

# University student surveys using chatbots: Artificial Intelligence conversational agents

**Professor Eric Atwell:** What I want to do today is to show you something about our research on chatbots. I wanted to ask you to have a look at a research paper. We have a European funded research project on looking into the use of chatbots in higher education. And I'm going to quickly go through some of the highlights so you can see what you want to read.

So, you might think a chatbot is simply something where I say something, and the computer says something, and I say something, and the computers says something to chat. But actually, the dialogue is much more complicated. You have to keep track of other things that are going on and ideally, you want to have interfaces to various other IT systems that are going on.

So, here's the paper. University student surveys using chatbots: Artificial Intelligence conversational agents. I'm just going to go through some of the highlights in the diagrams and then you can read it yourself later on. So, Noorhan Abbas and Tom Pickard, they're research fellows on the project. And Aisha Walker is my colleague in the School of Education who's involved in using chatbots for education. And I am a computer scientist.

The abstract is essentially that chatbots are useful in many applications. And ideally, we want to be able to use chatbots for one exact particular example. And it's the course evaluations. So, at the end of the course, for every module at Leeds University, you get sent a questionnaire and you have to fill in or for each of the questions answer on a scale of I think it's 1 to 5 whether you agree or disagree. The trouble with that is that you don't get much scope for really useful feedback. If you just know a scale of 1 to 5, OK you know overall people like each of these things or don't like these things.

Whereas what we'd like is to have some more narrative texts and real details of what you like and what you don't like. So, that's what the chatbot tries to do. And this has been used in four different universities and research centres in the UK, Leeds obviously, in Spain, in Croatia, and in Cyprus. The chatbots were developed by commercial companies in Norway and Sweden and then we use them in our universities. And we found that collecting coarse feedback this way gets better quality responses and potentially the students can think it's fun as well.

OK. Let's go down and have a look at some of the highlights of this. So, first of all, there's some research on getting feedback from students. If you just have a question and you have to respond it by a tick box, then there's two problems. One is that you don't get that much useful information out of it. Another problem is that some students will just get fed up with having to answer these things. If you get loads of questionnaires and each questionnaire's got 10 questions on it, then you may not respond at all. So, we end up getting very poor-quality responses compared to a face-to-face conversation or interview or even a telephone interview.

So, maybe what you could have...the trouble with telephone interviews is they're very time consuming and expensive to organise so why not have a chatbot do it instead? That was the overall idea. And that's the system called Hubert or Hubert.ai And ai is the top-level domain for AI research, apparently. And there are a number of other chatbots around. So, the first thing we do is have a look to see what other sorts of chatbots are there available for this sort of research.

So, very prominently, Amazon and Google and Microsoft have voice-driven chatbots. So, Amazon have Alexa, where you can speak rather than having to type things in. And are going to look at some of these examples in a minute. Amazon Web Services have a demonstrator called Ask LU. And LU stands for Lancaster University, not Leeds University, unfortunately. And it was developed where you can ask questions about your timetable or grades or when is my next lecture or things like that. It's not actually teaching but it's providing services.

And there are other chatbots which provide services. Many of these chatbots, or chatbot architectures, were originally developed for commercial services. So, for example, I have a NatWest bank account. And when I opened my online banking, there's the little Cora chatbot and I can ask questions like, I don't know, how do I close down my account? Or how do I make a complaint about NatWest? And sometimes it will give me a right answer.

And within universities, they realised that these sorts of architectures could be adapted to many student services. So, for example, Lola or Dena, there are systems for enrolment and admissions when you first join the University. Differ is a system that we're using on our research project, which helps students to, even before they come, to get to know each other. So, it helps with student engagement, getting students to feel that they're part of the University community.

