the model provides a theoretical and conceptual foundation for future mathematical and technical advancements.

3.1.5. Development of a Custom GPT-Based Robo-Advisor

For the analysis, design and development of the profiling process in ChatGPT-4o, a questionnaire and portfolio universe must be defined to propose a customer profile and possible portfolios. The model ChatGPT-4o was used as a tool for developing and implementing the GPT-based robo-advisor (OpenAI, 2025). Table 2 in Subsection 3.2.1 provides a set of typical questions that RAs ask in the investor profiling process, along with the common investment universe for portfolio construction in Subsection 3.2.2. These were used as a basis for the implementation in ChatGPT-4o-based RA.

For ease of application, some questions were adapted or removed. The question about which assets to invest in was omitted because the GPT-based RA is supposed to suggest the portfolio composition itself based on the investor profile. Regarding sustainability preferences, only the question of whether to invest sustainably was retained with the corresponding ESG (environmental, social, governance) ETFs of the asset classes stocks, bonds, and real estate. The rationale behind this lies in the design of general ESG instruments, which were intended to mitigate complexity arising from the aggregation of numerous smaller weightings of various instruments within portfolio compositions whilst facilitating the implementation of allocation.

Furthermore, the results of the analysis in Subsection 3.2.2 demonstrate that the presence of more detailed specific sustainability preferences does not exert any influence on the portfolio, given that only general ESG instruments are given consideration. The questions on tax residence and politically exposed individuals were also being removed, as they are too specific and not appropriate for the diversified portfolio composition in the context of implementation. These questions are intended to fulfil regulatory, tax, and other legal checks and obligations for opening a robo-advised investment account and not for the specific portfolio composition of assets with weightings. The remaining 32 questions with corresponding answer options, presented in APPENDIX 5, were loaded into ChatGPT-4o. The defined investment universe, consisting of 17 standard instruments derived from Subsection 3.2.2, is listed in APPENDIX 6.

The RA profiling was then implemented in ChatGPT-4o using the defined questionnaire with answer options and the investment universe. Subsequently, a sample of 100 investors is generated who complete the questionnaire and thus go through the profiling process in ChatGPT-4o. The 100 investors' output, which included all answers to each of the 32 questions, formed the data basis (32×100 matrix) for analysing the profiling and portfolio recommendation process of the GPT-based RA. This approach was performed through human-machine interaction, whereby numerous identical input prompts were executed one after the other to provide the GPT-based RA the instructions it needs to process the data. The process involved successively adding a random investor to the sample and running the analysis until determining the portfolio allocation. ChatGPT-4o was used in this study to create these investors due to the existing functionality of the AI model.

However, this step could also be handed over to the GPT-based RA in an alternative way, as the purpose of this analysis was to create and evaluate the investor profile through ChatGPT and to make portfolio recommendations, including specific allocations, based on the responses of individual investors, rather than to create a sample of investors. In addition, further requirements for investor responses were given via additional prompts to create a varied sample of investors for the creation of portfolio recommendations. The individual profiling runs for each investor using the GPT-based RA were necessary because of the high computational intensity involved, as well as the need for verification and potential

corrections before proceeding to the next investor. Consequently, a sample of 100 investors, each with their own profile, was created. They all received a portfolio recommendation from the GPT-based RA.

The results for the investor profile with the proposed portfolios by the GPT-based RA were then examined using descriptive and inferential statistics. The portfolio allocations were studied to understand how GPT-based RA allocates investors' assets, with a close focus on the equity components of the portfolios, as these are essential for determining portfolio risk (Asness, 2000; Bhansali & Hsu, 2012; Liew & Ajakh, 2020). Accordingly, the responses to questions on risk preference in relation to the equity share of portfolios were evaluated.

Subsequently, numerous multilinear ordinary least squares (OLS) regression analyses were performed to measure the effect of 15 independent variables derived from the questionnaire's responses to investigate the effects on the equity exposure Y (dependent variable) of the proposed portfolios. The modelling method and procedure for performing the OLS regression analyses, including filters and the selection process for the "best" models to be chosen, including prior z-scaling transformation, will be explained hereafter.

Due to the varying scales of the independent variables, and in order to avoid distorting the regression by dominant variables with greater variance, each observed question variable $X_i \in \{X_1, \dots, X_{15}\}$ was standardised via the transformation (z-scaling)

$$Z_i = \frac{X_i - \mu_i}{\sigma_i}, \quad (26)$$

where μ_i and σ_i are the corresponding empirical mean and standard deviation. Z-scaling standardises the numerical value of all independent variables (mean = 0, standard deviation = 1), enhances the condition of the design matrix, and facilitates the interpretation of the β 's as the effect per one-standard-deviation change, which is advantageous when the units vary significantly. Given the 100×15 matrix of standardised data

$$Z = (Z_{j,i})_{\substack{j=1,\dots,100, \\ i=1,\dots,15}}, \quad (27)$$

where $Z_{j,i}$ is the value of the observation (investor) $j \in \{1, \dots, 100\}$ on the independent variable $i \in \{1, \dots, 15\}$, the basic (multi)linear OLS regression model for the equity exposure Y_j is defined by

$$Y_j = \alpha + \sum_{i \in I} \beta_i Z_{j,i} + \varepsilon_j \quad (28)$$

with parameterisations α, β_i and a selection $I \subseteq \{1, \dots, 15\}$ of independent variables of the corresponding power set $\mathcal{P}(X_1, \dots, X_{15})$. The power set is therefore a set of all possible subsets of combinations of the independent variables, so that a regression model is defined for each combination (sum index set). Since the power set is formed from 15 elements, there are $|\mathcal{P}(X_1, \dots, X_{15})| - 1 = 2^{15} - 1 = 32,767$ non-empty combinations (subsets) and thus regression analyses that can be performed.

Instead of performing 32,767 equations, a best-subset selection can be achieved in (28) simply by setting the “beta” coefficients β_i individually and sequentially to zero or retaining them. For this purpose, a more efficient method was implemented: filtered hierarchical best-subset selection (FHBSS) with significance and variance inflation factor (VIF) constraints, where significance refers to the statistical significance of the independent variables and VIF to their variance inflation factors. The method is founded on the principles of forward selection and best-subset selection and is further enriched by integrating additional quality criteria to enhance efficiency and stability.

The hierarchical approach of the FHBSS was implemented in such a sequence that, commencing with the dimension $k = 1$ independent variables, all 15 simple linear OLS regressions were initially calculated. Subsequently, the dimension was successively increased by incrementing k by 1 ($k \rightarrow k+1$), meaning that in the next step, all pairs of independent variables were calculated, followed by all 3-tuples and then all 4-tuples. After each OLS regression in hierarchy level k , the filter process described below was applied to observe the significance and multicollinearities, as well as other quality criteria. The procedure terminated at the level of the hierarchy $k \leq 15$, wherein there were no further significant models or solely multicollinear models in which not all variables possessed a $VIF < 4$. For each hierarchy level k , the following filtering process was applied to the set of k -linear OLS regression bases (all OLS calculation models with k -tuples) $M_Y(k)$ using the following steps.

As a first filter $\mathcal{S}$ on the set of all k -dimensional models $M_Y(k)$, only those calculated models that contained statistically significant variables at a significance level of 2.5% were retained. This was determined via the two-sided t -test at $\alpha = 0.025$ ($df = 97$) for each coefficient, based on the OLS estimate and its standard

error ($p\text{-value} < 0.025$). Models were retained only if all included variables had absolute t -value $|t_i|$ exceeding the corresponding critical value (see limitations in 5.3.1). In addition, robust estimation using heteroskedasticity-consistent standard errors (HCSE) was also assessed and reported (HC3). Because HC3 provides valid inference under unknown forms of heteroskedasticity, formal heteroskedasticity tests are not required for calculating p -values. However, to characterise the residual structure, Breusch–Pagan tests (assuming normality) and White tests (robust to non-normality) were also conducted for heteroskedasticity. Residual normality was examined using the Shapiro–Wilk test as the primary guidance and a supplementary check via the Jarque–Bera test, which are informative for exact small-sample t -based inference. Moreover, it was also screened for influential observations via the maximum Cook’s distance. These diagnostics did not alter the HC3-based inferential procedure but provided complementary evidence on model specification and supported robustness checks across hierarchy levels k . The focus was therefore on robustness and reliability, ensuring that each variable made a meaningful contribution to explaining the dependent variable.

Multicollinearity was assessed using the variance inflation factor (VIF). As a second filter $\mathcal{V}$, only models in which all variables had a $VIF < 4$ were retained to improve the interpretability and stability of the models, as the presence of high multicollinearity leads to unreliable coefficient estimates and distortion of individual effects, which impairs the interpretability of the results.

As a third filter $\mathcal{R}$, the adjusted R^2 of the remaining models was evaluated, and the models with the highest values were selected to reflect the goodness of fit of the model, considering the number of variables. With each increasing dimension $k+1$ (i.e., additional explanatory variable), the minimum requirement of $\text{adj. } R^2$ was set to the maximum of $\text{adj. } R^2$ of the models with lower dimension k . This ensures that each model with an additional dimension must at least improve $\text{adj. } R^2$. This step of selecting the models with the highest adjusted R^2 helped to mitigate overfitting and preserve the reliability of the estimated effects.

The fourth filter $\mathcal{C}$ consisted of selecting the optimal model based on the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), where lower values were better. AIC and BIC balance model fit and complexity, with BIC tending to favour simpler models. This step served to validate that the

final models are parsimonious and reduced in complexity, and at the same time ensured maximum generalizability. If the selection of the best model according to the criteria presented was not clear or differences were marginal, up to three of the best results were presented to ensure that at least one “best” model was always specified.

The filtering process (29) can be expressed as a composite map of filter functions $\mathcal{S}, \mathcal{V}, \mathcal{R}, \mathcal{C}$ on the set of all k -dimensional models $M_Y(k)$ for Y , where $\mathcal{S}$ stands for the significant models (i.e., the selection of models in which all independent variables are significant at a level of 2.5%), $\mathcal{V}$ is the mapping for the selection of models in which all independent variables have a VIF < 4 in order to select the models with the lowest multicollinearities, $\mathcal{R}$ is for the selection of models with the highest (adjusted) R^2 in order to achieve the highest possible model quality, $\mathcal{C}$ for the selection of models with the lowest BIC and AIC, respectively:

$$\mathcal{C} \circ \mathcal{R} \circ \mathcal{V} \circ \mathcal{S}: M_Y \mapsto \mathcal{C} \left(\mathcal{R} \left(\mathcal{V} \left(\mathcal{S}(M_Y(k)) \right) \right) \right). \quad (29)$$

The filtering process (29) seems to be deliberately conservative and systematic (two-sided coefficient t-tests with HC3 robust standard errors, VIF for multicollinearity, adj. R^2 for model quality, and further diagnostics: Shapiro–Wilk and Jarque–Bera tests, White’s test, maximum Cook’s D). Nevertheless, FHBSS remains constrained by selection-induced bias/multiplicity (step-wise) and reliance on in-sample criteria, so stability and generalisability are not guaranteed. Moreover, although starting at low k is justified because VIF rises sharply with higher k , the hierarchical search can still miss higher-order joint effects. The details of the limitations of the FHBSS method are discussed in Subsection 5.3.1.

3.2. RESULTS

The first subsection, 3.2.1, looks at the findings of the subprocess for investor profiling, examining the questionnaires from all RAs in the sample. Then, Subsection 3.2.2 shares the results of the subprocess for asset allocation, looking at the portfolio structures with assets, weights and strategies. Subsection 3.2.3 documents the outcomes of the TF-IDF and cluster analysis, and the results are compared with those from Subsection 3.2.1. In Subsection 3.2.4, based on the previous results, a mathematical model of the RA process is developed, taking into account additional building blocks, such as rebalancing, investor profile updates,

and market dynamics. The analysis results of the custom ChatGPT-4o-based robo-advisor described in Subsection 3.1.5 are finally presented in Subsection 3.2.5.

3.2.1. Questionnaires for Investor Profiling of Robo-Advisors

Kordela (2018) identified 25 RAs in Germany in 2018, and Scherer and Lehner (2022) identified 30 in 2021, while this study identified 45 active RAs in July 2024, revealing that the market has become significantly more fragmented (see APPENDIX 7). While most RAs offer robo-advisory technology as a standalone service and partner with banks, brokers, and asset managers to meet regulatory requirements, the latter also offer their own RA services to their clients. There is considerable variation in the number of questions that RAs ask their clients as part of the profiling process, as well as the number of portfolios proposed and managed by different RAs on the market. Table 1 lists the names of the RAs from the sample, along with the number of portfolios and questions.

Table 1. Sample Overview of Selected Robo-Advisors in Germany

Robo-advisor	Questions	Portfolios
Mintos	36	4
Fidelity	32	7
ONVEST	32	5
Dr. Lux & Präuner Vermögensverwaltung	24	5
BW-Bank ON	23	5
growney	22	10
Minveo	22	10
Estably	20	10
peningar	20	3
easyInvesto	19	14
Triodos Impact Portfolio Manager	19	6
my-si	19	5
Solidvest	19	5
quirion	18	20
Zeedin	18	4
Gerd Kommer Invest	17	11
Liquid	15	31
LAIC ADVISOR	15	5

digiVestor	15	5
INNO INVEST	14	14
bevestor	14	10
cominvest	14	6
Oskar	14	5
easyfolio	14	3
VisualVest	13	14
Smavesto	13	8
inyova	13	5
Raisin Invest	13	4
Whitebox	12	27
ROBIN	12	16
investify	12	7
Warburg Navigator	12	7
Revolut	12	5
MONVISO	12	3
fintego	11	10
GLS onlineInvest	11	4
truevest	11	4
vividam	11	4
ginmon	10	20
Scalable Capital	9	37
Digital Invest Assets	9	3
Evergreen	6	10
Financery	6	5
Openbank	2	5
Peaks	2	4

Note. Author's elaboration: The sample comprises 45 robo-advisors operating in Germany in July 2024. Questions reports the number of questionnaire items extracted per provider; Portfolios reports the number of model portfolios recommended by each provider. Data were collected and coded by the author from providers' public disclosures and platform questionnaires.

Since the RAs are active on the German market, the questions are obviously in German. They have therefore been translated into English. A total of 687 questions were recorded and categorised into eight different content categories, covering a broad range of topics that are essential for creating investor profiles. During the profiling process carried out by the RAs and analysis of the questions, it was noticed that the questionnaires generally began with the objectives and reasons, then asked about the investment horizon and total amount or savings rate,

before moving on to specific preferences regarding expected returns, risk tolerance, sustainability or other preferences for the investment. The above sequence was typically followed by a survey of the investor's knowledge and experience of the financial markets and their general financial situation, consisting of questions on monthly income, expenses, existing (il)liquid assets and debts. This sequence was characteristic of the profiling process; for which reason, the questions were also categorised and assigned in this order.

The first content category identified across the 45 German RA questionnaires concerned the objectives and purpose of the investment. In the usual course of events, providers initiated the dialogue by requesting that clients specify their objectives concerning the capital at stake. These objectives might include, for instance, the preservation of capital, the attainment of long-term growth, the provision for retirement, or the funding of a particular purchase. The subsequent profiling exercise was anchored by such questions within an explicit goal hierarchy, thereby helping RAs distinguish between discretionary and liability-driven assets and enabling a preliminary consistency check between aspirations and the client's declared constraints. 51 questions were identified in this category.

A second category with 55 questions focused on the investor's investment horizon and age. The respondents were invited to provide details regarding the planned holding period, which was typically segmented into short, medium, and long intervals. They were also asked to disclose their current age or expected retirement date. In many cases, respondents were prompted to indicate the withdrawal period for the invested funds.

The third category, with 63 questions, captured the amount and frequency of deposits. Enquiries made by providers included the request for the initial lump sum, the size of any monthly contributions, and, less frequently, anticipated withdrawals. There were questions regarding investors' preferences for a one-off investment or regular investments (monthly, annual, etc.). Several questionnaires delved more deeply into the client's willingness to adjust their contributions upwards or downwards.

A fourth content category with 119 questions was dedicated to the investor's risk preferences. This broad term integrated classical risk-tolerance gauges, frequently Likert-style statements concerning comfort with loss, with quantitative

loss-capacity scenarios, such as the maximum one-year drawdown a client could withstand without liquidating. In several cases, RAs augmented these elements with psychometric proxies. To illustrate this, stochastic return paths were presented, and the investor was asked to select the line they would choose. Return–risk matrices, bar charts and other visualisations were used to present investors with a range of options for their return–risk profile.

A fifth category, containing 88 questions, was defined to assess the investor’s sustainability preferences regarding the application of environmental, social, and governance filters. The range of options included a simple yes/no tick-box and granular sliders, distinguishing exclusionary screens, best-in-class tilts, and thematic allocations. In some cases, specific ESG areas such as low-carbon investments or socially equitable jobs were also queried as conditions for the public limited companies within the investments.

The sixth content category comprised eight questions concerning other preferences that did not fall under any of the other categories. Examples of these preferences included a desire for human guidance (hybrid advice) alongside digital execution, specific language preferences, tax wrappers, and investment in certain industry sectors or regions.

The seventh content category, with 142 questions, concerned the investor’s knowledge and experience of financial markets. The RA providers were classified according to their prior exposure to instruments such as shares, ETFs, options, or foreign exchange, and their self-reported proficiency was measured. A comprehensive array of enquiries was posed, encompassing the respondents’ familiarity with financial markets, the comparison of investment risks, securities, trading, and transactions conducted, as well as their educational background, occupation, and the highest level of education attained. Additionally, the survey delved into their prior experience with asset management services.

Lastly, the eighth category comprised 161 questions intended to evaluate the investor’s financial situation. This encompassed current income, liquid reserves, outstanding liabilities, and the proportion of net wealth represented by the proposed investment. It was found that certain questionnaires requested a full balance-sheet snapshot, whereas others relied on coarse bands. Questions were also raised about how long investors can live off their existing assets. Table 2 summarises the results of the 34 typical questions of the eight content categories.

Table 2. Representative Questions by Content Category

Category	Questions
Objectives and purpose	What are your investment objectives? What would your money be invested in? How high is your savings target? Who do you want to invest money for?
Investment horizon and age	What is your investment horizon? How old are you? When do you plan to retire? When do you expect to withdraw some of the money you have invested?
Amount and frequency of deposits	Would you like to invest money on a one-off and/or regular basis? How much would you like to invest? (one-off and/or monthly) How much additional money could you invest each year?
Risk preferences	What loss in value of your investment are you willing to accept? How concerned are you about the risk of losing part of your assets? How willing have you been to take risks when making financial decisions in the past?
Sustainability preferences	Do you want to invest sustainably? Which sustainability (ESG) categories should the investment strategy include? What negative impacts should be avoided? In which areas should a positive contribution be encouraged?
Other preferences	Would you like to have an advisor at your side? Would you prefer to prioritise certain sectors and/or themes? Would you like to prioritise regions for your investment? Which project should receive your contribution?
Knowledge and experience	What knowledge and experience (in years) do you have of the following types of securities? How many trades have you made in the following asset classes in the last n years? Have you gained relevant capital market experience through training, studies, or work? What is your highest level of education? Which asset management services have you used in the past?

Financial situation	What is your monthly income and total monthly expenditure? What is the amount of your liquid assets and your total assets (including non-liquid assets)? How high are your liabilities? How long will your reserves last to cover your current expenses and obligations? What is your profession and main source of income? What is your tax residency? Are you a politically exposed person?

Note. Author's elaboration: This table shows eight content categories with a total of 34 representative questions, derived from the analysis of 687 questions from 45 robo-advisors operating in Germany as of July 2024. Although the wording of the questions in the robo-advisors' questionnaires varies, they address the same content and items as these representative questions.

Among the 45 RAs studied, 35 (77.78%) considered the objective and purpose of investment, 41 (91.11%) the investment horizon and age, 38 (84.44%) the amount and frequency of deposits, 44 (97.78%) risk preferences, 36 (80%) sustainability preferences, 41 (91.11%) knowledge and experience, 40 (88.89%) the financial situation, and 7 (15.56%) other preferences such as sector, theme, region, or bank selection as part of the profiling process to create the investor profile. 48.89% (22/45) of RAs asked at least one question within each category when other preferences were excluded. In 7 RAs where the amount and frequency of deposits were irrelevant to the profiling process, investors were able to determine the amounts themselves, regardless of the RA's recommendation.

The RAs asked a total of 15.27 questions on average during the profiling process across all eight question categories. The standard deviation of the number of questions was 6.96, and the range was 34, indicating that the variance was very high. On average, the categories "financial situation" (approx. 3.58), "knowledge and experience" (approx. 3.16), and "risk preferences" (approx. 2.64) were asked most frequently. It is noteworthy that these categories together comprised 380 of the 687 questions, which is more than 55% of all questions asked. Some RAs asked many questions in these categories. In the "risk preferences" category, the maximum number of questions was 16; in the "knowledge and experience" category, up to 22; and in the "financial situation" category, up to 12. In contrast, other categories, such as "objectives and purpose", "investment horizon and age", and "amount and frequency of deposits", contained a maximum of three questions

each. Table 3 shows the descriptive statistics for the number of questions in each of the eight question categories and for the entire sample.

Table 3. Descriptive Statistics of Question Counts by Content Category

Question category	Mean	Median	SD	Min	Max	Range	Sum
Objectives and purpose	1.13	1	0.81	0	3	3	51
Investment horizon and age	1.22	1	0.64	0	3	3	55
Amount and frequency of deposits	1.40	1	0.89	0	3	3	63
Risk preferences	2.64	2	2.70	0	16	16	119
Sustainability preferences	1.96	2	1.73	0	7	7	88
Other preferences	0.18	0	0.44	0	2	2	8
Knowledge and experience	3.16	2	3.49	0	22	22	142
Financial situation	3.58	3	2.58	0	12	12	161
Total	15.27	14	6.96	2	36	34	687

Note. Author's elaboration: Data are based on questionnaire items extracted from 45 robo-advisors (RAs) operating in Germany in July 2024 (total items = 687). Mean = arithmetic mean of item counts per RA in the category; SD = standard deviation; Median = sample median; Min/Max = observed minimum/maximum; Range = Max – Min; Sum = total number of items across providers. The Total row reports the across-category aggregates. Values are rounded to two decimals in the case of mean and SD.

Marked variation emerges in the quantity of questions individual RAs devote to each content category. Except for the “other preferences” category, 22 providers posed at least one item in every category. Among them, Mintos, Fidelity, ONVEST, Dr. Lux & Präuner, BW-Bank ON and growney recorded the highest overall question counts, albeit with differing emphases: Fidelity concentrated most heavily on risk-attitude items; Mintos and ONVEST prioritised assessments of investor knowledge, whereas Dr. Lux & Präuner placed greatest weight on clients’ financial circumstances. Table 4 groups the RAs into four frequency intervals according to the number of questions asked per category.

Table 4. Robo-Advisors by Question Count Intervals per Category

Question category	Questions-per-category intervals				Interval ranges for the number of questions per category
	1	2	3	4	
Objectives and purpose	10	21	12	2	Interval 1 = [0–0], Interval 2 = [1–1], Interval 3 = [2–2], Interval 4 = [3–3]
Investment horizon and age	4	28	12	1	Interval 1 = [0–0], Interval 2 = [1–1], Interval 3 = [2–2], Interval 4 = [3–3]
Amount and frequency of deposits	7	18	15	5	Interval 1 = [0–0], Interval 2 = [1–1], Interval 3 = [2–2], Interval 4 = [3–3]
Risk preferences	19	12	7	7	Interval 1 = [0–1], Interval 2 = [2–2], Interval 3 = [3–4], Interval 4 = [5–16]
Sustainability preferences (ESG)	9	12	11	13	Interval 1 = [0–0], Interval 2 = [1–1], Interval 3 = [2–2], Interval 4 = [3–7]
Other preferences	38	6	1	0	Interval 1 = [0–0], Interval 2 = [1–1], Interval 3 = [2–2], Interval 4 = $\emptyset$
Knowledge and experience	13	10	15	7	Interval 1 = [0–1], Interval 2 = [2–2], Interval 3 = [3–4], Interval 4 = [5–22]
Financial situation	9	15	11	10	Interval 1 = [0–1], Interval 2 = [2–3], Interval 3 = [4–5], Interval 4 = [6–12]

Note. Author's elaboration; Data are based on 45 robo-advisors (RAs) operating in Germany in July 2024. The table reports the number of RAs falling into each interval (1–4) of questions per category. The interval ranges are defined in the last column.

No RA consistently fell into the two upper intervals (3 and 4). Hence, there is no “all-round leader” in terms of question depth, even after excluding the sparsely used “other preferences” category. Conversely, no RA appeared solely in the lowest interval once that category is ignored. Financery, Peaks, ginmon, and Openbank occupied only the first or second interval, indicating the leanest questionnaires overall. At the opposite end, ONVEST featured in interval 4 for four categories and in interval 3 for one further category. Growney followed, reaching band 4 in three categories and interval 3 in two. Fidelity, Mintos and BW-Bank ON each attained interval 4 twice; Fidelity sat in interval 3 on four occasions, and Mintos and BW-Bank ON on three. Every remaining provider appeared in the top two intervals no more than four times in total.

It is worth noting that 22.22% (10/45) of RAs do not ask any questions in the first category regarding objectives, whereas 77.78% (35/45) collect data for the profile in this category, of which 28.89% (13/45) even asked two or three questions. 91.11% (41/45) of RAs ask questions about the investment horizon and age in the second category. Most of these (62.22%, or 28/45) asked only one question about either the horizon or, in the case of three RAs, the investor's age instead. The amount and frequency of deposits are recorded by 84.44% (38/45) RAs, 44.44% (20/45) two or three of which asked questions about the matter.

It is remarkable that 42.22% (19/45) of RAs only ask a maximum of one question about the risk preferences, including risk affinity and risk tolerance, which is surprising given the importance of this content category. Evergreen was the only RA that did not ask about risk matters at all. Nevertheless, they make portfolio proposals with allocations to assets with different risk weights. This fact alone shows that there is a massive weakness in the risk assessment. 80% (36/45) of RAs ask about sustainability preferences.

While 14 (31.11%) RAs asked three or more questions in the risk category, 13 (28.89%) RAs with three or more questions were interested in investors' sustainability preferences. This fact illustrates that investor preferences regarding sustainability and investment risk are apparently of almost equal importance to RAs. 91.11% (41/45) of RAs surveyed investors on their knowledge and experience of financial markets, with almost half asking three or more questions. In addition, 88.89% (40/45) of RAs took the investor's personal financial situation into account when allocating portfolios, with almost half enquiring about four or more aspects.

3.2.2. Portfolio Recommendations and Strategies of Robo-Advisors

Upon completion of the questionnaire during the profiling phase by investors, an RA generates a portfolio recommendation. Table 5 presents some descriptive results regarding the distribution of portfolios per RA. The numerous runs through the profiling processes for each of the 45 RAs delivered a total of 405 distinct portfolios, giving an average of 9 portfolios per RA (standard deviation = 7.55), with the smallest RA comprising three portfolios and the largest 37. The high variance in the number of portfolios shows that the granularity of the profiling process varies greatly between RAs and that the individualisation of

portfolio recommendations ranges from very low to a high variety of portfolios created depending on the investor profile. 77.78% (35/45) of RAs offered 10 portfolios or fewer. The distribution of portfolio counts was segmented into quartiles. The most common number of portfolios was five, which applied to 12 RAs. RAs in the top quartile (10 out of 45) together accounted for 204 portfolios, representing 50.37% of the full sample. The RAs with the greatest breadth of portfolios were: Scalable Capital (37), Liquid (31), Whitebox (27), quiron (20), ginmon (20), Robin (16), VisualVest (14), easyInvesto (14), INNO INVEST (14), and Gerd Kommer Invest (11).

Table 5. Descriptive Statistics of Portfolios per Quartile

Quartile	Portfolios	#RAs	Sum	Mean	SD	Min	Max	Range
1st Quartile (Lower 25%)	≤ 4	11	40	3.64	0.50	3	4	1
2nd Quartile (25%–50%)	= 5	12	60	5.00	0.00	5	5	0
3rd Quartile (50%–75%)	> 5 and ≤ 9	12	101	8.42	1.73	6	10	4
4th Quartile (Upper 25%)	> 10	10	204	20.40	8.58	11	37	26
Total	≥ 3	45	405	9.00	7.55	3	37	33

Note. Author's elaboration: Data are based on 45 robo-advisors (RAs) operating in Germany in July 2024. Quartiles are defined by the number of model portfolios offered per provider: 1st quartile (≤ 4 portfolios), 2nd quartile (= 5), 3rd quartile (> 5 and ≤ 9), and 4th quartile (> 10). Sum = total number of portfolios per quartile; Mean = average number of portfolios per RA in that quartile; SD = sample standard deviation; Min/Max = observed minimum and maximum; Range = Max–Min. The Total row shows summary statistics across all RAs (≥ 3 portfolios in total).

The portfolios proposed by the RA platforms were typically constructed from a mix of equity and fixed-income segments, almost always implemented through passive equity and bond ETFs. Every RA permitted allocations to shares and/or bonds, which formed the core building blocks of their portfolios. A minority diversified further: seven firms (15.56%) incorporated property exposure, commonly via real estate ETFs, and an identical number added other alternative assets. Cryptocurrencies featured in the line-up of only four providers (8.89%). Commodity holdings were available at 18 RAs (40%), while 21 RAs (46.66%) enabled clients to select sustainability-orientated (ESG) options, yielding 147

sustainable portfolios containing ESG instruments in total, an average of just over three per RA.

The composition analysis revealed standard instruments based on ETFs from equities (indices) and bonds (indices), which in turn were divided into developed and emerging markets. In addition, the associated ESG instruments were available at the RAs that had integrated ESG into their instrument pool. It was evident that a divergence of preferences was exhibited by sector equity ETFs, encompassing technology, healthcare, financials, energy and industrials, along with other sectors. Furthermore, the investment universe encompassed tangible assets, notably through real estate ETFs, along with commodities, which were principally constituted by precious metals, with gold being the predominant constituent. The use of alternative instruments, such as cryptocurrencies, was also observed. Derived from the identified asset classes and instruments, the following standard investment universe is utilised by RAs:

Stocks Global ETFs, Stocks Global ESG ETFs, Stocks Developed Markets (DM) ETFs, Stocks DM ESG ETFs, Stocks Emerging Markets (EM) ETFs, Stocks EM ESG ETFs, Bonds Global ETFs, Bonds Global ESG ETFs, Bonds DM ETF, Bonds DM ESG ETFs, Bonds EM ETFs, Bonds EM ESG ETFs, Real Estate ETFs, Real Estate ESG ETFs, Stocks Sector ETFs (e.g., Technology, Healthcare, Financials, Energy, etc.), Precious Metals, and Cryptocurrencies.

All recognised assets from the composition analyses of the RA portfolios can essentially be assigned to one of these asset (sub)classes and asset groups. The portfolio analysis yielded 11 different equity groups with assigned portfolios. The portfolio groups differ in their equity weighting by increments of 10%, beginning with a group that had 0% equities, followed by groups of portfolios with 20%, 30%, ... up to 100% equity exposure. Hence, equity weights spanned the full spectrum from 0% to 100%. On average, portfolios with a 50% equity allocation were most prevalent (mean = 1.33 per RA), whereas fully bond-based portfolios (0% equities) were rare (mean = 0.22). Many RAs supplied multiple portfolio compositions within the same equity group. Allocations of 50% and 70% equities appeared at 32 of the 45 RAs; 30% at 31 RAs; 20% and 80% at 26 RAs each; and 40% at 23 RAs, indicating strong clustering around these risk levels. Table 6 shows the descriptive statistics of the number of portfolios within the equity exposure groups.

Table 6. Descriptive Statistics of Portfolios by Target Equity Exposure

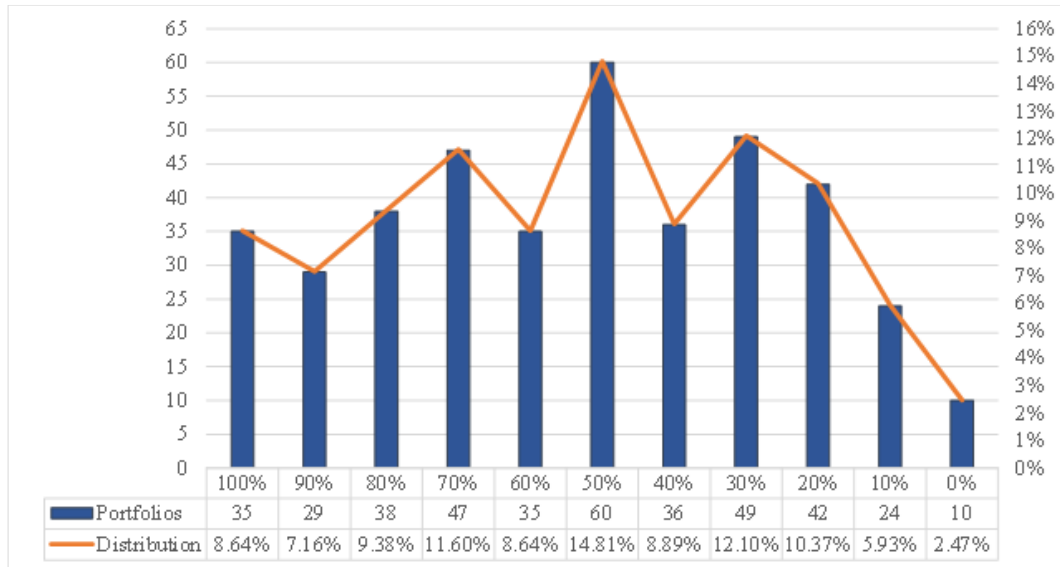
Target equity exposure	Mean	SD	Min	Max	Range	Sum
100%	0.7778	1.2227	0	6	6	35
90%	0.6444	0.8300	0	3	3	29
80%	0.8444	0.9524	0	4	4	38
70%	1.0444	0.8779	0	3	3	47
60%	0.7778	0.9508	0	4	4	35
50%	1.3333	1.5667	0	8	8	60
40%	0.8000	0.9909	0	4	4	36
30%	1.0889	0.9960	0	4	4	49
20%	0.9333	0.9630	0	3	3	42
10%	0.5333	0.8421	0	3	3	24
0%	0.2222	0.5988	0	3	3	10
Total	9.0000	7.5528	3	37	34	405
Of which ESG portfolios	3.2667	4.7213	0	24	24	147

Note. Author's elaboration: Data are based on 45 robo-advisors (RAs) operating in Germany in July 2024. Target equity exposure refers to the intended percentage of equities in the portfolio (0%–100%). Mean = arithmetic mean number of portfolios across RAs at each exposure level; SD = sample standard deviation; Min/Max = observed minimum and maximum; Range = Max – Min; Sum = total number of portfolios across all RAs. "Of which ESG portfolios" reports the subset explicitly labelled as environmental, social, and governance (ESG) portfolios. Values of Mean and SD are rounded to four decimals.

As illustrated in Figure 3, the 405 portfolios recommended by the 45 RAs are distributed across various equity exposure brackets. The largest group contained portfolios with a 50% equity exposure (60 of 405). In descending order, the next most common portfolios were those holding 30% equities (49 of 405), 70% (47), 20% (42), 80% (38), 40% (36), 100% (35), 60% (35), 90% (29), 10% (24) and, finally, 0% (10). The average equity exposure was approximately 53.16%. The distribution was almost symmetrical with a skewness of 0.042, tending towards higher equity allocations, and with a kurtosis of -1.007 , it was platykurtic (flat). 60.24% (244/405) of the portfolios contained an equity allocation of 50% or more. Approximately 36.79% of all RA portfolios analysed had high target equity exposures of 70% and above.

Equity allocations of 50%, 30% and 70% represented balanced, defensive, and more aggressive strategies from a risk perspective, respectively. For RAs that offered only three portfolios, such as easyfolio, these three allocations were exactly the equity percentages represented.

Figure 3. Distribution of Portfolios by Target Equity Exposure



Note. Author's elaboration: This diagram shows the frequency distribution of the equity exposure level of 405 proposed portfolios from 45 robo-advisors operating in Germany in July 2024.

The analysed RAs employed a diverse set of portfolio management strategies, ranging from purely quantitative to hybrid human-machine approaches, often combining different methods like Value-at-Risk models or the Black–Litterman framework to refine risk–return trade-offs. Table 7 shows the strategies mentioned by RAs in their management methods.

In two-thirds (30/45) of cases, the Modern Portfolio Theory (MPT) (mean–variance optimisation) was utilised as the underlying principle for strategic asset allocation. Furthermore, 46.47% (21/45) of RAs employed impact investing (ESG) as a means of creating sustainable portfolios. It was found that one-third (15/45) of RAs employed an active portfolio management approach, incorporating hedging techniques, while 28.89% (13/45) explicitly mentioned buy-and-hold strategies. A hybrid approach combining human oversight and management with algorithmic decision-making was adopted by 24.44% (11/45) of the RAs. Furthermore, 11.11% (5/45) of RAs implemented artificial intelligence investing (AI-based trading bots) for dynamic rebalancing or tactical allocation. An equal number of RAs applied value investing principles to portfolio composition. A total of 8.89% (4/45) of the RAs incorporated value-at-risk (VaR) models into their portfolio management strategies. 6.67% of RAs utilised Black–Litterman models (3/45), factor investments

(smart beta strategies) (3/45), or core–satellite instruments (3/45). One RA (2.22%) mentioned risk parity as a deployed portfolio strategy. The most common combinations consisted of RAs that used both MPT and impact investing (28.89%), MPT and buy-and-hold (20%), MPT and active management (15.56%), active and hybrid management (13.33%), or active and impact investing (11.11%), and MPT and hybrid (11.11%).

Table 7. Portfolio Strategies Disclosed by Robo-Advisors

Strategies	Robo-advisor
Modern portfolio theory (mean–variance optimisation)	Whitebox, quirion, ginmon, VisualVest, easyInvesto, INNO INVEST, Gerd Kommer Invest, fintego, growney, bevestor, Evergreen, Smavesto, Fidelity, investify, Warburg Navigator, cominvest, Oskar, Openbank, Financery, ONVEST, Revolut, BW-Bank ON, Raisin Invest, Peaks, GLS onlineInvest, Mintos, easyfolio, peningar, Digital Invest Assets, MONVISO
Impact investing (ESG)	Scalable Capital, Liquid, Whitebox, quirion, ginmon, VisualVest, easyInvesto, INNO INVEST, fintego, Estably, growney, bevestor, Minveo, Smavesto, Triodos Impact Portfolio Manager, cominvest, Oskar, my-si, inyova, vividam, GLS onlineInvest
Active trading	Liquid, INNO INVEST, bevestor, Evergreen, Minveo, Warburg Navigator, digiVestor, LAIC ADVISOR, Solidvest, truevest, Zeedin, vividam, peningar, Digital Invest Assets, MONVISO
Buy-and-hold	quirion, easyInvesto, INNO INVEST, Gerd Kommer Invest, growney, Triodos Impact Portfolio Manager, my-si, inyova, Dr. Lux & Präuner Vermögensverwaltung, Raisin Invest, Peaks, Mintos, easyfolio
Hybrid (human + algorithmic decisions)	Liquid, ROBIN, VisualVest, INNO INVEST, Estably, cominvest, BW-Bank ON, truevest, Zeedin, vividam, peningar
Artificial intelligence investing (AI-based trading bots)	VisualVest, INNO INVEST, Minveo, Smavesto, LAIC ADVISOR
Value investing	Gerd Kommer Invest, Estably, Triodos Impact Portfolio Manager, Solidvest, Dr. Lux & Präuner Vermögensverwaltung

Value-at-risk (VaR)	Scalable Capital, Whitebox, ROBIN, ONVEST
Black–Litterman	Scalable Capital, Liquid, Estably
Factor investing (smart beta)	ginmon, Gerd Kommer Invest, digiVestor
Core–satellite	Liquid, ROBIN, investify
Risk parity	my-si

Note. Author's elaboration: Strategies reported by 45 robo-advisors (RAs) operating in Germany in July 2024. Each cell lists the providers that explicitly disclose use of the corresponding strategy in their portfolio management methods. Abbreviations: ESG = environmental, social, and governance; AI = artificial intelligence; VaR = value at risk.

A total of 5 of the 45 RAs work with AI to manage their portfolios. The LAIC ADVISOR employs advanced artificial intelligence to deliver highly personalised, risk-optimised asset allocation (LAIC Vermögensverwaltung GmbH, 2025). The Digital Asset Platform 4.0 is a core component of the architecture, facilitating the processing of substantial datasets encompassing both structured and unstructured data. These datasets span a duration of 30 years, from capital market history to real-time ESG metrics. Machine learning models, in particular Bayesian Neural Networks (BNNs), facilitate probabilistic forecasting of market trends, quantification of uncertainty and support for dynamic portfolio adjustments (Mongwe et al., 2025; Syed et al., 2025; Ticknor, 2013). The RA's AI framework analyses over 125 million data points on a daily basis, generating forecasts for thousands of securities and constructing over one million unique portfolio constellations tailored to individual risk profiles. The RA also uses natural language processing (NLP) to handle large amounts of unstructured data such as text documents to identify patterns (Du et al., 2025; Xing et al., 2018). The RA platform of LAIC represents a significant advance on conventional RAs through its combination of AI-driven analytics and automated risk management, offering adaptive, ESG-compliant investment strategies for institutional investors and high-net-worth individuals (LAIC Vermögensverwaltung GmbH, 2025).

VisualVest incorporates AI into its investment process through “Malina” (Machine Learning for Investment Applications), an AI system developed by Union Investment (VisualVest GmbH, 2025). Malina employs statistical machine learning techniques to analyse large, diverse datasets, including exchange rate dynamics, inflation expectations, yield curve structures, and commodity price trends, integrating these with historical market patterns to generate probabilistic

forecasts. A key application of the model is to assess the likelihood of equity funds outperforming commodity funds within the VestFolios ETF portfolios over a six-month horizon. These forecasts inform dynamic portfolio re-weighting, enabling more responsive asset allocation. Notably, Malina's outputs are not implemented autonomously; they are subject to evaluation by an investment committee that incorporates broader fundamental analysis. This human-and-machine approach combines the scalability and pattern-recognition capabilities of AI with the contextual judgement of experienced portfolio managers, which is claimed by the RA to improve decision-making quality while maintaining strategic overview.

Smavesto employs a multi-layered AI architecture to deliver ETF-based, cost-efficient wealth management. Its system integrates three advanced models to optimise portfolio decisions (Smavesto GmbH, 2025). The Hidden Markov Model (HMM) identifies prevailing market trends by detecting patterns in historical price data, supporting precise timing of buy and sell actions (de Angelis & Paas, 2013; Nguyen & Nguyen, 2015; Roman et al., 2010). The Long Short-Term Memory (LSTM) network processes time-series data, trading volumes, and even financial news, retaining relevant information and discarding noise to forecast short- to medium-term performance (Bao et al., 2017; Fischer & Krauss, 2018). eXtreme Gradient Boosting (XGBoost) refines these forecasts by classifying trends and analysing technical indicators through ensembles of decision trees (Oukhouya et al., 2024; Wang & Guo, 2020). Following a major August 2024 upgrade, the AI can process larger datasets more rapidly, enhancing responsiveness during volatile periods (Smavesto GmbH, 2025). While the algorithms continuously monitor markets and generate recommendations, final execution rests with portfolio managers, ensuring a controlled synthesis of machine-driven analytics and human oversight.

Minveo applies a multi-layered AI architecture composed of several specialised models, each focused on distinct analytical tasks (Minveo AG, 2025). This design enables mutual monitoring between AIs and the detection of complex market interdependencies, enhancing forecast quality. Each of the > 900 portfolio positions is governed by a custom-built evaluation metric, continuously retrained to adapt to evolving market conditions, rather than applying uniform criteria. The system processes approximately two terabytes of market data daily, with the capacity to filter relevant signals from noise, particularly in volatile environments.

A self-learning neural network refines its decision-making process by reviewing and updating its heuristics daily, informed by past outcomes. Each position is assessed across a maximum of 23 variables, allowing Minveo to identify and decouple correlations, mitigate risk, and capture opportunities with high precision. Minveo's AI is complemented by human expertise with the aim of supporting a more disciplined, data-driven, and adaptive wealth management.

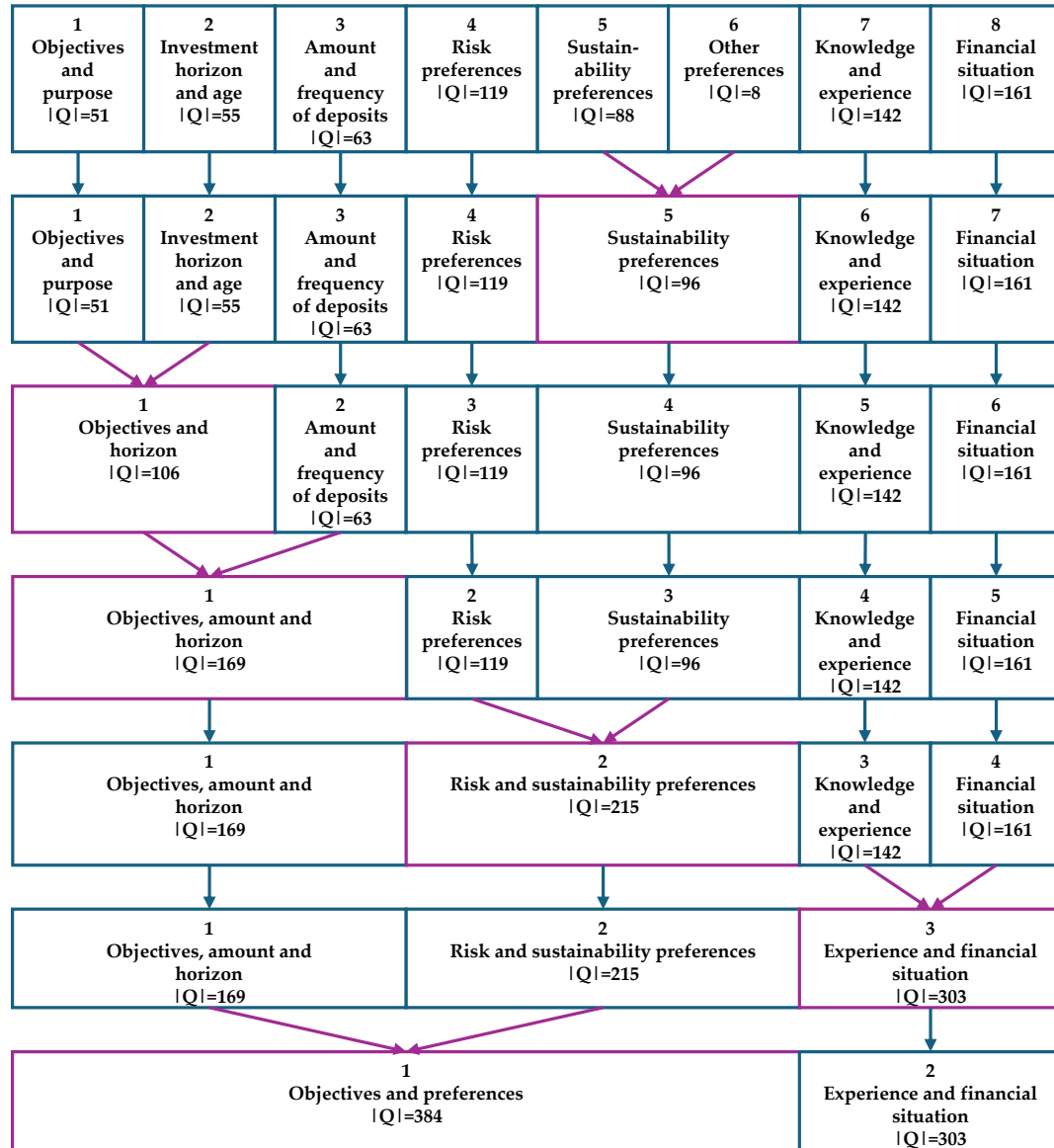
INNO INVEST operates a fully automated, AI-driven asset management system that integrates market analysis, signal generation, and trade execution in real time (Innovative Investment Solutions GmbH, 2025). At its core, a proprietary quantitative model integrates machine learning, neural networks and quantitative methods to analyse extensive volumes of market, sentiment and news data. The system employs a top-down approach, first evaluating macroeconomic factors such as growth, inflation, and monetary policy, then sector dynamics, and finally individual securities before issuing buy/sell signals within seconds. It is posited that three primary parameters (price, volume, and volatility) inform predictions of daily trading ranges. A risk range model, informed by behavioural psychology, integrates technical and sentiment analysis to optimise entry and exit points. It is customary for positions to be retained for a maximum period of 90 trading days. During volatile periods, the AI dynamically reallocates capital across various asset classes, including equities, bonds, and commodities. This approach is intended to minimise drawdowns while preserving long-term return potential. The role of the human quant specialist is to oversee research and continual refinement of the algorithms.

3.2.3. TF-IDF and Cluster Analysis of Robo-Advisors' Profiling Questionnaires

The number of documents corresponded to the number of questions, i.e. $|Q| = m = 687$, which were recorded in Subsection 3.2.1 and sorted into eight question categories. The vocabulary comprised $|V| = n = 1031$ terms. Therefore, $K_{max} = 8$ was set as the maximum number of categories. Given (18) and (19), the quadratic norm of the global mean weight vector was $\|\bar{x}\|^2 = 0.05799$ and total sum of squared errors was $TSSE = 2988.54$. According to (20), the sum of quadratic norm of the weight vectors was $\sum_{i=1}^m \|x_i\|^2 = 3028.38$. These key figures

represent K-invariant quantities and describe the total space. The categories were then gradually merged to reduce the number of categories. The procedure is shown in Figure 4.

Figure 4. Merging Procedure of Question Categories ($K = 8-2$)



Note. Author's elaboration: The diagram illustrates the stepwise merging of question categories by successively reducing the maximum number of categories from $K = 8$ to $K = 2$. The purple arrows represent the merging of the categories with the consolidation of all questions contained in the previous categories. |Q| = number of questions in the category. Data are based on 687 items from the profiling questionnaires of 45 robo-advisors operating in Germany in July 2024.

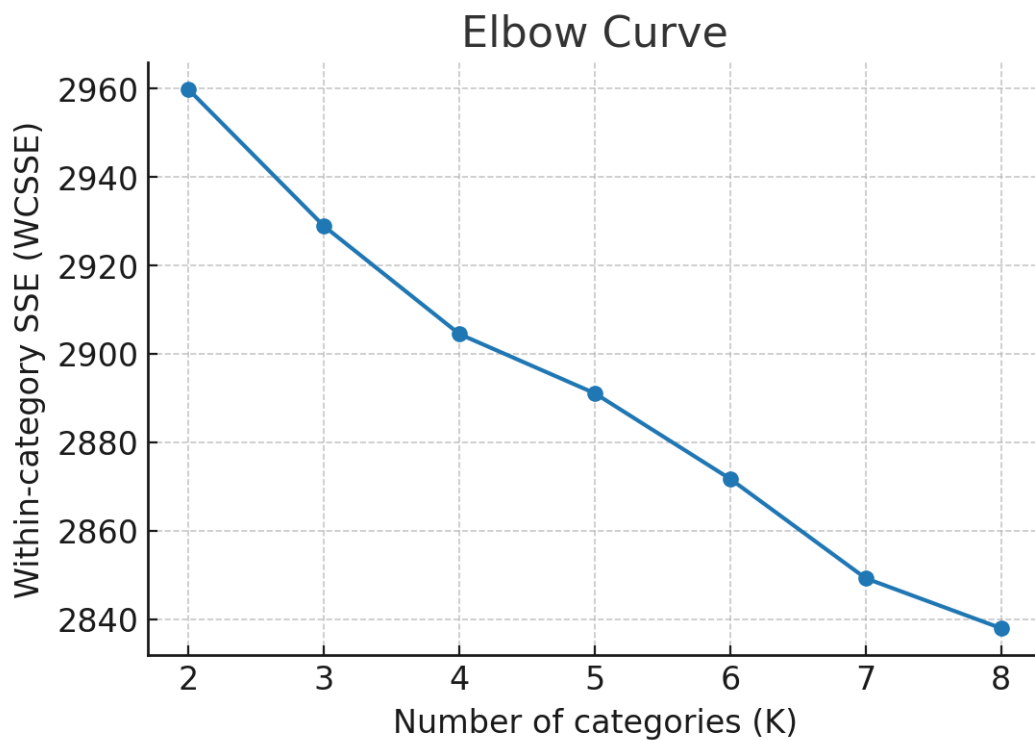
An initial reduction in the number of categories from $K = 8$ to $K = 7$ was achieved by merging the small category “other preferences” (comprising eight questions) into the “sustainability preferences” category, since the questions from the “other preferences” category also address investment topics (climate change, digitalisation, health) and certain industries and regions which are most closely related to investment preferences. Consequently, the category “sustainability preferences” generally covers investment preferences in a broader sense. A further reduction of the categories to $K = 6$ resulted in the merging of the categories “objectives and purpose” and “investment horizon and age” into a single category, termed “objectives and horizon”, covering both objectives and investment horizon.

A closer inspection of the categories revealed thematic overlaps and connections based on the questions. For instance, investment objectives also cover specific sums and the time until payout, such as pensions or house construction. The subsequent variant, comprising $K = 5$ categories, underwent further consolidation, with the categories “objectives and horizon” and “amount and frequency of deposits” being integrated into a larger category, labelled “objectives, amount and horizon”, which broadly encompasses all investment objectives, investment amounts and investment horizons. In a broader variant, comprising $K = 4$ categories, the categories for risk preferences and sustainability preferences were merged into a larger category for risk and sustainability preferences. The variant with $K = 3$ combines the categories “knowledge and experience” and “financial situation” into a new category termed “experience and financial situation”. The profiling analysis in Subsection 3.2.1 found that such questions were typically posed in sequence towards the conclusion of the process, with the objective of evaluating the investor’s experience and the general state of their financial situation.

The final variant with $K = 2$ categories merged the two categories “objectives, amount and horizon” and “risk and sustainability preferences” to form the composite category “objectives and preferences”. This category encompasses all questions related to objectives, amount, horizon and risk and investment preferences, i.e., questions that specifically concern the investment. The category “experience and financial situation” was retained as the second category, as it comprises all other questions that generally relate to the investor’s background in

terms of financial situation (e.g., income, household budget) and previous experience with capital markets and financial products. The plotted elbow curve is shown in Figure 5. Two turning points are visible between $K = 4$ and $K = 5$ between $K = 7$ and $K = 8$, which divide the graph into three sections: a convex section from $K = 2$ to between $K = 4$ and $K = 5$, followed by a concave section to the point between $K = 6$ and $K = 7$, and finally another convex section $K = 8$. This means that two “elbows” can be detected between $K = 4$ and $K = 5$ and likewise between $K = 7$ and $K = 8$. While the turning points are clearly recognisable due to the shape of the curve, the “elbows” are barely visible to the naked eye. On closer inspection, however, the first elbow in the convex area around $K = 4$ and a second elbow in the convex area around $K = 7$ are slightly recognisable. The numbers in the following Table 8 confirm the geometric recognition.

Figure 5. Elbow Curve of Within-Category SSE by Number of Categories (K)



Note. Author's elaboration: The elbow curve shows the relationship between the number of categories (K) and the within-category sum of squared errors (WCSSE) from the TF-IDF clustering algorithm applied to 687 profiling questions from 45 robo-advisors. Lower WCSSE indicates tighter clustering. The first elbow is marginally visible in the convex area at $K = 4$, and the second elbow in the convex area at $K = 7$. These can be identified by the steeper kinks at $K = 4$ and $K = 7$.

Table 8 shows the summary results of the scoring metrics WCSSE, BCSSE, CHI and the first and second differences of WCSSE for each of the categorisations $K = 2$ and $K = 8$. The WCSSE numbers confirmed the geometric interpretation of the elbow in the elbow curve from Figure 5. The central and backward difference of the second order of the WCSSE displayed positive local maxima between $K = 4$ and $K = 5$ as well as between $K = 7$ and $K = 8$. Between $K = 7$ and $K = 8$, the local maximum was slightly higher at 11.19 than between $K = 3$ and $K = 5$ at 11.02. A finer categorisation could therefore be $K = 7$ or $K = 8$, whereas categorisation with $K = 4$ or $K = 5$ could represent a coarser alternative. BCSEE is provided for informational purposes, increasing with K as TSSE is constant and WCSSE decreases.

The Calinski–Harabasz index (CHI) peaked at $K = 3$ with 6.9622 and then dropped with increasing K to 5.1490 at $K = 8$. According to the CHI, $K = 3$ offered the strongest ratio between BCSSE and WCSSE. For $K = 4$, $CHI = 6.5888$ was still close to the maximum, as the next best recommendation for a finer categorisation. From $K = 6$ to $K = 7$, there was another smaller peak with $CHI = 5.5426$. This finding coincides with the results of the elbow curve and WCSSE interpretation, shifted by $K-1$. Thus, depending on the key figure, two possible optimal categorisations can be identified between $K = 3$ and $K = 5$ (coarser categorisation) and between $K = 6$ and $K = 8$ (finer categorisation).

Table 8. Summary Metrics from TF-IDF Categorisation Variants ($K = 8-2$)

<i>K</i>	<i>WCSSE</i>	<i>BCSSE</i>	<i>CHI</i>	$\Delta WCSSE$	$\Delta^2 WCSSE$	$\delta^2 WCSSE$
2	2959.8505	28.6902	6.6398	-	-	-
3	2928.9157	59.6251	6.9622	-30.9349	-	6.5031
4	2904.4839	84.0568	6.5888	-24.4318	6.5031	11.0208
5	2891.0729	97.4679	5.7481	-13.4110	11.0208	-5.9699
6	2871.6919	116.8488	5.5420	-19.3809	-5.9699	-3.1109
7	2849.2001	139.3407	5.5426	-22.4919	-3.1109	11.1900
8	2837.8982	150.6425	5.1490	-11.3018	11.1900	-

Note. Author's elaboration: Results are based on TF-IDF clustering of 687 profiling questions from 45 robo-advisors in Germany as of July 2024. Metrics reported: WCSSE = within-category sum of squared errors; BCSSE = between-category SSE; CHI = Calinski–Harabasz index; $\Delta WCSSE$ = first-order (central) differences of WCSSE; $\Delta^2 WCSSE$ = second-order (central) differences; $\delta^2 WCSSE$ = second-order backward differences. Values are rounded to four decimal places.

Table 9 shows the results for the quadratic norm of the centroids $\|\mu_k\|^2$, the number of questions per category $|C_k|$, and the $WCSSE_k$ per category for the seven different categorisations $K \in \{2, \dots, 8\}$ at individual category level C_k . A high $\|\mu_k\|^2$ indicates that many questions share defining terms/N-grams, i.e. there is a clear thematic core, as this leads to low $WCSSE_k$ and thus high selectivity of the top terms. A low $\|\mu_k\|^2$ means that questions point in different directions, resulting in heterogeneous categories with higher $WCSSE_k$ and thus less distinct top term profiles.

Table 9. Centroid Norms, Cardinality, and WCSSE per Category ($K = 8-2$)

K	k	Category C_k	$\ \mu_k\ ^2$	$ C_k $	$WCSSE_k$
8	1	Objectives and purpose	0.4634	51	239.35
	2	Investment horizon and age	0.6834	55	249.95
	3	Amount and frequency of deposits	0.5126	63	343.61
	4	Risk preferences	0.1506	119	460.11
	5	Sustainability preferences	0.1425	88	253.31
	6	Other preferences	1.3737	8	60.48
	7	Knowledge and experience	0.1601	142	530.02
	8	Financial situation	0.1896	161	700.54
7	1	Objectives and purpose	0.4634	51	239.35
	2	Investment horizon and age	0.6834	55	249.95
	3	Amount and frequency of deposits	0.5119	63	343.41
	4	Risk preferences	0.1502	119	459.88
	5	Sustainability preferences	0.1409	96	326.04
	6	Knowledge and experience	0.1601	142	530.02
	7	Financial situation	0.1896	161	700.54
6	1	Objectives and horizon	0.3654	106	511.79
	2	Amount and frequency of deposits	0.5119	63	343.41
	3	Risk preferences	0.1502	119	459.88
	4	Sustainability preferences	0.1409	96	326.04
	5	Knowledge and experience	0.1601	142	530.02
	6	Financial situation	0.1896	161	700.54
5	1	Objectives, amount and horizon	0.3065	169	874.59
	2	Risk preferences	0.1502	119	459.88
	3	Sustainability preferences	0.1409	96	326.04
	4	Knowledge and experience	0.1601	142	530.02
	5	Financial situation	0.1896	161	700.54

4	1	Objectives, amount and horizon	0.3065	169	874.59
	2	Risk and sustainability preferences	0.0839	215	799.33
	3	Knowledge and experience	0.1601	142	530.02
	4	Financial situation	0.1896	161	700.54
3	1	Objectives, amount and horizon	0.3065	169	874.59
	2	Risk and sustainability preferences	0.0839	215	799.33
	3	Experience and financial situation	0.0958	303	1255.00
2	1	Objectives and preferences	0.1018	384	1704.85
	2	Experience and financial situation	0.0958	303	1255.00

Note. Author's elaboration: Results from TF-IDF clustering of 687 profiling questions from 45 robo-advisors, showing metrics per category across solutions with $K=8$ to $K=2$. $\|\mu_k\|^2$ = quadratic norm of the centroid vector, $|C_k|$ = number of questions assigned to category C_k , $WCSSE_k$ = within-category sum of squared errors. A higher $\|\mu_k\|^2$ indicates a more homogeneous category with clearer thematic core (lower $WCSSE_k$, whereas a lower $\|\mu_k\|^2$ indicates greater heterogeneity and less distinct term profiles. Values are rounded to four decimals in case of $\|\mu_k\|^2$ and to two decimals in case of $WCSSE_k$.

