**RESEARCH ARTICLE**                                                **Open Access**

# Platform-independent and curriculum-oriented intelligent assistant for higher education

Ramteja Sajja[1,3]* , Yusuf Sermet[3] , David Cwiertny[2,3,4,5] and Ibrahim Demir[1,2,3]

*Correspondence:
ramteja-sajja@uiowa.edu

[1] Department of Electrical Computer Engineering, University of Iowa, Iowa City, USA
[2] Department of Civil and Environmental Engineering, University of Iowa, Iowa City, USA
[3] IIHR Hydroscience and Engineering, University of Iowa, 300 S. Riverside Dr, Iowa City, IA 52246, USA
[4] Department of Chemistry, University of Iowa, Iowa City, USA
[5] Center for Health Effects of Environmental Contamination, University of Iowa, Iowa City, USA

## Abstract

Miscommunication between instructors and students is a significant obstacle to post-secondary learning. Students may skip office hours due to insecurities or scheduling conflicts, which can lead to missed opportunities for questions. To support self-paced learning and encourage creative thinking skills, academic institutions must redefine their approach to education by offering flexible educational pathways that recognize continuous learning. To this end, we developed an AI-augmented intelligent educational assistance framework based on a powerful language model (i.e., GPT-3) that automatically generates course-specific intelligent assistants regardless of discipline or academic level. The virtual intelligent teaching assistant (TA) system, which is at the core of our framework, serves as a voice-enabled helper capable of answering a wide range of course-specific questions, from curriculum to logistics and course policies. By providing students with easy access to this information, the virtual TA can help to improve engagement and reduce barriers to learning. At the same time, it can also help to reduce the logistical workload for instructors and TAs, freeing up their time to focus on other aspects of teaching and supporting students. Its GPT-3-based knowledge discovery component and the generalized system architecture are presented accompanied by a methodical evaluation of the system's accuracy and performance.

**Keywords:** Artificial Intelligence, Natural language processing, Machine learning, Transformers, GPT-3

## Introduction

One of the main causes of the knowledge disparities that lead to learning gaps among both undergraduate and graduate students is instructors' inability to communicate with these students in ways that suit the students' learning schedules and styles (Williamson et al., 2020). It has been widely shown in the literature that it is particularly effective to teach in ways that allow students to build conceptual understanding of the subject they are studying (Konicek-Moran & Keeley, 2015). This requires a certain degree of freedom and time for self-contemplation (Lin & Chan, 2018). Not surprisingly, allowing students to learn at their own pace positively contributes to a substantial increase in learning motivation and the development of creative thinking skills (Ciampa, 2014).

Sajja *et al. Int J Educ Technol High Educ*    (2023) 20:42

Page 2 of 24

A significant portion of students avoid or miss the opportunity to visit teaching assistants and instructors during office hours due to scheduling conflicts, the feeling of not being prepared, imposter syndrome, and shyness (Abdul-Wahab et al., 2019). Furthermore, most students study outside of regular work hours, which creates a need for assistance at odd times (Mounsey et al., 2013). The lack of immediate assistance can lead to discouragement and creates the feeling of being stuck despite the fact that many queries can be simply answered based on available material without in-depth expertise (Seeroo et al., 2021). Teaching assistants can sometimes fill this void, but they have their own responsibilities (e.g., classes, research, grading) which may render them unavailable during times such as exam weeks when the students need them most (Howitz et al., 2020). Thus, it would be extraordinarily helpful to develop new and more readily available forms of student assistance if this can be done without decreasing the time TAs and instructors have to spend on higher-level instruction (Mirzajani et al., 2016).

Information and communication tools and services play a crucial role in instructional technology and the learning process, enabling better knowledge dissemination and understanding. The purpose of this paragraph is to provide an overview of the current application areas of web technologies and AI in education and related domains, before moving on to discuss the potential of AI-based chatbots in higher education. Web technologies facilitate the delivery of curriculum in various fields, such as advanced modelling and analysis tools (Ewing et al., 2022), programming libraries (Ramirez et al., 2022), and the teaching of engineering ethics through serious games (Ewing & Demir, 2021). In the realm of AI, two primary areas of focus have emerged: information processing and knowledge communication. Deep learning models have been widely employed for tasks like image processing (Li & Demir, 2023), data augmentation (Demiray et al., 2021), synthetic data generation (Gautam et al., 2022), and modelling studies (Sit et al., 2021). However, the application of AI in information communication and delivery remains relatively under-explored, particularly in the engineering domain (Yesilkoy et al., 2022). Customized ontology-based smart assistants have seen successful implementation in public health care (Sermet et al., 2021) and environmental science (Sermet & Demir, 2018) studies. These examples demonstrate the potential value of AI-driven solutions in the educational context, paving the way for the development of chatbots that can address the communication challenges faced by students and instructors in higher education.

With the recent advancements (i.e., ChatGPT) in AI-based communication, there is a significant interest in the research of chatbots, which can be defined as intelligent agents (i.e., assistants) that have the ability to comprehend natural language queries and produce a direct and factual response utilizing data and service providers (Brandtzaeg & Følstad, 2017). Voice-based assistants are actively used in education, environmental science, and operational systems to access real-time data, train first responders (Sermet & Demir, 2022), and facilitate decision support coupled with other communication technologies like virtual and augmented reality (Sermet & Demir, 2020).

Technology companies have been taking the lead on operational virtual assistants integrated into their ecosystem which triggered a brand new and massive market that is forecasted to reach US\$ 11.3 billion by 2024 (IMARC Group, 2019). Several studies emphasize the potential chatbots hold to serve as the next-generation information communication tool and make the case for an urgent need for chatbot development

and adoption in their respective fields (Androutsopoulou et al., 2019; Miner et al., 2020; USACE, 2019; Vaidyam et al., 2019). However, the usage of chatbots for effective and reliable information communication is not widespread among the public, government, scientific communities, and universities (Schoemaker & Tetlock, 2017) and it is just starting to gain traction due to recent developments such as ChatGPT (OpenAI, 2022). Chatbots are increasingly being utilized across various applications, such as customer service (Pawlik et al., 2022) and educational contexts as a means of supporting teachers (Song et al., 2022). The adoption of virtual assistants within the context of the academic curriculum can help close the learning gaps identified above and, in the literature, (Hwang & Chang, 2021). Considering the prevalence of mobile phones and computers among students along with the recent remote-interaction culture that is gained during the pandemic, such technological and web-based solutions are relevant and needed more than ever (Iglesias-Pradas et al., 2021).

A recent report on the AI Market in the US Education Sector (TechNavio, 2018) emphasizes AI's focus on creating intelligent systems, discusses its increasing use in enhancing student learning, and states that intelligent interactive programs that are based on Machine Learning and Natural Language Processing help in overall learning of students. It is reported that the most significant market trend is the increased emphasis on chatbots (MindCommerce, 2019). The main aspect of how AI can be a vital tool in education is the utilization of AI in developing next-generation educational tools and solutions to provide a modern learning experience with the vision of personalized teaching, advising, and support (GATech2018; Ceha et al., 2021).

We propose an AI-augmented intelligent educational assistance framework that automatically generates course-specific intelligent assistants based on provided documents (e.g., syllabus) regardless of discipline or academic level. It will serve as a message-enabled helper capable of answering course-specific questions concerning scope and logistics (e.g., syllabus, deadlines, policies). The students can converse with the assistant in natural language via web platforms as well as messaging applications. The framework is conceived to address the listed issues and to unlock the immense potential of conversational AI approaches for education and enhancing the learning experience. Core benefits and advantages of the framework include the availability of assistance regardless of time, more TA and instructor time for advanced and customized advising, answers to time-consuming and repetitive questions, reduced human error due to miscommunication for course logistics, and accommodations for personal barriers, cultural, and disability-related issues (e.g., language barrier). A case study is conducted to quantitatively measure the proposed approach's efficacy and reliability within the context of the cited benefits.

In the context of complex trajectories, such as changes in degree programs or disciplines, students often face additional challenges in understanding the requirements, logistics, and subject policies associated with their new academic paths. By providing tailored information and support through an AI-augmented intelligent educational assistance framework, we aim to improve students' engagement and learning experiences and help them navigate these complex trajectories more effectively. This will enable them to make informed decisions about their academic paths, including the choice of subjects within a degree program or transitioning from one program to another.

