AI, Assessment, and the Student and Educator Experience: Reflections from a Celebrate Learning Week Workshop

Written By Emma Davy (PhD) and Noureddine Elouazizi (PhD)

This reflective essay highlights the key takeaways and reflections from the 2024 Celebrate Learning Week workshop on AI, assessment, and the student and educator experience. Below, we also share resources to support those in the teaching and learning community who may be grappling with the challenges of Generative AI tools and how they may affect their classes.

Context

Generative AI (GenAI) has been front of mind for many educators since interest in these tools exploded in November 2022 with the roll-out of the popular ChatGPT. In the face of unmitigated "hyping" in the media about the exaggerated capabilities of AI, many educational institutions (including UBC), faculties, departments, and individual educators have had much to consider and ponder when it comes to the use (or not) of Generative AI tools in their classroom. Do such tools add tangible educational value? Should faculty use these tools? Should students use these tools? If yes, *can* students even use these tools *effectively*? Should training to use these tools (and identify their limitations) be provided for educators and students? These questions, if left unattended to, might increase anxiety and even spark panic about the perceived (and at times misplaced) assumptions about the "god and/or ghost in the machine" imaginative powers attributed to Natural Language Generation systems (the type of AI that Generative AI is part of).

Intending to counterbalance the hype and enable an evidence-based and expertise-informed discourse around the value-add of GenAI for science education, we planned a workshop to address some of these questions and better equip educators to make informed decisions about Generative AI and their classes. We leveraged our combined expertise to develop this workshop: Noureddine has an extensive background in AI and is currently an Artificial Intelligence Senior Strategist at the Faculty of Science's Centre for Learning and Teaching (Skylight), and Emma is a Science Education Specialist in Chemistry (cross-appointed at the Department of Chemistry and Skylight) and teaches CHEM 300: Communicating Chemistry. Since Emma teaches a scientific writing and communication course, she has had to consider Generative AI tools carefully. Currently, Emma and her colleague, Chris Addison, hold a SOTL Seed Grant to support research into Generative AI and this discipline-based writing course. Emma and Chris have chosen to incorporate Generative AI training in CHEM 300 and allow students to use the tool provided they declare their use and provide their chat logs upon submission.

Celebrating Learning Week

We decided to offer the first run of the workshop as part of UBC's Celebrate Learning Week that took place from May 6th to 10th, 2024, at the UBC-V and UBC-O campuses. The theme of this Celebrate Learning Week was "Remembering the Human in the Loop" with the following description:

Generative AI is changing the way we think about education – it brings exciting possibilities and poses new challenges. But as we turn out attention to technology, remembering the people at the heart of our mission becomes increasingly vital. At the 2024 Celebrate Learning Week, we will explore how to retain our focus on the human aspects of education, while absorbing and exploring these new technological landscapes.

The full schedule of Celebrate Learning Week 2024 can be found here: https://celebratelearning-new.sites.olt.ubc.ca/files/2024/04/CLW-2024-Schedule-Final.pdf.

AI, Assessment, and the Student and Educator Experience

We delivered our session in person on the first day of Celebrate Learning Week (May 6th). We chose in-person delivery for two reasons: (1) we wanted to be able to easily mingle with participants and welcome them to engage with us, and we felt this would be more authentic in person, and (2) with only two facilitators and a maximum attendance of 55 people, we felt hybrid delivery would make it challenging for us to provide attentive support to all attendees. We asked participants to bring their laptop computers and ensure they had signed into a Generative AI tool. While our workshop was tuned toward participants from the Faculty of Science, everyone was welcome. We designed this session with an approximately 50/50 distribution of active learning components and presentation elements.
The schedule of our workshop was as follows:

1) Introduction (15 minutes): A (brief) discussion of AI, large language models (LLMs), and consequences to the educator (led by Noureddine).
2) Finding the limits of LLMs (15 minutes): Breakout activity to learn about the limits of LLMs (led by Noureddine).
3) Example from different fields (10 minutes): Some examples from Chemistry, Statistics, and Mathematics about using AI in student work and student assessment (led by Emma and Noureddine).
4) Short break (5 minutes)
5) Your own course/activity design breakout activity (30-35 minutes): Working with the provided guiding worksheets, consider how you can use AI to design your own assessment activities or how you can design an activity that will be assessed where students use AI.
6) Final wrap-up (10-15 minutes): Suggestions, questions, and future working groups!

