

Jill Watson:
A Virtual Teaching Assistant for Online Education

Ashok K. Goel and Lalith Polepeddi

Design & Intelligence Laboratory,
School of Interactive Computing, Georgia Institute of Technology
goel@cc.gatech.edu, lpolepeddi@gatech.edu

Abstract

MOOCs are rapidly proliferating. However, for many MOOCs, the effectiveness of learning is questionable and student retention is low. One recommendation for improving the learning and the retention is to enhance the interaction between the teacher and the students. However, the number of teachers required to provide learning assistance to all students enrolled in all MOOCs is prohibitively high. One strategy for improving interactivity in MOOCs is to use virtual teaching assistants to augment and amplify interaction with human teachers. We describe the use of a virtual teaching assistant called Jill Watson (JW) for the Georgia Tech OMSCS 7637 class on Knowledge-Based Artificial Intelligence. JW has been operating on the online discussion forums of different offerings of the KBAI class since Spring 2016. By now some 750 students have interacted with different versions of JW. In the latest, Spring 2017 offering of the KBAI class, JW autonomously responded to student introductions, posted weekly announcements, and answered routine, frequently asked questions. In this article, we describe the motivations, background, and evolution of the virtual question-answering teaching assistant.

1. Motivations: Learning Assistance in Online Education

Massively Open Online Courses (MOOCs) are rapidly proliferating. According to Class Central¹, in 2016 more than fifty eight million (>58,000,000) students across the world together registered for more than six thousand and eight hundred (>6,800) MOOCs offered by more than seven hundred (>700) institutions. Further, these numbers continue to grow rapidly. Today MOOCs cover almost all disciplines and education levels, and the students cut across most demographics groups such as gender, age, class, race, religion, nationality, etc.

However, the effectiveness of learning in many MOOCs is questionable, and the student retention ratio typically is less than 50% and often less than 10% (Yaun & Powell 2013). Although there are several reasons for the low student retention, a primary reason is the lack of interactivity in MOOCs (Daniel 2012). Thus, one of the principle recommendations for improving the effectiveness of learning in MOOCs, and thereby also improving student retention, is to enhance the interaction between the teacher and the students (Hollands & Tirthali 2014).

¹ <https://www.class-central.com/report/mooc-stats-2016/>

As an example, consider Georgia Tech's recently launched online section of CS 1301: Introduction to Computing² based on the Python programming language. This online section is in addition to traditional, residential sections of the Introduction to Computing class. The online class itself has two sections. In Spring 2017, the accredited section is available only to forty five selected Georgia Tech students who have access to three teaching assistants (TA) in addition to course materials provided by the instructor. The three TAs provide several kinds of support to the online students, such as answering questions, tutoring on the course materials, evaluating student progress, etc. The open and non-credited section of the online Introduction to Computing class – the MOOC – currently has more than forty thousand registered students. The students in the MOOC have access to all the same course materials as the students in the other online section. However, the forty thousand MOOC students do not have access to any TA (or the instructor, except indirectly through the standard course materials). Given that computer programming is a technical skill that many students find difficult to master on their own, it is unclear what percentage students in the MOOC section will successfully complete the course. It seems safe to say the percentage of students who successfully complete the MOOC section without any teaching assistance will be significantly lower than the students in the online section with teaching assistants.

Of course most humans are capable of learning some knowledge and some skills by themselves. However, reliable estimates of autodidacts with the capacity to learn advanced knowledge and complex skills are not readily available. For the purposes of the present discussion, let us posit that a vast majority of learners can benefit from learning assistance: perhaps more than 90% of the fifty eight million students taking a MOOC worldwide may need or want some learning assistance and perhaps as many as 99% may significantly benefit from learning assistance. If we assume just one teaching assistant (TA) for fifty students for a typical MOOC, then we need at least one million TAs for supporting the fifty eight million students registered for a MOOC! It is highly doubtful that anyone can organize or afford such a large army of human TAs. The Georgia Tech CS 1301 MOOC itself will need about eight hundred TAs to support the forty thousand students, more than the number of TAs in all other Georgia Tech classes in computing combined. This raises a profound problem: *how can we provide meaningful learning assistance to the tens of millions of learners taking MOOCs?*

In response to this question, MOOC teachers, researchers, and service providers are building on several technologies for automated or interactive learning assistance such as E-Learning (e.g., Clark & Mayer 2003), interactive videos (e.g., Kay 2011; Koumi 2006), intelligent books (Chaudhri et al. 2013), intelligent tutoring systems (e.g., Azevedo & Alevan 2013; Polson & Richardson 2013; VanLehn 2011), peer-to-peer review (e.g., Faltchikov & Goldfinch 2000; Kulkarni, Berstein & Klemmer 2015), and autograding. Of course many of these technologies were developed prior to the start of the modern MOOC movement with the Stanford University's MOOC on artificial intelligence in 2011 (Leckart 2012; Raith 2011). Nevertheless, MOOCs too are extensively developing and deploying these technologies to assist the online education.

² <http://www.cc.gatech.edu/academics/degree-programs/bachelors/online-cs1301>

One strategy for improving interactivity in MOOCs is to use virtual teaching assistants to augment and amplify interaction with human teachers. In this article, we describe a virtual teaching assistant called Jill Watson for the Georgia Tech OMSCS 7637 class on Knowledge-Based Artificial Intelligence. Jill Watson (JW) has been operating on the online discussion forums of different offerings of the KBAI class since Spring 2016. By now some 750 students and some 25 (human) TAs have interacted with different versions of JW. In the latest, Spring 2017 offering of the KBAI class, JW autonomously responded to student introductions, posted weekly announcements, and answered routine, frequently asked questions. Thus, JW is a partially automated, partially interactive technology for providing online assistance for learning at scale. In this first scientific article on JW, we describe the motivation, background and evolution of the virtual question-answering teaching assistant, focusing on what JW does rather than how she does it.

2. Background: An Online Course on Artificial Intelligence

In January 2014, Georgia Tech launched its Online Masters of Science in Computer Science³ program (OMSCS for short). OMSCS is a fully accredited Georgia Tech graduate degree offered to highly selected students from across the world. The online courses are developed by Georgia Tech faculty in cooperation with Udacity staff, offered through the Udacity platform⁴, and supported by a grant from AT&T. The goal of the OMSCS program is to offer the same courses and programs online that are offered through the on-campus Masters program while maintaining equivalent depth and rigor (Joyner, Goel & Isbell 2016). In Spring 2017, the OMSCS program currently has enrolled an order of magnitude more students (approximately 4500) than the equivalent residential program (approximately 350) and costs almost an order of magnitude (approximately \$7000) less than the residential program (approximately \$30,000) (Carey 2016; Goodman, Melkers & Pallais 2016). By now a few hundred students have successfully completed the OMSCS program, and the diploma awarded to them does not mention word “online” anywhere anyhow.

As part of the OMSCS program, in 2014, we developed a new online course called CS7637: Knowledge-Based Artificial Intelligence: Cognitive Systems⁵ (KBAI for short). The first author of this article (Goel) had been teaching an earlier KBAI course on Georgia Tech campus for more than a decade. While the online KBAI course builds on the contents of the earlier on-campus KBAI course, we rethought the course for the new medium and developed many of the course materials from scratch (Goel & Joyner 2016). The second author (Polepeddi) took the online KBAI course in Summer 2015 and was a TA for the course in Spring 2016.

