ChatGPT Dilemma: Effects of Generative AI on Higher Education in Systems Engineering

Emin Simsek
University of Southeastern Norway
+4791561048
eminsimsekm@usn.no

Gerrit Muller
University of Southeastern Norway
Telephone
Gerrit.muller@gmail.com

Kristin Falk
University of Southeastern Norway
+4797586909
Kristin.falk@usn.no

Abstract. This study explores the effects of ChatGPT on higher education in systems engineering. It focuses on how large language models influenced student learning and academic honesty before and after the introduction of ChatGPT 3,5. Comprehensive research is limited in the literature. This research uses surveys, experiments, and case studies to understand the role of AI in the Systems-Engineering Master's course at the University. . Most students used available AI tools in their homework, which helped improve the grades of their semester papers. The use of language models brings issues like plagiarism, the need for critical thinking, and low effort to write the term paper. This paper emphasizes the need for clear guidelines to ensure responsible use of AI and support ongoing skill development. The new guidelines can assist students and teachers in their learning and teaching processes, promoting ethical usage of AI and efficiency in their professional undertakings.

Keywords. Academic integrity, generative AI, higher education, large language model (LLM), ChatGPT, AI-assisted learning, systems engineering, guidelines, responsible use

Introduction

This paper investigates the impact of generative AI in higher education. "Generative artificial intelligence", or GenAI, is the term used to describe systems that autonomously create content. This content can range from text to photos and even audio and video ("Generative Artificial Intelligence," 2023).

Text-based GenAI is also called large language models (LLMs). There are several models or tools based on GenAI. Some of these created a "Google effect" regarding public impact. Digital conversational agents driven by artificial intelligence (AI) and natural language processing (NLP) are known as AI chatbots. (Dongbo et al., 2023). The first AI chatbot was ELIZA, which

was launched in 1966 (Murphy, 2023). ChatGPT is an AI chatbot or LLM that provides conversational service through machine learning using an extensive database (Spitzer, 2023). OpenAI launched ChatGPT 3.5 on November 30, 2022. There are AI tools after ChatGPT 3, like Google´s Gemini and Microsoft´s Bing chat. However, ChatGPT 3.5 made a critical impact on the public. 2024 there were approximately 180 million users, with 600 million active monthly visits. (Duarte, 2024). OpenAI trained ChatGPT 3 with 45 terabytes of text data by machine learning. (Cooper, 2023). After that, OpenAI launched ChatGPT 3.5 and ChatGPT 4 on March 14, 2023, with even more background data.

GenAI, like ChatGPT, is significantly affecting higher education. It changes how students see things, their willingness to use it, university policies, assessment procedures, and instructional assessment creation. However, there are concerns about its accuracy, privacy, and ethical implications (Chan & Hu, 2023). The concept of academic integrity explains moral behavior in learning environments, placing a strong emphasis on academic honesty, student-teacher interactions, and avoiding unfair benefits. (Zivcakova & Wood, 2014). Academic integrity requires avoiding plagiarism. The term plagiarism, which comes from the Latin word "plagiarius," is the dishonest use of someone's concepts, methods, or work without giving proper credit and selling it as one's own. (Mehić, 2013).

Universities worldwide have varying regulations related to GenAI. While many embrace the new technology for teaching and learning, others have banned it due to public opinion and academic reputation. According to Xiao et al. (2023), 67% of Universities used ChatGPT in teaching and learning. Yet, according to Schönberger, M. (2023), Urban, M. et al. (2023), and Rezaev and Tregubova (2023), there were concerns regarding how it will affect higher-order thinking abilities and academic integrity since relying too much on ChatGPT can reduce critical thinking and even result in academic dishonesty. Vicente-Yagüe-Jara et al. (2023) observed that students´ fluency, adaptability, and originality significantly improved with AI's help. According to several teachers, ChatGPT was a valuable tool for creating interactive courses, responding to inquiries, and offering individualized learning opportunities in the classroom. (Flaagan, 2023).

This research investigates the use of LLM in systems engineering education. Specifically, the problem concerns academic integrity regarding using GenAI for home exams at the master's level.

The University of South-Eastern Norway (USN) offers a "Systems Engineering" course at master's level. This is a compulsory course for the Innovation and Technology Management Program students. For the final exam, the students are asked to develop a project report based on a set of questions related to systems engineering. The students have ten weeks to answer the exam that tests various skills, including diagrams, literature, decision-making, creativity, conceptualizing, and analysis. In 2023, the exam questions were related to a) Introduction- problem definition, b) ConOps- as/is-to/be, c) Stakeholder analysis, d) Concept creation and selection, e) Verification and validation, f) Risk assessment. Teachers assess the following skills (5 points each). The project's quality and completeness are each section's score criteria.

The main author conducted an initial survey with thirteen teachers from USN. Five are associated to systems engineering, five to electrical and software engineering, and three are teaching business and innovation. Their primary concerns related to negative effects from AI were:

- academic integrity,
- the sufficiency and capability of LLMs in answering home exam questions
- the possibility of weakening competence development

- existing policies and guidelines

At the time of the research, there were no clear guidelines for the use of AI. Teachers wanted to allow students to use ChatGPT for their exams. However, they were uncertain whether the competence development would be the same. The university employees aimed to understand how students use GenAI and its implication on academic integrity. Of the respondents, only four agree that Gen AI, like ChatGPT, has improved exam responses, while six feel it negatively affect student learning. One teacher thinks it depends on the student. The remaining individuals are either neutral or lack experience. Eleven of the respondents asserted that they need a clear guideline or policy. Furthermore, eight of expressed concern over academic integrity.

The goal of this research is to understand the implications of GenAI for academic integrity. Specifically, we assess the adequacy of ChatGPT for the Systems Engineering home exam, whether it maintains academic integrity and influences the learning outcomes.

The proposed solution is a set of guidelines designed to help students and teachers to adapt and effectively use AI for home exams. These guidelines aim to improve AI-assisted learning and promote academic honesty. Furthermore, the guidelines are intended to foster competency development among students within AI-assisted learning environments, preparing them for future careers where they can use AI's conscious and responsible.

Research question that guides this study is: What measures could the university staff and students implement to ensure a beneficial use of generative artificial intelligence in systems engineering education at USN?
   a. According to educators, what are the criteria for assessing final papers in an introductory course called ´systems engineering´?
   b. What challenges and solutions have the literature proposed for enhancing students' learning in fields related to Systems Engineering?
   c. For what purposes are students and teachers in Systems Engineering currently using generative language models?
   d. How have the answers in the course papers on "systems engineering" changed after the introduction of generative artificial intelligence (ChatGPT)?

This paper begins with an overview of the research methods and design that includes a case study, surveys, and interviews. This is followed by a literature review that examines the use of GenAI in various educational fields, providing a broader context for the research. The next section presents a summary of numerical results, offering quantifiable insights into the study's findings. This is followed by a comparative analysis of exam papers pre- and post-ChatGPT introduction, shedding light on the impact of GenAI on student performance and writing style. The paper then proposes new guidelines for systems engineering course papers in the context of AI. This set of recommendations aims to guide educators and students in effectively integrating AI into their learning and assessment processes. The paper concludes with a discussion, drawing together the key findings related to the research questions. Finally, the conclusion provides a summary of the research, contributions, and potential for future study.

Research Methods and Design

To explore AI effects on the Systems-Engineering studies, this research takes a systems-thinking perspective, understanding each part, making connections, and seeing the holistic aspects. Figure 1 illustrates the approach that includes surveys, interviews, analysis of former exam papers and an experiment with ChatGPT generating an answer to the exam.

