

Identifying Educational Benefits in the Use of AI for Legal Studies

Well. Good afternoon. Thank you very much for having me here. Thank you for your talk. Thank you for talking about pathos and ethos, passion and morality, which is something that legal professionals should have if they want to be successful. I'm going to talk about my personal voyage in getting to know AI and how to use it. I've decided to start doing that because I don't trust AI. I don't trust the fact that it's so heavily funded and so extraordinarily sold to every institution that I don't know how much of that is true and how much of that is just fiction. So I decided to learn how to use it. I must acknowledge the help of William and Matthew from the TED. They have been wonderful in helping me, even in the middle of the night, and, of course, my students who have tolerated my experiments on them.

So in the first semester, I experimented in three seminars, doing focused, in my view, experiments trying to find particular aspects of the use of AI, which I'm going to describe, and deciding what is the optimal way of getting something out of it for my students. And in the second semester of this year, I have moved into the routine use of bots every week, and, actually, tomorrow is the deadline for an assessment based on AI to which I'm going to explain what the objectives for that are. I've also got the experience of using bespoke AI packages developed by some of the biggest legal publishers.

So as I've said, why am I curious? I distrust AI. I don't believe the AI hype. I believe there is a bias in what it can do. There is a logical objection to using it in law. It is based on language models, which are trained on the English language and on resources which relate to particular legal systems. So using it in other legal systems will be, by definition, problematic. But it is something that is used. Alright? So we cannot ignore it. I wanted to motivate my students. I wanted to help them prepare for class. Many of them come with only reading the PowerPoint. They have long reading lists, but they don't look at them. So is this a way through which I can make them do a little bit more thinking? They don't engage with case law. Case law is lovely if you like it, and terrible if you hate it. And they learn things, rather than thinking about the things they need to learn.

So these are the things. This is, perhaps you have seen that, this is the way the whole system works. Essentially, it's a bot which sits on top of ChatGPT, and you have a dataset that you can introduce there. So what I do is I enter a number of cases, and perhaps some text from trustworthy practitioners'

books to sit on top of the ChatGPT version, and I demand that legal referencing is always provided to the students. So whatever they get, if they get an answer, there must be a legal reference. I did start by planning to put lots and lots of materials on whatever cases I could find. That's impossible with the tools we have. There are some limits, but still, you can easily put 140-150 pages of text in there. The problem with law is that some of the cases are longer than 150 pages, and therefore, you do have a problem with that.

So this is what the students see, or used to see in the first semester. They would log in, they would select the course, and they would get to the prompt. The interesting thing about the prompt is that the answer is translated into Chinese, so they could ask in English and get the answer in Chinese, which was very helpful for some of my LLM students who came from the Mainland and did not have the language skills, which were competent to do this work quickly in class. So there are two courses. One is LLM, the other is LLB and JD. LLB is for undergraduates. LLM is for postgraduates. JD is for postgraduates, but for people who want to become lawyers. So this is really the professional or the preparation for professional courses, and perhaps there is a difference in the attitude of the students and the requirements. So the thing is we need to consider, for which students and at which level we should be using that, and I will indicate how this came about.

So in the first one, I did three seminars, three experiments. In the three seminars, you can see the numbers of students, so experiment one was about familiarization. I actually made the students use the bot in class. I had given them three Hong Kong court cases from the reading list. I explained to the students what the bot was doing so that they could understand what the difference is between the bot they're using at home, when they use ChatGPT, and I did an end survey to find out their feelings and their attitudes towards the work. In the second experiment, we got a very long case, 150 pages, a very important case, describing a particular aspect of the legal system for enforcement of foreign judgments in England. I've provided them with a list of starting questions so they can start looking into the case and try to find out what the aspects are. Basic questions. "Who were the judges? Who were the parties? Which court was the original decision? Which court decided? Whether it's enforceable?" And then, going to the reasons. "Why did they decide that?" So I actually guided them on what to ask. Of course, I didn't know what the answers would be. And then, I did an end survey to

see how they felt after using it again.

