LLM Based Embedded Value Trained ChatBot Framework for Personalized Academic Learning

Chandan Satwani[a], Vrashabh Patil, Kavya Morab, Santosh Pattar[b] and Prema T. Akkasaligar[c]

Department of Computer Science and Engineering, KLE Technological University's Dr. M. S. Sheshgiri College of Engineering and Technology, Belagavi, India

Keywords: LLM-based ChatBot, Personalized Academic Learning, Contextually Relevant Responses, Falcon-7B-Instruct Model, Data Acquisition and Processing.

Abstract: Latest developments in natural language processing and machine learning techniques have enabled the development of chatbots. However, these chatbots are generalized and not tuned for specific requirements of a particular domain. In an academic setting, both learners and teachers today require a specialized chatbot for their personalized teaching learning experience. In this regard, we propose a novel approach using Large Language Model (LLM), fine-tuned with embedded value training. It leverages contextual embeddings and semantic representation to provide tailored educational content. The experimental results demonstrate an improvement of 10%, 20%, and 30% for BERT F1, ROUGE, and BLEU scores respectively, when compared to generic ChatGPT 3.5 and Gemini AI chat applications. These results suggests effectiveness in use of academic specific chatbot in improving student engagement, comprehension and retention though personalized learning experience.

1 INTRODUCTION

In today's digital age, educational institutions are rapidly integrating technology into their teaching-learning process to enhance learning experiences. Among these technologies, chatbots powered by Large Language Models (LLMs) are gaining prominence(Yan et al., 2024). These chatbots are designed to assist in various educational activities, aiming provide immediate and precise responses to student and faculty, thereby improving the overall efficiency of academic interactions(Firth et al., 2020).

Various models and technologies have been proposed in the past to enhance chatbot capabilities. For instance, Montagna *et al.*(Montagna et al., 2023) introduced *Paperplain*, a tool aiding medical professionals in extracting relevant information from clinical papers using Named Entity Recognition (NER) models. Similarly, *Dramatron* (Mirowski et al., 2023) employs hierarchical story generation using the Chinchilla LLM, showcasing the potential of large language models in creative writing tasks. In healthcare, Omoregbe *et al.* (Omoregbe et al., 2020) developed a text-based medical diagnosis system utilizing NLP and fuzzy logic, demonstrating the adaptability of these technologies in diverse fields. These existing solutions highlight the versatility and potential of LLM-based chatbots in addressing domain-specific challenges.

Despite the advancements, several challenges persist in the implementation of LLM-based chatbots. One major issue is maintaining context over prolonged interactions, that leads to inaccurate or irrelevant responses(Florindi et al., 2024). Additionally, the performance of these chatbots heavily relies on the quality and diversity of their training data, that is often limited. These problems necessitate the development of more reliable and focused chatbot solutions tailored to specific domains, such as education (Šarčević et al., 2024).

To address these challenges, the proposed solution involves developing a chatbot specifically trained on the specific university syllabus, ensuring that it caters to the unique needs of students and faculties(Chen et al., 2024). The methodology includes gathering extensive course material from textbooks, student notes,

[a] https://orcid.org/0009-0003-7357-2894
[b] https://orcid.org/0000-0001-9029-5161
[c] https://orcid.org/0000-0002-2214-9389

Satwani, C., Patil, V., Morab, K., Pattar, S. and Akkasaligar, P. T.
LLM Based Embedded Value Trained ChatBot Framework for Personalized Academic Learning.
DOI: 10.5220/0013599800004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 2*, pages 669-675
ISBN: 978-989-758-763-4

and online resources, followed by training the LLM, creating a model capable of understanding and responding to academic queries accurately. The chatbot interface serves as the application layer, facilitating users to pose questions and obtain exact answers utilizing the trained model(Ooi et al., 2023).

The experimental results show remarkable results. The accuracy of responses is on par with, other renowned generative AI models such as ChatGPT 3.5(OpenAI, 2023) and Gemini AI(Google, 2023). Various testing metrics indicate a significant performance boost, with the chatbot effectively resolving queries from both the students and faculties. This success demonstrates the practical applicability of the proposed model and its potential to enhance academic interactions significantly. In this regards, the contribution are as follows.

