sciendo

European Journal of Open, Distance and E-Learning • 26(S1) • 2024 • 20-34
DOI: 10.2478/eurodl-2024-0013

European Journal of Open, Distance and E-Learning

Towards Generative AI for Course Content Production: Expert Reflections

Research Article

Thomas Daniel Ullmann*, Chris Edwards, Duygu Bektik, Christothea Herodotou, Denise Whitelock

Institute of Educational Technology, The Open University, UK

Abstract: The wide availability of generative artificial intelligence (AI) for content production has resulted in a growing interest in the area of education, particularly for course content production purposes. This research has mapped out a set of curriculum production tasks and illustrated how generative AI can support three important tasks: the development of course outlines and content, the drafting of assessment instructions and the mapping of learning outcomes to benchmark statements. We evaluated the outputs of the generative AI with five experts: a course production expert, an academic expert, two Learning Design experts and an AI expert. The results indicate that generative AI enabled the generation of plausible content skeletons for content and first drafts of relevant content to aid the course production team's brainstorming but also highlighted the importance of reviewing generated content. Our research indicates that generative AI can result in shifts in the delivery of these tasks.

Keywords: *Course production; Course content; Generative AI; ChatGPT; Natural language generation; Educational technology*

Introduction

Generative artificial intelligence (AI) has received great interest in educational research and the educational industry for using it in course content production purposes. This paper provides an overview of curriculum production tasks and presents the findings of an evaluation of the capability of generative AI on three core course content production tasks, namely the generation of course outlines and course content, the drafting of assessment instructions and the automated mapping of learning outcomes to benchmark statements.

A common discussion in learning forums is how AI can potentially impact the future of education at all levels, including K12 and higher education. References are made to the provision of personalised and bespoke learning, assessment and automated grading, student profiling and predictions of performance and intelligent tutoring systems (Zawacki-Richter et al., 2019). The use of AI is often seen as complementing and enhancing current teaching practices, especially when used in collaboration with teachers (Jianzheng & Xuwei, 2023). A recent systematic review of 40 articles noted a range of AI applications already tested in K12 and higher education, including expert systems for pedagogical planning, intelligent tutors providing customised material, guidance and feedback, machine learning algorithms predicting student attitudes, personalised learning systems facilitating student interactions and improving online learning, visualisations combined with virtual reality and a chatbot used as a student partner, with the student interest yet dropping after 1 week of using it (Zhang & Aslan, 2021). A similar review of 63 empirical studies in education clustered AI-related activities into three overarching areas: development of AI systems such as classifications, recommendation and deep learning; application of AI in three areas: immediate feedback generated from errors, fostering learners' reasoning abilities and adaptive learning systems customising themselves to students' needs and learning objectives; and integration through

* *E-mail: thomas.ullmann@open.ac.uk*

biofeedback, role-playing, immersive learning and gamification (Zhai et al., 2021).

This paper focuses on the application of generative AI in the design of online courses for distance-learning university students. The production of educational resources has been seen as one of the most common applications of AI in education. UK statistics showed that 42% of teachers used generative AI to help with school work, 62% used it to create curriculum resources and 42% for planning curriculum (The Open Innovation Team & Department for Education, 2024). It could be argued that AI can be used to support the production of online courses in terms of, for example, producing content and learning activities, creating personalised learning paths based on student performance and providing students with 24/7 support by answering questions related to content. AI could act as a co-designer in the production process by designing or updating course content, such as rubrics for assessment or focusing on achieving specific goals, such as making the curriculum more inclusive (UNESCO, 2023). Preliminary findings using ChatGPT in the production of scientific writing show that generative AI produces credible pieces of writing, yet with a mixture of true and fake data (Alkaissi & McFarlane, 2023). The study of Ullmann et al. (2024) stressed the importance of human content experts reviewing and validating any outputs of AI, mitigating risks related to inaccurate content.

In the next section, we review studies that used AI in the production of educational resources and material in an effort to assess the current state of the art in this area and define future directions.

AI in the production of educational resources

AI's influence is growing globally, with policy strategies often overlooking its educational implications. Various studies highlight AI's integration in schools and its potential to enhance learning through analytics and intelligent tools (Baker & Inventado, 2014; Chen et al., 2023; Fischer et al., 2020; Niemi, 2021). One of the promising applications of generative AI in education is its potential to assist educators in generating course materials. By synthesising and rephrasing existing content, generative AI can save time and effort for educators, allowing them to focus on more complex aspects of course design and pedagogy (Alshater, 2022; Terwiesch, 2023). Thus, governments invest in developing AI tools to benefit from this technology in content production. For example, the UK government invested up to £2 million in AI tools for Oak National Academy.[1] This online learning platform offers a suite of practical AI tools designed for and freely available to teachers, which will help develop new, free resources powered by AI for teachers, such as lesson planners and classroom quizzes.

There is increasing support for utilising generative AI in content production, ranging from creating instructional content to providing automated assessment feedback and supporting academic services (Pelletier et al., 2022). As stated in Bektik et al.'s (2024) literature review, generative AI can aid the content production process by creating prompts for *open-ended questions* that align with the learning goals; generating quality *rubrics* that clearly and concisely explain exactly what students need to accomplish; creating prompts for *formative assessment activities* that provide ongoing feedback to inform teaching and learning; helping in creating *lesson plans* for specific courses, developing customised resources and learning activities; carrying out *assessment and evaluation* and supporting the writing process of research; suggesting ideas for *graphical content*; creating ways to make content more engaging with *interactive exercises and simulations*; creating *analogies* to turn abstract concepts into concrete ones; generating *reflective questions* based on course readings or video scripts; generating student *learning tasks and feedback* based on existing course content; generating *engaging video scripts* based on existing written content, including suggesting on-screen graphics; and generating *scenarios or case studies* based on existing content (Herft, 2023; Liu et al., 2023; McInnes, 2023; Michel-Villarreal et al., 2023; Rahman & Watanobe, 2023). We have reviewed practical examples currently employed for such tasks and further examined examples of existing practices, which are discussed next.

