Q&AI: An AI-Infused Question Answering System for Online Learning

PATARA TRIRAT (20205642), KAIST, South Korea
SUBIN KIM (20180109), KAIST, South Korea
YOOSANG YOON (20160416), KAIST, South Korea
JISU KIM (20180162), KAIST, South Korea

Due to the massive increase of users (i.e., students) in online learning platforms, it is incredibly challenging for instructors, even with teaching assistants, to answer all the questions asked by students during their learning. Therefore, a well-performed and user-friendly question answering system is imperative to guarantee smooth and efficient learning experiences. To this end, this project proposes a new design and user experience for artificial intelligence (AI) based question answering system in the open online learning environment, such as educational channels on Youtube. We design our system to be sub-components of a bigger platform as an AI bot to spontaneously answer the student's question and provide a corresponding explanation. Our design can be easily embedded into any existing online learning platform. During the evaluation session, even though the testers seemed not familiar with the user interface at first, it is still a promising design given its responsiveness and explainability that make the study more efficient with an uninterrupted learning experience.

CCS Concepts: • **Information systems → Question answering**; • **Human-centered computing → Interface design prototyping**.

Additional Key Words and Phrases: Question Answering System, Online Learning, Human-AI Interaction

ACM Reference Format:
Patara Trirat (20205642), Subin Kim (20180109), Yoosang Yoon (20160416), and Jisu Kim (20180162). 2022. Q&AI: An AI-Infused Question Answering System for Online Learning. In *KAIST AI for Smart Life, December 20, 2021, Daejeon, South Korea.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

With a large number of users, mostly millions, who are using open learning platforms, such as Youtube and the Massive Open Online Course (MOOC) websites for their study, instructors cannot answer all the questions from their students. Consequently, discussion forums are leveraged to facilitate peer-to-peer learning. However, this approach has the potential of misleading each other with inaccurate information as well as the lack of responsibility and participation, thereby contributing to duplicate questions, interrupted learning experiences, and eventually, early dropouts [7]. A few studies developed an artificial intelligence (AI) based question answering (QA) model to mitigate the problems above. YouEDU [3] presented an approach that automatically detects confusion in MOOC forum posts and recommends video clips as answers in a specific course forum. Xiao-Shih Hsu and Huang [4] is the intelligent educational question answering bot made of Natural Language Processing (NLP) processes and a Random Forest model to answer learners' questions. While these approaches provide some answers to the problem, they primarily target students who actively participate in the course discussion forum. However, those students who use forums are a tiny part of all students

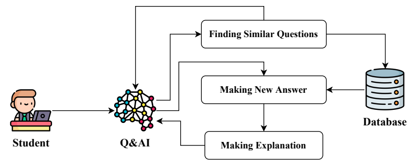


Fig. 1. Idea schema of the proposed Q&AI.

learning the course. Moreover, recent studies [2, 5] have also confirmed the significance of an automatic question-answering system with similarity-based methods to reduce the response time. Despite their responsiveness improvement, they have limited their studies in the discussion forum and did not design with an integrated framework for a seamless learning experience. In addition to the separated forum-based approaches, Wambsganss et al. [6] proposed ongoing design principles for a conversational QA system. It works similar to the automatic chatbot in the Facebook business page environment but is specialized for educational settings. Lastly, from another point of view, Aflalo [1] revealed that encouraging students to generate questions can improve students' ability to cope with questions, reduce test anxiety, and increase productivity in group learning.

