Thesis for The Degree of Doctor of Philosophy

From Stress to Strength: Well-Being and Resilience in Software Engineering

Cristina Martinez Montes

Division of Software Engineering
Department of Computer Science & Engineering
Chalmers University of Technology and Gothenburg University
Gothenburg, Sweden, 2026

From Stress to Strength: Well-Being and Resilience in Software Engineering

Cristina Martinez Montes

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5825
ISSN 978-91-8103-368-7
https://doi.org/10.63959/chalmers.dt/5825

Department of Computer Science & Engineering
Division of Software Engineering
Chalmers University of Technology and Gothenburg University
SE-412 96 Göteborg,
Sweden
Phone: +46(0)31 772 1000

This thesis has been prepared using LaTeX.
Printed by Chalmers Digitaltryck,
Gothenburg, Sweden 2026.

“”When we are no longer able to change a situation, we are challenged to change ourselves.”
- Viktor Frankl (Neurologist/Psychiatrist)

From Stress to Strength: Well-Being and Resilience in Software Engineering

Cristina Martinez Montes

Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

Software engineers face unique circumstances that shape a specific work context distinct from many other professions. They experience frequent stress due to tight deadlines, heavy cognitive load demands and the constantly changing technology they work with. Hence, it is necessary to pay special attention to engineers' well-being, stress management and resilience. General theories of well-being address several aspects that engineers face. However, due to their specific characteristics, these theories require adaptation to capture the distinct pressures and contextual demands of software engineering work. Moreover, current methodologies require refinement through data triangulation and context-sensitive approaches. Single-source data often falls short in capturing the full experiences, perceptions, and context of engineers.

This thesis aimed to develop a software engineering well-being framework that considers the field's unique circumstances. In addition, it sought to design, test and evaluate interventions targeting engineers' well-being and stress management. Finally, it also investigated a suitable methodological approach that incorporates data triangulation to better capture the complexity of software engineering contexts.

Various empirical methodologies were employed, including interventions, quasi-experiments, experiments, and surveys. The data were analysed using thematic and content analysis for the qualitative data, and descriptive, frequentist, and Bayesian statistics for the quantitative data.

The main outcomes are: First, results provide a context-specific software engineering well-being framework. Second, we present tailored interventions targeting stress and well-being, developed considering engineers' unique circumstances. Third, we propose a data-triangulation approach for data collection and analysis. Finally, they introduce a framework for integrating AI into qualitative data analysis.

The thesis contributions advance the state of the art by offering a framework that explains factors influencing the well-being of software engineers. This framework also offers policy recommendations and interventions to enhance work environments that support well-being. Finally, we advance human factors research with our data triangulation proposal and a hybrid qualitative data analysis framework.

Keywords

Well-being, Resilience, Stress, Software Engineers, Human Factors

List of Publications

This thesis is based on the following publications:

[A] C. Martinez Montes, B. Penzenstadler, R. Feldt "The Factors Influencing Well-Being in Software Engineers: A Mixed-Method Study"
Transactions on Software Engineering and Methodology (TOSEM), 2025. DOI: 10.1145/3770074.

[B] C. Martinez Montes, R. Khojah "Emotional Strain and Frustration in LLM Interactions in Software Engineering"
International Conference on Evaluation and Assessment in Software Engineering (EASE), 2025. DOI: 10.1145/3756681.3756951.

[C] B. Penzenstadler, R. Torkar, C. Martinez "Take a deep breath: Benefits of neuroplasticity practices for software developers and computer workers in a family of experiments"
Empirical Software Engineering Journal (EMSE), 2022. DOI: 10.1007/s10664-022-10148-z .

[D] C. Martinez Montes, B. Penzenstadler "Evaluating the Impact of a Yoga-Based Intervention on Software Engineers' Well-Being"
International Conference on Evaluation and Assessment in Software Engineering (EASE), 2025. DOI: 10.1145/3756681.3756950.

[E] C. Martinez Montes, D. Grassi, N. Novielli, B. Penzenstadler "A Multimodal Approach Combining Biometrics and Self-Report Instruments for Monitoring Stress in Programming: Methodological Insights"
Under submission to special issue on Empirical Software Engineering Journal (EMSE), 2025. https://doi.org/10.48550/arXiv.2507.02118.

[F] C. Martinez Montes, R. Feldt, C. Miguel Martos, S. Ouhbi, S. Premanandan, D. Graziotin "Large Language Models in Thematic Analysis: Prompt Engineering, Evaluation, and Guidelines for Qualitative Software Engineering Research"
Under submission to Transactions on Software Engineering (TSE), 2025.
https://doi.org/10.48550/arXiv.2510.18456.

Other publications

The below publications were written during my PhD as an exploration of tangential areas of my main research. Hence, they were not including in my PhD story line.

[a] C. Martinez Montes, J. Johansson, E. Dunvald "Factors Influencing Gender Representation in IT Faculty Programmes: Insights with a Focus on Software Engineering in a Nordic Context".
In Proceedings of the International Conference on the Foundations of Software Engineering (FSE Companion, Education Track), 2025, pp. 772–782. DOI: 10.1145/3696630.3727234.

[b] B. Penzenstadler, S. Motogna, P. Lago, C. Martinez Montes "Policy Making as Extension of Disseminating Research Results: Policy Influence Plan Canvas".
In B. Penzenstadler, K. Boudaoud, A. Di Marco & S. Caner-Yıldırım (Eds.), Actions for Gender Balance in Informatics Across Europe, 2025 (pp. 383–397). Springer, Cham. DOI: 10.1007/978-3-031-78432-3 16.

[c] S. Chand, C. Li, C. Martinez Montes, B. Cabrero-Daniel, J. Horkoff "Automating Requirements Review in the Automotive Sector: A Tailored AI Approach"
International Requirements Engineering Conference (RE), 2024 (Poster Track), (pp. 492-493). IEEE. DOI: 10.1109/RE59067.2024.00059.

[d] C. Martinez Montes, F. Sjögren, A. Klevfors, B. Penzenstadler "Qualifying and Quantifying the Benefits of Mindfulness Practices for IT Workers"
In 2024 10th International Conference on ICT for Sustainability (ICT4S) (pp. 272-281). IEEE. DOI: 10.1109/ICT4S64576.2024.00035.

[e] C. Martinez Montes, B. Penzenstadler "Piloting a well-being and resilience intervention in a course on digitalization for sustainability."
In The International Conference on Information and Communications Technology for Sustainability (ICT4S), Doctoral Symposium, Demos, Posters (Intl. Workshop on ICT4S Education), (pp. 105-118), 2023.

Research Contribution

In Paper A, my contributions were the conceptualisation, data curation and analysis, investigation, development of the research methodology, project administration, validation, visualisation, drafting, reviewing and editing the original manuscript.

In Paper B, I contributed to the conceptualisation, data curation and analysis, investigation, development of the research methodology, project administration, supervision, validation, visualisation, drafting, reviewing and editing the original manuscript.

For Paper C, I contributed to the formal analysis, data curation, validation and visualisation. I also participated in writing the original draft.

For Paper D, I was responsible for conceptualisation, data curation and analysis, investigation, development of the research methodology, project administration, validation, visualisation, drafting, reviewing and editing the original manuscript.

In Paper E, I was involved in the conceptualisation, data curation and analysis, investigation, development of the research methodology, project administration, validation, visualisation, drafting, reviewing and editing the original manuscript.

For Paper F, I participated in the conceptualisation, data curation and analysis, investigation, development of the research methodology, project administration, validation, visualisation, drafting, reviewing and editing the original manuscript.

In all the articles, my co-authors contributed in several roles, for example, Conceptualisation, Methodology, Validation, Supervision and Investigation. Moreover, particularly in Formal Analysis, papers involving qualitative data required my co-authors to be involved in the data analysis, thereby enhancing the reflexivity of the process. They also provided input reviewing, editing and refining the manuscript drafts.

Table presents my contributions to the appended papers following the CRediT (Contribution Roles Taxonomy) [1] criteria.

Role / Paper	A	B	C	D	E	F
Conceptualisation	✓	✓		✓	✓	✓
Methodology	✓	✓		✓	✓	✓
Software						
Validation	✓	✓	✓	✓	✓	✓
Formal analysis	✓	✓	✓	✓	✓	✓
Investigation	✓	✓		✓	✓	✓
Resources	✓	✓		✓	✓	✓
Data Curation	✓	✓	✓	✓	✓	✓
Writing – Original Draft	✓	✓	✓	✓	✓	✓
Writing – Review & Editing	✓	✓		✓	✓	✓
Visualization	✓	✓	✓	✓	✓	✓
Supervision		✓			✓	
Project administration	✓	✓			✓	
Funding acquisition						

Acknowledgment

A mi **F**amilia, and to all the minds that have accompanied, supervised, inspired, supported, listened, challenged, and motivated me, thank you all, especially those **W**ho were light when it was dark.

Contents

Abstract iii

List of Publications v

Other Publications vi

Research Contribution vii

Acknowledgement ix

1 Introduction 1

1.1 Research Focus . . . 3

1.2 Background . . . 6

1.2.1 Stress and the Nervous System . . . 6

1.2.2 Mindfulness-based practices . . . 8

1.2.2.1 Breathwork . . . 8

1.2.2.2 Yoga . . . 8

1.2.2.3 Journaling . . . 9

1.2.3 Emotions in SE . . . 9

1.2.4 Qualitative Data Analysis in SE . . . 10

1.3 Related Work . . . 10

1.3.1 Understanding Stress and Resilience in Software Engineers . . . 10

1.3.2 Support and Enhancement of Resilience and Well-Being in the Software Engineering . . . 11

1.3.3 Research on the Human Factors in Software Engineering . . . 12

1.4 Research Methodology . . . 13

1.4.1 Research Designs . . . 13

1.4.1.1 Quasi-experiments (interventions) . . . 13

1.4.1.2 Survey . . . 15

1.4.1.3 Experiment . . . 15

1.4.1.4 Exploratory Sequential Design . . . 15

1.4.2 Data Collection Methods . . . 16

1.4.2.1 Questionnaires . . . 16

1.4.2.2 Interviews and Focus Groups . . . 16

1.4.2.3 Biometric Data . . . 16

1.4.3 Data Analysis . . . 17

1.4.3.1 Quantitative Analysis Method . . . 17
1.4.3.2 Qualitative Analysis Method . . . 17
1.4.4 Reflexivity . . . 18
1.5 Ethical Considerations . . . 19
1.6 Research Results . . . 20
1.7 Discussion and Answers to the RQs . . . 25
1.7.1 Factors and conditions that influence stress, well-being, and resilience . . . 26
1.7.2 Approaches to foster sustained well-being . . . 30
1.7.3 Multimodal data triangulation and LLM-assisted analysis . . . 31
1.8 Limitations and Threats to Validity . . . 32
1.8.1 Scope of Applicability . . . 32
1.8.2 Internal Validity . . . 33
1.8.3 External Validity . . . 33
1.8.4 Construct Validity . . . 34
1.9 Conclusions . . . 34

2 Paper A: The Factors Influencing Well-Being in Software Engineers: A Mixed-Method Study 39
2.1 Introduction . . . 40
2.2 Background and Related Work . . . 41
2.2.1 Background . . . 41
2.2.1.1 The Conception of Well-being . . . 41
2.2.2 Related Work . . . 42
2.2.2.1 General Population . . . 42
2.2.2.2 Software Engineering Population . . . 43
2.3 Methodology . . . 44
2.3.1 Study design . . . 44
2.3.2 Population . . . 45
2.3.3 Data Collection . . . 45
2.3.3.1 Pilots of the data collection instruments . . . 45
2.3.3.2 Interviews . . . 45
2.3.3.3 Survey . . . 46
2.3.4 Data Analysis . . . 47
2.3.4.1 Interview Analysis . . . 47
2.3.4.2 Survey Analysis . . . 47
2.3.4.3 Reflexivity . . . 47
2.3.5 Ethical Considerations . . . 48
2.4 Results . . . 48
2.4.1 Interviews . . . 48
2.4.1.1 Theme 1: Individual Conception of Well-being . . . 48
2.4.1.2 Theme 2: Personal and Collaborative Factors . . . 50
2.4.1.3 Theme 3: Support and Recognition . . . 52
2.4.1.4 Theme 4: Work Environment and Culture . . . 55
2.4.1.5 Theme 5: Challenges and Stressors . . . 58
2.4.2 Survey Results . . . 60
2.4.2.1 Survey Respondent Demographics . . . 61

2.4.2.2 Associations Between Well-Being and Three-Level Variables . . . 61
2.4.2.3 Likert Scale Questions . . . 62
2.4.2.4 Open Questions . . . 63
2.5 Discussion . . . 69
2.5.1 Alignments in Quantitative and Qualitative Results . . . 70
2.5.1.1 Personal Practices . . . 70
2.5.1.2 Support from the Company and Peers . . . 71
2.5.1.3 Work Environment: Trust, Physical Well-being, and Compensation . . . 71
2.5.1.4 Equality, Equity, Diversity, and Inclusion (EEDI) . . . 71
2.5.1.5 Personal Life Situations . . . 72
2.5.1.6 Workload and Time Constraints . . . 72
2.5.1.7 Social Integration and Loneliness . . . 73
2.5.1.8 Tech Tools and Their Impact on Communication and Productivity . . . 73
2.5.2 Contrasting Quantitative and Qualitative Results . . . 73
2.5.2.1 Influence of Social Interactions on Well-being . . . 73
2.5.2.2 Recognition at Work . . . 74
2.5.2.3 Professional and Personal Growth Support from Companies . . . 74
2.5.2.4 Company Policies and Practices . . . 74
2.5.3 Comparison to Other Theories of Well-being Factors . . . 76
2.5.4 Policy Recommendations . . . 78
2.5.5 Validity Threats . . . 79
2.5.5.1 Internal validity . . . 79
2.5.5.2 External validity . . . 79
2.5.5.3 Construct validity . . . 80
2.5.6 Future Work . . . 80
2.6 Conclusion . . . 80