During the introduction, Noureddine provided an overview of what AI encompasses (including, but not limited to: natural language processing (NLP), knowledge representation, machine reasoning, computer vision, robotics, and machine learning) and specified that GenAI is one of the many subareas in AI. Many

of the tools that educators at UBC will be familiar with are natural language generation (NLG) systems and question-answering (dialogue/chat) systems (both of which fall under the umbrella of NLP). Zooming in on AI for/in education, Noureddine highlighted that the use of AI in/for education systems is not new and that classical efforts trace back more than 30 years. Examples range from intelligent tutoring systems (Conati (1999); Conati & VanLehn (2000), a.o.), automated curriculum design (Bull et al. (2018)), and AI-powered writing assistance (Li et al. (2020)), amongst others. With this framing about the value-add and important ways in which AI can enable education and learning, Noureddine focused the presentation on highlighting the limitations of LLMs with a specific focus on the architectural and inherent limitations in the design of LLMs, and mapped out the restrictive (micro and macro) implications of such limitations on assessment for science education. Some of the macro implications for educators, include:

1) *Understanding the limitations of AI systems*: It is vital that the educator can delimit situations where AI can help streamline part of the logistics of learning (e.g., information retrieval and organization) from situations where AI can hinder critical thinking and sense of inquiry, both of which are fundamental to thinking like a scientist.
2) *Shifts in the role of the educator*: The use of AI might induce a shift in the role of the educator, shifting the role more towards facilitating learning, scaffolding learning, and enabling critical thinking.
3) *Narrowing of curriculum*: By design, AI algorithms are designed to assess the easily quantifiable skills and knowledge aspects. This might narrow the aspects of the curriculum that enable critical thinking and deeper understanding and synthesis of complex concepts.
4) *Epistemic atrophy*: It is necessary to "bake" into the design of the assessment rail-guards to guard against the "epistemic atrophy", which might be caused by excessive and unprincipled use of AI to "consume" information (information is not insight).
5) *Depersonalization*: It is necessary to mitigate the depersonalization impact that could be created by AI-driven assessments that might lack the personal touch and individualized attention. AI algorithms (especially black box algorithms type) may prioritize objective metrics and quantifiable data in assessment processes, neglecting the subjective aspects of learning.

Our breakout activity allowed participants to play with the limitations of LLMs with a guided worksheet and activity. During this activity, a group of ~40 participants (a mixture of faculty and staff) were interacting with each other and excited to work on testing the limitations of the LLMs. The room was lively with people discussing their results and during our post-activity debrief, our conversation was robust. Participants were easily able to identify (reproduce) the LLMs' limitations and noted the difference between the paid and unpaid versions of such tools.

After the breakout activity, we shared experiences from the "pedagogical trenches", and we highlighted how several pioneering educators in the Faculty of Science examined and integrated the use of LLMs as part of their pedagogy to deliver teaching and enable the learning experiences of students across different disciplines. We focused on three classes from the Faculty of Science at UBC: CHEM 300 (Communicating Chemistry), MATH 220 (Mathematical Proof), and STAT (DSCI 100 (Data Science)). For CHEM 300, the instructors offered students basic Generative AI training (focused on the basics of how

the tool(s) work and how to engineer a prompt effectively) and completed an assignment using Generative AI. For MATH 220 (Mathematical Proof), the instructor experimented with tasking the LLM to work on several math problems. Some of the observations that the instructor established include: (1) when an LLM is asked to prove something that isn't true, it duly obliges, and (2) in other responses from the LLM, the proof could have been finished off way earlier, but the LLM continued with unnecessary wrong work. Another group of instructors conducted a study to evaluate the abilities of ChatGPT in the context of statistics and then develop guidelines for effective assessments in the ChatGPT era. They observed that the LLMs do answers some of the basic problems, but fail in some instances of elaborated problem solving. They decided to allow students to use the LLMs, and the instructor would then assess the performance of the student based on the part of the answers that did not depend on the use of the LLM, hence teasing apart and assessing separately the problem-solving segments provided by the LLM, and the ones provided by the students.

