ARTICLE IN PRESS

xxx (xxxx) xxx

Contents lists available at ScienceDirect

Geodata and AI

journal homepage: www.elsevier.com/locate/geoai

Assessing the usefulness of large language models in geotechnical education: EduHackathon at the third workshop on the future of machine learning in geotechnics (3FOMLIG)

Michele Calvello [a,*], Emilio Bilotta [b], Alessia Cuccurullo [c], Ana Heitor [d], Charles MacRobert [e], Enrico Soranzo [f], Marco Uzielli [g], Stephen Wu [h]

[a] Department of civil engineering, University of Salerno, Fisciano, Italy
[b] Department of Civil, Building and Environmental Engineering, University of Napoli "Federico II", Napoli, Italy
[c] Geomechanics Laboratory LGM, BATir Department, Université libre de Bruxelles, Bruxelles, Belgium
[d] School of Civil Engineering, University of Leeds, Leeds, UK
[e] Department of Civil Engineering, Stellenbosch University, Stellenbosch, South Africa
[f] Institute of Geotechnical Engineering, Department of Landscape, Water and Infrastructure, BOKU University, Vienna, Austria
[g] Department of Civil, Building and Environmental Engineering, University of Florence, Firenze, Italy
[h] Research Organization of Information and Systems, The Institute of Statistical Mathematics, The Graduate University for Advanced Studies (SOKENDAI), Tachikawa, Tokyo, Japan

A R T I C L E I N F O

Keywords:
Geotechnical engineering education
Pedagogical tools
Generative AI

A B S T R A C T

This article reports on the process and outcomes of the EduHackathon "Assessing the usefulness of Large Language Models in geotechnical education" held in Florence, Italy, as a session of the third Workshop on the Future of Machine Learning in Geotechnics (3FOMLIG) on October 16, 2025. Recognizing the increasing implementation of Large Language Models (LLMs) and Generative AI (Gen AI) for domain-specific geotechnical workflows, the event aimed to explore their potential utility in geotechnical engineering education. The EduHackathon was the culmination of a collaborative effort by a core group of geotechnical educators who presented and discussed with conference attendees a series of studies conducted, prior to the meeting, to evaluate LLM and Gen AI tools from both student and instructor perspectives. The article details aims and main outcomes of these studies, as well as the main conclusions of the discussions carried out in person, which were also driven by the results of a survey previously undertaken by 3FOMLIG participants. The findings suggest that while LLMs offer significant potential for enhancing interactive and reflective learning, they should be viewed as flexible pedagogical tools rather than autonomous instructors. It can be emphasized that the educational value of these tools depends heavily on careful integration with sound teaching principles and human supervision to mitigate potential risks, such as cognitive debt, inaccuracies, and the loss of critical information. Ultimately, the article proposes to keep exploring the potential benefits and the main limitations associated with the use of these tools within educational modules in soil mechanics and geotechnical engineering, where LLMs may serve as collaborative interlocutors to both geotechnical educators and students, in a sort of epistemic co-production with human expertise.

1. Introduction

Recent advances in large language models (LLMs) and generative AI (Gen AI) are beginning to translate from general-purpose language interaction tools to domain specific working tools, also including geotechnical engineering workflows (e.g., Wu et al., 2024). Research has also started exploring the impact of these tools on general education, yet very few studies have examined how it affects engineering students and instructors (e.g., Simelane and Kittur, 2024; Liu, 2025), and the specific field of geotechnical engineering education is still widely unexplored.

At the beginning of 2025, the first author of this paper identified the third Workshop on the Future of Machine Learning in Geotechnics (3FOMLIG), planned to be held a few months afterwards, as a perfect venue for addressing these issues with a group of fellow geotechnical

* Corresponding author.
E-mail address: mcalvello@unisa.it (M. Calvello).

https://doi.org/10.1016/j.geoai.2026.100118
Received 17 April 2026; Received in revised form 21 July 2026; Accepted 26 August 2026
Available online 5 September 2026

Please cite this article as: Michele Calvello et al., , https://doi.org/10.1016/j.geoai.2026.100118

engineers and educators. After some discussion with the organizers of the Workshop, the EduHackathon idea emerged and the work to make it happen started. This article briefly reports on the process followed and on the main outcomes of this event.

2. Purpose and structure of the EduHackathon

The EduHackathon "Assessing the usefulness of Large Language Models in geotechnical education" was held at the 3FOMLIG Workshop in Florence (Italy) on October 16, 2025. It was organized to explore the potential usefulness of LLMs and other Gen AI tools for geotechnical education purposes. The EduHackathon plenary session during the Workshop was conceived as the final part of a longer process involving a core group of geotechnical engineering educators (the authors of this article) who, in the months before 3FOMLIG, met regularly to interact on the specific educational topics to tackle, to jointly explore different LLM and Gen AI tools, to create reports on the issues tackled, and to discuss preliminary findings prior to the live sessions at the Workshop.

The EduHackathon session in Florence comprised two parts. In the first part, the members of the core group of conveners delivered eight short presentations showing the main results of a series of "experiments" conducted individually in the months preceding the Workshop. In the second part, a moderated floor discussion was held, starting from the results of a survey carried out among 3FOMLIG participants in the days preceding the Workshop.

The presentations of the first part of the session were grouped considering the issues tackled and the main purpose of the different experiments, so as to present to the audience, in the following order: the key features of the main LLM tools currently available to educators; studies addressing student learning; approaches addressing potential interactions with LLMs suitable for both students and teachers; LLMs as aids for geotechnical educators. The titles and the presenters of the eight invited talks follow.

1. Gen AI and LLM tools suitable for educational purposes (by Stephen Wu, The Institute of Statistical Mathematics, Tachikawa, Tokyo, Japan)
2. Students' misconceptions and conceptual barriers (by Charles MacRoberts, Stellenbosch University, Stellenbosch, South Africa)
3. Assisting student's learning: capturing diverse learning styles with NotebookLM (by Ana Heitor, University of Leeds, Leeds, UK)
4. Competence matters: assessing the quality of human-LLM interaction (by Marco Uzielli, University of Florence, Firenze, Italy)
5. Soil mechanics concepts "explained" by posing direct questions to textbooks (by Michele Calvello, University of Salerno, Fisciano, Italy)
6. Evaluation of LLMs in relation to source material and baseline knowledge (by Emilio Bilotta, University of Napoli "Federico II", Napoli, Italy)
7. Generative AI for exam preparation and feedback on student reports (by Alessia Cuccurullo, Université libre de Bruxelles, Bruxelles, Belgium)
8. Automated Grading with LLMs in Geotechnics (by Enrico Soranzo, BOKU University, Vienna, Austria)

The second part of the session started with an introductory talk (by Michele Calvello) aimed at setting the stage for the subsequent floor discussion (moderated by Emilio Bilotta). The presentations were centred around the aggregated outcomes of a short survey carried out among 3FOMLIG participants, comprising the following three open questions:

- If you have used generative AI for educational purposes, tell us something about your experiences (e.g., what was the purpose of using it what tools have you experimented with; what are your main comments about these tools)
- How do you think generative AI can be usefully adopted, in some way, within university courses?
- What are the main benefits and threats to students (undergraduate and postgraduate) stemming from the use of generative AI?

3. Invited talks

Given the pace of development in LLMs, it is worth highlighting that the findings reported should be interpreted in the context of the models' state-of-the-art that existed in the months preceding 3FOMLIG in 2025.

3.1. Gen AI and LLM tools suitable for educational purposes

The study aimed at surveying and contextualizing recent Gen AI and LLM tools that are potentially suitable for educational use, with particular reference to higher education and STEM (Science, Technology, Engineering, Mathematics) disciplines, such as geotechnical engineering. Rather than benchmarking a single platform, the study sought to frame LLM-based tools within established pedagogical theory, identify current technological trends, and discuss both opportunities and risks associated with their integration into teaching and learning practices.

