Fostering Pre-Service Teacher Reflection through AI-Based Feedback:

*From Understanding AI Acceptance to Developing Effective AI-Driven Feedback*

Der Philosophischen Fakultät und Fachbereich Theologie
der Friedrich-Alexander-Universität Erlangen-Nürnberg
zur
Erlangung des Doktorgrades Dr. phil.

vorgelegt von
Chengming Zhang

aus Zhejiang
Als Dissertation genehmigt
von der Philosophischen Fakultät und Fachbereich Theologie
der Friedrich-Alexander-Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung: 28.06.2024

Gutachter: Prof. Dr. Michaela Gläser-Zikuda
Prof. Dr. Sebastian Habig
Prof. Dr. Svenja Bedenlier
Acknowledgements

In September 2018, filled with uncertainty for the future and a touch of nervousness, I boarded a flight to Germany. As the plane gently touched down the runway at Düsseldorf Airport, the pre-dawn sky was illuminated by a sunrise, marking the beginning of a long-awaited life journey. Time passed unnoticed, and suddenly, it was the spring of 2024. At this moment, seated in my office in Nürnberg, I put the final period on my doctoral dissertation. Looking through the window at the slowly spreading evening sky, the beautiful scene seemed to whisper that this chapter of my journey was drawing close. In the glow of the sunset, I could see the years of effort and growth.

Completing a doctoral dissertation is not merely a solitary marathon but an exhilarating team football match. Throughout this beautiful game, I received immense support that helped me sprint towards a significant turning point. First and foremost, I extend my deepest gratitude to my advisor, Prof. Dr. Michaela Gläser-Zikuda. She opened the door for me to participate in the PetraKIP project and guided me through this unfamiliar land. As a Chinese student whose command of German was not yet fluent, this opportunity dramatically transformed my life trajectory. Under her guidance, I took my initial steps at an international conference, published my first paper, and eventually completed my doctoral degree—each phase enriched with her support. Furthermore, I am deeply grateful to Prof. Dr. Sebastian Habig and Prof. Dr. Svenja Bedenlier for their guidance and support during the final stage of my doctoral journey.

Additionally, I owe special thanks to my PetraKIP project team members, Dr. Florian Hofmann, Jessica Schießl, and Lea Plößl. Their expertise and selfless assistance in experimental design, data analysis, and dissertation writing proved invaluable throughout my research journey. I also appreciate the Free University of Berlin professors, Prof. Dr. Ralf Romeike, Prof. Dr. Christoph Benzmüller, Prof. Dr. Tim Landgraf, and Veronika Solopova. They introduced
me to computer science’s mysteries. They ignited my passion for unfamiliar fields such as artificial intelligence, machine learning, and neural networks.

Without a doubt, I am immensely grateful to all the members of our chair, especially Lisa Pösse, Ulrike von Urbanowicz, and Eva Weigand, for their assistance in implementing the mini-portfolio; without their expertise and enthusiasm, our project could not have been executed so smoothly. I also express my gratitude to my friends—Farrukh Kamran, Meltem Doganay, Yanfen Chen, Kailun Sha, Tianjing Qu, Longyu Chen, and Dr. Xining Wang. Thank you for your constant support and companionship over the past three years. You have been my unbreakable pillars on the path of scientific research, facing every challenge and obstacle.

Finally, I must express my deepest gratitude to my parents and family. They have not only fully supported my journey to study abroad in Germany but have also been the strongest supporters in my life’s journey. Behind every success lies their silent dedication; with every setback, they are my steadfast comfort. They share in every joy with me and bear every pain alongside me. On my academic path, through ups and downs, they have always been my most reliable haven.
Abstract

Learning from experiences is crucial for pre-service teacher to become professional teacher. As a bridge between theory and practice, reflection contributes to their professional development. However, reflection often requires exploring much tacit knowledge, which needs guidance from professionals. Indeed, many pre-service teachers do not receive adequate support due to lack of time for teacher educators and large student populations. Recently, advancements in artificial intelligence (AI) have led to the development of automated feedback systems. These systems provide prompt and relevant feedback to pre-service teachers, enhancing opportunities for professional development.

In the context of digital reform, this dissertation aims to explore the role of AI in fostering pre-service teacher reflection. It seeks to investigate the acceptance of AI within teacher education, enhance deep understanding of reflective writing, and develop suitable AI-driven feedback mechanisms. More specifically, the research objectives are threefold: (i) to investigate pre-service teachers’ AI acceptance and to determine whether gender plays a moderating role; (ii) to assess the quality of reflective writing and explore its vital predictive indicators; (iii) to develop an AI feedback system specifically designed to enhance the quality of pre-service teachers’ reflection. Through these objectives, the research seeks to provide innovative technological support for the professional development of pre-service teachers.

Employing a diverse array of research methods and incorporating multiple data types, this thesis systematically synthesizes and presents the findings from three publications. The results are structured to highlight the contributions of each study, providing a overview of the research outcomes. In the discussion section, I delve into the factors influencing AI acceptance and examine how gender moderates this process. Additionally, I also discuss the transparency and explainability of AI feedback algorithms. Based on these discussions, this thesis proposes a series of recommendations for improving future educational research.
Zusammenfassung


# Table of Contents

List of Figures .......................................................................................................................... III

List of Tables ............................................................................................................................ IV

List of Abbreviations ................................................................................................................. V

1. Introduction ............................................................................................................................ 1
   1.1 Motivation ......................................................................................................................... 1
   1.2 Research Objectives ......................................................................................................... 5
   1.3 Structure of the Dissertation ............................................................................................ 7

2. Literature Review .................................................................................................................. 9
   2.1 Reflective Writing in Teacher Education ........................................................................ 9
       2.1.1 Definition of Reflection ......................................................................................... 9
       2.1.2 Theoretical Models for Reflection ......................................................................... 11
       2.1.3 Impact of Reflective Writing on the Professional Development ......................... 14
   2.2 Integrating Artificial Intelligence into Teacher Education ........................................... 17
       2.2.1 Promises and Challenges of AI for Future Teachers ............................................ 17
       2.2.2 Acceptance of Artificial Intelligence Among Pre-Service Teachers ................. 19
   2.3 Feedback for Reflective Writing ..................................................................................... 21
       2.3.1 Theoretical Model for Effective Feedback Delivery ........................................... 21
       2.3.2 Artificial Intelligence Feedback Systems for Reflective Writing ....................... 23
   2.4 Mixed Methods in Research on Reflective Writing ....................................................... 26
       2.4.1 Mixed Methods in Educational Research ............................................................ 26
       2.4.2 Potential Strengths of Mixed Methods on Reflective Writings .......................... 27

3. Empirical Studies .................................................................................................................. 30
   3.1 PetraKIP Project ............................................................................................................. 30
   3.2 Research Design ............................................................................................................. 30
3.3 Summary of Research Results ...................................................................................... 31

3.3.1 Study 1 .................................................................................................................. 31
3.3.2 Study 2 .................................................................................................................. 32
3.3.3 Study 3 .................................................................................................................. 33

4. Discussion ..................................................................................................................... 35

4.1 Integration of Results ................................................................................................. 35
4.1.1 Determinants of AI Acceptance ........................................................................... 35
4.1.2 Design AI-based Automated Feedback For Reflection ......................................... 37
4.2 Practical Implications ................................................................................................. 40
4.3 Limitations and Future Work ..................................................................................... 43

5. Conclusion .................................................................................................................... 46

References ....................................................................................................................... 48

Appendix I: AI Acceptance Scale ................................................................................... VIII
Appendix II: List of All Publications ................................................................................. X
Appendix III: List of Publications for Dissertation ........................................................ XI

Publication P1: AI Acceptance Among Pre-Service Teachers ........................................ XI
Publication P2: Mixed Methods for Analyzing Reflective Writing ................................ XIII
Publication P3: Developing AI-Driven Feedback ............................................................. XIV
List of Figures

Figure 1 Summary of research objectives ................................................................. 7
Figure 2 Structure of the thesis ............................................................................... 8
Figure 3 Gibbs’ reflective cycle .............................................................................. 13
Figure 4 COACTIV model of professional competence ........................................... 15
Figure 5 Phases and subprocesses of self-regulation .............................................. 16
Figure 6 Mini-portfolio design and implementation ............................................... 30
List of Tables

Table 1 Depth model for reflection................................................................. 12
Table 2 Research summary for study 1 ............................................................ 32
Table 3 Research summary for study 2 ............................................................ 33
Table 4 Research summary for study 3 ............................................................ 34
Table 5 AI Acceptance Scale (German version).............................................. VIII
Table 6 Summary of publication P1 ................................................................. XII
Table 7 Summary of publication P2 ................................................................. XIII
Table 8 Summary of publication P3 ................................................................. XV
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>AIA</td>
<td>Artificial Intelligence Anxiety</td>
</tr>
<tr>
<td>AISE</td>
<td>Artificial Intelligence Self-Efficacy</td>
</tr>
<tr>
<td>ATT</td>
<td>Attitude Toward Using</td>
</tr>
<tr>
<td>AWA</td>
<td>Analytical Writing Assessment</td>
</tr>
<tr>
<td>BERT</td>
<td>Bidirectional Encoder Representations from Transformers</td>
</tr>
<tr>
<td>BI</td>
<td>Behavioural Intention</td>
</tr>
<tr>
<td>BOW</td>
<td>Bag-of-Words</td>
</tr>
<tr>
<td>DL</td>
<td>Deep Learning</td>
</tr>
<tr>
<td>GPT</td>
<td>Generative Pre-trained Transformer</td>
</tr>
<tr>
<td>ICT</td>
<td>Information and Communication Technology</td>
</tr>
<tr>
<td>JR</td>
<td>Job Relevance</td>
</tr>
<tr>
<td>LIWC2015</td>
<td>Linguistic Inquiry and Word Count 2015</td>
</tr>
<tr>
<td>LLMs</td>
<td>Large Language Models</td>
</tr>
<tr>
<td>MGA</td>
<td>Multi-Group Analysis</td>
</tr>
<tr>
<td>ML</td>
<td>Machine Learning</td>
</tr>
<tr>
<td>NLP</td>
<td>Natural Language Processing</td>
</tr>
<tr>
<td>PE</td>
<td>Perceived Enjoyment</td>
</tr>
<tr>
<td>PetraKIP</td>
<td>Persönliches transparentes KI-basiertes Portfolio für die Lehrerbildung</td>
</tr>
<tr>
<td>PEOU</td>
<td>Perceived Ease of Use</td>
</tr>
<tr>
<td>POS</td>
<td>Part-of-Speech</td>
</tr>
<tr>
<td>PU</td>
<td>Perceived Usefulness</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>RAG</td>
<td>Retrieval-Augmented Generation</td>
</tr>
<tr>
<td>SEM</td>
<td>Structural Equation Modeling</td>
</tr>
<tr>
<td>SN</td>
<td>Social Norms</td>
</tr>
<tr>
<td>TAM</td>
<td>Technology Acceptance Model</td>
</tr>
<tr>
<td>TAM2</td>
<td>Technology Acceptance Model 2</td>
</tr>
<tr>
<td>TAM3</td>
<td>Technology Acceptance Model 3</td>
</tr>
<tr>
<td>TF-IDF</td>
<td>Term Frequency-Inverse Document Frequency</td>
</tr>
<tr>
<td>TPB</td>
<td>Theory of Planned Behaviour</td>
</tr>
<tr>
<td>TRA</td>
<td>Theory of Reasoned Action</td>
</tr>
<tr>
<td>UTAUT</td>
<td>Unified Theory of Acceptance and Use of Technology</td>
</tr>
<tr>
<td>XIP</td>
<td>Xerox Incremental Parser</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 Motivation

Reflection is highly valued across various disciplines, leading to the development of several related concepts, including reflective writing, reflexivity, critical reflection, and critical thinking (e.g., van Beveren et al., 2018, p.2). This emphasis on reflection is particularly evident in the health professions, such as nursing, midwifery, dentistry, and broader medical education (Sandars, 2009; Winkel et al., 2017). Moreover, the relevance of reflection extends to the humanities and social sciences, impacting fields like teacher education, social work, the arts, and beyond (Beauchamp, 2015; Fook et al., 2006). Recognized as an essential educational strategy for professional development, incorporating reflective writing in higher education is notably widespread (Chan & Lee, 2021; Ryan, 2013). This widespread adoption underscores the critical role of reflection in enhancing an understanding of professional knowledge (Korthagen et al., 2001), fostering metacognitive skills (Desautel, 2009), and promoting self-awareness (van Beveren et al., 2018). By integrating reflective writing into curricula, educators aim to equip students with the critical thinking and analytical skills essential for lifelong learning and adaptability within their respective disciplines.

