



A Competency-Based Autonomous Coaching Framework Utilising Large Language Model Technology

István Vasas

A thesis submitted to the

University of Gloucestershire

In accordance with the requirements of the degree of

Doctor of Philosophy

January 2025

Word Count: 56.823

ABSTRACT **6**

DECLARATION OF ORIGINAL CONTENT **7**

ACKNOWLEDGEMENTS **8**

LIST OF ABBREVIATIONS **9**

LIST OF FIGURES **11**

LIST OF TABLES **12**

1. INTRODUCTION **13**

1.1 Project Background 13
1.2 Overall Aim and Research Objectives 16
1.3 Thesis Structure 17

2. LITERATURE REVIEW **19**

2.1 Introduction 19
2.2 Coaching 19
2.2.1 Definition 19
2.2.2 Coach Core Competences 22
2.2.3 Coaching Body of Knowledge Development 26
2.2.4 Boundary Issues 27
2.2.5 Challenges in Coaching Research 27
2.2.6 Future Trends 29
2.2.7 Ethical Issues 30
2.2.8 Benefits 32
2.2.9 Evaluation 33
2.3 Autonomous Coaching 38
2.3.1 Definition 38
2.3.2 Limitations of Existing Automated Coaching Solutions 39
2.3.3 Existing Autonomous Coaching Frameworks 42
2.3.4 Ethical Issues 44
2.3.5 Benefits 46
2.3.6 Evaluation 47
2.4 Technologies for Autonomous Coaching 48
2.4.1 Architecture Review 49
2.4.2 Emergence of Large Language Models 51
2.4.3 Limitations of the Large Language Models 52
2.4.4 Ethical Issues 54
2.4.5 Benefits 56
2.4.6 Evaluation and Testing 57
2.5 Chapter Summary 59
2.5.1 Research Gap 59
2.5.2 Originality of the Thesis 63

3. RESEARCH DESIGN **65**

3.1 Introduction 65
3.2 Research Philosophy 65
3.3 Research Hypotheses 66
3.4 Research Methodology 66
3.4.1 Framework Building Approach 66
3.4.2 Framework Components 68
3.4.3 Quality Considerations 70
3.4.4 Research Design 71
3.4.5 Framework Validation Approach 74
3.4.6 Methods for Data Collection and Analysis 75
3.4.7 Assumptions and Limitations 80

3.5 Ethical Issues 82
3.6 Chapter Summary 82

4. NEW FRAMEWORK BUILDING: KEY COMPONENTS 84

4.1 Introduction 84
4.2 Stakeholder Analysis 84
4.2.1 Approach 84
4.2.2 Stakeholder Identification 85
4.2.3 Stakeholder Mapping 88
4.2.4 User (Client) 89
4.2.5 Coach 90
4.2.6 Sponsor 91
4.2.7 Commissioner 92
4.2.8 Vendor (DevOps) 92
4.2.9 Service Providers 92
4.2.10 Accrediting Institute(s) 93
4.2.11 Researcher 93
4.2.12 Outcome 94
4.3 Macro Context Analysis 95
4.3.1 Approach 95
4.3.2 Social Factors 95
4.3.3 Technological Factors 98
4.3.4 Economic Factors 99
4.3.5 Environmental Factors 101
4.3.6 Political Factors 101
4.3.7 Legal Factors 102
4.3.8 Ethical Factors 104
4.3.9 Outcome 105
4.4 Micro Context Analysis 107
4.4.1 Approach 107
4.4.2 Coaching – Multiple-Perspective Model (MPM) 108
4.4.3 Autonomous Coaching 111
4.4.4 Outcome 112
4.5 Value Proposition Design 113
4.5.1 Approach 113
4.5.2 Customer Segment Profile 115
4.5.3 Value (Proposition) Map 118
4.5.4 Outcome 120
4.6 Objective Analysis 121
4.6.1 Approach 121
4.6.2 Objectives 121
4.6.3 Objective-Problem Relationship 123
4.6.4 Measurement Approach 125
4.6.5 Outcome 126
4.7 Constraint Analysis 127
4.7.1 Approach 128
4.7.2 Constraints 128
4.7.3 Outcome 130
4.8 Chapter Summary 130

5. NEW FRAMEWORK BUILDING: SOLUTION REQUIREMENT LIST 133

5.1 Introduction 133
5.2 Coaching Common Theme Analysis 133
5.2.1 Approach 133
5.2.2 Coaching Common Themes 136
5.2.3 Theme Groups 138
5.2.4 Outcome 139

5.3 Data Modelling 139
5.3.1 Approach 139
5.3.2 Data Element: User 140
5.3.3 Data Element: Outcome 141
5.3.4 Data Element: Conversation 142
5.3.5 Data Element: Reporting 143
5.3.6 Data Element: Coach 143
5.3.7 Data Element: Dialog 144
5.3.8 Outcome 148
5.4 Process Modelling 149
5.4.1 Approach 149
5.4.2 Process: Coaching Administration 149
5.4.3 Process: Coaching Engagement 151
5.4.4 Process: Coaching Session 152
5.4.5 Outcome 155
5.5 Deriving Solution Requirement List 155
5.5.1 Approach 155
5.5.2 Requirement: Functional Suitability 156
5.5.3 Requirement: Performance Efficiency 158
5.5.4 Requirement: Compatibility 159
5.5.5 Requirement: Usability 160
5.5.6 Requirement: Reliability 160
5.5.7 Requirement: Security 161
5.5.8 Requirement: Maintainability and Portability 161
5.5.9 Outcome 162
5.6 Chapter Summary 164

6. A NEW FRAMEWORK FOR AUTONOMOUS COACHING 166

6.1 Introduction 166
6.2 COBAC Initial Framework Description 166
6.2.1 Initial Constraint List 166
6.2.2 Initial Problem List 167
6.2.3 Initial Objective List 167
6.2.4 Initial Solution Requirement List 167
6.3 COBAC Reference Architecture 168
6.4 Framework Application for Implementation 170
6.5 Chapter Summary 171

7. DESIGN AND DEVELOPMENT OF THE PROTOTYPE OF THE PROPOSED FRAMEWORK 172

7.1 Introduction 172
7.2 DSRM Step 1: Identify Problem and Motivation 172
7.3 DSRM Step 2: Define Objectives 173
7.4 DSRM Step 3: Prototype Design and Development 175
7.4.1 Prototype Solution Requirements 175
7.4.2 Prototype Architecture 176
7.4.3 Prototype Technology Selection 177
7.4.4 Prototype Design 179
7.4.5 Prototype Development 180
7.4.6 Prototype Development Challenges 183
7.4.7 Outcome 184
7.5 DSRM Step 4: Prototype Demonstration - Data Collection 186
7.5.1 Initial Survey Design 186
7.5.2 Pilot Survey 189
7.5.3 Main Survey 189
7.6 Chapter Summary 191

8. DESIGN AND DEVELOPMENT OF AN OPTIMISED NEW FRAMEWORK 193

8.1 Introduction 193
8.2 Descriptive Statistics, Data Screening, Data Cleaning 193
8.3 Hypothesis 1 (H1) - Feasibility 194
8.4 Hypothesis 2 (H2) - Efficacy 195
8.5 Hypothesis 3 (H3) - Dialog Quality 197
8.6 Hypothesis 4 (H4) - Language Influence on Adoption Factors 198
8.6.1 Measurement Model 199
8.6.2 Structural Model 202
8.7 Proposed Optimised Framework 203
8.7.1 Optimised Constraint List 204
8.7.2 Optimised Problem List 204
8.7.3 Optimised Objective List 204
8.7.4 Optimised Solution Requirement List 205
8.8 Comparative Analysis 205
8.9 Chapter Summary 207

9. CONCLUSIONS AND FURTHER WORK 210

9.1 Main Achievements 210
9.2 Conclusion Regarding the Hypotheses 211
9.3 Contribution to the New Knowledge Generation 213
9.4 Limitations and Further Work 214
9.4.1 Context and Value Proposition Change 214
9.4.2 Scope Limitation 215
9.4.3 Methodology Limitation 216
9.4.4 Technology 216

REFERENCES 218

APPENDICES 237

Appendix A – EMCC codes 237
Appendix B – ICF codes 241
Appendix C – Common codes 244
Appendix D – Coaching Common Themes 245
Appendix E – Initial Solution Requirement List 249

Abstract

This study examines major challenges in the coaching industry, including the lack of a universally recognised definition, persistent ethical dilemmas, and reliance on non-empirical approaches that hinder data-driven AI coaching solutions. To address these issues, it proposes and develops the Competency-Based Autonomous Coaching (COBAC) framework, integrating rigorous scientific methods with the Large Language Model (LLM) technology. By automating the coaching process, COBAC aspires to provide scalable, ethically sound, and evidence-based coaching interventions that can be deployed across various contexts and user groups.

Guided by a positivist epistemology, this research undertakes a two-phase approach to develop and validate the COBAC framework. In the initial phase, document-centric methodologies and requirement engineering techniques establish a robust theoretical and practical foundation. Subsequently, the framework is evaluated using the Design Science Research Methodology, which includes the development of a software prototype tested by 123 participants in a cross-sectional, exploratory, quantitative study. Hypotheses are assessed through Partial Least Squares Structural Equation Modelling (PLS-SEM), providing insights into both the feasibility and perceived effectiveness of the framework and highlighting areas for further refinement.

The findings confirm that the COBAC framework's viability in operationalising established coaching competencies within an automated, LLM-based setting. However, linguistic and cultural dimensions influence user perceptions and dialogue quality, suggesting that culturally sensitive adaptations are necessary to enhance AI coaching solutions. Overall, this thesis offers a context-independent foundation for autonomous coaching that unites theoretical principles with state-of-the-art automation and paves the way for future innovations. By promoting scalable, evidence-based AI coaching tools, COBAC supports the broader coaching industry's shift towards ethically grounded, data-driven services accessible to diverse populations.

Declaration of original content

I declare that the work in this doctoral thesis was carried out in accordance with the regulations of the University of Gloucestershire and is original, except where indicated by specific reference in the text. No part of this doctoral thesis has been submitted as part of any other academic award.

Any views expressed in this doctoral thesis are those of the author and in no way represent those of the University.

Signed:

Date: 17th January 2025

DOI: 10.46289/JIND8345

© István Vasas (2025).
This thesis is copyright material and no quotation from it may be published without proper acknowledgement.

Acknowledgements

I extend my deepest gratitude to Professor Shujun Zhang for his guidance and mentorship throughout my doctoral journey. His invaluable advice, unwavering encouragement, and outstanding support have profoundly shaped both my project and my growth as a scholar.

I also wish to express my heartfelt thanks to my wife, Zsuzsanna, and my son, Martin, for their patience, understanding, and love. Their support has been a cornerstone of my resilience and success throughout this challenging yet rewarding endeavour.

I would like to acknowledge the assistance of ChatGPT in improving the quality of English text in this thesis.

List of Abbreviations

ACM	Amazon Certificate Manager
AI	Artificial Intelligence
API	Application Programming Interface
AVE	Average Variance Extracted
AWS	Amazon Web Services
BCRS	Brief Coaching Readiness Scale
BI	Behavioral Intention
BPMN	Business Process Model and Notation
CBT	Cognitive Behavioural Therapy
CDN	Content Delivery Network
CE	Coaching Effectiveness (as measure)
COBAC	Competence-Based Autonomous Coaching (as framework)
CR	Composite Reliability (as measure)
CSAT	Customer Satisfaction (as measure)
CSES	Core Self-Evaluation Scale
CSS	Cascading Style Sheets
DNS	Domain Name System
DRQ	Developmental Readiness Questionnaire
DSRM	Design Science Research Methodology
EE	Effort Expectancy (as measure)
EMCC	European Mentoring & Coaching Council
FC	Facilitating Conditions (as measure)
GAS	Goal Attainment Scaling
GCoE	Global Code of Ethics
HCI	Human Computer Interaction
HRD	Human Resource Development
HTML	HyperText Markup Language
ICF	International Coaching Federation
IEC	International Electrotechnical Commission
IES	Institute of Employment Studies
IREB	International Requirements Engineering Board
ISO	International Organization for Standardization
JS	JavaScript
JSON	JavaScript Object Notation
LG	Language (as participant group)
LLM	Large Language Model
ML	Machine Learning
MOOC	Massive Open Online Course
MPM	Multiple-Perspective Model
MTurk	Amazon Mechanical Turk
NLP	Natural Language Processing
OCEAN	Big Five personality trait model named after the five factors: openness to experience, conscientiousness, extraversion, agreeableness, neuroticism

OECD	Organisation for Economic Co-operation and Development
OMG	Object Management Group
PE	Performance Expectancy (as measure)
PLS-SEM	Partial Least Squares Structural Equation Modelling
PR	Perceived Risk (as measure)
RAG	Retrieval-Augmented Generation
RDS	Relational Database Service
SH	Shell Script
SI	Social Influence (as measure)
SQL	Structured Query Language
SSL	Secure Sockets Layer
UML	Unified Modeling Language
USR	Unsupervised and Reference-free metric
UTAUT	Unified Theory of Acceptance and Use of Technology
WAR	Web Archive
WEF	World Economic Forum

List of Figures

Figure 3.1: COBAC Components and Relationships 69
Figure 3.2: Initial Framework Building Stage Overview 71
Figure 3.3: Design Science Research Methodology (DSRM) 75
Figure 4.1: Stakeholders in the Context of Autonomous Coaching System 88
Figure 4.2: Macro Context Analysis - Overview 105
Figure 4.3: Customer Segment Profile - Coaching 116
Figure 4.4: Customer Segment Profile - Changes Introduced by Automation and LLM 117
Figure 4.5: Customer Segment Profile - LLM-based Autonomous Coaching 118
Figure 4.6: Value (Proposition) Map - LLM-based Autonomous Coaching 119
Figure 5.1: Top 8 ICF and EMCC Code Counts 137
Figure 5.2: COBAC Conceptual Data Model 148
Figure 5.3: Coaching Administration Process - Engagement Setup 150
Figure 5.4: Coaching Engagement Process 152
Figure 5.5: Coaching Session Process 153
Figure 5.6: ISO/IEC 25010 Quality Characteristics 156
Figure 6.1: COBAC Reference Architecture 168
Figure 7.1: Prototype Architecture - Changes 177
Figure 7.2: Prototype Backend Package Overview 181
Figure 7.3: Prototype User Interface 185
Figure 8.1: Coaching Effectiveness Summary Distribution Among Language/Country 196
Figure 8.2: Conceptual Model and Hypothesis 200

List of Tables

Table 2.1: Coaching Definition Overview 20
Table 2.2: Coaching Competence Framework Overview 23
Table 2.3: Coaching Measurement Taxonomy 34
Table 2.4: Existing Autonomous Coaching System Examples 41
Table 2.5: Ethical Issues Associated with Coaching Automation 45
Table 3.1: COBAC Component Overview 69
Table 3.2: COBAC Quality Criteria 71
Table 3.3: Applied Measures 78
Table 4.1: Initial Stakeholder List 94
Table 4.2: Initial Problem List Derived from the Macro Context Analysis 106
Table 4.3: Initial Problem List Derived from the Micro Context Analysis 112
Table 4.4: Objective-Problem Relationship 123
Table 4.5: Problems Not Covered by Objectives 125
Table 4.6: Measurement Approach Overview 126
Table 4.7: Initial Objective List 127
Table 4.8: Initial Constraint List 130
Table 5.1: Document Key Naming Patterns 135
Table 5.2: Coaching Common Theme Summary - Agreement Management 137
Table 5.3: Intentions in Coaching Conversation 146
Table 5.4: Key Activities of Autonomous Coaching Systems 155
Table 5.5: Solution Requirement Summary - Functional Suitability 157
Table 5.6: Initial Solution Requirement List 162
Table 6.1: COBAC Reference Architecture - Component Description 169
Table 7.1: Prototype Objective List - Changes 174
Table 7.2: Prototype Solution Requirement List - Changes 176
Table 7.3: Prototype Infrastructure Components in AWS 178
Table 7.4: Development Tools and Services 179
Table 7.5: Prototype Complexity 185
Table 7.6: Survey Structure 187
Table 7.7: Survey Questions 188
Table 7.8: Data Collection Pilot Phase - Challenges and Adjustments 189
Table 7.9: Data Collection Outcome Overview 190
Table 8.1: Participant Descriptive Statistics 193
Table 8.2: Coaching Effectiveness Average 195
Table 8.3: Dialog Quality Average 197
Table 8.4: Item Loadings, Reliability and Convergent Validity 201
Table 8.5: Discriminant Validity 202
Table 8.6: Path Coefficients in the Full PLS-SEM Model 202
Table 8.7: Path Coefficients by Language (LG) 203
Table 8.8: Optimised Problem List - Changes 204
Table 8.9: Optimised Objective List - Changes 205
Table 8.10: Optimised Solution Requirement List - Changes 205
Table 8.11: Comparative Analysis of AI Coaching Frameworks 207

1. Introduction

This chapter provides a short overview of the research background, aim, objectives, contribution to new knowledge, and describes the structure of the thesis.

1.1 Project Background

Coaching has developed from a limited social capital supporting practice into an exponentially growing and evolving field over the past four decades. Despite its triumphant growth, the evolution of coaching has encountered challenges, notably the ongoing discourse surrounding its precise definition. The diversity of definitions offered by esteemed scholars and leading organizations within the field (Association for Coaching, 2020; Chittum, 2013; Cox, Bachkirova, & Clutterbuck, 2010; Drake, 2018; EMCC, 2018; Gallwey, 2008; Grant, 2006a; ICF, 2019b; Peterson, 1996; Stober & Grant, 2006) illustrates the complexity and dynamism of coaching, yet it simultaneously highlights a major impediment: the lack of a unified and broadly accepted definition. This definitional ambiguity complicates the alignment of expectations between coaches and their clients and poses substantial challenges for research comparability and the application of coaching practices. It was argued that none of the prevailing definitions at the time of their investigation successfully fulfilled the major criteria indispensable to define a practice: distinctiveness and universality (Bachkirova & Kauffman, 2009). This situation underscores the necessity for a standardized and refined methodology in coaching.

The lack of consensus on a coaching definition not only affects research outcomes but also significantly impacts clients' expectations and perceptions of coaching, potentially undermining its credibility and deterring client participation. Furthermore, the reliance on expert opinions rather than empirical evidence in developing coaching competency frameworks signals a major area for research into autonomous coaching. Such research aims to ground coaching practices in scientifically validated methodologies, enhancing their legitimacy and effectiveness.

Additionally, the ethical considerations in coaching reflect the industry's focus on establishing clear ethical guidelines. However, the ongoing debates in ethical compliance and ethical code development suggest a need for more robust ethical frameworks. Autonomous coaching platforms incorporate the coaching definition and ethical standards into their algorithms and functionalities, hence clarity in these areas are vital prerequisites to develop such systems.

The challenges in coaching research, including the high costs associated with rigorous studies and the limitations imposed by small sample sizes and homogeneous populations, hinder the generalizability and applicability of research findings. This situation underscores the potential of autonomous coaching as a research avenue, offering scalable, cost-effective methods for gathering empirical data and refining coaching practices based on evidence-based outcomes.

The motivation to research autonomous coaching emerges from the need to address the definitional ambiguity, standardize competencies, enhance ethical practices coded in algorithms, and overcome the research limitations present in the current coaching industry. By leveraging technology and evidence-based methodologies, autonomous coaching presents an opportunity to advance the field, ensuring its practices are effective, reliable, and grounded in robust scientific principles.

The quest for autonomous coaching systems stems from the digital and automation revolution, which has profoundly affected various sectors, including the coaching industry. This surge in digitalization seeks to enhance the effectiveness and reach of coaching practices using advanced technologies such as Artificial Intelligence (AI), Machine Learning (ML) or Large Language Models (LLM). Historically, the coaching industry was characterised by a reliance on direct human interaction, which, while effective, presents limitations in terms of scalability, accessibility, and consistency. The advent of AI offers a promising avenue to overcome these challenges, proposing systems capable of delivering personalized coaching without the need for continuous human oversight.

The concept of autonomous coaching is grounded in the broader context of automation, where tasks traditionally performed by humans are increasingly being transferred to machines or algorithms. This shift is motivated by the potential for increased efficiency, the

ability to process and analyse vast amounts of data, and the promise of providing consistent, unbiased support. Furthermore, autonomous coaching systems hold the potential to democratize access to coaching services, making professional and personal development tools accessible to a wider audience regardless of geographical or financial barriers.

However, the development of autonomous coaching systems is not without its challenges. Issues of ethical concern, such as data privacy, algorithmic bias, and the integrity of coaching outcomes, necessitate careful consideration and rigorous scrutiny. The design and implementation of these systems require a multidisciplinary approach, incorporating insights from psychology, ethics, computer science, and coaching practice to ensure they are effective, fair, and respectful of users' rights and dignity.

Moreover, the integration of autonomous coaching systems into the existing ecosystem presents both opportunities and challenges. On one hand, they offer the potential to enhance human coaching by providing additional resources, supporting data analysis, and enabling a more personalized coaching experience. On the other hand, there is the risk of undermining the value of human connection and intuition that forms the cornerstone of traditional coaching relationships.

The pursuit of autonomous coaching is also fuelled by the desire to address the limitations of current automated coaching solutions, which often focus on non-core aspects of coaching or augment rather than replace human coaches. The goal is to create systems that can manage the entire coaching process, from goal setting to achieving outcomes, in a manner that rivals or surpasses human capabilities.

The development of autonomous coaching systems reflects a broader trend towards leveraging technology to enhance human capabilities and improve life quality. As such, it represents an exciting frontier in the intersection of technology and personal development, promising to reshape the landscape of coaching in the digital age. The research and development in this field are driven by a vision of creating accessible, effective, and ethical coaching solutions that harness the power of AI to empower individuals and organizations worldwide.

Lastly, the motivation for researching autonomous coaching technologies originates partially from the increasing recognition of the profound impact that technology choices have on the non-functional aspects of coaching systems. Historically, the focus in coaching automation was on functional capabilities, often overlooking how the underlying technology stack can influence operational costs and user accessibility. As AI becomes more integrated into coaching, there is a major need to understand how different technologies, including Large Language Models, can improve the coaching experience while adhering to ethical guidelines and constraints. The push towards digital platforms, voice recognition, and intelligent response generation signifies a shift toward more nuanced and sophisticated systems that can handle complex human interactions. Therefore, academic research in this area is crucial to address the design challenges, ethical considerations, and limitations of LLMs in the development of effective and responsible autonomous coaching solutions.

1.2 Overall Aim and Research Objectives

The overall aim of this thesis is to propose, design and develop a framework for competency-based autonomous coaching utilizing LLM technology. The newly designed and developed theoretical framework for the development of LLM-based autonomous coaching systems is referred to as the Competency-Based Autonomous Coaching (COBAC) framework throughout this thesis. The COBAC framework is aimed to fulfil the following high-level requirements:

1. Coaching Core Competence-Based: This approach is anchored in the latest coaching core competency frameworks published by major coaching accrediting organizations.
2. Scientific Research-Based: It integrates the most up-to-date scientific research from the coaching and automation fields.
3. Ethics-Oriented Design: Key ethical considerations from both the coaching and automation fields are carefully addressed in the design.
4. Practical Application: This framework is well-suited to serve as the foundation for the development of autonomous coaching systems.

The target group of the COBAC framework consists of coaching researchers and requirements engineers to support them deriving their own application specific frameworks based on their individual needs and software engineers, who may partake in development of autonomous coaching systems.

The research aimed to address the following research objectives:

1. Based on the literature review, to apply business process modelling techniques to analyse, model and optimise the processes of coaching and hence to produce the list of addressed problems, objectives, constraints, and solution requirements.
2. Based on (1) to propose a theoretical framework for building autonomous coaching services.
3. Based on (1) and (2), to design and develop an LLM-based software service prototype to demonstrate the feasibility of the framework.
4. To evaluate the proposed framework and developed software service prototype by measuring functional and non-functional characteristics based on feedback gathered from users.

1.3 Thesis Structure

The thesis presents a structured examination of the field, beginning with an abstract and acknowledgments, followed by the standard lists of abbreviations, figures, and tables.

- Chapter 1:
  - Introduces the research project.
  - Outlines the background, motivations, and objectives.
- Chapter 2:
  - Performs an extensive literature review.
  - Addresses definitions, competencies, and evolution of coaching and autonomous coaching.
  - Covers technological underpinnings and ethical implications.
- Chapter 3:
  - Details the research design.

    - Encompasses research philosophy, hypotheses, and methodology.
    - Considers ethical issues.
- Chapter 4:
    - Constructs a new framework through stakeholder and context analysis.
    - Designs the value proposition.
- Chapters 5 and 6:
    - Elaborate on the framework building and its validation with a prototype.
    - Offer a reference architecture and a detailed description of the new COBAC framework.
- Chapter 7:
    - Presents the prototype validation.
    - Includes problem identification, objective definition, and evaluation.
- Chapter 8:
    - Discusses the design and development of an optimised framework.
- Chapter 9:
    - Encapsulates the main achievements and contributions to knowledge.
    - Suggests avenues for further research.

2. Literature Review

2.1 Introduction

This chapter aims to review the existing literature with the goal to identify gaps related to the field of autonomous coaching systems. The review is based on the state of the world at the time of finalizing the literature review in 2023 June. It begins by examining coaching, a rapidly expanding practice in the realm of Human Resource Development (HRD) that has evolved into a new scholarly field over the last few decades. Following this, the chapter delves into the effects of digitalization and automation on the coaching industry, practice, and research. The third area is the technology review, which constitutes a major facet of this dissertation, as the latest developments of the Large Language Models have revolutionized the field of Natural Language Processing (NLP). Finally, the chapter concludes with a summary highlighting the identified research gaps, thesis originality, and significance of this research project.

2.2 Coaching

According to researchers, within just four decades, coaching has evolved from a specialised practice that supported human capital into a rapidly expanding and developing field (Koopman, Englis, Ehgrenhard, & Groen, 2021). Despite its triumphant trajectory, the evolution of coaching has encountered a significant impediment, rooted in the ongoing discourse surrounding its definition.

2.2.1 Definition

There is an array of coaching definitions, but there is no unified and broadly accepted definition because it is still a maturing and rapidly changing Human Resource Development practice. Table 2.1 lists a spectrum of definitions from esteemed scholars and preeminent organizations within the field.

Table 2.1: Coaching Definition Overview

"Coaching is thus, collaborative, individualised, solution-focused, results orientated, systematic, stretching, fosters self-directed learning, and should be evidence-based, and incorporate ethical professional practice" (Grant, 2006a)
"Coaching can be defined as the facilitation of mobility. It is the art of creating an environment, through conversation and a way of being, that facilitates the process by which a person can move toward desired goals in a fulfilling manner" (Gallwey, 2008)
"The process of equipping people with the tools, knowledge, and opportunities they need to develop themselves and become more effective" (Peterson, 1996)
"Coaching is a process focused on working with a person's needs, wants, goals, or vision for where they want to go, and then designing steps for getting there" (Stober & Grant, 2006)
"Coaching is a human development process that involves structured, focused interaction and the use of appropriate strategies, tools and techniques to promote desirable and sustainable change for the benefit of the client and potentially for other stakeholders" (Cox et al., 2010)
"A collaborative endeavour between a coach and a client (an individual or group) for the purpose of enhancing the life experience, skills, performance, capacities or wellbeing of the client. This is achieved through the systematic application of theory and practice to facilitate the attainment of the coachee's goals in the coachee's context." (Chittum, 2013)
"Helping people to identify, scaffold, and leverage the learning that is just beyond their current capabilities yet within their reach and necessary for the next step forward." (Drake, 2018)
"It is a professionally guided process that inspires clients to maximise their personal and professional potential. It is a structured, purposeful and transformational process, helping clients to see and test alternative ways for improvement of competence, decision making and enhancement of quality of life. Coach and Mentor and client work together in a partnering relationship on strictly confidential terms. In this relationship, clients are experts on the content & decision making level; the coach & mentor is an expert in professionally guiding the process." (EMCC, 2018)
"A collaborative solution-focused, results-orientated and systematic process in which the coach facilitates the enhancement of work performance, life experience, self-directed learning and personal growth of the coachee." (Association for Coaching, 2020)
"Partnering with clients in a thought-provoking and creative process that inspires them to maximize their personal and professional potential" (ICF, 2019b)

A comparable set of definitions, reflecting those available during the inquiry, was assembled, revealing that none of the widely accepted definitions successfully met the two essential conditions required to define a practice: distinctiveness and universality (Bachkirova & Kauffman, 2009). They also conclude that most of the existing definitions are based on a combination of four perspectives: process, purpose, context, and clientele. To circumvent potential misalignments of expectations between coaches and their clients, they advise coaches to describe these four dimensions explicitly, thereby offering a comprehensive framework to understand and engage with the practice.

The lack of consensus surrounding the definition of coaching presents several challenges within the industry. While a pragmatic approach that emphasizes the four dimensions — process, purpose, context, and clientele — might facilitate managing client expectations and help coaches in establishing a professional identity, it falls short in addressing issues of research comparability. Research outcomes concluded by different groups are difficult to compare when they use significantly divergent perspectives, methods, processes, models, or measure of success on what constitutes coaching. Consequently, research findings remain constrained within a narrow scope, despite the substantial resources — both financial and temporal — invested. These discrepancies could also obstruct the transferability of research findings and fuel a looming “replication crisis” similarly to what was developed in the behavioural science domain during the last decade (Shrout & Rodgers, 2018).

Moreover, the lack of unified and broadly accepted coaching definition extends beyond researchers and significantly impacting coaching clients. The lack of a standardized definition and a unified methodological approach can foster misconceptions and propagate false expectations among clients. In the worst case, it can fuel the perception that coaching is devoid of any scientific evidence or empirical rigor, thus undermining its credibility. It is, therefore, hardly surprising that the absence of "evidence that coaching works" is frequently cited among the principal obstacles deterring clients from participating in coaching, as per a recent report by the International Coach Federation (ICF, 2022a).

Lastly, prominent coaching researchers criticise that it can be also confusing to understand the boundaries of coaching as a multi-disciplinary body of knowledge for novice coaches who participated only in shorter training programs (Cox et al., 2010). For instance, the ICF definition of coaching (ICF, 2019b) could also cover mentoring or consulting activities, which is clearly not the intention of the ICF based on the detailed coach core competences.

The lack of consensus on the definition of coaching had a catalysing effect during the early formative decades of the coaching industry, as it allowed for a broader sense of inclusion, enabling many coaches to align their personal interpretations of coaching with the field, thereby attracting more members to coaching institutes.

Despite these limitations, this thesis utilises the definition put forth by the International Coach Federation. As a global leader in the coaching profession, the ICF exerts considerable influence on both coaching practice and research. Its impact on practice is manifested through its role in shaping the development of the largest coaching community worldwide. In the realm of research, the ICF's significance is illustrated by its numerous research studies, which it either conducts directly or sponsors. Hence, by explicitly employing the ICF's definition of coaching, this thesis aligns itself with one of the most influential and authoritative voices in the coaching field. This ensures that the ensuing discussions and analyses are grounded in a definition that is widely recognised and respected within the global coaching community.

2.2.2 Coach Core Competences

While a universally accepted definition for coaching remains elusive, the need to clearly set the boundaries of coaching is recognised by leading organizations within the coaching industry. Leading organizations employ a comprehensive, competency-based approach to separate coaching from other helping services such as mentoring, training, or counselling. Table 2.2 presents an overview of four competency frameworks that strive to define and standardize coaching. These frameworks were chosen to enable a comparison among the largest global institute (ICF), the leading regional institute (EMCC), and prominent national-level organizations, such as the UK's Association for Coaching (AC) and Australia's Standards Association of Australia (SAA).

Table 2.2: Coaching Competence Framework Overview

Title	Total Competencies	Structure	Institute	Reference
Updated ICF Core Competency Model	63	4 categories, 8 subcategories	International Coaching Federation	(ICF, 2019a)
EMCC competence framework v2	112	8 categories	European Mentoring and Coaching Council	(EMCC, 2015)
Coaching in Organizations	97	4 generic categories, 4 additional categories	Standards Association of Australia	(Chittum, 2013)
AC Coaching Competency Framework	66	9 generic, 3 executive coach categories	Association for Coaching	(Association for Coaching, 2012)

Over the past four decades, competency frameworks have found widespread application across various domains to identify the essential components that contribute to effective professional performance. For instance, competence is defined by Klemp as "a characteristic of an individual that underlies effective work performance" (Klemp Jr., 1980). There are two types of competencies discussed in the literature: threshold competency and distinguishing competency. “A threshold competency is a person´s generic knowledge, motive, trait, self-image, social role, or skill which is essential to performing a job, but is not causally related to superior job performance” (Boyatzis, 1991) whereas a distinguishing competency is required to do a superior job performance.

The structure of the above listed four competency frameworks reflects four significantly different perspectives on the coaching profession.

Firstly, the compilation of the latest version of the ICF core coach competencies (ICF, 2019a) is the result of a global study (ICF, 2020b) that began with semi-structured interviews in conjunction with seasoned coaches. These interviews were followed by two intensive two-day workshops with ten experienced coaches, designed to gather robust data. The harvested data underwent validation through a survey administered to a carefully chosen sample of 1300 participants—comprising both ICF-affiliated and non-ICF coaches worldwide. The published competency framework is organised in four major categories (foundation, co-creating the relationship, communicating effectively, and cultivating learning and growth),

which are further divided into eight subcategories (demonstrates ethical practice, embodies a coaching mindset, establishes, and maintains agreements, cultivates trust and safety, maintains presence, listens actively, evokes awareness, facilitates client growth).

Secondly, The EMCC competence framework v2 (EMCC, 2015) similarly structured into eight - albeit significantly different mentoring/coaching competence categories: understanding self, commitment to self-development, managing the contract, building the relationship, enabling insight and learning, outcome and action orientation, use of models and techniques, evaluation. The individual competencies under these main categories exhibit great variety, making their comparative analysis a task of high complexity. The complexity involved in the ICF and EMCC competency framework comparison is illustrated in a report, which was prepared based on the comparison of previous versions of the referenced competency documents (Lindvall, Holm, Ross, van Vlerken, & Margariti, 2013). The team responsible for analysing the previous versions of the competency framework documents deduced that 84% of the ICF's demonstrative behaviours could be identified within the EMCC framework, and this rate, when reversed, stood at 60%. Moreover, the researchers underscored the additional complication presented by terminological disparities, as the two organizations employ different terms even for foundational definitions such as 'evaluation' or 'client'. It was also noted that the respective documents can only be understood in the context of the accreditation process. For instance, the ICF framework - unlike the EMCC framework - features only observable demonstrative behaviours beyond section 'A' within a coaching conversation. This is because the ICF document serves not only as the foundation for coaching competence development training but also acts as a cornerstone for the coach accreditation process.

Thirdly, the Standards Association of Australia (Chittum, 2013) adopts a notably distinct approach in defining coaching competencies. Competences divided into “generic coaching competences and supporting knowledge bases” and “additional core competences” based on the four specific coaching types: skill coaching, performance coaching, developmental coaching, and remedial coaching.

Fourthly, the Association for Coaching (Association for Coaching, 2012) splits the competencies based on client groups into generic coaching and extends the list of competences with additional categories for executive coaching.

Despite these four diverse approaches to defining coaching competencies, an important caveat regarding their validity was raised (Boyatzis et al., 2022). They caution that most competency frameworks are primarily the product of gathered expert opinions, lacking substantial supporting evidence or the application of rigorous research methods. This suggests a need for a more evidence-based, empirically rigorous approach in the construction and validation of coaching competency frameworks.

In addition to the above-mentioned coaching competency frameworks, there is only a limited number of studies with published research methods and data. One notable example is (Horne, 2008) competency-based research in the area of managerial coaching, conducted with a sample size of 40. This study suggested a tentative coaching model for managers who integrate coaching skills into their managerial duties. Another study (A. S. Maxwell, 2017), examined executive coach competencies using a sample of 17 executive coach practitioners. From a proposed 39-item competency instrument, 22 were identified as contributing to the effectiveness of executive coaching. Common characteristics of both studies include small sample sizes and inconsistent findings, which echo the recommendation of several researchers (Nadeem, Garvey, & Down, 2021), (Griffiths & Campbell, 2008), (Boyatzis et al., 2022) that further evidence-based research of the core coaching competences is required to establish a scientific base for the coaching industry because the current contemporary frameworks are only legitimized by practitioners based on their personal experiences.

As indicated in sections 2.2.1 and 2.2.2, terminology, definitions, and competencies associated with coaching vary significantly from one geographic region to another and differ among accrediting organizations. The ramifications of these competency frameworks are substantial for the industry's future as they guide not only the curricula of coach training programs but also the accreditation of individual coaches and shape research activities, including this doctoral thesis. The competency frameworks discussed herein significantly contribute to this thesis, as they served as the initial basis for deriving the requirements for

the proposed autonomous coaching related theoretical framework. Each coaching competency framework is comprehensive, with significant overlapping elements. This indicates that, despite their distinctive features, there is a common set of competencies that few experts would dispute as forming the backbone of coaching. Since the objective of the COBAC framework is to focus on the most defining characteristics of coaching, the chosen approach was not to adopt one framework or another but to identify the intersection of the two most influential frameworks and use the shared competencies as the foundation for constructing the COBAC model.

2.2.3 Coaching Body of Knowledge Development

There is a consensus among scholars that the volume of academic research related to coaching has increased significantly over time (Grant, 2011; Koopman et al., 2021). According to Brock the history of coaching as an interdisciplinary field was initially based on body of knowledge from different domains, such as behavioural psychology, human development and motivation theories (Brock, 2012). Other sources are extending the list of root disciplines with psychotherapy, counselling, sociology, adult education, communication, management consulting, organizational development and training (Bachkirova, 2014; Brock, 2008; Stec, 2012). The knowledge transfer between these disciplines is not confined to the core concepts and theories alone; it also influences the research philosophy, methods, and applied measures.

However, surprisingly few studies assess whether a specific concept, proven to be effective in one field, can be transferred to coaching without losing its efficacy. For instance, De Haan at al. were validating the “working alliance” concept, which is originated in the field of psychotherapy (De Haan, Molyn, & Nilsson, 2020). These researchers concluded, based on a longitudinal, randomized-controlled-trial study, that the impact and effectiveness of the "working alliance" concept within the coaching context were limited to only the initial phase of the relationship.

Consequently, without efficient and effective methods for validating the transferability of these concepts and theories, the research gap between coaching practice and research is likely to continue expanding.

2.2.4 Boundary Issues

The lack of clear boundaries between psychotherapy and coaching leads to professional and ethical issues (Grant, 2006b; A. Maxwell, 2009). Even though the distinction between coaching and psychotherapy is well-understood in theory and only overlaps in practice, but at the same time there is a fundamental debate going on among scholars how to distinguish mentoring from coaching. For instance, the International Coaching Federation (ICF) is focusing only on coaching (ICF, 2019a), however its European counterpart the European Mentoring and Coaching Council (EMCC) defines the two terms separately (EMCC, 2018) but covers both in one competence framework (EMCC, 2015). Similarly, it was concluded that they are so closely related concepts and also developed historically in a similar way that they should be treated together (Koopman et al., 2021). Other researchers are suggesting to consider mentoring as a subdomain of coaching (Boyatzis et al., 2022).

Lastly, as an alternative approach some scholars suggest moving away from the artificially created discipline boundaries, rather focusing on potential opportunities to explore synergies, and adopting more client-focused practices (Bachkirova & Cox, 2004; Cavanagh & Buckley, 2014).

Given the lack of clarity and consensus, the research project covered in this thesis has intentionally excluded competences related to mentoring, concentrating solely on core coaching competences to reduce research complexity. A further rationale for this decision is the result of an ICF client awareness study, which found that only 26% of clients deemed industry or job experience within their domain as highly significant during the coach selection process (ICF, 2009), which indicates that mentoring related skills, competences and knowledge is not a top priority for the clients either.

2.2.5 Challenges in Coaching Research

Clients consider “lack of evidence that coaching works” among the top barriers to start engaging in a coaching service according to recent ICF global consumer studies (ICF, 2017a,

2022a). Conversely, Grant argued based on the analysis of 234 outcome studies published between 2000 and 2011 January that a significant share of research activities was dedicated in the last decades to outcome research (Grant, 2012). The research results of the outcome-based studies indicate overwhelming scientific consensus on a significant positive outcome of the coaching sessions (Athanasopoulou & Dopson, 2018; Grover & Furnham, 2016; Jones, Woods, & Guillaume, 2016; Theeboom, Beersma, & van Vianen, 2014). Despite these encouraging results, a growing gap persists between scientific research and practical application, a discrepancy that cannot be wholly attributed to the ongoing debate surrounding the definition of coaching. Further factors contributing to this disjunction warrant examination.

Firstly, Linley argues that the fundamental reason for the divergence between the research field and practice is the substantial cost associated with research, compounded by the high opportunity cost borne by participants (Linley, 2006). Additional costs are incurred in relation to the methodologies employed in conducting research. For example, a commonly used approach in coaching process research is predicated on the textual analysis of transcripts from human-to-human coaching sessions (Harald Geissler, 2009). The process of manually transcribing audio or video recordings of real-life coaching sessions incurs significant cost and considerably prolongs the duration of research.

Secondly, apart from direct financial considerations, a recurrent limitation of coaching process research is the restriction on sample size (Grant, 2011; Jones et al., 2016), which undermines the validity of studies significantly (Button et al., 2013).

Thirdly, akin to other behavioural research studies, most coaching research tends to be based on homogeneous populations, thereby limiting the scope for generalizing the findings (Henrich, Heine, & Norenzayan, 2010).

Fourthly, it was emphasized that scientifically-based coaching outcome evaluations are more important in the coaching domain compared with other professions because “coaching is not only difficult to evaluate on the basis of its intangibility, opacity and complexity, but also because of the confidential nature of the conversations” (Bachkirova, Spence, & Drake, 2016).

In conclusion, it was summarized the impact of these challenges by stating that “with these major barriers to research, one may begin to consider that it is surprising that any research is done at all!” (Linley, 2006).

Despite the pressing need to address these challenges, there appears to be a lack of concerted efforts within the scientific community to mitigate their effects. It is crucial to reduce the impact of these issues, as the gap between scientific research and practice will persistently expand in the absence of substantial improvements on these structural impediments. Should these issues remain predominantly unresolved, the evidence-based research approach risks losing its relevance, thereby undermining its potential to guide coach trainings and stimulate the further development of the coaching industry.

2.2.6 Future Trends

A contemporary snapshot of the evolving dynamics within the coaching industry was encapsulated in a report based on survey responses gathered from coach practitioners (Jonahtan Passmore, 2021). Key insights about future trends in coaching were identified, namely, the amplified role of coaching in group and team contexts, a growing demand for online coaching sessions, and the positive impact of coaching on client well-being. An earlier version of the report (Jonathan Passmore, Brown, Csigas, & European Coaching and Mentoring Research Consortium, 2017) merely acknowledged the well-being aspect as a recognised outcome of the coaching process, signalling a significant shift in the industry's focus.

To start with, group and team coaching, which entails engaging an entire team or a group of people with a shared objective in a coaching conversation, has become a ubiquitous tool for organizational development in the past decade. Not only coaching accreditation bodies but also organizational development associations are providing competency frameworks and certification programs. To name a few, ICF published the Team Coaching Competencies (ICF, 2021) in 2021 and the first pilot program to grant the “Advanced Certification in Team Coaching (ACTC)” certification was concluded in 2022. Likewise EMCC initiated the Team

Coaching Training Quality Award (TCQA) training program and the related Team Coaching Individual Accreditation (ITCA) (EMCC, 2020). These programs were established long after organizational development focused organizations have launched similar programs. For instance the Scrum Alliance launched its Certified Team Coach and Certified Enterprise Coach programs in 2015 (Scrum Alliance, 2015). The focus of this doctoral thesis was confined to individual coaching due to the significantly higher complexity of team and group coaching and its relatively sparse scientific base (Widdowson, Rochester, Barbour, & Hullinger, 2020).

After that, the digitalization of coaching service delivery was underway for more than a decade (Sherpa Consulting, 2018).

Lastly, the focus on well-being as a coaching outcome stems from the global decline in mental and physical health. (WHO, 2022).

In summary, the future trends identified by coaching practitioners can be explained by the ongoing recent changes in the world. However, none of them substantially address the major structural issues and challenges in the coaching field.

2.2.7 Ethical Issues

One area that has undergone significant and structural changes in recent decades is the ethical code used by coaching practitioners. The Global Convention on Coaching has prioritised the topic of clarifying the ethical factors that influence the formation of the coaching profession and one of their key recommendation was to create a universal ethical code from the five available codes (Stout-Rostron, 2009). As a result, the Global Code of Ethics (GCoE) was created within a decade and signed by all major organizations of the coaching field except the industry leader ICF (Association for Coaching & European Mentoring & Coaching Council, 2019), which maintains its own code of ethics (ICF, 2019b).

According to the 2018 Ethical Conduct and Compliance Report (Independent ICF Review Board, 2019) there were only 23 formal complaints in 2018. By comparison, other professional accreditation bodies in related fields often report higher volumes of annual

complaints. For example, the Nursing and Midwifery Council, which oversees multiple regulated professions, recorded over 5000 new fitness-to-practice concerns in the 2018–2019 period alone and a similar magnitude year-over-year (Nursing and Midwifery Council, 2019). This figure, while not directly analogous to the coaching profession, provide a sense of scale and illustrate that some professional bodies deal with substantially greater complaint volumes, partly due to their size and the robustness of their reporting systems. The small number of yearly reported issues indicates that only a fraction of ethical issues is reported through the accreditation bodies, implying that transparency over compliance with the ethical code remains a significant issue.

Despite the lack of transparency concerning breaches of the ethical code, ethics play a more prominent and visible role in coaching practice compared to other industries. This is evidenced by a global study, which found that 56% of coaches share their ethical code with their clients, and 32% communicate it at the start of the coaching conversation (Jonathan Passmore et al., 2017).

Even though the ethical codes published by major coaching accreditation bodies are quite extensive, there are ongoing debates among researchers about certain aspects not specifically covered in the current versions. For example, it was argued that while coaches might assist their clients in achieving individual goals, these individual objectives might also contribute to the accelerated "burning of the planet" (Boyatzis et al., 2022).
Similarly, Han has warned that burnout represents the pathological consequence of voluntary self-exploitation in achievement-focused societies (B.-C. Han, 2015) and coaching may inadvertently contribute to burnout by placing excessive emphasis on goals.

Simultaneously, Mercaldi engages in a discussion about the ethics of developmental coaching (Mercaldi, 2018), suggesting that at least six ethical aspects should be considered: involvement of the self, individual freedom, internalisation of problems that originates in external factors, lack of critique to the social context and contribution to the ‘society of achievement’.

Lastly Schermuly and Graßmann assert that the potential negative side-effects on both coaches and clients should be better understood and communicated (Schermuly & Graßmann, 2019).

Overall, this brief review of contemporary ethical debates demonstrates that increased transparency and evidence-based research could promote the development of the coaching field. It also underscores the importance of structuring the coaching process, research, and development of the field around ethical considerations, rather than focusing solely on "what works."

2.2.8 Benefits

The primary benefit for a client engaging in coaching is to accomplish or progress toward their goals. Coaching goals can differ widely, encompassing situational needs like guiding leaders or key employees through the initial months following a promotion, as well as focused skill development, such as improving interpersonal communication, mastering time management, or improving performance to achieve specific business objectives (ICF, 2009).

Similarly to the field of psychotherapy, there is a growing number of studies researching common factors, active ingredients, and input factors that influence coaching outcomes (De Haan, Duckworth, Birch, & Jones, 2013; De Haan, Grant, Burger, & Eriksson, 2016; De Haan et al., 2020; Molyn, 2020; van Coller-Peter & Adriana de Vries, 2022).

Beyond goal achievement, a meta-analysis on the effects of coaching conducted in the organizational context has confirmed significant positive effects on "performance and skills, well-being, coping, work attitudes, and goal-directed self-regulation" (Theeboom et al., 2014).

The same message was conveyed by an other researcher, who encouraged the coaching community to move beyond the functionalist view of measuring outcomes and effectiveness (Mulvie, 2015). Her research focuses on the executive coaching domain, where the outcome is not confined to the client's individual objectives and personal development, but also

impacts others in the organization and contributes to the coaching culture of the organization itself.

Moreover, the advantages of coaching are not solely restricted to individuals or groups. The majority of coaching practitioners also hold the belief that coaching exerts a significant influence on societal change (PwC & ICF, 2016).

Nonetheless, a substantial research gap is present between coaching competency frameworks and the generation of value. There's no guidance regarding when or how often to use the specified competencies, in what combination or sequence to apply them, and how they contribute to value generation. Contemporary coaching competency frameworks employ expressions to describe intermediate outcomes, such as insight, learning, action plans, or feedback, without establishing a connection to the generation of value.

2.2.9 Evaluation

Krapu suggests that a value proposition-based measurement, which includes both efficiency and effectiveness measures, might be required to genuinely evaluate coaching in its original context, rather than drawing conclusions about a business practice based on studies conducted in artificially constructed environments that do not replicate all business-critical factors. He also argues that a reductionist approach toward coaching might “invite a view of coaching as nothing more than a utilitarian set of tools to accomplish discrete, individual goals” (Krapu, 2016).

Despite the highlighted shortcomings of the reductionist approach, there is a wide variety of frequently applied measurements in the coaching literature. These can be classified in various ways; one aspect-based taxonomy is summarized in Table 2.3.

Table 2.3: Coaching Measurement Taxonomy

Measurement Aspect	**Description**
Frequency	Some measures can be repeated one time or many times during or after the coaching engagement to evaluate changes caused by coaching, depending on the study design and purpose.
Time horizon	Short-term and long-term measures are used to measure whether a change created by coaching was temporary or permanent; in other words, immediately after or a long time after the coaching engagement has concluded.
Data source	Information can be collected from different stakeholders, such as the coach, the client, coach commissioners or sponsors, or even external stakeholders who might not be directly involved in the coaching process but whose perspectives might be valuable. For example, a client's frequent negotiation counterparts might be asked about the client's performance as a negotiator.
Outcome	These measures assess changes related to the coaching engagement. Examples include goal attainment measures and client well-being measures. The Multiple-Perspective Model (Grant, 2017) can be a good basis to categorise areas of outcomes.
Process	These measures focus on identifying common and specific factors in coaching engagements to gauge their effectiveness and relationship to outcomes.
Validated vs custom	Validated measurements are instruments that are tested for reliability, validity, and sensitivity. Custom measures are specific to the study, client, client goal or client-context. For example, a database extract related to the number of successful sales concluded by the coaching client is a custom measure specific to the client's organisation.
Expert evaluation	An experienced coach evaluates interaction between another coach and a client based on observation (shadowing), recording or transcript of recording to measure level of compliance to pre-defined competences.

The following paragraphs provide a brief and by no means complete overview about what coaching relevant measures are applied frequently or discussed in the coaching scientific literature.

Suitable examples of coaching input measures include the Brief Coaching Readiness Scale (BCRS) (Franklin, 2005) and the Developmental Readiness Questionnaire (DRQ) (MacKie, 2015). The BCRS is a 17-item non-validated scaling instrument that measures mental constructs like readiness to change, while the DRQ is a 14-item questionnaire that assesses underlying constructs related to developmental readiness. Furthermore, the concept of "coachability", which originates from the sport industry was also adopted and used in the organizational and executive coaching context (Bacon, 2011). These input measurements for coaching are effective examples because the measured factors can predict goal attainment.

As the primary purpose of coaching is to achieve goals, it seems natural to use goal attainment related standardised measures. Even though the theory of goal setting and motivation is a well-researched area (Locke & Latham, 2002) there are not many applicable frameworks to measure it. Spence argues that Goal Attainment Scaling (GAS) is one of the applicable mental constructs in the field of coaching to satisfy the growing demand for evidence-based coaching (Spence, 2007). Although it was originally developed to evaluate mental health programs (Kiresuk & Sherman, 1968), its application was expanded in the last decades in many areas, such as psychotherapy, which is one of the most influential knowledge domain related to coaching. The implications of GAS was investigated in coaching context (Prywes, 2012) and used in various coaching contexts, for instance in a health coaching study (Spence, Cavanagh, & Grant, 2008) and a well-being study (O'Connor & Cavanagh, 2013).

An alternative validated and frequently used measure in applied psychology is the Core Self-Evaluation Scale (CSES) (Judge, Erez, Bono, & Thoresen, 2003), which is applicable for measuring the four major psychological traits: self-esteem, generalised self-efficacy, neuroticism and locus of control. There are many examples of its application in coaching specific studies: leadership coaching (MacKie, 2015) and health coaching (Tülüce & Kutlutürkan, 2018).

However, not only validated measures, but also goal-specific external measures can be used to assess outcomes. Sales performance, an external and goal-specific compound outcome measure was used in a study related to sales executive coaching (Libri & Kemp, 2006). It was defined as a weekly average of related measures: amount of client leads, client loan interviews, submitted loan applications and approved loan applications.

In addition to input and outcome-based measures, process measures are also used from both client and coach perspectives. A process-oriented measurement approach, coaching effectiveness (CE), is related to coaching as a complete process, including but not limited to the outcome. It uses a 4-point scale and assesses the created safe space, learnings, behaviour change, and overall outcome (De Haan et al., 2016). Unlike the coaching effectiveness questionnaire, the accreditation process for coaches conducted by accreditation institutes is robust and involves very detailed external assessments. For instance, as part of the ICF

accreditation process depending on the level of accreditation, among many other requirements, one or multiple audio recordings of live coaching sessions, along with the associated transcripts of the recordings, must be provided for assessment purposes. Multiple trained assessors, usually senior coaches, evaluate the recordings and transcripts, looking for evidence of demonstrative behaviours appropriate to the accreditation level (ICF, 2017b). This external competence assessment method is a manual, effort intensive approach but it can provide a rich source of insights related to coaching competencies that the coach may consider for future improvement. The method is also frequently used as part of coach competence training for skill development.

Lastly, measurements are not only used to measure efficiency or effectiveness of the coaching engagement, but also to provide content to the coaching conversations. For instance, psychometric testing, as a frequently applied instrument in coaching, offers many use-cases and benefits, such as increasing client self-awareness, fast-tracking coach and client at the beginning of the coaching engagement, providing structure for coaching conversations as a "question-aid", introducing new sources of data for stimulation, or measuring change at the end of the engagement (McDowall, 2017). Tests can measure personal preferences, values, skills, performance or personality traits, for instance the Big Five personality traits, also known as OCEAN (Gosling, Rentfrow, & Swann, 2003) with varying length (Furnham, 2008). A psychometric test measuring personality attributes might be used in any coaching conversations, but some assessments are specific to coaching goals. For example, in case of health coaching stress-reduction might be measured with the Perceived Stress Scale (Cohen, Kamarck, & Mermelstein, 1983).

There are a lot of critiques and challenges associated with the ways how, when and which coaching characteristics are measured.

To begin, one of the most frequently noted critiques regarding research quality in coaching outcome research is the scarcity of standardized and validated measures used in studies (Jonathan. Passmore, Peterson, & Freire, 2013). For instance, the Coaching Effectiveness Survey (CES) instrument was validated successfully (Tooth, Nielsen, & Armstrong, 2013) and

the researchers advocate the employment of this tool as a starting point to measure coaching effectiveness.

Following on from that, an obvious downside of measurements that they cost time, so they need to provide substantial value to compensate the associated direct cost and opportunity cost. McDowall argues that the added value of measurements amplifies when applied sparingly (McDowall, 2017). Concurrently, he underscores the evident research gap concerning the usefulness of multiple psychometric assessments within a single coaching process.

Finally, an apparent discrepancy arises when coaching is viewed alternately as a service or a process. Process researchers often use measurements after the coaching conversation as an administrative activity that serves a scientific purpose but does not belong to the core coaching service. However, this perspective seems incongruous with the European Mentoring and Coaching Council's (EMCC, 2015) stance, which not only includes evaluation as a competence but highlights it as a main competence category with four defined sub-competences. This categorization implies that measurements are indeed intrinsic to coaching, thereby challenging the perspective of decoupling the coaching process from measurements.

In conclusion, cultivating a culture of evidence-based development and research within the realm of coaching encounters an array of obstacles. These include the absence of a unified, unique, and widely accepted coaching definition; lack of scientifically validated competence frameworks; the underutilization of validated measures; and the prohibitive cost of research methods. Each of these factors cumulatively contribute to the continuously widening gap between coaching practice and research. Consequently, without a radical transformation in the approach to coaching research, the existing methodologies remain fundamentally unscalable.

