DESIGNING AI SOFTWARE FOR LARGE CLASSROOM ENGAGEMENT

INTERACTIVE LEARNING AT SCALE: LEVERAGING GENERATIVE AI TO IMPROVE ENGAGEMENT AND PARTICIPATION IN LARGE CLASSROOM SETTINGS

By STEPHANIE EVELINE KOEHL, BASC

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science - Computer Science

McMaster Univ.

MASTER OF SCIENCE - COMPUTER SCIENCE (2024)

Hamilton, Ontario, Canada (Dept. Computing and Software)

TITLE:	Interactive Learning at Scale: Leveraging Generative AI to Improve Engagement and Participation in Large Classroom Settings
AUTHOR:	Stephanie Eveline Koehl BASc (Human Behaviour), McMaster University
SUPERVISOR:	Dr. Christopher K Anand
NUMBER OF PAGES:	xiv, 152

Abstract

This thesis presents the design of an educational tool that enhances student engagement and interaction during group presentations in large classroom settings. Specifically, the study aimed to create a tool that streamlines the management of questions and participation, making the process more efficient and equitable for students and instructors.

The research explored three primary questions: (1) How can educational software be designed to increase engagement and participation during student presentations? (2) How can AI be used to assist in tasks traditionally performed by professors, such as managing Q&A sessions? (3) How does the application of design thinking, particularly the empathy stage, influence the development of effective educational tools?

Students provided ample feedback on improving the course and detailed explanations for their preferences. Qualitative methods including reflexive thematic analysis were used to process this volume of feedback. Descriptive statistics, confusion matrices, and Kappa scores were used to ensure the integrity of the analysis. An open-source large language model, Meta's LLaMA, was implemented to automate the selection and clustering of questions during student-led Q&A sessions, with these results compared against instructor-selected questions.

AI-driven question selection matched the effectiveness of instructor selections and enhanced efficiency, significantly reducing the logistical burden on educators while sustaining student engagement. Additionally, the research gathered extensive data on students' experiences within the university classroom, with particular attention to issues such as anxiety, group dynamics, and disengagement. A paper prototype was developed to address these challenges, leveraging AI to foster interaction and improve peer-to-peer communication.

These results have broader implications for educational technology, showing how AI could foster deeper student involvement and provide instructors with tools to manage participation effortlessly at scale, improving the overall learning experience.

Dedication

I owe many thank yous.

First, to my supervisor and friend, Dr. Christopher Anand, for seeing potential in me over a decade ago when I doubted I had what it took to pursue a technical degree. Your flexibility, enthusiasm and support in allowing me to pursue research interests beyond your immediate goals have been an uncommon gift in academia, and for that, I am deeply grateful.

To my family, especially Mom and Dad, and in memory of Baba and Zelda. To my Dad, for the countless study-break lunches that saved me from malnutrition and certain death. To Christina, for 15 years of unwavering support, laughter, and friendship which has been an anchor throughout my life. To Anthony, for being a sweet perfect angel gift of a boyfriend.

To my wonderful friends who kept me sane, especially Sara, Dien, Chris S., Emilie, Jimmy, Pedram, Milena, and Brendan. For Savannah, Madison, Charlotte, Audrey and the next generation of strong, brilliant, capable women.

By the time you read this, it will already be outdated. The rapid evolution of this field is its greatest challenge and what makes it most alluring. My hope is that this work serves as a snapshot of AI in education as it was in 2023. For me, it has served as a canvas for the ideas that captivated me during my academic career.

Engaging deeply with this project has been challenging and transformative, and I am profoundly grateful for the journey.

Table of Contents

Abstract **iii**

Dedication **v**

Abbreviations **xiii**

1 Introduction and Background **1**

1.1 Evolution of Classroom Dynamics 2

1.2 Developing Student-Centered Educational Tools 5

1.3 Research Questions 10

1.4 Contributions 11

2 Literature Review **13**

2.1 Research Methods 13

2.2 Using Dashboards to Improve Educational Outcomes 18

2.3 Large Language Model Ranking 19

2.4 Generative AI in Education 20

3 Methodology **28**

3.1 Research Design Overview 29

3.2 Feedback Collection 34
3.3 Qualitative Coding and Reflexive Thematic Analysis 36
3.4 Quantitative Analysis 41

4 Initial Prototype Design **48**
4.1 Course Design Overview 48
4.2 Design Features and Considerations 54
4.3 Summary 55

5 Prototype Feedback & Results **57**
5.1 Summary of Feedback from Instructional Staff 58
5.2 Summary of Findings from Students 60
5.3 Presentation of Questionnaire Data 61
5.4 Presentation of Pre-Agreement Data 107

6 Proposed Design: Interaction.ai **112**
6.1 Design Goals 112
6.2 Wireframes 114

7 Limitations and Future Work **139**
7.1 Critique and Alternatives to Design Thinking 139
7.2 Future Work 145
7.3 Conclusion 147

List of Figures

1.1 Stages of the Design Thinking Process 10

3.1 Histogram of inter-coder agreement as measured by Cohen's Kappa for all questions and codes identified by at least one rater. The small negative values indicate instances of no agreement on rare codes, highlighting the challenge of coding infrequent themes. However, for more common codes, the histogram shows strong inter-rater reliability, with the majority of Kappa values falling between 0.3 and 0.9, indicating moderate to high agreement. 43

3.2 Question 1.1 Pre-agreement confusion matrices for all used codes. Excludes 100% agreement or 100% disagreeement 47

5.1 Colour Legend 60

5.2 [Q1.1] In your opinion, what is the most valuable part of the university classroom experience? 62

5.3 [Q1.2] What is the most challenging part of the university classroom experience? What makes learning / achieving success challenging? What do you wish could be changed? 67

5.4 [Q1.3] How do you feel about the process of asking questions in class? What prevents you from asking questions in a traditional classroom setting? 68
5.5 [Q1.4] What makes a question a great question in your opinion? . . . 70
5.6 [Q1.5] Do you think asking questions improves your ability to learn? . 73
5.7 [Q1.6] Do you typically ask questions in class? 75
5.8 [Q1.7] Asking your question into a microphone in front of class 77
5.9 [Q1.8] Raising your hand in lecture 78
5.10 [Q1.9] Asking your questions on Microsoft Teams public chat 79
5.11 [Q1.10] Sending a DM to your TA / instructor 79
5.12 [Q1.11] Asking via anonymized form 80
5.13 [Q1.12] Which set was better (LLaMA vs Instructor) 82
5.14 [Q1.13] How much better was the set you chose? 85
5.15 [Q2.2] How did you find the way the group presentations were handled? 86
5.16 [Q2.3a] Why did you feel the group presentations were BETTER than other large classrooms? 88
5.17 [Q2.3b] Why did you feel the group presentations were ABOUT THE SAME as other large classrooms? 89
5.18 [Q2.3c] Why did you feel the group presentations were WORSE than other large classrooms? 91
5.19 [Q2.4] What are the biggest challenges you face in paying attention in class? What about in paying attention to the video presentations specifically? 93

5.20 [Q2.5] When you are having difficulty concentrating with other tasks at home or at school, what do you find helps you get back on track? . 94
5.21 [Q2.6] Which of the following would have helped you maintain focus? 96
5.22 [Q2.7] Did you watch any videos at home instead of in class? 98
5.23 [Q2.8] What prevented you from watching the videos beforehand? . . 100
5.24 [Q2.10] What is your preference for watching videos? 101
5.25 [Q2.11] Why is that your preference for watching videos? 103
5.26 [Q2.12] What of the following would you like to see in a future iteration of the way presentations and questions are handled in a large class setting? 104
6.1 User flow diagram 116
6.2 Student and instructor user interface for presentation mode. 118
6.3 Question generation flow 124
6.4 Engagement Settings Flow 127
6.5 Insights Flow 130
6.6 left student view, middle group view, right instructor view 132
6.7 Student flow for discussion mode 133
6.8 Discussion Mode: Group Table View 135
6.9 Discussion Mode: Instructor View 136
6.10 Instructor View. When a group indicates they need help a banner pop-up will appear 137

List of Tables

3.1 Code Reference Table A-G 38

3.2 Code Reference Table I-P 39

3.3 Code Reference Table Q-T 40

3.4 Kappa and Accuracy by Code for Question 1.1, excluding codes which were not used or resulted in perfect disagreement 45

5.1 Pre-Agreement Kappa and Accuracy Scores by Code for Questions 1.1 to 1.3, including codes which were identified at least once and did not have 100% disagreement within a question. 109

5.2 Pre-Agreement Kappa and Accuracy Scores by Code for Questions 1.4 to 2.3c, including codes identified at least once and did not have 100% disagreement within a question. 110

5.3 Pre-Agreement Kappa and Accuracy Scores by Code for Questions 2.4 to 2.11c, including codes identified at least once and did not have 100% disagreement within a question. 111

Abbreviations

A2L	Avenue to Learn – McMaster University's learning platform branded from Desire to Learn
AI	Artificial Intelligence
ChatGPT	Chat Generative Pre-trained Transformer – a language model interface developed by OpenAI
CS	Computer Science
CS 1JC3	Computer Science 1JC3 – Introduction to Computational Thinking, a first-year computer science course at McMaster University
DT	Design Thinking
GenAI	Generative Artificial Intelligence
GPT-3	Generative Pre-trained Transformer 3 – large language model released by OpenAI in 2020
GPT-4	Generative Pre-trained Transformer 4 – multimodal large language model released by OpenAI in March 2023

HASRL	Human-AI Shared Regulation
HCI	Human-Computer Interaction
LLM	Large Language Model
LLaMA	Large Language Model Meta AI – an open-source large language model released by Meta
ML	Machine Learning
MOOC	Massive Open Online Course
NLP	Natural Language Processing – a branch of artificial intelligence that enables computers to comprehend, generate, and manipulate human language
OpenAI	AI research and deployment company
Q&A	Question and Answer
RFA	Reflexive Thematic Analysis
SSRL	Socially Shared Regulation
TA	Teaching Assistant – a student hired to assist with course delivery
UI	User Interface
UX	User Experience

Chapter 1

Introduction and Background

For as long as classrooms have existed, the struggle to foster genuine student engagement has persisted as one of the most complex challenges faced by educators. Poor engagement drastically limits both the depth of learning and the quality of peer interaction, ultimately leading to negative outcomes for both students and teachers. This paper explores the process of designing software to improve engagement, particularly during student presentations, with the goal of improving material comprehension, quality of peer feedback, and minimizing friction and cognitive load for presenters.

As educational institutions increasingly turn to digital tools to enhance learning, there is a growing need for scalable technology that can support active participation. The rise of remote learning and hybrid models, exacerbated by the pandemic, has further highlighted the importance of tools that can facilitate interaction without increasing the burden on educators. This design study aims to address these issues by designing, testing, iterating, and proposing a technology-assisted solution to improve student interaction and feedback during presentations.

The motivation behind this research stems from the need to create more inclusive and engaging learning environments where students feel empowered to participate, ask questions, and interact with their peers despite the challenges posed by large class sizes. This work seeks to bridge the gap between traditional classroom methods and modern technologies, ensuring that students' voices are heard and their learning experiences are enriched.

1.1 Evolution of Classroom Dynamics

According to Chingos and Whitehurst [9], university classrooms have experienced substantial shifts over the past few decades in the student-to-instructor ratio. Factors including rising student enrollments and the push for economic efficiency have led to an increased prevalence of large classrooms, often accommodating hundreds of students. This shift brings with it several challenges, including, as Wang and Calvano [48] noted, diminished student participation, a reduction in personalized instruction, and, consequently, a less engaging learning environment overall.

Traditionally, classrooms have relied heavily on didactic teaching due to resource constraints, as it allowed for efficient dissemination of information to large numbers of students. Similar to how the Industrial Revolution introduced technologies that transformed manufacturing processes and increased productivity, modern technological advances, particularly in generative AI, present an opportunity to transform educational methodologies. These technological tools can enhance educational experiences by enabling more personalized and interactive learning without significantly increasing the workload for educators. This shift leverages the capabilities of AI to facilitate a more engaged and participatory learning environment, moving away from purely lecture-based formats to more dynamic, student-centred approaches.

1.1.1 Consequences of COVID-19 on Student Learning and Engagement

The COVID-19 pandemic was the largest disruption to education in history [42]. The abrupt and poorly resourced transition to remote learning significantly impacted student behaviour and well-being, reducing social and self-regulatory skills and increasing mental health challenges [15].

Engzell et al. [15] found that, during the shift to remote learning, students faced challenges related to self-motivation and time management, which persisted after returning to face-to-face instruction. The rise in social media usage has also contributed to these behavioural changes, with Paul et al. [29] finding that reduced time on social networks correlates with improved attention spans and academic performance. In large classroom environments, maintaining student engagement becomes increasingly challenging as individual attention is harder to monitor. Moreover, increases in social anxiety and depression [30] have further complicated students' re-adjustment to traditional classroom interactions, undermining group work efficacy (as found by Loades et al. [24]) and impacting emotional engagement, which is crucial for feelings of belonging and academic interest (as found by Delfino [12]). Emotional engagement, including students' feelings of belonging and interest in their studies, also impacts overall academic performance [12]. This combination of factors has led to a decreased likelihood of student participation [19, 15, 24, 30].

1.1.2 Significance of the Psychological and Social Dimensions of Student Engagement

The significance of these changes lies in the crucial role of psychological safety for learning and group dynamics. Psychological safety is a key factor in successful group dynamics, as highlighted by Google's Project Aristotle, which found that psychological safety is the most critical factor for team success [14]. This suggests that creating an environment where members feel safe to take risks and share openly is essential for optimal team performance. Additionally, dependability, structure and clarity, meaning, and impact are important components that contribute to healthy team dynamics.

Addressing student psychological barriers to engagement is essential, especially in large classrooms where students often feel intimidated. Creating an environment where students feel safe interacting tends to achieve better educational outcomes [55, 18]. Meeting students where they are and fostering socialization within the classroom is vital, as these experiences help students develop teamwork and social skills [12].

Improving student's social skills is an important goal of our education system. Classrooms serve as a crucial environment in fostering inclusion in various social spheres and broader cultural integration [40]. Educational environments act as pivotal social spaces where students learn to integrate into larger social and cultural contexts, essential for personal growth and professional development. The role of socialization extends beyond mere knowledge acquisition to encompass the development of social skills, teamwork abilities, and cultural adaptability, which are all vital for a well-rounded educational experience [2].

1.2 Developing Student-Centered Educational Tools

In this thesis, we will explore the design space for educational tools supporting peer interaction in large classrooms. We will focus on one technology (Generative AI), two design perspectives (data-driven versus value-driven design) and one design prescription (design thinking), and briefly define them to give context to our research questions.

1.2.1 Generative Artificial Intelligence

In response to these challenges in education, one avenue worth exploring is integrating Large Language Models (LLMs) and other generative artificial intelligence (GenAI) tools. LLMs like OpenAI's ChatGPT (Generative Pretrained Transformer) use transformer architectures to utilize self-attention mechanisms [45]. These models are trained on large datasets to predict the next word in a sentence, which, after extensive training, allows them to generate coherent and contextually appropriate text based on the input they receive. The recent developments in hardware capabilities and algorithmic efficiency have significantly reduced computational costs and improved training times, making these models more accessible and popular. This advancement, combined with their ability to generate human-like text, has led to their widespread application across various industries. At the time of this writing chatGPT is the fastest-growing consumer application in history [13], attracting over 100 million monthly active users within just two months of its launch. The influence of generative AI in the coming years cannot be overstated.

Tools like OpenAI's ChatGPT are being increasingly utilized in educational settings, demonstrating significant potential to transform how educational content is delivered and how students interact with it. These models can perform a variety of tasks that are beneficial in education, including generating educational content, assisting with administrative tasks, providing tutoring assistance, and even automating the creation of exams and quizzes. They also introduce risks, notably in facilitating plagiarism, inaccuracy with poor decision tracing, and encouraging a dependency that could undermine critical thinking and deep learning [13]. At present, we are at the start of a significant shift with a unique opportunity to direct its trajectory. Some experts suggest we may be living in the era of 'peak education' [17] as AI will disproportionately help those already educated and stunt the progress of novices who have a diminished incentive structure for educating themselves. As the presence of Generative AI grows, so does our need for strategic approaches to leverage its benefits while addressing potential drawbacks. The inability of our current institutional mechanisms to adequately detect AI-generated content exacerbates these challenges, as we cannot effectively screen submissions without potentially compromising student privacy and data security[6, 43].

A key challenge in academia lies in learning to effectively leverage AI tools, like LLMs, which, while powerful, require a nuanced understanding of how they operate. The prevailing skepticism surrounding AI often stems from concerns about the accuracy of AI-generated content and its potential to produce misleading or low-quality inputs. Much of this skepticism is linked to a lack of user expertise. Large Language Models are complex tools that require well-structured inputs—referred to as "prompts"—to generate meaningful and accurate responses. Although the technology itself has inherent limitations, it is important to recognize that vague or poorly designed prompts result in less useful or inaccurate outputs.

The issue is less about the inherent flaws in the technology and more about the user's ability to harness it effectively. This highlights the emerging importance of "prompt engineering," the skill of crafting precise and detailed inputs that elicit the best possible results from AI systems. As AI continues to integrate into academia, developing prompt engineering expertise will likely become a critical 21st-century skill, as it enhances educators' ability to search for relevant materials and improve productivity by engaging AI to generate refined outputs in less time. By improving our ability to ask the right questions, we can significantly enhance the quality and usefulness of AI.

It is also important to recognize that not every user needs to be an expert in prompt engineering to benefit from AI. Well-designed systems can abstract away complexity, allowing users to interact with AI tools without needing to understand or engage in the intricacies of the underlying algorithms. This allows for more intuitive and accessible interfaces where AI can assist in tasks like question generation, content summarization, or feedback collection without adding a cognitive load on the user. The focus then shifts from learning how to communicate with AI effectively to simply benefiting from the results, with the system handling the complexities in the background. This type of integration can streamline educational processes, offering support that is efficient, user-friendly, and capable of scaling across diverse learning environments.

The power of AI, particularly large language models (LLMs), lies largely in their ability to process and understand vast amounts of data – making them especially effective for search and filtering tasks. As these capabilities continue to advance, they are positioned to unlock opportunities we cannot forsee and fundamentally reshape how we interact with information and enhance decision-making across a wide range of fields.

1.2.2 Value-Driven and Data-Driven Design

Value-driven design is a holistic approach that embeds human values at the forefront of the technology development process. This method emphasizes the importance of aligning the product or service with the ethical standards, social norms, and overall well-being of its users and stakeholders [54]. This approach aims to ensure that technology supports the values held by its users and stakeholders. For instance, in the context of educational technology, value-driven design might focus on goals such as enhancing learning outcomes, promoting inclusivity, and maintaining academic integrity. The process is inherently qualitative, engaging with stakeholders via interviews, surveys, and participatory design sessions to gain a deep understanding of their values and needs, which then directly informs the development of the technology [37].

In contrast, data-driven design emphasizes the role of quantitative data in shaping design decisions. This approach relies heavily on metrics such as engagement, performance data, and usage patterns to guide the development process [47, 8]. Data-driven design is often used to optimize product features for maximum usability and efficiency, aiming to improve user experience based on empirical evidence. For example, a data-driven design in an e-commerce platform might focus on increasing conversion rates and time spent on the site by analyzing user interaction data and then tweaking the interface accordingly to achieve these metrics.

While data-driven design excels in precision and optimization, it may sometimes neglect the broader ethical and social implications of technology, focusing predominantly on numerical metrics [28]. Conversely, value-driven design's strength lies in its ethical grounding and focus on the user, though it might struggle with quantifying success and operational efficiency. This paper integrates value-driven principles with data-driven insights. Doing so allows technologists to develop technologies that not only achieve high performance but also resonate deeply with user values and societal norms [46].

1.2.3 Design Thinking

Design thinking has emerged as a robust methodology for addressing complex problems in various domains, including education. Rooted in a user-centric approach, this methodology has five stages as seen in Figure 1.1. Empathize, where designers understand users' needs through research; define, which focuses on articulating the core problems; ideate, involving brainstorming potential solutions; prototype, where tangible versions of solutions are created; and test, where these prototypes are evaluated and refined based on user feedback.

In the context of educational technology, design thinking proves particularly beneficial as it involves active collaboration with both stakeholders (students and educators) to identify and address their specific needs. By incorporating iterative feedback loops, this approach ensures that the resulting tools are not only functional but also enhance the overall learning experience. Building the right product along with building the product right. The application of design thinking to develop interactive classroom tools can lead to more effective and engaging educational environments, promoting better student participation and engagement through thoughtfully designed

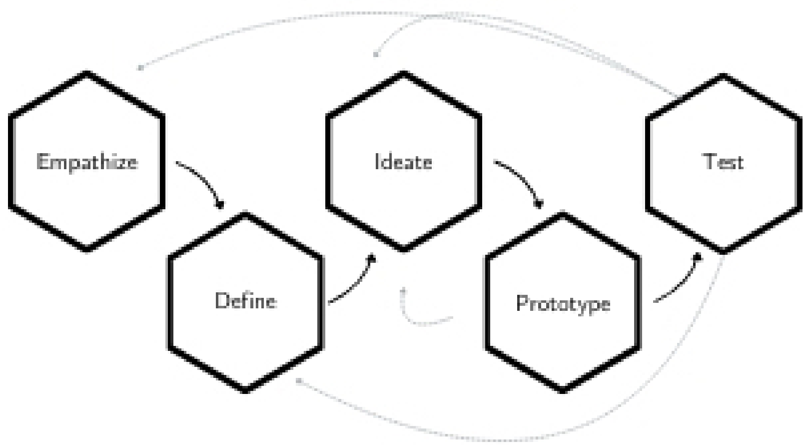


Figure 1.1: Stages of the Design Thinking Process

solutions that address real-world challenges.

1.3 Research Questions

RQ1 How can educational software be designed, using design thinking principles, to enhance student engagement while minimizing additional workload for instructors?

RQ2 How can AI be utilized in educational software to replicate or improve upon tasks traditionally performed by instructors, such as question selection, and how does its performance compare to that of experienced educators?

RQ3 What are the most effective strategies for using real-time feedback and question filtering systems to improve group dynamics and interaction during student presentations?

1.4 Contributions

The primary contribution of this work is the design of a tool to enhance student engagement in large classroom settings, specifically during student-led presentations. This software prototype, titled Interaction.ai (See Chapter 6), was developed through an interactive process and informed by qualitative and quantitative feedback from both students and instructors during the initial prototype testing.

One key element of this study is the comparative analysis of the effectiveness of Large Language Models (LLMs) and instructors in selecting relevant and insightful questions during student presentations. The results demonstrate that AI-driven question selection was not only as effective as instructor-selected questions but also more efficient, reducing the logistical burden on educators while maintaining high levels of engagement.

This thesis also presents an example of how the design thinking process can be adapted to the development of educational tools. In contrast to the traditional rapid prototyping cycles often seen in design thinking, the educational context required a more deliberate approach due to the constraints of the academic calendar. This study placed particular emphasis on the empathy phase, as understanding the needs and challenges of both students and instructors was crucial in shaping the tool's development. Feedback from one deployment of the course was used to inform changes in the subsequent iteration, highlighting the value of a user-centred approach even within the limitations of longer development cycles.

Additionally, this research provides a detailed exploration of student expectations and barriers to participation in large classroom settings. By analyzing both internal and external factors—such as anxiety, group dynamics, and social pressures—the study uncovers insights into what prevents students from asking questions and engaging actively with their peers. These findings informed the design of a tool that leverages technology to facilitate interaction, reduce cognitive and emotional barriers, and ultimately foster deeper engagement in the learning process.

Chapter 2

Literature Review

2.1 Research Methods

2.1.1 Design Thinking

Design thinking, a methodology originating in the mid-20th century, has its roots in the disciplines of architecture and industrial design. The Bauhaus movement, emerging in the early 1900s, significantly influenced design thinking by promoting the integration of art, craft, and technology [50]. Simon [39] originally conceptualized design as a systematic process in his seminal work *The Sciences of the Artificial*, which framed design thinking as a problem-solving approach distinct from scientific inquiry. Simon [39] was the first to mention design as a way of thinking and contributed many ideas throughout the 1970s that are now regarded as principles of design thinking. His work focused on making design scientific and explored whether human forms of thinking could be replicated by machines—an area highly relevant in today's design landscape [11].

Simon [39] emphasized the generation of satisfactory solutions quickly, rather than in-depth problem analysis to produce an optimal solution. This approach, described as 'satisficing' rather than optimizing, involves producing one of many possible satisfactory solutions instead of attempting to find a hypothetically optimal one. Cross [10] also observed this approach in other studies of design behaviour, including architects, urban designers, and engineers. The work of Cross [10] emphasizes the value of design as a discipline and its methodologies, and has contributed significantly to the academic foundation of design thinking.

Initially, design thinking aimed to improve the effectiveness and creativity of design processes in physical product development. Over time, its scope expanded, finding applications in diverse fields such as business, education, software, and healthcare. The establishment of institutions like Stanford University's d.school and the influence of design firms such as IDEO played crucial roles in popularizing design thinking beyond traditional design disciplines [21]. Design thinking also addresses the fragmentation of scientific disciplines into specialized fields, proposing a holistic approach to integrating these areas to solve new, complex problems in socio-technical systems [11].

In the realm of software development, design thinking has become a widely adopted approach due to its user-centered and iterative nature. In his Harvard Business Review article, Brown [7] highlights the importance of collaboration across disciplines and encourages organizations to adopt a design thinking mindset to foster innovation and create solutions that are both functional and emotionally resonant. It involves understanding the end-user's needs through empathy, defining clear problem statements, ideating multiple solutions, prototyping, and testing and is a way to encourage interdisciplinary collaboration. This methodology aligns well with agile software development practices, fostering innovation and ensuring that software solutions are both functional and user-friendly.

In the book *Creating with Code: An Introduction to Functional Programming, User Interaction, and Design Thinking*, Anand et al. [1] further illustrates how design thinking can be effectively integrated into software development. The authors emphasize the importance of empathy to understand user needs, the definition of clear problem statements, and iterative prototyping to refine solutions. They highlight practical exercises and real-world examples to demonstrate the application of design thinking principles in creating intuitive and user-friendly software.

Empathy is a cornerstone of the Design Thinking process, in which designers engage with users to understand their experiences, needs, and challenges. This understanding serves as the foundation for building solutions that are genuinely user-centered. In the context of educational software, where diverse student needs and learning environments must be accommodated, empathy allows designers to grasp the nuances of both student and instructor experiences, ensuring that the developed tools are relevant, inclusive, and effective.

According to Brown [7], empathy is about seeing the world through the eyes of the user, gaining insights that might not be immediately apparent through quantitative data alone. In education, this can involve understanding the pressures students face during presentations, the challenges they experience with engagement in large classrooms, and the frustrations instructors encounter when managing student participation.

Empathy also facilitates iterative improvement. As Brown [7] notes, by continually interacting with users throughout the design process, designers can refine their prototypes based on real-time feedback. This iterative process, driven by empathy, ensures that the software evolves in a way that remains aligned with user needs. Both cognitive and emotional factors impact user experience, empathy ensures that the developed tools support not just functional learning outcomes but also the emotional well-being of the users.

The benefits of incorporating design thinking in software development are manifold. It promotes a deep understanding of user requirements, leading to more intuitive and effective software products. Additionally, the iterative prototyping and testing phases enable rapid identification and resolution of design flaws, reducing the risk of costly errors later in the development process. By encouraging collaboration and creativity, design thinking helps software development teams generate innovative solutions that better meet user needs and market demands [7, 21].

2.1.2 Thematic Analysis

Thematic analysis has its roots in the broader tradition of qualitative research methodologies. Braun and Clarke [4] formally conceptualized thematic analysis as a distinct method in their seminal paper, which laid out a clear guide for using it. This work was pivotal in standardizing the approach and making it accessible to researchers across various disciplines.

Braun and Clarke [5]'s method involves identifying, analyzing, and reporting patterns (themes) within data. It is a flexible research tool for turning qualitative data into rich patterns. **Reflexive thematic analysis**, in particular, emphasizes the researcher's role in constructing themes—not just in finding them [5]. It allows researchers to interpret nuanced data through an iterative process that includes coding, generating themes, reviewing themes, defining and naming themes, and producing a report.

Several studies have successfully employed reflexive thematic analysis in contexts similar to this, emphasizing the usability and effectiveness of educational technologies. For instance, in Varanasi et al. [44]'s paper, thematic analysis was used to explore participants' emotional and cognitive responses when interacting with digital platforms. The insights gained were instrumental in designing interfaces sensitive to the user's socio-economic and emotional contexts, thus enhancing user engagement and satisfaction.

In his paper on how students perceive ChatGPT in educational settings Shoufan [38] utilized thematic analysis to uncover themes from student feedback and followed up with a survey to quantify these insights. This method allowed him to identify and address key areas of student concerns and interests regarding ChatGPT's application in learning environments. This method enriched the understanding of user requirements and experiences as well as ensuring that the tool development is grounded in actual user needs and preferences, facilitating a user-centred approach in educational technology design.

2.2 Using Dashboards to Improve Educational Outcomes

In the context of educational dashboards, Tissenbaum et al. [41]'s exploration of real-time visualization of student activities represents an important development in supporting classroom orchestration through technology. Their paper emphasizes how technology-enhanced environments can offer instructors immediate insights into student engagement and progress, enabling better decision-making during class. Unlike traditional dashboards, which focus on post-hoc reflections, the tools discussed in this research present live data, giving teachers real-time control without overwhelming them.

These dashboards are designed to reduce the "orchestrational load" (i.e., the cognitive burden) on educators by visualizing data in ways that are easily interpretable and actionable. This is particularly valuable in inquiry-based, open-ended learning environments, where students progress at different paces, making it difficult for teachers to gauge the needs of the class in real time.

The paper also points to critical questions about what data should be captured and how it should be visualized to support meaningful interventions. By creating real-time dashboards, teachers can ensure students remain on track while fostering collaboration and participation. The study underlines how such technologies must integrate seamlessly into a teacher's workflow, providing valuable insights without requiring excessive manual intervention.

The inclusion of these orchestrable technologies into the classroom illustrates the growing importance of harnessing telemetry data—not just for reflection, but for making informed decisions in real-time. This research is crucial to understanding how technology can assist educators in dynamically managing classroom activities and is a valuable reference point for other educational technology designs seeking to improve interaction, feedback, and engagement.

2.3 Large Language Model Ranking

In Webb et al. [49]'s paper "Emergent Analogical Reasoning in Large Language Models" the authors explore the capacity of Large Language Models (LLMs) like GPT-3 to perform zero-shot analogical reasoning tasks, a hallmark of human cognitive ability. Their research focuses on comparing LLM performance with human participants across several analogy-based reasoning tasks, including non-visual, text-based matrices.

The study finds that GPT-3 matched or surpassed human performance, demonstrating strong reasoning abilities with little direct task-specific training. This shows that LLMs have emergent reasoning capabilities, raising questions about their potential future use in diverse fields, including ranking tasks.

Analogical reasoning plays a central role in identifying patterns in diverse datasets, making it highly relevant to AI applications in education. Webb et al. [49]'s findings suggest that AI systems can make meaningful, reasoned selections without direct training on specific tasks. This capability not only highlights the future possibilities of LLMs in supporting instructional tasks but also underscores their ability to mimic human reasoning in structured decision-making scenarios.