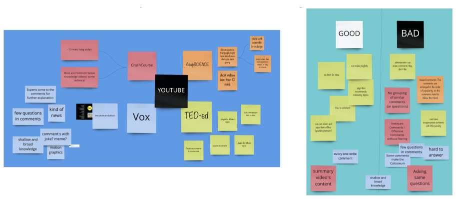
1.1 Main Idea

Noticeably, all previous studies made their QA systems in the dedicated settings. This setting causes an interrupted learning experience because students need to go back and forth between different screens. In this project, to provide a seamless learning experience and a high-quality answer on the open online learning platform, we extend prior research by incorporating the idea of an AI-powered automatic QA system with an integrated novel interface and interaction designs. We also adopt the idea of similarity-based question finding to enhance the speed and encourage peer-to-peer learning. However, we additionally design the system to have a voting mechanism so that only high-quality answers will remain. To realize the seamless learning experience, we designed a QA system, called Q&AI, as an AI bot embedded into the main video learning screen so that students can continue studying while making a question. Ultimately, to iteratively improve the quality of the answer and gain the users' trust, we also provide an interactive explanation of the given answer by mapping each question word to its corresponding words in the answer. To the best of our knowledge, this kind of explanation has never been done by other studies on the QA system for online learning environments. Notably, since we design our Q&AI to be sub-components of an entire user interface, it can be easily embedded or adopted by other existing learning platforms to enhance user experiences. Fig. 1 illustrates the overview idea of the proposed Q&AI.

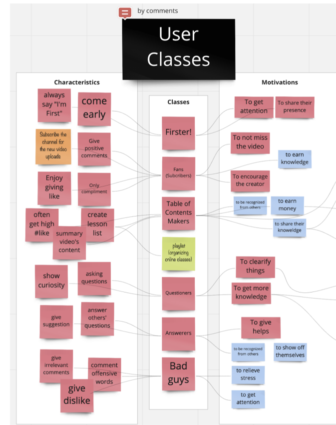
2 DESIGN PROCESS

2.1 Needfinding

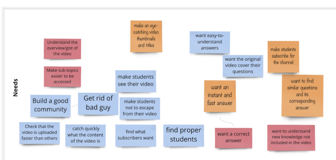
In the needfinidng phase, instead of dedicated learning platforms, e.g., edX, Coursera, or Future Learn, our team decided to study the educational channels on Youtube because we thought that these channels could better represent more comprehensive types of users (i.e., students). As presented in Fig. 2a, we selected four popular educational channels, namely CrashCourse, Vox, AsapScience, and TED-ed, for each person in our team. Accordingly, we found some good and bad sides of user behavior. As a result, we classify the users from the four channels into seven classes with their corresponding characteristics shown in Fig. 2b. Based on user behavior and their traits, we derived the various user



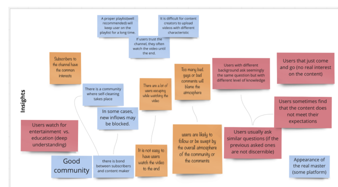
(a) Survey source and categorization.



(b) User classes.



(c) User needs.



(d) Insights.

Fig. 2. Observations from the needfinding phase.

needs, including building a good community, asking for an answer, and creating an accessible playlist. Finally, as shown in Fig. 2d, we extracted a few valuable insights by connecting the dots between what we have done so far. Consequently, we discovered that the most important and beneficial thing to the learning process is to correctly answer the users' questions in a timely manner.

2.2 Ideation

2.2.1 Point of View. We analyzed students who asked questions through comments. We were amazed to realize that they could not get the correct answer and waited a long time. Old questions were scrolled down, so many people could not see. It would be better to sort out questions that are actually related to the video's content and to separate questions from actual comments. We came up with an example persona: student A, a college student who tries to study AI through youtube, is subscribed to channel B. One day, he asked questions to the channel through the comment, but the question was scrolled down because of other comments unrelated to the video. It has been a week since he asked the question, but it is still not answered. We developed our ideation based on this persona.

2.2.2 How Might We. During the class activity, we came up with numerous how might we questions, but the following How Might We (HMW) questions are selected:

- **HMW 1**: HMW make students get an instant and reliable answer?
- **HMW 2**: HMW group similar questions together?
- **HMW 3**: HMW motivate students to share questions and answers freely?

2.2.3 *Solution Ideas.*

For each HMW, again, we came up with various solution ideas, but only the following ideas are voted to be the most useful ones as answers to the above HMWs.

- **Idea 1**: Apply ranking system/level-up and honor mark to indicate the quality of answers while encouraging peers to share their knowledge.
- **Idea 2**: Group similar questions together based on content similarity and recommend previously asked question group if applicable or detected.
- **Idea 3**: AI-based QA system with subsystems connecting to outside knowledge-base (e.g., Wikipedia) for additionally answering questions if the in-platform videos cannot answer.

