Building a Chatbot for Student Guidance

Hindi Elsa
National Institute of Applied Sciences
Toulouse, France
elsa.hindi@insa-toulouse.fr

Loubejac–Combalbert Jean-Philippe
National Institute of Applied Sciences
Toulouse, France
loubejac-com@insa-toulouse.fr

Bongibault Romain
National Institute of Applied Sciences
Toulouse, France
romain.bongibault@insa-toulouse.fr

Meetoo Anya
National Institute of Applied Sciences
Toulouse, France
anya.meetoo@insa-toulouse.fr

Hamdan Célian Hilal
National Institute of Applied Sciences
Toulouse, France
hilal.hamdan@insa-toulouse.fr

Rousseau Firmin
National Institute of Applied Sciences
Toulouse, France
firmin.rousseau@insa-toulouse.fr

***Index Terms*—large language model (LLM), small language model (SML), machine learning, chatbot, neural networks, transfer learning, energy, data scrapping, natural language processing, deep learning**

I. Introduction

The first conceptualization of the chatbot is attributed to Alan Turing, who raised the fundamental question: "Can machines think?". Through his work on the "Imitation Game" (now known as the Turing Test), Turing laid the groundwork for machines designed to simulate human-like conversation [1]. This early vision has since evolved into the chatbot, a computer program that communicates with people by providing answers to their questions. By processing natural language input—whether speech or text—the chatbot is able to generate intelligent, contextually appropriate responses, simulating a conversation with a human user [2].

Nowadays, the desire for such human-like machine communication is growing across various sectors, including customer service, healthcare, and education, where people increasingly expect seamless, conversational experiences. To reach this level of sophistication, chatbots have evolved significantly. It began with simple systems like ELIZA, created in the 1960s by Joseph Weizenbaum. ELIZA relied on simple keyword matching and predefined responses, which limited its ability to handle complex or varied conversations [3]. While primitive compared to modern chatbots, this system marked an important milestone, demonstrating the potential for machines to engage in dialogue and paving the way for more advanced conversational agents. In 1994, Michael Mauldin coined the term "ChatterBot" to describe these systems, a name that was later shortened to "chatbot" — the term is now widely used to refer to advanced conversational agents like IBM Watson, Amazon Alexa, and Apple Siri [4].

II. Related Work

A. Definitions

The technological advancements of recent decades accelerated chatbot development. Key breakthroughs in Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP) enabled chatbots to evolve from simple keyword-based systems to highly sophisticated conversational agents. ML algorithms allowed chatbots to learn from vast datasets, improving their ability to generate relevant responses. Subsequently, Deep Learning, particularly with neural networks like transformers, enhanced their capacity to understand and generate complex language, while NLP techniques refined their comprehension of grammar, context, and sentiment [5]. These combined advancements have made possible the creation of highly intelligent systems like ChatGPT, capable of engaging in nuanced, human-like conversations across a wide range of topics.

B. Existing LLMs and their architectures

Artificial intelligence models are generally based on *Dense Neural Networks*, which are mathematical models inspired by the functioning of the human brain. They consist of artificial neurons organized into layers: an input layer, hidden layers, and an output layer. Each neuron receives data, transforms it using parameters such as weights and biases computed after training, applies an activation function, and then transmits the result to the next neurons [6].

A Large Language Model (LLM) is characterized by its large size, measured by the number of parameters such as model weights, and its ability to process and generate text in a sophisticated manner. Introduced in 2017 by Google engineers [7], the Transformer was originally designed for machine translation. It is the architecture behind GPT (Generative Pre-trained Transformer) and all other LLMs that are now proliferating in the AI field.

The Transformer architecture consists of two main parts: the encoder and the decoder. It includes *Dense Neural Networks* (Feed Forward) and relies on a key mechanism: attention. The encoder transforms an input sequence composed of symbolic representations (words, characters, etc.) into a sequence of continuous representations which are numerical vectors that represent the properties of the symbols, their meanings, syntactic roles, contextual relationships, etc. The decoder then generates an output sequence composed of symbols (words), one element at a time. At each step, the model is auto-

regressive, meaning it uses the symbols previously generated as additional input to generate the next symbol.

ChatGPT, like other GPT models, is based on millions or even billions of parameters. These parameters enable the model to learn rich and complex representations of language. Table I illustrates the evolution of the number of parameters in OpenAI's GPT models since the release of the first version in 2018 [8] [9].

TABLE I
EVOLUTION OF THE NUMBER OF PARAMETERS SINCE THE FIRST VERSION OF OPENAI CHATGPT [8] [9]

Version	GPT-1	GPT-2	GPT-3	GPT-4
Parameters	117 millions	1.5 billions	175 billions	1.7 trillions

Other architectures for generative AI exist but are currently less efficient or are tailored to specific use cases, such as Recurrent Neural Networks (RNNs). Unlike traditional networks, RNNs have recurrent connections that allow them to retain information from previous states.

C. Training and Fine-Tuning

The training of LLMs, such as those based on the Transformer architecture, relies on several key elements: vast amounts of textual data, powerful computing infrastructure, significant energy consumption, and a skilled workforce for data cleaning and supervision.

Training consists of two distinct phases. The first phase is task-agnostic pre-training, where the model learns semantic representations of words across contexts using self-supervised techniques, such as auto-regressive language models and auto-encoders. This phase utilizes large-scale data, which may include text, text-image, or text-video pairs, to build foundational knowledge. The second phase is fine-tuning, where the model is adapted for specific tasks using smaller domain-specific datasets. This step allows the model to specialize in applications such as classification, structure prediction, or sequence generation, which will enhance its relevance and performance [10].

