Norwegian School of Economics

Bergen, Autumn 2023

Adopt or Abort?

Mapping students' and professors' attitudes towards the use of generative AI in higher education

Daniel Rasmussen and Tommy Karlsen

Supervisor: Mohammed Mardan

Master thesis, Economics and Business Administration

Major: Business Analysis and Performance Management; and Strategy and Management

NORWEGIAN SCHOOL OF ECONOMICS

Preface

This thesis is completed in the fall of 2023 as a requirement for our Master of Science degree in Economics and Business Administration at the NHH. It falls within the scope of the Business Analysis and Performance Management (BUS) and Strategy and Management (STR) programs.

The rollout of ChatGPT in November 2022 sparked our fascination with artificial intelligence, prompting us to explore this promising field further. We quickly realized the potential of AI in transforming various sectors, particularly education, which plays a pivotal role in societal value creation. Our literature review revealed that the education sector stands to benefit significantly from AI, particularly in enhancing productivity. This discovery led us to investigate AI's applications and implications in an educational setting, especially considering the impact of attitudes on the successful integration of such technologies. We noted that at NHH, the policy surrounding GenAI tools was unclear, leading to uncertainty among students and professors. This observation further cemented our choice to focus on the attitudes towards AI in higher education, as understanding these perspectives is critical for effective implementation.

We would like to thank our supervisor, Mohammed Mardan, for his invaluable contributions, including insightful ideas, constructive feedback, and unwavering support throughout the semester. His expertise greatly enriched the quality of our work. We are also grateful to Frank Mortensen and Arild Schanke for their input on the use of GenAI in higher education, and to Bram Timmermans for his perspectives on AI's role in this sector. His invitation to present and discuss our findings at a workshop with esteemed professors from NHH, Aalborg University, Gothenburg University, and Tampere University was an honor. Lastly, our thanks go to Joel Berge, Siv Skard, Aruna Tatavarthy, and Björn Schmeisser for their valuable feedback on our survey.

Abstract

This exploratory and descriptive study investigates the attitudes of students and professors at Norwegian School of Economics (NHH) towards the implementation of Generative Artificial Intelligence (GenAI) in higher education. The research is set against the backdrop of GenAI's rapid rise in popularity due to OpenAI's prominent service in ChatGPT, a development that has ignited extensive discussions within the educational sector. Grounded in a thorough literature review, we developed and administered an online survey, collecting quantitative data from 274 participants. Specific applications of GenAI explored include text generation for student work, GenAI as an assessment tool, and GenAI as a tool for virtual assistance.

Findings reveal a general trend of enthusiasm among students towards GenAI, contrasted with a slightly more cautious approach from professors, indicating a potential generational gap. Further, the study observes (1) a decline in approval for text generating features as they get more intrusive in the student's work, (2) a critical stance towards GenAI as an assessment tool, and (3) a preference for human advice while simultaneously recognizing the utility of GenAI as a virtual assistant tool.

In the last part, we propose reflection points and recommendations for future policy-making at NHH in regard to GenAI use, rooted in the earlier analysis and discussion. The general suggestion revolves around developing education-specific GenAI tools, drawing inspiration from University of Oslo's approach, as well as forming an AI-committee for ongoing evaluation of how the GenAI applications play out. Lastly, it must be highlighted that while there is interest in integrating GenAI in higher education, significant considerations must be addressed, including varied opinions, institutional goals, and the need for specialized tools and competencies. This research contributes to the understanding of GenAI's role in higher education and aspires to serve as a source of inspiration when forming future policies at NHH, and beyond.

Table of Contents

Table of Figures

Table of Tables

1 Introduction

GenAI, a branch of the broader field of AI, has recently attracted a lot of attention with the emergence of OpenAI's chatbot ChatGPT. Only two months after its launch in November 2022, the service reached a staggering 100 million users and notably passed the US bar exam scoring in the 90th percentile (Arredondo, 2023; Hu, 2023). In addition to sparking wider societal discourse, the academic community caught interest in the potential use of GenAI in higher education. More specifically, the technology has received mixed reactions from educational institutions, where some have opted to completely prohibit all use, while others proactively adopt it as an educational learning tool (UCL, 2023; UiO, 2023; Vassdal 2023).

Looking at the literature, we find preliminary research highlighting both benefits and challenges related to integrating GenAI in higher education. Some of the benefits include personalized feedback, accessibility, scalability, reduced barrier for educational support, and student engagement (Andreassen, 2023; Chan & Hu, 2023; Crompton et al., 2023; Labaddze et al., 2023; Mollick et al., 2023; Wollny et al., 2021). On the other hand, misleading and non-factual information, dishonest use, data privacy, lack of socialization, and reduced critical thinking, are often mentioned issues (Ayman et al., 2023; Chan & Hu, 2023; Montenegro-Rueda et al., 2023; Munthe et al., 2022).

Concerning research specifically gauging opinions towards the use of GenAI in higher education, we find two papers with contrasting results. First, Chan and Hu (2023) conducted a survey of 399 undergraduate and postgraduate students in Hong Kong, which revealed a generally positive attitude and familiarity towards the use of GenAI. Conversely, a study on Spanish economics and business students by Almaraz-López (2023) uncovered that a majority of participants expressed discomfort with AI concepts. These contrasts allude to varying degrees of familiarity and acceptance across different cultural and educational settings.

After conducting an extensive literature review, it becomes apparent that the academic resources available on this subject are scarce, which likely is a consequence of the technology's very recent surge in mainstream popularity. In this paper, we would therefore like to contribute in the effort of bridging this research gap by further exploring attitudes towards the adoption of GenAI in higher education. More specifically, we will be examining the perceptions of students and professors at NHH, which is the institution our research team is conducting the master thesis under.

At a broad level, our research is important because higher education plays a crucial role in society. It not only fosters individual intellectual development and social progression (Mokyr, 2002), but also has the responsibility to serve public purposes by enforcing social change (Shapiro, 2005). Brennan further discusses the social role of universities in a digital era, highlighting their importance in the development of labor market skills and citizenship. This is particularly relevant in our context because the labor market expects institutions like NHH to equip students with future-oriented skills like being proficient with using GenAI.

Narrowing down to Norway, our thesis is useful because the significance of AI in shaping Norway's digital future has been increasingly recognized by the government (Astrup, 2020). Although this technologically advanced country is uniquely positioned to effectively leverage AI (Nesse & Erdal, 2022; Parmiggiani & Mikalef, 2022), the Norwegian ministry of education and research has acknowledged that there still remain major steps before digital tools could add the desired quality in education (Kunnskapsdepartementet, 2021). In this effort, our research could possibly inform a strategy of how to appropriately integrate GenAI in a way that satisfies the mentioned desired educational quality.

Finally, we would like to highlight the importance of gauging attitudes as a necessary precursor to technological adoption. Chatterjee and Bhattacharjee (2020) demonstrate that attitudes significantly influence individuals' intentions to use AI in higher education. Similarly, Okonkwo et al. (2019) establish that user attitudes are key determinants in the adoption of software engineering products. These findings collectively suggest that assessing attitudes is a crucial preliminary step in technology adoption, providing a gauge of user readiness to embrace GenAI in a higher education context.

Research question: *What are students' and professors' attitudes towards implementing generative artificial intelligence in higher education? How do the perceptions differ between the two groups?*

2 Literature review

In this paper, we would like to map NHH students' and professors' attitudes towards adoption of GenAI in higher education. To do this, we must first understand (1) the current state of higher education in Norway, (2) what GenAI is, and (3) investigate the intersection between these domains. By reviewing existing literature, studies, and reports, this chapter aims to lay a foundation for understanding the complex dynamics at play in the perception and use of GenAI in an academic setting.

2.1 The current state of higher education in Norway

The Norwegian higher education system, as reported in government documents, is founded on principles of quality, accessibility, and research excellence (Ministry of Education and Research, 2022). This system encompasses a variety of predominantly publicly funded universities and specialized colleges, ensuring that education remains affordable for students, irrespective of financial background (Liu & Kong, 2023). This egalitarian approach is a cornerstone of the Nordic model of education, emphasizing inclusivity and equal opportunities for all learners. With rising labor market needs for educated workers as well as supportive government policies, participation rates in higher education have steadily increased from just a few percent in the 1960s, to around 37% in 2023 (Bleiklie, 2023; ssb.no, 2023).

Despite all the positive aspect, there are also significant challenges in the Norwegian higher education system, where digitalization is a particularly prominent topic. During the Covid-19 pandemic, students and professors were abruptly required to move from the traditional physical space to a digital one in order to maintain social distancing requirements during lockdown (Ratten, 2023). This accelerated an already growing change towards digitalization in higher education. The Norwegian ministry of education and research acknowledged that the sector was not ready for this sudden move, and that major steps remain before digital tools could add the desired quality in Norwegian higher education (Kunnskapsdepartementet, 2021). Nonetheless, the ministry also noticed the potential digital teaching could provide. Students and professors gained valuable hands-on experience of digital teaching and educational technology, such as Zoom and Microsoft Teams. While post-pandemic has shown a reintroduction of physical face-to-face classes, students' desires have now changed as we are seeing an increase in wanting hybrid learning (Ratten, 2023). Imran et al. (2023) confirm the effectiveness of hybrid learning and anticipate it having a significant impact in enhancing teaching in higher education in the post-pandemic landscape.

2.2 GenAI

2.2.1 Artificial intelligence and machine learning

To get an understanding of what GenAI is, we must first define the terms artificial intelligence and machine learning.

Artificial intelligence can be defined as a broad field of study that encompasses the creation of intelligent agents, which are systems that can reason, learn, and act autonomously (Russel & Norvig, 2009). These agents utilize techniques ranging from rule-based systems to complex neural networks. Core components of AI include search and optimization for task-solving, knowledge representation for information storage, and diverse learning methods such as machine learning and evolutionary algorithms (McCarthy, 2007; Nilsson, 1998). Through capabilities like natural language processing in chatbots and autonomous navigation in self-driving vehicles, AI's overarching aim is to emulate human-like intelligence across various contexts (University of Bergen, n.d).

Machine learning (ML), a subfield of AI, emphasizes the development of algorithms capable of learning from data (Domingos, 2012; Domingos, 2015; Mitchell, 1997). An essential aspect of ML is training algorithms using datasets, which are structured collections of examples related to a specific task. A typical application of ML can be seen in image classification. Here, an algorithm is trained with labeled images, and learns to categorize them based on pixel patterns (Cortes & Vapnik, 1995; Goodfellow et al., 2016). This pattern recognition is a crucial step in the learning process (Bishop & Nasrabadi, 2006). Once adequately trained, such algorithms can predict outcomes for new and unseen data. Today, ML algorithms have found their way into various services and products we use daily, from internet search engines to email spam filters.

2.2.2 GenAI and chatbots

Just as machine learning is a subset of the broader field of AI, GenAI is a branch of machine learning – see Figure 1. This subfield emphasizes the production of new content that is often almost indistinguishable from genuine data. At the forefront of this movement is the concept of Generative Adversarial Networks (GANs). Introduced by Goodfellow et al. (2014), GANs consist of two neural networks, the generator and the discriminator, which work in tandem. The generator crafts new data samples, while the discriminator evaluates their authenticity. Through iterative training, the generator becomes increasingly better at producing realistic data. GANs have paved the way for machines to generate content that mirrors human creativity, whether it is in the form of photo-realistic images, synthesized art, or lifelike chatbot responses.

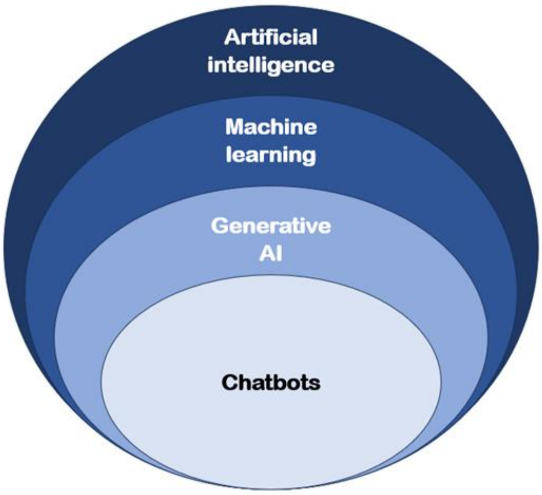


Figure 1 - The field of AI

Chatbots, a notable branch of GenAI, particularly in the context of higher education due to its potential applications as virtual assistants and text-enhancing tools (Chassignol et al., 2018), have greatly benefited from these advancements. Traditional chatbots relied on predefined decision trees or rule-based systems for responses (Shawar & Atwell, 2007). However, with the evolution of GenAI, there have been significant advancements beyond GANs, most notably the development of the Transformer architecture. This architecture, developed at Google Brain by Vaswani et al. (2017), employs a mechanism called "self-attention" that allows it to process whole sequences of words simultaneously rather than one at a time. This method enhances the understanding and generation of contextually relevant text, significantly improving the performance of language models. Thus, alongside GANs, the advent of the Transformer architecture represents another major stride in the development of GenAI, particularly influencing the capabilities and responsiveness of chatbots. Models such as OpenAI's GPT series, known as Large Language Models (LLMs), are based on this Transformer architecture and are not directly influenced by GAN principles. They demonstrate an ability to craft humanlike text. Hence, the integration of GenAI into chatbots represents a shift within AI from simple reactive mechanisms to systems that can proactively generate content, leading to increasingly adaptive, engaging, and humanlike digital interactions (Vinyals & Le, 2015).

2.3 Intersection between GenAI and higher education

In Norway and the Nordic countries, specific literature exploring the intersection of GenAI with higher education is currently limited. To bridge this gap and offer a comprehensive understanding, our analysis extends to a global perspective. This broader approach allows us to identify and understand emerging trends, challenges, and opportunities that GenAI presents in the educational sector.

We will start by reviewing literature on the current use and perceptions of GenAI in higher education, followed by an investigation of its possible applications.

2.3.1 Use of GenAI in higher education today

The introduction of OpenAI’s ChatGPT in November 2022, which notably passed the US bar exam scoring in the 90th percentile (Arredondo, 2023), sparked a widespread debate within the academic community on use of GenAI in higher education. While some institutions such as University of Oslo and University College London are exploring frameworks for application of these technologies (UCL, 2023; UiO, 2023), others have completely banned them due to ethical concerns (Vassdal, 2023). Currently, the integration of GenAI in higher education is still developing, with AI chatbots at the forefront. The primary use of these chatbots has been answering simple queries about class registrations, financial aid, and other administrative matters (Global Admissions, 2023). However, more advanced chatbots are also in development, such as University of Oslo’s student tailored version of OpenAI’s ChatGPT which has added privacy filters (UiO, 2023).

Although GenAI has not been widely integrated into higher education, it has gained substantial traction for personal use. Most notably is ChatGPT, which reached 100 million users just two months after its release (Hu, 2023). A report by Kantar Media (2023) suggests that approximately one in five students (18%) in Norway between the ages of 15 to 24 have used ChatGPT for educational needs (Zulic et al., 2023). Additionally, the EDUCAUSE QuickPoll reveals that 67% of respondents in higher education have used GenAI tools in their work during the 2022–23 academic year, and 83% agree that these technologies will profoundly change higher education in the next three to five years (McCormack, 2023). Moreover, a study in the

United States by the Walton Family Foundation (2023) finds that over half (51%) of teachers and one-third (33%) of students aged 12 to 17 use ChatGPT. Together, such reports indicate a global trend, where many students and professors are already using GenAI tools despite there being no formal integration of them into the educational system.

2.3.2 The effect of digital tools on learning

Munthe et al. (2022) examined the use of digital tools in Norwegian primary and high schools, in order to assess its effect in education. As of 2021, 98% of high school students had access to their own digital computers, presenting a prime opportunity to explore the learning outcomes facilitated by digital tools. The study found that the implementation of digital tools positively affects student learning in Norway. This positive impact is attributed to more engaging and personalized teaching, which in turn increases student motivation. Additionally, the study highlights the benefit of digital tools in providing instant feedback on complex issues, further enhancing learning through such technology. However, the research also points out some challenges, where some students report that the utilization of digital tools has led to technical difficulties. Moreover, students reported that the heightened use of digital tools could negatively impact the communication and interaction within the classroom.

In the context of GenAI, a study by Ayman et al. (2023) at The British University in Egypt provides a relevant perspective. Their research on ChatGPT and its learning implications, reveal similar benefits to those observed by Munthe et al. (2022). These include enhanced student motivation, engagement, and the provision of immediate and personalized feedback. Together, these insights highlight the potential impact of GenAI technologies in higher education, a topic that is particularly pertinent as Norwegian institutions adapt to an increasingly digitalized educational landscape.

2.3.3 Perceptions towards the use of GenAI in higher education today

Understanding the diverse attitudes towards GenAI in higher education is crucial for its effective integration. Darayseh et al. (2023) investigated the perceptions of science teachers in the UAE using the Technology Acceptance Model. Their study demonstrated no significant demographic differences in attitudes toward AI, suggesting a broadly uniform perception among educators in this field. However, the study also emphasized the need for further research into other disciplines and student acceptance of AI, highlighting a possibility of varied attitudes across different educational contexts.

Davis (1989) further highlights that successfully implementing new technology heavily relies on user acceptance. In the context of GenAI in higher education, Chan and Hu's (2023) conducted a survey of 399 undergraduate and postgraduate students in Hong Kong, which revealed an overall positive attitude and familiarity towards this technology, with perceived benefits including personalized learning support and assistance in writing. Yet, concerns about accuracy, transparency, and ethical issues were also noted, underscoring the complexities surrounding the adoption of GenAI.

In contrast, research by Almaraz-López (2023) on Spanish economics and business management students revealed a majority (59%) expressing discomfort with AI concepts, indicating varying levels of familiarity and acceptance across different cultural and educational settings. This study raises important questions about how GenAI is perceived and understood in different academic environments, implying that cultural and educational backgrounds may significantly influence attitudes towards technology.

2.3.4 Applications of GenAI in higher education

Based on our literature review, four applications of GenAI in higher education emerged as the most prominent. These include: (1) GenAI as virtual assistants, (2) GenAI as a tool for grading assignments and exams, (3) GenAI as a tool for text generation of student work, and (4) GenAI as a tool for administrative tasks. The ensuing sections will further explore these.

GenAI-powered virtual assistants

In higher education, the integration of GenAI as a personalized virtual assistant offers several advantages, particularly in emulating the key benefits of personal tutors (Labaddze et al., 2023). Personal tutors have been vital in higher education for providing individualized learning experiences, offering tailored guidance and enhancing student participation in their learning journey (Mollick & Mollick, 2023). In fact, Andreassen (2023) notes that such personalized learning can lead to a 22% increase in student performance. However, the high cost of personal tutors often makes them inaccessible to many students. GenAI presents a cost-effective and scalable solution in this context (Wollny et al., 2021).

While traditional educational settings struggle to offer personalized student feedback at scale (NTNU KTDiM, 2013), chatbots could bridge this gap by working as virtual assistants, offering its broad knowledge to meet the diverse learning needs and interests of students

(Labadze et al., 2023). These virtual assistants also provide instant feedback, essential in large classroom settings where individual attention is limited (Wollny et al., 2021). Further, their 24/7 accessibility allows students to seek help anytime, anywhere, fostering continuous learning and making education more inclusive.

Mollick & Mollick (2023) highlights the effectiveness of using GenAI as a tutor when students craft appropriate prompts, fostering cognitive thinking through interactive dialogues. Intelligent Tutoring Systems (ITS) further enhance this capability, allowing AI to tailor interactions based on comprehensive insights into individual student profiles and learning histories (Crompton & Burke, 2023). This approach is highly scalable, making it accessible to any student with internet access (Mollick & Mollick, 2023).

Although formal integration of chatbots as virtual assistants in higher education is still emerging, some smaller scale case studies have been conducted. One such example is a project conducted at UniDistance Suisse, where an AI tutor application was tested in a psychology class ($n = 51$) during a neuroscience course over a semester (Baillifard et al, 2023). This class was then compared to a similar course without AI assistance. The study found that students frequently engaging with the AI tutor achieved significantly higher grades. In particular, regular interaction with the AI tutor correlated with an average increase of up to 15 percentile points in grades, compared to students in the non-AI-assisted course.

GenAI as an assessment tool

Assessment in education is crucial for evaluating and enhancing student learning (Ramesh & Sanampudi, 2022). Ramesh and Sanampudu (2022) highlight that traditional methods are time-consuming and often lack consistency, a challenge aggravated by increasing student-to-teacher ratios. Historically, Automated Essay Scoring (AES) systems were limited to evaluating multiple-choice tests, but recent advancements in GenAI have enabled possibilities related to assessment and scoring of written essays (Mizumoto & Eguchi, 2023).

The reliability and accuracy of AES have been examined in a study by Mizumoto and Eguchi (2023), exploring the role of AI in enhancing AES. Their findings reveal that AI not only has the potential to reduce time in the grading process, but also ensures greater consistency in evaluation. They further highlight tools like ChatGPT as potentially transformative in writing evaluation and feedback, signifying a significant shift in academic assessment practices. Similarly, Chan and Hu (2023) identify AI as a possible solution to scale and reduce time spent on feedback. However, while these advancements are promising, they also necessitate a consideration of the challenges and limitations inherent in AI-assisted assessments, such as ensuring fairness and addressing biases in AI algorithms.

Moreover, Ayman et al. (2023) investigates the possibility of using AI as a complimentary tool for students and teachers in assessing. With GenAI's potential of providing immediate feedback, students can start the revision process much faster by pinpointing strengths and weaknesses in their work. For teachers, it can be used as a complementary for getting a second opinion in the evaluation process, as well as analyzing specific aspects of a student's text.

GenAI as a tool for generating text for assignments and exams

One of the key benefits of GenAI, as identified by students in Chan and Hu's (2023) study, is its proficiency in writing assistance. This capability extends to correcting grammar, rephrasing and restructuring sentences, and generating text, thereby aiding various stages of the writing process. Harunasari (2023) observes that tools like ChatGPT have proven effective in supporting students from formulating initial ideas to producing well-structured, final drafts. Additionally, Haleem et al. (2023) highlight GenAI's proficiency in verifying grammar and sentence structure, enhancing the quality of written academic work.

Despite these advantages, there are significant concerns regarding the potential academic misuse of GenAI tools. Montenegro-Rueda et al. (2023) highlight the importance of addressing issues related to dishonesty and unethical practices in academia, particularly where AI-generated text may be used inappropriately. This concern is echoed by Ayman et al. (2023), who point out instances where GenAI tools can generate content that is factually incorrect or misleading, all while being technically flawless texts.

The dual nature of GenAI as both a facilitator and a potential challenge in academic writing, underscores the need for a balanced approach in its adoption. Educators and institutions must navigate these complexities, ensuring that while the benefits of GenAI in enhancing student learning and writing skills are harnessed, robust measures are in place to prevent its misuse and maintain academic integrity.

GenAI as a tool for administrative tasks

The application of GenAI in administrative tasks within higher education is gaining increasing attention for its potential to enhance productivity and efficiency. In their literature review, Okonwo and Ade-Ibijola (2021) highlight that there are significant possibilities of streamlining administrative processes in academic institutions through the use of this technology. These AI-driven chatbots are capable of handling a wide range of routine queries from students, including inquiries about admission procedures, scholarship opportunities, tuition fees, class schedules, and more, thereby reducing the workload of administrative staff.

Further emphasizing this potential, Andreassen (2023) draws attention to a report by the McKinsey Global Institute which finds that automation in the higher education sector, particularly in administrative tasks, could lead to substantial cost savings. Specifically, it estimates an annual saving of 200 billion US dollars, largely attributed to reductions in labor costs associated with routine administrative functions.

Finally, the implementation of GenAI in these areas not only offers financial benefits but also contributes to a more responsive and efficient administrative system. By automating repetitive tasks, educational institutions can allocate human resources to more complex and impactful areas, enhancing overall institutional effectiveness (Okonwo & Ade-Ibijola, 2021). The deployment of GenAI-powered chatbots is a step towards a more streamlined, cost-effective, and student-centric administrative model in higher education.

Research question

Based on our literature review, we would like to further investigate the three following applications of GenAI in higher education: (1) text generation for student work, (2) assessment, and (3) virtual assistance. As our priority and research interests primarily lie with student learning, we decided to leave out the aspect related to administrative automation.

3 Methodology

In the upcoming chapter, we will outline our strategy for addressing the research question. This will include choice of (1) research design, (2) data collection, (3) data analysis and (4) ethical considerations.

3.1 Choice of research design

The research purpose of this study can be characterized as both exploratory and descriptive because the research question seeks to map students' and professors' attitudes towards use of GenAI in higher education. Contrary to explanatory research, we do not aim to establish any causalities, but rather describe correlations and characteristics. A descriptive approach is particularly suitable in our context because we found a moderate-to-low amount of academic literature on the topic in the literature review (Saunders et al., 2019). Consequently, we aim to contribute in establishing groundwork for future research to further advance and nuance the understanding of this subject. Additionally, since we do not set out to test any hypotheses, but instead take a bottom-up approach where we attempt to derive descriptions from the collected data, the study can be defined as inductive.

Further, the nature of our research design is primarily quantitative, as we have opted for surveys as the research strategy. The rationale behind this decision boils down to the fact that surveys, particularly those using Likert scales, are traditionally used to gauge opinions and attitudes – making them well-suited for our specific context (DeVellis, 2012; Streiner & Norman, 2014). Moreover, surveys are a preferred method in situations where (1) preexisting data is absent, and (2) one would like to gather quantitative data from a large number of respondents, as is the case in our research (Creswell & Creswell, 2017; Lewis et al., 2016; Sekaran & Bougie, 2016). Additionally, quantitative data provides uniform and quantifiable data, enabling a precise way of identifying and describing patterns, correlations, and characteristics.

3.2 Data collection

In the ensuing section, we will elaborate on (1) our sampling method, (2) the execution of data collection, (3) the questionnaire design and (4) ethical considerations.

3.2.1 Sampling

The population for our study consists of all business and administration students and professors in Norway, while the target population is limited to those at NHH.

Targeting NHH students and professors was a decision grounded in pragmatism. More specifically, we did not have the necessary resources to access contact details for students and professors at other higher education institutions such as University of Bergen or Western Norway university of Applied Sciences. Hence, we would not be able to send out the survey in a systematic and dependable manner.

To sample the target population, we chose a voluntary self-selection method. This decision was primarily driven by a concern of not getting enough respondents had we not made the survey open for all students and professors at NHH. Although there certainly are challenges and limitations with opting for a non-probabilistic sampling method, ensuring sufficient response rate was prioritized.

A further discussion on the implications of our methodological choices regarding sampling will follow in section 3.3 on data analysis methodology, where we will particularly touch upon the risks of self-selection bias.

3.2.2 Primary data collection

The timeframe for the data collection can be described as cross-sectional, because the data from each participant was recorded through completing a single questionnaire (Saunders et al., 2019). Further, the survey period was approximately 17 days, or 2.5 weeks. During this period, we issued a reminder to the participants at the beginning of the second week. When we experienced an entire day without new responses, we decided to conclude the data collection period.

To distribute the survey, we got access to two email lists: (1) one for all bachelor and master students, and (2) another for PhD students and professors. The former had a total of 3305 recipients, while the latter had 615. Consequently, we sent out an email to all the 3920 students and professors at NHH with a link to our survey, accompanied with an invitational text to nudge individuals to participate. Here it should be mentioned that we did not use any economic incentives, such as gift cards or other prizes, to increase response rates. This was a deliberate decision which was influenced by two factors. (i) Firstly, we wanted to keep the survey completely anonymous, for reasons we will touch upon later. Followingly, we could not ask for any contact details to reach out to any "winners". (ii) Secondly, we wanted to avoid a situation where individuals clicked through our survey with the single intention of attaining some sort of "prize". This could have led to rushed and false responses, subsequently distorting our data.

Our self-completed internet questionnaire could be conducted through a computer, mobile phone or tablet. This accessibility increased the likelihood of higher response rates, which we highlighted as notably important. Despite somewhat different formatting due to screen size, the experience was tested across various devices, and was considered as clear and user friendly before sending out to the target population. Although, it should be mentioned that the survey's accessibility also has its drawbacks, such as challenges related to interpretation of questions, verification of respondent identity, time of day, social environment, hunger level, and more (Saunders, 2019). However, we do not consider this a considerable issue granted we have a high number of respondents.

3.2.3 Questionnaire design

Structure

The questionnaire was divided into five main sections: (1) Profile data, (2) attitudes towards GenAI, (3) attitudes towards using GenAI for text generation in higher education, (4) attitudes towards using GenAI as a tool for assessment in higher education and (5) attitudes towards using GenAI as virtual assistants in higher education. More specifically, the survey began by collecting demographic data, then progressed to explore general perceptions of GenAI, and finally concentrated on the three specific areas of interest we outlined at the end of our literature review. See Appendix 1 for a complete view of the questionnaire.

All participants received a uniform survey as the questions were identical. However, we utilized the "display logic" feature of Qualtrics to ask follow-up questions based on the individual answers. For example, in the first section on profile data, we asked participants for their role at NHH, and depending on their response, different follow-up questions were triggered. If the individual responded "Master student", they received a follow-up question on what their profile was. Meanwhile, if they answered "Professor", they would be queried on which department they worked in.

We attempted to make the questionnaire as concise as possible to increase response rates, while including enough questions to get meaningful results. The survey comprised of 23 possible questions, including six follow-up questions. Depending on the responses, the participant would be presented 17 to 22 of these questions, which translated to an estimated completion time of approximately 5 to 10 minutes.

Question design

Likert scales are particularly well-suited for gauging attitudes (DeVellis, 2012; Streiner & Norman, 2014), and were therefore chosen as the primary design throughout our survey. By employing Likert scales, we can to some degree operationalize qualitative constructs such as "attitudes towards use of GenAI in higher education", and get measurable results (Saunders, 2019). Here, we utilized a five-point Likert scale where we posed a statement, followed by response-options which can be characterized as ordinal variables (measuring level). More specifically, we presented the following response-options: (1) "Disagree", (2) "Somewhat disagree", (3) "Neither agree nor disagree", (4) "Somewhat agree", and (5) "Agree".

In addition to the Likert scale questions, we also employed single-choice and multiple-choice questions. The former is suitable for querying demographics such as age and gender where the answer is singular and straightforward, while the latter is utilized when multiple answers can be correct, for questions such as: "What GenAI tools have you used?". The measuring level for these response variables varies depending on the question. For instance, response-options for gender and degree level can be characterized as nominal, while age is a ratio variable by nature (Saunders et al., 2019).

All questions had preset values for participants to choose from. Through pilot testing our survey with a smaller group of students and professors, we attempted to ensure comprehensive lists of response options for all questions (Saunders et al., 2019). Additionally, we added a last response option of "Other (please specify)", for many of the single and multiple-choice questions, in case the presented response options were not adequate for the participant.

Further, when designing each question in our survey, we had to keep in mind best practices for how to formulate ourselves. (1) First, it is important to use simple language, so the questions stay clear and unambiguous (Fink, 2013; Saunders et al., 2019). Given our respondents are business students and professors with varying familiarity with GenAI, a pilot tester of our survey highlighted the benefit of defining terminology related to our subject. We implemented this feedback by introducing each of our sections with concise definitions explaining crucial terms in the upcoming section. (2) Second, we attempted to use unbiased and open-ended language. For instance, one of our single-choice questions where: "What do you consider the most significant benefit of using GenAI in higher education?". This exemplifies a non-leading question. Although, it should be mentioned that for the Likert scale questions where we posed statements, somewhat biased language is unavoidable as the whole premise of the question is to address the participant's relation towards the statement. (3) Third, we also wanted to highlight the importance of avoiding double-barreled questions. Reusing the example of: "What do you consider the most significant benefit of using GenAI in higher education?", we asked a single question, steering away from bundling multiple inquiries into one. (4) Fourth, we employed the technique of posing multiple similar questions from different angles to see whether the responses stayed consistent, or gave counter-intuitive results (Patton, 2002; Podsakoff et al., 2003). For instance, we posed the two following Likert statements: "In a higher education context, implementing chatbots as virtual assistants will benefit student learning" and "In a higher education context, virtual assistants can help students achieve a better grade by serving as a complementary teaching tool". Here, we expect participants to have similar attitudes towards both statements, and the more datapoints we have on the same matter, the better it is for our findings validity. (5) Fifth, it is also worth mentioning a slight variation of the previous technique, where we ask highly similar questions with a single critical point of difference, seeking to attain a nuanced understanding of the dynamics we are investigating, by holding everything else "constant". We will come back to this in the discussion.

