

Bond University

Legal Education Review

Volume 35 Issue 1

2025

ChatGPT & Other Generative AI Tools as University Teaching Aids

Stuart Hargreaves

The Chinese University of Hong Kong

Follow this and additional works at: <https://ler.scholasticahq.com/>



© Copyright the authors.

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivative 4.0 International Licence](https://creativecommons.org/licenses/by-nc-nd/4.0/).

CHATGPT & OTHER GENERATIVE AI TOOLS AS UNIVERSITY TEACHING AIDS

STUART HARGREAVES *

I INTRODUCTION

Tools that allowed end-users to experiment with generative artificial intelligence (‘generative AI’) exploded into public consciousness in late 2022 and early 2023.¹ Generative AI refers to ‘computational techniques that are capable of generating seemingly new, meaningful content such as text, images, or audio from training data’.² McKinsey has estimated the impact of generative AI on global productivity to be in the trillions of USD.³ In part, this is because, unlike past technological advances that tended to displace forms of physical labour, the advance of generative AI is likely to dramatically reshape knowledge work as well.⁴ Higher education – both a form of knowledge work itself and the *sine qua non* of training future generations in it – will surely be impacted too. Initial fears about the impact of generative AI on higher education tended to focus on issues of academic integrity because of the seeming ease with which it could generate coherent and generally appropriate English text in response to user prompts.⁵ A

* Faculty of Law, The Chinese University of Hong Kong. I am grateful to all my students for their enthusiastic embrace of this project. Elements of this project have been presented and discussed at the following conferences: the “Asian Conference on Education” in Tokyo (Nov 2023), the “Teaching & Learning Expo” at CUHK (Dec 2023), the “Generative AI Insights Conference” at CUHK (Jan 2024), the “Artificial Intelligence in Higher Education” workshop at the University of Macau (Mar 2024), and the “Annual Conference on Assessment in Higher Education” at the University of Cumbria (Jun 2024). I thank all participants for their valuable comments and discussion.

¹ See, eg, Unigram Labs, ‘Generative AI Ecosystem Part I: End User Applications’ *Medium* (online, 20 June 2023) <https://medium.com/@unigram_labs/the-generative-ai-ecosystem-9a82f6fe7eae>.

² This refers to an artificial intelligence model able to generate new and appropriate content of varying forms in response to simple language-based inputs from a user: see, eg, Stefan Feuerriegel, Jochen Hartmann, Christian Janiesch and Patrick Zschech, ‘Generative AI’ (2023) 66 *Business & Information Systems Engineering* 111.

³ See, eg, ‘The Economic Potential of Generative AI: The Next Productivity Frontier’ *McKinsey Digital* (online, 14 June 2023) <<https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>>.

⁴ Carl Benedikt Frey and Michael Osborne, ‘Generative AI and the Future of Work: A Reappraisal’ (2023) 30(1) *Brown Journal of World Affairs* 1.

⁵ For popular media reports, see, eg, Bobby Hristova, ‘Some Students Are Using ChatGPT to Cheat – Here’s How Schools Are Trying to Stop It’ *CBC News* (online, 2 February 2023) <<https://www.cbc.ca/news/canada/hamilton/chatgpt-school->

number of scholars have therefore begun to consider the way in which all kinds of assessments might have to be redesigned in a world of universal access to generative AI tools.⁶ While I have addressed some of these issues in an earlier work,⁷ I do not intend to add further to that body of literature here. Instead, this paper considers the potential use of AI tools from the “other side of the lectern” – in other words, rather than worrying about inappropriate use by students, it explores possible ways in which academics can incorporate AI into their own pedagogy. In a sense, it is a form of autoethnography⁸ – in this paper I simply describe my experimentation with a variety of AI tools in my teaching. I do not intend my conclusions or suggestions to be prescriptive; rather hopefully they will inspire others to embark on their own forms of experimentation.

I argue that, as with other institutions, universities (including individual academics) remain in the critical *experimentation* stage of generative AI. With that in mind, what this paper represents is not a complete model but rather a single data point, describing experimentation with different generative AI tools in the delivery of a large lecture course in a law school. In Part 1, I explain the genesis of the project and the background to the choice of AI tools used. In Part 2, I explain three different ways in which I used AI tools for pedagogical ends – for lecture summarization, for quiz generation, and for a self-revision chatbot. In Part 3, I discuss the results of a student survey conducted at the end of the course regarding those tools. In Part 4, I suggest that, although those surveys indicate the AI-powered chatbot was the least successful of the three tools with which I experimented, it

cheating-1.6734580>; Caitlin Casidy, ‘Queensland Public Schools to Join NSW in Banning Students from ChatGPT’ *The Guardian*, (online, 22 January 2023) <<https://www.theguardian.com/australia-news/2023/jan/23/queensland-public-schools-to-join-nsw-in-banning-students-from-chatgpt>>; Maya Yang, ‘New York City Schools Ban AI Chatbot That Writes Essays and Answers Prompts’ *The Guardian* (online, 6 January 2023) <<https://www.theguardian.com/us-news/2023/jan/06/new-york-city-schools-ban-ai-chatbot-chatgpt>>; Stephen Marche, ‘The College Essay Is Dead’ *The Atlantic* (online, 7 Dec 2022) <<https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-writing-college-student-essays/672371/>>.

⁶ See, eg, David Ernesto Salinas-Navarro, Eliseo Vilalta-Perdomo, Rosario Michel-Villarreal and Luis Montesinos, ‘Designing Experiential Learning Activities with Generative Artificial Intelligence Tools for Authentic Assessment’ (2024) 21(4) *Interactive Technology and Smart Education* 708; Rajan Kadel, Bhupesh Kumar Mishra, Samar Shailendra, Samia Abid, Maneeha Rani and Shiva Prasad Mahato, ‘Crafting Tomorrow’s Evaluations: Assessment Design Strategies in the Era of Generative AI’ (2024) <<https://arxiv.org/abs/2405.01805>>; Jin Mao, Baiyun Chen and Juhong Christie Liu, ‘Generative Artificial Intelligence in Education and Its Implications for Assessment’ (2024) 68(1) *TechTrends* 58; Peter Matheis and Jubin Jacob John, ‘Reframing Assessments: Designing Authentic Assessments in the Age of Generative AI’ in Saadia Mahmud (ed) *Academic Integrity in the Age of Artificial Intelligence* (IGI Global, 2024).