Various companies and platforms have developed AI-powered tools since ChatGPT was launched in late 2022 to aid in curriculum generation, enhancing the course creation process for educators. We have reviewed these tools and their functionalities, including examining each tool's features, capabilities and specific contributions to educational content creation and course design. Table 1 summarises some notable developments in this field. The table encapsulates the functionalities and distinctive aspects of each AI tool or platform, illustrating the diverse ways AI is being integrated into

[1] https://labs.thenational.academy/

Table 1. Generative AI applications for content generation

Tool/platform	Description	Key features
Blackboard's AI Design Assistant[1]	Simplifies the creation of courses by assisting in structure construction, content creation and aesthetic design.	Generates learning modules, rubrics, question banks and assessments. Customisable content and aesthetics. Image insertion from Unsplash. Guidance on course structure.
Moodle Plugins[2]	Enhances Moodle with AI-powered features like ChatGPT, DALL-E, Stable Diffusion integrations, AI chat support, AI-generated images and AI-generated questions.	AI Connector for API services. AI chat block[3] for 24/7 support. AI Text to Image for Moodle file picker. AI Text to questions generator.
Khan Academy's Khanmigo[4]	An AI-powered guide aiding educators in lesson planning and providing student feedback. Acts as a writing coach and offers real-time progress insights.	Lesson planning assistance. Real-time student feedback. Collaboration and customisation tools for educators.
TeacherMatic[5]	An AI tool suite offering various educational tasks. Developed with input from educators, it streamlines material creation and course improvement.	Generates lesson plans, quizzes, rubrics and more. Course Improvements. Generator for personalised teaching enhancements.
Course AI[6]	A platform designed to simplify the online course creation process, making it accessible to a wide audience, including non-educators.	Intuitive interface with drag-and-drop functionality. Generates course outlines, quizzes, video scripts and content.
EduWeaver[7]	An open-source tool using AI to create course outlines and learning activities, enhance text content generation and integrate interactive elements.	Generates course outlines, sections, examples and quizzes. Utilises ChatGPT for instructional design advice.
Coursera[8]	Prototype AI engine assisting educators in content creation and learning pathway design, integrating Coursera content and user-uploaded materials.	Generates course structure and content. Incorporates public Coursera materials. Easy Author for course review and editing.

[1]AI Design Assistant (blackboard.com)
[2]Moodle plugins directory: AI Connector | Moodle.org
[3]Moodle plugins directory: OpenAI Chat Block | Moodle.org
[4]Khanmigo Education AI Guide | Khan Academy
[5]https://teachermatic.com/generators/
[6]AI Course Creator | Best Online Course Builder - CourseAI
[7]https://github.com/aneesha/eduweaver
[8]Product Demo: AI-Assisted Course Building - Coursera Conference 2023 - YouTube

educational technology to support teaching and learning processes.

These developments suggest that course designers could potentially create highly engaging course content simply by providing a generative AI engine with a few sophisticated text prompts that are aligned with student learning outcomes. For many course developers and subject matter experts, course creation is among the most time-consuming tasks they undertake. Overall, AI-based content-generation tools can speed up the course development process. However, since these AI-powered tools listed above are not open source (apart from EduWeaver), their operational mechanisms remain unclear, as does whether this speed-up process sacrifices the content quality. Consequently, there is a need to evaluate the capabilities of generative AI in producing course content, which is the focus of this paper.

Curriculum development tasks

To obtain a list of tasks associated with curriculum development, we reviewed available internal University documentation for the development of the curriculum. This included details of the business planning process, the decision-making process for curriculum development, academic governance, advice and guidance on various elements, including assessment, learning outcomes and review.

We found that curriculum development first appears within the business planning process. However, this is perhaps better viewed as a management rather than a curriculum development activity. Therefore, we have included a range of core activities that follow from the moment someone first begins to describe the actual aims and content of a course. Many will be familiar with writing educational aims and learning outcomes aligned to benchmarked statements and other standards. Most will also have considered course outlines and assessment strategies. Some people will be familiar with the use of personas to help focus on the needs of the intended learners. In some institutions where materials are intended for large numbers of students or for distance learning, there may be a house style and the need for different levels