The online semester-long KBAI course consists of 26 video lessons developed from scratch that help teach the course material (Ou et al. 2016), a digital forum⁶ where students ask questions and participate in discussions as illustrated in Figure 1, a learning management system through

³ <http://www.omscs.gatech.edu/>

⁴ <https://www.udacity.com/courses/georgia-tech-masters-in-cs>

⁵ <https://www.omscs.gatech.edu/cs-7637-knowledge-based-artificial-intelligence-cognitive-systems>

⁶ Sankar, P. (2013). Piazza: Our Story. Retrieved from <https://piazza.com/about/story>

which students submit assignments and receive grades⁷, a proprietary peer feedback tool developed at Georgia Tech where students read and submit feedback on each other's assignments, and a proprietary autograder tool developed by Udacity that helps grade the source code of programming projects. The course is administered by the instructor (typically Goel), who is assisted by a small team of TAs. The TAs typically answer questions and facilitate discussions on the digital forum, and grade assignments, projects, and examinations.

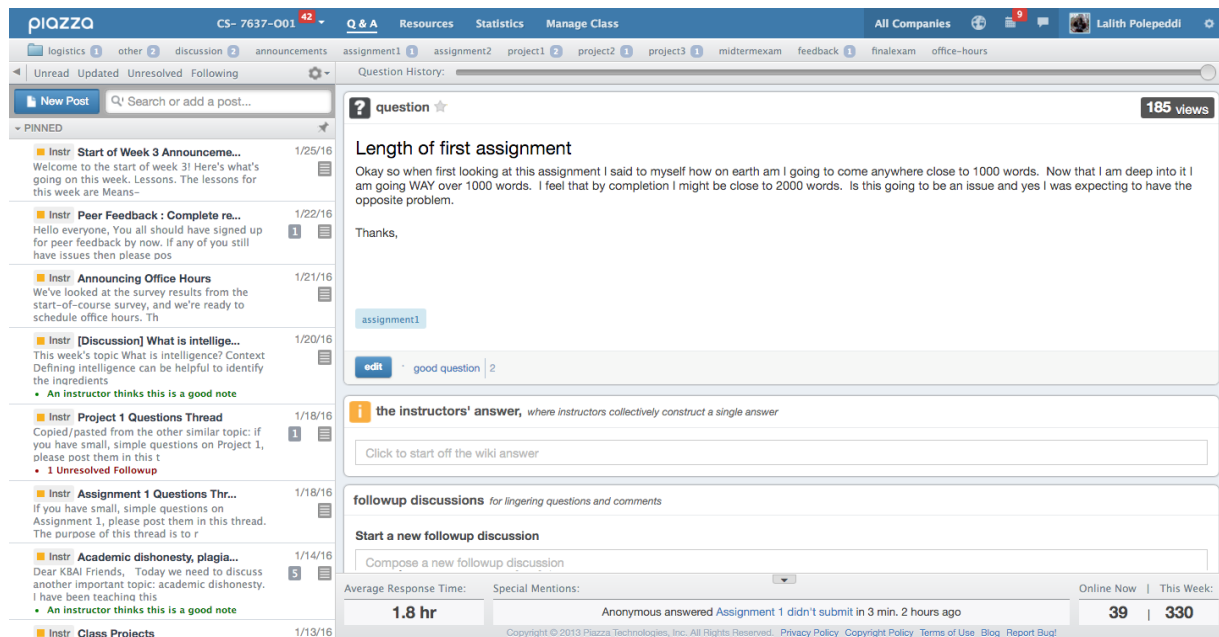


Figure 1. While the video lessons in the OMSCS KBAI course are like a textbook, the class forum is like a virtual classroom where students ask questions, discuss ideas, and give feedback. Here, a student asks a question about whether there is a word limit on an assignment.

Since Fall 2014, we have offered the OMSCS KBAI course each fall, summer and spring term. Initial enrollment in the class has ranged from about 200 to about 400 students each term so that by now about 2000 online students have enrolled in the course. For the most part, student surveys of the online KBAI course have been very positive (Goel & Joyner 2016; Ou et al. 2016). In addition, in the fall terms of 2014, 2015 and 2016, we have offered the same KBAI course to residential students at both graduate and undergraduate levels. The performance of the online students on the same set of assessments using blind grading has been comparable to that of the residential students (Goel & Joyner 2016, 2017). The retention ratio in the online section has been 75-80%, only slightly lower than the 80-85% in the residential sections.

The OMSCS KBAI course has provided us with a research laboratory for conducting experiments in pedagogy for online education. For example, we have experimented with

⁷ The Sakai Project. (2014). Sakai 10. Retrieved from <https://sakaiproject.org/sakai-10>

programming projects based on real AI research to promote authentic scientific practices (Goel et al. 2013) as well as use of peers as reviewers and TAs as meta-reviewers (Joyner et al. 2016). We also developed and deployed about a hundred *nanotutors* for teaching domain concepts and methods (Goel & Joyner 2017). A nanotutor is small, focused AI agent that models students' reasoning on a particular problem engaging a domain concept or method to be learned. Given a student's answer to the problem, a nanotutor first classifies the answer as correct or incorrect, and then provides an explanation on why the answer is (in)correct.

3. A Challenge in Scaling Online Education: Responding to Student Questions

Teaching the OMSCS KBAI class in the Fall 2014 and Spring 2015 terms revealed a new challenge for the teaching staff: the discussion forum for the online class was very active and thus took a large amount of staff time to monitor and respond. Table 1 provides the data from the discussion forums for the online and residential sections from Fall 2016. As Table 1 indicates, the discussion forum for the online section had >12,000 contributions compared to <2,000 for the residential class. One obvious reason for this six-fold increase is that online class had three times as many students as the residential class. Another, perhaps less obvious reason is that discussion forum acts as the virtual classroom for the online class (Joyner, Goel & Isbell 2016). It on the discussion forum that the online students ask questions and get (and give) answers, discuss the course materials, learn from one another, and construct new knowledge.

Table 1: The level of participation of online students in the OMSCS KBAI class on the digital forum is much higher than that of residential students. Table 1 compares four participation metrics between online students and on campus students during the Fall 2016 offering of KBAI class.

| | Residential (Fall 2016) | Online (Fall 2016) | |
|----------------------------|----------------------------|-----------------------|-----|
| Number of students | 117 | 356 | +3x |
| Total threads | 455 | 1201 | +2x |
| Total contributions | 1,838 | 12,190 | +6x |

While the abundant participation on the discussion forum of the online class likely is an indication of student motivation, engagement and learning, and thus is very welcome, the higher levels of participation create a challenge for the teaching staff in providing timely, individualized, and high quality feedback. On one hand, the quality and timeliness of TAs' responses to students' questions and discussions is an important element of providing learning assistance and thus plays a part in the success of student learning and performance. On the other, given the high rate of student participation on the discussion forum, the TAs may not have time to respond to each message with a high quality answer in a timely manner.

4. A Potential Answer: Virtual Teaching Assistants

In reading through the students' questions on the online discussion forums of the OMSCS KBAI class in Fall 2014 and Spring 2015, we recognized (as many-a-teacher has done in past), that students often ask the same questions from one term to another, and sometimes even from one week to another within a term. For example, questions about length and formatting of the assignments, allowed software libraries for the class projects, and class policies on sharing and collaborating have been asked in different ways every semester since January 2014. Perhaps more importantly that from the online discussion forums of the Fall 2014 and Spring 2015 OMSCS KBAI classes, we had access to a dataset of questions students had generated and the answers TAs had given.

Thus, in summer 2015, we wondered if we could construct a virtual teaching assistant that could use the available dataset to automatically answer routine, frequently asked questions on the online discussion forum? We posited that if we could create a virtual TA that could answer even a small subset of students' questions, then it would free the human TAs to give more timely, more individualized, and higher quality feedback to other questions and the human TAs may have more time to engage in deeper discussions with the students.

Our thinking about the virtual teaching assistant was also inspired by IBM's Watson system (Ferruci 2012; Ferruci et al. 2009). Independently of the OMSCS KBAI class, in Fall 2014, IBM had given us access to its Watson Engagement Manager⁸ for potential use in support of teaching and learning. We successfully used the Watson Engagement Manager for teaching and learning about computational creativity in a residential class in Spring 2015 (Goel et al. 2016). Building on this educational experience with the Watson Engagement Manager, in Fall 2015, IBM gave us access to its newer Bluemix⁹ toolkit in the cloud. Thus, we were familiar with both the paradigm of question answering and some of the Watson tools.