2.3 Planning

We did not make any strict plan or schedule for the design of this project. All the work was done mainly by just following the class activities. However, after the idea was finalized about the QA system, we started thinking about what kind of user interface (UI) we wanted to design and what kind of data we needed for our system. At this time, we did not have any concern on the ethical issues yet. As a result of this phase, we just prepared what to be integrated together based on the user's needs and our final goals of solving the problem of existing QA systems.

2.4 Data

About the data, we got the following results for each level according to the DIKM class activity with the goal of forming a system that can reduce the downtime of QA through comments. For the **Data** level, we needed text from users' comments and their corresponding metadata as well as video content of each video in the platform. For the next level, **Information**, we can extract the downtime of questions being answered through comments, the context and level of the questions in the comments, and the level of similarity between videos and comments. For the **Knowledge** level, we can further derive more practical knowledge. For example, the comments tend to get scrolled down because of other comments, users tend to ask similar questions in a particular topic if the previous one(s) did not get an answer, and questions that users ask tend to be similar (i.e., repetitive questions). Finally, at the highest level, the **Wisdom**, we can derive the insights. These insights include a ranking system to promote quality of peer-to-peer answers, similar question clustering for previously asked questions to reduce duplication, and external source referring, e.g., Wikipedia, for answering questions if there is no in-platform content (course video and peer answer) as an answer to ensure the instant response.

2.5 Model

The model that we propose should support interactive question-answering in an online video class. To explain in more detail, the role of AI in our model should do the following functions: (1) grouping similar questions for reducing duplication and making an answer faster, and (2) correctly answering user questions so that user can smoothly continue their learning. Then, the AI needs text from questions, corresponding answers, and auxiliary data from videos for each data. Through clustering and prediction, we desired to give the following outputs: (1) questions with their group number(s), i.e., groups of similar questions, and (2) expected answer (as a text sequence) to the user question.

The appropriate model for our AI design is the Natural Language Understanding model provided by Google or OpenAI, trained to group similar questions. BERT-based Question Answering model, trained with SQuAD 2.0 and comments data, to answer user questions. We can use such models for clustering similar questions, only to answer user questions based on text data and video contents. Also, possible side-effects could be that if the words used are similar. They are classified as the same question, Not for creating any text that is irrelevant or offensive to users.

According to evaluation metrics, the following metrics are used for task-level evaluation. (1) **Explainability** on why such an answer is given to the user question, (2) **Usability** by allowing a user to intuitively catch how to ask questions and retrieve answers without any tutorials, (3) **Trustworthiness** by allowing control on the granularity of the predicted answer, and (4) **Overall User Satisfaction** on the answer by receiving user feedback. Additionally, the model-level metrics are as follows. (1) Exact Match (EM) and F1 Score for QA Model, (2) Confidence Score for each answer, and (3) Dunn's Index (DI) and Silhouette Coefficient for Clustering Model.

2.6 Interaction

When the user asks a question through video comments, the AI analyzes the text from user questions and video content via NLU. Then, it finds and shows similar questions, if any, via clustering to answer the user. Also, it can directly give an answer via prediction to the user based on video content and user question. The roles of human users are that users choose the most similar question among the suggested similar questions and give feedback to whether it is helpful or not (e.g., Upvote or not). Specifically, AI recommends answers, and humans can choose and confirm the candidates(i.e., recommender system). The user first interacts with the system by typing a question. Then, AI recommends lists of similar questions with an answer. Accordingly, the user chooses the most similar question from the list. If there is no similar question, AI suggests the possible answer. In this process, the user evaluates the given answer whether it is helpful to them, and the user gives feedback to the given answer. If negative, the AI provides more answers within a given threshold. If positive, the user may continue with the class. Overall, users watch lectures through our service and ask questions. Our system provides a list of possible answers to the user, and the user chooses the best one. We can evaluate through user satisfaction surveys and interviews.