3.2.4 Ethical considerations

To conclude the data collection methodology section, we would like to discuss a couple of ethical considerations. We will start by highlighting the survey's anonymity, which was clearly stated in both the invitational email and the first page of the survey. Anonymity was important for us because we did not want participants to be hesitant in giving their true thoughts. We acknowledge that discussing the use of GenAI tools in a higher education context can be a sensitive topic for some students and professors, as it is unclear whether many of the applications are allowed to be used. We therefore wanted to reassure participants that there are no risks to doing our survey, because we cannot trace back information to identify individuals. For this, online surveys are a particularly well-suited research strategy, as a meta-analysis by

Gnambs and Kaspar (2015) finds that online surveys increase respondents' disclosure of sensitive topics. Additionally, it is also worth mentioning that anonymity can increase participation rates (Tourangeau & Yan, 2007), which we previously noted as highly important for us.

Further, we would also like to shed some light on the survey's voluntariness and transparency. Considering there was no monetary reward or "prize" for completing the survey, all respondents who engaged in it, presumably did so out of interest in the topic. Notably, we informed participants that one should only proceed if one wishes to participate in the study, and that one can quit at any time. Moreover, we also stated the purpose of the survey for transparency reasons, as well as providing our contact details should there be any questions. These efforts seek to create a "safe" environment for the participants to answer truthfully, without any limitations, and in an ethically sound manner.

3.3 Data analysis

In the following section we will start by detailing our data analysis methods, consisting of (1) data cleaning, (2) data measurement scale and (3) statistical methods. We will then move on to a discussion of our study's validity and reliability.

3.3.1 Data cleaning

The survey received 274 responses through Qualtrics, where 208 of those are utilized in our analysis. The data cleaning process involved two key criteria:

1. **Time-Based Filtering**: Responses taking under 150 seconds were excluded, based on our estimation that this duration was the minimum required to read, comprehend, and answer the survey questions. This criterion aims to ensure that respondents engaged meaningfully with the survey content, rather than hastily completing it. Notably, this time estimate did not account for reading the AI-related descriptions, assuming that respondents with prior AI knowledge could understand the questions without extensive reading.

2. **Completion Rate**: We also eliminated responses where less than 80% of the survey was completed. This threshold was chosen to ensure a comprehensive understanding and engagement with the survey.

Additionally, we calculated the variance of each respondent and inspected those who score below 0.7. This threshold was chosen due to an observable “gap”, which lead to an additional response removed. In total we removed 55 participants, based on the aforementioned criteria.

Furthermore, we excluded the “other” category from the participant role question, removing 11 responses. This category, mainly consisting of roles like postdoctoral researchers, research assistants, and emeriti, did not fit neatly into our study’s focus on students and teaching staff, thereby not aligning well with our research question.

Post-cleaning, the sample consisted of 208 respondents: 56 bachelor students (26.9%), 88 master students (42.3%), 17 PhD students (8.2%), and 47 professors (22.6%). This refined dataset forms the basis of our analysis.

3.3.2 Data Measurement Scale

In our study, we employed Likert-style questions, which are ordinal in nature. This implies a hierarchical ranking system in our survey, ranging from (1) “Disagree” to (5) “Agree”. Analyzing ordinal data presents unique challenges, as most statistical tests are designed for nominal data.

To conduct parametric tests on ordinal data, we utilized the interval scale assumption. This means that we treat our Likert-data as interval data. In practical terms, for our analysis, this means interpreting the midpoint between (1) “Disagree” and (3) “Neutral” as (2) “Somewhat disagree”. This approach is widely common in dealing with Likert-style data (Wu & Leung, 2017).

Wu and Leung (2017) suggest that increasing the number of Likert points can enhance normality in data analysis. However, a trade-off was observed in our study. Implementing more points in the Likert scale resulted in a less mobile-friendly survey experience in Qualtrics. To balance user experience with data integrity, we opted for a five-point scale, aiming to maximize participant response rates.

An alternative scale ranging from (1) strongly disagree to (5) strongly agree was considered to potentially increase normality. However, given the relatively nascent nature of the technology under study, we anticipated that strong attitudes might not yet be fully formed. It is generally observed that direct experience with the object increases the attitude strength (Jhangiani & Tarry, 2022). Additionally, there was skepticism about the uniformity of intervals between strongly agree, agree, and neutral. Therefore, a five-point scale, without the extremes of "strongly" agree or disagree, was deemed most appropriate for capturing the nuances of participants' attitudes towards emerging technology.

3.3.3 Statistical Methods

In this section, we will outline the statistical tests, conducted in R, applied during the data analysis phase of our research.

Descriptive Analysis

Descriptive statistics, which include both numerical and graphical methods, summarize and present data for easy comprehension (Gudivada, 2012). These techniques focus on central tendency and dispersion in a dataset (Saunders et al., 2019). Central tendency is described using mean and median, offering insights into typical responses. Dispersion is assessed through standard deviation, and minimum and maximum values, highlighting the data's range and variability. Additionally, histograms are used to graphically depict response distributions, especially useful for the Likert-style questions in our survey, as they clearly illustrate the frequency of each response.

Statistical testing assumptions

The validity of our statistical tests rests on three key assumptions, as outlined by Kothe (n.d.):

1. Normality

The assumption of normality requires that the data in each group should be normally distributed (Saunders et al, 2019). To test this assumption, we employed the Shapiro-Wilk test, a recognized method for assessing normality.

As outlined by Mishra et al. (2019), the null hypothesis for the Shapiro-Wilk test is that the data are normally distributed. A p-value less than 0.05 leads to the rejection of the null hypothesis, implying the data are not normally distributed. In our study, the Shapiro-Wilk test was conducted on all Likert-scale variables. The results led to the rejection of all null hypotheses, with p-values below 0.05, indicating a lack of normality.

To further assess normality, we also plotted the residuals of each Likert-scale variable on Q-Q plots. These graphical representations help visualize how closely the data follow a normal distribution. When combined with the results from the Shapiro-Wilk test, these plots provided a comprehensive basis for evaluating normality in our dataset. The conclusion was that none of the variables in our data set satisfied the normality assumption.

Despite the observed deviation from normality, our approach considers the sizeable number of observations in our dataset. The Central Limit Theorem argues that for sufficiently large sample sizes, the distribution of the sample mean will approximate a normal distribution, regardless of the population's distribution (Statistics LibreTexts, u.d.). This theorem generally becomes applicable with a sample size of 30 or more, a criterion met by most of our groups, with the exception of the PhD student group. Additionally, the Welch's t-test and Welch's ANOVA-test is regarded as robust against normality violations (Caldwell et al., n.d .) Hence, while we acknowledge the lack of normality, especially in smaller groups like the PhD students, the large sample sizes in other groups allow us to somewhat mitigate this concern. However, we exercise caution in interpreting results, particularly for the PhD student group, due to their smaller sample size.

2. Homogeneity

The homogeneity of variances assumption for ANOVA states that the variances within each group should be similar (Saunders et al., 2019). To assess this assumption, we conducted a Levene test, which is less sensitive for normality deviations than the more common test, Bartlett's test (McDonald, 2017).

Key assumptions of the Levene test include the random sampling of observations and their independence, with the latter also being a fundamental requirement for ANOVA, which will be discussed separately. The null hypothesis for the Levene test is that all groups have equal variances, implying homoskedasticity (Gatswirth et al., 2009). A p-value below 0.05 leads to a conclusion of heteroskedasticity.

In our study, the Levene test detected heteroskedasticity among some of the variables, leading us to conduct Welch's t-test and Welch's ANOVA, instead of the regular t-test and ANOVA-test, as it does not assume homoskedasticity.

3. Independency

The third assumption for our statistical tests is the independence of observations within each group (Saunders et al., 2019). This assumption implies that the responses (or residuals) from different participants are not influenced by each other and do not exhibit any systematic relationship.

Assessing independence is often context-dependent and can be challenging to test directly. As Kothe (n.d.) illustrates, a common violation of this assumption occurs when participants belong to multiple groups simultaneously, creating interdependencies in responses. In our study, however, such a scenario is not applicable. The distinct roles of our respondents (e.g., PhD students, master's students) are mutually exclusive by their nature, ensuring that a participant in one category cannot simultaneously be in another. This separation effectively upholds the independence assumption for our dataset.

Welch's two sample t-test

A two-sample t-test is a statistical method used to compare two independent groups, in our case, means (JMP, n.d.). The null hypothesis in the t-test is that there is no difference in means between the groups, whereas the alternative hypothesis is that there is a difference in means between the groups. In our study our specific hypothesis is as follows:

H0: $\mu_{Students} = \mu_{Professors/Lecturers}$

HA: $\mu_{Students} \neq \mu_{Professors/Lecturers}$

In our analysis we conduct Welch's t-test, as we detected heteroskedasticity in some of our variables. The Welch's t-test does not assume homoscedasticity, which results in it being more robust towards Type 1 error rates, when heteroskedasticity is detected (Delacre et al., 2017). Additionally, newer research argue for using Welch's t-test by default, compared to the traditional t-test, making it optimal also for our variables where heteroskedasticity was not detected (Delacre et al., 2017; West, 2021).

Welch's ANOVA

Analysis of Variance (ANOVA) is a statistical method used to determine whether there are any statistically significant differences between the means of two or more independent groups. While ANOVA is a parametric test typically suited for nominal data. In our study, we employed the one-way ANOVA, which assesses if the means of two or more groups differ significantly from each other. This test utilizes the F-statistic to compare the variance between groups to the variance within groups, as described by Kim (2017).

We utilized the Welch's ANOVA, as heteroskedasticity was detected in some of our variables. The Welch's ANOVA is an alternative to the traditional ANOVA that does not require the assumption of equal variances, making it more suitable for data where this assumption is violated (Declare et al., 2019). Additionally, newer studies argue that researchers should also use Welch's ANOVA by default, instead of the traditional ANOVA, making it appropriate also for variables where heteroskedasticity is not detected (Delacre et al., 2019).

Our Welch's ANOVA analysis was guided by the following hypotheses:

H0:$\mu_{Bachelor} = \mu_{Master} = \mu_{PhD} = \mu_{Professors/Lecturers}$

HA: at least one group mean is significantly different from the others

A high F-value indicates that the variance between the groups is greater than the variance within them, potentially indicating significant differences in group means and resulting in a low p-value (Saunders et al., 2019).

Post Hoc Analysis

Following a Welch ANOVA test, if the null hypothesis of equal means among groups is rejected, it becomes necessary to employ post-hoc tests (Frost, n.d.). These tests are crucial for identifying which specific groups that statistically significant differ, in terms of means.

Post-hoc tests, as explained by Frost (n.d.), systematically compare the means of each group pair in the dataset to determine statistically significant differences. The hypothesis tested for each pair is H0:$\mu_1 = \mu_2$ (null hypothesis), against HA:$\mu_1 \neq \mu_2$ (alternative hypothesis). This process is identical to conducting multiple t-tests, but with adjustments to account for the increased error rate due to multiple comparisons.

The more comparisons made, the higher the likelihood of a Type I error (false positive; Frost n.d.). While the Bonferroni correction is one method to address this, we chose to use the Games Howell test, because we found it easier to conduct in the R programming language. Additionally, the Games Howell test does not rely on the assumption of homogeneity of variances, making it more suitable for data where the Levene test has detected heteroskedasticity (Rusticus & Lovato, 2019).

3.3.4 Validity

Internal validity

Although internal validity, strictly speaking, refers to the extent to which a study can demonstrate a clear, causal relationship between the independent and dependent variables (Saunders et al., 2019), we can still discuss some relevant points related to our exploratory and descriptive research which does not seek to prove or disprove causality.

First, the cross-sectional timeframe is restricted to a static snapshot (Saunders et al., 2019). This limits our ability to investigate the temporal order of the various variables we have. For instance, we have a question querying whether the participants view GenAI as a progressive step towards the future of higher education, and another one querying how often they use GenAI tools. However, since all the information is captured at a single point of time (survey takes approximately 5-10 minutes), we do not know whether the first variable precedes the second, or vice versa. Although, as mentioned above, contrary to an explanatory research purpose we do not seek to establish causality, but rather describe statistics and correlations. Therefore, we do not consider the cross-sectional timeframe to significantly decrease the study's internal validity.

Second, designing questions for our survey in a way that mitigates biased responses, also directly relates to internal validity because it is tied to the accuracy and trustworthiness of the study's findings within its specific context (Saunders et al., 2019). Above, we detailed five best practices to follow when designing questions: (1) Simple language, (2) non-leading and open-ended formulations, (3) avoiding double-barreled questions, (4) posing multiple similar questions from different angles, and (5) nuancing them to uncover critical points of difference. In the context of avoiding biased responses, we would like to highlight the importance of using (2) non-leading language. By formulating the questions in a neutral manner, we increase the likelihood of participant-responses to truly reflect their true opinions, not distorted by guided formulations (Fink, 2013). Moreover, (4) posing multiple similar questions from different angles further increases the probability that participant-responses match their actual opinions because we get more datapoints on the same matter (Patton, 2002; Podsakoff et al., 2003). Combined, these efforts aim to reduce the amount of response biases, which consequently increases the study's internal validity.

Third, construct validity is also highly relevant when discussing internal validity (Saunders et al, 2019). As mentioned previously, our main approach to operationalize qualitative constructs is by employing Likert scales, due to it being the most natural way to map attitudes in a numeric and quantitative way (DeVellis, 2012; Streiner & Norman, 2014). For example, in the last section of our survey, we focused on GenAI as a virtual assistant tool. Here, we wanted to measure the attitudes NHH students and professors had towards whether they valued human advice above the advice of GenAI, or vice versa. By analyzing the data from our five-point Likert scale, we were able to assess whether there were significant differences between groups of respondents with varying demographic attributes, such as age and degree level. This was done for all Likert scale questions. Moreover, it should also be mentioned that we acquired and analyzed measurable results from single-choice questions as well. Single-choice questions were utilized when we wanted a singular and clear-cut answer, while Likert scales inherently represent more of a spectrum.

Following the thread of construct validity, a notable limitation to our study is that our questions have not been previously tested and validated in other research, which may reduce the internal validity (Groves et al., 2009). However, considering that (1) there is a scarce body of literature to validate from, and (2) that we have attempted to follow best practices when designing the questions, we thought it would still be worth opting for an approach where we *measure* untested constructs, instead of solely conducting qualitative research on the attitudes of NHH students and professors.

Lastly, it is also worth briefly mentioning that we will not go into the discussion of spurious effects, confounding variables, and necessary prerequisites to test causality, as this is beyond the scope of our exploratory and descriptive research purpose.

External validity

In the following section, we will examine to which extent our findings can be generalized to other populations beyond the specific context in which we conducted the research in (Saunders et al., 2019).

Starting with the sample size, our cleaned data set of 208 respondents has a total of 161 students and 47 professors. For each of these two broad categories we consider the number of respondents as acceptable. However, when dividing the students further up into 56 bachelor students, 88 master students and 17 PhD students, we observe that the PhD students are on the lower end of what is considered sufficient. We therefore need to be cautious when we discuss this group's generalizability. The same goes for splitting it in other configurations, such as by master majors and professor departments. However, if we choose to do so, it can still be both interesting and valuable to look at these statistics in some places because they can give us a vague understanding of the broader picture, despite the low generalizability.

The overall response rate for our survey was 6.99%. To calculate the response rate for students and professors, we must use NHH's annual report for 2022 (Norwegian School of Economics, 2022), because the mailing list with 615 recipients bundled PhD students together with professors. Although this report is a year old at the time of writing, we assume the number of positions has not changed greatly. According to this report, there were a total of 84 PhD positions at NHH. Assuming our mailing list followed this number, the response rate for students was 5.29%, and 10,55% for professors. According to Groves et al. (1995), there is no single acceptable response rate that can be applied to all studies, but rather depends on numerous factors such as research purpose and methodological design. In example, exploratory and descriptive research does not necessarily require the same level of precision as explanatory research where one is tasked with proving or disproving causality. Moreover, we did not find any statistics on average response rates for online surveys at NHH, or other higher education institutions in Norway, further complicating the clarity of what to expect. However, based on the results from one of our Likert-questions stating: "I think AI is an interesting field", 97% clicked "Agree" or "Somewhat agree" ($\bar{X}$= 4.73). This gives us a reasonable suspicion of there being some degree of self-selection bias at play here, where those who are particularly interested in conducting our survey also are people with disproportionately higher levels of interest in the technology.

The concern for self-selection bias in our dataset is further fueled by our underlying non-probabilistic sampling method, which did not randomize who received our online survey (Saunders et al., 2019). As mentioned in the section on sampling, the priority was to get a sufficient number of respondents, which could come at the cost of somewhat lower sample-representativeness. However, even with a probabilistic sampling method, there would still be a risk of self-selection bias.

We will end the discussion on external validity here, and further elaborate on the sample's generalizability in the following chapter when discussing profile data and demographics.

3.3.5 Reliability

Reliability refers to the replicability of the findings when the same methods are employed under consistent conditions over time, showcasing the consistency and stability of the measurements (Saunders et al, 2019).

In our research, we have attempted to be considerably transparent and elaborate on our methodology, to ensure a high level of reliability. By detailing everything from the overarching methodological choices, down to specific question-design-techniques and statistical analyses, we enable future researchers to replicate and scrutinize our study. Not only does this strengthen the integrity of our research, but the detailed methods also create a roadmap for subsequent studies that wish to extend our work, which lays at the core of exploratory research (Saunders et al., 2019).

As mentioned in the section on collection of primary data, our self-completed online survey is prone to both participation error and participation bias. These biases not only effect the internal validity as discussed above, but also the reliability. While convenient for respondents, the survey’s accessible nature, allowing participants to take it anywhere and at any time during the survey period, introduces the risk of participation error. For example, variables such as the time of day and the respondent’s hunger level can influence their responses (Saunders et al., 2019). In addition to factors altering the answers, there are also false responses caused by biases. An example of this would be the social desirability bias, where respondents taking the survey in public spaces might feel observed, which leads them to provide answers that they believe are socially acceptable, instead of their true opinions or experiences (Krumpal, 2013). Both participation error and participation bias result in distorted answers, which is an inherent measurement consistency problem, thereby decreasing the study’s reliability. However, we do not consider participation error and participation bias to have a significant effect on our dataset, and we believe that the number of responses will to some degree naturally even things out.

4 Analysis

In this chapter, we explore the data gathered from our Qualtrics survey to address our research question: "*What are students' and professors' attitudes towards implementing GenAI in higher education? How do the perceptions differ between the two groups?*" Our analysis is segmented into five key sections: (1) Profile Data, (2) General Attitudes, (3) Text Generation, (4) Assessment, and (5) GenAI-powered Virtual Assistants. Each section will dissect the responses from single-choice and Likert-scale statements in our sample.

We categorize our participants into two main groups for a layered analysis: (1) students (Bachelor and Master levels, excluding PhD) versus professors, and (2) a more detailed division of Bachelor students, Master students, PhD candidates, and professors. The first layer aims to reveal the overarching differences and similarities between the broader student body and professors, for which we will apply Welch's t-test due to the binary nature of the groups. The second layer seeks to delve into the nuances among the different academic levels, utilizing Welch's ANOVA to accommodate the analysis of four distinct groups. This layered approach is critical for identifying specific needs and perspectives at each academic level, thereby informing the development of nuanced and effective policies.

Mention the overarching structure: We first analyze, then discuss. We do this question by question in our survey, in chronological order.

4.1 Profile data

4.1.1 Demographics

As seen in Figure 2, the gender distribution for our sample size is 38.5% women and 61.5% men. Figure 3 further breaks this down based on the participants' "Role at NHH". The numbers for bachelor, master and PhD students, as well as the professors in our cleaned data set was detailed in the previous chapter, under the section of data cleaning. In the 2022 annual report for NHH, the women-percentage was 42% for bachelor students, 37% for master students, and 50% for PhD students (Norwegian School of Economics, 2022). These statistics closely resemble our data set, with the biggest deviation being 6 percentage points for master students. Further, the 2022 annual report also shows a 17% female proportion among professors, while ours is somewhat higher with 28%. The reason for the 11 percentage points gap is unclear, but one can speculate (1) whether self-selection bias would prefer this segment, (2) whether there is a maternal motivation to "help out young students" with their master thesis, or (3) whether this simply is a result of standard deviation.

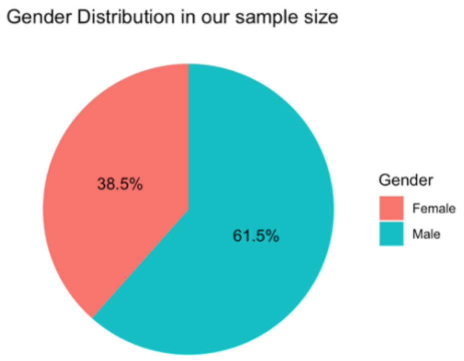


Figure 2 - Overall gender distribution

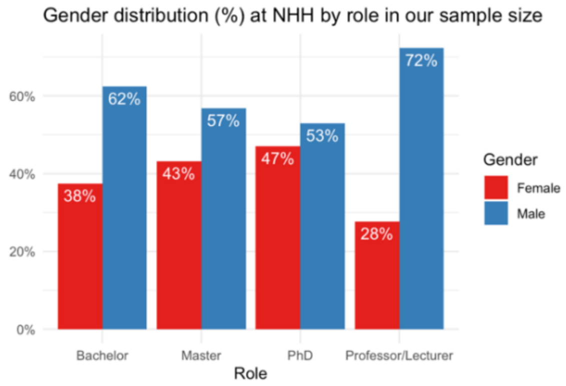


Figure 3 - Gender distribution across roles

Moving on to age, Figure 4 shows a left-skewed distribution towards a young sample size, predominantly below or equal to the age of 25 years. This was expected considering that there are far more students than professors. Additionally, the year Norwegians normally finish upper secondary school, they turn 19 years of age. Given that the master's degree takes five years to complete for most, nearly all students are 25 years old or below, for most of their stay at NHH, even when accounting for "gap years".

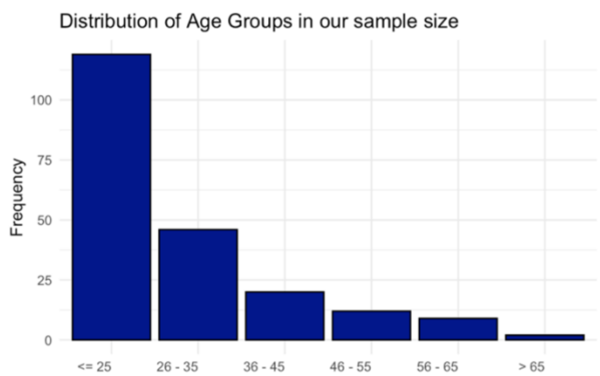


Figure 4 - Age distribution

Further, we would like to highlight the distribution of participants according to major and department affiliation. Firstly, it is important to note that this only applies for master students, PhD students, and professors. Starting with the master students, we observe that Business Analytics (BAN), Business analysis and performance management (BUS), Energy, Natural resources and the Environment (ENE), Financial Economics (FIE), and Strategy and Management (STRAT) are the largest categories with $n \geq 12$ (Figure 5). This is followed by Accounting and Auditing (MRR), consisting of 9 participants. Moreover, we see that the PhD students and professors in our sample are mainly affiliated with the department of Business and Management Science, Economics, and Strategy and Management (STRAT), with $n \geq 13$ (see Figure 6). This is then followed by Accounting, Auditing and Law (AAL), and Finance, with 8 participants in each department. All in all, the sample size is somewhat varied, although there are some seemingly underrepresented groups. These are Economic Analysis (ECO), Economics (ECON) and Marketing and Brand Management (MBM) for master profiles, and Professional and Intercultural Communication (communication) for departments. The true size of these seemingly underrepresented groups is unknown, but we assume they are generally smaller than the largest groups in our sample size. Despite this assumption, we still believe a notable implication of our varied group sizes in relation to major and department affiliation, is that the generalizability for the smaller groups is significantly lower. Although, considering our biggest groups have $n = 15$, we should be cautious with generalizations in general because

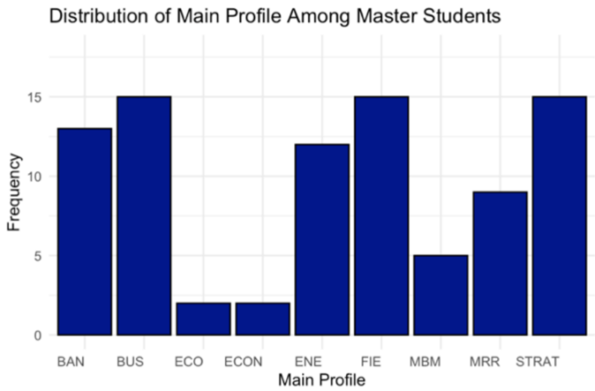


Figure 5 - Distribution of main profiles

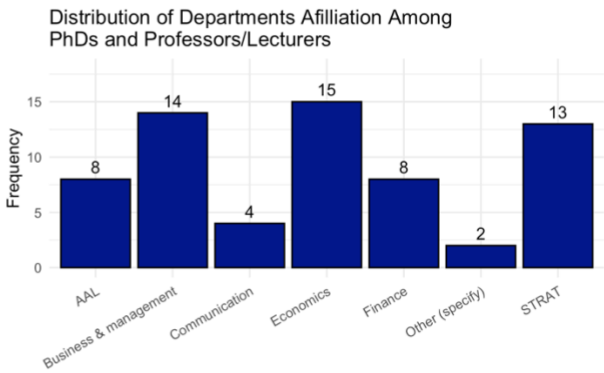


Figure 6 - Distribution of department affiliation

even the largest groups have a rather low number of respondents.

4.1.2 Generalizability

Lastly, we would like to end the section on profile data with a brief discussion on the generalizability of our sample. Now that we have had a closer look at the demographic breakdown, we can continue the discussion we left off in the section on external validity.

As noted in the chapter on methodology, the population for our study consists of all business and administration students and professors in Norway. We chose this population because we believe our target population of students and professors at NHH can be generalized to this broader context. More specifically, we considered (1) geography, (2) field of study, and (3) courses and curriculum as useful aspects to evaluate when formulating a suitable population.

At the broadest level, we found geography to be a convenient starting point when narrowing down our population. More specifically, we have limited ourselves to Norway due to a finding in the literature review which characterizes Norway as a country in a unique position to leverage AI effectively because of its status as one of the most technologically advanced and digitalized nations globally (Astrup, 2020). The claim is further backed by Parmiggiani and Mikalef (2022) who posits that Norway scores well in digitalization because Norwegians are considered early adopters of digital technologies and possess very good digital skills. This presumably also shapes the attitudes students and professors in Norway has towards new digital technology. Similarly, Sweden, Denmark, and Finland, are also considered as particularly well-developed in the digital space compared to other European countries, due to the development of the 4G/5G grid, internet use, ICT competence, and access to public digital services (Nesse & Erdal, 2022). However, since criteria (2) and (3) are not necessarily directly translatable to other Nordic countries to the same degree as this criterion, we decided to limit ourselves to Norway. Although, it is evident that there is no binary answer to what is generalizable and not, but rather a spectrum of options – this logic applies for the ensuing three criteria as well.

Second, we believe that it is reasonable to assume that different fields of study encompass students and professors who have greatly different attitudes towards use of GenAI in higher education. For example, the field of philosophy, electrical engineering, and machine learning all require different skills and personal attributes compared to those needed for business and administration at NHH, which is our target population. We would therefore argue that an individual's field of study significantly shapes their attitudes. One could for instance imagine how an organizational psychology major would share many of the same attitudes with NHH students, while a physiotherapy major would not to the same degree.

This divergence becomes noticeably evident when we think about the course selection and accompanying curriculum for different schools and fields of study. At NHH there is a set course selection and curriculum for all students at bachelor's and master's level because this school only offers the protected Norwegian degree of "civil economist", at the time of writing this in December 2023 (Norwegian School of economics, n.d). Despite there being elective courses as well as major specializations at master's, the education for bachelor and master students is highly uniform. Hence, generalizing our findings to other schools offering "civil economist" as a degree, like BI, which is another business school in Norway, would make the most sense. However, we would argue that the generalizability also extends beyond students and professors involved in a "civil economist"-program. Although the courses and curriculum for other business and administration degrees are somewhat different as there are multiple variations of it, as opposed to the more standardized "civil economist"-program, the course material and topics are fairly similar. Finally, another possible extension would be to generalize to student and professors involved with Industrial Economics and Technology Management (INDØK), which is a popular study program offered at Norwegian University of Science and Technology (NTNU, n.d). However, here the course selection and curriculum start to diverge significantly from the one at NHH, due to the strong emphasis on engineering and technology.

Taking a step back, we observe that there is no clear-cut answer as to what population our target population is representative of, but rather a dynamic discussion of varying degrees of generalizability, depending on how much the population diverges in its attributes. More specifically, we used the three criteria of (1) geography, (2) field of study and (3) course selection and curriculum as a guiding framework for our discussion. In the end, we concluded that our chosen population consists of all business and administration student and professors in Norway. Further, considering the gender, age, major and department affiliation distribution in our sample, we estimate the representativeness to be acceptable in relation to our target population. Hence, we would argue that the findings onwards, stemming from our sample, in fact can give valuable insights into our broader population.

4.2 Usage and general attitudes towards GenAI

In the following section we will (1) analyze the level of GenAI usage among our participants, (2) map the general attitudes towards the topic, and (3) present participants' perceptions on the most significant benefits and challenges of using GenAI in higher education.

4.2.1 Analysis

GenAI usage

Our survey explored the frequency of GenAI tool usage among participants. Remarkably, only 3.4% (7 participants) reported never having used GenAI tools, whereas 64.9% (135 participants) engage with these tools on a weekly or daily basis (Figure 7). Further analysis based on academic roles reveal a distinct pattern where bachelor, master, and PhD students lean towards more frequent usage, compared to professors who have more evenly distributed responses, indicating a varied adoption rate (Figure 9).

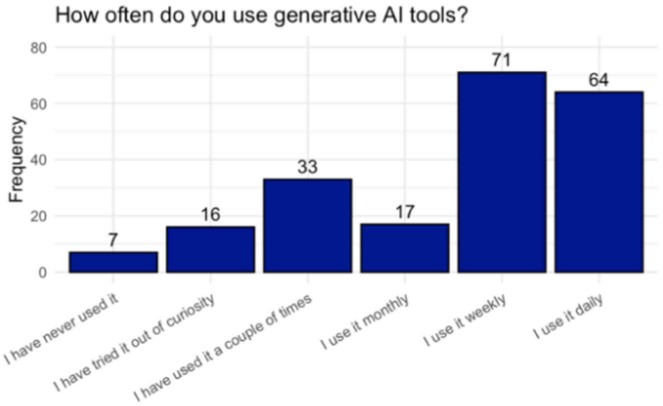


Figure 7 - Overall GenAI usage

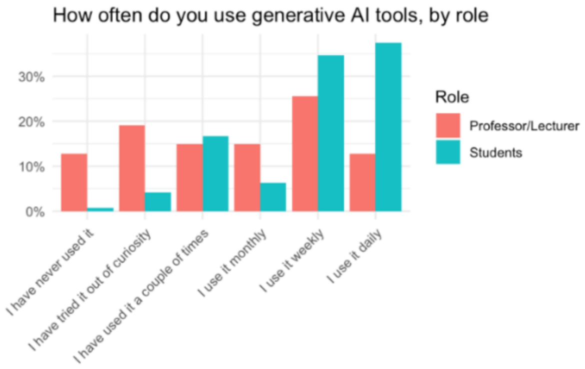


Figure 8 - GenAI usage by students (excl. PhD) and professors

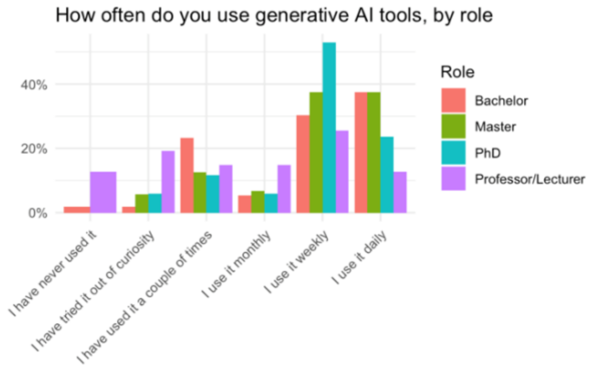


Figure 9 - GenAI usage by academic roles

General attitudes Likert statements

To investigate general attitudes towards GenAI, we had participants rate five statements using our five-point Likert scale. For the first statement, "I think AI is an interesting field", analysis reveals a high mean score of 4.73 ($SE = 0.56$; see Table 1). This suggests a strong agreement among most respondents, implying a keen interest in GenAI. Curiously, no respondent completely disagreed, while only three somewhat disagreed with the statement (Figure 10). Looking at the distribution across academic roles at NHH (Figure 10 and 11), the results of the Welch's t-test ($p = 0.978$) and Welch's ANOVA test ($p = 0.798$) indicate that this interest is uniformly distributed among students and faculty alike, regardless of their specific role.