These stories offered examples of how different educators had incorporated Generative AI into their classes. This culminated in our final activity: allowing our participants to work with each other and with the facilitators on either using Generative AI to design assessments or to design activities where students use Generative AI themselves. The participants were provided either an educator-centred handout or a student-centred handout to structure the activities.

To finish our session, we made time for participants to ask questions and fill in a Padlet with further questions and queries about Generative AI in science education. We created a OneDrive folder for further resource sharing and provided our workshop handouts and slides by email. We were thrilled by the level of engagement among participants and excited by this first offering of this workshop!

Some Reflections from the Workshop

From the Facilitators

With this workshop, we deliberately sought to strike a balanced perspective regarding how to approach the use of AI for assessment in the context of science education, such that:

1) AI is defined in a way that relates and translates to the context of science education.
2) The potentials of using AI are clearly identified. There are many opportunities to use AI to enhance aspects of assembling and designing learning contents, enabling interaction and engagement for learning, delivering learning, and assessing learning.
3) The limitations of using AI are identified for science education. It is only by appreciating the limitations of what AI tools can(not) offer that an educator would be able to set up realistic expectations for how AI can support teaching and learning.

The list of the attendees included people with a wide array of expertise and experience in AI: scholarly authorities in the AI field, faculty members who have experimented with the use of AI for teaching, faculty members who have never used AI, PhD students and post-docs working in the field of AI, administrators, and support staff. Having such a diverse group of attendees indicates that interest in this

offering cuts through diverse segments of stakeholders at UBC, in general, and at the Faculty of Science, in particular. However, these different groups might benefit from some form of a stratified, discipline-contextualized, and customized approach that caters to varying levels of pre-existing AI familiarity and knowledge.

We felt the amount of time that was allocated to the workshop with all its segments was not sufficient to delve into all the details. The session went by quickly, and did not allow enough time for the attendees and the facilitators to engage in more activities and discussions. In the future, the impact of the workshop could be maximized by offering it for a half-day or a full day with enough time allocated to interactive activities, such as group work, hands-on activities, and Q&A sessions.

Some of What Seemed to Have Caught the Attention of Workshop Attendees

The focus on "stories from the trenches" of science education pedagogy helped attendees relate to their experiences with the use of AI for education in their own contexts. We also noted:

1) The use of an evidence-based approach to integrate AI as part of assessment and learning, as was implemented by Emma Davy and Chris Addison, demonstrated one of the innovative and principled ways to use AI for assessment in the context of science education.
2) The use of AI in the context of Statistics courses (by Joel Östblom and team) and Mathematics courses (by Nahid Walji) highlighted the importance of benchmarking the (in)abilities of an LLM in resolving some of the problems in Stats and Maths, and therefore, the importance of integrating LLMs in a principled and guided manner to assess students' learning.
3) The use of AI and NLP-informed diagnostics, developed by Noureddine Elouazizi, to delimit the inherent (in)abilities in the reasoning of an LLM helped the attendees appreciate the need to understand the limitations of an LLM and the consequences of such limitations on the use of AI for assessment, and therefore, the need to handle the use of LLMs for assessment carefully to ensure a positive impact on teaching and learning experiences.
4) The interactive and question-and-answer components of the workshop fostered the participation of different attendees, enabling meaningful engagement and discussion of the way faculty members used AI as part of their teaching, the ways they gauged the impact and the ways they appreciated the potential and limitations of using AI in the context of higher education.
5) Different attendees expressed the desire that we offer more of the same type of workshops at the institution.

It is our intention to develop and execute a program that would strive to meet such needs.

Contact Information:

Dr. Emma Davy (Chemistry/Skylight): edavy@chem.ubc.ca
Dr. Noureddine Elouazizi (Skylight): noureddine.elouazizi@science.ubc.ca