INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Personality Simulacra

Niranjan Ajgaonkar[1], P Gautama[2], Samuel D Jonathan[3], Dr. Priyanka Bharti[4]

Reva University

Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bengaluru – 560064

Abstract: To develop and meticulously train an advanced language model capable of accurately emulating the personality and voice of a consenting individual. This cutting-edge model will serve as the foundation for creating an innovative platform. Depending on the prevailing needs and standards during the development lifecycle, the platform can take the form of either a desktop application or a web application. Users will have the unique opportunity to engage with a simulation of the specified person, experiencing their distinct personality traits and voice nuances in an interactive and engaging manner.

We aim to develop a virtual platform that faithfully recreates the authentic personas of willing individuals, spanning various professional domains such as technology, medicine, and finance. Through our innovative mechanism, individuals can grant consent to replicate their expertise. This replicated persona can then be accessed by others in their field, enabling the seamless availability of their knowledge and skills, whether they are physically present or not.

All in all, this requires rigorous large language model training through accumulation of perfectly collated datasets and information. It functions as a highly specific branch to a rather generic LLM model.

Keywords: Personality simulation, Large Language Models (LLMs), cognitive architectures, behavior modeling, artificial intelligence.

I. INTRODUCTION

In the contemporary digital landscape, characterized by the omnipresence of technology and the proliferation of online communication channels, the quest for authentic human interaction within artificial intelligence frameworks has emerged as a paramount challenge. Our project represents a pioneering endeavor to bridge this divide by developing an innovative platform that leverages advanced artificial intelligence techniques to replicate the nuanced personalities of consenting individuals. Through meticulous data gathering and sophisticated modeling, we aim to create a system capable of not only mimicking the linguistic patterns of individuals but also encapsulating their unique personality traits and communication styles. At its core, our project seeks to transcend the limitations of traditional chatbots and automated systems by imbuing our model with a deep understanding of human behavior and interaction dynamics. By harnessing the capabilities of Large Language Models (LLMs) and cutting-edge data processing methodologies, we envision a platform that offers users a truly immersive and personalized experience, fostering genuine connections in the digital realm. Moreover, by prioritizing ethical considerations such as informed consent and data privacy, we strive to establish a framework that not only empowers users but also safeguards their rights and autonomy in the digital sphere. The implications of our project extend far beyond the realm of virtual communication, with potential applications spanning a diverse array of industries and sectors. From revolutionizing customer service interactions in the corporate world to providing invaluable therapeutic support for individuals navigating the complexities of grief and loss, the transformative potential of our solution is boundless. By seamlessly integrating human-like conversational capabilities with ethical and user-centric design principles, we aim to set a new standard for AI-driven communication platforms, ushering in a future where technology enhances, rather than detracts from, the richness and authenticity of human interaction. Furthermore, our project underscores the importance of transparency and accountability in AI development, as we navigate the intricate ethical landscape surrounding the replication of human personalities. By fostering a collaborative and inclusive approach to technology, we aim to engage with stakeholders across diverse fields, including psychology, digital health, and data science, to ensure that our platform adheres to the highest standards of ethical conduct and user protection.

II. LITERATURE SURVEY

The paper published by Joon Sung Park, Joseph C. O`Brien and Carrie J. Cai [1] on April 7 2023 called Generative agents gives a brief disclosure to our idea; Believable proxies of human behavior can empower interactive applications ranging from immersive environments to rehearsal spaces for interpersonal communication to prototyping tools.

In this paper, we introduce generative agents--computational software agents that simulate believable human behavior. Generative agents wake up, cook breakfast, and head to work; artists paint, while authors write; they form opinions, notice each other, and initiate conversations; they remember and reflect on days past as they plan the next day.