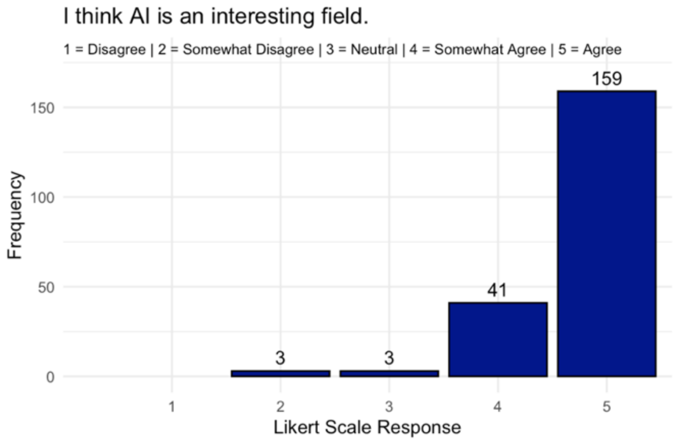


Figure 10 – Interest among all participants

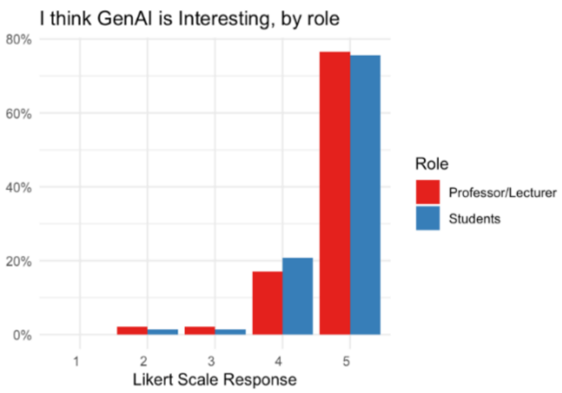


Figure 11 - Interest by students (excl. PhD) and professors

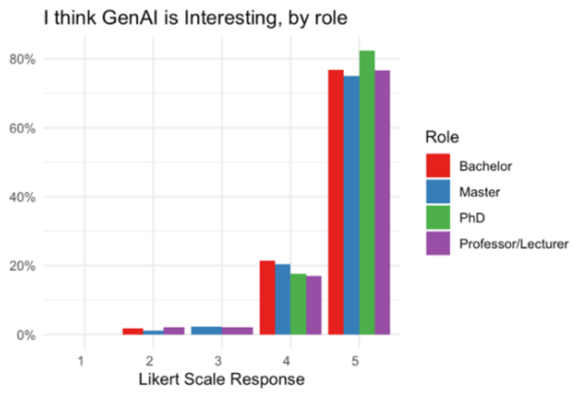


Figure 12 - Interest by academic roles

We also assessed the participants' confidence in using GenAI tools with the statement, "I am confident in using GenAI tools like ChatGPT". Responses gravitated predominantly towards the upper end of the Likert scale (Figure 13), resulting in a mean of 3.89 ($SE = 1.12$; see Table 1). This trend hints at a general confidence among participants in using these tools. A further role-based analysis revealed higher confidence among students ($\bar{X} = 4.04$), as opposed to professors ($\bar{X} = 3.36$). Welch's t-test shows a statistically significant difference in this confidence level within Layer 1 ($t = -3.064, p < 0.01$). Additionally, Welch's ANOVA identified significant variations in confidence levels across academic roles in Layer 2 ($F = 3.750, p < 0.05$), with Games-Howell post-hoc tests highlighting significant differences between master students and professors (difference $= 0.743, p < 0.01$), and between PhD students and professors (difference $= 0.756, p < 0.05$). A notable trend was also observed between bachelor students and professors, though not statistically significant ($p = 0.108$). These findings allude to a generally higher level of confidence in using GenAI tools among students compared to professors.

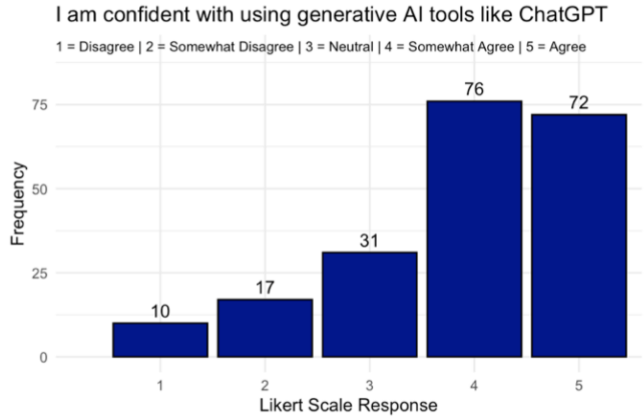


Figure 13 - Confidence among all participants

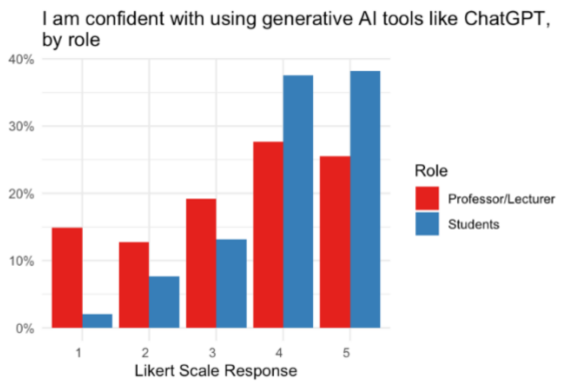


Figure 14 – Confidence by students (excl. PhD) and professors

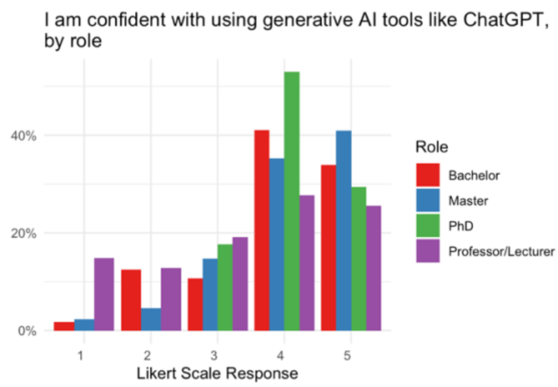


Figure 15 - Confidence by academic roles

In evaluating the statement, "I think GenAI will have a significant impact on higher education," we find a strong consensus among participants who mostly agree, which is reflected in a high mean of 4.62 ($SE = 0.64$; see Table 1). When examining the responses across different academic roles, a variance is observed (see Figure 14 and 15). While the p-value in Welch's t-test is not statistically significant, it is relatively low ($p = 0.09$), suggesting a possible difference in Layer 1. Similarly, the Welch ANOVA test for Layer 2 showed a p-value of 0.070 ($F = 2.386$), indicating a trend towards diverging opinions, although not reaching the conventional 5% significance level. A more detailed examination of the responses in Layer 2 (see Figure 15) indicates that bachelor and master students tend to agree more with the statement compared to PhD students and professors. Despite not being statistically significant, this trend points to differing perspectives on the impact of GenAI between students and faculty. These findings, while not conclusive, highlight a potential generational divide in perceptions of GenAI's role in a higher education context.

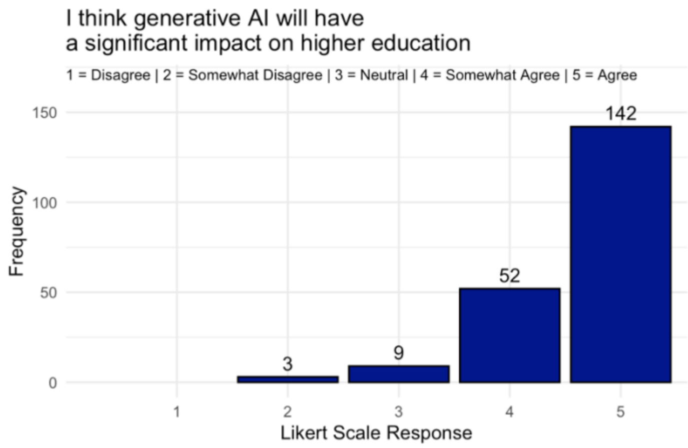


Figure 16 - Overall perceived impact on higher education

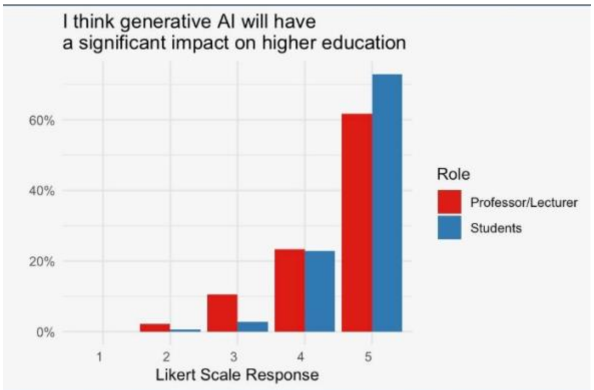


Figure 17 - Perceived impact by students (excl. PhD) and professors

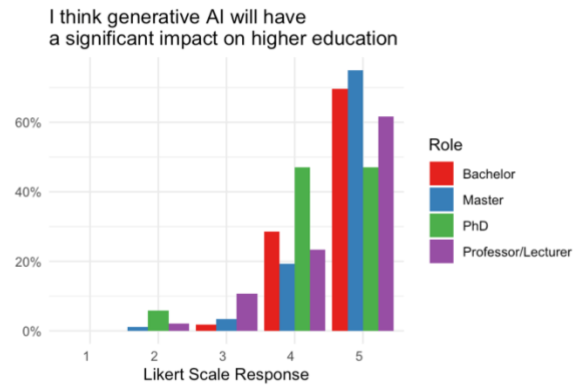


Figure 18 - Perceived impact by academic roles

In our fourth statement, we explored perceptions of GenAI as a progressive step in higher education. The general response suggests agreement, with a mean score of 4.06 ($SE = 1.04$; see Table 1), indicating a positive view of GenAI's role in the future of higher education. Dissecting these responses by academic roles, we observed notable nuances. A Welch t-test revealed a statistically significant difference ($t = -3.064,\ p < 0.05$) between the groups, with students generally holding more positive views than professors. Further analysis in Layer 2 illustrated that bachelor and master students tend to agree with the statement, while PhD students and professors were more reserved, often choosing "Somewhat agree". This distinction was statistically significant, as found by Welch ANOVA ($F = 3.603, p < 0.05$). However, the Games Howell post-hoc test did not identify significant differences among these groups. This lack of specific significant differences, despite an overall significant ANOVA result, implies a general trend towards positivity for GenAI in higher education, but without stark contrasts between specific academic roles. Together, the findings highlight a broadly positive perception of GenAI's potential in higher education across different academic roles, with a slight tendency for more optimism among students.

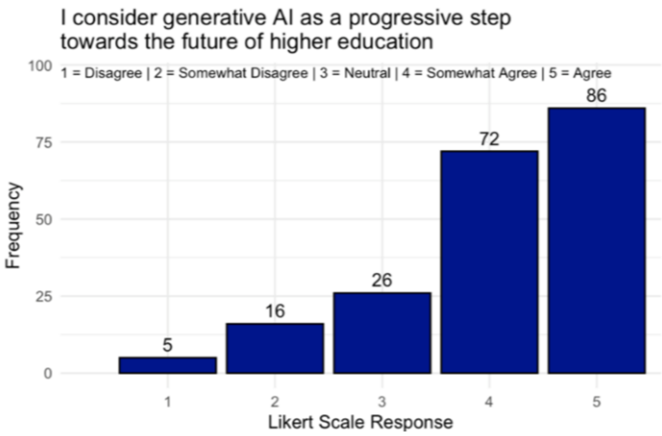


Figure 19 - GenAI as a progressive step, among all participants

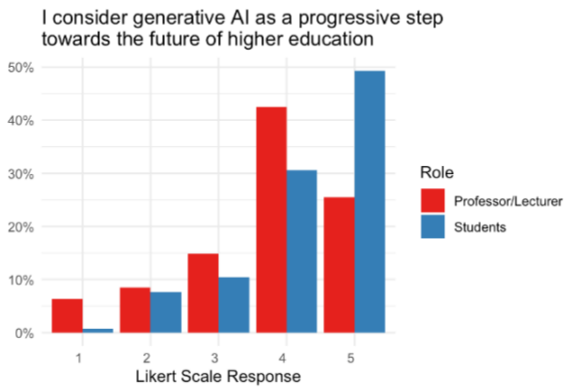


Figure 20 - GenAI as a progressive step, by students (excl. PhD) and professors

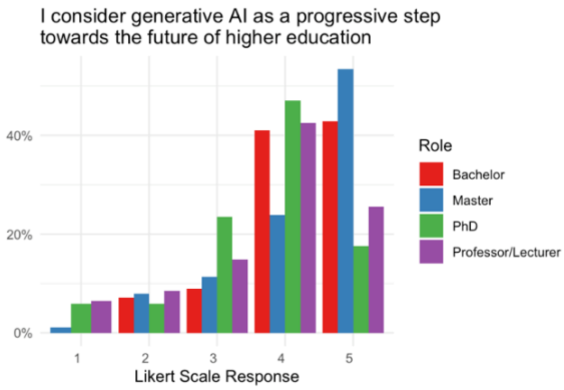


Figure 21 - GenAI as a progressive step, by academic roles

Lastly, we examined perceptions regarding GenAI as a potential danger in higher education. The responses, as depicted in Figure 22 and detailed in Table 1, reveal a divided stance among participants, signaled by a mean score of 3.35 and a standard deviation of 1.28. This suggests a wide range of views on the risks associated with GenAI. Breaking down the responses by academic roles, we find professors ($\bar{X} = 3{,}07$) appear to perceive GenAI as less of a danger compared to students ($\bar{X} = 3{,}39$) who express relatively stronger concerns about its potential risks. However, neither the Welch t-test ($p = 0.866$) nor the Welch ANOVA ($p = 0.125$) showed these differences to be statistically significant, indicating no substantial variation in risk perception across different academic roles.

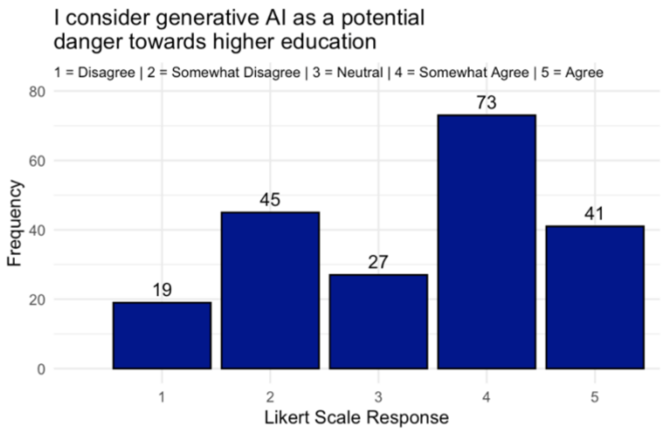


Figure 22 - Perceptions on GenAI's potentially being a danger towards higher education, among all participants

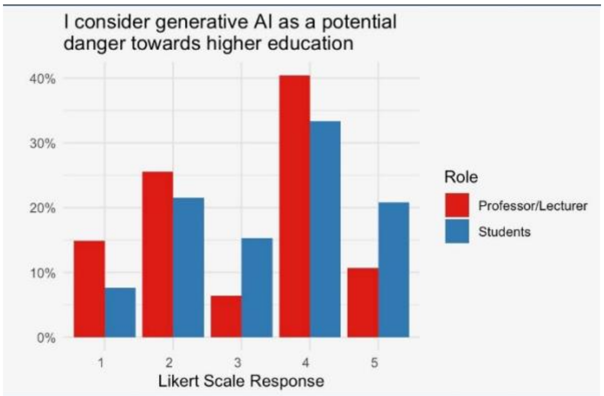


Figure 23 – Perceptions on GenAI's potentially being a danger towards higher education, by students (excl. PhD) and professors

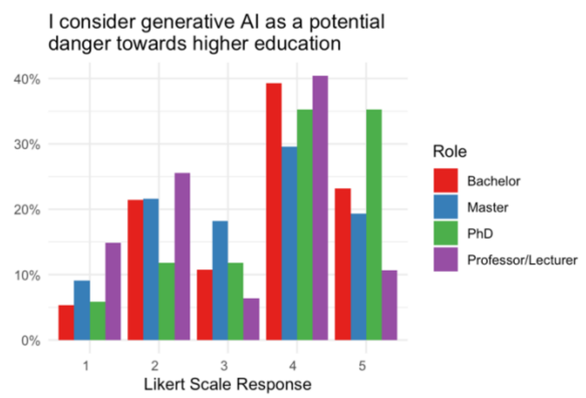


Figure 24 - Perceptions on GenAI's potentially being a danger towards higher education, by academic roles

In a separate single-choice question participants were asked to identify the most significant benefit of using AI in higher education. The respondents chose "Text generation – creating, rephrasing, and restructuring text" as the most significant benefit, while "Assessment – getting instant feedback on exams and assignments" was seen as the least important alternative (Figure 25). Interestingly, this ranking is consistent among both students and professors.

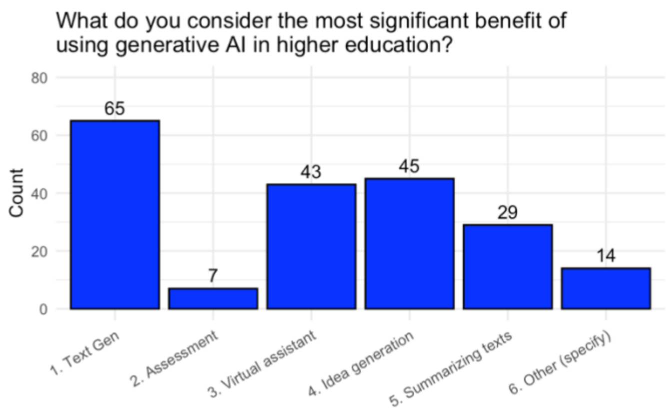


Figure 25 - Most significant benefit, among all participants

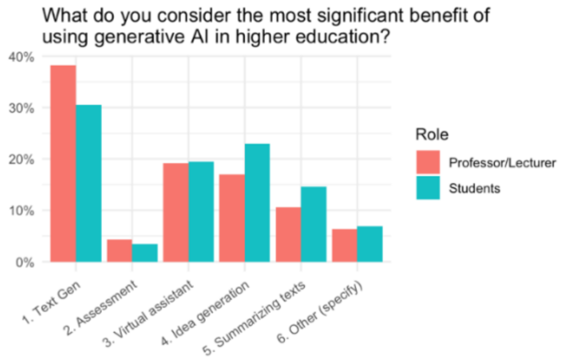


Figure 26 - Most significant benefit, by students (excl. PhD) and professors

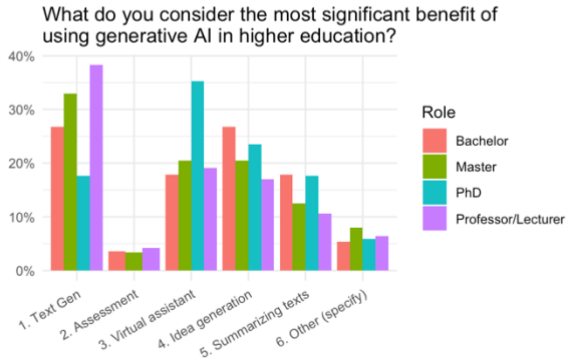


Figure 27 - Most significant benefit, by academic roles

Regarding challenges, participants most commonly identified “Reduced learning for students due to the AI doing the work” as the most significant concern (Figure 28). Additionally, “Receiving non-factual or misleading information due to engine inaccuracy” and “[…] due to inherent biases in GenAI’s dataset” were also highlighted by many. Conversely, “Data privacy” and “Reduced teaching quality due to AI doing the work” were considered as less significant challenges. *Therefore we will not discuss these aspects further in this paper*. The views regarding challenges appear to be shared across students and faculty members, suggesting a common understanding of GenAI’s implications in higher education (Figure 29 and 30).

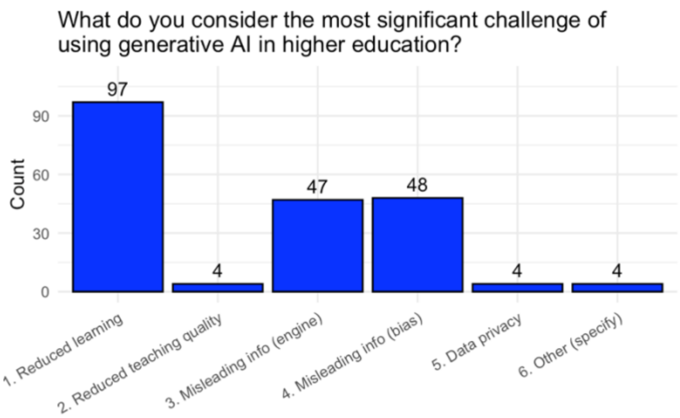


Figure 28 - Most significant challenge, among all participants

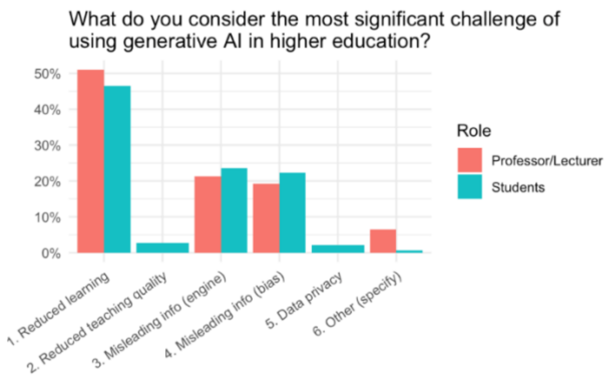


Figure 29 - Most significant challenge, by students (excl. PhD) and professors

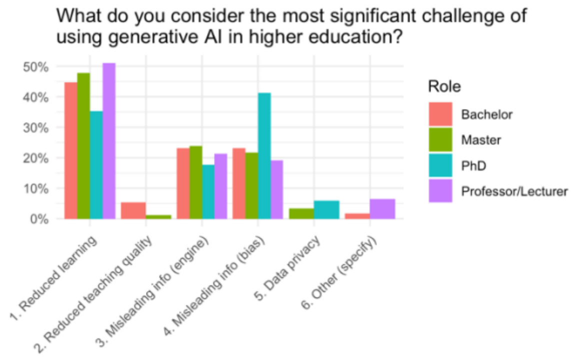


Figure 30 - Most significant challenge, by academic roles

4.2.2 Discussion

GenAI usage

The analysis of GenAI usage at NHH revealed intriguing patterns. A significant majority of the NHH sample, particularly students, are actively engaging with GenAI tools such as ChatGPT, as reported in our survey, either daily or weekly. More specifically, 72.7% of students report using these tools weekly and daily, contrasted with 38.3% of professors. This disparity highlights a potential generational gap in the adoption of emerging technologies in higher education.

In our literature review, we referenced a Kantar Media survey from the spring and summer of 2023, which found that 18% of Norwegian students aged 15 to 24 use ChatGPT weekly for educational purposes. While our survey did not explicitly target "educational needs", the high level of GenAI tool usage among the NHH students - 72.7% compared to 18% in the broader student population – points to more frequent use in NHH's academically and technologically inclined environment. Additionally, the rapid advancement of GenAI technologies since the Kantar survey may have further spurred this increased usage. However, this significant disparity in usage rates hints that NHH-specific factors, like a strong focus on technology in the curriculum or advanced digital literacy, could be influencing this trend. As noted earlier regarding the external validity of our study, it's important to consider that self-selection bias might have affected these findings, particularly for this aspect of the survey.

Regarding GenAI usage among teachers, we found no specific data from Norway to compare with. However, our literature review found a study from the United States reporting that 40%

of K-12 teachers use GenAI tools weekly. While the educational contexts of K-12 teachers in the U.S. and professors at NHH are quite different, the alignment of our findings with this reported figure is notable (38.3% at NHH and 40% for American K12-teachers). It suggests a growing trend of GenAI tool adoption among educators.

In conclusion, the high rate of GenAI tool usage among NHH students, professors and the general student population in Norway, points to a significant trend in the adoption of emerging technologies in higher education. This trend raises important considerations for the integration of GenAI tools in academic settings, potentially influencing teaching methodologies and student learning experiences.

Interest

In assessing the general attitudes towards AI, our survey revealed a considerably high level of interest among our NHH-participants. 77.2% fully agree that AI is an interesting field, with an additional 19.9% partially agreeing. This strong consensus extended across all academic roles at NHH. Contrary to our initial expectations of a moderate interest from professors, there has been a notable engagement with our research. Our interactions with faculty at NHH and the invitation to present our findings at a workshop in Aalborg, attended by professors from four Nordic universities, underscore a genuine academic curiosity and recognition of GenAI's relevance.

The heightened interest in AI may be partially linked to the technology's presence in today's discourse. With the emergence of a new generation of state-of-the-art chatbots, GenAI has been subject to extensive debates, including its use in educational contexts (OECD, 2023). The portrayal of platforms like ChatGPT as a sophisticated tool capable of both promoting learning as well as facilitating academic dishonesty (OECD, 2023), can be one possible explanation for the level of interest among NHH students and faculty alike. Additionally, the broader narrative framing AI as a cornerstone of the "fourth industrial revolution" (Schwab, 2016) and its highlighted benefits across various sectors (McKinsey & Company, 2023) might amplify this interest. For business students at NHH, the emphasis by potential employers, especially in sectors like consulting, on AI proficiency as a desirable skill could further fuel this enthusiasm (Workflow, n.d.).

While we must consider the possibility of selection bias influencing these results, the data nonetheless reflects a significant interest in GenAI at NHH. This interest, spanning students and professors, imply a readiness within the NHH community to engage with these technologies, potentially shaping future educational approaches and research directions. Such widespread enthusiasm could position NHH at the forefront of integrating AI into higher education, aligning with global technological trends and evolving job market demands.

Confidence

Our survey findings indicate a general confidence among participants in using GenAI tools. However, a notable distinction emerges when comparing confidence levels between students and professors. Specifically, we observed a statistically significant higher confidence in using these tools among students compared to professors.

A probable cause for these differences in confidence level can be linked to the reported frequency of GenAI tool usage, where students in particular reported more frequent use. The higher adoption rate could lead to greater familiarity, and thus higher confidence in using these applications. This hands-on experience with GenAI tools, whether in academic or personal contexts, might foster a comfort level that is less prevalent among professors.

In our literature review, studies showed mixed results regarding students' familiarity with AI technology, a factor closely linked to confidence (Fitzsimmons et al., 2020). Chan and Hu (2023) reported a high level of familiarity among students in Hong Kong, while Almaraz-López (2023) found business and education students at the University of Salamanca to be less familiar and even uncomfortable with AI concepts. Almaraz-López et al.'s (2023) study also suggested an average confidence level ($\bar{X}_{confidence} = 3.20$; five-point Likert-scale ranging from "Strongly disagree" to "Strongly agree") in post-graduation AI use. The NHH students tend to align more with the students in Hong Kong for confidence levels, rather than the students at University of Salamanca.

The confidence among Hong Kong students might be influenced by China's prominent role as a leading developer and user of digital technologies, especially AI (Ernst & Schaefer, 2023). Being in an environment where AI development is highly prominent likely contributes to greater familiarity and confidence. While Norway may not be at the forefront of AI engineering, its high level of digitalization, as highlighted in our literature review, could similarly foster a conducive environment for technology adoption in general. This context could contribute to explain why NHH students exhibit a level of confidence akin to their counterparts in Hong Kong, despite different AI development rates.

Curiously, our analysis also reveals that the interest in GenAI among our participants exceeds their confidence in using the technology ($\bar{X}_{interest} = 4.73$ vs. $\bar{X}_{confidence} = 3.89$; see Table 1). This disparity suggests a keenness to engage with GenAI, yet a hesitation or uncertainty in its practical application. This could indicate a gap in GenAI-related education and training. Further, presents NHH an opportunity to enhance its curriculum and provide additional resources to bridge this gap, ensuring that students and professors not only are interested in the topic, but also equipped with the skills and confidence to effectively utilize these tools.

Impact on higher education

In our study, we first examine students' and professors' views on the overall impact of GenAI on higher education, where we find that a majority believe the impact will be significant ($\bar{X}_{Impact} = 4.62$; see Table 1). Here, the data shows that both bachelor and master students perceive this impact as more pronounced than professors do. Secondly, we explored whether the perceived impact is regarded as positive (progressive step) or negative (potential danger). Intriguingly, master students view GenAI as a more progressive step towards the future, compared to professors. However, both these groups seem to be aligned in their perceptions regarding GenAI's potentially being a danger in higher education.

EDUCAUSE published a poll consisting of 100 000 of their higher education faculty and student members, where 83 % agreed that GenAI technology will have a significant impact on the education sector in the following three to five years (McCormack, 2023). Their sample is scattered across the globe, and they do not disclose the fraction of students, professors and others that have participated in this poll. However, the findings imply that the trend seen at NHH is not isolated, indicating a widespread belief that AI will significantly impact the education sector in the coming years. Likewise, the aforementioned Walton Family Foundation (2023) report also posits that students aged 12-17, and K-12 teachers believe GenAI tools will be essential in order to succeed in college.

Echoing the points made in the section on confidence, the varying levels of experience and exposure to GenAI might explain the different perceptions of its impact on higher education. More frequent engagement with the technology typically provides a better understanding of its potential impacts, benefits, and risks. For instance, if users employ GenAI in educational contexts and encounter issues, they may perceive its impact on the education sector as limited. Conversely, those who find it beneficial for their educational needs are likely to view its impact as more significant and as a progressive development. This suggests that the varied perceptions could stem from differing levels of positive and frequent interaction with GenAI, especially among students compared to professors. Furthermore, GenAI usage among professors tends to be more uniform, and the importance of effective prompting techniques, as highlighted by Roca (2023), cannot be overstated. These techniques, which are likely to improve with practice and proper AI training, are crucial for maximizing GenAI's benefits. It's conceivable that professors with limited GenAI experience may not fully appreciate its impact, particularly if they are not using the most effective prompting strategies. For instance, a simplistic approach where professors merely copy and paste exam questions into ChatGPT for evaluation can lead to unimpressive outcomes. This method often fails to leverage the tool's full potential due to a lack of nuanced prompting. In contrast, the application of more sophisticated and contextually relevant prompts, can substantially elevate the quality of the AI's responses. Such enhanced interactions not only demonstrate the capabilities of GenAI tools but also significantly influence their perceived effectiveness and utility in academic settings.

In examining the perceived danger level of GenAI in higher education, our findings show a consistent neutral stance among all academic groups at NHH. Students, who are generally more acquainted with this emerging technology, display a moderate level of concern. This cautious approach seems to be informed by their firsthand experiences with both the strengths and limitations of GenAI, giving them a well-rounded perspective on its potential risks. On the other hand, professors, despite their relatively limited direct experience with GenAI, bring a depth of knowledge about the higher education landscape and a comprehensive academic background. This expertise likely helps them in assessing the possible challenges and threats that GenAI may pose. As a result, their views align with the neutral stance observed among students, reflecting a balanced understanding of GenAI's implications in the academic environment.

Most significant benefit and challenge of using GenAI in higher education

In our survey, we delved into students' perceptions of the most significant benefits and challenges of AI in higher education. The most valued AI feature, according to our findings, was "Text generation – Creating, rephrasing, and restructuring text", followed by "Idea generation - Brainstorming with GenAI". These preferences could be tentatively attributed to the participants' likely exposure to text generation tools such as ChatGPT, suggesting that familiarity may play a role in perceiving benefits. Interestingly, “Virtual assistant - Personalized learning support” emerged as a significant benefit, despite potentially limited direct experimentation by many participants. This pattern mirrors the results of Chan and Hu (2023), who identified “Personalized and immediate learning support” and “Writing and brainstorming support” as prominent benefits in their study of GenAI tools in education. On the other hand, “Assessment - Instant assessments on exams and assignments” was less favored, possibly due to less frequent use among our respondents.

Moreover, the concerns about AI in education are reflective of a broader academic discourse. Students and professors at NHH, expressed worries towards “Reduced learning for students due to AI doing the work”. Furthermore, a significant number of participants also highlighted “Receiving non-factual or misleading information” due to potential inaccuracies in AI’s engine or biases in AI’s dataset, as a challenge of using GenAI in higher education. Likewise, students in Hong Kong, as documented by Chan and Hu (2023), expressed apprehensions about becoming overly dependent on AI and its potential to undermine university education, including concerns on AI’s accuracy. This similarity points to a possible global trend in the academic community, where there is an ambivalence towards AI: it is seen as a valuable educational tool but also as a potential risk to the integrity and depth of the learning experience.

