A PROTOTYPE OF A CHATBOT FOR EVALUATING AND REFINING STUDENT STARTUP IDEAS USING A LARGE LANGUAGE MODEL

Joseph Benjamin Ilagan
Ateneo de Manila University
Quezon City, Philippines
jbilagan@ateneo.edu

Jose Ramon Ilagan
Ateneo de Manila University
Quezon City, Philippines
jrilagan@ateneo.edu

May 20, 2023

ABSTRACT

Assessing the soundness of business models is a critical skill for aspiring entrepreneurs and is an essential part of entrepreneurship education. However, evaluating business models can be time-consuming, costly, and subjective. This study describes the design and the prototype of a chatbot as a conversational intelligent tutoring system that assesses and gives feedback on business model soundness using natural language processing techniques and GPT-3.5, a large language model (LLM) trained by OpenAI, to help student co-founders learn and refine their startup ideas. Our method involves indexing articles and rubrics for evaluating technology startup pitches by extracting word embeddings via the OpenAI API. The chatbot accepts descriptions of startup businesses from student co-founders through a Telegram chatbot, and these are formatted as prompts and then fed into GPT-3.5. The responses are formulated by GPT-3 based on another set of prompts instructing the bot to give feedback from three virtual panelists: 1) a harsh judge, 2) a neutral expert, and 3) an optimistic investor.

1 Introduction

Entrepreneurship education must be flexible and adaptive, but studies show that higher education does not have the needed flexibility [1]. Traditional teaching approaches to entrepreneurship have been based on business management education, which is inflexible and ineffective [2] and not interesting [3]. A startup is a temporary organization searching for a business model [4]. This definition of a startup is consistent with the working definition of entrepreneurship as a process in which firms search for, discover and exploit new profit opportunities by engaging in arbitrage or innovation activities [5]. Unlike established and traditional businesses, for startups, many unknowns about customers, markets, partners, and technology–essentially elements of the business model itself– must be discovered iteratively and incrementally. This paper is part of a larger set of studies using simulation for entrepreneurship education [6]. Educational software, known as computer-assisted instruction (CAI), is of low-quality [7] . An attempt to improve CAIs is in the form of Intelligent Computer-Aided Instruction (ICAI) [7], which, early on, was rules-based. As entrepreneurial education is non-linear, rules-based approaches may not be as fit as needed. Aside from starting and running a business while enrolled in university, recent entrepreneurship education and entrepreneurial experiential learning attempts involve the use of computer simulations to compress learning cycles, reduce time and cost [8], and model and illustrate how agent behaviors and interactions in a complex environment involving entrepreneurial ecosystems. At some point, startups will need to pitch their business ideas and models to panelists consisting of practitioners in technology, finance, investments, regulations, and various domains or industries. As it is expensive and infeasible to invite panelists frequently, simulating them to learn how to do proper pitching will be a promising option. Generative artificial intelligence (GAI) is an unsupervised or partially supervised machine learning framework that generates content using probability and statistics based on existing digital content such as text, video, images, and audio through training examples, thus learning their patterns and distribution [9]. A large language model (LLM) is a GAI and mathematical model of the statistical distribution of tokens in the vast public corpus of human-generated text [10]. From the training, LLMs can produce human-like language [11]. The tokens involved include words, parts of words, or individual characters, including punctuation marks. They are generative because we can sample and ask them questions [10]. GPT (Generative Pre-trained Transformer), an LLM-based system, is designed to generate or statistically-predict sequences of words, code, or other data, starting from a source input called the prompt [12]. GPT is based on a deep neural network architecture called a transformer [13, 14, 15], which trains large amounts of publicly-available data in parallel. They can exhibit creativity in writing from a paragraph to a full article convincingly on almost any topic [9]. A conversational agent or AI assistant based on LLM is ChatGPT, an interface to GPT [16]. The use of ChatGPT as chatbots in education has also been reported in blog posts and social media fields [17]. Educators can use ChatGPT to create interactive quizzes, lesson plans, and educational materials.

1.1 Objectives

This study describes an interface prototype to a technology startup coach simulator implemented as a chatbot. It discusses the methodology, architecture, and design based on a chatbot-based conversational intelligent tutoring system using a large language model (LLM) backend. The startup coach simulator allows students to formulate and get feedback on their startup idea pitch that discusses the business model, validation of their product-market fit, estimation of market potential, and the formulation of initial financial models.