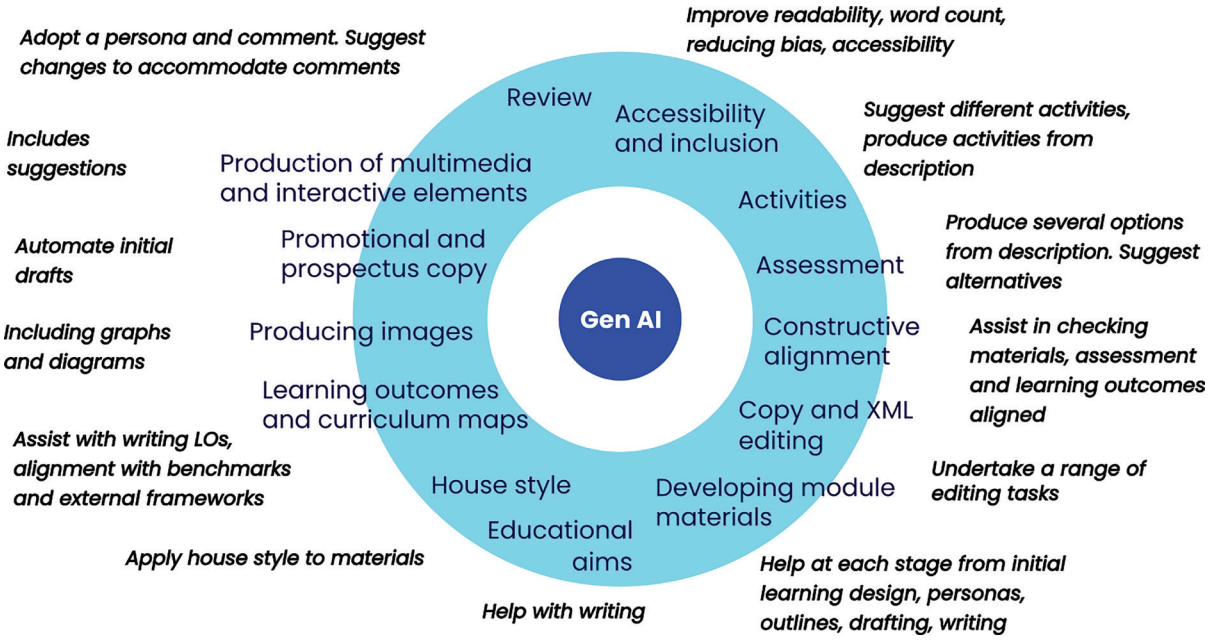

Figure 1. The potential for generative AI to support common curriculum tasks. (The inner circle is an alphabetical list of common curriculum tasks and the outer ring contains an example of how each might be supported by generative AI).

of editing. This editing might also include checking and standardising XML tags to ensure they are ready for a virtual learning environment (VLE). All these and more are set out in Figure 1. The details in the figure are not intended to be exhaustive but to indicate the range of activities that we might expect to see in curriculum development.

Figure 1 shows several of the main tasks associated with developing a curriculum and how generative AI might contribute to completing them. The list of tasks in the inner ring is a collation of items taken from our institutional references relating to curriculum development. The outer, italicised ring gives an example of how we might expect generative AI to support the completion of the task for each item. The items are arranged in alphabetical order. The cycle represents the iterative nature of the process of developing a curriculum.

Quality of curriculum tasks

Each of the tasks outlined above will result in tangible outcomes, and course authors, students and other stakeholders will have expectations regarding the quality associated with the produced materials. These expectations are often only partly made explicit in training materials and guidance for course authors but more often are part of the implicit knowledge of the course production team, who, over the years, have learned how to write high-quality course materials. The importance of this implicit knowledge is further amplified by research that indicates that, as of now, it is not clear how course quality is best conceptualised (Schiering et al., 2023). In this research, we rely on the expert knowledge of staff to judge the quality of the generated content. We included multiple experts to cover a wide range of perspectives on course quality. A further reason to work with experts is that there are currently no standards or benchmarks that allow us to judge the quality of the content for the specific tasks outlined above. The research on the various dimensions that potentially influence course quality (for example, see Lowenthal & Hodges, 2015; MacDonald & Thompson, 2005) looks at the quality of a course as a whole but not at quality dimensions for the individual tasks of course production, which was required for this research.

Research aims

In summary, the literature review showed that educational researchers and education companies see the potential of using generative AI to support curriculum production tasks and our curriculum development tasks mapping showed that many tasks could potentially benefit from generative AI. The literature review also revealed that the research around generative AI and course content production

is currently in an exploratory phase and more research is needed to understand the quality of the rapidly generated course content. In this context, this qualitative research aims to assess whether generative AI can support teachers in creating course content. This research is novel, timely and of high importance as the introduction of generative AI in the curriculum production process could significantly impact the education sector. Our research question is:

RQ: What is the quality of generative AI-created course content, according to experts?

In this paper, we report on our recent experience in using GPT-4 to assist with activities relating to a selection of three of these tasks. The three tasks are:

1. Produce outline and content of a course section
2. Produce instructions for an assessment
3. Adapting and mapping learning outcomes to benchmark statements

We chose these three tasks because they highlight important characteristics of generative AI in the context of curriculum production. We will use the findings, in response to the research question, to indicate any shifts in the delivery of course content production that the introduction of generative AI to the process may have.

Methods

We set up a ChatGPT instance using Microsoft's Azure OpenAI. For replication purposes of this research, we used the following parameters: ChatGPT 4.0 (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 did not fine-tune the model or use retrieval augmentation generation (RAG) on university content, meaning the generated content is from ChatGPT only.

For each of the tasks, we experimented with several prompts. A prompt is the user input which is sent to ChatGPT. The generative AI uses the prompt to return a completion/response to the user. As highlighted in Ullmann et al. (2024), this stage required trialling of different prompts. For this, we were considering available guidance, such as creating detailed prompts, adding examples, splitting larger tasks into smaller prompts, providing a recipe-like structure, etc. We also incorporated ideas from domain-specific prompts in the 'Prompts for Education' GitHub repository. The prompt developing process stopped when the prompt author deemed the completion as a plausible prompt response for the task. The content that was used for the expert evaluation, a new ChatGPT session was initiated and the completion of the final prompt was captured in a document.

For evaluating the quality of the prompt responses, we used a qualitative research approach, namely a synchronous online focus group of experts organised by a facilitator (Underhill & Olmsted, 2003). The facilitator interrogated five experts: a course production expert, an academic expert, two Learning Design experts and an AI expert. The experts had >50 years of combined expertise in curriculum production and the facilitator had >15 years of interdisciplinary expertise in education and computer science.

A week before the focus group session, the experts had the opportunity to familiarise themselves and comment individually on a document containing the final prompt responses for each task. On the day, the experts met for a one-hour-long open discussion guided by a facilitator. Within a structured discussion, the experts commented on the quality of the responses of the generative AI, considering the task specifics. In order to avoid leading the experts towards specific aspects of quality, the facilitator started the session with an open question: 'With your expertise when reading the text, please report any observations'. Additionally, the facilitator prepared several key questions to follow up in case the experts needed more direction: Is the text relevant? Does the text respond to the aims of the task? Did you spot any mistakes? Is the language appropriate? The document was available during the focus group and served as a stimulus during the discussion.

The facilitator captured all the experts' responses in summary notes, one for each task. The facilitator shared all the notes with the experts to confirm that they represented the discussion and all experts agreed with the content of the summary notes. The results section reports the main observations that the experts shared.