In summary, our investigation into GenAI’s use and overall sentiment reveals enthusiasm and engagement from students, with a slightly more tempered response from professors. While pinpointing exact reasons for this requires caution, these trends may point to a generational divide in adoption of this technology. Nevertheless, this generally positive reception is counterbalanced by concerns about GenAI potentially compromising the quality and integrity of higher education, reflecting wider global concerns and trends (Chan & Hu, 2023).

4.3 Text generation

4.3.1 Analysis

This section aims to outline participants’ perspectives on what they consider permissible chatbot-features in the context of writing exams and assignments in higher education.

The first feature is the use of chatbots for grammar correction. A majority of the respondents appear to endorse this application of AI, with an average approval rating of 4.32 ($SE = 1.12$; see Table 1). Specifically, 62.4% of participants agree and an additional 22.0% somewhat agree, whereas 5.9% disagree and 3.4% somewhat disagree, that AI should be allowed to assist in correcting grammar (see Figure 31). Further, when analyzing responses across different academic roles, we observed a consistent consensus across both layers (1 and 2), as indicated by similar mean scores and response distributions ($\bar{X}_{BA} = 4.12$, $\bar{X}_{MS} = 4.48$, $\bar{X}_{PhD} = 4.18$, $\bar{X}_{professor} = 4.30$; Figure 33). This uniformity in attitude is further supported by the results of Welch's t-test ($p = 0.866$) and ANOVA test ($p = 0.308$), both of which points to no statistically significant difference in opinion across the various roles.

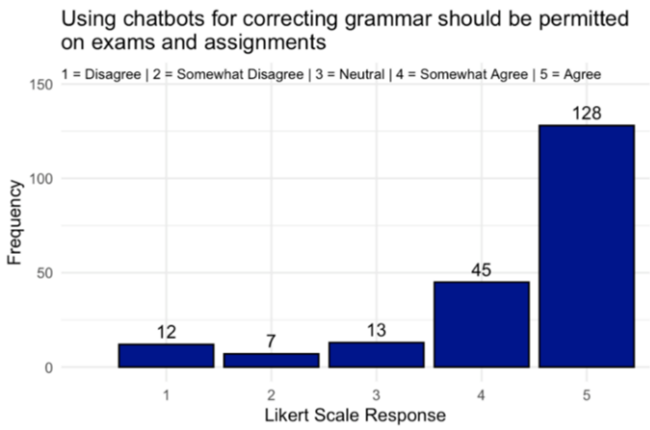


Figure 31 - Perceptions on correcting grammar, among all participants

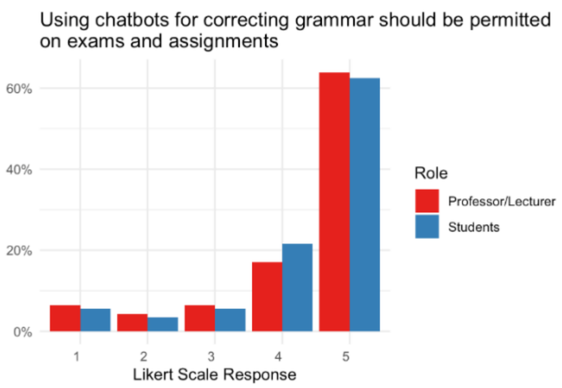


Figure 32 - Perceptions on correcting grammar, by students (excl. PhD) and professors

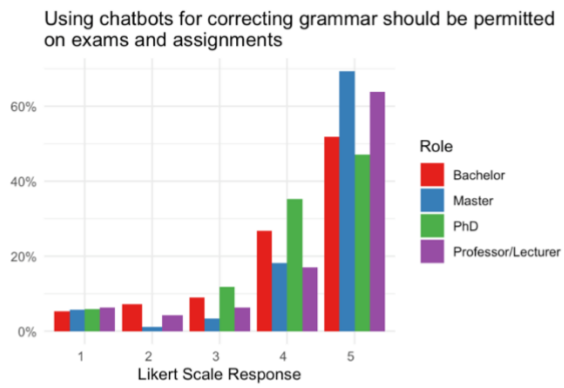


Figure 33 - Perceptions on correcting grammar, by academic roles

For the next feature, we investigated attitudes towards the use of chatbots for restructuring and rephrasing text to improve readability and structure in student work. Most respondents (68.1%) expressed approval (agree or somewhat agree) for permitting this feature in exams (Figure 34). However, this approval is marginally lower than for grammar correction, which is reflected in a mean score of 3.74 ($SE = 1.38$; see Table 1). Moreover, when analyzing differences between students and professors we observe minimal differences, as indicated by their means ($\bar{X}_{students} = 3.82$ versus $\bar{X}_{professors} = 3.67$) and Welch's t-test ($p = 0.533$). Delving into

Layer 2, the data reveals varying degrees of acceptance. PhD students exhibit a more neutral stance ($\bar{X} = 3.18$), showing less agreement towards this feature than bachelor ($\bar{X} = 3.76$), master ($\bar{X} = 3.86$), and professors ($\bar{X} = 3.67$). Despite these variations, the Welch's ANOVA test concludes with no statistically significant differences ($F = 1.21, p > 0.1$), suggesting a generally uniform perception across different academic roles at NHH regarding the use of chatbots for text restructuring and rephrasing in exams.

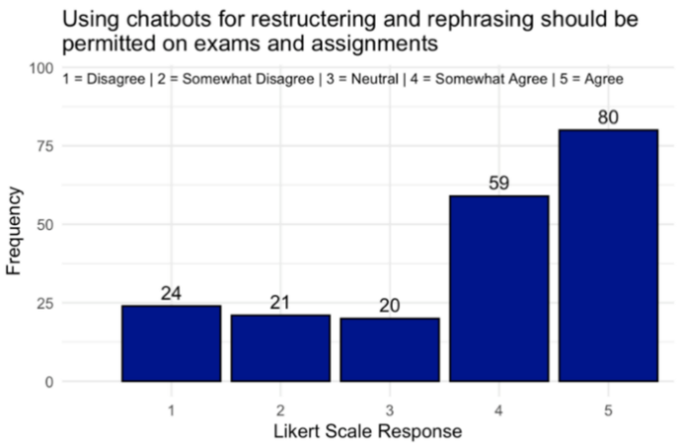


Figure 34 - Perceptions on restructuring text, among all participants

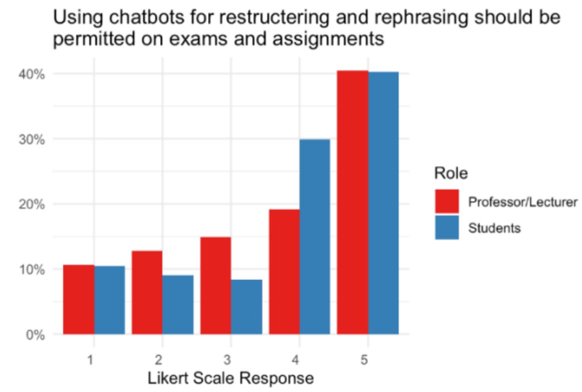


Figure 35 - Perceptions on restructuring text, by students (excl. PhD) and professors

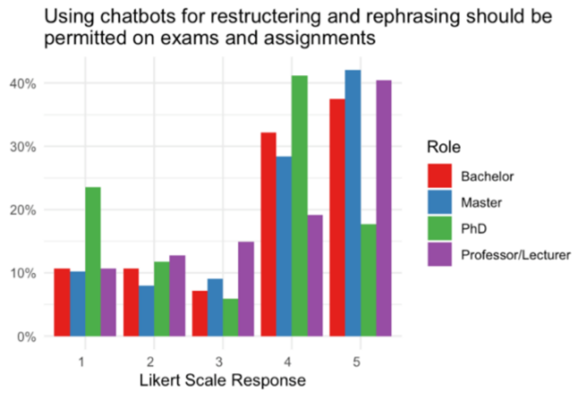


Figure 36 - Perceptions on restructuring text, by academic roles

Third, we examined the acceptability of utilizing GenAI tools for idea generation in academic work. The responses, as shown in Figure 37, are generally evenly distributed across the scale, with a slight tendency towards agreement, reflected by an average score of 3.20 ($SE = 1.42$; see Table 1). This distribution suggests a balanced yet cautiously positive view on the role of GenAI in facilitating idea generation. Furthermore, analyzing by academic roles, as seen in Figure 38 and 39, reveal variance in opinions among groups in Layer 1 and 2. A low p-value ($= 0.068$) in Welch's t-test indicates differences between students ($\bar{X} = 3.326$) and professors ($\bar{X} = 2.867$), hinting that students may be more receptive to the idea of AI-assisted idea generation. However, the p-value extracted from Welch's ANOVA ($p = 0.290$) does not support a significant variance across the broader layers, implying a more commonly shared sentiment among different academic roles.

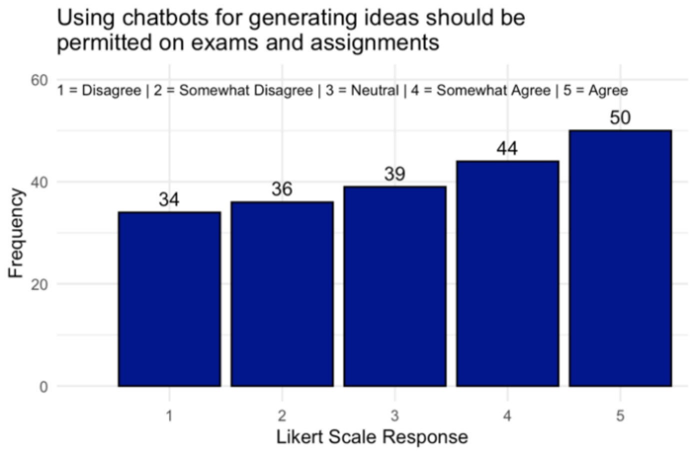


Figure 37 - Perceptions on generating ideas, among all participants

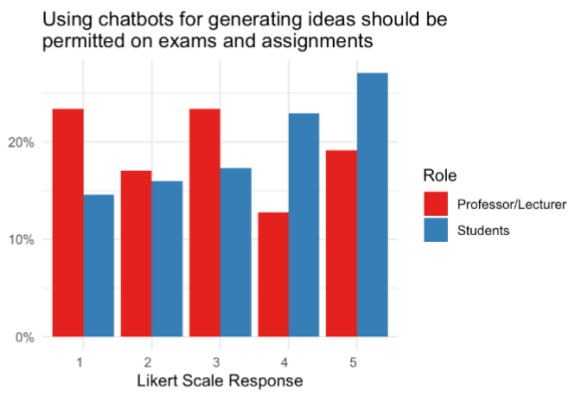


Figure 38 - Perceptions on generating ideas, by students (excl. PhD) and professors

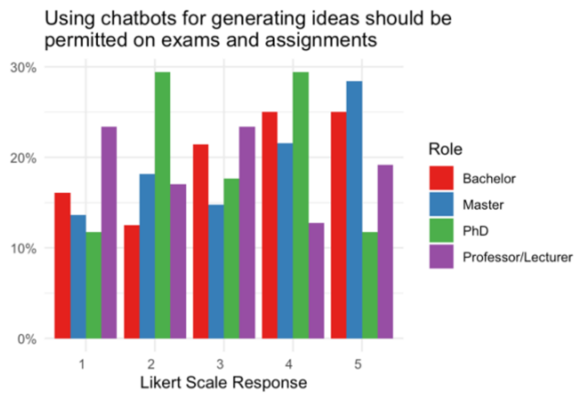


Figure 39 - Perceptions on generating ideas, by academic roles

In assessing attitudes towards allowing all functions of GenAI, including complete text generation for exams and assignments, we observed a predominant disapproval. As shown in Figure 40, 46.3% of respondents disagreed, and 18% somewhat disagreed, resulting in a low mean score of 2.20 ($SE = 1.41$; see Table 1). Further, when analyzing the responses in Layer 1, we found relatively small differences in opinion as indicated by Figure 41 and a non-significant Welch t-test ($p = 0.514$). However, in Layer 2, significant differences were observed – Nearly 80% of PhD students disagreed with complete text generation (see Figure 42), which is a noticeably higher rate than bachelor students (39.2%), master students (43.2%), and professors (48.9%). The Welch's ANOVA confirmed significant variances in opinions ($F = 3.48, p < 0.05$), with Games-Howell post-hoc tests pinpointing the differences to bachelor and PhD students ($\text{difference} = 0.887, p < 0.05$) and master and PhD students (difference $= 0.808, p < 0.05$). The results find that PhD students are more opposed to allowing full use of GenAI for academic work compared to bachelor and master students.

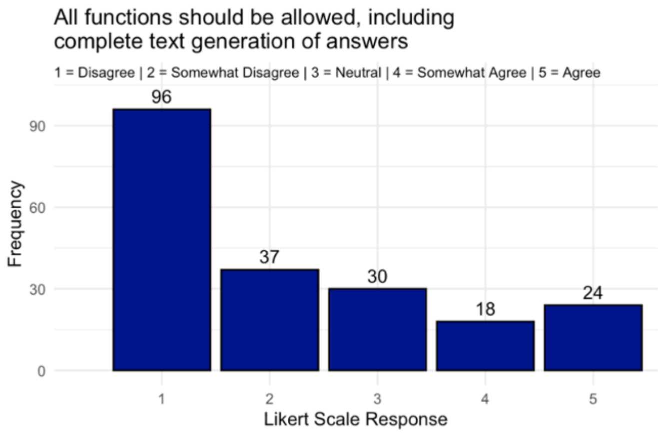


Figure 40 - Perceptions on all functions allowed, among all participants

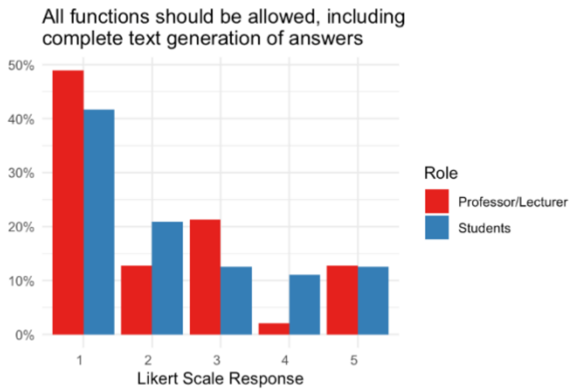


Figure 41 -Perceptions on all functions allowed, by students (excl. PhD) and professors

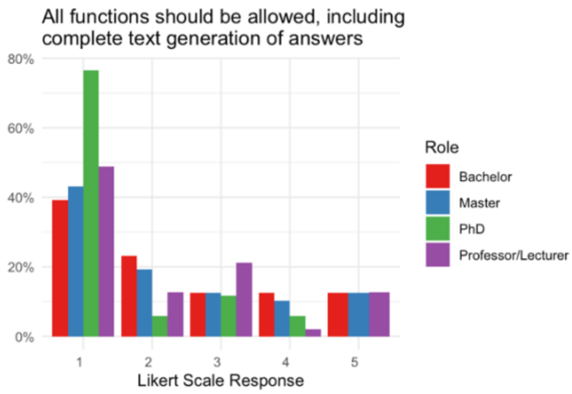


Figure 42 - Perceptions on all functions allowed, by academic roles

Conversely, when evaluating the stance on completely disallowing chatbot features, “No features should be allowed - Chatbots should not be permitted at all”, the average response leaned towards disagreement ($\bar{X} = 1.97$; $SE = 1.23$). A majority of 51.4% disagreed, and 19.6% somewhat disagreed, as opposed to 13.2% approving (5.9% agree and 7.4% somewhat agree). This trend implies a reluctance to totally prohibit chatbot use (Figure 43). Additionally, both the Welch t-test ($p = 0.514$) and ANOVA test ($p = 0.493$) revealed no statistically significant differences across academic roles, suggesting a general consensus against the complete ban of chatbot features for student work.

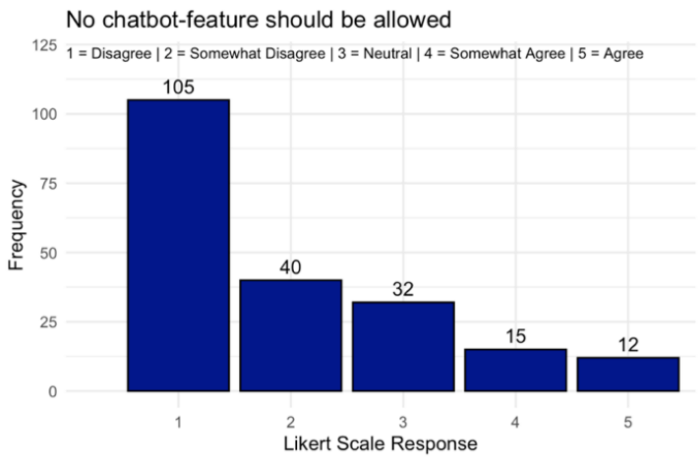


Figure 43 - Perceptions on no functions allowed, among all participants

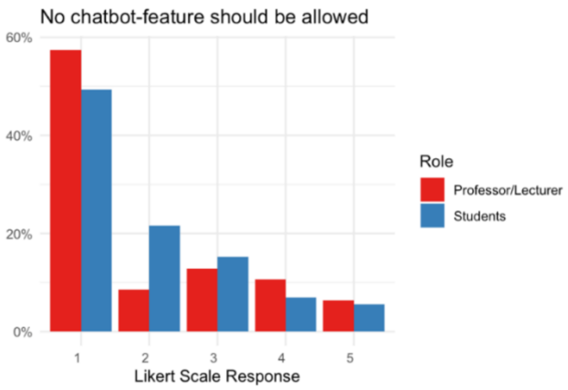


Figure 44 - Perceptions on no functions allowed, by students (excl. PhD) and professors

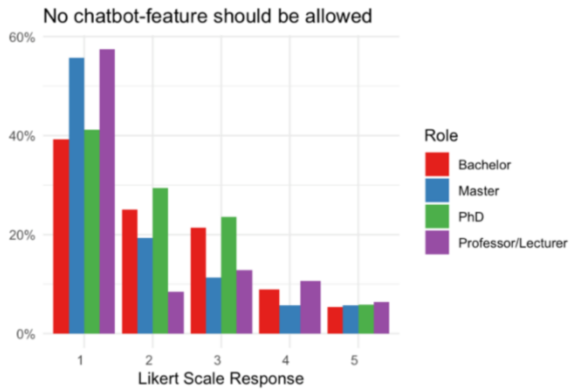


Figure 45 - Perceptions on no functions allowed, by academic roles

4.3.2 Discussion

The use of GenAI for text generation in student work is perhaps one of the most prominent topics in contemporary discussions when it comes to GenAI in education. As detailed in the survey, the purpose of this section is to map where students and professors draw the line between which features should be allowed during exams and assignments, and which features should not be allowed. According to our literature review, key capabilities of GenAI in writing assistance are: (1) grammar correction, (2) rephrasing and restructuring text for increased readability, (3) idea formation and (4) complete text generation (Chan & Hu, 2023; Haleem et al., 2023; Harunasari, 2023). Using the chronological order above, we attempted to make a spectrum which captures the extent to which GenAI interferes with student work, ranging from minor interference to complete text generation. Additionally, we had a last Likert statement addressing whether no GenAI features should be allowed at all.

Starting with the least amount of intervention, findings from our analysis suggest a broad acceptance for both students and professors regarding the use of GenAI for grammatical assistance. This was also true when we split students up into bachelor, master, and PhD, which reflects a consistent acknowledgment of the practical utility of GenAI in enhancing the quality of written work, regardless of the respondent's academic level. Considering that programs like Word already incorporate grammar correction tools, the broad acceptance is not surprising, as both students and professors are accustomed to using such features. However, while most respondents were positive to grammatical assistance, we must acknowledge the 9.56% who either disagreed or somewhat disagreed. This heterogeneity in attitudes highlights an important consideration when making future policies, which we will discuss further later in this section, as well as in the chapter on policy-making.

Moving on to the use case of rephrasing and restructuring text for increased readability, we still see a clear majority having a positive stance towards this application of GenAI. However, we curiously observe that the mean score for this is somewhat lower than for the previous question. This is consistent with our assumption that bigger interventions and text alterations will meet more opposition. A possible explanation here is that students and professors may perceive higher levels of assistance as less reflective of the students' true competence. On one hand, this concern is valid since increased automation could lead to less work and critical thinking. For instance, rephrasing and restructuring text for increased readability might be considered a crucial skill for students to learn, and if automated, students might not develop this skill to a sufficient extent. On the other hand, one could argue that the benefits of automation outweigh the drawbacks, because the total amount of work is not diminished, but rather redirected to more important areas. More specifically, one could imagine that assistance in text formatting enables students to spend more time on problem solving and decision-making. This discussion point is salient when making policies, because at its core, the question is what skills higher education institutions should prioritize in student development.

In contrast to the two previous GenAI features, generation of ideas introduces a more advanced level of assistance, emphasizing the creation of entirely new content. This shifts the focus from modifying existing student work to generating original ideas. Our analysis shows that opinions on this feature are evenly mixed, with a slight preference towards its use. Building on the previous discussion, it is arguable that generating ideas is a crucial aspect of student work, unlike text alteration which may be seen as more secondary. Consequently, this raises the question of whether idea generation is where one should draw the line for what is considered appropriate use of GenAI. On the other hand, it could be argued that using GenAI for idea generation is not about outsourcing a vital part of student work, but rather engaging in a collaborative process. In this process, students must apply critical thinking to decide which ideas to adopt and which to discard, using their human judgment to assess the ideas within the larger context of their task. Additionally, it is worth noting that GenAI could accelerate the ideation process by suggesting ideas that students might eventually think of themselves. This could help in overcoming writer's block, which is defined by Rose (2006) as the inability to start or continue writing due to reasons other than lack of skill or commitment. In other words, one could argue that GenAI can be a tool for enhancing, rather than replacing, the student's creative process.

Investigating the highest degree of GenAI involvement, we mapped students' and professors' attitudes towards allowing all GenAI features, including complete text-generated student work. Here, our analysis found a majority disapproving of this approach. Contrary to idea generation, this type of tool would not only allow generating new ideas and content, but also apply it directly in the context under question. Although one certainly could make this more collaborative and engaging by iterating the generated text using more prompts, it seems like the participants in our survey perceived this level of involvement as too overreaching. Intriguingly, we observe from the section on general attitudes that text generation was considered the most significant benefit of using GenAI in higher education, while reduced learning was regarded as the most significant challenge. Keeping in mind the general disapproval of allowing all GenAI features for text generation, it appears that respondents might perceive this level of GenAI presence as a threat to student learning, but would still like some features to be allowed.

Lastly, looking at the attitudes towards banning all chatbot-features, we see a slightly stronger disagreement for this statement compared to the previous concerning allowing all features. As touched upon above, this indicates a general preference for a balanced approach, favoring neither the total exclusion nor the unrestricted use of GenAI. However, it is not exactly clear as to where the line should be drawn. Taking a utilitarian approach, one could for instance make the decision based on which features had a mean leaning towards agreement, which in our findings would include grammatical correction, text alteration and idea generation. Nonetheless, this approach does not take into account two critical considerations: (1)

heterogeneity of answers and (2) the institution's goals for student learning. (1) Despite some statements having noticeably low or high mean scores, such as banning all features, allowing grammatical correction, and allowing all features, we curiously acknowledge that there still is a crowd of respondents who disagree with the majority vote. It is therefore important to further investigate and understand their point of view before implementing guidelines for the use of GenAI. (2) Secondly, it is essential to recognize that the policy-making cannot be exclusively guided by students' and professors' attitudes, but must also ensure that it serves the broader goals of student learning and development for the respective higher education institution.

In summary, our research finds a consistent decrease in approval ratings as more GenAI text generation features are introduced for use in exams and assignments. This trend supports the notion that increased GenAI involvement tends to meet with greater resistance. Further, the decision of which features to permit is complex due to the mixed attitudes of students and professors, as well as the need to align with the educational objectives of the respective higher education institution. This involves balancing the significance of different student skills, such as text alteration compared to idea generation, within the assessment criteria. Additionally, one could raise the question of whether the current goals for student development are keeping pace with technological advancements, particularly with the emergence of GenAI. This will be further discussed in the chapter on policy-making.

4.4 Assessment

4.4.1 Analysis

In this section, we aim to understand the attitudes of participants at NHH towards the use of GenAI as a tool for assessing assignments and exams in higher education. Here, we start by mapping professors' and PhD students' current use of this application, followed by an investigation of participants' perspectives on using it to assess graded and non-graded student work. Lastly, we will examine respondents' attitudes towards the fairness of using GenAI as an assessment tool.

Starting with the current use of GenAI for assessment, our survey sought insights into how professors and PhD students at NHH use GenAI as a complementary tool for (1) assessing assignments and exams, as well as (2) crafting questions for these assessments. The responses, detailed in Figure 46 and 47, indicate a low usage-level among these two groups. Interestingly, while no participants reported experimenting with GenAI out of curiosity for question-making, 24.5% acknowledged using it on a few occasions. This suggests that initial trials of GenAI in assessment contexts lead to repeated, though not extensive, use.

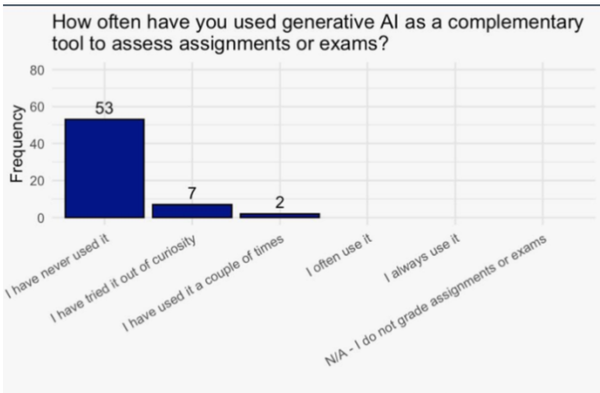


Figure 46 - GenAI as a complementary tool for assessing

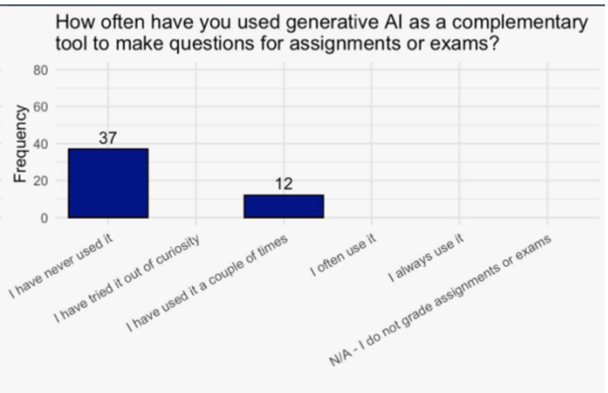


Figure 47 - GenAI as a complementary tool for question-making

Moving on to participants' attitudes towards using GenAI to assess graded and non-graded student work, we start by analyzing the use of GenAI as a complementary tool for assessing (1) graded and (2) non-graded exams and assignments. Regarding graded assessments, responses were mixed, with an average score of 2.93 ($SE = 1.33$; see Table 1). Notably, "Somewhat agree" emerged as the most frequent response, while "Agree" was the least common (Figure 48). For students and professors there seems to be small differences, in terms of distribution and means (Figure 49 and $\bar{X}_{students} = 2.88$, $\bar{X}_{professors} = 3.00$), signaling some openness to AI assistance, although with reservations. The small differences were not statistically significant, according to Welch's t-test ($p = 0.587$). An examination of the responses in Layer 2 (Figure 50) suggests that bachelor students are relatively less supportive of this application of AI, as indicated by their lower mean score compared to other groups ($\bar{X}_{BA} = 2.68$, $\bar{X}_{MS} = 3.01$, $\bar{X}_{PhD} = 3.12$, $\bar{X}_{professor} = 3.00$). However, the Welch ANOVA test, similarly to the Welch's t-test, found no statistically significant variation in attitudes across the groups ($F = 0.93, p = 0.466$).

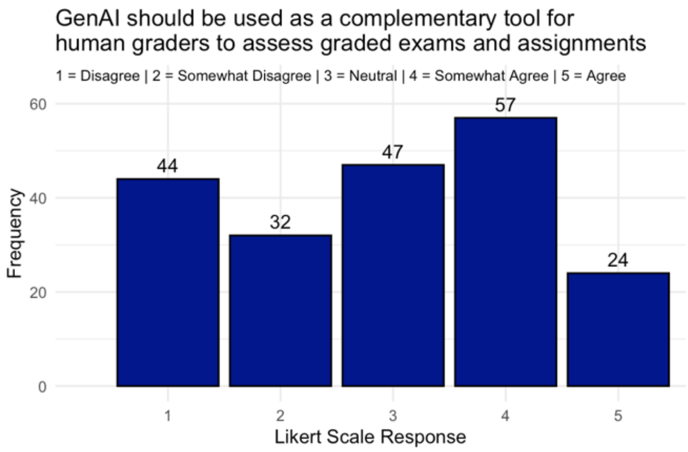


Figure 48 - Perceptions on using GenAI as a complementary tool for graded exams, among all participants

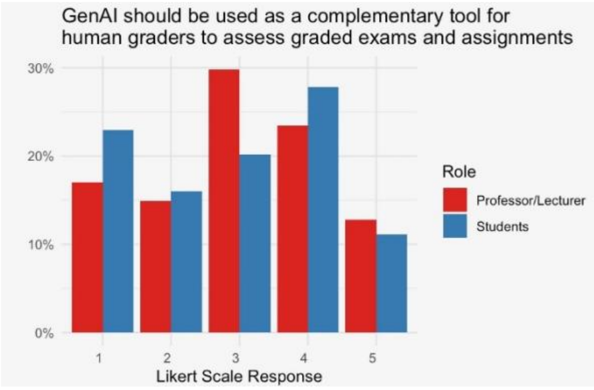


Figure 49 - Perceptions on using GenAI as a complementary tool for graded exams, by students (excl. PhD) and professors

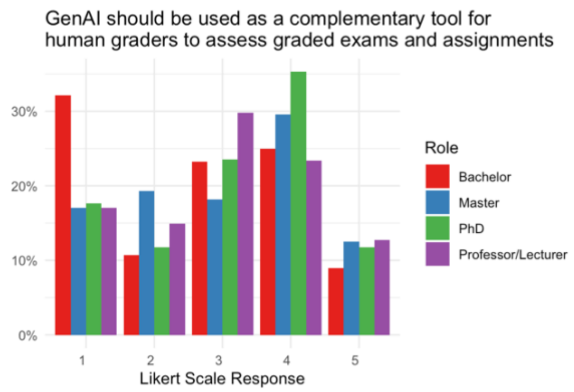


Figure 50 - Perceptions on using GenAI as a complementary tool for graded exams, by academic roles

In contrast, responses for non-graded assessments were more favorable, with an average approval rating of 3.45 ($SE = 1.30$; see Table 1). This positive inclination is visually represented in Figure 51, where the histogram displays a rightward skew, contrasting with the more evenly distributed responses for graded assessments (Figure 48). Further, when examining across roles, we observe a slight variance in opinions between students and professors (Figure 52), with students showing more agreement, this difference was not statistically significant as per Welch's t-test ($p = 0.301$). In Layer 2, master students appeared slightly more acceptable to the idea than their counterparts (Figure 53). However, similar to the Welch's t-test, the Welch ANOVA test indicated no statistically significant differences in attitudes ($p = 0.215$).