It became apparent that in the categorisation with $K=8$, the categories relating to experience and financial situation exhibited a relatively high degree of heterogeneity. In terms of knowledge and experience, for example, this heterogeneity was due to the fact that numerous different professional situations, industry experience, knowledge of capital markets, and even the specific number of transactions carried out per financial instrument were surveyed. With regard to the financial situation, many different questions were asked about income, expenses, reserves (both illiquid and liquid assets), liabilities, and so forth, covering various areas of the household budget. There was also a high degree of heterogeneity in risk preferences, as the questions on risk preferences were posed using different approaches such as loss aversion, risk–return matrices, emotional–psychological items, and scenario models, which were incorporated into the questionnaires using different linguistic terms. Other categories relating to investment objectives and horizons were significantly closer together. As the number of categories was reduced, the WCSSE per category and overall increased. A lower number of categories ($K < 5$) led to more heterogeneous, coarser clusters with increasingly diverse questions.

The typical questions q_i were those that had the smallest distances $\text{dist}(i, k)$ to the centroid μ_k in the respective category C_k , defined by (15). Table 10 shows the results of the top 5 questions in each category with $K = 8$ categories. Compared to

the typical questions per category in Table 2 from Subsection 3.2.1, it is evident that the content of the questions for each category was covered. Thus, the TF-IDF results for eight categories essentially underlined the selection of representative questions identified in Table 2. Since the number of questions per category was high and Table 10 only shows the top 5 questions according to the minimum distance to the category centroid, some questions only appeared later in the list after the top 5 and are therefore not listed in Table 10.

Table 10. Top Five Typical Questions per Category ($K = 8$)

Category	Questions
Objectives and purpose	What is your investment goal and how long do you want to invest your money? How long and for what do you want to invest your money? What do you want to invest money for? Do you want to achieve a savings goal with this investment? How high is your savings goal?
Investment horizon and age	How long do you want to invest your money? How long and for what do you want to invest your money? How long is your investment horizon? What is your investment horizon for your investment? How many years do you want to invest your money for?
Amount and frequency of deposits	How and in what amount would you like to start your sustainable investment? (one-time investment, monthly savings plan) Select the investment strategy and the investment amount for your new portfolio. How much money do you want to invest and/or save? Your investment amount? How much money would you like to invest each month?
Risk preferences	Capital markets are prone to fluctuation. What loss of value makes you nervous? With your investment amount of €10,000 over your investment period, there may be a loss of €652. Can you bear this possible capital loss? Do you find large price fluctuations with possible interim losses in the context of this investment make you nervous?

	Of course you are investing this money to make a profit. But what potential loss could you bear if the portfolio were to lose value? I would like to achieve higher returns and am willing to accept risks to do so.
Sustainability preferences	Please indicate whether environmental, social and good corporate governance issues should be taken into account in recommendations and investment decisions (so-called sustainability preferences). Do you want to invest in environmentally sustainable economic activities that contribute significantly to achieving environmental goals under the Taxonomy Regulation? Should your portfolio include investments in environmentally sustainable economic activities that contribute significantly to achieving environmental goals? Would you like to invest in environmentally sustainable economic activities that contribute significantly to achieving environmental goals? Would you like to specify a lower limit for investing in products that promote the further development of environmental and/or social issues through their investment strategy?
Other preferences	Would you like to add 5% of any of the following investment topics? (5% climate change, digitalization, health) Investment preferences? Do you want to favor regions in your investment? At which bank do you want your portfolio? Do you want to favor certain industries?
Knowledge and experience	Do you have knowledge and/or experience with foreign currency transactions (e.g. buying stocks/bonds in foreign currency)? Which of the following asset classes do you already have experience with? (Query on various investment instruments such as funds, bonds, stocks, etc.) Do you have knowledge and experience of at least 3 years in the following asset classes? In the knowledge section, experience in years, the number of trades and the level of experience with 7 financial instruments are queried.

	Have you gained relevant capital market experience through training, studies or work?
Financial situation	What is the amount of your current assets (assets less liabilities/financial obligations) in EUR? (cash assets, capital investments, real estate assets) How much of your monthly net income is freely available after deducting your monthly expenses (e.g. living expenses, rent, loan interest, amortization, etc.)? Free monthly liquidity (monthly income minus monthly obligations)? How long will your reserves last to cover your current expenses and obligations? We recommend keeping at least two months' salary for unforeseen expenses. How long can you live on your reserves beyond that?

Note. Author's elaboration: The table presents the five questions per category for the categorisation with $K = 8$ categories, with the smallest distance $\text{dist}(i, k)^2$ to the centroid μ_k of the respective category C_k (see Equation (15)). Results are based on TF-IDF clustering of 687 profiling questions from 45 robo-advisors. Although only the top five questions are shown, additional questions with greater distances were also assigned to each category. The results confirm that the eight TF-IDF categories align closely with the representative questions identified earlier (see Table 2).

For example, questions about age in the category “investment horizon and age” only appeared from positions > 10 , sorted by distance. This can be explained as follows: the TF-IDF algorithm groups highly similar questions together, thereby forming subclusters. Numerous similar questions about the investment horizon lead to dominance within one category and thus shift the weighting centroid in their direction, so that the subclusters relating to age, which occur in smaller numbers, lie slightly further away from the centre of the cluster.

Much of the question content overlapped, so that the distances to the category centroids, $\text{dist}(i, k)$, differed only minimally. To obtain different questions per category, it would be necessary not only to sort the top typical questions according to $\text{dist}(i, k)$, but also to select those that differ from the preceding question by a minimum distance. This would prevent the inclusion of questions with identical content but slightly different wording multiple times. For example, in the category “investment horizon and age”, fewer questions about the investment horizon would be included, as this question content appeared multiple times, but instead a question about age and retirement plans would also be added. This, in turn, led to

the results already achieved by the content analysis in Table 2, which is why they are not listed again here to avoid duplication. The previously categorised questions were therefore derived in a different way using the quantitative text-analytics method of TF-IDF.

Table 11 displays the results for $K = 4$. As already derived from Table 9, the areas of risk and sustainability clustered together in a combined category, with sustainability issues dominating the centre of the cluster. The numerous and important risk issues lie slightly further away from the category centroid. The other categories still covered the content clearly, as each of the topics was already close to the category centroid in the top five questions. A further reduction led to topic-related inaccuracies in categories. APPENDIX 4 shows the results for $K = 5$, where the distinction between the risk and sustainability categories still resulted in the relevant questions that these categories represented.

Table 11. Top Five Typical Questions per Category ($K = 4$)

Category	Questions
Objectives, amount and horizon	What is your investment goal and how long do you want to invest your money? How long do you want to invest your money to accumulate or optimize your wealth? How long and for what do you want to invest your money? How long do you want to invest your money? How and in what amount would you like to start your sustainable investment? (one-time investment, monthly savings plan)
Risk and sustainability preferences	Please indicate whether environmental, social and good corporate governance issues should be taken into account in recommendations and investment decisions (so-called sustainability preferences). Do you want to invest in environmentally sustainable economic activities that contribute significantly to achieving environmental goals under the Taxonomy Regulation? Would you like to specify a lower limit for investing in products that promote the further development of environmental and/or social issues through their investment strategy?

	Should your portfolio include investments in environmentally sustainable economic activities that contribute significantly to achieving environmental goals? Would you like to invest in environmentally sustainable economic activities that contribute significantly to achieving environmental goals?
Knowledge and experience	Do you have knowledge and/or experience with foreign currency transactions (e.g. buying stocks/bonds in foreign currency)? Which of the following asset classes do you already have experience with? (Query on various investment instruments such as funds, bonds, stocks, etc.) Do you have knowledge and experience of at least 3 years in the following asset classes? In the knowledge section, experience in years, the number of trades and the level of experience with 7 financial instruments are queried. Have you gained relevant capital market experience through training, studies or work?
Financial situation	What is the amount of your current assets (assets less liabilities/financial obligations) in EUR? (cash assets, capital investments, real estate assets) How much of your monthly net income is freely available after deducting your monthly expenses (e.g. living expenses, rent, loan interest, amortization, etc.)? Free monthly liquidity (monthly income minus monthly obligations)? How long will your reserves last to cover your current expenses and obligations? We recommend keeping at least two months' salary for unforeseen expenses. How long can you live on your reserves beyond that?

Note. Author's elaboration: The table presents the five questions per category for the categorisation with $K = 4$ categories, with the smallest distance $\text{dist}(i, k)^2$ to the centroid μ_k of the respective category C_k (see Equation (15)). Results are based on TF-IDF clustering of 687 profiling questions from 45 robo-advisors. Four categories were defined: objectives/amount/horizon, risk and sustainability preferences, knowledge and experience, and financial situation. Compared with the eight-category solution in Table 10, the risk and sustainability dimensions merged into a combined cluster, with sustainability issues dominating the centroid. This reduction illustrates that further lowering K can lead to loss of topical precision.

3.2.4. Mathematical Model of the Robo-Advisory Process

Initially, the modelling of the profiling subprocess is presented. To do this, the questionnaire mapping (“profiling”) is defined, which consists of a response space, profile space and a profiling operator. Let the onboarding questionnaire return a vector

$$Q = (q_1, \dots, q_n) \in \mathcal{S} = \prod_{i=1}^n \mathcal{S}_i, \quad (30)$$

where each component q_i may be categorical (e.g. experience), ordinal (e.g. risk affinity on a 1–5 scale) or metric (e.g. monthly net income). The type and number of questions with the selection options may vary depending on the RA and can be specified according to which investor data is to be collected during onboarding.

An RA could, for example, use the 34-item questionnaire from Table 2 developed in Subsection 3.2.1 for profiling. Since the 34 questions are assigned to the eight content categories, the onboarding questionnaire Q can be partitioned into categories $Q = Q_1 \dot{\cup} \dots \dot{\cup} Q_8$ in the same way. These then represent the categories of the questionnaire ($\mathcal{C}_j = Q_j$). This provides partial profiles $\mathcal{S}_j$. Categories include, for example, “risk preferences” or “financial situation”, as defined in Table 2. Through partitioning, RAs can categorise the profile immediately at the outset to later perform profile analyses within each category. This also simplifies the adjustment of the vector if, as part of a profile update, only the part of a Q_j in a category needs to be adjusted. Changes such as goals, risk preferences or financial situation would be directly visible in the partitioning and can be continuously monitored and analysed. This is explained further below by introducing dynamics.

For the profile space define a low-dimensional feature space

$$\Omega \subset \mathbb{R}^d, \quad \omega = (\rho, \tau, \lambda, \varepsilon, \dots), \quad (31)$$

whose coordinates capture the investor characteristics risk affinity ρ , investment horizon $\tau \in \mathbb{R}_{>0}$, liquidity preference λ , sustainability preference ε , and so on.

This profile space can also be partitioned as a set, thereby enabling the creation of low-dimensional feature spaces. This is particularly helpful when measuring the impact of a category with questions on the portfolio. Subspaces can significantly simplify simulations and calculations, as they reduce higher-dimensional problems to smaller, low-dimensional problems. This means that only

one or a few categories are considered, e.g. again, if changes to the profile vector relate to only one or a few questions. The partitioning implies that the Ω_i represent lower-dimensional feature spaces from $\mathbb{R}^{d_j}$, where $d_j \leq d$.

Then the profiling operator is defined by the deterministic (or probabilistic) mapping

$$\varphi: \mathcal{S} \rightarrow \Omega, \quad \varphi(Q) = \omega, \quad (32)$$

that translates raw answers from equation (30) into a point in profile space from equation (31). The profiling operator $\varphi|_{\mathcal{S}_j}$ can also be restricted to the subspace. Formally φ can be defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Stochastic uncertainty may enter through noisy self-assessment or randomised follow-up questions. This may occur, for instance, if the investor revises his or her preferences over time, thereby changing the answers to the investor's questions and, consequently, the profile in the space. In this case, the time indexing can be supplemented to obtain a profile space-time $\omega_t = (\omega, t)$. This profile space-time generalisation is carried out below, allowing a static profile ω to be assumed at this point for an initial understanding of the remaining modelling steps.

Next, the portfolio recommendation mapping (initial asset allocation mapping) is defined. Let

$$Z_t = (\zeta_t^1, \dots, \zeta_t^m) \in \mathbb{R}^m, \quad (33)$$

be a market state as a stochastic factor vector collecting macro, fundamental and technical indicators. The process $(Z_t)_{t \geq 0}$ is modelled, for example, as a Markov diffusion or regime-switching chain. With a static RA that assigns profiles directly to predefined portfolios regardless of market conditions, Z_t can be assumed to be deterministic (constant) assumed to be constant and omitted in the following model. Bourgeron et al. (2018) propose a basic ordinal variable definition as a rating scale with seven different grades (from strong bearish to strong bullish) to translate the market state into bets by portfolio managers, using the Black–Litterman model for transforming these grades into expected returns.

If $A = \{\alpha^1, \dots, \alpha^k\}$ is the investment universe of the RA containing k investable assets (financial instruments), then a portfolio is a weight vector

$$w_t \in \Delta^{k-1} = \left\{ w \in \mathbb{R}_{\geq 0}^k : \sum_{i=1}^k w^i = 1 \right\} \quad (34)$$

in the portfolio simplex Δ^{k-1} , where w^i is the weight corresponding to the asset α^i . The sum condition ensures that the weights of the portfolios together add up to 100%. The allocation operator

$$\psi: \Omega \times \mathbb{R}^m \rightarrow \Delta^{k-1}, \quad w_t = \psi(\omega, Z_t), \quad (35)$$

maps the investor profile from (31) and prevailing market state Z_t from (33) to a recommended portfolio w_t as defined in (34). This is the target allocation. Algorithms frequently instantiate ψ in (35) via mean–variance optimisation, Black–Litterman, risk-parity, or machine-learning surrogates (Beketov et al., 2018; Bianchi & Briere, 2021; Dai et al., 2021; Day et al., 2018; Kim, 2019; Rossi & Utkus, 2020).

The essential RA process steps are now derived in a narrower sense. However, the model allows for a intuitive generalisation for dynamic rebalancing and adaptive profiling, which will follow now to capture further process steps for the ongoing management of the portfolio. Let B_t denote observable investor actions (deposit, withdrawal, questionnaire revision). The stochastic investor profile process evolves as

$$\omega_t = \varphi(Q_t), \quad Q_t = u(Q_{t-}, B_t), \quad (36)$$

where u amends previous answers in line with new information. Q_{t-} indicates the left-hand limit (or “pre-jump” value) at the stated time, thus it stands for the value vector Q for the questionnaire before the event (update of questionnaires response by the investor) occurring at time t . In a purely discrete-time setting it can be replaced by Q_{t-1} for $t > 1$. For simplicity, a discrete grid for the rebalancing function is used. On a grid $G = \{t_i \in \mathbb{R}_{\geq 0} | 0 = t_0 < t_1 < t_2 < \dots\}$, the RA platform updates the portfolio by

$$w_{t_i} = \arg \min_{w_t \in \Delta^{k-1}} \left\{ \delta(w_t, \psi(\omega_{t_i}, Z_{t_i})) + \kappa \xi(w_t, w_{t_{i-1}}) \right\}, \quad (37)$$

where δ measures tracking error to the target allocation (deviation from the target portfolio), $\kappa \geq 0$ reflects the turnover tolerance and ξ captures transaction cost. The scalar κ governs the trade-off between fidelity to the target and cost of adjustment. The grid is dynamically defined by the events (e.g., changes in the market or investor requirements), as well as additional fixed default rebalancing re-evaluation of the RA. By making the time-index explicit on w_t , it becomes a sequence of control decisions in a discrete-time stochastic control problem. At each

time event time t_i , the system state $(\omega_{t_i}, Z_{t_i}, w_{t_{i-1}})$ is observed and the control w_{t_i} is chosen to minimise a one-step cost consisting of tracking error plus transaction cost. Thus, the pair (ω_t, w_t) constitutes the state of a controlled process, while the choice of weights is the control. Optimal robo-advisory can thus be posed as a discrete-time stochastic control problem or, under continuous trading, as a Hamilton–Jacobi–Bellman (HJB)-type optimisation subject to investor-specific utility (Cui et al., 2022; Khosravi, 2024; Monoyios, 2010, 2018; Tahvildari, 2021).

Articulating the RA process in a formal mathematical language yields tangible advantages for implementation and for research that extend well beyond a merely descriptive specification. The explicit representation of the end-to-end pipeline as mappings between well-defined spaces combined with a stochastic control formulation for rebalancing provides a common framework within which design choices can be justified *ex ante*, evaluated *ex post*, and iterated in a disciplined manner. In practical terms, this concept promotes internal transparency: every recommendation can be traced to inputs and operators, so the path from raw responses to portfolio weights is no longer a black box but a sequence of clearly stated transformations. This traceability facilitates auditability, model governance, and regulatory dialogue because assumptions, parameter values, and their effects are rendered explicit rather than embedded in undocumented codes or business rules.

A further practical gain arises from modularity. By separating the profiling operator from the allocation map and isolating the rebalancing objective into a tracking error term and a transaction cost term, architectural boundaries become precise. Each component can be substituted without re-engineering the rest: an alternative elicitation method or a new machine-learning estimator for expected returns can be introduced by replacing the relevant operator while preserving interfaces. In production systems this modularity supports versioning, A/B experimentation, and safe rollbacks. It also makes it feasible to offer differentiated strategies on the same platform, since the same control layer can govern portfolios produced by distinct allocation engines, provided the outputs remain elements of the same simplex.

Calibration and validation benefit from the fact that the objective function for rebalancing is explicitly parameterised. The turnover tolerance and cost kernel can be tuned against observed outcomes with the trade-off between tracking fidelity

and trading intensity quantified rather than intuited. Because the market state is modelled stochastically, and the profile can evolve with observable actions, end-to-end simulation becomes methodologically straightforward. Backtests can therefore be conducted under alternative market regimes and profile-evolution scenarios, producing comparative statistics for return, volatility, drawdown, tracking error, turnover, and realised costs. Sensitivity analysis with respect to parameters then yields stability diagnostics, while stress tests under extreme factor paths expose failure modes such as excessive churning or unintended concentration. The same machinery supports ongoing monitoring in production through drift detection in the distribution of questionnaire responses and market factors, with alarms triggered when the operating environment departs materially from the calibration regime.

The formalism materially improves explainability. Because each recommendation is the solution to a stated optimisation problem conditional on an investor profile and market state, explanations can be grounded in quantities that have been computed rather than inferred post hoc. Changes in weights over time can be attributed to movements in the market state, to revised profile inputs or to the operation of the turnover constraint. This form of mechanistic explanation aligns with the expectations of supervisory bodies and emerging standards for trustworthy AI while also enabling coherent investor communications that relate portfolio changes to observable drivers.

From an engineering perspective, the event-driven grid and the separation of state and control produce a clear execution loop for online decision-making. The system state at the last decision time is updated with new market and profile observations; the rebalancing problem is then solved subject to the chosen tolerance for deviation and cost. This loop admits pre-trade risk checks and guardrails expressible in the same language, for example, bounds on asset or sector exposure, floors on diversification or ceilings on one-period turnover. Crucially, such constraints can be enforced at optimisation time rather than as after-the-fact patches, reducing the risk of incoherence between business rules and portfolio construction.

The modelling also furnishes a platform for scientific enquiry. The control formulation connects RA to a literature in stochastic control and dynamic

programming (Cui et al., 2022; Khosravi, 2024; Monoyios, 2010, 2018; Tahvildari, 2021). Even where closed-form solutions are unavailable, structure can be exploited. Convexity of the tracking-error term and standard choices of transaction-cost penalties yield well-posed one-step problems; the resulting myopic policy can be studied for performance dominance under specified cost regimes. Approximate dynamic programming and model-predictive control become natural candidates for policy improvement, and their performance can be analysed with regret or sub-optimality bounds under stylised dynamics. The dynamics of the investor profile introduce an estimation problem that is amenable to Bayesian filtering or latent-state models, opening research avenues on the identifiability of preferences and on optimal questionnaire design. Because the questionnaire was partitioned into categories, one can study local perturbations and their marginal effects on allocations, enabling principled experiments that reduce the respondent burden while preserving inferential accuracy.

Interfacing with machine learning becomes cleaner when operators are defined on explicit domains and codomains. A supervised model that predicts components of the market state or expected returns has a defined role within ψ ; a model that maps partial questionnaire data to profile coordinates has a defined role within φ . This separation mitigates leakage between training and evaluation by ensuring that learning tasks are scoped and that performance metrics are aligned to their function in the overall decision. It also permits the rigorous inclusion of uncertainty from predictive models into downstream optimisation, for example, through scenario distributions or robust penalties that are calibrated to predictive residuals rather than arbitrary heuristics.

The framework facilitates rigorous comparison across RA designs. Two platforms can be compared not only on headline returns but also on the properties of their operators and cost functions, such as responsiveness to profile changes, sensitivity to regime shifts or efficiency in converting profile updates into actionable allocations. Because the state-control loop is made explicit, like-for-like experiments can be constructed where only one component differs. Such data is valuable for academic benchmarking and for due diligence by institutional partners who may integrate external allocation engines into their own delivery channels.

There are also benefits to governance, fairness and consumer protection. Explicit profiling and allocation mappings allow systematic testing for disparate impacts across demographic groups by holding market states constant and varying profile inputs along protected characteristics that are legally permissible to test. Where regulatory frameworks require demonstrable alignment between stated objectives and delivered outcomes, the model supports counterfactual analysis: “If an investor’s risk affinity had not changed, would the control policy have recommended the observed shift in exposure?” The same apparatus assists with the specification and enforcement of suitability and appropriateness rules as constraints that are evaluated at recommendation time rather than audited manually after the event.

Operational considerations are also supported. The model specifies the data required at each step, which assists in designing data pipelines and access controls consistent with privacy obligations. Because profile updates, market states and allocated weights are typed objects with associated timestamps, reproducibility is achievable through deterministic reruns of historical decision points. This enables post-incident reviews and reduces the cost of compliance responses that would otherwise rely on ad-hoc reconstructions.

Although the optimal control problem is not solved in closed form here (which is dependent on the explicit implementation of profiling questionnaires, strategies and other methods defined in the framework), the act of formalisation is itself valuable. It renders explicit the open questions that subsequent technical work must address, such as conditions under which an optimal stationary policy exists, approximation quality of heuristic controllers and the stability of feedback when profile dynamics are misspecified. These are substantive research problems that warrant a separate treatment, but their articulation within a working model ensures that future analytical results can be translated immediately into system improvements. In the interim, the framework admits practically effective policies, such as myopic rebalancing with calibrated turnover penalties, that can be deployed while more sophisticated controllers are developed. Fundamentally, the model supplies an organising principle for design, an evidential basis for calibration and monitoring, and a research scaffold that connects consumer profiling, predictive modelling and stochastic control.

3.2.5. ChatGPT-4o-Based Robo-Advisor

Based on the results of the typical questions from Table 2 in Subsection 3.2.1 and the cluster analyses in Subsection 3.2.3, the 32 distinct questions from APPENDIX 5 were used as a questionnaire for investor profiling by the GPT-based robo-advisor. The 17 instruments from APPENDIX 6, which were already developed as standard instrument (sub)classes in Subsection 3.2.2, served as the investment universe for asset allocation.

During the generation of the sample of 100 investors with portfolio recommendations, 12 (12%) errors occurred in portfolio allocation that deviated from 100%, with four cases each allocating 90%, 95%, and 110% of assets instead of 100%. These 12 errors were corrected by further prompts within ChatGPT-4o. Table 12 shows the descriptive results of the proportions, broken down by portfolio components for the entire sample of 100 investors. It is notable that real estate ETFs were only included in a few portfolios with a small share. When ChatGPT-4o was asked, it justified this decision with a general answer that this is due to investor preferences and that real estate is a niche product. Precious metals tended to appear when hedging, diversification, or asset preservation were selected as investment objectives. Cryptocurrencies were allocated across different risk affinities, but mostly to higher-risk preferences.

The AI made little use of the portfolio components Bonds EM (ESG) ETF and Bonds DM ESG ETF, preferring to invest in Bonds Global (ESG) ETF or Bonds DM ETF. Furthermore, purely sustainable portfolios were never created from ESG instruments, even if the investors wanted to invest sustainably, but always mixed portfolios of ESG and conventional financial instruments. Complementary, ESG instruments were completely avoided if the investor did not want to invest sustainably. The instruments with the highest average weight were Stocks Global ETF (16.6%), Stocks DM ETF (12.95%), Bonds Global ETF (11.9%), Bonds DM ETF (10%), Stocks EM ETF (8.15%) and Stocks Sector ETFs (7.65%). On average, ChatGPT-4o invested 59.4% in equity ETFs and 30.3% in bond ETFs across all 100 portfolios. The rest was distributed between cryptocurrencies (3.95%), real estate ETFs (3.35%) and precious metals (3%). Thus, the GPT-based RA preferred global equity and bond ETFs and developed markets over emerging markets and sector investments. Apart from traditional equities and bonds, alternative asset classes,

such as real estate, precious metals and cryptocurrencies, accounted for an average of only 10.3% of the allocations.

The detailed results per investor for the answers to the 15 question variables from Table 4, as well as the share of equities within the portfolios recommended by the ChatGPT-4o-based RA based on the sample of 100 investors, are presented in APPENDIX 11.

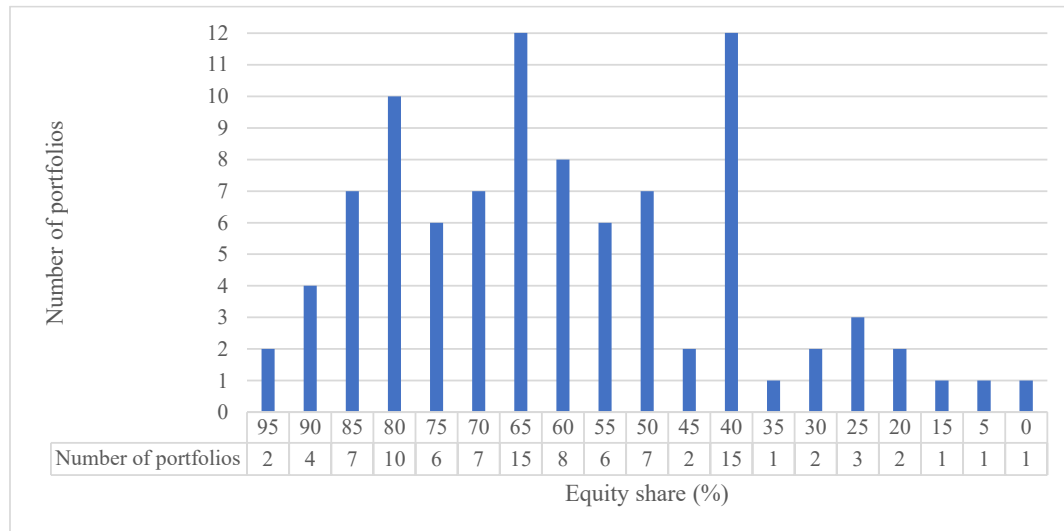
Table 12. Descriptive Statistics of Recommended Portfolio Composition

Portfolio component	Mean	Median	SD	Min	Max	Range
Stocks Global ETF	16.60%	17.50%	6.39%	0.00%	30.00%	30.00%
Stocks Global ESG ETF	5.95%	7.50%	6.22%	0.00%	20.00%	20.00%
Stocks DM ETF	12.95%	15.00%	4.56%	0.00%	25.00%	25.00%
Stocks DM ESG ETF	5.05%	5.00%	5.05%	0.00%	15.00%	15.00%
Stocks EM ETF	8.15%	10.00%	5.63%	0.00%	20.00%	20.00%
Stocks EM ESG ETF	3.05%	0.00%	4.26%	0.00%	15.00%	15.00%
Stocks Sector ETFs	7.65%	10.00%	7.70%	0.00%	20.00%	20.00%
Bonds Global ETF	11.90%	10.00%	8.49%	0.00%	30.00%	30.00%
Bonds Global ESG ETF	5.45%	0.00%	7.46%	0.00%	25.00%	25.00%
Bonds DM ETFs	10.00%	10.00%	8.67%	0.00%	35.00%	35.00%
Bonds DM ESG ETF	1.60%	0.00%	5.17%	0.00%	30.00%	30.00%
Bonds EM ETF	1.10%	0.00%	2.31%	0.00%	10.00%	10.00%
Bonds EM ESG ETF	0.25%	0.00%	1.10%	0.00%	5.00%	5.00%
Real Estate ETF	2.98%	0.00%	4.14%	0.00%	15.00%	15.00%
Real Estate ESG ETF	0.37%	0.00%	1.64%	0.00%	10.00%	10.00%
Precious Metals	3.00%	0.00%	3.55%	0.00%	10.00%	10.00%
Cryptocurrencies	3.95%	0.00%	6.75%	0.00%	25.00%	25.00%
Total Stocks ETFs	59.40%	65.00%	20.61%	0.00%	95.00%	95.00%
Total Bonds ETFs	30.30%	25.00%	20.73%	0.00%	90.00%	90.00%
Total Real Estate ETFs	3.35%	0.00%	4.33%	0.00%	15.00%	15.00%

Note. Author's elaboration: Descriptive statistics for the portfolio components recommended by the GPT-based robo-advisor for a sample of 100 investors. Mean = arithmetic mean of component weights across the 100 portfolios; Median = sample median; SD = sample standard deviation; Min/Max = observed minimum and maximum; Range = Max–Min. ETF = exchange-traded fund; DM = developed markets; EM = emerging markets; ESG = environmental, social, and governance. Values are reported as percentages and rounded to two decimal places. Totals are reported for major asset groups (Total Stocks ETFs; Total Bonds ETFs; Total Real Estate ETFs).

Figure 6 illustrates the distribution of the number of portfolios with each equity allocation within the portfolios for the 100 portfolio recommendations. The corresponding table is displayed in APPENDIX 8.

Figure 6. Distribution of GPT-Recommended Portfolios by Equity Share



Note. Author's elaboration: Bars show the number of portfolios recommended by the ChatGPT-4o-based robo-advisor at each equity share (%) for a sample of 100 portfolios (profiles). Values are listed in APPENDIX 8.

It is noticeable that portfolios with equity allocations of 65% (15/100) and 40% (15/100) occurred most frequently. These are followed by other high equity shares of 80% (10/100), 60% (8/100), 85% (7/100), 70% (7/100), 50% (7/100), 75% (6/100) and 55% (6/100). More than 89% of the portfolios had an equity share of at least 40%, and 51% of the portfolios had a high equity share of at least 65%. Portfolios with less than 50% equities (45% maximum) accounted for only 28% of the total sample.

Table 13 shows some descriptive statistics on the equity share of the portfolios, depending on the investor's answers to three questions about their risk profile: risk affinity (maximum loss acceptance), asset loss concern, and past risk willingness. The values range from 1 (lowest risk) to 5 (highest risk). The standard deviation of the equity exposure in the portfolios within the risk categories ranged between 5.99% and 16.77%, with the lower risk levels showing a higher variance. The average equity share rose from the lowest risk level to the second-highest risk level, reaching its maximum value, but then fell again to the highest risk level. This decrease was attributable to the fact that, at the highest risk preference level, the GPT-based RA invested part of the assets in cryptocurrencies as a higher-risk

investment. For example, the average equity share for the second-highest risk level (= 4) for past risk willingness was 74.38%, and the cryptocurrency proportion was 5.83%. In the highest risk level (= 5), however, the equity share fell to 69.44% in favour of a significantly increased investment in cryptocurrencies at 18.89%. However, the variance of the equity share decreased as the risk levels increased, resulting in more accurate portfolio allocations regarding stock-to-bond ratios. Lower risk levels had an average equity share between 31.82% and 41.52%, while the medium risk level was between 67.42% and 69.22%. The equity share in the higher risk levels was between 69.44% and 74.38%. The differences in the descriptive statistics of the low risk levels for asset loss concern and past risk willingness were identical, with only minor differences at the medium and higher risk levels. There was no difference between the minimum and maximum equity ratios for the last two risk variables. In general, the differences between the descriptive statistics for the risk variables were relatively small.

Table 13. Descriptive Statistics of Equity Share by Risk Profile

Variables and values	Portfolios	Mean	Median	SD	Min	Max	Range
Risk affinity (loss acceptance)							
5 (very high risk, > 35%)	15	71.67%	70.00%	7.94%	60%	85%	25%
4 (higher risk, < 35%)	18	74.17%	72.50%	10.33%	60%	95%	35%
3 (medium risk, < 20%)	33	67.42%	65.00%	16.64%	25%	90%	65%
2 (low risk, < 10%)	23	41.52%	40.00%	12.92%	20%	60%	40%
1 (very low risk, < 5%)	11	31.82%	40.00%	16.77%	0%	50%	50%
Asset loss concern							
5 (not concerned at all)	10	69.50%	70.00%	5.99%	60%	80%	20%
4 (slightly concerned)	23	73.91%	70.00%	10.22%	60%	95%	35%
3 (moderately concerned)	32	69.22%	70.00%	15.14%	45%	90%	45%
2 (very concerned)	24	40.83%	40.00%	13.08%	20%	60%	40%
1 (extremely concerned)	11	31.82%	40.00%	16.77%	0%	50%	50%
Past risk willingness							
5 (extremely willing)	9	69.44%	70.00%	6.35%	60%	80%	20%
4 (very willing)	24	74.38%	72.50%	9.92%	60%	95%	35%
3 (moderately willing)	32	68.75%	67.50%	15.03%	45%	90%	45%
2 (slightly willing)	24	40.83%	40.00%	13.08%	20%	60%	40%
1 (not willing at all)	11	31.82%	40.00%	16.77%	0%	50%	50%

Note. Author's elaboration: Statistics are computed on the 100 portfolios recommended by the GBPT-based robo-advisor. Risk items correspond to Q13 (risk affinity), Q14 (asset loss concern), and Q15 (past risk willingness); values run 1 (lowest risk) to 5 (highest risk). n = number of portfolios in each category. SD = sample standard deviation. Percentages are rounded to two decimals; Range = Max–Min.

According to Subsection 3.1.5, all the 15 variables X_i defined in Table 14 were standardised via z-scaling (26) prior to estimation according to (28). Hence, the standardised independent variables $Z_i = (X_i - \mu_i)/\sigma_i$ were used for the (multi)linear OLS regressions for modelling equity exposure Y of the recommended portfolios by the GPT-based RA.

Table 14. Independent Variables for Regressions Models of Equity Exposure

X_i	Independent variable X_i	Derived from question(s)	Unit [X_i]
X_1	Investment horizon	Q5	years
X_2	Age	Q6	
X_3	Retirement horizon	Q7	
X_4	Withdrawal horizon	Q8	
X_5	Risk affinity	Q13	1–5
X_6	Asset loss concern	Q14	
X_7	Past risk willingness	Q15	
X_8	Monthly income	Q25	EUR
X_9	Monthly expenditure	Q26	
X_{10}	Net monthly income	Q26a := Q25-Q26	
X_{11}	Liquid assets	Q27	
X_{12}	Total assets	Q28	
X_{13}	Liabilities	Q29	
X_{14}	Net assets	Q29a := Q28-Q29	
X_{15}	Coverage reserve	Q30	months

Note. Author's elaboration: Definitions and units of the independent (question) variables X_i used to model the dependent variable Y = equity share (percentage of equities in the recommended portfolio). Question numbers refer to the ChatGPT questionnaire in APPENDIX 5. Derived variables are shown explicitly (Q26a = net monthly income; Q29a = net assets). Monetary variables are reported in EUR. Regressions use Z-scaled variables (see (26)), so that X_i were standardised to Z_i prior to estimation.

The univariate regression analyses between equity exposures Y and the individual z-scaled 15 independent variables Z_i from (26) based on the question variables from Table 14 yielded the risk variables Z_3 (past risk willingness), Z_2 (asset loss concern) and Z_1 (risk affinity) as the highest correlated variables. The models qualities (adj. R^2 values) ranged 46.63–%44.20%. All other variables except for the variable coverage reserve also had a significant impact on the equity exposure variable Y at the 2.5% level, but with considerably lower model quality ($10.8\% < R^2 < 47.2\%$). The variable coverage reserve had a p-value of 0.033 (HC3 of

0.03), meaning that it represented edge significance due to the strict significance filter $\mathcal{S}$ of the FHBSS method (29). The detailed results are listed in APPENDIX 9.

The addition of a variable to become bilinear models provided substantially higher model qualities (adj. R^2) without multicollinearities. Many pairs of variables as vectors were practically linearly independent. The results of the three best bilinear models are presented in Table 15.

Table 15. Bilinear OLS Models Explaining Portfolio Equity Exposure

Metric	Var1= Z_7 (past risk willingness), Var2= Z_{15} (coverage reserve)	Var1= Z_6 (asset loss concern), Var2= Z_{15} (coverage reserve)	Var1= Z_2 (age), Var2= Z_{13} (liabilities)
alpha (intercept)	0.59400	0.59400	0.59400
beta Var1	0.14493	0.14508	-0.12819
beta Var2	0.05312	0.05790	0.08597
p-value Var1	< 0.00001	< 0.00001	< 0.00001
p-value Var2	0.00033	0.00011	< 0.00001
HC3 Var1	< 0.00001	< 0.00001	< 0.00001
HC3 Var2	0.00003	0.00001	< 0.00001
VIF Var1	1.00404	1.00936	1.00560
VIF Var2	1.00404	1.00936	1.00560
adjusted R^2	0.52828	0.52667	0.51212
AIC	-104.24542	-103.90535	-100.87712
BIC	-96.42991	-96.08984	-93.06161
Log-Likelihood	55.12271	54.95268	53.43856
Breusch-Pagan p	0.79586	0.62125	0.58480
White p	0.01245	0.00883	0.38401
Shapiro-Wilk p	0.21014	0.19705	0.62996
Jarque-Bera p	0.48178	0.44470	0.73149
Condition number	1.06554	1.10141	1.07763
Max Cook's D	0.08388	0.08569	0.09760

Note. Author's elaboration: Results from ordinary least squares (OLS) regressions modelling the portfolio equity share Y using pairs of Z-scaled variables (standardised to mean = 0, SD = 1). Columns report three selected two-variable models: (1) Var1 = Z_7 (past risk willingness), Var2 = Z_{15} (coverage reserve); (2) Var1 = Z_6 (asset loss concern, higher = less concern), Var2 = Z_{15} (coverage reserve); (3) Var1 = Z_2 (age), Var2 = Z_{13} (liabilities). A = intercept (baseline equity share, proportion); β Var1/Var2 = coefficient for the Z-scaled variables (interpretable as change in equity share in proportion units per 1 SD increase in the variable, e.g., $\beta = 0.14493 \approx +14.49$ percentage points). HC3 = p-value based on heteroskedasticity-consistent (HC3) robust standard

errors; VIF = variance inflation factor (collinearity diagnostic); AIC/BIC/Log-Likelihood = model fit statistics; Breusch–Pagan and White = tests for heteroskedasticity (p-values reported); Shapiro–Wilk and Jarque–Bera = normality tests for residuals (p-values); Condition number = collinearity/conditioning metric; Max Cook’s D = maximum Cook’s distance (influence).

It is crucial to emphasise that the statistical significance reported for the selected models should be regarded as indicative rather than confirmatory, as the stepwise best-subset procedure (FHBSS from Subsection 3.1.5) introduces selection bias and inflates the risk of type I error. Consequently, the p-values reflect in-sample associations within the chosen specifications and do not provide formal inferential evidence without appropriate post-selection adjustment. As a consequence, the results are in-sample fits according to the FHBSS method (see limitations in Subsection 5.3.1).

The first column of Table 15 contains the calculated metrics, and the subsequent columns contain the three selected models. The first model with the explanatory variables past risk willingness and coverage reserve had a model quality of adj. $R^2 = 52.83\%$, whereby, according to the beta coefficients, a standard deviation (SD) of the first variable was associated with 14.49% higher equity exposure, and the second variable was slightly weaker with 5.31% higher equity exposure. Both past risk willingness and coverage reserves therefore had a positive effect on equity exposure. There was no collinearity between the two variables according to VIF. The Shapiro–Wilk and Jarque–Bera tests showed approximately normally distributed residuals. The tests for homoscedasticity, according to Breusch–Pagan, showed no significant deviation at a 5% significance level, whereas the White test indicated mild heteroscedasticity. However, the robust estimation of standard errors HC3 also demonstrated substantial effects of both variables on equity exposure, in line with the p-values reported. Log-likelihood showed a high logarithmic probability of 55.12% that the estimated model generated the recommended equity exposure in the portfolios. The condition number indicated well-conditioned modelling (no collinearity), and maximum Cook’s D showed no extreme influence points.

The second model with asset loss concern instead of past risk willingness had a similar model quality of adj. $R^2 = 52.67\%$, with an SD of asset loss concern at +14.51% and an SD of coverage reserve at +5.79%, which had a slightly lower impact on equity exposure. The positive sign on the asset loss concern reflected the coding, since higher values meant less concern, which aligned with the other two

risk variables (risk affinity and past risk willingness), as defined by the scale (see also Table 13). Hence, the association was positive. For the purposes of interpretation, if the scale were reversed, the interpretation would also be reversed, meaning that a higher level of concern about loss would imply a lower equity ratio (negative correlation). Again, there was no multicollinearity and no extreme influence points. The residuals were approximately normally distributed according to Shapiro–Wilk and Jarque–Bera. The Breusch–Pagan tests showed no evidence of heteroscedasticity; however, the White test indicated mild heteroscedasticity. Despite this, even when using the reported robust (HC3) p-values, the model variables retained substantial and stable associations with the equity exposure.

The third model, which includes the variables age and liabilities, had a slightly lower explanatory power with an adjusted R^2 of 51.21%. The standard deviation (SD) of age had a -12.82% impact on equity exposure, while the SD of liabilities had an $+8.6\%$ impact. Age was negatively associated with equity exposure, while liabilities had a slightly positive effect. The younger the investors were, the more equity exposure was suggested by the GPT-based RA. Interestingly, however, debt had a positive impact on equity exposure. Both tests for homoscedasticity showed no evidence of heteroscedasticity. There was no evidence of collinearity, non-normal residuals, or other extreme influence points.

Hence, risk-related attitudes (past risk willingness or asset loss concern) had large, positive and high associations with equity exposure, on the order of $+14.50\%$ per SD. As described above, the scale of asset loss concern was reversed, i.e., higher loss concern had a negative effect on exposure. Liabilities and coverage reserves exhibited moderate to mild positive associations in the models, whereas age showed a pronounced negative association with the equity share of the portfolio.

When adding another explanatory variable for constructing three-factor models, higher model qualities, measured by adjusted R^2 , were obtained with low multicollinearity. In contrast to the selected bilinear models, which showed practically no collinearity (likely due to approximately linearly independent vectors), all trilinear models showed the onset of multicollinearities, which were nevertheless still comfortably low, well below the threshold ($VIF < 4$). The best three selected models are reported in Table 16.

Table 16. Trilinear OLS Models Explaining Portfolio Equity Exposure

Metric	Var1=Z₇ (past risk willingness), Var2=Z₁₄ (net assets), Var3=Z₁₅ (coverage reserve)	Var1=Z₆ (asset loss concern), Var2=Z₁₃ (liabilities), Var3=Z₁₄ (net assets)	Var1=Z₁ (investment horizon), Var2=Z₂ (age), Var3=Z₁₃ (liabilities)
alpha (intercept)	0.59400	0.59400	0.59400
beta Var1	0.17132	0.14618	0.05318
beta Var2	-0.05723	0.09661	-0.08826
beta Var3	0.09145	-0.05411	0.08560
p-value Var1	< 0.00001	< 0.00001	0.01437
p-value Var2	0.01264	0.00002	0.00008
p-value Var3	0.00002	0.02169	< 0.00001
HC3 Var1	< 0.00001	< 0.00001	0.04249
HC3 Var2	0.00833	0.00005	0.00039
HC3 Var3	< 0.00001	0.02090	< 0.00001
VIF Var1	1.56658	1.22662	2.28933
VIF Var2	2.64571	2.42677	2.29660
VIF Var3	2.19080	2.76436	1.00571
adjusted R²	0.55341	0.54690	0.53701
AIC	-108.75658	-107.31032	-105.15175
BIC	-98.33590	-96.88964	-94.73107
Log-Likelihood	58.37829	57.65516	56.57587
Breusch–Pagan p	0.94649	0.47911	0.11384
White p	0.23932	0.71281	0.43367
Shapiro–Wilk p	0.32409	0.89310	0.77238
Jarque–Bera p	0.45796	0.72922	0.92889
Condition number	2.91228	3.01020	2.65828
Max Cook's D	0.06833	0.06856	0.14571

Note. Author's elaboration: Results from ordinary least squares (OLS) regressions modelling the portfolio equity share Y using three Z-scaled variables (standardised to mean = 0, SD = 1). Columns report three selected trilinear models: (1) Var1 = Z₇ (past risk willingness), Var2 = Z₁₄ (net assets), Var3 = Z₁₅ (coverage reserve); (2) Var1 = Z₆ (asset loss concern; larger = less concern), Var2 = Z₁₃ (liabilities), Var3 = Z₁₄ (net assets); (3) Var1 = Z₁ (investment horizon), Var2 = Z₂ (age), Var3 = Z₁₃ (liabilities). α = intercept (proportion units; e.g., 0.594 $\approx$ 59.4% equity); β Var# = standardized coefficient. HC3 = p-value using HC3 standard errors; VIF = variance inflation factor; AIC/BIC/Log-Likelihood = model fit statistics; Breusch–Pagan and White = heteroskedasticity tests (p-values reported); Shapiro–Wilk and Jarque–Bera = residual normality tests (p-values); Condition number = conditioning/collinearity metric; Max Cook's D = maximum Cook's distance (influence).

The first model in Table 16., with the explanatory variables past risk willingness, net assets, and coverage, showed a high quality of adj. $R^2 = 55.34\%$. The first variable exerted the strongest association with equity exposure, corresponding to an increase of +17.13% of its standard deviation. This was followed by a smaller negative association for net assets (–5.72% of the SD) and a positive association for the coverage reserve (+9.15% of the SD), when controlling for the other two variables. Multicollinearity was comfortably low based on the VIFs of the condition number. No non-normality or extreme values were identified. Breusch-Pagan and White did not provide any indications on heteroscedasticity. The reported inference using robust (HC3) standard errors was appropriate. Past risk willingness and the coverage reserve were positively associated with equity exposure, whereas net assets showed a negative association, which is of particular note.

The second model in Table 16, containing the variables asset loss concern, liabilities, and net assets, had an adj. $R^2 = 54.69\%$. Here, too, multicollinearity was low. Breusch-Pagan and White again showed no evidence of heteroscedasticity. When considering the HC3 p-values, all three variables retained stable and meaningful contributions. An increase of one SD in asset loss concern (reversed scale) was associated with a +14.62% change in equity exposure, liabilities with a +9.66% change, and net assets with a –5.42% change. No evidence of non-normality or extreme values was detected.

In the third model, investment horizon (+5.32%), age (–8.83%), and liabilities (+8.56%)—all measured per standard deviation—were associated with equity exposure, yielding an adjusted R^2 of 53.7%. The model exhibited low multicollinearity, no indication of heteroskedasticity, and no evidence of non-normality or extreme values.

The addition of a fourth explanatory variable to create 4-linear models yielded slightly higher model qualities measured by adjusted R^2 at the expense of slightly increasing multicollinearities, which were nevertheless still low. Table 17 displays the two best models. The first four-factor model demonstrated low multicollinearity but did not exhibit any signs of heteroscedasticity, non-normality or extreme influence. All variables met the robustness criteria and displayed stable associations, with the following beta coefficients measured in the respective SD of

the variables: past risk willingness (+17.38%), liabilities (+7.07%), net assets (−9.74%), and coverage reserve (+6.96%). Accordingly, past risk willingness was strongly and positively associated with equity exposure, whereas liabilities and coverage reserve showed moderate positive associations, and net assets exhibited a moderate negative association. The model achieved an adjusted R^2 of 59.41%.

The second four-factor model, which replaced past risk willingness with asset loss concern, demonstrated robust significance, homoscedasticity, and normality, while maintaining controllable multicollinearity. The beta coefficients were as follows: asset loss concern (+16.84%), liabilities (+7.41%), net assets (−8.81%), and coverage reserve (+6.65%). The differences compared to the first four-factor model were marginal, with a model goodness of adj. $R^2 = 58.69$.

Table 17. 4-Linear OLS Models Explaining Portfolio Equity Exposure

Metric	Var1=Z ₇ (past risk willingness), Var2=Z ₁₃ (liabilities), Var3=Z ₁₄ (net assets), Var4=Z ₁₅ (coverage reserve)	Var1=Z ₆ (asset loss concern), Var2=Z ₁₃ (liabilities), Var3=Z ₁₄ (net assets), Var4=Z ₁₅ (coverage reserve)
alpha (intercept)	0.59400	0.59400
beta Var1	0.17376	0.16844
beta Var2	0.07070	0.07413
beta Var3	−0.09738	−0.08814
beta Var4	0.06956	0.06653
p-value Var1	< 0.00001	< 0.00001
p-value Var2	0.00154	0.00103
p-value Var3	0.00016	0.00053
p-value Var4	0.00110	0.00183
HC3 Var1	< 0.00001	< 0.00001
HC3 Var2	0.00149	0.00104
HC3 Var3	0.00016	0.00070
HC3 Var4	0.00010	0.00027
VIF Var1	1.56979	1.49815
VIF Var2	2.69934	2.70378
VIF Var3	3.51599	3.39917
VIF Var4	2.44947	2.42701
adjusted R ²	0.59413	0.58686

AIC	-117.36495	-115.59022
BIC	-104.33910	-102.56437
Log-Likelihood	63.68247	62.79511
Breusch–Pagan p	0.55135	0.35553
White p	0.56519	0.56981
Shapiro–Wilk p	0.94524	0.92947
Jarque–Bera p	0.92027	0.85485
Condition number	3.62410	3.55113
Max Cook's D	0.06152	0.05588

Note. Author's elaboration: Results from ordinary least squares (OLS) regressions predicting the portfolio equity share Y using four Z-scaled predictors (standardised to mean = 0, SD = 1). Column 1: Var1 = Z_7 (past risk willingness), Var2 = Z_{13} (liabilities), Var3 = Z_{14} (net assets), Var4 = Z_{15} (coverage reserve). Column 2: Var1 = Z_6 (asset loss concern; larger = less concern), Var2 = Z_{13} (liabilities), Var3 = Z_{14} (net assets), Var4 = Z_{15} (coverage reserve). α = intercept (proportion units; e.g., 0.594 $\approx$ 59.4% equity). β Var# = standardized coefficient (change in equity share in proportion units per 1 SD increase in the predictor; multiply by 100 to convert to percentage points). p = p -value (two-tailed); HC3 = p -value using HC3 heteroskedasticity-consistent standard errors; VIF = variance inflation factor; adj. R^2 = adjusted R-squared; AIC/BIC/Log-Likelihood = model fit statistics; Breusch–Pagan and White = heteroskedasticity tests (p -values); Shapiro–Wilk and Jarque–Bera = residual normality tests (p -values); Condition number = conditioning/collinearity metric; Max Cook's D = maximum Cook's distance (influence).

The FHBSS terminated after $k=4$ at the hierarchical level with $k=5$ dimensions, as none of the five-dimensional models either showed robust associations or the multicollinearities increased significantly, indicating the presence of linearly dependent variables, which would lead to overfitting and unstable models. Furthermore, the increase in dimension only resulted in marginal, but negligible, explanations of the variance of the equity exposure measured by (adj.) R^2 . In summary, across the k -linear OLS models, risk-attitude variables were most robustly associated with equity exposure. Past risk willingness and (reverse-coded) asset-loss concern consistently showed the largest positive coefficients ($\beta \approx +0.14$ to $+0.17$ per SD), with risk affinity also positive. Coverage reserve and liabilities displayed moderate positive associations ($\beta \approx +0.05$ to $+0.10$), whereas net assets ($\beta \approx -0.05$ to -0.10) and age ($\beta \approx -0.09$ to -0.13) were negative; investment horizon was smaller and positive ($\beta \approx +0.05$). Model fit increased with dimensionality: univariate adj. R^2 spanned 0.11–0.47 (top risk-attitude models ≈ 0.44 –0.47); bilinear models achieved 0.51–0.53; trilinear, 0.54–0.55; and four-variable models, 0.587–0.59. The results showed the dominance of risk-attitude factors alongside smaller contributions from assets/debt, reserve and horizon/age.

3.3. DISCUSSION

3.3.1. Scope, Categories and Typical Items of Investor Profiling

This subsection synthesises the profiling census to address research question RQ1 from Subsection 1.2, focusing on the scope, taxonomy and recurring content of investor questionnaires.

Where earlier studies, for example Kordela (2018), Tertilt and Scholz (2018), Scherer and Lehner (2022), and Oehler and Horn (2024), examined a far smaller set of German platforms, the present research identified and analysed 45 active RAs as of July 2024, highlighting the intensification of competition and fragmented market. In the pursuit of distinction within a highly competitive market, service providers have been observed to direct their attention towards distinct areas of expertise and distinctive value propositions. Some of these platforms prioritise sustainable investments, while others specialise in offering a diverse range of portfolios with customisation options. Some platforms emphasise simplicity, offering a limited number of portfolios, streamlined onboarding and profiling processes. Others employ a combination of human and automated RAs in hybrid models. This abundance of choice can leave retail clients unsure which service best suits their circumstances, because the distinctions are not always obvious to non-specialists. The assessment results indicate that, despite the evident surface differences, the providers demonstrate a considerable degree of common ground. The profiling questionnaires contain numerous overlapping items, and the resulting portfolios are dominated by low-cost (passive) equity and bond ETFs, a phenomenon that is both typical and well-known for robo-advisory (D'Acunto & Rossi, 2021; Fisch et al., 2019; Jung, Dorner, Glaser, et al., 2018).

Across the 45 RAs studied, 687 individual questionnaire items were identified. The larger sample, coupled with an increase in the average number and breadth of questionnaire items, allows the derivation of a richer taxonomy. Eight distinct categories were observable: “objectives and purpose”, “investment horizon and age”, “amount and frequency of deposits”, “risk preferences”, “sustainability preferences (ESG)”, “other preferences”, “knowledge and experience”, and “financial situation”. Other preferences included questions about preferred regions or sectors, for example. The order in which the categories were developed was also

reflected in the profiling process during questionnaire assessment, with numerous RAs structuring their questionnaires in line with this. A total of 34 representative questions were recognised from these eight categories. The 687 individual questions recorded underscore the widening scope and dispersion of profiling practices; greater complexity suggests that the underlying algorithms now have more granular data with which to gauge client needs and risk tolerance, potentially enabling more finely-tuned portfolio recommendations. On average, each RA provider posed 15.27 questions (median = 14), with a standard deviation of 6.96 and counts ranging from 2 to 36. This heterogeneity reflects the wide divergence in profiling algorithms: nevertheless, most providers centred their assessments on objectives, investment horizon or age, risk tolerance, sustainability preferences, experience, and financial situation. Risk appetite, experience, and financial status attracted the greatest depth of questioning.

Categories with the highest average counts were “financial situation” (approx. 3.58), “knowledge and experience” (approx. 3.16), “risk preferences” (approx. 2.64) and “sustainability preferences” (approx. 1.96). Risk-orientated, horizon-related, experiential and financial questions appeared in at least 89 % of platforms; no single RA dominated all categories, though some led several while others lagged across many. Compared to the results from previous scientific literature, risk preferences have expanded the portfolio dimensions, and sustainable portfolios are increasingly being offered. Nevertheless, the most common categories in the 45 RAs analysed were “risk preferences” (97.78%), “investment horizon and age” (91.11%), “knowledge and experience” (91.11%) and “financial situation” (88.89%). No RA covered all categories in terms of the number of questions, but the results show that there were some RAs that dominated in numerous categories and, conversely, RAs that had significantly fewer questions in many categories.

Faloon and Scherer (2017) recommended collecting household income statements and balance sheets and developing a framework for asset allocation based on these key figures. By analysing four German RAs, Scherer (2017) found that they failed to account for human capital, financial assets, liabilities or real estate when determining asset allocations. Tertilt and Scholz (2018) studied seven German RAs and similarly reported that data on the investor’s income, expenses,

job, dealing with financial decisions, experience with risky investments, and risk behaviour was not collected. This means that essential elements of household balance sheets, investment knowledge, and risk capacity were neglected by German RAs, despite financial regulatory requirements. In contrast to the previous observations, the findings of this thesis demonstrate that RAs are now asking more extensive questions about risk tolerance, household income statements and balance sheets (e.g., income, expenses, assets, liabilities), previous experience on financial markets, individual financial situation, job, and risk-related aspects (e.g., tolerance, affinity, preferences) (see Table 3 and Table 4). 77.78% of RAs asked questions about investment objectives and reasons, 91.11% about investment horizon or age, 84.44% regarding investment amounts and savings plans, 97.78% about risk preferences and risk behaviour, 80% concerning sustainability preferences, 91.11% on knowledge and experience in financial markets and 88.89% on the financial situation of investors.

A total of 119 risk-related questions were identified, compared to 88 sustainability-related questions among the 45 RAs. 31.11% of RAs asked three or more questions in the risk category, while 28.89% asked the same number of questions about investors' sustainability preferences. Therefore, based on the number of questions per RA, it can be concluded that investors' preferences for sustainable investments play a similarly important role to risk appetite. In previous studies, sustainability preferences were not identified in the profiling process of German RAs at all (Faloon & Scherer, 2017; Kordela, 2018; Scherer, 2017; Scherer & Lehner, 2022, 2025; Tertilt & Scholz, 2018).

The TF-IDF and cluster analysis validated that categorisation with $K = 7$ and $K = 8$ categories represented the finest categorisation with high selectivity and lower within-category squared sum of errors (WCSSE). This signifies that the questions within these categories were proximate to their respective category centroids. In general, the WCSSE was highest for the three categories of risk, experience, and financial situation, indicating a tendency for different questions to emerge within these categories. The variant with $K = 7$ encompassed the addition of eight further questions within the sustainability category. For instance, questions pertaining to region and sector were found to be closely aligned with the sustainability questions in a broader sense. The top five questions, based on the lowest distances to the category centroids, corresponded very well with the 34

representative questions from the eight categories that had previously been identified. It was observed that some of the 34 representative questions manifested subsequently, that is, following the presentation of the top 5 lowest distances to the category centroids. This phenomenon can be attributed to the fact that within the categories, certain questions predominate over others due to their numerical prevalence and similarity, thereby giving rise to subclusters. A thorough review of the categories was conducted, during which questions with identical content were removed. The remaining questions were then sorted according to their proximity to the category centroid. This process resulted in the identification of the 34 representative questions that had been previously outlined. The elbow curve in the graph's plot, denoted by $\Gamma(K, WCSSE(K))$, revealed two distinct elbows.

In addition to the finer categorisation between $K = 7$ and $K = 8$, as previously mentioned, a coarser categorisation between $K = 4$ and $K = 5$ was also found to be optimal, according to the elbow method. The Calinski–Harabasz index (CHI) recommended Peaks with categorisation $K = 3$ and again between $K = 6$ and $K = 7$. The categorisation for $K = 7$ did not differ significantly from $K = 6$, suggesting that this alternative could also be employed. However, in the coarser categorisations with $K < 5$, more detailed information was lost due to overly general categorisation when only the top questions close to the category centroid were selected. For instance, questions pertaining to risk and sustainability preferences exhibited a high degree of cohesion, with sustainability preferences occupying a prominent position within the top five due to their numerical prevalence, linguistic similarity, and dominance. In contrast, the risk categories emerged at a later stage, as the questions within this category manifested substantial linguistic distinctions, employing disparate terms such as “loss”, “risk”, and “nervous”. With smaller numbers of categories, such as $K = 3$ and $K = 2$, an increasing number of subclusters formed, rendering it impossible to distinguish between them in terms of content. This finding was made independently of CHI. With $K = 5$, it remained possible to achieve a clear distinction using the top five questions that were selected based on their category centroid distance as representative examples. This included the categorisation of investment objectives, investment amounts and investment horizons in comparison to $K = 7$. In consideration of the aforementioned findings, it is recommended that a refined categorisation with $K = 7$ or $K = 8$ be employed, complemented by a coarser categorisation with $K = 5$ as an alternative.

In comparison to this thesis' findings, So (2021) defined five categories based on 20 Hong Kong and overseas RAs with 180 questions, consisting of investment objectives and horizon, risk appetite, knowledge and experience, investor behaviour given loss, and risk-taking ability. There were no other categories, such as personal financial situation or sustainability preferences. Instead, So (2021) defined three risk categories, while this thesis combined them into one. However, the 119 questions from 44 RAs in this work demonstrated that a single risk category was sufficient to encompass the questions on risk, as the cluster analysis evidenced.

Scherer and Lehner (2022) developed five categories for investment objectives, investment horizon, wealth, risk aversion, and experience, based on 16 German RAs. The category "wealth" can be considered analogous to the category defined in this thesis for "amount and frequency of deposits". The remaining categories are comparable to those from the cluster analysis categorisation for $K = 7$, where in this thesis additional categories for financial situation and sustainability preferences arise due to a significant increase in new questions gathering additional investor profile information.

Conversely, Tertilt and Scholz (2018) proposed a three-category classification system comprising general information and two risk categories (risk capacity, risk tolerance). This classification was derived from a set of 13 RAs from Germany (7), the US (3), and the UK (3), yielding a total of 145 questions. The general category comprised a mixture of questions concerning the financial situation, the timing of retirement, or the job type. The category "risk capacity" included a combination of questions regarding the financial situation (capital and debt), the investment horizon, savings rates and the amount invested. The category "risk tolerance" consisted of questions regarding knowledge and experience, investor age, investment objectives, and risk and return preferences. The findings presented in this thesis demonstrated a more structured, refined, and accurate categorisation, with questions from a particular subject area grouped together instead of being dispersed across various categories. As indicated by the cluster analysis, subcategories emerge in scenarios where the number of categories is minimal, such as in cases where $K = 3$. Nevertheless, even with $K = 3$, a category definition of "objectives, amount, and horizon", "risk and sustainability", and "experience and financial situation" can provide a more refined categorisation, as the CHI also confirmed.