- *Domain-Specific Training:* The chatbot is specifically trained on a specific university data, ensuring high relevance and accuracy of responses.
- *Performance and Accuracy:* The model exhibits superior performance compared to traditional models, as evidenced by various testing metrics.

The rest of the paper is organized as follows. Section 2 introduces the problem statement and objectives, followed by the proposed methodology in Section 3. Section 4 discusses the implementations and results obtained. Finally, the paper concludes in Section 5.

2 PROBLEM STATEMENT

This section describes our problem statement and objectives of our work.

A. Problem Statement

Our goal is to design a university-specific chatbot that provides detailed, accurate responses to queries. Users interact with an interface, submitting questions that are processed by a value-trained model. This model, trained on diverse academic materials, generates precise, contextually relevant answers, ensuring immediate and effective support for students and faculty.

B. Objectives

The chatbot is designed for university students and faculty, addressing limitations such as maintaining context over long interactions, data quality, and privacy concerns. The objectives are:

1. *Support Stakeholder Queries:* The chatbot aims to address stakeholder queries efficiently, eliminating the need for manual searches.
2. *Provide Unambiguous and Precise Responses:* The chatbot must deliver swift, accurate, and clear answers, enhancing reliability and efficiency in academic support.

3 SYSTEM MODEL

This section provides a comprehensive overview of the architecture and operational workflow of the proposed chatbot as shown Fig 1. It starts by explaining the data collec- tion process, particularly how the training data is curated in alignment with the university's syllabus. It then outlines the embedding model used to convert this data into a structured format suitable for model training. Further, the model training phase is described, detailing the development and fine-tuning of the machine learning model.

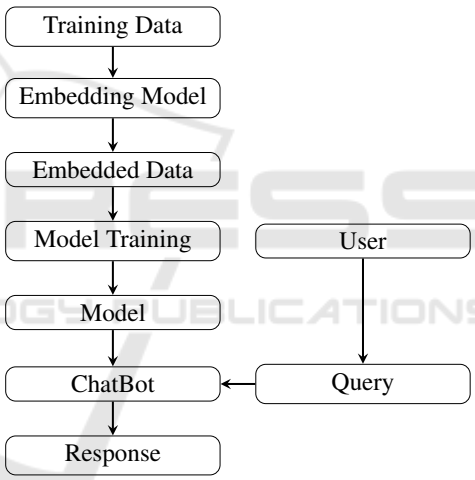

Figure 1: Flowchart of the proposed Chatbot System.

3.1 Model Description

The workflow begins with the acquisition of training data, specifically collected in alignment with the university's syl- labus. This data is processed by the embedding model, convert- ing it into structured embedded data. The embedded data then undergoes model training, resulting in a finely tuned machine learning model. Once the model is integrated into the chatbot system, users input their queries, and the chatbot leverages the trained model to promptly generate and deliver accurate results. This process ensures the chatbot to provide precise, unambiguous, and timely responses, significantly enhancing the academic support experience for all stakeholders.

3.2 Data Description

The dataset comprises a diverse collection of academic materials relevant to courses at a specific university. It includes textbooks, student notes, and online resources, primarily in PDF format. The dataset is carefully curated to ensure com- prehensive coverage of the course content, providing a rich source of information for embedding and subsequent use in the chatbot framework. Detailed descriptions and statistics of the dataset are presented in the Table1.

Table 1: Details of the Dataset.