1.2 Research Questions

What must be considered when developing and validating as virtual panelists simulator chatbot in the form of a conversational intelligent tutoring system to evaluate and provide valuable feedback on business model descriptions presented by student co-founders?
RQ1. How might we validate the efficacy of the proposed conversational virtual panelists in simulating the feedback provided by real-life panelists (harsh judge, neutral expert, and optimistic investor) based on test datasets of business model descriptions?
RQ2. How could the designed conversational intelligent tutoring system handle the complexity and variability of business model descriptions provided by potential student co-founders regarding understanding and giving relevant feedback?

2 Review of Related Work

2.1 Conversational Intelligent Tutoring Systems

Intelligent Tutoring Systems (ITS) [7] simulate human tutors to help students, especially the struggling ones, to improve their learning [18] with personalized, step-by-step guidance [19]. Unlike CAI, which usually marks a student's answer immediately, an ITS waits until the student has submitted a solution. It then marks individual steps as correct or incorrect. If incorrect, it conducts a debriefing, which discusses individual steps with the student [20]. Conversational Intelligent Tutoring Systems (CITS) are e-learning systems that deliver tutorial content through discussion, asking and answering questions, identifying gaps in knowledge, and providing feedback in the natural language [21]. They use natural language to cover concepts, break down the learning material into conversations, ask and answer questions, determine knowledge gaps, and provide contextual feedback and corrective interventions [21]. Examples of CITS software are AutoTutor [22] and Betty's Brain [23].

2.2 Generative AI, LLMs, and Conversational Interfaces

LLMs can solve a variety of natural language processing (NLP) tasks *zero-shot*, without relying on any training data for a given task simply conditioning the model on appropriate prompts [24, 25]. Related terms to zero-shot are *in-context learning* (ICL) and *few-shot learning* [26], which, in the context of LLMs, is the ability to learn from limited examples [24, 27]. *Chain-of-thought* (CoT) prompting induces LLMs to generate intermediate reasoning steps before answering [25]. *ChatGPT* [16] is a model trained to interact with GPT conversationally. It is itself trained on GPT-3.5 [25, 17] and through *reinforcement learning through human feedback*(RLHF). RLHF involves three steps: training a language model with supervised learning, collecting comparison data based on human preferences and training a reward model, and optimizing the language model against the reward model using reinforcement learning [25]. Prompting is much easier and cheaper than fine-tuning the whole model, especially if you only have a dozen training examples or cases [28, 29, 30]. Numeric representations of words as vectors [31] have been used for semantic search in LLMs [32]. The idea is that an input string or query is compared with an expected question. Tools to make it easier to work with documents for Q&A types of systems have emerged. They use numeric representations or embeddings and chain-of-thought prompting field [33] for language understanding [34, 33]. One open-source project utilizing these

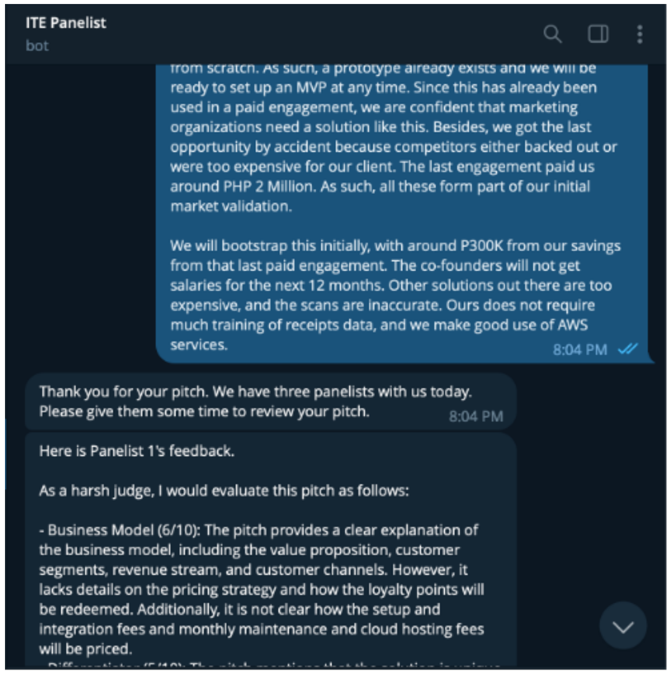

Figure 1: Screengrab of the Telegram virtual pitch session between student co-founder and virtual panelists.

techniques is Langchain [35]. This study will use a combination of prompt-based few-shot ICL and get the standard answers to the closest cosine similarity of pre-defined questions for the CITS portion.

