Accounting for student use of generative artificial intelligence in STEM education

Designing ethical and sustainability-driven course guidelines

Rik Baetens - 5619653
Master Thesis - Complex Systems Engineering and Management
To be defended on July 14 2026

Image source: (Tsolev, 2025)

TU Delft

Accounting for student use of generative artificial intelligence in STEM education

Designing ethical and sustainability-driven course guidelines

---

Master thesis submitted to Delft University of Technology in partial fulfilment of the requirements for the degree of

MASTER of SCIENCE

Complex Systems Engineering and Management
Faculty of Technology, Policy and Management
Delft University of Technology

Student name: Rik Baetens
Student number: 5619653
Thesis defence date: 14 July 2026

Graduation committee

Dr. Jolien Ubacht	Engineering Systems and Services	Chair
Dr. ir. Roel Dobbe	Engineering Systems and Services	First supervisor
Dr. ir. Fatima-Zahra Abou Eddahab-Burke	Multi-Actor Systems	Second supervisor

TU Delft Delft University of Technology

Preface

The thesis that lies in front of you has been written for graduation in Complex Systems Engineering and Management at Delft University of Technology. The dissemination showcases qualitative research developing guidelines in Science, Technology, Engineering, and Mathematics education to cope with students' use of GenAI. It thereby showcases my journey of the past five months, where I discovered how challenging the topic was, but also the impact it could make. The latter motivated me every time to continue, and I'm extremely grateful to share the work, hoping to support course managers at the TU Delft within the turbulent era of GenAI.

Moreover, I would want to show my appreciation for the guidance of my supervisors. Jolien, I would like to thank you for your supervision during the formal meetings we have had. Your role in other courses, such as the Master Thesis Research Design course, proved beneficial in guiding this thesis. Then, Roel, being supervised by you during my Bachelor's Thesis and allowing me to collaborate with you and others on research, have been valuable experiences. I am grateful to have had you on my side once again and want to thank you for all the effort you put into supervision. Fatima-Zahra, meeting you during the CoSEM Research Challenges course last year showed me your supervision skills. Once again, these proved to be very useful, for which I want to thank you. Then, thanks to all the interviewees who took part in this research. Without all of your help, the research would not be in its current state.

To conclude, this work also marks the end of my 5 years at Technology, Policy, and Management. This included a fascinating minor in Computer Science and challenging electives, which I am pleased to have followed. Also, along the way, I have met lots of people and made close friends, whom I sincerely want to thank for all the support and camaraderie over the past years. Thanks, and let's see how the future unfolds!

Rik Baetens
Delft, July 6 2026

Executive summary

Generative Artificial Intelligence (GenAI) plays a significant role in Science, Technology, Engineering, and Mathematics (STEM) education through students using it for their coursework. The technology has promising advantages, including saving time, a personalised learning experience, and a teacher or teaching assistant. Consequently, its high adoption in STEM education. Nevertheless, GenAI is an often intertwined concept; throughout the thesis, it is defined as a type of Deep Learning model that generates content based on a user prompt. Examples include text-assistants, ChatGPT, research assistants, NotebookLM, writing tools, Grammarly, coding assistants, GitHub Copilot and image generation, DallE.

However, this broad adoption of GenAI tools in education also has negative consequences. On one hand, ethical and legal problems arise around the way models are trained, collect and provide data, with their black-box nature. On the other hand, sustainability issues arise due to the high computational demand and its impact on skills such as problem-solving and critical thinking. Nonetheless, in STEM education, there is little support and understanding for lecturers to transform their courses to incorporate students' GenAI usage. Accordingly, lecturers lack an understanding of how to do so in combination with students' diverse behaviours due to a lack of standardisation on how to use GenAI tools pedagogically. Moreover, limited discussion has covered GenAI use on the course level. Together with the little attention on cognitive sustainability, the research objective is to create hands-on design guidelines for STEM courses to support course managers in addressing the GenAI usage of students for their coursework, ensuring constructive alignment is pursued. The corresponding main research question is:

Which ethical and sustainability-oriented design guidelines can inform STEM course design that addresses students' use of GenAI while retaining constructive alignment?

To gather an answer to the main research question, multiple methods have been adopted to ensure triangulation and adherence to Design Science (DS). First, a comprehensive literature review was conducted to establish a solid knowledge base grounded in scientific theories, following the rigor cycle. Next, semi-structured expert interviews were carried out to gain insights into the challenges faced by university course managers and to identify their needs during the relevance cycle. Subsequently, the design cycle began, developing guidelines and applying them to create a conceptual course redesign. These have been evaluated through another round of semi-structured expert interviews and changed accordingly to create the final set of guidelines with an implementation pathway.

The literature review highlighted various ethical and sustainability tensions. Ethical issues included plagiarism, cheating, and academic misconduct, as well as algorithmic unfairness and access inequalities. Sustainability concerns focused on ecological impacts from computational demands and the costs of training data centres. Additionally, cognitive sustainability was addressed, noting effects on Higher-Order Thinking Skills (HOTS) and potential neurological consequences. Moreover, through the literature review guidelines from previous Artificial Intelligence in Education (AIED) were identified in alignment with Constructive Alignment (CA) and a pedagogical framing. This included changes to assessment, such as integrating in-person assessment, learning objectives and activities through a focus on HOTS and Artificial Intelligence (AI) literacy.

Concerning the semi-structured expert interviews, the first round connected the theoretical tensions towards real-world perceptions of course managers. It was both an extension and a practical validation of the literature. Consequently, it highlighted monitoring and enforcement issues along with a lack of AI policy. In the meantime, through design requirements and course criteria, the Data Science in Technology, Policy, and Management (TPM) course was chosen to redesign according to the guidelines. This allowed a more tangible representation of the guidelines for course managers through CA, which formed the basis for the second round of interviews. The second round reflected the practical aspects of implementing the guidelines with their constraints and trade-offs, including feasibility and perceived appropriateness. In doing so, different codes were identified, ranging from making changes towards their own courses to student motivation. The ten final guidelines validated through this second round of interviews include:

1. Rethink foundational knowledge and skills*
For every course, identify the foundational knowledge and skills that students need to obtain and include those in the learning objectives to facilitate students in improving these skills. Examples: foundational programming knowledge and basic scientific literacy skills.

2. Adopt authentic assessments
For every course where this change is appropriate and feasible, incorporate authentic assessments that prepare students for the labour market while aligning with available course resources. Examples: replace a traditional report with one that includes a real-world scenario, as well as a midterm and final presentation or proposal.

3. Integrate on-site assessments
For every course where this change is appropriate and feasible, integrate an on-site assessment to assess foundational learning objectives and HOTS. Examples: an oral assessment or a physical closed-book exam. For larger courses, this may also include a lightweight pass/fail midterm or final or increasing the weight of the existing on-site assessment.

4. Ensure process over product within assessments
For every assessment where GenAI use is permitted, assessments should emphasise students' learning process alongside the final deliverable to reduce GenAI reliance and increase personalised guidance, if the course resources can support this change. Examples: drafts, proposals and presentations.

5. Enhance AI literacy†
For every assessment where GenAI use is permitted, courses should contribute to students' AI literacy development to foster responsible and effective use. Example: include a learning objective about fostering AI-literacy, which could involve how students should use GenAI tools or their general implications.

6. Provide rules on how to document GenAI use for assessment
For every assessment where GenAI use is permitted, provide clear guidance on how students should document GenAI use to increase reproducibility and reduce opportunities for academic misconduct through student accountability. Example: tool(s) used, prompts, purposes, and scope of integration.

7. Focus on developing HOTS
Courses should foster the development of HOTS to retain cognitive sustainability to benefit from GenAI as a scaffolded learning partner through critical engagement. Examples: balancing AI-integration with independent reasoning, discussions and reflective exercises to ensure students experience productive struggle.

8. Employ a human-centred approach to CA
Courses should employ a human-centred approach to CA that protects intellectual and social skills that may be compromised through GenAI use, if the course resources can support this change. Examples: integrate on-site assessments and class discussions to encourage meaningful interactions and enhance student engagement.

9. Create a fair learning environment concerning GenAI
For every assessment where GenAI use is permitted, foster students' understanding of how it may be used through equitable guidance and learning opportunities to diminish the skill gap between students. Examples: provide relevant content on the course platform or dedicate a section of the syllabus to explaining how to use GenAI for the course.

10. Keep the technology close
Currently, this is an institutional guideline for standardising an educational tool within the institution.

*For every guideline that changes a part of the CA triad: make the course compliant with CA by reflecting the change in the other parts of the triad

†As a caveat, within an educational institution, there is also a responsibility of increasing the AI literacy of lecturers.

Example: when an institutional GenAI tool is available, every course that permits GenAI use should align its guidance with that tool and encourage students to use it.

Altogether, these design guidelines provide an ethical and sustainability-oriented foundation for adapting STEM courses to address students' use of GenAI, while maintaining CA. The findings show that implementation is a multidisciplinary challenge. The guidelines must encompass both the technical aspects, which involve the infrastructure, and the societal considerations, which address high-level tensions. The literature has demonstrated that this is a complex issue, complemented by semi-structured interviews that reveal diverse perspectives among respondents, including views that differ from the literature, mainly rooted in the controversy around the topic. Outside of these tensions, the interviews also introduced practical issues of incorporating guidelines within an educational institution. Especially resource constraints, policy around course changes, and the controversy around integration of GenAI. Accordingly, the responsibilities were emphasised through a top-down approach. This example showed that the Executive Board of the university adopts the guidelines through an AI policy with a clear vision. To enhance flexibility, every faculty may write its own vision as an appendix to the central policy, tackling the multidisciplinary nature of the university. This enables the director of studies to be flexible and the policy to be practical for course managers.

A key academic contribution of this research is the creation of an artefact that addresses the gap in lecturers' understanding of how to manage the use of GenAI by students. This work also highlights a significant societal contribution by illustrating the responsibilities related to implementation and identifying practical actors within a STEM university. The main strength of this academic contribution lies in its comprehensive literature review, which integrates ethical considerations and sustainability challenges into specific themes. This approach facilitated a multidisciplinary perspective on the issue, incorporating often neglected insights from cognitive sustainability based on neurological research. Additionally, this thesis contributes to the discourse by including interview-based findings of course manager perceptions and institutional constraints when integrating guidelines concerning GenAI, ultimately aiming to advance research in AIED.

Contents

List of Figures

List of Tables

Acronyms

AI	Artificial Intelligence
AIED	Artificial Intelligence in Education
AIGC	Artificial Intelligence Generated Content
ANN	Artificial Neural Network
CA	Constructive Alignment
CoSEM	Complex Systems Engineering and Management
DS	Design Science
GenAI	Generative Artificial Intelligence
GPT	Generative Pre-trained Transformer
HOTS	Higher-Order Thinking Skills
ITS	Intelligent Tutoring Systems
LLM	Large Language Model
ML	Machine Learning
NLP	Natural Language Processing
RAG	Retrieval-Augmented Generation
STS	Socio-Technical System
STEM	Science, Technology, Engineering, and Mathematics
TPM	Technology, Policy, and Management

1 Introduction

This chapter offers the first insights into the dissemination, presented in five paragraphs. First, the motivation for choosing Generative Artificial Intelligence (GenAI) will be explained, providing the reader with essential knowledge about key terms. Then, §1.2 will outline the root of the problem, which will be translated into a research object in §1.3. Furthermore, §1.4 connects the thesis to the master's program Complex Systems Engineering and Management (CoSEM) while the final §1.5 serves as a reading guide throughout the thesis.

1.1 Motivation for Generative Artificial Intelligence

Although GenAI has been developed over several decades, a significant breakthrough occurred in late 2022 with the release of ChatGPT (Stephanou, 2025; OpenAI, nd). Its use is evident with a user base of 100 million users just after its launch (Hu, 2023), and a high adoption in Science, Technology, Engineering, and Mathematics (STEM) education (Dorji and Subba, 2025). Throughout the research, the term GenAI refers to: "a type of deep learning models trained to generate various types of contents conditionally based on the prompts" (Jung, 2023, p. 1039). Regarding its relevance, a characteristic is the prompting concept, which enables a user to guide a model's output. This allows STEM students to provide a prompt, which can be text, a document, or a picture, and receive generated content based on this user input. Such answers, often referred to as Artificial Intelligence Generated Content (AIGC), can be obtained from text, image, research, or coding assistants, as well as paraphrasing and writing tools. Consequently, students can use GenAI systems to support their daily coursework, influencing student learning and assessment. Hence, STEM education is transforming, where students who use this technology are exposed to its implications, leaving a substantial responsibility for course managers to adapt their courses.

To clarify the distinction between several often intertwined concepts, GenAI differs from Artificial Intelligence (AI), which can also refer to autonomous applications, and is defined as: "computing systems that are able to engage in human-like processes such as learning, adapting, synthesising, self-correction and use of data for complex processing tasks" (Popenici and Kerr, 2017, p. 2). Another popular term, distinct from GenAI, is foundation models, defined as "any model that is trained on broad data, that can be adapted to a wide range of downstream tasks" (Bommasani et al., 2021, p. 3). Lastly, a Large Language Model (LLM) is a subset of foundation models specifically focused on language-based tasks. Altogether, for example, the Generative Pre-trained Transformer (GPT) models are a class of LLMs that power applications such as ChatGPT. ChatGPT itself is a GenAI application built on top of one or more foundation models, which are themselves AI. Consequently, focusing on LLMs would neglect certain multimodal capabilities, such as text, image, code, and audio generation, offered by current applications like ChatGPT, NotebookLM, or GitHub Copilot. Meanwhile, foundation models or AI would be too broad, exceeding the scope. An overview of these different technological concepts is presented in Figure 3.2.

1.2 Motivation for STEM education

Throughout the thesis, the focus is placed on STEM education. One important reason is that GenAI tools such as ChatGPT are adopted extensively in STEM disciplines (Dorji and Subba, 2025). As a result, many opportunities and tensions surrounding the adoption of these tools have emerged in this educational context. Therefore, existing literature forms a strong foundation for analysing these opportunities and tensions within this form of education. In the context of literature, a scoping review conducted during the research proposal indicated that significant work has been conducted on Artificial Intelligence in Education (AIED), particularly in STEM education and programming. This establishes a solid academic foundation for this thesis and enhances access to relevant participants for data collection and evaluation due to the researchers' background.

Furthermore, Wang and Fan (2025) argue that GenAI tools like ChatGPT have the largest effect within STEM courses. They reason that this is due to the inherently complex nature of these courses. Such tools can, therefore, be utilised to explain a topic from various perspectives and generate comprehensive examples. At the same time, a study by Wang et al. (2023b) on AI assistants in computer science, highlighted that lecturers were willing to change their courses,

but lacked proper guidance. This highlights a clear gap within STEM education. Lastly, GenAI remains a technical application, as discussed in Section 3.2. Accordingly, STEM education provides a suitable environment for evaluating and testing interventions aimed at addressing the lack of guidance among course managers.

1.3 Problem statement

GenAI is used on a large scale by STEM students (Dorji and Subba, 2025), and impacts STEM and related courses the most (Wang and Fan, 2025). While it offers tools with great advantages, such as enhanced creativity and faster design iterations (Sandhaus et al., 2025), it also has negative consequences. Sustainability-wise, the bigger-is-better mindset (Ananthaswamy, 2023) has led to the creation of large models with billions of parameters (Raffel et al., 2023). Since both the training and utilisation of these models are closely tied to high energy consumption (Fontaine et al., 2024), this growth needs to be reconsidered. Concerning ethics, how the models are trained, collect and provide data is leading to privacy, data and equity concerns (Mattiello et al., 2025). Moreover, their complex architectures can be challenging to comprehend, which complicates individuals' ability to form their own opinions when evaluating these systems (Ferrer et al., 2021). In STEM education, this lack of understanding may undermine student performance due to over-reliance on falsified information Dong et al. (2025); Wang et al. (2023b). Additionally, there is great concern about the devaluation of degrees (Chan and Colloton, 2024; Staro minski-Uehara, 2026).

Still, in STEM education, there is limited guidance for course managers to adapt their courses based on GenAI usage by students to pursue learning goals (Wang et al., 2023b). Consequently, lecturers have a limited understanding of doing so (Prather et al., 2025; Nguyen, 2025; Rahman et al., 2025), while students show diverging behaviour due to a lack of frameworks on how to use GenAI (Lont, 2026; Ala et al., 2025). Additionally, research has already suggested incorporating AI ethics and related courses; there is limited discussion on how other courses should adapt to the rising use of GenAI by students. Hence, it is essential to shift from a curricular level of depth to a course-level approach. This will enhance generalisability since most research focuses on ChatGPT or specific educational fields, such as computer science. Lastly, integrating the cognitive sustainability implications of GenAI usage is crucial as extensive usage suggests a decrease in learning skills (Kosmyna et al., 2025).

1.4 Research objective and main research question

The thesis objective is to create hands-on design guidelines for STEM courses to support course managers in addressing the GenAI usage of students for their coursework, ensuring constructive alignment is pursued. This includes using such tools responsibly from ethical and sustainability perspectives. To achieve this, throughout this thesis, the following research question is presented:

Which ethical and sustainability-oriented design guidelines can inform STEM course design that addresses students' use of GenAI while retaining constructive alignment?

1.5 CoSEM suitability

For the master's program CoSEM, an integral part is about assessing the dynamic components of a Socio-Technical System (STS). These components are technical, institutional, economic and social aspects. Throughout the thesis, STEM education represents the STS, combining both societal aspects regarding educational value and technical aspects of designing and GenAI tools used within education. Concerning the technical components, the thesis assesses the GenAI tools used by STEM students, discovering their ethical and sustainability implications. Institutional and economic components relate to the rules and regulations. This includes both the use of these systems, the course requirements and the implementation of the artefact through institutional regulations. Social aspects encompass STEM students using GenAI, as well as lecturers who are indirectly influenced by their behaviour and learning methods.

1.6 Reading guide

The thesis is structured as follows: Chapter 2 highlights the research approach, which is the foundation of the dissemination. Then, Chapter 3 showcases state-of-the-art literature to understand how AIED evolved, the technology behind it, and why students use it within their courses. Chapter 4 describes current ethical and sustainability tensions of GenAI use in education. Next, Chapter 5 dives into contemporary pedagogical lessons of AIED, and translates this, with the tensions of the previous chapter, into a descriptive framework and design requirements. Chapter 6 discovers the last pedagogical lessons, before the set of guidelines, and the conceptual course design itself are constructed and discussed. Chapter 7 presents semi-structured interviews that explore course managers' perceptions of the use of GenAI in education, as well as their pedagogical views on the guidelines and design requirements. Following that, Chapter 8 analyses the results from various sections of the thesis with the limitations and recommendations for future research. Finally, Chapter 9 addresses the main research questions and sub-questions, combined with the academic and societal relevance of this work, a final reflection on the connection with the CoSEM programme, and an implementation pathway for educational institutions.

2 Research approach

Within this chapter, the approach towards the main research questions presented in Section 1.4 is showcased. To do so, §2.1 explains the design approach taken to answer the main research question. In addition, §2.2 outlines the sub-questions, which are smaller inquiries derived from the main research question, along with the methods used to address them and their limitations. Section §2.3 presents the research approach, where the sub-questions, methods, and deliverables are illustrated in relation to one another in the research flow diagram. Ultimately, §2.4 shows the process of the scoping literature review, which formed the basis largely for Chapter 3, Chapter 4, and Chapter 5.

2.1 Design approach

Based on the research question, one can consider either a design-oriented approach or an exploratory approach. The exploratory approach aims to generate insights into the sustainability perspective, as this area has been underexamined in current AIED research. In contrast, the objective of this research extends beyond merely analysing the contemporary situation; it involves developing and evaluating an artefact that will assist course managers in addressing the use of GenAI by students in their coursework. While it is important to acknowledge gaps in previous analyses, the primary focus of this thesis is to establish an artefact. Therefore, the thesis adopts a Design Science (DS) approach.

In DS, it is vital to communicate results with an end-user (Hevner et al., 2004). In this thesis, the end-users are course managers, since they can adapt courses based on the final design guidelines. Sometimes, the term "lecturers" is used to refer to all individuals who teach in a course, not just the one responsible for the course design. They are affected by changes in the course management, but more importantly, they are also influenced by tensions arising from student use of GenAI. DS research can be understood through three research cycles, as shown in Figure 2.1. This includes the Relevance Cycle, which connects the field context of the research with the DS activities. The Rigor Cycle links the DS activities with the knowledge base informing the research project. Finally, the central Design Cycle iterates between building and evaluating the design artefacts and processes (Hevner, 2007). Below the figure, each cycle will be connected to the research.

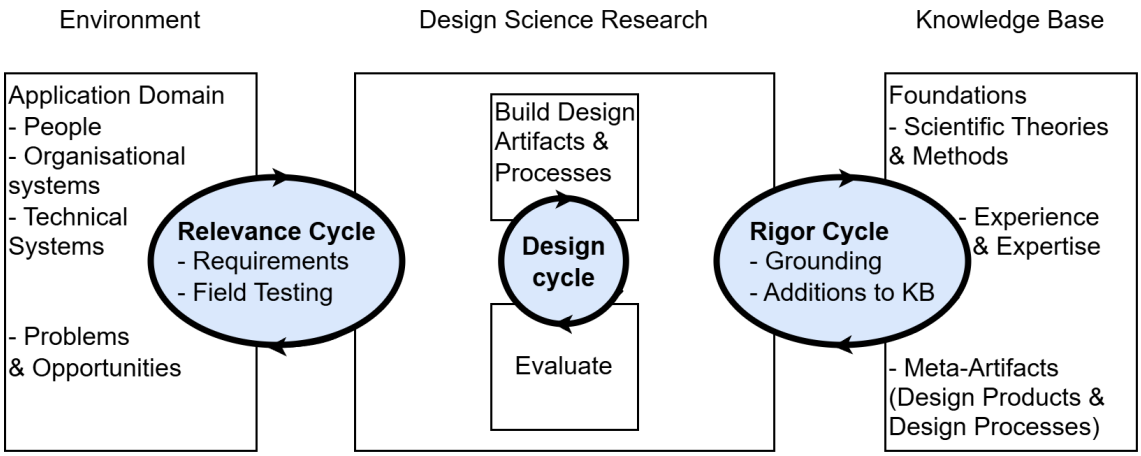

Figure 2.1: The DS Research Cycles based on (Hevner, 2007)

Effective design science research integrates relevance and rigor, as defined by Hevner (2007), by beginning with the identification of opportunities and challenges within real-world application environments. Hence, he describes both the requirements for the research and the acceptance criteria for evaluating the research results that need to be obtained within this cycle. This research does so through two rounds of semi-structured expert interviews with professors of the TU Delft. Course managers are involved in practical aspects of the issue and possess the educational expertise to tackle problems and opportunities related to the use of GenAI by students in STEM education. Moreover, since they are the end-users of the artefact, they can provide design requirements for the final design guidelines and evaluate them accordingly. While the Executive Board is generally responsible for expressing university-wide solutions through policy, this research will concentrate on

course managers and the course-level, as they can translate these solutions into course modifications.

Hevner (2007) defines two types of additional knowledge in the knowledge base: experience and expertise, defining state-of-the-art in the application domain of the research and the already existing artefacts and processes in this domain. Throughout the research, this is grounded in the extensive literature review. It will explore the ethical and sustainability tensions surrounding the use of GenAI by STEM students and extend previous AIED research to identify established course design guidelines. This ensures that the produced design will conform to guideline four, necessitating research contributions over routine designs (Hevner et al., 2004). This stage also includes meta-artefacts, which serve as a descriptive framework that connects key concepts from the literature review with a STEM course design. The combination of the relevance and rigor cycles facilitates a close connection to real-world application, allowing theoretical insights to be practically operationalised.

Within the middle block, requirements and input are gained from the relevance cycle, and evaluation theories and methods from the rigor cycle (Hevner, 2007). There is a clear emphasis on balancing designing and evaluating, so one does not spend all their time on either of them. In the research, this is evident through the conceptual course design and the second round of semi-structured expert interviews. Constructing the course design will demonstrate a balanced approach to design and evaluation by showing how the design guidelines can be applied to a course. Additionally, interviews will include another layer of validation as part of the ex-ante evaluation, crucial in DS (Hevner et al., 2004), to discover the pedagogical feasibility and validity of the design guidelines. This cycle provides room for iteration to create a fine-tuned artefact tailored to stakeholder needs.

2.2 Sub-questions and methods

To address the main research question, it has been broken down into smaller sub-questions. This approach allows for structured and manageable steps toward ultimately answering the main question. Each sub-question will include a brief explanation of its purpose, expected outcome, and the method used, which will help guide the understanding of the subsequent sub-question.

1. *How have the historical development of AI in education and technological characteristics of GenAI shaped its educational relevance for student use in STEM education?*

To start, the thesis examines the history of AI in education and the technology behind contemporary GenAI models. After doing so, it discusses why it is currently relevant for students in STEM education to use them for their coursework. Additionally, the history of AI in education will broaden the perspective of the researcher to be aware of earlier technological crises, such as the AI winter and dot-com bubble (Guest et al., 2025). Moreover, motives behind the first AIED developments can be discovered. This is combined with analysing the technological characteristics that explain high-level implications and give brief insights into how the models work to enhance understanding. This question will be assessed through a literature review, which will derive these drivers of the current adoption of GenAI by students. Overall, this offers insights into why and how we arrived at the current popularity of GenAI tools among STEM students.

2. *Which ethical and sustainability tensions arise from student use of GenAI in STEM education?*

To understand the necessity of addressing the current issue of high adoption, the ethical and sustainability concerns arising from students' use of GenAI are evaluated. A literature review reveals these issues in the current literature and connects them with the perspectives and examples of course managers. These perspectives are obtained as validation of scientific findings, not as additional knowledge extraction, through the first round of semi-structured interviews. The coding for the interviews will be inductive, with codes being derived directly from the data (Hecker and Kalpokas, nda). The interviewees for the semi-structured interviews were gathered through the researchers' network and extended via snowball sampling from those respondents. Together with sub-question 1, this dual approach provides a historical and technical foundation for the thesis while framing the research focus on the educational influences of GenAI. In doing so, it leads to a clear problem specification about what kind of tensions are present through an ethical and

sustainability viewpoint. Moreover, this may reveal practical implications arising from these tensions.

3. *Which pedagogical themes from prior AIED research can inform STEM course design in response to the ethical and sustainability tensions?*

Following the problem specification, the research examines concurrent literature on guidelines in education to identify pedagogical lessons, ensuring that it does not reinvent the wheel. A literature review is conducted to identify scientific lessons that will serve as a foundation for the design requirements. To expand on this foundation, the first round of semi-structured interviews was used to gather practical examples that inform the rationale behind the design requirements and as a check on the scientific findings. These insights are integrated with prior findings to create a descriptive framework which showcases connections between the tensions from sub-question 2 and STEM education through its pedagogical lessons. Furthermore, this framework highlights a pivoting point where uncontrollable tensions for course managers and education in general are discarded.

4. *How can the identified design guidelines be demonstrated through a conceptual STEM course design that integrates GenAI ethically and sustainably?*

This stage focuses on redesigning a course based on structural requirements and the preliminary guidelines. Constructive Alignment (CA) serves as an overarching guideline, as it is fundamental to course design and aids students in learning by following the proposed course process (Biggs, 2003). Therefore, it already assesses the feasibility of integrating the guidelines within an existing course while ensuring CA is retained. Thereby, it addresses the practical challenge faced by course managers of transforming guidelines into course modifications. In doing so, the course redesign demonstrates how the identified guidelines can be operationalised within a STEM course context. Moreover, this approach enables the course redesign to serve as a basis for evaluating the behavioural requirements of perceived feasibility and pedagogical suitability with the guidelines through semi-structured interviews regarding the course design.

5. *How do university course managers perceive the feasibility and pedagogical appropriateness of the proposed design guidelines, as reflected through the course design?*

Ultimately, a final set of guidelines can emerge by integrating the last findings from the second round of semi-structured validation interviews, as indicated in the last sub-question. These interviews will be the final validation method to ensure the guidelines are feasible, pedagogically appropriate, and align with the needs of course managers. The interviewees for the semi-structured interviews are selected based on the researchers' network and extended through snowball sampling from those respondents. The coding for these interviews will be deductive, where the guidelines and design requirements serve as predefined codes (Hecker and Kalpokas, nda). These sub-questions together address the main research question and help validate both some design requirements and the design guidelines.

2.2.1 Limitations and mitigations of the research methods and approach

This subsection discusses limitations associated with the selected data collection methods identified during the research design phase and DS as the research approach. These methodological limitations differ from those addressed in Chapter 8, which reflect issues encountered during the research. By addressing these limitations in advance, precautionary measures can be implemented to mitigate some of their effects. The literature review covers a significant portion of the first three sub-questions. This approach ensures that the research is firmly rooted in scientific theory, in accordance with the rigor cycle of DS. However, this leads to a substantial amount of information being obtained within this cycle, which may inherit publication bias present within the body of literature. Such bias is present in work that focuses on positive results. Hence, the researcher needs to be aware of this potential bias, identifying it where possible and tempering certain claims. Furthermore, the field is rapidly evolving, which is why non-traditional publication channels, such as those from arXiv, can be part of the review. Another limitation of the literature review is the search strings, which may introduce selection bias by influencing the studies retrieved. This is planned to be mitigated by including as many relevant synonyms as possible and employing two search strings.

In addition, the scientific base from the literature review may be incomplete due to a lack of practical stakeholder perspectives. Therefore, this large scientific foundation will be complemented by semi-structured interviews with STEM course managers. In this manner, combining findings from literature and interviews facilitates methodological triangulation. This approach may help reduce biases present in the existing literature, but it could also introduce new ones. One such bias arises from the researcher's positionality as a student, which can create challenges during interviews (Johannesson and Perjons, 2014). Additionally, the interviews may be influenced by social desirability bias (Hecker and Kalpokas, ndd), particularly since the questions focus on contemporary processes within the universities where the interviewees work, and the interviewer effect (Hecker and Kalpokas, ndc). To foster trust, a planned mitigation is to anonymise transcripts to the best of the researcher's ability. Furthermore, all transcripts and codes are planned to be returned to interviewees within two weeks for review, and the interview questions formulation as neutrally as possible (Hecker and Kalpokas, ndb).

Another issue with interviews is that respondents may alter their answers because they are aware that they are part of a study, which can make it difficult to gather truly authentic data (Hecker and Kalpokas, ndb). In this case, the same mitigation applies, since the goal is to create an environment favourable to open and honest responses. Lastly, the paradox of self-disclosure may occur, where interviewees slowly share more as the interview proceeds, and they get more comfortable (Hecker and Kalpokas, ndb). To ensure that respondents feel comfortable with the use of their data, all interview notes and transcripts will be sent back to each interviewee for review before being included in the thesis.

Lastly, DS research also has its limitations. Since the artefact implementation will be tailored to the educational context of the TU Delft, this may limit generalisability to other institutions. Therefore, as an initial step to improve generalisability, guidelines from Leiden University were also analysed during the process of creating the guidelines. Furthermore, the guidelines are based on an extensive literature review, grounding the artefact rationale in scientific theories that apply to multiple contexts. Additionally, DS involves design choices which are dependent on the researcher's positionality. To mitigate this influence, design choices are grounded in analytical frameworks, requirements, and semi-structured interviews.

2.3 Research flow

The research flow diagram, illustrated in Figure 2.2, outlines each research phase, including the corresponding research question, method, and input or output produced by the method. By doing so, it reflects the different phases of the DS research approach and oversight of this research.

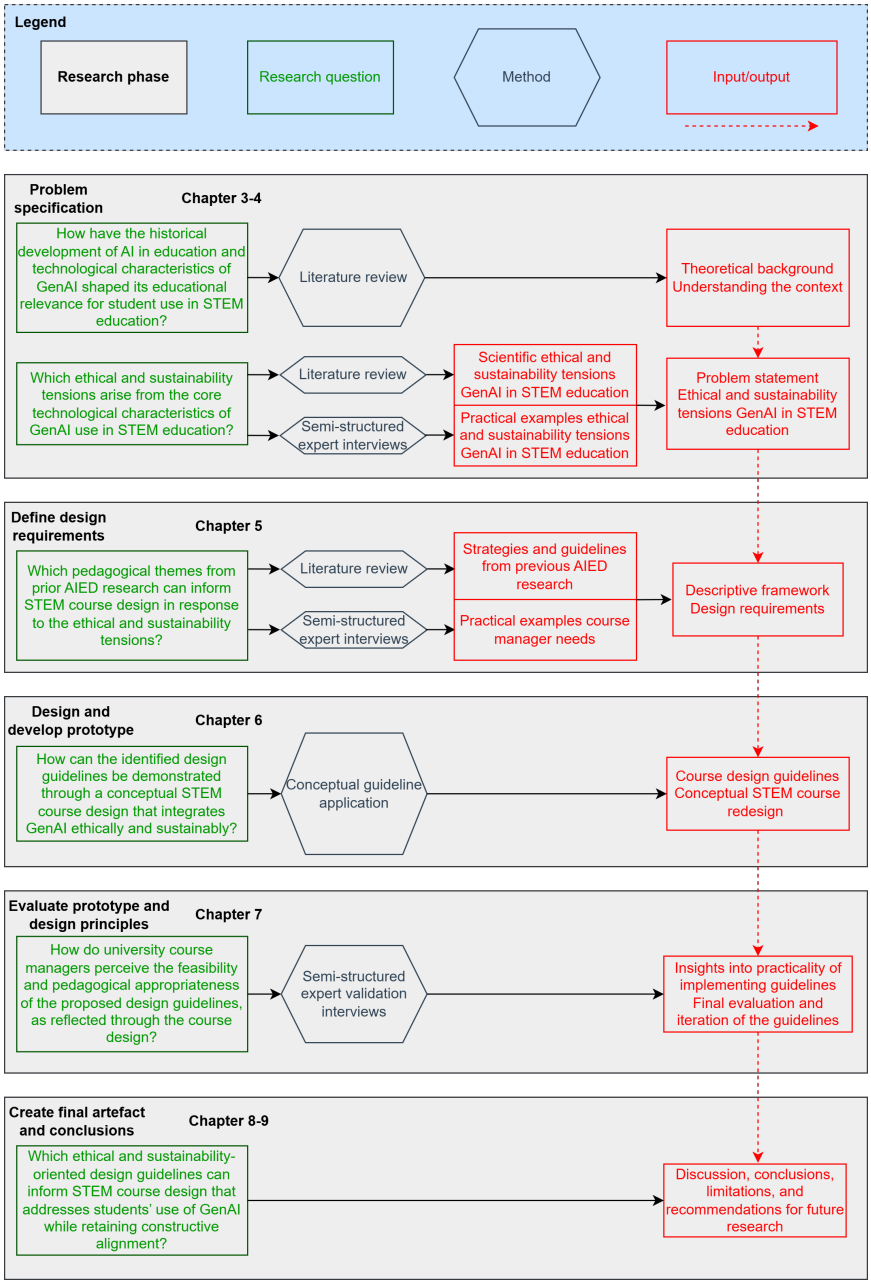

Figure 2.2: A visualisation of the full research flow of the thesis

After the identification of the research problem and defining the research approach, a literature review forms the informative foundation of the thesis, supported by semi-structured expert interviews. Firstly, the review discovers the necessary background information to understand how contemporary GenAI models work, the history of AIED and why students should use them (Chapter 3). Subsequently, the literature is utilised to examine the tensions that emerge in STEM education through students' use of GenAI to specify the problem (Chapter 4). Semi-structured interviews will ensure the problem is validated by mapping real-world examples from course managers onto the scientific tensions. Subsequently, these findings will be structured into a descriptive framework, using insights from the literature review on previous AIED strategies and guidelines as the foundation (Chapter 5). Within this chapter, the end-user findings from the first round of semi-structured interviews are combined with literature findings to create relevant design requirements. Next, the final pedagogical insights are drawn, the guidelines are created and used to redesign a STEM course based on these guidelines (Chapter 6). Following, another round of semi-structured interviews will validate the feasibility and appropriateness of the guidelines, as a basis for a final guideline iteration (Chapter 7). To conclude, the findings are discussed, conclusions drawn, limitations addressed, and future research proposed (Chapter 8 and 9).

2.4 Method of scoping literature review

For conducting the scoping literature review, an adapted version of the PRISMA approach was used (Page et al., 2021). Accordingly, the search process consisted of an *Identification*, *Screening*, *Eligibility*, and *Included* phase, as shown in Figure 2.3. The search process took place from March 18, 2026, to April 6, 2026.

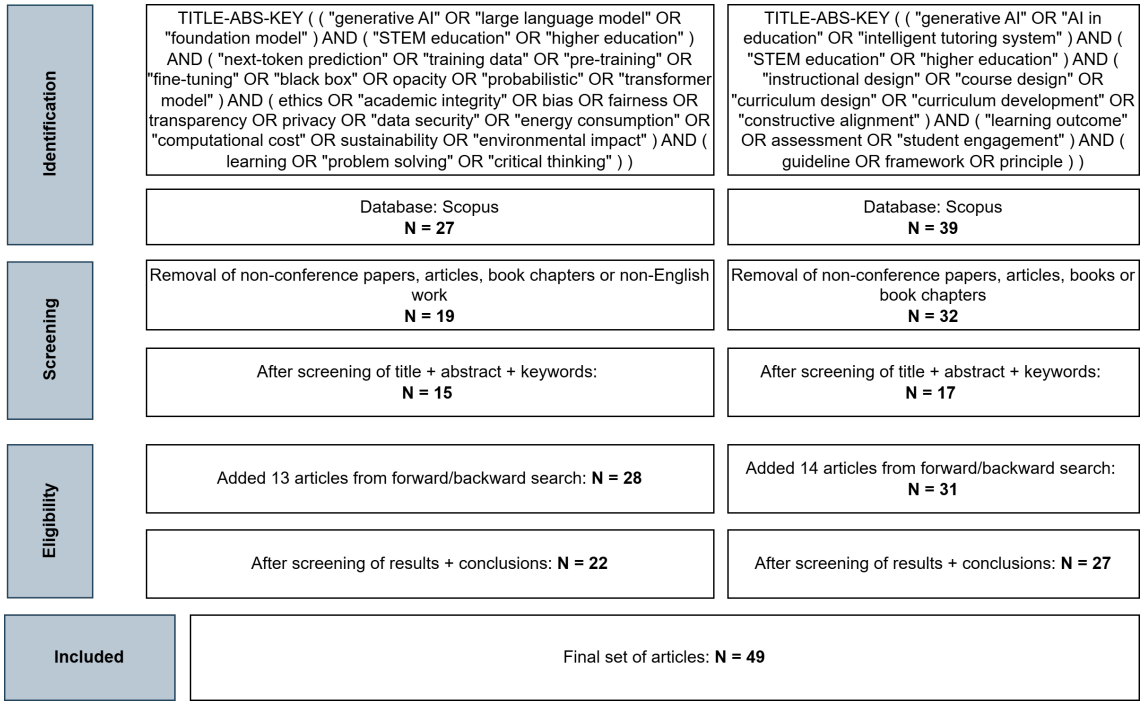

Figure 2.3: Scoping literature review selection process with adapted PRISMA approach

The selection process started by creating two search strings, shown in Appendix C: Search Strings. The first addressed the tensions caused by students' use of GenAI, while the second focused on existing guidelines to adapt education. Since literature employed multiple synonyms for most of the terms, these were also accommodated in the search strings. This held 27 and 39 results in Scopus. The initial screening focused on filtering reviews and conference reviews or non-English content. Further screening has been done based on the title, abstract and keywords to ensure compliance with the research context. Then, papers were added through snowballing based on the screened papers. The screening criteria are presented below. Lastly, after the final screening of results and conclusions, the final set of 49 papers emerged.

The search criteria for the screening and eligibility phase were as follows. The first two criteria were applied to papers from both substrings. The third criterion focuses on identifying ethical or sustainability tensions associated with GenAI related to the first search string, while the fourth criterion addresses discovering existing educational design responses that relate to the second search string.
- **Perspective**: The papers should focus on problems that arise from students' use of GenAI in their coursework. This meant that papers were excluded if they focused solely on integrating GenAI tools, unless the work addressed AI detection tools.
- **Type of education**: Studies should focus on STEM education or on educational and pedagogical approaches with clear relevance to STEM pedagogy. Therefore, papers that focused exclusively on non-STEM disciplines without a clear connection to STEM education were excluded.
- **Framing of the issues**: Papers should focus on either ethical or sustainability issues related to GenAI. Preference was given to studies that discussed how this affects students.
- **Solution space**: The focus of the paper should be on educational design related to learning outcomes, activities or assessment in the form of guidelines, frameworks, principles or similar design-oriented approaches. Studies that directly connected this towards GenAI use by students were given preference.

It is important to note that an additional database was not consulted, as the sources from Scopus provide sufficient information to address the sub-questions. Furthermore, to prioritise primary research contributions, literature reviews and conference reviews were excluded from the analysis. Since the oldest paper included in this review was published in 2023, there was no need to filter by publication year concerning the novelty of the topic. Finally, because the field is advancing rapidly, papers from non-traditional publication channels, such as arXiv, were also included.

3 Understanding GenAI use in STEM education

This chapter establishes a foundation for understanding the technical aspects of GenAI, its origins, and its significance in STEM education. The corresponding sub-question is: *How have the historical development of AI in education and technological characteristics of GenAI shaped its educational relevance for student use in STEM education?* Throughout this chapter, state-of-the-art literature guided this investigation. First, §3.1 explains the history of AI in education, showcasing technological developments over time. §3.2 addresses the technology behind concurrent GenAI tools, proposing a general description of content generation. Moreover, §3.3 presents the reasoning behind the relevance of students adopting GenAI. To finish, a sectional conclusion will discuss sub-question 1, as defined above. By doing so, it created a broad base to understand the tensions discussed in Chapter 4.

3.1 History of AI in education

The following subsection aims to showcase the history, including the earlier visions behind the technology. By doing so, it grounds contemporary GenAI models within the evolution of AIED, to understand both the intentions and growing capabilities. The history of AI in education shows the gradual improvement of technology and education goals, as shown in Figure 3.1. This dates back to the early 1950s when linear programs demonstrated step-by-step instructional sequences, which encountered issues with individualisation (Nwana, 1990). As the paper discusses, the following technologies were branching systems in the 1960s. These still had a certain number of fixed frames, but were able to select a frame based on the students' response. Notably, even back in the late 1960s and early 1970s, there were ideas of generating teaching materials through such systems. Also, during this period, concerns were about the growing dependency on science (Weizenbaum, 1976). However, a large group believed in intelligent tutoring (Nwana, 1990). According to the author, this meant that the computer would serve as a tutor, often referred to as Intelligent Tutoring Systems (ITS), combining both computer science, psychology and education and training. This movement led eventually to the first ITS, which were rule-based systems with adapted learning for specific knowledge domains (Adamakis and Rachiotis, 2025), as reflected in Figure 3.1.