Courses	Type	Materials	Size
Course–1	Theoretical	Notes + Text Book	2048 pages
Course–2	Theoretical	Online Resources + Notes	1980 pages
Course–3	Theoretical	Online Resources + Notes	1880 pages
Course–4	Laboratory	Notes + Online Resources	1500 pages

3.3 Embedding

The embedding process involves several stages to ensure the seamless integration of the instructor module within the LLM-based chatbot framework, tailored for a specific university. Initially, a comprehensive data collection phase is undertaken, gathering information from a variety of academic sources. These sources include textbooks, student notes, online materials, and any additional relevant resources related to specific courses. This data, often in PDF format, undergoes an extraction process to isolate the pertinent content required for embedding. Following data extraction, a crucial preprocessing stage is employed to cleanse the data. This involves removing noise, correcting inconsistencies, and standardizing the text to ensure it is in an optimal format for embedding. This step is vital to maintain the integrity and quality of the data, enabling effective embedding. The preprocessed data is then subjected to embedding algorithm using instructor-xl instructor model, transforming the textual information into numerical representations as shown in Table 2. These embeddings capture the semantic nuances and contextual meanings of the text, making them comprehensible for the LLM. The instructor module, now enriched with these sophisticated embeddings, is integrated into the chatbot framework. This integration allows the chatbot to utilize the embedded knowledge, providing personalized and contextually accurate academic assistance to students. The instructor module leverages the embedded data to generate detailed explanations, answer complex queries, and offer tailored learning support. This embedding process ensures that the chatbot is not only informed by a wide array of academic materials but also capable of delivering precise and relevant educational guidance, enhancing the learning experience for students at the specified university.

Table 2: Embedded Dataset Details.

CT	T	S	NC	UPC	NS	WFS
Course–1	Notes	60MB	10	7000–8500	60–65	12–14
	Text Book	60MB	12	7000–9000	60–65	12–14
Course–2	Online Res.	60MB	10	7000–8500	60–65	12–14
	Notes	60MB	10	7000–8500	60–65	12–14
Course–3	Online Res.	60MB	12	7000–9500	60–65	12–14
	Notes	60MB	10	7000–8500	60–65	12–14
Course–4	Notes	60MB	10	7000–8500	60–65	12–14
	Online Res.	60MB	10	7000–8500	60–65	12–14

CT: Course Type, T: Type of Material, S: Size of the dataset
NC: Number of chunks generated, UPC: Unique Word Per Chunk
NS: Number of sentences, WFS: Word Frequency Per Sentence

3.4 Training model

The training model serves as a cornerstone of the chatbot framework. It utilizes the embedded data and employs an advanced algorithm to generate a model proficiency in pro- cessing user queries and producing responses. Specifically, the Falcon-7B-Instruct model architecture (Almazrouei et al., 2023), an LLM for text- based chat applications, is leveraged. The model features 32 layers, a 4544-dimensional hidden space, 64-dimensional at- tention heads, a vocabulary of 65024, and supports 2048-token sequences. This model excels in text generation, prioritizing speed and efficiency over contextual depth, distinguishing it from other models. To tailor the model to specific require- ments, hyperparameters are adjusted, including temperature hyperparameter, token generation, and maximum length. This customization ensures optimal performance aligned with the specified needs. Further, the model undergoes training using university-specific data from the embedding module. Once the training phase is complete, the refined model is stored and integrated into the chatbot system. The resultant model, finely tuned to the specifications as shown in Table 3, is thus equipped to handle user queries with precision and generate accurate, relevant responses.

Table 3: Falcon - 7B LLM Specifications.

Hyperparameter	Value
Layers	32
d–model	4544
head–dim	64
Vocabulary	65024
Sequence length	2048

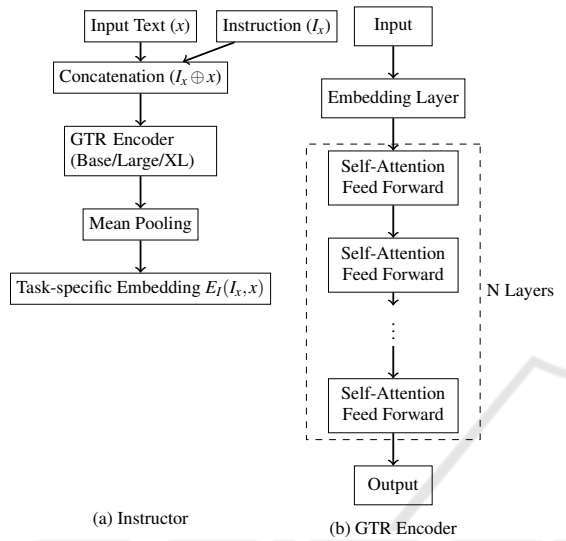

Figure 2: Instructor-XL Architecture.