The intent of each of these are as follows: Students' and teachers' surveys and literature help to define problems and requirements for the guideline. Student's exam papers help understand

AI's effects on home exams. While quantitative analysis of the papers gives an overview of the situation, qualitative analysis will help to see details and AI`s effects on student`s competence. Finally, interviews with three experts validate the designed guideline.

We performed an initial Survey answered by thirteen Academic Teachers. The purpose is to learn more about their views on GenAI integration related to innovation, systems engineering education, and academic integrity. The survey contains twenty Likert questions. The questions are related to teaching background, experiences with ChatGPT, how they perceive advantages and disadvantages of GenAI, worries related to academic honesty and privacy, and thoughts on the university's usage restrictions for GPT.

Case Study: We collected ten final exam papers from the ITM4200-Systems Engineering master course at USN, comprising four papers from 2022 and six from 2023. This home exam gives ten weeks to answer a 15-page report on system development on the topic of Fundamentals of Systems Engineering competencies.

ChatGPT 4 Experiment: As an experiment, the main author produced an exam paper for the 2023 exam together with ChatGPT. The author experimented with exam questions to ChatGPT 4 and copied and pasted all the answers into the exam paper. At the time of the research, ChatGPT 4 is the most developed and popular LLM. The author initially typed the following text string into ChatGPT: "This is a system engineering course homework. I will ask questions individually, but it will be a report about creating energy solutions. Make it detailed:" After that, the author asked each question from the "2023 exam questions. "The exam with ChatGPT took the author 1.5 hours to copy and paste one by one. The goal of this experiment is to validate the comparison between the pre and post-ChatGPT 3.5 (November 2022) eras. Should the implications observed over the years parallel those in the experimental paper, it will facilitate a more straightforward discussion of potential correlations.

Analysis: Thereafter, the main author analyzed and compared the 11 papers (Four from 2022, six from 2023, and one created in the ChatGPT experiment). The analysis considered average sentence length, word variety, plagiarism rate, AI detection rate, and unique word rates. A Python script, written by the author and supported by ChatGPT, were used to calculate the average sentence length and word variety. The average sentence length is the word count in every sentence, and the meaning of word variety is the number of different words.. The author also calculated the percentage of unique words using Grammarly, a useful metric for comparing different databases. The AI detection rate, as defined by 'Gptzero.com', signifies the frequency of supervised AI, where both AI and human data are used for training , usage identified in each exam paper."
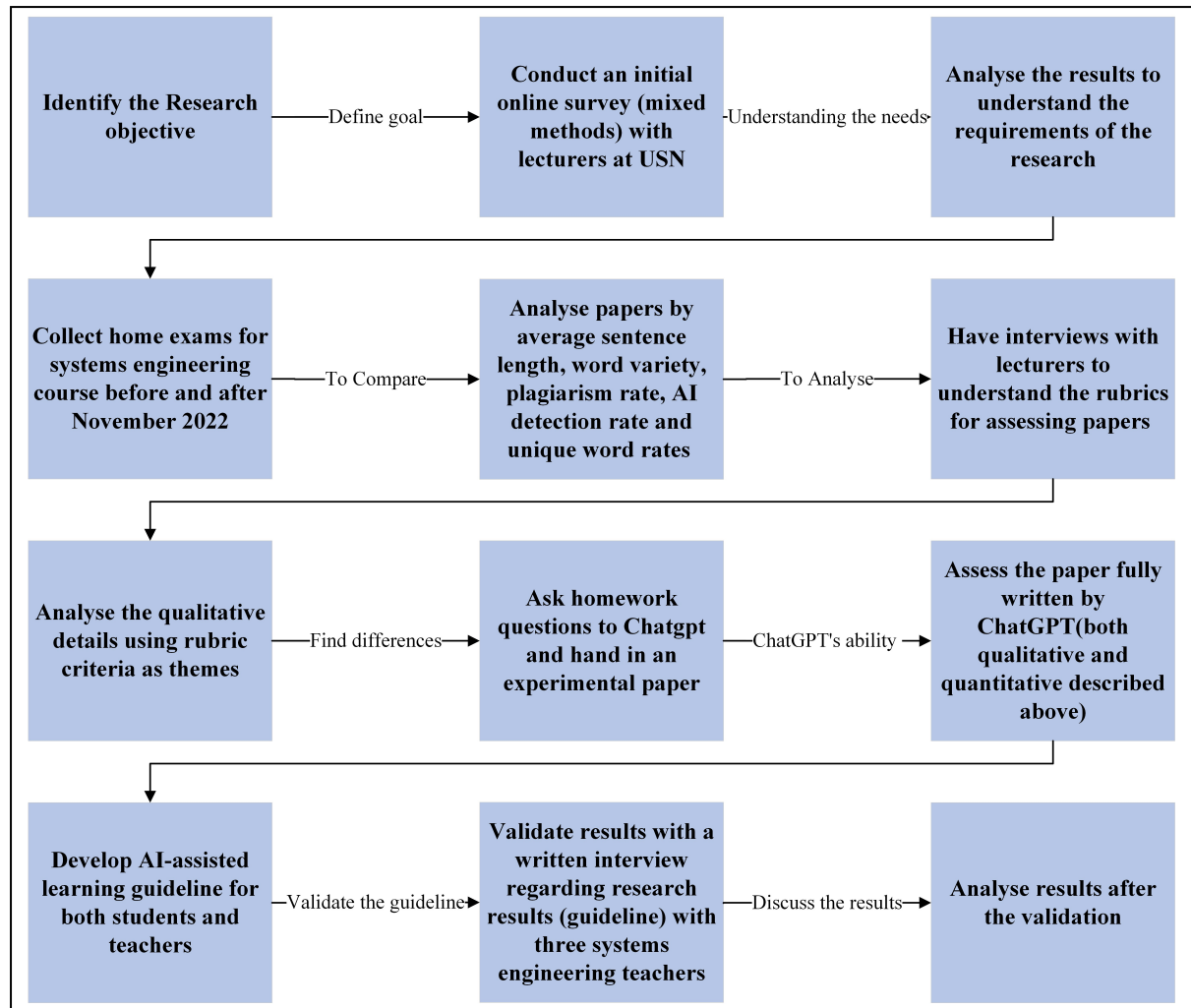
Figure 1. Research methods diagram (Drawing5.vsdx) (Simsek, E., 2024)