To enable generative agents, we describe an architecture that extends a large language model to store a complete record of the agent's experiences using natural language, synthesize those memories over time into higher-level reflections, and retrieve them dynamically to plan behavior. We instantiate generative agents to populate an interactive sandbox environment inspired by The Sims, where end users can interact with a small town of twenty-five agents using natural language. In an evaluation, these generative agents produce believable individual and emergent social behaviors: for example, starting with only a single user-specified notion that one agent wants to throw a Valentine's Day party, the agents autonomously spread invitations to the party over the next two days, make new acquaintances, ask each other out on dates to the party, and coordinate to show up for the party together at the right time. We demonstrate through ablation that the components of our agent architecture--observation, planning, and reflection--each contribute critically to the believability of agent behavior. By fusing large language models with computational, interactive agents, this work introduces architectural and interaction patterns for enabling believable simulations of human behavior.

Notably the paper published by Kwadwo Opong-Mensah, titled "Simulation of Human and Artificial Emotion" [2] on November 4th, 2020, called Artificial Emotion Simulation is a rather simplistic and rudimentary view on what we aspire to accomplish; The framework for Simulation of Human and Artificial Emotion (SHArE) describes the architecture of emotion in terms of parameters transferable between psychology, neuroscience, and artificial intelligence. These parameters can be defined as abstract concepts or granularized down to the voltage levels of individual neurons. This model enables emotional trajectory design for humans which may lead to novel therapeutic solutions for various mental health concerns. For artificial intelligence, this work provides a compact notation which can be applied to neural networks as a means to observe the emotions and motivations of machines.

The paper published by Valentin Lungu "Artificial emotion simulation model" [3], is regarding how Research in human emotion has provided insight into how emotions influence human cognitive structures and processes, such as perception, memory management, planning and behavior. This information also provides new ideas to researchers in the fields of affective computing and artificial life about how emotion simulation can be used in order to improve artificial agent behavior. This paper describes an emotion-driven artificial agent architecture based on rule-based systems that do not attempt to provide complex believable behavior and representation for virtual characters, as well as attempts to improve agent performance and effectiveness by mimicking human emotion mechanics such as motivation, attention narrowing and the effects of emotion on memory. To this end, our approach uses an inference engine, a truth maintenance system and emotion simulation to achieve reasoning, fast decision-making and intelligent artificial characters.

Also the paper published by Gerald Matthews a, Peter A. Hancock b, Jinchao Lin a, April Rose Panganiban c titled, "Evolution and revolution: Personality research for the coming world of robots, artificial intelligence, and autonomous systems" [4], discusses directions for future personality research. Cross-cultural research provides a model, in that both universal traits and those specific to future society are needed. Evolution of major "etic" trait models of today will maintain their relevance. There is also scope for defining a range of new "emic" dimensions for constructs such as trust in autonomy, mental models for robots, anthropomorphism of technology, and preferences for communication with machines. A more revolutionary perspective is that availability of big data on the individual will revive idiographic perspectives. Both nomothetic and idiographic accounts of personality may support applications such as design of intelligent systems and products that adapt to the individual.

III.   POSITIONING

1) *Problem Statement:* Developing a virtual platform capable of faithfully replicating the authentic personas of consenting individuals poses a significant challenge. The objective is to achieve precise accuracy in mirroring real-life personalities, ensuring an ethically sound framework based on informed consent. Moreover, the challenge extends to the potential scalability of this platform, especially when adapting it to cater to specific industry needs. Balancing the intricacies of individual consent, technological precision, and industry-specific customization creates a complex problem that demands innovative solutions and ethical considerations. In the corporate world, this may cause legality issues based on the personae used. Hence we have decided to cater to a singular business stream that admits full consent during its usage and is aware of its customary interactions.

2) *Product Position Statement:* A consenting person from any field (tech/it, medical, finance etc), can have his/her persona replicated and reproduced using this mechanism, built by us. Their expertise can then be readily available to the respective individual in their absence or presence. A working solution that is beneficial to the general public, is its usage in the medical field, as a consultant for patients during the clinic doctor's absence. Unlike a regular chatbot, this provides a nuanced human-like experience like none other.

