Teaching with Generative AI: moving forward with content creation

Thomas Daniel Ullmann[1], Duygu Bektik[1], Chris Edwards[1], Christothea Herodotou[1], Denise Whitelock[1]

[1] Institute of Educational Technology, The Open University, United Kingdom; thomas.ullmann@open.ac.uk, {firstname} dot {lastname} at open.ac.uk

Correspondence: Thomas Daniel Ullmann: thomas.ullmann@open.ac.uk

Abstract:

Generative AI is widely available and has raised the expectation that it will impact Education. Models, such as ChatGPT, can quickly produce plausible texts on a wide range of topics, and this capability may be of potential use in course content production. This paper selects several important course content production tasks, describes the prompts used, and assesses the quality of the automatically generated texts by a team of experts. Across all tasks, the Generative AI produced content that could help solve specific tasks by aiding with brainstorming, creating outlines, and adhering to particular writing guidance. In all cases, the content required adjustments and checking by human experts.

Keywords: course production; course content; generative AI; ChatGPT; natural language generation; educational technology

Introduction

Generative Artificial Intelligence, with its easy-to-use interfaces and widespread availability, has led us to investigate its potential for course content production. This paper provides expert reflections on the quality of AI-generated content for typical tasks during course production. It showcases prompts for supporting several course content production tasks, evaluates the quality of prompt completions, and outlines relevant prompt engineering strategies.

During the last few years, an increasing use of AI in Education, especially in developed countries, has been observed (Tahiru, 2021). Empirical studies have mainly focused on Higher Education and K-12 and investigated AI applications, including chatbots and expert systems, intelligent tutors, machine learning, personalised learning systems and visualisations in K12 and Higher Education (Zhang & Aslan, 2021).

Within the last year, Generative AI and tools such as ChatGPT became widely available and started another AI boom. Their capability to produce plausible text about a wide variety of topics in an interactive way sparked interest within the educational community as it potentially offers a variety of possibilities for Education (Bektik et al., accepted; Sharples, 2023). For example, Liu et al. (2023) reported that ChatGPT can support teachers in generating exercises, quizzes, and scenarios for student evaluation. The review by Chen et al. (2023) on AI in Education, using a qualitative research approach, found that AI systems have enabled the customisation and personalisation of curriculum and content according to students' needs, leading to improved learning experiences and overall educational quality.

Furthermore, various companies and universities have developed AI-powered tools, all of which claim to support the generation of curriculum. For example, several Moodle Plugins now support Generative AI. The EduWeaver research demonstration creates course content based on a few parameters (Aneesha, 2023). Blackboard's AI Design Assistant, TeacherMatic, Course AI, Coursera AI, and others offer support tools for teachers.

Despite potential benefits, AI does come with significant challenges, such as high cost and scalability, requiring new literacies and actionable guidelines for educators, ethical and data privacy issues related to student data, and inappropriate uses of AI (Zhai et al., 2021; Zhang & Aslan, 2021). ChatGPT, in particular, raises issues of integrity, accuracy, and credibility, such as producing scientific essays from a mixture of true and fabricated information (Alkaissi & McFarlane, 2023). Such observations highlight the importance of human input and expertise in moderating and validating AI outputs and the need for clear policies and guidelines on how AI-generated content should be managed (Adeshola & Adepoju, 2023).

The introduction of Generative AI has the potential to, at least, disrupt the processes in place for creating curriculum. Each institution will have its own framework in place and need to consider whether this technology has a role and, if so, what this role might be and what benefits can be achieved for an institution, its staff, and, of course, the students who study the courses.

The Open University is one of the largest online providers of Higher Education globally and an early developer of a team approach to curriculum production. This method brings together academics and other colleagues to produce tested, high-quality learning materials in various formats. Within the team, there have been several consistent roles, such as the academic lead, authors, editors, artists, and reviewers, and new roles have been created as formats evolved. These include multimedia developers and programmers. One of the main reasons for using this approach to curriculum development is that each course is intended to have a life of several years, and the investment is made to ensure high quality from the start.

The use of learning outcomes (LOs) became a requirement during the 2000s, and in 2013, the university codified the process of learning design (LD). This was partly to make it easier for colleagues, particularly those new to curriculum development, to make the right contribution. It also ensured the full integration of the constructive alignment (Biggs, 2014) between LOs, materials, and assessment and more rigorous consideration of the range of learners through personas. LD supports the full range of courses, including continuing professional development (CPD), microcredential, short, undergraduate, and masters level courses. As microcredentials are the newest format of courses, each of ours with 7.5 ECTS credits, we used examples of steps from creating these courses within this paper (see Table 1). This paper focuses on a few important curriculum development elements to begin revealing the value of Generative AI to Higher Education.