Figure 2 presents the analysis tools. The author classified the 2022 exam, 2023 exam, and 2024 ChatGPT experiment into three categories. He analyzed these per exam question using "Maxqda" for qualitative thematic analyses. ChatGPT also helped to analyze, but only partially and indirectly. Two university teachers assessed the 2024 ChatGPT experimental paper. We did this qualitative analysis to understand the sub-research question: How have the answers in the course papers on "systems engineering" changed after the introduction of generative artificial intelligence (ChatGPT)?
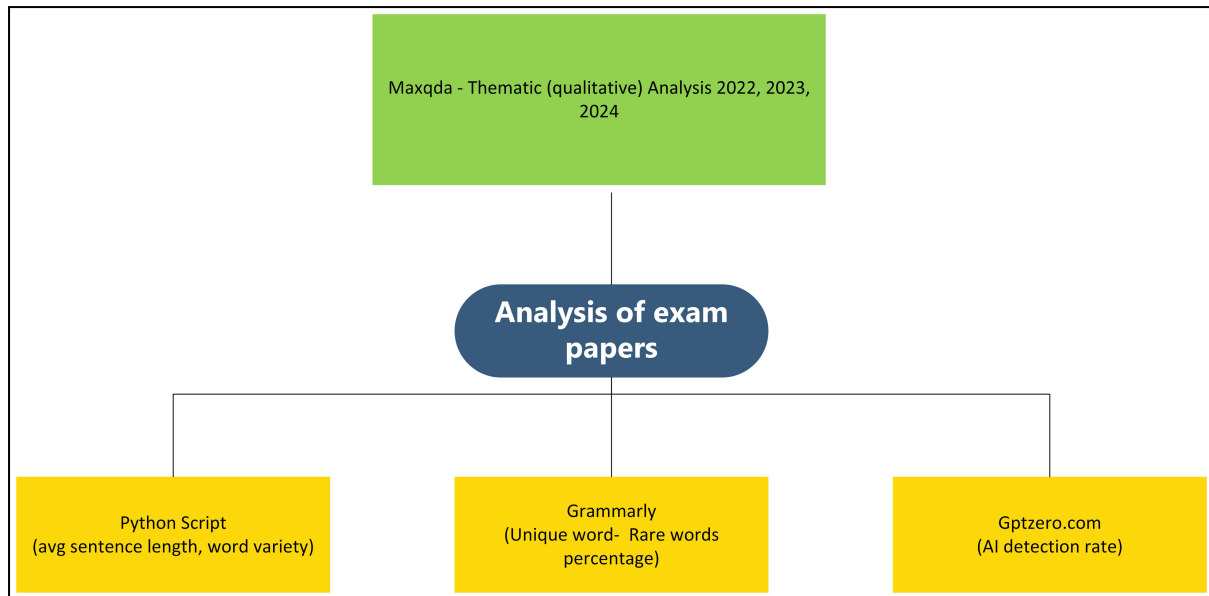
Figure 2. Research analysis

Student Questionnaires were answered in parallel to collecting the exam papers. Each student that gave in their student paper for research purposes also answered the questions were related to usage of Chatbots during their assignment, use cases, grade for their assignment, suitability of chatbot use in various exam steps, privacy concerns, and satisfaction.

Two experienced assessors evaluated the experimental exam paper produced by ChatGPT4. Their expertise in grading exams provided insights into the differences between content created by AI and that created by humans.

To validate our results, we performed three written interviews. Each included fifteen questions regarding the guidelines. The interviewees are the systems engineering course owner, course instructor, and senior instructor at USN. The questions comprise each part of the guidelines (ten questions) and the main changes in the new possible rubric design (five questions).

The literature review forms the foundation of this research. It encompasses works from related fields that contribute to the ever-growing body of knowledge in this area.

Literature Review

Literature review methodology. The literature review involves academic search tools like Google Academic, Web of Science, and Typeset.io. There was no time filter while searching. Scribbr and Google Scholar helped with the references. This research scanned around 50 articles; this literature shows over 15 articles. The keywords are "Language Models and Higher Education," "Generative AI and Teaching in Higher Education," "Systems Engineering and Language Models," "Learning and Language Models," "Learning and ChatGPT," "ChatGPT and Higher Education," "ChatGPT and Systems Engineering," "Language Learning Models and ChatGPT," "Effect of ChatGPT on Higher Education."

De Castro (2023) had a literature review article about the benefits and concerns of using ChatGPT in education. According to the study, using GenAI may improve student engagement and motivation, and the personalized learning experience with an immediate feedback system may enhance students' language skills, decision-making support, and data analysis. On the

other hand, privacy, academic integrity, and potential data biases are the concerns. Additionally, an ethical policy requires the use of GenAI in scientific papers. Another literature review paper by Alqahtani et al. (2023) delivers literature in two parts: teaching and researching. It shows that using AI tools in teaching may help analyze data, automate grading, and develop personalized learning plans. On the other hand, Cassidy (2023) focuses on the unjust use of AI. Universities that permit their students to use ChatGPT may have an advantage over universities that do not allow their students to use it. The author provides some examples, such as Australian universities trying to return paper and pen exams. Another suggestion in Shaw et al.'s article (2023) is not to give essay-writing exams but to give discussion and reasoning exams on already written ones.

Some papers focus on experimental studies. Meyer et al. (2024) tested the difference between ChatGPT 3 and 4 for creating Knowledge graphs with the same prompts. According to the survey, ChatGPT 3 creates more correct results than ChatGPT 4. However, ChatGPT 4 is better at using only defined properties. However, the results heavily depend on prompts. Jungherr (2023) experimented, asking 24 questions to ChatGPT 3.5 to understand the sufficiency. ChatGPT´s capacity decreases when a specific question is asked. It can be helpful to explore something, formulate a research question, conduct data analysis, structure a paper, or edit it. Frith (2023) described ChatGPT as a disruptive educational technology and asked ChatGPT to create paragraphs. Frith has concerns about overusing, accountability, and failure of current plagiarism detection programs. Frith suggests deep learning of the background of the LLMs and the data behind them as a solution while using them for educational purposes. Joshi et al. (2023) researched real exam questions on different types of questions about data algorithms, operating systems, database management systems, machine learning, GATE, and coding. According to the results, coding questions have 92% accuracy. The data algorithm questions have 70.1% accuracy because they mainly consist of true and false questions. However, database management system questions have 33% accuracy, 76% accuracy in design questions and 75% in true and false. However, there is 39% accuracy in numerical questions. (Joshi et al., 2023). In this case, ChatGPT does not seem helpful with numerical questions. Although it appears to be the easy part for the AI,

Some qualitative papers test grades and similarities. AlAfnan et al. (2023) have 30 tests with ChatGPT, but the research also graded and checked the Turnitin similarity. All grades are over 70%, only 4 have over 30% similarity, and most are under 15%. (AlAfnan et al., 2023). The main criticisms are the need for more detail and the answers being too general.

Deshpande & Szefer (2023) have researched exam papers for the computer engineering course exam for the University of Yale; they also ask students to upload another version of the exam paper with the written assistance of ChatGPT. However, they mentioned that it would not affect their grades and would be just for the research. In this case, it may decrease the power of the study. It collects papers for quizzes, exams, and lab exams. Interestingly, 100% of text-based Quiz 1 has an average grade of 2.7 without ChatGPT, which increases to 3.0 (the maximum point is 3.0). Assignment 3 has 72.6% figure-based and 27.3% text-based. 21.7 is the average grade without using ChatGPT (out of 25), and the average score dramatically decreases to 3 (out of 25). (Deshpande & Szefer, 2023).

De Vicente-Yagüe-Jara, M. et al. (2023) wrote an interesting paper comparing creativity between humans and artificial intelligence in universities. The author describes the indicators by the "creative imagination test": fluency, flexibility, and narrative originality. In some tests, fluency is three times higher in AI, flexibility is almost doubled, and narrative originality is more than three times higher. (De Vicente-Yagüe-Jara, M. et al., 2023).