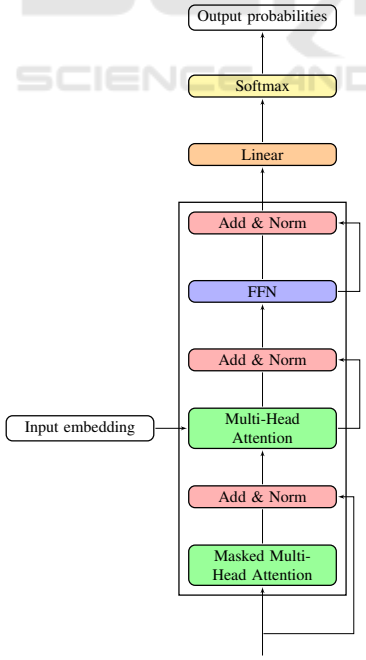

Figure 3: Falcon-7B Decoder Architecture.

3.5 Performance Evaluation Parameters

The following parameters are used to assess the proposed model's performance.

BERT–F1 Score: BERT–F1 Score is a metric that evaluates the quality of text generation by comparing the con- textual embeddings of the generated text and the reference text using BERT, a powerful language model. Unlike traditional metrics, BERT–F1 score captures subtle nuances in meaning and context, providing a more refined assessment of textual similarity and quality(Shankar et al., 2024). The equation for the BERT–F1 score is as follows.

$$Pr_{BERT} = \frac{1}{|\hat{y}|} \sum_{\hat{y}_j \in \hat{y}} \max_{y_i \in y} \overbrace{\mathbf{y}_i^\top \cdot \mathbf{y}_j}^{\text{cosine similarity}} \quad (1)$$

$$Re_{BERT} = \frac{1}{|y|} \sum_{y_i \in y} \max_{\hat{y}_j \in \hat{y}} \overbrace{\mathbf{y}_i^\top \cdot \mathbf{y}_j}^{\text{cosine similarity}} \quad (2)$$

$$BERTF_1 = 2 \cdot \frac{Pr_{BERT} \cdot Re_{BERT}}{Pr_{BERT} + Re_{BERT}} \quad (3)$$

where, Pr_{BERT} represents the precision score of BERTScore, while Re_{BERT} represents the recall score. The terms y_i and $\hat{y}_j$ are individual contextual embeddings. The dot product, $\mathbf{y}_i^\top \cdot \mathbf{y}_j$, reflects the cosine similarity between embeddings $\mathbf{y}_i$ and $\mathbf{y}_j$.

ROUGE Score: ROUGE measures the overlap of n-grams, word sequences, and word pairs between the generated output and reference text, indicating relevance and completeness(Shankar et al., 2024). The following equations show the calculation for ROUGE scores.

$$\text{ROUGE-1} = \frac{\text{Number of overlapping unigrams}}{\text{Total number of unigrams in the reference summary}} \quad (4)$$

$$\text{ROUGE-2} = \frac{\text{Number of overlapping bigrams}}{\text{Total number of bigrams in the reference summary}} \quad (5)$$

$$\text{ROUGE-L} = \frac{\text{LCS(C, R)}}{\text{Length of the reference summary}} \quad (6)$$

where LCS(C, R) is the length of the longest common subsequence between the candidate summary C and the reference summary R.