3 Paper B: Emotional Strain and Frustration in LLM Interactions in Software Engineering 83
3.1 Introduction . . . 84
3.2 Background and Related Work . . . 85
3.2.1 Large Language Models in Software Engineering . . . 85
3.2.2 Emotions Involved When Using Technology . . . 85
3.2.3 Emotions in Software Engineering Tasks . . . 87
3.3 Methodology . . . 88
3.3.1 Target Population . . . 89
3.3.2 Data Collection . . . 89
3.3.3 Data Analysis . . . 90
3.4 Results . . . 90
3.4.1 Respondents Demographics . . . 90
3.4.2 Usage of LLMs in Software Engineering Industry and Academia 90
3.4.3 Emotions During LLM Interaction . . . 91
3.4.4 Expectations When Interacting with LLMs . . . 93

3.4.5 Frustration Triggers in Software Engineering 95
3.4.6 Unmet Expectations' Impact on Motivation 97
3.4.7 User Actions for Improving LLM Interactions 98
3.5 Discussion 101
3.5.1 Frustrations in the Context of LLM Interaction 101
3.5.2 Impact of Frustrating Experiences on Motivation 102
3.5.3 Towards a Less-Frustrating User Experience 102
3.5.4 Threats to Validity 103
3.6 Conclusion 104

4 Paper C: Take a deep breath: Benefits of neuroplasticity practices for software developers and computer workers in a family of experiments 105
4.1 Introduction 106
4.2 Background 107
4.2.1 Context: Stress in Software Development and IT Work 107
4.2.2 Background: Breathing practices 108
4.2.3 Theory: Mindfulness and Mindfulness Attention Awareness . . 109
4.2.4 Related work: Well-being and Resilience for Engineers, Software Developers and IT Workers 110
4.3 Research Design 112
4.3.1 The intervention: Rise 2 Flow program 112
4.3.2 Research Questions 114
4.3.3 Population and Inclusion Criteria 115
4.3.4 Instrument Design 116
4.3.4.1 Entry/Exit Survey 116
4.3.4.2 Daily journal (for RQ1b) 119
4.3.4.3 Weekly survey: WHO-5 (for RQ2d) 119
4.4 Analysis Procedure 120
4.4.1 Demographics 120
4.4.2 Statistical Analysis of Instruments 121
4.4.2.1 The Data and Data Cleaning 122
4.4.2.2 Temporal Analysis of Entry vs. Exit 123
4.4.2.3 Temporal Analysis of Weekly and Daily Trends 124
4.4.3 Qualitative Analysis of Instruments: Thematic analysis 126
4.5 Results 126
4.5.1 Overall Engagement 127
4.5.2 RQ1: Changes in Mindfulness Attention Awareness and Daily Perceptions 130
4.5.2.1 Does the intervention bring about change in the participants Mindfulness Attention Awareness? (RQ1a) 131
4.5.2.2 How did the perceptions of life experience progress over time? (RQ1b) 135
4.5.3 RQ2: Changes in Well-being 137
4.5.3.1 Does the intervention lead to change in the participants' perceptions of positive and negative experiences? If so, how are their experiences affected? (RQ2a) 139

4.5.3.2 Does the intervention lead to change in the participants' psychological well-being? If so, how is it affected? (RQ2b) . . . 140
4.5.3.3 Does the intervention lead to change with regard to their positive thinking? If so, how is it affected? (RQ2c) 140
4.5.3.4 How does the well-being fluctuate and vary over the course of the intervention? (RQ2d) . . . 141
4.5.4 RQ3: Changes in Perceived Productivity and Self Efficacy . . . 142
4.5.4.1 Does the intervention lead to change in the participants perceived productivity? If so, how is it affected? (RQ3a) . . . 142
4.5.4.2 Does the intervention lead to change in the participants' self-efficacy? If so, how is it affected? (RQ3b) . 143
4.6 Discussion . . . 144
4.6.1 Significance . . . 144
4.6.2 Observations and Implications . . . 146
4.6.3 Limitations and Threats to Validity . . . 147
4.6.4 Relating Back to Theory in Psychology . . . 149
4.6.5 Future Work . . . 149
4.7 Conclusion . . . 150
4.8 Data Availability . . . 152
4.9 acknowledgements . . . 152

5 Paper D: Evaluating the Impact of a Yoga-Based Intervention on Software Engineers' Well-Being 153
5.1 Introduction . . . 154
5.2 Related Work . . . 155
5.2.1 The Effectiveness of Yoga in General . . . 155
5.2.2 The Effectiveness of Hatha Yoga . . . 156
5.2.3 Consequences for job performance and outcomes proposed . . . 156
5.2.4 Moderating Factors and Conditions for Mindfulness . . . 157
5.2.5 Well-being Interventions in SE . . . 157
5.3 Methodology . . . 157
5.3.1 Research Design . . . 157
5.3.2 Intervention . . . 158
5.3.3 Company, Population and Inclusion Criteria . . . 158
5.3.4 Data Collection . . . 159
5.3.4.1 Entry and Exit Survey . . . 159
5.3.4.2 Weekly Well-being Scale . . . 160
5.3.4.3 Focus Group . . . 160
5.3.5 Data Analysis . . . 160
5.3.5.1 Statistical Analysis of Instruments . . . 161
5.3.5.2 Qualitative Analysis of Focus Group . . . 161
5.3.5.3 Reflexivity . . . 162
5.3.6 Ethical Considerations . . . 162
5.4 Results . . . 162
5.4.1 Quantitative Analysis . . . 163

5.4.2 Thematic Analysis 164
5.4.2.1 Theme 1: Individual Benefits and Shared Reflections in Practice 165
5.4.2.2 Theme 2: Organisational Support and Logistical Challenges in Implementing the Programme 167
5.4.2.3 Theme 3: Perception and General Feedback 168
5.5 Discussion 168
5.5.1 Results in Context 169
5.5.2 Importance of the Study 170
5.5.3 Lessons Learned 171
5.5.4 Validity Threats 172
5.5.4.1 Internal Validity 172
5.5.4.2 External Validity 172
5.5.4.3 Construct Validity 172
5.5.4.4 Conclusion Validity 172
5.6 Conclusion 173
5.7 Acknowledgement 173

6 Paper E: A Multimodal Approach Combining Biometrics and Self-Report Instruments for Monitoring Stress in Programming: Methodological Insights 175
6.1 Introduction 176
6.2 Background and Related Work 177
6.2.1 Physiological Measures of Stress and Mental Load 178
6.2.2 Using Biometrics for Studying Cognitive and Affective States in Software Development 180
6.3 Methodology 181
6.3.1 Participants and Recruitment Strategy 182
6.3.2 Experiment Setup 182
6.3.3 Data Collection Methods 183
6.3.3.1 Biometric Sensors 183
6.3.3.2 Self-report Instruments 184
6.3.3.3 Post-Task Interview 184
6.3.4 Data Analysis 184
6.3.4.1 Psychometrics 185
6.3.4.2 Biometrics 185
6.3.4.3 Interviews 187
6.4 Results 187
6.4.1 Psychometrics 188
6.4.1.1 Closer Look per Participant 189
6.4.2 Biometrics 190
6.4.3 Thematic Analysis 192
6.4.3.1 Theme 1: Task Impressions: Engagement and Learning 192
6.4.3.2 Theme 2: Emotional Responses to Challenge and Uncertainty 193
6.4.3.3 Theme 3: Sources of Distraction and Discomfort 194
6.4.3.4 Theme 4: Coping Strategies and Adaptation 194

6.5 Discussion . . . 195
6.5.1 Main Contributions . . . 195
6.5.1.1 Evidence Supporting the Alignment of Biometrics and Psychometric Instruments . . . 195
6.5.1.2 Methodological Insights on Recruitment and Experimental Design . . . 200
6.5.1.3 General Lessons for Conducting Stress Studies . . . 201
6.5.1.4 Ethical Reflections on Stress Induction in Research . 202
6.5.2 Threats to Validity . . . 203
6.5.2.1 Internal Validity . . . 203
6.5.2.2 External Validity . . . 203
6.5.2.3 Construct Validity . . . 203
6.5.2.4 Ecological Validity . . . 204
6.6 Conclusion . . . 204
6.6.1 Summary . . . 204
6.6.2 Future Work . . . 204
6.7 Declarations . . . 205
6.7.1 Funding: . . . 205
6.7.2 Ethical approval: . . . 205
6.7.3 Informed consent: . . . 205
6.7.4 Author Contributions [all authors should be mentioned] . . . 205
6.7.5 Data Availability Statement . . . 205
6.7.6 Conflict of Interest . . . 206
6.7.7 Clinical Trial Number in the manuscript. . . . 206

7 Paper F: Large Language Models in Thematic Analysis: Prompt Engineering, Evaluation, and Guidelines for Qualitative Software Engineering Research 207
7.1 Introduction . . . 208
7.2 Background and Related Work . . . 209
7.2.1 Qualitative Data Analysis in Software Engineering . . . 209
7.2.1.1 Thematic analysis . . . 209
7.2.2 Early AI and ML in Qualitative Data Analysis . . . 210
7.2.3 LLMs in Qualitative Analysis . . . 211
7.2.4 Generative AI Tools and Frameworks in QDA . . . 211
7.3 Methodology . . . 213
7.3.1 Dataset . . . 213
7.3.2 Study Design: Mapping TA Phases to Human vs LLM Roles . 213
7.3.3 Prompting Strategy, Application and Evaluation . . . 216
7.3.3.1 Prompting Strategy . . . 218
7.3.3.2 Initial Coding (Phase 2) . . . 219
7.3.3.3 Generating, Reviewing, Refining and Naming Themes (Phases 3-5) . . . 220
7.3.4 Data analysis . . . 221
7.3.5 Evaluators Team . . . 221
7.3.6 Ethical Considerations . . . 222
7.4 Results . . . 222

7.4.1 Evaluation Phase 2: Human vs LLM (1a) 222
7.4.2 Evaluation Phase 2: LLM Codes Rubric-Based Evaluation (1b) 226
7.4.3 Evaluation Phase 3-5 Generating, Refining and Naming Themes (2a) 228
7.5 Discussion 229
7.5.1 LLMs as Analytical Assistants in SE Research 229
7.5.1.1 Human Oversight Remains Essential 231
7.5.2 Strengths and Limitations of LLM Outputs in Engineering Contexts 233
7.5.3 Strategies to Ensure Methodological Rigour 234
7.5.4 Implications for Empirical SE Methods 234
7.5.5 Threats to Validity 235
7.6 Conclusion 236
7.7 Authors' Contributions 237

Bibliography **239**

Appendix **277**

A Appendix - Paper A **277**
A.1 Data Collection Instruments 277
A.2 Survey Instruments 277
A.2.1 The MAAS instrument 277
A.2.2 The instruments SPANE, PWB, and PTS 277
A.2.3 Self Efficacy 279
A.2.4 Perceived Productivity 280
A.2.5 The WHO-5 instrument 282
A.3 Model designs 283
A.3.1 Gaussian Process model 283
A.3.2 Dummy variable regression model 284
A.4 Detailed Findings: Significant Effects of Other Predictors 285
A.4.1 Mindfulness Attention Awareness Scale 285
A.4.2 Scale of Positive And Negative Experiences 286
A.4.3 Psychological Well-Being 287
A.4.4 Positive Thinking Scale 288
A.4.5 Self Efficacy 289
A.4.6 Perceived Productivity 290
A.4.7 Predictor Number of Sessions 290

Chapter 1

Introduction

"*We cannot change the human condition, but we can change the conditions under which people work*" - James Reason.

Over the past two decades, human factors in software engineering (SE) have evolved from a peripheral concern to a central research theme [2–4]. More recently, the focus has been on psychological factors (a core part of the human condition) [5], including well-being.

Research from occupational and cognitive psychology has long established that well-being, stress regulation and resilience influence performance, motivation, and long-term health [6–10]. Prolonged or poorly regulated stress can erode well-being, while resilience processes help maintain functioning under pressure and support motivation and long-term health. In software engineering (SE), these principles require adaptation to the specific characteristics of engineers' work. SE poses different challenges compared to other fields [5] as it encompasses cognitive, emotional, and social aspects. For example, engineers' activities require a combination of creativity and autonomy [11], high cognitive load [12], problem solving [13], group dynamics [14], long periods of focused attention [15], and collaborative workflows [16]. Moreover, engineers engage in sustained problem-solving under conditions of uncertainty, time pressure, and rapid technological change [17, 18]. **These characteristics make software engineers particularly vulnerable to stress, emotional exhaustion, and burnout** [19]. Additionally, this specific combination of factors makes work conditions in SE distinct from those in general occupational contexts, creating a need to adapt psychological models and measures to this high-risk population.

Furthermore, the recent and rapid adoption of artificial intelligence (AI) tools, particularly large language models (LLMs), presents new concerns for well-being. These tools change several engineers' tasks. For example, how they search for information, design solutions, and code and debug [20]. Its use potentially reduces some forms of effort. However, it also introduces new challenges, such as overreliance on generated outputs, the need to be vigilant in detecting subtle errors, and the requirement to learn new workflows. Understanding how these technologies influence cognitive load, affect, and well-being is important for the future of sustainable software development.

At the same time, practical change requires going beyond identifying and measuring these factors. Improving well-being in practice requires developing and implementing interventions that address them. Promoting well-being is not merely an ethical imperative but a condition for sustainable development, retention, and innovation [21]. Nevertheless, there is limited empirical evidence on which approaches can be effectively developed, evaluated, and sustained to foster well-being in software engineering. Such approaches can only be effective if they are grounded in a precise understanding of the problem they seek to address and are evaluated using appropriate empirical methods.

Despite this growing recognition and research, **empirical evidence on the factors influencing stress, well-being, and resilience in software engineers remained fragmented**. This fragmentation limits theoretical integration and constrains the design of effective interventions. Existing studies have predominantly focused on single constructs such as happiness [8], sentiments and emotions [22], motivation [23] burnout [24] or productivity [25]. However, an integrated view of these factors is essential, since the interaction between them influences how software engineers experience and manage their work-related challenges.

Additionally, **current study methods often rely on a single data collection point**, typically surveys [3], and rarely integrate multiple data sources into a single analysis. This limits the comprehensiveness of the problem. Quantitative surveys capture correlations but lack context, while qualitative studies offer depth but are challenging to replicate and scale [26]. Some studies have started to apply data triangulation by incorporating biometric data. However, methodological standards for such multimodal research are still in development.

