PRELIMINARY PREPRINT VERSION: DO NOT CITE
The AAAI Digital Library will contain the published
version some time after the conference.

ChEdBot: Designing a Domain-Specific Conversational Agent in a Simulational Learning Environment Using LLMs

Andreas Martin[1], Charuta Pande[1], Hans Friedrich Witschel[1], Judith Mathez[2]
[1]FHNW University of Applied Sciences and Arts Northwestern Switzerland, Intelligent Information Systems Research Group, Riggenbachstrasse 16, 4600 Olten, Switzerland
[2]FHNW University of Applied Sciences and Arts Northwestern Switzerland, Institut Weiterbildung und Beratung, Bahnhofstrasse 6, 5210 Windisch, Switzerland
(andreas.martin|charuta.pande|hansfriedrich.witschel|judith.mathez)@fhnw.ch

Abstract

We propose conversational agents as a means to simulate expert interviews, integrated into a simulational learning environment: ChEdventure. Designing and developing conversational agents using the existing tools and frameworks requires technical knowledge and a considerable learning curve. Recently, LLMs are being leveraged for their adaptability to different domains and their ability to perform various tasks in a natural, human-like conversational style. In this work, we explore if LLMs can help educators easily create conversational agents for their individual teaching goals. We propose a generalized template-based approach using LLMs that can instantiate conversational agents as an integrable component of teaching and learning activities. We evaluate our approach using prototypes generated from this template and identify guidelines to improve the experience of educators.

Introduction

Conversational Agents (CAs) are frequently used in technology-assisted teaching and learning. Several studies have highlighted the benefits as well as the challenges of integrating CAs in different fields of education (Hwang and Chang 2023; Sandu and Gide 2019).

A challenge that is less addressed in the literature is the knowledge of technology, tools, and frameworks needed for developing any CA. Most examples of CAs in education have been developed either by interdisciplinary teams with knowledge of technology or by teams that know how to effectively use CA development tools or frameworks (Bahja, Hammad, and Butt 2020; Burkhard et al. 2022). We would like to explore how educators can develop CAs without undergoing technical training or being entirely dependent on people with technical knowledge.

As a part of the ChEdventure research project (based on the proposed approach in (Pande, Witschel, and Martin 2021)), we are developing a simulation-based learning environment where CAs will play the role of domain experts; we call them ChEdBots. Students should be able to interview ChEdBots and collect e.g., requirements to complete an assignment. In this way, students will be able to develop interviewing skills. It will also be possible to design ChEdBots that provide confusing or contradictory information (as stakeholders in the real world sometimes do) so that students will also learn to deal with such challenges.

Most educational CAs follow an agent-driven conversation where the CA provides pre-defined responses to the users' queries or takes the user on a pre-defined learning path (Kuhail et al. 2023). Such a design puts limitations on interactivity as well as user experience and will not be a suitable choice for the design of ChEdBot. The desired interaction style to simulate interviews is a user-driven conversation in which a student freely asks questions to a ChEdBot.

However, to design user-driven conversations, CAs need to be trained on huge amounts of data to be able to interpret different formulations of user queries. Pre-trained LLMs like GPTx are known to perform very well with the interpretation of user queries. Compared to traditional CA development approaches, it is easier and quicker to create a conversational system using LLMs as it involves intuitive techniques like providing instructions and knowledge in natural language, known as "prompts" (Goel et al. 2023; Meskó 2023). However, LLM-based CAs may not provide accurate domain-specific responses as they have a tendency to "hallucinate" (Rawte, Sheth, and Das 2023). In the case of ChEdBot, the responses should not only be domain-specific but also assignment-specific. Thus, a pre-trained LLM cannot be applied to ChEdBot off-the-shelf. On the other hand, as mentioned above, the somewhat random or creative nature of responses generated by LLMs can be seen as a "feature" when it comes to settings where students need to learn to deal with confusing or inconsistent answers of stakeholders.

We take the ChEdventure learning environment as an application scenario and explore how lecturers can quickly and easily design CAs to fit their teaching goals and assignment objectives without letting technology become a barrier. Thus, in this work, we address the following question:
How can LLMs be leveraged such that lecturers (educators) are empowered to design and develop an educational CA by investing minimum time and effort in learning CA development technology?

This paper is organized as follows: the motivation for our work is discussed in Section 1, followed by a discussion of the state-of-the-art related to our work in Section 2. In Section 3, we describe the design and configuration details of ChEdBot and the evaluation of the ChEdBot prototype in Section 4. In Section 5, we discuss how domain and common sense knowledge can be integrated into ChEdBot and its influence on learning and learners' experiences. Section 6 describes the next steps in the ChEdventure research project, followed by our conclusion in Section 7.