The remainder of this article is organized as follows: "Related work" section summarizes the relevant literature and identifies the knowledge gap. "Methods" section presents the methodology of the design choices, development and implementation of a course-oriented intelligent assistance system based on syllabi. "Case study design" section describes the case study design. "Results and discussion" section describes the preliminary results and provides benchmark results and performance analysis. "Conclusion and future work" section concludes the articles with a summary of contributions and future work.

## Related work

There have been several initiatives to leverage conversational interfaces in education systems and higher learning (Chuaphan et al., 2021; Hobert, 2019; Wollny et al., 2021). Georgia Tech pioneered a virtual teaching assistant (TA) named "Jill Watson" and reported inspiring results for student satisfaction (GATech, 2018). Additionally, many students were inspired and created their own chatbots that converse about the courses, exhibiting increased interest in AI tools. The positive impacts of cultivating a teaching motivation for individual learning are successfully demonstrated in the University of Waterloo's Curiosity Notebook research project, in which the students reported increased engagement upon conversation with an intelligent agent (i.e., Sigma) that asks Geology-related questions in a humorous manner (Ceha et al., 2021). Several universities have similar projects exploring AI's role in education including Stanford University (Ruan et al., 2019) and Carnegie Mellon University (Helmer, 2019). Further initiatives have been explored to utilize chatbots in certain aspects of campus life (Dibitonto et al., 2018; Duberry & Hamidi, 2021). Georgia State University developed a virtual assistant for admission support (i.e., Pounce) to incoming freshmen students. The Randomized Control Trial they implemented to assess effectiveness yielded that first-generation and underrepresented groups disproportionately benefited from the system which resulted in a decreased gap in graduation rate among different demographics (Hart, 2019). Furthermore, 94% of the students recommended GSU to continue the service, citing their satisfaction in receiving instant responses any time of the day without the feeling of being judged or perceived as unintelligent (Mainstay, 2021).

The process of creating a knowledge framework includes retrieving relevant documents and extracting answers from the retrieved documents (Zylich et al., 2020). One of the main documents that can be used to acquire course information to answer logistical questions is the syllabus (Zoroayka, 2018; Zylich et al., 2020). Chatbots can also be extended to other works such as helping students with technical issues and questions (Chuaphan et al., 2021). The limits of TA's human resources can be addressed with the help of these chatbots. A chatbot was deployed at Stanford University to respond to student inquiries for a class by compiling information from their participation in an online forum (Chopra et al., 2016). Similarly, a solution to augment the staffing shortages is with an AI Teaching Assistant (Ho, 2018). In addition to assisting with staff shortages, the virtual teaching assistants improve students' educational experiences (du Boulay, 2016). The chatbots can be developed internally using readily available open source tools (Zylich et al., 2020) or through the use of cloud-based language models (Benedetto et al., 2019; Chuaphan et al., 2021; Ranavare & Kamath, 2020).

Sajja *et al. Int J Educ Technol High Educ*    (2023) 20:42

Page 5 of 24

The literature review clearly shows the importance of chatbots and how they can be used in the educational setting. Based on the survey done by Abdelhamid & Katz, 2020, more than 75% of students who responded to the study said they have previously used a chatbot service or other comparable system. 71% of the students stated that they find it challenging to meet with their teaching assistants for a variety of reasons. More than 95% of students claimed that having a chatbot available would be beneficial for providing some of their questions with answers. Though the previous work puts forth limited-scope case studies that clearly serve as proof of potential and benefits of utilizing conversational approaches in the educational setting, a complete and multidisciplinary solution has not been introduced to transform teaching and learning. A major distinction of the proposed framework in contrast to relevant work is the ability to automatically generate a ready-to-use intelligent assistant based on dynamic input provided in the format of a textual document, such as a curriculum summary and syllabus. It is both independent from the field and the technology (e.g., learning management systems) used for content delivery. Furthermore, it relies on a Service-Oriented Architecture (SOA) to enable integration into any delivery channel.

## Methods

This section discusses the method used in the research, specifically focusing on natural language inference, syllabus knowledge model, system architecture, intelligent services, and framework integration. The method explores the use of language models such as GPT-3 and the generation of a knowledge graph based on syllabus templates. It also explains the system architecture of the VirtualTA framework, which consists of four major components. The intelligent services component utilizes deep learning-powered natural language tools, while the framework integration involves the incorporation of the VirtualTA system into various communication channels. Overall, this section provides a comprehensive overview of the approach taken to develop and implement the VirtualTA system.

### Natural language inference

In recent years, the rapid expansion of data volume was facilitated by technological innovation. A Forbes survey from a few years back revealed that 2.5 quintillion bytes of data were produced every single day. According to current estimates, unstructured data makes up more than 90% of all information stored (Kim et al., 2014). The introduction of language models as a foundation for numerous applications trying to extract useful insights from unstructured text was one of the major forces behind such research. In order to anticipate words, a language model analyzes the structure of human language. There are multiple available large language models such as BERT (Devlin et al., 2019), XLNet (Yang et al., 2020), RoBERTa (Liu et al., 2019), ALBERT (Lan et al., 2020), GPT-3 (Brown et al., 2020), GPT-2 (Radford et al., 2019), and PaLM (Chowdhery et al., 2022).

OpenAI provides the generative pre-trained Transformer 3 (GPT-3), an autoregressive language model that uses deep learning to generate writing that resembles that of a human. GPT-3 can be utilized off the shelf as well as by using a few-shot learning technique and fine-tuning the model in order to adapt it to any desired application area.GPT-3 has been pre-trained on a large quantity of text from the public

Sajja *et al. Int J Educ Technol High Educ*    (2023) 20:42

Page 6 of 24

internet resources, which may be regarded as few-shot learning. When given only a few instances, it can typically figure out what task you're attempting to complete and offer a convincing solution. Fine-tuning builds on fine-tuning learning by training on many more instances that can fit in the prompt, allowing to obtain higher outcomes. Once a model has been fine-tuned, it no longer needs examples in the prompt. Fine tuning also reduces expenses and makes lower-latency requests possible. We decided to choose GPT-3 because the model is cloud based and has a developer friendly API. GPT-3's Davinci is the biggest model in terms of parameters that's available to use by researchers and the general public.

### Syllabus knowledge model

It is crucial to pick the correct questions to pose to the chatbot in order to test its accuracy. The key questions that could arise from the course syllabus were extracted using the literature on syllabus templates because the goal of this research was to create a chatbot that could answer questions using the course logistics and policies from the syllabus. The important sections of a syllabus or course description is Course Information, Faculty Information, Instructional Materials and Methods, Learning Outcomes, Grading and Course Expectations, Policies, and Course Schedule (Hess et al., 2007; Passman & Green, 2009). Similarly critical sections such as disability statements, academic misconduct, inclusivity, accessibility and harassment; and optional information such as mental health resources can also be included in the syllabus (Wagner et al., 2022). Based on this literature, we included questions that were related to the following topics: (1) Course Information, (2) Faculty Information, (3) Teaching Assistant Information, (4) Course Goals, (5) Course Calendar, (6) Attendance, (7) Grading, (8) Instructional Materials, and (9) Course and Academic Policies. Course Information, Faculty Information and TA Information sections of the knowledge graph developed to encompass a standard syllabus in higher education is shown in Table 1. After analyzing all the main categories specified above, we have generated 36 questions to reflect the information included in these categories. We also used text and data augmentation techniques on these 36 questions to generate 120 questions in total to reflect the different ways a question could be asked. This text and data augmentation approaches are reflected in Table 6 in the form of competency questions.