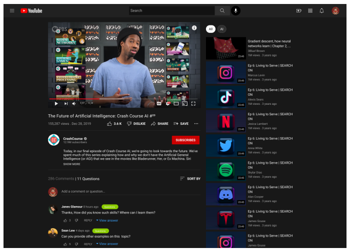
3 FINAL PROTOTYPE

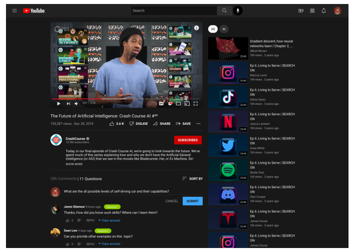
3.1 Scenario 1: Responsive QA with Interactive Explainability

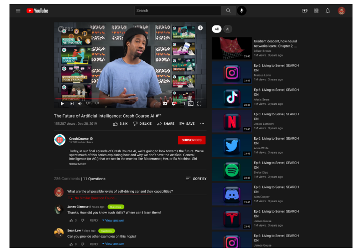
In this scenario, the task our system supports is answering the student's question. As depicted in Fig. 3, after the user enters their question, the Q&AI first analyzes the input text data to find similar questions. However, in this case, there is no similar question found. Accordingly, the system finds a new answer by extracting keywords from the user's question and video contents. Then, it generates the answer by finding the highest similarity scores between generated answers and the question. This can be done using a similarity metric, such as cosine similarity with a well-performed AI model (e.g., GPT-3) representing (embedding) the text data. After the answer is generated, the system makes an interactive explanation by utilizing a contribution score of each word to map the word in the question to the word in the answer, as shown in Fig. 3f.

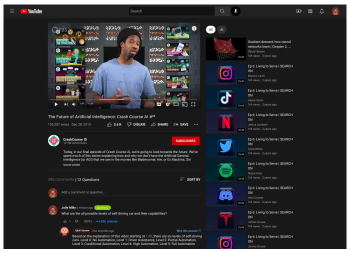
3.2 Scenario 2: Similarity-based Question Finding for Better Answers

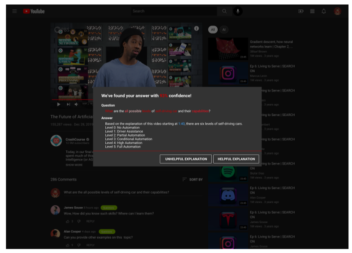
The second scenario is about enhancing answer quality and reducing question duplication that can hinder the chance of other students getting their answers. As shown in Fig. 4, similar to the previous scenario, the system analyzes the user input text, but in this case, it can find matching questions. The Q&AI then displays the found questions to the user so that they can select which answer they want to read. After user reads the chosen answer, they can further make an upvote or decide to ask their question as a new question right away if they think that the found answers are not


(a) Main video screen.


(b) User asks a question.


(c) System processes the question.


(d) User gets an answer.


(e) After user clicks "Why this ansewr".

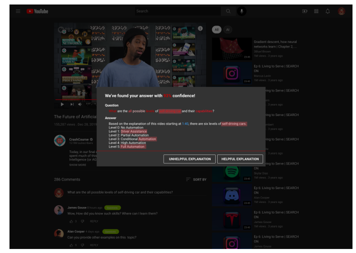
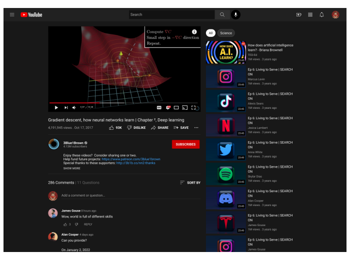
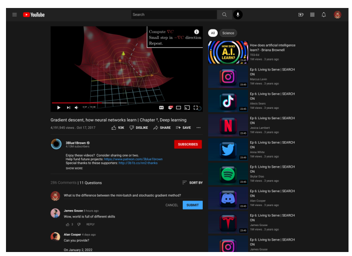
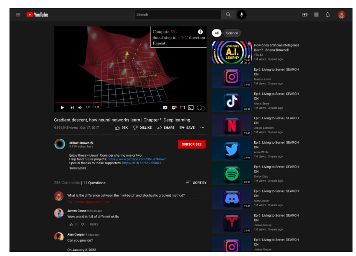
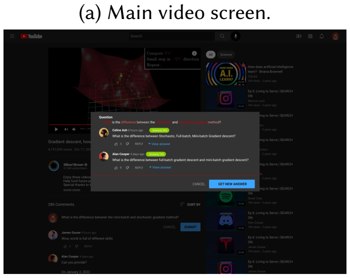

(f) User hovers on a word of the question.

