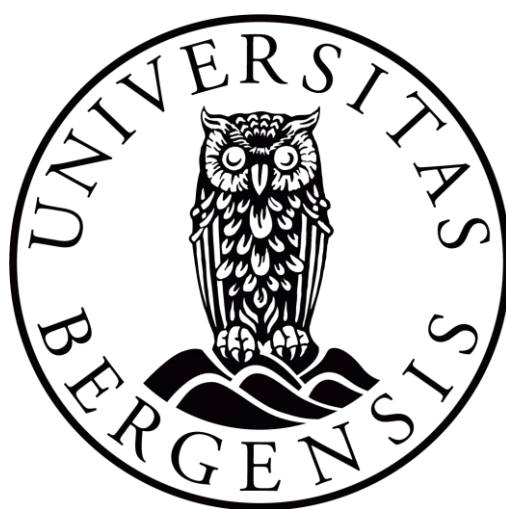


Large Language Models in L2 English Upper
Secondary Teaching: Exploring Teachers'
Understanding and Utilisation



Master's Thesis in English Didactics

Department of Foreign Languages

University of Bergen

Bendik Vik

May 2023

Abstract in Norwegian

Denne masteroppgåva har hatt som mål å undersøkje korleis lærarar forstår store språkmodellar (large language models, LLMs), korleis lærarane brukar desse i engelskundervisninga si, samt kva didaktiske val som rettleiar lærarane i integreringa av desse språkmodellane. Læreplanen i engelsk (LK20) legg vekt på at digitale ferdigheiter handlar om å kunne bruke digitale medium og resursar. Det er difor av stor interesse å undersøkje korleis LLMs vert nytta frå lærarane sin ståstad. Gjennom ei kvalitativ tilnærming ved bruk av djupneintervju, har fem engelsklærarar i vidaregåande skular i Noreg delt sine meiningar og praksisar. Eit teoretisk rammeverk for profesjonsretta digital kompetanse, saman med anna relevant teori som til dømes profesjonsutvikling, har blitt presentert i studien. Funna er følgeleg analysert i lys av dette teoretiske rammeverket og tidlegare forskning som kan knytast til funna. Studien syner mellom anna at LLMs i stor grad vert korrekt forstått av lærarane. Samstundes har dei samanfallande tankar og erfaringar kring etiske utfordringar ved bruken av desse modellane. Vidare syner studien at lærarane nyttar LLMs i ulike situasjonar i engelskundervisninga. Nokre av undervisningsopplegga lærarane viser til, fokuserer på skriving, kritisk tenking og refleksjon. Det kjem også fram at LLMs vert tatt i bruk som eit verktøy til bruk utanfor eiga undervisning.

Vidare syner analysen at medan lærarane aktivt brukar LLMs, så ligg det i stor grad føre ei integrering utan eit klårt fagdidaktisk grunnlag. Mangelen på ei fagdidaktisk operasjonalisering kan vidare stille spørsmål om legitimeringa av bruken av LLMs. Dette kan følgeleg føre til utfordringar med å balansere bruk og nytte av teknologien. Konsekvensen kan vere at det vert utfordrande å integrere slik digital teknologi på ein måte som tek vare på etiske omsyn, personvern og pedagogiske og fagdidaktiske rettesnorar, som alle bidreg til å forme engelskundervisninga. Kunstig intelligens er ikkje eit nytt fenomen i utdanningssektoren, men med framspringet av ChatGPT, har det synt seg viktigare enn nokon gong å reflektere dugeleg over eigen og andre sin bruk av denne typen teknologi. LLMs har til no vore relativt lite utforska i ein engelskfagleg kontekst i vidaregåande skule, og vidare forskning er imperativt for å forstå omfanget av både fordelar og negative konsekvensar ved bruken av dei.

Acknowledgements

First, I wish to express my gratitude to the participating teachers in this thesis, both in the interviews and the pilot stages. Without your insights and contributions, completing this thesis would have been difficult. I am grateful that you found the time in your busy schedules to participate in this study. Your valuable insights into your teaching made this research possible.

Furthermore, I want to thank my supervisor Kimberly Marie Skjelde. Your feedback, comments, and relaxing supervisions guided (and sometimes carried) me through what unsurprisingly turned out to be a very challenging project.

I would also like to direct the attention towards my fellow students. Since the first day we met five years ago, you have truly become a big part of my life. Thank you for the lengthy and enjoyable lunches we've shared, and thanks for enduring my awfully bad jokes.

Last but certainly not least, I owe deep appreciation to my family and friends. Your constant support and encouraging words have truly been invaluable and cannot be underestimated.

Bergen, May 2023

Bendik Vik

Table of Contents

Abstract in Norwegian	II
Acknowledgements	III
List of Abbreviations	VII
List of Tables	VIII
List of Figures	IX
List of Appendices	X
1 Introduction	1
1.1 Research Gap and Previous Research	1
1.2 Defining Artificial Intelligence	3
1.3 Large Language Models	4
1.4 Aims and Research Questions	5
1.5 Outline of Thesis	6
2 Theory and Background	8
2.1 Chapter Outline	8
2.2 Teacher Cognition	8
2.2.1 <i>Teacher Beliefs</i>	9
2.3 Teacher Professional Development	11
2.4 Professional Digital Competence	13
2.4.1 <i>Digital Didactics</i>	16
2.4.2 <i>Professional Digital Competence Framework for Teachers</i>	17
3 Methods and Materials	20
3.1 Chapter Outline:	20
3.2 Research Design	20
3.2.1 <i>Rationale for Choosing Qualitative Research Design</i>	21
3.3 Materials	22
3.3.1 <i>Designing the Interview Guide</i>	22

3.4	Participants and Recruitment	24
3.5	Conducting the Interviews	26
3.6	Data Analyses	27
3.6.1	<i>Transcribing the Interviews</i>	27
3.6.2	<i>Analysis of Interview Data</i>	28
3.7	Reliability	30
3.8	Validity	32
3.9	Ethical Considerations	34
4	Findings and Discussion	37
4.1	Chapter Outline	37
4.2	Teachers' Knowledge of LLMs	37
4.3	Implications of Using LLMs	39
4.3.1	<i>LLMs for Student Support</i>	39
4.3.2	<i>On Problems Integrating LLMs in Classroom Activities</i>	40
4.3.3	<i>Ethical Issues</i>	41
4.4	Classroom Activities Integrating LLMs	43
4.4.1	<i>LLMs as Tools for Writing</i>	43
4.4.2	<i>Adapting Learning Materials</i>	46
4.4.3	<i>Teacher-Pupil Interactions Discussing LLMs</i>	48
4.4.5	<i>LLMs as Time-Saving Mechanisms</i>	51
4.5	Didactical Considerations for Integrating LLMs	54
4.5.1	<i>Curiosity and Interest as Motivations for Integrating LLMs</i>	55
4.5.2	<i>Didactical Considerations in Practice</i>	58
4.5.3	<i>Absence of Didactical Considerations</i>	60
4.6	Limitations	63
4.7	Suggestions for Further Research	64
5	Conclusions	67

5.1 Revisiting the Research Questions	67
5.2 Didactical Implications.....	69
7. References	71
8. Appendices	83

List of Abbreviations

AI: Artificial intelligence

GPT: Generative, pre-trained, transformer

I: Interviewee

ICT: Information and communication technology

L1: First language

L2: Second language

LK20: The English subject curriculum

LLMs: Large language models

PDC: Professional digital competence

PfDK: Professional digital competence framework for teachers

R: Researcher

RETTE: Risiko og ETTErlevelse i forskningsprosjekter

T1: Teacher 1

T2: Teacher 2

T3: Teacher 3

T4: Teacher 4

T5: Teacher 5

TPACK: Technological Pedagogical Content Knowledge framework

List of Tables

Table 1.....	26
Table 2.....	29
Table 3.....	38
Table 4.....	43
Table 5.....	55

List of Figures

Figure 1..... 29

List of Appendices

Appendix A Interview guide (English Version).....	83
Appendix B Interview guide (Norwegian Version).....	87
Appendix C Information to participants	90
Appendix D Consent form	95
Appendix E Approval for using private recording equipment	96
Appendix F Examples of participation requests.....	97

1 Introduction

In November 2022, the world witnessed the significant advancements in language technology with the introduction of ChatGPT by OpenAI, a leading American artificial intelligence organisation. ChatGPT, essentially a large language model (LLM), quickly generated widespread discussion and analysis, particularly in the field of education. LLMs like ChatGPT are capable of generating seemingly factual information with a high degree of confidence, which has led to a multitude of apprehensions across a wide range of sectors. In the field of education, the integration of these expanding LLMs has sparked a growing discourse concerning the potential opportunities and challenges of utilising these models. Notably, the popularity of ChatGPT continues to rise every day. This trend is also evident in Norwegian contexts. According to a report by Kantar (2023), approximately 20% of people aged 15 to 24 frequently use ChatGPT for school related activities on a weekly basis. Furthermore, a survey conducted by Møgelvang et al. (2023) revealed varying attitudes among students and instructors at the Faculty of Mathematics and Natural Sciences at the University of Bergen. Despite being conducted shortly after the launch of ChatGPT, the survey indicated that instructors held significantly more pessimistic attitudes regarding the future implications of LLMs compared to students. Additionally, instructors reported limited enhancements in their task completion abilities as a result of LLMs integration (Møgelvang et al., 2023, pp. 9-10).

Concerning LLMs in education from a teacher's perspective, two essential questions arise: How do L2 English teachers integrate LLMs into their teaching, and to what extent do they validate this integration within the framework of English didactics? These are inquiries this study aims to investigate.

1.1 Research Gap and Previous Research

The integration of artificial intelligence (AI) in education is still slow compared to other fields, such as finance or engineering, and consequently, less studies considering the use of LLMs in education have been conducted (Salas-Pilco et al., 2022). A recent review of opportunities and challenges of LLMs in education highlighted that studies related to LLMs in education are still in an early stage, with few empirical studies investigating the use of effective learning designs and learning strategies (Hwang & Chang, 2023). From a teachers' perspective, research studies concerning LLMs are therefore often related to AI technologies in general. Kasneci et al. (2023) presented an overview on the current state of LLMs and their

applications. In their research, they highlighted a selection of research works addressing teachers' perspective on LLMs in education.

In a pilot study, Polak et al. (2022) found that the teachers from this study seemed to have a basic level of digital skills, but a low level of skills concerning AI. Along the same lines, a study of Nigerian teachers emphasised that motivation and readiness are key prerequisites for integrating AI technologies in education (Ayanwale et al., 2022). Additionally, perceived usefulness, perceived ease of use, and perceived trust in AI technologies are factors to consider when integrating AI technologies such as LLMs (Kasneci et al., 2023). Concerning teachers' different uses of LLMs, recent works have focused on assessment of student answers, adaptive feedback, and the generation of learning materials (Kasneci et al., 2023). Moore et al. (2022) demonstrated the effectiveness of a finetuned GPT3-model in evaluating student-generated answers, suggesting that LLMs could assist teachers in assessing the quality of students' texts. Additionally, the use of natural language processing models for generating adaptive feedback have found positive feedback on learning outcomes among students (Bernius et al., 2022; Sailer et al., 2023; Zhu et al., 2020). Moreover, Sarsa et al. (2022) highlighted the ability of AI technologies to generate programming tasks and provide automated feedback, while Qu et al. (2021) proposed a framework for automatically generating question-answer pairs for teaching materials. LLMs have also been utilised to create answers for multiple-choice questions (Raina & Gales, 2022; Rodriguez-Torrealba et al., 2022). Finally, Tack and Piech (2022) investigated the performance of conversational agents in educational dialogues, highlighting their potential but emphasising the need for further research to match human performance in assisting students effectively.

Moreover, Zawacki-Richter et al. (2019) conducted a systematic review of AI in higher education in the period of 2007-2018. This review study concluded that there is a weak connection between AI in education and pedagogy, a lack of critical reflection on positive outcomes and challenges of using AI in education, and that ethical concerns should be given substantially more attention (Zawacki-Richter et al., 2019). Another review study was conducted by Crompton and Burke (2022), monitoring the period 2016-2022. This review study concluded that there were four competence areas of AI that were central: Assessment and evaluation, predictive models, AI assistant, intelligent tutoring systems, and organising students' learning (Crompton & Burke, 2022).

Research studies on AI, as noted by (Krumsvik, 2023), have been retrospective in nature. Elstad (2023) propose similar views on this matter. The technology itself drives the

research, rather than the research driving advancements in the technology. Consequently, the rapid development of this type of technologies has resulted in humans becoming victims of these technologies without understanding the consequences of it (Krumsvik, 2023, p. 35). Technological tools are a huge part of our society and are used in our everyday life and in educational contexts. One of the issues that occur is that we use these technologies without knowing the full extent of consequences and positive outcomes. Moreover, identifying AI literacy has become important (Skulstad & Touileb, 2024)

The examples of some of the previous research done concerning LLMs in education indicate that there are still many knowledge gaps and uncertainties concerning successful and responsible integration of LLMs into teaching and learning processes. As noted by Kasneci et al. (2023), some of those knowledge gaps are concerned with customising models to specific needs, addressing biases in specific teaching contexts and dealing with ethical considerations and copyright issues. Addressing this pool of knowledge gaps and uncertainties requires “multidisciplinary evidence-based research and evaluation” (Kasneci et al., 2023, p. 4). In the field of education, LLMs stand at a crossroads where convenience, anticipations, and responsibilities converge (Skulstad & Touileb, 2024). At this crossroads, it seems important for teachers, pupils, and researchers to contribute to a more fundamental understanding of potential opportunities and challenges posed by LLMs, especially within the context of upper secondary schools. In light of these gaps, the purpose of this thesis is to investigate how L2 English teachers understand and utilise LLMs in the context of upper secondary schools in Norway.

1.2 Defining Artificial Intelligence

Digital skills are one of the four basic skills in the English subject curriculum of 2020 (Directorate for Education and Training, 2019), but was introduced firstly already in 2006 (Norwegian Ministry of Education and Research, 2006). Although not specifically mentioned in the definition of digital skills, AI technologies are included in the definition of digital skills (Skulstad & Touileb, 2024). For the purpose of this thesis and artificial intelligence in education, the following definition of AI is suggested:

Artificial intelligence (AI) systems act in the physical or digital dimension by perceiving their environment, processing and interpreting information and deciding the best action(s) to take to achieve the given goal. Some AI systems adapt their behaviour

by analysing how the environment is affected by their previous actions. (Norwegian Ministry of Local Government and Modernisation, 2020, p. 9)

This definition provided by Norwegian Ministry of Local Government and Modernisation (2020) suggests one general understanding of what AI entails. Moreover, Strümke (2023) noted that AI constitutes a field within the realm of computer science, directed towards the development of machines capable of exhibiting intelligent behaviour (p. 41). This approach could be perceived as far more operationalised, as it creates a link between human capabilities and intelligence with machines (Krumsvik, 2023). AI in education represents continuous and revolutionising technology, and the potential for AI in education have received increased attention (Holmes & Tuomi, 2022).

Defining AI in this study is relevant to provide a contextual understanding. Drawing upon the definitions of AI by the Norwegian Ministry of Local Government and Modernisation (2020) and Strümke (2023), AI is a broad field encompassing various subfields and technologies. Clearly defining the term AI helps understand the scope and focus of the current thesis. Moreover, conducted research on the field often focus on a general view on AI in education, rather than scoping specifically on the various subfields. It is therefore useful to provide a definition of AI to serve as a foundation for establishing the scope of the study's focus. With the establishment of a definition of AI, the subsequent focus shifts towards a specific subfield within the broader realm of AI, namely large language models.

1.3 Large Language Models

Large language models (LLMs) are in essence language machines with a transformer-based architecture using Reinforcement Learning with Human Feedback (RLHF) (Christiano et al., 2017). RLHF can simply be described as an approach where humans provide feedback when teaching a machine learning model to perform a task (Skulstad & Touileb, 2024). LLMs gained initial recognition in 2019 with the release of GPT-2. Subsequently, the introduction of GPT-3 occurred in 2020, followed by the emergence of perhaps the most renowned version, ChatGPT (GPT3.5), in November 2022 (Krumsvik, 2023, p. 17). The acronym GPT stands for *generative, pre-Trained, transformer*, highlighting key characteristics of the model: its capacity to generate text (generative), its training on extensive textual data (pre-trained), and its utilisation of transformer architecture to comprehend and produce text through acquired patterns from the textual data (transformer) (Strümke, 2023). Transformer architecture employs mechanisms of self-attention to effectively analyse the relevance of input data and

then produces an output sequence (Vaswani et al., 2017). The use of transformer architecture has significantly developed the field of Natural Language Processing (NLP) and has proven to be effective for analysing long sequences of words in texts. The self-attention mechanism identifies the most important parts of a text to generate suitable predictions such as the next possible words in a sequence. (Vaswani et al., 2017). In essence, LLMs are trained with a huge amount of data to comprehend and generate human language. By learning language patterns and human like communication, LLMs become models that can generate answers, solve tasks, and write coherent texts (Krumsvik, 2023).

The history LLMs, or language modelling, goes back some 70 years. In 1948, Claude Elwood Shannon investigated how to estimate word sequences using *n-grams* (Shannon, 1948). The N-gram language model essentially predicts “the most likely next word based on fixed numbers of preceding (*n*) words and actual word occurrences and count in some corpus” (Skulstad & Touileb, 2024). In essence, the language model could predict the next word in a sentence without any lexical or syntactical knowledge. Recent language models are substantially more complex than Shannon’s model due to the enormous training data, access to contextual information for each word in a sequence, and most importantly the RLFH (Skulstad & Touileb, 2024). Consequently, LLMs have impressive capabilities in language generation, such as answering questions, producing summaries, or write coherent texts. What is challenging is that LLMs generate with such confidence which results that non-experts or children may be misled by the generated texts. Therefore, one can assume that the capabilities of LLMs is a contemporary challenge for educational institutions on all levels in a variety of ways. These challenges pertain everyone in the field of education and the integration of them into educational contexts requires consideration of both the potential challenges and opportunities (Skulstad & Touileb, 2024).

1.4 Aims and Research Questions

The interest and motivation for conducting this study was to address and shed light upon the capabilities of language technology and machine learning that somewhat abruptly affected the field of education. Therefore, the purpose of this study is to understand how LLMs are understood and utilised among teachers within L2 English classrooms in Norway. More importantly, the study seeks to enhance didactical considerations of integrating this technology. The English subject curriculum of 2020 (LK20) highlight digital skills as one of four basic skills. Hereunder, the curriculum specifies the ability to use “digital media and resources” (Directorate for Education and Training, 2019, p. 4). Furthermore, it emphasises

that digital skills also requires “critical and reflected behaviour using digital forms of expression in English and in communication with others” (Directorate for Education and Training, 2019, p. 4). Although AI technologies are not explicitly mentioned in the definition of digital skills, they are nevertheless included (Skulstad & Touileb, 2024). AI technologies will presumably be a part of our lives both in workplaces and our everyday lives. Consequently, AI technologies may increasingly affect the field of education. Therefore, investigating how teachers operationalise LLMs seems to be of interest to the field of education. This thesis seeks to explore the phenomenon of LLMs in the EFL classroom through L2 English teachers’ beliefs, utilisations, and didactical considerations. The following research questions are proposed:

1. What do L2 English teachers in upper secondary schools report about their understanding and beliefs of the functionality and implications of large language models?
2. What specific classroom activities do L2 English teachers in upper secondary schools report when integrating large language models into their teaching?
3. What didactical considerations do L2 English teachers in upper secondary schools report guiding their decisions when integrating large language models into their teaching?

The research questions will be explored through a qualitative research design that includes the use of in-depth interviews to shed light upon teachers’ beliefs and experiences with LLMs, as well as how contemporary English didactics guide their integration of named models.

1.5 Outline of Thesis

This thesis consists of five chapters. Following the introduction, the second chapter (Chapter 2: Theory and Background) will present theory and research relevant for the context of the current study. The third chapter (Chapter 3: Methods and Materials) describes the methods and materials used to collect data material. Furthermore, the chapter addresses reliability, validity, and ethical considerations of the study. The penultimate chapter (Chapter 4: Findings and Discussion) will present the main findings concerning the research questions. Furthermore, the findings will be discussed in relation to the theoretical framework presented in Chapter 2, as well as relevant research studies. The study’s limitations and suggestions for further research will also be addressed at the end of the fourth chapter. The last chapter

(Chapter 5: Conclusions) provides a short summary of the thesis' findings and addresses possible implications for the context of L2 English teaching.

2 Theory and Background

2.1 Chapter Outline

The following chapter aims to provide a theoretical framework for the current study and a foundation for the analysis and discussion in Chapter 4. The selection of theoretical perspectives has been guided by both the research questions and insights from the teacher interviews. The chapter will begin by directing attention towards teacher cognition, encompassing teachers' beliefs, attitudes, and practices. Second, I will present theoretical perspectives on teacher professional development and professional digital competence. These two perspectives will be presented to contextualise the intertwined relationship between teacher development, digital competence, and the integration of LLMs in L2 English teaching. Lastly, governmental documents concerning digital competence will be presented to facilitate a discussion on the relationship between government-issued guidelines and contemporary utilisations of AI in L2 language education. Throughout the chapter, the theoretical perspectives will be contextualised concerning their relevance to the study.

2.2 Teacher Cognition

The purpose of this thesis is to research how teachers understand LLMs, in what classroom activities they integrate named models, and whether didactical considerations guide their choices of utilising them. One part of the study examines their experiences and beliefs about LLMs. It is thereby useful to define a theoretical framework concerning teacher cognition. Teachers are active, thinking decision-makers who have a prominent role in education. Borg (2015) defines teacher cognition as “what language teachers think, know and believe” (p.1). As noted by Borg (2019), teacher cognition encompasses a broad spectrum of aspects, including but not limited to teacher commitment (Al-Mahdy et al., 2018), teacher motivation (Richardson et al., 2014), and teacher identity (Burri et al., 2017). These aspects collectively contribute to what teachers “believe, think, perceive, or feel” Borg (2019, p. 1152). Moreover, research on teacher cognition has provided the concept with several distinctive terms that are intertwined and closely related to one another (Borg, 2015). Borg's definition to the term suggests that teachers are active and thinking individuals where decisions in their classroom practice are partially based on knowledge and experience. This broad definition suggests that “teachers have cognitions about all aspects of their work, and lists recurrent labels used to describe the various psychological constructs which I collectively refer to here as teacher cognition” (Borg, 2015, p. 48). Furthermore, Borg (2015) states that teacher cognition is a dynamic concept in constant development and that teacher cognition is

a “tacit, personally-held, practical system of mental constructs held by teachers and which are dynamic – that is defined and refined on the basis of educational and professional experiences throughout teachers’ lives” (Borg, 2015, p. 40). The mental constructs introduce a psychological approach with a set of concepts used in teacher cognition research (Borg, 2015, pp. 41-45). The different concepts, or labels, highlight teacher cognition as being complex, and Borg (2015, p. 47) defined a framework to conceptualise teacher cognition within teaching. The framework indicates that teachers have cognitions about all aspects of their work and display a wide-ranged set of characteristics which play a pivotal role in defining teacher cognition (Borg, 2015, p. 48). Additionally, the framework highlights the interconnected relationship between the schooling, professional coursework, contextual factors, and classroom practice. To operationalise the Borg’s framework (2015, p. 47) within the context of the current study, the focus will be on teacher beliefs. Section 2.2.1 delineates the concept of beliefs, examines challenges concerning teacher beliefs, and explaining its relevance to this study.

2.2.1 Teacher Beliefs

Within the teacher cognition framework provided by Borg (2015), the concept of teacher beliefs emerges. For over two decades, teacher beliefs have played a significant role in language teaching (Borg, 2017). This study aims to explore teachers’ beliefs regarding LLMs. Borg’s (2015) broad characterisation of teacher cognition mirrors the definition of teacher beliefs as being complex and not easily defined. There have been, however, numerous efforts to describe beliefs (Borg, 2017).