But that's not really teaching. So, there are also some researchers developing systems for helping to teach and learn. For example, Coding Tutor. This helps students in introductory programming courses. It sort of analyses the software and has an English language feedback about what you did wrong. You're probably aware that here at least, we do have automated grading systems for Python programming so that you can do the Python course then some of the grading is semi-automated. But the grading, the results are then said to the tutor, who then gives you a mark, whereas the Coding Tutor chatbot in addition gives feedback to the students in English that you could have redone it this way or something like that.

And there are other teaching chatbots, like Clive chatbot and Book Buddy chatbot for helping certain teaching tasks. The problem with teaching is that this is something we discovered that real teachers like me aren't just answering your questions but are doing a bit more than that. So, building in the background knowledge into that sort of chatbot can be harder than building the background knowledge for a chatbot which tells you when your next lecture is. Many of these things, sort of services, are typically done by access to databases of some sort, which you can imagine doing, having a chatbot interface to a database, whereas teaching and learning requires more knowledge somehow.

OK. So, what we did for our project was we tried to find one use case, which was fairly straightforward, which would be useful in teaching and that is getting feedback from students about the course. What do you think of it so far? How could we improve it? What should we be dropping from it? So, we had partners in the project University of Granada in Spain, University of Zagreb in Croatia, and the Centre of Excellence Centre in Cyprus, Science Centre of Excellence, and Leeds

University in the UK. And we all taught courses and at the end of the courses asked the students to give us some feedback on the courses.

A lot of times, the students didn't reply at all, and we got...most of the courses, because the project is in chatbot, most of the courses are computing-oriented or business-oriented because that's the sort of area where chatbots are taught so it's not all University students but a selection of classes in computing and business sort of areas. And there's a sample of how the system went. Hang on. Let me just try to make this a bit bigger so you can see a bit more clearly. "Beep boop. Here we go. By the way, just say 'go back' and correct me if I misunderstand you. Hi how old are you? Please type your age as a number." And then the student has to reply. And "What is your gender? Please choose male or female." This time, those are the only options.

So, you can see what is working well with course X and should continue in the same way, where course X is replaced with what the actual name of the course is. And the student is supposed to type something, not just yes or no or a score but more about what is actually good or bad. And then what could the teachers start doing? What could improve it? And what could the teachers stop doing that could improve course X? So, the notice that it's asking the same questions to all the students but there will be some variety.

And also, this is just the overall framework. So, if a student doesn't give you an appropriate response, then the chatbot can go into a dialogue to tease out more responses. So, that's the big difference. It's not just presenting the questions in a box for you to type into, but it is actually going into a conversation. And we'll see some examples of it in a minute. So, we managed to get quite a few examples from a number of different courses. You can see here, mainly in computing type areas or in business type areas, like taxation and business informatics.

So, these conversations were of varied length. So, here's the average depth of conversation. How many interactions were there? So, we saw from the template, you're basically asking a small number of questions but sometimes the student didn't give an appropriate answer or just said, if I asked you what was wrong with the course, if a student just said, nothing, well, the chatbot would try to say something like, are you sure? Can you tell it think about it a bit more? So, in drawing out the conversations, we saw it went up to about 30 turns, which is quite a long conversation.

Here's the ratings that the students gave back at the end of the chatbot asked, what do you think of course overall on a score of 1 to 10? And the students tended to like the course so most of the scores are around 9 or 10. The chatbot did try to get out more detailed, for example, are you absolutely, definitive 100% sure about that? And a smiley face to make sure that everything was OK.

Here's an example of a somewhat longer conversation just to illustrate that. Hubert says, "What could the teachers start doing that would improve it?" User replies, "I think everything is all right. Nothing to improve." Hubert says "Surely there must be something that could be improved." User says, "I can't remember anything bad." "Not yet? Let me repeat the question. What could the teacher start doing that would improve it?" "I can't remember anything bad." "Not yet. What could the teachers start doing that would improve it?"

In the end, the user is sort of cajoled into saying something. Maybe weekly tasks could be equal because they heard that some students got harder tasks than others. So, there, Hubert's actually trying to not just simply accept that the student is very happy but try to find something. So, it goes on

a bit more. And sometimes by the end of it, the user is drawn out into a conversation and says quite a lot. So, you can see that this is quite detailed feedback. And this is useful for the lecturer because they know what sorts of things they should at least improve for next year.