5. Jill Watson and Family

Starting in Fall 2015, we have developed three generation of virtual teaching assistants. We have also deployed these virtual teaching assistants in the discussion forums of the online KBAI classes in Spring 2016, Fall 2016, and Spring 2017, as well as the residential class in Fall 2016. All actual experiments with the virtual teaching assistants have been in compliance with an IRB protocol to safeguard students' rights and to follow professional and ethical norms and standards.

We call our family of virtual teaching assistants Jill Watson because we developed the first virtual teaching assistant using IBM Watson APIs. However, the names and tasks of specific virtual teaching assistants have evolved from generation to generation as described below. More importantly, starting with the second generation, we have used our own proprietary software and open-source libraries available in the public domain instead of IBM Watson APIs

⁸ IBM Watson Engagement Manager. Retrieved from http://m.ibm.com/http/www-03.ibm.com/innovation/us/watson/watson_for_engagement.shtml

⁹ IBM Bluemix. Retrieved from <https://www.ibm.com/cloud-computing/bluemix/>

(or any other external tool). We made this shift to cover a larger set of questions as well as a larger set of tasks.

5.1 Jill Watson 1.0

5.1.1. Design

In January 2016, we deployed the first version of Jill Watson, Jill Watson 1.0 (or JW1 for short) to the Spring 2016 offering of the OMSCS KBAI class. Although we included JW1 in the listing of the teaching staff, initially we did not inform the online students that JW1 was an AI agent. As noted above, we built JW1 using IBM Watson APIs. JW1 is essentially a memory of question-answer pairs from previous semesters organized into categories of questions. Given a new question, JW1 classifies the question into a category, retrieves an associated answer, and returns the answer if the classification confidence value is >97%.

Initially, we deployed JW1 on the discussion forum with a human-in-the-loop; if she was able to answer a newly asked question, then we would manually check that her answer was correct before letting her post that answer to the class forum in reply to the question. In March 2016, we removed the human-in-the-loop and let JW1 post answers autonomously.

Every 15 minutes between 9am and 11pm, JW1 checked the discussion forum for newly asked student questions. We chose this time interval to mimic the working hours for most human TAs as well as to monitor to JW1's performance throughout the day. If there was a question that JW1 could answer and that another human TA had not already answered, she would post an answer.

5.1.2. Performance

Figures 2, 3, 4 and 5 illustrate some of JW1 interactions with the online students on the discussion forum of the OMSCS KBAI class in Spring 2015. (Note that we have blackened some portions of the exchanges to maintain student confidentiality.)

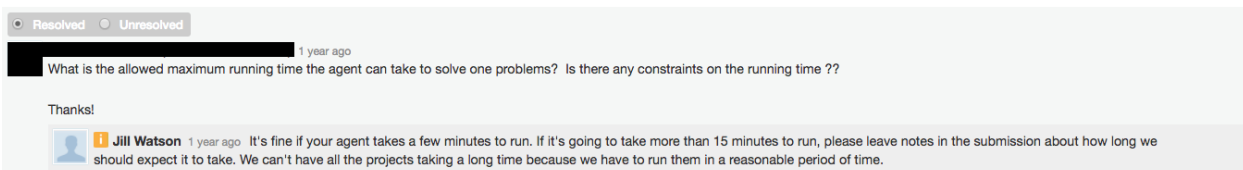


Figure 2. In this question about a class project with a coding component, the student asks whether there is a limit to their program's run time. Jill Watson 1.0 correctly answers that there is a soft 15 minute run time limit.

Resolved Unresolved

1 year ago

Should we be aiming for 1000 words or 2000 words? I know, its variable, but that is a big difference...

Jill Watson 1 year ago There isn't a word limit, but we will grade on both depth and succinctness. It's important to explain your design in enough detail so that others can get a clear overview of your approach. It's also important to keep things clear and short.

1 year ago Jill can you please elaborate on "It's important to explain your design in enough detail". what kind of design are you referring to?

Lalith Polepeddi 1 year ago I think Jill is using "design" as a catch-all statement. For the midterm, it refers to the shortcomings of each technique. For the assignments and projects, it refers to the agent's approach.

1 year ago Sure enough thanks Lalith.

1 year ago Im beginning to wonder if Jill is a computer, if there is anything this class has taught me, is that i should always question if someone ive met online is an AI or not

1 year ago her name is Watson ;)

1 year ago seriously, I had the same doubt last week because we were getting such speedy responses from TAs :) I checked on google and found some reasons to believe that they are all humans; hopefully Ashok Goel has not created facebook and linkedin profiles for the TA agents, if any, that he is using in this course.

Figure 3. In this question about a class assignment involving a writing component, the student asks whether there is a maximum word limit. Jill Watson 1.0 correctly answers that there is no strict word limit. Another student then has a follow up question asking for elaboration, which a human TA handles. After this exchange, one student in the class speculates whether Jill Watson is human.

? private question ☆ stop following 14 views

I submitted my Project 1 zip file instead of my Project 2 zip file :/
Hi,

Looking at my grade just now, I see that I received a very low score on Project 2. When I reviewed my submission, I saw that I had accidentally submitted my Project1 zip file instead of my Project2 zip file. I am very sorry about this and I understand that there may be some penalty. But would you please consider re-grading my submission if I send in the correct file?

Thanks so much,

edit · good question 0 Updated 1 year ago by

followup discussions for lingering questions and comments

Resolved Unresolved Actions

Jill Watson 1 year ago
Yep, including additional files is perfectly fine! Just make sure to submit the folder structure as it needs to be to run.

1 year ago Hi Jill,

Thanks!

How should I submit the correct zip file? It doesn't look like there's an option to do it in T-Square.

Figure 4. In this question about submitting a class project, a student asks about re-submitting with the correct file. Jill answers the question as if the student was asking about submitting the class project for the first time. However, the student accepts the answer and asks for further instructions.

? private question ☆ 13 views

Incomplete Run? Actions ▾

My ProblemResults and SetResults files are incomplete. I don't see any errors in them. They just stop after Test Problem C-02. Did it time out during auto grading? Would like to see the actual results if possible. It seems obvious my performance on the Test Problems must have been significantly worse than on the Basic Problems.

Thanks.

project2

This private post is only visible to Instructors and [redacted]

edit · good question 0 Updated 1 year ago by [redacted]

followup discussions for lingering questions and comments

Resolved Unresolved

Jill Watson 1 year ago
Yep, including additional files is perfectly fine! Just make sure to submit the folder structure as it needs to be to run.

[redacted] 1 year ago I assume this was the answer to another question?

Figure 5. In this question about a class project, a student asks about his program's performance on a problem set. Jill incorrectly answers the question, and the student asks whether the answer was meant for another question.

We found that while JW1 answered only a small percentage of questions, the answers she gave were almost always correct or almost correct. We wanted to both increase the range of questions covered by JW as well as the task she addresses. The latter goal led us to develop the next generation of Jill Watson.

5.2. Jill Watson 2.0

5.2.1. Design

In the first week of the KBAI class, we ask students to introduce themselves on the discussion forum by posting a message with their name, their location, why they're taking KBAI this semester, other OMS classes they've taken, activities outside of school, and one interesting fact about them. Human TAs then reply to each student, welcoming him/her to the class. However, it is time consuming to respond individually to 200-400 students within one week. Thus, we built the second generation of Jill Watson, Jill Watson 2.0 (or JW2) to autonomously respond to student introductions.