Related Work

In this section, we discuss the existing tools and technology used for developing conversational agents including LLMs. We then discuss some work that has been done to include common sense and domain knowledge in LLMs as well as some applications in the education domain.

Technology Burden of Conversational Agent Development

Common techniques to develop CAs are to use a *rule-based design* or a *probabilistic approach* to learning conversations from data using machine learning algorithms (Harms et al. 2018). The former approach usually follows a decision-tree-like structure that directs the conversation and is intuitive but rigid and time-consuming. Additionally, the rule-based approach may be augmented with natural language understanding to extract the "intents" and the "entities" from the user utterances that trigger the rules driving the conversation (Lam, Le, and Kalita 2020), also known as a *frame-based design* (Harms et al. 2018). Here, the performance of the CA is dependent on the number of training examples for intents and entities; the more training data, the better the understanding of the CA. The other approach, i.e., the probabilistic approach is entirely data-driven, and the conversational agent learns to predict an appropriate response based on the past conversations it was trained on. However, a data-driven approach to training a CA, especially in a closed domain CA, has limitations due to its requirement of a large amount of domain-specific training data (Su et al. 2020). Advanced abilities include interaction using text as well as voice, recognition of sentiments and emotions in user utterances, and information retrieval from knowledge sources.

Several CA development tools and frameworks exist that support features to create rule-based dialogues, text- and voice-enabled CAs as well as train a custom NLU component by defining custom intents and entities, e.g., Dialogflow [1], Voiceflow [2], Rasa [3]. More recently, these platforms have introduced features enabling the use of LLMs within conversations [4] and to leverage external sources of knowledge [5]. Although many of these tools are low-code platforms, i.e., do not require extensive programming knowledge, there is a considerable learning curve involved in learning tool usage as well as the technical concepts behind the features. Moreover, most CA development tools focus on CA technology and do not integrate features of creating educational CAs, thus leading to a separation between the two activities of defining educational requirements and developing a corresponding CA. In ChEdventure, we propose an integrated no-code environment dedicated to educational requirements, leveraging LLMs.

Common Sense and Domain Knowledge in LLMs

Despite a seemingly human-like ability to converse and respond to queries, there is a lack of consensus in literature on whether LLMs have reasoning abilities and common sense, and to what extent (Wei et al. 2022). To effectively respond to queries as humans do, machines require both common sense and domain knowledge. A widely used technique to impart common sense and domain knowledge to LLMs is to use prompting and in-context learning (Huang and Chang 2022). It is possible to nudge and restrict LLMS towards domain-specific responses using techniques like "zero/few-shot learning" and "chain of thought prompting". Kan et al. (2023) demonstrate how knowledge-aware prompts can help improve the performance of vision-language models. Baek, Aji, and Saffari (2023) use prompting to augment LLMs for domain-specific question answering using a knowledge graph. We find prompting techniques, generalized as prompt templates [6], appropriate for ChEdBot as they can be intuitively grasped by lecturers. Alternatively, pre-trained LLMs can be "fine-tuned" using domain-specific datasets; however, this technique is known to be expensive and time-consuming (Huang and Chang 2022).

LLMs in Education

Since the introduction of ChatGPT, several studies have reviewed the potential of using LLMs in various applications (Fraiwan and Khasawneh 2023). In the education domain, LLMs are prominently used in supporting various teacher and student activities. Tu et al. (2023) explore the potential of LLMs for teaching data science in an interactive way. Abd-Alrazaq et al. (2023) explore the potential in medical education to identify issues such as plagiarism, misinformation, overreliance, inequity, privacy, and copyright issues. Sridhar et al. (2023) use LLMs for curriculum design by generating learning objectives.

Although several studies see the benefits of LLMs, many are skeptical, especially in the domain of education. Larsén (2023) and Orenstrakh et al. (2023) argue how LLMs could be misused in exams and propose mitigation strategies. LLM-based tools like ChatGPT are readily accessible and have the ability to generate text in the form of reports, essays, papers etc. Jungherr (2023) warn that the use of LLMs for academic tasks like paper writing can affect skill development of students.