Skott (2014, pp. 18-19) delineates four foundational elements that highlight the different definitions of beliefs: they encompass ideas individuals consider to be true, they have cognitive and affective dimensions, they are stable and result from substantial social encounters, and they influence practice. Addressing that beliefs influence practice is of particular interest for this study. Skott (2014) notes that “beliefs are expected to significantly influence the ways in which teachers interpret and engage with the problems of practice” (p. 19). However, this claim has been challenged in empirical research. First, Stainton Rogers (2011) noted that beliefs are not good indicators of behaviour. Second, Borg (2017) suggests that the effects of beliefs may be incompatible, as teachers hold beliefs across various issues in education. Lastly, beliefs are challenging to observe, as they are often observed through examination of professed beliefs (what teachers say they believe) and enacted beliefs (beliefs in action) (Borg, 2017, p. 77).

Burns (1992, p. 58) describes beliefs as motivational forces driving teaching practices. Additionally, beliefs can be viewed as shaping teachers' teaching practice (Johnson, 1992), or guide teachers in how they think and behave (Borg, 2001). Consequently, beliefs emerge as constructs influenced by various cognitive processes such as interpretation and experience. In summary, beliefs are a concept that have received diverse definitions. Despite receiving diverse definitions, there is a consensus regarding beliefs' influence on teachers' viewpoints and judgements, thereby affecting their classroom practices (Pajares, 1992).

In conclusion, my research adopts the definition of teacher beliefs delineated by Skott (2014) and Borg (2001). Furthermore, I recognise the consensus among teachers, as noted by Pajares (1992), regarding the influential nature of teacher beliefs. Pajares' insights align with the foundational elements highlighted by Skott (2014) and Borg (2001), reinforcing the importance of understanding teacher beliefs as constructs that shape teacher practices. This study focuses on the relationship between teachers' beliefs concerning LLMs, and their integration of LLMs in their classroom practice, without delving into discrepancies between beliefs and practices. Based on the assertions above, I consider teachers' beliefs to encompass thoughts and ideas about all aspects of a teacher's work and consequently influence the teacher's practice. Additionally, the study seeks to explore multifaceted and nuanced aspects of teachers' utilisation of LLMs. Teachers' beliefs, that is, how they think, know, and believe, seem to have a significant role on their attitudes and behaviours toward integrating AI technology and LLMs into their teaching practices.

Teacher's beliefs are relevant to this study in several ways. Firstly, beliefs may affect teachers' actions in the classroom and also impact their awareness, attitudes, and teaching methods (Gilakjani & Sabouri, 2017, p. 78). Beliefs concerning LLMs and their choices for integrating LLMs in their teaching practice will thereby be relevant to examine. Second, beliefs guide teachers' decisions on lesson planning and shape their teaching practice according to the curriculum (Gilakjani & Sabouri, 2017, p. 80). Third, while possessing knowledge of technology is essential, it alone does not suffice if teachers lack the confidence to effectively utilise that knowledge in facilitating student learning (Ertmer & Ottenbreit-Leftwich, 2010, p. 261).

Although computer related variables, in general, continue to impact on teachers' ability to integrate technology, it is positive experiences with computers in the classroom context that build a teacher's belief in computer technology and confidence in its potential as an instructional tool. (Mueller et al., 2008, p. 1533)

The perceptive reader recognises that the reference above pertains to computer technology in general. Nevertheless, one could reasonably infer that this statement holds relevance within the specific context of LLMs and AI technology. While the underlying principles remain consistent, the context differs. Several suggestions for building technology confidence are presented in literature. For example, giving teachers time to interact with the technology seems to be useful for promoting technology confidence (Somekh, 2008). Additionally, there are research studies indicating that having small successful experiences, cooperating with knowledgeable peers, and participating in a professional learning community may promote confidence in using technology (Ertmer et al., 2006; Leftwich, 2007; Putnam & Borko, 2000).

2.3 Teacher Professional Development

Teacher professional development concerns how teachers learn to learn and how their knowledge is applied in practice (Avalos, 2011). From a teacher's perspective, professional development happens through a variety of forms such as participation in different courses, reflecting on their own teaching, observing other teachers' teaching, and formal and informal discussions of teaching practices with colleagues (Postholm, 2012). It is thereby reasonable to address teacher professional development as a multifaceted endeavour. To fund a theoretical framework for the current study, I will address perspectives from Postholm (2012). The main motivation for this is her opinions on professional development being connected to the cognitivist and constructivist paradigm. These paradigms highlight the learner in the learning process taking an active role (Postholm, 2012, p. 406). From qualitative analyses of what characterises good teachers, co-operating with teacher colleagues tend to increase competence and teachers' own professional development (Hagen & Nyen, 2009). The increase of competence and taking responsibility for professional development, is defined as 'adaptive exercise' (Hammerness et al., 2005). From these perspectives, co-operation with other teachers and learning both of and with them are valuable skills to have when promoting professional development.

From the theoretical review conducted by Postholm (2012), seven themes were developed to her findings in research on teacher' learning in school (pp. 412-419). For the current study, the following themes will be addressed: *Five characteristics of teachers' learning, individual and organisational factors in teachers' learning, and co-operation between external resource persons and teachers for teachers' learning*. First, she highlighted five characteristics for improving knowledge and skills of teachers, as identified by Desimone (2009): *content focus, active learning, coherence, duration, and co-operation*. This set of

characteristics may all be considered as crucial factors for teachers' professional development (Postholm, 2012). Second, concerning the importance of teacher autonomy, Postholm identified the themes *individual and organisational factors in teacher's learning* and *co-operation between external resource persons and teachers for teacher's learning*. The initial theme emphasised that teacher autonomy was crucial for the professional development of teachers, and that teacher autonomy was supported by teachers identifying and reflecting on their own learning objectives independently and in collaboration with other teachers (Postholm, 2012, pp. 412-413). The latter theme identified that professional teacher development had to come from teachers themselves (Postholm, 2012, p. 416). Hence, teachers must develop willingly and not feel obligated to do so. This approach thereby suggests that there must be an inner interest to develop, both on content focus and for professional development. Thirdly, co-operation between teachers within the same subjects and schools, and across different schools are important factors to consider regarding teachers' professional development (Postholm, 2012, p. 416). These aspects have proven to be essential as it promotes internal development, e.g., within the same school, and external collaboration between schools with often different teaching cultures. In this context, teaching cultures concern different views on teaching, varieties in how teachers co-operate. In essence, all teachers are obligated to follow the current subject curriculum (LK20), but the teacher autonomy in LK20 allows teachers to decide what methods they employ. Lastly, it seems crucial to establish a positive atmosphere and culture for self-development. Establishing a promoting learning culture for teachers' professional development is according to several research studies an important factor (Jurasaitė-Harbison & Rex, 2010; Kennedy, 2016; Rinke & Valli, 2010).

In relation to the current thesis, the use of LLMs in teaching correlates with the teacher professional development in a variety of ways. Firstly, looking at how teachers utilise LLMs may give indications of their professional development. Although professional development is a multifaceted concept, how they comprehend and use AI in their teaching practice may give indications about how teachers are willing to learn and adapt according to the themes identified by Postholm (2012). Moreover, focus on content means both "subject knowledge and knowledge about how pupils can acquire this knowledge" (Postholm, 2012, p. 412). Applying this perspective to the research questions, using LLMs to teach L2 English requires teachers to understand how these tools can be utilised from an English didactics' perspective. If the teachers utilise LLMs by being guided by English didactics, it may become easier to

legitimise the integration in their teaching practice. In this sense, investigating if there are didactical considerations guiding teachers' integration of language models will assumably reveal some novel aspects of their professional development. Secondly, how they co-operate with other teachers in the process of learning and utilising artificial intelligence tools is crucial as their professional development is highly dependent on which ways they collaborate with other teachers. Collaboration with other teachers may in this context be included in the second research question. This research question investigates which specific classroom activities they integrate LLMs in two ways. For one, it looks at which classroom activities they integrate them into. Second, it investigates whether the integration is an individual project, a project initiated by the school board or county, or a cross-school collaboration. Both these aspects are examined in this research question because how they utilise LLMs in their teaching practice might provide answers to how their teacher professionalism is developed.

2.4 Professional Digital Competence

Professional Digital Competence (PDC) has had an increased interest in Norwegian schools in recent years. This interest corresponds with a teacher role where the teacher is seen as an executor of the curriculum, a knowledge worker and designer of learning environments and trajectories (Lund, 2019, pp. 156-157). The term PDC was introduced in the Professional Digital Competence Framework for Teachers (see section 2.4.2) to give “substance and meaning to the concept of teachers' professional digital competence, and thereby establish a basis for competence development and improve the quality of this in the teaching profession” (Kelentrić et al., 2017, p. 2).

Whereas teacher professional development concerns how teachers learn to learn and how their knowledge is applied in practice (see section 2.3), professional digital competence relates to how these aspects are realised in digital platforms. In the educational context, digital competence represents a multifaceted angle of approach encompassing various dimensions, as mentioned by Krumsvik (2023). These dimensions include proficiencies in digital skills and knowledge, abilities of self-regulation, learning strategies, and elements pertinent to Bildung (Krumsvik, 2023, p. 268). Navigating in the digital world of 21st Century English teaching is thereby a challenging pathway.

Technology affects aspects both in our daily life and in education. The different ways in which we communicate and acquire knowledge may therefore depend on technological developments. In educational contexts, teachers play a crucial role. Ottestad et al. (2014, p.

248) proposed three aspects to describe teachers' PDC: *generic digital competence*, *didactic digital competence*, and *professional oriented digital competence*. How teachers understand and utilise technology in education are dependent on their digital competence and their ability to adapt the technology to subject matter. Thus, their digital competence consequently frames the productivity of these technologies. Lund et al. (2014) similarly argue that PDC involves teachers' and pupils' ability to make technologies appropriate and put them to productive use. They further underscore that technologies are understood differently depending on which school subject they are integrated in (Lund et al., 2014, p. 284). Norhagen et al. (2024) relate this to the didactics of each subject matter.

Although PDC is not an established concept or practice (Lund, 2019), Kelentrić et al. (2017) has tried to conceptualise PDC:

A professional, digitally competent teacher understands how digital developments are changing and expanding the content of subjects. The teacher understands how the integration of digital resources into learning processes can help to achieve competence aims in a subject, and to address the five basic skills. As a prerequisite for this, the teacher needs to develop their own digital skills. At the same time, the teacher needs to understand what pupils' digital skills entail, and how they can be fostered in the subjects. (Kelentrić et al., 2017, p. 4)

This definition of PDC underlines interconnection between teachers' and pupils' digital competence. Moreover, it fosters a responsibility for teachers to develop both their own and pupils' digital competence. This double role is emphasised in a research study conducted by Redecker (2017) where she advocated that teachers on the one hand have to guide students to develop their digital competence. Conversely, teachers must develop their digital knowledge to meet the needs of digital competence in 21st century teaching. This argument of teachers' double role also acknowledged by Pegalajar-Palomino (2018) where she argues that teachers' digital competence is crucial to improve their own professional development, which resultatively improves their students' learning processes. In essence, the digital competence of teachers is not solely complementary knowledge for teachers alone. It also represents a competence that directly benefits their students as well.

However, the definition provided by Kelentrić et al. (2017) raises questions concerning the terms 'skills' and 'competence'. Competence may be seen as a more comprehensive term than skills. It includes both the technical aspects related to hardware and

software, and cognitive aspects related to knowledge and education (Erstad, 2005). Skills may thereby be misleading to what the term entails, as the stated intentions behind the framework clearly points towards competences (Ottestad et al., 2014, p. 245). A definition of PDC that avoids the term ‘skill’ is the following proposed by Norhagen et al. (2024):

Professional Digital Competence is the comprehensive ability for teachers to effectively and responsibly use digital technologies in education, including integrating them into subject-specific teaching, designing educational activities, making informed decisions about tool use, guiding students in productive digital use, and considering societal and ethical implications of technology. (Norhagen et al., 2024, p. 5)

This definition is dynamic and can be shaped by the current state of knowledge and experiences from the educational field (Norhagen et al., 2024).

To meet the essential requirements to promote PDC among teachers, Rossi Cordero and Barajas Frutos (2018) highlighted teachers’ attitudes towards digital competence as crucial factors. The principal arguments of their research revolved around fostering constructive attitudes towards pedagogical change, coupled with effective implementation strategies within the school structure, as essential elements for successful promoting digital competence. Likewise, this perspective is reinforced by Postholm (2012) regarding professional development in a broader context (see section 2.3). In the pursuit of cultivating and developing teachers’ digital competence, it appears imperative that this pedagogical and didactical change relies upon commitment across various areas. Primarily, each teacher should exhibit a genuine wish for developing their own digital competence. As evidenced from Rossi Cordero and Barajas Frutos (2018) and Postholm (2012), this interest is not something to be coerced but rather to be inherently desired. Secondly, the interest in developing must come from an organisational structure as well. Thannimalai and Raman (2018) underscore the correlation between integrating technology among teachers and the pivotal role of leadership within the educational institutions. An absence of this mutual interest between teachers and institutions could affect development of PDC negatively. In essence, digital competence is, as pointed out by Fernández-Batanero et al. (2022) in their systematic review, “considered as a key factor to improve their professional development, enhancing their students’ teaching and learning processes” (p. 526). This relationship seems to require both teachers and educational institutions to focus on giving attention to technology from both a pedagogical and didactical point of view. Teachers must be included in the process of acquiring technological knowledge

and more importantly in the process of how these technologies are to be integrated in the teaching practice.

Within the scope of this thesis, the focus is on teachers' integration of AI technology, particularly LLMs, in their L2 English teaching. The assessment of their PDC is undertaken through the first research question, which investigates their understanding and beliefs of the functionalities and implications of LLMs. This investigation will, to some degree, reveal their digital competence concerning LLMs, yet the research question does not examine their digital competence in a broader context. Conversely, the second research question seeks to investigate how their digital competence is reflected through applying their competence in various instructional contexts. This inquiry encompasses not only individual classroom activities but also how they co-operate with other teachers and the educational institutions within which they operate. Lastly, the third research question revolves around elucidating the underlying motivations for implementing LLMs in their L2 English teaching. This line of inquiry seeks to reveal whether their implementation of LLMs is driven by a nuanced understanding of English didactics, or if it is merely a product of interest in the technology itself. Additionally, it will shed light on the nuanced connection between PDC and teaching practice decision-making.

2.4.1 Digital Didactics

For the current study, it is appropriate to introduce the concept *digital didactics*. Digital didactics is “an instructional theory of technology which puts the special focus towards the art of teaching in technology dense learning environments” (Krumsvik & Almås, 2009, p. 11). The theory focusses on relevant elements in technology-rich environments, emphasising that 21st century technology will dilute the effectiveness of teaching (Kongsgården & Krumsvik, 2019). A study conducted by Krumsvik et al. (2013) found that teachers' digital competence promoted pupils' learning outcomes. Additionally, the study found that teachers have general ICT-skills, but a lower pedagogical competency of ICT. These findings suggest that the biggest challenge for teachers is how to utilise ICT tools from a pedagogical point of view. Furthermore, Kongsgården and Krumsvik (2019) demonstrated that lesson design and teachers' didactical choices for utilising technology in the classroom were critical for promoting positive learning outcomes among upper secondary students. Concerning lesson design, the study also highlighted teachers' ability to develop good lessons plans and facilitate the learning processes in the classroom. Consequently, the teachers must operationalise their didactic choices by emphasising the use of technological tools as

cognitive aids for learners, rather than adopting a technological approach that solely prioritises access to technology (Dumont & Istance, 2010).

Operationalising digital didactics in this study is relevant as the study endeavours to shed light upon teachers' didactical considerations when integrating LLMs into their teaching practices. As noted by Dumont and Istance (2010), it is the didactics that should guide the use of technology and not the opposite. This approach requires teachers to have a comprehensive understanding of the curriculum, content knowledge, and understand how digital tools can be applied to promote learning outcomes (Bernacki et al., 2011). By emphasising digital didactics, this study acknowledges the important role of pedagogical and didactical strategies in using LLMs as educational resources. In essence, digital didactics, together with the *Professional Digital Competence Framework for Teachers* (see section 2.4.2) serve as a guiding framework for teachers as they integrate technology in their teaching practices. The interviews conducted in the study may thereby reveal whether teachers prioritise didactical considerations, or if there were other considerations guiding their integration.

2.4.2 Professional Digital Competence Framework for Teachers

As Professional Digital Competence (PDC) is a complex concept (see section 2.4), operationalising the term into L2 English teaching is of particular interest. In 2006, the *Technological Pedagogical Content Knowledge* (TPACK), a framework for describing “the type of teacher knowledge required to teach effectively with technology (Koehler et al., 2013, p. 2). In essence, TPACK highlights the intertwined relationship between technology, pedagogy, and content, and how these elements may guide the teacher to become professional digitally competent (Mishra & Koehler, 2006, p. 1017). TPACK guides teachers to understand certain topics for discussing educational technology in education. Additionally, the framework shows how content, pedagogy, and technology mutually influence each other (pp. 1045-1046). Moreover, research conducted in the United States using TPACK underscore that the intersection of pedagogical knowledge, content knowledge, and technology content must be considered to successfully integrate technology in teaching and learning (Norhagen et al., 2024).

In the context of Norwegian education, The Norwegian Centre for ICT in Education (an agency under the direct authority of the Norwegian Ministry of Education and Research) released the Professional Digital Competence Framework for Teachers (PfDK) in 2017. As

explicitly stated, “The Professional Digital Competence Framework for Teachers is a guidance document that...teachers...can use as a reference in their work on improving the quality of teacher education and systematic continuing professional development for teachers” (Kelentrić et al., 2017, p. 2). Furthermore, the framework attempts to illustrate the complexity of “knowledge, skills, and competencies in teachers’ professional practice that are associated with understanding the opportunities and challenges in today’s digital society” (Kelentrić et al., 2017, p. 2). In this context, the framework provides teachers with a shared conceptual reference point for understanding the implications of PDC (Norhagen et al., 2024). The PfDK was created in alignment with Norwegian national regulations for education, and draws upon international frameworks and evaluation tools for digital competence (Norhagen et al., 2024).

The framework can be utilised in three main domains: *developing common national frames and direction, planning and implementing initial and continuing teacher education, and evaluating and following up on teacher’s professional digital competence* (Kelentrić et al., 2017, pp. 2-3). The framework is essentially a guide including seven equally important areas, and the sum of the competence areas makes up a professional digitally competent teacher (Kelentrić et al., 2017, p. 3). Moreover, it emphasises a change of focus from product to process underscoring the importance of the process leading to a product rather than the product itself (Kongsgården & Krumsvik, 2019).

The competence areas especially relevant to this thesis are *ethics, pedagogy and subject didactics, and subject and basic skill*. Firstly, the framework states that “a professional, digitally competent teacher...has an insight into the legislation and ethical concerns, as well as development of pupil’s digital Bildung...in a digital and democratic society (Kelentrić et al., 2017, p. 6). In essence, teachers should acquire knowledge, skills and competencies concerning the ethical considerations surrounding digital aspects of teaching. In addition, teacher must address legal and ethical concerns, focussing on digital citizenship, privacy, and intellectual property (Norhagen et al., 2024). Secondly, the framework contains expectations for teachers to possess knowledge of “subject didactics relevant to the practice of their profession in a digital environment” (Kelentrić et al., 2017, p. 7). In line with this competence area, the teachers must employ a range of digital tools and materials, fostering reflective, engaging teaching practices (Norhagen et al., 2024). Having this knowledge, acquiring these skills, and possess this competence concerning subject didactics constitute a high degree of expectations for teachers to fulfil. It also constitutes a constant, critical evaluation and integration of digital teaching materials and resources. Lastly, a digital competent teacher

understands “how the integration of digital resources into learning processes can help to achieve competence aims in a subject, and to address the five basic skills” (Kelentrić et al., 2017, p. 4). Additionally, the teachers must enhance subject and basic skills through technological tools, intertwining digital literacy with reading, writing, numeracy, and oral skills. In a continuing process, the teacher must also develop their own and pupils’ digital abilities, aligning educational goals with digital fluency (Norhagen et al., 2024). As ‘digital skills’ is one of the basic skills in the English subject curriculum (Directorate for Education and Training, 2019, p. 4), an awareness of various forms of digital tools could contribute to PDC. Concerning the topic in this thesis, teachers’ PDC should include sufficient skills that allow them to critically utilise LLMs in a didactical and pedagogical manner.

3 Methods and Materials

3.1 Chapter Outline:

The primary objective of this study is to investigate the integration of LLMs in educational settings, particularly focusing on teachers' integration of this technology. To achieve this aim, the study addresses the following research questions:

1. What do L2 English teachers in upper secondary schools report about their understanding and beliefs of the functionality and implications of large language models?
2. What specific classroom activities do L2 English teachers in upper secondary schools report when integrating large language models into their teaching?
3. What didactical considerations do L2 English teachers in upper secondary schools report guiding their decisions when integrating large language models into their teaching?

The research questions examine teachers' understanding of LLMs, their application contexts, and how contemporary didactical considerations guide their utilisation. The chapter initiates by outlining the research design (section 3.2), explaining the rationale for adopting a qualitative approach in the current study (section 3.2.1). It proceeds to elaborate on the materials (section 3.3), the interview guide (section 3.3.1), participant recruitment (section 3.4), the teacher interviews (section 3.5) and the processes involved in the data analysis (section 3.6). Moreover, this chapter will attend to the reliability (section 3.7) and validity (section 3.8) of the current study. Lastly, I will encompass ethical and methodological considerations (3.9) concerning this research project.

3.2 Research Design

The current study is conducted as a qualitative study realised as in-depth interviews of L2 English teachers in the context of Norwegian upper secondary schools. Qualitative research is based on “qualitative data and tends to follow the exploratory mode of the scientific method” (Johnson & Christensen, 2020, p. 46). In qualitative research, there are many types of research designs depending on the type of study (Creswell & Creswell, 2023, p. 191). The research design I have chosen is a *phenomenological research design*, a research design that focusses on experiences of individuals about a specific phenomenon (Creswell & Creswell, 2023, p. 15; Johnson & Christensen, 2020, p. 48). In phenomenological research studies, it is common for researchers to gather data via in-depth interviews (Johnson &

Christensen, 2020, p. 425). In-depth interviews are one of the main methods of data collection used in qualitative research (Legard et al., 2003), and can be described as a form of conversation (Burgess, 2002). Unlike a normal conversation, however, the researcher in qualitative interviews should be the repository of detailed information and ask follow-up questions if greater clarity or depth is needed from the interviewee (Johnson & Christensen, 2020, p. 193). This type of interviews is useful when a researcher wants detailed information about a person's thoughts or want to explore an issue in depth (Boyce & Neale, 2006).

As emphasised by Johnson and Christensen (2020, p. 32), qualitative research is relevant to use when there exists a knowledge gap concerning a specific topic or phenomenon. One could therefore argue that it is crucial to “get close” to your objects to fill the gap about a phenomenon or topic when little is known about it. By conducting in-depth interviews with L2 English teachers, this research will examine the nuanced perspectives and personal encounters of teachers within the context of the Norwegian upper secondary school system. Moreover, the project aims to offer insight into various ways LLMs influence teaching practices and experiences of teachers, providing a rich qualitative understanding of a relatively unexplored terrain.

Like other research designs, qualitative studies also possess disadvantages. These disadvantages are explained in the sections below but essentially concerns aspects of reliability (section 3.5), and validity (section 3.6). These aspects are important to consider as they elucidate potential limitations of qualitative studies.

3.2.1 Rationale for Choosing Qualitative Research Design

I have chosen a qualitative research design realised as in-depth interviews of L2 English teachers for this study. The research questions guiding this study focus on understanding how teachers understand and use LLMs in their L2 English teaching. Qualitative research is selected to examine people's experiences and viewpoints, aiming to minimise the researcher's subjective meaning biases in interpreting the data (Creswell & Creswell, 2023, p. 193; Johnson & Christensen, 2020, p. 34). One of the advantages of qualitative research is the exploratory potential, as in-depth studies do not rely on previous literature or empirical data (Dörnyei, 2007, p. 39). Creswell and Creswell (2023, p. 21) similarly note that the qualitative method approach suitable when little research has been done on the field. LLMs have caused heated debates across the field of education about assessment and academic integrity since its release in November 2022 (Bozkurt, 2023; Krumsvik, 2023). However, the use of AI creates new possibilities for learning (Norwegian Directorate for

Education and Training, 2022). Furthermore, the impact of AI technologies such as LLMs on teaching experiences and learning potential are still at an early stage (Casal-Otero et al., 2023). Given the novelty and little scholarly attention paid to the topic of LLMs in L2 English teaching, I was motivated to explore this phenomenon in depth.