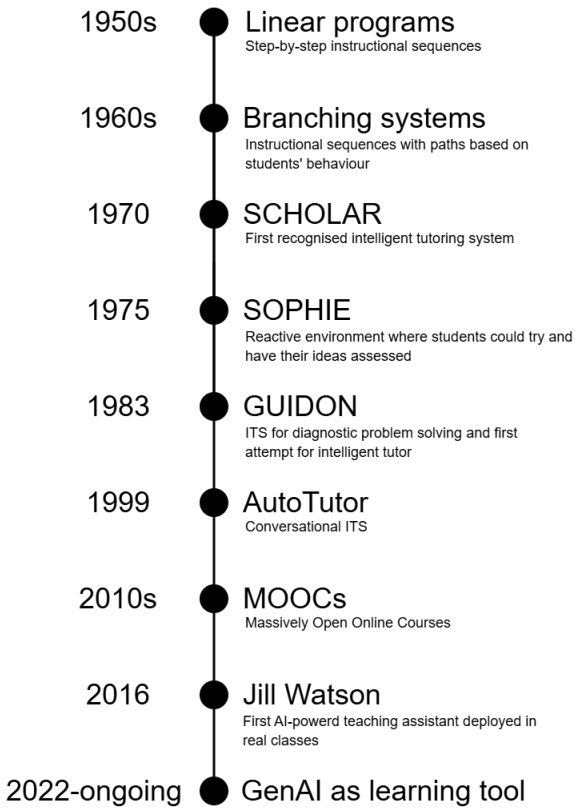

Figure 3.1: Relevant milestones in the history of AI in education, adapted from (Adamakis and Rachiotis, 2025)

This paragraph gives a brief overview of the most influential ITS in the history of education. The first recognised ITS was SCHOLAR, which was able to handle student questions about South American geography and able to generate instructional material (Carbonell, 1970). Then came SOPHIE in 1975, which was a system that could help to create experiments and debug its own ideas (Brown et al., 1975). Afterwards, GUIDON was constructed in 1983, enabling physicians to ask questions about a patient's case, which would then be critiqued (Clancey, 1983). To come even closer to human tutoring, AutoTutor was introduced in 1999, designed to assist students in an introductory computer literacy course through a talking head which served as a conversation partner (Graesser et al., 1999).

The next period emerged through significant advancements in Machine Learning (ML), data availability, Massively Open Online Courses, and Natural Language Processing (NLP), leading to personalised learning, automated grading, and plagiarism detection (Adamakis and Rachiotis, 2025). A remarkable use of AI in education was teaching assistant Jill Watson, based on IBM's Watson platform, able to answer students based on frequently asked questions (Goel and Polepeddi, 2016). Another influential moment was the creation of the GPT, being the first model based on the transformer architecture (Vaswani et al., 2017). These models amplified capabilities in natural language understanding and generation (Guo et al., 2025). However, the large turning point came with the release of ChatGPT in late 2022 (Stephanou, 2025). It represented the first GenAI tool that was available to a large audience and capable of generating academic-level output based on user prompts (Adamakis and Rachiotis, 2025). The authors argue that this change meant that AI usage in education shifted from the background towards an everyday tool used by students and lecturers.

Understanding the evolution of AI in education allows us to appreciate the early goals and technological advancements of these systems over time. This perspective helps us avoid a chrono-centric view that considers the current era as the most significant in history (Mishra et al., 2012). From systems with fixed frames, adaptive frames, towards ITS, AI teaching assistants, and finally

GenAI tools. While the goal of supporting human tutoring remains unchanged, the diversity of model capabilities has evolved. Where the early models were text-based and contained a single domain of knowledge, GenAI tools such as ChatGPT are capable of generating text, images, videos, and documents in every domain. Hence, current tools represent a stage in the speeding evolution of learning tools to support the learning process, which is crucial for governing them responsibly and effectively in STEM education due to its implications.

3.2 GenAI and its technology behind it

After discussing the technological improvements and timeline of AIED, it is essential to address some technical considerations. As discussed to be important in literature (Prather et al., 2025), this subsection tries to familiarise course managers with the technology behind popular GenAI tools. It does so by briefly discussing the most vital mechanisms powering these tools. Understanding the origin of model behaviour allows lecturers to comprehend the tensions and evaluate appropriate usage, which will be discussed further in Chapter 4. However, before discussing any technicalities, Figure 3.2 shows an overview of important concepts as an extension of the presentation in Chapter 1.

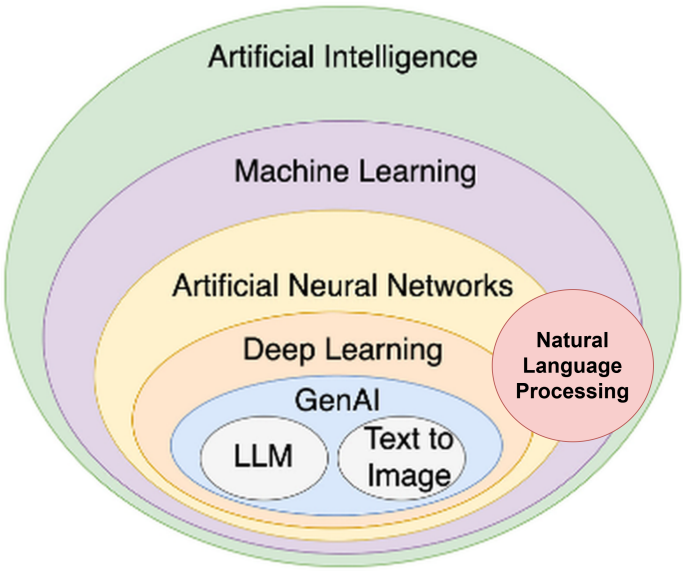

Figure 3.2: AI and its subdomains, adapted from (Chan and Colloton, 2024)

As can be seen in the Figure, the umbrella term is AI. Within this domain, a subfield is machine learning, which aims to enable machines to learn from data, thereby improving their performance on specific tasks without being programmed specifically to do so (Chiu, 2024). Going deeper are Artificial Neural Network (ANN)s, which withholds computational models inspired by the functioning of the human brain, with its synaptic connections between neurons (Chan and Colloton, 2024; Holmes et al., 2023). A class of ANNs are Deep Neural Networks, characterised by a significant amount of hidden layers, which form the foundation of Deep Learning. As mentioned in the introduction, GenAI is a type of deep learning model that generates content based on prompts (Jung, 2023). Moreover, NLP is a subfield of AI that focuses on language-related tasks, enabling computers to understand and generate human-like language (Bommasani et al., 2021). To end, the final layer includes LLMs such as GPT, which is a subclass of NLP models. GPT, the foundation of ChatGPT, is primarily focused on generating written text and has multimodal capabilities, including text-to-image generation.

In principle, the process of a text response to a prompt can be assessed in six different steps as presented in (Holmes et al., 2023). Initially, the user prompt is divided into smaller pieces, called tokens, which serve as the input for the GPT. Then, the GPT works as a decoder-only transformer to be able to predict words based on prior information. Afterwards, these predicted words are converted into readable text. This text is filtered through guardrails that ensure offensive

or unethical information is removed. Next, the previous three steps are repeated until an entire response is constructed. The last step, referred to as post-processing, ensures the text is readable by applying formatting, punctuation and other enhancements.

This paragraph goes deeper into how the steps listed above occur. Therefore, it aims to provide more context on how next-word prediction works, helping to understand, for example, why hallucinations can occur. At first, one might wonder how the input tokens are fed into the GPT and what happens afterwards before a word can be predicted. This process involves encoding each token and mapping it into a matrix so that the model can interpret the token values (Chan and Colloton, 2024). Additionally, the same study states that it incorporates positional encoding to determine the position of each token within the original input sentence. Then, the next phase includes embedding. Embedding ensures that the token-matrix constructed prior is mapped into a vector (Chan and Colloton, 2024). The study also mentions that such an embedding space is high-dimensional, where words with a similar definition are close to each other, helping the model to 'understand' text. Although the words within this space do not have a 'real' meaning, logical constructs can still be made, for example, King - Man + Woman = Queen. Next, the vector representations of the words are used to construct scores for certain words to be able to determine which one fits best in the context (Chan and Colloton, 2024). This is combined with the attention mechanism to identify long-range dependencies between words (Vaswani et al., 2017). For example, if earlier in the sentence a word is plural, 'was' must become 'were'. The combination of multiple of these attention heads allows the generation of scores through query, key and value parameters to determine the mathematically best successive word (Chan and Colloton, 2024).

To conclude, this technical subsection gives a general understanding of the mechanisms behind contemporary GenAI such as ChatGPT. For instance, the stochastic nature may suggest why hallucinations occur. This emergent effect is exacerbated by junk in the database, which is not 'fact-checked' by the model. It also provides insights into accepting the average AIGC with language becoming more generic, leading to the normalisation of mediocrity (Deng and Joshi, 2024). Also, it may give insights into how to teach the usage of GenAI effectively through prompt design techniques such as the tree of thoughts prompting, to utilise the underlying transformer architecture effectively (Prather et al., 2025). Moreover, the authoritative way of representing input to the user, even when it does not have a deep understanding of the subject matter (Rudolph et al., 2023), may lead to over-reliance on the models. This may lead to students being less likely to critically assess content and learn false information (Bull and Kharrufa, 2023). In its usage, many more tensions arise, which will be discussed in greater detail in the Chapter 4.

3.3 GenAI relevance for students in STEM education

To conclude this chapter on the educational significance of using GenAI, this subsection discusses why its adoption in education is beneficial, despite the argument that institutions can entirely prohibit its use. First of all, GenAI offers lots of benefits for both students and lecturers. Mentioned often as the predominant benefit is time saving (Randall et al., 2023; Kasneci et al., 2023; Cotton et al., 2024; Elkhodr and Gide, 2025). Although it seems natural for students to use it for this purpose, this can be tricky due to over-reliance and a loss of cognitive abilities, discussed more thoroughly later in Section 4.2.2. Another frequently mentioned benefit is a personalised learning experience, due to the flexibility of interaction with the tools (Randall et al., 2023; Bommasani et al., 2021; Zhai, 2022; Prather et al., 2023; Bollas and Bennett, 2026). Moreover, carrying the ideas behind early ITS, these tools can also serve as a teacher or teaching assistant (Randall et al., 2023; Zhai, 2022; Tran et al., 2025). Also, the philosophy of early ITS SOPHIE 'learn by doing' (Nwana, 1990) is seen through current models, allowing students to learn through experimentation and experience (Rudolph et al., 2023; Bollas and Bennett, 2026). It should also be noted that GenAI tools are widely adopted in the working field and are offered by academic publishing companies, such as Elsevier, through their LeapSpace (Elsevier, nd). Therefore, using it is valuable for industry and career readiness (Prather et al., 2025; Chan, 2023), averting the accumulation of the gap between education and industry.

Secondly, prohibiting would harm motivated students, since they will try not to use it, while unmotivated students do. This will exacerbate inequities between students (Randall et al., 2023;

Feng et al., 2024; Cotton et al., 2024). Many students will continue to use it despite prohibitions (Randall et al., 2023), and reliable monitoring is currently not achievable with existing detection tools. Besides that, regarding the learning goals for a course, it is better for course managers to clearly articulate how to use GenAI to achieve the benefit of the course rather than students deciding on their own. The latter will otherwise happen regardless since students have a positive perception about GenAI tools (Ben Otman et al., 2025; Elkhodr and Gide, 2025).

Lastly, it is difficult to enforce disabling it reliably. That is because you cannot monitor students outside of college; they will still use it when usage is banned (Randall et al., 2023), and current AI detection tools are not 100% accurate (Wang et al., 2023a; Yan, 2025; Mak et al., 2026; Erhardt et al., 2025). A study conducted by Weber-Wulff et al. (2023) examined twelve publicly available and two commercial, widely used detection tools for academic checking, which were tested on both English and non-English texts. They found that the results were neither accurate nor reliable, with a bias towards human-written classifications. This bias is a severe problem due to the inherent black-box nature of the detection models (Randall et al., 2023). A significant issue they noted was concerning content obfuscation, which severely reduced the tool's performance and compromised its robustness. Especially the latter makes this a burden for enforcement since students can use AI tools themselves to make a text look more human-like. Also, students worsen their text to make it look human-like (Lont, 2026), which is a bad practice. Hence, course managers cannot enforce it as a technically robust system. Yet, advancements in detection tools should not be neglected since new tools provide promising features (Oliveira et al., 2024).

3.4 Sectional conclusion

This section tries to answer the first research sub-question, as shown in Section 2.2. The sub-question is mentioned as follows: *How have the historical development of AI in education and technological characteristics of GenAI shaped its educational relevance for student use in STEM education?*

To understand the concurrent use of GenAI in education, it is essential to examine its history. The initial section of this chapter outlines the development of AI in educational contexts through several key milestones. It began with learning programs that displayed a limited number of frames in a predetermined sequence. From there, the evolution progressed to branching systems that adapted content based on user input, eventually leading to ITS and AI teaching assistants in classrooms. This journey led to the emergence of GenAI tools. This progression highlights that while the goal of supporting human tutoring has remained consistent, the capabilities and scale of usage have advanced significantly.

Additionally, Section 3.2 contextualises the term GenAI within a specific subdomain, highlighting its connections to other popular technical terms such as ML. It also explains the mechanisms and processes involved in converting a prompt into AIGC. This showed how, through next-word prediction, hallucinations can occur, but also why prompt engineering is important and why AIGC is presented authoritatively.

In conclusion, the ongoing pursuit of technology that enhances human tutoring, alongside rapid modern multimodal capabilities and stochastic next-word prediction in all knowledge domains of GenAI tools, has made these tools indispensable in students' daily work. This was emphasised in Section 3.3, which discusses other benefits, such as time savings and personalisation, as well as the challenges of prohibiting these tools.

4 GenAI tensions

To identify contemporary ethical and sustainability tensions from student use of GenAI in STEM education, a thorough literature review has been conducted. By doing so, this chapter addresses sub-question 2: *Which ethical and sustainability tensions arise from student use of GenAI in STEM education?* Firstly, §4.1 addresses the ethical implications, covering topics such as academic integrity, inequality and reliability. §4.2 will go over sustainability implications and address computational demand and cognitive debt. To end, §4.3 will assess sub-question 2, as spelt out above. This comprehensive overview establishes a solid foundation for course managers to implement changes within their courses, as discussed in Chapter 5.

4.1 Ethical implications

As addressed earlier, the use of GenAI also has its negative implications. It is critical to identify these to derive striking guidelines. Therefore, this section discusses prevalent ethical tensions one must be aware of as a course manager in STEM education. It does so through three themes. First, academic integrity and misconduct are discussed in terms of over-reliance on AIGC and cheating. Then, problems related towards fairness are shown, such as bias and inequality. Lastly, data-connected tensions are identified concerning transparency, data provenance and academic credibility. To complement the literature, interviews were conducted with TU Delft course managers to gain a deeper understanding of the university's context, as presented in Appendix G: Transcript deepening interviews.

4.1.1 Academic integrity and misconduct

Using GenAI can greatly reduce one's effort in constructing an assignment or any other course-related work. However, a first concern frequently mentioned in the literature is pressure on academic integrity (Deng and Joshi, 2024; Mak et al., 2026; Prather et al., 2023; Rahman et al., 2025; Chan and Colloton, 2024). It has even taken a step further, with authors stating it can lead to a devaluation of degrees (Cotton et al., 2024; Starominski-Uehara, 2026). Furthermore, Prather et al. (2025) agree that the availability of GenAI diminishes the barrier to cheating. This is supported by a lecturer who states that while cheating has always existed, its prevalence has significantly increased (G.3). Therefore, this tension suggests the importance of educational approaches that foster student accountability and responsible use of GenAI. Moreover, this complements earlier findings in Section 3.3, which indicated that lecturers struggle to monitor students, G.1, and highlight the shortcomings of current AI detection tools. The latter is critical, since it is challenging for humans to differentiate between AIGC and human-written text (Zdravkova et al., 2023; Kasneci et al., 2023; Susnjak and McIntosh, 2024; Cotton et al., 2024; Rudolph et al., 2023). Something that undermines student accountability and assessment integrity. Although, as shown through the example below, it might be clearly visible, obtaining a valid proof seems infeasible. Hence, this tension highlights the need for assessment approaches that remain meaningful in contexts where GenAI use is difficult to detect. Concerning the example, a lecturer noticed that the level of English between assignments and exams was significant; assignment texts can be fluent and polished, while exam answers may show weaker structure, reasoning, and expression (G.3). Still, proving AIGC is infeasible since a lecturer cannot provide a valid proof, and if you have no proof, you have no evidence (G.1; G.3).

Another tension is related to academic misconduct or dishonesty (Zdravkova et al., 2023; Oliveira et al., 2024; Zou, 2025). A great concern due to the difficulties of detecting its use, combined with the fact that students can use GenAI tools themselves to make their work look more human-like. Its use is often related towards plagiarism (Deng and Joshi, 2024; Borji, 2023; Cotton et al., 2024; Tran et al., 2025; An et al., 2025). However, Prather et al. (2023) assess the use of AI tools related to plagiarism differently. They state that GenAI almost always creates content based on its training data, which is comparable to an author generating new information based on existing literature. In this regard, a community member has not authored the AIGC, meaning it would not violate plagiarism of the original work that was fed into the model. Nonetheless, they still argue that in rare cases where GenAI would output someone else's work exactly, it would be classified as plagiarism. Thus, they take a position arguing that the user of the GenAI tool should always be considered the author of the work. Indirectly, this means that a GenAI tool should not be treated as an author or co-author of the produced work. This debate suggests that clear guidance

about whether and how GenAI can be used for a course is critical, further showcased by a teacher example. Dishonesty is further evident when students do not explain their use of GenAI beyond just improving grammar, which undermines student integrity (G.1).

4.1.2 Bias, fairness, inequality

Another concern is related to AIGC being biased, seen in several studies (e.g. Chan and Colloton 2024; Kasneci et al. 2023; Zou 2025; Cheng et al. 2025). It is essential to relate this type of bias towards algorithmic fairness, as GenAI perpetuates existing biases in the training data (Deng and Joshi, 2024). Bias originating from real-world societal structures within the data itself is challenging to eliminate and may lead to other forms of bias when attempting to address it. For example, one could argue that there is societal bias in the dataset if it presents the male-to-female ratio of full professors at the Delft University of Technology. However, if one were to change this by either removing male entries or taking a random sample of both genders, this would lead to representation bias since it no longer adheres to the true distribution. Still, this is a tricky discussion due to the inherent black-box nature of the models (Randall et al., 2023), making it difficult to solidify these statements.

A related issue is inequality, where not all students have access to the same quality GenAI tools (Adamakis and Rachiotis, 2025; Erhardt et al., 2025; Gavira Durón and Jiménez-Preciado, 2025), where some might not be able to overcome the financial constraints (Randall et al., 2023). Particularly concerning are the benefits that come with premium models, as certain individuals have access to these, while others do not (G.1). This creates a gap between students, which might put certain groups of individuals at a disadvantage who already carry financial stress. The increasing reliance on technology can widen the digital gap, as students who are more AI literate are likely to adapt to new technologies more quickly. This situation emphasises the importance of creating a fair environment where students' performance is not dependent on their financial resources. The term AI literacy is critical when assessing guidelines later on, and is defined as: "a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace" (Long and Magerko, 2020, p. 2). Therefore, STEM education must introduce this concept, which will be discussed later in Chapter 5.

4.1.3 Transparency, data provenance, academic credibility

The last subsection addresses risks based on the inherent black-box nature of the models, leading to, for example, a lack of transparency (Borji, 2023; Adamakis and Rachiotis, 2025; Zhai, 2022; Zdravkova et al., 2023). Such a lack of transparency prevents both students and lecturers from profoundly understanding how a certain output is generated. Combining this with both algorithmic and systematic bias seen in hallucinated output poses risks when students show over-reliance on this falsified information. On the one hand, they may learn incorrect information if not assessed critically, impacting their learning experience, raising epistemic concerns. On the other hand, it can also harm their academic credibility if such information is publicly articulated (Prather et al., 2023). This was the case this year with a chancellor of Ghent University incorrectly citing Albert Einstein, among others (The Brussels Times Newsroom, 2026). The latter solidifies why it is important to foster students' AI literacy in contemporary educational programmes. Also, it introduces risks regarding data privacy (Borji, 2023; Bommasani et al., 2021; Zhai, 2022). As Kasneci et al. (2023) mention, student data is often sensitive and personal. Although the operators of the models, such as OpenAI, must be transparent about data handling due to legislation like the General Data Protection Regulation, they still try to gather as much data as possible. They do this sometimes in ways that users are not aware of: a user must disable model improvement in their settings, as this feature is enabled by default when creating an account.

Lastly, the hallucinations and biases mentioned earlier impact the reliability of the model and, therefore, academic credibility. Jackson (2025) argues that LLMs are inherently limited as a reliable source of knowledge due to their stochastic nature. Something that was also concluded in the technical part of the literature review in Subsection 3.2. Hence, academic credibility in general is of great concern with GenAI tools, as earlier research on AI detection software has also noted these inconsistencies with the technology (Weber-Wulff et al., 2023). This was also noticed in student

output, where a lecturer observed that essays produced with GenAI were of subpar quality, often lacking clarity and depth (G.2). This emphasises the significance of AI literacy and Higher-Order Thinking Skills (HOTS) such as critical thinking, in STEM education, to overcome illusions of understanding (Messeri and Crockett, 2024).

4.2 Sustainability implications

In addition to the ethical implications, it is vital to address those related to sustainability. In the literature regarding the use of GenAI, the impacts regarding sustainability are often overlooked, despite their significant effects. Therefore, adopting this perspective is crucial for developing comprehensive guidelines later in the thesis. To establish this foundation, this subsection will address both ecological and cognitive sustainability concerns related to students using GenAI tools. Interview findings have been used as complementary material to connect findings from the relevance cycle with those from the rigor cycle, and can be found in (Appendix G: Transcript deepening interviews).

4.2.1 Ecological sustainability

When sustainability implications are assessed, most of the time this is regarding ecological sustainability, such as computational demand (Borji, 2023). In this light, Kasneci et al. (2023) argued the need for both energy-efficient hardware and infrastructure based on renewables, but also efficient algorithms to reduce the cost of training. However, an argument often posed is the trade-off between model size and accuracy (Mak et al., 2026). An article by (Ananthaswamy, 2023) questioned whether this is true; is in AI bigger always better? From the document, one can argue that larger models improve performance, but at some point, these benefits diminish. Restricted development due to sheer data set size, computing power, expenses, and electricity consumption are general concerns for this push for bigger models. Although the highest ecological costs occur during training and maintenance of data centres, a study by Luccioni et al. (2023) showed that BLOOM, with only 230,768 requests through its API, consumed 914 kWh of electricity over 18 days. Notably, this model featured 176 billion parameters, while models already exceed a trillion (Hudson et al., 2023). Yet, there is still room for improvement when comparing the models to the functioning of our own brain, which is larger and more complicated, but consumes way less energy (Ananthaswamy, 2023). And still, the rebound effect should not be neglected since the increase in efficiency can lead to an increase in consumption altogether (International Energy Agency, 2025).

4.2.2 Cognitive sustainability

An even less encountered point in the effects of extensive GenAI usage is regarding cognitive sustainability. Studies discuss that over-reliance on GenAI declines HOTS such as creativity, critical thinking, reasoning, and problem-solving (Zdravkova et al., 2023; Zou, 2025; Kasneci et al., 2023; Ala et al., 2025; Jackson, 2025). This was also noted by a respondent explaining that essay writing with GenAI involves less critical thinking compared to manual writing, due to the lack of proper desk research (G.2). This, in combination with less involvement, is detrimental, since they will not learn (G.1). Yet, neurological consequences are often disregarded. But, a study by Kosmyna et al. (2025) tested the effects of LLM-assisted writing on brain activity. The study consisted of 54 participants divided into a LLM only group, a search engine group and a brain-only group and made them write an essay in multiple sessions. In the fourth session, the LLM-group and the brain-only group switched. They used electroencephalography to record brain activity, combined with NLP and interviews after each session for evaluation. The results were staggering: the LLM-group showed under-engagement of alpha and beta networks, which are parts of the brain connected to creativity, working memory, and executive function (Wokke et al., 2019; Adamakis and Rachiotis, 2025). It also showed that the brain-only group had higher memory recall and prefrontal nodes with an enhanced feeling of ownership. The fact that the LLM-group performed worse shows what the authors call, cognitive debt: "a condition in which repeated reliance on external systems like LLMs replaces the effortful cognitive processes required for independent thinking" (Kosmyna et al., 2025, p. 151). Something highly related to the 'Google effect', which means that somebody relies on the Internet as a knowledge source rather than remembering it themselves (Gong and Yang, 2024). Therefore, it was rightly noted by a lecturer that students should not delegate their responsibility to AI, which often does not align with current use (G.3). Other studies in this field indicate that more

experienced users experience less cognitive load during web searches compared to those who are less experienced (Gwizdka, 2009). Additionally, Stadler et al. (2024) argue that students using LLMs have a lower cognitive load compared to those engaging in web searches that involve lower-quality reasoning and argumentation. Overall, this subsection highlights the importance of introducing educational approaches that safeguard HOTS and cognitive function in general.

4.3 Sectional conclusion

This section aimed to answer the second research sub-question, as shown in Section 2.2. The sub-question is mentioned as follows: *Which ethical and sustainability tensions arise from student use of GenAI in STEM education?*

This chapter presented both ethical and sustainability tensions students are exposed to when using GenAI tools. Beyond describing these tensions, the analysis identified a set of educational design implications, providing the foundation for the pedagogical themes in Chapter 5. Ethical implications included concerns of plagiarism, cheating, overuse, and circumventing detection tools, categorised as academic integrity and misconduct. This proposes changes needed to increase student accountability and responsible use, and assessments that remain intact if GenAI use is difficult to detect. Moreover, it discussed model-related ethical topics. This highlights issues of bias caused by algorithmic unfairness and inequality resulting from premium subscriptions. The latter proposes the need for an educational environment where financial resources are not the driver of the quality of student output. Lastly, it covered operational ethical concerns. An example is the black-box nature of diminishing transparency and understanding of how the output is generated, which is important for preventing reputational harm. Another implication concerned data provenance, where model operators do not always adhere to legislation regarding data collection. It concludes with assessing the reliability related to academic credibility of the technology due to its stochastic nature.

Finally, the sustainability implications of ecological and cognitive sustainability were discussed. This underscores the substantial electricity demand linked to model usage, a key aspect of environmental costs alongside training and data centre expenses. Additionally, a neurological study was discussed, assessing cognitive debt through research that examined the effects of LLM writing on brain activity. It showed these users had under-engagement within parts of the brain connected to creativity, working memory, and executive function. They also concluded that the brain-only group had higher memory recall and a feeling of ownership. Such insights suggest the importance of changing course pedagogy to safeguard the development of HOTS and cognitive function generally.

5 Identifying pedagogical themes

This chapter synthesises guidelines from prior AIED research into themes, supported by theoretical findings from the literature review and practical insights from semi-structured interviews with course managers. It combines the tensions discussed in the previous chapter with these themes to develop a framework for course managers' resources and create design requirements. This corresponds to sub-question 3: *Which pedagogical themes from prior AIED research can inform STEM course design in response to the ethical and sustainability tensions?* Therefore, §5.1 covers the themes and intentions found in contemporary literature to account for the implications of GenAI in education. §5.2 showcases the descriptive framework together with its goals and decisions. Then, §5.3 discusses and refines the design requirements for the artefact with their rationale and evaluation approaches for those supported by the findings. To conclude, §5.4 will discuss sub-question 3, as showcased above. Overall, this chapter outlines a design framework for the guidelines and course design, which is discussed in Chapter 6.

5.1 Guidelines from previous AIED research

After analysing and identifying multiple ethical and sustainability risks related to GenAI use, it is essential to find literature assessing contemporary AI-related guidelines. This will ensure that for the rest of the study, the best practices and ideas behind such policies can be adopted, and that the wheel will not be reinvented by giving authors proper credit. To do so, information and guidelines were extracted from the reviewed studies and iteratively synthesised into higher-level themes. Additionally, it provides insight into whether current literature emphasises learning activities, learning objectives, assessment, or more general rules. Additionally, it aligns with the literature necessitating clear guidelines on GenAI usage (Zhai, 2022; Cotton et al., 2024; Prather et al., 2023; Chan, 2023).

This chapter adopts a pedagogical framing of the identified tensions in STEM education, emphasising teaching, learning, and assessment practices at the course level. It is applied to findings from the literature review, evident in the categorisation based on CA below and in the descriptive framework, which outlines the demarcation of tensions used for further analysis in Figure 5.1. Nonetheless, such pedagogical practices are constrained by institutional conditions, including operational and governance dimensions, which are not modelled in the framework but discussed in the evaluation phase, among others, in the code On responsibility. By doing so, the scope of the analysis is bounded to the course level, while acknowledging the broader institutional context for integration of the guidelines.

5.1.1 Assessment

A regularly discussed point is whether contemporary assessment methods should be changed. All papers agree that online exams need proctoring and a disclosed environment to prevent unauthorised use of AI (Rudolph et al., 2023; Randall et al., 2023; Prather et al., 2025). However, Bollas and Bennett (2026) in their study on learner autonomy and engagement address the importance of refraining from creating an environment around AI that feels like prohibition and surveillance. They pose that this can be supported by being transparent about permissible and pedagogically meaningful AI use.

Assessment is also discussed in relation to which forms of assessment are still appropriate to reach learning goals. Notably, the most mentioned approach is to return to on-site assessment, such as oral assessments, coding interviews, and physical closed-book exams (Randall et al., 2023; Prather et al., 2023; Erhardt et al., 2025; Tran et al., 2025; Rudolph et al., 2023). Nonetheless, this depends on the type of course, since in a programming course, doing pen-and-paper does not align with the real-world (Randall et al., 2023). The same authors also argue that banning any type of AI from exams is a stopgap, and refer to mathematics exams that have also adapted to calculator use. Still, within a course, it is essential to ensure accountability, allowing students to defend and demonstrate their work and thinking processes (Karunaratne et al., 2025).

Lastly, assessments are discussed regarding their content and design. Multiple sources refer to the need for real-world context assessments (Tran et al., 2025; An et al., 2025; Randall et al.,

2023; Karunaratne et al., 2025). In their study, Wood and Moss (2024) state that real-world scenarios help through hands-on experience, which reinforces learning and displays the practical value GenAI can pose in instructional design. Chan and Colloton (2024) also state that students resonate more with assessments when they perceive them as authentic, for example, real-world tasks and projects (An et al., 2025).

5.1.2 Course learning objectives and activities

Outside of the assessments changing, learning objectives and activities need to be reconsidered as well (Prather et al., 2023). Clear learning objectives at the start of a course on GenAI integration help students and instructors understand its purpose, benefits, and challenges (Wood and Moss, 2024). This aligns with CA, which optimises conditions for quality of learning within education (Biggs, 2003). Biggs defines a good system as one in which all aspects of both teaching and assessment support higher-order learning processes. That aligns with recent literature addressing the relevance to focus more on HOTS within education (Prather et al., 2025; Randall et al., 2023; Elkhodr and Gide, 2025; Prather et al., 2023). Within the literature, this is sometimes referred to as soft skills (Karunaratne et al., 2025; Chan and Colloton, 2024). According to Wood and Moss (2024) and Holmes et al. (2023), combining this focus with the value of human instruction is essential for growth in HOTS and the protection of human agency. Referring to the components of AI literacy as shown by Adamakis and Rachiotis (2025), it is also important to make students aware of the risk of sharing sensitive, confidential or copyrighted material through public tools (Erhardt et al., 2025; An et al., 2025). Moreover, according to Tran et al. (2025), as a student, being able to evaluate is vital since students will otherwise not benefit from the technology. Something that corresponds towards reaching higher levels of Bloom's taxonomy, such as 'evaluation' (Bloom et al., 1956).

In addition, learning activities may change. For example, towards activities effectively using GenAI through teaching prompting techniques Prather et al. (2025). Important since literature suggests potential of assessments that balance AI-integration with independent reasoning (Ben Otman et al., 2025; Tran et al., 2025), and the development of GenAI-related skills (Holmes et al., 2023). Learning activities should also support prioritising the process rather than the final product (Ala et al., 2025; Chan and Colloton, 2024). In their work, Prather et al. (2025) argue that in computer science education, the emphasis on coding syntax is decreasing. Instead, the focus should shift to higher-level concepts such as algorithmic thinking and problem-solving. This aligns with the learning objectives centred on HOTS. Similarly, Elkhodr and Gide (2025) suggest that students should be encouraged to concentrate on analytical skills rather than passively consuming information. They recommend incorporating reflective exercises related to AIGC to improve learning outcomes. Moreover, Wood and Moss (2024) emphasise the importance of discussing the ethical use of GenAI to foster critical thinking among students.

5.1.3 Document AI use

Another frequently encountered topic is to document any use of AI in the appendix of an assignment (Prather et al., 2023; Erhardt et al., 2025; Adamakis and Rachiotis, 2025). Ross and Baines (2024) go into more detail and necessitate that students append screenshots of all their AI generated outputs to their work. This, however, creates additional tedious tasks for lecturers, contributing to their already heavy workload. They still advocate for specific citation guidelines, and if not adhered to, this will be handled as plagiarism. Concerning the content of such an appendix, Adamakis and Rachiotis (2025) suggest that it encloses tools used, prompts, purposes, and the scope of integration.

5.2 Descriptive framework

Chapter 3 discussed the history of AIED, provided deeper insights into the technology, and explained why it should be adopted in STEM education. Following that, Chapter 4 presented an overview of various contemporary issues arising from the current use of GenAI in students' coursework, building upon the foundations discussed in Chapter 3. It did so from both an ethical and a sustainability perspective. This chapter focuses on pedagogy and its relation to the identified tensions, and therefore, what changes course managers can make to address these. Hence, to demarcate, the implications and guidelines are divided into three dimensions: ethical, sustainability,

and pedagogical. Each dimension illustrates the contemporary perspective on its topic as presented in the literature. Therefore, it serves as an overview for course managers of implications in STEM education and literature-based guidelines to tackle them, illustrated in Figure 5.1. By doing so, it not only aligns the tensions towards the resources of course managers, but also serves as a foundation for the design requirements discussed in Section 5.3. The size of the dimensions does not indicate their relative importance. Rather, it reflects the number of themes identified in the literature that aligned with the course-level focus of this chapter, discussed in greater detail below.

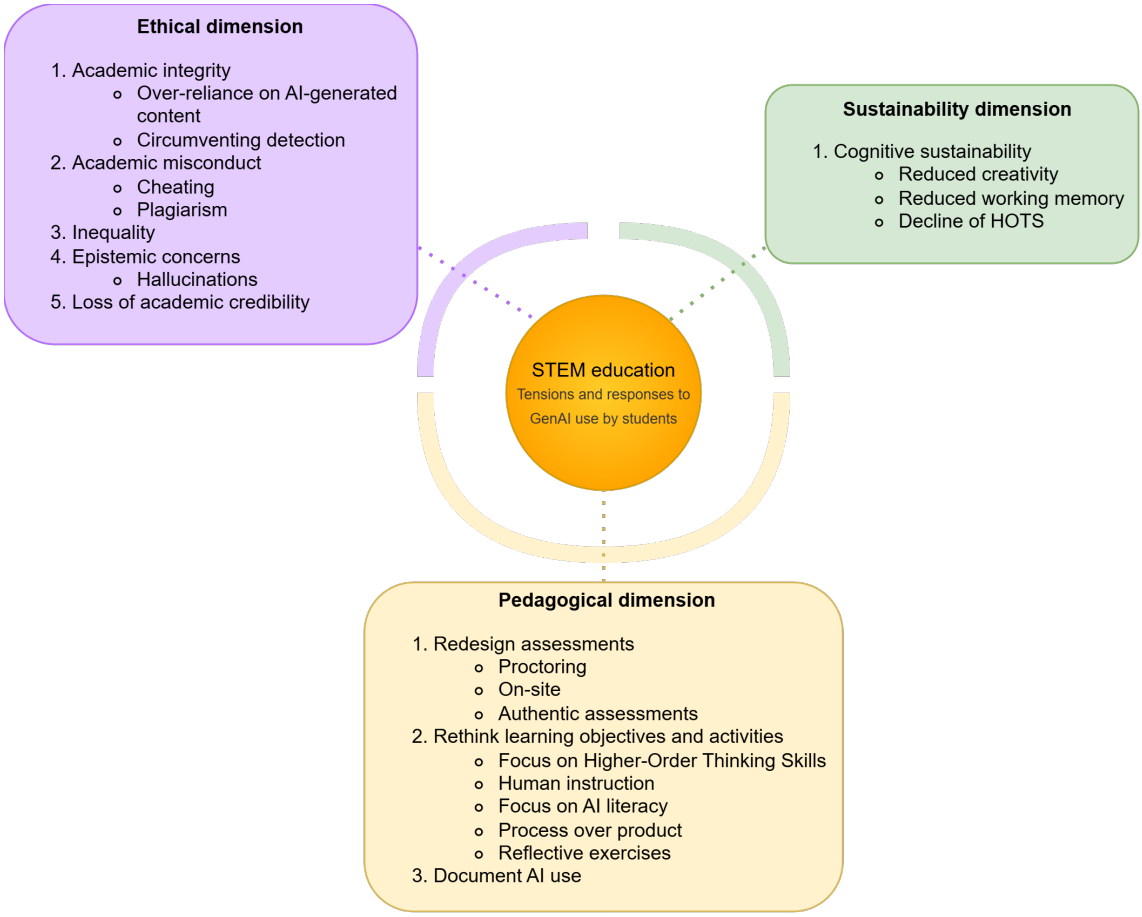

Figure 5.1: Descriptive framework showcasing the tensions in ethical and sustainability dimensions and responses in the pedagogical dimension within STEM education, inspired by (Chan, 2023)

For the ethical dimension, the framework introduces five different terms. These can be directly connected to Section 4.1, where they have been discussed in great detail. In short, this dimension highlights the awareness about the importance of knowing how to use GenAI tools pedagogically. However, it does not cover all the concepts related to the ethical implications. The focus is on pedagogy and course modifications, as the end-users are course managers. Therefore, the guidelines provided must be applicable for them to implement in their courses. For instance, algorithmic fairness is not taken into account, as it pertains to the model itself and cannot be managed within the educational context by course managers. The same applies to other tensions related to models and data, such as bias, transparency, and data provenance. These are controlled by large agencies, such as OpenAI, which are responsible for creating the LLMs. Although it must be noted that, currently, there are movements within the educational field to counter this partly (Npuls, nd).

Concerning the sustainability dimension, two concepts were identified in Section 4.2. This dimension emphasises another viewpoint towards implications that need to be understood by course managers. Ecological sustainability related to the use of GenAI models, training, and maintaining data centres, including cooling, is disregarded for further analysis because these factors are largely under the control of large tech corporations. Although legislation may restrict the construction of hyper-scale data centres in the Netherlands, this and other mitigations are not related to pedagogy or course

changes. Thus, cognitive sustainability is the critical concept, as it also relates more closely to the fundamental reasons why individuals navigate the educational system. When assessing it, one must realise that visible effects are long-term; therefore, it is crucial to address it already within education. For instance, consider an exaggerated scenario in which a student relies solely on GenAI to complete all their courses. In this case, the essential opportunity to learn and develop knowledge is compromised, harming the essence of education. To prevent this, course managers have a responsibility to adapt their courses so that this scenario becomes infeasible. Examples include a pen-and-paper or online proctored closed-book exam, which leads to the final dimension.

Ultimately, the pedagogical dimension underscores the need to transform teaching, learning and assessment practices at the course level. It highlights the themes identified throughout the literature in the previous subsection 5.1. To conceptualise them, they are connected to the tensions shown in the descriptive framework, Figure 5.1. This way of reasoning, by connecting changes towards tensions, is also employed when discussing the course redesign in subsection 6.5. The documentation of AI use potentially reduces academic misconduct by diminishing cheating and mitigating integrity risks associated with over-reliance. Although cheating relies on student honesty (I.2), an AI appendix provides a way to create accountability (I.1). Furthermore, mandatory documentation may raise students' awareness of their usage and increase the effort associated with relying on GenAI, potentially reducing over-reliance. Rethinking learning objectives by emphasising AI literacy and HOTS may help mitigate the loss of academic credibility and reduce over-reliance. An increase in skills, such as being able to critically evaluate AIGC, suggests minimisation of loss of academic credibility and reliance on this content, since students will be more critical about potential hallucinations. Furthermore, it may decrease the chances of plagiarism by informing students about the potential for direct copying of authors' work in AIGC and many modern GenAI tools that provide sources. A caveat to benefits connected to learning objectives is proper CA, as it guides students to achieve the learning objectives (Biggs, 2003). Redesigning assessments aims to maintain academic integrity and minimise overuse, consequently enhancing cognitive sustainability. The latter is further fostered through focusing on HOTS in courses, ensuring that students experience productive struggle (Vendrell and Johnston, 2026). Additionally, on-site assessments potentially reduce opportunities for circumvention and allow instructors to assess student performance. This may be complemented by learning activities, ensuring process over product, enhancing student accountability and progress. Still, after discussing these three themes, educational design implications concerning inequality and epistemic concerns related to hallucinations are not observed, leaving room for new guidelines addressing these tensions. One explanation may be that inequalities resulting from GenAI access or use are less visible to instructors than manifestations of academic misconduct, making them less likely to be addressed through course-level interventions.

To conclude, it is important to emphasise that the different dimensions of the descriptive framework do not operate in isolation. All of this occurs within STEM education, where ethical and sustainability tensions arise from students' use of GenAI, and the pedagogical dimension ties responsibility to course managers. A successful integration of students using GenAI in STEM education requires a commitment from both students and lecturers, as well as support from other institutional actors, including the Executive Board and the director of studies. Course managers are responsible for a course that adapts to the use of GenAI while maintaining CA, adhering to study objectives. Meanwhile, students are expected to use these tools responsibly and in accordance with educational guidelines or policy to achieve the learning objectives. By doing so, STEM education can help students become proficient in the use of GenAI. Nonetheless, it is essential to recognise how course managers can effectively implement innovations and enhance students with them, as past mediums like radio, computers, and the Internet have often received insufficient consideration (RudolphA, 2023). Therefore, to provide clearer guidance for course managers, Chapter 6 will offer specific course guidelines and a conceptual course redesign.

5.3 Design requirements

The work so far has focused extensively on the rigor cycle through a comprehensive state-of-the-art literature review and the relevance cycle through in-depth interviews. The demarcation towards a pedagogical framing in the descriptive framework, Figure 5.1, illustrates this framing applied to findings from the rigour and relevance cycle. It showed the importance of tensions within three

dimensions and explained how the pedagogical dimension can influence the others. Therefore, according to DS research, this forms the foundation of the design cycle, as shown in Figure 2.1.

