SteamPATHY: a Conversational STEAM and Empathetic Dataset Harnessed Using ChatGPT

Anonymous ACL submission

Abstract

Artificial intelligence, particularly large language models (LLMs), has made a significant impact across various domains, extending its reach into areas originally not tailored for its application, such as education and child emotional support. This expansion is driven by the LLMs' remarkable capacity to generate human-like content and engage in extended, coherent conversations. Besides, LLMs powered Chatbots can make STEAM education and empathetic virtual companions more accessible to children from different backgrounds. However, the process of fine-tuning a LLM model for this particular task faces a challenge due to the unavailability of readily accessible datasets focused on elementary school-level STEAM and empathy. Such sparsity of data can be solved by synthetically generating dataset samples using generative AI methods. While ChatGPT has been used in the past to generate datasets to finetune smaller and more computationally efficient LLMs on specific tasks, we propose in this paper to leverage ChatGPT prompting to generate *SteamPATHY*, student-caretaker interactions that cover STEAM and empathetic topics.

1 Introduction

In an educational environment that nurtures Science, Technology, Engineering, Arts, and Mathematics (STEAM), students are encouraged to apply critical thinking skills to devise innovative solutions for real-world problems (Yakman, 2008). Artificial Intelligence (AI) has undergone a profound transformation across various domains, with education emerging as a primary beneficiary. Its effectiveness, particularly in the realm of online teaching and learning, becomes evident through its ability to automate tasks for instructors and generate assessments (Seo et al., 2021).

Moreover, recent studies have highlighted the potential of child-friendly chatbots to serve as valuable resources when parents are unavailable. These chatbots offer children novel avenues for self-expression, which, in turn, can positively influence their mental well-being (Aarts et al., 2022; Cooper & Ireland, 2018; Santos et al., 2020; Zhang et al., 2022). Consequently, the integration of AI-powered chatbots into STEAM education merits thoughtful consideration. Such an approach not only enhances accessibility and engagement but also empowers children to articulate their thoughts and emotions freely.

The most recent advancements in AI-powered Chatbots have leaned heavily on large language models (LLM), as they have proven their ability to sustain longer and more coherent conversations compared to the conventional natural language processing techniques utilized in chatbot development (Lee, 2023). These LLM-based models can be finely tuned to create a STEAM-oriented, empathetic conversational chatbot, finely attuned to the academic and emotional needs of elementary school children. However, the primary challenge impeding the development of such LLM-powered STEAM and empathetic chatbots for children is the lack of datasets that can be employed to fine-tune these models into helpful assistants for children. To address this data sparsity issue, we propose utilizing synthetic data generation techniques to gather text in any desired format. LLMs like ChatGPT have already been employed successfully to generate data for fine-tuning smaller LLMs on specific tasks, such as following instructions and coding (Ding et al., 2023; Köksal

et al., 2023; Rohan et al., 2023; Ubani et al., 2023; Zhou et al., 2023). Our recommendation is to prompt ChatGPT to simulate conversations between a child and their caretaker on various STEAM and empathetic topics. It is crucial to note that all content generated by ChatGPT underwent thorough review to ensure factual accuracy and appropriateness before releasing it.

More specifically, the paper contributes to the literature a ChatGPT based approach for collecting elementary school level children dataset for STEAM education and empathetic conversations.

The rest of the paper is divided such that Section 2 presents related work on augmenting datasets with synthetic data produced by LLMs while Section 3 briefly describes the curated dataset SteamPATHY. Section 4 demonstrates how the dataset was collected and processed. Section 5 demonstrates the evaluations of the dataset. Section 6 presents limitations. Section 7 concludes the work.

2 Related Work

Table 1 provides a concise overview of existing children chatbots as documented in the literature. These chatbots primarily rely on conventional NLP techniques and utilize pre-defined sets of questions, answers, and conversation flows tailored for specific tasks. However, it's important to note that none of the chatbots listed in Table 1 offer comprehensive coverage of general STEAM topics and demonstrate empathy towards children's broader concerns.

To bridge this gap, there is a need for the development of children chatbots by harnessing the capabilities of LLMs. LLMs can be effectively trained to create chatbots capable of engaging in more extended, diverse, and coherent conversations while specializing in STEAM subjects and empathetic dialogues, as highlighted in (Lee, 2023).