D. Challenges in Data Security

Interactions between users and LLM-based systems often involve the exchange of sensitive information. Without robust safeguards, this raises concerns about confidentiality, exposing sensitive system prompts or user data due to prompt hacking attacks [11]. Additionally, LLMs are prone to unintentionally memorize sensitive details from training data or user interactions, potentially leading to privacy violations [12]. These risks are exacerbated by three types of prompt hacking [11]: *jail-breaking*, which bypasses the model's intended behavior to extract unauthorized information; *prompt injection*, which manipulates responses through malicious inputs; and *leaking*, which exploits system prompts or pre-loaded sensitive data. Such attacks compromise not only data confidentiality but also the system's reliability.

To address these challenges, privacy-preserving techniques like *differential privacy*, [12] minimize the risk of sensitive data memorization by ensuring that individual data points have negligible influence on model outputs [13]. For further protection, *silos* are used for data compartmentalization [12]. Additionally, *input and output filtering mechanisms* block harmful queries or responses, preventing runtime disclosure of sensitive data [14]. Finally, *robust training methods*, including adversarial training and the use of synthetic data [11] [14], improve model resilience to adversarial inputs and reduce vulnerabilities to prompt hacking.

E. Practical Applications of Chatbots

The practical applications of chatbots across various fields highlight their ability to quickly meet user needs while providing intuitive interactions. Here is a non-exhaustive list of research on their usage in different contexts:

1) Interactive Assistant for Students: The bilingual Student Interactive Assistant [15] enhances student experiences with features like viewing campus maps, setting reminders and providing Q&A support.

2) Chatbot for Cryptocurrency: I&C Chat [2] retrieves real-time prices of top cryptocurrencies and answers queries, simplifying access to financial data in a dynamic industry.

3) Chatbot for Smart Agriculture: A LINE chatbot [16] helps Thai farmers with crop advice and smart irrigation controls, achieving high user satisfaction despite its rule-based limitations.

These examples illustrate the adaptability of chatbots in addressing specific needs, whether by improving the educational experience, supporting financial exchanges, or modernizing agricultural practices. The ongoing development of intelligent chatbots promises to expand their reach and effectiveness across various domains.

REFERENCES

[1] A. M. Turing, "I.—COMPUTING MACHINERY AND INTELLIGENCE," *Mind*, vol. LIX, pp. 433–460, Oct. 1950.

[2] Q. Xie, D. Tan, T. Zhu, Q. Zhang, S. Xiao, J. Wang, B. Li, L. Sun, and P. Yi, "Chatbot Application on Cryptocurrency," in *2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, (Shenzhen, China), pp. 1–8, IEEE, May 2019.

[3] J. Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, pp. 36–45, Jan. 1966.

[4] M. L. Mauldin, "ChatterBots, TinyMuds, and the Turing test: entering the Loebner Prize competition," in *Proceedings of the Twelfth National Conference on Artificial Intelligence (Vol. 1)*, AAAI '94, (USA), pp. 16–21, American Association for Artificial Intelligence, 1994. event-place: Seattle, Washington, USA.

[5] S. Zheng, Z. Yahya, L. Wang, R. Zhang, and A. N. Hoshyar, "Multi-headed deep learning chatbot for increasing production and marketing," *Information Processing & Management*, vol. 60, p. 103446, Sept. 2023.

[6] R. Qamar and B. Ali Zardari, "Artificial Neural Networks: An Overview," *Mesopotamian Journal of Computer Science*, pp. 130–139, Aug. 2023.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," 2017. Version Number: 7.

[8] M. M. Mijwil, K. K. Hiran, R. Doshi, M. Dadhich, A.-H. Al-Mistarehi, and I. Bala, "ChatGPT and the Future of Academic Integrity in the Artificial Intelligence Era: A New Frontier," *Al-Salam Journal for Engineering and Technology*, vol. 2, pp. 116–127, Apr. 2023.

[9] A. Koubaa, "GPT-4 vs. GPT-3.5: A Concise Showdown," Apr. 2023.
[10] R. Li, D. Fu, C. Shi, Z. Huang, and G. Lu, "Efficient LLMs Training and Inference: An Introduction," *IEEE Access*, pp. 1–1, 2024. Conference Name: IEEE Access.
[11] B. Rababah, S. T. Wu, M. Kwiatkowski, C. K. Leung, and C. G. Akcora, "SoK: Prompt Hacking of Large Language Models," in *2024 IEEE International Conference on Big Data (BigData)*, pp. 5392–5401, Dec. 2024. ISSN: 2573-2978.
[12] A. Alabdulkareem, C. M. Arnold, Y. Lee, P. M. Feenstra, B. Katz, and A. Barbu, "SecureLLM: Using Compositionality to Build Provably Secure Language Models for Private, Sensitive, and Secret Data," June 2024. arXiv:2405.09805 [cs].
[13] E.-M. El-Mhamdi, S. Farhadkhani, R. Guerraoui, N. Gupta, L.-N. Hoang, R. Pinot, S. Rouault, and J. Stephan, "On the Impossible Safety of Large AI Models," 2022. Version Number: 2.
[14] T. Nguyen, H. Nguyen, A. Ijaz, S. Sheikhi, A. V. Vasilakos, and P. Kostakos, "Large language models in 6G security: challenges and opportunities," Mar. 2024. arXiv:2403.12239 [cs].
[15] S. Z. Sweidan, S. S. Abu Laban, N. A. Alnaimat, and K. A. Darabkh, "SIAAA-C: A student interactive assistant android application with chatbot during COVID-19 pandemic," *Computer Applications in Engineering Education*, vol. 29, pp. 1718–1742, Nov. 2021.
[16] P. Suebsombut, P. Sureephong, A. Sekhari, S. Chernbumroong, and A. Bouras, "Chatbot Application to Support Smart Agriculture in Thailand," 2023. Publisher: arXiv Version Number: 1.