The accessibility of participants was an important consideration when choosing the research design. The narrow scope of the study limited the potential interviewees, resulting in a small sample size. The small sample size suggests various barriers explaining why few teachers seemed to be using LLMs actively in their teaching. Firstly, a lack of awareness of potential benefits of integrating LLMs in their teaching might be present. Furthermore, teachers may lack necessary training or guidance to effectively integrate LLMs. Secondly, at the time of the recruiting process, using LLMs actively in teaching was prohibited which could also explain why few teachers seemed to be using them (see section 3.3.2). Thirdly, teachers may have been reluctant to participate due to time constraints since teaching is a demanding profession and teacher may feel they lack the time and resources to experiment with LLMs. These delineated barriers describing why few teachers seemed to be using LLMs actively in their teaching are probable suggestions. Nevertheless, due to the limited sample population, gathering quantifiable data became consequently difficult. As a result, I found a qualitative driven research design most suitable. Conducting in-depth interviews allowed for a more thorough exploration of the topic and compensated for the limitations imposed by the limited sample size. Additionally, these interviews facilitated a nuanced investigation into the relationship between L2 English teachers' understanding and use of LLMs.

3.3 Materials

The results gathered from five teacher interviews form the empirical data material. All the teachers were asked to bring along an example of a lesson plan where they utilised LLMs. Only one teacher brought a written example. The remaining teachers detailed the lesson orally during the interview. Nevertheless, all the task examples from the lesson plans were gathered and included in the data material. The interviewees were all in-service teachers of L2 English in Norwegian upper secondary schools.

3.3.1 Designing the Interview Guide

Qualitative interviews consist of open-ended questions and presuppose that qualitative data is provided (Johnson & Christensen, 2020, p. 193). Qualitative interviews enable researchers to delve into the interviewee's subjective experiences, gaining a deeper

understanding of their beliefs. Although the process of conducting qualitative interviews is a time-consuming and costly approach, the one-on-one interviews provide the researcher with personal qualitative data. Establishing rapport and trust is consequently crucial, as it encourages open and honest responses during the interviews (Johnson & Christensen, 2020). Patton (1987) argues that the purpose of gathering responses to open-ended questions is to “permit the evaluator to understand and capture the perspective of program participants without predetermining their perspective through prior selection of questionnaire categories” (p.11). To ensure honest responses regarding participants’ beliefs, it is essential to maintain a neutral and unbiased stance throughout the study, refraining from posing leading questions or engaging in any form of coercion.

The method employed for gathering qualitative data through interviews involved the development of an interview guide, which is provided in Appendix A. An interview guide is essentially a list of pre-prepared inquiries consisting of open-ended questions (Johnson & Christensen, 2020, p. 194). For the qualitative interviews, I have therefore designed an interview guide structured in four main parts, eliciting the four different topics: *background information* (part one), *teachers’ understanding of language models* (part two), *teachers’ experiences and beliefs about using language models* (part three), and *teachers’ reports of their didactical considerations* (part four).

Part one of the interview guide is merely an introductory part where the teachers explained their formal qualifications and background as teachers. Part two examines how the participants understand LLMs. It specifically asks for their detailed understanding of what LLMs are and examines their general beliefs about LLMs. Questions concerning ethical concerns and the concept of digital Bildung are also included in this part. Part three examines what type of teaching practices they integrated LLMs in. It also invites the teachers to reflect on their use of LLMs in their lesson plan, as well as reflecting on positive and negative impacts of using LLMs in their teaching. The last part of the interview guide concerns their didactical reasoning for choosing to integrate LLMs in their teaching. Moreover, it includes questions examining their motivations for integrating LLMs and whether integrating LLMs could be didactically legitimised.

In this sense, the research questions frame the questions in the interview guide. The questions in each part of the interview guide are also to some degree intertwined. For example, “Define your understanding of language models and explain their relevance to language learning in your own words” (Question 2.1, see Appendix A), and “Based on your teaching experiences

with language models, share your insights on how language models can be integrated as constructive tools for language learning” (Question 3.6, see Appendix A). Question 3.6 can be seen as a continuum from question 2.1 as the teachers are first asked about LLMs relevance to language learning. In question 3.6, their general viewpoints are expanded into how LLMs can be integrated successfully in the classroom. The intertwined relationship between the research questions and the different parts of the interview guide was a conscious choice. It provided both detailed results as well as a wider perspective about the teachers’ relations to LLMs.

3.4 Participants and Recruitment

The research questions in the current study specify that the focus is on English teachers as the intended participants. Consequently, the sampling technique used is a *non-random sampling technique* (Johnson & Christensen, 2020, p. 253). Every participant must be L2 English teachers in an upper secondary school context to be available for selection. Moreover, each participant must have some experience with LLMs. Amount of experience, however, is lenient as LLMs in education represent a relatively recent phenomenon. When a researcher intentionally recruits a criterion-based selection of participants, it is a case of *purposeful sampling* (Creswell & Guetterman, 2021, pp. 240-241; Johnson & Christensen, 2020, p. 259; Patton, 2015). Johnson & Christensen (2020) also name this type of sampling strategy *purposive sampling* and describes the method as when “the researcher specifies the characteristics of the population of interest and locates individuals with those characteristics”(p. 254). Both terms describe the identical phenomenon, but I choose to employ “purposeful sampling”, which aligns with the concept delineated by Patton (2015). For this current study, I have employed this non-random sampling technique to ensure that each participant met the relevant criteria to ensure that each participant fits the purpose of the study. This technique is best suited to fit the purpose of the study as I want in-depth perspectives about teachers’ beliefs and experiences about LLMs. By intentionally selecting participants who can offer valuable information, researchers can thereby ensure that the collected data is relevant to the research inquiry. Moreover, purposeful sampling may promote the exploration of diverse viewpoints and experiences within the same topic, leading to a nuanced understanding of the subject under investigation.

Non-random sampling techniques, including purposeful sampling, does however exhibit innate limitations. As with any non-random method, the “ability to generalise from a sample to a population on the basis of a single research study is severely limited” (Johnson & Christensen, 2020, p. 254). By selecting participants according to specific characteristics

pertinent to the study, the extension of results to a broader population becomes challenging. In the context of my study into teachers' knowledge, beliefs, and experience with LLMs, it is crucial to note that the study did not prioritise quantifiable data. The narrow scope on this specific group of participants made obtaining relevant quantified data impractical. The study's focus necessitates specific criteria for potential participants due to the newness of LLMs in teaching. Furthermore, considering the ethical concerns associated with these models, the selection of available participants is inherently limited. At the time of the recruitment process, teachers were prohibited from actively incorporating LLMs into their teaching practices with their pupils. This prohibition issued from the Norwegian government was primarily due to privacy concerns. Despite these limitations, I choose to use non-random sampling techniques as the aim of the study was to investigate teachers' knowledge, beliefs, and experience with LLMs.

For the current study, I employed diverse strategies for participant recruitment. Initially, I reached out to all the schools in my region by sending participation request via email, including all the schools the University of Bergen has collaborations with. Subsequently, I posted an open invitation in a Facebook group dedicated to Norwegian teachers interested in artificial intelligence. However, a lack of willingness to participate emerged, primarily due to challenges in availability and a limited number of teachers fitting the study's criteria. Consequently, I decided to explore opportunities in various regions across Norway. In total, 88 upper secondary schools in Norway were contacted. Regrettably, only 20 schools replied whereas most were declining the invitation to participate. Nevertheless, I successfully enlisted two teachers this way. I therefore contacted teachers at different schools directly to recruit the remaining teachers and found this approach to be more practical. The more direct approach resulted in three more teachers. Ultimately, at the conclusion of the recruitment process, I successfully enlisted five teachers across different regions of Norway. The L2 English teachers are viewed in Table 1.

Table 1*Overview of Participating Teachers*

Teacher (T)	Gender	Age	Teaching duration
Teacher 1 (T1)	Male	40-45	7 years
Teacher 2 (T2)	Female	25-30	Less than 1 year
Teacher 3 (T3)	Female	40-45	23 years
Teacher 4 (T4)	Male	35-40	5 years
Teacher 5 (T5)	Male	45-50	20 years

3.5 Conducting the Interviews

The interviews spanned a two-month period, with the first four interviews taking place in December 2023 and the final one in January 2024. Four interviews are conducted in person at the teachers' school, while the last one is conducted digitally via Zoom. All audio files and transcriptions are securely stored on a remote server provided by SAFE and the University of Bergen.

Each interview follows a standardised structure, as outlined in the interview guide (see Appendix A). However, certain questions are omitted during the interviews as some of the participants spontaneously addressed multiple questions in their responses. This phenomenon may indicate that certain questions may have been superfluous. Nevertheless, considering this phenomenon is not the case for all the interviewees, including these seemingly superfluous questions in the interview guide ensures comprehensive coverage of all the topics.

Regarding the language used during the interviews, all participants are given the autonomy to select either English or Norwegian for their interview sessions. Some of the teachers indicated a preference for expressing their opinions more clearly in their L1 language. Consequently, two interviews are conducted in Norwegian, while three are conducted in English. Variation in response pace and extent of responses is observed among the interviewees, with some providing quick answers and others offering more comprehensive responses. The duration of the interview therefore ranges from 44 to 75 minutes. The brevity of the shortest interview, however, is attributed to time constraints rather than a shortage of response material.

As outlined in the interview guide (see Appendix A), the interview guide included excerpts of definitions and government-issued documents. These were presented to each interviewee along with subsequent follow-up questions. After the initial interview, it became apparent that a different approach concerning this part of the interviews was necessary. Given the length of some excerpts, the interviewee encountered challenges in comprehending the content of the excerpts and the definitions while simultaneously responding to follow-up questions. Consequently, each interviewee in the four remaining interviews was provided with written examples of the excerpts and definitions. I experienced that the written examples were valuable for the interviewees, as it enhanced their capacity to comprehend the excerpts and respond more effectively to the follow-up questions.

Pertaining the role of being an interactive listener was a crucial part of the interviews. As an interactive listener, I aspired to maintain an inviting tone while being attentive to ambiguous responses, clarifying uncertainties, or providing follow-up questions. My impression is that the interviewees appeared comfortable which made this process a positive experience. I aspired to establish a conversational environment, recognising the interview as a jointly constructed situation (Brinkmann & Kvale, 2015, p. 35). Additionally, by being an interactive listener, I maintained to keep the interviews on track by bringing the interviewees back if the teacher went off on a non-relevant topic (Johnson & Christensen, 2020, p. 194).

3.6 Data Analyses

3.6.1 Transcribing the Interviews

The in-depth interviews were converted from oral to written format using transcription. Transcription is “the process of transforming qualitative research data, such as audio recordings of interviews...into typed text” (Johnson & Christensen, 2020, p. 543). The two most common ways of transcribing interviews are verbatim oral or a more formal written style (Brinkmann & Kvale, 2015, p. 207). For the current study, I transcribed the interviews in a written format, preserving the meaning while excluding redundant fillers, disfluencies, and inconsequential interjections. Another rationale behind transcribing the interviews into a formal written style was to make member-checking easier (see section 3.8). Using a formal written style for the transcriptions should be considered from the start to avoid interviewees being shocked of how their oral language may appear incoherent (Brinkmann & Kvale, 2015, pp. 213-214).

To make the process of transcribing more effective, I used the dictation function in Microsoft Word. Each recording was uploaded to the dictation function, and the recordings were consequently transcribed from oral speech to written text. Microsoft Word's dictate function was able to separate when the researcher and interviewee spoke, thus include punctuation. For the sake of convenience, each utterance where the interviewer spoke was labelled "R", and "I" for the interviewees. The function produced a surprisingly good reproduction of the oral speech, including punctuation, but it became apparent that revision was necessary. For example, the dictate function never transcribed "ChatGPT" correctly. I therefore read through the transcribed text while listening to the audio recording and corrected any misspellings or faults where relevant.

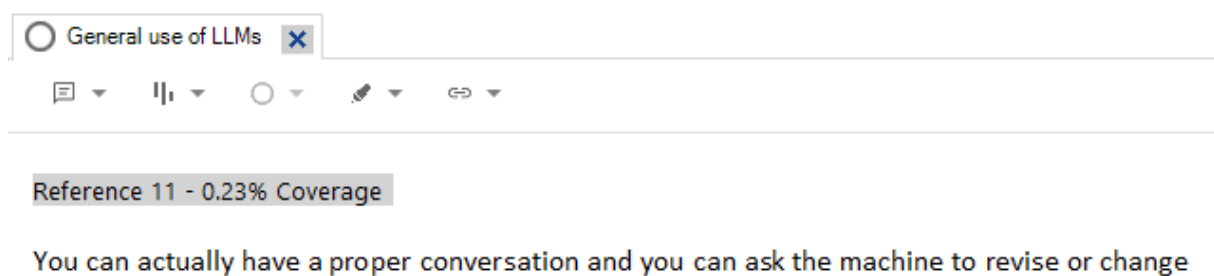
3.6.2 Analysis of Interview Data

The interview data are analysed using thematic analysis. As noted by Braun and Clarke (2006), thematic analysis is a "method for identifying, analysing and reporting patterns (themes) within data (p. 79). Themes capture important aspects of the collected data and seek to find a meaning or patterns within the data set (Braun & Clarke, 2006, p. 82). Before overarching themes could be identified, the material is coded and categorised. Coding refers to the process of providing an overview of the qualitative data material by linking segments of the text to keywords (Brinkmann & Kvale, 2015, p. 227). The software program NVivo 14 was used for the analysis and coding process in this study. Coding the data made it possible to identify patterns from the interviews and explore findings related to the research questions.

The initial categories for the coding were created according to the structure of the interview guide for the interview transcripts. For instance, I created the categories *digital competence*, *didactical considerations*, *motivations*, *challenges*, *general use of LLMs*, *understanding of LLMs*, *ethical concerns*, *digital Bildung*, and *training and guidance*. Developing codes in advance can be seen as *concept-driven* or *priori* coding (Brinkmann & Kvale, 2015, p. 227; Johnson & Christensen, 2020, p. 549). I found developing codes in advance useful in this study as it provided an overarching look of the data material for subsequential in-depth analysis. Table 2 represents the four phases of the qualitative analysis in this study.

Table 2*Overview of the Phases in the Qualitative Analysis of the Interviews*

Phase 1	Phase 2	Phase 3	Phase 4
Read through the transcriptions to identify any relevant segments related to the initial coding. The codes in this process are derive from the interview guide	Code the transcriptions in related to the research questions. Establish new sub-categories and new codes where relevant. Thematis, categorise, and organise coded statements.	Review and revise the codes and categories throughout the process.	Identify relations and themes concerning the research questions.

Figure 1*Example of Utterance Coded as Reference to General Use of LLMs.*

Next in line was attaching segments from the interview transcriptions to the pre-prepared codes. A segment is a meaningful unit of text and be anything from single words to larger passages (Johnson & Christensen, 2020, p. 544). An example of coding from the current study is observed in Figure 1. While reading through the transcriptions, I found that creating the code *teacher cognition* was necessary. Developing codes directly from the data can be seen as *inductive coding* or *data-driven coding* (Brinkmann & Kvale, 2015, p. 228;

Johnson & Christensen, 2020, p. 549). The inductive coding led to including teacher cognition to as a theoretical perspective in the thesis (see section 2.2).

3.7 Reliability

In academic studies and research, it is crucial to consider both reliability and validity. Reliability evolves around consistency and trustworthiness, and whether a finding from the study can be reproduced by other researchers (Brinkmann & Kvale, 2015, p. 281; Johnson & Christensen, 2020, p. 157). In essence, this revolves around whether interview subjects will provide the same answers to different interviewers. A study's reliability is important in several phases of the research process, e.g., the interviews, transcriptions, and analysis (Brinkmann & Kvale, 2015, p. 281). To uphold the reliability of the study, several actions were taken.

First, the interview underwent a pilot phase to ensure the reliability of using interviews as an instrument. A pilot may serve as an initial test of the research instrument, aimed at identifying any potential issues that may arise and could potentially compromise the integrity of the study (Van Teijlingen & Hundley, 2002). During the pilot, participants were asked to provide feedback, particularly if there were any ambiguous questions or questions that were challenging to understand. A notable issue arising during the pilot was the participants' lack of familiarity and experience with the subject matter in the interview guide. Consequently, assessing whether the questions in the interview guide effectively elicited the desired responses became a complex task. Some changes were, however, made concerning the wording of some of the questions in the interview guide. For example, the question "To what extent are you familiar with the technology and algorithms employed in large language models?" was changed to "Describe your familiarity with the technology and algorithms used in large language models" (Question 2.3, Appendix A). This adjustment aimed the question more clearly at examining their understanding of the algorithms used in LLMs, as opposed to a brief indication of their familiarity with the subject. It is noteworthy that despite the absence of prior knowledge about the topic, participants of the pilot interview were able to provide constructive feedback, such as the example above. For the remaining questions, the pilot interviewees demonstrated a clear and unambiguous understanding of the questions. However, a successful pilot is not a guarantee of success for the full-scale interview (Van Teijlingen & Hundley, 2002). This was something I had in mind when the interviews were conducted.

Second, the time of the day the interviews were conducted could affect the consistency of the results. Consequently, efforts were made to ensure that the interviews were conducted at approximately the same time. Yet, the teachers had demanding and divergent schedules. The interviews were therefore tailored to each teacher's availability. This led to a departure from a uniform interview timing, and the interviews were conducted both before noon and in the afternoon. Although the time variable may not be a crucial factor for the reliability of the study, it is necessary to be aware of the potential effects this may have.

Third, it is important to point out that an interview is a social interaction between the interviewer and the interviewee. Brinkmann and Kvale (2015) emphasise that the interviewee's statements are not merely collected; rather, they are collaboratively authored by the interviewer. My role as the interviewer in this study allowed me to influence which aspects to highlight based on the statements from my interviewees. Conversely, how the interviewees respond to the questions will affect how the interview develops. As pointed out by Brinkmann and Kvale (2015), "[t]he decisive issue is not whether to lead or not to lead but where the interview questions lead and whether they lead to new, trustworthy, and worthwhile knowledge" (p. 201). Although I had designed an interview guide with pre-prepared questions, the flexibility of the semi-structured interview allowed me to give more attention to specific topics through follow-up questions that emerged during the interviews. Furthermore, the transcriptions reveal that leading questions were used to confirm interpretations. For example, when the researcher asks, "So, for you it's all a matter of interest and motivation?" and the interviewee replies "Yeah, yeah". Drawing from this example, I find it crucial to emphasise the collaborative aspect in interviews. Including leading questions in qualitative interview can serve to assess both the reliability of the interviewees' answers and to verify the interpretations made by researchers (Brinkmann & Kvale, 2015, p. 200). Despite employing a semi-structured interview guide, I experienced that each interview was different to some extent. In certain cases, I posed leading questions for clarification. In other cases, I paid more attention to specific topics. Additionally, I employed member-checking to ensure that the interviewees agreed with the transcribed text of our conversation (see section 3.8).

Lastly, to uphold transparency in the research process, an appendix containing the interview guide is included (See Appendix A). To keep the anonymity of my participants, the transcriptions are not included as an appendix. Nevertheless, the provision of data collection and its procedures enhances the overall transparency of the study.

3.8 Validity

Whereas reliability concerns the consistency and trustworthiness of a finding, validity evolves around how accurate the interpretations of a finding are (Brinkmann & Kvale, 2015, p. 163; Creswell & Creswell, 2023, p. 213; Johnson & Christensen, 2020, p. 156). There are many different types of validity evidence one could collect in research studies, and “the best rule is to collect multiple sources of evidence” (Johnson & Christensen, 2020, p. 163). The concept of validity is a unified idea, and represents essentially “the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed purpose” (National Council on Measurement in Education & American Psychological Association, 2014, p. 14). One could thereby pose the question “are you measuring what you think you are measuring?” (Kerlinger, 1979, p. 138). In a broader perspective, validity also concerns to which degree the observations reflect the phenomena or variables in question (Cervone & Pervin, 2019, p. 48). Keeping these viewpoints in consideration, I have strived to be aware of potential threats that can affect the validity of the current study.

The materials for this study consist mainly of teachers’ reported beliefs and thoughts. Accurate interpretations are thereby essential as wrongly interpreted findings could weaken the validity of the study. To ensure that my interpretations coincided with the interviewees’ utterances, I interpreted their statements throughout the interviews and summarised the essence of their statement, e.g., “R: So your impression is that in addition to digital literacy, teachers and pupils should also acquire AI literacy?”. In the example of the interview of one of the teachers, I rephrased the answer and summarised the views to ensure that my interpretation of the utterances was correct. Rephrasing the interviewees’ answers can be one way ensuring validity as it can make it clear that the interviewer has understood the answer correctly (Brinkmann & Kvale, 2015). The crucial measure, however, to ensure that my interpretations were portrayed accurately, was to send the transcriptions back to the interviewees for member-checking.

Member-checking is one way of providing evidence for interpretive validity (Johnson & Christensen, 2020, p. 285). Interpretive validity concerns to which degree the researcher accurately understands the participants’ thoughts (Johnson & Christensen, 2020, p. 285). Creswell and Creswell (2023) clarify that member-checking entails providing interviewees with semi-refined findings, such as major findings and themes, rather than sending raw transcripts for verification (p. 213). For my study, however, I asked each participant to review the whole transcription to ensure that their opinions and intentions were reproduced correctly.

In this sense, sending raw transcripts instead of the semi-refined findings may limit my study's validity. Nevertheless, I found this member-checking necessary as teachers' beliefs and reports form the data material collected from the interviews. Furthermore, it is vital that the teachers validated the transcriptions to ensure they accurately reflected their viewpoints and interview experiences. For my study, no remarks were provided by either teacher regarding the transcriptions. It is important to highlight, however, that there was no formal verification confirming that the transcriptions were representative of their viewpoints. Despite the absence of explicit verification from the teachers, it is reasonable to infer that the transcriptions did indeed capture their perspectives. The absence of validation, nonetheless, poses as a potential risk to the overall validity of the study.

The role of the researcher is important to address when interpreting and analysing the data material. If a researcher carefully selects and interprets findings that coincide with what the researcher wants to find, we might come across a phenomenon called *researcher bias* (Johnson & Christensen, 2020, p. 284). This can be a threat to the study's validity, but it is possible to reduce the effects of researcher bias. A strategy for this is *reflexivity*, a process in which the researcher is engaging in critical self-reflection about potential biases (Johnson & Christensen, 2020, p. 284). For the current study, I am generally positive about artificial intelligence in education. It is a very exciting and relevant topic to investigate, but this positive stance may affect my interpretations of the findings. A complete objectivity is impossible, and complete subjectivity is undermining the credibility of the results (Johnson & Christensen, 2020, p. 47). Therefore, I focussed on balancing these aspects to understand and interpret the findings as objective as possible. This was done by always controlling, questioning, and theorising the results. Another way to reduce researcher bias is transparency with the collected data material and analysis procedures (see section 3.6). All materials used in this study, apart from the interview transcriptions, are therefore available as appendices.

In the context of this study, it is relevant to address external validity. External validity refers to when you want to generalise from a set of research findings to other people (Johnson & Christensen, 2020, p. 288). External validity is a weakness in qualitative research, and as mentioned, the participants for this study were not randomly selected (see section 3.4). The intention of this study is not to generalise findings, but to investigate whether any insights regarding LLMs emerge from L2 English teachers in upper secondary schools. These insights pertain to how teacher understand LLMs, in what ways they utilise LLMs, and how didactical considerations guide their integration of LLMs. Consequently, teachers from different

counties in Norway were recruited. If the teachers were recruited from one geographical area, it could be seen as a case study (see e.g., Johnson & Christensen, 2020, p. 49). Additionally, the gender and age ratios among the participants vary (see Table 1). The geographical variable, gender and age ratios will, however, not be discussed in relation to the results. Nevertheless, a geographically diverse sample together with a varied age and gender variable may provide with a wider range of teachers' beliefs and viewpoints.