3 Approach

The prototype's architecture involves indexing articles and rubrics for evaluating technology startup pitches by extracting word embeddings via the OpenAI API. The chatbot accepts descriptions of startup businesses from student co-founders through a Telegram chatbot, and these are formatted as prompts and then fed into GPT-3.5. The responses are formulated by GPT-3 based on another set of prompts instructing the bot to give feedback from three virtual panelists: 1) a harsh judge, 2) a neutral expert, and 3) an optimistic investor. The panelist chatbot uses a simple three-tier (i.e., database, HTTP API server, and user interface) web architecture. Telegram, the instant messaging app chosen as the user interface, sends events to FastAPI, the Python HTTP API server, through an HTTP webhook endpoint. When FastAPI receives user prompts through the webhook, it formats the prompts, sends them to gpt-3.5-turbo through the LangChain library, and sends the user the gpt-3.5-turbo response by calling the Telegram bot's API. The embeddings were stored in a vector database (originally ChromaDB, later PostgreSQL with the PGVector extension) and made available to LangChain. The prototype employs a three-step process: collecting articles, extracting text embeddings (or "indexing"), and testing prompting techniques iteratively [36] to simulate conversation with the student. LangChain's "question-answering" chains were used to induce gpt-3.5-turbo to act as a startup panelist. Ideas from YCombinator (e.g., founder-market fit, known viable startup business models, how startup competition panelists evaluate pitches, common mistakes to avoid) were encoded into text files converted into embeddings by text-embedding-ada-002. Additional text prompts were used to describe each of the three judges to influence how they would give feedback and rate the startup business model pitches. Figure 1 below shows a screen grab of the session dialogue to show the truncated startup pitch on top and a portion of the harsh judge's response below.

4 Discussion

RQ1. How might we validate the efficacy of the proposed conversational virtual panelists in simulating the feedback provided by real-life panelists (harsh judge, neutral expert, and optimistic investor) based on test datasets of business model descriptions? The intent is to have student startups use the virtual panelist simulator with their startup pitches and answer a few questions involving usefulness of the feedback. The typical qualitative questions would be the following: 1) How useful was the feedback for your startup? 2) What gaps did you realize your pitch had?

3) What changes will you incorporate in your next iteration? 4) Do you have suggestions on what to improve in the simulator?
RQ2. How could the designed conversational intelligent tutoring system handle the complexity and variability of business model descriptions provided by potential student co-founders regarding understanding and giving relevant feedback? Through careful and iterative prompt engineering, with tools like Langchain, LLMs will allow interaction with the virtual panelists through natural language. Startup co-founders will be able to deliver their pitch, while the LLM back-end will help the virtual panelists pick up the important points as instructed in the hidden prompts in the backend.

5 Conclusion

While testing the prototype with a few startup volunteers, the system seems ready for the next stage, involving validating and testing more formally with student startup co-founders. The next step for this study is to invite several student startups, have them enter the description of their startups as if they were pitching to the panelists, and provide feedback to the researchers as to the helpfulness of the chatbot in refining their business model.

References

[1] Harry Matlay and Jay Mitra. Entrepreneurship and learning: the double act in the triple helix. *The International Journal of Entrepreneurship and Innovation*, 3(1):7–16, 2002. Publisher: SAGE Publications Sage UK: London, England.

[2] Alain Fayolle and Benoit Gailly. From craft to science: Teaching models and learning processes in entrepreneurship education. *Journal of European Industrial Training*, 32(7):569–593, January 2008. Publisher: Emerald Group Publishing Limited.

[3] Mengze Wei. The Design of Intelligent Tutoring Systems Using College Students' Innovation and Entrepreneurship Education under the Background of Online Teaching. *Wireless Communications and Mobile Computing*, 2022, 2022. Publisher: Hindawi.

[4] Steve Blank. Why the lean start-up changes everything. *Harvard Business Review*, May 2013.

[5] R. Brent Ross. Entrepreneurial behaviour in agri-food supply chains: The role of supply chain partners. *Journal on Chain and Network Science*, 11(1):19 – 30, 2011.

[6] Joseph Benjamin Ilagan. The design and use of agent-based modeling computer simulation for teaching technology entrepreneurship. 2022.

[7] John R. Anderson, C. Franklin Boyle, and Brian J. Reiser. Intelligent tutoring systems. *Science*, 228(4698):456–462, 1985. Publisher: American Association for the Advancement of Science.

[8] Fernando Almeida and Jorge Simões. Serious games in entrepreneurship education. In *Encyclopedia of Information Science and Technology, Fourth Edition*, pages 800–808. IGI Global, 2018.

