A Preamble to the use of ChatGPT in Education

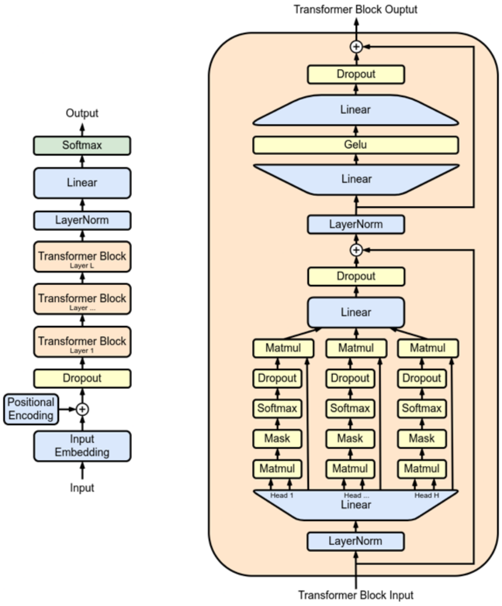
Sergi Bernet Andrés, Guillermo Brugarolas Sobejano, Míriam Calvo Gómez, Álvaro García Piquer, Juan Manuel García Sánchez, Elisabet Golobardes i Ribé, Jessie Caridad Martín Sujo, Antonio Rodríguez de la Torre, Nuria Valls Canudas, Xavier Vilasís Cardona
Smart Society Research Group, La Salle-Universitat Ramon Llull

Abstract: To contribute to the framing of the discussion on the use of ChatGPT in education, a brief explanation on its structure and technology is presented.
Keywords: NLP, ChatGPT, education.

INTRODUCTION

In the last weeks, a new concern has arisen within the teaching community related to the ability of the ChatGPT tool to carry out assignments and exams. ChatGPT is a chatbot from OpenAI Labs, belonging to both profit and non-profit companies under variations of the name OpenAI. A chatbot is a conversational application that simulates human conversation and communicates with end-users via chat. Chatbots have been for a long time in the market mainly used to provide answers about customer service, engagement, and support, and they were based on predefined rules and scenarios. Recently, new neural architectures have boosted the accuracy of. ChatGPT is one of these AI-based chatbots, and it has the ability and accuracy to create automated responses, while responding differently to each interaction, making the use of plagiarism checkers pointless. Many universities seek to anticipate the use of these tools through bans and specific detection strategies (Bracero, 2023; Peirón, 2023). Should the university community really discourage the use of chatbots, or should they integrate them into their daily lives like other tools that were a revolution at the time? In this paper we are going to introduce ChatGPT structure and evolution to ground the discussion on its use in the academic environment.

Figure 1. GPT structure.



CHATGPT EVOLUTION

OpenAI developed ChatGPT, which is a state-of-the-art natural language processing (NLP) model. NLP is a subfield of AI and computer science that deals with the interaction between computers and human languages. The goal of NLP is to enable computers to understand, interpret, and generate human language. NLP tasks include language translation, text classification, sentiment analysis, speech recognition, and language generation. It uses a combination of techniques from computer science, linguistics, and machine learning to accomplish its tasks. State-of-the-art techniques in NLP tasks involve complex deep neural models. The first model from OpenAI, called Generative Pre-Trained Transformer (GPT), was released in 2018 (Radford 2018). Its structure can be seen in figure 1. It relies heavily on transformers, which allow learning the contextual relationships between the words (or subwords) of a text.

Further evolutions of the model were GPT2 (Radford, 2019) and GPT3 (Brown, 2020). Current ChatGPT is based on GPT3.5. The current success of ChatGPT3.5 compared to its previous version is that now it is based on reinforcement learning (RL), which involves providing feedback to the model in the form of rewards of penalties. Specifically, the type of RL used is called reinforcement learning from human feedback, which enhances the RL agent's training by including humans in the training process.

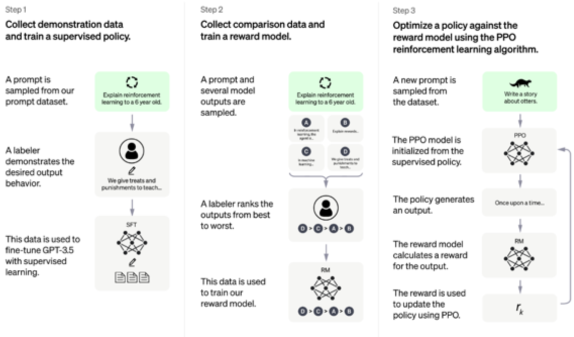
HOW CHATGPT3.5 WORKS?

The training process for ChatGPT3.5 involves three main steps (see Figure 2). First, it trains the model using unsupervised learning in a large dataset of text. This dataset includes books, articles, and websites. The model is trained to predict the next word in a sentence, or a dialogue given the previous words. The second step applies a reinforcement learning algorithm that uses human feedback to fine-tune the model. It allows humans to correct the model's output. Finally, the third step is to carry out the model learning from the provided feedback, updating its parameters so that it does not make the same mistake again. This process allows the model to learn from the feedback and improve its performance over time. It is important to note that the training data and knowledge cutoff for ChatGPT3.5 are from 2021, so the model cannot answer based on any feedback received after that date.

DISCUSSION and CONCLUSIONS

ChatGPT has the potential to be a powerful tool for natural language processing tasks and applications, with the ability to understand and generate human language in a natural way, and the ability to learn and improve from human feedback. However, ChatGPT answers based on the data used in its training, which implies that it is limited to this information. Moreover, ChatGPT was trained with texts from all over the world, past and present. This means that the same biases that exist in the data can also appear in the model, giving answers that discriminate against minorities. Apart from this, it could also spread false information or fake news, as ChatGPT makes even wrong answers sound convincingly correct. Although many of these limitations can be overcome due to its ability to continually learn from human-provided feedback, we cannot forget that ChatGPT answers are based on existing texts, so it does not understand or reason. This opens a new scenario in education, in which it is necessary to change the way of teaching and make pedagogical advances that promote reasoning and critical thinking.

Figure 2. Training process for ChatGPT3.5.


Step 1
Collect demonstration data and train a supervised policy.
A prompt is sampled from our prompt dataset.
Explain reinforcement learning to a 6 year old.
A labeler demonstrates the desired output behavior.
We give treats and punishments to teach...
SFT
This data is used to fine-tune GPT-3.5 with supervised learning.
Step 2
Collect comparison data and train a reward model.
A prompt and several model outputs are sampled.
Explain reinforcement learning to a 6 year old.
A
B
C
D
A labeler ranks the outputs from best to worst.
D > C > A > B
RM
This data is used to train our reward model.
D > C > A > B
Step 3
Optimize a policy against the reward model using the PPO reinforcement learning algorithm.
A new prompt is sampled from the dataset.
Write a story about otters.
PPO
The PPO model is initialized from the supervised policy.
The policy generates an output.
Once upon a time...
RM
The reward model calculates a reward for the output.
The reward is used to update the policy using PPO.
r_k

REFERENCES

Peirón, F. (2023, 6 de enero). Nueva York prohíbe el ChatGPT en sus escuelas. La Vanguardia.

Bracero, F., & Farreras, C. (2023, 15 de enero). ChatGPT revoluciona las aulas de arriba a abajo. La Vanguardia.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), 9.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.