In teacher education, the development of pre-service teachers into reflective practitioners is recognized as crucial for their continuous personal and professional development (Baumert & Kunter, 2013; Carlson et al., 2019; Darling-Hammond & Bransford, 2005). Reflection is a pivotal bridge between theoretical knowledge and practical application, enabling the seamless integration of abstract academic concepts with concrete teaching practices (Leonhard & Rihm, 2011). This integration is especially critical for novice and pre-service teachers, who frequently experience confusion and helplessness when navigating the complexities of actual teaching scenarios (Nguyen et al., 2014). Through consistent and deliberate reflection, they are
empowered to critically evaluate their teaching approaches, identifying strengths and areas for improvement (Rogers, 2001). Over time, this reflective writing can lead to the development of a tailored set of teaching philosophies and strategies, thereby enhancing the overall quality of their instruction (Berndt et al., 2017; Kultusministerkonferenz der Länder, 2019).

In the past few decades, reflective writing have been adopted in teacher education to facilitate pre-service teachers’ reflection (Cohen-Sayag & Fischl, 2012). Reflective writing include both written and oral formats. On the one hand, written tools such as portfolios (Gläser-Zikuda, 2012, 2015), reflective practicum reports, journals, and one-minute papers encourage the practice of written reflection; on the other hand, oral formats, including reflective dialogues, self-reports with recordings, and fishbowls, offer alternative avenues for reflective writing (e.g., Christof et al., 2018). As technological advancements have progressed, the tools for reflection have evolved from paper-based to web-based and, more recently, to mobile-based applications (e.g., Trent & Shroff, 2013; Xerri & Campbell, 2016). However, improving pre-service teachers’ reflective skills through reflective writing presents inherent limitations. Research conducted by Azimi et al. (2019) and Körkkö et al. (2016) has shown that students’ reflections often remain descriptive rather than evolving into critical analyses. Hence, based on reflective writing, it is necessary to rely on more instructional strategies to improve pre-service teachers’ reflection.

Given the above mentioned challenge, an instructional approach to initiating and supporting reflection for professional development involves using prompts (Hume, 2009; Shulman & Shulman, 2009). Prompts serve as a form of scaffolding (Pea, 2004), offering a structured means to support reflection by guiding individuals through the process with specific guideline questions (Lee, 2005). The effectiveness of prompting as a method to stimulate reflection across various disciplines has been demonstrated in prior research (e.g., Sobral, 2000; Van den Boom et al., 2004). Hume (2009) developed scaffolds informed by Shulman’s (1987) classification framework, with findings indicating that pre-service teachers’ reflective writing skills and
pedagogical content knowledge significantly improved when these scaffolds were utilized to assist their writing. However, Hume’s research (2009) also demonstrated a critical caveat: if these prompts were discontinued, the reflective skills among pre-service teachers would begin to decline once more.

To address this concern, the researchers proposed giving students consistent, timely, and constructive feedback for reflective writing (e.g., Quinton & Smallbone, 2010; Sargeant et al., 2009). However, the barrier to providing feedback on reflective writing is costly time and labor (Ullmann, 2015). Hence, many automated feedback systems for reflective writing have emerged in higher education in recent years. These systems employed a variety of techniques based on Natural Language Processing (NLP), ranging from rule-based approaches to Machine Learning (ML) and, more recently, Large Language Models (LLMs). Rule-based systems operate on a predefined criteria foundation to appraise a text’s reflective elements. For instance, pioneering research has utilized lexicon-based rules from resources like Linguistic Inquiry and Word Count 2015 (LIWC2015) (Liu et al., 2021), Analytical Writing Assessment (AWA) (Hanlon et al., 2021), and Xerox Incremental Parser (XIP) (Gibson et al., 2017) to analyze students’ reflective writing. With the progression of ML techniques, a faction of scholars has adopted supervised (Ullmann et al., 2019) and unsupervised (Chen et al., 2016) ML approaches to assess reflective writing, which also encapsulates deep learning (DL) (Nehyba & Štefánik, 2023). In the current landscape, models such as GPT (Generative Pre-trained Transformer) and BERT (Bidirectional Encoder Representations from Transformers) exhibit the capacity for language comprehension and generation. For instance, a recent study by Wulff et al. (2023) demonstrates the utilization of the BERT model in automating the classification of pre-service physics teachers’ reflective writing.

The emerging research on applying AI feedback systems in reflective writing highlights its nascent yet promising potential for educational research. In teacher education, deploying AI
feedback introduces challenges and possibilities that necessitate further investigation. Firstly, it is necessary to explore AI acceptance among pre-service teachers and identify the factors influencing their acceptance of these technologies (Pedró et al., 2019; Reiss, 2021). The previous research indicates that teachers frequently lack motivation to integrate Information and Communication Technology (ICT) and AI in educational settings (Collie & Martin, 2024). This presents a significant challenge in incorporating AI into teacher education programs and school curricula. Understanding their attitudes towards AI is essential, as it significantly influences the seamless integration of these technologies into educational frameworks and determines how AI can effectively fulfill its expected roles. Secondly, reflective writing is a complex and multifaceted process requiring high cognitive engagement and self-awareness (Sudirman et al., 2021). In the context of school environments and the interpretation of teaching experiences, it often lacks clear-cut “right” answers (e.g., Crespo, 2000). This inherent ambiguity makes it particularly challenging to identify and delineate the criteria for high-quality reflective writing. The subjective nature of reflection means that assessing its quality is not straightforward and requires a nuanced understanding of the various dimensions of reflective writing through different research methods. Lastly, while AI feedback on reflective writing in teacher education is still relatively limited, this area presents substantial potential for innovation and growth. There are two primary challenges currently confronting the field of research in the reflective writing feedback system. The first challenge relates to the datasets used to evaluate reflective writing. Many studies utilize datasets gathered through non-standardized methods, thereby compromising data quality and reliability. The second challenge involves methodological limitations, where most studies predominantly focus on analyzing single sentences, neglecting the assessment of entire documents. This approach is particularly problematic in reflective writing, which necessitates an evaluation of the overall coherence and the logical interconnection between narratives (Moon, 2013). These challenges underscore the
urgent need for future research to devise more sophisticated methodologies that enhance data quality and deepen the integrative analysis of such texts. A well-designed, transparent, and user-friendly AI feedback system will significantly support the professional development of pre-service teachers, enhancing their reflective skills and, by extension, their teaching effectiveness (Escalante et al., 2023; Leite & Blanco, 2020). In summary, this thesis explores pre-service teachers’ attitudes toward AI technology, addresses the complexities inherent in reflective writing, and advocates for developing effective AI feedback systems tailored to teacher education.

1.2 Research Objectives

The integration of AI feedback into a teacher education program aims to prepare pre-service teachers to become reflective practitioners. To realize this ambition, the dissertation outlines three research objectives, which are systematically organized within a structured framework (see. Figure 1). This framework dissects the aims into three distinct and more focused research objectives (RO1-RO3). Furthermore, to attain these objectives, a variety of well-established research methodologies have been employed, ensuring a robust and comprehensive approach to exploring the integration of AI feedback in reflective writing.

The first research objective (RO1) focused on AI acceptance by pre-service teachers and identifying the factors that influence this acceptance. To address this, the study revised a model based on the Technology Acceptance Model 3 (TAM3) (Venkatesh & Bala, 2008) to analyze pre-service teachers’ AI acceptance. The research utilized structural equation modeling (SEM) to investigate how various factors affect pre-service teachers’ behavioral intentions toward AI and their actual usage behaviors. Moreover, a multi-group analysis (MGA) was conducted to explore the nuances of AI acceptance further, specifically examining the impact of gender differences on AI acceptance.
The second research objective (RO2) focused on exploring and assessing quality indicators in reflective writing utilizing mixed methods research. This approach combined qualitative content analysis, computational linguistics analysis, and unsupervised ML. The first step was coding based on the reflective level framework proposed by Hatton and Smith (1995). This framework provided an approach to identify and categorize different levels of reflection in reflective writing, ranging from descriptive writing to critical reflection. Second, this study utilized LIWC2015 (Pennebaker et al., 2015) to extract linguistic features in reflective writing, including affective expressions, cognitive processes, social relations, and other aspects. Meanwhile, by utilizing BERTopic (Grootendorst, 2022), the study could automatically identify and extract themes from the text data. Finally, combining the reflective levels with the linguistic features and themes extracted through LIWC2015 and BERTopic further explored how these indicators work together to predict the quality of reflective writing.

The final research objective (RO3) was to develop an AI feedback system for reflective writing. In order to summatively evaluate the level of reflective writing, a variety of shallow ML algorithms, as well as different linguistic representations, including LIWC2015, Bag-of-Words (BOW), and Frequency-Inverse Document Frequency (TF-IDF) techniques, were employed. These methods provided basic lexical and structural information for text analysis. Further, the study introduced state-of-the-art pre-trained language models such as BERT (Devlin et al., 2018), RoBERTa (Liu et al., 2019), BigBird (Zaheer et al., 2020), and Longformer (Beltagy et al., 2020), which utilize deep learning techniques to capture complex linguistic features and contextual relationships in the text to improve the accuracy and depth of analysis. Ultimately, by comparing the performance of these different models in assessing reflective writing levels, the study aimed to identify the most effective methodology to provide a scientific basis for the quality assessment of reflective writing. This mixed methods approach
not only improves the accuracy of the assessment but also enriches the understanding of the depth and quality of reflective writing.