3.9 Ethical Considerations

Ethical considerations (or issues) play a pivotal role in educational research, both in the development and implementation of any research study (Johnson & Christensen, 2020, p. 119). Consequently, ethical considerations are an ongoing and integral component of the entire research process (Brinkmann & Kvale, 2015, p. 97). Considerations regarding professionalism and treatment of the participants are two aspects of research ethics that have been focused on for the current study (Johnson & Christensen, 2020, p. 120). In short, professionalism and treatment of the participants for the current study were considered via obtaining informed consent, participants' anonymity, and member-checking. Further elaboration of how professionalism and treatment of the participants were addressed is provided in the subsequent paragraphs. The ethical approach of *utilitarianism* dictates that judgements regarding the ethics of a study depend on the individual consequences for each participant, and the broader advantages that potentially might arise from the study results (Johnson & Christensen, 2020, pp. 119-120).

Furthermore, in qualitative studies, there is a persistent dilemma between the pursuit of in-depth knowledge and the careful consideration of ethical concerns. On one hand, researchers aim for thorough and probing interviews, which can potentially intrude upon the interviewee's privacy and comfort. Conversely, researchers endeavour to maintain a high level of respectability, but this cautious approach may limit the depth of the empirical material collected (Brinkmann & Kvale, 2015, p. 96). For the current study, I took several measures to uphold the research ethics.

First, the study was approved and registered in RETTE (Risiko og ETTErlevelse i forskningsprosjekter).

Second, obtaining an informed consent from each interviewee was a prerequisite for proceeding with the data collection. Informed consent should include information of any

factors that may affect their willingness to participate (Johnson & Christensen, 2020, p. 126). Moreover, the consent insures voluntarily participation and the right to withdraw from the study at any stage in the process, as well as detailed information regarding the purpose of the study and data access both during and after the study (Brinkmann & Kvale, 2015, p. 104). I included this information in an informed consent in all the participation requests. The consent form was created by using the template provided by the Norwegian Agency for Shared Services in Education and Research (Sikt). Included in the consent form were information regarding the study's objectives, content, and the participants role in the study. Additionally, the participants were also offered the option to withdraw from the study at any point to ensure their voluntary participation (see Appendix C). Ultimately, the interviewees signed a consent form where they confirmed that they had read the informed consent document (see Appendix C) and accepted to participate in the interviews (see Appendix D).

Third, protecting the participants' anonymity was important at every stage of the data collection and analysis. Confidentiality mandates that participants remain anonymous, securing their identities and statements from recognition by others (Brinkmann & Kvale, 2015, p. 106). Securing the identity and respecting the privacy of the participants is a crucial part of research ethics (Johnson & Christensen, 2020, p. 133) To uphold anonymity, each participating teacher in the study was assigned an anonymous label. These labels were assigned based on the order in which the teachers were recruited, thereby ensuring consistency and clarity in referencing each teacher. For example, Teacher 1 was assigned the label "T1", Teacher 2 as "T2", and so forth. This practice ensured the protection of teachers' anonymity while still allowing for clear identification and reference within the research findings, aligning with the ethical considerations provided by Brinkmann and Kvale (2015) and Johnson and Christensen (2020). Additionally, any references to location and upper secondary schools or universities, were removed from the transcriptions. Furthermore, all the data material were saved in password protected storage solution SAFE and managed using licenced software from the University of Bergen (e.g., NVivo).

Fourth, the transcribed interviews were member-checked (see section 3.8). In qualitative interviews, the participants may express information that they later wish they had not disclosed (Brinkmann & Kvale, 2015, p. 96). Therefore, the transcriptions were sent back to the interviewees to verify that the transcriptions were accurate reflections of their viewpoints. Giving the interviewees the opportunity to validate the accuracy of the transcriptions can act as a way of reducing *research misconduct*. As noted by Johnson and

Christensen (2020), research misconduct is a segment of professionalism and concerns fabrication, falsification or plagiarism of the research results (p. 121). To avoid research misconduct of this study, all transcriptions were sent back to the participants for validation. Any findings or results obtained from the data analysis procedures are thereby transparent and could be verified by the participants.

4 Findings and Discussion

4.1 Chapter Outline

This thesis has explored the utilisation of LLMs by teachers in the context of their L2 English teaching in upper secondary schools in Norway. The first aim of the study was to examine L2 English teachers' knowledge of LLMs, as well as their beliefs and reported implications of using LLMs in their teaching, e.g. ethical, didactical, and pedagogical implications. The second aim of the study was to elucidate specific classroom activities in which the teachers integrated LLMs in. The third aim of the study was to examine their motivations for integrating LLMs, as well as whether didactical considerations guided their integration of LLMs. Each research question will be addressed in turn, through a presentation of findings followed by a discussion of teachers' knowledge, classroom activities in which they utilised LLMs, and the didactical considerations guiding their integration of LLMs. During this process, the findings will be elucidated in relation with one another as well as their alignment with the theoretical concepts and research as outlined in Chapter 2. The initial subchapter will address the first research question, followed sequentially by the exploration of the second and third research question. Finally, relevant limitations of the study will be deliberated upon as well as suggestions for further research.

4.2 Teachers' Knowledge of LLMs

The qualitative in-depth interviews of five L2 English teachers in upper secondary schools revealed that four of the five teachers expressed similar understandings of the technology and algorithms within LLMs. The overview from Table 3 represents the teachers' understanding of LLMs.

Table 3*Overview of Teachers' Knowledge of LLMs*

Teacher	Expressed knowledge of LLMs
T1	I have a pretty good understanding for predicative models and how they are built up by neural networks. It reproduces data based on probability. (my translation).
T2	The first I think of is different ways to learn English, meaning the use of digital technology. (my translation)
T3	It just gives you what it thinks you want...goes through so many written documents and figures out what the next possible solution is. It's based on everything that have been published.
T4	It goes through statistics. What is the likelihood of that word being there? It has access to a lot of text.
T5	It's just basically immense volumes of text that a computer as analysed and basically is looking for patterns. It is a kind of pattern recognition technology.

T1's knowledge about LLMs was different from the other teachers. In terms of terminology, T1 expressed far more detailed descriptions of his knowledge of LLMs. He stated he had a good understanding of predicative models and how they are based on neural networks. Describing LLMs as built up by neural networks corresponds with theoretical perspectives that LLMs are constructed from neural networks characterised by an immense volume of parameters (Ray, 2023; Strümke, 2023). T3, T4 and T5 all shared similar assumptions of how LLMs are built. T3 and T4 described LLMs as figuring out what the next word in the line would be. Moreover, they stated that LLMs draw upon a diverse corpus of texts and leverage statistical analysis to determine the most probable succeeding word. T5 expressed similar views on how LLMs are built and described them as immense volumes of computer analysed text. What distinguished T5 from the other teachers was his acknowledgement that LLMs depend on "pattern recognition technology". Describing LLMs as a pattern recognition technology seems appropriate, given their capacity to absorb patterns that are woven into the structure of language (Mirchandani et al., 2023).

In contrast to the teachers' expressions of their knowledge of LLMs above, T2 did initially describe LLMs as different models (different ways) to learn English. From the theoretical perspectives on LLMs, this is a wrong understanding of LLMs. However, as the interview

progressed, T2 expressed a far more nuanced understanding of the implications of using LLMs in L2 English teaching (see section 4.3). T2 did not have general knowledge of LLMs or how are built up. On the other hand, T2 understood different ways of utilising them in classroom activities (see section 4.4).

These findings show that there is a general consensus among the majority of the teachers regarding their knowledge of LLMs, suggesting a nuanced understanding of technology competence among the interviewed teachers. The findings support Krumsvik (2023) dimensions of digital competence concerning digital skills and knowledge. Moreover, the findings correspond with the proposed aspects provided by Ottestad et al. (2014). Generic digital competence specifies the general knowledge and skills that teachers should obtain in order to function as digital teachers (Ottestad et al., 2014, p. 248).

4.3 Implications of Using LLMs

4.3.1 LLMs for Student Support

In the analysis of the participants' beliefs concerning the functionality and implications of LLMs, some interesting insights emerged. First, T1 highlighted the role of LLMs, thus AI tools in general, as a supplement to teacher feedback for his pupils. Rather using the language model as a substitute tutor, the teachers expressed a positive stance on integrating the language model to extend the teachers' pedagogical repertoire. By tailoring the LLMs to the individual pupils' academic proficiency level, it might be possible to utilise the potential functionality of LLMs in a constructive manner. The main reason for utilising the LLMs in this manner was to counter the lack of resource and time each teacher had to follow up each individual student.

To use a language robot to have a continuous dialogue that can provide my students with active feedback one-to-one. This is something that I neither have resources or the opportunity to do in a classroom (T1, my translation)

This perspective indicates a reflection of T1's belief in the importance of maintaining the teacher's central role in the educational process while utilising AI as a supportive tool. Providing feedback on aspects such as grammar, academic writing, punctuation, and voicing may be some of the areas a LLM could provide feedback effectively. This way of leveraging the capabilities of LLMs within classroom contexts is noted by (Kasneci et al., 2023). Moreover, by emphasising the role of AI as a supplement, T1 expressed a willingness to integrate AI into education in a way that he thought enhanced academic proficiency without

diminishing the role of human teachers. T1's beliefs on how LLMs could function as a conversation partner aligns well with the findings from Kuhail et al. (2023). In their study, Kuhail et al. (2023) found three main types of chatbots that relate to the teaching profession: teacher agents, peer agents, and motivational agents (Kuhail et al., 2023, pp. 990-991). The teacher agents offered guidance on content resources and engage in discussions on specific topics with the pupil (Kuhail et al., 2023, p. 990). Peer agents provided support and enabled students to request targeted assistance as needed (p. 991). Motivational agents were designed to respond to pupils' answers with a range of emotions aiming to motivate the pupils (p. 991). Additionally, a work by Tack and Piech (2022) investigated the capability of conversational agents to reply to a student in an educational dialogue. They found that the LLMs used were capable of replying to a student and generated dialogues that conveyed the impression that the models understood the learner (Kasneci et al., 2023). However, the models are well behind human performance when it comes to pedagogically helping the student (Tack & Piech, 2022). In their research, Tack and Piech (2022) investigated the pedagogical ability of both human and AI teachers. They found that the students revealed a preference for human-generated responses for their tests. It is notable, therefore, that the research Tack & Piech (2022) proposed that human teachers' responses to student tests were preferable as opposed to the AI generated response. From an education perspective, there are still many gaps concerning knowledge and uncertainties when it comes to appropriate integration of LLMs into teaching processes (Kasneci et al., 2023).

4.3.2 On Problems Integrating LLMs in Classroom Activities

Effective utilisation of AI technologies, such as LLMs, is dependent on teachers' pedagogical knowledge (Celik, 2023). Perhaps more importantly, the teachers revealed that how they teach this knowledge is crucial for their pupils' constructive integration of LLMs. As emphasised by T1, "pupils must understand what incorrect use looks like, and the teachers are responsible for teaching them that. That should be the main principle" (my translation). This report was similarly noted by T2 who stated that "the teacher must teach the pupils how to use LLMs and teach them how to reflect on how they are using LLMs" (my translation). To be able to utilise LLMs effectively and teach their pupils what constitutes effective utilisation of LLMs, teachers should acquire the necessary competence. Moreover, integrating LLMs into effective teaching practice requires understanding their capabilities and limitations (Kasneci et al., 2023). However, it is crucial to acknowledge that many teachers may lack the required competence to effectively integrate new technologies into their teaching practices

(Redecker, 2017). If teachers do not acquire the necessary competence to understanding LLMs capabilities and limitations, it stands to reason that their pupils will also struggle to do so.

Furthermore, a possible consequence of incorrect use of LLMs may be the need for changes in how teachers assess their pupils. T3 noted that the emergence of LLMs “changed the way I’m testing my students”, a viewpoint echoed by T5, who remarked that “we’re really in the process of changing a lot of things about how we teach and how we evaluate”. This observation seems to align with the broader discourse concerning LLMs and assessment, which have sparked heated debates since the release of ChatGPT in November 2022 (Bozkurt, 2023; Krumsvik, 2023).

4.3.3 Ethical Issues

The teacher interviews revealed numerous ethical considerations with using LLMs. First, teachers reported ethical considerations concerning the inherent biases of LLMs. T1 stated that “it [LLMs] can amplify bad views on topic such as religion, gun industry, and trust in government”. T2 explained that “it [LLM] is biased when it writes responses about American politics”. T5 stated that “the language model is biased by what it’s fed”, thereby implying that the data material the LLMs is trained inadequately. Meaning, the data material does not reflect a nuanced depiction of the society. T4 described the inherent bias in LLMs differently than the other teachers:

It doesn’t present all the sides in an argument. If it's controversial, let's take racism as an example. It is very biased to antiracism, which is of course fine, but you need to understand the racist arguments. (T4)

In his view, the LLM’s bias was not about amplifying outdated prejudices, but to avoid writing about problematic topics. This can be seen as a lack of *historical empathy*. Historical empathy can be described as the capacity to understand and relate to the thoughts, feelings, and experiences of people who lived in the past, considering their context and perspective rather than imposing present-day values and judgements (Roos, 2022). From research studies on LLMs, LLMs can perpetuate and amplify existing biases and unfairness in society, which can negatively impact teaching and learning processes and outcomes (Kasneci et al., 2023). If handled sensibly by the teacher, the challenges of biases can be insightful in learning and education scenarios to acquaint pupils with this risk of LLM application (Kasneci et al., 2023). Furthermore, by interacting with LLMs, individuals can develop a nuanced

understanding of the ethical considerations revolving around artificial intelligence (Krumsvik, 2023).

Second, the T1 and T5 revealed beliefs concerning copyright issues and data privacy and security. On addressing the privacy issues, T1 stated: “We cannot ask the students to log into ChatGPT and hand over their personal information, and teachers cannot upload students’ texts to let ChatGPT help us assess them” (my translation). Similarly, T5 noted that “It’s illegal for us to put students’ work into the bot”. T1 and T5’s viewpoints seem underscore their beliefs on the importance of privacy when integrating LLMs into their teaching. The use of LLMs in education have raised concerns about data privacy and security as student data is often sensitive and personal (Kasneci et al., 2023). These concerns have also been noted in Norwegian contexts (NOU 2022: 11). Questions concerning privacy issues are thereby crucial to address in the integration of LLMs (Krumsvik, 2023). Regarding copyright issues, it has been suggested that ChatGPT generates novel summaries with such precision that it seems implausible it has not been trained on the original content (Vallance, 2023; Aas, 2023). Acknowledging concerns on copyright issues and especially data privacy is crucial when comparing it to the professional digital framework for teachers (PfdK). The PfdK states that “a professional, digitally competent teacher...has an insight into the legislation and ethical concerns, as well as development of pupil’s digital Bildung...in a digital and democratic society” (Kelentrić et al., 2017, p. 6). Furthermore, it is imperative for teachers to actively address ethical concern surrounding privacy issues (Krumsvik, 2023). This implies that the teachers must navigate the digital ethics and focus on digital citizenship, privacy, and intellectual property (Norhagen et al., 2024). T1 and T5 did report that digital ethics were something they were addressing when integrating LLMs, thereby encompassing the content in the PfdK. In the context of ethical issues, the findings derived from the interviews revealed that the most prominent concerns revolved around LLMs’ inherent biases and privacy issues.

Examining teachers’ beliefs and interpretations of the functionality and implications of LLMs offered a nuanced perspective surrounding the integration of these technological tools in their teaching practice. The focus now transitions from exploring teachers’ understanding of LLMs, including their beliefs, and reported implications, to examining the various classroom activities in which the teachers integrate LLMs in.

4.4 Classroom Activities Integrating LLMs

The second aim of the study was to examine which different classroom activities L2 English teachers integrated LLMs into. The main findings regarding the second research question will be presented and discussed in relation to relevant theoretical perspectives. The overview from Table 4 represents the different classroom activities in which LLMs were integrated.

Table 4

Overview of Reported Classroom Activities Integrating LLMs

Teacher	Classroom activities
T1, T2, T4	Writing <ul style="list-style-type: none"> - Writing inspiration, providing examples of essay structures - Students revising own texts
T1, T5	Adapting learning materials <ul style="list-style-type: none"> - Summarising texts - Rewriting of texts adapted to students' academic level
T2, T3, T4	Teacher-pupil interactions <ul style="list-style-type: none"> - Having the pupils explore good and bad prompts - Teacher-pupil interactions discussing pros and cons of using LLMs
T1, T4, T5	LLMs as time-saving mechanisms <ul style="list-style-type: none"> - Creating assessment criteria - Task creation - Lesson planning and forming of these

4.4.1 LLMs as Tools for Writing

Using LLMs as tools for writing was explicitly reported by two of the five teachers. What was common in these findings was how the teachers utilised LLMs as a tool for their pupils' writing processes. T2 reported that she integrated LLMs as a source of writing inspiration. In the beginning of a writing process, she encouraged her pupils to utilise LLMs as a starting point for their writing:

I tell my students that those of you who struggle with creativity can use ChatGPT to get started. I've given my students prompts I've tested out beforehand, so I know that

they work. And then they can subconsciously understand how to formulate questions to the chatbot. (T2, my translation)

When pupils use generative artificial intelligence as a writing tool, the tool can be used in an *iterative* process (Kohnke et al., 2023). In this context, iterative refers to repeating a process numerous times with continuing feedback to enhance the final product (Elstad, 2023). T2 stated that “they [pupils] tend to ask wrong or inaccurate questions in the start, and this type of activity will make them more aware of the type of questions that provide the best answers” (my translation). In addition, T4 also used LLMs for creating a framework for what a written assignment could look like.

I gave the students an assignment with a month submission deadline...we worked together in class to get a framework for what the assignment could look like. The framework they got was average, so we obviously had to change it, but it was a nice starting point. (T4)

Similarly to T2’s utilisation, T4 used LLMs with his pupils at the beginning of a written assignment. These findings support previous research that has suggested that using LLMs as a inspirator at the beginning of a writing process could be a useful way of utilising AI tools (Elstad, 2023). Moreover, the output text produced by the chatbot could subsequently be revised by the pupils, thereby enabling them to put their personal voice to the text, e.g., rhetorical devices, arguments, wording, and the choice of which factual information the pupil find useful to include in the text (Elstad, 2023, p. 23). Using LLMs as an inspiration and revision tool seems in this context to serve as a productive utilisation of artificial intelligence. The LLMs can aid the pupils to identify grammatical errors, spelling, and syntax (Rathje et al., 2023). Additionally, they can provide suggestions to improve clarity, precision, and cohesion. All these elements may help the students improve the quality of their drafts and ultimately enhance the quality of the final product (Elstad, 2023). Furthermore, experimental studies shows that using LLMs can support productivity and enhance the quality of the final written product (Noy & Zhang, 2023). If used correctly, Elstad (2023) suggests that LLMs have a *functional enhancement* potential, meaning an enhancement that makes the writing process more effective. One correct way of utilising LLMs seems to rely on assessing the quality of the generated text. This was reported by T4 who emphasised the importance of always revising the output received from LLMs and stated that “we obviously had to change it [the output text], but it was a nice starting point”. Reporting that utilising LLMs in the beginning of the writing process to get his pupils started, may indicate an awareness of how

LLMs could make the process more effective. T4's emphasis on revising generated output aligns with the importance for teachers to use LLMs as a supplement to their classroom instruction, rather than a replacement (Kasneci et al., 2023). As noted by Lund et al. (2014), "PDC involves teachers not only appropriating technologies, but also making their learners appropriate them and put them to productive use" (Lund et al., 2014, pp. 283-284).

However, there are ongoing debates revolving around the accuracy of LLMs' responses (Kohnke et al., 2023). LLMs may give you a very impressive-sounding answer that turns out to be false (Bowman, 2022). Moreover, while LLMs have shown remarkable abilities in generating human-like text, it may occasionally produce incorrect or misleading information (Ray, 2023). Inaccurate responses, or worse factual wrong responses, are particularly an issue for learners who are incapable of noticing the quality of the LLMs' responses (Kohnke et al., 2023). Production of incorrect information was reported among the teachers. T1 reported that "it [LLMs] comes up with weird things that is absolutely wrong but writes very convincingly". Taking production of incorrect information into account, other consequences arise. For example, there is a risk of overreliance on AI if LLMs are used uncritically. As AI tools such as LLMs become extensively more advanced, an overreliance on the model may lead to a reduction in critical thinking and independent problem-solving skills (Ray, 2023). This was evidently reported by T1: "Even when they were told not to do this [use ChatGPT to finish the submission], they still took the shortcut...they hadn't read nor revised their text...consequently, they were not able to see that their text was bad" (my translation). The utterance from T1 described the learner's perspective. Nevertheless, one could draw lines from his experience towards a teacher's perspective. A teacher's digital didactics and didactical choices of when and how to integrate technologies in the classroom are the crucial factors of successful integration (Kongsgården & Krumsvik, 2019; Krumsvik, 2012). Furthermore, successful integration seems to also rely on changing the focus from product to process (Kongsgården & Krumsvik, 2019). When the pupils used LLMs to write their assignment for them, it seems that T1 recognised that his choice of integrating LLMs in this session was a poor didactical choice. This finding support previous research by Kongsgården and Krumsvik (2019) and Krumsvik (2012), suggesting the need for understanding when and how LLMs could be integrated constructively. Moreover, this finding may be interpreted to indicate that T1 was aware of how digital technologies can support the pedagogical and didactical aspect of teaching. These aspects rely on what to use, why, how, when, and where, as suggested by Lund and Hauge (2011) and Røkenes and Krumsvik (2016). A strong

overreliance, or dependence, on LLMs has inherent risks that pertains both to students and teachers (Kasneci et al., 2023). From a teacher's perspective, the effortlessly generated information could negatively impact their critical reasoning, problem-solving skills, and creative thinking (Kasneci et al., 2023). Assessing the decisions by AI technologies, such as LLMs, requires different types of pedagogical knowledge (Holmes et al., 2022). Consequently, it seems important to continuously monitor the quality of the output from the LLMs, and always refine accordingly.

T3 and T5 did not report any specific classroom activities in which they integrated LLMs for writing purposes. T3 expressed reluctance towards integrating LLMs in writing exercises, citing concerns such as, "If you are a language teacher and the point is to teach them how to write themselves, it can be quite troublesome if they just type it into a computer and get an answer". Similarly, T5 mentioned feeling uncomfortable and stated, "I haven't really explored that [using ChatGPT for revising texts]". From these statements, there may be reason to claim that T3 and T5 decided not to integrate LLMs for writing purposes due to perceived limitations of LLMs' pedagogical and didactical abilities. It seems that T3 and T5 recognised utilising LLMs for writing purposes potentially could undermine the learning outcomes of their writing skills. This can furthermore indicate an awareness of knowing when not to use LLMs. On the other hand, the decision not to integrate LLMs for writing purposes nevertheless raises questions about their PDC concerning LLMs. Effective integration of AI technologies is dependent on teachers' pedagogical knowledge (Celik, 2023). In fact, the role of technological and pedagogical knowledge is important in any successful integration of any technology (Mishra & Koehler, 2006). Their reluctance to integrate LLMs may therefore also indicate a lack of sufficient pedagogical and didactical knowledge to utilise AI technologies for writing purposes. Hence, their lack of PDC to exploit the opportunities of AI in education. Because AI has transformed the pedagogical knowledge teachers must have (Seufert et al., 2021), there may be reason to claim that T3's and T5's PDC currently lack the sufficient pedagogical and didactical competence to effectively integrate AI technologies in their L2 English teaching.