3) *Stakeholder Descriptions:* Currently, the individuals and organizations that can be targeted are seemingly endless. This works even as a live project deployed on multiple different servers. However, for the sake of formalities, we have decided to target the medical field/business as an environment to expand on our application. Qualified doctors can pass on their expertise through this model, for automating general consultancies, hence clinic's can make up the involvement phase as well as the market phase.

a) *Primary Stakeholders:* Since, the target stakeholders are hospitals/clinics, the user stakeholders are the patients who require general consultancy over trivial matters. The model can be scaled up to accommodate larger diagnosis, but this may be a matter of concern for most patients who try not to compromise their health with such matters.

b) *User Stakeholders:* These include patients who have suffered through a loss of a loved one/family member. The project may act as a therapeutic converser by capturing their style of talking and personality, therefore immortalizing it.

IV. PROJECT OVERVIEW

Our goal is to create a platform that faithfully reproduces the genuine personas of individuals who willingly participate, spanning diverse professional sectors such as technology, medicine, and finance. Through our innovative approach, individuals can provide their consent to replicate their expertise. This replicated persona becomes accessible to others within their field, ensuring the seamless availability of their knowledge and skills, regardless of their physical presence.

We plan to develop and train a sophisticated language model capable of precisely emulating the personality and voice of consenting individuals. Leveraging this model, we will establish a tailored platform, adaptable as either a desktop application or a web-based solution, designed to meet specific requirements during the development phase. Users will have the opportunity to engage with a simulated version of the consenting individual, replicating their [6]distinctive personality traits and voice characteristics in an interactive and immersive manner.

1) *Objectives:* Develop and train an advanced language model to accurately mimic the personality and voice of a consenting individual. Utilize this model to establish a platform, either as a desktop or web application, tailored to specific requirements during the development phase. This platform will enable users to interact with a simulated version of the consenting individual, replicating their unique personality.

2) *Goals:* Create and refine a sophisticated language model capable of precisely replicating the personality and voice of a willing participant. Employ this model to create a platform, be it a desktop application or a web-based system, customized according to specific requirements in the development phase. [5]This platform will empower users to engage with a virtual representation of the consenting individual, mirroring their distinct personality and voice characteristics for an immersive interactive experience.

3) *Feasibility Study:* This completely depends on the extraction, availability and processing of the required data fed into the language model, so as to train it to mimic certain personalities. This may be the longest step in achieving perfection in the model, as the rest of the platform only depends on hard-coding. The model used may change the feasibility of the entire project, as GPT-4 doesn't come in cheap. Even for generation of text, some LLMs have a significant recurring cost. Hence such models may only be used, if the LLaMa 2 model is insufficient or poses unforeseen roadblocks.

4) *Alternatives:* Variations are limited to the choice of Large Language Models (LLMs) and the voice encoding methods we employ. Presently, the available LLM options encompass LLaMa 2, GPT-4, PaLM, and BERT. The application of this model is multifaceted, with numerous focal areas and limitless possibilities, contingent upon the approval of the designated committee for voice and persona cloning in automated human responses.

5) *Budget*

a) The budget is a variable factor completely dependent on the Large Language Model used. LLMs such as GPT-4 have a 0.03\$ budget for 1K tokens used. Other models have similar pricing models. The only LLM which does not require a subscription is LLaMa. Unfortunately the working requirements for LLaMa are extremely vague and rather time consuming. The other factor is the availability of a mid-range GPU, on which the model is trained.

b) Another large chunk of the budget would be going towards renting a compute instance (cloud GPU), to run the entire large language model, as local GPUs may suffer through a system crash.

6) *Necessary Materials:* The model and requirements may be satisfied through an estimated specified budget. However, since training such models is an extremely resource intensive project, a setup/workstation with high specs may be required for the core processes. This workstation can be made to run in the background for training the LLM using our datasets. Such language models require a strong non-mobile GPU, which is one of the biggest current requirements for the functioning of this project. Other software modules include LangChain using Python for the most part, and preferably the LLaMa 2 model.