Nonetheless, before an artefact can be created and evaluated, design requirements must first be established, as illustrated in the research flow diagram in Figure 2.2. These requirements are derived from the objective grounded in the problems and opportunities identified during in-depth interviews with course managers and from the scientific theories presented in the literature. The main objective of the artefact is to support and guide course managers in understanding the implications of GenAI for education and translating these insights into the design of their STEM courses. By achieving this goal, the artefact addresses the knowledge gap of limited guidance and understanding identified in the problem statement. To operationalise this goal, it is divided into functional requirements. These are translated into behavioural requirements, about how the artefact should behave, to conclude the refinement with structural requirements specifying the artefact design. Consequently, if all requirements are met, the artefact provides the necessary support for feasible and practical implementation towards achieving the main objective. The high-level requirements grounded in the literature and deepening interviews are presented in Table 5.1, together with their rationale for inclusion and evaluation approach. The evaluation approach specifies how fulfilment of each requirement and impact can be assessed, for example, whether it supports enhancing AI literacy. They were derived directly from the requirements. Since each requirement relates to a subgoal, the corresponding evaluation criteria were designed to assess the extent to which that requirement can be achieved. To operationalise these criteria, potential evaluation approaches were identified through brainstorming and subsequently refined using institutional resources, such as EvaSys for post-course surveys, resulting in the final evaluation approach. If a requirement concerned an aspect that could be assessed through course manager perceptions, an ex-ante evaluation approach was adopted, as these aspects could be evaluated within this thesis through the semi-structured interviews.

Notably, it is essential to recognise that guidelines have been selected as the primary artefact, as described by Offermann et al. (2010). According to their definition, requirements limit the system design space by imposing specific properties on the system. For example, if I state that a website must be coded in Rust, this requirement restricts the design options because it eliminates the possibility of using other programming languages such as Python or Java. Therefore, guidelines are preferred in this context, as the aim of the artefact is to provide course managers with practical and flexible guidance and suggestions for designing STEM courses rather than to impose restrictions. Moreover, the evaluation approaches, as listed in Table 5.1, are mostly ex-post, demonstrating how the effectiveness of the artefact can be measured after implementing the guidelines in a real-world STEM context. However, within the scope of the thesis, the artefact evaluation will focus on ex-ante evaluation by perceived metrics through semi-structured interviews on a course design, evaluated upon in Chapter 7.

Table 5.1: High-level design requirements with their rationale and evaluation approach

Requirement	Rationale	Evaluation Approach
The artefact should support raising awareness about pedagogically responsible use of GenAI	Multiple respondents expressed the difficulty of enforcing GenAI usage (G.1; G.3), and the soft recommendation style not working (G.1). Together with the literature implications of over-reliance (3.2; 3.3; 4.1.3; 4.2.2)	The number of claims of GenAI misuse within courses reported through a post-course survey or a university entity, such as the Board of Examiners, with a student survey through self-awareness of pedagogically responsible GenAI use
The artefact should support AI literacy reflection	An interview participant noticed the importance of students being able to understand the capabilities and limitations of current GenAI tools (G.3), which can be concluded from the infrastructure, 3.2, and aligns with preparing students for the workforce (Prather et al., 2025; Chan, 2023)	A post-course student survey in adapted courses with questions with a reflection on students' capabilities and comfort level with GenAI tools
The artefact should be feasible for course managers to integrate into current classes	A respondent addressed the importance of the artefact, showcasing what does and does not work well for integration in course design, while providing implementation space (G.2), important for feasibility for the end-user	Course managers report perceived feasibility and pedagogical appropriateness of the guidelines through semi-structured validation interviews
The artefact should be able to be applied to courses from different faculties	An interviewee expressed the need for guidelines to be high-level due to the multidisciplinary nature of the university (G.3)	A percentage of the number of courses per faculty able to implement the guidelines within two years
The artefact should allow for robust course changes	An interviewee discussed moving towards a conceptual model, which was initially outside the scope of GenAI, but quickly shifted (G.1), and the notions of current TU Delft recommendations regarding GenAI not being really applicable, up-to-date, and vague (G.1; G.2)	The percentage of guidelines that need to be changed within a year after implementation
The artefact should cover the ethical and sustainability tensions within reach for course managers (5.1)	The literature expressed ethical problems through: academic integrity misconduct, inequality, epistemic concerns and loss of academic credibility, and concerning sustainability: cognitive sustainability (5.1)	Course managers report through a yearly survey the effectiveness of every integrated guideline on the identified tensions per course
The artefact should be applicable and fair to both motivated and unmotivated students	In the literature, it was seen that students will continue using the tools despite prohibitions, which may exacerbate inequities between students (Randall et al., 2023; Feng et al., 2024; Cotton et al., 2024)	Course managers report the perceived applicability of the guidelines for students with varying levels of intrinsic motivation through semi-structured validation interviews
The artefact should align with the needs of students	Students have a positive perception about GenAI tools (Ben Otman et al., 2025; Elkhodr and Gide, 2025). Therefore, for example, prohibition would not align	Questions in the post-course survey for students to assess whether the guidelines restrict them in their normal coursework

5.3.1 Refinement of requirements

To obtain further insights into the design requirements, the high-level ones from Table 5.1 are categorised as functional and behavioural and italicised to differentiate interview and literature-based from those created during the refinement process. Translating the requirements into the design of the artefact resulted in refining the different abstraction levels, as illustrated in Figure 5.2. This process provides traceability between requirements and the artefact while retrieving potential conflicts.

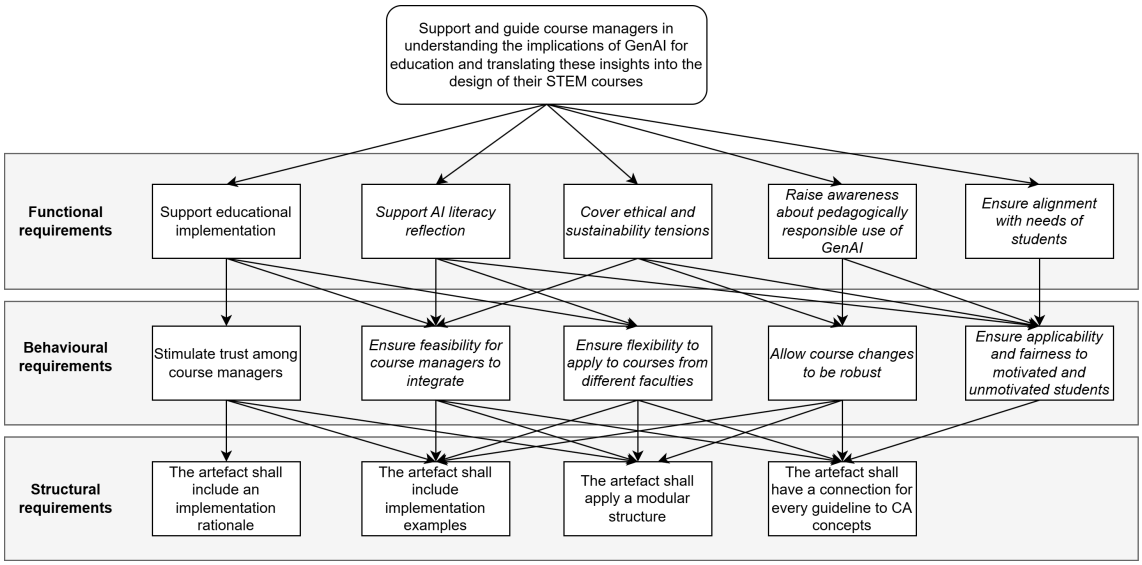

Figure 5.2: Refinement of the high-level design requirements, guided by the overall design objective

The table showcases the requirements identified during the refinement process. Although they collectively support the objective, the refinement process also revealed potential tensions between requirements. First, conflicts may arise from the functional requirement of aligning with the needs of students, as the other requirements are related to institutional objectives. As an example, the artefact should support raising awareness about pedagogically responsible use of GenAI, which may not align with students' current or preferred use. To address this, a requirement addressing applicability and fairness to both motivated and unmotivated students was incorporated, although its practical feasibility may be limited. Moreover, aligning the guidelines with CA concepts may constrain educational implementation, through limiting feasibility. This may require changes to the learning objectives, activities and assessments, which may not be feasible within institutional contexts, and therefore should be kept in mind when including implementation examples. Lastly, there is a trade-off between the feasibility for course managers to integrate the guidelines and their flexibility. If the guidelines are prescriptive, this may be overly specialised towards a certain type of course with the methods of this thesis. Thus, to balance this, both a modular structure and different implementation examples are adopted in the design. In doing so, the guidelines themselves are generic for flexibility purposes, while the implementation examples highlight practicality.

5.4 Sectional conclusion

Chapter five tried to answer the third research sub-question, as displayed in Section 2.2. The sub-question is as follows: *Which pedagogical themes from prior AIED research can inform STEM course design in response to the ethical and sustainability tensions?*

To learn from previous research on AIED and avoid redundancy, this chapter analysed previously proposed guidelines for students' use of GenAI. It did so with a focus on teaching, learning, and assessment practices at the course level. By doing so, three themes were identified, as shown in Figure 5.1 in the pedagogical dimension, and will be briefly discussed.

The first theme is concerning assessment. This involved proctoring online exams, transitioning back to in-person assessments, and integrating authentic assessments. Thus, tries to tackle concerns

about academic integrity by preventing over-reliance and cognitive debt. Regarding the latter, it is still important to reach higher levels of Bloom's taxonomy and perform those with responsible or no GenAI support.

However, assessment is only one part of CA. Therefore, it was valuable to see literature also discussing changing course learning objectives and activities, being the second theme. Within this theme, there was a focus on higher levels of Bloom's taxonomy through HOTS. Not only important to diminish cognitive debt, but also to benefit from the technology as much as possible. To benefit even more from the technology, it is important to focus on making students AI literate. This reduces loss of academic credibility and over-dependence, while enhancing the understanding of plagiarism potential.

The last established theme is the documentation of AI use. This requires a student to append for every assessment the tool(s) and prompts they used, with their purposes, and the scope of integration (Adamakis and Rachiotis, 2025). It aims to reduce academic misconduct, particularly focusing on cheating, plagiarism, and over-reliance on GenAI tools. Nonetheless, although these three identified pedagogical themes address many identified tensions, they undermine tensions related to inequality and epistemic issues related to hallucinations.

6 Educational Design

This chapter builds on the design requirements and guidelines identified in prior AIED research and translates these into a tangible course redesign. Therefore, it addresses the practical challenge faced by course managers of transforming guidelines into concrete course changes. Consequently, the underlying knowledge question concerns how the guidelines can be operationalised within a STEM course. The corresponding sub-question is: *How can the identified design guidelines be demonstrated through a conceptual STEM course design that integrates GenAI ethically and sustainably?* To start, §6.1 presents the pedagogical lessons from the literature review, after which §6.2 presents the resulting guidelines. §6.3 discusses the course selection criteria and defines the target course. Then, §6.4 covers the current content and structure of the course that will be redesigned. Based on this information, §6.5 presents the redesigned course by discussing improvements to the current course as well as elements that already align with the proposed guidelines. Lastly, §6.6 will address sub-question 4, defined above. Thereby, this chapter creates a design cycle that will be evaluated through semi-structured interviews in Chapter 7.

6.1 Pedagogical lessons assessment

Before exploring the guidelines and course redesign, it is essential to consider the pedagogical implications at the course level. The focus on the course level can be best explained through a point made by Deng and Joshi (2024). They state that the learning objectives are important when GenAI use would be considered as cheating. Imagine a course where a student has to write a document about a software they have engineered. In this case, using Grammarly would not be considered cheating. However, what if one of the learning objectives of the course is to practice syntactical rules like sentence structure or subject-verb agreement? Then, one may consider using Grammarly to be cheating. This example illustrates the distinction between the opportunity for concreteness of policy at a course or university-wide level. As an example, on a course level, students can be given proper and improper use cases for using GenAI (Deng and Joshi, 2024).

Moreover, prior research has focused on students' perceptions of assessment and has been explored further to support the relevance of the guidelines. Khan and Madden (2018) created a model of learning together that copes with fear of failure, poor testing history and lack of preparation. Since tests are sources of anxiety, they highlighted the benefits of partner learning in reducing anxiety about answering incorrectly. The study also showed that students preferred ungraded tests, which allowed them to focus on the learning concepts rather than thinking about what would be asked on the test. Moreover, in a case study at a Swedish University, Karunaratne et al. (2025) concluded that project work was scored as the most appropriate for enabling students' knowledge reflection. They also found that students vouch for combining assessments and having authentic assessments. Nonetheless, assessment does not always achieve its intended goal. In work by Halinen et al. (2014) and Ramsden (2003), lecturers noticed that students focus on memorising instead of understanding, to obtain a high mark.

Finally, the literature on lecturers' perspectives regarding assessment is examined to highlight the pedagogical insights that can be drawn from it. A study by Rawlusyk (2018) on assessment in higher education found that the most commonly used forms of assessment by lecturers are written papers, individual or group projects and presentations. He also noted that many educators continue to rely on tests as the primary form of assessment, with written examinations being the most popular. These assessments have advantages, as they require students to retrieve and apply knowledge on the spot, a skill that is essential in many professions (van Bergen and Lane, 2016). Chan and Colloton (2024) discuss the purpose of summative assessment, explaining that the resulting grades can motivate students to study, understand, and retain information. However, Halinen et al. (2014) point out that within summative assessments, educators often reuse the same or similar types of questions each year. This repetition is likely due to the grading workload faced by lecturers. Although this is especially the case for group assessments (Chan and Colloton, 2024). Nonetheless, students themselves perceive written exams as difficult to reflect knowledge acquisition (Karunaratne et al., 2025).

6.2 Overview of guidelines

This subsection translates the prior findings and design requirements into practical guidelines to realise the goal of supporting and guiding course managers in understanding the implications of GenAI for education and translating these insights into the design of their STEM courses. The guidelines are primarily derived from the functional requirements, as illustrated in Figure 5.2, as they reflect the findings and tensions from the previous chapters, which support the overall objective. For example, the requirement about supporting AI literacy reflection is transformed into a guideline about enhancing AI literacy. Those insights are combined with the pedagogical insights from the previous subsection and the previous chapter about approaches for assessment, learning objectives and activities to showcase the implementation rationale and examples. Moreover, current recommendations and support from the TU Delft are consulted (TU Delft Teaching Support, 2025, nd). Together with the domain knowledge of the researcher, this led to the creation of the guidelines presented below. For every guideline, an explanation and implementation examples are given in alignment with the structural requirements. These guidelines serve as the foundation for the course redesign, presented in Section 6.5. This redesign is the core for the validation interviews to validate and reflect on the guidelines and some behavioural requirements, which are discussed in Chapter 7.

To note, an overarching guideline and design requirement, not presented is CA, harmonising the learning objectives, activities, and assessment (Biggs, 2003). Biggs states that when aligned, it should be difficult for the learner to deviate from the process without learning appropriately. Moreover, he describes it as an approach to design, optimising the quality of learning. This provides a solid pedagogical foundation for the course redesign, ensuring that all components are aligned and that students meet the learning objectives. Vital, since when the condition of CA is not transformed towards potential GenAI use, it cannot hold.

Rethink foundational knowledge and skills
In an era where GenAI tools can outperform humans in certain activities, and are offered by scientific publishers (Elsevier, nd), education needs to return to its basics. Learners still need the foundational literacy, numeracy and basic scientific literacy skills to be able to assess the AIGC (Holmes et al., 2023). A great example is given by Poundstone (2017), arguing that a learner must always have a basic understanding of facts to avoid the Dunning-Kruger effect of not knowing ones own ignorance. As an example, he describes searching the capital of Colombia. Although this may seem common sense, what if you do not know that there is a country named Colombia, and that every country has a capital? Even with GenAI tools, which make it much simpler to find information, taking the garbage in, garbage out principle, this same principle can be reflected towards its use. Important over the long term due to LLMs triggering less cognitive load within students compared to traditional web searching, leading to lower quality reasoning and argumentation (Stadler et al., 2024).

Adopt authentic assessments
Including authentic assessment in a course where possible is vital to prevent students from going through the motions (Karunaratne et al., 2025). As outlined in Section 5.1.1, these assessments support by resembling professional or real-world contexts, helping students to resonate more with them (Wood and Moss, 2024; Chan and Colloton, 2024). To further boost student engagement and knowledge reflection, course managers should consider incorporating partner learning, formative assessments, and project work (Khan and Madden, 2018; Karunaratne et al., 2025). By doing so, course managers can create a learning environment where real-world problems and human values are in harmony. Something GenAI is less appropriate for (Holmes et al., 2023). Still, as with any change, course managers need to be cautious with incorporating group work, as this can be prone to free riding.

All of the above try to create something crucial, which is intrinsic motivation within students. This not only makes them more comfortable and confident but also more motivated to achieve the learning goals of a course. An example is when a course has a group project, allowing students to choose from a set list of topics or create their own. Additionally, such authenticity within a student's learning experience can enhance both engagement and relevance (Wood and Moss, 2024). Lastly, if the course load can be increased, the authentic assessments can be combined with others to have diversity within a course (Karunaratne et al., 2025).

Integrate on-site assessments
To test students on fundamental learning objectives and assess their HOTS, on-site assessments are key. As described in Section 5.1.1, this include amongst others, oral assessment, coding interviews and physical closed-book exams (Randall et al., 2023; Prather et al., 2023; Rudolph et al., 2023; Erhardt et al., 2025; Tran et al., 2025). These first three papers agree that if an exam is online, proper proctoring or a disclosed environment is needed. A caveat is that the type of assessment should still align with the learning objectives. As an example, Randall et al. (2023) states that a pen-and-paper exam for a programming course is misaligned since it does not represent a real-world scenario. Moreover, within on-site assessment, course managers must take into account that most cognitive tasks, for example, gathering information, have become less necessary due to GenAI (Lee et al., 2025). So, for oral assessment, asking about facts of a certain book is less appropriate than asking about an opinion of the impact of a certain event on the story. In this manner, such assessments go beyond memorisation. Nonetheless, this may increase the difficulty or workload of a course.

Ensure process over product
Within assessments, it is becoming more important to focus on the process, through intermediate output and development throughout an assignment, than on the final product. This may reduce the need for students to rely on GenAI tools due to more confidence in recurrent feedback (TU Delft Teaching Support, 2025). Something that can be enhanced by collaborative discussion and many feedback moments (Wood and Moss, 2024; Chan and Colloton, 2024). However, it can also allow course managers to monitor student performance better, allowing for more personalised guidance. This can be achieved through process-oriented assessment, such as drafts, presentations, or oral exams (Tran et al., 2025).

Enhance AI literacy
To support students in using GenAI pedagogically, they must become AI-literate. This has been greatly emphasised in a study by Adamakis and Rachiotis (2025). In the study, they found five core competencies: understanding AI, ethical awareness, data literacy, AI tool usage, and critical thinking. These components ensure students can use GenAI tools effectively and pedagogically responsibly. By doing so, they may see that GenAI should be used as an assistive rather than authoritatively, which scaffolds it as a learning partner (Vendrell and Johnston, 2026).

Concerning implementation, AI literacy can be implemented through ethics courses, where a brief overview can be provided in the syllabus. For example, there can be a small statement to never feed confidential or sensitive data in public GenAI tools (An et al., 2025). Also, the sustainability costs or effects on cognitive functions such as working memory or critical thinking can be briefly stated. This could be used as an argument for why certain parts of the course disallow GenAI use to overcome consequences, such as over-reliance. As an example, if using GenAI tools unstructured, it weakens your capacity to monitor and regulate your thinking (Vendrell and Johnston, 2026).

Moreover, since students lack frameworks on how to use GenAI (Ala et al., 2025; Lont, 2026), a large responsibility for a course is to guide them. Course managers can do so by dedicating a large part of the syllabus towards how to use GenAI tools optimally for the course. In this manner, they can both provide opportunities for showing prompting techniques, as well as connecting use towards the learning goals. These prompting techniques make students more efficient.

Clear guidance on how to document GenAI use
Since reproducibility is one of the key components of academic research, together with preventing academic misconduct, it is necessary to document your GenAI use. As discussed in Section 5.1.3, this can go as far as students appending screenshots of all their AIGC. However, this may pose pedantic work for lecturers, which is why records of only the prompts seem more reasonable. At Leiden university they have a similar mechanism for assessments where GenAI tools are permitted, where the records of prompts need to be archived by a student and given to the teacher on request (Leiden University, nd). As for other guidelines, they can be clearly stated in the syllabus or referenced in a university-wide document outlining the rules. Such a document may state, for example, that disclosed GenAI use must include the tool(s) used, prompts, purposes, and scope of integration (Adamakis and Rachiotis, 2025). It must also include what happens if the tools are not

used fairly. This ties back to the definition of fraud and the university's code of conduct, creating clarity for students.

Focus on developing HOTS
Referring to Bloom's taxonomy (Bloom et al., 1956), the higher levels become increasingly important for retaining cognitive sustainability and preventing overuse. As highlighted in 5.1.2, Tran et al. (2025) stated that evaluation is crucial to benefit from GenAI. A valid statement taking into consideration other points made about hallucinations and the capacity to easily find and articulate information neatly. It also aligns with the other literature discussed in the same subsection. An example was given by (Prather et al., 2025) stating that syntax is becoming less important compared to algorithmic thinking and problem-solving.

To achieve this focus, lots of learning activities have already been discussed in Section 5.1.2. Some include balancing AI-integration with independent reasoning (Ben Otman et al., 2025; Tran et al., 2025), while others emphasise discussions and reflective exercises (Wood and Moss, 2024; Elkhodr and Gide, 2025). Although there is no specific exercise that suits best, according to Vendrell and Johnston (2026), students must experience productive struggle, which underpins critical thinking. For example, this could take the form of a reflective group exercise that combines critical thinking with collaborative learning.

Protect human agency
Through education and hands-on courses, individuals develop a variety of skills. One of them is about intellectual skills, which can be compromised by using GenAI (Holmes et al., 2023). They state that GenAI should not overpower human thinking, which is important for developing HOTS (Wood and Moss, 2024). However, another essential skill that may be lacking is social skills, due to insufficient human communication (Holmes et al., 2023), which can lead to social isolation. It also links back towards earlier points, such as the normalisation of mediocrity and authority of AIGC. This emphasises preventing overloading with AIGC, and therefore enhances social interaction and exposure to creative output produced by humans (Holmes et al., 2023).

Create a fair learning environment concerning GenAI
Within a course, it is essential to have a fair learning environment where student output does not depend on financial resources. Regarding GenAI, ideally, this means that students all have access to the same advanced models. However, this is infeasible for the university to overcome due to the rapidly growing environment, and forcing students to use a single public tool is difficult. Still, education is trying to overcome this issue of inequality. An example is CS50.ai, a web application from Harvard University, which is similar to ChatGPT but with additional pedagogical guardrails (Liu et al., 2025). They state that this aligns more with their teaching philosophy of not providing a direct solution to a question, but guiding students. Although this was specifically for the CS50 course, it can also be created for the university as a whole.

Keep the technology close
Although the technological developments are not within reach for education to keep up, there is still space for adapting towards educational needs. Briefly discussed in the previous guideline was CS50.ai, used in the introductory programming course at Harvard. Such technology can decrease inequity concerning access, since it can be accessed through the university. Concurrently, initiatives with the same goal are going on with the surf AI hub (SURF, nd) and eduGenAI (Npuls, nd). This helps support an UNESCO measure of promoting universal connectivity and digital competencies (Holmes et al., 2023). Moreover, at the large operator level of OpenAI, initiatives have arisen, for example, a study mode for ChatGPT (OpenAI, 2025).

Still, for education, it is necessary to take something into one's own hands to counter the tech industry hype and harm. For example, creating their own guardrails or using Retrieval-Augmented Generation (RAG) with a vectorised database containing course materials (Liu et al., 2024). This will decrease epistemic concerns due to a lower hallucination rate on course-related material. However, the rapidly growing landscape of tools should be taken into account, which complicates such solutions due to new features and improved performance in commercial models.

Caveat: AI-detection tools
For education, a challenging discussion arises regarding the use of AI-detection tools in assessments. This would be an appropriate tool for preventing student overuse of tools, just as it is used concurrently with plagiarism checks. When the percentage exceeds a specific threshold, it will be flagged for discussion with the student. However, as outlined in Section 3.3, these systems are not very reliable and tend to produce a significant number of false positives (Weber-Wulff et al., 2023). Although this method of detecting AIGC is not accurate enough, its development should be monitored due to its impact. Until then, this technology remains in a grey area within education.

6.3 Course selection criteria

To test and evaluate the guidelines, they will be projected onto a current course given at the TU Delft. On one hand, this will test the guidelines by seeing whether they can be applied to a contemporary course. On the other hand, this will form the basis for the validation interviews, where course managers will assess these guidelines and the design requirements. Concerning the latter, the mapping from guidelines towards practical course changes gives interviewees the chance to assess its pedagogical appropriateness and feasibility. By doing so, it forms an evaluation point within the design cycle of DS research, as shown previously in 2.1. To find a suitable course, the course selection criteria are proposed in Table 6.1, with the last two being optional.

The course selection criteria were developed based on insights from the literature review and semi-structured interviews. For instance, these findings suggest that motivation is a key factor in determining whether students adhere more closely to course policies. Consequently, one of the criteria states that a course should be from a main Bachelor's programme rather than an honours or minor course, as these attract students who may be more motivated and academically skilled than the average student. Moreover, both sources indicate that it is challenging to track pedagogical GenAI use systematically. Therefore, courses with a limited number of students were preferred. During the process, a maximum of 150 students was adopted as a practical threshold to ensure the feasibility of incorporating supervised assessments, allowing for greater oversight of students' work.

Table 6.1: Course selection criteria and their rationale

Criteria	Rationale
A course with at least an exam and a project as assessment	Different types of assessments provide a clearer illustration of the guidelines. Enhancing a group project according to the guidelines is more tangible than creating a new project aligned with GenAI.
A Bachelor's course	Students face greater challenges in the early years of their studies with the responsible use of GenAI due to the transition from secondary school to university. This provides room for the guidelines to be applied more strictly and comprehensively. An example includes enhancing AI literacy.
A course from a main Bachelor's programme	A fair sample of students to test the guidelines' effectiveness on. That is because, for example, an honours or minor course includes more highly motivated students.
The course is not over 150 students	Adjustments made according to the guidelines facilitate smoother integration for ex-post evaluation purposes with a smaller group. Without them, there will be reduced control by lecturers and teaching assistants. This approach provides a gradual pedagogical introduction of GenAI tools.
A course where students can use GenAI, but also contains foundational knowledge or skills in its learning objectives	Introduces careful integration of the guidelines to foster the learning process while maintaining foundational knowledge and skills taught in the course, such as introductory programming or basic literacy skills.
Ideally, the course already has guidelines on GenAI use	By carefully considering such a document, the positive aspects and areas that require improvement can be more easily showcased compared to when a new document is created.
Ideally, the course manager is within close reach of the researcher	This, for example, allows me to discuss which parts of the course may or may not be included in the thesis for privacy concerns or other needs. Such rapid communication is crucial due to the time constraints of the thesis, as the course is not my own responsibility or work.

6.3.1 Chosen course

In alignment with the defined selection criteria in Table 6.1, the Data Science in Technology, Policy, and Management (TPM) course has been chosen. As further elaborated upon in the next subsection, it includes both an exam, a group project, practicals, and a reflection assignment. It is a first-year Bachelor's course given in the last quarter, Q4. However, it is not part of the main programme of the TPM Bachelor's, but a track course. Hence, it is essential to consider that this group may comprise more motivated students, and this bias may be inherent in the course design. Moreover, the course included 41 students this year. It contains foundational knowledge and skills in data science and ML using Python, as well as reflection skills on socio-technical problems. Moreover, it contains a document with rules on GenAI use in the course, see Appendix D: The use of chatGPT in assignments. Finally, the researcher can easily reach the course manager. However, this may introduce, among others, selection bias. Therefore, it is essential to recognise the technical nature of the course, concerning generalisability, and to understand that any interpretations or changes may be misinterpreted. Thus, the changes were approached as if they were applied to a random course.

6.4 Current course overview

This subsection provides a more detailed overview of the course being redesigned, which introduces data-driven analysis methods and algorithms to process data into useful, safe, and reliable information. It does so through its learning objectives, activities and assessment, aligning with CA.

Learning objectives
The learning objectives are taken directly from the syllabus and are as follows:
LO1: Formulating the problem and research question with multiple stakeholders

a) applying TB problem analysis to data science issues
b) justifying how data science and AI functionality can contribute to solving the problem
c) exploring a dataset to determine whether and how it can provide information for interesting questions or support decision-making
d) mapping out and substantiating the possibilities and limitations of data science / AI in advance

LO2: Operationalising a data analysis or AI system for research or decision-making:
a) selecting the appropriate modelling approaches and algorithms
b) determining the need and quality of data
c) processing datasets using modern data science scripts and statistical tools
d) building an AI model (rule-based or learning-based) for decision-making
e) summarising the key mathematical concepts for AI systems
f) applying relevant tests to evaluate a decision-making system against various criteria
g) identifying the limitations of the AI system's output and assessing its impact on accuracy and utility

LO3: Reflection – Communicating responsibly about the possibilities and limitations of a data-driven and algorithmic decision-making system, including:
a) using best practices for documenting analyses for reproducibility and accountability
b) reflecting on the ethical, legal, and societal implications of an AI system in terms of the research process and implementation
c) translating implications into possible considerations in problem formulation and system use

Learning activities
The course incorporates both in-class instruction and discussions as part of the learning activities. Discussions often involve small group break-out sessions where students can share their thoughts, which are then brought back to the larger class for further conversation. Additionally, each week, supplementary slides are uploaded to explain key concepts and, at times, feature tutorials in Jupyter Notebooks. These prepare students for practicals and the final exam. The same argument holds for reading the literature proposed each week. Additionally, it employs a variety of assessments, as discussed earlier. This encompasses practical exercises in Python, a group project, and reflective tasks.

Assessment
Concerning the assessment, the course is very diversified. It contains practical exercises in Python to assess coding data-analysis skills, accounting for 20% of the grade. Moreover, there is a group project, consisting of a project proposal, feedback sessions with teaching assistants, and a progress presentation, 30% of the final grade. The proposal and progress presentation account for 50% of this 30%, while the other 50% is the final report. Self-reflection exercises during the course are required for the exam and are considered formative. These prepare students to critically discuss the implications of AI use in societally challenging applications, and for the final individual reflection, which accounts for 20% of the grade. Finally, there is an online on-site exam where students may use a calculator and two A4 sheets, written on both sides with a font size of at least ten. To pass the course, a student must achieve a minimum of five for all summative assessments, with an average of at least 5.75.

6.5 Redesigned course

In the next subsections, the course redesign will be discussed. The goal is to create an example implementation of the guidelines, making the translation more tangible for course managers to consider during the validation interviews in Chapter 7. Initially, the key points that need to change are discussed to align with the guidelines. Then, some parts of the course are assessed that are already up to the guidelines standard. The latter is important for generalisation, as it provides more examples of how guidelines can be implemented in other courses. Regarding these changes, CA serves as the overarching framework, which means that all modifications in one category of its triad should be reflected in the others A quick overview of the guidelines and course redesign, with how every change relates to what guideline is shown in Appendix H: Validation interview pre-read.

Since the case course is a first-year BSc course, concerning GenAI use, the focus will be on developing independent reasoning capabilities to strengthen agency and self-regulation through integrating the guidelines. Among others, this will reduce the likelihood of experiencing the Dunning-Kruger effect (Poundstone, 2017). Within the STS of STEM education, important themes were identified, encompassing ethical, sustainability and pedagogical ones, as shown in Figure 5.1. Nevertheless, this also includes factors based on the institutional context, which are kept in mind during the change process but discussed in greater detail in Chapter 7.

To apply the guidelines, document analysis was conducted on the main course page of Brightspace and the study guide. This included reviewing the syllabus, a document on AI use, assignment descriptions, rubrics, practice exams, and lecture slides. During this process, each document was examined against the proposed guidelines, and notes were taken whenever a guideline was applicable. After this initial examination, all points were classified according to the categories of the CA triad, except for those related to the document on AI use, which are discussed separately. This allowed to identify dependencies between the different elements of the triad, ensuring changes in one category aligned with another. Lastly, for justification, where applicable, the proposed changes were linked back to findings from the literature. Together, this approach provides a method that course managers can use to apply the guidelines towards their courses.

6.5.1 Places for improvement

This subsection aims to apply the guidelines to the Data Science in TPM course. It therefore proposes improvements according to the guidelines through the components of CA. Moreover, it also addresses the document on AI use in the course separately.

Learning objectives
As discussed multiple times, evaluation and reflection on AIGC are crucial skills in benefitting from GenAI tools (Tran et al., 2025). Therefore, this general skill is added in learning objective three, d) reflecting on best practices and implications for adopting GenAI within the practicals and group project. Moreover, the same argument holds for the general skill of AI literacy that can be incorporated as a new learning objective: LO4 Enhancing AI literacy in line with (Adamakis and Rachiotis, 2025). With a) the ability to critically reflect on AIGC, b) the proficiency to work with contemporary GenAI tools, c) a basic understanding of the influence of good prompting and how to do so, and d) the ability to use GenAI pedagogically responsibly, adhering to academic integrity. Such a new learning objective not only fosters foundational skills within AI literacy, but also prepares students for the workforce (Prather et al., 2025; Chan, 2023). Hence, this shows how these skills can be incorporated into the learning objectives of a course.

Learning activities
To align with the new learning objective of developing GenAI skills throughout the course, new learning activities can be created. In the group project, students are responsible for selecting a topic. This can be converted into an in-class activity where students first think about a subject on their own and then brainstorm or refine their ideas using their favourite GenAI tool to find an appropriate dataset. By generating their own ideas initially, students still support the development of HOTS and AI literacy, as they need to evaluate the AIGC provided by the tools and compare it with their own knowledge and interests. This evaluation ensures students can benefit the most from using these tools (Tran et al., 2025). Moreover, this also supports students in recognising hallucinations, as the AIGC may contain non-existing datasets. Additionally, such an activity can be combined with prompt guidance from the course information, as shown in prompt (1) in the next subsection.

Looking towards the future, these GenAI tools may include ones provided by the TU Delft or companies such as Surf or Npuls. For the course, incorporating these tools would align closer with learning objectives through pedagogical guardrails. For example, instead of providing the answer to a coding question right away, it will first ask for further clarification. By doing so, the student will already critically think about what they do not understand, creating the rubber duck effect. In addition, it creates a fairer learning environment, since all university students have access to the same quality tool. As a university, we can learn from the implementation of CS50.ai, discussed earlier in the last two guidelines in Section 6.2. Such a pedagogical innovation

can alleviate privacy concerns by removing personally identifiable information from user prompts, incorporating pedagogical guardrails, and still function like a personal tutor to students (Liu et al., 2024). However, a case in Estonia demonstrates that student responsibility is crucial in such instances, as they would otherwise use the 'normal' version of ChatGPT, since it provides faster answers NOS (2026). Hence, AI literacy and the creation of intrinsic motivation are important to stimulate in the course to ensure pedagogically responsible use of new educational GenAI tools.

Assessment
Due to the change in learning objectives concerning reflection and AI literacy, this needs to be reflected in the assessment. Therefore, a development for the practicals is to add a coding question where students first have to provide their own solution, and then a GenAI solution. Such a reflective question could focus on discussing the differences and their impacts. Which is in accordance with literature stressing the importance of balancing AI-integration and independent reasoning (Ben Otman et al., 2025; Tran et al., 2025), and discussions and reflective exercises (Wood and Moss, 2024; Elkhodr and Gide, 2025). By doing so, it adheres to foundational skills defined in learning objective two, and fosters AI literacy. Since this is not an introductory programming course for Computer Science, the discussion does not need to address time or space complexity. Instead, it should emphasise the approach taken, since syntax is becoming less important due to GenAI (Prather et al., 2025). This approach encourages students to develop their problem-solving skills and understand their own and AIGC, progressing from the "understand" level to the "analyse" level in Bloom's revised taxonomy (Anderson and Krathwohl, 2001). To align with the GenAI guidelines of the course, an AI appendix should be included, consistent with the definition given later in the information on GenAI use in the course subsection.

Moreover, the group project contains a project proposal with a reflection. Currently, this calls to reflect on which parts of the problem can and cannot be answered through data science and AI techniques. Additionally, this may also include a reflection on where you would integrate GenAI within your project and why, while reflecting on the learning objectives. On the one hand, this increases AI literacy among students. On the other hand, this provides lecturers with feedback about the assessment's difficulty level. If students intend to use GenAI tools for a critical part of their assessment, as a course manager, you may have to revise CA to prevent this. For example, by adding a learning activity in the form of a mandatory on-site presentation. Although they could still use GenAI to construct this section, doing so may encourage them to understand the reasons for completing it themselves. Thereby, aligning with the evaluation of important learning goals and the protection of human agency.

Furthermore, the written exam guidelines can be adjusted slightly. Students are currently allowed to bring two double-sided A4 sheets with a minimum font size of ten to their exam. In an era dominated by GenAI, students may place all course information into these tools and request a 4-page summary covering all necessary topics for a written exam. By not writing a cheat sheet themselves, students miss developing essential HOTS, such as critical thinking, which is needed to determine what information they need for the exam. To overcome this, either one or two handwritten sheets may be taken to the exam.

Lastly, the coding practicals are summative assessments. While this may encourage students to complete their assessments, as grades motivate them to study (Chan and Colloton, 2024), it might also lead them to use GenAI tools to expedite the process and achieve a high mark. Moreover, contemporary search engines like Microsoft Bing provide AIGC when performing basic searches. As a lecturer, you cannot prevent students from looking at this content. For instance, if I were to search for how to add a string to a list and use the AIGC at the top of the page, I would need to reference this properly. Thus, to alleviate pressure on students, which is a primary motivation for GenAI use in the first place (Ala et al., 2025), this can be transformed into a formative assessment with a pass/fail. While this may decrease motivation to achieve high grades, it is worth questioning whether such motivation is truly effective. Research indicates that grades can divert students' focus from the actual learning objectives of a course. Lecturers have observed that students often prioritise memorisation over understanding in their quest for high marks (Halinen et al., 2014; Ramsden, 2003).

Information on GenAI use in the course
The course refers to a document about the use of AI tools, shown in Appendix D: The use of chatGPT in assignments. Although this document is a great starting point, and every course should have one, several improvements can be made. As an example, the benefits are fairly broad and not course-specific. Also, a point about brainstorming seems to be missing. In relation to the course, one example could be: brainstorming various ideas for utilising a dataset for the chosen topic. Another example might involve explaining the meanings of error messages encountered while coding during practical sessions or the group project (Prather et al., 2023). Furthermore, generating additional practice questions from existing exam material should be included as a benefit. By including these course-specific benefits, students resonate more with the benefits within the course.

Moreover, this document is strict about AI use, and does not recommend its use for assessments within the course. This would not align with enhancing AI literacy through AI tool usage competency (Adamakis and Rachiotis, 2025). Furthermore, this creates an environment of prohibition and surveillance, diminishing learner autonomy and engagement (Bollas and Bennett, 2026). Based on insights from the latter study, it is essential to be transparent about the meaningful use of AI in education. Consequently, the guide incorporates a section outlining the rules for using AI tools. However, an improvement ties back towards the previous point, that the benefits of GenAI should be connected with the course activities. This can be further improved by including example prompts that are optimised through the models' capabilities and operating principles for robustness, such as persona-based prompting. While the model architectures may evolve, principles of aligning prompts with how GenAI systems process context and instructions will remain applicable. Such an approach may also enhance students' prompting techniques and skills, ultimately boosting AI literacy. Combined, this may be stated as:

"As a first-year Bachelor student, I'm creating a report requiring explorative data analysis and a prediction model, either regression or classification. To do so, I need to find a dataset large enough for training ML models. Can you help brainstorm ideas for datasets about AI in warfare or related topics? Please provide a short description for each dataset. Verify the existence of the dataset by checking major database websites like Kaggle, Data.gov, or the UCI Machine Learning Repository."(1)

To guide students further, one can explain why this prompt is useful and relevant to the course. For example, this prompt utilises a strong feature of GenAI, brainstorming, and merges it with a carefully crafted prompt. The prompt employs a zero-shot prompting technique, incorporating the persona of a first-year Bachelor student, while also considering the context of the assignment, which is exploratory data analysis, along with the prediction model. It also includes constraints, such as the suitable dataset for training a ML model, whether it is a regression or classification problem, along with the verification requirements for the major databases. This example enables students to assess the GenAI output in relation to their own interests, thereby enhancing AI literacy and intrinsic motivation and engagement. It also assists them in choosing a suitable topic and database for their group project by providing guidance on how to use GenAI within the course, thereby addressing the gap of missing student guidance (Ala et al., 2025; Lont, 2026). Although one might think a similar example could be created regarding the use of GenAI tools to synthesise and apply course literature, this would undermine the foundational skills of scientific literacy (Holmes et al., 2023), which are crucial for first-year Bachelor students to develop.

Furthermore, the document states that the use of AI in assessment should be acknowledged and properly referenced. At the end of the assessment, a reflection section is required. This section should detail how AI was utilised in various parts, the prompts used, how the output was verified, and the measures taken to safeguard academic integrity. Yet, to align with the finding outlined in Section 5.1.3 and its guideline, this can be complemented by stating the tool(s) used, and changing 'how' AI was used towards 'why' AI was used. The tools serve for course managers as insight into what students use, while the why question triggers HOTS. Moreover, the proper referencing of the use is not elaborated upon. Thus, a solution could be to combine the acknowledgement, referencing, and reflection sections into a single AI appendix. To assist students, the document should provide a clear example along with a disclaimer stating that the chat history must be archived and made

available to the teacher upon request, inspired by Leiden University (Leiden University, nd). Such an example includes the outlined elements, thereby standardising the structure of the appendix, which saves time if it needs to be checked. Examples can be seen in Appendix B: AI-statement.

6.5.2 Already up to guidelines standard

In addition to the improvements, the course already follows various guidelines and principles. These are assessed below to serve as examples for other courses, as they show how certain guidelines can be integrated.

Learning objectives
What stands out in the learning objectives is the reflection part addressed in objective three and the identification in LO2 g. With GenAI tools becoming more capable of performing boilerplate tasks, and therefore being more used to do so, reflective skills are indispensable. Additionally, according to LO3 b, the reflection part is currently about ethical, jurisdictional and societal implications of AI systems, which is already improving students AI literacy.