Another literature review article by Shaw et al. (2023) focuses on using GenAI tools because ChatGPT, or what Shaw called "CheatGPT," may create fake facts or machine hallucinations. One of the concerns is plagiarism on text-based assignments, and as of 2023, there is no publicly available AI plagiarism checker (Shaw et al., 2023). Some papers have focused on describing flow diagram guidelines to help with the effective use of AI. Ngoc et al. (2023) had a case study for introducing the computer science course. They categorize the exams and exercises as theory, programming, and applied exercises. Then, the author collected the students' and lecturer´s perspectives and described the starting point as contextualizing and forming the prompts, evaluating the initial results taken from ChatGPT, comparing the reliability of its theory, and simulating and validating if it is coding. So, if the answers are not satisfying, stop using ChatGPT. (Ngoc et al., 2023). There is a similar paper but resulted with the more comprehensive guidelines by T. Pham et al. (2023). This paper is also from the same case study as Ngoc et al. (2023) and similar processes described in the guideline. It includes conceptualizing the prompt, controlling the answers, and validating.

According to the research by Kontovas (2024), "delve into," "embark on a journey," "in the realm of," and "it is essential to note that" and "seamlessly" are the most common phrases ChatGPT uses. However, the usage of these words is dramatically decreasing after 2023. It may be shown that ChatGPT 4 is not used anymore, or people are paraphrasing.

Table 1 summarizes the concerns and solutions made by the academic environment in the last few years to answer the second sub-question: b. What challenges and solutions have the literature proposed for enhancing students' learning in fields related to Systems Engineering?

Table 1. Summary of the challenges and solutions described in the literature.

Reference	Challenges	Solutions
Adar & Kandemir (2008)	Concern about academic integrity and ethics	Enhancement of soft skills in the education process
AlAfnan et al., (2023)	Lack of details. Too general answers. Informality of the learning from ChatGPT and academic integrity.	Take home exam style and teacher's assessment rubrics may change.
Alqahtani et al. (2023)	Being not comprehensive in literature review, lack of complexity and ethical considerations.	Using AI for assisting, accelerating, and facilitating literature, text generation, and data analysis.
Bower et al. (2024)	Unconscious effect of generative AI on education: teaching motivations and adaptations	Encouraging authentic tasks and real-world application in teaching
De Castro (2023)	Concerns over privacy and academic integrity and data manipulations.	AI use must be clearly declared. However, any use of AI is not breaching the Cambridge plagiarism policy.
De Vicente-Yagüe-Jara (2023)	LLM's sufficiency and qualitative impacts	Analyzed the originality, creativity, and flexibility by using a creative imagination test.
Eager & Brunton (2023)	Integration of AI tools to develop future skills and how to command effectively?	Emphasis on industry-specific skills and academic skills in assignments and development of a command process for AI.
Frith (2023)	Potential underdevelopment of student skills, overusing, accountability, and failure of current plagiarism detection programs.	It is suggested that students need more time to understand the algorithm and use it effectively.

Joshi et al. (2023)	ChatGPT does not consistently provide accurate explanations and answers. (Sufficiency of AI)	Giving the right prompt may increase accuracy. Also, students must use it as an assistant.
Jungherr (2023)	Misinformation and limited temporal coverage.	Asking questions from general to specific, using theory rather than a literature review.
Ngoc et al. (2023)	Integration of AI assisting learning	The diagram introduced for the use cases and purposes.
Shaw, Morfeld & Erren (2023)	Ethical considerations, misuse, and insufficiency of ChatGPT and competitive advantage of using it.	ChatGPT should be able to detect, or assessment style may change rather than giving an essay exam. However, a common policy is critical for equity.
Su, Scarinci & Cicirello (2023)	Challenges in integrating LLMs into systems engineering	Use in automatic generation of systems engineering diagrams from the written texts for the requirements
Szefer & Deshpande (2023)	Students are using AI chatbots but there can be some limitations.	ChatGPT may be used as a supplementary tool; teachers may apply laboratory-based or figure-based questions if they are willing to avoid them.
Walczak & Cellary (2023)	AI systems disrupt traditional education. Also, they have originality issues and 'hallucinations.'	Students should develop new skills like AI-human communication, critical thinking, etc., because they will replace the jobs they are preparing for.

Previous literature mainly focused on experimenting with ChatGPT or other LLMs. Articles highlight the importance of academic integrity after GenAI. Some exceptional papers focus on collecting first-hand data, which is data from students and their improvement after the ChatGPT revolution. The literature on how to use ChatGPT is primarily related to computer science. Systems Engineering should learn from the experiences of other domains. Some papers need to be more specific about the topic, and they are primarily short articles and opinion essays. Some papers focus on experimenting with ChatGPT with different prompts, and some papers about systems engineering focus on integrating LLMs into requirements management. Comprehensive research in systems engineering is limited.

Numerical Results

This section shows the numerical and some qualitative result from the initial survey on how the teachers envisage GenAI. Furthermore, it shows the initial answers from the students on their use of ChatGPT, and the initial numerical analysis of the student papers.

The research began with identifying the initial problem and gathering requirements for potential solutions by surveying ideas from teachers at USN. The survey revealed that while teachers are permitting students to use GenAI, they have several concerns

According to the teachers, universities should not restrict the use of GenAI for courses and exams. Yet, controlling and supporting it can be necessary, as 46% of the respondents propose some control over AI use.
- 38% think the university needs a more explicit statement for GenAI.
- 84% of the teachers require additional support to integrate the GenAI.
- 76% recommend that universities be able to verify the usage of GenAI for the exams.

The teachers also responded the open-ended question: "How do you see the problem in your own words and point of view?" Responses revealed concerns about AI's impact on learning, with some fearing that it could undermine critical thinking and problem-solving abilities, and suggesting methods like oral exams to counterbalance its use. There was a common sentiment

that guidelines are needed for grading AI-generated content and training students to use AI effectively. While some respondents did not see AI as problematic, others emphasized the importance of teaching students to use it constructively. Concerns were also raised about the quality of AI-generated answers, with teachers calling for regulation to ensure AI enhances, rather than hinders, learning.

The Case Study investigated student papers before and after ChatGPT 3.5. Table 2 shows numerical results. The indicators in the numerical analysis of the exam papers are the average sentence length, unique words, word variety, AI plagiarism rate, plagiarism, whether they use ChatGPT on assignments, and how they used it in their assignment. These indicators are to support thematic analysis with quantitative data.

In addition to the results from the initial analysis of the papers (Table 2), the author asked the students further questions concerning privacy, satisfaction, and usage. The answers reveals that 64% have no privacy concerns, and the average satisfaction level for using ChatGPT is 5.6 (out of 10). According to the student survey, 71% of the students used ChatGPT to help answering this home exam. 85% of the students are using it in general.

Figure 3 displays how the students have used the chatbot. Their use differs. 50% of the students used it for chatting while doing this home exams. 43% used it for language concerns, grammar, or spelling. On the other hand, only 14% of the students used it for literature reviews, 21% for concept creation, and 7% for decision-making during homework.