This thesis' new findings show that RAs apparently have included investor preferences for sustainability in their profiling processes, asking a high proportion of questions about ESG topics, which is similar to the proportion of questions about investment risk appetite in the questionnaire. In addition, there are significantly more specific questions in all other categories, such as financial situation or risk preferences. Hence, on average, there are more "know your customer" (KYC) questions, suggesting improved alignment with regulatory standards.

However, there remain notable weaknesses in implementation. Only around half of the RAs asked at least one question in each category. One-fifth of them ignored the investor's investment objectives. Half of the RAs asked a maximum of two questions about financial knowledge and experience. 40% maxed out at two questions or fewer about financial circumstances, such as income, expenses, and assets and liabilities. Moreover, the analysis of the questionnaires revealed that 42.22% asked a maximum of one question, and 68.89% asked up to three questions at most concerning risk preferences. Just 15.56% included at least five questions in the risk category. Risk tolerance and affinity are therefore only assessed to a limited extent by many RAs on the market.

The number of questions relating to investors' risk appetite can be considered insufficient in view of existing risk-related psychological research, which uses more advanced psychometric questionnaires to assess investors' risk behaviour, that have been validated in the fields of psychology and household finance. One such questionnaire is the 40-item Domain-Specific Risk-Taking (DOSPERT) scale, which has a revised version containing 30 items and uses Likert-type items to gauge risk-taking and risk perception (including a dedicated financial investment domain, which is distinct from gambling), showing to have robust factorial validity and reliability (Blais & Weber, 2006; Shou & Olney, 2020; Weber et al., 2002). The Grable-Lytton Financial Risk Tolerance Scale (GL-RTS) focuses explicitly on financial risk tolerance via scenario-based and self-assessment items; it is shorter (13 items), widely used in practice, and exhibits good internal consistency and construct validity (Grable & Lytton, 1999; Kuzniak et al., 2015; Lawrenson & Dickason-Koekemoer, 2020). General propensity measures include the Risk Propensity Scale (RPS) with 7 items and the General Risk Propensity Scale (GRiPS) with 8 items, both unidimensional, brief, and suited to field settings (Meertens & Lion, 2008;

Zhang et al., 2019). The 12-item Risk-Taking Index (RTI) offers a profile differentiated by domain (including finance) with demonstrated psychometric soundness (Häusler et al., 2018; Nicholson et al., 2005). Such instruments usually use summated Likert scoring and report reliability (e.g., Cronbach's α), and they have also been validated against behavioural criteria. They provide practical, evidence-based information for investor profiling in advisory contexts. RAs are advised to include validated, standardised psychometric questionnaires in their risk-related category and replace their existing ones, as these provide a more scientifically sound basis for risk profiling. This would make the categories of questions about risk affinity and preferences more comprehensive and accurate.

Apart from the professional assessment of investor risk profiling from a psychological perspective, there are also regulatory requirements for risk assessment. The German Securities Trading Act, the Wertpapierhandelsgesetz (WpHG), regulates the legal requirements for investment advice provided by financial institutions operating in Germany, which is also relevant for robo-advisory. § 31(4) WpHG obliges advisors to take into account the expertise, investment objectives and financial circumstances of their clients so that only financial instruments that correspond to the investor's risk tolerance are recommended. In doing so, the information must be collected and documented in accordance with § 34 WpHG within the framework of investor protection. A robo-advisor that fails to elicit, quantify, and integrate an investor's risk preferences also cannot comply with the European Union's (EU) core suitability and product-governance rules, nor can it credibly match clients to the standardised risk labels that now accompany retail investment products. Article 25(2) of Directive 2014/65/EU (MiFID II) establishes the legal foundation: before providing discretionary portfolio management or investment advice, a firm "shall obtain the necessary information [...] including [the client's] risk tolerance so as to enable the firm to recommend [...] instruments that are suitable [...] and in particular are in accordance with his risk tolerance and ability to bear losses" (ESMA, 2016; European Union, 2014). This statutory duty is principles-based and technology-neutral; consequently, it applies with full force to RAs as well as human advice models. The Guidelines on Certain Aspects of MiFID II Suitability Requirements by the European Securities and Markets Authority (ESMA) flesh out the obligation. The ESMA is the EU's independent supervisory authority for securities markets,

tasked with safeguarding investor protection and financial stability (ESMA, 2025). It develops the EU single rulebook (standards and guidelines), coordinates national regulators, and directly supervises certain entities such as credit rating agencies and trade repositories.

According to ESMA, firms must gather “up-to-date, accurate and sufficient information” on knowledge, experience, financial situation and investment objectives, “including their risk tolerance.” Without this data, “firms cannot provide investment advice and portfolio-management services” (ESMA, 2023, p. 5). The same document highlights specific challenges in automated channels: when advice is delivered through an online questionnaire, the responses “may be the sole basis for the robo-advice” (ESMA, 2023, p. 6), so investors must be told explicitly that their answers drive portfolio construction, and the questionnaire must be designed to detect inconsistencies and to avoid self-assessment bias. In other words, robust risk-profiling is not ancillary but central to algorithmic suitability.

Crucially, EU disclosure regulations have standardised the numerical risk scales used for retail instruments. Firstly, Undertakings for Collective Investment in Transferable Securities (UCITS) funds must display a Synthetic Risk and Reward Indicator (SRRI) on a scale of 1 to 7 in the Key Investor Information Document (KIID) (ESMA, 2009). All other packaged retail and insurance-based investment products (PRIIPs) carry a Summary Risk Indicator (SRI) with the same ranking, which aggregates market, credit and liquidity risk in the KID (European Union, 2017). These harmonised labels naturally bridge the gap between product risk and investor tolerance. An RA that captures a client’s willingness and capacity to bear loss can use an algorithm to map a “risk bucket” (e.g. “moderate” = SRI/SRRI 3-4) to eligible instruments, while simultaneously filtering out products whose risk class exceeds the client’s profile. Without reliable risk preference input, this mapping becomes ineffective, and the selection engine risks systematic misguiding.

It is recommended that RA providers revise their investor (risk) profiling and conduct further research into the investor’s risk profile. Using standardised psychometric questionnaires in the risk profiling process provides a professional understanding of the investor’s risk appetite. In addition, regulatory documents containing key figures, such as SRI, allow these to be divided into risk groups.

Based on this, risk-dependent portfolios can then be allocated using various equity ratios, for example. This approach is instrumental in determining the investor's risk profile with greater precision, enabling the construction of tailored portfolios that align with the investor's risk tolerance and specific requirements. One disadvantage is that investors may drop out during the onboarding process due to the length of the questionnaires. However, RA providers must decide whether to carry out recognised, psychologically grounded and regulatory-compliant investor risk profiling to better understand investors and create targeted, long-term portfolios, or focus on customer acquisition and optimising the customer journey to maximise their acquisition KPIs. According to the results of this study, most RAs currently pursue the latter strategy, striking a balance between the two. This suggests that RAs may provide poor advice, leading to portfolios that are unsuitable for investors.

Table 18 provides a summary of the key findings and recommendations for various stakeholders of the analysis of robo-advisors' profiling questionnaires.

Table 18. Key Findings from Robo-Advisors' Profiling Questionnaires

Key finding	Recommendation	Primary audience
Broad profiling scope: 687 items across 45 providers cluster into 8 categories (objectives, horizon/age, deposits, risk, ESG, other preferences, knowledge/experience, financial situation).	Maintain the 8-category structure as the minimum KYC spine; publish a short schema so clients and auditors can see what is collected and why.	Providers
Coverage is uneven: many firms ask ≤ 2 items on knowledge/experience and financial situation; only about half touch all categories.	Introduce category floor checks (≥ 1 scored item per category) and contradiction flags (e.g., high risk + very short horizon).	Providers
Risk and ESG questions are now common (risk $\approx 98\%$, ESG $\approx 80\%$); ESG depth remains variable.	Structure ESG into sub-preferences (e.g., climate, exclusions, impact) and connect each to instrument	Providers, Compliance

	filters/tilts; log reasons when ESG requests cannot be met.	
Psychometric quality varies; non-validated risk scales dominate.	Adopt validated scales (e.g., GL-RTS, DOSPERT), target $\alpha \geq 0.8$, and disclose scoring logic; use adaptive routing to limit burden.	Providers, Researchers
TF-IDF/clustering supports K = 7 and K = 8 categories; highest dispersion in risk, experience, financial situation.	Standardise wording for high-dispersion areas; curate item banks with tested phrasing to improve reliability.	Regulators/ Standard-setters
Mapping to product risk is not always explicit.	Align client risk grades with SRRI/SRI; show the bucket mapping (e.g., “moderate → SRI 3–4”) in the advice summary.	Providers, Compliance
Profiles are static at onboarding in many firms.	Schedule periodic re-profiling (annual or life events) and version profile changes for auditability.	Providers
Answers drive the entire portfolio mapping. Some RAs do not cover all important (5–8) categories. Some questions (such as those on sustainability preferences) are not yet implemented correctly.	Check that most categories are covered; if not, ask for the risk-bucket/SRRI mapping and why missing areas (e.g., liabilities or ESG) are not considered.	Investors/ Clients
Profiles change with circumstances.	Update profile after life events (job change, mortgage, dependants); keep answers consistent and avoid “gaming” risk questions.	Investors/ Clients

Note. Author’s elaboration: Findings are based on the content analysis of 687 questionnaire items from 45 robo-advisors active in Germany as of July 2024. Abbreviations: KYC = know-your-customer; ESG = environmental, social, and governance; TF-IDF = term frequency–inverse document frequency; GL-RTS = (example validated risk tolerance scale); DOSPERT = Domain-Specific Risk-Taking scale; SRRI/SRI = synthetic risk and reward indicators / standard risk indicators. Recommendations are practical suggestions for providers and their functional departments, regulators, researchers, and investors.

3.3.2. Portfolio Menus and Strategy Patterns

This subsection addresses research question RQ2 from Subsection 1.2., which concerns the composition of portfolio menus, the asset-class building blocks, weighting patterns, and prevailing strategies.

The portfolio composition analysis revealed portfolios consisting of broadly diversified passive ETF instruments (stocks and bonds), distributed globally across different geographical regions and markets, with classic market capitalisation-weighted passive ETFs on stock indices such as the MSCI All Country World Index, MSCI World Index, American indices such as the S&P 500, NASDAQ-100, etc. dominating. This observation is consistent with the findings of previous research, which shows that RAs favour diversified, low-cost passive ETFs (D'Acunto et al., 2019; Fisch et al., 2019; Jung, Dörner, Glaser, et al., 2018; Jung et al., 2019).

The composition analysis of the 405 portfolios showed an average of nine portfolios per RA, with a standard deviation of 7.55 and a range of three to 37 portfolios. This indicates a high degree of granularity in risk profiling: some RAs recommend three pre-built portfolios, while others recommend over 30. Hence, the variance in the recording of the actual investor profile through to the creation of the portfolio is very high. Notably, approximately half of all RAs have a maximum of five portfolios in their offering. This does not appear to be very granular either, given that the questionnaires contain numerous questions from different content categories. The analysis revealed that 77.78% (35/45) of RAs offered 10 portfolios or fewer, which is consistent with the 75% share reported by Scherer and Lehner (2022) based on 16 RAs. The portfolios in this study differ essentially only in terms of equity allocation share, as can be seen from the analysis results. Therefore, it is valid to criticise that the majority of RAs have a very limited range of options for portfolio creation and do not accurately classify investors to create individual portfolios. Instead, they offer a relatively low number of predefined ETF-diversified portfolios and use the questionnaires to categorise various investor profiles into a few broad groups, before offering a portfolio recommendation from the available selection. Hence, most RAs create their portfolios in advance and do not continuously recompile them for each input combination, as this would entail a significantly higher number of portfolios with much more variations regarding asset classes and weightings. In accordance with the assertions put forward by

Scherer and Lehner (2022), this is interpreted as evidence of a management logic based on an efficient frontier, instead of portfolio modelling with varying inputs evaluated from first principles. In line with the recommendation that RAs refine the determination of investor risk preferences, it is also recommended that the number of portfolios be significantly increased, offering a much greater variety of equity exposure allocations to accommodate different risk profiles.

The upper quartile of RAs accounted for over 50% of all portfolios, indicating a skewed distribution. Notwithstanding the fact that each recommendation was primarily constructed from equity- and bond-tracking ETFs, diversification has increased. A total of 46.67% of RAs offered ESG-orientated portfolios, comprising 147 portfolios and accounting for 35.31% of the total. Of these, 40% integrated commodities such as gold, 15.6% incorporated real estate funds or other alternative assets, and 8.9% provided exposure to cryptocurrency.

However, the range of ESG options remains limited, indicating a need for greater precision in aligning client sustainability preferences. Despite the presence of detailed sustainability preferences questions, no implicit differentiation in portfolio composition through sustainable instruments can be observed. It is evident that a proportion of RAs do not translate sustainability preferences into ESG instruments, or only do so to a limited extent, as their investment universe is restricted. Conversely, there is minimal additional differentiation between ESG instruments. The existence of ESG/sustainability instruments is evident, though the extent to which sustainability preferences are incorporated remains ambiguous. A customised selection of sustainability priorities, such as climate neutrality or social aspects, is not (yet) correctly and separately reflected in the composition of RAs in the portfolio with weighting. This is likely attributable to the considerable number of potential sustainability preferences, which renders product selection arduous for RAs due to the cost structure, liquidity, and the range of (potentially) less transparent sustainability instruments available. Thus, there is still potential for RAs to develop their ESG product offering. ESG-focused evidence suggests that low-carbon automated funds can outperform (Shan et al., 2022). Further research indicates that adoption hinges on trusted, personalised ESG portfolio construction (Chen et al., 2025) and that demand is concentrated among older, female, lower-wealth and lower-risk segments (Faradynawati & Söderberg, 2022). In addition,

there are regulatory risks, since, although almost half of the RAs examined offer ESG-orientated portfolios, these sustainability preferences are often reflected only to a limited extent and with little differentiation in the actual portfolio composition. Regulation (EU) 2019/2088 on Sustainable Finance Disclosure Regulation (SFDR) requires harmonised, transparent and product-related disclosures, as well as the systematic integration of sustainability risks and the consideration of material adverse effects (European Union, 2019). These are requirements that many of the RAs observed in the study appear to only partially meet.

Portfolios are typically characterised by risk bands based on equity exposure. This is a finding from the analysis of 405 portfolios of the 45 RAs, as 11 different risk groups emerged based on equity exposure interval limits. The interval limits were based on 10% increments from 0% equity to 100% equity share as target equity within the portfolio strategies. Scherer and Lehner (2023) also used this approach in their analyses, dividing RA portfolios into 10 equity exposure bins for comparison purposes. However, their equity allocations included not only equities, but all non-loan assets, i.e., equities, real estate, commodities and alternative instruments. In this thesis, though, the actual equity percentage was counted towards equity allocations, which allows for a more precise allocation of risk classes. During the investor conducted profiling process, it was noticed that an investor profile with a higher risk preference corresponds to proposals with portfolios with a higher equity share. It is therefore important not to dilute the equity exposure in portfolios with investments such as commodities, real estate, or other investments with a completely different risk profile. These patterns align with cross-sectional evidence by Boreiko and Massarotti (2020): recommended equity weights are primarily driven by the investor's risk profile, while the sheer number of model portfolios is not a significant determinant; cross-firm differences are better explained by platform asset-class expertise (equity vs. fixed income) and geography (with US RAs more equity-heavy than German peers) (Boreiko & Massarotti, 2020). Taken together, their results reinforce the emphasis on robust risk-to-portfolio mapping and transparent rationales for asset-class tilts, rather than catalogue expansion per se.

Thus, the equity exposures of the analysed portfolios in this thesis were subdivided into 10% increments ranging from 0% to 100% equity weightings. The most prevalent equity allocations were a 50/50 split between equities and bonds,

followed by 70% and 30% equity exposure. Despite the variation in construction techniques, RAs encompassed the full range of portfolio types, from defensive to balanced to growth-focused strategies. However, at the individual RA level, as previously criticised, there was a lack of portfolios that reflect these 11 risk categories. It is recommended that RA providers ensure that at least one portfolio with the respective equity exposure is offered for each of the 11 risk categories to cover at least 11 risk profiles. The analysis demonstrated that approximately half of the RAs did not meet this requirement. Consequently, there is a lack of RAs with a diverse range of equity exposure portfolios, which prevents the accurate reflection of investors' risk profiles. When considering the fact that, in addition to risk preferences, a myriad of other questions from the categories in the questionnaires are posed in the profiling process, the issue arises as to how these can be precisely translated into portfolio allocations. This raises doubt about the ability of the sheer number of different answer combinations to ultimately result in suitable portfolios for every investor profile.

In order to address each salient aspect captured by the onboarding questionnaire without allowing the catalogue of model portfolios to become unwieldy, the profiling dimensions can be sensibly grouped into a two-layer design. The pivotal issue under consideration is the fundamental risk decision that underpins the financial strategy, namely, the decision between equities and bonds. The creation of portfolios is predicated on 11 distinct basic portfolios, with equity allocations ranging from 0% to 100% in 10% increments. These 11 "core" mixes form the skeletal structure upon which further portfolio variations are superimposed. The additions of further parameters to the portfolio multiplies the portfolio combinations.

The subsequent layer provides overlays that modify the underlying mix without altering its stock-bond ratio. Firstly, the objectives and time horizon of the investor must be taken into consideration. Given the tendency of both elements to trigger analogous adjustments, such as shorter duration bonds, higher liquidity buffers, and different rebalancing rules, it is possible to merge them into three meta-variants: preservation for short horizons, balanced income for intermediate horizons, and growth for long horizons. The concept of sustainability preferences introduces a second fork, as each core mix is typically available in both a

conventional and an ESG-screened version. The possession of knowledge and experience results in a further bifurcation: providers could maintain a simplified portfolio with simpler instruments for novices and a more granular or alternative-tilted portfolio with more products for advanced users. This information can be combined with the other question categories to keep the number of portfolios low. Additional information is provided by the financial situation and the amount and frequency of deposits that help to determine the number of individual instruments, or which instruments are suitable for many or only a few trades in the context of rebalancing and deposits/withdrawals. For example, if there are more deposits and withdrawals and the financial situation is more volatile, it may be advisable to have fewer instruments and less expensive instruments in the portfolio, as this is likely to result in more trades being executed and deposits and withdrawals leading to higher costs when rebalancing. In conclusion, Applying the minimal set of genuine variants to the eleven risk buckets produces

$$11 (\text{risk}) \times 3 (\text{horizon}) \times 2 (\text{ESG}) \times 2 (\text{knowledge, frequency}) = 132 \quad (38)$$

portfolios. This number is representative of a lower bound that is both financially substantial enough to honour the information captured during profiling and sufficiently restrained to remain operationally feasible.

The analysis of the strategies and methods used by the RAs revealed that two-thirds used MPT for their asset allocation. Almost half applied impact investing methods by taking sustainable products (ESG) into account. One-third used active management techniques such as hedging, and approximately 29% used passive buy-and-hold strategies. Hybrid approaches with human oversight were mentioned by a quarter of the RAs. A few RAs used VaR, Black–Litterman, factor investing (smart beta) or core–satellite strategies. In comparison, Beketov et al. (2018) found that MPT was also the most commonly used methodological framework in 39.7% of RAs. Factor investing was found in 2.7% of RAs by Beketov et al. (2018), although this study showed the proportion to be 6.67%. Risk parity was found in 1.4% of cases by Beketov et al. (2018) and in 2.22% of cases in this study. Beketov et al. (2018) referred to sample portfolios as a strategy with a share of 27.4% and mentioned that the strategies were unknown, although it was assumed that this meant a simple compilation of portfolio instruments without much customisation. Constant portfolio weight was also mentioned as a method, which simply means rebalancing to keep the portfolio component weights

constant. However, this study found that rebalancing was used in all RAs to correct deviations from target equity exposures and thus maintain the weightings within the strategy. VaR and Black–Litterman were mentioned in Beketov et al. (2018) as additions to MPT but were not reported separately. Similarly, there was no documentation on impact investing (ESG), AI-driven strategies or hybrid approaches.

In this study, five AI-enabled RAs employed a variety of advanced techniques, ranging from probabilistic modelling to multi-model ensemble learning. A platform integrated Bayesian Neural Networks with natural language processing to generate probabilistic forecasts, quantify uncertainty, and process unstructured data. In addition, statistical machine learning was employed for macroeconomic and market analysis, with the aim of supporting probabilistic re-weighting of portfolios under human oversight. Multi-layered frameworks frequently combined Hidden Markov Model (HMM) for regime detection, Long Short-Term Memory (LSTM) networks for time-series and news-based forecasting, and eXtreme Gradient Boosting (XGBoost) for refined trend classification. Specific platforms functioned by means of multi-agent AI architectures, in which specialised models are responsible for monitoring each other. These models utilised customised, position-specific metrics and self-learning neural networks to process large volumes of data. At the most automated end of the spectrum, quantitative AI engines merged neural networks, sentiment and news analysis, and behavioural-psychology-based risk models to produce real-time buy/sell signals, dynamically reallocating capital across asset classes while aiming to minimise drawdowns and preserve long-term returns.

The outcomes indicate that the application of AI in robo-advisory is on the rise, a phenomenon that was not previously observed in earlier studies such as those conducted by Beketov et al. (2018). The AI methods utilised encompass sophisticated machine learning techniques and deep learning algorithms, in addition to dynamic mathematical models for financial market analysis and risk assessment. However, further research is required to determine whether these models ultimately offer an advantage in terms of risk-adjusted performance. In the context of this discussion, it can be assumed that human intervention will decrease and that, as postulated by the RAs, human oversight and monitoring will play an

increasingly important role in a hybrid human-and-machine approach alongside AI methods.

Table 19 provides a summary of the key findings and recommendations for various stakeholders based on the analysis of the robo-advisors' portfolios and strategies.

Table 19. Key Findings from Robo-Advisors' Portfolios and Strategies

Key finding	Recommendation	Primary audience
Menus are ETF-centric (global equities + EUR-hedged global bonds); equity weight is the primary differentiator.	Keep the low-cost beta core, but publish holdings, costs, and index exposures to aid comparability.	Providers
Granularity is limited at provider level ($\approx 50\%$ offer ≤ 5 portfolios; $77.78\% \leq 10$).	Guarantee at least 11 equity bands (0–100% in 10% steps) so every risk bucket is covered.	Providers
Many-to-few mappings from rich profiles to only few model portfolios.	Implement a two-layer design: 11 equity cores $\times$ overlays (horizon 3, ESG 2, novice/advanced 2).	Providers
ESG offer present but translation from preferences to holdings is often coarse.	Provide pure-ESG variants plus best-in-class/tilt/exclusion options; disclose ESG	Providers, Compliance
Modern portfolio theory (mean–variance optimisation) dominates; rebalancing is universal but the rule is often opaque.	Adopt and disclose threshold/volatility-aware rebalancing; report drawdown/recovery alongside	Providers
Diversifier sleeves (commodities/real estate) are selective; crypto appears in a minority.	Treat diversifiers as capped sleeves; require suitability checks and disclosures for high-volatility assets.	Providers, Compliance
Catalogue concentration: upper-quartile providers hold $>50\%$ of portfolios.	Test menu breadth vs comprehension; prioritise explainability over sheer catalogue size.	Providers, User Experience
The number of portfolios at different risk levels based on	Check that equity band exists (0–100% in 10% steps). If not,	Investors/ Clients

equity exposure is low for many providers; portfolio breadth may not cover the investor's risk bucket.	request a custom overlay or consider a provider with fuller coverage.	
ESG labelling may not equal ESG holdings.	Request the ESG look-through (indices, exclusions, tilts) and the rebalancing policy (calendar vs threshold) before accepting a mandate.	Investors/ Clients

Note. Author's elaboration: Findings are drawn from the composition analysis of recommended portfolios across the sample (405 portfolios from 45 robo-advisors; data collection: Germany, July 2024). Abbreviations: ETF = exchange-traded fund; ESG = environmental, social, and governance. Recommendations indicate practical actions; "Primary audience" suggests the stakeholder(s) responsible for implementation/action. Compliance and user experience refer to functional areas of robo-advisor providers. Where numeric cutoffs are proposed (e.g., "11 equity bands"), these are operational suggestions—providers should adapt them to regulatory requirements and internal suitability processes.

3.3.3. A Dynamic Formal Framework for Robo-Advice

With regard to the research question RQ3 from Subsection 1.2, the formal framework is discussed in this subsection in terms of its function as a generalisation of prior static formulations and as an operational scaffold for providers.

Moreover, the developed mathematical model of the entire RA process can be utilised for the purpose of further research and the advancement of existing profiling and portfolio management implementations of RAs. While earlier studies on the RA process, such as those by Faloon & Scherer (2017), Beketov et al. (2018) and Jung, Dorner, Glaser, et al. (2018) only described the building blocks and their connections within the RA process qualitatively or visually, they refrained from formal mathematical modelling of the process to map the individual building blocks. In contrast to Bourgeron et al. (2018), the framework is constructed to encompass the entirety of the RA process, rather than focusing exclusively on individual algorithmic components in the asset allocation subprocess. Bourgeron et al. (2018), concentrated on enhancing the numerical robustness of conventional mean–variance engines, primarily by reformulating quadratic-programming problems via the alternating direction method of multipliers (ADMM), and thus treated investor profiles and market inputs as exogenous. Likewise, Perrin and Roncalli (2020) surveyed large-scale optimisation techniques (coordinate descent,

proximal gradients, and Dykstra's algorithm), demonstrating their applicability to static portfolio allocation tasks.

In this thesis, by contrast, the modelling framework employs a rigorous two-stage mapping to depict the RA workflow. Firstly, the questionnaire is employed to transform raw answers into a concise profile vector, which records risk appetite, time horizon, liquidity needs, sustainability stance, and similar traits. The profiling operator may be either deterministic or stochastic when clients revise responses or supply imprecise self-assessments. Thereby, it captures both self-assessment noise and imputation uncertainty. This new feature is the probabilistic dynamics in profiling, meaning that the RA profiling process is not modelled statically in one direction, but can be restarted, for example, by revising or updating the investor profile.

Empirical evidence indicates that portfolio adjustments after robo-advice are group-specific: inexperienced users tend to persist with default allocations, whereas experienced users frequently readjust towards higher equity shares (Bräuer, 2021). These heterogeneous, state-dependent responses justify modelling investor preferences as dynamic and path-dependent. Accordingly, the developed model allows the investor profile and recommended risk level to evolve over time, with transition dynamics calibrated separately for inexperienced and experienced users and by prior equity exposure. This has generalised the static profile ω into a profile space-time ω_t , which is incorporated into the portfolio allocation and optimisation. This implies that even changes to individual questions can lead to completely different profiles and thus portfolio allocations. In this case, the RA could respond with a recalibration to align the portfolio with the new investor profile. However, for reasons of practicability and cost, the question arises as to how frequently the investor profile can be changed for the RA under the generally low fee regime of RAs. The RA could initiate this by means of regular evaluations at specific times and, if there are any changes, these would be implemented gradually through reallocations in the portfolio. No reallocation would take place in between. Alternatively, the RA could limit itself to a significant change in allocation, implied by a tolerance of the difference between the old and new allocation through the profile change, although in practice, the frequency would probably have to be limited in some way due to the costs involved.

Secondly, the allocation engine combines that profile with real-time market factors to select a weight vector inside the portfolio simplex, typically via mean–variance, Black–Litterman, or machine-learning techniques. Simultaneously, market dynamics are no longer abstracted into fixed moments; rather, a stochastic factor process, potentially a diffusion or regime-switching chain, is introduced, and a second operator yields time- and profile-dependent asset weights. The configuration is intentionally dynamic: upon the occurrence of a top-up, withdrawal, or adjustment to preferences on the part of the user, the profile is updated, and a new target mix is computed. In other words, rebalancing is triggered by events rather than by predefined fixed calendar dates alone. At the grid points, the system then rebalances, achieving a trade-off between tracking error and transaction costs. From this standpoint, robo-advisory can be conceptualised as a stochastic control problem of HJB-type: the platform perpetually directs the combined state of investor profile and portfolio towards the client-specific optimum, whilst adhering to market conditions and cost constraints. This end-to-end formulation permits rigorous propagation of profiling uncertainty through allocation decisions and enables modular interchange of profiling and allocation algorithms for empirical comparison.

The formalisation provides practical advantages that extend beyond a descriptive account of robo-advice. By expressing the workflow as an operator between well-defined spaces and casting rebalancing as an optimisation on an event-driven grid, each recommendation becomes traceable from inputs to portfolio weights. This improves internal transparency, auditability, and dialogue with supervisors because assumptions and parameters are explicitly stated rather than embedded in code or ad hoc rules. The architecture is modular because profiling, allocation and control are separated; components can therefore be upgraded or substituted without re-engineering the platform, which facilitates controlled experimentation, versioning and safe rollbacks. Calibration is rendered systematic: the turnover tolerance and cost kernel can be tuned against observed outcomes with the trade-off between tracking fidelity and trading intensity measured rather than intuited. Since market states and profiles are modelled stochastically, end-to-end backtests and stress tests follow naturally, allowing the propagation of profiling uncertainty through to allocations and supporting stability diagnostics and drift monitoring in production.

Explainability also improves with the model. Weight changes are the solution to a stated objective given the current state, so they can be attributed to profile updates, market moves or cost constraints in a manner consistent with supervisory expectations for trustworthy automation. The same language admits guardrails at optimisation time, such as exposure bounds, diversification floors and turnover caps, which reduces the risk of incoherence between business rules and portfolio construction. Operationally, the explicit state-control loop defines a reproducible execution cycle, clarifies data requirements and supports deterministic re-runs of historical decisions for review and compliance.

Scientifically, the framework links robo-advice to established literature in stochastic control (Cui et al., 2022; Khosravi, 2024; Monoyios, 2010, 2018). Even where closed-form dynamic solutions are unavailable, the one-step problems induced by the tracking-error and cost terms could be applied, permitting analysis of myopic and model-predictive policies and the study of their sub-optimality under stylised dynamics. The explicit treatment of profile dynamics opens research on preference inference, optimal questionnaire design and filtering of latent traits. Machine-learning estimators are cleanly integrated as sub-operators with defined domains and loss functions, which mitigates leakage, aligns validation metrics with their role in decision-making, and permits robust or distributionally aware optimisation that is calibrated to predictive uncertainty rather than heuristic buffers. The same formalism enables systematic fairness and suitability testing by varying profiles and holding market states fixed, with counterfactual explanations framed in the model's primitives.

It must be acknowledged that, while the optimal control problem has been fully articulated, its analytical solution and associated convergence proofs are not provided here. A complete mathematical treatment would demand an extensive excursion into stochastic control theory and viscosity-solution techniques, which lies beyond the present thesis' scope. Instead, the contribution lies in establishing the conceptual and mathematical scaffolding necessary for subsequent in-depth technical treatments and empirical implementations, enabling the implementation of effective policies immediately.

Table 22 provides a summary of the key findings and recommendations for various stakeholders based on the developed model of the robo-advisory process.

Table 20. Key Findings from the Robo-Advisory Model

Key finding	Recommendation	Primary audience
Two-operator design: dynamic profile and allocation conditional on investor/market states.	Build a modular chassis to A/B-test profiling and allocation engines under the same governance.	Providers, Engineers
Event-driven rather than calendar-only rebalancing.	Use event triggers (deposits, horizon change, regime flags) within tolerance bands; log triggers and outcomes.	Providers
Embedded risk guardrails (β/ρ limits, drawdown caps, turnover budgets).	Encode live risk controls and monitor breaches in near-real time; publish periodic risk adherence reports.	Providers, Risk
End-to-end auditability (inputs $\rightarrow$ states $\rightarrow$ trades).	Implement tamper-evident logs and regulator read-only viewers.	Providers, Compliance
Framework is an operational scaffold; analytical solutions are beyond scope.	Run engine bake-offs in a shared simulator; publish cost- and risk-adjusted deltas.	Providers, Researchers
Profiling uncertainty is propagated; robust design avoids brittle weights.	Apply robust optimisation or scenario ensembles to cushion parameter error.	Providers, Engineers
Governance clarity matters for suitability and trust.	Ask for a “Why this portfolio?” view showing how answers mapped to risk limits, rebalancing rules and holdings; request audit trail access in disputes.	Investors/ Clients
Rebalancing and risk controls affect realised outcomes.	Confirm event triggers and guardrails (e.g., max drawdown, beta range) in plain language; escalate unusual changes to a human adviser.	Investors/ Clients

Note. Author’s elaboration: Recommendations derive from the operational robo-advisory model developed in this study. Abbreviations: β = beta (market sensitivity); ρ = correlation; “Primary audience” indicates the stakeholder(s) responsible for implementing the recommendation/action. Compliance, engineers, researchers and risk refer to the roles and functional areas of robo-advisor providers. Implementation should respect applicable regulation and internal suitability processes.

3.3.4. GPT-Based Prototype and Empirical Patterns

This subsection finally discusses the GPT-based prototype and addresses research question RQ4 from Subsection 1.2 by examining implementation choices and the empirical allocation patterns observed.

This study implemented a GPT-4o-based robo-advisor (GPT-based RA) and analysed 100 personalised portfolio recommendations. The GPT-based RA displayed a clear preference for broad, low-cost exposure, with an average asset allocation of 59.4% to equity ETFs and 30.3% to bond ETFs. Alternative assets (real estate, precious metals, and cryptocurrencies) accounted for about 10.3% of assets. Stocks Global ETFs (16.6%) and Stocks DM ETFs (12.95%) received the largest single weights; by contrast, real-estate, sector tilts, and precious-metal allocations were small and infrequent. While this may be interpreted as prudent diversification, the literature also notes that large language models (LLMs) can inherit historical skews from their training corpora, necessitating human calibration to avert systematic under- or over-weighting of certain regions or sectors (Ko & Lee, 2024; Pathak & Bansal, 2024). Equity exposures clustered around 65% and 40%. Over 89% of the recommended portfolios had an equity share of at least 40%. 51% of portfolios held at least 65% equities, while portfolios with less than 50% equities comprised only 28% of the sample.

Initially, 12 allocation-sum errors occurred (90%, 95% or 110% of assets), but these were corrected through additional prompting. The sustainability findings are particularly salient. ESG instruments were never assembled into fully ESG-compliant portfolios (“green” portfolios as advertised by many of the RAs) even when investors expressed sustainable preferences. These patterns situate the GPT-based RA squarely within the emerging literature that documents a shift from deterministic allocators to LLM-enabled advisors able to synthesise unstructured preference cues and market narratives. Prior experiments show that prompting ChatGPT with scenario-specific macros reconfigures portfolios in economically sensible ways and can even outperform passive benchmarks *ex post* (Ko & Lee, 2024; Romanko et al., 2023). Yet, research cautions that output variability rises with ambiguous or contradictory inputs and that guardrails are required to stabilise recommendations (Oehler & Horn, 2024; Zhu et al., 2024). The small but non-trivial rate of allocation-sum errors and the need for prompt-based repairs in this thesis’

deployment align with this warning and reinforce the case for validation layers that automatically enforce accounting consistency before any advice is surfaced. Such consistency can be achieved by implementing an additional automatic correction process step, as in this study, so that the AI corrects itself in case of obvious errors to complete the allocation. It is also unclear why ESG instruments were not exclusively selected when investors prioritised sustainability. A practical implication is to implement “explanation triggers” whenever the system declines a user’s sustainability preference (e.g., “pure ESG portfolio not recommended because X, Y, Z”), alongside auditable evidence that the user’s objectives were faithfully interpreted. Proposals such as an AI Robo-Advice Licence (AIRAL) with bias-control and explainability scorecards would make such assurances verifiable across providers (Kofman, 2025; Pisoni & Díaz-Rodríguez, 2023).

Equity exposure standard deviations ranged between 5.99% and 16.77%, with lower risk levels showing higher variance, indicating more precise allocations with higher risk levels. Equity exposure ranged from 31.82 to 41.52% at the lowest risk level, 67.42-69.22% at the medium level, and 69.44-74.38% at the highest level. Average equity holdings increased from the lowest to the second-highest risk level, then fell again to the highest. Meanwhile, the proportion of cryptocurrencies significantly increased at the highest risk levels. For example, past risk willingness increased from just under 6% to just under 19% of the total portfolio, whereas the equity share fell by approximately 5% from the second-highest to the highest risk level. The crypto exposure was therefore partly financed by shifts from equities and bonds.

This substitution effect is consistent with findings that LLM-enabled advisors reallocate toward higher-volatility alternative assets when prompted with strong risk appetite or shock scenarios, effectively treating these assets as the marginal risk budget once equity exposure is already elevated (Ko & Lee, 2024; Romanko et al., 2023). At the same time, prior work cautions that advice quality is sensitive to contradictory or extreme risk cues and loss-aversion heuristics, which can yield outputs that appear counter-intuitive without an explicit rationale; hence the call for guardrails and explanation triggers to make such cross-asset substitutions intelligible to end-users (Oehler & Horn, 2024; von Walter et al., 2022; Zhu et al., 2023, 2024).

From a governance perspective, the EU's emerging approach to high-risk financial AI and reinforces this need for traceable audit logs that show why an RA reduced equities at the highest risk level and expanded cryptocurrencies so that suitability and client-preference mapping remain transparent (Montagnani et al., 2024). Moreover, because cryptocurrencies carry sustainability externalities, the equity-to-crypto rotation should be surfaced and, where relevant, constrained or justified against investors' ESG preferences (Gómez-Martínez & Medrano-Garcia, 2025). In general, transparent rationales and audit trails are not just cosmetic additions, but core safeguards: users should be able to see why a certain allocation has been made, and which data influenced the recommendation (Crisanto et al., 2024; Lee et al., 2024; Montagnani et al., 2024; Niszczoła & Abbas, 2023). This is also important considering psychological barriers such as reluctance to trust algorithmic decisions and fear of opaque decision-making processes hindering widespread acceptance (Baek & Kim, 2023; Dwork & Minow, 2022; Maier et al., 2022; von Walter et al., 2022). Not least, obstacles to the implementation of the European Commission's *Ethics Guidelines for Trustworthy AI* in financial advice also play a role when working with generative AI in robo-advisory (Brüggen et al., 2025). The diversified, ETF-heavy allocations and risk-contingent crypto substitution mirror a claim by Tahvildari (2025d) in a review that generative AI enables adaptive, real-time portfolio reconfiguration whilst posing challenges regarding explainability and suitability. ESG underemphasis likewise maps to ethical-regulatory gaps, reinforcing calls for licensing and audit trails to secure trust and compliance (Tahvildari, 2025d).

Human-in-the-loop (HITL) operating models, i.e. hybrid advice, can therefore be recommended based on the results of the GPT-based RA implementation. The combination of algorithmic scaling and expert review may produce better and safer outcome than either element alone (Cao et al., 2024; D'Acunto & Rossi, 2021; Khan et al., 2025; Pelster & Val, 2024). For practice, this implies pre-trade checks that verify accounting totals; stress test allocations against scenario libraries; and expose a concise, plain-language "why this portfolio?" accompanied by drill-down justifications. The latter is part of explainable AI (XAI), which is discussed in the literature to resolve the "black box" of incomprehensibility between AI input and output and increase traceability for

users (Barredo Arrieta et al., 2020; Tahvildari, 2025c; Vilone et al., 2024; Weber et al., 2024; Yue et al., 2023).

The equity share was strongly associated with risk-related attitudes. Across univariate and multivariate models, past risk willingness and (the reverse-coded) asset-loss concern exhibited large, positive effects on equity exposure (approximately between +14% and +17% per SD), while investment horizon, liabilities and coverage reserves were more moderate positive drivers. Age and net assets had a moderate negative association with equity exposure. Differences across risk variables, such as asset loss concern and past risk willingness, were minimal, with identical values at low levels and negligible variations at medium and high levels. Therefore, it is sufficient to focus on a risk variable when explaining the equity share recommended by ChatGPT-4o. Based on the model qualities, the best choice is past risk willingness.

The best two-factor model by quality for explaining equity exposure with $\text{adj. } R^2 = 52.83\%$ consisted of a combination of past risk willingness (+14.79% per SD) and coverage reserve (+5.31% per SD). An alternative bilinear model with $\text{adj. } R^2 = 51.21\%$ consisted of age (−12.82% per SD) and liabilities (+8.6%). The addition of net assets to the preferred two-factor model increased the goodness further to $\text{adj. } R^2 = 55.34\%$, with net assets having a moderate negative association (−5.72% per SD). In this trilinear model, the standardised effects of past risk willingness (+17.13%) and coverage reserve (+9.15%) on equity exposure were further amplified. Extending the alternative model by incorporating investment horizon produced a model quality of $\text{adj. } R^2 = 53.7\%$, with age (−8.83% per SD), liabilities (+8.56% per SD), and investment horizon (+5.32% per SD) each contributing to the equity exposure. Ultimately, expanding the primary specification to include net assets becoming a 4-linear model delivered the best fit (adjusted $R^2 = 59.41\%$), with standardised coefficients for past risk willingness (+17.38% per SD), liabilities (+7.07% per SD), net assets (−9.74% per SD), and coverage reserve (+6.96% per SD). No other four-factor model with the alternative variables met the requirements of significance and robustness.

Remarkably, liabilities had a positive moderate impact on equity exposure, while net assets had a negative moderate influence. This serves as a prime example of the previously mentioned issue, which seems counterintuitive without a clear

rationale. This inversion plausibly reflects the GPT-based RA's heuristic mapping of "liability-rich" profiles to longer horizons or "catch-up" return targets, alongside omitted capacity controls (e.g., income, debt-service ratios, collateral) and small-sample noise, rather than genuine risk-capacity logic. It also illustrates how generative systems may reweight contradictory cues and embed corpus biases unless explanation triggers and audit logs make the equity-debt trade-off explicit (Montagnani et al., 2024; Oehler & Horn, 2024; Pathak & Bansal, 2024; Zhu et al., 2023, 2024). The counterintuitive liabilities/net assets coefficients exemplify corpus-driven heuristics and opacity, emphasising the recommendations of Tahvildari (2025d): XAI, bias audits and hybrid human oversight. Accordingly, these coefficients should be interpreted cautiously and paired with bias checks and user-facing rationales that justify any equity increase despite high liabilities. This is consistent with the finding of Blaschke and Kriebel (2021) that "guidance customers" (those with lower literacy and income/assets, but who invest larger sums) require deeper elicitation, suggesting the adoption of adaptive questionnaires and increased oversight for these segments. Scherer and Lehner (2023, 2025) identified risk preference and investment horizon as the most important drivers of equity exposure. Risk preference also plays the most important role in GPT-based RA. The investment horizon has a moderate association with the equity exposure in the trilinear model. Referring to the criticism of previous research (Scherer, 2017; Scherer & Lehner, 2022, 2025; Tertilt & Scholz, 2018) that RAs hardly take financial circumstances such as household balance sheets and income into account, the implementation in ChatGPT shows that household balance sheet indicators such as coverage reserve, liabilities, and net assets are among the driving forces for equity exposure in portfolio recommendations, together with risk preference. This is an indicator that generative AI-based RAs could provide better profiling and more accurate equity recommendations by taking into account additional financial situation indicators of investors, thereby eliminating a previous weakness of conventional investor profiling in RAs.

The implementation shows a practical solution that accounts for the essential variables to determine the investor profile in the portfolio recommendation. In addition, the AIs offer the option of using complicated prompts to incorporate more customising into the profiling and allocation logic to create even more

individual portfolios for investors. Also, there is no longer a necessity for pre-built portfolios that are readily assigned to investors, as is the case with almost all the 45 RAs analysed. Instead, countless combinations can be created through weighting and asset types. The findings of this study also imply that the weaknesses can be addressed through better prompting as instructions on how the answers should be evaluated by the AI in terms of profiling. The evidence from the GPT-based RA implementation also supports the recommendations from prior findings, like Fieberg et al. (2023) or Oehler and Horn (2024). Fieberg et al. (2023) argue that ChatGPT-4 can provide financial advice and formulate investment portfolios aligned with an investor's profile despite limitations like home bias and insensitivities. They also indicate that the historical risk-adjusted performance of ChatGPT-4 is comparable to professional benchmarks, suggesting its utility as a back-office tool for advisors. Oehler and Horn (2024) show that ChatGPT's financial advice closely matches academic benchmarks and is better suited to different investor profiles than most other conventional RAs, indicating ChatGPT's superior potential as a reliable source of financial advice for retail investors.

Therefore, based on the findings of this study, providers of RAs can further develop their RAs by implementing LLMs such as ChatGPT, offer significantly more options for customisation, and move away from deterministic decision-based algorithms and incorporate the advantages of a learning generative AI, while still having the option of teaching the AI a certain level of individualisation by setting constraints to tailor their service.

Table 21 provides a summary of the key findings and recommendations for various stakeholders based on the implementation of the GPT-based robo-advisor.

Table 21. Key Findings from the GPT-Based Robo-Advisor

Key finding	Recommendation	Primary audience
Allocations centre on $\approx 59\%$ equities, $\approx 30\%$ bonds, $\approx 10\%$ alternatives; equity clusters $\approx 40\%$ and 65% .	Pre-define target bands per risk level and provide natural-language rationales for deviations.	Providers
Initial sum-to-100 errors required prompt fixes and	Enforce pre-trade validators (mass-balance, asset-class	Providers, Engineers

other logical checks for profile-to-portfolio mapping	bounds, SRRI/SRI fit); block advice until all checks pass.	
ESG requests not translated into pure-ESG portfolios.	Offer a pure-ESG mode or trigger an explanation (liquidity, costs, tracking error) with the closest-feasible alternative.	Providers, Compliance
Equity share drivers: risk willingness (+), coverage reserves (+); age/ net assets (-); liabilities show counter-intuitive (+).	Run bias audits; require capacity tests before any liability-driven equity increase; route edge cases to HITL review.	Providers, Risk
High-risk prompts substitute crypto for part of equity/bonds.	Cap crypto; require stress-tests and client acknowledgement; surface sustainability externalities.	Providers, Compliance
Outputs sensitive to ambiguous inputs; variance higher at low risk.	Use structured prompts, scenario libraries, and prompt versioning with drift monitoring.	Providers, Engineers
LLM advice is powerful but needs guardrails and explanations.	Treat LLM output as a first pass; verify 100% allocation, SRRI/SRI fit, and the reason for any crypto/ESG choices before accepting.	Investors/ Clients
ESG and high-volatility sleeves require clarity.	Ask for the rationale, caps and stress tests; opt out if these are not provided.	Investors/ Clients

Note. Author's elaboration: Findings and recommendations are based on the GPT-based robo-advisor experiment and composition analysis of 100 recommended portfolios. Abbreviations: ESG = environmental, social, and governance; SRRI/SRI = standard risk/reward indicator or synthetic risk and reward indicator; LLM = large language model; HITL = human-in-the-loop. Recommendations indicate typical responsible stakeholders for implementation/action. Compliance, engineers and risk refer to the roles and functional areas of robo-advisor providers.

**IV – EMPIRICAL
PERFORMANCE ANALYSIS
OF ROBO-ADVISORS IN
GERMANY**

IV -EMPIRICAL PERFORMANCE ANALYSIS OF ROBO-ADVISORS IN GERMANY

4.1. DATA AND METHODS

This section presents the data and methods applied for the performance analysis of robo-advisor portfolios. It consists of five consecutive steps. Subsection 4.1.1 introduces the data sources and a sample of RA portfolios with indexed price data and time periods. Subsequently, Subsection 4.1.2 defines the benchmarks and the corresponding benchmark groups to which the RA portfolios are assigned according to their equity exposures to enable benchmark comparisons between the RA portfolios. Subsection 4.1.3 defines and explains the performance measures used for the analyses.

The data collection in Subsection 4.1.1, the benchmark definitions in Subsection 4.1.2, and the definition of performance indicators in Subsection 4.1.3 are based on Tahvildari (2025a). While Tahvildari (2025a) focuses exclusively on calculating and comparing performance measures, the analyses in this chapter delve much deeper into RA portfolio performance analysis and extend beyond the scope of the aforementioned study.

In Subsection 4.1.4, methods from modern portfolio theory (mean–variance analysis) are employed to compare the return–risk characteristics of RA portfolios with those of their benchmark for each risk class (benchmark group with different equity levels) using sigma–mu diagrams. The Sharpe isolines and the capital market line (CML) are determined, as are the planar distances of the portfolios to the CML, and the empirical efficient frontier is visualised.

Furthermore, Subsection 4.1.5 extends beyond the scope of Tahvildari (2025a) by providing detailed correlation analyses between the CAPM key metrics (alpha and beta) and the correlation between RA and benchmark portfolios. Locally weighted scatterplot smoothing (LOWESS) is employed to visualise and study the behaviour of the key CAPM metrics according to systematic risks in greater detail. This allows for a more detailed performance analysis of the robo-advisors' portfolios.

4.1.1. Selection and Data Collection of the Sample

The results in Section 3.2 revealed a fragmented German RA market, with 45 active RA providers offering more than 405 portfolio strategies as of July 2024. However, given that the data collection for the performance analysis took place in May 2022, followed by data preparation and structuring for the analyses, there were significantly fewer RAs and portfolios for which public price data was available at the time of the data collection phase.

In addition, a common practice among many RAs is the publication of data with a lag of 1–2 months. Consequently, in May 2022, the maximum cutoff for February 28, 2022 (end of month) could only be achieved by ensuring a uniform time interval with complete price data. Moreover, the end of the sample period in February 2022 also marks the beginning of a new bear market cycle triggered by the war in Ukraine in 2022 and the drastic interventions of the US Federal Reserve by raising interest rates and reducing the money supply to fight high inflation, which has had a strong negative impact on the stock and bond markets ever since (Ahmed et al., 2022; Bounvou & Yatié, 2022; Federal Reserve Bank, 2022). This interval end is therefore justified not only on a technical-temporal basis but as well in terms of events and cycles, so as to avoid distorting the results of the analysis.

The selection of the data basis for the study was based on RAs operating in Germany for which historical data was available. Only portfolios of RAs for which monthly price or index data without missing values (no gaps) were publicly available on the websites of the RA providers, exchange-traded securities replicating the RA portfolio, or from providers such as extraETF (Isarvest GmbH, 2023), which document the performance of RA portfolios by means of real money deposits or replication of the ETF compositions of the RA portfolios, to provide normalised time series for performance comparison.

The monthly time basis was chosen for the analysis, as data on a daily or weekly basis is only available for very few RA portfolios. Since most RAs on the German market were only launched in 2016–2017, quarterly or annual data would provide too short a time series with too few data points for the analysis, and analyses during the year would be insufficient or impossible to carry out. The analysis should therefore be as granular as possible, while, for statistical reasons, providing the largest possible data basis is essential.

Given that most RAs were launched in 2016–2017, data on portfolio performance is only available for data collection from 2017 onwards in an acceptable sample size (e.g., more than 50 RA portfolios). For this reason, the start of the analysis period was limited to data from 31 January 2017 (end of month). This enabled data collection from 62 portfolios from 12 RAs.

As the COVID-19 crisis occurred during this period, a second subperiod was defined for the COVID-19 phase starting in December 2019 to measure the performance of the RAs during this downturn phase, with the associated stock market crash in March 2020 and the subsequent recovery phase (Huber et al., 2021; Mazur et al., 2021; Song et al., 2022). Since an increasing number of RAs had entered the German market after January 2017, there was a growing amount of data available on their portfolios. Almost three years later, in December 2019, this provided data on 86 portfolios from 16 RAs, thus giving a larger RA sample for analysis of the COVID-19 period. This justified examining the period beginning with the COVID-19 crisis to evaluate the performance of RAs during the crisis phase (see also RQ6 in Subsection 1.2).

Therefore, two time periods were used for the performance analyses. The first period τ_1 (“total period”) with a length of $|\tau_1| = 61$ months (62 time points) started on 31/01/2017 and ended on 28/02/2022. Only 62 portfolios of the 12 RAs fintego, quirion, easyfolio, Raisin Invest, Oskar, ginmon, Estably, truevest, growney, bevestor, digiVestor, peningar up to were examined in the first period of study, as the historical prices of other RAs, such as, Fidelity, Peaks, Evergreen, and wevest, started 31/01/2017. Although the time series was longer, it encompassed a smaller base of RA portfolios.

The second period, the “COVID-19” period, with a length of $|\tau_2| = 26$ months (27 time points) started on 31/12/2019 and ended on 28/02/2022 and considered 86 portfolios of 16 RAs. The RAs and their portfolios with the data coverage are given in Table 22.

Table 22. Sample of Robo-Advisors and Portfolios for Performance Analysis

Robo-advisor	Portfolios	Start date	End date	Observations
fintego	5	31.12.2013	28.02.2022	99
quirion	10	31.01.2014	28.02.2022	98
easyfolio	3	30.04.2014	28.02.2022	96
Raisin Invest	4	30.09.2014	28.02.2022	90
Oskar	5	31.12.2015	28.02.2022	75
ginmon	10	31.12.2015	28.02.2022	75
Estably	3	31.12.2015	28.02.2022	75
truevest	4	31.01.2016	28.02.2022	74
growney	5	30.04.2016	28.02.2022	71
bevestor	5	31.12.2016	28.02.2022	63
digiVestor	5	31.12.2016	28.02.2022	63
peningar	3	31.01.2017	28.02.2022	61
Fidelity	7	30.08.2018	28.02.2022	42
Peaks	4	31.10.2019	28.02.2022	29
Evergreen	10	31.12.2019	28.02.2022	27
wevest	3	31.12.2019	28.02.2022	27

Note. Author's elaboration: The table lists 16 robo-advisors (RAs) included in the performance analysis and the number of model portfolios per provider, together with the start/end dates of the gap-free monthly price or index series and the number of monthly observations available per provider. Inclusion criteria: publicly available, gap-free monthly series for each portfolio (sourced from RA websites, ETF replication/coverage providers such as and RA provider disclosures); data were collected in May 2022 and uniformly truncated on 28 February 2022 to ensure comparable time windows.

Notably, the extraction of price data over the total period, encompassing 62 points and 62 portfolios from 12 RAs, resulted in the generation of 3,844 data points. During the COVID-19 period, the remaining four RAs, with a total of 27 points in time multiplied by 24 portfolios, yielded an additional 648 data points. Therefore, the total number of data points was found to be 4,492. For some RAs, the data could not always be downloaded automatically via an interface and had to be copied manually. This is primarily due to the fact that certain RAs, for which no exchange-traded price data for the portfolios is available and which only publish the data on their websites, do not present the indexed price data in tabular form, but rather as a visual interactive chart, where each individual data point can only

be viewed manually by clicking on the individual time points on the chart. The process was time-consuming due to the necessity of subsequent double- and triple-checking and quality assurance steps to transfer the figures again to avoid data entry errors.

4.1.2. Benchmark Definitions and Assignments to Benchmark Groups

For the performance analysis, benchmarks were needed as a comparison that can be measured and evaluated with the corresponding RA portfolios (Crane & Crotty, 2018; Cremers et al., 2022; Hunter et al., 2014; Jeffers et al., 2021; Mateus et al., 2019). The benchmark portfolios were composed of an equity and a bond component. For the equity component, the global equity index MSCI All Countries World Index (MSCI ACWI) was used, which consists of developed and emerging equity markets (MSCI, 2022). The costs for a EUR-hedged ETF on the MSCI ACWI were assumed to be 0.47% p.a. (SSGA, 2022). For the bond (fixed income) component, the Vanguard Global Bond Index EUR Hedged Acc fund was used, which tracks the Bloomberg Global Aggregate Float Adjusted and Scaled Index as an accumulating market-weighted index of global government and corporate bonds with a maturity of more than one year (Bloomberg, 2022; Vanguard, 2022).

As equities represent the largest risk component, it was appropriate to define the benchmark allocation based on equity exposure as done by Helms et al. (2022) and previously in Subsections 3.2.2 and 3.2.5. A holdings analysis of each RA portfolio in Table 22 was conducted to determine the asset allocation for mapping it to an appropriate benchmark so that each portfolio was compared within its benchmark group. The RA portfolios were created mainly through passive ETFs investing in equities and bonds and were divided into 10 benchmark groups with integer multiples of 10% equity weightings. The remaining funds of the benchmark portfolios were allocated to the bond component. If the asset ratio exceeded the 3% threshold, a rebalancing was carried out to ensure that the composition of the benchmarks remained stable over time and that there was no distortion of the performance comparison. Accordingly, rebalancing was carried out on average once a year over the total period and represented a trade-off between a robust benchmark and a passive investment. This resulted in 10 benchmark portfolios with the weightings shown in Table 23.

Table 23. Benchmark Portfolio Definitions

Portfolio symbol $\pi^{(i,j)}$	Portfolio name	Target stock exposure	Target bond exposure	Number of rebalances
$\pi^{(0,1)}$	Benchmark 1 (BM1)	10%	90%	2
$\pi^{(0,2)}$	Benchmark 2 (BM2)	20%	80%	4
$\pi^{(0,3)}$	Benchmark 3 (BM3)	30%	70%	7
$\pi^{(0,4)}$	Benchmark 4 (BM4)	40%	60%	6
$\pi^{(0,5)}$	Benchmark 5 (BM5)	50%	50%	8
$\pi^{(0,6)}$	Benchmark 6 (BM6)	60%	40%	6
$\pi^{(0,7)}$	Benchmark 7 (BM7)	70%	30%	4
$\pi^{(0,8)}$	Benchmark 8 (BM8)	80%	20%	3
$\pi^{(0,9)}$	Benchmark 9 (BM9)	90%	10%	2
$\pi^{(0,10)}$	Benchmark 10 (BM10)	100%	0%	0

Note. Author's elaboration: Benchmark portfolios are two-asset reference portfolios composed of an equity component and a bond (fixed-income) component: Equity component: MSCI All Country World Index (MSCI ACWI; developed + emerging markets); Bond component: Vanguard Global Bond Index EUR Hedged Acc,

The definition of the benchmarks and the rebalancing function will now be expressed mathematically. Let $\pi^{(i,j)}$ denote the portfolios, where the first index $i \in \{0,1, \dots, 16\} =: I$ indicates the number of the benchmark ($i = 0$) or RA ($i \neq 0$) and the second index $j \in \{1,2, \dots, n_i\} =: J_i$ the number of the portfolio with the total number n_i of portfolios of the associated benchmark or RA. Let $w_{s,t}^{(i,j)}, w_{b,t}^{(i,j)} \in [0,1]$ denote the equity (stock) and fixed income (bond) weights for a portfolio $\pi_t^{(i,j)}$ at time t , respectively. For each benchmark $j = 1, \dots, 10$ define the target weight in equities and fixed income instruments by

$$\tilde{w}_s^{(0,j)} = 0.1j, \quad \tilde{w}_b^{(0,j)} = 1 - \tilde{w}_s^{(0,j)}. \quad (39)$$

A threshold rebalancing rule is imposed if

$$\delta_j := \left| w_{s,t}^{(0,j)} - \tilde{w}_s^{(0,j)} \right| > 0.03,$$

then the rebalancing functions applies

$$\gamma_{j,\delta} \left(w_{s,t}^{(0,j)} \right) := w_{s,t}^{(0,j)} - \mathbb{1}_{\{\delta > 0.03\}} \left(w_{s,t}^{(0,j)} - \tilde{w}_s^{(0,j)} \right), \quad (40)$$

where $\mathbb{1}$ is the indicator returning 1 if $\delta > 0.03$ and 0 otherwise. The definition of the rebalancing function appears more complicated than it is. In principle, it simply works by selling equity shares in the event of an overweight of 3% and investing

them in the bond portion until the target weighting is reached again, and in the event of an underweight of 3%, buying equity shares, which is financed by selling the complementary bond portion.

Now that the benchmarks have been defined, the next step is to determine how the RA portfolios are assigned to the benchmark groups. The assignment function allocates the portfolios to the benchmark groups according to their risk level based on their equity exposure and enables an appropriate comparison. In other words, each RA portfolio is assigned to the corresponding benchmark with the same target equity allocation based on its equity exposure. This allocation results in 10 disjoint benchmark groups, each containing the benchmark and the RA portfolios for that benchmark.

Following 3.2.2, 11 groups were specified, as it still included the portfolio without equity exposure, i.e., a pure fixed income bond portfolio. However, this defensive benchmark was irrelevant for the analysis as all RA portfolios had a target equity allocation of at least 10%.

Mathematically, this can be defined as follows: For the set of all and benchmark portfolios $\Pi = \{\pi^{(i,j)} \mid i \in I, j \in J_i\}$, each portfolio is assigned to its benchmark via a surjective function

$$\varphi: \Pi \rightarrow \{\pi^{(0,k)} \mid k = 1, \dots, 10\}, \quad \pi^{(i,j)} \mapsto \pi^{(0,k(i,j))}, \quad (41)$$

where $k := k(i,j)$ is determined by the target equity weight $\tilde{w}_s^{(i,j)}$ of the portfolio $\pi^{(i,j)}$ matching the benchmark's target equity weight $\tilde{w}_s^{(0,k)}$. Hence, the maps $\varphi(\pi^{(0,j)}) = \pi^{(0,j)}$, $j = 1, \dots, 10$ of the benchmarks are trivial. This leads to a decomposition of $\Pi = \bigcup_{k=1}^{10} \Pi_k$ in disjoint subsets $\Pi_k := \varphi^{-1}(\pi^{(0,k)})$ referred to as the benchmark group of the benchmark k with corresponding RA portfolios. The difference set $RA_k = \Pi_k \setminus \pi^{(0,k)}$ denotes the RA portfolios to the benchmark group Π_k , excluding the benchmark $\pi^{(0,k)}$ itself.

Table 24 displays the benchmark groups, along with the number of portfolios within each group. The portfolio names have undergone a renaming process and are accompanied by the name of the respective RA with an ascending number based on ascending target equity exposure (where possible) to facilitate ease of reading and assignment, as many RA portfolios have unintuitive names in practice.