Learning activities
The formative self-reflection exercises are based on real-world cases, adhering to the adoption of authentic assessment guidelines. They also need to be personal, creating a learning environment where GenAI is less appropriate for (Holmes et al., 2023). Moreover, they adopt questions for students to reflect on these cases and compare information from different sources, aligning with the focus on developing HOTS. Additionally, incorporating formative assessment into a course allows students to focus more on learning concepts rather than what will be asked on the test (Khan and Madden, 2018). The reflection exercises also cover topics such as algorithmic bias, which can be applied to GenAI, therefore also fostering AI literacy amongst students. Restricting GenAI in these self-reflections aims to preserve the foundational skill of learning objective three. Lastly, the in-class discussions adhere to the protection of human agency through human communication and authenticity since they are based on real-world cases.

Assessment
Concerning the assessment, the group project shows several components in compliance with the guidelines. A Swedish university study concluded that project work positively impacts students' ability to reflect on their knowledge (Karunaratne et al., 2025). The group project allows students to choose from a variety of topics, which fosters intrinsic motivation and helps them achieve the highest levels of Bloom's revised taxonomy. The selection of topics in real-world datasets also aligns with the adoption of authentic assessment. In addition, the project clearly emphasises process over product through process-oriented assessment (Tran et al., 2025). Both the project proposal and progress presentation account for half of the grade, while the peer reviews and feedback sessions with teaching assistants can foster students' confidence, which may reduce reliance on GenAI (TU Delft Teaching Support, 2025; Wood and Moss, 2024; Chan and Colloton, 2024). Moreover, the oral presentation examination aligns with the emphasis on on-site assessment, and human agency through improvement of social skills via human communication (Holmes et al., 2023).

Furthermore, the practical Python exercises do not only include syntax-loaded questions, but also reflective questions where students need to explain their design choices. This aligns with the development of HOTS through higher levels of Bloom's revised taxonomy, such as 'analyse'. Within these exercises, it is clearly mentioned why GenAI may not be used, and what happens if its use is suspected. This can safeguard the development of foundational coding skills, while also giving students arguments to raise awareness. The same principle is applied in the final reflection assessment, which ensures the development of reflective skills in alignment with learning objective three. Hence, HOTS development is preserved through a clear kick-off criterion. Lastly, the inclusion of a written examination has the benefit of testing students' ability to retrieve and apply knowledge on the spot, which is important for many professions (van Bergen and Lane, 2016).

6.6 Sectional conclusion

Chapter six tried to find an answer to the fourth research sub-question, as displayed in Section 2.2. The sub-question is as follows: *How can the identified design guidelines be demonstrated through a*

conceptual STEM course design that integrates GenAI ethically and sustainably?

This chapter proposed changes through document analysis against the guidelines, towards a contemporary university STEM course. While doing so, the overarching principle was CA, which is the base of a course to ensure the learning objectives, activities and assessments are aligned. The guidelines have been formed based on the technical foundation in Chapter 3, tensions in Chapter 4, prior AIED guidelines in Section 5.1, and finally the pedagogical lessons in Section 6.1. In the previous chapter, the demarcation aimed to provide guidelines in this chapter that course managers could easily apply.

By filtering courses based on a set of criteria, the Data Science in TPM course has been selected to redesign. Applying the guidelines to redesign this course demonstrated their applicability while also revealing several implementation considerations. The main opportunity identified was concerning enhancing AI literacy within the course. Nonetheless, it is essential to note that guideline implementation must be tailored to the educational context. Every course, and especially a first-year Bachelor's course should support rather than replace the development of foundational academic skills. Specifically for this course, GenAI could be suitable for specifying topics and finding appropriate datasets that satisfy course requirements. Yet, using it for synthesising and applying literature was considered less appropriate, as this may undermine the foundational skill of scientific literacy, and may hinder the development of learning objective three about reflection. Moreover, the redesign demonstrated that the effort required to implement the guidelines depends on the existing course structure. The presence of multiple assessment types and a course document addressing AI use allowed for smaller modifications, compared to courses lacking such elements. Still, this chapter made the guidelines more tangible by providing concrete examples through the course redesign, which were connected back to the foundation of CA.

7 Validation Interviews

This chapter progresses in the design cycle by being part of ex-ante evaluation through semi-structured validation interviews. By doing so, it evaluates both the guidelines and some behavioural design requirements. The corresponding sub-question is: *How do university course managers perceive the feasibility and pedagogical appropriateness of the proposed design guidelines, as reflected through the course design?* To answer this question, §7.1 contains information about the interview process. §7.2 showcases themes about the practicalities at play through the incorporation of the guidelines within a STEM course at the university level. Then, §7.3 highlights the validation through deductive coding of the guidelines and appropriate design requirements. Concluding the evaluation, §7.4 will cover the guidelines one more time by discussing them through the validation interview findings. Lastly, §7.5 addresses sub-question 5, as defined above. In doing so, this chapter concludes the final step within DS research possible within the time frame of the thesis.

7.1 Interview preparations

To validate the guidelines and behavioural design requirements, this subsection begins by explaining the process of gathering the interviewees and the preparations involved. In this second round, four interviews were conducted, recorded, and transcribed using Microsoft Teams. The transcripts were thoroughly reviewed line by line against the recordings, and relevant codes were assigned, aligning with open coding (Hecker and Kalpokas, nda). For practicalities, this process combines inductive coding, while deductive coding will be employed for the guidelines and design requirements. An overview of the resulting coding schemes is provided in Appendix F: Coding schemes.

Since the interviewees are anonymised, only limited information can be provided. The interviewees consist of a diverse group of individuals, primarily course managers from both Bachelor's and Master's programs, along with members of related communities. This group includes individuals from various faculties, encompassing technical, societal, and pedagogical courses. All participants were contacted through the researcher's network and their supervisors, with additional suggestions for other respondents provided by the interviewees themselves. Notes from the first round of interviews can be found in Appendix G: Transcript deepening interviews, while those from the second round are available in Appendix I: Transcript validation interviews. The questions themselves were formulated based on the interview round objective and split into themes. For the first round, this meant gaining insights into what lecturers need to cope with GenAI use by students and what is currently missing, as well as concurrent lecturer perceptions on GenAI use by students. The second round contains questions concerning the guidelines and behavioural design requirements to obtain insights into the course managers' perception. Concerning the latter, this is directly connected to the ex-ante validation outlined in Section 5.3. The questions are shown separately in Appendix E: Semi-structured questions, showcasing the connection towards identified tensions and both the guidelines and behavioural design requirements. Regarding the latter, this allowed for the ex-ante validation of the behavioural design requirements, as presented in Table 5.1. For the second round of interviews, a pre-read is constructed, as seen in Appendix H: Validation interview pre-read. This includes relevant information concerning the guidelines and conceptual course redesign, which was sent to respondents beforehand. If respondents did not have time to read it before the interview started, it began with a brief silent reading.

To note, although the first round of interviews was primarily for deepening purposes, some aspects were also relevant for discussing the incorporation of the guidelines. In doing so, both rounds of interviews are part of an ex-ante[‡] evaluation of the guidelines, behavioural design requirements, and tensions identified in the extensive literature review. Therefore, their goal is to connect theory towards practice and iterate the guidelines accordingly, but also discover other limitations within the university to cope with GenAI use by students.

[‡]Due to the guidelines not yet being implemented in a contemporary course, the validation was conducted through semi-structured interviews. Therefore, the ex-ante evaluation focuses on the perceived feasibility and pedagogical appropriateness of the guidelines rather than their effectiveness in practice

7.2 Practicalities of incorporating the guidelines

This subsection presents the identified practical restrictions and processes which hinder or complement the implementation of the guidelines. To code, themes were identified by reading the interview transcripts line-by-line and labelling relevant parts, aligning with open coding (Hecker and Kalpokas, nda). When a theme did not apply to any of the current codes but was still relevant to the practicalities, a new code was created. This aligns with inductive coding, where codes are derived directly from the data, which is closely related to grounded theory (Hecker and Kalpokas, nda; Glaser and Strauss, 1967).

On making course changes
For course managers to modify their courses at TU Delft, several implications need to be considered. Specifically, if a significant change is made to a course, a new course code must be obtained, and the updated course information must be uploaded before June 22nd (G.2). Hence, this respondent explained that when you make them a bit more vague, you leave yourself some room for implementation, confirmed by another respondent explaining that the implementation process can be a long one (G.1). This is further exacerbated by a respondent addressing a recurring issue that to make a change, you need to align with other colleagues. Which is hard, as reflected in the last quote, since people in general have diverse opinions on the effects and integration of AI

"(. . .) but then if you want anything, you need to align it with all your other colleagues, where, again, if many are not of the same mindset, you can actually not really make it aligned." - Appendix I.1

"So everybody has their own personal opinion about this, their own personal fears. Big tech or environment or just they think this is just a hype and we should just wait it over until the whole AI boom collapses." - Appendix I.4

Another discussed topic was changing the assessment. More specifically, the institutional and resource constraints that come with it. For instance, a course manager discussed the desire to include a midterm or an oral exam. However, they noted that this requires classrooms, surveillance during the exam, and teaching assistants to grade the results, explaining its infeasibility for large courses (G.3). While it would be ideal to return to these methods, they require additional effort, which should be taken into account alongside all guidelines concerning changes to assessment practices. Additionally, it is vital to consider the fair balance between the course load and the credits it represents, as this aspect was not addressed.

On changes towards their own courses
Multiple interviewees expressed some changes they have already made, or are planning to make, towards their courses to deal with or incorporate GenAI. All of these seem to align with the guidelines. One respondent emphasised the importance of students critically engaging with topics, which led to the modification of the essay format for an oral exam (G.2). This aligns with the focus on developing HOTS, protecting human agency and the integration of on-site assessment. Several changes were made to achieve the same goal. These included increasing the weight of the oral exam and introducing more in-class assignments (G.1). Additionally, the weight of the physical closed-book exam was increased, while the weight of at-home assignments was transformed to a pass/fail format (G.3). Yet, it was also seen that AI literacy was valued. In one of the courses, students were required to conduct customer interviews; however, most of the time, there was no opportunity to ask people on-site for practising purposes. To practice the interview, they were allowed to ask GenAI to behave as their intended customer, by providing a customer profile (G.2). This clearly shows that GenAI can be used as a scaffolded learning partner.

On responsibility
Concerning the responsibility for the guidelines, course managers seem to agree. The university has responsibility for providing high-level principles, with a clear support structure on how to integrate them into a course, together with providing the tools for lecturers to work with (G.3; G.1; G.2). While the lecturers have the responsibility to clearly state how to use GenAI for the course and the implementation of the guidelines (G.3; G.2; G.1). However, a problem noted in the quotes below by an interviewee showed that most university teachers are at the university for their individual

and research interests, and not to teach. The third quote further highlights why it is important, since students do not always show responsibility towards academic principles.

"I heard these complaints broader in the university, (...) there's a lot of influence of personal, individual, and research interests that may even go above these more foundational, broader-reaching questions." - Appendix I.1

"Yes, they teach but mostly people are here to research, and of course they choose to also teach, but many people didn't come here to primarily teach." - Appendix I.1.

"A hard thing is, I know when we did the homework assignments, (...) checking all the references. And again, the hit rate was surprisingly high on non-existing citations or claims or this claim of this is in this article and it wasn't actually in the article." - Appendix I.1.

However, a course manager may not feel responsible for the ethical dimension, as outlined in Figure 5.1. However, if the Executive Board of TU Delft establishes policy regarding GenAI, the directors of studies translate these into program-level learning objectives and educational policies. Ultimately, course managers are responsible for aligning their courses with these program objectives and policies. In this context, the Board of Examiners plays a crucial role in safeguarding the quality of assessments and ensuring procedures are fair and legally sound (G.3). Therefore, even if a course manager does not feel personally responsible, the university as an institution does have that responsibility. Consequently, through this bureaucratic top-down governance approach, course managers are still expected to act in accordance with this AI policy.

On institutional issues
Something encountered in multiple interviews is the notion of bureaucratisation within the university. A clear example was given concerning a fraudulent case (G.3). When a teacher suspects concerns, it reports to the Board of Examiners and explain the evidence of misconduct or why the submitted work does not meet assessment requirements. The lecturer must provide proof, because in doubt, the board says no, as also seen in the quote below.

"Where I've also heard from multiple boards of examiners also in other universities that the student tends to be very protected in this. So in these, unless you can make a fully hard claim that something truly has been there, there's no way that you will win that case. So many of these AI cases just are not admissible in that sense by the procedures we have now legalised, which means that maybe actually the board of examiners is not equipped to deal with this." - Appendix I.1

This inherent bureaucratised nature makes it difficult to change processes such as these to include a process for GenAI detection. This issue extends beyond the university itself; it includes how the government partly funds institutions based on the number of graduates and students enrolled within the nominal study duration (Rijksoverheid, nd). While this topic is outside the main focus of the thesis, the push for larger class sizes, aimed at ensuring more students graduate on time (I.1), highlights the need for a more structural change in STEM education within the Netherlands.

On generalisability
Since many universities and institutions have different faculties spanning different domains, the generalisability of the guidelines was also discussed by two respondents. One respondent suggested establishing a single central policy that would guide each faculty within the university in creating their own specific policies. To ensure flexibility among these faculty policies, this interviewee emphasised that any changes made by the faculty can be documented as an appendix to the central policy. In this manner, every faculty of an institution has the same base or rules, with some additional ones transformed towards its own domain. The last quote aligned with the behavioural requirements for the guidelines to be general to implement within the entire institution, but also feasible to be implemented at the course level.

"I think, it would be good to have one central policy, but still the faculties also need to develop one for themselves concerning their field and their domain." - Appendix I.4

"Okay, so this is our central part. It's very general. But now for your faculty, think about the impact and the implications for the programmes and write your own vision as like an appendix to our central policy." - Appendix I.4

"They have to be general, and they have to be open enough to cover the entire 8 faculties in TU Delft, and then allow flexibility for each course to be specific enough." - Appendix I.3

7.2.1 Subsectional conclusion

Regarding the practicalities, therefore, about feasibility, five different codes have been identified. Related to making course changes, the identified themes highlighted the long change process and difficulty of aligning the changes with colleagues due to the controversy around AI. Changes towards their own course, aligned with the guidelines, and included increasing the weight of oral or physical closed-book exams, together with coding interviews, and integration of GenAI within a course. Concerning responsibility, interviewees agreed that the university should provide high-level guidelines and tools, while course managers should translate them into their courses. Institutional issues arose concerning the bureaucratic nature of the fraud process and the difficulty in implementing it for AI cases, with awareness of the need for a more structural educational change for STEM in the Netherlands. Lastly, the generalisability code expressed the need for flexibility regarding the guidelines to be generally applicable in an educational institution at the course level.

7.3 Validation of the guidelines and behavioural design requirements

This subsection covers predefined code groups that directly apply towards the proposed guidelines and behavioural design requirements. Together with the previous subsection, this includes both the content and context of the guideline integration. Additionally, it will address the behavioural design requirements that can be evaluated through the semi-structured interviews, as presented in Table 5.1. The codes are directly connected to predefined guidelines and behavioural design requirements. This aligns with deductive coding, where data analysis is based on predefined codes (Hecker and Kalpokas, nda). Moreover, the coding has been done by going through the interview transcripts line-by-line and labelling them accordingly. In this manner, as much detail as possible could be covered, aligning with open coding (Hecker and Kalpokas, nda).

7.3.1 On guidelines

On rethinking foundational knowledge and skills
This guideline primarily focuses on course-level skills and knowledge, including foundational literacy, numeracy, and coding skills. As a result, one course manager suggested reevaluating the primary value of education and reconsidering the meaning of assessment within that context, as reflected in the authentic assessment guideline. A second respondent echoed this vision by encouraging a return to the basics of a course, such as the learning objectives and the methods used for assessment. As a course manager, you can consider how to integrate the use of GenAI in your course while ensuring alignment with these fundamental principles. Through that lens, they must revise their courses by rethinking learning objectives, activities, and assessments, as was demonstrated by the course redesign.

"Maybe not just foundational knowledge and skills, but also need foundational takes on what assessment means." - Appendix I.1

"I think going back to the learning objectives and rethinking those (. . .). If we're able to reconsider learning objectives, then from that I think we can think about, okay, how do we want to assess this? (. . .) but also like how can, how do we want to integrate AI in your learning objectives?" - Appendix I.4

Moreover, an interviewee mentioned that AI is shifting the necessary skills needed for the labour market, as seen in the first quote. This person also explained what these foundational skills are and how they have shifted over time in the second and third quotes. These encompass HOTS but also social skills. To ensure these skills are fostered within the university, another respondent acknowledged the need for supervised assessment. These encompass activities where students are

monitored, such as written exams or practicals performed in-class. In the last quote, this person expresses the benefit of such assessments in ensuring that every student improves the intended skills.

"AI has somehow shifted the skills you are necessary to be successful in the labour market. (...) We're using it as a steward, as an inspector." - Appendix I.3

"You need critical thinking, you need analytical thinking, you need creativity, you need other skills like collaboration, communication, resilience, productivity." - Appendix I.3

"The AI and large language models we had in 2022 or 2023, and they are vastly different, and they have grown in capabilities. But the way we use them (...) remains the same. (...). Even if you were to think of a GenAI five years from now or whatever it is, their skills of inspection or skills of stewardship remain." - Appendix I.3

"What are critical skills that we really want our students to be able to showcase independently? Let's go for a supervised assessment there. So we know that everybody's got it." - Appendix I.4

On authentic assessment
One interviewee expressed a positive vision towards the adoption of authentic assessments. However, a caveat is that many lecturers have not updated their assessments and knowledge yet, as showcased by the quotes below. So, many still think that old school reports pre-GenAI are still applicable to assess HOTS. Hence, this respondent also raised awareness of the importance of clearly emphasising what contemporary authentic assessments are.

"Authentic assessments, that's important. (...) it has always been important that students learn the skills that are readily acceptable in the labour market." - Appendix I.3

"Many lectures also don't recognise that the old school report that we used to do before 2022, before ChatGPT. (...). That report is still being used to assess critical, creative, and analytical thinking in a way that it was in the same way that pre-AI was being done, (...) that is not an authentic assessment any more." - Appendix I.3

On integrating on-site assessment
The integration of on-site assessments is recommended to ensure students meet specific learning objectives, as they do not have access to GenAI tools. Although respondents did not directly mention this guideline, throughout the example course changes from interviewees, this strategy was reflected. For instance, one respondent valued students being able to critically engage with the course topics, which led to a modification of the essay format for an oral exam (G.2). Another increased the weight of the oral exam and introduced more in-class assignments (G.1). Furthermore, another effective change was the inclusion of a formative coding interview. As stated in the quote below, this was enough for the course manager to know they understood what they did. While doing so, it was possible to assess whether they met the key learning objectives and developed HOTS.

"The only thing I changed was that I had a discussion with each group, (...), and I asked them to explain for me their code and the decisions that they made, and each of them should have talked. (...). If they know why they did something, that was more than enough." - Appendix I.2

Nonetheless, the integration of on-site assessment requires significant resources to assess, especially for large courses. A course manager expressed the intention to integrate a midterm or oral assessment, but concluded that this is infeasible due to classroom availability, surveillance during the exam, and teaching assistants for grading purposes (G.3). Another respondent addressed the same concern about integrating oral assessment, as seen in the quote below.

"And then also our idea to do orals was a problem because with twice as many people (...). I had four days and 1 1/2 times more students and one teacher down." - Appendix I.1

On process over product
A single participant briefly mentioned this guideline. The quote below shows a positive perception and its importance. According to this interviewee, this was the fact that co-creation is more accessible than ever, which aligns with prior findings of the large adoption scale and ease of use.

"AI is a tool, it is a calculator of evaluation, analysis, creativity. (...) the end result becomes (...) I don't want to say necessarily less important, but (...) co-creation becomes more accessible. (...). Putting a little bit more attention on the process, and not only the results." - Appendix I.3

On enhancing AI literacy
A single respondent addressed the guideline of enhancing AI literacy. This interviewee had, in general, a positive attitude towards the current way the TU Delft adopts this. Still, the need for a university policy on what AI literacy is for the university was expressed in the first quote. The second showed the perspective of the respondent on what AI literacy is, while the last one also expressed the need on the teacher's side to develop these skills before they can be taught effectively.

"I don't see that students are constrained in the use of AI and neither are educators, but we do need a university policy that guides us and allows us to know, that we are doing this equivalent TU Delft needs." - Appendix I.3

"Basically three things, (...). How did you figure out that it is correct, what the AI is telling you? Second, how did you verify that? And third, how did you act upon those verification processes?" - Appendix I.3

"Lecturers also need to develop their AI literacy skills. (...) being professional nowadays also involves updating yourself." - Appendix I.3

On documenting AI use
A controversial guideline was the clear guidance on how to document GenAI use. On the one hand, it was argued that such documentation is unrealistic, as it is too burdensome to document, and unmotivated students will construct it with GenAI tools regardless (G.1). Moreover, a respondent argued that the issue at hand revolves around student integrity and honesty: honest students will admit to using resources truthfully, while those who lack integrity will not be forthcoming.

"I just don't get it because it's about being honest. And if somebody is honest, they have used it honestly. If they're not honest, they won't report it honestly." - Appendix I.2

On the other hand, respondents were also positive about this guideline. One interviewee stated that there are currently no guidelines on it, with a template already being an improvement (G.2). Another respondent mentioned that it is really important concerning accountability and authenticity, as shown in the quote below.

"I think it's the only reasonable way you can ask for the accountability. (...) Because, let's say, not using this you'll keep with this feeling of whether what you see or not is authentic or not." - Appendix I.1

Still, one respondent provided a nuanced perspective, linking this to the learning objectives of the course in relation to CA. The following quote illustrates that the inclusion of an AI appendix may vary depending on the specific course.

"I think it goes back to the goal or the objective of your of your course. So again, if the learning objective also includes, let's say more the AI literature skills or working with AI skills, I think it will be beneficial to have some form of evidence saying, okay, how well did you prompt? - Appendix I.4

On developing HOTS
The respondent who specifically addressed this guideline recognised its importance. Nonetheless, this person also acknowledged that TU Delft is already making significant progress in this area. Current courses are designed to equip students with the essential toolkit of a graduate engineer.

Still, the respondent indicated that adjustments are needed in our teaching and assessment methods, as highlighted in the quote below. Examples can be drawn from other guidelines, such as adopting authentic assessments, prioritising process over product, and integrating on-site assessments.

"You could argue that although many of bachelor courses (. . .) it's part of building that toolkit of what the engineer is. The question is, how do you translate, you know, those higher order thinking skills? (. . .). And again, we can, but it requires a small tweak in the way we teach and we assess." - Appendix I.3

On protect human agency
The guideline on protecting human agency, whose name was taken from (Holmes et al., 2023), was considered rather broad and beyond the course level. Therefore, a course manager suggested with the first quote that within this guideline, the emphasis should be on why being in a learning situation is valuable. Furthermore, a course manager addressed the issue of social isolation in the second quote. This further highlights the need to protect human agency as student attendance declines (G.3; I.1; I.2).

"(...) with a guideline like protect human agency (...). So not all of the guidelines are necessarily operationalised in the context of teaching a particular course. (. . .) Strengthen the awareness of why being in a learning situation is valuable." - Appendix (I.1)

"It seems that you found your own friend, you found your own teacher, your everything. And then we are becoming more and more isolated, especially for students if they cannot really separate. (. . .). You need human to talk to, to hammer your ideas, to make you a better person, and they don't know this." - Appendix I.2

On creating a fair learning environment concerning GenAI
The guideline on the creation of a fair learning environment concerning GenAI focused specifically on increasing accessibility to advanced models. Although this is not the only parameter that accounts towards fairness in this regard, it is one of the few within reach for course managers to adopt in their courses. An interviewee highlighted the advantage of having a unified policy regarding a tool for the university, as demonstrated in the first quote below. Additionally, other fairness issues extend beyond the university level, affecting local and global contexts, surpassing the limits of the end-user. Yet, one course manager still raised this awareness in the second quote.

"Which is exactly the problem that we aim to prevent by a unified policy, which is to have everybody with the same AI tool, for example, your output as a student doesn't depend on how much can you pay a month." - Appendix I.3

"But still have some interventions that try to create a more equal playing field, which go beyond only offering the tool to everyone. (...) that's also one that I think could probably be richer in terms of how it ultimately will be worked out and what I now see." - Appendix I.1

Another aspect of inequality that extends beyond mere access to tools is AI literacy. A concern highlighted by a course manager indicates a significant disparity in literacy levels among students. This difference can be attributed to the fact that students who are enthusiastic about technology and find it intriguing are more willing to adopt it, in contrast to those who are more hesitant. This is a gap widened through little to no guidance from the university, as seen in the quote below.

"Not all the students are equipped the same way. I think that is a bigger problem as well. (...) but I think some students are not trained and we don't have any training anywhere in our curricula, like how to use this thing, what to ask, how to validate." - Appendix I.2

On pedagogical tooling
Pedagogical tooling, related towards the keep the technology close guideline, such as EduGenAI (Npuls, nd), was seen as a double-edged sword. In an ideal situation, every student would use such a tool when released. Thereby, all benefits such as keeping the technology close, creating a fair learning environment concerning GenAI, developing HOTS, and indirectly enhancing AI literacy would

be realised. However, as seen in the quotes below, as long as these tools do not meet the private company's standards and students are already accustomed to current tools, it will be difficult to change this. Outside of the fact that these can provide ways to cheat, as reflected in the fourth quote.

"So if there is a tool outside the prescribed tool that is giving you what you want better, you won't stick to the prescribed tools" - Appendix I.2

"While they will offer it to everyone, students will probably already be accustomed to using some of these other tools that they might have paid the premium fee for already." - Appendix I.1

"You will always find a way around when you know there's a tool that makes things easy for you" - Appendix I.2

"(...) like in software engineering for very long, we've been doing pair programming. (...) became the two students talk to their agents or their GenAI thing and then just mesh the things together and haven't actually spoken to each other." - Appendix I.1

Still, Npuls is working on addressing this issue, as highlighted in the following quote about EduGenAI. By storing interaction data, course managers can gain insights into student usage. Paired with an AI appendix, this could serve as a way to hold students accountable for their use of GenAI tools.

"They currently plan to have more sample environments where they can store such interactions. So, ideally, then you'd have the students kind of force working with the tool and then looking at the logs there in the end." - Appendix I.1

7.3.1.1 Subsectional conclusion

Concerning the validation of the guidelines, most of them were perceived as both effective and needed for the continuation of transforming education where students use GenAI tools. Nonetheless, some guidelines were only evaluated once. Still, they were perceived as important regarding the adoption of authentic assessment and the weighting of process over the product. Enhancing AI literacy and developing HOTS were also perceived as necessary, but in the case of the TU Delft, already partly integrated. For other guidelines, respondents showed shared interest and applicability. As an example, protecting human agency was valued because fewer people attend lectures and are, therefore, isolated, diminishing essential social skills. These skills are part of the guideline about rethinking foundational knowledge and skills, emphasising going back to what the skills are that we, as a university, want students to learn. Regarding the fair learning environment, multiple participants addressed the need for a standardised tool within the university.

However, some guidelines provoked mixed opinions among interviewees. Some participants expressed scepticism about the effectiveness of AI documentation, arguing that it ultimately relies on students' honesty. They noted that those who are not truthful will still find ways to utilise GenAI, and also even for creating this documentation in the first place. Others shared that this would be an improvement, even if it is only viewed through an accountability perspective or the ability to relate whether it is authentic. Furthermore, discrepancies concerning pedagogical tooling involved the effectiveness of the guideline. Multiple respondents agreed that students will continue using other, often 'better' tools, rooted in the argument that one will find a loophole if something makes it easier for them. Still, it was noted that current educational tools, such as eduGenAI, are showing promising developments in an environment where student use can be logged.

7.3.2 On behavioural design requirements

On perceived feasibility and pedagogical appropriateness
One dimension towards course changes from the guidelines is the student capacity of the course. Within a smaller course of around 50 students, you are less restricted by your context to make course changes, such as adding the previously discussed midterm. However, as one course manager stated in the quote below, there are main programme classes with over 400 students, with a large number not involved. This extends the discussion on conflicting design requirements of student

needs and institutional objectives by giving an example of what limitations within institutional contexts represent.

"So really, an issue is I think a course suffers from the restrictions of its more than 400 people. (...) half of them don't want to follow your course and will try to resist you. (...) with such restrictions and the budget (...), there is so much you can do really." - Appendix I.1

Another dimension concerns the containment of students. As discussed in the coding on pedagogical tooling, it is a challenge to ensure that students will actually use the prescribed tools. This complements earlier findings that showed the difficulty enforcing pedagogical use in Section 3.3 and design requirement conflicts around implementation. According to the quote below, it is therefore important to diminish the chances of finding a loophole.

"Where containing the students is going to be more and more difficult because they have all the tools they have for free and they can try it off. So whatever you put there, they will find a loophole to go around it." - Appendix I.2

A dimension regarding pedagogical appropriateness is missing AI literacy. As highlighted in Chapter 3, having a basic technological understanding is vital to use it properly. Not a problem only at the student level, but also at the course manager level, creating opportunities to change learning activities, as reflected in the course redesign. This is also reflected in the quotes below. Especially the last one that stresses the importance of improving hard skills.

"Still many people, students, even researchers, after a while using a GenAI tool, (...), forget that it's just predicting the next word. (...) they need to learn how to steer GenAI to actually do the deep research, give the proper context, what does context mean? So all these things we are missing." - Appendix I.2

"So, it also opens the door to be much more creative than before, because you can experiment a lot, you can brainstorm a lot with your tool. This is the part that we are not doing as much." - Appendix I.2

"You want to know how the algorithm goes, how much energy it's using, it's fantastic and everybody should know this." - Appendix I.3

Another topic was the level of implementation. Where the current guidelines focus on the university level, one course manager argued that for the policy to be striking, it needs to be not at a national, but European Union level, as seen in the quote below. For standardisation and monitoring of tools, this would be a way to create an EU-wide education GenAI tool. Scoping was also addressed in the second quote, showing that some guidelines are very large, entailing more than course design.

"(...) I think it should be something very, very higher level than university, not even national level. I think if there is a policy and if there's a tool being created that people need to use, it should be at (...) EU level." - Appendix I.2

"What I noticed is some of them are very large. As in, it entails a lot more than course design. For instance, enhancing AI literacy." - Appendix I.4

Lastly, as discussed earlier, respondents showed rather negative feelings towards current TU Delft support and recommendations towards GenAI in education, highlighting the pedagogical appropriateness of guidelines. Since a university-wide AI policy is missing, employees are left unsupported, as shown through the first quote below. The second quote further exacerbates the need and appropriateness for guidelines by necessitating a standardised GenAI tool. Another respondent expressed concerns about the lack of a vision for GenAI, which is detrimental when combined with the absence of university-wide guidelines, as discussed regarding course design.

"One of the reasons is that we don't clearly have a university-wide AI policy. So people are a little bit lost on what's happening or where are they going to align with." - Appendix I.3

"What type of large language models can we have at the university? Which ones are we using? How come? What type of access do we have? Like, currently, we don't have those." - Appendix I.3

"If I look at TU Delft, we don't have vision, we don't have central guidelines. And what I see now is that we don't have a common language or common ground that we can discuss course design or even the impact of AI in our courses. " - Appendix I.4

On perceived applicability of guidelines for varying levels of intrinsic motivation
For the effectiveness of the guidelines, one core component that arises throughout different interviews is student motivation. One course manager noted that motivated students who use AI responsibly are not a problem, while unmotivated students are, and face challenges due to competing pressures such as passing or graduating on time (G.3). Besides these students being less invested in course rules and regulations, according to this and other respondents (I.1; I.2), they also no longer attend the lectures. This is an underlying issue that harms the guideline of protecting human agency, necessitating the importance of in-class activities. However, this also raises resource and logistical issues, in addition to the complexity of the guidelines. Moreover, according to the quotes below, this student motivation is diverse within programs, due to its content and whether it is numerus clausus, and therefore not always reflected in student behaviour.

"And let's say again, what I do know, and probably it's also because of the skill of this particular major we have and the selection process before, we have quite a group of very pragmatic people that just want to get super good grades in the bachelor's to jump to a US institute or so for their masters, or that just want to get their diploma." - Appendix I.1

"We say that people are there to learn, but we don't always see it as intrinsically with the students." - Appendix I.1

As described above, intrinsic motivation is a large driver for students to engage pedagogically with GenAI and may lead to conflicting design requirements. Therefore, a design requirement was that the artefact should apply to different levels of intrinsic motivation. Nevertheless, the first quote from a course manager below suggests that the general feeling of the guidelines is focused on the motivated group of students. Hence, this quote again highlights the difficulty in applying guidelines to large, common courses with more unmotivated students. Still, another respondent viewed this behavioural design requirement differently concerning the responsibility. This individual argued, as indicated in the second quote below, that the university is solely responsible for creating an environment that enables students to develop and adopt foundational skills. However, it is ultimately the student's responsibility to take advantage of that environment. This raises the question of whether a university can effectively reach and motivate students who lack motivation. Nonetheless, this respondent argued that it is possible to do so through revising CA. When a course follows CA neatly, a student can almost only follow the course by also adopting the learning objectives, as seen in the other quotes below. This aligns closely with Biggs (2003), stating that with proper CA, it is difficult for the learner not to follow the process and learn. This last perspective was also shared by another interviewee, stating that when we go back to the basics, students may feel more intrinsically motivated to perform. A question about what education means and what we want to teach our students, combines a clear vision with CA within courses, as identified in the last quote.

"Some of the setups kind of assume good faith intents and motivation of the students. What I see in a big common course, because of course people also didn't choose it, and there is some resistance against the topic also." - Appendix I.1

"So the university is responsible for designing the environment that allow for students to foster those skills purposefully. That this student is responsible for taking advantage of those environments." - Appendix I.3

"Then that's the problem with what I'm talking about. It's the problem with constructive alignment. Then are we assessing the skill that we want to assess in the course?" - Appendix I.3

"It's about analytical thinking. And that analytical thinking should happen if you use AI or if not. (...) many lectures don't recognise this idea of constructive alignment and what they are assessing." - Appendix I.3

"I think vital to everything is that we, I think it's an opportunity for us to rethink what does education mean and what do we want to teach our students. (...) then I hope that all students feel more intrinsically motivated to do that." - Appendix I.4

7.3.2.1 Subsectional conclusion

Two design requirements were briefly evaluated. Regarding the perceived feasibility, outside of the positive attitudes, three constraints were mentioned: the size of the classes, the containment of students to adhere to the guidelines, and the implementation level necessary, which may surpass that of a university. Concerning pedagogical appropriateness, many interviewees agreed on its urgency due to a lack of AI literacy and the need for stricter and more appropriate AI policies and vision. Thereby, highlighting the appropriateness of the guidelines. Finally, on the perceived applicability for varying levels of intrinsic motivation, diverse opinions were shared. This included the guidelines currently assuming good faith and motivation of students. However, other respondents stated that students also have their own responsibility in adhering to the course rules, and that this is rooted in CA of a course. The latter aligns with CA being the overarching guideline, since it is the foundation of an effective course where learning objectives are reached.

7.4 Iterations on the guidelines

This subsection concludes the evaluation phase by discussing the final iterations and controversies around the guidelines from the semi-structured interviews. For each guideline, it is discussed whether changes are necessary and, where applicable, the final version is given. It also provides a brief discussion of the responsibilities important for actionability.

On rethinking foundational knowledge and skills
The interview findings support retaining the guideline as presented.

On authentic assessment
The findings from the interviews suggest that many lecturers have not updated their assessments and knowledge yet. Therefore, it is important to support lecturers by providing examples to expand on the current description of the guideline. This can include traditional written reports, which may be less suitable as authentic assessments, since students can generate substantial parts of the written deliverable with GenAI tools. So, for example, applying the process over product and integrating on-site assessment guidelines by adding a midterm presentation or oral exam based on this report, while making sure the topic is rooted in a real-world application. In doing so, a course manager can still assess the critical skills required for the labour market to ensure students have obtained the learning objectives.

On integrating on-site assessment
The interviews indicate that integrating on-site assessment is possible. Still, it is important to append a note that this modification should only be implemented to evaluate a specific learning objective, provided that the course resources can support this change. This approach ensures that lecturers and teaching assistants maintain a manageable workload, especially considering that course budgets can be limited.

On process over product
The interview findings support retaining the guideline description as presented. However, since this guideline is not course-specific yet, it has been changed to: Ensure process over product within assessments. A caveat is that this includes assessments where students can use GenAI.

On enhancing AI literacy
The interview findings suggest a shift in perspective regarding this guideline. Currently, it is centred around students, but should also include supporting lecturers. For TU Delft, this can be

achieved through courses offered by the teaching lab to obtain the University Teaching Qualification. Consequently, this support towards teachers and the responsibility for the teaching lab should be appended towards the current guideline description.

On documenting AI use
The interview findings indicate that the effectiveness of an AI appendix depends on student honesty, while it can also be seen as the only form of accountability concerning GenAI use. Nonetheless, as mentioned, current pedagogical GenAI tools such as eduGenAI are being developed, allowing for greater monitoring of students. Thus, the findings support that assessments where AI may be used should include an AI appendix in the deliverable, with the caveat that at some point responsibility lies within the hands of students. Besides this content clarification that should be appended to the current guideline description, the guideline is missing an action verb and therefore changed to: Provide rules on how to document GenAI use for assessment.

On developing HOTS
The interview findings support retaining the guideline as presented.

On protect human agency
The findings indicate that this guideline was abstract and should emphasise why being in a learning situation is valuable. Even though the content is unchanged, an explicit connection with CA would make the guideline more concrete for course managers. Concerning assessment, this can imply integrating an on-site assessment. Therefore, this guideline will be changed to: Employ a human-centred approach to CA.

On creating a fair learning environment concerning GenAI
Participants agreed about the need for a standardised tool within the university to ensure a student's output is not reliant on how much they can spend on it each month. Still, something that was not mentioned alongside the guideline was about AI literacy. The findings indicate that not all students are at the same level. In combination with the lack of literacy on how to use it, what to ask, and how to validate, this exacerbates the technological gap between students. Therefore, this guideline should also be presented in alignment with enhancing AI literacy to decrease this gap by equipping all students with a proper foundation. Examples include, concerning courses with assessment allowing its use, supporting students in pedagogically responsible use and how to assess AIGC critically. Consequently, the emphasis on increasing AI literacy among students to diminish the technological gap concerning GenAI should be added towards the guideline description.

On pedagogical tooling
Pedagogical tooling in the form of keep the technology close was perceived differently by interviewees. The problem with this guideline is that students may seek alternative tools if these provide more advanced functionality compared to the prescribed ones. Hence, as long as educational tools, such as eduGenAI, and AI detection tools, do not meet industry standards, all an institution can do is clearly specify the prescribed tool. This is especially important since the first perception of people about educational tooling will likely shape their opinion concerning its potential. Thus, the guideline is retained but with the caveat that it should explicitly acknowledge that the future steps depend on institutionally supported tools.

7.4.1 Responsibilities concerning implementation

Regarding the implementation and actionability of the guidelines, the responsibility is briefly addressed. One interview participant emphasised that the university should create an environment that allows students to develop essential skills (I.3). Additionally, as highlighted in other interviews, the university must support course managers by providing them with the necessary resources to implement course changes. Although budget constraints may pose challenges, whenever a midterm integration aligns with the guidelines for properly assessing foundational skills, it should be feasible to proceed, indicating responsibility at the course and institutional level.

The university Executive Board has responsibility for declaring the high-over guidelines as part of university policy. Then, the director of studies should translate this into programme-level learning objectives and policy. It is important to do this on the programme level due to the diverse studies

within an institution. Course managers should facilitate the course changes by aligning them with the programme objectives and policy. Therefore, the final guidelines should support a broader university-wide AI policy, where every faculty can append to the central policy if needed. Lastly, for the TU Delft specifically, the teaching lab has the responsibility for ensuring lecturers become AI literate when obtaining the University Teaching Qualification for understanding and teaching purposes.

7.5 Sectional conclusion

This chapter focused on a design cycle by evaluating the guidelines through semi-structured interviews with course managers and employees from relevant communities at the TU Delft. In doing so, another set of codes emerged regarding the practicalities of incorporating the guidelines. It also addressed the interview preparations and the pre-read for the second round of interviews. Thereby, it assesses the fifth research sub-question, as presented in Section 2.2. The corresponding sub-question is: *How do university course managers perceive the feasibility and pedagogical appropriateness of the proposed design guidelines, as reflected through the course design?*

The findings concerning the practicalities, and therefore, feasibility, highlight the trade-off between course modifications based on the guidelines and institutional possibilities. They suggest that modifications may require substantial additional resources, especially for larger courses. Moreover, they indicate that course changes in general are constrained by the change process within the institution and opinions from colleagues around the topic AI. Also, findings suggest that implementing a reliable procedure around claims of GenAI misuse faces barriers due to the difficulty of obtaining valid proof.

Furthermore, the guideline code groups indicate a positive perception towards the guidelines being appropriate. Nonetheless, the interviews suggest the need for institutional support to enable certain course modifications in alignment with the guidelines. For instance, a fair learning environment can be created when students have access to the same tools; this can be established by standardising a GenAI tool for coursework within the institution. This is a critical decision, as findings indicate that if other 'better' or user-friendly tools are available, students will find loopholes to go around this standardised tool. Hence, these insights suggest that the effectiveness of implementation within an educational institution depends on students' responsibility and cooperation, which may be driven by intrinsic motivation.

In addition, behavioural design requirements were evaluated. The perceived feasibility part of the first one aligned with findings from the practicalities. Therefore, both these findings suggest the challenge is not whether the guidelines are sound, but whether an institution has the resources required for effective implementation within its educational context. The pedagogical appropriateness of the first requirement also aligned with the predefined guideline codes, concerning the need for AI policy containing a clear vision. The final code on perceived applicability indicates that the effectiveness of the guidelines also relies on student motivation with proper CA within a course.

8 Discussion

This thesis aimed to create hands-on design guidelines for STEM courses to support course managers in addressing the GenAI usage of students for their coursework. In doing so, CA was the overarching framework that the guidelines must adhere to. The corresponding key findings, together with their interpretations and implications, will be discussed for both the literature review and the artefact and validation interviews in §8.1 and §8.2. Moreover, this chapter will discuss the limitations of the research in §8.3 and, to end, recommendations for future research in §8.4.

8.1 Literature review

The inclusion of GenAI in students' educational process highlights a structural change in the way we teach and assess. Results indicate that multiple implications happen from an ethical and sustainability viewpoint that inform the artefact. A clear disagreement in the literature was about plagiarism, often encountered as an issue with its use. Nevertheless, Prather et al. (2023) took a different view, stating that GenAI almost always creates content based on training data, comparable to an author writing based on existing literature. While it is uncommon for one person's work to exactly match another's, contemporary models offer sources that significantly reduce the risk of plagiarism. Still, because GenAI tools can sometimes create non-existent citations or hallucinate information, students need to take responsibility when using these tools to ensure proper academic citation. Moreover, some tensions are outside the reach of course managers to mitigate in their course design when considering a pedagogical framing. This is particularly true for those associated with the powerful large firms that develop the underlying LLMs and interfaces of contemporary GenAI tools. Ultimately, while education can influence the educational outcomes, many technological concerns remain dependent on decisions made by external developers. Thus, these were excluded from being addressed by the guidelines.