<table>
<thead>
<tr>
<th>Research Objectives</th>
<th>Research Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>To investigate the key factors influencing pre-service teachers’ AI acceptance</td>
<td>Structural Equation Modeling, Measurement Invariance &amp; Multi-group Analysis</td>
</tr>
<tr>
<td>To design AI-based feedback for reflective writing</td>
<td>Qualitative Content Analysis, Computational Linguistics &amp; RQ Kantop</td>
</tr>
</tbody>
</table>

![Figure 1](image1.png)  
**Figure 1** Summary of research objectives (modified from Zschech, 2020, p.7)

### 1.3 Structure of the Dissertation

To achieve the stated research objectives, the remainder of this thesis is structured into four main chapters. **Chapter 2** is a literature review that describes the concept of reflection, including its definition, theoretical models, and the significant role of reflection in promoting the professional development of pre-service teachers. In addition, this chapter discusses the application of AI in education, and the current state of pre-service teachers’ AI acceptance. Moreover, the review covers the summary of the feedback models, exploring the critical impact of effective feedback in developing pre-service teachers’ reflections, and how AI feedback has been developed and utilized to support research on reflective writing. Finally, this chapter briefly describes the potential benefits of mixed methods approaches in educational research and current analytical methodologies for reflective writing. In **Chapter 3**, the thesis delves into three empirical studies. The experimental context, participant characteristics, research methods employed, and significant findings are described in detail for each study. **Chapter 4** focuses on an in-depth discussion of the empirical studies’ findings. The implications and significance of these results in the field of teacher education and AI applications are analyzed here, and the implications of these findings for future research directions are also discussed, as well as the
limitations encountered during the research study. Finally, Chapter 5 summarizes the entire thesis. Figure 2 demonstrates the structure of the thesis.
2. Literature Review

2.1 Reflective Writing in Teacher Education

2.1.1 Definition of Reflection

The concept of reflection, initially rooted in optics, has been adopted into philosophical discourse later. Owing to its origin in optics, the term “reflection” frequently evokes the metaphor of a mirror, capturing the notion of an individual’s mirrored image (Zahn, 2007). John Dewey’s 1933 publication, “How We Think,” is celebrated as a seminal reflection work, amalgamating numerous forerunners’ insights. He pioneered the conceptualization of reflection as a cognitive process, describing it as “a deliberate, enduring, and meticulous evaluation of any given belief or presumed piece of knowledge, in consideration of the evidence that underpins it and the subsequent inferences it implies” (Dewey, 1933, p.9). Deeply influenced by John Dewey’s theories, Donald Schön underscored the importance of reflective writing in his seminal works, “The Reflective Practitioner” (1983) and “Educating the Reflective Practitioner” (1987). Schön introduced two pivotal types of reflective writing: reflection-in-action, which occurs concurrently with the action, and reflection-on-action, which happens after the event. These distinctions highlight the dynamic nature of reflection, emphasizing its role both during and after professional activities, thereby offering a nuanced understanding of how professionals can engage in continuous learning and improvement. However, reflection is complicated because it is heavily influenced by different philosophical concepts and motivations, making a precise definition of reflection complex (Akbari et al., 2010). Moreover, considerable differences exist among theorists, researchers, and teacher educators concerning the exact interpretation or application of reflection (e.g., van Beveren et al., 2018, p.2). Introducing new concepts like critical reflection, reflexivity, and critical thinking has further contributed to the ambiguity surrounding reflection within individual disciplines and across
them (Ecclestone, 1996; Fendler, 2003; Hatton & Smith, 1995). For instance, other scholars have argued that reflection is not just a cognitive process but also contains critical, psychological, and affective components (Cui et al., 2019). Mezirow (1991) articulated reflection as “the process of critically assessing the content, process, or premise(s) of our efforts to interpret and give meaning to an experience” (p.104). This definition emphasizes critically evaluating one’s interpretative efforts across various dimensions. Boyd and Fales (1983, p.99) offered a complementary perspective, describing reflection as “the process of internally examining and exploring an issue of concern, triggered by an experience, which creates and clarifies meaning in terms of self, and results in a changed conceptual perspective.” This highlights the reflective journey prompted by experiences that shape and redefine one’s understanding of self. Further enriching the discourse, Boud et al. (2013) acknowledged the significance of the affective component of reflection, noting its potential to precipitate specific experiences. Due to the complexity of the definition of reflection, Nguyen et al. (2012) undertook a systematic review and defined reflection as “the process of engaging the self in attentive, critical, exploratory, and iterative interactions with one’s thoughts and actions, and their underlying conceptual frame, with a view to changing them and with a view on the change itself” (p.48). This definition of reflection reflects a multidimensional understanding that emphasizes the individual’s active participation in the reflective process. It reveals that reflection is not merely a review of past behavior but a process that involves self-observation, critical analysis, and continuous iteration. Meanwhile, this definition highlights reflection’s dynamic and constructed nature, pointing to the possibility of personal and professional development through reflective writing.
2.1.2 Theoretical Models for Reflection

Expanding upon the foundational definition, scholars have developed various reflective models designed to elucidate the complexities inherent in the reflective process (Calderhead, 1992; Gore & Zeichner, 1991; Hatton & Smith, 1995). These models incorporate a variety of approaches to both facilitate and assess reflection, utilizing tools such as frameworks (Ward & McCotter, 2004), rubrics (Miller-Kuhlmann et al., 2016), and coding schemas (Poldner et al., 2014). Through providing structured guidelines, these models not only augment the effectiveness of individual reflection but also furnish educators and researchers with methodologies for measuring and scrutinizing the quality of reflective writing. Reflective models are categorized based on their primary focus into depth or breadth models (e.g., Ullmann, 2019). Depth models concentrate on the qualitative aspects of reflection, emphasizing profound thinking and personal insight. In contrast, breadth models employ a multidimensional and process-oriented analytical perspective, prioritizing the scope and diversity of reflection. These models highlight the importance of considering a broad spectrum of dimensions within the reflective process, thereby fostering a comprehensive engagement with various aspects of experiences. Each model type offers distinct advantages for reflective writing: depth models enhance the quality and profundity of introspective insights. In contrast, breadth models extend the range of considerations and perspectives explored during reflection. Collectively, they establish a comprehensive framework for reflection, accommodating varied needs and goals within reflective writing and facilitating a richer, more nuanced understanding and application of reflection across different contexts.

In response to the depth model of reflection, scholars such as Ip et al. (2012), Kember (1999), and Mezirow (1991) have outlined a range of reflections varying from non-reflective to highly reflective levels. Among these, the reflective level model introduced by Hatton and Smith in 1995 has gained widespread recognition as a seminal depth model in teacher education.
They identified four distinct levels of reflection: starting with descriptive writing, which encompasses merely descriptive notes; progressing to descriptive reflection, where justifications are offered either from the author’s perspective or through reference to pertinent literature; advancing to dialogic reflection, characterized by the cultivation of an internal dialogue to explore various possibilities; and culminating in critical reflection, which considers the broader historical, social, and political contexts, evaluating the ramifications of individual actions and informing decision-making accordingly. Table 1 shows a specific description of this model.

Table 1 Depth model for reflection (cf. Hatton & Smith, 1995, pp.48-49; Fütterer, 2019, p.250)

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0:Descriptive Writing</td>
<td>A situation (action, behavior...) is described. No efforts to classify or explain exist. Because reflective processes were defined as metacognitive, a mere description does not represent a reflective process.</td>
</tr>
<tr>
<td>1:Descriptive Reflection</td>
<td>Situations are either justified (personal judgment, perspective), or feelings, optional perspectives, or influential variables are reported, but without connecting them or considering their contextual embedding. Personal assumptions are presented.</td>
</tr>
<tr>
<td>2:Dialogic Reflection</td>
<td>Different perspectives, influencing factors, and justifications for situations are identified. Perspectives are weighed in an intra-personal dialogue. For this to happen, subjective theories and beliefs must become conscious. Competing perspectives are weighed up, leading to judgment.</td>
</tr>
<tr>
<td>3:Critical Reflection</td>
<td>It is recognized that both situations and the identified perspectives, influential factors, and rationales are embedded in and influenced by a broader context (including historical, social, and political). Values and norms of the professions goals are also challenged, and institutional expectations are included.</td>
</tr>
</tbody>
</table>
The breadth model’s application is illustrated through the research contributions of scholars such as Kolb (1984), Mansvelder-Longayroux et al. (2007), and Poldner et al. (2014). Among the most emblematic breadth models is the Reflective Cycle developed by Gibbs (1988) (see Figure 3), which outlines six stages to provide a structured approach for composing and analyzing reflective writings. These stages include: (1) **Description**: offering a concise summary of the events that transpired; (2) **Feelings**: expressing personal feelings and emotional responses to the events; (3) **Evaluation**: critiquing the positive and negative aspects of those responses; (4) **Analysis**: delving into the understanding of the event and its implications; (5) **Conclusion**: drawing general or specific conclusions from the analysis; (6) **Action Plan**: devising a strategy for addressing similar situations in the future. This cyclical framework enables a comprehensive exploration of experiences from multiple dimensions, encouraging a thorough and nuanced reflection. The Reflective Cycle by Gibbs facilitates a systematic dissection of events. It promotes a proactive stance towards personal and professional growth by encouraging individuals to contemplate and plan future actions based on reflective insights.

Figure 3 Gibbs’ reflective cycle (Gibbs, 1988, p.50)
2.1.3 Impact of Reflective Writing on the Professional Development

Teacher professional development means *how they learn to learn and how they apply their knowledge in practice to support pupil learning* (Avalos, 2011, p.11). The COACTIV model provides a framework for teachers’ professional competencies in German-speaking countries (Baumert & Kunter, 2013). This model delineates these competencies into several pivotal domains: *professional knowledge, beliefs and values, self-regulation,* and *motivation.* Figure 4 describes the specific elements of COACTIV. From this model, it is clear that self-regulation is considered one of the critical components of teachers’ professional competence. Furthermore, Zimmerman (2002) categorizes the self-regulated learning process into three interconnected phases: *the forethought phase, the performance phase,* and *the self-reflection phase.* Within this theory, reflection emerges as a fundamental element, enabling learners to derive insights from their experiences and fine-tune their learning strategies, optimizing their learning outcomes. Figure 5 shows the self-regulated learning model. In summary, self-regulation is critical to teachers’ professional competence, as delineated in the COACTIV model. Zimmerman’s self-regulation model further elucidates that reflection is vital in fostering teachers’ ongoing development. This study posits that enhancing teachers’ reflective capacity can significantly contribute to their professional development. This assertion is grounded in the theories above, which highlight the value of reflection in the educational domain. Through reflective writing, teachers can critically assess and rectify deficiencies in their pedagogical strategies, thereby elevating the overall quality of their instruction. In addition, this model’s applicability extends beyond in-service teachers to encompass pre-service teachers (Krauss et al., 2017), suggesting its broad relevance in education.
Reflective writing is inherently subjective and is usually done in the first person. During this process, the writer not only records how events occurred but delves into the impact of these experiences on him or her, the lessons learned, and how these experiences will shape his or her future thinking and actions (e.g., Jasper, 2005). Therefore, reflective writing is not only a way to achieve learning (or conceptual change) through the development of thinking but also a process of shaping and understanding knowledge (Usher et al., 1999). Through this in-depth personal exploration, authors can reconceptualize their behaviors and beliefs, promoting continued professional and academic self-development. The significance of reflective writing in advancing professional development in teacher education is well-documented across various empirical studies (Avalos, 2011). For instance, Körkkö et al. (2016) analyzed the portfolios of 13 pre-service teachers and determined that supporting these individuals’ reflective skills enhanced their professional development. This finding highlights the criticality of developing reflective skills within teacher education programs. Moreover, Poldner (2014) thoroughly examined 34 pre-service teachers’ reflective writings over two semesters, focusing on the depth
and thematic content of their argumentation. The study concluded that prompting students to integrate argumentative content, engage in dialogic exchanges, and embrace transformative learning within their reflective writing was pivotal for their professional advancement. Similarly, Cohen-Sayag and Fischl (2012) explored the development of pre-service teachers’ reflective writing ability and its correlation with their teaching effectiveness. They discovered that, in the first semester, the student’s success in practical teaching assignments correlated positively with their ability to provide descriptive and comparative reflections. By the second semester, successful teaching experiences were only associated with critical reflections, indicating that students who achieved this reflection level could improve their teaching practices. These studies collectively highlight the vital role of reflective writing in teacher education, showing its substantial influence on the professional development of educators. Such evidence provides crucial insights for educational policymakers and teacher educators, emphasizing the value of incorporating reflective writing into teacher education curricula to foster more effective and reflective educators.