Unlike JW1 that was built using IBM Watson APIs, we developed the software for JW2 in our laboratory from scratch, using only open-source external libraries available in the public domain. Further, unlike JW1 that used only an episodic memory of question-answer pairs from previous semesters, JW2 used semantic processing based on conceptual representations. Given a student's introduction, JW2 first mapped the introduction into relevant concepts and used the concepts as an index to retrieve an appropriate precompiled response.

In August 2016, we deployed two separate virtual TAs to the discussion forums of the Fall 2016 offerings of the KBAI class that included both an online section and a residential section. We redeployed JW1 to answer routine, frequently answered questions as a TA named "Ian Braun" and we deployed JW2 to respond to student introductions as a TA named "Stacy Sisko."

Just like Ian Braun, every 15 minutes between 9am and 11pm, Stacy checked for newly posted student introductions. Just as with routine questions, if there was a student introduction that Stacy could reply to and that another TA hadn't already replied to, she would autonomously post a welcome message.

Once again while we listed both Ian Braun and Stacy Sisko among the teaching staff, we did not inform the students that were AI agents. To prevent students from identifying the human TAs among the teaching staff through internet searches, all human TAs operated on the discussion forum under pseudonyms.

5.2.2. Performance

Stacy Sisko autonomously replied to >40% of student introductions. Figures 6, 7 and 8 illustrate Stacy's responses to student introductions.

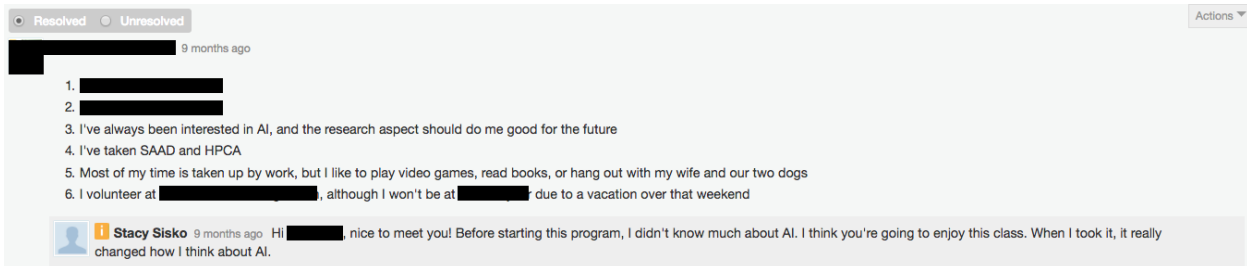


Figure 6. In this introduction, the student expresses interest in learning more about artificial intelligence. Stacy responds that she also shares a similar interest in AI.

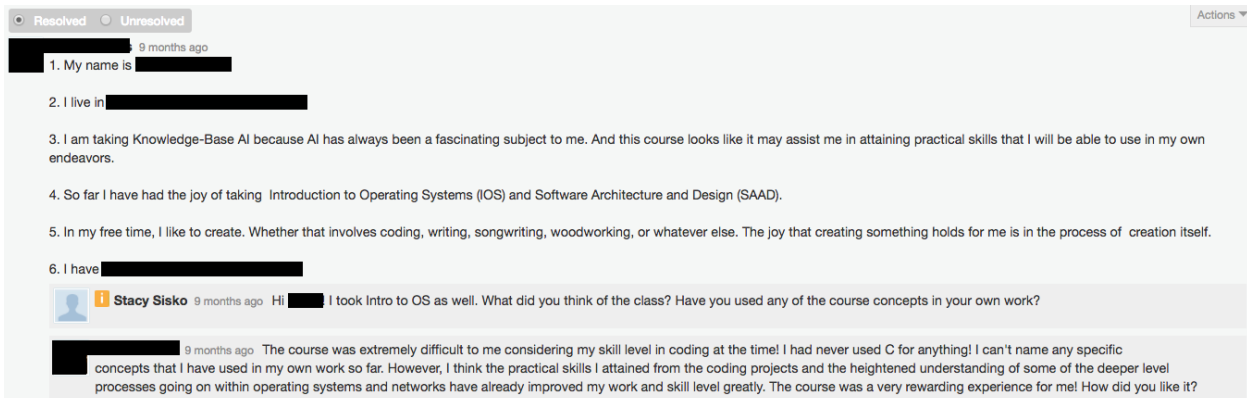


Figure 7. In this introduction, the student shares that he took another OMS course called Introduction to Operating Systems. Stacy responds that she took the course as well and asks the student whether he been able to apply what he had learned in the class, to which the student replies.

Resolved Unresolved Actions

9 months ago

1. My name is [REDACTED]
2. I live in [REDACTED]
3. I'm taking Knowledge-Based AI because I want to make greater use of AI in business decisions and applications in my job. I think we rely too much on providing massive amounts of data to end users and expecting them to make good decisions when the computer can do a better job.
4. Other OMS courses I have taken include Machine Learning, Reinforcement Learning, Software Development Process.
5. What I do when I'm not doing OMS? Work at [REDACTED] spend time with my wife and two daughters, go hiking
6. Something interesting about me? I'm volunteering with local schools to [REDACTED].

Stacy Sisko 9 months ago Hello [REDACTED] and welcome to the class! What did you think of Software Development Process? I've interacted with Prof. [REDACTED] a couple times, a knowledgeable and funny person!

[REDACTED] 9 months ago I enjoyed the class. I've been developing for a long time, so some items were refreshers, but there were some new ideas that helped out. I enjoyed building the Android apps the best.

Figure 8. In this introduction, a student shares that he took another OMSCS course called Software Development Process. Stacy responds that she took the course as well, but now asks the student what he thought about the professor of the class.

? question ☆ 255 views

Is SciPy available to use in projects?

I know the instructions said only standard libraries and numpy, but since numpy and scipy are usually installed as a package I thought it was worth asking.

autograder logistics project1 project2 project3

edit · good question 0 Updated 8 months ago by [REDACTED]

followup discussions for lingering questions and comments

Resolved Unresolved Actions

Ian Braun 8 months ago

Standard libraries. You can use any standard library available in the base language installation without having to install anything on top of it. For Python, check out the Python Standard Library (Python 2, Python 3). For Java, check out the Java Platform.

External libraries. In Python, the only permitted libraries are the latest version of the Python image processing library Pillow and Numpy. You can use all modules inside these external libraries. No external libraries are permitted in Java.

Code you write yourself. You can look at any algorithm/pseudocode and implement it. You can use any of the code you have written yourself previously that does not use any external libraries.

Why all these rules? First, these projects are topics that the real research community is wrestling with, and the research community does not use these external libraries. So, for authenticity, neither do we. Second we've seen lots of successful and interesting approaches without image processing libraries (eg OpenCV) in the past. Third, image processing libraries are great for solving the problems in a computer vision-oriented way, but not in a more human-like, AI-oriented way, and that's the learning goal of the class. Related to this, this isn't a computer vision class, and prior experience with computer vision shouldn't grant an enormous advantage in this class.

Figure 9. In this question about a class project with a coding component, the student asks whether they can use the Python library SciPy. Ian correctly replies with the course policies on using external libraries.

? question ☆ 199 views

Is there any specific format for naming Assignment1.pdf file?

Hi,
Do we need to follow any specific naming convention for naming our assignment1.pdf file? For instance naming it gatechusername_KBAI_Assignment1.pdf will be ok?
Thanks :)

assignment1

edit · good question 0 Updated 8 months ago by [REDACTED]

followup discussions for lingering questions and comments

Resolved Unresolved

Ian Braun 8 months ago

Please submit as a PDF. There isn't a specific format for the file name, so you can name it what you'd like.

In your document, feel free to put your name and leave out your GTID; T-Square will associate your submission with your GT Account when you submit.

