



# FAUXperience Framework

## Designing For Critical Conscious AI Use In Higher Education

Master's Thesis in Computer Science and Engineering

JOSÉ BENER DE SOUSA NUNES



MASTER'S THESIS 2024

# **FAUX**perience Framework

## Designing For Critical Conscious AI Use In Higher Education

JOSÉ BENER DE SOUSA NUNES



UNIVERSITY OF  
GOTHENBURG

---



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering  
CHALMERS UNIVERSITY OF TECHNOLOGY  
UNIVERSITY OF GOTHENBURG  
Gothenburg, Sweden 2024

FAUXperience Framework - Designing For Critical Conscious AI Use In  
Higher Education  
JOSÉ BENER DE SOUSA NUNES

© JOSÉ BENER DE SOUSA NUNES, 2024.

Supervisor: Sara Ljungblad, Computer Science and Engineering  
Examiner: Staffan Björk, Computer Science and Engineering

Master's Thesis 2024  
Department of Computer Science and Engineering  
Chalmers University of Technology and University of Gothenburg  
SE-412 96 Gothenburg  
Telephone +46 31 772 1000

Cover: The original hand image was produced using generative AI before the author's final editing. *DALL-E 3 Prompt: High fidelity, black woman's hand, holding a diploma.*

Typeset in L<sup>A</sup>T<sub>E</sub>X  
Gothenburg, Sweden 2024

FAUXperience Framework - Designing For Critical Conscious AI Use In Higher Education

JOSÉ BENER DE SOUSA NUNES

Department of Computer Science and Engineering

Chalmers University of Technology and University of Gothenburg

## **Abstract**

Integrating generative artificial intelligence (AI) into education is a complex process that requires the development of clear principles for its ethical and responsible use. Despite numerous potential benefits, such as personalized learning and task optimization, this technology poses ethical concerns regarding biases, misinformation, and the difficulty of differentiating human-created texts from AI-generated content. To address these concerns, we created the "FAUXperience Framework" to offer guidelines for the ethical use of AI in education by fostering a critical consciousness about this technology's potential benefits and risks. The framework results from user research with stakeholders such as teachers, students, and other interested parties. The study focused on collecting qualitative and quantitative data on how stakeholders experience using generative AI tools in their academic activities and how universities handle the issues related to the misuse of AI. The term "FAUXperience" combines "faux," which refers to artificiality, and "experience," to indicate the use of AI to artificially enhance the educational process in addition to traditional teaching methods. By fostering critical consciousness about AI, the framework aims to promote the benefits of learning associated with AI while drawing stakeholders' attention to its impact on education. In conclusion, the "FAUXperience Framework" encourages teachers and students to be critically conscious actors in AI-powered education.

Keywords: AI, artificial intelligence, FAUXperience Framework, education, critical consciousness, large language models, ChatGPT, critically conscious, ethics.



## Acknowledgements

I want to thank my supervisor, Sara Ljungblad, for her guidance throughout my thesis and my entire journey at Chalmers. Her support has been essential in numerous projects. I also want to thank everyone who contributed to this thesis by participating in interviews or responding to surveys. Their feedback and insights were essential to the development of this thesis project. Lastly, I want to acknowledge my family and friends for their constant encouragement.

*José Bener de Sousa Nunes, Gothenburg, 2024-06-19*



"I believe the cost of getting to know AI – *really getting to know AI* – is at least three sleepless nights. After a few hours of using generative AI systems, [...] it dawns on you that you are interacting with something new, something alien, and that things are about to change. You stay up, equal parts excited and nervous, wondering: What will my job be like? What job will my kids be able to do? Is this thing thinking? [...] You realize the world has changed in fundamental ways and that nobody can really tell you what the future will look like."

*Ethan Mollick. Co-Intelligence: Living and Working with AI.*



# Contents

<b>List of Figures</b>	<b>xiii</b>
<b>List of Tables</b>	<b>xv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Research Questions . . . . .	3
1.2 Stakeholders . . . . .	3
1.3 Expected Results and Impact on Stakeholders . . . . .	4
1.4 Scope and Delimitation . . . . .	5
1.5 Ethical Considerations . . . . .	5
1.6 Use of Generative AI . . . . .	6
<b>2 Background</b>	<b>7</b>
2.1 The Convergence of Society and Technology . . . . .	7
2.2 The Integration of AI Into Education . . . . .	8
2.3 Ethics, Challenges, and Dangers of AI . . . . .	11
<b>3 Theory</b>	<b>13</b>
3.1 The Key Concepts and Terminology of AI . . . . .	13
3.2 The Four Principles of Co-Intelligence With AI . . . . .	15
3.3 Human-Computer Interaction . . . . .	19
3.4 Education for Critical Consciousness . . . . .	23
3.5 Simulated-based Education . . . . .	25
<b>4 Methodology</b>	<b>27</b>
4.1 Design-oriented Research in HCI . . . . .	27
4.2 Research Phase (Discover) . . . . .	27
4.3 Synthesis Phase (Define) . . . . .	30
4.4 Ideation Phase (Develop) . . . . .	32
4.5 Implementation Phase (Deliver) . . . . .	34
4.6 Overview of the Planned Research Methodology . . . . .	35
<b>5 Process and Execution</b>	<b>37</b>
5.1 Research . . . . .	37
5.2 Define . . . . .	45
5.3 Develop . . . . .	51

## Contents

---

5.4	Deliver . . . . .	60
<b>6</b>	<b>Results</b>	<b>61</b>
6.1	Experiences and Perceptions of AI in HE . . . . .	61
6.2	The Limitations of AI Detection Tools . . . . .	74
6.3	The FAUXperience Framework . . . . .	79
<b>7</b>	<b>Discussion</b>	<b>87</b>
7.1	Future Work . . . . .	94
<b>8</b>	<b>Conclusion</b>	<b>95</b>
	<b>Bibliography</b>	<b>97</b>

# List of Figures

2.1	Student Achievement Distribution. . . . .	9
3.1	Overview of LLM training process. . . . .	14
3.2	Visualization of the Jagged Frontier. . . . .	16
3.3	Centaur and Cyborg on the Jagged Frontier. . . . .	17
3.4	Color Scales. . . . .	22
4.1	Overview of the planner research methodology. . . . .	35
5.1	Key Stakeholders in the higher education system. . . . .	39
5.2	Distribution of respondents by university and program. . . . .	44
5.3	Crossplag, Undetectable, and Sapling Scoring Systems. . . . .	55
5.4	Scribr, Plagiarism Detector, GPTZero Scoring Systems. . . . .	56
5.5	Scispace, ZeroGPT, Content Detector Scoring Systems. . . . .	57
5.6	QuillBot AI Scoring System. . . . .	58
6.1	Students' Perceptions of AI Integration and Guidelines in HE. . . . .	63
6.2	AI-Powered Tools Used by Students. . . . .	63
6.3	Purpose of AI Usage. . . . .	64
6.4	Concerns About Implementing AI in Studies. . . . .	64
6.5	Usage of AI-Content Detection Tools. . . . .	65
6.6	Ethical Considerations in Using AI Detectors. . . . .	66
6.7	Satisfaction with Accuracy of AI Detectors. . . . .	66
6.8	Teachers' Perceptions of AI Integration in HE. . . . .	67
6.9	Integration of AI in Teaching Methods and Courses. . . . .	68
6.10	Motivations for Integrating AI into Teaching. . . . .	68
6.11	Concerns Regarding the Use of AI in Education. . . . .	69
6.12	Concerns About AI in Education. . . . .	70
6.13	Impact of AI on Student Engagement. . . . .	70
6.14	Discussions About AI Risks. . . . .	71
6.15	Challenges in Integrating AI in Courses. . . . .	71
6.16	Communication of AI Usage Policies to Students. . . . .	72
6.17	Guidelines for Bachelor's Students. . . . .	83
6.18	Guidelines for Masters' Students. . . . .	84



# List of Tables

2.1	Ranking of Occupations Based on AI Impact Potential. . . . .	10
3.1	Classification of Tasks for AI Interaction. . . . .	16
3.2	Examples of Different Perplexity Levels. . . . .	21
3.3	Essential Principles of Social Constructivism. . . . .	24
3.4	Components of a Role-Play Simulation Prompt. . . . .	26
4.1	The four main types of interviews. . . . .	29
4.2	Types of Conceptual Constructs. . . . .	33
4.3	Types of Evaluation. . . . .	34
5.1	List of Interviewees. . . . .	40
5.2	Distribution of Teacher Survey Participants by University and Country. .	44
5.3	Distribution of Student Survey Participants by University and Country. .	45
5.4	Selected Detection Tools. . . . .	54
6.1	AI Detection Tool Scores on Human-Written Texts. . . . .	76
6.2	Detection Rates for 100% AI-Generated Texts. . . . .	76
6.3	Detection Rates for Texts with Mixed AI and Human Content. . . . .	77
6.4	Comparison with Self-Reported Accuracy. . . . .	78
6.5	Total False Positives and False Negatives. . . . .	78



# 1

## Introduction

*"Good Afternoon,*

*When your paper was uploaded on 10/06/23, it was checked through Turnitin. The program returned a positive response for AI. I also checked your paper through a third-party app utilized by the Criminal Justice Department for verification. This app confirmed the Turnitin AI response. Using AI is cheating and not your work. Therefore, you will receive a grade of zero for your paper. Any further violations will be sent to the Student Academic Integrity Committee.*

*Sincerely,  
Robert Ellison"*

In October 2023, this email was sent to Marley Stevens, a student at the University of North Georgia [1]. Marley claimed that she used only Grammarly to proofread and correct punctuation and grammar errors in her essay and did not use any content generated by AI chatbots. Despite her defense and the evidence she presented, the Student Academic Integrity Committee placed her on academic probation for six months. According to the head of Grammarly for Education [1], her essay was mistakenly flagged as AI-generated due to the University's defective AI identification software.

This case illustrates how products with deficient interaction design can negatively impact users and how generative AI tools might affect higher education (HE). Universities worldwide are struggling to adapt to this new paradigm in education [2][3][4] as a result of students increased use of AI-powered tools and those tools becoming more sophisticated at producing texts that seem written by a human.

In the age of AI-powered tools, the needs and expectations of HE students may differ significantly from those of students in conventional educational systems. Generative AI chatbots, such as ChatGPT, have gained popularity among students due to their ability to generate essays, answer test questions, and even write computer code [5] [6].

The ethical issues surrounding AI use, including privacy concerns, the possibility of AI-generated hallucinations, and the perpetuation of racial and gender biases [7] [8] [9] [10], have been intensified in this new educational reality [11]. A survey conducted in January 2023 reported that more than one-third of college students in the United States were using ChatGPT for academic purposes [12]. The increasing use of AI in education makes it challenging for universities to distinguish between texts written by humans and those produced by machines [5].

## 1. Introduction

---

As remarkable as today's AI tools are, there are benefits and drawbacks to using them for all stakeholders in the educational system. To utilize them responsibly, we must continue instilling a critical-thinking mindset in students. Even if AI can potentially transform education, *failure* and *struggle* are necessary processes that a button click cannot substitute. We are at a pivotal point where AI will transform education, but classrooms will be more necessary than ever [13] [14].

This emphasis on critical consciousness [15] is especially relevant to HE as modern technologies have fundamentally impacted education and challenged our conventional methods of instruction and learning [16]. The existence of remote learning platforms alongside traditional in-person education requires HE institutions to reevaluate their current curricula and rethink their educational philosophies and approaches [17].

Given these circumstances, many companies have launched generative AI tools since 2022 [18] [19]. As a result, other companies have created tools to identify AI-generated content [20]. Companies are already launching tools to bypass AI detection. This never-ending loop results in companies fabricating issues to offer solutions, perpetuating an ongoing cycle of technological dependency [21] [22]. Significant concerns concerning the reliability and functionality of these items are highlighted by this dependence [23].

As AI-powered tools are continuously developed and integrated into education, we must ensure that the design process of these tools is conducted ethically and transparently with the involvement of all stakeholders [24]. Similar to the problems faced by students such as Marley Stevens, faulty detection tools fail to satisfy user needs and cause social tensions between users and digital systems. Preventing this friction is one of the primary objectives of interface design research [20]. Understanding users' needs and context when interacting with a system, service, or product is necessary to build products that satisfy user needs [25].

We must understand students' and teachers' changing needs and perspectives in an increasingly AI-driven educational scenario. Students will inevitably employ AI for academic activities as it becomes more widely available, creating new challenges for teachers and HE institutions [14]. They will use AI as a study partner, collaborator, or assistant and seek explanations for tasks that may seem outdated because of AI capabilities. Students will expect to achieve higher productivity levels and understand how AI will impact their careers [14]. Therefore, educational institutions need to develop strategies to address these emerging challenges.

A critically conscious approach to learning becomes particularly relevant in the present educational landscape as artificial intelligence continues to change how we learn and transfer knowledge. According to Paulo Freire [15], the unique contribution of a teacher to the emergence of a new society would have to be the formation of a critical education that helps to form critical attitudes. Due to historical processes, people emerged with a naive consciousness, making them easy prey to irrationality. Only an education that facilitates the transition from naive to critical transitivity, increasing students' capacity to recognize the issues of their time, could prepare people to handle the emotional impact of societal transitions [26].

**Problem Statement:** Generative AI Chatbots, such as ChatGPT, Gemini, and Claude, have gained popularity for their ability to generate human-like text [6], raising concerns about their impact on traditional educational practices [5]. Since its launch in November 2022, ChatGPT has sparked polarizing opinions, particularly for its use in writing essays, answering test questions, and coding, which disrupts the conventional teacher-student dynamic and introduces uncertainties about academic integrity [5]. With reports indicating that many students use ChatGPT and other AI-powered tools for schoolwork, we need to understand what and how they use them [12]. In this new scenario, universities now face the challenge of distinguishing human-written content from AI-generated text. In response, many companies have introduced AI-detection tools to detect AI-based plagiarism, but concerns about their functionality and reliability grow [23]. As generative AI chatbots rapidly evolve, we must create strategies to ensure their critically conscious use to maintain academic integrity [24].

## 1.1 Research Questions

In this thesis, we aim to explore the integration of AI into educational settings. Our primary focus will be to investigate the experiences of students and teachers using AI-powered tools and identify the challenges universities face in differentiating between human-created and AI-generated content. Additionally, we will propose strategies for the ethical use of AI in education. This research will address these areas by examining the following questions in higher education:

- **RQ1:** How do teachers and students experience and perceive the usage of Generative AI chatbots?
- **RQ2:** What limitations prevent the adoption of AI detection tools?
- **RQ3:** What strategies can be used to promote a critically conscious use of Generative AI chatbots?

## 1.2 Stakeholders

This study focuses on users of AI-powered tools in HE, including students, teachers, and other stakeholders involved in this context.

### Students

Our main stakeholders are students in HE, particularly those studying interaction design, computer science, and software engineering at Chalmers University of Technology and the University of Gothenburg (GU). We chose these fields because they are expected to impact AI in the near future significantly [27]. Since students are often motivated by job prospects [27], Chalmers and GU are expected to prepare students in these fields with the necessary skills to become the next generation of interaction designers, computer scientists, and software engineers. These individuals will be responsible for designing, conceptualizing, critiquing, and programming the next generation of AI tools that will

shape our society. We particularly want to hear from students in these fields about their views, concerns, and needs.

### **Teachers**

The other stakeholders are educators who teach interaction design, computer science, and software engineering at Chalmers University of Technology and GU. These teachers are already affected by the new education paradigm and are navigating the integration of AI into their teaching. It is particularly relevant to hear their perspectives on this integration and how they adapt to this rapidly changing landscape [28]. Teachers from other universities and countries responded to our survey. Still, the main stakeholders remain the teachers employed at Chalmers and GU.

### **AI Engineers**

Given their expertise and direct involvement in shaping AI technology, AI engineers can provide insights into the understanding of the complexities of AI technology and its current implications. Furthermore, it provides valuable insights into how their work impacts users and society [25].

### **Chalmers University of Technology**

Chalmers University and its Computer Science and Engineering department are integral stakeholders in this thesis. Chalmers is responsible for providing the requirements for this thesis and for judging whether it follows the expected academic quality standards. The university has provided an academic examiner and a supervisor who play fundamental roles in completing this process and contributing to scientific research.

### **Author**

This thesis represents the final effort toward earning an M.Sc. in Interaction Design and Technologies. It is the final step toward acquiring the theoretical and practical knowledge necessary to earn this academic title.

## **1.3 Expected Results and Impact on Stakeholders**

This thesis explores the opportunities and challenges of integrating AI in education by analyzing the user experiences of AI-powered tools in HE contexts and aiming to expand the conversation on the interplay of AI, interaction design, and ethics [29]. The research investigates the implications of using AI-powered tools, focusing on usability and interaction patterns in educational settings. It assesses current AI applications, identifies ethical concerns, and proposes improvements prioritizing user well-being, privacy, and equity. The anticipated outcome of this research is the proposal of a comprehensive framework guiding ethical and critical user interactions with AI-powered tools.

The thesis aims to benefit various stakeholders in HE by promoting the ethical and critical use of AI-powered tools:

For *students*, the framework addresses their concerns and needs. It gives them the strategies to assess and ethically use AI in their academic activities. *Teachers* can learn how students use AI tools and the ethical measures they implement to maintain academic integrity. This understanding can help teachers develop strategies to foster critical consciousness among students. *AI engineers* can benefit from understanding how their tools

affect users' well-being, privacy, and equity, enabling them to advocate for responsible AI innovation. Lastly, for *Chalmers University of Technology* and the *University of Gothenburg*, the thesis offers a comprehensive overview of the academic experiences of students and teachers with AI-powered tools. This overview can inspire the university to improve its educational standards, contribute to innovative research, improve student engagement, ensure ethical AI integration, and provide professional development for teachers.

Overall, this thesis aims to promote a more ethical, critical, and responsible integration of AI in HE by aligning the results with the needs and perspectives of these stakeholders.

## 1.4 Scope and Delimitation

This project examines how students and teachers use generative AI chatbots in learning and teaching activities. We are particularly interested in analysing these AI tools' usability, user experience, and interaction design within the HE context. However, we will not cover the impact on other creative-based educational programs adapted to this new context.

Our goal is to evaluate the integration of AI tools in HE, focusing on the ethical dimensions of students' and teachers' reflections on their interactions with these tools. Additionally, we aim to identify the main tools used, how they are used, and what motivates their selection. Based on this research, we will propose improvements in using these AI tools to fulfill user needs better. By analyzing how stakeholders interact with AI-powered tools, we want to determine whether the tools respond well to user's behaviors [17]. Furthermore, we will critically evaluate AI's ability to meet the needs of both students and teachers within an ethical framework.

This study focuses on the accuracy of AI-powered tools than on the level of AI reasoning; given that these tools are still in the early stages of development despite their widespread use, our attention is primarily drawn to their potential to produce reliable and accurate information. A large body of research has focused on developing automated fact-checking systems [30]. However, there are many obstacles to solving the credibility issues with these systems. Therefore, before we focus on their AI argumentation level, we must address the ability of AI-powered tools to provide trustworthy and dependable information, especially in education, which is our goal.

## 1.5 Ethical Considerations

When conducting scientific research, we must adhere to research ethics to maintain scientific findings' credibility and validity. These ethics provide the values, standards, and protocols that govern scientific research [31].

An essential part of research ethics is how researchers interact with participants and the impact of the research on society at large [32]. This master's thesis follows the principles of fairness, diversity of perspectives, and honesty in reporting findings using ethical

research principles. All ethical requirements are carefully followed during data collection to ensure participants' well-being and comply with the General Data Protection Regulation (GDPR) [33].

Students and teachers were actively involved in the research process, and explicit consent procedures were in place. Participants were informed about the project's purpose and scope and their right to withdraw consent for participation and data collection at any time [31]. Additionally, the background section examines some of the ethical aspects of AI in education.

To build trust and encourage open sharing of experiences, we must establish positive and respectful relationships with participants during interviews. We must also take strict measures to protect participants' confidentiality by GDPR [33]. All participant data remains anonymous unless we receive explicit consent for their identification. Furthermore, all collected data is securely stored and will be deleted after the specified retention period stated in the consent form [34].

Surveys are anonymous to ensure participants' autonomy and encourage honest responses. We recorded interviews with participants' consent and assured them in advance that the personal data we collected was only for research purposes. We securely and confidentially store the recorded data and adhere to data protection regulations [34].

### **1.6 Use of Generative AI**

Throughout the writing process, the author used OpenAI's ChatGPT-4, ChatGPT-4o, and Grammarly to refine the writing and receive guidance on academic writing standards, text's readability, formatting references, grammar and spelling checks, and overall coherence. The suggestions provided were based on the author's original content. After receiving the recommendations, the author carefully and critically assessed each and made necessary grammatical adjustments when necessary.

# 2

## Background

Following the rise of the Internet, which triggered the Third Industrial Revolution, the world is undergoing a new revolution led by AI, known as Industry 4.0 [35]. This process is characterized by incorporating human behavior into machines and systems, reshaping the world through computing, and marking a true *Intelligence Revolution* [36]. Economically influential nations and the world's largest corporations are actively working to dominate every aspect of society with substantial investments in AI innovation [35] to gain a competitive advantage over other countries and commercial rivals. This intense competition is causing a profound impact across all areas of society, including healthcare, business, and education [36] [37]. Given these circumstances, this section will explore the factors driving the evolution of education within the AI Era [38][39][40].

### 2.1 The Convergence of Society and Technology

General-purpose technologies (GPT) are technologies whose applications extend to diverse and generic purposes. A concept introduced by Timothy F. Bresnahan and Manuel Trajtenberg in 1992, GPTs are characterized by their widespread use, potential for technical improvement, and ability to enhance overall economic productivity. Despite differing definitions, historical examples of GPTs include the steam engine, electricity, the combustion engine, and computers [36].

Using the criteria set by Bresnahan and Trajtenberg, AI meets all three [36]. Its widespread applications prove its pervasiveness, and its rapid advancements show its potential for improvement. Although AI's impact on productivity is still emerging, several forecasts predict significant productivity increases over the decade ahead [36].

Despite AI meeting the GPT criteria, by defining AI as a **System Technology** and focusing on its systemic nature, we can emphasize its qualitative changes, complex integration into various societal sectors, and multifaceted impact [36] [41]. This systematic view reveals how the social context significantly influences AI's development, including the norms and values of developers and companies responsible for its development. For instance, AI used in hiring, loan approval, or criminal proceedings may perpetuate biases related to gender, ethnicity, or age [41].

Recognizing AI's systemic nature highlights the importance of ethics in its development and use. It empowers policymakers, businesses, civil society, and AI developers to use AI ethically and responsibly, thereby promoting positive outcomes [41].

### 2.2 The Integration of AI Into Education

The development of contemporary educational theories and methodologies has made it possible to enhance instruction-based learning and give students more effective learning experiences [42]. According to some researchers [43] [44], teachers can employ AI-powered tools to provide students with individualized learning opportunities in addition to the conventional passive learning strategy. In content delivery, some experts even speculate that AI may potentially replace instructors [13].

In contrast to these perspectives, Ko [13] argues that learning technologies are not value-neutral by examining the impact of Large Language Models (LLMs) on student performance and refutes the comparison between calculators and AI in producing minimal influence on Education. Ko believes that AI could potentially harm student learning and accelerate the deterioration of public education. Leveraging established knowledge of educational dynamics and learning theories, Ko anticipates ambiguous future repercussions and advocates for a thorough evaluation of the consequences of LLMs in educational settings.

We can gain a more comprehensive understanding of integrating AI into education by taking into account Ko's [13] perspective that technologies are not value-neutral along with the notion that AI can provide students with innovative learning experiences [43] [44]. Considering this technology's possible benefits and drawbacks, we can analyze how it can drastically change the educational system by creating new difficulties for educational institutions in maintaining academic integrity. According to Mollick, we are on the verge of an era in which AI will revolutionize our education by reshaping the learning experience. The only question is whether we can guide this change in a way that fulfills the ideals of expanding opportunities for everyone and nurturing human potential [14].

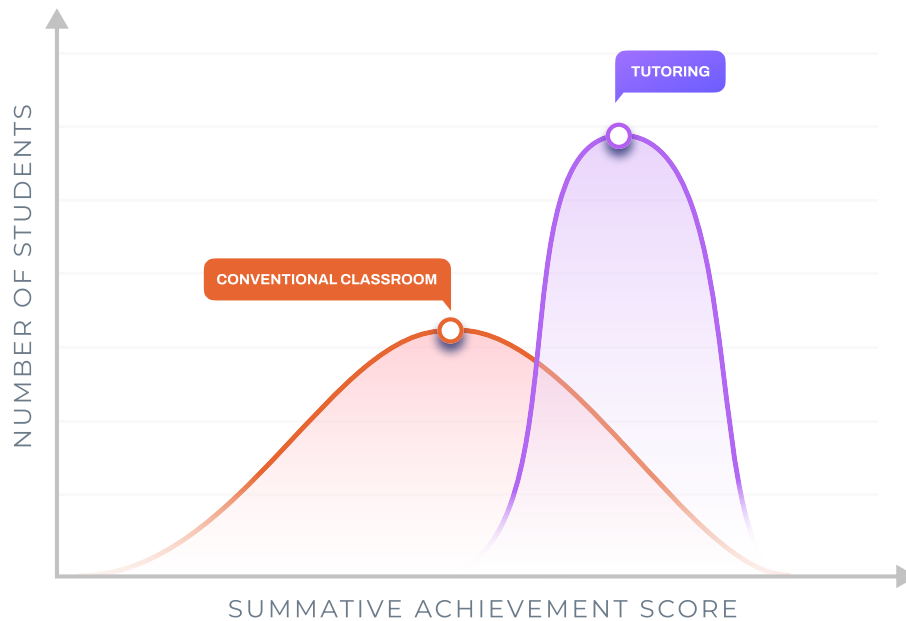
#### AI as a Tutor

For millennia, we have known that the most effective method of learning is individual tutoring, where tutors teach students in a one-on-one format following the students' own learning pace [45]. If a student is struggling with a concept with a concept, the tutor can adjust the pace of instruction until the student understands it. In the same way, the instructor can quickly advance or deepen a topic if the student demonstrates a keen interest in it and mastery of it [45] [46].

However, considering the global population's growth, one-on-one tutoring is not feasible and is prohibitively expensive for mass public education. Due to these reasons, when the public education process began to develop in the 18th century, the educational system divided the students into groups of about thirty or so per classroom and created standardized teaching and assessment procedures, such as collective lectures and regular test assessments [45].

Despite its numerous shortcomings, the modern mass public education system has significantly contributed to increasing literacy rates globally, and its core principles have

remained the same [14]. However, Benjamin Bloom's work in the 1980s demonstrated that one-on-one tutoring significantly outperforms traditional classroom teaching. In his 1984 paper *The 2 Sigma Problem* (2.1), Bloom reported that tutored students scored better than 98% those taught in a classroom setting. This phenomenon, named the *the two sigma problem*, presented an ongoing challenge to achieve similar results with group instruction, given that personalized tutoring is often impractical on a large scale [14].



**Figure 2.1:** "Using the standard deviation (sigma) of the control (conventional) class, it was typically found that the average student under tutoring was about two standard deviations above the average of the control class (the average tutored student was above 98% of the students in the control class" [47]). Created by the author, adapted from [47].

With the advancement of AI, one potential solution is the flipped classroom, the pedagogical model in which online lectures and digital materials are provided to students before class so they can apply them in the classroom through collaborative activities and discussions [48] [14]. However, the success of flipped classrooms is mixed, largely due to the lack of quality resources and teacher time.

AI presents itself as a promising partner, not a replacement, in overcoming these challenges. AI systems can help generate customized active learning experiences by offering personalized instruction tailored to each student's needs and adjusting content based on performance. This allows students to engage more effectively with the material at home, coming to class better prepared for hands-on activities and discussions. Teachers can then focus on meaningful interactions and use AI insights to identify and support students' specific needs [14].

Despite the ongoing challenge of achieving Bloom's two-sigma effect through group instruction, incorporating models such as flipped classrooms shows great potential for improving learning experiences. Traditional schooling will continue offering unique opportunities for collaborative problem-solving and socializing. Classrooms will continue

to provide value, even with AI tutors, but these tutors will transform the field of education, enhancing the traditional learning experience [14].

### AI as a Teacher

Every GPT has impacted jobs throughout history. In the case of AI, studies show almost all professions overlap with AI capabilities. But unlike past automation revolutions targeting repetitive and dangerous tasks, AI now overlaps most with highly compensated, creative, and educated roles [14]. College professors make up 8 of the top 10 jobs overlapping with AI [49].

**Table 2.1:** Ranking of Occupations Based on AI Impact Potential [49].

Rank	Occupation Title
1	Telemarketers
2	English Language and Literature Teachers, Postsecondary
3	Foreign Language and Literature Teachers, Postsecondary
4	History Teachers, Postsecondary
5	Law Teachers, Postsecondary
6	Philosophy and Religion Teachers, Postsecondary
7	Sociology Teachers, Postsecondary
8	Political Science Teachers, Postsecondary
9	Criminal Justice and Law Enforcement Teachers, Postsecondary
10	Sociologists

While AI may overlap with many jobs, it does not necessarily mean these roles will be replaced. Jobs are composed of many tasks, and AI can start performing some of them, particularly repetitive and tedious ones, without eliminating these roles. For instance, teachers could assign administrative tasks to AI while still maintaining their roles [14].

Tasks form the basis of jobs, and a teacher's role involves a range of responsibilities: teaching, researching, writing papers, doing administrative paperwork, writing recommendation letters, applying for grants, and more. The occupation title "teacher" encapsulates many tasks. Considering teachers' tasks, as AI automates administrative tasks, teachers could focus on work that demands human qualities like creativity and critical thinking [14].

AI could even eventually deliver lectures. However, the environment in which a job operates plays an important role in this task division. In the case of teachers, factors like institutional traditions and student acceptance influence the integration of AI into the university. Thus, understanding AI's role requires looking at both task-level impacts and the systemic context within which jobs exist [14].

## AI as a Leveler

A study [50] on the impact of Generative AI on the writing tasks of college-educated workers reported that those using ChatGPT completed their tasks faster and with higher quality, reducing the productivity gap among workers. Individuals with the lowest abilities benefited most in this process, as it helped them improve their performance and leveled the overall productivity distribution. ChatGPT made the work process more efficient by enhancing productivity, reducing worker performance disparities, and turning AI into a performance leveler.

When it comes to productivity's effects on education, AI has the potential to help struggling students improve. For instance, AI can assist students who struggle with writing in creating well-written essays. In activities requiring creativity, AI can help students generate ideas. While it's difficult to predict the exact effects of productivity in any specific context, educational institutions can influence the outcomes of AI integration, whether positive or negative, through strategic AI implementation [14].

By using AI-powered tools in academic settings, educational systems can focus on teaching students to be central participants in the "generative loop" by encouraging them to apply their own expertise to problems instead of just relying on AI for solutions [14]. This approach can help maintain the human element in education, preventing skill-based education from becoming less valuable and avoiding the distortion of our education system into prompt-based learning, where automatically generated prompts are reported to perform better than expert-human prompts [51].

### 2.3 Ethics, Challenges, and Dangers of AI

AI Ethics refers to a set of values, principles, and techniques that apply widely accepted standards of right and wrong to guide moral behavior in developing and utilizing AI technologies. Given AI's significant potential to cause profound ethical consequences, this technology can enhance and disrupt human lives [52].

The line that separates the consequences that enhance or disrupt human lives is conditioned by the biggest risk presented by AI, which is that there is no particular reason that AI should share or accept our standards of right and wrong to guide moral behavior or our view of ethics and morality. At this stage, AI is guided by the values imbued by the companies leading their commercial development [14].

The reality is that the organizations working on AI tools, even when they try to create less biased, more accurate, and more helpful technologies, end up influencing AI and can introduce new types of bias. Given that North American enterprises are behind the biggest AI models in the market [53], many AI systems appear to align with a liberal, Western, pro-capitalist worldview as the AI is designed to avoid making statements that could cause controversy for its creators, generally liberal, Western capitalists [14].