⁷ Stuart Hargreaves, “‘Words Are Flowing out Like Endless Rain into a Paper Cup’: ChatGPT & Law School Assessments”, (2023) 33 *Legal Education Review* 69.

⁸ See, eg, Tony E Adams et al, ‘Autoethnography’ in Jorg Mathes et al (eds) *The International Encyclopedia of Communication Research Methods*, (Wiley-Blackwell, 2017); David Butz and Kathryn Besio, ‘Autoethnography’ (2009) 3 *Geography Compass* 1660.

is nonetheless the one that best offers a glimpse into a tantalizing future of AI-assisted pedagogy.

II PROJECT GENESIS

I developed a keen interest in the possible use of AI in higher education in early 2023, but I felt that my own University was behind the curve. While other institutions were developing rules and procedures and even providing students with access to generative AI tools for free, we had done comparatively little. I had a general suspicion that *many* students were using these tools already, and an anonymous survey I conducted in June indicated that roughly 2/3 of our students were indeed using generative AI tools in the context of their studies.⁹ By the beginning of the 2023-24 academic year my Faculty had adopted formal policies governing the use of generative AI tools by students (with a default of essentially forbidding their use) – but there was no guidance or advice on possible ways in which academic staff could ethically incorporate them into the classroom. I have argued previously that there is much to be gained by assessing a student’s ability to use new AI tools in ethical and appropriate ways that might help prepare them for a world of employment in which they will be expected to show appropriate competencies¹⁰ There are obvious benefits in a law school to assessing the ability of students to both use AI tools and understand their limits as part of the preparation for a legal career. In a course with a small number of students, one might imagine teaching students to use AI to help structure an argument or to aid in certain forms of research. But incorporating such forms of learning and assessment into a larger lecture course (at least as they are run in my Faculty) did not seem realistic, due to resource constraints. Allow me to set out the context.

For the autumn of 2023, I was assigned to teach “Principles of Hong Kong Constitutional Law”, a large lecture style course in which 199 students were registered. These students were divided into three large groups, and each group was given a separate three-hour lecture each week, for thirteen weeks. So, I taught the same lecture three times each week, to around 70 students a time. This course is a required one for all students who wish to qualify as a barrister or solicitor in Hong Kong.¹¹ The Law Society of Hong Kong mandates that “core” courses such as

⁹ Results on file with the author.

¹⁰ Stuart Hargreaves, ‘Words Are Flowing out Like Endless Rain into a Paper Cup’: ChatGPT & Law School Assessments’, (2023) 33 Legal Education Review 69.

¹¹ Following completion of a qualifying law degree such as an LLB or JD, all prospective barristers and solicitors in Hong Kong must complete the Postgraduate Certificate in Laws, or ‘PCLL’ – a one-year skills-based programme delivered by each of the local law schools. Entry into the PCLL is contingent upon completion of a law degree from a Hong Kong institution which includes a series of ‘pre-requisite’ courses, or a qualifying common law degree from another jurisdiction along with conversion courses in specified areas. See, eg, ‘Becoming a Solicitor’ *The Law Society of Hong Kong*, (online, February 2023) <https://www.hklawsoc.org.hk/-/media/HKLS/Home/Maintain-Standards/Admission-Standards/Booklet_on_becoming_a_solicitor_202302.pdf>.

this be examined by an in-course examination worth no less than 50% of the final grade.¹² While in theory this means I am free to adopt other assessment methods for the rest of the final grade, in practice this does not happen; the typical mode of assessment for large lecture courses in my Faculty is a 100% in-class final exam. While there is ample literature arguing convincingly *against* such high-stakes summative assessments,¹³ it persists in many law schools both in and out of Hong Kong¹⁴ for a straightforward reason: resources, or rather a lack thereof. The provision of teaching assistants in law schools is much less common than it is in other departments or faculties within universities.¹⁵ Trying to “reduce the stakes” by, for example, adding in two shorter written assessments alongside the summative exam required by the professional body would have required me to personally assess roughly 600 pieces of work for this course alone.

This context drove the goals I had when I decided to experiment with AI tools as teaching aids. My intent was not to lessen my workload, but rather to leverage them to allow me to achieve pedagogical goals which otherwise were out of reach. In particular, I wanted to add formative assessment components to the course. Formative assessment is sometimes referred to as ‘assessment for learning’,¹⁶ and ‘refers to frequent, interactive assessments of student progress and understanding to identify learning needs and adjust teaching appropriately’ rather than the credentialling of acquired knowledge associated with summative assessment.¹⁷ Research in this area suggests that formative assessments

¹² ‘Benchmarks for the PCLL’, *The Law Society of Hong Kong* (online, June 2007) <https://www.hklawsoc.org.hk/-/media/hkls/pub_e/news/societyupdates/20070907.pdf>.

¹³ See, eg, Peter T Knight, ‘Summative Assessment in Higher Education: Practices in Disarray’ (2002) 27(3) *Studies in Higher Education* 275; Lorrie A Shepard, ‘The Role of Assessment in Learning Culture’ (2000) 29(7) *Educational Researcher* 4; Rich Tapper, *The Problem of High Stakes Assessment in Public Education* (Report, 1997); Sarah French, Ashton Dickerson and Raoul A Mulder, ‘A Review of the Benefits and Drawbacks of High-Stakes Final Examinations in Higher Education’ (2023) 88 *Higher Education* 893.