2.4 Generative AI in Education

Generative AI presents both opportunities and challenges for education, particularly in higher education settings. Large Language Models (LLMs) such as GPT-4 are increasingly integrated into teaching and assessment. Generative AI in education primarily refers to applications powered by large language models (LLMs), such as GPT-4. These LLMs are machine learning algorithms designed to process vast datasets, generating predictions and responses based on user prompts. Although generative AI also includes models for images and videos, LLMs are central to the challenges and opportunities in tertiary education today.

Lodge et al. [25] explores the rapidly evolving landscape of generative AI and its implications for higher education through the creation of a research agenda. The authors emphasize that, while AI in education has been researched for decades, the unexpected emergence of tools like ChatGPT has outpaced prior predictions. This sudden advancement has left educational institutions grappling with the implications, requiring a strategic research agenda to address the changes. And, in their opinion, previous scholarship is not helpful: "It is evident that much of the scholarly discussion and debate from as recently as [2022] is now sorely outdated given the emergence of tools such as ChatGPT." The paper calls for an in-depth examination of how generative AI reshapes learning, assessment, and policy within tertiary education.

Topics of interest in Lodge et al. [25]'s work span assessment [53], learning and teaching through and with AI [22], and the technical and ethical aspects of AI relevant to education [27].

Zawacki-Richter et al. [53] conducted a systematic review of research conducted between 2007 and 2018 regarding AI in higher education. Despite AI's longstanding presence in education research, the paper highlights that educators still face challenges in understanding how to meaningfully integrate AI into pedagogical practices. The review aims to clarify the existing landscape of AI applications, focusing on the intersection between educational technology and AI while questioning where educators fit into the conversation.

The review shows that most research comes from STEM fields, particularly computer science, with quantitative methods being the most frequently used in empirical studies. One of the critical reflections of the paper is the lack of integration between AI research and pedagogical theories, as well as the minimal discussion around ethical challenges and risks associated with AI in educational settings.

Zawacki-Richter et al. [53] also emphasize the potential risks of AI in higher education, such as concerns around data privacy, the replacement of human educators, and the ethical implications of relying heavily on automated systems for teaching and assessment. The paper underscores the need for future research that more closely aligns AI's capabilities with pedagogical goals while addressing the ethical considerations of AI-enhanced learning environments. It calls for educators to play a more active role in shaping the development and deployment of AI in education to ensure it is used in a way that supports both learning and teaching effectively.

The widespread availability of generative AI has drastically reduced the effort and risk associated with cheating while making detection significantly harder, if not impossible. This shift presents a critical challenge for educators, as traditional methods of assessment are being undermined. However, this situation also offers a unique opportunity to rethink the purpose of education. It invites educators to reconsider the nature of learning, how knowledge is acquired, and how to best prepare students to use generative AI tools responsibly in both academic and professional settings. Instead of focusing solely on detection, educators can explore ways to integrate these tools into curricula, fostering critical thinking and ethical usage that align with the demands of modern work and life. For one thing, AI is likely to be a core tool in their toolkit during their careers.

In their work, Kuka et al. [22] explore the transformative impact of artificial intelligence (AI) in higher education. Through a scoping literature review, the authors synthesize key AI-related technologies such as machine learning, learning analytics, and data mining, while highlighting the pervasive yet often invisible role these technologies play in enhancing teaching and learning. AI's applications range from automation and predictive analytics to personalized learning and intelligent tutoring systems, all contributing to improved educational experiences. However, the chapter also stresses the growing importance of 21st-century skills—such as critical thinking, collaboration, and computational literacy—for both students and educators, as they adapt to this digital transformation.

Kuka et al. [22] underscore how AI can seamlessly operate behind the scenes, offering recommendations or automating routine tasks, but they also highlight the potential risks associated with its integration. The paper concludes by asserting that higher education institutions must adapt their curricula to prepare students for a rapidly changing, AI-driven labor market. This involves equipping students and educators with digital literacy skills that encompass not just technical proficiency but also critical thinking and ethical decision-making in the context of AI technologies.

In "Human and Artificial Intelligence Collaboration for Socially Shared Regulation in Learning," Järvelä et al. [20] explore the potential for combining human and AI capabilities to improve learning regulation, specifically in socially shared regulation (SSRL). They introduce a hybrid model of human-AI shared regulation (HASRL) that emphasizes the strengths of both human and AI systems in managing cognitive, emotional, and social aspects of learning. The study argues for using AI to assist in routine tasks, freeing humans to focus on more complex and adaptive aspects of learning, like goal-setting and collaboration.

Järvelä et al. [20] highlight the potential for AI in learning environments to enhance student self-regulation by providing real-time insights and fostering collaboration. By examining how human-AI partnerships could support socially shared regulation, they propose using AI in tandem with human expertise to create more effective educational interventions. This paper suggests that combining AI's efficiency with human creativity could unlock new opportunities for educational research, especially in terms of developing tools that support collaborative learning and deeper regulatory processes in both short-term and long-term learning cycles.

This work underscores the importance of hybrid intelligence—leveraging both human and AI strengths—to create learning systems that are both effective and ethically sound. The integration of multimodal data collection and learning analytics tools shows the promise of AI in advancing the learning sciences and facilitating more meaningful interactions between students and teachers.

Generative AI, particularly large language models (LLMs) like GPT-4, are becoming increasingly integrated into educational environments. While these models have the potential to enhance learning, students must be made aware of their limitations, such as generating misleading or incorrect responses. By critically evaluating AI outputs, students can develop essential skills that will serve them in academic contexts and their future careers.

2.4.1 Educational Software to Improve Participation

Fan et al. [16] developed a tool called CourseMIRROR, a mobile learning platform designed to enhance instructor-student interactions in large classrooms. Using NLP techniques, CourseMIRROR streamlines the process of collecting in-situ reflections from students after each lecture. The system provides immediate feedback on the quality of these reflections, scaffolding students to compose more specific, higher-quality responses, which are then summarized and presented to both students and instructors. This process helps instructors identify areas of confusion and adapt their teaching strategies accordingly.

By facilitating timely submission of reflections, CourseMIRROR directly addressed participation challenges in large classrooms. The feedback loop not only scaffolded students' reflection-writing process but also increased their engagement by ensuring that their reflections were pedagogically valuable. The system's ability to summarize student reflections and present significant insights to both instructors and students enhanced communication and understanding of learning progress. This real-time interaction encouraged active involvement from students, thereby improving the overall learning environment.

This work is also notable for its scalability. Fan et al. [16] demonstrated the system's robustness across various STEM subjects and its ability to handle large classroom environments with minimal disruptions. By automating much of the reflection and feedback cycle, the tool reduced the manual workload for instructors and helped them tailor their teaching strategies to address the specific areas where students struggled. CourseMIRROR thus serves as a model for how AI and NLP can be integrated into educational software to promote engagement, reduce friction, and streamline student-instructor interaction in large classroom settings.

Sarsa et al. [35] explore the potential of OpenAI Codex in generating programming exercises and code explanations for introductory programming courses. By leveraging natural language generation, Codex creates programming problems, sample solutions, and code explanations that can be directly integrated into courses. The study found that most of the generated content was novel and relevant, but it still required some human oversight to ensure accuracy, especially with corner cases and faulty test cases.

The work highlights the benefits of using AI-driven tools like Codex for scaling the creation of exercises in large courses, where manually designing diverse and high-quality programming exercises is a significant challenge. Codex's ability to generate programming problems, solutions, and accompanying explanations saves instructors time and effort, allowing them to focus more on other pedagogical tasks. While some generated outputs may need minor adjustments, the results suggest that AI tools can be valuable in streamlining the course design process, creating a scalable and efficient way to address the increasing demand for personalized and active learning experiences.

The use of generative AI models, like Sarsa et al. [35]'s work with Codex, provides an automated mechanism for scaling educational resources without sacrificing quality, supporting the broader objective of leveraging AI to improve classroom interaction and participation. By reducing instructors' workload and providing immediate, scalable feedback to students, tools like Codex could play a key role in fostering active learning and engagement, much like the systems discussed throughout my research.

Similarly, the work of Rai et al. [31] explores how generative AI tools, such as OpenAI's GPT models, can be used to generate assessments for Massive Open Online Courses (MOOCs). The authors emphasize the potential of AI to quickly produce various types of assessments, such as quizzes and self-evaluation exercises, which can be useful for both educators and learners. By experimenting with prompts related to a foreign language course, the study shows how educators can create assessments efficiently using AI tools, though the need for oversight and content verification is stressed.

One of the major findings of the paper is that with careful prompting, generative AI tools can create relevant, scalable assessments that suit the diverse needs of MOOCs. Given the constraints of traditional methods, AI enables educators to generate assessments for large online classrooms more rapidly. However, the paper also points out potential limitations, including concerns over the creativity of instructors and the need for verifying AI-generated content to avoid errors or hallucinations—incorrect outputs provided by the AI.

Rai et al. [31]'s work highlights the balance between the benefits and challenges of using generative AI in education, specifically in assessment creation. While generative AI tools can streamline assessment design, some educators and institutions may resist their adoption due to concerns about academic integrity and the loss of instructor creativity. The paper argues that, despite these challenges, generative AI is a powerful complement to traditional assessment methods, particularly in the context of online learning environments like MOOCs, where scalability and efficiency are critical. Just as MOOCs benefit from AI in assessments, classroom settings can similarly use AI to streamline participation management and encourage more interactive learning experiences.

Chapter 3

Methodology

This chapter outlines the methods and procedures used in this study to enhance student engagement in a large university classroom setting. Conducted during the Fall 2023 semester in a first-year computer science course at McMaster University with 197 students, the study explores the effectiveness of various technological interventions to encourage interaction and participation, particularly during group presentations.

This chapter discusses the methods employed, including an overview of the study's phases and the qualitative and quantitative research techniques utilized, such as user interviews, surveys, and feedback loops with both students and instructional staff. Additionally, it addresses the application of design thinking principles, reflexive thematic analysis, and assessment techniques like confusion matrices and Kappa scores, which were used to evaluate inter-rater agreement before and after consensus. The chapter will also explore the rationale behind these methodological choices and their implications for educational technology research.

3.1 Research Design Overview

3.1.1 Course Context

This study was conducted during the fall semester of 2023 in a first-year computer science course at McMaster University, with 197 enrolled students. The study's primary objective was to improve student engagement, particularly in the context of group presentations. The course structure required students to select their presentation groups and topics, which spanned a wide range of core computer science concepts and their applications to society, business, and science.

Students chose topics such as emerging technologies like generative AI and cryptocurrency, as well as the societal impacts of innovations like social networks. These topics were supported by accessible source materials; for example, probabilistic graphical models were introduced using "The Book of Why" by Judea Pearl. While this diversity of topics provided broad exposure to relevant issues, it also posed a risk of reducing student engagement.

Each group was required to deliver a pre-recorded presentation followed by a live Q&A session. Presentations could be watched in advance or during class sessions, and active participation, primarily through asking questions, was integrated into the course's grading scheme. This approach encouraged students to engage with each other's presentations, but it also resulted in a high volume of questions per group, with some groups receiving up to 168 questions. Managing and filtering these questions within a limited timeframe (typically less than 10 minutes) presented a significant logistical challenge.

3.1.2 Problem Identification

This research was catalyzed by the 1JC3 professor's concerns regarding low levels of engagement and participation in the classroom, specifically in the context of group presentations. Recognizing these challenges, the professor sought to explore potential solutions that could enhance student interaction and streamline the process of question collection and handling. This situation underscored the need for a systematic study to address these critical educational barriers.

3.1.3 Study Design

One key element of the study was the evaluation of AI-curated questions versus instructor-curated questions to improve student interaction during group presentations and Q&A sessions. Traditional methods of question selection, such as opting for the most recent or shortest submissions, were insufficient for ensuring fairness or relevance. Alternative methods, including peer-upvoted questions or random selection, were considered but did not fully address the problem. Thus, the study introduced two strategies for filtering questions: 1) professor-filtered, non-sophisticated methods, and 2) advanced, AI-based filtering using LLaMA, an open-source large language model by Meta. LLaMA was used locally on a secure desktop to process and filter the questions, offering a novel solution that balanced effectiveness with concerns about privacy and data control.

The iterative nature of the prototype design allowed for continuous refinement of the methods used to enhance engagement, with insights gained from this process informing the development of future prototypes. Chapter 5 of this thesis explores these findings in greater detail.

The ultimate goal of this study was twofold: first, to develop an application that encourages students to actively engage with their peers' presentations and enhance their understanding and retention of the material. Second, the study aimed to create a tool that simplifies engagement, making participation feel more effortless for both students and instructors (or audience and presenter). By integrating traditional and AI-driven methods, the study sought to address logistical challenges while fostering a systematic and equitable approach to managing student interactions and participation.

3.1.4 Design Thinking

Design thinking is an iterative, human-centred approach to problem-solving that emphasizes understanding users' needs, generating a wide range of creative ideas, and continuously refining solutions. This methodology unfolds across five stages: empathy, define, ideate, prototype, and test. Each stage is interconnected, with feedback loops ensuring the developed solution remains aligned with user requirements.

The process begins with the **empathy** phase, which is the foundation of design thinking and where this research placed significant emphasis. The goal of this stage is to deeply understand the problems students and instructors face, not only through direct interaction but also by observing and analyzing their behaviours in educational settings. In this study, empathy was operationalized through interviews, surveys, and feedback loops with both students and instructional staff. These activities aimed to identify pain points in classroom engagement, such as low participation during presentations and ineffective question management during group discussions. Engaging users early in the process ensured that the prototype addressed authentic problems rather than assumptions about what users might need.

Following this, the **define** stage synthesizes the gathered information to craft a clear and actionable problem statement. It refines the understanding of user needs and challenges, focusing the project on human-centred solutions.

The **ideation** stage then encourages the generation of a broad spectrum of ideas, prioritizing quantity over quality to maximize creativity and reduce the likelihood of overlooking unconventional solutions. This emphasis on volume encourages a diverse range of perspectives and fosters innovative thinking. In this research, ideation was operationalized through student questionnaires (particularly 5.26). Discussions with course staff also explored potential feature options, allowing for a collaborative brainstorming process that ensured both students' and instructors' perspectives were considered in the development of new functionalities.

The fourth stage, **prototyping**, converts the best ideas into scaled-down, workable models. This step focuses on refining and converging the wide range of potential solutions. These prototypes, whether simple physical forms or interactive digital models, allow for iterative testing and refinement based on real-world application and usability. In this research, an initial prototype was developed (See Chapter 4), and the second is explored in Chapter 6.

Finally, the **test** stage involves evaluating prototypes in real-world settings typically by actual users. This phase is crucial for obtaining feedback and assessing the effectiveness of the solutions. Insights from this stage can lead to further refinements, or even revisiting earlier stages to perfect the approach. The findings from this testing phase, including key insights and challenges, are explored in detail in Chapter 5, offering a clear view of how initial user interactions shaped the next prototype.

In the context of educational technology, design thinking offers a structured yet flexible framework for creating innovative solutions that enhance user experience—in this case, improving student engagement and participation in large classroom settings.

The decision to employ design thinking in this research stemmed from its strong emphasis on user-centric design, making it particularly appropriate for the development of educational software. Unlike more traditional approaches that may prioritize technological capabilities, design thinking starts by focusing on the real-world experiences of users, ensuring that the technology developed serves their actual needs. This approach allows the development process to remain grounded in the users' experiences—students and instructors alike.

Empathy was not just confined to the early stages of this research but remained an ongoing component. The iterative nature of the design thinking process meant that student and instructor feedback was continually sought, with their insights directly influencing each phase of the prototype's development. For instance, during initial testing, students highlighted difficulties with asking questions, which led to the refinement of features aimed at improving question management and reducing social anxiety. These insights were crucial in ensuring that the software remained relevant, practical, and user-friendly throughout its evolution.

By focusing on empathy, the research also recognized the cognitive and emotional dimensions of learning. Understanding how students felt when engaging with the software—whether it reduced anxiety or facilitated more active participation—was just as important as the technical functionality. Instructors, too, provided valuable feedback on how the tool could ease their workload while promoting more equitable student participation.

Ultimately, design thinking was chosen because it ensures that technology does not exist in isolation from its users. Its emphasis on empathy ensured that the software developed in this research was tailored to meet the specific needs of students and instructors, making it a more effective tool for improving classroom engagement.

3.2 Feedback Collection

3.2.1 Instructional Staff Interviews

As part of the iterative design process, a series of three structured prototype review sessions were conducted with instructional staff to gather feedback on both the course structure and the prototype itself. The first session focused on identifying challenges within the course that could be addressed through a software-based solution, setting the foundation for initial design decisions. The second session concentrated on evaluating the functionality and impact of the initial prototype, with staff reflecting on how its features compared to their prior experiences teaching this or similar courses. It also discussed potential future improvements and new features that could be incorporated into the final prototype. During this session, feedback centred on observed classroom challenges that persisted despite the initial prototype and how additional features could address those issues.

The final session was a review session for the proposed prototype where staff had the opportunity to give feedback on the final design. Instructional staff provided insights into how the revised design might alleviate these challenges and anticipated how the proposed changes could influence student engagement and classroom dynamics. These review sessions played a crucial role in refining the prototype, ensuring that it was aligned with the practical needs of educators while addressing key pedagogical challenges.

The staff interviews were conducted via Microsoft Teams, allowing for flexible scheduling while ensuring the conversations remained secure. These semi-structured interviews began with broad questions about the course's progress, the instructors' overall experiences, observations, and any challenges they encountered. From there, the discussions shifted to specific areas such as the effectiveness of the current teaching tools, the level of student engagement, and the dynamics of classroom interactions, all informed by insights gained during the initial, broader phase of the interview. Other key discussion points included the impact of remote learning on teaching methods, student participation, the use of supplementary materials, and any difficulties with the existing anonymous question-asking system.

The primary focus was to gain a clear understanding of the firsthand experiences of the instructional staff with the initial prototype. This approach allowed us to identify recurring themes and specific pain points to ensure that the new software would address the real needs and concerns of the instructional staff.

3.2.2 Student Questionnaires

Two surveys were administered throughout the course to collect data to inform the development of more effective engagement strategies. The participation rate was favourable with 163/197 students participating in at least one of the surveys.

The first questionnaire aimed to assess the intrinsic and extrinsic factors influencing student participation, particularly in the context of asking questions in large classroom settings. It consisted of 14 questions: one true or false, eight multiple-choice, and five short-answer questions. The second survey was conducted after the presentations were completed. It aimed to evaluate the initial prototype, investigate the flipped classroom model, measure the impact of videos with automatically generated transcripts on accessibility and course structure modifications, and gather feature suggestions. It also consisted of 14 questions: seven multiple-choice and seven short-answer questions. A copy of the questionnaires can be found in the appendix 7.3.

The findings of these questionnaires are explored in detail in Chapter 5.

3.3 Qualitative Coding and Reflexive Thematic Analysis

The size of the course and the number of short answer questions presented a large amount of qualitative data to sift through. 1,182 written responses totalling 35,418 words and 205,889 characters were anonymized and systematically analyzed multiple times through reflexive thematic analysis[5]. The final analysis resulted in 2,161 codes applied to the written answer responses signifying an average of approximately 1.8 codes per response. Reflexive thematic analysis, in particular, emphasizes the researcher's role in constructing themes – not just in finding them[5]. In our context, it was used to transform qualitative questionnaire responses into quantifiable insights that inform the design and functionality of an educational tool. By analyzing the qualitative data gathered from surveys and interviews, the research could systematically categorize data into thematic clusters. This approach allowed for the identification of recurring ideas or sentiments that could influence design decisions, thus bridging qualitative data with quantitative needs for software development.

3.3.1 Thematic Codes

A set of 39 codes were developed to categorize and analyze the qualitative data from student responses. These codes and their corresponding descriptions are detailed in Tables 3.1, 3.2 and 3.3. Each code was defined and chosen to reflect significant themes relevant to student engagement, the dynamics of question-asking, and classroom interaction patterns. The process involved an initial reading of the responses to identify recurring themes, followed by a discussion with a committee to refine and finalize the codes. This approach ensured that the codes were both representative of the data and useful for highlighting specific aspects of student experiences and behaviours.

Each code was carefully defined to capture significant themes that emerged from the data, particularly those related to student engagement, question-asking behaviours, and patterns of classroom interaction. The development of these codes followed a structured process, starting with an initial reading of the responses to identify recurring themes. This was followed by several iterations of discussion and refinement with a committee to ensure that the codes were grounded in the data and tailored to highlight the most relevant aspects of student experiences.

The coding process involved two independent raters who each applied the codes to the data in a blind coding exercise, where neither rater had access to the other's responses. This method was chosen to ensure objectivity and minimize bias. After completing the initial round of coding, the two raters convened to resolve any disagreement between their responses. Cohen's Kappa, discussed further in Section

Table 3.1: Code Reference Table A-G

Code	**Description**
ACCOUNT	Students who mention an issue with accountability to their group
ANON	Students who positively talked about the anonymous nature of question asking
ANX	Students who talked about nervousness / social anxiety
BREAK	Students who noted they wanted to take breaks
CAPT	Students who mentioned captions / transcript
CODE	Students who wanted more help on the coding portion of the presentation (from peers, examples, or more lecture time explaining)
COMMON	Referring to questions that other students have as well (relatable) or that other students would want to know
CONCISE	Referring to questions that are short and to the point / well worded
DEBAT	Students who wanted a debate between groups
DEEPEN	Referring to questions that deepen the understanding or clear uncertainty of a topic, taking new perspectives or enhancing critical thinking
ELMpos	Students who talked positively about the Elm programming language used in the course
ELMneg	Students who complained or spoke negatively about Elm
ENGL	Students who mention challenges with English
FOCUS	Students who had difficulty with distractions in class or maintaining focus during the presentations
GPT	Answers that are suspected to be written by ChatGPT or a generative AI

Table 3.2: Code Reference Table I-P

Code	**Description**
INTERpos	Students who talked positively about interacting with other students
INTERneg	Students who had a negative experience with other students
IRL	Students who mentioned the importance of in-person learning
LENdec	Students who noted the presentations were too long
LENinc	Students who noted the presentations were too short
LIVE	Students who preferred to see a live presentation instead of prerecorded
NOISE	Students who complained about the noise level in the classroom or about the audio quality
ONL	Students who preferred the prerecorded presentations or mentioned positive aspects of hybrid or virtual learning
OUTCOME	Students' expected outcomes
PACE	Students who mentioned the pace of the course (i.e. the fast pace of the course)
PERS	Students who mentioned the challenges of different learning styles / wanting personalized learning
PREREQ	Students who talked about not feeling properly prepared for the course / not beginner friendly
PROCESS	Students expectations for the process of the course
PROF	Students who talked about the importance of the role of the professor

Table 3.3: Code Reference Table Q-T

Code	**Description**
QAinc	Students who wanted more time for the Q&A period
QAmov	Students who wanted the time of the Q&A period moved (in some cases immediately after the presentations)
RELEVANT	Referring to questions that are contextualized to the topic discussed or have applications in the real world
RESOURCE	Students who talked about a lack of resources (time with prof / TA, online resources, etc.)
ROUNDpos	Students who liked sitting at the round tables in the active-learning classroom or mentioned the ease of socialization due to the classroom set-up
ROUNDneg	Students who didn't like sitting in the round tables for some aspect of the course
RUBR	Students who complained about the rubric / structure / format of the course
SIZE	Students who complained about the size of the classroom
TIMEMGMT	Students who talked about time management
TIME	Students who indicated a temporal element to the (recorded) presentations (pausing, rewinding, replaying, etc)

3.4.1, was calculated to assess the inter-rater reliability and provide a statistical measure of agreement between the two coders. As expected, some disagreements arose, often due to minor oversights or coder fatigue, which the Kappa score reflected.

Through a discussion, the raters were able to resolve all differences, often uncovering cases where rare codes were being interpreted differently. This dialogue not only improved the consistency of the coding but also provided further insights into the nuances of certain themes, allowing for a more precise application of the codes. Ultimately, this process resulted in a perfect inter-rater agreement, ensuring the reliability and robustness of the thematic analysis.

3.4 Quantitative Analysis

3.4.1 Cohen's Kappa

To quantitatively analyze the qualitative data, the responses were clustered into thematic codes, following the Reflexive Thematic Analysis approach 3.3. To assess the reliability of the coding process and measure the agreement between two independent raters, we used Cohen's Kappa coefficient (κ), a robust statistical measure that accounts for the probability of chance agreement [23]. Unlike simple percent agreement, Cohen's Kappa adjusts for the likelihood that coders might agree purely by chance, offering a more accurate evaluation of inter-rater reliability .

Cohen's Kappa (κ) is defined by the following formula:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where p_o is the observed agreement proportion, and p_e is the hypothetical probability of chance agreement. For binary decision tasks, which was used in this research, Cohen's Kappa is calculated by the following simplified formula:

$$\kappa = \frac{2 \times (TP \times TN - FP \times FN)}{(TP + FP) \times (FP + TN) + (TP + FN) \times (FN + TN)}$$

Where the components are defined as:

- **TP (True Positive)**: The count of instances where both raters independently applied the same code, agreeing on a positive classification.
- **TN (True Negative)**: The count of instances where neither rater applied the code, agreeing on a negative classification.
- **FP (False Positive)**: The count of instances where the first rater applied the code but the second did not.
- **FN (False Negative)**: The count of instances where the first rater did not apply the code but the second did.

This formula quantifies the agreement between two raters, while adjusting for the possibility of random agreement. By incorporating both instances of agreement and disagreement, Cohen's Kappa (κ) provides a more nuanced measure of inter-rater reliability. It is particularly suitable for binary classification tasks, where accurate and consistent identification of themes is crucial to the integrity of the analysis. The resulting Kappa values offer insight into the reliability of the coding process, with higher values indicating stronger agreement between raters.

To further evaluate inter-coder reliability, we computed Cohen's Kappa for all codes identified by at least one rater, with the results visualized in the histogram

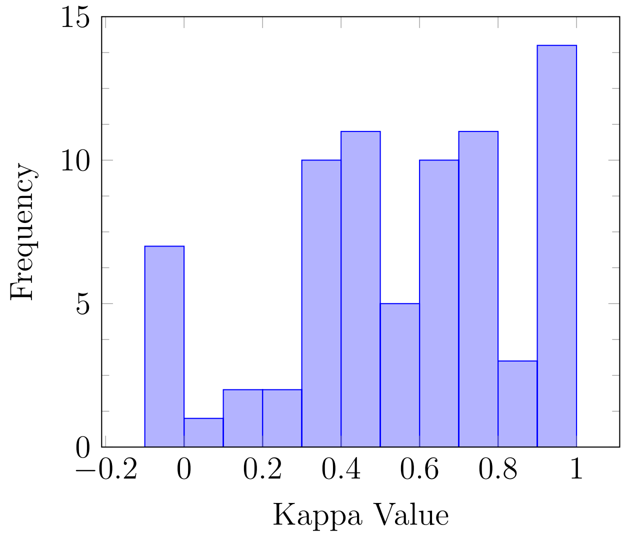


Figure 3.1: Histogram of inter-coder agreement as measured by Cohen's Kappa for all questions and codes identified by at least one rater. The small negative values indicate instances of no agreement on rare codes, highlighting the challenge of coding infrequent themes. However, for more common codes, the histogram shows strong inter-rater reliability, with the majority of Kappa values falling between 0.3 and 0.9, indicating moderate to high agreement.

seen in Figure 3.1. The distribution of Kappa values provides insight into the level of agreement across different thematic codes. As depicted, the majority of Kappa values range between 0.3 and 0.9, representing moderate to substantial agreement between raters. The presence of small negative values reflects instances where no agreement was reached, which primarily occurred for very rare codes. This distribution suggests that while disagreement was minimal and mostly confined to infrequent codes, common themes demonstrated high levels of consistency between raters. Such patterns confirm the overall reliability of the thematic coding process, as inter-rater agreement was strongest for the codes that appeared most frequently in the data.

3.4.2 Accuracy

Accuracy was calculated to assess the overall agreement between the two raters before any discussion took place. It measures the proportion of correctly classified instances, both true positives (TP) and true negatives (TN), relative to the total number of classifications. The formula used for accuracy is as follows:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Population}}$$

This metric provides a straightforward, intuitive way to evaluate the level of consistency between raters in applying codes to the data. By summing the true positives (cases where both raters applied the same code) and the true negatives (cases where neither rater applied the code), we can determine how often the raters agreed, regardless of whether the code was applied. Accuracy gives a general sense of how reliably the raters performed, though it does not account for the possibility of chance agreement. Instead this is captured by Cohen's Kappa, which is discussed in detail in Section 3.4.1.

To provide a clear illustration of the accuracy scores across codes, a subset of results from question 1.1 has been included in Table 3.4. This table displays the Cohen's Kappa values alongside the accuracy scores for each code, showing both agreement metrics side-by-side. The full set of results for all questions can be found in Chapter 5, where we provide a comprehensive analysis of the inter-rater reliability across the entire dataset.

	Code	Kappa	Accuracy
Question 1.1	INTERpos	0.8321783105	92.1%
	ONL	1	100%
	IRL	0.3784530387	94.1%
	ROUNDpos	0.648721921	96.1%
	PROF	0.6856677524	86.2%
	OUTCOME	-0.0221987315	91.4%

Table 3.4: Kappa and Accuracy by Code for Question 1.1, excluding codes which were not used or resulted in perfect disagreement

3.4.3 Multiple Choice and True / False

In addition to the qualitative analysis, a quantitative approach was applied to multiple-choice and true/false questions from the survey data. The more structured nature of these questions allowed for a straightforward statistical analysis, contrasting with the complexity of coding qualitative responses.

For effective visualization, bar charts and pie charts were used to illustrate the distribution of responses and trends within the data. These graphical representations made it easier to interpret patterns in the responses and communicate the quantitative findings clearly. By employing these visual tools, the analysis became more accessible, allowing for a quicker identification of key insights related to student engagement and question-asking dynamics.

3.4.4 Confusion Matrices

To visualize the discrepancies in coding before reaching consensus, a selection of confusion matrices for question 1.1 have been included as a representative example. For brevity, only a subset of these matrices is presented, as including all matrices for every question would have unnecessarily filled many pages. However, it was still valuable to incorporate a few as confusion matrices offer a clear visual representation of how well codes aligned between raters, and they help identify where inconsistencies occurred. This visualization allows for a deeper understanding of how frequently both raters agreed or disagreed on specific codes before coming to an agreement.

The confusion matrices are visualized as heat maps, where the intensity of each quadrant reflects the frequency of that particular classification (e.g. true positives or false negatives). These heat maps provide a more intuitive understanding of inter-rater agreement, allowing us to quickly assess where raters aligned and where inconsistencies occurred. By focusing on the most common codes, we were able to observe patterns that were crucial for refining the coding process and enhancing the overall reliability of the analysis.