Table 24. Definition of Benchmark Groups with Portfolios

Benchmark group	Portfolio names	Portfolios
Π_1	BM1, fintego 1, quirion 1, ginmon 1, bevestor 1, digiVestor 1, Fidelity 1	7
Π_2	BM2, quirion 2, ginmon 2, growney 1, Fidelity 2, Fidelity 3, Evergreen 1, Evergreen 2, wevest 1	9
Π_3	BM3, fintego 2, quirion 3, easyfolio 1, Raisin Invest 1, ginmon 3, growney 2, bevestor 2, digiVestor 2, peningar 1, Peaks 1, Evergreen 3, Evergreen 4	13
Π_4	BM4, quirion 4, ginmon 4, Estably 1, Fidelity 4, Evergreen 5, wevest 2	7
Π_5	BM5, fintego 3, quirion 5, easyfolio 2, Raisin Invest 2, Oskar 1, ginmon 5, growney 3, bevestor 3, digiVestor 3, Peaks 2, Evergreen 6, Evergreen 7	13
Π_6	BM6, quirion 6, Oskar 2, ginmon 6, Estably 2, truevest 2, peningar 2, Fidelity 5, Evergreen 8, Evergreen 9	10
Π_7	BM7, fintego 4, quirion 7, easyfolio 3, Raisin Invest 3, Oskar 3, ginmon 7, growney 4, bevestor 4, digiVestor 4, Peaks 3, Evergreen 10	12
Π_8	BM8, quirion 8, Oskar 4, ginmon 8, truevest 1, truevest 3, truevest 4, peningar 3, Fidelity 6	9
Π_9	BM9, fintego 5, quirion 9, Oskar 5, ginmon 9, bevestor 5, digiVestor 5, Fidelity 7, Peaks 4, wevest 3	10
Π_{10}	BM10, quirion 10, Raisin Invest 4, ginmon 10, Estably 3, growney 5	6

Note. Author's elaboration: The table groups robo-advisor (RA) portfolios into ten benchmark groups $\Pi_1 \dots \Pi_{10}$ for performance comparison. Each group contains the reference benchmark portfolio (BM1–BM10) and the RA portfolios assigned to that benchmark based on holdings analysis and closest target equity exposure. To improve readability, RA portfolios were renamed using the provider's name plus an ascending ordinal (e.g., "fintego 1", "fintego 2") ordered by increasing target equity exposure where possible; original provider labels are retained in the dataset. The Portfolios column reports the number of portfolios in each group including the benchmark (hence total portfolios = 10 benchmarks + 86 RA portfolios = 96). Inclusion criteria: portfolios had gap-free monthly price/index series for the selected sample and were mapped to the integer 10% equity-step benchmark whose target weight most closely matched the portfolio's equity exposure.

4.1.3. Definitions of Performance Measures

As some of the performance measures required a risk-free interest rate r^f the one-month EURIBOR was used due to its consistency with the monthly time series and its suitability for asset pricing models (EMMI, 2022; Ghosh et al., 2020; Spears, 2019). The performance measures introduced here for this benchmarking were also applied by Tahvildari (2025b). The data on the EURIBOR is provided by the financial data service provider Refinitiv (LSEG, 2022). If $S_t^{(i,j)}$ is the price of the portfolio $\pi^{(i,j)}$ at time $1 \leq t \in \tau$ for a monthly interval τ , then the return of the portfolio $\pi^{(i,j)}$ is calculated by

$$R_t^{(i,j)} := \frac{S_t^{(i,j)}}{S_{t-1}^{(i,j)}} - 1. \quad (42)$$

The cumulative return (CR) of the portfolio $\pi^{(i,j)}$ between t_1 and t_2 for $t_2 > t_1$ is

$$CR_{t_1, t_2}^{(i,j)} := \frac{S_{t_2}^{(i,j)}}{S_{t_1}^{(i,j)}} - 1. \quad (43)$$

The annualised return (AR) is

$$AR_\tau^{(i,j)} := \sqrt[|\tau|]{\left(1 + CR_\tau^{(i,j)}\right)^{12}} - 1 = \left(1 + CR_\tau^{(i,j)}\right)^{\frac{12}{|\tau|}} - 1, \quad (44)$$

where $|\tau| = 1 + \max_{t \in \tau}\{\tau\} - \min_{t \in \tau}\{\tau\}$ is the length of the interval τ . The annualised volatility ($AVol$) is

$$AVol_\tau^{(i,j)} := \hat{\sigma}_\tau^{(i,j)} \sqrt{12} \quad (45)$$

with $\hat{\sigma}_\tau^{(i,j)}$ as the empirical standard deviation of (42). The maximum drawdown (MDD) is

$$MDD_\tau^{(i,j)} := \max_{t_2 \in \tau} \left[\min \left\{ 0, \frac{S_{t_2}^{(i,j)}}{\max_{t_2 \geq t_1 \in \tau} S_{t_1}^{(i,j)}} - 1 \right\} \right] \leq 0. \quad (46)$$

The MDD gives the maximum loss in a period τ and represents the scenario of investing at the high point and selling at the low point. The quotient of the CR and the absolute value of MDD is the return-to-drawdown risk (RDR):

$$RDR_\tau^{(i,j)} := - \frac{CR_\tau^{(i,j)}}{MDD_\tau^{(i,j)}}. \quad (47)$$

The RDR is a risk-adjusted performance measure (RAPM), over a longer period of time, in which the CR is set in relation to the MDD . The negative sign comes from

the fact that the MDD is not positive. If the MDD is zero, the RDR is set equal to the CR . While RDR is an asymmetric RAPM, the annualised-return-to-volatility risk ratio (AVR) defines a symmetric RAPM:

$$AVR_{\tau}^{(i,j)} := \frac{AR_{\tau}^{(i,j)}}{AVol_{\tau}^{(i,j)}}. \quad (48)$$

Analogously, AVR is set equal to AR if $AVol$ is zero. The recovery time (RT) since the MDD is calculated by

$$RT_{\tau}^{(i,j)} := \min_{l < t \in \tau} \left\{ t - l \mid S_{l+(t-l)}^{(i,j)} \geq S_h^{(i,j)} \right\}, \quad (49)$$

for $l, h \in \tau$ satisfying the identity

$$MDD_{\tau}^{(i,j)} := \frac{S_h^{(i,j)}}{S_l^{(i,j)}} - 1. \quad (50)$$

The recovery time represents the time until the asset recovers from the maximum drawdown. It is not always possible to calculate the recovery time because if the minimum in (49) is not assumed, i.e. the quantity is empty, then for reasons of calculability, the recovery time is set to the maximum remaining length of the time interval. Another important symmetrical RAPM is the Sharpe ratio (SR) defined and revised by William Sharpe (1998):

$$SR^{(i,j)} := \frac{\overline{ER}^{(i,j)}}{\hat{\sigma}^{(i,j)}} = \frac{\frac{1}{|\tau|} \sum_{t \in \tau} (R_t^{(i,j)} - r_t^f)}{\sqrt{\frac{1}{|\tau| - 1} \sum_{t \in \tau} (ER_t^{(i,j)} - \overline{ER}^{(i,j)})^2}} \quad (51)$$

with the excess return (premium) $ER^{(i,j)} := R_t^{(i,j)} - r_t^f$. Since the empirically calculated Sharpe ratio is a biased estimator that overestimates the actual Sharpe ratio, as proven by Miller and Gehr (1978), an unbiased correction factor was applied for reasons of consistency, resulting in the modified Sharpe ratio:

$$mSR^{(i,j)} := SR^{(i,j)} \cdot \frac{\Gamma\left(\frac{|\tau| - 1}{2}\right)}{\Gamma\left(\frac{|\tau| - 2}{2}\right)} \sqrt{\frac{2}{|\tau| - 1}}. \quad (52)$$

However, even for the shorter COVID-19 period, the sample size of 27 data points results in a bias error of approximately 3% in the original Sharpe ratio, meaning that the effects are minimal. In addition, the CAPM was applied to calculate Jensen's α and β by OLS linear regressions (Fama & French, 2004; Jensen, 1968; Jensen et al., 1972; Sharpe, 1964):

$$R_t^{(i,j)} - r_t^f = \alpha^{(i,j)} + \beta^{(i,j)} (R_t^{(0,k)} - r_t^f) + \varepsilon_t^{(i,j)}, \quad (53)$$

where $\varepsilon^{(i,j)}$ is a discrete white noise process and $R_t^{(0,k)}$ the return of the benchmark portfolio $\pi^{(0,k)}$. The (excess) Treynor ratio (ETR) is computed using the quotient of Jensen's α and β (Hübner, 2005):

$$ETR^{(i,j)} := \frac{\alpha^{(i,j)}}{\beta^{(i,j)}}. \quad (54)$$

So far, the performance indicators for comparison have been defined. The following two subsections describe the methodical procedures for further in-depth analyses based on the performance measures previously provided. These include, on the one hand, a classic mean–variance analysis in a (σ, μ) -diagram with Sharpe isolines, the capital market line and the efficient frontier, and, on the other hand, correlation analyses between CAPM measures with geometric visualisations.

4.1.4. Mean–Variance Analysis Method

Furthermore, a mean–variance analysis was performed, focusing on graphical representation in the classic (μ, σ) -diagram (mean-standard deviation diagram) with Sharpe isolines and the empirical efficient frontier, to illustrate the excess returns of the portfolios and benchmarks together with their standard deviations (Markowitz et al., 2000; Markowitz, 1976; Petters & Dong, 2016; Sharpe & Markowitz, 1989; Steinbach, 2001). For the mean of excess return $\mu := \overline{ER}$ ($\mu^{(i,j)} := \overline{ER}^{(i,j)}$) and its standard deviation $\sigma := \hat{\sigma}(\sigma^{(i,j)} := \hat{\sigma}^{(i,j)})$, a portfolio $\pi^{(i,j)}$ dominates another portfolio $\pi^{(k,l)}$, if

$$(\mu^{(i,j)} \geq \mu^{(k,l)}) \wedge (\sigma^{(i,j)} \geq \sigma^{(k,l)}) \wedge ((\mu^{(i,j)} > \mu^{(k,l)}) \vee \sigma^{(i,j)} > \sigma^{(k,l)}), \quad (55)$$

where the first conditions define the weak dominance in all objectives and the second condition strict improvement in at least one. The empirical efficient frontier can then be defined by the dominant portfolios through arbitrary curve interpolation. The characteristic Sharpe isolines are

$$SR_l = \frac{\mu}{\sigma} \Leftrightarrow \mu(\sigma) = SR_l \sigma, \quad (56)$$

where SR_l defines the slope of the isoline. For $l \in [0,1]$ and the tangent slope SR_{tang} , the l -isoline is $SR_l = l \cdot SR_{tang}$. The capital market line (CML) ($l = 1$) is given by

$$\mu_{CML}(\sigma) = SR_{tang} \sigma \quad (57)$$

with the (100% tangency) Sharpe ratio isoline determined by $SR_{tang} = \mu_{tang}/\sigma_{tang}$. The distance (error) from a portfolio $\pi^{(i,j)}$ in the (σ, μ) -plane to the CML is

$$d^{(i,j)} = \frac{\text{abs}(SR_{tang}\sigma^{(i,j)} - \mu^{(i,j)})}{\sqrt{SR_{tang}^2 + 1}}, \quad (58)$$

where $\text{abs}()$ is the absolute value. The proof is provided in APPENDIX 10. The sum of squared distances (SSD) (sum of squared errors relative to the CML) of a subset of portfolios $K \subseteq \Pi$ is

$$SSD = \sum_{\pi^{(i,j)} \in K} d^{(i,j)^2}, \quad (59)$$

with the root sum of squared distances $RSS = \sqrt{SSD}$ and the root mean squared errors

$$RMSE = \frac{RSS}{\sqrt{|K|}}, \quad (60)$$

where $|K|$ is the cardinality of K (count of portfolios in K).

4.1.5. Correlation Analysis Method

In an additional analysis based on the CAPM ratios, the relationship between α and β and the (Pearson) correlation coefficient of excess returns

$$\rho^{(i,j,k)} = \text{Cov} \frac{(ER^{(i,j)}, ER^{(0,k)})}{\hat{\sigma}^{(i,j)} \hat{\sigma}^{(0,k)}} \quad (61)$$

were also calculated for all RA portfolios $i \neq 0$. It was intended to provide further clarification of any under- or outperformance of RAs according to CAPM. Furthermore, based on a normality test (Shapiro-Wilk test), the normal distributions of $\alpha^{i,j}$, $\beta^{i,j}$, and $\rho^{i,j,k}$ are evaluated (González-Estrada et al., 2022; González-Estrada & Cosmes, 2019; Yap & Sim, 2011). Based on the normality test, in addition to the Pearson correlations between α and β , and, α and ρ , Spearman's rank correlation coefficient

$$\varrho_{\alpha,\eta}^{(i,j)} = \text{Cov} \frac{(\mathcal{R}[\alpha^{(i,j)}], \mathcal{R}[\eta])}{\sigma_{\mathcal{R}[\alpha^{(i,j)]}} \sigma_{\mathcal{R}[\eta]}} \quad (62)$$

were computed for both cases $\eta \in \{\beta^{(i,j)}, \rho^{(i,j,k)}\}$, where $\mathcal{R}$ denotes the rank variables (Abdullah, 1990; Croux & Dehon, 2010; Iman & Conover, 1979).

Spearman's ρ was used as a robust test for monotonic relationships if normality is not given. Moreover, 10,000 iterations of bootstrap resampling were employed to reconstruct the empirical distribution of Spearman's ρ , thereby deriving robust 95% confidence intervals without recourse to stringent distributional assumptions and rendering the estimates resistant to outliers and distributional skew (Chernick et al., 2011; James et al., 2023). To visualise the relationship between α and β , as well as between α and ρ , locally weighted scatterplot smoothing (LOWESS) was applied (Cleveland, 1979; Gijbels & Prosdocimi, 2010). This method fits a locally weighted linear regression at each point, offering sufficient flexibility to capture monotonic trends without overfitting. The default smoothing factor 0.66 was used, meaning that each local regression draws on 66% of the data to strike a balance between capturing genuine curvature and suppressing noise. Additionally, three robustness iterations were applied. After an initial fit, outlying observations receive reduced weight in subsequent passes, enhancing resistance to anomalies. These standard settings are conventional for exploratory visualisation, ensuring that the smoothed curve reliably reflects the underlying relationship between Jensen's α and the chosen metrics β and ρ while remaining robust to extreme values (Cleveland, 1979; Cleveland & Devlin, 1988).

4.2. RESULTS FOR CALCULATED PERFORMANCE MEASURES

This section presents the analysis results for the performance measures defined in Subsection 4.1.3 for the two periods as well as for the RA portfolios and benchmarks described in Subsections 4.1.1 and 4.1.2.

4.2.1. Total Period (31/01/2017–28/02/2022)

Table 25 displays essential descriptive statistics of the calculated performance measures for all benchmark and RA portfolios RA_i per benchmark group Π_i for the total period τ_1 between 31/01/2017 and 28/02/2022. Substantial differences in performance metrics were observed across all performance indicators within the respective benchmark groups. In addition, with one exception, all performance indicators of the RAs in all benchmark groups were, on average, worse than those of the corresponding benchmark. The exception to this was the benchmark group,

designated as Π_1 , where the RA portfolios based on the CAPM ratios (Jensen's α and Treynor ratios ETR) exhibited a marginal improvement in performance, albeit on average. All other performance measures for the RAs within this group lagged behind the BM1 benchmark. Moreover, the minimum and maximum values revealed that certain RA portfolios within specific benchmark groups outperformed the benchmark, as evidenced by a few performance metrics. In summary, the RAs demonstrated underperformance in comparison to the benchmark. Not only did the RAs show significant differences in performance among themselves, but when measured against the benchmark at the individual RA level, no RA consistently managed to maintain its performance across benchmark groups.

Table 25. Descriptive Statistics of Performance Measures (Total Period)

Portfolio statistics		<i>mSR</i>	<i>AR</i>	<i>MDD</i>	<i>RDR</i>	<i>AVR</i>	<i>RT</i>	α	<i>ETR</i>
Benchmark 1		0.6167	1.8366%	-4.35%	2.2287	0.5604	4.00	0.0000%	0.0000%
RA_1	Mean	0.4908	0.6888%	-6.19%	0.7300	0.2152	19.40	0.0217%	0.0316%
	SD	0.1060	0.8624%	1.62%	0.7306	0.2406	8.11	0.0557%	0.0688%
	Min	0.3135	-0.7583%	-9.01%	-0.4214	-0.1871	10.00	-0.0471%	-0.0460%
	Max	0.5696	1.4235%	-5.09%	1.4630	0.3915	30.00	0.0803%	0.1068%
Benchmark 2		0.5779	2.8242%	-5.07%	3.0012	0.7128	6.00	0.0000%	0.0000%
RA_2	Mean	0.4824	1.5674%	-7.13%	1.1612	0.3926	11.33	-0.0315%	-0.0334%
	SD	0.0340	0.4125%	0.65%	0.3378	0.1002	2.31	0.0514%	0.0541%
	Min	0.4479	1.1745%	-7.87%	0.9165	0.3137	10.00	-0.0840%	-0.0867%
	Max	0.5158	1.9971%	-6.67%	1.5466	0.5053	14.00	0.0187%	0.0215%
Benchmark 3		0.5222	3.8848%	-6.77%	3.1590	0.7847	6.00	0.0000%	0.0000%
RA_3	Mean	0.4338	2.5551%	-8.28%	1.7340	0.4936	11.44	-0.0200%	0.0611%
	SD	0.0663	1.0618%	1.63%	0.9635	0.1979	2.01	0.1760%	0.3851%
	Min	0.3228	1.1923%	-11.25%	0.8025	0.2681	10.00	-0.2144%	-0.1916%
	Max	0.5452	4.6851%	-6.08%	4.0332	0.9106	16.00	0.3607%	1.0533%
Benchmark 4		0.4728	4.8272%	-8.85%	3.0586	0.7992	7.00	0.0000%	0.0000%
RA_4	Mean	0.3259	4.0361%	-17.79%	1.4450	0.4643	15.33	-0.2584%	-0.1651%
	SD	0.1248	1.0140%	11.27%	0.4985	0.1766	4.16	0.2879%	0.1389%
	Min	0.1864	3.1716%	-30.79%	0.9451	0.2799	12.00	-0.5806%	-0.3062%
	Max	0.4269	5.1522%	-10.70%	1.9420	0.6318	20.00	-0.0265%	-0.0285%
Benchmark 5		0.4359	5.8779%	-10.74%	3.1380	0.8107	10.00	0.0000%	0.0000%
RA_5	Mean	0.3660	4.5719%	-12.56%	2.0682	0.5863	10.78	-0.0902%	-0.0819%
	SD	0.0525	1.3557%	1.56%	0.7207	0.1520	3.46	0.1094%	0.1093%
	Min	0.2921	2.5982%	-14.62%	1.3462	0.4228	3.00	-0.2455%	-0.2308%
	Max	0.4544	6.6310%	-10.35%	3.3950	0.8657	15.00	0.0712%	0.0753%
Benchmark 6		0.4086	6.8880%	-12.59%	3.2019	0.8139	10.00	0.0000%	0.0000%
RA_6	Mean	0.3254	4.9167%	-14.99%	1.8497	0.5287	14.50	-0.1594%	-0.1538%
	SD	0.0434	2.1079%	2.48%	0.7213	0.1655	9.40	0.1004%	0.0744%
	Min	0.2728	1.8854%	-18.82%	0.6897	0.2491	7.00	-0.3144%	-0.2235%

	Max	0.3900	7.9260%	-11.42%	2.6934	0.7513	33.00	-0.0318%	-0.0320%
Benchmark 7		0.3847	7.9508%	-14.56%	3.2650	0.8102	10.00	0.0000%	0.0000%
RA_7	Mean	0.3252	6.2681%	-16.20%	2.2617	0.6021	10.89	-0.0819%	-0.0750%
	SD	0.0313	1.7123%	1.62%	0.7048	0.1087	3.44	0.1277%	0.1368%
	Min	0.2930	3.6425%	-18.35%	1.5145	0.4666	3.00	-0.2488%	-0.2391%
	Max	0.3848	9.5588%	-13.17%	3.5393	0.7955	15.00	0.1058%	0.1516%
Benchmark 8		0.3715	9.0265%	-16.20%	3.4061	0.8190	10.00	0.0000%	0.0000%
RA_8	Mean	0.2778	4.7415%	-17.58%	1.5474	0.4466	20.29	-0.1888%	-0.2208%
	SD	0.0500	2.2041%	2.66%	0.7374	0.1938	13.74	0.0890%	0.1071%
	Min	0.2011	1.7049%	-20.87%	0.4439	0.1640	10.00	-0.3082%	-0.4016%
	Max	0.3387	7.7530%	-14.69%	2.5890	0.6980	48.00	-0.0922%	-0.0926%
Benchmark 9		0.3587	10.1247%	-17.96%	3.5238	0.8207	10.00	0.0000%	0.0000%
RA_9	Mean	0.2972	7.3756%	-20.02%	2.1887	0.5841	11.83	-0.1272%	-0.1089%
	SD	0.0191	1.6872%	2.78%	0.4892	0.0746	3.31	0.1557%	0.1793%
	Min	0.2654	4.5966%	-22.99%	1.6374	0.4926	6.00	-0.3350%	-0.3195%
	Max	0.3197	9.0052%	-14.88%	2.7966	0.6742	15.00	0.1106%	0.1978%
Benchmark 10		0.3466	11.0204%	-19.67%	3.5662	0.8121	10.00	0.0000%	0.0000%
RA_{10}	Mean	0.2827	8.3662%	-22.74%	2.2385	0.5833	11.60	-0.2040%	-0.2013%
	SD	0.0236	1.1938%	1.66%	0.4347	0.0743	2.51	0.0699%	0.0628%
	Min	0.2607	7.1726%	-24.39%	1.7480	0.5081	8.00	-0.2768%	-0.2611%
	Max	0.3193	9.6997%	-20.90%	2.8355	0.7062	14.00	-0.0980%	-0.0988%

Note. Author's elaboration: This table reports descriptive statistics for performance metrics of benchmark and robo-advisor portfolios within each benchmark group Π_i over the total analysis period τ_1 (31 Jan 2017–28 Feb 2022). For each benchmark group the first row shows the benchmark values; subsequent rows summarise the RA portfolios in that group (Mean, SD, Min, Max). The performance measures are defined in Subsection 4.1.3.

A deeper insight into the performance comparison was provided by examining the portfolios within the individual benchmark groups using performance indicators. To this end, the results were sorted to identify the RA portfolios that outperformed their benchmark.

Table 26 shows how many portfolios of each RA outperformed their benchmark for each performance indicator, stating the benchmark abbreviation, for the entire period. Multiple entries mean that the RA had several portfolios within this benchmark group that outperformed their benchmark. Only RA Raisin Invest managed to outperform the BM5 benchmark in all metrics within the benchmark group Π_5 . Apart from the recovery time, the RA also outperformed BM3 in all other metrics in Π_3 . According to a few metrics, it also outperformed BM7. Only ginmon managed to beat the benchmarks in the low equity groups Π_1 and Π_2 using the CAPM. Bevestor managed to outperform the benchmark in the three groups Π_5 , Π_7 , and Π_9 for some metrics. DigiVestor beat the benchmark in the CAPM key figures in five groups but otherwise only in three benchmark groups when measured using the MDD. Estably's portfolios only beat the benchmark with two

portfolios in terms of *AR* or *RT*. Peningar and truevest, on the other hand, only outperformed the benchmark according to *MDD*. The remaining RAs lagged behind the benchmark in all key figures.

Table 26. Robo-Advisors Outperforming Benchmarks (Total Period)

Robo-advisor	<i>mSR</i>	<i>AR</i>	<i>MDD</i>	<i>RDR</i>	<i>AVR</i>	<i>RT</i>	α	<i>ETR</i>
Raisin Invest	BM3, BM5, BM7	BM3, BM5	BM3, BM5, BM7	BM3, BM5	BM3, BM5	BM5	BM3, BM5, BM7	BM3, BM5, BM7
ginmon							BM1, BM2	BM1, BM2
bevestor		BM5, BM7		BM7		BM5, BM7, BM9	BM7	BM7
quirion			BM3				BM1, BM3	BM1, BM3
digiVestor			BM5, BM7, BM9				BM1, BM3, BM5, BM7, BM9	BM1, BM3, BM5, BM7, BM9
Estably		BM4, BM6				BM6, BM10		
peningar			BM6, BM8					
truevest			2x BM8					

Note. Author's elaboration: For each robo-advisor (RA) and performance metric, the cell lists the benchmark(s) (BM = benchmark portfolio) of the respective benchmark group(s) in which the RA had one or more portfolios that outperformed the benchmark over the total analysis period τ_1 (31 Jan 2017–28 Feb 2022). Outperformance is defined as a more favourable observed value on the metric compared with the benchmark (higher = better for *mSR*, *AR*, *RDR*, *AVR*, *RT*, α , and *ETR*; for *MDD* a less negative drawdown is considered better). Multiple occurrences (e.g., "2x BM8") indicate the number of distinct RA portfolios within that benchmark group that outperformed on the given metric. Blank cells indicate no RA portfolio in that sample outperformed the corresponding benchmark on the metric.

Table 27 summarises the number and proportion of portfolios per RA that outperformed their benchmark. Of the 62 RA portfolios over the entire period, only between 2 (3.2%) and 13 (21%) of the portfolios managed to outperform their benchmark, depending on the performance measure. The RAs performed particularly poorly according to the risk-adjusted performance measures (RAPMs) such as *mSR*, *RDR*, and *AVR*. Merely the CAPM figures α and *ETR* show that some portfolios of five RAs outperformed. The RAs fintego, Oskar, growney, truevest, and easyfolio underperformed their benchmarks on all measures for all their portfolios. The results revealed significantly higher risks (drawdowns, volatilities)

and longer recovery times RT for the RAs in comparison to their benchmarks. The maximum drawdown for almost all portfolios occurred in the COVID-19 crash in March 2020. Thus, the recovery time was measured from this low point. Eight RAs had portfolios that outperformed the benchmark in at least one performance indicator in any benchmark group. This underperformance is particularly evident when RAs have many portfolios

Table 27. Portfolios Outperforming Benchmarks (Total Period)

Robo-advisor	Portfolios	mSR	AR	MDD	RDR	AVR	RT	α	ETR
ginmon	10	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	20% (2)	20% (2)
quirion	10	0% (0)	0% (0)	10% (1)	0% (0)	0% (0)	0% (0)	20% (2)	20% (2)
bevestor	5	0% (0)	40% (2)	0% (0)	20% (1)	0% (0)	60% (3)	20% (1)	20% (1)
digi-Vestor	5	0% (0)	0% (0)	60% (3)	0% (0)	0% (0)	0% (0)	100% (5)	100% (5)
fintego	5	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
Oskar	5	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
growney	5	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
Raisin Invest	4	75% (3)	50% (2)	75% (3)	50% (2)	50% (2)	25% (1)	75% (3)	75% (3)
truevest	4	0% (0)	0% (0)	50% (2)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
Estably	3	0% (0)	67% (2)	0% (0)	0% (0)	0% (0)	67% (2)	0% (0)	0% (0)
peningar	3	0% (0)	0% (0)	67% (2)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
easyfolio	3	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
Total	62	5% (3)	10% (6)	18% (11)	5% (3)	3% (2)	10% (6)	21% (13)	21% (13)

Note. Author's elaboration: For each robo-advisor (RA) the table reports the total number of portfolios examined (column Portfolios) and, for each performance metric, the proportion and raw count of that RA's portfolios that outperformed the corresponding benchmark over the total analysis period τ_1 (31 Jan 2017–28 Feb 2022). Values are shown as percent where (n) is the number of portfolios (from the RA's Portfolios total) with better observed performance than the benchmark on that metric. Outperformance criteria: higher is better for mSR , AR , RDR , AVR , RT , α , and ETR ; for MDD a less negative drawdown is considered better. Zeros indicate no portfolios from that RA outperformed the benchmark on the given performance metric.

Table 28 summarises the results of the number and proportions of portfolios per benchmark group that outperform their benchmark. It supplements the previous results and also shows that RA portfolios in mixed portfolios with an equity component of 30%, 50% and 70%, tended to have a higher probability of outperforming their benchmark, which is also attributable to the fact that most RA portfolios are included in these groups, as they roughly represent the most common risk groups for defensive/risk-averse, balanced, and more aggressive/risk-

seeking investments. Apart from this, the results overall underscored the underperformance of RA in their respective benchmark group.

Table 28. Portfolios Outperforming per Benchmark (Total Period)

Bench- mark	Port- folios	mSR	AR	MDD	RDR	AVR	RT	α	ETR
BM1	5	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	60% (3)	60% (3)
BM2	3	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	33% (1)	33% (1)
BM3	9	11% (1)	11% (1)	22% (2)	11% (1)	11% (1)	0% (0)	33% (3)	33% (3)
BM4	3	0% (0)	33% (1)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
BM5	9	11% (1)	22% (2)	22% (2)	11% (1)	11% (1)	22% (2)	22% (2)	22% (2)
BM6	6	0% (0)	17% (1)	17% (1)	0% (0)	0% (0)	17% (1)	0% (0)	0% (0)
BM7	9	11% (1)	11% (1)	22% (2)	11% (1)	0% (0)	11% (1)	33% (3)	33% (3)
BM8	7	0% (0)	0% (0)	43% (3)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
BM9	6	0% (0)	0% (0)	17% (1)	0% (0)	0% (0)	17% (1)	17% (1)	17% (1)
BM10	5	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	20% (1)	0% (0)	0% (0)
Total	62	5% (3)	10% (6)	18% (11)	5% (3)	3% (2)	10% (6)	21% (13)	21% (13)

Note. Author's elaboration: For each benchmark portfolio (BM1–BM10) the table reports the number of RA portfolios examined in that group (Portfolios) and, for each performance metric, the proportion and raw count of portfolios in the group that outperformed the benchmark over the total analysis period τ_1 (31 Jan 2017–28 Feb 2022). Outperformance is defined as a more favourable observed value than the benchmark (higher = better for mSR, AR, RDR, AVR, RT, α , and ETR; for MDD a less negative drawdown is considered better). Values are shown as percent (n) where n is the number of portfolios in the benchmark group that exceeded the benchmark on the metric. Zeros indicate no portfolios in the group outperformed on that metric.

4.2.2. COVID-19 Period (31/12/2019–28/02/2022)

Table 29 summarises the descriptive statistics of the performance measures of the portfolios for the COVID-19 period from 31 December 2019 to 28 February 2022. On average, the RA portfolios again underperformed their respective benchmarks in all benchmark groups with large differences, except for benchmark group Π_3 , where the RAs outperformed their benchmark slightly on average

according to *ETR*. Minima and maxima showed that there were still some RA portfolios outperforming.

Table 29. Descriptive Statistics of Performance Measures (COVID-19 Period)

Portfolio statistics		<i>mSR</i>	<i>AR</i>	<i>MDD</i>	<i>RDR</i>	<i>AVR</i>	<i>RT</i>	α	<i>ETR</i>
Benchmark 1		0.4484	0.5768%	-4.35%	0.2883	0.1363	4.00	0.0000%	0.0000%
RA_1	Mean	0.3406	-0.6288%	-5.72%	-0.1607	-0.1266	12.83	-0.0206%	-0.0328%
	StDev	0.1073	1.6232%	1.67%	0.4491	0.3218	7.36	0.0739%	0.1061%
	Min	0.1325	-3.6220%	-9.01%	-0.8529	-0.6978	5.00	-0.1305%	-0.2161%
	Max	0.4330	0.7951%	-4.61%	0.2954	0.1788	25.00	0.0931%	0.1073%
Benchmark 2		0.4278	1.7461%	-5.07%	0.7542	0.3366	6.00	0.0000%	0.0000%
RA_2	Mean	0.3766	0.8877%	-6.95%	0.3091	0.1603	10.13	-0.0664%	-0.0204%
	StDev	0.1112	0.4838%	2.63%	0.1939	0.0843	2.17	0.1790%	0.1851%
	Min	0.2458	0.0170%	-11.74%	0.0055	0.0035	6.00	-0.3645%	-0.2295%
	Max	0.5714	1.7148%	-3.34%	0.6588	0.2919	14.00	0.1751%	0.2939%
Benchmark 3		0.4099	3.1742%	-6.77%	1.0352	0.4964	6.00	0.0000%	0.0000%
RA_3	Mean	0.3517	1.6415%	-7.64%	0.5075	0.2571	10.92	-0.0162%	0.3944%
	StDev	0.0703	1.1180%	1.98%	0.3633	0.1744	2.02	0.2267%	1.5747%
	Min	0.2531	-0.5283%	-11.25%	-0.1555	-0.0904	8.00	-0.3224%	-0.2689%
	Max	0.4714	3.2520%	-4.34%	0.9515	0.4774	16.00	0.5750%	5.3750%
Benchmark 4		0.3730	4.1750%	-8.85%	1.0466	0.5345	7.00	0.0000%	0.0000%
RA_4	Mean	0.3016	2.9941%	-13.80%	0.6361	0.3479	12.83	-0.1972%	-0.0894%
	StDev	0.1063	0.7745%	8.90%	0.3185	0.1595	3.82	0.3741%	0.2207%
	Min	0.1224	1.8619%	-30.79%	0.1325	0.0679	10.00	-0.8830%	-0.4073%
	Max	0.4115	3.9956%	-5.51%	0.9988	0.5496	20.00	0.1374%	0.2068%
Benchmark 5		0.3547	5.4889%	-10.74%	1.1433	0.5905	10.00	0.0000%	0.0000%
RA_5	Mean	0.3077	4.0544%	-11.47%	0.8101	0.4137	10.42	-0.0564%	-0.0511%
	StDev	0.0498	2.1805%	2.57%	0.4012	0.1661	3.06	0.1599%	0.1669%
	Min	0.2204	0.7356%	-14.62%	0.1547	0.0913	3.00	-0.3683%	-0.3340%
	Max	0.3722	9.3155%	-6.62%	1.4559	0.6284	15.00	0.1640%	0.1969%
Benchmark 6		0.3406	6.7143%	-12.59%	1.2013	0.6248	10.00	0.0000%	0.0000%
RA_6	Mean	0.2749	4.1008%	-13.64%	0.7193	0.3710	11.33	-0.1706%	-0.1344%
	StDev	0.0544	1.6070%	3.92%	0.3499	0.1467	2.92	0.2509%	0.2055%
	Min	0.1828	0.8669%	-18.82%	0.1375	0.0921	7.00	-0.7189%	-0.4518%
	Max	0.3459	6.0685%	-7.67%	1.2130	0.5620	17.00	0.1281%	0.1907%
Benchmark 7		0.3274	8.0137%	-14.56%	1.2486	0.6478	10.00	0.0000%	0.0000%
RA_7	Mean	0.2801	6.2468%	-15.29%	0.9404	0.4710	10.45	-0.0615%	-0.0640%
	StDev	0.0386	3.1474%	2.74%	0.4766	0.1588	3.30	0.2239%	0.2271%
	Min	0.2193	1.6635%	-18.35%	0.2763	0.1600	3.00	-0.3226%	-0.3138%
	Max	0.3386	14.0014%	-8.67%	1.9678	0.6933	15.00	0.4793%	0.4503%
Benchmark 8		0.3228	9.4009%	-16.20%	1.3270	0.6803	10.00	0.0000%	0.0000%
RA_8	Mean	0.2252	4.0775%	-17.69%	0.5100	0.2923	13.25	-0.2913%	-0.3397%
	StDev	0.0560	3.2917%	2.62%	0.3861	0.2294	2.60	0.1544%	0.2104%
	Min	0.1145	-1.6179%	-20.87%	-0.1787	-0.1290	10.00	-0.6226%	-0.7874%
	Max	0.2834	7.5085%	-14.34%	0.9493	0.5340	18.00	-0.1504%	-0.1498%
Benchmark 9		0.3169	10.7844%	-17.96%	1.3836	0.7006	10.00	0.0000%	0.0000%
RA_9	Mean	0.2749	9.4809%	-19.19%	1.2287	0.5559	10.33	-0.0464%	-0.0550%

	StDev	0.0498	5.2598%	3.14%	0.9857	0.2347	3.87	0.3501%	0.3391%
	Min	0.2164	2.6820%	-22.99%	0.3967	0.2223	4.00	-0.4112%	-0.3900%
	Max	0.3778	21.6364%	-14.34%	3.6863	1.0493	15.00	0.7477%	0.6938%
Benchmark 10		0.3088	11.8477%	-19.67%	1.3960	0.7017	10.00	0.0000%	0.0000%
RA_{10}	Mean	0.2301	7.1600%	-22.74%	0.7245	0.4103	11.60	-0.3665%	-0.3377%
	StDev	0.0454	1.8830%	1.66%	0.2408	0.1413	2.51	0.2984%	0.2100%
	Min	0.1561	4.4934%	-24.39%	0.4096	0.1899	8.00	-0.8905%	-0.6962%
	Max	0.2784	9.8052%	-20.90%	1.0752	0.5810	14.00	-0.1442%	-0.1451%

Note. Author's elaboration: This table reports descriptive statistics for performance metrics of benchmark and robo-advisor portfolios within each benchmark group Π_i over the COVID-19 analysis period τ_2 (31 Dec 2019–28 Feb 2022). For each benchmark group the first row shows the benchmark values; subsequent rows summarise the RA portfolios in that group (Mean, SD, Min, Max). The performance measures are defined in Subsection 4.1.3.

Table 30 shows for the COVID-19 period how many portfolios of the respective RA outperformed their benchmark for each performance measure by mentioning the benchmark abbreviation. Several entries again indicate that there were RAs within this benchmark group that had multiple portfolios that outperformed their benchmark. The RAs wevest and Peaks outperformed BM9 in all key performance indicators of the benchmark group Π_9 . Peaks also outperformed BM7 according to all calculated key figures. bevestor managed to outperform the benchmark in three groups, Π_5 , Π_7 , and Π_9 , in some key figures. Evergreen outperformed many of its benchmark groups according to several performance indicators, with Evergreen represented multiple times overall with 10 portfolios and in several benchmark groups. Quirion performed well in groups Π_1 and Π_3 based on several key figures, while Peaks outperformed Π_5 according to all performance measures except for MDD . Apart from that, although more RA portfolios performed better in the COVID-19 period than in the overall period, most RAs still underperformed their benchmark.

Table 30. Robo-Advisors Outperforming Benchmarks (COVID-19 Period)

Robo-advisor	<i>mSR</i>	<i>AR</i>	<i>MDD</i>	<i>RDR</i>	<i>AVR</i>	<i>RT</i>	α	<i>ETR</i>
quirion		BM1	BM3	BM1	BM1		BM1, BM3	BM1, BM3
wevest	BM4, BM9	BM9	BM9	BM9	BM4, BM9	BM9	BM4, BM9	BM4, BM9
Peaks	BM5, BM7, BM9	BM3, BM5, BM7, BM9	BM7, BM9	BM5, BM7, BM9	BM5, BM7, BM9	BM5, BM7, BM9	BM5, BM7, BM9	BM5, BM7, BM9
bevestor		BM5, BM7, BM9		BM5, BM7	BM5, BM7	BM5, BM7, BM9	BM5, BM7, BM9	BM5, BM7, BM9
Fidelity		BM1						
Raisin Invest			BM3, BM5, BM7				BM3, BM5	BM3, BM5
Evergreen	2x BM2, 2x BM3, BM4, 2x BM5, BM6		2x BM2, 2x BM3, BM4, 2x BM5, 2x BM6, BM7	BM6, BM7			2x BM2, 2x BM3, BM4, 2x BM5, 2x BM6, BM7	2x BM2, 2x BM3, BM4, 2x BM5, 2x BM6, BM7
Oskar							BM5	BM5
ginmon							BM1	BM1
digiVestor			BM5, BM7, BM9					
peningar			BM6, BM8					
Estably						BM6, BM10		
truevest			2x BM8, BM8					

Note. Author's elaboration: For each robo-advisor (RA) and performance metric, the cell lists the benchmark(s) (BM = benchmark portfolio) of the respective benchmark group(s) in which the RA had one or more portfolios that outperformed the benchmark over the COVID-19 analysis period τ_2 (31 Dec 2019–28 Feb 2022). Outperformance is defined as a more favourable observed value on the metric compared with the benchmark (higher = better for *mSR*, *AR*, *RDR*, *AVR*, *RT*, α , and *ETR*; for *MDD* a less negative drawdown is considered better). Multiple occurrences (e.g., "2x BM8") indicate the number of distinct RA portfolios within that benchmark group that outperformed on the given metric. Blank cells indicate no RA portfolio in that sample outperformed the corresponding benchmark on the metric.

The number and proportion of portfolios per RA that outperformed their benchmark during the COVID-19 period are shown in Table 31. Of the 86 RA portfolios analysed, only 8 (9.3%) to 24 (27.9%) outperformed their benchmark, depending on the performance measure, which represented a low proportion. The RAs performed particularly poorly according to the RAPMs, such as *mSR*, *RDR* and *AVR*. Only the CAPM metrics α and *ETR* showed that some portfolios of eight RAs outperformed. Remarkably, Peaks outperformed its benchmarks with all 10

portfolios according to both the CAPM metrics α and ETR as well as MDD . In addition, eight of the Peaks portfolios outperformed their benchmarks according to the modified Sharpe ratio. Furthermore, wevest and Peaks outperformed the benchmark in most key figures with a high percentage. Fidelity, fintego, Oskar, growney, and easyfolio were unable to outperform their benchmark with any of their portfolios in terms of one of the most important performance indicators. The results show significantly higher risks (drawdowns, volatilities) since the low point in March 2020 and longer recovery times RT for the RAs compared to their benchmarks. RAs were therefore unable to limit the risks for investors and needed longer recovery times.

Table 31. Portfolios per RA Outperforming Benchmarks (COVID-19 Period)

Robo-advisor	Portfolios	mSR	AR	MDD	RDR	AVR	RT	α	ETR
quirion	10	0% (0)	10% (1)	10% (1)	10% (1)	10% (1)	0% (0)	20% (2)	20% (2)
Evergreen	10	80% (8)	0% (0)	100% (10)	20% (2)	0% (0)	0% (0)	100% (10)	100% (10)
ginmon	10	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	10% (1)	10% (1)
Fidelity	7	0% (0)	14% (1)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
fintego	5	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
bevestor	5	0% (0)	60% (3)	0% (0)	40% (2)	40% (2)	60% (3)	60% (3)	60% (3)
Oskar	5	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	20% (1)	20% (1)
digi-Vestor	5	0% (0)	0% (0)	60% (3)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
growney	5	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
Peaks	4	75% (3)	100% (4)	50% (2)	75% (3)	75% (3)	75% (3)	75% (3)	75% (3)
Raisin Invest	4	0% (0)	0% (0)	75% (3)	0% (0)	0% (0)	0% (0)	50% (2)	50% (2)
truevest	4	0% (0)	0% (0)	50% (2)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
west	3	67% (2)	33% (1)	33% (1)	33% (1)	67% (2)	33% (1)	67% (2)	67% (2)
peningar	3	0% (0)	0% (0)	67% (2)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
Estably	3	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	67% (2)	0% (0)	0% (0)
easyfolio	3	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
Total	86	15% (13)	12% (10)	28% (24)	10% (9)	9% (8)	10% (9)	28% (24)	28% (24)

Note. Author's elaboration: For each robo-advisor (RA) the table reports the total number of portfolios examined (column Portfolios) and, for each performance metric, the proportion and raw count of that RA's portfolios that

outperformed the corresponding benchmark over the COVID-19 analysis period τ_2 (31 Dec 2019–28 Feb 2022). Values are shown as percent where (n) is the number of portfolios (from the RA's Portfolios total) with better observed performance than the benchmark on that metric. Outperformance criteria: higher is better for mSR, AR, RDR, AVR, RT, α , and ETR; for MDD a less negative drawdown is considered better. Zeros indicate no portfolios from that RA outperformed the benchmark on the given performance metric.

Table 32 summarises the results of the number and proportions of portfolios per benchmark group that outperformed their benchmark. Like the total period, very few RA portfolios managed to beat their benchmark, even if this occurred more frequently, e.g., in benchmark groups Π_3, Π_5, Π_7 , and Π_9 . This was particularly evident in the maximum drawdown in the middle equity-weighted benchmark groups, where around one third of portfolios had a lower drawdown than the benchmark. This also improved some RAPMs, as they limit the risk of loss. However, the recovery time demonstrated that the RA portfolios were unable to recoup their losses as quickly as the benchmarks. On average, the RA portfolios therefore underperformed the benchmark significantly.

Table 32. Portfolios Outperforming per Benchmark (COVID-19 Period)

Bench-mark	Port-folios	mSR	AR	MDD	RDR	AVR	RT	α	ETR
BM1	6	0% (0)	33% (2)	0% (0)	16% (1)	16% (1)	0% (0)	33% (2)	33% (2)
BM2	8	25% (2)	0% (0)	25% (2)	0% (0)	0% (0)	0% (0)	25% (2)	25% (2)
BM3	12	16% (2)	8% (1)	33% (4)	0% (0)	0% (0)	0% (0)	33% (4)	33% (4)
BM4	6	33% (2)	0% (0)	16% (1)	0% (0)	16% (1)	0% (0)	33% (2)	33% (2)
BM5	12	25% (3)	16% (2)	33% (4)	16% (2)	16% (2)	16% (2)	50% (6)	50% (6)
BM6	9	11% (1)	0% (0)	33% (3)	11% (1)	0% (0)	11% (1)	22% (2)	22% (2)
BM7	11	9% (1)	18% (2)	36% (4)	27% (3)	18% (2)	18% (2)	27% (3)	27% (3)
BM8	8	0% (0)	0% (0)	37% (3)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
BM9	9	22% (2)	33% (3)	33% (3)	22% (2)	22% (2)	33% (3)	33% (3)	33% (3)
BM10	5	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	20% (1)	0% (0)	0% (0)
Total	86	15% (13)	12% (10)	28% (24)	10% (9)	9% (8)	10% (9)	28% (24)	28% (24)

Note. Author's elaboration: For each benchmark portfolio (BM1–BM10) the table reports the number of RA portfolios examined in that group (Portfolios) and, for each performance metric, the proportion and raw count of portfolios in the group that outperformed the benchmark over the COVID-19 analysis period τ_2 (31 Dec 2019–28 Feb 2022). Outperformance is defined as a more favourable observed value than the benchmark (higher = better for mSR, AR, RDR, AVR, RT, α , and ETR; for MDD a less negative drawdown is considered better).

Values are shown as percent (n) where n is the number of portfolios in the benchmark group that exceeded the benchmark on the metric. Zeros indicate no portfolios in the group outperformed on that metric.

4.3. MEAN–VARIANCE ANALYSIS WITH SHARPE ISOLINES, CAPITAL MARKET LINE AND EFFICIENT FRONTIER

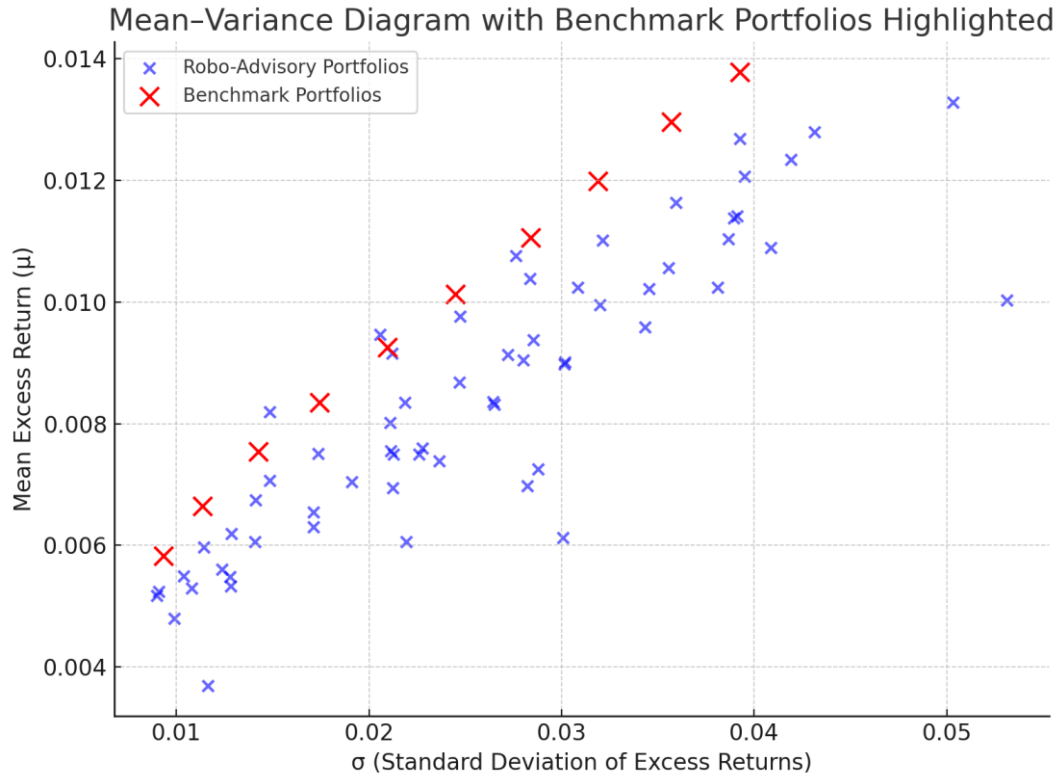
This section provides the results from the mean–variance analysis based on the method described in Subsection 4.1.4.

4.3.1. Total Period (31/01/2017–28/02/2022)

Figure 7 displays the mean–variance diagram in the (σ, μ) -space for all portfolios of the total period, with the benchmark portfolios highlighted in red. Apparently, the portfolios are arranged in an approximately linear relationship to each other. The benchmark portfolios are observed to be positioned approximately on a straight line (upper right section of a hyperbola in (σ, μ) -space, upper right section of a parabola in (σ^2, μ) -space). As illustrated in the diagram, the slope between the benchmark portfolios BM1 and BM2 at $\mu \approx 0.006$ and $\sigma \approx 0.01$ indicates that the tangential portfolio is situated within the region $\mu < 0.006$ and $\sigma < 0.01$.

It is apparent that almost all RA portfolios lie below the line approximately defined by the benchmark portfolios, which illustrates the underperformance of the RAs below the capital market line. Most of the RA portfolios lie below the empirical efficient frontier, which can be illustrated approximately as the upper branch of a hyperbola in the diagram, although the curvature is barely discernible in the diagram due to the subtle differences. This is illustrated in Figure 8. In addition, there are numerous portfolios with extreme return volatilities (high standard deviations) and constant excess returns, which stand out as underperformers. The standard deviations of some portfolios are multiple times higher than those of the benchmark portfolios, which in turn illustrates the mostly very low Sharpe ratios from Subsection 4.2.1. The geometric visualisation in Figure 7 thus clearly shows that the benchmark portfolios essentially dominate the RA portfolios based on the pareto dominance rule (55). Nevertheless, there are a few RA portfolios in a few benchmark groups dominating, as already indicated some performance measures Subsection 4.2.1.

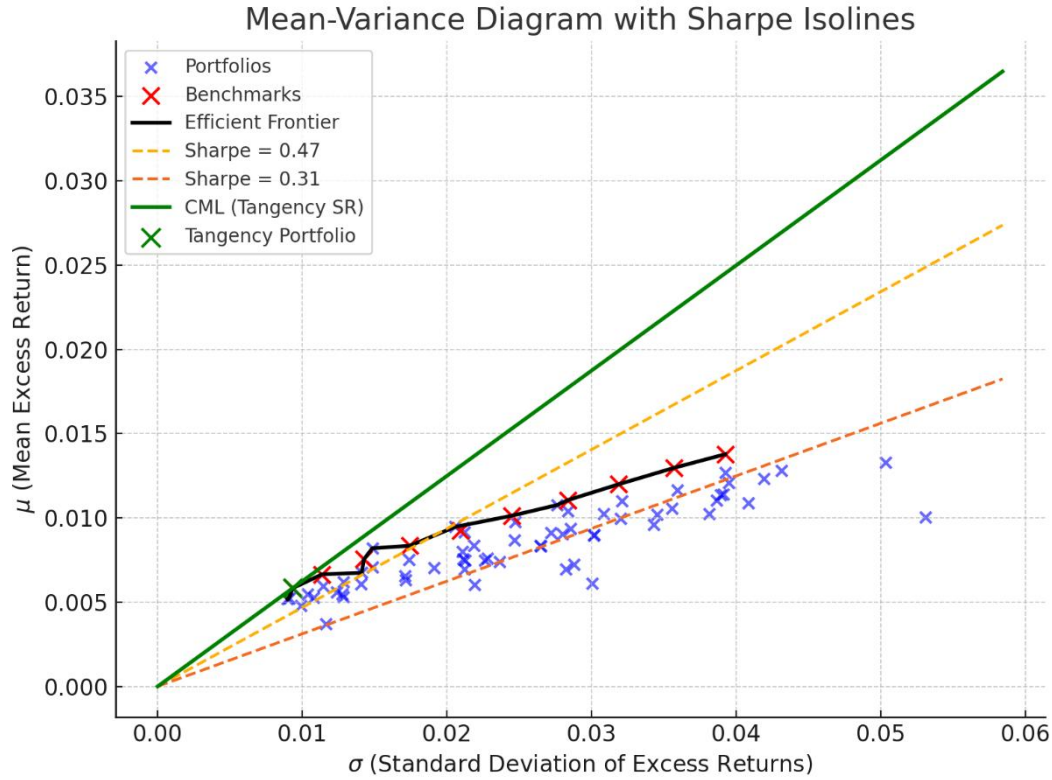
Figure 7. Mean–Variance Diagram of Portfolios (Total Period)



Note. Author's elaboration: Each marker shows a portfolio's mean excess return (μ) against the standard deviation of excess returns (σ) over the total analysis period τ_1 (31 Jan 2017–28 Feb 2022); benchmark portfolios are highlighted in red while robo-advisor portfolios are shown in blue. μ = arithmetic mean of monthly excess returns (portfolio return minus risk-free rate); σ = standard deviation of monthly excess returns.

Figure 8 presents the Sharpe ratio isolines and the capital market line with the tangency portfolio. The “jagged” curve is the empirical efficient frontier, an empirical (concave) hull of a finite set of observed portfolios (a step-wise envelope of discrete data points), rather than the smooth theoretical frontier derived from continuous optimisation in classical mean–variance theory (without a risk-free asset). The tangential portfolio is $(\sigma_{tang}, \mu_{tang}) \approx (0.009325, 0.005824)$ with $SR_{tang} = \mu_{tang}/\sigma_{tang} \approx 0.6245$, which corresponds to BM1. The 50% and 75% Sharpe ratio isolines are given by scaling $\mu_{50\%} = 0.5 \cdot SR_{tang} \approx 0.3123\sigma$ and $\mu_{75\%} = 0.75 \cdot SR_{tang} \approx 0.4684\sigma$, respectively. The CML is identical to the 100% Sharpe ratio isoline $\mu_{CML} = SR_{tang}\sigma$. The CML is thereby defined by the benchmark portfolio BM1.

Figure 8. Mean–Variance Diagram with Sharpe Isolines (Total Period)

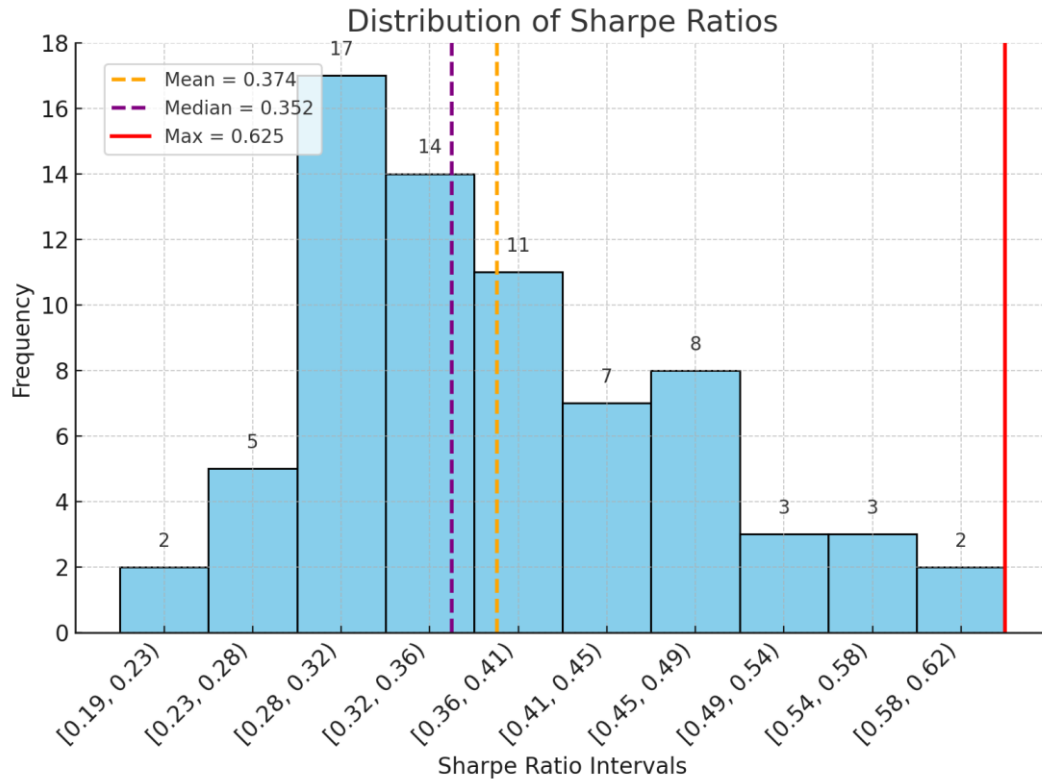


Note. Author's elaboration: The plot shows portfolio mean excess return (μ) against the standard deviation of excess returns (σ) for the total analysis period τ_1 (31 Jan 2017–28 Feb 2022). Blue markers = robo-advisory portfolios; red markers = benchmark portfolios. The black jagged curve is the empirical efficient frontier (concave hull of observed portfolios). Orange dashed lines show Sharpe-ratio isolines at 50% and 75% of the tangency Sharpe ratio; the green line is the capital market line (CML) defined by the tangency portfolio.

Figure 9 shows the frequency distribution of the Sharpe ratios for the RA and benchmark portfolios. The Sharpe ratios were centred around a mean of 0.374 and a slightly lower median of 0.352, indicating a modest rightward tilt in the distribution. The positive skewness of 0.67 suggested that a few higher outliers are pulling the average upwards, while the slightly negative kurtosis of -0.08 reflected a distribution that is slightly flatter and less peaked than a normal curve. Dispersion was moderate, with a standard deviation of 0.095, and the overall range spanned from a minimum of 0.189 to a maximum of 0.625. The diagram in Figure 9 shows the wide range of Sharpe ratios resulting from the performance differences between the RAs. Remarkably, there are portfolios with Sharpe ratios up to 1.67

times the mean. A closer look at the individual Sharpe ratios reveals that 8 out of 10 benchmark portfolios (all except BM9 and BM10) are above average, with BM9 and BM10 only slightly below the overall average.

Figure 9. Distribution of Sharpe Ratios of Portfolios (Total Period)



Note. Author's elaboration: Histogram of realised Sharpe ratios (SR) for the analysed portfolios (robo-advisor and benchmark portfolios) over the total analysis period τ_1 (31 Jan 2017–28 Feb 2022).

Table 33 shows averages, standard deviations, minima and maxima of Sharpe ratios (SR) per RA and the Benchmark. In addition, there are several portfolios from the RAs Raisin Invest, fintego, ginmon, digiVestor and growney that perform better than average in terms of the Sharpe ratio. A total of 30 of 72 portfolios were above the average Sharpe ratio, 22 of which were RA portfolios. Of these, only two RA portfolios in the benchmark groups with more than 50% equity share were able to achieve an above-average Sharpe ratio. These were Oskar in Π_6 and Raisin Invest in Π_7 . Concerning the benchmark, these were the portfolios BM6, BM7 and BM8. This suggests that it is increasingly difficult to exceed the average Sharpe ratio with a higher equity share. This coincides with the law of diminishing marginal returns,

which states that additional risk-taking yields decreased per-unit incremental returns, which is also reflected in the concave-hyperbolic structure in the (σ, μ) -plane (Brue, 1993; Faccarello, 2016; Maiti, 2021; Markowitz, 1976; Mondello, 2023; Sharpe & Markowitz, 1989; Shephard & Färe, 1974).

Table 33. Descriptive Statistics of Sharpe Ratios (Total Period)

Robo-advisor	Average of SR	SD of SR	Min of SR	Max of SR
Benchmark	0.45529	0.09620	0.35105	0.62453
Raisin Invest	0.43133	0.09801	0.32335	0.55214
fintego	0.40488	0.09514	0.30551	0.52930
ginmon	0.39638	0.10141	0.28561	0.57649
digiVestor	0.38895	0.06968	0.31629	0.48496
growney	0.37872	0.08077	0.29216	0.48975
Total	0.37420	0.09534	0.18877	0.62453
Oskar	0.37186	0.04263	0.32375	0.43148
peningar	0.36795	0.04187	0.33602	0.41536
quirion	0.36417	0.10575	0.26654	0.57686
easyfolio	0.32595	0.03697	0.29784	0.36783
bevestor	0.30629	0.01492	0.29455	0.32691
Estably	0.24821	0.05332	0.18877	0.29183
truevest	0.24477	0.03021	0.20366	0.27624

Note. Author's elaboration: Per-provider summary statistics of realised Sharpe ratios (SR) computed over the total analysis period τ_1 (31 Jan 2017–28 Feb 2022). Columns report the sample mean SR, the sample standard deviation of SR, and the minimum and maximum SR observed across the portfolios attributed to each provider.