A central argument of the presentation was that the educational value of LLMs depends primarily on how they are used, rather than on their technical capabilities alone. This position was framed through reference to long-standing critiques of passive learning models, notably the "banking model" of education articulated by Freire (1970), in which learners are treated as passive recipients of information. When used uncritically, LLMs risk reinforcing this model by providing ready-made answers that discourage inquiry and reflection. Conversely, when integrated thoughtfully, conversational AI systems can support more active and dialogic forms of learning by encouraging students to formulate questions, articulate reasoning, and explore alternative explanations.

The presentation situated contemporary LLM tools within the broader evolution of pedagogical agents. Earlier work on animated and conversational agents emphasized personalization, scaffolding, and learner engagement through dialogue. Advances in LLMs extend these ideas by enabling open-ended natural language interaction, adaptive feedback, and domain-flexible tutoring. Emerging trends include LLM-based pedagogical agents (Yusuf et al., 2025) and multi-agent educational systems (Zhang et al., 2025), in which different agents may assume complementary instructional roles such as tutor, challenger, or reflective guide. These developments are particularly relevant in technical disciplines, where iterative explanation and conceptual refinement are essential for meaningful learning.

A survey of currently available platforms illustrated how these concepts are being translated into practice. General-purpose systems such as ChatGPT are increasingly deployed at institutional scale to support student assistance, content generation, and administrative tasks. Education-oriented tools, such as Khanmigo (Shetye, 2024), emphasize Socratic guidance and curriculum alignment, while other platforms target specific functions such as step-by-step problem solving in STEM subjects. In parallel, domain-restricted or classroom-oriented LLMs seek to improve safety and reliability by grounding responses in curated or institutionally controlled content. Although more advanced systems incorporating features such as citation, self-reflection, or multi-agent reasoning remain largely at a research or prototype stage, they point to possible future directions for educational AI.

Alongside these opportunities, the presentation emphasized emerging empirical evidence regarding potential risks to cognitive engagement. Recent experimental studies have shown that while LLM-based assistants can improve short-term productivity in tasks such as writing, sustained reliance on such tools may reduce effortful thinking, weaken recall, and diminish learners' sense of ownership over their work. Kosmyna et al. (2025) introduce the notion of "cognitive debt" to describe the cumulative cost of outsourcing cognitive processes to AI systems, reporting reduced neural engagement and increased linguistic homogeneity among participants who relied heavily on LLM assistance. These findings highlight the importance of designing educational uses of LLMs that actively promote reasoning and reflection rather than substitution of cognitive effort.

In the context of geotechnical education, these considerations are particularly significant. Engineering judgment and conceptual understanding cannot be replaced by automated responses, yet LLMs may offer valuable support when used as tools for clarification, comparison, and exploration. For example, they may assist students in articulating misconceptions, examining alternative interpretations of concepts, or reflecting on solution strategies, while also supporting educators in generating prompts, examples, and alternative explanations aligned with active learning objectives.

The main conclusion of this study, echoed throughout the Edu-Hackathon discussion, is that LLMs should not be regarded as autonomous instructors, but as flexible pedagogical tools whose educational impact depends on careful integration with sound teaching principles. Used thoughtfully, they have the potential to support more interactive and reflective learning experiences; used uncritically, they risk undermining the very cognitive processes that education seeks to cultivate. In this sense, the key question is not whether LLMs should be used in geotechnical education, but how they can be employed in ways that genuinely enhance learning rather than merely offering convenience.

3.2. Students' misconceptions and conceptual barriers

The LLM tool used in this study was the Enterprise version of Microsoft Co-Pilot, as this is increasingly integrated into Microsoft products, which are ubiquitous in many higher education settings. Two experiments were conducted, the first was to elicit misconceptions or conceptual barriers students may have around specific geotechnical concepts, and the second was to test whether LLMs can predict potential misconceptions or errors students may make with assessment questions.

Engineering education misconceptions are generally conceptual or procedural. Conceptual errors occur when students have incomplete or conflicting mental models, often due to prior beliefs, struggles with threshold concepts, language ambiguities, or mismatches between daily experience and scientific principles. Research indicates these misconceptions persist because students fit new knowledge into existing flawed frameworks, and they are reinforced by repetition, making them hard to correct without targeted instruction (Streveler et al., 2014). Procedural misconceptions involve errors in carrying out steps or using formulas, rather than misunderstanding concepts. In geotechnical engineering, tasks like computing bearing capacity or predicting settlement require multi-step calculations. Students can forget steps, apply methods incorrectly or simply make calculation errors. Research shows conceptual and procedural knowledge reinforce each other, so weaknesses in one area can lead to mistakes in the other (Rittle-Johnson and Schneider, 2015). Procedural errors are common in geotechnical engineering education because of complex calculations, unit conversions, and integrating lab data with analysis (Pinho-Lopes and Powrie, 2020).

Experienced educators are often familiar with student misconceptions. However, this experiment was conducted to see if LLMs would be beneficial to new educators or educators preparing to teach a new subject. Prior to LLMs, online chat boards (e.g., www.eng-tips.com, www.engineersedge.com and www.reddit.com) provided a way to identify misconceptions. Being open source, these chat boards are likely included in the databases used to train LLMs.

Prompt design followed an iterative approach which started from a broad prompt "What concepts in soil mechanics do undergraduate students find most difficult to understand?" with subsequent prompts narrowed down to more specific concepts based on responses received. The general drift of the chat went from general concepts, to exploring misconceptions around slope stability, limit equilibrium, methods of slices, porewater pressure, shear strength, to the prompt evaluated in this paper, "Give examples of student misconceptions surrounding cohesive and cohesionless soils?" To generate a broad range of misconceptions, further prompts included, "Please provide additional examples for the above", "Please provide more examples of the above misconceptions" and "Can you provide 10 more misconceptions around cohesive and cohesionless soils." This prompting resulted in 28 misconception and reality statements, many of which were repetitive. Some examples are given in Table 1. Only one instance of the chat was evaluated. Responses were reviewed by the author (Charles MacRobert) based on over 10 years teaching experience and relevant literature. These responses reveal that misconceptions largely revolve around everyday use of the terms cohesive (i.e., sticky, stable and solid) and cohesionless (i.e., not bound or weak) compared to their technical use. While this is not an unknown misconception (Santamarina, 2015; Wesley, 2015) the prompts can help educators frame their discussion of the behaviour of fine and coarse grained soils (and perhaps away from using the terms cohesive and cohesionless).

The second prompt (or pair of prompts) was to provide Co-Pilot with verbatim exam questions and ask "I am going to provide you with an exam question. I would like you to identify potential errors students could make in answering this question. I would like to be able to point these out during lectures." Provided responses grouped errors into conceptual, procedural and interpretation-related issues and ended with a suggested list of lecture highlights. Only one chat was evaluated.

The first question required students to determine whether two designs (tabulated values of foundation dimensions, soil strengths and loads) were satisfactory to stated ultimate limit state design requirements. The suggested lecture highlights are provided verbatim below. Three of these suggestions (1, 2 and 4) are common errors that students make.

1. Watch out for whether you're in total or effective stress territory.
2. Always ask: where is the groundwater? It changes everything.
3. Don't just plug into a formula—check if it's the right one for the soil condition.
4. DA1a and DA1b are not interchangeable—know your limit states.
5. State your assumptions clearly—especially when data is missing or implied.

The second question required students to calculate the allowable bearing pressure for a square footing (dimensions provided) using the Schmertammn method based on tabulate CPT results. Suggested lecture highlights are provided verbatim below. The suggestions were relevant but also generic (e.g., item 2 was not relevant as CPT data had already

Table 1
Examples of student misconceptions suggested by Co-Pilot.