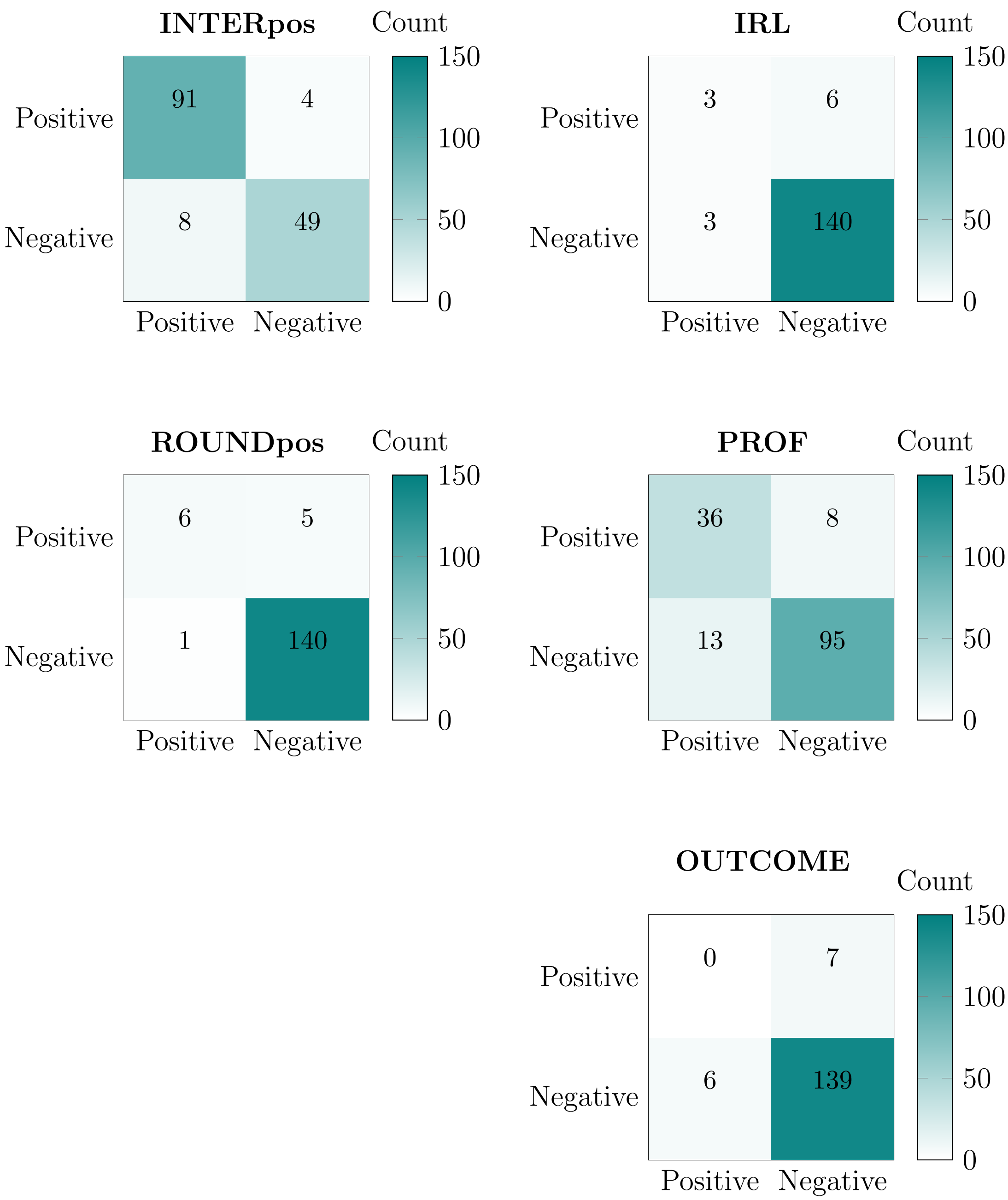


Figure 3.2: Question 1.1 Pre-agreement confusion matrices for all used codes. Excludes 100% agreement or 100% disagreeement

Chapter 4

Initial Prototype Design

The first iteration of the prototype for this study was implemented in a first-year computer science classroom at McMaster University during the fall semester of 2023. This prototype introduced a combination of prerecorded group presentations, question collection via the Desire2Learn platform (branded as Avenue to Learn at McMaster), and the use of Meta's large language model, LLaMA, to support question selection for class discussion. The overall goal was to explore how to enhance student participation and engagement in large classroom settings, while maintaining an emphasis on flexibility, privacy, and inclusivity.

4.1 Course Design Overview

The initial prototype was designed around the existing course infrastructure with additional technological support provided by Avenue to Learn and the LLaMA model. Avenue to Learn was the platform through which students accessed course content and submitted assignments and questions. Microsoft Teams (MSTeams), was the platform where students could access live classes and recordings, and interact with classmates and instructors. The prototype aimed to address the specific needs of a large, lecture-based computer science course, in which maintaining active student engagement and facilitating meaningful interaction posed significant challenges. The design of this prototype was largely the result of interviews with the course's instructional staff.

4.1.1 Prerecorded Presentation Access

A significant component of the initial prototype was the integration of various online elements designed to provide students with flexibility and autonomy in how they accessed and interacted with course content. The course made use of prerecorded group presentations, which were created by student groups and made available online through Microsoft Teams and Avenue to Learn. This hybrid approach allowed students to view the presentations either during class time or asynchronously, offering them greater control over their schedules.

The prerecorded presentations introduced several key features that enhanced the learning experience. Students had the ability to pause, rewind, and re-watch segments of the videos, allowing for a more personalized pace of learning. For those who preferred faster content consumption, the platform also supported viewing the presentations at increased playback speeds (e.g., 1.5x or 2x speed). These features were particularly beneficial for students who needed to revisit complex concepts or were managing busy schedules that made live attendance difficult.

In addition to the video content, the presentations' transcripts were made available. These transcripts allowed students to follow along with the presentation text, search for specific sections, or review content without needing to watch the entire video again. This feature not only catered to different learning styles but also added important accessibility options for students who were hard of hearing or had a language barrier, such as international students. By providing both video and text-based materials, the course ensured that students with diverse needs could access and engage with the content in a way that best suited their abilities and preferences, fostering a more inclusive learning environment.

Students could access these resources from any location with an internet connection, making the presentations highly accessible whether they chose to watch in class, at home, or on the go. This flexibility was a key design consideration, especially given the diverse schedules and commitments of first-year students. Some students commented on enjoying watching the videos while commuting to class. By allowing students to engage with the material at their convenience, the prototype aimed to reduce the barriers to participation and ensure that all students, regardless of their circumstances, could stay engaged with the course.

This design, leveraging prerecorded videos, transcripts, and flexible viewing options, provided a foundation for a more accessible and student-centred approach to learning. It acknowledged the diverse needs of the students, allowing them to learn at their own pace and revisit materials as necessary, which is critical in large, content-heavy courses like first-year computer science.

The Q&A portion, however, was conducted live during class. After watching the presentations, students could participate in a live question-and-answer session where the presenting group addressed questions submitted by their peers. These Q&A sessions were also recorded and made available afterward for students who could not attend in real-time. However, students were not able to submit questions after the live session, and any questions not asked during the session would not be addressed by the presenting group. This structure was designed to incentivize engaged participation and encourage students to attend the live sessions.

4.1.2 Question Submission via Avenue to Learn

One of the primary ways students interacted with the presentations was through submitting questions. Using Avenue to Learn (A2L), students submitted questions for each presentation as they watched through a form linked to each group's presentation. The question portal was only available during class time to ensure students were asking while the groups were available to answer their questions. The platform required that submissions be linked to the student's ID. However, during the live sessions, all submitted questions were presented anonymously to the classroom, so the other students and presenting groups could not see who had asked each question. This allowed for privacy and reduced potential social anxiety while reducing irrelevant or inappropriate comments.

From the instructor's perspective, this system provided a dual benefit: instructors could see the names of the students who submitted questions, allowing them to track participation and provide positive feedback, but students did not face the potential negative consequences of having their questions singled out in front of the class.

4.1.3 Question Selection and Live Q&A

In addition to manually selecting questions, the course design incorporated LLaMA, Meta's large language model, to assist with selecting a representative sampling of questions. Hundreds of questions were submitted for each group since 197 students could ask up to three questions. The course design incorporated LLaMA, Meta's large language model. The LLM was run locally on a graduate student's private computer, ensuring that student data remained secure and private, and did not leave the university's infrastructure. This was a key design choice to address any concerns about privacy, although the submissions did not contain any sensitive information.

The LLM's role was to perform a sorting task: after students submitted their questions via Avenue to Learn, a graduate student inputted the list of questions into LLaMA where the LLM would analyze and select 5-8 of the most interesting and representative questions conducive to class discussion. The professor also chose 5-8 questions manually, however, this took much longer, preferred early submissions, since questions were posted based on a single live read-through, and was cognitively resource intensive. This process allowed for a comparison between human and AI-curated questions and explored the potential for AI to augment instructional tasks.

The language model used was the 13-billion-parameter version of LLaMA, deployed without any specific training or fine-tuning for the task. It was utilized straight out of the box, and simply given a prompt designed to direct its selection process. The prompt used was as follows: "After this set of instructions, I will give you a large list of questions related to a book on [topic]. Your goal is to filter and select 5-8 questions from the following list that are representative, interesting and most likely to prompt a good discussion. Do not change the wording of the questions at all, simply select the questions and give it back to me." This basic application of the LLM demonstrated how off-the-shelf AI tools could be integrated into classroom settings to enhance educational experiences.

Once the questions had been selected by both the professor and the LLM, they were presented anonymously during the live Q&A sessions. This setup allowed the presenting group to engage with a curated set of questions without being overwhelmed by hundreds of submissions.

The anonymity of the questions promoted a low-pressure environment, encouraging students to submit without fear of judgment or embarrassment. As only a limited number of questions could be addressed in the live setting, students received implicit positive feedback when their questions were selected, while those whose questions weren't chosen did not face any negative feedback.

The main motivation behind this solution was to address the practical and cognitive challenges of sifting through hundreds of questions in a live environment. Without such a system, if students were tasked with manually sorting the questions, they would likely select from the most recent or randomly choose from partway down the list, without the time or energy to meaningfully engage with the entire set of submissions. This approach would be inherently biased, favouring more recent questions and disadvantaging others.

Alternative solutions, such as allowing students to vote on questions, were considered but ultimately rejected. Voting introduces additional social pressures, which could lead to anxiety for students whose questions might be ignored or rated poorly by their peers. Moreover, voting tends to favor questions that are submitted earlier, simply because they have more time in the framework to gather votes. The LLM-based approach, while not perfect, addressed many of these concerns by ensuring a more thoughtful and representative selection process. It provided a more equitable platform, promoting anonymity and minimizing social pressure while ensuring that the selected questions were meaningful and conducive to further discussion.

4.2 Design Features and Considerations

4.2.1 Flexibility in Content Delivery

A key goal of the prototype was to provide students with greater flexibility in how they engaged with course material. The ability to view presentations asynchronously helped students manage their time and allowed them to absorb content at their own pace. This flexibility was particularly valuable in a large first-year course, where students were often adjusting to the increased autonomy of university-level learning.

4.2.2 Balancing Anonymity and Accountability

The design of the question submission process struck a balance between anonymity and accountability. While instructors could see which students submitted questions, the anonymity of the questions in the classroom fostered a safer space for students to express themselves. This hybrid model provided positive reinforcement for students whose questions were selected while avoiding the negative consequences of public rejection for those whose questions were not chosen.

4.2.3 Integrating AI into the Classroom

Integrating AI into educational settings has been met with significant apprehension from both instructors and students. Concerns range from AI systems potentially replacing human roles in teaching, to fears of academic dishonesty, such as students using large language models (LLMs) like ChatGPT to complete assignments for them. However, introducing LLaMA into this prototype provided a unique opportunity to demonstrate how AI could be incorporated behind the scenes to enhance classroom dynamics without encouraging misuse or undermining academic integrity.

In this case, the LLM was tested as an assistant to the instructor. It performed a task, sorting through and selecting questions from hundreds of student submissions, that placed a considerable burden on the professor. To make this possible, the professor watched every presentation video in advance so the class time was available for reading the questions as they were submitted. Thirty hours of instructor time could have been saved by using the LLM, and relieved the cognitive load on the instructor while maintaining the human oversight necessary to ensure that discussions remained pedagogically sound.

Since the AI was tasked with sorting and selecting questions, rather than producing answers or academic content, it could assist the teaching process without interfering with student learning. This allowed the AI to enhance instructional efficiency while preserving the integrity of the educational experience. When carefully implemented, AI can augment the educational process by automating repetitive tasks such as sorting and selecting content, allowing instructors to focus on more meaningful, human-centred aspects of teaching.

4.3 Summary

The initial prototype introduced several innovative features to improve student participation and interaction in large courses. By combining prerecorded group presentations, a structured question submission process via Avenue to Learn, and AI-assisted question selection using LLaMA, the prototype sought to address key challenges associated with large classroom environments. Importantly, the prototype was designed with privacy and flexibility, ensuring students had control over how they engaged with the material while fostering a classroom culture of inclusivity and anonymity.

These features laid the groundwork for future iterations of the prototype, which would further refine the integration of AI tools, enhance the flexibility of content delivery, and continue to address the needs of both students and instructors in large academic settings.

Chapter 5

Prototype Feedback & Results

The design thinking process emphasizes the importance of building the right product instead of building the product right. User feedback is an integral part of the development process. User feedback is taken and used to iterate and improve upon the product and create another version. Then the cycle continues.

As part of the empathize phase of the design thinking process, we sought feedback from two groups of stakeholders: students and instructional staff. This was done through user feedback in two different ways. User interviews which were done for all the senior instructional staff, and questionnaire data for the students.

We interviewed all of the senior instructional staff, including two graduate teaching assistants (TAs) and one instructor. Given the course's nearly 200 students, interviewing each individually would have been impractical, so we conducted surveys via Avenue to Learn, the course's online learning management software. Participation was high, with around eighty percent of students responding to at least one questionnaire.

5.1 Summary of Feedback from Instructional Staff

Question and Answer Timing and Audio Challenges: Instructional staff reported that time constraints significantly impacted the depth of discussions, particularly during lengthy class periods that also included lab sessions. Audio projection issues necessitated makeshift solutions like strategically positioning microphones near computers, although classroom noise was not generally disruptive.

Attendance Influences: External factors, including public transportation strikes, impacted in-person attendance, prompting a shift towards increased virtual participation.

Question Submission System: The system allowed students to watch presentations in advance and submit questions early, which fostered more thoughtful and well-articulated inquiries. Early submissions were more likely to be selected for discussion.

Barriers to Engagement: Non-native English speakers often relied on tools like ChatGPT for interpreting or formulating questions, indicating a reliance on AI tools that could hinder genuine understanding. Anonymity in question submission was highlighted as significantly beneficial in boosting participation among students with anxiety, helping to maintain a focus on content rather than the fear of judgment.

Medium and Message: Two of the presentation books, *Grasp*[34] and *Failure to Disrupt*[32], discussed attempts to transform education using technology. Students in these groups made connections between the ideas discussed in the books, the active classroom, and the flipping of the classroom in the sense that presentations were pre-recorded and that students were responsible for selecting and presenting course material, and the emphasis on developing teamwork and communication skills. Many other books addressed ethical issues, to the point that many students felt they needed to address ethical issues in their questions. In these ways, students connected the content and structure of the course. Many students brought up the importance of these skills to their future careers in discussions with the professor.

5.1.1 Design and Feature Suggestions for Prototype 2

Live Transcription and Question Extraction: Proposals were made to integrate live transcription features to automatically capture and summarize questions during live discussions. This would potentially include using large language models to enhance the accuracy and relevance of captured content.

Enhanced Interaction Protocols: A "speaking stick" protocol was suggested to regulate turn-taking during discussions, which would aid in improving the accuracy of live transcriptions by reducing crosstalk and overlapping conversations.

Analytical Tools for Question Assessment: The concept of visually organizing questions on a graphical plane was discussed, where questions of similar themes could cluster together, aiding in the analytical assessment and fostering deeper discussion on related topics. Additionally, a matrix for evaluating the depth and relevance of questions was proposed to encourage students to refine their inquiries into more impactful contributions.

Balancing Anonymity with Engagement: While anonymity was recognized as a critical feature for encouraging participation, there was also a consensus on the need for features that would encourage students to gradually overcome communication apprehensions, thereby fostering a more engaging and interactive learning environment.

Figure 5.1: Colour Legend

5.2 Summary of Findings from Students

Student data was collected through the quantitative and qualitative analysis of the questionnaire data. In the framework of Design Thinking, this represents our first opportunity to empathize with the key stakeholders, the students. It also represents feedback on Prototype 1, described in Chapter 4.

All pie charts use the colour key for codes are presented in Figure 5.1.

5.3 Presentation of Questionnaire Data

This section presents the data collected from the two questionnaires used to gather student feedback. For clarity, the notation “Question 1.x” is used to refer to specific questions from the first questionnaire, and “Question 2.x” denotes questions from the second questionnaire. For example, “Question 2.6” references the sixth question of the second questionnaire.

5.3.1 Question 1.1

Question 1.1 sought to identify what students considered to be the most valuable aspect of the university classroom experience. See Figure 5.2. Through thematic analysis, it became evident that positive social interactions were central to the students’ perceptions of value. This finding is significant because, while educational technology often focuses on content delivery and automation, the human elements of learning—particularly social engagement—remain a vital component of the classroom experience.

The most prominent theme, coded as INTERpos, with 103 responses, underscored the importance of peer interactions. Students emphasized the value of collaborating with classmates, engaging in discussions, and exchanging ideas. These interpersonal dynamics were seen as essential for learning, as they fostered a sense of community and provided opportunities for deeper exploration of the material. Collaborative learning, where students actively engage with one another, supports cognitive development by exposing them to different perspectives and encouraging critical thinking. It also aligns with socio-constructivist theories, which argue that learning is inherently social, and knowledge is constructed through interaction with others.

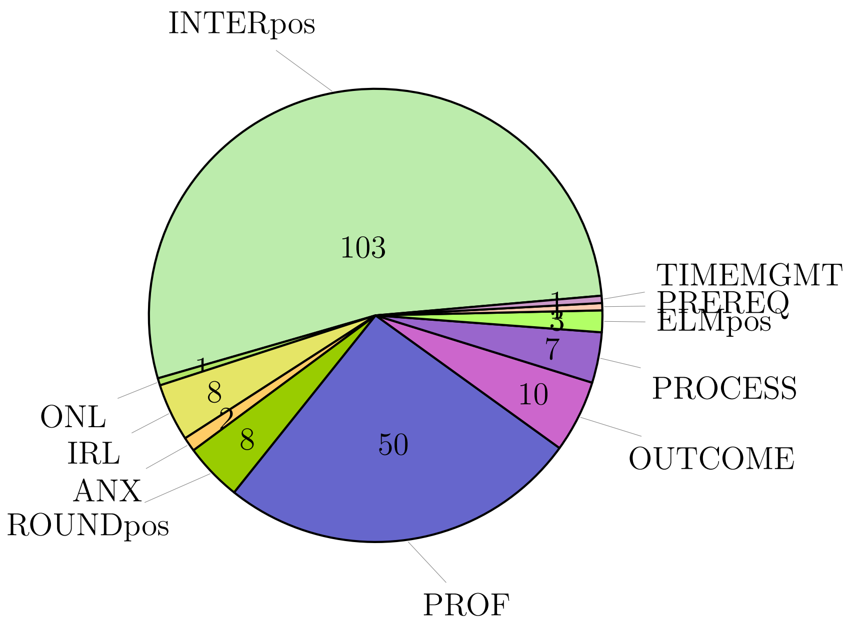


Figure 5.2: [Q1.1] In your opinion, what is the most valuable part of the university classroom experience?

The second most frequent theme, PROF, with 50 responses, highlighted the critical role that professors play in shaping the classroom experience. Many students mentioned that the quality of instruction, as well as direct communication with professors, significantly impacted their learning outcomes. Professors are seen not only as transmitters of knowledge but also as facilitators who guide discussions, provide feedback, and foster a supportive environment. This suggests that students greatly value the expertise and engagement of their instructors, viewing them as key figures in their educational journey.

Interestingly, only a small number of students (8 responses under IRL and 1 under ONL) focused on the format of learning itself, with few distinguishing between in-person and online learning as a key factor in their overall experience. This suggests that while the mode of delivery can be important, the quality of interactions—both with peers and professors—far outweighs the specific format in terms of perceived value.

Other noteworthy but less prominent themes included OUTCOME (10 responses), which reflected students' focus on the end results of their education, such as skills acquired or grades earned, and PROCESS (7 responses), where students emphasized the importance of the learning process itself over outcomes.

These findings point to a need for educational technologies that prioritize and support the social aspects of learning. Tools that facilitate peer-to-peer interaction, real-time discussions, and accessible communication with professors could enhance student engagement and create a more dynamic and fulfilling classroom experience. While technology can enhance content delivery, these results suggest that any successful educational tool must also consider how to replicate or amplify the human elements of learning that students find most valuable.

In an era dominated by Massive Open Online Courses (MOOCs) and the overwhelming availability of educational resources online, the social component of the university experience has emerged as one of its most distinctive and irreplaceable features. While MOOCs provide accessible and often high-quality content, they largely lack the interpersonal dynamics that define traditional university settings. The ability to engage in face-to-face discussions, collaborate on projects, and form meaningful connections with peers and professors adds depth to the learning process that cannot be replicated through asynchronous or solitary study. This social interaction fosters not only academic growth but also personal development, creating a sense of community and belonging that enriches the university experience. The human element—peer engagement, mentorship, and shared learning—thus remains a core advantage of the university environment, differentiating it from the purely content-focused approach of online education platforms.

By focusing on interpersonal dynamics and professor-student relationships, educational technologies can better align with the aspects of the university classroom experience that students hold in the highest regard.

5.3.2 Question 1.2

The responses to **Question 1.2** "What is the most challenging part of the university classroom experience? What makes learning/achieving success challenging? What do you wish could be changed?" provided valuable insights into students' obstacles in the university environment. See Figure 5.3. The most frequently cited challenge was time management (TIMEMGMT), with many students expressing difficulty balancing the demands of academic life with personal and extracurricular commitments. This is particularly relevant in the post-pandemic context, where shifts in learning environments have exacerbated struggles with motivation and self-regulation, as discussed in the work of Engzell et al. [15] in Chapter 1.

The second most cited theme was the pace (PACE) of content delivery, with students often finding the speed overwhelming and difficult to manage alongside other tasks. This may be interconnected with time management concerns, where students feel pressured by rapid information delivery and struggle to keep up. Resources (RESOURCE) also emerged as a notable challenge, reflecting a perceived lack of access to teaching staff, academic support, or learning materials that could hinder effective learning.

Several responses identified class size (SIZE) and focus (FOCUS) as key environmental factors affecting student success. Large class sizes can feel intimidating, while distractions within the classroom hinder concentration. Concerns about rubrics (RUBR) also emerged, noting that unclear grading criteria and expectations created frustration. Anxiety (ANX) was frequently noted as a psychological barrier to participation, particularly for students hesitant to speak up in class. Anxiety can also point to uncertainty about course expectations and performance pressure. In larger classrooms, students may feel anonymous or hesitant to engage, fearing they'll ask inappropriate questions that will elicit negative responses from their peers. Anxiety is also linked to time management challenges and the fast pace of content delivery, which heightens stress for those struggling to keep up.

These themes can be categorized into two broader concerns: logistical and psychological barriers (TIMEMGMT, PACE, RESOURCE, ANX, PERS) and environmental factors (SIZE, FOCUS). Together, they reveal a complex network of challenges that students face in large classroom settings, requiring a multifaceted response. Solutions should address not only the physical environment but also the support structures for time management, pacing, and mental well-being.

The implications for instructional design underscore the need for flexible, inclusive, and supportive educational structures that cater to diverse student needs and learning styles. It is interesting to note that INTER appears as both a positive and a negative, despite the question being framed to elicit responses about barriers and challenges. While many students viewed peer interaction as essential for developing non-technical skills, they also identified it as a source of frustration (e.g., when peers fail to complete required work) and anxiety (e.g., due to shyness or language barriers). Understanding these nuanced perspectives is crucial for refining teaching methodologies and designing learning spaces that promote effective learning outcomes.

5.3.3 Question 1.3

In **Question 1.3**, students highlighted anxiety (ANX) as the primary deterrent from asking questions in class. See Figure 5.4. Anxiety in this context often stems from fear of judgment, the pressure of speaking in front of large groups, and concerns over appearing uninformed in front of peers. For post-secondary educators, understanding this anxiety is critical because it represents a significant barrier to active learning and dialogue, both of which are essential to fostering deeper understanding and critical thinking.

This fear of public exposure, particularly in large lecture halls where students may feel anonymous yet scrutinized, often leads to a form of learned silence, where students opt out of participating even when they have valuable questions or need clarification. The social dynamics of the classroom—such as students' perceptions of their peers' judgment or the professor's potential reaction—can intensify these feelings, making it more difficult for students to engage in spontaneous inquiry. Moreover, anxiety-related barriers to asking questions mean that instructors may lose key feedback on how well students are grasping the material. Without questions, professors may falsely assume comprehension, which can result in the continuation of misconceptions or gaps in knowledge.

While encouraging students to practice speaking up in environments with social pressure is important for building confidence and communication skills, it is equally

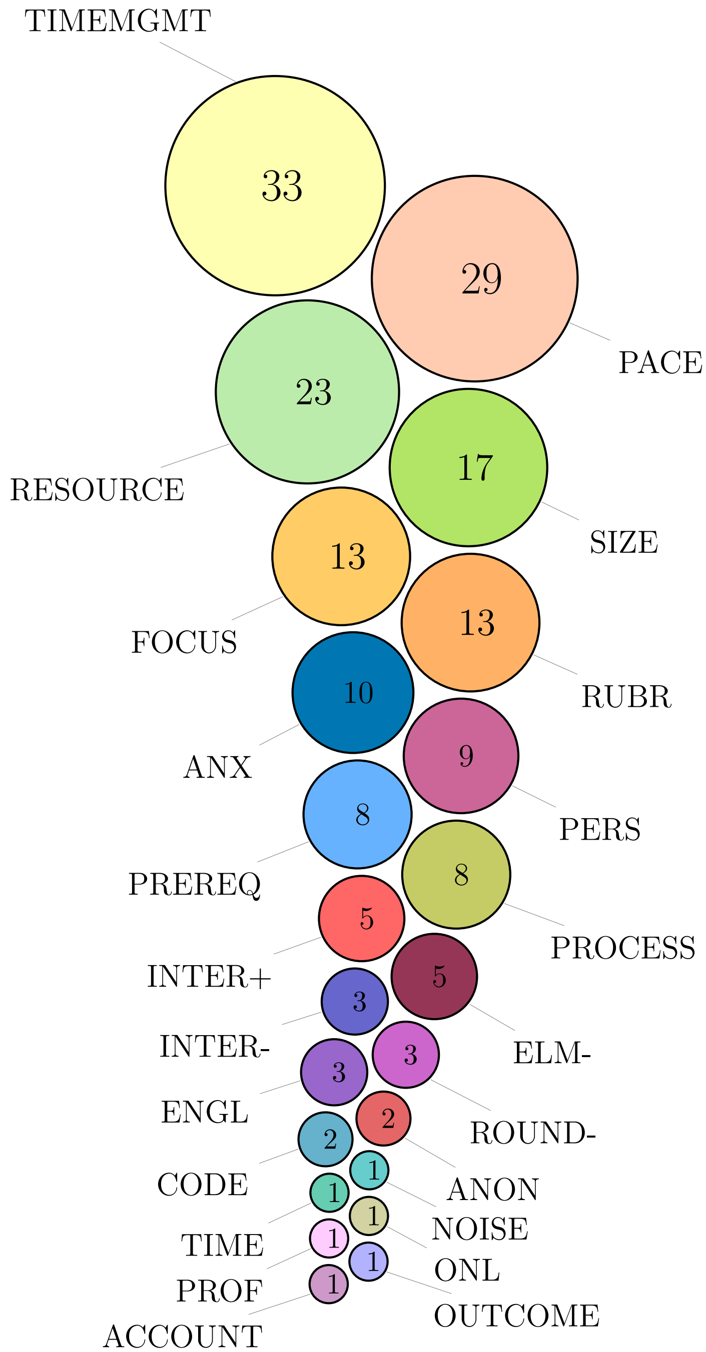


Figure 5.3: [Q1.2] What is the most challenging part of the university classroom experience? What makes learning / achieving success challenging? What do you wish could be changed?

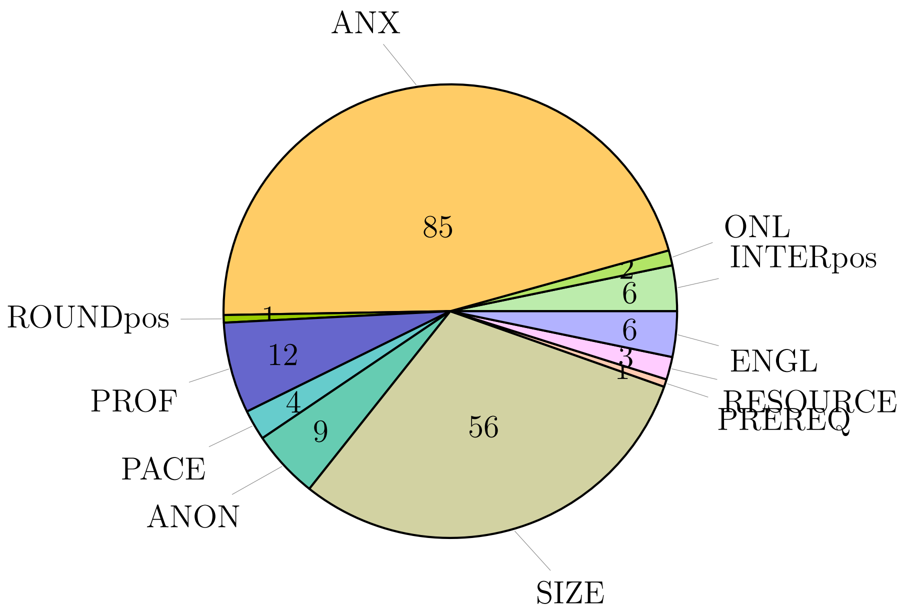


Figure 5.4: [Q1.3] How do you feel about the process of asking questions in class? What prevents you from asking questions in a traditional classroom setting?

essential to design solutions that cater to their varying comfort levels. By accommodating the diverse needs and capabilities of all students, we can foster a more inclusive environment that promotes active engagement and participation, ensuring that every student feels supported and empowered to contribute. This approach not only encourages more engaged learners but also helps to create a classroom atmosphere where all voices can be heard, regardless of individual differences in comfort and confidence. The data indicates a strong interest in features such as anonymous question submission, which would allow students to engage in a way that feels safer and more accessible. This approach could reduce the anxiety associated with participation, thereby increasing inclusivity and accessibility in the classroom.

The SIZE theme also emerged as a key factor, suggesting that the large scale of lecture halls may contribute to students' reluctance to ask questions. This could be linked to the impersonal nature of large classrooms and is particularly relevant given the rise in social anxiety among students [24, 30].

Additionally, other concerns related to the themes of PROF, PACE, and ANON highlight the importance of the professor's approachability, the pace of material delivery, and the desire for anonymity in influencing students' willingness to ask questions. The professor's approachability is critical for engagement because it directly impacts the comfort level students feel when seeking clarification or engaging in discussions. An approachable professor creates an environment where students feel safe to voice concerns, ask questions, and participate in discussions with less intimidation or fear of judgment.