Moreover, the literature review findings suggest that sustainability-related tensions are often overlooked. This might be due to their long-term nature, which can make them less visible compared to other tensions, such as academic misconduct. Consequently, this poses risks of impacts being observable at a point where they are difficult to reverse. Yet, the papers that covered sustainability mostly involved ecological sustainability through electricity demand. Still, this does not consider another identified sustainability tension: cognitive sustainability. Its importance became evident through a well-established paper that showed excessive reliance on GenAI tools hinders the enhancement of crucial brain functionality, such as memory, creativity and executive function (Kosmyna et al., 2025). This tension highlights the importance of coordinated responses from education to maintain cognitive development. In doing so, education can stimulate development of HOTS and cognitive function, while preparing students for the labour market.

To identify what course changes course managers can make, themes from prior AIED have been analysed through a pedagogical framing. On the one hand, the findings emphasise the importance of CA by aligning assessments with learning objectives and activities to facilitate a proper learning experience. On the other hand, they are also grounded in multiple well-established literature sources, which implies consensus in the field. Still, after synthesising the literature, some identified tensions, such as epistemic concerns regarding hallucinations and inequality, remain unaddressed. This may be because issues such as inequality resulting from a lack of GenAI access are less visible to lecturers.

The last part of the literature expressed pedagogical lessons for assessment. Findings indicate that many students tend to prioritise memorisation over understanding in their pursuit of high grades (Halinen et al., 2014; Ramsden, 2003). Therefore, this suggests a tension between the goals of CA and student approaches to assessment. Especially since a study by Chan and Colloton (2024) describes that summative assignments are used to motivate students to understand and retain information. This may only be exacerbated by using GenAI tools without necessary engagement. Hence, GenAI tools serve as a trigger for course managers to reconsider how they can achieve CA within their courses. To guide them and move from state-of-the-art literature to practical recommendations, the next subsection will discuss the artefact and validation interviews.

8.2 Artefact and Validation interviews

The artefact in the form of guidelines has been used to transform the findings from the relevance and rigor cycle into a concrete design. Accordingly, the goal was to support and guide course managers in understanding the implications of GenAI for education and translating these insights into the design of their STEM courses. A first finding complementary to the literature is the practicality of implementing guidelines concerning the controversy around AI. The diverse opinions observed throughout the interviews reflect this broader controversy, which implies that course changes concerning GenAI may be constrained by not only technicalities but also diverging beliefs regarding the technology. This may constrain the implementation of guidelines such as enhancing AI literacy since course managers can advocate against its adoption.

Additionally, the literature suggests that incorporating oral evaluations, coding interviews, and closed-book exams can enhance the assessment of essential skills. However, to extend these findings, the interviews indicate that the successful implementation of these assessment changes relies on several factors, including class size, the availability of teaching assistants and institutional rules. Accordingly, the budget and resources of course managers may not be appropriate to facilitate this. Hence, the university or institution plays a significant role in supporting course managers to transform their courses. This institutional responsibility should be taken into account concerning guidelines such as adopting authentic assessment, integrating on-site assessment and ensuring process over product since these may require substantial institutional support.

Moreover, student motivation was a commonly discussed topic, specifically towards the effectiveness of pedagogical tooling. Although the need for a standardised GenAI tool was raised, it was also argued that students will always try to find alternatives as long as more advanced tools are available compared to the prescribed ones. These findings extend previous work regarding students' behaviour towards educational methods, as through assessment, where students obtain high grades by prioritising memorisation over understanding (Halinen et al., 2014; Ramsden, 2003). Both of these arguments indicate the need for course managers to realign CA for their courses. This is because proper CA, according to Biggs (2003), complicates the situation where the learner can follow the process without obtaining the learning goals.

Furthermore, the documentation of AI use sparked different opinions among respondents compared to the literature. As an example, Adamakis and Rachiotis (2025) argue that such an AI appendix would enhance transparency and preserve academic integrity and originality. On the one hand, in contrast to this study, participants imply that this integration is inherently related to student honesty, where unmotivated students will construct it with GenAI regardless. On the other hand, some also agreed that it is the only reasonable way to ask for accountability and may align with the learning objectives, particularly if those concerns are skills related to working with GenAI. These differences compared to the literature may be because they employed a theoretical state-of-the-art study, while this thesis also adopts a practical perspective through the relevance cycle, which consists of semi-structured interviews. This suggests that rules concerning such an appendix should be implemented carefully and concretely for students to be properly included in their assessments. And still, the disagreement may indicate that such a measure is not enough for mitigating academic misconduct and ensuring academic integrity due to diverging student motivation.

Lastly, an example from a course manager highlighted the impact of hallucinations on potential academic misconduct. A sample taken from a homework assignment revealed a high incidence of non-existent citations and claims that were not found in the sources (I.1). This reinforces the literature where authors presented various perspectives on academic misconduct through plagiarism, an issue inherently linked to the use of GenAI. Still, some authors approached the topic with a more nuanced view, comparing AIGC to the writing style of a real author based on existing literature. Differences in perspectives on this topic may arise because of the controversy surrounding it; individuals opposed to the technology might be motivated to view its use as plagiarism. This highlights the importance of AI literacy, enabling students to evaluate AIGC critically and responsibly.

8.3 Research limitations

Some limitations occurred during this thesis. The first limitation is the selection and background of interviewees. All interviewees were from the networks of the researcher and supervisors, which inherently introduced selection bias. Although they represented different faculties and genders, all respondents were employees from the TU Delft. Consequently, the first round of interviews in the relevance cycle and the second for validation are inherently shaped by the organisational context of the TU Delft. This may hinder the extent to which the findings can be transferred to other institutional contexts. Ideally, both interview studies would have been conducted across multiple countries and universities.

Secondly, the interview sample included seven participants, with three in the first round and four in the second round. For instance, because of this sample, not all guidelines were discussed, or they were only addressed by a single interviewee during the validation round. As a consequence, the validation of certain guidelines relied on a limited diversity of perspectives.

The third limitation is that the interviews, along with their corresponding codes, were conducted by a single researcher. Thus, the researcher's positionality as a student within the university may have influenced the analysis. To mitigate this issue, all interview transcripts and codes were sent back to the interviewees for their review. However, it would have been preferable for multiple researchers to be involved in this process.

The fourth limitation pertains to the choice of course for redesign. The Data Science in TPM course is being taught by one of my supervisors, which introduces potential selection bias. Although other selection criteria were considered, this ultimately led to the choice of a rather technical course, which may have limited the generalisability of the guidelines.

As a fifth limitation, the first research question specifically focused on the historical development of AIED, the technological characteristics of GenAI, and its relevance for adoption. Consequently, broader political-economic perspectives, such as the influence of commercial interests of AI companies on the adoption of GenAI in education, were not examined in depth. While aspects related to commercial AI companies were considered when discussing model-related ethical and ecological concerns associated with model training and data centres, a more comprehensive political-economic analysis falls beyond the scope of the study.

Finally, although the research was conducted from a student's perspective, it did not include enough student viewpoints. The primary focus was on the perspectives of course managers regarding students' use of generative AI for their coursework and its implications for course modifications. However, the research highlighted that students' intrinsic motivation is a significant factor in adhering to rules. Therefore, this omission may have led to missing factors influencing students' behaviour and compliance with course regulations. This is therefore also recommended for future research in the next section.

8.4 Recommendations for future research

The recommendations are divided into two categories: those for future research and those for practical steps moving forward. Future research recommendations have been identified based on gaps observed during the process, while the practical recommendations suggest actions to continue based on this research. Additionally, these practical recommendations are addressed toward specific stakeholders.

8.4.1 Future research

This research serves as a foundation for future studies. First and foremost, student motivation emerged as a crucial factor during the validation interviews. It was concluded that the majority of students no longer attend campus regularly. Nonetheless, the research focused mostly on the teacher perspective, as discussed in 8.3. Therefore, future research could involve this student perspective to validate why students do not attend lectures.

Secondly, a new interview study should be conducted on students' use of GenAI for their coursework in another country, involving multiple universities. This study would provide deeper insights into whether the proposed guidelines are effective for various teaching methods and tools across different countries. Therefore, this would evaluate and adapt the artefact in different educational settings to assess its generalisability, which may have been constrained by the DS approach. During the thesis research, it was noted that a Harvard class was already utilising pedagogical GenAI tools, while secondary schools in Estonia were testing their own version of ChatGPT. Similarly, in the Netherlands, tools like EduGenAI have been developed. These developments underscore the importance of comparative international research in examining whether different educational contexts require distinct strategies. Moreover, this creates an opportunity to discover whether an EU-level GenAI education tool is feasible and desirable amongst other things.

The third recommendation involves the under-representation in the literature of the effects of GenAI usage on cognitive functioning, discussed under the concept of cognitive sustainability. Although much literature discusses the effects on cognitive processes, such as critical thinking and problem-solving, a gap appears to exist concerning neurological studies as the empirical foundation for these claims.

As a fourth opportunity for future research, the existing perspectives from this study could be complemented by a more political-economy analysis of GenAI in education. This could include examining the commercial incentives and business strategies of AI companies that may shape the dissemination of GenAI tools among students. For example, this may involve investigating platform dependence, broader market dynamics and how these compare to educational initiatives involving GenAI.

8.4.2 Practical steps forward

Firstly, it is recommended that future research focus on developing and evaluating a framework or artefact that supports the integration of guidelines. This could take the form of a master's thesis for the TU Delft, where the final guidelines from this dissertation are transformed into a university-wide AI policy and vision. Additionally, the framework should include a comprehensive guide for effectively implementing these strategies into courses, while ensuring that the policy can be applied across all faculties. This implementation guide may be based on the responsibilities and actionability outlined in this thesis. It is also essential to provide support for teachers regarding GenAI. Therefore, conducting interviews will be important to ensure that the policy and vision are relevant and aligned with the needs of course managers. This approach will facilitate a clear integration of the guidelines into institutional AI policy, making the process streamlined for educational institutions.

Moreover, it is recommended that course managers apply the guidelines to a STEM course for a pilot to collect data for the ex-post validation of the remaining design requirements, as outlined in Table 5.1. This process could be part of the master's thesis mentioned in the previous recommendation, or it could be a separate project. By doing this, the evaluation of the artefact would be strengthened with valuable feedback, ultimately enhancing the guideline quality and consequent policy. To facilitate implementation, the guidelines could first be tested in a smaller course as a pilot project before being applied to larger, more common courses.

Finally, a recommendation is to revise CA for courses in the era of GenAI when applying the guidelines, especially concerning assessments. While standardisation through an AI policy would be beneficial, course managers can start familiarising themselves with the topic and consider making course adjustments in the meantime, which can be part of the pilot discussed above, based on the guidelines presented and in accordance with CA. This is important because, throughout the literature and interviews, it has been demonstrated that all course changes are rooted in the concept of CA. In the literature, Deng and Joshi (2024) explain that learning objectives play a crucial role in determining whether the use of certain GenAI tools is considered cheating. This perspective on CA was shared by interviewees, who emphasised key questions about the primary value of assessment, the evolving nature of effective assessments, and the importance of ensuring that students achieve the learning objectives.

9 Conclusion

This chapter concludes the thesis by first giving answers to the research sub-questions in §9.1. Then, §9.2 showcases the conclusions to the main research question. §9.3 addresses the relevance of the research from both an academic and societal perspective. Finally, §9.4 will express the final reflections based on the connection with the CoSEM programme.

9.1 Answers sub-questions

1. *How have the historical development of AI in education and technological characteristics of GenAI shaped its educational relevance for student use in STEM education?*

The literature findings indicate that AIED historically pursued the goal of human tutoring, combined with technological advancements, increasing adaptability and accessibility. Therefore, the creation of GenAI tools can be seen as an advancement of this trajectory, where it differs from its predecessors in its wide availability. Moreover, their technological characteristics can explain their educational benefits and limitations. Multimodal capabilities and the large knowledge base enable personalised support for STEM students, while stochastic next-word prediction may introduce hallucinations. Hence, the evolution of AIED, together with the technical capabilities and broad knowledge base, shaped its educational relevance by making personalised learning support more accessible. This explains why these tools are widely adopted in students' daily coursework.

2. *Which ethical and sustainable tensions arise from student use of GenAI in STEM education?*

The findings indicate that ethical tensions do not only cover academic integrity and misconduct, but also model-related and operational issues. Therefore, they suggest that educational institutions that adopt GenAI use are dependent on design choices made by external developers. This leads to implications that are not within the control of educational institutions, such as the inherent black-box nature, which creates transparency issues. Nonetheless, findings also suggest design implications concerning increasing student accountability and responsible use in combination with assessments that remain intact if GenAI detection remains infeasible. Furthermore, investigating the sustainability aspect did not only indicate tensions concerning ecological, but also cognitive sustainability. Especially the latter imposed the need for educational change, as overusing LLMs is associated with a decline of essential cognitive abilities, including memory, creativity, and executive functioning.

3. *Which pedagogical themes from prior AIED research can inform STEM course design in response to the ethical and sustainability tensions?*

The identified themes indicate a strong connection with CA, therefore used as a foundation to address the relation of the themes towards the identified tensions. Findings from the assessment theme highlight a push to return to supervised in-person assessment as a response to academic integrity and misconduct. This seems to be rooted in an observation that students' GenAI use cannot be controlled. To address an earlier tension of cognitive sustainability, the literature responds by emphasising HOTS and making students AI literacy. The latter suggests the importance of a foundational skill; critically assessing AIGC. Nevertheless, combined with the documentation of AI, using the identified pedagogical guidance gives limited attention to those related to inequality and epistemic concerns through hallucinations.

4. *How can the identified design guidelines be demonstrated through a conceptual STEM course design that integrates GenAI ethically and sustainably?*

The overarching principle of CA in combination with document analysis allowed for a standard to both evaluate and integrate the guidelines. For finding an experiment course, multiple course criteria were created, leading to Data Science in TPM as output. The application of guidelines to redesign the course demonstrated a trade-off between their applicability to the educational context and the existing course structure. Although enhancing AI literacy was seen as the main opportunity, the implementation was affected by the educational context; within the first-year

Bachelor's course, development of foundational knowledge and skills, such as scientific literacy, should not be compromised by GenAI use. Additionally, the existing course structure influenced the amount of redesign required due to the presence of multiple assessment types and a course document addressing AI use.

5. *How do university course managers perceive the feasibility and pedagogical appropriateness of the proposed design guidelines, as reflected through the course design?*

The interview findings indicate that the guidelines were generally perceived as pedagogically appropriate. Nonetheless, their feasibility depends on institutional resources, support, and policies combined with colleague alignment and student cooperation. Overall, there is a trade-off between the extent to which the guidelines can be incorporated and institutional constraints, such as class size and available resources. Consequently, the extent to which the guidelines can be integrated into a course depends on the institutional context in which they are applied.

9.2 Answer main research question

Which ethical and sustainability-oriented design guidelines can inform STEM course design that addresses students' use of GenAI while retaining constructive alignment?

Ten different guidelines have been constructed and validated through semi-structured interviews. CA was seen as an overarching guideline, as this concept forms the basis of any course design. Therefore, this has not been included as a separate guideline. Hence, the recognised guidelines are:

1. Rethink foundational knowledge and skills[§]
For every course, identify the foundational knowledge and skills that students need to obtain and include those in the learning objectives to facilitate students in improving these skills. Examples: foundational programming knowledge and basic scientific literacy skills.

2. Adopt authentic assessments
For every course where this change is appropriate and feasible, incorporate authentic assessments that prepare students for the labour market while aligning with available course resources. Examples: replace a traditional report with one that includes a real-world scenario, as well as a midterm and final presentation or proposal.

3. Integrate on-site assessments
For every course where this change is appropriate and feasible, integrate an on-site assessment to assess foundational learning objectives and HOTS. Examples: an oral assessment or a physical closed-book exam. For larger courses, this may also include a lightweight pass/fail midterm or final or increasing the weight of the existing on-site assessment.

4. Ensure process over product within assessments
For every assessment where GenAI use is permitted, assessments should emphasise students' learning process alongside the final deliverable to reduce GenAI reliance and increase personalised guidance, if the course resources can support this change. Examples: drafts, proposals and presentations.

5. Enhance AI literacy[¶]
For every assessment where GenAI use is permitted, courses should contribute to students' AI literacy development to foster responsible and effective use. Example: include a learning objective about fostering AI-literacy, which could involve how students should use GenAI tools or their general implications.

6. Provide rules on how to document GenAI use for assessment
For every assessment where GenAI use is permitted, provide clear guidance on how students should document GenAI use to increase reproducibility and reduce opportunities for academic misconduct

[§]For every guideline that changes a part of the CA triad: make the course compliant with CA by reflecting the change in the other parts of the triad

[¶]As a caveat, within an educational institution, there is also a responsibility of increasing the AI literacy of lecturers.

through student accountability. Example: tool(s) used, prompts, purposes, and scope of integration.

7. Focus on developing HOTS
Courses should foster the development of HOTS to retain cognitive sustainability to benefit from GenAI as a scaffolded learning partner through critical engagement. Examples: balancing AI-integration with independent reasoning, discussions and reflective exercises to ensure students experience productive struggle.

8. Employ a human-centred approach to CA
Courses should employ a human-centred approach to CA that protects intellectual and social skills that may be compromised through GenAI use, if the course resources can support this change. Examples: integrate on-site assessments and class discussions to encourage meaningful interactions and enhance student engagement.

9. Create a fair learning environment concerning GenAI
For every assessment where GenAI use is permitted, foster students' understanding of how it may be used through equitable guidance and learning opportunities to diminish the skill gap between students. Examples: provide relevant content on the course platform or dedicate a section of the syllabus to explaining how to use GenAI for the course.

10. Keep the technology close
Currently, this is an institutional guideline for standardising an educational tool within the institution. Example: when an institutional GenAI tool is available, every course that permits GenAI use should align its guidance with that tool and encourage students to use it.

Altogether, these design guidelines provide an ethical and sustainability-oriented foundation for adapting STEM courses to address students' use of GenAI, while maintaining CA. The findings show that implementation is a multidisciplinary challenge. For guidelines to be effective, both the technical side, through the infrastructure, and the social side, through the high-level tensions, are indispensable. This makes it a complex problem, complemented by semi-structured interviews, which revealed diverse perspectives among respondents and differences from findings in the literature, mainly rooted in the controversy around the topic. Outside of the tensions, this also introduced practical issues of incorporating guidelines within an educational institution. In particular, resource constraints and institutional context, together with the controversial nature of GenAI, complicate implementation. Hence, responsibility for implementation was positioned within a top-down approach, where the Executive Board of the university adopts the guidelines through an AI policy with a clear vision. To make this approach more flexible, every faculty writes its own vision as an appendix to the central policy, tackling the multidisciplinary nature of the university. This approach enables flexibility among faculties at the director of studies level and practicality for course managers. The thesis aimed to explore literature gaps and support course managers in their course design during the contemporary GenAI era. Both the academic and societal relevance are further explained in the next subsection.

9.3 Relevance of the research

9.3.1 Academic relevance

First, through the literature review, it has become clear that lecturers have a limited understanding and guidance on how to address the use of GenAI by students (Prather et al., 2025; Nguyen, 2025; Rahman et al., 2025; Wang et al., 2023b). Consequently, a study by Ala et al. (2025) noted the lack of constructive guidance for students concerning GenAI and, more generally, the need for institutional policies about GenAI use and literacy. In response, this thesis addresses knowledge gaps by exploring the technical aspects of current GenAI models to understand their broader implications, including issues like hallucinations. By examining practical applications and course-level responses to GenAI use, this thesis contributes to the AIED literature by identifying pedagogical considerations for course design.

Secondly, throughout the extensive literature review, it has been observed that there is limited research in the field of AIED that addresses sustainability. Although cognitive sustainability is

sometimes discussed, it lacks a solid foundation in neurological research, except for a study by Kosmyna et al. (2025). Consequently, this thesis argues for the inclusion of cognitive sustainability in future evaluations of the use of GenAI in education, combined with the need for further neurological research to strengthen the theoretical foundation. Given the current uncertainty regarding the long-term consequences of excessive use, this thesis highlights the relevance of considering precautionary measures.

Third, this thesis contributes to the existing literature in the field of AIED (e.g. Tran et al. 2025; Prather et al. 2023) by providing additional field research findings on how STEM university course managers perceive the use of GenAI in education. In doing so, it highlights potential institutional challenges that may arise when implementing guidelines within a STEM-oriented university. Specifically, the research aligns with previously identified resource issues, such as class size noted by Tran et al. (2025). Yet, the findings reveal additional resource constraints, such as the availability of teaching assistants and course budgets. Moreover, the thesis identifies additional constraints related to policies for course modifications at the TU Delft, diverse student motivation, and the implications of the controversial nature of a topic like GenAI for making course changes.

9.3.2 Societal relevance

Firstly, the thesis showcased an example of how the responsibilities concerning the implementation of AI policy within a STEM university can be mapped. By doing so, it revealed the role of actors, such as the Executive Board, high-level AI policy, the director of studies, programme learning objectives and policy, and the course managers, course changes. Therefore, other educational institutions with an equal bureaucratic structure may draw on this structure when implementing AI policy, combined with programme and course-level modifications. Additionally, the resulting guidelines may help institutions that struggle to comprehend and govern GenAI use by students.

Moreover, the complexity of pedagogically responsible GenAI adoption in education extends beyond solely educational considerations. Based on the model-related ethical tensions illustrated in Figure 5.1, such as data provenance and algorithmic fairness, popular commercial models are shaped by decisions made by their developers. Similar dependencies can be observed based on recognised ecological costs associated with training models and maintaining data centres, which include cooling. Consequently, an educational institution adopting these GenAI tools operates within a broader technological ecosystem over which it has limited influence. In this context, the guideline to keep the technology close can be interpreted as a partial response to these dependencies, emphasising the importance of institutional oversight and control, as reflected in initiatives such as eduGenAI. While these pedagogical models may continue to rely on externally developed LLMs, the integration of governance and pedagogical guardrails and a self-developed interface may enhance control over the technology.

Lastly, the semi-structured interviews suggest that current education was not designed for technologies such as GenAI. Examples include a change of what contemporary authentic assessment is (I.3), GenAI fraud cases not being admissible (I.1), and the need to revisit CA for your courses (I.3; I.4 with changes that already have been performed by the interviewees. This implies that GenAI within this context is a technology that made existing challenges within the current educational structures more visible. Moreover, it highlighted the trade-off between course changes and resource constraints and the importance of timely graduation, given that Dutch universities are funded in part based on the number of graduates and students enrolled within the nominal study duration. Hence, this thesis also raised awareness of a trade-off between the quality of education, resources, class sizes and revenue structures. The challenges identified in the interviews may indicate the need for broader discussion regarding structural educational change within STEM education in the Netherlands.

9.4 Reflection on the link with the CoSEM programme

This thesis is connected to the CoSEM programme, specifically the Information and Communication track (I&C). During this research, it became evident that the adoption of GenAI in STEM education is not solely a technical challenge. While technical solutions are important, their effectiveness depends on, among others, institutional regulations and social acceptance from relevant actors,

such as course managers. This became evident during the semi-structured interviews, in which participants identified several institutional practicalities that constrain the implementation of the guidelines. These findings highlighted the trade-offs between technical possibilities and social and institutional environments. Consequently, this thesis approached the problem within the STS rather than as an isolated technical issue, reflecting a core principle of the CoSEM programme. Moreover, the systems engineering component was reflected in the Design Science Research approach, which has been part of the I&C track of the master's programme. Using this research approach, this thesis designed guidelines, grounded in a comprehensive literature review and demonstrated a course redesign based on the guidelines. This supports course managers in understanding how to develop courses that retain CA and account for student use of GenAI.

9.5 How educational institutions could use the final artefact

Although this thesis provides guidelines, the findings suggest that seamless integration involves multiple trade-offs and thus depends on institutional support and governance beyond the guidelines themselves. Therefore, initially, an educational institution wanting to incorporate the guidelines at the course level should support course managers by providing them with the necessary resources to implement course changes. Still, implementation outside of a small-scale pilot is not recommended before a university-wide AI policy has been established, as this can provide institutional standardisation required for consistent application of the guidelines.

To facilitate institutional implementation, earlier practical recommendations for future research suggest developing and evaluating a framework or artefact that further supports integration of the guidelines. This implementation framework should build upon the institution's AI policy and vision. This policy and vision can operationalise the responsibilities, guidelines and implementation strategies outlined in this research within the institution's context. As part of future research, it is recommended to include ex-post validation of the remaining design requirements outlined in Figure 5.3 to strengthen the findings further. This could include a small-scale course pilot. Within this pilot, it is recommended that course managers apply the guidelines and revise their CA for the course accordingly. For instance, document analysis was applied within the thesis for the course redesign.

Finally, after such a policy and vision concerning AI is in place within the institution, it should be adopted by the highest governing board. For instance, within the TU Delft, this is the Executive Board. Subsequently, the director of studies should translate this university-wide policy into programme-level learning objectives and policy. Furthermore, course managers should align their courses with the programme objectives and policy. Ideally, this would be supported by the implementation framework or artefact from the research defined above for seamless integration. Finally, enhancing AI literacy should also support teachers in guiding students effectively, which requires institutions to include this in their Teaching Qualification programs. For example, the TU Delft can dedicate this responsibility to the teaching lab. Together, these practical recommendations provide an implementation pathway that supports educational institutions in adapting the proposed guidelines while accounting for student use of GenAI in education.

References

Adamakis, M. and Rachiotis, T. (2025). Artificial intelligence in higher education: A state-of-the-art overview of pedagogical integrity, artificial intelligence literacy, and policy integration. *Encyclopedia*, 5(4):180. https://doi.org/10.3390/encyclopedia5040180.

Ala, M., Shahid, S., Mahmud, S., Mohyuddin, S., and Kaur, K. (2025). Perceived influence of genai on student engagement in online higher education. *Journal of Applied Learning & Teaching*, 8(2):67–80. https://doi.org/10.37074/jalt.2025.8.2.16.

An, Y., Yu, J. H., and James, S. (2025). Investigating the higher education institutions' guidelines and policies regarding the use of generative ai in teaching, learning, research, and administration. *International Journal of Educational Technology in Higher Education*, 22(1):10. https://doi.org/10.1186/s41239-025-00507-3.

Ananthaswamy, A. (2023). In AI, is bigger always better? *Nature*, 615(7951):202–205. https://doi.org/10.1038/d41586-023-00641-w.

Anderson, L. W. and Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition*. Addison Wesley Longman, Inc.

Ben Otman, S., Ben Otman, S., and Karahan Adalı, G. (2025). The impact of generative AI on university students' learning experience: A study on cognitive and affective outcomes. *Journal of Information and Organizational Sciences*, 49(2):329–344. https://doi.org/10.31341/jios.49.2.10.

Biggs, J. (2003). Aligning teaching for constructing learning. *Higher Education Academy*, 1(4):1–4.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., Krathwohl, D. R., et al. (1956). Handbook i: cognitive domain. *New York: David McKay*, pages 483–498.

Bollas, A. and Bennett, C. (2026). *Designing Curricula for Learner Autonomy and Engagement*. Springer. https://doi.org/10.1007/978-3-032-13520-9.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N. S., Chen, A. S., Creel, K. A., Davis, J., ..., and Liang, P. (2021). On the opportunities and risks of foundation models. *ArXiv*. https://doi.org/10.48550/arxiv.2108.07258.

Borji, A. (2023). A categorical archive of chatgpt failures. *arXiv preprint arXiv:2302.03494*. https://doi.org/10.48550/arXiv.2302.03494.

Brown, J. S., Burton, R. R., and Bell, A. G. (1975). Sophie: A step toward creating a reactive learning environment. *International Journal of Man-Machine Studies*, 7(5):675–696. https://doi.org/10.1016/S0020-7373(75)80026-5.

Bull, C. and Kharrufa, A. (2023). Generative ai assistants in software development education: A vision for integrating generative ai into educational practice, not instinctively defending against it. *arXiv preprint arXiv:2303.13936*. https://doi.org/10.48550/arXiv.2303.13936.

Carbonell, J. R. (1970). Ai in cai: An artificial-intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine Systems*, 11(4):190–202. https://doi.org/10.1109/TMMS.1970.299942.

Chan, C. K. Y. (2023). A comprehensive ai policy education framework for university teaching and learning. *International journal of educational technology in higher education*, 20(1):38. https://doi.org/10.1186/s41239-023-00408-3.

Chan, C. K. Y. and Colloton, T. (2024). *Generative AI in higher education: The ChatGPT effect*. Routledge. https://doi.org/10.4324/9781003459026.

Cheng, A., Calhoun, A., and Reedy, G. (2025). Artificial intelligence-assisted academic writing: recommendations for ethical use. *Advances in Simulation*, 10(1):22. https://doi.org/10.1186/s41077-025-00350-6.

Chiu, T. K. (2024). *Empowering K-12 education with AI: Preparing for the future of education and work*. Routledge.

Clancey, W. J. (1983). Guidon. *Journal of Computer-Based Instruction*, 10(1-2):8–15.

Cotton, D. R., Cotton, P. A., and Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in education and teaching international*, 61(2):228–239. https://doi.org/10.1080/14703297.2023.2190148.

Deng, X. and Joshi, K. (2024). Promoting ethical use of generative ai in education. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 55(3):6–11. https://doi.org/10.1145/3685235.3685237.

Dong, X., Wang, Z., and Han, S. (2025). Mitigating learning burnout caused by Generative Artificial Intelligence misuse in higher education: A case study in programming language teaching. *Informatics*, 12(2):51. https://doi.org/10.3390/informatics12020051.

Dorji, C. and Subba, P. B. (2025). Exploring the use of Generative Artificial Intelligence (GenAI) in teaching, learning, and assessment of STEM subjects. *Educational Innovation and Practice*, 11(1):1–15. https://doi.org/10.17102/eip.11.2025.01.

Elkhodr, M. and Gide, E. (2025). The sage framework for developing critical thinking and responsible generative ai use in cybersecurity education. *Discover Education*, 4(1):517. https://doi.org/10.1007/s44217-025-00935-3.

Elsevier (n.d.). *Elsevier Researcher Academy: Navigate your research journey with confidence*. Accessed: 2026-05-10, from https://researcheracademy.elsevier.com/.

Erhardt, C., Kullenberg, H., Grigoriadis, A., Kumar, A., Christidis, N., and Christidis, M. (2025). From policy to practice: the regulation and implementation of generative ai in swedish higher education institutes. *International Journal for Educational Integrity*, 21(1):21. https://doi.org/10.1007/s40979-025-00195-6.

Feng, T. H., Denny, P., Wuensche, B., Luxton-Reilly, A., and Hooper, S. (2024). More than meets the ai: Evaluating the performance of gpt-4 on computer graphics assessment questions. In *Proceedings of the 26th Australasian Computing Education Conference*, pages 182–191. https://doi.org/10.1145/3636243.3636263.

Ferrer, X., van Nuenen, T., Such, J. M., Cote, M., and Criado, N. (2021). Bias and discrimination in AI: A cross-disciplinary perspective. *IEEE Technology and Society Magazine*, 40(2):72–80. https://doi.org/10.1109/MTS.2021.3056293.

Fontaine, J., Shahid, A., and De Poorter, E. (2024). Towards a wireless physical-layer foundation model: Challenges and strategies. In *2024 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1–7. IEEE. https://doi.org/10.1109/ICCWorkshops59551.2024.10615509.

Gavira Durón, N. and Jiménez-Preciado, A. L. (2025). Exploring the role of ai in higher education: a natural language processing analysis of emerging trends and discourses. *The TQM Journal*. https://doi.org/10.1108/TQM-10-2024-0376.

Glaser, B. G. and Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine, Chicago.

Goel, A. K. and Polepeddi, L. (2016). *Jill Watson: A Virtual Teaching Assistant for Online Education*. https://repository.gatech.edu/entities/publication/4a7ab707-3311-4c67-a3c5-c309873d75f5.

Gong, C. and Yang, Y. (2024). Google effects on memory: a meta-analytical review of the media effects of intensive internet search behavior. *Frontiers in public health*, 12:1332030. https://doi.org/10.3389/fpubh.2024.1332030.

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., Group, T. R., et al. (1999). Autotutor: A simulation of a human tutor. *Cognitive Systems Research*, 1(1):35–51. https://doi.org/10.1016/S1389-0417(99)00005-4.

Guest, O., Suarez, M., Müller, B., van Meerkerk, E., Beverborg, A. O. G., de Haan, R., Elizondo, A. R., Blokpoel, M., Scharfenberg, N., Kleinherenbrink, A., et al. (2025). *Against the Uncritical Adoption of'AI'Technologies in Academia*. https://doi.org/10.5281/zenodo.17065099.

Guo, Z., Tang, F., Luo, L., Zhao, M., and Kato, N. (2025). A survey on applications of large language model-driven digital twins for intelligent network optimization. *IEEE Communications Surveys & Tutorials*. https://doi.org/10.1109/COMST.2025.3568637.

Gwizdka, J. (2009). Assessing cognitive load on web search tasks. *The Ergonomics Open Journal*. https://doi.org/10.48550/arXiv.1001.1685.

Halinen, K., Ruohoniemi, M., Katajavuori, N., and Virtanen, V. (2014). Life science teachers' discourse on assessment: A valuable insight into the variable conceptions of assessment in higher education. *Journal of Biological Education*, 48(1):16–22. https://doi.org/10.1080/00219266.2013.799082.

Hecker, J. and Kalpokas, N. (n.d.a). *Coding Interviews*. Accessed: 2026-05-25, from https://atlasti.com/guides/interview-analysis-guide/coding-interviews.

Hecker, J. and Kalpokas, N. (n.d.b). *Disadvantages of Interviews in Research*. Accessed: 2026-05-25, from https://atlasti.com/guides/interview-analysis-guide/disadvantages-of-research-interviews.

Hecker, J. and Kalpokas, N. (n.d.c). *Interviewer Effect*. Accessed: 2026-06-01, from https://atlasti.com/guides/interview-analysis-guide/interviewer-effect.

Hecker, J. and Kalpokas, N. (n.d.d). *Social desirability bias | definition & mitigation*. Accessed: 2026-05-25, from https://atlasti.com/guides/interview-analysis.

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian journal of information systems*, 19(2):4. http://aisel.aisnet.org/sjis/vol19/iss2/4.

Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design science in information systems research. *MIS quarterly*, pages 75–105. https://doi.org/10.2307/25148625.

Holmes, W., Miao, F., et al. (2023). *Guidance for generative AI in education and research*. Unesco Publishing. https://doi.org/10.54675/EWZM9535.

Hu, K. (2023). ChatGPT sets record for fastest-growing user base – analyst note. *Reuters*. By Krystal Hu; updated February 2, 2023.

Hudson, N. C., Pauloski, J. G., Baughman, M., Kamatar, A., Sakarvadia, M., Ward, L., Chard, R., Bauer, A., Levental, M., Wang, W., et al. (2023). Trillion parameter ai serving infrastructure for scientific discovery: A survey and vision. In *Proceedings of the IEEE/ACM 10th International Conference on Big Data Computing, Applications and Technologies*, pages 1–10. https://doi.org/10.1145/3632366.3632396.

International Energy Agency (2025). *Energy and AI*. https://iea.blob.core.windows.net/assets/dd7c2387-2f60-4b60-8c5f-6563b6aa1e4c/EnergyandAI.pdf.

Jackson, J. (2025). "who's doing the thinking here?": A pedagogy-first approach to integrating large language models in higher education. In *2025 IEEE Global Engineering Education Conference (EDUCON)*, pages 1–3. IEEE. https://doi.org/10.1109/EDUCON62633.2025.11016560.

Johannesson, P. and Perjons, E. (2014). Design science. *An introduction to Design Science, Springer*, 10. https://doi.org/10.1007/978-3-030-78132-3.

Jung, K.-H. (2023). Uncover this tech term: foundation model. *Korean Journal of Radiology*, 24(10):1038. https://doi.org/10.3348/kjr.2023.0790.

Karunaratne, T., Ferati, M., et al. (2025). Harnessing the generative ai wave towards fair and diverse higher education assessments: A comprehensive analysis through an innovative lens of students. In *European Conference on e-Learning*, pages 476–484. Academic Conferences International Limited. https://doi.org/10.34190/ecel.24.1.4259.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, 103:102274. https://doi.org/10.1016/j.lindif.2023.102274.

Khan, A. and Madden, J. (2018). Active learning: a new assessment model that boost confidence and learning while reducing test anxiety. *International Journal of Modern Education and Computer Science*, 10(12):1–9. https://doi.org/10.5815/ijmecs.2018.12.01.

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., and Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv preprint arXiv:2506.08872*, 4. https://doi.org/10.48550/arXiv.2506.08872.

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., and Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI conference on human factors in computing systems*, pages 1–22. https://doi.org/10.1145/3706598.3713778.

Leiden University (n.d.). AI Guidelines for teachers & support staff. https://www.staff.universiteitleiden.nl/binaries/content/assets/ai-gidsen/ai-guide-for-teachers—new.pdf.

Liu, R., Zenke, C., Liu, C., Holmes, A., Thornton, P., and Malan, D. J. (2024). Teaching cs50 with ai: leveraging generative artificial intelligence in computer science education. In *Proceedings of the 55th ACM technical symposium on computer science education V. 1*, pages 750–756. https://doi.org/10.1145/3626252.3630938.

Liu, R., Zhao, J., Xu, B., Perez, C., Zhukovets, Y., and Malan, D. J. (2025). Improving ai in cs50: Leveraging human feedback for better learning. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1*, pages 715–721. https://doi.org/10.1145/3641554.3701945.

Long, D. and Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–16. https://doi.org/10.1145/3313831.3376727.

Lont, N. (2026). Kunnen TU-studenten nog zonder ChatGPT? 'Soms voel ik me er schuldig over'. *Delta*. https://delta.tudelft.nl/article/kunnen-tu-studenten-nog-zonder-chatgpt-soms-voel-ik-me-er-schuldig-over.

Luccioni, A. S., Viguier, S., and Ligozat, A.-L. (2023). Estimating the carbon footprint of bloom, a 176b parameter language model. *Journal of machine learning research*, 24(253):1–15. http://jmlr.org/papers/v24/23-0069.html.

Mak, J., Nakatumba-Nabende, J., Clear, T., Clear, A., Abluwi, I., Andrei, O., Angeli, L., MacNeil, S., Oyelere, S. S., Hale Rattigan, M., Sheard, J., and Zhu, T. (2026). Navigating the ethical and societal impacts of Generative AI in higher computing education. In *Proceedings of the 2025 Working Group Reports on Innovation and Technology in Computer Science Education*, *ITiCSE-WGR 2025*, page 313–359. Association for Computing Machinery. https://doi.org/10.1145/3760545.3783975.

Mattiello, H., Mattiello, D., and Wittberg, V. (2025). Driving sex-gender equity and ethical integration in Edu X. 0: Harnessing GenAI for human-centric innovation, responsibility, and industry X. 0 (When X. 0= 5.0). In *2025 IEEE Global Engineering Education Conference (EDUCON)*, pages 1–10. IEEE. https://doi.org/10.1109/EDUCON62633.2025.11016475.

Messeri, L. and Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature*, 627(8002):49–58. https://doi.org/10.1038/s41586-024-07146-0.

Mishra, P., Group, D.-P. R., et al. (2012). Rethinking technology & creativity in the 21st century: Crayons are the future. *TechTrends*, 56(5):13. https://doi.org/10.1007/s11528-012-0594-0.

Nguyen, K. V. (2025). The use of generative ai tools in higher education: Ethical and pedagogical principles. *Journal of Academic Ethics*, 23(3):1435–1455. https://doi.org/10.1007/s10805-025-09607-1.

NOS (2026). *Estland wil AI-sprong maken op scholen met speciale versie ChatGPT*. https://nos.nl/artikel/2609339-estland-wil-ai-sprong-maken-op-scholen-met-speciale-versie-chatgpt.

Npuls (n.d.). *EduGenAI*. https://npuls.nl/edugenai.

Nwana, H. S. (1990). Intelligent tutoring systems: an overview. *Artificial Intelligence Review*, 4(4):251–277. https://doi.org/10.1007/BF00168958.

Offermann, P., Blom, S., Schönherr, M., and Bub, U. (2010). Artifact types in information systems design science–a literature review. In *International conference on design science research in information systems*, pages 77–92. Springer. https://doi.org/10.1007/978-3-642-13335-0_6.

Oliveira, E., Mohoni, M., and Rios, S. (2024). Towards explainable authorship verification: an approach to minimise academic misconduct in higher education. In *International Conference on Artificial Intelligence in Education*, pages 87–100. Springer. https://doi.org/10.1007/978-3-031-64315-6_7.

OpenAI (2025). *Introductie van de studiemodus*. https://openai.com/nl-NL/index/chatgpt-study-mode/.

OpenAI (n.d.). *ChatGPT*. https://openai.com/nl-NL/index/chatgpt/.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *bmj*, 372. https://doi.org/10.1136/bmj.n71.

Popenici, S. A. and Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and practice in technology enhanced learning*, 12(1):22. https://doi.org/10.1186/s41039-017-0062-8.

Poundstone, W. (2017). *Head in the Cloud: Dispatches from a Post-fact World*. Simon and Schuster.

Prather, J., Denny, P., Leinonen, J., Becker, B. A., Albluwi, I., Craig, M., Keuning, H., Kiesler, N., Kohn, T., Luxton-Reilly, A., et al. (2023). The robots are here: Navigating the Generative AI revolution in computing education. In *Proceedings of the 2023 working group reports on innovation and technology in computer science education*, pages 108–159. https://doi.org/10.1145/3623762.3633499.

Prather, J., Leinonen, J., Kiesler, N., Gorson Benario, J., Lau, S., MacNeil, S., Norouzi, N., Opel, S., Pettit, V., Porter, L., et al. (2025). Beyond the hype: A comprehensive review of current trends in generative ai research, teaching practices, and tools. *2024 Working Group Reports on Innovation and Technology in Computer Science Education*, pages 300–338. https://doi.org/10.1145/3689187.3709614.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2023). Exploring the limits of transfer learning with a unified text-to-text transformer. *ArXiv*. https://doi.org/10.48550/arXiv.1910.10683.

Rahman, G., Almutairi, E. A. A., Almutairi, M., and Mudhsh, B. A. (2025). The impact of ai-powered text generation tools on the critical thinking skills of undergraduate students. *Journal of University Teaching and Learning Practice*, 22(7):1–21. https://doi.org/10.53761/srkt0510.

Ramsden, P. (2003). *Learning to teach in higher education*. routledge.

Randall, N., Wäckerle, D., Stein, N., Goßler, D., and Bente, S. (2023). What an ai-embracing software engineering curriculum should look like: An empirical study. *IEEE Software*, 41(2):36–43. https://doi.org/10.1109/MS.2023.3344682.