While STEAM-related content is abundantly available in sources such as Wikipedia scraped datasets and select children's websites like kids.kiddle.co, it's important to recognize that this data remains unstructured. Consequently, it may not serve as a suitable foundation for training an LLM to effectively engage in STEAM discussions with children.

Furthermore, datasets like "daily_dialog," "prosocial-dialog," and "Customer Support on Twitter" are indeed conversational in nature. However, they predominantly cover typical adult conversations, which do not equip an LLM with the necessary insights to address a child's emotional needs, as noted in references (*Allenai/Prosocial-Dialog · Datasets at Hugging Face*, 2022; *Daily_dialog · Datasets at Hugging Face*, 2017; Axelbrooke, 2017).

In light of these considerations, the primary challenge in developing such a chatbot lies in the absence of datasets that showcase caretaker-child interactions related to STEAM subjects and empathetic conversations. Addressing this data gap will be crucial to the successful creation of chatbots tailored to engage with children effectively in these domains.

One solution to overcome the lack of a dataset is to use AI assisted synthetic data generation techniques. Ubani et al. (Ubani et al., 2023) augmented training sets in low resource scenarios using data obtained from prompting LLMs such as ChatGPT, and showed that with task specific ChatGPT prompts it is possible to outperform other state of the art approaches for data augmentation. Furthermore, such large LLMs have been prompted in the literature to produce synthetic instruction datasets that were used to successfully improve the performance of smaller LLMs. Figure 1 summarizes the most prominent LLM generated and augmented instruction datasets found in literature (Ding et al., 2023; Köksal et al., 2023; Rohan et al., 2023).

While it may be tempting to use these datasets to finetune a STEAM specialized and empathetic chatbot for children, such datasets follow a non-conversational instruction-answer paradigm, and manually inspecting such datasets shows that a lot of the target responses are challenging for children to read. In fact, running the Flesch-Kincaid Grade (Kincaid et al., 1975) on such datasets shows that they require a 9th US grade reading level to easily understand their content, which is not suitable for elementary school children.

For these reasons, this paper uses ChatGPT to produce *SteamPATHY*, a dataset specialized for STEAM and conversational for elementary school children.

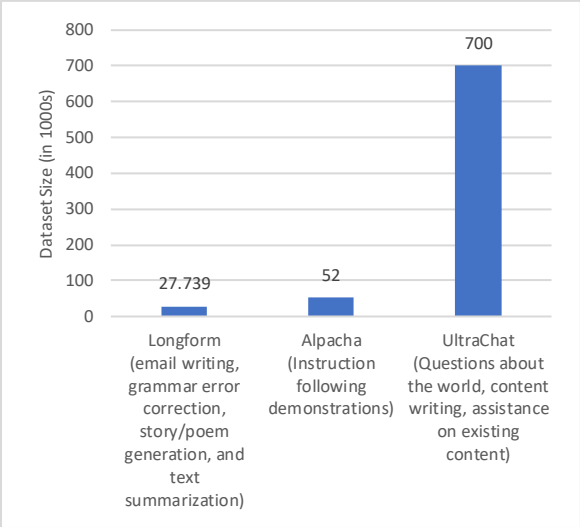

Figure 1 Summary of LLM generated instruction datasets

Authors	Contribution	Methods	Citations
Cooper et al. (2018)	Chatbot that can aid children on the autism spectrum.	AIML programming language to define chat patterns.	31
Liu et al. (2022)	Chatbots as book talk companions for children.	Google Actions Console to produce a Chatbot system with questions and answers on books.	42
Aarts et al. (2022) (Aarts et al., 2022)	Help children with sleeping disorders.	Predefined a set of conversations a child can have with Snoozy based on interviews	4
Mageira et al. (2022)	Teach high school students cultural content in the English or French languages.	Snatchbot website and predefined knowledgebase	62
Rajwal (2023)	A proposed framework for a Chatbot that can help improve the emotional intelligence of children.	Google's DialogFlow and Assistant and a knowledgebase of conversations.	-

Table 1 Children chatbots literature summary

3 The Dataset

The curated dataset *SteamPATHY* consists of 308 conversations between a student and a teacher. 210 of these conversations are STEAM related (hereafter Steam), and 71 conversations are about empathetic situations (hereafter Pathy) where the student may need guidance. Each conversation contains around 7 turns on average so around 2000 samples were prepared by formatting the dataset such that the chat history is the input and the teacher's response given the chat history is the output. Given the advancements in parameter efficient finetuning techniques, such a small dataset could be sufficient for orient an LLM to be a child's STEAM and empathetic companion (Xu et al., 2023).