It's reasonable to expect a correlation between pace and approachability as students who feel that the material is being covered too quickly may hesitate to ask for clarification or pace adjustment if they perceive the professor as unapproachable or dismissive. In contrast, when students perceive the professor as approachable, they are more likely to speak up if they find the pace overwhelming, thereby providing the professor with the opportunity to adjust the delivery to better suit the class's needs. This dynamic interaction can help prevent the negative cycle of confusion and disengagement that might occur in a less supportive classroom environment. Understanding and addressing the interplay between these factors is crucial for creating a learning environment that encourages active participation and accommodates the diverse needs of students.

These findings suggest that creating a more supportive classroom atmosphere, potentially through anonymous questioning options or more personalized interactions

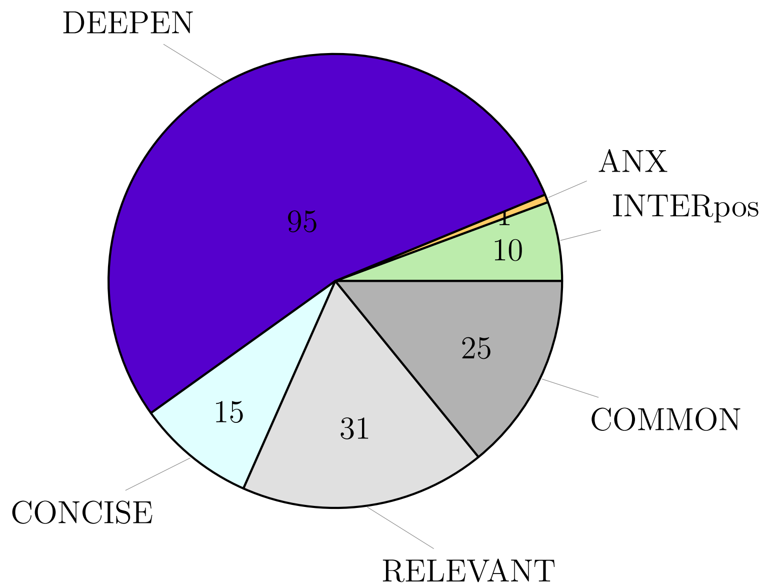


Figure 5.5: [Q1.4] What makes a question a great question in your opinion?

with instructors, could help overcome these barriers. Addressing both the psychological and environmental factors that inhibit participation is essential for fostering an environment that encourages active student engagement, ultimately enhancing the overall educational experience in large classroom settings.

5.3.4 Question 1.4

Question 1.4 investigates what students perceive to be the key qualities of a great question. See Figure 5.5. Understanding this is crucial for guiding the development of AI-driven educational tools that can not only filter questions effectively but also coach students on how to formulate them. The goal was to discern the most important factors that contribute to what students consider a meaningful and impactful question so that the AI could use these criteria when sorting and selecting questions.

The data indicated a strong preference for questions that deepen (DEEPEN) understanding, with 95 responses falling into this category. Students clearly valued inquiries that go beyond surface-level information, reflecting a desire to explore topics in greater depth. This aligns with constructivist learning theories, which suggest that active engagement with complex, thought-provoking material enhances learning. By encouraging critical reflection and promoting metacognitive skills, deeper questions help students better regulate their own learning processes and strategies.

Relevance (RELEVANT) was another significant factor, highlighted by 31 respondents. Students preferred questions that were directly related to the course material, emphasizing the importance of coherent contributions that follow the flow of class discussions. Relevant questions help maintain focus, reinforcing key concepts and preventing distractions or off-topic detours. This underscores the role of well-timed, pertinent inquiries in creating a productive learning environment, benefiting both the questioner and their peers.

Concise (CONCISE) and common (COMMON) questions were less emphasized but still present in the feedback. Fifteen students mentioned the value of concise questions, recognizing that brevity helps maintain the flow of a lecture and avoids overcomplicating discussions. Twenty-five students highlighted common questions—those that many others are likely wondering as well. These types of questions reflect the importance of social learning dynamics, where students benefit from hearing questions that they may be too shy or unsure to ask themselves. Common questions promote a shared learning experience, bridging individual concerns with collective classroom understanding.

The feedback suggests that while students appreciate brevity and shared concerns, their primary preference is for questions that enhance learning through deeper analysis and critical engagement. These types of questions are seen as the most effective for driving rich classroom discussions and promoting a deeper understanding of the material.

Translating these insights into the design of educational tools involves prioritizing features that encourage meaningful student participation and guide them toward constructing thoughtful, relevant inquiries. For instance, the AI tool could include mechanisms that provide real-time feedback on the quality of questions, helping students refine their queries to be more concise and aligned with the subject matter. Over time, this could foster an environment where critical thinking is not only encouraged but actively facilitated by technology.

Incorporating these criteria into the AI's question-filtering algorithm would ensure that the tool aligns with students' learning needs and preferences, ultimately enhancing the educational experience by promoting more insightful and relevant classroom discussions.

5.3.5 Question 1.5

Question 1.5 investigates students' perceptions of how asking questions contributes to their learning process. See Figure 5.6. An overwhelming majority, 148 out of 152 respondents (97.4%), affirmed that asking questions is vital to their academic growth. This near-unanimous consensus underscores the importance of fostering a classroom environment that actively encourages and facilitates student inquiries. The ability to ask questions helps reinforce course material and prompts students to engage more deeply, critically analyze the subject matter, and clarify any areas of confusion. These elements are key to effective learning, as they stimulate cognitive processes beyond rote memorization, enhancing comprehension and long-term retention.

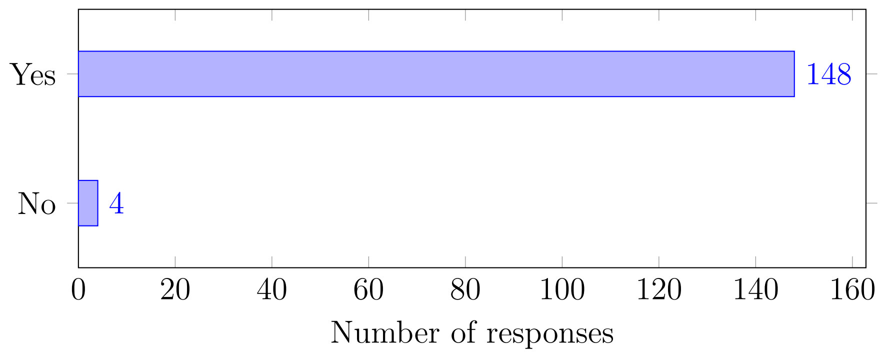


Figure 5.6: [Q1.5] Do you think asking questions improves your ability to learn?

The strong endorsement of question-asking aligns with constructivist learning theories, which posit that active student participation, particularly through inquiry, is central to constructing meaningful knowledge. Asking questions helps students challenge their assumptions, seek clarification, and explore the broader implications of their learning, all of which contribute to a richer and more dynamic learning experience. This finding suggests that any educational tool designed to support student engagement should prioritize mechanisms that make asking questions easier, more accessible, and less intimidating.

Simplifying the question-asking process is essential, not only because students recognize its value but because they actively seek opportunities to deepen their understanding through inquiry. The challenge, however, lies in overcoming the barriers that prevent students from asking questions—such as anxiety, fear of judgment, or the impersonal nature of large classroom settings, as revealed in earlier responses. Therefore, the design of an AI-driven educational tool must focus on creating a user-friendly, approachable interface that encourages frequent and meaningful student participation. Such a tool could include features like anonymous question submission or real-time feedback to reduce the anxiety associated with speaking up in class, thereby fostering a more inclusive learning environment.

This question, when paired with subsequent inquiries about the barriers to asking questions, provides deeper insight into the complexities of the student mindset. While students overwhelmingly agree that asking questions is crucial to their learning, they are often hindered by a variety of psychological and environmental factors. This disconnect between the recognized value of inquiry and the actual practice of asking questions reflects a broader issue in classroom dynamics: students want to engage and understand the material more fully, yet external and internal pressures prevent them from doing so.

The implications for educational design are clear: tools that streamline and demystify the question-asking process can significantly enhance student participation. By reducing friction points—whether through anonymity, intuitive interfaces, or scaffolding mechanisms that guide students in formulating their questions—an AI-driven system can create a more supportive learning environment. This, in turn, could help students overcome the barriers that inhibit their engagement, allowing them to take full advantage of inquiry as a means of deepening their understanding.

5.3.6 Question 1.6

Question 1.6 examines the frequency with which students ask questions during class, revealing a significant discrepancy between students' recognition of the value of asking questions and their actual participation in classroom discussions. See Figure 5.7. Despite the earlier overwhelming acknowledgment that question-asking enhances learning, students tend to engage in this practice infrequently. As illustrated in the

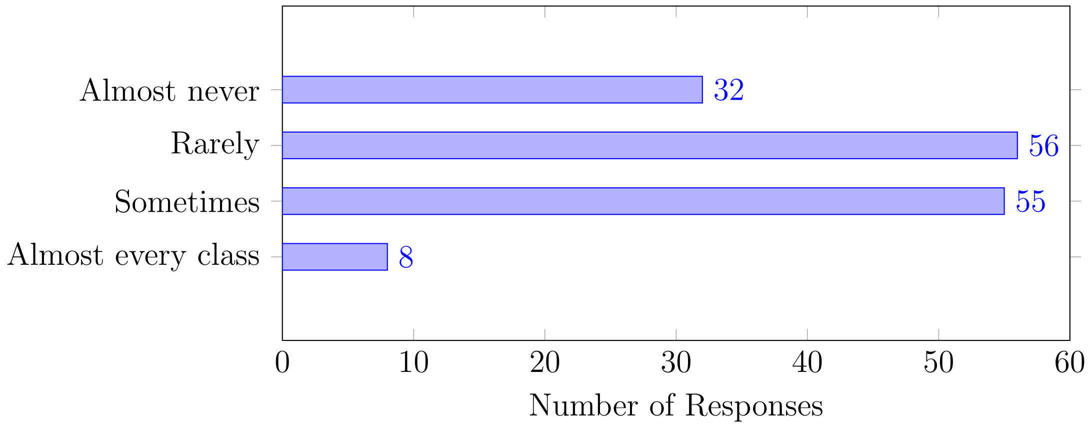


Figure 5.7: [Q1.6] Do you typically ask questions in class?

graph, the majority of respondents indicated that they ask questions “Sometimes” (55 responses) or “Rarely” (50 responses). Only a small fraction—8 students—reported asking questions “Almost every class,” while 32 students stated they “Almost never” engage in question-asking.

This contrast between understanding and action suggests that while students are aware of the benefits of asking questions, several factors inhibit their actual participation. Potential barriers, such as the size of the classroom, fear of peer judgment, social anxiety, and concerns about the relevance or quality of their questions, likely contribute to this gap. These barriers align with earlier findings in the survey, where students expressed reluctance due to anxiety (ANX) and perceived the classroom environment (SIZE) as an obstacle to active engagement. Additionally, the fear of appearing uninformed or asking a “bad” question may lead students to remain silent, despite knowing that asking questions would improve their learning outcomes.

This data highlights the importance of designing educational tools and classroom strategies that lower these barriers and create a more conducive environment for student participation. Tools that enable anonymous question submission or provide structured opportunities for students to ask questions without fear of judgment could significantly improve engagement. By removing the social pressure associated with public questioning, such solutions would help bridge the gap between students' understanding of the value of inquiry and their actual participation.

Furthermore, this discrepancy suggests that simply telling students that asking questions is valuable is insufficient; instead, instructors and educational tools must actively foster an environment where students feel safe and encouraged to ask questions. This could involve more personalized interactions with instructors, opportunities for asynchronous questioning, or digital platforms that normalize frequent questioning without putting students on the spot.

By understanding the factors that hinder students from participating, educators can design tools and strategies that emphasize the importance of inquiry and facilitate it in practice.

5.3.7 Question 1.7 - 1.11

To understand how various question-asking methodologies influence student willingness to engage during lectures, questions 7 through 11 of the first questionnaire presented students with different scenarios to test its effects. Each scenario explored a distinct approach: asking questions via a microphone in front of the class, raising hands during a lecture, using a public chat on Microsoft Teams, sending direct messages to teaching assistants or instructors, and submitting questions in an anonymized form. The goal was to identify the most effective methods to encourage student engagement and refine educational tools accordingly.

For **Question 1.7**, perhaps unsurprisingly, many students indicated they would

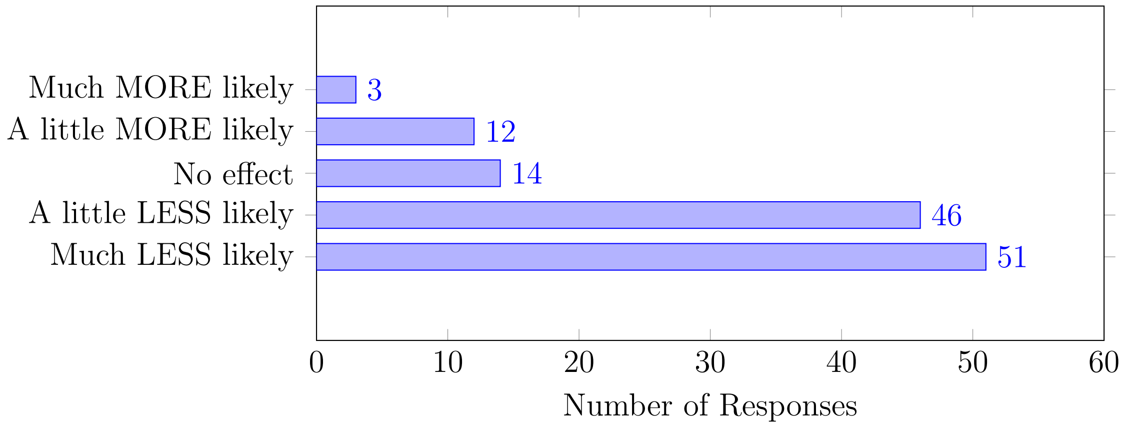


Figure 5.8: [Q1.7] Asking your question into a microphone in front of class

be "Much Less Likely" to ask questions this way, suggesting a general discomfort with public speaking and social anxiety. See Figure 5.8. This method was also the most active of the question-asking modalities, presenting the highest barrier to entry. It required students to physically get up from their seat and move to a different location in order to ask their question, adding an additional layer of discomfort. The combination of public visibility and the physical act of moving in front of the class creates significant friction for many students, discouraging participation. This finding points to the need to reduce public visibility in question-asking mechanisms to alleviate anxiety and facilitate greater student engagement.

In **Question 1.8**, responses showed a more balanced curve, with many students indicating they were "About the Same" likelihood to ask questions compared to other methods. See Figure 5.9. This reflects a moderate level of comfort with this traditional approach, likely because it has been ingrained in their educational experiences from a young age. However, despite this familiarity, students still reported not asking questions frequently in class, suggesting that comfort with the method alone doesn't necessarily translate into active participation.

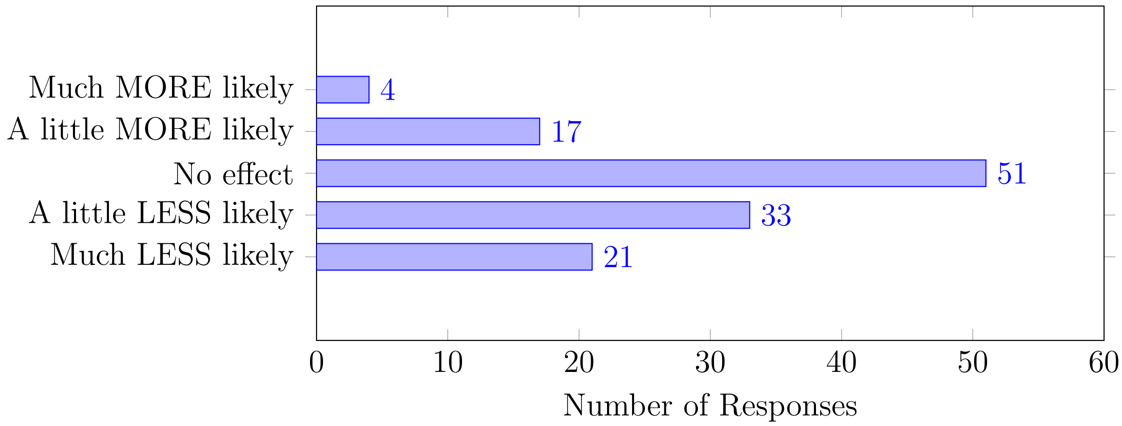


Figure 5.9: [Q1.8] Raising your hand in lecture

One potential factor is inertia, the effort required to initiate the act of raising one's hand, especially in large classrooms where the fear of drawing attention or being judged by peers is heightened. The ease of staying passive often outweighs the desire to engage, despite the perceived comfort with the method. This indicates a need to reduce the barriers, both psychological and behavioral, to lower the inertia of participating through hand-raising, or to introduce alternative methods that more naturally integrate participation into the classroom dynamic.

Question 1.9 explored asking questions via a public chat on Microsoft Teams, a digital collaboration platform commonly used for remote or hybrid classroom settings. See Figure 5.10. The responses leaned towards "A Little More Likely" or "No Effect" on the likelihood of asking questions. While this method reduces the physical act of public speaking, it still retains public visibility, which may explain why some students showed reservations about this method.

The key concern here seems to be the public nature of the chat. Even in a digital space, students may hesitate to ask questions due to concerns about how their questions will be perceived by their peers. This indicates that while digital platforms

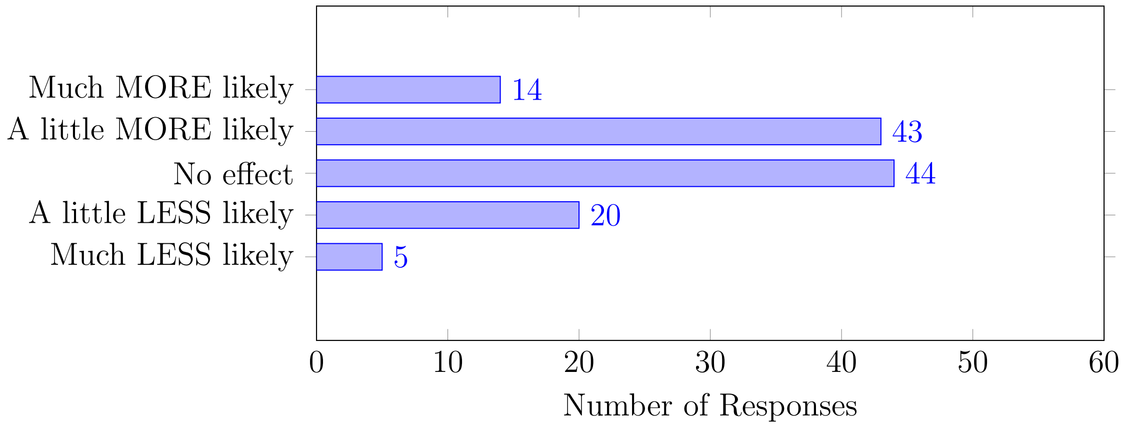


Figure 5.10: [Q1.9] Asking your questions on Microsoft Teams public chat

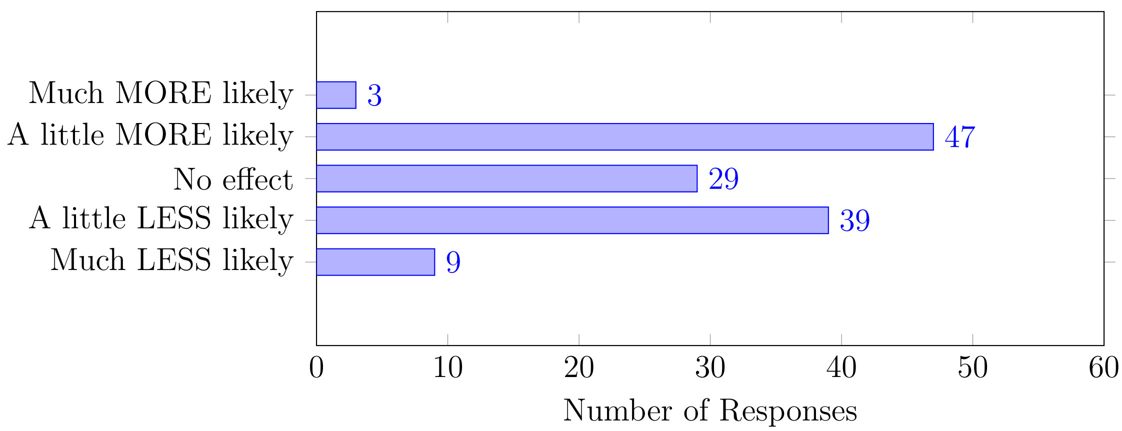


Figure 5.11: [Q1.10] Sending a DM to your TA / instructor

like Teams may facilitate easier communication, the public aspect of the platform still serves as a barrier to some students.

Question 1.10 examined the method of sending a direct message (DM) to a teaching assistant (TA) or instructor. See Figure 5.11. Responses were notably divided, with some students feeling "More Likely" to ask questions this way, while others felt "Less Likely." This split highlights the mixed levels of comfort students have with direct communication with instructional staff.

This division could be linked to previous findings on the theme of professor approachability 5.3.3. Students who perceive the professor or TA as approachable and open may feel comfortable engaging via direct messages, viewing it as a more private, low-pressure way to communicate. Conversely, if the instructor is seen as unapproachable or distant, students may avoid this method altogether. The variability in responses suggests that direct messaging could be an effective tool for some students but is highly dependent on the interpersonal dynamics between students and staff.

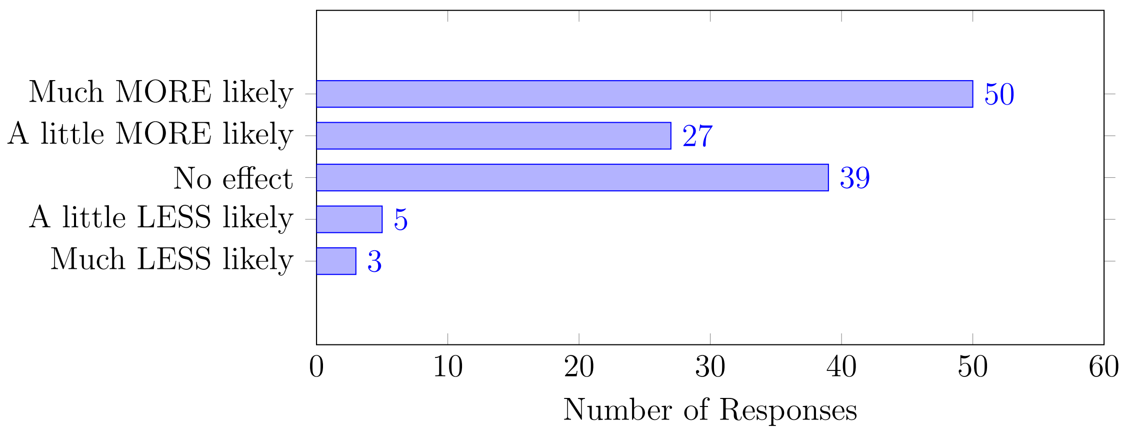


Figure 5.12: [Q1.11] Asking via anonymized form

Question 1.11 addressed asking questions through an anonymized form, and this method was overwhelmingly the most preferred, with many students indicating they would be "Much More Likely" to ask questions this way. See Figure 5.12. The anonymity of the method reduces social anxiety and the fear of peer judgment, making it a low-barrier option for students who might otherwise hesitate to ask questions.

The preference for anonymous question submission also aligns with themes related to reducing friction between the desire to ask a question and the act of doing so. Anonymity removes the need for students to worry about how they will be perceived, giving them time to formulate their thoughts without the immediate pressure of speaking in public. This method clearly has the most success in increasing question-asking behavior, at least among the options provided.

One important aspect worth exploring is self-moderation. In public forums, where questions are tied to a student's identity, there is a natural tendency to moderate inappropriate or irrelevant questions. An anonymous system, while reducing anxiety, could open the door to a flood of less useful questions. Here, AI could play a crucial role in filtering questions for relevance and quality, reducing the burden on instructional staff while enhancing the student experience.

Reflection on Question Design and findings Q1.7-1.11

It is worth noting that there may have been some confusion regarding the design of this question. Some students appeared to interpret it as asking how likely they would be to use each method compared to traditional question-asking practices, such as raising a hand in class, while others compared these methods to the experimental way questions were asked during this study. The intended comparison was between these modalities and the experimental setup, not traditional classroom practices.

This confusion likely stems from a lack of clarity in the question design. More context should have been provided to explicitly frame the scenario being referenced, ensuring that students understood that the question was asking for a comparison with the experimental setup. However, despite this ambiguity, the data still reveals important trends. Despite the design error, we can observe meaningful patterns in student preferences and behaviours that offer valuable insights into how different question-asking methods impact student participation.

These findings provide valuable insights into student question-asking behaviour.

Methods reducing public visibility, especially anonymized forms, are most effective at increasing student engagement. In contrast, more public methods, such as using a microphone or participating in a public chat, discourage students from asking questions. This evidence underscores the importance of offering diverse, low-pressure methods for students to engage in classroom discussions. By incorporating these findings into the design of educational tools, instructors can create environments that are more inclusive and conducive to active student participation.

5.3.8 Question 1.12

In **Question 1.12**, we examined student preferences between two sets of questions presented during a lecture: Set A, which was selected by the instructor, and Set B, selected by LLaMA (Locally Run Large Language Model). See Figure 5.13. The questions were drawn from the same lecture material, with selections made randomly using a random number generator to eliminate bias.

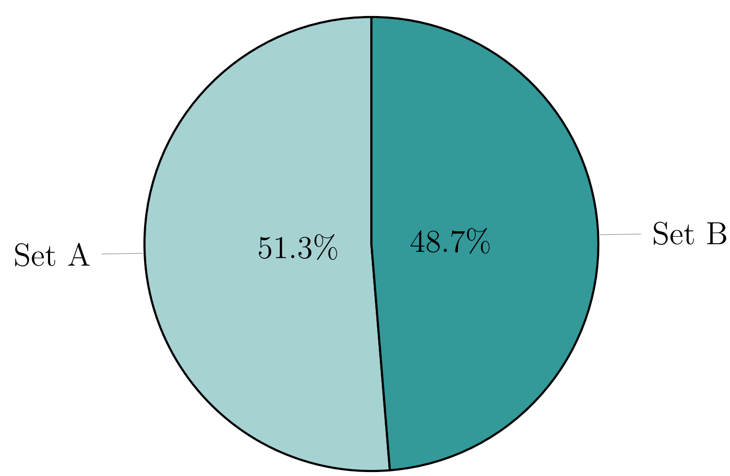


Figure 5.13: [Q1.12] Which set was better (LLaMA vs Instructor)

The purpose of this question was to conduct a preliminary assessment of the efficacy of AI-selected questions compared to those chosen by an instructor. Our goal was to determine whether there would be any significant difference in student preferences between the two, without attempting to conduct a rigorous, detailed study. Existing research already indicates that large language models are quite capable of generating relevant and insightful questions. This survey aimed to ensure there would be no obvious issues with using AI for this task in an educational setting.

The survey involved 151 respondents, and the results were closely split:

- **Instructor-selected questions (Set A)**: 77 votes
- **LLaMA-selected questions (Set B)**: 74 votes

The near-equal distribution of preferences suggests that students did not overwhelmingly favour one method of question selection over the other. This indicates that AI-generated questions are on par with those selected by instructors in terms of quality and relevance, based on student feedback. Importantly, the results do not show any indication that using AI to select questions would lead to adverse educational outcomes, suggesting that it can be a viable and effective method for enhancing classroom engagement.

While this was not an in-depth study, the findings support the potential for integrating AI tools like LLaMA into the question-asking process without compromising educational quality. Further research could explore how AI-generated questions could complement instructor-led questioning to provide a more dynamic and inclusive learning environment.

5.3.9 Question 1.13

Question 1.13 aimed to follow up on the findings from Question 1.12 by asking students how much better they perceived the set they selected. See Figure 5.14. The intent was to verify whether a clear preference drove the choices made by the students or whether they were relatively minor distinctions. We wanted to ensure that the selection of one set over the other was not purely arbitrary and to determine if any significant trends emerged from student feedback.

The survey responses indicated the following distribution of preferences:

- **31 respondents** chose "Significantly better"
- **107 respondents** chose "A little better"
- **14 respondents** selected "I chose at random"

These results confirm that the majority of students (107 out of 152) felt that the set they chose was only marginally better, with very few choosing at random. This suggests that the students were indeed discerning in their selections and that their preferences between AI-generated (LLaMA) and instructor-selected questions were not overwhelmingly strong.

For the respondents who indicated a "Significantly better" preference, the split between the two sets was relatively balanced. Of the 31 students who reported this stronger preference, **18** favoured Set A (instructor-selected), while **13** favoured Set B (LLaMA-generated), reflecting a 58% to 42% split. This small difference does not suggest a substantial bias toward either question set, further confirming that students did not show a marked preference for one method over the other.

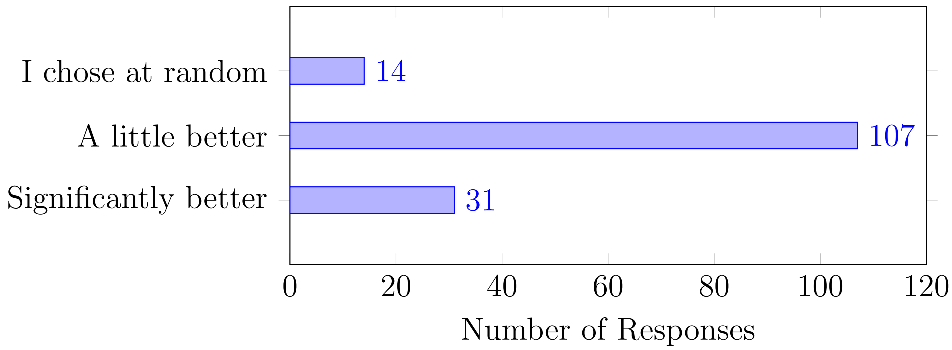


Figure 5.14: [Q1.13] How much better was the set you chose?

Crucially, these findings also help verify that students were not choosing their preferred set at random or without thought, as only 14 respondents indicated random selection. This strengthens the conclusions drawn from Question 1.12: both AI and instructor-generated questions are similarly effective in supporting student learning, and the slight edge in preferences does not point to any significant educational disadvantage for using AI tools in this context.

By examining the magnitude of student preferences, this question ensures that the subtle differences observed between question sets were valid and not the result of arbitrary selection. It reinforces the conclusion that AI-generated questions can complement traditional methods without diminishing the quality of classroom interaction.

5.3.10 Question 2.2

Question 2.2 asked students to evaluate how group presentations were handled in the experimental condition compared to other large classrooms. See Figure 5.15. The responses showed a largely positive reception. Of the 49 respondents, 26 (53%) found the group presentations to be "Better than other large classrooms," 18 (37%) rated them “About the same,” and only 5 (10%) rated them “Worse.”