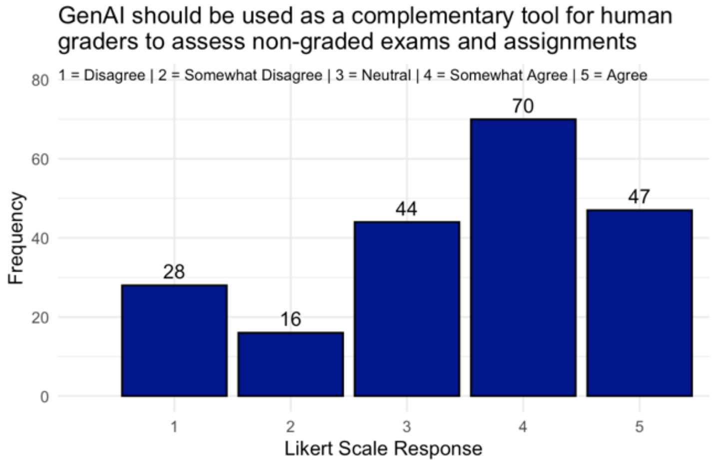


Figure 51 - Perceptions on using GenAI as a complementary tool for non-graded exams, among all participants

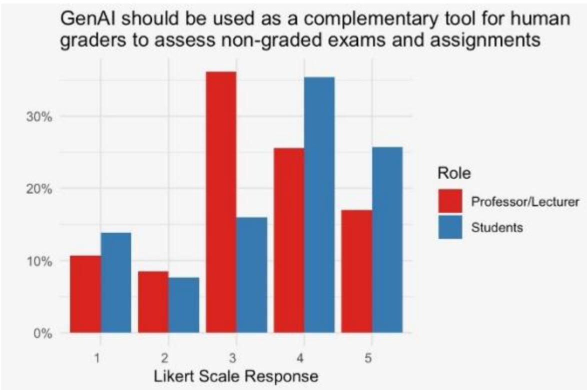


Figure 52 - Perceptions on using GenAI as a complementary tool for non-graded exams, by students (excl. PhD) and professors

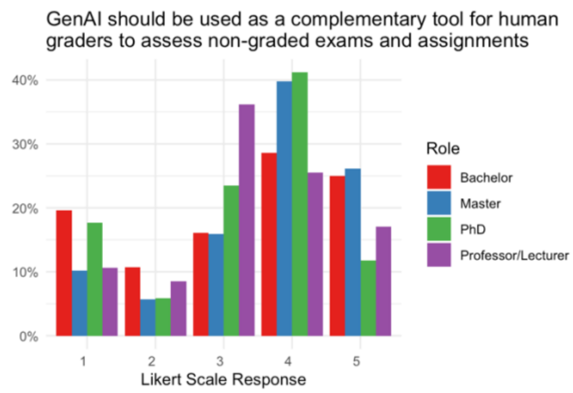


Figure 53 - Perceptions on using GenAI as a complementary tool for non-graded exams, by academic roles

In parallel to our exploration of text generation, we assessed participants' comfort levels with GenAI's involvement in assessment. We stated: "GenAI should be used to assess graded exams and assignments without human intervention". The majority of respondents were clearly opposed to this statement, which was reflected in a low mean score of 1.49 ($SE = 0.88$; see Table 1), where 82% of respondents disagreed (Figure 54). Moreover, when examining the distribution across roles in Layer 1, we observed homogenous responses (see Figure 55), supported by a high p-value in Welch's t-test, indicating no significant differences ($p = 0.617$). However, in Layer 2, PhD students appear more opposed than the other groups (Figure 56). The ANOVA analysis reveals a statistically significant difference ($F = 1.32, p < 0.05$). However, somewhat uncanny the Games Howell test did not find any statistically significant pairs on a 5% significance level.

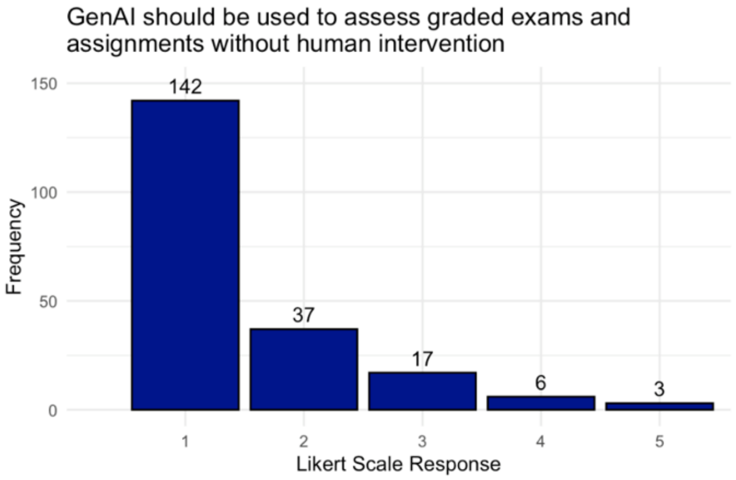


Figure 54 - Perceptions on using GenAI without human intervention for graded exams, among all participants

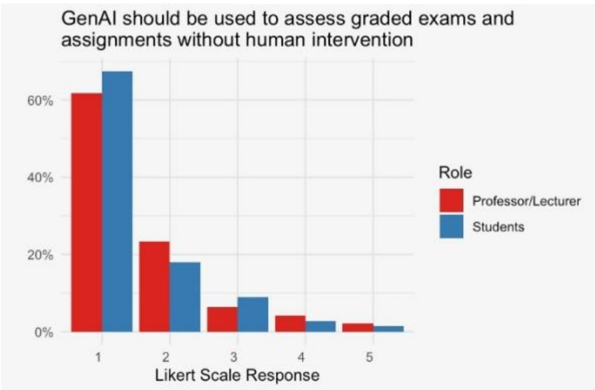


Figure 55 - Perceptions on using GenAI without human intervention for graded exams, by students (excl. PhD) and professors

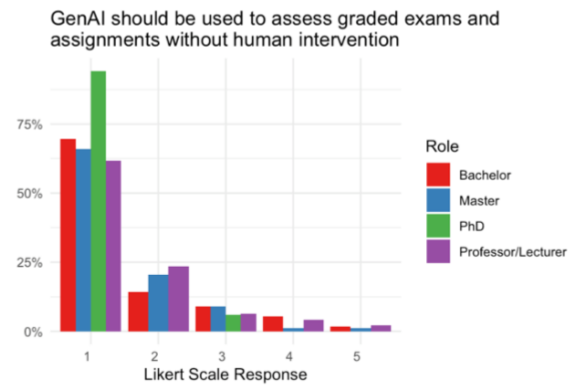


Figure 56 - Perceptions on using GenAI without human intervention for graded exams, by academic roles

Similarly, we looked at attitudes for non-graded exams and assignments. Here, while still predominantly negative, the disagreement was slightly lower, indicated by a higher mean of 1.98 ($SE = 1.16$; see Table 1) and a left-skewed distribution (Figure 57). This suggests a slightly more open attitude towards AI's role in non-graded assessments. Like graded assignments, the opinions in Layer 1 were shared, as shown by the distribution in Figure 58, and a high p-value in Welch's t-test ($p = 0.422$). For Layer 2, PhD students exhibited greater disagreement than other groups ($\bar{X}_{BA} = 1.95$, $\bar{X}_{MS} = 2.05$, $\bar{X}_{PhD} = 1.18$, $\bar{X}_{professor} = 2.17$; and Figure 59). Welch's ANOVA revealed significant differences in attitudes ($F = 6.69, p < 0.001$), with Games-Howell pairing the significant differences to PhD and (1) bachelor students (estimate$= -0.769, p = 0.01$), (2) master students (estimate$= -0.871, p < 0.01$), and (3) professors (estimate$= -0.997, p = 0.001$).

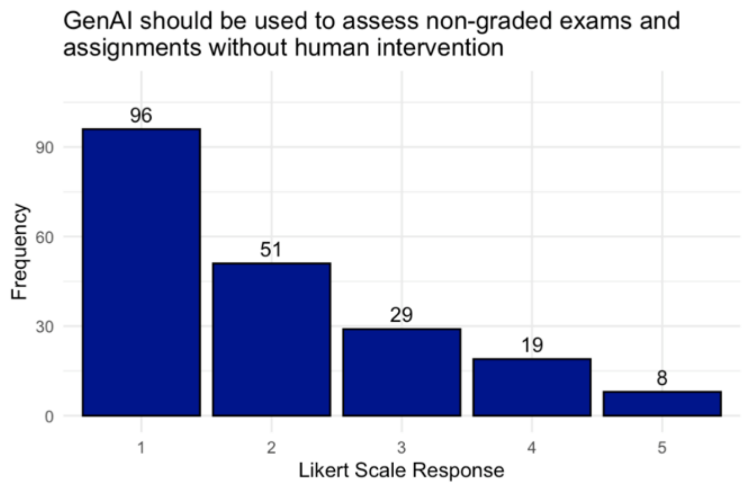


Figure 57 - Perceptions on using GenAI without human intervention for non-graded exams, among all participants

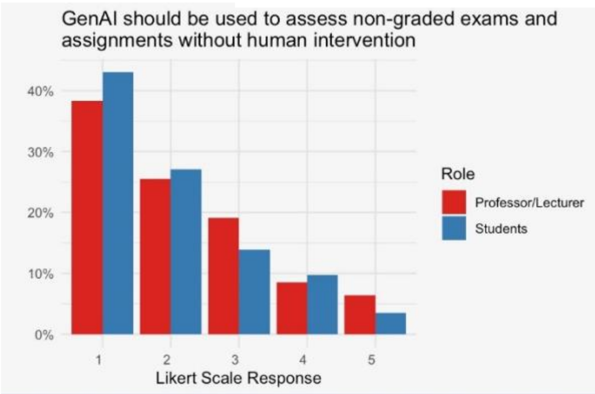


Figure 58 - Perceptions on using GenAI without human intervention for non-graded exams, by students (excl. PhD) and professors

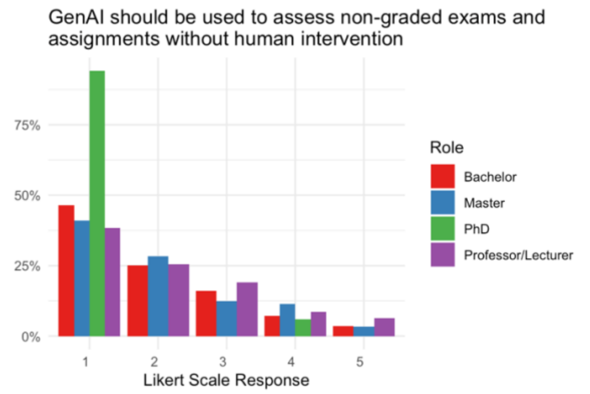


Figure 59 - Perceptions on using GenAI without human intervention for non-graded exams, by academic roles

In the final part of our analysis on assessment, we focused on attitudes towards the fairness of feedback provided by GenAI. As shown in Figure 60, a modest number of participants agreed that GenAI would offer fairer feedback (3.41%), with the majority taking a neutral stance (40%; $\bar{X} = 2.65$). This neutrality suggests uncertainty or skepticism within the academic community about AI's ability to deliver unbiased evaluation. Furthermore, we observe slight variances in opinions among students and professors (Figure 61), where the former seem more inclined to believe in GenAI's fairness compared to professors, a difference statistically supported by Welch's t-test ($t = -2.435,\ p < 0.05$). Moreover, a statistically significant variance in views was observed in the ANOVA test for groups in Layer 2 ($F = 3.28, p < 0.05$). The Games Howell post-hoc analysis identified a significant difference between master students and professors ($\text{difference} = 0.536, p < 0.05$), implying that master students are more optimistic than faculty in viewing GenAI as capable of providing more fair feedback.

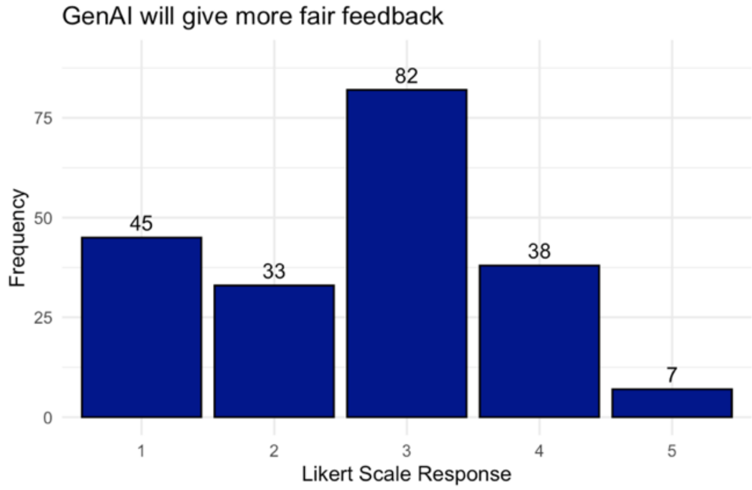


Figure 60 - Perceptions on fairness of GenAI vs. human feedback, among all participants

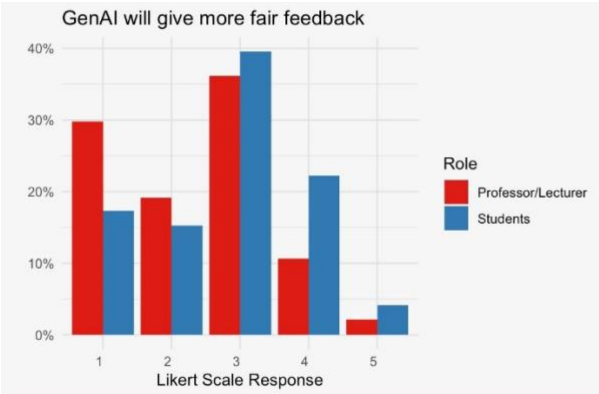


Figure 61 - Perceptions on fairness of GenAI vs. Human feedback, by students (excl. PhD) and professors

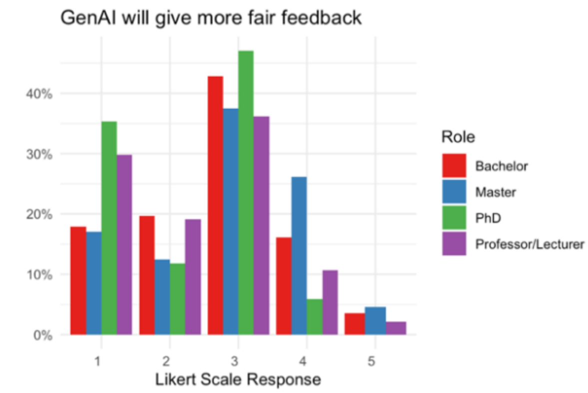


Figure 62 - Perceptions on fairness of GenAI vs. Human feedback, by academic roles

4.4.2 Discussion

Before we start the discussion, it is important to note that using GenAI for assessment is not something most students and professors have experience with, contrary to the more familiar use case of text generation. We must therefore be careful with how we interpret the responses.

In the analysis, we observe that an overwhelming majority of professors and PhD students have never used GenAI as a complementary tool for assessing student work. It is therefore evident that there is a potential for testing the viability of these tools in the context of assessment. As mentioned in the literature review, contemporary methods are time-consuming and often lack consistency (Ramesh & Sanampudi, 2022). However, despite recent advancements in GenAI, enabling more sophisticated assessment capabilities (Mizumoto & Eguchi, 2023), there is still a lack of prominent tools which can be used by professors (and PhD students) to assess student work. Followingly, it makes sense that usage of GenAI for this purpose is low, considering that there are no standardized ways of doing it. Those who have tried, have presumably had to make their own frameworks, or simply adopt a “try and fail”-approach. This consequently leads to the same inconsistency as in the traditional methods for assessment, defeating a major benefit with using GenAI for assessment.

Similarly, we observe that most professors and PhD students have not used GenAI as a complementary tool to make questions for assignments and exams. Interestingly, there were (1) nine fewer responses for this question compared to the previous, and (2) a higher proportion who had tried using GenAI for this purpose. (1) It is unclear as to why we got less responses for this question. Considering the fact that more people were comfortable with saying that they have employed the technology for this purpose, we do not believe that the lower response rate is caused by the question being more sensitive than the previous. (2) A possible explanation for why a higher proportion of participants had tried GenAI for making questions for assignments and exams, as opposed to using it as a complementary tool for assessing, might be related to the complexity of the respective task. One could for instance imagine it being a more straightforward process to brainstorm possible questions related to a topic using simple prompts, rather than inserting vast amounts of student work into the model and ask it to give a comprehensive evaluation based on certain criteria. First, it can be challenging to insert a lot of text into chatbots such as ChatGPT, because there is a certain word limit per prompt. However, even if one would have used an extension of ChatGPT which allows this, or another program all together, it is still not obvious how much time one would save by using it, compared to going over the analysis manually to quality check. Together, these hurdles, as well as others we might have missed, presumably make the assessment process more complex than the process of creating questions for exams and assignments.

Moving on to the Likert statements, the analysis generally finds that both students and professors are more open to using GenAI as a complementary tool for non-graded exams and assignments, compared to graded ones. A possible explanation for this is that the importance and risks related to graded work is significantly higher than for non-graded work. However, with means of 2.93 and 3.45 for graded and non-graded contexts respectively, we would not consider the overall attitude as particularly approving.

Concerning use of GenAI to assess student work *without human intervention*, the data reveals that both students and professors had noticeably negative attitudes towards this approach, with means of 1.49 and 1.98 for graded and non-graded contexts respectively. Despite being overwhelmingly negative, we interestingly observe a somewhat more positive attitude for the non-graded, underscoring the possibility of risk being an important driver for perceptions on adoption. Moreover, the results generally indicate that some level of human oversight is important for students and professors when it comes to assessing student work. An interesting point of discussion here is whether this is linked to the phenomena of *algorithm aversion*, which is the preference for human forecasting above statistical algorithms irrespective of the algorithm accuracy (Dietvorst et al., 2015; Jussupow et al., 2020). Curiously, in a follow-up paper to their seminal work from 2015, Dietvorst et al. (2018) found that people are more likely to use an algorithm if they are allowed to tweak its decisions slightly, suggesting that the aversion to algorithms can be mitigated by allowing some level of human intervention. However, Logg et al. (2019) found contrasting results from six experiments showing that laypeople adhere more to advice when they think it comes from an algorithm, as opposed to a person. Although, this effect (*algorithm appreciation*) diminished when they had to choose between an algorithm's estimate and their own, and when they possessed expertise in forecasting. Paradoxically, experienced professionals who regularly make forecasts relied less on algorithmic advice than laypeople, which adversely reduced their accuracy.

Finally, we will discuss the results on whether NHH students and professors think GenAI will provide more fair feedback on student work, compared to traditional assessment methods. We observe that most respondents either have a neutral stance or somewhat disagree/agree. In fact, this is the only question in our survey where "Neither agree nor disagree" was the most common response. Although we will not speculate extensively as to why this is, we would like to come full circle and highlight what we mentioned in the very beginning of the discussion on GenAI for assessment – Most students and professors do not have any prior experience on this topic, which was evident in the two opening questions mapping professors' and PhD students' usage of this application of GenAI. While bachelor and master students did not receive these questions, it is reasonable to assume that they are similarly inexperienced with this particular use case. Followingly, the lack of experience could explain why respondents do not feel strongly about GenAI's potential for objective and fair feedback. Nonetheless, the analysis interestingly finds a significant difference between the attitudes of master students and professors, with master students being more positive towards GenAI's ability to give fair feedback. Looking at Figure 62, we see that the master students' response rate for "Somewhat agree", is similar to the professors choosing "Disagree". Lastly, although one could argue that professors have a better understanding of what fair feedback is, we must keep in mind the findings of Logg et al. (2019) stating that seasoned professionals who frequently engage in forecasting relied less on algorithmic advice than laypeople, a tendency which negatively impacted their prediction precision – More on this in the next chapter on policy making.

In summary, there is a prevailing sense of caution regarding the adoption of GenAI for assessment purposes, particularly in scenarios lacking human supervision. We also observe a lack of experience with this particular use case of GenAI, and mixed attitudes towards the fairness of AI-generated assessment.

4.5 GenAI-powered chatbots as virtual assistants

4.5.1 Analysis

In this section, we explore participants' perspectives on employing GenAI as virtual assistants in higher education settings. Our analysis is structured as follows: (1) we begin by discussing the results from the survey's single-choice questions, which focus on identifying the primary benefits and challenges associated with this application of GenAI, (2) we then delve into examining students' preferences regarding problem-solving assistance, and (3) finally, we assess the overall attitudes towards this topic.

Perceived benefits and challenges of using GenAI-powered virtual assistants

Starting with the attitudes towards the most significant benefit of this application, Figure 63 shows that the participants are divided, where the majority highlight "Asking questions anytime", "Instant feedback", and "Reduced barrier to ask for help" as the most prominent benefits. In Layer 1, the students and professors seem to share opinions, except for "Asking questions anytime", which was more popular among students (Figure 64). This also seems to be the case in Layer 2 where we see a similar trend. Notably, PhD students overwhelmingly chose "Reduced barrier to ask for help" as the greatest benefit, hinting they perceive challenges in seeking assistance in their current academic environment.

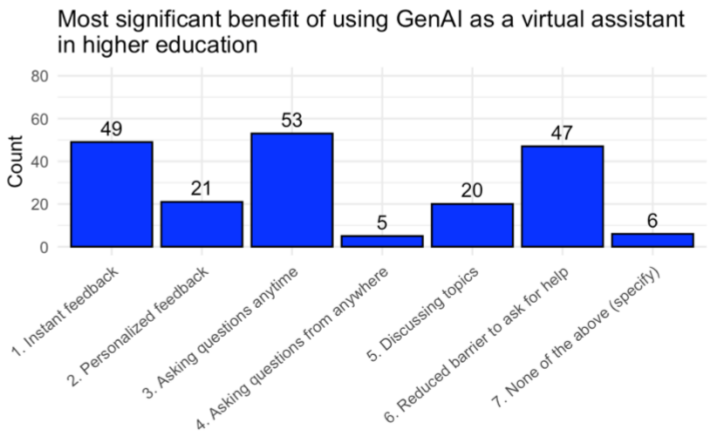


Figure 63 - GenAI virtual assistant benefit, among all participants

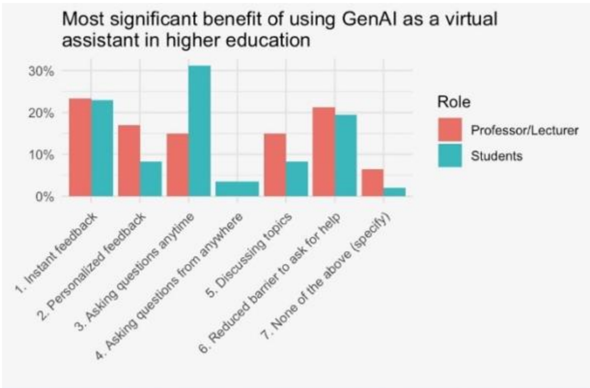


Figure 64 - GenAI virtual assistant benefit, by students (excl. PhD) and professors

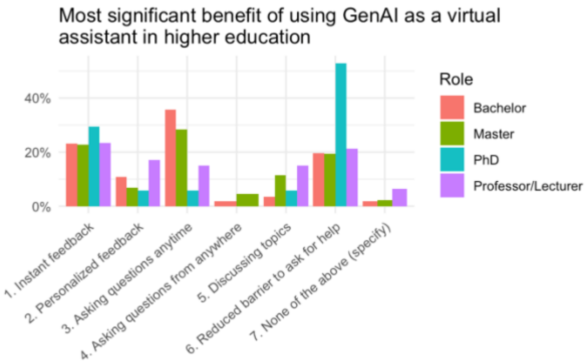


Figure 65 - GenAI virtual assistant benefit, by academic roles

Regarding challenges, most respondents identified the risk of misleading information as the most significant concern. Specifically, 62.3% attributed this to either (1) engine inaccuracy or (2) inherent biases in the GenAI's dataset, as depicted in Figure 66. Additionally, the data shows minimal variation in responses in Layer 1 and 2, as illustrated in Figure 67 and Figure 68. Notably, we find that a larger percentage of professors and PhD students, in comparison to bachelor and master students, expressed concern that the adoption of GenAI-powered virtual assistants could lead to decreased interaction between students and professors.

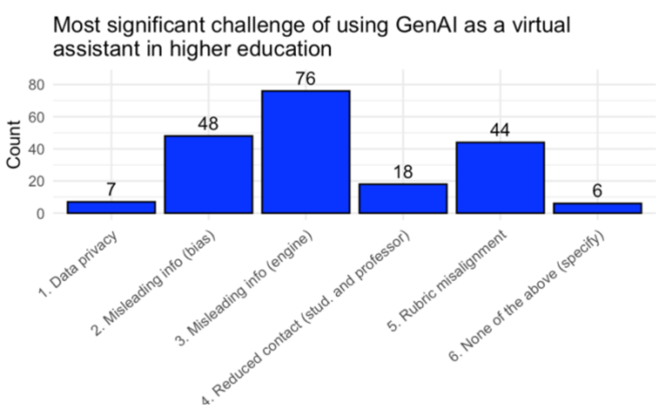


Figure 66 - GenAI virtual assistant challenge, among all participants

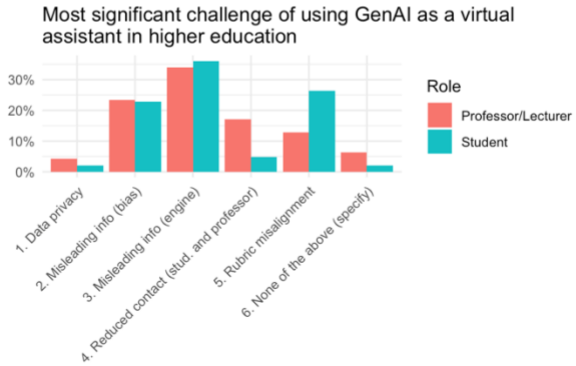


Figure 67 - GenAI virtual assistant challenge, by students (excl. PhD) and professors

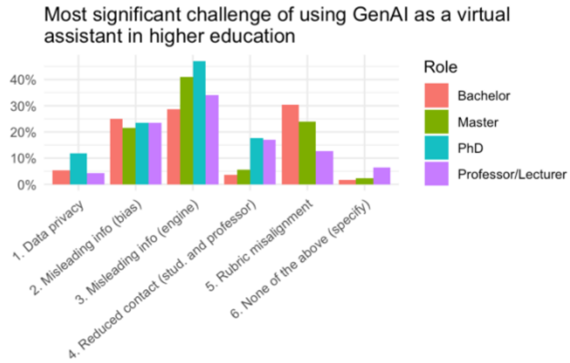


Figure 68 - GenAI virtual assistant challenge, by academic roles

Students' preferences for problem-solving assistance

Secondly, we investigated students' preferences for problem-solving assistance, specifically their inclination to interact with a chatbot versus (1) a professor, (2) a student assistant, and (3) a fellow student. This question was exclusive to students (including PhD), thus making Layer 1 irrelevant to analyze.

The findings indicate a clear preference for human interaction, particularly with professors, over chatbots ($\bar{X} = 2.37, SE = 1.30$; see Table 1). This preference is visually underscored by a left-skewed distribution (Figure 70), reflecting a general consensus across all student groups (including PhD), further supported by the ANOVA results ($p = 0.874$).

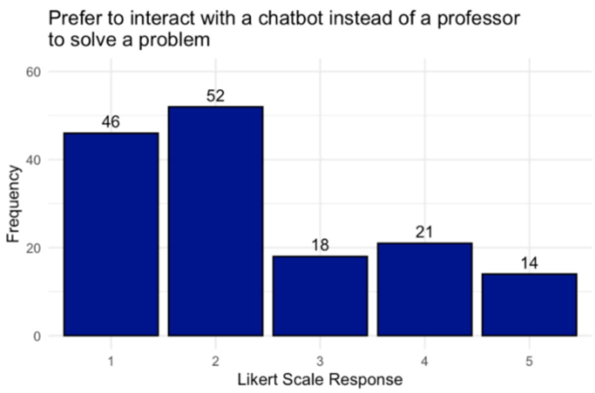


Figure 69 - Chatbot vs. professor support

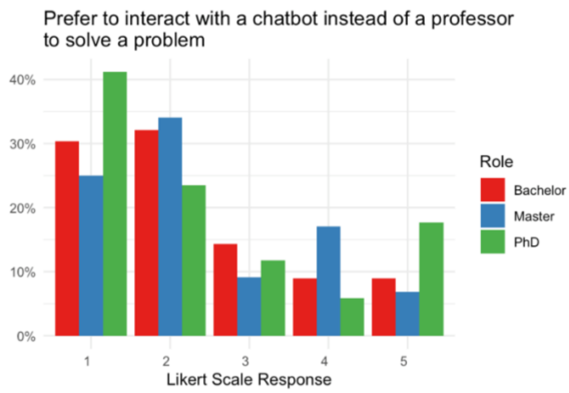


Figure 70 - Chatbot vs. professor support, by students

When considering interaction with a chatbot instead of a student assistant, the responses where slightly more neutral ($\bar{X} = 2.66, SE = 1.31$; see Table 1). While the histogram (see Figure 72) suggests minor group variations, with PhD students being less favorable ($\bar{X} = 2.18$) compared to bachelor ($\bar{X} = 2.62$) and master students ($\bar{X} = 2.79$), these differences were not statistically significant per Welch ANOVA ($p = 0.239$).

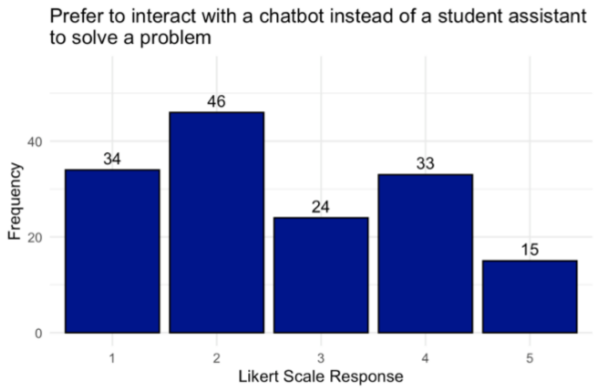


Figure 71 - Chatbot vs. student assistant support

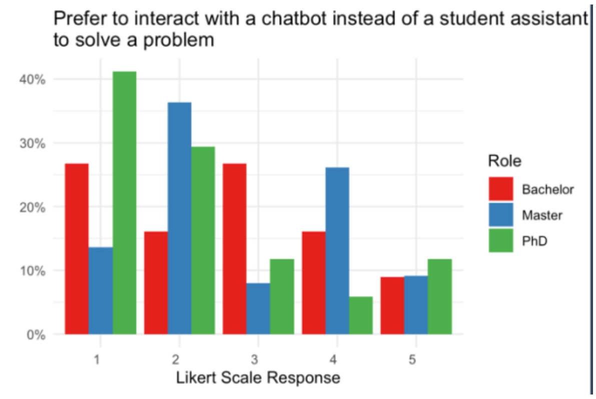


Figure 72 - Chatbot vs. student assistant support, by students

Regarding interacting with a chatbot versus fellow students, the average response indicated a mild disagreement ($\bar{X} = 2.32,\ SE = 1.23$; see Table 1), with the response distribution being relatively uniform across different student levels (Figure 74). The ANOVA analysis further supports this observation of equal means ($p = 0.867$).

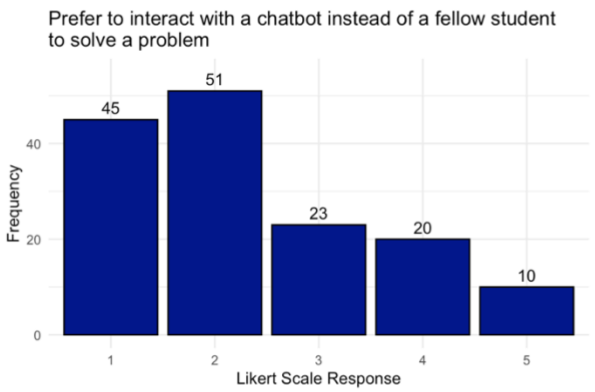


Figure 73 - Chatbot vs. fellow student support

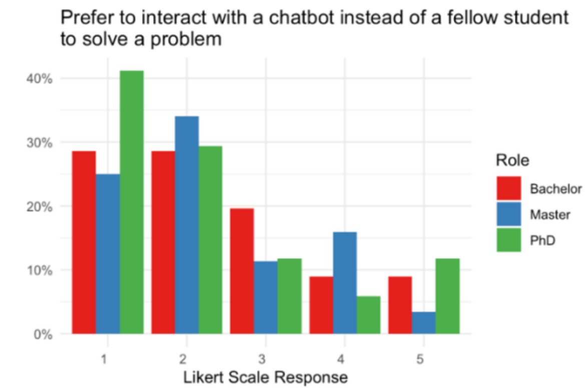


Figure 74 - Chatbot vs. fellow student support

General attitudes towards the use of GenAI-powered virtual assistants in higher education

Moving on to mapping the general attitudes for GenAI-powered virtual assistants in higher education, we presented participants with three statements to address using a Likert scale. The first statement assessed the extent to which participants value human advice over that provided by GenAI. A significant portion of our respondents expressed a preference for human advice, with 49.5% agreeing and 33.3% somewhat agreeing, resulting in a high average score of 4.27 ($SE = 0.88$; see Table 1). Further, when analyzing the responses across different academic roles, we observed a minor variance between students and professors (Figure 76), with a relatively low p-value (0.072). Here, professors tend to agree more than students. In Layer 2, PhD students mirrored the agreement level of professors (Figure 77). However, neither Layer 1 nor Layer 2 showed statistically significant differences in these attitudes, as indicated by Welch's t-test and Welch's ANOVA ($p = 0.343$).