As validation for the need of *SteamPATHY*, we compared the dataset to the Alpaca dataset - which is one of the most popular instruction datasets available on Huggingface (*Hugging Face – The AI Community Building the Future.*, n.d.), on several metrics. Finally, we classified our data to understand its distribution across topics and sentiments. We classified the data by first finding the contextualized embeddings for our class words and the conversations, and then assigning each conversation to the class word it is closest to in the embeddings space. The classes and the distribution of the chats among them can are shown in the Appendix.

Furthermore, we had five reviewers inspect the generated data to ensure the information quality

and conversational style. Section 5 details the reviewers' evaluation process and results.

Figure 2 and Figure 3 show the word maps for the Steam and Pathy conversational parts of the *SteamPATHY* dataset. The most prominent words in Steam include make, use, create, learn, heard, and know all of which are associated with the STEAM pedagogy that encourages critical thinking. Secondary words include wow, yes, thank, fun, great question, try, tell, sure, and keep asking, which are positive reinforcement words used in the context of engaging conversations. On the other hand, PATHY is dominated by phrases such as talk, feel, feeling, teacher, help, feel better, and feel nervous, because this subpart of the *SteamPATHY* dataset focuses on conversations where the child is expressing his/her feelings to his/her teacher regarding a specific situation and then seeking the teacher's help and advice.

For the sake of comparison, Figure 4 shows the word map for the Alpaca dataset. The most prominent words include explain, create, make, generate, provide, describe, list, and find, which are used in imperative sentences to order the model to follow an instruction and produce a result. Furthermore, unlike our STEAM dataset, the Alpaca dataset lacks the conversational words discussed previously.

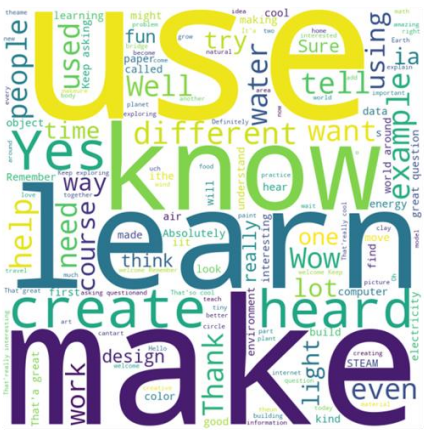

Figure 2 The word map for Steam of the SteamPATHY dataset

Figure 3 The word map for PATHY of the SteamPATHY dataset

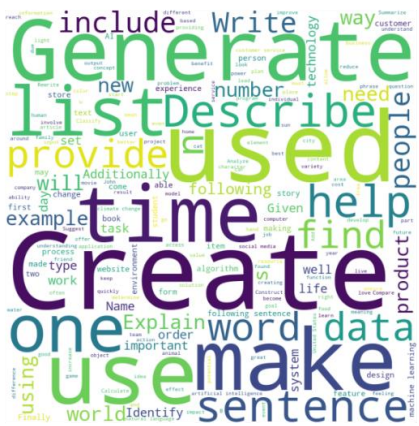

Figure 4 The word map for the Alpaca dataset

Moreover, Table 2 compares our dataset to the Alpaca dataset on the average Flesch-Kincaid US

Grade readability, the average target response length, and entropy. Our dataset achieves a 5.4 Flesch-Kincaid score meaning that it is readable by 5th grade students, while the Alpaca dataset requires 9th grade reading level. We also compared the entropy of our dataset to that of Alpaca. Since our dataset has similar high entropy levels to Alpaca, we have a preliminary indication that the quality of the English text and redundancy levels in both datasets is similar. This combined with the significantly shorter target chatbot responses of our dataset makes it more appealing for finetuning an LLM Chatbot to converse with lower elementary school children.

	SteamPATHY	**Alpaca**
Flesch-Kincaid Grade	**5.4**	9.3
Target Response Length (words)	**33.37**	44.18
Entropy	4.64	4.6

Table 2 The average Flesch Kincaid grade, response length, and entropy for the STEAMPATH and the Alpaca datasets

Furthermore, the Steam and PATHY data were classified as shown in Tables 3 and 4. The clusters can be found in Figure 5 and Figure 6 of the Appendix.