### System architecture

The VirtualTA framework can be partitioned into four major cyber components with specialized functions (Fig. 1). The first component can be attributed to curating and indexing appropriate classroom resources for information that falls within the scope of a syllabus, as described in Table 1. The second component contains the cyber framework to create, serve, and manage the smart assistant. It includes server management, API access from the perspectives of both students and instructors, data analytics, and smart assistant management. The Intelligent Service component is concerned about the deep learning-powered natural language tools that are provided under the umbrella of VirtualTA framework (e.g., inference and intent mapping, emotion detection and lightening the mood with witty yet helpful responses). Finally, the integration component is

**Table 1**  Knowledge graph for a syllabus in higher education

| Category | Syllabus Element |
| --- | --- |
| Course information | Course name |
| | Course number |
| | Credit hours |
| | Location and class times |
| Faculty information | Name |
| | Contact information |
| | Office location |
| | Office hours |
| TA/teaching assistant information | Name |
| | Contact information |
| | Office location |
| | Office hours |
| Course goals and/or objectives | Course objectives |
| | Expectations from the course |
| Course calendar | Due dates |
| | Assignment dates |
| Attendance and classroom behavior | Attendance policy |
| | Expected classroom behavior |
| Grading and assignments | Grading criteria |
| | Tentative exam schedule |
| Instructional materials | Textbooks |
| | Other required materials for the course |
| Policies | Late assignments |
| | Academic dishonesty |
| | Disability statement |
| | Freedom of speech |
| | Makeup policy |
| | Mental health resources |
| | Absences for religious holy days |



**Fig. 1**  System architecture and components of VirtualTA

Sajja *et al. Int J Educ Technol High Educ*     (2023) 20:42

Page 8 of 24

concerned with communication channels the smart assistant can be served from and entails the appropriate protocols, webhooks, and software.

The sub-section titled "Intelligent Services" describes the process of generating a course knowledge model to power the VirtualTA system. It explains how the raw syllabus document is parsed and processed using GPT-3 to extract relevant information. On the other hand, the sub-section titled "Framework Integration" focuses on the integration of the VirtualTA system into various communication channels. It describes the centralized web-based cyberinfrastructure used for data acquisition, training deep learning models, and storage of course-specific information.

### Intelligent services

*Course knowledge model generation* In order to power the VirtualTA, the raw syllabus document needs to be parsed to create a knowledge model (Fig. 2). The process for knowledge model generation entails utilizing GPT-3 to attempt to find relevant snippets out of unstructured text by using the competency questions provided in Table 6. The retrieved information for each syllabus element is curated and stored in a JSON file. Upon post-processing and validation, the resulting knowledge model is used to fine-tune the model for question answering.

*Pre-processing* Once the syllabus file is parsed, the generated data is split into smaller pieces or documents to lower the cost to use the GPT-3 model and lower the latency time. The data was initially divided into 2000 characters, but this led to increased API request costs. We created our final version of the code to divide the syllabus content into documents with 200 characters without compromising accuracy or the model's affordability. It was made sure that the split does not divide a word into different syllables.

*Post-processing* When the extracted text from the syllabus does not contain the information the question was intended for, the GPT-3 model may sometimes return irrelevant snippets to the asked question. In some cases, the model can return partial answers as well; partial in the sense that the response has the correct information, yet it is not complete (e.g., returns the information of 3 TAs out of 5 total TAs listed on the course description). To address and resolve these edge cases, the instructors (e.g., teaching assistant, faculty) are provided with the initial draft of the automatically populated knowledge graph and validate the information or modify as needed before the graph can be fed to the model for question answering. This is a one-time process, where the instructor(s) or TAs can go through the template at the start of the semester, to check the proposed answers by the model, and modify the knowledge base with accurate information. Throughout the semester, this workflow can be repeated as needed if major changes occur in the syllabus.

*Question answering (QA)* The questions answering process relies on the provided course knowledge model and two models to understand the intent, map it to the
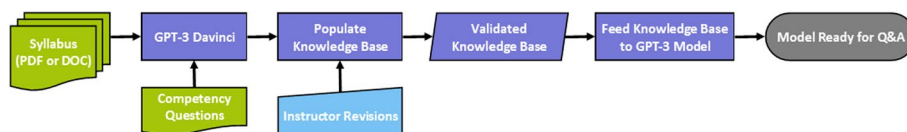


**Fig. 2** Knowledge base population process with instructor revision

Sajja *et al. Int J Educ Technol High Educ*     (2023) 20:42

Page 9 of 24

requested resource, and produce a natural language response in the form of a to-the-point and concise answer. The question-answering framework from GPT-3 by OpenAI works in two parts. The first part of the QA process is the search model, which is used to search the provided documents. This model then lists the most applicable documents to answer the given question. For this, we created a fine-tuned model rather than using the models provided by OpenAI. Our fine-tuned model is trained on the *Stanford Question Answering Dataset* (SQuAD) dataset. Both the training and validation datasets have 1314 entries each for our fine-tuned model. The second part of the QA process is the completion model, which is a built-in model provided by OpenAI named "text-davinci-002". Davinci is the most capable model family provided by OpenAI. This model is good at understanding content, summarization and content generation. The completion model is used to generate the answer from the documents provided by the Search Model.

As a way to expand the accessibility as well as the human-like and empathetic interactions of the system in a way that it can seamlessly serve with an approachable persona, several enhancements were devised and implemented. The question-answering system operates in a variety of languages. This was accomplished through the use of GPT-3's language translation capabilities. Students can ask a question in any language supported by GPT-3, which we then translate to English, send to VirtualTA, receive an answer in English, and finally translate back to the language the question was asked in. VirtualTA can be tailored to the demands of the students by fine-tuning the model using the questions asked by the students and answers provided by the model. This customization can be done at the domain or course level. The sentiment of a student's question can be analyzed by VirtualTA, and if negative emotions or stress are identified, the system will give positive comments or optimistic messages to lighten the situation and point them towards appropriate available resources.

### Framework integration

The framework is founded upon a centralized web-based cyberinfrastructure for data acquisition, training deep learning models, storage and processing of course-specific information, as well as to host the generated chatbots for use in communication channels. The cyberinfrastructure entails a NGINX web server, NodeJS-based backend logic, a PostgreSQL database, accompanied with caching, and user and course management mechanisms. The core intelligent assistant is created based upon the Service-Oriented Architecture, allowing its plug-and-play integration into any web platform with webhooks. Several integrations have been realized as part of this paper to showcase the system's utility, although it can potentially be integrated into numerous channels (e.g., augmented and virtual reality applications, automated workflows, learning management systems).

*Web-based conversational interface* To make asking questions and receiving responses easier, a web-based chatbot application user interface (UI) has been developed. The UI designed by Palace C was modified for this development using standard JS. Through the API we developed, VirtualTA's replies may be retrieved. The user sees this response on the chatbot provided by VirtualTA. This may be included into any web-based chatbot by using the API we established to receive an answer and utilize it as the chatbot's response.

This procedure enables VirtualTA's functionality to be incorporated into any web-based conversational bot.

*Social platforms* A Discord bot is created to allow students to include the VirtualTA into workspaces they already use for specific courses to facilitate easy access to pertinent information. The availability of VirtualTA into social messaging platforms students already utilize allows for easy adoption of the system as well as an organic and friendly interaction. This is due to the fact that students are already familiar with similar technologies (Benedetto et al., 2019).

*Smart apps and devices* VirtualTA is integrated to work with Google Assistant. We have created an API to return an answer, when asked a question regarding a course. We have used Google DialogFlow to integrate VirtualTA as a third-party action on Google Assistant. Students can access VirtualTA using Google Assistant on their mobile phones, smart home devices, Android TV, Android Auto and smartwatches. This integration has been deployed in the test environment on this platform and screenshots of these implementations are shared in Sect. 4 below.

## Case study design

In order to establish the accuracy and performance of the VirtualTA, an assessment was conducted by collecting 112 syllabus files from a variety of institutions and domains, including Engineering, Math, Physics, History, Computer Science, English, Art, Business, Philosophy, Arabic, Anthropology, Accounting, Chemistry, Music, and Economics. We removed 12 of these files because the syllabi were in image format and text extraction from the image could hinder the benchmark of VirtualTA's capabilities. Hence, a case study is designed upon 100 syllabi and in two phases to assess the performance for (1) extracting data from syllabi and (2) mapping user questions to the extracted syllabi data. It is important to note that this study was not conducted in an actual classroom setting, but rather in a controlled environment for the purposes of evaluating the performance and effectiveness of the VirtualTA system. Nonetheless, the results of this study provide valuable insights into the potential of VirtualTA as an effective educational tool for supporting student learning and engagement.