4.4.2 Adapting Learning Materials

Utilising LLMs to adapt learning resources to pupils' academic level was evidently reported by two of the teachers. T1 stated that:

I've shown my students how to adapt the texts they're reading, rewriting them to make them easier to understand...I had an academically difficult text and used AI to

translate it into English. The translated English would then be adapted to different levels. For example, I prompted the chatbot to adapt the text to a 16-year-old student with a B2 English level. Consequently, the chatbot adapted according to the preferences. (T1)

The situation described by T1 seems to be a good example of how LLMs could be incorporated effectively in L2 English teaching. A reported use of LLMs is that it can adjust the complexity of a dialogue or text to make it more suitable for beginners or advanced learners (Kohnke et al., 2023). Moreover, creating customised learning resources tailored to address pupils' specific areas of challenges are one way of leveraging the capabilities of LLMs (Kasneci et al., 2023). Utilising LLMs to adapt learning resources is similarly noted by Krumsvik (2023). Adapting learning resources to academic level was also reported by T4:

It's a great way to summarise or simplify difficult texts. Many of the prompts that I've given are like "this is for high school students", or "this is too difficult, can you make it easier", and vice versa "this is too easy, can you make it more difficult". If you know how you can talk to the chatbot, I think it's easier to adapt whatever the result is, and you get something that you can actually use. (T4)

LLMs can be used to adapt learning materials to adjust the materials based on pupils' progress and performance (Baidoo-Anu & Ansah, 2023; Kasneci et al., 2023). In instances where a pupil encounters difficulty comprehending the learning material, teachers can use LLMs to adjust the proficiency level, provide personal learning materials, and suggest customised learning paths (Kuhail et al., 2023). The PfdK also notes the importance of adapting content to pupils' individual needs as a part of being a professional digitally competent teacher (Kelentrić et al., 2017). However, the quality of T4's prompts is subject to investigation. Leveraging the capabilities of LLMs often involves creating effective prompts (Elstad, 2023). This process, often called *prompt engineering*, is a competence needed to interact effectively with LLMs, as prompts essentially serve as instructions given to LLMs to ensure desired qualities in generated output (White et al., 2023). However, prompts could be susceptible to several pitfalls including ambiguity, bias reinforcement, overfitting, lack of context, ethical considerations, unintended side effects, and unrealistic dependency on limitations (Giray, 2023). To mitigate these challenges associated with inaccurate generated prompts, critical evaluation of the generated content, striving for clarity, inclusivity, and alignment with ethical standards are preferred (Giray, 2023, p. 2633). Consequently, addressing T4's examples of the prompts "easy" and "difficult" becomes imperative, as they appear to fall short of the

standards of accuracy proposed by Giray (2023). This deficiency may have didactical implications for T4, as LLMs lack the capacity to discern what is easy or difficult, and for whom. Therefore, elucidating T4's lack of accurate prompt engineering is crucial, as proper prompt engineering and high-quality prompts are fundamental for ensuring the quality of generated output and fostering effective interaction with LLMs. On the other hand, it is plausible to consider that T4's use of "easy" and "difficult" as prompts may have been solely for visualisation, rather than accurate representations of the prompts used in interactions with LLMs.

In light of T1's statements of prompting the chatbot to "adapt a text to a 16-year-old student with a B2 English level", concerns arise regarding the reliability and accuracy of LLMs in general. First, it seems disconcerting that T1 places trust in LLMs to accurately adapt a text for a B2 English reading level. While studies suggest that LLMs possess the capability to tailor learning materials to match pupils' academic level (Kasneci et al., 2023), contradictory evidence indicates that LLMs may generate human-like but inaccurate information (Ray, 2023). LLMs' responses are derived from pattern learned from training data, but this does not guarantee their ability to encompass the wide range of factors influencing an individual's B2 English proficiency. Prompting LLMs to act as someone or adapt learning material for a specific purpose, can be seen as what White et al. (2023) call *the persona pattern*. The persona pattern is essentially a prompt engineering process in which the LLMs' output take a certain perspective (White et al., 2023). In T1's example, a 16-year-old student with a B2 level of English was the adopted perspective. This approach, however, raises the question regarding ethical considerations. Similar to the example from T4, a LLM may lack comprehensive material of what constitutes a B2 English level, nor does it necessarily reflect the wide characteristics and abilities of all 16-year-old pupils eligible for this academic level. In conclusion, while LLMs offer potential benefits of adapting learning material, careful considerations and assessment of the quality of the adapted learning material seems to be crucial for responsible and effective use. Additionally, this finding suggests that assessing the quality of the adapted learning material seems equally important as being able to comprehend the content itself.

4.4.3 Teacher-Pupil Interactions Discussing LLMs

T3 reported that she facilitated a groups discussion to let the pupils differentiate between AI and human generated content.

I placed them in groups to do a written assignment, write it as an article, and include sources... Then they were told to do the exact same task, but now they were only going to use ChatGPT and copy/paste... Afterwards, they handed in both their own and the version generated by ChatGPT. I printed them, shuffled the copies, and handed them back to the students. Their task now was to distinguish between the AI generated texts and those that were written by themselves. (T3)

The pupils were also tasked to assess the quality of the AI generated text. According to T3, the pupils noted that the referencing in the texts generated by ChatGPT was “off”, and they also commented on repeated paragraph structures. Her pupils’ ability to distinguish between AI and human generated content contradicts the finding in a study conducted by Merine and Purkayastha (2022). In their study, they reported that students were not able to differentiate between AI and human generated texts. The reported example from T3 contains, however, inherent limitations. T3’s pupils were in upper secondary schools in Norway. The participants in Merine and Purkayastha were higher education students. Consequently, the academic level of the AI and human generated texts was different.

Another interesting finding from T3’s statements are her pupils’ assessment of the texts generated by ChatGPT. Her pupils noted a dissatisfaction with the content generated by the LLM. T3 stated the repeating paragraph structure in the text created by ChatGPT as having “the same structure like a model would do”. Additionally, she reported that “[the] first thing my pupils noticed was the fact that ChatGPT had given the same title to two of the fake ones”. Assessing student satisfaction with AI-generated responses has been studied by (Nguyen et al., 2021). In their study, more than 98% were satisfied with the AI generated response from a LLM (p. 6). T3 stated her pupils’ assessment of ChatGPT made them realise some of the limitations to ChatGPT. The group activity she facilitated, seem to have prompted a deeper reflection on the capabilities and limitations of LLMs in language learning. Moreover, Skulstad and Touileb (2024) proposed this type of activity as being an unproblematic way of using LLMs (p. 146). The exercise may have encouraged her pupils to experience a more nuanced understanding of LLMs’ role as a supplement to, rather than a replacement for, human created content.

Whereas T3 facilitated a writing exercise to highlight potentials and limitations of LLMs, T2 organised a classroom discussion where the students conversated about differentiating between real content and AI created content:

I divided them into groups and handed out a bunch of AI generated photos. I asked them to place the photos they thought were AI generated into one folder and the real photos in another. Some groups understood my intention, and other groups divided the photos about 50/50. Subsequently, I showed them an article about AI generated photos. The article contained the AI generated photos I handed out to them, but I did not show them the photos in the article. Afterwards, they were given they choice to rearrange the photos in the folders again, and all the groups moved every photo to the AI generated folder. (T2)

The group discussion encouraged T2's pupils' to critically examine the benefits and ethical considerations of using LLMs in language learning. T2 reported that "this activity introduced my pupils to AI and they received insight in how difficult it is to distinguish between real and fake". Studies have been made on making LLMs innovations available for student evaluation (Merine & Purkayastha, 2022; Nguyen et al., 2021; Song et al., 2023). The student evaluations involved students differentiating AI-generated from human-generated content (Merine & Purkayastha, 2022; Song et al., 2023). Merine & Purkayastha's (2022) findings support the reported viewpoints from T2. In their study, Merine and Purkayastha (2022) found that expert-level graduate students could not distinguish clearly between AI-generated and student-generated summaries of fairly complex scientific articles (p. 572). The authors tasked students to differentiate between AI and human generated texts. T2's pupils differentiated between AI generated and real photos. Consequently, there are evident obvious methodological and academical differences between this study's findings and Merine & Purkayastha (2022). Merine & Purkayastha's (2022) students were unable to differentiate between AI and human-generated text. Similarly, many of T2's pupils were unable to distinguish between real and AI generated photos. Nevertheless, the underlying principle remain consistent: the challenge of distinguishing between AI and human-generated content.

The dialogue also offered her pupils the opportunity to reflect on their use of LLMs, making informed decisions and promoting digital citizenship. Several researchers argue the importance of self-regulation where pupils learn to take initiative in their own learning processes thus in collaboration with each other (Brookhart, 2007; Nicol & Macfarlane-Dick, 2006; Pintrich, 2000; Zimmerman, 2008). By facilitating classroom discussions and group work addressing potentials and challenges with using LLMs, T2 and T3 thereby allowed for a learner centred lesson where their pupils could interact with each other. These findings may indicate their beliefs on prioritising peer interaction as a constructive learning activity when

integrating LLMs. By facilitating classroom discussions and group work addressing the potentials and challenges associated with using LLMs, T2 and T3 demonstrated a learner-centred approach (see e.g., Brown, 2003; Weimer, 2013). As teacher beliefs may affect teachers' actions in the classroom (Gilakjani & Sabouri, 2017), these activities may reflect their cognitive processes and beliefs regarding language learning, digital competence, and LLMs. For example, it seems that T2 and T3 hold beliefs about the importance of collaborative and student-centred learning when interacting with LLMs. Their decision to structure their lessons around these types of activities furthermore suggests a conscious consideration of approaches that priorities active student participation and peer interaction. Therefore, the emphasis placed by T2 and T3 on peer interaction seems to underscore the importance of aligning learning activities with their pedagogical beliefs to create effective and engaging learning experiences for their pupils. Moreover, T2 and T3 allowed for their pupils to promote critical thinking by designing an exercise where the students could discuss differentiating between real and artificial content generated by AI. Collaborative learning has been suggested to enhance critical thinking (Bradshaw & Lowenstein, 2011; Fung & Howe, 2014; Gokhale, 1995). The reports from T2 and T3 indicate that the group discussions concerning the capabilities and limitations of using ChatGPT furthermore encouraged the students to critically evaluate the LLMs utility, thus fostering critical thinking skills and digital skills which are central aims of the curriculum (Directorate for Education and Training, 2019). Additionally, through these classroom activities, they addressed ethical concerns and focussed on developing digital citizenship among their pupils. According to the PfDK, focussing on these elements are essential to be professional digital competent teachers (Kelentrić et al., 2017, p. 6). Concerning digital didactics, T2 and T3 facilitated a learner centred approach focussing on using LLMs as cognitive tools, rather than a technology based approach where the focus is on accessing technology (Dumont & Istance, 2010).

4.4.5 LLMs as Time-Saving Mechanisms

The teacher interviews revealed that the integration of LLMs within their teaching practices extended beyond task-based student activities and directly addressed workload concerns among teachers. T5 reported that:

I've used ChatGPT for making multiple choice tasks. Those tasks are very laboursome and intensive to do manually. I'm not really a fan of multiple-choice tasks, but recent exams in English were full of these multiple-choice tasks. Me and my colleagues were

discussing how we should incorporate these multiple-choice tasks in the lessons because it's so much work putting together these tasks. Then ChatGPT comes along, and it could put together ten multiple-choice questions for a text in two seconds. (T5)

The construction of multiple-choice tasks was described by T5 as “very laboursome and intensive to do manually”. Automating time-consuming tasks such as the example described by T5 could potentially play a vital role in reducing teachers’ heavy workloads (Carroll et al., 2022). More specifically, question-generation such as the multiple-choice tasks is one suggested way of using LLMs to save time and consequently reduce workload among teachers (Bulut & Yildirim-Erbasli, 2022; Kurdi et al., 2020; Olney, 2023). Qu et al. (2021) proposed a framework to automatically generate question-answer pairs. This can be used in the creation of for example reading comprehension tasks (Kasneci et al., 2023). Moreover, in a recent work by Dijkstra et al. (2022), the researchers used LLMs to generate multiple-choice questions and answers for reading comprehension tasks. Dijkstra et al. (2022) argue that automated generations of quizzes such as the multiple-choice tasks both reduces the burden of designing them manually, as well as being a helpful tool for students to train and test their knowledge.

In the current study, the integration of LLMs from a workload reduction perspective was also related to creating assessment criteria and lesson plans. T4 stated that:

This [the use of LLMs] is a case of saving time. You can use the [same] assessment criteria in other assignments, so you save time. Furthermore, you save time on preparing lessons, or the majority of the lesson. So the one of the main uses of ChatGPT for me is that it saves time. It can give you a framework to work on when preparing your lessons. (T4)

T4 reported that he used LLMs to create assessment criteria for both written and oral assignments. Drawing from the stipulated competence aims outlined in the English curriculum, T4 delegated the language model to formulate assessment criteria aligned with the competence aims. Subsequently, he evaluated and refined these criteria to ensure alignment with the specified competence aims. Moreover, T4 used ChatGPT for lesson planning, a process that many teachers may find time consuming (Merritt, 2016). Kasneci et al. (2023) suggest that LLMs can assist teachers in the creation of lesson plans and activities. T4 used LLMs to create a framework for most of his lessons. Like the assessment criteria creation, he revised the lesson plan generated by LLMs to ensure its accuracy (Kohnke et al., 2023).

Assessing and revising the output received from LLMs are crucial stages in verifying the output's quality (Han et al., 2023; Lo, 2023).

Additionally, T4 made the process of creating lesson plans more efficient. One reported way of using LLMs is that it can assist teachers in the creation of lesson plans and activities (Kasneci et al., 2023). T4's refinement of assessment criteria and lesson plans indicates a reflective practice. By always revising the output from the language model, and fine-tuning the final product, they also demonstrate the interconnection between technical knowledge and utilisation of these technologies. T4's decision to employ LLMs for the creation of assessment criteria reflects his awareness of aligning assessment practices with curriculum objectives. Through his actions, T4 demonstrated an understanding of the role of assessment in promoting student learning and the need for clear and coherent criteria to evaluate student performance accurately. Teachers' use of digital technology differs because they need to consider how such technologies can support the design of teaching and learning activities as well as pedagogical and didactical aspect including what to use, why, how, when, and where (Lund & Hauge, 2011; Røkenes & Krumsvik, 2016).

Observing that the teachers utilised LLMs in both task creation and developing assessment criteria highlights a multifaceted competence that extends beyond mere technical competence. It reflects a nuanced understanding of pedagogical and didactical principles, and a strategic approach to utilise LLMs for workload reduction. As mentioned, some of the teachers were using the LLMs potential for rapid generation of multiple-choice exercises and development of grammar and vocabulary tasks. This utilisation seemed to be driven by the recognition that creating such tasks manually entails considerable time investment. This strategic use was recognised by many of the teachers. Furthermore, as identified by T4, utilised LLMs to create assessment criteria for both written and oral assignments. Drawing from the competence aims outlined in the curriculum, T4 prompted the LLM to formulate assessment criteria aligned with these aims. Subsequently, he evaluated and refined these criteria to ensure alignment with the specified competence objectives. Similarly, this approach of revising the output from the language model was conducted by the teachers when they asked the language model to create multiple-choice, vocabulary, and grammar tasks. There are validated language tests teachers can use that have been created by researchers (see e.g., Alderson, 2009; Council of Europe Council for Cultural Co-operation, 2001; Im & Cheng, 2019). These ways of using LLMs suggest it was all about enhancing the teachers' own efficiency. By adopting LLMs in their own work, the language models served as a

pragmatic solution for time management. In essence, the burden of workload associated with instructional preparation was alleviated. Furthermore, the nuanced application of LLMs in task creation and assessment development underscore that they utilised these digital tools to show digital proficiency.

The integration of LLMs in task creation indicates that some of the teachers possess a nuanced understanding of their instructional needs. Furthermore, it indicates that they are able to use LLMs in appropriate contexts. By using LLMs to generate tasks such as multiple-choice exercises and grammar and vocabulary tasks, the teachers demonstrated an awareness of the time constraints related to create tasks manually. Moreover, they exhibited an awareness of balancing practical considerations regarding workload and resource allocation, as well as instructional objectives.

4.5 Didactical Considerations for Integrating LLMs

The third research question examined what didactical considerations teachers in upper secondary schools reported as guiding their decisions when integrating LLMs into their teaching practice. In essence, this research question examined their motivations for using LLMs in their teaching practice. The overview from Table 5 represents the teachers' motivations for integrating LLMs into their teaching. In general, findings revealed that teachers used LLMs out of curiosity.

Table 5*Overview of Teachers' Motivations for Integrating LLMs*

Teacher	Motivations for integrating LLMs
T1	Firstly, we know this is something that is here to stay. It will be a part of our everyday lives whether we want it or not...With AI (LLMs), I saw the opportunity for those who struggle extra to produce where they couldn't earlier. (my translation)
T2	I saw the effect it had when using it myself. Additionally, I enjoy discussing ethical problems. I also think it's good for my pupils to acquire knowledge of how to use them [LLMs]. (my translation)
T3	I think what made me interested was that you saw examples where language models had been used, and I wanted to know just for my own sake "how good this model really is?". And OK, what's the shortcomings? How can I use it? How can a student use it? What are the pitfalls of using it as well?
T4	It's a tool I wanted to introduce them to. And maybe they can use it in the future, but how will it go? Someone will say "thank you, done!", others will actually learn...What can we do here? Can we learn something? How can it be utilised? Like what works and what doesn't work?
T5	I've always been kind of interested in technology and I have this curiosity about it, but it's not driven by my professional life, it's more because of my private interests. This [technology] is something that I've used in teaching and that's been a topic in my teaching for actually longer than ChatGPT.

4.5.1 Curiosity and Interest as Motivations for Integrating LLMs

The predominant motivation reported by the interviewed teachers for incorporating LLMs in their L2 English teaching was curiosity and interest.

T1's predominantly reported motivation for integrating LLMs in his teaching was regarding his recognition of that LLMs and AI in education is something that is here to stay. He furthermore expressed that AI will be a prominent part of people's daily lives and noted that "if we close our eyes and let it loose, we would be left with the negative consequences and none of the potential positive outcomes with AI" (my translation). T2's motivation of

using it came from her own experience. Experiencing that using LLMs in her own work made her motivated for integrating LLMs in her own teaching. Furthermore, T2 expressed an interest in discussing ethical issues with her pupils, and additionally noted her beliefs on her pupils acquiring the necessary competence to utilise LLMs. T3 and T4 reported similar views on their motivations for integrating LLMs. Whereas T3's predominant motivation was to find out for herself the potential usefulness of LLMs in her teaching, T4 was motivated by the potential capabilities of LLMs. T5, on the other hand, stated that his motivation of using LLMs in his teaching was merely driven by private interests, and not his professional life.

A pilot study with European teachers indicated positive attitudes towards AI within educational settings, and a high motivation to integrate AI related content at school (Kasneci et al., 2023). Motivation and positive attitudes towards promoting PDC among teachers, coupled with positive attitudes towards pedagogical change and effective implementation strategies, are crucial factors for successful integration of digital technologies (Rossi Cordero & Barajas Frutos, 2018). The cultivation and development of teachers' PDC seem to rely on commitment across various areas. Firstly, teachers must have an inherent desire to develop their PDC (Postholm, 2012; Rossi Cordero & Barajas Frutos, 2018). The findings from four of five teacher interviews support this stance. T1's recognition of the presence of LLMs and AI technologies suggests an awareness of both their evolving role in education and society at large. Moreover, his reports were understood how technological developments are changing and expanding, aligning with the PFDK provided by (Kelentrić et al., 2017). T2 and T5 were reportedly motivated to integrate LLMs due to personal experiences with using LLMs. T2 furthermore stated that "when I realised how much ChatGPT had helped me in my own writing, I was eager to teach my pupils' to use it as well". Conversely, T5's integration of LLMs was driven by professional considerations, suggesting beliefs on the lack of pedagogical implications if integrating LLMs. Nevertheless, it seems T5 was more curious if LLMs could have a positive pedagogical effect on his pupils. In contrast, T3 and T4 reported that their motivations were rooted in a desire to explore the functionalities of LLMs, rather than a direct consideration of their potential impact on their own teaching and their pupils' learning outcomes. While the teachers' motivations may differ, these insights shed light on the diverse factors driving teachers' integration of LLMs in their classroom activities.

There is a proposed correlation between teachers' integration of technology and the role of educational institutions (Thannimalai & Raman, 2018). A lack of mutual interests between teachers and leadership can therefore negatively impact development. Concerning

the professional guidance and training to integrate LLMs, all the teachers reported that they largely had to learn independently about both the potential positive outcomes and potential challenges. T1 reported that teachers at his school were invited to a few seminars but noted limited learning outcomes from these sessions. Furthermore, he emphasised that “learning about these AI technologies is equally important to actively using them” (my translation). Similarly, T2 stated that the courses she and her colleagues attended were not very helpful. T3 remarked that the teachers had “not received any help whatsoever”. T4 stated “not really”, and T5 stated “absolutely not”. In contrast, T2 noted positive experiences with cooperating with a digital pedagogist at her school. T2 was the only teacher who mentioned having a digital pedagogist at the school, highlighting the benefits of such support. The reported statements from the interviewed teachers raise concerns about their professional development and PDC. Teachers’ reports of minimal co-operation with school leadership or higher authorities suggest why they felt left to navigate the implications, functionalities, and limitations on their own. It is plausible that the lack of co-operation is not due to leadership’s disinterest in AI technologies but rather a mutual lack of understanding regarding what the opportunities and challenges AI integration entails. Therefore, it seems imperative that the motivation for integrating LLMs and other AI technologies arises from a shared responsibility between teachers and leadership. Collaboration between teachers within and across schools, as well as with leadership, is crucial to promote professional development (Postholm, 2012). Moreover, several studies have suggested the importance of establishing a promoting learning culture for teachers’ professional development (Jurasaitė-Harbison & Rex, 2010; Kennedy, 2016; Rinke & Valli, 2010). These findings thereby suggest that promoting a shared learning culture among teachers and staff could improve the quality of professional development, PDC, and consequently, the integration of AI technologies.

Additionally, three of the five teachers reported a general interest in technology as incentives for integrating LLMs in their teaching. T1 stated that he originally was educated within the field of IT, and that this education was the source from which his interest in technology came from. Moreover, T1 had held some lectures for the teachers at his school teaching them about artificial intelligence and LLMs. Education within the field of IT education was also reported by T5, although this education was only a single semester. T4 had no formal qualifications within IT, yet he had a “spiking interest” in technology from which his fascination of LLMs came out from. It stands to reason that interests may affect the competence of an individual. Having an interest in technology may consequently be beneficial

for their motivation to learn about and integrate LLMs in their teaching. T1, T4, and T5 stood out to be the teachers most interested in technology. They were far more opinionated on their beliefs concerning the utilisation and implications of integrating LLMs. This finding may indicate that interest in technology could affect their beliefs on the potential implications and limitations of LLMs in their teaching. However, this is difficult to determine as beliefs are not good indicators of behaviour (Stainton Rogers, 2011). Nevertheless, whether their beliefs on the functionalities and implications of LLMs were rooted from an interest in technology, the effective use of AI technologies dependent on teachers' pedagogical knowledge (Celik, 2023). Alongside knowledge of pedagogy, the subject didactics should guide the use of technology, as noted by Dumont and Istance (2010). Moreover, the lesson design and teachers' didactical choice for utilising technologies are critical for promoting learning outcomes (Kongsgården & Krumsvik, 2019).

4.5.2 Didactical Considerations in Practice

T2's integration of LLMs was influenced by didactical considerations. Within the context of the English subject curriculum of 2020, which emphasise the significance of developing writing skills as one of the basic skills (Directorate for Education and Training, 2019, p. 4), T2 reported that her utilisation was motivated by using LLMs as a constructive tool to support pupils' writing processes.

As a relatively new educated teacher, I automatically think with a didactical lens in my lesson designing. For example, I've used the "think-pair-share" model actively to get my students to discuss the use of LLMs. Discussing positive and negative sides with everything we do is something I focus on in my teaching. And this will also help them for their writing process and to get them started. (T2, my translation)

Writing is an active process where the student must learn to make successful choices, both conscious and unconscious, learning to operate in a variety of genres and discourses Skulstad (2020, p. 117). Recognising this complex process, T2 discussed the positive and negative sides of using LLMs with her pupils. This communicative discussion seems to align with Stone (2013) where the author described a Vygotskian-inspired analysis of scaffolding. According to Vygotsky, learning first takes place on a social level before it takes place on an individual level (Van de Pol et al., 2010). In the views of Stone (2013), scaffolding is seen as a interpersonal process in which both participants are active participants. The pupils are thereby not a passive participant in the teacher-student interaction. T2 utilised this teacher-student interaction through the *think-pair-share model*, which is described as a learning

strategy to promote critical thinking (Kaddoura, 2013). T2 stated that this interactive classroom discussions led to “develop them to think critically about why we do it the way we do” (my translation).