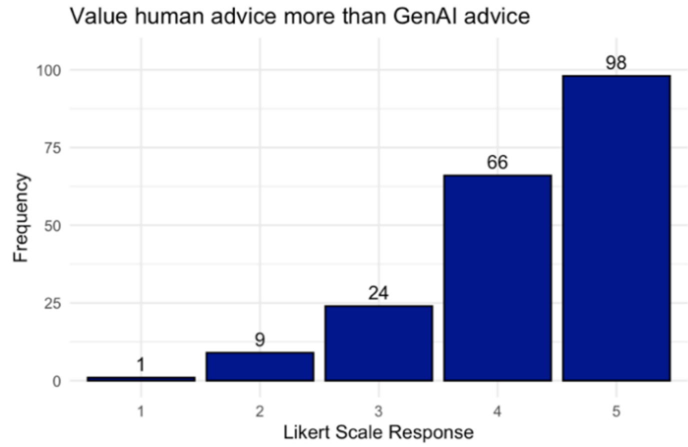


Figure 75 - Value of human vs. chatbot advice, among all participants

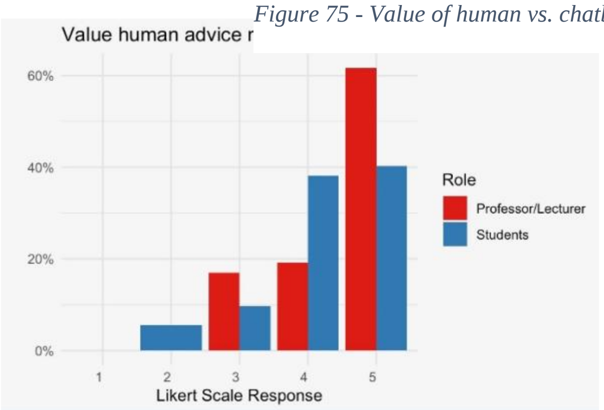


Figure 76 - Value of human vs. chatbot advice, by students (excl. PhD) and professors

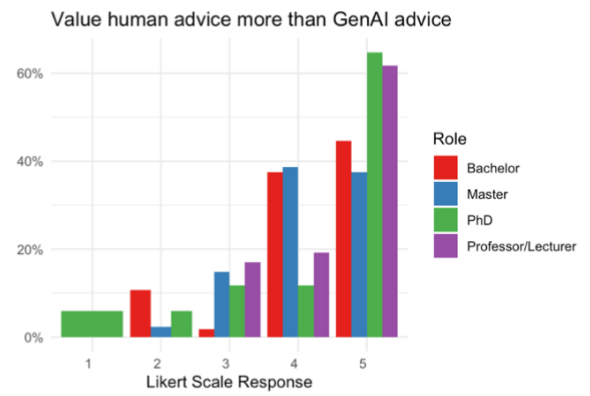


Figure 77 - Value of human vs. chatbot advice, by academic roles

The next statement queried respondents' attitudes towards whether implementing chatbots as virtual assistants would benefit student learning. Our data finds generally positive responses, with a mean of 3.90, where 52% of respondents somewhat agree and 22.7% fully agree with the statement. Moreover, when analyzing the distribution of these opinions across academic roles, we noticed a variance in attitudes (Figure 79 and 80). In Layer 1, students ($\bar{X} = 4.04$)

exhibit more positivity towards the learning outcomes of using chatbots compared to professors ($\bar{X} = 3.63$). This difference is statistically significant, as indicated by Welch's t-test ($t = -2.782,\ p < 0.01$). Further examination in Layer 2 showed that bachelor and master students shared similar views, while PhD students aligned more with professors' perspectives (Figure 80). The ANOVA test revealed a statistically significant variation among these groups ($F = 4.86, p < 0.01$), with post-hoc comparisons pinpointing significant differences between master students and professors (difference $= 0.455, p < 0.05$), as well as between master and PhD students (difference $= 0.615, p < 0.05$).

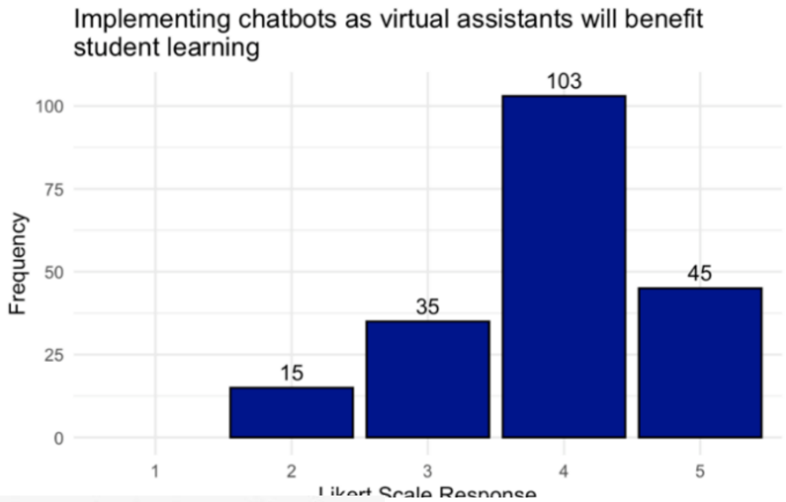


Figure 78 - Perceptions on chatbots' ability to benefit student learning, among all participants

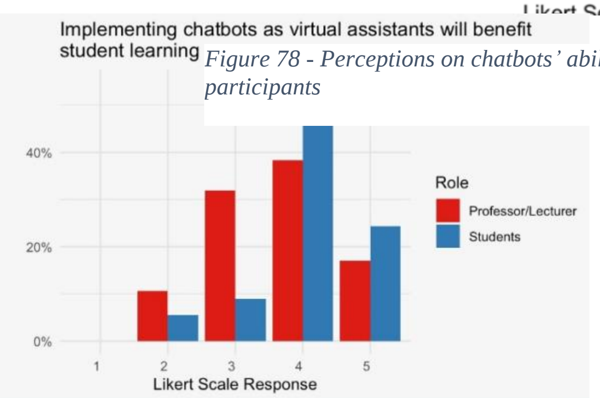


Figure 79 - Perceptions on chatbots' ability to benefit student learning, by students (excl. PhD) and professors

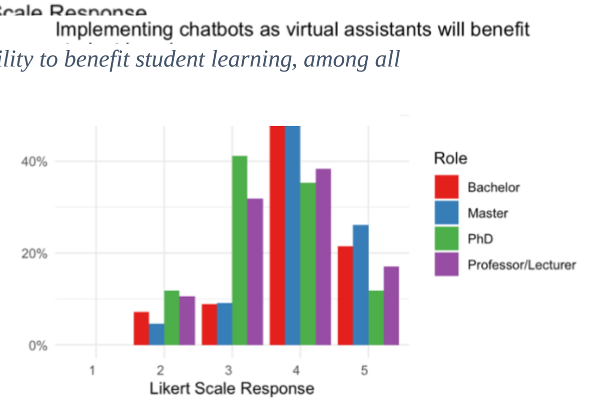


Figure 80 - Perceptions on chatbots' ability to benefit student learning, among all participants, by academic roles

Our last statement in the survey asked participants whether they believe virtual assistants can help students achieve a better grade by serving as a complementary teaching tool. As shown in Figure 81, the responses were predominantly positive, with 31.3% of respondents agreeing and 47.4% somewhat agreeing, leading to an overall mean of 3.99 ($SE = 0.96$; see Table 1). Notably, when examining the responses by academic role (Figure 82), it becomes evident that students have a higher belief in GenAI-powered chatbots' ability to improve their grades compared to professors. This difference is statistically significant, as concluded by Welch's t-test ($t = -2.892,\ p < 0.01$). Furthermore, we observe a similar pattern in Layer 2 as in the previous statement, where bachelor and master students share their views and are more inclined towards agreement than PhD students and professors. This difference in opinion is statistically significant, as indicated by the ANOVA results ($F = 3.71, p < 0.05$). The Games Howell test further pairs the significant difference in means to master students and professors (difference $= 0.506, p < 0.05$). Additionally, when excluding PhD students, a significant difference also emerged between bachelor students and professors (difference $= 0.461, p < 0.05$).

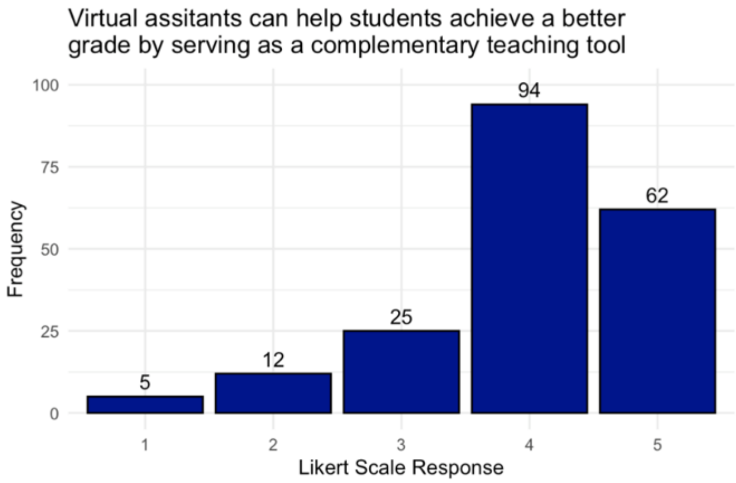


Figure 81 - Perceptions on chatbots' ability to help students achieve a better grade, among all participants

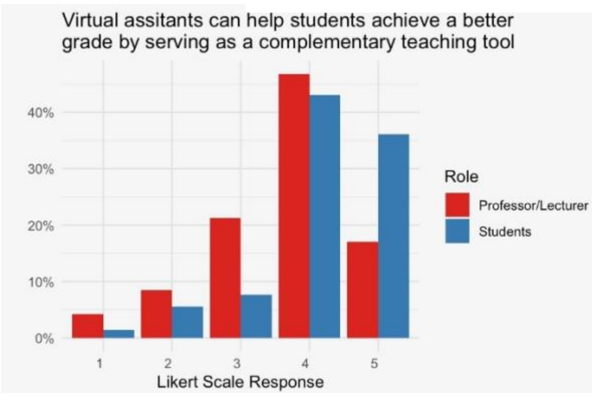


Figure 82 - Perceptions on chatbots' ability to help students achieve a better grade, by students (excl. PhD) and professors

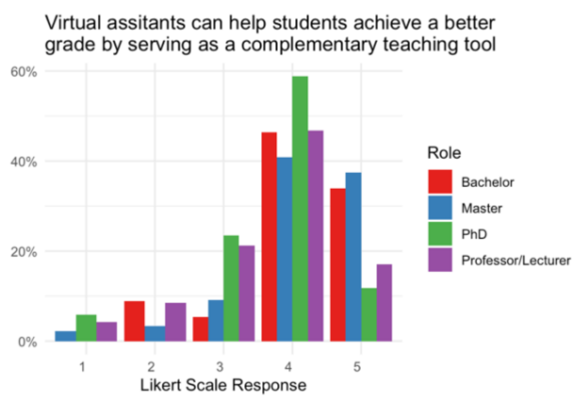


Figure 83 - Perceptions on chatbots' ability to help students achieve a better grade, by academic roles

4.5.1 Discussion

Looking at the advantages of implementing GenAI as virtual assistants in higher education, the analysis finds that time-related benefits seem to be highly valued by our participants. With “Asking questions anytime” and “Instant feedback” emerging as the most significant perks in our dataset, it appears that both students and professors agree that the convenience and efficiency offered by this technology is crucial. This aligns with findings from the literature review, suggesting that such availability has the potential to enhance the learning experience by providing immediate support, as well as creating a more flexible educational environment (Wollny et al., 2021).

Additionally, our data reveals that "Reduced barrier to ask for help" is another key benefit, receiving comparable support to the first two advantages. This suggests an existing gap in accessibility to learning support, which is an issue also pointed out in the literature review. The review noted that due to the high student-to-professor ratio, individual attention is often lacking in contemporary teaching methods (Wollny et al., 2021). However, we interestingly find that PhD students disproportionately perceive "Reduced barrier to ask for help" as the biggest benefit, despite the generally lower student-to-professor ratio at the PhD level. It would be intriguing to further investigate as to why this is in future research, but we will end the discussion on this point here.

Lastly, we observe that "Personalized feedback" and "Discussing topics" got a relatively low number of votes, despite much of the literature highlighting these aspects to be amongst the most prominent advantages with GenAI as virtual assistants (Andreassen, 2023; Crompton & Burke, 2023; Labaddze et al., 2023; Mollick & Mollick, 2023). Here, we see that students in particular seem less focused on these perks, in contrast to professors who place greater value on them (Figure 64). Considering that the three most highly regarded features all revolve around accessibility, it seems like student's regard the *quantity* of learning support as the primary benefit of GenAI tools.

As for the challenges of implementing GenAI as virtual assistants in higher education, the analysis shows that non-factual or misleading output appears to be the primary concern for both students and professors. More specifically, the apprehension mostly lies with engine inaccuracy, indicating a cautiousness towards the technology's precision. Although the emergence of this new generation of sophisticated chatbots have yielded impressive results, the presence of information inaccuracy is probably not going to disappear in the near future. This underscores the importance of understanding the underlying technology in terms of both its capabilities as well as its limitations. In doing so, the relevant higher education institution can better assess who would be an appropriate and satisfactory GenAI-provider for their institution.

Moving on to students' attitudes towards problem-solving assistance, we find relatively consistent views across all three Likert statements, implying a collective preference for interacting with humans over chatbots, irrespective of the three human alternatives we presented. This also aligns with the results from the Likert statement: "I value human advice more than generative AI advice", where an overwhelming majority agreed (82.8% somewhat agreed or fully agreed).

Despite this human preference, we also observe that participants mostly agree with GenAI benefiting student learning as a complementary tool to the current human assistance (Figure 78). Notably, bachelor and master students share this view, while PhD students and professors are somewhat more reserved. This could possibly reflect varying degrees of exposure and comfort with emerging technologies among these groups. Bachelor and master students, being more recent entrants into the academic world, might be more open and adaptable to new technological tools in education, while professors and PhD students, with their longer tenure in academia, might approach these technologies with more caution and skepticism. Interestingly, our section on general attitudes found that respondents considered "Reduced learning due to AI doing the work" to be the most significant challenge with using GenAI in higher education. It is therefore crucial that the complementary learning support provided by GenAI chatbots engage the students in critical thinking (Ayman et al., 2023; Crompton & Burke, 2023; Labaddze et al., 2023; Mollick & Mollick, 2023; Munthe et al., 2022), instead of taking over all the work. This ties back to our discussion on text generation, and the importance of collaboration rather than replacement.

Similarly, we see a generally positive attitude towards the statement: "Virtual assistants can help students achieve a better grade by serving as a complementary teaching tool". Again, bachelor and master students are somewhat more approving about this than PhD students and professors, one could therefore apply the same logic and argumentation as in the discussion above. Here, we would like to highlight a paper from our literature review by Bailifard et al. (2023), which found that students in a psychology class (n = 51) with an AI tutor achieved significantly higher grades (up to 15 percentage points) compared to peers taking a similar course without AI assistance. This finding provides quantitative support for the perceptions of students and professors in favor of leveraging chatbots as a complementary educational tool. However, we must also acknowledge the respondents holding the opposing view, including a noteworthy portion of professors in particular. As the implementation of GenAI is still in its early stages, it remains premature to draw any conclusions of the effectiveness of chatbots as virtual assistants, despite promising preliminary research. In fact, one could argue that the opinions of professors should carry more weight than those of students, considering their extensive experience in academic settings and deeper understanding of the nuances of educational guidance.

In summary, it appears that respondents in our survey value accessibility-related features for GenAI as virtual assistants, while non-factual or misleading information is perceived as the most significant challenge of this application. Furthermore, we observe that students consistently prefer human advice more than those given by GenAI. Nevertheless, this preference does not negate their recognition of GenAI's usefulness as a complementary tool to existing teaching methods. Curiously, we find statistical significance between students and professors, where the latter group have a more cautious approach towards use of GenAI-powered virtual assistants.

5 Informing future policies

In this chapter, we will provide some thoughts and recommendations for how NHH could shape future policies on GenAI and its use for educational purposes. Although we specifically look at NHH, we aspire for these guidelines to also serve as a valuable reference for other higher education institutions as well.

Structure wise, we start by presenting the current state of policies at NHH, followed by reflections and recommendations for each of our three focus areas – Consisting of (1) text generation, (2) assessment and (3) chatbots as virtual assistants. Since there currently is a limited body of literature on the topic, we do not have the necessary resources to provide extensively detailed policies. Therefore, we will instead propose overarching recommendations based on the findings in our research.

5.1 Current state at NHH

The current policies concerning use of GenAI at NHH are seemingly unclear. According to an article by Studvest, Stig Tenold, the Vice-Chancellor at NHH, stated that employing a robot in the creation of exam responses is considered cheating (Sauesund & Lia, 2023). Judging by this remark, it appears that all chatbot-features are prohibited in the context of text generation during exams. However, we could not find any further details on this matter, outside of the brief comment in the mentioned article.

In an effort to gain more clarity, we contacted Tenold directly through email, where we received some additional insights into GenAI policies at NHH. In his response, the Vice-Chancellor indicated that the guidelines are currently in progress. He also emphasized that due to the significant differences across various academic disciplines, the NHH board believes it should be left to each department to determine the appropriateness of GenAI usage. Simultaneously, it is important that instructors have a clear understanding of the limits to their latitude, and make students aware of what constitutes acceptable use of GenAI. This strategy aims to strike a balance between overarching institutional guidelines, and flexibility at the departmental level. The underlying argument for this policy is that different types of courses, such as business mathematics compared to strategy courses, demand distinct skill sets. Followingly, the learning support and examination tools need to be specifically adapted to suit these varied requirements.

Moreover, our research team was invited to discuss our preliminary findings at a workshop in Aalborg on GenAI in business education, where we got to share these views in front of professors and PhD students from NHH, Aalborg University, Gothenburg University, and Tampere University. Intriguingly, several participants expressed reservations to this approach, arguing that delegating decision-making to individual departments would lead to a lack of cohesion in the overarching institutional policy framework. While there are some general guidelines applicable across all departments, the local autonomy might amplify the current uncertainty and ambiguity in the policies concerning use of GenAI, as well as creating inconsistencies in student education and assessment. Consequently, one could imagine that this lack of clarity and dependability could undermine the overall quality and integrity of the educational programs offered at NHH.

In addition to the conflicting views on department autonomy related to policy-making, discussing NHH's student development goals can further complicate the picture. Although contemplating the ultimate educational purpose of the institution is beyond the scope of our thesis, we would like to highlight the school's development agreement for 2023-2026, outlined in the annual report of 2022 (Norwegian School of Economics, 2022). Here, we find four performance indicators related to student development:

1. Increased diversity and engagement among students
2. Academic development that contributes to sustainable value creation
3. Excellent learning environment and educational methods that emphasize student-active forms of teaching
4. High labor market relevance

In our context, we would argue that the third and fourth objectives are particularly relevant. For instance, one could employ chatbots as virtual assistants in a classroom setting as an innovative approach to student-centered teaching methods (third objective). This could create an excellent learning environment where the benefits of human advice, which were highly valued in our analysis, are complemented by the scalable accessibility of GenAI (Wollny et al., 2022). Further, the fourth objective of high labor market relevance ties back to the discussion on which student skills to emphasize developing. Here, we could draw a parallel to the reflections made on where one should draw the line on text generating features. Although it appears that none are currently permitted, according to Tenold's statement in Studvest

(Sauesund & Lia, 2023), it is worth considering whether certain features actually should be learned to leverage in response to evolving labor market demands. As an example, the ability to brainstorm in tandem with a virtual assistant might emerge as a valuable skill in the future workforce, particularly in industries that increasingly rely on collaboration between humans and AI, such as in tech and creative sectors.

On the other hand, there are reasons to be careful with the widespread use of GenAI tools like chatbots in education. A key concern is that relying too much on AI might weaken essential skills like critical thinking and problem-solving (Bogdanović-Dinić, 2023), important in various careers beyond tech and creative fields. Also, overusing GenAI could lessen the important benefits of human interaction in the educational process. This overuse might also diminish the role of teachers (Ghamrawi et al., 2023), who play a crucial part in guiding and inspiring students, affecting their overall academic growth. Moreover, despite an increasing focus on AI-competencies in the labor market, interpersonal skills and human judgement are still highly valued, yet they may not be adequately developed in an AI-focused learning environment.

In summary, there is a noticeable ambiguity in the existing NHH policies regarding the use of GenAI for educational purposes. While the Vice-Chancellor has stated in a Studvest article that GenAI use during exams is forbidden (Sauesund & Lia, 2023), detailed guidelines for its use in other contexts are not readily available. Additionally, we observe conflicting views among professors about whether future policies should permit individual departments to customize their own GenAI usage guidelines. Finally, it is important to recognize that the adoption of GenAI at NHH could have both positive and negative impacts on the institution's student development goals.

5.2 Text generation

Despite having identified and detailed multiple risks related to allowing text-generating features for student work, we must acknowledge the practical challenge of prohibiting its use. This is especially true for term papers, assignments and home exams. From a pragmatic standpoint, one could therefore argue that it would make more sense to find ways of adopting the technology, rather than significantly increasing the efforts for detecting its use.

A possible approach would be to follow the steps of UiO, by tailoring an existing chatbot software service to the specific academic needs of NHH students and professors (UiO, 2023).

This way, the institution would have more control of the features being used in student work, a decision which could be informed by the data in our research and NHH's student development goals. Although such a strategy would require resources for programming, integration, and training, one could argue that the widespread use of chatbots in academia and the labor market is inevitable, making it better to have a proactive relationship to this emerging technology, instead of being reactive.

Concerning training, our research indicates that professors, like students, are interested in GenAI tools; however, they exhibit relatively less confidence in using these tools compared to their students. Either way, general training needs to be offered to everyone, ensuring that students and professors who are not confident and proficient with these tools receive adequate development. Further, a paper by Okonkwo and Ade-Ibijola (2021) states that: "In a study on the adoption of AI in higher education, Chatterjee and Bhattacharjee (2020) finds that individual's behavioral intentions to use AI in higher education are influenced by their attitudes. Likewise, another research on the adoption of software engineering product proved that user attitude influences the adoption of software tools (Okonkwo et al., 2019)." (Okonkwo & Ade-Ibijola, 2021, p. 7). In light of this, one could argue that the adoption and training of GenAI at NHH will likely be welcomed, because our data showed an overwhelming majority of participants being interested in the field of GenAI, as well as perceiving GenAI as a progressive step towards the future of higher education. However, we should mention that the attitudes of professors were somewhat more reserved, indicating that this segment might be more challenging to get onboard. This resistance should not be underestimated, and an appropriate dialog with professors where they can express potential concerns is necessary.

As found in the analysis on GenAI as an assessment tool, it appears that the implementation should first be tested in non-graded work, particularly for non-graded assignments. From a risk perspective, this would also make the most sense in the case of text generation for student work. Moreover, to mitigate some of the downsides of employing chatbots for this application, we propose revising contemporary approaches to (1) question-making and (2) assessing assignments. In example, the questions themselves could encourage students to use specific prompts, and then critically evaluate the results. This way, instructors could take greater control of how students use the chatbots, as well as assessing the work based on students' ability to critically evaluate the output quality of GenAI in relation to the course curriculum. Moreover, by incorporating the process of detailing methods and use of GenAI into the evaluation rubric, one could incentivize transparency and thereby reduce potential misuse. Such an approach is just one of many examples of how one could proactively adopt GenAI at NHH.

In summary, we propose to experiment with implementation of chatbots for text generation in non-graded assignments. By doing so, NHH can gather more data on what works and what does not, all while minimizing the negative consequences. This proactive approach is grounded in the inevitable widespread use of GenAI in academia and the labor market, as well as seeing opportunities to leverage the benefits of this emerging technology. In fact, our research notably revealed that students and professors highlighted text generation as the most significant benefit of using GenAI in higher education, which according to insights from Okonkwo et al. (2019), likely would make the implementation well-received. However, we would highly encourage to keep an ongoing dialog with critics, particularly professors, as considering all viewpoints is salient. Lastly, to balance out the increased use of GenAI in student work, we recommend keeping school-exams for courses where skills related to GenAI use are less relevant, like introductory bachelor courses which aim to assess foundational understanding of core economic concepts.

5.3 Assessment

In contrast to our position on using GenAI for text generation in student work, we advise against employing GenAI for assessment purposes. The rational for this recommendation is a lack of approval for all four statements we posed in our survey:

1. GenAI as a complementary tool for non-graded assessments ($\bar{X} = 3.45$)
2. GenAI as a complementary tool for graded assessments ($\bar{X} = 2.93$)
3. GenAI without human intervention for non-graded assessments ($\bar{X} = 1.98$)
4. GenAI without human intervention for graded assessments ($\bar{X} = 1.49$)

Although the first option has a mean leaning towards favor, we do not consider this sufficient enough to justify developing a program for GenAI assessment at NHH. Contrary to the use case of text generation, we do not see any prominent tools that can meet the standards necessary for assessment at NHH. Given its nascent stage, we would instead encourage monitoring the advancements of this technology.

In addition to the technical software challenges, staff training emerges as another significant hurdle. Our data reveals very limited use of GenAI for assessment purposes, which might lead back to the scarcity of specialized tools. As explored in previous chapters, most attempts to integrate GenAI in assessments likely involved conventional chatbots, such as ChatGPT. However, the lack of established frameworks for utilizing these chatbots in assessments presumably led professors to devise their own methods, highlighting a gap in structured guidance for effectively leveraging GenAI as a useful evaluative tool. Therefore, it would be necessary to extensively train professors at NHH on how to use GenAI for assessment in a consistent manner across various academic disciplines. Considering the lack of necessary GenAI tools, and a limited body of literature on this topic, we consequently advise against initiating such an assessment program at NHH.

Finally, we would like to highlight that our research revealed professors being more opposed to perceiving GenAI as capable of providing fair feedback, compared to students. Bearing in mind that professors have significantly more experience with assessing student work, as opposed to students themselves, we would argue that it is reasonable to place greater emphasis on their opinions for this particular matter. Although, we must remember the findings of Logg et al. (2019), which stated that seasoned professionals, who frequently engage in forecasting, relied less on algorithmic advice than laypeople, a tendency to which negatively impacted their prediction precision.

Summing up, we find several reasons to caution against a potential implementation of GenAI for assessments at NHH. These include lacking (1) approval from students and professors, (2) specialized software, and (3) necessary staff competencies.

5.4 GenAI-powered virtual assistance

We generally recommend an integration of GenAI-powered virtual assistance for students at NHH. Similar to the approach for text generation, we propose tailoring an existing chatbot software to the specific needs of students and professors at this institution (UiO, 2023). Since platforms like ChatGPT offer capabilities which could be used for both text generation as well as for virtual assistance, it would be possible to make a unified model for both purposes. This model could be adjustable, with features that can be enabled or disabled to address different academic needs.

The reason we advocate for integrating GenAI-powered virtual assistants stems from our research indicating a generally favorable opinion towards this technology among our study participants. While there was a preference for human advice, respondents expressed interest in virtual assistants as a supplementary resource. Notably, students particularly valued the availability of virtual assistants, a benefit also highlighted by Wollny and colleagues in 2021 (Wollny et al., 2021). Additionally, this technology is highly scalable, and could contribute to solving the issue of a high student-to-professor ratio. Lastly, we must also highlight the previously touched upon fact that implementing chatbots as virtual assistants could support the third student development objective of NHH, by serving as an education tool which emphasizes student-active forms of teaching (Norwegian School of Economics, 2022). Specifically, we could tie this to GenAI's notable benefit of personalized feedback, where students would get tailored feedback on their academic work and various questions (Andreassen, 2023; Crompton & Burke, 2023; Labaddze et al., 2023; Mollick & Mollick, 2023).

Nevertheless, we must also be aware and cautious of the potential drawbacks accompanied by implementing GenAI as virtual assistants at NHH. Although there are several concerns, we would like to shed light on the importance of critical thinking. An overreliance on GenAI for study or project work could potentially hinder the development of students' critical thinking (Bailey, 2023). The virtual assistants must therefore be programmed in such a way that they encourage and facilitate independent thinking rather than providing direct answers. This could be achieved by designing the system to pose guiding questions, offer hints, or suggest resources instead of giving outright solutions. Such an approach would help maintain a balance between leveraging the efficiency of GenAI and nurturing the essential skill of critical analysis in students. By doing so, we can harness the benefits of GenAI while minimizing the risks associated with dependency.

Additionally, it is also worth considering the risks related to misinformation. In our survey, this emerged as the biggest issue with GenAI as a virtual assistant, particularly in regard to engine inaccuracy. Walter Pohl, a NHH professor at the department of finance, has expressed concerns about the use of ChatGPT in an article published at NHH Bulletin (Pohl, 2023). Here, he highlights its limitations when it comes to understanding and solving complex problems accurately. While the tool can provide plausible answers, it may not always offer the depth and precision required for academic work. This underscores the need for students to critically evaluate the information provided by GenAI assistants and avoid overreliance. It is therefore salient to address these issues and ensure responsible use, to minimize the mentioned risks.

In summary, we propose implementing virtual assistants tailored to the academic needs of students at NHH. Although there are multiple risks to this, we believe GenAI can be a valuable educational tool, as evident by the perceptions of participants in our study. More specifically, it could help solve the issue of a high student-to-professor ratio, by increasing the availability of personalized educational support (Andreassen, 2023; Crompton & Burke, 2023; Labaddze et al., 2023; Mollick & Mollick, 2023; Wollny et al., 2021). This implementation would also align with NHH's overarching student development goals for 2023-2026 (Norwegian School of Economics). However, careful consideration is warranted to address the associated risks, while ensuring responsible use.

6 Conclusion

In this master's thesis, we sought to map the attitudes of NHH students and professors on GenAI in higher education, as well as investigating their differences. We began by conducting a literature review, where we looked at (1) the current state of higher education in Norway, (2) the technology of GenAI, and (3) the intersection between these domains. The literature revealed that use of GenAI in higher education is still in nascent stages, with four applications emerging as notably promising: (i) Text generation for student work, (ii) assessment, (iii) virtual assistance, and (iiii) administrative tasks. Prioritizing student learning, we focused on the first three applications in our study. Further, we employed an exploratory and descriptive research method, using online surveys to gather quantitative data from a non-probabilistic sample of NHH students and professors. This data was later analyzed through various statistical tests, followed by an extensive discussion of the findings. Finally, grounded in these insights, we shared some reflections and recommendations for how one could go about shaping future GenAI policies at NHH, with the aspiration to also work as a point of inspiration for other institutions as well.

We generally find that the adoption of GenAI in higher education is characterized by an approval from students, while professors have a somewhat more cautious stance, suggesting a possible generational gap in embracing this technology. For text generation, our data reveals a decline in approval as its use expands, indicating increased resistance to more extensive GenAI integration. The decision of which features to allow is complex, due to varied opinions among students and professors. Additionally, one must also consider this in the broader context of institutional student development goals, as well as questioning whether the overarching goals themselves should change to match the pace of the technological innovations. Concerning assessment, we observe less positive attitudes towards this application. Although participants slightly favor the idea of GenAI as a complementary tool for assessing non-graded student work, the overall sentiment is critical. Moreover, we also find a lack of experience with such a use case, which may allude to the scarcity of specialized GenAI tools for this purpose. Lastly, while it appears that students consistently prefer human advice over its machine counterparts, the preference does not negate their recognition of GenAI's usefulness as a complementary tool to existing teaching methods. Here, participants particularly value accessibility-related features for GenAI as virtual assistants, while non-factual or misleading information is perceived as the most significant challenge.

Finally, we find that NHH’s current policies on GenAI are unclear, and that professors seem to have conflicting views on how to move forward. Our general recommendation is to mirror the approach of UiO by developing education-specific versions of existing GenAI tools, in order to explore applications for text generation and virtual assistance (UiO, 2023). Given the inevitable widespread use of tools like ChatGPT, we would argue that it is better to proactively experiment with such integrations in low-risk environments, rather than expending significant resources on prohibition. Simultaneously, we would encourage creating a committee which actively monitors what works and what does not, by maintaining an ongoing dialogue with both proponents and critics. As for GenAI in assessment, we find several reasons to caution against this implementation. These include lacking: (1) Approval from students and professors, (2) specialized software, and (3) necessary staff competencies.