Table 1: Course production elements

Course production elements	Brief description
The big question and learning outcomes	What is the course's big question? What are course learning outcomes?
Assessment strategy	How many assessment points are needed? What is the topic and format of the assessment? What are the grading criteria?
Course outline	What is the big question of each week? Which topics are covered in each week?
Course content and activities (weekly)	Types of activities following the Learning Design (LD) framework Content format: e.g., audio, video, text, quiz, forum.
Assessing and enhancing inclusive language	Use prompts from the Inclusive Curriculum Design Tool to assess the quality of produced material.
Respond to reviewers' comments and queries.	Improve content accordingly
Review content for permissions (media, etc)	Acknowledge the use of GenAI if applicable.

Copy editing	Final check for typos, tone, style, etc

The literature review and the latest developments of eLearning vendors showed that several researchers and practitioners expect that Generative AI will influence teaching practices, including course material production. However, it is unclear how Generative AI can support teachers to create course content. In this paper, we show how Generative AI can support important elements of course content production and report on expert evaluations of the quality of the generated content.

Method

For each course production task outlined in Table 1, we envisioned a role for the AI within a course writing team. Our aim was not to replace the course writing team but to explore tasks where AI could potentially be of use. For each task, one of the authors generated and adapted several prompts and produced a first list of observations. We instantiated ChatGPT via Microsoft's Azure OpenAI with version 4.0 of ChatGPT (1106-Preview, standard deployment, capacity of 10K tokens per minute, temperature of 0.5, max tokens set to 800, top p set to 0.95, frequency penalty and presence penalty were set to 0, and there was no stop flag). We used ChatGPT via the university's approved Microsoft Azure subscription, as this guaranteed that copyrighted and otherwise protected university materials would not be used to train future models of ChatGPT. We did not fine-tune the model on the university data or use Retrieval Augmented Generation (RAG), meaning that all contents generated come from ChatGPT, not the university.

For each task, we constructed prompts for ChatGPT. A prompt is a message that is sent to the AI engine and to which the AI generates a completion. The completion is the response from the AI in continuation of the prompt text. We wrote several prompts for each task until we found a usable completion. Not every prompt returned a satisfying response. For example, the completion was vague or captured a different intent. Therefore, some experimentation was required to construct successful prompts. Our prompt engineering took into account available general guidance but also took inspiration from domain-related prompts publicised by other researchers. In line with general guidance, we used the following principles: using detailed information, providing examples, splitting complex tasks into several smaller tasks[1], chain of thought prompting (i.e. provision of recipe-like instructions), and adding/repeating instructions at the end of a prompt[2]. We also consulted domain-specific prompts, such as those suggested in the GitHub repository 'Prompts for Education, ' which showcases several prompts, such as a prompt for planning lessons or assignment ideation. The domain-specific prompts were particularly useful as they provided a sense of the structure and depth of a good prompt.

Once this phase of the work was completed, the new text was shown to a course production expert, an academic expert, two Learning Design experts, and an AI expert, who commented on the quality of the prompts and their suitability in teaching and the initial observations and agreed on their interpretations. The results section illustrates the prompts for selected tasks, a summary of the completion, and the main observations shared by the expert team.

Results

We show generative AI prompts and responses to three of the elements of course production (see Table 1), which are about generating big questions, learning activities, and inclusive curriculum. Due to the word limit, we selected only three of the cases that highlight and discuss important characteristics of generative AI pertinent to the context of course production.

Task: Big questions of a course and learning outcomes

The topic chosen for the course was Data Science. One of the first tasks among a physical writing course team is to discuss the salient content and define the learning outcomes. Here, we envisioned the AI producing an initial

[1] https://platform.openai.com/docs/guides/prompt-engineering/six-strategie

[2] https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/advanced-prompt-engineering?pivots=programming-language-chat-completions

list of big learning questions and learning outcomes for the course team to review. The prompts and responses from ChatGPT to illustrate the progress made in this respect are produced as a dialogue in abbreviated form below.

We trialled several prompts, such as one big prompt for the whole task or using brief descriptions. After some experimentation, we found that using two prompts and adding extra information, such as asking for generating examples, yielded better completions.

The first prompt was: 'For a data science apprenticeship course, I am looking for big ideas in this area to hook my students in. Can you give me examples of such ideas?'. ChatGPT's response started with this opening statement: 'Certainly! Data science is a dynamic and impactful field with many big ideas that can captivate students. Here are several concepts that could serve as hooks to engage your apprentices' and then listed fifteen big Data Science ideas such as 'Ethics and Bias in Data: Discuss the ethical considerations of data collection and analysis, including privacy concerns and the potential for Bias in algorithms that can lead to discrimination', or 'Quantum Computing and the Future of Data Science'.