2.3 Autonomous Coaching

The pervasive wave of digitalization and automation, identified as a significant megatrend (Vukanović, 2018), markedly influencing the coaching industry among numerous other sectors. This segment delineates the ramifications of digitalization and automation on the coaching industry, encapsulating its practical and research-oriented dimensions.

2.3.1 Definition

Automation, as defined by the Oxford Learner's Dictionary, is "the utilization of machines and computers to perform tasks previously executed by humans" (Oxford University Press, 2020). This phenomenon manifests on a spectrum, with ten discrete levels spanning from no automation to autonomous operations without human intervention, as described in a proposed model (Parasuraman, Sheridan, & Wickens, 2000). These scholars designate the highest level of this continuum as autonomous or fully automated systems.

The digital transformation and incremental automation within the coaching service sector was evolving for over a decade (Sherpa Consulting, 2018). Nonetheless, the scholarly discourse surrounding technologically augmented coaching exhibits a plethora of overlapping terminologies. Terms such as "online-coaching" (Koch, 2012), "distance coaching" (L. Wang & Wentling, 2001), "e-coaching" (H Geissler, Hasenbein, Kanatouri, & R., 2014; Kamphorst, 2017), "technology assisted coaching" (Bachkirova et al., 2016) and more recently, "AI coaching" abound.

"AI coaching" was also defined as "a machine-assisted, systematic process to help clients set professional goals and construct solutions to efficiently achieve them" (Graßmann & Schermuly, 2021). Despite its restrictive nature of the definition — where the term "professional" unnecessarily narrows the scope to business coaching, thus omitting life coaching — this contemporary term seems to set the conceptual basis best for the emergent breed of autonomous coaching systems.

As discussed above, the research about coaching automation is still very immature therefore, no consistent terminology is used in the research community for the varying automation levels. This thesis uses the term “autonomous coaching system” or “AI coaching system” to refer to a system that provides an end-to-end coaching service for a human client without any interference by a human other than the client. Herein, 'end-to-end' denotes a service encompassing all cardinal functions from the initial client engagement through to the conclusion of the coaching relationship.

2.3.2 Limitations of Existing Automated Coaching Solutions

Over a decade ago, the dawn of coaching automation emerged with a pledge to alleviate coaches from non-core coaching related activities. This innovation promised a suite of tools and platforms designed to let coaches focus on their core competencies, thereby, propelling the coaching profession into a new era. Some notable advancements include:

1. Distance Coaching Enablement: Platforms such as Zoom, Microsoft Teams, and Skype have bridged geographical divides, making remote coaching a reality.
2. Smart Sensor-Based Training Data Collection: in amateur sport coaching context - “personal informatics applications are increasingly available for amateur endurance athletes to record and monitor their performance and training” (Wakefield, Neustaedter, & Hillman, 2014).
3. Automated Psychometric Test Execution: This automation assists in gauging client's psychological metrics efficiently (Harper, 2008).
4. Feedback Procurement: Tools for gathering 360-feedback from groups have streamlined the feedback process, providing a wealth of insight for coaching improvement (Özgen, Alabart, & Medir, 2008).
5. Marketing and Client-Coaches Matching: Platforms like CoachHub or Coaching.com have emerged to facilitate a seamless pairing of clients and coaches (CoachHub, 2020; Coaching.com, 2024).

The refinement of these tools has continued to advance, capturing substantial productivity and efficacy gains in automating supporting tasks. The majority of repetitive, administrative,

or otherwise mundane tasks encompassing an end-to-end human coaching service were partially or fully automated in recent years.

The development of the second type of automation, which aims to augment a human coach, has commenced slowly due to the scarcity of contemporary examples or research on tools that collaborate with coaches for client benefit. Even if there is a hybrid business model like the initial commercial offering of the PocketConfidant app (PocketConfidant, 2023) suggests, the company proposes a symbiotic existence of human and algorithmic models, albeit with ambiguous interaction mechanisms. Further insights are provided by the Rocky.ai app (Techstars, 2020), which offers for coaches a dedicated service with functions like: 'daily morning mindset' or daily evening reflection. They claim to target the following benefits for coaches: increased follow through, boost coaching result, higher client retention, touch point with clients in between sessions, personal support at the right time and scaling the coaching program of the coach. The only scientifically grounded example for this second type of automation is combining telephone coaching with an internet-based programme (H Geissler et al., 2014). The authors conclude that the virtual coaching tool used during a telephone conversation compensate for missing visual cues inherent in telephone coaching, hence comparing the telephone coaching with the telephone coaching based on virtual tools were reducing the ambivalence associated with telephone coaching on the coach side. This research, however, predominantly focuses on the client, leaving the coach's perspective less explored. Unlike the initial non-core coaching competence-oriented tools, the augmentation-oriented tools do not seem to offer much value for the coach, however there is a notable gap of knowledge due to lack of sufficient data and research to make any meaningful conclusion.

The third type of automation is related to the autonomous systems that focus on the core competences of a human coach and aims to replace a human coach. There are several examples of standalone autonomous coaching services as it is shown on the Table 2.4.

Table 2.4: Existing Autonomous Coaching System Examples

Topic	Description
Food coach	Rupert, "whose goal is to coach people who wish to improve their food lifestyle" (Casas, Mugellini, & Khaled, 2018).
PocketConfidant	"make personal growth, reflection, learning and change, available to everyone" (PocketConfidant, 2023)
Mental health	Wysa, "AI-enabled, empathetic, text-based conversational mobile mental well-being app" (Inkster, Sarda, & Subramanian, 2018)
Cognitive behavioural therapy (CBT)	Woebot, "self-help content derived from CBT principles in a conversational format with a text-based conversational agent" (Fitzpatrick, Darcy, & Vierhile, 2017)
Leadership coaching and development	Rocky.AI, "5 minute leadership coaching and self-reflection questions" (Techstars, 2020)
Habitica	Gamified goal management (Habitica, 2020)

These examples focus on a specific target segment and lacks an abstract, theoretical model for quick adaptation of these solutions in other domains. Moreover, there is ambiguity about what does the system consider under "coaching", the applied coaching competencies, the workings, and limitations of these systems.

As a recent review about the virtual coaching systems in the healthcare domain concludes: "Transition to fully automated in home coaching systems, however, is proving to be in its early stages and many obstacles still remain before holistic coaching solutions with real-time evaluations and personalized feedback can become widely available for unsupervised in home use" (Tsiouris, Tsakanikas, Gatsios, & Fotiadis, 2020).

Moreover, Facebook's AI Research organization highlighted the complexity of automating human interaction, as an adept conversationalist seamlessly blends various skills such as providing engaging talking points, active listening, displaying knowledge, empathy, personality, while maintaining a consistent persona (Roller et al., 2020). Technical and methodological challenges in this domain suggest that despite recent advancements, achieving a level of conversation akin to human interaction in automated systems may not occur in the near future (Li, Weston, & Roller, 2019). However, this notion was challenged by the emergence of Large Language Model technology, particularly with the launch of OpenAI's ChatGPT chatbot in November 2022 (Schulman et al., 2023).

One of the last enduring challenges remains the application adoption because even if the autonomous coaching application is working according to the highest functional and non-functional standards, there are several adoption factors to be considered in the design of the application as proposed by multiple studies: performance expectancy, effort expectancy, social influence, facilitating conditions and perceived risks (Terblanche & Cilliers, 2020; Terblanche & Kidd, 2022). A similar study affirmed some of these concerns and introduced positive attitude toward the application as an additional factor (Terblanche, Molyn, Williams, & Maritz, 2023). Lastly, anthropomorphism should be also considered when designing an autonomous coaching application as it enhances client satisfaction significantly (Weber, Lömker, & Moskaliuk, 2021).

This concise literature review on existing automations in the coaching domain reveals several limitations of modern solutions: unclear communication on the purpose of the solutions - whether they are automating partial non-core coach tasks, augmenting a human-to-human coaching process or designed to completely replace the human coach as a standalone solution; lack of clarity on the conceptual basis, coach competencies and scientific evidence underpin the solutions; and lack of transparency on what the applications designed to do exactly and why.

2.3.3 Existing Autonomous Coaching Frameworks

The literature review about the design features of conversational agents in the eHealth domain (ter Stal, Kramer, Tabak, op den Akker, & Hermens, 2020) suggests that there is no shortage of approaches to identify the key considerations for an effective conversational system, however the overall field is considered immature.

The Designing AI Coach (DAIC) framework (Terblanche, 2020) stands out as a positive exemplar, which proposed a design framework based on four principles of effective human coaching, five design principles and eight design considerations. It is grounded in the scientific exploration of the coaching domain, showcasing an exceptional level of transparency and an evidence-based approach in tracing the framework elements. Several subsequent publications shed light on additional conceptual elements not covered in the initial study,

such as the high-level process and conversation flow (Terblanche & Kidd, 2022) based on the Facebook Messenger platform, which later transitioned to the Telegram platform (Terblanche, Molyn, Haan, & Nilsson, 2022). The framework is designed with a single-purpose in mind: goal-achievement and to evaluate AI coaches it proposes: “perhaps researchers could use existing coaching competency frameworks, such as those of the International Coach Federation (ICF), as a guide to evaluate AI Coaches” (Terblanche, 2020).

Despite the significant and unique virtues of the DAIC design framework, there are several critiques. Firstly, if the evaluation is to be grounded in a coaching competency-based approach, it should be reflected across various facets of the application design: process flow, data architecture, and requirements. Secondly, the value proposition of an AI coach should extend beyond mere goal attainment, encompassing the full spectrum of values that human coaches create, combined with the advantages delivered through automation.. Thirdly, the proposed model does not exhibit a systematic application of contemporary requirements engineering best practices and standards, leaving room for enhancement in model quality attributes such as completeness and consistency. Fourthly, the platform switch from Facebook Messenger to Telegram underscores a belated acknowledgment of confidentiality concerns, despite trust and ethical considerations being identified as core elements in the initial model. Establishing traceability from the objectives, requirements through the technology selection could have likely prevented this change.

None of the other proposed models and frameworks come close in terms of comprehension despite considering the shortcomings of the DAIC model; however, some provide various extensions, alternative perspectives, or additional insights into modelling an autonomous coaching system. For instance, insights were provided into some aspects of the applied requirements in the form of the “Structural Model of the Effects of Results-Oriented Individual Coaching” (Weber et al., 2021) while others provide a proposed “loosely coupled architecture of an e-coaching system” (Ochoa & Gutierrez, 2018). A more extensive study on a counselling system offers a high-level view on module hierarchy, dialog strategy architecture, and Natural Language Processing-based utterance generation process (S. Han, Lee, Lee, & Lee, 2013). A specialized coaching and training system is even patented in the

United States (Obispo et al., 2016), with the patent description providing insights into use-cases, modules, architecture, and key processes.

Admittedly, designing autonomous coaching systems is a high-complexity task. The example models above exhibit a level of conceptualization that lacks modern requirements engineering practices, completeness, and traceability of objectives, requirements, design principles, and architecture. The consequence of not having a robust, systematic approach to modelling is that the current autonomous coaching ecosystem remains subject to fundamental flaws, limiting its utility in advancing evidence-based scientific discovery on how to enhance the efficacy of AI coaching.

2.3.4 Ethical Issues

As discussed in Section 2.2.7, the development of autonomous coaching systems requires a holistic approach that goes beyond evaluating the efficacy of the solution ("what works"). It is imperative to integrate ethical considerations into the design phase of these solutions. Automation introduces a myriad of individual and societal ethical dilemmas, augmenting the already existing ethical concerns to human-to-human coaching. This sentiment is echoed by a recent report illustrating a disheartening reality concerning the current status of applied ethics: a lack of transparency on how AI supported decisions are made, a struggle to get executive support for AI ethic practices and an absence of monitoring for fairness and ethics (Zoldi, 2021).

On the coaching front, the confidentiality of client data has emerged as a paramount ethical concern in coaching, particularly with the exponentially growing use of audio-video platforms for delivering human-to-human coaching services (ICF, 2023). Practical measures to enhance the protection of sensitive client data in cyberspace are well-documented (Fouracres, 2022).

At the individual client level, the exploration is in its infancy regarding how technology usage is diminishing our collective attention span (Lorenz-Spreen, Mønsted, Hövel, & Lehmann, 2019), inducing physiological alterations such as “text neck syndrome” (Baokar, Shrikrishna &

Wamane, Vikas & Taware, Ganesh & Jadhav, 2018), and adversely affecting our well-being (Shakya & Christakis, 2017).

Table 2.5: Ethical Issues Associated with Coaching Automation

Ethical Issues	**Reference**
reveal non-humanness self-disclosure on extent and limits of capabilities transparency on confidentiality asks for consent to start the coaching process algorithmic transparency explainability of AI models and system design AI biases AI system training data and process flow quality control	(ICF, 2022c)
privacy fairness considerations personal autonomy new source of authority in the medical or health context	(Kamphorst, 2017)
problematization of unhealthy living (health context) autonomy lack of robustness transparency	(Niezen, Adams, Nadya, & Vedder, 2016)
privacy autonomy liberty fairness and equality responsibility authenticity	(Anderson & Kamphorst, 2014)

Another recurrent ethical concern related to automation is bias, specifically over-trust in automated systems (Wagner, Borenstein, & Howard, 2018) and algorithmic bias inherent in the systems, methods, or data (Akter et al., 2021).

At the societal level, job displacement, alteration of job responsibilities, and the requisite re-skilling are the primary concerns (Hawksworth, Berriman, & Goel, 2018). To navigate these challenges, researchers and technologists in the field propose AI governance principles (Future of Life Institute, 2017).

The literature on coaching automation lacks consensus regarding the core ethical issues, as depicted in Table 2.5.

Several recurring themes emerge from the collection above, including autonomy, privacy, and transparency; however, each source also identifies unique aspects, some of which are context- or industry-dependent. Consolidating these ethical issues, tailoring them to the application's context, and providing solutions for each are the ongoing responsibilities of the autonomous coaching system designer. As Tzafestas articulates this gap: "there is still a strong need to develop more practical, and easy to understand and apply, ethical rules and codes for the designers, professionals, and users" (Tzafestas, 2018).

2.3.5 Benefits

Beyond the key ethical concerns and drawbacks summarized in Section 2.3.4, automation also brings a multitude of benefits to clients and society. Neumeier, Wolf, and Oesterle have categorized typical benefits of digital transformation based on a meta-analysis of 29 articles and studies into five categories: customer, business model, business processes, application systems, and services and infrastructure (Neumeier, Wolf, & Oesterle, 2017). The researchers identified 38 distinct benefits, such as customer experience, service quality, innovative services, cost reduction, process flexibility, knowledge management and customer insights.
Moving closer to the coaching domain, a recent qualitative study related to digital interventions in health care highlighted the following four benefits as perceived by clients: “convenience and flexibility for interventionists and recipients, support for implementing evidence-based interventions with fidelity, scaling and improving access to interventions, and getting a foot in the door despite stigma and disenfranchisement” (Marcu et al., 2022).

Moreover, in the augmented coaching context, improved self-resilience was named as a key value for clients (Ellis-Brush, 2021), whereas a research on AI coaching (Terblanche et al., 2022) lists the following potential benefits: new modes of treatment, the ability to reach currently excluded populations, improve patient response, free up limited resources, accessibility, vastly reduced cost and scalability. In addition to cost savings, on-the-job support, improved self-revelation, non-judgmental nature of chatbots, shorter coaching units, increased control of the coaching process by the client, and automated documentation benefits were pointed out (Weber et al., 2021). They further note that when a problem needs to be addressed urgently, e-coaching can provide significant value.

However, in the organizational environment the benefits was proposed to be divided into two categories: benefits for the employees and benefits for the company (Bersin, 2021). Benefits for the employee are identified as on-the-job development, better digital communication and increased emotional intelligence. For the company among operational aspects culture change, performance improvement, engagement and positive behaviour change were named.

In conclusion, this short review shows that even though the benefits of human-to-human coaching is a well-researched area, there is very limited, and sometimes overlapping or even divergent views on the exact benefits of digitizing the coaching process. There is even less focus among contemporary researchers on how to realise or measure the identified values of AI coaching.

2.3.6 Evaluation

Although automation enhances the value proposition of coaching, it does not alter it entirely. Hence, it is observed that the existing coaching outcome and process measures remain applicable even in human-to-AI coaching scenarios. Common coaching measures such as goal attainment and task difficulty (Terblanche et al., 2022), goal-attainment, goal-awareness, honesty, efficacy (Casas et al., 2018), and satisfaction, trust, effectiveness (Weber et al., 2021) continue to be assessed in the context of AI coaching.

Moreover, technological metrics can serve as new data sources. For instance, variables such as the average duration of conversation, target achievement, user satisfaction, and overall benefit were evaluated to appraise chatbot characteristics (Tavanapour & Bittner, 2018). Another example is the study that measures task attentiveness and personality traits from mouse movements (Meidenbauer, Niu, Choe, Stier, & Berman, 2023).

Concurrently, information technology has also established metrics, theories, and concepts for evaluating information systems. Often, when metrics are developed from the information technology perspective, the Unified Theory of Acceptance and Use of Technology (UTAUT) is

employed (Venkatesh, Morris, Davis, & Davis, 2003). The first version of UTAUT contained 4 measures: performance expectancy, effort expectancy, social influence, and facilitating conditions. The UTAUT2 (Arenas-Gaitán, Peral-Peral, & Ramón-Jerónimo, 2015) extended this list with the price/value, hedonic motivation and habit measures. Finally, the UTAUT-3 (Bhatnagr & Rajesh, 2023) added the concepts of perceived risk and personal innovativeness in IT. Due to its high validity, UTAUT-based surveys were utilised in various autonomous coaching related studies (Terblanche & Cilliers, 2020; Terblanche et al., 2023).

Recently, with the release of "The Artificial Intelligence (AI) Coaching Capability Assessment Guidelines" (ICF, 2022c), a new form of evaluation — compliance testing — has emerged. The authors recommend evaluating AI coaching systems through the defined 62 capabilities to make informed decisions. While this document serves as an initial conversation starter, a comprehensive compliance evaluation process can be developed around it in the future.
Furthermore, certain metrics aim to identify weaknesses in technological capabilities. For instance, chatbot dialog management, a relatively weak capability, is assessed through human preference on communication by employing the ACUTE-Eval method (Li et al., 2019) or, alternatively, the "Self-Chat ACUTE-Eval" method (Roller et al., 2020).

In conclusion, no standardized, domain-specific measures for AI coaching have emerged thus far. Therefore, the integration of established coaching measures with standardized information technology evaluation metrics remains a recurring theme in scholarly publications.

2.4 Technologies for Autonomous Coaching

The role of technology stack or architectural elements was historically overlooked in the development and research of coaching automation. This disregard stems from the perception that the chosen cloud platform, programming language, or chatbot framework employed for implementation does not markedly influence the functional suitability characteristics of the resulting solution.

Contrarily, the selection of technology invariably exerts a substantial influence on non-functional aspects. As discussed in the previous section, cost reduction emerges as the principal advantage of AI coaching as recognised by numerous scholars. The architectural complexity, choice of cloud platform, applied technology frameworks, and integration of external data sources can significantly affect both the implementation and operational costs. Accessibility, another pivotal non-functional requirement in the AI coaching realm, is heavily contingent on the frontend technology approach, be it via web application, mobile application, or a hybrid of the two. This section is devoted to a review of technology landscape, particularly its substantial impact on the functional and non-functional characteristics of Large Language Models within the coaching automation domain.

2.4.1 Architecture Review

The predominant method for implementing autonomous coaching systems is facilitated through digital chat platforms, utilizing chatbots. Chatbots represent a thoroughly researched domain within information technology, operating on a simple principle: they receive messages in natural human language, associate them with a knowledge base, and generate responses (Bernardini, Sônego, & Pozzebon, 2018). To perform these functions, most chatbots incorporate at least three key components: a knowledge base, a Natural Language Processing module, and a dialog manager.

For instance, a proposed conceptual model and architecture of the e-coaching system indicates two types of operations: reflective and persuasive (Ochoa & Gutierrez, 2018). The ethical concerns surrounding digital persuasion techniques, particularly autonomy, are well-examined within the realm of information technology (Lembcke, Engelbrecht, Brendel, & Kolbe, 2019). This model underscores the significance of system designers' comprehension of coaching, given that persuasion contradicts the ethical guidelines upheld by many contemporary coaching organizations.

Another example from a related field is a “Counseling Dialog System” is the dialog strategy architecture and Natural Language Processing related utterance generation process (S. Han et al., 2013), which contains a novel component: the emotion detector. This module's

inclusion signifies a design intent not only to engage with the communication's content but also to acknowledge the emotional sentiment, aligning with the core competencies outlined in modern coaching frameworks.

Additionally, the system architecture of a career guidance chatbot exhibits several innovations: voice based input and output processing, working with intent and connection with external data (Ohm & Bhavani, 2019). Working with voice-based communication does not alter the system architecture radically, since most voice based conversational systems use speech-to-text and text-to-speech addons for the initial and final step of the communication but leave the rest of the system architecture work with text-based messages. The conversation deconstructing from utterance to intention is also a typical step in conversational systems, similarly to the reverse way of utterance generation from system intention. However, the integration of external data hints at information sharing, a practice not considered ideal among professional coaches who advocate for a reflective, non-directive coaching approach centred on the client’s thoughts, emotions, and beliefs rather than external information.

In contrast, the system architecture of an automated e-coaching system (Pratiwi et al., 2018) incorporates several coaching process-specific components, such as a self-reported questionnaire manager, goal-template rules, and a PA coaching engine. Although such nomenclature could potentially foster communication among coaching researchers, it may pose challenges for system designers and developers attempting to map these components to the system's functionality due to its deviation from the standard naming conventions of applied information technology.

Conversely, the architecture of most commercial systems, such as evoach (Paulick, 2023), PocketConfidant (PocketConfidant, 2023) or coachvici (Terblanche, 2023) remain undisclosed, precluding any analysis and consequently, hindering the transparency of AI-based systems.

Based on these examples, it can be concluded that the evolution of the system architecture in the AI coaching domain tends to gravitate toward rule-based expert systems. The variety

of examples also indicates a lack of publicly available and debated AI coaching-specific reference architecture, which could serve as a foundation for these individual developments. Ultimately, the advent of Large Language Models has overshadowed these rule-based NLP solutions, rendering this knowledge gap less urgent and relevant. Given that Large Language Model-based systems adopt a fundamentally different architectural approach, the subsequent section delves deeper into the proliferation of LLMs and their relevance within the conversational system domain.

2.4.2 Emergence of Large Language Models

A Large Language Model (LLM) is an artificial intelligence model, often built on transformer architecture, trained on huge amounts of text and uses billions of parameters to learn complex language patterns and generate coherent, contextually relevant responses (Brown et al., 2020).

According to the Artificial Intelligence Index Report 2023 (Clark & Perrault, 2023), out of 35 significant Machine Learning models released in 2022, only three originated from academia, while 32 were products of industry-funded research. The report attributes this stark imbalance to several factors: access to substantial amounts of data, computational capacity, and financial resources to support research. OpenAI is one among many privately funded industry entities that released a new Machine Learning model (Schulman et al., 2023); however, its ChatGPT chatbot, created the most significant ripple both in the industry and academia. ChatGPT is a text generation language model that is trained for natural and formal language provided by OpenAI (OpenAI, 2024b), which is based on the generative pre-trained transformer (GPT) architecture.

With the release of ChatGPT version 3.5, the scientific community is swiftly moving to assess the implications of this ground-breaking technology across all coaching-related domains, including education (Kasneci et al., 2023), health care (Sallam, 2023), human resource development (Budhwar et al., 2023), sport psychology (Szabo, 2023), and management science (Nugroho et al., 2023).

Regardless of the domain or industry, researchers emphasize the following impactful capabilities of Large Language Models: ability to translate across all high resource languages, significantly improved human-computer conversation experience through better understanding and more relevant and coherent responses, access to large amount of information, ability to rewrite, summarize, and analyse textual information.

Coaching research is no exception to this trend: experimentation has begun to integrate LLMs with behavioural change in activity coaching (Vardhan et al., 2023), and the technology's potential is being assessed by leading coaching researchers (Jonathan Passmore & Woodward, 2023). Similarly, commercial AI coaching platform vendors have begun to explore this technology; for example, evoach has introduced new ChatGPT-based products: AI Assistant, AI Coach and AI Learning Reflection (Paulick, 2023).

Concurrently, new architectural patterns are emerging on the technological front, facilitating effective and efficient integration of the new LLMs through APIs within the traditional application development landscape (Bornstein & Radovanovic, 2023).

There appears to be a consensus among researchers and industry stakeholders that conversational systems have proven to be effective and efficient in coaching (Singh et al., 2023), even prior to the widespread adoption of LLMs. The introduction of the LLM capabilities will bring a new era to this domain, provided that the associated risks, ethical concerns, architectural challenges, and technological limitations are addressed adequately.

2.4.3 Limitations of the Large Language Models

Some well-known security concerns associated with Natural Language Processing apply to Large Language Models as well, such as SQL injection (Pedro, Castro, Carreira, & Santos, 2023) or the utilization of LLMs for cyber-attacks (Goldstein et al., 2023). However, within the realm of coaching, there are unique limitations associated with LLMs that necessitate consideration in the design of AI coaching applications.

Firstly, modern LLM implementations aim to maximise the number of tokens utilised for interaction with the models (OpenAI, 2024c). A single token is approximately equivalent to four characters of common English text (OpenAI, 2024b). For instance, in the case of ChatGPT-4 with a maximum token limit of 8192, the size of input data—which includes conversational context, instruction length, and generated response—cannot exceed roughly 32,000 characters. This limitation implies that an AI coaching system must effectively manage dialogue and contextual information, as without adaptive measures, the overall conversation length is severely restricted. Consequently, an out-of-the-box LLM, despite its potential, currently falls short of fulfilling the myriad requirements that a human coach can meet.

Secondly, the training and operation of an LLM demand substantial resources: computational capacity and a large volume of training data (Xue, Fu, Zhou, Zheng, & You, 2023). Given that enhanced accessibility is a primary benefit of AI coaching solutions, cost reduction, considering these resource constraints, can only be achieved if the AI coaching system leverages an external LLM service provider. Dependency on an external vendor carries significant implications concerning data privacy, cost, service availability, maintainability, failure tolerance, and other non-functional requirements that necessitate attention.

Thirdly, although LLMs are trained on a multitude of natural languages, the depth of language capabilities fluctuates in accordance with the volume of data employed for training the models. For instance, Google's Gemini solution supports over 40 languages (Google, 2024b), yet considering that 53.3% of all websites are authored in English (W3Techs, 2023), a substantial imbalance likely exists regarding the capabilities of LLMs across different languages. While LLMs present an opportunity to extend AI coaching to non-English speaking populations, the quality of AI coaching solutions may vary across languages.

Fourthly, despite the enhanced reasoning capabilities of ChatGPT-4, the risk of hallucination remains a concern (Alkaissi & McFarlane, 2023), cautioning against complete and blind reliance on these systems. Although this aspect may not appear severe given that reflective, non-directional coaching is not centred on imparting new information to clients, safeguards must be instituted to mitigate the likelihood of inaccurate responses.

In conclusion, this concise review shows the primary limitations intrinsic to LLMs, underscoring the complex nature of LLM-based AI coaching solutions. Despite an active research community exploring the limitations of LLM-based systems, research on how these limitations impact AI coaching systems, the extent of their impact, and the design decisions necessary to address them is scarce, even as these systems are being deployed in public service.

2.4.4 Ethical Issues

Like in case of the LLM-specific technical limitations, there are numerous well-researched ethical concerns associated with LLM-based applications, which are also pertinent in the context of AI coaching.

On an individual level, privacy concerns dominate the scientific literature, as confidential coaching conversation-related information must be shared with a third-party organisation operating the LLM. Contemporary LLM providers offer varying approaches and levels of privacy to address this major issue: Gemini Apps Privacy Hub offers rights for deletion and some level of control over privacy (Google, 2024a), OpenAI offers an option for opting out of using the client data for training their algorithms (OpenAI, 2024a), and Amazon Bedrock is offering full encryption both in transit and at rest (Amazon, 2024). According to the research community privacy concerns coupled with other LLM-specific security concerns, such as jailbreaking (Deshpande, Murahari, Rajpurohit, Kalyan, & Narasimhan, 2023; Nardo, 2023) and prompt injection (Pedro et al., 2023) should be addressed in the AI coaching systems. The most frequently noted LLM-specific security risks are social engineering, malware creation, AI Package hallucination, propaganda spreading, misinformation, or training data poisoning (Wu, Duan, & Ni, 2023). Depending on the AI coaching system context and exact objectives many of these security risks might not be relevant.

On a broader societal level, the key LLM-specific ethical concerns revolve around copyright, bias, environmental issues, and equality.

Regarding copyright, an ongoing debate exists among researchers, authors, and legal representatives concerning the ownership of texts generated through LLMs based on massive amounts of data harvested from the internet. In the realm of generated images and audio content, traditional methods like watermarking ensure protection of rights; however, similar protection procedures for text generation are still in the experimental stage (Yao, Lou, Ren, & Qin, 2023). Depending on the outcomes of the first lawsuits (BBC, 2023; The Verge, 2023) and the reliance of specific AI coaching systems on text generation capabilities, addressing copyright issues may be necessary.

Unlike copyright issues, fairness and bias-related concerns are unavoidable in any LLM-based system. The spectrum of bias-related concerns ranges from stereotypical and derogatory associations (Basta, Costa-jussà, & Casas, 2019; Kurita, Vyas, Pareek, Black, & Tsvetkov, 2019), and intersectionality to socio-political bias (Tan & Celis, 2019). The fundamental underlying issue is that LLMs are trained on vast amounts of data scraped from the internet, with no scalable, robust method to sanitise the data from these characteristics. Some approaches aim to improve the current situation, such as the application of abuse detection algorithms (Mishra, Yannakoudakis, & Shutova, 2018) or injecting LLMs with helpful, honest, and harmless textual prompts (Askell et al., 2021). However, until the fundamental issue is resolved, workarounds need to be implemented in systems reliant on LLM capabilities.

Another ethical concern relates to the environmental impact associated with the high resource consumption of deep learning-based NLP systems (Strubell, Ganesh, & Mccallum, 2019). Other researchers recommended reporting training time and other parameters to enable comparability of different models, providing access to models for researchers and urging researchers to prioritise high efficiency when selecting algorithms. Along the same line a systematic reporting mechanism is suggested by another research group (Henderson et al., 2020) to make energy and carbon footprint transparent. The 2023 AI Index Report compares the environmental impact of selected Machine Learning models using the Power Usage Effectiveness metric, a typical measure used to evaluate the efficiency of data centres, and concludes that GPT-3 appears to emit the most carbon among all the investigated models (Clark & Perrault, 2023). Resource efficiency is directly linked to environmental impact; hence, it must be prioritised when designing and developing AI coaching systems.

Lastly, concerns regarding equality and fairness are further side-effects of contemporary language model development, because a predominantly English-based LLM development will mainly benefit privileged societies, while the environmental impact will predominantly affect marginalised communities (Bender, Gebru, McMillan-Major, & Shmitchell, 2021). Ensuring that access to AI systems is provided not only to English-speaking societies seems to be one of the natural consequences of this concern.

In summary, this brief literature review on LLM-specific ethical concerns concludes that identifying and addressing these issues at a very early stage in the design of AI systems is essential.

2.4.5 Benefits

Despite the technical limitations and ethical concerns posed by the application of Large Language Models, utilising LLM-based technology offers unique benefits beyond the automation-related advantages discussed in Section 2.3.5. However, there appears to be no research yet within the coaching field that explicitly describes these benefits or how to reconcile the known ethical concerns, technical limitations, and the LLM-generated value.

Various research has revealed that LLM-based systems offer a broad array of capabilities that mimic human behaviour: learning, reasoning, summarising, generating new content, analysing sentiment, and much more (Brown et al., 2020; Bubeck et al., 2023; Nori, King, McKinney, Carignan, & Horvitz, 2023; Park et al., 2023; Radford et al., 2019; Schulman et al., 2023; Z. Wang, Xie, Ding, Feng, & Xia, 2023; West, 2023).

Some of these capabilities are central to coaching, such as working with emotions based on sentiment analysis or summarising, which is recognised as a core coaching competence by most contemporary coach core competence frameworks. Hence, LLM systems seem to enhance the core coaching-related capabilities of AI coaching systems.

Additionally, other researchers argued that the three key limitations of pre-LLM chatbots are the limited context awareness, limited scale and limited text generation ability (Koubaa, Boulila, Ghouti, Alzahem, & Latif, 2023). Along the same line, others named incorporating context, coherent personality, evaluation of models, intention and diversity as the key challenges of early-stage chatbots (Britz, 2016). These limitations of early version chatbots were significantly enhanced with the advent of GPT-based algorithms, suggesting that the overall dialogue quality may be meaningfully improved when deploying these technologies.

Furthermore, employing LLM-based technology can contribute to addressing the current prevalent English language bias, as most AI coaching systems were developed in English only. Although the current development of LLM systems is far from being language-agnostic (Joshi, Santy, Budhiraja, Bali, & Choudhury, 2020), they appear to be proficient enough to create the first AI coaching systems operating in at least the top 30 high-resource languages. This advancement could significantly enhance the accessibility of AI coaching systems, foster diversity, and enable the large-scale research of cultural aspects of coaching.

If only these three key benefits – coaching core competence improvement, dialogue quality enhancement, and multi-language support – are leveraged from the vast capabilities of LLMs, they could propel the current AI coaching systems to a new level.

2.4.6 Evaluation and Testing

Since Large Language Models represent a new generation of Natural Language Processing technologies, the pre-LLM age evaluation methods and metrics discussed in Section 2.3.6 remain applicable to this emerging technology. However, additional metrics, closely linked to how LLMs generate content, warrant consideration.

Toxicity, a well-established concept in NLP, has garnered significant attention from researchers due to the tendency of pre-trained language models to generate toxic content, even from seemingly innocent prompts, as compared to earlier NLP solutions. For instance, a spiralised group of researchers proposed a 100k test data set to validate toxicity of language models (Gehman, Gururangan, Sap, Choi, & Smith, 2020).

Similarly, other researchers have provided test sets for social bias (Nangia, Vania, Bhalerao, & Bowman, 2020) and even new measurement related concepts are developed like “regard” (Sheng, Chang, Natarajan, & Peng, 2019) to advance the progress of building less biased language models. Despite these advancements, a consensus within the research community regarding several fundamental bias-related concepts remains elusive. Some argue that certain characteristics are culture-dependent while others are universal; hence, the pursuit of universal "standards" is perceived as overly ambitious (Fiske, 2017; Sczesny, Bosak, Neff, & Schyns, 2004). A critical meta-analysis within the NLP bias domain further underscored this message and suggested enhancements to the quality of research in the domain, emphasising the need for clear articulation on how bias is conceptualised, which behaviours are deemed harmful, to whom, and why (Blodgett, Barocas, Daumé, & Wallach, 2020).

In the context of LLM-driven systems, such as AI coaching, clear articulations are imperative, particularly as these measures are typically employed by automated auditing and diagnostic systems. Nonetheless, constructing such systems necessitates precise algorithmic definitions, a feat unlikely to be realised soon without addressing the ambiguity within the bias research domain. Given that LLMs are facilitating access for non-English speaking populations, these systems invariably interface with a diverse array of cultures, necessitating decisions in implementing "universal" bias and fairness checking algorithms.

In conclusion, the quality of content generated by Large Language Models hinges on three primary factors: the characteristics of the training data, such as size and quality; the instruction; and the LLM model itself. Upon fine-tuning these major aspects, the model's output can be further refined through additional diagnostics, filtering, or content regenerating mechanisms. This necessitates a profound comprehension of which specific bias, fairness, and diagnostic measures are associated with coaching conversations and dependent on the application's context. Consequently, a specifically designed, easily adaptable architectural composition is warranted to address these continuously and rapidly evolving concerns.

2.5 Chapter Summary

This section provides a summary that highlights the identified research gaps associated with the LLM-driven AI coaching domain. It outlines the originality of the thesis and the significance of this research project.

2.5.1 Research Gap

It was concluded a decade ago that "coaching is generally an under-explored and under-researched HRD-related practice" (Ellinger & Kim, 2014) and the status quo has not significantly improved since. The research highlights several gaps in the field of coaching, suggesting that despite its success and growth, coaching remains under-researched with several issues unaddressed. The coaching community has not effectively changed perceptions about the profession's credibility and scientific basis. Here are the key points regarding the coaching industry:

1. Definition and Boundaries: There is no unified, universally accepted definition of coaching, and it often overlaps with other modalities.
2. Ethical Standards: The development of ethical codes remains opaque.
3. Scientific Research: There is a significant lack of evidence-based research concerning core coaching competencies.
4. Emerging Trends: New trends like team coaching do not address fundamental issues in the field.
5. Ethical Concerns: Ethical issues, such as the risk of burnout and negative side-effects of coaching, remain unaddressed.

Furthermore, the following structural issues in coaching research contribute to these gaps:

1. Evidence-Based Culture: There is a struggle to foster a culture of evidence-based research in coaching.
2. Research Characteristics Missing: Important aspects like transferability studies and outcome-based evidence are lacking.

3. Standardization and Rigor: There is a scarcity of standardized measures and rigorous research methods.
4. Link to Value Generation: The connection between coaching outcomes and value generation is not well understood.

Similarly, the key gaps in the literature regarding autonomous coaching are summarised here:

1. Inconsistent Terminology: Lack of consistent terminology for automation levels in coaching (e.g., "online-coaching," "AI coaching," "technology-assisted coaching") creates conceptual confusion.
2. Focus on Non-Core Coaching Tasks: Existing automated solutions primarily focus on automating administrative and non-core coaching tasks, with limited development in augmenting core coaching competencies or fully autonomous coaching systems.
3. Limited Coach-Centric Research: Existing studies predominantly address client outcomes while neglecting the coach perspective, especially concerning augmentation-oriented tools.
4. Lack of Theoretical Models for Autonomous Systems: Absence of a robust, abstract, and theoretical model to adapt autonomous coaching solutions across domains.
5. Evaluation Metrics Gap: No standardized, domain-specific metrics for evaluating AI coaching systems. Integration of coaching measures with established IT metrics remains underexplored.
6. Ethical Considerations: Limited practical frameworks for integrating ethical principles (e.g., transparency, consent, bias mitigation) into the design of autonomous coaching systems.
7. Design Framework Immaturity: Existing frameworks lack completeness and adherence to contemporary requirements engineering practices.
8. Adoption Barriers: Limited research on factors influencing user adoption of autonomous coaching applications (e.g., performance expectancy, effort expectancy, anthropomorphism).
9. Technological and Methodological Challenges: Insufficient progress in achieving human-like conversational interaction in autonomous coaching systems.

10. Benefits Realization and Measurement: Lack of focused research on how to measure or realize identified benefits of AI coaching (e.g., scalability, cost reduction, client engagement).
11. Ethical and Practical Issues in Data Management: Concerns about data privacy, algorithmic transparency, and explainability in AI coaching systems remain under-addressed.
12. Cross-Domain Applicability: Existing solutions are often narrowly targeted (e.g., mental health, habit management) with limited adaptability for broader or novel coaching domains.

Lastly, the gaps associated with the LLM technology and AI coaching specifically:

1. Underrepresentation of Technology Stack Influence: Lack of exploration of the impact of specific technology stack or architectural elements on the functional and non-functional characteristics of autonomous coaching systems.
2. Absence of AI Coaching-Specific Reference Architectures: No publicly available or standardized AI coaching-specific reference architectures to guide development, especially considering Large Language Models (LLMs) replacing rule-based NLP systems.
3. Integration of Ethical Guidelines in System Design: Insufficient research on integrating ethical considerations, such as autonomy and non-directive coaching, into the system design of autonomous coaching solutions.
4. Impact of LLM Limitations on AI Coaching Systems: Limited research addressing how LLM-specific limitations (e.g., token constraints, dependency on external vendors, language capability disparities, hallucination risks) affect AI coaching applications.
5. Data Privacy and Security Concerns: Inadequate studies on the implications of privacy and security risks (e.g., third-party vendor reliance, data encryption) within the context of AI coaching systems.
6. Bias and Fairness Challenges: Lack of consensus on bias conceptualization and measurement standards, especially for multilingual and culturally diverse AI coaching environments.

7. Environmental Impact of LLMs: Limited focus on the environmental impact of LLMs (e.g., energy consumption, carbon footprint) and strategies to mitigate it in AI coaching solutions.
8. Effectiveness of AI Coaching Across Languages: Sparse research on the variability in coaching quality due to the uneven language proficiency of LLMs across high-resource and low-resource languages.
9. Adaptation to Cultural Contexts: Limited exploration of how LLM-based AI coaching can address diverse cultural norms and expectations within coaching conversations.
10. Ethical Implications of Content Generation: Minimal research on reconciling ethical concerns (e.g., copyright, misinformation, bias) with the benefits of LLMs in autonomous coaching systems.
11. Evaluation Metrics for LLM-Based Systems: Lack of domain-specific evaluation metrics tailored to LLM-driven coaching systems, particularly regarding toxicity, fairness, and bias in conversational contexts.
12. The Role of LLMs in Enhancing Core Coaching Competencies: Absence of explicit studies evaluating how LLM capabilities improve core coaching competencies, such as emotion recognition, summarization, and reflective questioning.
13. Scalability and Accessibility in Multilingual Settings: Limited research on the scalability of AI coaching systems in multilingual and multicultural environments to ensure accessibility and inclusivity.
14. Adaptability of Architectural Designs: Need for adaptable architectural designs that address the rapidly evolving limitations, capabilities, and ethical concerns associated with LLM-based AI systems.

The compounding effect of these research gaps on all three levels - coaching, automation, LLM - is accumulating, hence it is vital to invest resources in researching the LLM-based autonomous coaching systems.

2.5.2 Originality of the Thesis

The aim of this thesis is to design and develop a theoretical framework for the development of LLM-based autonomous coaching systems that encompass all functions linked to core coaching competences for the advantage of clients, coaching researchers, autonomous coaching system developers, and the broader coaching community.

The thesis makes major contributions to the field of autonomous coaching systems, highlighting innovative approaches and transformative implications:

1. Exploring a Competence-Based Autonomous Coaching Framework: The thesis introduces the first framework explicitly designed around core coaching competencies for autonomous coaching systems. This novel approach ensures alignment with established coaching practices, providing a robust foundation for both research and practical applications.
2. Enabling Context-Independent Applications: By abstracting the framework from specific coaching goals, this research ensures adaptability and applicability across diverse coaching scenarios. This flexibility allows the framework to be utilised in various industries and contexts, enhancing its potential impact.
3. Innovating Through Framework-Based Prototyping: The developed framework-based prototype demonstrates the feasibility of evidence-based, large-scale insight generation. This prototype serves as a proof of concept for using advanced technologies like Large Language Models (LLMs) in scalable, autonomous coaching applications.
4. Ensuring Framework Traceability and Adaptability: The framework's structured design emphasizes clear linkages among components, facilitating future modifications and improvements. This traceability ensures that the framework can evolve alongside advancements in technology and coaching research.
5. Advancing Research in LLM-Based Coaching Systems: Investigating the use of LLM technology within autonomous coaching marks a relatively unexplored research area. This thesis not only bridges this gap but also establishes a foundation for future advancements in coaching-related technology, ensuring that AI is leveraged responsibly and effectively.

6. Addressing Global Challenges Through Scalable Coaching Solutions: Through automation and scalability, the framework offers a pathway to democratize coaching. This contribution is particularly essential in addressing pressing global challenges, enabling individuals across diverse socio-economic backgrounds to access the transformative benefits of coaching.

3. Research Design

3.1 Introduction

This chapter encompasses the pivotal considerations in designing the research project that influence the processes of framework building, optimization, and validation. The research methodology is structured in two distinct stages to achieve the research outcomes and validate the four hypotheses:

1. COBAC Framework Construction: Predicated on document-centric techniques, such as textual analysis and the employment of requirements engineering-related analytical methods, the framework is derived from the existing coaching core competences and the available scientific research.

2. Framework Validation with Service Prototype: The proposed COBAC framework serves as the foundation to develop a functional software prototype and to validate the framework, adopting the Design Science Research Methodology (DSRM) (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2008).

3.2 Research Philosophy

Technology research is traditionally conducted from a scientific epistemological position utilising experiments to support the argumentation, hence being perceived as an empirical method of knowledge generation. This contrasts markedly with the coaching research field, which embraces a plethora of approaches borrowed from psychotherapy (Bachkirova et al., 2016). The research paradigm of this paper aligns with the ontological assertions of the prevalent positivistic approach, which not only predominates in the technology domain but is also recurrently employed in coaching process research.

3.3 Research Hypotheses

Pursuant to the research outcome, the following hypotheses are considered for the thesis to ascertain the value creation potential of the newly designed and developed COBAC framework:

1. Hypothesis 1 (H1) - Feasibility: Developing an autonomous coaching system based on the COBAC framework is feasible.
2. Hypothesis 2 (H2) - Effectiveness: The COBAC-based autonomous coaching system is effective.
3. Hypothesis 3 (H3) - Dialog Quality: The dialog quality is independent of the high-resource language used for the conversation.
4. Hypothesis 4 (H4) - Language Influence on Adoption Factors: The language of the coaching conversation does not influence the factors contributing to the intention to use the coaching chatbot.

The four hypotheses can be categorized in two distinct segments. H1 ascertains that the newly devised framework transcends an abstract theoretical framework, aligning with the nomenclature, structure, and content pertinent to the information technology domain, thereby being efficacious for the development of new autonomous coaching systems. This hypothesis is implicitly tested through the prototype construction.

Key factors, such as effectiveness, language influence on adoption factors and dialog quality, are identified in Chapter 2 to largely distinguish the value proposition of the LLM-based system compared with all other existing frameworks and systems. These facets are targeted by H2, H3, and H4 respectively.

3.4 Research Methodology

3.4.1 Framework Building Approach

The methodology for constructing the initial theoretical framework relies on a documentation-centric, textual analysis-based approach. The documentation-oriented approach was proposed for creating the initial framework for several reasons:

1. Developing a well-designed research project to derive the core coaching competences based on primary sources and traditional, manual research approach poses a Herculean task, with resource requirements far exceeding those typical of a PhD thesis.
2. With the introduction of automation, an autonomous coaching system can collect feedback from numerous users across thousands of conversations without substantial resource demands. Core competences can be synthesised using statistical methods based on this user feedback and the analysis of digital conversation transcripts. Hence, achieving a perfect initial starting point is not as critical from the perspective of the end-result, given this process of empirical investigation with multiple iterations based on data collected through automated systems.
3. Thus, three decades of explicit knowledge about coaching, albeit largely based on expert opinions, knowledge transferred from other domains, and secondary sources, serves as a sufficient foundation for initiating evidence-based exploration with automation.

Most of the analytical procedures and notations applied in the framework construction phase are deemed industry standards in the domain of requirements engineering. Approaches from the requirements engineering field are employed as they offer a spectrum of potential methods, practices, and research-based conceptual frameworks for theorising the COBAC, ensuring that the constructed framework is apt for subsequent system development.

In the realm of requirements engineering, models exhibit three crucial properties (Frühauf et al., 2017): pragmatic, representation and reduction.

1. Pragmatic: The COBAC is designed and developed primarily to advance the theoretical knowledge base of the research domain, and to ensure its pragmatic suitability for practical development purposes.
2. Representational: The proposed COBAC framework mirrors reality by offering conceptual elements of an autonomous coaching service that aids users in effectively achieving their goals, grounded in the current realities of the human-to-human coaching process.

3. Reductionist: The framework distils the represented reality not only to the boundary of the coaching field, but also in terms of content by spotlighting only the most critical components. While a commercial autonomous coaching system might have thousands of individual objectives, constraints, and requirements, the framework intends to cover only those items representing the core, defining, and common elements of coaching, irrespective of its context, specific application area, or coaching school.

3.4.2 Framework Components

IREB defines requirements engineering as "The systematic and disciplined approach to the specification and management of requirements with the goal of understanding the stakeholders' desires and needs and minimizing the risk of delivering a system that does not meet these desires and needs" (Glinz, 2014). This definition of requirements engineering suggests that the role of a requirements engineer is to collaborate with stakeholders to derive system requirements from their desires and needs. An alternative approach is outlined by the authors of value proposition design (Osterwalder, Pigneur, Bernarda, & Smith, 2014), a highly influential requirements elicitation practice within the digital product management community. This model advocates focusing on customers' pains, gains, and jobs rather than desires and needs, and centring on gain creators, pain relievers, and accordingly deriving requirements for products and services. This method aptly reflects the paradigm shift experienced over the last two decades by the agile transition within the software engineering community from a system-oriented requirements gathering practice towards explicit value creation. It also aligns with the design science research model methodology employed in the second stage of this thesis, which succinctly encapsulated: "whereas natural science tries to understand reality, design science attempts to create things that serve human purposes" (March & Smith, 1995).

A requirements engineering model (Loenhoud, Lauenroth, & Steiger, 2017) was devised as a conversation starter within the requirements engineering community to amalgamate agile value creation approaches with traditional requirements engineering practices. This model introduces a triad of problems, goals, and solutions, and explicates the relationships among

them. It provided the inspiration to craft a slightly modified model version that aligns well with the structure and terminology of the DSRM methodology. Figure 3.1 presents a visual representation of the COBAC components and the relationships among them, while Table 3.1 details the individual components utilised in the COBAC framework.

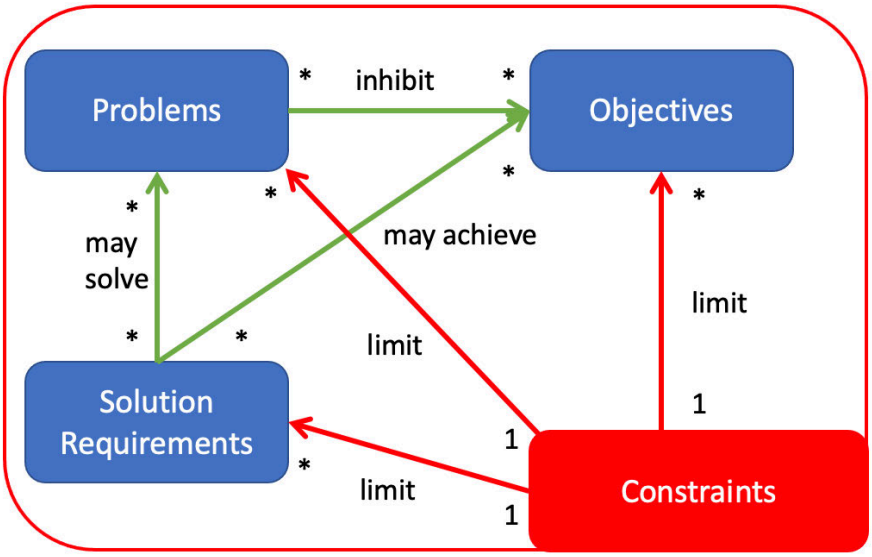


Figure 3.1: COBAC Components and Relationships

Table 3.1: COBAC Component Overview

Component Name	**Component Definition, Description**
Problems	A problem is a negatively experienced situation or state in the stakeholder context (Loenhoud et al., 2017).
Objectives	Objectives provide the basis for measuring whether progress is being made towards solving the problem (Object Management Group (OMG), 2015).
Solution Requirements	A requirement is "(1) A condition or capability needed by a user to solve a problem or achieve an objective. (2) A condition or capability that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed documents" (IREB, 2020).
Constraints	An extended definition applied based on the IREB definition (IREB, 2020): a constraint limits the objective space, the problem space and also the solution space beyond what is needed or desired by the stakeholders.

Problems and objectives emanate from the stakeholders' context, and solution requirements are derived based on them. One or more problems can obstruct one or more objectives, and an objective might enable another objective. One or more requirements may achieve objectives and may solve problems. Constraints limit one or more elements of this construct, yet they also demarcate the boundary of the problem, objective, and solution spaces.

Augmenting the model with constraints is necessary because of three reasons (Loenhoud et al., 2017):

1. Constraints aid in maintaining focus on the main problems and objectives, which is crucial in a high-complexity domain such as human interactions or coaching.
2. Constraints render underlying assumptions and limitations transparent and explicit.
3. Making the context boundary and system boundary explicit with a mental concept is imperative for delivering a theoretical model with high modifiability and extensibility.

The relationships between these mental constructs provide an effective means to easily identify the specific items that belong together, facilitating their removal from the framework or replacement with different items, thereby enhancing the framework's modifiability. For instance, should there arise a need for a new automated coaching framework that offers a virtual reality version of the coach, - a likely future scenario considering current advancements in the metaverse and generative AI fields - only the meta-communication and communication-related problems, objectives, solution requirements, and constraints need to be amended, while the rest of the framework remains intact. This example illustrates how component relationships contribute to high modifiability by enabling traceability.

3.4.3 Quality Considerations

IREB recommends five quality criteria for requirement documents: completeness, unambiguity and consistency, clear structure, modifiability and extensibility, and traceability (Frühauf et al., 2017). IREB did not define these aspects as measurable quantities, hence they were considered only as a guidance through the model building process without adherence test or compliance checks.

The quality criteria for the framework were summarised in Table 3.2 based on the IREB recommendations, along with specific descriptions explaining how the framework approached each criterion.

Additionally, the International Requirements Engineering Board (IREB) advocates two primary formal rules for enhancing the readability of requirements articulated in natural languages:

employing concise sentences and encapsulating a singular requirement or thought within each sentence. These principles are applied consistently throughout the framework construction, encompassing all framework components, not solely the requirements. The elements represent high-level concepts, necessitating a more detailed specification for each component prior to the commencement of the actual implementation.

Table 3.2: COBAC Quality Criteria

Quality Criterion	Description
Completeness	Completeness, in the sense of identifying all critical and common attributes for a viable autonomous coaching service, underpins the theoretical work.
Unambiguity and consistency	The framework describes the basic concepts without ambiguity by using the domain specific terms and concepts. The elements of the framework are consistently fitting together.
Clear structure	The structure of the framework was detailed exhaustively in the preceding section.
Modifiability and extensibility	The extent to which the framework can be effectively and efficiently tailored for varying or evolving contexts is an essential characteristic, given the rapidly changing nature of the field.
Traceability	Requirements can be traced back to their origins through applied keys and references.

The initial framework building stage utilises a literature review as the principal research methodology, alongside various modelling techniques, to conceptualise the COBAC elements in alignment with the predefined quality criteria.

3.4.4 Research Design

To elucidate the COBAC framework construction process, Figure 3.2 provides a synopsis of the employed methods, modelling techniques, and their outcomes.

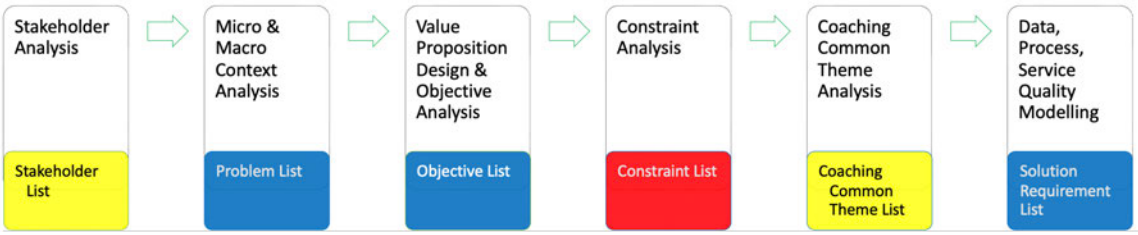


Figure 3.2: Initial Framework Building Stage Overview

The purpose of the illustrated analytical strategy is to describe the COBAC framework components in list format: Problem List, Objective List, Constraint List, and Solution Requirement List. Besides the framework elements, coloured blue and red in accordance with the colour scheme detailed in Section 3.4.2, two intermediary outputs are designated and highlighted in yellow: Stakeholder List and Coaching Common Theme List. The literature review, a systematic approach for aggregating and synthesising prior research, serves as an effective means to explore issues within the research domain (Snyder, 2019), thereby being the chosen research methodology throughout this phase.

The stakeholder analysis leverages literature as a source to identify key stakeholders and their characteristics in relation to autonomous coaching. The stakeholder analysis has three stages: the identification of key stakeholders, stakeholder mapping from human-to-human coaching to the autonomous coaching context, and the detailed analysis of the identified stakeholder. Conducting a stakeholder analysis is essential as a precursor to problem analysis, since problems are non-existent without stakeholders, as defined in Section 3.4.2. The outcome of the stakeholder analysis is the creation of the Stakeholder List, which informs the subsequent analytical step, micro-context analysis.