Misconception	Reality
Cohesive soils are always strong and stable.	While cohesive soils like clay can have high shear strength when dry, they can lose strength significantly when wet due to increased pore water pressure.
Cohesive soils do not experience significant volume changes.	Cohesive soils, particularly clays, can undergo significant volume changes due to swelling and shrinkage when moisture content changes.
Cohesive soils do not drain at all.	Cohesive soils have low permeability, meaning they drain slowly, but they do eventually allow water to pass through.
Cohesionless soils have no strength.	Cohesionless soils derive their strength from friction between particles. They can be very stable under certain conditions, especially when well-compacted.
Cohesionless soils do not retain water.	While cohesionless soils have high permeability and allow water to drain quickly, they can still retain some water within the pore spaces.
Cohesionless soils cannot support heavy loads.	Cohesionless soils can support heavy loads due to their high frictional strength, especially when compacted.

been interpolated). Nevertheless, these generally align with errors observed in assessment responses.

1. Misidentifying the depth of influence
2. Incorret interpolation of CPT data
3. Confusing depths: Ground level vs. Footing base
4. Misapplying the Schmertmann Equation
5. Incorrect use of Elastic Modulus
6. Ignoring the unit weight in overburden correction
7. Misinterpreting the allowable settlement
8. Confusing Net vs Gross Bearing Pressure
9. Ignoring the hint (limits to bearing pressure)

In conclusion, this simple experiment showed that Co-Pilot can be used to elicit misconceptions around basic principles. No particularly novel misconceptions were identified that could not be identified using other sources (e.g., talking to other educators, reading literature, or teaching). Nevertheless, it was particularly revealing that most errors identified for the exam questions were procedural in nature, suggesting conceptual understanding is not being assessed. This is considered the most useful outcome of the experiment.

3.3. *Assisting student's learning: capturing diverse learning styles with NotebookLM*

A study was carried out to explore the capability and feasibility of using common large language model tools, such as Google NotebookLM, to facilitate a student's learning process by capturing diverse learning styles. Recent research has shown that the use of LLMs can offer a transformative potential in education by improving accessibility, inclusivity, and individualized learning experiences (e.g., Lopez-Gazpio, 2025). For instance, Fig. 1 shows how the main challenges in inclusive education can be overcome using LLM as language tools, to develop adaptive content and feedback, for multimodal content, or to allow flexible delivery. This study focused mainly on the development of multimodal content that can cater for diverse learning styles, which is a critical issue to assist in closing success gap for students from diverse

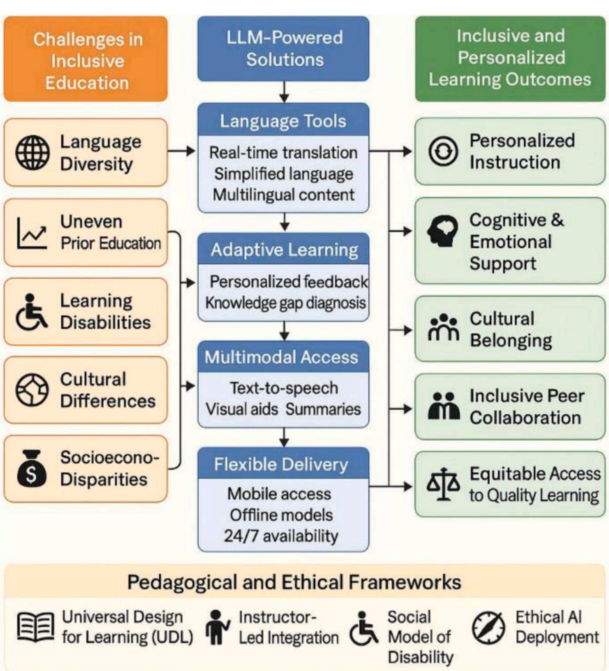

Fig. 1. Schematic illustration of the integration of large language models into inclusive education (From: Lopez-Gazpio, 2025, reproduced with permission under CC BY 4.0).

backgrounds, in terms of language diversity and cultural differences, as well as in overcoming barriers associated with learning disabilities and socioeconomic inequalities.

There are many learning styles systems that have been proposed in the pedagogical literature. One of the simplest models (named VARK) captures four sensory modalities for information intake: visual, auditory, reading and kinaesthetic (Fleming and Baume, 2006). While this model offers good insights into how students learn, more complex systems, addressing perception and processing within a learning cycle have been proposed. For instance, Kolb (1984) identified four learning styles based on preferences, i.e., diverging (feeling/watching), assimilating (watching/thinking), converging (thinking/doing), and accommodating (doing/feeling). Among these, the converging style, which represents learners that excel at problem solving, decision making, and the practical application of ideas was found to be more common for engineers. More recently, Felder and Solomon, (2008) expanded the Kolb learning style categories into active/reflective (AR), sensing/intuitive (SI), visual/verbal (VR), and sequential/global (SG). Within this framework, engineering students have a greater alignment with active, sensing, visual, and sequential learners.

Recent surveys conducted across higher education institutions in the UK (Neves et al., 2025) reveal that around 70% of students prefer multimodal learning, utilizing a combination of styles depending on the context. Notably, 79% of students believe they learn better when taught in their preferred style. Despite these preferences, empirical evidence supporting the effectiveness of matching teaching methods to learning styles remains limited. These statistics underscore the importance of acknowledging student preferences (Gaikawad and Patil, 2017) while also emphasizing the need for flexible and inclusive teaching strategies that cater to a broad range of learning needs. AI in engineering education offers opportunities for the development of adaptive learning systems that dynamically adjust content and teaching methodologies to suit individual students' learning styles (Timms, 2016). Furthermore, AI-driven systems can identify individual learning preferences through continuous assessment and tailor content delivery in real time, thereby creating a more personalized and effective learning environment (Unesco, 2019; Isaza Dominguez et al., 2024). Despite these perceived benefits, the challenge of instructor supervision to avoid inaccuracies and hallucinations persists.

In this study, only one tool was evaluated, i.e., NotebookLM. This tool is well suited for the development of content that suits a variety of learning styles, and it allows some level of instructor supervision, such as the selection of sources, thus enhancing the accuracy of the answers and contributing to a reduction in the likelihood of AI errors and hallucinations. The version tested, powered by Gemini 2.5, supports the development of a variety of resources, including: mind maps, context-rich video overviews, infographics catering to visual and spatial learners. Simultaneously, conversational interfaces can be used by auditory and verbal learners through dialogue-based exploration, especially by means of audio overviews. To audit knowledge retention and self-assessment aligned with kinaesthetic and reflective learners, NotebookLM also produces flash cards and quizzes. The use of these resources, if properly supervised by educators, may also facilitate learning progress tracking.

NotebookLM was tested within an MSc module taught at the University of Leeds, i.e., CIVE5160 – Geotechnical Characterisation and investigation. The notebooks sources were carefully curated by the instructor and then shared with the students. Four notebooks were developed on the following topics associated with soil characterisation: particle size distribution, plasticity limits, hydraulic conductivity, and effective stress in unsaturated soil. The purpose of these was to complement resources delivered in a flipped manner (i.e., video lecture and in person workshops) by allowing students to explore more advanced topics, mainly focusing on the limitations of the tests adopted for soil characterisation. Anecdotal feedback from the small group of students involved (12 students) indicate that they all engaged with the notebooks, they mostly welcomed the audio overviews, while they were less interested in mind maps, quiz, flash cards and infographics. It should be noted that the content covered in the module is supported by other modules in the program, where more fundamental soil mechanics concepts are addressed. Further research with wider groups of students is of course needed to draw any conclusions on how to possibly integrate these tools more seamlessly in asynchronous module activities to help student learning geotechnical engineering concepts.

In summary, technology is reshaping the landscape of higher education in the UK. Hybrid learning models, combining in-person and digital instruction, have become the norm. Students increasingly value flexibility, interactivity, and access to digital resources. Tools such as NotebookLM which provides some level of adaptive learning, have the potential to enhance the educational experience by providing personalized and engaging content. These developments highlight the importance of integrating technology into teaching to meet evolving student needs and preferences.