V. METHODOLOGY

The methodology includes creating and refining a sophisticated language model capable of precisely replicating the personality and voice of a willing participant. To then employ this model to create a platform, be it a desktop application or a web-based system, customized according to specific requirements in the development phase. This platform will empower users to engage with a virtual representation of the consenting individual, mirroring their distinct personality for an immersive interactive experience.

This completely depends on the extraction, availability and processing of the required data fed into the language model, to train it to mimic certain personalities. This may be the longest step in achieving perfectness in the model, as the rest of the platform only depends on hard-coding, once the dataset has been secured, the model is trained.

QLoRA[7] is a highly efficient approach to fine-tune an LLM that allows you to train on a single GPU, and it's become ubiquitous for these sorts of projects. It combines 2 techniques:

1) Quantization, i.e. using an LLM with weight precision reduced to 4 bits

2) And, LoRA (low-rank adaptation of LLMs) which freezes model weights and injects new layers, vastly reducing the number of weights to be trained.

Configuration of the hyperparameters[8] in the QLoRA implementation is an unstable step, and may require extra configuration based on the user's set and needs.

The model used may change the feasibility of the entire project, as GPT-4 doesn't come in cheap. Llama 2 - 7b is the perfect model which can provide a balanced experience throughout the entire training process. Huggingface consists of multiple versions of this model suited to train our dataset and fine tune it even more.

A. This will Occur in 4 Separate Phases

1) Phase 1: Considerable time will be invested in gathering data, which is likely to be sourced internally, providing a robust yet accessible foundation for our project. The task of compiling and structuring this data falls within our responsibility, given its entirely innovative and project-specific nature. We have gone through our entire message log history on Discord, Instagram and other P2P applications, and collated all of them into 1 file for processing. A python script which converted all of this raw data into readable format was run, to expand its functionality. The data was grouped based on the sender and receiver and the type of response one receives after a certain input.

2) Phase 2: Subsequently, the model will undergo training tailored to this unique dataset. Large Language Models (LLMs) inherently possess a broad scope, making the training process time-intensive. Utilizing Llama 2, we can streamline the development of an application specific to this model iteration. A specific text model was trained on this dataset crafted previously. All of the parameters and nodes were configured to suit the needs of the input conversation and the desired generative output. Training an LLM requires a considerable amount of GPU power, hence different methods were utilized (local training/cloud training). These methods required some amount of funding including rent costs, and model costs (Both open source and premium models were used).

3) Phase 3: Once the model was trained, it was connected onto our own Web Text Generator UI, for prompts and responses. Colab's T4 GPU was able to generate a trained model onto a HuggingFace account within 8 hours of training, whilst a local RTX 2060 SUPER took over 13 hours. Hence a rented cloud GPU was utilized, to not run out of session limits whilst training the model.

4) Phase 4: The result now albeit a little unfiltered with its language, was able to generate accurate responses, based on the individual's online persona. Multiple prompts relating to our daily events, and general events were submitted to the model, and it came up with responses that mirrored our actual conversation.

VI. MODULES IDENTIFIED

A. Data Source

The key modules consist of the data that will be sourced and created by the team members, in regard to training the model for a considerable amount of time. This phase consists of PyTorch, and HuggingFace as the base piece of technology for creating and loading an entire script for the project-based data. A full sectioned conversation from P2P chat applications serve as a decent starting point for acting as datasets.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue VI June 2024- Available at www.ijraset.com

B. Model Selection

Once the model has been sourced from HuggingFace, we can use LangChain for creating an application around this trained instance of the model, which allows us to integrate the project into the LLM infrastructure. The model itself may go 2 separate ways, one of them being the open source Llama 2 model, and the other being the enterprise GPT-4 model. This depends on availability of resources and budget.