Similarly, **interventions to enhance well-being (e.g., mindfulness-based or resilience-training programs) are rare and typically limited to short-term pilots** with small sample sizes [27]. More importantly, improving well-being in software engineering requires approaches that go beyond isolated individual practices. Moreover, prior research suggests that supporting well-being in knowledge-intensive work requires broader approaches that consider organisational conditions [9]. However, there remains a limited empirical understanding of how such multi-level approaches (spanning individual and organisational factors) can be designed, combined, and evaluated within SE contexts.

To effectively address these limitations, a qualitative approach is necessary. Consequently, this raises another question: how to rigorously and consistently analyse rich qualitative data as studies grow in scale and complexity? Recently, researchers have been exploring whether LLMs can support qualitative analysis (or parts of it). For example, to generate deductive codes [28] or create themes [29]. However, as this line of inquiry is still emerging, its methodological foundations remain unsettled. The implementation of LLMs to assist qualitative analysis presents opportunities and epistemological risks. Partial automation can augment analytical rigour, but it also threatens interpretive validity if used uncritically. Hence, clear methodological strategies are needed to integrate LLM-assisted analysis without compromising rigour and transparency.

This thesis addresses these gaps by studying the well-being of software engineers through an integrated, empirical, and reflexive research agenda. It combines psychological theory, human factors research, and software engineering methodologies.

1.1 Research Focus

The thesis follows a progression argument, moving from explanation to action and, ultimately, to methodological contributions. Hence:

In-depth problem analysis → Approaches (actions) targeting the problem → Methodological proposals to study the problem

The first step was to explore and **study the problem (RQ1)**, so we developed an empirical framework to explain the factors that influence well-being in software engineers. Then, with a more precise understanding of the problem, we could suggest actions. In this thesis, those **actions (RQ2)** were translated into interventions, policy recommendations, organisational guidelines, and design proposals for chatbots and LLMs. It is important to note that the interventions served beyond just testing tailored programmes for stress management. They also informed the need to strengthen and improve measurement, completeness, and rigour in qualitative research, specifically in human factors. Finally, we addressed and proposed **strategies for studying human factors in SE (RQ3)**. We suggested and tested the inclusion of biometric data in mixed-methods studies and proposed a hybrid framework to integrate LLMs as research assistants.

The work is organised around three overarching Research Questions(RQs):

RQ1.What are the factors and conditions that influence stress, well-being, and resilience in engineers?

This question aims to identify the main contributors and hindrances to stress, well-being, and resilience. It attempts to directly address the current trend of treating each factor in isolation. It is driven to detect and integrate these multiple factors, explaining their interaction at different levels (individual, team, and organisational) and their influence on the well-being of software engineers. As stated in the introduction, much of the current research studies general psychosocial factors without distinguishing between fields or professions. Given the specific cognitive, social, and organisational characteristics of software engineering activities, this thesis aims to explore and compare these factors in a context-sensitive and empirically grounded manner. One of the goals is to develop a framework grounded in the characteristics, practices, and environments of the software engineering population. The resulting framework is intended to provide a structured view of existing and newly identified factors. Our vision is for the framework to guide future empirical research and to inform the design of interventions, educational practices, and organisational policies that are better aligned with the realities of software engineering work and learning contexts.

RQ2.What approaches can be developed and evaluated to foster sustained well-being among software engineers?

The goal of this question is to propose informed mindfulness interventions based on the data and results from the previous question. We aimed to develop approaches grounded in the specific stressors, work practices, and contextual constraints identified within SE settings. Mindfulness-based interventions are selected as the primary focus because they directly target attentional regulation, cognitive reactivity, and emotional awareness. These aspects are particularly relevant to SE work, which is characterised by prolonged cognitive effort, high mental load, frequent interruptions, and persistent problem-solving demands. Moreover, mindfulness interventions can be applied at the individual and team level and are adaptable to diverse work contexts, including remote and time-constrained environments. Additionally, these types of interventions are a low-cost and scalable alternative that can be integrated into existing practices with minimal disruption.

Beyond the design, this question also addresses how these interventions can be more effectively measured and evaluated, as well as the challenges that may hinder their implementation and success. With RQ2, we want to bridge the gap between theory and well-being interventions applicable in real-world SE environments.

RQ3. How can multimodal data triangulation and LLM-assisted analysis be used to develop rigorous methodological strategies for studying human factors in software engineering?

This question examines how triangulating multiple types of data (psychometric instruments, interviews and physiological measures) can strengthen the study of human factors in SE. It focuses on improving validity, depth, and interpretive robustness. Each data modality captures different aspects of stress, well-being, and cognitive experience, and their combined use allows the identification of convergent, complementary, or conflicting evidence that would be hard to observe through single-method approaches.

It also investigates how LLMs can be integrated as analytical assistants to support qualitative data analysis while maintaining methodological rigour, transparency, and ethical safeguards. We want to assess if LLMs can augment human analysis while preserving the subjective and reflexive nature of qualitative data analysis. This question seeks to examine strategies for maintaining transparency, traceability of analytical decisions, and consistency with established qualitative methods. One of the goals is to critically evaluate the benefits and limitations of multimodal triangulation and LLM integration. This RQ aims to develop practical methodological strategies that are empirically robust and suitable for human factors research in SE.

To answer these questions, six empirical studies (Papers A–F) were conducted:

- Paper A (Well-Being Factors): Mixed-method exploration of the determinants of software engineers' well-being.
- Paper B (Emotional Strain by AI): Investigation of emotional strain in human–LLM interaction.
- Papers C (Breathwork Intervention) and D (Yoga Intervention): Quasi-experimental mindfulness interventions using breathwork, yoga, and journaling.
- Paper E (Multimodal Methodology): Multimodal stress study combining biometric, self-report measures and interviews.
- Paper F (AI 4 Thematic Analysis): Advancing human–AI collaboration in qualitative data analysis.

These studies propose a multi-level (considering individual, team and organisation) and multi-modal (integrating different data sources) perspective on well-being in SE. Figure 1.1 shows how the papers group to answer the previous RQs. Each paper is presented with an icon with its main contribution written below.

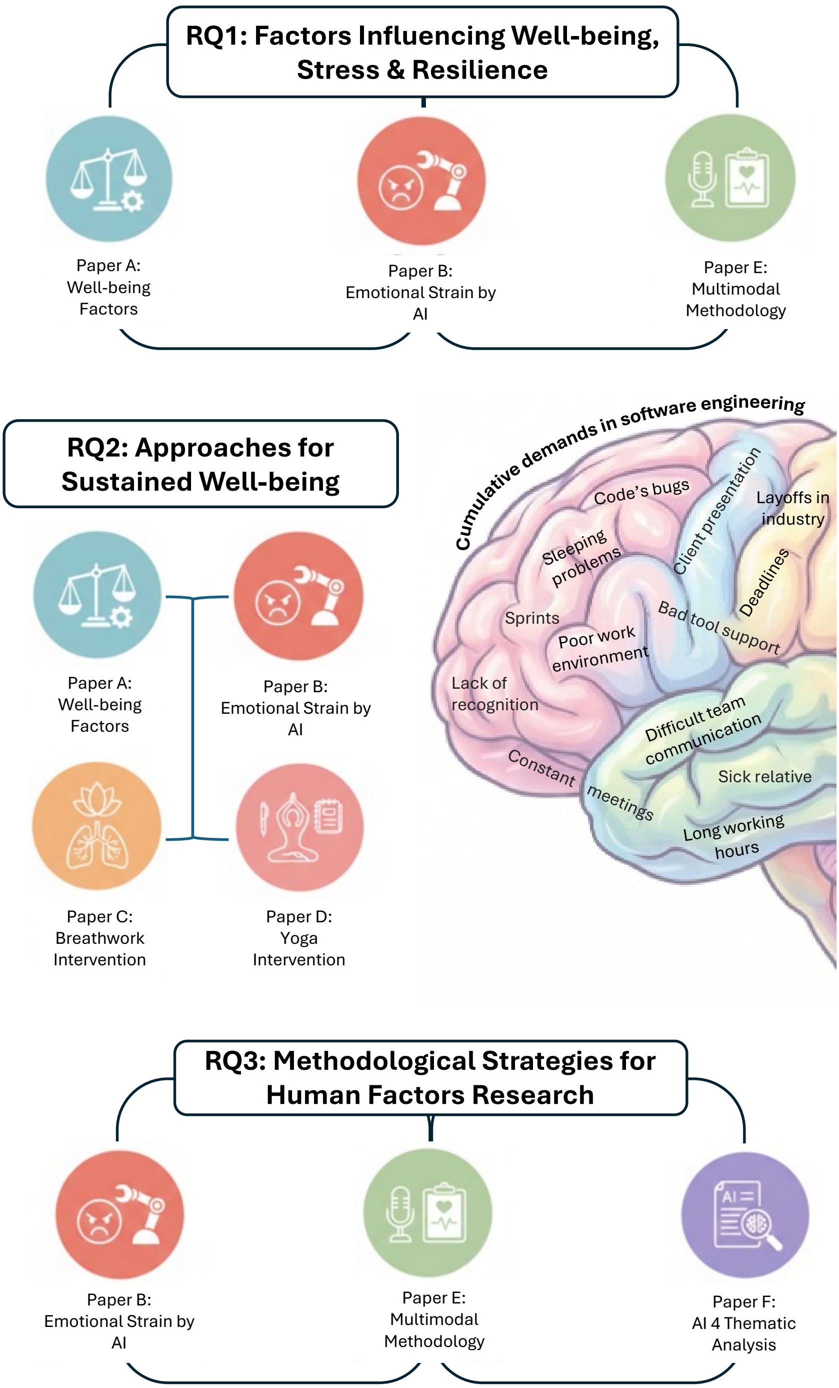

Figure 1.1: Thesis overview. It shows how the papers answer each RQ. The brain in the centre simulates what triggers stress in software engineers.

1.2 Background

This section provides definitions of the main concepts used in the thesis, an introduction to the biology involved in stress, and a brief overview of mindfulness practices.

Table 1.1 presents the definitions based on the American Psychological Association (APA) [30].

Concept	**Definition**
Awareness	- Being consciously able to notice, recognise, or understand something, and being able to describe it accurately. It refers to the state of being conscious of what is happening around or within oneself.
Mindfulness	- A state of enhanced awareness of the present moment, including one's sensations, thoughts, bodily states, consciousness, and environment, while fostering an attitude of acceptance without judgment or reaction.
Neuroplasticity	- The ability of the nervous system to change its structure or function in response to experience or environmental stimulation.
Resilience	- The process and outcome of successfully adapting to difficult or challenging life experiences, primarily through mental, emotional, and behavioural flexibility and adjustment to external and internal demands.
Stress	- The body's response to events that demand an individual to adjust or employ coping strategies. These events can arise from external situations or internal factors. It influences almost every system in the body, shaping how people behave and feel.
Well-being	- A state of happiness and contentment, with low levels of distress, overall good physical and mental health and outlook, or good quality of life

Table 1.1: Summary of the most important terms used in the thesis.

1.2.1 Stress and the Nervous System

This subsection provides a brief explanation of the biological aspects of stress in the human body [31].

Stress is a biological process that prepares the body to respond to demands. When the brain perceives a situation as challenging or threatening, either physical (e.g., danger, injury) or psychological (e.g., work pressure, fear), the amygdala detects it and sends a distress signal to the hypothalamus. The hypothalamus acts as a control centre that activates the sympathetic nervous system (SNS). Then, the SNS triggers the adrenal medulla (a part of the adrenal glands located on top of the kidneys) to release adrenaline (epinephrine) into the bloodstream. When this happens, several physiological changes occur in the body, including increased heart rate, elevated blood pressure, faster breathing, redirection of blood flow to the muscles, and heightened alertness. If the challenge continues, the hypothalamic–pituitary–adrenal (HPA) axis releases cortisol, which helps the body sustain attention, mobilise energy, and remain prepared.

Under short-term conditions, the stress response is adaptive. The parasympathetic nervous system (PNS), largely via the vagus nerve, restores the body once the demand has passed by reducing heart rate, supporting digestion, and lowering stress-related neuroendocrine activity [32].

However, modern environments create a different kind of demand. In knowledge-intensive work, such as in SE, stress comes from continuous cognitive load, task switching, ambiguity in roles, interruptions, overload of work and prolonged mental effort [33, 34]. These situations require the brain to regulate attention, energy, and physiological resources repeatedly. This ongoing regulatory activity is described as allostasis: the process of adjusting bodily systems to meet current and anticipated demands [35–37].

When these adjustments are triggered too often, or recovery is insufficient, the body accumulates allostatic load, a measurable strain on physiological systems [36]. Elevated cortisol levels over time can weaken immune function, impair memory and decision-making, and increase inflammation [38, 39]. Repeated sympathetic activation can contribute to hypertension, sleep difficulties, and anxiety, while excessive epinephrine exposure can strain the cardiovascular system [38, 39]. Brain areas involved in focus, planning, and emotional regulation (such as the prefrontal cortex and hippocampus) also become less efficient under chronic load [40].

Since recovery depends heavily on parasympathetic activity, the vagus nerve is essential in helping the body return to baseline. Higher vagal tone is associated with quicker physiological recovery, improved emotional regulation, and better stress resilience [32].

In this thesis, the implemented interventions targeted the vagus nerve.

Technostress

Technostress is defined as a particular type of stress caused by the use of information and communication technologies (ICT) [41]. Software engineers are particularly prone to face technostress due to their close daily interaction with technology [42]. This interaction is defined by continuous development, use, and modification of complex, rapidly evolving technological systems [19, 43]. While technostress affects many technology-intensive professions, software engineering is characterised by persistent cognitive demands, frequent interruptions, tight deadlines, and ongoing pressure to adapt to new tools, frameworks, and programming paradigms [44, 45]. These characteristics make software engineers especially susceptible to technostress and its specific dimensions, including techno-overload, techno-complexity, and techno-uncertainty. These dimensions have been associated with increased psychological tension and emotional exhaustion [41]. Hence, besides the average stressors of daily life, there are more and closer risk factors for the engineering population.

Moreover, the increased use of AI in SE tasks adds to the current equation. Only a few works have specifically examined the negative impact of AI tools. However, existing studies report that constant use can cause emotional dysregulation and social withdrawal [46, 47]. Since AI use also produces positive effects, such as increased productivity, it becomes challenging for users to notice when the advantages shift into pressure or emotional strain. This makes the risks of technostress harder to detect in daily work.