T2’s instructional design of the think-pair-share model indicated aspects of pre-task training. She facilitated smaller activities (pre-task training) wherein students actively engaged with LLMs. Bernacki et al. (2011) argue the importance of pre-task training as an important factor for pupil’s learning outcomes. Moreover, they underscore the significance of following up pupils’ learning and teachers’ use of technology. T2 noted that through these interactive activities, her pupils were not only exposed to the capabilities of LLMs but also made them “more able what type of questions will give the best answers” (my translation). This approach where T2’s pupils learned to distinguish between optimal prompts to receive desired responses suggests an interesting aspect. It underscores T2’s commitment to empowering students with the necessary abilities to navigate and leverage LLMs effectively as constructive tools to enhance their writing skills. Additionally, this classroom activity may seem to have assisted T2’s pupils in developing critical thinking skills, which is convenient given the increased attention on critical thinking in education (Skulstad, 2023). This didactical choice emphasises T2’s nuanced understanding of LLMs capabilities as a dynamic learning tool within the classroom. By fostering small activities with the pupils, they were exposed to parts of the functionalities of LLMs and presented with a set of crucial skills in formulating proper prompts. This aspect of the instructional design underscores T2’s commitment to cultivate pupils’ critical thinking when engaging with LLMs. Importantly, she exhibits a nuanced exhibition of her digital competence. A teacher must understand how to integrate digital resources into learning processes to aid the pupils achieving the competence aims (Kelentrić et al., 2017). In this context, however, the current study does not provide conclusive evidence regarding whether the T2’s pupils successfully attained the competence aims outlined in the English subject curriculum. This limitation arises from the thesis’ focus on examining the utilisation of LLMs rather than monitoring pupils’ grade or academic results. Nevertheless, Kongsgården and Krumsvik (2019) pointed in their study towards a correlation between teachers’ didactical choices of using technology and pupils’ academic results (Kongsgården & Krumsvik, 2019, p. 157). Despite the absence of direct assessment of pupils’ academic performance, T2’s utilisation of LLMs nonetheless indicate correlations with principles of professional digital competence.

Firstly, T2's demonstration of the proposed dual role of Redecker (2017) was evident. T2's classroom activity showcased that she was aware of her own digital competence alongside her efforts to cultivate her pupils' digital skills. The dual responsibility underscores T2's recognition of the intertwined relationship between T2's and her pupils' digital competence. Secondly, T2's classroom practice aligns with the assertions of Pegalajar-Palomino (2018) regarding the role of digital competence in teachers' professional development. The theoretical framework (see section 2.4.2) delineated emphasises the importance of a clear and foundational understanding of digital competence as a prerequisite for effective professional practice. T2's conscientious integration of LLMs in her activities may indicate her a commitment to ongoing professional development and fostering digital skills among both teachers and pupils. Moreover, T2's approach to teacher professional development aligns with some of the characteristics identified by Desimone (2009) for improving knowledge and skills. She demonstrated awareness of content focus by actively reminding her pupils of the learning objective, such as writing five-paragraph essays and utilising LLMs to scaffold their writing process. Furthermore, T2 actively engaged her pupils with LLMs through several small-scale activities across a longer period in an effort to ensure her pupils' familiarity with LLMs, reportedly to make them "understand how the language model can be used appropriately" (my translation). Additionally, Hwang and Chen (2023) propose using LLMs as a learning peer, facilitating interaction and collaboration between pupils and LLMs. This approach aligns with Skulstad and Touileb's (2024) suggestion of using LLMs for process-oriented writing (see e.g., Skulstad, 2020 on process-oriented writing). In sum, T2's classroom activity exemplifies a multifaceted approach to teacher professional development, incorporating theoretical insights and practical strategies to foster both her own and her pupils' digital competence.

4.5.3 Absence of Didactical Considerations

Despite the motivations discussed, the teacher interview revealed an absence of didactical considerations in the teachers' integration of LLMs in their L2 English teaching. While curiosity and interest were powerful drivers, this absence raised questions about the pedagogical and didactical rationale for integrating LLMs.

First, the absence of didactical considerations in teachers' decision-making processes suggest a potential disconnection between technology adoption and pedagogical objectives. I suggest that this disconnection is not necessarily due to teachers' incapability to incorporate LLMs, but rather that LLMs do not possess a didactical function at this moment (Gilje, 2023).

On whether it is possible to legitimise the use of LLMs in terms of English didactics, T1 stated “as it is now, it is not possible” (my translation). When questioned about what is required for legitimising the use of LLMs in terms of English didactics, T1 stated:

First the development of LLMs must stabilise...Number two is that there are many, many great teachers out there who will not be able to take the step with AI...There is not a sufficient understanding of what an LLMs actually is...In other words, there are a lot of prerequisites before LLMs can be didactically and pedagogically legitimised for language learning (T1, my translation)

In contrast, T4 expressed a bombastic opinion on whether it is possible to legitimise the use of LLMs:

So, you should be a competent digital teacher. ChatGPT is just another tool. You legitimize it with the same view as you legitimize any tool you use in your teaching: If you think it will enhance your students’ knowledge and abilities, it’s fair to use it. And then you have to abide that those benefits are from the potential risks: that the students take shortcuts. (T4)

Both T1 and T4 viewpoints aligns well with Krumsvik (2012) who argue that it is not technology, but the teachers digital didactic competence that is the key when integrating technology in teaching. Teachers should integrate digital resources with pedagogical and subject knowledge (Kelentrić et al., 2017). While one could say that technological advancements have revolutionised teaching and learning, their efficacy seem to hinge on teachers’ ability to use them purposefully and effectively. Furthermore, when using digital technology in teaching, the teachers must consider how the technology can support the learning activities (Lund & Hauge, 2011; Røkenes & Krumsvik, 2016). This stance can underscore the need for a balanced approach to technology integration where digital tools supply and enrich traditional teaching methods rather than replacing them.

T2 was the only teacher who revealed that her integration of LLMs in her L2 English teaching practice was guided by didactical considerations (see section 4.5.2). Her instructional design may be viewed as a representation of a forward-thinking pedagogical and didactical paradigm that integrates with the emerging technologies of AI. By adapting these technologies into the educational curriculum, T2 showed a nuanced understanding of one way of utilising LLMs appropriately. The other teachers interviewed, did however exhibit their professional teacher and their professional digital competence in another way.

Deciding to be reluctant to incorporate LLMs into their teaching practice may indicate that they are aware of the lack of LLMs lack of subject didactics. The first part of solving an issue is to recognise when there is one. From an international perspective, it has been suggested that AI and LLMs may be adequate resources for both teachers and pupils when designing and forming lessons (Crompton & Burke, 2022; Holmes & Tuomi, 2022). Using LLMs sets two premises. Pupils must develop their competence and understanding of how they can be used to promote learning. Conversely, teachers must focus on using variants of LLMs if they can affect pupils' learning. Consequently, one potential rationale behind the decision to withdraw from using LLMs in their teaching practice may stem from a nuanced awareness of pedagogical and didactical needs and instructional objectives. Teachers who prioritise a deep understanding of curriculum requirements and student learning outcomes may want to use other instructional methods that align more closely with the specific competencies targeted in the curriculum. By not integrating LLMs, these teachers may demonstrate an approach where their pedagogical and didactical decisions are guided by a comprehensive understanding of educational objectives rather than a superficial pursuit of AI technologies.

Furthermore, they seem to have found that other instructional methods that prioritise pedagogical efficacy over the integration of technology for technology's sake (Lund et al., 2014). T3 reported that "I'm not interested in how smart ChatGPT is or how well the model works. I'm interested in how smart my students can be". Rather than using LLMs because of its popularity, T3 exhibited a commitment to employ instructional strategies that have been empirically validated and demonstrated to promote language learning. By prioritising evidence-based approaches over technological trends, she underscored her dedication to facilitate meaningful and impactful learning strategies for her pupils. Additionally, this correlates well with theoretical perspectives on teacher cognition (Borg, 2015). Based on her experience as a teacher, she believed that her employed instructional strategies worked better than using LLMs as a supplemental tool in her teaching. Drawing upon her beliefs and experience as a teacher, supported with her teacher professional development, she legitimised her choice to refrain from using LLMs.

4.6 Limitations

The findings from the current study should not be taken as representative of teachers' beliefs, professional digital competence (PDC), nor their didactical competence of integrating LLMs. Based on a single research study, the ability to generalise the findings becomes severely limited (Johnson & Christensen, 2020). Consequently, it is difficult to make quantitative predictions (Johnson & Christensen, 2020). The choice of methods for a research design often involves balancing between the ideal and the practical (Borg, 2019). The sample size for the study was limited to five participants, and one could always assume that a more extensive sample size could potentially enhance the insights gained from this study. However, the research design for this study was affected by what was achievable by one student within two semesters. Moreover, the small sample size restricted the generalisability of the findings to a broader population of teachers. Therefore, the findings from this study should be taken as indications that can be explored in further research.

The novelty of being a researcher revealed room for improvement in the analysis of the interviews. In qualitative interviews, the researcher is the main instrument for obtaining insights, emphasising the importance of a researcher's integrity (Brinkmann & Kvale, 2015). During the interviews, I focused on following the interview guide and being as objective as possible. However, there were occasions when I relieved myself from the interview guide and phrased questions differently. Additionally, I considered some of the questions to be repetitive, and some responses to be superfluous. Although some questions may be considered repetitive, the questions were often rephrased for clarity. The superfluous responses seem to be a result of the participants spontaneously providing materials that was interesting, but not relevant for the scope of the current study. Moreover, there will always be inherent subjectiveness when conducting interviews and analysing the data. In the process of conducting interviews, I may phrase questions differently. For the analysis, I may interpret the data material differently than other researchers. Additionally, the data analysis revealed that more follow-up questions could have contributed to receiving more accurate descriptions of their didactical considerations of integrating LLMs.

The lack of empirical data, research studies, and theoretical perspectives to relate to the findings are limitations to the current study. LLMs in L2 English teaching is a relatively unexplored topic. Establishing a clear and coherent link between this study's findings and research thereby became challenging. To address this issue, I applied relevant empirical data, research studies, and theoretical perspectives I found to be closely aligned with my findings.

However, the selection of theoretical perspectives and research was inherently subjective, based on my interpretation of the data analysis process. Consequently, it is important to acknowledge that other researchers may identify different theories or research as more applicable.

The centre of attention concerns a complex and broad field of L2 English teaching, namely digital technology, and artificial intelligence. The study has focused on teachers' perspectives targeting their understanding and utilisation of LLMs. It should be noted that the scope of the current study sheds light upon the field of LLMs in L2 English teaching from teachers' perspective. How their pupils perceived the integration of LLMs was thereby solely based on teachers' viewpoints. This limited scope of inquiry may limit the overall validity of the findings.

The transcriptions are not included in the appendices due to privacy concerns. Protecting each participant's privacy and addressing associated ethical concerns thereby impacted the study's reliability and validity. Reliability, which rely on the consistency and trustworthiness of findings (Brinkmann & Kvale, 2015; Johnson & Christensen, 2020), was affected because the absence of transcriptions means other researchers cannot fully replicate the study. The study's validity depended on how accurate my interpretations of the findings were. Validity, defined by the accuracy of interpretations of the findings, also faced challenges. Without access to the transcriptions, validating the interpretations becomes difficult. However, the validity was strengthened by reflexivity and member-checking. Ultimately, protecting participants' anonymity and securing their identities and statements from recognition by others prioritised over maintaining maximum transparency.

4.7 Suggestions for Further Research

The findings from this study have essentially investigated how L2 English teachers understand and utilise LLMs in their teaching. This study's sample is limited and qualitative, which makes for several things for further studies. Further research could assess this study's topic using a larger sample from a quantitative perspective to examine whether the indications revealed in this study, e.g., different classroom activities in which LLMs are integrated, are generalisable to a larger population. This way, further studies could validate or discard the main findings from this study. Additionally, further studies could employ methods like observation or video recording to qualitatively analyse in more detail.

From the teacher interviews, several avenues for further research emerged that were beyond the scope of the current study. First, there was a consensus among the teachers to explore the relationship between digital competence and AI competence. While the current study provided insights and indications concerning how and why LLMs are used in the classroom, further studies could delve deeper into whether the integration of LLMs indeed fosters a positive learning environment. Relatedly, it would be beneficial to explore what specific knowledge about AI in education teachers should acquire to effectively integrate LLMs and other AI technologies into their L2 English teaching practices. From the perspective of prompt engineering, developing a proficiency of this competence seems to be positive for gaining a deeper understanding for the capabilities and limitations of LLMs. A further exploration of subject matter and content related prompt engineering would presumably benefit a larger population of L2 English teachers. Such explorations could shed light on the competencies necessary for teachers to navigate the constantly evolving landscape of technology-enhanced learning effectively. Second, it would be interesting to examine the perceived experiences of the integration of LLMs from a learners' perspective. This study only collected teachers' perspectives and their perceived reports on the functionalities and implications of LLMs. Examining learners' perceived experiences could potentially deviate from teachers' reports, and perhaps elucidate any discrepancies between the teachers' and learners' perspectives. Third, all the teachers revealed that they had no professional guidance or training to constructively use LLMs in their teaching. Despite the increased focus on LLMs in education and the evolving and promising implications for students and teachers, it seems from teachers' reports that they were left to themselves. This raises questions of who are responsible for a pedagogical and didactical integration of LLMs in their teaching.

Additionally, it would be of interest to change the scope of further research to encompass how LLMs could be integrated in both teaching and assessment. Given the inherent limitations of privacy concerns, uploading pupils' texts in LLMs is thereby a potential threat to protecting the learners' security. Would it be possible for the Norwegian government to create their own LLM where there is a total transparency of how user's personal information is used? Is there even a need for LLMs to assess the quality of pupils' texts? Which qualities should be focused on when assessing pupils' knowledge and capabilities of a given content? What are we learning in school, to produce a product or learning to produce a product? AI in education has been said to revolutionise the way we teach, which in some cases may be a threat to maintaining the core values embedded in the

curriculum and pedagogical literature. It seems that it is useful to change the focus from how AI in education change teaching practices, to when, why and where to integrate AI technologies in our teaching practices.

5. Conclusions

To conclude, the research findings will be revisited to shed light on some of the study's significant findings. Furthermore, didactical implications of the current study's findings will be addressed in relation to digital competence and LLMs in L2 English teaching.

5.1 Revisiting the Research Questions

This study has collected valuable insights into L2 English teachers' understanding of LLMs, their utilisation of LLMs, and their didactical considerations for integrating LLMs into their teaching through an analysis of teacher interviews and examples of classroom activities.

The first research question explored L2 English teachers' understanding of LLMs and their beliefs regarding their use in teaching. The interviews revealed varying levels of understanding among the teachers. While four of the teachers described LLMs as predictive models, one teacher also described LLMs as being based on neural networks. One teacher initially misinterpreted LLMs but developed nuanced perspectives concerning the limitations and implications over the course of the interview. The interviewed teachers generally expressed beliefs in LLMs' potential as supplements rather than substitutes for human teachers. Moreover, they highlighted ethical concerns related to biases, data privacy, and copyright issues.

The second research questions investigated the classroom activities in which L2 English teachers integrated LLMs. In the analysis of the data material, different classroom activities and preparatory activities were reported. Concerning classroom activities, the teachers integrated LLMs in various ways. Integrating LLMs in writing activities was evidently reported among two of five teachers. Utilising LLMs to generate inspiration and creating frameworks for written assignments was reported among the teachers. Moreover, the analysis revealed that they facilitated exercises for their pupils to utilise LLMs for revising their own texts. Two teachers reported collaborative tasks wherein the teachers facilitated groups discussions for their pupils. In these discussions, the pupils were collaborating in exploring different prompts for desirable generated output. More importantly, the teachers allowed their pupils to try to differentiate between human and AI generated content. These collaborative activities were reported by the respective teachers to promote their pupils' critical thinking on the opportunities and limitations of LLMs, fostering an understanding of LLMs role as supplementary tools in language learning. Additionally, the teachers reported that the group discussions facilitated pupils' critical evaluation of the benefits and ethical

issues concerning the integration of LLMs. In sum, these varied activities were by the teachers considered beneficial towards promoting their own PDC and their pupils' PDC.

Moreover, the integration of LLMs extended beyond student tasks to directly address workload concerns among teachers. The teachers reported that they used LLMs for creating various tasks, particularly multiple-choice, vocabulary and grammar exercises. Additionally, LLMs were employed for developing assessment criteria and lesson plans. This utilisation of LLMs suggests a strategic approach to reduce workload for teachers. Furthermore, it underscores the teachers' nuanced understanding of instructional needs, balancing practical considerations with pedagogical objectives. In essence, the multifaceted integration of LLMs within L2 English teaching reflected their strategic incorporation of LLMs to enhance efficiency in their own work, promote critical thinking among their pupils, and align their teaching practice with pedagogical and curricular objectives. However, it seems crucial to emphasise that their integration of LLMs in their L2 English teaching was not primarily driven by didactical considerations.

The third research question investigated the didactical considerations guiding L2 English teachers' decisions when integrating LLMs into their teaching practices. The teachers' interviews elucidated that the primary motivations for integrating LLMs were curiosity and interest in technology. While curiosity and interest were reported, didactical considerations were less prominent. Only one teacher explicitly reported didactical considerations, aligning the integration of LLMs with curriculum objectives to enhance pupils' writing skills. The teacher reports. Three of five teachers reported a personal interest in technology advancements. For two of the teachers, this interest was also accompanied by formal IT education. All the teachers, however, shared a desire to explore the functionalities of LLMs, expressing an interest in how the capabilities could be leveraged regarding their pedagogical beliefs. However, the absence of didactical considerations raises questions about the pedagogical and didactical rationale for integrating LLMs in their L2 English teaching. Given the novelty of LLMs and AI technologies in education and the lack of professional guidance and training, it seems plausible that the reported integration of LLMs was not motivated by English didactics. Therefore, it seems imperative that further research is necessary to explore the opportunities of developing PDC to constructively integrate AI technologies in the EFL classroom.

5.2 Didactical Implications

This study sheds light on how teachers understand and operationalise LLMs in the context of L2 English teaching. One implication of this study is that it can provide insight into how L2 English teachers utilise this digital technology in their teaching. As highlighted by Casal-Otero et al. (2023), experiences with teaching and learning with AI are still in their infancy. Consequently, it seems crucial that substantial research considering potential challenges, possible opportunities and pedagogic and didactic effects of AI technologies is necessary. Acknowledging the importance of understanding how to leverage the capabilities of AI technologies, such as LLMs, can support a focus on productive and effective integration them in EFL classrooms.

Furthermore, the study introduces some novel dimensions to the pedagogical and didactical strategies the interviewed teachers employed for a constructive integration of LLMs. While ChatGPT and other LLMs have shown impressive capabilities in various domains in education, the field of education have experienced, and will continue to experience, a necessity for balancing between the potential opportunities and challenges related to the inclusion of LLMs. Although the current study has revealed various ways of integrating LLMs in L2 English teaching, the majority of the integration was not guided by didactical considerations, but due to a general interest in technology itself. However, through the analysis and discussion of the findings from this study, the study provides indications that the integration of LLMs in education can indeed be legitimised in terms of English didactics. Ensuring beneficial usage of LLMs across a diverse group of teachers and pupils seems to rely on approaching LLMs and other AI technologies with a critical eye, leveraging the opportunities while maintaining the values embedded in pedagogical and didactical literature (Skulstad & Touileb, 2024).

Promoting PDC, professional development, AI literacy, and digital literacy seem to be essential competences for cultivating critical and reflective behaviour concerning digital technologies. The English subject curriculum underscores that “Digital skills in English involve being able to use digital media and resources” (Directorate for Education and Training, 2019, p. 4). This basic skill applies not only to pupils but also teachers, highlighting teachers’ need for acquiring the necessary competence to aid their pupils navigating in the AI landscape. Considering beliefs on AI technologies as existing on a continuum, ranging from AI enthusiasts to AI sceptics, it seems plausible that teachers and pupils may position themselves anywhere along this spectrum. However, regardless of one’s beliefs on AI

technology, the influence of AI technologies on the field of education is undeniable. The findings in this study may indicate that there is indeed a need for learning about the opportunities and challenges presented by AI technologies to successfully integrate them in teaching practices. In this context, the 'AI train' has become a popular metaphor in discourses concerning AI technology. Faced with the opportunities and challenges brought by AI technologies, the field of education is confronted with a choice: To remain stationary or to embark on the journey.

7. References

- Al-Mahdy, Y., Emam, M. M., & Hallinger, P. (2018). Assessing the contribution of principal instructional leadership and collective teacher efficacy to teacher commitment in Oman. *Teaching and teacher education*, 69, 191-201.
- Alderson, J. C. (2009). Test review: Test of English as a foreign language™: Internet-based test (TOEFL iBT®). *Language Testing*, 26(4), 621-631.
<https://doi.org/https://doi.org/10.1177/0265532209346371>
- Avalos, B. (2011). Teacher professional development in teaching and teacher education over ten years. *Teaching and teacher education*, 27(1), 10-20.
- Ayanwale, M. A., Sanusi, I. T., Adelana, O. P., Aruleba, K. D., & Oyelere, S. S. (2022). Teachers' readiness and intention to teach artificial intelligence in schools. *Computers and Education: Artificial Intelligence*, 3, 1-11.
- Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52-62.
- Bernacki, M. L., Aguilar, A. C., & Byrnes, J. P. (2011). Self-regulated learning and technology-enhanced learning environments: An opportunity-propensity analysis. *Fostering self-regulated learning through ICT*, 1-26.
- Bernius, J. P., Krusche, S., & Bruegge, B. (2022). Machine learning based feedback on textual student answers in large courses. *Computers and Education: Artificial Intelligence*, 3, 1-16.
- Borg, S. (2015). *Teacher cognition and language education : research and practice*. Bloomsbury.
- Borg, S. (2017). Teachers' beliefs and classroom practices. In P. Garrett & J. M. Cots (Eds.), *The Routledge handbook of language awareness* (pp. 75-91). Routledge.
- Borg, S. (2019). Language Teacher Cognition: Perspectives and Debates. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 1149-1170). Springer International Handbooks of Education. https://doi.org/https://doi.org/10.1007/978-3-030-02899-2_59
- Bowman, E. (2022, December 19). *A new AI chatbot might do your homework for you. But it's still not an A+ student*. NPR. Retrieved April 29 from <https://www.npr.org/2022/12/19/1143912956/chatgpt-ai-chatbot-homework-academia>

- Boyce, C., & Neale, P. (2006). *Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input* (Vol. 2). Pathfinder International.
- Bozkurt, A. (2023). Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education*, 18(1), 198-204.
- Bradshaw, M. J., & Lowenstein, A. J. (2011). Debate as a teaching strategy. *Innovative teaching strategies in nursing and related health professions*, 163.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.
- Brinkmann, S., & Kvale, S. (2015). *InterViews: learning the craft of qualitative research interviewing* (3rd ed.). Sage.
- Brookhart, S. M. (2007). Expanding views about formative classroom assessment: A review of the literature. *Formative classroom assessment: Theory into practice*, 22, 43-62.
- Brown, K. L. (2003). From teacher-centered to learner-centered curriculum: Improving learning in diverse classrooms. *Education*, 124(1), 49-54.
- Bulut, O., & Yildirim-Erbasli, S. N. (2022). Automatic story and item generation for reading comprehension assessments with transformers. *International Journal of Assessment Tools in Education*, 9(Special Issue), 72-87.
- Burgess, R. G. (2002). *In the field: An introduction to field research*. Routledge.
- Burri, M., Chen, H., & Baker, A. (2017). Joint development of teacher cognition and identity through learning to teach L2 pronunciation. *The Modern Language Journal*, 101(1), 128-142.
- Carroll, A., Forrest, K., Sanders-O'Connor, E., Flynn, L., Bower, J. M., Fynes-Clinton, S., York, A., & Ziaei, M. (2022). Teacher stress and burnout in Australia: examining the role of intrapersonal and environmental factors. *Social Psychology of Education*, 25(2), 441-469.
- Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: a systematic literature review. *International Journal of STEM Education*, 10(29). <https://doi.org/https://doi.org/10.1186/s40594-023-00418-7>
- Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, 138. <https://doi.org/https://doi.org/10.1016/j.chb.2022.107468>