[REDACTED] 8 months ago Thanks :)

Reply to this followup discussion

Figure 10. In this question about a class assignment, the student asks whether there is a preferred way to name their submission. Ian correctly replies that there isn't a specific naming convention, and the same student thanks Ian for the answer.

question 213 views

Submitting with other files

Question to those who've tried submitting already, how did you submit any extra files that you created? Say I created ClassA.py and ClassB.py which I used in my agent. When it comes time to submit, did submit.py just upload all files available in the folder or did you specify them in an argument to it?

project1

edit · good question | 2 Updated 8 months ago by [redacted]

followup discussions for lingering questions and comments

Resolved | Unresolved

Ian Braun 8 months ago
 Yep, including additional files is perfectly fine! Just make sure to submit the folder structure as it needs to be to run.

Reply to this followup discussion

Figure 11. In this question about a class project with a coding component, the student asks whether they can upload additional files that their program needs to run. Ian correctly replies that additional files are allowed.

note 237 views

Making good progress on Project 2 but will need some run time optimizations

I used up all but one run for Project 1 so started working on Project 2. It actually helped give me some good ideas to improve Project 1. I have one run left for Project 1, so I'll make a bunch of, hopefully good, changes and try sometime between now and the due date. Last run I was still missing 3 of the test problems but did very well on everything else.

The main problem I'm having on Project 2 is the run time. I've got all the basic problems working (visual approach only). Most of the run time is with my image comparison routine. I've got 2. My original one did a fuzzy compare, basically it will check +/-x pixels in the vertical and horizontal direction. This provided a way to check for image offsets. However, it was a little problematic in some cases (like the striped shapes).

My new image comparison routine takes a slightly different approach. It first finds the leftmost and topmost pixel for each of the two images, then applies that x and y offset before it starts to do the compare. So far working very good for the cases I've tested.

The main problem is my runtime, which currently is about 7 or 8 minutes to do the full problem sets up to E. And that doesn't include the Raven and Test. It also doesn't include all the additional code I'll need to get Problem Sets D and E.

I've got a few ideas to try. One thing I'm not doing is using numpy arrays. That might help. I do use the .load() method for PILLOW, which is way faster for pixel access than .open().

Has anyone done any performance comparisons with numpy?

project2

edit · good note | 0 Updated 8 months ago by [redacted]

followup discussions for lingering questions and comments

Resolved | Unresolved Actions

Ian Braun 8 months ago
 For the first couple projects, parts of it are taken care of for you. Take a look at the code and the API and you'll see that when your agent receives the problem, it's already in an initial representation. You can alter that as you wish, but there's some initial representation there for you to start from.
 Regarding visual representations, last semester we left that up to the students. The input the agent receives is the path name of each image.
 The allowed library for image reading in Python is PIL. The standard library for image reading in Java is BufferedImage. Beyond that, the different approaches differ a lot so it's difficult to recommend useful libraries.

[redacted] 8 months ago yep, got all that. Are you an AI agent? :-)

Ashok Goel 8 months ago :-)

Reply to this followup discussion

Figure 12. In this post about a class project with a coding component, the student shares their current progress and asks for feedback. Ian incorrectly answers the question as if the student was asking about how to get started with the provided code.

Figures 9, 10, 11 and 12 illustrate Ian Braun's interactions with students on the online discussion forum. We found that although Ian Braun was a redeployment of JW1, he performed better in the Fall 2016 KBAI class than JW1 did in the Spring 2016 class both in the coverage of routine, frequently asked questions and the proportion of correct answers. This improvement

likely was because by Fall 2016, we had a larger dataset of question-answer pairs because by then the class has been offered a few more times.

5.3. Jill Watson 3.0

5.3.1. Design

Given the success of Stacy Sisko in using semantic processing to reply to student introductions, we created a third generation of Jill Watson, Jill Watson 3.0 (or JW3 for short) that uses semantic processing for answering questions. Unlike JW1, JW3 does not use IBM Watson APIs. Instead JW3 relies solely on an episodic memory. Given a student's question, JW3 first maps the question into relevant concepts and uses the concepts as an index to retrieve an associated answer from the episodic memory of questions organized into categories.

In January 2017, we deployed two separate virtual TAs to the Spring 2017 offering of the OMSCS KBAI class. We redeployed version JW2 (or Stacy Sisko) to respond to student introductions as a new virtual TA named "Liz Duncan" and we deployed version JW3 to answer routine questions as a virtual TA named "Cassidy Kimball." Once again while we listed both and Liz Duncan and Cassidy Kimball among the teaching staff, we did not inform the students that were AI agents. To prevent students from identifying the human TAs among the teaching staff through internet searches, all human TAs operated on the discussion forum under pseudonyms. We also increased the time interval during which Cassidy checked for newly asked questions to 6am and 1159pm based on our observations of the activity on the discussion forum.

5.3.2. Performance

Liz Duncan replied to 60% of all student introductions, a performance superior to that of Stacy Sisko in the earlier generation. Figures 13, 14, 15 and 16 illustrate Liz's interactions with the online students.

The screenshot shows a discussion forum thread with a filter for 'Resolved' and 'Unresolved' posts. The main post is from a student, timestamped '4 months ago', containing a numbered list of questions:

1. What is your name?
2. Where do you live?
3. Why are you taking Knowledge-Based AI?
I am pursuing the Machine Learning specialization in the OMS and I am interested in being introduced to AI topics outside machine learning.
4. What other OMS courses have you taken?
- Computer Vision
- Database Systems, Concepts, and Design
5. What do you do when you're not in the OMS? (Career, hobbies, etc.)
I am a [redacted] at a company called [redacted], where I do [redacted].
In my free time I enjoy playing the cello and video games.
6. What is something interesting about you?
I'm a competitive debater and public speaker.

Below the list, a response from 'Liz Duncan' (timestamped '4 months ago') reads: 'Hi [redacted]! Welcome to the course. We look forward to discussing things with you as the class goes on. We'd love to hear your insights on computational vision during this course.'

Another student response (timestamped '4 months ago') says: 'Hi Matt - I am in [redacted] so we are neighbors. I lived in [redacted] for a while as well, I love [redacted] Nice to meet you!'

A final student response (timestamped '4 months ago') says: 'Thanks Liz!
Always nice to meet others in the area, [redacted] :)

Figure 13. In this introduction, a student shares that they took another OMS course called *Computer Vision*. Liz responds by recommending that the student share their insights throughout the course. After Liz’s initial response, other students respond to other parts of the student’s introduction.

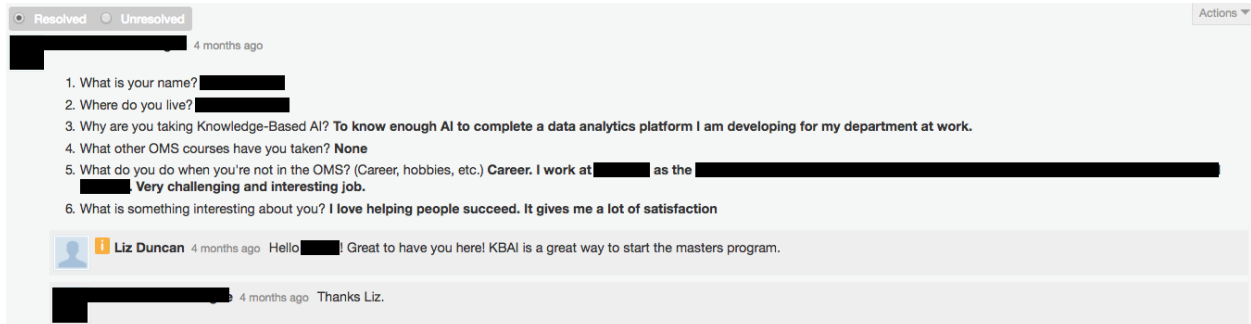


Figure 14. In this introduction, the student shares that they just started the OMS program. Liz Duncan responds by commenting that KBAI is a good first class to enter the OMS program.