We chose a common scenario for each of the three tasks: working towards creating a course about 'Data science'. While we chose this specific area, we believe the prompts are general enough to be adapted to other

course topics by replacing 'Data Science' specific cues with another course topic.

Results

We report the prompts and excerpts of the prompt completions for each of the three tasks: generating course outlines and sections, generating assessment instructions and mapping learning outcomes to benchmark statements. For each task, we outline the experts' main observations.

Produce outline and content of a course section

One of the core tasks of a course production team is preparing various course outlines. These outline the structure and content of the course. The actual writing of course content often follows this initial step, drafting content and refining it iteratively for each of the outline's points.

We tasked the AI with producing an outline of content suggestions based on a subject-specific big idea and draft content for one of the suggestions. The course team provides the central theme and the AI generates a course outline based on that idea. Ullmann et al. (2024) presented a prompt that suggested big ideas for a specific course topic. These big ideas are meant to hook students and create interest in the topic. The prompt presented there can inform this prompt by using one of the generated big ideas.

We developed the following prompting sequence that consisted of two prompts to create a course outline and a follow-up prompt, which created course content based on a pre-defined structure for one of the outline's points. The first prompt set up the interaction and asked the user to specify a big idea for the course content:

> 'You are assisting an apprenticeship teacher in creating course content for a single week of an online apprenticeship course.
>
> Start by asking the teacher about the big question for this week's course.
>
> Once the big question is described in detail, prepare an outline for the week of the course using headings and provide a short description of the course content under each header.
>
> After presenting the course content, ask the teacher for feedback, whether to expand it, whether you should generate another idea, or whether the teacher is satisfied'.

The AI composed a completion consisting of three major sections. The first section asked the user to provide the week's big idea or central theme. In the second section, the AI showed an exemplary course outline for 7 days of the week, starting with an introduction, several content sections and, towards the end of the week, a group activity and a reflection and question-and-answer section. In the third section, the AI drew up a section exemplifying how it would ask the user for feedback.

The follow-up prompt provided the big idea for the week: 'The big idea that my students should work on this week is Ethics and bias in data'. This idea was one of the big ideas generated by another prompt reported by Ullmann et al. (2024).

The completion followed the structure that the AI proposed in the previous completion. It restated the big idea: 'What are the ethical considerations and potential biases in data collection, analysis and usage, and how can we address them?' This followed seven sections for each day of the week. For example, it proposed for Day one the following:

> 'Day 1: Introduction to Ethics and Bias in Data
>
> - Defining data ethics and understanding its importance.
> - Types of bias that can affect data: sampling bias, measurement bias, etc.
> - Real-world implications of unethical data practices and biased analyses'.

Similarly, it proposed for the other days of the week. Day two was about ethical data collection, day three was about the identification and mitigation of bias, day four was about ethical decision-making, day five was about data governance and policy, day six proposed a workshop on ethical problem-solving and day seven was about reflecting on the week and a question-and-answer session.

The final prompt asked the AI to produce course content about one of the bullet points of the previously generated outline. The prompt also specified a special structure asking to include a surprising example, sections about explaining the topic and a section aimed at aiding the learners to reflect on the topic:

> 'Please create course content about the topic: Defining data ethics and understanding its

> importance. Start with a surprising example demonstrating the importance of the topic. Then, explain the topic in three paragraphs. Then, summarise the content in one paragraph. Finally, help the student to critically reflect on the topic by creating three reflective questions'.

The AI created a completion that consisted of three requested parts. As for the example, it proposed the following:

> 'Imagine a scenario where a seemingly harmless fitness tracking app collects data on users' physical activities, locations and health metrics. This data is then sold to insurance companies without users' explicit consent, leading to increased premiums for those identified as high-risk based on the collected data. The users are unaware of this breach of trust and the ethical implications it carries, highlighting the critical need for ethical standards in data handling'.

It then explained in three paragraphs what data ethics is about, its importance and the consequences of ignoring data ethics. This is followed by a summary and three reflective questions.

As outlined in the methods section, the prompts and completions were sent to five experts who commented on them individually. Afterwards, the experts met to discuss their comments in a structured one-hour-long meeting.

Regarding outlining generating prompts 1 and 2, the experts found that the AI's proposed outline can help the course team draft a first outline, which they then need to adapt further and contextualise considering the whole course. They also thought that the outline might help to provide another perspective on the course outline, which may help adapt an existing outline. The experts found an implicit pedagogical structure embedded in the response of the AI, which can help the learners understand the materials. The proposed sequence of contents started with an introduction to the topic. This was followed by several content-specific sections, which also build from the more concrete issues, such as the ethics of data collection, to more abstract topics, such as data governance. After the content-laden sections, a group and reflective exercise were suggested to be useful in strengthening the learning of the contents. This structure can help to guide the learner, starting from introductory concepts and moving them to more advanced concepts and strengthens the learning with activities and exercises at the end of the course block.

Regarding content-generating prompt three, the experts found that the generated text provided a first draft of texts which then needed further adaption and change. The generated example, the three content paragraphs and the summary all contain valid points to make, but as they stand, would need further development. This is tied to another observation, the created text lacks citations. This is likely a limitation of the prompt compared to that of generative AI, as these systems can produce text, including citations. Furthermore, the generated text would need to be checked for inaccuracies. For example, the AI proposed the following sentence when defining data ethics (example was shortened): 'Data ethics is a branch of ethics (…). It encompasses a wide array of ethical issues, including privacy, accuracy and consent, particularly (…)'. Here, it was not clear why accuracy was mentioned along with privacy and consent. Instead of accuracy, a better point to mention could have been bias and fairness.

The next investigated task was the production of instructions for an assessment.

Produce instructions for an assessment

Writing instructions for assessments is a core part of the course production team. Students benefit from clear instructions. Here, we envisioned the AI producing draft instructions for an assessment. The course production team specifies the topic of the assessment and reviews and edits the outputs generated by the AI.