BM1, as the tangential portfolio, had a Sharpe ratio distance of zero. The average distance tended to increase, which was also observed by the deviation measures SSD, RSS, and RMSE. The RMSE rose from 0.002925 in Π_1 by a factor of approx. 5.5 to 0.016074 in Π_{10} . As the equity exposure in the benchmark groups increased, the Sharpe ratio therefore declined. The larger deviations within the benchmark group Π_4 were due to the small number of portfolios, as one RA portfolio, Estably 1, distorted the result with a relatively large distance. Table 34 shows some descriptive statistics of the distance of the Sharpe ratios of the portfolio to the CML within the (σ, μ) -plane, aggregated per benchmark group.

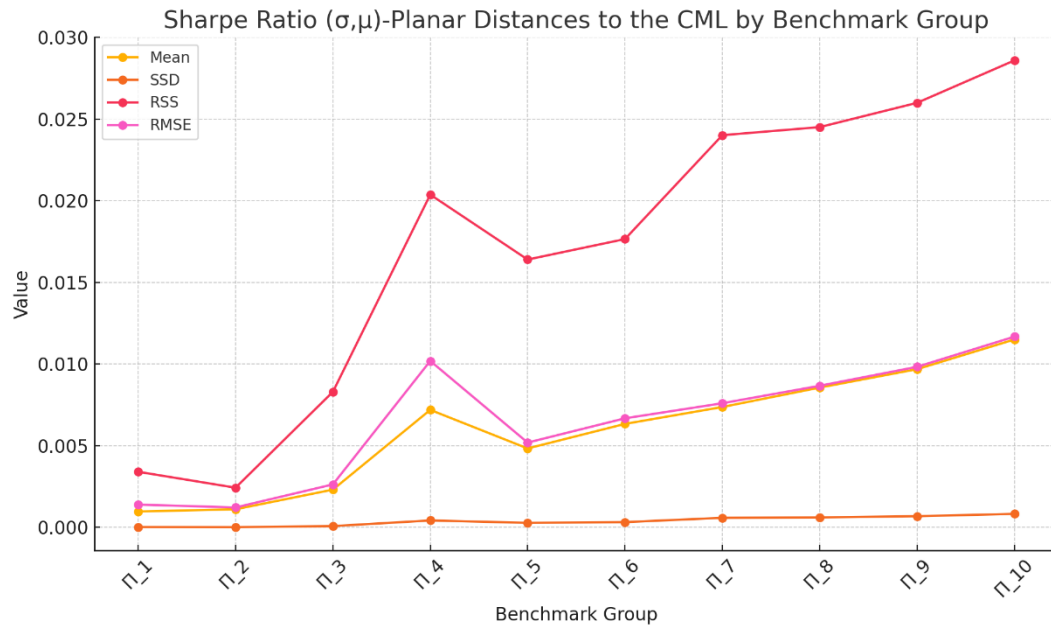
Table 34. Descriptive Statistics of the Distances to the CML (Total Period)

BM Group	Count	Mean	SD	Min	Max	Range	SSD	RSS	RMSE
Π_1	6	0.00096	0.00109	0.00000	0.00303	0.00303	0.00001	0.00340	0.00139
Π_2	4	0.00110	0.00059	0.00038	0.00179	0.00141	0.00001	0.00242	0.00121
Π_3	10	0.00230	0.00132	0.00091	0.00536	0.00445	0.00007	0.00829	0.00262
Π_4	4	0.00719	0.00833	0.00216	0.01963	0.01747	0.00042	0.02037	0.01019
Π_5	10	0.00483	0.00200	0.00287	0.00963	0.00676	0.00027	0.01640	0.00519
Π_6	7	0.00633	0.00228	0.00437	0.01104	0.00666	0.00031	0.01765	0.00667
Π_7	10	0.00737	0.00195	0.00550	0.01199	0.00649	0.00058	0.02402	0.00760
Π_8	8	0.00856	0.00148	0.00666	0.01073	0.00407	0.00060	0.02451	0.00867
Π_9	7	0.00968	0.00181	0.00691	0.01173	0.00482	0.00068	0.02600	0.00983
Π_{10}	6	0.01150	0.00220	0.00910	0.01539	0.00628	0.00082	0.02860	0.01167

Note. Author's elaboration: Descriptive statistics of the distance between each portfolio's realised mean excess-return point (μ, σ) and the capital market line (CML) by benchmark group Π_i , over the total period τ_1 .

The distance metrics are illustrated in the planar plane in Figure 10. Due to slight concavity (average positive second derivative of the distance metrics when outliers in Π_4 are excluded), the Sharpe ratios were on the hyperbolic path already pointed in Figure 7.

Figure 10. Distances to CML per Benchmark Group (Total Period)



Note. Author's elaboration: Plotted metrics per benchmark group Π_i for distances of monthly excess-return (μ, σ) points to the capital market line (CML) for portfolios over the total period τ_1 (31 Jan 2017–28 Feb 2022).

Table 35 shows some descriptive statistics of the distance of the Sharpe ratios of the portfolio to the CML within the (σ, μ) -plane, aggregated per RA. The differences between the average distances between the RAs varied considerably. The higher the equity exposure differences between the portfolios of an RA, the greater the average deviations. This was already shown in Figure 10. Most RAs had significantly more suboptimalities in terms of mean–variance distances than the benchmark.

Table 35. Distance Metrics to the CML by Robo-Advisor (Total Period)

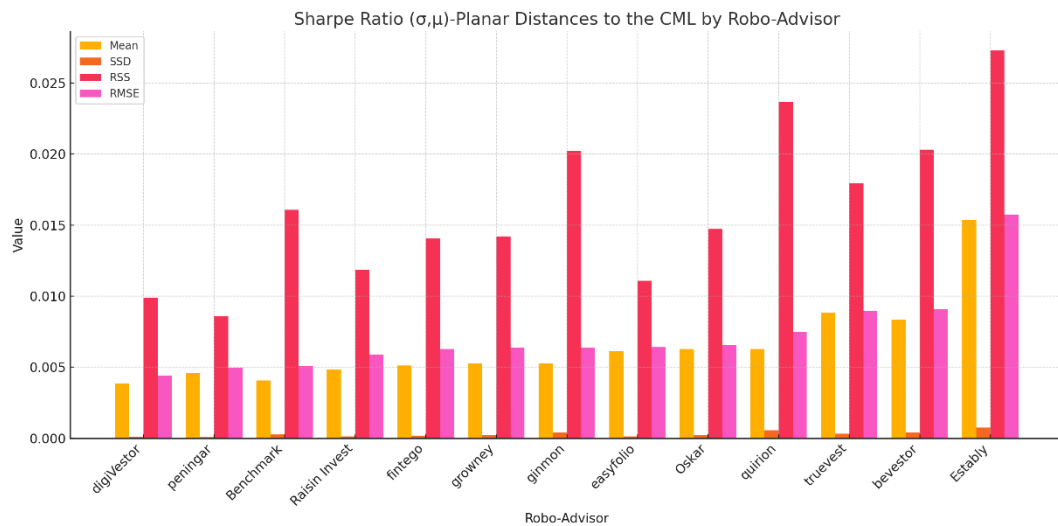
Robo-advisor	Count	Mean	SD	Min	Max	Range	SSD	RSS	RMSE
Benchmark	10	0.00407	0.00321	0.00000	0.00910	0.00910	0.00026	0.01607	0.00508
Estably	3	0.01535	0.00430	0.01104	0.01963	0.00859	0.00074	0.02727	0.01575
Oskar	5	0.00627	0.00225	0.00347	0.00917	0.00569	0.00022	0.01472	0.00659
Raisin Invest	4	0.00483	0.00394	0.00091	0.01003	0.00911	0.00014	0.01182	0.00591
bevestor	5	0.00835	0.00399	0.00303	0.01199	0.00896	0.00041	0.02030	0.00908
digiVestor	5	0.00386	0.00238	0.00117	0.00691	0.00574	0.00010	0.00987	0.00441
easyfolio	3	0.00612	0.00232	0.00373	0.00835	0.00463	0.00012	0.01109	0.00640
fintego	5	0.00511	0.00408	0.00084	0.01069	0.00985	0.00020	0.01404	0.00628
ginmon	10	0.00527	0.00382	0.00037	0.01111	0.01074	0.00041	0.02021	0.00639
growney	5	0.00529	0.00392	0.00124	0.01098	0.00974	0.00020	0.01419	0.00635
peningar	3	0.00461	0.00220	0.00228	0.00666	0.00438	0.00007	0.00858	0.00495
quirion	10	0.00628	0.00430	0.00036	0.01240	0.01204	0.00056	0.02366	0.00748
truevest	4	0.00883	0.00176	0.00648	0.01073	0.00425	0.00032	0.01793	0.00896

Note. Author's elaboration: For each robo-advisor (and the benchmark row) the table reports descriptive statistics of the planar distance between a portfolio's realised point (σ, μ) and the capital market line (CML) defined by the tangency Sharpe ratio over the total period τ_1 (31 Jan 2017–28 Feb 2022).

Figure 11 provides a bar chart sorted by RMSE. DigiVestor had the lowest RMSE, followed by peningar and the benchmark. In terms of mean distance, the benchmark came in second place. For the benchmark with 10 portfolios with different equity exposures ranging from 10% to 100%, the distances were higher because there were significantly more portfolios in the higher equity exposure segments. In contrast, peningar and digiVestor achieved smaller RMSE to the CML, with just three and five portfolios, respectively, and relatively high Sharpe ratios. It should be noted that only two RAs, quirion and ginmon, were represented exactly once in all 10 benchmark groups, thus covering every equity exposure from

10% to 100% in 10% increments. And these two RAs performed significantly worse than the benchmark according to the mean–variance analysis. Consequently, a comparison of the results must be made, considering the number and representation of RA portfolios in the specific benchmark groups. A detailed comparison revealed that all five digiVestor portfolios and three peningar portfolios underperformed the benchmark according to the Sharpe ratios. Taking these results into account, the benchmark achieved better performance than all RAs on average.

Figure 11. Diagram of Distances to the CML by Robo-Advisor (Total Period)



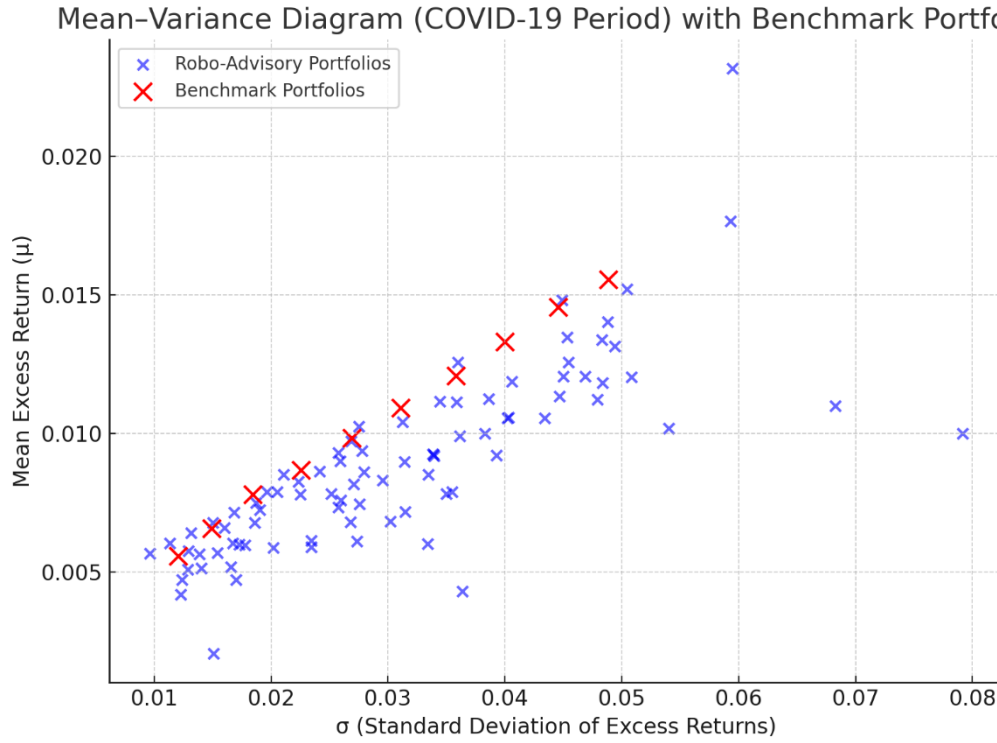
Note. Author's elaboration: Bars show four distance metrics computed per robo-advisor from the monthly excess-return points (σ, μ) over the total period τ_1 (31 Jan 2017–28 Feb 2022). CML = capital market line; metrics defined in 4.1.4.

4.3.2. COVID-19 Period (31/12/2019–28/02/2022)

Figure 12 displays the mean–variance diagram for all portfolios of the COVID-19 period, with the benchmark portfolios highlighted. The portfolios are arranged in an approximately linear relationship to each other. The benchmark portfolios are observed to be positioned approximately on a straight line (upper right section of a hyperbola in (μ, σ) -space). The slope at $(\mu, \sigma) \approx (0.007, 0.015)$ indicates that the tangential portfolio is situated within the region $\mu < 0.007$ and

$\sigma < 0.015$. Here, too, it is apparent that most RA portfolios lie below the line defined by the benchmark.

Figure 12. Mean–Variance Diagram of Portfolios (COVID-19 Period)

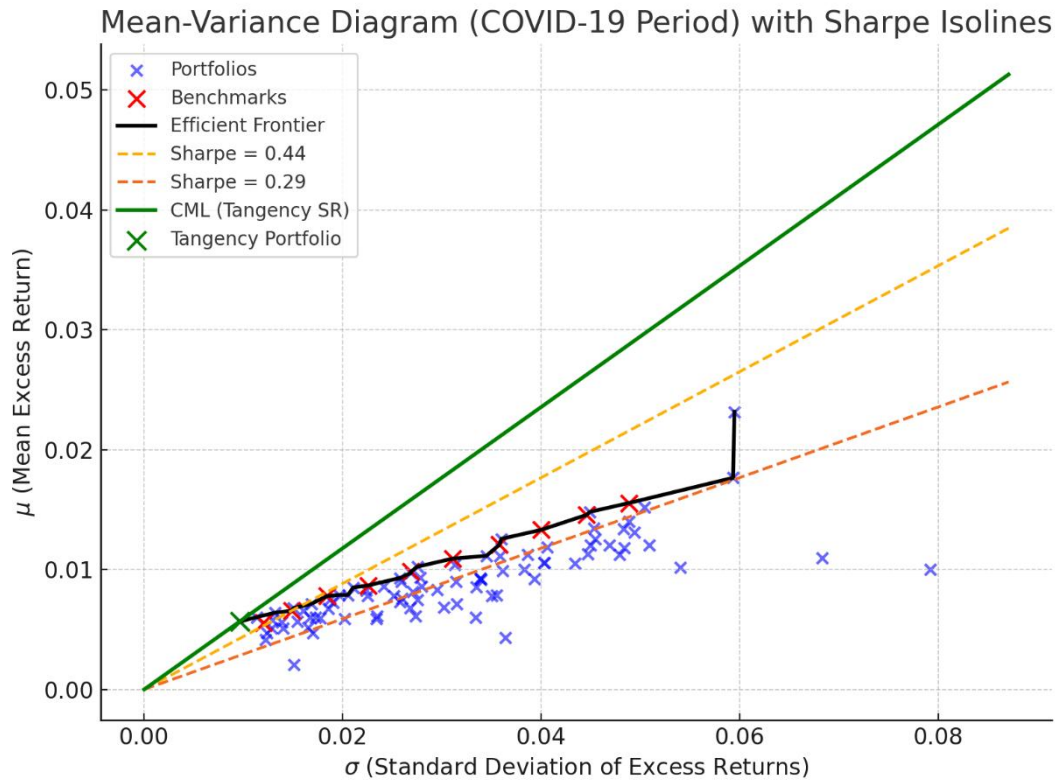


Note. Author's elaboration: Each marker shows a portfolio's mean excess return (μ) against the standard deviation of excess returns (σ) over the COVID-19 analysis period τ_2 (31 Dec 2019–28 Feb 2022); benchmark portfolios are highlighted in red while robo-advisor portfolios are shown in blue. μ = arithmetic mean of monthly excess returns (portfolio return minus risk-free rate); σ = standard deviation of monthly excess returns.

However, there are a few RA portfolios that dominate, as illustrated by their proximity to the efficient frontier in Figure 13, together with the Sharpe isolines, CML and the tangential portfolio. There is one dominant portfolio from wevest that, as an outlier, distorts the empirical efficient frontier as the last point, which is the result of interpolation. The tangential portfolio is given by the point $(\sigma_{tang}, \mu_{tang}) \approx (0.009614, 0.005665)$ with $SR_{tang} = \mu_{tang}/\sigma_{tang} \approx 0.5893$ which corresponds to the maximum Sharpe ratio and is equal to the one of the RA portfolio Evergreen 1. The 50% and 75% Sharpe ratio isolines are given by the means $\mu_{50\%} = 0.5 \cdot SR_{tang} \approx 0.2946\sigma$ and $\mu_{75\%} = 0.75 \cdot SR_{tang} \approx 0.4419\sigma$,

respectively. The CML is identical to the 100% Sharpe ratio isoline $\mu_{CML} = SR_{tang}\sigma$, defined by Evergreen 1.

Figure 13. Mean–Variance Diagram with Sharpe Isolines (COVID-19 Period)

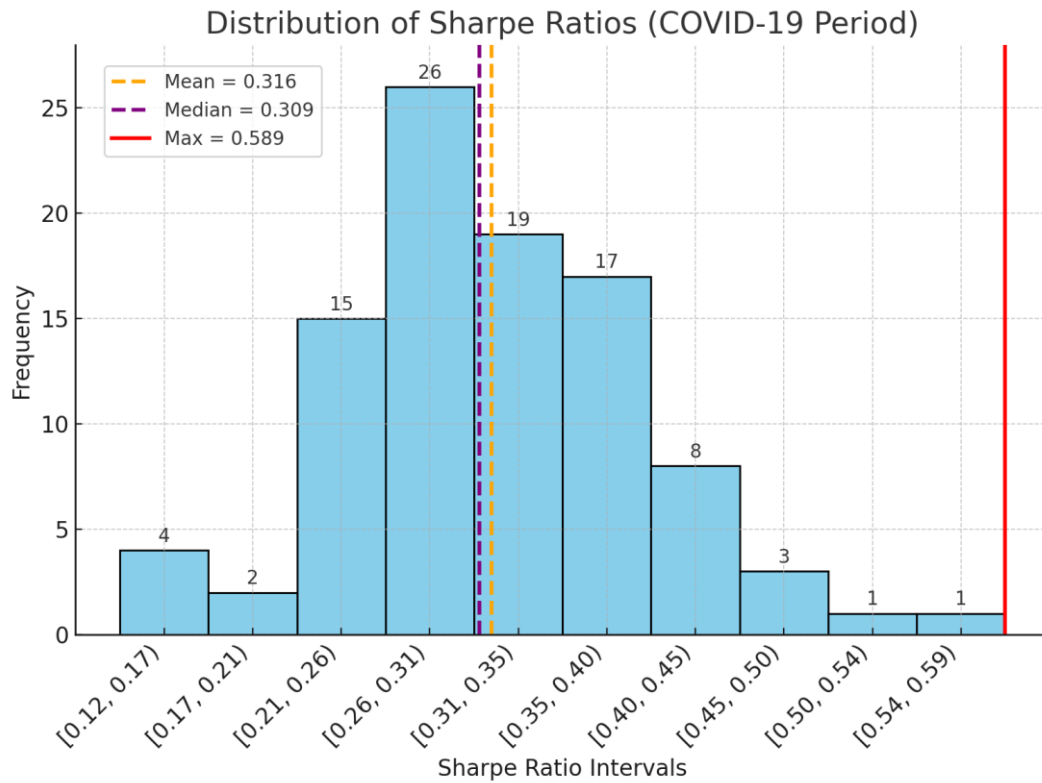


Note. Author's elaboration: The plot shows portfolio mean excess return (μ) against the standard deviation of excess returns (σ) for the COVID-19 analysis period τ_2 (31 Dec 2019–28 Feb 2022). Blue markers = robo-advisory portfolios; red markers = benchmark portfolios. The black jagged curve is the empirical efficient frontier (concave hull of observed portfolios). Orange dashed lines show Sharpe-ratio isolines at 50% and 75% of the tangency Sharpe ratio; the green line is the capital market line (CML) defined by the tangency portfolio.

Figure 14 shows the frequency distribution of the Sharpe ratios for the RA and benchmark portfolios. The Sharpe ratios were centred around a mean of 0.316 and a slightly lower median of 0.309, suggesting a fairly symmetric central tendency. A positive skewness of 0.34 indicated a slight rightward pull, with some portfolios achieved notably higher Sharpe ratios. Meanwhile, a kurtosis of around 0.94 suggested heavier tails than those of a normal distribution. Moderate dispersion was evident with a standard deviation of 0.082, and the spread from minimum to maximum of 0.471 underscored variability across strategies. Table 36 shows averages, standard deviations, minima and maxima of Sharpe ratios (SR)

per RA. In addition, there are several portfolios from the RAs Evergreen, wevest, Peaks, Raisin Invest, and fintego that perform better than average in terms of SR.

Figure 14. Distribution of Sharpe Ratios of Portfolios (COVID-19 Period)



Note. Author's elaboration: Histogram of realised Sharpe ratios (SR) for the analysed portfolios (robo-advisor and benchmark portfolios) over the COVID-19 analysis period τ_2 (31 Dec 2019–28 Feb 2022).

Table 36 shows averages, standard deviations, minima and maxima of Sharpe ratios (SR) per RA and the Benchmark. A total of 45 of 96 portfolios were above the average Sharpe ratio, 35 of which were RA portfolios. Of these, only eight RA portfolios in the benchmark groups with more than 50% equity share were able to achieve an above-average Sharpe ratio. These were Evergreen and Oskar in Π_6 , Peaks, Evergreen, and Raisin Invest in Π_7 , and Peaks and wevest in Π_9 . For the benchmark, these were the portfolios BM6, BM7, BM8, and BM10. Compared to the total period, a few more RA portfolios beat the average Sharpe ratio. Nevertheless, it remains evident that as the equity share increases, the number of portfolios with above-average Sharpe ratios decreases.

Table 36. Descriptive Statistics of Sharpe Ratios (COVID-19 Period)

Robo-advisor	Average of SR	SD of SR	Min of SR	Max of SR
Evergreen	0.42759	0.08573	0.33677	0.58926
wevest	0.38466	0.02223	0.36035	0.40397
Benchmark	0.37440	0.05135	0.31846	0.46243
Peaks	0.36302	0.03045	0.33005	0.40071
Raisin Invest	0.34313	0.04900	0.28710	0.40111
fintego	0.33019	0.05389	0.27671	0.39565
Total	0.31574	0.08221	0.11813	0.58926
Oskar	0.31489	0.03360	0.27674	0.36175
ginmon	0.30827	0.05199	0.24470	0.38093
growney	0.30519	0.04512	0.25730	0.36597
quirion	0.30421	0.07478	0.23447	0.44652
peningar	0.28998	0.02090	0.27217	0.31298
Fidelity	0.28610	0.05468	0.25348	0.40684
digiVestor	0.26389	0.04827	0.22321	0.34078
bevestor	0.26070	0.07045	0.13661	0.30133
easyfolio	0.24080	0.01782	0.22726	0.26098
truevest	0.18578	0.04935	0.11813	0.22301
Estably	0.15857	0.03126	0.12618	0.18857

Note. Author's elaboration: Per-provider summary statistics of realised Sharpe ratios (SR) computed over the COVID-19 analysis period τ_2 (31 Dec 2019–28 Feb 2022). Columns report the sample mean SR, the sample standard deviation of SR, and the minimum and maximum SR observed across the portfolios attributed to each provider.

Table 37 shows some descriptive statistics of the distance of the Sharpe ratios of the portfolio to the CML within the (σ, μ) -plane, aggregated per benchmark group. Evergreen 1 in benchmark group Π_2 , as the tangential portfolio, had a Sharpe ratio distance of zero. The average distance tended to increase, which was also observed by the deviation measures SSD, RSS, and RMSE. The RMSE rose from 0.001387 Π_1 by a factor of approx. 8.4 to 0.011674 in Π_{10} . The larger deviations within the benchmark group Π_4 were due to the small number of portfolios, as one RA portfolio, Estably 1, distorted the result with a relatively large distance. Compared to the overall period, the average distances of the portfolios in the (σ, μ) -plane were significantly higher.

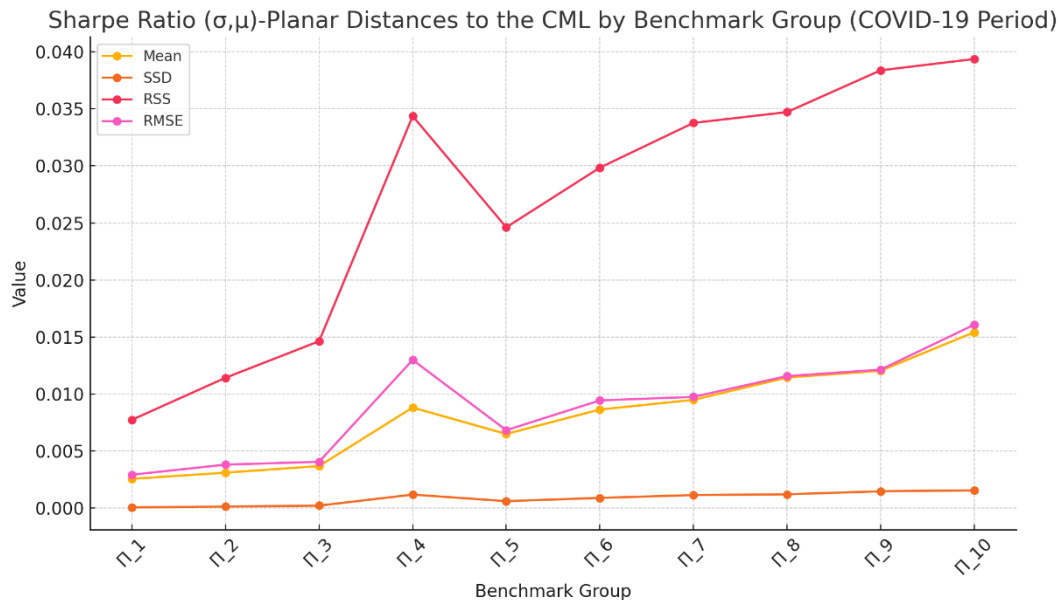
Table 37. Descriptive Statistics of the Distances to the CML (COVID-19 Period)

BM Group	Count	Mean	SD	Min	Max	Range	SSD	RSS	RMSE
Π_1	7	0.00256	0.00152	0.00132	0.00588	0.00456	0.00006	0.00774	0.00293
Π_2	9	0.00310	0.00234	0.00000	0.00775	0.00775	0.00013	0.01143	0.00381
Π_3	13	0.00368	0.00178	0.00117	0.00756	0.00639	0.00021	0.01463	0.00406
Π_4	7	0.00881	0.01031	0.00239	0.03157	0.02918	0.00118	0.03437	0.01299
Π_5	13	0.00650	0.00219	0.00363	0.01139	0.00776	0.00061	0.02462	0.00683
Π_6	10	0.00864	0.00401	0.00484	0.01864	0.01380	0.00089	0.02985	0.00944
Π_7	12	0.00949	0.00234	0.00604	0.01490	0.00885	0.00114	0.03377	0.00975
Π_8	9	0.01144	0.00185	0.00883	0.01477	0.00594	0.00121	0.03472	0.01157
Π_9	10	0.01204	0.00163	0.01002	0.01464	0.00462	0.00147	0.03838	0.01214
Π_{10}	6	0.01542	0.00499	0.01139	0.02520	0.01380	0.00155	0.03937	0.01607

Note. Author's elaboration: Descriptive statistics of the distance between each portfolio's realised mean excess-return point (μ, σ) and the capital market line (CML) by benchmark group Π_i , over the COVID-19 period τ_2 .

The distance metrics are illustrated in Figure 15. Measured by the RMSE, these were between 28% and 218% higher in the COVID-19 period. Due to the high drawdowns, the numerous RA portfolios achieved significantly worse Sharpe ratios.

Figure 15. Distances to CML per Benchmark Group (COVID-19 Period)



Note. Author's elaboration: Plotted metrics per benchmark group for distances of monthly (μ, σ) points to the capital market line (CML) for portfolios over the COVID-19 period τ_2 (31 Dec 2019–28 Feb 2022).

However, as Evergreen's most defensive portfolio had the highest Sharpe ratio, it defined a CML to which all other portfolios on the (σ, μ) -plane had an even greater distance.

Table 38 shows some descriptive statistics of the distance of the Sharpe ratios of the portfolio to the CML within the (σ, μ) -plane, aggregated per RA. The distances of the portfolios to the CML varied significantly depending on the RA, with most RAs performing significantly worse than the benchmark.

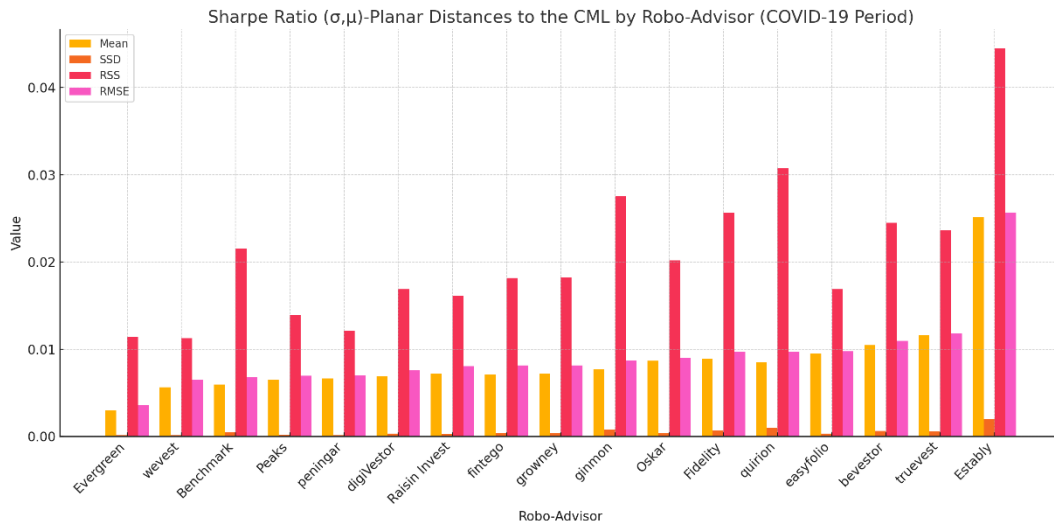
Table 38. Distance Metrics to the CML by Robo-Advisor (COVID-19 Period)

Robo-advisor	Count	Mean	SD	Min	Max	Range	SSD	RSS	RMSE
Benchmark	10	0.00594	0.00352	0.00132	0.01139	0.01008	0.00047	0.02156	0.00682
Estably	3	0.02514	0.00647	0.01864	0.03157	0.01294	0.00198	0.04449	0.02568
Evergreen	10	0.00301	0.00212	0.00000	0.00604	0.00604	0.00013	0.01144	0.00362
Fidelity	7	0.00890	0.00414	0.00217	0.01375	0.01158	0.00066	0.02563	0.00969
Oskar	5	0.00868	0.00275	0.00527	0.01223	0.00697	0.00041	0.02017	0.00902
Peaks	4	0.00645	0.00295	0.00319	0.01002	0.00683	0.00019	0.01387	0.00694
Raisin Invest	4	0.00717	0.00419	0.00303	0.01271	0.00968	0.00026	0.01608	0.00804
bevestor	5	0.01045	0.00368	0.00588	0.01490	0.00902	0.00060	0.02449	0.01095
digiVestor	5	0.00689	0.00344	0.00262	0.01104	0.00842	0.00029	0.01688	0.00755
easyfolio	3	0.00949	0.00271	0.00663	0.01202	0.00539	0.00029	0.01687	0.00974
fintego	5	0.00706	0.00450	0.00215	0.01300	0.01086	0.00033	0.01816	0.00812
ginmon	10	0.00768	0.00433	0.00222	0.01435	0.01213	0.00076	0.02754	0.00871
growney	5	0.00716	0.00431	0.00270	0.01340	0.01071	0.00033	0.01818	0.00813
peningar	3	0.00665	0.00266	0.00393	0.00925	0.00532	0.00015	0.01211	0.00699
quirion	10	0.00847	0.00502	0.00159	0.01545	0.01386	0.00094	0.03072	0.00971
truevest	4	0.01160	0.00252	0.00862	0.01477	0.00614	0.00056	0.02361	0.01180
wevest	3	0.00563	0.00398	0.00330	0.01022	0.00692	0.00013	0.01125	0.00650

Note. Author's elaboration: For each robo-advisor (and the benchmark row) the table reports descriptive statistics of the planar distance between a portfolio's realised point (σ, μ) and the capital market line (CML) defined by the tangency Sharpe ratio over the COVID-19 period τ_2 (31 Dec 2019–28 Feb 2022).

Figure 16 provides a bar chart sorted by RMSE. Only two RAs, Evergreen and wevest, were able to achieve average low distances and higher Sharpe ratios than the benchmark. While wevest only had three portfolios, Evergreen had 10 portfolios, three of which had the three highest Sharpe ratios. In addition, Evergreen's 10 portfolios were in six benchmark groups, i.e., they were represented twice in four benchmark groups and were therefore not represented in the benchmark groups with higher equity exposure of 80%, 90%, and 100%, which tended to have higher distances to the CML, contributed to Evergreen's lower average distances. At portfolio level, Evergreen outperformed the benchmark in all portfolios according to mean-variance comparison. Wevest managed to outperform with two out of three portfolios according to the Sharpe ratio.

Figure 16. Diagram of Distances to the CML by Robo-Advisor (COVID-19 Period)



Note. Author's elaboration: Bars show four distance metrics computed per robo-advisor from the monthly excess-return points (σ, μ) over the COVID-19 period τ_2 (31 Dec 2019–28 Feb 2022); Metrics defined in 4.1.4.

4.4. CORRELATION RESULTS BETWEEN CAPM KEY FIGURES

This section provides the results from the correlation analysis, the method of which is described in Subsection 4.1.5.

4.4.1. Total Period (31/01/2017–28/02/2022)

Regarding the total period, the Shapiro-Wilk test provided the test statistic 0.9691 with a p-value of 0.1191 for the distribution of Jensen's α of the RA portfolios, which means that normal distribution cannot be rejected at a significance level of 5%. For the distribution of the Pearson correlation ρ of the excess returns between RA portfolios and their benchmark, the test statistic was 0.7598 with a p-value $\ll 0.001$, so that normal distribution is rejected as a hypothesis. It should be recalled that the term "excess return" $ER = R - r^f$ in this context denotes the return difference of the RA portfolios in relation to the risk-free interest rate (individual premium), as delineated in (51) under Subsection 4.1.3, as opposed to their outperformance relative to the benchmark.

The Shapiro-Wilk test provided the same result for the distribution of β , with a test statistic of 0.8645 (p-value $\ll 0.001$). Thus, only α was approximately normally distributed, but ρ and β were not. In this case, the Spearman rank correlation ϱ was also calculated. The results are given in Table 39. There was a slight but significant negative correlation between α and ρ . Between α and β there was a strong negative correlation, i.e., higher β tend to lead to lower α .

Table 39. Correlation Tests Between CAPM Metrics (Total Period)

Variable pair	Pearson's r	Pearson's p	Spearman's ϱ	Spearman's p
(α, ρ)	-0.3647	0.00357	-0.2882	0.02314
(α, β)	-0.7115	< 0.00001	-0.5883	< 0.00001

Note. Author's elaboration: Sample: $n = 62$ robo-advisor (RA) portfolios with monthly excess-return series over the total analysis period τ_1 (31 Jan 2017–28 Feb 2022). Columns report Pearson correlation coefficient r with two-sided p-value (Pearson's p) and Spearman rank correlation ϱ with two-sided p-value (Spearman's p). Variable definitions: α = Jensen's alpha (CAPM excess return, in percentage points); ρ = Pearson correlation of the portfolio excess returns with the benchmark excess returns; β = CAPM beta (market sensitivity).

Table 40 shows the results of bootstrapping for the 95% confidence interval (CI) with 10,000 iterations of Spearman's ϱ between Jensen's α and the correlation

ρ of the excess returns of the RA portfolios and the benchmark, as well as between α and β according to the CAPM calculations. A moderate but significant negative relationship was observed between Jensen's α and correlation ρ with Spearman's $\rho = -0.2882$, together with its 95% confidence interval $CI_{\alpha,\rho} = [-0.5371, -0.0041]$, indicating that RA portfolios that are more closely aligned with their benchmark tend to deliver lower risk-adjusted outperformance, measured by α . An even stronger inverse relationship was found for β with Spearman's $\rho = -0.5883$ and 95% confidence interval $CI_{\alpha,\beta} = [-0.7587, -0.3623]$, suggesting that higher systematic risk β exposure generally results in lower Jensen's α .

Table 40. Bootstrapped Spearman's ρ of the CAPM Metrics (Total Period)

Variable pair	Spearman's ρ	Confidence interval (CI)	
		2.5% boundary	97.5% boundary
(α, ρ)	-0.2882	-0.5371	-0.0041
(α, β)	-0.5883	-0.7587	-0.3623

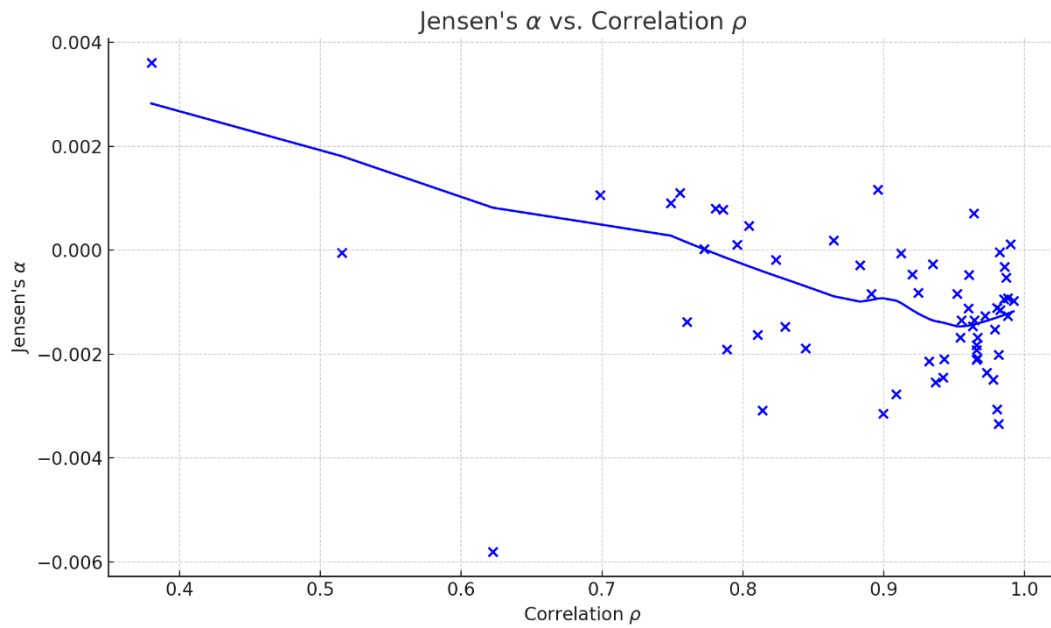
Note. Author's elaboration Spearman rank correlations ρ between Jensen's α and (i) the Pearson correlation ρ of portfolio excess returns with their benchmark excess returns, and (ii) CAPM β , with 95% confidence intervals (CIs) obtained by nonparametric bootstrap (10,000 iterations, percentile method). Sample: $n = 62$ robo-advisor (RA) portfolios with monthly excess-return series over the total analysis period τ_1 (31 Jan 2017–28 Feb 2022). Reported CI bounds are the 2.5% and 97.5% percentile limits from the bootstrap distribution. Variable definitions: α = Jensen's alpha (CAPM excess return, percentage points); ρ = Pearson correlation of portfolio and benchmark excess returns; β = CAPM beta (systematic market sensitivity).

The locally weighted scatterplot smoothing (LOWESS) fit in the Jensen's α versus correlation ρ plot in Figure 17 displays a gentle downward slope, which mirrors the more moderate negative Spearman's rank correlation $\rho \approx -0.29$ and suggests that minor increases in correlation result in a modest decrease in α . The distribution ρ had a high central tendency, with a mean of 0.892 and a median of 0.943, emphasising that most RA portfolios closely tracked their benchmarks.

The standard deviation of 0.120 suggested moderate dispersion around the mean, while the range spanned from 0.380 to 0.992, illustrating occasional occurrence of portfolios with relatively low tracking, which can also be seen in Figure 17. The strong positive kurtosis (approx. 5.52) implied a leptokurtic distribution with heavier tails and a sharper peak than normal. The negative

skewness (approx. -2.10) highlighted a left-skewed distribution with a longer tail of lower correlation values.

Figure 17. Jensen's α vs. ρ of CAPM with LOWESS (Total Period)



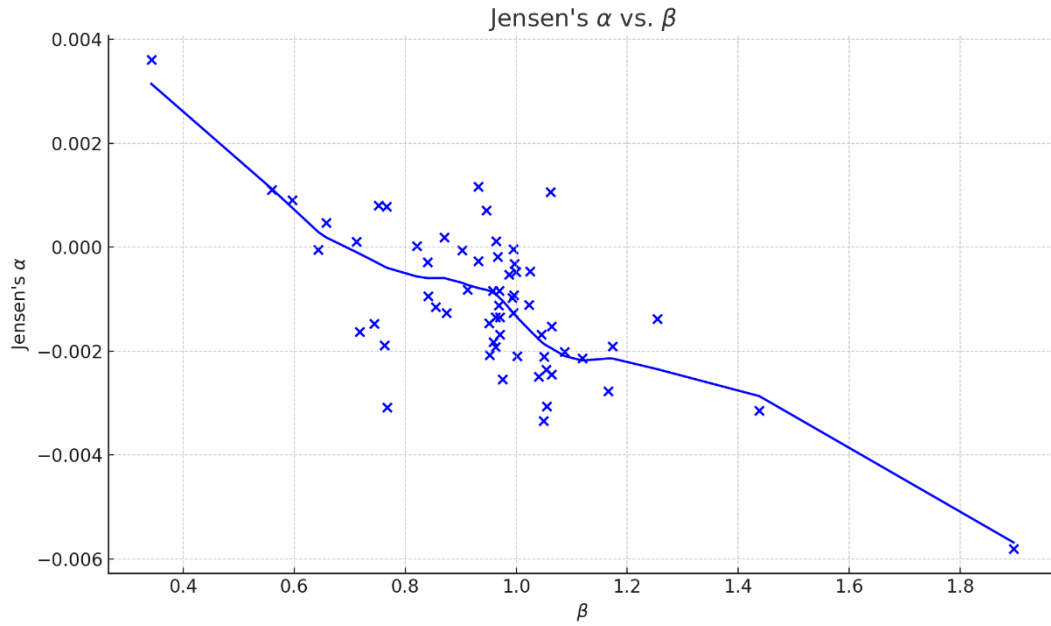
Note. Author's elaboration: Scatterplot of Jensen's α (vertical axis; CAPM excess return in proportion units) against the Pearson correlation ρ of portfolio excess returns with the corresponding benchmark excess returns (horizontal axis) for $n = 62$ robo-advisor (RA) portfolios over the total period τ_1 (31 Jan 2017–28 Feb 2022). A locally weighted scatterplot smoothing (LOWESS) curve is overlaid to visualise the conditional trend.

Similarly, the Jensen's α versus β plot in Figure 18, the LOWESS curve descends approximately linearly across most of the β -range, thereby reaffirming the strong negative Spearman correlation ($\rho \approx -0.59$). Minor deviations from strict linearity at the extremes suggest the possibility of nonlinear effects, yet the overall downward slope remains clear.

The distribution of β centred around a mean of 0.95 and a median of 0.967. These results again showed that most RA portfolios closely tracked their benchmark while having slightly less systematic risk relative to the benchmark. Moderate variability is reflected by a standard deviation of 0.210, with values ranging from 0.342 to 1.896. The high kurtosis (approx. 6.92) also denoted a leptokurtic distribution, while the positive skewness (approx. 1.07) indicated a right-skewed shape driven by a few portfolios with unusually high market sensitivity. Consequently, correlation ρ and systematic risk β implied a high

dependence on the benchmark, which was accompanied by increasingly underperformance measured by α .

Figure 18. Jensen's α vs. β of CAPM with LOWESS (Total Period)



Note. Author's elaboration: Scatterplot of Jensen's α (vertical axis; CAPM excess return in proportion units) against CAPM β (horizontal axis) for $n = 62$ robo-advisor (RA) portfolios over the total analysis period τ_1 (31 Jan 2017–28 Feb 2022). A locally weighted scatterplot smoothing (LOWESS) curve is overlaid to visualise the conditional trend.

4.4.2. COVID-19 Period (31/12/2019–28/02/2022)

Concerning the COVID-19 period, the Shapiro-Wilk test provided the test statistic 0.9242 with a p-value $\ll 0.001$ for the distribution of Jensen's α of the RA portfolios, which means that normal distribution is rejected at a significance level of 5%. For the distribution of the Pearson correlation ρ of the excess returns between RA portfolios and their benchmark, the test statistic was 0.6371 with a p-value $\ll 0.001$, so that normal distribution is also rejected as a hypothesis. The Shapiro-Wilk test provided the same result for the distribution of β , with a test statistic of 0.8704 (p-value $\ll 0.001$). Hence, α , ρ , and β were not normally distributed. In this case, the Spearman rank correlation ϱ was also calculated. The results are given in Table 41. There was a slight but significant negative correlation

between α and ρ . Between α and β there was a strong negative correlation, i.e., higher β tend to lead to lower α .

Table 41. Correlation Tests Between CAPM Metrics (COVID-19 Period)

Variable	Pearson's r	Pearson's p	Spearman's ρ	Spearman's p
(α, ρ)	-0.3034	0.0045	-0.2542	0.0182
(α, β)	-0.5710	< 0.0001	-0.4632	< 0.0001

Note. Author's elaboration: Sample: $n = 86$ robo-advisor (RA) portfolios with monthly excess-return series over the COVID-19 analysis period τ_2 (31 Dec 2019–28 Feb 2022); Columns report Pearson correlation coefficient r with two-sided p -value (Pearson's p) and Spearman rank correlation ρ with two-sided p -value (Spearman's p). Variable definitions: α = Jensen's alpha (CAPM excess return, in percentage points); ρ = Pearson correlation of the portfolio excess returns with the benchmark excess returns; β = CAPM beta (market sensitivity).

Table 42 shows the results of bootstrapping for the 95% confidence interval (CI) with 10,000 iterations of Spearman's ρ between Jensen's α and the correlation ρ of the excess returns of the RA portfolios to the benchmark, as well as between α and β according to the CAPM calculations. A moderate but significant negative relationship was observed between Jensen's α and correlation ρ with Spearman's $\rho = -0.2542$, together with its 95% confidence interval $CI_{\alpha, \rho} = [-0.4539, -0.0354]$, indicating that RA portfolios that are more closely aligned with their benchmark tend to deliver lower risk-adjusted outperformance, measured by α . An even stronger inverse relationship was found for β with Spearman's $\rho = -0.4632$ and 95% confidence interval $CI_{\alpha, \beta} = [-0.6534, -0.2391]$, suggesting that higher systematic risk β exposure generally results in lower Jensen's α .

Table 42. Bootstrapped Spearman's ρ of the CAPM Metrics (COVID-19 Period)

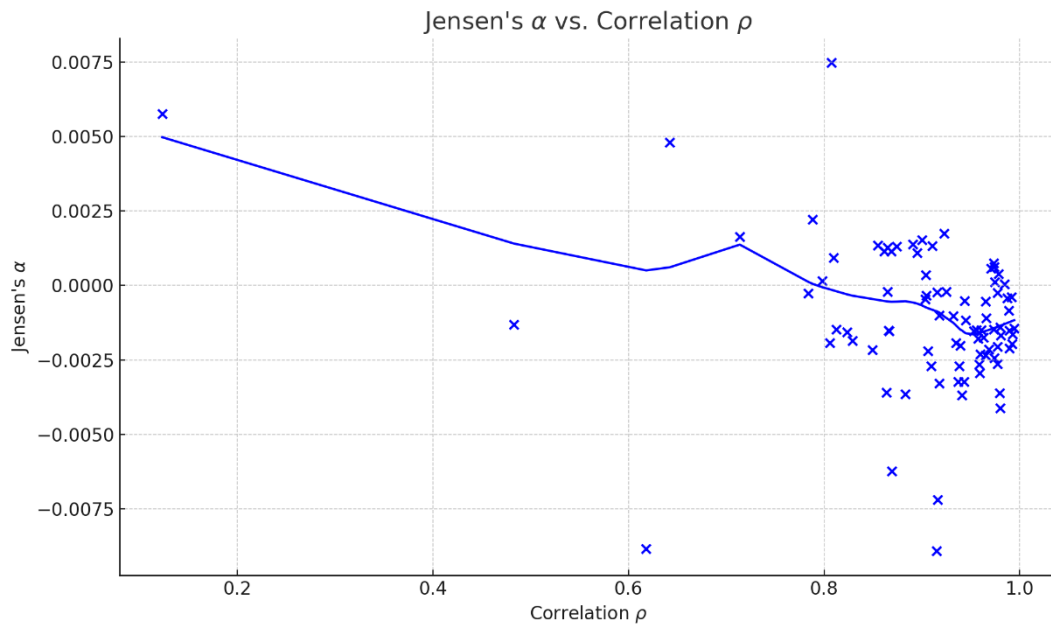
Variable	Spearman's ρ	Confidence interval (CI)	
		2.5% boundary	97.5% boundary
(α, ρ)	-0.2542	-0.4539	-0.0354
(α, β)	-0.4632	-0.6534	-0.2391

Note. Author's elaboration: Spearman rank correlations ρ between Jensen's α and (i) the Pearson correlation ρ of portfolio excess returns with their benchmark excess returns, and (ii) CAPM β , with 95% confidence intervals (CIs) obtained by nonparametric bootstrap (10,000 iterations, percentile method). Sample: $n = 86$ robo-advisor (RA) portfolios with monthly excess-return series over the COVID-19 analysis period τ_2 (31 Dec 2019–28 Feb 2022). Reported CI bounds are the 2.5% and 97.5% percentile limits from the bootstrap distribution. Variable definitions: α = Jensen's alpha (CAPM excess return, percentage points); ρ = Pearson correlation of portfolio and benchmark excess returns; β = CAPM beta (systematic market sensitivity).

In the plot of Figure 19 of Jensen's α against correlation ρ , the LOWESS line slopes downward from left to right, thus pointing to a moderate inverse association, as indicated by Spearman's $\rho \approx -0.25$. Like the analysis of the total period, portfolios that are more tightly coupled to the benchmark tend to exhibit lower α values. It is evident that minor deviations, such as a slight uptick around $\rho \approx 0.75$, are due to non-linear subtleties.

On average, the correlations ρ between RA portfolios and their benchmarks were very high with a mean of 0.899, and a median of 0.933. The standard deviation of 0.122 reflected moderate dispersion, though the range spanned from 0.123 to nearly perfect alignment at 0.994. The distribution was markedly leptokurtic (approx. 19.89), indicating exceptionally heavy tails and a sharp central peak. It was also strongly left-skewed (approx. -3.80), showing that there were more high correlation values than low ones, with a few extreme outliers.

Figure 19. Jensen's α vs. ρ of CAPM with LOWESS (COVID-19 Period)

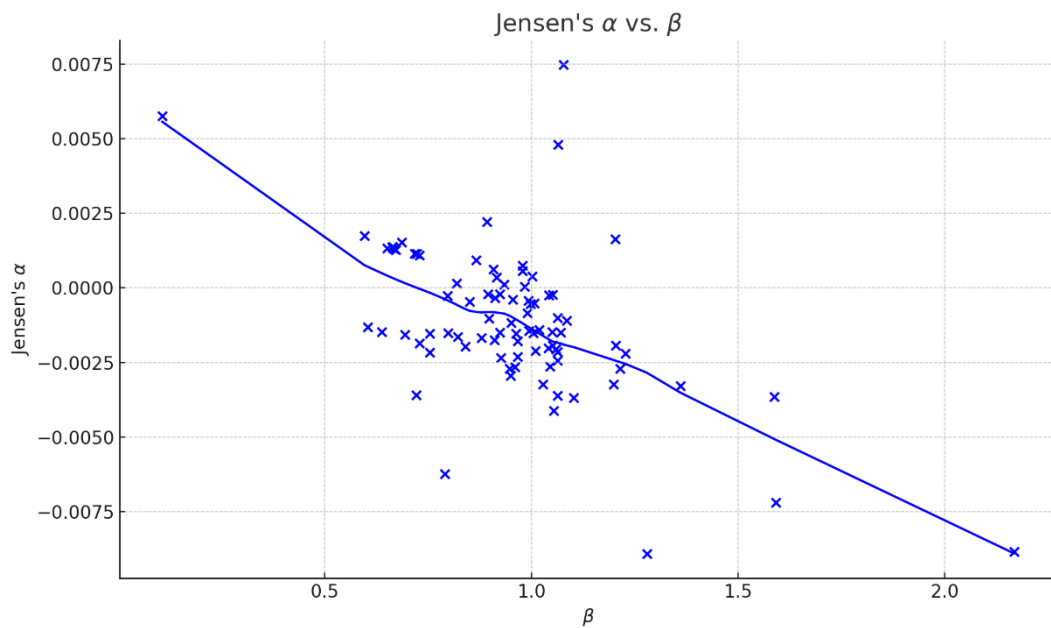


Note. Author's elaboration: Scatterplot of Jensen's α (vertical axis; CAPM excess return in proportion units) against the Pearson correlation ρ of portfolio excess returns with the corresponding benchmark excess returns (horizontal axis) for $n = 86$ robo-advisor (RA) portfolios over the COVID-19 analysis period τ_2 (31 Dec 2019–28 Feb 2022). A locally weighted scatterplot smoothing (LOWESS) curve is overlaid to visualise the conditional trend.

Analogously, the LOWESS in Figure 20 demonstrates a significant decline in Jensen's α when β increases, illustrating the negative relationship as stated by Spearman's $\rho \approx -0.46$.

The distribution of β centred on means of 0.953 and 0.961, demonstrating close tracking of the benchmark. Moderate dispersion was reflected by a standard deviation of 0.249, with betas ranging widely from 0.107 to 2.168. The marked positive skewness (approx. 1.20) indicated a right-tailed distribution driven by a few portfolios with higher systematic risk. Meanwhile, the elevated kurtosis (approx. 7.43) denoted a leptokurtic distribution with heavy tails and a sharp peak. Correlation ρ and systematic risk β thus showed a high dependence on the benchmark, which was accompanied by increasingly underperformance measured by α .

Figure 20. Jensen's α vs. β of CAPM with LOWESS (COVID-19 Period)



Author's elaboration: Scatterplot of Jensen's α (vertical axis; CAPM excess return in proportion units) against CAPM β (horizontal axis) for $n = 86$ robo-advisor (RA) portfolios over the COVID-19 analysis period τ_2 (31 Dec 2019–28 Feb 2022). A locally weighted scatterplot smoothing (LOWESS) curve is overlaid to visualise the conditional trend.

4.5. DISCUSSION

This section discusses the findings of the performance analysis for the two examined periods, to address the research questions RQ5 (total period 31/01/2017–28/02/2022) and RQ6 (COVID-19 period 31/12/2019–28/02/2022) from Subsection 1.2.

Previous studies examined a smaller sample with shorter time series and significantly fewer portfolios (e.g., five German RAs with a single equity group with a 50:50 stock-to-bond ratio in Puhle (2019)) or smaller heterogeneous samples with only simulated but no real data from different countries, such as Helms et al. (2022). This work thus contributes to research in this area by conducting a broader performance analysis study with longer time series of real realised data (2017–2022) between numerous portfolios with different equity shares grouped according to individual benchmarks (risk levels) from a larger RA sample.

In both periods examined in this chapter, the RAs analysed significantly underperformed appropriately constructed benchmarks, with pronounced dispersion within each benchmark cohort. Thus, the answer to both research questions RQ5 and RQ6 concerning underperformance is essentially the same, with the following nuances to be discussed below.

Concerning RQ5, for the entire review period from January 2017 to February 2022, 12 RAs (62 portfolios) outperformed the benchmark on average only in the CAPM statistics (Jensen's α , Treynor ratio) in the 10% equity group; in all other metrics and allocation groups, they underperformed the benchmark. Another exception was the benchmark group with 30% equity exposure, where the RAs were only slightly ahead of the benchmark on average in terms of the Treynor ratio, but not in terms of Jensen's α . Only one provider from the group with a 50% equity allocation outperformed its benchmark in all calculated indicators. Although isolated portfolios in several buckets outperformed their benchmarks in certain metrics, no consistent outperformance was observed. Drawdown and volatility of RA portfolios were significantly higher than the benchmarks, which weighed on risk-adjusted returns, a weakness that was evident during the COVID-19 sell-off in March 2020. Nevertheless, allocations of 30%, 50%, and 70% equities occasionally contain portfolios that performed better on at least some metrics.

With regard to RQ6: The COVID-19 subsample (December 2019 to February 2022, 16 RAs, 86 portfolios) showed a similar picture. The average underperformance persisted, even though more portfolios appeared in the top category, simply because the universe was larger. The only exception was the 30% equity cohort, which was evaluated using a RAPM measure. At least two RAs performed better in every metric within the 90% equity group, and one even achieved this result in the 70% group. Here, too, depending on the metric, there were sometimes relative winners among the 50%, 70%, and 90% equity weightings. Longer recovery times and deeper declines after the March 2020 crash suggested that many RAs failed to limit losses in a timely manner, reallocated too late, and, as a result, benefited less from the rapid recovery, leading to systematic average deficits.

These results already suggest that RA portfolios underperform their benchmarks. They are unable to beat their benchmarks according to most RAPMs and benchmark groups. Only a few RAs manage to outperform some equity exposure portfolios according to certain performance metrics. However, this does not speak for the performance of the RA provider if it offers several portfolios, as outperformance is then only a matter of luck and depends on chance whether an investor is offered the portfolio that outperforms after risk profiling. Since, though, almost all RA portfolios significantly underperform, outperformance is unlikely if investors are randomly assigned to a portfolio. Outperformance at the overall RA market level is even more unlikely if a random investor randomly selects an RA and gets a portfolio recommendation, which is partly random, depending on the information provided in the profiling. It can therefore be said with a high degree of certainty that it will underperform the benchmark.

The larger drawdowns in the COVID-19 market crash are one of the biggest weaknesses of RAs, as they were unable to manage the risks associated with COVID-19. If even a simple, naive benchmark with threshold rebalancing can manage such crises better, the question arises as to what RA risk management has to offer. It therefore appears that the smart risk models of RAs, as advertised by the providers, are not so smart after all. But the fact is that RAs also underperformed in bull markets and sideways markets before the crash, even significantly more than after COVID-19. The underperformance may be attributed to different asset allocation engines, fee structures, rebalancing rules, poorer risk management

strategies, and a lack of stop-loss and reallocation limits, as well as other design decisions. A precise understanding of which RA portfolio performed as it did, at which point in time, and for what reason, is a limitation of this study, as this would require data at the individual transaction level.

The mean–variance analysis for the entire period revealed that the most defensive benchmark, BM1 (with a 10% equity allocation), had the tangential portfolio with the highest Sharpe ratio. The RA portfolios' Sharpe ratios varied substantially, as did their distances to the efficient frontier and CML. Most benchmarks were on or close to the empirical efficient frontier. As the equity share increased, the RA and benchmark portfolios moved further away from the CML on the mu-sigma plane. Distance analysis from the CML showed that the benchmark had the lowest overall distances according to RMSE and, according to mean–variance theory, also offered the most efficient portfolios on average. All other RAs underperformed consequently. A similar pattern emerged during the COVID-19 period. Here, a defensive RA portfolio (Evergreen 1), with 20% equity exposure, defined the CML with the highest benchmark. On average, two RAs (Evergreen and Wevest) were able to remain closer to the CML and thus outperform the benchmark. However, both of these had portfolios with low to medium equity exposure, which explains why they were closer to the CML. Conversely, all other RAs underperformed their benchmark on average due to greater distances from the CML. It can be inferred from both periods that RAs underperform their benchmark according to mean–variance analysis. More efficient portfolios tend to have lower equity weightings. From a mean–variance theory perspective, the benchmark is a more efficient choice than the RAs.

The analysis of the correlation between Jensen's α from the CAPM and the correlation of the excess returns of the RA portfolios and those of the benchmark was moderately negative with a Spearman's ρ of approximately -0.29 . The negative correlation between α and β was even stronger, with a Spearman's ρ of approximately -0.5883 . The LOWESS in Figure 17 and Figure 19 illustrated this negative correlation geometrically. The averages and medians of Pearson's correlation ρ for RAs were 0.89 and 0.94, respectively, and those for β were 0.95 and 0.97, respectively, highlighting how closely the RA portfolios systematically tracked their benchmarks. The negative Spearman's rank correlation was even

more pronounced in the analysis of the COVID-19 period, where ϱ was approximately -0.30 between Jensen's α and correlation ρ , and approximately -0.57 between Jensen's α and β . This finding was backed up by the LOWESS curve in Figure 19 and Figure 20. The Pearson correlations ρ (mean ≈ 0.90 , median ≈ 0.93) and the systematic risk β (mean ≈ 0.95 , median ≈ 0.96) were both close to the benchmark.