3.4. Competence matters: assessing the quality of human-LLM interaction

3.4.1. Conceptual rationale and research questions

The use of LLMs by undergraduate students, postgraduate students, and researchers for independent study, completing assignments, and supporting research work is increasing rapidly. Within this increased interaction, users often assume LLMs to be inherently authoritative sources of knowledge. However, is this always true? This presentation focused on the work-in-progress development, conceptual framing, and pilot testing of a comparative experiment designed to evaluate the quality of human-LLM interaction in geotechnical engineering education and research. The central hypothesis underlying the activity is that the quality of LLM output does not solely depend on the LLM but emerges from the interaction between model capabilities and user competence. This hypothesis raises two guiding research questions: (1) to what extent is the correctness and structure of an LLM’s output driven by the quality of the question it received? (2) do the model’s responses scale in sophistication when interacting with more competent users? In short, “can the answer only be as good as the question”?

3.4.2. Methodological structure

The project can be conducted according to a stepwise process involving sequentially: (1) framing and briefing; (2) human-LLM prompt interaction sessions with a focus on a geotechnical topic at various competence levels; (3) data collection and anonymization to comply with privacy-related aspects; (4) qualitative and statistical single-blind expert evaluation of outputs; (4) reflection on results and debriefing; and (5) reporting and scientific dissemination. The empirical data generated allow: (1) comparison of human-LLM interaction quality across varying competence levels; (2) identification of LLM-human interaction features that enhanced or hindered learning; and (3) evaluation of whether competence-adaptive prompting could increase pedagogical and research value.

3.4.3. Parameterization of interaction quality

The quality of human-LLM interaction was projected along three dimensions; namely: (1) epistemic quality: the degree to which LLM responses were technically correct, logically consistent, and aligned with accepted geotechnical principles; (2) communicative quality: the clarity, structure, transparency, and thoroughness of the explanations provided; and (3) pragmatic quality: the usefulness of the output for learning, including its capacity to guide reasoning, consolidate understanding, provide methodological steps, etc. A shared semi-quantitative quality rating indicator scale was defined in the range 1 (very low quality) to 5 (very high quality).

3.4.4. Pilot experiment

A mock pilot application was conducted to explore the feasibility of the framework and to provide initial evidence of competence-dependent interaction effects. The pilot experiment should be interpreted as a controlled simulation, in which the LLM was explicitly instructed to adopt different user competence profiles. More specifically, the LLM was asked to simulate three interactions with fictitious users representing three distinct competence levels; namely: (1) an undergraduate student (UGS); (2) a postgraduate student (PGS); and (3) a researcher (RES). All three interactions started from the identical prompt: “How can I design a foundation on sand?” The resulting transcripts exhibited substantial variation across the three simulated user competence levels. In the UGS case, the model offered general guidelines, introductory explanations, and simplified conceptual descriptions. In the PGS case, the model’s explanations became more technical, referring to intermediate design methods and geotechnical considerations. In the RES case, the model produced detailed methodological outlines, identified assumptions, and framed the response within advanced analytical reasoning and technical aspects including spatial variability and uncertainty. The pilot study demonstrated that the model autonomously modulated the epistemic and communicative depth of its responses based on user competence. It confirmed the underlying hypothesis that identical prompts could lead to markedly different outputs depending on the inferred expertise of the interlocutor. Differences were not limited to contents; they also involved changes in structure, terminology, problem decomposition, and explanation style. In the RES simulation, for instance, the LLM tended to adopt a more reflective and analytical tone, explicitly referencing methodological choices and conditional frameworks while in the UGS-competence simulations, the model emphasized foundational concepts and procedural clarity.

3.4.5. LLM self-evaluation

Following completion of the comparative interactions, the LLM was asked to perform a structured self-evaluation of its own outputs in the form of a quality matrix relying on the previously described indicators of epistemic, communicative, and pragmatic quality. This step highlighted the capability on the part of the LLM to articulate an internally coherent assessment of its strengths and limitations. Although such self-evaluation is, of course, itself subject to expert scrutiny, it offered indications that LLMs could assist humans not only through the provision of content but also through explicit reasoning audits, thereby supporting metacognitive development.

3.4.6. Concluding remarks and expected developments

The preliminary activities performed suggest a series of potential pedagogical and scientific benefits. Pedagogical benefits may include: (1) enhanced epistemic awareness among students, who would observe how the progressive acquisition of competence would reshape the quality and structure of LLM outputs, thereby triggering a virtuous educational spiral; (2) improved prompting skills, providing learners with enhanced capability to engage critically with LLMs; and (3) repurposing LLMs as collaborative learning tools rather than answer-providing machines. However, the same results also suggest the need for caution, as inappropriate or unsupervised interaction could mislead inexperienced users. Research benefits may include: (1) acquisition, interpretation, and reporting of systematic data on competence-dependent variations in human-LLM interaction quality; (2) development of novel methodologies for evaluating AI-assisted learning in geotechnical engineering contexts; and (3) contributions to the emerging research field of AI-in-education for the geotechnical discipline. Further research could support the development of a framework for integrating LLMs into geotechnical training in a manner that may reinforce human expertise and promote reflective, data-literate engineering practice. The concept of “epistemic co-production” is proposed as a foundation for future educational models in which LLMs serve as collaborative interlocutors rather than unilateral information providers.

3.5. Soil mechanics concepts "explained" by posing direct questions to textbooks

The LLM tool that was used in this study is Google NotebookLM. This tool was chosen because it has the following two key features: the answers are generated by considering only the content of specific sources uploaded by the user; links are provided to the paragraphs of the source material used to generate the sentences in the answer. The objective of the study was twofold: i) evaluating the possibility to get good explanations of basic soil mechanics concepts by directly interrogating a specific textbook using an LLM; ii) comparing how different textbooks address and discuss the same concepts (Calvello, 2025).

The two books used in the study are: (i) Principles of Geotechnical Engineering, 9th ed (Das, 2017); (ii) Mecânica dos Solos, 6th ed (Matos Fernandes, 2024). Even if the second book is written in Portuguese, English was used for the questions and answers in the chat of NotebookLM, as this tool can "read" text written in languages other than English, at the same time providing the answers in the language chosen by the user for the prompts. To generate the chats, two different NotebookLM projects have been produced, having as source each single soil mechanics textbook (divided in chapters to avoid the possible truncation of the source text by the LLM). The same questions were posed to the two textbooks, exploring three soil mechanics concepts, as follows:

1. Shear strength
   - What is the shear strength of a soil?
   - Is effective cohesion a fitting parameter of the Mohr-Coulomb failure criterion or a true property of the soil?
   - Can soil shear strength and soil failure be used as synonyms?
   - Are there different types of soil failure?
   - Can soils exhibit different behaviours upon reaching the maximum shear strength?
2. Consolidation
   - What is consolidation?
   - Soil compression and soil consolidation are the same process?
   - How does creep differ from consolidation?
   - What are the main mathematical laws governing primary consolidation?
3. Volumetric compression
   - What is soil volumetric compression?
   - Is there a difference between soil compressibility and soil deformability?
   - Are soil settlements related to compressibility or deformability?
   - How can one compute the soil settlements due to foundations?