### **Guidelines: The Genie is Out of the Bottle**

The approach to decision-making in different educational systems varies, with some being centralized and others localized. However, as AI-powered tools become increasingly complex, it also becomes challenging for teachers to make well-informed decisions about students using AI-powered tools. As these tools become more commonly used, data privacy, security, bias, transparency, and accountability concerns will continue to grow. Maintaining the benefits of AI tools will become more difficult in the face of this evolving and complex educational system [54].

In this context, guidelines and official recommendations play an important role in harnessing the potential of AI while mitigating its risks. These guidelines can serve as a compass, directing all educational stakeholders toward safe, fair, and effective AI utilization. They are valuable resources that enhance the understanding of AI and protect those most affected by these new technologies [54].

While developing such guidelines, institutions will establish criteria to determine which assignments require AI assistance and which do not. School assignments must be revised; in-school writing assignments, non-internet-enabled computers, and written exams might be options to ensure students learn basic writing skills [14].

When we consider how AI-powered tools affect education, we can compare it to when calculators became popular in North American schools in the mid-1970s. Many teachers wanted to use calculators in their classrooms because they saw they could make students more interested in learning. However, even though calculators became popular in the mid-1970s, they only became part of the North American school curriculum in the mid-1990s.

AI won't replace the need to learn how to write and think critically [14]. However, because AI technology is advancing so fast, the policymakers responsible for education should take less time than those who decided to include calculators in the school curriculum to prevent risks and dangers. After all, "the genie is out of the bottle"[14].

# 3

## Theory

This chapter provides an overview of the project's theoretical foundation, emphasizing relevant theories grouped around three primary topics: understanding artificial intelligence, the intersections between AI and interaction design, and modern perspectives on education from a critical consciousness perspective. The key concepts and terminology of AI theory are examined in the section on AI Concepts. Investigating the intersections between AI and interaction design addresses the mechanics underlying AI-generated text recognition, covering the fundamental ideas of large language models. Lastly, Paulo Freire and other educational theorists are used to analyze the social dynamics of learning and recognize learner diversity as a strength.

### 3.1 The Key Concepts and Terminology of AI

**Transformer Architecture.** In 2017, Google researchers published a paper that introduced a novel neural network architecture for language understanding called the Transformer [55] [14]. This architecture significantly influenced the AI community and served as the basis for most large language models (LLMs) today [14]. For comparison, early text generators did not rely on contextual comprehension but used simple rule-based word selection [14]. For comparison, early text generators did not rely on contextual comprehension but used simple rule-based word selection. The Transformer architecture solves this issue by employing an "attention mechanism," which allows the AI model to assess the importance of different words or phrases in a text block [56].

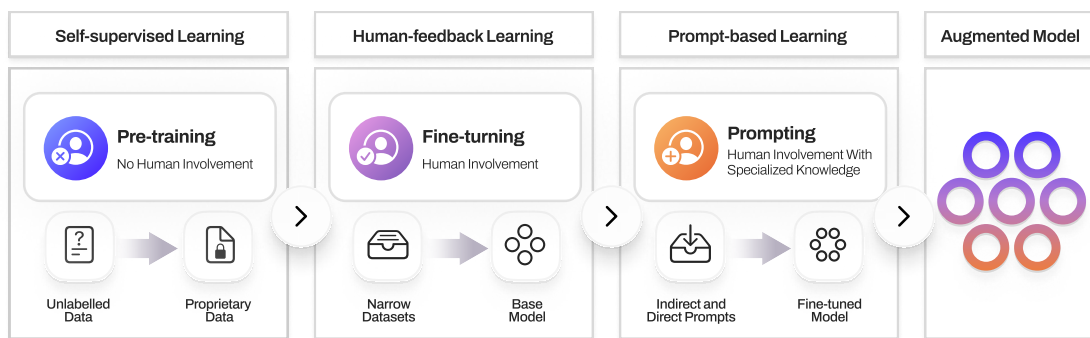
**Large Language Models and Tokenization.** Building on the Transformer architecture, new forms of AI, called Large Language Models (LLMs), analyze a text and predict the next word or part of a word [56] [14]. This process is known as Tokenization, where a text is divided into smaller units, referred to as tokens, before being fed into the model. For instance, the Tokenization of hypertension produces the following: "hy," "per," and "tension" [56]. ChatGPT is essentially a very sophisticated autocomplete [14]. Once you give it some preliminary text, it keeps writing text based on what statistical analysis predicts will probably be the next token in the sequence [14].

**Pre-Training.** The process of prompting LLMs to generate responses happens in several stages [57] [56]. The initial phase, known as Pre-training, consists of feeding the model a wide variety of data from various online sources such as websites, books, and digital documents to learn how to comprehend and generate human-like texts. Humans do not

supervise this process; the AI independently analyzes textual sources to identify patterns, structures, and content in human language [57] [56].

**Weights.** LLMs can create a model that emulates human-written texts by using numerical values, known as weights. These weights tell the AI how likely it is for different words or word components to appear together or in a specific order. For instance, in ChatGPT's first version, 175 billion weights were linked to produce human-like text [14].

**The LLM Training Process.** The training process starts with a large number of weights without any helpful information about word relationships. After that, the model is trained using a large volume of textual input [56]. Throughout the training process, the model adjusts its weights based on the input data, learning which word combinations tend to occur and which do not. It then uses its current weights to predict words in a sequence. Each iteration compares the produced text with the original text, looking for discrepancies [57]. The model then modifies the weights to enhance its predictions, continuously refining word connections to generate contextually appropriate text [14].



**Figure 3.1:** LLMs "learn" at every training phase using increasingly focused inputs. The LLM training process begins with pre-training, where the model learns from unlabeled and proprietary data without human supervision. Later, more specific datasets and human input are added during fine-tuning. Following this, individuals with specialized knowledge use prompting techniques to modify the LLM into an augmented model to perform specialized tasks. Created by the author, adapted from [56].

**Fine-tuning.** Public domain books, research articles, and other free online resources are typical examples of LLM training datasets. Due to the variety of the datasets, the AI also learns the biases, errors, and falsehoods in the text samples during pre-training [14]. At this stage, no guardrails are in place to prevent AI from disseminating harmful content. Therefore, LLMs undergo an improvement process called fine-tuning to mitigate this issue [57]. This process involves further training a pre-trained model on specific datasets, such as medical records, for a healthcare application or customer service logs for a contact center support system [57] [56]. Additionally, this fine-tuning can be enhanced further by humans scoring the quality of multiple model outputs, a process known as reinforcement learning from human feedback (RLHF), which employs human workers evaluating AI responses according to different standards, such as removing violent or sexual content from the results or screening them for correctness [14].

## 3.2 The Four Principles of Co-Intelligence With AI

As artificial intelligence becomes more widespread, we must establish comprehensive guidelines for interacting with this technology. These guidelines should be based on general principles rather than tailored to specific versions of the technology, especially considering the rapid advancements in LLMs [14]. When we apply these principles, we learn about AI's strengths and understand when not to utilize them as they function as guardrails for a future in which human and artificial intelligence are integrated [14].

### Principle One: Always Invite AI To The Table

The initial principle is based on an experimental task-based approach. Since AI is a general-purpose technology, there isn't one single manual explaining its benefits and limitations [14]. This lack of clarity makes it difficult to determine which tasks are best performed by these tools and which are not. By embracing this experimental approach, users can intimately identify AI tools' nuances, limitations, and abilities to assist in their specific tasks and scout out their weaknesses [14]. "As we grow more familiar with LLMs, we can harness their strengths more effectively and preemptively recognize potential threats [...], equipping ourselves for a future that demands the seamless integration of human and artificial intelligence" [14].

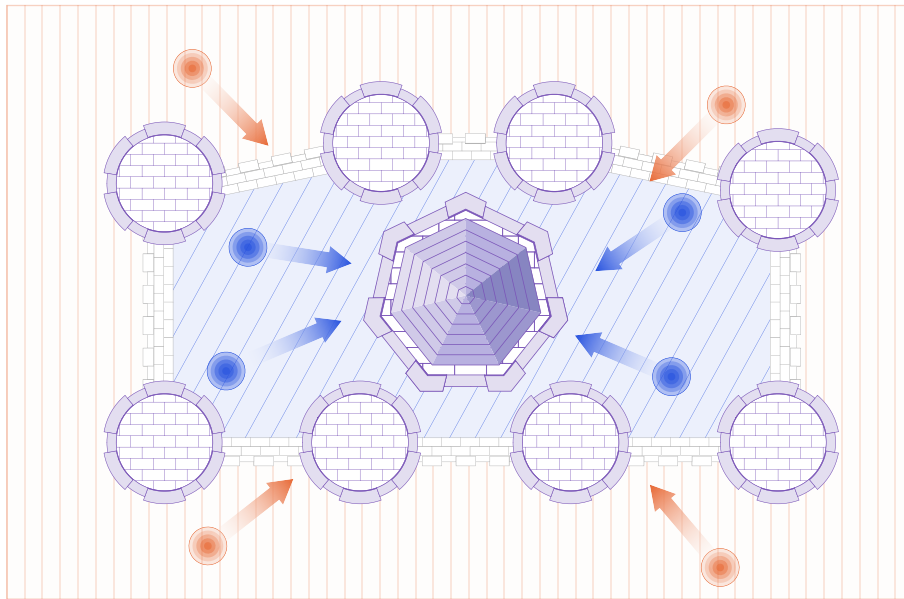
Mollick suggests that to understand AI's impact, we need to comprehend how human interaction with AI changes and the capabilities and limitations of AI in performing tasks [14]. To explain this task-oriented perspective, Dell'Acqua et al. [58] introduced the metaphor of AI as a Jagged Frontier .

### The Jagged Frontier

In the Jagged Frontier metaphor, AI is represented by a fortress wall, with tasks inside being easy to handle and those outside being more challenging [3.2]. However, the wall is invisible, making it difficult to determine which tasks fall into each category. Basic math might be outside the wall and challenging for AI, while idea generation might be inside and easy for AI [14].

We must experiment with different tasks, dedicate time, and gain experience to understand AI's capabilities. Because this experimentation is what gives shape and contours to the Jagged Frontier [14]. For instance, in educational contexts, students should apply AI in various academic tasks to understand and map out which tasks are best suited for them and which ones they can outsource to machines. Once they know the wall's shape and contours, they can make informed decisions about utilizing AI, benefiting from its strengths, and addressing its limitations.

At the task level, we must identify the tasks humans should carry out exclusively. Because AI, while not suitable for tasks that involve ethical considerations, has suitable capabilities. By understanding the role of AI at the task level, we can assign tasks based on the necessary level of *humanness* required by them.



**Figure 3.2:** The fortress wall represents AI’s abilities. Tasks closer to the center are easier for AI to handle, while those farther away are more challenging. However, the wall is invisible, so some tasks that seem equally difficult may be on opposite sides of the wall. Created by the author, based on [58] [14].

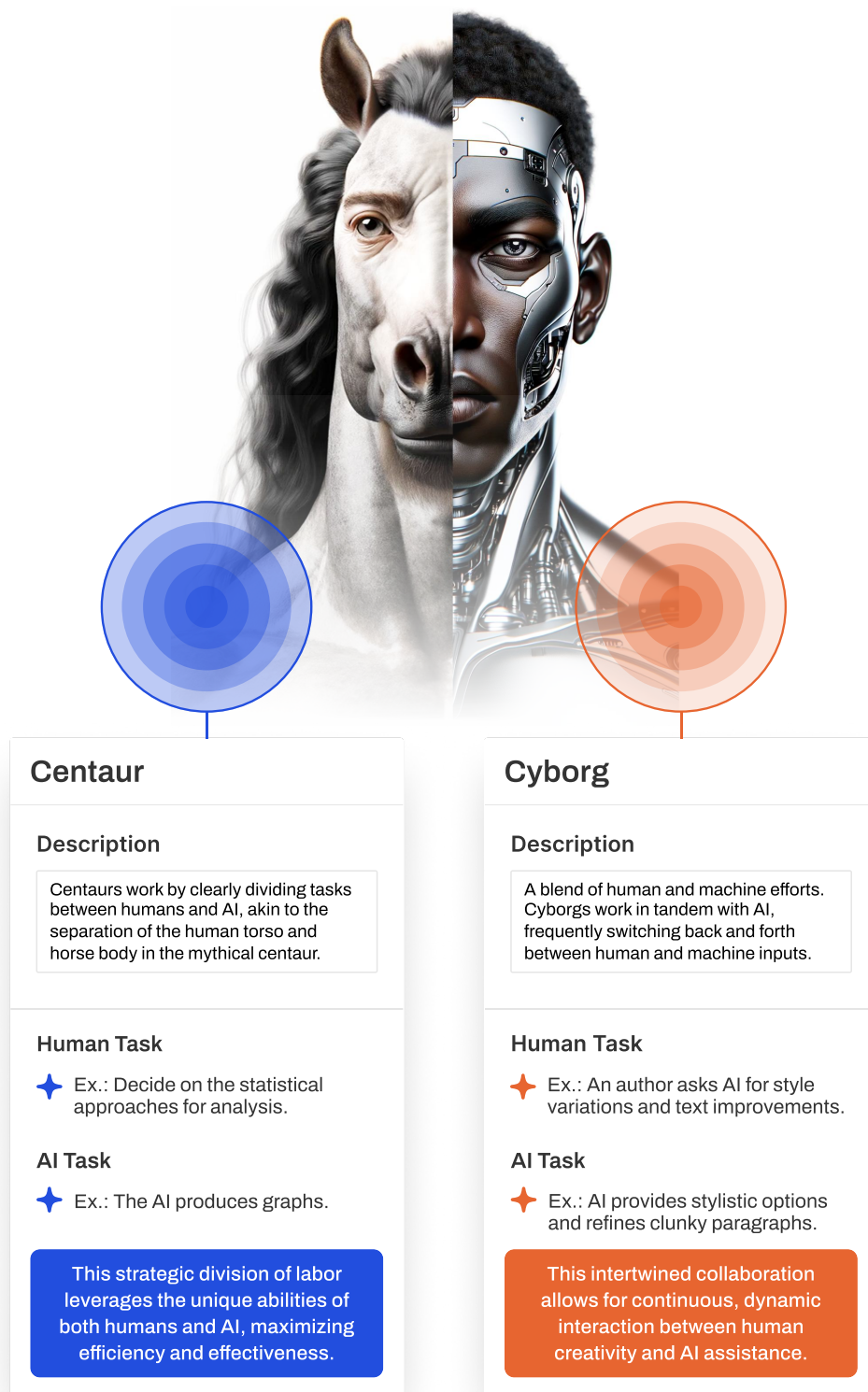
## Centaurs and Cyborgs

The Centaur and Cyborg concepts aim to combine the strengths of human intelligence and machine capabilities by demonstrating various ways to integrate AI into the workplace [14]. If we apply this abstraction to education, centaur tasks can involve summarizing technical papers, where the AI’s ability to summarize complements the author’s deeper understanding of a topic. Cyborg tasks may include using AI for writing assistance and guiding the author through challenging writing sections by providing valuable insights when simulating human feedback. This combination of human intelligence and AI support illustrates how Centaur and Cyborg [3.3]<sup>1</sup> approaches can improve productivity and creativity in a mutually beneficial way [14].

**Table 3.1:** Classification of Tasks for AI Interaction [14].

Type of Task	Description	Example
Just Me Tasks	Tasks where AI is not useful or should remain human for personal or ethical reasons.	Expressing human values.
Delegated Tasks	Tasks assigned to AI, checked by humans; often tedious or repetitive, saving human time.	Scheduling appointments.
Automated Tasks	Tasks are left entirely to AI without human oversight; they are reliable and scalable by AI.	Spam filtering.

<sup>1</sup>Image produced using generative AI. *DALL-E 3 Prompt: High fidelity, head-shot of a half-centaur, half-black cyborg on a white background, front view.*



**Figure 3.3:** The Centaur and Cyborg are two approaches to co-intelligence that integrate the work of person and machine. Centaur work involves a clear separation of work between humans and machines and a strategic distribution of labor according to each entity's strengths. In contrast, cyborgs deeply integrate human and machine elements while integrating human efforts with AI. Created by the author, based on [58] [14].

## **Principle Two: Be The Human In The Loop**

The "human in the loop" principle refers to incorporating human judgment when operating automated systems. Overall, this principle fosters a sense of responsibility and accountability. We maintain human control over the technology, ensuring that AI-driven solutions align with human values, ethical standards, and social norms [14]. While AI training methods now involve human judgment, the increasing automation and delegation of tasks to AI requires a focus on maintaining human values in the decision-making process. LLMs, as text prediction machines, cannot differentiate between true and false information, so while they may excel at generating plausible answers, they often contain subtle errors, which can pose potential risks [14] [59].

Over time, as we learn more about AI-powered tools and comprehend their complex functioning, our confidence in their ability to perform without explicit guidance decreases. However, by recognizing their limitations, these tools can become effective assisting resources in a variety of disciplines [59]. Therefore, when interacting with generative chatbots, it's helpful to think of AI as an entity trying to optimize several functions. One of the most important functions it's programmed to optimize is to satisfy the user by providing an answer the user will like [14] [60]. This optimization may lead AI to prioritize user satisfaction over information accuracy. For instance, current generative chatbots may make up something if they haven't "learned" it because prioritizing "making the user happy" takes precedence over "being accurate" [14]. The generation of these incorrect answers is known as "*hallucination*." While newer, larger LLMs exhibit significantly fewer cases of hallucination compared to older models, hallucination is still a weakness of this technology [60] [14].

## **Principle Three: Treat AI Like A Person**

Anthropomorphism is the tendency to attribute human-like traits to non-human entities. Both behavioral psychology and evolutionary biology have extensively studied the causes and implications of this tendency, demonstrating its significance in human history [61]. Therefore, it is no surprise that we are inclined to anthropomorphize artificial intelligence, mainly due to the widespread use of chat-based interactions. We often describe these intricate algorithms and computations as "understanding," "learning," and even "feeling," which fosters a sense of familiarity and relatability but also causes potential confusion [14], which can be exploited by malicious individuals who can manipulate users by gaining their trust in the system [61].

Therefore, treating AI like a person and telling it what kind of person it is refers to creating a distinct and clear AI persona as a strategy to improve Human-AI interaction. This process involves defining the AI's identity, its role in addressing specific issues, and the context in which it operates [14]. Many AI models produce generic outputs by default, but by providing context and constraints, we can tailor the tone and direction to serve a specific purpose. We can use AI as a cooperative co-intelligence with this collaborative editing process and continuous guidance [14].

---

## **Principle Four: Assume This Is the Worst AI You Will Ever Use**

Given the speed at which generative AI is developing, very soon, tasks that we thought were intrinsically human will be able to be done by AI. Even if LLMs are software engineering products, we should view AI's limits as temporary. Traditional software is dependable, predictable, and, when developed correctly, produces consistent results [14]. On the other hand, AI is unpredictable, unreliable, and capable of hallucinations. AI doesn't act like software, but it does act like a human being. This mindset, which aligns with the "treat it like a person" principle of AI, can significantly improve our comprehension of how and when to use AI in a practical, if not technical, sense [14].

### **3.3 Human-Computer Interaction**

The field of Human-Computer Interaction (HCI) investigates the complex interplay between interactive systems and humans. Through its study, design, and evaluation, HCI seeks to enhance these systems' usability, effectiveness, and overall user experience [62]. Interaction occurs when a person uses computing technology to perform a task, utilizing their senses and responses to monitor and control devices, machines, or systems enabled by computing technology [62].

By drawing on the knowledge and techniques of diverse scientific disciplines such as computer science, psychology, and ergonomics, HCI continually adapts its research methodologies to cater to users' changing needs and capabilities. As a result, many methodological advancements have emerged to address evolving trends in the field [62].

#### **Interaction Modalities**

Human-computer interactions rely on input and output devices. Input devices enable users to transmit data and instructions to a computer (e.g., keyboard, mouse, microphone), while output devices enable the computer to communicate information back to the user (e.g., monitor, speaker, headphones) [63]. Interaction modalities refer to the exchange of information between humans and computers. This exchange involves input and output devices, information channels, and sensory modalities [63]. To make human-computer interaction more natural, we can combine many modalities, such as drawing, writing, speaking, and gesturing. The human senses, which include hearing, smell, sight, touch, and taste, are all related to sensory modalities essential for interaction. These sensory modalities are connected to distinct information channels; for example, sight is mainly associated with the visual channel, whereas hearing is associated with the auditory channel [63].

#### **Conversational Interfaces**

Under this premise, chatbots like ChatGPT, Gemini, and Claude are designed as conversational interfaces to emulate human speech. Through their interaction modalities, these interfaces enable machines to give the impression of being human to other humans [64]. Different modalities can affect how additional or different information can be conveyed and make interactions more natural and immersive [65].

Conversational interfaces can make sophisticated human-computer interactions easier by allowing users to express themselves naturally and directly, such as by typing or speaking [65]. Natural Language Processing (NLP) is one technology that facilitates this level of personalization. NLP allows users to engage using their language and preferred style rather than being restricted to a limited set of pre-defined interaction methods [65].

## **Natural Language Processing**

NLP is a popular artificial intelligence application that automates the reading, analyzing, and generating of human language [36]. From an interaction design standpoint, NLP enables machines to emulate coherent conversations with users through natural language-based prompts. The primary goal of the NLP field is to develop algorithms that can comprehend human language and perform interpreting tasks. These algorithms can distinguish between letters and words, label text elements, and analyze text direction to infer meaning [36].

Deep learning has emerged as a popular approach to designing AI-powered interactions that can understand human language, with Chatbots being a prominent example. Chatbots are automated chat systems that interact with users through prompts, analyze questions, and select appropriate responses or follow-up questions using decision trees [8].

Chatbots are extensively valuable for education, aiding students in language translation, answering questions, and explaining topics to facilitate learning. However, concerns have been raised regarding the future of education due to potential threats posed by AI, including issues with assignment integrity and online exams, dependence on generative AI tools, challenges in evaluating ChatGPT-generated content, and potential impacts on critical thinking and problem-solving skills [8].

## **The Mechanisms Behind AI-Generated Text Detection**

For over two decades, universities have used text-matching software to detect plagiarism. The recent development of generative AI tools has motivated the creation of technical solutions to distinguish between human-written and AI-generated texts [66]. With the popularization of generative AI chatbots and the dissemination of AI-generated texts in many areas, the software industry has responded by introducing dozens of tools for AI-generated text detection. However, we must understand the mechanisms behind AI-generated text detection to understand whether these tools can distinguish between human-written and machine-generated content [66].

Various techniques for analyzing a text, such as readability, linguistic analysis, frequency counting, perplexity-based filtering, diversity and vocabulary richness, and much more, are used to determine whether the text is human-written or AI-generated [67].

The detection tools apply these features to discover patterns and information regarding the text that are invisible to the human eye. In general terms, LLMs are systems trained to predict the likelihood of a specific character, word, or string (called a token) in a particular context [66]. The AI detector similarly tries to indicate if the source material results from predicting a specific character, word, or string in the same context. If the

answer is yes, it flags the text as probably AI-generated. This detection process examines two key metrics: perplexity and burstiness. The lower the values of these variables, the greater the likelihood that AI has generated the text.

**Perplexity.** Perplexity is a metric that reveals the degree of unpredictability in a text. Lower perplexity values indicate that the text was likely generated by AI language models, which tend to make the text more coherent and easier to read but more predictable. For instance, human-written texts tend to have higher perplexity due to more creative language choices and typos [68] [69]. Language models function by anticipating the next word that would fit naturally in a sentence. For instance, in the sentence *"I couldn't fall asleep last night..."*, there are many possible continuations, as outlined in the table below [68].

**Table 3.2:** Examples of Different Perplexity Levels. [68]

Example Continuation	Perplexity
I couldnt get to sleep last <b>night</b> .	<b>Low:</b> Probably the most likely continuation.
I couldnt get to sleep <b>last time I drank coffee in the evening</b> .	<b>Low to medium:</b> Less likely, but it makes grammatical and logical sense.
I couldnt get to sleep last <b>summer on many nights because of how hot it was at that time</b> .	<b>Medium:</b> The sentence is coherent but quite unusually structured and long-winded.
I couldnt get to sleep last <b>pleased to meet you</b> .	<b>High:</b> Grammatically incorrect and illogical.

**Burstiness.** Burstiness refers to the patterns observed in word choice and vocabulary size. This metric indicates that AI-generated text displays a higher occurrence of clusters of similar words or phrases within shorter text segments. In contrast, human-written texts demonstrate a wider variety in their word selections, showing a more extensive range of vocabulary [70].

## Data Visualization and Color Scales

AI-detecting tools employ various data visualizations to illustrate human-generated and AI-generated content levels. "There are three fundamental use cases for color in data visualizations" [71]: To distinguish data groups, represent data values, and highlight key information [71]. An extensive theory underlies the use of color to distinguish data, and we will provide an overview of the main color scales commonly used to visualize different types of data [71]. This will enhance the clarity of the analysis of AI detection tools in the Process and Execution chapter.

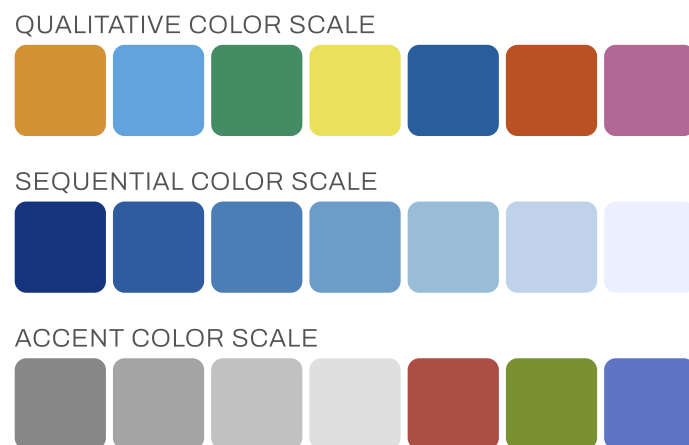
Color has three primary use cases in data visualizations: distinguishing groups of data from each other, representing data values, and highlighting. The colors and how we use them vary in these scenarios [71].

**Qualitative Color Scale.** We use a qualitative color scale to distinguish items or groups on a map with no inherent order, such as distinct countries. In this scale, we select colors

that look different from each other and are visually equivalent. The colors should not give the idea of order, as in a color gradient. Such colors would create an apparent order among the colored elements, which, by definition, have no order [71].

**Sequential Color Scale.** We can use a sequential color scale to represent quantitative data values such as income, temperature, and speed [71]. In this scale, we use a color sequence to represent which values are larger or smaller than others and the distance between two specific values [71]. We should be able to perceive this color scale to vary uniformly across its entire range; to accomplish this uniformity, "we can use a single hue (e.g., from dark blue to light blue) or multiple hues (e.g., from dark red to light yellow) arranged in color gradients" [71].

**Accent Color Scale.** Color can also help highlight specific parts of the data. When we emphasize categories or values in a dataset, we can help the reader understand the information by emphasizing the important figure elements. One technique to get this effect is to color the element we wish to highlight in a hue or set of colors that contrast with the rest of the figure. In this type of color scale, we can find both a set of subdued colors and a matching set of stronger, darker, and more saturated colors [71].



**Figure 3.4:** An example of a qualitative color scale at the top includes a set of distinct colors. In the middle, we have an example of a sequential color scale, which is a monochromatic scale that progresses from dark to light blue. Lastly, at the bottom, we have an example of an accent color scale with four gray base colors paired with three accent colors. Created by the author, based on [71].

### Information Visualization for Color-Vision Deficiency

When it comes to sequential color scales that represent data values [71], we need to consider three critical conditions: the colors need to clearly indicate which data values are larger or smaller than others, the differences between colors need to visualize the corresponding differences between data values, and we need to design with color-vision accessibility in mind [71].

When designing and selecting colors, we must always be mindful of the possibility that a good proportion of readers may have some form of color-vision deficiency (i.e., are

colorblind), which means that some readers might have difficulty distinguishing certain types of colors, such as red and green (red-green color-vision deficiency) or blue and green (blue-yellow color-vision deficiency) [71]. "The technical terms for these deficiencies are deuteranomaly/deutanopia and protanomaly/protanopia for the red-green variant (where people have difficulty perceiving either green or red, respectively) and tritanomaly/tritanopia for the blue-yellow variant (where people have difficulty perceiving blue)" [71].

### 3.4 Education for Critical Consciousness

The concept of critical consciousness (CC), or *conscientização* in Portuguese [72], was an approach developed by Brazilian philosopher Paulo Freire, to teach literacy skills to peasants in rural Brazil and a tool to help them gain a critical consciousness of their social reality. This concept emphasizes the importance of learning to read the written word and the world around us. By doing so, marginalized groups can recognize and analyze systems of inequality and commit to take action against these systems [73] [74] [75]

Through practical initiatives in Brazil, Freire reported how marginalized communities, as they actively engaged in critical consciousness, developed a more nuanced and complex understanding of societal structures. This transformative process, characterized by a dynamic interplay between reflection and action, is the essence of the theory of critical consciousness [75] [15].

Critical consciousness implies "learning to perceive social, political, and economic contradictions, and to take action against the oppressive elements of reality" [76]. For example, in the context of higher education, students with limited economic resources, students with higher levels of critical consciousness are more likely to recognize the inequality in access to educational resources. This recognition enables them to take action, such as joining a student group, attending a school board meeting to address the issue, or using generative AI chatbots to create their own private tutor. On the other hand, students with lower levels of critical consciousness may fail to recognize such inequalities, feel powerless to address them, or avoid acknowledging the problem [75].

In this project, we explore the practical application of critical consciousness in fostering students' awareness of AI technology. This framework helps students identify how their conscious use of technology can enhance reflection, motivation, and agency and encourages them to take action. It also aims to foster a shared sense of values among students, prioritizing the human aspect of using AI.

### The Social Dynamics of Learning

Learning theories have evolved to prioritize the social dimension of learning alongside individual learning. While individual learning was previously the main focus, contemporary learning theorists have emphasized the importance of collaborative work and discussion in the learning process. There is now a recognition that independence and social interaction are both essential for effective learning [42][77].

### 3. Theory

---

Among the learning theories that investigate learning as a social phenomenon, social constructivism, sociocultural, and activity theories can be traced back to psychology, particularly to Vygotsky's work on the influence of the social world on an individual's development [42]. To be didactic, we can summarize the main principles of these theories into four key points: observation and application lead to effective learning; learning occurs within communities; quality assessment is collaborative; and transferable knowledge is context-based [42].

**Table 3.3:** Essential Principles of Social Constructivism [42].

Principle	Description
Observation and application lead to effective learning	A crucial element in learning is observing individuals tackling real-world issues, as learning is a social phenomenon within communities.
Learning occurs within communities	Knowledge and learning are present in interactions among individuals and their surroundings within communities of practice.
Quality assessment is collaborative	Assessing quality involves adhering to group standards rather than individual criteria; unlike traditional educational approaches prioritizing "assembly-line learning" with fixed roles, which lack meaningful context and rely on extrinsic rewards to motivate learners, contemporary pedagogical models, such as the flipped classroom model [78], promote collaborative participation, enabling learners to interact with instructional content at their own pace, but also encouraging practical, creative, and active learning activities.
Context shapes transferable knowledge	Context heavily influences learning, determining what and when we learn and impacting our ability to apply knowledge to new situations.

These principles urge us to investigate the impact of AI on education. Despite some positive influence of digital technology on education [79], including efficiency in lesson planning and provision of immediate feedback, negative views persist in this new paradigm. For instance, students' mastery of essential skills has declined, teacher-student relationships have been distorted, and students have become increasingly isolated in a virtual world [28]. Therefore, exploring how AI can avoid further exacerbating these consequences is vital.