7 Bibliography

Al Darayseh, A. (2023). Acceptance of artificial intelligence in teaching science: Science teachers' perspective. *Computers and Education: Artificial Intelligence*, *4*, 100132. https://doi.org/10.1016/j.caeai.2023.100132

Almaraz-López, C., Almaraz-Menéndez, F., & López-Esteban, C. (2023). Comparative Study of the Attitudes and Perceptions of University Students in Business Administration and Management and in Education toward Artificial Intelligence. *Education Sciences*, *13*(6), Article 6. https://doi.org/10.3390/educsci13060609

Andreassen, T. W. (2023, May 30). *HVORFOR JEG ØNSKER KI I HØYERE UTDANNING VELKOMMEN*. NHH Bulletin: https://www.nhh.no/nhh-bulletin/artikkelarkiv/2023/mai/hvorfor-jeg-onsker-ki-i-hoyere-utdanning-velkommen/

Arredondo, P. (2023, April 19). *GPT-4 Passes the Bar Exam: What That Means for Artificial Intelligence Tools in the Legal Profession*. Stanford Law: https://law.stanford.edu/2023/04/19/gpt-4-passes-the-bar-exam-what-that-means-for-artificial-intelligence-tools-in-the-legal-industry/

Astrup, N. (2020, January 14). *The National Strategy for Artificial Intelligence* [Plan]. Government.No; regjeringen.no. https://www.regjeringen.no/en/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/

Ayman, S. E., El-Seoud, S. A., Nagaty, K. A., & Karam, O. (2023). The Impact of ChatGPT on Student Learning/performing. *ResearchGate*, 9. https://doi.org/DOI: 10.13140/RG.2.2.28890.11205

Bailey, J. (2023). AI in Education. *Education Next*. https://www.educationnext.org/a-i-in-education-leap-into-new-era-machine-intelligence-carries-risks-challenges-promises/

Baillifard, A., Gabella, M., Lavenex, P. B., & Martarelli, C. S. (2023). *Implementing Learning Principles with a Personal AI Tutor: A Case Study* (arXiv:2309.13060). arXiv. http://arxiv.org/abs/2309.13060

Bishop, C. M., & Nasrabadi, N. M. (2006). Pattern recognition and machine learning (Vol. 4, No. 4, p. 738). New York: springer.

Bleiklie, I. (2023, September 25). *Norwegian higher education futures*. Springer Link: https://link.springer.com/article/10.1007/s10734-023-01107-8#Sec1

Bogdanović-Dinić, S. (2023, April 13). *Critical Time for Critical Thinking: Is AI Making Us Dull?* - HTEC. https://htecgroup.com/is-ai-making-us-dumb/

Brennan, J. (2004). The social role of the contemporary university: Contradictions, boundaries and change. Ten years on: Changing education in a changing world, 22-26.

Caldwell, A. R., Lakens, D., Parlett-Pelleriti, C. M., Prochilo, G., & Aust, F. (n.d.). *Chapter 12 Violations of Assumptions | Power Analysis with Superpower*. https://aaroncaldwell.us/SuperpowerBook/violations-of-assumptions.html

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, *20*(1), 43. https://doi.org/10.1186/s41239-023-00411-8

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: a narrative overview. Procedia Computer Science, 136, 16-24.

Chatterjee, S., & Bhattacharjee, K. K. (2020). Adoption of artificial intelligence in higher education: A quantitative analysis using structural equation modelling. Education and Information Technologies, 25, 3443-3463.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20, 273-297.

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, *20*(1), 22. https://doi.org/10.1186/s41239-023-00392-8

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Davis, F, D., (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly, Vol. 13, No. 3*. pp. 319-340. https://doi.org/10.2307/249008

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of student's t-test—Eindhoven University of Technology research portal.

International Review of Social Psychology, *30*(1), 92–101. https://research.tue.nl/en/publications/why-psychologists-should-by-default-use-welchs-t-test-instead-of-

Delacre, M., Leys, C., Mora, Y. L., & Lakens, D. (2019). *Taking Parametric Assumptions Seriously: Arguments for the Use of Welch's F-test instead of the Classical F-test in One-Way ANOVA* (1). *32*(1), Article 1. https://doi.org/10.5334/irsp.198

DeVellis, R. F. (2012). Scale development: Theory and practice (4th ed.). *Sage Publications*.

Dietvorst, B., J. P. Simmons & C. Massey. (2015). "Algorithm aversion: People erroneously avoid algorithms after seeing them err." *Journal of Experimental Psychology: General, 144*(1), 114–126

Dietvorst, B., J. P. Simmons & C. Massey. (2018). "Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them." Management Science, 64(3), 1155–1170.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78-87.

Domingos, P. (2015). The master algorithm: How the quest for the ultimate learning machine will remake our world. *Basic Books*.

Ernst, E., & Schaefer, K. (n.d.). *The rise of AI in China – Digital technologies and their regulation | The Future of Work Podcast*. https://voices.ilo.org/podcast/the-rise-of-ai-in-china--digital-technologies-and-their-regulation

Fink, A. (2013). How to ask survey questions (5th ed.). *SAGE Publications*.

Finans Norge. (2023, October 17). *Ny digitaliseringsminister: – Etterlengtet*. https://www.finansnorge.no/artikler/2023/10/ny-digitaliseringsminister--etterlengtet/

Fitzsimmons, C. J., Thompson, C. A., & Sidney, P. G. (2020). Confident or familiar? The role of familiarity ratings in adults' confidence judgments when estimating fraction magnitudes. *Metacognition and Learning*, 15(2), 215–231. https://doi.org/10.1007/s11409-020-09225-9

Frąckiewicz, M. (2023, September 12). How AI Intelligent Tutoring Systems are Empowering Educators and Enhancing Learning Outcomes. *TS2 SPACE*. https://ts2.space/en/how-ai-intelligent-tutoring-systems-are-empowering-educators-and-enhancing-learning-outcomes/

Frost, J. (n.d.). *Using Post Hoc Tests with ANOVA - Statistics By Jim*. https://statisticsbyjim.com/anova/post-hoc-tests-anova/

Gastwirth, J. L., Gel, Y. R., & Miao, W. (2009). The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice. *Statistical Science, 24*(3), 343–360. https://doi.org/10.1214/09-STS301

Ghamrawi, N., Shal, T., & Ghamrawi, N. A. R. (2023). Exploring the impact of AI on teacher leadership: Regressing or expanding? *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-12174-w

Gnambs, T., & Kaspar, K. (2015). Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. Behavior research methods, 47, 1237-1259.

Global Admissions. (2023). *Exploring the Use of AI in Universities*. Global Admissions: https://www.globaladmissions.com/blog/exploring-the-use-of-ai-in-universities/

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). Survey methodology (Vol. 561). John Wiley & Sons.

Gudivada, V. N. (2017). 2.2.1.1 Descriptive Statistics. In *Data Analytics for Intelligent Transportation Systems*.

Han, P. (2023, June 15). *Tutor, TA, Talent Scout: The Coming GenAI Revolution in Higher Education.* Medium: https://medium.com/aimonks/tutor-ta-talent-scout-the-coming-genai-revolution-in-higher-education-400282d68793

Hu, K. (2023). ChatGPT sets record for fastest-growing user base - analyst note. Reuters. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing- user-base-analyst-note-2023-02-01/

Imran, R., Fatima, A., Salem, I., & Allil, K. (2023, April 17). *Teaching and learning delivery modes in higher education: Looking back to move forward post-COVID-19 era.* ScienceDirect: https://www.sciencedirect.com/science/article/pii/S1472811723000435

Jhangiani, D. R., & Tarry, D. H. (2022). *4.1 Exploring Attitudes*. https://opentextbc.ca/socialpsychology/chapter/exploring-attitudes/

JMP. (n.d.). *The t-Test*. Statistics Knowledge Portal. https://www.jmp.com/en_ch/statistics-knowledge-portal/t-test.html

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion.

Kantar media. (2023). Medietrender Ung 2023. https://kantar.no/rapporter/medietrender-ung-2023/

Kim, T. K. (2017). Understanding one-way ANOVA using conceptual figures. *Korean Journal of Anesthesiology*, *70*(1), 22–26. https://doi.org/10.4097/kjae.2017.70.1.22

Kjørstad, E. (2023, August 24). *Bør ChatGPT og kunstig intelligens brukes i utdanning? Ja, mener forsker*. Forskning.no: https://forskning.no/kunstig-intelligens-utdanning/bor-chatgpt-og-kunstig-intelligens-brukes-i-utdanning-ja-mener-forsker/2239431

Kothe, D. N. (n.d.). *Chapter 14 Comparing several means (one-way ANOVA) | Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.6.1)*. https://learningstatisticswithr.com/book/anova.html

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. Quality & quantity, 47(4), 2025-2047.

Kunnskapsdepartementet. (2021). *Strategi for digital omstilling i universitets- og høyskolesektoren.* Oslo: Kunnskapsdepartementet.

Kunnskapsdepartementet. (2023, September 7). *Regjeringen med milliardsatsing på kunstig intelligens* [Pressemelding]. Regjeringen.no; regjeringen.no. https://www.regjeringen.no/no/aktuelt/regjeringen-med-milliardsatsing-pa-kunstig-intelligens/id2993214/

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. International Journal of Educational Technology in Higher Education, 20(1), 56.

Liu, D., & Kong, S. (2023). Graduate employability and the labour-market relevance of Norwegian higher education: Perspectives from students. In The SAGE Handbook of Graduate Employability (p. 433). SAGE Publications.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90-103.

McCarthy, J. (2007). What is artificial intelligence?. *http://jmc.stanford.edu/articles/whatisai/whatisai.pdf*

McCormack, M. (2023). EDUCAUSE QuickPoll Results: Adopting and Adapting to Generative AI in Higher Ed Tech. EDUCAUSE Review. https://er.educause.edu/articles/2023/4/educause-quickpoll-results-adopting-and-adapting-to-generative-ai-in-higher-ed-tech

McDonald, J. H. (2017, June 27). *4.5: Homoscedasticity and Heteroscedasticity*. Statistics LibreTexts. https://stats.libretexts.org/Bookshelves/Applied_Statistics/Biological_Statistics_(McDonald)/04%3A_Tests_for_One_Measurement_Variable/4.05%3A_Homoscedasticity_and_Heteroscedasticity

McKinsey & Company. (2023, August 25). *Generative AI Business Functions*. http://ceros.mckinsey.com/generative-ai-business-functions

Ministry of Education and Research. (2022). *Meld. St. 5 (2022–2023) Long-term plan for research and higher education 2023–2032*. The Government. https://www.regjeringen.no/en/dokumenter/meld.-st.-5-20222023/id2931400/?ch=1

Mitchell, T. M. (1997). Machine learning.

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive Statistics and Normality Tests for Statistical Data. *Annals of Cardiac Anaesthesia*, *22*(1), 67–72. https://doi.org/10.4103/aca.ACA_157_18

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, *2*(2), 100050. https://doi.org/10.1016/j.rmal.2023.100050

Mokyr, J. (2011). The gifts of Athena: Historical origins of the knowledge economy. Princeton University Press.

Mollick, E., & Mollick, L. (2023, September 25). *Part 2: AI as Personal Tutor*. Harvard Business Publishing Education: https://hbsp.harvard.edu/inspiring-minds/ai-as-personal-tutor

Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M., & López-Meneses, E. (2023). Impact of the Implementation of ChatGPT in Education: A Systematic Review. *Computers*, *12*(8), Article 8. https://doi.org/10.3390/computers12080153

Munthe, E., Erstad, O., Njå, M. B., Gilje, Ø., Amdam, S., Moltudal, S., & Hagen, S. B. (n.d.). Digitalisering i grunnopplæring; kunnskap, trender og framtidig kunnskapsbehov. *Utdanningsdirektoratet*, 136. https://www.udir.no/tall-og-forskning/finn-forskning/rapporter/digitalisering-i-grunnopplaringen-bedre-muligheter-for-laring/

Nesse, P. J., & Erdal, O. B. (2022, July). Smart Digitalization in Nordic Cities and Municipalities Through Internet of Things. In *Economics and Finance Readings: Selected Papers from Asia-Pacific Conference on Economics & Finance, 2021* (pp. 33-55). Singapore: Springer Nature Singapore.

Nilsson, N. J. (1998). Artificial intelligence: a new synthesis. *Morgan Kaufmann*.

Norwegian School of Economics. (2022). *Annual report 2022*. https://www.nhh.no/contentassets/5b66c71995f24fb684740286cb8525d1/norwegian-school-of-economics_annual-report-2022.pdf

Norwegian School of Economics. (n.d.). Bachelor i økonomi og administrasjon. https://www.nhh.no/studier/bachelor-i-okonomi-og-administrasjon/

Norwegian University of Science and Technology (NTNU). (n.d.). Industriell økonomi og teknologiledelse. https://www.ntnu.no/studier/mtiot

NTNU KTDiM. (2013, November 29). *Kvalitet, tilgjengelighet og differensiering i grunnutdanningen i matematikk (KTDiM).* NTNU:

https://www.ntnu.no/documents/1268430158/0/IME_KTDIM_1027107_READONLY.pdf/9b01b76e-3477-4cd3-92e2-48b827fe743b

OECD. (2023). *Generative AI in the classroom*: From hype to reality? OECD. https://one.oecd.org/document/EDU/EDPC(2023)11/en/pdf

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *ScienceDirect*. https://doi.org/10.1016/j.caeai.2021.100033

Okonkwo, C. W., Huisman, M., & Taylor, E. (2019). The adoption of m- commerce applications: Rural dwellers perspectives. *12th, IADIS, International conference*. Information systems.

Parmiggiani, E., & Mikalef, P. (2022). The case of Norway and digital transformation over the years. *Digital Transformation in Norwegian Enterprises*, 11.

Patton, M. Q. (2002). Qualitative research and evaluation methods (3rd ed.). *Sage Publications*

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903. https://doi.org/10.1037/0021-9010.88.5.879

Pohl, W. (2023). *Er ChatGPT en fuskemaskin?* NHH Bulletin. https://www.nhh.no/nhh-bulletin/artikkelarkiv/2023/april/er-chatgpt-en-fuskemaskin/

Ramesh, D., & Sanampudi, S. K. (2022). An automated essay scoring systems: A systematic literature review. *Artificial Intelligence Review*, *55*(3), 2495–2527. https://doi.org/10.1007/s10462-021-10068-2

Ratten, V. (2023, February 17). *The post COVID-19 pandemic era: Changes in teaching and learning methods for management educators.* ScienceDirect: https://www.sciencedirect.com/science/article/pii/S1472811723000150

Roca, R. B. de. (2023, April 5). What is #prompting in AI and why is it crucial for effective within AI communication? *Medium*. https://roger-basler-de-roca.medium.com/what-is-prompting-in-ai-and-why-is-it-crucial-for-effective-within-ai-communication-d8cd4c91bfbd

Rose, M. (2009). *Writer's block: The cognitive dimension*. SIU Press.

Rusticus, S., & Lovato, C. (2019). Impact of Sample Size and Variability on the Power and Type I Error Rates of Equivalence Tests: A Simulation Study. *Practical Assessment, Research, and Evaluation*, *19*(1). https://doi.org/10.7275/4s9m-4e81

Russell, S. J., & Norvig, P. (2009). Artificial intelligence: A modern approach (3rd ed.). Pearson Education.

Sauesund, L., & Lia, L. M. (2023). Dette mener studiestedene om ChatGPT under eksamen. *Studvest*. https://www.studvest.no/nyhet/dette-mener-studiestedene- om-chatgpt-under-eksamen/121071

Saunders, M., Lewis, P., & Thornhill, A. (2019). Research Methods for Business Students (8th ed.). *Pearson Education Limited*.

Schwab, K. (2016, January 14). *The Fourth Industrial Revolution: What it means and how to respond*. World Economic Forum. https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/

Sekaran, U., & Bougie, R. (2016). Research methods for business: A skill-building approach (7th ed.). *Wiley*

Shawar, B. A., & Atwell, E. (2007). Chatbots: are they really useful?. *Journal for Language Technology and Computational Linguistics*, 22(1), 29-49.

Shapiro, H. T. (2009). A larger sense of purpose: Higher education and society. Princeton University Press.

Statistics LibreTexts. (2019, April 20). *7.1: The Central Limit Theorem for Sample Means*. Statistics LibreTexts. https://stats.libretexts.org/Bookshelves/Applied_Statistics/Introductory_Business_Statistics_(OpenStax)/07%3A_The_Central_Limit_Theorem/7.01%3A_The_Central_Limit_Theorem_for_Sample_Means

Streiner, D. L., & Norman, G. R. (2014). Health measurement scales: A practical guide to their development and use (5th ed.). *Oxford University Press*.

ssb.no. (2023). Students in higher education. https://www.ssb.no/en/utdanning/statistikker/utuvh

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys: The impact of data collection mode, question wording, and social context. *Psychological Science*, 18(3), 235-240.

University of Bergen. (n.d.). What is artificial intelligence? *https://www.uib.no/en/ai/152317/what-artificial-intelligence*

UCL. (2023, September 12). *Using generative AI (GenAI) in learning and teaching*. UCL: https://www.ucl.ac.uk/teaching-learning/publications/2023/sep/using- generative-ai-genai-learning-and-teaching

UiO. (2023). *GPT UiO*. Universitet i Oslo: https://www.uio.no/tjenester/it/gpt/

Vassdal, S. (2023, January 21). Forbyr ChatGPT. *TV2*. https://www.tv2.no/nyheter/utenriks/forbyr-chatgpt/15443825/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Vinyals, O., & Le, Q. (2015). A neural conversational model. arXiv preprint arXiv:1506.05869.

Walton Family Foundation (2023). 'ChatGPT Used by Teachers More Than Students, New Survey from Walton Family Foundation Finds'. *Walton Family Foundation*. https://www.waltonfamilyfoundation.org/chatgpt-used-by-teachers-more-than-students-new-survey-from-walton-family-foundation-finds.

West, R. M. (2021). Best practice in statistics: Use the Welch t-test when testing the difference between two groups. *International Journal of Laboratory Medicine*, *58*(4). https://doi.org/10.1177/0004563221992088

Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet?-a systematic literature review on chatbots in education. Frontiers in artificial intelligence, 4, 654924.

Workflow. (n.d.). Workflow by service now. *The Impact of AI on Implementation Consultant Skills*. https://www.servicenow.com/workflow/learn/impact-ai-implementation-consultant-skills.html

Wu, H., & Leung, S. O. (2017). Can Likert Scales be Treated as Interval Scales?—A Simulation Study. *Journal of Social Service Research*, 1–6. https://doi.org/10.1080/01488376.2017.1329775

Zulic, A.; Reikerås, M.; Granbo, K.. (2023, June 16). *1 av 5 elever bruker ChatGPT til skolearbeid*. NRK: https://www.nrk.no/kultur/na-sier-1-av-5-norske-elever-at-de-bruker-chatgpt-til-skolearbeid-1.16442993

8 Appendix

Appendix 1: Questionnaire

[redacted] Wed 11/8/2023 2:51 PM

Hi fellow student!

NHH and most other schools are currently making policies on use of generative AI such as ChatGPT. We happen to write our master thesis on this topic and invite all students to take a short and completely anonymous survey on your opinions about this. The data we gather can be helpful in informing the policies that are currently in making, so your opinions are greatly appreciated and important.

Here is a link to the short survey:
https://qualtricsxmxmndlmjbg.qualtrics.com/jfe/form/SV_bjDbRSazXs7bMjA

If you have any questions, just shoot us a message.

Thank you in advance for your participation! 😄

All the best,
[redacted]

Introduction

Purpose
This survey seeks to map the attitudes of students, lecturers and administrative staff at NHH towards adoption of generative artificial intelligence (AI) in higher education. More specifically, we are interested in how generative AI can be used as a tool for (1) generating text-answers for assignments and exams, (2) grading assignments and exams, and (3) teaching through virtual assistants.
Generative AI for research and administrative tasks are outside of this study's interests.

Procedures, anonymity and voluntary participation
The survey takes about 5-10 minutes to complete, and is ***completely anonymous***.
Participation is voluntary, and you can quit at any time.

Risks and benefits
There are absolutely no risks in answering this survey, as it is anonymous. A potential benefit of this survey is that the results might inform future NHH policies on use of generative AI - Your answers are therefore valuable and greatly appreciated.

Questions
If you have questions about this study you may contact any of the members of the research team:

[redacted]

Agreement to participate
If you agree that you

- have read this text
- do not have any questions regarding participation
- are at least 18 years of age
- wish to participate in this study

Press the NEXT >> below to begin the survey.

→

Profile data (1 of 5)

What is your gender?

- Male
- Female
- Non-binary
- Prefer not to say

What is your age?

- <=25 years old
- 26-35 years old
- 36-45 years old
- 46-55 years old
- 56-65 years old
- >65 years old

What is your role at NHH?

- Bachelor student
- Master student
- PhD student
- Professor/Lecturer
- Other (please specify)

→

What is your (main) profile?

ACC

BAN

BUS

ECO

ECON

FIE

MBM

NBD

STRAT

MRR

Other (please specify)

→

Follow up question for master students

What department are you in?

Accounting, auditing and law

Business and management science

Economics

Finance

Professional and intercultural communication

Strategy and management

Other (please specify)

→

Follow up question for professors and PhD

What is generative artificial intelligence (AI)?

Generative AI refers to a type of artificial intelligence that can create new content, such as images, music, or text, by learning from existing data and patterns. Online tools such as ChatGPT for text creation and Midjourney for image creation are examples of generative AI. For these tools, you write a prompt, and they generate the desired content based on that instruction.

→

General attitudes towards generative AI (2 of 5)

In this section, we want to map your experience and general attitudes towards the field of generative AI

How often do you use generative AI tools like ChatGPT, Bard, Midjourney etc.? This can be either for work or private use.

- I have never used it
- I have tried it out of curiosity
- I have used it a couple of times
- I use it monthly
- I use it weekly
- I use it daily

→

What tools have you used (multiple answers are allowed)

ChatGPT

BingChat

Midjourney

Bard

DALL-E

DeepArt

RunwayML

ChatPDF

Canva (specifically using the magic design function)

Other (please specify)

→

This question was displayed for every participants who did not answered “I have never used it” on the previous question

Assess the following statements

	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree
I think generative AI is an interesting field	○	○	○	○	○
I am confident with using generative AI tools like ChatGPT	○	○	○	○	○
I think generative AI will have a significant impact on higher education	○	○	○	○	○
I consider generative AI as a progressive step towards the future of higher education	○	○	○	○	○
I consider generative AI as a potential danger towards higher education	○	○	○	○	○

→

What do you consider the most significant benefit of using generative AI in higher education?

Text generation - Creating, rephrasing and restructuring text

Assessment - Getting instant assessments on exams and assignments

Virtual assistant - Personalized learning support (i.e., how to improve your text)

Idea generation - Brainstorming with generative AI

Summarizing texts (i.e., skimming academic papers, articles and textbooks)

Other (please specify)

What do you consider the most significant challenge of using generative AI in higher education?

Reduced learning for students due to AI "doing the work"

Reduced teaching quality due to AI "doing the work"

Receiving non-factual or misleading information due to engine inaccuracy

Receiving non-factual or misleading information due to inherent biases in the generative AI's dataset

Data privacy

Other (please specify)

→

Text generation (3 of 5)

A chatbot (i.e., ChatGPT) is a software program designed to stimulate conversation with human users by taking text-prompts as input, and generating text answers in return.

In this section, we want to map where you draw the line for which chatbot-features should be allowed to use for exams and assignments in higher education, and which features should not be allowed.

For exams and assignments, the following chatbot-features should be a permitted tool:

	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree
Correcting grammar	◯	◯	◯	◯	◯
Restructuring and rephrasing text to enhance readability and structure	◯	◯	◯	◯	◯
Generating ideas	◯	◯	◯	◯	◯
All functions should be allowed, including complete text generation of answers	◯	◯	◯	◯	◯
No features should be allowed - Chatbots should not be permitted at all	◯	◯	◯	◯	◯

→

Assessment (4 of 5)

Generative AI can be used to analyze and assess the quality of work for a given assignment or exam, as well as providing an explanation for the assessment.

In this section, we want to map your attitudes towards the use of generative AI as a tool to assess assignments and exams in higher education.

Have you used generative AI as a complementary tool to assess assignments or exams?

- I have never used it
- I have tried it out of curiosity
- I have used it a couple of times
- I often use it
- I always use it
- N/A - I do not grade assignments or exams

Have you used generative AI as a complementary tool to make questions for assignments or exams?

- I have never used it
- I have tried it our of curiosity
- I have used it a couple of times
- I often use it
- I always use it
- N/A - I do not make questions for assignments or exams

The question above displayed only for lecturers and PhD students

Assess the following statements on the use of generative AI for assessment in higher education

	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree
Generative AI should be used as a complementary tool for human graders to assess ***graded exams and assignments***	○	○	○	○	○
Generative AI should be used as a complementary tool for human graders to assess ***non-graded assignments***	○	○	○	○	○
Generative AI should be used to ***grade exams and assignments*** without human intervention	○	○	○	○	○
Generative AI should be used to assess ***non-graded assignments*** without human intervention	○	○	○	○	○
Generative AI will give more fair feedback than human graders	○	○	○	○	○

Virtual assistant (5 of 5)

The text-generating and assessment capabilities of generative AI can also be applied to create virtual assistants which could answer student questions on various academic topics. These assistants could be instructed to "act" as a personal tutor, which instead of simply answering questions, could ask questions guiding which actively promote critical thinking and learning through engagement.

In this section, we want to map your attitudes towards the use of generative AI as virtual assistants in higher education.

What do you consider the most significant benefit of using generative AI as a virtual assistant in higher education?

- Personalized feedback
- Instant feedback
- Asking questions anytime
- Asking questions from anywhere
- Possibility of discussing topics with somebody (creates dialogue)
- Reduced barrier to ask for help
- None of the above (please specify)

What do you consider the most significant challenge of using generative AI as a virtual assistant in higher education?

- Non-factual or misleading information due to engine inaccuracy
- Non-factual or misleading information due to inherent biases in the generative AI's dataset
- Reduced contact between students and professors
- Rubric misalignment - Virtual assistants might focus on different aspects of course material than what professors find important for grading
- Data privacy
- None of the above (please specify)

Assess the following statements

	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree
I would prefer to interact with a chatbot instead of a **professor** to solve a problem	○	○	○	○	○
I would prefer to interact with a chatbot instead of a **student assistant** to solve a problem	○	○	○	○	○
I would prefer to interact with a chatbot instead of a **fellow student** to solve a problem	○	○	○	○	○

In a higher education context, [...]

	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree
I value human advice more than generative AI advice	○	○	○	○	○
implementing chatbots as virtual assistants will benefit student learning	○	○	○	○	○
virtual assistants can help students achieve a better grade by serving as a complementary teaching tool	○	○	○	○	○

→

Appendix 2: Descriptive statistics among the academic roles

Overall Summary Statistics of Likert Scale Responses						
Statement	Count	Mean	Median	Standard Deviation	Min	Max
I think GenAI is an interesting field	206	4.73	5	0.56	2	5
I am confident with using GenAI tools like ChatGPT	206	3.89	4	1.12	1	5
I think GenAI will have a significant impact on higher	206	4.62	5	0.64	2	5
I consider GenAI as a progressive step towards the future of higher education	205	4.06	4	1.04	1	5
I consider GenAI as a potential danger towards higher education	205	3.35	4	1.28	1	5
GenAI should be a permitted tool for correcting grammar	205	4.32	5	1.12	1	5
GenAI should be a permitted tool for restructuring	204	3.74	4	1.38	1	5
GenAI should be a permitted tool for generating ideas	203	3.20	3	1.42	1	5
GenAI should be a permitted tool in all its functions	205	2.20	2	1.41	1	5
GenAI should not be a permitted tool	204	1.97	1	1.23	1	5
For graded exams and submissions: GenAI should be a complimentary tool to assess	204	2.93	3	1.33	1	5
For non-graded exams and submissions: GenAI should be a complimentary tool to assess	205	3.45	4	1.30	1	5
For graded exams and submissions: GenAI should be used to grade without human intervention	205	1.49	1	0.88	1	5
For non-graded exams and submissions: GenAI should be used to grade without human intervention	203	1.98	2	1.16	1	5
GenAI will give more fair feedback than human graders	205	2.65	3	1.12	1	5
Prefer to interact with a chatbot instead of a professor to solve a problem	151	2.37	2	1.30	1	5
Prefer to interact with a chatbot instead of a student assistant to solve a problem	152	2.66	2	1.31	1	5
Prefer to interact with a chatbot instead of a fellow student to solve a problem	149	2.32	2	1.23	1	5
I value human advice more than generative AI advice	198	4.27	4	0.88	1	5
implementing chatbots as virtual assistants will benefit student learning	198	3.90	4	0.84	2	5
VAs can help students achieve a better grade by serving as a complementary teaching tool	198	3.99	4	0.96	1	5

Table 7 - Descriptive statistics for the whole sample (BA + MS + PhD + professors)

Layer 1: Descriptive statistics

Students Summary Statistics of Likert Scale Responses						
Statement	Count	Mean	Median	Standard Deviation	Min	Max
I think GenAI is an interesting field	143	4.72	5.0	0.56	2	5
I am confident with using GenAI tools like ChatGPT	142	4.04	4.0	1.01	1	5
I think GenAI will have a significant impact on higher	143	4.69	5.0	0.56	2	5
I consider GenAI as a progressive step towards the future of higher education	142	4.22	4.5	0.97	1	5
I consider GenAI as a potential danger towards higher education	142	3.39	4.0	1.25	1	5
GenAI should be a permitted tool for correcting grammar	142	4.34	5.0	1.11	1	5
GenAI should be a permitted tool for restructuring	141	3.82	4.0	1.34	1	5
GenAI should be a permitted tool for generating ideas	141	3.33	4.0	1.42	1	5
GenAI should be a permitted tool in all its functions	142	2.31	2.0	1.44	1	5
GenAI should not be a permitted tool	142	1.96	1.5	1.21	1	5
For graded exams and submissions: GenAI should be a complimentary tool to assess	141	2.88	3.0	1.35	1	5
For non-graded exams and submissions: GenAI should be a complimentary tool to assess	142	3.52	4.0	1.34	1	5
For graded exams and submissions: GenAI should be used to grade without human intervention	142	1.51	1.0	0.88	1	5
For non-graded exams and submissions: GenAI should be used to grade without human intervention	140	2.01	2.0	1.15	1	5
GenAI will give more fair feedback than human graders	142	2.80	3.0	1.11	1	5
Prefer to interact with a chatbot instead of a professor to solve a problem	134	2.37	2.0	1.27	1	5
Prefer to interact with a chatbot instead of a student assistant to solve a problem	135	2.73	2.0	1.29	1	5
Prefer to interact with a chatbot instead of a fellow student to solve a problem	132	2.34	2.0	1.21	1	5
I value human advice more than generative AI advice	135	4.21	4.0	0.86	2	5
implementing chatbots as virtual assistants will benefit student learning	135	4.04	4.0	0.77	2	5
VAs can help students achieve a better grade by serving as a complementary teaching tool	135	4.14	4.0	0.91	1	5

Table 8 - Descriptive statistics for Students (BA + MS)

Professors and Lecturers Summary Statistics of Likert Scale Responses						
Statement	Count	Mean	Median	Standard Deviation	Min	Max
I think GenAI is an interesting field	46	4.72	5.0	0.62	2	5
I am confident with using GenAI tools like ChatGPT	47	3.36	4.0	1.39	1	5
I think GenAI will have a significant impact on higher	46	4.48	5.0	0.78	2	5
I consider GenAI as a progressive step towards the future of higher education	46	3.74	4.0	1.14	1	5
I consider GenAI as a potential danger towards higher education	46	3.07	4.0	1.32	1	5
GenAI should be a permitted tool for correcting grammar	46	4.30	5.0	1.19	1	5
GenAI should be a permitted tool for restructuring	46	3.67	4.0	1.42	1	5
GenAI should be a permitted tool for generating ideas	45	2.87	3.0	1.46	1	5
GenAI should be a permitted tool in all its functions	46	2.15	1.5	1.41	1	5
GenAI should not be a permitted tool	45	1.96	1.0	1.35	1	5
For graded exams and submissions: GenAI should be a complimentary tool to assess	46	3.00	3.0	1.28	1	5
For non-graded exams and submissions: GenAI should be a complimentary tool to assess	46	3.30	3.0	1.19	1	5
For graded exams and submissions: GenAI should be used to grade without human intervention	46	1.59	1.0	0.96	1	5
For non-graded exams and submissions: GenAI should be used to grade without human intervention	46	2.17	2.0	1.23	1	5
GenAI will give more fair feedback than human graders	46	2.35	2.5	1.10	1	5
Prefer to interact with a chatbot instead of a professor to solve a problem	0	NaN	NA	NA	Inf	-Inf
Prefer to interact with a chatbot instead of a student assistant to solve a problem	0	NaN	NA	NA	Inf	-Inf
Prefer to interact with a chatbot instead of a fellow student to solve a problem	0	NaN	NA	NA	Inf	-Inf
I value human advice more than generative AI advice	46	4.46	5.0	0.78	3	5
implementing chatbots as virtual assistants will benefit student learning	46	3.63	4.0	0.90	2	5
VAs can help students achieve a better grade by serving as a complementary teaching tool	46	3.65	4.0	1.02	1	5