1.2.2 Mindfulness-based practices

Mindfulness-based practices are activities or interventions where the individual intentionally observes their own body and mind in the present moment [48, 49]. It implies an attitude of openness, curiosity, acceptance, and non-judgment.

Practising mindfulness has been shown to have a positive effect on emotionally stressful situations and to increase immune system response [50–52]. For this thesis, we selected specific practices that could be adapted to the formats and organisations in which they were implemented. I elaborate on each practice, breathwork, yoga and journaling, in the following sub-sections.

1.2.2.1 Breathwork

Breathwork is the practice of regulating the way one breathes [53]. It encompasses specific breathing rhythms and patterns focused on promoting mental, emotional, and physical well-being [54]. By adjusting their breathing patterns, individuals can quickly influence the interaction between the respiratory system and the brain regions that regulate behaviour, thinking, and emotions [55]. Moreover, controlled breathing rhythms can promote harmony in brainwave activity. Intentionally slowing the breath aligns with brain electrical patterns, enhancing communication across different brain regions [53].

Research in psychiatry has found that breathwork can improve obsessions, depression, anxiety, trauma, inattention and compulsions [56]. Additionally, specific breathing exercises have proven to help reduce stress levels [57].

Breathwork has historical and traditional roots, including Tibetan Buddhism, yoga, and ancient practices such as Pranayama. Over time, numerous contemporary methods and exercises have been developed, yet the traditional and modern approaches remain widely practised today. As it has gained attention, several studies have examined its effects in different populations. In this thesis, the focus was on IT workers.

1.2.2.2 Yoga

The word "Yoga" has its origin in the Sanskrit root yuj, which means "to join" or "to unite". This refers to the union of mind and body. It originated in ancient India, seeking to achieve the union of the individual self and the transcendental self [58].

The Western adaptation and manifestation include physical exercises and postures combined with the regulation of respiration and meditation [59]. The focus is mostly on isometric exercise and stretching.

Several studies have demonstrated the positive effects of yoga on overall health, particularly on stress regulation. For example, a study by Ross [60] comparing the benefits of exercise and yoga found that, in healthy and diseased populations, yoga was more effective than exercise in improving various health-related measures. Furthermore, some studies have gone beyond and researched its effects on specific brain regions. For instance, Li et al. [61] found that in different phases of practising yoga, long-term yoga practitioners showed higher blood oxygen levels in the dorsolateral prefrontal cortex compared with short-term practitioners. Participants also reported better task mastery and showed lower ventrolateral prefrontal activation. Finally, they exhibited higher blood oxygen levels in the orbitofrontal and ventrolateral prefrontal regions compared to short-term practitioners.

This lays a more scientific ground for the benefits of practising yoga.

1.2.2.3 Journaling

Journaling is the practice of writing down thoughts, feelings, experiences, or reflections. When done reflectively, it encourages introspection and emotional discharge. Studies have found that it increases self-awareness, self-exploration and release of pent-up emotions [62]. It also improved individuals' physical and emotional well-being [63].

Journaling is commonly used in nursing studies to enhance reflective practice and as a technique for active learning. Among the reported benefits are the development of critical thinking, making connections between experiences, and discovering meaning [64]. Researchers have adapted journaling to digital formats, for instance, e-journaling. This format has the same effects as traditional journaling [65]. King and LaRocco [65] implemented e-journaling, achieving positive results with their students and demonstrating that the transition from paper to electronic does not hinder outcomes. In this thesis, we used e-journaling, as it was implemented in an online course; therefore, it was not possible to gather paper journals from around the world.

1.2.3 Emotions in SE

In this thesis, we used the American Psychological Association's (APA) definition of emotion [66]. Although numerous theoretical perspectives define emotions in different ways, this formulation was chosen because it explicitly integrates these experiential, behavioural, and bodily components into a single, coherent construct. The APA describes emotion as *"a complex reaction pattern, involving experiential, behavioural, and physiological elements, by which an individual attempts to deal with a personally significant matter or event"*.

Software engineering is an intensely human, cognitive, and social activity; hence, emotions are present in it. Engineers continuously engage in problem solving, learning, communication, and coordination, all of which are influenced by emotional states. At the individual level, emotions have been linked to other constructs such as productivity [67], personality [68], and collaboration [69], showing that emotions permeate to team and organisational levels. SE relies heavily on collaboration through activities such as code reviews, meetings, and issue discussions, where emotional expressions influence trust, conflict, and decision-making. Hence, understanding emotions in this specific context contributes to healthier teams, more sustainable work practices, and improved engineering outcomes.

With the previous in mind, we incorporated existing emotion frameworks into SE tasks. For example, for Paper B specifically, we used the Emotions Wheel created by Gloria Willcox [70] to guide the participants' emotions classification. This tool, frequently employed in therapeutic and self-reflective contexts, assists individuals in naming and differentiating their feelings with greater specificity. The wheel is organised in concentric layers: at its centre lie six foundational affective categories (happy, sad, angry, scared, strong, and calm). Progressing outward, each of these core categories branches into increasingly fine-grained descriptors, offering a structured way to capture the nuance and complexity of emotional experience.

1.2.4 Qualitative Data Analysis in SE

Qualitative data is crucial in SE for examining the human and organisational dimensions of engineering work [71]. Through the close interpretation of rich, non-numerical evidence, researchers can uncover themes, explanations, and relationships that are not accessible through quantitative methods [71, 72]. SE qualitative datasets frequently combine technical records with materials centred on human experience. For instance, code review discussions, architecture decision logs, or incident communication channels alongside interviews or observational notes. Working with these types of sources explains the need for careful handling of dataset size, consistent analytic reasoning across researchers, and documented chains of interpretation to ensure the trustworthiness of the findings.

In SE studies, thematic analysis is one of the most popular methods to analyse qualitative data [73, 74] (see steps in methodology section 1.4.3.2). We chose the Braun and Clarke guidelines [75] for their clear and simple steps to develop a hybrid framework in Paper F. The framework proposes a collaboration between human researchers and Large Language Models (LLMs) to automate steps 2 to 5 of the analysis process.

1.3 Related Work

This section provides an overview of related work as a big scope og the main problem. The detailed related work is discussed in each paper separately.

1.3.1 Understanding Stress and Resilience in Software Engineers

In the general literature, primarily from psychology, various frameworks aim to explain the study, development, and measurement of well-being and resilience. Examples of these well-being frameworks are: Gallup's Five Elements of Well-being [76], Seligman's Five Pillars of Well-being [77] and Michaelson's Five components [9]. However, these frameworks are not population-specific. Software engineers, like many other subgroups, have specific characteristics that influence how well-being and resilience are experienced and maintained in their context. They also have specific stressors that need to be taken into account. Factors such as high cognitive demands, frequent task switching, collaborative yet often distributed work environments, and rapid technological change shape their stressors and coping mechanisms in unique ways. Consequently, applying general psychological frameworks without adaptation may overlook critical occupational, social, and organisational dynamics.

Previous studies, SE population-specific, have focused on specific areas. For example, emotions of: (un)happiness [8], feeling overwhelmed [78], frustration [79], reasons for negative emotions in agile contexts [80] and emotions recognition in software development [81]. Productivity, such as successful environments on software teams [82], satisfaction and perceived productivity [83]. Additionally, there are other sub-areas, for example, inclusivity, empathy, and supportive work environments [84,85], everyday interpersonal challenges [86], and burnout [24]. However, there were no integrative frameworks that consider the interaction of these aspects when this thesis project began; currently, there are two besides our proposal, which are discussed next.

A comprehensive understanding of the well-being and resilience of software engineers, as well as their main stressors, requires moving beyond isolated constructs. It is essential to study different dimensions of interaction and include specific contexts. Wong et al. [87] proposed one of the first integration models explaining mental well-being, considering different levels (individual, team, and organisation) of interaction. Nevertheless, this framework is US-focused, which limits generalisability and fails to integrate different cultural contexts. In one step forward, Godliauskas and Šmite [88] conducted a literature review analysing 44 studies that reached populations from 42 countries. They proposed a theory about the Predictors and outcomes of software engineers' well-being. While this represents a significant step towards a multidimensional conceptualisation, the study was based solely on secondary data, relying predominantly on quantitative, cross-sectional surveys. As acknowledged by the authors, this limits causal inference and overlooks the nuanced, context-dependent experiences of software engineers. The analysis has limitations in capturing the lived experiences, interpretations, or mechanisms underlying the relationships identified.

Understanding well-being is a complex task that requires context-sensitive and empirically grounded insights. To address the previous limitations, this project's thesis conducted a mixed-methods study (Paper A) combining interviews and surveys. It also included participants from different contexts and countries. The data was triangulated to obtain statistical generalisability and in-depth contextual understanding. Moreover, it was then complemented and updated with a study (Paper B) focusing on the use of AI in daily SE tasks [89].

Addressing this gap enables the development of context-sensitive models and interventions that more accurately reflect the lived experiences of software engineers and support their sustainable well-being in the workplace.

1.3.2 Support and Enhancement of Resilience and Well-Being in the Software Engineering

A limited number of interventions within software engineering have explored well-being enhancement through mindfulness-based practices, yet the available studies consistently report positive effects. For instance, Heijer et al. [90] examined mindfulness in agile software teams through a two-month intervention involving short, three-minute mindfulness exercises during daily stand-up meetings. Conducted across eight companies and involving more than sixty participants, the study reported enhanced perceived effectiveness, decision-making, and listening skills. One limitation noted by the authors, however, was the use of non-standardised questionnaires. Research by Bernardez et al. [91, 92] has found that mindfulness interventions positively influence the mental well-being and self-perception of software engineers. Complementary to these findings, Romano et al. [93] investigated the effects of an eight-week Mindfulness-Based Stress Reduction (MBSR) programme among software developers from a multinational company in Italy. Participants who underwent MBSR reported reduced stress and improved focus.

Collectively, these studies suggest that mindfulness-based interventions hold considerable potential for enhancing well-being and focus within software development contexts. Nevertheless, no more interventions have been done in SE contexts. One important characteristic of the previous interventions is the limited number of participants. In the appended studies of this thesis, we discussed the same challenges. Participation rates tended to decline over time, which, although common in longitudinal or behavioural interventions, still poses difficulties in maintaining engagement and ensuring sufficient statistical power. This attrition can influence the robustness and generalisability of findings, as participants who remain engaged may differ systematically from those who discontinue. To address the challenges, it is essential to design interventions that are flexible, minimally intrusive, and better integrated into existing work routines. These aspects, among others, are discussed in our yoga intervention (Paper D) study [93].

1.3.3 Research on the Human Factors in Software Engineering

Research on human factors in SE focuses on understanding how cognitive, affective, social, and organisational factors influence software development activities. Prior work has studied engineers' motivation [94, 95], personality [96, 97], and job satisfaction [83, 98]; cognitive load and program comprehension [99, 100]; collaboration and communication in teams [101, 102]; decision-making [103] and expertise development [104], creativity [105] among others.

Recently, researchers have started exploring stress, burnout, well-being [19, 27, 106, 107] and empathy [108]. These factors have been shown to influence individual productivity, software quality and long-term sustainability of development teams.

Studies in this area have increasingly adopted mixed-methods approaches to gain a deeper and more complete understanding of the subject.

Quantitative studies, such as large-scale surveys, have been widely used to identify relationships between psychological factors and productivity or well-being [8, 83]. However, these studies often provide limited contextual understanding. To address this, several authors have integrated qualitative methods, for instance, interviews and observations, to interpret developers' subjective experiences.

A growing body of work also incorporates physiological and behavioural measures to complement self-reported and qualitative data. Studies have combined electroencephalography (EEG), electrodermal activity (EDA), and heart-rate monitoring with surveys and interviews to investigate cognitive load, affect, and engagement during programming tasks [109–111].

The triangulation of data produces more comprehensive and ecologically valid insights into developers' experiences and perceptions. A central part of this triangulation is the analysis of qualitative data (QD). QD enables an in-depth examination of the non-technical dimensions of software development [71]. Through systematic interpretation of rich, non-numerical data, researchers uncover patterns, meanings, and insights [71, 72]. QD is especially valuable for understanding software processes, tool adoption, and organisational or technical contexts.

Given the potential of Large Language Models (LLMs) to process substantial amounts of textual data, several authors have explored their use for the analysis of QD [28, 29, 112–114]. However, these studies face several challenges, including limited transparency and explainability, a lack of systematic evaluation on SE data, insufficient methodological rigour, and a narrow model scope. To address these limitations, this thesis proposes a human–LLM collaborative framework for thematic analysis (TA). This is the first study (Paper F) to incorporate tailored rubrics for evaluating the quality of codes and themes and to compare LLM-generated results with human-coded themes.

Thus, our contribution is to consolidate a triangulated approach that combines surveys, behavioural/physiological signals, and qualitative data (Paper E). Additionally, to extend the triangulation with a transparent implementation of LLMs as assistants in qualitative data analysis.

1.4 Research Methodology

This section explains the research methodology used in the studies presented in this thesis. Each study employed a mixed-methods approach to obtain a comprehensive, coherent picture of the researched phenomena.

Mixed methods research involves collecting quantitative and qualitative data and integrating them to get a comprehensive analysis of the research problem [115–117]. Scholars like Cresswell [116] claim that all methods have biases and weaknesses, and by combining more than one method, these biases can be neutralised. In this thesis, we applied a **triangulation strategy**, where numerical data from surveys and biometric data are enriched by the lived experiences captured in participant journals and interviews. Table 1.2 shows how each paper is linked to the thesis research questions, as well as the design and study methods. Each design, data collection and analysis method is explained next.

1.4.1 Research Designs

The research design serves as a broad structure or strategic framework that connects the research questions to the empirical data collection and analysis. It is based on the kind of explanation the researcher wants to deliver from the study [117].

1.4.1.1 Quasi-experiments (interventions)

Quasi-experiments are a type of experimental design used to investigate whether a direct causal link can be established between the independent variable (in this thesis context, an intervention) and the dependent variable [116, 118]. Quasi-experiments are positioned between the strict control of true experiments and the great flexibility of observational studies. It is often used when randomisation of groups and a control group cannot be implemented [119]. Hence, participants self-select into the treatment group. This presents challenges for internal validity, as the lack of randomisation means the groups may not be equivalent at the outset [119]. Pre-existing differences between the groups, known as selection bias, can confound the results, making it difficult to attribute any observed effect solely to the intervention [119]. Therefore, while quasi-experimental designs are efficient for real-world research, their conclusions

Table 1.2: Overview of the included papers in the thesis, the research questions they answer, their design, and their data collection and analysis methods. All the papers followed a mixed-methods approach.