A third of the Steam dataset revolves around science, another third is about engineering and technology, and the last third is about art and mathematics. While art is often neglected in STEAM education (C.-Y. Liu & Wu, 2022), we ensured that more than 20% of the topics revolve around art. Only 13.3% of the chosen STEAM topics revolve around mathematics because of the difficulty we faced getting ChatGPT to produce reliable mathematics outside basic theories. Therefore, we decided that mathematics shouldn't be a primary goal for an LLM finetuned on this dataset as shown in Table 3.

As for the gathered empathetic situations in PATHY, about 70% of the situations we gathered result in a child having a negative sentiment. We choose such a distribution because we believe that finetuning a LLM capable of correctly guiding a child through a difficult situation is a more critical and challenging task than finetuning a LLM to encourage the child in positive events.

Topic	**Count**	**Percentage (%)**
Science	69	32.8
Technology	39	18.6
Engineering	29	13.8
Art	45	21.4
Mathematics	28	13.3

Table 3 The distribution of the STEAM topics over science, technology, engineering, art, and mathematics in Steam of the SteamPATHY dataset

Sentiment	**Count**	**Percentage (%)**
Excited	6	8.5
Upset	8	11.3
Proud	6	8.5
Frustrated	17	23.9
Anxious	25	35.2
Indifferent	9	12.7

Table 4 The distribution of empathetic situations over various sentiments in the PATHY subdataset of SteamPATHY

4 Building SteamPATHY

4.1 Curating a Dataset Using ChatGPT

In this section we explain how *SteamPATHY*, and values-targeted dataset were prepared using ChatGPT. While instruction generation prompting asks ChatGPT to produce an instruction, a response to an instruction, or both, our methodology relies on prompting ChatGPT to generate conversations by alternating between the roles of a curious elementary school child needing assistance and an empathetic, kind, and encouraging teacher.
To automate the generation process, a python script that uses ChatGPT API (gpt-3.5-turbo API) was used to send the prompts to ChatGPT, collect and store the responses.

4.2 STEAM

The first step to curating the Steam data was collecting a list of STEAM topics that would spark the curiosity of a child, and that children are usually very curious about. We collected a list of around 210 topics by first querying children websites (https://www.kiddle.co/, https://www.sciencekids.co.nz/, and https://academickids.com/), and then prompting ChatGPT to generate related topics. The topics include but are not limited to fossils, space, AI,

biomimicry, origami, mosaic art, and climate change.

The second step was to generate conversations that simulate the child-teacher interaction regarding those STEAM topics. The teacher had to be kind, emotionally connect with the child, encourage his/her curiosity, and answer any follow up questions the child may have. Also, since the child is in elementary school, the questions and answers are expected to be short. To keep the child safe, the teacher should not provide the student with any links to the internet.

The following prompt was used to generate the STEAM conversations between a child and his teacher:

> *Generate a conversation between a teacher chatbot named SAHAR and an elementary school student. SAHAR, the chatbot is a STEAM education teacher, she is very nice, funny and she connects emotionally with her students. The student is curious and young, he will ask a lot of follow-up questions. SAHAR should thoroughly answer the student's questions with short, simple and STEAM inspired answers while trying to spur his/her curiosity. Make the conversation as creative as possible, and remember, short answers with a lot of mind-stimulating follow up questions. SAHAR should say words of encouragement from time to time. SAHAR and the student know each other, no need for introductions. SAHAR doesn't have access to the internet and she cannot send the student any links or documents. The student hears about a STEAM topic and s/he's curious to know more about it. The conversation should include a scenario on how the student knew about this specific topic.*

The <<TOPIC>> token is replaced each time with one of the 210 curated topics and the gpt-3.5-turbo API is called to generate a new conversation.

4.3 PATHY

A similar approach to the one presented in the previous section was used to curate the empathetic data. ChatGPT was asked to compile a list of situations where children would want to express themselves, discuss their achievements or failures, or seek advice for a problem they are facing. A list of about 70 situations was compiled. The following prompt was used in generating the dataset. The list includes situations such as being subjected to bullying, failing a test, being asked to read or perform in front of class, making a new friend, getting praised for good behavior, or having their academic achievements recognized.

> *Generate a conversation between SAHAR and an elementary school student. SAHAR is a compassionate chatbot, that connects deeply with the students, gives them the floor to express their feelings on a specific situation, and will sometimes help them find a solution to a specific problem. SAHAR will, if needed offer emotional support to the student. The student will talk to SAHAR about the following: <<SITUATION>>. SAHAR and the student know each other, no need for introductions. Limit the conversation to 1000 tokens.*

The <<SITUATION>> token is replaced by one of the 70 situations covered in the curated list and related to empathy.