¹⁴ See, eg, Alex Steel, ‘Rethinking Assessment in Law’ in *Critical Legal Education as a Subversive Activity* (Routledge, 2022); Judith Welch Wegner, ‘Law School Assessment in the Context of Accreditation: Critical Questions, What We Know and Don’t Know, and What We Should Do Next’ (2018) 67(2) *Journal of Legal Education* 412; L Danielle Tully, ‘What Law Schools Should Leave Behind’ (2022) 2022(4) *Utah Law Review* 837; Gerald F Hess, ‘Heads and Hearts: The Teaching & Learning Environment in Law School’ (2002) 52(1/2) *Journal of Legal Education* 75; Andrea A Curcio, ‘Assessing Differently and Using Empirical Studies to See if it Makes a Difference: Can Law Schools Do It Better?’ (2009) 27 *Quinnipiac Law Review* 899.

¹⁵ This is, in part, because there are a comparatively small number of students in any institution who study law at the postgraduate level and realistically be able to help with the assessment of work of students completing their first law degree. Even those who advocate for the wider deployment of teaching assistants in law schools assume they would be drawn from the existing ‘first law degree’ student pool and so would only be used to do things like help guide small group conversations, etc: see Jay M Feinman, ‘Teaching Assistants’ (1991) 41(2) *Journal of Legal Education* 269.

¹⁶ Dylan Wiliam, ‘What is Assessment for Learning?’ (2011) 37(1) *Studies in Educational Evaluation* 3.

¹⁷ ‘Formative Assessment: Improving Learning in Secondary Classrooms’, *OECD* (online, November 2005) <<https://www.oecd.org/education/ceri/35661078.pdf>>.

can improve learning outcomes.¹⁸ Formative assessment theory can in large part be traced to the instructional design work of Sadler, who argued that summative assessment can have no impact on learning – it comes at the end of the process as a matter of evaluation, often leading to certification.¹⁹ In contrast, formative assessment is about ‘shap[ing] and [improv]ing student’s competence’ by judging the quality of student responses.²⁰ Though the formative assessment literature frequently considers it in the context of the primary and secondary classroom, the ideas are equally applicable in higher education.²¹

What matters for formative assessment is the impact of the assessment tool rather than its specific mode. It need not only be graded written work such as a mid-term exam: ‘any instructional activity that allows teachers to uncover the way students think about what is being taught and that can be used to promote improvements in students’ learning can serve a formative purpose’.²² This can include various forms of self-assessment²³ as well as instructor-developed low-stakes quizzes.²⁴ Conventional scores are not always necessary – for example, a ‘traffic light’ approach that allows student to self-report their understanding as ‘green’, ‘amber’, or ‘red’ can still provide the instructor with sufficient information to alter their approach to the delivery of material as necessary.²⁵ As Furtak puts it, formative assessment can refer broadly to the ‘material artifacts, tasks, or tests that *teachers and students* use in classrooms to elicit student thinking during the course of instruction.’²⁶ Effective tools, then, are both those that allow teaching staff and students alike to accurately gauge student understanding in order to adjust their approach as required. Done properly, formative assessment can create a kind of feedback loop: Teachers use feedback to make programmatic decisions with respect to readiness, diagnosis and remediation; students use it to monitor the strengths and weaknesses of their performances, so that aspects

¹⁸ See, eg, Paul Black and Dylan Wiliam, ‘Assessment and classroom learning’, (1998) 5(1) *Assessment in Education: Principles Policy & Practice* 7.

¹⁹ Royce D. Sadler, ‘Formative assessment and the design of instructional systems’ (1989) 18(2) *Instructional Science* 119, 120.

²⁰ Ibid.

²¹ Rebecca Morris, Thomas Perry, and Lindsey Wardle, “Formative assessment and feedback for learning in higher education: a systematic review” (2021) 9 *Review of Education* e3292.

²² Elise Trumbull and Andrea Lash, *Understanding Formative Assessment* (Report, 2013) <https://www2.wested.org/www-static/online_pubs/resource1307.pdf>.

²³ Alexandra Cosi, Núria Voltas, José Luis Lázaro-Cantabrana, Paula Morales, Mireia Calvo, Silvia Molina, and María Ángles Quiroga, ‘Formative Assessment at University Through Digital Technology Tools’ (2020) 24(1) *Profesorado* 164.

²⁴ Rebecca Morris, Thomas Perry and Lindsey Wardle, ‘Formative Assessment and Feedback for Learning in Higher Education: A Systematic Review’ (2021) 9:e3292 *Review of Education* 1.

²⁵ James W Popham, ‘Defining and Enhancing Formative Assessment’ (Speech, Annual Large-Scale Assessment Conference, 15 September 2006) 4.

²⁶ Erin Marie Furtak, ‘Formative Assessment’ Routledge Resources Online (online, 30 May 2022) <<https://www.taylorfrancis.com/entries/10.4324/9781138609877-REE62-1/formative-assessment-erin-marie-furtak?context=troe>>.

associated with success or high quality can be recognized and reinforced, and unsatisfactory aspects modified or improved.²⁷

I sought to use generative AI tools to help create this kind of feedback loop, creating the space for both me and my students to develop a better sense of their understanding of the lecture material *throughout* the course, rather than waiting simply for a final grade. This would allow both sides to modify our approaches in the hopes of improving ultimate outcomes.

III THREE WAYS OF DEPLOYING GENERATIVE AI IN A LARGE LECTURE COURSES

With this background, I considered possible ways in which I could adopt different AI tools in the delivery of the course that would provide students with opportunities for informal formative assessment. In the end, some of them were useful in the delivery of course material as well, again suggesting that there is much to be gained by academic staff experimenting with different ways of using AI. Below I describe three techniques in which I used generative AI tools: lecture transcription & summarization, question generation for “no-stakes” quizzes, and student self-revision via chatbot. After each, I also provide the results of a student survey regarding each tool.²⁸ There were 10 questions in total, asking about the students’ perceptions of the value each of the three usages of AI adopted on a Likert scale, whether they would like to see similar tools used in other courses, and freeform questions allowing students to leave general comments or views.

