Project Title

AI-based teaching for complex knowledge and communication skills (AI-Teacher)

Full Name of the Team Leader(s)

Jan C. Zoellick (Charité), Yannik Müller (Charité), Nicolas Bioul (UCLouvain), Benoit Macq (UCLouvain), Arthur Pisvin (UCLouvain), Viktoria Nagy (Circle U), Jean-François Rees (Biology)

Corresponding leaders: jan.zoellick@charite.de, nicholas.bioul@uclouvain.be

Profile of the Team Leader(s)

Jan Zoellick is a psychologist and researcher in the field of digital health. He is PI for methods and evaluation in several projects testing telehealth applications in realistic settings using randomised controlled trials. His previous work focused on using clinical decision support systems in medical education (Kafke et al., 2023). He is also an experienced lecturer in more than 250 teaching units in the past eight years. Jan is active in Circle U at CU.ecr promoting postdocs. He was a visiting scholar at the Department of Population Health Sciences at King's College London in autumn 2022, and he will be a visiting lecturer at Université Paris Cité in autumn 2025; both contacts established through Circle U.

Yannik Müller is a medical student and doctoral candidate. His research focuses on new technologies for medical education. He is a member of the student council and deals with educational matters in the universities' Education Commission. In addition, he is involved as a student representative in the curriculum organisation.

The UCLouvain team combines expertise in educational innovation (Nicolas Bioul), artificial intelligence and digital pedagogy (Benoît Macq), cognitive biology (Jean-François Rees), and European academic collaboration (Viktoria Nagy). The interdisciplinary composition of the team ensures a strong foundation for both theoretical exploration and practical implementation.

Project Abstract

Teaching complex content and examining learning outcomes have been important tasks for higher education. Examples include communication models and techniques such as motivational interviewing or breaking bad news in the medical setting as well as integration and establishing connections between concepts in science such as biology. Multiple-choice questionnaires (MCQs) are undoubtedly a fast and scalable method for evaluating large cohorts of students. However, they have been widely criticized for their limitations. MCQs often assess factual recall rather than deep understanding, critical thinking, or problem-solving abilities. To address these shortcomings, Jean-François Rees and his team at UCLouvain developed a novel assessment method that combines efficiency with the capacity to evaluate higher-order thinking. This method is based on concept maps created by the instructor, in which half of the concepts and their interconnections are hidden (Guisset et al., 2019). Students are tasked with placing the missing concepts—provided in a list—into the correct positions. In doing so, they actively revisit and reinforce their understanding of both concepts and their relationships. This process can be enhanced through peer discussion, and over 80% of students reported a better grasp of the material following the evaluation.

In this project, we aim to use artificial intelligence (AI) to support and develop the learning methods further in two use cases:

- Use case 1 focuses on training medical students in communication skills in motivational interviewing and breaking bad news. We will refine a prompting pipeline for a large language model (LLM) that we have developed involving patients presenting different symptoms and needs. Students acting as physicians engage in conversations with the AI that is prompted as a patient with COPD (MI) and (2) a patient with lung cancer (breaking bad news). We will refine this pipeline with audio-visual aspects guided by concept maps as an underlying educational method.
- Use case 2 will develop an AI-assisted protocol that enables teachers to easily generate concept maps from their course content. A teacher - or a student - could prompt a text (e.g. a chapter or several chapters of the course) and be helped in the generation of a concept map that could be used for generating exercises.

Our project will thus generate valuable insights into Human-AI interaction in the context of higher education. The developed tools can be easily integrated into the curriculum of other institutions, and we will provide access to our code and the data in accordance with open science and good scientific practice guidelines.

Project Objectives

Our objective is to test the use of AI-tools in education. Specifically, we will:

- analyse the usability and efficacy of LLMs for teaching communicative skills in medical education
- develop an existing prompting pipeline further to include audio-visual components for enhanced immersion
- augment existing concept map assessments with AI-generated scaffolding, suggestions, or feedback (via RAG)
- support metacognitive skills through AI-driven CoT prompting to help students explain links and justify choices.
- create a hybrid evaluation model combining peer learning, CCàT (Cartes Conceptuelles à Trous), and AI interventions.
- assess learning gains and usability compared to traditional CCàT-only methods.
- ensure scalability for large classes and multilingual settings within Circle U.

Project Dataset

Use case 1: Working with the API of OpenAI's GPT, we have built an infrastructure and chat interface to collect the responses between the user and GPT using Flask. Our dataset consists of json files that contain socio-demographic information by the user, the interaction between the user and GPT, feedback on the interaction and its subjective evaluation by the student, as well as a list of questionnaires as outcomes, e.g., self-rated competence in communication skills or technology acceptance and commitment. We start collecting data in June 2025, and plan to have data from 100

conversation for the TRAIL workshop.
Use case 2: Concept maps have been created for all chapters and some combinations of them from an animal biology course. More of them will be created by students from a textbook such as the Campbell's Biology, which is one of the most used biology textbook in universities.

Background Information

Guisset et al. have shown that CCàT-based peer learning evaluations can enhance learning outcomes, engagement, and concept retention, even in large cohorts. However, this method still relies heavily on the instructor's design and manual knowledge scaffolding. At the same time, LLMs with RAG capabilities allow for real-time content generation grounded in trusted sources, while CoT reasoning helps elicit deeper student thinking. Combining these innovations holds promise for expanding concept-based learning and formative assessment in higher education.