Table 9 - Descriptive statistics for Professors and lecturers

Layer 2: Descriptive statistics

Bachelor Summary Statistics of Likert Scale Responses						
Statement	Count	Mean	Median	Standard Deviation	Min	Max
I think GenAI is an interesting field	56	4.73	5.0	0.56	2	5
I am confident with using GenAI tools like ChatGPT	56	3.93	4.0	1.06	1	5
I think GenAI will have a significant impact on higher	56	4.68	5.0	0.51	3	5
I consider GenAI as a progressive step towards the future of higher education	56	4.20	4.0	0.88	2	5
I consider GenAI as a potential danger towards higher education	56	3.54	4.0	1.22	1	5
GenAI should be a permitted tool for correcting grammar	56	4.12	5.0	1.18	1	5
GenAI should be a permitted tool for restructuring	55	3.76	4.0	1.36	1	5
GenAI should be a permitted tool for generating ideas	56	3.30	3.5	1.40	1	5
GenAI should be a permitted tool in all its functions	56	2.36	2.0	1.43	1	5
GenAI should not be a permitted tool	56	2.16	2.0	1.20	1	5
For graded exams and submissions: GenAI should be a complimentary tool to assess	56	2.68	3.0	1.39	1	5
For non-graded exams and submissions: GenAI should be a complimentary tool to assess	56	3.29	4.0	1.46	1	5
For graded exams and submissions: GenAI should be used to grade without human intervention	56	1.55	1.0	0.99	1	5
For non-graded exams and submissions: GenAI should be used to grade without human intervention	55	1.95	2.0	1.13	1	5
GenAI will give more fair feedback than human graders	56	2.68	3.0	1.06	1	5
Prefer to interact with a chatbot instead of a professor to solve a problem	53	2.30	2.0	1.28	1	5
Prefer to interact with a chatbot instead of a student assistant to solve a problem	53	2.62	3.0	1.32	1	5
Prefer to interact with a chatbot instead of a fellow student to solve a problem	53	2.38	2.0	1.27	1	5
I value human advice more than generative AI advice	53	4.23	4.0	0.95	2	5
implementing chatbots as virtual assistants will benefit student learning	53	3.98	4.0	0.80	2	5
VAs can help students achieve a better grade by serving as a complementary teaching tool	53	4.11	4.0	0.89	2	5

Table 10 - Descriptive statistics for Bachelor students

Master Summary Statistics of Likert Scale Responses						
Statement	Count	Mean	Median	Standard Deviation	Min	Max
I think GenAI is an interesting field	87	4.71	5.0	0.57	2	5
I am confident with using GenAI tools like ChatGPT	86	4.10	4.0	0.98	1	5
I think GenAI will have a significant impact on higher	87	4.70	5.0	0.59	2	5
I consider GenAI as a progressive step towards the future of higher education	86	4.23	5.0	1.03	1	5
I consider GenAI as a potential danger towards higher education	86	3.29	3.5	1.27	1	5
GenAI should be a permitted tool for correcting grammar	86	4.48	5.0	1.05	1	5
GenAI should be a permitted tool for restructuring	86	3.86	4.0	1.34	1	5
GenAI should be a permitted tool for generating ideas	85	3.34	4.0	1.44	1	5
GenAI should be a permitted tool in all its functions	86	2.28	2.0	1.44	1	5
GenAI should not be a permitted tool	86	1.84	1.0	1.20	1	5
For graded exams and submissions: GenAI should be a complimentary tool to assess	85	3.01	3.0	1.32	1	5
For non-graded exams and submissions: GenAI should be a complimentary tool to assess	86	3.67	4.0	1.23	1	5
For graded exams and submissions: GenAI should be used to grade without human intervention	86	1.48	1.0	0.81	1	5
For non-graded exams and submissions: GenAI should be used to grade without human intervention	85	2.05	2.0	1.16	1	5
GenAI will give more fair feedback than human graders	86	2.88	3.0	1.13	1	5
Prefer to interact with a chatbot instead of a professor to solve a problem	81	2.42	2.0	1.27	1	5
Prefer to interact with a chatbot instead of a student assistant to solve a problem	82	2.79	2.0	1.27	1	5
Prefer to interact with a chatbot instead of a fellow student to solve a problem	79	2.32	2.0	1.17	1	5
I value human advice more than generative AI advice	82	4.20	4.0	0.79	2	5
implementing chatbots as virtual assistants will benefit student learning	82	4.09	4.0	0.76	2	5
VAs can help students achieve a better grade by serving as a complementary teaching tool	82	4.16	4.0	0.92	1	5

Table 11 - Descriptive statistics for Master students

PhD Summary Statistics of Likert Scale Responses						
Statement	Count	Mean	Median	Standard Deviation	Min	Max
I think GenAI is an interesting field	17	4.82	5	0.39	4	5
I am confident with using GenAI tools like ChatGPT	17	4.12	4	0.70	3	5
I think GenAI will have a significant impact on higher	17	4.35	4	0.79	2	5
I consider GenAI as a progressive step towards the future of higher education	17	3.65	4	1.06	1	5
I consider GenAI as a potential danger towards higher education	17	3.82	4	1.24	1	5
GenAI should be a permitted tool for correcting grammar	17	4.18	4	1.07	1	5
GenAI should be a permitted tool for restructuring	17	3.18	4	1.51	1	5
GenAI should be a permitted tool for generating ideas	17	3.00	3	1.27	1	5
GenAI should be a permitted tool in all its functions	17	1.47	1	0.94	1	4
GenAI should not be a permitted tool	17	2.00	2	1.12	1	5
For graded exams and submissions: GenAI should be a complimentary tool to assess	17	3.12	3	1.32	1	5
For non-graded exams and submissions: GenAI should be a complimentary tool to assess	17	3.24	4	1.30	1	5
For graded exams and submissions: GenAI should be used to grade without human intervention	17	1.12	1	0.49	1	3
For non-graded exams and submissions: GenAI should be used to grade without human intervention	17	1.18	1	0.73	1	4
GenAI will give more fair feedback than human graders	17	2.24	3	1.03	1	4
Prefer to interact with a chatbot instead of a professor to solve a problem	17	2.35	2	1.54	1	5
Prefer to interact with a chatbot instead of a student assistant to solve a problem	17	2.18	2	1.38	1	5
Prefer to interact with a chatbot instead of a fellow student to solve a problem	17	2.18	2	1.38	1	5
I value human advice more than generative AI advice	17	4.24	5	1.25	1	5
implementing chatbots as virtual assistants will benefit student learning	17	3.47	3	0.87	2	5
VAs can help students achieve a better grade by serving as a complementary teaching tool	17	3.71	4	0.92	1	5

Table 12 - Descriptive statistics for PhD students

Appendix 3: A selection of normality tests showing lack of normality

"I think generative AI is an interesting field"

```
        Shapiro-Wilk normality test

data:  residuals(ANOVA_L1_AI_interesting)
W = 0.55379, p-value < 2.2e-16
```

"I am confident with using generative AI tools like ChatGPT"

```
        Shapiro-Wilk normality test

data:  residuals(ANOVA_L1_AI_confident)
W = 0.92026, p-value = 4.079e-09
```

"I think generative AI will have a significant impact on higher education"

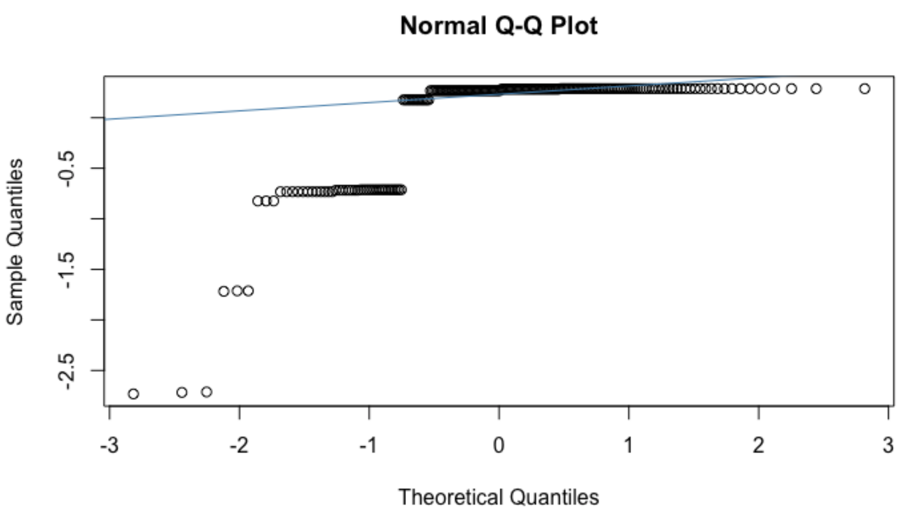
```
        Shapiro-Wilk normality test

data:  residuals(ANOVA_L1_impact)
W = 0.74843, p-value < 2.2e-16
```

"I consider generative AI as a progressive step towards the future of higher education"

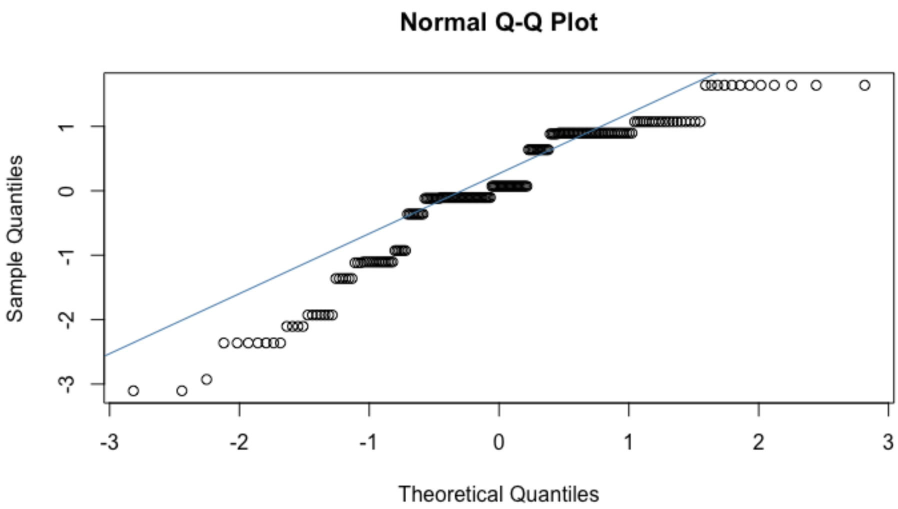
```
        Shapiro-Wilk normality test

data:  residuals(ANOVA_L1_future)
W = 0.87396, p-value = 4.923e-12
```

…

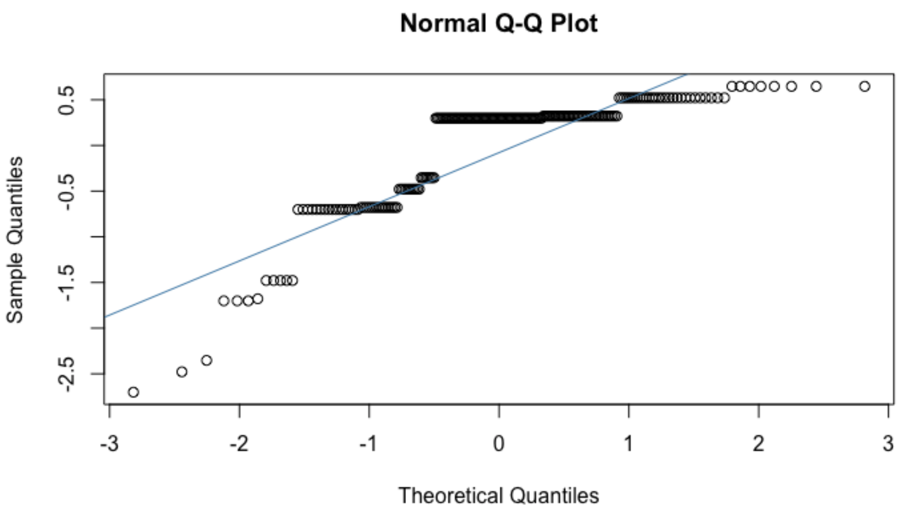
"I think generative AI is an interesting field"



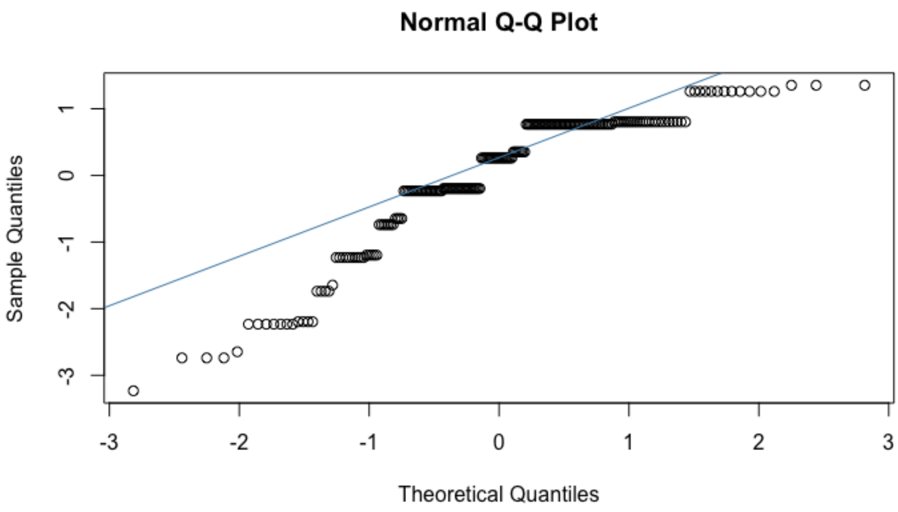
"I am confident with using generative AI tools like ChatGPT"



"I think generative AI will have a significant impact on higher education"



"I consider generative AI as a progressive step towards the future of higher education"



…

Appendix 4: A selection of Levene's test indicating heteroskedasticity

"I am confident with using GenAI tools"

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value   Pr(>F)
group   3  4.6257 0.003751 **
      202
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

"All functions should be allowed"

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value  Pr(>F)
group   3  2.7596 0.04334 *
      201
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

"GenAI assess without human intervention"

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value   Pr(>F)
group   3  5.2253 0.001712 **
      199
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix 5: Welch's t-test for Layer 1

Section 1: General attitudes

"I think GenAI is an interesting field"

```
	Welch Two Sample t-test

data:  L1_AI_interesting by Role
t = -0.028085, df = 70.308, p-value = 0.9777
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.2079885  0.2022117
sample estimates:
mean in group Professor/Lecturer            mean in group Student
                        4.717391                         4.720280
```

"I am confident with using GenAI tools like ChatGPT"

```
	Welch Two Sample t-test

data:  L1_AI_confident by Role
t = -3.0636, df = 62.981, p-value = 0.003216
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -1.1128284 -0.2341899
sample estimates:
mean in group Professor/Lecturer            mean in group Student
                        3.361702                         4.035211
```

"I think GenAI will have a significant impact on higher"

```
	Welch Two Sample t-test

data:  L1_AI_confident by Role
t = -3.0636, df = 62.981, p-value = 0.003216
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -1.1128284 -0.2341899
sample estimates:
mean in group Professor/Lecturer            mean in group Student
                        3.361702                         4.035211
```

"I consider GenAI as a progressive step towards the future of higher education"

```
        Welch Two Sample t-test

data:  L1_future by Role
t = -2.5597, df = 67.171, p-value = 0.01273
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.8528143 -0.1055446
sample estimates:
mean in group Professor/Lecturer           mean in group Student
                         3.73913                          4.21831
```

"I consider GenAI as a potential danger towards higher education"

```
        Welch Two Sample t-test

data:  L1_danger by Role
t = -1.453, df = 73.012, p-value = 0.1505
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.7639147  0.1197016
sample estimates:
mean in group Professor/Lecturer           mean in group Student
                        3.065217                         3.387324
```

Section 2: Text generation

"GenAI should be a permitted tool for correcting grammar"

```
        Welch Two Sample t-test

data:  S3L_grammar by Role
t = -0.16953, df = 72.115, p-value = 0.8659
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.4297174  0.3623567
sample estimates:
mean in group Professor/Lecturer           mean in group Student
                        4.304348                         4.338028
```

"GenAI should be a permitted tool for restructuring"

```
        Welch Two Sample t-test

data:  S3L_restructure by Role
t = -0.6269, df = 73.305, p-value = 0.5327
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.6217466  0.3241826
sample estimates:
mean in group Professor/Lecturer           mean in group Student
                        3.673913                         3.822695
```

"GenAI should be a permitted tool for generating ideas"

```
        Welch Two Sample t-test

data:  S3L_ideas by Role
t = -1.8561, df = 72.551, p-value = 0.0675
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.95309060  0.03394166
sample estimates:
mean in group Professor/Lecturer          mean in group Student
                        2.866667                       3.326241
```

"GenAI should be a permitted tool in all its functions"

```
        Welch Two Sample t-test

data:  S3L_all by Role
t = -0.65504, df = 77.307, p-value = 0.5144
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.6370020  0.3216316
sample estimates:
mean in group Professor/Lecturer          mean in group Student
                        2.152174                       2.309859
```

"GenAI should not be a permitted tool"


```
        Welch Two Sample t-test

data:  S3L_all by Role
t = -0.65504, df = 77.307, p-value = 0.5144
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.6370020  0.3216316
sample estimates:
mean in group Professor/Lecturer          mean in group Student
                        2.152174                       2.309859
```


Section 3: Assessment

"For graded exams and assignments: GenAI should be a complimentary tool to assess"

```
        Welch Two Sample t-test

data:  S4L_complementary_graded by Role
t = 0.54598, df = 80.314, p-value = 0.5866
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.3188703  0.5600051
sample estimates:
mean in group Professor/Lecturer          mean in group Student
                        3.000000                       2.879433
```

"For non-graded exams and assignments: GenAI should be a complimentary tool to assess"

```
        Welch Two Sample t-test

data:  S4L_complementary_non_graded by Role
t = -1.0412, df = 84.694, p-value = 0.3008
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.6307704  0.1972126
sample estimates:
mean in group Professor/Lecturer            mean in group Student
                        3.304348                         3.521127
```

"For graded exams and assignments: GenAI should be used to grade without human intervention"

```
        Welch Two Sample t-test

data:  S4L_no_human_graded by Role
t = 0.50193, df = 71.437, p-value = 0.6173
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.2375146  0.3973431
sample estimates:
mean in group Professor/Lecturer            mean in group Student
                        1.586957                         1.507042
```

"For non-graded exams and assignments: GenAI should be used to grade without human intervention"

```
        Welch Two Sample t-test

data:  S4L_no_human_non_graded by Role
t = 0.8085, df = 72.273, p-value = 0.4215
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.2443975  0.5779379
sample estimates:
mean in group Professor/Lecturer            mean in group Student
                        2.173913                         2.007143
```

"GenAI will give more fair feedback than human graders"

```
        Welch Two Sample t-test

data:  S4L_fair_feedback by Role
t = -2.4351, df = 76.685, p-value = 0.01721
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.82707330 -0.08290833
sample estimates:
mean in group Professor/Lecturer            mean in group Student
                        2.347826                         2.802817
```

Section 4: GenAI powered virtual assistants

"I value human advice more than generative AI advice"

```
        Welch Two Sample t-test

data:  S5L_human_advice by Role
t = 1.8231, df = 84.672, p-value = 0.07181
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.02258166  0.52081032
sample estimates:
mean in group Professor/Lecturer            mean in group Student
                        4.456522                         4.207407
```

"Implementing chatbots as virtual assistants will benefit student learning"

```
        Welch Two Sample t-test

data:  S5L_chatbot_benefit_studentlearning by Role
t = -2.7821, df = 68.717, p-value = 0.006967
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.7109063 -0.1171130
sample estimates:
mean in group Professor/Lecturer            mean in group Student
                        3.630435                         4.044444
```

"VAs can help students achieve a better grade by serving as a complementary teaching tool"

```
        Welch Two Sample t-test

data:  S5L_chatbot_better_grade by Role
t = -2.8924, df = 71.032, p-value = 0.00507
alternative hypothesis: true difference in means between group Professor/Lecturer and group Student is not equal to 0
95 percent confidence interval:
 -0.8253636 -0.1517701
sample estimates:
mean in group Professor/Lecturer            mean in group Student
                        3.652174                         4.140741
```

Appendix 6: Welch’s ANOVA for Layer 2

Section 1: General attitudes

“I think GenAI is an interesting field”

```
        One-way analysis of means (not assuming equal variances)

data:  L1_AI_interesting and Role
F = 0.33809, num df = 3.000, denom df = 68.474, p-value = 0.7978
```

"I am confident with using GenAI tools like ChatGPT"

```
        One-way analysis of means (not assuming equal variances)

data:  L1_AI_confident and Role
F = 3.7497, num df = 3.000, denom df = 68.625, p-value = 0.01482
```

"I think GenAI will have a significant impact on higher"

```
        One-way analysis of means (not assuming equal variances)

data:  L1_impact and Role
F = 1.7821, num df = 3.000, denom df = 59.578, p-value = 0.1603
```

"I consider GenAI as a progressive step towards the future of higher education"

```
        One-way analysis of means (not assuming equal variances)

data:  L1_future and Role
F = 3.1946, num df = 3.00, denom df = 62.03, p-value = 0.02953
```

"I consider GenAI as a potential danger towards higher education"

```
        One-way analysis of means (not assuming equal variances)

data:  L1_danger and Role
F = 1.9865, num df = 3.000, denom df = 63.078, p-value = 0.125
```

Section 2: Text generation

"GenAI should be a permitted tool for correcting grammar"

```
        One-way analysis of means (not assuming equal variances)

data:  S3L_grammar and Role
F = 1.2249, num df = 3.000, denom df = 62.541, p-value = 0.3081
```

"GenAI should be a permitted tool for restructuring"

```
        One-way analysis of means (not assuming equal variances)

data:  S3L_restructure and Role
F = 1.0363, num df = 3.000, denom df = 61.472, p-value = 0.3829
```

"GenAI should be a permitted tool for generating ideas"

```
        One-way analysis of means (not assuming equal variances)

data:  S3L_ideas and Role
F = 1.2754, num df = 3.000, denom df = 64.039, p-value = 0.2904
```

"GenAI should be a permitted tool in all its functions"

```
        One-way analysis of means (not assuming equal variances)

data:  S3L_all and Role
F = 3.4812, num df = 3.000, denom df = 69.769, p-value = 0.02034
```

"GenAI should not be a permitted tool"

```
        One-way analysis of means (not assuming equal variances)

data:  S3L_none and Role
F = 0.81039, num df = 3.000, denom df = 63.244, p-value = 0.4928
```

Section 3: Assessment

"For graded exams and assignments: GenAI should be a complimentary tool to assess"

```
        One-way analysis of means (not assuming equal variances)

data:  S4L_complementary_graded and Role
F = 0.86111, num df = 3.000, denom df = 63.081, p-value = 0.466
```

"For non-graded exams and assignments: GenAI should be a complimentary tool to assess"

```
        One-way analysis of means (not assuming equal variances)

data:  S4L_complementary_non_graded and Role
F = 1.5333, num df = 3.000, denom df = 62.498, p-value = 0.2147
```

"For graded exams and assignments: GenAI should be used to grade without human intervention"

```
        One-way analysis of means (not assuming equal variances)

data:  S4L_no_human_graded and Role
F = 3.0912, num df = 3.000, denom df = 72.955, p-value = 0.03224
```

"For non-graded exams and assignments: GenAI should be used to grade without human intervention"

```
        One-way analysis of means (not assuming equal variances)

data:  S4L_no_human_non_graded and Role
F = 6.6853, num df = 3.000, denom df = 70.634, p-value = 0.0004893
```

"GenAI will give more fair feedback than human graders"

```
        One-way analysis of means (not assuming equal variances)

data:  S4L_fair_feedback and Role
F = 3.2384, num df = 3.000, denom df = 63.877, p-value = 0.02782
```

Section 4: GenAI-powered virtual assistants

"Prefer to interact with a chatbot instead of a professor to solve a problem"

```
        One-way analysis of means (not assuming equal variances)

data:  S5L_students_interact_professor and Role
F = 0.13565, num df = 2.000, denom df = 42.227, p-value = 0.8735
```

"Prefer to interact with a chatbot instead of a student assistant to solve a problem"

```
        One-way analysis of means (not assuming equal variances)

data:  S5L_students_interact_studentassistant and Role
F = 1.4781, num df = 2.000, denom df = 43.373, p-value = 0.2393
```

"Prefer to interact with a chatbot instead of a fellow student to solve a problem"

```
        One-way analysis of means (not assuming equal variances)

data:  S5L_students_interact_student and Role
F = 0.143, num df = 2.000, denom df = 42.562, p-value = 0.8672
```

"I value human advice more than generative AI advice"

```
        One-way analysis of means (not assuming equal variances)

data:  S5L_human_advice and Role
F = 1.133, num df = 3.000, denom df = 59.014, p-value = 0.3432
```

"Implementing chatbots as virtual assistants will benefit student learning"

```
        One-way analysis of means (not assuming equal variances)

data:  S5L_chatbot_benefit_studentlearning and Role
F = 4.3787, num df = 3.000, denom df = 60.879, p-value = 0.007418
```

"VAs can help students achieve a better grade by serving as a complementary teaching tool"

```
        One-way analysis of means (not assuming equal variances)

data:  S5L_chatbot_better_grade and Role
F = 3.3942, num df = 3.000, denom df = 62.461, p-value = 0.02326
```

Appendix 7: Games-Howell tests for Layer 2

"I am confident with using generative AI tools like ChatGPT"

```
# A tibble: 6 × 8
  .y.             group1   group2              estimate conf.low conf.high p.adj p.adj.signif
* <chr>           <chr>    <chr>                  <dbl>    <dbl>     <dbl> <dbl> <chr>
1 L1_AI_confident Bachelor Master                0.176    -0.285    0.637  0.752 ns
2 L1_AI_confident Bachelor PhD                   0.189    -0.401    0.780  0.826 ns
3 L1_AI_confident Bachelor Professor/Lecturer   -0.567    -1.21     0.0811 0.108 ns
4 L1_AI_confident Master   PhD                   0.0130   -0.529    0.555  1     ns
5 L1_AI_confident Master   Professor/Lecturer   -0.743    -1.34    -0.141  0.009 **
6 L1_AI_confident PhD      Professor/Lecturer   -0.756    -1.45    -0.0569 0.029 *
```

"I consider generative AI as a progressive step towards the future of higher education"

```
# A tibble: 6 × 8
  .y.       group1   group2             estimate conf.low conf.high p.adj p.adj.signif
* <chr>     <chr>    <chr>                 <dbl>    <dbl>     <dbl> <dbl> <chr>
1 L1_future Bachelor Master               0.0361   -0.385    0.457  0.996 ns
2 L1_future Bachelor PhD                 -0.549    -1.33     0.231  0.237 ns
3 L1_future Bachelor Professor/Lecturer  -0.457    -0.997    0.0822 0.126 ns
4 L1_future Master   PhD                 -0.585    -1.36     0.189  0.185 ns
5 L1_future Master   Professor/Lecturer  -0.493    -1.02     0.0352 0.076 ns
6 L1_future PhD      Professor/Lecturer   0.0921   -0.741    0.925  0.99  ns
```

"All functions should be allowed, including complete text generation of answers"

```
# A tibble: 6 × 8
  .y.     group1   group2             estimate conf.low conf.high p.adj p.adj.signif
* <chr>   <chr>    <chr>                 <dbl>    <dbl>     <dbl> <dbl> <chr>
1 S3L_all Bachelor Master              -0.0781   -0.721    0.565  0.989 ns
2 S3L_all Bachelor PhD                 -0.887    -1.69    -0.0874 0.025 *
3 S3L_all Bachelor Professor/Lecturer  -0.205    -0.945    0.535  0.887 ns
4 S3L_all Master   PhD                 -0.808    -1.56    -0.0598 0.03  *
5 S3L_all Master   Professor/Lecturer  -0.127    -0.807    0.554  0.962 ns
6 S3L_all PhD      Professor/Lecturer   0.682    -0.145    1.51   0.139 ns
```

"Generative AI should be used to grade exams and assignments without human intervention"

```
# A tibble: 6 × 8
  .y.                 group1   group2             estimate conf.low conf.high p.adj p.adj.signif
* <chr>               <chr>    <chr>                 <dbl>    <dbl>     <dbl> <dbl> <chr>
1 S4L_no_human_graded Bachelor Master              -0.0768   -0.490    0.337  0.962 ns
2 S4L_no_human_graded Bachelor PhD                 -0.436    -0.905    0.0327 0.077 ns
3 S4L_no_human_graded Bachelor Professor/Lecturer   0.0334   -0.472    0.539  0.998 ns
4 S4L_no_human_graded Master   PhD                 -0.359    -0.753    0.0350 0.085 ns
5 S4L_no_human_graded Master   Professor/Lecturer   0.110    -0.325    0.545  0.91  ns
6 S4L_no_human_graded PhD      Professor/Lecturer   0.469    -0.0173   0.956  0.062 ns
```

"Generative AI should be used to assess non-grade exams and assignments without human intervention"

```
# A tibble: 6 × 8
  .y.                     group1   group2             estimate conf.low conf.high p.adj p.adj.si
* <chr>                   <chr>    <chr>                 <dbl>    <dbl>     <dbl> <dbl> <chr>
1 S4L_no_human_non_graded Bachelor Master                0.102   -0.414     0.617 0.956 ns
2 S4L_no_human_non_graded Bachelor PhD                  -0.769   -1.39     -0.145 0.01  **
3 S4L_no_human_non_graded Bachelor Professor/Lecturer    0.228   -0.392     0.849 0.771 ns
4 S4L_no_human_non_graded Master   PhD                  -0.871   -1.46     -0.285 0.002 **
5 S4L_no_human_non_graded Master   Professor/Lecturer    0.127   -0.453     0.707 0.94  ns
6 S4L_no_human_non_graded PhD      Professor/Lecturer    0.997    0.323     1.67  0.001 ***
```

"Generative AI will give more fair feedback than human graders"

```
# A tibble: 6 × 8
  .y.               group1   group2             estimate conf.low conf.high p.adj p.adj.signif
* <chr>             <chr>    <chr>                 <dbl>    <dbl>     <dbl> <dbl> <chr>
1 S4L_fair_feedback Bachelor Master                0.205   -0.283   0.693   0.693 ns
2 S4L_fair_feedback Bachelor PhD                  -0.443   -1.23    0.344   0.429 ns
3 S4L_fair_feedback Bachelor Professor/Lecturer   -0.331   -0.895   0.233   0.422 ns
4 S4L_fair_feedback Master   PhD                  -0.648   -1.42    0.120   0.119 ns
5 S4L_fair_feedback Master   Professor/Lecturer   -0.536   -1.07   -0.00512 0.047 *
6 S4L_fair_feedback PhD      Professor/Lecturer    0.113   -0.698   0.923   0.981 ns
```

"Implementing chatbots as virtual assistants will benefit student learning"

```
# A tibble: 6 × 8
  .y.                               group1   group2             estimate conf.low conf.high p.adj p.adj.signif
* <chr>                             <chr>    <chr>                 <dbl>    <dbl>     <dbl> <dbl> <chr>
1 S5L_chatbot_benefit_studentlearning Bachelor Master                0.104   -0.255    0.464  0.873 ns
2 S5L_chatbot_benefit_studentlearning Bachelor PhD                  -0.511   -1.17     0.146  0.168 ns
3 S5L_chatbot_benefit_studentlearning Bachelor Professor/Lecturer   -0.351   -0.802    0.100  0.183 ns
4 S5L_chatbot_benefit_studentlearning Master   PhD                  -0.615   -1.25     0.0200 0.06  ns
5 S5L_chatbot_benefit_studentlearning Master   Professor/Lecturer   -0.455   -0.867   -0.0424 0.025 *
6 S5L_chatbot_benefit_studentlearning PhD      Professor/Lecturer    0.160   -0.522    0.842  0.919 ns
```

"Virtual assistants can help students achieve a better grade by serving as a complementary teaching tool"

```
# A tibble: 6 × 8
  .y.                    group1   group2             estimate conf.low conf.high p.adj p.adj.s
* <chr>                  <chr>    <chr>                 <dbl>    <dbl>     <dbl> <dbl> <chr>
1 S5L_chatbot_better_grade Bachelor Master               0.0453   -0.370    0.461  0.992 ns
2 S5L_chatbot_better_grade Bachelor PhD                 -0.407    -1.10     0.290  0.395 ns
3 S5L_chatbot_better_grade Bachelor Professor/Lecturer  -0.461    -0.967    0.0453 0.088 ns
4 S5L_chatbot_better_grade Master   PhD                 -0.453    -1.13     0.225  0.278 ns
5 S5L_chatbot_better_grade Master   Professor/Lecturer  -0.506    -0.981   -0.0318 0.032 *
6 S5L_chatbot_better_grade PhD      Professor/Lecturer  -0.0537   -0.782    0.675  0.997 ns
```