There are two important disclaimers to make. First, I have no affiliation with any of the companies or services described in the next part. I selected them in a very informal way – I simply tried to read as much as I could online about the various new AI tools coming on stream and then brainstormed ways in which I could use them. If I decided a tool might be useful and the university did not already have some of kind license agreement with that service, then I paid for a subscription out of my own pocket (the short timeline I had between deciding to start this project and when the course would actually begin made application for grant funding implausible). Second, to repeat I have no intent here to be prescriptive. I am not claiming these are the only or best ways in which AI may be implemented in the delivery of large lecture courses in a university setting – I am simply describing my own experience and, again, I encourage all academics to play around with AI tools and find what works and what does not work in their particular context.

²⁷ Ibid 120-1.

²⁸ The survey was conducted using Microsoft Forms, with students completing it outside of class time. Of 199 students, 99 participated in the survey. It was conducted in the final week of the term (that is, prior to the exam). All responses were voluntary and anonymous, and students were informed of the purpose of the survey. There was no way for me to tell which students responded and which did not, and a decision to participate or not could have no impact on the student’s final grade. For the substantive answers provided by students, I have made no effort to proofread or edit the text so as to eliminate the possibility of incorrectly stating their views.

A *Technique 1: Lecture Transcription & Summarization*

OtterAI²⁹ is a tool that transcribes recorded audio into text. While recordings are made on a user's local device, the audio is uploaded to the cloud where OtterAI's algorithms analyze it in order to generate a transcript. I used my phone to record the audio of my lectures, and I was able to download a transcript in raw txt form for further processing. Inexperienced with AI-powered transcription, I found the process easy and surprisingly accurate. It did, however, sometimes struggle with the names of pieces of legislation and frequently struggle with Chinese party names in case titles. In such cases, I would go back through the transcript to correct where necessary before further processing.

I made a choice to not provide these transcripts directly to students for several reasons. The first was due to accuracy – while as noted overall I found the accuracy to be relatively good, if I wanted to provide direct transcripts to students, I think there would be an obligation for me to ensure the transcript was a completely accurate representation of what I said. Missing out a word like “not” or mishearing what I said could result a transcript that was a very inaccurate representation of the law or my views. Given the course dealt with what in Hong Kong are sometimes considered politically sensitive issues, this was problematic. But going through the transcript line by line in detail before providing it to students would have taken an unnecessary amount of time, and to an extent defeated the purpose of relying on AI tools as pedagogical assistants. The second reason was that I did not want students to stop taking notes or paying attention because they thought they would be provided with a ‘perfect’ transcript of each lecture regardless of whether they attended. My goal was to provide them with a good *summary* of each lecture that they could then compare with their own notes to see if their own understanding was on track.

Generative AI tools like OpenAI's ChatGPT³⁰ are believed to be quite effective at not only generating new text in response to natural language prompts, but also summarizing existing material. In summer 2023, when I was developing my plans for this course, the University had agreed to provide staff with free access to GPT 3.5.³¹ However, my initial experimentation with summarizing lecture transcripts was unsuccessful. GPT 3.5's context window was relatively small, and it would simply stop summarizing after a certain amount of text. GPT4 was much more effective, however.³²

²⁹ *OtterAI* (online) <<https://otter.ai/home>>.

³⁰ ChatGPT, *OpenAI* (online) <<https://chatgpt.com>>

³¹ Dylan Butts, ‘Microsoft to Bring OpenAI to Eight Hong Kong Universities to Deploy Generative AI Tech in Education’ *South China Morning Post* (online, 29 August 2023) <<https://www.scmp.com/tech/tech-trends/article/3232691/microsoft-bring-openai-eight-hong-kong-universities-deploy-generative-ai-tech-education>>.

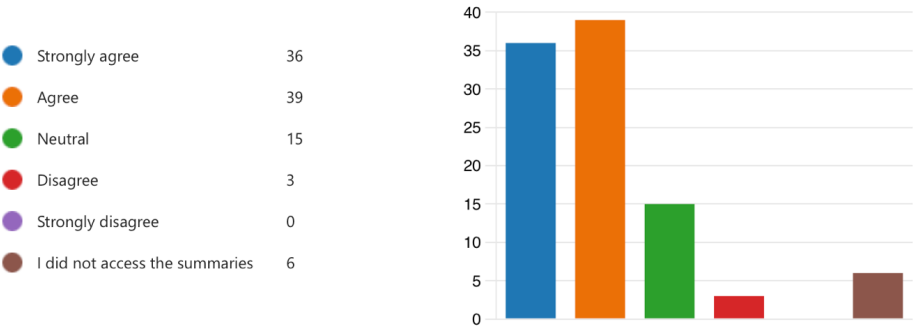
³² ‘GPT-4 is OpenAI's Most Advanced System, Producing Safer and More Useful Responses’, *OpenAI*, online: <<https://openai.com/index/gpt-4/>>; I had to pay for access to GPT4 at the time the project ran.

After some experimentation I settled on the following prompt: ‘You are a law professor teaching graduate students. Based on the transcript of the lecture provided, create a 2- or 3-page summary of that lecture from which they may study. Ensure key points are summarized’.

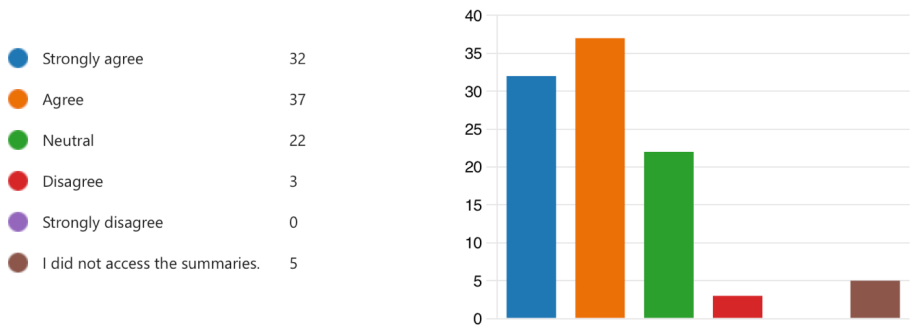
I then took the resulting output, pasted it into a Word document, and made *that* document available to students using Blackboard, the learning management system in use at my institution. I included a disclaimer at the top of each summary that it was AI generated and that I did not endorse it for accuracy, and that students should decide for themselves the extent they wished to rely on it after comparing it with their own notes. Despite that disclaimer, I did review the summaries for any egregious errors – but this was a very quick process since each was only a couple of pages of bullet points.