The study allowed to highlight both positive and negative features of the performed interaction. On the positive side, it can be stated that both the objectives of the study have been met, as the answers from NotebookLM provide a good synthesis of the general concepts addressed (starting with a straightforward to-the-point short sentence and ending with a final synthesis paragraph), leaving the user with the possibility to directly read the parts of source material linked to them. Moreover, the multi-language ability of LLMs allow users (geotechnical instructors) to consult and interrogate, in English, textbooks written in other languages they may not know. Particularly interesting answers come from prompts explicitly asking differences between things (e.g., processes, parameters) as the answers typically produce useful comparison lists. On the negative side, one must be aware that the answers are based on the sources uploaded as well as on previous answers within the conversation (with the latter appearing to have a significant weight) and thus the same question may produce slightly different answers depending on when it is asked within a conversation. At the time of the experiment, figures were not processed, and therefore answers were based only on figure captions and on text commenting them, and the equations were not always written (both in the reference text and in the answers) in an understandable format.

The findings of this study can be useful to geotechnical educators wishing to consult different sources to improve the way they explain specific soil mechanics concepts to students, e.g., to corroborate knowledge and to explore new ways of dealing with these concepts (which may be alternative but not conflicting or mutually contradictory). Moreover, the natural conversational style typical of LLMs may be considered a valuable feature for both students and educators to effectively explore a textbook in relation to a soil mechanics concept. Indeed, the answers themselves may give hints to the user on how to "keep on chatting" with the LLM to explore subtopics of interest or to dive deeper into a specific issue that was mentioned.

3.6. Evaluation of LLMs in relation to source material and baseline knowledge

The study was guided by three core research questions:

1. Can the AI tool discern a logical narrative and conceptual flow from graphically-intensive lecture slides?
2. Is the information summarised by the AI factually correct and grounded in the provided source materials?
3. Does the AI enhance the source material by effectively incorporating its own "baseline knowledge" for further elaboration and clarification?

To investigate these questions, the study employed two distinct exercises designed to test the capabilities and limitations of LLMs in different educational scenarios. The first exercise focused on the AI's ability to summarize content from a closed-source environment, using only the provided lecture materials. The second exercise expanded the scope to compare the performance of a closed-source tool against an AI-powered search engine that actively browses open scientific literature.

3.6.1. Exercise 1: evaluating AI-generated audio summaries from lecture slides

This experiment was designed to assess an LLM's capacity to interpret and summarize complex course materials. The tool selected for this task was Google NotebookLM, with a specific focus on its "audio summary" feature. According to the tool's online manual, this feature generates what are described as "*lively and in-depth discussions between two AI hosts*" based exclusively on user-uploaded source documents.

Source Material: the inputs for the AI were PDF slideshows from a course on geotechnical engineering for tunnels. These presentations, ranging from 30 to 120 slides each, contained a typical academic mix of content, including bullet points, short descriptive sentences, equations, tables, and numerous figures.

Process: a separate Notebook LM project was created for each lecture of the course. The corresponding PDF slideshow was uploaded into each project, and the tool was prompted to generate one audio summary for each distinct lecture. The resulting audio files were then analysed for quality, accuracy, and coherence.

The audio summaries generated by NotebookLM from lecture slides produced surprisingly positive results. The AI was able to generate comprehensive dialogues in fluent English that successfully captured the logical flow and overarching themes of the lectures. This output was stylistically sophisticated, as the two AI hosts were observed to mimic the interaction between two different point of view, like in a real discussion between peers. Rather than simply reciting the text from the slides, the tool demonstrated a notable degree of analytical capability; they "*did not learn the lesson by heart*" but clearly show autonomous processing of the information received. This was further enhanced by the AI's ability to draw upon its own background knowledge base to make sophisticated parallels and comparisons, framing complex technical ideas in more accessible terms.

Despite these strengths, the tool exhibited several significant limitations. The most critical weakness was the complete loss of information contained in figures, as the AI was unable to process graphical content at the time the exercise was carried out (latest versions of the tool have overcome this limitation). Furthermore, the summaries were often very general, providing a high-level overview without capturing the necessary technical depth and detail required for engineering students. Finally, the tool offered no mechanism for users to interact or provide feedback to correct lexical imprecisions or improve the tool's use of specialized technical vocabulary.

3.6.2. Exercise 2: comparing closed-source vs. literature-grounded AI tools

The second experiment aimed to compare the educational outcomes of two functionally different AI tools. The first was NotebookLM, operating in a closed environment with user-provided textbooks; the second was Consensus, an AI-powered academic search engine designed to ground its responses in published scientific literature.

Source Material: the key distinction lies in the information sources. The results of the study described in Section 3.5 using NotebookLM were used as a benchmark, since the literature sources was provided by the user, thereby restricting its operation to the content of two soil mechanics textbooks uploaded by the researcher. In contrast, Consensus was designed to browse a large corpus of existing scientific papers.

Process: the same set of questions posed to NotebookLM in the study of Section 3.5, focusing on the core geotechnical topic of "shear strength", was prompted to Consensus. The performance of Consensus was then evaluated based on the quality, structure, and documentation of its responses.

The Consensus tool demonstrated exceptional performance. Its outputs were consistently well-documented, accurate, and thoughtfully structured. Responses were broken down into logical sections, typically including a concise framing statement, a series of detailed discussion points, and a concluding summary.

The strength of Consensus was its rigorous grounding in scientific research. Every statement and claim made by the tool was explicitly supported by citations from valid journal papers, and a complete list of references was provided with each answer. This ensures a high degree of reliability and academic integrity. The tool consistently used a minimum of 10 papers to formulate a response and, in cases where sufficient literature was not immediately available, it would proactively suggest alternative question phrasing to improve the search results and yield a more detailed answer. The only minor issue identified was a single response that covered a broader spectrum than strictly soil mechanics, touching upon domains such as engineering geology, geology, and soil science. However, the tool itself offered suggestions for refining the query to better circumscribe the application domain.

3.7. Generative AI for exam preparation and feedback on student reports

This presentation reported on two exploratory studies examining the use of GenAI and LLMs as pedagogical tools to enhance both teaching and learning experiences. LLMs are already being studied for practical geotechnical applications, such as supporting problem-solving and engineering reasoning, but their use in engineering education remains less explored. Most AI applications in education focus on tutoring or writing support, whereas engineering education aims to develop conceptual understanding, reasoning skills, and the ability to interpret physical processes, which are essential competencies for engineers.

The contribution addressed this gap by examining how GenAI may support two recurring educational bottlenecks: (i) the need for cognitively demanding, exam-oriented preparation materials and timely, detailed feedbacks on students' projects and reports; (ii) instructors' limited time to continuously produce diverse, high-quality, conceptually rich learning material and to provide prompt, constructive feedback, especially on long technical reports. Consistently with the broader EduHackathon perspective, the studies position LLMs not as autonomous instructors but as support tools integrated within human-guided pedagogical workflows.

3.7.1. Study 1: AI-generated deep-reasoning questions for exam preparation

The first study explored the use of an LLM to generate deep-reasoning questions aimed at promoting higher-order thinking in exam preparation. A chapter from a geotechnical engineering textbook was uploaded to Google's NotebookLM, selected for its ability to ground outputs in user-provided sources. The system was prompted to generate questions requiring analysis, synthesis, comparison of mechanisms, and application of concepts, rather than factual recall. Prompts were iteratively refined to encourage scenario-based reasoning and justification of assumptions. The resulting questions were distributed as supplementary material for students preparing for examinations.

This approach extends previous EduHackathon uses of NotebookLM focused on concept explanation and textbook interrogation (Sections 3.3 and 3.5) by shifting the tool toward an assessment-oriented function, where the objective is not explanation but activation of reasoning processes under exam-like conditions. In this sense, the LLM acts as a generator of cognitive stimuli, rather than a provider of solutions.