Many studies on conversational agents for educational simulation and their effects on teaching can be found in the literature (Sullivan, Albright, and Khalid 2021; Brubacher et al. 2015; Cybulski, Parker, and Segrave 2006; Debnath and Spoletini 2020; Göbl, Kriglstein, and Hlavacs 2021; Schnurr and MacLeod 2021; Scarpellini and Lim 2020), but an investigation on how LLMs can be utilized for this purpose is limited. Van Buren (2023) and Xu et al. (2023) use

[1] https://cloud.google.com/dialogflow

[2] https://www.voiceflow.com/

[3] https://rasa.com/product/rasa-platform/

[4] https://www.voiceflow.com/blog/ai-assist-llms-voiceflow

[5] https://www.voiceflow.com/product/knowledge-base

[6] https://www.pinecone.io/learn/series/langchain/langchain-prompt-templates/

LLMs to simulate expert personas. We use a similar approach to create unique personas for the stakeholders as well as inject an element of humor into the conversations, but we go beyond persona design to simulate an interview-like dialog with multiple turns between the stakeholder and the student. Chen et al. (2023) have simulated dialogues between a psychiatrist and patients using LLM to investigate potential in medical education.

ChEdBot Design

ChEdBot, a key component of the ChEdVenture didactics research project, represents a novel approach in educational technology. It leverages a multi-agent system, enabling university lecturers to create virtual business stakeholders. These agents are designed to interact with students, simulating real-world business workshop experiences. The system ingeniously combines a frame-based dialogue system for structured conversation management and domain knowledge application, with a Large Language Model (LLM)-based system for dynamic, persona-driven interactions.

Frame-based Dialogue System Configuration

The backbone of ChEdBot's interaction model is its frame-based dialogue system. This system is meticulously structured to handle various aspects of a conversation:

1. Utterances: These are user queries that initiate the dialogue process. The students will need to plan which utterances should be used (e.g. in the form of an interview guideline).
2. Context Activation: This feature allows the system to maintain coherence in dialogue by understanding the current context of the conversation, e.g. the previous questions asked by the students.
3. Intent Recognition: Using natural language processing, the system interprets the intent behind user utterances, ensuring relevant responses. Intents are linked to responses that include domain knowledge. For instance, a CA may represent an employee in a company. When the students ask a question about the tasks that the employee has to do, this may activate a known intent (let us call it "tasks") to which a response containing a list of tasks is linked. This response will be passed (as a prompt) to an LLM, together with the original question, to generate an adapted / varied response.
4. Dialogue Variables: These variables store information extracted from user utterances, aiding in personalized and context-aware responses. Entity recognition, while not currently implemented, is considered for future integration to enhance the system's understanding of user utterances.
5. Response Generation: Leveraging an LLM, the system generates responses that not only align with the domain knowledge, but are also dynamically tailored to the conversation.

LLM-Agent-Based System Configuration

The incorporation of an LLM-based system introduces advanced capabilities to ChEdBot:

1. Message History/Memory: This component retains a record of previous interactions, enabling the system to provide contextually rich responses.
2. Prompt Templates with Variables: These templates, customizable with various variables, guide the LLM in generating responses that are both relevant and diverse.
3. LLM Temperature Setting: This feature controls the level of randomness in responses, balancing between predictable and creative outputs.

ChEdBot Configuration

The configurability of ChEdBot is a cornerstone of its design, empowering lecturers to tailor the system to their specific educational needs. Central to this customization process is Figure 1, which presents various tables that guide the configuration of the multi-agent system.

In Table a., of Figure 1, lecturers have the flexibility to define a range of agents, each with a distinct name and role, mimicking the diverse stakeholders in business scenarios. This table also facilitates the setting of parameters that are crucial for the functioning of the agents, such as the size of the message history or memory, which is instrumental for the LLM in generating context-aware responses. Moreover, lecturers can establish a default context for interactions, particularly useful when an intent is not directly referenced at the conversation's outset. Additionally, the table includes a feature for setting a threshold for intent recognition, ensuring a fallback mechanism activates when the recognition probability falls below this threshold.

The personalization of the agents is further enhanced in Table b., where lecturers can define distinct personas for each agent. This feature is vital in creating a realistic and engaging interaction experience, as it allows each agent to embody a unique character, reflecting the varied personalities encountered in real-world business environments.

In Table c., the system's flexibility is evident in the provision for lecturers to craft and modify prompt templates. These templates are not just static scripts; they are dynamic constructs that can incorporate multiple variables, including those from the dialogue system configuration and system-specific elements like the last utterance, predefined responses, and the current persona. This level of customization enables the creation of responses that are not only relevant to the dialogue but also rich in variety and depth. For example, the "temperature" parameter can help in configuring the randomness or creativity in ChEdBot's responses, which we consider as an important feature to simulate real-world ambiguities in communication due to fragmented knowledge and individual communication styles.

Table d., in Figure 1 of the ChEdBot system, plays a pivotal role in shaping the interactive dialogue experience. This table is where lecturers craft the intricate details of the dialogue flow, ensuring each agent not only understands and responds to user inputs accurately, but also maintains a coherent and engaging narrative throughout the interaction.