But I also used a pre-use quiz, which tested substantive knowledge with questions about the case. “Do you know which court decided that? Do you know where the case came from? Do you know what were the aspects?” So a number of questions. And then they did the same quiz after they finished the work with the ChatGPT. And the third one was doing a tutorial problem to apply their knowledge. I provided cases and ordinances from Hong Kong, gave them the problem, and let them try it, either by dumping the whole problem into the AI or by asking specific questions to break it apart. And again, an end survey.

The starting explanation was telling them how an LLM model works, outlining that, and then telling them that I have banned their access to the internet. So they were looking for specific, good quality materials. Okay? However, even if they were looking at good quality materials, as a matter of fact, there were things coming from the training of the model. So a particular case did not have the term “punitive damages”, but the answer of the model came out that the reason the judgment was not recognized was because of, it referred to, “punitive damages”, which are the kind of damages that are used in the United States. Okay. So then the question I asked the bot, “Does ‘punitive damages’, the term, appear in the text?” And the answer was “No”. So the next question is, “Why? Why do you use it?” It said, “It came from my training”. Okay, so you cannot isolate the background. That goes back to my first comment that it is based on specific data, and you cannot get away from that.

So these are some of the survey results. Most of the students that answered the survey have used AI before. Many of the students have not prepared by reading the reading list, going through the reading list, but most of them found the use of the AI helpful for them. And you can see some of the comments at the bottom line. These are the types of guidance that I gave them for questions. I’m not going to let you read all of it. You can go through them very quickly. And this is what the second survey’s outcome was. So, “Did you find it helpful in understanding the facts and circumstances?” 78% indicated “Yes”, and 18% “Partly”, and the legal issues were almost the same. Okay, so most of the students found the use of the bot good, and they provided some comments, which are, in general, in support. But the question is, “Did the students actually learn?” They thought they learned. They felt

more comfortable, but did they actually learn anything? So there are different types of questions, and I'm not, again, going to go through them. The first one was, "Have they read the case?" The second was factual questions, "Which court decided that, and which court was deciding whether it's going to be enforced?" And then we asked some more advanced knowledge, which, I don't think it's the time to discuss. So then I compared the answers.

And the problem here was the following. I was a bit surprised about that. If you look at the number of correct answers, the number of correct answers was higher. Okay, so the students could give more correct answers, but at the same time, they gave more wrong answers as well. So if you look at the number of wrong answers before and after, okay, they gave more wrong answers as well, which means that, I think, they are becoming more confident. They have a source where they read something and they believe in this, so they're more confident to answer, but they haven't actually learned something, because the information that comes from the bot may have been wrong and they haven't realized that. So increased confidence, but the percentage doesn't change. Slightly changed upwards, okay, and the survey is, again, positive. They think they've learned by using the bot. 50% of them thought that they were motivated and they saved time. There are about a fifth of the students who say, "I could have done it without the bot faster". I'm not sure if that's correct. They wouldn't, be able to read 150 pages faster than the 20-30 minutes they did the test on. So the chatbot was found useful in motivating them, but no evidence that it actually helped them. Provisioning the appropriate questions to start them was, I think, crucial.

And the point is, how do you actually make them aware what they get from the bot is not necessarily correct? Okay, and that's what I'm trying to do in the second semester. This class now consists of undergraduates, third year and fourth year undergraduates, plus JD students. JD students have done a degree, have completed a degree in Hong Kong or another place already. They are very good students trying to become professionals. So they are seasoned students. They know how to study law. Some of them have been doing it for three years, and therefore, they are less dazzled by the use of AI. So first of all, there is less response to my prompting to use the bots and some very critical comments. "The bot fails this course." "It lacks depth and relevance." "The responses are shallow and generic." "It cannot support critical thinking." "It offers no meaningful insights." "I personally prefer

not to use AI, and have never used AI before because I learned better without it and using traditional studying.” This is the message, actually, that we started using when they first came out. Right?

