

# CHATGPT AS A CO-PILOT FOR ASSESSMENT DESIGN REFINEMENT: AN EXPLORATORY STUDY

Faouzi Kamoun

ESPRIT School of Engineering, ESPRIT School of Business

Aymen Ben Brik, Ibtihel Rebhi, Salsabil Besbes, Heni Abidi, Asma Baghdadi, Rym  
Ammar

ESPRIT School of Business

## ABSTRACT

The evaluation of students' attainment of course learning outcomes is a fundamental aspect of a successful engineering program, exemplified by CDIO Standard 11 'Learning Assessment'. However, earlier research has identified a prevailing gap in assessment competency among faculty. Rooted in established interdisciplinary concepts and theories, this study aims to explore the usage of ChatGPT-4 as a co-pilot to guide faculty in assessment design refinement. To achieve this goal, we adopt a conversational analysis approach, contextualizing our study within the settings of the final exam of the senior course "Wireless Sensor Networks", offered at ESPRIT School of Engineering. We propose a framework to guide the implementation of the conversational analysis method. Our research results illustrate the merits, potentials, and limitations of using ChatGPT as a co-pilot to assist faculty in refining the assessment design process. It also brings into evidence the importance of keeping a 'human in the loop' perspective during the faculty-ChatGPT assessment co-creation activities. Our study can pave the way for further research on other potential applications of "Human-AI co-creation" and augmented man-machine intelligence in a CDIO engineering education.

## KEYWORDS

Learning assessment, CDIO, Generative AI, ChatGPT, Engineering education, co-creation, Standards: 10, 11.

## INTRODUCTION

Assessing the extent to which engineering students achieve their program learning outcomes upon graduation is of paramount importance for educational effectiveness, continuous improvement and for meeting accreditation requirements (Ali, 2023). Program Learning Outcomes (PLOs) are generally expressed as broad statements that capture the cumulative disciplinary knowledge, skills, and abilities that engineering students should demonstrate upon completing their educational program. For assessment, PLOs are often mapped to individual courses within the program and the PLOs-Courses mapping matrix showcases how individual courses, collectively, contribute towards achieving the overarching goals set by the program learning outcomes. Unlike the broader PLOs, Course Learning Outcomes (CLOs) are specific to individual classes, focusing on personal and interpersonal skills, knowledge areas, and product, process, system, and service building skills (CDIO, 2020) that students are expected to master within a course, and these contribute towards the fulfillment of the program's broader educational aims.

The design of effective learning assessment instruments to probe students' achievement of course learning outcomes is among the key pillars of effective engineering programs, as reflected for instance by CDIO standard 11 (CDIO, 2020).

Effective course learning assessment goes beyond evaluating students' competencies and grading their academic performance. Its primary added value lies in guiding instructors to continuously refine their teaching material and methods, thus providing a 360-degree loop of critical feedback to continuously improve students' learning and faculty teaching. For example, if an assessment shows that students are struggling with a specific learning outcome, it incites the instructor to revisit the related teaching material and approaches, providing a potentially better learning opportunity for the next cohort.

Despite the importance of designing effective learning assessments in engineering education, earlier research (e.g. Palm, 2019; Villarroel et al., 2018; Fook & Sidhu, 2010) revealed that not all faculty members possess the required competencies in designing authentic learning assessments that are aligned with the stated course objectives and learning outcomes. While the challenges of designing authentic assessments that are valid, reliable, and engaging are well documented (Heritage, 2022), some faculty are still struggling with developing appropriate assessment instructions and establishing clear scoring rubrics. In fact, while some faculty may demonstrate superior performance as subject matter experts, they sometimes lack the necessary knowledge and skills to effectively design authentic assessments. Guskey (2015) argued that faculty professional development programs often privileged teaching methods and technology integration over assessment practices. In its coverage of the actions needed to enhance faculty teaching competence, CDIO standard 10 (CDIO, 2020) emphasizes "assessment of student learning" as an integral teaching competence and highlights the need for faculty development in this area.

For the past years, there has been a growing interest in the application of Generative AI (GAI) and AI-driven conversational models to enhance students' learning and faculty teaching. Previous research has suggested that AI chatbots can effectively assist faculty in creating exam questions and grading students' responses (see, for example, Cheung et al., 2023; Daun & Brings, 2023; Grévisse, 2023; Skrabut, 2023, Chapter 4). However, these studies have mainly focused on showcasing how intelligent natural language processing tools can assist faculty in proposing practice exercises and a variety of assessments, including interactive quizzes, multiple choice exam questions, and fill-in-the-blank test questions. Such an approach tends to favor the passive consumption of information among faculty, without enhancing their creativity and critical thinking skills in designing effective assessment instruments. In addition, earlier studies were mainly focused on leveraging GAI for assessment creation, primarily to liberate faculty time for what is perceived as “more critical” educational endeavors such as preparing course material and providing feedback to students (Skrabut, 2023, Chapter 4).