At the core of functionality is the meticulous process of intent identification. Lecturers define specific intents, along with exemplary utterances that trigger these intents. This

Agent	Name	Description	History	Context	Threshold
Demo	Demo	This is a demo agent.	0	Start	0.8
Kryptowerk	Kryptowerk	Case Kryptowerk AG by Ullrich et al. (2020)	0	Smart	0.1

a. Define agents for the stakeholders

Agent	Role	Persona
Demo	CEO	Spricht wie eine strenge Geschäftsführerin.
Demo	Pirat	Spricht wie ein Pirat.
Demo	IT	Spricht wie ein IT-Nerd.
Demo	Neutral	Spricht ganz neutral.

b. Define persona and role for the agents

Agent	Prompt	Template	Temperature
Demo	Utterance	{History} ## System: {Persona} ## Human: {Utterance} Response: {Response} ## Passe die Message "Response" an "Human" an. ## AI:	0.2
Demo	Continuation	System: {Persona} ## Response: {Response} ## Passe die Message "Response" an. ## AI:	0.2

c. Available Prompt Templates

Agent	Intent	Context	Utterance	Response	Continuation	Contextualisation	Role	Prompt	Fallback	Variable
Demo	1	Start	Hallo	Wie ist dein Name?	2	Start	CEO		3	Name
Demo	2	Start		Hallo {Name}, was möchtest du machen?	3	Start	CEO		3	
Demo	3	Start		Möchtest du A oder B erledigen?		Option	CEO	Continuation	3	
Demo	4	Option	Ich möchte A erledigen	Ja, A ist super!	6	Task A	Pirat	Utterance	3	
Demo	5	Option	Ich möchte B erledigen	Auch B ist gut.	10	Task B	IT	Utterance	3	
Demo	6	Task A		Was hast du für eine Frage zu A?		Task A	Pirat	Continuation	3	
Demo	7	Task A	Ich möchte was über XYZ wissen.	Ah ja, über XYZ kann ich dir folgendes erzählen: Bla bla XYZ bla bla. Task A		Task A	Pirat	Utterance	3	
Demo	8	Task A	Ich möchte was über ABC wissen.	Ah ja, über ABC kann ich dir folgendes erzählen: Bla bla ABC bla bla. Task A		Task A	Pirat	Utterance	3	
Demo	9		Tschüss	Bye	3		Neutral		3	
Demo	10	Task B		Was hast du für eine Frage zu B?		Task B	IT		13	
Demo	11	Task B	Ich möchte was über XYZ wissen.	Ah ja, über XYZ kann ich dir folgendes erzählen: Bla bla XYZ bla bla. Task B		Task B	IT	Utterance	13	
Demo	12	Task B	Ich möchte was über ABC wissen.	Ah ja, über ABC kann ich dir folgendes erzählen: Bla bla ABC bla bla. Task B		Task B	IT	Utterance	13	
Demo	13	Task B		Sorry {Name}, das habe ich nicht verstanden.	10	Task B	IT	Continuation	13	

d. Define dialogs for the agents

Figure 1: ChEdBot Design Template for Lecturers

careful mapping is essential for the system to recognize and categorize user queries accurately. The precision in this process ensures that each agent responds in a manner that is contextually appropriate and aligned with the intended educational objectives.

Another significant aspect is the concept of context activation, or contextualization. Here, lecturers specify how an agent should shift its conversational focus based on recognized intents. This feature is crucial for maintaining a logical and seamless flow of dialogues. It allows for dynamic shifts in conversation, mirroring real-life interactions where topics and focus areas change naturally over time. By defining these context shifts, lecturers can simulate realistic and complex dialogue scenarios, enhancing the learning experience for students.

Moreover, the system allows for the specification of hard-coded continuations in the dialogue. This aspect of the configuration is akin to setting a predetermined path in the conversation, where specific responses from an agent lead to a new set of intents being activated. This mechanism can be particularly useful in guiding the conversation towards specific learning outcomes or ensuring that certain educational points are addressed in the dialogue.

In addition to these features, it also facilitates the activation of corresponding roles, reflecting different personas for each agent. This capability enables each agent to not just deliver content, but to do so with a personality that aligns with its designated role. The use of personas adds depth to the interactions, making the simulation more immersive and relatable for students.