These results suggest that RA portfolios that correlate more strongly with their benchmark show increasing underperformance according to CAPM. In other words, the higher the degree to which the RA portfolios track the benchmark, the greater the statistical underperformance of the RAs. The systematic coupling through higher market risk (β) therefore means that there are fewer opportunities for RAs to generate superior returns (α) through allocation decisions. The fact that ρ and β fulfil this relationship is also subject to the definition of systematic risk

$$\beta^{(i,j)} = \rho^{(i,j,k)} \frac{\hat{\sigma}^{(i,j)}}{\hat{\sigma}^{(0,k)}}$$

in the CAPM (53). This finding has clear implications for portfolio construction. It is evident that excessive monitoring of benchmarks and the implementation of smart beta strategies are counterproductive to the attainment of systematic outperformance. Elevated $\beta > 1$ appears to constrain active return generation even more; a phenomenon that was predominant in the market crash caused by the emergence of COVID-19, resulting in even greater underperformance. The supplementary fees levied by RAs, which are appended to the fees of the underlying (ETF) instruments, persist in exerting a negative influence on the performance of systematic strategies in relation to the benchmark. Tighter coupling to market movements leaves less scope for skilled managers to exploit undervaluation. To harness these insights, RAs platforms should consider reducing their reliance on high-correlation or high-beta positions and instead pursue genuine diversification across uncorrelated asset classes. A superior mean-variance optimisation strategy, characterised by enhanced diversification relative to the benchmark with reduced volatility measured by $\hat{\sigma}^{(i,j)}$, in conjunction with a reduced correlation ρ of the RA portfolios relative to the benchmark, has the potential to result in a substantial reduction of β . Furthermore, the implementation of an active alpha strategy has the potential to enhance performance. By implementing strategies such as these, it is theorised that they can reduce overall

systematic risk and tracking error, thereby enhancing the potential for active alpha generation. Concurrently, a concentration on α and β (Treynor ratio) optimisation will serve to further enhance risk-adjusted performance. While the theoretical framework appears promising, the practical implementation of RAs has reveals significant shortcomings. A substantial proportion of these individuals have not been able to demonstrate success in achieving the desired outcomes.

In an earlier study, Puhle (2019) examined five RA portfolios with a 50% equity share and showed that these underperformed a similar benchmark consisting of a global equity and bond ETF. This thesis extends these observations to all equity groups with up to 10% increments, concluding that RAs generally underperform their benchmark over a longer period of time. However, during the 2020 market crash, some RAs managed to limit investor losses. While Puhle (2019) used one portfolio with a 50% equity share from each RA, which still showed distortions over time due to a lack of rebalancing in his constructed benchmark, this thesis robustly supports the conclusion of underperformance through threshold rebalancing. This is because the equity deviations of the constructed benchmark were gradually corrected during the study period. Puhle (2019) enquires about the reasons behind the significant variance. The numerous different portfolio strategies and methods in Chapter III point to one reason for the variance in performance between RAs, which was also identified in Chapter IV. Furthermore, it is noted that performance variance between portfolios within a selected RA is also significant, which Puhle (2019) was unable to investigate due to the limited selection of one portfolio per RA. It can therefore be concluded that the RAs have inconsistent performance, with their portfolios having different risk levels (equity shares). Accordingly, there is no consistent quality within a fixed RA selection. Depending on the risk level, performance can be significantly worse, average, or relatively satisfactory compared to the benchmark. This finding makes it even more difficult for investors to evaluate and select an RA.

In contrast, Tao et al. (2021) reported that US RAs (2016–2019) outperformed equity, fixed-income, money-market and hybrid funds, as well as three indices, in terms of adjusted Sharpe, reward-to-VaR, and multi-factor α . This divergence is likely driven by differences in design and context, such as the use of different comparators (funds/indices vs. bespoke benchmarks), geographical differences (US

vs. Germany), differences in market phases (pre-Covid bull market vs. the period affected by the pandemic), differences in product definitions (pooled RA portfolios vs. managed portfolios) and differences in measurement choices (weekly, value-weighted US data with US Treasuries as the risk-free rate vs. monthly EURIBOR). In short, Tao et al. (2021) addressed robo-vs-fund dominance in an early-rising US market, whereas this thesis tests benchmark-relative and risk-adjusted performances between RAs.

In comparison to the results of this thesis Lee (2021) analysed Korea up to March 2021 using a multifactor model by Fama–French plus explicit rebalancing tests. This likewise revealed substantial dispersion: no RA provider simultaneously delivers abnormal excess returns, close market tracking and favourable rebalancing; market sensitivity is generally low; and RAs for securities show relatively stronger selection, while rebalancing is weak across types. Therefore, the conclusions by Lee (2021) do not contradict those of this thesis.

Helms et al. (2022) used bootstrapped simulations to compare 15 RAs internationally, reporting that there was no single RA that consistently outperformed the others. Rather, they found that each market had a particular platform that was favoured, notably Schwab Intelligent in the US, Wealthify in the UK and Scalable in Germany. Rankings were found to be broadly stable across Sharpe, VaR and Sortino ratios and were only modestly affected by fees and rebalancing choices. The methodological differences between the benchmark-relative evaluation in this thesis and the simulated pairwise comparisons in Helms et al. (2022) are a plausible explanation for the divergence. This thesis emphasises absolute benchmark efficiency, whereas Helms et al. (2022) emphasise relative performance within a heterogeneous set based on simulated bootstrapping data. Helms et al. (2022) conclude that the specific design of RAs is a more important driver of performance than their geographical location and that most RA providers failed to deliver consistently superior risk-adjusted performances. The performance analysis results of German RAs in this thesis also support this conclusion. As such, the data reinforces the underperformance findings.

Integrating the conclusions of the performance analysis from previous studies, which also covered other regions, with the findings of this study that focused on the German market, it can be inferred that the risk-adjusted underperformance of RAs compared to benchmarks is not confined to any specific

region, as analogous conclusions have been reached by independent studies. Consequently, this thesis's findings suggest that RA underperformance is a matter of design and portfolio strategy rather than jurisdiction.

During data collection, it was observed that, although all selected RAs disclose charts, price data, or returns for the purpose of performance assessment, they provide almost no information on risk-adjusted performance indicators, such as CAPM indicators, Sharpe ratios, or returns in relation to drawdowns. Furthermore, the absence of benchmarks in the disclosure of comparisons is a common occurrence. Alternatively, 'pseudo' benchmarks are sometimes adopted, though these frequently exhibit disparities in composition, including significantly divergent equity proportions and, by consequence, risk levels. The evidence presented here, in conjunction with the observed underperformance, gives rise to the suspicion that RAs may be deliberately manipulating this representation with a view to gaining a competitive and marketing advantage. This suggests that RAs may prioritise strategies that focus on maximising customer conversions and short-term AUM as opposed to emphasising sustainable performance and long-term growth.

Therefore, it can be recommended, based on the findings of this thesis, that RA providers are urged to improve their risk-adjusted performance if they are to survive in the long term and deliver a value proposition to their clients. Benchmark tests must be carried out for the strategies, and weak phases must be analysed in detail in order to optimise the strategies. RA providers should use better risk models, especially in crash phases. Early exit or reduction of equity exposures and re-entry after a major downturn in the equity markets can generate positive alpha relative to the benchmark. It is also a question of market timing, but the benchmark shows that even without market timing, very strong performance can be achieved with naive threshold rebalancing. This advantage is also due to the well-known cost averaging effect (Brennan et al., 2005; Kirkby et al., 2020; Marshall, 2000). During the COVID-19 crash, equities declined in value, as did their weight in the benchmarks, so that rebalancing sold bonds and purchased equities at lower prices. After equity valuations recovered and rose, the equity exposures in the benchmarks were sold again above the threshold at a profit. Using this method, the benchmarks

bought equities low and sold them high. Such profit-taking through the cost averaging effect can also generate alpha.

Table 43 provides a summary of the key findings and recommendations for various stakeholders based on the performance analysis results.

Table 43. Key Findings from the Performance Analysis

Key finding	Recommendation	Primary audience
Persistent average underperformance vs matched benchmarks across most equity buckets and metrics; only isolated pockets of outperformance (RQ5, RQ6).	Re-engineer product sets to deliver clear, risk-bucketed value: either (i) benchmark-like outcomes at lower all-in fees, or (ii) demonstrable alpha at the same risk. Publish side-by-side performance vs the correct equity-matched benchmark for every portfolio.	Providers
Higher drawdowns and volatility; slower post-March-2020 recovery (RQ6).	Embed drawdown limits, volatility targeting, and crisis playbooks (rules for de-risk/re-risk). Report max drawdown and time-to-recovery live.	Providers, Risk
Benchmarks sit on/near the empirical efficient frontier; many RA portfolios plot below the CML—distance widens at higher equity shares (RQ5, RQ6).	Design to the CML: minimise σ and tracking error at a given return; publish distance-to-CML metrics per bucket; prioritise lower-equity sleeves where efficiency is empirically stronger.	Providers
CAPM diagnostics: α is negatively related to β and benchmark correlation ρ ; high β and ρ plus fee drag $\rightarrow$ negative α (RQ5, RQ6).	Manage β and ρ as live constraints (target ranges, alerts); broaden uncorrelated sleeves; if the strategy hugs the benchmark, price accordingly.	Providers
Occasional winners exist within 30/50/70/90% buckets, but provider-level	Standardise assignment logic and show provider-wide dispersion charts so clients see	Providers, Compliance

consistency is weak; assignment risk for clients (RQ5, RQ6).	variability across your line-up; avoid marketing based on single-portfolio anecdotes.	
Threshold rebalancing benchmarks handled the COVID shock better (buy low/sell high via bands) (RQ6).	Replace calendar-only rules with threshold/volatility-aware rebalancing; disclose bands and band-breach behaviour; back-test vs 2020-style shocks.	Providers
Underperformance pattern appears design-driven rather than jurisdiction-specific (RQ5, RQ6, synthesis with literature).	Run method bake-offs (MPT, BL, factor, risk-parity, vol-targeting) in a shared simulator; adopt the best-in-bucket engine per equity band and publish results.	Providers, Researchers
Sparse disclosure of risk-adjusted metrics; use of “pseudo” benchmarks and marketing focus on absolute returns rather than risk-adjusted performance metrics (RQ5, RQ6).	Mandate a standard template per portfolio: Sharpe, Treynor, α , β , RDR, MDD, recovery time; require benchmark declarations that match equity share and hedging.	Regulators/ Standard-setters
Fee stacks materially depress outcomes when portfolios track the benchmark closely (RQ5, RQ6).	Tie fees to tracking tolerance: low-fee wrappers for benchmark-like mandates; performance-linked fees only where ex ante capability for α is evidenced.	Providers
Data granularity limits causal attribution; RA portfolios are “black boxes” without trades/positions (RQ5, RQ6).	Provide holdings- and trade-level audit trails (at least monthly; ideally daily) to supervisors/research partners; introduce advice provenance logs.	Providers, Regulators
Heterogeneity of methods (rebalancing rules, risk controls, overlays) likely drives dispersion (RQ5, RQ6).	Document engine settings (β /q targets, band widths, hedging) and link them to outcomes in per-bucket scorecards; iterate designs where dispersion is value-destructive.	Providers

Drawdown management is the critical differentiator in crises (RQ6).	Before onboarding, request the provider's MDD and recovery history for your equity bucket; avoid mandates without clear crisis rules.	Investors/ Clients
Investors face low odds of beating the benchmark via typical RA portfolios (RQ5, RQ6).	Prefer providers that publish per-bucket risk-adjusted scorecards and distance-to-CML; where portfolios track and fees are non-trivial, consider a low-cost benchmark replicator.	Investors/ Clients

Note. Author's elaboration: This table summarises key empirical findings from the benchmarked performance analysis of robo-advisor (RA) portfolios and provides targeted recommendations and primary audiences (Providers, Risk, Compliance, Regulators, Researchers, Investors). Compliance, researchers and risk refer to the roles and functional areas of RAs. Results and recommendations are based on monthly portfolio and benchmark data between January 2017 and February 2022 of 16 RAs with 86 portfolios (see Section 4.1). Performance diagnostics referenced include: CML = capital market line (mean–variance tangency line); α = Jensen's alpha (CAPM excess return); β = CAPM beta (systematic market exposure); ρ = Pearson correlation of portfolio excess returns with the benchmark; MDD = maximum drawdown; RDR = recovery-to-drawdown ratio (time/effort to recover from drawdown); SR/Sharpe = Sharpe ratio (see Subsection 4.1.3). Research-question pointers (RQ5, RQ6) refer to the study's stated research questions (see Section 1.2). Recommendations are pragmatic: they reflect design and governance changes (product design, rebalancing, disclosure, guardrails, fee alignment, auditability and independent simulation/bake-offs) justified by observed patterns of (i) average underperformance versus equity-matched benchmarks, (ii) wider distances to the CML at higher equity weights, and (iii) negative associations of α with β and benchmark tracking.

V – CONCLUSIONS

V - CONCLUSIONS

This concluding chapter first recapitulates the objectives of the dissertation, and the research gaps it set out to address. The thesis pursues an integrated programme that maps contemporary robo-advisory practice in Germany, formalises the underlying workflow, prototypes an AI-enabled implementation, and evaluates realised performance against transparent, risk-matched benchmarks.

The first objective is descriptive and taxonomic: to provide a current, market-wide account of investor profiling as implemented by RAs active in Germany. Prior knowledge lacked a validated, comparable question taxonomy and an up-to-date view of what providers actually ask prospective clients. The thesis, therefore, sets out to enumerate questions, organise them into content categories, and identify typical items and omissions, thereby building a shared vocabulary for profiling across firms (RQ1).

The second objective concerns portfolio construction in practice. The German market lacked granular, holdings-based evidence on the menus that investors are actually assigned to, the breadth of equity-exposure bands, and the extent to which sustainability and alternative assets are reflected in live allocations. Hence, the study mapped asset-class building blocks, weights and risk buckets across a large cross-section of portfolios and documented disclosed strategy families (e.g., mean-variance engines, overlays and hybrid governance). This approach addresses the gap between questionnaire breadth and portfolio granularity and enables risk-consistent benchmarking later on (RQ2).

A third objective is methodological: to bridge the divide between business-operational accounts of RA and formal financial mathematics. Existing descriptions typically stop with qualitative block diagrams or static optimisers. The thesis therefore developed a coherent mathematical framework that treats profiling as an operator on the client state, couples it to an allocation operator under market dynamics, and embeds event-driven rebalancing. This structure generalises prior static formulations and supplies a modular chassis for subsequent algorithmic and operational development (RQ3).

Fourth, the field lacked a concrete, scalable implementation of an LLM-driven RA that ingests market-realistic questionnaires, maps responses to investable universes, and enforces legal-technical guardrails. To fill this operational gap, the dissertation implemented a GPT-based prototype with a defined instrument universe and validation checks and examined empirical allocation patterns together with the relationship between profiling variables and equity exposure as a risk proxy (RQ4).

The fifth objective addresses realised outcomes. Evidence on German RAs' risk-adjusted performance remained fragmentary, with small samples, short horizons, heterogeneous comparators, and limited insight into behaviour during stress. The study therefore assembled a benchmark-consistent, bucketed dataset of portfolios, constructed low-cost reference portfolios matched by equity exposure and currency hedge, and evaluated outcomes across complementary lenses (risk-adjusted performance metrics, mean–variance geometry, and CAPM diagnostics) over a longer horizon (January 2017–February 2022) (RQ5).

A sixth, related objective is to isolate crisis-phase behaviour using common methods. The market shock associated with the COVID-19 period provided a natural experiment to observe drawdowns, rebalancing responses and recoveries within matched equity buckets. By holding measurement and benchmarking constant, the thesis aimed to characterise how RAs performed relative to simple reference strategies during acute stress (December 2019–February 2022) (RQ6).

Collectively, these objectives tackle four classes of gaps: (i) informational opacity in profiling questions and portfolio menus (RQ1–RQ2); (ii) the absence of a unified, dynamic formalism for the full advice-to-allocation pipeline (RQ3); (iii) the lack of an operational blueprint for LLM-enabled advice with auditable outputs (RQ4); and (iv) the scarcity of benchmark-matched, risk-class-specific evidence on realised performance in calm and crisis regimes (RQ5–RQ6). The remainder of the conclusion synthesises findings against these aims, draws implications for regulators, investors and providers, acknowledges methodological bounds, and sets out a research agenda that extends the contribution of each question into cumulative, replicable evidence. Section 5.1 provides a summary of the key findings of the dissertation, organised according to the research questions. Section 5.2 summarises the practical implications and recommendations according to the most important stakeholders: regulators, robo-advisory investors and robo-

advisory providers. Section 5.3 outlines the limitations of this thesis, organised by Chapters III and IV. Finally, Section 5.4 outlines the outlook and implications for future research.

Chapter III synthesises the empirical analysis of investor profiling and portfolio strategies by 45 robo-advisors in Germany, together with mathematical modelling of the RA process and a GPT-based RA prototype implementation. The chapter covers:

- i. a market-wide analysis of investor-profiling questionnaires used by 45 RAs active in Germany as of July 2024;
- ii. an assessment of their portfolio menus and strategy disclosures;
- iii. a quantitative text-analytics validation of the questionnaire taxonomy (TF-IDF) and cluster analysis of the full corpus of 687 questions;
- iv. a formal mathematical model of the end-to-end RA process; and
- v. the implementation and evaluation of a GPT-4o-based RA on a sample of 100 synthetic investors.

Chapter IV synthesises the evidence from the empirical performance analysis of 16 RAs with 86 portfolios and provides answers to the three research questions that guided the inquiry into German robo-advisors' outcomes between January 2017 and February 2022. The chapter covers the following topics:

- i. assembling a transparent dataset of monthly portfolio indices for 16 providers and 86 portfolios, mapping each portfolio to one of 10 equity-exposure benchmark groups and defining a set performance measures, with evaluations both over the total period 31/01/2017–28/02/2022 and the COVID-19 subperiod from 31/12/2019 to 28/02/2022;
- ii. a complementation by a mean–variance study with Sharpe isolines and
- iii. a correlation analysis of Jensen's alpha with beta and benchmark co-movement.

The principal insights are structured according to the research questions RQ1–RQ6 outlined in Subsection 1.2.

5.1. SUMMARY AND KEY FINDINGS

5.1.1. Robo-Advisory Questions in the Profiling Process

RQ1: What and how many questions and content categories are asked by the robo-advisors active in Germany as part of investor profiling, and what differences, similarities, and typical questions can be gained from this?

The market scan established that Germany's RA landscape in July 2024 is both larger and more fragmented than prior studies suggested, with forty-five active providers compared with far smaller samples in earlier work. Superficial differentiation is widespread, with some firms emphasising sustainability, others offering a broad catalogue of portfolios and customisation, and still others focussing on simplicity, low entry thresholds, streamlined onboarding, and a subset of advertising hybrid human-machine models. Yet beneath this variety, a common pattern emerges: profiling questionnaires converge on overlapping themes, and recommended portfolios are dominated by low-cost passive equity and bond ETFs.

Of the 687 questions collected, eight categories of recurring content emerged:

1. objectives and purpose;
2. investment horizon and age;
3. amount and frequency of deposits;
4. risk preferences;
5. sustainability preferences;
6. other idiosyncratic preferences;
7. knowledge and experience; and
8. financial situation.

On average, providers asked 15.27 questions, but there was substantial dispersion (standard deviation: 6.96; range: 2–36). The market therefore exhibits both convergence and divergence: almost all firms now ask about risk, horizon, experience and finances, yet the number and phrasing of items within these domains vary sharply across providers. The depth of questioning was highest for financial situation (3.58 questions per RA), knowledge and experience (3.16), risk preferences (2.64), and sustainability preferences (1.96). Risk-orientated, horizon-

related, experiential and financial questions appeared in at least 89 % of platforms; no single RA dominated all categories, though some led several while others lagged across many.

Compared to the results from previous scientific literature, risk preferences have expanded the portfolio dimensions, and sustainable portfolios are increasingly being offered. Coverage was high for risk (97.78%), investment horizon and age (91.11%), knowledge and experience (91.11%), and financial situation (88.89%). Contrary to earlier findings that German platforms neglected household balance sheets, the present sample shows a significant expansion of the KYC scope to include income, expenditure, assets, and liabilities, as well as a more nuanced elicitation of risk attitude and ESG preferences. Based on the number of questions per RA, it can be concluded that investors' preferences for sustainable investments play a similarly important role as risk tolerance, which is a novelty compared to previous studies, where sustainability preferences were not identified in the profiling process of German RAs.

At the same time, critical weaknesses persist. One-fifth of providers did not ask about investment objectives; half asked two or fewer knowledge and experience items; and 42.22% asked only one risk question. Only 15.56% posed five or more risk items. The discussion linked this shortfall to the difference between ad hoc risk questions and validated psychometric instruments, such as the GL-RTS, DOSPERT and related scales. It also linked the uneven depth to regulatory obligations under WpHG and MiFID II, noting that online responses may form the sole basis for robo-advice and must therefore be comprehensive, internally consistent and free from bias if suitability is to be defensible.

The TF-IDF and cluster analysis provided an independent quantitative validation of the inductive taxonomy. Using a vocabulary of 1,031 n-grams over the 687 questions, candidate partitions from two to eight clusters were evaluated with within-cluster sums of squared errors and the Calinski–Harabasz index. Two granularities emerged as most useful. A fine categorisation with seven to eight clusters achieved the tightest groupings and highest selectivity; a coarser alternative with five clusters preserved content meaning with manageable information loss.

The analysis also revealed why risk, experience, and financial situation items are harder to standardise: they are linguistically diverse and spawn subclusters, whereas goal and horizon questions cluster tightly. The representative “top five” questions nearest each category centroid closely matched the eight categories identified by qualitative coding, confirming that the taxonomy is not an artefact of subjective judgement. This quantitative corroboration underpinned the call to replace eclectic risk questions with validated psychometrics, which would likely reduce heterogeneity in the risk cluster, improve comparability across providers, and help meet supervisory expectations for robust, bias-resistant profiling.

5.1.2. Portfolios, Strategies and Asset Allocations of Robo-Advisors

RQ2: What portfolio strategies and specific compositions of asset classes and weightings are proposed by the robo-advisors active in Germany based on the results of investor profiling, and what are the differences, similarities, and typical portfolios that can be derived from this?

The repeated journeys through all 45 profiling processes yielded 405 distinct portfolios, with an average of nine per provider and high variance (standard deviation 7.55; range 3 to 37). Portfolios were predominantly constructed from passive equity and bond exchange-traded funds (ETFs), frequently global or developed-market trackers, with a minority of providers incorporating commodities, real estate, or, infrequently, cryptocurrencies. Equity exposure was identified as the primary risk dial. The menus traversed the entire 0%–100% equity range in 10-percentage-point increments, and portfolios were distributed across well-defined bands: The most prevalent proportion of equity involved was 50%, with 30% and 70% following closely behind. Proportions of 40%, 20%, 80%, and 100% were also notably represented. The mean value of equity shares across the corpus was just above the midpoint, and more than half of all portfolios contained at least 50% equities. ESG-labelled variants were available at nearly half of the firms, yielding 147 sustainable portfolios; commodities appeared at about two in five firms, real estate ETFs at roughly one in six, and crypto at fewer than one in ten.

Consequently, the cross-section evinces both breadth and constraint. The breadth of the range is evident in the variety of risk bands and the growing

presence of ESG toggles and alternative sleeves. Constraint manifests in the impoverished mapping from extensive questionnaires into a limited menu of stock-bond mixes. The analysis revealed that 77.78% (35/45) of RAs offered 10 portfolios or fewer. Approximately half of the providers offered a maximum of five model portfolios, which compresses heterogeneous profiles into broad risk buckets and leaves many questionnaire dimensions without a visible imprint on the allocation. The discussion posited the hypothesis that this pattern reflects an efficient-frontier logic. This hypothesis suggests that providers maintain a small number of equity bands and allocate clients to these bands accordingly, rather than recomputing weights from first principles for each profile.

In terms of a pragmatic response, the study put forward a two-tier menu configuration, comprising eleven “core” equity bands ranging from 0% to 100%, with the incorporation of overlays for the following categories: horizon (short, intermediate, long), ESG (on/off) and knowledge or trading-frequency tiers. This process affords approximately 132 meaningful configurations, a sufficient quantity to honour profiling information without compromising the manageability of the catalogue. In instances where dynamic instantiation is operationally advantageous, the same logic can be implemented without the proliferation of static stock keeping units (SKUs).

The investigation of the strategy disclosures confirms the dominance of classical mean–variance thinking from modern portfolio theory (MPT) and the gradual layering of complementary methods and operating models. A substantial proportion of the providers—two-thirds, to be precise—referenced modern portfolio theory or mean–variance optimisation as their strategic asset-allocation engine. Impact investing, or ESG mandates, was observed in approximately half of the firms, typically manifested as parallel initiatives rather than pervasive shifts. A third of the participants explained active overlays, frequently employing terms like “hedging” or “tactical rebalancing”.

The presence of buy-and-hold language was observed in approximately 30% of the firms, indicating a rebalancing of policy towards fixed targets rather than tactical shifts. The integration of human oversight (hybrid character), where committees review or ratify model outputs, was the focus of approximately one-quarter of the RA providers. A smaller number of RAs cited specific quantitative

frameworks, with a few naming Black–Litterman or value-at-risk; a similar number referenced factor or smart-beta building blocks or core–satellite structuring; and one provider described risk parity. Five providers disclosed the use of AI, spanning statistical learning for macro-signal extraction, multi-model ensembles for regime detection and forecasting, and partially or fully automated signal-to-execution chains.

The most prevalent pairings were MPT with ESG, MPT with buy-and-hold; and MPT with active overlays. Collectively, these pairings underscore a pragmatic synthesis: mean–variance provides the skeletal structure, while ESG toggles, tactical tilts and governance overlays provide differentiation. The nascent application of AI methods by RAs suggests a likely future direction of machine-assisted dynamic allocation under human oversight. However, the current evidence is insufficient to draw conclusions about risk-adjusted performance.

5.1.3. Dynamic Mathematical Modelling of the Robo-Advisory Process

RQ3: How can the robo-advisory process with all its building blocks be formalised and dynamically extended into a coherent mathematical framework that generalises prior static formulations and equips providers with a foundation for subsequent algorithmic and operational development?

The thesis contributes a coherent end-to-end framework that integrates onboarding, allocation, and rebalancing within a single mathematical structure and generalises prior static formulations. The initial mapping process involves the transformation of the questionnaire into a low-dimensional profile vector. This vector encapsulates risk appetite, time horizon, liquidity, sustainability stances, and associated traits. The definition of this profiling operator incorporates an allowance for stochasticity, with the objective of facilitating the representation of self-assessment noise, imputations, and genuine preference changes. The second mapping takes that profile together with a market-state vector and returns target portfolio weights in the portfolio simplex. It is deliberately agnostic about the engine used and thus accommodates mean–variance, Black–Litterman, factor-based or machine-learning surrogates.

The third element is an event-driven control layer. Decisions regarding rebalancing are determined on a grid defined by observable events, including deposits and withdrawals, profile revisions, and market-regime signals. These decisions are made independent of calendar considerations, and the objective is to achieve balances that are proximate to the target mix in relation to the cost of adjustment under operational constraints. Two dynamic generalisations are central to this study: Profiles are subject to evolution over time, transitioning from a static entity to a dynamic process. The propagation of profiling uncertainty extends to allocations, and the decision grid is influenced by investor and market events, aligning with the management of portfolios in low-fee environments. This structure unifies and extends earlier block-diagram descriptions and static optimisations.

Addressing RQ3, the principal learning is that a unified state–operator–control formalisation renders the RA workflow both analysable and falsifiable. Once the process is couched in mappings between explicit spaces, claims about performance, suitability and robustness become testable statements about how well the operators and cost terms approximate the economic reality they intend to capture. The framework thus transforms an otherwise heuristic pipeline into a set of estimable objects, allowing the empirical researcher to identify which components drive outcomes in production—estimation of client state, estimation of market state, or the policy that trades off tracking and costs.

A further insight is that the separation between estimation and decision facilitates disciplined experimentation. Because the control layer depends solely on the current state and declared penalties, any alternative profiling procedure or market model can be swapped in and evaluated under invariant decision rules. This enables out-of-sample comparisons that are not confounded by changes elsewhere in the stack and supports the construction of shared benchmarks for allocation engines and rebalancing heuristics. The same separation clarifies governance: data lineage, parameter choices and policy constraints can be versioned and audited, and historical decisions can be reproduced for supervisory review.

The formalism also yields concrete measurement strategies. The tracking-error term provides a natural notion of deviation from the

client-appropriate target, while the transaction-cost term provides a measurable budget for frictions. Their interaction defines an observable frontier along which platform settings can be tuned. Inferences on the turnover tolerance and cost kernel can proceed based on realised trades and market conditions, and profile dynamics can be calibrated based on observed revision patterns. As uncertainty in profiling is propagated end-to-end, sensitivity analyses and stress tests become methodologically coherent, allowing variance in questionnaire responses to be traced through to dispersion in allocations and trading activity.

The proposed methodology offers providers a modular blueprint, which involves maintaining a versioned profile state, computing targets as a function of the profile and market state, and rebalancing on events within disclosed tolerance bands and turnover ceilings. Furthermore, it equips researchers with a uniform framework upon which diverse allocation engines and rebalancing rules can be evaluated under shared assumptions. The dynamic formulation reveals new research questions. Event-trigger design becomes an optimisation problem under frictions; profile updating can be studied as a filtering task for latent preferences; and the embedding permits fairness, suitability and explainability assessments to be expressed in the same primitives that govern portfolio construction. These directions indicate how the framework generalises static accounts to a program of cumulative empirical and theoretical work.

5.1.4. Implementation and Analysis of a GPT-Based Robo-Advisor

RQ4: How can a GPT-based robo-advisor be implemented using a suitable questionnaire and investment universe, and what empirical patterns emerge in an investor deployment regarding portfolio composition and the relationships between profiling variables (risk-preference items and household balance-sheet indicators) and equity exposure as a risk proxy?

The study constructed a GPT-4o-driven system around a 32-item questionnaire and a realistic investment universe and then subjected 100 synthetic investors to profiling and allocation. The resulting allocations exhibited clear regularities. The AI-based system favoured broad, low-cost exposures, with average weights of approximately 59.4% in equity ETFs and 30.3% in bond ETFs,

and modest holdings in alternatives such as real estate, precious metals and crypto that together summed to about 10%. Within equities, global and developed-market trackers exhibited the highest average weights, while emerging-market and sector exposures were comparatively smaller and more variable. The equity weights were found to be concentrated within the 65% and 40% ranges; a significant proportion of portfolios, exceeding 89%, maintained a minimum equity allocation of 40%, while slightly more than 50% held a minimum of 65%.

The treatment of sustainability was asymmetric. When investors indicated a preference for sustainable investing, the system generated mixed portfolios that combined ESG and conventional instruments rather than exclusively ESG portfolios. Conversely, when sustainability was deselected, ESG instruments disappeared entirely. This phenomenon is analogous to the market-wide discrepancy between detailed sustainability questionnaires and the relatively coarse ESG options present in many menu architectures. The implementation process was accompanied by several challenges. A total of twelve initial allocations did not satisfy the fundamental “sum to one” verification criterion (90%, 95% or 110% of assets), and these were subsequently rectified through the use of subsequent prompts. The discussion interpreted the above findings as decisive evidence for the need for hard validation layers—mass-balance checks, whitelisted instruments and asset-class caps—between model output and client-facing advice.

The regression analysis linked questionnaire variables to equity exposure as a risk proxy. Univariate models demonstrated that risk-attitude items—past willingness to take risk, concern about losses and stated loss acceptance—were by far the strongest single variables, each explaining close to half the variance in equity shares, where, due to the correlation, one of the risk variables is sufficient for profiling. Multivariate models served to reinforce this hierarchical structure, thereby introducing a greater degree of complexity and nuance. The optimal two-factor specification, which combined past risk willingness with the coverage reserve (in months of expenses), explained more than half of the variance. The addition of liabilities and net assets to the model, extending it to four factors, resulted in an increase in adjusted R^2 to just under 60%.

The findings indicated that age was consistently and significantly negative, while longer horizons and larger liquidity buffers were moderately positive. Two

coefficients proved to be counterintuitive: liabilities were entered positively and net assets negatively. The study posits that this inversion is probably indicative of a combination of small-sample noise, omitted capacity constraints such as debt-service ratios and collateral, and corpus-driven heuristics that can map “liability-rich” profiles to longer horizons or “catch-up” targets in the absence of guardrails and human review.

A concomitant phenomenon was the substitution of cryptocurrencies for equities at the highest stated risk levels. Once equity weights had already increased, the system utilised alternatives to allocate the marginal risk budget, thereby reducing the equity percentage even as the overall portfolio risk increased. These behaviours are economically interpretable, yet they underscore the necessity for explanation triggers whenever recommendations diverge from intuitive capacity logic or when sustainability preferences are only partially honoured.

In the context of production systems, the LLM-based advisory solution is unambiguous. The implementation of pre-trade validations that enforce accounting and eligibility constraints, the use of post-hoc rationales that cite specific answers and decision rules, and the establishment of triage thresholds that route contradictory or high-risk profiles to human review are all effective methodologies. Through individual configuration and supplementary specifications, RAs can integrate LLM-based systems into their profiling and portfolio recommendation processes instead of using static and less individualised off-the-shelf solutions. Below is a detailed explanation of what the LLM-based implementation entails for RAs. The consequences of LLM-based robo-advisory systems for RAs will now be explained in detail.

From a provider’s perspective, strategic significance lies in the productisation of this capability as a governed, auditable service. Firstly, regarding architecture, prompts, constraint sets and instrument catalogues should be regarded as version-controlled artefacts. These should be released through change management with blue/green rollouts and automatic rollback on breach of predefined quality gates. One potential solution would be to implement a system of “advice provenance” ledgers. This process would involve the recording of the model/version, prompt template, input profile vector, guardrails triggered, validation results, and any human overrides. The implementation of such a system would create a defensible audit trail and accelerate root-cause analysis.

Additionally, RAs should implement model-risk management tailored to advisory outputs. Beyond the standard accuracy checks, RAs may define operational key performance indicators (KPIs), such as zero mass balance errors, the percentage of recommendations that fit declared risk buckets, the stability of recommended beta and correlation versus targets, the honour rate of stated ESG preferences, the exception/override rate, and the time-to-resolution for triaged cases. Continuous monitoring with drift alarms and periodic scenario re-certification (stress, liquidity, concentration) should be embedded in production.

RAs should then reengineer the operating model and position the LLM as a “first-pass allocator”, overseen by a specialist review desk with clearly defined service-level agreements. RAs have the capacity to capture reviewer rationales in structured fields, thereby facilitating iterative prompt/guardrail improvements and introducing tiered service levels (fully automated for plain-vanilla profiles; hybrid for edge cases) to align cost-to-serve with client complexity.

Furthermore, RAs should revisit portfolio engineering and governance, encode concrete caps and eligibility lists as machine-readable policy (e.g., crypto and leverage limits, duration bands, and issuer/sector concentration limits), and make them client-visible through layered explanations and “what-if” views. The utilisation of explainable artificial intelligence (XAI) within LLMs facilitates the elucidation of their positions. It is submitted that RAs would benefit from adopting this approach. For instance, in scenarios where the sustainability option is selected, RAs are required to implement deterministic mapping rules and a “closest-feasible” protocol, accompanied by explicit, auditable deviations.

LLMs facilitate the customisation of RAs at scale. To convert this into a competitive advantage, price should be calculated according to the delivered capability (e.g., lower fees for benchmark-hugging mandates, premium tiers for genuinely bespoke, constraint-aware portfolios). Live metrics on drawdowns, recoveries, and mandate adherence should be published.

5.1.5. Performance of German Robo-Advisors (31/01/2017–28/02/2022)

RQ5: How do RAs in Germany perform in terms of their (risk-adjusted) performance compared to benchmarks between January 2017 and February 2022?

Measured across 62 portfolios from 12 providers that span the full period, RAs underperform their equity-bucketed benchmarks on average in nearly all groups and for almost all indicators. In the most defensive bucket, which contains 10% equity, there is only a narrow average advantage in CAPM-based statistics; however, other indicators worsen even in this case. When the analysis counts genuine outperformance at the portfolio level, the proportions are small. Depending on the metric, only between 3% and 13% of portfolios, or about 5% to 21% of the sample, beat their matched benchmark. This discrepancy is most evident in symmetric and asymmetric risk-adjusted measures, such as modified Sharpe, annualised return-to-volatility, and return-to-drawdown ratios. Several providers emerge as notable exceptions within specific categories. Raisin Invest, for instance, is a clear example of this within the 50%-equity group. However, this pattern is not consistent across categories or metrics, and no provider consistently dominates within their line-up.

The mean–variance analysis elucidates the rationale behind the observed averages. In the total period (σ, μ) -lane, the benchmark set essentially traces the empirical efficient frontier and defines the capital market line, which is tangent to the most defensive benchmark. Most RA portfolios are situated below this line, frequently exhibiting comparable average excess returns but elevated standard deviations, which mechanically diminish their Sharpe ratios and consequently place them at a greater distance from the capital market line. As equity shares rise, the distances to the capital market line grow and dispersion widens. It can thus be concluded that higher-risk RA portfolios tend to be less efficient than the corresponding higher-equity benchmarks. A complementary CAPM correlation study establishes a connection between these geometry-of-returns observations and econometrics: Jensen's alpha has been demonstrated to be inversely correlated with both beta and the correlation of excess returns with the benchmark. In practical terms, portfolios that shadow the benchmark more closely and portfolios that load more heavily on systematic market risk, earn lower risk-adjusted excess

returns. Throughout the total period under consideration, the average correlation with the benchmark is very high, and the average beta hovers near one, which leaves little room for positive alpha once additional platform-level fees and frictions are paid.

5.1.6. Performance of German Robo-Advisors over the COVID-19 period (31/12/2019–28/02/2022)

RQ6: How do RAs in Germany perform in terms of their (risk-adjusted) performance compared to benchmarks in times of crisis, taking the period between the outbreak of the COVID-19 pandemic in December 2019 and the beginning of the war in Ukraine in February 2022?

The COVID-19 window expands the sample to all 86 portfolios and sharpens the test because all providers navigate the same shock and rebound. The averages again favour the benchmarks in almost all groups and measures, with a modest exception in the 30%-equity bucket for a Treynor-style ratio. Portfolio-level outperformance becomes slightly more frequent than over the total period—9% to 21% of RA portfolios outperform their benchmark, depending on the metric—because the larger universe contains a few strategies that managed the crisis better. There are notable provider-specific pockets of strength: Peaks and wevest have achieved significant progress in their clean sweeps in high-equity buckets across multiple indicators. In contrast, Evergreen has demonstrated a notable capacity to deliver superior Sharpe ratios within lower-equity mixes. Notwithstanding, the general outlook for RAs remains negative. It is evident that drawdowns are more pronounced for the average RA portfolio in comparison to the matched benchmark. Furthermore, recovery times from the March 2020 trough are extended, resulting in risk-adjusted ratios that subsequently lag. These findings suggest that the majority of providers adopt a risk-averse strategy during periods of decline, implement risk-mitigating measures too gradually during periods of growth, or both.

The mean–variance geometry during COVID-19 reinforces the diagnosis. The tangency portfolio is no longer a benchmark but rather a defensive RA portfolio with approximately 20% equity, which becomes the empirical capital market line. The consequence of this phenomenon is that distance metrics for the rest of the

universe increase relative to the full period, often substantially so in the higher-equity groups. The correlation study replicates the pattern observed over the entire interval: alpha declines as beta and benchmark co-movement rise. The average correlation and beta remain high, indicating that a significant proportion of RA portfolios mirror their bucketed benchmark behaviour. However, during periods of stress, these portfolios experience amplified losses and protracted recoveries. Many RA portfolios appear to have failed to participate in this cycle, unlike the conventional rebalancing strategy of benchmarks that involves purchasing equities during market weakness and selling during periods of strength.

5.2. PRACTICAL IMPLICATIONS AND RECOMMENDATIONS FOR ACTION

5.2.1. For Regulators and Standard-Setters

The evidence supports a series of practice-orientated conclusions that are intended to speak to regulators and the wider market, to clients, and to providers themselves. Furthermore, each recommendation is connected to the weaknesses and strengths identified in the discussion. For regulators and standard-setters, the existence of an empirically validated profiling taxonomy suggests a path toward harmonisation. It is recommended that a reference scheme be implemented, comprising seven to eight categories for fine profiling, in addition to a compact five-category variant for lighter journeys. This would enhance the comparability of services across providers and mitigate investor confusion.

A concrete supervisory instrument would be a supervisory profiling taxonomy and reporting standard (SPTRS) for robo-advisory onboarding. The core of the schema is comprised of an eight-category core profiling schema (CPS-8) that fixes common headings and minimum content. These include objectives and purpose; investment horizon and age; amount and frequency of deposits; risk preferences; sustainability (ESG) preferences; other preferences (e.g., regions/sectors); knowledge and experience; and financial situation (income, assets, liabilities, and liquidity buffers). In the context of streamlined digital journeys, the use of a compact core profiling schema (CCPS-5) facilitates the aggregation of these elements into five distinct categories: objectives/horizons, risk preferences, knowledge/experience, financial capacity, and ESG. This approach

ensures the preservation of comparability while concurrently mitigating redundancy and complexity.

The standard would specify measurement rules rather than prescribe any house view. Each category would contain at least one scored item; the risk module would rely on a validated psychometric instrument with disclosed scoring and reliability; contradiction and completeness checks would be embedded (for example, high risk tolerance paired with a very short horizon); and re-profiling would be versioned and event-triggered as well as periodic. The development of a machine-readable data dictionary is imperative for standardising item wording, response scales, and allowable transformations. Client risk grades would be explicitly mapped to product risk labels (e.g., SRRI/SRI) within the advice summary. ESG would be elicited as sub-preferences, such as exclusions, thematic tilts and impact, linked to instrument filters or tilts. A short justification is recorded whenever a preference cannot be honoured.

To establish a framework of accountability, providers would be required to submit a quarterly, machine-readable profiling report. This report would need to cover category coverage, the use of validated scales, contradiction-flag rates and resolutions, ESG honour rates, reprofiling cadence, and the logic that maps profiles to risk buckets and eligible instruments. The logic would be reflected in a one-page client profile statement, expressed in straightforward language. The SPTRS would thus replace heterogeneous questionnaires with a common taxonomy, validation layer, and disclosure protocol, improve like-for-like assessments across robo-advisory firms, reduce investor confusion, and leave ample scope for innovation in how providers operationalise each module and integrate it with their portfolio engines.

It is evident that a significant number of platforms operationalise risk as equity exposure bands. Therefore, it is crucial to explicitly link this harmonised profiling to the numerical risk labels already present in product disclosures. This could be, for instance, the SRRI for UCITS and the SRI for PRIIPs. This would ensure that willingness and capacity are translated into instrument eligibility and automatic exclusion of products whose risk class exceeds the client's profile. Supervisory expectations should also acknowledge the operational logic of event-triggered rebalancing. In circumstances where profile revisions, deposits or

withdrawals, or market regime signals trigger decisions, the rules governing tolerance bands, turnover ceilings, and tax consequences should be disclosed and auditable. Additionally, given that LLM-enabled advice introduces output variability, it is recommended that a minimal compliance envelope encompass immutable audit logs that capture the rationale behind allocation choices, pre-trade validations that enforce mass balance and instrument constraints, and built-in explanation routines whenever the system declines to implement a user preference, such as a pure-ESG portfolio.

Drawing upon the findings of the performance analysis outlined in this thesis, regulators and standard-setters are advised to incorporate performance considerations into investor protection guidelines. A thorough examination of the value propositions and performance claims articulated by RAs themselves is also recommended. Standardising the benchmark for equity exposure and requiring disclosure for all RA portfolios prior to client investment would facilitate greater transparency and enhanced investor protection. In addition, the proposed legislation would encourage competitiveness and value measured in terms of performance, as it would allow RAs with a proven track record of strong performance to demonstrate their competitive advantage.

Conversely, RAs with a documented history of underperformance would be motivated to innovate and optimise their strategies for performance to maintain their competitive standing in the market. Accordingly, RAs would be precluded from self-proclaimed designation as the “best robo-advisor” whilst employing pseudo-benchmarks that fail to reflect the composition of their portfolios, with a view to misleading investors into the erroneous belief of outperformance. This phenomenon can be illustrated by an RA using a benchmark in a bull market that has a lower equity component than their portfolio. Given that equities generally outperform bonds in such markets, the RA portfolio would outperform the benchmark with a lower equity component (and a larger bond component), thereby creating the impression of genuine outperformance. It constitutes yet another reason why the disclosure of risk-adjusted performance is important, as almost all RAs analysed for the performance comparison highlight charts and returns during investor profiling and on websites, whereas drawdowns, Sharpe ratios or CAPM metrics are rare.

Hence, market discipline will likely push toward clearer disclosure of benchmark selection, live publication of drawdown and recovery statistics, and price pressure on offerings that do not demonstrably improve a client's risk–return position within a given equity group. Regulators, industry bodies, and consumer advocates can accelerate comparability by standardising reporting along equity-exposure buckets and by recommending common cost-inclusive performance templates.

5.2.2. For Robo-Advisory Clients and Investors

For clients and investors, the practical message is that the questionnaire is not a formality but rather the core input of the advice engine. Precision and internal consistency in answers have been shown to have a significant impact on allocations, particularly with regard to risk appetite, time horizons and balance-sheet buffers. Clients should expect and demand a clear mapping from their inputs to the recommended portfolio, explicit statements of how product risk labels were applied, and transparent rationales for any partial implementation of sustainability preferences. As equity shares remain the prevailing risk instrument, investors are advised to meticulously scrutinise whether extreme risk declarations are precipitating substitutions into higher-volatility alternatives.

Moreover, investors should enquire about the ramifications of these substitutions on downside risk and insist on straightforward, unambiguous explanations. Hybrid human-in-the-loop operating models merit particular consideration for complex cases characterised by multiple objectives, narrow emergency reserves, or high leverage, as the integration of an automated scale and expert review more effectively addresses conflicting signals than automation alone.

Furthermore, investors are advised to consider whether it constitutes a viable alternative for the construction of their self-built passive ETF portfolio (“do-it-yourself”), predicated on the profiling of RAs. It should be noted that providers characteristically combine a limited number of financial instruments, primarily focusing on broadly diversified global equity and bond ETFs. Investors can therefore also save on the costs of RAs. Clients should also verify whether their responses are having the anticipated effect on the suggested portfolio. For instance, in the context of sustainability preferences, numerous RAs do not implement these

correctly. Investors are recommended to tend to prefer RAs that have longer questionnaires with more questions, as the additional data is likely to result in a better capture of their profile. However, it remains to be seen how well the profile is translated into a tailor-made portfolio.

For clients and their advisors, the findings of the performance analysis results suggest the need for more rigorous due diligence. The first step is to consider performance in the context of the matching equity bucket rather than in isolation. Second, risk-adjusted measures and recovery profiles should be prioritised over point returns because the latter are flatter in bull phases and conceal fragility in sell-offs. A third approach involves the examination of live or audited beta and correlation measures. It is considered uneconomical to pay an additional platform fee for a portfolio that mostly tracks a low-cost benchmark yet experiences deeper losses. Where a provider publishes risk-adjusted performance measures, investors can use proximity to the benchmark as an indicator of efficiency. In instances where such transparency is lacking, investors are advised to request it or to favour providers that supply it. Finally, in times of crisis, clients should expect to see a coherent explanation for the sequence of rebalancing decisions and, ideally, evidence that the rules were implemented in a manner that buy into weakness and trim into strength rather than the reverse.

5.2.3. For Robo-Advisory Providers

For RA providers, the implications are operational as well as methodological. Risk elicitation should move from idiosyncratic items to validated psychometrics. The implementation of this process will result in the tightening of the dispersion observed in the risk cluster. Furthermore, it will enhance stability over time and strengthen the link to product risk bands. The formal φ - ψ -control architecture developed in the thesis provides a concrete blueprint. Maintaining a versioned profile vector, computing target weights as a function of the profile and the market state, and rebalancing on events under a cost-aware objective together allow providers to honour profiling information while controlling turnover and taxes. In parallel, LLM-specific guardrails need to be institutionalised. Strict constraints must enforce allocations that sum to one, whitelists must limit instruments to those

permitted by the mandate, and asset-class caps must prevent unbounded risk concentrations.

Because the discussion highlighted partial translation of ESG preferences into portfolios, providers should connect the sustainability section of the questionnaire to concrete allocation switches and disclose the performance, cost, and tracking-error implications of ESG choices relative to conventional baselines. The portfolio set architecture should also evolve. Rather than freezing a handful of model portfolios, providers can adopt the two-layer design: ensure at least one portfolio per equity band from 0% to 100%, then apply overlays for horizon, ESG, and knowledge to produce about 132 meaningful configurations. Where operational constraints make such breadth difficult, the event-driven engine can instantiate bespoke weights at recommendation time without proliferating static SKUs. Finally, hybrid oversight should be designed into workflows. Thresholds that route recommendations for human review, such as contradictory inputs, thin liquidity buffers, high leverage or unusually high alternative-asset weights, are a direct response to the implementation anomalies revealed by the GPT-based experiment, and they align with the literature's call for explainable, accountable AI in finance.

Combining the conclusions of the performance analyses from other studies, some of which also included other regions such as South Korea, the US and the UK, with the findings from this study based on the German market, it can be concluded that the risk-adjusted underperformance of RAs relative to benchmarks does not appear to be specific to any particular region, as independent studies have come to the same or similar conclusions. Rather, based on the findings from Chapter III together with those from Chapter IV, it can be assumed that RAs simply fail to track or even outperform their benchmarks with a low tracking error due to their portfolio strategies and methods.

For the German RA market, the headline implication from the performance analysis is a credibility gap. When a simple two-asset benchmark with transparent, infrequent rebalancing beats the average RA portfolio in both calm and crisis, the burden of proof shifts to providers to show where value is created. During stress episodes, platforms can empirically document controlled drawdowns, and brisk recoveries will attract flows at the expense of those that cannot.

RA platform architectures should be engineered to reduce unnecessary co-movement with the equity-bond benchmark in the same risk bucket. That does not imply indiscriminate market timing; it means pursuing genuine diversification at the same equity share through instruments that reduce volatility and correlation, which lowers beta and enlarges the scope for alpha without changing the investor's nominal risk class. Furthermore, the benchmarks' simple threshold rebalancing outperforms most RA variants in the COVID-19 crisis because it systematises buying low and selling high. Providers should therefore audit their rebalancing logic and, where appropriate, replace calendar-only rebalancing with threshold- or dispersion-aware rules that are explicitly backtested for drawdown behaviour.

In addition, drawdown and recovery metrics must be treated as first-class design constraints, not *ex post* curiosities. Explicit drawdown limits, loss-recovery playbooks, and volatility-targeting overlays would have clipped the left tail in March 2020 and accelerated the climb back toward prior peaks. Beta and correlation should be managed as live risk controls rather than accepted outputs. If monthly beta drifts above target, the engine should be able to pivot toward lower-beta sleeves within the same equity share and toward bond exposures with better shock absorption, all within the portfolio's declared mandate. RA providers should measure, publish, and govern against the risk-adjusted performance measures like mean–variance, Sharpe ratios, distance-to-CML diagnostics used in this thesis. Designing for proximity to the capital market line within each equity bucket disciplines both selection and risk budgeting and creates a transparent yardstick for ongoing improvement.

The rebalancing process should transition from a calendar cadence to a dynamic rule-based system that is both threshold-based and volatility-aware. This system should be subjected to testing for drawdown and recovery consequences under shock scenarios. It is also recommended that other RAPMs, such as those defined in Subsection 4.1.3, be considered, including return-to-drawdown risk (RDR) and annualised return-to-volatility risk (AVR), as well as CAPM indicators such as alpha, beta and Treynor ratios. This approach aims to provide a potentially unbiased picture of benchmarked performance from multiple perspectives for management purposes, with guardrails that trigger small, mandate-consistent adjustments rather than infrequent, large ones. It is recommended that the asset-selection menus within each equity bucket be expanded where necessary to create

alternative low-correlation sleeves. The creation of such sleeves would ensure that the nominal equity share is constant while lowering total portfolio variance. A review of fixed-income building blocks is necessary to assess their duration and credit-beta behaviour in stress, given that the crisis exposed the fragility of generic bond sleeves as crash diversifiers.

Lastly, pricing must be aligned with delivered value. If the strategy is designed to hug a benchmark, a low-fee wrapper is the honest proposition; if the promise is outperformance, the design must include the capabilities that historically correlate with alpha generation in this dataset: disciplined, shock-responsive rebalancing, live control of beta and correlation, and a portfolio architecture that can move along the efficient frontier rather than away from it. Where portfolios behave as high correlation, high beta clones of the benchmark, investors will capture a superior risk-adjusted outcome by holding the benchmark directly at a lower cost; the only defensible response is to either deliver demonstrable alpha at the same risk or deliver the same risk at a lower fee drag.

5.3. LIMITATIONS

5.3.1. Chapter III: Profiling and Portfolios—Data Scope, Disclosure Bias, and Interpretation

The aforementioned conclusions and implications based on the empirical analysis of investor profiling and portfolio recommendations done in Chapter III should be interpreted in consideration of the following limitations, which impose restrictions on the generalisability of the results and necessitate a cautious approach when interpreting the findings. Initially, the market scope was national and time-specific. Given the sector's dynamism, the snapshot may fail to capture subsequent entry, exit, or redesign of offerings, and it does not address practices in other jurisdictions where regulatory requirements, product sets and investor habits differ. To establish external validity, replication across countries and time would be required. As the dataset only includes RAs that were active in Germany at the time of the market scan in July 2024, it is subject to survivorship bias. This means that companies that have left the market, for example through liquidation, mergers and acquisitions or business cessation, are systematically omitted. This selection

process tends to produce upward-biased estimates of typical provider quality and resilience while underestimating downside risks and heterogeneity, since poorly performing or otherwise distressed platforms are less likely to be included in the sample. Therefore, inferences about the prevalence of particular questionnaire and portfolio practices should be interpreted with caution and not extrapolated to the full historical universe of entrants and failures. Incorporating delisted providers or archival records in future work would be necessary to assess the magnitude and direction of this bias more precisely.

Secondly, the data underpinning the questionnaire and portfolio analyses were drawn from publicly observable journeys and disclosures. It is evident that multiple passes were utilised to explore branching logic. However, it is impossible to guarantee that every conditional path or private suitability rule was surfaced, especially in cases where firms gate advanced questions behind funding commitments or identity verification. Although the process of translating German-language items into English is meticulously executed, it introduces a modicum of risk with regard to semantic drift, which has the potential to impact tokenisation and centroid proximity within the framework of the TF-IDF analysis. The labels employed by firms, such as “MPT”, “AI” or “buy-and-hold”, may not fully disclose the intricacies of internal processes; they were regarded as good-faith signals rather than audited statements of methodology.

Thirdly, although there was a strong convergence between inductive coding and TF-IDF clustering, methodological choices in text analytics – such as stop-word lists, n-gram ranges, and minimum document frequencies – have the capacity to influence feature sets and distances. The elevated within-cluster dispersion evident in the categories of risk, experience and financial situation is indicative of authentic linguistic heterogeneity, a phenomenon with the potential to obscure latent subclusters. The coding was performed by a single researcher; inter-rater reliability was not measured and would be a worthwhile addition in future work to mitigate subjective bias. K-means clustering using Lloyd’s algorithm was carried out as part of the research for this thesis to refine the clustering, but it became apparent that the manual effort required to further break down the clusters into too many subclusters using whitelists and blacklists undermined the purpose of more objective text analytics validation (Hamerly & Drake, 2015; Lloyd, 1982; Ostrovsky et al., 2012; Sinaga & Yang, 2020; Xu & Lange,

2019). It would therefore only mean a manual reproduction of the clustering already carried out instead of a triangulated analysis using a TF-IDF method.

Fourthly, the portfolio corpus captured labelled strategies at the time they were revealed to the researcher. Market fluctuations between captures and intermittent tactical overlays may have induced minor variations in weights, thereby compromising the integrity of like-for-like comparisons. The apparent similarity in instrument labels across providers can be deceptive, as it hides crucial differences in tracking indices, fees and ESG construction. These variations may be pertinent in practical applications, but they were not covered in the present mapping exercise.

Fifthly, the formal modelling contribution, while unifying profiling, allocation, and event-driven rebalancing in a stochastic-control frame, stopped short of deriving analytical solutions or establishing convergence properties for specific numerical schemes. Moreover, it is important to note that the system under discussion abstracted from significant production frictions. Such frictions included, but were not limited to, lot-level tax accounting, market impact, slippage, multi-account householding and cross-goal interactions. The framework is intended to serve as a guide for engineers, rather than a comprehensive implementation manual.

Sixthly, the GPT-based RA experiment utilised a synthetic sample of one hundred investors, designed to span a diverse range of profiles. The resulting cross-sectional relationships, particularly the counterintuitive signs on liabilities and net assets, should be regarded as diagnostics of model behaviour rather than as causal estimates of human preferences. The evaluation concentrated on allocation patterns, internal consistency, and correlates of equity exposure. It did not seek to back-test performance or risk properties of the recommended portfolios under common market assumptions. The inherent variability in the output of LLMs gives rise to non-stationarity: the same prompts can result in slightly different allocations when the model is run multiple times or updated, which suggests the importance of continuous monitoring in live settings. Furthermore, during the finalisation of this thesis, OpenAI's latest GenAI model, ChatGPT-5, was released, which far outperforms older versions, such as ChatGPT-4o, in various areas of intelligence testing with new features (Artificial Analysis Inc., 2025; OpenAI, 2025).

This could not be covered in the analysis, which is obvious given the rapid development in the field of GenAI. It is assumed that ChatGPT-5 reduces error rates and, thanks to features such as “thinking”, “reasoning”, “deep research”, etc., offers significantly more application possibilities and better outcomes for the use of this technology in robo-advisory.

Seventhly, equity share is merely a surrogate for portfolio risk. The incorporation of more sophisticated measures, such as the expected shortfall under a shared capital market simulator, would undoubtedly add depth to the present study. However, such measures were beyond the scope of the present study. These limitations do not compromise the fundamental contributions; rather, they delineate the domain within which the findings are most reliable and indicate where further research is required before policy or engineering choices can be made.

Eighthly, the applied FHBSS method applied in Subsection 3.2.5 was implemented with coefficient t-values, HC3 robust standard errors, and residual checks (Shapiro–Wilk, Jarque–Bera, Breusch–Pagan, White, and max Cook’s D), which mitigates many routine pitfalls of OLS screening. The filtering function $\mathcal{S}$ (significance) requires all coefficients to be individually significant at $\alpha = 2.5\%$ using OLS standard errors and robust HC3 errors. The rule is stringent, eliminating models whose covariates are jointly important yet individually imprecise, as it guards against spurious inclusions, but is vulnerable to post-selection multiplicity, because the same α is applied after searching hundreds and thousands of models (depending on k), and OLS SEs can understate uncertainty under heteroskedasticity. Searching in a large model space renders nominal p-values anti-conservative and can inflate apparent effect sizes (the “winner’s curse”). This is a critical limitation that must be taken into account when interpreting the results. The classical t-tests and p-values are predicated on the assumption that a single model has been specified a priori and that precisely one set of tests has been run, as opposed to the alternative of testing hundreds of variable combinations and then selecting the “best” one.

Moreover, insisting on universal individual significance risks false negatives under collinearity; variables may be jointly informative while failing individual tests. The multicollinearity filter $\mathcal{V}$ ($VIF < 4$) meaningfully constrains multicollinearity, and starting from low k is methodologically sound because VIF

typically escalates with higher k . However, the hard threshold is still somewhat arbitrary, and the hierarchical path can prune models whose effects emerge only when additional, correlated variables are included. The hierarchical low- k strategy introduces path dependence, meaning that models whose effects materialise only jointly at higher k be pruned early, and in combination with the requirement that all coefficients be significant at $\alpha = 2.5\%$ increases the risk of false negatives under collinearity. The OLS main-effects specification cannot rule out omitted-variable bias or functional-form misspecification, which robust errors do not resolve.

The filter $\mathcal{R}$ (monotonic adjusted R^2 improvement) provides an in-sample fit screen, yet remains sensitive to noise and does not guarantee out-of-sample generalisability; the stopping rule may thus be myopic. The filter $\mathcal{C}$ (AIC/BIC) penalises complexity but is applied after prior in-sample filters, leaving model uncertainty unresolved when near-ties occur. All criteria are evaluated in-sample; without cross-validation or a hold-out set, out-of-sample performance and selection stability cannot be guaranteed. In short, despite robust errors and extensive diagnostics enhancing reliability, the combination of multiple search, in-sample criteria, and path dependence constitutes the principal, remaining limitations.

5.3.2. Chapter IV: Performance—Sampling Frame, Benchmark Design, and Inference

The empirical design of the methodology in Chapter IV is intentionally transparent and conservative; however, several limitations restrict the scope of inference. Firstly, as it consists of monthly observations, the dataset cannot fully capture intra-month path dependence, peak-to-trough dynamics, or transaction timing. Consequently, maximum drawdowns and recovery times are measured on a monthly basis, which tends to understate true extremes and can obscure the precise sequencing of rebalancing responses in March 2020.

Secondly, the sample comprises 16 providers and 86 portfolios for the COVID-19 window, falling to 12 providers and 62 portfolios for the full period because several offerings launched later. While this amount is substantial for a manually curated universe with publicly verifiable index series, it remains a subset of the German market and excludes products lacking continuous, gap-free data.

Constraints of this kind in the selection process raise the possibility of survivorship and disclosure bias, since underperforming portfolios are more likely to be delisted or publish only chart images rather than downloadable tables, which complicates extraction and verification. Platforms that exited, discontinued portfolios or ceased disclosure between January 2017 and February 2022 are systematically omitted, which introduces survivorship bias. Consequently, the reported risk-adjusted outcomes characterise the surviving cross-section rather than the full population of strategies launched during this period. Consequently, average performance may be overstated, and return volatility and drawdown risk may be understated. All inferences should therefore be interpreted as being conditional upon survival and public disclosure at the time of the data freeze in March–May 2022. In several cases, historical series had to be manually copied from interactive charts despite repeated checks, and the possibility of transcription errors cannot be entirely eliminated.