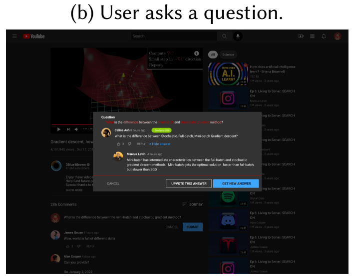
Fig. 3. Scenario 1: Asking a new question with an explainable answer.


(a) Main video screen.


(b) User asks a question.


(c) System found similar questions.


(d) System shows the similar questions.


(e) User sees an answer.

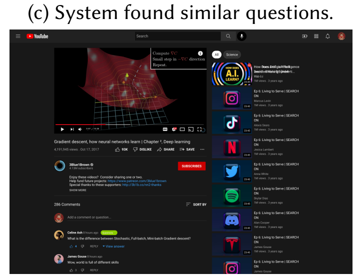

(f) User upvotes the answer.

Fig. 4. Scenario 2: System found the similar questions.

helpful or do not solve their problems. The upvote button here not only make the answered question become the top of the comment/question list to remind the user, but also a quality assurance for the top-quality answer and a signal for other users that this kind of question has already been asked several times so that they can get the answer immediately without wasting their time.

3.3 Data Used and Ethical Issues

In this project, we expect to collect the content text and metadata, as well as video transcripts from user comments and course videos, respectively. These datasets will be used to train the similarity-based question clustering and QA model. Examples of these data are: (1) **text** (e.g., "who invented convolutional neural networks?") of user's comment, (2) **metadata** (e.g., datetime: 2021/10/26 17:00:00, video_title: "Intro to AI", etc.) of each comment, and (3) **course video** (e.g., title, topic tags, duration, and captions) of each video in the platform.

Regarding the data collection process, we will use in-platform data storage system for comments and videos. Then, to make labeled subsets of the datasets, we may use the human workforce or crowdsourcing approach for creating or selecting a reasonable correct answer to each question.

Possible Data Quality Issues. Firstly, many random users can ask questions, and we do not know if it is a different person or the same person with different IDs. Therefore, we cannot directly identify the user's information. In addition, we cannot guarantee whether the given data, such as a user question, is actually complete for our QA system. The redundancy could be another possible issue, even though we propose the Q&AI to reduce the duplication of questions by design. Lastly, some noise from irrelevant comments and non-educational videos may occur and degrade the model performance.

Ethical and Privacy Concerns. The data privacy of the users could be a problem. We need permission to use random users' comments, which can be hard to achieve from every user. Some comment data may have user-sensitive information, for instance, such as the location and name. Also, it may violate the rights of the channel (or provider) if we use the data related to the video content without permission. Intellectual property is another problem because the information that AI can crawl from the internet and use for answers can be ambiguous.

Proposed Solutions. To solve the above issues, there are several potential solutions as follows. We can first ask for permission beforehand to the channels and users we want the data from, requiring consent and the terms of agreement form. We can also provide anonymity of the information and comments that we use and exclude user-sensitive information from the analysis/training of the model, as it may not affect the final accuracy of the system. In addition, we should try to answer questions based mainly on the content itself, i.e., sole text data.

4 EVALUATION AND DISCUSSION

4.1 Evaluation

As a first lesson during the testing session, we found that even though we adopted the native Youtube UI with our embedded design, users could not intuitively understand our design and were curious why we needed a modal design to answer a question. They suggested making a bot as a standard reply like on Youtube. To this end, we incorporate their advice with our first idea. As a result, shown in the previous section, we make the main answer be the standard reply while keeping the modal design for interactive explanation and similar question exploration. We believe that it is better to separate different functionality with different UI, and thus the user can focus on each component easily. In addition, some parts of our design were ambiguous because they lacked labels, such as comment vs. question list. Here, we decided to add labels for each component with its corresponding number of items.