We used two prompts. The first prompt set up the interaction with the course production team. It asked the course writing team to provide information about the course and assessment topic. It also specified that the AI should generate assessment instructions. The final prompt that we created was:

> 'You are assisting an apprenticeship teacher in creating instructions and grading criteria for an assessment in a distance learning context.
>
> Start by asking the teacher about the subject matter of the course and a detailed description of the topic of the assessment.
>
> Once the assessment is described in detail, prepare detailed instructions for the students of the assignment.

After presenting the details, ask the teacher for feedback'.

This prompt combined several steps into one. The idea was to let the AI generate a sequence of responses, which would guide the user through the steps.

The AI created a detailed completion of the prompt. It started by outlining the purpose of the interaction: 'I am looking forward to assisting you with the creation of the assessment instructions and grading criteria for your course'. It then detailed information that the user was supposed to provide: 'What is the subject matter of the course?' and 'Could you give a detailed description of the topic that the assessment will cover?'. In addition to the information that was explicitly specified in the prompt, the AI added five additional user input requests, such as 'What are the learning objectives that this assessment aims to evaluate?', 'Are there any specific guidelines or requirements that the students must follow (e.g., word count, citation style, presentation length)?' and others. The experts found that the additional information requested by the AI would have been useful for specifying the assessment guidance. Still, they also noted that it would have substantially increased the amount of time to retrieve and prepare the requested information.

The AI has also already drafted an example of what the assessment instructions could potentially look like. It starts with a greeting message to the students: 'Dear Students, As part of your [Course Name], you are required to complete an assessment that will evaluate your understanding of [Assessment Topic]'. This is followed by a template proposing six sections for the assignment instructions, including sections such as 'Assignment Type: [e.g., Research Essay/Project/ Presentation]', 'Topic Description: [Provide a detailed description of the topic and what aspects need to be covered]', submission deadline, submission method and others.

As a response to the completion, we chose the following second prompt providing a course topic, subject of the assignment and learning outcomes:

'The course topic is about Ethics and Bias in data. The subject of the assignment is Ethical considerations when collecting data. The learning outcomes are (a) the development and presentation of justification, rationale and argument; (b) your ability to communicate complex ideas in an appropriate and focussed manner; and (c) your ability to select and critically appraise appropriate resources on a given topic'.

The AI created a completion that followed the template AI had proposed earlier, consisting of an introduction, assignment type, topic description, learning objectives, requirements, submission deadline and submission method. The AI chose an essay as an assignment type. The generated introduction introduces the chosen topic with the following words:

'Dear Students,
As part of your course on Ethics and Bias in Data, you are required to complete an assessment that will demonstrate your understanding and critical analysis of ethical considerations when collecting data. This assignment will help you develop your ability to justify and rationalise ethical decisions, communicate complex ideas effectively and critically appraise resources related to data ethics'.

After this introduction, it offers a short description of the topic:

'You are to write an essay that discusses the ethical considerations involved in the collection of data. Your essay should cover key ethical principles, potential biases that may arise in data collection and strategies to mitigate these issues. You should also include real-world examples to support your analysis'.

This is followed by the three learning objectives outlined in the prompt, some requirements such as word length, a section to add information about the deadline and submission method.

Overall, the experts noted that the generated text presented a good start for drafting the assignment instructions. The assignment creators would need to think about adding more details and depth. A different prompt might potentially open up more ideas for assessments, for example, by asking it to provide multiple options for the module team to consider.

The experts commented on the additional information that the AI asked the user to add, which was

not explicitly mentioned in the prompt. They were all seen as good suggestions, such as learning objectives or guidelines. However, it was also noted that providing all the AI-requested information also adds to the user's workload.

The experts noted that the AI also created an exemplary template for the prompt response without explicitly specifying this in the prompts. The experts found that it could be useful for the course writer to see such templates to structure the assessment instructions. This can serve as the starting point for creating a template that fits the requirements of the particular assessment instructions, making use of one of the capabilities of generative AI, which is to write content according to a template structure.

Adapting and mapping learning outcomes to benchmark statements

One of the challenges for anyone creating a curriculum within higher education is to ensure alignment with external benchmarks. An aspect of this challenge is creating appropriate learning outcomes and objectives that both fit with external requirements and the internal criteria set down by the institution. Within UK higher education, one available qualification is the Level 7 Research Scientist Apprenticeship (https://www.instituteforapprenticeships.org/apprenticeship-standards/research-scientist-v1-0). This is a master's level degree. Within the specification is a set of 23 statements of intended learning objectives, often referred to as knowledge, skills, and behaviors (KSBs) because they are arranged within three categories: knowledge, skills and behaviours. If a university creates a programme to support this degree, it has to demonstrate that it delivers these KSBs. This creates an immediate problem for universities which have developed an internal model specifying a different number of learning outcomes or categories. Like many universities, ours keeps numbers relatively low, aiming for around 12 and has four categories: knowledge and understanding (KU), cognitive skills (CS), key skills (KS) practical and professional skills (PS).

Given this context, we tasked the generative AI with aligning the learning outcome categories of the university with the 23 externally specified apprenticeship learning outcomes. We envisioned that the output of this process could be reviewed by the curriculum team and used to identify the degree of alignment between internal and external learning outcomes. We developed a series of prompts towards the solution. We chose a conversational approach in which the activity was broken into steps in which each step could be checked and revised before progressing to the next. The first step was to provide the generative AI with the 23 KSB statements. Hence, the initial prompt was:

'The following is a list of KSBs, these are descriptors I would like to ask you to perform some actions on. Each KSB comprises a paragraph of text. It begins with an identifier in the form', 'Kx:', 'Sx:', or 'Bx:'. Where 'x' is an integer. The descriptor follows the colon. The list is: 'The generative AI responded with':

> 'Sure, I'm ready to assist with the actions related to your list of KSBs. Please go ahead and provide the list and let me know what specific actions or tasks you would like me to perform on each KSB'.