1 *Student Feedback – AI-generated lecture summaries*

(a) *Question 1 - The AI summaries helped me better understand the lecture material.*



(b) *Question 2 - The AI summaries were useful for my exam revision.*



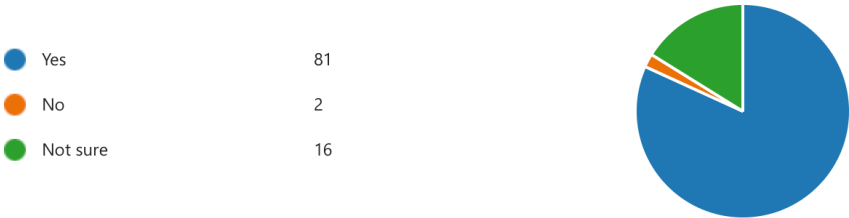
(c) *Question 3 – Please describe how you used the AI summaries and any feedback you have.*

- 1 good!
- 2 Save time for doing revision
- 3 Use to make notes
- 4 Clearly read for the materials
- 5 I used them for quick review on a class
- 6 May receive some important points of the lecture and may communicate with lecturer effectively.
- 7 Read weekly
- 8 To help with understanding the structure and specific questions
- 9 Help review every week's key point, convenient and timesaving
- 10 to use them as reference to compare with my own notes
- 11 Help with the consolidation of the notes
- 12 to have a quick review after class
- 13 Thorough the second-round review
- 14 I was worried it wouldn't be correct and I thought I would still need to read the course materials on my own so I did not access it
- 15 I used it instead of the recordings, which saved a lot of trouble for me. I think it's kinda concise and not that detailed, which might be a little bit concerning, but it would be backed up by the powerpoint, so not a very big deal.
- 16 help me to formulate the structure of the course, and I can compare my notes with it.
- 17 To prepare my notes
- 18 It helps me make the structure of the constitutional law system, and also helps me better understand the complicated parts.

- 19 Normally, at the end of the week, after I make my own notes, I would double-check them with the summaries provided by the professor to make any supplement. And the summaries also provide a clear logic and structure of each topic.
- 20 use AI to review the class content It's useful
- 21 I think it is very helpful. Because there is so much content in the slides and through the summary i can capture the main important point in the class to help me review.
- 22 The summaries were good to read before going over the topic. It helped me understand the topic before filling in the details
- 23 Basically to get a general and rough idea of what has been gone through in the lecture. The summaries tend to be rather short and brief. It is useful for students who have attended the lecture to retain memory of what they got out of the lectures.
- 24 very good
- 25 After doing my own notes, I like to see the summary and try to supplement my notes.
- 26 use it as a guideline where detail will be added my me in reviewing and note revision
- 27 Used as a general outline to organise my note.
- 28 I use the AI generated summaries to take out and highlight keywords that help me memorize for exam
- 29 It helps me understand the course structure before getting into details
- 30 Used as a summary within revision to draw key concepts from topics.
- 31 I personally feel like because this course is about essay writing, which requires a lot of details knowledge to the course. Therefore I feel like a summarized tool is not that important specifically for this class.
- 32 I haven't read the Ai summaries yet but I will do so as part of my revision. They will be read before reading the entire PowerPoint slides for the lectures.
- 33 I read a few times but stopped using it as the AI summaries are too brief
- 34 I used it to help me in making my own outline and for reviewing purpose after class
- 35 It's a decent summary, but it was not elaborate enough to be useful in revisions. Suit better as a pre-lecture study to get a better sense of the class and better understand the substantive material in class
- 36 I used AI-generated summaries to compare with my own notes during the midterm revision to see what I have missed or misunderstood.
- 37 The fact that the summaries were not read and approved by the professor made me disinclined to rely on using them.
- 38 I used it for the first 4 weeks since at that time i was so overwhelmed by constitutional laws
- 39 I use it to revise as it gives out a brief summary on the topics.

- 40 Only check on the AI summary when Im confused about the concepts after reading through the all materials. But It seems not that useful to help my understanding.
- 41 prepare myself for lecture
- 42 Given that the summaries are not guaranteed to be accurate, although it helps summarising the professor's speech, it still feels risky to really study them.
- 43 Provides an overall understanding of the topic
- 44 It does highlight what the professor lectured in a very streamlined way.
- 45 As a overview, it's good to generate brief idea of the topic
- 46 The summary provides a road map to the class teaching.
- 47 Have a preliminary idea of general structure of the course
- 48 Read as a summary of lecture materials , sometimes not quite clear
- 49 If the AI notes can be cleaned up and checked for accuracy, that would be helpful. Otherwise, not sure how accurate the information is and will need to spend time cross checking against slides/notes from class

(d) *Question 4 – Would you like AI summaries to be provided in other courses?*



B *Technique 2: AI question-generation for “no-stakes” quizzes*

I used GPT4 to process my own lecture notes and asked the system to generate a series of quiz questions based on those notes in different styles (multiple choice, true or false, short answer, traffic light). I generally asked for ten questions and then chose the five that I felt were most useful, sometimes tweaking them slightly. On occasion I would manually develop a different question myself if I felt some key aspect I wanted to review had not been covered by the AI tool. In advance of each lecture, I uploaded those questions to Mentimeter,³³ a web-based tool that allowed me to present the questions to students ‘live’ during the final fifteen minutes of the lecture. Students used their mobile devices or laptops to answer the questions anonymously after entering

³³ *Mentimeter* (online) <<https://www.mentimeter.com>>.

in a code into Mentimeter.³⁴ No records were kept and the answers provided (or a decision to not participate in the quiz at all) had no impact on a student's grade. At the close of the quiz, I was able to instantly present the results on the classroom projector.