C. QLoRA implementation along/ Training Sequence

The script trains the model for a surmountable amount of time, and automatically uploads it privately to HuggingFace, with checkpoints for code interruptions. This would end up being a completely unique model based off the Llama 2 7b LLM. This completely depends on the extraction, availability and processing of the required data fed into the language model, so as to train it to mimic certain personalities. This may be the longest step in achieving perfectness in the model, as the rest of the platform only depends on hard coding, once the dataset has been secured, the model is trained.

D. Text Generator UI

A Web Text Generator UI will be used to upload the model, for experimentation and accuracy testing. Parameters for RAM and storage requirements can be configured within the text generator itself, as the model runs on the host's system using their resources.

VII. PROJECT IMPLEMENTATION

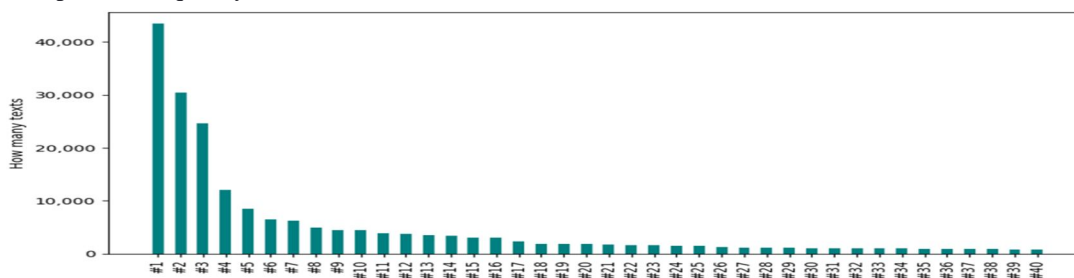*1) Initial text chat phrasal frequency:*

Fig.1. Text Chat Phrasal Frequency

The senders have been hashed and their message frequency has been plotted.

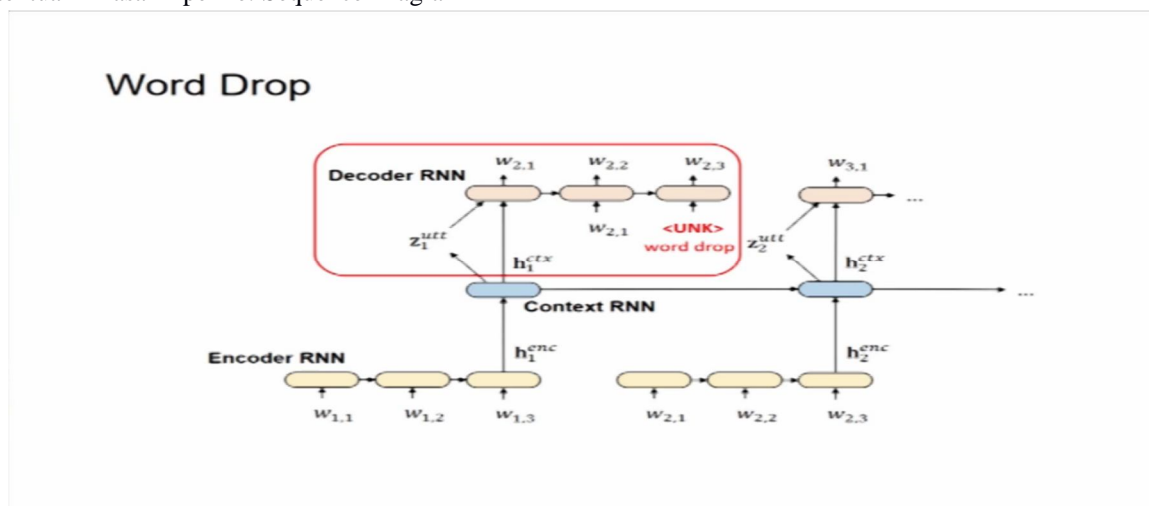*2) Contextual Phrasal Pipeline: Sequence Diagram*