Figure 5 Phases and subprocesses of self-regulation (Zimmerman, 2002, p.67)
2.2 Integrating Artificial Intelligence into Teacher Education

2.2.1 Promises and Challenges of AI for Future Teachers

AI is driving significant advancements in the digitization of the education sector. These advancements manifest in various forms, including analyzing student behavior, predictive assessments, automated grading, and providing personalized learning (Zawacki-Richter et al., 2019). More recently, generative AI technologies such as ChatGPT, Midjourney, and DALL-E have been widely integrated into educational settings. These technologies foster innovative approaches to personalized teaching and learning and signify a significant shift in educational methodologies (Baidoo-Anu & Ansah, 2023; Lo, 2023). The deployment of these tools has demonstrated their effectiveness in handling complex educational tasks. However, this integration also presents several challenges.

The integration of AI in education offers numerous advantages, significantly enhancing student learning outcomes, administrative efficiency, and the development of teachers’ competencies. For instance, AI-based educational technologies analyze vast student data, including test scores, homework submissions, and interaction records. This enables educators to pinpoint students’ strengths and weaknesses across various subjects, facilitating targeted instructional support. DreamBox Learning (https://www.dreambox.com/), an AI-driven adaptive platform specializing in mathematics, tailors its content and difficulty according to each student’s pace and ability. When students struggle with specific mathematical concepts, such as addition or subtraction, the platform automatically presents relevant exercises and resources to address these challenges. Moreover, the application of AI extends beyond cognitive aspects to include the analysis of multimodal data, yielding deeper insights into non-cognitive factors like learners’ metacognition, emotions, and motivation (e.g., Molenaar et al., 2023). Regarding educational management, AI dramatically reduces teachers’ workload by offering
real-time feedback capabilities (e.g., Chen et al., 2020). For example, Duolingo (https://www.duolingo.com/), a language learning application, leverages AI to assess students’ exercises instantaneously, promptly identifying mistakes and suggesting corrections. Furthermore, AI-enhanced tools and platforms equip teachers with extensive educational resources and innovative learning methodologies. Nearpod (https://nearpod.com/), an interactive learning platform incorporating AI, significantly improves classroom interactions and learning experiences. It enables educators to design, distribute, and manage course content while boosting student engagement through dynamic engagement and feedback mechanisms.

In conclusion, AI technology streamlines the teaching and learning processes and opens new avenues for educators and learners alike, underscoring its transformative potential in reshaping the educational landscape.

However, recent research indicates that despite the widespread adoption of technology, many teachers maintain a low acceptance of new technological tools, potentially obstructing the effective implementation of AI (Istenic et al., 2021; Kaban & Ergul, 2020). One primary concern is the technical knowledge and competence required to utilize AI effectively. For instance, some educators have voiced apprehensions about integrating such technologies. As noted by Wang & Cheng (2021), a teacher expressed: “I’m neither an old-school teacher nor a techsavvy one. But my experience suggests new technologies could bring many unknowns or difficulties. As I haven’t learned enough about AI, I won’t claim to have full confidence in the AI initiative. I won’t know if the requirement would exceed my capacity, but my overall feeling is that using AI to teach is too early for massive adoption” (I–C2-TS-1) (p.8). This reluctance is driven partly by anxiety regarding new educational technologies (Zimmerman, 2006) and partly by resistance to deviating from traditional teaching methods (Tallvid, 2016). Insufficient technological knowledge might lead teachers to question the pedagogical benefits of such innovations. However, more recent studies, such as those by Nazaretsky et al. (2022),
demonstrate that transparent explanations of AI’s decision-making processes and clear communication of its benefits can mitigate teachers’ concerns and enhance their confidence in using AI technologies. Therefore, it is crucial to provide continuous professional development and training to empower educators with the necessary skills to leverage AI effectively in their teaching practices. Furthermore, potential bias in the design and implementation of AI systems is a significant concern in educational settings (Baker & Hawn, 2022). When AI is utilized for student assessments, teachers must be vigilant about possible gender or racial biases that could lead to unfair results. It is essential for educators to recognize these biases and implement measures to mitigate their effects, thereby ensuring fair and inclusive education (Akgun & Greenhow, 2022). Additionally, the increasing reliance on AI can alter traditional teacher-student relationships (e.g., Alam & Mohanty, 2022; Chiu et al., 2023). The use of technology may decrease face-to-face interactions, potentially impacting the teacher’s role as a mentor. To counterbalance this, teachers should maintain direct contact with their students through regular videoconferences or real-time interactive sessions.

2.2.2 Acceptance of Artificial Intelligence Among Pre-Service Teachers

In the past few years, with a marked acceleration in the last two, the education sector has witnessed a substantial transformation driven by generative AI, notably the GPT series from OpenAI (Baidoo-Anu & Ansah, 2023; Mhlanga, 2023). These advanced LLMs offer innovative opportunities for the educational community, reshaping traditional pedagogical approaches and learning environments (Fuchs, 2023; Kasneci et al., 2023). For instance, applications of generative AI employing LLMs enhance the assessment process and provide customized guidance tailored to individual student needs, significantly augmenting the learning experience (Hooda, 2022). However, integrating AI technology in education sparks diverse opinions among educators regarding its role in classrooms. The sector faces ongoing exploration and
challenges in effectively leveraging AI technologies to enrich the learning experience (Chan, 2023; Pedró et al., 2019). Hence, to navigate the increasing adoption of AI-driven educational tools in the future, it is essential to comprehend the acceptance of AI among pre-service teachers and identify the factors influencing this acceptance.

The Technology Acceptance Model (TAM) is extensively acknowledged and employed in academic research, particularly for analyzing the adoption of educational technology. This model has proven instrumental in assessing the acceptance and adoption behaviors of both pre-service (Baydas & Goktas, 2017; Koutromanos et al., 2015) and in-service (Choi et al., 2022) teachers towards AI technologies. TAM, integrating essential aspects of the Theory of Reasoned Action (TRA) (Ajzen & Fishbein, 1980) and the Theory of Planned Behaviour (TPB) (Ajzen, 1985), has evolved through subsequent versions like Technology Acceptance Model 2 (TAM2) (Venkatesh & Davis, 2000) and Technology Acceptance Model 3 (TAM3) (Venkatesh & Bala, 2008), as well as the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al., 2003). These developments signify the model’s adaptability to the dynamic and complex landscape of technology. Within this framework, variables such as Perceived Ease of Use (PEOU), Perceived Usefulness (PU), Attitude Toward Using (ATT), and Behavioural Intention (BI) are emphasized for their significant roles in describing and predicting technology acceptance behaviors (Marangunić & Granić, 2015).

In teacher education, research on AI acceptance typically employs the TAM and its derivatives. A notable study by Al Darayseh (2023) investigated the acceptance of AI among 83 science teachers. The findings indicated a high level of acceptance, which positively correlated with factors such as self-efficacy, PEOU, perceived benefits, ATT, and BI. Although the effects of anxiety and stress on these variables were not significant, these insights are crucial for understanding AI technology acceptance by in-service teachers. Furthermore, a study by Choi et al. (2023) demonstrated that in-service teachers’ pedagogical beliefs and perceived trust
in AI significantly influence their BI toward AI educational tools. These studies offer valuable perspectives on in-service teachers’ attitudes towards AI; however, there is a notable paucity of research concerning pre-service teachers’ acceptance of AI technology. This gap underscores a lack of understanding about how pre-service teachers perceive and accept AI in educational contexts. The attitudes and acceptance levels of pre-service teachers are pivotal for the future development of educational technology, as they will shape subsequent teaching practices. Consequently, conducting systematic research on pre-service teachers’ acceptance of AI technologies is essential for crafting effective educational policies and designing robust teacher training programs. This approach will help ensure that future educators are well-prepared to integrate AI into their teaching effectively and ethically.

2.3 Feedback for Reflective Writing

2.3.1 Theoretical Model for Effective Feedback Delivery

Feedback is conceptualized as information that modifies learners’ behavior and thinking (e.g., Hattie & Timperley, 2007; Shute, 2008). This process aids learners in identifying their strengths and weaknesses, bridging the gap between their current performance and objectives (Biber et al., 2011). Consequently, feedback is pivotal in facilitating learning and skill development (e.g., Thurlings et al., 2013) while bolstering individual self-confidence and motivation (Fong & Schallert, 2023). Hattie and Timperley (2007) developed the most influential feedback model. Hattie and Timperley (2007) conceptualized feedback as information provided by an agent (e.g., teacher, peer, self) to modify the learner’s thinking or behavior to enhance learning. They claimed that effective feedback should address three fundamental questions: 1) “Where am I going?”, 2) “How am I going?” and 3) “Where to next?”, aligning with their feedback model’s three levels: task level (feedback concerning the task, for instance, a student’s performance on a specific assignment or skill), process level (feedback about the strategies and their efficacy
needed to complete the task), and self-regulation level (feedback that fosters students’ self-assessment and self-regulatory skills, promoting self-directed learning). Moreover, Hattie and Timperley (2007) underscored the importance of timely, specific, and pertinent feedback, enabling students to leverage this information to optimize their learning trajectory. They emphasized that effective feedback transcends mere identification of errors or correct responses; it should cultivate a deeper understanding and mastery of the subject matter, encouraging students to introspect about their learning journey, capitalize on their strengths, and address their shortcomings. In addition, Shute (2008) delineates various forms of feedback content, categorizing them by their complexity. The research primarily focuses on three types of feedback: (1) *Knowledge of Results* (KR), which confirms the accuracy of a response (e.g., “your answer is correct/incorrect”); (2) *Knowledge of Correct Responses* (KCR), which provides the correct answer and more information about task (e.g., “the correct answer is A/B, etc.”); and (3) *Elaborative Feedback* (EF), offering comprehensive information, such as a strategic hint, an explanation, or a worked example (e.g., “your answer is correct/incorrect because...”). This classification aids in understanding the impact of varying feedback levels on learner outcomes.

Feedback is pivotal in the educational process, significantly enhancing student learning and development across multiple dimensions. When teachers deliver timely, specific, and constructive feedback, it not only improves learning efficiency and effectiveness (Pereira et al., 2016) but also boosts motivation (Gan et al., 2021) and fosters students’ self-regulation and metacognitive skills (Callender et al., 2016; Labuhn et al., 2010). However, in higher education, the challenge of increasing class sizes contributes to heavier workloads for instructors, thereby constraining their capacity to provide personalized and timely feedback (e.g., Banihashem et al., 2024). To address this issue, researchers and developers have introduced various solution strategies, notably AI-based feedback systems (e.g., Zawacki-Richter et al., 2019). These
systems are gaining traction within the educational community for their ability to automate the feedback generation process. This automation enables educators to manage large classes while delivering customized and immediate feedback on students’ learning activities. Furthermore, the application of AI-driven systems not only alleviates the workload of teachers but also enhances the learning process by providing accurate and personalized feedback. The development of AI-powered feedback tools tailored to specific pedagogical tasks is crucial for elevating the quality and efficiency of education. Such tools optimize students’ learning experiences and equip teachers with robust pedagogical support, thus significantly enhancing modern educational practices and adapting them to evolving educational demands.

2.3.2 Artificial Intelligence Feedback Systems for Reflective Writing

In recent years, many automated feedback systems have been generated in education to meet the needs of various learning activities. These systems range from pre-defined rules (Pardo et al., 2018) to those leveraging pre-trained language models (MacNeil et al., 2023). The recent emergence of ChatGPT is an essential milestone for AI feedback (e.g., Dai et al., 2023; Banihashem et al., 2024). Specifically to the domain of reflective writing, Ullmann (2019) has cataloged a variety of automated analysis methodologies that encompass dictionary-based, rule-based, and ML-based approaches. These innovations hold substantial potential for enhancing the feedback process in reflective writing, offering a more personalized and scalable approach to student support.