Paper	RQ	Design	Data Collection	Data Analysis
Paper A: Well-being Factors	1: Factors 2: Approaches	Exploratory Sequential Design	Questionnaire & Interviews	Statistics & Thematic & Content Analysis
Paper B: Emotional Strain	1: Factors 2: Approaches 3: Methodology	Survey	Questionnaire (Open & Closed)	Descriptive Statistics & Content Analysis
Paper C: Breathwork Intervention	2: Approaches	Quasi-Experimental	Questionnaire & Journals	Bayesian Inference & Thematic Analysis
Paper D: Yoga Intervention	2: Approaches	Quasi-Experimental	Questionnaire & Focus Groups	Descriptive Stats & Thematic Analysis
Paper E: Multimodal Approach	1: Factors 3: Methodology	Experiment	EEG (Biometric), Questionnaire & Interviews	Statistics & Thematic Analysis
Paper F: LLM for Qualitative Analysis	3: Methodology	Experiment	Questionnaire (Open & Closed)	Statistics & Content Analysis

about causality must be interpreted with caution, acknowledging the potential for alternative explanations.

We used two interventions in the form of quasi-experiments in this thesis. To increase validity and rigour in our quasi-experiments, we followed the implementation guidelines by Maciejewski [119]. Both studies (Papers C and D) implementing quasi-experiments used a one-group pre-test and post-test design. That is, data was collected from the same single group at three time points: before the intervention (pre-test), during its implementation, and after its completion (post-test). The difference lay in the psychometric instruments used for data collection and the methods employed for data analysis. Both intervention programmes had a mindfulness practice as a core practice. Each programme is explained next:

Online Intervention Rise 2 Flow (R2F). It was designed to help build mental and emotional resilience and enhance well-being. It was based on a yogic breathing practice called Pranayama. This practice was combined with two other mindfulness practices, journaling and meditation. The technique is a three-part breath through the mouth, practised while lying down. A certified facilitator guided sessions.

We implemented two rounds of the R2F, lasting 12 and 8 weeks, respectively. All sessions were held online once a week. The participants were IT workers who joined from 26 countries. The data collected were quantitative, coming from day ratings and psychometric instruments (such as entry and exit surveys), and qualitative from participants' journals. For the quantitative analysis, we used Bayesian analysis, and for the qualitative data, we used thematic analysis by Braun and Clark's [120] guidelines. The results are published in Paper C.

In-person Industry Intervention. The yoga intervention was done in a software development company. It was a weekly practice for eight weeks. The target population were software engineers. Participants had a 45-minute Hatha yoga session every Wednesday from 8:00 to 8:45, taught by a yoga instructor. These sessions focused on the principles of Hatha yoga, incorporating physical postures, breathing exercises, and relaxation techniques (5 min). The data collection was done using psychometric instruments to create a survey and obtain quantitative data. Additionally, for qualitative data, we organised a focus group. The results were published in Paper D. The data analysis was conducted using descriptive statistics and thematic analysis, as described by Braun and Clarke [120]. The results were published in Paper D.

1.4.1.2 Survey

Surveys offer a quantitative description of trends, traits, attitudes, and opinions of a population [116, 121]. By systematically collecting responses from a defined sample, surveys enable the aggregation and statistical analysis of self-reported data. This approach is particularly suitable for examining patterns across participants and for comparing responses across predefined variables or conditions. In this thesis, a survey research design was employed to gain an understanding of participants' perceptions, opinions, and psychological constructs in various study contexts. By using a survey design, we could identify patterns, averages, and correlations.

1.4.1.3 Experiment

To examine the relationships between variables, we employed controlled experiments. In experimental studies, one or more independent variables are deliberately manipulated in order to observe their effects on dependent variables while holding other factors constant [122]. We carefully controlled task conditions, standardised instructions, and consistent evaluation procedures. These guidelines informed the selection of variables, the structuring of experimental conditions, and the interpretation of results. For the experiment involving human participants, we followed established experimental design principles as outlined by Brysbaert [123], including careful control of task conditions, standardised instructions, and consistent evaluation procedures. For the experiment involving large language models (LLMs), we aimed to compare human and LLM outputs in the same task. The goal was to identify and measure similarities and differences in the performance of a specific analysis method.

1.4.1.4 Exploratory Sequential Design

This design employs a two-phase mixed-methods research approach, where the researcher collects and analyses qualitative data first to explore a topic, identify key themes or variables, and develop a theory, framework, or instrument [116]. Then, the qualitative results directly inform and guide the following quantitative phase (a survey or experiment) to test, generalise, or validate those initial findings on a larger scale. In this design, it is essential to focus on the appropriate qualitative findings to build a solid and useful foundation for the second phase (quantitative), and to select the correct sample and analysis methods.

1.4.2 Data Collection Methods

In this section, the specific techniques and instruments used to gather evidence as dictated by the research design are explained.

1.4.2.1 Questionnaires

We primarily used two different types: the first was creating the questionnaire from scratch, following Stol and Fitzgerald guidelines [124]. This first type employed open-ended, closed-ended, and Likert questions and was for exploratory purposes. The second type was a questionnaire made of psychometric instruments. This type was mainly used to measure changes after an intervention.

Psychometric instruments are tools designed to measure psychological constructs, attitudes, and behaviours in a systematic and quantifiable manner [125]. The main reasons to use them in this thesis were: validity, which refers to the accuracy with which an instrument measures the intended construct; reliability, reflecting the stability and reproducibility of measurements over time; and responsiveness, indicating the instrument's sensitivity to detect meaningful changes [126].

1.4.2.2 Interviews and Focus Groups

To gather views, opinions and perceptions of our study's participants, we used semi-structured interviews [116]. We chose the semi-structured format to allow our participants flexibility in expressing themselves and capturing unexpected yet valuable insights. Two different semi-structured interviews, open questions in questionnaires, and one focus group were used. Since the studies were exploratory, aiming to understand participants' experiences, perspectives, and emotions, **interviews** were the most suitable approach [117]. The **focus group** was chosen since we had two goals for that study. First, we wanted to know our participants' experiences in the intervention. Second, we aimed to investigate the interaction between the intervention's organisers. This method enabled a dynamic exchange of ideas, reflections, and shared experiences [127]. It provided deeper insights into the collective understanding of the intervention's design, delivery, and perceived impact.

We designed semi-structured interview guides to conduct the interviews and the focus group. This type of interview guide allowed respondents to expand on their answers, and me, as an interviewer, to ask follow-up questions and explore topics in depth.

1.4.2.3 Biometric Data

Biometric data is unique information about a person's physical (fingerprints, face, iris), physiological (heart rate, DNA), or behavioural (voice, typing rhythm, gait) characteristics [128]. We collected Electrodermal Activity (EDA) and Heart Rate Variability (HRV) using a wearable wristband and Electroencephalogram (EEG) using a Neurosity Crown device for one of the studies.

Human factors and their characteristics are a core part of all the studies in this thesis; hence, obtaining reliable and objective data is essential for our studies [129]. Therefore, we decided to collect biometric (EDA, EEG and HRV) information to go beyond self-reported answers. Our goal was to understand participants' responses, emotions, and cognition in a more in-depth and accurate manner.

1.4.3 Data Analysis

This section explains how the two types of data, qualitative and quantitative, were analysed in the included studies.

1.4.3.1 Quantitative Analysis Method

We used the quantitative data in two ways: first, to identify trends, traits, and overall perceptions, which were presented as means, medians, modes, and standard deviations. For this goal, we used descriptive statistics to summarise and describe our datasets. We also used these summaries to create visualisations that give readers an overview of our results. Second, we applied inferential statistical methods to analyse changes in responses over time at multiple temporal points(entry vs. exit, daily, and weekly trends).

We employed non-parametric and parametric frequentist tests to seek differences between groups. Similarly, we also studied the relationship among variables (Mann–Whitney U, Kruskal–Wallis and Spearman). We also performed Bayesian analyses to examine responses in three ways: (1) temporal analysis for each instrument at t0 versus t1 (entry vs. exit), (2) temporal analysis of daily trends, and (3) temporal analysis of weekly trends.

1.4.3.2 Qualitative Analysis Method

For the qualitative data, we used two methods: **Reflexive Thematic Analysis** and **Content Analysis**.

For **Reflexive Thematic Analysis**, we followed Braun and Clarke's [75] guidelines, which consist of six steps, as explained below.

1. Familiarisation with the Data: Reading and re-reading the data.

2. Generating Initial Codes: Systematically going through the data to label meaningful segments (codes) relevant to the RQs.

3. Generating Initial Themes: Grouping and organising the codes into broader, potential themes that reflect meaningful patterns across the dataset.

4. Reviewing Themes: Evaluate the potential themes against the coded data, the entire dataset and the RQs. Refine, combine, split, or discard themes to ensure they are coherent and distinct.

5. Refining, Defining and Naming Themes: Identify the themes' central idea, define and name each theme and evaluate the general structure.

6. Writing the Report: Write the story following the themes' narrative and using quotes to answer the study's RQs.

Similarly, for **Content Analysis**, we followed the guidelines by Kuckartz and Radiker [130]; the seven steps are explained below.

1. Initiating Text Work: Read and highlight important passages and write "case summaries" to grasp the overall context.

2. Developing Main Categories: Create broad codes based on the study's RQs.

3. First Coding Cycle: Code the entire material using these broad main categories.

4. Inductive Sub-categorisation: Create sub-categories within the main categories, focusing on the central categories for the study.

5. Second Coding Cycle (coding data with sub-categories): Re-code the entire dataset using the now-complete system of main and sub-categories.

6. Simple and Complex Analysis: We chose to create visualisations and data display in this step.

7. Analysis and Presentation: Compare categories across cases and write up your results.

1.4.4 Reflexivity

I am a PhD student with a background in behavioural and social sciences. My background shaped my epistemological positioning, choice of research questions, preference for mixed-methods approaches, and emphasis on participants' subjective experiences.

I acknowledge that my lack of a software engineering background made interpreting domain-specific terminology, situating findings within software engineering practices, and articulating technically grounded practical implications challenging at the beginning of my PhD. I relied on my supervisors' technical knowledge, professional experience and insights to support interpretation and contextualisation.

Through prolonged engagement with software engineering research communities, repeated interaction with practitioners, and sustained immersion in the domain throughout the PhD, my positionality shifted. While I remain professionally grounded in social and behavioural sciences, I have developed substantial domain familiarity and sensitivity to software engineering practices, norms, and constraints. As a result, I came to occupy a hybrid insider–outsider position, functioning as a domain insider regarding software engineering culture and concerns, while retaining an external disciplinary perspective.

This hybrid positionality had methodological implications. My background supported sensitivity to affective, cognitive, and well-being-related aspects of participants' accounts. Meanwhile, my developing domain knowledge helped me with a more nuanced interpretation of software engineering-specific practices and constraints. At the same time, maintaining an external disciplinary stance supported critical questioning of the field's assumptions. Reflexive engagement with this shifting positionality was therefore central to data interpretation, analytic decision-making, and the development of practical implications.

1.5 Ethical Considerations

This thesis topic core was human factors; hence, having humans in all the studies was imperative to gather data and evidence to create a basis for informing the development of guidelines, policies, interventions, and programmes. To ensure the safety and privacy of participants in studies involving humans, oversight by an Institutional Review Board is necessary. For the included studies, we consulted the Swedish Ethics Review Authority (etikprövningsmyndigheten). We obtained ethical approvals from etikprövningsmyndigheten [131]. We also followed the guidelines of Chalmers University for ethical research and The National Institutes of Health's seven principles of ethics for human subjects research [132]:

1. Social and clinical value. Our research aims to improve the well-being of our target population, software engineers.

2. Scientific validity. We employed rigorous and appropriate scientific methods that contributed to the body of evidence for the SE field.

3. Fair subject selection. We invited a diverse range of software engineers to participate in our studies. We did not limit participation based on age, sex, race/ethnicity, or sexual orientation.

4. Favourable risk-benefit ratio. We tried to minimise any risks or discomfort for our participants. For example, in the interventions, we informed participants of potential discomfort, such as a dry throat, and provided them with advice on what to do after the intervention.

5. Independent review. We review the interventions and their corresponding activities. We also piloted all survey and interview guides to prevent any misunderstandings or identify activities or questions that could be considered socially, racially, and/or ethnically inappropriate.

6. Informed consent. For all studies, we collected written informed consent from all participants. They received an explanation of the study's objectives, methods, and potential risks, as well as their right to withdraw from the study at any time.

7. Respect for potential and enrolled subjects. To keep our participants' privacy, all data was anonymised and securely stored. During the interventions and experiments, we monitored our participants to ensure they were not experiencing any discomfort.

We also offered compensation without undue inducement. Participants in Paper C received a donation to a charity of their choice as a token of appreciation. Similarly, participants in Paper E received a voucher for a meal or drink, while those in the control group of Paper D received a gift card. We followed the arguments by Wikilson and Moore [133] and concluded that gift cards did not bias or put our participants at risk.

1.6 Research Results

This section presents the results in the form of a summary of the included papers. Each paper starts with the main goal or motivation, then an overview of the findings and finally the implications. The full papers are in the coming chapters. The papers follow the argument presented in the Research Focus Section 1.1.

Paper A - Well-being Factors

The primary motivation of this paper was to empirically explore the factors, from the software engineers' perspective, that influence their well-being. We wanted to gather experiences from engineers from various parts of the world to gain a comprehensive picture and also to make our model contextually relevant.

The methodology that best suited this paper's goal was a mixed-method combining surveys and interviews. We interviewed 16 software engineers in Sweden to get deep insights and experiences in their daily work. Then, we created the survey in three languages (English, Spanish and Portuguese) to reach a large number of respondents to examine whether the themes identified in the interviews were reflected more broadly across the international sample.

Findings: A framework that identifies the main factors shaping well-being, that considers individual perceptions of well-being, interpersonal and collaborative relations, workplace support and recognition, organisational culture, and stressors arising within software engineering. The framework is presented in Figure 1.2.