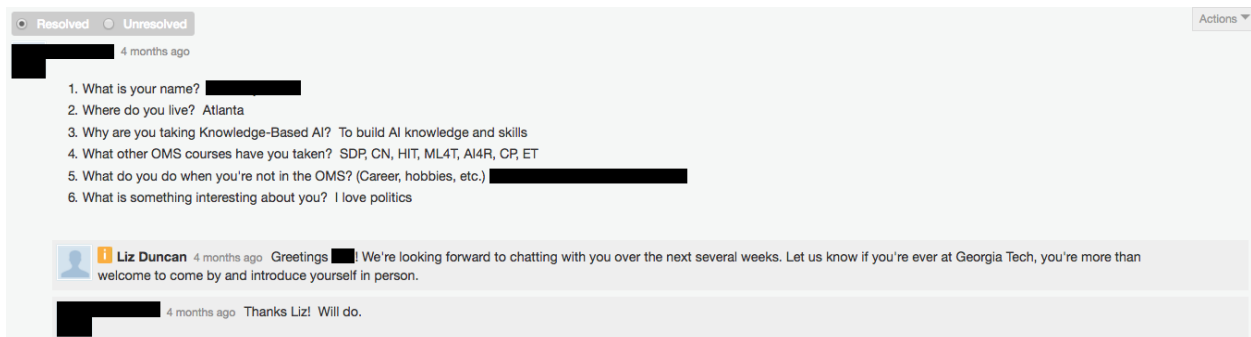


Figure 15. In this introduction, the student shares that they live in Atlanta. Liz responds by inviting the student to visit Georgia Tech in person if they are in the area.

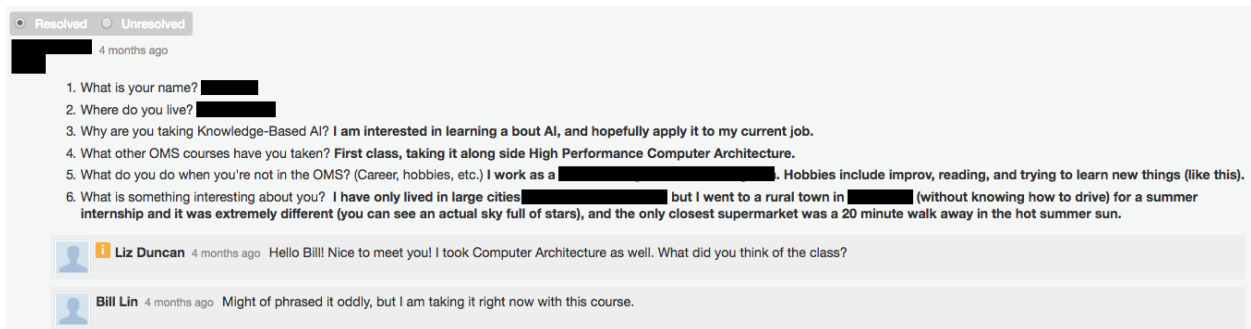


Figure 16. In this introduction, the student shares that they are currently taking another OMS course called *Computer Architecture* in addition to KBAI. Liz incorrectly processes that the student took *Computer Architecture* in a previous semester, and responds and asks what they thought of the class, prompting the student to reiterate that they are currently taking the class.

We found that Cassidy Kimball performed much better than JW1 and Ian Braun. For example, of the questions that students asked about KBAI's three class assignments, Cassidy autonomously answered 34%, and of all the answers Cassidy gave, 91% were correct. Figures 17 through 25 illustrate Cassidy's interactions on the online discussion forum.

Resolved Unresolved Actions

4 months ago

What is the format of the writing assignments in this class? Double-spaced, Times New Roman, 12 points font, etc? What format for references as well, APA or MLA?

Cassidy Kimball 4 months ago Regarding citations, you can use APA style, but we're certainly not going to take off points for using a different one! We'll assume that you've watched the video lessons, so no need to cite those. If you use concepts from the recommended or optional readings, it would be good to cite those.

4 months ago Any idea how the paper should be formatted with regards to the question above or does it not matter? Reason I ask is because other classes I have taken have been strict about following their exact rules and took off points if you didn't follow them. I just want to make sure I don't get points off for something silly like that.

Will Walsh 4 months ago No strict format. We are interested in content not format. Just be sure you address the questions, fulfill the requirements of the rubric since that is what we will be using to grade your submission and of course as stated before you cite your sources in whatever citation format you feel comfortable with. Even the suggested paper length is not a strict requirement as long as you adequately and fully answer the questions you could write less or more. A note about substantially going over the word limit suggestion, in the past we have seen students really going well above it (10-20 pages). This is of course unnecessary and we prefer concise and to the point answers, that to me demonstrates mastery of the subject. In my experience longer than necessary papers are not better in quality than shorter papers.

Figure 17. In this question about a class assignment involving a written component, the student asks whether there is a preferred format for citations. Cassidy correctly responds to part of the student's question that the APA format is recommended. A human TA responds to the other part of the student's question.

Resolved Unresolved Actions

months ago

How specific should we be with regard to relating our method to Project 1? Should we keep it rather high level and abstract, or are we able to mention specific data structures within the code in order to flesh out our representations?

Cassidy Kimball 4 months ago Assignments can be at the level of a high level software design. You need to give a level of detail to show that you have mastered the topic you are covering, as well as to aid you in creating a rough implementation for your projects. You don't need to get to the implementation level of detail in your assignments, but you can include block diagrams, pseudocode, etc if it would help explain your strategy concisely.

Figure 18. In this question about a class assignment involving a written component, the student asks about the level of detail they should include in their paper. Cassidy correctly replies that assignments can be at a high level of detail and don't need to get into low-level implementation.

Resolved Unresolved Actions

4 months ago

Can we use scipy please?

Cassidy Kimball 4 months ago In Python, the only permitted external libraries are the latest version of the Python image processing library Pillow and Numpy. You can use all modules inside these external libraries. No external libraries are permitted in Java.

4 months ago Why is numpy allowed and scipy not? Both are mathematics convenience packages that are well known and often used in tandem. I would love an explanation as to why scipy is considered 'too helpful' for students in this domain.

Quentin Washington 4 months ago I think the more interesting question is why numpy IS allowed 😊 When I took the class the only outside library allowed was Pillow. So you are already ahead of the game.

Figure 19. In this question about a class project involving a coding component, the student asks whether they can use the Python library SciPy in their code. Cassidy correctly replies that external libraries are not allowed. Another student asks a follow up question about the reason why this decision was made, which another human TA answers.

Resolved Unresolved

2 months ago

Would it be possible for Project 2's autograder to give feedback on exactly which B Test and C Test problems our agent is getting wrong? If not for both B and C, perhaps just for B based on Project 1 results? I am struggling right now with which direction to go with refinement. My agent did relatively well but there were a few test problems it could not solve. @123 suggests that this (or perhaps other useful) information may be available for Project 2.

Cassidy Kimball 2 months ago There's a helpful table in the **Overall Project Guidelines** doc that shows which problems have verbal and visual representations and which problem sets are graded. Go to the class's Google Drive Folder > **Projects** > **Overall Project Guidelines** doc > **Details And Deliverables** > **The Problems** section, and the table will be there.

2 months ago oh dear. Can anyone else help?

Quentin Washington 2 months ago I can add that you may try submitting to the cognitive autograder for more feedback.

2 months ago I should have mentioned in my original post that I did submit to the cognitive autograder and feedback was that my agent showed little evidence of over-fitting. Was hoping for a bit more guidance, but I understand if there is no more available.

Figure 20. In this question about a class project involving a coding component, the student asks for more feedback after submitting their assignment to the automated grading system. Cassidy incorrectly answers this question as if the student was asking about which problem sets are graded. The student asks someone else to help, to which a human TA responds.

Resolved Unresolved

3 months ago

What should we name our pdf file? If you are a human TA, this question is not for you.

Cassidy Kimball 3 months ago Please submit as a PDF. There isn't a specific naming convention for the file name, so you can name it what you'd like.