Drawing on established interdisciplinary concepts and approaches like Human-AI Interaction (HAI) and Augmented Intelligence (Aul), as well as educational theories and approaches including co-creation, constructivism, and connectivism, this study seeks to explore the potential usage of GAI as a mentor to guide faculty in refining the assessment design process. We position the usage of GAI as co-pilot for assessment design as another potential initiative to further bridge the prevailing gap in assessment competency among faculty. To achieve this objective, we adopt a qualitative research methodology that uses a conversational analysis approach, taking ChatGPT-4 as an illustrative example. We contextualized our study within the setting of the senior engineering course “Wireless Sensor Networks” taught at ESPRIT School of Engineering.

In this paper, the term ‘co-pilot’ is used metaphorically to describe the role of ChatGPT in assisting faculty with refining the assessment design. This usage should not be confused with Microsoft’s Co-Pilot, a GAI tool launched in 2023.

The remainder of this paper is organized as follows: Section 2 presents a literature review of related studies. Section 3 details the research methods and procedures. Section 4 presents the results of our study, while section 5 provides a detailed discussion of these findings. Finally, section 6 presents a summary of the key research results, their implications, and some suggestions for future research.

## **LITERATURE REVIEW AND RESEARCH CONTRIBUTIONS**

AI is poised to transform teaching and learning in higher education, with applications ranging from profiling and prediction, intelligent tutoring systems, assessment and evaluation, and adaptive systems and personalization (Bond et al., 2023). Bond et al. (2023) conducted a meta systematic review of the applications of AI in higher education. The study identified several earlier contributions addressing the usage of GAI in assessment and evaluation. These earlier studies were however mainly focused on automating and/or enhancing the evaluative aspects associated with automated grading (Algabri et al., 2021), and personalized feedback to students (Ouyang et al., 2022).

Table 1 summarizes some of the related earlier contributions.

Table 1. Summary of Related Contributions

Source	Research Method	Key Findings
Fergus et al. (2023)	Case study	While ChatGPT can generate good responses for some chemistry questions, it has limitations when it comes to application and interpretation questions, especially those involving non-text information.
Gamage et al. (2023)	Literature survey	While ChatGPT can be useful in providing personalized feedback, support for research and data analysis, and automated administrative services, it poses some challenges associated with academic integrity, reliability issues, and the potential for biases and falsified information.
Chaudhry et al. (2023)	Case study	ChatGPT can provide solutions to the assessment tools used for assessing students' learning outcomes.
Kolade et al. (2023)	Case study	ChatGPT can generate original academic essays with high-quality but has limitations in providing concise referencing.
Rybak & Tytenko (2023)	Conceptual study	A high-level conceptual model was developed to guide ChatGPT generate tests and quizzes. The proposed model has not been tested or validated
Daun & Brings (2023)	Case study	ChatGPT can be used for the automated assessment of coding exercises, and it can create sample software engineering exam exercises, hence assisting faculty to cope with the tight time constraints in searching for potential assessment questions.
Nikolic et al. (2023)	Case study	Cross-examining ChatGPT responses to existing assessment prompts from ten subjects showed that ChatGPT did pass some subjects and excelled with some assessment types. Careful finetuning of the input prompts can help in enhancing the reliability of ChatGPT responses.

As may be seen, most of the reported studies on using GAI for assessment (see also the work of Meo et al., 2023; Friederichs et al., 2023; Fergus et al., 2023) have focused on evaluating the reliability of AI-driven chatbots in correctly answering quizzes, and exam questions. Fewer studies have focused on using GAI for assessment design and creation. More importantly, studies like those reported in (Skrabut, 2023, chapter 4; Cheung et al., 2023; Grévisse, 2023), mainly viewed ChatGPT as a time-saver for faculty, having the potential to quickly generate a test bank of questions (predominantly multiple choices) and answers in few seconds. Not only does such an approach undermines the role of assessment in higher education, but it also hinders critical thinking among faculty by stimulating passive consumption of information.

This research makes the following two main contributions:

- Based on established interdisciplinary concepts, theories, and approaches that are detailed later in this section, our study explores the usage of GAI as a value-added tool and a mentor to guide faculty in refining the assessment design process. In doing so, it emphasizes the importance of human-AI co-creation in bridging the gap in assessment competency. Instead of promoting the passive consumption of information, the proposed approach aims to stimulate some higher-order thinking skills among faculty when they actively engage in meaningful discussions and collaborative conversations with AI-driven chatbots.
- We support this exploratory study with a qualitative research methodology, which employs a systematic and iterative conversational analysis approach to mimic the interaction with a professional educational expert.