4.4 Dividing the Conversations to sub-conversations

The Chatbot should be trained to generate the appropriate response to the child's last question given the context from previous questions and their corresponding answers. As such, feeding the entire conversation to the model is not optimal, and the conversation must be divided into sub-conversations. The input and the labels for the sub-conversations we used are as follows:

1. Input: the input sample is comprised of a question asked by the student in each conversation preceded by the needed context, which is provided as the previous questions and answers from the same conversation. The context and question are enclosed with beginning and end of question tokens (<BOQ> and <EOQ>).
2. Label: the target output is the response to the student's question given the previous questions and answers as context. The target output will be enclosed with beginning and end of answer tokens (<BOA> and <EOA>).

The <BOQ>, <EOQ>, <BOA>, and <EOA> tokens are used for the following reasons. First, the <BOQ> and <EOQ> tokens bound the question

and context and help the model understand where the question finishes so that it can start generating the response. Without the <EOQ> token, the model may decide to generate a response that adds to the student's question before answering it. Second, the <EOA> token helps the model understand where their response needs to be terminated so that the model does not generate text beyond the needed response. Third, even in the case where the model generated text after the <EOA> token, only text before the <EOA> can be presented to the user in post processing.

Applying this methodology allowed for generating several datapoints from a single conversation. Therefore, the final dataset is comprised of about 2000 samples.

5 Dataset Evaluation

We relied on five reviewers with different backgrounds and professional English proficiency and recruited as part of the same research team to give their opinion about the *SteamPATHY* dataset. Table 5 summarizes the credentials of our reviewers.

Reviewer	Age	Background	English proficiency
1	21	Computer and Electrical Engineer	Professional
2	24	Computational Physics	Professional
3	24	Computer Engineer	Professional
4	21	Computer Science and Engineering	Professional
5	40	Biophysicist and AI	Professional

Table 5 Reviews' credentials

We asked the reviewers questions that are common to the STEAM and empathy parts of the datasets and additional questions that are specific to each part.

The common questions aimed to assess if the conversations generated remained on coherent within the topic, had any redundancy, and were simple to read.

For the STEAM partition, we were interested in knowing if the answers provided by the Chatbot were also factual for an elementary school level.

As for the Empathy partition, our concerns were that the chatbot might not validate the child's feelings, or it might suggest infeasible or dangerous solutions for the child. Therefore, our questions targeted those concerns.

We also asked the reviewers to give suggestions about any chats that they rate poorly. These suggestions were then used to amend the poor chats to ensure the quality of all chats in the dataset.

Table 6 summarizes the questions that were asked to the reviewers about the dataset samples.

	Question	Answer Format	Available Choices
Common Questions	The Conversation is on Coherent.	Yes/No	Yes or No
	The Conversation is not Redundant.	Yes/No	Yes or No
	The Language is Simple.	Yes/No	Yes or No
STEAM Questions	The Content is Factual for an Elementary School Level	Yes/No	Yes or No
Empathy Questions	The chatbot validates the child's feelings in the situation.	Yes/No	Yes or No
	The chatbot suggests viable solutions to the child's problem	Yes/No	Yes or No or Not Applicable
	The chatbot's suggestions do not put the child in danger	Yes/No	Yes or No

Table 6 Questions Summary

Table 7 summarizes the average scores per reviewer per question. We can see that the scores are overall high with no clear discrepancies between reviewers.

	Reviewers % Yes					
Question	**1**	**2**	**3**	**4**	**5**	**Average**
The Conversation is on Coherent.	100.0	98.8	100.0	98.3	98.8	**95.72**
The Conversation is not Redundant.	94.3	92.4	97.6	96.9	97.4	**94.24**
The Language is Simple.	84.1	93.3	100.0	95.0	98.8	**96.16**
The Content is Factual for an	99.5	100.0	97.6	98.6	85.1	**97.46**

Elementary School Level						
The chatbot validates the child's feelings in the situation.	98.6	100.0	100.0	98.6	90.1	**95.78**
The chatbot suggests viable solutions to the child's problem	98.6	100.0	100.0	94.4	85.9	**95.58**
The chatbot's suggestions do not put the child in danger	93.4	98.6	98.6	95.7	91.6	**95.72**