We then entered the KSBs in the form:

'K1: Subject specific knowledge: A deep and systemic understanding of a named/recognised scientific subject as found in an industrial setting, such as biology, chemistry or physics, found in the nuclear, food manufacture, pharmacology or energy production sectors, at a level that allows strategic and scientific decision making, while taking account of interrelationships with other relevant business areas/disciplines'.

The prompt included a list of all KSBs copied from the apprenticeship standards website. For brevity, the whole list is not shown here.

The generative AI responded:

> 'Thank you for sharing the detailed list of KSBs. These descriptors are quite comprehensive and cover various aspects essential in a professional setting, especially in scientific and industrial environments.
>
> To assist you further, could you please specify the actions or tasks you would like me to perform with respect to these KSBs? For example, are you looking for an analysis of each KSB, suggestions on how to develop these skills, or perhaps a way to integrate these into a training programme or job description? Let me know how I can assist you with these KSBs'.

We asked the generative AI:

In our institution, we use four categories of learning outcomes: KU; CS); KS; practical and PS. Please produce the most suitable statements that could be used within these four categories to cover all the concepts, skills and ideas set out in the text of the KSBs. Please use the abbreviations of the learning outcome categories, minimise the number of statements and write in the form 'At the end of your study you will…' but omitting these words.

The generative AI responded:

'Certainly! I'll create concise statements for each of the four learning outcome categories based on the concepts, skills and ideas set out in the KSBs.

KU (Knowledge and Understanding)

1. Possess a deep and systemic understanding of scientific subjects such as biology, chemistry, or physics, particularly in industrial contexts like the nuclear, food manufacture, pharmacology, or energy production sectors'.

In the remainder of the response, generative AI outlined 11 more learning outcomes, resulting in 3 within each category (KU1, KU2, KU3, CS1, CS2, etc.).

Our final prompt was:

'Produce a table with the KSBs as the first column, these 12 learning outcome statements (identified by their codes) as the header row and a 1 in each cell to identify where each KSB links to each of the learning outcomes'.

Generative AI produced the following table (we added the highlighting to rows with no matches):

The experts found that generative AI can support developing learning outcomes in the required internal format and style and in alignment with a defined set of external requirements.

They also found that the tool readily produces a tabular representation of a mapping between two lists of

KSBS	KU1	KU2	KU3	CS1	CS2	CS3	KS1	KS2	KS3	PS1	PS2	PS3
K1	1	0	0	1	0	0	0	0	0	0	0	0
K2	1	0	0	1	0	0	1	0	1	0	1	0
K3	0	1	0	0	0	0	0	1	0	0	0	0
K4	0	0	0	0	1	0	0	0	0	1	0	0
K5	0	0	0	0	0	1	0	0	0	0	0	1
K6	0	0	0	0	0	1	0	0	0	0	0	1
K7	0	0	1	0	0	0	0	0	0	0	0	0
K8	0	0	0	0	0	0	0	0	1	0	0	0
S1	1	0	0	0	1	0	0	0	0	1	0	0
S2	0	0	0	0	0	1	0	0	0	0	0	1
S3	0	0	1	0	0	0	1	0	0	0	1	0
S4	1	0	0	1	0	0	1	0	0	0	0	0
S5	1	0	0	1	0	0	1	0	0	0	1	0
S6	0	0	0	0	1	0	0	0	0	0	0	0
S7	0	0	0	0	1	0	0	0	0	1	0	0
S8	0	0	0	0	0	0	0	0	1	0	0	0
B1	0	0	0	0	0	0	1	0	0	0	0	0
B2	0	0	0	0	0	0	0	0	0	0	0	0
B3	0	1	0	0	0	0	0	1	0	0	0	0
B4	0	0	0	1	0	0	1	0	0	0	0	0
B5	0	0	0	0	0	0	0	0	0	0	0	0
B6	0	0	0	0	0	0	0	0	0	0	0	0
B7	0	0	0	0	0	0	0	0	1	0	0	0

CS, cognitive skills; KS, key skills; KSBs, knowledge, skills, and behaviors; KU, knowledge and understanding; PS, professional skills.

statements. They noted that not all the KSB items were found to match. We will consider this below.

The original list of KSBs had three categories and generative AI successfully arranged the generated learning outcomes into the four internal categories as requested. The allocations made are also appropriate, with each learning outcome fitting well into the allocated category and with no other category offering a better home. These learning outcomes are written to an appropriate level, matching that of the KSBs and with none looking out of place.

Considering the quality of the mapping, the experts found that generative AI successfully produced a table in the way requested. To compare the automatically generated mapping with an expert mapping, one of the experts created a mapping by hand. On comparing both mappings, the expert found that all the KSBs were covered by the internal learning outcomes, although the text of four of the learning outcomes could be improved with small changes. Those that could be improved include the three cells where generative AI did not identify a link. Overall, the manual mapping agreed with the large language model (LLM) on 86% of the cells for the three KSBs. A small number, four, of the links indicated by the LLM were completely incorrect.

Overall, the experts found that generative AI can provide a strong starting point for developing a set of compliant learning outcomes. Through very simple prompts, the AI generated a workable list of learning outcomes in minutes rather than the morning or afternoon it might otherwise be expected to take. This provided a good foundation text for further development. The mapping produced by the LLM was also done in a couple of minutes and is helpful as a point of comparison for one's own expert mapping. The points of disagreement provide opportunities to consider the rationale of our choices and would ultimately lead to improvements in the wording.

Discussion

The literature review showed that there is little research into the extent to which generative AI can support curriculum tasks and a need for more empirical evidence in using generative AI for course content production. This research investigated the usefulness of generative AI in supporting curriculum production. To achieve this, it chose three core tasks from a comprehensive curriculum production framework and evaluated the quality of AI-generated responses. The chosen tasks are fundamental to the production of curriculum and, therefore, represent a realistic test bed. They were the generation of courses, outlines and writing of course sections, the production of instructions for an assessment and mapping learning outcomes to benchmark statements.