[9] David Baidoo-Anu and Leticia Owusu Ansah. Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Available at SSRN 4337484*, 2023.

[10] Murray Shanahan. Talking About Large Language Models, February 2023. arXiv:2212.03551 [cs].

[11] Grant Cooper. Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *Journal of Science Education and Technology*, pages 1–9, 2023. Publisher: Springer.

[12] Luciano Floridi and Massimo Chiriatti. GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30:681–694, 2020. Publisher: Springer.

[13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[14] Md Rahaman. Can chatgpt be your friend? emergence of entrepreneurial research. *Emergence of Entrepreneurial Research (February 18, 2023)*, 2023.

[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[16] Introducing ChatGPT, 2023.

[17] Ahmed Tlili, Boulus Shehata, Michael Agyemang Adarkwah, Aras Bozkurt, Daniel T. Hickey, Ronghuai Huang, and Brighter Agyemang. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1):15, 2023. Publisher: Springer.

[18] Zhiqiang Cai, Xiangen Hu, and Arthur C. Graesser. Authoring conversational intelligent tutoring systems. In *Adaptive Instructional Systems: First International Conference, AIS 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings 21*, pages 593–603. Springer, 2019.

[19] Stephen Atlas. ChatGPT for higher education and professional development: A guide to conversational AI. 2023.

[20] Kurt VanLehn. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational psychologist*, 46(4):197–221, 2011. Publisher: Taylor & Francis.

[21] Mike Holmes, Annabel Latham, Keeley Crockett, and James D. O'Shea. Tomorrow's Learning: Involving Everyone. Learning with and about Technologies and Computing, 11th IFIP TC 3 World Conference on Computers in Education, WCCE 2017, Dublin, Ireland, July 3-6, 2017, Revised Selected Papers. *IFIP Advances in Information and Communication Technology*, 515:251–260, 2017.

[22] Sidney D'mello and Art Graesser. AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(4):1–39, 2013. Publisher: ACM New York, NY, USA.

[23] Gautam Biswas, James R. Segedy, and Kritya Bunchongchit. From design to implementation to practice a learning by teaching system: Betty's Brain. *International Journal of Artificial Intelligence in Education*, 26:350–364, 2016. Publisher: Springer.

[24] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc., 2020.

[25] Chengwei Qin, Aston Zhang, Zhuosheng Zhang, Jiaao Chen, Michihiro Yasunaga, and Diyi Yang. Is ChatGPT a General-Purpose Natural Language Processing Task Solver?, February 2023. arXiv:2302.06476 [cs].

[26] Jonathan Bragg, Arman Cohan, Kyle Lo, and Iz Beltagy. FLEX: Unifying Evaluation for Few-Shot NLP. In *Advances in Neural Information Processing Systems*, volume 34, pages 15787–15800. Curran Associates, Inc., 2021.

[27] Zihao Zhao, Eric Wallace, Shi Feng, Dan Klein, and Sameer Singh. Calibrate before use: Improving few-shot performance of language models. In *International Conference on Machine Learning*, pages 12697–12706. PMLR, 2021.

[28] Iz Beltagy, Arman Cohan, Robert Logan IV, Sewon Min, and Sameer Singh. Zero- and Few-Shot NLP with Pretrained Language Models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 32–37, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[29] Tianyu Gao, Adam Fisch, and Danqi Chen. Making Pre-trained Language Models Better Few-shot Learners, June 2021. arXiv:2012.15723 [cs].

[30] Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)*, 53(3):1–34, 2020. Publisher: ACM New York, NY, USA.

[31] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26, 2013.

[32] Japa Sai Sharath and Rekabdar Banafsheh. Question Answering over Knowledge Base using Language Model Embeddings. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, July 2020. ISSN: 2161-4407.

[33] Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. Interleaving Retrieval with Chain-of-Thought Reasoning for Knowledge-Intensive Multi-Step Questions, December 2022. arXiv:2212.10509 [cs].

[34] Simon Ott, Konstantin Hebenstreit, Valentin Liévin, Christoffer Egeberg Hother, Milad Moradi, Maximilian Mayrhauser, Robert Praas, Ole Winther, and Matthias Samwald. ThoughtSource: A central hub for large language model reasoning data, February 2023. arXiv:2301.11596 [cs].

[35] Harrison Chase. LangChain, October 2022. original-date: 2022-10-17T02:58:36Z.

[36] Boshi Wang, Xiang Deng, and Huan Sun. Iteratively Prompt Pre-trained Language Models for Chain of Thought. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 2714–2730, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.