Recent research indicates that ML-based approaches are popular for analyzing reflective writing within educational settings. The application of these techniques spans a broad spectrum, encompassing traditional ML algorithms (Jung & Wise, 2020; Ullmann, 2019), deep learning frameworks (Carpenter et al., 2020; Nehyba & Štefánik, 2023), and pre-trained language models (Wulff et al., 2023). In traditional ML algorithms, the literature suggests that Random
Forest algorithms, SVMs, and Naïve Bayesian classifiers are among the most effective for analyzing reflective writing. The Random Forest algorithm has been validated by the research of Beigman Klebanov et al. (2017), Kovanović et al. (2018), and Liu et al. (2019) for its robustness and accuracy in handling such classification tasks. Similarly, SVMs have been demonstrated by Fan et al. (2017) and Liu et al. (2019) to be highly effective, benefiting from their capacity to manage high-dimensional data spaces. Lastly, the utility of Naïve Bayesian classifiers in this context is supported by the findings of Poon et al. (2017), who highlight their efficiency and relative simplicity. Further, the research conducted by Nehyba and Štefánik (2023) underscores the enhanced capabilities of deep learning models in classifying reflective writing. Their findings reveal a notable accuracy range of 76.56-79.37% in samples with low confidence levels, and an impressive 97.56-100% accuracy in instances where the model has high confidence. The study by Wulff et al. (2023) significantly contributes to analyzing reflective writing by demonstrating the efficacy of the BERT (Bidirectional Encoder Representations from Transformers) model in the domain of reflective writing analysis. Their findings indicate that BERT begins surpassing the performance of other models after training on just 20% to 30% of the available data. This efficiency not only underscores the practicality of implementing BERT in educational assessment but also points to its potential to significantly refine the categorization process of reflective writing, thereby providing educators with powerful tools for analyzing and supporting student learning.

Regarding the content of feedback, current automated feedback also relies on the theory of reflection. Some studies have given automated feedback based on reflective levels. For instance, Jung et al. (2022) analyzed the reflection level of 369 dental students by examining 1,500 reflective essays and categorizing them as non-reflective, shallow, or deep. Similarly, Liu et al. (2019) implemented a binary classification system to analyze 301 pharmacy students’ reflective statements about their clerkships, successfully distinguishing between reflective and
non-reflective responses. Such feedback mechanisms enable educators to precisely assess students’ current reflective capabilities and establish specific developmental objectives. Several studies also provide feedback from multiple dimensions. For instance, Ullmann (2019) analyzed 76 student essays encompassing 5,080 sentences to assess various dimensions, including reflection, experience, feeling, belief, difficulty, perspective, learning, and intention. This multidimensional approach allows for a more comprehensive evaluation of students’ reflective writing, capturing a broader range of cognitive aspects. Moreover, some studies have achieved a more comprehensive reflection assessment by integrating previous theories. For example, Solopova et al. (2023) employed Gibbs’ Cycle of Reflection model (1988) to evaluate and provide feedback on the various cognitive stages of reflection. They also utilized Fleck and Fitzpatrick’s hierarchical model (2010) to conduct a nuanced assessment of reflection levels. Additionally, their feedback model incorporates a composite rating that includes affective responses, reflective themes, and linguistic criteria. This multifaceted approach offers a multidimensional perspective that deepens the understanding of learners’ reflective processes, thereby enhancing the accuracy and utility of the assessment.

While algorithms and feedback mechanisms have been extensively explored, practical applications still present significant challenges. From an algorithmic standpoint, sentence-based annotation is commonly employed in AI-driven feedback for reflective writing (e.g., Beigman Klebanov et al., 2017; Kovanović et al., 2018; Liu et al., 2019; Ullmann et al., 2019). However, this method may fail to capture the entire text’s overarching themes and contextual nuances. Reflective writing typically encompasses deeply personal experiences and emotions, necessitating a comprehensive understanding of the full text (e.g., Schön, 1983, 1987). Additionally, sentence-based analysis might lead AI systems to overemphasize grammar and keywords at the expense of grasping the text’s deeper meanings and emotional subtleties. Thus, shifting the AI’s focus from sentence-level to text-level analysis could yield more deep and
holistic feedback. Moreover, as researchers increasingly turn to deep learning and pre-trained language models (Nehyba & Štefánik, 2023; Wulff et al., 2023), the transparency of these algorithms often remains underaddressed in educational settings. Transparent algorithms enable users to comprehend how decisions are made by the AI, thereby building trust in the system (Choi et al., 2023). They also facilitate easier identification and correction of errors or biases by researchers and users, making transparency a critical factor in enhancing the effectiveness and acceptance of AI in education. Regarding feedback content, although theoretical models are well-implemented, current mechanisms primarily address the question “How am I doing?” without sufficiently tackling “Where to next?” (e.g., Hattie & Timperley, 2007). There is a compelling need to integrate interdisciplinary expertise from pedagogy, psychology, and artificial intelligence to develop more nuanced and effective feedback mechanisms. This approach would address immediate performance and guide future development, thus enriching the learning experience.

2.4 Mixed Methods in Research on Reflective Writing

2.4.1 Mixed Methods in Educational Research

A mixed-methods research design is an approach that involves collecting, analyzing, and integrating both quantitative and qualitative data within a single study to address a research problem (Creswell, 2015). Mixed methods research combines quantitative and qualitative approaches and provides a powerful way to address complex research questions. Specifically, Mixed methods research is categorized into three main strategies: sequential exploratory strategies, sequential explanatory strategies, and concurrent mixed methods approach, each with its unique application advantages and potential limitations (Hagenauer et al., 2023; Tashakkori & Teddlie, 2010).
Firstly, the sequential exploration strategy begins with the use of qualitative research methods to explore new research areas or complex phenomena, and then utilizes the qualitative findings to guide subsequent quantitative research. This strategy is particularly suited to emerging fields and helps researchers form and refine hypotheses and reveal deeper meanings from the data to provide direction for quantitative research. However, this strategy requires a larger investment of time and resources, and the preliminary findings of qualitative research may compromise the objectivity of quantitative research. Second, the sequential explanatory strategy, on the other hand, consists of a quantitative study to collect data, followed by the use of qualitative research to deepen and explain the quantitative results. The advantage of this strategy is that the quantitative data provide a broad perspective, while the subsequent qualitative research helps to deepen the understanding of the statistical phenomena, and is particularly helpful in interpreting complex data, such as statistical anomalies or unexpected trends. However, this strategy is also time- and resource-intensive, and the qualitative research phase may be limited by the prior quantitative results, affecting the depth and breadth of its exploration. A concurrent mixed-methods approach conducts both quantitative and qualitative research and aims to cross-validate and enhance the overall explanatory power of a study through different data sources (Greene, 2007). This approach saves time and provides a more comprehensive and multi-perspective view, enhancing the validity and reliability of the study. However, integration of data can be a challenge, especially when the two approaches produce conflicting results. In addition, this approach requires that the researcher be proficient in both quantitative and qualitative research methods.

2.4.2 Potential Strengths of Mixed Methods on Reflective Writings

Research on reflective writing has largely used qualitative content analysis, frequently employing small sample sizes to conduct an in-depth analysis of the writing’s qualitative
attributes, constituent elements, thematic content, and others. Noteworthy contributions to this method include Cochran et al. (2013) and Dunne (2019), who concentrated on analyzing the depth of reflection in their respective studies. Similarly, Alt et al. (2022), Bell et al. (2011), and Bowman (2021) contributed by examining the specific components that constitute reflective writing. Additionally, Pryjmachuk et al. (2019) and Rojí et al. (2017) identified emergent topics within the corpus of reflective writings under consideration. Qualitative methodologies are indispensable for eliciting detailed insights into individual differences, cultural factors, and complex psychological underpinnings in this context. Nevertheless, it is crucial to note that the circumscribed scope and particularity of the contexts studied in qualitative research often limit the generalizability of the findings to broader demographic cohorts.

Reflective writing is textual data, so more and more studies are starting to use computational linguistics to analyze and assess it. For example, the LIWC tool emerges as a recurring choice for research methodology, prominently featuring in investigations by Cui et al. (2019), Savicki and Price (2015, 2021), and Springer and Yinger (2019). Complementing the utility of the LIWC tool, the AWA tool has been adopted in scholarly works by Hanlon et al. (2021) and Shum et al. (2016). Beyond these, keyword analysis is another widely employed methodological approach, evidenced by its utilization in the research undertakings of Chong et al. (2019) and Ullmann (2019). Linguistic attributes have also been systematically analyzed through POS tagging, as exhibited in the works of Gibson et al. (2017) and Chong et al. (2019). Adding another layer of methodological richness, the study conducted by Wise et al. (2020) explored co-occurrence networks within the scope of reflective writing. Computational linguistics can efficiently process large amounts of textual data and analyze reflective writing from different periods or groups of people in order to identify significant trends and patterns. In addition, computational methods provide a more objective means of analysis that significantly reduces human bias and error compared to manual assessment. However,
computational linguistics has limitations in dealing with complex semantic content, especially in understanding complex linguistic expressions such as metaphor and irony in human writing. This may lead to misinterpretation of the true intent of the text. Therefore, when relying on computational linguistics for text analysis, additional mechanisms are needed to recognize and process such complex linguistic expressions to ensure the accuracy and reliability of the analysis results.

Given the limitations of qualitative content analysis and computational linguistics approaches, using a mixed-method approach combining the strengths of quantitative and qualitative research may provide a more comprehensive and in-depth analytical perspective. Mixed methods allow the researcher to focus on the data’s quantitative and qualitative aspects, thus providing a more comprehensive analysis. For example, a researcher may use quantitative methods to identify common linguistic patterns and frequencies when analyzing reflective writing. In contrast, qualitative methods delve into the meaning of the text and the emotions behind it. In addition, quantitative data can be used to validate the generalizability and breadth of qualitative analysis, while qualitative data provide the necessary depth of interpretation and context for quantitative results. The complementary nature of this approach not only enhances the credibility of the research but also deepens its depth. Researchers can explore issues at multiple psychological, social, and linguistic levels through this combination of methods, resulting in deeper insights and more comprehensive understanding.
3. Empirical Studies

3.1 PetraKIP Project

The data for the thesis came from the PetraKIP (Persönliches transparentes KI-basiertes Portfolio für die Lehrerbildung) project. The project aims to develop an AI-based portfolio system to enhance the processes of reflection and self-regulation learning in teacher education. The project undertook the development of the AI system by employing techniques such as NLP and ML, while studying the acceptance of AI among pre-service teachers. Additionally, the functional effectiveness of the implemented AI system was also evaluated.

3.2 Research Design

The thesis was conducted in the context of a mini-portfolio within the teacher education curriculum at a German university during the winter semester of 2021/2022, the summer semester of 2022, and the winter semester of 2022/2023. The two self-learning modules selected from the school education curriculum were “Pedagogical Diagnostics” and “Classroom Management.” This e-portfolio comprised learning materials, assignments, and reflective writings. Additionally, the study materials included pre-recorded videos, presentation slides, and recommended readings. These educational resources were available to students for both preparation before classes and review afterward.

Figure 6 Mini-portfolio design and implementation
The data for this thesis were categorized into two main types: questionnaire data and reflective writing text data. The data for Study 1 were collected via questionnaires during the winter semester of 2021/2022, resulting in a total of 452 valid responses. The data for Study 2 were derived from reflective writings completed in two self-study modules also during the winter semester of 2021/2022, yielding 198 valid entries. For Study 3, the data encompassed all reflective writings across three semesters (the winter semester of 2021/2022, the summer semester of 2022, and the winter semester of 2022/2023), from which 1043 valid entries were obtained.