Importantly, Mentimeter allowed me to not only show my students the correct answers, but also how everyone else had answered (that is, how many people had chosen option a, option b, option c, etc, or what 'short answers' they had typed in). This allowed students to gauge their understanding of the week's material but also to informally discuss with their peers the reason they chose a particular answer – many visibly relaxed and laughed when they realised many people had made the same error as they had. This allowed me to give immediate feedback and also served as a springboard for discussion to further consolidate learning of the content of that particular lecture. If I could see that 2/3 of the students were selecting the same incorrect answer, then I knew that some aspect of the lecture had not been delivered in the way that I had hoped – and I then had a chance to immediately correct it. It was an ideal example of formative assessment, and as the surveys below reveal, the students appeared to greatly appreciate it.

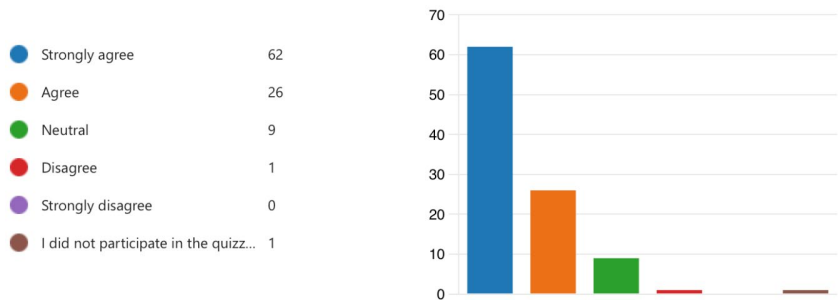
Three caveats are worth noting. First, it is not strictly necessary to rely on AI tools to generate quiz questions. The benefits are in speed or in scale. AI greatly simplifies the process, particularly as the desired number of questions grows or if there is a desire to avoid recycled questions year on year. Second, while GPT4 was quite effective at generating multiple-choice questions from my lecture notes, it was less so at creating "short answer" questions that required students to synthesize various elements. It was also relatively poor at developing "true or false" questions. This reflects the inherent weaknesses of large language models like GPT in their current iteration – while unquestionably impressive and of real value, they are still to an extent 'stochastic parrots'.³⁵ At the risk of over-simplifying, they generate text based on the probabilities and parameters contained in their training datasets. At the time of writing (January 2024), they cannot perform "critical analysis" because they have no basis for locating critique; they lack the concept of truth and falsity. Third, if you request the AI to generate a relatively small number of questions it is likely that it will miss out at least one or two central ideas contained within the lecture, and so in those scenarios instructors would likely want to still want to include some of their own questions as well. Nonetheless, the result of my experiment was that the use of generative AI to help me with question generation (and Mentimeter for quiz delivery) meant that with relatively low effort I was able to offer weekly informal formative assessment for over 200 students.

³⁴ While I did not formally survey the students about this, as far as I could tell there were no students who came without either a laptop or mobile device such as a tablet or smartphone.

³⁵ Emily M Bender, Timnit Gebru, Angelina Macmillan-Major and Shmargaret Shmitchell, 'On the Dangers of Stochastic Parrots: Can Large Language Models Be Too Big?' (2021) *FACCT'21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* 610.

1 *Student Feedback – AI-generated quizzes*

(a) *Question 5 – The quizzes helped me gauge my own understanding of the material.*



(b) *Question 6 – Please provide any feedback you have on the quizzes.*

- 1 Help make relevant knowledge more available and really provide a concrete basis for learning much content
- 2 able to see other people's questions
- 3 Very fun and useful for class review
- 4 very efficiency and encouraging communication
- 5 Let me know things that i don't understand.
- 6 Great
- 7 good to clear concepts learned in the lectures
- 8 It guarantees the chance to ask some longer questions.
- 9 It helps me efficiently go through the key points learned in this class and clarifies the questions that might be confusing
- 10 No comments
- 11 quick review after each class; when i find answers wrong, i will check the notes & clear the wrongs
- 12 It can quickly check any misunderstanding concepts I have, and then strengthen the good concepts for me.
- 13 It was a bit misleading to me that I thought facts were the most important and I was totally missing the big picture until W6
- 14 It helps summarize the key points and main focus of the class.
- 15 quick and clear review of the course, very convenient
- 16 Kind of exciting and straight forward to see the potential points in the materials

- 17 It is interesting to see others' answers.
- 18 The quizzes help me to review some important facts about the topic. It does help me to clarify some confusing points.
- 19 It's useful to go over what has learned in class
- 20 very good, it can help me cover some important points in the class.
- 21 They were a fun, interactive way of seeing what I had understood from the lecture itself
- 22 It's a nice quick way to check whether students have grasped the basic idea from materials used in lectures.
- 23 Very good
- 24 Perfect and interesting quiz at the end of each lecture
- 25 Just general good wrap up of course material
- 26 Helps me solidify concepts after a lecture
- 27 It can consolidate the newly learnt concept
- 28 Mentimeter quizzes provided very nice overview at end of class to sum up material. Also appreciated the use of the menti live questions.
- 29 I think it's helpful to assess how much I learn in a class period and what I should focus on after class if I get some of the questions wrong.
- 30 It is a quick but effective check on the key concepts, as well as fun and know the the standing of the rest of the class
- 31 I really like it , it help me retain the knowledge learnt in class
- 32 Seem to good enough easier to gather class participation and attention
- 33 I hope instructor can upload a summary of all quiz questions on BB as well to give students a full picture of the weekly important topics. These quizzes are very helpful for me to assess my mastery of the lecture.
- 34 The quizzes get people involved and on their feet about content they've just learnt.
- 35 Really helpful
- 36 It was great as it examines my knowledge on the concepts
- 37 In-class realtime check on understanding of the course material is very efficient and helpful.
- 38 It was a very useful tool to check my understanding of the materials.
- 39 fun
- 40 They are useful to see how classmates answer the questions and to give more confidence to ourselves.
- 41 It's a quick review of the lecture materials - very interactive and useful!
- 42 Great way of checking your understanding at the end of the class.
- 43 It immediately refreshed the memories

- 44 Having fun and increase motivation
- 45 Identify my weaknesses
- 46 Simple reflection of key points in the class and quick feedback about my understanding
- 47 Quite good, remind us of details discussed during classes, but seemingly many tools can do the same
- 48 Loved it! Fun and reinforced what was covered during class. Also helpful it was reviewed at the end of class

(c) Question 7 – Would you like these kinds of quizzes to be adopted in other courses?