Table 7 Reviewers' evaluation results for the SteamPATHY dataset

6 Limitations

While this study provides a human evaluated dataset that can help train an empathetic chatbot that can discuss STEAM topics with a child, it has several limitations. First, the dataset curated is not exhaustive to every STEAM and empathetic topic a child may want to discuss. Therefore, a chatbot trained on this dataset is not guaranteed to be hallucination free when asked about topics it is not exposed to. Second, the generation and evaluation process are not scalable since it requires human evaluators to proofread and give suggestions about the text generated by ChatGPT to guarantee factuality and safety. This limited the size of our dataset. Finally, the none of our evaluators where from literary or childcare backgrounds. Perhaps having reviewers from such backgrounds may shed light on further improvements that can be made to the dataset.

7 Conclusion

To make STEAM education more accessible to lower elementary school students from various backgrounds and given that the role of a child's caretaker extends beyond education to providing emotional support, and compassion, we developed *SteamPATHY*, a dataset that can be used to finetune the training of an LLM based chatbot capable of engaging with the child in a compassionate, informative tone while answering with correct and unbiased responses. To that end, we used ChatGPT to curate the *SteamPATHY* dataset. Comparing our dataset to an available instruction dataset our dataset is more readable for children, and has more conversational and empathy components, making it more suitable for finetuning an LLM to become a STEAM based and empathetic companion for a child. Future works will focus on discovering more scalable ways of constructing such datasets to guarantee a wider coverage of topic.

References

Aarts, T., Markopoulos, P., Giling, L., Vacaretu, T., & Pillen, S. (2022). Snoozy: A Chatbot-Based Sleep Diary for Children Aged Eight to Twelve. *Interaction Design and Children*, 297–307. https://doi.org/10.1145/3501712.3529718

allenai/prosocial-dialog · Datasets at Hugging Face. (2022). https://huggingface.co/datasets/allenai/prosocial-dialog

Axelbrooke, S. (2017). *Customer Support on Twitter*. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/8841

Cooper, A., & Ireland, D. (2018). Designing a Chat-Bot for Non-Verbal Children on the Autism Spectrum. *Studies in Health Technology and Informatics*, *252*.

daily_dialog · Datasets at Hugging Face. (2017). https://huggingface.co/datasets/daily_dialog

Ding, N., Chen, Y., Xu, B., Qin, Y., Zheng, Z., Hu, S., Liu, Z., Sun, M., & Zhou, B. (2023). *Enhancing Chat Language Models by Scaling High-quality Instructional Conversations*.

Hugging Face – The AI community building the future. (n.d.). Retrieved December 28, 2023, from https://huggingface.co/datasets?other=instruction-finetuning&sort=downloads

Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*.

Köksal, A., Schick, T., Korhonen, A., & Schütze, H. (2023). *LongForm: Optimizing Instruction Tuning for Long Text Generation with Corpus Extraction*.

Lee, A. (2023). *What Are Large Language Models Used For and Why Are They Important? | NVIDIA Blog*. https://blogs.nvidia.com/blog/2023/01/26/what-are-large-language-models-used-for/

Liu, C.-C., Liao, M.-G., Chang, C.-H., & Lin, H.-M. (2022). An analysis of children' interaction with an AI chatbot and its impact on their interest in reading. *Computers & Education*, *189*, 104576. https://doi.org/https://doi.org/10.1016/j.compedu.2022.104576

Liu, C.-Y., & Wu, C.-J. (2022). STEM without art: A ship without a sail. *Thinking Skills and Creativity*, *43*, 100977. https://doi.org/https://doi.org/10.1016/j.tsc.2021.100977

Mageira, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. (2022). Educational AI Chatbots for Content and Language Integrated Learning. *Applied Sciences (Switzerland)*, *12*(7). https://doi.org/10.3390/app12073239

Rajwal, S. (2023). Design of a Chatbot for Four- to Ten-Year-Old Children Based on Emotional Intelligence. In D. Gupta, A. Khanna, S. Bhattacharyya, A. E. Hassanien, S. Anand, & A. Jaiswal (Eds.), *International Conference on Innovative Computing and Communications* (pp. 675–683). Springer Nature Singapore.