Furthermore, the table allows lecturers to define variables that can be used within dialogues. These variables can store and utilize information provided by users during interactions, enabling personalized and dynamic responses that adapt to the specifics of the conversation (the simplest example is to store the name of the user and use it in the conversation). This customization extends to the activation of prompt templates, designed to utilize the LLM effectively. Lecturers can choose specific templates to be triggered under certain conditions, ensuring that the LLM-generated responses are not only relevant but also creatively varied, adding an element of unpredictability and motivation to the interactions. The example demonstrated in Figure 1 shows a dialog in German, thus highlighting the multilingual ability of this ap-

Agent	Intent	Context	Utterance	Response	Continuation	Contextualisation	Role	Prompt	Fallback	Variable
Kryptowerk	14	Smart	Hallo	Wie ist Ihr Name?	15	Smart	CEO		20	UserName
Kryptowerk	15	Smart		Hallo {UserName}, mein Name ist Herrmann Smart, CEO der Kryptowerk AG, die seit 1998 Software- und Hardwarekomponenten für Computer und Smartphones herstellt.	16	Smart	CEO		20	
Kryptowerk	16	Smart		Mit über 2000 Mitarbeiter*innen und internationaler Präsenz sind wir stetig am Wachsen.	17	Smart	CEO		20	
Kryptowerk	17	Smart		Daher haben wir Sie beauftragt, insbesondere im Bereich Human Ressource, unsere Prozesse zu modellieren und zu optimieren.	18	Smart	CEO		20	
Kryptowerk	18	Smart		Daher beauftragen wir Sie zur Aufzeichnung der Prozesse des Kompetenzmanagement.	19	Smart	CEO		20	
Kryptowerk	19	Smart		Unsere Mitarbeiter*innen am Standort Deutschland werden Ihnen alle Fragen beantworten.	20	Smart	CEO		20	
Kryptowerk	20	Interview		Mit wem möchtesten Sie ein Intreview führen?	21	Interview	CEO		20	
Kryptowerk	21	Interview		Abteilungsleiter HR?		Interview	CEO		20	
Kryptowerk	22	Interview	Abteilungsleiter HR	Abteilungsleiter HR ist Franz Hubert.	23	Interview	CEO		20	
Kryptowerk	23	Interview		Herr Hubert kennt den gesamten Prozess bestens, hat jedoch keine detaillierten Kenntnisse.		HR	CEO		20	
Kryptowerk	24	HR	Was bezweckt der Prozess?	Kompetenzmanagement macht die Skills und Fähigkeiten der Mitarbeiter*innen sichtbar.	25	HR	HR		20	
Kryptowerk	25	HR		Dadurch können wir gezielt fördern und erkennen, wo neue Talente benötigt werden, um effizient und wirtschaftlich zu arbeiten.		HR	HR		20	
Kryptowerk	26	HR	Tschüss	Ich hoffe ich konnte Ihnen was über das Kompetenzmanagement sagen.	20	Interview	HR		20	

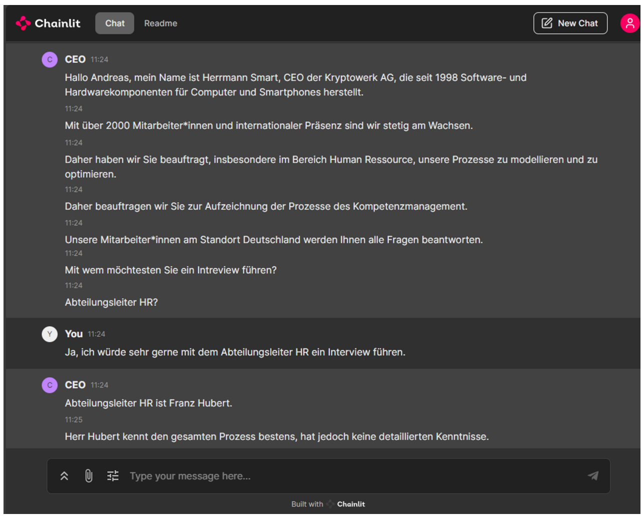

Figure 2: Overview of final implementation of FAQ system

proach; the same template can be used to design a conversational agent in multiple languages.

In essence, the orchestrator of the dialogue experience in ChEdBot provides lecturers with a powerful tool to design intricate, responsive, and engaging conversational agents. Through this comprehensive configuration, ChEdBot stands as a sophisticated platform for creating interactive learning environments that mimic real-world interactions, thereby enriching the educational experience for students. Figure 2 shows how the dialogs defined in the ChEdBot template have been instantiated into a live interaction using the tool Chainlit [7].

ChEdBot Evaluation

A preliminary evaluation of the ChEdBot prototype was carried out with three lecturers. Some pre-conditions that should be met before starting to work on the ChEdBot design template are as follows:

1. The lecturers should start by defining the learning goals of the assignment. This helps in defining the dialogues such that each learning goal is met.
2. The chosen teaching case should be appropriate for ChEdventure didactics environment, i.e., it should have multiple business stakeholders, it should be possible to simulate it as a workshop where the stakeholders provide requirements of the assignment through interviews, and the final solution should involve modeling tasks to realize the requirements, e.g., business processes, use cases, data models, etc.
3. The lecturers should get familiarized with the template and the terminologies used. Although technical knowledge is not required, it is necessary to understand the meaning of technical terms like intent, context, temperature etc., and their impact on the resulting CA.