Our research is rooted in and guided by the following established concepts, theories, and frameworks:

- Augmented Intelligence (Aul) and co-creation (Dellermann et al., 2021; Xue et al., 2022): Aul is a subset of AI that emphasizes a collaborative model where humans and AI work together to enhance cognitive performance, problem solving and decision making through a co-creation process. In doing so, people and AI work iteratively to refine tasks, goals, and solutions. Aul can also enable humans to acquire new ideas and perspectives that they might not have considered on their own. It also promotes the "human in the loop" design principle to emphasize human oversight, ensuring that AI enhances rather than substitutes human capabilities.
- Human-AI Interaction (HAI) framework (Abedin et al., 2022; Rezwana & Maher, 2023; Sreedharan, 2023): HAI refers to the dynamic interaction between humans and AI systems. HAI is crucial to AI as it shapes how humans engage with AI in an effective and efficient manner. In the context of our study, HAI provides some useful guidelines that we followed to optimize our interactions with ChatGPT during the assessment co-design experiments. These include keeping clear, specific, and engaging prompts, avoiding information overloading, providing feedback, and seeking explanations for the generated responses, cross-examining and critiquing the validity and accuracy of the generated responses, finetuning and revectoring the conversations when needed to make them useful, usable, and productive. To this regard, Ng et al. (2023) highlighted the growing need for faculty to equip themselves with adequate digital competencies to make best usage of AI in their teaching, learning and assessment. Understanding the distinction among GAI (AI algorithms capable of generating new content), Aul, and HAI, as defined above, is useful to apprehend the layered approach of our study. We employ GAI as the specific type of AI technology, we adopt HAI to set the stage for effective human-centered AI use and we draw from Aul to advocate for a "human-in-the-loop" design approach.
- Constructivism (Jonassen, 1999): Our research has been inspired by the principles of constructivism whereby faculty actively construct knowledge through meaningful interaction with ChatGPT to co-design assessments. In doing so, it relies on the important concepts of active engagement, iterative interactions, and dynamic feedback. Some earlier studies have underscored ChatGPT's role in fostering constructivism learning by enabling learners to actively negotiate meaning, co-construct knowledge, engage in self-reflective metacognition activities, and participate in dynamic scaffolded learning activities (Rasul et al., 2023; Hatmanto & Sari, 2023; Tabib & Alrabeei, 2024).
- Connectivism (Siemens, 2004): The collaborative nature of the faculty-ChatGPT interaction in the co-design of assessments is in alignment with the connectivism principle that knowledge networks and collective intelligence facilitated by ChatGPT play an important role in stimulating a dynamic learning process.

## METHODS AND PROCEDURES

### *Methods*

As an exploratory study for the potential usage of AI-driven conversational models as co-pilots for assessment design refinement, we opted for a qualitative research methodology that uses a Conversational Analysis (CA) approach. Conversational Analysis (CA) is an approach developed within the field of sociology that analyzes the mechanisms and patterns of human interactions in natural conversations. It emphasizes the role of talk-in-interaction and how conversational turns are constructed among participants to achieve mutual understanding (Sidnell & Stivers, 2013).

We adapted the CA approach from its traditional social science conversational context, towards a human-AI perspective, with ChatGPT mimicking the interaction with a professional educational expert. The CA approach fits well with the aims of this study for the following two main reasons:

- It incarnates the iterative nature of the conversation that takes place between the researcher and ChatGPT, reflecting the adaptation and learning that take place during the conversation. In fact, the CA approach captures the evolvement of ChatGPT's responses to the researcher's prompts over time, based on the follow-up prompts, requests for refinement and clarifications, and ChatGPT's learning from previous interactions.
- It can showcase how ChatGPT handles misunderstandings, disagreements, or incorrect information, and how it corrects itself.

We contextualized our study within the setting of the first-year graduate engineering course "Wireless Sensor Networks", taught at ESPRIT School of Engineering. This course aims to introduce students to the main concepts and challenges associated with the design and deployment of Wireless Sensor Networks (WSNs) and the applications of these networks in the context of smart-cities and Internet of Things (IoT).

The course is assessed using the following instruments:

- Classwork: Presentation & WSN simulation labs (30%)
- Team project (20%): Analysis, design and implementation of a WSN solution
- Final exam (50%)

The course has the following CLOs:

*CLO.1.* Explain the basic concepts and applications of wireless sensor networks.

*CLO.2.* Describe wireless sensor network types, components, operations, challenges, and sensor node architectures.

*CLO.3.* Demonstrate understanding and usage of WSN standards and protocols (physical, data link, network, and application layers).

*CLO.4.* Appraise the usage of WSN simulations.

*CLO.5.* Investigate potential deployment of WSN in the context of smart cities/ IoT.

*CLO.6.* Conceive, design and implement a complete WSN solution.

To keep the scope of this study to a manageable level, we focused on the final exam as an illustrative assessment instrument, hence excluding the classwork and team project.

The final exam has been designed primarily to evaluate students' understanding of key WSN concepts and principles (CLOs 2 & 3). It also included an element that assesses students' ability to conceptualize and design a small-scale WSN solution (CLO6). The original exam developed by the faculty consisted of 7 essay-type questions and one case-study. It is a closed-book exam, set for a duration of 1.5 hours.