Students generally perceived the AI-generated questions as challenging, exam-relevant, and helpful in identifying gaps in understanding. The situational framing of questions supported integration of theoretical knowledge with practical geotechnical contexts. From an instructional perspective, many questions required students to justify assumptions, interpret mechanisms, or link multiple concepts, thereby promoting deeper conceptual processing. However, limitations were observed. Some questions were excessively complex or ambiguously phrased, occasionally allowing multiple interpretations. Quantitative and calculation-based aspects were underrepresented, highlighting challenges that general-purpose LLMs still face in technical domains. This imbalance underscores the need for instructor oversight to filter, adjust, and complement AI-generated content. The process also revealed that instructors' prompt design strongly influences output quality, making educator involvement essential in shaping the pedagogical value of the material. By using the AI to generate questions rather than answers, the system supports active reasoning and helps mitigate concerns about "cognitive debt" associated with AI overreliance. The tool therefore stimulates, rather than replaces, cognitive effort.

3.7.2. Study 2: generative AI for feedback on student reports

The second study examined whether GenAI could assist supervisors in providing structured formative feedback on technical projects and reports. Previously graded student reports (high and low quality) were uploaded to NotebookLM, which was asked to produce comments on clarity, structure, and content depth. Students received both instructor-provided feedback and AI-generated feedback, and their perceptions were collected qualitatively.

The AI proved effective in extracting and organizing information and in identifying explicit gaps related to clarity and structure. It rapidly produced coherent, well-structured feedback. However, its comments remained largely surface-level. In contrast, instructor feedback was consistently perceived as more valuable, particularly for its ability to interpret technical content, evaluate the soundness of reasoning, and provide discipline-specific, experience-based insights. The AI lacked scientific judgment, could not interpret figures or data, and did not offer critical evaluation of methodological choices or engineering assumptions. Students described the AI feedback as "useful for readability but superficial," whereas instructor comments were seen as essential for improving the technical and conceptual quality of the work. Interestingly, sections the AI failed to interpret often corresponded to unclear writing, suggesting that AI misunderstanding functioned as a form of diagnostic clarity feedback. Thus, while AI feedback supported improvement in structure and readability, it did not replace the deeper evaluative role of the instructor. These findings are coherent with observations in other EduHackathon experiments involving complex instructional materials (e.g., Section 3.6), where LLMs demonstrated strong abilities in structuring and rephrasing information but limited capacity for domain-specific evaluation.

It should be noted that these studies were conducted within the normal teaching activities of the course. Owing to the limited time available and the need to provide students with authentic learning materials for their immediate educational needs, a controlled comparison between AI-generated and instructor-generated materials was not feasible. Consequently, the evaluation was based primarily on qualitative student feedback and instructor observations. Future studies should include controlled experimental designs and objective learning performance indicators to more rigorously assess the educational effectiveness of AI-assisted materials.

3.7.3. Educational implications

Together, the two studies suggest that LLMs can serve as cognitive scaffolding tools in geotechnical education. In the first case, AI-generated questions expand the range and diversity of reasoning-oriented preparatory material. In the second, AI-generated feedback provides a preliminary layer of linguistic and structural support that can be refined by instructors. Across both contexts, the LLM performs most reliably in tasks involving generation, restructuring, and linguistic mediation of knowledge, while remaining dependent on human oversight for epistemic evaluation, contextual interpretation, and professional judgment. Rather than replacing instructor-designed activities or feedback, AI outputs function best as a first-pass support layer within a hybrid workflow where educators ensure depth, correctness, and disciplinary rigor. Consistent with broader EduHackathon findings and the educational AI literature, the results reinforce the view of LLMs as partners in learning processes - tools that can enhance reflection, feedback cycles, and engagement when integrated with sound pedagogical design, but not substitutes for expert judgment.

3.8. Automated grading with LLMs in geotechnics

The presentation unfolded as a case study rather than a declaration of ready-made practice, with a clear disclaimer that no official grading was conducted and prototypes were used solely to explore feasibility under controlled conditions. The following strict boundaries were set from the beginning: transparency, human oversight, limited scope, continuous monitoring, and bias-aware design. As in the corresponding peer-reviewed article (Soranzo, 2025) the same arc is traced from concept to method to measured outcomes, detailing how cosine similarity and fine-tuned LLMs can be combined to support assessment while documenting their limits within an educational environment.

The work began by building the content scaffolding. A retrieval-augmented generation (RAG) pipeline (Lewis et al., 2020), drawing from a curated corpus of geotechnical sources, was used to generate questions and "correct" answers aligned with current practice, creating reliable anchors for evaluation and feedback. To train and test grading methods on a realistic spectrum of performance, synthetic student answers were produced with tailored ChatGPT prompts, deliberately spanning good, poor and off-topic responses; this strategy enriched the data with the variability necessary to teach a model the difference between nuanced correctness and plausible error. With those pieces in place, the team deployed a web application that offered students immediate feedback by comparing their responses to the reference answers via cosine similarity, assigning grades based on similarity thresholds and logging interactions to support analysis and iterative improvement (Fig. 2).

Two grading approaches were tested in parallel. Cosine similarity offered speed and interpretability, translating proximity of meaning into grade bands; yet, when evaluated against lecturer ground truth, it tended to blur distinctions, frequently misclassifying answers toward intermediate grades and underperforming on the extremes, as revealed by confusion matrix analyses and sub-0.5 metrics for precision, recall, F1 and accuracy. In contrast, an open-source LLM fine-tuned on triplets of Question, Answer and Grade showed a dramatic lift in performance,

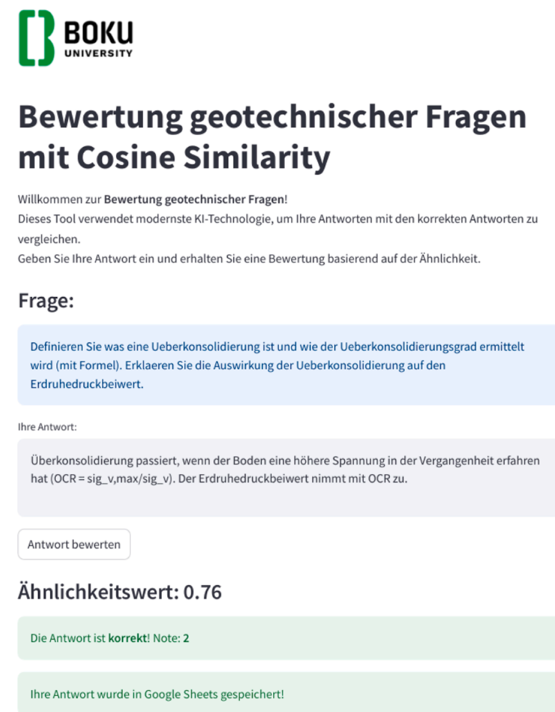

Fig. 2. Screenshot of the Web application for dynamic student feedback. The question "Define overconsolidation, how OCR is calculated and its influence on K0" is answered with "Overconsolidation occurs when the soil experienced a higher stress in the past (OCR = sig_v,max/sig_v). K0increases with OCR". The cosine similarity score and the grade achieved are 0.76 and 2, respectively, corresponding to "good" in the Austrian grading system (Soranzo, 2025).

achieving near-perfect classification on training and strong generalisation on testing, with accuracy and F1 scores hovering around 98% and 97.5%, respectively and substantially fewer errors across classes than the similarity-based baseline.

The results on real student data brought nuance. When automatic grades were compared with lecturer assessments, discrepancies emerged, likely rooted in unseen questions and the model's limited confidence on certain responses, underscoring the importance of broader, representative data and the value of active learning via continued collection through tools that students actually use. The team also extended the approach to anonymised technical reports on foundation and levee design, obtaining moderate performance on validation: accuracy around 0.714 and F1 near 0.656. These results are promising but they also show the need for stronger pre-processing, augmentation and possibly different architectures before relying on the system in practice.

Beyond numbers, the educational implications were twofold. First, RAG-enabled generation proved useful for creating tailored, source-linked learning materials; second, automated grading systems showed potential to deliver consistent, objective assessments and, crucially, immediate feedback that helps students identify and close knowledge gaps well before exams. The study's ethical spine ran through privacy-preserving anonymisation, informed student consent, fairness in training, transparency of decision-making, and clear documentation of capabilities and limitations. These principles align with evolving EU frameworks that classify educational AI systems as high-risk and therefore demand human oversight and robust governance (European Union Regulation, 2024). In addition, the work emphasised interpretable outputs, including similarity scores and confidence distributions, to support accountability in grading decisions.