C *Technique 3: Student self-revision through chatbot*

The first two techniques were deployed right from week one; the third not until the course was about two thirds over. About halfway through the term, I became increasingly aware of open-source large language models such as Meta’s LLaMA.³⁶ I wondered whether it would be possible to fine-tune³⁷ an open-source model on my own course content, including my lecture notes, the slides, past exams, etc. After discussion with a computer science PhD student who happened to be taking law courses as well, it became apparent that this was hopelessly naïve on my part, at least with the limited time and resources I had. I likely would have needed to access to university computer resources for the fine-tuning and to make any resulting model available to students, as well as several funded research assistants to help with the process.³⁸ With only one month left in the course at that point, I abandoned the idea.

³⁶ ‘Discover the possibilities with Meta Llama’, *Meta* (online) <<https://llama.meta.com>>.

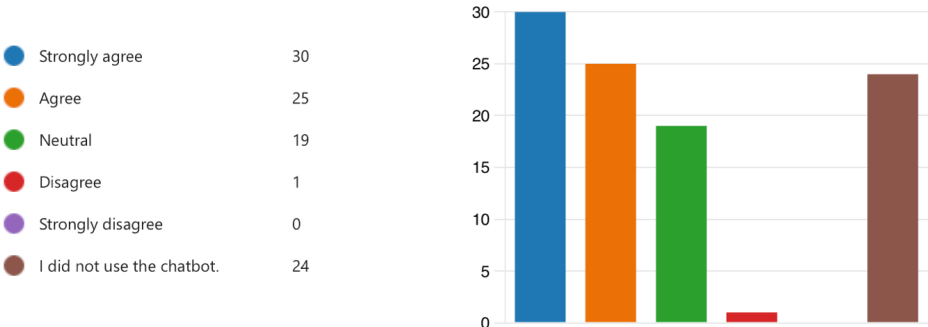
³⁷ Shri Varsheni, ‘Guide to Fine-Tuning Open Source LLM Models on Custom Data’, *StackAbuse*, (online 6 July 2023) <<https://stackabuse.com/guide-to-fine-tuning-open-source-llms-on-custom-data/>>.

³⁸ Nathan Lambert, Louis Castricato, Leandro von Werra and Alex Havrilla ‘Illustrated Reinforcement Learning from Human Feedback (RLHF)’, *Hugging Face* (online, 9 December 2022) <<https://huggingface.co/blog/rlhf>>.

However, discussions with my University’s information technology centre (ITSC) led to a more modest solution – a basic chatbot that would allow students to make queries of the course’s slides only. I was not involved at all with its technical development.³⁹ The chatbot was made available to students only in week 10 (of 13) and was very limited in scope. Students could ask basic questions such as “what was the central holding in [case X]?” and the chatbot would provide an answer drawn directly from the text of the material I provided if I had mentioned case X in it and specifically used the phrase “central holding”. It could not perform any kind of deeper analysis or reasoning nor supplement its answers with any information that might be found in the training data of a large language model, open source or not. While (as the student feedback shows in the next section) this was the least effective pedagogical use of AI with which I experimented, I think it nonetheless has the most promise and I consider why in Part 4.

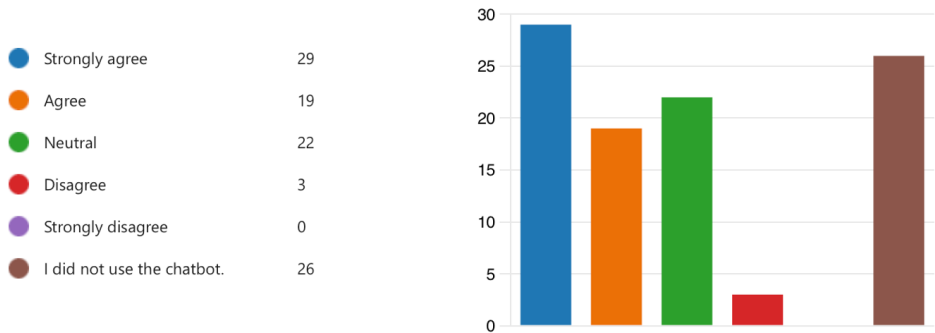
1 *Student Feedback - Chatbot*

(a) *Question 8 – The Chatbot helped me better understand the lecture material.*



³⁹ ITSC used a variety of tools to develop and host the chatbot, including FlowiseAI <<https://flowiseai.com>>, Azure OpenAI <<https://azure.microsoft.com/en-us/products/ai-services/openai-service>>, Render <<https://render.com>>, and Pinecone <<https://www.pinecone.io>>. I was not involved in the technical process.

(b) Question 9 – The Chatbot was useful for my exam revision.



(c) Question 10 – Please provide any feedback you have on the Chatbot.