### **Procedures**

The general framework that guided our conversational analysis approach is depicted in Figure 1. The framework has been developed based on thorough desktop research on best practices in interacting with AI-driven conversational models and it has been guided by the concepts, theories and frameworks described in the previous section. We adopted a systematic and iterative conversational approach to mimic the interaction with a professional educational expert. In addition, recognizing that well-defined ChatGPT prompts are the recipe for a successful conversation, we took special care to ensure that each prompt is clear, concise, specific, and focused. Further, to avoid overwhelming ChatGPT with lengthy details, we decided to address each exam question individually.

As may be seen in Figure 1, we started by providing ChatGPT with the general context of the intervention, including general information about the course, its level, description, content, and CLOs among others. We also stated our goals, objectives, and expectations, with a special emphasis on conveying our aim to collaborate to co-design and iteratively refine each exam question. We prompted ChatGPT to act as a senior engineering professor and a leading pedagogical expert. We entered each final exam question, one at a time, along with its associated CLO and Bloom's cognitive level. We leveraged ChatGPT's analytical capabilities to prompt it for critiques, insights, and feedback. We regularly sought clarifications by posing follow-up questions and specifying constraints. We cross-examined the validity and the appropriateness of the generated ChatGPT responses, and iteratively refined the original questions until we achieved satisfaction with the final outcomes. Occasionally, we had to replace an exam question with a better alternative suggested by ChatGPT. We also performed a final iteration to eliminate potential overlaps among the revised questions. At the end, we prompted ChatGPT to generate the answer for each validated question, along with its grading rubric. These were also iteratively refined until validated.

Throughout our interactions, we maintained a journal to capture details of the iterative aspects of our intervention: Queries, ChatGPT's responses, critique of the responses, follow-up prompts, subsequent refinements and modifications, and the rationale behind them, among others. An extract of the journal can be found in ([Prompt\\_Journal, 2023](#)) on GitHub.

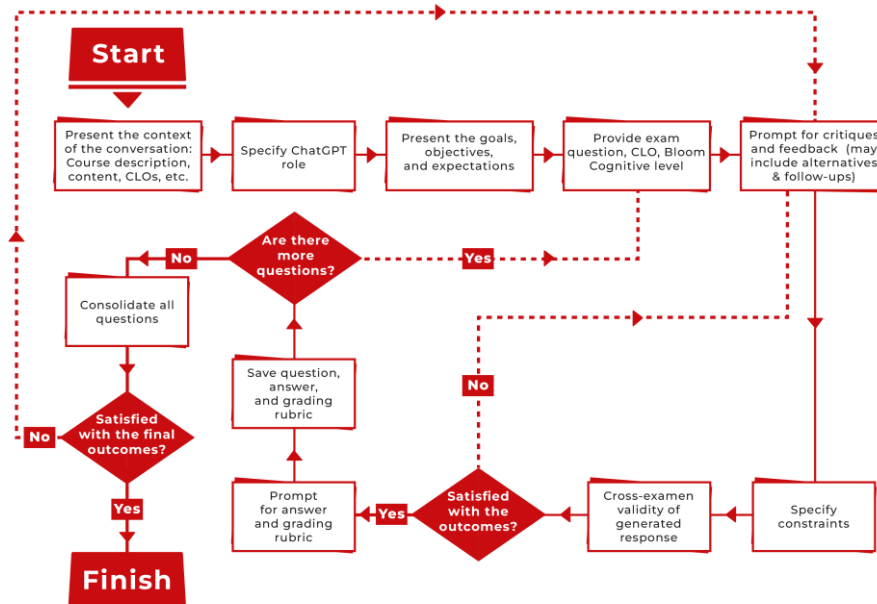


Figure 1. Framework of the Researchers' Conversation with ChatGPT

## RESULTS

Because of lack of space, we will only report herein sample of the faculty-ChatGPT interactions. The reader is referred to (Prompt\_Journal, 2023) for full details regarding the conversations that took place to refine the other 7 exam questions.

 You:

Your role: Act as {senior Engineering Professor and a leading pedagogical expert} . As a {co-pilot}, I would like you to collaborate with me (the instructor) to {co-design} and {refine} each proposed exam question, one at a time. Just say OK.

 ChatGPT:

OK.

 You:

Your task: I would like you to {critique, refine, and provide constructive feedback} on each submitted final exam question, taking into account the corresponding CLO, Bloom's cognitive level, and the appropriateness of the question in terms of clarity, specificity, and alignment with best assessment practices. Are you ready to start with the first question?

 ChatGPT:

Yes, I'm ready to start with the first question. Please go ahead and share it.

 You:

The first question Q1: {What kinds of design and resource constraints does a wireless sensor node have?}. This question is meant to assess CLO.2. It is associated with Bloom's cognitive level of {understanding}.

 ChatGPT:



Q1 Critique and Refinement:

Original Question: "What kinds of design and resource constraints does a wireless sensor node have?"