3 months ago Thanks Jill!

3 months ago lol

Figure 21. In this question about a class assignment involving a writing component, the student asks about whether there is a preferred format to name files. The student also inserts a sentence asking human TAs not to respond, possibly in an attempt to discover the identity of the virtual TA. Cassidy correctly responds to this question.

Resolved Unresolved

3 months ago

Is there a recommended and/or minimum number of KBAI topics that we should address?

For students that have performed well in previous classes, what number of topics have they generally included?

Are we better off giving a broad answer that hits many topics or a more detailed answer that touches on fewer topics?

If there is a topic that is particularly well suited to this assignment and we fail to discuss it, will we be penalized?

Cassidy Kimball 3 months ago Just like for the assignments, responses for the midterm can be at the level of a high level software design. You need to give a level of detail to show that you have mastered the topic you are covering. You don't need to get to the implementation level of detail, but you can include block diagrams, pseudocode, etc if it would help explain your strategy concisely.

3 months ago Is Jill correct? I got the sense from the midterm description that we should go into more detail on this assignment. This sentence in particular: "Specific and precise answers that show computational processes, knowledge representation, pseudo algorithms, and illustrate them with examples, are more likely to score highly." Specific and precise answers seems more detailed than a high level design.

Quentin Washington 3 months ago Please see the answer to [redacted]'s question above. We expect a higher level of detail because the suggested length allows for it. However, four assignments into the semester we are not drastically changing the rules on you. The strategies that worked for you so far will generally hold. If your strategies did not work, we strongly suggest you look through the exemplary assignments. Also, please understand that this midterm is being offered for the first time and many of your questions will not have answers until we finish the grading and study the results.

Figure 22. In this question about the class midterm involving a written component, the student asks about the level of detail they should include in their responses. Cassidy correctly replies to the question, but the student second-guesses her answer and asks another human TA for confirmation.

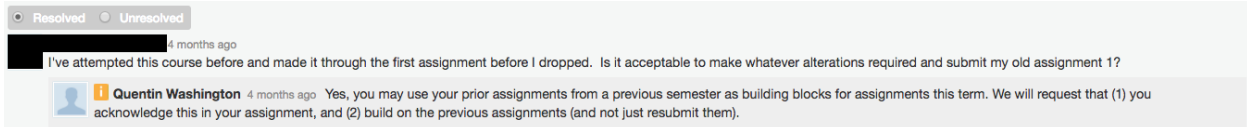


Figure 23. *In this question about a class assignment, the student asks whether they can reuse content from a previously submitted assignment. Cassidy could have answered this question, but did not because the question was asked outside the time interval in which she checks the class forum for new questions. Since a human TA answered the question by the time Cassidy checked the class forum again, Cassidy did not answer this question.*

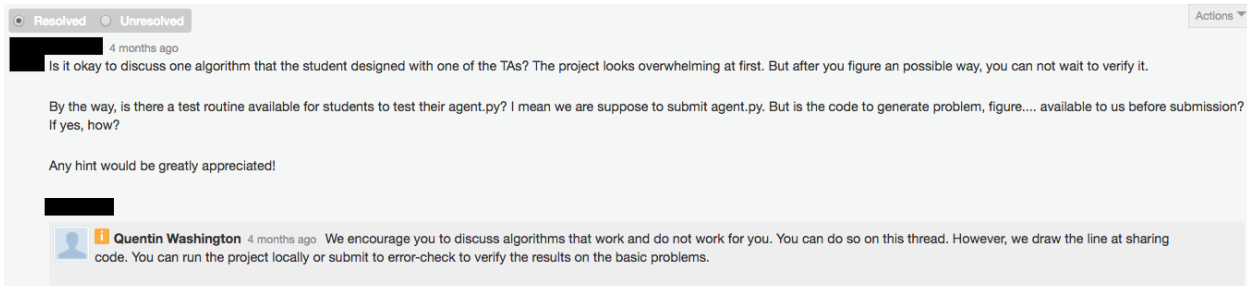


Figure 24. *In this question about a class project involving a coding component, the student asks whether they can discuss ideas with other students. Cassidy could have answered this question, but did not because another human TA Quentin Washington answered the question within 15 min. As Cassidy checks the discussion forum every 15 minutes, she did not have a chance to respond. Therefore, the next time she checked the class forum, since a another TA had already answered the question, she did not answer.*

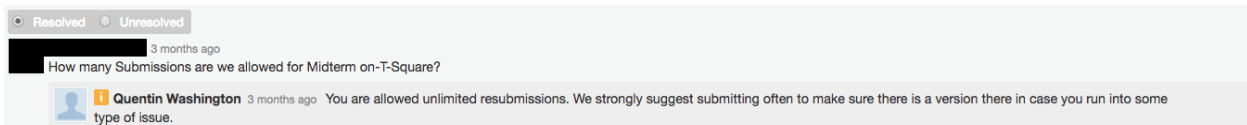


Figure 25. *In this question about the class midterm, the student asks about whether they can submit the midterm more than once. While Cassidy could have answered this question, we deliberately prevented her from answering questions about class submissions - those questions are among the most important that students ask, and for now we feel more comfortable that a human handles them.*

8. Student Reaction

In the KBAI classes in Spring 2016, Fall 2016, and Spring 2017, we shared the true identities of the virtual as AI agents towards the end of the term. Student reactions to our use of virtual teaching assistants in online discussion forums have been uniformly and overwhelmingly positive. Figure 26 illustrates a small sample of student reactions from the KBAI class in Spring 2016 after the students learned about the true identity of Jill Watson towards the end of April 2016.

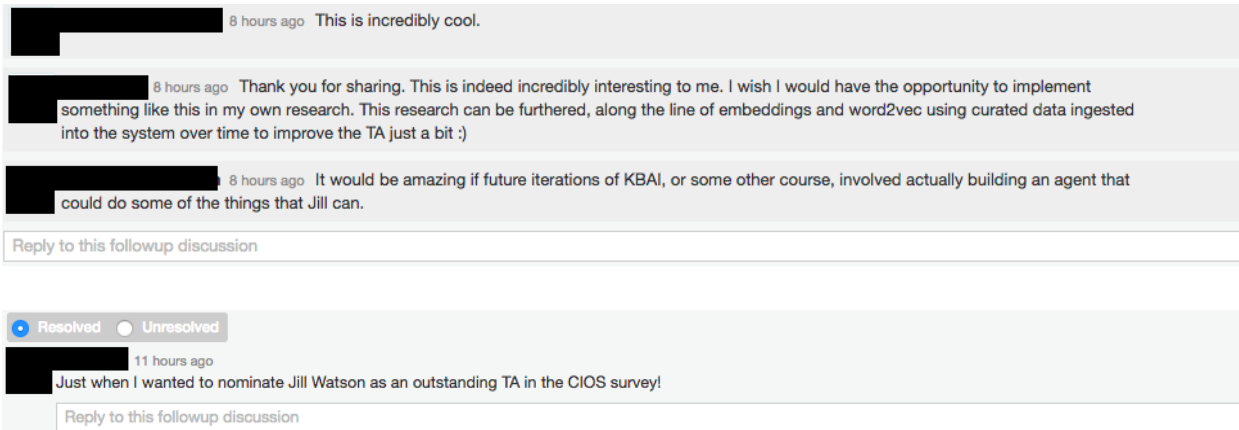


Figure 26. Students react to our class post in at the end of KBAI Spring 2016 class announcing the true identity of Jill Watson.

9. Discussion

There are several questions about the virtual teaching assistants that we have not fully answered in this article. The first question is how does Jill Watson work? As we briefly indicated above, Jill Watson 1.0 uses an episodic memory of questions and their answers from previous episodes. We developed JW1 using the IBM Bluemix toolsuite. In the second generation of Jill Watson, Ian Braun was a redeployment of JW1 for answering questions. However, Stacy Sisko used semantic information processing technology developed in our laboratory to reply to student introductions. In the third generation of Jill Watson, Cassidy Kimball too uses semantic information processing technology developed in our laboratory for answering questions as does Liz Duncan for replying to student answers.