- 1 very efficiency and encouraging communication
- 2 Might find the same question that i have
- 3 Really great to answer questions
- 4 can help me review the materials in a quick way, save time on flipping notes/ slides
- 5 Assisted in my revision for the final with relatively detailed contents from the chatbot
- 6 Actually AI could provide a brief of the concepts, but not too much details. Therefore, it helps in some way.
- 7 I didn't use it, but I believe if it's used, it would be really helpful to the class. I mean, why not?
- 8 They can be used to answer some questions but the answer may not be accurate or even wrong.
- 9 I have not used yet.
- 10 sorry I have not used the chatbot
- 11 Very helpful! I asked some questions like the comparison between the different interpretation methods and other things. It can give a good summary and provide the main point in the class. However, I also once asked one question regarding the difference between the Decisions and interpretations of NPCSC. But it cannot give an answer.
- 12 The chatbot helped me fill in small gaps in my knowledge
- 13 N/A
- 14 good
- 15 Just hard to use

- 16 The chatbox sometimes does not provide me with exact answers that I want, e.g. does the living tree approach equal purposive approach?
- 17 Can consolidate all the questions from other students and induce discussion
- 18 Chatbot AI questions very basic in understanding and the answers which provided were easily found from within lecture material
- 19 I did not use the chatbot provided by CUHK because I already have a subscription to a more powerful chatbot.
- 20 We cannot go back to check the questions and responses
- 21 N/a
- 22 It would be great if the chatbox questions and answers could be provided after class
- 23 quite useful
- 24 Great to have it so that we can post live questions during the lecture.
- 25 None
- 26 Not sure which chatbot you referring to

IV SURVEY ANALYSIS

The responses to the survey questions make it clear that students generally responded positively and enthusiastically to various uses of AI within this course. 75% of respondents agreed or strongly agreed that the use of AI summaries helped them better understand the lecture material, and 69% agreed or strongly agreed that those same summaries were useful for their exam revision. Students described them as helpful, time-saving, and convenient. They found the AI summaries provided a clear overview of the material, and they were able to compare them with their own notes as a reference, as well as pointing them in the direction of the key ideas to focus upon. They were very useful, then, as a tool that allowed students to engage in self-revision and self-assessment in an informal manner. No wonder, then, that 81% of them wished to see similar AI summaries provided in other courses.

The AI-generated quizzes deployed through Mentimeter were also extremely successful as far as the students were concerned. 89% of students agreed or strongly agreed that the quizzes helped them gauge their own understanding of the course material. The decision to deploy the quizzes in a collective, “no-stakes” approach during class time was highly effective. Students describe participating in such quizzes as fun, interactive, and exciting. They found them clear and convenient, making knowledge more available to them and helping guide their own decisions about where to focus their review efforts. I was able to leverage AI to provide students with highly effective weekly opportunities to engage in informal self-assessment and reflection upon the material – with very little extra effort on my part. That 92% of the respondents said they wished to see these kinds of quizzes adopted in other courses is testament to the value they provided.

The surveys suggested that the students were less keen, however, on the third tool adopted – the AI-powered course Chatbot. Less than half the respondents agreed or strongly agreed that it helped them understand the course material better or was useful in their exam revision, while around a quarter reported they never used it. Students seemed more ambivalent as to its value as compared to the other AI tools used. I consider this relative lack of success down to several factors. First, the obvious limitations of the Chatbot (it could only respond with information directly from the PowerPoint slides that formed its knowledge base, and only then if exact phrasing was used) meant it was not that much more sophisticated than a search engine applied to that material. Second, that Chatbot was only completed and made available to the students relatively late in the course, meaning they had less time to experiment with it. Indeed, the survey indicated that many students had not even realized there was a Chatbot for them to use; late in the term, students are perhaps overwhelmed with the various emails they receive. Finally, students appeared skeptical of the Chatbot's accuracy. At the start of the course, I had provided students with a link to my earlier research on ChatGPT and law school assessments, noting its propensity to 'hallucinate'. It is plausible that students associated the Chatbot with ChatGPT, even though the systems were distinct (ironically, they appeared to have no such qualms with the AI summaries, which *were* created in part with the same technology that powered ChatGPT).

V CONCLUSION

This paper has described experimenting with different generative AI tools in the delivery of a large lecture course in a law school. Though the results are by no means intended to be prescriptive, the basic concepts should be applicable across any discipline that continues to rely at least in part on a traditional "lecture" as a component of course delivery. Three techniques were identified: AI-assisted lecture summarization, AI-assisted quiz generation, and an AI-powered chatbot. The first two were achieved using commercially available tools at relatively low cost, and require no particularly advanced technical skills to deploy. The last was more dependent on computing expertise. Student surveys indicated that the generation of summaries and quizzes was particularly positive, and clearly offer a straightforward route to creating positive feedback loops of formative assessment. Moreover, regardless of class size these AI supported forms of assessment can be deployed by a single instructor alone, without requiring any kind of teaching assistant or similar support. The third usage – the chatbot – was less successful, however it is easy to imagine that a more robust version of the concept could be transformative for course delivery and assessment in higher education. The problem was with implementation, rather than concept. An effective tool that could respond to students' queries 24/7 in a helpful and patient way, grounded in a complete

understanding of the material would be an incredible boon for student-led learning.

We remain in the experimental stage of AI in higher education, and to repeat, I present nothing here as a prescriptive solution – simply the results of my own experimentation. I do not mean to suggest that every academic needs or even should incorporate AI into their teaching. But, the reality is that AI tools are certain to shape both the learning environment and the careers of our students. We owe it to our students to become familiar with the strengths and weaknesses of those tools, and one of the best ways to do this is to experiment with them. I do strongly encourage each of you to at least explore them – if only to get a better feel for the new waters in which your students are swimming. They, after all, will have no choice; the AI tide seems unlikely to recede. Even during the brief time of this project, the ability of generative AI tools rapidly improved: frequency of hallucinations lowered, context windows increased, and Large Language Models began to show evidence of genuine analytic reasoning. Universities and individual academics ought to take seriously not only the possible risks these tools pose to our traditional modes of content delivery and assessment, but also the potential such technology can bring into the higher education classroom.

VI EPILOGUE

The relative success of these experiments led me to expand upon them. After writing this paper I successfully applied for a grant to further explore the way in which commercial AI tools can be leveraged as teaching aids. In academic year 2024-25, I used AI to provide my students with video summaries of lectures presented by what I call “teaching avatars”, podcasts with two AI-hosts discussing the week’s material in a conversational way, a significantly refined course Chatbot, and a simulated client interaction powered by audio AI. A subsequent paper will detail these experiments as well as student reactions gathered through surveys.