Critique:

Clarity and Specificity: The question is somewhat open-ended, which might lead to varied and possibly incomplete responses. It doesn't specify the context or type of wireless sensor network, which can have different constraints based on their applications.

Alignment with CLO.2: This question aligns well with CLO.2 as it explores the understanding of components and operations of WSNs, but it could be more targeted to extract specific knowledge relevant to sensor node architectures.

Bloom's Cognitive Level: It currently addresses the 'Understanding' level appropriately but could be enhanced to touch upon 'Application' by asking for real-world examples or implications.

Refinement: To make this question more effective, consider adding a specific context or application. For instance:

Revised Question: "Considering a wireless sensor network deployed in a smart city environment, identify and describe the design and resource constraints typically faced by a sensor node in this context. How do these constraints impact the architecture and functioning of the sensor node?"

This revision:

Provides a clear context (smart city), prompting students to think about specific applications.

Encourages students not only to list constraints but also to discuss their impact, deepening their understanding and application of the concepts.

Are you ready for the next question or is there anything else to refine in this one?



You:

Thank you. Can you suggest a modification to question Q1 to target the same objective but reframe it indirectly in the context of an application, targeting a higher-order cognitive level?



ChatGPT:

Certainly! To reframe question Q1 in the context of an application and target a higher-order cognitive level, we can shift the focus to evaluating or creating scenarios based on the given constraints. This approach will elevate the cognitive demand from merely understanding to analyzing, evaluating, or even creating, which are higher levels in Bloom's Taxonomy.

Modified Question:

"Imagine you are tasked with designing a wireless sensor network for monitoring environmental conditions in an agricultural setting. Describe the specific design and resource constraints you would need to consider for the sensor nodes in this application. Additionally, propose a solution on how you would address one of these constraints in your design. Justify your choices based on the operational requirements and limitations of the sensor nodes."

This modification:

Places students in a real-world scenario, requiring them to apply their knowledge in a practical context (higher-order thinking).

Asks for a description of constraints (Analysis), a proposal for a solution to one of these constraints (Synthesis/Creativity), and justification of their choices (Evaluation).

Ensures that students not only understand the concepts but also can apply and evaluate them in a specific application, aligning with CLO.2 and targeting a higher cognitive level.

Would you like to proceed with refining another question, or is there anything else to adjust in this one?

## DISCUSSIONS

After engaging in dialogue with ChatGPT to assess and refine the proposed exam questions, we came to the following main observations:

- ChatGPT was extremely useful in critiquing our proposed questions and in suggesting viable alternatives. ChatGPT proved highly valuable in transforming straightforward questions, initially set at the comprehension cognitive level, into more challenging ones aimed at higher cognitive levels, such as application-type tasks.
- ChatGPT did mimic a skilled academic mentor by providing sound, clear and specific justifications when providing feedback and suggesting new alternatives.
- ChatGPT's answers to some of our proposed exam questions were very enlightening and these enabled us to enrich the course's content.
- When prompted to integrate aspects related to the United Nations Sustainable Development Goals (UN SDGs) into the case study question, ChatGPT provided an interesting and well justified proposal. The same applied when we prompted ChatGPT to propose exam questions that are aligned with one of the CDIO standards (e.g. last page in [Prompt\\_Journal, 2023](#)). It should be noted that while UN SDGs are not explicitly reflected in the current WSN course learning outcomes, our intention to embed SDG concepts into the case study question stems from the broader program educational aim to promote interdisciplinary learning and to highlight the relevance of technological innovations like WSNs in addressing pressing sustainability challenges. This also aligns with CDIO Optional Standard 1 on Sustainable development (CDIO, 2022).
- ChatGPT performed very poorly when analyzing a picture associated with an exam question. Refer to the sample case on page 8 in ([Prompt\\_Journal, 2023](#)). A similar very mediocre performance was also noticed when suggesting a graphical solution (ibid, page 18). In particular, the image generated by ChatGPT as an answer to one of our proposed exam questions was quite unusual and unexpected.

As faculty, we found the conversations with ChatGPT very meaningful, interesting, and informing. These conversations also stimulated our curiosity, as well as our analytical, critical, and creative thinking. We also recognized the necessity of maintaining a 'human in the loop' perspective during our interactions with ChatGPT, given the occasional inaccuracies in the automated responses. This human insight perspective is crucial for several reasons: First, it acknowledges the fact that ChatGPT can make mistakes or generate inappropriate responses. Second, it recognizes that when ChatGPT deviates from the context of the assigned task, human intervention is needed to guide it back on track, ensuring relevance of the response. Third, our feedback can help ChatGPT to continuously improve its training model and hence enhance its reliability.

Our research findings provide evidence on the merits of the collaboration between humans and AI in co-creation and new idea generations, as suggested by the Aul and co-creation frameworks (Dellermann et al., 2021; Xue et al., 2022). Our research also illustrates how knowledge can be constructed during the exam co-design collaboration with ChatGPT through active engagement, iterative interactions, and feedback. This process aligns with the principles of constructivism (Jonassen, 1999).