BLEU Scores: BLEU evaluates the precision of the generated text by comparing it to one or more reference texts, focusing on coherence and accuracy. The BLEU score is calculated using the following equations:

$$\left\{ \text{Precision}_n = \frac{\text{Number of n-grams in the candidate that appear in the references}}{\text{Total number of n-grams in the candidate}} \right. \tag{7}$$

$$BP = \begin{cases} 1 & \text{if Candidate Length} > \text{Reference Length}, \\ e^{\left(1 - \frac{\text{Reference Length}}{\text{Candidate Length}}\right)} & \text{if Candidate Length} \leq \text{Reference Length}. \end{cases} \tag{8}$$

$$\text{BLEU} = \text{BP} \times \exp\left(\frac{1}{N}\sum_{n=1}^{N} \log \text{Precision}_n\right) \tag{9}$$

The score ranges for BLEU, ROUGE, and BERT score all span from 0 to 1, with higher scores indicating better quality: BLEU scores above 0.5 are considered good for translation quality, ROUGE scores above 0.6 signify good overlap with reference texts, and BERT scores above 0.9 reflect high contextual similarity.

4 IMPLEMENTATIONS AND PERFORMANCE ANALYSIS

The experiments are conducted on Google Colab, which provides access to CPUs, GPUs like Tesla K80, T4, P4, and P100, TPUs, up to 12 GB RAM, and integrated Google Drive storage. The environment uses libraries such as langchain, torch, transformers, json, and matplotlib for efficient model training, text processing, and data visualization.

Data is collected from various sources, including textbooks, notes, and online materials, forming a comprehensive dataset of approximately 700 megabytes. This data is formatted into PDFs and processed using the PyPDF2 module in Python to extract textual information. The extracted text is then fed into the proposed embedding module, that performs meaningful embeddings based on provided instructions, resulting in a context-rich, high-quality training dataset for the LLM.

The Instructor-XL model as shown in Fig. 2 enerates task-specific embeddings by merging text inputs with task instructions, handling multiple tasks without needing extra training. It adds an extra fine-tuning layer and processes data in overlapping chunks to ensure high quality and ac- curate relationships. The model is trained with General Text Representation (GTR) architecture for multitask approach on 330 diverse datasets. The model fine-tunes embeddings with contrastive loss to separate, related from unrelated text pairs. It excels in tasks like classification, information retrieval, and semantic similarity. Similarity between texts is determined by cosine similarity of their embeddings. Higher the similarity indicates greater textual closeness.

Table 4: Sample Course Queries.

Query	Question	Answer
Q1	Discuss the concept of operational amplifiers (op-amps) in analog electronics?	Op-amps are electronic circuits that amplify small signals and convert them into large signals that can be used within a device.
Q2	What are key components of computer organization and architecture?	The key components of computer organization and architecture include processor (CPU), memory (RAM/ROM), storage devices (hard drives/SSDs), control unit (CPU's processor), and input/output devices (keyboard/input devices).
Q3	How does the concept of eigenvalues and eigenvectors relate to Principal Component Analysis (PCA) in statistics and data analysis?	Eigenvalues determine the variance captured by each principal component, while eigenvectors define the direction of these components. Eigenvectors, as principal components, are linear combinations of original data points that capture the most variance.
Q4	Explain the concept of matrix rank and its significance in linear algebra?	A matrix rank is used to solve problems that involve computing the eigenvectors and eigenvalues of a matrix, and thus, matrix ranks have played a key role in the development of modern linear algebra theory and applications.

This embedded dataset as shown in Table 2 is used to train the customized Falcon-7B model, fine-tuned to retain the meaning of the text during the embedding process, ensuring accurate and contextually relevant responses. The final chatbot system integrates this well-trained model with an interface. The user submits the query, to the model to generate pre- cise and relevant responses, enhancing academic support for stakeholders at the specified university. The Table 4 shows the sample query asked to the model by the stakeholders.

When the user asks a query to the chatbot model, the chatbot processes the input using an embedding model. The embedding model converts the text into a structured format and uses the Instructor-XL model to generate meaningful embedding. The embedded data is then used to train the Falcon-7B-Instruct model, fine-tuned with embedded value training. The trained model generates a response based on the input query.