Taking the pre-requisites into account, each lecturer identified a teaching case from their respective modules. The teaching cases contained text-based description of the background, context and problems in the case scenario. Some cases also included the details of the stakeholders and their interviews. When these details are not included in a case, the lecturers had to identify them themselves. The lecturers were verbally explained the details of the ChEdBot template, i.e., the tables shown in Figure 1. They were also provided with a filled template as a simple example of a Demo ChEdBot. Using this information, the lecturers were asked to fill out the template by identifying the personas for the stakeholders, intents, contextualization, and variables for the dialogues depending on the respective teaching goals.

The exercise of creating ChEdBot prototypes was carried out as a workshop where a pedagogical expert guided the lecturers from a didactical viewpoint whereas a Conversational AI expert was available to answer any technical queries. The observations, queries, and corresponding expert responses were documented to develop guidelines for future ChEdBot design exercises. Table 1 summarizes the feedback of the lecturers from the perspectives of both conversation design and technical implementation with the corresponding action points.

The results of the ChEdBot prototype evaluation will be used to design further workshops with lecturers from FHNW School of Business and carry out a comprehensive evaluation of ChEdventure didactics environment. In this evaluation, besides ChEdBot, the lecturers will also configure the simulation environment where ChEdBots will be situated.

Incorporation of Domain and Commonsense Knowledge

In the design of ChEdBot, a key aspect is the integration of domain-specific and commonsense knowledge to enrich the educational experience. This integration enhances the authenticity and relevance of interactions, making them not only informative but also reflective of real-world scenarios.

1. **Embedding Domain Knowledge**:
   The primary method of embedding domain knowledge is through intent recognition and response generation processes. Lecturers can infuse responses with detailed, accurate information relevant to each recognized intent, ensuring that every interaction is both informative and contextually appropriate. For instance, in a business simulation, responses may include insights into market trends, financial strategies, or operational management, depending on the user's query.

[7] https://docs.chainlit.io/get-started/overview

Observation and Feedback	Identified Action
It is important to control the dialogue such that only appropriate questions from the students should be answered.	Design *Fallback intents* to handle irrelevant, generic and out of context questions
Since the dialogue is open-ended, it should lead students to asking questions that have appropriate responses. Otherwise, it is likely that the students lose interest in the conversation.	The dialogue design should ensure that the answer to one question gives a hint to the next question.
It is possible that the students do not ask all the relevant questions that are required to complete the assignment.	The lecturers should explicitly include the *scope* of the ChEdBot in dialogues, e.g., all important points the ChEdBot can answer as a response to an intent that will help students in formulating questions.
Handovers between stakeholders should be considered in dialogue design.	Handovers involve communication between multiple agents. Where required, pass technical information like *context* and *variables* during the handover.
Intents should be unique and a *context* is required at the start of the conversation.	Include as a mandatory guideline.
Technical requirements can be further simplified.	Provide pre-configured ChEdBot templates with example dialogues for reference.
Dialogue design and generation can be simplified.	Use AI support (e.g., ChatGPT) to generate dialogues from the description of teaching cases.

Table 1: ChEdBot Prototype Evaluation Results

2. **Contextualization with Domain-Specific Scenarios**:
   Context activation is tailored to reflect realistic scenarios within the educational domain – going beyond e.g. the often-used case teaching. By modeling contexts after actual situations, such as business negotiations or scientific experiments, the system applies commonsense knowledge in a domain-relevant manner. This approach enhances the learning process and prepares students for practical applications in their field. The chance to conduct actual conversations with varied responses makes the simulations based on our CAs much more practice-oriented and real than e.g. in-class case discussions can achieve.
3. **Personalization with Domain-Relevant Variables**:
   The system's ability to use variables for generating personalized responses is further enhanced by incorporating domain-specific data. These variables hold information relevant to the subject, allowing for responses that are not just contextually appropriate, but also deeply ingrained with domain knowledge.
4. **Customized Prompt Templates**:
   Prompt templates are crafted to include domain-specific information, guiding the LLM in generating responses that are both informative and relevant to the subject. This customization ensures that the LLM's vast knowledge base is effectively harnessed to provide accurate and domain-appropriate information.
5. **Persona Development with Commonsense Knowledge**:
   The development of personas involves the incorporation of commonsense knowledge which comes from the teaching case chosen by the lecturers. This approach ensures that each persona not only represents a specific role within the domain but also interacts in a way that is consistent with everyday understanding of the role in the context of the chosen case. The personas thus become more relatable and realistic, enhancing the interactive learning experience.
6. **LLM Training and Fine-Tuning**:
   As a potential possibility, to further enhance the system's capability, the LLM integrated into ChEdBot can be specifically trained or fine-tuned using datasets rich in domain knowledge and commonsense information. This specialized training will allow the LLM to better understand and respond to queries in a manner that is both accurate and contextually relevant.
7. **Continuous Learning through Feedback**:
   As a long-term approach, we foresee implementing a feedback loop where interactions are periodically reviewed and updated by domain experts to ensure that the system continually evolves. This process will help maintain the accuracy and relevance of the content, aligning it with the latest developments and understandings in the field.