Fig.2. Sequence Diagram

3) Class Diagram

This section consists of graphical data captured by Wandb.ai, whilst running the model. Most features have not been included to secure local system privacy
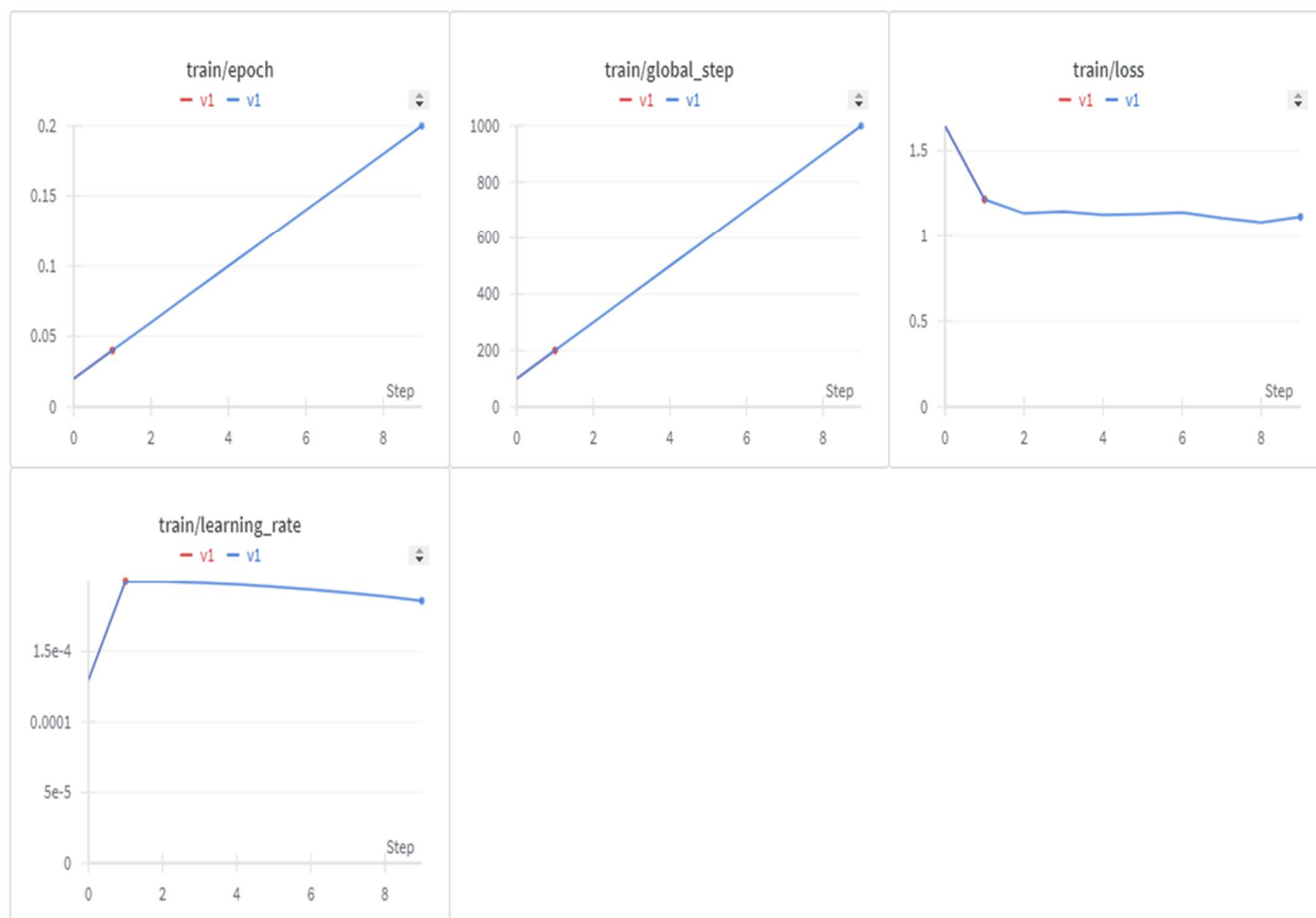

Fig.3. Model Training, loss and learning rate.