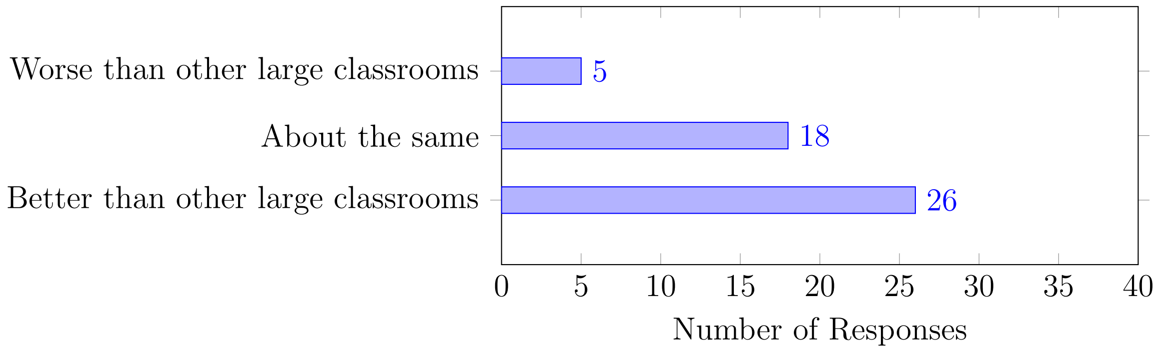


Figure 5.15: [Q2.2] How did you find the way the group presentations were handled?

Two key interventions in this experiment were the use of pre-recorded presentations instead of live delivery and the option for students to ask questions via an anonymous form, which were then filtered and presented for a live Q&A session. Pre-recording presentations can reduce performance anxiety for students and ensure that technical or logistical issues don’t disrupt the flow of the session. It also allows for more thoughtful and polished presentations, as students can review and perfect their delivery before submission.

The use of an anonymous question form and a moderated Q&A period is particularly noteworthy. This format reduces the social pressures often associated with large classrooms, where students may feel hesitant to ask questions publicly. By filtering questions and presenting them anonymously, the intervention lowers barriers to participation and encourages more meaningful engagement with the material. This approach aligns with broader themes in the research that suggest reducing the fear of peer judgment and creating safer spaces for inquiry can lead to deeper student involvement. The social risk of asking a question is dramatically reduced for the student.

The 37% who found the presentations "About the same" could reflect a preference for live interaction or a feeling that pre-recorded presentations lack the dynamic element of spontaneous question-and-answer sessions. However, with only 10% indicating a worse experience, it appears that the majority of students responded well to this method. This result suggests that pre-recorded presentations, paired with structured, anonymous Q&A sessions, may offer a viable solution to some of the challenges inherent in large classroom presentations, promoting a more equitable and accessible learning environment.

5.3.11 Question 2.3a

Question 2.3a looks at data from students who felt the group presentations were better than those in other large classrooms. See Figure 5.16. The most notable factors were the online elements of the course (18 mentions), the reduction of social anxiety (6 mentions), and positive social interactions (4 mentions). The online format, which included pre-recorded presentations and the ability to join live sessions, gave students more flexibility and control over their learning experience. Anonymous question-asking, in particular, played a critical role in making students feel more comfortable participating, as it reduced fear of judgment and increased engagement.

Additionally, the course's design effectively reduced social anxiety by allowing students to ask questions without speaking publicly. The structured group interactions, including Q&A sessions and round table discussions, further enhanced participation and helped create a more inclusive and balanced learning environment. Other factors like instructor involvement and clear rubrics also contributed to the perceived improvement over traditional large classroom settings. Overall, the integration of online

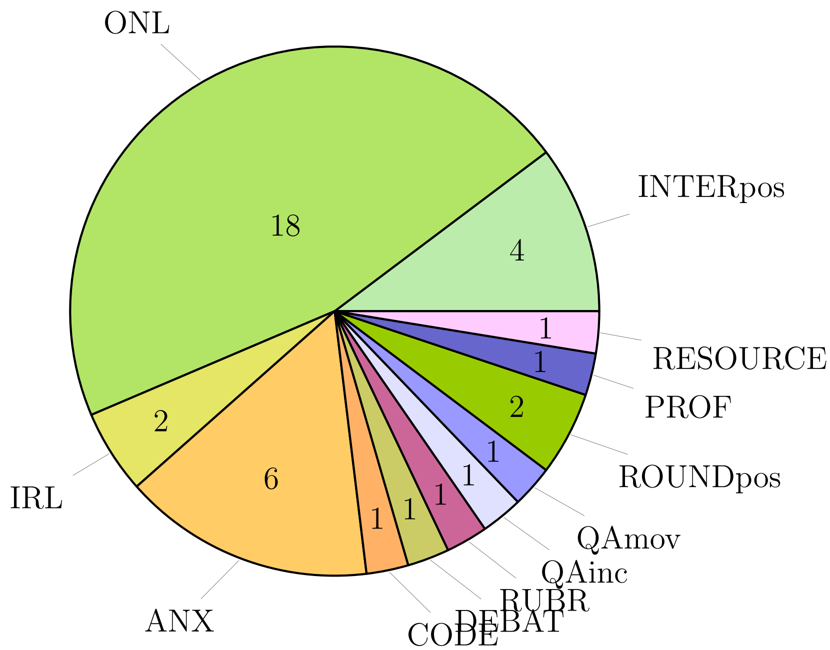


Figure 5.16: [Q2.3a] Why did you feel the group presentations were BETTER than other large classrooms?

tools and thoughtful course design were key in making the group presentations more engaging and accessible for students.

5.3.12 Question 2.3b

Question 2.3b looks at data from students who felt the group presentations were "about the same" as those in other large classrooms. See Figure 5.17. A few respondents (3 mentions) cited positive interaction during presentations, but this was not viewed as particularly distinct from their experiences in other courses. Some students (2 mentions) preferred live presentations, expressing that the pre-recorded format diminished the dynamism typically found in real-time interactions. This suggests that while the online and pre-recorded elements were appreciated by some, others felt they lacked the spontaneity and engagement of live presentations.

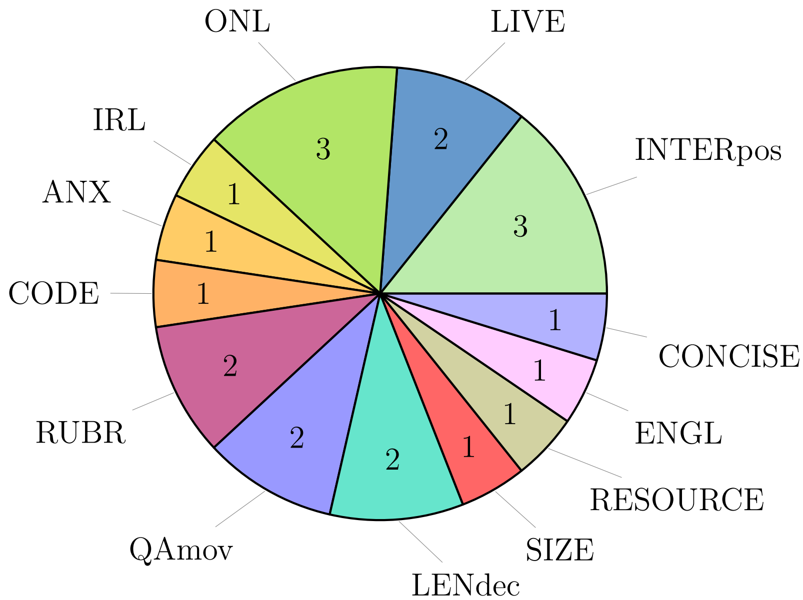


Figure 5.17: [Q2.3b] Why did you feel the group presentations were ABOUT THE SAME as other large classrooms?

Other students (3 mentions) acknowledged the value of the online format, yet did not see it as a significant improvement over traditional setups. Some respondents raised concerns about the presentation length and how it impacted their engagement. The large class size (1 mention) also remained a limiting factor, with some students feeling that this was comparable to other large lecture-based courses, where individualized attention and interaction were similarly constrained.

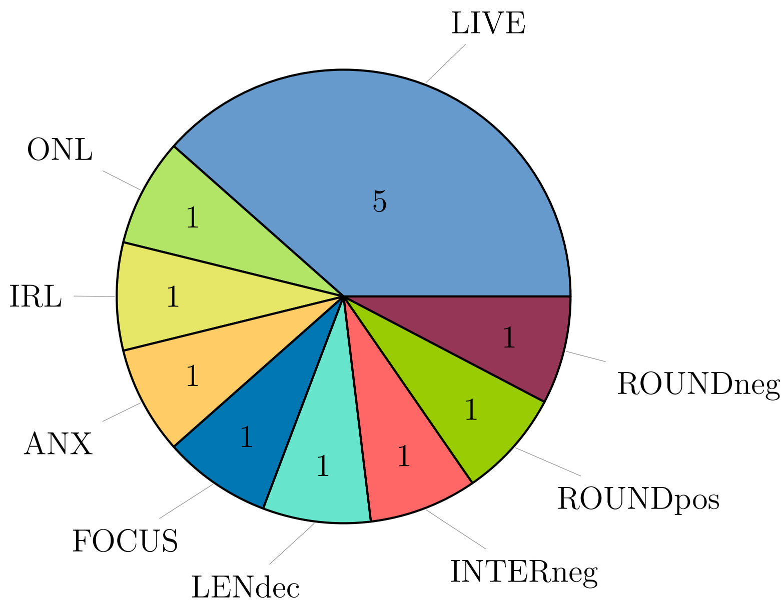
Moreover, factors like language barriers (1 mention of English proficiency concerns) and the concise nature of presentations (1 mention) were seen as common in large classrooms, reducing the distinctiveness of the experience. The overall data suggests that while the course had several positive features, many students did not perceive a significant departure from their prior experiences, viewing these as standard components of large lecture environments. However, it is worth noting several of the students in this category mentioned they did not feel they had enough university experience to respond to this question thoughtfully.

5.3.13 Question 2.3c

In **Question 2.3c** data emerges from students who felt the group presentations were "worse" than those in other large classrooms. See Figure 5.18. All students mentioned their dislike for the prerecorded format, preferring live presentations instead (5 mentions). They felt that prerecorded presentations made it harder to maintain focus (1 mention), were less engaging, and questioned the value of coming into class to watch a video. The lack of live interaction was a key factor in their dissatisfaction, as they wanted more of an in-person experience that felt dynamic and engaging.

Additionally, some students noted other concerns, such as a negative social interaction with their group (1 mention) and feeling that the structure didn't facilitate engaging discussions (1 mention). One student pointed out that the presentations didn't allow for enough focus, making it difficult to remain attentive, while another mentioned dissatisfaction with the length of the presentations. Although one respondent cited a positive aspect of the roundtable format (1 mention), another felt it was a negative, indicating mixed reactions to that particular setup.

Overall, the lack of live presentations and in-person interaction was the primary factor driving the perception that these presentations were less effective compared to traditional large classroom settings.



]

Figure 5.18: [Q2.3c] Why did you feel the group presentations were WORSE than other large classrooms?

5.3.14 Question 2.4

The data from **Question 2.4** reveals several notable challenges students face in maintaining attention both in class and during video presentations. See Figure 5.19. The most frequently mentioned issue, with 18 respondents citing it, was difficulty focusing (FOCUS). This challenge is a common one in educational settings, especially when students are required to engage with content for extended periods. A significant number of students (16 mentions) also highlighted classroom noise (NOISE) as a distraction, indicating that external factors in the learning environment can severely impact concentration, particularly in larger, less controlled spaces.

In relation to the video presentations specifically, several students (9 mentions) mentioned the availability and quality of learning resources (RESOURCE), suggesting that inadequate or poorly organized materials can hinder their ability to follow the content effectively. Some students (7 mentions) expressed frustration with the length of the presentations (LENdec), indicating that overly long or monotonous videos make it difficult for them to stay engaged. Additionally, 8 students pointed out negative experiences with the roundtable format (ROUNDneg), showing that not all aspects of the design were universally well-received, particularly when it came to maintaining attention in a video-based format.

Other challenges related to video presentations include the difficulty of staying engaged in an online environment (ONL) and the need for breaks (BREAK) to manage attention spans during longer presentations. Some students (3 mentions) expressed a preference for live interactions (LIVE) and noted that prerecorded presentations lack the spontaneity and interactivity that help sustain focus in real-time settings. Captions (CAPT) were another challenge, with 3 students noting issues related to the accessibility of the video content or perhaps difficulties in following along.

Finally, some respondents mentioned more individualized challenges such as anxiety (ANX), time management (TIMEMGMT), and issues with the pacing of the videos (PACE). These findings highlight that both environmental and personal factors contribute to students' attention difficulties and suggest that a more dynamic, interactive, and accessible format may be necessary to address these diverse needs and improve engagement with the content.

5.3.15 Question 2.5

The responses to **Question 2.5** reveal a range of strategies students use to regain focus when they encounter concentration difficulties at home and school. See Figure 5.20. The most commonly mentioned strategy, with 24 responses, was taking a break

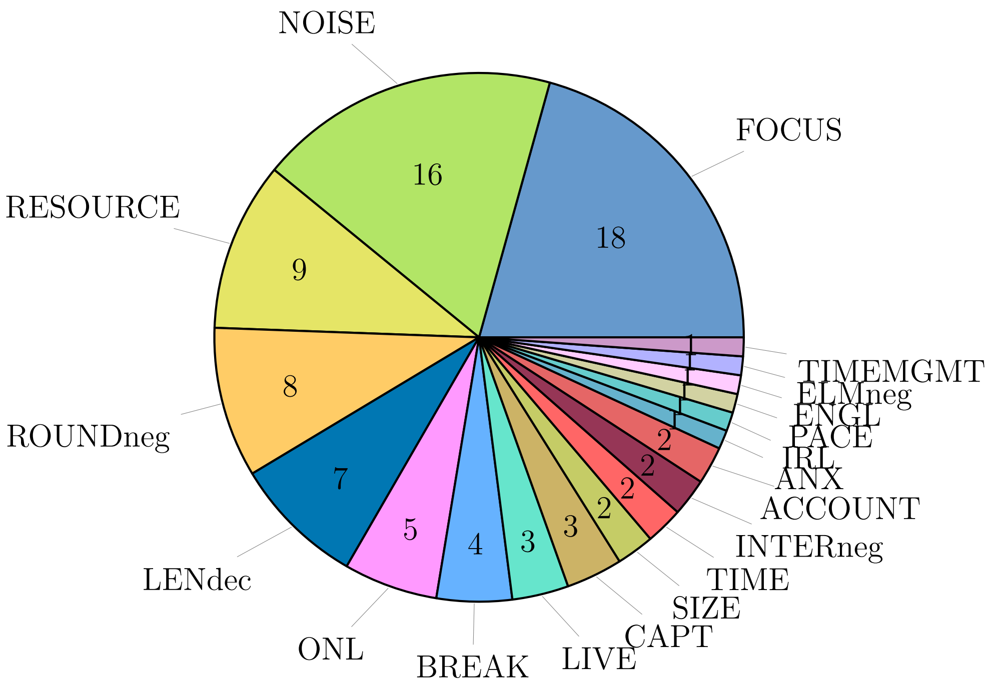


Figure 5.19: [Q2.4] What are the biggest challenges you face in paying attention in class? What about in paying attention to the video presentations specifically?

(BREAK). Several students emphasized that taking a brief pause of 5–20 minutes helps them refresh mentally before returning to their tasks. One student mentioned, “Taking a small break away from school content to relax the mind so that I can come back with a more clear mind.” Breaks appear to be a key tool in managing attention and mental fatigue, allowing students to reset before continuing their work. Designing educational software and experiences while recognizing our natural rhythms of attention helps to create a learning environment that fosters academic success and mental well-being.

Positive interaction with peers and study environments was another strategy frequently mentioned (INTERpos) with 13 mentions. Several respondents discussed the importance of accountability partners, friends, or professors who help them stay on

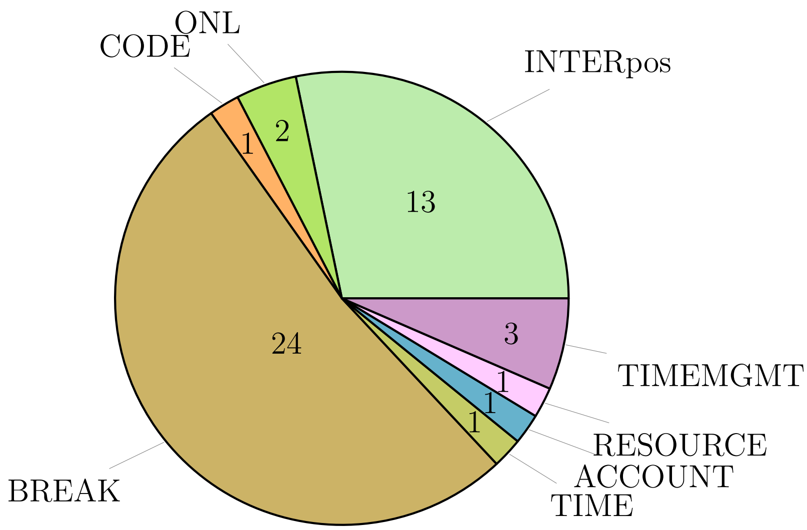


Figure 5.20: [Q2.5] When you are having difficulty concentrating with other tasks at home or at school, what do you find helps you get back on track?

track. Positive social interaction includes social engagement, working with peers, or even informal discussions that help clear mental blocks. Such interactions provide a fresh perspective or alleviate feelings of isolation during challenging tasks, contributing to improved concentration.

A few students also mentioned time management (TIMEMGMT) and online resources (ONL), indicating that setting clear schedules and using digital tools can help keep them on track. These methods may be particularly important for students juggling multiple responsibilities or who rely on external structure to maintain focus. Other methods, such as accountability (ACCOUNT) and ensuring the availability of resources (RESOURCE), highlight the importance of having external support systems and materials in place to overcome distractions and stay engaged.

In summary, the data suggests that a combination of short breaks, social interaction, and effective time management are key strategies students use to overcome attention difficulties, with breaks being the most favored method by a significant margin.

Designing educational software and learning experiences with the realities of student needs—such as the need for breaks, social interaction, and effective time management—can significantly enhance student engagement, retention, and overall success. As the responses show, strategies like taking breaks, changing study environments, and using structured scheduling systems help students regain focus and improve productivity. Acknowledging these human experiences in academia is essential to creating tools and systems that support diverse learning styles and challenges. When we design with the student in mind, recognizing their natural rhythms of attention and the benefits of peer support, we create a learning environment that fosters academic success and mental well-being, making the educational experience more enriching and sustainable.

5.3.16 Question 2.6

For **Question 2.6**, students selected from several options that could improve their ability to maintain focus during the course. See Figure 5.21. Additionally, an "other" option was provided, in which some students chose to write their suggestions. Of the three options provided, the most popular suggestion was longer breaks, with 47 students selecting this option. This highlights the importance of mental rest in sustaining attention during longer academic sessions, underscoring the need for courses to integrate adequate breaks to avoid cognitive fatigue. As seen in other responses, students often find short breaks helpful in resetting their focus, and this finding reinforces the importance of allowing students time to recharge during intensive learning experiences.

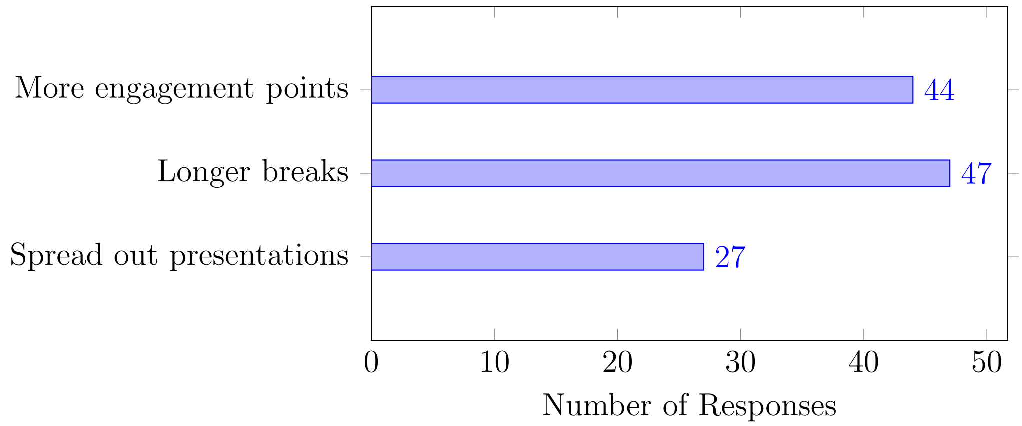


Figure 5.21: [Q2.6] Which of the following would have helped you maintain focus?

The other two suggestions, spreading out the presentations (27 votes) and providing more engagement points (44 votes), also received considerable attention. Spreading out presentations implies that students may have felt overwhelmed by the volume of content delivered in a short timeframe, which hindered their ability to maintain focus. This could indicate a desire for more time to delve deeper into the material, allowing for a more thorough exploration of each topic rather than skimming the surface of multiple subjects in quick succession. Students may have been hoping for more meaningful engagement with the content, and spacing out presentations would provide greater opportunities for reflection, analysis, and discussion.

This idea aligns with earlier feedback from students about what constitutes a good question 5.5, where they identified "depth" as an important component of their educational experience. The fact that students valued depth in their learning and wanted more time to explore subjects could be resurfacing here. It suggests a desire for more substantial engagement with fewer topics, allowing them to fully grasp the material rather than rushing through a broad range of books in a limited time. Meanwhile, more engagement points—such as interactive elements or participatory activities—could help foster active learning, ensuring students remain focused and connected to the material. These findings collectively suggest that students want better pacing and the opportunity to engage with topics more deeply, enhancing both focus and educational value.

In addition to the primary options, several students wrote custom suggestions for improving focus. Some expressed that presentations could be shorter, making it easier to stay engaged throughout. One student pointed out that live presentations might encourage more focus by reducing distractions, such as phone use, and fostering immediate interaction with group members. One student mentioned that assigning more homework could help with retention, allowing students to engage with the material outside of class in a structured manner. These additional responses suggest that designing presentations with clarity, interactivity, and brevity in mind could further enhance student focus and engagement.

5.3.17 Question 2.7

Question 2.7 looks at the location frequency of video watching. See Figure 5.22. The data reveals that many students watched the videos at home rather than in class. The majority of respondents (52 students) reported watching the videos at home "sometimes," while 23 indicated they did so "most of the time." This suggests that many students preferred the flexibility of watching prerecorded presentations on their schedule, a key benefit of the online format. This flexibility may have allowed them to manage their time more effectively or watch the videos in an environment that was more conducive to their focus and learning preferences. Additionally, students who watched the videos at home may have appreciated the ability to pause, rewind, or revisit challenging sections of the material.

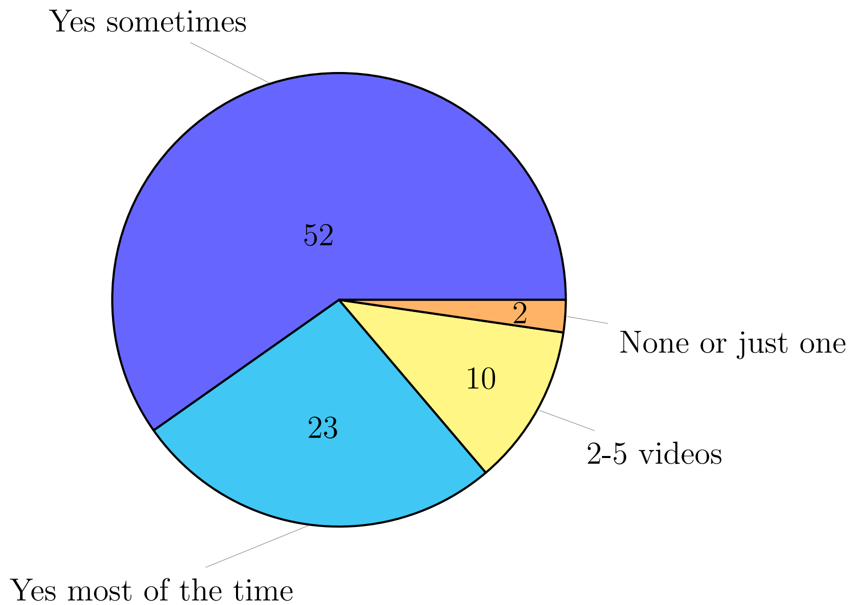


Figure 5.22: [Q2.7] Did you watch any videos at home instead of in class?

Interestingly, a smaller group of students, 10 in total, reported watching only 2-5 videos at home, and just two students watched none or only one. While most embraced the option to watch videos remotely, this minority suggests that some students still valued the in-class experience or felt they learned better in a live setting. These findings introduce potential reasons for this preference, which we delve into more deeply in subsequent questions. In the upcoming analysis, we explore student suggestions and feedback, including the desire for more structured live interactions and the benefits of prerecorded flexibility, offering further insight into the diverse learning approaches within the classroom.

One notable point from the data in question 2.7 is the presence of some contraindicated responses. While many students indicated they watched the videos at home "most of the time" or "sometimes," further along in the questionnaire, some claimed they rarely, if ever, watched the videos. This inconsistency suggests that even though students were told their grades wouldn't be impacted by their responses, the presence of their email on the survey might have influenced their answers. It's possible that some students were concerned about facing consequences for not watching the videos and may have over-reported their engagement in the course material.

5.3.18 Question 2.8 - 2.9

The responses to **Question 2.8** reveal a range of reasons that prevented students from watching the videos beforehand, with the most common being schedule constraints (46 responses). See Figure 5.23. This suggests that many students struggled to find time in their personal or academic schedules to engage with the videos outside of class hours. This finding aligns with broader challenges in post-pandemic educational experiences of time management and self-discipline. With the shift to more flexible, often hybrid learning environments, students are increasingly responsible for managing their own study schedules outside the structure of traditional classroom settings. This autonomy, while beneficial in theory, has exposed a significant struggle among students to balance their academic responsibilities with personal obligations. This finding aligns with wider trends observed in post-pandemic education, where students face difficulties organizing their time effectively, leading to procrastination or inconsistent engagement with course materials.

This course was a first-year computer science class taken in the fall semester. For most students, this means it marked their first experience with the increased autonomy that university education demands. Many students in the course were transitioning from high school, where schedules are typically more structured and closely

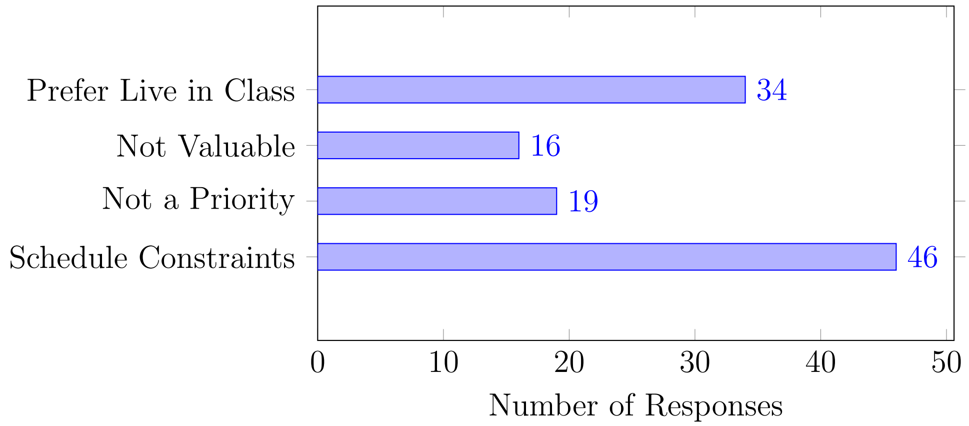


Figure 5.23: [Q2.8] What prevented you from watching the videos beforehand?

monitored by teachers. The move to a university setting—especially in a course with prerecorded videos, flexible viewing times, and independent study requirements—was a significant adjustment. This reinforces the importance of designing educational experiences that provide flexibility and support mechanisms to guide students in time management and self-regulation.

The difficulty of balancing autonomy with academic expectations is a common theme in education, but it's especially acute for students new to this environment. Without prior experience managing large blocks of independent study time, many students may have struggled to stay on top of their coursework. The need to independently organize and prioritize tasks—something not often emphasized in high school—became a considerable obstacle in staying engaged with the course materials, further emphasizing the importance of providing support structures to help students navigate this transition.

Additionally, many students (34 responses) indicated that they preferred to watch the presentations live in class rather than beforehand. This preference for live engagement could be tied to the desire for a more dynamic, real-time experience, where students feel more connected to the material. Nineteen students reported that watching the videos was not a priority for them. Sixteen students felt that the videos did not provide enough value to warrant their time.

Several custom responses provided additional insight, with some students admitting that they were either unaware that the videos were available before class or that they occasionally watched them, though not consistently. These additional responses suggest that clearer communication regarding the availability and importance of the videos might help improve engagement.

5.3.19 Question 2.10 - 2.11

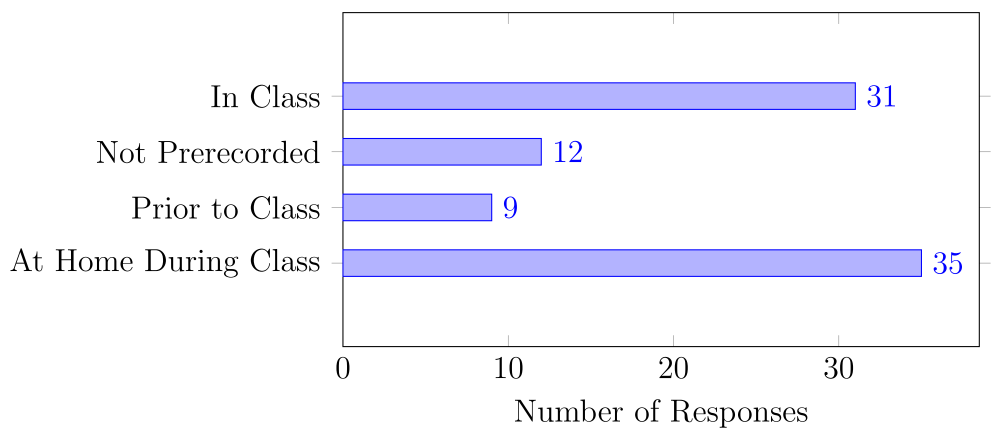


Figure 5.24: [Q2.10] What is your preference for watching videos?

Question 2.10 and **Question 2.11** examined student preferences for watching prerecorded videos. See Figure 5.24 and Figure 5.25. The largest group (35 students) preferred to watch the videos at home during the scheduled class time, while 31 students opted to watch the videos in class. Only 9 students preferred watching the videos prior to class, and 12 students expressed a desire for the videos to not be prerecorded at all, preferring live presentations instead. These preferences reveal key insights into how students engage with flexible, autonomous learning environments and suggest varying degrees of comfort with self-directed study.

For those who preferred to watch the videos at home during class time, the thematic analysis revealed that time management (7 mentions) and focus (3 mentions) were central reasons for this preference. These students likely found that being able to manage their time while still sticking to the class schedule helped them stay on track without disrupting their other commitments. Additionally, some students mentioned that watching the videos in a quieter home environment (2 mentions) allowed for fewer distractions compared to the noise often experienced in a physical classroom.