In response to our central research question, 'What is the quality of generative AI-created course content according to experts?', we found that for tasks 1 and 2, the generative AI was able to generate a first plausible structure or skeleton for the content. The experts found that it was most useful in helping to generate first drafts and aid with brainstorming ideas. The points brought forward by the AI were also seen as a useful perspective from a third person, which the course content production team might find useful to compare their ideas with. The experts also noted that for the outline creation task, the generative AI produced a structure that could potentially aid learners as the content was introduced, starting with concrete and abstract concepts together with activities and exercises to strengthen learning.

The third task showed that generative AI could potentially be used to assist with the development of learning outcomes and with checking the alignment between institutional learning outcomes and external benchmark statements. Prior to the availability of LLMs, this task would have been a classical text classification case requiring large annotated datasets and considerable technical expertise to work with them. The inroads demonstrated through the presented example show that, with care, generative AI could potentially assist in the completion of this kind of classification task without that training data or expertise.

For all three cases, the experts noted that with the current version of ChatGPT, generated contents cannot be used as they are and that they would need to be verified with varying degrees of adaptation to make them suitable for the particular course and university context.

The findings suggest that the introduction of generative AI for course content production can result in shifts in the delivery of these tasks. We note that generative AI can have particular benefits in the drafting stages of course content. With its capability of creating course outlines (see also Aneesha, 2023) and instruction drafts, course production teams can rapidly experiment with content ideas. The automatically

generated content can serve the writer as a third perspective to consider when drafting content. Some of the generated contents may seem useful and worth further exploring, while other contents may not seem fit or require reworking. It is important that targeted prompts, as they have been presented in this paper, can aid the brainstorming of the course content production team and, therefore, potentially can help their productivity.

Furthermore, the introduction of generative AI can help accelerate content search tasks, such as finding fitting and surprising examples or ideas about topics (see also Jeon & Lee, 2023). Previously, these tasks had to be executed with standard search engine technology, requiring the exploration of the top search results. Generative AI can instead provide content summaries for various questions. While these responses still need verification, which might require traditional search engine technology, the automatically generated prompt responses can already contain enough clues to productively find suitable content.

The results presented in this paper also indicate that generative AI can generate plausible but not always correct information (see also Baidoo-Anu & Ansah, 2023; Kim et al., 2023). When using generative AI, this characteristic may lead to shifts in the work of experts from writing every word to ensuring that content suggestions are accurate and true, saying exactly what the subject expert wants to say, and that unwanted bias and plagiarism are both avoided. We expect that all the usual reviewing points, arranged at critical moments of the course production, will continue.

As highlighted in our literature review, the three examples presented in this paper are tasks that are core to the course content production writing process. With this research, we contribute to a nascent research field, which is prone to grow rapidly, of exploring generative AI for teaching and learning. The curriculum development task list presented in our literature review can be used to categorise similar work and help identify areas that require further research.

It is useful to highlight that the studied curriculum production tasks apply to a wide range of subject areas and contexts. All three production tasks are tasks that course production teams in all subject areas have to engage with. The prompts developed here are generic enough to be applied to other subject areas. Subject experts in their respective areas can use these prompts and start tailoring them to their specific context by adapting and extending them. In this process, the experts make their knowledge about the quality of curriculum tasks explicit in their prompts. This process of explicating curriculum production knowledge may contribute to the research about how course quality is best conceptualised (Schiering et al., 2023).

We also note that with the rapid development of generative AI and other AI technologies, we will likely see a further acceleration of the discussed shifts and new shifts depending on confidence in the correctness of the outputs.

Every method has its benefits and drawbacks. Here, we conducted a single focus group session with various experts as it can already reveal most of the themes about a topic (Hennink et al., 2019). Further focus groups can be beneficial, although the likelihood of finding new topics declines with each session. The insights obtained from this research about the quality of generated course content can contribute towards establishing standards and guidance, but considering other sources of information beyond course experts will help to advance the discussion around generative course content quality.

Conclusion

The rigorous systematic investigation and publication of studies associated with the different facets of the course production process, especially assessment, using generative AI is now a pressing concern to the Higher Education community. This is because universities such as the University of Michigan request staff before the fall (autumn) term to review their learning objectives concerning the knowledge, skills and values of generative AI that students will need to succeed in their learning objectives.[2] The integration of AI tools will also mean a reevaluation of Course Design.

Other universities are producing guidelines for AI use, as illustrated by Moorhouse et al. (2023), which cover three main areas: academic integrity, advice on assessment design and communication with students. There are suggestions for teachers about assessment design, one of which is running assessment tasks through generative AI to check the extent to which the

[2] https://genai.umich.edu/guidance/faculty/redesigning-assessments

tool can accomplish the task together with developing students' AI literacy.

More evidence is required to ascertain whether a first draft of course content should be produced using generative AI. There is an argument that this approach quickly aids a deeper reflection and it is quicker to recognise what is not required or unsuitable for a particular course employing this method. It also raises the issue that our paper illustrates about the use of prompts and future research into production of specialised prompt engines. Should we even be contemplating an AI prompt pedagogy?

This latter point moves the argument into developing appropriate methodologies for evaluating generative AI course content and assessments. How can we come to an agreement on standards of outputs, and what would this look like for Higher Education Quality Assurance Agencies? Recommendations for a future research agenda include appraising new learning outcomes for both students and teachers, such as generative AI literacy and relating generative AI skills for future student employment (Chiu, 2024).

We have been using AI for some time to provide adaptative learning, but we need to explore and test whether productive and accurate personalised feedback is possible with generative AI and what sort of corpora will need to be generated to obtain this goal when universities could well want to keep their course content as a critical business asset. Another area to be explored is the threat of reputational risk to Higher Education (HE)if a new *modus operandi* is not adequately evaluated. Variables such as scale, ethics and student agreement, together with their use in administrative tasks like preparing data for examination boards, require attention. More importantly, in moving in this direction, we should always keep in mind whether we are supporting student learning and assisting with developing skills required for future employment.