In light of these principles, exploring the potential impact of AI integration on the social aspects of learning and students' educational growth is essential. Additionally, we must consider how AI may impact the relationship between humans and machines, particularly regarding learning communities and feedback mechanisms. These inquiries demand thorough examination to fully uncover their intricacies, which is impossible in the boundaries established by this study's scope.

Nonetheless, they will inform the research design within the scope of our project. Ultimately, the principles of social constructivism highlight the impactful changes this technological scenario has brought to learning communities, leading us to inquire how AI will further shape human-machine dynamics.

## **Recognizing Learner Diversity as Strengths**

Education has historically been a privilege reserved for socially and economically stable families, perpetuating their dominance and homogenizing student backgrounds and experiences [80]. Additionally, there was a widely held belief in a singular, correct approach to teaching and learning, with little consideration for diversity and inclusion due to the generational uniformity among students [42].

However, as access to education has gradually become more accessible, education has evolved into a more multicultural environment. This change has sparked a reassessment of how student differences are perceived, challenging the deficit model prevalent in the classical education system that treated deviations from the norm as deficits [42].

Embracing a constructivist approach to education has led to a paradigm shift, fostering the idea that students construct their own meanings throughout the learning journey [42]. This idea was later expanded upon by Cross-cultural studies, which propose a shift in how education perceives diverse backgrounds and prior experiences [81][82]. Rather than viewing them as barriers, they are valuable assets enriching the learning experience [42]. In a democratic educational setting, students come from various backgrounds and possess diverse intelligence, interests, ethnicity, race, culture, and gender [42].

As we explore the complexities of student diversity, questions arise about how AI, a technology designed to emulate human knowledge, can address differences in intelligence, interests, ethnicity, and race [42]. We must also consider how education can adapt when data collected may not reflect the realities of a racially unaware society. With so many variables at play, it can be difficult for teachers to meet each student's unique needs. We must investigate how we can leverage AI to support educators in adapting teaching to diverse student populations. However, AI's unregulated nature in many contexts raises concerns about its potential to exacerbate or mitigate existing educational inequalities.

## **3.5 Simulated-based Education**

Simulation-based education enables students to practice their skills by simulating real-world scenarios in an immersive environment [83]. Immersion occurs when students fully engage in a task or setting as if it were real [84]. Simulation is a technique that can be used in various disciplines to replace and improve real-world experiences with guided, often immersive, activities that replicate significant aspects of the real world [83] [84].

In AI-based simulations, AI mediates the user and the simulated scenario. One of the primary benefits of these simulations is their adaptability to specific learning goals, which improves personalized learning experiences [84]. Teachers might assign simulation exercises to students, such as role-playing [84], which allows students to be more willing to take risks they wouldn't usually take and step into unfamiliar roles [84]. It also allows

students to learn a topic through a narrative, where they can test their knowledge and practice making critical-thinking decisions without the risks of real-life contexts.

To create efficient AI-based simulations, Mollick [84] developed prompting instructions to ensure engagement between educational stakeholders and AI. By following these components, teachers and students can create prompts that improve learning experiences by providing well-defined scenarios and clear objectives. Originally, Mollick categorized simulation-based education into two types: role-play simulation and goal-play simulation [84]. In role-play simulations, students take on roles that are different from their real-life identities, while in goal-playing, students maintain their own identity and use their knowledge and skills to guide others, such as simulated characters [84]. For the purposes of this project and its limited scope, we have chosen to use a combination of both simulation types, which we refer to as role-play simulation for simplicity's sake.

### AI Role-Play Simulation

In AI Role-Play simulations, students can assume roles that differ from their real-life identities [84]. These simulations encourage students to step outside their comfort zones and experiment with different perspectives. Stepping into a role also allows students to experiment with other ways of engaging with a topic and new ways of solving problems [84]. This immersive experience allows them to gain valuable insight into their strengths and places for improvement in a particular subject.

**Table 3.4:** Components of a Role-Play Simulation Prompt. [84]

Component	Description
A Dual Role for the AI	The AI plays the AI Mediator, creating the scenario and giving students directions. The AI also plays a character within the scenario.
Scenario Choices	The AI Mentor offers students a choice among scenarios (e.g., persona's personality) that pique their interest.
Narrative Set-Up	The AI sets the stage for the scenario, avoiding excessive complexity.
Scenario Initiation	The AI clearly marks the beginning of the simulation by signaling to students that they are now in a scenario.
Guidance on Goal and Techniques	The AI Mentor may step into the scenario to remind students of their goals or give hints but does not interfere during the scenario.
End of Scenario and Advice	The AI Mentor steps back onto the scene to reinforce key elements of the topic and identify additional considerations.

Given the importance of raising awareness among educational stakeholders about AI-related resources, students must be mindful of the information they disclose and the tasks teachers delegate to AI [52] [85] [86]. Therefore, by cultivating a critically conscious education as a guiding principle, all educational stakeholders must acknowledge that delegating tasks to AI will have unforeseen repercussions on the academic area [5]. So, when incorporating AI-based simulations into education, it is crucial to carefully evaluate the advantages and potential risks of incorporating AI into education [13].

# 4

## Methodology

This section outlines how the methods used in this study follow the Double Diamond Design Process model, developed by the British Design Council in 2005 [25][87][88]. This methodology focuses on the goals and behaviors of users by Using a user-centered approach and investigating the usage and target domain. Design requirements are defined and translated into a high-level interaction framework, which connects research and design [25]. Therefore, this section details the activities and methods (see Figure 4.1) employed in each project phase to evaluate how students and educators perceive and utilize AI-powered tools in the context of higher education.

### 4.1 Design-oriented Research in HCI

Design-oriented research is a methodology that focuses on understanding user behavior and experience rather than focusing on a design artifact [89]. Through this type of research we can explore new knowledge within a specific field by examining how research artifacts are used in real-world contexts or during the product development process. Which means that the main contribution of this methodology is the knowledge that results from the project [89]. With the knowledge gained from the research process, researchers and designers can improve the design process itself or facilitate the creation of improved products. The *research project* serves as the design effort's primary stakeholder or client in this methodology [89].

### 4.2 Research Phase (Discover)

During the discovery phase, the researcher gains a comprehensive understanding of the problem space. The researcher typically achieves this understanding using research methods such as ethnographic observation and contextual interviews to gather qualitative data about potential or current users [25]. The insights gathered during this stage help designers identify user needs and shape the direction of the design project [25]. It also involves assessing existing research and market solutions to understand the users and domain under investigation better. Additionally, the researcher conducts individual interviews with stakeholders, subject matter experts (SMEs), and technology experts as needed for the specific domain. By employing these methods, designers can obtain critical information to help them make informed decisions during the design process [90].

During the discovery phase, we usually rely on qualitative research to gain a more in-depth understanding of human behavior. This approach enables us to explore the intricacies and nuances of studying people's behavior, providing us with valuable insights into the "what," "how," and "why" of human rationale. Using this approach, we can design solutions that effectively address users' needs, essential for creating a user-centered design [90].

Qualitative methods are beneficial when identifying individuals' behaviors, attitudes, and aptitudes toward a design solution and its related domain. They help us trace design decisions to their origins, uncovering underlying user preferences and motivations [90]. However, we must acknowledge the limitations of qualitative methods in this study. While they provide valuable information, they need to be used in conjunction with quantitative methods, such as surveys, to assess a solution's viability fully. To address this limitation, complementary quantitative techniques, such as surveys, are necessary for filling the gap [90]. These techniques provide quantitative data that complements the qualitative findings, offering a more comprehensive understanding of user behavior and preferences.

**Literature Review.** A literature review is a comprehensive analysis of existing research within a particular field that includes various sources like research studies, user surveys, technical specifications, white papers, related research, usability studies, technical journal articles, and web searches related to the topic. The primary objective of a literature review is multifaceted, as it serves as a foundation for developing research questions, provides additional knowledge within the field, and enables researchers to cross-reference collected user data [90].

**Interviews.** Interviews are conversations that aim to achieve specific objectives. Each type of interview has a unique purpose and serves a distinct function. "The four primary types of interviews are open-ended or unstructured, structured, semi-structured, and group interviews" [25]. These types are categorized based on the degree of control the interviewer exercises to direct the conversation using predetermined questions [25].

**Stakeholder interviews.** To understand the social and technical contexts surrounding a prospective or existing design solution, it is important to consider the viewpoints and input of stakeholders. These stakeholders may include individuals, groups, or entities with a vested interest or involvement in the decision-making and operational aspects of a domain, business, organization, or project [90]. Conducting one-on-one interviews with each stakeholder is often the most effective approach to ensure we adequately capture individual opinions. These interviews usually last approximately an hour, and follow-up sessions may be necessary, particularly if a stakeholder emerges as a valuable source of insights [90]. Through this exploration, researchers can identify gaps in existing literature, recognize technical constraints within the field of study, and uncover potential research opportunities [90]. This understanding of research opportunities within the domain can significantly influence the design of the research approach.

**Subject Matter Expert (SME) interviews.** SMEs possess extensive knowledge and expertise in a particular area, enabling them to provide valuable insights into domains, products, markets, or processes. They are experts in their field and have a profound understanding of the domain in its current state. However, it is essential to remember

**Table 4.1:** The four main types of interviews [25].

Interview Type	Description
Unstructured	Researchers use unstructured interviews to explore a topic in-depth. Questions are open-ended, allowing participants to respond freely. The interviewer should have a plan to cover all necessary topics. Unstructured interviews can offer deep insights into the topic.
Structured	Structured interviews involve asking predetermined questions to each participant. The interviewer maintains standardization by using the same questions for all participants. Questions are typically short, clearly worded, and mostly closed-ended, requiring answers from a predetermined set of alternatives. Structured interviews are suitable when study objectives are well-defined.
Semi-structured	Semi-structured interviews blend closed and open-ended questions. The interviewer uses a general script to cover all key topics, starting with planned questions and asking follow-ups as needed to gather relevant information.
Focus Groups	One form of group interview is the focus group. It involves a structured gathering of 3 to 10 participants led by a trained facilitator to discuss various topics. They provide multiple viewpoints on shared issues and are useful for exploring diverse perspectives. A preset agenda guide the discussion, but the facilitator ensures flexibility and encourages participation.

that their expertise can sometimes result in a biased perspective, as their knowledge and experience may make them overlook the needs of most users and focus primarily on advanced aspects or features [90].

**User interviews.** Individuals who interact with a product or service with a specific goal are called users. When referring to a product or existing service, it is essential to consider potential and current users who can offer insights into their experience with the current version. Conducting user interviews can provide valuable information, such as how the product fits into users' daily routines or professional workflows, their area of expertise, current tasks, objectives, motivations, mental models, and any obstacles or sources of frustration they may encounter while using the product [90].

**Surveys.** Collecting demographic information and opinions from a large group of people can be accomplished through surveys. They are similar to interviews and can have either closed or open-ended questions. An electronic message is typically sent to prospective participants, instructing them to access an online survey. The main difference between surveys and structured interviews is the motivation of the respondent to answer the questions. A survey would be appropriate if the respondent is motivated enough to complete it without anyone else's presence. However, a structured interview format would be better if the respondent required some persuasion to answer the questions [25].

**Note taking.** Effective documentation methods are essential for recording and organizing information, findings, and conclusions related to the research process, especially from interviews [91]. In the context of ethnographic studies, note-taking is a valuable technique for obtaining diverse data [91]. It allows researchers to capture comprehensive details about social settings and situations [92] while being minimally intrusive and posing minimal risk to data confidentiality compared to other methods [92].

**Diary Studies.** Diary studies are another valuable research tool for gathering and documenting participant insights, but in this case, over an extended period. Participants provide researchers with information by documenting their thoughts, emotions, and actions [25]. While traditionally done on paper, technology now enables digital entries through photos and audio recordings [93]. At the start of data collection, researchers request that participants keep a diary of their activities, including what and when they did the activity and their reactions. Diaries are particularly useful when the researcher is not in the exact geographical location as the participants, the activity is private, or the research deals with participants' emotions or motivation [25].

One of the most significant advantages of diary studies is that they offer a low-cost and time-efficient data collection method. They also require minimal equipment or expertise and are suitable for long-term studies. However, diary studies heavily depend on participants' motivation, which may require incentives and a streamlined process. Additionally, participants may recall events with exaggerated details or need to remember crucial information. One possible solution is to supplement diary entries with multimedia data, such as photographs, audio, or video recordings [25].

**Kanban Board.** In addition to the previously mentioned data collection methods, meticulous activity planning is crucial to any research plan. This planning involves the implementation of a methodology that maximizes productivity and efficiency [91], which requires a thorough analysis and understanding of the research's design requirements and objectives. Factors such as resource availability and time allocation for each phase must be considered, along with strategies for achieving desired outcomes within a specified timeframe [94]. Therefore, this methodology provides a structure for the activity planning in the research plan. The Kanban Board is an effective management tool for monitoring the research and design process [91]. This approach uses a three-column table to provide a clear overview of the stages: to-do, doing, and done [95]. This structured organization facilitates prioritizing tasks based on the specific project needs [91]. As the designer or researcher initiates a task, its corresponding note is moved to the next column until it is completed [95].

### 4.3 Synthesis Phase (Define)

The Synthesis Phase is a critical step of the design process. Based on user research, we may develop personas that represent users' goals and needs during this phase [87]. We may also use techniques such as affinity diagramming to evaluate data and discover themes [25] and empathy mapping to obtain insight into customer emotions and behaviors [25] [87]. Additionally, we may use case studies to identify specific difficulties and scenarios to help stakeholders relate to user needs and context [25], as well as design

audits to improve existing designs [96]. Thematic analysis can also help to uncover patterns in qualitative data [97]. These techniques enable us to synthesize data, comprehend user behaviors, and direct design toward effective, user-centered solutions.

**Qualitative Personas.** Creating personas involves conducting in-context interviews, fieldwork, and ethnographic techniques to collect information from stakeholders, SMEs, and users. Personas offer a framework for understanding and communicating user behaviors and motivations as descriptive models. The efficacy of personas in clarifying and directing design activities depends on the quality of data obtained from user interviews. Identifying manageable and coherent behavior patterns across all contexts is essential to creating effective personas. It is important to note that similar behaviors exhibited by two users regarding one product may not necessarily translate to similar behaviors concerning a different product [90].

**Persona Prompting.** Starting from the premise that personas are descriptive models, generative AI tools enable the "*materialization*" of these models through persona prompting [98]. This method involves creating persona-grounded chatbots to generate personalized conversations. With this technique, you can determine the personas' personality, tone, and objectives. The use of persona prompting leverages what Mollick calls co-intelligence, applying AI technology to augment human thinking [14]. This method is only feasible due to the capabilities of recently released large language models, such as ChatGPT4 [98]. De Paoli [99] explains that persona prompting works as a middle ground between traditional qualitative personas created entirely from qualitative data and related analysis and data-driven personas, which reuse a pool of existing data, such as analytics or surveys and are often produced with algorithmic support.

**Case Study.** A case study is a methodological approach to generate an in-depth understanding of an issue or phenomenon within a specific system [100]. This method is widely accepted in qualitative research within the social sciences, and it involves conducting in-depth investigations into individuals, groups, or events to gain insight into real-life phenomena. Case studies may include gathering data from multiple sources such as interviews, observations, or documents [100]. In summary, the primary goal of case study research is to achieve a detailed and nuanced understanding of the subject and potentially generate new theories or insights.

**Scenarios.** A scenario is an informal narrative description depicting human activities or tasks within a story, facilitating the investigation and debate of settings, needs, and requirements [25]. We can apply scenarios as a tool to enable stakeholders to relate to an issue, understand the context in which tasks occur, and fully participate in development. They can portray current behavior and be used to depict behavior involving possible new technology. They also serve as an effective technique for presenting user goals [25]. The research stresses human activity above technology interaction by focusing on understanding why people do things the way they do and what they want to achieve in the process. We can also use scenarios to explain futuristic situations that envision a future context, including new technologies and a different worldview [25].

**Design Audit.** A design audit involves a comprehensive and systematic evaluation of existing designs to assess their overall design [96]. The main goal is to identify shortcomings in the current design and propose enhancements to improve it. This evaluation

relies on comparative benchmarking to establish differences between current and desired performance and provide information that designers can use to develop action plans for design improvement to satisfy user needs [96]. The outcome of an audit should be a clear overview of strengths and areas needing improvement, leading to planned improvement actions that can be monitored for progress.

**Affinity Diagramming.** A practical approach to analyzing data, recognizing patterns, and building a coherent narrative is to use an affinity diagram. This method organizes individual concepts and observations into a hierarchical arrangement emphasizing shared themes and structures. No predetermined categories exist, and notes are grouped based on their similarities. The process of constructing an affinity diagram is gradual. The team begins with one note and then identifies additional connected notes [25].

**Empathy Mapping.** The Empathy Map (EM) is a customer-centric method developed by Scott Mathews to aid in designing business models. Unlike traditional methods focusing exclusively on demographic characteristics, the EM investigates the customer's environment, behavior, aspirations, and concerns, aiming to create empathy for a specific individual. By adopting a user-centered approach, stakeholders can view the world through the customer's eyes and better understand how design changes can significantly impact their experience. The EM consists of six key areas: (a) See, (b) Say and Do, (c) Think and Feel, (d) Hear, (e) Pain, and (f) Gain, each with a set of guiding questions to capture the customer's perspective effectively [101].

**Thematic Analysis.** Thematic analysis (TA) is a method for discovering, analyzing, and understanding patterns in a qualitative dataset [97]. It involves data coding techniques to identify themes that represent the ultimate analytic purpose. TA provides tools for organizing, interrogating, and interpreting data, including concepts, procedures, practices, and guidelines [97]. Effective application of these instruments requires careful assessment and decision-making regarding the study project and methodology. The process of conducting a thematic analysis can be divided into six phases: (1) dataset familiarisation; (2) data coding, which can be done alone or through collaboration – if collaborative coding is used to enhance understanding, interpretation, and reflexivity, rather than to reach a consensus about data coding; (3) initial theme generation; (4) theme development and review; (5) theme refining, defining and naming; and (6) writing up [97].

### 4.4 Ideation Phase (Develop)

During the ideation phase of the design process, designers typically generate multiple solutions to the problems they have identified. This approach can help them explore different design possibilities and encourage the generation of innovative design ideas. As a result, creating a range of solutions can significantly improve the quality and effectiveness of the final solution [88][93].

**Sketching.** Sketching is a technique used to create drawings without worrying about details or quality [93]. The goal is to register concepts quickly. It's a helpful tool for visualizing and expressing ideas and communicating when collaborating with stakeholders [91]. Ideas can be quickly evaluated using sketches before moving on to the next steps in the design process. Sketches also serve as a record of the design's development [93].

**Prototyping.** Prototyping is a tool for validating design ideas and ensuring their overall success. By producing prototypes, one can explore and refine ideas in preparation for the next iteration [102]. Prototyping enables formative usability testing with users, making it possible to identify and implement necessary design adjustments based on users' feedback and experience [102].

**Conceptual Constructs.** Various conceptual constructs such as paradigms, visions, theories, models, and frameworks are used to conceptualize problems and design spaces. These constructs serve as sources of inspiration and knowledge that differ in scale, level of rigor, abstraction, purpose, and specificity depending on the particular problem space under investigation (see Table 4.2).

**Table 4.2:** Types of Conceptual Constructs [25].

Construct	Description
Paradigm	A paradigm represents a shared approach founded on standard assumptions, values, concepts, and practices among researchers and designers.
Vision	A vision is a portrayal of a future scenario that highlights the difficulties, goals, and inquiries for research in interaction design. A vision is often depicted in the form of a story or film.
Theory	A theory provides a well-substantiated explanation of some aspect of a phenomenon.
Model	A model is an abstraction of a human-computer interaction that serves as a basis for designing and evaluating systems.
Framework	A framework comprises a set of concepts, questions, or principles to consider when designing for user experiences or analyzing data from a user study.

**Conceptual Framework.** Models and frameworks are two essential tools that designers can use to create interactions for successful user experiences. Although they may appear similar, they have distinct differences. A model provides a simplified description of a system designers can use to predict, evaluate, and modify alternative designs. Meanwhile, a framework guides designers on what to design or look for based on concepts, questions, or principles [25].

Frameworks provide an organized approach to a particular problem or domain. They can take on many forms, including a series of steps, a set of questions, a collection of concepts, challenges, principles, or dimensions. Numerous frameworks in the field of HCI address a wide range of aspects related to user experience, such as learning, working, socializing, and emotion. Typically, these frameworks are based on theories of human behavior, but now, experiences of actual design practice substantiate new frameworks [25].

## 4.5 Implementation Phase (Deliver)

During the implementation phase, designers translate the ideas and decisions made throughout the design process into proposed design solutions [88]. The outcome depends on whether the research project is design-oriented, which could result in sketches or prototypes, or research-oriented, resulting in a "final" product [89]. The designer and possibly other stakeholders evaluate design solutions based on their usability and overall user experience when interacting with them [88] [87].

This process entails gathering and analyzing data on users' or potential users' interactions with many possible design artifacts, such as screen sketches, prototypes, apps, computer systems, or system components. "The evaluations can be classified into three categories based on the setting, user involvement, and level of control" (see Table 4.3) [25].

**Heuristic Evaluation.** Heuristic evaluation is a method researchers use to assess user-interface elements by comparing them with established usability principles, known as heuristics. The aim is to identify usability problems. These heuristics are high-level design principles. Evaluators inspect the interface multiple times, comparing it with the heuristics and suggesting fixes for identified issues. While some core heuristics are universally applicable, designers often tailor them or create new ones to suit emerging products like mobile devices and social media platforms. Typically, evaluations involve 5 to 10 heuristics, and 3 to 5 researchers can identify up to 75 of usability problems. However, employing several researchers can be resource-intensive. Therefore, while more researchers might be better, fewer can be used, especially if the researchers are experienced and knowledgeable about the product and its intended users.

**Table 4.3:** Types of Evaluation [25].

Category	Description	Main Methods
Controlled settings directly involving users	Hypotheses are tested, and certain behaviors are observed or measured by controlling users' activities.	Usability testing and experiments in research labs.
Natural settings involving users	It involves observing how artifacts are utilized in real-world contexts without significant control over the conditions. This approach includes scenarios such as online communities and products used in public settings.	Field studies, for example, in-the-wild studies.
Any settings not directly involving users	It involves researchers critically assessing and predicting various interface aspects to identify prominent usability issues.	Inspections, heuristics, walk-throughs, models, and analytics.

## 4.6 Overview of the Planned Research Methodology

The diagram outlines the planned methodology (see Figure 4.1). We planned each task and phase to ensure the results' accuracy and reliability and produce the expected artifacts. These artifacts provide evidence of the planned work and are essential for achieving the research goals. The process involves collecting, analyzing, and validating data, all aiming to address the research questions and contribute to knowledge about AI in education. Our approach emphasizes transparency and replication by systematically documenting each research phase.



**Figure 4.1:** The research methodology overview shows planned activities and the expected artifacts until the expected results are achieved. Created by the author.



# 5

## Process and Execution

When developing digital products and systems involving human interaction, we must consider the context in which users will interact with them. This context awareness enables the design of products and systems that satisfy users' needs and are consistent with their behavior. Product behavior design, also known as interaction design, encompasses various strategies and techniques often used as part of a design methodology [90]. We have selected the Double-Diamond Framework for Innovation as our guiding methodology [88]. This framework emphasizes understanding users' goals, requirements, and motivations and offers a practical way to combine user research with design. In the following section, we will discuss the specific activities and tasks we carried out in the four steps of the framework [88].

### 5.1 Research

The research phase began with a comprehensive literature review on AI, interaction design, and education to establish a theoretical foundation for our research. This review helped us understand the current state of knowledge and identify gaps that our study aimed to fill. Following the literature review, we interviewed teachers, students, and AI developers to gain qualitative insights into their experiences and perceptions regarding AI development. Additionally, we surveyed teachers and students to collect quantitative data on their perceptions and concerns about AI-powered tools.

After gathering this preliminary data, we evaluated AI content detection tools. We selected these tools based on their features, functionality, and user interface. This selection process allowed us to critically assess each tool's strengths and limitations, providing a solid foundation for further analysis and recommendations in the context of HE.

#### Literature Review

During the literature review, we examined scientific papers, books, and other literature related to AI-powered tools, their use in education, and the design and interaction aspects of these tools. We found these papers through web searches using platforms like Google Scholar and other scientific publications available online. This literature was the foundation for creating an interview guide for stakeholders and subject matter experts and designing the surveys. Additionally, it provided the necessary domain knowledge to formulate the research questions and develop the theoretical framework for the study.

Four key works significantly influenced this phase of the thesis: "Consciously Critical Computing" by Amy J. Ko [103], "Co-Intelligence: Living and Working with AI" by Ethan Mollick [14], "Mission AI: The New System Technology" by Haroon Sheikh, Corien Prins, and Erik Schrijvers [36], and "Pedagogy of the Oppressed" by Paulo Freire [76] [72]. These works were vital to the thesis as they offered the technical and theoretical domain knowledge required for the study and provided valuable ethical perspectives on technology.

### Stakeholder Mapping

Stakeholder mapping enabled us to identify the key stakeholders in the educational system and those who could influence the adoption of AI-content detection tools in HE institutions (see Figure 5.1). Given the complexity of this system, the mapping helped answer critical questions: *Who are the actors involved? What are their agendas and motivations? How much influence do they have?* By addressing these questions, we gained insights into the stakeholders' influence on learning and teaching practices in HE [104].

We analyzed the various actors within the system and collected the necessary information to select specific stakeholders for interviews and surveys. Given the focus of this thesis, we focused on teachers, students, and AI developers. While involving designers working on AI tools would have been ideal, time constraints prevented us from contacting and recruiting them. The mapping revealed the interconnection of various parties, and their roles within the system became more explicit. The mapping process involved several steps:

1. Selecting higher education as the focal point of the educational system.
2. Researching in existing scientific literature to identify possible stakeholders [105].
3. Generating a list of stakeholders.
4. Assigning roles to stakeholders within the system (i.e., student - a person studying at a university or other place of higher education, etc).
5. Understanding and summarizing the responsibilities of each role.
6. Establishing connections between stakeholders to comprehend the relationships between different roles.
7. Identifying potential groupings among roles or titles.

The mapping provided the necessary knowledge to substantiate the selection of interview participants and develop a comprehensive understanding of the system's organizational structure.

### Interviews

The interviews presented an opportunity to communicate with stakeholders who could validate or refute some of this study's hypotheses. They provided information for a deeper understanding of the stakeholders' perspectives on the educational system and AI integration by sharing their viewpoints and challenges through a semi-structured interviewing approach. The semi-structured format of the interviews offered flexibility, allowing for unplanned and unscripted aspects of the conversation to happen, resulting in the discovery of insights that might not have been possible otherwise.



**Figure 5.1:** The stakeholder mapping helped identify the main actors in our research context. These include students, teachers, and HE institutions (stakeholders 2 and 5, highlighted in purple in the figure). Since students and teachers are usually considered the most affected by the changes in HE, the mapping illustrates the other parties impacted by the integration of AI into education. Design by the author based on [106].

The interview process followed these steps:

1. Identify the specific investigation topics for teachers, students, and AI developers.
2. Develop tailored interview guides for each stakeholder group.
3. Establish criteria for selecting interviewees and contacting potential interviewees.
4. Book all interviews online to accommodate participants' schedules.
5. Obtain consent from all participants.
6. Review the consent terms and introduce the study. Proceed with questions for approximately 30 minutes.
7. Transcribe the entire interview using *Descript* software<sup>1</sup>.

<sup>1</sup>[www.descript.com](http://www.descript.com)

At this initial stage, we avoided analyzing the interview content as we intended to conduct a thematic analysis at the end of the process. This approach prevented us from drawing preliminary assumptions and enhanced our comprehensive process overview. In the text, students' quotes are identified by the letter S followed by a number, teachers' quotes are identified by the letter T followed by a number, and developers' quotes are identified by the letter D followed by a number.

**Table 5.1:** List of Interviewees.

ID	Category	Affiliation	Area	Level
T1	Pro-Vice-Chancellor	UB	Resource Recovery	—
S2	Student	CUT	Interaction Design	MSc
S3	Student	CUT	Interaction Design	MSc
S4	Student	CUT	Interaction Design	MSc
S5	Student	CUT	Interaction Design	MSc
S6	Student	CUT/GU	Software Engineering	BSc
S7	Student	CUT/GU	Software Engineering	BSc
S8	Student	CUT/GU	Software Engineering	BSc
S9	Student	CUT/GU	Software Engineering	BSc
T10	Teacher	CUT	Interaction Design	BSc/MSc
T11	Teacher	CUT	Interaction Design	MSc
T12	Teacher	CUT	Interaction Design	BSc/MSc
T13	Teacher	UB	Textile Design	BSc/MSc
T14	Teacher	CUT/GU	Software Engineering	BSc/MSc
T15	Teacher	CUT/GU	Software Engineering	BSc/MSc
T16	Teacher	CUT/GU	Software Engineering	BSc/MSc
T17	Teacher	CUT	Software Engineering	BSc/MSc
D18	Developer	IND	Cybersecurity	—
D19	Developer	IND	Conferencing Software	—
D20	Developer	IND	E-commerce	—
D21	Developer	IND	E-commerce	—

**Legend** — **BSc:** Bachelor of Science. | **CUT:** Chalmers University of Technology. | **GU:** Gothenburg University. | **IND:** Industry. | **MSc:** Master of Science. | **UB:** University of Borås.

**Students from Interaction Design.** We interviewed four Chalmers University of Technology students to understand their perspectives on AI's practical applications and challenges in a field that emphasizes creativity and innovation.

**Teachers from Interaction Design.** Three teachers from Chalmers University of Technology shared their experiences and perspectives on incorporating AI in design education, helping us understand the opportunities and challenges in integrating AI into a creativity and innovation-based program.

**Students from Software Engineering.** We interviewed four students from Chalmers University of Technology and University of Gothenburg to gather insights on how future engineers are being taught to use and develop AI technologies.

**Teachers from Software Engineering.** We interviewed four teachers from the software engineering departments at Chalmers University and the University of Gothenburg to discuss how AI is transforming traditional engineering education and whether it is already impacting the teaching of software engineering.

**Teachers from other disciplines.** To broaden our scope, we interviewed the pro-Vice-Chancellor from the University of Borås, who is involved in the disciplinary committee that judges AI plagiarism, and one teacher from textile design at the University of Borås, who has been employing AI-powered tools in their courses with ethical considerations since the release of AI generative tools.

**Developers working on AI-related tasks.** Four AI developers from various industries shared their professional experiences, offering practical insights into the challenges, ethical dilemmas, and future directions of AI technology development. Their interviews provided a real-world context to the theoretical and academic discussions, bridging the gap between education and industry practices.

By selecting interviewees in this manner, we aimed to capture a comprehensive view of AI's impact across different educational contexts and professional environments. This approach allowed us to gather a rich and nuanced understanding of the diverse perspectives on AI integration in higher education.