Rohan, T., Ishaan, G., Tianyi, Z., Yann, D., Xuechen, L., Carlos, G., Percy, L., & B., H. T. (2023). *Alpaca: A Strong, Replicable Instruction-Following Model*. Stanford CRFM. https://crfm.stanford.edu/2023/03/13/alpaca.html

Santos, K. A., Ong, E., & Resurreccion, R. (2020). Therapist vibe: Children's expressions of their emotions through storytelling with a chatbot. *Proceedings of the Interaction Design and Children Conference, IDC 2020*. https://doi.org/10.1145/3392063.3394405

Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education*, *18*(1). https://doi.org/10.1186/s41239-021-00292-9

Ubani, S., Polat, S. O., & Nielsen, R. (2023). *ZeroShotDataAug: Generating and Augmenting Training Data with ChatGPT*.

Xu, L., Xie, H., Qin, S.-Z. J., Tao, X., & Wang, F. L. (2023). *Parameter-Efficient Fine-Tuning Methods for Pretrained Language Models: A Critical Review and Assessment*.

Yakman, G. (2008). STEAM Education: An Overview of Creating a Model of Intergrative Education. *Pupils Attitudes Towards Technology (PATT-19) Conference : Research on Technology, Ionnovation, Design & Engineering Teaching*, *53*(9).

Zhang, Z., Xu, Y., Wang, Y., Yao, B., Ritchie, D., Wu, T., Yu, M., Wang, D., & Li, T. J.-J. (2022). StoryBuddy: A Human-AI Collaborative Chatbot for Parent-Child Interactive Storytelling with Flexible Parental Involvement. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3491102.3517479

Zhou, C., Liu, P., Xu, P., Iyer, S., Sun, J., Mao, Y., Ma, X., Efrat, A., Yu, P., Yu, L., Zhang, S., Ghosh, G., Lewis, M., Zettlemoyer, L., & Levy, O. (2023). *LIMA: Less Is More for Alignment*.

A Clustering the Dataset

Below in the next page we show the clusters of the Steam and EMPATHY dataset partitions.

A.1 Steam Topics

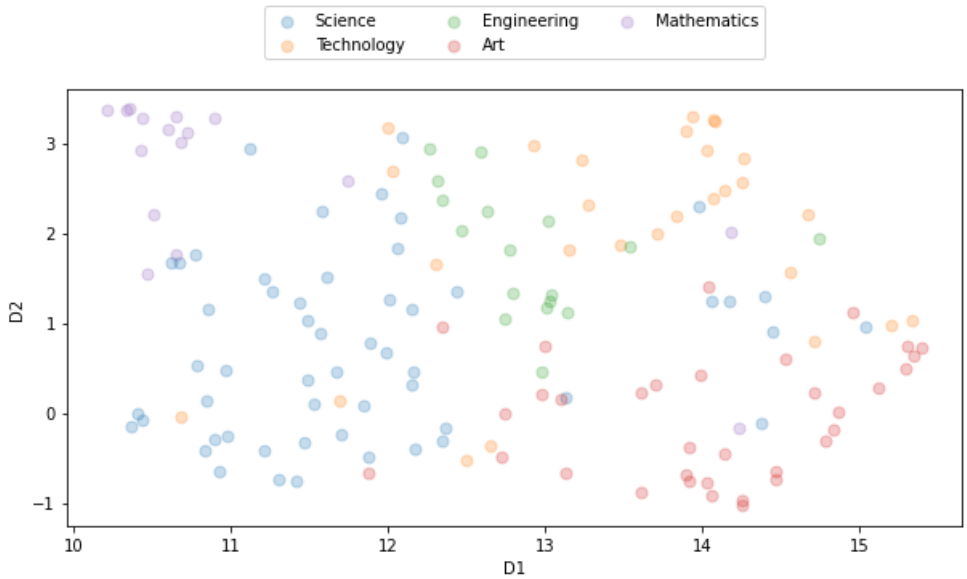

Figure 5 Steam topics clustering

A.2 Empathetic Situations (PATHY)