### Phase 1—knowledge extraction

We chose 38 files from the 100 syllabus files we collected. We asked VirtualTA 36 questions we chose based on main categories for every syllabus file. The goal of this is to measure the accuracy of the bot on the frequently asked questions. We had three parameters we are collecting from this study: number of questions answered correctly, number of questions answered incorrectly, and number of questions partially answered. These parameters can be seen below in the template that is created for one of the courses and is in JSONL format.

Before edits

```
{"QUESTION":"What is the name of the course?","ANSWER":"BUS 100","isTrue":"Change this to TRUE or FALSE or PARTIAL"}
{"QUESTION":"What is the course number?","ANSWER":"The course number is BUS 100.","isTrue":"Change this to TRUE or FALSE or PARTIAL"}
```

{"QUESTION":"How many credit hours is this course worth?","ANSWER":"This course is worth 3 credit hours.","isTrue":"Change this to TRUE or FALSE or PARTIAL"}

After edits

{"QUESTION":"What is the name of the course?","ANSWER":"Introduction to Business","isTrue":"FALSE"}

{"QUESTION":"What is the course number?","ANSWER":"The course number is BUS 100.","isTrue":"TRUE"}

{"QUESTION":"How many credit hours is this course worth?","ANSWER":"This course is worth 3 credit hours.","isTrue":"TRUE"}

Once we collect the answers from the bot on all the 36 questions for the syllabus file, we store these results in a JSONL file format. In the file we have three fields including question, answer and isTrue flag. The question field contains the question asked to the bot, the answer field contains the answer we got from the bot and the isTrue is whether the given answer is correct or incorrect. We manually went through all the 38 JSONL files and checked the answers with the actual syllabus file and change the isTrue field to "TRUE" if the answer given by the bot is correct, "FALSE" if the answer given by the bot is incorrect and "PARTIAL" if the answer given by the bot is partially correct. When an answer is false/incorrect, in addition to making the isTrue "FALSE", we also change the answer to the correct version/information. These manual corrections were done to use this information for our second phase of testing. The manual corrections explained above could be seen in After Edits, this shows the corrections made and isTrue field set to either "FALSE"," TRUE" or "PARTIAL" and these changes have been highlighted in green.

### Phase 2—question answering

We use the manually corrected templates created in Phase 1. Here in this phase of testing we increase the questions asked from 36 to 70. This was done using text augmentation to test the model's question and answering performance on different question asking techniques or structures. Each question has at least one other variation except for two questions. The rationale for these two questions, "How do I submit my assignments?" and "when is the final exam?", left out from data augmentation because we couldn't discover a sensible and good approach to supplement or augment these queries. We had three parameters collected during this study: number of questions answered correctly, number of questions answered incorrectly, and number of questions partially answered.

Once we collect the answers from the bot on all the 70 questions for the template file created in Phase 1, we store these results in a JSONL file format. In the file, we have three fields including question, answer and isTrue flag. The question field contains the question asked to the bot, the answer field contains the answer we got from the bot and the isTrue field is whether the given answer is correct or incorrect. We manually went through all the 38 JSONL files and checked the answers with the actual syllabus file and change the isTrue field to "TRUE" if the answer given by the bot is correct, "FALSE" if the answer given by the bot is incorrect and "PARTIAL" if the answer given by the bot is partially correct.

## Results and discussion

In this section, we present the results and discussion of our study on the integration and performance evaluation of the VirtualTA system. The section is divided into two main subsections: communication channels and performance evaluation. The communication channels subsection outlines the various platforms on which the VirtualTA system was integrated, including popular messaging platforms, learning management systems, and mobile applications. It discusses how the integration of the VirtualTA system with these platforms helped to improve access to course-related information and reduce the logistical workload for instructors and TAs.

The second subsection, performance evaluation, is further divided into two parts: knowledge extraction and question answering. The knowledge extraction section evaluates the accuracy and effectiveness of the system's knowledge discovery component, which is based on the GPT-3 language model. The question-answering section evaluates the system's ability to provide accurate and relevant responses to student queries.

### Communication channels

The UI for the web platform shown in Fig. 3 has been adapted from (Palace, 2021). Figure 3 shown below is a web platform designed using vanilla JavaScript. It shows select Competency Questions asked to the model and the answers returned by the model for a history course. Figure 4 shows the integration of VirtualTA with Discord. In the figure, the questions asked to the model and the responses given by the model for a STEM course specifically a CS course are provided.

Figure 5 shows the integration of VirtualTA with Google Assistant. The questions are asked to VirtualTA using voice. The command "talk to Virtual T.A." is needed to connect Google Assistant to the third-party action VirtualTA. These figures show the questions asked and the answers returned by VirtualTA for a history course.

Figure 6 illustrates the capability of language translation in VirtualTA. The VirtualTA support languages for Spanish, French, and German are shown in (A), (B), and (C),
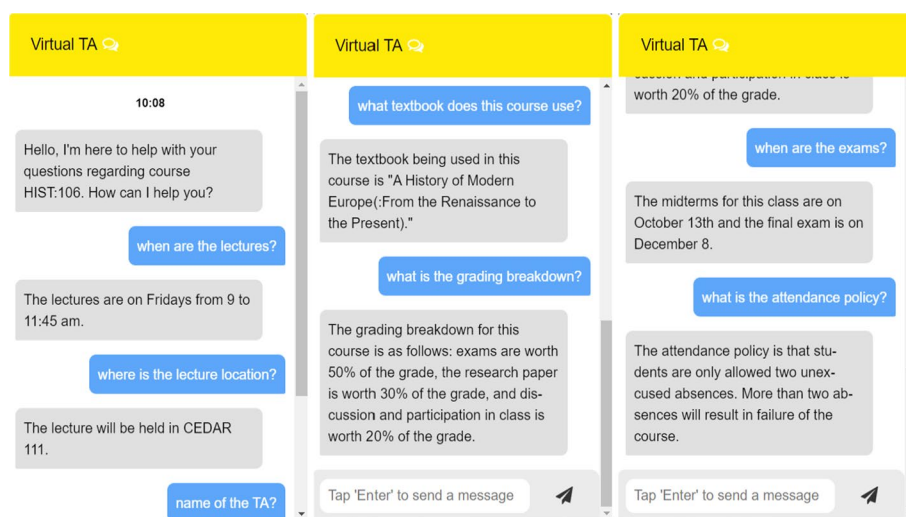


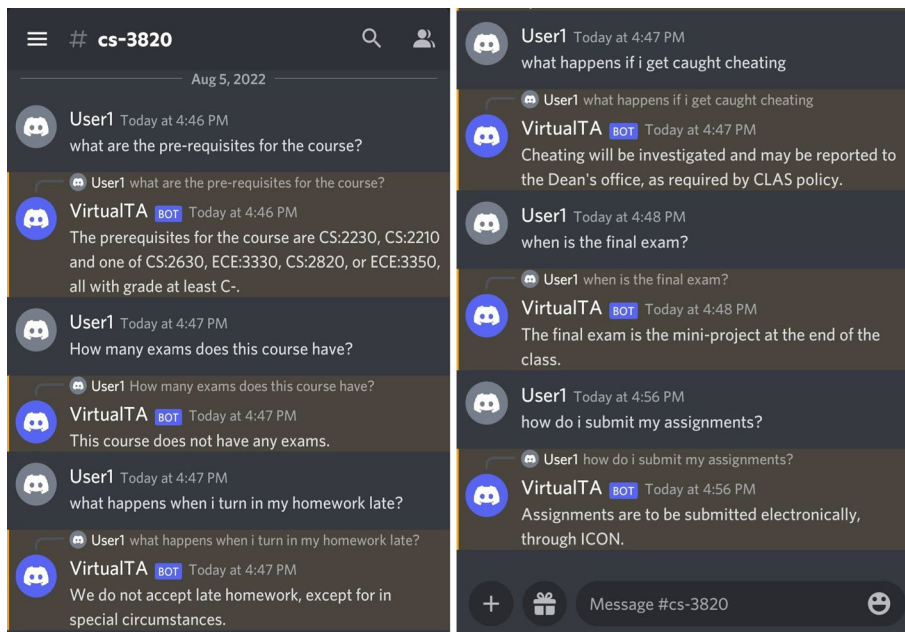**Fig. 3** Web based chatbot user interface with questions and answers