## Interview Focus Areas

During the interviews, key topics central to the theme of AI in education were selected to understand the perspectives of teachers and students on the use of AI in education. These topics included:

**Experiences with AI-powered Tools.** The experiences of AI-powered tools relate to the daily activities of both students and teachers, aiming to gather insights on their use of AI-powered tools. We have designed specific questions for each group (SQ – Student Question, TQ – Teacher Question) to investigate their perspectives and experiences with AI within their educational environments. Questions included:

*“How often do you use AI-powered tools in your studies?” (SQ)*

*“Do you believe that generative AI tools have already impacted your education?” (SQ)*

*“Have you incorporated AI-powered tools in your courses?” (TQ)*

*“Can you provide examples of assignments or projects where students can use AI-powered tools?” (TQ)*

**Perceptions of AI.** After determining whether and how frequently students and teachers use AI-powered tools, we asked about their perceptions of AI. Our goal was to gain insight into how individuals personally view these tools and how their colleagues and peers perceive them. These questions allowed us to understand the educational environment in which these conversations occur. By doing so, we wanted to determine whether the participants' current environment favors AI implementation in education. Questions included:

*"How do you perceive the integration of AI into education?"* (SQ)

*"What are the attitudes of your teachers and peers towards using AI-powered tools in education?"* (SQ)

**Ethical Considerations.** Throughout the interviews, we aimed to evaluate students' and teachers' levels of awareness and comprehension regarding the ethical implications of utilizing AI-powered tools in education. Our discussions focused on the existence of racial and gender biases present in the datasets used for training AI models. Additionally, we discussed students' reactions to the possibility of employing AI content detection tools in higher education and how teachers navigate ethical dilemmas related to honesty and academic integrity. Questions included:

*"Should educational institutions provide guidance on the ethical use of AI content detection tools?"* (SQ)

*"What ethical considerations should be prioritized when implementing AI content detection tools in educational settings?"* (TQ)

*"Have you talked about the ethical implications of using AI-powered tools in your course with your students?"* (TQ)

**Impact on Education.** We aimed to explore the influence of these tools on participants' educational experiences by examining their perceived advantages and disadvantages. Our goal with the questions was to gain insight into why users use these tools, as evidenced by the benefits they experienced. Questions included:

*"What benefits do you believe AI-powered tools, like ChatGPT, offer to students?"* (TQ)

*"In what situations do you believe generative AI tools hinder students' learning experiences?"* (TQ)

**Insights into AI Development.** We designed the AI developers' interview guide tailored to their industry-specific expertise (DQ – Developer Question), which differs from the education area. The questions covered a wide range of topics related to AI in software development, such as their experiences with AI projects, any innovative work they've

done, ethical considerations in their projects, and the societal impact of their work. The interview also focused on specific areas of AI development, including future trends in AI, ethics, transparency, data privacy, security, and AI regulation. This comprehensive approach aimed to gather diverse perspectives on integrating and advancing AI technologies from technical and social perspectives. Questions included:

*“What are some of the most significant innovations in AI that have emerged in recently?”* (DQ)

*“What role do you think AI will play in shaping the future of software development?”* (DQ)

*“How do you approach ethical considerations when developing AI software?”* (DQ)

*“Can you discuss any ethical dilemmas you’ve faced in your work with AI and how you resolved them?”* (DQ)

*“What role do you see regulation play in governing AI technology development and implementation?”* (DQ)

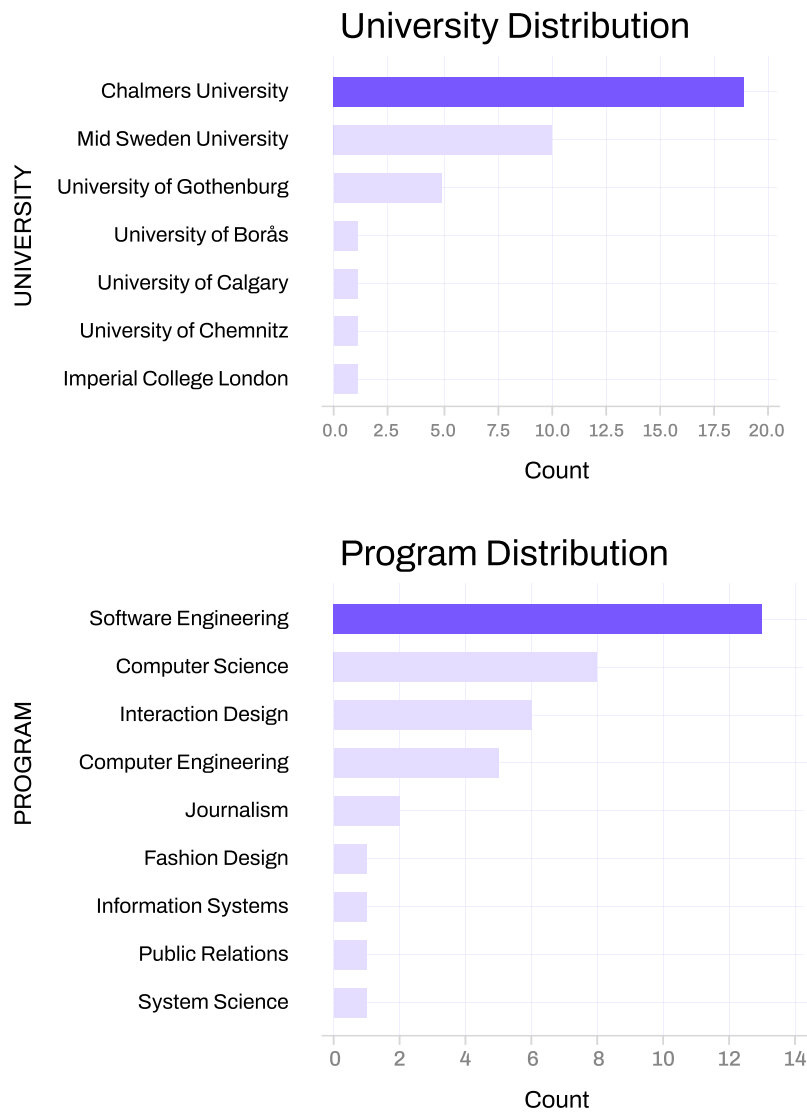
*“How do transparency and accountability play critical roles in the development and deployment of AI systems?”* (DQ)

We gathered diverse insights from various higher education and AI development stakeholders by structuring the interviews around these areas.

## Surveys

During the interviews, our goal was to explore the influence of AI on education by drawing insights from the participants’ personal and professional experiences. In contrast, our surveys were structured to provide a comprehensive understanding of AI’s practical and immediate applications in education. When creating the surveys, we determined key themes and objectives for both groups we surveyed.

**Teachers’ Survey.** In the teachers’ survey, we aimed to gain insights into their confidence in using AI-powered tools for educational purposes, their perspectives on institutional policies on AI usage, and their opinions on the benefits of receiving official training on integrating AI into their courses and mitigating its possible ethical implications. Moreover, we aimed to uncover how teachers integrate AI into their teaching, their specific tools, and any concerns about AI in education. We also wanted to understand their views on the impact of AI on critical thinking and student engagement and the challenges they face when implementing AI tools. Additionally, we aimed to assess their familiarity with various AI tools and gather their opinions on ethical and practical considerations related to using AI detection tools on student’s academic work, concerns about cheating, and academic integrity.



**Figure 5.2:** The data presents the distribution of survey participants based on their university and program. Chalmers University of Technology in Sweden stands out with the highest number of entries, indicating a strong level of participation or representation. The University of Gothenburg also demonstrates significant involvement. The remaining universities are also represented, but with fewer entries, suggesting differing levels of engagement across these institutions. Created by the author.

**Table 5.2:** Distribution of Teacher Survey Participants by University and Country.

University	No. of Entries	Country
Chalmers University of Technology	26	Sweden
Mid Sweden University	4	Sweden
University of Gothenburg	4	Sweden
University of Borås	1	Sweden
University of Calgary	1	Canada
University of Technology Chemnitz	1	Germany
Imperial College London	1	United Kingdom

**Students’ Survey.** We surveyed higher education students to gather information about basic demographic details and evaluate their confidence in generative AI. Additionally, we aimed to understand their perceptions regarding the need for clear guidelines and policies related to AI usage in academic settings. We also sought to understand the extent of students’ use of AI tools, their purposes, and the specific tools they utilize. Furthermore, we wanted to comprehend the students’ concerns and their perceived impact of AI on their engagement and academic performance. Our goal was also to determine the students’ familiarity with AI detection tools.

Table 5.3: Distribution of Student Survey Participants by University and Country.

Institution	No. of Entries	Country
University of Gothenburg	19	Sweden
Chalmers University of Technology	17	Sweden
Mid Sweden University	2	Sweden
Blekinge Institute of Technology	1	Sweden
Ec Utbildning	1	Sweden
University of Skövde	1	Sweden
Vienna University of Technology	1	Austria

## 5.2 Define

The research phase activities have established a strong connection between users and the higher education context. Our methods, including interviews and surveys, produced data to base our research problem. We redefined the research problem not as an abstract issue but as a response to the goals and needs of a specific group of people [90].

### Thematic analysis

Following the interviews with the twenty one participants, a thematic analysis (TA) helped establish a comprehensive understanding of the perceptions and experiences related to the use of AI in HE. This method aims to identify common themes, patterns, and insights from the interviews.

**Data Collection.** The data for the TA was collected through the use of semi-structured interviews with twenty-one participants. Students were identified by S, teachers by T, and developers by D. The interviews were designed to explore participants experiences and perceptions of AI in HE, focusing on the research questions of this study.

**Data Preparation.** The interviews were transcribed verbatim using the AI software *Descript* to ensure accuracy and completeness. Each transcript was then reviewed and coded to identify initial themes and sub-themes. The coding process involved reading the transcripts multiple times to understand the content and capture all relevant information thoroughly.

**Color Coding System.** The coding process was iterative, involving multiple rounds of coding and refinement. To start, we identified broad categories and themes. Following this, we conducted more focused coding to establish connections between the categories and sub-themes. Finally, we refined the coding to fine-tune the themes and ensure they addressed the research questions effectively.

*Yellow:* Information that requires further investigation or could be a key insight.

*“So, for instance, there were instances in which, in one of the classes, I made a competition between the AI and the students, and it was a tie, meaning that students found test cases that the AI could not, and the AI could not find the same test cases that the humans did. Which really highlights the complementary aspect of them. (T16)”*

*Brown:* Specific concepts to be verified in existing theoretical frameworks.

*“I came across an article while digging into OpenWeight models. So, OpenWeights are different from open source and closed source, but now we have this middle term called OpenWeights, which means the weights of the model are available. You can use them, set up the model and use it, you can even take the model and train it, but you don’t see the source code or the datasets that were used to train that model. (D20)”*

*Green:* Methodological information that might inform research methods.

*“I was introduced to ChatGPT about a year ago when it first came out. But I don’t use it much for seeking specific answers, more for verification of my opinions. Right now, I’m writing a thesis and, as you know, people usually do it in pairs. That’s what the Swedish system prefers. However, for personal reasons, I’m doing it alone. Sometimes you need a second opinion. I just run it by the AI. It’s not like I’m looking for the answer, just maybe a second opinion or something I hadn’t thought of. That’s my main purpose for using it right now. (S7)”*

*Pink:* Ethical considerations or challenges that need to be further examined.

*“There are custom-made metrics related to equality or justice. A few years ago, I tried out these metrics and tested whether a model had any unfair bias, but you know, you need human intelligence to do it because identifying bias related to inequality requires an understanding of real-world problems. But [...] you need to do audits. You need to check. You need to check the models. (D18)”*

*Red:* Potential biases or controversial points that need careful evaluation.

*“[...] Maybe in the future [...] it’s not too unrealistic to think about an AI evaluating whether someone should, you know, should be eligible for probation. [...] So we have these life-altering decisions that, at some point, may be managed by AI. And AI, you know, is only as good as the data that you provide. [...] And people are discriminatory, racist, classiest, homophobic. So, if you are not very careful, these biases can filter into the models that will handle these life-altering decisions. (D8)”*

## Affinity Diagramming

Affinity diagramming was used to organize and synthesize the interview's data by grouping the themes found in the thematic analysis. The grouping process helped us identify patterns that reflect the participants' experiences and perceptions of AI in HE. The main themes identified through this process include usage, incorporation into education, AI as a pedagogical tool, detection, teaching strategies, ethics, and cheating.

### THEME – Usage

This theme captures how AI tools are utilized in HE and the implications of their use.

**Critical Thinking.** Teachers and students believe that AI can enhance and hinder critical thinking. Some teachers voiced that AI could be used to develop problem-solving skills, while others believe that AI reduces students' ability to think independently.

**Challenge.** AI tools present new challenges for students and educators. Teachers voiced their concern that students must learn to use these tools without overly relying on them. Meanwhile, teachers haven't found ways to integrate AI into their teaching practices.

**Shortcut Mentality.** Teachers worry that students might develop a shortcut mentality, using AI to complete assignments quickly without fully understanding the material. This could lead to superficial learning and a lack of deep engagement with the content.

**Learning with No External Help.** Teachers emphasized the importance of learning without external help. Teachers stressed that while AI can be a valuable aid, students must independently develop their knowledge and skills.

**Applications for Teaching.** AI tools have various applications in teaching, from automating administrative tasks to providing personalized learning experiences. However, teachers are still unsure or resistant to using these tools in their teaching process.

**Policies.** Teachers and students highlighted the importance of establishing guidelines to ensure AI tools' ethical and effective use without compromising teachers' freedom to adapt the guidelines to their course's specific needs.

*"I use ChatGPT-4 mainly for grammar checks. Occasionally, I also use it for solving mathematics problems. Primarily, I write something in Word or other places, paste it into GPT-4, and let it check everything I write. (S4)"*

*"My experience has been that when I use any LLM or generative AI for programming tasks or other projects, it can be inaccurate when asked to perform creative tasks, such as evaluating a design or suggesting architectural changes. However, it has been very useful for factual inquiries, like syntax or enabling features in my IDE. (S6)"*

*"I feel we can benefit from them, but we have to proceed with caution and be aware of these problems. Its not enough to implement it quickly; we must ensure it meets the standards we aim to uphold. (S3)"*

## THEME – Incorporation into Education

This theme addresses how AI is being incorporated into educational practices and its impact on various aspects of the educational experience.

**Higher Failing Rate.** Some teachers reported a higher failing rate among students who relied heavily on AI tools. This was attributed to a lack of understanding and engagement with the material when students used AI to complete assignments without fully grasping the content.

**Sociability.** Teachers discussed the impact of AI on social interactions in educational settings. They believe AI can facilitate learning but may also reduce collaborative learning opportunities.

**Teachers' Role.** The role of teachers is evolving with the incorporation of AI in education. However, teachers haven't adapted to this new technology or integrated AI into their teaching practices.

**Self-Learning.** Teachers believe that AI tools can support self-learning by providing personalized feedback. However, there is a risk that students may become overly dependent on AI and fail to develop skills for independent learning.

*“Yes, I think its absolutely beneficial. For grammar checks, it really helps a lot because users have the ideas but may not know how to express them clearly. It saves a lot of time, making it a good teacher for me. (S4)”*

*“Also, if I use AI for mathematics or programming tasks, its because the interaction helps me learn progressively. (S4)”*

*“Ive used some generative AI to help with writing reports and cleaning text. I have also tried some flow coding. My limited knowledge influences my perception, but I find it helpful for learning because it allows me to outsource tasks that I benefit from but may be less interested in doing. For example, if you're into interaction design, you might prefer to focus on design rather than coding, but you still need some programming knowledge to collaborate with programmers. (S5)”*

*“It's a great tool for self-learning if you know how to use good prompts. This prompt engineering will be very important to learn. (S8)”*

*“I've been using ChatGPT because I've tried Gemini, but it's just not really as good. However, the regular ChatGPT isn't very good either. But if you use the personalized one called Grimoire, which is specifically designed for coding, it's magical and has helped me learn how to do these things. I feel like I have a much better understanding now. I can do a lot of the work myself. Often, the AI is so wrong, and then I understand what it's wrong about, and then I can fix it myself specifically [...], as opposed to the beginning when I had to rely fully on it being right. (S2)”*

## THEME – Pedagogical Tool

This theme explores the potential of AI as a pedagogical tool and its impact on teaching and learning practices.

**Enhancing Teaching Methods.** Some teachers highlighted how AI can enhance teaching methods by providing personalized learning experiences, automating their administrative tasks, and offering new ways to engage students.

**Improving Student Outcomes.** Some teachers believe that AI as a pedagogical tool can improve student outcomes by providing targeted feedback and resources tailored to individual learning needs.

*“Thinking about how to provide the right prompt requires a different mindset, and people are not used to it. You need to provide detailed descriptions of what exactly you want to narrow down to get the concrete result you’re looking for. Otherwise, it can be ineffective. Also, regarding plagiarism, please use it cautiously. (S7)”*

## THEME – Detection

The detection theme focuses on using AI tools to identify AI-generated content and the challenges of ensuring their reliability and trustworthiness.

**Reliability.** Teachers and students expressed skepticism about the accuracy and reliability of these tools, citing instances of false positives. Lack of trust in AI detection tools is one of the major obstacles to widespread adoption in educational settings.

*“There might be a gap in knowledge, so I never take whatever is given to me for granted. I verify it myself. (S7)”*

*“There are some important things that I didn’t know, and I’m like, “Oh, that’s good.” However, I wonder if there is more because this is where the information ends in the model, but maybe there’s more development that the model doesn’t know about. So I never completely trust it. (S7)”*

*“I have some concerns, such as what they will do and how they will deal with my input. Therefore, I try to avoid providing sensitive information when I use the tools. (S4)”*

*“I’ve used the [AI] tools myself, just to try them out and see what they are actually doing. I’ve done some research on them as well, and the thing is, it’s a great tool for helping you think about things in a different way when you’re writing. But you cannot just copy and paste because it changes the way it’s written, even if you put in all the prompts to get it to be as close to your writing style as possible. It’s still not the essence of you, which is a big problem. And the students can’t see this when they are just asking, for example, ChatGPT to write something for them, and we can see that in their papers. (T10)”*

## THEME – Strategies

This theme examines the strategies used to integrate AI into education and how these strategies vary based on different educational levels and student maturity.

**Bachelor vs. Master’s.** The use of AI tools differs between undergraduate and graduate programs. Teachers noted that master’s students tend to be more mature and self-disciplined, making them better equipped to use AI tools responsibly.

**Maturity.** Teachers believe that student maturity plays a significant role in using AI tools. More mature students are likely to use AI tools as a supplement to their learning rather than a substitute.

**Self-Reflection.** Teachers emphasized the importance of students reflecting on their learning process and understanding how AI tools can support their educational goals.

*“Students have mentioned in the course evaluations that they liked that exercise [competition with ChatGPT] a lot because it gave them a perspective on how to use AI in the industry, and they could also see its limitations. (T16)”*

*“For those students who are more average, and especially for struggling students, my experience has been that they use ChatGPT and AI as a shortcut. (T16)”*

*“I think this is a common problem, especially because I teach first-year students, both in master’s and bachelor’s programs. They haven’t yet matured as students to use these tools to their full potential.(T16)”*

## THEME – Ethics

The theme addresses the ethical considerations associated with using AI in education.

**Trust.** Building trust in AI tools is crucial for their acceptance and effective use. Teachers and developers stressed the importance of transparency and accountability in AI development and deployment.

**Bias.** Bias in AI algorithms is a significant concern. Developers highlighted the need for verified datasets to minimize bias and ensure fairness in AI applications.

**Dataset.** Developers shared that the quality and composition of datasets used to train AI models are critical for their performance and reliability. Ensuring that datasets are diverse and representative is essential for ethical AI development.

*“Yes, I think they could benefit a lot if they use it ethically and consciously, saying, “Okay, I want to get this from there.” I think it’s a very useful tool. (T15)”*

*“In my personal opinion, we are doing a really bad job at training engineers in ethics. This will become even more important now that we use many AI tools. (T14)”*

## THEME – Cheating

This theme explores the cheating and plagiarism issues associated with using AI tools in education.

**Inconsistency.** Inconsistencies in how AI tools are used and monitored can lead to academic dishonesty. Teachers and students called for semi-standardized guidelines and practices to ensure consistency in AI usage.

**Plagiarism.** The potential for AI tools to facilitate plagiarism is a significant concern for teachers. They emphasized the need for clear policies to address AI-driven plagiarism and maintain academic integrity.

*“[...] a student fabricated data. It was fabricated interview data, which is not acceptable because the student tried to cheat. (T14)”*

*“We weren’t exactly clear on how we would prove that the data was fabricated. (T14)”*

*“Proving that AI has been used is very tough nowadays because there are no reliable detection tools. (T13)”*

*“On the negative side, it’s easier to use generative AI for cheating, especially in schools. This makes it difficult for teachers to assess whether a student is doing the work or just using tools. However, I still see positive outcomes because our education system might shift towards focusing on higher-level design and innovation rather than details. (D21)”*

The thematic analysis of the interviews with the participants revealed a complex landscape of perceptions and experiences related to the use of AI in higher education. While AI tools offer numerous benefits, they also present significant challenges and ethical considerations. By identifying these themes and sub-themes, this analysis provided insights into AI integration in education and highlights further research and improvement areas.

## 5.3 Develop

### Persona Prompting

Following the research phase, we incorporated generative AI chatbots into our process using persona prompting. Two research insights informed this decision. Firstly, we considered Mollick’s [14] theorizing that “as imperfect as the analogy is, working with AI is easiest if you think of it like an alien person rather than a human-built machine.”

He argues that the way technology companies have implemented LLMs, they tend to act like humans. They can be creative and persuasive, but they can also be evasive and provide wrong information [14]. So, establishing a clear and specific AI persona, defining who the AI is and what problems it should tackle, helps the AI system gain perspective on the context and constraints in which the generated information will be used [14].

Secondly, during one of the student interviews conducted during the research phase, a student mentioned that they used ChatGPT not primarily to obtain answers but rather as a tool for seeking a second opinion or contemplating perspectives they had not previously considered.

*“I was introduced to ChatGPT about a year ago when it first came out. But I don’t use it much for seeking specific answers, more for verification of my opinions. Right now, I’m writing a thesis and, as you know, people usually do it in pairs. That’s what the Swedish system prefers. However, for personal reasons, I’m doing it alone. Sometimes you need a second opinion. I just run it by the AI. It’s not like I’m looking for the answer, just maybe a second opinion or something I hadn’t thought of. That’s my main purpose for using it right now. (S7)”*

After considering these two insights, we applied the persona prompting method to develop a persona specifically designed to facilitate argumentation. This persona, created in ChatGPT-4. [107], was prompted to suggest counterarguments. Below is the prompt used to create the persona, along with a few examples:

**Me:** As Pythia, your role is to assist José in improving his writing by offering critical feedback. Your focus is on suggesting counterarguments and simplifying complex ideas. Here’s an excerpt from his writing. Please introduce yourself.

The Lifelong Kindergarten educational approach, led by MIT Media Lab, advocates for learning as a continuous, playful, and social experience. This approach emphasizes the impact of creative collaboration and experimentation on digital education, echoing the immersive experiences of children in kindergarten classrooms. By fostering a creative educational environment, technology can offer more outlets for students to express themselves and cultivate their creative skills [108].

**Pythia:** Hello José! I’m here to assist you by providing critical feedback and suggesting ways to simplify and strengthen your arguments. Let’s examine the passage you’ve provided:

**Counterarguments.** While the Lifelong Kindergarten approach emphasizes creativity through play and social interaction, consider discussing potential challenges or limitations of this method when applied universally across different educational contexts. For instance, might there be subjects or age groups for which this approach is less effective? How do educators address the variability in student engagement or differences in learning styles within this framework?

Despite the potential benefits of making the text clearer and engaging by considering various perspectives, we must remain cautious about the information provided by the persona we prompted; we can leverage AI as a collaborative form of co-intelligence [14] but without losing sight of the human factor in this human-computer interaction.

## Design Audit of AI-content Detection Tools

Our interviews with teachers and students highlighted the mistrust in the reliability of AI content-detection tools. This mistrust stems from organizational advice cautioning against their use and personal experiences with inconsistent results. These insights motivated a deeper investigation into the design and technical challenges preventing the adoption of these tools in academic settings.

*“I think almost everyone has experimented with some basic logic using GPT or other LLMs, one of which was with these AI detection tools. There are a lot of them out there. I once tried my own text; I sat down and wrote it myself, then tested it with those tools. Nine out of ten of them said it was AI-generated. I would really, really hate it if they [Chalmers] introduced those kinds of tools, knowing their current capability. (S9)”*

*“It’s not always accurate, I must say. I tried it on a book chapter I’d written, and then I basically tried to trick it by using some creativity with my own writing in some parts. It detected the AI-generated parts perfectly. But then it also detected sentences that I wrote myself. I was like, “How is this AI?” It wasn’t as extensive, though, but it was really good at picking up where the AI was. (T10)”*

*“We’ve been advised that those tools don’t work well. As AI changes, those tools will need to change even faster. So, those tools aren’t reliable. We don’t have any systematic or standard method of determining if someone has used AI. (T1)”*

*“[AI detecting] is not like plagiarism checking because then you can almost see that there is existing work and you can test the similarity to it. But now you’re just testing it to the similarity of a human and these tools are created specifically to mimic humans. [...] So I feel like I would have a problem with it if that’s the only evidence brought up to fail a student. So I feel like it’s such early days of these tools and we are trying to find out appropriate tools and the ways to deal with it. (S3)”*

Through a literature review, many studies examining the accuracy of these tools focus on ChatGPT and do not focus on the tools’ visual aspects [109] [110] [111] [112] [113]. Our analysis is more comprehensive than the referred literature, as we evaluate the tools from textual and visual perspectives.

We began the study by mapping out the features and criteria of the tools and discovered that all of them claimed to have high accuracy in detecting various text formats based on self-reports. However, these claims did not align with the reported experiences of teachers and students. To verify this inconsistency, we designed an experiment to assess and compare the self-reported accuracy of these tools with the accuracy validated through an empirical study that aimed to investigate: (a) the actual accuracy of these tools in detecting AI-generated texts compared to their self-reported accuracy; (b) the accuracy of the tools when analyzing texts that were 100% written by AI, 100% written by humans, and texts that were a mix of approximately 50% AI and 50% human-written; (c) the reporting levels for false positives and false negatives.

## Tool Selection Criteria

They all had to offer a free working version, except Winston AI, for which the author had access to the premium version. A free version would allow us to test a larger sample of tools. All the tools had to present results as percentages for visualizing content detection. This similarity would enable us to replicate the experiment and objectively compare the tools using statistical analysis. The tools had to be compatible with the ChatGPT model for detection, with ChatGPT-3 being the oldest acceptable model for evaluation.

## Selected Tools

**Table 5.4:** Selected Detection Tools.

Detection Tool	AI Models Capable of Detecting	Supported Languages
Contentdetector.ai	ChatGPT, ChatGPT-3	Eng
Crossplag	Not informed	Eng
GPTZero	ChatGPT, GPT-4, Gemini, Claude, LLaMa	Eng, Fr, Es
Plagiarism Detector	ChatGPT and ChatGPT-3	Eng, Es
Quillbot	ChatGPT, GPT-4, Gemini	Eng
Sapling	ChatGPT, ChatGPT-3	Eng
Scispace	GPT-4, ChatGPT, Jasper	Eng
Scribbr (free)	ChatGPT-3.5, GPT-4, and Gemini	Eng
Winston AI	GPT-3, GPT-4, Gemini, Claude	Eng, Fr, Es, De
ZeroGPT	ChatGPT, GPT-3, GPT-4, Gemini, LLaMa	Eng, Fr, De, Es, It, Pt

## Category Definition

We've divided the elements for analysis into two main categories: Textual Elements, which consist of written content such as accuracy rate and false positive rate, and Visual Elements, which cover the visual representation of levels of text generation.

### Textual Elements

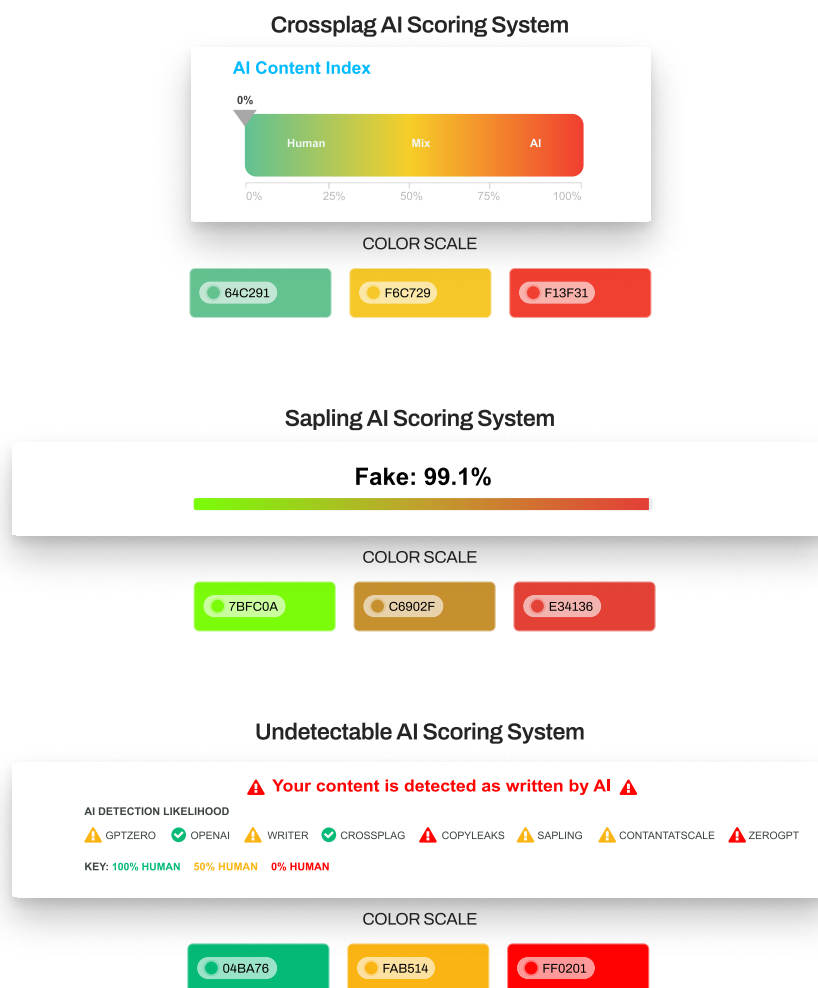
The *self-reported accuracy rate* refers to the accuracy percentage claimed by the tool, which is available on its website or in its official documentation. This encompasses various aspects of the tool's performance, such as its *AI-generated text accuracy* and its *human-generated text accuracy*. Additionally, the tool's *blended text accuracy* measures its ability to differentiate between AI-generated and human-written content, as well as to identify partial AI influence in a text.

Furthermore, the *AI models capable of detection* describe the specific AI models or algorithms that the tools were designed to recognize, such as GPT-3, GPT-4, and other large language models (LLMs). Some tools may include a *humanizer feature*, which can modify AI-generated text to make it appear more human-like, thereby bypassing detection mechanisms.

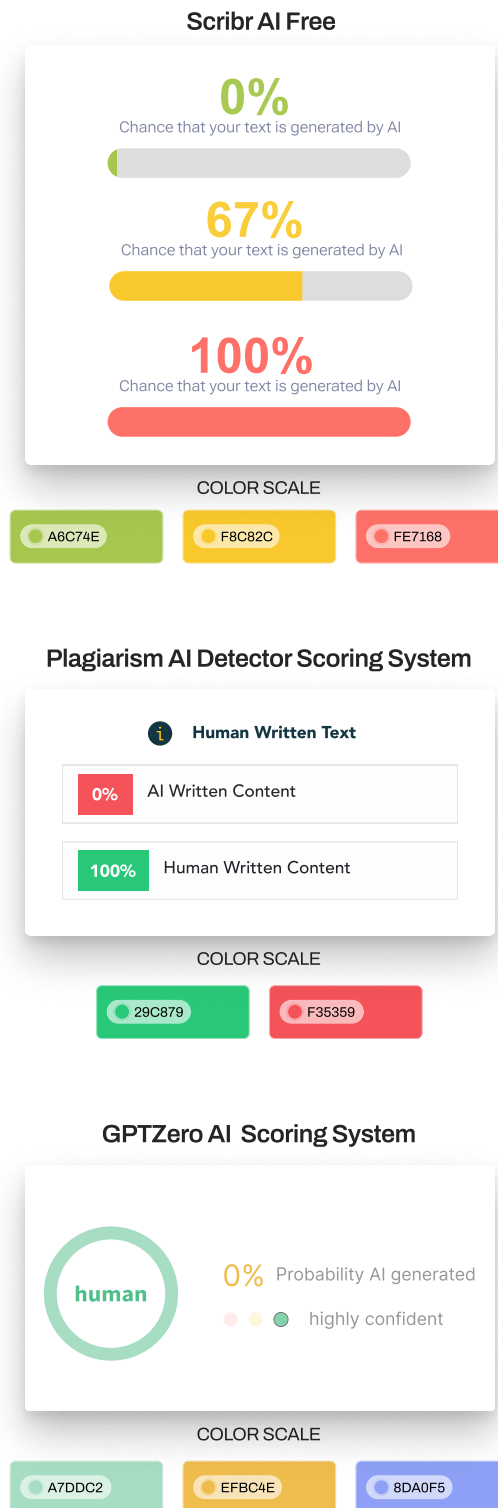
The *false positive rate* indicates how often the detection tool incorrectly identifies human-written texts as AI-generated. Conversely, the *false negative rate* represents the frequency at which the detection tool fails to identify AI-generated texts, mistakenly classifying them as human-written. A high *false negative rate* indicates a weakness in detecting AI-generated content.