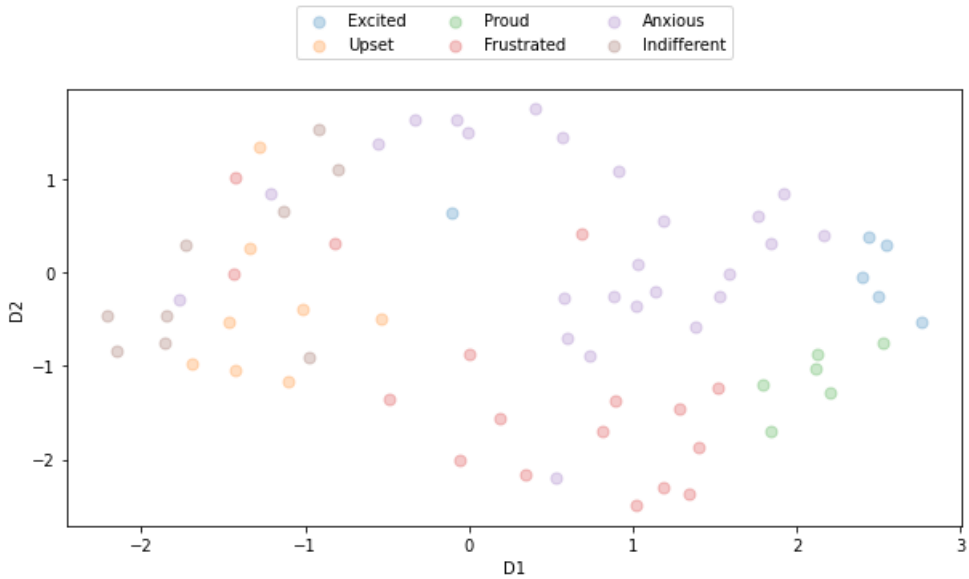

Figure 6 Empathetic Situations Clustering

B Dataset Examples

Below are examples of STEAM and empathetic conversations curated using our prompts.

B.1 STEAM Conversation

Student: Hi Sahar! Today I read about something called Artificial Intelligence. Can you tell me more about it?

SAHAR: Hi there! That's a great question! Artificial Intelligence, or AI for short, is a type of computer technology that allows machines to think and learn like humans. It's pretty amazing stuff!

Student: How do computers think and learn?

SAHAR: Well, just like humans learn by gathering information and experience, computers use algorithms, or sets of rules, to analyze data and make decisions based on that information. It's like a big puzzle, and the computer is trying to figure out the best solution.

Student: That sounds complicated. How do people create AI?

SAHAR: Creating AI requires a lot of programming and data inputting. Scientists and engineers design complex algorithms and build neural networks to teach the computer how to learn and make decisions on its own. It's a lot of hard work, but the end result is something truly amazing!

Student: What are some examples of AI that people use in everyday life?

SAHAR: You may not realize it, but AI is used in all sorts of things we use every day! It's in our phones, our video games and even in our online shopping experiences. AI helps us find the best deals, make recommendations about what to buy, and even detects fraud in our online transactions.

Student: That's really cool! Can I learn more about AI?

SAHAR: Definitely! There is so much to learn about AI, and as you continue to explore STEAM topics, you'll learn even more about this exciting field. Keep asking questions and keep exploring, and you'll be amazed at what you can discover!

Student: Thank you, Sahar! You always make learning so much fun!

SAHAR: You're welcome, kiddo! Keep up the great work and never stop being curious!

B.2 Empathetic Example

Student: Hi SAHAR, I don't know what to do. I feel really miserable at school because some of my classmates keep bullying me and excluding me.

SAHAR: Hi there, I'm sorry to hear that. That sounds really tough. Can you tell me more about what's been going on?

Student: Well, they make fun of me and call me names, and sometimes they push me or take my things. And other times they just ignore me and leave me out of things.

SAHAR: That must make you feel really lonely and hurt. Have you talked to anyone about what's been happening?

Student: I told my teacher, but she didn't do anything. And I don't want to tell my parents because I don't want to worry them.

SAHAR: That's understandable, but it's important to have someone you can trust to talk to. Maybe there's someone else you could talk to at school, like a counselor or another teacher?

Student: I don't know. I'm afraid they won't believe me or they'll think I'm making a big deal out of nothing.

SAHAR: It's valid to feel that way, but you deserve to feel safe and happy at school. It might be worth a try to talk to someone. And remember, you're not alone. There are other people who have gone through similar situations.

Student: Okay, I'll try talking to my counselor. But what can I do in the meantime when I'm around my classmates?

SAHAR: It's not your fault that they're treating you badly. But sometimes it can help to try and be assertive and stand up for yourself if you feel safe to do so. And if that doesn't work, it's okay to try to avoid them and spend time with other people who treat you kindly.

Student: Thank you, SAHAR. I feel a little better talking about it.

SAHAR: You're welcome. Remember, I'm here for you and I care about how you feel.