Future Work

When we generalise the findings from Table 1, three main areas for future work emerge:

1. **Simplification**: It will be necessary to design a user-friendly environment for teachers to design CAs – where a number of simplifications are desirable (see the last two rows in Table 1.
2. **Risk mitigation**: as suggested by the first two rows in Table 1, lecturers see risks of students e.g. asking inappropriate questions or getting stuck in a conversation. Further risks may emerge when the resulting CAs are applied in a real educational setting.

3. **Advanced contextualisation**: while contextualisation is already foreseen in our current solution, further advancements become necessary when complex simulations are to be designed, involving e.g. several CAs.

Conclusions

In this paper, we have discussed opportunities for realistic educational simulation of interviews / business workshops from which students can learn e.g. interview strategies and the analysis of sometimes confusing or contradicting statements of conversational agents. These opportunities arise because of the possibility to prompt large language models in a way that mimics real conversation, providing variety, persona personalities, and occasional hard-to-interpret statements.

Our approach ensures that these realistic features are complemented with the possibility for teachers to control what information / knowledge conversational agents will reveal to the students, to manage context adequately and to configure personas with suitable traits. Starting from a first implementation of our approach, we have been able to gather some first insights and identify areas for future work as described in the previous section. In the current part, we focused on the tasks from the educators' point of view. However, in the final part of the project, ChEdBots will be integrated into classroom teaching and evaluated to assess their effect on students' learning in comparison to traditional case-based teaching methods.

Acknowledgments

This work has been funded by the FHNW Hochschullehre 2025 under the project number HSLE2025 TP3 ChEdventure.

References

Abd-Alrazaq, A.; AlSaad, R.; Alhuwail, D.; Ahmed, A.; Healy, P. M.; Latifi, S.; Aziz, S.; Damseh, R.; Alrazak, S. A.; Sheikh, J.; et al. 2023. Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions. *JMIR Medical Education*, 9(1): e48291.

Baek, J.; Aji, A. F.; and Saffari, A. 2023. Knowledge-Augmented Language Model Prompting for Zero-Shot Knowledge Graph Question Answering. *arXiv preprint arXiv:2306.04136*.

Bahja, M.; Hammad, R.; and Butt, G. 2020. A user-centric framework for educational chatbots design and development. In *HCI International 2020-Late Breaking Papers: Multimodality and Intelligence: 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22*, 32–43. Springer.

Brubacher, S. P.; Powell, M.; Skouteris, H.; and Guadagno, B. 2015. The effects of e-simulation interview training on teachers' use of open-ended questions. *Child abuse & neglect*, 43: 95–103.

Burkhard, M.; Seufert, S.; Cetto, M.; and Handschuh, S. 2022. Educational Chatbots for Collaborative Learning: Results of a Design Experiment in a Middle School. *International Association for Development of the Information Society*.

Chen, S.; Wu, M.; Zhu, K. Q.; Lan, K.; Zhang, Z.; and Cui, L. 2023. LLM-empowered Chatbots for Psychiatrist and Patient Simulation: Application and Evaluation. *arXiv preprint arXiv:2305.13614*.

Cybulski, J. L.; Parker, C.; and Segrave, S. 2006. Touch it, feel it and experience it: Developing professional IS skills using interview-style experiential simulations. *ACIS 2006 Proceedings*, 85.

Debnath, S.; and Spoletini, P. 2020. Designing a virtual client for requirements elicitation interviews. In *Requirements Engineering: Foundation for Software Quality: 26th International Working Conference, REFSQ 2020, Pisa, Italy, March 24–27, 2020, Proceedings 26*, 160–166. Springer.

Fraiwan, M.; and Khasawneh, N. 2023. A Review of ChatGPT Applications in Education, Marketing, Software Engineering, and Healthcare: Benefits, Drawbacks, and Research Directions. *arXiv preprint arXiv:2305.00237*.