The micro and macro context analyses, predicated on the literature review and the Stakeholder List, aim to pinpoint the perceived problems and issues relevant to the research topic. Identification of problems are important, because problem framing is often considered a prelude for defining value of software products (Wieringa, Gordijn, & Eck, 2004). Regarding the context analysis the term ‘context’ within the requirement engineering domain is defined as “network of thoughts and meanings needed for understanding phenomena or utterances” (IREB, 2020). Just like the previous and subsequent steps, the contextual analysis focus on the essential aspects without aspiring to a comprehensive analysis of the context. The result of this phase is the documentation of identified issues related to autonomous coaching in the form of a Problem List artefact. Each identified problem is allocated an identifier for traceability purposes, adhering to the naming convention ‘P.<n>’, where ‘P’ denotes problem and ‘n’ represents an incremental number. Analogous naming conventions are implemented for other framework components. Numerous models in the business management domain, such as Porter’s Five Forces, SWOT analysis, or the STEEPLE model - an extended variant of

the traditional PEST model (Aguilar, 1967) - offer systematic approaches to analysing organisational contexts. The thematic macro context analysis employs the STEEPLE model, wherein the acronym represents the following contextual factors: social, technological, economic, environmental, political, legal, and ethical. After the macro context analysis, the micro context analysis, largely based on Grant's Multiple-Perspective Model (Grant, 2017), concludes the contextual investigation to collect the associated problems of stakeholders in autonomous coaching into a Problem List.

The value proposition design and objective analysis step is based on the Problem List, and where textual analysis falls short, the literature is also used to compile a consistent list of objectives with associated measures. The outcomes of the value proposition design and objective analysis, the list of objectives, are extracted from the value map. Analogous to the Problem List, a key is assigned to each objective with the following naming convention: O.<n>, where 'n' denotes a continuously growing sequence.

Constraints identified during the previous steps are consolidated in a Constraint List and further constraints are defined to create tacit knowledge about the limitations of the COBAC framework to limit the system boundary, the problem, and the objective space. Each constraint is assigned a key with the following naming convention: C.<n>, where 'n' denotes a continuously growing sequence.

The coaching common theme analysis aims to compile a list of behaviours and characteristics that form a detailed understanding of the core elements of coaching. These elements are recognised by both professional ICF and EMCC coaches as defining characteristics of the coaching process.

The final step of the analysis aims to create the solution requirements. Since the advent of the software industry, the two principal conceptual building blocks of software were encapsulated by the equation: "algorithms plus data structures equals programs" (Wirth, 1976). In alignment with this paradigm, the design phase encompasses data structure analysis to ascertain which data entities are indispensable for supporting the algorithms and a process analysis to identify the algorithms that underpin the autonomous coaching system. A

conceptual data model serves as the blueprint of the actual data structure implementation, which might be defined by lower-level data models, such as logical or physical data model. A hallmark of the COBAC framework is its conceptual essence, thus lower-level data models are not conceived at this step. The guiding principle for the data modelling is to define the minimum set of data structures that are required and not to deliver a specific solution. The content of the conceptual data model is crafted based on the Coaching Common Themes and the initial objectives. The conceptual data model is visualised using an industry-standard diagram type, the UML Class Diagram (Object Management Group (OMG), 2017). After the data modelling, business process modelling techniques are used to define the relevant algorithms involved in conducting coaching and they are visualised using the Business Process Model and Notation (BPMN) version 2 (Object Management Group (OMG), 2011) . In this context an algorithm is defined as "a list of instructions for solving a problem" (Cambridge Business English Dictionary, 2020). Literature review is used to provide feasibility context from the following fields: Natural Language Processing, Human Computer Interaction (HCI), and Machine Learning. Finally, solution requirements address the query of how a particular objective can be attained. Solution requirements, which formulate the last remaining component of the COBAC framework, are elicited systematically based on the conceptual data model, the coaching process model, and the coaching common themes, employing the terminology and taxonomy proposed in the Product Quality Model of the ISO 25010:2011 (ISO/IEC, 2011) standard. Solution requirements are intertwined with problems, common coaching themes, and objectives through their references. Each solution requirement is assigned a key with the following naming convention: SR.<n>, where 'n' represents a continuously growing sequence.

3.4.5 Framework Validation Approach

The second stage of the thesis adopts the six-step Design Science Research Methodology (DSRM) (Peffers et al., 2008), as depicted in Figure 3.3.

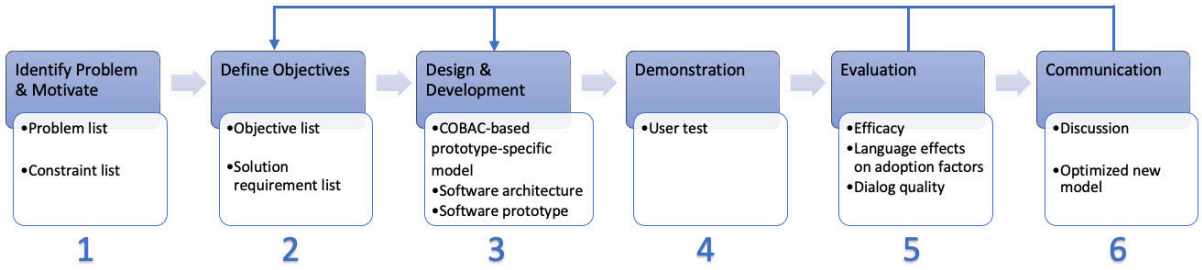


Figure 3.3: Design Science Research Methodology (DSRM)

This methodology was chosen for the thesis because the framework should serve as a foundation for developing autonomous coaching systems. Simultaneously, the four hypotheses cannot be validated without a working software.

Based on the COBAC framework synthesized in the first stage of the thesis, a prototype-specific framework is designed and developed encompassing all framework elements: Problem List, Constraint List, Objective List, and Solution Requirement List. The prototype is at the proof-of-concept level; hence it encapsulates all essential aspects of COBAC that is required for hypothesis validation, although there are disparities in invested resources for development and operations, level of detail, robustness, and maturity. As part of the design activities, a prototype-specific architecture is created that is most suitable for the purpose of the proof-of-concept, and key design decisions are documented in the thesis.

The development is undertaken by the author of the thesis, and it relies on technology choices deemed most effective for utilising the developer's resources, skills, knowledge, and experience.

3.4.6 Methods for Data Collection and Analysis

Upon conclusion of the development, the autonomous coaching system was made available online for research participants to initiate a coaching session, averaging around 10 minutes.

Participants were recruited mainly via Amazon's MTurk and the Clickwork platforms, due to their proven ability to yield inexpensive yet high-quality data (Buhrmester, Kwang, & Gosling,

2011), facilitating data collection in multiple high-resource languages. Participant selection criteria were defined as acceptance rate above 98% to improve data quality. Moreover, eligibility was contingent upon language and location: English with location in the USA, German with location in Germany, and Spanish in Spain. Language criteria intended to address the language-related hypotheses (H3, H4), and location was intended to enhance data quality. Unlike human coaching, where regular client awareness studies are conducted, there is no evidence of any study addressing whether the selected participants accurately represent users of AI coaching applications. As a result, no additional selection criteria could be established to further refine participant selection. Comparable studies boasted sample sizes of 226 (Terblanche & Cilliers, 2020; Terblanche & Kidd, 2022), and 105 and 134 in study 1 and 2 respectively (Terblanche et al., 2022). Given the practical limitations posed by participant recruitment cost, this study was aimed to 120 participants, equally distributed among the three language groups.

A cross-sectional survey was designed for completion by participants following the chatbot interaction. All measured variables employed a 7-point Likert scale to collect quantitative data. Alongside the conversation language related data gathering, questions from the Coaching Effectiveness (De Haan et al., 2016) questionnaire, questions from the standardized UTAUT survey with 5 measures (Venkatesh et al., 2003), and additional 5 questions concerning the dialogue quality derived from the USR model (Mehri & Eskenazi, 2020) were incorporated to address hypotheses H2, H3, and H4, respectively. Furthermore, additional supplementary questions were appended to the questionnaire to mitigate the risk of inattention, a significant challenge in crowdsource-based research (Aguinis, Villamor, & Ramani, 2021). Data garnered from these questions were solely utilised to filter out invalid responses and were not employed in hypothesis testing. Adhering to other research recommendations, a pilot test with ten participants was conducted prior to the full study to minimise the risks of technical and human errors.

The Coaching Effectiveness (De Haan et al., 2016) was selected as a basis of the questions for H2 testing because of its simplicity and the questionnaire covers not only outcomes but also process measures. However, only three questions from the 4-item questionnaire were utilised, as question #3 refers to a coaching outcome - altered behaviour - between the

coaching session and survey completion, was deemed irrelevant for this study and consequently omitted. The reason of omission was related to the survey filling context: the survey is filled right after the coaching session, hence there is no time allowed for behaviour change.

A tailored version of the UTAUT was selected for the research because it is considered a wide-spread approach for validating technology acceptance and use (Dwivedi, Rana, Jeyaraj, Clement, & Williams, 2019) and because it was already used with success in various autonomous coaching related studies (Terblanche & Cilliers, 2020; Terblanche et al., 2023). The first version of UTAUT survey with 4 measures (Venkatesh et al., 2003) was selected over more recent iterations, namely UTAUT2 (Arenas-Gaitán et al., 2015) and UTAUT-3 (Bhatnagr & Rajesh, 2023), as they would have unnecessarily extended the survey duration without contributing significant value in this specific AI coaching context. However, the Perceived Risk (PR) measure is added from UTAUT-3 based on the successful application and proven relevance of this construct in the AI coaching context (Terblanche & Cilliers, 2020).

Table 3.3: Applied Measures

Measure	Description	Source
Coaching Effectiveness (CE)	A combined measure of 3 aspects out of the originally defined 4: reflective space, insights, success of coaching journey. The 3 measured data points are averaged to compile a single value.	Coaching Effectiveness (De Haan et al., 2016)
Performance expectancy (PE)	The degree to which an individual believes that using the system will help to attain gains.	UTAUT (Venkatesh et al., 2003)
Effort expectancy (EE)	The degree of ease associated with the use of the system.	UTAUT (Venkatesh et al., 2003)
Social influence (SI)	The degree to which an individual perceives that important others believe he or she should use the new system.	UTAUT (Venkatesh et al., 2003)
Facilitating conditions (FC)	The degree to which an individual believes that an organizational and technical infrastructure exists to support use of the system.	UTAUT (Venkatesh et al., 2003)
Behavioral intention (BI)	A measure of an individual's perceived likelihood or relative strength of intention to perform a particular behaviour.	UTAUT (Venkatesh et al., 2003)
Perceived Risk (PR)	The degree to which the user of chatbot coaching believes that s/he may be exposed to certain types of risks. Includes breaches of confidentiality, trust issues, and perceptions that AI is a threat towards humans.	UTAUT extension in AI context (Terblanche & Cilliers, 2020)
Language (LG)	Conversation language of the coaching session, which is designed for the study as English, Spanish and German.	MTurk input parameter (Premium Qualifications)
Dialog Quality (USR)	The degree, to which a dialog is considered understandable, natural, maintaining context and interesting.	USR (Mehri & Eskenazi, 2020)

To summarize, Table 3.3 provides an overview of the name, description, and source of the applied measures.

The system generated also operational costs, which could be employed to estimate average conversation costs and the magnitude of environmental impact. However, such analysis transcends the scope of this thesis.

Following the data collection, a data validation step was undertaken to ensure the inclusion of valid data in subsequent analyses. Data sets were excluded in the pilot phase if any of the following criteria were met:

1. Attrition: Incomplete data set due to attrition.
2. Attention Check: Incorrectly answered attention testing questions.

No specific data was collected for H1 as it was implicitly validated through prototype demonstration.

H2 validation necessitated the statistical averaging of all provided responses to the three Coaching Effectiveness questions to ascertain the coaching effectiveness for each conversation. The same approach is used for H3 validation based on the measures related to the dialogue quality.

In alignment with similar studies, partial least square structural equation modelling (PLS-SEM) was the selected analytical approach to validate the effect of language on the intention to use the coaching chatbot for H4 validation.

The results of the data analysis, coupled with the documentation of framework building and prototype development, were integrated into an updated list of objectives, constraints, and solution requirements to present the optimised framework.

3.4.7 Assumptions and Limitations

There are several assumptions and limitations associated with the described approach.

The initial framework building assumes that phenomena observed in human-to-human coaching relationships may be similarly useful and significant for an autonomous coaching service. As discussed previously, it is reasonable to assume transferability at this stage, as it is only the initial framework. Measured data is anticipated to yield evidence-based insights that can contribute to an optimised framework.

The quality of the literature appears to be another limiting factor. A considerable portion of the research data is provided by institutes with vested interests in portraying coaching positively, such as the International Coaching Federation (ICF) or the European Mentoring & Coaching Council (EMCC). A second, less biased source of information emerges from major management consulting firms like PwC, McKinsey, or Deloitte, as they are vested in 'selling future change', regardless of the specific area. A third category of research source is led by organisations with no direct commercial interests, such as the World Economic Forum (WEF) or the Institute of Employment Studies (IES). All three categories are accepted as sources of information for the purpose of framework building as, while the results may vary in scale, the research findings are generally well aligned in terms of direction.

The complexity entailed in the competency framework comparison is highlighted in a report compiled based on the comparison of previous versions of the referenced competency documents (Lindvall et al., 2013). The team conducting the analysis on the previous versions of the competency documents, devoid of the ethical codes, concluded that 84% of ICF's demonstrative behaviours can be found in the EMCC, and the rate is 60% in a reverse comparison. An additional challenge arises from the terminology, as the two organisations utilise different terms even for fundamental definitions such as evaluation or client. Given these complications, the applied research method is not suitable for understanding the complex relationships of the coaching competencies. Nor is it applicable for identifying objectives that are mutually exclusive and collectively exhaustive (C. Y. Lee & Chen, 2018), an approach that is typical for gathering objectives in the field of requirement engineering.

Nonetheless, it serves to define all essential elements of coaching, a notion few would challenge regardless of affiliation with a specific professional institute, scientific community, or coaching school.

There are various challenges associated with the crowdsource platform based data collection approach (Aguinis et al., 2021): inattention, self-misrepresentation, self-selection bias, attrition, English language fluency, non-naivete, participant interactions through communities, web robots, social desirability bias and perceived researcher unfairness. These challenges can impact multiple factors such as internal validity, external validity, construct validity, and statistical conclusion validity. The research design was formulated with these challenges in mind; however, it is inevitable that some of these aspects influenced the data collection stage.

Another limitation is associated with the development effort. One of the drawbacks of automation is that the initial resources, such as skills, cost, and time, can be significant to deliver the automated system. The focus of the prototyping was to deliver a version that fulfils the predetermined set of constraints, objectives, and solution requirements; it was not intended to build a fully functioning, commercially ready product, which might have impacted efficacy measures. After the initial framework creation, all subsequent iterative improvements of the framework are recommended to be based on empirical data gathering. Hence, after this initial validation step, the document-centric approach to derive framework elements is losing its relevance. It is assumed that the COBAC framework can be further refined with additional iterations using evolving software; however, the thesis is limited only to the first iteration due to practical reasons.

The direct costs associated with the research are also not negligible, which can be divided into three categories: LLM operational costs, cloud infrastructure cost, and cost of recruiting participants. These costs also limit the research duration, sample size, and the overall validity of the research.

3.5 Ethical Issues

Despite these limitations, ethical considerations were not sacrificed to conduct the research. The basis of research ethics follows the ethical guidelines of the "Research Ethics: A Handbook of Principles and Procedures" (University of Gloucestershire, 2018). In line with the six principles of the handbook, the pillars of the research project are autonomy, beneficence, non-maleficence, justice, fidelity, and academic freedom.

1. Autonomy: The user data gathering of the second stage targets only the adult population and requests informed consent before the experiment takes place. Privacy concerns, such as data confidentiality and anonymity, are upheld as part of the research process.
2. Beneficence: The research aim is to contribute to knowledge for the benefit of the coaching community, and in a broader context to the world, by enhancing and making coaching, an already applied and effective HRD practice, accessible to more people.
3. Non-maleficence: The risk of potential harm is analysed and incorporated in the design of the newly designed and developed framework as well as in the developed software prototype.
4. Justice: The equity of participant treatment is ensured primarily by anonymity and automation
5. Fidelity: Honesty, integrity, and trust are fundamental values incorporated in the research design.
6. Academic freedom: The research project that underpins this thesis is self-funded, without involvement or influence by any organization.

3.6 Chapter Summary

Chapter 3 described the research design for developing and validating a novel framework (COBAC) aimed at fostering autonomous coaching systems. A distinctive aspect of this study is its positivistic epistemological stance from the perspective of technology research, presenting a contrast to the psychotherapeutic approaches predominantly found in coaching research. The study proposes four hypotheses to validate the COBAC framework's feasibility, effectiveness, language influence on adoption factors, and dialog quality in autonomous

coaching, aiming to contribute a robust framework for future developments of AI coaching systems.

Initially, the COBAC framework was constructed through document-centric methodologies and established requirements engineering approaches, drawing insights from existing coaching competencies and scholarly work. The proposed framework consists of components, such as problems, objectives, solution requirements, and constraints to address stakeholders' concerns. The quality criteria of the framework are defined based on International Requirements Engineering Board guidelines, emphasizing the importance of clarity, consistency, modifiability, and traceability in ensuring the robustness and adaptability of the framework in a rapidly evolving domain. Potential biases and complexities are acknowledged in the framework building process, recognizing the presumptive transferability of human coaching dynamics to autonomous coaching, and the intricacies involved in competency framework comparisons, while reaffirming the methodology's utility in capturing the essential elements of coaching.

Following this, a software prototype is developed based on the COBAC framework and validated using the Design Science Research Methodology. The development, grounded in the researcher's expertise, led to an online autonomous coaching system tested with approximately 120 participants across three language groups via Amazon's MTurk platform. Post-session surveys, employing various measures from existing frameworks, collect data to evaluate the coaching effectiveness and dialog quality. Data analysis utilises Partial Least Square Structural Equation Modelling (PLS-SEM) to validate hypotheses, particularly focusing on language effects on adoption factors. The thesis acknowledges limitations like MTurk-associated challenges, development constraints, and research costs while ensuring adherence to ethical guidelines from the University of Gloucestershire (University of Gloucestershire, 2018), emphasizing values like autonomy, beneficence, and justice throughout the research process.

4. New Framework Building: Key Components

4.1 Introduction

Chapter 4 addresses the key components of the COBAC framework - stakeholders, problems, objectives, and constraints - that characterise an autonomous coaching system. Following the reductionist framework building approach, the analyses in Chapter 4 are aiming at deriving from the literature only the most influential elements, without impacting largely the representativeness of the COBAC framework.

4.2 Stakeholder Analysis

As discussed in Section 3.4.4, the stakeholder and context analysis aim to identify the most critical and influential problems of key stakeholders, that underpin the research. This section addresses the question of who might be directly or indirectly involved in a coaching relationship, both in terms of human-to-human interactions and autonomous coaching processes. This conceptualisation step is crucial for understanding which players influence an effective coaching process, as problems and objectives cannot be interpreted without considering the subjective perceptions of the stakeholders.

4.2.1 Approach

The stakeholder analysis consists of three stages: identifying key stakeholders, mapping these stakeholders from traditional human-to-human coaching to autonomous coaching, and conducting a detailed analysis of each stakeholder. This analysis is fundamental to understanding the problems at hand, which are inherently linked to stakeholders as outlined in a specific section of the literature. The result of this process is a Stakeholder List, which serves as a precursor to a more focused analysis of the macro-context and micro-context.

4.2.2 Stakeholder Identification

The two primary stakeholders associated with human-to-human coaching are the coach and the client. Client is defined as “the individual or team/group being coached, the coach being mentored or supervised, or the coach or the student coach being trained” (ICF, 2019b). Similarly, the EMCC defines the as “denotes anyone using the services of a mentor/coach/supervisor in return for monetary considerations” (EMCC, 2018). The coach "is an expert in establishing a relationship with people in a series of conversations with the purpose of: serving the clients to improve their performance or enhance their personal development or both, choosing their own goals and ways of doing it and interacting with each person or group by applying one or more relevant methods, according to standards and ethical principles set up by EMCC and other professional associations" (EMCC, 2018)

In addition to the fundamental roles of coach and client, indirect roles are also involved in coaching. The majority (53%) of clients are sponsored clients globally (PwC & ICF, 2016), indicating that there is an organisation in the background with its own interest, usually represented by a senior manager or a line manager who also has specific expectations from the coaching process. Two roles stand out in this regard: sponsor and coach commissioner. A sponsor is defined by ICF as “the entity (including its representatives) paying for and/or arranging or defining the coaching services to be provided” (ICF, 2019b). In contrast, “coach commissioners are individuals within organisations who are charged with the responsibility for commissioning coaching” (Jonathan Passmore et al., 2017). Commissioning, in the coaching context, refers to supporting the coach selection process and helping with administrative tasks throughout the coaching process. Commissioners are usually represented by an employee from the HR department. Unlike the sponsor, who introduces new interests from the organisation to the process, the coach selection and administration can also occur in a private context. According to the EMCC report (Jonathan Passmore et al., 2017), when asked who the primary client in an organisational setting is, 58% said the individual client, 7% identified the commissioner, 2% named the sponsor, 30% selected all of equal importance, and 3% responded ‘whoever is paying’. This data underscores the importance of these stakeholders in a business context.

Coaches can bolster their credibility with clients through accreditation provided by coach accreditation bodies. The most prominent international accreditation bodies include the International Coaching Federation (ICF), the European Mentoring and Coaching Council (EMCC), the International Association of Coaching (IAC), and the Center for Credentialing & Education (CCE). They offer accreditation based on some or all the following criteria: training completion, professional experience, knowledge-related exams, participation in mentor coaching, adherence to enforceable ethical codes, and commitment to continued education. The ICF is the leading coaching accreditation body with 52,905 members globally in 168 countries as of October 2022 (ICF, 2022b), while the EMCC Directory lists 16,281 members (EMCC, 2022) during the same period. These coaching accreditation bodies further support their members with services such as funding relevant research, preparing supporting material for practitioners, helping with ethical issues, maintaining a client-facing member directory, helping establish local chapters of the organisation, and accrediting coaching training organisations.

Training organisations play a pivotal role in a coach's initial journey to hone their coaching skills. However, their role does not extend to imposing new needs or requirements on the coaching process or the coach, so this role is not included in the COBAC framework.

Much like training, coaching supervision is another process linked with coaching that aids the personal development of the coach. The EMCC defines it as “the interaction that occurs when a mentor or coach brings their coaching or mentoring work experiences to a supervisor in order to be supported and to engage in reflective dialogue and collaborative learning for the development and benefit of the mentor or coach, their clients and their organisations” (EMCC, 2016). The EMCC mandates one hour of supervision for every 35 hours of practice (EMCC, 2016). However, data from practitioners reveals that 34% of coaches do not undergo supervision (Jonathan Passmore et al., 2017). Neither the Association for Coaching (AC) nor the International Coaching Federation (ICF) has a similar requirement, but the ICF recognises 10 hours of coaching supervision towards credential renewal (ICF, 2020a). A recent study on the state and future of coaching supervision concludes that “although most in the industry agree that supervision plays a valuable role, few agree about what that role should actually

be” (Tkach & DiGirolamo, 2017). Given the nascent nature of this role and its limited use in practice, it is not included in the COBAC framework.

Lastly, while the role of researchers is somewhat indirect, there are few processes and stakeholders in the coaching context that remain uninfluenced by them:

1. From a coach's perspective, market research data provides invaluable insights into market trends and competition. Research on coaching processes and outcomes informs them about best practices in the ever-evolving coaching domain.
2. Clients and coach commissioners seek service consistency; thus, research on coaching processes and outcomes is indirectly pertinent to them.
3. Sponsors need to justify their return on investment, making coaching outcome research crucial.
4. Accreditation institutes anchor their operations in research: competency models, ethical codes, and indirectly, accreditation requirements are all research-based.

The key stakeholders identified in a human-to-human coaching context are client, coach, sponsor, coach commissioner, accreditation institute(s), and coaching researchers.

In contrast to the coaching context, the two primary stakeholders in the realm of system engineering are the users and the system developers. According to the ITIL 4 standards, user is defined as “a person who uses services” (AXELOS, 2019). As systems grow more complex, additional roles emerge, such as support, which focuses on system stability, unlike developers who introduce changes and enhancements. These roles have converged into cross-functional teams in recent decades due to the rise of the DevOps concept and its newer incarnation: DevSecTestOps. While development and operational tasks are increasingly automated, these roles are still predominantly occupied by humans.

In the age of cloud computing, it's nearly impossible to operate a system without the involvement of service providers at various stages of system development or operations. A service provider is defined as “role performed by an organization in a service relationship to provide services to consumers” (AXELOS, 2019). Service providers can fulfil various functional roles. For example, an LLM-provider facilitates access to Large Language Model functionality

via APIs, while a cloud service provider can supply the infrastructure needed for system operation and development.

The key stakeholders identified in the system engineering context are User, DevOps, Service Provider(s).

4.2.3 Stakeholder Mapping

The interests and functions of the roles identified in the coaching context continue, to some extent, after automation. Stakeholders from the coaching context are mapped to the autonomous coaching context. Figure 4.1 depicts the new configuration of stakeholders in the form of a UML Use-Case diagram.

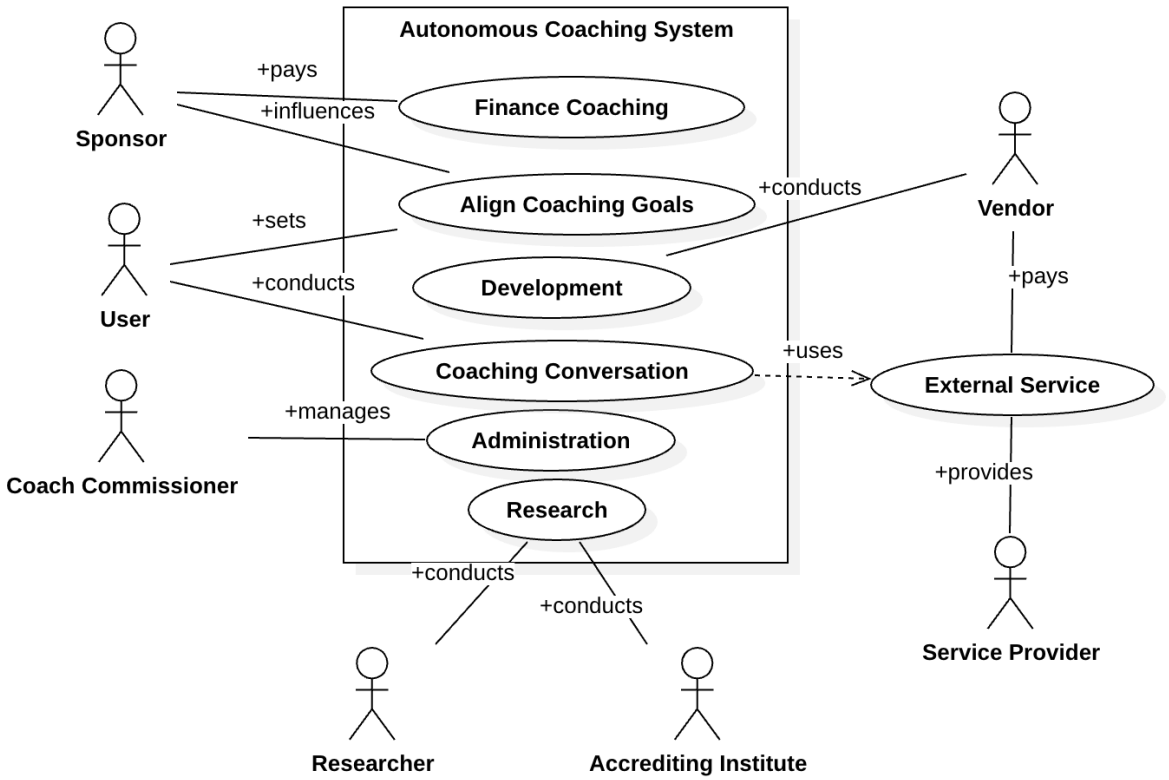


Figure 4.1: Stakeholders in the Context of Autonomous Coaching System

4.2.4 User (Client)

Within system engineering, the coaching client is termed a 'user'. This transition from a human-to-human coaching client to using an application not only necessitates a different communication modality, including meta-communication, but also presents opportunities for coaching anytime and anywhere.

Information regarding coaching clients is limited due to the confidentiality inherent in coaching relationships. While some sparse statistical data is available, it is typically gathered indirectly via surveys filled out by coaches, reflecting what coaches are permitted and willing to disclose. Based on this limited data, a brief analysis of five key characteristics is presented: client's role, age, mental health, coaching objective, and value orientation.

The ICF Global Coaching Study surveyed coach practitioners about their clients, revealing: "managers were mentioned most frequently (29%). Almost one in four (23%) said they mainly coached executives, with a further 19% selecting personal clients" (PwC & ICF, 2016). The top three roles held by clients were manager, executive, and 'business owner/entrepreneur', with 'personal client' in fourth place.

The same study reported age distribution as follows: 5% under 25 years, 18% aged 26-34, 37% aged 35-44, 31% aged 45-54, and 10% older than 54.

A report on the European coaching landscape highlighted the primary focus of coaching as "coaching clients to help with work performance" (Jonathan Passmore et al., 2017), suggesting that coaching predominantly occurs in a business context. An ICF study from the same year confirmed this, listing the top three coaching areas as optimising individual/team performance, improving communication skills, and increasing productivity (ICF, 2017a).

Data on clients' mental health is scarce. However, a life-coaching study with 107 participants found 25 participants exhibited high levels of psychological distress (Green, Oades, & Grant, 2005). This suggests that the proportion of coaching clients with significant mental health issues might exceed that of the general population. There's a debate in coaching research

about introducing mental health screenings (Cox et al., 2010). This is due to concerns over coaches' ability to support clients with mental health challenges (McKelley & Rochlen, 2007), despite some studies suggesting potential benefits (Lerner et al., 2015).

Value orientation was a major theme in system engineering over the past two decades, influenced by movements and approaches like DevOps, agile, and lean. This value-seeking behaviour is also evident in AI coaching, significantly influencing user intent to use the application (Terblanche & Cilliers, 2020).

The underlying assumption that client characteristics remain the same in AI coaching context as in human-to-human coaching might be wrong, because introducing automation can change the characteristics of the clients. For example, offering coaching services digitally could shift the target demographic from middle-aged managers to a broader, younger audience, akin to trends seen in the MOOC market (Bayeck, 2016).

This literature review suggests that coaching is versatile in its global applications. However, a typical coaching client often exhibits the following traits: aged between 35 and 55, seeks business coaching in a managerial role to enhance business-related skills or performance, has a high likelihood of stress, and demonstrates a strong value orientation when using AI coaching services.

4.2.5 Coach

Unlike the client, the role of the coach in the autonomous coaching context is replaced by the autonomous coaching system. However, understanding the characteristics of a typical human coach from the human-to-human coaching context is crucial. This understanding can offer initial insights into the skills, behaviours, and knowledge levels that need to be mapped to an algorithm during the initial stage of the automation process. Over time, the significance of understanding the human coach may diminish in the AI coaching context as algorithms learn and evolve based on the increasing amount of client-generated data.

There is a wealth of information available about coaches compared to clients. This is due to several reasons: market research on coaches is financed and conducted by well-funded coaching training and accrediting institutes; coaches are keen to understand their competition in a saturated market; and there is no binding confidentiality agreement with a third party.

According to the 2016 ICF Global Coaching Study (PwC & ICF, 2016), a typical coach practitioner resides in North America or Western Europe, uses coaching skills either as a professional coach or a manager, and over half are aged between 45 and 59 years. Additionally, 68% have completed more than 125 hours of coaching-specific training.
A study focusing solely on Europe (Jonathan Passmore et al., 2017) confirms the magnitude of these figures and adds that most coaches are not members of any coaching organisation. If they are, the ICF is the most likely choice, followed by national coaching federations and the EMCC. Almost half have between 1 and 7 years of coaching experience, suggesting that coaching in Europe is still a relatively new trend. Only 10-30% of their time is dedicated to coaching, indicating that it is often a part-time role rather than a full-time job. Continuous professional development is a priority, with around 60% of coaches investing between 16 and 60 hours annually in this activity.

Sperry estimated that, although coaching heavily relies on practices derived from psychotherapy, only about 20% of executive coaching practitioners possess a degree in psychology (Sperry, 2004).

In summary, a typical coach, based on generalised data, resides in North America or Western Europe, is well-trained, dedicates significant time to continuous development, is an experienced business professional who practices coaching part-time, and is often not affiliated with coaching organisations or holds a degree in psychology.

4.2.6 Sponsor

It is a very much under researched area of coaching considering that most coaching engagements take place in business context where sponsor is a likely stakeholder. Among the

very few studies that focus on sponsors of coaching engagements, the crucial moment related research stands out (De Haan & Nieß, 2015), which provides insights into the expectations of sponsors, but not much information is available about the characteristics of this stakeholder group.

4.2.7 Commissioner

In contrast to the outcome-focused approach of sponsors, coach commissioners are deeply involved in the operational aspects of coaching engagements. The main concerns of the coach commissioner role encompass procurement, contracting, management, and evaluation of the quality and impact of coaching (Jonathan Passmore & Woodward, 2023). Therefore, skills in these areas are of paramount importance.

4.2.8 Vendor (DevOps)

A broader organizational form of DevOps teams is covered by the terminology vendor, which refers to the practical reality that system engineers are usually represented by companies, that sell AI coaching services. At the time of writing the thesis there seems to be no research about AI coaching system vendors, but a recent study on the coaching education listed the following “AI Coach / Conversation bots” in the market segment and provider review: AI Coaching, AIMY (CoachHub), CAI (EZRA), Evoach, Pocket Confidant, rocky.ai (Jonathan Passmore & Woodward, 2023).

4.2.9 Service Providers

For LLM-based AI coaching services, most of these elements must be coordinated with a cloud service provider and LLM provider to guarantee that the AI coaching system delivers added value to the clients. According to the definitions provided by ITIL 4, the primary concern pertains to service provisioning, service consumption, and service relationship management. Service provisioning encompasses activities such as the management of the provider's resources, which are configured to deliver the service; ensuring users have access to these resources; fulfilling agreed-upon service actions; service level management; and continuous improvement. Service consumption involves the management of the consumer's resources

required to utilise the service, actions performed by users, and the acquisition of goods (if necessary). Service relationship management denotes the collaborative activities undertaken by both the service provider and the service consumer to ensure continuous value co-creation based on agreed and available service offerings (AXELOS, 2019).

4.2.10 Accrediting Institute(s)

Accrediting Institutes can also be viewed as service providers. However, unlike technology service providers that interact directly with the developed system, accrediting institutes engage with DevOps teams using requirements, regulations, and guidance documents. They are identified as stakeholders undergoing significant change because providing such requirements or guidance necessitates the accrediting institutes to invest effort in developing this AI coaching support, guidance, and regulatory capability. An illustrative example of such a document is "The Artificial Intelligence (AI) Coaching Capability Assessment Guidelines" (ICF, 2022c), which targets AI coaching developers, coaching practitioners, buyers, professional bodies, and researchers.

4.2.11 Researcher

A recent OECD report on the impact of digitalisation on research notes that "the life sciences (with the exception of pharmaceutical) and the physical sciences (other than engineering) report the greatest effort to make data and/or code usable by others" (OECD, 2019a). This observation is also relevant to coaching researchers. Although the challenging and labour-intensive task of generating transcripts from audio or video recordings was automated due to the rise of AI-driven tools like TranscribeMe, Otter, and Trint, the issue of coaching conversation confidentiality persists. Competency-based AI coaching systems produce more accessible data by anonymising large-scale sensitive conversation data or by offering a mechanism to manage the use of sensitive data, ensuring confidentiality. This advancement can initiate profound changes on various fronts: it can bring the era of evidence-based coaching for the benefit of both AI systems as well as human coaching context; allow the application of new methodologies, data collection, and analytical methods; and enhance the

validity of research. This significant shift is the rationale for identifying this group as a stakeholder undergoing substantial change.

4.2.12 Outcome

Table 4.1 summarises the identified stakeholders and their key characteristics based on the stakeholder analysis in Section 4.2. The stakeholder characteristics column does not aim to establish strict boundaries for including or excluding stakeholders in the study. Instead, it outlines typical traits to guide further analysis based on these features, like the concept of personas in user research.

Table 4.1: Initial Stakeholder List

Key	Name	Stakeholder characteristics
S.1	User (Client)	- aged between 35 and 55 - holds a managerial position - receives coaching in business coaching context - aims to improve performance and skills - likely to experience significant stress - strong value orientation when using AI coaching services
S.2	Coach	- resides in North America or Western Europe - well-trained and dedicate significant time to continuous development - an experienced business professional who practices coaching part-time - most are not members of coaching organisations nor possess a degree in psychology
S.3	Sponsor	- values client behavioural changes and improvements in client communication - considers the return-on-investment aspect
S.4	Commissioner	- proficient in skills related to procurement, contracting, management, and evaluation of the quality and impact of coaching
S.5	Vendor (DevOps)	- limited information availability of key players, such as AI Coaching, AIMY (CoachHub), CAI (EZRA), Evoach, Pocket Confidant, rocky.ai
S.6	Service Providers	- key roles: service provisioning, service consumption, and service relationship management
S.7	Accrediting Institutes	- capable of interacting with the DevOps teams using requirements, regulations, and guidance documents
S.8	Researcher	- values access to large amount of coaching related sensitive data

4.3 Macro Context Analysis

4.3.1 Approach

To define the value proposition and objectives of the AI coaching system covered in the COBAC framework, it is crucial first to identify the broad and specific context boundaries, the involved stakeholders, and their needs. Some of these identified problems are conflicting; for example, while the high cost of coaching is a problem for the client, it naturally represents value for the coach. Some issues affect all involved parties; for example, the lack of regulations poses a service quality risk for the client, while simultaneously making the service selling process more challenging for qualified coaches. To simplify these concerns, only one stakeholder is associated with each problem — the primary stakeholder — indicating who is perceiving the problem or being impacted the most by the stated problem.

Since the macro-context is constantly changing, the impact of these changes on AI coaching systems needs to be continuously assessed to ensure the system remains relevant, effective, efficient, and ethically conforming to the latest research insights.

As discussed in Section 3.4.4, the thematic macro context analysis of autonomous coaching systems utilises the STEEPLE model to cover the fundamental contextual factors in the subsequent sections: social, technological, economic, environmental, political, legal, and ethical.

4.3.2 Social Factors

The perception and awareness of coaching in society are controversial. According to the ”2022 ICF Global Consumer Awareness Study” (ICF, 2022a), which is based on a global survey incorporating 30,727 participants from 30 countries, 28% are ‘very aware’ of coaching, 44% are ‘somewhat aware’, and 27% are ‘not aware’ of coaching. However, when the participants who answered ‘very aware’ or ‘somewhat aware’ were presented with the definition of five organisational support professions (coaching, mentoring, consulting, training, and

counselling), only 25% selected the correct definition of coaching, a decline from 30% compared with the 2017 version of the same study (ICF, 2017a). This indicates confusion about what coaching is and is not, even within the group of people who believe they know what coaching is.

The top barriers to coaching have also significantly increased, comparing the survey results between 2017 and 2022. Among participants with at least some awareness of coaching and who said they would not participate in coaching, the following reasons topped the list: 'I am at a time in my life where coaching is not important' (43% in 2017, not listed in 2022), 'I can't afford coaching' (26% in 2017, 49% in 2022), 'Coaching is expensive' (58% in 2022, not measured in 2017), 'I do not need coaching' (24% in 2017, 49% in 2022), and 'I do not believe there is evidence that coaching works' (15% in 2017 and 26% in 2022) (ICF, 2017a, 2022a).

As discussed in the stakeholder analysis section, the users of an autonomous coaching service might have a slightly changed profile compared with human-to-human coaching regarding the average age of users. Consequently, the younger generation might broaden the scope of coaching topics from business skill development to a great variety of other topics, situations, and skills that are more of a concern to this age group. The impact of targeting a different consumer segment impacts not only the scope of coaching topics but also influences requirements due to the different consumer behaviours and expectations.

Consumer behaviour related to digital products has completely shifted in the last decades. Since the launch of the iPhone in 2007, for the first time in history, everyday people had access to technology that is more advanced from a usability point of view than the technologies available for corporations. This change led to elevated user expectations, one of the driving forces behind the paradigm shift in the field of software design patterns, such as 'mobile first' or responsive design. The rapidly evolving user experience in the private sphere of smartphone users remained unmatched by most corporate software systems that struggle to compete, even though some suggest that "investment in mobile first applications by business and government, is likely to grow" (P. Lee & Calugar-Pop, 2019).
A second area where expectations have skyrocketed is related to availability: information needs to be available quickly, everywhere, and all the time. This has led to significant growth

in new forms of employment, such as remote-only jobs or the phenomenon of digital nomads. Changes related to availability expectations have led to an increase in independent workers and "a new development regarding independent work is the growth of the so-called gig economy, which involves people providing their labour services through on-demand digital marketplaces" (Lund et al., 2019). Any form of dependency is not acceptable due to the high impact on the effectiveness of remote working, whether it is time-dependency, location-dependency, or dependency on a person. Smartphone users were empowered to access information anytime and anywhere instantly; however, it has also impacted the availability of workers: it is increasingly difficult to resist the notion of always being connected.

Furthermore, it raised employee expectations related to corporate infrastructure that provides this connectivity because organisations play an important role in the enablement of social connectedness (Sethi, Stubbings, Gratton, & Brown, 2017).

Hyper-connectedness was a prerequisite for the emergence of social media, which has fundamentally changed not only how we communicate with each other but an OECD report suggests that "constant connectivity is changing attitudes and behaviour in people's personal life, with the transfer of part of social relations online and the blurring of work and leisure time" (OECD, 2019b). A survey about digital transformation concludes that "instant messaging and social networking account for the majority of time spent online" (OECD, 2019a). Social media and instant messaging are just two of the most widespread recent examples that have changed how we interact with each other. Coaching accreditation bodies and training organisations provide social connectedness to coaching practitioners; however, there is a lack of social connectedness among coaching clients due to the confidential and siloed nature of coaching conversations.

Consumer acceptance of new technologies has also accelerated, considering that it took several decades to embrace the telephone, but less than a decade to do the same with smartphones globally. The diversity of the top frequently used digital services that did not exist a couple of years ago, such as Twitter, Snapchat, or Instagram, also indicates faster progress in consumer acceptance (Comscore, 2017). People show signs of being able to connect even with socially underdeveloped robots on an emotional level, as observed in a

military context related to Explosive Ordnance Disposal (EOD) units. “EOD units have been known to promote their bots to titles such as Staff Sergeant, award them Purple Hearts, and even hold funerals for their fallen brothers in robotic arms. They assign them personalities, dispositions” (Rose, 2014).

The trends of technological development combined with demographic changes have staged a situation where a new generation has entered the workforce, who are not only “born digital” but for whom digital is their natural language. Companies need to adapt their brand, culture, systems, processes, and services to attract and retain talented workers from this generation, and the domain of coaching is no exception. The Human Resource Development industry is at the beginning of the journey to embrace these fundamental consumer changes, hence there is a big opportunity for improvement.

The major trends and factors of the social context can be summarized as poor awareness of coaching; barriers to participation are irrelevance, and lack of scientific evidence; expected coaching goal diversity due to the shifting coaching target segment; unmatched software usability expectations; unmatched availability expectations related to time, location of service usage; dependency reduction; lack of social connectedness function among coaching clients; unexploited opportunity of human capability to connect emotionally with algorithmic agents and increased adaptation speed to new technologies.

4.3.3 Technological Factors

As discussed in Sections 2.3.2 and 2.3.3, the current models for AI coaching are not adaptable and exhibit severe limitations, rendering them inadequate to address the high complexity involved in human interactions. Even when applications are developed based on specialized models, the choice of technology can be a limiting factor. For instance, pre-LLM conversational technologies offer poor dialogue quality, and most of the currently running AI coaching systems do not utilise LLMs.

At the same time, the scaling of the newest technologies, already deployed in other domains, remains sluggish. For instance, according to the 2019 Executive Coaching Survey (Sherpa Coaching, 2019), only 32% of coaching sessions were provided through in-person meetings.

Coaching services delivered through some forms of audio-video-based channel have increased from 0% in 2007 to 35% in 2019, replacing other forms of interaction such as telephone, email, chat, or face-to-face meetings. The digitalisation of the communication channel was driven by socio-economic factors, which are not specific to coaching: lowering the invested time and cost associated with physical presence and the shift to remote work. Even with the digitalisation of the communication channel, the coaching industry remains generally a low-tech but growing industry, making it ideal for digital disruption.

Coaching currently faces challenges due to its low-tech nature, the absence of adaptable autonomous models, the subpar performance of deployed conversational technologies in addressing high-complexity human interactions, and the untapped potential of technologies from other fields.

4.3.4 Economic Factors

Despite its challenges, coaching is a growing industry from an economic perspective: “the estimated global total revenue from coaching in 2015 was $2.356 billion USD, representing a 19% increase over the 2011 estimate” (PwC & ICF, 2016). Training of coaches and building coach competence is typically provided through a program, which combines theoretical lectures and coaching exercises over several months. This approach has many benefits; however, one disadvantage is that it is difficult to scale coaching competence: it takes a long time, is human resource intensive, and is thus expensive.

The high cost of building coach competence, combined with the opportunity cost of well-trained coach candidates, results in a relatively high hourly rate for coaches. Coaching fees vary greatly based on location, experience, and other contextual factors; however, according to a study conducted in Europe, most coaches working in a corporate setting charge between 100 and 400 Euro per hour (Jonathan Passmore et al., 2017). The direct cost of an average coaching process might be between 500 Euro and 4000 Euro, considering that a complete process might consist of 5 to 10 coaching sessions, not counting the opportunity cost to the client.

The prohibitive cost of a typical coaching service, considered by HRD professionals to be an effective way of skill development, is hardly affordable for blue-collar workers. Market research confirms that organisations tend to hire coaches only for their management or highly valued subject matter experts (PwC & ICF, 2016). The current form of human-to-human coaching is contributing to social inequality through these economic dynamics, even if the trend is positive, as coaching "once offered only to top management, coaching is now also used with middle and lower levels of management and offered to individuals and groups with various life and work-related coaching issues" (Bachkirova et al., 2016). Massive open online courses (MOOCs) certainly support the democratisation of education by providing accessibility to information for those who are unable to afford education (Dillahunt, Wang, & Teasley, 2014) ; however, automating trainings alone is not enough to make a difference. A similar shift is expected in the coaching domain, analogous to the disruptive change caused in the training field by the exponential growth of online trainings. A significant breakthrough to expand the coaching target segment from high-earning leaders and specialists to individuals with less earning power cannot happen unless the prohibitive cost of a coaching service is dramatically reduced. Once this major barrier is addressed, coaching might contribute to solving inequality rather than reinforcing it.

At the same time, technological breakthroughs over the last few decades have contributed to the rise of automating typical human tasks, leading to a significant increase in global economic productivity. It is estimated that global gross domestic product (GDP) growth could be up to 14% higher in 2030 due to automation (Rao & Verweij, 2017). The impact of automation varies depending on the current development of the country, the industry, the education level of workers, and the age group of the workers. For example, one study conducted in Germany concluded that "more robot exposed workers are even more likely to remain employed in their original workplace, though not necessarily performing the same tasks, and the aggregate manufacturing decline is solely driven by fewer new jobs for young labor market entrants" (Dauth, Findeisen, Südekum, & Wößner, 2017). Despite the education sector, where coaching as a human development practice belongs, being estimated as one of the least prone to automation (Hawksworth et al., 2018), there is growing pressure to increase the productivity of the sector due to the soaring demand for upskilling or cross-skilling caused by the continuously spreading automation. Overall, technological breakthroughs are causing rapid

shifts in the job market, resulting in an exponentially increasing demand for effective Human Resource Development practices.

However, financial incentives of technology development do not currently encourage solving social problems in many cases, and sometimes they even exacerbate existing problems, as is frequently discussed in relation to issues such as fake news (J. A. Braun & Eklund, 2019) or addictive software design patterns (Neyman, 2017).

In conclusion, the coaching industry is economically growing, but it faces challenges due to the high costs and time-intensive nature of training coaches, resulting in high hourly rates for services that are predominantly accessible to management and highly valued experts only. It contributes significantly to social inequality, while the sector also faces pressure to increase productivity amidst rising demand for upskilling due to automation, despite being one of the least prone sectors to automation. Current financial incentives for technology development often fail to address social problems and can intensify issues.

4.3.5 Environmental Factors

As discussed in Section 2.4.4, the relatively high resource consumption of NLP systems contributes significantly to the environmental impact of AI coaching systems. However, there is no unified reporting and measurement metrics applied in the industry to create awareness of the problem. Without regulations, economic incentives, or any other form of incentives, the prioritization of resource-efficient technology remains at the discretion of the software vendor.

4.3.6 Political Factors

Despite comprehensive research on the economic and social impacts of automation, there is a lack of political discourse in the Western world about automation and artificial intelligence. As Yuval Noah Harari framed the situation of artificial intelligence and politics: “even though in the academic world and private business people were aware of the potential at least, in the political field and public discourse you hardly heard anything about it” (Anthony, 2018).

With the launch of ChatGPT, several concerns of this applied technology have surfaced in the public conversation, such as job security, ethical concerns, and safety risks. However, a meaningful debate, for instance in a political election campaign, remains to be seen.

The State of AI Report 2023 lists topics under the section politics, such as regulatory approaches, global governance, chip war, compute capacity, funding, AI warfare, election manipulation, culture and value wars, and job security (Benaich & Air Street Capital, 2022). This list shows the spectrum of topics that are impacted by AI and illustrates that political activities are ongoing, but decisions and progress remain untransparent.

In the more specific context of educational technology, despite the proven potential of addressing inequality in society, there is also little progress. Inequality is described as “the defining challenge of our time” by Barack Obama (Newell, 2013), and a book related to the impact of education on inequality has summarized the current status as follows: “Americans like to believe that theirs is the land of opportunity, but the hard facts are that children born into poor families in the United States tend to stay poor and children born into wealthy families generally stay rich” (Smeeding, Jäntti, & Erikson, 2011).

Considering the lack of public conversation, politics does not play a significant role in the development of AI coaching now, but it might play a vital role in the future of AI development through influencing legislation.

4.3.7 Legal Factors

Coaching is a self-regulated industry globally, and it is explicitly registered in the EU’s database for self-regulation initiatives in "The Professional Charter for Coaching, Mentoring, and Supervision of Coaches, Mentors, and Supervisors" (EMCC, ICF, AC, & SFC, 2021). From a legal perspective, this means that a coach is not required to have a license or mandatory accreditation from a professional body; however, general regulations such as data protection law might apply depending on the specific country or industry. Coaching bodies are well-positioned to be subject matter experts (SMEs) for regulating this profession one day, as 53% of coaches believe that coaching should become regulated and among those who were

unsure, 84% felt that coaching bodies were best placed as regulators (PwC & ICF, 2012). One of the consequences of self-regulation is that, from a client's point of view, there is a significant service quality risk when it comes to the selection of coaches. This risk is significant, as 'questionable expertise of coaches' was named as a reason in 53% of the cases related to early coaching contract termination (Tompson et al., 2008). Without coaching-specific laws, the guiding principles of coach behaviour at the time of writing this thesis are market logic, ethical norms provided by coaching accreditation institutes, and general laws.

One example of generic legislation that impacts coaching is the proliferation of data protection-related laws. Just as automation was not regulated until recently, there was a growing concern about the data privacy of citizens in the European Union, leading to the introduction of the General Data Protection Regulation (GDPR) (European Parliament and of the Council, 2016) to make organisations accountable for managing privacy-related data. Confidentiality plays an important role in building and maintaining trust in the coaching process (Palmer & Gyllensten, 2007), but unlike in the human-to-human coaching situation, it is realised in a technology-driven platform primarily through the application of information security-related measures. Global legislation is far behind in regulating automation-related technologies, despite their increasingly important role and ethical implications (Benaich & Air Street Capital, 2022).

Europe seems to play a leading role in regulating the technology space, as after launching comprehensive GDPR regulation in 2016, an AI Act is also scheduled to be released in the upcoming years (European Parliament and of the Council, 2021).

However, at the time of writing this thesis, there are only two legally relevant aspects that need to be considered: the service quality risk posed on clients due to the unregulated coaching industry, and data protection law compliance.

4.3.8 Ethical Factors

Lastly, Chapter 2 dedicates three sections to ethical concerns, encompassing coaching (2.2.7), autonomous coaching (2.3.4), and technology-specific (2.4.4) ethical issues. These are summarized here to complete the STEEPLE macro-context analysis.

Section 2.2.7 underscores key ethical concerns in coaching, including the imperative for a universal ethical code that continuously incorporates ongoing debates on aspects not covered by the latest research on ethical issues. It also addresses the issue of underreported ethical breaches, the potential contribution of coaching to societal issues such as burnout and the 'society of achievement,' and the necessity for increased transparency in coaching practices.

Section 2.3.4 highlights the key ethical concerns in the development of autonomous coaching systems. It emphasizes the need for a holistic approach that integrates ethical considerations into the design phase, addressing both individual and societal dilemmas. These include the lack of transparency, absence of executive support, and the need for monitoring fairness and ethics. It also covers the protection of sensitive client information in cyberspace, the impact of technology on physiological and mental health, biases related to over-trust in automated systems and algorithmic bias, job displacement, the need for re-skilling, and a lack of consensus on core ethical issues. Recurring themes in this section include autonomy, privacy, and transparency.

Section 2.4.4 outlines key ethical concerns in LLM-based applications, particularly in AI coaching. It discusses individual privacy issues arising from sharing confidential information with third-party LLM operators, security risks, and broader societal challenges. These include copyright debates, biases in generated content, environmental impacts from resource-intensive NLP processes, and inequality issues stemming from predominantly English-based LLM development, which benefits privileged societies while disproportionately affecting marginalized communities.

4.3.9 Outcome

As demonstrated in the thematic macro-context analysis, a wide spectrum of issues was identified from various perspectives, influencing both key stakeholders and the AI coaching process as it is summarized in Figure 4.2.

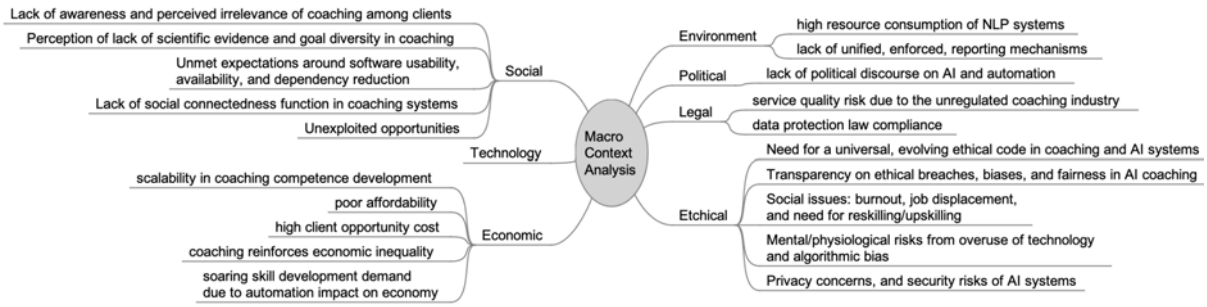


Figure 4.2: Macro Context Analysis - Overview

As discussed in Section 4.2.2, in its most simplistic form, the two essential stakeholders in a coaching engagement are the client and the coach. Analogously, in the AI coaching context, the two essential stakeholders are the user and the vendor developing the AI coaching system. As the Macro Context Analysis demonstrated, the literature, however, addresses more specific stakeholders and their needs. All these problems and needs also exist in a simplified two-stakeholder context; however, these aspects are represented either by the user or the vendor. For instance, a typical need of a sponsor in the context of an organisational coaching engagement is the minimisation of costs. If the coaching is paid for by the client, that cost awareness is represented by the client. Similarly, all the identified needs and problems can be attributed to one of the two primary stakeholders: the client (user) and the coach (AI coaching system). Table 4.2 contains the detailed list of all macro-contextual problems identified in the macro context analysis using the simplified, primary stakeholder assignment.