Table 2. Initial analysis and comparison of eleven papers

Year	Paper	Grades	Average Sentence length	Unique words	Word variety	Plagiarism rate	AI Plagiarism	using ChatGPT	What was their use case?
2022	Paper1	C	18.7	23 %	578	2 %	2 %	No	N/A
2022	Paper2	A	22.7	18 %	750	4 %	1 %	No	N/A
2022	Paper3	A	20.6	20 %	766	7 %	2 %	No	N/A
2022	Paper4	C	24.3	18 %	605	5 %	1 %	No	N/A
2023	Paper1	C	18.7	14 %	1092	8 %	82 %	Yes	Better writing, concept and idea creation
2023	Paper2	A	14.5	16 %	944	1 %	11%	Yes	Asking specific questions/Chatting
2023	Paper3	A	14.7	15 %	761	4 %	82 %	Yes	Asking specific questions/Chatting, concept and idea creation
2023	Paper4	A	22.1	19 %	1118	2 %	1%	Yes/No	Asking specific questions/Chatting
2023	Paper5	C	12.6	19 %	785	3 %	16 %	Yes	Literature review, idea concept creation, Chatting
2023	Paper6	A	18.3	18 %	1089	7 %	40 %	No	Better writing

2024	Experi-mental Paper1	C (As-sessed by 2 teach-ers)	16.5	20 %	1078	4 %	100 %	Yes (100 %)	All
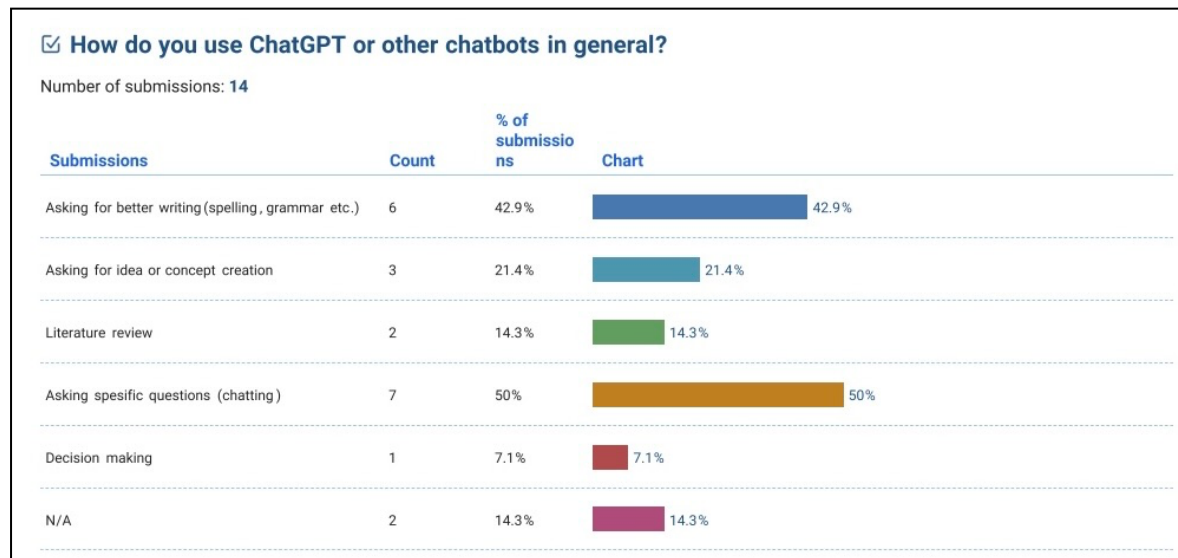
Figure 3. Chatbot use cases for students

Comparing Student Papers and the Experimental Paper

Thise section compares the exam papers from 2022, 2023 (when ChatGPT 3.5 was available) and the experimental paper generated in 2024 by ChatGPT4. Some of the core results were already presented in Table 2. We start by looking at the results per year.

The 2022 exam papers is used as a reference for the standard quality. The average sentence length for these papers ranged from 18.7 to 24.3 words, and the percentage of unique words ranged from 18% to 23%. The students we interviewed did not really use chatbots in "GPTzero"'s AI detector checker. The average plagiarism percentage was between 1% and 2%.

The 2023 exam papers demonstrate notable changes. The AI plagiarism rates increased from an average of 1% to 38%. We have a reason to believe that this is resulting from the increased use of ChatGPT in exams. There was no significant increase in plagiarism rates and a slight decrease in unique word usage to 14%; compared to 2022. The average sentence length de-creased from 21 to 16 words, and the unique word percentage decreased from 19% to 16%. Students, on average, used 674 different words in 2022 and 964 in 2023. However, word vari-ety as a percentage is decreasing because papers are significantly longer.

The 2024 experimental paper generated by ChatGPT-4 verifies the "GPTZero" s AI plagiarism rate with 100%. The variety in word usage is 1078 words. We saw a similar increase from 2022 to 2023. Academic plagiarism rates are similar to the exam papers from 2023. Recent literature Kontovas (2024) revealed that the most commonly used phrases by ChatGPT were "delve into," "embark on a journey," "in the realm of," "it is essential to note that," and "seamlessly".

However, in the student papers these terms were not that prominent. Yet, "it is important to note that" was used 25 times in the four 2022 exam papers and 43 times in the six in 2023.

Two teachers assessed the 2024 experimental paper to be a lower C according to the Norwegian grading system, where an average student should obtain a C). This is below average in both 2022 and 2023. The two teachers are normally teaching this course, and one of them also graded the 2022 and 2023 exams. Criteria for the assessment were the paper's alignment, clarity, coherence, depth, accuracy, engagement, critical thinking, and originality, according to the assessors. The exam did not explicitly contain these criteria. The first teacher observed that the 2024 experimental paper looked good. It is well organized and theoretically sound. References and literature are accurate. However, diagrams are beautiful and not functional. The paper does not provide explanations or precise numbers. Verifications are not completed. Risks are too general and just listed without sound explanations. Teacher 2 has similar critics. The teacher expected to see the text repeating questions. But in this paper, all questions jump to the solutions. Diagrams are just nice pictures. Numbers do not have concrete explanations. There is also a lack of originality. Verifications are not clear and helpful. The paper could have take a B grade with small touches, but the grade would be a C.

Interestingly, when ChatGPT assesses papers, it likes its own paper better. Still, the author presents two teachers' evaluations for the experimental paper.

Table 3 contains a generalized analysis of the exam papers. One column for each year, and one row for one sub-question of the exam. The sub-questions are related to introduction, concept creation, stakeholder analysis, concept creation, verification method, and risk assessment. These are the parts that are similar for both years' exams. The table contains the two teachers' comments for the 2024 paper and the main author's observations for all the other papers. Comments are not valid for each paper but for general comparison. We use this to answer the sub-research question: d. How have the answers in the course papers on "systems engineering" changed after the introduction of generative artificial intelligence (ChatGPT)?

Table 3. Thematic analysis of the exams according to the current rubric and each question

Criteria	2022 Exam papers	2023 Exam papers	ChatGPT Exam Paper
a) Introduction-	Explains SDG but often lacks depth coverage of its components.	Detailed introduction on SDG, covering its components and relevance.	Clear and well-structured, strong alignment with energy goals. But lack of details and starts with solution.
a) Problem Statement	Generally clear but could be more concise and specific; often lacks depth in problem articulation.	Clear and well-articulated problem statement with specific details and context.	Problem defined too general and background information is too short.
b)Concept creation as-is, to-be	Present but often lacks detailed descriptions and context.	Detailed with thorough context and descriptions.	As-is/to-be is listed directly but its. CONOPs diagram is just a picture and is not good enough.
c)Stakeholder analysis	Requirements are listed but not thoroughly explained or aligned with stakeholder needs.	Detailed and well-aligned with stakeholder needs, providing clear and specific requirements.	Broad identification but it is just listed and described in one sentence.

d)Concept creation and decision making	Described with Pugh Matrix but often lack full exploration of limitations and detailed comparisons.	Thoroughly described with detailed Pugh Matrix analysis, including comparisons and limitations.	While deciding the solar panel concept, criteria pointed with similar grades in Pugh matrix. So, decision is not explained well.
e)Verification method	Described but need more detail and traceability to stakeholder requirements.	Detailed and with clear traceability to stakeholder requirements.	Logical and includes explanations. However, it is only verifying design and prototype phases.
f) Risk assessment	Risks are mentioned but not thoroughly analyzed, with limited discussion on mitigations.	analysis of pros, cons, and detailed mitigations with thorough discussion.	Risks are well listed and explained but all the risks are just financial, social and climate based. There is no technical risk identified.