Rawlusyk, P. (2018). Assessment in higher education and student learning. *Journal of Instructional Pedagogies*, 21.

Rijksoverheid (n.d.). Financiering hoger onderwijs. Accessed: 2026-05-23, from https://www.rijksoverheid.nl/onderwerpen/financiering-onderwijs/financiering-hoger-onderwijs.

Ross, E. A. and Baines, J. (2024). Treading water: new data on the impact of ai ethics information sessions in classics and ancient language pedagogy. *Journal of Classics Teaching*, 25(50):181–190. https://doi.org/10.1017/S2058631024000412.

Rudolph, J., Tan, S., and Tan, S. (2023). Chatgpt: Bullshit spewer or the end of traditional assessments in higher education? *Journal of applied learning & teaching*, 6(1):342–363. https://doi.org/10.37074/jalt.2023.6.1.9.

RudolphA, J. (2023). Editorial 5 (si1): Technology and education: Innovation or hindrance. *Journal of Applied Learning & Teaching*, 6(1). https://doi.org/10.37074/jalt.2022.5.s1.1.

Sandhaus, H., Gu, Q., Parreira, M. T., and Ju, W. (2025). Co-Designing with transformers: Unpacking the complex role of GenAI in interactive system design education. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*, pages 1228–1243. https://doi.org/10.1145/3715336.3735805.

Stadler, M., Bannert, M., and Sailer, M. (2024). Cognitive ease at a cost: Llms reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior*, 160:108386. https://doi.org/10.1016/j.chb.2024.108386.

Starominski-Uehara, M. (2026). *AI is degrading the value of a university degree.* Accessed: 2026-06-05, from https://blogs.lse.ac.uk/impactofsocialsciences/2026/05/28/ai-is-degrading-the-value-of-a-university-degree/.

Stephanou, C. (2025). *GenAI history: A living timeline of generative AI breakthroughs.* https://helixiora.com/genai-history-a-living-timeline-of-generative-ai-breakthroughs/.

SURF (n.d.). AI-hub. https://www.surf.nl/themas/artificial-intelligence/projecten-en-samenwerkingen/ai-hub.

Susnjak, T. and McIntosh, T. R. (2024). Chatgpt: The end of online exam integrity? *Education Sciences*, 14(6):656. https://doi.org/10.3390/educsci14060656.

The Brussels Times Newsroom (2026). *UGent chancellor Petra De Sutter caught using AI in opening speech.* https://www.brusselstimes.com/1909581/ugent-chancellor-petra-de-sutter-caught-using-ai-in-opening-speech.

Tran, T. M., Bakajic, M., and Pullman, M. (2025). Teacher's pet or rebel? practitioners' perspectives on the impacts of chatgpt on course design. *Higher Education*, 90(3):777–800. https://doi.org/10.1007/s10734-024-01350-7.

Tsolev, N. (2025). *The effect of ChatGPT on students' learning performance: insights from a meta-analysis.* Accessed: 2026-06-20, from https://genaisa.eu/the-effect-of-chatgpt-on-students-learning-performance-insights-from-a-meta-analysis/.

TU Delft Teaching Support (2025). AI chatbots in unsupervised assessment. https://www.tudelft.nl/teaching-support/educational-advice/assess/guidelines/ai-chatbots-in-unsupervised-assessment.

TU Delft Teaching Support (n.d.). *Design AI-resilient assessments.* https://www.tudelft.nl/teaching-support/educational-advice/ai-in-education/ai-in-assessment/design-ai-resilient-assessments.

van Bergen, P. and Lane, R. (2016). Should we do away with exams altogether? No, but we need to rethink their design and purpose. *The Conversation.*

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Arxiv.* https://doi.org/10.48550/arXiv.1706.03762.

Vendrell, M. and Johnston, S.-K. (2026). Scaffolding critical thinking with generative ai: Design principles for integrating large language models in higher education. *Computers and Education: Artificial Intelligence*, page 100572. https://doi.org/10.1016/j.caeai.2026.100572.

Wang, J. and Fan, W. (2025). The effect of chatgpt on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanities and Social Sciences Communications*, 12(1):1–21. https://doi.org/10.1057/s41599-025-04787-y.

Wang, J., Liu, S., Xie, X., and Li, Y. (2023a). Evaluating aigc detectors on code content. *arXiv preprint arXiv:2304.05193*. https://doi.org/10.48550/arXiv.2304.05193.

Wang, T., Díaz, D. V., Brown, C., and Chen, Y. (2023b). Exploring the role of AI assistants in computer science education: Methods, implications, and instructor perspectives. In *2023 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 92–102. IEEE. https://doi.org/10.1109/VL-HCC57772.2023.00018.

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., and Waddington, L. (2023). Testing of detection tools for ai-generated text. *International Journal for Educational Integrity*, 19(1):1–39. https://doi.org/10.48550/arXiv.2306.15666.

Weizenbaum, J. (1976). *Computer Power and Human Reason: From Judgment to Calculation*. W. H. Freeman.

Wokke, M. E., Padding, L., and Ridderinkhof, K. (2019). Creative brains show reduced mid frontal theta. *bioRxiv*, page 370494. https://doi.org/10.1101/370494.

Wood, D. and Moss, S. H. (2024). Evaluating the impact of students' generative ai use in educational contexts. *Journal of Research in Innovative Teaching & Learning*, 17(2):152–167. https://doi.org/10.1108/JRIT-06-2024-0151.

Yan, Q. (2025). A study on self-supervised pretraining of large language models on higher education text data. In *2025 2nd International Conference on Intelligent Computing and Robotics (ICICR)*, pages 242–246. IEEE. https://doi.org/10.1109/ICICR65456.2025.00049.

Zdravkova, K., Dalipi, F., and Ahlgren, F. (2023). Integration of large language models into higher education: a perspective from learners. In *2023 International Symposium on Computers in Education (SIIE)*, pages 1–6. IEEE. https://doi.org/10.1109/SIIE59826.2023.10423681.

Zhai, X. (2022). Chatgpt user experience: Implications for education. *Available at SSRN 4312418*. http://dx.doi.org/10.2139/ssrn.4312418.

Zou, L. (2025). Chatgpt in improving college english learning outcomes: A comparative study of traditional vs. ai-assisted coursework. In *Proceedings of the 2025 2nd International Symposium on Artificial Intelligence for Education*, pages 367–370. https://doi.org/10.1145/3775073.3775132.

Appendices

A Ethics statement

The thesis followed ethical guidelines in line with the TU Delft Human Research Ethics Committee application. That is because I'm a master's student who collected data from Human Research Subjects through semi-structured interviews. Hence, during the thesis, personal data has been collected. This data has been securely stored throughout the thesis in the TU Delft OneDrive, and only anonymised interview transcripts were added to the appendix. All interviewed participants were digitally informed before the interview via an informed consent form. This form included the purpose of the study, its storage, the data handling process, when the data will be destroyed, and the possibility of being excluded from the study. Moreover, all interview transcripts and codes were sent towards the interviewees for feedback within a week after the interview. Concerning AI or GenAI tools, ChatGPT and Grammarly with its Grammarly AI, have been used, as stated in the Appendix B: AI-statement.

B AI-statement

OpenAI's GPT-5.4 was used through the GenAI application ChatGPT to find the search strings for the literature review. First, I defined keywords myself to safeguard academic intregrity, for example for string one: generative artificial intelligence OR infrastructure AND transformer architecture OR next-token prediction AND fairness OR academic integrity quality OR transparency OR bias OR privacy and data security OR knowledge indistinguishability OR ethics OR responsible artificial intelligence OR computational demand OR energy consumption OR over-reliance OR problem-solving OR critical thinking OR neurological OR resistance to change OR sustainability OR environmental impact. And for search string two: generative artificial intelligence AND STEM education OR teaching OR constructive alignment OR curriculum OR course design OR pedagogy. Afterwards, I gave some background information: "You are a student conducting a literature review for an MSc thesis with two different search strings within academic research towards accounting for students' use of GenAI. These should find answers to the following research sub-questions: {sub-question 2 and 3 of the research}. Be very critical, here are already some terms which you should consider {terms defined above}." I also had to be critical since it gave terms not relevant to the research, or too broad, leading to too many search results.

In addition, OpenAI's GPT-5.4 has been adopted to formulate sub-question 1. After writing the literature review, it felt like too much information was included in a single chapter. So, I decided to move the history of AI in education, the technology behind it, and why we should use it in the first place to a separate chapter. For finding the sub-question, I used the prompt: "Within a student's MSc thesis research that tries to design guidelines to cope with GenAI use by students. The literature review assesses both ethical and sustainable implications of using GenAI. It also discusses topics and guidelines from current AI-in-education research. But before that, there is a method overview of the literature review, something about the history of AI in education, why GenAI should be used in the first place, and about the technology behind it. How would you define a sub-question that addresses this, and becomes a new sub-question 1 with this as the rest of the sub-questions for the study: current sub-question 2-5. I was thinking something like: How did the state-of-the-art of the history of AI in education, its benefits, and the technology behind it, create relevance for students to use it within STEM education?" I used an example at the end to safeguard academic integrity.

Furthermore, OpenAI's GPT-5.4 has been consulted for reviewing my interview protocols. This was especially useful since it felt like another person could critique/evaluate my thoughts. Within the dialogue, I made sure to keep the questions related to my goal of the interview within the study context. Hence, I was able to use it assistive and not authoritative way to safeguard academic integrity. The prompt I used: "As an MSc student who wants to do a small interview study to gain insights into what lecturers need to cope with GenAI use by students and concurrent lecturer perceptions on GenAI use by students. Predict the duration of the following number of questions, and be critical if the themes and questions are clear: my initial set of themes and questions".

Moreover, OpenAI's GPT-5.4 has been utilised to convert online sources to BibTeX, to be able to easily integrate them in the main.bib of Overleaf. This included prompts such as: "Can you convert this source to BibTeX when the style is apalike: https://delta.tudelft.nl/article/kunnen-tu-studenten-nog-zonder-chatgpt-soms-voel-ik-me-er-schuldig-over?" This process involved being cautious, since sometimes the model hallucinated, and the author or date was incorrect. Therefore, I always had to check within the source whether all the fields were corresponding.

Also, OpenAI's GPT-5.5 came in helpful to create tables in Overleaf. This would be tedious work to do on my own and, therefore, a waste of time. To do so, I've used a prompt such as: "Without changing any words, I think it is valuable to put these requirements into a table with a column for requirements, and a column for function. Convert it to a LaTeX table: {my initial set of requirements}". Yet, it was needed to be cautious, since for this specific prompt the AIGC contained my text in the requirements column and still wrote something in the function column. Accordingly, this information had to be removed.

Lastly, during the preparation of this work, I used Grammarly and the Grammarly Assistant

to improve my spelling and grammar. Concerning the Grammarly Assistant, I solely used it with "Improve it" or "Shorten it" to rewrite my text, and only did this when I was unable to improve the text myself. Other input to the tool would not safeguard academic integrity, in my opinion. After using this tool/service, I reviewed and edited the content as needed, and I take full responsibility for the content of my research proposal.

C Search Strings

This appendix presents the two search strings used for the scoping literature review, which formed the basis largely for Chapter 3, Chapter 4, and Chapter 5. Both are shown in the format used for the advanced search in Scopus. The first search string was defined as follows: TITLE-ABS-KEY (("generative AI" OR "large language model" OR "foundation model") AND ("STEM education" OR "higher education") AND ("next-token prediction" OR "training data" OR "pre-training" OR "fine-tuning" OR "black box" OR opacity OR "probabilistic" OR "transformer model") AND (ethics OR "academic integrity" OR bias OR fairness OR transparency OR privacy OR "data security" OR "energy consumption" OR "computational cost" OR sustainability OR "environmental impact") AND (learning OR "problem solving" OR "critical thinking")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "ch"))

Another search string was adopted as follows: TITLE-ABS-KEY (("generative AI" OR "AI in education" OR "intelligent tutoring system") AND ("STEM education" OR "higher education") AND ("instructional design" OR "course design" OR "curriculum design" OR "curriculum development" OR "constructive alignment") AND ("learning outcome" OR assessment OR "student engagement") AND (guideline OR framework OR principle)) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "bk") OR LIMIT-TO (DOCTYPE , "ch"))

D The use of chatGPT in assignments

This appendix presents a document for rules on AI for the Data Science in TPM course. It is not a direct copy since some PII has been removed for privacy purposes.

The use of AI tools in (non-invigilated) assessments

About AI tools[‖]

AI tools like the chatbot 'chatGPT' can produce convincing answers to certain questions, depending on their nature. However, their output is not always reliable: outputs can contain convincingly presented factual errors. Furthermore, their training data can be outdated. It is also important to note that most chatbots do not list their resources. Other disadvantages can be attributed to the output of AI tools, such as potential discriminatory outcomes, as AI tools inherit biases present in the training data.

On the positive side, chatbots can help with for instance:

- Checking grammar and spelling in a text
- Generating ideas by listing information from different sources in an accessible way
- Summarising or expanding texts and concepts
- Converting text to a different competence level (for instance from a master level to a bachelor level)
- Suggesting alternatives for key words to include in search queries e.g. when searching for literature

The use of AI tools in (non-invigilated) assessments in TB142Ic

In this course, we make use of non-invigilated assessments such as writing group papers or reports. As lecturers we are assuming that AI tools are used by our students, especially because these services are easy to use, easy accessible, and currently free of charge. Therefore, we would like to provide you as a student with clear rules on how AI tools can be used during this course.

These rules are provided because prohibiting the use of AI tools is not feasible or realistic. However, it is strongly recommended that AI tools are not used for assessments in TB142Ic. After all, the assessments are designed to assess students' understanding of the content, as well as their proficiency of various academic and professional skills (analysis, critical thinking, argumentation, research skills, reflection, etc.). Including AI-generated texts in assessments therefore compromises the learning objectives of the assessments, as well as the learning objectives of the overall TB Bachelor programme. Furthermore, as a student you are the responsible one for the content you submit. Making inadequate use of ChatGPT does not uphold this responsibility. However, if you do choose to make use of AI tools in non-invigilated assessments in this course, the following rules apply.

Rules when using AI tools

Under no circumstances, the AI generated output should be copied without critically evaluating it. If AI tools are used, they merely should serve as an additional aid/tool, and not as a replacement of the actual work that needs to be done. It is key that AI generated output is critically evaluated, which can only be done if there is a clear understanding of the actual content, relevance, and context of the output. Mere replication or verbatim copying of AI generated output without thoughtful and critical evaluation, a clear understanding, and the necessary adaptation, is not acceptable and will be considered as a violation of academic integrity.

Furthermore, it is key that the output generated by AI tools is validated. This validation should be done by searching for and including reliable (academic) sources. A lack of this validation might lead to the introduction of incorrect or incomplete content in your assessments.

¶

[‖]See https://www.tudelft.nl/teaching-support/didactics/assess/guidelines/ai-chatbots-in-unsupervised-assessment for more information 2

If AI is used in an assessment, it is needed to acknowledge this use and properly reference to this use. After all, it is the norm in academia that it is substantiated how you came to certain statements, insights, conclusions, etc. The acknowledgment of and reference to the use of AI is also needed to ensure the distinction between the students' original ideas and those provided by AI, and to check whether the student critically checked the output of the AI-generated outcome.

Besides including an acknowledgement and reference to the used AI tools, it is needed to include a reflection section at the end of your assessment. This reflection should include: an explanation of how you used AI and for which parts, the prompts you used to get the results, how you verified the generated output, and how you safeguarded academic integrity.

Compliance to these rules

Disregarding the above rules can have an effect on the quality of the learning process as a student. But the potential effects can even go further. After all, during your studies, you are being educated and trained for your future professional career.In this professional career, you might have the (end) responsibility for certain actions and decisions taken. This responsibility implies that society should be able to rely on academia to deliver students that, if they make use of AI tools, they do it in a responsible way (by following the above formulated rules). During the study career it is therefore crucial to use AI tools in a responsible way, so this responsible use also becomes the norm during the professional career.

If during this course, there are indications that the above formulated rules have not been followed, there will be referral to the Board of Examiners. After all, failure to follow these rules is in violation of academic honesty policies and can be seen as fraud ("fraud is taken to mean any act or omission by a student that makes it fully or partially impossible to properly assess the knowledge, insight and skill of that student or another student" **).

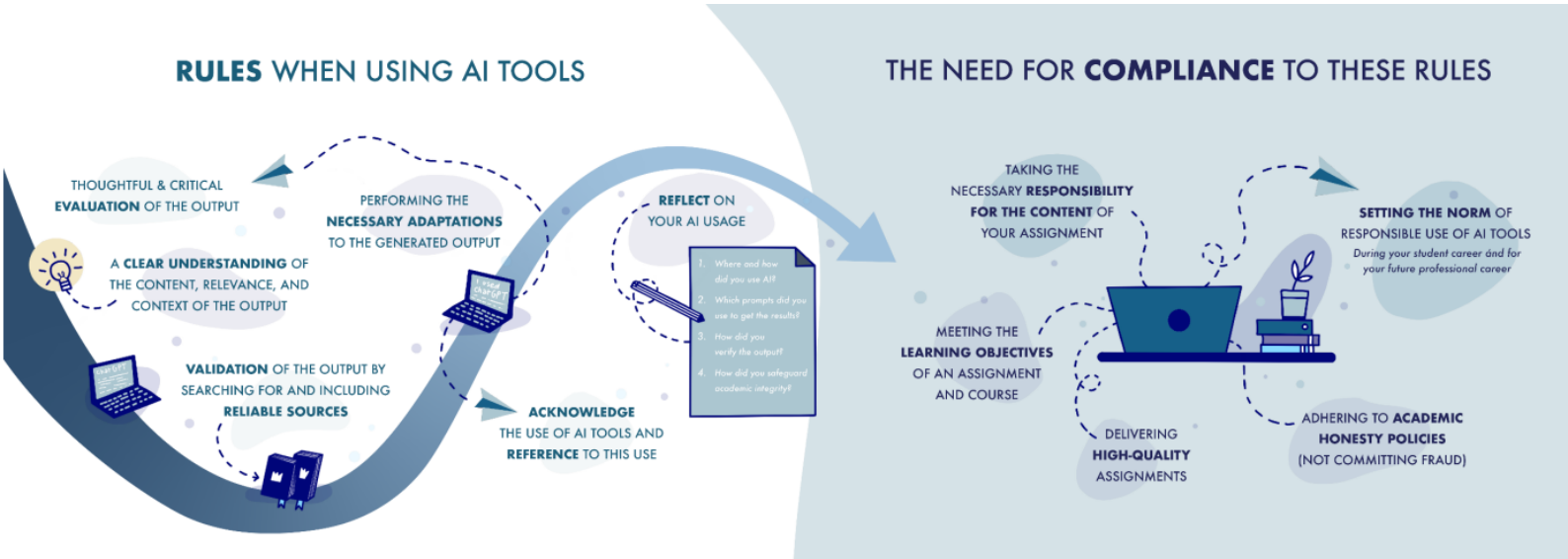

Figure D.1: Responsible use of AI through rules and its compliance

**Source: model Rules and Guidelines Board of Examiners BSc / MSc 2023-2024, article 7.1

E Semi-structured questions

This appendix shows the questions for both interview rounds one and two, which were both semi-structured. Hence, in the interview notes and transcripts, additional text or questions may appear from the interviewer.

E.1 Questions round 1

GenAI experience & perception
Do you have a positive or negative feeling towards students currently using GenAI in (your) courses, if that is the case?

Does current student use of GenAI align with how you would like it to be used?

Have you encountered or observed concrete problems with students using GenAI for their coursework?

Do you experiment with GenAI yourself for your courses, and if so, for what purpose?

Values & needs
What core values of your course do you think are at stake in terms of learning goals or assessment methods with student GenAI use?

What guidance within the university is currently available that would help you adapt your courses to deal with GenAI use by students? (from Teaching support, there is prompt training for GenAI, AI literacy training and workshops for designing AI-resilient assessment)

What do you think is missing?

Teaching & responsibility
Would you be comfortable teaching students how to use GenAI or guide them or at least encourage them to use it within your courses?

Do you think it is the responsibility of the university, the lecturer or the student to decide how to work with GenAI?

Policy & enforcement
Does this perception differ for you depending on the type of course or assessment method?

Have you encountered a situation where you suspected overuse of GenAI by a student? If so, what did you do?

Did you already or plan to implement guidelines or rules in your course(s) to cope with students' use of GenAI?

If implemented already, how effective were they in the sense that you reached your intended goal?

Do you have any questions or additional side notes for me?

E.2 Questions round 2

Perception guidelines and course redesign
When having read the pre-notes, what are the first things that caught your eye?

Guideline feasibility
Do you feel these guidelines respond adequately to the challenges and opportunities of student

GenAI use?††

Do you feel these guidelines apply to students with different levels of intrinsic motivation?‡‡

How feasible would these guidelines be for you to incorporate in your courses?

What support or policy changes from the university would guide you in dealing with GenAI use by students?

Implementation
If necessary: In an earlier interview, a lecturer raised scepticism about requiring an AI appendix stating information about tools used, prompts, purpose. What is your opinion on such an appendix?

If you would like to implement these guidelines for your courses, what institutional steps would you need to take?

What would be an ideal way for implementing such guidelines, if you would do so in the first place?

Do you have any questions or additional side notes for me?

††From interview 3 changed to: Do you feel these guidelines respond adequately to the ethical and sustainability challenges, but also the opportunities of student GenAI use?

‡‡Added from interview 3 onwards

F Coding schemes

This appendix details the development of the open codes and code groups. Open coding was used to identify codes in the deepening interviews, while for the validation interviews were coded deductively using predefined code groups. Within these predefined groups, additional supporting open codes were identified where relevant. Table F.1 relates to subsection 7.2, while Table F.2 relates to subsection 7.3. The validation interviews were coded using a deductive approach; therefore, the first column of Table F.2 contains the predefined coding groups. In both tables, superscript 1 indicates that a code was derived from the deepening interviews, while superscript 2 signifies that a code was derived from the validation interviews.

F.1 Practicalities codes

Table F.1: The codes identified in the deepening and validation interviews concerning practicalities, grouped into 5 groups

Code Group	**Open Code**
On making course changes	Course changes deadline[1] Course change process[1] Resource constraints[1] Impact of colleagues' mindsets[2] Impact of diverse personal opinions[2]
On changes towards their own courses	Increase weight assessment [1] More on-site assignments[1] GenAI supported assessment[1]
On responsibility	University high-level guidelines[1] Lecturers GenAI use within their course[1] Board of Examiners safeguarding quality of assessments[1] Influence of individual research interests[2] Students committing plagiarism[2]
On institutional issues	Challenging fraud process[1,2] Difficulty implementing AI misuse procedures[2] Pressure towards larger-scale courses[2]
On generalisability	Central policy[2] Faculties vision integration[2] General for faculties and flexible for specificity[2]

F.2 Guidelines and design requirements codes

Table F.2: The codes identified in the deepening and validation interviews regarding the guidelines and design requirements, divided into 12 groups

Predefined code Group	**Open Code**
On rethinking foundational knowledge and skills relevance	Assessment purpose[2] Rethink learning objectives[2] Shifted skills labour market[2] Examples foundational knowledge and skills[2] Gatekeeping foundational skills[2]
On authentic assessment market	Importance of learning skills labour [2] Lack of assessment development[2]
On integrating on-site assessment	Change towards on-site assessment[1]

Predefined code Group	**Open Code**
	Resource constraints[1,2]
On process over product	Accessibility of co-creation[2]
On enhancing AI literacy	Access to AI tools[2] Elements AI literacy[2] Need for lecturer AI literacy necessity[2]
On documenting AI use	Template is an improvement[1] Student motivation[1] Student honesty[2] Accountability mechanism[2] Alignment with course learning objectives[2]
On developing HOTS	Adaptation of pedagogy[2]
On protect human agency	Value of the physical learning process[2] Social isolation[2]
On creating a fair learning environment concerning GenAI	Standardised AI tool[2] Lack of AI training in curricula[2]
On pedagogical tooling (keep the technology close)	Quality private company tools[2] Accustomed to contemporary tools[2] Loopholes with private company tools[2] Prompt monitoring[2]
On perceived feasibility and pedagogical appropriateness	Class size[2] Difficulty enforcing pedagogical tooling[2] Insufficient AI literacy[2] Implementation level[2] Existing institutional support[2]
On perceived applicability of guidelines for varying levels of intrinsic motivation	Unmotivated students challenges[1] Lower lecture attendance[1,2] Educational ambitions[2] Hidden intrinsic motivation[2] Learning environments[2] Importance CA[2] Changing views on education[2]

G Transcript deepening interviews

This appendix contains the reports of the deepening interviews. Although the interviewees are anonymous, they are presented per interviewee. This round includes three interviews. [I] refers to text being said by the interviewer, while [R] refers to text being said by the respondent.

G.1 Interview 1

GenAI experience & perception
[I]: Do you have a positive or negative feeling towards students currently using GenAI in (your) courses, if that is the case?
[R]: I have a rather negative feeling. It's a big change, and we are not fully prepared. The current structures are no longer working; constructive alignment is changing. Difficult to assess, and difficult to teach because assessment is a motivator. Not saying it should not be used, we need to learn some skills and knowledge. Bachelor's courses are currently compromised.

[I]: Does current student use of GenAI align with how you would like it to be used?
[R]: Not yet. Trying to offer a lot of advise towards students. Trying to have in-class assignments. AI is not supposed to be used in the assignments for one course. I tell them to what extent AI can be used, for example, for information, and put a footnote if used for text generation. None of the students did it; structuring and improving grammar was their reasoning. Uncomfortable situations, cannot trust the integrity of students. End up grading with a sort of best trust. Grades are up on .8, and 1 from last year and the year prior. Did they really become better, or did AI play a significant role? No, I feel like the soft recommendation style is not working.

[I]: Have you encountered or observed concrete problems with students using GenAI for their coursework?
[R]: Yeah. One of the problems is that they do not learn a thing. A course is designed for learning; the fact that you are less involved is detrimental. In a course, students have debates, which are formative. It's a small group where debate skills are tested. Last year, the models were not great, but a lot of involvement. Now, a better model, but the presentation is worse. Explanation, interpretation, they are less confident at all. A bachelor's student makes a difference, too. Do not want to grade them on an English presentation.

[I]: Do you experiment with GenAI yourself for your courses, and if so, for what purpose?
[R]: A little bit. Use it for my own research. For courses specifically, I ask to generate some answers for models. Then, I take ideas to make the model for the assessment. Can you generate answers for the model? I do not generate exam questions. Sometimes, give feedback on the draft questions. Test the quality of exam questions, not generate them. Case studies, can you think of a case where this applies?

Values & needs
[I]: What core values of your courses do you think are at stake in terms of learning goals or assessment methods with student GenAI use?
[R]: Written assignment is the most vulnerable, essay style. Excel calculation exercises are going to be the next in line. To the extent that they are using it, they are not learning from it. Not really testing the essay, but more model; thought the visual would be out of scope for AI. The fences we put around them worked for some time, but now they can go over them. Good faith as a teacher. Maybe some students will, but some will not, which is a little unfair. Took a nice course on the Teaching & Learning Services (TLS) on GenAI in assessment. They talked about students' roles, what I am willing to teach them. Role as a builder, and as a vetter. A builder is someone who can create with their own hands, rather than a signature under a project, like is it good enough. With GenAI, you can already say it is good enough. You want to test student skills as a builder, not a vetter. Go to AI safe environments for testing builder, but how could you test for vetter? Have they cross-referenced their output? If you want to act as a vetter, you also need the skills of a builder. Typical assignment during the course tests is being a vetter before a builder, which does not work. It has to be the other way around.

[I]: What guidance within the university is currently available that would help you adapt your courses to deal with GenAI use by students? (from Teaching support, there is prompt training for GenAI, AI literacy training and workshops for designing AI-resilient assessment)
[R]: Not an expert in educational science. There is a TLS website with recommendations, but no TU Delft-wide guidelines. The point about random oral checks is a bit difficult. Some of it is not really applicable/up-to-date.

[I]: What do you think is missing?
[R]: Needs to be firmly developed. Currently expanding on the builder versus vetter idea. Documenting how you have prompted AI is unrealistic. Too much to document. The 'lazy' students will still do it with AI.

Teaching & responsibility
[I]: Would you be comfortable teaching students how to use GenAI or guide them or at least encourage them to use it within your courses?
[R]: Something I would consider doing more. I say in which ways I want them to use it. If you have to read a lot, you may create an abstract of a book chapter, but also read the book chapter. No way to enforce it for me at all. Skills I'll still be testing manually, which includes lots of good faith. I am testing their skill as a builder, but AI is currently able to be the builder. Cannot be soft boundaries, should be very structural.

[I]: Do you think it is the responsibility of the university, the lecturer or the student to decide how to work with GenAI?
[R]: University to the current level. Implementation at the level of the lecturer. The university should have a clear structure for the builder and vetter. If this course is testing being a builder, then do it this way and test this way. Currently working towards a grading rubric for students' AI work. If we have an environment with AI, we should consider using full AI. Documenting AI use may die down, at the bare minimum, with current students. Needs to move to oral examination for thesis.

Policy & enforcement
[I]: Does this perception differ for you depending on the type of course or assessment method?
[R]: One of the courses involved coding. The moment you use it for one thing, you lose hope. Ultimately, I would like it to be very black and white. Or use it entirely, or not. Both for bachelor's and master's it can be applied in the same way.

[I]: Have you encountered a situation where you suspected overuse of GenAI by a student? If so, what did you do?
[R]: Heard of people doing it. The Board of Examiners (BoE) cannot do anything with it. You cannot prove the use. Did about plagiarism in the past, and students really tried to debate. Then, BoE would be on the student side; could it be the case that this has happened unintentionally? A really difficult process.

[I]: Did you already or plan to implement guidelines or rules in your course(s) to cope with students' use of GenAI?
[R]: I have not fully figured it out. Thought of increasing the weight of the oral exam part of it. Something that is to be finalised because it is a long process. Assessment changes have to be done a year in advance. Delighted that people were in class learning on the spot. If not attending, not part of a group, no grade. Was good to move out of AI design to do things in class. In the thesis work, it is difficult to work with; it is a big topic. Documentation is a bit utopian; you can generate it with AI. No defence in bachelor's thesis so far. Supervising in good faith. I want people to have the skills to talk about it.

[I]: If implemented already, how effective were they in the sense that you reached your intended goal?
-

[I]: Do you have any questions or additional side notes for me?
[R]: Curious about the course design. Feel free to send the results. What might be within reach, firm

guidelines are difficult to establish. But best practices for teachers to tweak courses are important. Opportunity for educational support on the Brightspace to keep the learning going. Currently couple of working groups on AI in Education.

G.2 Interview 2

GenAI experience & perception
[I]: Do you have a positive or negative feeling towards students currently using GenAI in (your) courses, if that is the case?
[R]: Slightly negative I would say. So, for instance, I asked students to write an essay. Some of them, around half of them, were generated by AI. When I read them, it is a waste of time. They do not learn how to do proper desk research; they just ask AI. Is less efficient than when doing it manually. Less critical thinking is involved.

[I]: Does current student use of GenAI align with how you would like it to be used?
[R]: Well, they can use it, let's say, as a copilot. They should not let their work be steered by the AI tool. I think it is hard to enforce.

[I]: Have you encountered or observed concrete problems with students using GenAI for their coursework?
[R]: Well, I think it is quite a problem. Generating an article by AI is rubbish. There is no genuine text any more, the same with project reports. Yes, they can learn from AI by generating ideas, but students should learn by creating themselves. I think I do not know. I was teaching in Pakistan, and students were also using it a lot. Also, for video generation, they were really vague/high-level. That should not be enough to pass a university course.

[I]: Do you experiment with GenAI yourself for your courses, and if so, for what purpose?
[R]: Yes, I have done some ideation. When you want to set up a new class activity, think from AI. Some good examples from AI that you can work out and try in class. You have to filter a lot still. It gave me some quite obvious ideas. They refresh your mind about what your overall knowledge of the world is. But if you want to outline a course or workshop, you still need to give it your personal touch. You still need to think for yourself. You ask AI to come up with exercises, usually the lower levels of Bloom's taxonomy. AI is mostly appropriate for the lower levels of the taxonomy.

Values & needs
[I]: What core values of your course do you think are at stake in terms of learning goals or assessment methods with student GenAI use?
[R]: Well, I decided to skip the essay. I wanted to let students think more critically about topics, so they do a deep dive. The essay does not work out here. Next year, try to work with oral exam. Can put things into writing, but they need to be able to explain skills face-to-face.

Before June 22, you have to upload the new course information. In general, the course stays the same, so you do not have to ask for a new course code. If there is a big change in the course, however, then you have to go for a new course code. So, not applicable if they have to do the same, but they are not writing, but presenting.

Still fine doing it years in advance, when doing the course, you have to stick to the prescribed courses. If you are doing it a bit more vaguely, you leave yourself room for yourself to change the implementation a bit. If it is written like you have an individual written exam and a group assessment, that will not change. Saw a course where the students had to do an AI assessment for the group work. For the project, they go abroad in the second quarter. In the first quarter, they have customer interviews. Most of the times not an option to ask the people on-site. So they can practice the interview style. To the AI, ask questions to behave as a certain customer, and give it a customer profile. Students get some ideas by doing it this way. The trick is to verify it with real-world people.

[I]: What guidance within the university is currently available that would help you adapt your courses to deal with GenAI use by students? (from Teaching support, there is prompt training for

GenAI, AI literacy training and workshops for designing AI-resilient assessment?
[R]: I think there was this from the Teaching Lab, a web page on the use of AI. I looked into it, received a message about it, and looked up the website. Was quite vague, to be honest. No strict policy on the use of AI in the university. Just acknowledge that it is difficult to cope with.

[I]: What do you think is missing?
[R]: Would not be in favour of strict guidelines, I want to have some room. Can raise awareness by coming up with examples. What works well does not work well. Also, there is no tool to track the use of AI. When I went to Pakistan, they had a tool to check on AI. We need to be able to use the trace of AI. The same as the plagiarism checker.

Teaching & responsibility
[I]: Would you be comfortable teaching students how to use GenAI or guide them or at least encourage them to use it within your courses?
[R]: Sure, but then they also need to reflect on it.

[I]: Do you think it is the responsibility of the university, the lecturer or the student to decide how to work with GenAI?
[R]: The teacher, who decides what goes into the course, is the course manager. A teacher can be a separate unit. The university should enable and provide the tools for the teachers to work with.

Policy & enforcement
[I]: Does this perception differ for you depending on the type of course or assessment method?
[R]: Yeah, maybe, not sure if it matters. Do not make a distinction between bachelor's and master's courses. Yeah, I think when you steer students, for example, with the interview, it is fine to use AI. But if they have to write a report, they have to do it themselves. It is important that they still learn from it, and not copy-paste it.

[I]: Have you encountered a situation where you suspected overuse of GenAI by a student? If so, what did you do?
[R]: Yeah, so it's the case of the essays. Yeah, they just get a fail. It is not meeting the quality standards. Then they have to repair, and I tell them that I suspect AI use. That is not the key; the key is that the quality standard is not met. More of a warning that they used AI. They do the AI statement on how they used AI. Yeah, I think student integrity is definitely at stake. Let their AI statements be checked by AI; they just give prompts and copy-paste the output.

I think it is good to have an AI statement. Should be included in the making of the work. You should have declared how you have used it.

[I]: Did you already or plan to implement guidelines or rules in your course(s) to cope with students' use of GenAI?
[R]: I have to think about it this month. The courses take place in September. 22nd of June is an important date for change. Teaching one course, was AI suspicion, have to change the course. The other course moves from a written essay towards an oral examination.

[I]: If implemented already, how effective were they in the sense that you reached your intended goal?
[R]: In general, positive about AI use within the interviews. Also, the AI appendix. No clear guidelines on that yet, with regard to a template, would be an improvement.

[I]: Do you have any questions or additional side notes for me?
[R]: No, not really. I would like to know the results, what kind of tool or guidelines. Quality of education, I think we need to take a stand. We do not want to arrive in a situation where we have an assessment, students use AI, we grade with AI, who is working?

G.3 Interview 3

This interview transcript has not been included in the appendix at the request of the interviewee.

H Validation interview pre-read

This appendix provides the content that was given to the interviewees in round two in advance. Due to the questions, the idea was to read this information before the interview started to get an idea of the guidelines and how they can be applied towards a contemporary course.

H.1 Course pre-read

This document provides theoretical guidelines applied to a current Technology, Policy, and Management Bachelor course at the TU Delft. It does so through the concept of constructive alignment, addressing assessment, learning objectives, and activities. Moreover, it shows what is already up to the guideline standard, for generalisability of these examples to other courses.

Guidelines*
(1) Adopt authentic assessments
(2) Enhance AI literacy
(3) Ensure process over product
(4) Clear guidance on how to document Generative AI (GenAI) use
(5) Focus on developing Higher-order thinking skills (HOTS)
(6) Protect human agency
(7) Integrate on-site assessments
(8) Rethink foundational knowledge and skills
(9) Create a fair learning environment concerning GenAI
(10) Keep the technology close

Current course information
First-year TPM Bachelor course at the TU Delft, around 40-60 students. The course is part of the Information & Communication track of TPM; students chose to take part in this course.† It introduces data-driven analysis methods and algorithms to process data towards useful, safe, and reliable information.

Assessment
- Practical exercises in Python to assess coding data-analysis skills (20%)
- Group project, consisting of a project proposal, feedback sessions with teaching assistants, and a progress presentation (30%)
- Self-reflection exercises during the course are required for the exam, but are formative. These prepare students to be able to critically discuss the implications of AI use in societally difficult applications and for the final individual reflection (20%)
- Online on-site exam where students may use a calculator and two A4 sheets, written on both sides with at least font size ten (30%)

Learning objectives
(1) LO1: Formulating the problem and research question with multiple stakeholders
  a) applying TPM problem analysis to data science issues
  b) justifying how data science and AI functionality can contribute to solving the problem
  c) exploring a dataset to determine whether and how it can provide information for interesting questions or support decision-making
  d) mapping out and substantiating the possibilities and limitations of data science/AI in advance
(2) LO2: Operationalising a data analysis or AI system for research or decision-making:
  a) selecting the appropriate modelling approaches and algorithms
  b) determining the need and quality of data
  c) processing datasets using modern data science scripts and statistical tools
  d) building an AI model (rule-based or learning-based) for decision-making

*The guidelines are numbered for reference in this appendix, and, therefore, not necessarily in the order of the main text.

†This last sentence and addition of the number of students has been added based on insights from the first validation interview

e) summarising the key mathematical concepts for AI systems
f) applying relevant tests to evaluate a decision-making system against various criteria-
g) identifying the limitations of the AI system's output and assessing its impact on accuracy and utility

(3) LO3: Reflection – Communicating responsibly about the possibilities and limitations of a data-driven and algorithmic decision-making system, including:
a) using best practices for documenting analyses for reproducibility and accountability
b) reflecting on the ethical, legal, and societal implications of an AI system in terms of the research process and implementation
c) translating implications into possible considerations in problem formulation and system use

Learning activities

- In-class instruction and discussions
- Supplementary slides are uploaded to explain key concepts
- Practical exercises
- Group project
- Reflection exercises (personal, so no AI use)

Places for improvement

Assessment

- Change written exam A4 sheets to be handwritten (ensures critical process of determining what to include on the sheets (5)
- Change practicals to pass/fail (alleviate pressure on students so they will not use GenAI and learn fundamental coding skills (8))
- Add a coding question to the practicals where students have to solve first on their own, and then an AI solution with a reflection question (balance AI-integration and independent reasoning and AI skills (2) + keeps foundational coding skill (8) + move from understand to analyse level in Bloom's revised taxonomy (5))
- Include in reflection of project proposal to reflect on how you would integrate GenAI if you were allowed to (enhances AI literacy (2) + protection of human agency (6))

Learning objectives

Both are to adhere to constructive alignment due to the change in the practicals and the group project described above in alignment with (2).

- Add point d) to LO3: reflecting on best practices and implications for adopting GenAI within the practicals and group project
- Add LO4: a) the ability to critically reflect on AIGC, b) the proficiency to work with contemporary GenAI tools, c) a basic understanding of the influence of good prompting, and d) the ability to use GenAI pedagogically, adhering to academic integrity.

Learning activities

- To adhere to the new learning objective, in the group project, students choosing a topic can change to an in-class activity where students brainstorm with a GenAI tool to find an appropriate topic and dataset (enhancing AI literacy (2) + HOTS through evaluation of the AI-generated content (5))
- For the example above, integrate pedagogical AI, such as the Surf AI hub, EduGenAI, GPT-NL, or something the TU Delft may develop in the future (pedagogical guardrails, such as not giving immediate answers (5) + fairer learning environment through equal usage (9) + keeping the technology close (10)) Information on GenAI use in the course
- Make benefits of GenAI course-specific so students resonate more with the benefits and for guidance (e.g. brainstorming various ideas for utilising a dataset for the chosen topic or explaining meanings of coding error messages or generating additional practice questions from practice exams, (2))
- Less strict on AI use in the course, but more tangible guidance, outside of reflection exercises (enhance AI literacy (2))
- Guide students with example prompts and explanations of prompt quality (2)
- Clearer guidance on documentation of AI use with a template for how students should create their AI appendix (4) + (5)

Already up to guidelines standard

Assessment

- Group project with own selection of topic (Intrinsic motivation + highest level of Bloom's revised taxonomy 'create' (5) + authentic assessment since students can choose a real-life dataset (1))
- On-site presentation (improves social skills through human communication (6) + (7))
- Project proposal and progress presentation (ensure process over product (3) + enhances students' confidence, reducing reliance on GenAI)
- Practicals (lower and higher levels of Bloom's taxonomy (5) + foundational coding skills in Python (8))
- Reflection assessment (AI use as kick-off criterion → HOTS should be developed without AI (2) + (5) + authentic due to real-world cases (1))
- Written examination (benefits of testing students' capability to retrieve and apply knowledge on the spot, necessary for many professions (7)) Learning objectives
- Learning objective 3 (GenAI able to perform boilerplate tasks → reflection becoming indispensable (8))
- Learning objective 2g + 3b (foster AI literacy (2)) Learning activities
- Self-reflection exercises (authentic due to real-world cases (1) + reflection which develops HOTS (5) and is a foundational skill (8) + topics such as algorithmic bias improving AI literacy (2))
- In class discussions (protection of human agency through human communication (6) + authenticity because they are based on real-world cases (1))

I Transcript validation interviews

This appendix contains the transcripts of the semi-structured validation interviews. Although the interviewees are anonymous, they are presented per interviewee. This round includes four interviews. [I] refers to text being said by the interviewer, while [R] refers to text being said by the respondent.

I.1 Interview 1

Perception guidelines and course redesign
[I]: When having read the pre-notes, what are the first things that caught your eye?
[R]: Well I think first of all, I would like to learn a bit more about the context of the course. So how large is the cohort expected to be here?