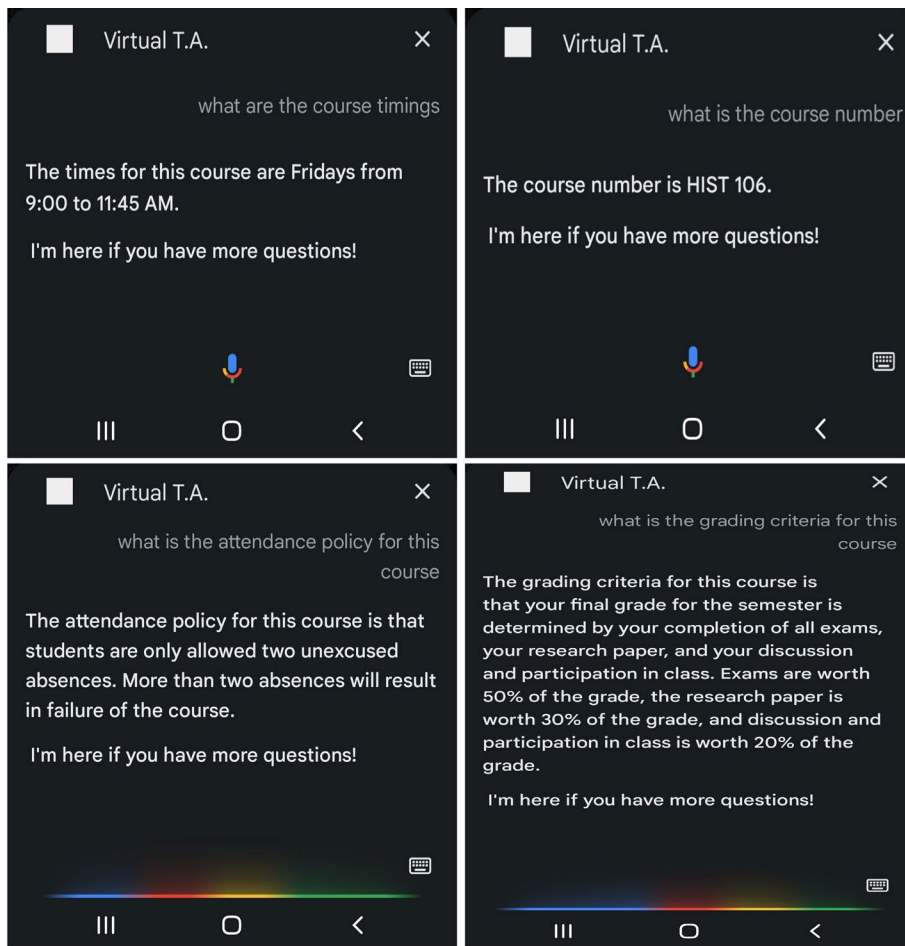**Fig. 4** Integration of VirtualTA with Discord social media application



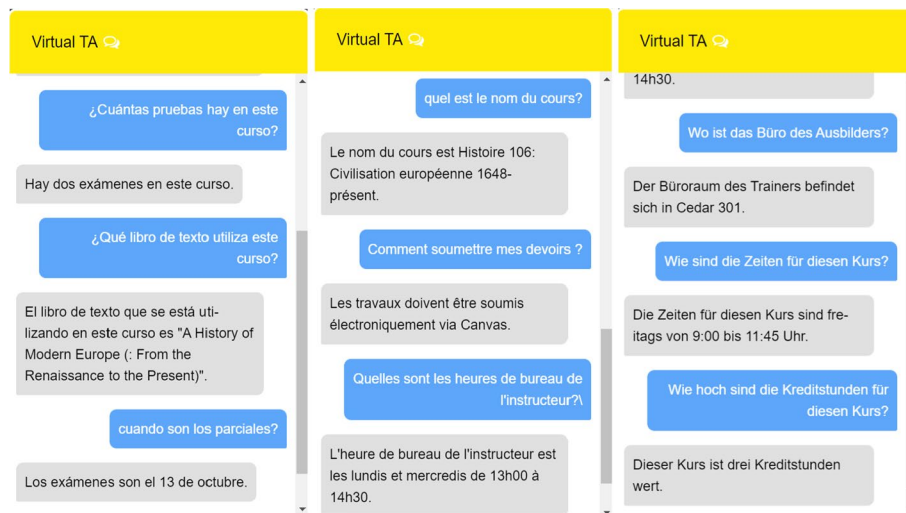**Fig. 5** Integration of Virtual TA with Google Assistant application

**Fig. 6** VirtualTA language translation capabilities

respectively. Any language that GPT-3 supports can be used by the user to ask a query, and VirtualTA will respond in that language.

The possibilities of VirtualTA's sentiment analysis are displayed in Fig. 7. The model responds with the standard response and also in a humorous or witty way to lighten the situation if it determines that the user is asking a question with a negative emotion or feeling. Figure 7 shows some examples of sentiment analysis in VirtualTA. Private information, including the instructor's email address, has been obscured in Fig. 7 (center). As illustrated in Fig. 7, our system generates two types of responses. The first is a standard reply that includes only the answer to the student's question. The second type of response, triggered only when negative emotions are detected in
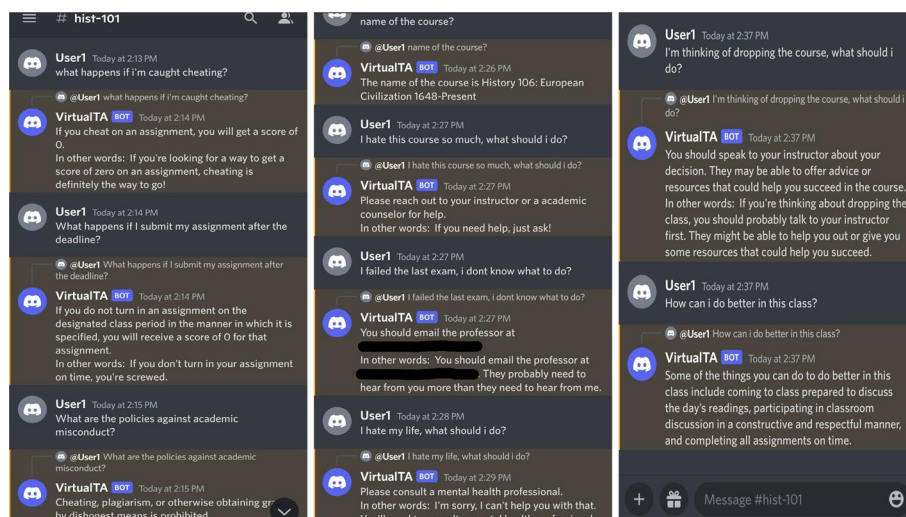


**Fig. 7** VirtualTA sentiment analysis for a history course

the student's query, begins with the phrase "In other words" and serves to alleviate tension and lighten the situation. The implementation of this experimental feature is designed to enhance the overall user experience and demonstrate a sensitivity to the emotional state of the student. By identifying negative emotions in the question and responding in a manner that acknowledges and addresses those emotions, we aim to improve the efficacy of the system and promote a more positive and productive learning environment.

### Performance evaluation

To quantify the model's effectiveness, precision (Eq. 1), recall (Eq. 2), and f1-score (Eq. 3) metrics have been selected for this imbalanced classification problem with multiple classes as formulated below (Sokolova & Lapalme, 2009). *n value* in the equations below represents the number of different questions in the FAQ (i.e., classes). For computing results using Eqs. 1–3 (Sokolova & Lapalme, 2009), we have used the criteria listed below and computed two sets of results where one includes "PARTIAL" as correct/TruePositive and the other does not consider "PARTIAL" as correct/TruePositive.

$$\text{Precision(multiclass)} = \frac{\sum_0^n TP}{\sum_0^n TP + \sum_0^n FP} \tag{1}$$

$$\text{Recall(multiclass)} = \frac{\sum_0^n TP}{\sum_0^n TP + \sum_0^n FN} \tag{2}$$

$$\text{f1} - \text{score(multiclass)} = \frac{2 \times precision \times recall}{precision + recall} \tag{3}$$

The aim of the testing phase is to optimize the precision and recall values to build an accurate and complete system, however a trade-off evaluation is necessary (He & Ma, 2013). Depending on the specific use case, it may be necessary to optimize for precision in order to provide highly accurate answers, or to optimize for recall in order to match as many questions as possible while minimizing the sacrifice of accuracy. For this use case, we tried to maximize the precision value for the model.