Students who preferred to watch the videos in class (31 votes) were influenced primarily by their desire for positive social interaction (7 mentions), with many feeling that learning alongside peers contributed to a better experience. Focus (7 mentions) also emerged as a key factor, as students found it easier to stay engaged in a structured environment rather than independently. Some students mentioned that the live aspect (1 mention) added value to the learning experience, making it feel more immersive. This preference for in-class learning highlights the ongoing need for structure and interaction, even in a flexible, prerecorded learning format.

Among the smaller group who preferred watching the videos prior to class (9 votes), focus (2 mentions) and time management (2 mentions) were the two key drivers. These students seemed to favour a proactive approach, where viewing the material in advance allowed them to engage in discussions more meaningfully during class time.

Finally, those who did not want prerecorded videos at all (12 votes) cited both focus (2 mentions) and the importance of interactive learning experiences (2 mentions)

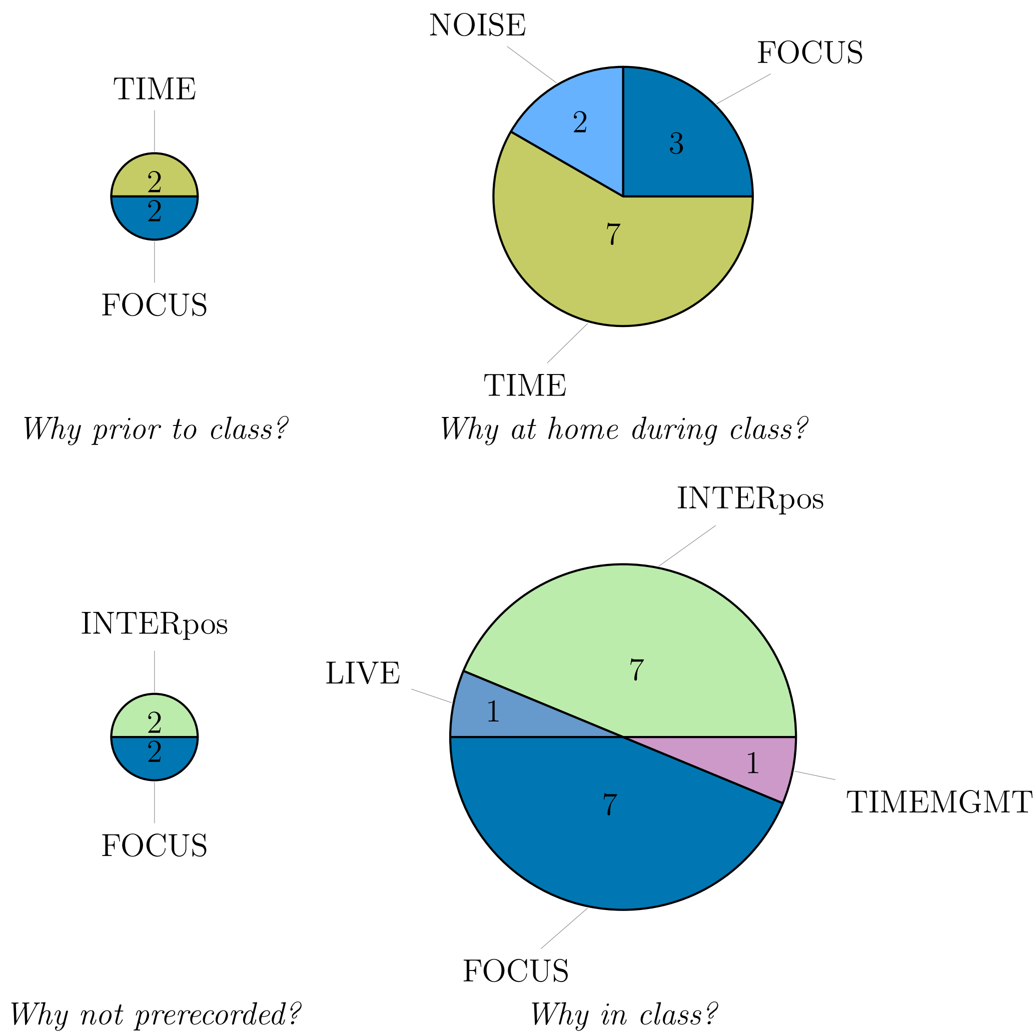


Figure 5.25: [Q2.11] Why is that your preference for watching videos?

as reasons, indicating that some students feel more engaged and accountable when content is presented live rather than asynchronously.

This variety in preferences underscores the need for course designs that accommodate diverse learning styles and provide students with multiple ways to engage with the material.

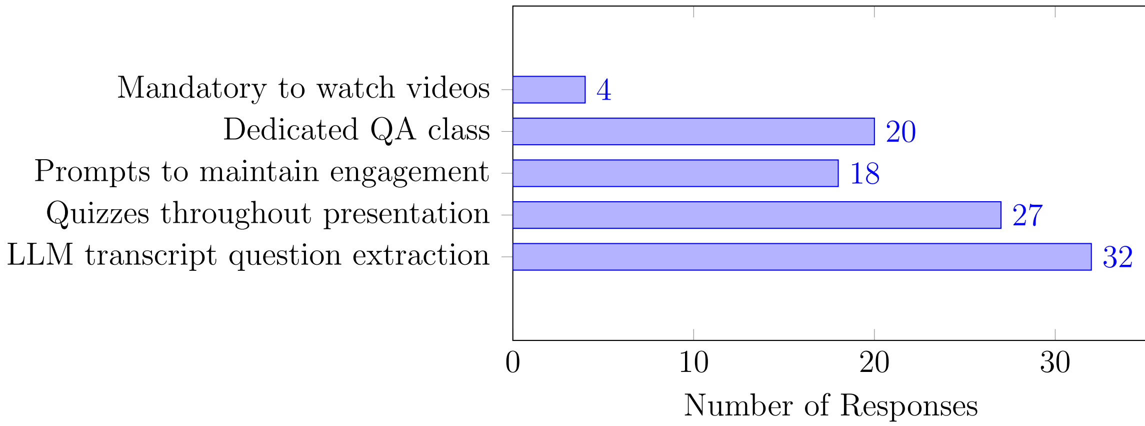


Figure 5.26: [Q2.12] What of the following would you like to see in a future iteration of the way presentations and questions are handled in a large class setting?

5.3.20 Question 2.12

Question 2.12 explored features students would like to see in the next iteration of the prototype. See Figure 5.26. The most popular suggestion, with 32 votes, was the implementation of an LLM (large language model) transcript extraction feature. This would allow for questions asked at each table to be transcribed and extracted from live conversations, reducing the need for students to manually type out questions. This highlights a desire for technology that can enhance real-time engagement without adding extra effort on the student's part, which could also improve participation by allowing students to focus on the discussion rather than multitasking. This also suggests that students are open to automation in facilitating discussion and highlights that they don't mind the idea of having an AI tool listening to their table discussions. Privacy, surprisingly, wasn't a major concern here, which is an interesting note considering the current discourse around surveillance in educational technologies.

Quizzes embedded throughout the presentations were another popular option, with 27 students endorsing this feature. This finding further reinforces that students see value in having regular check-ins to reinforce comprehension and engagement. On a similar note, students proposed interactive prompts to maintain attention, such as requiring them to click through the presentation (18 votes). This aligns with the need for active learning strategies that help mitigate distractions.

Following closely behind, 20 students suggested having a dedicated question-answering or presentation-watching class. This structure could help streamline the engagement process by providing dedicated time to focus on either content delivery or clarification. The idea of requiring students to watch the videos beforehand, with class time dedicated solely to answering questions, garnered less support, with only 4 votes. This suggests that while some students may benefit from this approach, the majority might struggle with time management or prefer more flexible viewing options.

Additionally, students offered several written-in suggestions that echoed the broader themes of engagement and flexibility. One suggestion emphasized the importance of offering multiple ways for students to engage with presentations, suggesting that personalization and autonomy are key components of effective learning. Some students advocated for smaller group discussions after presentations, rather than large-class discussions, indicating a preference for more intimate, focused interactions. Others wanted shorter videos and fewer questions, as well as a mix of live questions and typed questions to promote a more dynamic and responsive learning environment. These suggestions underscore the importance of designing educational tools that cater to diverse preferences and promote deeper engagement through both technology and social interaction.

The data gathered here suggests students value interactive and participatory features that make the learning experience more engaging and streamlined. Whether it's through automation, live interaction, or increased accountability through quizzes, future iterations of this prototype should aim to balance flexibility with the need for consistent student participation and engagement. This feedback offers clear directions for future improvements, keeping student preferences and engagement at the forefront of design.

5.4 Presentation of Pre-Agreement Data

5.4.1 Kappa and Accuracy Tables by Code

The tables in this section 5.1, 5.2, 5.3 focus on the pre-consensus data, offering insight into the raw inter-rater reliability scores, including Kappa and accuracy scores for each code. For more information about the methods used see Chapter 3.

A Kappa score of 1 indicates perfect agreement, 0 represents agreement that is no better than chance, and negative scores suggest systematic disagreement. Accuracy score, in contrast, reflects the percentage of agreement between the raters without adjusting for chance.

The tables display data for all codes excluding codes that showed perfect disagreement or were not used at all. Perfect disagreement occurred when one rater applied a particular code to a response, while the other rater did not apply the code at all. This discrepancy was often due to the complexity of managing 39 distinct codes, where a rater might overlook a specific code's existence during the initial blind coding phase.

Negative Kappa scores emerged in cases where both raters applied a code, but never on the same question, indicating a lack of shared understanding or criteria for that specific code.

This data also serves as a baseline for assessing how consensus discussions improved inter-rater reliability, showing the effectiveness of collaboration in the coding process. Including this pre-consensus data ultimately adds transparency to the study's findings by ensuring that any coding inconsistencies are addressed and accounted for.

By examining this pre-consensus data, we gain a clearer picture of the challenges involved in achieving consistent coding, especially when working with a complex coding scheme like the one in this study. The data highlights areas where adjustments and discussions were most necessary to reach consensus.

	Code	Kappa	Accuracy
Question 1.1	INTERpos	0.8321783105	92.1%
	ONL	1	100%
	IRL	0.3784530387	94.1%
	ROUNDpos	0.648721921	96.1%
	PROF	0.6856677524	86.2%
	OUTCOME	-0.0221987315	91.4%

	Code	Kappa	Accuracy
Question 1.2	INTERpos	0.905707196	99.3%
	ANX	0.4735915493	93.4%
	FOCUS	0.3077513431	92.1%
	RUBR	0.2323232323	92.1%
	INTERneg	0.7967914439	99.3%
	ROUNDneg	0.4950166113	98.7%
	PROF	-0.01732673267	92.8%
	ELMneg	0.4898819562	97.4%
	PACE	0.7091346154	91.4%
	ANON	0.4950166113	98.7%
	SIZE	0.3294367694	88.8%
	PREREQ	0.3259423503	97.4%
	RESOURCE	-0.01178451178	89.5%
	ENGL	0.7967914439	99.3%

	Code	Kappa	Accuracy
Question 1.3	INTERpos	0.793877551	98.7%
	ANX	0.5302901748	75.0%
	ROUNDpos	1	100%
	PACE	-0.005543237251	96.1%
	ANON	0.190412783	94.7%
	SIZE	0.5769694534	80.9%
	PREREQ	0.4950166113	98.7%
	ENGL	0.793877551	98.7%

Table 5.1: Pre-Agreement Kappa and Accuracy Scores by Code for Questions 1.1 to 1.3, including codes which were identified at least once and did not have 100% disagreement within a question.

	Code	Kappa	Accuracy
Question 1.4	INTERpos	-0.005917159763	94.1%
	DEEPEN	0.4251831182	69.1%
	CONCISE	0.7242647059	94.7%
	RELEVANT	0.3048600884	83.6%
	COMMON	0.5525204867	84.2%

	Code	Kappa	Accuracy
Question 2.3a	INTERpos	0.6421568627	94.6%
	ONL	0.3232876712	62.2%
	ANX	0.6858789625	91.9%
	DEBAT	1	100%
	PROF	1	100%

	Code	Kappa	Accuracy
Question 2.3b	INTERpos	0.7832167832	96.8%
	LIVE	0.6516853933	96.8%
	ONL	0.3673469388	90.3%
	ANX	1	100%
	RESOURCE	1	100%

	Code	Kappa	Accuracy
Question 2.3c	LIVE	0.75	87.5%
	FOCUS	0.6	87.5%
	ANX	1	100%
	ROUNDneg	1	100%

Table 5.2: Pre-Agreement Kappa and Accuracy Scores by Code for Questions 1.4 to 2.3c, including codes identified at least once and did not have 100% disagreement within a question.

Question 2.4

Code	**Kappa**	**Accuracy**
LIVE	0.7932011331	98.6%
ONL	0.4186795491	93.2%
FOCUS	0.4615877081	83.6%
LENdec	0.333767927	90.4%
BREAK	1	100%
TIME	0.6604651163	98.6%
INTERneg	0.4895104895	97.3%
ROUNDneg	-0.01264044944	86.3%

Question 2.5

Code	**Kappa**	**Accuracy**
INTERpos	0.12109375	84.0%
ONL	0.4897959184	97.3%
BREAK	0.8390455531	93.3%

Question 2.11a

Code	**Kappa**	**Accuracy**
FOCUS	0.5294117647	90.6%
TIME	0.8181818182	93.8%

Question 2.11b

Code	**Kappa**	**Accuracy**
FOCUS	1	100.0%
TIME	0.6956521739	85.7%

Question 2.11c

Code	**Kappa**	**Accuracy**
INTERpos	0.7824074074	91.7%
FOCUS	0.5681818182	79.2%

Table 5.3: Pre-Agreement Kappa and Accuracy Scores by Code for Questions 2.4 to 2.11c, including codes identified at least once and did not have 100% disagreement within a question.

Chapter 6

Proposed Design: Interaction.ai

This chapter presents the second prototype design of an interactive classroom tool, Interaction.ai, developed in response to extensive feedback gathered from initial testing, interviews, survey data, and discussions outlined in the previous chapter. The design focused on improving engagement, cultivating positive social interaction and improving educational outcomes while minimizing the burden on the instructor. Detailed wireframes of each interface within the system, capturing both student and presenter modes, are delineated to illustrate the functionality of user interaction flows. This design iteration aims to balance between technological innovation and practical usability, adding features not simply because they are possible, but because they meet needs identified during the first prototype feedback collection period.

6.1 Design Goals

In this subsection, we outline the key design goals achieved by Interaction.ai. Each goal is crafted to address specific challenges within large classroom settings, enhancing both teaching and learning experiences.

1. **Enable Active Student Participation:** Create tools which allow every student to participate in large classes where students present course material. Change the current expectation (many students can only rarely be heard, so students listen passively) by lowering the barrier to participation through technology. In a conventional lecture, real-time polls and quizzes are used, but how should this be adapted for a more engaged classroom? This design goal should respond to the number one trend in user feedback: anxiety preventing them from fully participating in large classes.

2. **Improve Group Dynamics:** In large classes in which instructional staff cannot monitor, let alone facilitate discussions in every group, use technology to create a license to and expectation to participate in group discussions. For example, a digital version of a "talking stick" could not only identify the current speaker, but give feedback on time per speaker and ensure that everyone gets equal opportunity to speak. This design goal should respond to the second trend in user feedback: the desire to share ideas and discuss reactions to class material within a known peer group.

3. **Minimized Instructor Workload:** Having observed that LLaMA could curate questions nearly as well as the instructor, and that participation in reflection through posing questions was very high, use technology to automate administration of question curation and other tasks which give every student a chance to participate. For example, in addition to curating to achieve depth and breadth in questions, memory of previous choices can be used to ensure

that every student has a great chance of at least some of their questions being used in class. This design goal responds to the feedback from students that question curation (e.g., to encourage deepening of knowledge) is very desirable, and feedback from instructional staff that manual curation is resource-intensive.

4. **Instant, Actionable Feedback** Use technology to gather anonymous feedback on the clarity and interest of each part of their presentation so that it can be iteratively refined. This goal responds to the common desire for increased discussion time, which is only achievable by shortening presentations. Instructors believe this can be facilitated by focusing presentations on the most relevant and most clearly presented material.

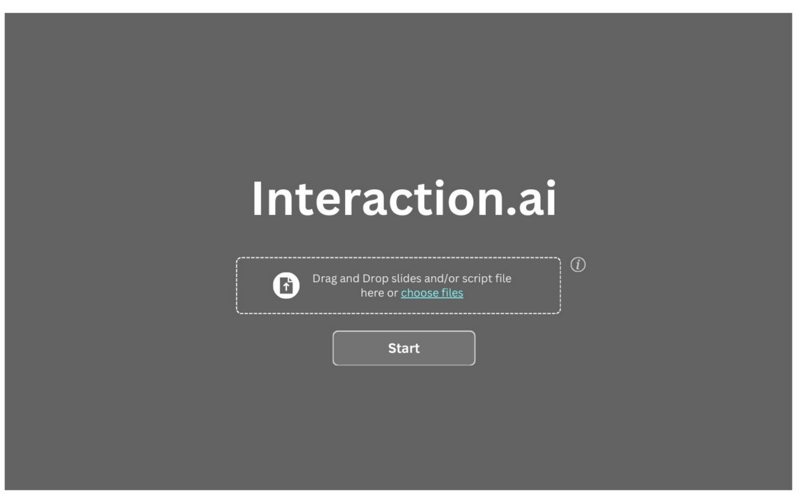
6.2 Wireframes

The proposed design of Interaction.ai includes detailed wireframes of each interface within the system, specifically tailored for both student and presenter modes. These designs are meant to showcase the functionality of user interaction flows, ensuring that the tool is intuitive and effective across various educational scenarios.

The application is designed for versatility in educational delivery, accommodating both synchronous and asynchronous learning environments. It is well-suited for live lectures as well as for settings where presenters prefer to use prerecorded presentations. This flexibility makes it an ideal resource not only for university classrooms but also for workshops and seminars aimed at boosting engagement and interaction among participants. Whether facilitating a live seminar or providing a platform for students to interact with prerecorded lecture content, this app supports a range of educational activities, including group discussions and tutorial sessions. This adaptability ensures that the tool can enhance learning experiences in diverse educational contexts, from formal university settings to more informal learning workshops and seminars.

Figure 6.1 presents a comprehensive user flow diagram for Interaction.ai, displaying the array of interactive options available to users within the application. This diagram serves as a visual map of the entire user journey, detailing each possible action and decision point across different user roles, including students and instructors. While individual screens and their functionalities will be depicted and discussed separately below, this user flow diagram provides an overarching view of how users navigate through the application

Upon launching the app, the user is met with the below landing screen. It prompts users to upload their presentation slides and accompanying script files by either dragging and dropping them into the designated area or using the file selection dialog. This interface is designed to ensure ease of use, allowing users to quickly set up their presentation.



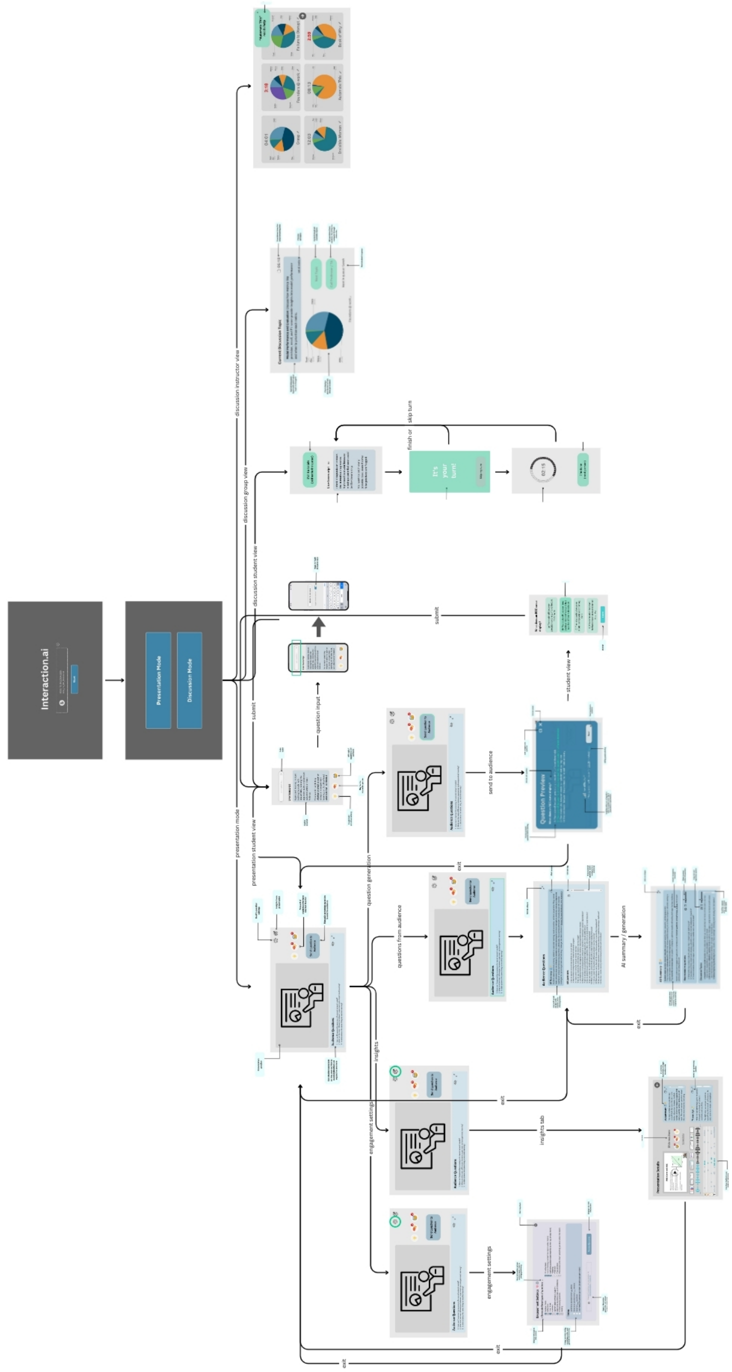
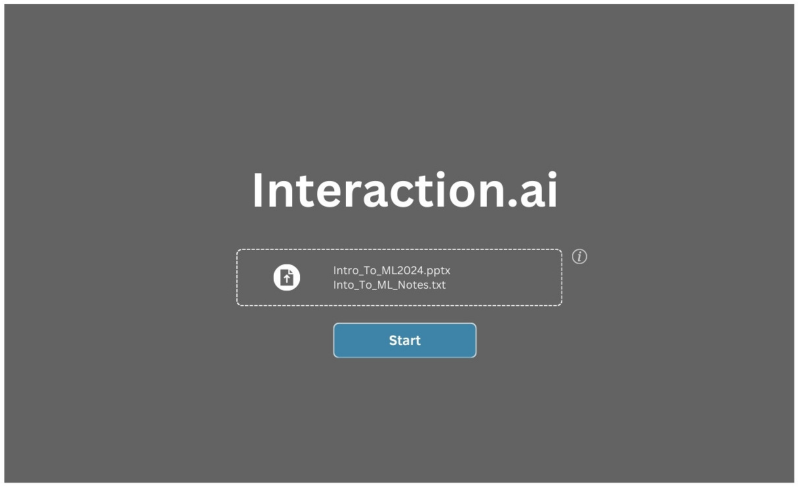
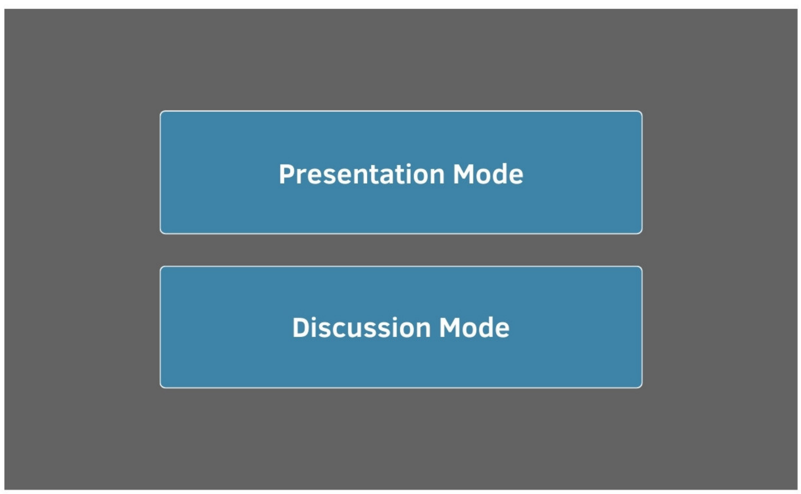


Figure 6.1: User flow diagram

Once the files are uploaded, the option to start is unlocked.



Afterwards a screen appears allowing users to select their desired mode.



In 'Presentation Mode,' instructors can initiate a lecture-style session that leverages AI for question analysis, sentiment capturing, and instant engagement generation, such as context-driven questions aimed at increasing class interaction. This mode also provides a detailed analysis post-lecture to assess overall engagement and comprehension. Conversely, 'Discussion Mode' is designed for interactive sessions,

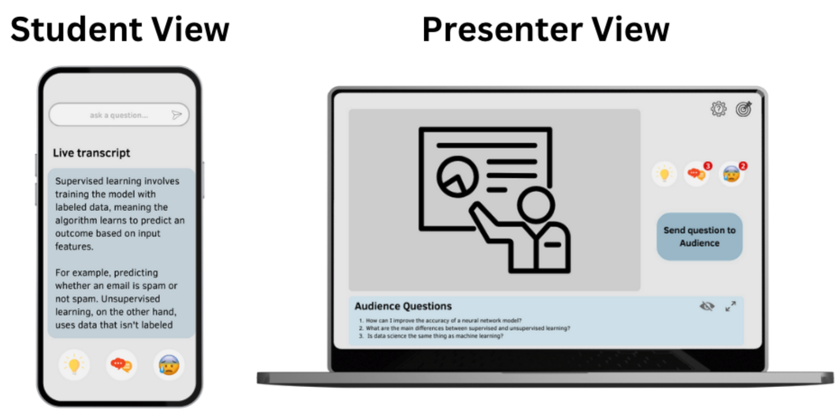


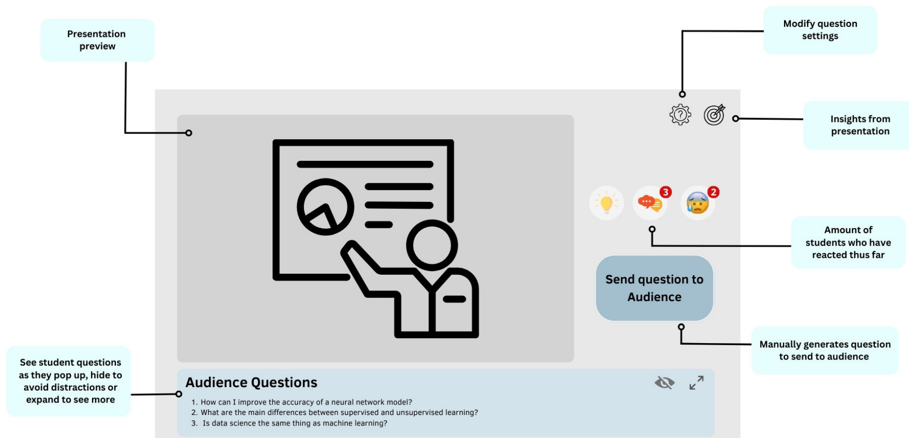
Figure 6.2: Student and instructor user interface for presentation mode.

promoting in-depth discussion and collaborative learning with tools like the digital talking stick for regulated turn-taking, a countdown timer, and graphs that monitor and visualize student talk time. These features collectively ensure a dynamic and responsive learning environment tailored to the needs of both instructors and students.

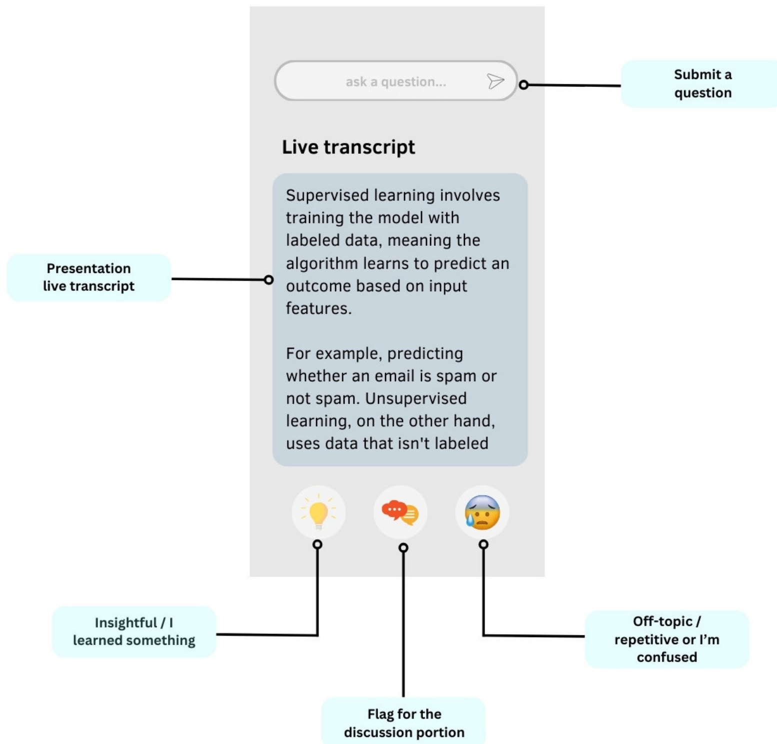
6.2.1 Presentation Mode

Figure 6.2 showcases the dual-view functionality of the application in Presentation Mode for the presenter and the participant.