3.3 Summary of Research Results

3.3.1 Study 1

*Research Questions (RQs):*

**RQ1:** To what extent do the pre-service teachers’ AI acceptance support the hypothesized relationships in the proposed research model?

**RQ 2.1:** To what extent do pre-service teachers respond differently concerning gender when measuring their acceptance of AI?

**RQ 2.2:** What are the gender differences in the latent mean of each construct?

**RQ 2.3:** Do gender differences have a moderating effect on the relationship of variables in the proposed research model?

*Methods:*

**Theoretical model:** TAM3 (Venkatesh & Bala, 2008).

**Measurement instrument:** Perceived Usefulness (PU, four items, \( \alpha = 0.88 \)), Perceived Ease of Use (PEOU, four items, \( \alpha = 0.76 \)), Artificial Intelligence Self-Efficacy (AISE, four items, \( \alpha = 0.87 \)), Artificial Intelligence Anxiety (AIA, three items, \( \alpha = 0.91 \)), Perceived Enjoyment (PE, three items, \( \alpha = 0.86 \)), Social Norms (SN, two items, \( \alpha = 0.90 \)), Job Relevance (JR, three items, \( \alpha = 0.90 \)), and Behavioral Intention (BI, two items, \( \alpha = 0.66 \)) (Stephan, 2021). Detailed information is available at Table 5 AI Acceptance Scale (German version).

**Sample:** \( N = 452 \); of which 325 female students (71.90 %), winter semester 2021/2022
**Research methods:** Structural equation modeling (SEM), measurement invariance, latent mean differences, multigroup analysis (MGA)

**Key Results:**
- The results of the study showed support for eight of the nine hypotheses; of these, PEOU ($\beta = 0.297^{***}$) and PU ($\beta = 0.501^{***}$) were identified as the main predictors of pre-service teachers’ intentions to use AI.
- In addition, latent mean difference results indicated that AIA ($z = -3.217^{**}$) differed significantly by gender.
- Notably, gender moderated the path effects of AIA on PEOU ($p = 0.018^*$).

**Publication P1**

Appendix II: Table 6

Table 2 Research summary for study 1

### 3.3.2 Study 2

**Research Questions (RQs):**

**RQ1:** On which levels of reflection can reflective writings of pre-service teachers on profession-related topics be described and evaluated?

**RQ2:** How do the psycholinguistics of reflective writings differ among pre-service teachers at different levels of reflection?

**RQ3:** How do pre-service teachers at different levels of reflection differ in the topics of their reflective writings?

**Methods:**

**Framework:** Levels of reflection (Hatton & Smith, 1995; Fütterer, 2019).

**Participants and Data Corpus:** $N = 105$ (female 66.67%), 198 Reflective Writings.

**Research Methods:** Qualitative content analysis* (Gläser-Zikuda et al., 2020; Mayring, 2014), LIWC2015 (Pennebaker et al., 2015), BERTopic (Grootendorst, 2022).

*Cohen’s Kappa was very good, with 0.97 for the topic of pedagogical diagnostics and 0.96 for classroom management.
Key Results:

- The qualitative content analysis indicated that participants primarily engaged in descriptive and low-level reflective writing.
- Further, computational linguistic analysis demonstrated that the use of affective and cognitive terms varied across different levels of reflection, with a higher frequency of these terms associated with deeper levels of reflection.
- Additionally, the results from BERTopic suggested that the reflective content primarily focused on learning materials and gradually transitioned towards affective and motivational themes as the level of reflection increased.

Publication P2

Table 3 Research summary for study 2

3.3.3 Study 3

Research Questions (RQs):

RQ 1: To what extent do shallow machine learning models employing diverse language representations demonstrate effectiveness in classification reflective writing of pre-service teachers?

RQ 2: What is the performance of pre-trained language models when employed to classify reflective writing of pre-service teachers?

RQ 3: How do shallow machine learning models compare to pre-trained language models in terms of their effectiveness in classifying reflective writing of pre-service teachers?

Methods:

Dataset*: N = 1043 reflective writings, winter semester 2021/22, summer semester 2022, and winter semester 2022/23, word count: M = 51.38, SD = 143.08.

Data Annotation: Hatton and Smiths’ theory (1995) and Fütterer’s (2019) adapted coding scheme for classification, qualitative content analysis (Gläser-Zikuda et al., 2020).

Classification Algorithm: Shallow Machine Learning (Decision Trees, Random Forests, Support Vector Machines, Ridge Classifier, SGD Classifier, XGB Classifier, and Gradient
Boosting Classifier) & Pre-trained Language Models (BERT, RoBERTa, Longformer, and BigBird).

*300 reflective writings (28.76% of the total) were randomly selected for a second coding round, Cohen’s kappa coefficients of 0.67, 0.66, and 0.73.

**Key Results:**

- The average accuracy achieved by the shallow ML is typically under 60%.
- LIWC2015 generally outperforms both BoW and TF-IDF across various algorithms.
- These accuracy levels mark a considerable improvement, a rise of 12% to 16%, in comparison to the shallow ML models.
- Longformer had best performance with an accuracy of 77.22%.

**Publication P3**


Appendix II: Table 8

Table 4 Research summary for study 3
4. Discussion

This chapter synthesizes findings from three empirical studies that explore pre-service teachers’ AI acceptance and the design of AI-based feedback systems for reflective writing. Initially, it identifies critical determinants influencing pre-service teachers’ AI acceptance. It also examines the role of gender as a moderating factor, investigating whether there are differences in AI acceptance between male and female pre-service teachers and the implications of these differences. Furthermore, the chapter outlines practical strategies for designing AI-based feedback systems tailored for reflective writing. Lastly, it addresses potential limitations of the current research and proposes directions for future studies.

4.1 Integration of Results

4.1.1 Determinants of AI Acceptance

The study targeting pre-service teachers’ AI acceptance revealed significant variables derived from the TAM literature. Study 1 emphasized that PU and PEOU are the main factors influencing an individual’s decision to accept and utilize AI. PEOU is defined as pre-service teachers’ perceptions of the simplicity or complexity of using AI systems (Venkatesh & Davis, 2000). Conversely, PU pertains to the degree to which pre-service teachers believe that employing AI will enhance their professional performance (e.g., Lee et al., 2005). These findings align with earlier theoretical frameworks (Davis et al., 1989) and empirical studies across various fields (e.g., Scherer & Teo, 2019). A more recent study by Schiavo et al. (2024) highlights that PEOU represents the importance of user interaction, whereas PU is more closely linked with task-specific relevance. They critique the portrayal of AI as offering “natural and intuitive” interactions, arguing that this perspective overlooks the user’s experience and perception during use (Schiavo et al., 2024, p.8). Our results, indicating that PU substantially impacts BI more than PEOU, support this critique. Even if an AI system’s interaction design
appears “natural,” its inefficiency or clumsiness in executing essential tasks can produce a predominantly negative user perception. When developing AI functionality in education, it is crucial to prioritize usefulness. This implies that AI systems should not only be capable of completing established educational tasks but should also do so in a manner that aligns with educational goals and positively impacts learners. The effectiveness of these systems must be measured against their ability to enhance learning outcomes and support educational objectives, ensuring that the integration of AI contributes meaningfully to the educational process.

In addition, the study identified a gender difference in AIA, revealing that female pre-service teachers exhibited greater anxiety towards AI technology than their male pre-service teachers. According to Bandura’s theory of self-efficacy, an individual’s confidence in their ability to successfully execute specific tasks significantly influences their performance and motivation (Bandura & Adams, 1977). This concept is particularly relevant in technology fields, including AI, where women often report lower self-efficacy than men (e.g., Huffman et al., 2013; Vekiri & Chronaki, 2008). Gender biases are entrenched from childhood through various socialization agents, including family, the educational system, media, and peers (Eagly & Wood, 2012). Women are often socialized with the stereotype that science and technology are not suited for them, which can dampen their interest and confidence in these fields (Dai et al., 2020; González-Pérez et al., 2020). Furthermore, the limited visibility and representation of women in science and technology and a lack of female role models constitute significant barriers that restrict women’s entry into these fields (Cheryan et al., 2011). Such underrepresentation makes it challenging for women to find mentors and inspiration, hindering their ability to envision themselves succeeding in these domains. The absence of relatable success stories may lead them to perceive science and technology as unsuitable career paths, adversely affecting their self-efficacy and, ultimately, their career choices. Female teachers play a crucial role in the K-12 education system. However, excessive anxiety regarding AI can negatively impact their
professional development. This anxiety not only hinders their personal growth but can also adversely affect the K-12 education system as a whole. Firstly, AI anxiety can hinder female teachers’ ability to learn and effectively utilize this technology, subsequently limiting their professional development. In an era where technology is increasingly integral to education, discomfort with these changes can compromise their competitiveness in the educational sector. Secondly, if teachers feel unconfident or anxious about utilizing AI technologies, it may decrease both the frequency and effectiveness with which they integrate these tools into their teaching practices. This reluctance can, in turn, detrimentally affect the quality of their teaching and the learning outcomes of their students. These challenges underscore the need for targeted support and training programs to help alleviate anxiety and build competence in AI technologies among female educators. This support can empower them to harness the full potential of AI, enhancing their teaching effectiveness and ultimately improving student learning outcomes.

4.1.2 Design AI-based Automated Feedback For Reflection

In Study 2, we conducted an exploratory analysis of reflective writing. Our study yielded a critical insight: pre-service teachers exhibit relatively weak reflection, underscoring the necessity of providing structured feedback. Furthermore, our findings indicate a significant positive correlation between the frequency of cognitive and affective terminology usage in reflective writing and the overall quality of the reflections. These results are consistent with prior research (Cui et al., 2019; Lin et al., 2016). This correlation highlights the importance of emotional and cognitive expressions in enhancing the depth and effectiveness of reflective writing. Cognitive terminology, such as “understand,” “analyze,” and “reason,” typically signals abstract thinking, logical reasoning, and conceptual integration. Their presence in a text indicates the extent to which the writer engages in complex cognitive operations, thus reflecting the development of higher-level cognitive activities (e.g., Lucas et al., 2019). Conversely,
affective terms like “happy,” “confused,” and “frustrated” expresses the writer’s emotional
depth and engagement in the reflective process, highlighting emotional evaluations and
responses to personal experiences (Cushion, 2016). Our study highlights that the adept use of
cognitive and affective terms showcases the depth of the author’s cognitive and emotional
processing and enhances our understanding of their reflective writing.

The findings from Study 2 provided empirical support for developing more transparent
and interpretable ML algorithms, which we further explored in Study 3. In this subsequent study,
we employed features such as BoW, TF-IDF, and linguistic indicators extracted using
LIWC2015 for training shallow ML models. The results indicated that indicators derived from
LIWC2015 were the most effective. LIWC2015, a psycholinguistic tool, analyzes text for
elements related to emotion, social relationships, cognitive processes, and psychological
dynamics (Pennebaker et al., 2015). It categorizes and analyzes text using detailed labels, such
as positive emotion, negative emotion, and cognitive processes. This enables a more accurate
capture of subtle differences in student writing and provides a refined assessment of reflective
writing quality. Additionally, the classification criteria of LIWC2015 are grounded in extensive
psychological and linguistic research, providing a robust theoretical foundation for each
category (cf. Gottschalk & Glaser, 1969; Stone et al., 1966; Weintraub, 1989). This grounding
not only enhances the transparency of the study but also improves its interpretability, allowing
researchers to theoretically explain why specific categories correlate with the quality of
reflective writing or other psychological variables.