Acknowledgements

During the preparation of this work, the authors used Microsoft Azure OpenAI ChatGPT to create the prompts for course content production. After using this tool, the authors cited the tool outputs, reviewed and edited the content as needed and took full responsibility for the publication's content.

References

Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus, 15*(2), e35179. https://doi.org/10.7759/cureus.35179

Alshater, M. (2022). Exploring the role of artificial intelligence in enhancing academic performance: A case study of ChatGPT (SSRN Scholarly Paper 4312358). https://doi.org/10.2139/ssrn.4312358

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

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI, 7*(1), 52–62. https://doi.org/10.61969/jai.1337500

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics* (pp. 61–75). Springer.

Bektik, D., Ullmann, T. D., Edwards, C., Herodotou, C., & Whitelock, D. (2024). *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. https://oro.open.ac.uk/98012/

Chen, Y., Jensen, S., Albert, L. J., Gupta, S., & Lee, T. (2023). Artificial intelligence (AI) student assistants in the classroom: Designing chatbots to support student success. *Information Systems Frontiers, 25*(1), 161–182. https://doi.org/10.1007/s10796-022-10291-4

Chiu, T. K. F. (2024). Future research recommendations for transforming higher education with generative AI. *Computers and Education: Artificial Intelligence,* 100197. https://doi.org/10.1016/j.caeai.2023.100197

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education, 44*(1), 130–160. https://doi.org/10.3102/0091732X20903304

Hennink, M. M., Kaiser, B. N., & Weber, M. B. (2019). What influences saturation? Estimating sample sizes in focus group research. *Qualitative Health Research, 29*(10), 1483–1496. https://doi.org/10.1177/1049732318821692

Herft, E. (2023). *A teacher's prompt guide to ChatGPT aligned with 'What Works Best' guide*. https://usergeneratededucation.files.wordpress.com/2023/01/a-teachers-prompt-guide-to-chatgpt-aligned-with-what-works-best.pdf

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies, 28*(12), 15873–15892. https://doi.org/10.1007/s10639-023-11834-1

Jianzheng, S., & Xuwei, Z. (2023). Integration of AI with higher education innovation: Reforming future educational directions. *International Journal of Science and Research (IJSR), 12*(10), 1727–1731. https://doi.org/10.21275/SR231023183401

Kim, J. K., Chua, M., Rickard, M., & Lorenzo, A. (2023). ChatGPT and large language model (LLM) chatbots: The current state of acceptability and a proposal for guidelines on utilization in academic medicine. *Journal of Pediatric Urology, 19*(5), 598–604. https://doi.org/10.1016/j.jpurol.2023.05.018

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

Lowenthal, P., & Hodges, C. (2015). In search of quality: Using quality matters to analyze the quality of massive, open, online courses (MOOCs). *The International Review of Research in Open and Distributed Learning, 16*(5), 83–101. https://doi.org/10.19173/irrodl.v16i5.2348

MacDonald, C. J., & Thompson, T. L. (2005). Structure, content, delivery, service, and outcomes: Quality e-Learning in higher education. *The International Review of Research in Open and Distributed Learning, 6*(2). https://doi.org/10.19173/irrodl.v6i2.237

McInnes, R. (2023). *Embracing generative AI as a co-design partner for designing and developing courses*. https://www.timeshighereducation.com/campus/embracing-generative-ai-codesign-partner-designing-and-developing-courses

Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Education Sciences, 13*(9), Article 9. https://doi.org/10.3390/educsci13090856

Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open,* 100151. https://doi.org/10.1016/j.caeo.2023.100151

Niemi, H. (2021). AI in learning: Preparing grounds for future learning. *Journal of Pacific Rim Psychology*, 1–12. https://doi.org/10.1177/18344909211038105

Pelletier, K., McCormack, M., Reeves, J., Robert, J., & Arbino, N. (2022). *2022 EDUCAUSE horizon report, teaching and learning edition*. EDUCAUSE. https://library.educause.edu/resources/2022/4/2022-educause-horizon-report-teaching-and-learning-edition.

Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences, 13*(9), Article 9. https://doi.org/10.3390/app13095783

Schiering, D., Sorge, S., Tröbst, S., & Neumann, K. (2023). Course quality in higher education teacher training: What matters for pre-service physics teachers' content knowledge development? *Studies in Educational Evaluation,* 101275. https://doi.org/10.1016/j.stueduc.2023.101275

Terwiesch, C. (2023). *Would chat GPT get a Wharton MBA?* [White Paper]. The Wharton School, The University of Pennsylvania. https://mackinstitute.wharton.upenn.edu/2023/would-chat-gpt3-get-a-wharton-mba-new-white-paper-by-christian-terwiesch/

The Open Innovation Team & Department for Education. (2024). *Generative AI in education: Educator and expert views*. https://www.gov.uk/government/publications/generative-ai-in-education-educator-and-expert-views

Ullmann, T. D., Bektik, D., Edwards, C., Herodotou, C., & Whitelock, D. (2024, August 28). Teaching with generative AI: Moving forward with content creation. *Learning in the age of AI: Towards imaginative futures*. Annual conference on European Distance and E-Learning Network, EDEN 2024, Graz, Austria. https://doi.org/10.5334/uproc.157

Underhill, C., & Olmsted, M. G. (2003). An experimental comparison of computer-mediated and face-to-face focus groups. *Social Science Computer Review, 21*(4), 506–512. https://doi.org/10.1177/0894439303256541

UNESCO. (2023). Harnessing the era of artificial intelligence in higher education: A primer for higher education stakeholders. https://unesdoc.unesco.org/ark:/48223/pf0000386670

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

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), 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,* 100025. https://doi.org/10.1016/j.caeai.2021.100025