Implications: This paper provides policy guidelines and recommendations for organisations to support the well-being of engineers. For research, our framework can inform the design of well-being interventions and future empirical studies.

Paper A's findings align partially with other well-being frameworks, such as Gallup [76], the Perma model [77], and Michaelson's framework [9]. However, Paper A focuses on software engineers and considers the characteristics of their working context. Regarding previous works on the SE context, Paper A considered engineers' answers from different countries, in contrast to Wong et al. [87], whose work focuses on the USA. Furthermore, Wong et al. examined internal self-reported well-being experiences, while Paper A also considers external factors such as company culture and peer support. Moreover, Paper A's results are based on survey and interview data, unlike the work of Godliauskas and Šmite [88], which relied on literature reviews. This difference is substantial because primary data offers a more direct and contextually grounded basis for developing a well-being framework than secondary synthesis alone.

This paper contributes to answering RQ1 by defining the first of the proposed frameworks, which explores the factors influencing the well-being of engineers as a distinct population. By explicitly grounding these influences in a SE context, the framework points to what should be considered relevant when studying well-being-related phenomena in this domain. It also answers RQ2 by proposing guidelines for organisations and policy recommendations to support the well-being and good practices.

Paper B - Emotional Strain in LLM Interactions

Paper B extends the conceptual framework of Paper A into the technological sphere, demonstrating how AI-based tools can introduce an additional layer of stress. Although the technological aspect was mentioned in Paper A, given the current updates and state of AI, it was necessary to explore this area in greater depth. In this paper, the source of strain shifts from workload or interpersonal tension to the interaction between human expectations and machine behaviour.

We used a survey to gather engineers' experiences interacting with LLMs. We obtained 62 answers and analysed them using content analysis. We used Wilcox's Emotions Wheel to categorise participants' answers and conceptualise their emotions.

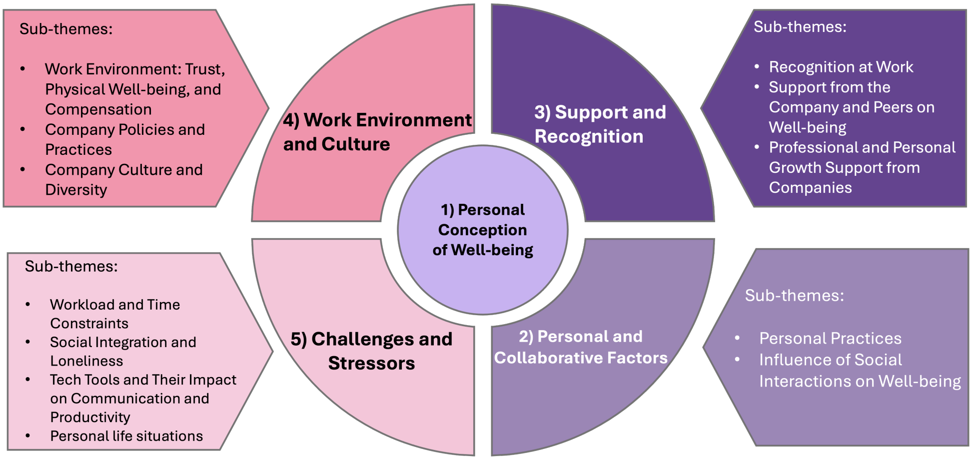

Figure 1.2: Themes from Paper A showing the framework of factors influencing the well-being of software engineers. These factors were later compared to the survey answers.

Findings: Software engineers using LLMs in their daily tasks reported distinct emotional responses. These emotional answers ranged from curiosity and satisfaction to frustration, disappointment, and guilt when these systems produced incorrect, misleading, or verbose outputs. Several participants' experiences were manifestations of techno-frustration [134], a specific form of techno-stress driven by perceived inefficacy, cognitive dissonance, and loss of control during digital interactions. Importantly, participants displayed adaptive resilience strategies, such as refining prompts, cross-verifying outputs, or switching tools. However, repeated failures or "hallucinations" led to cumulative strain.

Implications: Based on our results, we proposed recommendations for designing tools that reduce stress associated with user interaction. Similarly, we argue that companies need to prioritise employees' emotional intelligence training to cope with techno-stress.

The psychological effects of using LLMs are a relatively new area of research. Most studies focus on how LLMs are used at work [20, 135, 136], and how they influence workers' efficiency and efficacy [137, 138]. However, their impact on emotions remains largely underexplored.

Paper B is the first study to investigate emotional responses in LLM–human interaction specifically. Recently, only the work by Maitipe [139] explored the psychological impact of LLMs on IT professionals. However, the emotions involved in this interaction are not yet well studied, and their influence on workers' well-being and performance remains largely unexplored. Emotions have a direct impact on stress, satisfaction, and overall mental health. Understanding how LLMs influence feelings like frustration, anxiety, or confidence helps design systems that support users rather than harm their well-being [89].

Paper B contributes to answering RQ1, RQ2, and RQ3 by extending the investigation of well-being and stress into the technological domain through the lens of human–LLM interaction. It identifies how AI-based tools introduce new forms of techno-frustration and cognitive demand. For RQ2, Paper B proposes design-oriented and organisational recommendations to mitigate emotional strain and support sustainable tool use in practice. Furthermore, Paper B contributes to RQ3 by exemplifying the adaptation of a psychological instrument for emotion classification to the SE context.

Paper C - Breathwork Intervention

This paper presents the results of a breathwork intervention. The intervention was the implementation of R2F p.14, a programme designed to teach breathwork to IT workers in weekly online meetings. The goal was to help participants manage stress, increase well-being, and develop resilience. R2F's Thursday sessions followed this structure: 1) It started with participants answering questions on the weekly self-development topic. They received the topic on Monday and had time to reflect and answer the questions. 2) The breathing practice for three rounds of seven minutes. 3) 20-minute relaxation. 4) Aftercare suggestions (e.g., to hydrate well) and time for participants' questions. The data collection was done before (with the entry survey), during (with a written journal and a weekly survey) and after the intervention (with an exit survey). The entry and exit survey was created with the following psychometric instruments:

- Mindfulness Attention Awareness Scale (MAAS) [140]
- The Scale of Positive and Negative Experience (SPANE) [141]
- The Psychological Well-Being scale (PWB) [141]
- The Positive Thinking Scale (PTS) [141]
- Perceived Productivity instrument (HPQ) [142]
- Self-Efficacy instrument [143]

Thematic analysis was used to analyse the qualitative data, and a temporal analysis was conducted for the quantitative data (for each instrument, comparing entry vs. exit surveys, as well as daily and weekly trends).

Findings: The results indicated that the R2F programme may help improve participants' mindfulness attention awareness, well-being, and self-efficacy.

Implications: We identify three types of implications. For policy, organisations need to create concrete actions on mental health awareness. For research, our programme proved to be effective; more modalities need to be tested and adapted to in-person interventions. For practice, modelling healthy well-being habits is more effective than only talking about them. Hence, managers and teachers should demonstrate these habits in their daily work to foster a healthier and more supportive environment.

Paper C presented an online intervention, which enabled participants from various locations around the world to be reached. Unlike other in-person interventions [92, 93, 144, 145], the online format allowed participants to practise breathwork to manage stress more effectively and increase their well-being and resilience. This was particularly important since the interventions were done during pandemic times. In the software engineering context, no other study has targeted this population in a large-scale, remote intervention that combines stress-management techniques with outcomes related to well-being and resilience.

This paper addresses RQ2 by empirically evaluating R2F to support stress regulation and well-being among software engineers and IT workers. It provides evidence on how structured, recurring practices can foster individual resilience and emotional regulation. The findings inform how such interventions can be designed, measured, and implemented in real-world settings, explaining their potential benefits and practical considerations for sustained engagement.

Paper D - Yoga Intervention

This paper presents the second intervention appended in this thesis. To measure the effectiveness of yoga in improving general well-being among software engineers, we implemented an eight-week yoga programme. The intervention was done in collaboration with a Swedish software company. We collected quantitative data using the following psychometric instruments as entry and exit surveys:

- The Schutte Self-Report Emotional Intelligence Test (SSEIT) [146]
- The 14-Item Resilience Scale (RS-14) [147]
- Short Form Self-Regulation Questionnaire (SSRQ) [148]
- Self-Transcendence Scale (STS) [149]
- The Flourishing Scale (FS) [141]
- Brief Resilient Coping Scale (BRCS) [150]

Additionally, we used the WHO-5 Well-being Index [151] as a weekly survey and a focus group with the organisers to collect qualitative data. One of the objectives was to have a control group, which was not possible due to the small number of volunteers.

Findings: Results from the psychometrics did not reveal any statistically significant differences between the entry and exit surveys. However, the qualitative results showed participants experienced positive effects after the sessions. Our conclusions focused on how contextual factors, in this specific case, layoffs, critical deliveries, and non-work demands, time pressure, emotional intensity, and schedule disruptions associated with Christmas celebrations can mitigate the positive effects of yoga.

Implications: We shared lessons learned that can inform future mindfulness interventions in the workplace. Particularly, the importance of tailoring interventions to consider the context and unique needs of participants is one of the main implications.

Paper D is the first study involving software engineers practising yoga. Previous interventions in software engineering have mainly used mindfulness as the primary practice. For example, Bernardez et al. [92] ran three controlled experiments with students practising mindfulness and focusing on conceptual modelling. Paper D instead involved engineering practitioners and investigated general well-being, making the context and population markedly different. One more study by Romano et al. [93] introduced a Mindfulness-Based Stress Reduction (MBSR) program. However, they only focused on collecting qualitative data. In contrast, Paper D employed a mixed-methods design, combining quantitative data from validated psychometric instruments with qualitative insights from focus groups. This provided a more robust and comprehensive assessment of the intervention's impact. Finally, Bernardez et al. [144] continued their experiments by having software workers participate in mindfulness practices. In Paper D, yoga included mindful awareness but extended beyond it through physical postures and breath-based exercises, offering a more comprehensive approach than mindfulness alone.

Paper D also contributes to RQ2 by complementing Paper C's findings by illustrating how contextual and organisational factors can mediate the effectiveness of individual-level well-being interventions. The lessons learned from this paper aim to strengthen approaches to fostering sustainable well-being. They also refine the understanding of the conditions necessary for well-being interventions to succeed in practice.

Paper E - Multimodal Methodology

The motivation of this paper was to introduce a physiological dimension by utilising biometric data to measure stress, mental workload and emotional responses during programming tasks. We designed an experiment to expose participants to stress associated with limited time constraints while programming.

We combined three data sources to collect data during and after the programming tasks:

- Biometrics: EEG (electroencephalography), EDA (electrodermal activity), and HRV (heart rate variability) sensors.
- Validated psychometric scales: Perceived Stress Scale (PSS-10) [152], Short Stress State Questionnaire (SSSQ) [153] and NASA Task Load Index (NASA TLX) [154].
- Interviews.

The data was analysed using thematic analysis for the interviews and descriptive analysis and T-tests for the quantitative data.

Findings: The psychometric results did not show any differences when comparing the tasks with and without time limitations. However, while participants claimed to feel relaxed or neutral, EDA data showed micro-level spikes in arousal, especially during time-constrained tasks.

Implications: Based on our findings, we proposed guidelines and considerations for research in stress, mental workload and emotions in SE. Paper E's findings challenge the reliability of self-report instruments in isolation. It also emphasises that stress in SE can be non-conscious or cumulative, accumulating over repetitive micro-stressors like debugging or code compilation delays. Hence, it also evidences the importance of multimodal methods for accurately capturing human experiences in technical contexts.

Paper E contributes to answering RQ1 and RQ3 by advancing the understanding of stress-related phenomena, revealing discrepancies between self-reported experiences and physiological indicators, and demonstrating that stress may be subtle, cumulative, or non-conscious. Methodologically, the paper addresses RQ3 by proposing guidelines for integrating biometric data, psychometric instruments, and qualitative interviews, thereby strengthening the rigour and interpretive depth of empirical human-factors research in SE.

Paper F - LLMs as Analytical Assistants

Paper F aimed to advance qualitative data analysis methodology by integrating LLMs as analytical assistants. To achieve this, we designed an experiment to compare human and LLM-generated steps 2 (creating initial codes) to 5 (creating, naming, and refining themes) of thematic analysis by Braun and Clarke. We used 15 interviews from a previous study that had already been coded and analysed by human researchers. First, we created a prompt for the LLM to create initial codes. We compared those codes with the human-made and asked external experts to evaluate them using a tailored rubric. Then we had the LLM create themes and also compared them with those of the human researchers. These themes were also systematically evaluated using a rubric designed based on Braun and Clarke's guidelines.

Findings: Evaluators preferred LLM-generated codes 61% of the time over the human ones. They found them analytically useful for answering the research question. However, evaluators also pointed out the limitations of LLM codes and themes.

Implications: - A reproducible approach integrating refined, documented prompts with an evaluation framework to operationalise Braun and Clarke's reflexive TA.
- An empirical comparison of LLM- and human-generated codes and themes in software engineering data.
- Clear guidelines for integrating LLMs into qualitative analysis, preserving methodological rigour.

Figure 1.3 illustrates our proposal for integrating LLMs in thematic analysis, specifically delimiting LLM and human activities. LLMs act as assistants in steps 2 (creating initial codes) and 3-5 (creating, naming, and refining themes). Researchers supervise, evaluate and refine the steps done by LLMs. Steps 1 (familiarising with the data) and 6 (writing the final report) remain entirely done only by the researcher.

Despite several studies aiming to automate qualitative analysis methods, either partially or fully [28, 29, 112], Paper F extends the literature by evaluating LLM-generated outcomes in terms of interpretive depth, theme coherence, and alignment with the RQs. In this paper, we also provided rubrics to assess the quality of codes and themes, which can be applied to data generated by LLMs or by human researchers.

Paper F addresses RQ3 by advancing methodological strategies for qualitative data analysis in SE through the integration of LLMs as analytical assistants. The study contributes concrete guidelines for maintaining methodological rigour, transparency, and reflexivity when incorporating AI into qualitative research workflows.