However, when comparing shallow ML methods that utilize LIWC2015 with pre-trained
language models, the findings indicate that although LIWC2015 provides intuitive and
interpretable outputs, pre-trained language models generally demonstrate superior predictive
accuracy and overall efficacy. Pre-trained language models, trained on extensive corpora, excel
in comprehending language’s deep structures and complex semantics. These models utilize
sophisticated network architectures, such as Transformers, to extract high-level and multi-layered features, far surpassing the capabilities of traditional methods like BoW, TF-IDF, or LIWC2015. Consequently, pre-trained models offer significant advantages in handling the diversity and complexity of language. Unlike tools such as LIWC2015, which rely on a fixed vocabulary, pre-trained language models interpret word meanings dynamically based on context. This capability is crucial for understanding the polysemy and complex emotions students may express in reflective writing. However, the primary drawback of using pre-trained language models lies in their opaque decision-making processes. With their intricate network structures and vast parameter sets, models like BERT obscure the rationale behind specific decisions (Devlin et al., 2018). The core issue stems from the difficulty educators and learners face in understanding and verifying the decision mechanisms of these models, which in turn impacts the trust in and adoption of these technologies. This opacity impedes the establishment of trust and poses substantial barriers to the practical use of these systems. Firstly, educators often lack the technical expertise required to interpret outputs from complex models, and the “black box” nature of these systems can foster skepticism. Moreover, the absence of sufficient transparency prevents educators from validating the effectiveness and accuracy of the recommendations or decisions made by these models. Such validation is essential for assessing student performance and making instructional adjustments. Furthermore, in an educational environment where fairness and ethics are paramount, the opacity of these decision-making processes could obscure potential biases, thus hindering educators’ efforts to guarantee equitable treatment of all students.

According to the results of Study 3, we faced a dilemma when choosing the appropriate AI algorithm for providing feedback on reflective writing. Shallow ML models based on LIWC2015, while exhibiting excellent transparency and interpretability, underperformed by 17% compared to pre-trained language models. Although pre-trained language models demonstrated
higher effectiveness, their “black box” nature could potentially increase pre-service teacher anxiety, thereby reducing their willingness to use AI-based educational tools. However, after a comprehensive evaluation, we opted for the latter—pre-trained language models. This decision was based on the consideration that the accuracy and reliability of AI predictions are crucial for educational applications. While AI-related anxiety might affect users’ perceptions of tool usability (PEOU), our study found that the performance advantage (PU) of a tool has a greater impact on users’ acceptance of AI than usability (PEOU). Therefore, during the decision-making process, we prioritized performance advantage. It should be noted, however, that no single AI model is universally applicable to all educational contexts. The specific choice of model depends on the particular needs of the research, available resources, and the anticipated implementation outcomes. In this decision-making process, it is necessary to consider a variety of factors and make a comprehensive trade-off.

4.2 Practical Implications

This section summarizes the practical implications of three empirical studies, each addressed separately. The first study provides targeted recommendations to enhance the acceptance of AI-based educational systems among pre-service teachers. It highlights the critical role of PEOU in AI adoption, advocating for AI system designers and developers to focus on creating intuitive user interfaces and interactions. The goal is to design AI systems that are straightforward and user-friendly, thereby minimizing the learning curve and enhancing users’ willingness to engage with the technology. Additionally, the study emphasizes the importance of PU. It recommends conducting a comprehensive needs analysis to ascertain students’ requirements and preferences for utilizing AI systems. This analysis should explore students’ expectations of how AI can support their learning, the essential functionalities, and the specific learning objectives they wish the AI system to facilitate. Moreover, this thesis addresses gender differences, particularly the heightened anxiety towards AI among female pre-service teachers,
by proposing tailored training and support initiatives. It advocates for specialized training programs that enhance pre-service teachers’ AI-related competencies and mitigate their technophobia, fostering greater trust and acceptance of AI technologies. Echoing the findings of Nazaretsky et al. (2022), the thesis underscores the efficacy of AI skills training in reducing teachers’ anxiety about AI, suggesting that such educational interventions are crucial in promoting the adoption of AI within educational contexts.

For the second empirical paper, the above findings can provide practical advice for pre-service teachers on how to become reflective practitioners, for teacher educators on how to diagnose reflective writing, and for researchers on how to train ML algorithms using linguistic indicators. First, pre-service teachers should understand that developing reflection is not a one-time event but a continuous and challenging process that requires consistent practice. Reflective writing enables them to better understand their teaching practice and professional growth, supporting their ongoing development throughout their careers. The study suggests that pre-service teachers can pay more attention to writing reflections when affective and cognitive descriptions are present. This is because using these terms implies a deeper integration of past experiences. This process necessitates pre-service teachers to critically examine their experiences and beliefs while actively engaging in self-reflection and self-awareness. Furthermore, teacher educators may offer various supports, such as prompts, to aid pre-service teachers in developing their reflective writing abilities. Nevertheless, pre-service teachers must understand that they should not depend excessively on these prompts when engaging in reflective writing. Prompts should serve as tools to facilitate reflection and not as templates to rely upon. Refrain from relying on prompts can result in superficial reflections, which are often characterized by homogenous themes and a lack of in-depth exploration of personal experiences. Teacher educators may provide continuous training in reflection for pre-service teachers, not only in theoretical courses but also in practical courses, due to the low level of reflection
exhibited by students. In supporting reflection, some teacher educators design prompts to aid students in better reflection. However, it is essential to note that the design of these prompts should be brief and should allow room for individual interpretation. Furthermore, when evaluating reflective texts, it is crucial to pay special attention to the level of depth in students’ reflections when affective and cognitive vocabulary is present. The frequency of such language is often an indicator of more profound engagement with personal experience and critical thinking, which is a fundamental aspect of reflective writing. Therefore, teacher educators should encourage pre-service teachers to use affective and cognitive vocabulary in their reflective writing and use it as a criterion for assessment. Next, timely feedback is critical to support pre-service teachers’ ability to engage in reflective writing. Teacher educators should provide timely and constructive feedback on not only how to improve reflection but also how to be a good teacher. Lastly, our study has identified several linguistic indicators that can predict the level of reflective writing among pre-service teachers, which can guide researchers in developing machine learning algorithms to support reflective writing assessment. This has important implications for teacher education, as machine learning algorithms can provide objective and reliable assessments of reflective writing, allowing teacher educators to focus on providing feedback and supporting pre-service teachers’ development.

The impact of this third empirical study extends beyond the boundaries of teacher education into the broader field of education. By incorporating diverse ML algorithms, particularly the pre-trained language models discussed in the study, students were able to receive more accurate and immediate feedback. This feedback plays a crucial role in enhancing students’ reflective writing skills, thus positively impacting teaching and learning outcomes. For educators, utilizing an automated system to provide accurate feedback improves productivity and reduces the stress of their work. This allows teachers to devote more time and
resources to other critical instructional tasks such as curriculum design, classroom interactions, and individual tutoring of students.

**4.3 Limitations and Future Work**

For future improvement of this thesis, several vital aspects warrant attention. Firstly, the AI acceptance scale necessitates a comprehensive redesign. While the current research utilized a scale modified from the TAM3, it became evident that AI acceptance exhibits distinct characteristics compared to traditional technology acceptance (e.g., Castelo et al., 2019; Li & Huang, 2020). Additionally, the variables examined were confined to those within the TAM3 framework, neglecting AI-specific attributes such as transparency, compatibility, and potential risks (Fan et al., 2020; Man et al., 2020; Weitz et al., 2021). To address these gaps, future research should not only extend the TAM3 model to incorporate additional external variables but also strive to develop novel scales tailored specifically for AI acceptance evaluation. Moreover, while this study employed quantitative methods to uncover factors affecting pre-service teachers’ acceptance of AI, including notable gender differences, it did not delve into the deeper motivations and underlying reasons for these phenomena. Future research should incorporate qualitative methodologies, such as interviews, to better understand the observed disparities and the factors driving them.

Secondly, current research has predominantly focused on task-level feedback (e.g., Hattie & Timperley, 2007), with a noticeable shortfall in examining feedback related to self-regulated level. Such assessments are vital for an in-depth understanding of students’ metacognitive skills and self-regulated learning processes during teaching and learning activities. Reflection is a multifaceted process that includes cognitive, emotional, and psychological dimensions (Boud et al., 2013; Kember et al., 1999). Incorporating self-regulation level assessment metrics into feedback gives educators a holistic view of students’ ability to monitor their learning progress, establish goals, and modify cognitive strategies accordingly. Furthermore, concerning technical
implementation, reliance on pre-trained language models for classification presents challenges, including limitations in real-time knowledge updating, interpretability, traceability, resource demands, and operational efficiency (Min et al., 2023). In contrast, retrieval-based methods, like Retrieval-Augmented Generation (RAG) (Lewis et al., 2020), offer dynamic information retrieval from documents or knowledge bases, enabling access to the most current information and delivering more timely and pertinent responses to users. The integration of retrieval components in RAG bolsters the timeliness and relevance of the provided information and improves the model’s interpretability, allowing users to understand the basis of the model’s outputs—specifically, the documents or data from which the answers are derived. In addition, developing LLMs tailored for reflective writing tasks represents a promising direction for enhancement, assuming adequate resources are available. Such a model, trained on an extensive corpus of pedagogical literature encompassing various educational theories like constructivism, behaviorism, and cognitivism, would possess an advanced capacity for deep text comprehension and generation, showcasing superior linguistic processing skills. These specialized LLMs would not only deliver highly relevant and precise feedback but also engage with students in a more nuanced and human-like manner. The above approaches can avoid the problem of bias caused by unbalanced training data.

Lastly, this thesis did not conduct an empirical study into the efficacy of AI-based feedback. Future studies could consider deploying quasi-experimental designs to contrast the impacts of AI-generated feedback with those of teacher-provided feedback. This methodological approach would enable a more nuanced exploration of AI feedback’s potential benefits in enhancing learning outcomes, fostering students’ self-regulated learning skills, and boosting motivation. Furthermore, it is essential for upcoming research to delve into students’ acceptance of AI-based feedback. This involves examining students’ perceptions and evaluations of AI feedback’s effectiveness and practicality and understanding how these
perceptions influence their learning behaviors and outcomes. Undertaking such research will equip educators with crucial insights, guiding them to incorporate AI feedback into their pedagogical strategies more seamlessly.
5. Conclusion

In the wave of digital transformation, educational technologies, particularly AI, are becoming increasingly essential for fostering the professional development of pre-service teachers. Research has shown that reflection is critical to enhancing students’ self-regulation and thereby professional development. However, the level of reflection among pre-service teachers is low and fostering their reflection remains a significant challenge in teacher education. Leveraging AI to support student reflection may be a potential solution. Therefore, this thesis integrates three empirical studies designed to (i) identify the determinants of pre-service teachers’ AI acceptance, (ii) analyze the depth and linguistic indicators of reflective writing, and (iii) target the design of AI feedback mechanisms. These studies synergistically contribute to a deeper understanding of reflective writing and foster reflection in pre-service teachers to prepare them for their future careers.