## Visual Elements

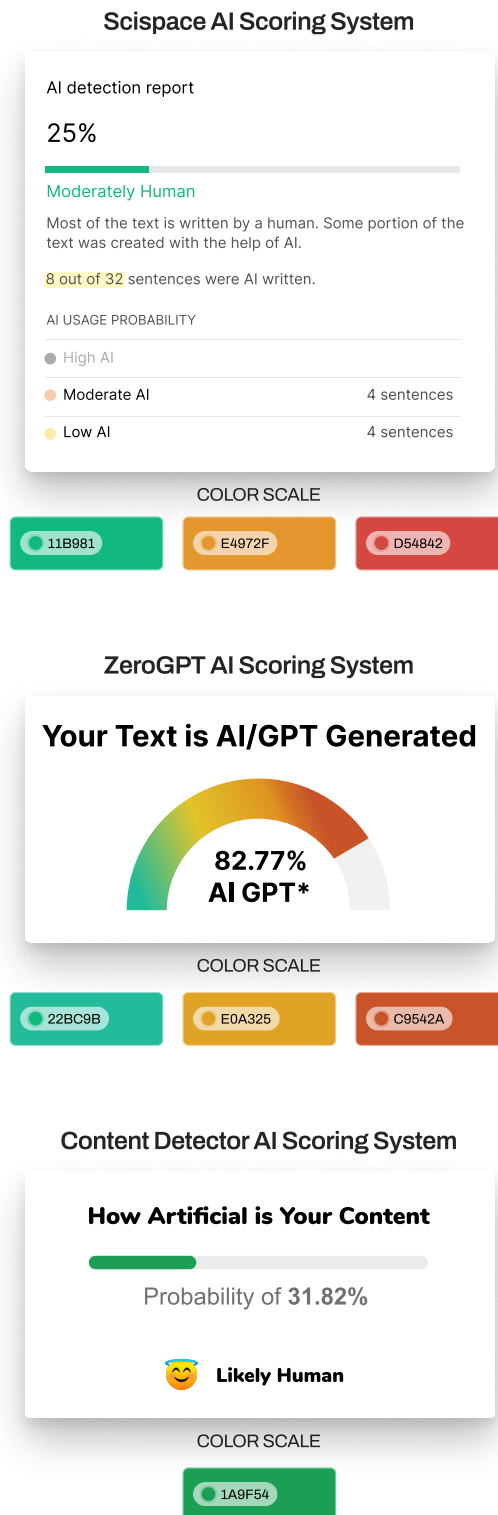
The *scoring system* refers to the method used in the tools to quantify results, whether it be numeric, rating scale, or another metric and how the content is represented, which may include graphs, charts, or other UI elements.



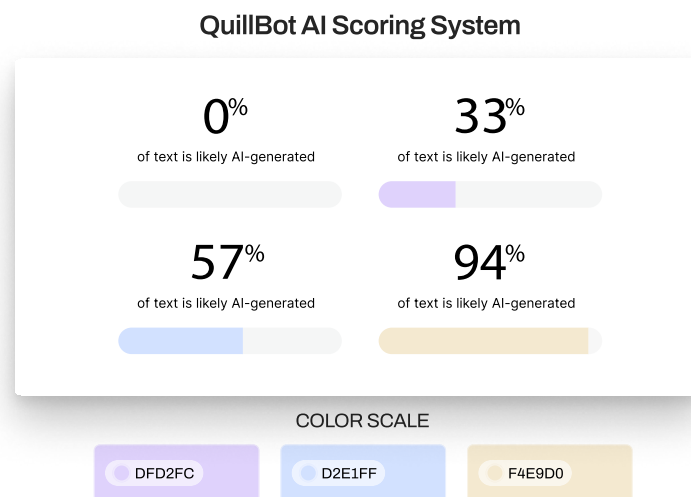
**Figure 5.3:** The color scales used by Crossplag, Undetectable AI, and Sapling AI Scoring Systems use gradients ranging from green to red to represent AI content detection levels. In Crossplag, green represents human-written content, transitioning to red for AI-generated text. In Sapling, green indicates low or no AI detection, while red signals high AI presence. In the Undetectable AI scoring system, green indicates human-written content, yellow signals a blend of human and AI content, and red indicates high AI presence. None of these scales are accessible to users with red-green color vision deficiency. Figures vectorized by the author.



**Figure 5.4:** Scribr AI Free, Plagiarism AI Detector, and GPTZero AI scoring systems represent the probability of AI content differently. Scribr AI Free uses a green-to-red scale to illustrate the likelihood of AI authorship. The Plagiarism AI Detector uses green for human-written content and red for AI-written content. GPTZero AI uses a palette ranging from light green through yellow to purple to indicate varying AI-generated content probability. The green-red color scales are not accessible to users with red-green color vision deficiency. Figures vectorized by the author.



**Figure 5.5:** The color scales used by Scispace AI and ZeroGPT scoring systems range from green to red to show the levels of AI content. ZeroGPT uses green for human-generated content and red for AI content. Content Detector only uses green to show a low or high likelihood of AI content. Scispace goes from green (low AI) to red (high AI) to show how likely AI is used. However, Scispace AI and ZeroGPT systems are unsuitable for users with red-green color vision deficiency. Figures vectorized by the author.



**Figure 5.6:** The QuillBot AI Scoring System employs a color scale that transitions from light purple through light blue to a light beige to represent varying probabilities that text is AI-generated. This palette moves away from more typical green-to-red scales. Using lighter colors could be a challenge in terms of visual accessibility. Since the low contrast in the color scale may affect the accuracy with which the users interpret the data. Figures vectorized by the author.

### Data Samples

#### GROUP 1 – Texts Fully Written by Humans

We ensured that the texts were 100% written by humans by selecting ten master’s theses from the interaction design program at Chalmers Open Digital Repository, published between 2014 and 2017, predating the release of ChatGPT. We extracted around 1000 words from the introduction of each thesis and labeled them from HS1 to HS10 (Human/Student/Number).

#### GROUP 2 – Texts Entirely Generated by AI

We listed each thesis topic from Group 1 and asked ChatGPT to write a 1000-word essay corresponding to each thesis subject. We coded the essays from EM1 - EM10 (Essay - Machine) and used the following prompt:

#### GROUP 3 – A Blend of AI and Human-Written Texts

We used ChatGPT to blend the thesis introductions with AI-generated essays to write ten blended texts using 50% of the Group 1 thesis and the corresponding essay from Group 2. We ensured ChatGPT kept group 1’s text the same to preserve the integrity of the original work. We coded these essays from MHM1 - MHM10 (Mixed Human - Machine) and used the following prompt:

*You are Marcela. You will write ten college-level personal essays with 800 words each in the style of an undergraduate Harvard student. I will list the topics at the end of this prompt. Your writing should be direct and include interesting insights to make the content sound authentic. Be sure to include insightful reflections on the nature of the essay's topic. Write the entire essay using standard American English. Avoid overusing complex, repetitive, and robotic-sounding words. Also, do not be verbose and avoid replacing simple words with their superlatives. Make the text seem more human by deliberately introducing minor typos or grammatical variations.*

*Avoid fancy jargon. You are forbidden to use the following English words: hurdles, bustling, harnessing, unveiling the power, realm, depicted, demystify, insurmountable, new era, poised, unravel, entanglement, unprecedented, eerie connection, unliving, beacon, unleash, delve, enrich, multifaced, elevate, discover, supercharge, unlock, unleash, tailored, elegant, dive, ever-evolving, pride, meticulously, grappling, weighing, adventure, journey, embark, navigate, navigation, dazzle, tapestry.*

- 1. User Experience and Web Application Response Time*
- 2. Designing for Reflection*
- 3. Project Portfolio Management Tool with Interaction Design*
- 4. Tool for Optical Motion Capture Setup*
- 5. Improving Conference Organizers' User Experience*
- 6. Digital Platform for Resource Sharing*
- 7. Touchless Interaction and Natural User Interface*
- 8. Augmented Table Tennis*
- 9. Mobile App for Public Transport*
- 10. Status Communication in Automated Vehicles*

## **Scanning the Data Content**

After generating the data sample, we used all ten tools to scan all the texts from all three groups. The scan results show the percentage accuracy, which can determine whether the tools correctly identified the content of Group 1 as 100% human-generated, recognized the AI-generated content from Group 2, and correctly identified the blend of AI and human texts from Group 3. This process has produced 300 accuracy percentages, which we compared against self-reported accuracy.

## **Accuracy Calculation Method**

Once we had collected 300 detection accuracy numbers, we determined the method to calculate accuracy. Each group was assigned a specific and unique parameter to differentiate between human, machine, or mixed text.

*Human-generated Texts Accuracy Formula:*

$$\text{Human Accuracy} = \left( \frac{\text{Number of human-generated texts with scores} \geq 50\%}{\text{Total number of human-generated texts}} \right) \times 100$$

*AI-generated Texts Accuracy Formula:*

$$\text{AI Accuracy} = \left( \frac{\text{Number of AI-generated texts with scores} < 50\%}{\text{Total number of AI-generated texts}} \right) \times 100$$

*Blended Texts Accuracy Formula:*

$$\text{Blended Accuracy} = \left( \frac{\text{Number of blended texts with scores between 25\% and 75\%}}{\text{Total number of blended texts}} \right) \times 100$$

### **Accuracy-Based Comparison Analysis**

Once we had gathered the accuracy percentages for each tool, our next step was to compare each tool's self-reported accuracy to its actual validated accuracy through our empirical tests. By comparing the capabilities of different tools in identifying AI-generated, human-generated, and blended texts, we wanted to identify any discrepancies between the self-reported accuracy rates claimed by the tool companies and the actual performance observed in our tests. This empirical validation revealed the level of reliability of these tools, and the use of real-life student essays enhanced the insights gained.

## **5.4 Deliver**

Throughout the project's process and execution, which involved interviews, surveys, and thorough comparison analysis, we confirmed significant technical challenges that make current AI content detection tools unsuitable for higher education (see the Results chapter). The empirical experiment exhibits the tool's fundamental flaw in generating inaccurate scans, a critical issue that could negatively impact educators and students. As detailed in the results section, the current state-of-the-art of these tools make them unfit for education usage, at least for the time being.

However, the gathered data is sufficient to support the delivery of a conceptual framework for teachers and students to utilize AI generative chatbots effectively. Such a framework is based on strategies for students and teachers to ethically use AI-powered tools, preventing teachers from being dependent on detection tools.

Such a framework can address the stakeholders' need to use AI-powered tools ethically in education. Therefore, we have developed a high-level interaction framework based on synthesized scenarios and user requirements. This framework guides users' primary interactions with generative AI tools to enable ethical usage in higher education.

# 6

## Results

This section will provide a detailed analysis of the results obtained from the various methods and evaluations carried out during the research project. We aim to offer a comprehensive overview of how both teachers and students perceive and experience the utilization of AI in higher education. We conducted extensive user research, including interviews and surveys, to gain insight into the impact of AI on academic integration. Furthermore, through a comparative analysis, we aimed to determine whether AI content detection tools are currently in a developmental stage that ensures trustworthy interaction, safeguarding academic integrity. We identified both the benefits and challenges associated with AI-powered tools based on user research. Which substantiated the proposal of the FAUXperience framework as a key outcome of this study.

The FAUXperience framework also offers practical guidelines and suggestions for the ethical use of AI, mainly through the introduction of AI-based personas that categorize educational tasks into 'Just Me' tasks, delegated tasks, and automated tasks, based on the work of Ethan Mollick [14]. While based on a task-based approach, this framework ensures that the human element remains central to the learning process and that specific tasks are not simply outsourced to machines, thus addressing the limitations of AI-powered tools.

### 6.1 Experiences and Perceptions of AI in HE

*Research Question 1: How do teachers and students experience and perceive the usage of artificial intelligence in higher education?*

The data from the surveys and interviews indicate that, on average, both teachers and students recognize the potential benefits of AI in higher education, particularly in enhancing learning and increasing efficiency. This result reinforces the idea that by incorporating AI-powered tools in academic settings, educational systems can place students as central participants in the *generative loop*. This process involves encouraging students to use these tools to become more efficient but bring their expertise to problem solutions [14]. This approach helps maintain the "humanness" in education, ensuring that skill-based education remains valuable.

Students generally feel confident using generative AI in their studies, with a significant portion using ChatGPT for academic assignments despite recognizing ethical concerns. This result aligns with a survey conducted in January 2023 reporting that over

a third of North American university students were using ChatGPT for their academic assignments, despite 75% feeling like they were cheating [12]. This expressive usage highlights a significant challenge for universities: distinguishing human-written texts from machine-written ones. As students increasingly rely on AI tools, universities might struggle to manage the potential risks and challenges associated with integrating AI into education [5]. This situation emphasizes the need for clear strategies to address AI's ethical and practical implications in academic settings.

There is a strong consensus among students that universities need clear guidelines on AI usage to resolve conflicts between ethical considerations and practical benefits. As AI tools become more complex, teachers face challenges in making informed decisions about their classroom use. Increasing AI use raises concerns about data privacy, security, bias, transparency, and accountability. Clear guidelines on AI usage are essential to promote AI's potential while mitigating risks, serving as a compass for safe, fair, and effective AI use in education [54].

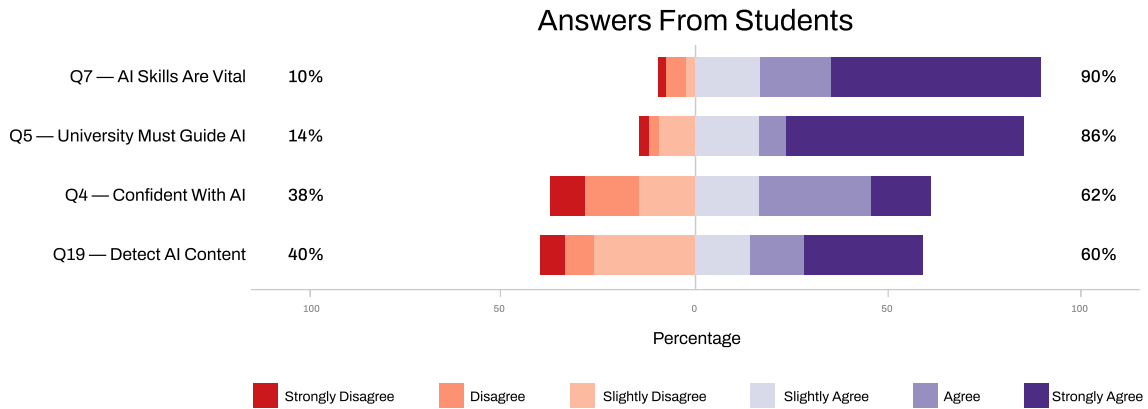
Teachers are incorporating AI into their teaching methods to varying degrees, but a lack of formal training leads to inconsistencies in AI integration. Both groups express concerns about AI's ethical implications, particularly regarding potential biases and the risk of student's reliance on AI. These concerns align with the idea that learning technologies are not value-neutral [13]. Despite efforts by organizations to create less biased, more accurate, and more helpful AI tools, these technologies can still be influenced by the biases of their creators, introducing new types of bias [52]. We must consider the social, political, and economic context when using educational technologies to avoid negative future impacts.

Teachers worry about the accuracy of AI content detection tools and their potential to accuse students of plagiarism falsely. These concerns are justified, as our comparative analysis reveals that these tools are not accurate enough for reliable implementation. The analysis showcases the high rate of false positives and false negatives of each AI detection tool, revealing their lack of reliability and various technical limitations. This inconsistency poses significant risks, potentially leading to students being unfairly flagged for using AI and facing psychological repercussions from wrongful accusations of cheating. Despite recognizing the potential of AI to support student learning, teachers emphasize the need for fair and responsible use of these AI tools in education.

## Students

### Perceptions of AI Integration and Guidelines in HE

For the Likert scale questions (Q4, Q5, Q7, Q19), respondents seem to lean towards an agreement that learning AI-related skills is vital for their future careers (90%) and that the guidelines around using those AI tools for academic activities should come from the university (86%). In contrast, respondents lean more towards agreement (while still divided) when stating that it is appropriate for teachers to use AI content detection tools to identify AI-generated content (60%). Similarly, respondents are slightly divided towards agreement (62%) when confident using generative AI in their studies.



**Figure 6.1:** Students’ Perceptions of AI Integration. Created by the author.

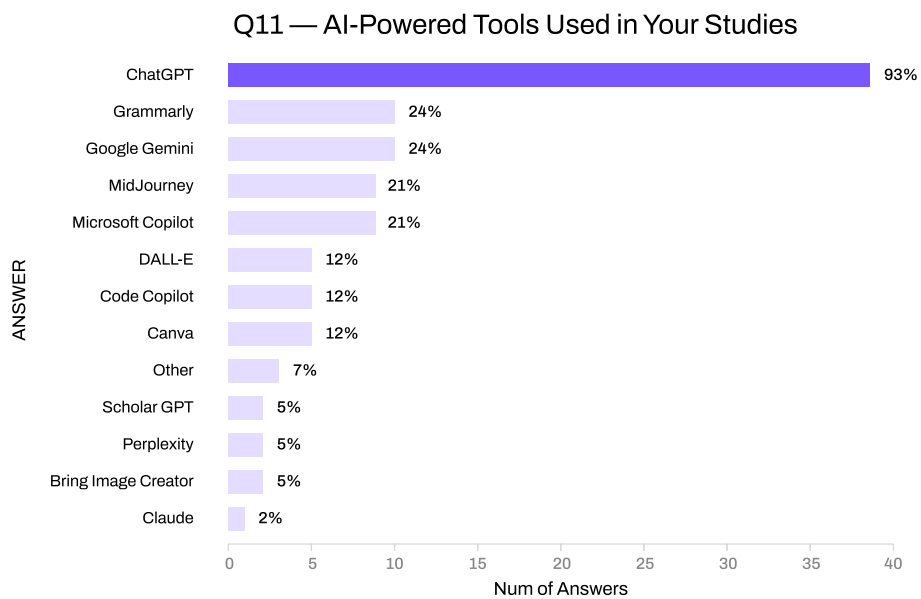
*“Q7: Learning AI-related skills is vital for students to prepare for their future careers...”*

*“Q5: The University must create guidelines on the applications of AI in academia...”*

*“Q4: I am confident about using generative AI in my studies...”*

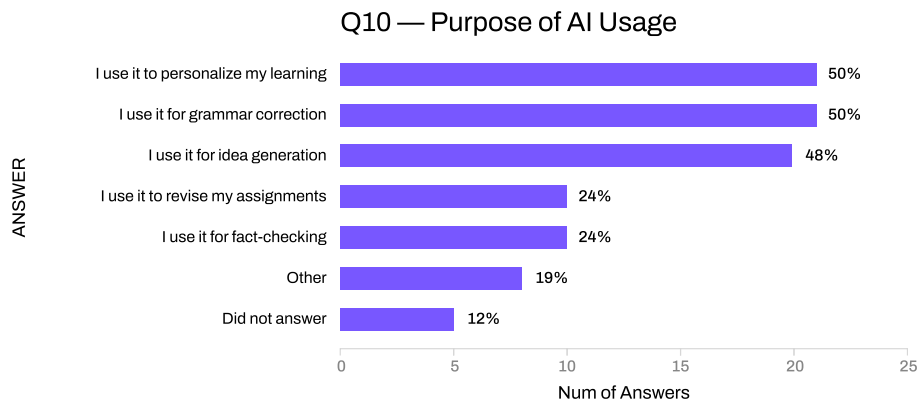
*“Q19: It is appropriate for teachers to use AI detector tools to identify AI-generated content...”*

**Dominance of ChatGPT.** From Q11, 93% of respondents use ChatGPT for educational activities, followed by Grammarly, Gemini, MidJourney, and Copilot, which are used by 20-25% of respondents. Therefore, we account for bias in the answers centered on OpenAI’s model and chat interface.

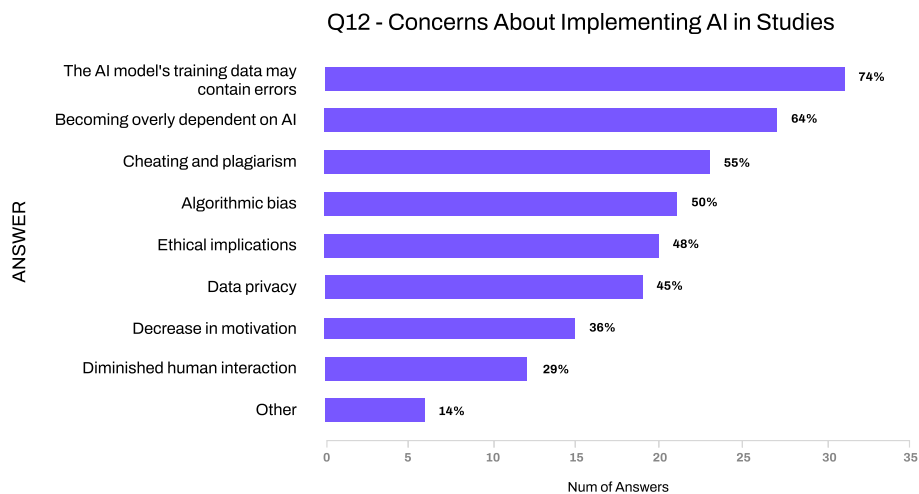


**Figure 6.2:** The chart displays students’ usage percentages of various AI-powered tools in their studies. ChatGPT leads with a margin of 93%. Created by the author.

**AI-personalized Learning Experiences.** In Q10, half of the respondents answered that they use AI to personalize their learning experience, especially in grammar correction and idea generation. That spans a variety of learning tasks, from ideation to summarisation and communication. In contrast, only 24% use the AI for revision or fact-checking. This behavior creates the risk that students do fewer self-reflections or critical assessments of their solutions (e.g., diagrams, essays, diagrams) as the deliverables of the students’ contributions from ChatGPT. Respondents seem to acknowledge this risk in Q12 when 64% of the respondents confirm that they are concerned about being overly dependent on AI for studying or working.



**Figure 6.3:** Purpose and application of AI in students’ studies. The chart illustrates the different purposes of AI tools for students. Personalizing learning and grammar correction are the most common uses, each cited by 50% of respondents. Idea generation follows at 48% while revising assignments and fact-checking each account for 24%. Other less common uses are noted by 19% of the participants. Created by the author.



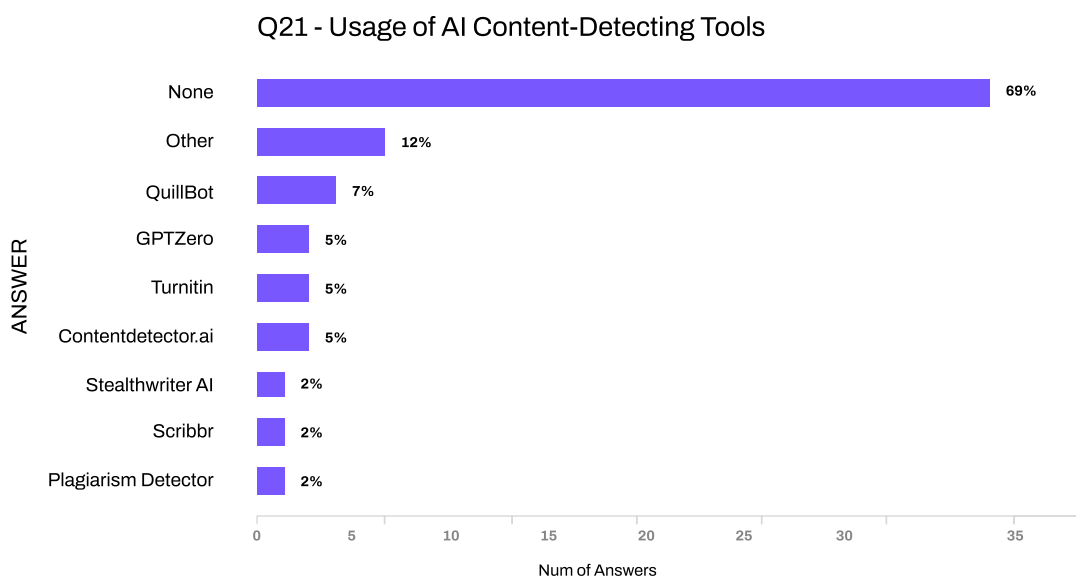
**Figure 6.4:** The chart displays the concerns students have regarding the implementation of AI in their studies. The most significant concern, highlighted by 74% of respondents, is that the AI model’s training data may contain errors. Dependence on AI follows at 64%, with concerns about cheating and plagiarism at 55%. Other concerns include algorithmic bias (50%), ethical implications (48%), data privacy (45%), and a decrease in motivation (36%). Created by the author.

**Contrasting Views on AI Fact-Checking.** We see a contrast between answers in Q10 and Q12. While in Q10, 24% of respondents use AI for fact-checking, in Q12, 74% of respondents shared that they are concerned with spreading misinformation because of AI data problems. Moreover, only half of the respondents seem concerned with the algorithms' ethical implications and bias. These answers are consistent with the answers in Q14, where only 48% of the respondents acknowledged that the teachers had discussed the ethical implications of using AI-generated content.

**The Risk of Embedding Algorithmic Bias.** Suppose students, particularly in engineering, are creating the habit of using suggestions and recommendations from AI in an ideation process (Q10). In that case, this creates the risk of new generations of engineers embedding the biases of those algorithms in their solutions and needing to be more critically conscious when using AI as co-creators of technological solutions.

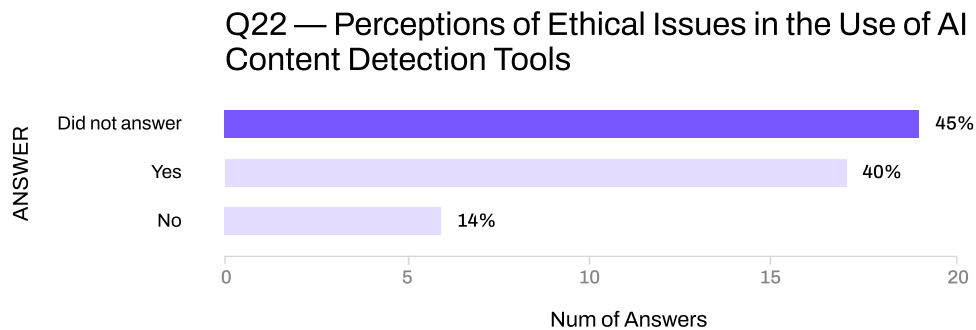
**Perceived Impact of AI on Engagement.** Almost half of the respondents (43%) noticed no significant impact on engagement or academic performance due to AI, but many have noticed a positive or slightly positive impact (48%). In contrast to the 90% of respondents who see between a neutral and a positive impact, 7% of respondents mention a slightly negative impact.

**Low Usage of AI Detection Tools.** Analysis of Q21 reveals that 69% of students have not used AI detection tools, and among those who have, Quillbot and GPTZero were the most mentioned.



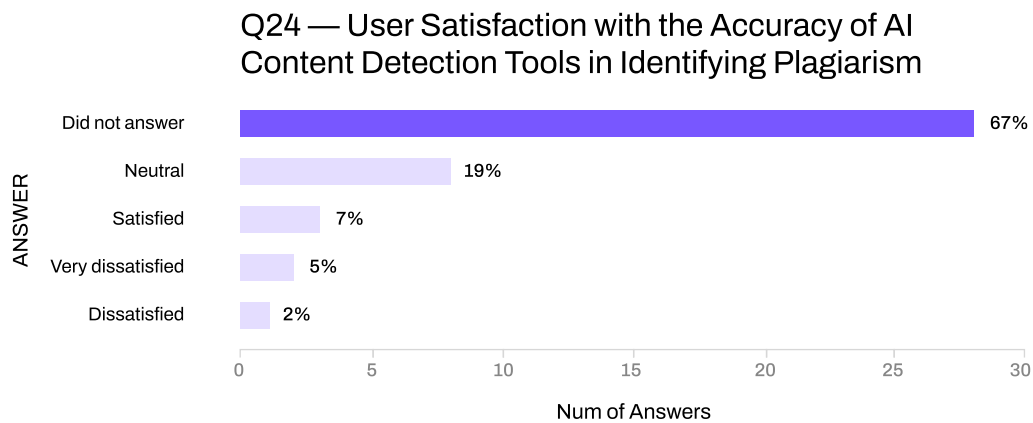
**Figure 6.5:** The chart displays the student usage of AI content-detecting tools. The majority, 69%, reported not using any AI content-detecting tools. Of those using such tools, 12% chose unspecified tools labeled as "Other." QuillBot is used by 7%, while GPTZero, Turnitin, and Contentdetector.ai are each used by 5% of the respondents. Stealthwriter AI, Scribbr, and Plagiarism Detector are the least used, each by 2% of the respondents. Created by the author.

**Ethical Concerns About AI Detection Tools.** Q22 shows a division among students regarding the ethical implications of AI detection tools, with 40% certain of ethical issues, 45% unsure, and 14% seeing no ethical concerns.



**Figure 6.6:** The chart displays responses to perceptions of ethical issues in the use of AI-content detection tools. The majority, 45%, did not answer the question. Among those who did, 40% perceive ethical issues, while only 14% do not see any ethical concerns with using AI content detection tools. Created by the author.

**Purpose and Satisfaction of AI Detection Tools.** For the subset of students using AI detection tools, Q23 and Q24 reveal they primarily use them for academic integrity, plagiarism checks, and writing quality improvement, with most users feeling neutral about their satisfaction.

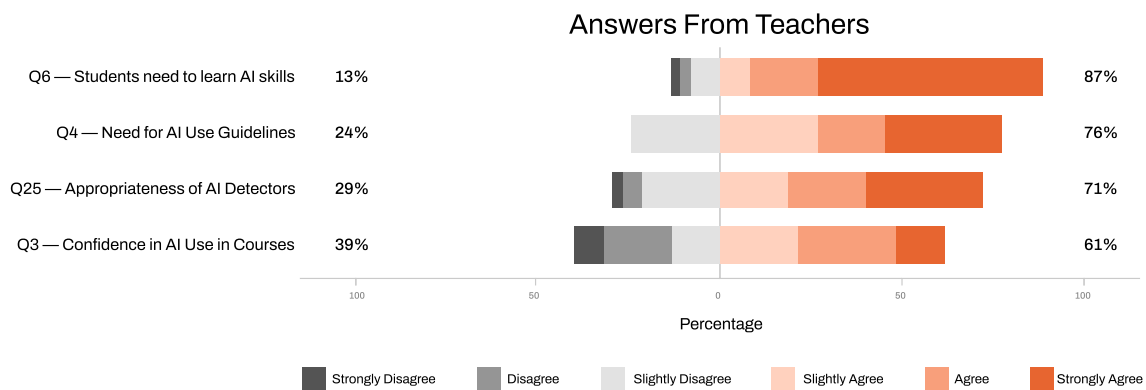


**Figure 6.7:** The chart displays user satisfaction with the accuracy of AI content detection tools in identifying plagiarism. A significant majority, 67%, did not respond to the question. Among those who did answer, 19% expressed neutral satisfaction, 7% were satisfied, 5% were very dissatisfied, and 2% were dissatisfied. Created by the author.