[I]: It's a pretty small group, I think around 50 to 70 students.
[R]: Okay, okay. And how adversarial is this group deemed to be?

[I]: So within TPM you have a track. So this is the information and communication track. So that's more for students. Yeah, that chose the track and it's more of like the more of the technical side of TPM. So more of the communication, digital communication.
[R]: OK, so so then they at least willingly choose this course.

[I]: Yes, definitely.
[R]: Yeah, that was my first question, because some of the points that I see here, and some of the setups kind of assume good faith intents and motivation of the students, and. Yeah, well, what I see in this big common course, because of course people also didn't choose it, and there is some resistance against the topic also. So it could be a very, yeah, diverse group of responses you could get back, and there are some that will deliberately try to kind of sideline you. So then then you'd need to be a bit more restrictive probably on potential abuse, but maybe here then it is probably less risky.

[I]: I think that's definitely a valid point. Is there something also you saw, like, from the content of the guidelines? Is there something that caught your eye, like, okay, for example, this is something that's really hard to implement, or this is nice, or just some some general comments on that?
[R]: Um. Well. I'm the keep the technology close is I think an interesting one. It's, of course, also already linked, I think, in other points, so why is it a separate point? I'm curious. Um. And the fair learning environment, we just in the pre-discussion, we had a bit of a discussion on on to what extent fairness is or isn't influenceable in the situation you're in. I do think at the intervention level and maybe even at prompting levels or so. That there could be some potential waste of still handling it in some, like not fully doing an algorithmic fairness optimisation, but still have some interventions that try to create a more equal playing field, which go beyond only offering the tool to everyone. Uhm, so that's also one that I think could probably be a richer in terms of how it ultimately will be worked out and what I now see.

[I]: Yeah, so my general idea behind 9 and 10 was from a course from Harvard, CS50, introductory programming course. Where they conducted their own sort of middleware layer. So what they did is, uhm, they used Azure Open AI models and then they made their own sort of interface where they had a large system prompt with some pedagogical guidelines and some guardrails. So for example, when a user prompted personal information, it would filter that out, but also it wouldn't give immediate answer. It would just first ask, what do you think yourself? And so, it was from that idea that I thought, okay. That's nice if we could implement Npuls and like with the EduGenAI is working on that service working on an AI hub currently and we have I think GPT-NL and now as well. So I think those initiatives are really important to sort of keep up with because if we could implement such a sort of layer within education, on one hand you would increase the availability to students, because you don't have a premium model, that would just be a model that's available to all students. So from that idea was the the fairness point. And to keep the technology close was more also to sort of counter the really large harm and also sort of Silicon Valley, just growing technology base that you just want to keep it close to just sort of have something in your own hands as well, because there's a, it's a bit of a controversial topic whether these models are fair itself and in general. So when you're just using the models that are being created there, that also creates sort of a technology,

you're just as education system as a whole also giving to that sort of technology hype and harm that's going on currently.
[R]: Well, this is an interesting one because somewhere there is a privacy panel where EduGenAI is discussed. Uhm, but I know someone that was there with the people who helped negotiating that and they've quite openly also already been conceding that they would assume probably even while they will offer it to everyone, students will probably already be accustomed to using some of these other tools that they might have paid the premium fee for already. So while you do offer the tools to everyone, actually their assumption is that still, secretly or not secretly, probably the usage of these other tools on premium paid accounts will probably still just continue as is so that even in that, you just have an extra alternative and you help more people go there. But just because it's so convenient and comfortable to do the other thing, things are running anyway, will people actually make a change or not? I wonder if that's been a consideration also for you.

[I]: Oh, yes, definitely. There was a, I think it was a small article about how they did it currently in Estonia. So there are also some I think that was in secondary education. And then there were also students that said, yeah, sometimes I just use the normal version of ChatGPT because that one gives answers more promptly and quicker. So that's, I think, sort of the the whole, I don't know, thought process behind all of the guidelines, because you can have really nice guidelines and a really nice model, but if no one's going to use it. Then you still don't have anything, like nobody's going to change. So that's at least what I hope in my study that I can sort of get across is also the message of this whole sort of cognitive side as well, because I hope it sort of raises awareness witin students that why are you here in education in general, if you would just solely use GenAI? Like, why are you here then? Because the whole idea is to learn.
[R]: So then why are you here question? I think also is important. I mean, I don't see it as much in the guidelines, but I do think it's also the bigger theme that's going on. And let's say again, what I do know, and probably it's also because of the skill of this particular major we have and the selection process before, we have quite a group of very pragmatic people that just want to get super good grades in the bachelor's to jump to a US institute or so for their master's, or that just want to get their diploma. Because then they have a TU Delft, and for the rest they don't care, they just want the points that's it. So, so we have quite a few of those that actually much less seem inclined to really think apart from just scoring points. And this must have been quite hard to convince them otherwise, especially also as the skill they're facing. So that's at least been something that we've seen in the major. We say that people are there to learn, but we don't always see it as intrinsically with the students. And if that intrinsic motivation aspect should maybe also be a bit more explicit to them. Uhm. Yeah, I'm not sure if we have actual empirical data on that, but it's at least something that I've often heard colleagues being surprised about. Myself also noticing sometimes, yeah.

[I]: Yeah, I definitely think that's an important point. And that's also why I chose to, that's not within the pre-read, but before I go into the course, I also have sort of pedagogical lessons. So that goes more into like the really constructive alignment side and how do we get intrinsic motivation of students. So that's also why the authentic assessment is in there because what was seen in other studies is that when you have real world cases and you can let students, for example, choose their own topic, they're more, in general, they have more intrinsic motivation because it's actually something that interests them instead of just a generic topic that is given to them and they just go ahead and do this. Then, it feels more like going through the motions instead of actually sort of thinking about it yourself and actually. So that's sort of the layer behind all these guidelines is actually about this. How do we sort of grasp this intrinsic motivation within students that they actually want to sort of cope with GenAI themselves, because we can sort of, because within the study, this feels like a sort of top-down study. So how do we sort of enforce it from a university and especially I focus on within the course, so from the lecturer perspective, but yeah, also from the bottom, you of course have to sort of grasp something. For them to actually see the value of this.
[R]: No, and there I just wonder, like, with a guideline like protect human agency, that seems some more a value driven one, that's a bit more abstract. So not all of the guidelines are necessarily operationalised in the context of teaching a particular course. So that's just been one other thing that I think maybe it could be a bit more forefront, on like strengthen the awareness of why being in a learning situation is valuable or something like that. That's probably too many words still, but something along such lines, because that I don't see as much at this moment. Um. Like

most of them do seem a bit more production-oriented in some way. Where I do think there is an undertone of that. And maybe also the rethinking of foundational knowledge and skills, that's also an interesting one, because also on that one, in our major, I've seen a lot of trying to rely on other inventories or other perspectives to legitimise you should or shouldn't do something. So, for example, when we do curriculum reforms, we tend to look at the guidelines of professional instances. Uhm, and of course, those are slow because they adapt slowly. So this degree of reflection, I mean, there is some ethical component in some of these curricula guidelines. That's been one of the reasons why I mean, I did my course, but it's not that worked out. It's not that clear. It's not that clear if it should be core or not. I feel there is now a more urgent push for this, and maybe not just foundational knowledge and skills, but also need foundational takes on what assessment means, which we are also touching upon. And that's just been something that I collectively also in the country, in other levels of schools as well, see being noticed and at the same time full paralysis on whether we can change anything. So just also because accreditation and quality assurance is very restrictive in how it currently goes and there seems no capacity to truly do a complete system change on that. So, that's also just been one thing that intrigues me where I wonder, I think within if you have enough people that are interested in this and want to push this, of course you can pilot something. But there could probably be also a more deeper reason to really look into this, that will need external actors as well to help with that to really make it valid enough.

Guideline feasibility

[I]: Do you feel these guidelines respond adequately to the challenges and opportunities of student GenAI use?

[R]: Yes, so, so I think if that part is more fore fronted, then yes. So some of them are I said a bit more operational, and I think in line with how everyone that tries to do this. Like, for example, the guidance on documenting your use. That's something everyone does. Uhm, but yeah, surely some of them need to push towards this, you know, looking a bit broader.

[I]: How feasible would these guidelines be for you to incorporate in your courses?

[R]: So again, it very much depends on the restrictions your context gives you. So one issue with a course is that I'd like way more reflection and supervision qualitatively, and I cannot operationalize it given the skill. So really, an issue is I think a course suffers from the restrictions of its more than 400 people. There is no real cap on how many people there are, half of them don't want to follow your course and will try to resist you. And we have the traditional testing ways of assessment. So with such restrictions and the budget you have with the TAs, there is so much you can do really. And that's really been something where I've been sorry that it was so restrictive and that the apparatus in that sense really has been guiding. When in the earlier, I might have told you when we spoke first, that that's initially when we spoke about this course becoming mandatory in the new curriculum. My first proposal had been to make this a year-long reflective course. So to actually have smaller units over the year, even some extracurricular optional activities, just to allow more continuous exposure and engagements. Maybe having, of course, some more centralised point where you do for the final assessment, but you build basically over a complete year. And then by the time you end your year and start towards going towards graduation, hopefully you've seen enough to then completely pass the thing. And what would have made it way more manageable then is that you have like smaller exposure over longer time. So you can pace a bit more. How to do it, people can choose to show up or not, but then still you kind of spread it out. And initially, there was some openness to it, and then we failed the first accreditation attempt because of bureaucratic paperwork issues. It was not us, it was just a document missing. Uhm, but then when we failed that and we had to go to a resit, then of course everyone got very conservative. And at that point we were told, no, no way you are going to be just the one quarter course, traditional like all the other courses. And at least, okay, can we then be as late as possible in the year because you want to have people being exposed for longer. Okay. Even in that, the size of the course was unspecified. So initially they said it will stay same size as the masters used to be. And then also our idea to do orals was a problem because we had twice as many people, my fellow co-teacher falling out because he became a father and exam week being in the Easter week, so Good Friday fell out too. I had four days and 1 1/2 times more students and one teacher down. And that's again where at that point where you just survive your course. And, and that's been a pity. Because even then, I kind of liked the idea of speaking to my students back then. And now, then we also had to set the protocol in March for the next year. So the course was still running, when

we had to figure out how to do this year. And we did this written assignment group assignment thing, and now, of course, that's a lot of extra work for us as teachers to correct, and I cannot converse with the students to check why would you say it this way. And again, that's not optimal. And then again, with these restrictions, I don't see any other way in which I can do this. So, With the course of 50 to 70 students, you might still be able to do that, and I think that's really good, and I think you should. But if you're in this templates and you have this skill? Yeah, you save whatever you can save. That was maybe a bit of a rant on the current circumstances, but this is.

[I]: What support or policy changes from the university would guide you in dealing with GenAI use by students?
[R]: Well, so, so definitely, so, so the space to rethink those foundational knowledges and skills, and also the space to probably then change the way we do assessments, and currently those are to rather dogmatic things you have to work with. So in terms of financial knowledge and skills, we set our learning objective course every year with the study guide overall, you can kind of change things. But even in that, even the idea of constructive alignment is super bureaucratic in nature in terms of how it's being reported on and acknowledged for, right? So having the matrix fully fitting in some way goes almost above also qualitatively seeing if it actually works or not. So whatever I claim and whatever I account for makes the thing fitting. And experimentation as a consequence is a bit harder. And then again, at the scale of 400 people in your course, you cannot actually afford it anymore. So, I'm not sure even smaller skill classes you can still try it in some simpler ways. But even in that, it almost feels like you should almost also deliberately be obstruct enough in your learning objectives or so to have your matrix being there and all the assessments being fitting and not over bureaucratize it to still have the freedom to do these things where, yeah, it also feels a bit strange to have to also even work around that in some way. Then it is your only binding way of showing if you didn't accomplish whatever you were supposed to do. Also for students, also in case challenges may come up on whether they weren't or were properly assessed. And even that even the knowledge on why do we need certain knowledge in the curriculum also for me feels like. I think quite a lot of that in practise still boils down to individual assessments of teachers, really, like typically how I've seen curriculum reforms happening. Honestly, typically, it did seem that everyone wanted their course in which they hope to get a thesis student afterwards, which would be research of their interest, right? So even there, you see that that often, and I heard these complaints broader in the university, that the way curriculum is designed, yeah, tends to almost, there's a lot of influence of personal, individual, and research interests that may even go above these more foundational, broader-reaching questions. And an interesting thing about this particular one is that it is a holistic question, right? It's a holistic question, also not clearly technical. So I'm not sure if with TPM, it might more naturally fit. Let's say in EEMCS, it's also challenging that this is a topic that then is seen as almost TPM. So it's the extra soft thing you have to do, because otherwise legal people get mad at you, so, so, so even in founding what's needed here or doing it properly or showing that this is also asking for a skill set that asks beyond what we normally do, at least in my positioning, that that has been conflicting a lot with how the average colleague sees their mission and what students are supposed to teach or what we should spend curriculum hours on. So I know even some colleagues of mine have been upset with 5EC go to this course. Because those five ECTS could have been spent otherwise on more content. So that question, I wonder if there is space, because at central level, I know in our educational vision, in the strategies and how teaching can be positions, you do see momentum of teachers and strategic leaders that seem to want to push here. This idea of responsible engineering and the place more in society, the quoting of Biesta as well, is a lot in our strategy documents. And here I again wonder might TPM have a bit more of a fertile ground for this because it might be closer to some things that I think the faculty will be sensitive for. And at the same time, is there enough space to truly get foundational or will you be restricted to always having to do that within the small scope of your course? Or if you don't, is there enough equipment also to do it more broadly? Because that's the funny thing with our curriculum, that's beyond the course. It's actually also become an entering curriculum level. So also on the thesis, there needs to be some student can reflect on position of the work in society. It's been a curriculum-level thing, but then I have no clue. My colleagues, I think, are not equipped to actually assess that, so they fill checks, but I don't think it's proper. I don't have the capacity to check every single thesis, and now we do it this way. So it's bureaucratized, but I don't understand for the quality of what's actually going on there, given the degree of resistance I can see.

Implementation

[I]: If necessary: In an earlier interview, a lecturer raised scepticism about requiring an AI appendix stating information about tools used, prompts, purpose. What is your opinion on such an appendix?

[R]: I think it's the only reasonable way you can ask for the accountability. It's not optimal. At the same time, it's a way to acknowledge that people can use it, but you expect them to be accountable and transparent on that. Because, let's say, not using this you'll keep with this feeling of whether what you see or not is authentic or not. Again, this is a bit coloured by the experiences in my particular course. I unfortunately don't have the space to talk through every piece with every student. So I've heard in some institutes that, for example, for coding exercises, they've changed the assessments to really interviews that the students come with code, but they have to talk through the code. So it's not really automatically assessed any more. It doesn't really matter if it passes or not, but in the process, can you talk how you got here? Why is this constructed that way? That's how they enlighten them, the practicals. In the situation that I've been at, if you have like 400 potential reports coming your way, of course, in practise not everyone submits, but the 100 group reports we now also have, right? A hard thing is, I know when we did the homework assignments, I had my TAs, at the lecture, checking all the references. And I knew that something that capacity-wise you can't always do, but I suggest for this one, because I asked referencing, trace them all of the sample you have, see if they exist or not. And again, the hit rate was surprisingly high on non-existing citations or claims or this claim of this is in this article and it wasn't actually in the article. And it's, again, unmanageable to fully police on that in the scale of homework you get. I do think it's one of the earliest slippery slopes that we seem to tolerate because of this now in terms of if people authentically actually understand how to process information. I'm really struggling with how to do that properly. And there, I think at least then having some accountability statements on, I used it to do literature search, but this is how I then did it, gives me hopefully a few more hints provided they are authentic about it. That if they say I use it for brainstorming, okay, but if I use it to summarise the paper, that that might be a hint to then look more closely at what went on there. Um. And especially also with the prompting, it does force to hopefully make it psychologically a bit harder to instigate prompts of just giving the whole answer, because you kind of know you cannot really come up with that as a student, right? So So I think that's the best level you can get it. Then again, looking also at the project reports of this year, because we asked such an accountability statement at the end. Honestly, I think students have been way more abstract and vague than what they really were. So we asked for them showing prompts. I hardly get prompts back. I can't really have some failing on that with this scale again. All the instruction, yeah, there again, there's so much you could do, and that that's maybe where something like EduGenAI could help, because there they currently plan to have more sample environments where they can store such interactions. So, ideally, then you'd have the students kind of force working with the tool and then looking at the logs there in the end.

[I]: I think that's a really interesting one, because that was something I talked about as well. It's also like you have student integrity on one hand, but yet then of course you have to trade-off between the privacy as well. Like you can't really sort of police. And that's also something we do not want to encourage in general in education, right? But I was curious, this is of course one point of the story, but do you think whenever something happens like this, you could go to the board of examiners, right, to sort of say that there is a case. Is that something that happens a lot? That, for example, you say, okay, AI is overused? Because somebody earlier also said that when I go to the board of examiners for plagiarism, the first thing they say is: Are you really sure? That doesn't feel like sort of an environment where they sort of want to punish students, but first protect the students. Are you really, really sure? And then, but of course with AI, it's more difficult to really say, okay, this is a clear AI use. Whether with plagiarism, you can literally say, okay, this is the source and this is the student work. So do you think we could sort of improve maybe in that process to sort of change it a bit, so yeah, because I think you have to be strict, like this is a point I don't think where we could be soft, like this has to be a strict guideline to raise awareness, within students, because I know a lot of students from my own perspective and experience still use it because they know that we cannot check it from like an educational perspective. So do you have any ideas maybe into that direction?

[R]: So an issue with board of examiners is board of examiners also is very bureaucratized in a

particular role and mission and in how also the complaints or challenges go, as we've seen, if, if you. If board of examiner rules and you want to proceed against that and appeal, you go straight on to the Raad van State, right? There is this strange judicial situation where very quickly it escalates in a very legalised way. Where I've also heard from multiple boards of examiners also in other universities that the student tends to be very protected in this. So in these, unless you can make a fully hard claim that something truly has been there, there's no way that you will win that case. So many of these AI cases just are not admissible in that sense by the procedures we have now legalised, which means that maybe actually the board of examiners is not equipped to deal with this. Of course, there's been some attempts, but I've heard nearly all of them fail, even while it seemed rather clear that the student doesn't understand just because of how this procedure works. So that's again a bigger institutional systemic question, I think. Even there also one other interesting side effect in terms of burden is I've also heard in multiple schools and also from Board of Examiners and even seeing it, again, the scale and types of courses we have that when students appeal, they now also legalise it quicker because they have more tuning to make nicer cases. So we do have to, if there is a challenge, to the grading or two in assessments, go through more texts that we also have to verify more closely. Where, yeah. things just escalate in some form because of this. So that's also been something I think that the Board of Examiners has been quite burdened with. And, yeah, handling all those complaints, it's a lot of work. So, so it is something we acknowledge and that we see, and that at the same time, I feel it also costs so much energy and fighting, which we couldn't do, kind of knowing that in the end, the final committee will probably not give the right to you. But there, I'm not sure if that's the route or if there should be other routes to somehow encourage that more. And I'm with you, I'm really not wanting to police, just also because students are autonomous beings that I don't want to interfere with per se in that way. Yet, at the same time, you do want to somehow be able to. Still. have to report with them in some way, right? And finding that and finding the right means to do that. How we can properly do that? That's again a question. It was something that I hoped in the initial setup of my course being this year, of course, and sometimes doing an excursion to the Hague or something like that, right? That's at least that could be a way to more informally make that go without it being awkward in the context of learning. But that type of that type of setup again at our scale, you cannot do it, maybe at the smaller skill you could, so. Yeah, this generally also is something that I think is a big challenge. Also generally the comfort that these students give for students. I recall that, actually, I had a nice chat with someone. And he was still in his final year after he retired from executive board, still teaching in physics major in the bachelors. And at some point, he came to me terrified because I was speaking about also this intrinsic motivation question again in the talk that he invited me to, that he had this mechanics course he was giving, and now he got pushback that students told him, like, we'd rather study at home with ChatGPT than listening to you in this hall. And if you dare saying that as a first year to someone like, right? Something is off, right? And that's the thing that we also do see, that in either a refusal to come here or to rather stay at home and do your things with your chatbots. Probably, there's a social comfort in there that some students really value it, but is not so good for them down the road.

[I]: I think that's sort of sorry for interrupting, but that's sort of at least something from my perspective that's just the whole push of like the social media thing that's going on for like years on end that we're more getting into social isolation because. And that's why protect human agency, I think, is sort of one of the core normative guidelines of this work is that's really important because we're sort of getting more into a social isolation. And then I think that's also being reflected when in university people, because to be honest, I think also myself, I can be way more efficient at home studying the slides with ChatGPT definitely. But I also go here to the lectures to see other people, socialise a bit, maybe ask some questions to the lecturer. I mean and that's sort of also a social value in that sense.

[R]: But there indeed you need to also make that social value clear enough. And also there inclusion, especially to cohorts of students that on average tends to be a bit more to themselves out here, right? That's a big cliche of Delft, but I think it's true. And I had it also last year, I might told you that, that last year, originally with our course, we did one hour lecturing, one hour discussion. And then in the discussion hour, I really had students just leaving the room because they didn't want to talk to other people. I couldn't force them to stay. I did complaints on it in my education assessment, in the end of the course assessment. And there again, it's like, you're supposed to go through this and I cannot force you and I don't want to force you. But how can I encourage you,

right? I have no clue how to do that. And there, I think that goes beyond a single course. There are, I think, the ones who are more tasked for the social fabric should take that up. And I'm not sure if that's more the studentenwelzijn people, those type of people, if it's a study association or so. Um. Like, it, it's akward to do that also has a teacher of a single course, so even with a course like this also, I don't think the socialisation aspect is very easy to enforce, and at the same time. Finding more space for that also I think is very important in being an antidote against discomfort of just being efficient at home and then quickly getting there and too many incentives, of course, in the current world are pushing to being there be quick and efficient and get good grades to reach the end and find your next qualifications. Especially if the market also seems to kind of drop that in this way, so.

[I]: Yeah, it's really interesting because I've seen also a lot of, there was, I think, it was an article written from somebody at the TU Delft who did a small interview study. And he also found out that we see sort of a movement where soon also move away from spending a lot of time on their studies. So for example, there was a student that stated that, okay, I now work like part-time, work like 32 hours because I can just sort of do all my school work in two days. So, it's really difficult to make assessments where GenAI is not applicable. And I think that's also sort of a stopgap because, okay, then it may work this year, but next year it may not work. You sort of want to get there in some way that you sort of can do something where you like cannot use GenAI or at least, and that's of course also why I include in the guidance like the oral assessment and more of like the on-site assessment because that still is like a trigger for students. Okay, there's no way you can use GenAI on the test. Like, there's no way for you to do that. So you still need to be able to do this because when you do everything at home with GenAI and you're going to be there on the test, then you know what's going to happen. But then there's also the argument again that, okay, tests raise anxiety as well as presentations. And then sort of you always have a trade-off within like this big ball of normative and social values. That of course goes maybe way beyond my studies, but that's just something I find particularly interesting within this case that you always have these trade-offs whenever you're going to implement these guidelines.
[R]: So maybe also just as a very operational idea of coding, like in software engineering for very long, we've been doing pair programming. So, if you kind of enforce such a structure of a person to another person in a physical controlled environment, if you could, and this is a bachelor's course, so there you can actually have mandatory attendance still imposed. Maybe having some of these moments that two students have to work on this and actually much more look at the trace of how they interact with one another. But the issue is this only works if you can actually watch it on site. Because there, I've heard also, again, some colleagues that had an assignment like that that happened whenever students wanted and that now the last few years became the two students talk to their agents or their GenAI thing and then just mesh the things together and haven't actually spoken to each other. Um, so, so you really need to to explicitly encourage or maybe even police that that particular part. Well, there could be a way to at least work with each other. Yeah, and one issue is also that that I've seen the, how do you say that the positive people on these technology support also then say, well, but these exams are something of old times, and in reality you always will have your tools there, right? So maybe even in that, like, if we should admit it. If then again, it would be more about about the prompting or the questioning, so, so, like, what we should now then have as prompts, what we should now then try to get as answers. And then it says that in some way. But again, it opens way bigger questions on longer term, thinking of it.

[I]: Yeah, definitely more of like the value of questions. What do we really want to get across as education in general? And that's also why I really think foundational skills and knowledge is in there, because education as a whole sort of has to sort of go back into like, what is really sort of the foundation of why students are here at the first place? Because a really interesting example is, of course, the use of a calculator, right? Of course, in the 80s, when the first graphical calculators were there, there was a lot of pushback for that because, oh, we wouldn't be never able to do like 1 + 1, you see numbers everywhere in society, right? When they go to the grocery store, it's like everywhere. But then now it is integrated in a way that in your first years of elementary school, you don't have calculators at all. But then at some point you do. So of course, it's looking like way into the future, but maybe there could also be something with like GenAI, because now course, for example, my own cohort, I only used ChatGPT and text assistants from like my third year of my bachelor's, like two or three years ago maybe. But now we sort of get into a system where toddlers

are almost using it, you know. So I think that's sort of within the whole educational system also something we should look into because I definitely think it could be integrated in some way, but currently it's just a big blob and you don't have any control. But that's just something I always try to compare it with, which was also why I included sort of the history of AI in education.
[R]: That's an interesting one on the calculator, because I do think the Dutch students, in terms of their math knowledge, are internationally quite behind some others, right? And maybe because this has been so much. Embrace together with the way the assessments were also rather linguistic, for example, or rather applied to. Which has been something we've been seeing in admissions to our bachelor's programme that we feel it would be if it would be truly about about only quality of tests, then we would hardly admit any Dutch students because the Romanians are better. That's the reality. So, this is also a thing that that's we have seen in the place of it and our society does function. That is not an issue. But then still, it's in terms of the international playing field or what others do and there you see that the ones who do make our assessments easier in the admission. They've had way more, yeah, mind training in terms of theoretical foundations is stronger. They don't always need the assistance, and. Yeah, then again, in terms of curricular fit, or what you need, or how to get there. Um arguably, since we've done the selection, our throughput is better, and students are more motivated. So, in that sense, it has been convenient, but it also says something about who we would and wouldn't admit, right, who we wouldn't work for.

[I]: If you would like to implement these guidelines for your courses, what institutional steps would you need to take?
[R]: Yeah, so I think you really need to back in that you can do it in this way. I think it's very important that they should wouldn't be left on their own to figure it out in their own course and then make it fit in the bigger picture. Um, so that at least will meet the ones who are responsible for the educational management and the leadership to really also encourage and push for that. And even in that, I'm. I'm also just curious in terms of. Let's say. What intrigued me now lately in the budget cutting plans. Is, is, it's also never been super clear to us as staff. What the financial implications are at the end of running our courses? And apparently, of course, much of our stable funding comes out of the ECTS we convert and the way we've been inviting that. Let's say that me teaching a 400 course is very good for my group if these students get a their diploma on time. So I see within computer science, actually, there is a push for larger scale courses because it's convenient because it brings in more money in the end. And if that in your management becomes your incentive, of course, again, it's a bit dangerous that you risk becoming super pragmatic about this. So in such a playing field, the resistance of let's take it easier actually skill down, it's hard to convince people who are super convinced of this being a good idea. And I know that different people are different on this. I know that in our ultimate vision, we actually do want to go more in this reflective side, but to really get that going needs a lot of joined, truly wanting to go there. And even in getting them things accepted or not, I fear that. Currently also, a lot of the agreements on indeed how to shape a curriculum, how to do these things, it's much as super tentative, but then if you want anything, you need to align it with all your other colleagues, where, again, if many are not of the same mindset, you can actually not really make it aligned. Then again, I've heard that in more situations and more curricula across this campus. So, I'd really think this is highest leadership that should push this down and our leadership and then also should stand for that and push it down. More so than just waiting whatever random loose researchers ultimately decide together. Because then it becomes too much dividing it by in a certain way across your interests. And I'd also be curious there how it would look for TPM given that also financially, I think that faculty also has been struggling. So I'm not sure what incentives teachers even here on this end, like they have been cutting down such small courses across minors sometimes or so. But also in terms of staffing hours, getting the time to develop something else. Also that thinking beyond your general frame is something that often we do in our own time because we care, but not something we get scheduled time for. And then there is this teaching academy that tries to at least be a bit of a well, safe space where people can go and work together, but then again, it's extra work that you have to do together. That often, I think, if you go as a voluntary choice, you as an individual are free to do, but it much less feels like really systematically. There is to support it and really properly settle on this and follow through courses.

[I]: What would be an ideal way for implementing such guidelines, if you would do so in the first place?

[R]: If you propose the guidelines then that will happen and people will say, oh very nice, you came with these documents and 3 pick it up and the rest will leave it be. I've been lately thinking that probably. Hopefully, the alignment on this should be even as another level with places that are more open to this I think universities have generally also been in a strange bind between. Yes, they teach but mostly people are here to research, and of course they choose to also teach, but many people didn't come here to primarily teach. And that also makes things more difficult, because I think even the motivation of teachers, I think their heart is not primarily there honestly. So that also doesn't help in really pushing this forward. And I've really been lately wondering if in other types of education there might be more openness towards this, because their heart is more primarily in that role. Honestly, I had better conversations on this with primary and secondary schools and in other levels of tertiary education. So this has been very interesting that I haven't anticipated that, but I really think. Yeah, and then also there they have a bit more space. I also, lastly, was speaking to, a bestuurssecrataris, and he's been a bit of like an interim manager in many places. And he's been concerned about this, but he, let's say, within the realm of a school, he can do a lot of things and the teachers will kind of follow if they are aligned and it's easier to align them in some way. Similarly, I've been quite surprised at the energies I've seen at MBO and HBO level. So also there, I've seen a bit more openness when I also once spoke for an MBO school educational level. About like, ethics and AI, I immediately saw all the teachers that worked on citizenship that is a curricular interest at that level to see what they could pick up and use in their courses or integrate there. And that willingness just to hear this be inspired and try something. I'm not sure, maybe it's also less bureaucratized than with us and maybe it's a bit more manageable and these people again are there to really teach and work with the young people. But just that mindset that they were willing to incorporate something. Here it tends to be a lot of political alignment before you get people there. Again, because I feel that that level, the universities of applied sciences are a bit more agile in trying things out. Also, a bit better rooted of the society, because they do attract a more representative, not an elite group. But even there, I've seen in terms of what they try with education or with tooling, or they seem to have moved faster. I see also more pilots there on reflection and reflective chatbots, or so being implemented in multiple hogescholen that I haven't quite yet seen here. And again, there, I think for them, they've way more primarily seen education as their true primary process. Where here, it's mostly being used, as soon as there's a financial discussion going on, that we're reminded that it is our primary process. But normally it's not really what you are encouraged to do so. If it stays like this, I don't know. Like universities are themselves transforming and culturally trying to position themselves better. But in the history, let's say the classical university didn't really care. And I think that also is happening right now if we can really move forward and push forward. Because hat also made our systems and institutes less open to that.

[I]: Do you have any questions or additional side notes for me?
[R]: Well, maybe just a detailed thing, I like some interventions, like the handwritten exam A4 sheets. On the one hand, I like the idea and I'm pretty sure it will get a little bit of resistance. And there again, it's something where, of course, you can still ask ChatGPT and then just copy it with your hands. And still, of course, the fact you need to spend the time to write it out maybe will help a little bit. But it's a proxy for what you actually want to achieve, right? Um, yeah, and then I'm not too sure if it could be a way to do it, and especially with little time, it's probably an easy intervention. But this idea of critically determining what to include, even that's something you could you could even in homework, try, right? That's also we tried like try to keep the answer within this limit. Or if you go over have a very good reason too, but just go alert. There are ways to control for it, I think, throughout. I've been a little questioning about on how to reflect on how you would integrate GenAI if you were allowed to. In that, this does seem to frame positively towards the technology, so let's say the ones that don't want to say, I don't want to use this. They then also have to think of how they might want to admit it. And that's also been, of course, a general question of what do we do with the refusers? Because we have seen it in our grading system so far. Typically, they're worse off. It's a principle choice, but they will not be valued for that. And there again, I feel it's a bit disbalanced how that now goes. So I'm not fully sure if there should necessarily be an encouragement aspect there, or if only if you choose to use it, and that there also is sufficient acknowledgement if people choose not to use it. Because that's been a bit of a funny thing, right? Lately, if actually there are grammar mistakes and issues and the writing is awkward, that teachers might tend to be happier because they're sure that the student actually did the work themselves. Then again, arguably in terms of standards, the work will be of lower

quality in terms of then getting into the knowledge elsewhere. So how to also handle that? That's an open question. And maybe one thing with me that I think did help and inspire other people is also. Maybe with this in reflection, also focusing a lot on what shouldn't be. So, having people also think about the more negative situations about the catastrophic cases, the dystopic cases. So not only informatively only talking about nice things. Especially where you would have them some of the risk that they just tell you politically correct answers, because they think that is want you want to hear. So, giving some room to actually also tell them not to do that or to deliberately make a mistake and say, why, it's a mistake might help to take them a little out of the performance and should only be good mindset. And also, GenAI is much as optimised through that, so it is naturally a bit more of an antidote.

I.2 Interview 2

Perception guidelines and course redesign
[I]: When having read the pre-notes, what are the first things that caught your eye?
[R]: Yeah, I think generally it's interesting for me. I mean, not necessarily interesting. It's important that somebody is looking into this. Still, what I feel is the same thing that I feel with everything else that is going around this. The focus is so much about how to use GenAI in the sense of how to catch GenAI, how contain the students in the use of GenAI, how create policies for the use where containing the students is going to be more and more difficult because they have all the tools they have for free and they can try it off. So whatever you put there, they will find a loophole to go around it. So that is my biggest thing when it comes to GenAI like. Can we let it go free and then think about something else? Or what is it that you're giving the students basically to satisfy them with our policies and our tools?

[I]: Do you think that we can let it go free?
[R]: I think we have let it free. Like I mean, it doesn't matter what we build around them. You have been a student, you're still a student. You will always find a way around when you know there's a tool that makes things easy for you. The same is in everything else that comes with GenAI. Even in research, we use GenAI for writing. And then we spend a lot of time to make it look not like AI. I think the students do the same.

[I]: Yeah, they do. What I encountered in previous talks was also that we sort of went above like the policy and also more of like the normative values. And I think what it all sort of comes down to is like student motivation, because I myself am a pretty motivated student. So when I look into all these tensions, personally, I don't want to overuse it because I know the cognitive and the ethical sort of downsides of it. But I think, yeah, that's sort of the underlying mechanism. We should really try to also come up with guidelines that really create intrinsic motivation for students.
[R]: I very much have heard that what we are missing govern, teaching students how to govern their own views rather than trying to catch them. I have also talked to a lot of students, and like you, they also say that they understand that the reliance is too much and they don't want it. And but still, it's very efficient. Some of them have already started creating their own rules of how to use this. So I do this, then I do step by step my own solution, then I use the code, you know, Claude, for example, or ChatGPT to help me, but they don't have the tools. This is, I think, what we need to give them, that like understand what's happening, and still many people, students, even researchers, after a while using a GenAI tool, and they're getting better and better, forget that it's just predicting the next word. So they take it for granted, but that knowledge should really be hardcoded, and why I am getting this answer. And they need to learn how to steer GenAI to actually do the deep research, give the proper context, what does context mean? So all these things we are missing. And also, we forget that what drives the students, like it's motivation, it's emotion. It's really a lot of deep behavioural thing. And we prescribe the same thing for everyone. So I think we need to give them the right tool rather than telling them what not to do and evaluating the use of AI.

[I]: Yeah, I think that's definitely a valid point. Something I tried to, it's not really clear from the guidelines, but point 10, a little bit point 9 as well. I created those because at Harvard, they have an introductory programming course, CS50, you've may have heard of it. And then they implemented their own sort of interface/middleware layer. So what they did is you have a connection with the Azure OpenAI tool, and then they created an interface and they had like a large system prompt. So

for example, don't immediately give all answers to students first pass, you know, a little bit about, okay, what do you really want to know, and then just have a little bit more of an interaction. But also, when I, for example, put in some personal identifiable information, it filters that out. So that's currently also going on, you might be aware, within Npuls, they're working on EduGenAI and other things. So that was maybe something I thought, okay, we can at least have a tool that hopefully a lot of people are going to use. But then I talked to somebody that said, well, why they should be used.
[R]: I tried something similar in my class. It was just a study I developed an AI tutor. And what they told me is that sometimes they were really picking at the student next to them who was not using to see what ChatGPT is telling them. So if there is a tool outside the prescribed tool that is giving you what you want better, you won't stick to the prescribed tools. And I don't think any university is capable of building something Claude level or ChatGPT level. So we're going to fail. It's more like how to use the other things that exist better, I think.

[I]: Yeah, I definitely agree. There was an NOS article about this where they did a pilot in Estonia where they created their own layer on top of ChatGPT, but then still students were like, sometimes I just use the normal version because it's way faster. And because like the underlying GPT model is not changing. For example, if I use a LLama model or I use some kind of GPT, that's not changing. So still like the large corporates are making those underlying LLMs. But it's really, of course, a pedagogical interface is way more interacting, not giving entire answers and then I think that then we have another problem that students that have already used it are, yeah, in some way also thinking, okay, I use it, I get an entire answer. So when it's not the case, it's way more difficult also, like psychologically, I think.
[R]: Not all the students are equipped the same way. I think that is a bigger problem as well. They, they, some students are very good. I mean, let's think about prompting, but how they get the prompt together, but I think some students are not trained and we don't have any training anywhere in our curricula, like how to use this thing, what to ask, how to validate. What are the added options like for many of the citizens is like a search engine not really using it to its ultimate capacity. Like, I mean, think about Claude. It has so many different layers that you can optimize. Again, how to use the memory, how to give the instructions, how to define a persona that works with you. Not everybody's using them and this is frightening in the sense that we always talked about this digital gap. It's being wired in a different way. So it's not really how good you are a programmer, for example I mean, it's how good you're a prompt engineer or how good you're using the GenAI tool. And since we are not providing the education, I think it's a bit unfair on the students to be evaluated on something that they are not really trained on.

Guideline feasibility
[I]: Do you feel these guidelines respond adequately to the challenges and opportunities of student GenAI use?
[R]: I, yeah, no, I don't think so, to be honest. I feel at the university level, we kind of sort of panicking. And we want to get back to the old days, so we are divided, designing all the guidelines around our old mentality and the fear that. We realise at some point we need to change all the educational contents. And that is a heavy load and very difficult and our teachers are not also educated. It's not only educated, the students are not dedicated using GenAI ultimately, or very well, the same is with the teachers. And then they need to create new content, which respects a new word, which is very difficult. We don't know it. We try to avoid it and that is, I think, the primary reason around designing all these guidelines. We want to keep whatever we have and build things around them. I think we need to accept at some point that we need to change everything.

[I]: Yeah, it's also definitely something I've encountered in prior interviews is that this, of course, is not something you can encounter within one institution as well. Like there are social values, normative values. It's such a big thing. You can, of course, also have critique on like the whole educational system within the Netherlands that we as education are getting paid for students that graduate, but what's really like sort of the core value of education, I think we sort of have to take a step back. That's why I wanted to incorporate the, I think I put something like the, go back to the foundational knowledge and skills. I think that's a really important one, to sort of take a step back and really think like, What do we sort of want to graduate the students on? Do you have maybe any thoughts on that or anything you want to share about that guideline or? Yeah, well, sort of what do you think that currently is like really the foundational knowledge and skills within education when

Gen AI is there? Do you think like that is changing or?
[R]: It's not only about GenAI is more than like, I mean, the access to knowledge and how fast they're changing. It's like kind of somebody that give a student, ask us to memorizing. I think I say it in my classes as well, that the way they think, the way they work with these things is more important thing at the moment. First of all, thinking, problem solving. These are the things that you'll carry on no matter what tools you're using. The rest you can force it up. I understand mathematical thinking, for example, these are like playing with the things and learning how to think, building your mental structure rather than feeding information. The information can be found and the information is going to change. The validity is going to change as well. What is valid today is not valid three months from now, potentially. So focusing so much, and this is the frightening part because we don't know what that is exactly. I mean, before I knew I need to teach them calculus, linear algebra, then a bit of programming, and then like all these things it's not the same way anymore. It's like, even when you're teaching programming, it's not about the syntaxes anymore. It's about like, how you design an algorithm, how you verify the output you're getting, how you start a problem. How do you find the normal idea that you see is a double-edged sword. GenAI can create a mediocre generation and they all do the same basically copy-pasting the output of GenAI. So, it also opens the door to be much more creative than before, because you can experiment a lot, you can brainstorm a lot with your tool. This is the part that we are not doing as much. I mean, most of them are there. They just need to steer the students to just do the way that they actually start playing a lot. The computational resource is there. The discussion partner is there. Things are faster. You don't want to spend like hours finding a semicolon missing in your code. We need to use it somehow more efficiently.

[I]: How feasible would these guidelines be for you to incorporate in your courses?
[R]: Oh. There is another thing with these guidelines, that first of all, there are policies, you can decide not to use them. And, sometimes they are more difficult to be implemented rather than not being implemented, and it becomes an unsuccessful fight for something that you cannot do. To be honest, I think it should be something very, very higher level than university, not even national level. I think if there is a policy and if there's a tool being created that people need to use, it should be at, I don't know, EU level. But it needs that level of governance. Not some, but what happens now is like you find two or three enthusiastic teachers who sit together and write policies for the courses that are actually very different in nature. They just use their own experience with whatever they have to create the policies. But one course or the other are very different when it comes to use of AI. So, I don't know, I'm not such a big fan of existing policies.

[I]: What support or policy changes from the university would guide you in dealing with GenAI use by students?
[R]: For me, I think it should be EU level at least. Okay, so Claude is a bit more advanced in all of them because it takes a larger context. And if you play with it, it will come to an understanding of you with all the shared memories. So it guides you better, and if you are more specific it can use your own cues to guide you, how you learn, how not to learn. So it basically is a behavioural architecture in a sense, rather than just giving feedback. I think we should have something really equal about the project that they designed at the students. Redesign how they are learning the way they are learning using GenAI. These are the things that make me trigger. These are my initiation properties. These are my avoidance behaviour strategies. So GenAI can use you actually help you to go around and do a better learning experience. This is a frightening in some level because it actually gets a lot of data from you and profiles you in a sense. That is why you really need that higher level thing, something that is only on your computer, for example, that nobody can use. The right that you can delete your data whenever you want and nobody has access. So this is really, really a huge thing. And the worst thing is that is happening at the moment. Because we are using this free versions of the tools, they profile us already without any governance around them. I think the thing is happening, we're afraid of admitting that it is happening, so we don't do anything about it.