Exam papers from 2022 had insufficient problem descriptions and decision-making procedures, and they were not comprehensively covered, in-depth, or in line with stakeholder requirements. The papers from 2023, on the other hand, significantly improved with thorough introductions, precise problem definitions, thorough background, and detailed stakeholder analysis. The experimental 2024 (ChatGPT paper) had general problem descriptions, minimal stakeholder analysis, detailed but poorly shown concept generation, limited verification focused on design and prototype stages, and precise but broad introductions. Technical risk factors do not explicitly evaluate risks.
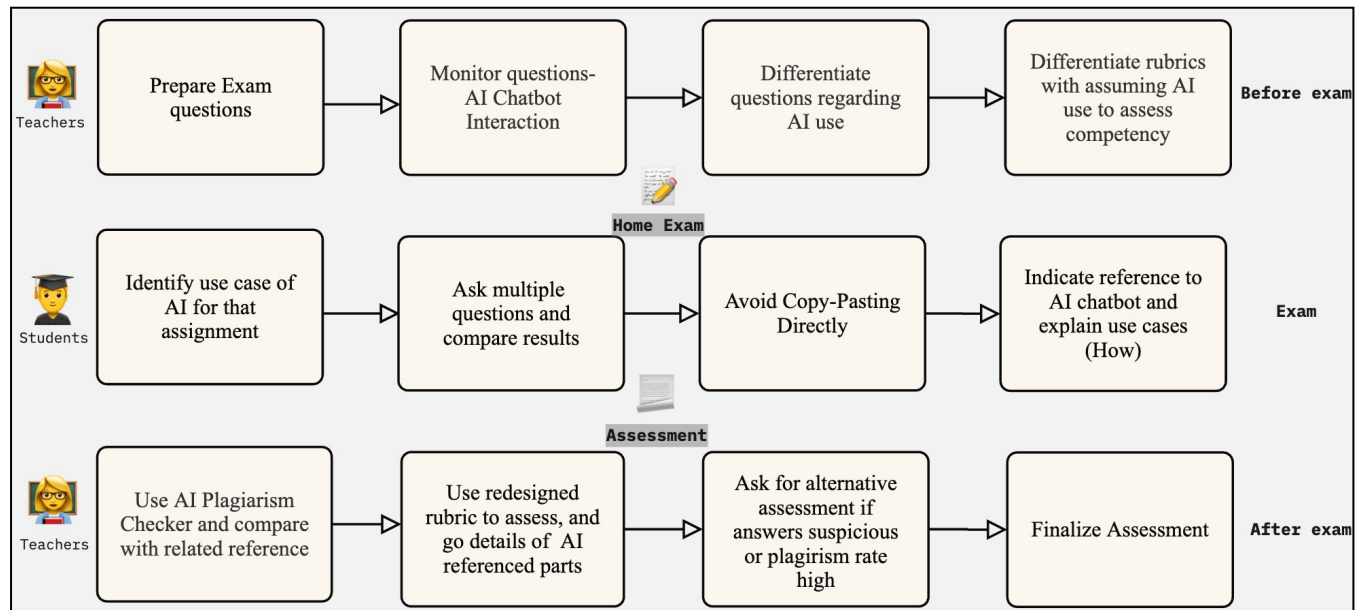
Guidelines for AI-Assisted Learning

The we propose the below guidelines for students and teachers using LLMs responsibly in an academic environment. The main aim of the AI-assisted learning guidelines is to maximize the positive impacts of LLMs on learning. They are based on literature review and our research.

The guidelines depicted in Figure 4 are specifically designed to aid in the effective use of large language models for home exams. These guidelines are divided into three sections: pre-exam, during the exam, and post-exam. Both the pre and post-exam sections outline the teachers' responsibilities, while the section pertaining to the exam itself is directed towards the students.

The Guideline:

- Before home exams, teachers should explore the questions and see how AI Chatbots can solve problems and redesign the exam. This will help teachers update and improve their exam questions. Teachers should adopt home exam questions to generative AI based on the assumption that students use AI.

- During the exam, students must decide on the use case of AI tools and explicitly mention both the use of tools and use cases. They should avoid copying/pasting anything directly. They need to ask the same question with different prompts for the best answer, as the literature mentions that prompting is critical for practical use. Mentioning the AI tool using the same logic as referencing a single article can be a requirement for academic integrity.

- After the exam, teachers should check AI plagiarism rates and Turnitin plagiarism rates. However, teachers should only use AI plagiarism results for suspicion. If suspicion is

high, teachers can ask to take alternative exams (oral exams, pen-and-paper exams, etc.). Teachers can determine their percentage rate regarding AI plagiarism.

Figure 4. AI-Assisted Learning Guidelines

Teachers should test the exam questions during preparation. If they are easy to solve without human interaction, they should change the questions, or teachers should change the questions from the entire question to solve to the structured exam guides through the process. Secondly, teachers can adopt rubrics according to the assessed competency and consider AI use from the beginning. Students should ask the same question several times with different prompts to obtain diverse insights. This will help them get unique answers instead of simply copying and pasting comments from AI chatbots. Students should mention how and when they used which AI tool and cite it as a reference before sending it.

Evaluation

This research involves a preliminary validation of our research with three Systems Engineering teachers responding to a series of fifteen questions. The teachers identified several significant areas where they believe AI impacts education, and suggested modifications to current procedures. Their perspectives on the results are detailed in Table 4. The teachers observed that the content of the AI-generated assessment could often be generic or irrelevant, leading some to modify exams to include more case-based and application-oriented assessments (Interviewee 1, Interviewee 3). While current detection tools are considered unreliable, there was consensus among experts that they should be mandatory (Interviewee 1, Interviewee 2). For experts 1 and 3 , oral exams were suggested as an alternative to AI plagiarism detection tools, which despite being useful, are only partially reliable.

Table 4. Validation of results

Written Interview	Interview 1	Interview 2	Interview 3
Changes in assessment practices	I have seen several cases where students send in a lot of "blabla" that apparently was produced by AI.	Assessment – enforced midway-feedback, introduced KPP's Dynamic Behavior.	Selected a more application-based exam where students needed to address a relevant case.

Challenges in AI declaration	We have too little experience.	It is possible to sense the use of AI if they don't declare it.	Requested as part of the academic honesty declaration.
AI plagiarism detection	It is more a detector of suspicious text, not fully trusted.	Yes, some frequent use of repeated words in the text could be a sign.	Plagiarism should be checked as part of the assessment.
Adapting rubrics for AI use	I don't see a need to adapt the rubric.	There shall be different rubrics for AI-supported answers.	More application-based examples in rubrics might help.
Alternative exams for AI misuse	I see oral exams as an alternative, but it increases workload.	Yes, physical Viva-Voce or multiple-choice exams without AI.	There should be a follow-up method to check competency.
Creativity and critical Thinking	Critical thinking and creativity should not change, rubric doesn't need adaption.	CONOPS development is a place for creativity.	Originality based on relevant case-based assessment is key.
Guidelines for AI use in exams	A digital assistant that doesn't leak data could be helpful.	Not planning to use AI in teaching, but active learning may benefit.	AI can be useful if treated as a co-teacher guiding the student.
Data security concerns	Prevent data leaks and have open discussions about AI use.	USN has a policy on AI tool usage that should be followed.	We advise students on risks, especially company-based examples.