After training for 1 epoch the 2nd epoch resulted in significant over-fitting immediately with eval loss leaping up.

Fig.4. Training and runtime of the model

The runtime tabular statistics were returned after training for a total of 1000 steps, each with a checkpoint of 500. Noticeably the second checkpoint's runtime was significantly longer than the first.
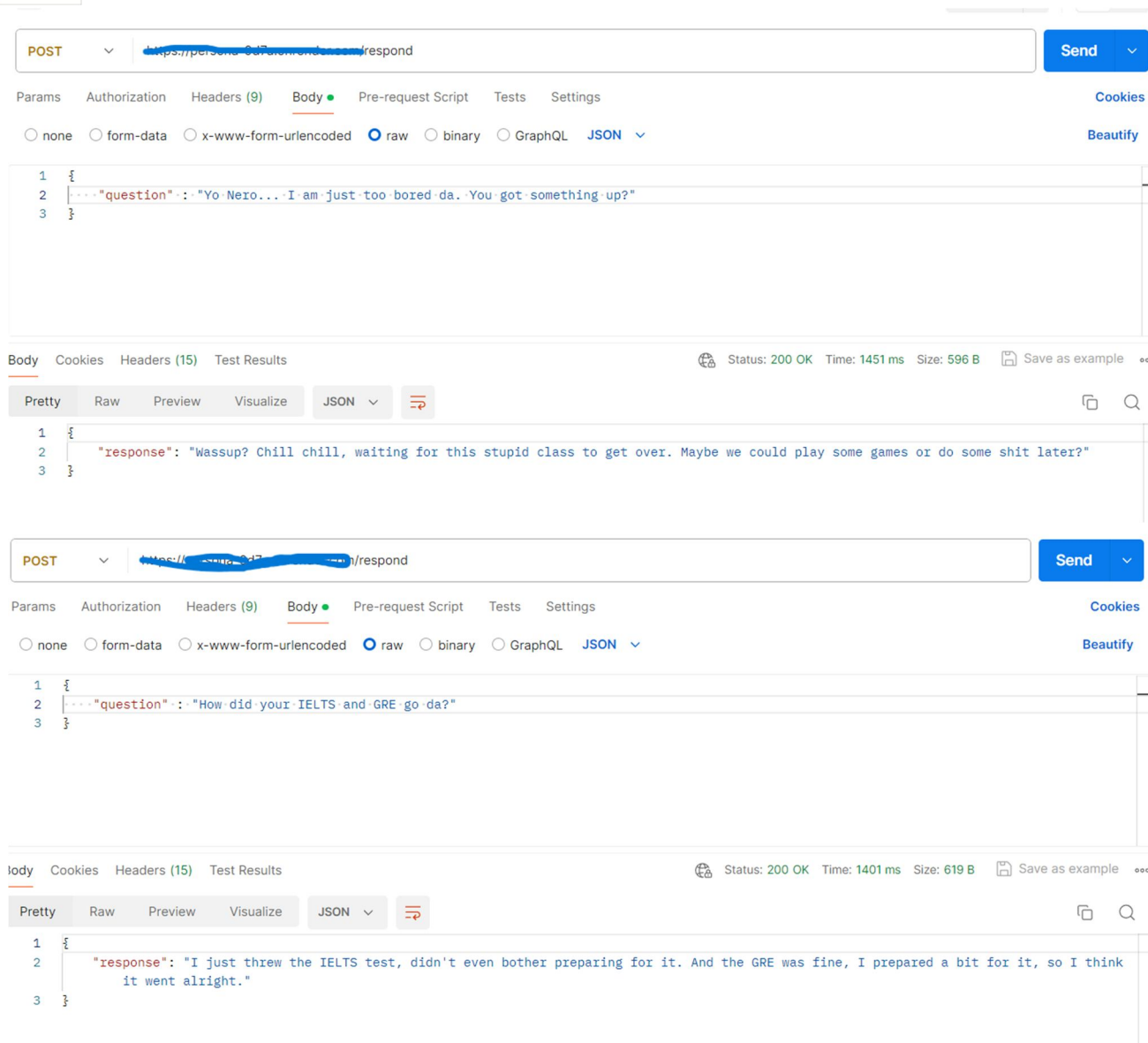
VIII. PROPOSED SOLUTION