Table 4.2: Initial Problem List Derived from the Macro Context Analysis

Key	Problem Summary	Primary Stakeholder	Source
P.1	poor awareness of coaching among coaching clients	User	Macro Social
P.2	barrier: coaching is perceived irrelevant	User	Macro Social
P.3	barrier: perception of lack of scientific evidence	User	Macro Social
P.4	coaching goal diversity due to the shifting coaching target segment	User	Macro Social
P.5	unmatched software usability expectations	User	Macro Social
P.6	unmatched availability expectations related to time	User	Macro Social
P.7	unmatched availability expectations related to location of service usage	User	Macro Social
P.8	unmatched expectations on dependency reduction	User	Macro Social
P.9	lack of social connectedness function	User	Macro Social
P.10	unexploited opportunity of increased adaption speed to new technologies	User	Macro Social
P.11	unexploited opportunity of human capability to connect emotionally with algorithmic agents	User	Macro Social
P.12	low-tech nature	User	Macro Technology
P.13	absence of adaptable autonomous models	Vendor	Macro Technology
P.14	subpar performance of deployed conversational technologies	Vendor	Macro Technology
P.15	high-complexity human interactions	Vendor	Macro Technology
P.16	untapped potential of technologies from other fields	Vendor	Macro Technology
P.17	scalability issues in coaching competence development	User	Macro Economic
P.18	client direct cost high: poor affordability of coaching	User	Macro Economic
P.19	client opportunity cost high	User	Macro Economic
P.20	coaching reinforces economic inequality	User	Macro Economic
P.21	soaring skill development demand due to automation impact on economy	User	Macro Economic
P.22	financial incentives do not encourage solving social problems	Vendor	Macro Economic
P.23	high resource consumption of NLP systems	Vendor	Macro Environment
P.24	lack of unified, enforced, reporting mechanisms	User	Macro Environment
P.25	lack of political discourse with voters on AI and automation	User	Macro Political
P.26	service quality risk posed on clients due to the unregulated coaching industry	User	Macro Legal
P.27	data protection law compliance	Vendor	Macro Legal

P.28	need for a universal, continuously evolving ethical code that includes latest research	Vendor	Macro Ethical
P.29	transparency on underreported ethical breaches in coaching practice	User	Macro Ethical
P.30	social issue: burnout	User	Macro Ethical
P.31	integration of ethical considerations right from the design phase in AI coaching systems	Vendor	Macro Ethical
P.32	AI coaching system: creating ethical transparency	Vendor	Macro Ethical
P.33	AI coaching system: absence of executive support of ethical concerns	Vendor	Macro Ethical
P.34	AI coaching system: lack of monitoring ethical concerns, such as fairness	Vendor	Macro Ethical
P.35	weak protection of sensitive client information in cyberspace	Vendor	Macro Ethical
P.36	mental and physiological risks caused by over usage of technologies	User	Macro Ethical
P.37	over trust in autonomous systems	User	Macro Ethical
P.38	algorithmic bias	Vendor	Macro Ethical
P.39	social issue: job displacement	User	Macro Ethical
P.40	social issue: addressing soaring need for reskilling, upskilling	User	Macro Ethical
P.41	lack of scientific consensus on core ethical issues: autonomy, privacy, and transparency	Vendor	Macro Ethical
P.42	privacy related to sharing data with third party, such as LLM operator	User	Macro Ethical
P.43	LLM security risks	Vendor	Macro Ethical
P.44	social issue: LLM copyright	Vendor	Macro Ethical
P.45	social issue: bias elated to LLM generated content	User	Macro Ethical
P.46	social issue: environmental impact of resource intensive LLM operations	User	Macro Ethical
P.47	social issue: inequality due to English language bias	User	Macro Ethical

4.4 Micro Context Analysis

4.4.1 Approach

In contrast to research on macro-context analysis, data pertaining to the micro-context analysis of coaching, autonomous coaching systems, and LLM-driven AI coaching systems are considerably scarcer. Section 4.4 endeavours to conduct a preliminary systematic micro-context analysis to discern key issues inherent to the field. However, acknowledging the current research limitations implies that this analysis can, at best, offer insights for the development of the initial COBAC framework. The analysis concentrates on two facets: understanding the existing human-to-human coaching micro-context to facilitate the

transference of this knowledge to the initial AI coaching system and acquiring information about the autonomous coaching micro-context.

Like the macro-context analysis, our comprehension of the autonomous coaching micro-context is continuously evolving. Consequently, it necessitates the regular undertaking of micro-context analysis to continuously refine the COBAC framework and the AI coaching systems.

The coaching-related micro-context analysis is based on Grant's Multiple-Perspective Model (MPM) (Grant, 2017), selected for its comprehensive nature in assessing various aspects of the coaching process and the categorization provided by the model enables a systematic, structured analysis.

4.4.2 Coaching – Multiple-Perspective Model (MPM)

The Multiple-Perspective Model classifies the issues into one of the following ten areas: a – attributes of client, b – client's goals, c – client's immediate personal system, d – attributes of coach, e – coach's goal for coaching, f – coach' immediate personal system, g – coach-client relationship, h – sponsor's influence and experience, I – stakeholder's influence and experience, and j – distal system impact on coaching.

Regarding the 'attributes of the client', numerous studies have explored how the client's personal factors influence coaching outcomes, such as (MacKie, 2015) or (McCormick & Burch, 2008). However, there is limited research on how the client changes in attributes not directly targeted during coaching. From a control theory perspective, it is suggested that clients should "become more effective self-regulators and effectively engage in goal setting and feedback seeking, among other things, in order to facilitate development" (Gregory, Beck, & Carr, 2011). Other concepts include change readiness (Franklin, 2005) and developmental readiness (MacKie, 2015). While the 'attributes of the client' aspect may offer valuable insights for future versions of the COBAC framework, the current evidence base appears inconclusive.

There is a wide consensus on the role of the 'client's goal' in coaching. "Goal-setting remains a largely unquestioned element of coaching practice" (David, Clutterbuck, & Megginson, 2014), with goal-related competencies included in the competence frameworks of both ICF (ICF, 2019a) and EMCC (EMCC, 2015). The coach's role as a facilitator of client goal setting is crucial in a coaching relationship (De Haan et al., 2013). The scientific literature is very rich when it comes to the effect of goals on performance, the goal theory, and associated problems with goal setting. Locke has summarized these aspects in an article (Locke, 1996), which identifies key pitfalls related to goals as following: inappropriate difficulty level, avoidance of generic goals, lack of commitment due to importance or attainability, level of self-efficacy, response to failure and task strategy, lack of feedback to show progress, lack of confidence, avoiding wrong direction, insufficient effort, inconstant actions, poor action planning quality due to lack of knowledge or high pressure. An additional aspect that plays a significant role in the coaching context is that goals during coaching conversations are likely to change significantly (Clutterbuck, 2008).

The generalisation effect of how workplace coaching affects the client's private life, and vice versa, is a frequently discussed topic among practitioners but remains under-researched in the coaching domain. Given the insufficient level of scientific knowledge, the 'client's immediate personal system' is not considered in the initial version of the COBAC framework.

A study with 20 purchase executives identified 'coach attributes' contributing to exceptional coaching experiences, including "credibility, empathy and respect, holding the professional self, diagnostic skill and insight, approach flexibility and range, working to the business context, a philosophy of personal responsibility, and skilful challenging" (Dagley, 2010). Additionally, there is also mounting evidence (Rubie-Davies, 2009) that high coach expectations may influence outcomes, highlighting its importance as a major coach attribute.

The 'coach's goal' and motivation for coaching are under-researched areas, surprising given the extensive research in sport coaching (Grant, 2017). As the goals of an autonomous system in an automated context differ significantly from a human coach's motivation, this area is not discussed further in this study due to its lack of relevance and scientific evidence.

The impact of coaching on the coach's life is an important topic for all stakeholders. However, the 'coach' immediate personal system' is not directly relevant to the initial COBAC framework building.

Regarding the 'coach-client relationship', a study about common factors of successful coaching engagements concluded that "a picture emerged of a client valuing the relationship with and the qualities of the coach, while making little distinction between specific interventions of that same coach" (De Haan, Culpin, & Curd, 2011). Another study related to coach relationship confirmed the same and concluded that a client view of coach-client relationship had a central role in success (De Haan et al., 2013). Additionally, Grant also highlights the importance of a supportive, challenging relationship and the coach's ability to maintain the client's focus on goals (Grant, 2017). However, coaching can have unwanted side-effects, and in extreme cases, lead to dropout, necessitating side-effect management for a balanced framework. "Research on client dropout in business coaching is scarce even though dropouts can have consequences for clients, coaches, organizations, and the validity of coaching research" (Schermuly, 2018). The same study concluded that the four variables influencing the dropout are the following: change motivation of client, coach-client relationship quality, client neuroticism, unwanted side-effects of clients. There was a mismatch identified between coach and client, questionable coach expertise, employee inability to change, and unmeasurable return-on-investment as top reasons for early contract termination (Tompson et al., 2008).

Regarding the 'sponsor's influence and experience' aspect, a crucial moment related study (De Haan & Nieß, 2015) found that sponsors primarily evaluate coaching success based on return-on-investment and visible change, such as client behavioural changes and communication improvements. Therefore, the primary focus areas for this role include client behavioural changes, improvements in client communication, and the return-on-investment aspect. Valentine adds the interest of gathering insights to improve future developmental environments (Valentine, 2019).

The 'stakeholder's influence and experience' category includes commissioners and accreditation bodies, as identified in Section 4.2. While accreditation bodies play a crucial role

in shaping the coaching industry, their influence is indirect, and research on their interests is lacking. Therefore, their needs are not considered for the initial COBAC framework. On the other hand, the commissioner's responsibility is related to administrative tasks associated with the coaching process, such as initial coach selection, contracting, management, and evaluation of the quality and impact of coaching, hence any difficulties that relate to administration is of a high importance for this role.

A social network analysis based study has concluded regarding the category of 'distal system impact on coaching' that "the closer any member of the network was identified as being connected to those who received coaching, the more likely they were to experience positive increases in wellbeing" (O'Connor & Cavanagh, 2013). This study indicates that even though the focus of most stakeholders is related to the client's development, it has a 'ripple effect' in the network of the client. However, due to methodological challenges of the network analysis, this topic is not pursued within the automation context further.

4.4.3 Autonomous Coaching

As discussed in Sections 2.3 and 2.4, there are numerous unexploited opportunities to harness the benefits of automation, including cost reduction, scalability, location independence, service availability, and service consistency. While these factors can be framed as challenges, it is imperative that they are addressed individually during the value proposition design stage of framework building.

Additionally, Section 2.3.3 sheds light on the following key issues regarding the limitation of autonomous coaching systems: unclear communication on the purpose of the solutions; lack of clarity on the conceptual basis, coach competencies and scientific evidence underpinning the solutions; and lack of transparency on what the applications designed to do exactly and why.

Owing to the scarcity of reliable studies, a comprehensive analysis of the micro-context of autonomous coaching would not yield conclusive results for the initial stage of framework

building. However, it is likely to form a crucial foundation for the future development of the COBAC framework and related AI coaching systems.

4.4.4 Outcome

Table 4.3: Initial Problem List Derived from the Micro Context Analysis

Key	Problem Summary	Primary Stakeholder	Source
P.48	client's goal: pitfalls - inappropriate difficulty level, avoidance of generic goals, lack of commitment due to importance or attainability, level of self-efficacy, response to failure and task strategy, lack of feedback to show progress, lack of confidence, avoiding wrong direction, insufficient effort, inconstant actions, poor action planning quality due to lack of knowledge or high pressure	User	Micro MPM
P.49	client's goal: poor goal change management	User	Micro MPM
P.50	coach attributes: lack of key attributes - credibility, empathy, respect, holding the professional self, diagnostic skill and insight, approach flexibility and range, working to the business context, a philosophy of personal responsibility, and skilful challenging, high coach expectation	Vendor	Micro MPM
P.51	coach-client relationship: lack of variety - supportive and challenging	Vendor	Micro MPM
P.52	coach-client relationship: keep the client's focus on the goal	Vendor	Micro MPM
P.53	coach-client relationship: poor side-effect management	Vendor	Micro MPM
P.54	coach-client relationship: poor dropout management potentially due to change motivation of client, coach-client relationship quality, client neuroticism, unwanted side-effects of clients, mismatch between coach and client, questionable expertise of coach, inability of employee to change and return-on-invest not easily measurable	Vendor	Micro MPM
P.55	sponsor: lack of client visible change in behaviour and communication	User	Micro MPM
P.56	sponsor: poor return-on-invest	User	Micro MPM
P.57	sponsor: lack of information on how to improve future coaching engagements	User	Micro MPM
P.58	stakeholder's influence and experience: administration difficulties related to coach selection, contracting, management, and evaluation of the quality and impact of coaching	User	Micro MPM
P.59	unexploited opportunities to utilise the benefits of automation, such as cost, scalability, location independence, service availability, and service consistency	User	Micro Automation
P.60	automation limitation: unclear communication on the purpose of the solutions	User	Micro Automation
P.61	automation limitation: lack of clarity on the conceptual basis	User	Micro Automation
P.62	automation limitation: lack of information on the coach competencies and scientific evidence	User	Micro Automation
P.63	automation limitation: lack of transparency on what the applications designed to do exactly and why	User	Micro Automation

Table 4.3 contains the outcome of the Micro Context Analysis based on the structure of the Problem List used for presenting the outcome of the macro context analysis in Section 4.3.9 using the simplified primary stakeholder concept.

The outcomes of the Macro Context Analysis and the Micro Context Analysis are presented in the form of the Initial Problem List. The items represent problems or needs attributed, in the simplified context, to either the User or the Vendor stakeholders. However, a weight cannot be derived from the literature at this stage of the process, as there is practically no data to suggest the prioritisation of the items.

4.5 Value Proposition Design

Following the exploration of the problem space, the subsequent phase in framework building involves analysing the specific and detailed boundaries and objectives of the research topic. This analysis provides the foundation for the solution requirements in the final stage of the COBAC framework building.

The value proposition "describes the benefits customers can expect from your products and services" (Osterwalder et al., 2014). The value proposition design is a precursor to setting explicit objectives for any IT system. Thus, before establishing the objectives, it is deemed best practice to define what the system aims to offer, drawing upon the client characteristics outlined in the preceding chapters.

4.5.1 Approach

The selected approach is based on the Value Proposition Canvas, a widely utilised tool in the process of constructing consistent value propositions for products and services (Osterwalder et al., 2014). The authors propose a three-step process to establish the foundation of a business centred on a service or product: the Environment Map, Value Proposition Canvas, and Business Model Canvas (Osterwalder & Pigneur, 2010). The aim of this thesis is to develop a theoretical framework rather than establish a business; therefore, the distinctly business-oriented methodologies applied in the Environment Map and Business Model Canvas are not

incorporated in this thesis, notwithstanding their value and acclaim within the start-up sphere. Moreover, of the three types of fitness (problem-solution fit, market fit, and business model fit), the emphasis is exclusively on the first for analogous reasons.

Two primary approaches exist concerning the initial point and orientation of value proposition exploration: the "technology push", where the inception of the investigation lies in technological innovation, and the "market-pull" approach, commencing the value proposition exploration with the customer profile. The customer segment profile analysis in this thesis employs both methodologies to formulate a cohesive hypothetical perspective on the probable value proposition, considering both the potential of the technology and the insights into the current coaching client segment.

The foundational data for the value proposition design encompasses insights discovered in the literature review on the benefits of coaching, autonomous coaching, and LLM-based systems, as well as the problem analysis results presented in Section 4.3 and 4.4.

The adopted terminology of customer jobs, gains, pains, pain relievers, and gain creators serve as the foundational pillars of the design process outcome, the Value Proposition Canvas. The subsequent steps are undertaken to achieve this outcome:

1. Customer Segment Profile:
    a. The coaching context-related Customer Segment Profile is formulated based on the prospective coaching advantages pinpointed in the literature review and the coaching-associated user challenges identified in the problem analysis.
    b. Modifications to the profile are recognised considering the literature review and problem analysis related to automation and LLM technology.
    c. Integration of all the information into the Customer Segment Profile for an LLM-based autonomous coaching service.
2. The Value (Proposition) Map is structured around the automation and LLM-based competencies, addressing the principal elements identified in the Customer Segment Profile.
3. The Value Proposition Canvas is compiled as the combination of the Customer Segment Profile and the Value (Proposition) Map.

4.5.2 Customer Segment Profile

In the initial phase of the value proposition design process, data from Section 2.2.8 was collected, providing a concise summary of the benefits of coaching. From an individual client's viewpoint, coaching offers multiple advantages, addressing a myriad of personalised objectives ranging from situational support during transitions and specific skill enhancement to wider influences on performance. Besides achieving objectives, coaching was demonstrated to enhance coping strategies (resilience, mindfulness, self-enhancing attributions), well-being, self-regulation, and attitude.

The context analysis in Section 4.3 and 4.4 summarised the major client limitations of coaching, excluding the situational, process-related, automation, social, and ethical implications as follows: perception of lack of scientific evidence (P.3), high direct client cost (P.18), high client opportunity cost (P.19), risk to service quality for clients due to the unregulated nature of the coaching industry (P.26), and administrative challenges (P.58).

Figure 4.3 provides a graphic representation of the gains, challenges, and responsibilities linked with a coaching client profile, following the terminology and visual vocabulary of the Customer Segment Profile of the Value Proposition Canvas (Osterwalder et al., 2014).

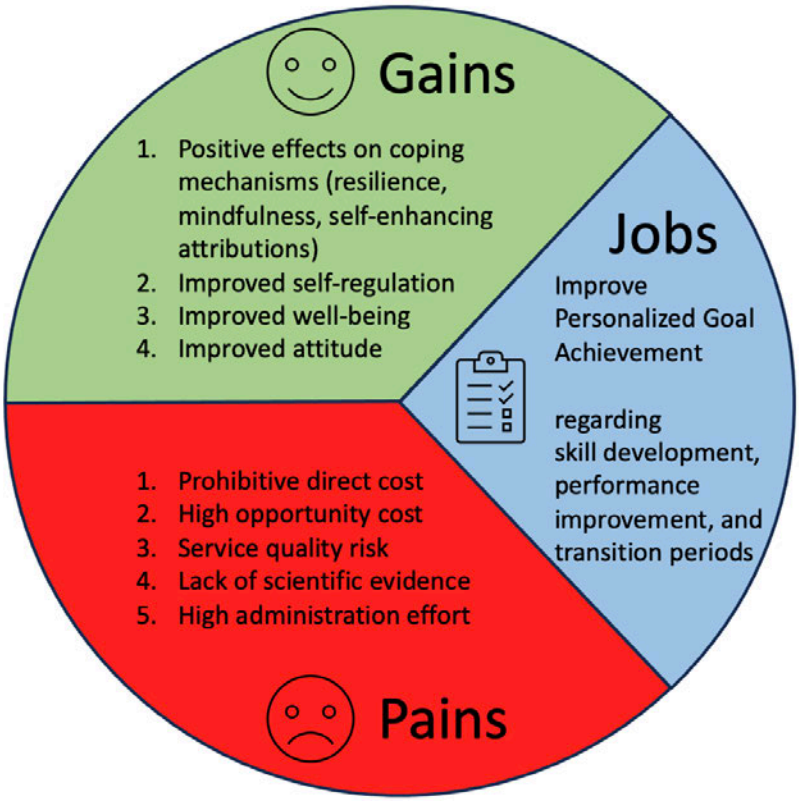


Figure 4.3: Customer Segment Profile - Coaching

Although the literature review offered insights into relevant coaching clients concerns, but it did not indicate their importance. It is also impossible to characterise them based solely on the existing literature review without a more comprehensive client analysis. However, this level of detail is adequate for initiating the framework construction, and future information collection iterations will likely modify the value proposition model based on actual customer insights.

Subsequently, Section 2.3.5 explored the advantages of autonomous coaching, concluding that from a user's perspective, autonomous coaching presents numerous benefits, including significant cost reductions, scalability, and enhancements in service quality and processes. Users can engage in a non-judgemental coaching environment, enjoying the independence, albeit with subpar dialogue quality with non-LLM based technologies.

The context analysis in Section 4.3 and 4.4 concerning the primary client issues of autonomous coaching, excluding the situational, process-related, social, and ethical

implications, can be summarised as follows: software usability risk (P.5), flexibility (P.6, P.7), independence (P.8), social connectedness (P.9), cost (P.18), opportunity cost (P.19), transparency (P.24, P.29), physiological and psychological risk (P.36), and over-reliance on automated systems (P.37).

Similarly, Section 2.4.5 enumerated the advantages for users of integrating LLM into autonomous coaching solutions. The LLM technology simulates human behaviour in learning, reasoning, and sentiment analysis, directly enhancing essential coaching competences. Moreover, it overcomes previous chatbot limitations by offering superior context recognition and text generation capabilities, culminating in a richer and superior dialogue experience. The progress in language support reduces the bias towards the English language, broadening accessibility and allowing users to communicate with the AI coaching application in their native tongue. In essence, LLM-based autonomous coaching markedly improves core coaching skills, dialogue quality, and supports multiple languages. In relation to the deployment of LLM-based technology, privacy risk (P.42) appears to be the predominant concern that should be incorporated into the customer segment profile.

Figure 4.4 captures the modifications stemming from these two supplementary contexts: automation and LLM-technology enablement.

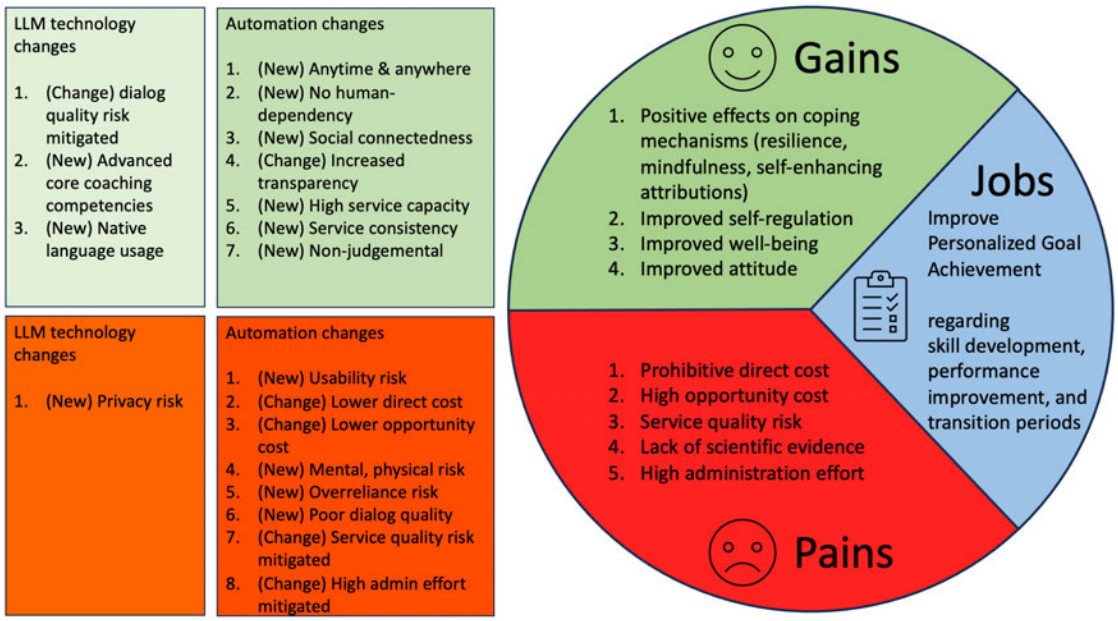


Figure 4.4: Customer Segment Profile - Changes Introduced by Automation and LLM

The terms "New" and "Change" prefixed to the descriptions in the boxes on the left side specify if the element is a new addition to the model due to the inclusion of new contexts or if an existing component was modified by the introduction. For example, Pain #5 "High administration effort" was mitigated by automation.

Figure 4.5 presents a comprehensive perspective after integrating all the identified modifications, portraying a potential customer segment profile for those utilising an autonomous coaching system.

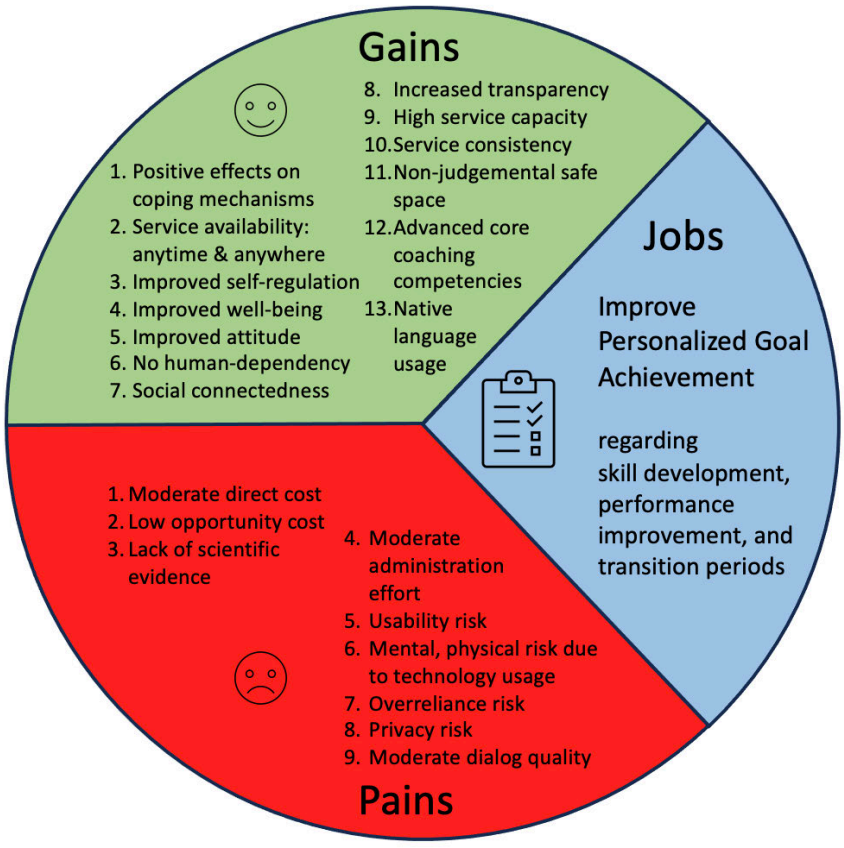


Figure 4.5: Customer Segment Profile - LLM-based Autonomous Coaching

4.5.3 Value (Proposition) Map

The second component of the Value (Proposition) Map is associated with the decomposition of the value proposition into offered products or services, gain creators, and pain relievers. Based on the analysis from previous chapters, an LLM-based AI coaching application

constructed on evidence-based coaching competencies and adhering to the highest ethical standards can effectively enhance the client's personalised goal attainment (highlighted in blue). Gain creators, such as coaching competencies, cloud infrastructure, community platform, reporting, data-driven methodology, and LLM technology, each contribute to the gains depicted in Figure 4.6 (green area). Similarly, aspects like cost-efficiency, service flexibility, application of validated scientific theories, digital administrative processes, continuous improvements in usability features, promotion of user autonomy, and adherence to security standards are integral to addressing the pain points identified in the literature (highlighted in red).

Figure 4.6 presents an integrated representation of these elements using the visual language of the Value (Proposition) Map.

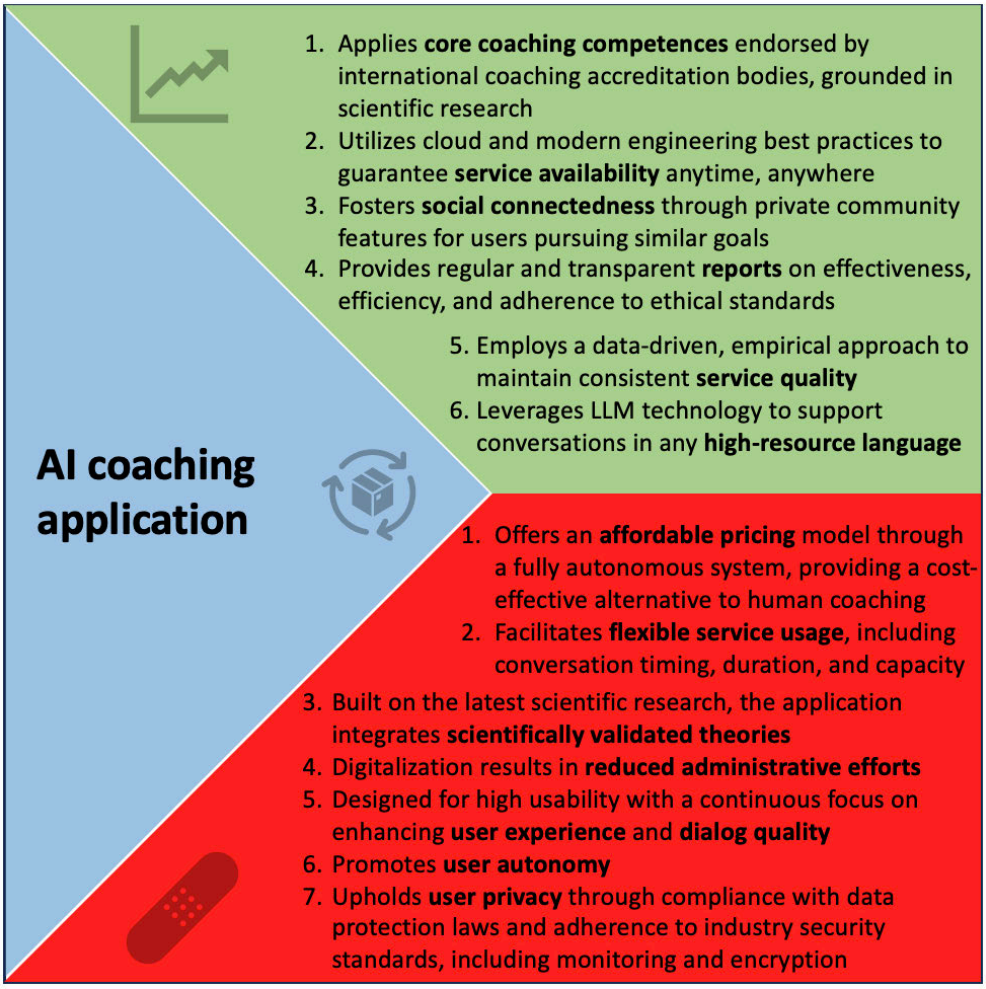


Figure 4.6: Value (Proposition) Map - LLM-based Autonomous Coaching

4.5.4 Outcome

In summary, the result of the value proposition design integrates both Figure 4.5 and Figure 4.6, which together formulate the content of the Value Proposition Canvas related to LLM-based autonomous coaching from a user's perspective.

4.6 Objective Analysis

The Value (Proposition) Map (Figure 4.6) reveals the potential objectives that an effective autonomous coaching system should accomplish. According to the Object Management Group (Object Management Group (OMG), 2015), objectives lay the foundation for measuring progress towards addressing the problem. Thus, a major additional element to formulate objectives are related to the discovery of how to measure solving the addressed problems.

4.6.1 Approach

The objectives stem from the Value Proposition Map. The following steps are undertaken to synthesise the Initial Objective List, which is the result of the objective analysis phase:

1. Pinpointing the primary objectives derived from the Value (Proposition) Map.
2. Objectives are put into context of items listed in the Initial Problem List.
3. Ascertaining any unresolved issues.
4. Outlining potential metrics for each objective.

To ensure traceability, the conclusion of the objective analysis comprises a list of clearly defined, concisely formulated objectives accompanied by the subsequent attributes:

1. Key: Values commence with the "O." prefix, mirroring the method used in problem identification, followed by sequential numbering.
2. Problem Reference: Enumerates the related keys from the Initial Problem List corresponding to each objective.
3. Measurement: A compilation of potential approaches to measure each objective.

4.6.2 Objectives

The Value Proposition Map (Figure 4.6) lists, in the blue segment titled "products and services," the AI coaching application as a digital product addressing the task of "improving personalised goal achievement." Consequently, the primary objective is maximising goal attainment (O.1).

Additionally, within the green segment of gain creators, the following elements are presented succinctly:

- Applies core coaching competences (O.2).
- High availability (O.3).
- Fosters social connectedness through private community features for users pursuing similar goals (O.4).
- Regular and transparent reporting on effectiveness, efficiency, and adherence to ethical standards (O.5).
- Service consistency continuous improvement (O.6).
- High-resource language enablement (O.7).

Meanwhile, in the red segment representing pain relievers, the following items are enumerated:

- Minimising direct cost (O.8)
- Service usage flexibility: conversation timing, duration, and capacity (O.9)
- Continuous integration of scientifically validated theories, concepts (O.10)
- Minimizing administrative efforts (O.11)
- User experience continuous improvement (O.12)
- Dialog quality continuous improvement with the most effective technology (O.13)
- User autonomy fostering (O.14)
- User privacy upholding (O.15)

The primary objective of goal attainment, paired with the mentioned gain creators and pain relievers, forms the initial objectives designated for the COBAC framework. These objectives are neither mutually exclusive nor ranked in order of priority, and they may occasionally be at odds. For instance, enhancing dialogue quality (O.13) might necessitate multiple requests to the LLM service provider, increasing operational costs (O.8). Balancing these conflicting objectives will be at the discretion of the system designers.

4.6.3 Objective-Problem Relationship

After finalising the objectives, the next step is to link defined problems to the list of objectives, which is summarized in Table 4.4.

Table 4.4: Objective-Problem Relationship

Key	Problem Reference	Objective Summary
O.1	P.48	maximising goal attainment
O.2	P.11, P.17, P.21, P.28, P.30, P.31, P.40, P.49, P.50, P.51, P.52, P.53, P.54, P.55, P.56, P.57, P.60, P.63	applies core coaching competences
O.3	P.6	high availability
O.4	P.9	fosters social connectedness through private community features for users pursuing similar goals
O.5	P.24, P.29, P.32, P.34, P.38, P.45	regular and transparent reporting on effectiveness, efficiency, and adherence to ethical standards
O.6	P.26	service consistency continuous improvement
O.7	P.47	high-resource language enablement
O.8	P.18, P.23, P.46	minimising direct cost
O.9	P.7, P.8, P.19, P.36	service usage flexibility: conversation timing, duration, and capacity
O.10	P.3, P.41, P.61, P.62	continuous integration of scientifically validated theories, concepts
O.11	P.58	minimizing administrative efforts
O.12	P.5	user experience continuous improvement
O.13	P.14, P.15, P.16	dialog quality continuous improvement with the most effective technology
O.14	P.37	user autonomy fostering
O.15	P.27, P.35, P.42, P.43	user privacy upholding

Maximising goal attainment (O.1) directly addresses potential pitfalls associated with client goals (P.48). However, the emphasis on core coaching competencies (O.2) intersects with numerous challenges, including adherence to ethical codes (P.28), micro-contextual elements of coaching (P.50 to P.57), and those of automation (P.60, P.63).

The proposed AI coaching system's high availability (O.3) tackles the issue of unmatched time-related availability expectations (P.6). The feature promoting social connectedness (O.4) counters the absence of such functionality (P.9). Reporting (O.5) aims to shed light on multiple facets such as ethical dilemmas and biases (P.24, P.29, P.32, P.34, P.38, P.45). Continuous

service quality enhancement (O.6) addresses risks posed by an unregulated coaching sector (P.26), while high-resource language enablement (O.7) seeks to contribute on solving the imbalance caused by English language bias (P.47).

In terms of pain relievers, the objective to reduce costs (O.8) correlates with several problem facets, from service affordability to environmental impact due to resource-intensive NLP systems (P.18, P.23, P.46). Service usage flexibility (O.9) aligns with issues like location dependence and mental risks (P.7, P.8, P.19, P.36). Infusing the system with scientifically grounded theories and concepts (O.10) seems to address multiple perspectives, for example the top user's barrier to use the service due to the perception of lack of scientific evidence, clarity of the conceptual basis of the solution, and research related concerns (P.3, P.41, P.61, P.62). While reducing administrative burdens (O.11) is paramount in an organisational context, it remains a concern for individual users in the private context. Efforts to constantly elevate user experience (O.12) respond to heightened user expectations (P.5). Striving for the highest dialog quality (O.13) is a side-effect of the poor dialog quality characteristics of the early technologies applied in the field and tries to address the high complexity involved in human interactions (P.14, P.15, P.16). Fostering user autonomy (O.14) addresses the pitfalls of over-relying on automation (P.37) but also aligns with modern coaching principles. Lastly, upholding user privacy (O.15) underscores the importance of data protection, client confidentiality, and research challenges (P.27, P.35, P.42, P.43).

The outcome of the analysis, the Initial Objective List, enumerates all objectives, and all the associations with the described problems. However, it omits unrelated items. Table 4.5 lists the items from the Initial Problem List, which are not related to objectives, hence remaining unaddressed by the COBAC framework.

Table 4.5: Problems Not Covered by Objectives

Key	Problem Summary	Reason for exclusion
P.1	poor awareness of coaching among coaching clients	This is a market condition, not addressable within the context of coaching.
P.2	barrier: coaching is perceived irrelevant	This is a market condition, more specialized offerings that restricts the goal-attainment objective to a specific field, such as time management skill improvement or well-being improvement, might contribute on addressing the issue.
P.4	coaching goal diversity due to the shifting coaching target segment	This is a market condition, not addressable within the context of coaching.
P.10	unexploited opportunity of increased adaption speed to new technologies	This is a market condition, not addressable within the context of coaching.
P.12	low-tech nature	One of the aims of the thesis to contribute on this problem, but the COBAC framework itself cannot address this issue.
P.13	absence of adaptable autonomous models	One of the aims of the thesis to contribute on this problem, but the COBAC framework itself cannot address this issue.
P.20	coaching reinforces economic inequality	One of the aims of the thesis to contribute on this problem, but the COBAC framework itself cannot address this issue.
P.22	financial incentives do not encourage solving social problems	This is a market condition, not addressable within the context of coaching.
P.25	lack of political discourse with voters on AI and automation	This is a market condition, not addressable within the context of coaching.
P.33	AI coaching system: absence of executive support of ethical concerns	The issue is an organizational issue, which is only addressable in the business context, not in the problem-solution fitness level, which is the focus of the thesis.
P.39	social issue: job displacement	This is a socio-political issue that requires AI governance posed by regulations, hence not addressable within the context of the thesis.
P.44	social issue: LLM copyright	This is a market condition, not addressable within the context of coaching.
P.59	unexploited opportunities to utilise the benefits of automation, such as cost, scalability, location independence, service availability, and service consistency	The problem is defined too broadly, but the COBAC framework incorporates these items on an individual opportunity level.

4.6.4 Measurement Approach

The objectives in Section 4.6.2 vary significantly. Some can be easily measured using standardised and validated coaching-related measures, while others necessitate continuous technical diagnostics and monitoring. For each objective, at least one, and in some instances,

multiple approaches were proposed. Table 4.6 summarises the measurement approaches suggested from the literature review in Chapter 2 for monitoring the progress of achieving the COBAC objectives.

Table 4.6: Measurement Approach Overview

Measurement Approach	**Description**
GAS - Goal Attainment Scaling (Kiresuk & Sherman, 1968)	A standardized and validated measure in the coaching domain to measure degree of goal attainment.
CE - Coaching Effectiveness (De Haan et al., 2016)	It uses a 4-point scale and assesses the created safe space, learnings, behaviour change, and overall outcome associated with a coaching engagement.
USR - UnSupervised and Reference-free metric (Mehri & Eskenazi, 2020)	It serves as a measurement for dialog quality.
UTAUT (Venkatesh et al., 2003)	Measuring various factors of user adaptation related to technology.
Feature availability	It is a binary measure of the functionality availability.
Monitoring	Monitoring is an IT-industry standard diagnostical approach to continuously measure characteristics of an operational system.
CSAT - Customer satisfaction	IT industry standard approach to measure customer satisfaction of the overall service experience.
External assessment	Expert review to ensure a characteristic is met by the system.

4.6.5 Outcome

The COBAC objectives were summarised in this section based on the previous detailed objective identification, how they relate to the items of the Initial Problem List, and the appropriate measurement approaches for them. Table 4.7 contains a unique key, a list of problem references, measurement approaches, and a summary for each initial objective.

Table 4.7: Initial Objective List

Key	Problem Reference	Measurement	Objective Summary
O.1	P.48	GAS, UTAUT performance expectancy (PE)	maximising goal attainment
O.2	P.11, P.17, P.21, P.28, P.30, P.31, P.40, P.49, P.50, P.51, P.52, P.53, P.54, P.55, P.56, P.57, P.60, P.63	feature availability, CE	applies core coaching competences
O.3	P.6	monitoring	high availability
O.4	P.9	UTAUT social influence (SI)	fosters social connectedness through private community features for users pursuing similar goals
O.5	P.24, P.29, P.32, P.34, P.38, P.45	feature availability, CE, monitoring	regular and transparent reporting on effectiveness, efficiency, and adherence to ethical standards
O.6	P.26	monitoring	service consistency continuous improvement
O.7	P.47	feature availability, USR	high-resource language enablement
O.8	P.18, P.23, P.46	CSAT	minimising direct cost
O.9	P.7, P.8, P.19, P.36	feature availability, CE	service usage flexibility: conversation timing, duration, and capacity
O.10	P.3, P.41, P.61, P.62	external assessment	continuous integration of scientifically validated theories, concepts
O.11	P.58	CSAT	minimizing administrative efforts
O.12	P.5	UTAUT effort expectancy (EE)	user experience continuous improvement
O.13	P.14, P.15, P.16	USR, monitoring	dialog quality continuous improvement with the most effective technology
O.14	P.37	feature availability, CE	user autonomy fostering
O.15	P.27, P.35, P.42, P.43	UTAUT perceived risk (PR)	user privacy upholding

4.7 Constraint Analysis

After clarifying the objectives, it is essential to explicitly define the boundaries of the COBAC framework. In this thesis, constraints serve three purposes: maintaining focus, ensuring transparency of assumptions, and facilitating the framework's adaptability. Constraints related to problems, objectives, and solution requirements are specified to make feasibility considerations both transparent and explicit. Given that feasibility constraints change rapidly alongside the exponential development in technology-related research, the model can be easily adjusted should any of the underlying constraints become invalid.

4.7.1 Approach

To identify and make constraints explicit, an analysis of the preceding chapters is essential. This involves first reviewing the thematic literature in Chapter 2, then examining the characteristics identified in the Stakeholder List, followed by the Initial Problem List. The last component of the COBAC framework, the solution requirements, was not incorporated into the constraint analysis yet.

The output of the constraint analysis, termed the Initial Constraint List, shares a similarity with the Initial Objective List in that it has a unique key to facilitate the mapping of constraints for traceability. Each identified constraint is assigned a key that follows the naming pattern of C.<n>, where 'C' denotes constraint.

4.7.2 Constraints

The literature review in Chapter 2 was approached with a focus on coaching, autonomous coaching, and LLM-based autonomous coaching. As the aim of the COBAC framework is to achieve full automation in coaching, this introduces a specific boundary summarised as follows:

- Full automation, in this context, signifies the goal to entirely replace the coach with an autonomous system. This contrasts with most academic studies and industry stakeholders that advocate for a hybrid model.
- Although team and group coaching are an emerging and rapidly evolving domain, it was deemed beyond the scope of this study due to its inherent complexities.

Furthermore, constraints identified from the stakeholder analysis include:

- The assumption that the client is an adult, simplifying the modelling process by excluding demographics requiring special consideration, such as children.
- The client does not suffer from any mental disorder.
- Contrary to market research, the coaching process is assumed not to occur within an organisational context. The client is believed to singularly represent all interests, which in an organisational setting may be dispersed among multiple individuals, such as a sponsor or a coach commissioner. Representing interests of indirect stakeholders like researchers, training, or accrediting institutes are excluded from the initial framework formulation. However, future research should encompass and integrate findings related to these stakeholders into subsequent framework versions.

Additionally, the macro-context analysis highlighted several factors relevant to the coaching industry. However, to minimise the initial model's complexity, these were not incorporated. Deliberate attempts to influence the client's broader context, including social networks, political dimensions, or economic-commercial interests, were not accounted for beyond the already specified objectives. Notwithstanding the numerous commercially and legally relevant guidelines that may shape a coach's behaviour during a coaching relationship - such as pertinent laws or tax regulations - these are not unique to the coaching process and, as such, are not the central focus of this study. The sole directive intrinsic to the coaching industry is the coaching-specific ethical code, integrated into the objectives via the application of core coaching competences (O.2). The effects on the client's immediate social circle, including family and close friends, are also regarded as outside the study's scope, except for the aim of fostering social connectedness (O.4).

In relation to the objectives, the analysis presumes that only the core coaching competences, as defined by coaching competency frameworks for a non-directive, reflective coaching style, are under consideration. This excludes supporting business competencies such as marketing or sales. The sole deviation is the administrative process, as it significantly overlaps with some core coaching competences. Other approaches, including counselling, teaching, mentoring, or training, were also determined to be beyond the scope.

4.7.3 Outcome

Table 4.8 enumerates the constraints identified from the literature review, stakeholder analysis, micro- and macro-context analysis, and objective analysis.

Table 4.8: Initial Constraint List

Key	Reference	Constraint Summary
C.1		Focus on full automation to replace coach with autonomous system, not on a hybrid model
C.2		Team and group coaching is out of scope
C.3	S.1	Client assumed to be an adult individual
C.4	S.1	Client without any mental disorder
C.5	S.1	Coaching is not conducted in an organisational context
C.6	S2, S.3, S.4, S.5, S6, S.7, S.8	Client represents all interests alone, interests of other stakeholders assumed to be represented by the client
C.7		Influencing broader context (social network, political, economic-commercial interests) not considered beside the identified specific objectives
C.8		Country-specific directives like laws, or taxation are not the focus, only the coaching-specific ethical code is considered
C.9		Impact on closer network of the client (e.g., family, friends) is out of scope, beyond the social connectedness feature
C.10	O.2	Only core coaching competences from modern coaching competency frameworks that define a non-directive, reflective style of coaching is considered

4.8 Chapter Summary

Section 4.2 provided an in-depth analysis of the stakeholders involved in both human-to-human coaching relationships and autonomous coaching processes, emphasizing the importance of understanding the various players and their characteristics. It categorized stakeholders into primary roles such as coach, client, sponsor, and commissioner, and considers the impact of accrediting institutes, training organizations, and researchers. The section further explored the transition of these roles in the context of autonomous coaching systems, highlighting the changes required and the new stakeholders introduced in this evolving domain.

Section 4.3 provided a comprehensive macro context analysis of AI coaching using the STEEPLE model, covering social, technological, economic, environmental, political, legal, and

ethical factors. It underscored the complexity and interrelatedness of these factors, highlighting the challenges and opportunities for the autonomous coaching industry.

Section 4.4 delved into the micro-context analysis of coaching, autonomous coaching systems, and LLM-powered AI coaching systems, acknowledging the scarcity of data available in these areas. The section aimed to provide initial insights for developing the COBAC framework, focusing on three main aspects: the existing human-to-human coaching micro-context, the autonomous coaching micro-context, and the researcher perspective. The analysis of the coaching micro-context is based on Grant's Multiple-Perspective Model (MPM), covering ten areas, and highlights the need for continuous adjustment of the COBAC framework and AI coaching systems as understanding of the micro-context evolves and the micro-context changes over time.

The value proposition design in Section 4.5 describes the benefits customers can expect from products and services and is crucial before setting objectives for an IT system. The approach employed is based on the components of the Value Proposition Canvas with an emphasis on the problem-solution fitness. The outcome of the design integrates the Customer Segment Profile and the Value (Proposition) Map to form the Value Proposition Canvas for LLM-based autonomous coaching.

Section 4.6 describes the process and results of an objective analysis for an effective autonomous coaching system based on the Value (Proposition) Map. Objectives are formulated by pinpointing primary objectives from the Value (Proposition) Map, contextualizing them with the Initial Problem List, identifying unresolved issues, and outlining potential measurement approaches. The approach to measure each objective varies, with some using standard coaching measures and others requiring continuous technical monitoring.

Lastly, Section 4.7 details the boundaries and assumptions of the COBAC framework. Constraints help maintain the framework's focus, ensure transparency, and aid its adaptability. The analysis uses literature review, the Stakeholder List, and the Initial Problem List to develop an Initial Constraint List. Notable constraints include focusing solely on full

automation in coaching, excluding team and group coaching, and assuming clients are adults without mental disorders, not within an organizational context.

Although relationships among the discussed issues can be important for the coaching research and practitioners, the detailed and complete analysis and synthetization of these links were not pursued in this study, as it offers no value for the initial stage of the automation. Capturing most of the underlying issues was sufficient to build the initial framework for coaching automation because subsequent observations of the working system can offer much more reliable and evidence-based insights to further develop the COBAC framework.

The following chapter delves into the process of gathering solution requirements, building upon the constraints and objectives previously discussed to shape the actionable features of the COBAC framework.

5. New Framework Building: Solution Requirement List

5.1 Introduction

The preceding chapters have delineated the stakeholders involved in the coaching context, identified problems, articulated the objectives that an autonomous system must fulfil to address these problems, and defined constraints within the problem-solution space. This chapter delves into how these objectives may be attained through an in-depth analysis of common coaching themes, data modelling, process modelling, and the synthesis of solution requirements.

5.2 Coaching Common Theme Analysis

The coaching common themes answer the question of what is meant by coaching in a more detailed way that is required for the realization of autonomous coaching systems. This analysis transcends simplistic definitions of coaching by identifying thematic clusters within human-to-human coaching dialogues that encapsulate the concept of coaching. This method underpins the initial COBAC framework and, consequently, the development of the first service prototype. Nevertheless, as the volume of empirical data collected from an operational system and client-system interactions expands, there is scope to transition from a document-centric framework building strategy to an empirical one, predicated on data amassed during system operation.

5.2.1 Approach

The text-based analysis is grounded in the following documents:

a) The ICF Core Competency Model (ICF, 2019a)
b) The ICF Code of Ethics (ICF, 2019b)
c) The EMCC Competence Framework (EMCC, 2015)
d) The EMCC Competence Framework Glossary (EMCC, 2018)

e) The Global Code of Ethics for coaches, mentors, and supervisors (Association for Coaching & European Mentoring & Coaching Council, 2019)

Item a) is the result of a large-scale study (ICF, 2020b) and encapsulates demonstrative behaviours of coaching practice. Item (b) provides guidance on potential ethical issues associated with coaching. The relationship between item a) and b) is that the first includes reference to the second and states that adherence to the ethical code is a core competence. Moreover, item b) includes a glossary for the basic terms utilised in both documents. Thus, items a) and b) can be considered one complete set of information, which includes coaching competences, ethical code, and defining terminology.

Items c), d), and e) encapsulate similar functions within the EMCC domain: term definitions, competences, and ethical rules. Item e) is referred to as 'global' because, at the time of this thesis's composition, eight major global coaching organisations, excluding the ICF, had endorsed the document. A recommendation emerged from the Global Coaching Convention in 2008 to unify the competences and ethical codes of coaching related organisations (Stout-Rostron, 2009), but it remains unrealised, which makes the scientific investigation significantly more complex.

The aforementioned documents are also instrumental in the coach accreditation process. For example, unlike item c), item a) solely encompasses demonstrative behaviours post section 'A' that are observable during a coaching dialogue, as this document underpins not only coaching competence development training but also the accreditation process for evaluating coaches.

Several rationales underscore the exclusive consideration of artefacts for analysis from the leading coaching organisations, ICF and EMCC:

1. The coaching core competency frameworks are predicated on scientific research, albeit based on secondary sources.
2. By utilising a competency-based approach, they describe desired coaching behaviour, making them comparable.

3. They represent two segments of coaching with the highest number of practitioners. Consequently, the analysis is streamlined by focusing solely on these organisations, which collectively embody most global coaching practitioners.

The coaching common theme analysis is conducted based on the following process:

i. Combining c) and d) into an EMCC theme list and assigning codes for theme identification.
ii. Merging a) and b) into an ICF theme list and assigning codes for theme identification.
iii. Identifying coaching common themes based on the assigned codes between the ICF theme list and the EMCC theme list.
iv. Uncommon items are acknowledged as good coaching practice but not as defining elements of coaching and are thus excluded from initial COBAC framework building.
v. Common coaching themes are grouped by original competence categories or theme origin.
vi. If reframing is necessitated, the common coaching themes are reframed to suit the context of automated coaching; alternatively, if a theme is deemed irrelevant for autonomous coaching, it is omitted.

The first three steps are a classic example of a textual, thematic analysis employing an inductive approach based on assigned codes (V. Braun & Clarke, 2006). The succeeding three steps facilitate the transformation of identified common theme groups into solution requirements in an automated coaching context in subsequent phases.

Table 5.1: Document Key Naming Patterns

Sample Key	**Pattern description**
ICF_CC_A.1.1	"ICF_CC" prefix refers to items in a) with the following naming pattern: ICF_CC_<section_letter>.<section_number>.<sub-point>
ICoE.E.1	"ICoE.E" prefix refers to items in b) with the following naming pattern: ICoE.E.<numbering_in_section_4>
EMCC_CF_001	"EMCC_CF" prefix refers to items in c) with the following naming pattern: EMCC_CF_.<numbering_in_parenthes_after_competence_descriptions>
GCoE.4.3	"GCoE" prefix refers to items in e) with the following naming pattern: GCoE.<section_number>.<sub-point>

Table 5.1 provides an overview of the key naming patterns used for the analysed documents to support the traceability requirement of the COBAC framework. The above patterns are used to assign a unique key to each item in the listed documents to provide a basis for traceability during the textual analysis.

5.2.2 Coaching Common Themes

In alignment with the naming convention detailed previously, each element within documents c) and e) was allocated a set of codes for theme identification. These codes, either single or hyphenated words such as "confidentiality," "goal-setting", "raise-awareness," or "client-emotion," are derived from the terminology used within the text. Appendix A details the specific codes assigned to each item, corresponding to the aggregated item list related to the EMCC documents.

This methodology was similarly employed to merge documents a) and b) into an ICF theme list, with codes assigned accordingly, except in cases where a code from the EMCC document already defined the same concept, in which instance the pre-existing EMCC code was applied. Appendix B presents a comprehensive breakdown of the codes linked to artefacts sourced from ICF.

The identification of prevalent coaching themes is based on a comparative analysis of the assigned codes. The presence of a code in both lists denotes a shared theme. In the absence of such correspondence, the theme is regarded as a best practice rather than a defining trait of coaching. Figure 5.1 exhibits the top eight codes from ICF and EMCC, ordered by frequency of occurrence, with confidentiality and contract-related activities being illustrative of shared themes.

This comparison also highlights intrinsic differences in concept and terminology. Notably, "feedback" ranks as the foremost entry in the EMCC list, yet it is completely absent from ICF documentation. Appendix C contains the complete list of common codes identified as part of the code comparison.

Row Labels	Count of ICF code list sorted with duplicates
confidentiality	6
contract-closure	5
client-context	4
client-experience	4
action-planning	4
contract-management	4
client-belief	4
support	4

Row Labels	Count of EMCC code list sorted with duplicates
feedback	8
interest-conflict-management	5
confidentiality	5
contract-management	5
client-autonomy	4
supervision	4
insight	4
contract-closure	4

Figure 5.1: Top 8 ICF and EMCC Code Counts

Despite the shared nature of these themes, variations in terminology are evident. The ICF terminology was adopted to reframe these themes, ensuring uniformity, and mitigating potential ambiguity amongst the items. For instance, "Establishes and Maintains Agreements," a section heading (ICF_CC_B.3) in ICF documents, corresponds to a competence category termed "managing the contract" within the EMCC competency framework. Descriptions of competence within this category showcase use-cases and exemplary behaviours related to the establishment, maintenance, alteration, and ultimately, the termination of agreements within the coaching relationship. Table 5.2 encapsulates the findings related to the 'agreement management' theme group, alongside the codes employed in the textual analysis, a unique identifier, and a succinct summary of each item.

Table 5.2: Coaching Common Theme Summary - Agreement Management

Key	Theme Group	Code	Common Theme Summary
CT.24	agreement management	contract-management	managing clear agreements
CT.25	agreement management	location	agreement on location of service
CT.26	agreement management	scheduling	agreement on session scheduling: date/time, frequency, duration
CT.27	agreement management	contract-closure-anytime	service termination conditions: any time indicated by the client but subject of agreement termination conditions
CT.28	agreement management	contract-closure	service termination conditions: client value gained from coaching is shifting
CT.29	agreement management	contract-closure-interest	service termination conditions: unresolvable interest conflict
CT.30	agreement management	competence-boundary	service termination conditions: coach competence boundary reached

Subsequent subsections describe the theme groups identified through this analytical process.

5.2.3 Theme Groups

Items associated with the theme group “code of ethics” has no relevance about coaching efficiency, this theme group is focused on what is right or wrong from an ethical point of view. Topics, such as confidentiality, interest conflict, referral to other professionals or continuous professional development belong to this area. According to the analysis abiding a coaching specific code of ethic is a core competence of coaching. Despite being an essential characteristic, ethical code compliance is not a demonstrative behaviour that can be observed during a coaching session directly, hence the related items are added under the “Engagement Setup”.