Göbl, B.; Kriglstein, S.; and Hlavacs, H. 2021. Conversational Interfaces in Serious Games: Identifying Potentials and Future Research Directions based on a Systematic Literature Review. *CSEDU (1)*, 108–115.

Goel, R.; Hidey, C.; Mohammad, H. R.; Muddireddy, P. R.; and Liu, F. 2023. LLM-based Task-oriented Dialog System with Few-shot Retrieval Augmentation.

Harms, J.-G.; Kucherbaev, P.; Bozzon, A.; and Houben, G.-J. 2018. Approaches for dialog management in conversational agents. *IEEE Internet Computing*, 23(2): 13–22.

Huang, J.; and Chang, K. C.-C. 2022. Towards reasoning in large language models: A survey. *arXiv preprint arXiv:2212.10403*.

Hwang, G.-J.; and Chang, C.-Y. 2023. A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, 31(7): 4099–4112.

Jungherr, A. 2023. Using ChatGPT and Other Large Language Model (LLM) Applications for Academic Paper Assignments.

Kan, B.; Wang, T.; Lu, W.; Zhen, X.; Guan, W.; and Zheng, F. 2023. Knowledge-aware prompt tuning for generalizable vision-language models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 15670–15680.

Kuhail, M. A.; Alturki, N.; Alramlawi, S.; and Alhejori, K. 2023. Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1): 973–1018.

Lam, K. N.; Le, N. N.; and Kalita, J. 2020. Building a Chatbot on a Closed Domain using RASA. In *Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval*, 144–148.

Larsén, S. 2023. Creating Large Language Model Resistant Exams: Guidelines and Strategies. *arXiv preprint arXiv:2304.12203*.

Meskó, B. 2023. Prompt engineering as an important emerging skill for medical professionals: tutorial. *Journal of Medical Internet Research*, 25: e50638.

Orenstrakh, M. S.; Karnalim, O.; Suarez, C. A.; and Liut, M. 2023. Detecting llm-generated text in computing education: A comparative study for chatgpt cases. *arXiv preprint arXiv:2307.07411*.

Pande, C.; Witschel, H.-F.; and Martin, A. 2021. ChEdventure-A Chatbot-based Educational Adventure Game for Modeling Tasks in Information Systems. In *CONVERSATIONS 2021–the 5th International Workshop on Chatbot Research*, 1–9.

Rawte, V.; Sheth, A.; and Das, A. 2023. A survey of hallucination in large foundation models. *arXiv preprint arXiv:2309.05922*.

Sandu, N.; and Gide, E. 2019. Adoption of AI-Chatbots to enhance student learning experience in higher education in India. In *2019 18th International Conference on Information Technology Based Higher Education and Training (ITHET)*, 1–5. IEEE.

Scarpellini, I.; and Lim, Y. 2020. Role-based design of conversational agents: Approach and tools. In *HCI International 2020–Late Breaking Posters: 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I 22*, 366–375. Springer.

Schnurr, M.; and MacLeod, A. 2021. *Simulations and student learning*. University of Toronto Press.

Sridhar, P.; Doyle, A.; Agarwal, A.; Bogart, C.; Savelka, J.; and Sakr, M. 2023. Harnessing llms in curricular design: Using gpt-4 to support authoring of learning objectives. *arXiv preprint arXiv:2306.17459*.

Su, H.; Shen, X.; Xiao, Z.; Zhang, Z.; Chang, E.; Zhang, C.; Niu, C.; and Zhou, J. 2020. Moviechats: Chat like humans in a closed domain. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, 6605–6619.

Sullivan, A.; Albright, G.; and Khalid, N. 2021. Impact of a Virtual Role-play Simulation in Teaching Motivational Interviewing Communication Strategies to Occupational Therapy Students for Readiness in Conducting Screening and Brief Interventions. *Journal of Higher Education Theory and Practice*, 21(2): 39–49.

Tu, X.; Zou, J.; Su, W. J.; and Zhang, L. 2023. What Should Data Science Education Do with Large Language Models? *arXiv preprint arXiv:2307.02792*.

Van Buren, D. 2023. Guided scenarios with simulated expert personae: a remarkable strategy to perform cognitive work. *arXiv preprint arXiv:2306.03104*.

Wei, J.; Tay, Y.; Bommasani, R.; Raffel, C.; Zoph, B.; Borgeaud, S.; Yogatama, D.; Bosma, M.; Zhou, D.; Metzler, D.; et al. 2022. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*.

Xu, B.; Yang, A.; Lin, J.; Wang, Q.; Zhou, C.; Zhang, Y.; and Mao, Z. 2023. ExpertPrompting: Instructing Large Language Models to be Distinguished Experts. *arXiv preprint arXiv:2305.14688*.