The outcome/solution obtained from this project has prospects in almost every sector be it "big tech", medical or psychological. A fully communicating model that acts like a specific person is obtained after a continuous and successful run. This can then be catered to behave like any desirable human, or any desirable characteristic portrayed by the masses. Its responses are much more stable and specific, when compared to raw LLMs that are simply prompted to act like someone else.

1) This is a rather sizable upgrade to the current status website chatbots, that end up returning back to the same responses when pushed to the limit.

2) Another prospect includes secretarial work, which may take some burden off a full time secretary/nurse working in hospitals and other companies

3) This project may also act as therapy/relief for individuals who have lost a loved one, and wish to communicate with them, further immortalizing their personality.

Fig.5. Deployed model Replies for questions.

IX. CONCLUSION

This project has a good amount of originality to it, as most other variations only exist in the form of unapplied research or as visual interactions between two artificially designed personalities. Accomplishing full fledged use of this technology is also rather simple for a common person and does not require an unnecessary amount of setup for their own purposes. The model adopts the tone and characteristics of individuals, even with everyday topics. This was one of our stress tests that were applied to find out the growth and scope of the trained model. The only roadblock that may occur whilst utilizing this project is acquiring consent, for training purposes, but since the model that is generated using the code is uploaded privately to your account, this may not be an issue in the long run.

The project, hence, tries to accomplish a simple, yet distinct goal in the field of Generative AI modules, where the power of LLMs can be leveraged to suit anyone's specific or non-specific needs and wants.

X. FUTURE SCOPE

Currently, this project may be termed as a glorified chatbot that emulates real people, but there are a lot of future prospects that can be applied to this project that can widen its functionality even more.

1) One such addition includes, real time voice cloning, which can easily be implemented onto the model prompt, to read out the generated messages, using that specific individual's voice. Hence the model would be trained on that person's style of speaking, and their voices. Voice emulation is not exactly a novel concept as of now, but it may provide some amount of nuance to this project

2) Another extension includes a fully operational and encrypted website, where people can directly upload a raw version of their sources. This data would then be processed and then a model would automatically be generated for private or public use. However the current status of the project only extends to an offline desktop application.

3) A final stride would be to increase the vertical scaling of the model to output even more accurate responses. The current Llama 2 7b model is the weakest one out of the lot, and can definitely benefit from extra power provided by 35b or even 70b models. Such models are just as expensive as they are powerful.

REFERENCES

[1] Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, Michael S. Bernstein "Generative Agents: Interactive Simulacra of Human Behavior" April 7, 2023

[2] Kwadwo Opong-Mensah "Simulation of Human and Artificial Emotion" November 4, 2020

[3] Valentin Lungu "Artificial emotion simulation model"July 4, 2010

[4] Gerald Matthews a, Peter A. Hancock b, Jinchao Lin a, April Rose Panganiban c, Lauren E. Reinerman-Jones a, James L. Szalma b, Ryan W. Wohleber "Evolution and revolution: Personality research for the coming world of robots, artificial intelligence, and autonomous systems" February 1, 2021

[5] Jemine, Corenti, Louppe, Gilles "Automatic Multi Speaker Voice Cloning"2019

[6] Ye Jia, Yu Zhang, Ron J. Weiss, Quan Wang, Jonathan Shen, Fei Ren, Zhifeng Chen, Patrick Nguyen, Ruoming Pang, Ignacio Lopez Moreno, Yonghui Wu "Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis" Jan 2, 2019

[7] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen "LoRA: Low-Rank Adaptation of Large Language Models" 17 June 2021

[8] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, Luke Zettlemoyer "QLoRA: Efficient Fine Tuning of Quantized LLMs" 23 May 2023

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)