1.7 Discussion and Answers to the RQs

This thesis used mixed-method and multi-modal studies to integrate psychological aspects, organisational perspectives, and SE practice to explain and theorise how stress, resilience, and well-being are experienced and shaped in contemporary software

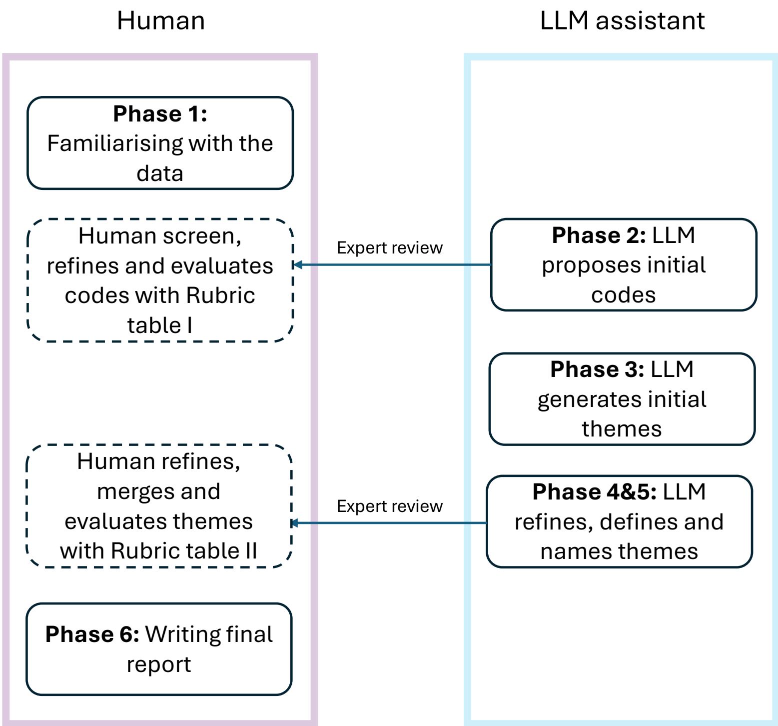

Figure 1.3: Proposal of implementation of LLM in thematic analysis (TA). LLM contributes to TA Phases 2–5 as an assistant; the human leads Phases 1 and 6 and gates progression using rubric-based evaluations. Dashed boxes indicate areas that require human evaluation and refinement.

engineering contexts. It conceptualises software engineers' well-being as an emergent property of interacting individual, social, and technological systems. Our results from empirically examining mindfulness-based interventions, emotional strain in human–LLM interaction, and the limitations of single-source self-report data show the potential and constraints of current approaches to supporting engineers' well-being.

The next paragraphs elaborate on the thesis's contributions to the state of the art in software engineers' well-being, resilience and research on human factors.

1.7.1 Factors and conditions that influence stress, well-being, and resilience

Results from papers A, B, and E answer RQ1 and show that well-being and stress in software engineers are multifactorial. Several factors from different contexts and systems interact and shape how software engineers feel and cope with stress. These factors operate at various levels: individual, interpersonal, and organisational. This makes them a bioecological and sociotechnical phenomenon, where individual characteristics and coping strategies are continuously influenced by social relations, organisational conditions, and the technological ecosystem in which work is taking place [155, 156]. Figure 1.4 shows the Individual–Social–Technological System (ISTS) integrated framework proposed in this thesis, which presents the factors influencing well-being and stress management in software engineering, derived from Papers A and B.

The ISTS framework (in Figure 1.4) shows well-being as the outcome of a balance between stressors and resources arising simultaneously from three interdependent systems: individual resources, social and organisational environment, and technological ecosystems.

We believe that well-being is a product of continuous interaction between personal capacities, social structures, and technological conditions. It is dynamic, interactive, cyclical, and domain-crossing, and it changes under the influence of any of the three spheres. Hence, any changes in any one system can ripple into the others, reshaping the overall balance. Several elements belong to more than one system. For example, mental workload is both an individual resource and part of the technological ecosystem, as both areas add to how engineers experience and manage their cognitive demands.

The ISTS framework's spheres are:

Individual Resources: This domain captures personal factors (habits, physical and emotional practices, life circumstances and the individual's conception of well-being) that shape people's starting point for managing stress. This sphere is also present in several other well-being frameworks; however, we give more importance to the work-related aspects, for example, we consider the mental workload.

Social and Organisational Environment: This sphere addresses the relationships, mainly at work, of engineers. Specifically, it considers social interaction and integration, company policies and culture, company and peers support, and recognition. Our framework considers these aspects as direct influences on well-being, rather than as background context or indirect influences, as in Michaelson's [9] and Seligman's [77] model.

Technological Ecosystem: For this sphere, we consider task demands, digital tools, automation and AI/LLM interactions, and technology in general as mediators in workflows. They also shape cognitive load, time pressure, attention fragmentation, and the overall stress in engineers' work. The technological ecosystem introduces constraints and affordances that can either amplify or reduce stress, making it an essential component of any well-being framework in software engineering. We use the term 'ecosystem' to emphasise the interrelation and interdependency of the system's component elements. Despite its relevance, this sphere is not addressed in key well-being frameworks such as those proposed by Seligman [77], Gallup [76], and Michaelson [9].

Previous well-being frameworks [9, 76, 77] focus on the general population and view well-being as either a psychological or life domain outcome, but rarely as an integration of both. We consider the interactions of the primary systems present in software engineers' lives.

Moreover, these frameworks are also context-independent. In contrast, our framework considers SE-specific context characteristics, such as sustained cognitive load, sociotechnical collaboration, rapidly evolving technologies, chronic time pressure, and

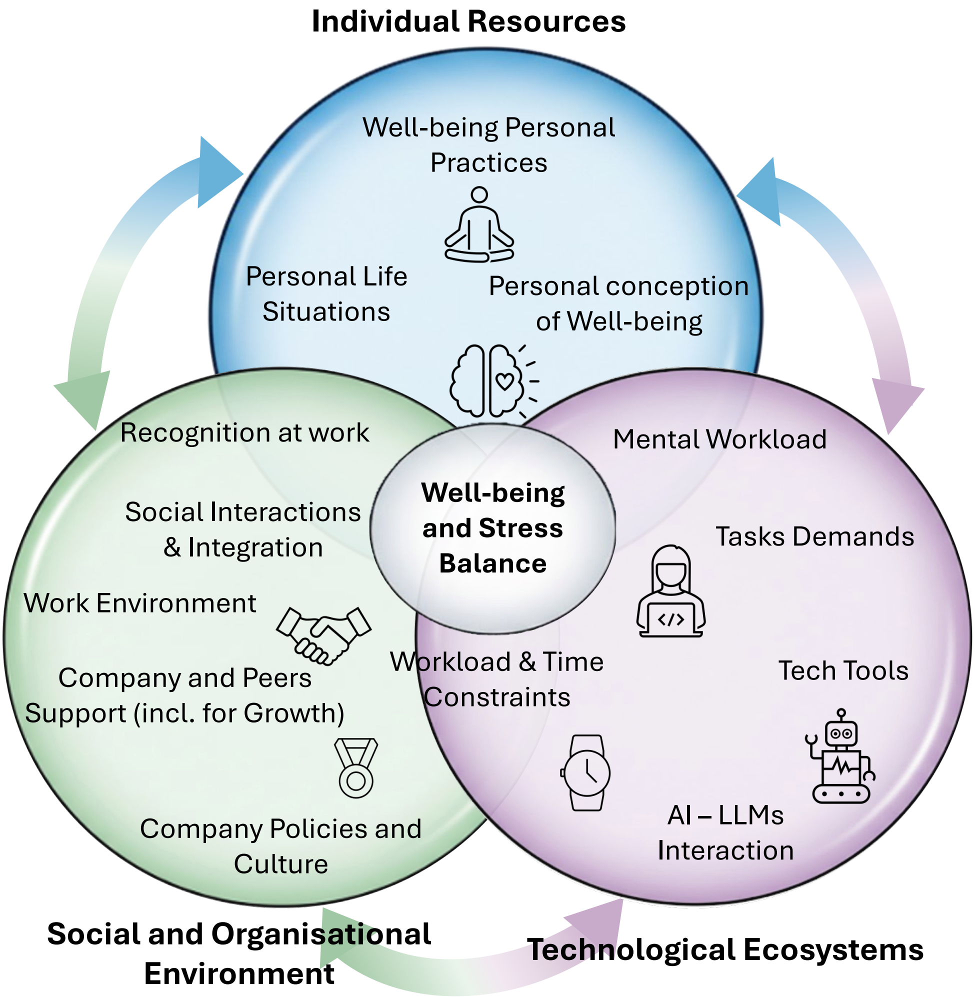

Figure 1.4: The Individual–Social–Technological System (ISTS) integrated framework of well-being and stress balance in software engineering. The framework illustrates how individual resources, social and organisational environments, and technological ecosystems jointly shape well-being and stress. The arrows indicate dynamic, cyclical interactions, whereby changes in one domain can propagate across the system.

particular cognitive demands.

Our individual-social-technological system views well-being as an emergent property, shifting the trend in traditional frameworks from seeing it as a set of dimensions [157]. For example, in our framework, individual habits influence how technology is used. At the same time, technology shapes organisational expectations. This explains how well-being arises within an individual-social-technological system, and not only what well-being looks like for engineers.

At a theoretical level, the ISTS framework proposes the following claims to explain how well-being emerges from the configurations and interactions among its spheres.

Specifically, the framework implies that (i) the effectiveness of individual regulation strategies is contingent on organisational enabling conditions, (ii) technological systems influence well-being indirectly by reshaping the distribution of cognitive and temporal demands across levels, and (iii) stress and resilience outcomes are shaped by alignment or misalignment between individual capacities, organisational structures, and technological practices. These claims articulate boundary conditions and interaction logics that can be operationalised and tested in future empirical research.

In Paper A, we explained that well-being depends on how these conditions converge to either balance or overload the engineer's mental and emotional resources [106]. For example, engineers are prone to experience higher well-being when they are recognised at work, have clearly defined roles, and work within inclusive and transparent leadership [158, 159]. However, when the situation is different, with deadlines tightening, interruptions multiplying, and context switching becoming continuous, the influences of this particular context can challenge individual resources to cope with stress [106]. Furthermore, the constantly changing technological context adds to the previous overview. The use of AI, particularly the increased interaction with LLMs, introduces new stressors that differ from traditional forms of technical frustration [89, 160]. Errors, hallucinations, or mismatched outputs require engineers to rephrase prompts, verify content, and correct the system's responses. This verification work becomes an additional source of demand that fragments attention and increases effort, especially when it must be performed under time pressure. The accumulation of such microfailures generates sustained emotional friction, contributing to fatigue over time.

Different factors influence at various levels, yet their effects are not simply additive. Individual traits such as resilience or self-awareness may buffer stress, but their effectiveness depends on whether organisational and social contexts enable their expression [106]. For instance, supportive leadership or psychologically safe environments can transform individual coping efforts into collective resilience. On the contrary, rigid hierarchies or conflicting demands can neutralise even strong personal resources. This interconnectedness implies that well-being in software engineering cannot be fully supported through individual-level interventions alone.

Structural and cultural aspects, such as unclear expectations, unstable workflows, and performance-driven norms, can perpetuate stress regardless of personal coping abilities [155]. Therefore, improving well-being and resilience requires systemic adjustments that align organisational practices, interpersonal dynamics, and individual strategies [106]. Stress recovery and coping are thus a manifestation of how well the surrounding system supports sustainable human functioning within the technological and social realities of software development.

One more factor to consider is the "not so visible" manifestation of stress [161]. In Paper E, the results indicated that stress, as measured by self-reported instruments, may not be accurately captured. However, electrodermal activity showed more frequent phasic peaks, indicating subtle physiological arousal. Since software engineers often work under high mental workload, they may not consciously recognise their stress. Additionally, it is important to consider variability across individuals. In Paper E's experiment, not all participants displayed the same physiological sensitivity. Hence, stress cannot be inferred solely from uniform group-level trends.

Well-being and resilience are maintained when organisational systems and technologies work together to stabilise demands [106, 162]. This prevents mental workload from escalating into sustained physiological strain. Resilience, then, is the visible outcome of an environment. This environment spans individual, organisational, and team levels, providing opportunities for recovery. It also enables software engineers to sustain performance without incurring hidden physiological costs.

Understanding this interdependence is essential for designing work environments and technologies that foster sustained well-being and resilience in software engineering.

1.7.2 Approaches to foster sustained well-being

Building on the previous question, a coherent program for sustained well-being must consider three key areas: individual regulation capacity, organisational context [163] and AI interaction.

At the individual level, Papers C and D investigated the effectiveness of mindfulness-based practices in cultivating resilience and emotional regulation. Paper C provided evidence that breathing practices had a positive effect on participants when done weekly. Notably, the study also demonstrated that sustained, structured engagement is essential. Short-term exercises produce transient effects, whereas weekly facilitated sessions promote the internalisation of well-being habits.

However, the results from Paper D (Yoga Intervention) were not as positive as those from Paper C. This study exposed a central challenge in workplace interventions, contextual interference from ongoing work stressors and organisational culture. Participants worked on their individual well-being but were overcome by the work context. They reported positive subjective experiences in qualitative data. These perceived benefits were strongly influenced by organisational and temporal context, including workload pressure and external stressors. Hence, sustained well-being also depends on organisational integration [164] to create organisational conditions that support individual participation.

Paper A complements these findings and suggests organisational strategies and policies that strengthen interventions. Multi-level interventions are more effective than isolated wellness initiatives. Furthermore, Papers B (LLM Frustration) and E (Biometrics) point toward technological design interventions. Paper B argues for "emotionally intelligent" digital tools that recognise user frustration and adapt accordingly. Paper E supports biofeedback-informed environments, where physiological data can alert individuals or teams to early signs of stress overload.

In the big picture, sustained well-being in software engineering requires a three-pronged strategy:

- Individual empowerment through evidence-based practices (e.g., breathwork, mindfulness, yoga).
- Organisational commitment to supportive structures (e.g. recognition, fairness, manageable workload).

- Technological empathy, where tools and systems are designed to minimise frustration and cognitive overload.

To conclude and answer RQ2, based on my results, approaches that address more than one level, with a particular focus on the individual, are more effective in enhancing well-being. Structured programmes with mindfulness-based practices at the individual level can be feasible to implement in SE contexts [27,145,165]. However, their effectiveness appears to be context-dependent, varies across practices, and is not consistently captured by quantitative measures. Implementation time (time of day and season of year) is a crucial factor to consider, as is the inclusion of supporting activities that enhance the effects of mindfulness-based interventions. Additionally, at the organisational level, practices and policies should be tailored to foster sustained well-being.