Teaching communication skills is a complex task, particularly for techniques such as motivational interviewing (MI) and breaking bad news that require practice and experience. MI consists of four basic and four advanced techniques that involve affirming, reflective listening or change talk (Miller & Rollnick, 2012). For breaking bad news, the SPIKES model (setting, perception, invitation, knowledge, emotions, strategy&summary) has been demonstrated as a valuable technique (Mahendiran et al., 2023). High communication skills are essential for physicians to practice shared decision making for patient-centred care. Several studies have used self-rated competence in communication skills as outcomes in their educational studies (Bauzon et al., 2021; Lück et al., 2022).

LLMs offer a new way of learning. Particularly the role prompts enable interactive learning experiences that would otherwise need real patients, lecturers, or actors. Previous studies have used LLMs to provide medical advice for forum posts that was rated more empathetic than the human responses in the forum by doctors (Ayers et al., 2023). Other studies show that interactions with LLMs reduce conspiracy beliefs demonstrating efficacy related to reliable knowledge that supports learning and reflexion (Costello et al., 2024). These studies act as proofs of concept how LLMs can be integrated in a learning and reflecting practice. We will apply these pipelines of using LLMs to education and develop a novel tool for learning that can be integrated in any curriculum with few resources.

Bibliographic References

Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., Faix, D. J., Goodman, A. M., Longhurst, C. A., & Hogarth, M. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine*, *183*(6), 589-596. https://doi.org/10.1001/jamainternmed.2023.1838

Bauzon, J., Alver, A., Ravikumar, V., Devera, A., Mikhael, T., Nauman, R., & Simanton, E. (2021). The impact of educational resources and perceived preparedness on medical education performance. *Medical science educator*, *31*(4), 1319-1326. https://doi.org/10.1007/s40670-021-01306-x

Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, *385*(6714), eadq1814. https://doi.org/10.1126/science.adq1814 #

Gibbs, G. & Simpson, C. (2005). Conditions Under Which Assessment Supports Students' Learning. Learning and Teaching in Higher Education.

Guisset, M. et al. (2022). Évaluations par Cartes Conceptuelles à trous et apprentissage par les pairs. UCLouvain.

Kafke, S. D., Kuhlmey, A., Schuster, J., Blüher, S., Czimmeck, C., Zoellick, J. C., & Grosse, P. (2023). Can clinical decision support systems be an asset in medical education? An experimental approach. *BMC Medical Education*, *23*(1), 570. https://doi.org/10.1186/s12909-023-04568-8

Lewis, M. et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS.

Lück, S., Kienle, R., Langenbeck, S., Sehy, V., & Freytag, J. (2022). Motivierende Gesprächsführung mit Simulationspatient_innen lehren und prüfen. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *50*(3-4), 159-167. https://doi.org/10.1026/1616-3443/a000637

Mahendiran, M., Yeung, H., Rossi, S., Khosravani, H., & Perri, G.-A. (2023). Evaluating the effectiveness of the SPIKES model to break bad news–a systematic review. *American Journal of Hospice and Palliative Medicine®*, *40*(11), 1231-1260.

Miller, W. R., & Rollnick, S. (2012). *Motivational interviewing: Helping people change*. Guilford press.

Novak, J.D. (2008). The Theory Underlying Concept Maps and How to Construct Them.

Wei, J. et al. (2022). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. arXiv.

Work Plan

Day(s)	Task	Lead(s)
Day 1	Kick-off meeting: project briefing, team role assignment, tech onboarding (RAG, CoT)	All, supervised by mentors
Day 2	Literature review on CCàT, RAG, and CoT in educational contexts; initial concept framing	Research subgroup
Day 3	Use case 1: building a dataset of conversations for analyses Use case 2: extraction of one or two CCàT examples and associated knowledge base (use case 2);	Content subgroup (with Rees and Zoellick)
Day 4	Use case 1: Analysing audiovisual needs and possibilities Use case 2: Implementation of a basic RAG pipeline (e.g. using Haystack, LangChain or LlamaIndex)	Tech subgroup (Arthur Pisvin)
Day 5	Design of Chain-of-Thought prompting templates for AI-supported student reasoning; refined prompting pipelines	Pedagogy subgroup (Bioul, Nagy, Zoellick, Müller)
Weekend	Integration: link CCàT map with RAG-CoT support (e.g. student clicks a blank → AI suggests or justifies a concept) Reflecting on the first week	Cross-group pairing
Day 6	Internal testing: simulate student interactions, gather feedback on usability	All
Day 7	Refinement of UI/UX flow (if applicable); adjustment of RAG sources or CoT scaffolding	Tech + pedagogy subgroup
Day 8	Pilot test with external users (e.g. 2 medical students for use case 1 and 2 biology students for use case 2), collect qualitative feedback	Bioul, Rees, Nagy, Zoellick, Müller

Day 9	Analyze user feedback and learning traces (concepts used, justifications, errors)	All
Day 10	Final presentation prep: demo, poster, short video or slides summarizing the prototype	All

Expected Outcomes

Scientific paper

Reproducible GitHub repository

Toolbox

Domain(s) of Application

Education

Scientific Theme(s)

Human-AI interaction

AI in Higher Education

Generative AI

TRAIL Grand Challenge(s)

NA

Additional Comments

This project bridges pedagogical innovation with cutting-edge AI research and aligns with Circle U's goals of fostering student agency, interdisciplinary collaboration, and open innovation in teaching. It may also serve as a pilot for a broader Circle U teaching toolkit on AI-enhanced learning.