[I]: Yeah, I definitely agree. It's sometimes frightening as well when I think about like all the personal information, but also like all the course information. I don't think I've ever seen a student asking towards a teacher, what can I put in GenAI and what not from like a course perspective. So like all the course materials just going to be thrown into the free tools. So I think that's I think

more toward going towards at least what I was supposed to go over with like the, I think it was the second guideline, it was about the AI literacy. I think that's really important because that's more of a general perspective on really for students because that's my focus, but also of course of course managers and teachers. It's really important to understand how these models work, but really understand sort of what their implications are.
[R]: Okay, do you think you're going to stop student doing, even if you give them the literacy? I mean, dumping all the course information, for example, in the tool. Even if they have the literacy, because that is how it works. If you don't give enough information, you don't get the good output. You need to give a very detailed context.

[I]: That is true. That is definitely true. But I do think it will still, of course, then we go back into like something we discussed earlier of the intrinsic motivation, right? Really motivate students will listen to the teachers and will actually listen to what they have to say. But then I do think it will definitely help when there are sort of general guidelines on creating a persona, it may be difficult, but at least some guidance. So what I proposed in the document as well is I think I gave an example prompt of somewhere in the syllabus, just give an example. For example, what I did for the course was, okay, you have to brainstorm for a topic. So you can choose your topic on your own and then give an example prompt where you create a really well-defined persona with like all different elements that are really about. Creating a sort of brainstorming session, make that a learning activity, maybe in class, so students are going towards a more pedagogical use of GenAI, I think that is where we currently lack, is that for students, it's just we use it, whatever.
[R]: Yeah, no, I agree. We don't give them enough teaching. How to use it and what does it mean actually? What each elements that you're adding to that prompt mean for the GenAI that you're using?

[I]: I think that's definitely important. Of course, there will still be the batch of students that don't care, that don't go to class. I think that sort of the whole system is changing where like a couple of years ago, maybe 10, 20 years ago. Everybody was going to campus. The classes were like 10, maybe 30 students. You had personal feedback and now we have like sort of this really large batch, maybe some course even have 400 students.
[R]: Especially in bachelor courses, this is crazy. You don't see students in the class. Master students are still surprisingly very good they're coming to the classes. Bachelor students, yeah, I mean, you see three, four students out of like 60 students in the class. Because they don't realise what they're missing in the end.

[I]: Do you have anything to say on that? Is that something we should, I should maybe incorporate somewhere in my research or maybe in my discussion session? Because I think the fact that people are working now like 32 hours a week and going to class like maybe once in a whole quarter, I think that's an underlying issue as well.
[R]: Yes, I can tell you something about that. It's about, there's a book, it's about Solaria, I think under the sun, or something, I don't remember the name. There's a planet called Solaria in the future that the AI is so advanced, or they didn't know AI at that point, the computers are so advanced, that people are very comfortable living with their own computers or robots. And in that world, meeting people was frightening. So they were only meeting virtually. Somebody died of meeting somebody with a heart attack. So this is a very, and I feel it more and more these days, and the more powerful these tools are becoming. It seems that you found your own friend, you found your own teacher, your everything. And then we are becoming more and more isolated, especially for students if they cannot really separate. Especially younger ones who just start university. They're all fun activities outside that is not as fun in the classroom. So they use this to study and they don't know what they are missing because this collaboration, being together, building the community, a lot of brainstorming, all these things have a lot of value that you don't get at. You need human to talk to, to hammer your ideas, to make you a better person, and they don't know this. This is the most frightening part for me. Otherwise, I like AI. I think it's helping you a lot. It gives you the tools and opportunities you didn't have before. You can grow much more than before now. It actually, if everybody uses how to use a pair, it can diminish a lot of inequalities. But there is a fear that people get more isolated. I don't know what to do about that because this needs some sort of education, but the people that age don't listen that much about these sorts of talks as well. So I honestly don't know how to fix that, but that is a problem that can happen.

Implementation
[I]: If necessary: In an earlier interview, a lecturer raised scepticism about requiring an AI appendix stating information about tools used, prompts, purpose. What is your opinion on such an appendix?
[R]: I don't know because I think we are using AI to make our life easier and faster. With that, it's like what I said, that we ask AI to write something for us and then we spend hours making it sound like human. It's within that line for me, and I mean, on bottom line, I think everybody's using AI these days, and if they don't use it honestly, they won't write honestly either. I am not a big fan. I know many people do that and ask in the classes to write the use of your AI. I just don't get it because it's about being honest. And if somebody is honest, they have used it honestly. If they're not honest, they won't report it honestly.

[I]: Yeah, I think it's again two sides. You have the batch of students that say they haven't used it at all and only for like grammar, but you know they didn't use it only for grammar. But also at some point you have to integrate some form of accountability.
[R]: For me, it goes the other way around. If you build the thing that we discussed, that the students know what they are doing and how they are doing it, use it properly and realise that using it properly is much better than using it ad hoc. Then I don't need to hear what they did. Probably my formal evaluation should be in the sense that they decide however they want to learn, and I just check if they learn.

[I]: If you would like to implement these guidelines for your courses, what institutional steps would you need to take?
[R]: I know there's a lot of work going on and some guidelines come now and then, but also as recommendations, not really. I was not involved in creating any policies.

[I]: What would be an ideal way for implementing such guidelines, if you would do so in the first place?
[R]: I don't know. For me, the ideal thing would be to first make sense of things rather than understand the policy and ask people. As I said, I feel it is not inclusive. Not all the opinions, not all the courses, not. Not the real things are addressed, as long as this is the case. I think, I know, not many people are going to be willing to use the policy at all.

[I]: Do you have maybe some last insights on what you think we could do to make it more effective in the sense that we really want to create like intrinsic motivation of students? Is there anything you changed from other years or?
[R]: Not really. The thing with that specific course, for some reason is that the students are really very enthusiastic about their own projects. And I feel that they, the thing you were saying, like it's about their motivation. They really like to do it. I talk to them, I mean, even they're very open about the use of their AI. At some point I told one of the groups that, but it's very complex. Yeah, that's okay. That part Claude we'll use, as we need to build the structure around the okay. I mean, the only thing I changed was that I had a discussion with each group, if a moment during the lab, the last lab session, and I asked them to explain for me their code and the decisions that they made, and each of them should have talked. So for me that was enough. If they know why they did something, that was more than enough. I didn't care about code per se, about like why the code was architectured this way, what each block did, what else could have they done and why they chose this specific way of doing things. That's the only thing that changed. Yeah, that is the only thing actually I care about this course. They know what models they're using, they know how to refine that, they know why they're doing certain things with the data, why they do it at this step and not the later step. But that was the only thing and they knew what they did.

[I]: Do you have any questions or additional side notes for me?
-

I.3 Interview 3

Perception guidelines and course redesign
[I]: When having read the pre-notes, what are the first things that caught your eye?

[R]: Your learning objectives. I want to know what is the purpose of your course, to understand what's my position here.

[I]: And about the guidelines, did you have any things you found interesting or that you want to know more about or things that you say, okay, this is really not feasible to implement. This is not really grasping the problem we currently have. Do you have any thoughts on that.
[R]: They're all possible. All the guidelines from one to ten. Careful adoption, most importantly, in the university policy on AI, which we currently lack. And earlier you told me that, you know, different people, different faculties, different educators have slightly different policies in their courses on AI. One of the reasons is that we don't clearly have a university-wide AI policy. So people are a little bit lost on what's happening or where are they going to align with. Yeah. So I think that could answer partially that idea that you had.

[I]: Yeah, so do you think it is an appropriate approach to sort of construct these high over guidelines, but also focus on the course level to really make it more tangible for lectures, because I think currently what I at least heard from other interviewees is that when you have those high over guidelines, they're a little bit vague and they're really high over because of course you cannot make guidelines for only one class. They have to be for a university. So do you think then that it is a good approach to sort of do it at a course level, because that's more tangible for lecturers, then they can say, okay, I see that you've, for example, implemented the authentic assessment. And that means, okay, I can implement for my project maybe a real world case instead of a generic case that I've done for the past couple of years.
[R]: Yeah, but what I mean with policy is usage that is standardised throughout. Like, you know, that can, you know, how can I say this. We need to understand what is our stance on the use of AI and potentially what tools we do offer to our audiences, be that educators or students. And yeah, of course, they have to be general, and they have to be open enough to cover the entire 8 faculties in TU Delft, and then allow flexibility for each course to be specific enough. Having said that, its still possible for those policies to exist. Like, for example, what type of large language models can we have at the university? Which ones are we using? How come? What type of access do we have? Like, currently, we don't have those. We're not supposed to use certain types of AI. Nobody knows which ones, or which ones can we. The ones that we can, like for example, Microsoft 365 Copilot, it's a little bit limited in the amount of tokens that we can use. So again, you know, deep collaboration, deep research is limited. So I'm not saying, you know, when we talk about policies, it's not about course specific ideas. It is about what do we know about what does it mean to be an engineer of the future? Yeah, what does it mean to be an educator of the future? How do we treat AI in education so students match the demand of the labour market. Yeah, how is the labour market treating AI and how is the labour market expecting students to deal with AI? These are things that policies can already tackle. I believe in things that I also believe they will eventually tackle it when they come out. Because it's also very complex, you know, as policies, but yeah.

[I]: Is that also the way it's currently being taught towards lecturers? Because I think it's a really interesting point you make about the labour market, because of course, also within my current job, I can use AI to whatever extent I want. But for me, it's always, okay, that's a private market, and we are in the public sector here within education. So, what is your opinion on that?
[R]: Well, my question to you is, why are you doing a bachelor degree or a master's degree?

[I]: To be able to, I guess, extend my knowledge, but the higher order thinking skills, like the problem solving, you know, skills in that way, because sort of the facts, it's more about raising awareness rather than really being able to recognise and, you know, understand all little details and just know from like five years ago what the small detail was in the book. It's more about being aware and then being able to apply it and evaluate on it.
[R]: That's why you are doing a bachelor and a master's.

[I]: Yep, that's why I'm here.
[R]: Okay, okay. So what do you plan to do afterwards? How do you plan to apply those skills? What is your actionable plan once you have all these skills? What do you do with it?

[I]: Well, I actually apply them also towards my generative AI usage. So for example, when

I know I have the knowledge of the architectures of the models, so then I can apply that within prompting techniques. So I know how to ask the model for certain information because I know how the architecture behind the models work. But also the reflective skills are really coming in handy because when you have to reflect on the output, those higher order thinking skills are really important, again, to being able to critically evaluate on that.
[R]: Yeah, that's true. But where I'm going is that you eventually, I mean, how can I say this? What you're saying is true, you know, you're doing this because you want to learn certain things, you have to put them in practice, but you eventually need a job, unless you're rich. You get me? So, of course, you know, you need critical thinking, you need analytical thinking, you need creativity, you need other skills like collaboration, communication, resilience, productivity. All these things are fantastic. But eventually you need to apply them on a real world scenario where you can tackle societal problems or environmental problems. And you want to, because it all other things that you're interested in, but the point is that, as you mentioned, you need those first applied. You need those higher-order thinking skills and the question is, how do we manage to acquire those skills, but students have skills, analysis, evaluation, creativity, but that they're still able to use AI tools the way the labour market demands. Now, I'm not saying that we should prefer students only to meet the labour market demands, because that would be all very, you know, reductionist. But I'm saying that it's an important thing to consider as we educate people. Because what I say, when we make surveys, 99% of students think they do a bachelor, get a certain place in society to get a job, to have some actionable careers, to have some place in society to make their lifestyle and quality of life better. So it's important that you possess those skills that match that. Now, AI has somehow shifted the skills you are necessary to be successful in the labour market. And as you say, it is very important to have a creativity, to have analysis. And that is important to use AI, like your own evaluation skills. You always need to verify. That's actually most of the things that happen in the labour market. We're using it as a steward, as an inspector. But again, to judge those outputs, you need analytical related critical thought. But you also need other skills, that responsibility. You know, are you responsible enough to say, yes, I agree, with this output. No, I don't agree with this output. Because one thing is to have analytical skills and understand where the AI is telling you crap. But another thing is to take action in a correct way. So I believe that also other skills that go concomitantly to AI and a need to also be developed. So it's not only about developing a tool. You get me, the hard skills and how the AI work. Like, like you did, that is very nice. You want to know how the algorithm goes, how much energy it's using, it's fantastic and everybody should know this. I know the ethical uses and where the information comes from and data and all this. But another question is, who are you becoming as an engineer? Yeah, because now you can do many things with AI, but should you? Can you, you know, what are your ethical perspectives, your perspectives? Those are the kind of skills that I find even more important than just using the tool per se.

Guideline feasibility
[I]: Do you feel these guidelines respond adequately to the ethical and sustainability challenges, but also the opportunities of student GenAI use?
[R]: Well, like clear guidance on how to document environment, sometimes only a policy is necessary because. Again, you know, if we want to build the engineer of the future is also a AI literature engineer. But the question is, okay, I can, but what are the platforms, what are the tools that we are using for that? Because what I'm currently seeing now is that since we don't have a standardised tool, standardised technology that we can all use. I see that many people use their own tools or private tools. You are using your own private tool. I am using the one that the university provides, but sometimes maybe also a private tool. So, it is becoming, which tool do you have for the work you are currently doing? Which is exactly the problem that we aim to prevent by a unified policy, which is to have everybody with the same AI tool, for example, your output as a student doesn't depend on how much can you pay a month. Okay. So, this is also very important. If I can pay 22 euros a month, okay, then I have a pro system. Maybe if I can't, or don't want to, then the same analytical quality. So, having a standardised policy, would also standardise certain usage. But then on the other hand. You know, we all have books. Do you buy your book or not? These are the things that we know. I mean, we need to make a decision. These are the things that are important or were important for my opinion, in 2024, but in 2026, we need to make it. We have to make it actionable. You know, what's happening, which tools are we using. But again, not for the tool per se, because you already know how AI works. You know, probably even better than me already in that, but again, the really a question is, what are those soft skills that we're building around the

technology? I don't want to bring this so much, but again, to the part of the engineer that you have become. What's your ethical confidence? What is it that you are going to do with the power that you have at your fingertips now? Because maybe two years from now, we will all have access to models like Claude Mythos. I mean, in two years, we all have that. We literally have the power on our fingertips. So then you see the question is not about, I mean, analytical, critical thinking, have always been important. But the real question is, what are you doing with those skills? How are you putting them in action? Putting that into a societal benefit, environmental benefit, or.

[I]: Do you feel these guidelines apply to students with different levels of intrinsic motivation?
[R]: Yeah, authentic assessments, that's important. You know, we need assessments that, I don't know what you mean by authentic, but what I mean by authentic is that they're able to assess the skills that we promised our students. That's what I understand by authentic.

[I]: Yeah, authentic is about real-world assessment, so really applying your skills towards real-world cases.
[R]: Okay, so about the relevance. Oh, yeah, that's super important. Indeed, you know, it has always been important that students learn the skills that are readily acceptable in the labour market. There's been throughout the year some complaints about the labour market. Some students don't have the skills that are necessary to meet that. With AI, that might be exacerbated. So authentic assessments are very important. Enhance AI literacy, I don't think we need to enhance it. I think TU Delft, in that sense, might be doing quite a good job in AI literacy. I don't see that students are constrained in the use of AI and neither are educators, but we do need a university policy that guides us and allows us to know, that we are doing this equivalent TU Delft needs. And should process over products? Yes, that is very important. Why? Because now, again, AI is a tool, it is a calculator of evaluation, analysis, creativity. So, you know that the end result becomes every time you know. I don't want to say necessarily less important, but you know, co-creation becomes more accessible. But then the question is, how did that co-creation process happen? You know, how are you inspecting the problems? How are you analysing what the AI is telling you, and so that processing becomes crucial, so somehow we need to move onto something. Putting a little bit more attention on the process, and not only the results, the process, or actually the results will do a reflection of the process, which has always been now more, because now AI can produce your report in three seconds. So, yeah, the question is. I get you. Basically three things, you know. How did you figure out that it is correct, what the AI is telling you? Second, how did you verify that? And third, how did you act upon those verification processes?

[I]: For me personally, I always do cross-validation. So when it's an important thing, I always Google it myself afterwards. Or at least find another source and do a cross-referencing or cross-reference it with my sort of own knowledge.
[R]: So, that's the verification process. The first one is that feeling, as you say, that analytical, critical thinking that makes you that truth seeker, that's like, sometimes you normally, but then there's like a truth detector pass it that you think, and you think I need to double check that stuff. That's the second process, what you're saying, the verification. And then there's a third process that makes you the steward, and that's the decision that you make in that, whole trap detector. That's the most important thing, because the previous one you're just being an inspector. You're checking what the AI is saying but in the third one, you have to make a decision on what the AI provided you. You know, to give you, a lame example, when the AI says this bridge is ready to be built, what are you going to do? That's what I'm doing. So you verify that everything was okay, but are you stamping that and saying, yes, me, verify this problem and you'll allow the bridge to be built. And that requires responsibility that requires other things. So, all these are the about AI literacy.

[I]: But within a course, do you think that unmotivated students would apply the sort of policy that is set by a course manager? Because of course, about authentic assessment is more of like, that's either there or it's not there. Like you do the project or you don't do the project. But with some of these guidelines, it's also about, for example, documenting AI use. Then, it's really about trusting the students. Our students honest. So that's sort of a dimension behind all these guidelines. It's also about, okay, motivated students, of course, they will listen to policy that the course manager says because they want to learn, they're enthusiastic, and they know why they're there. But you also

have the students that don't go to class any more. Work 32 hours a week, and. Do you have an opinion on how to sort of tackle that group with guidelines or with policy such as this?
[R]: Same guidelines, I mean, we create education for educational systems that benefit students, you know, that we are able to create students that are reflective and analytical, responsible, proactive. And, of course, we create education with the respect to constructive alignment. Authentic assessment, as you call it. And within that, we use principles of effective teaching, you know, and that's why we also need to AI, we need to tell them what they need to do. And that's why, again, you know, assessments or projects that tackle real-world scenarios. But beyond that, also the students are free to decide how much they are motivated or not. It's up to them to decide how much they want to build those skills. So the university is responsible for designing the environment that allow for students to foster those skills purposefully. That this student is responsible for taking advantage of those environments.

[I]: The reason why I'm asking this as well is because what other teachers see within their courses is that they clearly see separation between students that use AI a lot and that don't, and they also see that reflected in their grades. And that's, of course, which makes it a little bit more difficult because it's a little bit unfair when you're on the side where you want to use it responsibly, but you get a lower grade with putting in like two or three weeks of work compared to somebody who does it within a day, changes it a bit, and also passes the course with a higher grade.
[R]: Then that's the problem with what I'm talking about. It's the problem with constructive alignment. Then are we assessing the skill that we want to assess in the course? So, let me give you an example. If my aim is for students build analytical thinking then my teaching should be around analytical thinking with or without the use of AI. The assessment should be about analytical thinking, and I should test that skill. Now, the course is about a skill, then you should assess around skills. So, I mean, this might sound very simplistic, but the point is, again, are we assessing what we are promising the students. If, for example, I'm saying that I'm going to assess analytical thinking and the student that uses AI gets a higher grade just cause, then that means that there is a problem. That means that there's a constructive alignment, I'm not really assessing what I promised the students that I'm assessing. So you also need to understand what is it that you are promising students. It's like, for example, if I tell you by the end of this course, we're going to do, I don't know, change the wheels of a car. And then at the end I say, yeah, but you know, you use the blue shirts, and I really don't like that. I only like the ones that use red shirts. That is unfair because, you know, the course is not about how we dress. This course is not about, you know the course is about changing a wheel. It's about analytical thinking. And that analytical thinking should happen if you use AI or if not. And that's where education becomes a little bit tricky, because many lectures don't recognise this idea of constructive alignment and what they are assessing. Many lectures also don't recognise that the old school report that we used to do before 2022, before ChatGPT. And I mean, if you figure out the report was universal true in education, whatever it is, if you want to report. Then, that is the problem. That report is still being used to assess critical, creative, and analytical thinking in a way that it was in the same way that pre-AI was being done, yeah that needs to change, that is not an authentic assessment any more. If during 2022, you did the report, forgetting that you maybe paid another person to do it, that could happen. Let's imagine that you were honest, just to bring the honesty of students. You know, because without AI doesn't mean that everything is honest. So you have to be. But the point is that pre-AI, you would do that report, and because of that report, you would be able to demonstrate your analytical skills, how you were able to deal with data, how you were able to construct a graphic, and how you were able to do things like that. Now, all that is not possible with a report. So again, if a lecturer is assessing analytical skills with that same report, and there are no clear guidelines on AI. That is another method that does not work any more. You get it. So if I really want to assess analytical thinking, maybe I have to bring it to the classroom and go, oh, exam, to really observe the analytical skill things. Or you have to create spaces within the report where I see the three things that I told you. How is your truth detector seeking? Second, how are you verifying that information showing me? And 3rd, how are you acting based on that verification? All that has to be logged as part of the process for me to understand that you did analytical thinking. So, you see, I can still asses the same skills, but I have to change how to that. So the authentic assessment has to change.

[I]: How feasible would these guidelines be for you to incorporate in your courses?

[R]: I mean, your guidelines currently, how can I say this? There are, and don't get me wrong with what I say, but there are ideas. You know, guidelines not only tell me create fair learning, but tell me how I could create fair learning. So, if, for example, you give me this 10. Probably most lecturers will tell you, we know this. The real question is, how do we translate, create fair learning into action? The question is, how do we, how do adopting authentic assessments becomes a reality in my course? I already give you the example about the report, but not everybody understands these nuances. So you can give these guidelines, but one thing is to learn, give people a guideline or things to tell them how they could potentially operate this in their classrooms.

[I]: Yes, so that was my idea behind sort of the prototype course and the sort of rules I listed below to make it more tangible to really link them back towards the guidelines, for example, within the course. This course I've given, I applied, for example, the process over products. Because there was a group project and then we already had a lot of feedback moments. So then I also say, okay, you see that this guideline is already incorporated in this course because there's already a lot of feedback and we already see that there's a project proposal, there's a progress presentation. So that is sort of the more of the support from that side that I sort of try to apply the guidelines towards an already existing course and show, okay, this is already in line with the guidelines. Okay, here, maybe we can improve. For example, AI literacy is really something that was not really part of the course. So that was, for example, something I tried to improve in that sense. So that was sort of my idea. But of course, I'm restricted within the timeline of my thesis. But ideally, I want to make like a whole support framework, but I think that's more of a future research. But that's sort of how I try to approach it.
[R]: Like for example, number 5, focus on developing higher order think and skill. I think the University does good work on that already. Well, I think engineers in general graduate with very powerful analytical, let's call them higher order thinking skill. You could argue that although many of bachelor courses are, used as improvement, but it's part of building that toolkit of what the engineer is. The question is, how do you translate, you know, those higher order thinking skills? Because, again, you know, engineers are great at analysing, they are great at creating, are great at critising and professions continue developing those same skills despite the use of AI, well in spite the use of AI. And again, we can, but it requires a small tweak in the way we teach and we assess.

[I]: What support or policy changes from the university would guide you in dealing with GenAI use by students?
[R]: I'm going to limit myself to large language models, yeah, because you know AI.

[I]: Yeah, I'm talking about generative, so then I'm talking about ChatGPT, but also coding assistance like Copilot.
[R]: Perfect, so we want to have a clear usage on which tool do we use within the university? We have at the emails, and we currently use Microsoft e-mail. You have your own emails, and we currently have copilot, but it's not. Like the pro version, you know, you run out of tokens, you know, you can use it partially, you can really not. So I would want to have something that, again, everybody benefits, you know, like a standardised tool where, irrespective of your sex, race, or income, we have the same opportunities to use the same tool rather than you, because you can pay 20 bucks a month, have a better tool than me. So I think that's the first thing that you have to standardise. You know, it doesn't matter if it touch ChatGPT or whatever it is, or our own. But you know what, you get me? I mean, you as a student needs an environment where you can practise analysis using AI, which you have got. What platform is that? We don't know. So I think that's the most important thing.

[I]: They're really interesting. Always, in essence, how are we sort of going to enforce that, because we, of course, cannot control students.
[R]: So, and I understand that you know students are also free to use whatever is it that, but at least if I say what we have is, just to give you an example, Microsoft 360 and we're going to have a full license, at least I know, when I can do that for data privacy, at least I know that you can use it for my own research project when I'm not going to run out of tokens. There is nothing more annoying that suddenly the window says, you run out of tokens. I mean, come on. You get me, and most importantly, it gives you the equality, equality for all the students. You know, the output is not going to be dependent on how much they pay monthly. So, at this point, that for me is critical

that we have a standalone AI tool. Yeah, what we currently have is something analogous to my teacher at Brightspace, and otherwise, just imagine it. You need a standardised platform.

[I]: Yeah, I definitely agree. I also incorporated it, and I think it was a guideline 9. About the inequality, yeah, provide a standardised tool where Npulse is working on EduGenAI, Surf is working on an AI hub, because then you don't have that inequity of students not even having access to a tool, not being able to pay, and that is reflected in the grades, et cetera, et cetera, et cetera.
[R]: It doesn't really matter what the tool is, you know, we could have our own private tool, but something that is also reflective of that we could deny building on, so students understand how AI is used in professional scenarios as well. Sorry, that for me is critical. Let me give you another one that is important. Lecturers also need to develop their AI literacy skills. If lectures don't develop their AI literacy skills, and I'm not saying that everybody should teach AI literacy, don't get me wrong. I know lectures should be free. But being professional nowadays also involves updating yourself, which is professional development, and I think it's important that lectures also to have a little bit of AI literacy, so the universe should also offer training opportunities or upskilling opportunities for lecturers. Sometimes we forget about it. Or are they supposed to do fair assessments? They're supposed to do all these guidelines? Yeah. If you don't really know how AI works or how it's used in professional settings.

Implementation
[I]: If you would like to implement these guidelines for your courses, what institutional steps would you need to take?
[R]: Not all of them.

[I]: What would be an ideal way for implementing such guidelines, if you would do so in the first place? In respective on how AI changes. Compare what we had in 2022 and 2023. If you think about it the purpose becomes the same. Skills of inspection and stewardship remain. I think we can create guidelines that are flexible to accommodate for the future. What are the skills of the engineer of the future. How are we updating education to meet those demands. Economy hinges on trust. That why skills like communication and elaboration is important. Soft skills are becoming more important.
[R]: I think, you know, respectful of how AI changes. The AI, you know, and the AI and large language models we had in 2022 or 2023, and they are vastly different, and they have grown in capabilities. But the way we use them and their potential remains the same. Think about it. Even if you were to think of a GenAI five years from now or whatever it is, their skills of inspection or skills of stewardship remain. So I think we can still create guidelines that are fast enough and wide enough to create a specific membrane or a framework that really allow you to operate within the flexibility to accommodate the future. Like you, for example, in 2022, I say, well, we need guidelines in which students are going to do one, two, three. And AI literacy, they will do this with AI. They will probably do the same today. You get me. So I don't know. Yeah. I think the real question is, what are the skills of the engineer of the future? How do we update our education to meet those demands? Because there's a lot of change happening in education now because of AI. But because of that, there's also a lot of changes elsewhere that ideal learning tackle on the same degree of urgency. Again, the idea of productivity, the idea of resilience. You know, for example, skills and resilience probably know, yeah, but it's very important. We don't build resilience with plans or systems. We build resilience while investing in people, learning, and trust. That's how economy, perhaps at a certain point, it hinges on trust, it hinges on how information flows through the e-mail. That's why it's like communication and collaboration are so important. It's not only about how you know, siloed as an engineer, and how you're able to put that into a network that allows the flow of information with other people to create a product or create a service. If those skills are not being prompted in students, they're also not getting the education that will prepare them for that analytical thinking, creative thinking. How are you supposed to be analytical enough if you cannot consider contexts, collaborations, communications with others? So again, you know, the analytical skill not only in the context of you as an independent person, but in the context of where we are operating. It is in the team, is the team operating for what purpose, for what values, for the environment, the society? It's like a philosophical thing, yeah? All those still are necessary. They even become more and more relevant in the age of AI, because, again, AI can know do many things that the old school engineer could, so these soft. I do not like to use the word soft, so we

understand becoming even more important today than that they were in 2019.

[I]: Do you have any questions or additional side notes for me?
[R]: Let me check your guidelines for a last time.

[I]: Yeah, so I think what you really tried to say within the previous points is really, I think it was guideline 8 about rethinking foundational knowledge and skills. I think that goes back towards like, what is it we want to teach an engineer for the future?
[R]: Exactly. How are we making that a reality in the classroom? Because for everybody it is clear that there is analysis to the game. How do we teach that, how do we assess that in a way that is fair, valid, reliable, despite the use of AI? And I think that can happen in respect if you use AI or not. OK. Like, training. You know, here at the university, I do not know if you know, there's the university teaching qualification. You know, that you have a PhD and post-doc doesn't mean that you necessarily are a great teacher. So lecturers have to go through this university teaching qualification processes. So they upscale teaching. Right, and that's necessary. Yeah, I think there should be a new idea on how will we prep, you know, AI in education. It needs to be new modules in which lecturers are able to tackle this, because we cannot expect that now lecturer suddenly, you know, how to tackle this by themselves. So, we also need to provide scaffolds that support their teaching.

I.4 Interview 4

Perception guidelines and course redesign
[I]: When having read the pre-notes, what are the first things that caught your eye?
[R]: I think the guidelines itself, the first part, with the I would say tips and tricks maybe, they were very clear. I think they are, what I noticed is some of them are very large. As in, it entails a lot more than course design. For instance, enhancing AI literacy. Of course, an institutional or even maybe a national development that we should all take into account. And I really thought that, you know, I think what I did, I think the overall conclusion of these guidelines looked very solid. And I think also what I see in practice, things that we come up when discussing with lectures like, what should I do about my course design or about my assessment or learning objectives? I was only wondering how the, if there was a reason for the order.

[I]: Not necessarily.
[R]: So, for instance, I think with the Uh, like rethinking foundational knowledge and skills, and that could come before adopting authentic assessments, as in more like we have to go back to the basics. I think that's one of the overall conclusions I think you make in your research. So I think having, yeah, because as a person who saw this the first time, I think it will be helpful if like you read this as like these steps that you take. So having like the the one of the first major things to do would be helpful. Yeah.

Guideline feasibility
[I]: Do you feel these guidelines respond adequately to the ethical and sustainability challenges, but also the opportunities of student GenAI use?
[R]: Hmm. I find it difficult to say also because like for me, it's such a short period like that GenAI is like used as such a scale. So really seeing the effects long term, I think, is something that we all have to wait for, I guess. But I think we can put a lot of things in place that could help our students continue to think critically, because I think that's explained as I think one of the major issues of lectures see, they're afraid that their critical ability, which is for a lot of people, the essence of the university programme is going to waste. So I think having these principles perhaps, right, that the assessment is also takes that critical note into account or even like, right, making sure that students continue to be accountable for what they hand in. Those could be small steps. And I think I read this in the guidelines, but you know them better than I do. So I think in short, the effect is something that we have to see in the coming years, but I think for now we can do small tricks to ensure.

[I]: Do you think guidelines are an appropriate way of doing that? So really about the way I proposed it, having guidelines on a university-wide level, but then also focus on the course level,

because that's of course where they need to be integrated.
[R]: Sure, definitely. Yes, because what I now see is that we don't have, if I look at TU Delft, we don't have vision, we don't have central guidelines. And what I see now is that we don't have a common language or common ground that we can discuss course design or even the impact of AI in our courses. For instance, every session on AI was just completely different and I, but I think people didn't leave with what was supposed to aspire them because in these sessions we would find ourselves discussing even do we want to have AI? Should we take it into account? And we still have some discussions about those still those very important, but also fundamental parts, then we're not progressing into implementing these guidelines. So having a clear framework, I think everybody is very much in need of that because if we have, let's say, we just have the document where it says, okay, TU Delft things, AI, blah blah blah. Let's say we embrace AI. That already gives room for people to think, okay. AI is going to stay, I have something to do about it, I guess. What? That's the next question. Having a vision and perhaps a policy, or whatever you want to call it before is vital, and then having guidelines to make it more tangible for lecturers what that actually means.

[I]: Yeah, I definitely agree. This was also something I've encountered in other interviews and as well in my study is that there's also no centralised tool. So we're just saying, okay, AI is there. Some people try to enforce it, others don't, but which tool may I use? May I use Copilot? May I use ChatGPT? May I use Perplexity, Claude? What am I supposed to use? So, I definitely think that's sort of a step behind all these guidelines, but that's maybe you're aware of that within Npuls they're working on EduGenAI, but also within Surf on the Surf AI Hub, so that could be a step forward, but that's of course looking into the future, and hopefully I can sort of end on within my sort of recommendations.
[R]: And I think it also could like using that also goes back to that fundament of having a vision. You know, Amsterdam started with the University of Amsterdam started with the UvA chat, which was the first thing before Npuls took over. But they had a clear vision. Lik in the beginning, they said, we're not going to, we're going to ban AI. Nobody's going to use AI when ChatGPT first came out. But then they quickly changed and said, okay, we allow AI to be used, but we want it in a controlled environment for like all these reasons that, you know, you also named sustainability and data protection and all that. So there is a vision behind that, right, to have that implemented. So I'm very curious to see what's, yeah, what will be it for us. Because the same with children, I guess, when you parent or have a classroom, you need boundaries. You need some set of rules and within the boundaries, kids or people can do whatever they, you know feel like. But we don't have these boundaries now and people are just doing things and they find themselves lost. So yeah, having it. I think this would be a very good first step to towards that.

[I]: Do you feel these guidelines apply to students with different levels of intrinsic motivation?
[R]: Yeah, that's a tricky one, eh? Yeah, because I think apart from motivation, I think what I see now with students is that there are also so many high stakes, especially costs of living. And I think for our international students, they want their degree fast, but I need it fast, because for x amount of reasons. So, having AI like could also be for some raising confidence. I have something, English is not my native language, but I have something that could help me with that, or like I have something that could actually speed up my study, be more effective and allow me to pass it so I can achieve my goals. Um. In terms of motivation. I think vital to everything is that we, I think it's an opportunity for us to rethink what does education mean and what do we want to teach our students. And, maybe a little bit too idealistic, I think, but once we have that covered well and we have the, we know the benefit and we are able to communicate the benefit or the reason for a particular course or particular skill that they're working on, I don't know, then I hope that all students feel more intrinsically motivated to do that. And for some courses, of course, that is easier maybe said than done. Like we also have a mandatory course, of course. But apart from motivation, we could also think about, okay, what are critical skills that we really want our students to be able to showcase independently? Let's go for a supervised assessment there. So we know that everybody's got it. And for maybe the things that build up over time, we could take a portfolio or like putting, having students really the responsibility and accountability of their own growth, and making sure that it's visible. So I think there are ways to keep the motivation going, even though there will be a take home example or an assignment.

[I]: Yeah, I definitely agree on that because I think that's also something I've included into the

guidelines as well as what I try to accomplish is really, you know, what are sort of like really the rethink of the foundational skills and knowledge. I think it's really important because that sort of brings us back towards the question like what do we want within education? And then I propose also using then oral assessment and using an on-site assessment to be able to assess that, because then you can really assess as a teacher that somebody developed that skill, but then you of course have like the logistical problems, the resource problems, like those of course also arise, so that makes it like a really broad and difficult problem.
[R]: Definitely, yeah, and I think that that comes. Very important to take those resources into account, definitely, because what I think the oral suggestion is great, and I think what I see is that lectures from the large classrooms, let's say 900 students from Werktuigbouwkunde, are looking for other ways how to implement this effectively, but the budget cuts are also like something right? We we're saying with AI, focus like we want to focus on the process, so I think that will that will be more important, I think. But then there is this fragmented reality of universities having less money, like TAs are being cut. So if we really say, let's say in our vision, it's my idea. AI is here to stay. That means that we have to change our courses and then we just have to change our assessments. We need to focus more on process. We have to enable that also within our resources, within our teaching staff to establish that firmly, because otherwise it would just be a paper reality.

[I]: How feasible would these guidelines be for you to incorporate in your courses?
[R]: Resources and, apart from like the institutional vision, I think having that as a as a team. So, within the programme of a program, there are they are called differently everywhere. But within a programme we have dedicated lectures within the courses. So I think they all have to feel the same vision or have the same vision on a. We have, like, then it's not really okay, we have to do this, but they see the added value or they know that this is the way to go, otherwise it will still be like most people in person, and I think that's a very tricky part. Maybe you've also spoken with, yeah you've spoken with so many professor, so everybody has their own personal opinion about this, their own personal fears. Big tech or environment or just they think this is just a hype and we should just wait it over until the whole AI boom collapses. But, like, we. Yeah, you have to be on board. You have to have the same ideas, because otherwise it will not work.

[I]: And I think that's a difficult part of this, what I heard from other course managers, like, we are within a very bureaucratised university. People are here not to teach. They are mostly there for their own research. So there's so many different factors that make it so complicated to do this because even within current course adoption and changes, you already see that course managers struggle because even their own fellow colleague within the course doesn't want to change a certain thing. So imagine having that on a higher level with such a controversial topic like AI.
[R]: That's huge. It's so. I'm happy I, I'm not like leading some sort of program. What is it is, it's a challenge for sure, for sure.

[I]: But I definitely think that goes back into the vision, because when you sort of can get together on the same vision, that's sort of the basis, I think, to sort of.
[R]: And you can make agreements and you can also honestly say, but this may sound harsh, but this also, I think, maybe goes for students who want to study here. Okay, this is our stance on AI. Are you on board or not? Then maybe this is not the place to work or study here. I think it will come to that point, because if we really want to be effective, and we still want to be effective in achieving our learning objectives in an authentic manner, then we have to be, we can't do it half-half.

[I]: What support or policy changes from the university would guide you in dealing with GenAI use by students?
[R]: So the first thing. I think going back to the learning objectives and rethinking those, because they, you have probably read about constructive alignment, right? So learning objectives are the base of it. So if we, if we're able to reconsider learning objectives, then from that I think we can think about, okay, how do we want to assess this and how do we, how do we. Should we teach this? And not just in like, but also like how can, how do we want to integrate AI in your learning objectives? That's I think maybe a second question. If they're with AI, there are certain tasks within the learning objectives now that students may be able to delegate to AI sort of writing, brainstorming. But for industry, we also have to prepare them for a life with AI. So that's a second step. How comes AI into playing with the learning objectives, not just as a tool, but really as an

end-product.

[I]: Yeah, I definitely agree on that sense because of course, how do we want to use AI? Like how's the labour market using AI? But also, yeah, for example, when a course focuses on improving your English, then maybe you don't want to use Grammarly to write something for you. But when I want to make a report on a software project, then, whoever cares. So I think that's sort of like constructive alignment is sort of the overarching guideline towards like everything because I'm focusing on the course level and on the course level I think constructive alignment should always be the basis. But also that's very bureaucratised in some way. So, you always come back towards sort of a, as you say, sort of boundaries or at least like a sort of strict way of implementing it in some way. So, I wonder whether we can sort of still grasp or gather sort of some sort of flexibility in that sense, because it's always within the universe is always so standardized, like we need to have a standardised tool, we need to have standardised guidelines. And then you still want that flexibility for course management.
[R]: And creativity as well. Yeah, definitely. And I think from for course, for I think the impact of AI is also different per faculty. And I think in like IDE or architecture, so within their studios and their design, like they're very much focused on students going through a certain learning or thinking process, right? So that is maybe different than computer science. Though, of course, was coming. Well, yeah. So that might differ. So we could have, I think, it would be good to have one central policy, but still the faculties also need to develop one for themselves concerning their field and their domain. I think we've now done that, with quite similarly, with what we call an assessment framework. So I think, what was it, prior to, yeah, it was maybe in COVID, around COVID. We had all these, like the board of examiners in the faculties had like certain regulations, but there wasn't a central policy on what does assessment look like at TU Delft, and can we make an agreement on what good, valid and fair assessment is? For instance, can be the bonus points or what are like exemptions? Like all that nitty gritty stuff. People just did it. And I think around COVID, it was felt that there was a pressure to write things down, to have a central policy on assessment that everybody could agree with. And then the assignment was given to faculty, saying, Okay, so this is our central part. It's very general. But now for your faculty, think about the impact and the implications for the programmes and write your own vision as like an appendix to our central policy. And perhaps that will be a way around to, with that central policy, still the faculties would also need to think on their level.

Implementation
[I]: If necessary: In an earlier interview, a lecturer raised scepticism about requiring an AI appendix stating information about tools used, prompts, purpose. What is your opinion on such an appendix?
[R]: It's a tricky one. I think it goes back to the goal or the objective of your of your course. So again, if the learning objective also includes, let's say more the AI literature skills or working with AI skills, I think it will be beneficial to have some form of evidence saying, okay, how well did you prompt? But then that's a form of assessment as well, almost. I think for other courses, I think indeed. Yeah, I struggle with that as well, because I think what do you got, what are students going to do? They're going hand in 50 pages of their prompts and their questions. Are you going to read it and what are you going to do with it, right? And again, does it really say something about what de have done with AI? Apart from discussing, or next to discussing, why do we teach and what do we teach and all that, maybe we should also think about again, what is an academic integrity, like our academic integrity, and what does that look like? I think we very much with assessments and moved into the check boxes. But what is it really about? It's about students also showcasing what they can, and if they can't do it, then there are other ways also maybe to come to that goal. So yeah, I would, I think I would go for the honesty card. But, yeah, depending on the learning objective. But if you have a good one, then I would gladly be here as well.

[I]: If you would like to implement these guidelines for your courses, what institutional steps would you need to take?
[R]: If I remember correctly, we have to, if you want to change your course, your course objectives, you have to let your programme director know first of course because he or she is responsible for the entire alignment and then you are directed to the board of examiners and they also check and see and then there's this whole process of filling it in in the format for study guide. Putting it on there, and designing the whole course.

[I]: What would be an ideal way for implementing such guidelines, if you would do so in the first place?
[R]: Yeah, so leave more room for, for changes that are more, that need a fast response. Yeah, I think. Because on one hand, I think the process needs to be even more tight. Because if we really want to have this ensured in all courses, and that requires a lot of talks together. And we really have to understand that the courses are not an island, but they're building on things, especially if we also next to like, like we teach skills such as collaboration and communication. Those are often grasped into learning lines. So courses are part of a learning line and they contribute to like a longer learning objective in the final attainment, um, final attainment level. But I think if we have that for, if we also say we need to prepare our students for a work with AI that requires some AI literary skills, that's another topic that we need to be aware of, which courses correspond with that learning objective. So that would require even a stricter thing. So I think having the process for learning objectives of the course. Could be all right. Maybe I see the most intermediate change is in the way we assess. Because maybe last year we could say, hey, you can for an assessment just submit a video. But now if I put all that into Claude or ChatGPT with an image of myself, that tool can generate it for me, right? So that's something that changes. I see assessment formats or haven't been able to be more adaptable. But I would, I would say, maybe for for the course learning objectives. Yeah. Keep it tight.

[I]: Do you have any questions or additional side notes for me?
[R]: Oh, I'm very interested into reading it. If you're ready and yeah. Enjoy reading your insights because I think we're all learning. Everybody is here, so the more insights and the more, the better.