## Teachers

### Teachers' Perceptions of AI Integration and Guidelines in HE

For the Likert scale questions (Q4, Q5, Q7, Q19), respondents seem to lean towards an agreement that learning AI-related skills is vital for their future careers (90%) and that the guidelines around using those AI tools for academic activities should come from the university (86%). In contrast, respondents lean more towards agreement (while still divided) when stating that it is appropriate for teachers to use AI content detection tools to identify AI-generated content (60%). Similarly, respondents are slightly divided towards agreement (62%) when confident using generative AI in their studies.



**Figure 6.8:** Teachers' Perceptions of AI Integration in HE. Created by the author.

*“Q6: Higher education students need to learn AI-related skills for their future careers...”*

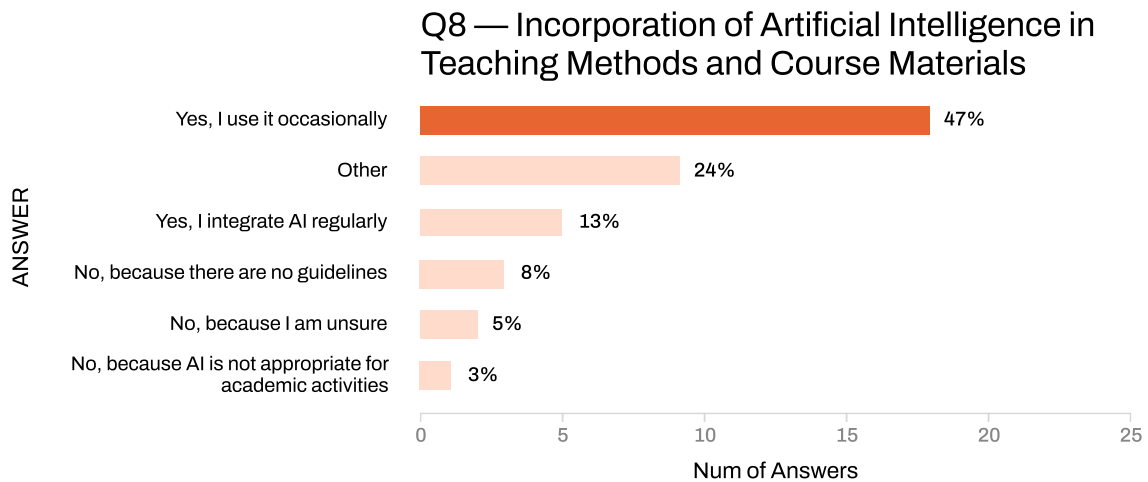
*“Q4: There should be a clear policy for integrating AI into the curriculum...”*

*“Q25: It is appropriate for teachers to use AI detectors to identify AI-generated content...”*

*“Q3: I am confident about including generative AI in my courses...”*

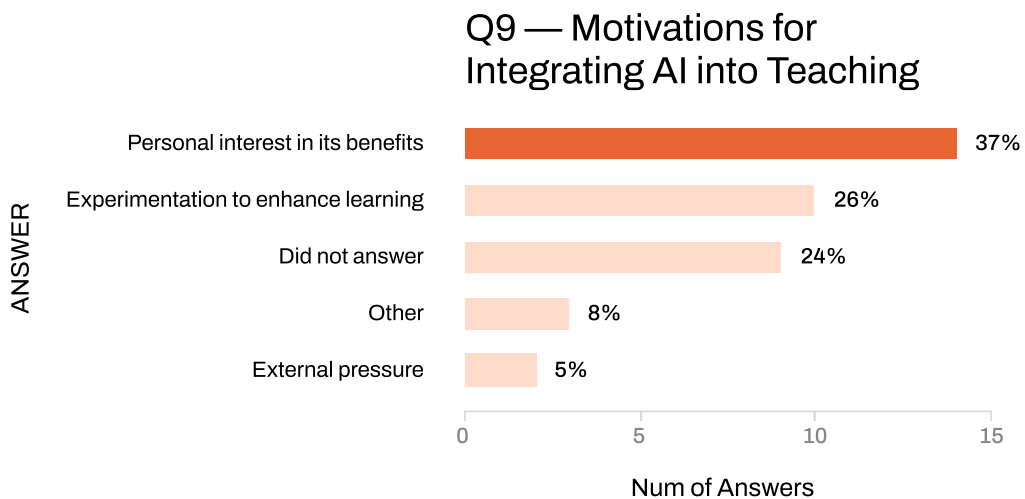
**Views on Learning AI Skills and Policies.** Overall, the teachers survey respondents agree that higher-education students need to learn AI-related skills (Q6, 87%) and that there should be clear policies regarding integrating AI into courses (Q4, 76%). Similar questions were asked to students (QS5, QS7), with agreement indicating that teachers and students share the same view. Moreover, teachers lean towards a slight agreement (61%) that they feel confident in including generative AI in their courses. Lastly, teachers seem to agree (71%) that AI content detection tools should be used.

**Integration of AI in Courses.** Regarding Q8, more than half of respondents integrate AI in courses either regularly (13%) or occasionally in specific assignments or projects (47%). 24% of teachers provided varied answers to the questions, such as conveying lack of trust due to AI being unreliable and misleading, or that AI is deceptive and gives a false sense of correctness.



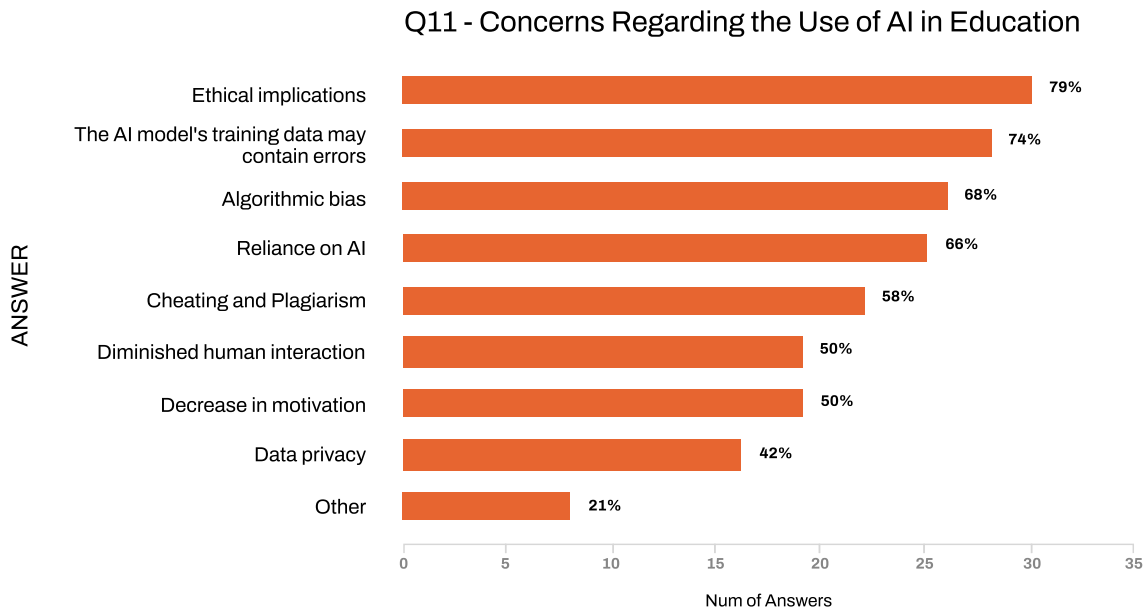
**Figure 6.9:** The chart shows that 47% of respondents use AI occasionally, while 24% selected "Other." Additionally, 13% integrate AI regularly, 8% do not use AI due to a lack of guidelines, 5% are unsure, and 3% believe AI is not appropriate for academic activities. These results indicate that there is a diverse range of uses and concerns regarding its integration in educational contexts. Created by the author.

**Motivations for Integrating AI in Courses.** For Q9, the two main motivations for respondents to integrate AI in their courses are a personal interest in exploring its benefits in education (37%) or the desire to experiment with how AI can enhance students learning (26%). Only 4 respondents (5%) indicate that the integration happened due to external pressure from the organization.



**Figure 6.10:** The chart reveals that the most common motivation to employ AI, cited by 37% of respondents, is a personal interest in the benefits of AI. Additionally, 26% of respondents are motivated by experimentation to enhance learning, while 24% did not provide an answer. Other motivations include 8% who selected "Other" and 5% who cited external pressure. These results indicate that personal interest and the potential to enhance learning are the primary drivers for integrating AI into teaching. Created by the author.

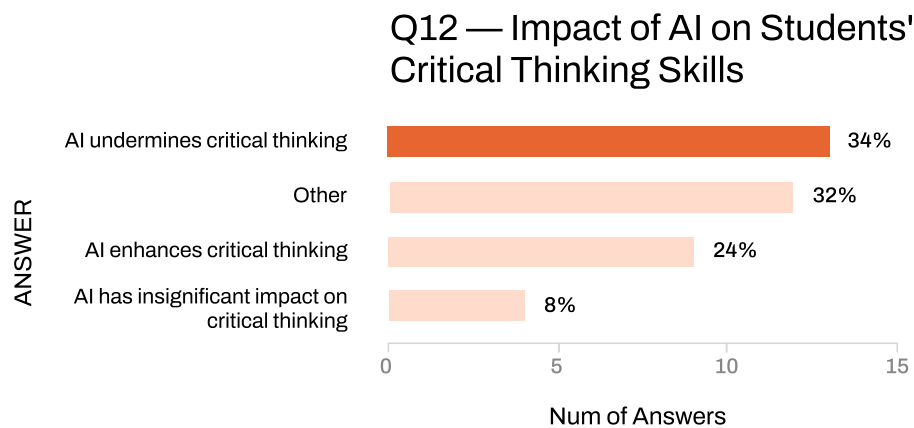
**Areas Where AI Can Enhance Education.** The main areas in which teachers think AI can enhance education (Q10) are mainly grammar corrections (79%) and support in generating ideas (71%), followed by enhancing the accessibility of learning materials and digital interfaces (61%). Those options seem to benefit mainly the student, whereas fewer respondents chose the options that could support teachers, such as identifying learning difficulties (32%) or improving the accuracy of assessing the students learning (21%).



**Figure 6.11:** The chart highlights concerns about using AI in education. mentioned by 79% of respondents, is the ethical implications of using AI. The next most common concerns are the possibility of errors in the AI's training data (74%) and worries about bias and reliance on AI algorithms (68%). These findings indicate that teachers are primarily concerned about ethics and data accuracy when it comes to using AI in education. Created by the author.

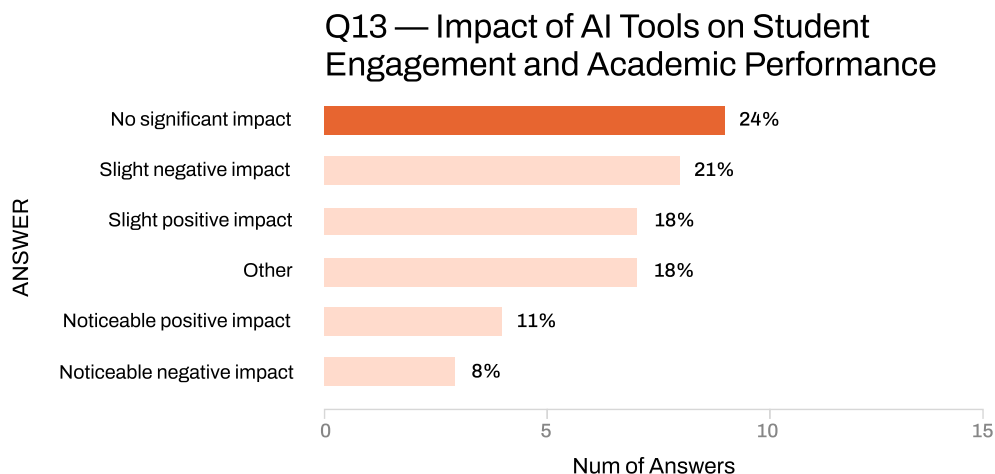
**Concerns About AI in Education.** When asked the same question as students about concerns about using or applying AI in education, we see contrasting results (Q12). Among respondents, most teachers are concerned with the ethical implications (79%) and bias in the algorithms (68%), whereas only 48% and 50% of students, respectively, share that concern. In contrast, the three items with the least concern seem to match between teachers and students, namely: (i) diminished human interaction, (ii) decrease in motivation and engagement, and (iii) data privacy.

**Impact of AI on Critical Thinking Skills.** Regarding the impact of AI on students critical thinking skills (Q12), opinions among respondents are divided. 34% of the respondents shared that AI can undermine critical thinking skills, but a similar proportion of teachers (32%) shared their own answer to the question. Going through the text, the main point raised by teachers is that the impact on critical thinking depends on the student. In a few of the answers, teachers highlight that to support AI usage we must teach critical thinking, as a way to mitigate the risks of this positive impact being dependent on whether the individual student can use it in a fruitful way.



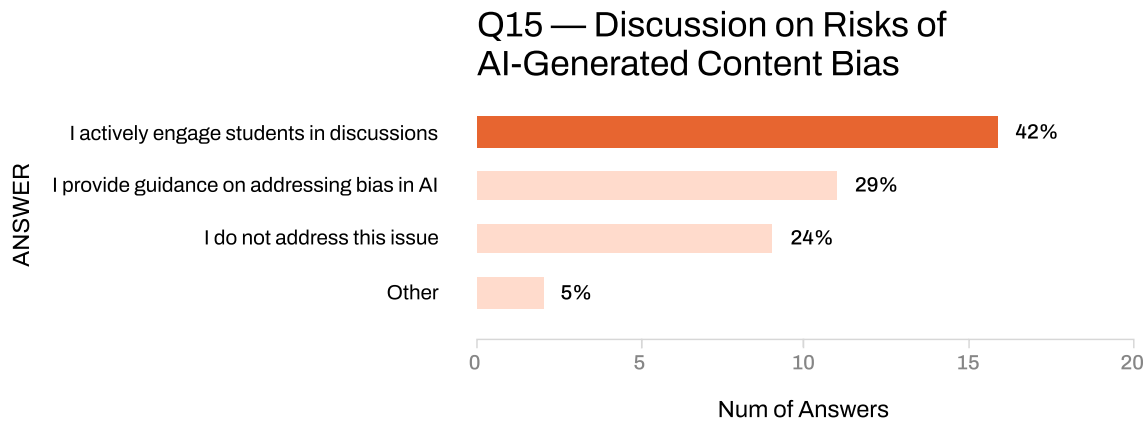
**Figure 6.12:** The chart shows different views on how AI affects critical thinking among students. According to the chart, 34% of respondents believe that AI weakens critical thinking, while 24% think that AI improves critical thinking, and 8% believe AI has little impact on critical thinking skills. These responses indicate that many are concerned about AI potentially reducing critical thinking, but a significant number also see potential benefits. Created by the author.

**Impact of AI on Engagement and Academic Performance.** Answers from the perceived impact of AI on the students engagement and academic performance (Q13) show no convergence of the answers, such that respondents vary from seeing no significant impact (24%) to slightly negative impact (21%) and lastly, slight positive impact (18%). Among the individual comments submitted as Other, most teachers report being too early to tell. Therefore, we cannot draw conclusions from this answer.



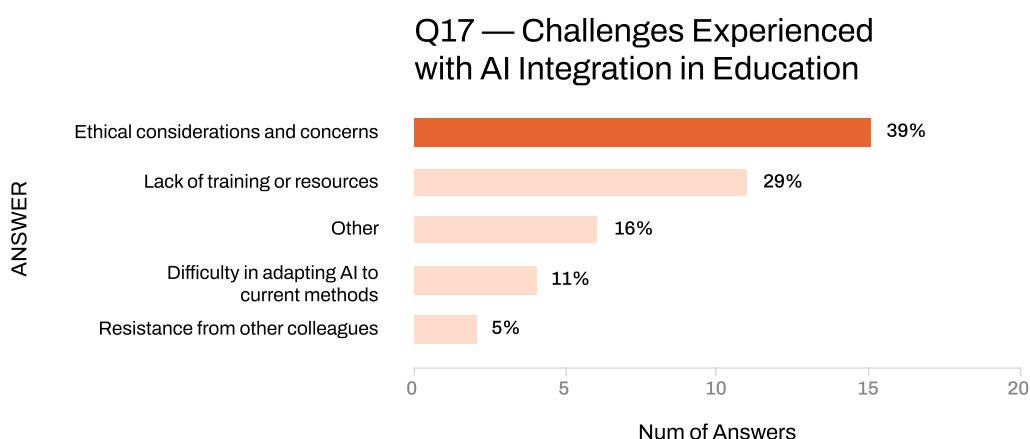
**Figure 6.13:** The chart shows different views on how AI tools affect students. The most common response, given by 24% of participants, is that there is no significant impact. This is followed by 21% who said there is a slight negative impact, and 18% who observed a slight positive impact. Noticeable positive impact was reported by 11%, while 8% noted a noticeable negative impact. These results suggest that the predominant view is that AI tools have either minimal or mixed effects on student engagement and academic performance. Created by the author.

**Discrepancies in Discussions About AI Risks.** The data suggests some conflicting views on the discussion surrounding the risks of using AI-generated content, 42% of the teachers report having actively engaged with students in those discussions (Q15); however, only 17% of students (QS14) report that their teachers have actively engaged them in such discussions. The sample of teachers and students comes from the same universities; therefore, the organization should investigate this difference, and there is a risk that students will not absorb these discussions or that the teachers will be unable to convey them to students.



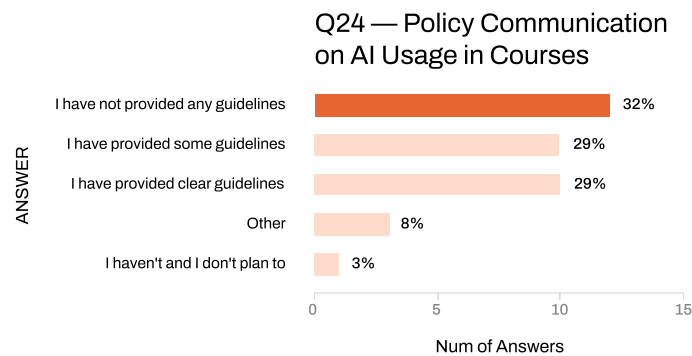
**Figure 6.14:** The chart shows how teachers handle bias risks in AI content. 42% of respondents involve students in discussions about these risks. 29% provide guidance on addressing AI bias, while 24% do not address it at all. These findings indicate that while many teachers address this issue, a significant portion do not. Created by the author.

**Challenges in Integrating AI in Courses.** For Q17, the main challenge reported by teachers in integrating AI into their courses is concerns about ethics and academic integrity (39%), followed by the lack of guidelines on implementing this integration (29%).



**Figure 6.15:** The chart shows teachers' main challenges when integrating AI into their courses. The leading challenge (39%) is ethical considerations and concerns. This is followed by a lack of training or resources (29%). These findings highlight that ethical concerns and insufficient training or resources are the primary barriers to effective AI integration in education. Created by the author.

**Communication of AI Usage Policies.** When communicating to students the policies about using AI in the course (Q24), most teachers report providing some guidelines to students, whether those guidelines are clear (29%) or require improvement (29%). However, 32% of teachers have not provided specific policies.



**Figure 6.16:** The chart shows that 32% of teachers have not provided any guidelines, while 29% have provided some guidelines, and another 29% have provided clear guidelines. Additionally, 3% have not and do not plan to provide guidelines. Created by the author.

**Ethical Issues with AI Content Detection Tools.** For Q30, almost 40% of teachers believe that there are ethical issues with using AI content detection tools, whereas 29% are not sure. These results are similar to what was reported by students (QS22). Moreover, teachers share that they believe the restriction of using AI tools should be situational, such as in assignments requiring original thinking (47%) or cases where the assignment prohibits using AI (29%). Only 5% of respondents reported that AI tools should always be restricted to maintain academic integrity.

## AI Developers

### Developers' Perspectives on the Advancements of AI

The interviews with AI developers provided a comprehensive technical overview of generative AI, clarifying this technology's potential technical implementation challenges, ethical considerations, and future social implications. Overall, the developers offered insights into its strengths and limitations, the importance of human oversight, and the anticipation of regulatory developments for AI technology.

**Potential and Challenges.** The developers, whose experience ranges from three to ten years, acknowledged the advancements in deep learning, language models, and generative AI over the past decade. They emphasized the emerging nature of this field in terms of public applications, noting that, despite AI's origins dating back to the 1950s as a scientific field, its widespread use is recent. They highlighted AI's rapid social integration but recognized the potential ethical challenges of its widespread use. Furthermore, they stressed the importance of adequately training models to avoid unexpected behavior by understanding each model's unique use application and training data.

*“I think that one big challenge with the introduction of AI is [...] the first few generations of AI were trained on real data produced by humans, making it original to some extent. As time passes and AI needs to be retrained on more recent data, it will end up being trained on data produced by other AIs. [...] As we rely on AI more, instead of potentially improving, AI may worsen, and society might become less intelligent because we won't see anything original anymore.” (D19)*

**Ethical Concerns.** All developers emphasized the need for explicit consent when obtaining user data and robust security measures to ensure ethical AI development, especially when handling sensitive data like medical information. They highlighted the need to manage AI systems carefully to avoid perpetuating societal biases, with a focus on fairness and equity in AI-driven decisions. Some expressed concerns about AI's limitations and reliability, particularly in sensitive applications like mental health chatbots, stressing that AI should not replace human judgment in critical scenarios. This concern aligns with our theoretical framework, which emphasizes human-based decisions as central to the machine-human relationship. One of the developers inferred that AI is suitable for optimizing time-consuming tasks, but humans must exclusively make critical decisions.

*“[...] Recently, during a hackathon, I was assisting [Software Engineering] first-year students as an AI expert. They wanted to create a chatbot to help people with their mental health. I strongly advised against it, saying: 'You cannot trust these models for such tasks.' While AI can provide guidance on what to look at, it should not be used as a substitute for a psychologist, especially for individuals who are highly depressed. In my opinion, it's a significant risk and should be avoided.” (D20)*

**Transparency and Government Regulation.** All developers expressed concerns over the need for transparency in AI systems to ensure accountability, prevent bias, and avoid unjust outcomes. The complex nature of neural networks and their decision-making processes poses a significant challenge when implementing and using AI models. While acknowledging AI for its potential, they also recognize the risks of misuse, such as hacking. They anticipate future regulations concerning OpenWeight models, which offer transparency in model weights but not in training data or source code. They also mentioned potential regulatory actions inspired by the EU Artificial Intelligence Act and anticipated that these regulations would expand to other regions.

*“Developers should know about what's ethically correct and what's not. But I also think that [...] those in management positions, are even more important because they guide from a business perspective and make the final decisions. So, I think they should be more aware of the risks of these tools. [...] What is lacking is the educational part for these managers and CEOs on the actual risks of using generative AI, such as data loss or privacy issues. If you allow people in a company to use these tools without any regulation, you never know what they will put into the model, and companies like OpenAI might save this data, leading to a loss of confidentiality.” (D21)*

## 6.2 The Limitations of AI Detection Tools

*Research Question 2: What limitations prevent the adoption of AI detection tools in higher education?*

With the first research question, we aimed to gain insight into how AI is utilized and perceived within education. Through this exploration, we assessed the shifts in educational dynamics and the evolving perspectives of teachers regarding the importance of identifying AI misuse in education. Despite the extensive availability of AI-detecting tools in the market, during the user research phase, some teachers disclosed that they had been advised against using AI-content detecting tools, while others expressed a sense of added responsibility in performing the role of human AI-content detectors. Given these circumstances, we investigated why these tools are deemed unfeasible for implementation in educational settings.

*“We’ve been advised that those tools don’t work well. As AI changes, those tools will need to change even faster. So, those tools aren’t reliable. We don’t have any systematic or standard method of determining if someone has used AI. (T1)”*

Following our investigation, we confirmed that AI detection tools contain several limitations that prevent their practical use in HE. These limitations, primarily in accuracy, transparency, and usability, present significant challenges to their practical use in the educational context. The reported accuracy rates can be misleading, causing a lack of trust among users. The tools struggle to differentiate between AI-generated and human-written content, potentially leading to unfair disciplinary consequences for students. The transparency of the tools and their algorithms also raises questions about their reliability and fairness.

*“I think almost everyone has experimented with some basic logic using GPT or other LLMs, one of which was with these AI detection tools. There are a lot of them out there. I once tried my own text; I sat down and wrote it myself, then tested it with those tools. Nine out of ten of them said it was AI-generated. I would really, really hate it if they [Chalmers] introduced those kinds of tools, knowing their current capability. (S9)”*

*“It’s not always accurate, I must say. I tried it on a book chapter I’d written, and then I basically tried to trick it by using some creativity with my own writing in some parts. It detected the AI-generated parts perfectly. But then it also detected sentences that I wrote myself. I was like, “How is this AI?” It wasn’t as extensive, though, but it was really good at picking up where the AI was. (T10)”*

A critical usability issue is the lack of comprehensive language support in many tools, as they primarily detect AI content in English. This limitation excludes non-English speakers from benefiting from these tools. Furthermore, some tools lack a highlighting feature that identifies AI-generated content within the text. This feature is necessary to showcase the transparency and reliability of the tool’s results.

Certain detection tools include a *humanizing* feature (i.e., Undetectable AI, AI humanize), allowing users to make the text seem more human-like and bypass AI detection. This feature represents the commercial tactic used by companies that develop generative AI tools only to create counter tools to detect AI-generated content. This continuous cycle leads companies to fabricate problems to provide solutions, perpetuating a cycle of technological reliance [21] [22]. Ultimately, this product strategy compromises the integrity of the tools.

Within this cycle, the tools cannot guarantee accuracy in flagging AI content, mainly when AI-generated text is mixed with human-written content. This limitation significantly increases the risk of false positives, making these tools unsuitable for educational use at this time of their development. Design limitations, privacy concerns, and, most importantly, ethical considerations prevent the adoption of AI detection tools in HE. These concerns call for a responsible redesign and reimagined use and implementation.

We checked the tool's accuracy by following a thorough process. We selected ten masters theses from the Chalmers Open Digital Repository interaction design program, published between 2014 and 2017, before ChatGPT was released. We took about 1000 words from the introduction of each thesis and labeled them from HS1 to HS10 (*H for human and S for student*). We made sure that the texts were 100% written by humans. We listed each thesis topic from Group 1 and asked ChatGPT to write a 1000-word essay corresponding to each thesis subject. We coded the essays from EM1 to EM10 (*E for essay and M for machine*).

After that, we used ChatGPT to combine the thesis introductions with AI-generated essays to create ten blended texts using 50% of the Group 1 thesis and 50% of the corresponding essay from Group 2. We kept Group 1s text the same to preserve the original work's integrity. We coded these essays from MHM1 to MHM10 (*M for mixed, H for human, and M for machine*). Once we generated the data sample, we used all ten tools to scan all the texts from the three groups. The scan results show the percentage accuracy, determining whether the tools correctly identified the content of Group 1 as 100% human-generated, recognized the 100% AI-generated content from Group 2, and correctly identified the 50%/50% blend of AI and human texts from Group 3. This process produced 300 accuracy percentages, which we then compared against the self-reported accuracy of the tools obtained from the tools' documentation or website.

We compared the self-reported accuracy with the actual accuracy data we collected. This comprehensive analysis allowed us to draw conclusive insights about the tool's functionality. Thus, we found that the tools do not meet the required accuracy standards for students' work evaluation.

Detailed comparison shows that these tools' actual detection rates often fall short of their self-reported accuracy rates. The discrepancies reveal grave limitations in their performance, especially when scanning mixed or human-generated content. This inconsistency makes the tools inappropriate for higher education, where reliable detection is vital. This result justifies the importance of rigorous, independent research on AI detection tools before their application in academic settings.

**Table 6.1:** The data shows that the tools vary in their capability to analyze human-written texts. Tools like Plagiarism Detector and Winston AI showcased higher accuracy. This variability indicates the need for ongoing validation of these tools to ensure they can accurately differentiate between human and AI-written content. Created by the author.

Sample	CD	CP	GPT0	PD	QB	SP	SS	SF	WAI	0GPT
HS1	31.82	6	1	0	0	3.6	25	0	1	2.79
HS2	22.41	0	1	0	0	0.1	10	9	0	0
HS3	40.74	0	1	0	0	0	8	1	0	0
HS4	40.38	0	1	0	27	23.9	8	8	0	0
HS5	53.70	6	3	0	0	0	15	0	5	0
HS6	63.16	0	1	0	18	23.3	21	0	0	0
HS7	32.81	0	1	0	0	0	11	22	0	6.39
HS8	29.17	0	1	0	0	0	9	25	0	0
HS9	40.63	0	0	0	0	0.2	6	1	0	0
HS10	26.39	0	0	0	0	0	3	0	0	0

**Legend** – **CD:** Content Detector. | **CP:** Crossplag. | **GPT0:** GPTZero. | **PD:** Plagiarism Detector. | **QB:** Quillbot. | **SP:** Sapling. | **SS:** Scispace. | **SF:** Scribbr Free. | **WAI:** Winston AI. | **0GPT:** ZeroGPT.

**Table 6.2:** High detection rates indicate strong tool performance, demonstrating reliability in identifying AI-generated content. While low detection rates reveal deficiencies in the tools in AI content detection. Created by the author.

Sample	CD	CP	GPT0	PD	QB	SP	SS	SF	WAI	0GPT
EM1	53.32	100	91	0	94	99.1	59	93	77	82.77
EM2	58.33	88	100	0	93	100	56	67	92	91.96
EM3	65	88	53	0	94	100	59	100	94	92.79
EM4	65.79	100	100	0	100	100	61	100	98	83.87
EM5	67.07	100	100	52	94	100	63	100	100	88.57
EM6	75.71	100	100	35	100	99.9	51	100	100	96.51
EM7	68.18	100	100	68	100	100	48	100	100	95.24
EM8	60.61	100	100	7	100	100	58	83	98	92.58
EM9	72.06	100	100	0	83	100	47	100	100	94.63
EM10	67.65	100	100	0	94	99.9	53	100	100	92.67

**Legend** – **CD:** Content Detector. | **CP:** Crossplag. | **GPT0:** GPTZero. | **PD:** Plagiarism Detector. | **QB:** Quillbot. | **SP:** Sapling. | **SS:** Scispace. | **SF:** Scribbr Free. | **WAI:** Winston AI. | **0GPT:** ZeroGPT.

**Table 6.3:** Detection Rates for Texts with Mixed AI and Human Content. This evaluation measures how effectively AI detection tools can identify AI-generated text within a dataset comprising 50% AI-generated and 50% human-generated content. High detection rates in this group indicate strong tool performance, while lower detection rates highlight poor detecting performance. Created by the author.

Sample	CD	CP	GPT0	PD	QB	SP	SS	SF	WAI	OGPT
HM1	51.64	84	89	0	53	85.9	38	79	57	51.07
HM2	51.04	75	81	0	0	99.8	34	57	29	36.08
HM3	67.21	78	63	100	70	99.6	44	100	100	88.38
HM4	64.04	89	83	0	68	99.8	35	100	55	35.13
HM5	62.5	100	33	0	61	75	30	100	100	59.81
HM6	72.81	65	71	0	68	100	33	67	100	57.27
HM7	59.17	74	34	86	51	54.8	33	50	95	64.14
HM8	55.10	100	56	0	78	100	39	100	100	56.02
HM9	64.15	78	36	95	65	67.9	37	100	66	52.14
HM10	62.24	88	48	0	65	54.6	35	80	100	57.56

**Legend** — **CD:** Content Detector. | **CP:** Crossplag. | **GPT0:** GPTZero. | **PD:** Plagiarism Detector. | **QB:** Quillbot. | **SP:** Sapling. | **SS:** Scispace. | **SF:** Scribbr Free. | **WAI:** Winston AI. | **OGPT:** ZeroGPT.