Motivating pre-service teachers to adopt AI-based educational tools is crucial. The findings of this thesis suggest that teacher educators should first address the specific educational needs of their students. Meanwhile, the development of AI models should also align with educational goals. Moreover, building and maintaining trust in AI technologies is essential, particularly in enhancing female students’ confidence and reducing their anxiety. Building on this understanding, the thesis proposed developing an AI-based feedback system for reflective writing. First, exploratory analysis revealed that students’ cognitive and affective engagement during reflective writing could effectively predict their level of reflection. Second, this thesis proposed utilizing these cognitive and affective terms as core indicators in ML algorithms, which would be compared with different pre-trained language models to create a more transparent and effective educational tool. This approach is expected to facilitate deeper learning through precise feedback and provide practical pedagogical support for pre-service teachers.
On the downside, it is essential to recognize that the current results are still an initial nature. Firstly, the content of AI-based feedback must be further developed and optimized to meet the specific needs of students, particularly in terms of providing targeted advice. Secondly, the effectiveness of this feedback urgently requires systematic validation and evaluation, particularly concerning their practical utility in enhancing students’ reflection. Despite these limitations, these initial findings offer valuable insights into AI’s application in education and inspire other researchers and practitioners to explore new research and practice avenues. They establish a solid foundation for subsequent studies, which augurs well for the potential of AI in educational technology and its extensive impact on educational practices.
References


53


**Development: The Role of Smart Technologies in Achieving Development Goals** (pp. 387-409). Cham: Springer Nature Switzerland. [http://dx.doi.org/10.2139/ssrn.4354422](http://dx.doi.org/10.2139/ssrn.4354422)


62


## Appendix I: AI Acceptance Scale

Table 5 AI Acceptance Scale (German version)

<table>
<thead>
<tr>
<th>Construct</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perceived Usefulness</strong></td>
<td><strong>PU1</strong> Die Nutzung von KI-basierten Systemen erhöht meine Produktivität.</td>
</tr>
<tr>
<td></td>
<td><strong>PU2</strong> Die Nutzung von KI-basierten Systemen verbessert meine Effektivität</td>
</tr>
<tr>
<td></td>
<td><strong>PU3</strong> Die Nutzung von KI-basierten Systemen empfinde ich für mich als nützlich.</td>
</tr>
<tr>
<td></td>
<td><strong>PU4</strong> Die Nutzung von KI-basierten Systemen unterstützt mich bei der Erfüllung meiner Aufgaben.</td>
</tr>
<tr>
<td><strong>Perceived Ease of Use</strong></td>
<td><strong>PEOU1</strong> KI-basierte Systeme empfinde ich als nutzerfreundlich.</td>
</tr>
<tr>
<td></td>
<td><strong>PEOU2</strong> Die Bedienung eines KI-basierten Systems erfordert keine hohe geistige Anstrengung</td>
</tr>
<tr>
<td></td>
<td><strong>PEOU3</strong> Ich empfinde es als einfach, bei KI-basierten Systemen das umzusetzen, was ich gerade tun möchte.</td>
</tr>
<tr>
<td></td>
<td><strong>PEOU4</strong> Die Bedienung eines KI-basierten Systems ist klar und verständlich.</td>
</tr>
<tr>
<td><strong>AI Self-Efficacy</strong></td>
<td><strong>AISE1</strong> Ich kann KI-basierte Systeme problemlos nutzen, auch wenn niemand mir gesagt hat/hätte, was ich tun muss.</td>
</tr>
<tr>
<td></td>
<td><strong>AISE2</strong> Ich kann KI-basierte Systeme problemlos nutzen, auch wenn ich nur eine integrierte Hilfe-Funktion hätte.</td>
</tr>
<tr>
<td></td>
<td><strong>AISE3</strong> Ich kann KI-basierte Systeme problemlos nutzen, auch wenn mir niemand gezeigt hat/hätte, wie man das KI-basierte System nutzt.</td>
</tr>
<tr>
<td></td>
<td><strong>AISE4</strong> Ich kann KI-basierte Systeme problemlos nutzen, auch wenn ich nicht bereits ähnliche KI-basierte Systeme genutzt habe/hätte.</td>
</tr>
<tr>
<td><strong>AI Anxiety</strong></td>
<td><strong>AI1</strong> Mit KI-basierten Systemen zu tun zu haben macht mich nervös</td>
</tr>
<tr>
<td></td>
<td><strong>AI2</strong> KI-basierte Systeme wecken bei mir ein Gefühl des Unwohlseins.</td>
</tr>
<tr>
<td></td>
<td><strong>AI3</strong> KI-basierte Systeme lösen bei mir Anspannung aus.</td>
</tr>
<tr>
<td><strong>Perceived Enjoyment</strong></td>
<td><strong>PE1</strong> Mir bereitet es Vergnügen, KI-basierte Systeme zu nutzen.</td>
</tr>
<tr>
<td></td>
<td><strong>PE2</strong> Die aktuelle Aufbereitung von KI-basierten Systemen ist angenehm</td>
</tr>
<tr>
<td>PE3</td>
<td>Mir macht es Spaß mit Hilfe von KI-basierten Systemen zu lernen.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>Subjective Norm</strong></td>
<td>SN1</td>
</tr>
<tr>
<td></td>
<td>SN2</td>
</tr>
<tr>
<td><strong>Job Relevance</strong></td>
<td>JR1</td>
</tr>
<tr>
<td></td>
<td>JR2</td>
</tr>
<tr>
<td></td>
<td>JR3</td>
</tr>
<tr>
<td><strong>Behavioral Intention</strong></td>
<td>BI1</td>
</tr>
<tr>
<td></td>
<td>BI2</td>
</tr>
</tbody>
</table>

*Note. The five-point Likert scale ranged from 1 (strongly disagree) to 5 (strongly agree).*
Appendix II: List of All Publications

(*Articles for doctoral dissertation)


X
## Appendix III: List of Publications for Dissertation

### Publication P1: AI Acceptance Among Pre-Service Teachers

<table>
<thead>
<tr>
<th>Title</th>
<th>Acceptance of artificial intelligence among pre-service teachers: a multigroup analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Chengming Zhang, Jessica Schießl, Lea Plößl, Florian Hofmann, Michaela Gläser-Zikuda</td>
</tr>
<tr>
<td>Link</td>
<td><a href="https://doi.org/10.3390/educsci13121213">https://doi.org/10.3390/educsci13121213</a></td>
</tr>
<tr>
<td>Ranking</td>
<td>JCR: Q1</td>
</tr>
<tr>
<td>2-year Impact Factor: 8.6</td>
<td></td>
</tr>
<tr>
<td>5-year Impact Factor: 9.4</td>
<td></td>
</tr>
<tr>
<td>Abstract</td>
<td>Over the past few years, there has been a significant increase in the utilization of artificial intelligence (AI)-based educational applications in education. As pre-service teachers’ attitudes towards educational technology that utilizes AI have a potential impact on the learning outcomes of their future students, it is essential to know more about pre-service teachers’ acceptance of AI. The aims of this study are (1) to discover what factors determine pre-service teachers’ intentions to utilize AI-based educational applications and (2) to determine whether gender differences exist within determinants that affect those behavioral intentions. A sample of 452 pre-service teachers (325 female) participated in a survey at one German university. Based on a prominent technology acceptance model, structural equation modeling, measurement invariance, and multigroup analysis were carried out. The results demonstrated that eight out of nine hypotheses were supported; perceived ease of use ($\beta = 0.297^{<em><strong>}$) and perceived usefulness ($\beta = 0.501^{</strong></em>}$) were identified as primary factors predicting pre-service teachers’ intention to use AI.</td>
</tr>
</tbody>
</table>
Furthermore, the latent mean differences results indicated that two constructs, AI anxiety \( z = -3.217^{**} \) and perceived enjoyment \( z = 2.556^* \), were significantly different by gender. In addition, it is noteworthy that the paths from AI anxiety to perceived ease of use \( p = 0.018^* \) and from perceived ease of use to perceived usefulness \( p = 0.002^{**} \) are moderated by gender. This study confirms the determinants influencing the behavioral intention based on the Technology Acceptance Model 3 of German pre-service teachers to use AI-based applications in education. Furthermore, the results demonstrate how essential it is to address gender-specific aspects in teacher education because there is a high percentage of female pre-service teachers, in general. This study contributes to state of the art in AI-powered education and teacher education.

Table 6 Summary of publication P1
**Publication P2: Mixed Methods for Analyzing Reflective Writing**

<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Evaluating Reflective Writing in Pre-Service Teachers: The Potential of a Mixed-Methods Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Chengming Zhang, Jessica Schießl, Lea Plößl, Florian Hofmann, Michaela Gläser-Zikuda</td>
</tr>
<tr>
<td><strong>Link</strong></td>
<td><a href="https://doi.org/10.1186/s41239-023-00420-7">https://doi.org/10.1186/s41239-023-00420-7</a></td>
</tr>
<tr>
<td><strong>Ranking</strong></td>
<td>CiteScore: Q1</td>
</tr>
<tr>
<td></td>
<td>2-year Impact Factor: 3.0</td>
</tr>
<tr>
<td></td>
<td>5-year Impact Factor: 2.8</td>
</tr>
<tr>
<td><strong>Abstract</strong></td>
<td>Reflective writing is a relevant aspect of pre-service teachers’ professionalization. Evaluating reflective writing in teacher education is demanding due to a shortage of resources. Hence, this study explores the practical possibilities of evaluating reflective writing using a mixed-methods approach to analyze reflective writing from 198 pre-service teachers at a German university. We used qualitative content analysis, computational linguistic approaches, and BERTopic. Results of qualitative content analysis results indicated primarily descriptive and low-level participants’ reflective writing. Next, computational linguistic analyses revealed that affective and cognitive terminology utilization differed across varying levels of reflection, with a higher frequency of such terms correlating with deeper levels of reflection. BERTopic results showed that reflective content mainly centered on learning materials and shifted toward affective and motivational themes related to higher levels of reflection. This study demonstrates that reflective writing can be evaluated across reflection levels and cognitive, affective, and thematic dimensions, combining qualitative content analysis, computational linguistic approaches, and BERTopic.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Table 7 Summary of publication P2</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>XIII</td>
</tr>
</tbody>
</table>
**Publication P3: Developing AI-Driven Feedback**

<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Classification of reflective writing: A comparative analysis with shallow machine learning and pre-trained language models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Chengming Zhang, Florian Hofmann, Lea Plößl, Michaela Gläser-Zikuda</td>
</tr>
<tr>
<td><strong>Link</strong></td>
<td><a href="https://doi.org/10.1007/s10639-024-12720-0">https://doi.org/10.1007/s10639-024-12720-0</a>.</td>
</tr>
<tr>
<td><strong>Ranking</strong></td>
<td>JCR: Q1</td>
</tr>
<tr>
<td></td>
<td>2-year Impact Factor: 5.5</td>
</tr>
<tr>
<td></td>
<td>5-year Impact Factor: 5.3</td>
</tr>
<tr>
<td><strong>Abstract</strong></td>
<td>Reflective writing holds critical importance, for example, in higher education and teacher education, yet promoting students’ reflective skills has been a persistent challenge. The emergence of revolutionary artificial intelligence technologies, notably in machine learning and large language models, heralds potential breakthroughs in this domain. The current research on analyzing reflective writing hinges on sentence-level classification. Such an approach, however, may fall short of providing a holistic grasp of written reflection. Therefore, this study employs shallow machine learning algorithms and pre-trained language models, namely BERT, RoBERTa, BigBird, and Longformer, with the intention of enhancing the document-level classification accuracy of reflective writings. A dataset of 1,043 reflective writings was collected in a teacher education program at a German university (( M = 251.38 \text{ words}, \ SD = 143.08 \text{ words} )). Our findings indicated that BigBird and Longformer models significantly outperformed BERT and RoBERTa, achieving classification accuracies of 76.26% and 77.22%, respectively, with less than 60% accuracy observed in shallow machine learning models. The outcomes of this study contribute to refining document-level classification of reflective writings and</td>
</tr>
</tbody>
</table>

XIV
have implications for augmenting automated feedback mechanisms in teacher education.

Table 8 Summary of publication P3
Eidesstattliche Erklärung


Nürnberg, den 04.05.2024

Chengming Zhang