Similarly, items associated with the theme group "Communication Before Coaching" are included in the "Engagement Setup" category. Items such as explaining the role of coaching, the coaching process, potential benefits, and communicating the code of ethics must be clearly articulated before commencing the coaching process. In contrast to topics related to the Code of Ethics, these items are observable, demonstrable items.
Conversely, items related to the "Agreement Management" theme group describe ongoing activities within the coaching process. This theme group includes administrative arrangements such as scheduling, as well as conditions for terminating the relationship.
The "Coach Behaviour" theme group encompasses behaviours that coaches must exhibit in a coaching context. Central to the coaching process, examples include working with a client's emotions and actively fostering client autonomy.

Lastly, the "Partnering Behaviour" theme group encapsulates behaviours performed collaboratively by coach and client, such as establishing goals or planning actions. This theme group is also integral to the coaching process.

5.2.4 Outcome

Appendix D presents a comprehensive list of all common coaching themes, complete with associated codes, unique keys, and theme groups.

5.3 Data Modelling

Social interactions, including coaching, necessitate sophisticated data models about the communicants and their associated attributes such as emotions, knowledge, and context. Section 5.3 provides an analysis and a detailed description of the conceptual data model associated with COBAC.

5.3.1 Approach

An LLM-based autonomous coaching system utilises three data sources: system configuration, user input data, and LLM-generated information. While future versions might address the context length limitations, at the time of writing the thesis LLMs lacked memory and have limited context length for conversations; thus, they were unsuitable for storing client or conversational data. Consequently, each conversation was structured such that the AI coaching system supplies all relevant contextual information, with the LLM responsible for analysing input data and generating user messages. User input data shared through interactions, or their absence falls into two categories: content and metadata. Content refers to information explicitly provided by the user, while metadata encompasses information that, though not expressly exchanged, includes cues about the client or context. Metadata examples include application screen deactivation, possibly indicating distraction, or the duration of pauses between communication sequences, which may suggest contemplation. All information not originating from user interaction or LLM generation must be stored in the AI coaching system. Section 5.3 focuses on conceptualising a data model encompassing all vital data elements associated with coaching engagements.

To conceptualise the data model, elements were identified from common coaching themes: user, outcome, conversation, reporting, coach, and dialogue. After analysing these elements, a conceptual data model is distilled and presented.

5.3.2 Data Element: User

In the context of digital services, the user, central to coaching engagements, has numerous associated attributes, from accessing services to facilitating effective coaching conversations.

The attribute 'authentication' pertains to data structures essential for user identification. Privacy (CT.7), security (CT.8), and confidentiality (CT.31) are paramount in coaching, as they form the foundation to build trust for effective collaboration. Actual implementation data structures might include username and password combinations or biometric identifiers, such as fingerprint technology (Google, 2020), or external services, like Facebook credential logins (Facebook, 2020). Authentication-related attributes are mandatory for multiple interactions with a client and non-anonymous conversations.

Unlike authentication, 'authorization' concerns resource or service access for authenticated users. It relates to the coaching business model (CT.22) and service cost (O.8).

Additionally, a client's lifecycle encompasses various stages, from a prospective client, following initial alignment (CT.17 - CT.23), to a regular client within an ongoing coaching relationship, and ultimately, the conclusion of the engagement. These transitions, indicated in the administration-related objective (O.11) and service termination conditions (CT.27, CT.28, CT.29), necessitate maintaining the client's status with appropriate attributes. Lifecycle changes also involve service access limitations; for instance, access to the initial coaching conversation is contingent upon communicating and obtaining client consent for service fundamentals (CT.17 - CT.23).

Moreover, managing coaching agreements, such as clarifying the coaching direction (CT.62), maintaining the conversation's focus (CT.68), and facilitating changes as indicated by the

client (CT.60), is essential for achieving the goal attainment objective (O.1). Therefore, a data structure mapping client needs to manageable information is essential.

Furthermore, gathering contextual information about the client is integral to building an autonomous coaching service, as a professional coach should be sensitive to the client's context (CT.40), which can influence the coaching conversation.

Working with client patterns related to emotions (CT.41), thinking (CT.45), beliefs (CT.39), or language (CT.43) is a key coaching aspect. Recognising and utilising these patterns in and across coaching conversations necessitates data structures within the user data entity.

Lastly, service usage agreements, such as session scheduling (CT.26), date, time, frequency, and duration, or session location (CT.25), require efficient management throughout the coaching relationship. Objectives indicating the need for efficient agreement management include service usage flexibility (O.9) and minimising administration effort (O.11). Although not applicable to digital services with high availability and location independence, other forms of service usage agreements are necessary, such as predefined client commitments to minimum weekly service usage or fixed session durations, analogously to traditional human-to-human coaching services.

5.3.3 Data Element: Outcome

Managing client relationships, akin to the management of coaching outcomes, is a core competence in a coaching relationship at multiple levels. The user entity and outcome entity share a one-to-many relationship, as a single user may have numerous outcomes, but each outcome is uniquely associated with one user due to the user-specific nature of these outcomes. Essential attributes of outcomes, directly and explicitly linked based on coaching common themes, include type, level, description, measure, and status.

The coaching common themes refer to various forms of coaching outcomes: client value (CT.15), awareness (CT.54, CT.63), goal (CT.64), outcome (CT.65, CT.66, CT.67), action (CT.69,

CT.70, CT.71, CT.72), insight (CT.76), and idea (CT.77). There are numerous types of coaching outcomes, potentially categorized based on the attribute type.

Certain outcome types are specific to a session (CT.67), while others pertain to the overarching coaching engagement (CT.65). Distinguishing between different levels of outcomes is crucial, as managing them is a core coaching competence.
In addition to type and level, a clear textual description is essential for all outcome types, whether they involve a new insight, an action item, or a behavioural change.

A professional coach partners with the client to define and reconfirm both session level (CT.73) and coaching engagement level (CT.74) outcomes. The proposed attribute 'measure' identifies the means to evaluate the achievement of the specified outcome.

Lastly, actionable, measurable outcomes necessitate monitoring to support the client's progress enhancement (CT.79). It is feasible only when the outcomes are defined and reconfirmed. Monitoring hinges on the status attribute, which may indicate various stages of achievement, such as identification, commitment, ongoing implementation, or completion.

5.3.4 Data Element: Conversation

In the context of autonomous coaching, human-computer conversation is the primary method for defining outcomes. The user entity and conversation entity are in a one-to-many relationship, as a user may have multiple conversations, but each conversation is unique to a user since group coaching falls outside the scope of the framework design (C.2).

The conversation entity encompasses all messages across conversations; thus, a 'session' attribute is necessary to collate messages within a conversation or coaching session.
Moreover, a conversation constitutes an exchange between coach and client, making it imperative to document the message originator, either the user or the autonomous coaching system, as well as the messages themselves. Recording the sequence of messages is essential for reconstructing a conversation. While any numeric data type could fulfil this need, the coaching common themes of scheduling (CT.26) and session focus management (CT.68)

necessitate recording each message's timestamp, rendering separate sequence recording redundant. The 'scheduling' attribute (CT.26) reflects conversation constraints, such as the user's commitment to the conversation's duration. Lastly, the attributes of user and LLM meta-data capture all non-verbal cues derived from specific communications. For instance, in an audio-based system, this might include pauses or intonation, while in a video-chat system, it could encompass facial expressions or other visual indicators. Similarly, LLM-generated messages may be enhanced with emotional patterns and additional meta-data.

5.3.5 Data Element: Reporting

The 'reporting' entity pertains to the collated data from three entities: user, outcome, and conversation. Collecting individual user information (O.5) is vital to cater to client needs and sustain high-quality coaching conversations (O.13), but data on typical user patterns, challenges, or goals may also offer valuable insights for the individual client. Data mining of conversation content is crucial for adhering to the coaching common theme of continuous professional development (CT.12). This could be achieved in various ways within an autonomous coaching service, notably through individual user data collection, aggregation, and analysis to enhance service effectiveness or efficiency. The reporting entity's attributes are broadly defined as each one could be a focal point for data collection, and additional compound information may be deduced based on specific needs. Given the intricate interrelations in data collection, the reporting entity has a many-to-many relationship with the connected entities.

5.3.6 Data Element: Coach

Coach behaviour is the primary focus of automation, although its contribution to the outcomes of engagement is controversial in research. Traditional human-to-human coaching involves at least two stages of behavioural alignment: To begin, the client selects a coach, thereby choosing a particular knowledge base, skill set, abilities, personality, values, beliefs, and behavioural patterns; finally, the coach's behaviour adapts (CT.35) to the client during cooperation. While this two-stage process need not be replicated in autonomous coaching systems, any system behaviour projects a "coach personality" to the user whether it is

consciously posed or not. A constraint of pre-LLM dialogue agents was their inability to display a consistent personality (S. Zhang et al., 2018), which made the conversation less engaging. An inconsistent or unexpected "coach personality" could harm the working relationship and the coaching outcomes, as a client-coach mismatch is a top reason for early contract termination (P.54).

Supporting a data-driven approach to "coach personality" selection at the coaching relationship's start are two arguments: CT.59, a competence on self-awareness of the coach, and the promotion of client autonomy (CT.37, O.14) by allowing users to choose the default behaviour set in the engagement. No technical solution leads personality data modelling; however, conceptually, a 'profile' attribute can define the coach's default behavioural patterns. For instance, Coach A's profile may involve asking 60% challenging questions (CT.36) and expressing empathy in 20% of messages (CT.48), while Coach B may have a profile with 10% challenging questions and 40% empathic statements. Adapting behaviour to the client is essential (CT.35) but selecting a coach profile offers an additional way to influence default behaviour.

5.3.7 Data Element: Dialog

Lastly, dialogue management also requires data. The mainstream research community about conversational agents are focusing on two use-cases: open-domain conversational agents (Roller et al., 2020), which are able to talk just about anything and closed-domain agents that are based on a specific knowledge and usually they target to solve a specific task (Nakano & Komatani, 2019). Both types require interpreting user messages and using a knowledge base to guide conversation or act, like booking travel. However, coaching conversations differ as they aim to support clients by discussing their goals, with less emphasis on knowledge transfer. A significant challenge for human coaches is to avoid giving advice, as Stainer warns against becoming an "advice-giver/expert/answer-it/solve-it/fix-it" (Bungay Stanier, 2016).

Therefore, a sophisticated method of knowledge model building is not required to deliver a coaching conversation because most of the complex thinking work is done by the client. It

makes the initial coaching automation significantly easier compared to other conversational agents where knowledge building and exact information exchange is the focus.

AI-driven conversational agents need large data sets for algorithm training. Thought leaders in the field note that AI requires a digital foundation and unique data, without shortcuts (Bughin et al., 2017). Due to severe shortcomings of manual research methods, and the confidential nature of coaching conversations (CT.31), there is a lack of research data for developing autonomous coaching services using modern LLM technologies. Starting without data requires a robust dialog management algorithm to support LLMs in generating relevant content. When sufficient quality data is available, specialized LLMs can be trained, and the conversational knowledge base model can be adjusted to the new automation paradigm.

5.3.7.1 Communication Knowledge Base

Coaching conversations comprise three overlapping areas: coaching-specific (O.2), coaching-goal-related domain (O.1), and human social norms. Coaching-specific communications follow core competencies, while the coaching-goal-related domain concerns the conversation's subject. Social norms may include greetings or small talk at the conversation's start and end, conforming to the shared culture between coach and client. Embedding knowledge, empathy, and personality into conversations is well-studied in Natural Language Processing, but blending these seamlessly is relatively new in Machine Learning (Smith, Williamson, Shuster, Weston, & Boureau, 2020). Listening (CT.38) and empathy (CT.48) are essential for both general and coaching conversations. Social norms in digital interactions are like face-to-face, assuming repeated and synchronous interactions, which are coaching key characteristics (Garas, Garcia, Skowron, & Schweitzer, 2012). Communication aspects of social norms are not detailed in coaching competencies, assuming that coaches are already effective communicators, focusing instead on distinct coaching aspects. For high-quality autonomous coaching dialogues (O.13), it's crucial to consider general communication aspects alongside coaching competencies.

Considering LLM-driven autonomous coaching, individual messages by the user or LLM form the conversation. The data model for a knowledge base can be split from an application perspective into intentions and response recognition patterns.

5.3.7.2 Intentions

Intentions are attributes that determine specific behaviours an automated conversational agent must exhibit to facilitate an effective conversation. The intentions can be categorized in three categories: generic conversational intentions that could occur in any human conversation, coaching conversation specific intentions and coaching goal specific intentions.

Table 5.3: Intentions in Coaching Conversation

Common Theme Key	**Intentions in Coaching Conversation**
CT.33	keeping client responsible
CT.36	challenging client
CT.37	promoting client autonomy
CT.46	showing responsiveness to the client
CT.48	showing empathy
CT.53	showing emotions
CT.52	understanding of key issues
CT.54, CT.76	raising awareness / evoking insight
CT.55	reflecting on interactions with client
CT.56	reflecting on coaching
CT.57	showing respect
CT.59	showing coach personality
CT.64	establishing goals
CT.67	defining session outcome
CT.73	defining measure of success for session outcome
CT.67	reconfirming session outcome
CT.73	reconfirming measure of success for session outcome
CT.68	managing session focus
CT.69, CT.70, CT.71, CT.72	action planning
CT.75	exploring beyond current thinking
CT.77	facilitating idea generation
CT.78	reframing
CT.79	enhancement of client progress
CT.66	identifying potential coaching outcome
CT.65	establishing overall coaching outcome
CT.74	defining measure of success for coaching engagement outcome
CT.74	reconfirming measure of success for coaching engagement outcome

To start with, intentions specific to social norms in a generic conversation are confined to initiating and concluding conversations and not elaborated more in the COBAC framework.

After that, intentions in coaching conversation are extrapolated from the 'coach behaviour' and 'partnering behaviour' thematic clusters within the coaching common themes (Appendix D). Table 5.3 summarises these coaching conversation specific intentions without coaching common themes that are not related to a behaviour or intention demonstrated during a coaching conversation, such as compliance to ethical standards or arranging the logistics of a coaching session.

Lastly, intentions linked to the domain-specific knowledge base, unlike social norms, do not introduce new elements but refine a subset of coaching intentions. For example, promoting client autonomy (CT.37) is a broad coaching intention that may not need customisation for specific goal domains. However, raising awareness (CT.54) could be detailed for domain-specific aspects, such as best practices and concepts, each serving as a foundational block for effective coaching dialogue. Intentions are determined by user need (CT.35, CT.60, CT.62), user context (CT.35, CT.40), time (CT.26), coach profile (CT.11), and exception handling.

5.3.7.3 Response patterns

Response recognition patterns are conversational structures necessary to discern user intentions for dialogue management or exceptional case identification.

Message patterns associated with general conversations based on social norms for recognising implicit feedback (W. N. Zhang, Li, Cao, & Liu, 2018) or emotion recognition (Poria, Majumder, Mihalcea, & Hovy, 2019), are useful for dialogue management. Exceptional case management refers to patterns that are indicative of a conversation derailment, such as client language mismatches or breaches in professional relationship boundaries (CT.5).

In addition to generic patterns, coaching conversations require specific response recognition patterns. Interpreting client responses requires basic response interpretation (CT.46). Emotion recognition algorithms can leverage patterns to classify client emotions to enable

showing emotions (CT.53). Identifying conversation boundaries, such as expected duration (CT.26), is also vital. Finally, exceptional case recognition on the coaching dialogue level might contain patterns related to testing ethical boundaries (CT.13) or recognising when the actual effect of a question on a client may be different than what was originally intended (Tomm, 1988).

Domain-specific response patterns pertain to questions or question sets, such as evaluating domain-specific questionnaire responses.

5.3.8 Outcome

In summary, Figure 5.2 illustrates the entities and their interrelations within the conceptual data model via a UML-based visualisation of the analysed data elements.

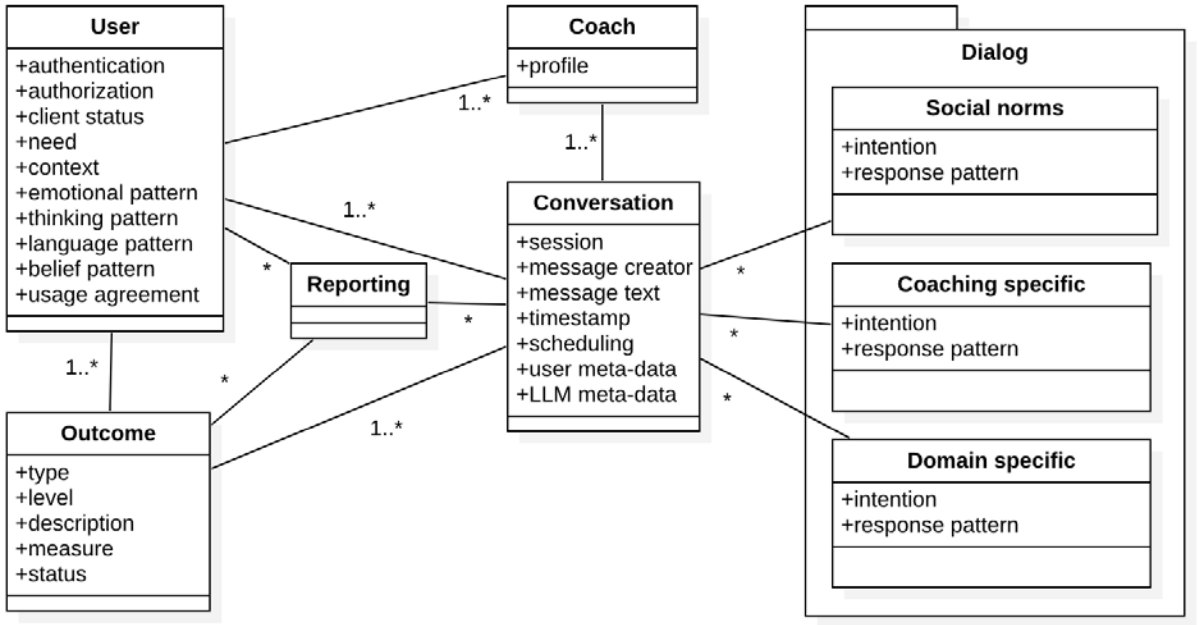


Figure 5.2: COBAC Conceptual Data Model

As discussed, numerous data elements are essential for designing an autonomous coaching service, encompassing commercial terms, client privacy, logistics, client-generated content, interaction metadata, coach-generated content, reporting, and dialogue management. The COBAC conceptual data model offers a birds-eye overview of this complexity.

The subsequent section describes the most significant processes that use the discussed data elements to deliver the autonomous coaching service.

5.4 Process Modelling

In Section 5.4 on process modelling, the focus shifts to the identification of major coaching processes, aiming to derive algorithms essential for coaching automation.

5.4.1 Approach

Process analysis relies on market research and outcomes of previous analyses, such as the Initial Objective List, Coaching Common Themes, Initial Constraint List, and the Initial Problem List. To ensure clarity and precision in visualizing key processes, Business Process Model and Notation (BPMN) version 2 (Object Management Group (OMG), 2011) was employed.

This process modelling exercise encompasses three core areas: coaching administration, coaching engagement, and coaching session processes. This scope ensures a comprehensive view of the coaching industry's operational framework. A concise summary is created, encapsulating the key processes and their associated activities, all crucial within an autonomous coaching environment. The coaching process model is constructed with a high level of generalization, deliberately chosen to keep the COBAC framework technology and communication channel agnostic. This approach avoids overly detailed process analysis, which could excessively narrow the solution space and render the COBAC framework less adaptable to various contexts.

5.4.2 Process: Coaching Administration

The coaching administration process (O.11) is best described by dividing the engagement into three phases: engagement setup, ongoing administration, and engagement closure. Although some activities fall within coaching core competences, the administration process attracts minimal academic research interest. However, it is often targeted by market research. In an

automation context, neglecting any part of the complete process is unviable, as a fully automated system depends on end-to-end user journeys, covering each mandatory step.
Due to a lack of academic research, initial setup phase information was derived from market research studies and coaching common themes (Appendix D). Figure 5.3 illustrates a typical engagement setup process using BPMN.

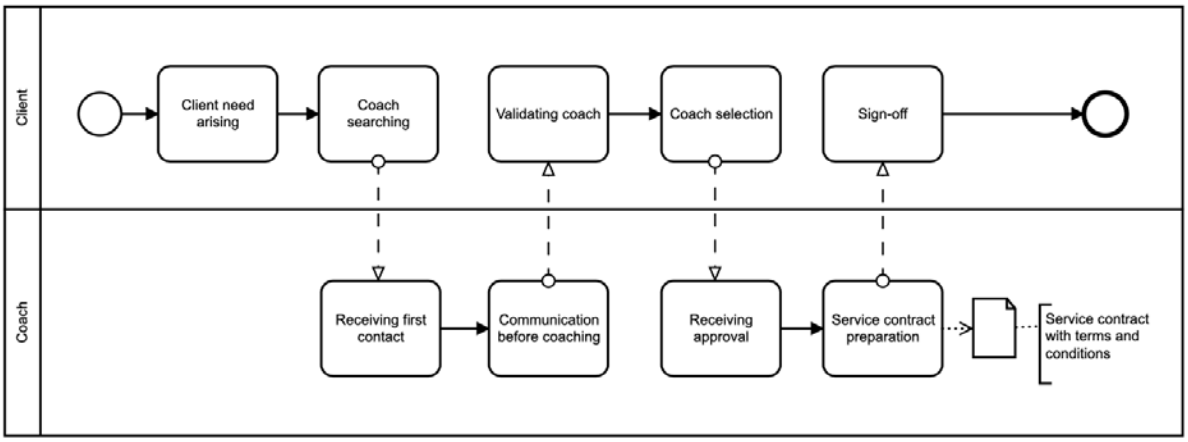


Figure 5.3: Coaching Administration Process - Engagement Setup

Figure 5.3 depicts a complete and successful engagement setup process, although deviations or breakdowns can occur. The first step in contracting a coach is identifying the client's need for personal change or learning. The first step toward closing a service contract with a coach is to identify the need of the client about a personal change or learning. As the problem analysis showed client awareness of coaching is low (P.1) and coaching is not perceived as a relevant, important (P.2) or scientifically founded (P.3) modality to address goal achievement. Additionally, poor coaching service affordability (P.18) can lead to early process termination. Market research indicates inefficient economic conditions: 46% of clients search for coaches based on personal referrals, and 58% engage the first coach they find (ICF, 2009). From the coach's perspective, after the first client contact, an appointment is arranged to discuss needs and coaching fundamentals: code of ethic (CT.17), potential benefits (CT.18), role (CT.19), modality (CT.20), process (CT.21), service terms (CT.22), and confidentiality limits (CT.23). If the coach is selected, a service contract is created and signed by both parties. The impact of automation on the engagement setup process is associated with tasks in the coach lane of

Figure 5.3, which need automation but are unlikely to change significantly as they are like tasks in other digitalised services.

Similarly, ongoing administration involves financial process management, such as payment scheduling, invoicing, and charging, which remain unchanged even when conducted digitally. Scheduling (CT.25, CT.26) and rescheduling coaching sessions are also part of ongoing administration. Initial agreements typically cover high-level logistics, like session frequency, location, or digital communication channels, and duration. Cancellation terms, usually in the service contract, significantly affect service profitability. Individual session arrangements, based on these conditions, are not necessary for an automated system, which eliminates time (P.6), location (P.7), and coach dependency (P.8), which allows for exploring alternative business models.

Lastly, an engagement can terminate early upon client request (CT.27), value shifts (CT.28), unresolvable conflicts (CT.29), or when coach competence boundaries are reached (CT.30). Except for unresolvable conflicts, these situations should also be managed in an automated context. Engagement closure typically involves evaluating the coach (EMCC_CF_069), overall process (EMCC_CF_028), and outcomes (EMCC_CF_071), which can be automated with no significant task changes.

5.4.3 Process: Coaching Engagement

Following the setup of engagement, the initial coaching session commences, addressing the client's needs. The coach collaborates with the client to establish the session outcome (CT.15, CT.54, CT.63, CT.64, CT.65, CT.66, CT.67, CT.69, CT.70, CT.71, CT.72, CT.76, CT.77), which is elaborated in the section pertaining to the conceptual data model. The client implements the session outcome within their context, and the results, alongside any new learning accrued during this phase, become the foundation for the subsequent session, in conjunction with the overarching goal of the coaching engagement (CT.74). This iterative process continues through to the concluding session, as illustrated in Figure 5.4.

The process of coaching engagement remains unaltered by automation, although minor adjustments may be made by the client owing to the enhanced accessibility of IT systems, potentially leading to more frequent sessions and shortened durations.

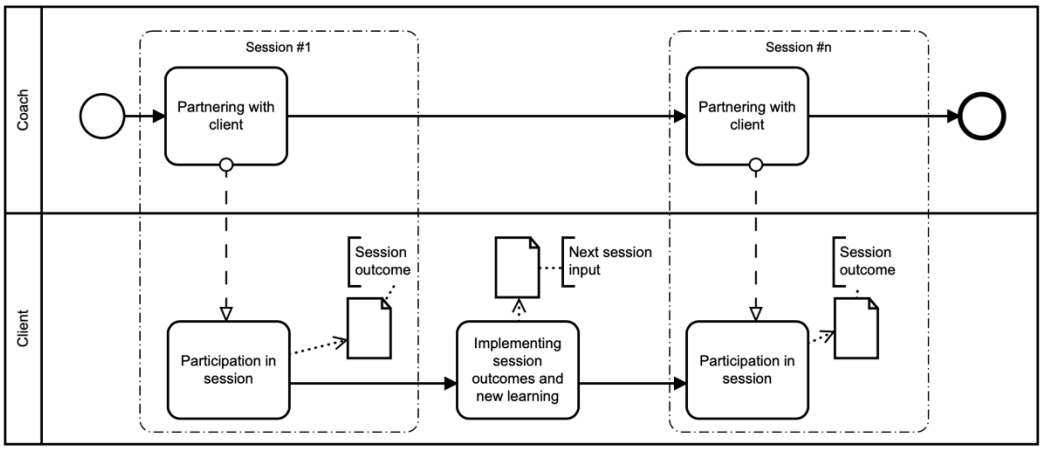


Figure 5.4: Coaching Engagement Process

5.4.4 Process: Coaching Session

A universally applicable, step-by-step description of coaching sessions is non-existent due to the iterative and dynamic nature of coaching. Nonetheless, this does not preclude the conceptualisation of the process, which requires abstraction at a higher level while managing numerous parameters and intricate data configurations. Figure 5.5 presents a generalised process for a coaching session conducted between an autonomous coaching service and a client.

The client initiates the conversation, triggering the autonomous coaching service's dialogue management process. In response, the service provides initial coaching messages aimed at establishing rapport, clarifying goals, and encouraging the client's input. The client's replies prompt repeated cycles of interpretation, data model updates, intention identification, and the generation of new coaching messages. With each exchange, the service tailors its guidance based on the client's evolving needs and context. This adaptive cycle continues until the service delivers concluding messages, summarising insights, and reinforcing next steps.

At the core of the dialogue management subprocess are four key components. The system interprets the client's message, extracting intent, emotional cues, and relevant themes. Next,

it updates the data model, an evolving repository that stores session history and personal context. This refined understanding informs the coach's intentions, shaping subsequent guidance. Finally, the service generates the next set of messages, translating these intentions into supportive, context-aware communication.

The entire interaction forms a coherent coaching conversation. By continuously cycling through interpretation, data updates, intention setting, and message generation, the service maintains a dynamic, personalised dialogue that helps the client progress towards their goals.

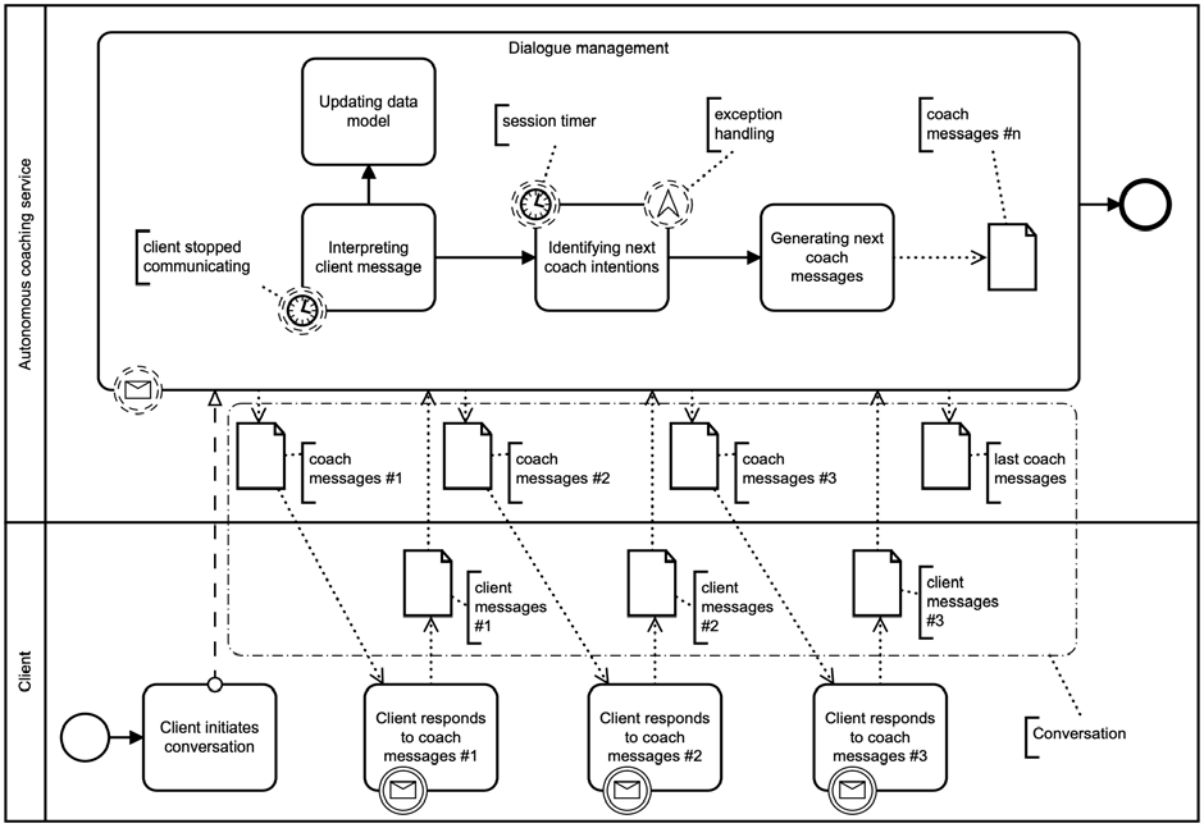


Figure 5.5: Coaching Session Process

5.4.4.1 Interpreting client message

The 'interpreting client message' subprocess begins with the signal to initiate a conversation or when the client pauses communication for a specified duration after sending messages, as denoted by the timer's non-interrupting boundary event. This stage utilises response recognition patterns from the conversational knowledge base to deduce information from

the client's direct and meta-communication. Client messages, which provide data on various attributes relating to the user, outcome, and conversation entities, are analysed, and converted into data structures suitable for subsequent processing.

5.4.4.2 Updating data model

The user, outcome, conversation, and reporting entities in the data model can be updated based on the extracted information from client messages. Each client utterance might reveal more contextual information about these entities, which should be categorized and as a knowledge base gathered for future utterance generation. For instance, one message might clarify and add more details to the conversation goal, which should result in the updated outcome record in the database.

5.4.4.3 Identifying next coach intentions

A list of coach intentions must be determined for the next set of coach messages, based on the updated data model. For example, the intentions leading to the generation of “coach messages #1” include greeting, demonstrating respect to the client (CT.57), and remaining observant of the client (CT.38). The “last coach messages” may include saying goodbye and managing the emotions of the coach (CT.53). There are many factors that influence the selected intention, they may include the client need, client context, timing of the message, conversation history, exceptional case, or coach profile.

5.4.4.4 Generating next coach messages

The task of generating the next set of coach messages stems from the identified list of coach intentions in the previous step and the content of the updated data model. For instance, the intentions of greeting, demonstrating respect to the client (CT.57), and remaining observant to the client (CT.38) from the previous section may lead to the following message: “Hi! This is the fifth time we've spoken this month; I enjoy supporting you as you invest time in achieving your goals. How are you today?” The data for the three sentences in the message originate

from the client context. The first two intentions do not require a client response, whereas the third intention prompts a response from the client.

5.4.5 Outcome

The coaching process model is derived from the objectives, common coaching themes, and linked research review from three parallel processes: coaching administration, coaching engagement, and coaching session processes. Table 5.4 summarises the key activities associated with these processes.

These activities must be managed and balanced throughout the service usage lifecycle, as they interact with each other and form the foundation of the coaching client experience.

Table 5.4: Key Activities of Autonomous Coaching Systems

Key	Process	Sub-Process	Activity
PR.1	Administration	Engagement Setup	Communication about the need of the client and fundamental elements of coaching
PR.2	Administration	Engagement Setup	Contracting
PR.3	Administration	Ongoing	Charging
PR.4	Administration	Closure	Evaluation of the coach, the overall coaching process, and overall engagement outcomes
PR.5	Engagement		Partnering with the client to manage the coaching engagement
PR.6	Session		Dialog Management

5.5 Deriving Solution Requirement List

The final modelling technique applied to derive the solution requirements, the last missing component of the COBAC framework, relates to service quality.

5.5.1 Approach

The COBAC solution requirements are derived from the coaching common themes (Appendix D), the Initial Objective List, the conceptual data model, and the coaching process model, following the structure of the 'System/Software Product Quality model' defined in ISO/IEC 25010:2011 (ISO/IEC, 2011) as depicted in Figure 5.6.



Figure 5.6: ISO/IEC 25010 Quality Characteristics

Solution requirements are derived considering the best practices defined by IREB (Frühauf et al., 2017) with the exception of the formal criteria associated with requirement formulation: "formulate only one requirement per sentence" (Frühauf et al., 2017). It is not followed to keep complexity and size of the COBAC framework manageable. The proposed solution requirements contain high-level conceptual considerations that require further refinement and tailoring to the actual context of the service at the time of detailed system design, which may result in high-level solution requirements being divided into many more granular sub-requirements. Consequently, a working instance of an actual autonomous coaching service might implement thousands of low-level solution requirements.

Solution requirements presented in Section 5.5 include only key and summary fields to ensure readability, but a complete list of solution requirements together with the objective reference, coaching common theme reference and problem reference is available in the Appendix E. Additional information is provided to describe the connection between the solution requirements and other conceptual elements for traceability and extensibility purposes.

5.5.2 Requirement: Functional Suitability

The functional suitability requirement characteristics is defined as the "degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions" (ISO/IEC, 2011). Table 5.5 contains a summary of the COBAC solution requirements covering the functional considerations and derived from the coaching common themes, data modelling outcome, and process modelling outcome.

Table 5.5: Solution Requirement Summary - Functional Suitability

Key	Solution Requirement Summary
SR.1	An openly accessible professional ethical code is created with at least the following topics: ethical code breach reporting, relationship boundaries, potential interest conflicts, data management, honest system capability representation, continuous development, discrimination, intellectual property, user referral and technology specific ethical concerns
SR.2	The user needs to be informed before the first coaching conversation at least about the professional code of ethic, coaching value, coaching modality, coaching process, terms and conditions and confidentiality limits
SR.3	The user is required to accept the terms and conditions before the first coaching conversation takes place
SR.4	The user can initiate a coaching conversation after acceptance of terms and conditions any time
SR.5	The status of the user is tracked throughout the service lifecycle of the user with at least the following stages: new user, active coaching user, inactive coaching user
SR.6	Managing information about the user needs is required throughout the service lifecycle of the user
SR.7	Managing information about the user context is required throughout the service lifecycle of the user
SR.8	User pattern collection and usage in interactions is required for the following areas: emotion, thinking, belief and language
SR.9	Service usage agreement is managed on two levels: committed time related to the service in general and individual coaching conversation duration
SR.10	Outcomes are collected from interactions with the user related to individual coaching conversations as well as for the overall coaching engagement
SR.11	At least the following attributes are associated and managed related to outcomes: type, level, description, measure and status
SR.12	Conversation history is maintained on an individual message level with at least the following attributes: message creator, message, timestamp, scheduling, session, and meta-data
SR.13	Aggregated statistical data in the form of reporting about users, conversations and outcomes is collected and used for continuous service operation and optimisation
SR.14	The user can choose from a set of coach profiles that describes behavioural preferences of the coaching service
SR.15	Dialog management related knowledge base is maintained for general conversations following social norms with at least the following attributes: intentions and response recognition patterns
SR.16	Dialog management related knowledge base is maintained for conducting coaching conversations with at least the following attributes: intentions, and response recognition patterns
SR.17	Dialog management related knowledge base is maintained for conducting user need specific conversations with at least the following attributes: intentions, and response recognition patterns
SR.18	System messages are only generated when the coaching conversation is coming to its end or when the user stopped communicating to avoid interruptions
SR.19	User messages are interpreted based on the response recognition patterns defined in the conversational knowledge base
SR.20	The information in the data model is updated based on the interpreted messages

SR.21	Interpreted messages and the updated information in the data model is used to identify next coaching intentions
SR.22	Conversational intentions are determined at least on the following factors: user need, user context, time, coach profile, and exception handling
SR.23	The system demonstrates the following intentions: acknowledging and keeping user responsible, challenging the user, actively promoting user autonomy, remaining observant, appropriate language usage, remaining responsive, integrating verbal and non-verbal communication, remaining empathetic and impartial, using a wide range of questioning, ensuring understanding of key issues, displaying emotions, raising awareness of the user, reflecting on the interactions with the user, respect, building trust
SR.24	The system supports the user in the following areas: defining user's need, building awareness of the impact on the user's wider context, establishing and reconfirming goals, overall outcomes, conversation outcomes and related measures of success, managing conversation focus, planning actions, exploring beyond current thinking, evoking insight, generating ideas, reframing, and enhancement of user progress
SR.25	Information about the implementation of coaching conversation outcomes between conversations and new learning from the implementation is used as an additional input for the following coaching conversations
SR.26	Seeking help from outside sources when necessary
SR.27	Provides private community features for users pursuing similar goals
SR.28	A coaching conversation is terminated based on user request anytime or when the planned duration is reached
SR.29	Engagement closure is managed in planned case as well as early termination case, such as shifting user value, unresolvable interest conflict, or reaching system competency boundary

The functional suitability characteristics consist of functional completeness, correctness, and appropriateness. The above-listed solution requirements are complete in the sense that they represent a minimum set of high-level requirements to fulfill the stated objectives and contribute to solving the referenced problems derived from the context analysis. They are considered correct because they were derived from literature analysis and from documentation, which are viewed as cornerstones of coaching competences. Finally, they are appropriate because they were derived from a consistent conceptual data model and process model that covers all identified common coaching themes.

5.5.3 Requirement: Performance Efficiency

Another requirement characteristics is performance efficiency, which is defined as “performance relative to the amount of resources used under stated conditions” (ISO/IEC, 2011). This characteristic consists of time behaviour, resource utilization, and capacity.

The response time within a coaching conversation is the only time-critical aspect that is specific to the autonomous coaching system. In general, silences between utterances are person-specific and can influence a conversation dramatically. A study has shown that participants in a non-coaching context tend to synchronize the length of silence (Fors, 2015). In a coaching conversation, the pause between the client messages and the coach messages is made artificially long but not too long to give sufficient space for the user to think without projecting the perception of being left alone. Because of this reason, the message response time of the system in a coaching conversation needs to be set higher than the observed response time of the user. This requirement is in line with the following coaching common themes: adapting the coaching behaviour to the client (CT.35), remaining observant to the client (CT.38), and remaining responsive to the client (CT.46).

The sub-characteristics "resource utilization" is defined as the "degree to which the amounts and types of resources used by a product or system, when performing its functions, meet requirements" (ISO/IEC, 2011). This characteristic is not specific to the coaching software context; however, it influences the ethical concern about environmental concerns (P.23, P.46), as well as poor resource utilization is inversely correlated with the cost of the service (O.8).

Lastly, regarding the sub-characteristics of capacity, it is tacit knowledge that the human coach's capacity at any moment in time is a maximum of one client; hence, the question of capacity is not even raised in a traditional human-to-human coaching context. On the other hand, system scalability to support a varying and high number of parallel user interactions is an essential added value of autonomous systems that support the user with independence (O.14) and the method to contribute to the minimization of direct cost (O.8). Since this aspect influences the low-level service design, a related solution requirement needs to be introduced.

5.5.4 Requirement: Compatibility

The compatibility requirement characteristics is defined as the "degree to which a product, system or component can exchange information with other products, systems or

components, and/or perform its required functions, while sharing the same hardware or software environment" (ISO/IEC, 2011). It consists of co-existence and interoperability subcategories.

There is no indication in the literature that co-existence is relevant in the sense of sharing the digital environment with other software; however, an actual implementation of the COBAC framework might consider adding requirements to support the hybrid model of using an autonomous coaching system to provide augmentation to the services of a human coach. However, this is out of scope for the current study.

Interoperability, on the other hand, is required to ensure a smooth information exchange with the third party LLM system.

5.5.5 Requirement: Usability

The usability characteristics covers many subcategories, such as appropriateness recognizability, learnability, operability, user error protection, user interface aesthetics and accessibility (ISO/IEC, 2011).

A user interface is required for an autonomous coaching service; however, usability requirements related to an audio-video based autonomous coaching service are very different compared with a chat-based coaching system. Apart from naming the relevant sub-characteristics, solution requirements related to usability are not specified further in the technology agnostic COBAC framework because user interface design is highly dependent on the applied technology and the communication channel used between the system and the user. A requirement must be included in the requirement characteristics usability to ensure that the user experience is continuously improved (O.12).

5.5.6 Requirement: Reliability

Reliability consists of four sub-characteristics: maturity, availability, fault tolerance, and recoverability. High availability is required to enable the client to initiate coaching

conversations at any time (O.3) and to maximise the user experience through uninterrupted interactions (O.12). This is another differentiator compared to coaching services provided by a human coach. A related solution requirement is required since availability requirements can influence the fundamental design of the software and technology choice. The other three sub-characteristics are not directly relevant or specific to the coaching context or the objectives.

5.5.7 Requirement: Security

Security considerations are essential aspects of coaching, but not all the five following sub-characteristics are specific to an autonomous coaching system: confidentiality, integrity, non-repudiation, accountability, and authenticity. Non-repudiation and accountability in the context of information security are omitted, considering the partnering nature of coaching engagements.

On the other hand, confidentiality (CT.31, CT.23) is required during communication with the user and related to data management processes within the limitations of the pertinent law (CT.32). Lastly, the system requires at least one user authentication and authorization method to check the identity of the user (CT.7, CT.8) and provision the agreed services.

5.5.8 Requirement: Maintainability and Portability

Maintainability and portability requirement characteristics, such as modularity, reusability, analyzability, modifiability, testability, adaptability, installability, and replaceability, are related to the operational aspects of the software system. Operational aspects impact the objective of minimization of direct cost (O.8) because operational costs take the largest share of the total cost of digital service ownership (Malhotra & Chug, 2016), hence generic solution requirements need to be included to address these items in the actual service design.

5.5.9 Outcome

The summary of Section 5.5, the complete list of the initial solution requirements, can be found in Table 5.6. Additionally, the same table with references to the contextual problems, objectives, coaching common themes, and software quality model characteristics can be found in Appendix E.

Table 5.6: Initial Solution Requirement List

Key	Solution Requirement Summary
SR.1	An openly accessible professional ethical code is created with at least the following topics: ethical code breach reporting, relationship boundaries, potential interest conflicts, data management, honest system capability representation, continuous development, discrimination, intellectual property, user referral and technology specific ethical concerns
SR.2	The user needs to be informed before the first coaching conversation at least about the professional code of ethic, coaching value, coaching modality, coaching process, terms and conditions and confidentiality limits
SR.3	The user is required to accept the terms and conditions before the first coaching conversation takes place
SR.4	The user can initiate a coaching conversation after acceptance of terms and conditions any time
SR.5	The status of the user is tracked throughout the service lifecycle of the user with at least the following stages: new user, active coaching user, inactive coaching user
SR.6	Managing information about the user needs is required throughout the service lifecycle of the user
SR.7	Managing information about the user context is required throughout the service lifecycle of the user
SR.8	User pattern collection and usage in interactions is required for the following areas: emotion, thinking, belief and language
SR.9	Service usage agreement is managed on two levels: committed time related to the service in general and individual coaching conversation duration
SR.10	Outcomes are collected from interactions with the user related to individual coaching conversations as well as for the overall coaching engagement
SR.11	At least the following attributes are associated and managed related to outcomes: type, level, description, measure, and status
SR.12	Conversation history is maintained on an individual message level with at least the following attributes: message creator, message, timestamp, scheduling, session, and meta-data
SR.13	Aggregated statistical data in the form of reporting about users, conversations and outcomes is collected and used for continuous service operation and optimisation
SR.14	The user can choose from a set of coach profiles that describes behavioural preferences of the coaching service
SR.15	Dialog management related knowledge base is maintained for general conversations following social norms with at least the following attributes: intentions and response recognition patterns
SR.16	Dialog management related knowledge base is maintained for conducting coaching conversations with at least the following attributes: intentions, and response recognition patterns

SR.17	Dialog management related knowledge base is maintained for conducting user need specific conversations with at least the following attributes: intentions, and response recognition patterns
SR.18	System messages are only generated when the coaching conversation is coming to its end or when the user stopped communicating to avoid interruptions
SR.19	User messages are interpreted based on the response recognition patterns defined in the conversational knowledge base
SR.20	The information in the data model is updated based on the interpreted messages
SR.21	Interpreted messages and the updated information in the data model is used to identify next coaching intentions
SR.22	Conversational intentions are determined at least on the following factors: user need, user context, time, coach profile, and exception handling
SR.23	The system demonstrates the following intentions: acknowledging and keeping user responsible, challenging the user, actively promoting user autonomy, remaining observant, appropriate language usage, remaining responsive, integrating verbal and non-verbal communication, remaining empathetic and impartial, using a wide range of questioning, ensuring understanding of key issues, displaying emotions, raising awareness of the user, reflecting on the interactions with the user, respect, building trust
SR.24	The system supports the user in the following areas: defining user's need, building awareness of the impact on the user's wider context, establishing and reconfirming goals, overall outcomes, conversation outcomes and related measures of success, managing conversation focus, planning actions, exploring beyond current thinking, evoking insight, generating ideas, reframing and enhancement of user progress
SR.25	Information about the implementation of coaching conversation outcomes between conversations and new learning from the implementation is used as an additional input for the following coaching conversations
SR.26	Seeking help from outside sources when necessary
SR.27	Provides private community features for users pursuing similar goals
SR.28	A coaching conversation is terminated based on user request anytime or when the planned duration is reached
SR.29	Engagement closure is managed in planned case as well as early termination case, such as shifting user value, unresolvable interest conflict, or reaching system competency boundary
SR.30	Message response time of the system in a coaching conversation is higher than the response time of the user
SR.31	The resource utilization aspect needs to be considered for software design decisions
SR.32	System scalability is required to support a high and varying number of parallel user interactions
SR.33	Effective interoperability is ensured with the LLM system
SR.34	Appropriateness recognizability, learnability, operability, user error protection, user interface aesthetics and accessibility aspects are required to be considered at software design and continuous development
SR.35	High availability is required to provide partnership in coaching conversations anytime
SR.36	The system requires at least one user authentication and authorization method
SR.37	Confidentiality is ensured during communication with the user and related to data management processes within the limitations of the pertinent law
SR.38	Modularity, reusability, analysability, modifiability, testability, adaptability, installability and replaceability are required to be considered at software design and continuous development

5.6 Chapter Summary

Section 5.2 discusses the necessity of detailed coaching themes for developing autonomous coaching systems, which goes beyond simple definitions to include common themes identified in human-to-human coaching conversations. This thematic analysis uses foundational documents from leading coaching organizations ICF and EMCC to design and develop a comprehensive COBAC framework for coaching competence, incorporating ethical guidelines and terminologies. It emphasizes a textual analysis method that assigns codes to key competencies and ethical codes, which are then compared to identify common coaching themes. These coaching themes are essential for building the initial COBAC framework, which will evolve with more empirical data from system use.

Section 5.3 outlines the conceptualization of a data model for an autonomous coaching system powered by a Language Model. It explains the need for complex data models to capture user interactions, including emotions, knowledge, and context, with a focus on authentication, authorization, and user lifecycle management. The model integrates user input, system configuration, and LLM-generated information, emphasizing the handling of content and meta-data, and the importance of privacy and confidentiality for trust-building in coaching. Finally, it discusses the necessity of structuring data to support user outcomes, conversation management, reporting, coach behaviour, and dialogue management, highlighting the many-to-many relationships between data elements and the overall complexity of an autonomous coaching service.

Section 5.4 on process modelling outlines the development of algorithms for automating major coaching processes, utilizing BPMN for clear visualization of complex processes. It covers the coaching administration, engagement, and session processes to ensure a comprehensive operational framework. The modelling seeks to maintain generality for broader applicability, avoiding over-specialization that could limit the utility of the framework. The section also discusses the significance of complete process coverage for automated systems, citing market research and common themes in coaching as bases for the construction of the framework.

The COBAC framework's service quality modelling adopts the ISO/IEC 25010:2011's 'System/Software Product Quality model' to derive solution requirements. Functional suitability is a key requirement, ensuring the system meets specified needs, with a focus on maintaining a balance between detailed design and readability. Performance efficiency, another vital requirement, emphasizes the system's time behaviour, resource utilization, and capacity, tailored to enhance the autonomous coaching service's value. The Initial Solution Requirement List (Appendix E) integrate high-level conceptual considerations, requiring further refinement for specific service contexts, potentially leading to thousands of sub-requirements in practice.

6. A New Framework for Autonomous Coaching

6.1 Introduction

The components of the COBAC framework were identified in Chapters 4 and 5: Problem List, Objective List, Constraint List, and Solution Requirement List. Chapter 6 contains the description of these initial elements of the COBAC framework to provide a consistent overview of the proposed framework and a Reference Architecture to guide the implementation.

6.2 COBAC Initial Framework Description

6.2.1 Initial Constraint List

As defined in the overview of COBAC components (Table 3.1), constraints limit the problem, objective, and solution spaces. The Initial Constraint List (Table 4.8) represents this aspect within the COBAC framework.

The constraints for autonomous coaching that limit the problem space define the client as an adult individual without mental disorders, not seeking coaching within an organisational context. It assumes the client is the sole representative of all interests, without consideration for broader social, political, or economic contexts, except for specified objectives. Country-specific laws and taxation are excluded, focusing solely on coaching-specific ethical standards, and the immediate social network's impact on the client is not addressed, except in terms of social connectedness. Furthermore, objectives and solution requirements are also limited by excluding group and team coaching approaches, and using only core coaching competencies from modern coaching competency frameworks that define a non-directive, reflective style of coaching.

6.2.2 Initial Problem List

Considering these constraints, the Initial Problem List is created based on the outcomes of the macro context analysis (Table 4.2) and micro context analysis (Table 4.3). The problems associated with the macro-context span across user and vendor stakeholders and encompass social, technological, economic, environmental, political, legal, and ethical domains. In contrast, the micro-context-related issues focus on the direct interactions and experiences of the key stakeholders identified, such as users, vendors, sponsors, and commissioners within the coaching process.

6.2.3 Initial Objective List

The Initial Objective List (Table 4.7), linked with the Initial Constraint List (Table 4.8) and the Initial Problem List (Tables 4.2 and 4.3) through the constraint identifiers and problem identifiers, provides the key objectives of the autonomous coaching system in the COBAC framework. The Initial Objective List emphasizes several key objectives to ensure effective service to their users. These autonomous coaching systems strive to maximise goal attainment by applying core coaching competencies, ensuring high availability, and fostering social connectedness among users with similar objectives. They prioritise regular, transparent reporting to track effectiveness and adherence to ethical standards while continuously improving service consistency. Furthermore, they aim to enable support for high-resource languages, minimise direct costs, and offer flexible service usage concerning conversation timing, duration, and capacity. The integration of scientifically validated theories is an ongoing process, as is the effort to minimise administrative work. Enhancing the user experience remains a constant goal, alongside improving dialogue quality through the latest effective technologies. Above all, fostering user autonomy and upholding user privacy are paramount to the system's integrity and trustworthiness.

6.2.4 Initial Solution Requirement List

Based on the identified constraints, problems, and objectives, and using various modelling approaches, the Initial Solution Requirement List (Table 5.6) was synthesized. The solution

requirements cover both functional and non-functional requirements, such as performance, compatibility, usability, reliability, security, and maintainability characteristics. The extended version of Table 5.6 can be found in Appendix E with additional references to other COBAC components: constraints, problems, and objectives. Highlighting references among the COBAC framework components is crucial to enable traceability and improve the adaptability of the framework for specific contexts or purposes.

6.3 COBAC Reference Architecture

Figure 6.1 illustrates a high-level reference architecture based on the COBAC initial framework using UML visual representation. It incorporates all identified objectives from the Initial Objective List and maps all items from the Initial Solution Requirement List to a blueprint of the autonomous coaching system.

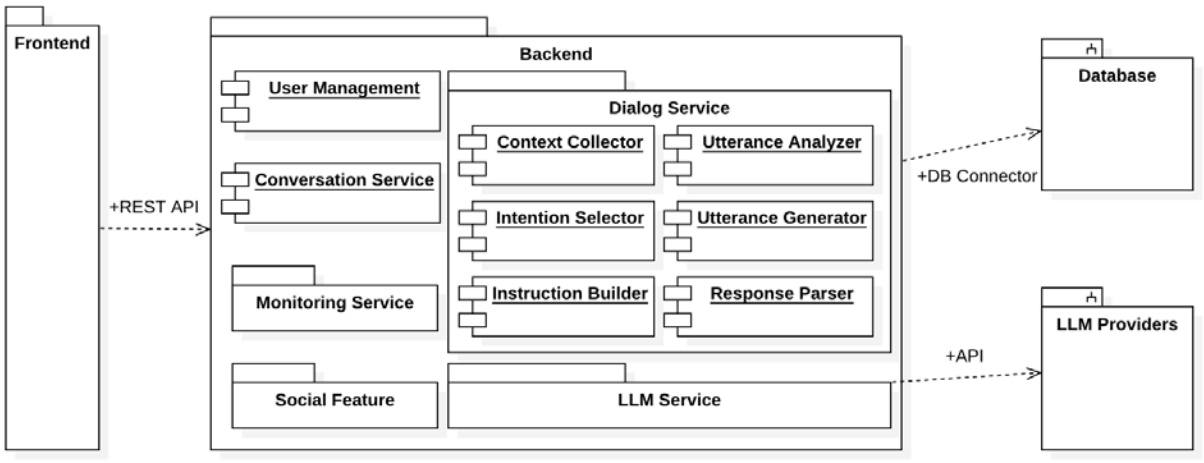


Figure 6.1: COBAC Reference Architecture

The COBAC Reference Architecture demonstrates the component structure of a software system designed to facilitate autonomous coaching, based on a three-tier design framework (frontend, backend, database) with an additional external component: "LLM Providers". The diagram applies various design and architectural principles, such as Separation of Concerns, modularity, API-first, and scalability. These principles are integral to creating a system that is both efficient and capable of evolving over time to accommodate new requirements and challenges.

Furthermore, Table 6.1 summarizes the components used and provides a brief description of each.