For each of the 38 syllabus files utilized in the testing phase, these performance values were calculated and analyzed. VirtualTA prioritizes the accuracy of the responses over giving an answer. We want to provide the most accurate results to the students. It is better to respond with "Answer not found" than to give an incorrect answer. Especially where the incorrect answer could misinform a student and can lead to them missing office hours or homework deadlines. When the model is unsure of the answer or is unable to locate pertinent documents, it responds to the user as "Response not found," which prevents it from providing the user an incorrect answer and allows them to double-check the answer outside of VirtualTA with instructor or TA.

**Table 2** Accuracy for Phase 1 testing results

| Category name | Number of questions per category | Accuracy % | Accuracy with partial content % |
|---|---|---|---|
| Course information | 6 | 63.2 | 69.7 |
| Faculty information | 4 | 60.5 | 73.0 |
| TA information | 4 | 49.3 | 57.2 |
| Course goals | 2 | 90.8 | 94.7 |
| Course calendar | 3 | 90.4 | 92.1 |
| Attendance | 2 | 93.4 | 97.4 |
| Grading | 4 | 63.8 | 71.0 |
| Instructional materials | 2 | 92.1 | 96.0 |
| Policies | 9 | 95.0 | 95.6 |
| Overall | 36 | 76.5 | 81.6 |

**Table 3** Performance metrics for Phase 1 testing

| Performance metrics | Without partial | Includes partial |
|---|---|---|
| Recall | 0.64 | 0.69 |
| Precision | 0.79 | 0.81 |
| f1-score | 0.71 | 0.74 |

### *Knowledge extraction*

For the knowledge extraction phase, Tables 2 and 3 provide the measured metrics (i.e., accuracy, precision, recall, f-1 score). The common problems faced by the model is in the "Teaching Assistant Information" section, and this could be due of many reasons, and we identified some of these cases as the reason for lower accuracy in our case study as follows: (1) when a course or syllabus has multiple TAs (Teaching Assistants), the model is failing to detect all the TAs from the text provided, (2) when a course doesn't have a TA listed; the model fills in the TA questions with the instructor's information. For instance: when there is no TA for the course and the user asks, "When are the TA's office hours?" the model replies back with the instructor's office hours, (3) the formatting of some syllabuses we tested on were really confusing or messy. This resulted in the model missing simple questions such as the course number or course name. In calculating the results discussed in this section, we considered all of the above cases (1–3) as Incorrect/False. Phase 1 is a study conducted on VirtualTA by asking 36 questions, selected from 38 syllabus files, to measure its accuracy in answering frequently asked questions. The study collected data on the number of questions answered correctly, incorrectly, and partially. The results were recorded in a JSONL format.

### *Question answering*

For the question answering phase, Tables 4 and 5 provide the measured metrics (i.e., accuracy, precision, recall, f-1 score). The common problems faced by the model in this phase is in the "Course Information" section, and this could be because of many edge

**Table 4** Performance metrics for Phase 2 testing

| Performance metrics | Without partial | Includes partial |
| --- | --- | --- |
| Recall | 0.87 | 0.88 |
| Precision | 0.96 | 0.96 |
| f1-score | 0.91 | 0.92 |

**Table 5** Accuracy values for Phase 2 testing

| Category name | Number of questions per category | Accuracy % | Accuracy with partial content % |
| --- | --- | --- | --- |
| Course information | 12 | 82.2 | 83.6 |
| Faculty information | 8 | 97.4 | 99.0 |
| TA information | 8 | 98.7 | 98.7 |
| Course goals | 4 | 98.0 | 99.3 |
| Course calendar | 5 | 99.5 | 100 |
| Attendance | 4 | 98.7 | 100 |
| Grading | 7 | 95.5 | 98.5 |
| Instructional materials | 4 | 96.0 | 97.4 |
| Policies | 18 | 99.0 | 99.6 |
| Overall | 70 | 95.3 | 96.5 |

cases, and we identified two main cases as the reason for lower accuracy in our case study as follows: (1) when the number of credit hours are not given in the context, the model tries to calculate the credit hours based on the number of lectures per week; (2) there is a small chance where the model can give different answers to similar questions, this could be based on the style of questioning. In calculating these results discussed in this section, we considered all of these above cases as Incorrect/False. In Phase 2 of testing, VirtualTA was asked 70 questions using text augmentation techniques to measure its performance in answering questions with different structures. The questions were based on manually corrected templates from Phase 1, with two questions excluded due to difficulty in augmentation. Data was collected on the number of questions answered correctly, incorrectly, and partially.

## Conclusion and future work

In this research, we designed an automated system for answering logistical questions in online course discussion boards, third party applications or educational platforms and highlighted how it can aid in the development of virtual teaching assistants. Specifically, the project's aims include enhancing course content quality and individualized student advising by delegating the time-consuming repetitive duties of instructors to virtual assistants and mitigating inequality among students in accessing knowledge to narrow retention and graduation gaps. Additionally, by providing support for students navigating complex trajectories, such as changes in degree programs or

disciplines, the virtual assistant can facilitate better decision-making and a smoother transition between academic paths. This research was conducted under controlled circumstances rather than in an actual classroom, hence while the results may not fully mirror an authentic educational environment, they nonetheless provide significant insights into the capabilities of VirtualTA as a tool for bolstering student learning and engagement.

Through this architecture, VirtualTA can be integrated with third-party applications to enable access from a variety of intermediaries, such as web-based systems, agent-based bots (such as Microsoft Skype and Facebook Messenger), smartphone applications (such as smart assistants), and automated web workflows (e.g., IFTTT, MS Flow). Users will find it simple to access VirtualTA through any communication channel that is familiar to them and that they feel comfortable using. Additionally, it enables any number of users enrolled in the course to access the system. We want to expand upon our existing approach to include course content in addition to the syllabus or administrative information.

While chatbots can be helpful in reducing the workload of instructors and TAs in education, it is important to acknowledge that they cannot completely replace human interaction and support. However, with proper development and implementation, chatbots can be effective tools to enhance the learning experience for students. As mentioned in the results section, certain types of questions are more likely to cause errors in chatbot responses. Educators and developers of chat-based educational platforms should be aware of these potential pitfalls and take steps to minimize errors. This may include incorporating natural language processing (NLP) algorithms to identify and flag potentially confusing or ambiguous questions before they are sent to a chatbot or human responder.

Future studies can focus on further enhancements to the AI-augmented intelligent educational assistance framework to better support students' complex trajectories, including personalized advising based on their academic paths, integration with learning management systems to provide more comprehensive support for degree program transitions, and developing methods to better understand and cater to the diverse needs of students in different disciplines or academic levels. By addressing these areas, the framework can play a pivotal role in helping students succeed in their higher education journeys, regardless of the complexities they may encounter along the way.

Furthermore, several future studies are possible including (1) a case study with students for a semester-long course in multiple fields/departments, (2) integrating the VirtualTA with learning management systems, (3) creating course content assistance and search, quiz and flash card mechanisms, (4) integration to other main-stream communication channels, (5) personalized communication to the pace and language for the student's level of understanding, (6) improving the system accuracy and performance by fine-tuning the model on bigger datasets, (7) developing methods to understand different question asking techniques, and (8) integrating necessary course information directly from learning management systems.

Sajja *et al. Int J Educ Technol High Educ*      (2023) 20:42

Page 19 of 24

# Appendix

See Table 6.