Thirdly, the self-constructed benchmark design is another limitation. The performance study uses a two-asset, hedged global equity and global bond construct with threshold rebalancing to reflect the core exposures prevalent in the portfolios themselves. The choice of the MSCI ACWI equity sleeve and a EUR-hedged global aggregate bond sleeve, together with an assumed ETF-level cost, is defensible but not unique. Other reasonable choices, such as unhedged equity, regionally tilted equity, shorter-duration bonds or the inclusion of commodities, would produce slightly different efficient frontier shapes and could alter small-sample rankings within specific categories, although the general pattern of average underperformance is unlikely to reverse. Portfolios are assigned to benchmark groups using stated target equity shares and holdings analysis. However, providers may drift or revise these targets over time. A fully time-varying mapping was not feasible on monthly data without exhaustive position histories.

Fourthly, the CAPM framework is used parsimoniously to ensure clear interpretation. It does not attempt multi-factor attribution to separate market beta from size, value, or momentum tilts. A richer factor model might explain some of the variation in outcomes and improve the economic interpretation of alpha. That said, multiple comparisons across buckets and metrics increase the risk of coincidental findings of outperformance. No portfolio or provider is deemed “superior” based on a single metric in a single bucket.

Fifthly, the sample ends in February 2022, thus avoiding the structural break at the onset of the war in Ukraine and the subsequent global rate-hiking cycle. While this arrangement preserves internal comparability, the chapter does not cover the subsequent period of stagflationary and high-interest rate regime dynamics in which bond sleeves (fixed-income securities) behaved very differently to the pre-2022 regime. It is conceivable that market conditions, product ranges, and individual algorithms have undergone evolution since that time. There are also RAs that have removed their financial product prices from publication after 2022, which means that structural breaks may occur if an expanded data basis is included. Nevertheless, these limitations suggest that alternative time frames for certain RAs may offer more favourable outcomes. Extending the dataset to include this phase would undoubtedly be informative and could alter the relative standings within specific buckets.

Sixthly, a further limitation of this study concerning the causes of underperformance is that it does not provide precise information regarding the timing and reasons behind the performance of specific portfolios managed by a particular RA during a given period. For this purpose, it would be ideal to disclose all portfolio positions and transactions, as well as the portfolio's closing prices for the day. If the portfolios are constructed exclusively with exchange-traded products, it would be necessary to obtain the historical data directly from external sources, along with all the portfolio positions and transactions, on a daily basis at the very least. Such an approach would allow the portfolio to be replicated daily and examined in greater detail. However, this is challenging to impossible for many RAs due to the limited data available and the low level of disclosure by RA providers. RAs, therefore, remain a black box, with investors needing a fair amount of luck to beat the benchmark.

5.4. OUTLOOK AND IMPLICATIONS FOR RESEARCH

The results and the discussion suggest several avenues for future work that are both practically relevant and scientifically tractable. The first strand is comparative and longitudinal. In order to ascertain whether the German patterns, expanded KYC scope, equity-band-centric menus, partial ESG translation and nascent AI adoption are universal or idiosyncratic, it would be necessary to

replicate the questionnaire census and portfolio-menu mapping in other jurisdictions and repeat the exercise over time. The establishment of a panel of providers would facilitate the linkage of alterations in profiling depth or menu breadth to subsequent outcomes such as client acceptance, churn and realised risk taking.

A secondary strand pertains to psychometrics and risk capacity. The integration of validated instruments, such as GL-RTS or DOSPERT, directly into the profiling map, with subsequent calibration of their scores to product risk classes, has the potential to reduce noise and assist in fulfilling suitability duties. The integration of capacity proxies, including income stability, debt-service ratios, liquidity buffers and collateral, could be achieved through the implementation of hard constraints or cost terms in the control objective. Randomised controlled trials could ascertain the efficacy of such enhancements in reducing misclassification and complaints compared to conventional one- or two-item risk questions, as well as the extent to which they modify the distribution of equity exposures.

A third strand is of an architectural nature. In lieu of the propagation of static model portfolios, providers have the option of adopting dynamic instantiation. This involves the implementation of an event-driven control engine to compute bespoke weights for each recommendation and subsequently projecting these onto a limited number of tradable sleeves that reflect core exposure, ESG tilts and liquidity. Field experiments could compare the client's comprehension and satisfaction with the current status quo of fixed menus. Concurrently, the two-layer catalogue design proposed here—comprising eleven equity cores with overlays for horizon, ESG and knowledge—could be implemented as a pragmatic lower bound and evaluated for operational feasibility and client understanding.

A fourth strand focuses on the implementation of individual investment preferences in allocations and the associated accountability mechanisms. These include, for example, investment preferences in specific sectors, regions, asset classes and sustainable investments. For instance, research should design, implement, and test mechanisms that ensure sustainability and bind the allocation in ways clients can verify, including pure-ESG variants, best-in-class tilts, and explicit exclusions. In cases where the system does not provide a pure-ESG portfolio, it should be obligated to furnish a concise, auditable rationale that addresses liquidity, concentration, cost, and tracking-error considerations. The

repercussions of such “explanation triggers” for trust, adoption, and long-term adherence merit meticulous examination.

A fifth strand focuses on the integration of guardrails within the framework of LLM-enabled advice. A standard battery of pre-trade validations, including mass balances, instrument eligibility, asset class caps, and suitability filters, should be paired with post-hoc explanations that cite specific answers and decision rules. Bias audits and model-drift monitoring require operational definitions and key performance indicators. A library of adversarial profiles, such as those characterised by contradictory preferences, thin emergency buffers, and high leverage, could be constructed for the purpose of stress testing behaviour, refining triage thresholds for human review, and comparing models. Future work should therefore test dynamic rebalancing with scenario-prompt libraries (Ko & Lee, 2024), integrate progressive explainable AI (XAI) (heat maps, sensitivity plots, risk-trajectory visualisations) with literacy-aware tutorials (Fatima & Chakraborty, 2024; Yue et al., 2023; Zhu et al., 2023, 2024), and evaluate HITL workflows that document when and why human overrides occur (Khan et al., 2025; Romanko et al., 2023). Attention to anthropomorphic cues should remain cautious: they can encourage engagement among prevention-focused users but do not substitute for substantive transparency (Baek & Kim, 2023).

A sixth strand pertains to the subjects of performance and risk under common assumptions. In order to facilitate a fair comparison of methods, it is necessary to benchmark allocation engines (i.e., mean–variance, Black–Litterman, factor-based and machine-learning surrogates) within the same formal chassis and under a shared market simulator. The introduction of transaction-cost models, slippage and tax-lot constraints would facilitate the determination of the true cost of event-driven versus periodic rebalancing, thereby assisting in the calibration of tolerance bands. The extension of the framework to multi-account, multi-goal households would serve to test the hypothesis that simple equity-band decomposition remains adequate in the presence of debt schedules, insurance and human-capital risk within the state space.

Another strand is governance engineering. It is recommended that advice provenance be elevated to a first-class object. This would entail capturing prompts, intermediate states, constraint activations and final allocations in tamper-evident

logs. Such logs would support both internal audit and supervisory review. The implementation of open artefacts has the potential to expedite progress in this domain. The creation of an anonymised corpus of profiling questions with category labels, the development of a reference schema for the profile vector, and the establishment of a baseline code for the dynamic control engine under stylised market models would facilitate reproducibility and incremental improvement. When considered collectively, these strands propel the field from static menus and opaque mappings to dynamic, understandable, event-aware advisory systems that are demonstrably suitable, auditable, and responsive to investor preferences and constraints.

Based on the findings of the performance analysis, future work could enhance external validity and increase diagnostic power in four areas. The first is temporal expansion and regime segmentation. Extending the dataset to cover the period up to 2025 would allow the inflation shock and rate-hike cycle to be incorporated, enabling more regime comparisons and post-Ukraine tightening. With this extended timeframe, rolling windows could track time-varying Sharpe ratios, betas, and distances to the capital market line. Persistence tests could also determine whether any provider exhibits consistent outperformance rather than sporadic spikes. The second axis is granularity. Where available, daily data would refine the measurement of drawdown and recovery and allow event-study designs around rebalancing dates. If providers or data vendors could supply position-level histories, holdings-based decomposition could separate the effects of allocation, selection and timing and attribute slippage to specific decisions rather than to an undifferentiated “engine”.

A third avenue is model richness. Replacing the single-factor CAPM with multi-factor specifications, such as those developed by Fama, French and Carhart, would enable us to test whether apparent alpha is in fact compensation for factor tilts and whether underperformance is concentrated in portfolios with unrewarded exposures. Parallel risk-adjusted metrics that emphasise downside, including the Sortino ratio, omega ratio and conditional drawdown at risk, would complement the Sharpe ratio-centric lens and might realign rankings in buckets where return distributions are visibly skewed. The mean–variance geometry can also be expanded into a scenario-based frontier, in which volatility and correlation are

conditioned on macro regimes. This allows for a more realistic efficient frontier against which to judge live portfolios.

A fourth direction is experimental improvement. Empirical performance analysis suggests specific levers, such as threshold rebalancing, volatility targeting and beta-and-correlation controls, that can be prototyped and tested out-of-sample using the same equity buckets and instrument universes. A controlled simulation could compare calendar-only rebalancing with threshold rules at several bands, with and without drawdown overlays, and record the consequences for alpha, the Sharpe ratio, and recovery times under shock paths calibrated to March 2020 and 2022–2023. Where providers advertise distinctive construction methods, such as factor sleeves, risk parity, or hedging overlays, stratified comparisons could identify which method families are systematically closer to the capital market line in matched buckets. This raises the legitimate question of the extent to which robo-advisors' strategies have an impact on risk-adjusted performance.

Finally, research can examine pricing and governance as performance drivers. A cross-sectional study that regresses risk-adjusted outcomes on fee stacks, disclosure quality, and the presence of live risk controls would reveal whether investors consistently receive what marketing promises: either benchmark-like behaviour at a benchmark-like cost or demonstrable, risk-aware deviations that justify their expense.

These extensions would transform a static snapshot into a richer, regime-aware map of RA performance, linking construction choices to outcomes over time. This would provide regulators, investors, and providers with the evidence they need to enhance portfolio design and market accountability.

VI – BIBLIOGRAPHICAL REFERENCES

VI - BIBLIOGRAPHICAL REFERENCES

- Abbas, S. K. (2024). AI Meets Finance : The Rise of AI-Powered Robo-Advisors. *Journal of Electrical Systems*, 20(11), 1011–1016. <https://doi.org/10.52783/JES.7359>
- Abdullah, M. Bin. (1990). On a Robust Correlation Coefficient. *Journal of the Royal Statistical Society Series D: The Statistician*, 39(4), 455–460. <https://doi.org/10.2307/2349088>
- Abraham, F., Schmukler, S. L., & Tessada, J. (2019, February 26). Robo-Advisors: Investing Through Machines. *World Bank Research and Policy Briefs No. 134881*. <https://papers.ssrn.com/abstract=3360125>
- Agarwal, S., & Ann, T. C. (2024). Personal Finance Applications of Fintech. In H. K. Baker, G. Filbeck, & K. Black (Eds.), *The Emerald Handbook of Fintech: Reshaping Finance* (pp. 235–248). Emerald Publishing. <https://doi.org/10.1108/978-1-83753-608-520241029/FULL/XML>
- Ahmed, S., Hasan, M. M., & Kamal, M. R. (2022). Russia–Ukraine crisis: The effects on the European stock market. *European Financial Management*. <https://doi.org/10.1111/eufm.12386>
- Ali, H., & Aysan, A. F. (2023). What will ChatGPT Revolutionize in Financial Industry? *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4403372>
- Allie, J., West, D., & Willows, G. (2016). The value of financial advice: An analysis of the investment performance of advised and non-advised individual investors. *Investment Analysts Journal*, 45, S63–S74. <https://doi.org/10.1080/10293523.2016.1201292>
- Al-Obaydy, W. N. I., Hashim, H. A., Najm, Y. A., & Jalal, A. A. (2022). Document classification using term frequency-inverse document frequency and K-means clustering. *Indonesian Journal of Electrical Engineering and Computer Science*, 27(3), 1517–1524. <https://doi.org/10.11591/ijeecs.v27.i3.pp1517-1524>
- Altrock, S., Mention, A. L., & Aas, T. H. (2023). Being Human in the Digitally Enabled Workplace: Insights From the Robo-Advice Literature. *IEEE*

- Transactions on Engineering Management*.
<https://doi.org/10.1109/TEM.2023.3291820>
- Apesteguia, J., Oechssler, J., & Weidenholzer, S. (2020). Copy Trading. *Management Science*, 66(12), 5608–5622. <https://doi.org/10.1287/MNSC.2019.3508>
- Arenas-Parra, M., Rico-Pérez, H., & Quiroga-Garcia, R. (2024). The emerging field of Robo Advisor: A relational analysis. *Heliyon*, 10(16), e35946. <https://doi.org/10.1016/J.HELİYON.2024.E35946/ASSET/89FE2AE5-DF2E-4315-92DD-BEEE2FDAC91F/MAIN.ASSETS/GR14.JPG>
- Arnott, A. C., Lucas, A., Culloton, D., Carter, D., Denis, G., Kathman, D., Templeton, E., Tran, L. A., & Gallagher, C. (2023). 2023 Robo-Advisor Landscape - Our take on the digital advice industry and the best options for individual investors. <http://global.morningstar.com/equitydisclosures>
- Artificial Analysis Inc. (2025). LLM Leaderboard - Comparison of over 100 AI models from OpenAI, Google, DeepSeek & others | Artificial Analysis. <https://artificialanalysis.ai/leaderboards/models>
- Arwas, A., & Soleil, K. (2016). Robo-Advice 2.0: The Next Generation. *Journal of Financial Transformation*, 43, 30–36. <https://ideas.repec.org/a/ris/jofitr/1567.html>
- Asness, C. S. (2000). Stocks versus Bonds: Explaining the Equity Risk Premium. *Financial Analysts Journal*, 56(2), 96–113. <https://doi.org/10.2469/FAJ.V56.N2.2347;WGROU:STRING:PUBLICATION>
- Baek, C. (2021). A study on consumer strategy of artificial intelligence service using importance-satisfaction analysis. *Global Business and Finance Review*, 26(2), 110–120. <https://doi.org/10.17549/GBFR.2021.26.2.110>
- Baek, T. H., & Kim, M. (2023). AI robo-advisor anthropomorphism: The impact of anthropomorphic appeals and regulatory focus on investment behaviors. *Journal of Business Research*, 164, 114039. <https://doi.org/10.1016/J.JBUSRES.2023.114039>
- Bafna, P., Pramod, D., & Vaidya, A. (2016). Document clustering: TF-IDF approach. *International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT 2016*, 61–66. <https://doi.org/10.1109/ICEEOT.2016.7754750>

- Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLOS ONE*, 12(7), e0180944. <https://doi.org/10.1371/JOURNAL.PONE.0180944>
- Barile, D., Secundo, G., Mariani, M., & Brandonisio, A. (2025). A new era: managing green investments through Robo-Advisors. *Management Decision*. <https://doi.org/10.1108/MD-06-2024-1268/1259057/A-NEW-ERA-MANAGING-GREEN-INVESTMENTS-THROUGH-ROBO>
- Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/J.INFFUS.2019.12.012>
- Beketov, M., Lehmann, K., & Wittke, M. (2018). Robo Advisors: quantitative methods inside the robots. *Journal of Asset Management*, 19(6), 363–370. <https://doi.org/10.1057/s41260-018-0092-9>
- Belleflamme, P., Omrani, N., & Peitz, M. (2015). The economics of crowdfunding platforms. *Information Economics and Policy*, 33, 11–28. <https://doi.org/10.1016/J.INFOECOPOL.2015.08.003>
- Bhansali, V., & Hsu, J. (2012). The Risk in Risk Parity: A Factor Based Analysis of Asset Based Risk Parity. *Article in The Journal of Investing*, 21(3), 102–110. <https://doi.org/10.3905/joi.2012.21.3.102>
- Bholowalia, P., & Kumar, A. (2014). EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. *International Journal of Computer Applications*, 105(9), 975–8887.
- Bianchi, M., & Briere, M. (2021). Robo-Advising: Less AI and More XAI? In *SSRN Electronic Journal*. Elsevier BV. <https://doi.org/10.2139/SSRN.3825110>
- Bisong, E. (2019). Introduction to Scikit-learn. In E. Bisong (Ed.), *Building Machine Learning and Deep Learning Models on Google Cloud Platform* (pp. 215–229). Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-4470-8_18
- Blais, A.-R., & Weber, E. U. (2006). A Domain-Specific Risk-Taking (DOSPERT) scale for adultpopulations. *Judgment and Decision Making*, 1(1), 33–47. <https://doi.org/10.1017/S1930297500000334>

- Blaschke, J., & Kriebel, J. (2021). Robo Advisory Customer Groups: Who Requires Advice? *Die Unternehmung*, 75(3), 397–410.
- Bloomberg. (2022). *Bloomberg Fixed Income Indices*. <https://www.bloomberg.com/professional/product/indices/bloomberg-fixed-income-indices/#/ucits>
- Boreiko, D., & Massarotti, F. (2020). How Risk Profiles of Investors Affect Robo-Advised Portfolios. *Frontiers in Artificial Intelligence*, 3, 60. <https://doi.org/10.3389/frai.2020.00060>
- Bose, D., & Srinivasan, S. K. (2024). Artificial Intelligence-Powered Competency Framework for Robo-Advisory Services in the Era of Digital Economy. In A. Khang (Ed.), *AI-Oriented Competency Framework for Talent Management in the Digital Economy: Models, Technologies, Applications, and Implementation* (pp. 141–152). CRC Press. <https://doi.org/10.1201/9781003440901-9/ARTIFICIAL-INTELLIGENCE-POWERED-COMPETENCY-FRAMEWORK-ROBO-ADVISORY-SERVICES-ERA-DIGITAL-ECONOMY-DEBANJALEE-BOSE-SAKTHI-SRINIVASAN>
- Bougou, W., & Yatié, A. (2022). The impact of the Ukraine–Russia war on world stock market returns. *Economics Letters*, 215, 110516. <https://doi.org/10.1016/j.econlet.2022.110516>
- Bourgeron, T., Lezmi, E., & Roncalli, T. (2018). Robust Asset Allocation for Robo-Advisors. In *SSRN Electronic Journal*. Elsevier BV. <https://doi.org/10.2139/SSRN.3261635>
- Bradbury, T. (2014). Robo-advice is coming : What it means, who will buy it - and why | Professional Planner. *Professional Planner*, 69, 40–41. <https://search.informit.org/doi/abs/10.3316/informit.832903631381744>
- Bräuer, K. (2021). Nudged into Better Portfolios and Lower Risk: Robo-Advice and Savings Decisions. In *SSRN Electronic Journal*. Elsevier BV. <https://doi.org/10.2139/SSRN.3927860>
- Brennan, M. J., Li, F., & Torous, W. N. (2005). Dollar Cost Averaging. *Review of Finance*, 9(4), 509–535. <https://doi.org/10.1007/S10679-005-4999-X>
- Brenner, L., & Meyll, T. (2020). Robo-advisors: A substitute for human financial advice? *Journal of Behavioral and Experimental Finance*, 25, 100275. <https://doi.org/10.1016/J.JBEF.2020.100275>

- Brue, S. L. (1993). Retrospectives: The Law of Diminishing Returns. *Journal of Economic Perspectives*, 7(3), 185–192. <https://doi.org/10.1257/JEP.7.3.185>
- Brüggen, L., Gianni, R., de Haan, F., Hogreve, J., Meacham, D., Post, T., & van der Werf, M. (2025). AI-Based Financial Advice: An Ethical Discourse on AI-Based Financial Advice and Ethical Reflection Framework. *Journal of Public Policy and Marketing*, 44(3), 436–456. https://doi.org/10.1177/07439156241302279/SUPPL_FILE/SJ-PDF-1-PPO-10.1177_07439156241302279.PDF
- Cao, S., Jiang, W., Wang, J., & Yang, B. (2024). From Man vs. Machine to Man + Machine: The art and AI of stock analyses. *Journal of Financial Economics*, 160, 103910. <https://doi.org/10.1016/J.JFINECO.2024.103910>
- Cao, X., De Zwaan, L., & Wong, V. (2025). Building trust in roboadvisory: technology, firm-specific and system trust. *Qualitative Research in Financial Markets*. <https://doi.org/10.1108/QRFM-02-2024-0033/1255881/BUILDING-TRUST-IN-ROBO-ADVISORY-TECHNOLOGY-FIRM>
- Caramanis, C. (2025, January 8). *What are the “Financial Engines”?* Edelman Financial Engines, LLC. <https://www.edelmanfinancialengines.com/education/investment-management/financial-engines>
- Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1), 57–82. <https://doi.org/10.1111/J.1540-6261.1997.TB03808.X>
- Chen, A., Wang, S., Mehta, A. M., Asif, M., Xu, S., & Shahzad, M. F. (2025). FinTech adoption for ESG integration through robo advisors, personalization, and perceived trust. *Scientific Reports*, 15(1), 1–13. <https://doi.org/10.1038/S41598-025-17046-6>;SUBJMETA=172,704,844;KWRD=ENVIRONMENTAL+SCIENCES,ENVIRONMENTAL+SOCIAL+SCIENCES
- Cheng, X., Guo, F., Chen, J., Li, K., Zhang, Y., & Gao, P. (2019). Exploring the Trust Influencing Mechanism of Robo-Advisor Service: A Mixed Method Approach. *Sustainability* 2019, Vol. 11, Page 4917, 11(18), 4917. <https://doi.org/10.3390/SU11184917>

- Chernick, M. R., González-Manteiga, W., Crujeiras, R. M., & Barrios, E. B. (2011). Bootstrap Methods. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 169–174). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_150
- Choi, W. C., & Chang, C. I. (2025). *ChatGPT-5 in Education: New Capabilities and Opportunities for Teaching and Learning*. Preprints. <https://doi.org/10.20944/PREPRINTS202508.0684.V1>
- Christian, H., Mikhael, , Agus, P., & Suhartono, ; Derwin. (2016). Single Document Automatic Text Summarization using Term Frequency-Inverse Document Frequency (TF-IDF). *ComTech: Computer, Mathematics and Engineering Applications*, 7(4), 285–294. <https://doi.org/10.21512/COMTECH.V7I4.3746>
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368), 829–836. <https://doi.org/10.1080/01621459.1979.10481038>
- Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83(403), 596–610. <https://doi.org/10.1080/01621459.1988.10478639>
- Cocca, & Teodoro. (2016). Potential and Limitations of Virtual Advice in Wealth Management. *Journal of Financial Transformation*, 44, 45–57.
- Coşkun, E. A. (2022). The Role of Artificial Intelligence in Investment Decisions and Applications in The Turkish Finance Industry. *Gaziantep University Journal of Social Sciences*, 21(4), 2208–2222. <https://doi.org/10.21547/JSS.1101615>
- Crane, A. D., & Crotty, K. (2018). Passive versus Active Fund Performance: Do Index Funds Have Skill? *Journal of Financial and Quantitative Analysis*, 53(1), 33–64. <https://doi.org/10.1017/S0022109017000904>
- Cremers, K. J. M., Fulkerson, J. A., & Riley, T. B. (2022). Benchmark Discrepancies and Mutual Fund Performance Evaluation. *Journal of Financial and Quantitative Analysis*, 57(2), 543–571. <https://doi.org/10.1017/S0022109021000119>
- Crisanto, J. C., Leuterio, C. B., Prenio, J., & Yong, J. (2024). *FSI Insights on policy implementation No 63 Regulating AI in the financial sector: recent developments and main challenges*. www.bis.org/emailalerts.htm.

- Croux, C., & Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods and Applications*, 19(4), 497–515. <https://doi.org/10.1007/S10260-010-0142-Z/METRICS>
- Cui, X. Y., Li, D., Qiao, X., & Strub, M. S. (2022). Risk and Potential: An Asset Allocation Framework with Applications to Robo-Advising. *Journal of the Operations Research Society of China*, 10(3), 529–558. <https://doi.org/10.1007/S40305-022-00400-0/METRICS>
- D’Acunto, F., Prabhala, N., & Rossi, A. G. (2019). The Promises and Pitfalls of Robo-Advising. *The Review of Financial Studies*, 32(5), 1983–2020. <https://doi.org/10.1093/rfs/hhz014>
- D’Acunto, F., & Rossi, A. G. (2021). Robo-Advising. In R. Rau, R. Wardrop, & L. Zingales (Eds.), *The Palgrave Handbook of Technological Finance* (pp. 725–749). Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-65117-6_26
- D’Acunto, F., & Rossi, A. G. (2023). Robo-Advice: Transforming Households into Rational Economic Agents. *Annual Review of Financial Economics*, 15(Volume 15, 2023), 543–563. <https://doi.org/10.1146/ANNUREV-FINANCIAL-110921-013217/CITE/REFWORKS>
- Dai, M., Jin, H., Kou, S., & Xu, Y. (2021). Robo-advising: a dynamic mean-variance approach. *Digital Finance*, 3(2), 81–97. <https://doi.org/10.1007/S42521-021-00028-4/METRICS>
- Darren, T. (2015). I, Robo-Adviser? Creating the Blended Adviser Experience - ProQuest. *Journal of Financial Planning*, 28(1), 17–19. <https://www.proquest.com/openview/3d4ec8fe4ded3a7aa25d2d82f6778e1f/1?pq-origsite=gscholar&cbl=4849>
- Day, M. Y., Lin, J. T., & Chen, Y. C. (2018). Artificial intelligence for conversational robo-advisor. *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018*, 1057–1064. <https://doi.org/10.1109/ASONAM.2018.8508269>
- de Angelis, L., & Paas, L. J. (2013). A dynamic analysis of stock markets using a hidden Markov model. *Journal of Applied Statistics*, 40(8), 1682–1700. <https://doi.org/10.1080/02664763.2013.793302;JOURNAL:JOURNAL:CJAS20;CTYPE:STRING:JOURNAL>

- Dey, R. K., & Das, A. K. (2023). Modified term frequency-inverse document frequency based deep hybrid framework for sentiment analysis. *Multimedia Tools and Applications*, 82(21), 32967–32990. <https://doi.org/10.1007/S11042-023-14653-1/METRICS>
- D'Hondt, C., De Winne, R., Ghysels, E., & Raymond, S. (2020). Artificial Intelligence Alter Egos: Who might benefit from robo-investing? *Journal of Empirical Finance*, 59, 278–299. <https://doi.org/10.1016/J.JEMPFIN.2020.10.002>
- Dias, F. S. (2025). An Introduction to Robo-Advisors, Notable Fintech Implementations, and Underlying Theory. In A. Zarifis & X. Cheng (Eds.), *Fintech and the Emerging Ecosystems* (pp. 285–300). Springer, Cham. https://doi.org/10.1007/978-3-031-83402-8_11
- Dorfleitner, G., & Hornuf, L. (2019). FinTech Business Models. *FinTech and Data Privacy in Germany*, 85–106. https://doi.org/10.1007/978-3-030-31335-7_5
- Du, K., Zhao, Y., Mao, R., Xing, F., & Cambria, E. (2025). Natural language processing in finance: A survey. *Information Fusion*, 115, 102755. <https://doi.org/10.1016/J.INFFUS.2024.102755>
- Dwork, C., & Minow, M. (2022). Distrust of Artificial Intelligence: Sources & Responses from Computer Science & Law. *Daedalus*, 151(2), 309–321. https://doi.org/10.1162/DAED_A_01918
- Dziawgo, T. (2021). Wealth tech impact on wealth management sector. *EUROPEAN RESEARCH STUDIES JOURNAL*, 24(3B), 141–151. <https://doi.org/10.35808/ERSJ/2463>
- Eickhoff, M., Muntermann, J., & Weinrich, T. (2017). What do FinTechs actually do? A Taxonomy of FinTech Business Models. *ICIS 2017 Proceedings*. <https://aisel.aisnet.org/icis2017/EBusiness/Presentations/22>
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107–115. <https://doi.org/10.1111/J.1365-2648.2007.04569.X>
- EMMI. (2022). *Euribor® Methodology*. <https://www.emmi-benchmarks.eu/benchmarks/euribor/methodology/>
- ESMA. (2009). *Annex to CESR's technical advice on the level 2 measures related to the format and content of KID disclosures for UCITS: methodology for the calculation of the synthetic risk and reward indicator*.

- <https://www.esma.europa.eu/document/annex-cesr%E2%80%99s-technical-advice-level-2-measures-related-format-and-content-kid-disclosures>
- ESMA. (2016). Article 25 Assessment of suitability and appropriateness and reporting to clients | European Securities and Markets Authority. In *European Securities and Markets Authority*. <https://www.esma.europa.eu/publications-and-data/interactive-single-rulebook/mifid-ii/article-25-assessment-suitability-and>
- ESMA. (2023). *Guidelines on certain aspects of the MiFID II suitability requirements*. <https://www.esma.europa.eu/document/guidelines-certain-aspects-mifid-ii-suitability-requirements-1>
- ESMA. (2025). *European Securities and Markets Authority*. <https://www.esma.europa.eu/>
- European Union. (2014). Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU (recast) (Text with EEA relevance). *Official Journal of the European Union*, 57(L 173).
- European Union. (2017). Commission Delegated Regulation (EU) 2017/653 of 8 March 2017. *Official Journal of the European Union*, 60(L 100). https://eur-lex.europa.eu/eli/reg_del/2017/653/oj/eng
- European Union. (2019). Regulation (EU) 2019/2088 of the European Parliament and of the Council of 27 November 2019 on sustainability-related disclosures in the financial services sector (Text with EEA relevance). *Official Journal of the European Union*, 62(L 317). <https://eur-lex.europa.eu/eli/reg/2019/2088/oj/eng>
- Faccarello, G. (2016). Anne-Robert-Jacques Turgot (1727-1781). In G. Faccarello & H. D. Kurz (Eds.), *Handbook on the History of Economic Analysis: Volume I: Great Economists since Petty and Boisguilbert* (Vol. 1, pp. 73–82). Edward Elgar Publishing Ltd. <https://doi.org/10.4337/9781785366642.00018>
- Faloon, M., & Scherer, B. (2017). Individualization of Robo-Advice. *The Journal of Wealth Management*, 20(1), 30–36. <https://doi.org/10.3905/jwm.2017.20.1.030>
- Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business*, 38, 34–105. <https://www.jstor.org/stable/2350752>

- Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 47(2), 427. <https://doi.org/10.2307/2329112>
- Fama, E. F., & French, K. R. (2004). The Capital Asset Pricing Model: Theory and Evidence. *Journal of Economic Perspectives*, 18(3), 25–46. <https://doi.org/10.1257/0895330042162430>
- Faradynawati, I. A. A., & Söderberg, I. L. (2022). Sustainable Investment Preferences among Robo-Advisor Clients. *Sustainability*, 14(19), 12636. <https://doi.org/10.3390/SU141912636>
- Fatima, S., & Chakraborty, M. (2024). Adoption of artificial intelligence in financial services: The case of robo-advisors in India. *IIMB Management Review*, 36(2), 113–125. <https://doi.org/10.1016/J.IIMB.2024.04.002>
- Federal Reserve Bank. (2022). *Federal Reserve Board - Home*. <https://www.federalreserve.gov/>
- Ferri, R. (2014). Now Anyone Can Be a Robo-Advisor. *Journal of Financial Planning*, 27(12), 28–28. <https://www.proquest.com/openview/82ac79c979d6535b0f209d649609646c/1?pq-origsite=gscholar&cbl=4849>
- Fieberg, C., Hornuf, L., Meiler, M., & Streich, D. (2025). *Using Large Language Models for Financial Advice* (11666; Working Paper). <https://doi.org/10.2139/SSRN.5133294>
- Fieberg, C., Hornuf, L., & Streich, D. (2023). *Using GPT-4 for Financial Advice*. https://www.researchgate.net/publication/371789526_Using_GPT-4_for_Financial_Advice
- Fisch, J. E., Labouré, M., & Turner, J. A. (2019). The Emergence of the Robo-advisor. In J. Agnew & O. S. Mitchell (Eds.), *The Disruptive Impact of FinTech on Retirement Systems* (pp. 13–38). Oxford University Press. <https://doi.org/10.1093/oso/9780198845553.003.0002>
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669. <https://doi.org/10.1016/J.EJOR.2017.11.054>
- Foerster, S., Linnainmaa, J. T., Melzer, B. T., & Previtero, A. (2017). Retail Financial Advice: Does One Size Fit All? *The Journal of Finance*, 72(4), 1441–1482. <https://doi.org/10.1111/JOFI.12514>

- Fortagne, M. A., Reith, R., Nguyen, K., & Lis, B. (2023). Technology Acceptance of Neo-Broker Applications: An Empirical Investigation. *International Journal of Innovation and Technology Management*, 20(5). <https://doi.org/10.1142/S0219877023500293>
- Ghosh, B., le Roux, C., & Verma, A. (2020). Investigation of the fractal footprint in selected EURIBOR panel banks. *Banks and Bank Systems*, 15(1), 185–198. [https://doi.org/10.21511/bbs.15\(1\).2020.17](https://doi.org/10.21511/bbs.15(1).2020.17)
- Gijbels, I., & Prosdocimi, I. (2010). Loess. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(5), 590–599. <https://doi.org/10.1002/WICS.104>;WEBSITE:WEBSITE:WIRES;WGROU:STRING:PUBLICATION
- Go, E. J., Moon, J., & Kim, J. (2020). Analysis of the current and future of the artificial intelligence in financial industry with big data techniques. *Global Business and Finance Review*, 25(1), 102–117. <https://doi.org/10.17549/GBFR.2020.25.1.102>
- Gomber, P., Koch, J. A., & Siering, M. (2017). Digital Finance and FinTech: current research and future research directions. *Journal of Business Economics*, 87(5), 537–580. <https://doi.org/10.1007/s11573-017-0852-x>
- Gómez-Martínez, R., & Medrano-Garcia, M. L. (2025). AI technology for developing Bitcoin investment strategies based on altcoin trends. *Sustainable Technology and Entrepreneurship*, 4(1), 100087. <https://doi.org/10.1016/J.STAE.2024.100087>
- González-Estrada, E., & Cosmes, W. (2019). Shapiro–Wilk test for skew normal distributions based on data transformations. *Journal of Statistical Computation and Simulation*, 89(17), 3258–3272. <https://doi.org/10.1080/00949655.2019.1658763>;PAGE:STRING:ARTICLE/CHAPTER
- González-Estrada, E., Villaseñor, J. A., & Acosta-Pech, R. (2022). Shapiro-Wilk test for multivariate skew-normality. *Computational Statistics*, 37(4), 1985–2001. <https://doi.org/10.1007/S00180-021-01188-Y/TABLES/6>
- Google. (2025). *Robo-Advisor*. <https://www.google.com/search?q=Robo-Advisor>
- Grable, J., & Lytton, R. H. (1999). Financial risk tolerance revisited. *Financial Services Review*, 8(3), 163–181. [https://doi.org/10.1016/S1057-0810\(99\)00041-4](https://doi.org/10.1016/S1057-0810(99)00041-4)

- Gupta, P., & Tham, T. M. (2019). Fintech: The new DNA of financial services. In *Fintech: The New DNA of Financial Services*. De Gruyter. <https://doi.org/10.1515/9781547400904-206>
- Hackethal, A., Haliassos, M., & Jappelli, T. (2012). Financial advisors: A case of babysitters? *Journal of Banking & Finance*, 36(2), 509–524. <https://doi.org/10.1016/J.JBANKFIN.2011.08.008>
- Haddad, C., & Hornuf, L. (2019). The emergence of the global fintech market: economic and technological determinants. *Small Business Economics*, 53(1), 81–105. <https://doi.org/10.1007/s11187-018-9991-x>
- Hamerly, G., & Drake, J. (2015). Accelerating Lloyd's Algorithm for k-Means Clustering. In M. E. Celebi (Ed.), *Partitional Clustering Algorithms* (pp. 41–78). Springer, Cham. https://doi.org/10.1007/978-3-319-09259-1_2
- Hao, J., & Ho, T. K. (2019). Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language. *Journal of Educational and Behavioral Statistics*, 44(3), 348–361. https://doi.org/10.3102/1076998619832248/SUPPL_FILE/DS_10.3102_1076998619832248.ZIP
- Hasselwander, M. (2024). Digital platforms' growth strategies and the rise of super apps. *Heliyon*, 10(5). <https://doi.org/10.1016/J.HELİYON.2024.E25856>
- Häusler, A. N., Kuhnen, C. M., Rudorf, S., & Weber, B. (2018). Preferences and beliefs about financial risk taking mediate the association between anterior insula activation and self-reported real-life stock trading. *Scientific Reports*, 8(1), 1–13. <https://doi.org/10.1038/S41598-018-29670-6;TECHMETA=36,59;SUBJMETA=1409,2649,2811,378,477,631;KWRD=DECISION,HUMAN+BEHAVIOUR>
- Havrlant, L., & Kreinovich, V. (2017). A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation). *International Journal of General Systems*, 46(1), 27–36. <https://doi.org/10.1080/03081079.2017.1291635;CTYPE:STRING:JOURNAL>
- Helms, N., Hölscher, R., & Nelde, M. (2022). Automated investment management: Comparing the design and performance of international robo-managers.

- European Financial Management*, 28(4), 1028–1078.
<https://doi.org/10.1111/EUFM.12333>
- Hoechle, D., Ruenzi, S., Schaub, N., & Schmid, M. (2017). The Impact of Financial Advice on Trade Performance and Behavioral Biases. *Review of Finance*, 21(2), 871–910. <https://doi.org/10.1093/ROF/RFW032>
- Hoechle, D., Ruenzi, S., Schaub, N., & Schmid, M. (2018). Financial Advice and Bank Profits. *The Review of Financial Studies*, 31(11), 4447–4492. <https://doi.org/10.1093/RFS/HHY046>
- Horn, M., Schneider, J., & Oehler, A. (2024). Do transactions on social trading platforms predict the stock market behavior of the aggregate private sector? *Finance Research Letters*, 66, 105668. <https://doi.org/10.1016/J.FRL.2024.105668>
- Huber, C., Huber, J., & Kirchler, M. (2021). Market shocks and professionals' investment behavior – Evidence from the COVID-19 crash. *Journal of Banking and Finance*, 133, 106247. <https://doi.org/10.1016/j.jbankfin.2021.106247>
- Hübner, G. (2005). The Generalized Treynor Ratio. *Review of Finance*, 9(3), 415–435. <https://doi.org/10.1007/s10679-005-2265-x>
- Hunter, D., Kandel, E., Kandel, S., & Wermers, R. (2014). Mutual fund performance evaluation with active peer benchmarks. *Journal of Financial Economics*, 112(1), 1–29. <https://doi.org/10.1016/j.jfineco.2013.12.006>
- Iannarone, N. G. (2018). Computer as Confidant: Digital Investment Advice and the Fiduciary Standard. *Chicago-Kent Law Review*, 93. <https://heinonline.org/HOL/Page?handle=hein.journals/chknt93&id=153&div=&collection=>
- Iman, R. L., & Conover, W. J. (1979). The use of the rank transform in regression. *Technometrics*, 21(4), 499–509. <https://doi.org/10.1080/00401706.1979.10489820>
- Innovative Investment Solutions GmbH. (2025). *INNO INVEST*. <https://inno-invest.de/einsatz-von-ki/>
- Isarvest GmbH. (2023). *extraETF*. <https://extraetf.com/de/wissen/robo-advisor-markt-in-deutschland>
- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Resampling Methods. In G. James, D. Witten, T. Hastie, R. Tibshirani, & J. Taylor (Eds.),

- An Introduction to Statistical Learning with Applications in Python* (pp. 201–228). Springer, Cham. https://doi.org/10.1007/978-3-031-38747-0_5
- Jeffers, J., Lyu, T., & Posenau, K. (2021). The Risk and Return of Impact Investing Funds. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3949530>
- Jensen, M. C. (1968). The Performance of Mutual Funds in the Period 1945-1964. *The Journal of Finance*, 23(2), 389–416. <https://doi.org/10.2307/2325404>
- Jensen, M. C., Black, F., & Scholes, M. S. (1972). The Capital Asset Pricing Model: Some Empirical Tests. In M. Jensen (Ed.), *Studies in the Theory of Capital Markets* (pp. 79–121). Praeger Publishers. <https://ssrn.com/abstract=908569>
- Jin, X., & Han, J. (2017). K-Means Clustering. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning and Data Mining* (pp. 695–697). Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7687-1_431
- Jokanovic, B. (2024). Wealthtech: Future Trends and Approaches in Global Wealth Management. *The First International Conference FUTURE-BME 2024 (Forging the Future: Pioneering Approaches in Business, Management and Economics Engineering to Overcome Emerging Global Challenges)*, 544–550. <https://doi.org/10.24867/FUTURE-BME-2024-060>
- Jones, C. L., & Sharpe, W. F. (2010). *The Intelligent Portfolio : Practical Wisdom on Personal Investing from Financial Engines*. John Wiley & Sons, Incorporated. <https://www.wiley.com/en-us/The+Intelligent+Portfolio%3A+Practical+Wisdom+on+Personal+Investing+from+Financial+Engines-p-9780470228043>
- Jung, D., Dorner, V., Glaser, F., & Morana, S. (2018). Robo-Advisory: Digitalization and Automation of Financial Advisory. *Business and Information Systems Engineering*, 60(1), 81–86. <https://doi.org/10.1007/s12599-018-0521-9>
- Jung, D., Dorner, V., Weinhardt, C., & Pusmaz, H. (2018). Designing a robo-advisor for risk-averse, low-budget consumers. *Electronic Markets*, 28(3), 367–380. <https://doi.org/10.1007/s12525-017-0279-9>
- Jung, D., Glaser, F., & Köpplin, W. (2019). Robo-advisory: Opportunities and risks for the future of financial advisory. In *Contributions to Management Science* (pp. 405–427). Springer. https://doi.org/10.1007/978-3-319-95999-3_20
- Kaya, O., & von Martens, T. (2020). *German robo-advisors: March of the machines driving passive investments*. https://www.dbresearch.com/PROD/RPS_EN-

- PROD/PROD0000000000505172/German_robo-advisors:_March_of_the_machines_drivin.pdf?&realload=fI26AZdX8GtkwLyZEOEepP/073YkQEKMN8J3YTKmIfPLeFBrntmxh9i1CRfQkeI
- Kern, A. (2017). Wikifolio: Social Trading. *FinTechs*, 189–198. https://doi.org/10.1007/978-3-658-14187-5_20
- Khan, F., Bartáková, G. P., Almadhor, A., Qayyum, A., Abeer, K., & Durrani, A. (2025). Evaluating the capacity and limitations of generative AI in financial decision making. *Computer Standards & Interfaces*, 93, 103965. <https://doi.org/10.1016/J.CSI.2024.103965>
- Khan, R., Qian, Y., & Naeem, S. (2019). Extractive based Text Summarization Using K-Means and TF-IDF. *International Journal of Information Engineering and Electronic Business*, 11(3), 33–44. <https://doi.org/10.5815/IJIEEB.2019.03.05>
- Khosravi, F. (2024). Transforming Investment Advisory Services Through Artificial Intelligence: A Study on Robo-Advisors and Algorithmic Portfolio Management. *Nuvern Applied Science Reviews*, 8(9), 1–8. <https://nuvern.com/index.php/nasr/article/view/2024-09-04>
- Kim, S. (2019). Robo-Advisor Algorithm with Intelligent View Model. *Journal of Intelligence and Information Systems*, 25(2), 39–55. <https://doi.org/10.13088/JIIS.2019.25.2.039>
- Kinder, G. (2015). Life Planning and the Robo-Adviser Role. *Journal of Financial Planning*, 28(6), 34–35.
- Kirkby, J. L., Mitra, S., & Nguyen, D. (2020). An analysis of dollar cost averaging and market timing investment strategies. *European Journal of Operational Research*, 286(3), 1168–1186. <https://doi.org/10.1016/J.EJOR.2020.04.055>
- Kitces, M. E. (2015). Advisers + Technology: Better than Either Alone? - ProQuest. *Journal of Financial Planning*, 28(1), 30–31. <https://www.proquest.com/openview/fbfebde081a4204d768191c582d82b11/1?pq-origsite=gscholar&cbl=4849>
- Kleverlaan, R., Wenzlaff, K., Zhao, Y., van de Glind, P., & Roux, E. (2021). *Current State of Crowdfunding in Europe*. https://research-portal.uu.nl/files/112544929/CrowdfundingHub_Current_State_of_Crowdfunding_in_Europe_2021.pdf

- Knewton, H. S., & Rosenbaum, Z. A. (2020). Toward understanding FinTech and its industry. *Managerial Finance*, 46(8), 1043–1060. <https://doi.org/10.1108/MF-01-2020-0024/FULL/XML>
- Ko, H., & Lee, J. (2024). Can ChatGPT improve investment decisions? From a portfolio management perspective. *Finance Research Letters*, 64, 105433. <https://doi.org/10.1016/J.FRL.2024.105433>
- Kofman, P. (2025). Scoring the Ethics of AI Robo-Advice: Why We Need Gateways and Ratings. *Journal of Business Ethics*, 198(1), 21–33. <https://doi.org/10.1007/S10551-024-05753-5/FIGURES/3>
- Kordela, D. (2018). Robo-advisors in asset management – the experience from Germany. *European Journal of Service Management*, 28, 151–157. <https://doi.org/10.18276/ejsm.2018.28/1-19>
- Kramer, O. (2016). Scikit-Learn. In O. Kramer (Ed.), *Studies in Big Data* (Vol. 20, pp. 45–53). Springer, Cham. https://doi.org/10.1007/978-3-319-33383-0_5
- Kumar, I. (2025). Evolution of financial advisory market with the advent of robo-advisors. *Managerial Finance*, 51(5), 831–856. <https://doi.org/10.1108/MF-03-2024-0240>
- Kuzniak, S., Rabbani, A., Heo, W., Ruiz-Menjivar, J., & Grable, J. E. (2015). The Grable and Lytton risk-tolerance scale: A 15-year retrospective. *Financial Services Review*, 24(2), 177–192. <https://doi.org/10.61190/FSR.V24I2.3240>
- Kwok, N. (2022). Choosing a WealthTech ecosystem: Where decisions are not simple. *Journal of Digital Banking*, 7(1), 62–69. <https://doi.org/10.69554/iyan7539>
- LAIC Vermögensverwaltung GmbH. (2025). LAIC ADVISOR. <https://laic.de/laic-advisor>
- Lawrenson, J., & Dickason-Koekemoer, Z. (2020). A model for female South African investors' financial risk tolerance. *Cogent Economics and Finance*, 8(1). <https://doi.org/10.1080/23322039.2020.1794493;WGROU:STRING:PUBLICATION>
- Lee, D. K. C. ;, Guan, C. ;, Yu, Y. ;, Kuo, D., Lee, C., Guan, C., Yu, Y., & Ding, Q. (2024). A Comprehensive Review of Generative AI in Finance. *FinTech*, 3(3), 460–478. <https://doi.org/10.3390/FINTECH3030025>

- Lee, I., & Shin, Y. J. (2018). Fintech: Ecosystem, business models, investment decisions, and challenges. *Business Horizons*, 61(1), 35–46. <https://www.sciencedirect.com/science/article/pii/S0007681317301246>
- Lee, S. (2021). Analysis on Robo-Advisor Performance in Korea. In *SSRN Electronic Journal*. Elsevier BV. <https://doi.org/10.2139/SSRN.4345757>
- Liew, J. K. S., & Ajakh, A. (2020). Volatility-Adjusted 60/40 versus 100—New Risk Investing Paradigm. *Journal of Risk and Financial Management* 2020, Vol. 13, Page 190, 13(9), 190. <https://doi.org/10.3390/JRFM13090190>
- Likas, A., Vlassis, N., & J. Verbeek, J. (2003). The global k-means clustering algorithm. *Pattern Recognition*, 36(2), 451–461. [https://doi.org/10.1016/S0031-3203\(02\)00060-2](https://doi.org/10.1016/S0031-3203(02)00060-2)
- Linnainmaa, J. T., Melzer, B. T., & Previtero, A. (2021). The Misguided Beliefs of Financial Advisors. *The Journal of Finance*, 76(2), 587–621. <https://doi.org/10.1111/JOFI.12995>
- Liu, C. Z., Sheng, Y. X., Wei, Z. Q., & Yang, Y. Q. (2018). Research of Text Classification Based on Improved TF-IDF Algorithm. *2018 IEEE International Conference of Intelligent Robotic and Control Engineering, IRCE 2018*, 69–73. <https://doi.org/10.1109/IRCE.2018.8492945>
- Liu, L. X., & Zhang, L. (2008). Momentum Profits, Factor Pricing, and Macroeconomic Risk. *The Review of Financial Studies*, 21(6), 2417–2448. <https://doi.org/10.1093/RFS/HHN090>
- Lloyd, S. P. (1982). Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129–137. <https://doi.org/10.1109/TIT.1982.1056489>
- Lo, A. W., & Ross, J. (2024). Can ChatGPT Plan Your Retirement?: Generative AI and Financial Advice. In *SSRN Electronic Journal*. Elsevier BV. <https://doi.org/10.2139/SSRN.4722780>
- LSEG. (2022). *Refinitiv Eikon*. <https://eikon.thomsonreuters.com/index.html>
- Lu, F., Huang, L., & Li, S. (2023). ChatGPT, Generative AI, and Investment Advisory. *SSRN Electronic Journal*. <https://doi.org/10.2139/SSRN.4519182>
- Maier, T., Menold, J., & McComb, C. (2022). The Relationship Between Performance and Trust in AI in E-Finance. *Frontiers in Artificial Intelligence*, 5, 891529. <https://doi.org/10.3389/FRAI.2022.891529/BIBTEX>

- Maiti, M. (2021). Efficient Frontier and Portfolio Optimization. In M. Maiti (Ed.), *Applied Financial Econometrics* (pp. 89–111). Palgrave Macmillan, Singapore. https://doi.org/10.1007/978-981-16-4063-6_4
- Markowitz, H. M. (1976). Markowitz Revisited. *Markowitz Revisited, Financial Analysts Journal*, 32(5), 47–52. <https://doi.org/10.2469/faj.v32.n5.47>
- Markowitz, H. M. ., Todd, G. Peter., & Sharpe, W. F. . (2000). *Mean-variance analysis in portfolio choice and capital markets*. Frank J. Fabozzi Associates. <https://www.wiley.com/en-us/Mean-Variance+Analysis+in+Portfolio+Choice+and+Capital+Markets-p-9781883249755>
- Marsden, M., Zick, C. D., & Mayer, R. N. (2011). The Value of Seeking Financial Advice. *Journal of Family and Economic Issues*, 32(4), 625–643. <https://doi.org/10.1007/S10834-011-9258-Z/METRICS>
- Marshall, P. S. (2000). A STATISTICAL COMPARISON OF VALUE AVERAGING VS. DOLLAR COST AVERAGING AND RANDOM INVESTMENT TECHNIQUES. *Journal of Financial and Strategic Decisions*, 13(1), 87–99.
- Martínez-Cruz, R., López-López, A. J., & Portela, J. (2025). ChatGPT vs state-of-the-art models: a benchmarking study in keyphrase generation task. *Applied Intelligence*, 55(50), 1–25. <https://doi.org/10.1007/S10489-024-05901-4/METRICS>
- Mateus, I. B., Mateus, C., & Todorovic, N. (2019). Review of new trends in the literature on factor models and mutual fund performance. In *International Review of Financial Analysis* (Vol. 63, pp. 344–354). Elsevier Inc. <https://doi.org/10.1016/j.irfa.2018.12.012>
- Matt, H. (2015). Robo-Advisers: What It Means to You. *Journal of Financial Planning*, 28(2), 26. https://openurl.ebsco.com/EPDB%3Aagcd%3A9%3A10957033/detailv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Aagcd%3A100778214&crl=c&link_origin=scholar.google.com
- Mayring, P. (2015). Qualitative Content Analysis: Theoretical Background and Procedures. In *Approaches to Qualitative Research in Mathematics Education* (pp. 365–380). Springer, Dordrecht. https://doi.org/10.1007/978-94-017-9181-6_13

- Mazur, M., Dang, M., & Vega, M. (2021). COVID-19 and the march 2020 stock market crash. Evidence from S&P1500. *Finance Research Letters*, 38, 101690. <https://doi.org/10.1016/j.frl.2020.101690>
- Meertens, R. M., & Lion, R. (2008). Measuring an Individual's Tendency to Take Risks: The Risk Propensity Scale1. *Journal of Applied Social Psychology*, 38(6), 1506–1520. <https://doi.org/10.1111/J.1559-1816.2008.00357.X>
- Miller, R. E., & Gehr, A. K. (1978). Sample Size Bias and Sharpe's Performance Measure: A Note. *The Journal of Financial and Quantitative Analysis*, 13(5), 943. <https://doi.org/10.2307/2330636>
- Miller, R. L. (2015). Rogers' Innovation Diffusion Theory (1962, 1995). In M. N. Al-Suqri & A. S. Al-Aufi (Eds.), *Information Seeking Behavior and Technology Adoption: Theories and Trends* (pp. 261–274). IGI Global Scientific Publishing. <https://doi.org/10.4018/978-1-4666-8156-9.CH016>
- Minveo AG. (2025). *Minveo*. <https://minveo.de/>
- Moncecchi, G. (2013). *Learning scikit-learn: Machine Learning in Python* Raúl Garreta. www.packtpub.com
- Mondello, E. (2023). Optimal Portfolio. In E. Mondello (Ed.), *Applied Fundamentals in Finance* (pp. 145–185). Springer Gabler, Wiesbaden. https://doi.org/10.1007/978-3-658-41021-6_5
- Mongwe, W. T., Mbuva, R., & Marwala, T. (2025). Bayesian Machine Learning in Quantitative Finance. In *Bayesian Machine Learning in Quantitative Finance*. Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-88431-3>
- Monoyios, M. (2010). Utility-based valuation and hedging of basis risk with partial information. *Applied Mathematical Finance*, 17(6), 519–551. <https://doi.org/10.1080/13504861003650883>;REQUESTEDJOURNAL:JOURNAL:RAMF20;WGROU:STRING:PUBLICATION
- Monoyios, M. (2018). *OPTIMAL INVESTMENT, VALUATION AND HEDGING*. <https://www.maths.ox.ac.uk/system/files/attachments/ovih1%20print%20in%20colour.pdf>
- Montagnani, M. L., Najjar, M. C., & Davola, A. (2024). The EU Regulatory approach(es) to AI liability, and its Application to the financial services

- market. *Computer Law & Security Review*, 53, 105984. <https://doi.org/10.1016/J.CLSR.2024.105984>
- Montmarquette, C. (2015). The Value of Financial Advice. *ANNALS OF ECONOMICS AND FINANCE*, 16(1), 69–94.
- Moritz, A., & Block, J. H. (2016). Crowdfunding: A Literature Review and Research Directions. In D. Brüntje & O. Gajda (Eds.), *Crowdfunding in Europe* (pp. 25–53). Springer, Cham. https://doi.org/10.1007/978-3-319-18017-5_3
- Moulliet, D., Stolzenbach, J., Völker, T., & Wagner, I. (2016). *Robo-Advisory in Wealth Management: Same name, different game: a look at the German Robo-Advisor landscape*. https://www2.deloitte.com/content/dam/Deloitte/de/Documents/financial-services/Robo_No_2.pdf
- MSCI. (2022). *End of day index data search*. MSCI Inc. <https://www.msci.com/end-of-day-data-search>
- Nguyen, N., & Nguyen, D. (2015). Hidden Markov Model for Stock Selection. *Risks*, 3(4), 455–473. <https://doi.org/10.3390/RISKS3040455>
- Nicholson, N., Soane, E., Fenton-O’Creevy, M., & Willman, P. (2005). Personality and domain-specific risk taking. *Journal of Risk Research*, 8(2), 157–176. <https://doi.org/10.1080/1366987032000123856>
- Niszczoła, P., & Abbas, S. (2023). GPT has become financially literate: Insights from financial literacy tests of GPT and a preliminary test of how people use it as a source of advice. *Finance Research Letters*, 58, 104333. <https://doi.org/10.1016/J.FRL.2023.104333>
- Nourallah, M., Öhman, P., & Amin, M. (2023). No trust, no use: how young retail investors build initial trust in financial robo-advisors. *Journal of Financial Reporting and Accounting*, 21(1), 60–82. <https://doi.org/10.1108/JFRA-12-2021-0451>
- Oehler, A., & Horn, M. (2024). Does ChatGPT provide better advice than robo-advisors? *Finance Research Letters*, 60, 104898. <https://doi.org/10.1016/J.FRL.2023.104898>
- Oehler, A., Horn, M., & Wendt, S. (2016). Benefits from social trading? Empirical evidence for certificates on wikifolios. *International Review of Financial Analysis*, 46, 202–210. <https://doi.org/10.1016/J.IRFA.2016.05.007>

- Oehler, A., & Schneider, J. (2023). Social trading: do signal providers trigger gambling? *Review of Managerial Science*, 17(4), 1269–1331. <https://doi.org/10.1007/S11846-022-00560-6/TABLES/12>
- OpenAI. (2025). *ChatGPT* | OpenAI. <https://openai.com/chatgpt/overview/>
- Ostrovsky, R., Rabani, Y., Schulman, L. J., & Swamy, C. (2012). The effectiveness of Lloyd-type methods for the k-means problem. *Journal of the ACM*, 59(6), 28:1-28:22. <https://doi.org/10.1145/2395116.2395117;PAGE:STRING:ARTICLE/CHAPTER>
- Oukhouya, H., Kadiri, H., El Himdi, K., & Guerbaz, R. (2024). Forecasting International Stock Market Trends: XGBoost, LSTM, LSTM-XGBoost, and Backtesting XGBoost Models. *Statistics, Optimization & Information Computing*, 12(1), 200–209. <https://doi.org/10.19139/SOIC-2310-5070-1822>
- Paracampo, M.-T. (2022). Robo-Advisor, consulenza finanziaria e profili regolamentari: quale soluzione per un fenomeno in fieri? *RIVISTA TRIMESTRALE DI DIRITTO DELL'ECONOMIA*, 4(supplemento 1), 256–284. <https://ricerca.uniba.it/handle/11586/195656>
- Pathak, A., & Bansal, V. (2024). AI as decision aid or delegated agent: The effects of trust dimensions on the adoption of AI digital agents. *Computers in Human Behavior: Artificial Humans*, 2(2), 100094. <https://doi.org/10.1016/J.CHBAH.2024.100094>
- Pedregosa, F., Michel, V., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <http://scikit-learn.sourceforge.net>
- Pelster, M., & Val, J. (2024). Can ChatGPT assist in picking stocks? *Finance Research Letters*, 59, 104786. <https://doi.org/10.1016/J.FRL.2023.104786>
- Perrin, S., & Roncalli, T. (2020). Machine Learning Optimization Algorithms & Portfolio Allocation. In E. Jurczenko (Ed.), *Machine Learning for Asset Management: New Developments and Financial Applications* (pp. 262–328). John Wiley & Sons. <https://doi.org/10.1002/9781119751182.ch8>

- Petters, A. O., & Dong, X. (2016). Markowitz Portfolio Theory. In A. O. Petters & X. Dong (Eds.), *An Introduction to Mathematical Finance with Applications* (pp. 83–150). Springer, New York, NY. https://doi.org/10.1007/978-1-4939-3783-7_3
- Phoon, K., & Koh, F. (2018). Robo-advisors and wealth management. *Journal of Alternative Investments*, 20(3), 79–94. <https://doi.org/10.3905/jai.2018.20.3.079>
- Pisoni, G., & Díaz-Rodríguez, N. (2023). Responsible and human centric AI-based insurance advisors. *Information Processing & Management*, 60(3), 103273. <https://doi.org/10.1016/J.IPM.2023.103273>
- Polasik, M., Widawski, P., & Lis, A. (2022). Challenger bank as a new digital form of providing financial services to retail customers in the EU Internal Market. In *The Digitalization of Financial Markets* (pp. 80–100). Taylor & Francis. <https://doi.org/10.4324/9781003095354-5>
- Puhle, M. (2019). The Performance and Asset Allocation of German Robo-Advisors. *Society and Economy*, 41(3), 331–351. <https://doi.org/10.1556/204.2019.41.3.4>
- Puschmann, T. (2017). Fintech. *Business and Information Systems Engineering*, 59(1), 69–76. <https://doi.org/10.1007/s12599-017-0464-6>
- Reher, M., & Sun, C. (2019). Automated Financial Management: Diversification and Account Size Flexibility. *SSRN Electronic Journal*. <https://doi.org/10.2139/SSRN.2872134>
- Renaux, M.-C., Rudnicki, M., Berniere, A., & Lorain, F. (2019). *How will robo-advisors reshape Asset Management?* INITIO (Square Group). <https://de.scribd.com/document/489644491/INITIO-White-Paper-Robo-Advisors-incl-Interview-CFR>
- Republic Europe. (2025). *Invest online in startups via equity crowdfunding* | Republic Europe. <https://europe.republic.com/>
- Restrepo Echavarría, R. (2025). ChatGPT-4 in the Turing Test. *Minds and Machines*, 35(8), 1–10. <https://doi.org/10.1007/S11023-025-09711-6/METRICS>
- Revolut Technologies Inc. (2025). *Robo-Advisor* | Revolut Germany. Revolut Bank UAB. <https://www.revolut.com/en-DE/robo-advisor/>
- Röder, F., & Walter, A. (2019). What drives investment flows into social trading portfolios? *Journal of Financial Research*, 42(2), 383–411. <https://doi.org/10.1111/JFIR.12174>

- Roman, D., Mitra, G., & Spagnolo, N. (2010). Hidden Markov models for financial optimization problems. *IMA Journal of Management Mathematics*, 21(2), 111–129. <https://doi.org/10.1093/IMAMAN/DPP009>
- Romanko, O., Narayan, A., Kwon, R. H., Romanko, O., Narayan, A., & Kwon, R. H. (2023). ChatGPT-based Investment Portfolio Selection. *Operations Research Forum*, 4(91). <https://arxiv.org/abs/2308.06260v1>
- Rossi, A. G., & Utkus, S. P. (2020). Who Benefits from Robo-advising? Evidence from Machine Learning. In *SSRN Electronic Journal*. Elsevier BV. <https://doi.org/10.2139/SSRN.3552671>
- Roy, S. K., Das, N. N., Savita, Bhasin, N. K., Kadyan, S., & Ananthi, R. K. (2024, October 23). Empowering Robo-Advisors: Data-Driven Mutual Fund and Stock Market Price Prediction with Deep Learning Techniques. *2024 3rd International Conference on Electrical, Electronics, Information and Communication Technologies, ICEEICT* 2024. <https://doi.org/10.1109/ICEEICT61591.2024.10718557>
- Rühr, A., Streich, D., Berger, B., & Hess, T. (2019). A classification of decision automation and delegation in digital investment management systems. *Proceedings of the Annual Hawaii International Conference on System Sciences, 2019-January*, 1435–1444. <https://doi.org/10.24251/hicss.2019.174>
- Sarica, S., & Luo, J. (2021). Stopwords in technical language processing. *PLOS ONE*, 16(8), e0254937. <https://doi.org/10.1371/JOURNAL.PONE.0254937>
- Scherer, B. (2017). Algorithmic portfolio choice: lessons from panel survey data. *Financial Markets and Portfolio Management*, 31(1), 49–67. <https://doi.org/10.1007/S11408-016-0282-8/METRICS>
- Scherer, B., & Lehner, S. (2022). Trust me, I am a Robo-advisor. *Journal of Asset Management*, 24(2), 1–12. <https://doi.org/10.1057/s41260-022-00284-y>
- Scherer, B., & Lehner, S. (2025). What drives robo-advice? *Journal of Empirical Finance*, 80, 101574. <https://doi.org/10.1016/J.JEMPFIN.2024.101574>
- Schouten, A. (2018). How a Digital Architecture Can Lead to Tangible Business Results. In *The WealthTech Book* (pp. 97–100). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119444510.CH24>

- Schueffel, P. (2016). Taming the Beast: A Scientific Definition of Fintech. *Journal of Innovation Management*, 4(4), 32–54. https://doi.org/10.24840/2183-0606_004.004_0004
- Schulz, N. (2007). Ladies and gentlemen, start your (financial) engines: small investors bull high, sell low, and pay too little attention to expenses. Nick Schulz explains how a web-based firm, started by Nobel prize-winning economist William Sharpe, is trying to help them. *The American* (Washington, DC), 1(6), 44–46. <https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=19328117&v=2.1&it=r&id=GALE%7CA168820762&sid=googleScholar&linkaccess=fulltext>
- Severino, F., & Thierry, S. (2022). Robo-Advisors: A Big Data Challenge. *Big Data in Finance: Opportunities and Challenges of Financial Digitalization*, 115–131. https://doi.org/10.1007/978-3-031-12240-8_7/COVER
- Shahbaz, N. (2018). WealthTech – Business as Unusual. In S. Chishti (Ed.), *The WealthTech Book* (pp. 19–22). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119444510.CH4>
- Shan, S., Umar, M., & Mirza, N. (2022). Can robo advisors expedite carbon transitions? Evidence from automated funds. *Technological Forecasting and Social Change*, 180, 121694. <https://doi.org/10.1016/J.TECHFORE.2022.121694>
- Sharma, S., & Lertnuwat, N. (2016). THE FINANCIAL CROWDFUNDING WITH DIVERSE BUSINESS MODELS. *Journal of Asian and African Social Science and Humanities*, 2(2), 74–89. <https://www.aarcentre.com/ojs3/index.php/jaash/article/view/66/271>
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), 425–442. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>
- Sharpe, W. F. (1992). ASSET ALLOCATION: MANAGEMENT STYLE AND PERFORMANCE MEASUREMENT An Asset class factor model can help make order out of chaos. *Journal of Portfolio Management*, 18(2), 7–19. <http://www.stanford.edu/~wfsarpe/art/sa/sa.htm>
- Sharpe, W. F. (1998). Morningstar's risk-adjusted ratings. *Financial Analysts Journal*, 54(4), 21–33. <https://doi.org/10.2469/faj.v54.n4.2195>

- Sharpe, W. F., & Markowitz, H. M. (1989). Mean-Variance Analysis in Portfolio Choice and Capital Markets. *The Journal of Finance*, 44(2), 531. <https://doi.org/10.2307/2328607>
- Shephard, R. W., & Färe, R. (1974). The Law of Diminishing Returns. In W. Eichhorn, R. Henn, O. Opitz, & R. W. Shephard (Eds.), *Production Theory - Proceedings of an International Seminar Held at the University at Karlsruhe May – July 1973* (pp. 287–318). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-80864-7_17
- Shou, Y., & Olney, J. (2020). Assessing a domain-specific risk-taking construct: A meta-analysis of reliability of the DOSPERT scale. *Judgment and Decision Making*, 15(1), 112–134. <https://doi.org/10.1017/S193029750000694X>
- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Sironi, P. (2016). FinTech Innovation: From Robo-Advisors to Goal Based Investing and Gamification. In P. Sironi (Ed.), *FinTech Innovation*. Wiley. <https://doi.org/10.1002/9781119227205>
- Smavesto GmbH. (2025). *Smavesto*. <https://www.smavesto.de/>
- So, M. K. P. (2021). Robo-Advising Risk Profiling through Content Analysis for Sustainable Development in the Hong Kong Financial Market. *Sustainability*, 13(3), 1306. <https://doi.org/10.3390/su13031306>
- Song, R., Shu, M., & Zhu, W. (2022). The 2020 global stock market crash: Endogenous or exogenous? *Physica A: Statistical Mechanics and Its Applications*, 585, 126425. <https://doi.org/10.1016/j.physa.2021.126425>
- Spears, T. (2019). Discounting collateral: quants, derivatives and the reconstruction of the ‘risk-free rate’ after the financial crisis. *Economy and Society*, 48(3), 342–370. <https://doi.org/10.1080/03085147.2018.1525153>
- Spedding, V. (2002). Investment management accessible to all. *Quantitative Finance*, 2(3), 172–173. <https://doi.org/10.1088/1469-7688/2/3/601>
- SSGA. (2022). *SPDR MSCI ACWI EUR Hdg UCITS ETF (Acc)*. https://www.ssga.com/de/en_gb/institutional/etfs/funds/spdr-msci-acwi-eur-hdg-ucits-etf-acc-spp1-gy

- Statista. (2025). *Robo-Advisors - Worldwide* | Statista Market Forecast. <https://www.statista.com/outlook/fmo/wealth-management/digital-investment/robo-advisors/worldwide>
- Steinbach, M. C. (2001). Markowitz revisited: Mean-variance models in financial portfolio analysis. *SIAM Review*, 43(1), 31–85. <https://doi.org/10.1137/S0036144500376650>;REQUESTEDJOURNAL:JOURNAL:SIREAD;ISSUE:ISSUE:10.1137/SIREAD.2001.43.ISSUE-1;WGROUPE:STRING:PUBLICATION
- StreetAuthority, L. (2016, August 26). *Profit From The Biggest Trend In The Investing Business* | Nasdaq. StreetAuthority, LLC. https://www.nasdaq.com/articles/profit-biggest-trend-investing-business-2016-08-26?utm_source=chatgpt.com
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, 336(1), 012017. <https://doi.org/10.1088/1757-899X/336/1/012017>
- Syed, F. A., Fang, K. T., Kiani, A. K., Shoaib, M., & Raja, M. A. Z. (2025). Design of Neuro-Stochastic Bayesian Networks for Nonlinear Chaotic Differential Systems in Financial Mathematics. *Computational Economics*, 65(1), 241–270. <https://doi.org/10.1007/S10614-024-10587-4>/METRICS
- Tahvildari, M. (2021). *Forward indifference valuation and hedging of basis risk under partial information*. <https://arxiv.org/pdf/2101.00251>
- Tahvildari, M. (2025a). A Quantitative Performance Analysis of Robo-Advisors in Germany. *Global Business and Finance Review*, 30(8), 1–18.
- Tahvildari, M. (2025b). Analysis of Investment Opportunities in the Gaming Stock Market Using Social Trading. *Risk Governance and Control: Financial Markets & Institutions*, 15(1, special issue), 204–214. <https://doi.org/10.22495/RGCV15I1SIP6>
- Tahvildari, M. (2025c). Analysis of Investor Profiling and Portfolio Recommendations by Robo-Advisors in Germany. *Global Business and Finance Review*, 30(8), 19–33.