The Table 5 hows queries asked to both ChatGPT-

3.5 and the proposed model during testing phase, with ratings provided by the stakeholders to the responses. The proposed model demonstrates superior performance compared to ChatGPT-3.5, with more accurate and context-aware responses

Table 5: Course Queries and Ratings

Query	Chat-GPT 3.5	Proposed Model
Q1	4.3	4.5
Q2	4.0	4.4
Q3	4.2	4.6
Q4	3.8	4.4

In the chatbot implementation, fine-tuning of the LLM is performed to optimize response time efficiency. From Fig. 4 we find that extensive experimentation and rigorous opti- mization leads to a significant reduction in response time. The Fig. 4 The Fig. 4 shows response generation time ranging from 2 to 15 seconds. This accomplishment not only validates the tuning process but also highlights the model's capability to deliver prompt and accurate answers. Thus, the chatbot demonstrates the practical application and efficiency of the optimized LLM model in real-world scenarios.

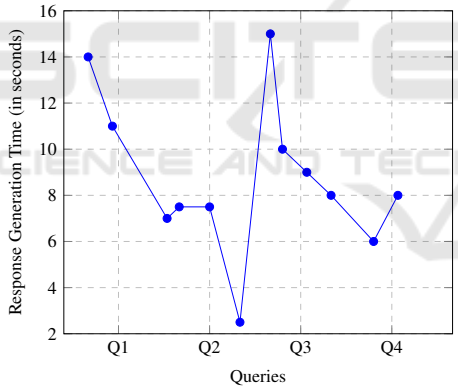

Figure 4: Response Times of LLM Model.

The BERT F1 score helps ensure query responses are contextually accurate and meaningful. The Fig.5 shows comparison of BERTF1Score for proposed method ChatGPT,GeminiAI. From the Fig 5 it is found that the proposed model outperforms ChatGPT-3.5 and Gemini AI in machine translation, text summarization, and dialogue generation. This leads to a significant impact on educational chatbots, as the proposed model can provide more precise, relevant, and high- fidelity interactions, enhancing personalized learning experi- ences and ensuring students receive high-quality support

ROUGE and BLEU are evaluation metrics for assessing the quality of generated text in natural language processing. The Fig.6 shows that the proposed model achieves parity with GPT-3.5 and outperforms Gemini AI by 30% due to its exceptional ability to generate contextually relevant and comprehensive responses, as reflected in the high ROUGE scores as shown in Fig. 6. The high BLEU scores further highlight the model's precision and fluency in language generation as shown in Fig. 7. his superior performance enhances the effectiveness of educational chatbots by ensuring they provide coherent, accu- rate, and contextually appropriate support, thereby improving the overall learning experience for students.

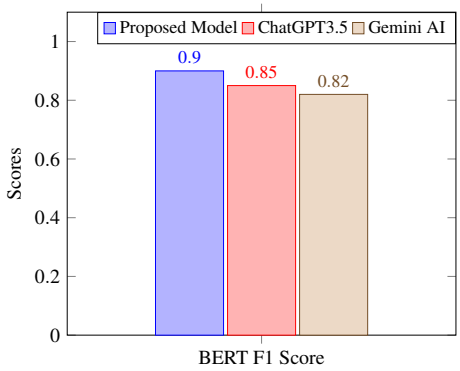

Figure 5: Comparison of Bert F1 Scores.

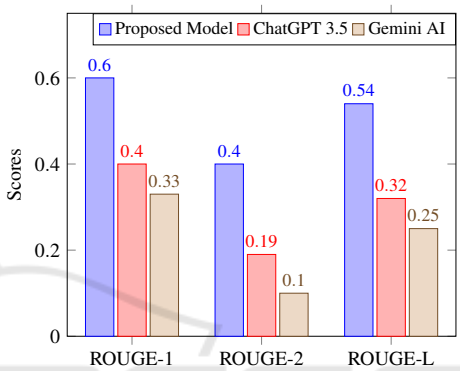

Figure 6: Comparison of ROUGE Scores

The results obtained are a direct outcome of the modifi- cations implemented. These enhancements give the proposed model an edge over others, making it unique and poised for great success in the future.