The second prompt then drilled into the generated Ethics and Bias topic: 'Can you create learning outcomes for the topic Ethics and Bias in Data?' and ChatGPT responded with eight examples of learning outcomes, such as 'Understand the Ethical Principles in Data Science' with the two learning outcomes 'Recognise the fundamental ethical principles that guide data science practices, including respect for privacy, consent, transparency, and accountability', and 'Discuss the importance of ethical considerations throughout the data lifecycle, from collection to analysis and reporting'. Another suggested learning outcome was, for example, 'Reflect on Personal and Professional Responsibility'.

The team of experts concluded that all the presented topics did express big Data Science ideas and that a course production team could use the generated topics as an initial set of potential topics for a course. They also noted that not all topics would be suitable for a specific course, and the course creator would need to decide whether to include or exclude a topic, such as 'Quantum computing'.

The second prompt created learning outcomes for one of the ideas, 'Ethics and Bias in Data'. The experts concluded that AI developed several plausible learning outcomes that may help the course creator with the ideation of learning outcomes. They also noted that the learning outcomes are not written to the standards required by the university and would need to be changed, for example, to adhere to the learning outcomes for the qualification or the levels framework.

Task: Create learning activities

During course production, the learning designers encourage the course team to use a variety of learning activities, especially more active activities, such as communicative or practice activities. For this task, we envision the AI proposing learning activities for a given topic that follows a specific Learning Design framework (Rienties & Toetenel, 2016) for the course team to consider.

After trialling several prompts, we found three successive prompts as useful for the task. The first prompt set up the interaction by asking the teacher about the subject matter. The second prompt provided information about the topic for which the assistant should find suitable activities. The third prompt asked the assistant to find activities for each of the five activity types of the LDS activity framework.

Here, we show the final prompt, which started with 'Can you please create one activity idea for the following five activity types:' followed by a list of all four activity types, such as 'Communicative activity: Students discuss with others by communicating, debating, discussing, arguing, sharing, reporting, collaborating, presenting, and questioning'.

The AI responded with ideas for all five activity types. For the communicative activity type, it suggested an Ethics Fishbowl Discussion with the following activity: 'Organise a 'fishbowl' discussion where a small group of students sits in an inner circle and discusses a topic related to ethics and Bias in data while the rest of the class observes. After a set time, new students rotate into the discussion. Topics can include the ethical use of data in different industries, the balance between innovation and privacy, and how to mitigate Bias in machine learning models'. Another generated activity was a practice activity about a Data Ethics Role-Play Simulation.

The expert assessment was that the AI could generate a list of plausible ideas for each activity type, yet these can be generic. For example, the example above does not provide details of how group discussion outcomes should be recorded and merged across teams. The course team would need to evaluate the usefulness of the ideas for their course and produce additional guidelines not included in the response. Also, some of the generated ideas may not be applicable within the context of a distance learning institution where students communicate mostly online.

Task: inclusive curriculum

The task assigned to the AI is to adapt produced materials according to the university's guidelines on inclusive curriculum. The inclusive curriculum guidance recommends a comprehensive checklist, such as considering student and staff feedback, assessing diversity, and access. Not all tasks listed in this guidance seemed suitable for Generative AI, but there were promising ones, such as writing in Plain English or using inclusive language. Here, we show how Generative AI was used to make a text from a United Nations tutorial on inclusive language[3] more inclusive. We chose the text from the tutorial as it allows us to compare the output of the Generative AI with the expert annotations.

The prompt began with instructions that stated that the example text should be rewritten. It was followed by the text to rewrite: 'Rewrite this text using inclusive language and explain it: Anyone who sees the terrible images of Country Y can only be touched in his heart. The crisis is manmade. (…). She met with the most relevant parties, (…)' (note that the prompt was abbreviated). The completion of the prompt returned this rewritten text: 'Anyone who sees the devastating images from Country Y can only be moved at heart. The crisis is a result of human actions. (…). They engaged with key stakeholders (…)'. Furthermore, the agent explained that 'his heart' and 'man-made are gendered expressions and can be replaced with expressions that do not assume the gender of a person.

The experts noted that the AI made all cases more gender-inclusive and that the explanations were useful as they aided in understanding which cases had been made more gender-inclusive.

Discussion and conclusions

Our paper explored the opportunities of this emerging technology to produce course materials based on important elements of a course production framework and reported the assessment of teaching experts on three common tasks.