Regarding the upvote answer feature, it was a helpful vs. unhelpful button, which was not meaningful enough for users to understand why an answer is helpful or not and what the effect is after clicking the buttons. Therefore, we changed the (un)helpful buttons to the upvote button since this is the already wide-known keyword and is more understandable. Also, about the explanation of the answer, it did not have the interaction part, but users told us that they could not clearly understand what the intensity of color was trying to convey. Hence, we added the interaction part to show the connection between question and answer. Lastly, as a critical lesson learned during this session, we realized that using a more familiar UI instead of making an entirely new design is better for users to understand.

As future improvements, we have thought about two things. First, since the UI of comments and questions may still not be obvious enough to grasp at a glance, we are considering redesigning this component. The second one is about designing the notification workflow for the unsolved questions that users provide feedback on.

4.2 Discussion

The biggest problem in non-real-time non-face-to-face education is that the teacher cannot closely monitor the students' reactions. Thus, it is not easy to convey what they want to teach accurately and check whether it has been delivered well. In this sense, our Q&AI system can check whether the learning process is going well in online settings. Students can immediately receive answers, making it easier to understand the content of the class. Also, students feel free to ask questions that are difficult to ask when learning face-to-face. This data is also valuable for teachers. Information about which QA has gone through can be a clue for understanding the level of learners and where they are interested. In addition, teachers can get a hint on how to do online lectures more effectively.

The social value of Q&AI is that everyone can get a high-quality education wherever they are and whenever they want. It is related to accessibility. According to google education, because of COVID-19, online lectures have appeared, and these online lectures educate the people who are excluded from education due to safety issues. There are some worries if this kind of service is commercialized. Because it is an online lecture, students only get filtered lectures, so they only can get a limited field of knowledge, not comprehensive. To address this problem, systematic help to educate a broader area is necessary.

5 CONCLUSION

This paper presents a new design for the AI-infused question-answer system in the online learning environment. To realize our ideas into the prototypes, we went through the design process, including need-finding, ideation, planning, data, model, and interaction design. Also, we show our unique aspects by presenting the two main scenarios on how our system design can effectively answer the user questions. Lastly, the discussions from the evaluation session and potential implications are given as our concluding remarks.

REFERENCES

[1] Ester Aflalo. 2021. Students generating questions as a way of learning. *Active Learning in Higher Education* 22, 1 (2021), 63–75.

[2] Abhishek Agarwal, Nikhil Sachdeva, Raj Kamal Yadav, Vishaal Udandarao, Vrinda Mittal, Anubha Gupta, and Abhinav Mathur. 2019. EDUQA: Educational domain question answering system using conceptual network mapping. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 8137–8141.

[3] Akshay Agrawal, Jagadish Venkatraman, Shane Leonard, and Andreas Paepcke. 2015. YouEDU: Addressing Confusion in MOOC Discussion Forums by Recommending Instructional Video Clips. In *Proceedings of the 8th International Conference on Educational Data Mining, EDM 2015, Madrid, Spain, June 26-29, 2015*. 297–304. http://www.educationaldatamining.org/EDM2015/proceedings/full297-304.pdf

[4] Hao-Hsuan Hsu and Nen-Fu Huang. 2018. Xiao-Shih: The Educational Intelligent Question Answering Bot on Chinese-Based MOOCs. In *2018 17th IEEE Int. Conference on Machine Learning and Applications (ICMLA)*. IEEE, 1316–1321.

[5] Atul Sahay, Smita Gholkar, and Kavi Arya. 2019. Selection-based Question Answering of an MOOC. *arXiv preprint arXiv:1911.07629* (2019).

[6] Thiemo Wambsganss, Leonie Haas, and Matthias Soellner. 2021. Towards the Design of a Student-Centered Question-Answering System in Educational Settings. (2021).

[7] Diyi Yang, Robert E. Kraut, and Carolyn Penstein Rosé. 2016. Exploring the Effect of Student Confusion in Massive Open Online Courses. In *Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016, Raleigh, North Carolina, USA, June 29 - July 2, 2016*. 8. http://www.educationaldatamining.org/EDM2016/proceedings/paper_303.pdf