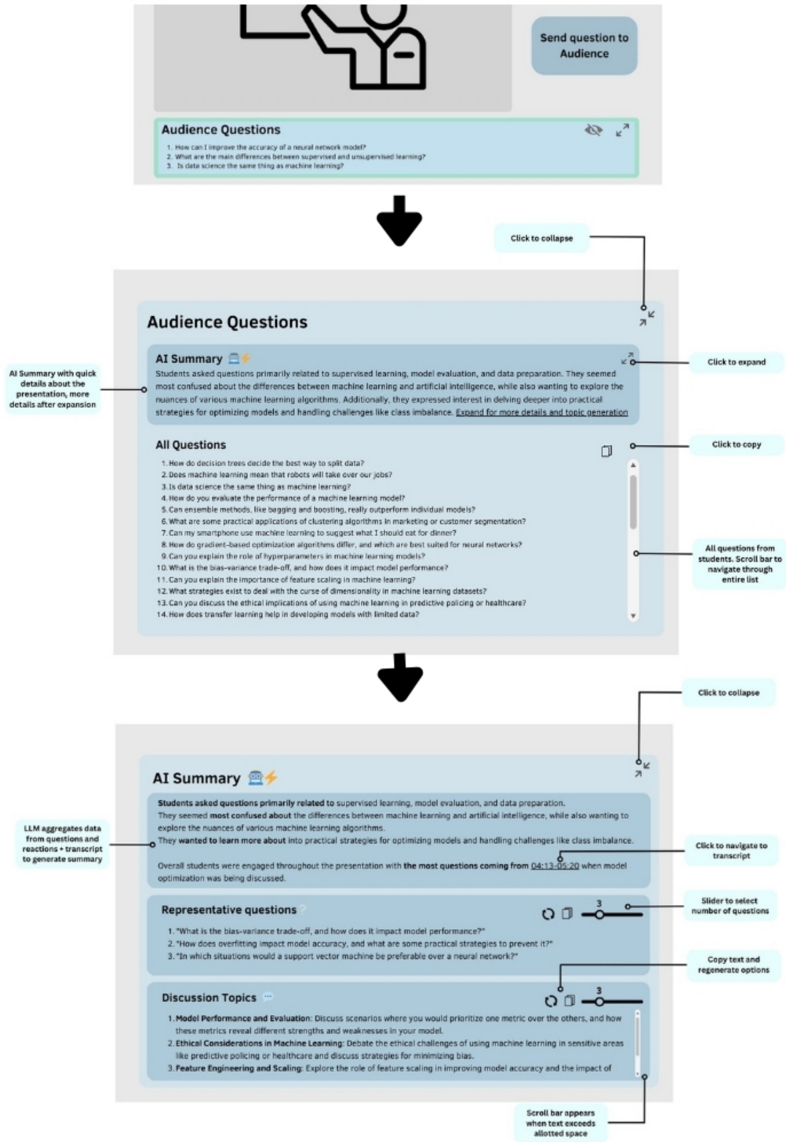
The main display of the presenter view for presentation mode acts as a dashboard, allowing the user to monitor real-time reactions and audience questions. It provides the option to instantly generate context-specific questions to stimulate engagement. These features are customizable to meet the presenter's requirements, and comprehensive feedback is available post-presentation through the insights tab. These features allow instructors to adapt their delivery based on immediate feedback and boost interaction with their students. More details about each of these features with specific screenshots to follow.



The student view showcases a streamlined interface designed around the most frequently requested features. Students can access a live transcript (a feature especially beneficial for those who face challenges with English comprehension), ask questions anonymously, and provide one-click feedback on the presentation. The feedback is designed to highlight specific segments as either particularly engaging, confusing, or worthy of discussion. This design significantly enhances the educational experience by facilitating easier navigation and enabling real-time interaction and targeted feedback. The interactive setup not only improves comprehension but also actively involves students in the learning process, creating a dynamic environment which engages students.



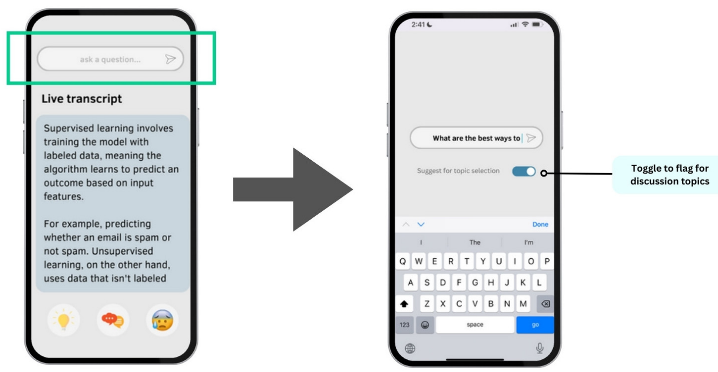
The audience collected question feature is a revised version of the main feature tested in the first prototype.



In the survey data, students reported a higher propensity to ask questions when afforded anonymity. This preference is largely driven by heightened social anxiety, a trend that has escalated post-pandemic [24, 30]. Reflecting on feedback from our instructor interviews, it became clear that instructors and presenters needed a quick low-effort way to make use of these student questions. Whether a presenter receives just a handful of questions or hundreds in a brief period, the system is equipped to manage both extremes. For the latter instance, the interface allows for the quick selection of a representative sample of questions, enabling presenters to address them efficiently within limited time frames. For example, if an instructor had 5 minutes at the end of lecture to answer questions, they could ask the system to pick one or two of the most representative questions that will aid in deepening understanding of the lecture content. The selection criteria used by the large language model (LLM) to identify these questions is based on survey data we collected from first year students asking them to define what constitutes a great question.

The presenter can use the AI-generated summary to quickly grasp the primary concerns and topics students are most curious or confused about. This summary aggregates the questions raised by students and highlights recurring themes or areas of particular interest or confusion. Presenters can quickly identify key points of discussion, tailor their responses more effectively, and ensure that they address the most pertinent issues raised during the session. This feature not only saves time but also enhances the relevance and impact of the presenter's follow-up, facilitating a deeper connection with the audience and fostering a more interactive and responsive educational environment. This aligns with another significant theme from the data: positive social interaction. These interactions are identified by students as the most important factor in the university classroom experience (See Figure 5.2). By facilitating tailored responses and focused discussions based on the AI summary, presenters can enhance these interactions by easily cultivating responsiveness.

The large language models act as a tool to analyze and interpret the questions within the context of the ongoing presentation or lecture, providing presenters with actionable insights with an emphasis on speed. For more in-depth insights to help revise future iterations of the presentation, or inform gaps before the next lesson, the insights tab is particularly useful. However, this current question analysis feature excels in summarizing only the audience questions which can be helpful during the presentation.



Students have a simple interface to ask questions with just a textbox and a toggle to flag the question as a discussion topic for presentations with a question and answer or tutorial component. The LLM takes these into consideration when picking discussion topics.

The question generation module in this application addresses a common challenge in large classroom settings: maintaining student engagement during extended lectures. Traditional transmission-style teaching often fails to involve students actively and leads to reduced learning outcomes [52], especially during lengthy sessions without breaks or interactive checkpoints. These features help transition from traditional transmission models (where information delivery is unidirectional) to a more interactive, constructivist approach.

Upon pressing the “Send question to Audience” button, shown in Figure 6.3, this feature displays a pop-up with a preview of the question to be sent to the audience.

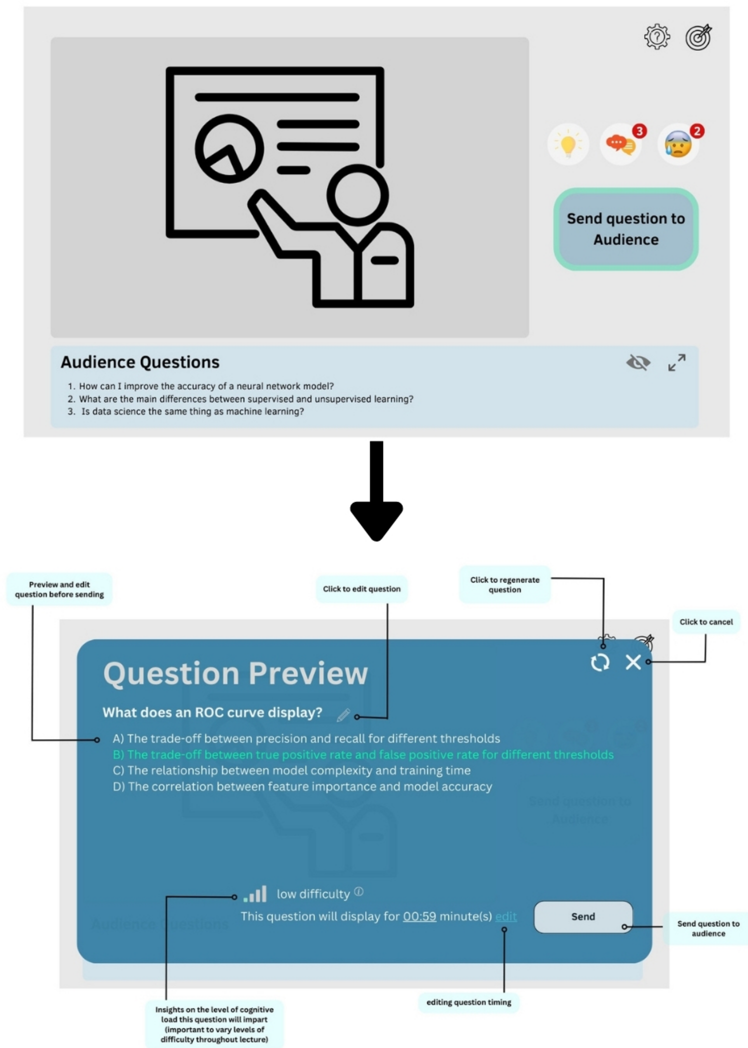
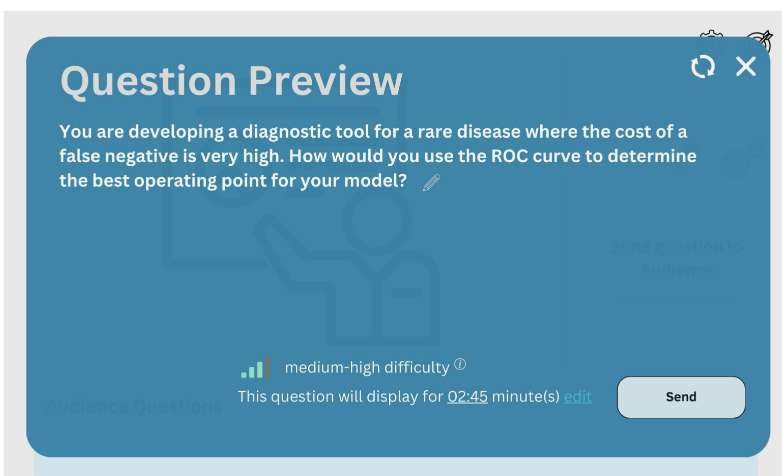


Figure 6.3: Question generation flow

This question can be automatically generated based on the content displayed at the current point in the slides or script. If the question does not meet the presenter's expectations, there is an option to regenerate the question. This allows for the selection of various question types, such as summary, multiple choice, lateral thinking, brain dump, and more depending on the customization specified in the engagement settings. See the engagement settings flow Figure 6.4 for more details on this.

The presenter can also edit the questions or the choices provided in multiple-choice formats with the pencil icon. This adaptability ensures that the questions align well with the lecture's objectives and the audience's comprehension level. It also provides an additional level of protection from GenAI hallucinations.

The interface offers insights into the cognitive load or difficulty level of each question, aiding presenters in maintaining a balanced cognitive challenge throughout the lecture. This feature recognizes the importance of varying the complexity of questions to optimize student engagement and learning outcomes.



Additionally, presenters can adjust the duration for which each question is displayed, providing flexibility to accommodate different discussion lengths and depths.

Once finalized, the question can be dispatched to all students, facilitating real-time interaction and engagement.

This module not only supports the seamless integration of interactive questions into presentations but also empowers educators to tailor their instructional approach. By enabling easy adjustments and offering a range of question types and difficulty levels, it enhances the educational experience, making lectures more engaging and responsive to student needs. This module leverages insights from educational theory to intersperse higher-order and lower-order cognitive questions throughout lectures, ensuring cognitive engagement without overwhelming students. For instance, while a continuous "brain dump" after each slide might lead to student fatigue, strategically placed questions can enhance retention and understanding.

Feedback from instructors highlighted a need for more manageable integration of quizzes and questions to enhance learning and maintain engagement. The interviews also unveiled a need for a product like this as existing university tools were deemed too cumbersome and time-consuming—especially given their other responsibilities.

Users can customize the types of questions generated—ranging from multiple-choice to more complex problem-solving queries—according to the lecture content and their pedagogical goals. The system intelligently varies question types based on user preferences and the structure of the lecture material.

To counter potential inaccuracies or hallucinations common in generative AI, this module focuses on question generation rather than evaluation. This approach harnesses AI's capability to reduce administrative burdens while fostering intellectual engagement.

Additionally, timing settings allow instructors to automate question generation at

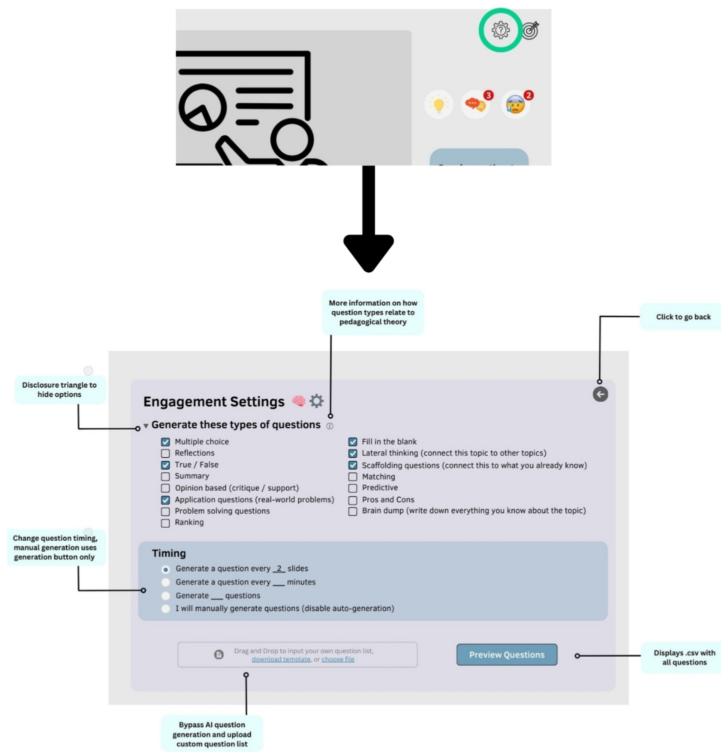
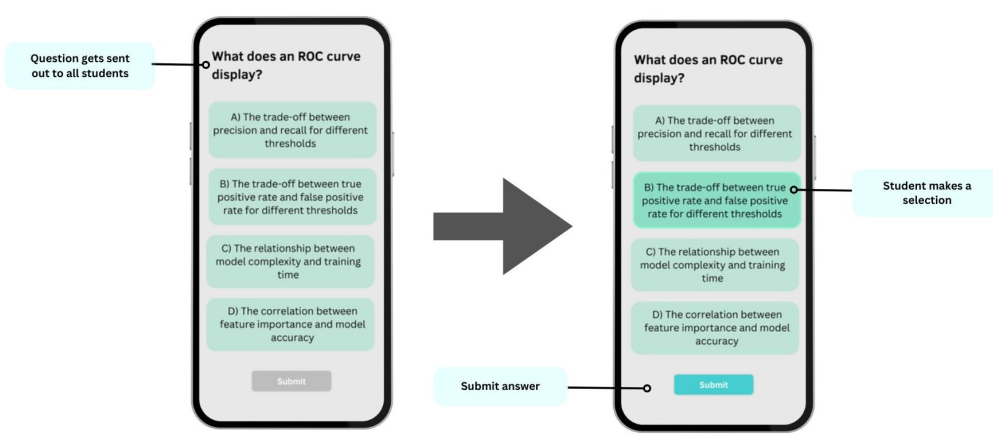


Figure 6.4: Engagement Settings Flow

specific intervals, suitable for those who might not remember to manually introduce questions during a lecture. Alternatively, presenters can initiate questions manually, aligning with natural breaks or transitions within the lecture flow.

For those who prefer to prepare in advance, the system supports the uploading of pre-vetted questions. A downloadable template guides users on how to format these questions effectively. Before each session, instructors can preview and reorder the auto-generated questions based on the uploaded slides and scripts, ensuring they align with the lecture's learning objectives and flow. This flexibility makes the Engagement Settings module a powerful tool for enhancing interactivity and learning in large-scale educational settings.

In the student interface, the design adapts to match the type of question being posed, ensuring a seamless user experience. The example below illustrates a multiple-choice question interface where students can select their answer and submit their response. For questions that require a written reply, the interface transforms to display a long-format text box, facilitating straightforward submission of longer answers. This adaptive interface design ensures that students can focus on the content of the questions without navigation difficulties.



The Insights module shown in Figure 6.5 serves as a comprehensive feedback mechanism for educators and presenters, enabling a deeper analysis of their lectures or presentations. This module synthesizes data from various sources to provide actionable insights, making it an essential tool for continuous improvement in educational settings.

Upon completion of a presentation, this module aggregates data from pre-uploaded slide content, real-time transcripts (which may diverge from the planned script), student reactions, and questions posed during the session. By analyzing these diverse data streams, the Insights module identifies key areas where attendees felt confused or found the content particularly enlightening. It also examines the timing and context of questions to pinpoint gaps in understanding and topics that may require further elaboration.

A unique feature of this module is its ability to provide both macro and micro-level feedback. Presenters can view aggregate reaction data to gauge the overall engagement and comprehension of their lecture, or they can drill down to specific sections. By selecting a particular slide or segment of the transcript, users can access detailed insights related to that portion of the presentation. The AI-powered summary updates dynamically, offering specific feedback on points of confusion and interest, along with recommendations for content adjustment or additional coverage areas.

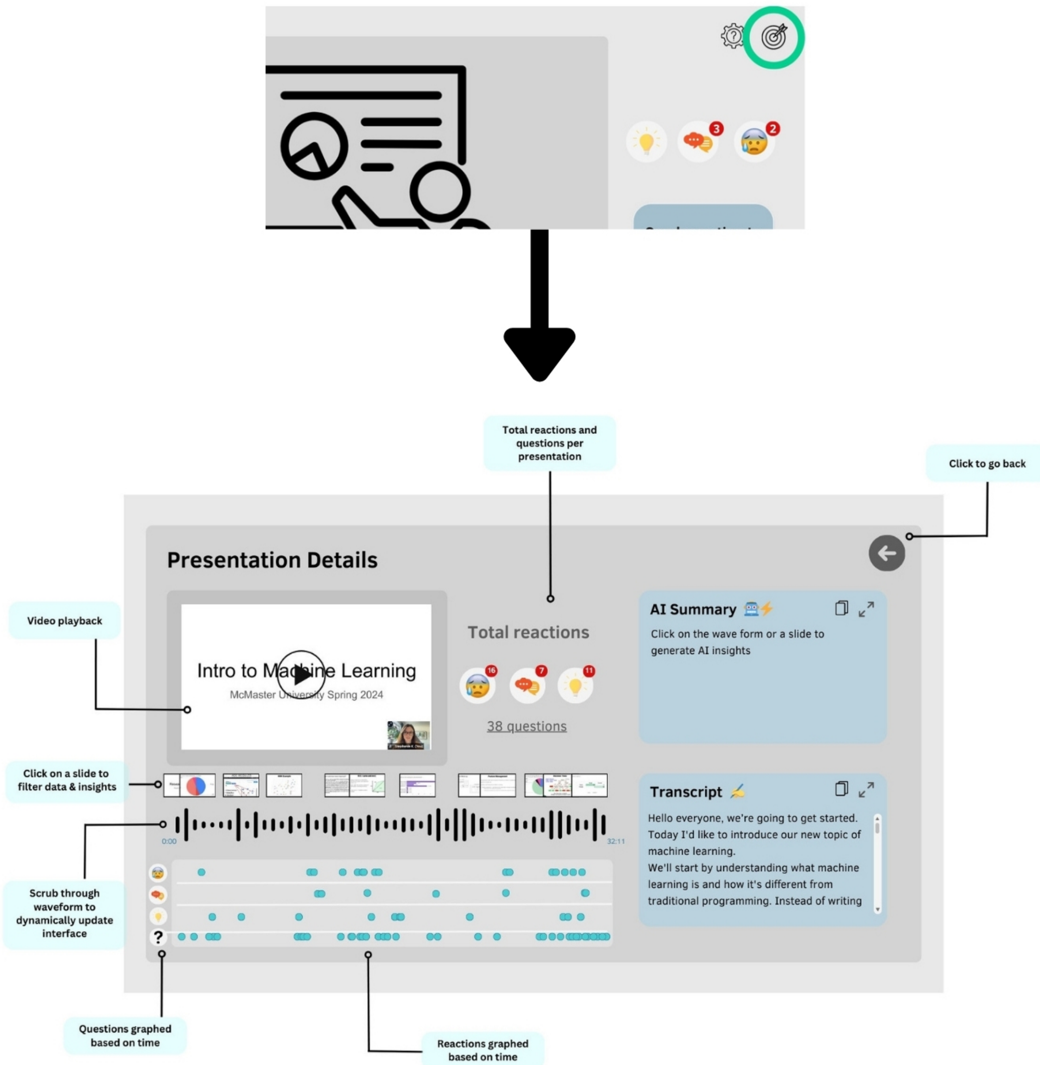
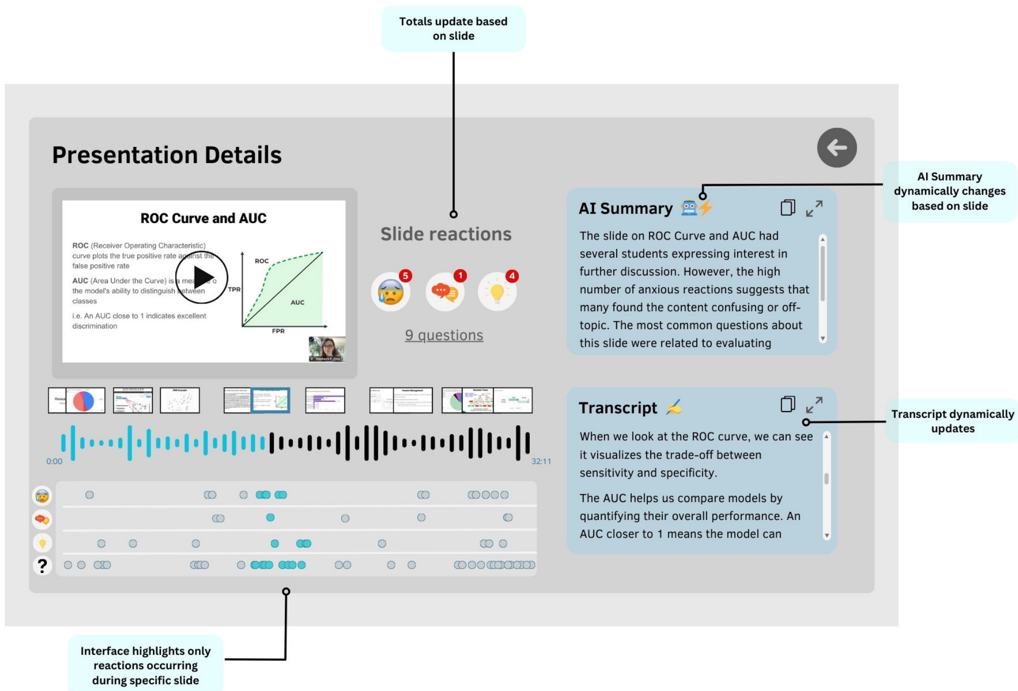


Figure 6.5: Insights Flow



This granular approach allows educators to tailor their teaching strategies more effectively, addressing specific learner needs and curiosities. It also aids in the strategic planning of future lessons, ensuring that subsequent presentations are more aligned with student learning requirements and interests. Overall, the Insights module transforms passive lecture delivery into an interactive, adaptive learning experience that enhances both teaching efficacy and student engagement.

6.2.2 Discussion Mode

Discussion Mode in the Interaction.ai platform is designed for settings including workshops, Q&A sessions, and tutorials. This mode strategically addresses the needs for positive social interaction and equitable participation, identified through extensive feedback and data analysis from previous user surveys and interviews. Each interface will be explored in detail, illustrating how they cater to the distinct needs of students, groups, and instructors. This section will cover the functionalities designed to support a variety of discussion formats, whether in academic settings or more informal educational workshops. Each mode is crafted to optimize interaction, encourage equitable participation, and manage the dynamics of group discussions, ensuring that all voices are heard and valued in the learning process.

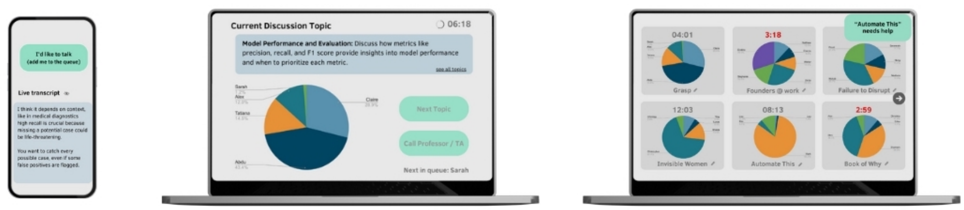


Figure 6.6: left student view, middle group view, right instructor view

In Student Mode, the application maintains essential features from Presentation Mode, including the live transcript. In catering to a diverse range of students, it was important to design for their unique needs and preferences. For this feature we're designing for the student who likes having a transcript to increase comprehension, and we're also designing for the student who gets easily distracted and wants a minimal interface – so we have the ability to hide the transcript. One of the additions of discussion mode's student view is the digital talking stick, a feature designed to alleviate social anxiety by encouraging equitable low barrier to entry participation. This tool not only enables students to signal their willingness to contribute by 'raising' their virtual hand but also includes the option to pass, accommodating students who may prefer not to speak at that moment.

The digital talking stick system is integrated with a countdown timer that governs the duration of each student's turn to speak. This ensures that speaking time is distributed fairly among all participants, a feature rooted in our commitment to fostering a balanced and inclusive dialogue environment. Some of the comments from the students that dictated negative experiences with their peers included group members talking too much or not participating at all – both scenarios this feature aims to address. The timer helps manage the flow of conversation, preventing any single participant from dominating the discussion, thereby aligning with our data-driven insights that emphasize the importance of fair and meaningful interaction. The Student Mode interface is designed to keep students engaged but not overwhelmed with features that support a structured yet flexible discussion format. Catering to the diverse needs of students and directly impacts their satisfaction with their educational experience. These elements collectively ensure that each student can engage in the learning process in a manner that respects their comfort levels yet encourages active participation with the material.

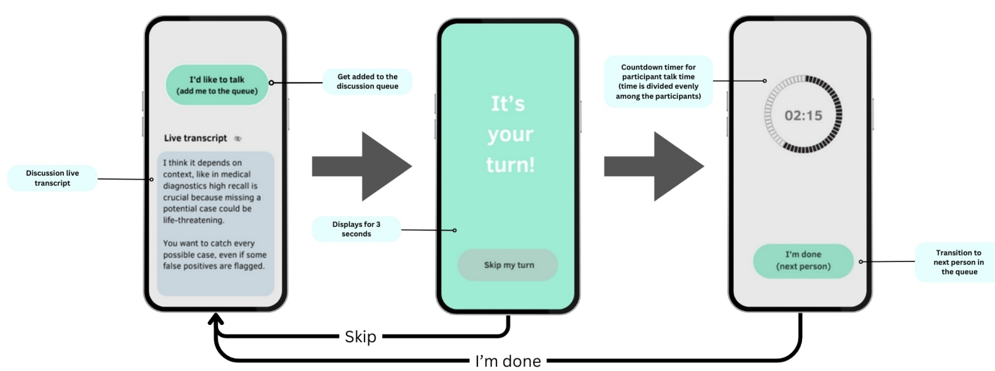


Figure 6.7: Student flow for discussion mode

Group Mode in the Interaction.ai platform is designed to enhance the group discussion experience by providing detailed, measurable insights into participation dynamics. This mode is built around the concept that having quantifiable attributes of performance, such as the amount and distribution of talk time, can significantly improve the coaching and management of group interactions.

One of the central features of Group Mode is the participation pie chart, a visual tool that displays the distribution of speaking time among group members. This chart makes it immediately apparent who is contributing most to the discussion and who may be holding back. By providing this level of transparency, the platform helps identify potential issues of dominance or reticence within the group that might not be evident even to the participants themselves. This feature is crucial as it allows both students and instructors to recognize and address imbalances early in the discussion, ensuring that all voices are heard and valued.

Group Mode also includes real-time management features that enhance the flow and relevance of discussions. It displays the current discussion question prominently and provides options for the group to switch to other topics, fostering a dynamic exchange that keeps the discussion fresh and engaging while allowing students options for what they want to engage with. Students can see a list of upcoming topics and collectively choose which ones to tackle next, giving them a sense of ownership and investment in the course of their dialogue.

Another innovative feature of Group Mode is the integration of a help request system. Students can request assistance from the instructor without disrupting the flow of discussion, which democratizes the attention each student receives. This system is designed to queue help requests in order of receipt, ensuring that assistance is rendered fairly and efficiently. This addresses a common challenge in classroom settings where instructors might miss hand raises or cues from quieter students, thus promoting an equitable learning environment.

By tracking and analyzing discussion metrics, Group Mode provides actionable insights that can significantly impact learning outcomes. For example, if a student consistently talks too much or too little, these behaviours can be addressed through targeted feedback and coaching. Instructors can use the data to encourage more balanced participation and to coach students on effective communication skills. Additionally, students themselves can reflect on their participation metrics, gaining awareness of their own behaviours and learning to adjust their engagement in group settings accordingly.

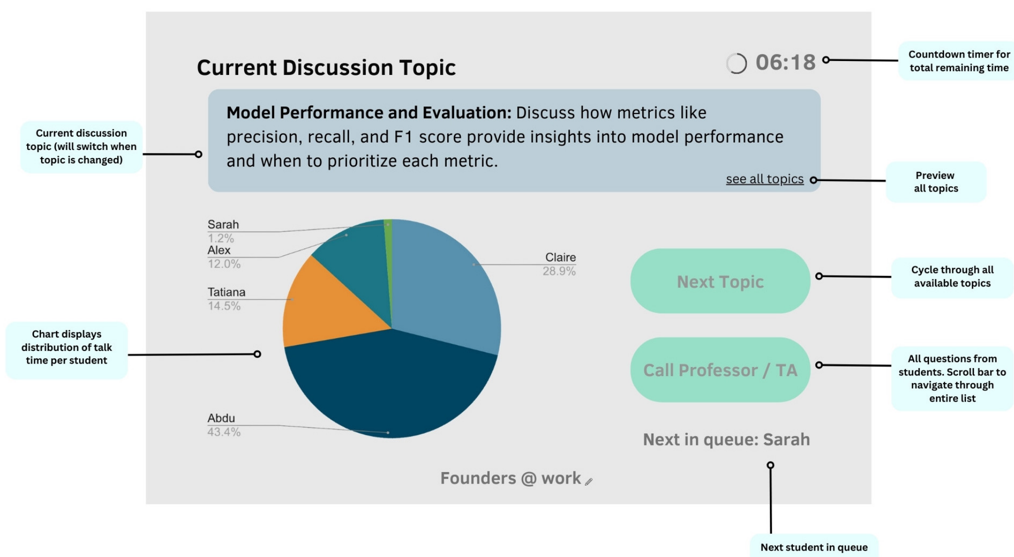


Figure 6.8: Discussion Mode: Group Table View

For instructors, Group Mode works in tandem with Instructor Mode, which offers a comprehensive at-a-glance view of class performance and participation metrics.

Instructors can monitor participation from a central dashboard that shows the participation charts for each group, allowing them to intervene strategically when they notice participation imbalances or other issues.