As mentioned by Kamoun et al. (2023), these meaningful engagements between the human and the AI-driven conversational model are most likely to become life-long learning competencies for faculty.

## CONCLUSION AND SUGGESTIONS FOR FUTURE RESEARCH

This study was motivated by earlier research findings that highlighted the prevailing gap in assessment competency among faculty and the need for creative and effective solutions to address this shortcoming. It adopted a conversational analysis approach, within the context of a senior engineering course's final exam, to showcase the potentials and limitations of using ChatGPT-4 as a co-pilot in assisting faculty to refine the assessment design process.

Unlike most related previous studies that promoted the passive consumption of information among faculty by simply soliciting ChatGPT to propose assessment questions, our approach promotes collaboration and co-creation. It uses ChatGPT as a co-pilot for assessment design refinement, while keeping a 'human in the loop' perspective.

Our research contributes to advancing the body of knowledge in engineering education by providing evidence that, when properly orchestrated, active engagement and meaningful interactive conversations with GAI systems can actively promote higher order thinking skills among faculty, enabling them to collaborate with GAI chatbots to co-create assessment tasks that reflect better creativity and enhanced effectiveness. This can pave the way for further research on other potential applications of "Human-AI co-creation" and augmented man-machine intelligence in a CDIO engineering education.

This study has several limitations: First it is of limited scope, being focused on a single course, which also makes the results not necessarily generalizable to other educational contexts. Second, being focused exclusively on the final exam, our study does not consider other forms of assessment instruments such as projects. Third, our study does not incorporate student feedback or perspectives on the outcomes of the co-created assessment.

Research is underway to address some of the limitations mentioned above.

## FINANCIAL SUPPORT ACKNOWLEDGEMENTS

*The authors received no financial support for this work.*

## REFERENCES

- Abedin, B., Meske, C., Junglas, I., Rabhi, F. & Motahari-Nezhad, H.R. (2022). Designing and Managing Human-AI interactions. *Information Systems Frontiers*, 24, (pp.691–697).  
<https://doi.org/10.1007/s10796-022-10313-1>
- Algabri, H. K., Kharade, K. G., & Kamat, R. K. (2021). Promise, Threats, And Personalization In Higher Education with Artificial Intelligence. *Webology*, 18(6), (pp. 2129–2139).
- Ali, Q. I. (2023). Surveying Different Student Outcome Assessment Methods for ABET Accredited Computer Engineering Programs. *Research Reports on Computer Science*, 2(1), (pp. 56–76).
- Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S.W., & Siemens, G. (2023). A Meta Systematic Review of Artificial Intelligence in Higher Education: A Call for Increased Ethics, Collaboration, and Rigour. Pre-print.
- CDIO (2020). *CDIO Standards 3.0*, 10 June 2020. Available at  
<http://cdio.org/files/CDIO%20STANDARDS%203.pdf>.
- CDIO. (2022). *CDIO Optional Standards 3.0*. 12 July 2022. Available at  
<http://www.cdio.org/content/cdio-optional-standards-30>