The final message mirrored the presentation's prudence. Automated grading of open-ended geotechnical questions with LLMs is feasible if supported by sufficient, diverse data and if deployed through carefully designed prototypes that both serve learners and bootstrap the training pipeline. Report grading appears promising but remains a step away from classroom-ready reliability. Clear legal requirements, institutional policies and instructor training are essential to move from demonstration to responsible use.

4. Floor discussion

The floor discussion started with the presentation of the main results of a short survey that circulated a few days prior to the workshop among 3FOMLIG registered participants with different personal experiences in Gen AI use (Fig. 3). Table 2 synthesises the main answers of the 22 respondents. Overall, the respondents find Gen AI tools useful for enhancing both teaching and learning, especially in coding, content creation, and conceptual explanation, and a valuable complementary tool that can enrich teaching and learning through interactivity, personalization, and creative problem-solving. At the same time, the need for responsible usage, proper verification, ethical guidance, and critical engagement is emphasized, and all these activities are perceived as essential to potentially enhance students' learning through AI tools rather than hinder it.

The floor discussion that followed was very well-participated by students, industry representatives, and researchers alike, leading to an overall perception of a very strong interest by all these categories. Direct past and current experiences, as well as speculations on the future role of AI and LLMs in geotechnical education, were shared freely and constructively. It was particularly interesting to witness the perspective of undergraduate and postgraduate students both as geotechnical learners and digital citizens, as well as their interactions with educators with short- and long-term experience. AI is indubitably perceived as an inevitable technological shift that can accelerate learning in some instances, e.g., assisting with coding, yet it was also clearly stated that it cannot replace fundamental laboratory work or the human supervision necessary to identify model hallucinations. The widespread use of AI also requires updated assessment methods beyond traditional reports, for instance maintaining or increasing in-person exams whenever appropriate. Moving from apprehension to action, it was pointed out that both teachers and students must become proficient in AI literacy to ensure the technology serves as an enhancer of human expertise rather than a substitute for foundational knowledge. To this aim, a shift of classroom time toward active learning was also argued, calling for educators to explicitly work with students on the how to properly use LLMs with ethical safeguards and verification built in, for instance making model assumptions and basic statistical tools intelligible to them or promoting conversations with LLMs as learning artifacts.

Based on the floor discussion and the survey results presented, several themes and keywords emerged. In relation to core educational benefits, the following keywords can be highlighted: accelerated learning; personalized education, i.e., the ability to tailor learning to individual student needs; coding assistance; interactive support for students and teachers. On the side of concerns, the most cited ones were: potential loss of critical thinking capabilities, for instance in connection to students relying too heavily on AI outputs or using LLMs to summarize didactic material rather than engaging in critical analysis; possible long-term effects on students' foundational skills, including reading and writing; hallucination, i.e., the possibility of models generating and disseminating incorrect information. Requirements for effective use included the following needs: supervision by human educators; clear guidelines on how and when to use these tools; feeding Gen AI tools with proper domain-specific information; effective prompt engineering.

5. Concluding remarks

The EduHackathon proved to be a very interesting opportunity to address the multi-faceted current and expected future role of LLMs and other Gen AI tools for geotechnical education purposes. The variety of topics contributed by the working group participants is in itself a statement to the potential of these tools to enhance geotechnical education if suitably planned, correctly calibrated, and competently assessed. The presentations and the floor discussions highlighted the centrality of the interaction between humans and LLMs and rejected the view of a unilateral flow of information and a passive role of learners. Therefore, there appears to be significant space for further research on the role of generative AI in geotechnical engineering educational. To this aim, the ongoing transformation which is guiding the discipline towards an increasingly data-centric paradigm offers a favourable background and provides a further stimulus to a competent, critical, and open-minded adoption of LLMs and Gen AI tools in geotechnical engineering education. Finally, considering that LLM technologies are evolving rapidly, the findings from the eight experiments and the floor comments from the third Workshop on the Future of Machine Learning in Geotechnics—all conducted in 2025—must be temporally contextualized.

CRediT authorship contribution statement

Michele Calvello: Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization. **Emilio Bilotta:** Writing – review & editing, Writing – original draft, Investigation. **Alessia Cuccurullo:** Writing – review & editing, Writing

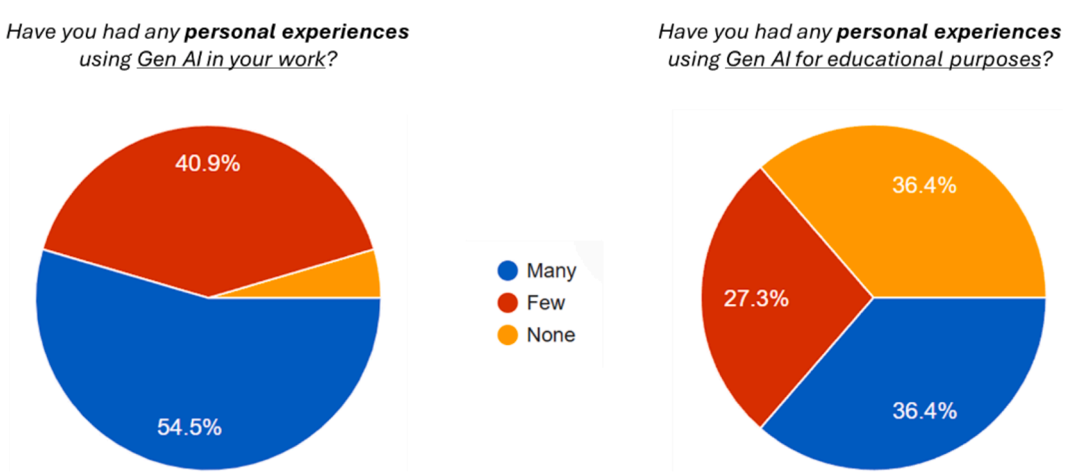

Fig. 3. Experiences in using Gen AI for work and educational purposes of the 22 respondents to a survey sent to 3FOMLIG registered participants a few days prior to the workshop.

Table 2
Main results of the survey sent to 3FOMLIG registered participants.