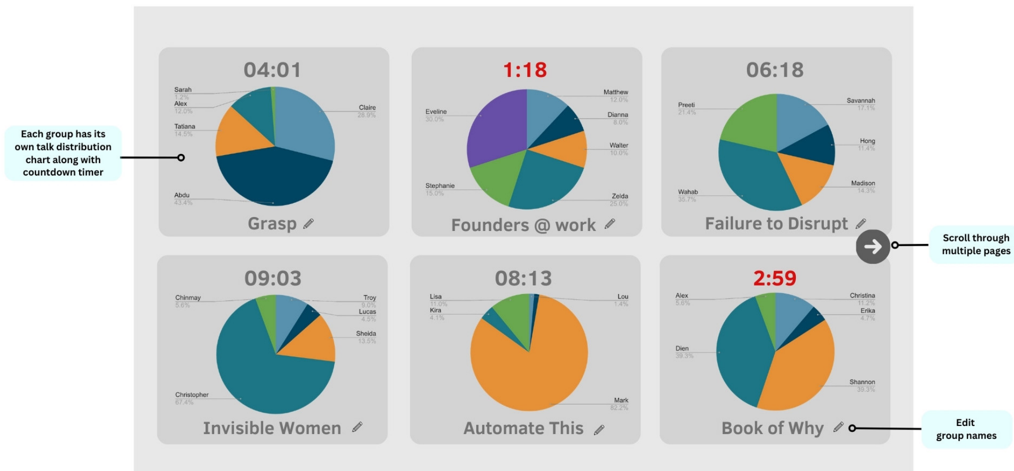


Figure 6.9: Discussion Mode: Instructor View

In addition to participation charts, Instructor Mode includes queue banners that show the speaking order of students. This feature is particularly beneficial in managing turn-taking and ensuring that all students have equal opportunities to request help. The banners update in real time, providing a clear sequence of participation.

The mode is also equipped with adjustable countdown timers, a feature that enhances the management of group discussions, especially in asynchronous learning environments where groups may start the discussion phase at different times. Whether a group starts their discussion earlier or later due to the varied pacing of video playbacks or personal scheduling, the countdown timer ensures that each group has a set, equitable amount of time to discuss. This standardization helps maintain a uniform structure across all discussions, regardless of when they start – though the interface is dynamic enough to support all organizational structures.

Ultimately, Instructor Mode empowers educators to enhance the effectiveness of group discussions. By providing a holistic view of class performance and detailed insights into each group's interaction, instructors can make informed decisions that foster a more inclusive and balanced educational environment. This mode not only supports the logistical aspects of teaching large classes but also contributes significantly to the goal of equitable student engagement and active participation in learning.

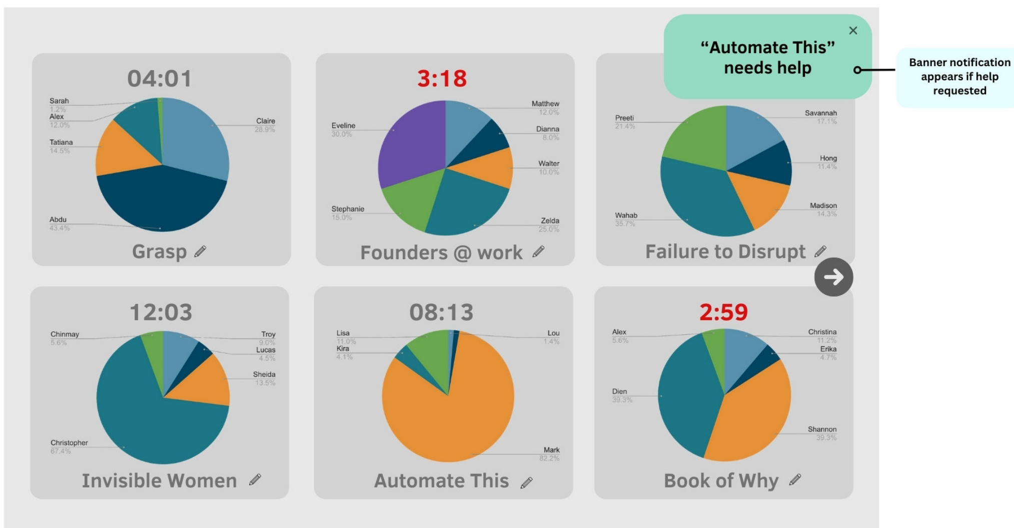


Figure 6.10: Instructor View. When a group indicates they need help a banner pop-up will appear

By integrating these modes, Discussion Mode not only facilitates robust interaction but also leverages technology to make educational experiences more engaging, fair, and responsive to student and instructor needs. It's a strategic enhancement that directly responds to the evolving dynamics of modern classrooms, ensuring that every student feels heard, valued, and actively involved in their learning journey.

The design of Interaction.ai's wireframes and its integrated features reflects a profound commitment to understanding and addressing the needs of end users. Through leveraging an extensive dataset, we have developed deepened empathy with our stakeholders, which has guided the creation of a tool that enhances the educational experience for both instructors and students. By incorporating technologies such as generative AI and large language models, Interaction.ai is positioned at the forefront of educational technology. It offers powerful, user-friendly solutions that streamline the teaching and learning processes, making sophisticated educational interactions more accessible and effective. Our hope is that this software not only improves the classroom dynamics but also sets a new standard for the integration of technology in education, fostering environments where learning is continuously enriched and effortlessly managed.

In conclusion, Interaction.ai has been meticulously designed with the end user in mind, utilizing a robust dataset to deeply understand and align with their needs. This application is envisioned to transform the classroom experience for both instructors and students by harnessing the latest advancements in generative AI and large language models. The integration of these technologies aims to enhance learning dynamics significantly, offering both real-time interaction and extensive analytical capabilities. It is our hope that this tool not only improves the educational landscape by making classrooms more interactive and responsive but also sets a new standard for the integration of technology in education, fostering environments where both teaching and learning can thrive.

Chapter 7

Limitations and Future Work

7.1 Critique and Alternatives to Design Thinking

Design Thinking is widely recognized for fostering creativity, aligning solutions with user needs, and encouraging cross-disciplinary collaboration. Its focus on empathy and experimentation makes it a particularly effective framework for addressing complex problems where human behaviour and preferences are central. Adopting Design Thinking in this research stems from its widespread use along with my prior experience with the method and the success I have observed in other contexts. However, while it remains a versatile and impactful approach, it is not without limitations. Other methodologies, such as Lean Startup, Agile Development, and Systems Thinking, offer alternative perspectives with unique advantages [26]. This section explores these approaches, considering their potential application to this research and providing a critical comparison to the Design Thinking framework.

7.1.1 Agile

Agile Development originated in software engineering as a framework to manage projects in dynamic and fast-paced environments. Its core principles, flexibility, collaboration, and iterative processes, are facilitated through tools like scrums, sprints, and Kanban boards [3]. Agile's primary goal is to adapt to changing requirements while consistently delivering value to the end user. By breaking projects into smaller, manageable cycles, Agile ensures regular feedback and allows teams to pivot quickly when needed.

Compared to Design Thinking, Agile focuses less on the exploratory phase of innovation, such as problem framing and ideation, and more on the operational phases of product development [26]. It excels in environments that require continuous refinement and implementation, where execution and efficiency are paramount. In contrast, Design Thinking prioritizes understanding user needs, reframing assumptions, and fostering creativity during the early stages of problem-solving. This creativity, however, can sometimes overlook practical constraints like cost, time, and feasibility, making solutions less executable in the operational phase, particularly when scaling ideas for larger user bases.

While Design Thinking focuses on gaining a deep understanding of the user [7], Agile ensures steady progress by delivering concrete solutions throughout the development process. Its emphasis on execution helps prevent teams from lingering too long in the problem-framing phase, addressing a critical weakness of Design Thinking in contexts that demand continuous, actionable delivery. However, Agile's focus on iterative development and rapid delivery can lead to short-term fixes rather than more comprehensive, long-term solutions, especially when broader contextual factors are neglected.

Unlike Design Thinking, which encourages reframing problems and challenging assumptions, Agile typically assumes that the problem space is already well-defined. This assumption can limit its effectiveness in addressing complex or ambiguous challenges that require a more in-depth exploration of user needs and systemic factors.

Despite these differences, the frameworks can complement each other. Agile's iterative cycles benefit from the insights and direction derived from Design Thinking's exploratory stages, such as understanding the user and framing the problem. Agile's structured approach to delivery can enhance the implementation of solutions generated during Design Thinking's creative and empathetic problem-solving phases.

Integrating Agile with Design Thinking can help bridge these gaps. Design Thinking's emphasis on empathy, creativity, and problem reframing can guide Agile's iterative cycles, ensuring that execution remains aligned with user-centred insights. Conversely, Agile's structured delivery methods provide a disciplined pathway for implementing the ideas generated through Design Thinking. This integration suggests the potential for a hybrid approach, though challenges remain in managing transitions between frameworks, particularly in how user feedback and problem framing are coordinated. Such considerations are particularly relevant in areas like educational technology, where user engagement and adaptability are essential for success.

7.1.2 Lean Startup

Lean Startup is a methodology aimed at improving the success rate of new ventures by emphasizing rapid iteration through minimum viable products (MVPs) and feedback loops. The cycle of "build, measure, learn" is designed to minimize waste, reduce market risk, and accelerate time to market [33]. By focusing on the continuous development of MVPs and incorporating real-world validation, Lean Startup allows businesses to test their assumptions early and adjust quickly based on market feedback. This makes the methodology particularly beneficial for startups or organizations operating with limited resources, as it prioritizes efficiency and the elimination of unnecessary processes. Lean Startup's key advantage lies in its ability to validate products quickly, helping to ensure that businesses are on the right track before committing significant resources.

In contrast, the abductive reasoning that underlies the Design Thinking process focuses more on the early stages of problem-solving, such as understanding user needs, framing and reframing assumptions, and generating creative solutions through empathy and observational research [26]. While both methodologies value iteration, Lean Startup is more outcome-focused, with a specific emphasis on market readiness and product-market fit. Lean Startup's approach is rooted in validating hypotheses through real-world testing, often in the form of MVPs, which can quickly confirm or disprove assumptions about customer needs and preferences. This emphasis on efficiency and quick pivots based on customer feedback is particularly advantageous in markets that require rapid adjustments or in resource-constrained environments where time and funding are limited.

One of Lean Startup's main advantages over Design Thinking is its focus on measurable metrics and tangible outcomes. Lean Startup emphasizes learning from real-world data, allowing teams to validate their ideas with customers early on in the process. This data-driven approach ensures that the solutions being developed are grounded in actual market demand, making Lean Startup particularly effective in environments that require quick decision-making and minimal resource investment. In contrast, Design Thinking often leans heavily on qualitative data, such as interviews and ethnographic research, which can sometimes delay decision-making and limit the immediacy of the feedback loop.

However, Lean Startup also has its downsides. Its emphasis on rapid iteration and MVPs may sometimes prioritize speed over depth, leading to solutions that are prematurely scaled or lack long-term sustainability. Additionally, Lean Startup's focus on market validation can sometimes overlook the broader social or cultural context that Design Thinking takes into account, potentially resulting in products that meet market demand but do not align with deeper user needs or systemic considerations. While Lean Startup's data-driven approach helps to minimize market risk, it can inadvertently limit creativity and exploration, particularly in fields or problems that require deeper empathy and understanding of user behaviour beyond what can be quickly tested in an MVP.

Lean Startup and Design Thinking offer distinct advantages depending on the context in which they are applied. Lean Startup's focus on rapid testing, real-world validation, and efficiency makes it an effective tool for startups and projects that require quick market feedback and validation. However, Design Thinking's emphasis on empathy, user-centred design, and exploration of ill-defined problems allows for a broader scope in addressing complex, multifaceted issues. Both approaches share iterative principles but differ significantly in their priorities and methodologies, making each more suitable for different stages of the product development lifecycle.

7.1.3 Systems Thinking

Systems Thinking is an approach to problem-solving that views complex problems as part of an interconnected system rather than isolated issues. It emphasizes understanding the relationships between various components of a system, recognizing that changes in one part of the system can have cascading effects on the whole. By considering feedback loops, patterns, and interdependencies, Systems Thinking enables a more holistic understanding of problems. This methodology is particularly beneficial in addressing complex, long-term challenges, as it allows for a comprehensive analysis of the underlying structures and forces that shape system behaviors [51]. The key advantage of Systems Thinking is its ability to model and understand how various elements within a system interact, providing a clearer picture of the whole and informing more sustainable and effective solutions.

One of the main advantages of Systems Thinking over Design Thinking is its ability to address complex, multi-dimensional issues by considering the interconnectedness of system components. For example, in tackling social, environmental, or organizational problems, Systems Thinking allows for a deeper analysis of the various factors influencing the issue, such as feedback loops, delayed effects, and unintended consequences [36]. This broader understanding can lead to more sustainable solutions that take into account the long-term impact of changes within the system. Design Thinking, while valuable in its focus on human-centred solutions, may not always account for these systemic complexities, potentially leading to solutions that are effective in the short term but fail to address root causes or long-term consequences.

However, the broad and complex nature of Systems Thinking can also be a drawback. The emphasis on understanding the entire system can lead to analysis paralysis, where the sheer number of interconnections and variables overwhelms decision-making. Unlike Design Thinking, which fosters creativity and ideation in a more defined problem space, Systems Thinking requires extensive analysis, which can delay action and result in less immediate, tangible outcomes. Additionally, the focus on systemic relationships may lead to solutions that are difficult to implement or scale in real-world contexts, where practical constraints such as time, resources, and political factors must be considered.

While Systems Thinking offers a powerful framework for addressing complex, multi-layered problems by considering the whole system, it may be less practical in situations that demand quick, iterative, user-centred solutions. Design Thinking, by focusing on human-centred problem-solving and iterative prototyping, is well-suited for contexts where empathy and immediate user needs are the priority. However, Systems Thinking's holistic view allows for a deeper understanding of the broader context, making it particularly valuable in addressing long-term, complex issues that extend beyond individual user needs or product features.

7.2 Future Work

A key next step involves refining the prototype Interaction.ai, and implementing it in a university classroom. Doing so would help evaluate its broader impact on student participation, engagement, and instructor workload. This phase can be supported by applying the frameworks discussed in this chapter (Agile, Lean, and Systems Thinking) which are well-suited to guide the implementation process. While this thesis has demonstrated the potential of AI-driven tools to enhance student engagement, the full impact on classroom dynamics, particularly in large and diverse courses, remains an area for further exploration. Implementing the system across multiple classrooms will also provide valuable insights into its ability to support various instructional styles, adapt to different course formats, and meet the specific needs of both instructors and students.

Future work will also seek to contribute to the broader discourse surrounding AI in education by exploring how AI tools like Interaction.ai can complement traditional pedagogical approaches rather than replace them. This will involve a deeper investigation into how AI can be used to enhance instructor-student interactions, improve engagement during presentations, and foster a more collaborative learning environment. By continuing to evaluate the system's effectiveness in real-world educational settings, the research will inform best practices for integrating AI into classrooms while preserving the human elements of teaching. This work will also contribute to the growing body of knowledge on the design, deployment, and impact of educational AI systems.

Finally, the potential for scaling Interaction.ai beyond the context of this study will be explored. This includes investigating its applicability in other educational settings, such as K-12 schools, online learning platforms, or professional development environments. By adapting the system to different teaching contexts and user demographics, future research could expand the reach and impact of Interaction.ai, making it a versatile tool for improving engagement and feedback across various educational levels. The insights gained from these efforts will provide a comprehensive framework for designing and implementing AI-driven solutions that enhance both teaching and learning in diverse educational contexts.

7.3 Conclusion

The primary contribution of this thesis is the design of a novel prototype, Interaction.ai (See Chapter 6.2), informed by design thinking principles and qualitative feedback. This work explores how innovative educational technology can transform traditional learning environments into more dynamic and interactive spaces.

This work also compared the performance of AI and human approaches in question selection during group presentations. The results demonstrated that AI-driven methods could achieve outcomes similar to those of instructors while significantly improving efficiency. These findings underscore the potential of AI to complement traditional teaching methods, offering scalable solutions that address both operational and pedagogical challenges. The integration of reflexive thematic analysis provided deep insights into students' experiences, highlighting barriers to participation, such as social anxiety and group dynamics.

While this research achieved its primary goals, it also revealed certain limitations. For instance, the implementation of design thinking was constrained by the cyclical nature of course delivery, limiting opportunities for rapid iteration. The study focused on specific classroom settings and disciplines, which may not fully capture the diversity of educational contexts. Future research should aim to address these limitations by exploring broader applications of the prototype and integrating complementary frameworks, such as Agile or Lean Startup, to enhance adaptability and scalability.

Looking ahead, this work paves the way for exciting opportunities in educational technology. Future iterations of the prototype could include more advanced AI features, such as sentiment analysis and adaptive feedback, to further support both students and instructors. Expanding the implementation of Interaction.ai across various classroom contexts will allow for a more comprehensive evaluation of its impact on participation, engagement, and overall learning outcomes.

As AI continues to reshape education, it is met with both enthusiasm and skepticism. Concerns about data privacy, algorithmic bias, and the potential over-reliance on technology highlight the need for a cautious approach. Many educators worry about the unintended consequences of integrating AI into classrooms, including the risk of diminishing human-centred learning or replacing critical aspects of teaching. History has shown that technology, when used thoughtfully, can significantly enhance education – just as the printing press and the internet revolutionized knowledge dissemination, AI holds the potential to do the same.

Rather than viewing AI as a replacement for human educators, it should be embraced as a powerful tool to augment teaching and learning. By automating administrative tasks and personalizing learning experiences, AI can free educators to focus on fostering critical thinking, creativity, and collaboration. The challenge lies in using AI responsibly, addressing its limitations, and ensuring it aligns with ethical practices. By preparing educators and students to work with AI effectively, we can harness its strengths to create more inclusive, efficient, and engaging learning environments, ultimately making education more accessible and impactful for all.

Bibliography

[1] C. Anand, G. Dulai, L. Yao, M. Arief, O. D'Mello, S. Menon, and C. Schankula. *Creating With Code, An Introduction to Functional Programming, User Interaction, and Design Thinking.* Fondation STaBL Foundation, 2023. ISBN 978-1-73886-950-3.

[2] M. B. Ayalew. Socialization experiences among undergraduate students in higher learning institutions (hli). In *Higher Education-New Approaches to Accreditation, Digitalization, and Globalization in the Age of Covid*. IntechOpen, 2022.

[3] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, et al. Manifesto for Agile Software Development. `https://agilemanifesto.org/`, 2001. Accessed: 2025-01-06.

[4] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[5] V. Braun and V. Clarke. *Thematic analysis.* American Psychological Association, 2012.

[6] B. Brinkman. An analysis of student privacy rights in the use of plagiarism detection systems. *Science and engineering ethics*, 19:1255–1266, 2013.

[7] T. Brown. Design Thinking. *Harvard business review*, 86(6):84, 2008.

[8] M. Cantamessa, F. Montagna, S. Altavilla, and A. Casagrande-Seretti. Data-driven design: the new challenges of digitalization on product design and development. *Design Science*, 6:e27, 2020.

[9] M. M. Chingos and G. J. Whitehurst. Class size: What research says and what it means for state policy. *Washington, DC: Brookings Institute. Retrieved June*, 5:2016, 2011.

[10] N. Cross. Designerly ways of knowing. *Design studies*, 3(4):221–227, 1982.

[11] R. F. Dam and Y. S. Teo. The history of design thinking. *Interaction Design Foundation - IxDF*, May 20 2022. URL `https://www.interaction-design.org/literature/article/design-thinking-get-a-quick-overview-of-the-history`.

[12] A. P. Delfino. Student engagement and academic performance of students of partido state university. *Asian Journal of University Education*, 15(1):n1, 2019.

[13] J. Dempere, K. Modugu, A. Hesham, and L. K. Ramasamy. The impact of chatgpt on higher education. In *Frontiers in Education*, volume 8, page 1206936. Frontiers, 2023.

[14] C. Duhigg. What Google learned from its quest to build the perfect team. *The New York Times Magazine*, 26(2016):2016, 2016.

[15] P. Engzell, A. Frey, and M. D. Verhagen. Learning loss due to school closures during the covid-19 pandemic. *Proceedings of the National Academy of Sciences*, 118(17):e2022376118, 2021.

[16] X. Fan, W. Luo, M. Menekse, D. Litman, and J. Wang. Scaling reflection prompts in large classrooms via mobile interfaces and natural language processing. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, pages 363–374, 2017.

[17] A. Hamilton, D. Wiliam, and J. Hattie. The future of ai in education: 13 things we can do to minimize the damage. *Working Paper*, 2023.

[18] Harvard Graduate School of Education. Why psychological safety matters in class, 2022. URL `https://www.gse.harvard.edu/news/uk/22/09/why-psychological-safety-matters-class`. Accessed: 2024-05-17.

[19] B. Hollister, P. Nair, S. Hill-Lindsay, and L. Chukoskie. Engagement in online learning: student attitudes and behavior during covid-19. In *Frontiers in education*, volume 7, page 851019. Frontiers Media SA, 2022.

[20] S. Järvelä, A. Nguyen, and A. Hadwin. Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology*, 54(5):1057–1076, 2023.

[21] U. Johansson-Sköldberg, J. Woodilla, and M. Çetinkaya. Design thinking: Past, present and possible futures. *Creativity and innovation management*, 22(2):121–146, 2013.

[22] L. Kuka, C. Hörmann, and B. Sabitzer. Teaching and learning with ai in higher education: a scoping review. *Learning with Technologies and Technologies in Learning: Experience, Trends and Challenges in Higher Education*, pages 551–571, 2022.

[23] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33:159–174, 1977.

[24] M. E. Loades, E. Chatburn, N. Higson-Sweeney, S. Reynolds, R. Shafran, A. Brigden, C. Linney, M. N. McManus, C. Borwick, and E. Crawley. Rapid systematic review: the impact of social isolation and loneliness on the mental health of children and adolescents in the context of covid-19. *Journal of the American Academy of Child & Adolescent Psychiatry*, 59(11):1218–1239, 2020.

[25] J. M. Lodge, K. Thompson, and L. Corrin. Mapping out a research agenda for generative artificial intelligence in tertiary education. *Australasian Journal of Educational Technology*, 39(1):1–8, 2023.

[26] P. Micheli, S. J. Wilner, S. H. Bhatti, M. Mura, and M. B. Beverland. Doing design thinking: Conceptual review, synthesis, and research agenda. *Journal of Product innovation management*, 36(2):124–148, 2019.

[27] A. Nguyen, H. N. Ngo, Y. Hong, B. Dang, and B.-P. T. Nguyen. Ethical principles for artificial intelligence in education. *Education and Information Technologies*, 28(4):4221–4241, 2023.

[28] C. O'neil. *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown, 2017.

[29] J. A. Paul, H. M. Baker, and J. D. Cochran. Effect of online social networking on student academic performance. *Computers in human behavior*, 28(6):2117–2127, 2012.

[30] N. Racine, B. A. McArthur, J. E. Cooke, R. Eirich, J. Zhu, and S. Madigan. Global prevalence of depressive and anxiety symptoms in children and adolescents during covid-19: a meta-analysis. *JAMA pediatrics*, 175(11):1142–1150, 2021.

[31] L. Rai, C. Deng, and F. Liu. Developing massive open online course style assessments using generative ai tools. In *2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT)*, pages 1292–1294. IEEE, 2023.

[32] J. Reich. *Failure to disrupt: Why technology alone can't transform education.* Harvard University Press, 2020.

[33] E. Ries. *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses.* Crown Currency, 2011.

[34] S. Sarma and L. Yoquinto. *Grasp: The science transforming how we learn.* Anchor, 2021.

[35] S. Sarsa, P. Denny, A. Hellas, and J. Leinonen. Automatic generation of programming exercises and code explanations using large language models. In *Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 1*, pages 27–43, 2022.

[36] P. M. Senge. *The fifth discipline: The art and practice of the learning organization*. Broadway Business, 2006.

[37] K. Shilton. Engaging values despite neutrality: Challenges and approaches to values reflection during the design of internet infrastructure. *Science, Technology, & Human Values*, 43(2):247–269, 2018.

[38] A. Shoufan. Exploring students' perceptions of chatgpt: Thematic analysis and follow-up survey. *IEEE Access*, 2023.

[39] H. A. Simon. *The Sciences of the Artificial, reissue of the third edition with a new introduction by John Laird*. MIT press, 2019.

[40] V. Terziev and S. Vasileva. The role of education in socialization of an individual. *Proceedings of INTCESS 2022- 9th International Conference on Education & Education of Social Sciences*, 2022. doi: 10.51508/intcess.202272.

[41] M. Tissenbaum, C. Matuk, M. Berland, L. Lyons, F. Cocco, M. Linn, J. L. Plass, N. Hajny, A. Olsen, B. Schwendimann, et al. Real-time visualization of student activities to support classroom orchestration. In *ICLS 2016 Proceedings*. Singapore: International Society of the Learning Sciences, 2016.

[42] United Nations. Education during covid-19 and beyond, 2020. URL `https://unsdg.un.org/resources/policy-brief-education-during-covid-19-and-beyond`. Accessed: April 2024.

[43] Vanderbilt University. Guidance on ai detection and why we're disabling turnitin's ai detector, Aug 2023. URL `https://www.vanderbilt.edu/`

`brightspace/2023/08/16/guidance-on-ai-detection-and-why-were-disabling-turnitins-ai-detector/`. Accessed: 2024-05-17.

[44] R. A. Varanasi, D. Siddarth, V. Seshadri, K. Bali, and A. Vashistha. Feeling proud, feeling embarrassed: Experiences of low-income women with crowd work. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2022.

[45] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[46] W. Venters, E. Oborn, and M. Barrett. A trichordal temporal approach to digital coordination. *MIS quarterly*, 38(3):927–A18, 2014.

[47] D. Vlah, A. Kastrin, J. Povh, and N. Vukašinović. Data-driven engineering design: A systematic review using scientometric approach. *Advanced Engineering Informatics*, 54:101774, 2022.

[48] L. Wang and L. Calvano. Class size, student behaviors and educational outcomes. *Organization Management Journal*, 19(4):126–142, 2022.

[49] T. Webb, K. J. Holyoak, and H. Lu. Emergent analogical reasoning in large language models. *Nature Human Behaviour*, 7(9):1526–1541, 2023.

[50] N. F. Weber. The bauhaus at 100: science by design. *Nature*, 572(7768):174–175, 2019.

[51] D. Wright and D. H. Meadows. *Thinking in systems*. Earthscan, 2008.

[52] C. Xie, M. Wang, and H. Hu. Effects of constructivist and transmission instructional models on mathematics achievement in mainland china: A meta-analysis. *Frontiers in psychology*, 9:1923, 2018.

[53] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur. Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1):1–27, 2019.

[54] P. Zhang and D. F. Galletta. *Human-computer interaction and management information systems: Foundations*, volume 5. ME Sharpe, 2006.

[55] M. Zulfiquar and S. Zamir. The role of classroom culture and psychological safety in efl students' engagement. *Frontiers in Education*, 2022. URL `https://www.frontiersin.org/articles/10.3389/feduc.2022.740039/full`. Accessed: 2024-05-17.

Appendix

First Questionnaire

1. In your opinion, what is the most valuable part of the university classroom experience?

2. What is the most challenging part of the university classroom experience? What makes learning / achieving success challenging? What do you wish could be changed?

3. How do you feel about the process of asking questions in class? What prevents you from asking questions in a traditional classroom setting?

4. What makes a question a ”great question” in your opinion?

5. Do you feel asking questions improves your ability to learn?

   ☐ Yes

   ☐ No

6. Do you typically ask questions in class?

   ☐ Yes, almost every class

☐ Yes, sometimes

☐ Rarely

☐ Almost never

7. Evaluate the following scenarios compared to how we collected questions for the presentations (through avenue quizzes). How would each change your likelihood of asking a question in class?

   (a) You were instructed to ask your question into a microphone in front of class.

   ☐ Much LESS likely to ask the question

   ☐ A little LESS likely to ask the question

   ☐ No effect (about the same as avenue quizzes)

   ☐ A little MORE likely to ask the question

   ☐ Much MORE likely to ask the question

   (b) You were instructed to raise your hand in lecture to ask your question.

   ☐ Much LESS likely to ask the question

   ☐ A little LESS likely to ask the question

   ☐ No effect (about the same as avenue quizzes)

   ☐ A little MORE likely to ask the question

   ☐ Much MORE likely to ask the question

   (c) You were instructed to ask your question in the public chat on Microsoft Teams.

   ☐ Much LESS likely to ask the question

☐ A little LESS likely to ask the question

☐ No effect (about the same as avenue quizzes)

☐ A little MORE likely to ask the question

☐ Much MORE likely to ask the question

(d) You were instructed to send your question as a private message to a TA or instructor.

☐ Much LESS likely to ask the question

☐ A little LESS likely to ask the question

☐ No effect (about the same as avenue quizzes)

☐ A little MORE likely to ask the question

☐ Much MORE likely to ask the question

(e) You were instructed to send your question in an anonymized form.

☐ Much LESS likely to ask the question

☐ A little LESS likely to ask the question

☐ No effect (about the same as avenue quizzes)

☐ A little MORE likely to ask the question

☐ Much MORE likely to ask the question

8. Which question set is better?

☐ Question Set A

☐ Question Set B

9. How much better was the set you chose?

- ☐ Significantly better. The question set was much more engaging and interesting.
- ☐ A little better. The question set I chose was slightly more interesting.
- ☐ I chose a set at random because they were unsubstantially different.

10. Do you have any additional comments? (optional)

11. Would you be open to a research assistant following up with you to better understand your answers?

    - ☐ I'd be open to a 10 minute audio-only call
    - ☐ I'd be open to filling out another survey
    - ☐ I'm not interested at this time

Second Questionnaire

1. What is your email?

2. How did you find the way the group presentations were handled in class?

   ☐ Better than in other large classrooms

   ☐ About the same

   ☐ Worse than in other large classrooms

3. Why did you feel that way? What would you like to see done differently?

4. What are the biggest challenges you face in paying attention in class? What about in paying attention to the video presentations specifically?

5. When you are having difficulty concentrating with other tasks at home or at school, what do you find helps you get back on track?

6. Which of the following would have helped you maintain focus?

   ☐ More engagement points (quizzes and check-ins during class)

   ☐ Longer breaks in between topics (2-5 minute exercise break, time to socialize)

   ☐ Having presentations more spread out

   ☐ Other (write here...)

7. Did you watch any videos at home instead of class?

   ☐ Yes, most of the time

- ☐ Yes, sometimes
- ☐ 2-5 videos
- ☐ None (or just one)

8. What prevented you from watching the videos beforehand?
   - ☐ My schedule constraints
   - ☐ Not enough of a priority
   - ☐ I didn't find it valuable to do so
   - ☐ I preferred to watch them live / in class
   - ☐ Other...
9. Why did you feel that way?
10. Did you prefer having the videos played in class or watched at home?
    - ☐ In class
    - ☐ At home during class time
    - ☐ At home prior to class
    - ☐ I would have preferred the videos to not be pre-recorded (presented live)
11. Why did you prefer that method?
12. What of the following would you like to see in a future iteration of the way presentations and questions are handled in a group class setting?
    - ☐ Recording a transcript of questions asked at each table and extracting questions from live conversations (instead of typing)

- ☐ Quizzes throughout the presentation
- ☐ Prompts to click through to maintain engagement (i.e. press space bar to keep playing the video)
- ☐ Have a dedicated question answering class / a dedicated presentation watching class
- ☐ Requiring students to watch videos before and just answer questions in class
- ☐ Other:

13. Any other comments about how this class and others could be improved? (comments on hybrid learning? comments on how you're finding university in general? what are you struggling with?)

14. Would you be interested in a 10 minute audio interview to better explain your opinion?

    - ☐ Yes
    - ☐ Yes but next semester only
    - ☐ No thank you!