**Table 6** List of competency questions for the syllabus knowledge graph

| Category | Syllabus element | Competency questions |
|---|---|---|
| Course information | Course name | What is the name of the course? |
| | | What is the course's title? |
| | | What is the course's name? |
| | Course number | What is the course number? |
| | | What is the number of the course? |
| | | What is the course ID? |
| | Credit hours | How many credit hours is this course worth? |
| | | What are the credit hours for this course? |
| | | What is the credit hour value of this course? |
| | Location and class times | Where is the location of the lecture? |
| | | Where will the lecture be held? |
| | | What will be the lectures location? |
| | | What times are the lectures? |
| | | When do the lectures take place? |
| | Prerequisites/Co-requisites | Does this course have any prerequisites/corequisites? |
| | | Are there any prerequisites or corequisites for this course? |
| | | Is there a pre-requisite or a co-requisite for this course? |
| Faculty information | Name | What is the name of the instructor/professor? |
| | | What is the name of the professor or instructor? |
| | | What is the professor's name? |
| | Contact information | How can I contact the instructor? |
| | | What is the best way for me to get in touch with the instructor? |
| | | What is the contact information of the Instructor? |
| | Office location | Where is the office of the instructor? |
| | | Where can I find the instructor's office? |
| | | What is the location of the instructor's office? |
| | Office hours | When are the instructor's office hours? |
| | | What are the instructor's office hours? |
| | | How long is the instructor's office open? |

**Table 6** (continued)

| Category | Syllabus element | Competency questions |
|---|---|---|
| TA/teaching assistant information | Name | What is the name of the TA/Teaching Assistant? |
| | | What is the TA's/Teaching Assistant's name? |
| | | What is the name of the TA? |
| | Contact information | How can I contact the TA? |
| | | What is the best way for me to get in touch with the TA? |
| | | What is the TA's email? |
| | Office location | Where is the office of the TA? |
| | | Where can I find the TA's office? |
| | | What is the location of the TA's office? |
| | Office hours | When are the TA's office hours? |
| | | What are the TA's office hours? |
| | | How long is the TA's office open? |
| Course goals and/or objectives | Course objectives | What are the course objectives? |
| | | What are the goals of the course? |
| | | What are the objectives of this course? |
| | Expectations from the course | What are the expectations from the course? |
| | | What can you anticipate from the course? |
| | | What are the course's objectives? |
| Course calendar | Due dates | When are Assignments due? |
| | | When are the deadlines for assignments? |
| | | When must assignments be completed? |
| | Assignment dates | When are assignments released? |
| | | When do assignments become available? |
| | | When will assignments be made available? |
| Attendance and classroom behavior | Attendance policy | What is the attendance policy for this course? |
| | | What is the course's attendance policy? |
| | | What is this class's attendance policy? |
| | Expected classroom behavior | Is there any expected classroom behavior? |
| | | Is there any standard classroom conduct? |
| | | Is there any anticipated behavior in the classroom? |

**Table 6** (continued)

| Category | Syllabus element | Competency questions |
| --- | --- | --- |
| Grading and assignments | Grading criteria | What is the grading criteria for this course? |
| | | What is the grading breakdown? |
| | | What are the grading guidelines for this course? |
| | Tentative exam schedule | How many exams does this course have? |
| | | When are the exams? |
| | | When will the exams be held? |
| | | When is the final exam? |
| Instructional materials | Textbooks | What textbook does this course use? |
| | | What textbook is being used in this course? |
| | | What textbook is used in this class? |
| | Other required materials for the course | What are the other materials this course uses? |
| | | What additional materials does this course use? |
| | | What additional materials does this course make use of? |
| Policies | Late assignments | What happens when I turn in my assignment late? |
| | | What happens if I submit my assignment after the deadline? |
| | | What happens if I miss the deadline for submitting my assignment? |
| | Academic dishonesty | What is the cheating policy? |
| | | What are the policies on Academic dishonesty? |
| | | What are the policies against academic misconduct? |
| | Disability statement | What are the accommodations provided for a student with a disability? |
| | | What kind of accommodations are made for a student who has a disability? |
| | Freedom of speech | How does this class accommodate Free speech and Expression? |
| | | How does this class allow for free speech and expression? |
| | Makeup policy | What is the makeup policy for assignments? |
| | | What is the policy on assignment makeup? |
| | Mental health resources | Are there any resources for mental health? |
| | | Is there any information about mental health? |
| | | Are there any mental health resources? |
| | Absences for religious holy days | How do absences for holy days be considered? |
| | | How are absences from holy days taken into account? |

Sajja *et al. Int J Educ Technol High Educ*     (2023) 20:42

Page 22 of 24

## Declarations

**Competing interests**
The authors declare that they have no competing interests.

### References

Abdelhamid, S., & Katz, A. (2020). Using Chatbots as Smart Teaching Assistants for First-Year Engineering Students. 2020 First-Year Engineering Experience. https://peer.asee.org/using-chatbots-as-smart-teaching-assistants-for-first-year-engineering-students

Abdul-Wahab, S. A., Salem, N. M., Yetilmezsoy, K., & Fadlallah, S. O. (2019). Students' Reluctance to Attend Office Hours: Reasons and Suggested Solutions. *Journal of Educational and Psychological Studies [JEPS], 13*(4), 715–732.

Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. (2019). Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly, 36*(2), 358–367.

Benedetto, L., Cremonesi, P., & Parenti, M. (2019). A Virtual Teaching Assistant for Personalized Learning (arXiv:1902.09289). arXiv. https://doi.org/10.48550/arXiv.1902.09289

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. In International conference on internet science (pp. 377–392). Springer.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language Models are Few-Shot Learners (arXiv:2005.14165). arXiv. https://doi.org/10.48550/arXiv.2005.14165

Ceha, J., Lee, K. J., Nilsen, E., Goh, J., & Law, E. (2021). Can a Humorous Conversational Agent Enhance Learning Experience and Outcomes?. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1–14).

Chopra, S., Gianforte, R., & Sholar, J. (2016). Meet Percy: The CS 221 Teaching Assistant Chatbot. *ACM Transactions on Graphics, 1*(1), 8.

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. W., Sutton, C., Gehrmann, S., Schuh, P., Shi, K., Tsvyashchenko, S., Maynez, J., Rao, A., Barnes, P., Tay, Y., Shazeer, N., Prabhakaran, V., … Fiedel, N. (2022). PaLM: Scaling Language Modeling with Pathways (arXiv:2204.02311). arXiv. https://doi.org/10.48550/arXiv.2204.02311

Chuaphan, A., Yoon, H. J., & Chung, S. (2021). A TA-Like Chatbot Application: ATOB. In Proceedings of the EDSIG Conference ISSN (Vol. 2473, p. 4901).

Ciampa, K. (2014). Learning in a mobile age: An investigation of student motivation. *Journal of Computer Assisted Learning, 30*(1), 82–96.

Demiray, B. Z., Sit, M., & Demir, I. (2021). DEM Super-Resolution with EfficientNetV2. arXiv preprint arXiv:2109.09661.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (arXiv:1810.04805). arXiv. https://doi.org/10.48550/arXiv.1810.04805

Dibitonto, M., Leszczynska, K., Tazzi, F., & Medaglia, C. M. (2018). Chatbot in a campus environment: design of LiSA, a virtual assistant to help students in their university life. In International Conference on Human-Computer Interaction (pp. 103–116). Springer.

du Boulay, B. (2016). Artificial Intelligence as an effective classroom assistant. *IEEE Intelligent Systems, 31*(6), 76–81. https://doi.org/10.1109/MIS.2016.93

Duberry, J., & Hamidi, S. (2021). Contrasted media frames of AI during the COVID-19 pandemic: a content analysis of US and European newspapers. *Online Information Review., 45*, 758.

Ewing, G., & Demir, I. (2021). An ethical decision-making framework with serious gaming: A smart water case study on flooding. *Journal of Hydroinformatics, 23*(3), 466–482.

Ewing, G., Mantilla, R., Krajewski, W., & Demir, I. (2022). Interactive hydrological modelling and simulation on client-side web systems: An educational case study. *Journal of Hydroinformatics, 24*(6), 1194–1206.

GATech, Georgia Institute of Technology Commission on Creating the Next in Education. (2018). Deliberate Innovation, Lifetime Education. Retrieved from http://www.provost.gatech.edu/commission-creating-next-education

Gautam, A., Sit, M., & Demir, I. (2022). Realistic river image synthesis using deep generative adversarial networks. *Frontiers in Water, 4*, 10.

Hart, K. (2019). How a chatbot boosted graduation rates at Georgia State. Retrieved from https://www.axios.com/chatbot-colleges-academic-performance-ff45cb79-1fe1-485c-ae24-aa88d088c067.html

He, H., & Ma, Y. (2013). *Imbalanced learning: foundations, algorithms, and applications*. John Wiley & Sons.