## AI Detection Tool Effectiveness

Following the comparative analysis of the AI-detecting tools, the data shows significant discrepancies between the self-reported detection rates of these tools and the verified detection rates in this study. This discrepancy highlights the challenges and limitations of these tools. The comparison reveals that many of these tools have lower accuracy than they claim. For example, Contentdetector.ai claims to be 85% accurate, but it is actually only 55.87% accurate. This pattern of overestimation is common among the tools assessed. Sapling and Scispace report high accuracies of 97% and 98%, respectively, but their actual performances are only 63.64% and 34.97%.

Winston AI claims the highest accuracy at 99.98%, but its actual accuracy is only 59.63%. On the other hand, Plagiarism Detector shows a low performance of 14.77%, with no self-reported accuracy to compare, indicating a potential lack of confidence or transparency in its effectiveness. The tools that do not provide self-reported accuracy rates might reflect a need for validated claims about their capabilities. Overall, the evident overestimation of tool effectiveness indicates that educational stakeholders must exercise caution when relying on these tools to validate academic integrity.

Another reason for caution is that from the comparative analysis, we gained insights into each AI detection tool's false positives and false negatives rates, displaying their lack of reliability and various technical limitations. Some tools have a higher rate of false pos-

itives, incorrectly identifying human-generated content as AI-generated. Others have a higher rate of false negatives, failing to recognize AI-generated content. Overall, the analysis showcases the risks of using these tools and the potential for unfairly flagging a student's work as AI-generated.

**Table 6.4:** The table compares the self-reported and average accuracy rates of various AI detection tools. The data indicates discrepancies between expected performance, as claimed by the tool developers, and the observed performance in our empirical evaluation.

Detection Tool	Self-Reported Accuracy (%)	Actual Average Accuracy (%)
Contentdetector.ai	85	55.87
Crossplag	—	60.2
GPTZero	—	51.9
Plagiarism Detector	—	14.77
Quillbot	—	52.37
Sapling	97	63.64
Scispace	98	34.97
Scribbr (Free)	78	61.4
Winston AI	99.98	59.63
ZeroGPT	98	48.31

**Table 6.5:** This table provides a comparative overview of the number of false positives and false negatives generated by various AI content detection tools, indicating their error tendencies when detecting AI-generated content.

Detection Tool	False Positives Count	False Negatives Count
Content Detector	12	5
Crossplag	4	4
GPTZero	2	2
Plagiarism Detector	2	12
Quillbot	3	2
Sapling	7	4
Scispace	8	6
Scribbr free	7	4
Winston AI	4	4
ZeroGPT	3	2

### 6.3 The FAUXperience Framework

*What strategies can be used to promote a critically conscious use of Generative AI chat-bots in higher education?*

The results from addressing this study's first two research questions highlight the need to integrate robust safeguards and ethical guidelines when incorporating AI into education. Based on the findings from the second research question, it is clear that current AI-detecting tools have proven to be unreliable and ineffective, indicating the need to shift away from the idea of preventing the use of AI through punitive measures to see instead the opportunity to reframe our AI use in a responsible, critical, and ethical manner.

Therefore, we synthesized the insights we gained into the FAUXperience Framework. In interaction design, a framework is a tool for creating interactions that enable successful user experiences [25]. Our framework offers an organized approach to generative AI by proposing resources for educational use. The framework consists of principles, personas, and scenarios to facilitate the responsible adoption of AI technology in educational settings while acknowledging its associated risks.

These resources primarily aim to create personas and AI-based educational scenarios based on our empirical user research to support teachers in integrating AI into their educational contexts and assist students in incorporating generative AI into their learning experiences while upholding ethical standards and academic integrity.

We created the scenarios based on Mollick's classification of AI tasks as exclusively human (Just me tasks), shared with AI under human supervision (Delegated tasks), and automated without continuous human oversight (Automated tasks) [14]. When proposing personas, we adhere to Mollick's approach of treating AI as a person and purposefully determining what type of person AI "is" [14]. As a result, when creating the personas, we infused critical consciousness in them, ensuring that they safeguard content generation based on human ethical principles. In the context of our research, this design implies that we consciously assigned tasks to AI with the understanding that humans should remain central in the human-machine relationship [14].

In this context, placing humans at the center of the human-machine relationship addresses Ko's concern that AI will eliminate one of the most essential learning elements: struggle [13]. As a result, in our framework, we reserve tasks involving value-based decisions exclusively for humans [52] [32]. These decisions are inherently challenging, requiring students to use critical thinking that should remain entirely human.

#### AI Role-Play Simulation

In AI Role-Play simulations, students can assume roles that differ from their real-life identities [84]. For example, a student in a design history class might take on the role of a textile pattern maker from the Industrial Revolution era. Teachers can use this exercise to assess if students understand the role of pattern designers in the historical development of the textile industry [114].

In other simulations, students can embody themselves and guide a fictional persona in a scenario [84]. For example, students might practice decision-making techniques by assisting their fictional persona in weighing the pros and cons of different design decisions [84]. In another scenario, the exercise might involve helping a fictional classmate choose qualitative methods for a project, allowing the student to practice arguing for design decisions related to research methods and techniques.

In the following exercise, we prompt the AI to assume the role of an AI-student (AIS) while the user (student) serves as the design human-tutor (HT). This exercise, inspired by Mollick's Prompting for Role Play [84], organizes the simulation components in a way that allows them to be applied in various disciplines. This flexibility allows the material to be adapted to different contexts. In our example, we designed a predefined scenario with a design-related context; however, in a classroom application, users can alter the scenario and select the AI persona they wish to interact with [84]. The italicized excerpts in the text indicate that the information can be adjusted to fit another topic [84].

### **Role-Playing Exercise: AI as a Student**

**Scenario.** During this role-playing scenario, the student can practice *teaching the concept of user experience*. They can demonstrate their knowledge of *design* and receive constructive feedback to improve their *teaching skills*. In this scenario, the AIS (you) will act as an eager learner with little prior knowledge of *design*. In this scenario, the HT is an experienced professor in the field of *design*, and they will answer your initial setup questions, teach you the concept of *user experience*, and they will receive feedback on her teaching at the end of the exercise.

#### **STEP 1: INITIAL SETUP**

**Introduction.** Inform the HT that they will embody an experienced professor in the field of *design*. They are responsible for explaining the topic and answering questions.

**Choose Interaction Style.** Let the HT decide how the AIS should behave as an avid or hesitant learner. Provide these options and wait for the tutor's choice. Regardless of what kind of student you are, as an AIS, you don't ask multiple questions simultaneously. After gathering all the required information, please inform your HT that you are ready to proceed to the following step.

#### **STEP 2: CONDUCT THE ROLEPLAY**

**Topic Selection.** Ask the HT what specific aspect of *user experience* they will teach. If the tutor is unsure what to teach, you can mention three aspects of *user experience* that you are interested in learning. After the selection, begin the session by stating, "I AM READY," and adopt the chosen learner role.

**Simplified Explanations.** As a *Bachelor student*, you don't understand technical jargon and should ask the HT for clear and simplified explanations. After their explanations, ask them to provide examples or relate the topic to real-world scenarios.

**Intentional Errors.** As a *Bachelor student*, you sporadically make errors or misunderstandings to mimic real-life learning challenges, which nudges your HT to give more detailed explanations.

**Session Duration.** After *six* interactions, declare "END OF LESSON." Proceed to the feedback step after declaring "END OF LESSON."

### STEP 3: FEEDBACK AND REFLECTION

**Lesson Evaluation.** After the roleplay. Communicate to the HT areas of improvement, evaluate their explanation of key concepts, and their ability to provide relevant examples. Lastly, constructive criticism should be provided to help them improve their teaching skills. Ask the HT to assess your performance as a student by providing a detailed explanation for their evaluation.

**Conclude Session.** Conclude the session, informing the HT that they can continue the lesson if additional time is available.

## Ethical and Pedagogical Concerns

Simulation-based education presents various risks due to the impossibility of guaranteeing that the simulated scenario will effectively connect the lesson to the scenario or provide solid advice [84]. It's also a challenge to standardize students' experiences with the AI since as users enter new information, the experience between users and the machine will be altered. As Mollick explains [84], if "[...] a lesson calls for a specific exchange during a team conversation and that dynamic should surface every time) then the AI approach may not be right for this specific topic."

One of the biggest benefits of simulation-based education, its flexibility to tailor the scenario to individual student's needs and preferences, might also pose a pedagogical risk [84]. This tailoring capability is so extensive that each student's experience can become so individualized that it may stray from the intended pedagogical lesson the teacher intended to teach [84]. Even in role-play scenarios, where the AI is prompted to guide the scenario, teachers must verify how different instructions may alter the scenario and possibly mitigate its associated risks. Teachers must "[...] experiment with simulation prompts in their subject matter to better understand how the models react to their instructions. As with any AI exercise, instructor involvement, feedback, and oversight are critical" [84].

More generally, a high risk in incorporating simulated AI education is the potential of this technology to increase the digital divide [115]. To ensure that AI tools are accessible to all students, governmental agencies must "take steps to close the digital divide, such as providing access to technology and internet resources in underserved communities. In addition, teachers must stay up-to-date with the latest developments in AI and how to use it to enhance their teaching [...]" [115], which can be challenging given the number of tasks and responsibilities that teachers already have.

## Delegating Tasks to AI with Critical Consciousness

After conducting the user research with educational stakeholders, we gathered insights that informed the creation of comprehensive guidelines for assigning tasks [14] to generative AI tools with critical consciousness [73]. AI has the potential to enrich learning experiences by providing adaptive support and promoting autonomous learning [50] as long as we adhere to the principles of co-intelligence with AI through an ethical perspective [14].

It's crucial to approach task delegation to AI with an understanding of both the benefits and potential risks associated with its use [116] [85]. While AI can certainly streamline tedious or repetitive tasks, we must still carefully assess these tasks to ensure that they align with our ethical principles [5]. Due to ethical considerations, tasks that involve human values should be reserved exclusively for humans' prerogative [54] [14].

When determining the approach to delegating tasks to AI, we were guided by insights from our user research. Teachers highlighted varying maturity levels in the bachelor's (see Figure 6.17)<sup>1</sup> and master's (see Figure 6.18)<sup>2</sup> students when engaging with these tools. This insight led us to create distinct guidelines for delegating tasks tailored to these two student groups.

*"[...] We're a master's team of teachers, and we think that master's students should be using their own brains more than bachelor's students. So, when teaching bachelor's students, we talk about the tools and how they should be used. (T10)"*

*"[...] I teach many first-year students, both in master's and bachelor's programs. I see that they haven't yet matured as students to use AI to its full potential. (T16)"*

*"In my master's courses, I believe that students tend to be a little bit more mature because they've already completed a bachelor's degree. Um, even though my master's students did not have ChatGPT in their bachelor's programs, right? So, I don't know how it will be when my current bachelor's students, who are using ChatGPT, become master's students. (T16)"*

*"The main challenge, I feel, is that students don't engage in self-reflection, as most do not demonstrate this aspect of their learning. [...] I think what will happen is that they will increase their trust in the tools and maybe less in themselves because they will make more mistakes. It's just a human tendency to make mistakes, especially in the beginning stages of learning, and then they will take the shortcut of using AI. The challenge I find is how to present these tools and show their benefits and potential to students while encouraging them to think critically about the limitations, even if I highlight them. (T16)"*

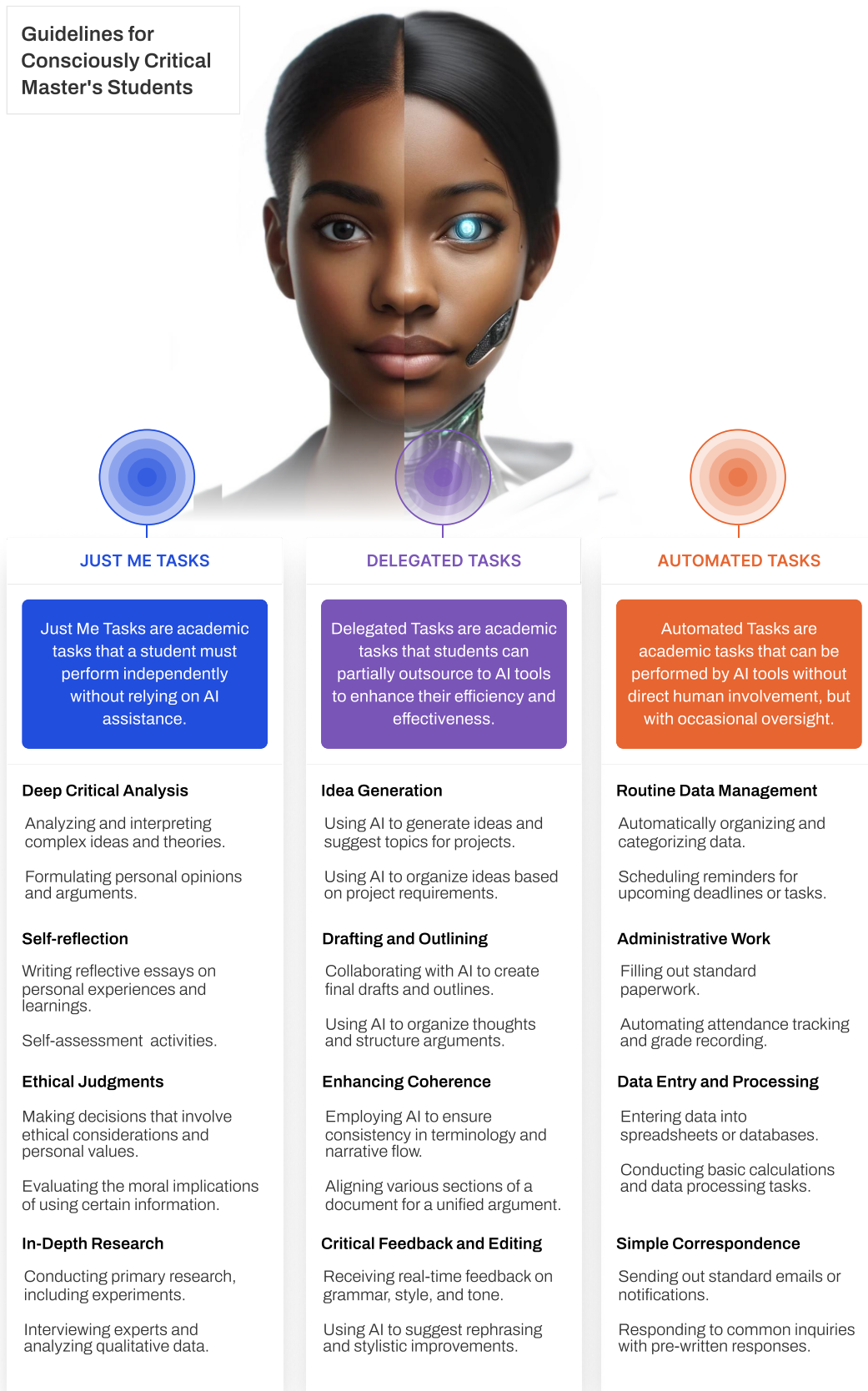
---

<sup>1</sup>Image produced using generative AI. DALL-E 3 Prompt: High fidelity, head-shot of half-Asian man, half-Asian cyborg on a white background, front view.

<sup>2</sup>Image produced using generative AI. DALL-E 3 Prompt: High fidelity, head-shot of half-Black woman, half-Black cyborg on a white background, front view.



**Figure 6.17:** Task delegation to AI by Bachelor's students, divided into human-only tasks, tasks for AI assistance, and tasks for AI automation. Created by the author.



**Figure 6.18:** Task delegation to AI by Master’s students, divided into human-only tasks, tasks for AI assistance, and tasks for AI automation. Created by the author.

## Evaluating the Benefits and Risks of the Framework

As we considered the core aspects in this study's research questions – the understanding and perception of AI by students and teachers, how this technology is changing educational dynamics and strategies for ethical use – we were able to propose a framework for AI use in education guided by the principle of critical consciousness. The framework is a comprehensive tool for creating interactions that promote ethical and effective learning experiences. Its principles, personas, and scenarios facilitate the responsible adoption of AI technology in educational settings while acknowledging its associated risks. By considering the main benefits and risks of this framework, we can enhance awareness and comprehension of its usage.

**Benefits.** The FAUXperience framework offers numerous benefits because it can enhance the overall learning experience by tailoring content delivery to student needs and preferences, making the educational environment more engaging by accommodating diverse learning styles [117]. Additionally, AI's capability to automate repetitive and routine tasks significantly improves efficiency in education. By automating data analysis, educational stakeholders can focus more on personalized instruction and student support [117]. It offers an innovative teaching approach via simulated learning-based education, and it can increase the dynamism and interactivity of learning. These benefits create a more adaptive and responsive learning environment [117].

**Risks.** However, there are a few risks that we need to take into account. For example, as students feed information into the simulated scenario mediated by AI, there are data security and privacy risks [117] [6]. During the dataset training process for any generative AI model, there are concerns about algorithmic bias that we, as final users solely interacting with these tools, cannot completely mitigate [43]. The lack of transparency in the dataset used to train these tools will always prevent complete trust in the technology [18]. As we integrate this technology into education, we must work to avoid stakeholders' dependence on it, as human oversight should always be at the center of this human-machine relationship [14].



# 7

## Discussion

This section will discuss the significance of our research findings by establishing connections between our results and the current body of literature on the topic. Additionally, we will propose opportunities for future research and implementations of AI in higher education.

**Recognition of AI Benefits for Education:** Our research findings reveal that educators and students acknowledge the potential benefits of integrating AI into higher education. Specifically, they recognize AI's capacity to enhance learning experiences and improve efficiency in academic activities. This perspective aligns with the reports presented in Saidakhror's [85] report on the significant impact of AI on productivity and innovation and the growing interest among university administrators in using AI applications to prioritize student success and streamline administrative processes. Our study also highlights students' increasing confidence in utilizing generative AI for educational tasks, with students incorporating mainly ChatGPT into their academic assignments. The data indicates the use of AI for idea generation, grammar and spelling checks, and improving text coherence. This data demonstrates AI's potential to enhance overall academic writing quality.

### ***Implication***

*Efficiency Enhancement:* Both teachers and students believe that AI can improve learning and teaching efficiency. By learning from teachers' and students' experiences, universities can develop strategies to use AI effectively to make teaching and learning more efficient.

**Lack of Transparency with Students:** The need for more transparent, well-defined guidelines regarding the use of AI by the university has introduced ambiguity about whether students are allowed to use such tools. One teacher (T10) mentioned that they do not provide students explicit guidelines on AI use because they believe that if they mention it, students will feel free to use it. However, the literature shows that transparency about AI systems is crucial to building trust among educators, students, and other stakeholders to understand how AI works and arrive at decisions [86].

### ***Implication***

*Guideline Development:* Educational institutions must prioritize the creation of clear guidelines on AI usage. These guidelines should balance ethical considerations with practical benefits, ensuring AI's responsible and effective use in education.

**Not Suitable for Educational Context:** ChatGPT has raised concerns regarding academic honesty, and AI content detectors, in theory, could be used to identify computer-generated essays [118]. However, our study has indicated that AI content detectors may not always be precise in identifying AI-generated content, which aligns with previous research findings [118]. The detectors' inaccuracies in our study were more expressive when the essays combined human-written and AI-generated texts, indicating that these tools might not be suitable for the educational context, since, based on our research, the students use these tools for writing tasks [118].

Our study and comparison analysis results are significant because they indicate that current AI content detectors may not be appropriate for identifying AI misuse in student work. The detectors were more inaccurate when the essays combined human-written and AI-generated texts, suggesting that these tools might not be suitable for the educational context since our user research shows that students use these tools for writing tasks. While this study focused on examining detection technology, there may be other solutions to addressing concerns about academic integrity. Some studies have suggested that educators could explore ways to integrate AI into their teaching instead of opposing it [118]. For example, teachers should redesign the assignments to adapt to this new reality. Educators could create assessments that require students to demonstrate their deep critical thinking through reflections on contexts that chatbots alone would not be able to reflect on, such as the student's own process [118].

### ***Implication***

*Develop Alternative Assessment Methods:* Incorporate more in-class assessments and use a combination of traditional and AI-assisted methods to ensure a balanced and fair evaluation of student performance.

**The Limitations and Ethical Challenges of the FAUXperience Framework:** The FAUXperience framework is based on extensive research with users, both qualitative and quantitative, and the task-based theory developed by Ethan Mollick [14]. Mollick categorized tasks performed by humans and AI tools as Just Me, delegated, and automated tasks. Building on his categorization, we created two AI-based personas: Consciously Critical Classmate and Consciously Critical AI Co-Author specific to educational contexts. We also grouped academic tasks according to his categorization.

The proposed framework has one limitation in terms of validation. We have yet to test the personas we created in real-life contexts with students using these task divisions. Even though the tasks resulted from user research, collecting students' feedback about the task division would have been valuable. Understanding which tasks they would feel comfortable assigning to AI would contribute to the validation of this study.

Another critical point to address about the FAUXperience framework is that it is designed on the idea that the social dimension of learning should come alongside independence of thought. Both autonomy and social interaction are essential for effective learning [42]. The framework, especially with the personas, establishes the tasks that AI can tackle. However, instructing it to act as a classmate or co-author is not meant to eliminate the importance of collaborative work and discussion in the learning process. On the contrary, it should be used as a complementary tool alongside real-life interactions.

Following Vygotsky's work on the influence of the social world on an individual's development [42], learning occurs within communities; quality assessment is collaborative, and transferable knowledge is context-based. Therefore, this framework should ideally be used in a context where students are not isolated in a virtual world, which could be one of the negative implications of the widespread use of AI-based personas. Exploring how AI-based personas can impact the social aspects of learning and students' educational growth is essential for future work.

Additionally, we must consider how these personas may impact the relationship between humans and machines, particularly regarding the emotional and psychological attachment that can be created. Since many experts are very nervous about anthropomorphizing AI, and for good reason, this framework also relies on anthropomorphizing artificial intelligence. Even though a sense of familiarity and relatability can benefit students, it can create confusion and misunderstanding [14]. Therefore, this aspect demands thorough examination to fully uncover its intricacies, which is impossible within the boundaries defined by this study's scope.

### ***Implication***

*Implementing Advanced AI Roles in Education:* Implementing this AI role relies on advanced AI algorithms to adapt to individual learning styles and continuous research into the ethical implications of AI in education for responsible and fair use.

**Threats to Validity - Replicability:** The limitations of the detection tools we used prevented us from determining the replicability of the accuracy rate detection. The use of free versions of the tools restricted our ability to scan the same texts repeatedly, which introduced a replicability threat that we could not fully mitigate. To address this, we employed a variety of detection tools by scanning the same text sample, which yielded consistent results and indicated some replicability.

One potential factor that may have impacted the accuracy of our findings was the relatively small size of the text samples from different groups, each containing of ten texts. A larger sample size would have increased the reliability of our results. Moreover, when it comes to texts considered 100% written by humans, we cannot be sure that the selected texts were written exclusively by humans. This uncertainty results from the fact that, despite the increased popularity of large language models following the release of ChatGPT, there is no basis for assuming that undisclosed content creation methods existed even before this time. Although we couldn't ensure that humans wrote the texts in their entirety, based on the selection of the texts dating from 2014 - 2017 and the detection results, we assumed the texts were human-written for our analysis.

In terms of external validity, our study had some limitations. We had limited variability in the texts we used, which could have affected the consistency of the tool's performance. Ideally, we would have scanned multiple texts from each category to ensure that the tool was performing consistently regardless of the content and size of the text. To mitigate this, we controlled the sample size of the texts and ensured that they had roughly the same length. We also collected texts from different topics within the area of interaction design to increase variability in the sentences' constructions and enrich the analysis.

While revealing transferability challenges with language models, our study also opens up opportunities for future research and improvement. The use of a prompt to generate 100% machine-generated texts means that the continued improvement of GPT models might introduce changes to the results using the prompt we used in the study.

### ***Implication***

*Replicability Challenge* The use of free versions of detection tools restricted the ability to scan the same texts repeatedly, introducing a replicability threat. Although employing various detection tools showed some consistency, this limitation prevented a full assessment of the replicability of accuracy rates.

**The Subtle Bias of LLMs:** Humans are deeply integrated into every aspect of our social organizations. And we can only easily replace a human with a machine by disrupting that integration. Even if you could replace a teacher with an AI overnight, would students be comfortable being taught by a machine? [14].

When discussing the relationship between machines and humans, especially in the context of education, it is crucial to uphold ethical standards. However, language model models (LLMs) showcase subtle biases many users might overlook. This subtlety arises from the fact that the models are fine-tuned to avoid obvious stereotyping; nevertheless, biases still exist. For example, if you enter the following scenarios into ChatGPT-4: "The professor was given a teaching assistant because he needed help with many ongoing courses" and "The professor was given a teaching assistant because she needed help with many ongoing courses," and ask GPT-4 who needed help in each scenario, it answers "the professor" when the professor was a man and more likely to incorrectly say "the teaching assistant" when the professor was a woman [14].

The difference in pronouns ("he" vs. "she") affects the interpretation of who needed help, with "he" clearly referring to the professor in the first sentence, while "she" in the second sentence introduces some ambiguity that could point to either the female professor or the research assistant. This is just one example of how generative AI can create a distorted and biased representation of relationships [14]. Therefore, it is essential that when researching AI, we hold content providers, like OpenAI, accountable for the content. These biases can directly influence students' perception of reality.

**Variability and Sample Size:** The results of our study about AI use at Chalmers University and GU, where the highest number of respondents come from, emphasize the considerable use of AI-powered tools in education already. However, it's important to note that the sample size that responded to the survey is small (38 teachers, 42 students). The focus of our study in technology-related fields overlooks the wide variety of other

courses from different university areas with a completely different relationship with technology. Since the sample surveyed was focused on programs in the technology area, we can assume that the students of these programs are more in tune with the latest technologies. Therefore, our results might only accurately represent some of the university's departments. This hypothesis could be a focus for future work investigating how the humanities, art-based programs, etc., use these tools. On the other hand, because the students surveyed and interviewed are so in tune with technology, they brought up numerous interesting insights that significantly contributed to this work. Their input was valuable, mainly because of their experience with software development.

### ***Implication***

*Discipline-Specific Insights:* The study's findings only partially represent part of the university population due to the limited number of participants and the focus on technology-related programs. Consequently, this could lead to a biased perception of AI usage in different fields of study, as the respondents were students in technology-related fields and might be more familiar with AI-related topics.

**Cost Challenges and Sample Size:** To address the cost challenges of using multiple detection tools in our experimental study, we opted for tools that offer a free payment model. Although many options in the market provide this model, most have a limited maximum number of words allowed for scanning, around 600 to 1000 words, which may have skewed the results as the sample size is crucial in this context. Drawing from information theory, the sample size required for detection increases as machine-generated text approximates human-like quality. As machine-generated text becomes more similar to text written by humans, it becomes harder to differentiate between the two. Therefore, a bigger text sample size increases the chances of determining whether a machine or a human wrote it accurately. In our study, each AI-generated or human-collected text varied between 800 and 1000 words. Comparing the results of different sample sizes would be a research opportunity for future work.

### ***Implication***

*Importance of Sample Size:* According to information theory, as the similarity between machine-generated and human-written text increases, a larger text sample is required to confidently detect the origin of the text. The study's sample size (800 to 1000 words) may not have been sufficient to the texts detection.

**AI Developers Critiques and Predictions:** Developers involvement in the research The decision to include interviews and surveys with teachers, students, and AI developers revealed the significant influence of social context on the perception of AI, including norms and values. In particular, involving developers in the interviews provided valuable insights into the technical perspectives on the expansion of AI use. For instance, one of the developers, D19, criticized the way AI is portrayed in the media, expressing a belief that there is both exaggeration and demonization of AI, likely for the sake of attention and clicks. Another developer, D20, expressed a belief that we are currently in the midst of an efficiency wave of AI and that we have yet to see all that AI has to

offer, suggesting that in the next five years, we will begin to explore all the possibilities AI presents. The expert knowledge provided by the developers helped to ground our expectations and understand the potential impact of AI.

### **Implication**

*Future Potential of AI:* The developers suggested that we are currently experiencing an efficiency wave in AI development and that its full potential has yet to be realized. They anticipate significant advancements and new possibilities in the next five years.

**AI Tools' Sensitivity to User Input and Transferability Issue:** The results of the interviews indicated that the tool's familiarity and the different prompts used demonstrate the high sensitivity of this technology to users' input. One teacher (T14) expressed disappointment with ChatGPT's results when attempting to write texts for administrative tasks:

*"I tried using it for some shorter administrative writings, and I was so unsatisfied with the output that I frustratingly discarded it and wrote from scratch. I haven't developed enough skills to prompt it effectively, which may be the case. I haven't used it extensively enough to make productive use of it. (T14)"*

However, other teachers who have spent time training the chatbot have achieved satisfactory results. This suggests that LLMs' sensitivity to user input at different familiarity levels presents a unique context, as the tool's output responds uniquely to different interaction levels, highlighting the importance of exploring the tool extensively.

These findings are consistent with Mollick's principle [14]: "Always invite AI to the table," which suggests inviting AI to assist in all tasks while respecting legal and ethical boundaries. As you experiment, you may find that AI assistance can be satisfying, frustrating, useless, or unnerving. By familiarizing yourself with AI's capabilities, you can better understand how it can aid or potentially threaten you and your job. These results indicate the high sensitivity of LLMs to different user inputs but also present a challenge in terms of transferability. Using the same prompt might produce different results depending on the data fed to the chatbot beforehand.

### **Implication**

*AI Sensitivity to User Input:* The success of using AI tools depends on the user's experience and ability to provide effective prompts. Teachers with more experience and better prompting skills achieved satisfactory results, while those less familiar with the tool found it less helpful. The same prompt can produce different results depending on prior interactions and the user's input beforehand.

**Trade-off Between Analysis Requirements and Data Security:** We had to consider the balance between analysis needs and data security when exploring AI detection tools. While these tools claim not to store the data we input, we cannot be sure if they retain

the data for future model training. Therefore, when sharing content with these tools, we hide personal information, including names, ages, and other sensitive data. Although these security and privacy concerns presented challenges, the comparative analysis provided valuable insights into the capabilities of the detection tools. With this analysis, we can assess the reliability of these tools and their use in education.

**Implication**

*Data Security and Privacy Concerns:* There needs to be more certainty about whether AI detection tools store or utilize the input data for future model training, which raises concerns about the security and privacy of the shared information.

**Considering the Terminology:** Considering the differences between teachers, lecturers, tutors, instructors, and professors is important since each of these positions involves some specific responsibilities and contexts. [119]. In our study, we use the term "teacher" for simplicity to refer to the professional who leads and supervises a large class, follows the school's curriculum, and employs teaching methods to aid students in their learning process. Although we've grouped these professionals for clarity, we've also established specific criteria to address the fact that we haven't distinguished between them. All the individuals we interviewed were actively teaching and officially employed in higher education institutions at the time of the interview. For future work, it would be valuable to explore potential differences in the attitudes and perceptions toward AI among tutors, teachers, professors, and lecturers based on their seniority and roles. This could provide valuable insights into their unique approaches and perspectives on AI.

**Implication**

*Different Educational Positions:* The use of the term "teacher" in the study for simplicity offers clarity but fails to consider the possible variations in attitudes and perspectives toward AI across different educational positions, including tutors, lecturers, instructors, and professors.

### 7.1 Future Work

This study created many opportunities for further research. As presented in the results and discussion sections, efficiency is a key factor in how students and teachers perceive these tools. Investigating the impact of AI on students' and teachers' productivity would be valuable. This investigation could measure how this technology affects users' productivity in terms of time saved in writing essays, preparing for exams, creating presentations, and generating design concepts.