Table 6.1: COBAC Reference Architecture - Component Description

Component	Description
Frontend	The frontend describes the user interface through which the system interacts with the user. It can be implemented by using various communication media, such as virtual or augmented reality, chat applications, video streaming, or voice-only interfaces. Additionally, this component is tasked with collecting meta-communication related information, including distractions, geographic location, and device parameters for displaying the user interface.
Backend	The backend implements the system's overall functionality and hosting its key components.
Database	This component stores and manages all system-related data and ensures accessibility via a standardized database connector interface.
LLM Providers	An LLM Provider is an external component that facilitates access to one or more Large Language Models' functionality through an Application Programming Interface (API). The Backend may engage multiple providers simultaneously to deliver autonomous coaching services; hence, the component name is pluralized.
User Management	This component manages the user lifecycle and parameters associated with the user. Typical functionalities include user creation, authentication, authorization, managing user preferences, and user deletion.
Conversation Service	Managing conversation-level information, this component's functionalities involve initiating conversations based on parameters such as language or planned duration, and managing the conversation's objective, summary, and outcomes throughout the dialogue process.
Monitoring Service	This service monitors the system's health at both the technical and content levels. Technical monitoring includes service availability, capacity, and performance, while content monitoring may involve diagnosing conversations for potential ethical concerns or bias.
Social Feature	Aligned with the objective (O.4) to implement social connectedness, this could encompass various features, such as forums, enabling one-on-one conversations among users with similar goals, or a Q&A board with commenting capabilities.
Dialog Service	Responsible for managing the complexities of individual conversations.
Context Collector	This component gathers all contextual information necessary for the LLM to analyse or generate responses to user utterances. Contextual information may include, but is not limited to, the user’s objectives, beliefs, emotions, previous and ongoing conversations, and prior user utterances.
Utterance Analyzer	Tasked with analysing the user's latest communication to manage the dialogue. The analysis may address conversation exceptions, such as detecting improper language, assessing utterance lengths, or identifying attempts to attack the system, such as through SQL injection. This component also ascertains the user's specific intentions for continuing the conversation and the emotions and beliefs associated with their utterances.
Intention Selector	Derives a set of intentions from the analysed user communication to determine the next steps in the conversation. The applied intentions are directly linked to identified

	common coaching themes, such as demonstrating respect (CT.57), establishing goals (CT.64), or managing session focus (CT.68) in the event of a distraction.
Utterance Generator	Generates a specific system utterance in the language of the conversation based on the selected intentions.
Instruction Builder	The preparation of specific LLM instructions, incorporating contextual information, is required for analysing user utterances and generating responses.
Response Parser	Transforms the LLM's response into information interpretable and utilizable by the system.
LLM Service	Facilitates communication with the LLM(s), such as establishing API connections, handling errors, and logging interactions with the LLM for diagnostic purposes.

6.4 Framework Application for Implementation

A new autonomous coaching system can be implemented based on the COBAC framework components and the Reference Architecture presented in Section 6.2 and Section 6.3.

The proposed process for implementation within a specific coaching context is applied in Chapter 7 and described in detail for one specific coaching context. However, the steps are summarised here:

1. Review and adjustment of stakeholders and their specific needs and problems: This step results in an updated and more specific Problem List. For instance, if the application is intended to address the developmental needs of leaders transitioning into executive roles, the characteristics of the user stakeholders would be refined, alongside the specific problems the system is designed to address.
2. Adjustment of the Objective List, Constraint List, and Solution Requirement List: These lists need to be revised to reflect the new coaching context.
3. Refinement of the Reference Architecture: Based on the new context and system requirements, the high-level conceptual Reference Architecture must be refined to develop a system-specific architecture suitable for implementation.

The outcomes of these activities serve as the input and foundation for standard software engineering processes used to design and develop the specific system.

6.5 Chapter Summary

The COBAC framework, outlined in Chapter 6, presents a structured approach for autonomous coaching, focusing on adult individuals without mental disorders and not within an organizational context. The framework is introducing constraints that shape the scope of the autonomous coaching system and exclude broader socio-political factors, group coaching methods, and non-coaching related legalities. The Initial Problem List considers a range of stakeholder concerns across various domains, while the Initial Objective List aims to maximise the effectiveness of the coaching system through core competencies, availability, and ethical standards. Solution requirements, synthesized using various models, address both functional and non-functional aspects of the system, ensuring reliability, security, and maintainability. The comprehensive connections among framework components facilitate traceability and adaptability, as documented in the extended tables and Appendix E.

The COBAC Reference Architecture presented in Section 6.3 outlines a software system for autonomous coaching, depicted through a UML visual diagram, following the COBAC initial framework. The architecture is structured into three tiers: frontend, backend, and database, alongside an external component, LLM Providers. Key design principles such as Separation of Concerns and modularity ensure the system's efficiency and scalability. Key components include User Management, Conversation Service, Monitoring Service, and several backend services like Context Collector, Utterance Analyzer, and Intention Selector that work together to manage conversations and integrate Large Language Models via APIs for dynamic autonomous coaching.

7. Design and Development of the Prototype of the Proposed Framework

7.1 Introduction

To validate the suitability of the COBAC framework for implementing an autonomous coaching system, as well as the COBAC Reference Architecture outlined in Chapter 6, a particular use case was selected. Chapter 7 demonstrates the customisation of the COBAC framework components and the COBAC Reference Architecture for this use case, following steps 1 to 4 of the Design Science Research Methodology.

7.2 DSRM Step 1: Identify Problem and Motivation

Section 7.2 is dedicated to Activity One of the Design Science Research Methodology (DSRM), which encompasses two main areas: “define the specific research problem and justify the value of a solution” (Peffers et al., 2008).

As discussed in Section 4.5.2, the COBAC framework can enhance personalised goal achievement in skill development, performance improvement, and transition periods. However, maintaining a generic application complicates communication with users about the precise benefits of the application. Conversely, a narrowly defined goal may render the application irrelevant to many users. To strike a balance, time management skill development was chosen as the specific goal for the prototype for the following reasons:

1. Relevance: Time management is a sufficiently broad topic, and most working adults can identify opportunities to enhance their time management skills.
2. Scope: Among coaching practitioners, time management improvement is often seen as a 'cover topic' because it typically encompasses underlying specific issues, such as poor prioritisation, the inability to say no, a lack of confidence, or general communication challenges.

Time management is also a suitably specific topic to initiate a coaching engagement, allowing for the testing of the prototype. Therefore, the research problem was defined as 'improving personalised goal achievement in time management skills', and the value of the solution for users lies not only in improved time management but also in the additional 'Gains' identified in the Customer Segment Profile – Coaching (Figure 4.3).

Given that the narrower coaching context does not conflict with any existing items in the Initial Problem List from the COBAC framework, the narrowed context can be integrated by adding an additional item that addresses the aspect of time management improvement (p. 64). This revised Problem List is referred to as the Prototype Problem List throughout the remainder of the document.

7.3 DSRM Step 2: Define Objectives

Section 7.3 covers step two of the Design Science Research Methodology (DSRM) model, which delineates the objectives of this stage as “infer the objectives of a solution from the problem definition and knowledge of what is possible and feasible” (Peffers et al., 2008).
The purpose of building the prototype is to gather data to validate the four hypotheses defined in Section 3.3. For hypothesis validation, it is sufficient to construct an application that is adequate for enabling this purpose, which means the introduction of the following additional constraints:

1. Application Complexity
    - C.11 - To keep complexity manageable, only functionalities that directly relate to at least one of the four hypotheses need to be implemented.
2. Process Limitation
    - C.12 - A lightweight Coaching Administration process is necessitated that fulfils the research data gathering purpose and adheres to ethical norms; however, a comprehensive Coaching Administration process is not required for the prototype as delineated in Section 5.4.2.
    - C.13 - The research design is aimed at a single coaching session; thus, the prototype should concentrate on the Coaching Session process (Section 5.4.4), and elements of

the objectives and solution requirements that cover Coaching Engagement (Section 5.4.3) related aspects may be omitted.

The adjusted Constraint List, which encompasses these additional items (C.11, C.12, C.13), is referred to as the Prototype Constraint List throughout the remainder of the document.

Table 7.1 illustrates the modifications made to the Initial Objective List to formulate the Prototype Objective List, predicated on the elements in the Prototype Problem List and the Prototype Constraint List.

Table 7.1: Prototype Objective List - Changes

Key	Initial Objective	Prototype Objective	Change Description
O.1	Maximising goal attainment	Maximising goal attainment to improve time management	Adjusted to the specific coaching goal: time management (P.64)
O.2	Applies core coaching competences	Applies core coaching competences related to an individual coaching session	Adjusted to consider the single session constraint (C.13)
O.3	High availability	High availability	No change
O.4	Fosters social connectedness through private community features for users pursuing similar goals	-	Due to the single session constraint (C.13), this feature is not required in the prototype.
O.5	Regular and transparent reporting on effectiveness, efficiency, and adherence to ethical standards	Transparent reporting on effectiveness, efficiency	Effectiveness and efficacy are measured as part of the data collection, but they are not regular. Continuous adherence to ethical standard is not related to the hypotheses (C.11), hence removed.
O.6	Service consistency continuous improvement	-	Due to the single session constraint (C.13), this feature is not required in the prototype.
O.7	High-resource language enablement	English, Spanish, and German languages are enabled for conversation	The research design aims to validate only these three languages (C.11). Other languages are not tested.
O.8	Minimising direct cost	Minimising direct cost	No change
O.9	Service usage flexibility: conversation timing, duration, and capacity	Service usage flexibility: conversation timing, and capacity	Duration flexibility in the prototype is not required because it influences the data collection costs and distorts the

			data comparability; hence a fixed session duration was applied during data collection.
O.10	Continuous integration of scientifically validated theories, concepts	Continuous integration of scientifically validated theories, concepts	No change
O.11	Minimizing administrative efforts	Minimizing administrative efforts	No change but considering C.11 in the solution requirements.
O.12	User experience continuous improvement	-	Continuous improvement in a single session design is not required (C.13).
O.13	Dialog quality continuous improvement with the most effective technology	-	Continuous improvement in a single session design is not required (C.13). A dialog quality is aimed at the prototype phase that is good enough for a meaningful single session.
O.14	User autonomy fostering	User autonomy fostering	No change
O.15	User privacy upholding	User privacy upholding	No change

The solution requirements can be derived for the prototype based on the Prototype Problem List, Prototype Constraint List, and the Prototype Objective List.

7.4 DSRM Step 3: Prototype Design and Development

The third step “includes determining the artifact’s desired functionality and its architecture and then creating the actual artifact” (Peffers et al., 2008).

7.4.1 Prototype Solution Requirements

Regarding the adjustments to the objectives, Table 7.2 presents the solution requirements that necessitated modification. Items that remained unchanged are not included.

The changes highlighted in the solution requirements impact the prototype architecture; thus, the reference architecture cannot be utilised and requires adjustments.

Table 7.2: Prototype Solution Requirement List - Changes

Key	Initial Solution Requirement	Prototype Solution Requirement
SR.9	Service usage agreement is managed on two levels: committed time related to the service in general and individual coaching conversation duration	(changed) Service usage agreement is managed by agreement to the coaching conversation duration.
SR.10	Outcomes are collected from interactions with the user related to individual coaching conversations as well as for the overall coaching engagement	(changed) Outcomes are collected from interactions with the user related to individual coaching conversations.
SR.11	At least the following attributes are associated and managed related to outcomes: type, level, description, measure, and status	(changed) At least the following attributes are associated and managed related to outcomes: description
SR.13	Aggregated statistical data in the form of reporting about users, conversations and outcomes is collected and used for continuous service operation and optimisation	(removed)
SR.14	The user can choose from a set of coach profiles that describes behavioural preferences of the coaching service	(removed)
SR.25	Information about the implementation of coaching conversation outcomes between conversations and new learning from the implementation is used as an additional input for the following coaching conversations	(removed)
SR.27	Provides private community features for users pursuing similar goals	(removed)
SR.29	Engagement closure is managed in planned case as well as early termination case, such as shifting user value, unresolvable interest conflict, or reaching system competency boundary	(removed)
SR.39		(added) Coaching conversation languages are English, German, and Spanish.
SR.40		(added) The conversation is optimised to improve time management skills of the user.

7.4.2 Prototype Architecture

Even though changes and additional requirements do not have an impact on the high-level architecture, the removal of SR.13 and SR.27 renders the Monitoring Service and Social Feature components irrelevant in the prototype, as illustrated in Figure 7.1.

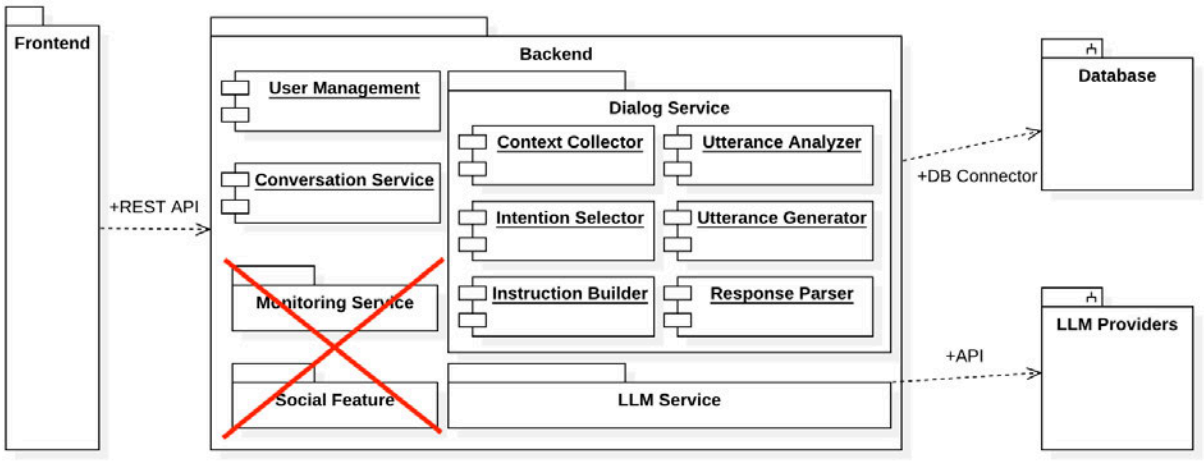


Figure 7.1: Prototype Architecture - Changes

7.4.3 Prototype Technology Selection

Upon establishing the Prototype Architecture, the subsequent phase in designing the application involves choosing the key technologies for the development phase. This selection is guided by principles outlined below:

1. Resource utilization
    a. Effective Development: The selected technology stack should support rapid and efficient development.
    b. Cost: Preference is given to low-cost solutions where feasible.
    c. Fitness to the Developer: The choice of technology is influenced by the developer's skill set and knowledge, as the development is undertaken by the researcher.

2. Security
    a. Security: The adoption of modern, industry-standard solutions that meet the highest security-related technology standards is essential.
    b. Privacy: A high level of privacy is necessary to ensure the security of coaching conversation data, both in transit and at rest.

Considering these principles, Table 7.3 outlines the AWS-based infrastructure components deployed to deliver the prototype's functionality.

Table 7.3: Prototype Infrastructure Components in AWS

Component Name	Description	Region
Route 53	DNS service for the domain	Global Service
CloudFront	CDN service to enable fast access from remote regions	Global Service
S3	Block storage - website static files, website log files, and WAR file for the Elastic Beanstalk instance	eu-central-1 (Frankfurt)
ACM (AWS Certificate Manager)	SSL service for secure communication	eu-central-1 (Frankfurt)
RDS (Relational Database Service)	PostgreSQL db.t2.micro instance; Engine Version 11.16	eu-central-2 (Zurich)
Elastic Beanstalk	Platform: Tomcat 10 with Corretto 17 running on 64bit Amazon Linux 2023/5.1.0	eu-central-1 (Frankfurt)
EC2 for Load Balancing + Elastic Beanstalk	Elastic Computing capacity & OS: Linux t3.micro	eu-central-1 (Frankfurt)
EC2 for RDS	Elastic Computing capacity & OS: Linux t3.micro	eu-central-2 (Zurich)

AWS was chosen as the cloud service provider due to its Elastic Beanstalk service, which allows a developer to upload a Web Application Archive (WAR) artifact. This artifact encompasses the complete front-end and back-end code and enables the deployment in an application server environment within seconds. It also enables the execution of the code with minimal intervention and configuration. The services listed in Table 7.3 reflect a minimalist approach to maintain the infrastructure as simple as possible. Each component provides a specific, indispensable service required for an autonomous coaching service. Data at rest are stored in data centres located in Switzerland, known for their stringent data protection policies. All other services, where regionality can be configured, are deployed in German data centres due to the unavailability of such services in Switzerland. No third parties, other than the LLM provider and cloud service provider, were granted access to sensitive data, upholding the privacy expectations. Consequently, no third-party platform was involved in constructing the front-end or the dialog management system.

Furthermore, Table 7.4 summarizes the tools and services applied throughout the development process.

Table 7.4: Development Tools and Services

Tool / Service	Description
Eclipse	Integrated Development Environment
LM Studio	An application to manage and run LLMs locally
LLaMa 2	An open source LLM that can be run locally
GitHub	Repository to manage source code
ChatGPT	Utilised as a developer co-pilot to speed up the development process

7.4.4 Prototype Design

The overall design approach was predicated on industry-standard best practices and design patterns, except for three areas: dialogue management, LLM prompt management strategy, and LLM model selection.

To begin, the dialogue management solution was developed around the key components of the Dialogue Service depicted in Figure 7.1. Context collection in the prototype was confined to previous utterances within the single conversation, the conversational goal, a prototype persona description of the autonomous AI coach, and communication guidelines. The Utterance Analyzer component retrieved data from the LLM regarding the last utterance, covering 10 aspects pertinent to managing the dialogue. These aspects included the user's intention with the last utterance, exceptions in the conversation that might derail the dialogue, and the detection of ethical concerns, such as the presence of discriminatory language. The Intention Selector identified one or two intentions for the next generated utterance based on the analytical data from the LLM, as listed in Table 5.3, specific to the session. The Utterance Generator and Instruction Builder components initiated a second API call to the LLM to procure utterances generated in the user's language. The Response Parser component interpreted the JavaScript Object Notation (JSON) response from the LLM, which was generated on the user language. This cycle repeated from the initial opening statement produced by the conversational agent to the conversation's conclusion.

Following on from that, the prompt engineering approach for the prototype remained simplistic but incorporated various techniques to enhance the quality, response time, and cost-efficiency of the LLM's response:

1. Utilising ChatGPT to optimise prompts involved sending examples of complete prompts to ChatGPT, requesting suggestions on how to refine the prompts to be as concise as possible while eliciting the correct response.
2. JSON variables: All requests to the LLMs were constructed to elicit a valid JSON-formatted response, which facilitated effective parsing and reduced prompt ambiguity.
3. Retrieval-Augmented Generation (RAG): Context-specific information was integrated into individual prompts. For example, the initial conversation goal was included in the conversation analysis request once it was aligned with the user.
4. Few-shot learning involved providing the LLM with examples that precisely demonstrated the expected outcome. The number of examples provided varied between two and six, depending on the complexity of the specific variable.

Finally, the decision regarding which LLM model to utilise for the prototype was postponed until the last moment, due to the rapidly evolving LLM landscape and the escalating low-level requirements emanating from the prototype's development and testing. The final prototype utilised GPT-4-1106-preview for utterance analysis-related calls and GPT-3.5-turbo-1106 for utterance generation.

7.4.5 Prototype Development

The delivered components could be divided into three major areas:

1. Frontend
2. Backend
3. Operational Scripts

The frontend structure focuses on providing multi-language support, application styling, functionality via JavaScript, and various image assets for a rich user experience. The application’s frontend file structure is organised as follows:

1. HTML Files: These files serve as the different language versions or pages of the application:

- index.html: The main HTML page that includes the consent collection.

- Language-specific HTML files like coachapp_time_de.html, coachapp_time_en.html, coachapp_time_es.html: These represent the same application in different languages.
- step_03.html: A final, third step to incorporate the survey related data collection.
- error.html: A custom error page for handling exceptions.

3. Images: This folder contains various image assets used throughout the application. They include icons of different sizes for web app icons, and favicons.
4. Styles: This folder holds the styling files for the application, with app.css as the main CSS file used to style the app's user interface.
5. Scripts: This directory includes JavaScript files that add interactivity and functionality to the application, with the main functionality covered in the app.js file.
6. Other Files and Folders:
   - WEB-INF: This folder contains configuration files like web.xml, which defines the application's settings, routes, and servlets.
   - manifest.json: A JSON file is used to define metadata for the web app, such as its name, icons, theme colour, and more, to enhance its appearance when installed as a Progressive Web App (PWA).

The backend structure is modular and layered, with clear separation between data access, business logic, and request handling, ensuring scalability and maintainability as shown in Figure 7.2.

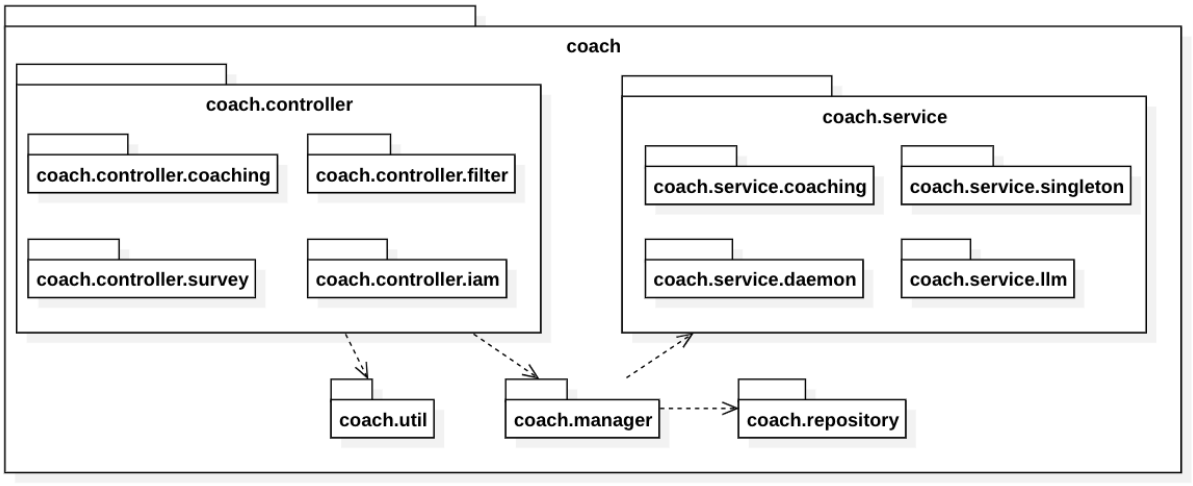


Figure 7.2: Prototype Backend Package Overview

The backend is organised into several JAVA packages, each handling distinct responsibilities:

1. coach: The root package that contains all the sub-packages and core classes for the application.
2. controller: This package handles incoming HTTP requests and maps them to the appropriate services and actions. It is divided into sub-packages like iam (for user authentication and management), survey (for survey-related functionality), and coaching (for coaching-related functionalities). Each sub-package contains controllers, which process user interactions.
3. manager: The manager package is responsible for coordinating higher-level logic, combining functionality from the repositories and services. For instance, UserManager manages user-related operations, while ConversationManager oversees coaching conversations.
4. repository: This package contains various repositories for interacting with the database and performing CRUD operations. It includes entity-specific repositories such as UserRepository, SurveyRepository, and ConversationRepository. It also holds the entity classes like User, Conversation, Utterance, and Intention, which define the data models for the application.
5. service: This package contains the core business logic, such as handling dialogues, managing LLM interactions, and daemon tasks. Key services include DialogService, LLMService, and CoachingService, which implement the main functionalities of the application. The daemon sub-package deals with background tasks like handling reports or translations.
6. util: This package provides utility classes that support the application, such as TemplateFileReader, JSONFlatter, and ValueExtractorFromJSON, which help in processing and managing various data formats.
7. service/singleton: Contains singleton classes, such as ContextSingleton, ensuring shared state and centralized access across the application. These classes provide functionality for maintaining consistent states, like user context or intentions.

Lastly, the operational scripts are responsible to deploy the application WAR file that is created by the maven build tool through the Eclipse IDE and to create the Database schema via DDL scripts.

7.4.6 Prototype Development Challenges

There were various challenges during the initial development and testing of the prototype, among which three stand out in terms of severity, probability, and impact.

To start with, the response times of the LLMs vary significantly among different models. The LLM providing the highest quality answers to complex prompts, which include a significant amount of contextual information necessary for leading a high-quality coaching conversation, is excessively slow for conducting a conversation with a human. Models were providing response times for the initial individual requests ranging between 15 and 100 seconds; however, the ideal response time for leading a smooth, human-like conversation in a chat format is around 8 seconds. As a result, significant refactoring and optimization were undertaken before the final version of the prototype was released. Measures to address the problem included prompt simplification, executing daemons that run in the background parallel to the ongoing conversation, and building a complex logic to manage examples used for realizing the few-shot learning approach.

After that, testing the prototype also proved challenging due to the high costs associated with testing and the non-deterministic nature of the generated conversation. Although these aspects are relevant for all LLM-based conversational agents, in the case of coaching, the system's response is only remotely related to any specific information provided by the user. The impact of this issue was that only manual testing was applied throughout the development phase, which is acceptable for a prototype; however, a scalable, maintainable system requires automated testing to provide consistent quality. At the time of writing this thesis, there seem to be multiple approaches in the experimental stage to cover automated testing of LLM-based conversational agents, though they are in a very early stage.

Lastly, the LLMs are not mature yet for a real-time application context due to multiple issues:

1. Service availability: LLMs appear to refuse individual requests randomly throughout the day with error messages, such as "internal server error". This phenomenon indicates that the LLM providers are not yet able to provide a high-availability

infrastructure. There are two possible approaches to overcome the problem: resending the request, which would result in a slower response time for the user or sending the request to an alternative provider. Neither of these approaches was implemented in the prototype.

2. Response structure corruption: Although most responses are composed based on the predefined response structure, which is also underpinned with examples, some answers fail to follow these expectations and are thus unusable as a response. As an improvement, a more robust and failure-tolerant response parser was implemented to recover parts of the response even if the JSON structure was corrupted.
3. Response language: Despite explicit instructions in the prompt that the response should be in a specific language, the LLM responses were not adhering to these requirements. To address the problem, instead of providing response examples for few-shot learning solely in English, the response examples were translated once into all target languages. This ensured that while the request language remained English, the example answers were already provided in the expected target language, resulting in higher language adherence.
4. LLM API structure differences: The system was tested with multiple LLMs from different vendors, such as OpenAI, Anthropic, and Mistral. However, at the time of testing, each model required a different API request, response, and header structure. For the provider selection tests, a more robust request preparation component was implemented to hide the complexities of managing communication with different target systems.

7.4.7 Outcome

As a result of these major adjustments, the final version of the prototype was able to handle the coaching conversations despite the challenges described in the previous section. Figure 7.3 shows three screenshots using the opening utterances of the three high-resource languages: English, German, and Spanish.

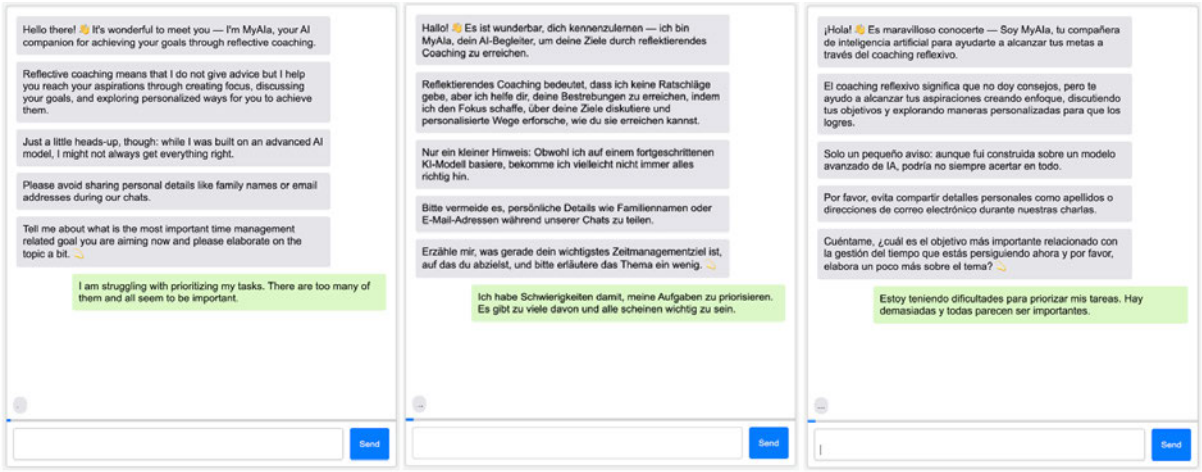


Figure 7.3: Prototype User Interface

To produce this result, a combination of programming and other specialized languages and technologies was employed in the development and operation of the application. Specifically, this included the Java programming language; SQL as a query language; HTML for structuring the content; JavaScript (JS) as a client-side scripting language; Shell script (SH) for command-line automation; and CSS for styling. These tools worked together to enable the system's functionality, presentation, and interaction. Table 7.5 shows a distribution of the used technologies, grouped by the number of lines of code written and the number of files.

The numbers related to the prototype complexity contain not only the conversational agent-related code but also the functionality related to the consent gathering and the implementation of the survey to collect data for the hypothesis validation after the development phase is concluded.

Table 7.5: Prototype Complexity

Languages , Technology	**Lines of Code**	**Number of Files**
Java	7'203	83
SQL	4'430	8
HTML	1'659	15
JS	438	1
SH	238	4
CSS	189	1
	14'157	112

7.5 DSRM Step 4: Prototype Demonstration - Data Collection

After the third stage of the Design Science Research Methodology (DSRM), which involved prototype development and design, the fourth stage covers the activity to “demonstrate the use of the artifact to solve one or more instances of the problem” (Peffers et al., 2008).

7.5.1 Initial Survey Design

The cross-sectional survey design employed a three-step process:

1. Consent Page: The landing page of the study was a language-specific page containing all major aspects of the study, designed to inform participants and obtain consent for participation. Information provided included the research aim, approach, participation and technical prerequisites, task related process, estimated duration, termination conditions, task description, task completion success criteria, payment terms, ethical standards and risks, limits of confidentiality, and contact information.
2. Chatbot Page: Upon clicking the "I consent" button on the consent page, the language-specific chatbot page loaded as illustrated in Figure 7.3. The conversation was set for 10 minutes, with leftover time measurement adjusted for inactivity. After the 10-minute conversation, a "thank you" page appeared, concluding this step with a button directing the user to the first survey page.
3. Survey Pages: The survey was implemented manually in English, without the use of external survey tools. Completion of the final question generated a code with a static prefix for validation purposes and a dynamically generated user identifier to enable the filtering out of information related to invalid datasets.

Table 7.6 provides an overview of the 43-question survey structure utilised for data collection.

Table 7.6: Survey Structure

Survey Block	Number of Questions	Question Type	Description
1	1	Open	Attention check: General feedback on the usefulness about the chatbot experience. Not used for the hypothesis validation, but it was included based on the suggestion originated at the MTurk best practices (Aguinis et al., 2021).
2	3	Likert scale	Coaching Effectiveness related questions (De Haan et al., 2016)
3	1	Likert scale	Attention check
4	8	Likert scale	The first block of the UTAUT (Venkatesh et al., 2003) related questions.
5	1	Likert scale	Attention check
6	14	Likert scale	The second block of the UTAUT (Venkatesh et al., 2003) related questions.
7	1	Open	Attention check
8	8	Likert scale	The third, final block of the UTAUT (Venkatesh et al., 2003) related questions.
9	1	Open	Attention check
10	5	Likert scale	USR (Mehri & Eskenazi, 2020) related questions.

Lastly, Table 7.7 contains the detailed questions grouped by the measurement constructs.

Table 7.7: Survey Questions

Question Block	Description
Coaching Effectiveness	I have been successful in creating reflective space for me. I have been successful in creating new insights for me. I would consider this upcoming coaching journey successful.
UTAUT	Using the chatbot would help me to achieve my goals more quickly. Using the chatbot would enhance my effectiveness to achieve my goals. Using the chatbot enhances my motivation to achieve my goals. I would find the chatbot useful for improving my skills to achieve my goals. I would find the chatbot useful for my future coaching needs. Overall, I believe the chatbot is easy to use. Using the chatbot for coaching takes too much time. I would find it easy to communicate with the chatbot. Learning to use the chatbot for coaching is easy for me. My coaching activities with the chatbot are clear and understandable. I would find it easy to achieve goals with the chatbot. People who are important to me think that I should use a chatbot for coaching. People who can affect my behavior use a chatbot for coaching. I know other people who use a chatbot for coaching. I think that using a chatbot for coaching is fashionable. I have the resources necessary to use the chatbot for AI coaching. I have the knowledge necessary to use the chatbot for AI coaching. I think that using the chatbot for coaching fits well with the way I like to develop myself. I have easy and regular access to the internet. I have a convenient place where I could use the chatbot for AI coaching. If I have problems using the chatbot for coaching, I could solve them very quickly. I feel that confidentiality concerns inhibit my willingness to use the chatbot for coaching. I believe my privacy is adequately protected when using the chatbot. I do not trust the chatbot as a coach. I prefer a human coach above the chatbot coach. I believe it is risk-free to use the chatbot. I intend to use the chatbot in my future coaching interventions. I predict I will use the chatbot for coaching. I plan to use the chatbot for coaching in the next 2 months. I would recommend the coaching chatbot to a friend.
USR	Do you find the chat messages generated by the chatbot understandable? Do the chat messages generated by the chatbot seem to be something that a person would naturally say? Do the chat messages generated by the chatbot serve as a valid continuation of the preceding conversation? Do you find the chat messages generated by the chatbot interesting? Given your previous answers, what is your overall impression of the dialog quality?

7.5.2 Pilot Survey

Based on the survey question described in Table 7.7, a pilot data collection phase was planned, involving 10 participants using the English version of the chatbot to assess the overall data collection process. Table 7.8 summarises the key findings and reaction to the outcomes of the pilot phase.

Table 7.8: Data Collection Pilot Phase - Challenges and Adjustments

Problem Description	**Reaction / Adjustment**
The qualitative, open-ended questions and the analysis of the initial conversations confirmed P.1 in the Prototype Problem List, indicating poor awareness of coaching among clients, with some clients persistently prompting the chatbot for advice.	The task description, task title, consent page, and opening statements in the chatbot were all adjusted to emphasise the chatbot's approach (reflective coaching, not providing advice).
Participants requested an immediate response in the event of a failed attention test, to avoid wasting time on a test where their responses would ultimately be invalidated.	The survey workflow was modified so that if an attention test response was invalid, the process would be terminated immediately.
Multiple participants were completing the survey by consistently selecting the same response (for example, always choosing 'strongly agree').	Although attention tests were supposed to detect this behaviour, there were instances where attention tests were answered correctly but other responses were not. Therefore, an additional validation step was implemented that invalidated the dataset when the same response was chosen nine times consecutively.
The spending limit for the OpenAI API service was reached, despite automated charging being configured. This did not occur immediately, resulting in the disruption of a complete conversation.	Prior to the main test, the OpenAI account was charged manually in advance to mitigate the risk of this failure. Additionally, measures were taken to further enhance the fault tolerance of the application.

7.5.3 Main Survey

Considering the variety of challenges encountered during the pilot phase, a staggered approach to data collection was adopted to reduce risks associated with the data collection stage. Data were collected in multiple batches, with each consisting of 10-15 participants, sequentially initiated based on the English version of the chatbot. This method also had the added advantage of accommodating participants with atypical work schedules, such as those only available on weekends, thereby enhancing the diversity of the participant pool. Data

collection for the study commenced in December 2023 and concluded in February 2024, incorporating several weeks' pause due to the year-end holiday season.

During the main data collection phase, an issue was identified concerning language groups. Although Amazon Mechanical Turk provides the option to select participants with German language skills, there was negligible response to the tasks posted for this demographic. Consequently, an alternative crowdsourcing platform, Clickworker, was utilised, facilitating the inclusion of German-speaking participants.

The results of the data collection phase are summarised in Table 7.9, detailing the number of abandoned conversations, completed conversations, instances of invalid survey completion, and the overall valid dataset.

Table 7.9: Data Collection Outcome Overview

Language	Initiated Conversations	Abandoned Conversations	Conversations Completed	Invalid Survey	Valid Data Set
English	92	16	76	26	50
German	69	15	54	13	41
Spanish	67	18	49	17	32
Total	228	49	179	56	123

The final sample size for the study, reflecting the total valid dataset available for analysis to validate the four hypotheses, is 123. 21.49% of all initiated conversations with the chatbot were abandoned and the rate of abandonment was consistent across the three languages: Spanish (26.86%), German (21.73%), and English (17.39%). Of the 179 participants who completed the conversation, 56 did not fill the surveys. The initial objective was to obtain at least 40 valid survey results for each language group, which was successful in case of the English and German groups, but it was not successful for the Spanish group even though the data collection was ongoing for several weeks.

In addition to the quantitative survey data set, application log files and a database dump were also collected to facilitate a more in-depth analysis of the underlying data, if necessary for hypothesis validation.

7.6 Chapter Summary

Chapter 7 establishes the validation of the COBAC framework for an autonomous coaching system through the Design Science Research Methodology (DSRM), specifically applied to a prototype tailored to time management skill development.

Section 7.2 identifies the problem of enhancing personalized goal achievement in time management, selected for its wide applicability and representation of a spectrum of underlying issues such as prioritization. This section updates the Initial Problem List to include time management, forming the Prototype Problem List.

Section 7.3 outlines the prototype's objectives and constraints, focusing on a single coaching session and pertinent functionalities for hypothesis testing, which leads to the formation of the Prototype Constraint List. Adaptations made to the Initial Objective List to accommodate the prototype's narrowed focus, resulting in the Prototype Objective List that directs the prototype's solution requirements.

Section 7.4 describes the design and development process of the prototype. It details changes to the prototype's solution requirements, noting adjustments and removals of certain requirements and the addition of new language capabilities. Despite alterations not impacting the high-level architecture, the elimination of certain solution requirements has made some components, like the Monitoring Service and Social Feature, obsolete. The technology selection for the prototype is informed by considerations of resource utilization, cost, fitness for the developer's skill set, security, and privacy, with an AWS-based infrastructure being chosen for its effectiveness. The design phase addresses dialogue management, prompt management, and model selection, with specific strategies for improving LLM response quality and efficiency. Challenges encountered include varying response times of different LLMs, testing difficulties due to cost and non-deterministic conversation generation, and the immaturity of LLMs for real-time applications. The final prototype, overcoming these hurdles, utilises multiple programming languages and reflects comprehensive complexity as indicated by the distribution of lines of code across different languages.

Section 7.5 describes a three-step process for a cross-sectional survey design beginning with a consent page, followed by a chatbot interaction, and concluding with survey pages, which included a 43-question structure detailed in Table 7.6. The pilot survey in Section 7.5.2 involved 10 participants and led to adjustments in the chatbot's explanation of its coaching role, immediate feedback on failed attention tests, and additional validation to prevent uniform response patterns, addressing the limitations encountered. The main survey conducted in batches to enhance participant diversity and accommodate different schedules, faced challenges with German language participant response rates, leading to the use of Clickworker to source participants. The data collection, completed between December 2023 and February 2024, resulted in 123 valid datasets out of 228 initiated conversations.

8. Design and Development of an Optimised New Framework

8.1 Introduction

Chapter 8 covers two aspects of the thesis: the data analysis, which is referred to in Step 5 of the Design Science Research Methodology (DSRM) as Prototype Evaluation, and the incorporation of the learnings from the prototyping phase into the new optimised framework. The fifth stage of the DSRM approach covers the activity of "observe and measure how well the artifact supports a solution to the problem" (Peffers et al., 2008). The data analysis part is structured around four hypotheses, while the enhancement of the framework follows the framework components.

8.2 Descriptive Statistics, Data Screening, Data Cleaning

The total number of participants who provided a complete and valid response to the survey is 123, which is distributed among the three languages and countries as Table 8.1 shows.

Table 8.1: Participant Descriptive Statistics

Language	Country	Valid Data Set	% of Total Population
English	USA	50	40.65%
German	Germany	41	33.33%
Spanish	Spain	32	26.02%
Total		123	100%

Apart from the fact that the participants were sourced exclusively from crowd work platforms (Clickworker, MTurk), no additional information was gathered about the participant demographics.

Considering that the nature of the gathered data is quantitative with a fixed value range (1–7), data validation was implemented in the data gathering user interface, which did not allow progression with the survey without entering a valid value. As a result, there was no impermissible data in the gathered survey results.

Similarly, it was technically not possible to skip questions unless the participant abandoned the questionnaire completely. Consequently, there was no need to compensate for missing data.

Regarding participant misconduct, such as suspicious response patterns like straight-lining, two measures were applied:

1. If a participant set the same value for nine consecutive items, the survey response was considered invalid.
2. For each response, a standard deviation was calculated, and the minimum value was 0.46, which is well above the 0.25 threshold commonly used for survey response validation.

The only post-collection data manipulation that was required concerned the reverse-scored items. These related to the items EE2, PR1, PR3, and PR4, where the questions were formulated in such a way that the response would measure the opposite of the measured item. In these cases, the response values had to be reversed by applying the following algorithm:

8 – 'Participant response value' = 'new value'

8.3 Hypothesis 1 (H1) - Feasibility

Hypothesis 1 (H1), which posited the feasibility of developing an autonomous coaching system based on the COBAC framework, was not evaluated through statistical data or surveys but rather through the presentation of a user-testable chatbot prototype.

Based on the adjustments of the initial COBAC framework elements, which resulted in the Prototype Problem List, Prototype Constraint List, Prototype Objective List, and Prototype Solution Requirement List, it was feasible to develop the chatbot. This process of adjustment showcased not only that the framework is suitable for the practical development purposes, but also that it is highly modifiable.

However, the methodological limitation of testing the chatbot in a single conversation per participant and for only 10 minutes restricts the generalizability of the conclusion that the COBAC framework is feasible for developing autonomous coaching agents. Despite these limitations, Hypothesis 1 (H1) is considered confirmed.

The technological challenges discussed in Section 7.4.6 should inform the improvement of future iterations of the COBAC framework.

8.4 Hypothesis 2 (H2) - Efficacy

Hypothesis 2 (H2), which proposed that the COBAC-based autonomous coaching system is effective in achieving coaching outcomes, was evaluated by collecting data on Coaching Effectiveness (CE) measures.

Table 8.2 displays the overall average for the three measured Coaching Effectiveness items (CE1, CE2, CE3) alongside language-based averages.

Table 8.2: Coaching Effectiveness Average

Language	**Coaching Effectiveness Average**	**CE1 (reflective space)**	**CE2 (new insight)**	**CE3 (successful journey)**
English	5.253	5.2	5.3	5.3
German	3.894	4.2	3.9	3.5
Spanish	4.979	4.9	5.1	4.9
Overall Average	4.729	4.8	4.8	4.6

The agreement scale for all three items had potential descriptive texts with corresponding numeric values: Very Strongly Disagree (1), Strongly Disagree (2), Disagree (3), Neutral (4), Agree (5), Strongly Agree (6), and Very Strongly Agree (7). The overall average indicates that participants' perceptions of the chatbot's 10-minute coaching conversation's effectiveness were on average closest to "Agree". The variance in the averages for the three items was low, indicating no specific aspect measured by the items significantly influenced the results.

However, a notable difference exists in the average for the German participant group compared with the other groups. Figure 8.1 provides a detailed language-based summary, illustrating the distribution of the three selected options by individual languages. The range of individual measurement item values was between 1 and 7, meaning the potential sets of values for all three measures span from 3 to 21.

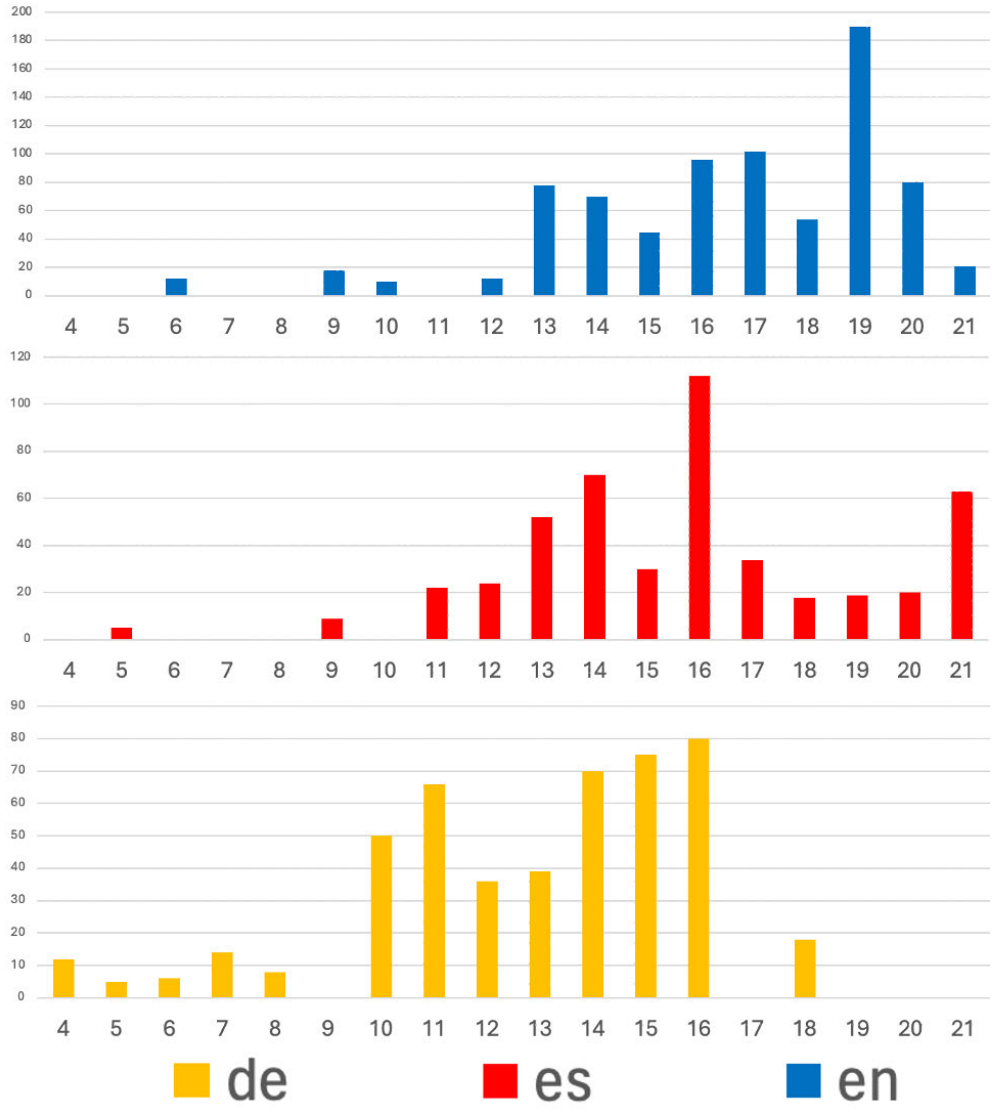


Figure 8.1: Coaching Effectiveness Summary Distribution Among Language/Country

Conversational data analysis did not reveal any specific reason for this discrepancy; however, the significant variance in the German group's responses warrants future research.

While several other studies have utilised the Coaching Effectiveness measure in human-to-human and AI coaching contexts (De Haan, Gray, & Bonneywell, 2019; De Haan et al., 2020;

Terblanche et al., 2022), comparing results is not methodologically sound due to significant differences in study design. These differences include longer coaching durations, longitudinal approaches, homogenous participant groups (e.g., executives, business school students), and the use of more mature AI coaching platforms developed over many years as opposed to presenting a service prototype. Of the aforementioned studies, only one - a human-to-human near-randomized controlled trial in the healthcare sector - published an actual measured value of 5.02 at the first point of measurement (De Haan et al., 2019).

Despite these challenges in comparability, H2 is deemed valid, considering that even a 10-minute exposure to an AI coaching prototype resulted in an overall agreement among participants regarding the prototype's effectiveness.

8.5 Hypothesis 3 (H3) - Dialog Quality

Hypothesis 3 (H3), which posited that the dialogue quality provided by the COBAC-based system is independent of the high-resource language used for the conversation, was validated through the five measurement items delineated by the USR model. These are characterised by understandability (USR1), naturalness (USR2), valid conversation continuation (USR3), being interesting (USR4), and general dialogue quality (USR5).

Table 8.3 illustrates the overall average of all measured USR items, as well as the language-based averages.

Table 8.3: Dialog Quality Average

Language	USR Average	USR1	USR2	USR3	USR4	USR5
English	5.6980	5.76	5.59	5.41	5.76	5.96
German	3.7707	5.07	3.83	3.34	3.44	3.17
Spanish	5.1125	5.91	4.78	5.06	4.94	4.88
Overall Average	4.9122	5.57	4.8	4.64	4.78	4.76

All five measured items utilised a 7-option scale, with specific text representing the measured quality. For example, in the case of USR1 (understandability), the option text and assigned values were as follows: Not Understandable at All (1), Very Difficult to Understand (2),

Somewhat Difficult to Understand (3), Neutral (4), Somewhat Easy to Understand (5), Easy to Understand (6), and Very Easy to Understand (7).

The overall average indicates that understandability (USR1) was rated by participants significantly better compared to all other characteristics, with valid conversation continuation (USR3) being the worst-performing indicator, albeit not far behind the others. The overall average of 4.9122 suggests that while there is room for improvement in dialogue quality, a basic prototype can still achieve a (somewhat) acceptable level of dialogue quality.

Comparison of these results with other studies is not feasible, as at the time of writing this thesis, no other study utilising USR in an AI coaching context to measure conversation quality could be found.

Like the Coaching Effectiveness, the German participant group rated the prototype significantly lower than the other groups across all indicators. The characteristic with the smallest discrepancy was understandability; however, all other aspects were rated significantly lower. Contrary to the Coaching Effectiveness rating, the Spanish group also rated the non-understandability characteristics significantly lower compared with the English language group. The data unequivocally indicates that dialogue quality is not independent of the high-resource language used for the conversation; thus, H3 is rejected.

8.6 Hypothesis 4 (H4) - Language Influence on Adoption Factors

Finally, Hypothesis 4 (H4) posits that the language of the coaching conversation does not significantly influence the factors contributing to the intention to use the COBAC-based chatbot. The partial least squares structural equation modelling (PLS-SEM) was utilised for analysis because this approach has become crucial for efficiently managing complex and non-normal data in business and social sciences research (Joe F. Hair, Sarstedt, Hopkins, & Kuppelwieser, 2014). It is also used in all three studies that theorised the UTAUT framework (Arenas-Gaitán et al., 2015; Bhatnagr & Rajesh, 2023; Venkatesh et al., 2003) and this analytical tool is employed in several AI coaching research (Terblanche & Cilliers, 2020;

Terblanche & Kidd, 2022). Version 4.1.0.0 of the SmartPLS 4 software (Ringle, Wende, & Becker, 2024) was used for the model assessment with a 5% significance level ($p < .05$) to validate the hypothesis. PLS-SEM undergoes a dual-phase evaluation process, namely specifying the measurement model and assessing the structural model (Ringle, Sarstedt, Mitchell, & Gudergan, 2020).

8.6.1 Measurement Model

According to the first version of the UTAUT model (Venkatesh et al., 2003), Behavioural Intention (BI) is influenced by Performance Expectancy (PE), Effort Expectancy (EE), Social Influence (SI), and Facilitating Conditions (FC). Perceived Risk (PR), as defined by UTAUT3 (Bhatnagr & Rajesh, 2023), was an added construct because risks associated with confidentiality and privacy are considered significant in the specific context of coaching. The employed model was further extended with the construct Language (LG), to examine the difference in these factors across language groups.

Other researchers suggested 0.6 as a minimum value for the factor loadings (Joseph F Hair, Black, Babin, Anderson, & Tatham, 1998), hence the following indicators were removed from the model: EE2 (0.153), FC4 (-0.03), and FC5 (0.328). Figure 8.2 illustrates the updated measurement model that was used to validate Hypothesis 4.

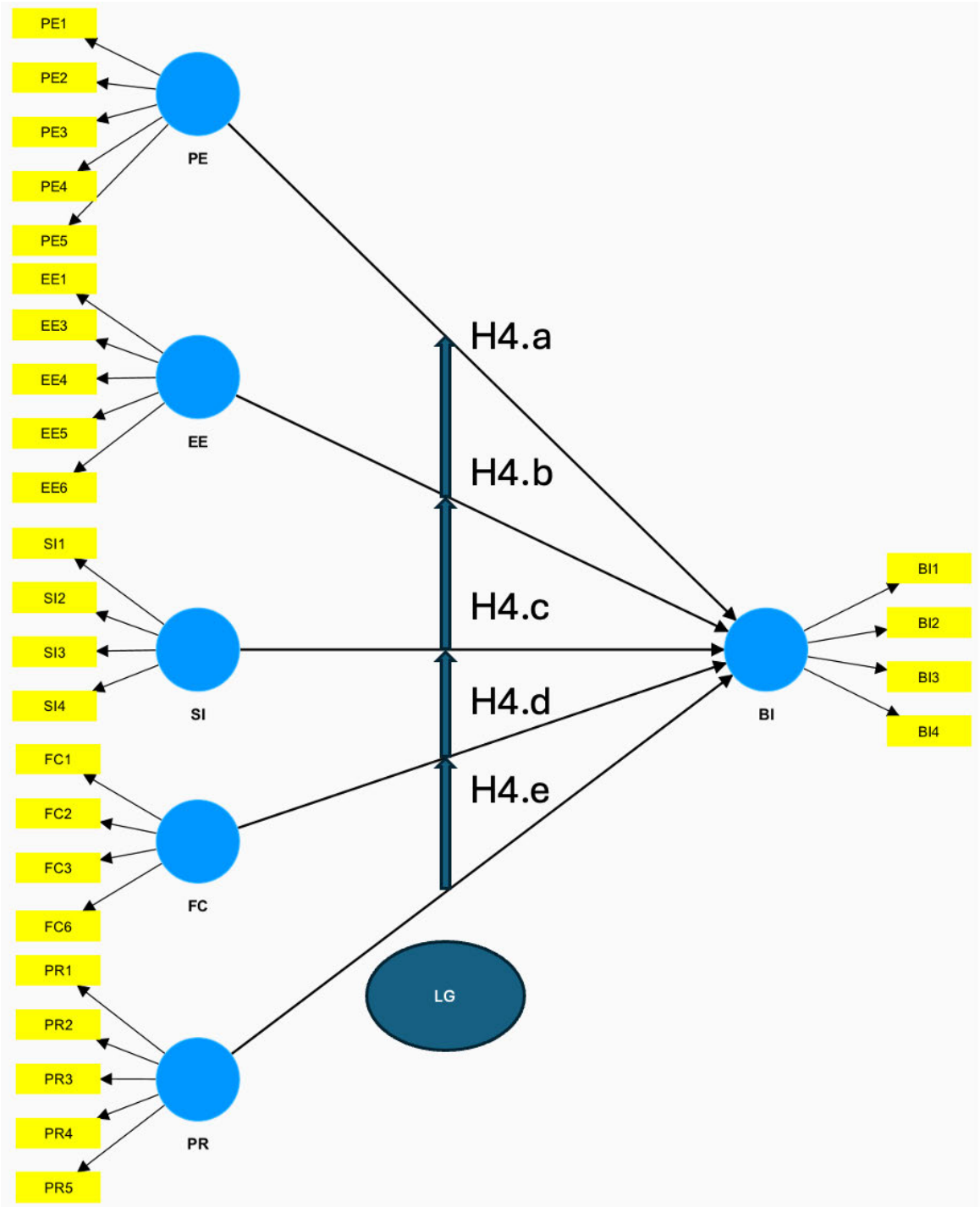


Figure 8.2: Conceptual Model and Hypothesis

The PLS-SEM algorithm was performed based on the updated model, which produced the outer loadings, Cronbach's Alpha, Composite Reliability (CR), and Average Variance Extracted (AVE) values shown in Table 8.4. Values in brackets show the analysis result without the removal of the three items, indicating an improvement of the overall construct. Perceived Risk (PR) was conceptualised as a formative construct due to its composition from distinct risk

components that collectively contributed to an overall risk assessment. Consequently, this approach meant that reliability metrics for perceived risk are not presented.

Table 8.4: Item Loadings, Reliability and Convergent Validity

Item	Outer Loadings	Cronbach's Alpha	Composite Reliability (CR)	Average Variance Extracted (AVE)
Behavioral Intention (BI)		0.933 (0.933)	0.936 (0.936)	0.834 (0.834)
BI1	0.910			
BI2	0.953			
BI3	0.896			
BI4	0.893			
Effort Expectancy (EE)		0.841 (0.792)	0.898 (0.899)	0.597 (0.501)
EE1	0.740			
EE3	0.794			
EE4	0.690			
EE5	0.798			
EE6	0.833			
Facilitating Conditions (FC)		0.736 (0.722)	0.802 (0.768)	0.536 (0.368)
FC1	0.655			
FC2	0.712			
FC3	0.809			
FC6	0.744			
Performance Expectancy (PE)		0.953 (0.953)	0.954 (0.954)	0.842 (0.842)
PE1	0.924			
PE2	0.916			
PE3	0.904			
PE4	0.925			
PE5	0.919			
Social Influence (SI)		0.883 (0.883)	0.896 (0.896)	0.742 (0.742)
SI1	0.892			
SI2	0.908			
SI3	0.885			
SI4	0.751			

The Cronbach's Alpha and Composite Reliability (CR) indicators were above 0.7 and the average variance extracted (AVE) values were above 0.5, hence the model is considered acceptable from the construct validity and reliability perspective.

The discriminant validity criterion suggested by Fornell and Larcker (Fornell, C., & Larcker, 2016) was also tested, as shown in Table 8.5.