From this first trial of the Microsoft Azure AI Chat technology, we found that the AI can produce content that could support towards completing the outlined set of tasks across all activities. The AI can aid with brainstorming ideas towards solving the task, creating an outline, and assessing adherence to specific writing guidance (such as inclusive language). In all cases, the experts noted that the content would need adjustments and checking, ranging from small sections to rewriting generated ideas to make them more suitable for an online and distance learning university context.

The breadth of the AI responses suggests that this technology can work for a broad range of purposes as a generalist. However, we still need to examine to which degree it can adapt to context-specific problems and work as a specialist incorporating subject-specific, pedagogical, and university-specific knowledge. Initial insights in this paper suggest that the AI can consider more specialised information for its responses with the right prompting.

Our evaluation of these findings is that, as the technology stands, it can best be considered an assistive technology that can support various roles but that, at best, it can have an advising capacity, as any output needs to be evaluated for integrity, veracity, and accuracy. Therefore, it looks likely that our curriculum development team model will continue but with the addition of a writing agent.

This leads our work into the realms of social science and how cognitive machines can work productively with one another. More importantly, how can cognitive agents assist learning? One previous example is Open Essayist

[3] https://www.un.org/en/gender-inclusive-language/assets/pdf/EN-Toolbox-Apply-the-guidelines-to-a-text_(self-paced).pdf

(Edwards et al., 2022; Whitelock et al., 2017), where Pask's (1975) Conversation theory was used to summarise a student's draft essay. This meant the student could check from the summary whether this was what they intended to write for this essay. The theory behind this action is that if a person understands what is being explained to them by a teacher, they can repeat it back to that teacher. In this case, the AI agent repeats what has been written as a summary. Now, with the introduction of Generative AI agents, does it mean we need to define learning in these new contexts more precisely? Are new theories required that are open to test? New horizons are open to Education, but we should not neglect the ethical concerns surrounding copyright, manipulation, and data privacy in our future agendas.

References

Adeshola, I., & Adepoju, A. P. (2023). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments*, 1–14. https://doi.org/10.1080/10494820.2023.2253858

Alkaissi, H., & McFarlane, S. I. (2023). Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus*. https://doi.org/10.7759/cureus.35179

Aneesha. (2023). *Eduweaver* [Jupyter Notebook]. https://github.com/aneesha/eduweaver

Bektik, D., Ullmann, T. D., Edwards, C., Herodotou, C., & Whitelock, D. (accepted). *AI-Powered Curricula: Unpacking the Potential and Progress of Generative Technologies in Education*. Annual Conference on European Distance and E-Learning Network, EDEN 2024, Graz, Austria.

Biggs, J. (2014). Constructive alignment in university teaching. *HERDSA Review of Higher Education*, *1*, 5–22.

Chen, Y., Jensen, S., Albert, L., J., Gupta, S., & Lee, T. (2023). Artifcial intelligence (AI) student assistants in the classroom: Designing chatbots to support student success. *Information Systems Frontiers*, 161–182.

Edwards, C., Whitelock, D., Cross, S., & Wild, F. (2022, June 22). *Towards autonomous, confident academic writing: Supporting students in developing their writing through a stand alone online tool*. AHE2022, Manchester, UK. https://ahenetwork.org/ahe-conference-2021/

Liu, M., Ren, Y., Nyagoga, L. M., Stonier, F., Wu, Z., & Yu, L. (2023). Future of education in the era of generative artificial intelligence: Consensus among Chinese scholars on applications of ChatGPT in schools. *Future in Educational Research*, *1*(1), 72–101. https://doi.org/10.1002/fer3.10

Pask, G. (1975). *Conversation, Cognition and Learning: Cybernetic Theory and Methodology*. Elsevier Science Ltd.

Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules. *Computers in Human Behavior*, *60*, 333–341. https://doi.org/10.1016/j.chb.2016.02.074

Sharples, M. (2023). Towards social generative AI for education: Theory, practices and ethics. *Learning: Research and Practice*, *9*(2), 159–167. https://doi.org/10.1080/23735082.2023.2261131

Tahiru, F. (2021). AI in Education: A Systematic Literature Review. *Journal of Cases on Information Technology*, *23*(1), 1–20. https://doi.org/10.4018/JCIT.2021010101

Whitelock, D., Twiner, A., Richardson, J. T. E., Field, D., & Pulman, S. (2017). What Types of Essay Feedback Influence Implementation: Structure Alone or Structure and Content? In D. Joosten-ten Brinke & M. Laanpere (Eds.), *Technology Enhanced Assessment* (pp. 181–196). Springer International Publishing. https://doi.org/10.1007/978-3-319-57744-9_16

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J.-B., Yuan, J., & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, *2021*, 1–18. https://doi.org/10.1155/2021/8812542

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, *2*, 100025. https://doi.org/10.1016/j.caeai.2021.100025