Survey question (number of replies)	Main themes addressed	Mentioned in the replies
1. If you have used Gen AI for educational purposes, tell us something about your experiences (14 answers)	Common Tools Used	ChatGPT (by far the most mentioned), DeepSeek, NotebookLM, GitHub Copilot, Bing Copilot, Canva AI, Genspark, Doubao, Local LLMs
	Primary Educational Uses	Teaching support: preparing lecture materials, slides, quizzes, and interactive exercises Learning enhancement: clarifying difficult concepts, summarizing notes, exploring new topics Coding assistance: debugging, generating or checking Python scripts, especially in geotechnical/data science contexts Research support: literature reviews, generating ideas, developing or refining methodologies Visualization and creativity: generating images, diagrams, or mind maps
	Perceived Benefits	Saves significant time (typing, coding, preparing materials) Enhances creativity and clarity in teaching materials Provides accessible examples and explanations Supports individualized or interactive learning experiences Promotes efficiency and engagement
	Main Concerns / Limitations	Accuracy and reliability: output sometimes incorrect or inconsistent Need for verification: users emphasize fact-checking, especially in engineering contexts Language inconsistency: occasional variation in quality and coherence Overreliance risk: AI should assist, not replace critical thinking or expertise
2. How do you think Gen AI can be usefully adopted, in some way, within University courses? (22 answers)	Pedagogical Integration	As an interactive assistant to guide students through complex topics or coding exercises To personalize learning, offering adaptive explanations and feedback To generate exercises, simulations, case studies, and alternative explanations As a support tool for writing, data analysis, and literature review
	Skill Development Goals	Foster critical thinking, ethical awareness, and AI literacy Encourage reflection on AI-generated content (e.g., comparing AI vs. human work) Develop prompting and verification skills
	Conditions for Effective Use	Must include clear guidelines, supervision, and transparency Should align with learning objectives and academic integrity standards Requires training for both students and educators to use tools responsibly
	Use Cases Mentioned	Coding support for non-programmers Simulation and design tasks in engineering Visualization (images, videos, summaries) Student engagement and communication improvement
3. What are the main benefits and threats to students (undergraduate and postgraduate) stemming from the use of GenAI? (22 answers)	Main Benefits	Accelerated learning and efficiency: faster access to information, feedback, and coding help Personalized education: tailored exercises, adaptive explanations, and individualized learning paths Enhanced research capability: improved literature review, data handling, and conceptual synthesis Creativity and curiosity: AI as a catalyst for exploration and idea generation Accessibility: supports diverse learners and helps those with disabilities
	Main Threats	Overreliance and loss of critical thinking: risk of accepting AI outputs without understanding or verification Erosion of foundational skills: decline in reading, writing, and problem-solving ability Academic integrity risks: plagiarism, lack of originality, unethical use of AI-generated work Knowledge inaccuracies: confident but incorrect outputs leading to misinformation Digital divide: unequal access to AI tools could widen learning inequalities Mental and cognitive impact: reduced attention span, overdependence, and possible loss of curiosity

– original draft, Investigation. **Ana Heitor:** Writing – review & editing, Writing – original draft, Investigation. **Charles MacRobert:** Writing – review & editing, Writing – original draft, Investigation. **Enrico Soranzo:** Writing – review & editing, Writing – original draft, Investigation. **Marco Uzielli:** Writing – review & editing, Writing – original draft, Investigation. **Stephen Wu:** Writing – review & editing, Writing – original draft, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Calvello, M. (2025). Teaching soil mechanics core concepts: Mapping the structure of university textbooks worldwide. In *Proc. International Conference on Geotechnical Engineering Education GEE 2025 "Charting the path toward the future* (pp. 387–400).

Das, B. M., & Sobhan, K. (2017). *Principles of geotechnical engineering, si edition* (9th ed.). Boston, USA: Cengage.

European Union Regulation. (2024). (EU) 2024/1689 of the European parliament and of the council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (artificial intelligence act) (text with EEA relevance). PE/24/2024/REV/1. Accessed 27/03/2026 http://data.europa.eu/eli/reg/2024/1689/oj.

Felder, R. M., & Solomon, B. A. (2008). Learning styles and strategies. *Educational Technology Research and Development, 56*(3), 313–332.

Fleming, N., & Baume, D. (2006). *Learning styles again: VARKing up the right tree!, educational developments* (pp. 4–7). SEDA Ltd.

Freire, P. (1970). *Pedagogy of the oppressed*. Herder and Herder.

Gaikwad, H. V., & Patil, S. K. (2017). Analysis of learning styles of engineering students for improving engineering education. *Journal Engineering Education Transform, 30*, 44–49.

Isaza Domínguez, L. G., Velasquez Clavijo, F., Robles-Gómez, A., & Pastor-Vargas, R. (2024). A sustainable educational tool for engineering education based on learning styles, AI, and neural networks aligning with the UN 2030 Agenda for Sustainable Development. *Sustainability, 16*, 8923.

Kolb, D. A. (1984). *Experiential learning: experience as the source of learning and development* (p. 41). Englewood Cliffs: Prentice-Hall.

Neves, J., Freeman, J., Stephenson, R., & Rowan, A. (2025). Student academic experience survey. Advance HE report. Accessed 27/03/2026 https://www.advance-he.ac.uk/knowledge-hub/student-academic-experience-survey-2025.

Kosmyna, N., Hauptmann, E., Yuan, Y.T., Situ, J., Liao, X.-H., Beresnitzky, A.V., … Maes, P. (2025). Your brain on ChatGPT: accumulation of cognitive debt when using an AI assistant for essay writing task. Accessed 27/03/2026 https://arxiv.org/abs/2506.08872.

Lewis, P. S. H., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., … Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks Accessed 27/03/2026 https://arxiv.org/abs/2005.11401.

Liu, C. (2025). A comprehensive review of applications of AI technologies in higher engineering education. *Discover Education, 4*, 528.

Lopez-Gazpio, I. (2025). Integrating large language models into accessible and inclusive education: Access democratization and individualized learning enhancement supported by generative artificial intelligence. *Information, 16*(6), 473.

Matos Fernandes, M. (2024). Mecânica dos Solos. *Conceitos e princípios fundamentais* (6th ed.). Porto, Portugal: U.Porto Press.

Pinho-Lopes, M., & Powrie, W. (2020). Feedback to students on soil mechanics laboratory reports – Why use virtual technology if you can have a productive real dialogue? In M. Pantazidou, M. Calvello, & M. Pinho-Lopes (Eds.), *Proc. 5th International Conference on Geotechnical Engineering Education, GEE 2020* (pp. 228–240). ISSMGE.

Rittle-Johnson, B., & Schneider, M. (2015). Developing conceptual and procedural knowledge of mathematics. In R. Cohen Kadosh, & A. Dowker (Eds.), *The oxford handbook of numerical cognition*. Oxford University Press.

Santamarina, J. C. (2015). Short communication: (What) to teach or not to teach – that is the question. *Geotechnical Research, 2*(4), 135–138.

Shetye, S. (2024). An evaluation of Khanmigo, a generative AI tool, as a computer-assisted language learning app. *SALT Journal, 24*(1), 38–53.

Simelane, P. M., & Kittur, J. (2025). Use of generative artificial intelligence in teaching and learning: Engineering instructors. *Perspectives. Computer Applications in Engineering Education, 33*, Article e22813.

Soranzo, E. (2025). Large language models for automated grading in geotechnics. *Machine Learning and Data Science in Geotechnics, 1*(1), 124–144.

Streveler, R. A., et al. (2014). Conceptual change and misconceptions in engineering education. In *Cambridge handbook of engineering education research, Ch5* pp. 83–102).

Timms, M. J. (2016). Letting artificial intelligence in education out of the box: Educational cobots and smart classrooms. *International Journal Artificial Intelligence Education, 26*, 701–712.

UNESCO. (2019). How can artificial intelligence enhance education?. Available at https://www.unesco.org/en/articles/how-can-artificial-intelligence-enhance-education#:~:text=AI%20has%20the%20potential%20to,order%20to%20improve%20learning%20outcomes accessed 27/03/2026.

Wesley, L. (2015). Short communication: (What) to teach or not to teach – from theory to practice. *Geotechnical Research, 2*(4), 139–147.

Wu, S., Otake, Y., Mizutani, D., Liu, C., Asano, K., Sato, N., Saito, T., Baba, H., Fukunaga, Y., Higo, Y., Kamura, A., Kodama, S., Metoki, M., Nakamura, T., Nakazato, Y., Shioi, A., Takenobu, M., Tsukioka, K., & Yoshikawa, R. (2024). Future-proofing geotechnics workflows: Accelerating problem-solving with large language models. *Geotechnical & Geological Engineering, 42*, 307–324.

Yusuf, H., Money, A., & Daylamani-Zad, D. (2025). Pedagogical AI conversational agents in higher education: A conceptual framework and survey of the state of the art. *Educational Technology Research and Development, 73*, 815–874.

Zhang, Z., Zhang-Li, D., Yu, J., Gong, L., Zhou, J., Hao, Z., Jiang, J., Cao, J., Liu, H., Liu, Z., Hou, L., & Li, J. (2025). Simulating classroom education with LLM-empowered agents. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics* (pp. 10364–10379). Association for Computational Linguistics.