Table 8.5: Discriminant Validity

	BI	EE	FC	PE	SI
BI	**0.913**				
EE	0.730	**0.772**			
FC	0.720	0.779	**0.732**		
PE	0.853	0.742	0.655	**0.918**	
SI	0.798	0.596	0.616	0.806	**0.861**

The difference of -0.007 between 0.772 (column and row EE) and 0.779 (column EE and row FC) indicates that the EE and FC constructs have a very low risk of not capturing distinct concepts. Considering the low numerical difference, the fact that the survey questions indeed conceptualise different aspects, and that the constructs are validated in other studies, this discrepancy is accepted.

8.6.2 Structural Model

The R^2 value of 0.833 for BI supports the model's in-sample predictive power (Sarstedt, Ringle, Henseler, & Hair, 2014), as it is above the required level of .10 (Falk & Miller, 1992). The significance of direct paths is determined through the Bootstrap resampling method with 5000 resamples and the results are presented in Table 8.6.

Table 8.6: Path Coefficients in the Full PLS-SEM Model

Item	Original Sample (O)	Standard Deviation (STDEV)	T Statistics (|O/STDEV|)	p Value	Path Coefficient
EE	0.073	0.098	0.744	0.457	0.073
FC	0.166	0.069	2.416	0.016	0.166
PE	0.358	0.112	3.186	0.001	0.358
PR	0.178	0.082	2.166	0.030	0.178
SI	0.252	0.062	4.097	0.000	0.252

The path coefficients of the full model indicate a very strong direct relationship between Performance Expectancy (PE) and Social Influence (SI) to Behavioural Intention (BI) ($p \leq .001$), a strong relationship ($p < .05$) for Facilitating Conditions (FC) and Perceived Risk (PR), and a negligible relationship for Effort Expectancy (EE). This result aligns with previous findings from a study investigating the moderating effects of age and gender in an AI coaching context (Terblanche & Kidd, 2022).

However, when the data sets are grouped by language (LG), significant differences emerge, as shown in Table 8.7.

Table 8.7: Path Coefficients by Language (LG)

	English		Spanish		German	
Item	**p Value**	**Path Coefficient**	**p Value**	**Path Coefficient**	**p Value**	**Path Coefficient**
EE	0.171	0.242	0.247	0.204	0.818	-0.036
FC	0.829	-0.037	0.136	0.206	0.998	0.000
PE	0.501	0.141	0.589	-0.108	0.000	0.518
PR	0.830	0.043	0.101	0.346	0.031	0.343
SI	0.002	0.562	0.001	0.419	0.151	0.173

Considering the 5% significance level ($p < .05$), no single factor can be consistently confirmed across all language groups. For example, Social Influence is significant in the English and Spanish groups but not in the German group. Simultaneously, Perceived Risk (PR) exhibits a relatively low path coefficient in the English group but is significant elsewhere. Consequently, Hypothesis 4 (H4) and all its sub-items (H4.a, H4.b, H4.c, H4.d, H4.e) are rejected.

The significant differences in the data might be attributed more to cultural than linguistic differences. Although the UTAUT model was validated against cross-cultural differences (Oshlyansky, Cairns, & Thimbleby, 2007), other researchers were successfully extending the UTAUT model with additional constructs (Alkhwaldi et al., 2023). Additional research is required to determine whether the validated constructs that perform well in simple websites or applications can also be applied to AI coaching applications targeting multi-lingual and multi-cultural audiences.

8.7 Proposed Optimised Framework

Based on the learning derived from constructing the prototype and the collated data, several elements of the COBAC framework ought to be revised. For the sake of completeness, Section 8.7 comprises all components of the Optimised COBAC framework with a description of what was enhanced relative to the initial COBAC framework elements.

8.7.1 Optimised Constraint List

The Constraint List remained unaltered due to the absence of additional input from the prototyping phase.

8.7.2 Optimised Problem List

Nevertheless, the Optimised Problem List necessitates additional items based on the insights from the prototype development and testing phase as summarised in Table 8.8.

Table 8.8: Optimised Problem List - Changes

Key	Problem Summary	Primary Stakeholder	Source
P.64	LLM Limitation: response time of contemporary LLMs is inadequate	Vendor	Prototyping
P.65	There is no established best practice related to testability of information infused coaching agents	Vendor	Prototyping
P.66	LLM Limitation: contemporary LLM providers are offering substandard service availability, which is inacceptable for a conversational agent	Vendor	Prototyping
P.67	LLM Limitation: the response structure of contemporary LLMs can be inconsistent	Vendor	Prototyping
P.68	LLM Limitation: the response language of contemporary LLMs can mismatch with the language specified in the prompt	Vendor	Prototyping
P.69	LLM Limitation: there are significant API structure differences between LLM Provider's API	Vendor	Prototyping
P.70	Coaching conversations are sensitive to cultural aspects	Vendor	Prototyping

8.7.3 Optimised Objective List

The additional Problem List Items are also assigned to the Optimised Objective List, and the objective summary of item O.7 was expanded to include the cross-cultural dimension in Table 8.9.

Table 8.9: Optimised Objective List - Changes

Key	Problem Reference	Measurement	Objective Summary
O.3	P.6, P.64, P.66	monitoring	high availability
O.6	P.26, P.64, P.65, P.66, P.67, P.68	monitoring	service consistency continuous improvement
O.7	P.47, P.70	feature availability, USR	high-resource language enablement with consideration of cross-cultural differences

8.7.4 Optimised Solution Requirement List

The alterations to the Initial Solution Requirement List based on the adjusted objectives and problems are also encapsulated in Table 8.10.

Table 8.10: Optimised Solution Requirement List - Changes

Key	Solution Requirement Summary
SR.30	Message response time of the system in a coaching conversation is higher than the response time of the user within the boundaries of technical feasibility
SR.33	Effective interoperability with the LLM system is ensured by establishing a robust service integration and monitoring that addresses at least the following aspects: response time, fault tolerance, service unavailability, API structural disparities, and non-compliance with prompting
SR.39	A robust test framework is essential for resource-efficient testing of the application
SR.40	The system should adjust its behaviour depending on the conversational language used and the user's demographics

Lastly, the reference architecture remains intact because none of the problems, objectives, solution requirements, or constraints have had an impact.

8.8 Comparative Analysis

The existing autonomous coaching frameworks discussed in Section 2.3.3 exhibit various strengths and limitations in their approaches to designing effective AI coaching agents. The comparative analysis of these frameworks presented in Section 8.8 reveals insights into the following vital aspects of an AI coaching framework:

1. Traceability: Whether the components of the proposed framework can be traced back to scientific literature, coaching competency frameworks, or problem analysis.
2. Coaching Competency: This characteristic shows whether the framework is based on a modern coaching competency framework.

3. Technology Best Practice: This characteristic shows if the framework considers standardized practices and terminology applied in the field of information technology.
4. Architecture: This characteristic shows if the framework provides a high-level technology architecture.
5. Comprehensiveness: This characteristic describes if the framework is based on a defined methodological approach or a systematic analysis to provide a comprehensive view of the subject.
6. Generalizability: This characteristic describes whether the framework underpins a purpose-built, limited-scope use case or can be applied to various coaching goals.
7. Evidence-based: This characteristic indicates if the framework is utilised to build a functional coaching system and the feedback from this endeavour is incorporated into the model.

Using these aspects, Table 8.11 compares the referenced frameworks in Section 2.3.3, the initial COBAC framework presented in Section 6.2, and the optimised COBAC framework demonstrated in Section 8.7. Even though the objectives of some of these frameworks are not specific to AI coaching, comparing these frameworks according to these aspects provides insight into the current level of maturity of the AI coaching frameworks.

Here is the list of frameworks and studies that are considered as part of the comparative analysis:

1. Designing AI Coach (Terblanche, 2020)
2. Design Features of Embodied Conversational Agents in eHealth (ter Stal et al., 2020)
3. Architecting E-Coaching Systems (Ochoa & Gutierrez, 2018)
4. Counseling dialog system with 5W1H extraction (S. Han et al., 2013)
5. Coaching and training system and method for integrated monitoring, managing, supporting, scheduling and prompting of clients (Obispo et al., 2016)
6. COBAC (initial) – Section 6.2
7. COBAC (optimised) – Section 8.7

Table 8.11: Comparative Analysis of AI Coaching Frameworks

Aspect	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Traceability	No	Yes	No	No	No	Yes	Yes
Coaching Competency	Yes	No	No	No	No	Yes	Yes
Technology Best Practice	No	No	Yes	Yes	Yes	Yes	Yes
Architecture	No	No	Yes	Yes	Yes	Yes	Yes
Comprehensiveness	No	Yes	No	No	No	Yes	Yes
Generalizability	Yes	No	No	No	Yes	Yes	Yes
Traceability	No	Yes	No	No	No	Yes	Yes
Evidence-based	Yes	No	No	Yes	No	No	Yes

As Table 8.11 shows, there are many good approaches that embody many vital characteristics of a high-quality AI coaching framework. However, most of them have major shortcomings too. In this context, the optimised COBAC framework is proposed to be the most mature framework considering the analysed characteristics.

8.9 Chapter Summary

Section 8.1 discusses the dual focus of Chapter 8 on data analysis as part of Prototype Evaluation in Step 5 of the Design Science Research Methodology (DSRM), as well as the enhancement of the framework based on insights from the prototyping phase. The data analysis is organised around four hypotheses, and the framework improvement follows its individual components, in accordance with the definition of the fifth DSRM stage.

Section 8.2 shows that 123 people finished the survey and Table 8.1 breaks down their numbers by language and country. Data validation during the data collection process ensured no invalid entries, eliminating the need for data cleaning for missing values or impermissible data, while response validation involved discarding responses with nine consecutive identical values. Reversal of values was required for reverse-scored items, using an algorithm to adjust the participant response values.

Hypothesis 1 (H1) suggested the feasibility of creating an autonomous coaching system with the COBAC framework, confirmed by developing a modifiable chatbot prototype despite

methodological limitations in testing. Technological challenges outlined in Section 7.4.6 are recommended for refining future COBAC framework iterations.

Hypothesis 2 (H2) is supported by the data showing participants generally agreed that the COBAC-based autonomous coaching system was effective in achieving coaching outcomes, with an overall Coaching Effectiveness average close to "Agree”. The data, however, indicated a significant variance in the responses of the German participant group, suggesting the need for further investigation into this discrepancy.

Hypothesis 3 (H3) proposed that the dialogue quality of a COBAC-based chatbot is language-independent and was assessed using the USR model's five criteria. However, the data showed significant variations in dialogue quality across different languages, with English scoring the highest and German the lowest, leading to the rejection of H3.

Hypothesis 4 (H4) posits that language does not significantly impact user adoption intentions for a COBAC-based chatbot, employing Partial Least Squares Structural Equation Modeling (PLS-SEM) with SmartPLS 4 software for hypothesis validation, citing its importance in business and social sciences for handling complex data. Section 8.6.1 extends the UTAUT model by adding Language as a construct to examine its influence on user adoption factors, removing indicators with low factor loadings to refine the model. Finally, Section 8.6.2 asserts the model's predictive power with an R2 value of 0.833 for Behavioral Intention, using Bootstrap resampling for path coefficient significance testing, indicating model acceptance based on reliability and validity metrics. The conclusion regarding Hypothesis 4 suggests that language significantly influence the factors contributing to the intention to use the COBAC-based chatbot, hence H4 is rejected.

Section 8.7 discusses the enhancements made to the Optimised COBAC framework, derived from prototype development learnings, including the need to revise elements to address LLM limitations and cultural sensitivities. The Optimised Problem List was expanded to include issues such as response time inadequacy, lack of best practices for testability, substandard service availability, inconsistent response structures, language mismatches, API differences, and cultural sensitivities in coaching conversations. Objectives and solution requirements

were adjusted to incorporate these issues, focusing on high availability, service consistency, cross-cultural dimensions, and robust interoperability with LLM systems.

Finally, the comparative analysis in Section 8.8 evaluates various AI coaching frameworks, highlighting their strengths and limitations across several key aspects, including traceability, coaching competency, technology best practices, architecture, comprehensiveness, generalizability, and evidence-based approach. Table 8.11 compares these frameworks, revealing that while many possess essential qualities of high-quality AI coaching frameworks, they also have significant shortcomings. The optimised COBAC framework, however, is identified as the most mature and comprehensive among the analysed frameworks, encompassing a wide range of essential characteristics.

9. Conclusions and Further Work

In Chapter 9, the findings of the study are delineated, commencing with an overview that synthesises and reassesses the diverse topics underscored in the debates. This is followed by a recapitulation of the responses to the research questions. A discussion on the theoretical and practical implications of the thesis ensues, which also considers the limitations of the study and potential areas for further investigation.

9.1 Main Achievements

The literature review presented in Chapter 2 uncovers the current landscape and the challenges that permeate the field of coaching automation. Notably, the coaching industry is insufficiently explored within academic spheres, contributing to its perceived diminished credibility and lack of a robust scientific foundation by clientele. A primary concern is the lack of a universally acknowledged definition for coaching. This absence fuels issues concerning professional boundaries with adjacent practices and complicates the evolution of a unified ethical code. Furthermore, there is a conspicuous scarcity of evidence-based research centring on the fundamental competencies of coaching, which erodes its professional integrity.

Current trends, such as team coaching, overlook these elemental issues, and fresh ethical dilemmas, inclusive of the risk of burnout and potential adverse impacts of coaching, were not sufficiently examined. The research methodology within this domain is compromised by structural limitations, encompassing the dearth of transferability studies, scarcity of outcome-based evidence, absence of standardized measurements, and a deficiency in stringent methodologies. These factors contribute to an escalating disjunction between practical application and research, endangering the pertinence of an evidence-based approach within coaching.

The assimilation of automation into coaching introduces further complexities. These include the inconsistent application of terminology, a lack of transparency in automated solutions,

the absence of a standardised reference architecture, and a shortfall in information on automation-related ethical issues, such as bias and confidentiality.

Moreover, the emergence of Large Language Models poses distinct challenges. The academic sector struggles to parallel commercially sponsored research due to resource constraints. Contemporary LLMs necessitate divergent architectural strategies in contrast to conventional rule-based systems and confront issues like the generation of inaccurate information, inherent biases, and extensive ethical concerns. These challenges underscore the urgent necessity for dedicated resources and innovative methodologies to propel the domain of coaching automation forward.

The contributions of this thesis can be interpreted as a considerable progression within this milieu, as this foundational study on competency-based, LLM-driven autonomous coaching has yielded the following outcomes:

1. Devised the inaugural research-oriented framework (COBAC) for the conceptualisation and development of the next generation of autonomous coaching systems.
2. Introduced a high-level reference architecture predicated on the cutting-edge Natural Language Processing technology, namely Large Language Models.
3. Validated four hypotheses and derived new insights through the utilisation of the COBAC framework-based prototype.

9.2 Conclusion Regarding the Hypotheses

This thesis has endeavoured to evaluate the COBAC framework's feasibility for developing autonomous coaching agents. The results, albeit limited by methodological constraints, shed light on the framework's potential and areas needing refinement. The examination was constrained by testing the chatbot in a singular conversational instance per participant and restricting each interaction to a mere 10 minutes. These constraints inevitably limit the broad applicability of the findings. Nevertheless, the results garnered provide valuable insights into the current state and future trajectory of the COBAC framework's development.

Hypothesis 1 posited that the COBAC framework is a viable approach for creating autonomous coaching agents. Despite the methodological limitations, the data corroborates H1, signifying a positive step towards the realization of effective AI coaching prototypes.

Hypothesis 2 (H2) concerns the AI coaching prototype's perceived effectiveness. The consensus among participants, who experienced only 10 minutes of interaction, supports the validity of H2. This suggests that even limited engagement with the AI prototype yields a favourable impression of its coaching abilities, reinforcing the potential of the COBAC framework.

The investigation into linguistic factors revealed intriguing disparities. The German participant group consistently rated the prototype lower on all indicators except understandability, suggesting a linguistic or cultural influence on the perception of AI coaching effectiveness. Furthermore, the Spanish group's ratings also diverged from those of the English language group, except for understandability. These findings lead to the rejection of Hypothesis 3 (H3), indicating that dialogue quality is indeed affected by the language of interaction, contrary to what H3 suggested.

Hypothesis 4 and its sub-components (H4.a through H4.e) delve into specific factors across different language groups. At the 5% significance threshold, not one of the factors demonstrated uniformity throughout all the groups. Notably, while Social Influence was significant for English and Spanish speakers, it did not hold true for German speakers. Similarly, Perceived Risk showed a low path coefficient in the English group but was significant in others. The rejection of H4 and its sub-items suggests that cultural dimensions, rather than linguistic factors alone, may exert a more profound influence on the adoption of AI coaching agents.

In conclusion, while the COBAC framework shows promise, the nuanced differences across cultural and linguistic lines underscore the necessity for a more culturally sensitive approach in the development of AI coaching agents. The results implore a deeper investigation into how cultural contexts shape the interaction with and perception of AI technologies.

9.3 Contribution to the New Knowledge Generation

The research advances theoretical applications in the field of Human Resource Development by addressing several major research gaps in coaching and its automation, particularly using Large Language Models. As outlined in Section 1.2, the main aim of the research was to propose a novel theoretical framework - Competency-Based Autonomous Coaching (COBAC) - for the development of LLM-based autonomous coaching systems.

The high-level requirements of the framework defined in the Section 1.2 were addressed as follows:

1. Coaching Core Competence-Based: The thesis aligns with contemporary coaching core competencies by introducing the Competency-Based Autonomous Coaching (COBAC) framework, ensuring that the framework integrates the foundational principles of professional coaching and addresses well-researched problems in the coaching industry.
2. Scientific Research-Based: By utilising the latest advancements in Large Language Models (LLMs) and addressing research gaps in coaching and its automation, the thesis provides a scientifically validated framework for autonomous coaching systems, promoting evidence-based practices in the coaching industry.
3. Ethics-Oriented Design: At various stages ethical concerns were incorporated in the research design, including but not limited to the Micro and Macro Context Analysis, and Objective Analysis.
4. Practical Application: The practical contributions of this thesis are substantial because it tested a prototype within real-world settings, demonstrating the framework's applicability and effectiveness. To support the practical applicability, it ensures traceability and easy adaptability of the framework components and explores the relatively uncharted application of LLM technology in the autonomous coaching domain, establishing the groundwork for future technological advancements in coaching.

In addition to the initially defined objective in Section 1.2 to deliver a list of addressed problems, objectives, constraints, and solution requirements further deliverables were necessary to achieve the objective, such as the Stakeholder List or the Value Proposition Map.

Lastly, the results of validating the hypothesis not only advanced the research field but also clarified the limitations and outlined directions for further work.

9.4 Limitations and Further Work

Even though this thesis has added to the creation of new knowledge, a significant number of limitations and areas remain that necessitate additional research.

9.4.1 Context and Value Proposition Change

Context and contextual changes remain the most influential factors for managing limitations and determining the direction of future research.

The macro-context is rapidly changing regarding all analysed aspects in Section 4.3: social, technological, economic, environmental, political, legal, and ethical. The most significant emerging macro-contextual changes appear to be related to the breakthrough in AI regulation by the AI Act (European Parliament and of the Council, 2021) and the growing concern over the energy consumption of the AI technologies (Desislavov, Martínez-Plumed, & Hernández-Orallo, 2023), which includes training, fine-tuning, testing, and operating these systems.

The rate of change, albeit significantly slower compared with the macro-context, is also accelerating. Key trends are associated with changing coaching goals, for instance, the increase of burnout indicates the growing demand to improve work-life balance and skill development through coaching (ICF, 2022a).

Lastly, the focus of the thesis was on individual goal achievement because it is considered the primary value targeted by individual coaching clients. However, in the currently rapidly changing world where society fails to adapt its behaviours quickly enough to combat the

impact of looming global threats, such as climate change, pollution, or energy shortages, the value of accelerated social change might become a more prominent value generated by autonomous coaching systems in the future.

Each of these contextual and value proposition-related changes can be incorporated into the proposed structure of the COBAC framework using the same approach as demonstrated with the building of the prototype. However, more research is required to precisely understand the impact of these changes on the framework's elements.

9.4.2 Scope Limitation

To start with, the scope of the study was explicitly limited by constraints defined in the Constraints List. One direction for further research could involve exploring the impact of these constraints, for instance, by not limiting the discussion to adults, healthy clients, or by examining the effects of providing group or team coaching modalities. The prototype was also restricted to the coaching goal of developing time management skills, and it remains to be seen how different coaching goals affect the efficacy of the coaching competencies.

After that, a fundamental principle of the COBAC framework is its basis on a fixed set of coaching competencies. A natural progression for coaching automation research is to investigate the impact of individual competencies or combinations of competencies on coaching outcome measures. Such research could provide an evidence-based foundation for a new competency framework, potentially revolutionizing the coaching industry for both humans and autonomous agents. This study would necessitate thousands of coaching conversations, employing a machine learning approach to identify patterns in the large dataset. Chittum proposed a competency framework that included generic core coaching competencies and additional core competencies (Chittum, 2013), differentiated by a single parameter: the specific type of coaching. The findings from a large-scale autonomous coaching study that validates coaching competencies could result in a similar differentiated structure, but utilising hundreds of parameters instead of one. This could make the coaching competency framework highly effective for machines, yet potentially complex for human coaches.

The user interface in the prototype stage was limited to a textual chatbot interface; however, exploring other human-computer interfaces could enhance the accessibility and interaction of autonomous coaching systems. This exploration would require in-depth research into processing data related to meta-communication and developing the capability to exhibit a robust spectrum of meta-communication elements.

Finally, while the study was confined to coaching, many of the applied competencies are relevant in other fields. Therefore, this approach should not be restricted to coaching conversations alone. A new frontier for research could be to apply the competency-based approach in boundary modalities, such as counselling, mentoring, or even psychotherapy.

9.4.3 Methodology Limitation

However, the most significant limitation of the study's scope and methodology was related to the researcher's resources: the external cost of conducting the study. These constraints substantially affected participant characteristics, including the number of participants, diversity, cultural differences, and geographical distribution. Consequently, these factors influenced the research design, resulting in a cross-sectional, non-controlled participant-based study. This design was suitable for testing the four hypotheses to gain preliminary insights into the COBAC framework. However, it may be adjusted in future research. Additionally, the measurement approach was constrained to a 43-question quantitative survey due to the same resource limitations.

9.4.4 Technology

Technological development not only impacts the design and creation of new autonomous coaching systems, but it also fosters the emergence of novel concepts and methods for testing LLM-driven applications. There are several promising approaches that could be relevant to the coaching context. For instance, the use of synthetic data generated by LLMs (Lorge et al., 2024) or the agent-based model (Durante et al., 2024), in which LLMs power the autonomous coaching system and could be employed to simulate clients and evaluate conversation

quality. This closed-loop feedback cycle can lead to significant and rapid improvements in dialogue quality and reduce the need for recruiting human participants. However, more coaching-specific research is necessary in the quality assurance domain of autonomous coaching systems, as many of the latest methods developed for data-driven, LLM-based conversational agents may not be suitable for the coaching context. This is because an autonomous coaching system does not provide information one-on-one in the conversation, unlike the advisory or service chatbots that have become widespread in recent years.

Another potential research avenue is the integration of traditional machine learning algorithms with LLMs to enhance specific task performance, such as sentiment analysis or pattern recognition, to improve coaching competencies. The current study relied solely on LLM Providers for all communication related tasks, yet there is scope for further technological advancements by incorporating established, traditional AI components into the coaching process.

The evolution of LLMs continues, and while the majority of modern LLMs are built on the transformer architecture (Vaswani et al., 2017), new techniques could lead to significant performance improvements. An example is the application of the Mixture of Experts concept demonstrated in the Mixtral LLM (Jiang et al., 2024). Further enhancements could be realised by exploring the effects of fine-tuning on model performance, with newer models likely to improve both response quality and response time—two crucial issues in the autonomous coaching field.

Finally, the widespread adoption of LLMs may fundamentally change how we interact with machines. Current research in this field includes the Rabbit R1 pocket companion (Rabbit Inc., 2024), which has the potential to transform the mobile application ecosystem by enabling user interaction via an LLM rather than traditional applications.

References

Aguilar, F. J. (1967). *Scanning the business environment*. *Studies of the modern corporation*. New York: Macmillan.

Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk Research: Review and Recommendations. *Journal of Management*, *47*(4), 823–837. https://doi.org/10.1177/0149206320969787

Akter, S., McCarthy, G., Sajib, S., Michael, K., Dwivedi, Y. K., D'Ambra, J., & Shen, K. N. (2021). Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management*, *60*(102387).

Alkaissi, H., & McFarlane, S. I. (2023). Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus*, *15*(2), 2–5. https://doi.org/10.7759/cureus.35179

Alkhwaldi, A. F., Al-Qudah, A. A., Al-Hattami, H. M., Al-Okaily, M., Al-Adwan, A. S., & Abu-Salih, B. (2023). Uncertainty avoidance and acceptance of the digital payment systems: a partial least squares-structural equation modeling (PLS-SEM) approach. *Global Knowledge, Memory and Communication*, *ahead-of-p*(ahead-of-print). https://doi.org/10.1108/GKMC-07-2022-0161

Amazon. (2024). Data encryption - Amazon Bedrock. Retrieved February 5, 2024, from https://docs.aws.amazon.com/bedrock/latest/userguide/data-encryption.html

Anderson, J., & Kamphorst, B. (2014). Ethics of E-Coaching : Implications of Employing Pervasive Computing to Promote Healthy and Sustainable Lifestyles. In *IEEE international conference on pervasive computing and communication workshops (PERCOM WORKSHOPS)* (pp. 351–356). IEEE.

Anthony, A. (2018). Yuval Noah Harari: 'The idea of free information is extremely dangerous.' Retrieved May 1, 2024, from https://www.theguardian.com/culture/2018/aug/05/yuval-noah-harari-free-information-extremely-dangerous-interview-21-lessons

Arenas-Gaitán, J., Peral-Peral, B., & Ramón-Jerónimo, M. A. (2015). Elderly and internet banking: An application of UTAUT2. *Journal of Internet Banking and Commerce*, *20*(1), 1–23.

Askell, A., Jones, A., Chen, A., Drain, D., Dec, C. L., Joseph, N., … Olsson, C. (2021). A General Language Assistant as a Laboratory for Alignment. Retrieved September 23, 2023, from https://arxiv.org/pdf/2112.00861.pdf

Association for Coaching. (2012). AC Coaching Competency Framework. Retrieved February 27, 2024, from https://cdn.ymaws.com/www.associationforcoaching.com/resource/resmgr/Accreditation/Accred_General/Coaching_Competency_Framewor.pdf

Association for Coaching. (2020). Coaching Defined - Association for Coaching. Retrieved March 29, 2020, from https://www.associationforcoaching.com/page/CoachingDefined

Association for Coaching, & European Mentoring & Coaching Council. (2019). The Global Code of Ethics For Coaches, Mentors, and Supervisors. Retrieved May 4, 2020, from https://app.box.com/s/8s3tsveqieq6vr6n2itb0p9mpsxgcncd

Athanasopoulou, A., & Dopson, S. (2018). A systematic review of executive coaching outcomes: Is it the journey or the destination that matters the most? *The Leadership Quarterly*, *29*(1), 70–88.

AXELOS. (2019). ITIL® Foundation Glossary terms and definitions. Retrieved October 17, 2023, from https://purplegriffon.com/downloads/resources/itil4-foundation-glossary-

january-2019.pdf
Bachkirova, T. (2014). Role of coaching psychology in defining boundaries between counselling and coaching. In *Handbook of coaching psychology* (pp. 371–386). Routledge.
Bachkirova, T., & Cox, E. (2004). A bridge over troubled water bringing together coaching and counselling. *International Journal of Mentoring and Coaching*, *2*(1), 1–10.
Bachkirova, T., & Kauffman, C. (2009). The blind men and the elephant: using criteria of universality and uniqueness in evaluating our attempts to define coaching. *Coaching: An International Journal of Theory, Research and Practice*, *2*(2), 95–105.
Bachkirova, T., Spence, G., & Drake, D. (2016). *The SAGE Handbook of Coaching*. (T. Bachkirova, G. Spence, & D. Drake, Eds.). London: SAGE Publications Asia-Pacific Pte Ltd.
Bacon, T. R. (2011). Measuring the Effectiveness of Executive Coaching. Retrieved May 1, 2024, from https://www.kornferry.com/insights/this-week-in-leadership/170-measuring-the-effectiveness-of-executive-coaching
Baokar, Shrikrishna & Wamane, Vikas & Taware, Ganesh & Jadhav, R. (2018). Text Neck Syndrome. In *Innopharm3: 3rd International Conference on Academic and Industrial Innovations* (p. 11797). Tal-Baramati.
Basta, C., Costa-jussà, M. R., & Casas, N. (2019). Evaluating the Underlying Gender Bias in Contextualized Word Embeddings. Retrieved September 23, 2023, from https://arxiv.org/pdf/1904.08783
Bayeck, R. Y. (2016). Exploratory study of MOOC learners' demographics and motivation: The case of students involved in groups. *Open Praxis*, *8*(3), 223–233.
BBC. (2023). Game of Thrones author sues ChatGPT owner OpenAI. Retrieved September 22, 2023, from https://www.bbc.com/news/technology-66866577
Benaich, N., & Air Street Capital. (2022). State of AI Report 2023. Retrieved October 13, 2023, from https://www.stateof.ai
Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots : Can Language Models Be Too Big ? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency. 2021.* (pp. 610–623). https://doi.org/10.1145/3442188.3445922
Bernardini, A. A., Sônego, A. A., & Pozzebon, E. (2018). Chatbots: An Analysis of the State of Art of Literature. In R. D. Araújo (Ed.), *1st Workshop on Advanced Virtual Environments and Education* (p. 1). Florianópolis, SC, Brazil: Sociedade Brasileira de Computação.
Bersin, J. (2021). Coaching at Scale. Retrieved May 1, 2024, from https://joshbersin.com/research/diversity-leadership-culture/
Bhatnagr, P., & Rajesh, A. (2023). Neobanking adoption – An integrated UTAUT-3, perceived risk and recommendation model. *South Asian Journal of Marketing*, *ahead-of-p*(ahead-of-print). https://doi.org/10.1108/sajm-06-2022-0040
Blodgett, S. L., Barocas, S., Daumé, H., & Wallach, H. (2020). Language (Technology) is Power: A Critical Survey of "Bias" in NLP. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics* (pp. 5454–5476). Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.485
Bornstein, M., & Radovanovic, R. (2023). Emerging Architectures for LLM Applications. Retrieved September 18, 2023, from https://a16z.com/emerging-architectures-for-llm-applications/
Boyatzis, R. E. (1991). *The competent manager: A model for effective performance*. John

Wiley & Sons.
Boyatzis, R. E., Hullinger, A., Ehasz, S. F., Harvey, J., Tassarotti, S., Gallotti, A., & Penafort, F. (2022). The Grand Challenge for Research on the Future of Coaching. *Journal of Applied Behavioral Science*, *58*(2), 202–222. https://doi.org/10.1177/00218863221079937
Braun, J. A., & Eklund, J. L. (2019). Fake News, Real Money: Ad Tech Platforms, Profit-Driven Hoaxes, and the Business of Journalism. *Digital Journalism*, *7*(1), 1–21.
Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.
Britz, D. (2016). Deep Learning for Chatbots, Part 1 – Introduction. Retrieved September 25, 2023, from https://dennybritz.com/posts/wildml/deep-learning-for-chatbots-part-1/
Brock, V. G. (2008). *Grounded theory of the roots and emergence of coaching*. *International University of Professional Studies*.
Brock, V. G. (2012). *Sourcebook of Coaching History* (2nd ed.). CreateSpace Independent Publishing Platform.
Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … Amodei, D. (2020). Language models are few-shot learners. In *34th Conference on Neural Information Processing Systems* (Vol. 33, pp. 1877–1901). Vancouver, Canada.
Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., … Zhang, Y. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. Retrieved September 25, 2023, from http://arxiv.org/abs/2303.12712
Budhwar, P., Chowdhury, S., Wood, G., Aguinis, H., Bamber, G. J., Beltran, J. R., … Varma, A. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Human Resource Management Journal*, *33*(June), 606–659. https://doi.org/10.1111/1748-8583.12524
Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., … Trench, M. (2017). Artificial intelligence: The next digital frontier? Retrieved July 2, 2020, from https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/how-artificial-intelligence-can-deliver-real-value-to-companies
Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980
Bungay Stanier, M. (2016). *The coaching habit : say less, ask more & change the way you lead forever*. Toronto: Box of Crayons Press.
Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475
Cambridge Business English Dictionary. (2020). algorithm. Retrieved June 11, 2020, from https://dictionary.cambridge.org/dictionary/english/algorithm
Casas, J., Mugellini, E., & Khaled, O. A. (2018). Food diary coaching chatbot. In *Adjunct Proceedings of the 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2018 ACM International Symposium on Wearable Computers* (pp. 1676–1680). Singapore: ACM.
Cavanagh, M., & Buckley, A. (2014). Coaching and mental health. In *The complete handbook of coaching* (pp. 405–417). Sage London.
Chittum, R. (2013). *Coaching in Organizations*. *On Becoming a Leadership Coach* (Committee). Sydney: SAI Global Limited. https://doi.org/10.1057/9781137344137_18

Clark, J., & Perrault, R. (2023). Artificial Intelligence Index Report 2023. Retrieved May 1, 2024, from https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf

Clutterbuck, D. (2008). Are you a goal junkie? *Training Journal*, (May), 43–46.

CoachHub. (2020). CoachHub. Retrieved April 8, 2020, from https://coachhub.io/for-coachees/

Coaching.com. (2024). Coaching.com. Retrieved February 28, 2024, from https://www.coaching.com/

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A Global Measure of Perceived Stress. *Journal of Health and Social Behavior*, *24*(4), 385–396.

Comscore. (2017). Global Digital Future in Focus - 2018 International Edition. Retrieved July 2, 2020, from http://adepa.org.ar/wp-content/uploads/2018/03/Global-Digital-Future-in-Focus-2018.pdf

Cox, E., Bachkirova, T., & Clutterbuck, D. (2010). *The complete handbook of coaching* (First Edit). Los Angeles: SAGE.

Dagley, G. R. (2010). Exceptional executive coaches: Practices and attributes. *International Coaching Psychology Review*, *5*(1), 63–80.

Dauth, W., Findeisen, S., Südekum, J., & Wößner, N. (2017). German Robots – The Impact of Industrial Robots on Workers. Retrieved July 1, 2020, from http://doku.iab.de/discussionpapers/2017/dp3017.pdf

David, S., Clutterbuck, D., & Megginson, D. (2014). Goal orientation in coaching differs according to region, experience, and education. *International Journal of Evidence Based Coaching & Mentoring*, *12*(2), 134–145.

De Haan, E., Culpin, V., & Curd, J. (2011). Executive coaching in practice: What determines helpfulness for clients of coaching? *Personnel Review*, *40*(1), 24–44.

De Haan, E., Duckworth, A., Birch, D., & Jones, C. (2013). Executive coaching outcome research: The contribution of common factors such as relationship, personality match, and self-efficacy. *Consulting Psychology Journal*, *65*(1), 40–57.

De Haan, E., Grant, A. M., Burger, Y., & Eriksson, P. O. (2016). A large-scale study of executive and workplace coaching: The relative contributions of relationship, personality match, and self-efficacy. *Consulting Psychology Journal*, *68*(3), 189–207.

De Haan, E., Gray, D. E., & Bonneywell, S. (2019). Executive coaching outcome research in a field setting: A near-randomized controlled trial study in a global healthcare corporation. *Academy of Management Learning and Education*, *18*(4), 581–605. https://doi.org/10.5465/amle.2018.0158

De Haan, E., Molyn, J., & Nilsson, V. O. (2020). New findings on the effectiveness of the coaching relationship: Time to think differently about active ingredients? *Consulting Psychology Journal*, *72*(3), 155–167. https://doi.org/10.1037/cpb0000175

De Haan, E., & Nieß, C. (2015). Differences between critical moments for clients, coaches, and sponsors of coaching. *Coaching Researched: A Coaching Psychology Reader*, *10*(1), 38–57.

Deshpande, A., Murahari, V., Rajpurohit, T., Kalyan, A., & Narasimhan, K. (2023). Toxicity in CHATGPT: Analyzing Persona-assigned Language Models. Retrieved September 23, 2023, from https://arxiv.org/pdf/2304.05335.pdf

Desislavov, R., Martínez-Plumed, F., & Hernández-Orallo, J. (2023). Trends in AI inference energy consumption: Beyond the performance-vs-parameter laws of deep learning. *Sustainable Computing: Informatics and Systems*, *38*(January 2022), 100857.

https://doi.org/10.1016/j.suscom.2023.100857
Dillahunt, T., Wang, Z., & Teasley, S. D. (2014). Democratizing higher education: Exploring MOOC use among those who cannot afford a formal education. *International Review of Research in Open and Distance Learning*, *15*(5), 177–196.
Drake, D. B. (2018). *Narrative coaching: The definitive guide to bringing new stories to life*. CNC Press.
Durante, Z., Sarkar, B., Gong, R., Taori, R., Noda, Y., Tang, P., … Huang, Q. (2024). An Interactive Agent Foundation Model. Retrieved May 1, 2024, from http://arxiv.org/abs/2402.05929
Dwivedi, Y. K., Rana, N. P., Jeyaraj, A., Clement, M., & Williams, M. D. (2019). Re-examining the Unified Theory of Acceptance and Use of Technology (UTAUT): Towards a Revised Theoretical Model. *Information Systems Frontiers*, *21*(3), 719–734. https://doi.org/10.1007/s10796-017-9774-y
Ellinger, A. D., & Kim, S. (2014). Coaching and Human Resource Development: Examining Relevant Theories, Coaching Genres, and Scales to Advance Research and Practice. *Advances in Developing Human Resources*, *16*(2), 127–138.
Ellis-Brush, K. (2021). Augmenting Coaching Practice through Digital Methods. *International Journal of Evidence Based Coaching and Mentoring*, *1*(S15), 187–197. https://doi.org/10.24384/er2p-4857
EMCC. (2015). EMCC competence framework v2. Retrieved March 29, 2020, from https://emcc1.app.box.com/s/4aj8x6tmbt75ndn13sg3dauk8n6wxfxq
EMCC. (2016). EMCC guidelines for supervision. Retrieved May 21, 2020, from http://www.emccouncil.org/src/ultimo/models/Download/7.pdf
EMCC. (2018). Emcc Competence Framework Glossary. Retrieved March 29, 2020, from http://www.emccouncil.org/webimages/EMCC/EMCC_Competence_Framework.pdf
EMCC. (2020). Team Coaching Individual Accreditation (ITCA). Retrieved October 18, 2020, from https://www.emccglobal.org/accreditation/itca/
EMCC. (2022). Directory - EMCC. Retrieved October 17, 2022, from https://www.emccglobal.org/directory/?record_type=1&full_name=&email=
EMCC, ICF, AC, & SFC. (2021). The Professional Charter for Coaching, Mentoring, and Supervision of Coaches, Mentors, and Supervisors. Retrieved October 24, 2023, from https://www.eesc.europa.eu/sites/default/files/files/150_-_private_act_-_professional_charter.pdf
European Parliament and of the Council. (2016). Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation – GDPR). Retrieved May 8, 2020, from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679
European Parliament and of the Council. (2021). The Artificial Intelligence Act. Retrieved October 24, 2023, from https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106(COD)&l=en
Facebook. (2020). Facebook Login for the Web with the JavaScript SDK. Retrieved June 12, 2020, from https://developers.facebook.com/docs/facebook-login/web/
Falk, R. F., & Miller, N. B. (1992). *A primer for soft modeling*. University of Akron Press.
Fiske, S. T. (2017). Prejudices in Cultural Contexts: Shared Stereotypes (Gender, Age) versus Variable Stereotypes (Race, Ethnicity, Religion). *Perspect Psychol Sci*, *12*(5), 791–799.

https://doi.org/10.1177/0022146515594631.Marriage
Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, *4*(2), e7785. https://doi.org/10.2196/mental.7785
Fornell, C., & Larcker, D. F. (2016). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50.
Fors, K. L. (2015). *Production and Perception of Pauses in Speech*. University of Gothenburg.
Fouracres, A. J. S. (2022). *Cybersecurity for Coaches and Therapists: A Practical Guide for Protecting Client Data* (1st ed.). Routledge.
Franklin, J. (2005). Change Readiness in Coaching: Potentiating Client Change. In *Evidence-based coaching: Volume 1* (pp. 193–200). Brisbane: Australian Academic Press.
Frühauf, K., Fuchs, E., Glinz, M., Grau, R., Hood, C., Houdek, F., … Beatty, J. (2017). IREB Certified Professional for Requirements Engineering-Foundation Level. Retrieved May 25, 2020, from https://www.ireb.org/content/downloads/2-syllabus-foundation-level/ireb_cpre_syllabus_fl_en_v222.pdf
Furnham, A. (2008). Relationship among four Big Five measures of different length. *Psychological Reports*, *102*(1), 312–316.
Future of Life Institute. (2017). AI Principles. Retrieved September 28, 2023, from https://futureoflife.org/2017/08/11/ai-principles/?cn-reloaded=1&cn-reloaded=1
Gallwey, T. (2008). Coaching - The Inner Game. Retrieved March 29, 2020, from https://theinnergame.com/coaching/
Garas, A., Garcia, D., Skowron, M., & Schweitzer, F. (2012). Emotional persistence in online chatting communities. *Scientific Reports*, *2*(1), 1–8. https://doi.org/10.1038/srep00402
Gehman, S., Gururangan, S., Sap, M., Choi, Y., & Smith, N. A. (2020). REALTOXICITYPROMPTS: Evaluating neural toxic degeneration in language models. https://doi.org/10.18653/v1/2020.findings-emnlp.301
Geissler, H, Hasenbein, M., Kanatouri, S., & R., W. (2014). E-Coaching: Conceptual and empirical findings of a virtual coaching programme. *International Journal of Evidence Based Coaching and Mentoring*, *12*(2), 165–186.
Geissler, Harald. (2009). Die inhaltsanalytische „Vermessung“ von Coachingprozessen. In *Coachingwissen* (pp. 95–127). Birgmeier.
Glinz, M. (2014). A Glossary of Requirements Engineering Terminology, Version 1.6 May 2014, (October).
Goldstein, J. A., Sastry, G., Musser, M., Participants, W., Adler, S., Avin, S., … Goldstein, J. (2023). Generative Language Models and Automated Influence Operations : Emerging Threats and Potential Mitigations. Retrieved September 20, 2023, from https://arxiv.org/pdf/2301.04246.pdf
Google. (2020). Android 10 for Developers. Retrieved June 12, 2020, from https://developer.android.com/about/versions/10/highlights#privacy_for_users
Google. (2024a). Gemini Apps Privacy Hub. Retrieved February 27, 2024, from https://support.google.com/gemini/answer/13594961?hl=en
Google. (2024b). What Gemini Apps can do and other frequently asked questions. Retrieved February 27, 2024, from https://gemini.google.com/faq
Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528.
Grant, A. M. (2006a). A personal perspective on professional coaching and the development

of coaching psychology. *International Coaching Psychology Review*, *1*(1), 12–22.
Grant, A. M. (2006b). An Integrative Goal-Focused Approach to Executive Coaching. In *Evidence based coaching handbook: Putting best practices to work for your clients* (pp. 153–192). John Wiley & Sons Inc.
Grant, A. M. (2011). Workplace, Executive and Life Coaching: An Annotated Bibliography from the Behavioural Science and Business Literature. Retrieved May 1, 2024, from http://libraryofprofessionalcoaching.com/wp-app/wp-content/uploads/2011/11/Coaching_Bio_-1st-JAN_2011_GRANT.pdfT.pdf
Grant, A. M. (2012). The Efficacy of Coaching. *The Wiley-Blackwell Handbook of the Psychology of Coaching and Mentoring*, (October 2012), 13–39. https://doi.org/10.1002/9781118326459.ch2
Grant, A. M. (2017). Coaching as evidence-based practice: The view through a multiple-perspective model of coaching research. In *The SAGE handbook of coaching* (pp. 62–84). Sage Publications, Inc.
Graßmann, C., & Schermuly, C. C. (2021). Coaching With Artificial Intelligence: Concepts and Capabilities. *Human Resource Development Review*, *20*(1), 106–126. https://doi.org/10.1177/1534484320982891
Green, S., Oades, L. G., & Grant, A. M. (2005). An Evaluation of a Life-Coaching Group Program: Initial Findings From a Waitlist Control Study. In *Evidence-based coaching: Volume 1* (Vol. 1, pp. 127–142). Bowen Hills, Qld, Australia: Australian Academic Press.
Gregory, J. B., Beck, J. W., & Carr, A. E. (2011). Goals, feedback, and self-regulation: Control theory as a natural framework for executive coaching. *Consulting Psychology Journal*, *63*(1), 26–38.
Griffiths, K. E., & Campbell, M. A. (2008). Regulating the regulators: Paving the way for international, evidence-based coaching standards. *International Journal of Evidence Based Coaching and Mentoring*, *6*(1), 19–31.
Grover, S., & Furnham, A. (2016). Coaching as a developmental intervention in organisations: A systematic review of its effectiveness and the mechanisms underlying It. *PLoS ONE*, *11*(7), 1–41. https://doi.org/10.1371/journal.pone.0159137
Habitica. (2020). Habitica. Retrieved April 9, 2020, from https://habitica.com/
Hair, Joe F., Sarstedt, M., Hopkins, L., & Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. *European Business Review*, *26*(2), 106–121. https://doi.org/10.1108/EBR-10-2013-0128
Hair, Joseph F, Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate data analysis* (5th ed., Vol. 5). Upper Saddle River.
Han, B.-C. (2015). *The burnout society*. Stanford University Press.
Han, S., Lee, K., Lee, D., & Lee, G. G. (2013). Counseling dialog system with 5W1H extraction. In *SIGDIAL 2013 - 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Proceedings of the Conference* (pp. 349–353).
Harper, A. (2008). Psychometric tests are now a multi-million-pound business: what lies behind a coach's decision to use them? *International Journal of Evidence Based Coaching and Mentoring*, *November*(Special Issue No. 2), 40–51. Retrieved from https://researchportal.coachfederation.org/MediaStream/PartialView?documentId=2088
Hawksworth, J., Berriman, R., & Goel, S. (2018). Will robots really steal our jobs ? An international analysis of the potential long term impact of automation. Retrieved July 2, 2020, from https://www.pwc.co.uk/economic-services/assets/international-impact-

of-automation-feb-2018.pdf
Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020). Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. *Journal of Machine Learning Research*, *21*(1), 1–43.
Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83.
Horne, K. A. (2008). An Analysis of the Coaching Competencies of Managers in the Eastern Cape. Retrieved May 1, 2024, from https://uir.unisa.ac.za/bitstream/handle/10500/2937/dissertation_horne_a.pdf?sequence=1&isAllowed=y
ICF. (2009). ICF Global Coaching Client Study. Retrieved May 3, 2020, from http://www.gos-coaching.ch/pdf/2009-ICF Global survey.pdf
ICF. (2017a). 2017 ICF global consumer awareness study: Executive summary. Retrieved May 3, 2020, from https://carlyanderson.com/wp-content/uploads/2017-ICF-Global-Consumer-Awareness-Study.pdf
ICF. (2017b). ICF core competencies rating levels. Retrieved July 2, 2020, from https://coachfederation.org/app/uploads/2017/12/ICFCompetenciesLevelsTable.pdf
ICF. (2019a). Updated ICF Core Competency Model. Retrieved February 10, 2020, from https://coachfederation.org/app/uploads/2019/11/ICFCompetencyModel_Oct2019.pdf
ICF. (2019b, September). ICF Code of Ethics. Retrieved March 30, 2020, from https://coachfederation.org/app/uploads/2020/01/ICF-Code-of-Ethics_final_Nov12.pdf
ICF. (2020a). Coaching Supervision. Retrieved February 27, 2024, from https://coachfederation.org/app/uploads/2020/02/CoachingSupervision_onesheet_FEB2.pdf
ICF. (2020b). ICF Core Competency Updates : Webinar Conversations with Coach Training Providers. Retrieved April 15, 2020, from https://coachfederation.org/app/uploads/2020/01/CC-Update-Webinar-FAQs-for-Coach-Trainers-Jan.-2020.pdf
ICF. (2021). ICF Team Coaching Competencies: Moving Beyond One-to-One Coaching. Retrieved October 18, 2022, from https://coachingfederation.org/app/uploads/2021/10/Team-Coaching-Competencies_10.4.21.pdf
ICF. (2022a). 2022 ICF Global Consumer Awareness Study. Retrieved February 27, 2024, from https://coachingfederation.org/research/consumer-awareness-study
ICF. (2022b). Professional Coaches Membership and Credentialing Fact Sheet. Retrieved October 17, 2022, from https://coachingfederation.org/app/uploads/2022/10/October2022_FactSheet.pdf
ICF. (2022c). The Artificial Intelligence (AI) Coaching Capability Assessment Guidelines. Retrieved May 1, 2024, from https://coachingfederation.org/
ICF. (2023). 2023 ICF Coaching Snapshot: Future of Coaching. Retrieved September 3, 2023, from https://coachingfederation.org/research/global-coaching-snapshots
Independent ICF Review Board. (2019). 2018 Ethical Conduct and Compliance Report 2018 Complaint Summary Fact Sheet. Retrieved May 4, 2020, from https://coachfederation.org/app/uploads/2019/07/2018EthicalConductandComplianceReport.pdf

Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR MHealth and UHealth*, *6*(11), 1–14.

IREB. (2020). A glossary of requirements engineering terminology. Retrieved April 11, 2020, from https://www.ireb.org/content/downloads/1-cpre-glossary/ireb_cpre_glossary_17.pdf

ISO/IEC. (2011). Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models. Retrieved January 15, 2020, from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=35733

Jiang, A. Q., Sablayrolles, A., Roux, A., Mensch, A., Savary, B., Bamford, C., … Sayed, W. El. (2024). Mixtral of Experts. Retrieved May 1, 2024, from http://arxiv.org/abs/2401.04088

Jones, R. J., Woods, S. A., & Guillaume, Y. R. F. (2016). The effectiveness of workplace coaching: A meta-analysis of learning and performance outcomes from coaching. *Journal of Occupational and Organizational Psychology*, *89*(2), 249–277. https://doi.org/10.1111/joop.12119

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics* (pp. 6282–6293). India. https://doi.org/10.18653/v1/2020.acl-main.560

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2003). The core self-evaluations scale: Development of a measure. *Personnel Psychology*, *56*(2), 303–331.

Kamphorst, B. A. (2017). E-coaching systems: What they are, and what they aren't. *Personal and Ubiquitous Computing*, *21*(4), 625–632. https://doi.org/10.1007/s00779-017-1020-6

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*(2023), 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kiresuk, T. J., & Sherman, R. E. (1968). Goal attainment scaling: A general method for evaluating comprehensive community mental health programs. *Community Mental Health Journal*, *4*(6), 443–453.

Klemp Jr., G. O. (1980). *The Assessment of Occupational Competence. Final Report: I. Introduction and Overview. McBer and Co., Boston Mass.* Washington, DC.

Koch, B. (2012). ‚onlineCoaching': ein geschriebener Dialog unabhängig von Zeit und Raum. In *E-Coaching und Online-Beratung : Formate, Konzepte, Diskussionen* (Springer V, pp. 87–100). Wiesbaden : VS Verlag für Sozialwissenschaften.