- Tahvildari, M. (2025d). Integrating generative AI in Robo-Advisory: A systematic review of opportunities, challenges, and strategic solutions. *Multidisciplinary Reviews*, 8(12), e2025379. <https://doi.org/10.31893/multirev.2025379>
- Tao, R., Su, C. W., Xiao, Y., Dai, K., & Khalid, F. (2021). Robo advisors, algorithmic trading and investment management: Wonders of fourth industrial revolution in financial markets. *Technological Forecasting and Social Change*, 163, 120421. <https://doi.org/10.1016/J.TECHFORE.2020.120421>
- Tertilt, M., & Scholz, P. (2018). To Advise, or Not to Advise—How Robo-Advisors Evaluate the Risk Preferences of Private Investors. *Journal of Wealth Management*, 21(2), 70–84. <https://doi.org/10.3905/jwm.2018.21.2.070>
- Tiberius, V., Gojowy, R., & Dabić, M. (2022). Forecasting the future of robo advisory: A three-stage Delphi study on economic, technological, and societal implications. *Technological Forecasting and Social Change*, 182, 121824. <https://doi.org/10.1016/J.TECHFORE.2022.121824>
- Ticknor, J. L. (2013). A Bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications*, 40(14), 5501–5506. <https://doi.org/10.1016/J.ESWA.2013.04.013>
- Torno, A., Metzler, D. R., & Torno, V. (2021). Robo-What?, Robo-Why?, Robo-How? – A Systematic Literature Review of Robo-Advice. *PACIS 2021 Proceedings*. <https://aisel.aisnet.org/pacis2021/92>
- Tsai, S. C., & Chen, C. H. (2022). Exploring the Innovation Diffusion of Big Data Robo-Advisor. *Applied System Innovation*, 5(1), 15. <https://doi.org/10.3390/ASI5010015>
- Uhl, M. W., & Rohner, P. (2018). Robo-advisors versus traditional investment advisors: An unequal game. *Journal of Wealth Management*, 21(1), 44–50. <https://doi.org/10.3905/JWM.2018.21.1.044>
- Vanguard. (2022). *Global Bond Index Fund - EUR Hedged Acc.* <https://www.de.vanguard/professionell/anlageprodukte/fond/anleihen/9197/global-bond-index-fund-eur-hedged-acc>
- Vilone, G., Sovrano, F., & Lognoul, M. (2024). On the Explainability of Financial Robo-Advice Systems. In L. Longo, S. Lapuschkin, & C. Seifert (Eds.),

- Communications in Computer and Information Science: Vol. 2156 CCIS* (pp. 219–242). Springer, Cham. https://doi.org/10.1007/978-3-031-63803-9_12
- VisualVest GmbH. (2025). *VisualVest*. <https://www.visualvest.de/>
- von Walter, B., Kremmel, D., & Jäger, B. (2022). The impact of lay beliefs about AI on adoption of algorithmic advice. *Marketing Letters*, 33(1), 143–155. <https://doi.org/10.1007/S11002-021-09589-1/FIGURES/2>
- Wang, D., & Nguyen, K. H. (2015). Idiosyncratic risks, bailout, and financial crisis: A copula approach. *Global Business and Finance Review*, 20(1), 53–70. <https://doi.org/10.17549/GBFR.2015.20.1.53>
- Wang, X., & Xu, Y. (2019). An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index. *IOP Conference Series: Materials Science and Engineering*, 569(5), 052024. <https://doi.org/10.1088/1757-899X/569/5/052024>
- Wang, Y., & Guo, Y. (2020). Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost. *China Communications*, 17(3), 205–221. <https://doi.org/10.23919/JCC.2020.03.017>
- Weber, E. U., Blais, A. R., & Betz, N. E. (2002). A Domain-specific Risk-attitude Scale: Measuring Risk Perceptions and Risk Behaviors. *Journal of Behavioral Decision Making*, 15(4), 263–290. <https://doi.org/10.1002/BDM.414;WGROU:STRING:PUBLICATION>
- Weber, P., Carl, K., Valerie, Hinz, O., Oliver, Carl, K. V., Hinz, O., & Weber, P. (2024). Applications of Explainable Artificial Intelligence in Finance—a systematic review of Finance, Information Systems, and Computer Science literature. *Management Review Quarterly* 2023 74:2, 74(2), 867–907. <https://doi.org/10.1007/S11301-023-00320-0>
- West, J. (2012). Financial advisor participation rates and low net worth investors. *Journal of Financial Services Marketing*, 17(1), 50–66. <https://doi.org/10.1057/FSM.2012.4/METRICS>
- Wonglimpiyarat, J., & Yuberk, N. (2005). In support of innovation management and Roger's Innovation Diffusion theory. *Government Information Quarterly*, 22(3), 411–422. <https://doi.org/10.1016/J.GIQ.2005.05.005>

- Xing, F. Z., Cambria, E., & Welsch, R. E. (2018). Natural language based financial forecasting: a survey. *Artificial Intelligence Review*, 50(1), 49–73. <https://doi.org/10.1007/S10462-017-9588-9/METRICS>
- Xu, J., & Lange, K. (2019). Power k-Means Clustering. *Proceedings of the 36th International Conference on Machine Learning*, 6921–6931. <https://proceedings.mlr.press/v97/xu19a.html>
- Yap, B. W., & Sim, C. H. (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12), 2141–2155. <https://doi.org/10.1080/00949655.2010.520163;PAGE:STRING:ARTICLE/CHAPTER>
- Yue, T., Au, D., Au, C. C., & Iu, K. Y. (2023). Democratizing Financial Knowledge with ChatGPT by OpenAI: Unleashing the Power of Technology. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4346152>
- Zhang, D. C., Highhouse, S., & Nye, C. D. (2019). Development and validation of the General Risk Propensity Scale (GRiPS). *Journal of Behavioral Decision Making*, 32(2), 152–167. <https://doi.org/10.1002/BDM.2102>
- Zhang, L., Pentina, I., & Fan, Y. (2021). Who do you choose? Comparing perceptions of human vs robo-advisor in the context of financial services. *Journal of Services Marketing*, 35(5), 634–646. <https://doi.org/10.1108/JSM-05-2020-0162/FULL/XML>
- Zhao, F. F., He, H. J., Liang, J. J., Cen, J., Wang, Y., Lin, H., Chen, F., Li, T. P., Yang, J. F., Chen, L., & Cen, L. P. (2025). Benchmarking the performance of large language models in uveitis: a comparative analysis of ChatGPT-3.5, ChatGPT-4.0, Google Gemini, and Anthropic Claude3. *Eye (London, England)*, 39(6), 1132–1137. <https://doi.org/10.1038/S41433-024-03545-9;SUBJMETA=160,3161,3177,648,692,699,706;KWRD=EDUCATION,UVEAL+DISEASES>
- Zhao, Q., & Fränti, P. (2014). WB-index: A sum-of-squares based index for cluster validity. *Data & Knowledge Engineering*, 92, 77–89. <https://doi.org/10.1016/J.DATAK.2014.07.008>
- Zhu, H., Sallnäs Pysander, E. L., & Söderberg, I. L. (2023). Not transparent and incomprehensible: A qualitative user study of an AI-empowered financial

- advisory system. *Data and Information Management*, 7(3), 100041.
<https://doi.org/10.1016/J.DIM.2023.100041>
- Zhu, H., Vigren, O., & Söderberg, I. L. (2024). Implementing artificial intelligence empowered financial advisory services: A literature review and critical research agenda. *Journal of Business Research*, 174, 114494.
<https://doi.org/10.1016/J.JBUSRES.2023.114494>

VII – APPENDIXES

VII - APPENDIXES

APPENDIX 1. Proof of the TSSE–Global Mean Norm Identity

The aim is to prove the identity (20) between TSSE (18) and the squared norm of the global mean vector $\|\bar{x}\|^2$. Starting with the definition of TSSE (18) leads to

$$\begin{aligned}
 TSSE &= \sum_{i=1}^m \|x_i - \bar{x}\|^2 = \sum_{i=1}^m \sum_{j=1}^n (x_{i,j} - \bar{x}_j)^2 = \sum_{i=1}^m \sum_{j=1}^n (x_{i,j}^2 - 2x_{i,j}\bar{x}_j + \bar{x}_j^2) \\
 &= \sum_{i=1}^m \sum_{j=1}^n x_{i,j}^2 - 2 \sum_{j=1}^n \bar{x}_j \sum_{i=1}^m x_{i,j} + \sum_{j=1}^n \sum_{i=1}^m \bar{x}_j^2 \\
 &= \sum_{i=1}^m \|x_i\|^2 - 2m \sum_{j=1}^n \bar{x}_j^2 + m \sum_{j=1}^n \bar{x}_j^2 = \sum_{i=1}^m \|x_i\|^2 + m\|\bar{x}\|^2,
 \end{aligned}$$

which proves the relation (20).

APPENDIX 2. Blacklist of Tokens for TF-IDF Preprocessing

This appendix lists the comma-separated tokens removed from the raw robo-advisor question data prior to the TF-IDF analysis:

000, 10, 12, 20, want, like, t like, don t, t, g, t, e g, e, 1, 12, 20, 3, 5, 10 000, 2, 7, 8.

APPENDIX 3. Python Script for TF-IDF Question Analysis

```

1. # -*- coding: utf-8 -*-
2. # TF-IDF and Cluster Analysis for Categorisation of Questions
3. """
4. This documented Python script:
5. 1) Builds a TF-IDF representation for 687 Robo-Advisor questions
6.    using 1-4 grams with scikit-learn English stop words plus a
7.    custom blacklist.
8. 2) Computes category centroids, each question's squared Euclidean
9.    distance to its category centroid, WCSSE per category, total
10.   WCSSE, TSSE, BCSSE, and CHI.
11. 3) Saves all results into one Excel workbook, incl. a
12.    "Data_Dictionary" sheet explaining every column.
13.
14. Notes
15. - TF and IDF follow the formulas:
16.   TF(v_j, q_i) = count(v_j, q_i) / sum_u count(u, q_i)

```

```

17. IDF(v_j) = ln(m / (n_j + 1))
18. TFIDF(v_j, q_i) = TF * IDF
19. - n-grams are computed after removing stop words and blacklist
20. terms, mirroring scikit-learn common practice.
21. - TF-IDF is implemented "manually" (no scikit-learn vectorizers).
22. """
23.
24. # -----
25. # 0) Imports (std Python + pandas/numpy; sklearn only for stops)
26. # -----
27. import pandas as pd
28. import numpy as np
29. import math
30. import re
31. from collections import Counter
32. from sklearn.feature_extraction.text import ENGLISH_STOP_WORDS
33.
34. # -----
35. # 1) Load inputs
36. # -----
37. # Expect:
38. # - RA_categories_questions.xlsx
39. #   (sheet "RA_categories_questions" with columns: category, question)
40. # - RA_custom_blacklist.xlsx (first column contains blacklist terms)
41. path_questions = "/data/RA_categories_questions.xlsx"
42. path_blacklist = "/data/RA_custom_blacklist.xlsx"
43.
44. # Read data
45. df_q = pd.read_excel(path_questions, sheet_name="RA_categories_questions")
46. df_bl = pd.read_excel(path_blacklist)
47.
48. # Basic validation
49. assert {"category", "question"}.issubset(
50.     df_q.columns
51. ), "Expected columns: 'category' and 'question'."
52.
53. # -----
54. # 2) Parameters and stop words
55. # -----
56. # TF-IDF and n-gram configuration
57. min_df = 2 # a term must appear in at least 2 documents
58. max_df_prop = 0.8 # a term may appear in at most 80% of documents
59. ngram_min, ngram_max = 1, 4
60.
61. # Combine sklearn English stop words with custom blacklist (first column)
62. custom_blacklist = set(
63.     str(t).strip().lower()
64.     for t in df_bl.iloc[:, 0].dropna().tolist()
65.     if str(t).strip() != ""
66. )
67. combined_stops = set(ENGLISH_STOP_WORDS) | custom_blacklist
68.
69. # -----
70. # 3) Preprocessing and n-grams
71. # -----
72. # Token pattern: alphanumerics only (A-Z, a-z, 0-9)
73. token_pattern = re.compile(r"[A-Za-z0-9]+")
74.
75. def preprocess(text: str) -> list:

```

```

76.     """
77.     Lower-case, tokenise into alphanumerics, remove combined stop
78.     words / blacklist. Return remaining tokens.
79.     """
80.     text = "" if pd.isna(text) else str(text).lower()
81.     tokens = token_pattern.findall(text)
82.     tokens = [t for t in tokens if t not in combined_stops]
83.     return tokens
84.
85. def generate_ngrams(tokens: list, n_min: int = 1, n_max: int = 4) -> list:
86.     """
87.     Build space-joined n-grams for n in [n_min, n_max].
88.     Example: ["investment", "horizon"] ->
89.     "investment", "horizon", "investment horizon".
90.     """
91.     grams = []
92.     L = len(tokens)
93.     for n in range(n_min, n_max + 1):
94.         if L >= n:
95.             grams.extend(
96.                 [" ".join(tokens[i:i+n]) for i in range(L - n + 1)]
97.             )
98.     return grams
99.
100. # Apply preprocessing and n-gram generation
101. tokenized_docs = df_q["question"].apply(preprocess).tolist()
102. docs_terms = [
103.     generate_ngrams(toks, ngram_min, ngram_max) for toks in tokenized_docs
104. ]
105.
106. # -----
107. # 4) Build vocabulary with DF filtering
108. # -----
109. m = len(docs_terms) # number of documents
110. df_counts = Counter()
111. for terms in docs_terms:
112.     # count a term once per document for DF
113.     df_counts.update(set(terms))
114.
115. max_df_abs = math.floor(max_df_prop * m)
116. vocab_terms = [
117.     term for term, dfc in df_counts.items()
118.     if (dfc >= min_df and dfc <= max_df_abs)
119. ]
120. vocab_terms.sort()
121. vocab_index = {t: idx for idx, t in enumerate(vocab_terms)}
122. n_vocab = len(vocab_terms)
123.
124. # -----
125. # 5) Compute IDF and long TF-IDF table
126. # -----
127. idf = np.zeros(n_vocab, dtype=float)
128. df_series = np.zeros(n_vocab, dtype=int)
129. for term, j in vocab_index.items():
130.     n_j = df_counts[term] # document frequency of term
131.     df_series[j] = n_j
132.     idf[j] = math.log(m / (n_j + 1))
133.

```

```

134. records = []
135. for i, terms in enumerate(docs_terms):
136.     # Term counts within document i (for terms that survived filters)
137.     counts = Counter(t for t in terms if t in vocab_index)
138.     if not counts:
139.         continue
140.     total = sum(counts.values()) # denom for TF normalisation
141.     for term, c in counts.items():
142.         j = vocab_index[term]
143.         tf = c / total if total > 0 else 0.0
144.         tfidf = tf * idf[j]
145.         if tfidf == 0:
146.             continue
147.         records.append(
148.             (
149.                 i, # doc_id
150.                 df_q.loc[i, "category"], # category label
151.                 df_q.loc[i, "question"], # original question
152.                 term, # 1- to 4-gram term
153.                 counts[term], # raw count in the document
154.                 total, # sum of token counts in document
155.                 tf, # term frequency
156.                 idf[j], # inverse document frequency
157.                 tfidf # TF-IDF weight w_(j,i)
158.             )
159.         )
160.
161. tfidf_long = pd.DataFrame(
162.     records,
163.     columns=[
164.         "doc_id", "category", "question", "term", "term_count",
165.         "doc_len_terms", "tf", "idf", "tfidf"
166.     ]
167. )
168.
169. # -----
170. # 6) Category centroids  $\mu_k$ 
171. # -----
172. # Sum of TF-IDF per (category, term)
173. cat_term_sum = (
174.     tfidf_long
175.     .groupby(["category", "term"])
176.     .agg(S=("tfidf", "sum"))
177.     .reset_index()
178. )
179. docs_per_cat = (
180.     df_q.groupby("category")["question"]
181.     .count()
182.     .rename("n_docs")
183.     .reset_index()
184. )
185. centroids_long = cat_term_sum.merge(
186.     docs_per_cat, on="category", how="left"
187. )
188. centroids_long["mu_term"] = (
189.     centroids_long["S"] / centroids_long["n_docs"]
190. ) #  $\mu_{(k,j)}$ 
191.
192. # -----

```



```

193. # 7) Norms (squared) of docs and centroids
194. # -----
195. doc_norm2 = (
196.     tfidf_long.assign(tfidf2=lambda d: d["tfidf"]**2)
197.     .groupby("doc_id")["tfidf2"]
198.     .sum()
199.     .rename("doc_norm2")
200.     .reset_index()
201. )
202.
203. centroid_norm2 = (
204.     centroids_long.assign(mu2=lambda d: d["mu_term"]**2)
205.     .groupby("category")["mu2"]
206.     .sum()
207.     .rename("centroid_norm2")
208.     .reset_index()
209. )
210.
211. # -----
212. # 8) Dot product  $x_i \cdot \mu_k$  with its category centroid
213. # -----
214. doc_dot_centroid = (
215.     tfidf_long.merge(
216.         centroids_long[["category", "term", "mu_term"]],
217.         on=["category", "term"],
218.         how="inner"
219.     )
220.     .assign(prod=lambda d: d["tfidf"] * d["mu_term"])
221.     .groupby("doc_id")["prod"]
222.     .sum()
223.     .rename("dot_with_centroid")
224.     .reset_index()
225. )
226.
227. # -----
228. # 9) Squared distances and WCSSE
229. # -----
230. doc_info = (
231.     tfidf_long[["doc_id", "category", "question"]]
232.     .drop_duplicates()
233.     .merge(doc_norm2, on="doc_id", how="left")
234.     .merge(doc_dot_centroid, on="doc_id", how="left")
235.     .merge(centroid_norm2, on="category", how="left")
236. )
237.
238. doc_info["squared_distance_to_centroid"] = (
239.     doc_info["doc_norm2"]
240.     - 2.0 * doc_info["dot_with_centroid"]
241.     + doc_info["centroid_norm2"]
242. )
243.
244. wcsse_by_cat = (
245.     doc_info
246.     .groupby("category")["squared_distance_to_centroid"]
247.     .sum()
248.     .rename("WCSSE_category")
249.     .reset_index()
250. )

```

```

251. WCSSE_total = float(wcsse_by_cat["WCSSE_category"].sum())
252.
253. # -----
254. # 10) TSSE, BCSSE and CHI
255. # -----
256. # Global mean vector  $\bar{x}$  (component-wise):
257. global_term_sum = (
258.     tfidf_long
259.     .groupby("term")["tfidf"]
260.     .sum()
261.     .rename("sum_w_over_docs")
262.     .reset_index()
263. )
264. global_term_sum["xbar_term"] = (
265.     global_term_sum["sum_w_over_docs"] / m
266. )
267. xbar_norm2 = float((global_term_sum["xbar_term"]**2).sum())
268.
269. sum_doc_norm2 = float(doc_norm2["doc_norm2"].sum()) #  $\sum ||x_i||^2$ 
270. TSSE = sum_doc_norm2 - m * xbar_norm2
271. BCSSE = TSSE - WCSSE_total
272.
273. K = int(docs_per_cat.shape[0])
274. if WCSSE_total > 0 and (K - 1) > 0 and (m - K) > 0:
275.     CHI = (
276.         (m - K) * BCSSE
277.     ) / (
278.         (K - 1) * WCSSE_total
279.     )
280. else:
281.     CHI = float("nan")
282.
283. # -----
284. # 11) Assemble tables for export
285. # -----
286. meta_rows = [
287.     ("n_documents (m)", m),
288.     ("n_vocabulary (|V|)", n_vocab),
289.     ("n_categories (K)", K),
290.     ("min_df", min_df),
291.     ("max_df_prop", max_df_prop),
292.     ("ngram_min", ngram_min),
293.     ("ngram_max", ngram_max),
294.     ("stopwords_source",
295.      "scikit-learn ENGLISH STOP WORDS + custom blacklist"),
296.     ("sum( $||x_i||^2$ ) over all docs", sum_doc_norm2),
297.     (" $||\bar{x}||^2$ ", xbar_norm2),
298.     ("TSSE", TSSE),
299.     ("WCSSE_total", WCSSE_total),
300.     ("BCSSE = TSSE - WCSSE_total", BCSSE),
301.     ("CHI = ((m-K)*BCSSE)/((K-1)*WCSSE_total)", CHI),
302. ]
303. meta_df = pd.DataFrame(meta_rows, columns=["key", "value"])
304.
305. vocab_df = pd.DataFrame(
306.     {
307.         "term": vocab_terms,
308.         "document_frequency": df_series,
309.         "df_proportion": df_series / m,

```

```

310.         "idf_ln_m_over_df_plus1": idf,
311.     }
312. ).sort_values(
313.     ["document_frequency", "term"], ascending=[False, True]
314. ).reset_index(
315.     drop=True
316. )
317.
318. centroids_table = centroids_long[
319.     ["category", "term", "n_docs", "mu_term", "S"]
320. ].copy()
321. centroids_table = centroids_table.rename(
322.     columns={"S": "sum_tfidf_over_category"})
323. )
324.
325. category_metrics = (
326.     centroid_norm2
327.     .merge(docs_per_cat, on="category", how="left")
328.     .merge(wcsse_by_cat, on="category", how="left")
329.     .rename(columns={"n_docs": "n_docs_in_category"})
330.     .sort_values("category")
331.     .reset_index(drop=True)
332. )
333.
334. doc_distances = doc_info[
335.     [
336.         "doc_id", "category", "question", "doc_norm2",
337.         "dot_with_centroid", "centroid_norm2",
338.         "squared_distance_to_centroid"
339.     ]
340. ].sort_values(
341.     ["category", "doc_id"]
342. ).reset_index(
343.     drop=True
344. )
345.
346. top_terms_per_category = (
347.     centroids_table
348.     .sort_values(["category", "mu_term"], ascending=[True, False])
349.     .groupby("category")
350.     .head(50)
351.     .reset_index(drop=True)
352. )
353.
354. global_mean_table = global_term_sum[
355.     ["term", "xbar_term"]
356. ].sort_values(
357.     "xbar_term", ascending=False
358. ).reset_index(
359.     drop=True
360. )
361.
362. # -----
363. # 12) Data dictionary (lookup sheet)
364. # -----
365. data_dict_rows = [
366.     ("Sheet", "Column", "Meaning / How to interpret"),
367.     ("meta_and_metrics", "key", "Name of parameter or metric."),

```

```

368.     ("meta_and_metrics", "value", "Value of parameter or metric."),
369.     ("questions", "category", "Assigned category (8 categories)."),
370.     ("questions", "question", "Original questionnaire item."),
371.     ("vocabulary", "term", "n-gram (1-4 tokens) after filters."),
372.     ("vocabulary", "document_frequency",
373.      "Docs where the term appears at least once."),
374.     ("vocabulary", "df_proportion",
375.      "document_frequency divided by m."),
376.     ("vocabulary", "idf_ln_m_over_df_plus1",
377.      "IDF = ln(m / (n_j + 1)); higher = rarer."),
378.     ("tfidf_long", "doc_id", "Zero-based index of the question."),
379.     ("tfidf_long", "category", "Category of the question."),
380.     ("tfidf_long", "question", "Original question text."),
381.     ("tfidf_long", "term", "n-gram found after preprocessing."),
382.     ("tfidf_long", "term_count", "Raw count of the term in question."),
383.     ("tfidf_long", "doc_len_terms",
384.      "Sum of term counts, TF denominator."),
385.     ("tfidf_long", "tf", "Term Frequency = count / doc_len_terms."),
386.     ("tfidf_long", "idf", "IDF = ln(m / (n_j + 1))."),
387.     ("tfidf_long", "tfidf", "TF × IDF weight."),
388.     ("centroids_mu", "category", "Category label k."),
389.     ("centroids_mu", "term", "Vocabulary term (n-gram)."),
390.     ("centroids_mu", "n_docs", "Docs in category |C_k|."),
391.     ("centroids_mu", "mu_term",
392.      "Centroid component  $\mu_{(k,j)}$  (avg TF-IDF)."),
393.     ("centroids_mu", "sum_tfidf_over_category",
394.      "Sum  $S(v_j, C_k)$  before  $|C_k|$ ."),
395.     ("category_metrics", "category", "Category label k."),
396.     ("category_metrics", "centroid_norm2", "Squared norm  $||\mu_k||^2$ ."),
397.     ("category_metrics", "n_docs_in_category", "Docs in  $|C_k|$ ."),
398.     ("category_metrics", "WCSSE_category",
399.      " $\sum_{i \in C_k} ||x_i - \mu_k||^2$ ."),
400.     ("doc_distances", "doc_id", "Zero-based index of question."),
401.     ("doc_distances", "category", "Category of question."),
402.     ("doc_distances", "question", "Original question text."),
403.     ("doc_distances", "doc_norm2", "Squared norm  $||x_i||^2$ ."),
404.     ("doc_distances", "dot_with_centroid",
405.      "Dot product  $x_i \cdot \mu_k$ ."),
406.     ("doc_distances", "centroid_norm2",
407.      "Squared norm  $||\mu_k||^2$ ."),
408.     ("doc_distances", "squared_distance_to_centroid",
409.      " $||x_i - \mu_k||^2 = ||x_i||^2 - 2x_i \cdot \mu_k + ||\mu_k||^2$ ."),
410.     ("top_terms_per_category", "category", "Category label k."),
411.     ("top_terms_per_category", "term", "Vocabulary term (n-gram)."),
412.     ("top_terms_per_category", "n_docs", "Docs in  $|C_k|$ ."),
413.     ("top_terms_per_category", "mu_term",
414.      "Centroid component (avg TF-IDF)."),
415.     ("top_terms_per_category", "sum_tfidf_over_category",
416.      "Sum  $S(v_j, C_k)$  before normalisation."),
417.     ("global_mean_vector", "term", "Vocabulary term (n-gram)."),
418.     ("global_mean_vector", "xbar_term",
419.      "Component of global mean  $\bar{x}$ ."),
420. ]
421. data_dict = pd.DataFrame(
422.     data_dict_rows[1:], columns=data_dict_rows[0]
423. )
424.
425. # -----
426. # 13) Write Excel

```

```

427. # -----
428. out_path = "/data/RA_TFIDF_centroids_metrics_documented.xlsx"
429. with pd.ExcelWriter(out_path, engine="xlsxwriter") as writer:
430.     # Optional README
431.     pd.DataFrame(
432.         [("This workbook documents all interim and final results", "")]
433.         , columns=["note", ""]
434.     ).to_excel(writer, sheet_name="README", index=False)
435.
436.     # Core sheets
437.     meta_df.to_excel(
438.         writer, sheet_name="meta_and_metrics", index=False
439.     )
440.     df_q.to_excel(
441.         writer, sheet_name="questions", index=False
442.     )
443.     vocab_df.to_excel(
444.         writer, sheet_name="vocabulary", index=False
445.     )
446.     tfidf_long.to_excel(
447.         writer, sheet_name="tfidf_long", index=False
448.     )
449.     centroids_table.to_excel(
450.         writer, sheet_name="centroids_mu", index=False
451.     )
452.     category_metrics.to_excel(
453.         writer, sheet_name="category_metrics", index=False
454.     )
455.     doc_distances.to_excel(
456.         writer, sheet_name="doc_distances", index=False
457.     )
458.     top_terms_per_category.to_excel(
459.         writer, sheet_name="top_terms_per_category", index=False
460.     )
461.     global_mean_table.to_excel(
462.         writer, sheet_name="global_mean_vector", index=False
463.     )
464.     data_dict.to_excel(
465.         writer, sheet_name="Data_Dictionary", index=False
466.     )
467.

```

Note. Author's elaboration: This appendix lists the documented Python code used to build a TF-IDF representation of 687 robo-advisor questions (1–4-grams; scikit-learn English stop words plus a custom blacklist from APPENDIX 2), compute category centroids, and derive clustering diagnostics. Term weights follow TF, IDF, and TFIDF = $TF \times IDF$ with m documents and n_j document frequency. For category k with centroid μ_k and documents x_i , the within-category dispersion WCSSE, and total within-category SSE (WCSSE) are computed. Total scatter TSSE with global mean $\bar{x}$, between-category scatter BCSSE = $TSSE - WCSSE$, and the Calinski–Harabasz index (CHI) are calculated. Parameter choices (e.g., N-gram range, min_df, max_df) and all summary values are exported to the workbook. TSSE is invariant across category variants (given the same TF-IDF space), while WCSSE decreases monotonically with K .

APPENDIX 4. Top Five Typical Questions per Category (K = 5)

Category	Questions
Objectives, amount and horizon	What is your investment goal and how long do you want to invest your money? How long and for what do you want to invest your money? How long do you want to invest your money? How long do you want to invest your money for capital accumulation or optimization? How many years do you want to invest your money for?
Risk preferences	Capital markets are prone to fluctuation. What loss of value makes you nervous? With your investment amount of €10,000 over your investment period, there may be a loss of €652. Can you bear this possible capital loss? Do you find large price fluctuations with possible interim losses in the context of this investment make you nervous? Of course you are investing this money to make a profit. But what potential loss could you bear if the portfolio were to lose value? I would like to achieve higher returns and am willing to accept risks to do so.
Sustainability preferences	Please indicate whether environmental, social and good corporate governance issues should be taken into account in recommendations and investment decisions (so-called sustainability preferences). Do you want to invest in environmentally sustainable economic activities that contribute significantly to achieving environmental goals under the Taxonomy Regulation? Should your portfolio include investments in environmentally sustainable economic activities that contribute significantly to achieving environmental goals? Would you like to specify a lower limit for investing in products that promote the further development of environmental and/or social issues through their investment strategy? Would you like to invest in environmentally sustainable economic activities that contribute significantly to achieving environmental goals?

Knowledge and experience	Do you have knowledge and/or experience with foreign currency transactions (e.g. buying stocks/bonds in foreign currency)? Which of the following asset classes do you already have experience with? (Query on various investment instruments such as funds, bonds, stocks, etc.) Do you have knowledge and experience of at least 3 years in the following asset classes? In the knowledge section, experience in years, the number of trades and the level of experience with 7 financial instruments are queried Have you gained relevant capital market experience through training, studies or work?
Financial situation	What is the amount of your current assets (assets less liabilities/financial obligations) in EUR? (cash assets, capital investments, real estate assets) How much of your monthly net income is freely available after deducting your monthly expenses (e.g. living expenses, rent, loan interest, amortization, etc.)? Free monthly liquidity (monthly income minus monthly obligations) How long will your reserves last to cover your current expenses and obligations? We recommend keeping at least two months' salary for unforeseen expenses. How long can you live on your reserves beyond that?

Note. Author's elaboration: The table presents the five questions per category for the categorisation with $K = 5$ categories, with the smallest distance $\text{dist}(i, k)^2$ to the centroid μ_k of the respective category C_k (see Equation (15)). Results are based on TF-IDF clustering of 687 profiling questions from 45 robo-advisors.

APPENDIX 5. Questionnaire for GPT-Based Robo-Advisor

#	Questions	Answer options
Q1	What are your investment objectives?	Wealth accumulation, Retirement planning, Preserve assets, Secure liquidity, Generate passive income, Funding a major purchase (e.g., home, car), Diversification of assets, Hedging against inflation or economic downturns

Q2	How high is your savings target?	in EUR
Q3	Who do you want to invest money for?	Myself, My family, My children's education, Future generations, Charitable causes
Q4	What are your personal financial goals?	Buying a home, Paying off debt, Financial independence, Saving for education, Planning for major purchases
Q5	What is your investment horizon?	investment horizon in years
Q6	How old are you?	Age
Q7	When do you plan to retire?	in years, already retired
Q8	When do you expect to withdraw some of the money you have invested?	in years
Q9	Would you like to invest money on a one-off and/or regular basis?	One-off, Regular (monthly), Both one-off and regular
Q10	How much would you like to invest one-off?	in EUR
Q11	How much would you like to invest monthly?	in EUR
Q12	How much could you invest in addition per year?	in EUR
Q13	What risk (loss in value) of your investment are you willing to accept (risk affinity or loss acceptance)?	1 (very low risk, <5%), 2 (low risk, <10%), 3 (average risk, <20%), 4 (higher risk, <35%), 5 (very high risk, >35%)
Q14	How concerned are you about the risk of losing part of your assets (asset loss concern)?	1 (extremely concerned), 2 (very concerned), 3 (moderately concerned), 4 (slightly concerned), 5 (not concerned at all)
Q15	How willing have you been to take risks when making financial decisions in the past (past risk willingness)?	1 (not willing at all), 2 (slightly willing), 3 (moderately willing), 4 (very willing), 5 (extremely willing)
Q16	Do you want to invest sustainably?	yes, no
Q17	Would you like to have an advisor at your side?	yes, no
Q18	Would you like to prioritize certain sectors and/or themes?	Technology, Healthcare, Financials, Consumer

		Discretionary, Consumer Staples, Industrials, Energy, Utilities, Real Estate, Materials, Cryptocurrencies, No preference
Q19	Would you like to prioritize regions for your investment?	Developed Markets, Emerging markets, No preference
Q20	What knowledge and experience (in years) do you have of the following types of securities?	years for stocks, bonds, ETFs, commodities
Q21	How many trades have you made in the following asset classes in the last 3 years?	number of trades per asset class (stocks, bonds, ETFs, commodities)
Q22	Have you gained relevant capital market experience through training, studies, or work?	Yes, through formal education, through professional experience, through both education and professional experience, No
Q23	What is your highest level of education?	High school diploma, Associate degree, Bachelor's degree, Master's degree, Doctorate, Professional certification
Q24	Which asset management services have you used in the past?	Financial advisory services, Robo-advisors, Full-service brokerage, Discount brokerage, None
Q25	What is your monthly income?	in EUR
Q26	What is your total monthly expenditure?	in EUR
Q27	What is the amount of your liquid assets?	in EUR
Q29	What is the amount of your total assets (including non-liquid assets)?	in EUR
Q29	How high are your liabilities?	in EUR
Q30	How long will your reserves last to cover your current expenses and obligations?	in months or years
Q31	What is your profession?	Student, Employed, Self-employed, Retired, Unemployed

Q32	What is your main source of income?	Salary/Wages, Business income, Investment income, Pension, Other

Note. Author's elaboration: This table shows the 32-item questionnaire used to implement the ChatGPT-based robo-advisor for investor profiling and portfolio recommendation. Each question is paired with the corresponding predefined answer options. The questions are based on the typical profiling questions identified from 687 items across 45 robo-advisors operating in Germany in July 2024.

APPENDIX 6. Investment Universe for the GPT-Based Robo-Advisor

Asset class	Assets (instruments)
Stocks (Equities)	Stocks Global ETF, Stocks Global ESG ETF, Stocks Developed Markets (DM) ETF, Stocks DM ESG ETF, Stocks Emerging Markets (EM) ETF, Stocks EM ESG ETF, Stocks Sector ETFs
Bonds (Fixed-income)	Bonds Global ETF, Bonds Global ESG ETF, Bonds DM ETF, Bonds DM ESG ETF, Bonds EM ETF, Bonds EM ESG ETF
Real Estate	Real Estate ETF, Real Estate ESG ETF
Commodities	Precious Metals
Alternatives	Cryptocurrencies

Note. Author's elaboration: The table presents the investment universe defined for implementing the ChatGPT-based robo-advisor. It includes 17 instruments across equities, fixed income, real estate, commodities, and alternatives. The universe is derived from the instrument sets observed in 405 portfolios managed by 45 robo-advisors operating in Germany in July 2024.

APPENDIX 7. Logos of 45 Robo-Advisors in Germany



Note. Author's elaboration: Logos represent the 45 robo-advisors active in Germany in July 2024. Logos were collected from publicly available sources on provider websites.

APPENDIX 8. Portfolios by Equity Share of the GPT-Based Robo-Advisor

Equity exposure	Number of portfolios
95%	2
90%	4
85%	7
80%	10
75%	6
70%	7
65%	15
60%	8
55%	6
50%	7
45%	2
40%	15
35%	1
30%	2
25%	3
20%	2
15%	1
5%	1
0%	1

Note. Author's elaboration: This table shows the number of portfolios per equity exposure based on the composition analysis of 100 recommended portfolios from the GPT-based robo-advisor. For example, an equity exposure of 60% means that portfolios with an equity share of 60% were suggested for 8 profiles.

APPENDIX 9. Simple OLS Regression Results of the GPT-Based Robo-Advisor

Independent Variable Z_i	α	β	p	HC3	R ²	Corr. ρ	Breusch-Pagan p	White p	Shapiro-Wilk p	Jarque-Bera p	Max Cook's D
Past risk willingness	0.594	0.142	0.000*	0.000*	0.472	0.687	0.381	0.681	0.120	0.540	0.117
Asset loss concern	0.594	0.140	0.000*	0.000*	0.458	0.677	0.335	0.610	0.121	0.477	0.116
Risk affinity	0.594	0.138	0.000*	0.000*	0.448	0.669	0.229	0.305	0.225	0.771	0.115
Net monthly income	0.594	0.126	0.000*	0.000*	0.374	0.611	0.600	0.018	0.291	0.372	0.354
Age	0.594	-0.122	0.000*	0.000*	0.349	-0.591	0.419	0.067	0.280	0.747	0.091
Withdrawal horizon	0.594	0.116	0.000*	0.000*	0.318	0.564	0.085	0.097	0.339	0.780	0.107
Monthly income	0.594	0.115	0.000*	0.000*	0.314	0.560	0.053	0.001	0.068	0.279	0.229
Investment horizon	0.594	0.115	0.000*	0.000*	0.312	0.559	0.080	0.131	0.546	0.731	0.129
Retirement horizon	0.594	0.115	0.000*	0.000*	0.311	0.557	0.270	0.039	0.263	0.808	0.096
Monthly expenses	0.594	0.091	0.000*	0.000*	0.195	0.442	0.001	0.002	0.018	0.173	0.130
Liabilities	0.594	0.076	0.000*	0.000*	0.137	0.371	0.122	0.302	0.061	0.174	0.091
Total assets	0.594	0.076	0.000*	0.000*	0.137	0.370	0.036	0.088	0.054	0.186	0.088
Net assets	0.594	0.073	0.000*	0.000*	0.126	0.354	0.049	0.113	0.061	0.189	0.090
Liquid assets	0.594	0.068	0.001	0.000*	0.108	0.329	0.111	0.114	0.004	0.084	0.096
Coverage reserve	0.594	0.044	0.033	0.030	0.045	0.213	0.879	0.831	0.001	0.047	0.086

Note. Author's elaboration: Results are based on ordinary least squares (OLS) regressions between Z-scaled independent variables Z_i (see Table 12) and the portfolio equity exposure Y in 100 portfolios proposed by the ChatGPT-4o-based robo-advisor. α = intercept; β = regression coefficient; p = p-value; HC3 = robust heteroskedasticity-consistent standard error (type 3); R^2 = coefficient of determination; ρ = Spearman's rank correlation; Diagnostic tests: Breusch-Pagan test for heteroskedasticity, White's test for heteroskedasticity, Shapiro-Wilk and Jarque-Bera tests for normality, Cook's D for influence. $p < .001$ is marked with an asterisk (*).

APPENDIX 10. Proof of the Distance from the Capital Market Line

This appendix proves the Euclidian distance (58) of portfolios in the (σ, μ) -plane from the capital market line (CML). For a line $\mu = m\sigma$ with slope m , any perpendicular line to it must have a slope equal to the negative reciprocal, i.e., $-1/m$, because the product of the slopes of two perpendicular lines in the plane is -1 . A line of slope $-1/m$ passing through $(\sigma^{(i,j)}, \mu^{(i,j)})$ fulfils the points-slope form

$$\mu - \mu^{(i,j)} = -\frac{1}{m}(\sigma - \sigma^{(i,j)}), \Leftrightarrow \mu(\sigma) = -\frac{1}{m}\sigma + \left(\mu^{(i,j)} + \frac{\sigma^{(i,j)}}{m}\right) \quad (63)$$

The foot of the perpendicular is the unique point where this line meets the original line $\mu = m\sigma$. Set the right-hand sides equal:

$$m\sigma^* = -\frac{1}{m}\sigma^* + \left(\mu^{(i,j)} + \frac{\sigma^{(i,j)}}{m}\right), \Leftrightarrow \sigma^* = \frac{m\mu^{(i,j)} + \sigma^{(i,j)}}{m^2 + 1} \quad (64)$$

Then $\mu^* = m\sigma^*$ gives

$$\mu^* = \frac{m^2\mu^{(i,j)} + m\sigma^{(i,j)}}{m^2 + 1}. \quad (65)$$

The squared distance between $(\sigma^{(i,j)}, \mu^{(i,j)})$ and (σ^*, μ^*) (after some algebra) is

$$(\sigma^{(i,j)} - \sigma^*)^2 + (\mu^{(i,j)} - \mu^*)^2 = \frac{(m\sigma^{(i,j)} - \mu^{(i,j)})^2}{m^2 + 1}. \quad (66)$$

Taking the positive square root gives the perpendicular distance

$$d^{(i,j)} = \frac{|m\sigma^{(i,j)} - \mu^{(i,j)}|}{\sqrt{m^2 + 1}}, \quad (67)$$

which proves (58).

APPENDIX 11. Profiling Sample for the GPT-Based Robo-Advisor

Investment horizon [yrs]	Age [yrs]	Retirement horizon [yrs]	Withdrawal horizon [yrs]	Risk affinity [1–5]	Asset loss concern [1–5]	Past risk willingness [1–5]	Mthly income [EUR]	Mthly expense [EUR]	Net mthly income [EUR]	Liquid assets [EUR]	Total assets [EUR]	Liabilities [EUR]	Net assets [EUR]	Reserve [mth]	Equity
10	35	30	5	3	3	3	4,000	2,500	1,500	30,000	150,000	20,000	130,000	12	75%
15	50	15	10	2	2	2	6,000	4,000	2,000	50,000	400,000	100,000	300,000	15	50%
20	40	25	15	4	4	4	5,000	3,000	2,000	40,000	250,000	50,000	200,000	10	80%
5	45	20	3	1	1	1	3,500	2,000	1,500	20,000	100,000	30,000	70,000	10	0%
25	30	35	20	4	4	4	8,000	4,000	4,000	80,000	300,000	50,000	250,000	20	95%
15	50	15	10	2	2	2	6,500	3,500	3,000	60,000	400,000	100,000	300,000	18	40%
10	55	10	5	2	2	2	5,500	3,500	2,000	50,000	350,000	80,000	270,000	15	30%
12	42	20	7	3	3	3	7,000	4,000	3,000	60,000	280,000	70,000	210,000	15	80%
20	35	30	15	4	4	4	7,500	3,500	4,000	80,000	350,000	50,000	300,000	20	70%
8	60	0	4	1	1	1	4,500	3,000	1,500	70,000	250,000	30,000	220,000	24	40%
15	45	20	10	3	3	3	6,500	4,000	2,500	50,000	350,000	75,000	275,000	12	85%
10	50	15	5	2	2	2	5,500	3,500	2,000	70,000	300,000	80,000	220,000	18	30%
25	30	35	20	4	4	4	7,500	4,000	3,500	80,000	350,000	50,000	300,000	20	95%
10	40	25	7	3	3	3	6,000	3,500	2,500	60,000	250,000	70,000	180,000	15	85%
15	50	15	10	2	2	2	6,500	4,000	2,500	70,000	300,000	80,000	220,000	18	60%
20	35	30	15	4	4	4	6,500	3,500	3,000	60,000	300,000	50,000	250,000	18	75%
5	45	20	3	1	1	1	4,500	3,000	1,500	60,000	200,000	50,000	150,000	18	40%

12	55	10	7	2	2	2	5,000	3,500	1,500	50,000	250,000	60,000	190,000	14	35%
15	40	25	10	3	3	3	6,000	3,500	2,500	50,000	250,000	70,000	180,000	14	85%
20	45	20	15	3	3	3	6,500	4,000	2,500	70,000	350,000	80,000	270,000	18	75%
25	35	30	20	4	4	4	7,000	3,500	3,500	60,000	300,000	50,000	250,000	18	80%
7	48	17	4	1	1	1	5,000	3,500	1,500	60,000	200,000	50,000	150,000	18	40%
20	45	20	15	3	3	3	6,500	4,000	2,500	70,000	350,000	80,000	270,000	18	75%
10	50	15	7	2	2	2	5,500	3,500	2,000	50,000	250,000	70,000	180,000	14	55%
15	42	20	10	3	3	3	7,000	4,000	3,000	60,000	300,000	75,000	225,000	15	85%
10	60	0	5	1	1	1	4,000	3,000	1,000	70,000	200,000	30,000	170,000	24	40%
20	35	30	15	4	4	4	7,000	3,500	3,500	60,000	300,000	50,000	250,000	18	80%
15	50	15	10	2	2	2	5,500	3,500	2,000	60,000	280,000	60,000	220,000	17	60%
12	55	10	7	2	2	2	5,500	3,500	2,000	50,000	250,000	60,000	190,000	14	60%
15	42	20	10	3	3	3	7,000	4,000	3,000	60,000	300,000	75,000	225,000	15	85%
8	37	28	4	1	1	1	4,800	3,200	1,600	45,000	190,000	40,000	150,000	12	50%
18	30	35	12	4	4	4	5,500	3,200	2,300	55,000	250,000	70,000	180,000	17	65%
5	48	17	3	1	1	1	4,800	3,200	1,600	55,000	200,000	40,000	160,000	16	40%
12	45	15	10	2	2	2	5,000	3,500	1,500	60,000	220,000	70,000	150,000	14	50%
20	40	25	15	3	3	3	8,500	5,000	3,500	80,000	450,000	100,000	350,000	18	80%
10	55	10	7	2	2	2	6,000	4,500	1,500	100,000	350,000	70,000	280,000	22	40%
15	38	25	10	3	3	3	6,000	3,500	2,500	50,000	220,000	60,000	160,000	14	70%
12	52	10	8	2	2	2	6,500	4,200	2,300	80,000	400,000	90,000	310,000	18	55%
20	40	25	15	4	4	4	8,000	4,500	3,500	90,000	400,000	100,000	300,000	20	80%

15	50	15	10	2	2	2	5,500	3,500	2,000	60,000	280,000	70,000	210,000	17	60%
18	35	30	12	3	3	3	5,500	3,200	2,300	55,000	250,000	60,000	190,000	14	70%
8	60	0	5	1	1	1	4,500	3,500	1,000	70,000	250,000	30,000	220,000	20	40%
20	45	20	15	3	3	3	6,500	4,000	2,500	75,000	350,000	50,000	300,000	16	90%
10	55	10	7	2	2	2	5,000	3,500	1,500	70,000	250,000	40,000	210,000	18	40%
20	42	23	12	3	3	3	7,500	4,200	3,300	65,000	350,000	70,000	280,000	15	90%
5	38	27	3	1	1	1	4,500	3,000	1,500	55,000	200,000	30,000	170,000	18	40%
12	40	20	8	3	3	3	6,500	4,000	2,500	65,000	300,000	60,000	240,000	15	90%
10	50	10	7	2	2	2	5,000	3,500	1,500	60,000	240,000	50,000	190,000	17	40%
15	35	30	12	3	3	3	6,000	4,000	2,000	65,000	300,000	60,000	240,000	15	90%
10	55	10	7	2	2	2	5,500	3,500	2,000	70,000	250,000	40,000	210,000	18	40%
20	30	35	15	5	5	5	8,500	4,500	4,000	80,000	400,000	70,000	330,000	16	80%
25	32	33	20	5	5	5	10,000	5,500	4,500	100,000	500,000	80,000	420,000	18	75%
15	28	40	10	5	4	4	12,000	6,000	6,000	120,000	600,000	100,000	500,000	20	65%
20	33	30	15	5	5	5	9,000	5,000	4,000	120,000	700,000	90,000	610,000	18	70%
25	30	35	20	5	5	5	7,500	4,000	3,500	85,000	450,000	80,000	370,000	16	65%
20	35	30	15	5	5	5	9,500	5,000	4,500	100,000	600,000	75,000	525,000	18	70%
25	29	36	18	5	5	5	10,000	5,500	4,500	120,000	700,000	90,000	610,000	18	75%
20	32	30	15	5	4	4	9,500	5,000	4,500	90,000	550,000	80,000	470,000	18	80%
25	34	30	18	5	4	4	8,500	4,500	4,000	90,000	500,000	70,000	430,000	16	85%
20	36	29	15	5	4	4	7,500	4,000	3,500	85,000	500,000	60,000	440,000	18	85%
15	48	17	12	3	3	3	6,800	4,200	2,600	70,000	380,000	60,000	320,000	14	65%
8	52	13	5	2	2	2	4,800	3,200	1,600	50,000	200,000	30,000	170,000	15	40%

22	33	34	16	4	3	4	8,000	4,800	3,200	85,000	550,000	90,000	460,000	16	80%
12	45	20	10	2	2	2	5,500	3,500	2,000	65,000	350,000	40,000	310,000	14	40%
10	40	25	8	3	3	3	6,000	4,200	1,800	80,000	350,000	55,000	295,000	14	55%
30	31	36	25	4	4	4	10,500	6,000	4,500	110,000	600,000	85,000	515,000	18	80%
18	42	23	10	3	3	3	7,000	4,000	3,000	75,000	420,000	60,000	360,000	16	55%
25	35	30	20	4	4	4	9,000	5,000	4,000	95,000	550,000	70,000	480,000	17	60%
10	50	17	7	2	2	2	4,200	3,200	1,000	50,000	250,000	30,000	220,000	12	25%
20	38	30	15	3	3	3	7,500	4,500	3,000	80,000	420,000	55,000	365,000	14	60%
30	28	37	25	5	5	5	11,000	5,800	5,200	130,000	720,000	85,000	635,000	20	65%
15	27	38	12	4	4	4	2,800	1,800	1,000	7,000	22,000	3,000	19,000	4	65%
20	30	35	18	2	2	2	2,500	1,800	700	6,000	18,000	4,000	14,000	4	40%
18	41	24	12	3	3	3	4,800	3,200	1,600	30,000	160,000	35,000	125,000	9	60%
25	33	35	20	4	4	4	6,500	4,200	2,300	40,000	210,000	45,000	165,000	10	65%
12	39	26	8	3	3	3	4,500	3,200	1,300	20,000	180,000	50,000	130,000	6	45%
30	29	36	20	5	4	4	7,000	4,000	3,000	35,000	240,000	30,000	210,000	7	65%
20	43	22	15	3	3	3	5,000	3,600	1,400	25,000	180,000	40,000	140,000	7	65%
25	26	39	20	3	4	3	2,800	1,900	900	9,000	32,000	5,000	27,000	4.5	65%
17	45	22	12	3	3	3	6,200	4,000	2,200	40,000	300,000	65,000	235,000	8	55%
30	27	38	25	4	4	4	3,500	2,200	1,300	14,000	65,000	6,000	59,000	6	65%
15	40	25	10	3	3	3	5,400	3,600	1,800	25,000	180,000	30,000	150,000	6	45%
12	58	7	5	3	2	2	4,500	3,500	1,000	60,000	320,000	45,000	275,000	9	25%
7	61	0	3	1	1	1	2,800	2,300	500	20,000	110,000	10,000	100,000	7	5%

25	32	35	20	3	3	3	4,000	2,600	1,400	15,000	90,000	10,000	80,000	6	55%
10	44	21	6	2	2	2	4,200	3,000	1,200	30,000	160,000	25,000	135,000	10	20%
30	28	37	25	5	5	5	4,500	2,800	1,700	25,000	100,000	12,000	88,000	8	65%
18	42	23	10	3	3	3	4,300	3,100	1,200	20,000	180,000	35,000	145,000	6	50%
30	26	40	20	4	4	4	3,000	2,000	1,000	8,000	30,000	4,000	26,000	4	65%
20	41	24	12	3	3	3	4,800	3,200	1,600	30,000	180,000	35,000	145,000	8	50%
35	29	38	25	5	5	5	5,200	3,500	1,700	22,000	120,000	15,000	105,000	6	60%
22	38	27	18	3	3	3	4,500	3,100	1,400	22,000	150,000	20,000	130,000	7	50%
12	49	16	7	2	2	2	3,800	2,800	1,000	30,000	160,000	18,000	142,000	10	20%
28	31	34	20	4	4	4	4,000	2,600	1,400	18,000	85,000	10,000	75,000	6	70%
20	43	24	12	3	3	3	4,200	3,000	1,200	25,000	140,000	30,000	110,000	8	50%
30	27	38	20	5	5	4	4,200	2,700	1,500	16,000	70,000	12,000	58,000	6	70%
15	45	20	10	2	2	2	4,000	3,000	1,000	28,000	160,000	20,000	140,000	9	25%
25	30	37	20	4	4	4	4,100	2,900	1,200	20,000	100,000	12,000	88,000	6	65%
8	59	6	4	1	1	1	3,000	2,500	500	35,000	120,000	20,000	100,000	8	15%
22	40	27	15	3	3	3	4,500	3,200	1,300	24,000	130,000	20,000	110,000	7	65%

Note. Author's elaboration: The table presents raw responses of 100 simulated investors to 15 profiling variables and the corresponding equity shares of portfolios recommended by the ChatGPT-based robo-advisor. Equity share (%) denotes the proportion of equities in the proposed portfolio. Risk variables (risk affinity, asset loss concern, past risk willingness) were measured on 5-point Likert scales. Financial variables are expressed in EUR.