Potentially, you could ask this feedback question at partway through the course. But typically, what happens is you get some feedback, and this goes into teaching next year. And of course, it goes back into, I have an annual academic review with my head of school and head of teaching and learning. And they get to find out something about what really works and what doesn't work for them in terms of not just this one module but all the modules I'm teaching.

And as well as one other thing we tried to do is well, is saying, what did you think of course and what did you not think of course? At the end, we asked, OK, how did you like this type of evaluation compared to a regular survey? That was, what did you think of Hubert? And the user may well say, very interesting and unique or things like that. And we also try to get a score for that. On a scale of 1 to 10, would you give it a 10? And then the user generally thinks it's fun or it's nice. So, you can see some of these, as well as asking about the course, we asked about Hubert. Because we want for the European Union, they want to know...they're funding the project...they want to know if it's very good.

Yeah, the depth of the conversation is the count of both the questions and responses. How many terms were taken? So, it does include Hubert's turns as well. So, let's say...I suppose you want to say how many things did the student say? More or less half that. But the conversation is both turns. This user said "This way is totally different with regular one. However, it is interesting." So, they liked...it's really interesting. Better than filling a form. So, the students like this and furthermore, they're more likely to interact with this. If they're given yet another survey questionnaire form to fill in, they might just ignore it or not answer it.

One unexpected side effect is that because we were asking computing students, some of them started to say actually, I just want to see if I can break the chatbot. They answered the questions, yeah, but they were also interested in trying out the chatbot and trying to see if it was really intelligent by asking difficult questions.

So, we have some examples of having answered the questions, Hubert says, "Thank you. Those are all the questions I had. You may close this window now. Have a great day". The user says "bitch" to rile the chatbot, see what happens. Hubert comes back and says, "A poor choice of words never improved anything." Hubert the bot, he knows that bitch is not a nice thing to say. So, the user says, "Tell me a joke." Hubert does tell them a joke. "Knock." "Who's there?" "Hey there. I don't have any questions for you right now. Please come back later." The user doesn't think that's a very good joke. "Sick joke, man." But at least he appreciates it is a joke.

Sometimes the user tries to ask Hubert question just to see what will happen. So, here we have Hubert says, "I don't have any questions right now. Please come back later." The user says, "coronavirus. What does coronavirus cause?" The user does know what coronavirus causes. They just want to see if Hubert knows. Hubert says, "I'll try to answer that. May be totally wrong. Give me a sec. SARS-CoV-2 virus." So, Hubert has got some sort of answer, even though it's not a very sensible answer.

We get longer examples of conversations. Here's a clever one. User says, "What is it?" Silly question. Hubert says, "I'll try to answer that, but I may be totally wrong. Give me a sec. What Is It is a 2005

American surrealist film written, edited, co-produced, and directed, blah, blah, blah.”? So, even a silly question like what is it? Hubert can come back with, what it's basically done is done a Google search of "what is it" to try to find some answer to that question.

So, there's examples of student responses that were challenging for Hubert. So, in general, it worked out nicely. Another nice feature is that students could change their minds and modify their answers. In conclusion, the chatbot was able to guide the students to much more constructive answers and many students volunteered positive perceptions of the use of Hubert. They really thought it was good using Hubert rather than just having a questionnaire. And students would engage with the chatbot and even give friendly responses to Hubert and Hubert was friendly in return.

Sometimes students try to break the system. They would try to give challenging responses and then Hubert tended to either just give a silly reply or treat it as a joke. Or Hubert could ask them to modify their questions to see if it worked. So, at least one of these responses, one of the interactions was very long and it was just the student keeping over and over again, trying to break the chatbot. OK. I think that's it. Thank you very much for listening. Goodbye.

[END]