Koopman, R., Englis, P. D., Ehgrenhard, M. L., & Groen, A. (2021). The Chronological Development of Coaching and Mentoring: Side by Side Disciplines. *International Journal of Evidence Based Coaching and Mentoring*, *19*(1), 137–151. https://doi.org/10.24384/3w69-k922

Koubaa, A., Boulila, W., Ghouti, L., Alzahem, A., & Latif, S. (2023). Exploring ChatGPT Capabilities and Limitations : A Critical Review of the NLP Game Changer. https://doi.org/10.20944/preprints202303.0438.v1

Krapu, T. (2016). Coaching from a philosophy of science perspective. *Philosophy of*

Coaching: An International Journal, *1*(1), 8–20.
Kurita, K., Vyas, N., Pareek, A., Black, A. W., & Tsvetkov, Y. (2019). Measuring Bias in Contextualized Word Representations. Retrieved September 23, 2023, from https://arxiv.org/pdf/1906.07337
Lee, C. Y., & Chen, B. S. (2018). Mutually-exclusive-and-collectively-exhaustive feature selection scheme. *Applied Soft Computing Journal*, *68*(2018), 961–971.
Lee, P., & Calugar-Pop, C. (2019). Plateauing at the peak The state of the smartphone. Retrieved February 27, 2024, from https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/technology-media-telecommunications/deloitte-uk-plateauing-at-the-peak-the-state-of-the-smartphone.pdf
Lembcke, T.-B., Engelbrecht, N., Brendel, A. B., & Kolbe, L. (2019). To nudge or not to nudge: Ethical considerations of digital nudging based on its behavioral economics roots. In *Proceedings of the 27th European Conference on Information Systems (ECIS)* (pp. 1–17).
Lerner, D., Adler, D. A., Rogers, W. H., Chang, H., Greenhill, A., Cymerman, E., & Azocar, F. (2015). A randomized clinical trial of a telephone depression intervention to reduce employee presenteeism and absenteeism. *Psychiatric Services*, *66*(6), 570–577.
Li, M., Weston, J., & Roller, S. (2019). ACUTE-EVAL: Improved Dialogue Evaluation with Optimized Questions and Multi-turn Comparisons. Retrieved May 8, 2020, from https://arxiv.org/abs/1909.03087
Libri, V., & Kemp, T. (2006). Assessing the efficacy of a cognitive behavioural executive coaching programme. *International Coaching Psychology Review*, *1*(23), 9–20.
Lindvall, P., Holm, S., Ross, A., van Vlerken, A., & Margariti, M. (2013). ICF & EMCC Competence Frameworks: A comparative approach. Retrieved June 3, 2020, from http://www.undici.fr/articles/ICF EMCC CFW Competence comparison - Final report.pdf
Linley, P. A. (2006). Coaching research: Who? What? Where? When? Why? *International Journal of Evidence Based Coaching and Mentoring*, *4*(2), 7.
Locke, E. A. (1996). Motivation through conscious goal setting. *Applied and Preventive Psychology*, *5*(2), 117–124.
Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*(9), 705–717.
Loenhoud, H. VAN, Lauenroth, K., & Steiger, P. (2017). The goal is to solve the problem. *Requirements Engineering Magazine*, (3), 1–9.
Lorenz-Spreen, P., Mønsted, B. M., Hövel, P., & Lehmann, S. (2019). Accelerating dynamics of collective attention. *Nature Communications*, *10*(1), 1759.
Lorge, I., Joyce, D. W., Taylor, N., Nevado-Holgado, A., Cipriani, A., & Kormilitzin, A. (2024). Detecting the Clinical Features of Difficult-to-Treat Depression using Synthetic Data from Large Language Models. Retrieved May 1, 2024, from http://arxiv.org/abs/2402.07645
Lund, S., Manyika, J., Segel, L. H., Dua, A., Hancock, B., Rutherford, S., & Macon, B. (2019). Future of work in America. Retrieved July 3, 2020, from https://www.mckinsey.com/featured-insights/future-of-work/the-future-of-work-in-america-people-and-places-today-and-tomorrow
MacKie, D. (2015). The effects of coachee readiness and core self-evaluations on leadership coaching outcomes: A controlled trial. *Coaching: An International Journal of Theory, Research and Practice*, *8*(2), 120–136. Retrieved from

https://www.tandfonline.com/doi/abs/10.1080/17521882.2015.1019532
Malhotra, R., & Chug, A. (2016). Software Maintainability: Systematic Literature Review and Current Trends. *International Journal of Software Engineering and Knowledge Engineering*, *26*(8), 1221–1253.
March, S. T., & Smith, G. F. (1995). Design and natural scienceresearch on information technology. *Decision Support System*, *15*(4), 251–266.
Marcu, G., Ondersma, S. J., Spiller, A. N., Broderick, B. M., Kadri, R., & Buis, L. R. (2022). The Perceived Benefits of Digital Interventions for Behavioral Health: Qualitative Interview Study. *Journal of Medical Internet Research*, *24*(3), e34300. https://doi.org/10.2196/34300
Maxwell, A. (2009). How do business coaches experience the boundary between coaching and therapy/counselling? *Coaching: An International Journal of Theory, Research and Practice*, *2*(2), 149–162.
Maxwell, A. S. (2017). *Essential Executive Coaching Competencies for Enhancing Executive On-the-Job Performance: A Modified Delphi Study*. Walden University.
McCormick, I., & Burch, G. S. J. (2008). Personality-Focused Coaching for Leadership Development. *Consulting Psychology Journal*, *60*(3), 267–278.
McDowall, A. (2017). The Use of Psychological Assessments in Coaching and Coaching Research. In *The SAGE handbook of coaching* (pp. 627–644). Sage Publications, Inc.
McKelley, R. A., & Rochlen, A. B. (2007). The practice of coaching: Exploring alternatives to therapy for counseling-resistant men. *Psychology of Men and Masculinity*, *8*(1), 53–65.
Mehri, S., & Eskenazi, M. (2020). USR: An unsupervised and reference free evaluation metric for dialog generation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics* (pp. 681–707). https://doi.org/10.18653/v1/2020.acl-main.64
Meidenbauer, K. L., Niu, T., Choe, K. W., Stier, A. J., & Berman, M. G. (2023). Mouse movements reflect personality traits and task attentiveness in online experiments. *Journal of Personality*, *91*(2), 413–425. https://doi.org/10.1111/jopy.12736
Mercaldi, G. (2018). Is Developmental Coaching Morally Acceptable? *Philosophy of Coaching: An International Journal*, *3*(2), 60–72.
Mishra, P., Yannakoudakis, H., & Shutova, E. (2018). Neural Character-based Composition Models for Abuse Detection. Retrieved May 1, 2024, from https://arxiv.org/abs/1809.00378
Molyn, J. (2020). The Role and Effectiveness of Coaching in Increasing Self-Efficacy and Employability Efforts of Higher Education Students. In *MIT LINC 2019* (Vol. 3, pp. 178–187). Boston.
Mulvie, A. (2015). *The Value of Executive Coaching*. London & New York: Taylor & Francis Books.
Nadeem, I., Garvey, B., & Down, M. (2021). The Adequacy of Competency Frameworks for Coaching Academic Deans: A Critical Review. *International Journal of Evidence Based Coaching and Mentoring*, *19*(2), 3–23. https://doi.org/10.24384/ntr0-xf66
Nakano, M., & Komatani, K. (2019). A Framework for Building Closed-Domain Chat Dialogue Systems. Retrieved July 3, 2020, from http://arxiv.org/abs/1910.13826
Nangia, N., Vania, C., Bhalerao, R., & Bowman, S. R. (2020). CrowS-Pairs: A challenge dataset for measuring social biases in masked language models. https://doi.org/10.18653/v1/2020.emnlp-main.154
Nardo, C. (2023). The Waluigi Effect (mega-post). Retrieved September 21, 2023, from

https://www.lesswrong.com/posts/D7PumeYTDPfBTp3i7/the-waluigi-effect-mega-post
Neumeier, A., Wolf, T., & Oesterle, S. (2017). The Manifold Fruits of Digitalization - Determining the Literal Value Behind. In *13. Internationalen Tagung Wirtschaftsinformatik (WI 2017), St. Gallen* (pp. 484–498).
Newell, J. (2013). Obama: income inequality is "defining challenge of our time." Retrieved April 6, 2020, from https://www.theguardian.com/world/2013/dec/04/obama-income-inequality-minimum-wage-live
Neyman, C. (2017). A Survey of Addictive Software Design. Retrieved May 8, 2020, from http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1127&context=cscsp
Niezen, M., Adams, S. A., Nadya, P., & Vedder, A. H. (2016). Socially Robust eCoaching. Retrieved May 1, 2024, from https://www.researchgate.net/publication/299455962_Socially_Robust_eCoaching_Dealing_with_the_ethical_and_legal_preconditions_for_eCoaching_acceptance
Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). Capabilities of GPT-4 on Medical Challenge Problems. Retrieved September 25, 2023, from http://arxiv.org/abs/2303.13375
Nugroho, S., Sitorus, A. T., Habibi, M., Wihardjo, E., Iswahyudi, S., Informatika, P. S., … Jember, U. (2023). The Role of ChatGPT in Improving the Efficiency of Business Communication in Management Science. *Jurnal Minfo Polgan*, *12*(1), 1482–1491.
Nursing and Midwifery Council. (2019). Annual Fitness to Practise Report 2018–2019. Retrieved December 15, 2024, from https://www.nmc.org.uk/globalassets/sitedocuments/annual_reports_and_accounts/ftpannualreports/nmc-fitness-to-practise-report-2019-singles-linked-contents.pdf
O'Connor, S., & Cavanagh, M. (2013). The coaching ripple effect: The effects of developmental coaching on wellbeing across organisational networks. *Psychology of Well-Being: Theory, Research and Practice*, *3*(1), 2.
Obispo, S. L., Minicucci, A., Luis, S., Obispo, S. L., Minicucci, A., & Luis, S. (2016). Coaching and training system and method for integrated monitoring, managing, supporting, scheduling and prompting of clients. Retrieved May 1, 2024, from https://patents.google.com/patent/US9473593B2/en
Object Management Group (OMG). (2011). Business Process Model and Notation (BPMN). Retrieved May 1, 2024, from https://www.omg.org/spec/BPMN/2.0.2/PDF
Object Management Group (OMG). (2015). Business Motivation Model Specification. Retrieved April 11, 2020, from http://www.omg.org/spec/BMM/1.3/
Object Management Group (OMG). (2017). OMG Unified Modeling Language. Retrieved June 11, 2020, from https://www.omg.org/spec/UML/About-UML/
Ochoa, S. F., & Gutierrez, F. J. (2018). Architecting E-Coaching Systems: A First Step for Dealing with Their Intrinsic Design Complexity. *Computer*, *51*(3), 16–23. https://doi.org/10.1109/MC.2018.1731079
OECD. (2019a). *Measuring the Digital Transformation: A Roadmap for the Future*. *Measuring the Digital Transformation*. Paris: OECD Publishing. Retrieved from https://www.oecd.org/publications/measuring-the-digital-transformation-9789264311992-en.htm
OECD. (2019b). Society at a Glance 2019 OECD social indicators. Retrieved July 3, 2020, from https://www.oecd-ilibrary.org/docserver/soc_glance-2019-en.pdf?expires=1554852449&id=id&accname=ocid53016272&checksum=6777238BF6A43281648509F3F14691D3

Ohm, A., & Bhavani, K. (2019). Chatbot for Career Guidance Using AI. *International Journal of Computer Sciences and Engineering*, *7*(6), 856–860.
OpenAI. (2024a). How your data is used to improve model performance. Retrieved February 27, 2024, from https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance
OpenAI. (2024b). Introduction - OpenAI API. Retrieved February 27, 2024, from https://platform.openai.com/docs/introduction
OpenAI. (2024c). Models - Open AI API. Retrieved February 27, 2024, from https://platform.openai.com/docs/models
Oshlyansky, L., Cairns, P., & Thimbleby, H. (2007). Validating the Unified Theory of Acceptance and Use of Technology (UTAUT) tool cross-culturally. In *Proceedings of HCI 2007: The 21st British HCI Group Annual Conference* (Vol. 2, pp. 1–4). University of Lancaster. https://doi.org/10.14236/ewic/hci2007.67
Osterwalder, A., & Pigneur, Y. (2010). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers (Strategyzer)* (Kindle Edi). Hoboken, New Jersey: John Wiley & Sons, Inc.
Osterwalder, A., Pigneur, Y., Bernarda, G., & Smith, A. (Designer). (2014). *Value proposition design : how to create products and services customers want*. Hoboken: John Wiley & Sons.
Oxford University Press. (2020). Oxford Advanced Learner's Dictionary. Retrieved April 26, 2020, from https://www.oxfordlearnersdictionaries.com/definition/english/automation?q=automation
Özgen, S., Alabart, J. R., & Medir, M. (2008). Implementation of a 360-degree feedback assessment process for the development of the leadership competence of senior engineering students. In *36th European Society for Engineering Education (SEFI)* (pp. 1–8). Brussels.
Palmer, S., & Gyllensten, K. (2007). The coaching relationship: An interpretative phenomenological analysis. *International Coaching Psychology Review*, *2*(2), 168–177.
Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*(3), 1–12.
Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative Agents: Interactive Simulacra of Human Behavior. https://doi.org/10.1145/3586183.3606763
Passmore, Jonahtan. (2021). Future Trends in Coaching. Retrieved May 1, 2024, from https://emccdrive.emccglobal.org/api/file/download/nbjeQXTfcblSe1yaHQotwVoQzy2htUOapI06VomX
Passmore, Jonathan., Peterson, D. B., & Freire, T. (2013). *The Wiley-Blackwell handbook of the psychology of coaching and mentoring*. Chichester, West Sussex, UK: Wiley Blackwell.
Passmore, Jonathan, Brown, H., Csigas, Z., & European Coaching and Mentoring Research Consortium. (2017). The state of play in European coaching & mentoring - Executive report. Retrieved July 3, 2020, from https://researchportal.coachfederation.org/MediaStream/PartialView?documentId=3359
Passmore, Jonathan, & Woodward, W. (2023). Coaching education: Wake up to the new

digital and AI coaching revolution! *International Coaching Psychology Review*, *18*(1), 58–72. https://doi.org/10.53841/bpsicpr.2023.18.1.58
Paulick, A. (2023). evoach. Retrieved September 12, 2023, from https://www.evoach.com/
Pedro, R., Castro, D., Carreira, P., & Santos, N. (2023). From Prompt Injections to SQL Injection Attacks : How Protected is Your LLM-Integrated Web Application ? Retrieved September 20, 2023, from https://arxiv.org/abs/2308.01990
Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2008). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–78.
Peterson, D. B. (1996). Executive coaching at work: The art of one-on-one change. *Consulting Psychology Journal: Practice and Research*, *48*(2), 78–86.
PocketConfidant. (2023). PocketConfidant. Retrieved September 12, 2023, from https://pocketconfidant.com/individuals/coaching-plans/
Poria, S., Majumder, N., Mihalcea, R., & Hovy, E. (2019). Emotion Recognition in Conversation: Research Challenges, Datasets, and Recent Advances. *IEEE Access*, *7*(2019), 100943–100953. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8764449
Pratiwi, P. S., Tjondronegoro, D. W., Xu, Y., Li, Y., Trost, S., & Clanchy, K. (2018). Automated e-Coaching System Architecture Framework for Promoting Physical Activity. In *2018 International Conference on Intelligent Autonomous Systems, ICoIAS 2018* (pp. 88–91). https://doi.org/10.1109/ICoIAS.2018.8494163
Prywes, Y. (2012). *Examining the Influence of Goal Attainment Scaling on Changes in Goal Attainment in a Coaching Versus Non-Coaching Context*. Columbia University. Retrieved from https://academiccommons.columbia.edu/doi/10.7916/D8CC16QH/download
PwC, & ICF. (2016). 2016 ICF Global Coaching Study (EXECUTIVE SUMMARY). Retrieved May 3, 2020, from https://coachfederation.org/app/uploads/2017/12/2016ICFGlobalCoachingStudy_ExecutiveSummary-2.pdf
Rabbit Inc. (2024). Rabbit - Home. Retrieved May 1, 2024, from https://www.rabbit.tech/
Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. Retrieved September 25, 2023, from https://insightcivic.s3.us-east-1.amazonaws.com/language-models.pdf
Rao, A. S., & Verweij, G. (2017). What's the real value of AI for your business and how can you capitalise ? Retrieved May 3, 2020, from https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf
Ringle, C. M., Sarstedt, M., Mitchell, R., & Gudergan, S. P. (2020). Partial least squares structural equation modeling in HRM research. *International Journal of Human Resource Management*, *31*(12), 1617–1643. https://doi.org/10.1080/09585192.2017.1416655
Ringle, C. M., Wende, S., & Becker, J.-M. (2024). SmartPLS 4. Retrieved May 1, 2024, from https://www.smartpls.com
Roller, S., Dinan, E., Goyal, N., Ju, D., Williamson, M., Liu, Y., … Weston, J. (2020). Recipes for building an open-domain chatbot. Retrieved May 8, 2020, from http://arxiv.org/abs/2004.13637
Rose, B. (2014). The Sad Story of a Real Life R2-D2 Who Saved Countless Human Lives and Died. Retrieved April 7, 2020, from http://gizmodo.com/5870529/the-sad-story-of-a-

real-life-r2-d2-who-saved-countless-human-lives-and-died
Rubie-Davies, C. M. (2009). Teacher expectations and labeling. In *International Handbook of Research on Teachers and Teaching* (pp. 695–707). Boston: Springer.
Sallam, M. (2023). Practice : Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare*, *11*(6), 887.
Sarstedt, M., Ringle, C. M., Henseler, J., & Hair, J. F. (2014). On the Emancipation of PLS-SEM: A Commentary on Rigdon (2012). *Long Range Planning*, *47*(3), 154–160. https://doi.org/10.1016/j.lrp.2014.02.007
Schermuly, C. C. (2018). Client Dropout from business coaching. *Consulting Psychology Journal*, *70*(3), 250–267.
Schermuly, C. C., & Graßmann, C. (2019). A literature review on negative effects of coaching–what we know and what we need to know. *Coaching*, *12*(1), 39–66. https://doi.org/10.1080/17521882.2018.1528621
Schulman, J., Zoph, B., Kim, C., Hilton, J., Menick, J., Weng, J., … Hesse, C. (2023). Introducing ChatGPT. Retrieved August 30, 2023, from https://openai.com/blog/chatgpt
Scrum Alliance. (2015). Become A Scrum Alliance Certified Coach. Retrieved October 17, 2020, from https://www.scrumalliance.org/get-certified/become-a-coach
Sczesny, S., Bosak, J., Neff, D., & Schyns, B. (2004). Gender stereotypes and the attribution of leadership traits: A cross-cultural comparison. *Sex Roles*, *51*(11–12), 631–645. Retrieved from https://boris.unibe.ch/75304/1/2004_Sex Roles_Sczesny%2C Bosak%2C Neff%2C %26 Schyns_Gender stereotypes and the attribution of leadership traits.pdf
Sethi, B., Stubbings, C., Gratton, L., & Brown, J. (2017). Preparing for tomorrow's workforce. Retrieved May 3, 2020, from https://www.pwc.com/gx/en/people-organisation/pdf/pwc-preparing-for-tomorrows-workforce-today.pdf
Shakya, H. B., & Christakis, N. A. (2017). Association of Facebook Use with Compromised Well-Being: A Longitudinal Study. *American Journal of Epidemiology*, *185*(3), 203–211. https://doi.org/10.1093/aje/kww189
Sheng, E., Chang, K. W., Natarajan, P., & Peng, N. (2019). The woman worked as a babysitter: On biases in language generation. https://doi.org/10.18653/v1/d19-1339
Sherpa Coaching. (2019). Executive Coaching Survey Report. Retrieved January 15, 2020, from https://www.sherpacoaching.com/pdf_files/2019_Executive_Coaching_Survey_Summary_Report.pdf
Sherpa Consulting. (2018). The 2018 Executive Coaching Survey. Retrieved February 27, 2024, from https://de.scribd.com/document/434685457/2018-Executive-Coaching-Survey-Executive-Summary
Shrout, P. E., & Rodgers, J. L. (2018). Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis. *Annual Review of Psychology*, *69*(1), 487–510. https://doi.org/10.1146/annurev-psych-122216-011845
Singh, B., Olds, T., Brinsley, J., Dumuid, D., Virgara, R., Matricciani, L., … Maher, C. (2023). Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. *Npj Digital Medicine*, *6*(1), 1–10. https://doi.org/10.1038/s41746-023-00856-1
Smeeding, T. M., Jäntti, M., & Erikson, R. (2011). *Persistence, privilege, and parenting : the comparative study of intergenerational mobility* (1st ed.). New York: Russell Sage Foundation.

Smith, E. M., Williamson, M., Shuster, K., Weston, J., & Boureau, Y. (2020). Can You Put it All Together: Evaluating Conversational Agents' Ability to Blend Skills. Retrieved July 3, 2020, from http://arxiv.org/abs/2004.08449

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*(1), 333–339.

Spence, G. B. (2007). GAS powered coaching: Goal Attainment Scaling and its use in coaching research and practice. *International Coaching Psychology Review*, *2*(2), 155–167.

Spence, G. B., Cavanagh, M. J., & Grant, A. M. (2008). The integration of mindfulness training and health coaching: an exploratory study. *Coaching: An International Journal of Theory, Research and Practice*, *1*(2), 145–163.

Sperry, L. (2004). *Executive Coaching: The Essential Guide for Mental Health Professionals* (1 edition). Routledge.

Stec, D. (2012). Using history to comprehend the currency of a passionate profession. *Journal of Management History*, *18*(4), 419–444. https://doi.org/10.1108/17511341211258756

Stober, D. R., & Grant, A. (2006). *Evidence based coaching handbook : putting best practices to work for your clients*. Hoboken, N.J.: John Wiley & Sons.

Stout-Rostron, S. (2009). The global initiatives in the coaching field. *Coaching: An International Journal of Theory, Research and Practice*, *2*(1), 76–85.

Strubell, E., Ganesh, A., & Mccallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. Retrieved September 22, 2023, from https://arxiv.org/pdf/1906.02243.pdf%22%3EWachstum

Szabo, A. (2023). ChatGPT is a Breakthrough in Science and Education but Fails a Test in Sports and Exercise Psychology. *Baltic Journal of Sport and Health Sciences*, *1*(128), 25–40. https://doi.org/10.33607/bjshs.v1i128.1341

Tan, Y. C., & Celis, L. E. (2019). Assessing Social and Intersectional Biases in Contextualized Word Representations. In *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)* (pp. 1–12). Vancouver, Canada.

Tavanapour, N., & Bittner, E. A. C. (2018). Automated Facilitation for Idea Platforms: Design and Evaluation of a Chatbot Prototype. *International Conference on Information Systems (ICIS) Proceedings*, (October), 1–9.

Techstars. (2020). Rocky.AI. Retrieved April 9, 2020, from https://www.rocky.ai

ter Stal, S., Kramer, L. L., Tabak, M., op den Akker, H., & Hermens, H. (2020). Design Features of Embodied Conversational Agents in eHealth: a Literature Review. *International Journal of Human Computer Studies*, *138*(1), 102409. https://doi.org/10.1016/j.ijhcs.2020.102409

Terblanche, N. (2020). A design framework to create artificial intelligence coaches. *International Journal of Evidence Based Coaching and Mentoring*, *18*(2), 152–165. https://doi.org/10.24384/b7gs-3h05

Terblanche, N. (2023). coachvici. Retrieved September 12, 2023, from https://www.coachvici.com/

Terblanche, N., & Cilliers, D. (2020). Factors that influence users' adoption of being coached by an Artificial Intelligence Coach. *Philosophy of Coaching: An International Journal*, *5*(1), 61–70.

Terblanche, N., & Kidd, M. (2022). Adoption Factors and Moderating Effects of Age and Gender That Influence the Intention to Use a Non-Directive Reflective Coaching

Chatbot. *SAGE Open*, *12*(2), 1–16. Retrieved from https://journals.sagepub.com/doi/full/10.1177/21582440221096136
Terblanche, N., Molyn, J., Haan, E. de, & Nilsson, V. O. (2022). Comparing artificial intelligence and human coaching goal attainment efficacy. *PLOS ONE*, *17*(6), e0270255. https://doi.org/10.1371/JOURNAL.PONE.0270255
Terblanche, N., Molyn, J., Williams, K., & Maritz, J. (2023). Performance matters: students' perceptions of Artificial Intelligence Coach adoption factors. *Coaching: An International Journal of Theory, Research and Practice*, *16*(1), 100–114. https://doi.org/10.1080/17521882.2022.2094278
The Verge. (2023). Microsoft, GitHub, and OpenAI ask court to throw out AI copyright lawsuit. Retrieved September 22, 2023, from https://www.theverge.com/2023/1/28/23575919/microsoft-openai-github-dismiss-copilot-ai-copyright-lawsuit
Theeboom, T., Beersma, B., & van Vianen, A. E. M. (2014). Does coaching work? A meta-analysis on the effects of coaching on individual level outcomes in an organizational context. *Journal of Positive Psychology*, *9*(1), 1–18. https://doi.org/10.1080/17439760.2013.837499
Tkach, J., & DiGirolamo, J. (2017). The state and future of coaching supervision. *International Coaching Psychology Review*, *12*(1), 49–63.
Tomm, K. (1988). Interventive interviewing: Part III. Intending to ask linear, circular, strategic, or reflective questions? *Family Process*, *27*(1), 1–14.
Tompson, H. B., Bear, D. J., Dennis, D. J., Vickers, M., London, J., & Morrison, C. L. (2008). Coaching - A Global Study of Successful Practices Coaching - Current Trends and Future Possibilities 2008-2018. Retrieved July 3, 2020, from https://www.opm.gov/WIKI/uploads/docs/Wiki/OPM/training/i4cp-coaching.pdf
Tooth, J. A., Nielsen, S., & Armstrong, H. (2013). Coaching effectiveness survey instruments: Taking stock of measuring the immeasurable. *Coaching: An International Journal of Theory, Research & Practice*, *6*(2), 137–151. https://doi.org/10.1080/17521882.2013.802365
Tsiouris, K. M., Tsakanikas, V. D., Gatsios, D., & Fotiadis, D. I. (2020). A Review of Virtual Coaching Systems in Healthcare: Closing the Loop With Real-Time Feedback. *Frontiers in Digital Health*, *2*(September). https://doi.org/10.3389/fdgth.2020.567502
Tülüce, D., & Kutlutürkan, S. (2018). The effect of health coaching on treatment adherence, self-efficacy, and quality of life in patients with chronic obstructive pulmonary disease. *International Journal of Nursing Practice*, *24*(4), 1–11.
Tzafestas, S. G. (2018). Ethics in robotics and automation: a general view. *International Robotics & Automation Journal*, *4*(3), 229–234.
University of Gloucestershire. (2018). Research Ethics: A Handbook of Principles and Procedures. Retrieved July 3, 2020, from http://resources.glos.ac.uk/currentstudents/research/ethics/index.cfm
Valentine, M. (2019). Contextual Coaching: Leveraging Context for Alignment in the System. *Philosophy of Coaching: An International Journal*, *4*(1), 93–106.
van Coller-Peter, S., & Adriana de Vries, D. J. (2022). Towards building theory on coachee readiness. *International Journal of Evidence Based Coaching and Mentoring*, *20*(1), 50–65. https://doi.org/10.24384/HDS1-2J35
Vardhan, M., Hegde, N., Nathani, D., Rosenzweig, E., Karthikesalingam, A., & Seneviratne, M. (2023). Infusing behavior science into large language models for activity coaching.

Retrieved September 19, 2023, from https://www.medrxiv.org/content/10.1101/2023.03.31.23287995.full.pdf
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. Retrieved May 1, 2024, from https://arxiv.org/pdf/1706.03762.pdf
Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425–478.
Vukanović, Z. (2018). The influence of ict megatrends on global megatrends. *Informatologia*, *51*(1–2), 43–52.
W3Techs. (2023). W3Techs.com - Usage Statistics and Market Share of Content Languages for Websites, September 2023. Retrieved September 20, 2023, from https://w3techs.com/technologies/overview/content_language
Wagner, A. R., Borenstein, J., & Howard, A. (2018). Computing ethics overtrust in the robotic age. *Communications of the ACM*, *61*(9), 22–24.
Wakefield, B., Neustaedter, C., & Hillman, S. (2014). The informatics needs of amateur endurance athletic coaches. In *Conference on Human Factors in Computing Systems - Proceedings* (pp. 2287–2292).
Wang, L., & Wentling, T. L. (2001). The relationship between distance coaching and the transfer of training. In *Proceedings from the Academy of Human Resource Development Conference* (Vol. 1, p. 212).
Wang, Z., Xie, Q., Ding, Z., Feng, Y., & Xia, R. (2023). Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study. Retrieved September 25, 2023, from http://arxiv.org/abs/2304.04339
Weber, U., Lömker, M., & Moskaliuk, J. (2021). The Human Touch: The Impact of Anthropomorphism in Chatbots on the Perceived Success of Solution Focused Coaching. *Management Revue*, *32*(4), 385–407. https://doi.org/10.5771/0935-9915-2021-4-385
West, C. G. (2023). Advances in apparent conceptual physics reasoning in GPT-4. Retrieved September 25, 2023, from http://arxiv.org/abs/2303.17012
WHO. (2022). Staggering backsliding across women's, children's and adolescents' health revealed in new UN analysis. Retrieved October 18, 2022, from https://www.who.int/news/item/18-10-2022-staggering-backsliding-across-women-s--children-s-and-adolescents--health-revealed-in-new-un-analysis
Widdowson, L., Rochester, L., Barbour, P. J., & Hullinger, A. M. (2020). Bridging the team coaching competency gap: A review of the literature. *International Journal of Evidence Based Coaching and Mentoring*, *18*(2), 35–50. https://doi.org/10.24384/z9zb-hj74
Wieringa, R., Gordijn, J., & Eck, P. van. (2004). Value framing: a prelude to software problem framing. In *Proceedings of the 1st International Workshop on Advances and Applications of Problem Frames* (pp. 75–84).
Wirth, N. (1976). *Algorithms plus data structures equals programs*. *Englewood Cliffs*. New Jersey: Prentice-Hall.
Wu, X., Duan, R., & Ni, J. (2023). Unveiling Security, Privacy, and Ethical Concerns of ChatGPT. Retrieved September 23, 2023, from https://arxiv.org/pdf/2307.14192.pdf
Xue, F., Fu, Y., Zhou, W., Zheng, Z., & You, Y. (2023). To Repeat or Not To Repeat: Insights from Scaling LLM under Token-Crisis. Retrieved September 20, 2023, from http://arxiv.org/abs/2305.13230
Yao, H., Lou, J., Ren, K., & Qin, Z. (2023). PromptCARE : Prompt Copyright Protection by

Watermark Injection and Verification. Retrieved September 23, 2023, from https://arxiv.org/pdf/2308.02816.pdf

Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D., & Weston, J. (2018). Personalizing dialogue agents: I have a dog, do you have pets too? In *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)* (Vol. 1, pp. 2204–2213).

Zhang, W. N., Li, L., Cao, D., & Liu, T. (2018). Exploring implicit feedback for open domain conversation generation. In *32nd AAAI Conference on Artificial Intelligence, AAAI 2018* (pp. 547–554).

Zoldi, S. (2021). It's 2021. Do You Know What Your AI Is Doing? Retrieved September 20, 2023, from https://www.fico.com/blogs/its-2021-do-you-know-what-your-ai-doing

Appendices

Appendix A – EMCC codes

The following table contains assigned codes to identify themes related to the EMCC Competence Framework (EMCC, 2015) and the EMCC Competence Framework Glossary (EMCC, 2018).

Key	**Codes**
EMCC_CF_001	coach-behaviour
EMCC_CF_002	diversity
EMCC_CF_003	effective-communication
EMCC_CF_004	value-behaviour belief-behaviour
EMCC_CF_005	coaching-skill-practice evaluation-skills
EMCC_CF_006	explain-role
EMCC_CF_007	explain-benefit explain-benefit-context
EMCC_CF_008	confidentiality
EMCC_CF_009	session-outcome-discussion
EMCC_CF_010	behaviour-impact-coaching
EMCC_CF_011	respect dignity
EMCC_CF_012	rapport
EMCC_CF_013	coach-language
EMCC_CF_014	trust non-judgemental commitment
EMCC_CF_015	client-specific-learning
EMCC_CF_016	understanding-check
EMCC_CF_017	active-listening
EMCC_CF_018	explain-effective-questioning
EMCC_CF_019	feedback style-useful style-acceptable style-meaningful
EMCC_CF_020	client-autonomy offer-perspective offer-idea
EMCC_CF_021	coaching-outcome clarify-outcome review-outcomes goal-setting
EMCC_CF_022	goal-alignment goal-alignment-context
EMCC_CF_023	options
EMCC_CF_024	solution-selection
EMCC_CF_025	noting track-progress review-progress client-progress
EMCC_CF_026	client-autonomy client-autonomy-after-session
EMCC_CF_027	model model-approach
EMCC_CF_028	monitor-effectiveness reflect-effectiveness
EMCC_CF_029	feedback feedback-request
EMCC_CF_030	feedback feedback-receive feedback-accept
EMCC_CF_031	coach-self-understanding model-human-behaviour reflect-practice
EMCC_CF_032	psychological-process-impact adapt-behaviour
EMCC_CF_033	empathy client-emotion
EMCC_CF_034	continuous-personal-development
EMCC_CF_035	supervision
EMCC_CF_036	supervision-effectiveness-evaluation
EMCC_CF_037	code-of-ethic

EMCC_CF_038	explain-modalities explain-benefit explain-benefit-context
EMCC_CF_039	contract-management
EMCC_CF_040	scheduling location
EMCC_CF_041	explain-coaching-process explain-coaching-style
EMCC_CF_042	competence-boundary referral-others-competence contract-closure
EMCC_CF_043	client-unfit
EMCC_CF_044	client-preference
EMCC_CF_045	contract-management contract-closure
EMCC_CF_046	empathy support
EMCC_CF_047	level-of-trust trust
EMCC_CF_048	emotional-state client-emotion
EMCC_CF_049	adapt-language adapt-behaviour
EMCC_CF_050	independence client-autonomy
EMCC_CF_051	effective-listening-block effective-listening
EMCC_CF_052	non-verbal-communication communication-content
EMCC_CF_053	thinking-pattern client-pattern
EMCC_CF_054	connecting-feeling-behaviour-performance
EMCC_CF_055	questioning-range
EMCC_CF_056	new-idea-creation
EMCC_CF_057	feedback perspective rapport responsibility
EMCC_CF_058	impartial options
EMCC_CF_059	review understanding commitment commitment-action
EMCC_CF_060	action-planning support resourcing contingencies
EMCC_CF_061	client-preference action-identification
EMCC_CF_062	responsibility responsibility-decision responsibility-action responsibility-learning
EMCC_CF_063	potential-barrier action-planning
EMCC_CF_064	commitment-outcome commitment-goal commitment-action
EMCC_CF_065	review-progress review-outcomes review-goals plan-adjustment
EMCC_CF_066	model-coaching
EMCC_CF_067	tool-usage
EMCC_CF_068	client-context-tool
EMCC_CF_069	feedback feedback-process-client
EMCC_CF_070	evaluation-process
EMCC_CF_071	evaluation-outcome
EMCC_CF_072	evaluation-coaching
EMCC_CF_073	coach-self-understanding model-based structured-input reflect-experience
EMCC_CF_074	self-regulation
EMCC_CF_075	review-belief review-attitude review-skill reflect-belief reflect-attitude reflect-skill
EMCC_CF_076	skill-gap knowledge-gap attitude-gap structured-process
EMCC_CF_077	explore-idea explore-themes explore-model
EMCC_CF_078	learning-into-practice evaluation-goal evaluation-process
EMCC_CF_079	feedback feedback-request feedback-peer
EMCC_CF_080	contract-management contract-ethic

EMCC_CF_081	client-identification-emotional client-identification-therapeutic
EMCC_CF_082	client-emotion client-mood client-language client-pattern client-belief non-verbal-communication
EMCC_CF_083	client-responsiveness client-attentiveness
EMCC_CF_084	raise-awareness exploration insight technique
EMCC_CF_085	feedback challenge raise-awareness insight responsibility-action
EMCC_CF_086	non-verbal-communication
EMCC_CF_087	questioning-range insight
EMCC_CF_088	coach-language reframe challenge
EMCC_CF_089	holistic understanding insight
EMCC_CF_090	client-context uncertainty possibility constraint
EMCC_CF_091	client-context-wide
EMCC_CF_092	technique-range method-range
EMCC_CF_093	method-range commitment commitment-outcome commitment-goal commitment-action
EMCC_CF_094	explore-client-change experimentation self-discovery
EMCC_CF_095	change-resistance
EMCC_CF_096	model-approach model-coaching approach-substantiation
EMCC_CF_097	model tool technique
EMCC_CF_098	approach-critique
EMCC_CF_099	insight-synthesis personal-evidence
EMCC_CF_100	reflect-client-interaction reflect-coaching
EMCC_CF_101	reflect-paradigm paradigm-impact
EMCC_CF_102	continuous-professional-development research research-evaluation
EMCC_CF_103	referral-others self-referral
EMCC_CF_104	client-need-boundary
EMCC_CF_105	tactic-explanation client-reaction
EMCC_CF_106	client-need-complexity
EMCC_CF_107	thinking-shift behaviour-shift
EMCC_CF_108	adapt-behaviour adapt-technique session-focus
EMCC_CF_109	unique-approach evaluation-model learning practice-learning supervision
EMCC_CF_110	tool-invention system-invention effectiveness
EMCC_CF_111	evaluation-knowledge
EMCC_CF_112	comment-theme comment-trend comment-idea evaluation-knowledge
GCoE.2.1	code-of-ethic
GCoE.2.2	code-of-ethic-sharing report-breach
GCoE.2.3	explain-coaching-process explain-coaching-terms
GCoE.2.4	client-need
GCoE.2.5	explain-coaching-process explain-method
GCoE.2.6	client-autonomy independence contract-management contract-duration
GCoE.2.7	environment
GCoE.2.8	interest-conflict-management client-need-first
GCoE.2.9	qualification-representation
GCoE.2.10	explain-benefit honestly accurately
GCoE.2.11	qualification-representation intellectual-property

GCoE.2.12	pertinent-law
GCoE.2.13	confidentiality
GCoE.2.14	confidentiality explain-confidentiality-limit
GCoE.2.15	confidentiality security privacy
GCoE.2.16	supervision supervision-confidentiality
GCoE.2.17	child vulnerable-adult
GCoE.2.18	communication-boundary cultural-boundary
GCoE.2.19	romantic-relationship
GCoE.2.20	client-exploitation
GCoE.2.21	interest-conflict-management professional-relationship
GCoE.2.22	interest-conflict-management
GCoE.2.23	interest-conflict-management interest-conflict-relationship
GCoE.2.24	interest-conflict-communication contract-closure-interest
GCoE.2.25	contract-management contract-closure contract-closure-anytime
GCoE.2.26	referral-others-value contract-closure
GCoE.2.27	confidentiality interest-conflict-management client-exploitation follow-up
GCoE.2.28	client-transfer
GCoE.3.1	professionalism
GCoE.3.2	respect-other-professionals
GCoE.3.3	diversity
GCoE.3.4	discrimination
GCoE.3.5	unconscious-bias
GCoE.3.6	discrimination-challenge
GCoE.3.7	discrimination discrimination-inadvertent
GCoE.3.8	equality-training diversity-training continuous-ethical-development
GCoE.3.9	breach-complaint
GCoE.3.10	report-breach
GCoE.3.11	pertinent-law
GCoE.3.12	indemnity-insurance
GCoE.4.1	referral-others-competence competence-boundary
GCoE.4.2	fit-for-practice healthy-for-practice
GCoE.4.3	supervision seeking-help
GCoE.4.4	supervision-relationship
GCoE.4.5	ethical-dilemma breach-discussion
GCoE.4.6	continuous-professional-development
GCoE.4.7	community-contribution
GCoE.4.8	feedback evaluation-system

Appendix A - Codes – EMCC Competency Framework, Global Code of Ethics

Appendix B – ICF codes

The following table contains the codes associated with the updated ICF core competency model released in 2019 October (ICF, 2019a) and the ICF Code of Ethics (ICF, 2019b).

Key	Codes
ICoE.E.1	explain-benefit explain-coaching-process confidentiality explain-confidentiality-limit explain-coaching-terms
ICoE.E.2	contract-management contract-role contract-responsibility contract-right responsibility
ICoE.E.3	confidentiality pertinent-law
ICoE.E.4	communication-channel
ICoE.E.5	confidentiality explain-confidentiality-limit
ICoE.E.6	interest-conflict interest-conflict-management
ICoE.E.7	confidentiality security privacy
ICoE.E.8	referral-others-value contract-closure benefit-change-recognition
ICoE.E.9	contract-management contract-closure contract-closure-anytime contract-closure-noreason
ICoE.E.10	interest-conflict interest-conflict-management
ICoE.E.11	power-difference-management
ICoE.E.12	third-party-communication
ICoE.E.13	service-consistency
ICoE.E.14	code-of-ethic breach-discussion report-breach
ICoE.E.15	code-of-ethic-support-personnel
ICoE.E.16	continuous-professional-development continuous-personal-development continuous-ethical-development
ICoE.E.17	competence-boundary contract-closure referral-others-competence
ICoE.E.18	interest-conflict interest-conflict-management contract-closure-interest
ICoE.E.19	privacy-ICF-Member
ICoE.E.20	qualification-representation
ICoE.E.21	explain-benefit explain-coaching-process explain-role
ICoE.E.22	code-of-ethic code-of-ethic-sharing
ICoE.E.23	communication-boundary cultural-boundary
ICoE.E.24	romantic-relationship
ICoE.E.25	discrimination equality cultural-sensitivity
ICoE.E.26	intellectual-property
ICoE.E.27	competence-boundary research-reporting
ICoE.E.28	client-context-wide
ICF_CC_A.1.1	integrity honesty
ICF_CC_A.1.2	client-identity client-context client-experience client-value client-belief
ICF_CC_A.1.3	client-language coach-language respect
ICF_CC_A.1.4	code-of-ethic
ICF_CC_A.1.5	confidentiality pertinent-law
ICF_CC_A.1.6	explain-modalities
ICF_CC_A.1.7	referral-others
ICF_CC_A.2.1	responsibility responsibility-decision

ICF_CC_A.2.2	continuous-professional-development
ICF_CC_A.2.3	reflect-coaching
ICF_CC_A.2.4	client-context coach-context coach-culture client-culture
ICF_CC_A.2.5	coach-self-understanding intuition
ICF_CC_A.2.6	self-regulation
ICF_CC_A.2.7	session-preparation
ICF_CC_A.2.8	seeking-help
ICF_CC_B.3.1	explain-coaching-process explain-role
ICF_CC_B.3.2	explain-coaching-terms responsibility professional-relationship contract-management
ICF_CC_B.3.3	contract-management logistics scheduling location confidentiality contract-closure
ICF_CC_B.3.4	goal-setting action-planning coaching-outcome
ICF_CC_B.3.5	compatibility
ICF_CC_B.3.6	session-outcome-discussion
ICF_CC_B.3.7	session-outcome-discussion client-belief client-need
ICF_CC_B.3.8	session-outcome-discussion evaluation-session evaluation-coaching evaluation-outcome
ICF_CC_B.3.9	session-focus time-focus
ICF_CC_B.3.10	plan-adjustment session-focus outcome-focus
ICF_CC_B.3.11	contract-closure positive-coach-experience
ICF_CC_B.4.1	client-context
ICF_CC_B.4.2	respect client-identity client-perception client-style client-language adapt-behaviour
ICF_CC_B.4.3	respect acknowledgement client-talent client-insight client-effort
ICF_CC_B.4.4	empathy support coach-interest
ICF_CC_B.4.5	support acknowledgement feeling-expression perception-expression problem-expression belief-expression suggestion-expression
ICF_CC_B.4.6	openness transparency coach-vulnerability trust
ICF_CC_B.5.1	client-focus client-responsiveness client-attentiveness empathy
ICF_CC_B.5.2	coach-interest
ICF_CC_B.5.3	self-regulation
ICF_CC_B.5.4	client-emotion
ICF_CC_B.5.5	comfortable-not-knowing
ICF_CC_B.5.6	silence pause reflection
ICF_CC_C.6.1	client-context client-identity client-experience client-value client-belief
ICF_CC_C.6.2	understanding-check reflect-client-interaction review-interaction summarize-interaction
ICF_CC_C.6.3	non-verbal-communication communication-content
ICF_CC_C.6.4	client-emotion client-energy-shift non-verbal-communication client-behaviour
ICF_CC_C.6.5	communication-content non-verbal-communication
ICF_CC_C.6.6	client-pattern client-trend behaviour-pattern
ICF_CC_C.7.1	client-experience
ICF_CC_C.7.2	challenge raise-awareness insight
ICF_CC_C.7.3	questioning-range client-thinking client-value client-need client-belief

ICF_CC_C.7.4	client-explore self-discovery
ICF_CC_C.7.5	client-experience
ICF_CC_C.7.6	client-progress identify-success
ICF_CC_C.7.7	adapt-behaviour
ICF_CC_C.7.8	thinking-pattern client-pattern emotion-pattern behaviour-pattern identify-factors
ICF_CC_C.7.9	new-idea-creation action-planning
ICF_CC_C.7.10	reframe
ICF_CC_C.7.11	insight share-feeling share-observation impartial
ICF_CC_C.8.1	integration-insight integration-learning client-context-wide client-behaviour
ICF_CC_C.8.2	goal-setting action-planning accountability
ICF_CC_C.8.3	acknowledgement support client-autonomy
ICF_CC_C.8.4	review-outcomes review-learning
ICF_CC_C.8.5	action-planning support resourcing potential-barrier
ICF_CC_C.8.6	review-learning review-insight
ICF_CC_C.8.7	celebrate
ICF_CC_C.8.8	session-closure

Appendix B - Codes – ICF Core Competency Model, ICF Code of Ethics

Appendix C – Common codes

The following list contains the common codes in the ICF and in the EMCC documents.

Common codes
action-planning adapt-behaviour breach-discussion challenge client-attentiveness client-autonomy client-belief client-context client-context-wide client-emotion client-language client-need client-pattern client-progress client-responsiveness coach-language coach-self-understanding coaching-outcome code-of-ethic code-of-ethic-sharing communication-boundary communication-content competence-boundary confidentiality continuous-ethical-development continuous-personal-development continuous-professional-development contract-closure contract-closure-anytime contract-closure-interest contract-management cultural-boundary discrimination empathy evaluation-coaching evaluation-outcome explain-benefit explain-coaching-process explain-coaching-terms explain-confidentiality-limit explain-modalities explain-role goal-setting impartial insight intellectual-property interest-conflict-management location new-idea-creation non-verbal-communication pertinent-law plan-adjustment potential-barrier privacy professional-relationship qualification-representation questioning-range raise-awareness referral-others referral-others-competence referral-others-value reflect-client-interaction reflect-coaching reframe report-breach resourcing respect responsibility responsibility-decision review-outcomes romantic-relationship scheduling security seeking-help self-discovery self-regulation session-focus session-outcome-discussion support thinking-pattern trust understanding-check

Appendix C - Common Codes in ICF and EMCC

Appendix D – Coaching Common Themes

The following list contains the summary of coaching common themes together with the assigned theme groups and common codes used during the analysis.

Key	Theme Group	Code	Common Theme Summary
CT.1	code of ethic	code-of-ethic	abiding the defined professional code of ethic
CT.2	code of ethic	report-breach	reporting when the coach (or other professionals) breaches the code of ethics
CT.3	code of ethic	breach-discussion	challenging, discussing when the coach (or other professionals) breaches the code of ethics
CT.4	code of ethic	professional-relationship, romantic-relationship	managing relationship boundaries: physical intimacy
CT.5	code of ethic	communication-boundary, cultural-boundary	managing relationship boundaries: culturally sensitive interactions
CT.6	code of ethic	interest-conflict-management	managing potential or actual professional interest conflict
CT.7	code of ethic	privacy	ensuring privacy in data management and communication
CT.8	code of ethic	security	ensuring security in data management and communication
CT.9	code of ethic	qualification-representation	representing honestly the level of coaching competence and coaching experience
CT.10	code of ethic	continuous-ethical-development	continuous development of code of ethic and managing issues related to the code of ethic
CT.11	code of ethic	continuous-personal-development	continuous development of projected coach personality, related communication, and behaviour
CT.12	code of ethic	continuous-professional-development	continuous development of coaching related communication and behaviour
CT.13	code of ethic	discrimination	avoiding discrimination on any grounds in all operations
CT.14	code of ethic	intellectual-property	honouring the intellectual property of others

CT.15	code of ethic	referral-others, referral-others-value	client referral to other professionals when client value gained from coaching is shifting
CT.16	code of ethic	referral-others, referral-others-competence	client referral to other professionals when coach competence boundary reached
CT.17	communication before coaching	code-of-ethic-sharing	communication of commitment to the defined professional code of ethic
CT.18	communication before coaching	explain-benefit	communication of potential value created by coaching
CT.19	communication before coaching	explain-role	communication of what coaching is related to the client
CT.20	communication before coaching	explain-modalities	communication of coaching modality difference to other supporting professions
CT.21	communication before coaching	explain-coaching-process	communication of the coaching process
CT.22	communication before coaching	explain-coaching-terms	communication of coaching service terms and conditions
CT.23	communication before coaching	explain-confidentiality-limit	communication of limits of confidentiality
CT.24	agreement management	contract-management	managing clear agreements
CT.25	agreement management	location	agreement on location of service
CT.26	agreement management	scheduling	agreement on session scheduling: date/time, frequency, duration
CT.27	agreement management	contract-closure-anytime	service termination conditions: any time indicated by the client but subject of agreement termination conditions
CT.28	agreement management	contract-closure	service termination conditions: client value gained from coaching is shifting
CT.29	agreement management	contract-closure-interest	service termination conditions: unresolvable interest conflict

CT.30	agreement management	competence-boundary	service termination conditions: coach competence boundary reached
CT.31	coach behaviour	confidentiality	ensuring confidentiality about personal data and communications
CT.32	coach behaviour	pertinent-law	ensuring confidentiality limit based on pertinent law
CT.33	coach behaviour	responsibility	keeping client responsibility: decision, action, learning approach
CT.34	coach behaviour	responsibility-decision	acknowledging that clients are responsible for their own choices
CT.35	coach behaviour	adapt-behaviour	adapting the coaching behaviour to the client
CT.36	coach behaviour	challenge	challenging the client
CT.37	coach behaviour	client-autonomy	actively promoting client autonomy, independence from coach
CT.38	coach behaviour	client-attentiveness	remaining observant to the client
CT.39	coach behaviour	client-belief	working with the client's belief
CT.40	coach behaviour	client-context	is sensitive to the client's context
CT.41	coach behaviour	client-emotion	working with the client's emotion
CT.42	coach behaviour	coach-language	using appropriate language to the client
CT.43	coach behaviour	client-language	working with the client language
CT.44	coach behaviour	client-pattern	working with the client patterns
CT.45	coach behaviour	thinking-pattern	noticing client thinking patterns
CT.46	coach behaviour	client-responsiveness	remaining responsive to the client
CT.47	coach behaviour	communication-content	integrating communication content with non-verbal communication
CT.48	coach behaviour	empathy	remaining empathetic to the client
CT.49	coach behaviour	impartial	remaining impartial
CT.50	coach behaviour	non-verbal-communication	recognizing non-verbal communication
CT.51	coach behaviour	questioning-range	using a wide range of questioning
CT.52	coach behaviour	understanding-check	ensuring understanding of key issues
CT.53	coach behaviour	self-regulation	managing the emotions of the coach
CT.54	coach behaviour	raise-awareness	raising awareness of the client
CT.55	coach behaviour	reflect-client-interaction	reflecting on the interaction with the client
CT.56	coach behaviour	reflect-coaching	reflecting on the coaching

CT.57	coach behaviour	respect	demonstrating respect to the client
CT.58	coach behaviour	seeking-help	seeking help from outside sources when necessary
CT.59	coach behaviour	coach-self-understanding	using awareness of the self of coach
CT.60	coach behaviour	plan-adjustment	changing coaching direction based on indication from the client
CT.61	coach behaviour	trust	building trust with the client
CT.62	partnering behaviour	client-need	partnering with the client to define the client's need
CT.63	partnering behaviour	client-context-wide	partnering with the client to build awareness of impact on the client's wider context
CT.64	partnering behaviour	goal-setting	partnering with the client to establish goals
CT.65	partnering behaviour	coaching-outcome	partnering with the client to establish overall coaching outcome
CT.66	partnering behaviour	review-outcomes	partnering with the client to identify potential outcomes
CT.67	partnering behaviour	session-outcome-discussion	partnering with the client to clarify and reconfirm session outcome
CT.68	partnering behaviour	session-focus	partnering with the client to manage focus of the session
CT.69	partnering behaviour	action-planning	partnering with the client to effectively plan actions
CT.70	partnering behaviour	potential-barrier	partnering with the client to effectively plan actions by identifying barriers
CT.71	partnering behaviour	support	partnering with the client to effectively plan actions by identifying support
CT.72	partnering behaviour	resourcing	partnering with the client to effectively plan actions by identifying resources
CT.73	partnering behaviour	evaluation-outcome	partnering with the client to define and reconfirm measures of success for coaching session outcome
CT.74	partnering behaviour	evaluation-coaching	partnering with the client to define and reconfirm measures of success for coaching engagement outcome
CT.75	partnering behaviour	self-discovery	partnering with the client to explore beyond current thinking
CT.76	partnering behaviour	insight	partnering with the client to evoke client insight
CT.77	partnering behaviour	new-idea-creation	partnering with the client to generate ideas
CT.78	partnering behaviour	reframe	partnering with the client in reframing
CT.79	partnering behaviour	client-progress	partnering with the client to enhance client progress

Appendix D - Coaching Common Themes

Appendix E – Initial Solution Requirement List

The following table contains the initial version of the solution requirement summary related to an autonomous coaching service.

Key	Problem Reference	Common Theme Reference	Objective Reference	Quality Model Characteristics	Solution Requirement Summary
SR.1	P.20, P.28, P.29, P.32, P.37, P.38, P.39, P.42, P.44, P.45, P.46, P.47, P.60, P.61, P.62, P.63	CT.1, CT.2, CT.3, CT.4, CT.5, CT.6, CT.7, CT.8, CT.9, CT.10, CT.11, CT.12, CT.13, CT.14, CT.15, CT.16	O.2, O.10	Functional suitability	An openly accessible professional ethical code is created with at least the following topics: ethical code breach reporting, relationship boundaries, potential interest conflicts, data management, honest system capability representation, continuous development, discrimination, intellectual property, user referral and technology specific ethical concerns
SR.2	P.1, P.2, P.3, P.31, P.32	CT.17, CT.18, CT.19, CT.20, CT.21, CT.22, CT.23	O.2, O.11	Functional suitability	The user needs to be informed before the first coaching conversation at least about the professional code of ethic, coaching value, coaching modality, coaching process, terms and conditions and confidentiality limits
SR.3	P.31, P.58	CT.17, CT.18, CT.19, CT.20, CT.21, CT.22, CT.23	O.2, O.11	Functional suitability	The user is required to accept the terms and conditions before the first coaching conversation takes place
SR.4	P.6	CT.37	O.2, O.11	Functional suitability	The user can initiate a coaching conversation after acceptance of terms and conditions any time
SR.5			O.2, O.11	Functional suitability	The status of the user is tracked throughout the service lifecycle of the user with at least the following stages: new user, active coaching user, inactive coaching user
SR.6	P.48	CT.62	O.2	Functional suitability	Managing information about the user needs is required throughout the service lifecycle of the user
SR.7		CT.40, CT.63	O.2	Functional suitability	Managing information about the user context is required throughout the service lifecycle of the user
SR.8	P.11	CT.39, CT.41, CT.43, CT.44, CT.45	O.2, O.7	Functional suitability	User pattern collection and usage in interactions is required for the following areas: emotion, thinking, belief and language

SR.9	P.6	CT.26	O.2, O.9	Functional suitability	Service usage agreement is managed on two levels: committed time related to the service in general and individual coaching conversation duration
SR.10	P.56	CT.64, CT.65, CT.66, CT.67	O.2	Functional suitability	Outcomes are collected from interactions with the user related to individual coaching conversations as well as for the overall coaching engagement
SR.11	P.48	CT.64, CT.65, CT.66, CT.67, CT.73, CT.74	O.2	Functional suitability	At least the following attributes are associated and managed related to outcomes: type, level, description, measure, and status
SR.12		CT.47, CT.55	O.2	Functional suitability	Conversation history is maintained on an individual message level with at least the following attributes: message creator, message, timestamp, scheduling, session, and meta-data
SR.13	P.24, P.32	CT.12	O.2, O.5, O.6	Functional suitability	Aggregated statistical data in the form of reporting about users, conversations and outcomes is collected and used for continuous service operation and optimisation
SR.14	P.50, P.51, P.52, P.53, P.54	CT.11, CT.59	O.2	Functional suitability	The user can choose from a set of coach profiles that describes behavioural preferences of the coaching service
SR.15		CT.4, CT.5	O.2	Functional suitability	Dialog management related knowledge base is maintained for general conversations following social norms with at least the following attributes: intentions and response recognition patterns
SR.16		CT.*	O.2	Functional suitability	Dialog management related knowledge base is maintained for conducting coaching conversations with at least the following attributes: intentions, and response recognition patterns
SR.17	P.48	CT.62, CT.64, CT.65, CT.66, CT.67	O.2	Functional suitability	Dialog management related knowledge base is maintained for conducting user need specific conversations with at least the following attributes: intentions, and response recognition patterns
SR.18		CT.26	O.2	Functional suitability	System messages are only generated when the coaching conversation is coming to its end or when the user stopped communicating to avoid interruptions
SR.19		CT.38, CT.46	O.2	Functional suitability	User messages are interpreted based on the response recognition patterns defined in the conversational knowledge base
SR.20		CT.38, CT.46	O.2	Functional suitability	The information in the data model is updated based on the interpreted messages

SR.21			O.2	Functional suitability	Interpreted messages and the updated information in the data model is used to identify next coaching intentions
SR.22		CT.11, CT.26, CT.35, CT.40, CT.60, CT.62	O.2	Functional suitability	Conversational intentions are determined at least on the following factors: user need, user context, time, coach profile, and exception handling
SR.23	P.36	CT.33, CT.34, CT.36, CT.37, CT.38, CT.42, CT.46, CT.47, CT.48, CT.49, CT.50, CT.51, CT.52, CT.53, CT.54, CT.55, CT.56, CT.57, CT.61	O.2	Functional suitability	The system demonstrates the following intentions: acknowledging and keeping user responsible, challenging the user, actively promoting user autonomy, remaining observant, appropriate language usage, remaining responsive, integrating verbal and non-verbal communication, remaining empathetic and impartial, using a wide range of questioning, ensuring understanding of key issues, displaying emotions, raising awareness of the user, reflecting on the interactions with the user, respect, building trust
SR.24	P.49	CT.62, CT.63, CT.64, CT.65, CT.66, CT.67, CT.68, CT.69, CT.70, CT.71, CT.72, CT.73, CT.74, CT.75, CT.76, CT.77, CT.78, CT.79	O.1	Functional suitability	The system supports the user in the following areas: defining user's need, building awareness of the impact on the user's wider context, establishing and reconfirming goals, overall outcomes, conversation outcomes and related measures of success, managing conversation focus, planning actions, exploring beyond current thinking, evoking insight, generating ideas, reframing, and enhancement of user progress
SR.25	P.55, P.56, P.57	CT.62	O.2	Functional suitability	Information about the implementation of coaching conversation outcomes between conversations and new learning from the implementation is used as an additional input for the following coaching conversations
SR.26		CT.58	O.2	Functional suitability	Seeking help from outside sources when necessary
SR.27	P.9		O.4	Functional suitability	Provides private community features for users pursuing similar goals
SR.28	P.58	CT.27	O.9, O.11	Functional suitability	A coaching conversation is terminated based on user request anytime or when the planned duration is reached

SR.29	P.58	CT.27, CT.28, CT.29, CT.30	O.11, O.14	Functional suitability	Engagement closure is managed in planned case as well as early termination case, such as shifting user value, unresolvable interest conflict, or reaching system competency boundary
SR.30		CT.35, CT.38, CT.46		Performance efficiency	Message response time of the system in a coaching conversation is higher than the response time of the user
SR.31	P.23, P.46		O.8	Performance efficiency	The resource utilization aspect needs to be considered for software design decisions
SR.32	P.6, P.7, P.17, P.59		O.8, O.14	Performance efficiency	System scalability is required to support a high and varying number of parallel user interactions
SR.33	P.12, P.14, P.16		O.8, O.13	Compatibility	Effective interoperability is ensured with the LLM system
SR.34	P.5		O.12	Usability	Appropriateness recognizability, learnability, operability, user error protection, user interface aesthetics and accessibility aspects are required to be considered at software design and continuous development
SR.35	P.6, P.7	CT.38, CT.46	O.3, O.12	Reliability	High availability is required to provide partnership in coaching conversations anytime
SR.36	P.27, P.35, P.41, P.42, P.43	CT.7, CT.8, CT.23, CT.31, CT.32	O.15	Security	The system requires at least one user authentication and authorization method
SR.37	P.27, P.35, P.41, P.42, P.43	CT.23, CT.31, CT.32	O.15	Security	Confidentiality is ensured during communication with the user and related to data management processes within the limitations of the pertinent law
SR.38	P.18, P.59		O.8	Maintainability	Modularity, reusability, analysability, modifiability, testability, adaptability, installability and replaceability are required to be considered at software design and continuous development

Appendix E - Initial Solution Requirement List