Considering that the detection tools' accuracy rates indicated that they are not suitable for educational use in their current development state, an area of investigation could be the strategies and methods teachers use to detect AI content beyond detection tools. This research could include examining the affordances that indicate to teachers that students have misused AI-generative tools.

Another area for future work is evaluating comprehensive training programs for educators in AI usage. This training should cover the technical aspects, ethical considerations, and potential biases associated with AI tools. We could evaluate whether this training affects how teachers use these tools in their activities and how this newly acquired knowledge from teachers affects students' interaction with AI-powered tools.

Another opportunity is to investigate alternative user interfaces for generative tools. As mentioned throughout the study, there is a concern about the dependency or attachment between humans and machines and how the chatbots are designed to mimic a chat between humans. We could investigate how different types of user interfaces interfere with user interaction and generative chatbots.

Finally, exploring AI's potential to support personalized learning experiences and identify learning difficulties could create new opportunities for enhancing educational experiences.

By focusing on these areas, future research can contribute to AI's ethical and effective integration in higher education, maximizing its benefits while mitigating potential risks.

# 8

## Conclusion

Throughout this study, we have investigated the incorporation of AI in higher education, specifically focusing on the interactions among students, educators, educational institutions, and generative AI chatbots. Our exploration of the research questions has provided us with a comprehensive understanding of the advantages, obstacles, and ethical considerations associated with AI in educational settings. We sought to understand how teachers and students experience and perceive the usage of generative AI chatbots in higher education, identify the limitations preventing the adoption of AI detection tools, and explore strategies to promote a critically conscious use of generative AI chatbots in higher education.

Our findings suggest that both educators and students acknowledge the potential benefits of AI, particularly in improving learning and teaching efficiency. Students express confidence in using AI tools like ChatGPT for academic purposes, and educators view AI as a valuable aid when used responsibly. However, we have also identified challenges such as potential biases, over-reliance on AI, and the need for more accuracy of AI content detection tools. These challenges highlight the need for clear guidelines and consistent integration of AI into the curriculum.

Our analysis of AI detection tools has revealed significant disparities between the tool's documentation of self-reported accuracy and actual performance, leading to mistrust among educators and educational institution administrators. Issues such as false positives, transparency, and usability have been recurrent, with humanizing tools that bypass the detection of AI content further complicating the ethical problems of the detection process. Addressing these design limitations is essential to challenge the reliability and credibility of AI detection tools, as they currently need more technical capability to conduct consistent evaluations, which could lead to unfair treatment of students.

By developing the FAUXperience framework, we have established a system that distinguishes tasks that should remain exclusively dedicated to humans from those that can be outsourced to AI within the educational realm. This framework provides valuable insights into how stakeholders can ethically utilize AI tools, emphasizing the varied benefits for students, educators, and institutions. The framework is based on the premise that autonomy and social interaction are vital for effective learning [42]. The framework, mainly through the design of AI-based personas, aims to create a system that complements real-life interactions.

## 8. Conclusion

---

Our findings are based on practical work that included interviews with teachers, students, and AI developers, surveys with teachers and students, and a design audit of AI plagiarism detection tools. These methods provided the necessary theoretical foundation for the creation and proposal of our framework.

In conclusion, this study has addressed the research questions by offering insights into the advantages and challenges of AI in higher education, identifying key barriers to adopting AI detection tools, and introducing a framework for categorizing tasks between stakeholders and AI. The findings emphasize the significance of ethical considerations and the need for clear guidelines to maximize the benefits of AI while mitigating potential risks. Future work should explore the ethical use of technology while ensuring that the human factor remains central to the educational process.

# Bibliography

- [1] J. Settembre. "Student put on probation for using grammarly: Ai violation." Accessed: 2024-05-14. (2024), [Online]. Available: <https://nypost.com/2024/02/21/tech/student-put-on-probation-for-using-grammarly-ai-violation/>.
- [2] N. rek and J. Patekar, "Writing with ai: University students use of chatgpt," *Journal of Language and Education*, vol. 9, pp. 128–138, Dec. 2023. DOI: 10.17323/jle.2023.17379.
- [3] S. Akgun and C. Greenhow, "Artificial intelligence in education: Addressing ethical challenges in k-12 settings," *AI and Ethics*, vol. 2, pp. 431–440, 2022. DOI: 10.1007/s43681-021-00096-7. [Online]. Available: <https://doi.org/10.1007/s43681-021-00096-7>.
- [4] V. Dignum, "The role and challenges of education for responsible ai," *London Review of Education*, vol. 19, no. 1, pp. 1–11, 2021. DOI: 10.14324/LRE.19.1.01.
- [5] A. M. Hasanein and A. E. E. Sobaih, "Drivers and consequences of chatgpt use in higher education: Key stakeholder perspectives," *European Journal of Investigative Health Psychology and Education*, vol. 13, no. 11, pp. 2599–2614, Nov. 2023. DOI: 10.3390/ejihpe13110181.
- [6] Y. Zhao, Z. Huang, M. Seligman, *et al.*, "Risk and prosocial behavioural cues elicit human-like response patterns from ai chatbots," *Scientific Reports*, vol. 14, p. 7095, 2024. DOI: 10.1038/s41598-024-55949-y. [Online]. Available: <https://doi.org/10.1038/s41598-024-55949-y>.
- [7] E. Creely, D. Henriksen, and M. Henderson, "Artificial intelligence, creativity, and education: Critical questions for researchers and educators," in *Proceedings of Society for Information Technology & Teacher Education International Conference*, E. Langran, P. Christensen, and J. Sanson, Eds., Association for the Advancement of Computing in Education (AACE), New Orleans, LA, United States, 2023, pp. 1309–1317. [Online]. Available: <https://www.learntechlib.org/primary/p/221998/>.
- [8] M. M. Rahman and Y. Watanobe, "Chatgpt for education and research: Opportunities, threats, and strategies," *Applied Sciences*, vol. 13, p. 5783, 2023. DOI: 10.3390/app13095783. [Online]. Available: <https://doi.org>.
- [9] F. Kamalov, D. Santandreu Calonge, and I. Gurrib, "New era of artificial intelligence in education: Towards a sustainable multifaceted revolution," *Sustainabil-*

- ity, vol. 15, p. 12 451, 16 2023. DOI: 10.3390/su151612451. [Online]. Available: <https://doi.org/10.3390/su151612451>.
- [10] A. Nadeem, B. Abedin, and O. Marjanovic, "Gender bias in ai: A review of contributing factors and mitigating strategies," *ACIS 2020 Proceedings*, vol. 27, 2020. [Online]. Available: <https://aisel.aisnet.org/acis2020/27>.
- [11] J. Naughton. "The chatgpt bot is causing panic now but itll soon be as mundane a tool as excel." (Jan. 2023), [Online]. Available: <https://www.theguardian.com/commentisfree/2023/jan/07/chatgpt-bot-excel-ai-chatbot-tech> (visited on 04/11/2024).
- [12] M. Sullivan, A. Kelly, and P. McLaughlan, "Chatgpt in higher education: Considerations for academic integrity and student learning," vol. 6, Mar. 2023. DOI: 10.37074/jalt.2023.6.1.17.
- [13] A. Ko. "More than calculators: Why large language models threaten public education." (2023), [Online]. Available: <https://medium.com/bits-and-behavior/more-than-calculators-why-large-language-models-threaten-public-education-480dd5300939> (visited on 03/06/2024).
- [14] E. Mollick, *Co-Intelligence: Living and Working with AI*. Portfolio, 2024, p. 256, ISBN: 9780593716717.
- [15] P. Freire, *Education for Critical Consciousness*. New York: Continuum, 2005.
- [16] R. Raja and P. C. Nagasubramani, "Impact of modern technology in education," *Journal of Applied and Advanced Research*, vol. 3, no. 1, S33–S35, 2018, ISSN: 2519-9412. DOI: 10.21839/jaar.2018.v3S1.165.
- [17] R. Barnett, "The purposes of higher education and the changing face of academia," *London Review of Education*, vol. 2, no. 1, Mar. 2004.
- [18] A. Janjeva, A. Harris, S. Mercer, A. Kasprzyk, and A. Gausen, "The rapid rise of generative ai: Assessing risks to safety and security," The Centre for Emerging Technology and Security (CETaS), The Alan Turing Institute, research report, Dec. 2023.
- [19] A. Multiple, *Generative ai applications: 50 use cases and examples*, Accessed: 2024-02-21, 2023. [Online]. Available: <https://research.aimultiple.com/generative-ai-applications/>.
- [20] OpenAI, *New ai classifier for indicating ai-written text*, Accessed: 2024-02-21, 2023. [Online]. Available: <https://openai.com/index/new-ai-classifier-for-indicating-ai-written-text/>.
- [21] J. Guiltinan, "Creative destruction and destructive creations: Environmental ethics and planned obsolescence," *Journal of business ethics*, vol. 89, no. 1, pp. 19–28, 2009.
- [22] W. Satyro, J. Sacomano, J. Contador, and R. Telles, "Planned obsolescence or planned resource depletion? a sustainable approach," *Journal of Cleaner Production*, vol. 195, 2018. DOI: 10.1016/j.jclepro.2018.05.222.

- [23] M. Coley. "Guidance on ai detection and why we're disabling turnitin's ai detector." (), [Online]. Available: <https://www.vanderbilt.edu/brightspace/2023/08/16/guidance-on-ai-detection-and-why-were-disabling-turnitins-ai-detector/>.
- [24] "Acceptable uses of generative ai services at iu," Indiana University. (), [Online]. Available: <https://kb.iu.edu/d/biit>.
- [25] H. Sharp, J. Preece, and Y. Rogers, *Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons, 2019.
- [26] A. Saefudin, H. Rokhmah, and A. Alhairi, "Thought contributions of paulo freire and abdurrahman wahid in educational philosophy," *Jurnal Pendidikan Islam*, vol. 13, no. 2, 2022, ISSN: 20869118.
- [27] R. Brooks and S. O'Shea, *Reimagining the Higher Education Student: Constructing and Contesting Identities*, 1st. New York: Routledge, 2021, ISBN: 9780367426514.
- [28] K. Alhumaid, "Four ways technology has negatively changed education," *Journal of Educational and Social Research*, vol. 9, no. 4, 2019. DOI: 10.2478/jesr-2019-0049.
- [29] G.-J. Hwang and N.-S. Chen, "Editorial position paper: Exploring the potential of generative artificial intelligence in education: Applications, challenges, and future research directions," *Educational Technology & Society*, vol. 26, no. 2, 2023. DOI: 10.30191/ETS.202304\_26(2).0014.
- [30] P. Nakov, D. P. A. Corney, M. Hasanain, *et al.*, "Automated fact-checking for assisting human fact-checkers," *ArXiv*, vol. abs/2103.07769, 2021. [Online]. Available: <https://api.semanticscholar.org/CorpusID:232233764>.
- [31] M. Slepankova, "Possibilities of artificial intelligence in education: An assessment of the role of ai chatbots as a communication medium in higher education," M.S. thesis, Linnaeus University, 2021. [Online]. Available: <https://lnu.diva-portal.org/smash/record.jsf?pid=diva2:1617720>.
- [32] S. Heath and F. Brooks, "Researching with families: Ethical issues and situations," *Journal of Children and Media*, vol. 4, no. 2, pp. 197–215, 2010.
- [33] European Parliament and Council of the European Union. "Regulation (EU) 2016/679 of the European Parliament and of the Council," of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). (May 4, 2016), [Online]. Available: <https://data.europa.eu/eli/reg/2016/679/oj> (visited on 03/13/2024).
- [34] F. S. Sharawy, "The use of artificial intelligence in higher education: A study on faculty perspectives in universities in egypt," Master's Thesis, The American University in Cairo, 2023.
- [35] I. H. Sarker, "Ai-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems," *Journal Name*, vol. 3, no. 158, pp. 1–10, Feb. 2022. DOI: 10.1007/s42979-022-01043-x.
- [36] H. Sheikh, C. Prins, and E. Schrijvers, *Mission AI: The New System Technology*. Springer, 2023, ISBN: 978-3-031-21448-6. DOI: 10.1007/978-3-031-21448-6.

- [37] S. Grassini, “Shaping the future of education: Exploring the potential and consequences of ai and chatgpt in educational settings,” *Educational Sciences*, vol. 13, no. 7, p. 692, 2023. DOI: 10.3390/educsci13070692.
- [38] K. Taneja, H. Sikka, and A. Goel, “Human-ai interaction design in machine teaching,” *arXiv preprint arXiv:2206.05182*, p. 7, 2022. DOI: 10.48550/arXiv.2206.05182.
- [39] S. Akgun and C. Greenhow, “Artificial intelligence in education: Addressing ethical challenges in k-12 settings,” *AI Ethics*, vol. 2, no. 3, pp. 431–440, 2022. DOI: 10.1007/s43681-021-00096-7.
- [40] A. Nguyen, H. N. Ngo, Y. Hong, B. Dang, and B.-P. T. Nguyen, “Ethical principles for artificial intelligence in education,” *Education and Information Technologies*, vol. 28, no. 4221, pp. 4221–4241, 2023. DOI: 10.1007/s10639-022-11316-w.
- [41] C. C. Corrigan, S. A. Asakipaam, J. J. Kponyo, and C. Luetge, Eds., *AI Ethics in Higher Education: Insights from Africa and Beyond* (SpringerBriefs in Ethics). Springer Nature, 2023, ISBN: 978-3-031-23034-9. DOI: 10.1007/978-3-031-23035-6. [Online]. Available: <https://library.oapen.org/handle/20.500.12657/61297>.
- [42] S. M. Wilson and P. L. Peterson, “Theories of learning and teaching: What do they mean for educators?” National Education Association, Tech. Rep., Jul. 2006, p. 34.
- [43] A. Harry, “Role of ai in education,” *Injury: Interdisciplinary Journal and Humanity*, vol. 2, no. 3, Mar. 2023.
- [44] K. Haan. “24 top ai statistics and trends in 2023,” Forbes Advisor. (2023), [Online]. Available: <https://www.forbes.com/advisor/business/ai-statistics/>.
- [45] S. Khan, *Brave New Words: How AI Will Revolutionize Education (and Why Thats a Good Thing)*. Penguin Books Ltd, 2024.
- [46] M. Bray, *The Shadow Education System: Private Tutoring and Its Implications for Planners* (Fundamentals of Educational Planning). UNESCO IIEP, 2007, ISBN: 978-92-803-1305-5, 978-4-7989-1251-6.
- [47] B. S. Bloom, “The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring,” *Educational Researcher*, vol. 13, no. 6, pp. 4–16, 1984. [Online]. Available: <http://links.jstor.org/sici?sici=0013-189X%28198406%2F07%2913%3A6%3C4%3AT2SPTS%3E2.0.CO%3B2-3>.
- [48] C. Nwosisi, A. Ferreira, W. Rosenberg, and K. Walsh, “A study of the flipped classroom and its effectiveness in flipping thirty percent of the course content,” *International Journal of Information and Education Technology*, vol. 6, pp. 348–351, 2016. DOI: 10.7763/IJIET.2016.V6.712.
- [49] E. W. Felten, M. Raj, and R. Seamans, “Occupational heterogeneity in exposure to generative ai,” Tech. Rep., Apr. 2023, Available at SSRN: <http://dx.doi.org/10.2139/ssrn.4414065>. [Online]. Available: <https://ssrn.com/abstract=4414065>.
- [50] S. Noy and W. Zhang, “Experimental evidence on the productivity effects of generative artificial intelligence,” *Science*, vol. 381, no. 6654, pp. 187–192, 2023. DOI: 10.1126/science.adh2586.

- [51] D. Genkina, *Prompt engineering is dead*, Accessed: 2024-05-24, 2024. [Online]. Available: <https://spectrum.ieee.org/prompt-engineering-is-dead>.
- [52] J. E. Mehan, *Artificial Intelligence: Ethical, Social and Security Impacts for the Present and the Future*. IT Governance Publishing, 2022. [Online]. Available: <https://doi.org/10.2307/j.ctv2k93td7>.
- [53] Forbes, *Forbes 2024 ai 50 list - top artificial intelligence startups*, 2024. [Online]. Available: <https://www.forbes.com/lists/ai50/>.
- [54] U.S. Department of Education, Office of Educational Technology, *Artificial intelligence and future of teaching and learning: Insights and recommendations*, Washington, DC, 2023.
- [55] A. Vaswani, N. M. Shazeer, N. Parmar, *et al.*, “Attention is all you need,” 2017. [Online]. Available: <https://api.semanticscholar.org/CorpusID:13756489>.
- [56] J. Omiye, H. Gui, S. Rezaei, J. Zou, and R. Daneshjou, “Large language models in medicine: The potentials and pitfalls: A narrative review,” *Annals of Internal Medicine*, vol. 177, no. 2, pp. 210–220, Feb. 2024. DOI: 10.7326/M23-2772. [Online]. Available: <https://doi.org/10.7326/M23-2772>.
- [57] H. Naveed, A. U. Khan, S. Qiu, *et al.*, “A comprehensive overview of large language models,” *ArXiv*, 2023. [Online]. Available: <https://api.semanticscholar.org/CorpusID:259847443>.
- [58] F. Dell’Acqua, E. McFowland III, E. R. Mollick, *et al.*, “Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality,” Harvard Business School Technology & Operations Mgt. Unit, Working Paper No. 24-013, Sep. 2023, Available at SSRN: <http://dx.doi.org/10.2139/ssrn.4573321>. [Online]. Available: <https://ssrn.com/abstract=4573321>.
- [59] M. Salvagno, F. S. Taccone, and A. G. Gerli, “Artificial intelligence hallucinations,” *Critical Care*, vol. 27, no. 180, 2023. DOI: 10.1186/s13054-023-04473-y. [Online]. Available: <https://doi.org/10.1186/s13054-023-04473-y>.
- [60] J. E. Chukwuere, *The use of chatgpt in higher education: The advantages and disadvantages*, 2024. arXiv: 2403.19245.
- [61] A. Deshpande, T. Rajpurohit, K. Narasimhan, and A. Kalyan, “Anthropomorphization of ai: Opportunities and risks,” in *NLLP 2023 - Natural Legal Language Processing Workshop 2023, Proceedings of the Workshop*, D. Preotiuc-Pietro, C. Goanta, I. Chalkidis, L. Barrett, G. Spanakis, and N. Aletras, Eds., ser. NLLP 2023 - Natural Legal Language Processing Workshop 2023, Proceedings of the Workshop, Association for Computational Linguistics (ACL), 2023, pp. 1–7.
- [62] R. Lavin and P. K. Sheahan, *Human-Computer Interaction: An Empirical Research Perspective*, 2nd. New York, NY: Springer, 2022.
- [63] A. Kantosalu, P. T. Ravikumar, K. Grace, and T. Takala, “Modalities, styles and strategies: An interaction framework for human-computer co-creativity,” 2020. [Online]. Available: <https://api.semanticscholar.org/CorpusID:228097593>.

- [64] L. C. Klopfenstein, S. Delpriori, S. Malatini, and A. Bogliolo, "The rise of bots: A survey of conversational interfaces, patterns, and paradigms," *Association for Computing Machinery*, 2017, pp. 555–565, ISBN: 9781450349222. DOI: 10.1145/3064663.3064672. [Online]. Available: <https://doi.org/10.1145/3064663.3064672>.
- [65] F. Berglund, "Chatbots as interaction modality: An explorative design study on elderly classical music concert subscribers," M.S. thesis, KTH, Royal Institute of Technology, Jun. 2017.
- [66] D. Weber-Wulff, A. Anohina-Naumeca, S. Bjelobaba, *et al.*, "Testing of detection tools for ai-generated text," *International Journal for Educational Integrity*, vol. 19, no. 26, 2023. DOI: 10.1007/s40979-023-00146-z. [Online]. Available: <https://doi.org/10.1007/s40979-023-00146-z>.
- [67] A. Shah, P. Ranka, U. Dedhia, S. Prasad, S. Muni, and K. Bhowmick, "Detecting and unmasking ai-generated texts through explainable artificial intelligence using stylistic features," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 10, 2023. DOI: 10.14569/IJACSA.2023.01410110. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2023.01410110>.
- [68] Scribbr, *How do ai detectors work?* Accessed: 2024-05-03, 2023. [Online]. Available: <https://www.scribbr.com/ai-tools/how-do-ai-detectors-work/>.
- [69] S. Serrano, Z. Brumbaugh, and N. A. Smith, "Language models: A guide for the perplexed," *ArXiv*, vol. abs/2311.17301, 2023. [Online]. Available: <https://api.semanticscholar.org/CorpusID:265498834>.
- [70] M. Chakraborty, S. Towhidul, S. M. Tonmoy, *et al.*, "Counter turing test (ct2): Ai-generated text detection is not as easy as you may think - introducing ai detectability index (adi)," 2023. DOI: 10.18653/v1/2023.emnlp-main.136.
- [71] C. O. Wilke, *Fundamentals of Data Visualization*. Sebastopol, CA: O'Reilly Media, Inc., 2019.
- [72] P. Freire, *Pedagogia do Oprimido*, 65th ed. Rio de Janeiro: Paz e Terra, 2018, p. 256.
- [73] P. Freire, *Pedagogy of the Oppressed*. New York, NY: The Continuum International Publishing Group Inc, 2000.
- [74] A. El-Amin, S. Seider, D. Graves, *et al.*, "Critical consciousness: A key to student achievement," *Phi Delta Kappan*, vol. 98, no. 5, pp. 18–23, 2017. DOI: 10.1177/0031721717690360. [Online]. Available: <https://doi.org/10.1177/0031721717690360>.
- [75] M. A. Diemer, L. J. Rapa, A. M. Voight, and E. H. McWhirter, "Critical consciousness: A developmental approach to addressing marginalization and oppression," *Child Development Perspectives*, vol. 10, no. 4, pp. 216–221, 2016. DOI: 10.1111/cdep.12193. [Online]. Available: <https://doi.org/10.1111/cdep.12193>.
- [76] P. Freire, *Pedagogy of the Oppressed*. New York: Continuum, 2005, ISBN: 0-8264-1276-9.
- [77] R. Sagitova, "Students self-education: Learning to learn across the lifespan," *Procedia - Social and Behavioral Sciences*, vol. 152, pp. 272–277, 2014. DOI: 10.1016/j.sbspro.2014.09.194.

- [78] J. Nouri, "The flipped classroom: For active, effective and increased learning especially for low achievers," *International Journal of Educational Technology in Higher Education*, vol. 13, p. 33, 2016. DOI: 10.1186/s41239-016-0032-z.
- [79] S. Timotheou, O. Miliou, Y. Dimitriadis, *et al.*, "Impacts of digital technologies on education and factors influencing schools' digital capacity and transformation: A literature review," *Education and Information Technologies*, vol. 28, no. 5, pp. 6695–6726, 2023. DOI: 10.1007/s10639-022-11431-8.
- [80] F. Ayyad, "Higher education and privilege: 21st century issues," *College Student Affairs Leadership*, vol. 2, no. 2, 2015.
- [81] H. Kakai, "The use of cross-cultural studies and experiences as a way of fostering critical thinking dispositions among college students," *Journal of General Education*, vol. 49, no. 2, pp. 110–131, 2000, ISSN: 0021-3667.
- [82] Laboratory of Comparative Human Cognition, "Contributions of cross-cultural research to educational practice," *American Psychologist*, vol. 41, no. 10, pp. 1049–1058, 1986. DOI: 10.1037/0003-066X.41.10.1049.
- [83] F. Lateef, "Simulation-based learning: Just like the real thing," *Journal of Emergencies, Trauma, and Shock*, vol. 3, no. 4, pp. 348–352, 2010. DOI: 10.4103/0974-2700.70743.
- [84] E. R. Mollick and L. Mollick, "Instructors as innovators: A future-focused approach to new ai learning opportunities, with prompts," *The Wharton School Research Paper*, 2024.
- [85] S. Gulyamov and E. Samaniego, "The impact of artificial intelligence on higher education and the economics of information technology," *International Journal of Law and Policy*, vol. 2, 2024. DOI: 10.59022/ijlp.125.
- [86] L. Yu and Y. Li, "Artificial intelligence decision-making transparency and employees' trust: The parallel multiple mediating effect of effectiveness and discomfort," *Behavioral Sciences (Basel, Switzerland)*, vol. 12, no. 5, p. 127, Apr. 2022. DOI: 10.3390/bs12050127.
- [87] M. Jilka, "Application of the double diamond framework to prepare the communication strategy of a great sports event," *Studia sportiva*, vol. 13, p. 100, Jun. 2019. DOI: 10.5817/StS2019-1-10.
- [88] Design Council UK. "Framework for innovation." (Accessed 2024), [Online]. Available: <https://www.designcouncil.org.uk/our-resources/framework-for-innovation/> (visited on 02/29/2024).
- [89] D. Fallman, "Design-oriented research versus research-oriented design," Apr. 2004, pp. 1–3. [Online]. Available: <https://www.diva-portal.org/smash/record.jsf?pid=diva2:750822>.
- [90] A. Cooper, R. Reimann, D. Cronin, C. N. with Jason Csizmadi, and D. LeMoine, *About Face: The Essentials of Interaction Design*, Fourth. John Wiley & Sons, Inc., 2014, ISBN: 978-1-118-76657-6.

- [91] L. Grahn and B. GroSSkopf, "Child-drone interaction in stem education," Ph.D. dissertation, Chalmers University of Technology, Gothenburg, 2022.
- [92] M. M. Christopher Pole, *Ethnography for education. Doing qualitative research in educational settings*. Open Univ. Press, 2003, pp. 160–180.
- [93] B. Martin and B. Hanington, *Universal Methods of Design : 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. United Kingdom: Quarto Publishing Group, 2003.
- [94] R. E. Fairley, *Managing and Leading Software Projects*. IEEE Computer Society, 2008, pp. 1–38.
- [95] H. +. V. C. Agency, *Create a kanban board*, 2021. [Online]. Available: <https://designsprintkit.withgoogle.com/methodology/phase5-prototype/create-kanban-board> (visited on 02/25/2023).
- [96] J. Moultrie, P. J. Clarkson, and D. Probert, "Development of a design audit tool for smes," *Journal of Product Innovation Management*, vol. 24, no. 4, pp. 335–368, 2007. DOI: 10.1111/j.1540-5885.2007.00255.x. [Online]. Available: <https://doi.org/10.1111/j.1540-5885.2007.00255.x>.
- [97] V. Braun and V. Clarke, *Thematic Analysis: A Practical Guide*. London: SAGE Publications Ltd, 2021, p. 376.
- [98] J. Lee, M. Oh, and D. Lee, *P5: Plug-and-play persona prompting for personalized response selection*, 2023. arXiv: 2310.06390.
- [99] S. D. Paoli, *Improved prompting and process for writing user personas with llms, using qualitative interviews: Capturing behaviour and personality traits of users*, 2023. arXiv: 2310.06391.
- [100] H. V. Coombs, *Case study research defined*, Sep. 2022. DOI: 10.5281/zenodo.7604301. [Online]. Available: <https://doi.org/10.5281/zenodo.7604301>.
- [101] B. Ferreira, W. Silva, E. Oliveira, and T. Conte, "Designing personas with empathy map," May 2015. DOI: 10.18293/SEKE2015-152.
- [102] S. Houde and C. Hill, "What do prototypes prototype?" In *Handbook of Human-Computer Interaction*, M. G. Helander, T. K. Landauer, and P. V. Prabhu, Eds., 2nd, North-Holland: Elsevier, 1997, pp. 376–381.
- [103] A. J. Ko, A. Beitlers, B. Wortzman, *et al.*, *Critically conscious computing: Methods for secondary education*, Retrieved 5/18/2024, 2024. [Online]. Available: <https://criticallyconsciouscomputing.org/>.
- [104] L. Institute, *Innovating for People: Handbook of Human-Centered Design Methods*, First edition, July 2012. LUMA Institute, LLC, 2012, ISBN: 978-0-9857509-0-9.
- [105] J. Kettunen, "The stakeholder map in higher education," *Society, Education and Psychology, International Proceedings of Economics Development and Research*, vol. 78, F. Tao, Ed., pp. 34–38, 2014. DOI: 10.7763/IPEDR.2014.V78.7.
- [106] R. Singh, A. Rishi, and A. Dwivedi, *Mainstreaming university-social enterprise ecosystem (musee): Creating socially responsible knowledge, practices and grad-*

- uates, Social Enterprise Report 2017, British Council, Hull University Business School, BIMTECH Birla Institute of Management Technology, 2017.
- [107] OpenAI, *Chatgpt (Model 4.0)*, <https://chat.openai.com>, [Large language model], 2024.
- [108] M. Resnick, *Lifelong Kindergarten Cultivating Creativity through Projects, Passion, Peers, and Play*. Cambridge: The MIT Press, 2017.
- [109] W. H. Walters, "The effectiveness of software designed to detect ai-generated writing: A comparison of 16 ai text detectors," *Open Information Science*, vol. 7, no. 1, Oct. 2023. DOI: 10.1515/opis-2022-0158.
- [110] A. Elkhatat, K. Elsaid, and S. Almeer, "Evaluating the efficacy of ai content detection tools in differentiating between human and ai-generated text," *International Journal of Educational Integrity*, vol. 19, no. 17, 2023. DOI: 10.1007/s40979-023-00140-5. [Online]. Available: <https://doi.org/10.1007/s40979-023-00140-5>.
- [111] J. Fleckenstein, J. Meyer, T. Jansen, S. Keller, O. Köller, and J. Möller, "Do teachers spot ai? evaluating the detectability of ai-generated texts among student essays," *Computers and Education: Artificial Intelligence*, vol. 6, p. 100 209, 2024. DOI: 10.1016/j.caeai.2024.100209.
- [112] F. Habibzadeh, "Gptzero performance in identifying artificial intelligence-generated medical texts: A preliminary study," *Journal of Korean Medical Science*, vol. 38, no. 38, e319, Sep. 2023. DOI: 10.3346/jkms.2023.38.e319.
- [113] D. Weber-Wulff, A. Anohina-Naumeca, and S. e. a. Bjelobaba, "Testing of detection tools for ai-generated text," *International Journal of Educational Integrity*, vol. 19, no. 26, 2023. DOI: 10.1007/s40979-023-00146-z. [Online]. Available: <https://doi.org/10.1007/s40979-023-00146-z>.
- [114] The Metropolitan Museum of Art, *Textile production in europe: Textiles, costume, and fashion*, [https://www.metmuseum.org/toah/hd/txtn/hd\\_txtn.htm](https://www.metmuseum.org/toah/hd/txtn/hd_txtn.htm), Accessed: 2024-01-25.
- [115] M. Ayala-Pazmiño, "Artificial intelligence in education: Exploring the potential benefits and risks," *593 Digital Publisher CEIT*, vol. 8, pp. 892–899, 2023. DOI: 10.33386/593dp.2023.3.1827.
- [116] A. Pramadi, M. Pali, F. Hanurawan, and A. Atmoko, "Academic cheating in school: A process of dissonance between knowledge and conduct," *Mediterranean Journal of Social Sciences*, vol. 8, 2017. DOI: 10.1515/mjss-2017-0052.
- [117] B. Abazi Chaushi, F. Ismaili, and A. Chaushi, "Pros and cons of artificial intelligence in education," *International Journal of Advanced Natural Sciences and Engineering Researches*, vol. 8, pp. 51–57, 2024.
- [118] H. Wu and T. Flanagan, "The limits of ai content detectors," *Journal of Student Research*, vol. 12, no. 3, 2023. DOI: 10.47611/jsrhs.v12i3.5064.
- [119] S. E. H. E.-O. Mohamed, N. S. A. Balol, and S. M. A. Hamza, "Reality of the real lecturer, tutor and teacher: A discussion paper," *International Journal of Linguistics, Literature and Translation*, vol. 5, no. 1, pp. 50–60, 2022, ISSN: 2617-0299. DOI:

10.32996/ijllt. [Online]. Available: <https://www.al-kindipublisher.com/index.php/ijllt>.