- Cervone, D., & Pervin, L. A. (2019). *Personality: Theory and Research* (14th; EMEA ed.). Wiley.
- Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30. <https://doi.org/https://doi.org/10.48550/arXiv.1706.03741>
- Council of Europe Council for Cultural Co-operation. (2001). *Common European framework of reference for languages : learning, teaching, assessment*. Cambridge University Press.
- Creswell, J. W., & Creswell, J. D. (2023). *Research design : qualitative, quantitative, and mixed methods approaches* (Sixth.; International student. ed.). Sage.
- Creswell, J. W., & Guetterman, T. C. (2021). *Educational research : planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson Education Limited.
- Crompton, H., & Burke, D. (2022). Artificial intelligence in K-12 education. *SN Social Sciences*, 2(113). <https://doi.org/https://doi.org/10.1007/s43545-022-00425-5>
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational researcher*, 38(3), 181-199. <https://doi.org/https://doi.org/10.3102/0013189X08331140>
- Dijkstra, R., Genç, Z., Kayal, S., & Kamps, J. (2022). Reading Comprehension Quiz Generation using Generative Pre-trained Transformers. iTextbooks@ AIED, Directorate for Education and Training. (2019). *Curriculum in English (ENG01-04)*. Established as a Regulation. The National curriculum for the Knowledge Promotion 2020. Retrieved from <https://www.udir.no/lk20/eng01-04?lang=eng>
- Dörnyei, Z. (2007). *Research methods in applied linguistics : quantitative, qualitative, and mixed methodologies*. Oxford University Press.
- Dumont, H., & Istance, D. (2010). Analysing and designing learning environments for the 21st century. In H. Dumont, D. Istance, & F. Benavides (Eds.), *The nature of learning: Using research to inspire practice* (pp. 19-34). OECD Publishing. <https://doi.org/https://doi.org/10.1787/9789264086487-en>
- Elstad, E. (2023). *Læreren møter ChatGPT*. Universitetsforlaget.
- Erstad, O. (2005). *Digital kompetanse i skolen : en innføring*. Universitetsforlaget.
- Ertmer, P. A., Ottenbreit-Leftwich, A., & York, C. S. (2006). Exemplary technology-using teachers: Perceptions of factors influencing success. *Journal of computing in teacher*

- education*, 23(2), 55-61.
<https://doi.org/https://doi.org/10.1080/10402454.2006.10784561>
- Ertmer, P. A., & Ottenbreit-Leftwich, A. T. (2010). Teacher technology change: How knowledge, confidence, beliefs, and culture intersect. *Journal of research on Technology in Education*, 42(3), 255-284.
<https://doi.org/https://doi.org/10.1080/15391523.2010.10782551>
- Fernández-Batanero, J. M., Montenegro-Rueda, M., Fernández-Cerero, J., & García-Martínez, I. (2022). Digital competences for teacher professional development. Systematic review. *European Journal of Teacher Education*, 45(4), 513-531.
<https://doi.org/https://doi.org/10.1080/02619768.2020.1827389>
- Fung, D., & Howe, C. (2014). Group work and the learning of critical thinking in the Hong Kong secondary liberal studies curriculum. *Cambridge Journal of Education*, 44(2), 245-270. <https://doi.org/https://doi.org/10.1080/0305764X.2014.897685>
- Gilakjani, A. P., & Sabouri, N. B. (2017). Teachers' Beliefs in English Language Teaching and Learning: A Review of the Literature. *English Language Teaching*, 10(4), 78-86.
<https://doi.org/http://doi.org/10.5539/elt.v10n4p78>
- Giray, L. (2023). Prompt engineering with ChatGPT: a guide for academic writers. *Annals of biomedical engineering*, 51, 2629-2633.
<https://doi.org/https://doi.org/10.1007/s10439-023-03272-4>
- Gokhale, A. A. (1995). Collaborative learning enhances critical thinking. *Journal of Technology Education*, 7(1).
- Hagen, A., & Nyen, T. (2009). *Kompetanseutvikling for lærere (Utdanning 2009 - læringsutbytte og kompetanse)*. Statistisk sentralbyrå.
https://www.udir.no/globalassets/upload/forskning/5/utdanning_2009.pdf
- Hammerness, K., Darlin-Hammond, L., Grossman, P., Rust, F., & Shulman, L. (2005). How teachers learn and develop. *The Design of Teacher Education Programs*, 358-389.
- Han, Z., Battaglia, F., Udaiyar, A., Fooks, A., & Terlecky, S. R. (2023). An explorative assessment of ChatGPT as an aid in medical education: Use it with caution. *Medical Teacher*, 1-8. <https://doi.org/https://doi.org/10.1080/0142159X.2023.2271159>
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., & Bittencourt, I. I. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 504-526.
<https://doi.org/https://doi.org/10.1007/s40593-021-00239-1>

- Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542-570.
<https://doi.org/https://doi.org/10.1111/ejed.12533>
- Hwang, G.-J., & Chang, C.-Y. (2023). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, 31(7), 4099-4112.
<https://doi.org/https://doi.org/10.1080/10494820.2021.1952615>
- Hwang, G.-J., & Chen, N.-S. (2023). Editorial position paper: Exploring the potential of generative artificial intelligence in education: Applications, challenges, and future research directions. *Educational Technology & Society*, 26(2).
- Im, G.-H., & Cheng, L. (2019). The test of English for international communication (TOEIC®). *Language Testing*, 36(2), 315-324.
<https://doi.org/https://doi.org/10.1177/0265532219828252>
- Johnson, B., & Christensen, L. B. (2020). *Educational research : quantitative, qualitative, and mixed approaches* (7th.; International student. ed.). SAGE.
- Jurasaitė-Harbišon, E., & Rex, L. A. (2010). School cultures as contexts for informal teacher learning. *Teaching and teacher education*, 26(2), 267-277.
<https://doi.org/https://doi.org/10.1016/j.tate.2009.03.012>
- Kaddoura, M. (2013). Think pair share: A teaching learning strategy to enhance students' critical thinking. *Educational Research Quarterly*, 36(4), 3-24.
- Kantar. (2023). *Medietrender Ung 2023 - ny rapport om ungdommens medievaner*. Kantar.
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., & Hüllermeier, E. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, 103, 102274.
<https://doi.org/https://doi.org/10.1016/j.lindif.2023.102274>
- Kelentrić, M., Helland, K., & Arstorp, A.-T. (2017). *Professional Digital Competence Framework for Teachers*. The Norwegian Centre for ICT in Education Retrieved from <https://www.udir.no/in-english/professional-digital-competence-framework-for-teachers/>
- Kennedy, M. M. (2016). How does professional development improve teaching? *Review of educational research*, 86(4), 945-980.
<https://doi.org/https://doi.org/10.3102/0034654315626800>
- Kerlinger, F. N. (1979). *Behavioral research : a conceptual approach*. Holt, Rinehart and Winston.

- Koehler, M., Mishra, P., Akcaoglu, M., & Rosenberg, J. (2013). The Technological Pedagogical Content Knowledge Framework for Teachers and Teacher Educators. In *ICT Integrated Teacher Education: A resource book*.
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *Relc Journal*, 54(2), 537-550.
<https://doi.org/https://doi.org/10.1177/00336882231162868>
- Kongsgården, P., & Krumsvik, R. J. (2019). Lærerens didaktiske valg i et teknologirikt læringsmiljø: En kasusstudie av selvregulert læring blant elever i videregående skole. *Nordic studies in education*, 39(2), 142-163.
- Krumsvik, R. (2012). The digital school and teacher education in Norway. In R. Schulz-Zander, B. Eickelmann, H. Moser, H. Niesyto, & P. Grell (Eds.), *Jahrbuch Medienpädagogik 9* (pp. 455-480). VS Verlag für Sozialwissenschaften.
https://doi.org/https://doi.org/10.1007/978-3-531-94219-3_20
- Krumsvik, R., & Almås, A. G. (2009). The digital didactic. In R. Krumsvik (Ed.), *Learning in the network society and the digitized school* (pp. 107-139). Nova Science Publishers.
- Krumsvik, R. J. (2023). *Digital kompetanse i KI-samfunnet: et blick på hvordan kunstig intelligens preger livene våre* (1st ed.). Cappelen Damm akademisk.
- Krumsvik, R. J., Egeland, K., Sarastuen, N. K., Jones, L. Ø., & Eikeland, O.-J. (2013). *Sammenhengen mellom IKT-bruk og læringsutbytte (SMIL) i videregående opplæring : hvilken sammenheng er det mellom IKT-bruk og (intendert, subjektivt, og objektivt) læringsutbytte i videregående opplæring : sluttrapport*. Universitetet i Bergen, Institutt for pedagogikk.
- Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and information technologies*, 28, 973-1018.
<https://doi.org/https://doi.org/10.1007/s10639-022-11177-3>
- Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2020). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, 30, 121-204.
<https://doi.org/https://doi.org/10.1007/s40593-019-00186-y>
- Leftwich, A. T. O. (2007). *Expert technology-using teachers: Visions, strategies, and development* [Purdue University].
- Legard, R., Keegan, J., & Ward, K. (2003). In-depth interviews. In J. Ritchie & J. Lewis (Eds.), *Qualitative research practice: A guide for social science students and researchers* (1st ed., pp. 138-169).

- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4). <https://doi.org/What> is the impact of ChatGPT on education? A rapid review of the literature
- Lund, A. (2019). PhD revisited: The teacher as interface. In U. Rindal & L. M. Brevik (Eds.), *English Didactics in Norway: 30 years of doctoral research* (pp. 140-161). Universitetsforlaget. <https://doi.org/10.18261/978-82-15-03074-6-2019-08>
- Lund, A., Furberg, A., Bakken, J., & Engelién, K. (2014). What Does Professional Digital Competence Mean in Teacher Education? *Nordic Journal of Digital Literacy*, 2014, 281-299. <https://doi.org/10.18261/ISSN1891-943X-2014-04-04>
- Lund, A., & Hauge, T. E. (2011). Designs for teaching and learning in technology-rich learning environments. *Nordic Journal of Digital Literacy*, 6(4), 258-271. <https://doi.org/https://doi.org/10.18261/ISSN1891-943X-2011-04-05>
- Merine, R., & Purkayastha, S. (2022). Risks and benefits of AI-generated text summarization for expert level content in graduate health informatics. In *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)* (pp. 567-574). IEEE. <https://doi.org/10.1109/ICHI54592.2022.00113>
- Merritt, E. G. (2016). Time for teacher learning, planning critical for school reform. *Phi Delta Kappan*, 98(4), 31-36. <https://doi.org/https://doi.org/10.1177/0031721716681774>
- Mirchandani, S., Xia, F., Florence, P., Ichter, B., Driess, D., Arenas, M. G., Rao, K., Sadigh, D., & Zeng, A. (2023). Large language models as general pattern machines. <https://doi.org/https://doi.org/10.48550/arXiv.2307.04721>
- Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record*, 108(6), 1017-1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Møgelvang, A., Bjune, A. E., Coelho, R., Cotner, S., Grellscheid, D., & Ellingsen, S. (2023). *Initial (March 2023) uses and perceptions of ChatGPT in a sample of students and instructors at the University of Bergen (UiB), Norway* (978-82-93789-09-3). <https://hdl.handle.net/11250/3106151>
- Moore, S., Nguyen, H. A., Bier, N., Domadia, T., & Stamper, J. (2022). Assessing the quality of student-generated short answer questions using GPT-3. In I. Hilliger, P. J. Muñoz-Merino, T. De Laet, A. Ortega-Arranz, & T. Farrell (Eds.), *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption* (pp. 243-257). Springer. https://doi.org/https://doi.org/10.1007/978-3-031-16290-9_18

- Mueller, J., Wood, E., Willoughby, T., Ross, C., & Specht, J. (2008). Identifying discriminating variables between teachers who fully integrate computers and teachers with limited integration. *Computers & education*, 51(4), 1523-1537.
<https://doi.org/https://doi.org/10.1016/j.compedu.2008.02.003>
- National Council on Measurement in Education, & American Psychological Association. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Nguyen, T. T., Le, A. D., Hoang, H. T., & Nguyen, T. (2021). NEU-chatbot: Chatbot for admission of National Economics University. *Computers and Education: Artificial Intelligence*, 2. <https://doi.org/https://doi.org/10.1016/j.caeai.2021.100036>
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, 31(2), 199-218. <https://doi.org/https://doi.org/10.1080/03075070600572090>
- Norhagen, S., Krumsvik, R., & Røkenes, F. M. (2024). Developing professional digital competence in Norwegian teacher education: a scoping review. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1363529>
- Norwegian Directorate for Education and Training. (2022). *Sluttrapport: Digitalisering gir bedre muligheter for læring [Final report: Digitalisation gives improved possibilities for learning]*. <https://www.udir.no/tall-og-forskning/finn-forskning/rapporter/digitalisering-i-grunnpopplaringen-bedre-muligheter-for-laring/>
- Norwegian Ministry of Local Government and Modernisation. (2020). *National Strategy for Artificial Intelligence*. Retrieved from <https://www.regjeringen.no/en/dokumenter/nasjonalt-strategi-for-kunstig-intelligens/id2685594/>
- NOU 2022: 11. (2022). *Ditt personvern – vårt felles ansvar — Tid for en personvernpolitikk* <https://www.regjeringen.no/no/dokumenter/nou-2022-11/id2928543/>
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187-192.
<https://doi.org/https://doi.org/10.1126/science.adh2586>
- Olney, A. M. (2023). Generating multiple choice questions from a textbook: Llms match human performance on most metrics. In *Empowering Education with LLMs*. AIED Workshops.

- Ottestad, G., Kelentrić, M., & Guðmundsdóttir, G. (2014). Professional Digital Competence in Teacher Education. *Nordic Journal of Digital Literacy*, 9, 243-249.
<https://doi.org/10.18261/ISSN1891-943X-2014-04-02>
- Patton, M. Q. (1987). *How to use qualitative methods in evaluation* (2nd ed., Vol. 4). Sage.
- Patton, M. Q. (2015). *Qualitative research & evaluation methods : integrating theory and practice* (4th ed.). Sage.
- Pegalajar-Palomino, M. C. (2018). Information and communication technologies and inclusive teaching: Perceptions and attitudes of future early childhood and primary education teachers. *Problems of education in the 21st century*, 76(3), 380-392.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In *Handbook of self-regulation* (pp. 451-502). Elsevier. <https://doi.org/10.1016/B978-012109890-2/50043-3>
- Polak, S., Schiavo, G., & Zancanaro, M. (2022). Teachers' perspective on artificial intelligence education: An initial investigation. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (pp. 1-7). Association for Computing Machinery. <https://doi.org/https://doi.org/10.1145/3491101.3519866>
- Postholm, M. B. (2012). Teachers' professional development: A theoretical review. *Educational research*, 54(4), 405-429.
<https://doi.org/https://doi.org/10.1080/00131881.2012.734725>
- Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational researcher*, 29(1), 4-15.
<https://doi.org/https://doi.org/10.3102/0013189X029001004>
- Qu, F., Jia, X., & Wu, Y. (2021). Asking questions like educational experts: Automatically generating question-answer pairs on real-world examination data.
<https://doi.org/https://doi.org/10.48550/arXiv.2109.05179>
- Raina, V., & Gales, M. (2022). Multiple-choice question generation: Towards an automated assessment framework. <https://doi.org/https://doi.org/10.48550/arXiv.2109.05179>
- Rathje, S., Mirea, D.-M., Sucholutsky, I., Marjeh, R., Robertson, C., & Van Bavel, J. J. (2023). GPT is an effective tool for multilingual psychological text analysis.
<https://doi.org/https://doi.org/10.31234/osf.io/sekf5>
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*. <https://doi.org/https://doi.org/10.1016/j.iotcps.2023.04.003>

- Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu*.
- Richardson, P. W., Karabenick, S. A., & Watt, H. M. G. (2014). *Teacher motivation : theory and practice*. Routledge.
- Rinke, C., & Valli, L. (2010). Making adequate yearly progress: Teacher learning in school-based accountability contexts. *Teachers College Record*, 112(3), 645-684.
<https://doi.org/https://doi.org/10.1177/016146811011200302>
- Rodriguez-Torrealba, R., Garcia-Lopez, E., & Garcia-Cabot, A. (2022). End-to-End generation of Multiple-Choice questions using Text-to-Text transfer Transformer models. *Expert Systems with Applications*, 208.
<https://doi.org/https://doi.org/10.1016/j.eswa.2022.118258>
- Røkenes, F. M., & Krumsvik, R. J. (2016). Prepared to teach ESL with ICT? A study of digital competence in Norwegian teacher education. *Computers & education*, 97, 1-20.
<https://doi.org/https://doi.org/10.1016/j.compedu.2016.02.014>
- Roos, M. (2022). *Å forstå fortiden og formidle for framtiden : historisk kunnskap for lærerutdanningen*. Universitetsforlaget.
- Rossi Cordero, A. S., & Barajas Frutos, M. (2018). Digital competence and educational innovation: Challenges and opportunities. *Profesorado-Revista de Curriculum y formacion de profesorado*, 22(3), 317-339.
- Sailer, M., Bauer, E., Hofmann, R., Kiesewetter, J., Glas, J., Gurevych, I., & Fischer, F. (2023). Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning. *Learning and Instruction*, 83, 101620. <https://doi.org/https://doi.org/10.1016/j.learninstruc.2022.101620>
- Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). Artificial intelligence and learning analytics in teacher education: A systematic review. *Education Sciences*, 12(8), 569.
<https://doi.org/https://doi.org/10.3390/educsci12080569>
- Sarsa, S., Denny, P., Hellas, A., & Leinonen, J. (2022). *Automatic generation of programming exercises and code explanations using large language models* (Vol. 1).
<https://doi.org/https://doi.org/10.1145/3501385.3543957>
- Seufert, S., Guggemos, J., & Sailer, M. (2021). Technology-related knowledge, skills, and attitudes of pre-and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior*, 115.
<https://doi.org/https://doi.org/10.1016/j.chb.2020.106552>

- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379-423.
- Skott, J. (2014). The promises, problems, and prospects of research on teachers' beliefs. In H. Fives & M. G. Gill (Eds.), *International handbook of research on teachers' beliefs* (1st ed., pp. 13-30). Routledge.
- Skulstad, A. S. (2020). Teaching writing. In A.-B. Fenner & A. S. Skulstad (Eds.), *Teaching English in the 21st century : central issues in English didactics* (2nd edition. ed., pp. 117-140). Fagbokforlaget.
- Skulstad, A. S. (2023). *Text creation in English teaching*. Universitetsforlaget.
- Skulstad, A. S., & Touileb, S. (2024). Large Language Models and their usage in EAL education. In A. S. Skulstad (Ed.), *Current issues in English teaching* (pp. 139-160). Fagbokforlaget.
- Somekh, B. (2008). Factors affecting teachers' pedagogical adoption of ICT. *International handbook of information technology in primary and secondary education*, 449-460. https://doi.org/https://doi.org/10.1007/978-0-387-73315-9_27
- Song, W., Hou, X., Li, S., Chen, C., Gao, D., Sun, Y., Hou, J., & Hao, A. (2023). An intelligent virtual standard patient for medical students training based on oral knowledge graph. *IEEE Transactions on Multimedia*, 25. <https://doi.org/10.1109/TMM.2022.3205456>
- Stainton Rogers, W. (2011). *Social psychology* (2nd ed.). McGraw-Hill Education.
- Stone, C. A. (2013). What is missing in the metaphor of scaffolding? In *Learning relationships in the classroom* (pp. 156-167). Routledge.
- Strümke, I. (2023). *Maskiner som tenker : algoritmenes hemmelighet og veien til kunstig intelligens*. Kagge forlag.
- Tack, A., & Piech, C. (2022). The AI teacher test: Measuring the pedagogical ability of blender and GPT-3 in educational dialogues. <https://doi.org/https://doi.org/10.48550/arXiv.2205.07540>
- Thannimalai, R., & Raman, A. (2018). The Influence of Principals' Technology Leadership and Professional Development on Teachers' Technology Integration in Secondary Schools. *Malaysian Journal of Learning and Instruction*, 15, 203-228. <https://doi.org/10.32890/mjli2018.15.1.8>
- Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in Teacher–Student Interaction: A Decade of Research. *Educational psychology review*, 22, 271-296. <https://doi.org/https://doi.org/10.1007/s10648-010-9127-6>

- Van Teijlingen, E., & Hundley, V. (2002). The importance of pilot studies. *Nursing Standard (through 2013)*, 16(40), 33.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Weimer, M. (2013). *Learner-centered teaching: Five key changes to practice*. John Wiley & Sons
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. <https://doi.org/https://doi.org/10.48550/arXiv.2302.11382>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16(39), 1-27. <https://doi.org/https://doi.org/10.1186/s41239-019-0171-0>
- Zhu, M., Liu, O. L., & Lee, H.-S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & education*, 143. <https://doi.org/https://doi.org/10.1016/j.compedu.2019.103668>
- Zimmerman, B. J. (2008). Goal setting: A key proactive source of academic self-regulation. In B. J. Zimmerman & D. H. Schunk (Eds.), *Motivation and self-regulated learning* (pp. 267-295). Routledge.

8. Appendices

Appendix A Interview guide (English Version)

Interview Guide

Part 1 – Background Information

- 1.1 Can you share details about your formal qualifications as a teacher, including degrees and any relevant certifications you hold? (Degrees, MA, BA, etc.)
- 1.2 In your current role, which educational level and program are you teaching? (e.g. English VG1 VS)
- 1.3 Could you elaborate on your role and the subjects you teach within this program?
- 1.4 How many years have you been teaching English? Could you highlight the aspects of teaching English that you find most rewarding?

Part 2 – Understanding of language models

- 2.1 Define your understanding of language models and explain their relevance to language learning in your own words.
- 2.2 Discuss your familiarity with various types of language models used in education. Did you choose any specific language model instead of any other?
- 2.3 Describe your familiarity with the technology and algorithms used in large language models.
- 2.4 Can you share your thoughts on some ethical concerns of using language models such as ChatGPT, Bing Chat, or Bard in L2 English teaching.

*If the participant identified some specific ethical concerns, delve deeper into those issues.

**If the participant does not mention any ethical concerns, introduce some concerns to them and discuss them.

Some ethical concerns:

- Privacy Concerns
- Bias

- Cultural Sensitivity
- Unintended Consequences of Automated Feedback
 - Are they learning anything?
- Lack of Critical Thinking Skills
- Limited Cultural Context

2.5a Are you familiar with what the concept of *Digital Bildung* entails? How would you define the term “Digital Bildung” based on your understanding?

2.5b *“Digital Bildung refers to the integrated development of the individual as a whole person, maturing in a digital culture. It therefore entails actively developing a person’s social, cultural, and practical competence in interaction with the digital environment, and being able to link their own digital experiences to the world around them. It also entails a personal maturity, that enables each individual to act in line with social expectations and ethical norms in a digital culture, as well as to reflect critically, and make well-considered and independent decisions”*

How do you perceive this definition of Digital Bildung and your understanding of the concept?

2.5c

The *Professional Digital Competence Framework for Teachers* states:

“A professional, digitally competent teacher...has an insight into the legislation and ethical concerns, as well as development of pupils’ digital Bildung...in a digital and democratic society.”

Based on this excerpt, how do you believe the integration of large language models in your teaching can contribute to the development of teachers’ Digital Bildung?

Part 3 – Your experiences and beliefs about using language models when teaching.

3.1 Describe your general use of language models in teaching, specifying situations like writing sessions, group work, oral presentations, inspiration, idea generation, etc..

3.2 You were asked to bring with you an example of a lesson plan/a written description of a segment of your teaching where you used language models. Please detail this lesson where you used language models, including learning objectives and how the language model supported the lesson.

3.3 Reflect on the lesson you described. Share your overall impressions, highlighting any expected or unexpected positive outcomes or challenges.

3.4 Discuss the challenges you've faced when using language models in lessons, including issues related to technology, student engagement, or other aspects of the teaching process.