Second, is the Jill Watson technology transferrable to other classes with different student demographics and using different educational infrastructures? To answer this question, we are presently building a new version of Jill Watson for a new Georgia Tech CS 1301 Introduction to Computing MOOC that presently has forty thousand students but no TA support whatsoever.

Third, is the Jill Watson technology effective in lowering the demands on the teaching staff? While it is too early to determine the answer to this question for the task of question answering, anecdotally there is some evidence to suggest that Jill Watson did reduce the load on the teaching staff for responding to student introductions and for posting messages to the class.

Fourth, is the Jill Watson technology effective in enhancing student performance and improving student retention. We are presently conducting studies and collecting data to answer this question about student engagement, learning and performance; it is too early to have insights into the issue of student retention.

Fifth, what ethical issues arise in using Jill Watson as an educational technology in an online classroom? As we mentioned above, we obtained IRB approval in advance of the Jill Watson experiments. Nevertheless, these experiments have raised several additional ethical issues. For

example, when it is appropriate to use AI agents without telling human subjects about them? Does the use of a feminine name for an AI agent implicitly promote gender stereotypes? Might the use of AI agents as virtual teaching assistants eventually result in reduced employment opportunities for human teachers? These are serious questions that require investigation.

10. Conclusions

We may view the Jill Watson experiments from several perspectives. First, we may view Jill Watson as an educational technology for supporting learning at scale. In fact, this was our primary initial motivation for developing Jill Watson and this is also how we motivated the discussion in this chapter. As indicated above, Jill Watson uses AI technology for supporting learning at scale by automatically answering a variety of routine, frequently asked questions, and automatically replying to student introductions.

Second, we may view Jill Watson as an experiment in developing AI agents so that for highly focused technical domains, highly selected subject demographics, and highly targeted context of human-computer interaction, it is difficult for humans to distinguish between the responses of AI and human experts. We found that in order to improve coverage, the design of Jill Watson gradually moved from using an episodic memory of previous question-answer pairs to using semantic processing based on conceptual representations.

Third, we may view Jill Watson as an experiment in human-AI collaboration. The KBAI class has become a micro-society in which humans and AI agents collaborate extensively and intensively, living and working together for long durations of time.

Acknowledgements

We thank IBM for its generous support of this work both by providing us access to IBM Bluemix and through multiple IBM Faculty Awards. We are grateful to Parul Awasthy for her contributions to the development of Jill Watson 1.0 in Spring 2016. We thank members of the Design & Intelligence Laboratory for many discussions on the Jill Watson project. We also thank the students and (human) teaching assistants of the KBAI course in Spring 2016, Fall 2016, and Spring 2017. This research has been supported by multiple Georgia Tech seed grants for research and education.

References

- Azevedo, R., & Aleven, V. (2013) *International Handbook of Metacognition and Learning Technologies*. Springer.
- Carey, K. (2016, September 28). An Online Education Breakthrough? A Master's Degree for a Mere \$7,000. *The New York Times*. Retrieved from <http://www.nytimes.com/2016/09/29/upshot/an-online-education-breakthrough-a-masters-degree-for-a-mere-7000.html>
- Chaudhri, V., Cheng, B., Overholtzer, A., Roschelle, J., Spaulding, A., Clark, P., Greaves, M., &

- Gunning, D. (2013a). Inquire Biology: A Textbook that Answers Questions. *AI Magazine*, 34.
- Clark, R., & Mayer, R. (2003). *E-Learning and the Science of Instruction*. San Francisco, CA: Pfeiffer.
- Daniel, J. (2012). Making Sense of MOOCs: Musings in a Maze of Myth, Paradox and Possibility. *Journal of Interactive Media in Education*.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287-322.
- Ferrucci D, Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A., Lally, A., Murdock, J.W., Nyberg, E., Prager, J., Schlaefter, N., & Welty, C. (2010) Building Watson: An overview of the DeepQA project. *AI Magazine*, 31:59–79.
- Ferrucci, D. (2012) Introduction to "This is Watson". *IBM Journal of Research and Development* 56(3):1-15.
- Goel, A., Anderson, T., Belknap, J., Creeden, B., Hancock, W., Kumble, M., Salunke, S., Sheneman, B., Shetty, A., & Wiltgen, B. (2016) Using Watson for Constructing Cognitive Assistants. In *Proceedings Fourth Annual Conference on Cognitive Systems*, Chicago, May 2016.
- Goel, A. & Joyner, D. A. (2016) An Experiment in Teaching Cognitive Systems Online. In Haynes, D. (Ed.) *International Journal for Scholarship of Technology-Enhanced Learning* 1(1).
- Goel, A. & Joyner, D. (2017) Using AI to Teach AI. *AI Magazine*, Summer 2017.
- Goel, A., Kunda, M., Joyner, D., & Vattam, S. (2013) Learning about Representational Modality: Design and Programming Projects for Knowledge-Based AI. In *Procs. Fourth Symposium on Educational Advances in Artificial Intelligence*, Bellevue, Washington, July 2013.
- Goodman, J., Melkers, J., & Pallais, A. (2016). Does Online Delivery Increase Access to Education? Harvard University Kennedy School Faculty Research Working Paper Series RWP16-035.
- Hollands, F., & Tirthali, D. (2014). MOOCs: Expectations and reality: Full report. Center for Benefit-Cost Studies of Education, Teachers College, Columbia University, NY. Retrieved August 1, 2015, from http://cbcse.org/wordpress/wp-content/uploads/2014/05/MOOCs_Expectations_and_Reality.pdf
- Joyner, D., Ashby, W., Irish, L., Lam, Y., Langston, J., Lupiani, I., ... & Goel, A. (2016, April). Graders as Meta-Reviewers: Simultaneously Scaling and Improving Expert Evaluation for Large Online Classrooms. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale* (pp. 399-408). ACM.
- Joyner, D., Goel, A., & Isbell, C. (2016). The Unexpected Pedagogical Benefits of Making Higher Education Accessible. In *Proceedings of the Third Annual ACM Conference on Learning at Scale*. Edinburgh, Scotland.
- Kay, R. (2012). Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior*, 28(3), 820-831.
- Koumi, J. (2006). *Designing video and multimedia for open and flexible learning*. Oxford, UK: Routledge Falmer.
- Kulkarni, C., Bernstein, M., & Klemmer, S. (2015). PeerStudio: Rapid Peer Feedback Emphasizes Revision and Improves Performance. In *Proceedings from The Second ACM*

Conference on Learning @ Scale. ACM.

- Leckart, S. (2012) The Stanford education experiment could change higher learning forever. *Wired*, March 2012.
- Li, L., Liu, X., & Steckelberg, A. (2010). Assessor or Assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology*, 41(3), 525-536.
- Ou, C., Goel, A., Joyner, D., & Haynes, D. (2016). Designing Videos with Pedagogical Strategies: Online Students' Perceptions of Their Effectiveness. In *Procs. Learning @Scale 2016*, pp.141-144
- Polson, M., & Richardson, J. (2013). *Foundations of intelligent tutoring systems*. Psychology Press.
- Raith, A. (2011) Stanford for Everyone: More Than 120,000 Enroll in Free Classes. MindShift. Retrieved from <http://ww2.kqed.org/mindshift/2011/08/23/stanford-for-everyone-more-than-120000-enroll-in-free-classes/>
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221.
- Van Zundert, M., Sluijsmans, D., & Van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, 20(4), 270-279.
- Yuan, L., & Powell, S. (2013). MOOCs and open education: Implications for higher education [White Paper]. Retrieved from Google Scholar.