5 CONCLUSIONS

The integration of chatbots in education presents a signifi- cant opportunity to enhance personalized learning experiences for students. However, existing language models and chatbots often face challenges in accurately understanding and respond- ing to the

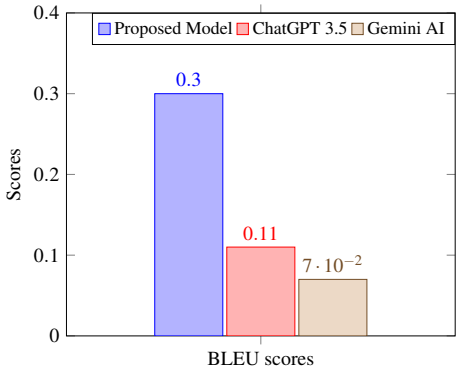

Figure 7: Comparison of BLEU Scores.

nuanced needs of learners, leading to inconsistencies and gaps in educational delivery. The proposed approach leverages an LLM fine-tuned with embedded value training, addresses these challenges by ensuring more contextually rele- vant and comprehensive responses. The superior performance of the model, as demonstrated by high ROUGE and BLEU scores, indicates its proficiency in generating coherent and precise language, surpassing both GPT-3.5 and Gemini AI. These results underscore the model's potential to improve educational outcomes by providing accurate and meaningful interactions. Future work will focus on further refining the model to enhance its adaptability and scalability, ensuring it can cater to a diverse range of educational contexts and needs. This continued development aims to solidify the role of advanced chatbots in shaping the future of personalized academic learning.

REFERENCES

Almazrouei, E., Alobeidli, H., Alshamsi, A., Cappelli, A., Cojocaru, R., Debbah, M., Goffinet, É., Hesslow, D., Launay, J., Malartic, Q., et al. (2023). The falcon series of open language models. *arXiv preprint arXiv:2311.16867*.

Chen, J., Lu, X., Du, Y., Rejtig, M., Bagley, R., Horn, M., and Wilensky, U. (2024). Learning agent-based modeling with llm companions: Experiences of novices and experts using chatgpt & netlogo chat. In *Proceedings of the CHI Conf. on Human Factors in Computing Systems*, pages 1–18. ACM.

Firth, J. A., Torous, J., and Firth, J. (2020). Exploring the impact of internet use on memory and attention processes. *International Journal of Environmental Research and Public Health*, 17(24).

Florindi, F., Fedele, P., and Dimitri, G. M. (2024). A novel solution for the development of a sentimental analysis chatbot integrating chatgpt. *Personal and Ubiquitous Computing*, pages 1–14.

Google (2023). Gemini AI. https://gemini.google.com/. Text-based AI Model.

Mirowski, P., Mathewson, K. W., Pittman, J., and Evans, R. (2023). Co-writing screenplays and theatre scripts with language models: Evaluation by industry professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM.

Montagna, S., Ferretti, S., Klopfenstein, L. C., Florio, A., and Pengo, M. F. (2023). Data decentralisation of llm-based chatbot systems in chronic disease self-management. In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*, pages 205–212.

Omoregbe, N. A. I., Ndaman, I. O., Misra, S., Abayomi-Alli, O. O., Damaševičius, R., and Dogra, A. (2020). Text messaging-based medical diagnosis using natural language processing and fuzzy logic. *Journal of Healthcare Engineering*, 2020:1–14.

Ooi, K.-B., Tan, G. W.-H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., Dwivedi, Y. K., Huang, T.-L., Kar, A. K., Lee, V.-H., et al. (2023). The potential of generative artificial intelligence across disciplines: Perspectives and future directions. *Journal of Computer Information Systems*, pages 1–32.

OpenAI (2023). ChatGPT (v3.5). https://chat.openai.com. Large Language Model.

Shankar, S., Zamfirescu-Pereira, J., Hartmann, B., Parameswaran, A. G., and Arawjo, I. (2024). Who validates the validators? aligning llm-assisted evaluation of llm outputs with human preferences. *arXiv preprint arXiv:2404.12272*.

Šarčević, A., Tomičić, I., Merlin, A., and Horvat, M. (2024). Enhancing programming education with open-source generative ai chatbots. In *2024 47th MIPRO ICT and Electronics Convention (MIPRO)*, pages 2051–2056.

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., and Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1):90–112.