Helmer, J. (2019). Carnegie Mellon shares $100 million in teaching research and resources. University Business. Retrieved from https://universitybusiness.com/carnegie-mellon-shares-100-million-in-teaching-research-and-resources

Hess, K., Falkofske, J., & Young, B. (2007). Syllabus Template Development for Online Course Success. 3.

Ho, F. (2018). TA-bot: An AI agent as a Teaching Assistant using Google's Conversational Technologies. https://doi.org/10.13140/RG.2.2.34344.06408

Hobert, S. (2019). Say hello to 'coding tutor'! design and evaluation of a chatbot-based learning system supporting students to learn to program.

Howitz, W. J., Thane, T. A., Frey, T. L., Wang, X. S., Gonzales, J. C., Tretbar, C. A., Seith, D. D., Saluga, S. J., Lam, S., Nguyen, M. M., Tieu, P., Link, R. D., & Edwards, K. D. (2020). Online in no time: Design and implementation of a remote learning first quarter general chemistry laboratory and second quarter organic chemistry laboratory. *Journal of Chemical Education, 97*(9), 2624–2634.

Hwang, G. J., & Chang, C. Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, 1–14.

Iglesias-Pradas, S., Hernández-García, Á., Chaparro-Peláez, J., & Prieto, J. L. (2021). Emergency remote teaching and students' academic performance in higher education during the COVID-19 pandemic: A case study. *Computers in Human Behavior, 119*, 106713.

IMARC Group. (2019). Intelligent Virtual Assistant Market: Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2019–2024 (Report ID: 4775648). https://www.researchandmarkets.com/reports/4775648/intelligent-virtual-assistant-market-global

Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. *Communications of the ACM, 57*(3), 78–85.

Konicek-Moran, R., & Keeley, P. (2015). *Teaching for conceptual understanding in science*. NSTA Press, National Science Teachers Association.

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2020). ALBERT: A Lite BERT for Self-supervised Learning of Language Representations (arXiv:1909.11942). arXiv. https://doi.org/10.48550/arXiv.1909.11942

Li, Z., & Demir, I. (2023). U-net-based semantic classification for flood extent extraction using SAR imagery and GEE platform: A case study for 2019 central US flooding. *Science of the Total Environment, 869*, 161757.

Lin, F., & Chan, C. K. (2018). Examining the role of computer-supported knowledge-building discourse in epistemic and conceptual understanding. *Journal of Computer Assisted Learning, 34*(5), 567–579.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach (arXiv:1907.11692). arXiv. https://doi.org/10.48550/arXiv.1907.11692

Mainstay. (2021). Georgia State University supports every student with personalized text messaging. Retrieved from https://mainstay.com/case-study/how-georgia-state-university-supports-every-student-with-personalized-text-messaging/

Mind Commerce. (2019). Virtual Personal Assistants (VPA) and Smart Speaker Market: Artificial Intelligence Enabled Smart Advisers, Intelligent Agents, and VPA Devices 2019–2024. https://mindcommerce.com/reports/virtual-personal-assistant-market/

Miner, A. S., Laranjo, L., & Kocaballi, A. B. (2020). Chatbots in the fight against the COVID-19 pandemic. *NPJ Digital Medicine, 3*(1), 1–4.

Mirzajani, H., Mahmud, R., Ayub, A. F. M., & Wong, S. L. (2016). Teachers' acceptance of ICT and its integration in the classroom. Quality Assurance in Education.

Mounsey, R., Vandehey, M., & Diekhoff, G. (2013). Working and non-working university students: Anxiety, depression, and grade point average. *College Student Journal, 47*(2), 379–389.

OpenAI. (2022). Optimizing Language Models for Dialogue [Web log post]. Retrieved April 10, 2023, from https://openai.com/blog/chatgpt

Palace, C. (2021). *Federicocotogno/mscbot* [CSS]. https://github.com/federicocotogno/mscbot (Original work published 2022)

Passman, T., & Green, R. A. (2009). Start with the Syllabus: Universal Design from the Top. *Journal of Access Services, 6*(1–2), 48–58. https://doi.org/10.1080/15367960802247916

Pawlik, Ł., Płaza, M., Deniziak, S., & Boksa, E. (2022). A method for improving bot effectiveness by recognising implicit customer intent in contact centre conversations. *Speech Communication, 143*, 33–45.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. 24.

Ramirez, C. E., Sermet, Y., Molkenthin, F., & Demir, I. (2022). HydroLang: An open-source web-based programming framework for hydrological sciences. *Environmental Modelling & Software, 157*, 105525.

Ranavare, S. S., & Kamath, R. S. (2020). Artificial intelligence based chatbot for placement activity at college using DialogFlow. *Our Heritage, 68*(30), 10.

Ruan, S., Jiang, L., Xu, J., Tham, B. J. K., Qiu, Z., Zhu, Y., Murnane, E.L., Brunskill, E., & Landay, J. A. (2019). Quizbot: A dialogue-based adaptive learning system for factual knowledge. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–13).

Schoemaker, P.J. and Tetlock, P.E. (2017). Building a more intelligent enterprise. MIT Sloan Management Review.

Seeroo, O., & Bekaroo, G. (2021). Enhancing Student Support via the Application of a Voice User Interface System: Insights on User Experience. In Proceedings of the International Conference on Artificial Intelligence and its Applications (pp. 1–6).

Sermet, Y., & Demir, I. (2018). An intelligent system on knowledge generation and communication about flooding. *Environmental Modelling & Software, 108*, 51–60.

Sermet, Y., & Demir, I. (2020). Virtual and augmented reality applications for environmental science education and training. In New Perspectives on Virtual and Augmented Reality (pp. 261–275). Routledge.

Sermet, Y., & Demir, I. (2021). A semantic web framework for automated smart assistants: A case study for public health. *Big Data and Cognitive Computing, 5*(4), 57.

Sermet, Y., & Demir, I. (2022). GeospatialVR: A web-based virtual reality framework for collaborative environmental simulations. *Computers & Geosciences, 159*, 105010.

Sit, M., Demiray, B., & Demir, I. (2021). Short-term hourly streamflow prediction with graph convolutional gru networks. arXiv preprint arXiv:2107.07039

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management, 45*, 427–437.

Song, D., Oh, E. Y., & Hong, H. (2022). The impact of teaching simulation using student chatbots with different attitudes on preservice teachers' efficacy. *Educational Technology & Society, 25*(3), 46–59.

TechNavio. (2018). Artificial Intelligence Market in the US Education Sector 2018-2022 (Report No. 4613290). Retrieved from https://www.researchandmarkets.com/research/pc2rfv/artificial.

USACE. (2019). Virtual Assistant Technology Holds Promise for USACE. Engineer Update—the Official Newsletter of the U.S. Army Corps of Engineers, 8 November. Alexandria, Virginia.

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *The Canadian Journal of Psychiatry, 64*(7), 456–464.

Wagner, J. L., Smith, K. J., Johnson, C., Hilaire, M. L., & Medina, M. S. (2022). Best practices in syllabus design. *American Journal of Pharmaceutical Education*. https://doi.org/10.5688/ajpe8995

Williamson, B., Eynon, R., & Potter, J. (2020). Pandemic politics, pedagogies and practices: Digital technologies and distance education during the coronavirus emergency. *Learning, Media and Technology*, 107–114.

Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet?-A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence*, 4.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2020). XLNet: Generalized Autoregressive Pretraining for Language Understanding (arXiv:1906.08237). arXiv. https://doi.org/10.48550/arXiv.1906.08237

Yeşilköy, Ö. B., Yeşilköy, S., Sermet, M. Y., & Demir, I. (2022). A comprehensive review of ontologies in the hydrology towards guiding next generation artificial intelligence applications. *EarthArxiv*. https://doi.org/10.31223/X5SS74

Zoroayka, S. (2018). Design and implementation of a chatbot in online higher education settings. *Issues in Information Systems*. https://doi.org/10.48009/4_iis_2018_44-52

Zylich, B., Viola, A., Toggerson, B., Al-Hariri, L., & Lan, A. (2020). Exploring automated question answering methods for teaching assistance. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial intelligence in education* (pp. 610–622). Springer International Publishing. https://doi.org/10.1007/978-3-030-52237-7_49

## Publisher's Note