Some students are reluctant to use it, despite the fact that I'm telling them, “Professionals use it, so you have to learn how to use it properly.” So the question is how we should deal with that. So I gave them an assessment that has three components. The first is to solve a mini problem by using the bot, using two bots actually. If students follow that, they have 24 hours to fix their assessments. One of the bots is open to ChatGPT, alright. They can see the internet. They don't know about it, but they can see the internet. The other bot has some materials from different jurisdictions that have contradictory rules. Okay, so it's clear that whichever way, whichever bot you use, you are not going to get the right answer. Okay, unless you specify some parameters. So that's the first point. The second point is, I want them to give me a description of their interaction between the bot and themselves. What did you ask? What did you get back? Is it right? Where do you suspect the bot is wrong? And why do you suspect it's wrong? I've seen the work of two groups, and they have done very good work. 40-50 pages of extracts with comments of where they think it's going wrong, so they are getting aware of the limitations of the bot. And the third one is to do what you suggested, to develop a statement that explains what they have done in solving the problem, in using the bot, so that they cannot be ethically accused that this is not their work. Okay, and I've asked them and encouraged them to look on the internet, because many universities and publishers are doing that, and develop it and adjust it to themselves. So I hope this is going to work for them. This is really about building the skills there.

Now, I have some doubts about all of that, right, and perhaps you have ideas of how to improve it. First of all, when I test the knowledge of the students, and I say, “Well, they haven't learned”, I haven't really measured whether the students learn when I teach them. I haven't asked them to do a quiz before and after I speak, perhaps they are doing their learning more from the bot than they learn from me. So that's one of the things. Perhaps learning is really a much longer process, and you cannot actually test it with this pre-quiz and after-quiz. I'm very much aware of that, okay, but I don't know of a better way of assessing that. Perhaps interviews are a different way of doing that. What I am certain of is that many of them are reluctant to interact with a bot, with AI. They are reluctant to ask questions. They are reluctant to ask ridiculous questions. Okay. I've encouraged them to put, “Was

Donald Trump accused in this case?” All right. Try things that are completely extraordinary, and see what the AI says, because you shouldn't be afraid of that. But most importantly, to go back, ask for the reference, because that's important for law, and go and check whether the referencing is correct.

I think we need to explain to them what the technology does. Okay, it's about correlation of words. It's not an intelligent thing that actually puts phrases together. And I think the thing that I've learned, and I was not really aware of that, but it was explained to me by colleagues in TED, is that the technology has a criterion where it constructs an answer and does not necessarily search all your materials. So it goes in and finds the first answer it can. It's like opening a book, finding a text that looks like your answer, and taking it out. And then you think it's the answer. You don't look at the whole book, right? That for law is a big mistake, alright, because it's the detail that actually matters.

I am very concerned about AI's basic mistakes in one of the professional packages that is actually out for work. I asked a very simple question, “Who were the judges in this particular case?” Okay, there were five judges. The system listed four correctly and one incorrectly. The fifth that was incorrectly listed was a lady, was a woman. It was not that it was sexist. It was that the title of this lady was Baroness. Okay, it wasn't Lord or Lady. It was Baroness, and the AI couldn't identify that. And that makes me very, very worried about using that seriously, for doing anything useful, professionally. For learning, it does motivate. It's a game. It helps the students get excited about some things. They test their knowledge. They are perhaps less afraid to ask the bot questions than the teacher. But there was one comment from a student that surprised me. When we were doing the testing class, the student said, “This is too slow.” I said, “Why are you saying this is slow? It takes less than 30 seconds to get an answer to a very complex question, I guess.” “Yes, but you will give that to me in five seconds.” So the students really prefer the personal interaction, although, of course, we cannot provide it for all the students all the time. So thank you very much. I was a bit late, but I'm sorry for that. Okay, thank you.