- Chaudhry, I. S., Sarwary, S. A. M., El Refae, G. A., & Chabchoub, H. (2023). Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector in a New Era of ChatGPT—A Case Study. *Cogent Education*, 10(1), 2210461. <https://doi.org/10.1080/2331186X.2023.2210461>
- Cheung, B. H. H., Lau, G. K. K., Wong, G. T. C., Lee, E. Y. P., Kulkarni, D., Seow, C. S., ... & Co, M. T. H. (2023). ChatGPT Versus Human in Generating Medical Graduate Exam Multiple Choice Questions—A Multinational Prospective Ptudy (Hong Kong SAR, Singapore, Ireland, and the United Kingdom). *PLoS One*, 18(8), e0290691. <https://doi.org/10.1371/journal.pone.0290691>
- Daun, M., & Brings, J. (2023). How ChatGPT will change software engineering education. *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1* (pp. 110-116).
- Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., & Ebel, P. (2021). The Future of Human-AI Collaboration: a Taxonomy of Design Knowledge for Hybrid Intelligence Systems. arXiv preprint arXiv:2105.03354. <https://doi.org/10.48550/arXiv.2105.03354>
- Fergus, S., Botha, M., & Ostovar, M. (2023). Evaluating Academic Answers Generated Using ChatGPT. *Journal of Chemical Education*, 100(4), (pp. 1672-1675). <https://doi.org/10.1021/acs.jchemed.3c00087>
- Fook, C. Y., & Sidhu, G. K. (2010). Authentic Assessment and Pedagogical Strategies in Higher Education. *Journal of Social Sciences*, 6(2), (pp. 153-161). <https://doi.org/10.3844/jssp.2010.153.161>
- Friederichs, H., Friederichs, W. J., & März, M. (2023). ChatGPT in Medical School: How Successful is AI in Progress Testing?. *Medical Education Online*, 28(1), 2220920.
- Gamage, K. A., Dehideniya, S. C., Xu, Z., & Tang, X. (2023). ChatGPT and Higher Education Assessments: More Opportunities Than Concerns?. *Journal of Applied Learning and Teaching*, 6(2).
- Grévisse, C. (2023). Comparative Quality Analysis of GPT-Based Multiple Choice Question Generation. In *International Conference on Applied Informatics* (pp. 435-447). Cham: Springer Nature Switzerland.
- Guskey, T. R. (2015). *On Your Mark: Challenging the Conventions of Grading and Reporting*. Solution Tree Press.
- Hatmanto, E. D., & Sari, M. I. (2023). Aligning Theory and Practice: Leveraging Chat GPT for Effective English Language Teaching and Learning. *E3S Web of Conferences* (Vol. 440, pp. 1-18). EDP Sciences. <https://doi.org/10.1051/e3sconf/202344005001>.
- Heritage, M. (2022). *Formative Assessment: Making It Happen in the Classroom* (2nd ed.). Corwin Press. <https://doi.org/10.4135/9781452219493>
- Jonassen, D. H. (1999). Designing Constructivist Learning Environments. In C. M. Reigeluth (Ed.), *Instructional-design Theories and Models: A New Paradigm of Instructional Theory* (2<sup>nd</sup> ed., pp. 215–239). NY: Lawrence Erlbaum Associates.
- Kamoun, F., El Ayeb, W., Jabri, I., Sifi, S., & Iqbal, F. (2023). Knowledge, Attitude, and Perception Towards ChatGPT Among University Students and Faculty: A Preliminary Exploration. *The 12th European Conference on Education (ECE2024)*. London. UK. (pp. 247–266).
- Kolade, O., Owoseni, A., & Egbetokun, A. (2023). Assessment on Trial? ChatGPT and the New Frontiers of Learning and Assessment in Higher Education. 01-22. Retrieved from: [https://www.researchgate.net/profile/Oluwaseun-Kolade/publication/369974458\\_Assessment\\_on\\_trial\\_ChatGPT\\_and\\_the\\_new\\_frontiers\\_of\\_learning\\_and\\_assessment\\_in\\_higher\\_education/links/64376dc720f25554da29ad9f/Assessment-on-trial-ChatGPT-and-the-new-frontiers-of-learning-and-assessment-in-higher-education.pdf](https://www.researchgate.net/profile/Oluwaseun-Kolade/publication/369974458_Assessment_on_trial_ChatGPT_and_the_new_frontiers_of_learning_and_assessment_in_higher_education/links/64376dc720f25554da29ad9f/Assessment-on-trial-ChatGPT-and-the-new-frontiers-of-learning-and-assessment-in-higher-education.pdf) (accessed on 15 December 2023).
- Meo, S. A., Al-Masri, A. A., Alotaibi, M., Meo, M. Z. S., & Meo, M. O. S. (2023). ChatGPT Knowledge Evaluation in Basic and Clinical Medical Sciences: Multiple Choice Question Examination-based Performance. *Healthcare*, 11(4), (p. 2046). <https://doi.org/10.3390/healthcare11142046>
- Ng, D.T.K., Leung, J.K.L., Su, J. et al. (2023). Teachers' AI Digital Competencies and Twenty-first Century Skills in the Post-pandemic World. *Education Tech Research Dev*, 71, (pp. 137–161). <https://doi.org/10.1007/s11423-023-10203-6>

- Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G. M., Grundy, S., ... & Sandison, C. (2023). ChatGPT Versus Engineering Education Assessment: A Multidisciplinary and Multi-institutional Benchmarking and Analysis of this Generative Artificial Intelligence Tool to Investigate Assessment Integrity. *European Journal of Engineering Education*, 48(4), (pp. 559–614). <https://doi.org/10.1080/03043797.2023.2213169>
- Ouyang F, Zheng L., & Jiao P (2022). Artificial Intelligence in Online Higher Education: A Systematic Review of Empirical Research from 2011 to 2020. *Education and Information Technologies*, 27(6), (pp. 7893-7925). <https://doi.org/10.1007/s10639-022-10925-9>.
- Palm, T. (2019). Performance Assessment and Authentic Assessment: A Conceptual Analysis of the Literature. *Practical Assessment, Research, and Evaluation*, 13(1). <https://doi.org/10.7275/0qpc-ws45>
- Prompt\_Journal (2023). [https://github.com/AlforU/Prompt\\_Journal/blob/main/Prompt%20Journal.pdf](https://github.com/AlforU/Prompt_Journal/blob/main/Prompt%20Journal.pdf)
- Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ... & Heathcote, L. (2023). The Role of ChatGPT in Higher Education: Benefits, Challenges, and Future Research Directions. *Journal of Applied Learning and Teaching*, 6(1), (pp. 1-16). <https://doi.org/10.37074/jalt.2023.6.1.29>
- Rezwana, J., & Maher, M. L. (2023). Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-creative Systems. *ACM Transactions on Computer-Human Interaction*, 30(5), (pp. 1-28). <https://doi.org/10.1145/3519026>
- Rybak, M., & Tytenko, S. (2023). Leveraging ChatGpt for Educational Text Analysis Aiming Assessment Generation. *Modern Engineering and Innovative Technologies*, (29-01), (pp. 112-116). <https://doi.org/10.30890/2567-5273.2023-29-01-045>
- Sidnell, J., & Stivers, T. (Eds.). (2013). *The Handbook of Conversation Analysis*. Wiley-Blackwell.
- Siemens, G. (2004). Connectivism: A Learning Theory for the Digital Age. *International Journal of Instructional Technology and Distance Learning*, 2(1), (pp. 3-10). <https://doi.org/10.3109/0142159X.2016.1173661>
- Skrabut, S. (2023). *80 Ways to Use ChatGPT in the Classroom: Using AI to Enhance Teaching and Learning*. Stan Skrabut (February 10, 2023).
- Sreedharan, S. (2023). Human-Aware AI—A Foundational Framework for Human-AI Interaction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 37 (13). <https://doi.org/10.1609/aaai.v37i13.26822>
- Tabib, F.M., & Alrabeei, M.M. (2024). Can Guided ChatGPT Use Enhance Students' Cognitive and Metacognitive Skills?. In: Al-Marzouqi, A., Salloum, S.A., Al-Saidat, M., Aburayya, A., Gupta, B. (eds) *Artificial Intelligence in Education: The Power and Dangers of ChatGPT in the Classroom* (pp. 143-154). *Studies in Big Data*, vol 144. Springer, Cham. [https://doi.org/10.1007/978-3-031-52280-2\\_10](https://doi.org/10.1007/978-3-031-52280-2_10)
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic Assessment: Creating a Blueprint for Course Design. *Assessment & Evaluation in Higher Education*, 43(5), (pp. 840-854). <https://doi.org/10.1080/02602938.2017.1412396>
- Xue, J., Hu, B., Li, L., & Zhang, J. (2022). Human—machine Augmented Intelligence: Research and Applications. *Frontiers of Information Technology & Electronic Engineering*, 23(8), (pp. 1139-1141).

## BIOGRAPHICAL INFORMATION

**Faouzi Kamoun** is a Professor, Director of the Research Office at ESPRIT, and the Deputy General Manager of ESPRIT School of Business. His most recent research focuses on the ethical, legal, and social implications of emerging technologies. Recently, he has been involved in projects exploring the impact of Generative AI models on cybersecurity as well as on students' learning and faculty's teaching, and assessment.

**Aymen Ben Brik** is an Assistant Professor and the head of the IMA department at Esprit School of Business. He earned his MSc degree in Arithmetic and Number Theory from the Faculty of Sciences of Tunis. Currently, his research focuses on applied mathematics for data science, as well as the application of Generative AI in higher education.

**Ibtihel Rebhi** is currently an Assistant Professor of Statistics at ESPRIT School of Business. She holds a Ph.D. in Applied Mathematics from the National Engineering School of Tunis. Her research interests include data science and statistical modeling applied to consumer behavior research and Engineering education. Ibtihel worked on several multidisciplinary research projects in biology, agronomy, and IT.

**Salsabil Besbes** holds a Ph.D. degree in Electrical Engineering from the National School of Engineering of Tunis. Since September 2020, she has been an Assistant Professor of computing science at Esprit School of Business. Her research interests include machine learning, artificial intelligence, pedagogical innovation, and digital signal processing.

**Asma Baghdadi** holds a Ph.D. degree in Information Sciences and Technologies. Since January 2022, she has been an Assistant Professor at Esprit School of Business. Her research interests focus on exploring the nexus between information technologies and healthcare. She is currently working on research projects related to the applications of GAI in higher education. Asma is affiliated with REGIM & LISSI Labs, as well as the AI4U research group.

**Heni Abidi** is an Assistant Professor of Mathematics at ESPRIT School of Business. He holds a Ph.D. in Applied Mathematics from the Faculty of Science in Tunis. His research interests include AI applications in Higher Education, stochastic analysis, optimal control, and numerical approximation of stochastic differential equations with applications in finance.

**Rym Ammar** holds a Ph.D. in Management from the University of Sousse. She is currently as Assistant Professor at ESPRIT School of Business. As member of the IA4U research team, Rym has recently been involved in several multidisciplinary research projects within the applied domains of AI in education, urban planning, and tourism.

### Corresponding author

Faouzi Kamoun  
Ecole Supérieure Privée d'Ingénierie et de  
Technologies (ESPRIT)  
Lot No. 46, Z.I. Chotrana II,  
Ariana, Tunisia, 2088.  
[faouzi.kammoun@esprit.tn](mailto:faouzi.kammoun@esprit.tn)



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).