3.5 Understanding the student perspective is crucial in evaluating the effectiveness of any tools used in teaching. Can you elaborate on the challenges your students have faced when interacting with language models in your lessons. Consider aspects such as how you prepared them and if you discussed learning aims, potential positive effects, or consequences.

3.5b In your experience, did students use language models as an easy road to the finish line, or were they able to utilize the potential for language learning?

3.6 Based on your teaching experiences with language models, share your insights on how language models can be integrated as constructive tools for language learning.

3.7a Have you undergone any formal training or received professional guidance related to using language models in your teaching? Discuss the formal training or professional guidance you've received regarding the use of language models in teaching. If not, explain how you acquired the necessary skills and knowledge for effective incorporation into your lessons.

3.7b Share your opinion on how teachers should acquire the necessary knowledge to effectively use language models in L2 English teaching.

3.8a Based on your experiences as a teacher, how do you perceive the positive impacts of incorporating large language models into your teaching practice?

3.8b On the flip side, based on your experiences as a teacher, how do you perceive the concerns or challenges of incorporating language models into your L2 English teaching practice?

3.9 Building on the ethical considerations discussed earlier, do you believe there are specific ethical guidelines or precautions that should be followed when integrating large language models into L2 English teaching?

Part 4 – Questions of your didactical considerations when integrating language models in your teaching

4.1 What motivated you to integrate language models into your L2 English teaching?

*Were there specific educational goals, student needs, or teaching challenges that prompted the incorporation of these tools? Did you do it because it was a “hot topic” in schools and education?

4.2 The *Professional Digital Competence Framework for Teachers* states:

“A professional, digitally competent teacher possesses knowledge of subject didactics relevant to the practice of their profession in a digital environment.”

Based on this excerpt, describe your integration of language models from a didactical point of view. How would you say it’s possible to legitimate the use of language models in terms of English didactics?

4.3 Share the didactical considerations you focused on when integrating language models into your teaching.

4.4 Lastly, based on your use of language models in teaching, how do you depict the future of L2 English teaching and language models?

Appendix B Interview guide (Norwegian Version)

Intervjuguide

Del 1 - Bakgrunnsinformasjon

1.1 Kan du gi detaljar om dine formelle kvalifikasjonar som lærar, inkludert grader og eventuelle relevante sertifiseringar du har? (Grader, MA, BA osv.)

1.2 Kva for eit utdanningsnivå og program underviser du på i di noverande stilling? (f.eks. engelsk VG1 VS)

1.3 Kan du utdjupe rolla di og kva fag du underviser i på dette programmet?

1.4 Kor lenge år har du undervist i engelsk? Kan du trekke frem de aspekta ved engelskundervisninga som du synes er mest givande?

Del 2 – Forståing av språkmodellar

2.1 Definer korleis du forstår språkmodellar og forklar med egne ord kva betydning dei har for språklæring.

2.2 Diskuter din kjennskap til ulike typar språkmodellar som du brukar i undervisninga. Har du vald éin bestemt språkmodell i staden for ein annan?

2.3 Beskriv kjennskapen din til teknologien og algoritmane som blir brukt i språkmodellar.

2.4 Kan du dele dine tankar om nokre etiske bekymringar eller tankekors ved bruk av språkmodellar som ChatGPT, Bing Chat eller Bard i L2-engelskundervisningen?

*Om deltakaren identifiserte nokre spesifikke etiske bekymringar, kan du gå djupare inn i desse.

**Om deltakaren ikkje nemner nokre etiske bekymringar, kan du presentere nokre av dem og diskutere dem.

Nokre etiske bekymringar:

- Bekymringar knytt til personvern
- Fordommar/bias
- Kulturell sensitivitet
- Utsikta konsekvensar av automatiserte tilbakemeldingar

- Lærer dei noko?
- Mangel på evne til kritisk tenking
- Avgrensa kulturell kontekst

2.5a Kjenner du til kva omgrepet «digital danning» (Digital Bildung) inneberer? Korleis vil du definere omgrepet "digital danning" basert på di forståing?

2.5b "Digital danning (Digital Bildung) er ein prosess der eit menneske formar identiteten sin i ein digital kontekst. Det inneberer å aktivt utvikle sin sosiale, kulturelle og praktiske kompetanse i samspel med dei digitale omgivnadane og å kunne knytte eigne digitale erfaringar til verda omkring seg. Det inneberer også ei personleg modning som set den enkelte i stand til å handle i tråd med sosiale forventningar og etiske normer i en digital kultur, samt å reflektere kritisk og fatte veloverveide og sjølvstendige avgjersler (kjelde: Senter for IKT i utdanninga).

Korleis oppfattar du denne definisjonen av digital danning og di forståing av omgrepet?

2.5c

I rammeverket for profesjonell digital kompetanse for lærarar står det:

«En profesjonell, digitalt kompetent lærer...har innsikt i lovverk så vel som etiske problemstillinger knytt til digital danning og deltaking i det digitale og demokratiske samfunnet. Læreren bidreg til å utvikle elevane si digitale dømmekraft, forståing og evne til å handle i tråd med dette»

Basert på dette utdraget, korleis trur du at integrering av store språkmodellar i undervisninga kan bidra til å utvikle lærarane si digitale danning?

Del 3 - Dine erfaringar og oppfatningar om bruk av språkmodellar i undervisninga.

3.1 Beskriv din generelle bruk av språkmodellar i undervisninga, og spesifiser situasjonar som skriveøker, gruppearbeid, munnlege presentasjonar, inspirasjon, idégenerering osv.

3.2 Du blei bedt om å ta med deg eit eksempel på en øktplan/ei skriftleg skildring av ein del av undervisninga din der du brukte språkmodellar. Beskriv denne leksjonen der du brukte språkmodellar, inkludert læringsmål og korleis språkmodellen støtta økta.

3.3 Reflekter over økta du skildra. Fortel om heilskapsinntrykket ditt, og framhev eventuelle forventa eller uventa positive resultat eller utfordringar.

3.4 Diskuter utfordringane du har møtt når du har brukt språkmodellar i undervisninga, inkludert problem knytt til teknologi, elevengasjement eller andre aspekt ved undervisningsprosessen.

3.5 Å forstå elevperspektivet er avgjerande for å kunne evaluere effektiviteten av verktøy som blir brukt i undervisninga. Kan du utdjupe kva utfordringar elevane har møtt når de har interagert med språkmodellar i undervisninga? Tenk på aspekt som korleis du førebudde dem, og om du diskuterte læringsmål, potensielle positive effektar eller konsekvensar.

3.5b Opplevde du at elevane brukte språkmodellar som ein enkel veg til målet, eller var de i stand til å utnytte potensialet for språklæring?

3.6 Basert på dine undervisningserfaringar med språkmodellar, kan du si noko om korleis språkmodellar kan integrerast som konstruktive verktøy for språklæring?

3.7a Har du gjennomgått formell opplæring eller fått fagleg rettleiing knytt til bruk av språkmodellar i undervisninga? Diskuter den formelle opplæringa eller faglege rettleiinga du har fått om bruk av språkmodellar i undervisninga. Dersom du ikkje har fått noko form for opplæring, forklar korleis du har tileigna deg dei nødvendige ferdigheitene og kunnskapen for å kunne bruke dei effektivt i undervisninga.

3.7b Kan du dele tankar om korleis lærarar bør tileigne seg den nødvendige kunnskapen for å kunne bruke språkmodellar effektivt i L2-engelskundervisningen?

3.8a Basert på dine erfaringar som lærer, korleis oppfattar du dei positive effektane av å integrere store språkmodellar i undervisninga?

3.8b På den andre sida, basert på dine erfaringar som lærer, korleis oppfattar du bekymringane eller utfordringane ved å integrere språkmodellar i L2-engelskundervisningen?

3.9 Med utgangspunkt i de etiske vurderingane/bekymringane som er diskutert tidlegare, meiner du at det finnes spesifikke etiske retningslinjer eller forholdsreglar som bør følgast når ein integrerer store språkmodellar i L2-engelskundervisningen?

Del 4 - Spørsmål om dine didaktiske vurderingar når du integrerer språkmodellar i undervisninga di

4.1 Kva motiverte deg til å integrere språkmodellar i din L2-engelskundervisning?

* Var det spesifikke pedagogiske mål, elevbehov eller undervisningsutfordringar som førte til at du tok i bruk disse verkøya? Gjorde du det fordi det var et "hot topic" innan skule og utdanning?

4.2 I rammeverket for profesjonell digital kompetanse for lærarar står det:

"En profesjonell, digitalt kompetent lærer har... fagdidaktisk kunnskap relevant for profesjonsutøvinga i digitale omgivnader."

Basert på dette utdraget, beskriv korleis du integrerer språkmodellar frå et didaktisk synspunkt. Korleis vil du seie at det er mogleg å legitimere bruken av språkmodellar når det gjelder engelskdidaktikk?

4.3 Fortel kva for didaktiske vurderingar du fokuserte på da du integrerte språkmodellar i undervisninga.

4.4 Til slutt, basert på din bruk av språkmodellar i undervisninga, korleis ser du for deg framtida for L2-engelskundervisning og språkmodellar?

Appendix C Information to participants

Informasjonsskriv til forskingsprosjektet «Large Language Models in L2 English Upper Secondary Teaching: Exploring Teachers' Understanding and Utilisation»

Dette er eit informasjonsskriv der du som mottakar blir førespurd om å delta i eit forskingsprosjekt der føremålet er å undersøke korleis engelsklærarar forstår språkmodellar, og korleis bruken av desse blir operasjonalisert i samband med undervisning i engelsk på vidaregåande skule. I tillegg vil prosjektet utforske kva didaktiske vurderingar som ligg til stades i bruken av språkmodellar i undervisning. Dette skrivet gir deg informasjon om måla for prosjektet og kva ei eventuell deltaking inneber for deg.

Føremål

Føremålet med forskingsprosjektet er å undersøke korleis lærarar forstår språkmodellar, og korleis dei nyttar språkmodellar i undervisninga si i den norske vidaregåande skulen. I tillegg skal prosjektet undersøke kva for didaktiske grunngevingar som ligg i botn for å nytte språkmodellar i undervisninga. Forskingsprosjektet er ein del av ei masteroppgåve innanfor engelsk fagdidaktikk. Målgruppa for forskingsprosjektet er lærarar som underviser i engelsk på vidaregåande skular. Prosjektet skal gjennomførast som ein kvalitativ studie realisert gjennom djupneintervju av lærarar i vidaregåande skule. Prosjektet har det mål å avdekke korleis lærarar forstår språkmodellar, korleis dei brukar dei, og om dei er didaktisk medvitne på bruken av språkmodellar.

Opplysningane som vert samla inn i gjeldande forskingsprosjekt skal berre brukast som datamateriale for masteroppgåva. Masteroppgåva i seg sjølv kan bli publisert offentleg i Universitetet i Bergen sine databasar, og kan følgjeleg bli referert til av andre. Alle opplysningar om deg som deltakar vil bli anonymisert og det vil ikkje vere mogleg å spore attende til deg på noko vis. Det er likevel avgjerande å presisere ansvaret av teieplikta både i materialet som blir innhenta frå intervjusituasjonen. Med andre ord er det essensielt at det ikkje førekjem opplysningar som kan identifisere einskildpersonar eller avsløre teiepliktig informasjon. Eg som prosjektansvarleg vil ta omsyn til dette i arbeidet med innsamla materiale.

Kven er ansvarleg for forskingsprosjektet?

Institutt for Framandspråk ved det Humanistiske Fakultet ved Universitetet i Bergen er ansvarleg for personopplysningane som blir behandla i prosjektet. Masterstudent Bendik Vik er ansvarleg for forskingsprosjektet. Masteroppgåva vert skriven som ein del av lektorutdanning i framandspråk ved Institutt for Framandspråk ved Universitetet i Bergen.

Kvifor får du spørsmål om å delta?

Du får spørsmål om å delta fordi du er ein lærar som underviser i engelsk på vidaregåande skule.

Kva inneber det for deg å delta?

Om du vel å delta i dette prosjektet inneber det at du deltek i eit intervju á 45-60 minutt. Sjølv intervjuet vil bli utforma og leia av prosjektansvarleg. Intervjuet er delt inn i fem delar:

- Del 1 – Noko bakgrunnsinformasjon om deg.
- Del 2 – Di forståing av språkmodellar
- Del 3 – Dine erfaringar og tankar kring bruk av språkmodellar i engelskfaget.
- Del 4 – Spørsmål knytt til didaktiske vurderingar kring bruken av språkmodellar.

Intervjua vert gjennomført med lydopptak. Lydopptaka vert transkribert. Lydopptaka blir oppbevart og lagra forsvarleg i Universitetet i Bergen sitt lagringssystem, og sletta med omsyn til gjeldande personvernregelverk. Intervjuet vil ta ca. 45-60 minutt.

Personopplysningar som blir samla inn om deg er namn og bakgrunnsinformasjon (noverande arbeid). Namn er berre for å skilje dei aktuelle deltakarane frå kvarandre. Alt som blir nemnd i intervjuet vert anonymisert og det er såleis ikkje mogleg å spore noko av innhaldet i intervjuet attende til deg som deltakar. Kvar deltakar i prosjektet vil bli gitt eit pseudonym. Den einaste staden den aktuelle deltakaren sitt namn vil syne, er underskrift på samtykkeskjema. Dette skjemaet vil ikkje bli gjort tilgjengeleg for andre, og er berre ein formalitet.

Du som deltakar i prosjektet står fritt til å velje om intervjuet skal gjennomførast på norsk eller engelsk. Om du vel å gjennomføre på norsk, vil eg vere naud til å omsette den norske transkripsjonen til engelsk. Du som deltakar vil bli tilsendt denne transkripsjonen for validering.

Det er frivillig å delta

Dette er frivillig å delta i prosjektet. Om du vel å delta, kan du som deltakar når som helst trekke attende samtykket utan årsak. Følgjeleg vil alle personopplysningar verte sletta. Det vil ikkje ha negative konsekvensar for deg om du ikkje ønskjer å delta, eller om du vel å trekke deg på eit seinare tidspunkt. Deltakarar i prosjektet står fritt til å la vere å svare på spørsmål i intervjuet dersom ynskjeleg.

Kort om personvern

Eg vil berre bruke opplysningane om deg til føremåla skildra i dette skrivet. Eg behandlar personopplysningane konfidensielt og i samsvar med personvernregelverket. Du kan lese meir om personvern på neste side

Med venleg helsing

Bendik Vik

Kimberly Marie Skjelde

Masterstudent

Rettleiar

Personvernet ditt – korleis vert dine opplysningar handsama og brukt?

Eg vil berre bruke opplysningane om deg til føremåla skildra i dette skrivet. Dine opplysningar bli handsama konfidensielt og i samsvar med gjeldande personvernregelverk. Det blir tatt lydopptak av intervjuet, og desse lydopptaka blir handsama og sletta i samsvar med gjeldande personvernregelverk. Masterstudent Bendik Vik og rettleiar Kimberly Marie Skjelde vil ha tilgang til opplysningane som vert samla inn i forskingsprosjektet. Vi vil sikre at ingen uvedkommande får tilgang til personopplysningar ved å anonymisere alt materiale som vert samla inn, samt lagre datamateriale forsvarleg i Universitetet i Bergen sitt lagringssystem.

Kva gir oss rett til å handsame personopplysningar om deg?

Eg handsamar opplysningar om deg basert på ditt samtykke. Gjeldande prosjekt er registrert i RETTE - System for Risiko og ETTERlevelse. Behandling av personopplysningar i forskingsprosjekt og studentoppgaver ved UiB.

Rettane dine

Så lenge du kan identifiserast i datamaterialet, har du rett til:

- Innsyn i kva for opplysningar eg behandlar om deg, og å få utlevert ein kopi av desse opplysningane.
- Å få retta opplysningar om deg som er feil eller på noko vis misvisande.
- Å få sletta personopplysningar om deg.
- Å sende klage til Datatilsynet om behandlinga av dine personopplysningar.

Eg vil gje deg ei grunngjeving dersom eg meiner at du ikkje kan identifiserast, eller at rettane dine ikkje kan utøvast.

Kva skjer med personopplysningane dine når forskingsprosjektet vert avslutta?

Prosjektet vil etter planen avsluttast seinast 30.09.2024. Ingen namn eller personlege opplysningar vil bli nemnde i oppgåva. Anonymiserte opplysningar, innsamla materiale og lydopptak frå intervju blir sletta etter at masteroppgåva er levert og godkjend.

Spørsmål

Dersom du har spørsmål til studien, ønskjer å vite meir om rettane dine eller nytte deg av rettane dine, ta kontakt med:

- Masterstudent Bendik Vik på mail: bendik.vik@student.uib.no, eller på telefon: 992 79 265.
- Rettleiar Kimberly Marie Skjelde på mail: kimberly.skjelde@uib.no
- Personvernombod ved Universitetet i Bergen: Janecke Helene Veim. Mail: personvernombud@uib.no

Appendix D Consent form

Samtykkeerklæring

Eg har motteke og forstått informasjon om prosjektet «Large Language Models in L2 English Upper Secondary Teaching: Exploring Teachers' Understanding and Utilisation » og har fått høve til å stille spørsmål. Eg samtykker til:

- Å delta i kvalitativt djupneintervju.

Eg samtykker til at opplysningane mine kan behandlast fram til prosjektet er avslutta.

Namn (Skriv tydeleg), signatur, dato

Appendix E Approval for using private recording equipment



UNIVERSITETET I BERGEN

Institutt for framandspråk

Til den det måtte vedkome

Dato

10.10.2023

Stadfesting ved bruk av privat opptaksutstyr

Institutt for framandspråk stadfester med dette at Bendik Vik 11.02.1997 er student ved masterprogrammet i Lektorprogram i Framandspråk Engelsk ved Institutt for framandspråk, Universitetet i Bergen.

I samband med gjennomføring av intervju til masteroppgåva, treng Bendik Vik å nytte privat opptaksutstyr. Institutt for framandspråk stadfester med dette at vi godkjenner bruken av privat opptaksutstyr.

Desse forholda ligg til grunn for stadfestinga

- studenten må setje seg inn i relevant regelverk, og følge dette
- studenten må bruke ei sikker løysing for handsaming av personopplysingar, som til dømes SAFE (Sikker Adgang til Forskingsdata og E-infrastruktur)
- persondata skal så raskt som mogleg fjernast frå privat eining og ikkje delast utover det som er tillate i regelverket/godkjenninga av prosjektet

Nyttige lenker

[SAFE](#)

[Datatilsynet - Personvernregelverket](#)

Denne stadfestinga skal signerast av student og administrasjonssjef ved Institutt for framandspråk.

student

administras • nssjef,
Institutt for f amandspråk

Institutt for framandspråk
Telefon 55582340

post@if.uib.no

Postadresse
Postboks 7805

5020 Bergen

Besø4Qdresse
Sydnesplassen 7

5007 Bergen

Appendix F Examples of participation requests

1. Participation request to schools

Hei!

Du får denne e-posten i forbindelse med mitt masterprosjekt om kunstig intelligens og engelskfaget. Eg søker hjelp til å verve aktuelle kandidatar til å vere informantar til mitt masterprosjekt.

Sjølv om eg potensielt ber om mykje, håpar eg at du kan ha kjennskap til nokon som kan vere aktuelle til å delta i prosjektet.

Målgruppe for informantar: **Engelsklærer i studieførebuande klassar i vidaregåande skule som har noko/litt/mykje erfaring med bruk av kunstig intelligens (språkmodellar) i undervisninga si.**

Kort skildring av prosjektet:

Mitt namn er Bendik og er masterstudent på lektorutdanninga ved Universitetet i Bergen. Eg arbeider for tida med mitt masterprosjekt i engelsk fagdidaktikk.

I prosjektet ønsker eg å undersøke og forstå lærarar sitt perspektiv på bruk av kunstig intelligens og språkmodellar i undervisninga si.

Gjennom eit djupneintervju ønsker eg å sjå på din forståing av språkmodellar, korleis du integrerer språkmodellar, og kva didaktiske vurderingar som ligg til grunn for valet ditt om å nytte språkmodellar i undervisninga di. Intervjua vil bli gjort opptak av og transkribert for vidare analyse av datamaterialet.

Personvern

Opplysningane som vert samla inn i gjeldande forskingsprosjekt skal berre brukast som datamateriale for masteroppgåva. Masteroppgåva i seg sjølv kan bli publisert offentleg i Universitetet i Bergen sine databasar, og kan følgjeleg bli referert til av andre. Alle opplysningar om deg som deltakar vil bli anonymisert og det vil ikkje vere mogleg å spore attende til deg på noko vis. Det er likevel avgjerande å presisere ansvaret av teieplikta både i

materialet som blir innhenta frå intervjusituasjonen. Med andre ord er det essensielt at det ikkje førekjem opplysningar som kan identifisere einskildpersonar eller avsløre teiepliktig informasjon. Eg som prosjektansvarleg vil ta omsyn til dette i arbeidet med innsamla materiale.

I forbindelse med mitt masterprosjekt lurar eg på om du kjenner til nokon som underviser engelsk på studieførebuande ved vidaregåande skule som kan vere aktuelle kandidatar til å delta i eit intervju med meg om akkurat deira forhold til språkmodellar som undervisningsverktøy?

Kanskje kan du som lesar dette hjelpe meg? Om du meiner denne e-posten ikkje rakk fram til riktig mottakar, kan du kanskje rettleie meg til meir aktuelle mottakarar? Kanskje kjenner du ein lærar eller to som er i målgruppa?

Har du spørsmål til prosjektet mitt, så gjerne ta kontakt for en uforpliktande samtale.

Håpar på eit kjekt svar.

Alt godt til alle flinke lærarar og andre tilsette rundt om kring på skulane!

Beste helsing,

Bendik Vik

Masterstudent på Lektorutdanning i Framandspråk ved Universitetet I Bergen

2. Participation request directly to teachers

Hei!

Håpar alt står bra til! Eg har ein førespurnad til deg.

I forbindelse med mitt masterprosjekt om kunstig intelligens og engelskfaget, så søker eg hjelp til aktuelle kandidatar som kan vere informantar til mitt masterprosjekt.

Eg forstår at du og andre lærarar har mykje å gjere på i lærarkvardagen. Sjølv om eg potensielt ber om mykje, håpar eg at du kanskje kan tenke deg å delta, eventuelt ha kjennskap til nokon som kan vere aktuelle til å delta i prosjektet.

Målgruppe for informantar: **Engelsklærer i vidaregåande skule som har noko/litt/mykje erfaring med bruk av kunstig intelligens (språkmodellar) i undervisninga si.**

Kort skildring av prosjektet:

Eg arbeider for tida med mitt masterprosjekt i engelsk fagdidaktikk.

I prosjektet ønsker eg å undersøke og forstå lærarar sitt perspektiv på bruk av kunstig intelligens og språkmodellar i undervisninga si.

Gjennom eit djupneintervju ønsker eg å sjå på di forståing av språkmodellar, korleis du integrerer språkmodellar, og kva didaktiske føringar som ligg til grunn for valet ditt om å nytte språkmodellar i undervisninga di. Intervjua vil bli gjort opptak av og transkribert for vidare analyse av datamaterialet.

Personvern

Opplysningane som vert samla inn i gjeldande forskingsprosjekt skal berre brukast som datamateriale for masteroppgåva. Masteroppgåva i seg sjølv kan bli publisert offentleg i Universitetet i Bergen sine databasar, og kan følgjeleg bli referert til av andre. Alle opplysningar om deg som deltakar vil bli anonymisert og det vil ikkje vere mogleg å spore attende til deg på noko vis. Det er likevel avgjerande å presisere ansvaret av teieplikta både i materialet som blir innhenta frå intervjusituasjonen. Med andre ord er det essensielt at det ikkje førekjem opplysningar som kan identifisere einskildpersonar eller avsløre teiepliktig informasjon. Eg som prosjektansvarleg vil ta omsyn til dette i arbeidet med innsamla materiale.

Kanskje kan du hjelpe meg? Og kanskje kjenner du ein lærar eller to som er i målgruppa?

Har du spørsmål til prosjektet mitt, så gjerne ta kontakt for en uforpliktande samtale.

Håpar på eit kjekt svar.

Alt godt til deg!

Beste helsing,

Bendik Vik

Masterstudent på Lektorutdanning i Framandspråk ved Universitetet I Bergen