The teachers had different views regarding redesigning the rubrics for grading. Interviewees 2 and 3 argued that adopting rubrics focusing on creativity and application-based examples may be needed to better assess AI-assisted work. In contrast, Interviewee 1 did not need a change.

The three Systems-Engineering teachers agreed that critical thinking and creativity should remain at the center of grading rubrics to ensure AI does not undermine students' abilities. Finally, there was a general concern about data security, particularly confidentiality when using AI tools, especially about student and business data (Interviewee 1, Interviewee 3).

Discussion

This paper is a result of an analysis at one educational program, with emphasis on evaluating student's papers and soliciting the opinions of professors and students at a specific point in time. As the field of AI rapidly evolves, the conclusions drawn here may change over time. Existing literature already points to factors that negatively affect academic integrity, such as co-author issues including ghostwriting and concerns about the reliability and quality of AI-generated research.

The central research question is broken down into four sub-questions that contribute to the understanding of the main query. The first sub-question is about the key criteria for assessing final papers in an introductory 'systems engineering' course. According to our research, educators focus on "alignment," "clarity and coherence," "depth and accuracy," "engagement and interactions," "critical thinking," and "originality and creativity" in their assessments.

The second sub-question addresses the challenges and solutions proposed in literature to enhance student learning in systems engineering. These include concerns about over-reliance on AI, its reliability, and possible negative effects on learning. However, the research also includes guidelines for AI-assisted learning, emphasizing the promotion of critical thinking, AI-human communication skills, and the application of real-world applications and lab exams for students.

The third sub-research question investigates how students and teachers in Systems Engineering currently utilize generative language models. Findings indicate that students primarily use chatbots to improve their writing skills and ask specific questions, accounting for 50% of usage. They also use AI to generate ideas, review literature, and aid in decision-making. Teachers allow students to use chatbots, likening it to the use of calculators as support tools. It's also worth noting that just 14% of students used such LLMs for literature review, 21% for concept creation, and 7% for decision-making during homework in 2022.

The fourth and final sub-question examines the changes in exam responses after the introduction of GenAI tools like ChatGPT. The analysis indicates a decrease in average sentence length and a greater variety in word use in 2023 papers with ChatGPT usage. Additionally, these papers presented a more detailed introduction, precise problem definitions, and thorough background, with a slight improvement in grades.

The research's validity is somewhat limited due to its small sample size, potential biases, and constraints imposed by General Data Privacy Rules. However, the honesty of the students in reporting their ChatGPT use, the comparisons made with existing literature, and different surveys add to the reliability of the findings. All students have English as their second language with Norwegian as their mother language.

The research acknowledges the risks associated with the rapid development of AI and the need for ongoing updates in this context. It offers a comprehensive perspective and a structured approach for integrating AI tools into home exams, thereby providing valuable insights for future research. It emphasizes the importance of clear policies governing AI usage in academic settings and the need for continual updates in research to keep pace with the rapid development of AI.

The primary research question is: What measures could university staff and students implement to ensure a beneficial use of generative artificial intelligence in systems engineering education at USN? The AI-assisted learning guidelines answer the main question and briefly require teachers to monitor and differentiate exam questions. Students must use AI chatbots responsibly for assignments by not copying answers directly and explaining their use of AI tools.

Teachers identified a need for clear policies and concepts to govern AI usage in academic settings. Teachers want to verify the use of AI and see it explicitly and be able to confirm with external tools. However, teachers may use AI plagiarism checkers before assessing the papers. Students already use ChatGPT, which improved their grades and provided easy, fast, and tailored access to information.

Surprisingly, students and teachers do not consider privacy an issue while chatting with LLMs. However, the Confidentiality of personal and corporate information may be a problem. Teachers should be able to inform students that it may be an issue. Some chatbots, like ChatGPT, have temporary chat options that do not use your chats to train their intelligence.

The main problem of the teachers is about "how" to integrate GenAI in exams as a supportive tool in students' competence development. In this case, the AI-assisted learning guidelines emphasize a structured approach for integrating AI tools like ChatGPT in-home exams for teachers and students.

Conclusion

This research presents guidelines for the use of generative AI in systems engineering education at USN. The study identified criteria for assessing final papers, including creativity, fluency, flexibility, and originality. The research revealed specific challenges, such as the over-reliance on AI, its reliability, and possible adverse effects on learning.

The research highlighted that the use of ChatGPT has led to more accessible language use and slightly better grades in student exams. However, the validity of the research is limited due to a small sample size and possible biases. The study also acknowledged the risks associated with rapid AI development, implying the need for continual research updates.

First-hand data from post- and pre-ChatGPT 3.5 show the AI effects directly. This paper has the correct timing for observing AI-assisted learning without any rules. In this case, it will help relieve concerns regarding uncontrolled AI use and the ChatGPT dilemma.

Future research should involve larger, more varied samples to enhance validity and reliability. It should also continuously update to keep pace with the rapid development of AI. Implementing the proposed guidelines at the University of Southeast Norway could serve as a valuable case study for change management. Understanding how systems engineering students applies LLM has a clear parallel to the professional use of AI in systems engineering. It is a research topic of great interest at the University of Southeast Norway. This research is a starting point for increasing efficiency in systems engineering education and profession with AI-assisted learning.

Acknowledgement

References

Adar, N., & Kandemir, M. C. (2008). M-learning tools for palm device: M-test and m- exercise. Paper presented at the 8th International Educational Technology Conference, Anadolu University, Eskisehir.

AlAfnan, M. A., Dishari, N. S., Jovic, N. M., & Lomidze, N. K. (2023). ChatGPT as an Educational Tool: Opportunities, Challenges, and Recommendations for Communication, Business Writing, and Composition Courses. *Journal of Artificial Intelligence and Technology*. https://doi.org/10.37965/jait.2023.0184

Alqahtani, T., et al.2023, "The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research," *Research in Social and Administrative Pharmacy*, vol. 19, no. 8, pp. 1236–1242, viewed https://doi.org/10.1016/j.sapharm.2023.05.016

Andreas, Jungherr. (2023). Using ChatGPT and Other Large Language Model (LLM) Applications for Academic Paper Assignments.   doi: 10.31235/osf.io/d84q6

Bower, M., et al. (2024). How should we change teaching and assessment in response to increasingly powerful generative Artificial Intelligence? Outcomes of the ChatGPT teacher survey. Education and Information Technologies. Retrieved from https://doi.org/10.1007/s10639-023-12405-0

Cassidy, C. (2023, January 10). Australian universities to return to 'pen and paper' exams after students caught using AI to write essays. *The Guardian*. https://www.theguardian.com/australia-news/2023/jan/10/universities-to-return-to-pen-and-paper-exams-after-students-caught-using-ai-to-write-essays

Chan, C. K. Y., & Hu, W. (2023). Students' Voices on Generative AI: Perceptions, benefits, and challenges in Higher education. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2305.00290

Cooper, K. (2023, September). *OpenAI GPT-3*. springboard.com. https://www.springboard.com/blog/data-science/machine-learning-gpt-3-open-ai/

De Castro, C. A. (2023c). A Discussion about the Impact of ChatGPT in Education: Benefits and Concerns. *Journal of Business Theory and Practice*, *11*(2), p28. https://doi.org/10.22158/jbtp.v11n2p28

Deshpande, S., & Szefer, J. (2023, July). Analyzing ChatGPT's aptitude in an introductory computer engineering course. In *2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE)* (pp. 1034-1038). IEEE.