1.7.3 Multimodal data triangulation and LLM-assisted analysis

Papers B, E and F provide a multi-modal, mixed, rigorously transparent methodology for investigating human factors in SE. The combinations of results answer RQ3. These papers' methodologies propose that to better understand human aspects, it is imperative to cover the depth and breadth of the object of study. It is also important to collect subjective and objective data observations. Gathering qualitative and quantitative data will provide a complete overview, helping to understand subjective experiences, measurable indicators, and contextual factors.

By combining mixed methods, each paper contributed a layer of study: the emotional and experiential dimension, the physiological dimension, and the analytic reproducibility of qualitative interpretation. For example, in Paper B, I used the Willcox Emotion Wheel [166] to map engineers' emotional states, adapting a psychological tool to SE contexts, which allowed quantification of affective states without oversimplification. The adaptation of a qualitative instrument from psychology helped us to translate complex emotional states into a structured format suitable for quantification. This helped us to integrate experiences with measurable data, allowing us to write more informed recommendations.

In Paper E, I implemented a multi-modal data triangulation. The experiment combined psychometric instruments, biometric sensing (EEG, EDA, HRV), and qualitative interviews. The design aimed to overcome the limitations of self-reports, offering a more objective and continuous measure of stress. Part of the goal was to test whether the tailored survey could reliably measure stress and mental workload, and whether the results were supported by the biometric data. However, the experiment was not conclusive and surfaced important ethical and interpretive challenges. It demonstrated the need to utilise more sensitive tools for measuring stress and mental workload. In this paper, psychometric instruments alone underestimated subtle strain, since they were not able to capture it. EDA captured variations that occur below conscious awareness [167], and the interviews showed that stress was present in the participants. The results point to the need to implement a measurement that accounts for individual variability and subtle traits of the object under study. Regarding the ethical aspects, creating higher levels of stress in participants to the point that the instruments could capture it was not permissible, as deliberately inducing harmful stress would violate research ethics. We raised this point in Paper E [161] and discuss its implications and limitations.

Finally, Paper F advanced qualitative methodology by integrating AI into the analytical process. Qualitative data forms a central component in exploring and understanding human factors. In this paper, we introduced an LLM-assisted collaborative framework designed to enhance transparency, reproducibility, and scalability in qualitative analysis. We also developed rubrics to systematically evaluate the quality of generated codes and themes. By documenting each step of the analytical process (from data preparation and coding to theme development), Paper F addressed a persistent challenge in QD analysis: the opacity of qualitative reasoning and replicability [168, 169]. Making this process publicly available, following clear guidelines rather than only presenting final results, further strengthens external validity.

We acknowledge the recent methodological critiques and concerns regarding the use of LLMs in qualitative analysis, particularly within reflexive approaches such as thematic analysis. Scholars, including Braun and Clarke, have cautioned that uncritical or automated use of such tools risks undermining reflexivity, interpretive depth, and researcher accountability, potentially reducing qualitative analysis to a mechanistic coding exercise rather than an analytic process grounded in meaning-making [170]. The hybrid framework proposed in this thesis explicitly aligns with these critiques by positioning LLMs as analytical assistants rather than analysts [171]. In Paper F, LLMs are deliberately constrained to support specific phases of thematic analysis under continuous human supervision. At the same time, interpretive authority, reflexive judgement, and theoretical sensitivity remain the responsibility of the researcher.

The study of human factors encompasses multiple variables that can compromise the robustness of empirical findings. For example, even when triangulating data, biases still exist, instruments can fail to identify nuances, and it is not possible to control for all confounding variables [172]. Recognising these boundaries is itself a methodological principle: validity in human-factors research depends as much on documenting what the data cannot show as on what it reveals.

To summarise, these studies implemented and propose a multi-level triangulation strategy: behavioural and emotional data, physiological indicators, and interpretive rigour in thematic analysis. A methodology that is a system of cross-validation, where each data type complements the interpretation of the others and is reflexively documented.

1.8 Limitations and Threats to Validity

This section presents a general scope and overview of the threats and biases encountered during our studies. In addition, each paper has a detailed section explaining its specific threats.

1.8.1 Scope of Applicability

The findings and frameworks proposed in this thesis should be interpreted in light of the sampling and contextual characteristics of the empirical studies. Across the included papers, participation was largely voluntary, which may have led to an overrepresentation of individuals already interested in well-being, mindfulness-based practices, or reflective approaches to work. Consequently, the ISTS framework (and the other frameworks in the papers) may place greater emphasis on resources and coping strategies salient to such populations. At the same time, these frameworks might underrepresent experiences of software engineers who are sceptical of, constrained from, or disengaged from well-being initiatives.

This limitation is particularly relevant for the intervention studies (Papers C and D), where self-selection and attrition may bias outcomes toward participants who were able or motivated to sustain engagement over time. Similarly, the stress responses observed in the experiment in Paper E were collected in a controlled academic setting and with a limited organisational range. Stress manifestations and physiological sensitivity may differ in environments characterised by chronic time pressure, lower psychological safety, or different organisational cultures.

As such, the findings are examples of how subtle or non-conscious stress responses can emerge under specific task conditions. These examples may not represent stress in all SE contexts.

This thesis's contributions are likely to transfer to SE contexts that share comparable characteristics: knowledge-intensive work, sustained cognitive load, and partial organisational openness to reflective or preventive well-being approaches. In contrast, contexts marked by extreme workload, limited employee autonomy, or low tolerance for non-production activities may require additional organisational or structural interventions beyond those examined here.

1.8.2 Internal Validity

Internal validity concerns whether the observed effects in each study can be confidently attributed to the intended intervention or condition rather than to confounding factors. Across the studies in this thesis, several potential threats were identified. In the intervention studies (Papers C and D), it was difficult to isolate the effects of the breathwork and yoga practices from the influence of group interaction or the novelty of participating in a community activity. Some participants emphasised the social component as a significant contributor to their perceived well-being, which complicates causal interpretation. Furthermore, the absence of control groups (considering also Paper E) limited our ability to distinguish intervention effects from spontaneous changes or placebo-like influences. In the biometric experiment (Paper E), individual differences in physiological reactivity introduced additional internal validity concerns, as not all participants exhibited comparable sensitivity to stress induction. To mitigate these issues, the studies used pre–post comparisons, triangulation of data, and followed standardised measurement protocols to enhance the credibility of observed effects. We also acknowledge that controlling for confounding variables was not possible given our quasi-experimental settings.

1.8.3 External Validity

External validity refers to the extent to which findings can be generalised beyond the studied samples, contexts, and tasks. The thesis collected data from industrial and academic settings, with participants ranging from professional software engineers to computing students. However, several limitations affect generalisability. Recruitment relied on personal and professional networks, academic mailing lists, and online channels, which may have resulted in an overrepresentation of participants who were already interested in mindfulness or well-being. Similarly, self-selection bias may have favoured individuals with positive attitudes toward the interventions, potentially inflating observed benefits. In the multimodal experiment, participants were purposively sampled, which may not represent the broader population of software developers. Because of the previous, the results may not translate to the "average" sceptical software engineer. To strengthen external validity, findings were interpreted with caution, and the triangulation of diverse samples, covering organisational, experimental, and educational contexts, was employed to identify consistent patterns across settings.

1.8.4 Construct Validity

Construct validity addresses whether the empirical indicators used in the studies truly capture the theoretical concepts of interest, such as stress, well-being, and resilience. Several steps were taken to safeguard construct validity. Several studies employed validated psychometric instruments recommended in occupational and psychological research. In the other cases, for example, for Papers A and B, we included the definitions of the concepts being explored. This helped participants to develop a shared understanding of key constructs such as frustration and emotional strain, thereby reducing the risk of misinterpretation. Nonetheless, measurement limitations were noted: participants occasionally reported survey fatigue, which might have reduced response accuracy, and psychometric instruments sometimes failed to capture subtle, non-conscious stress reactions. Paper E explicitly addressed this by integrating physiological measures (EEG, EDA, HRV) with self-reports and qualitative interviews, revealing discrepancies between subjective and objective indicators.

1.9 Conclusions

This thesis analysed and theorised the well-being of software engineers. It considered stress and resilience as central determinants for exploration and intervention. Through six empirical studies employing mixed, qualitative, and physiological methods, it contributes to a psychological, organisational, and methodological understanding of human factors in SE. It proposes evidence-based strategies to foster sustained well-being at individual and systemic levels.

Across the mixed-method studies (Papers A, B, and E), we identify and delimit the **factors influencing stress, well-being, and resilience of software engineers (RQ1).** The findings demonstrate that well-being is an emergent property of a sociotechnical ecosystem. Individual coping capacities, interpersonal and team relations, and the broader organisational and technological environment influence stress and recovery processes. The integrated model proposed in Figure1.4 illustrates the interactions between these environments.

Our goal is that by having a comprehensive picture of the conditions that sustain, hinder, and enhance well-being, organisations can implement empirically informed guidelines, policies, interventions and strategies to improve engineers' well-being.

For research in this area, we aim for our framework to offer direction for future studies to explore specific areas deeper and have a point of reference to confirm, contrast, and challenge our results.

Finally, by offering a framework that integrates several spheres beyond the work environment, we aim to raise awareness that interpretations of workplace stress and resilience must account for influences that originate in the broader life domain, which can significantly alter how engineers respond to pressures at work.

This thesis results inform **approaches to foster sustained well-being (RQ2)** in the form of interventions, policies and organisational recommendations. All of these have one common goal: to support the well-being of software engineers. The intervention studies provide empirical evidence that mindfulness-based practices (breathwork, yoga, and journaling) can support stress regulation and emotional balance in software professionals. The results confirm improvements in several areas, while also indicating that the workplace context strongly mediates these outcomes. Important to note that the absence of statistically significant quantitative effects in some intervention settings (Paper D) does not indicate failure, instead points to the sensitivity of well-being outcomes to organisational context, timing, and measurement choices. Individual-level practices are most effective when accompanied by supportive organisational policies and cultures.

Multiple stakeholders can use RQ2 results to inform practice and design decisions. Researchers and companies can apply the findings to design, implement, or adapt interventions to support engineers' well-being. Specifically, we provide recommendations of DOs and DON'Ts for tailoring mindfulness-based interventions to SE contexts, as well as organisational guidelines and policy-level recommendations. Additionally, designers and developers can utilise our findings to inform the design of chatbots with smoother interaction patterns, thereby reducing the triggers of frustration.

Regarding the **methodological strategies for studying human factors (RQ3)**, this research integrates triangulated, multi-modal evidence data (combining psychometrics, physiological data, and qualitative insights) to overcome the limitations of single methods. We made visible that this triangulation, although essential, comes with challenges that must be explicitly acknowledged and analytically addressed. Part of our goal was to report and analyse misalignments across data sources explicitly and to give visibility to null results. Future studies can consider the lessons learned from our experiences and implement experiments that avoid our struggles. Similarly, our results seek to raise awareness among researchers and practitioners of the importance of considering acute and long-term stress when planning studies or interventions.

Moreover, we also proposed an LLM-supported framework for qualitative analysis and created guidelines to improve the transparency, reproducibility, and scalability of the integration. Qualitative researchers can implement the framework and use the rubrics as metrics. The rubrics can even be used alone to guide manual coding and theme development, particularly among new users of thematic analysis.

RQ3 results are particularly useful for making visible the ethical implications of stress (and health-related topics) research, especially when combining intrusive measurements in non-heavy health-related fields, such as SE.

To conclude, with RQ3, we aimed to provide researchers with concrete methodological tools and considerations for studying human factors in SE more rigorously and responsibly.

Future Work

Building on the findings of this research, future work will test and refine the proposed ISTS integrated framework in different contexts. For example, although we attempted to include engineers from several countries, we acknowledge that there is still a need to explore other regions. Future studies will therefore extend data collection to underrepresented regions to examine the robustness and transferability of the framework across different sociotechnical environments and organisational cultures. The application of the framework to design and evaluate assessments and interventions for the organisational environment is also a next step in our studies. Coming studies can also explore the addition of new concepts at the individual level, such as emotional intelligence, need for cognition, and empathy. The social and organisational environment, as well as the technological ecosystem, can be further amplified in future studies by adding new layers and elements that adapt to specific contexts and companies.

Another direction is the investigation of role-specific experiences within software engineering. The studies included in this thesis intentionally adopted a broad view of the SE population. Future work could examine how the framework manifests differently for developers, testers, technical leads, product owners, and engineering managers. These roles are embedded in distinct sociotechnical positions, with different cognitive demands, accountability structures, and exposure to interruptions. Understanding these differences would enable more targeted interventions and refine the framework's sensitivity to organisational role structure.

About the interventions, we aimed to implement more follow-up studies, with a primary focus on exploring diverse mindfulness-based interventions. We also want to recruit larger cohorts and include control groups. For the implemented interventions, we will use follow-up measurements to assess their sustained effects on stress-related outcomes. In addition, future work can explore adaptive interventions that respond to contextual signals, such as workload peaks or project deadlines. Future work could also investigate individual differences as moderators of intervention effectiveness. For example, to study how factors such as prior experience with mindfulness practices, baseline stress regulation capacity, cognitive style, or attitudes toward technology influence outcomes.

Regarding the immediate work pipeline, for our methodology work, we plan to experiment with adapting content analysis using AI and eventually testing more qualitative data analysis methods. We also plan to test our hybrid framework with new interviews, in a way that we analyse them following our guidelines without comparing them to human outcomes. This will give us a good picture of how the framework works when we do not use a human benchmark. Another direction is to study the epistemic risks and boundary conditions of AI assistance in interpretive research.

For data triangulation, we aim to include EEG data in SE tasks and collect it in real-world software development scenarios. Further studies can align biometric data with self-report instruments and recruit larger cohorts. One of the goals of this comparison was to assess how accurately questionnaires (particularly psychometrics) capture the variability in the software engineering population. We argued in this thesis that theories need to be adapted; there is also the possibility that psychometric instruments need to be adapted to the SE population. Currently, we are exploring the collection of blinking data to test its effectiveness as a proxy measure for identifying stress in SE human tasks.