De-Vicente-Yagüe-Jara, M.-I., et al. (2023). Writing, Creativity, and artificial intelligence. ChatGPT in the university context. Comunicar, 31(77). Retrieved from https://doi.org/10.3916/c77-2023-04

Dongbo, M., Miniaoui, S., Fen, L., Althubiti, S. A., & Alsenani, T. R. (2023). Intelligent chatbot interaction system capable for sentimental analysis using hybrid machine learning algorithms. *Information Processing & Management,* 60(5), 103440. https://doi.org/10.1016/j.ipm.2023.103440

Duarte, F. (2024, June 8). Number of ChatGPT users (Jun 2024). *Exploding Topics*. https://explodingtopics.com/blog/chatgpt-users

Eager, B., & Brunton, R. (2023). Prompting higher education towards AI-Augmented Teaching and Learning practice. Journal of University Teaching and Learning Practice, 20(5). Retrieved from https://doi.org/10.53761/1.20.5.0

Flaagan, T. (2023). Using ChatGPT to generate computer science problem sets. *International Journal of Integrating Technology in Education*, 12(02), 19–29. Retrieved from https://doi.org/10.5121/ijite.2023.12202

Frith, K. H. (2023). ChatGPT: Disruptive Educational Technology. Nursing Education Perspectives, 44(3), 198–199. Retrieved from https://doi.org/10.1097/01.nep.0000000000001129

Generative Artificial Intelligence. (2023). In *Washington, DC: World Bank eBooks*. https://doi.org/10.1596/39959

G, S. (2023, September 20). *AI detector by SciSpace — Catch ChatGPT text in seconds*. SciSpace Resources. https://typeset.io/resources/ai-detector-academic-writing/

Grammarly: free AI writing assistance. (n.d.). https://app.grammarly.com/

How Grammarly's Performance Reports Make You a Stronger writer | Grammarly Spotlight. (2020, May 27). Grammarly Spotlight: How to Use Grammarly's Performance Reports | Grammarly. https://www.grammarly.com/blog/grammarly-performance-report/

Joshi, I., Budhiraja, R., Dev, H., Kadia, J., Ataullah, M. O., Mitra, S., Kumar, D., & Akolekar, H. D. (2023). ChatGPT in the Classroom: An analysis of its strengths and weaknesses for solving undergraduate computer science questions. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2304.14993

Kontovas, C. (2024, April 30). *Christos Kontovas on LinkedIn: #junk #papers #chatgpt #common #phrases #publish #academia #delve #into. . .* https://www.linkedin.com/posts/xkontovas_junk-papers-chatgpt-activity-7191194304441143296-JqG_?utm_source=share&utm_medium=member_desktop

Kulcsár, G., et al., (2022b). Natural Language understanding of systems engineering artifacts. *INCOSE International Symposium*, *32*(1), 1373–1387. https://doi.org/10.1002/iis2.13001

Mehić, B. (2013). Plagiarism and self-plagiarism. Bosnian Journal of Basic Medical Sciences, 13(3), 139. https://doi.org/10.17305/bjbms.2013.2344

Meyer, L., Stadler, C., Frey, J., Radtke, N., Junghanns, K., Meissner, R., Dziwis, G., Bulert, K., & Martin, M. (2024). LLM-assisted Knowledge Graph Engineering: Experiments with ChatGPT. In *Informatik aktuell* (pp. 103–115). https://doi.org/10.1007/978-3-658-43705-3_8

Ngoc, T. N., Tran, Q. N., Tang, A., Nguyen, B., Nguyen, T., & Pham, T. (2023). AI-assisted learning for electronic engineering courses in high education. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2311.01048

OpenAI. (2024). *ChatGPT* (October 2024 version) [Large language model]. https://chat.openai.com/

Otter voice meeting notes. (n.d.). Otter Voice Meeting Notes. https://otter.ai/home

Pham, T., Nguyen, T. B., Ha, S., & Ngoc, N. T. N. (2023). Digital transformation in engineering education: Exploring the potential of AI-assisted learning. *Australasian Journal of Educational Technology*, *39*(5), 1–19. https://doi.org/10.14742/ajet.8825

Rezaev, A. V., & Tregubova, N. D. (2023). ChatGPT and AI in the Universities: An Introduction to the near future. *Vysšee Obrazovanie V Rossii*, *32*(6), 19–37. https://doi.org/10.31992/0869-3617-2023-32-6-19-37

Schönberger, M. (2023). ChatGPT in higher education: the good, the bad, and the University. *Universitat Politecnica de Valencia*, *10*.

Shaw, D., Morfeld, P., & Erren, T. (2023). The (mis)use of ChatGPT in science and education. *EMBO Reports*, *24*(7). https://doi.org/10.15252/embr.202357501

Spitzer, M. (2023). ChatGPT. *Nervenheilkunde*, *42*(04), 192–199. https://doi.org/10.1055/a-1948-8785

Su, Z., et al. (2023). Natural Language Processing for systems engineering: Automatic generation of Systems Modelling Language diagrams. Knowledge-Based Systems, 259, 110071. Retrieved from https://doi.org/10.1016/j.knosys.2022.110071

The best AI Detector for Machine Learning | GPTZERO. (n.d.). GPTZero. https://gptzero.me/machine-learning

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2023). Can ChatGPT Improve Creative Problem-Solving Performance in University Students?. PsyArXiv, Center for Open Science.

Walczak, K., & Cellary, W. (2023). Challenges for higher education in the era of widespread access to generative AI. The Poznań University of Economics Review, 9(2). Retrieved from https://doi.org/10.18559/ebr.2023.2.743

Wikipedia contributors. (2024, September 13). Systems engineering. Wikipedia. https://en.wikipedia.org/wiki/Systems_engineering

Zivcakova, L., & Wood, E. (2014). Examining instructional interventions: Encouraging academic integrity through active learning approaches. In New frontiers of educational research (pp. 191–205). https://doi.org/10.1007/978-3-642-55352-3_8

Biography

Emin Simsek. He holds a Master of Science in Innovation and Technology Management with a specialization in Systems Engineering from the University of Southeastern Norway. He earned his bachelor's degree in business administration from Ankara Yildirim Beyazit University, where he distinguished himself academically and participated in an international exchange program at Mondragon Uniberstitae. His research focuses on integrating emerging technologies with systems engineering, as exemplified by his master's thesis, "Effects of Using Generative AI in Systems Engineering." Emin's interdisciplinary background is further enriched by professional experience as an Audit Associate at PwC and as a Retail Operations Associate at Inditex Norge, experiences that

inform his commitment to bridging theoretical research and practical application.

Gerrit Muller. worked from 1980 until 1999 in the industry at Philips Healthcare and ASML. Since 1999, he has worked in research at Philips Research, the Embedded Systems Institute, and TNO in Eindhoven. He received his doctorate in 2004. In January 2008, he became a full professor of systems engineering at University of South-Eastern Norway in Kongsberg (USN), Norway. He continues to work at TNO in a part-time position. Since 2020, he is INCOSE Fellow and Excellent Educator at USN.

Kristin Falk. has lead technology teams in start-ups, SME and large corporations primarily in the energy industry. She has been in the industry for more than twenty years, where she has introduced complex systems for integrated operations, subsea processing, and automated drilling. She is currently working with Systems Engineering at the University of South-Eastern Norway. Her research focus is 'how to create systems fit for purpose in a volatile, uncertain, complex, and ambiguous world'.