1. Introduction

With the advancement and inspection of intelligent hardware and software, also known as intelligent agents, artificial intelligence(AI) is steadily embodied in our daily lives. From manual labor to intricate operations, intelligent agents are proficient in carrying out an extensive range of tasks. One of the greatest fundamental and well-known examples of intelligent Human-computer interaction, a Chabot is a typical example of an AI system.

**BACKGROUND**

1.1

By substituting or enhancing human support agents with artificial intelligence(AI) and other automation technologies that can communicate with end users through chat. A Chabot is a conversant program that integrates or enhances support from human agents with customer service and support.

Joseph Weizenbaum, an American designer, built ELIZA in 1966. It seemed like a straightforward chatbot that was capable of responding to the user's data and was able to fulfill the Turing evaluation, which interprets whether machinery can imitate interaction with humans. ELIZA, on the contrary hand, was forced to shut down after it realized that the program had numerous flaws in programming including the fact it frequently responded to queries that lacked a conceivable solution.

The inaugural chatbot successfully completing the Turing Evaluation was given the acronym ALICE, or Automatic Linguistic Computerized Examiners. In 1995, Joseph Weintraub and David Smith were the people who created it. ALICE sought to interface with other individuals online by impersonating a person. It would be considered to have fulfilled the Turing Evaluation if a person couldn't tell whether they were communicating with an artificial intelligence or someone else. The name "ALICE" was motivated by Alice in Wonderland since, during one specific interaction with another individual, this bot's anticipation and speech pattern mirrored those of the character from Alice in Wonderland. The core intellect of ALICE focuses on the framework known as Artificial Intelligence mark-up language (AIML), which enables programmers to articulate the fundamentals of chatbot expertise [10]. ALICE uses an intuitive pattern-matching method. Chatbots have been generated and rendered accessible by using messaging apps, such as Smarter Child in 2001. Virtual assistants like the ones made by Apple, Cortana by Microsoft, the Alexa voice assistant from Amazon, Assistant from Google, and Watson, from IBM,  were developed as a subsequent phase.

1.3.Types of chatbot

Fig:

**1. keyboard chatbot:**

In contrast to menu-based chatbots communicate keyword-based chatbots are able to cater to what users enter before reacting effectively. These virtual assistants implement configurable vocabulary and NLP (natural language processing) technology to identify response prompts in the argument to figure out how they react to the users. If the key phrase analysis functionality breaks or people necessity additional support when seeking the correct reply, such bots for conversation have the benefit of simply offering directives accessible through clicking navigational skills push buttons. When the automated system fails to understand keywords in the manually filled-in feedback, this is a superior solution.

**2. hybrid bot:**

fortunately, a hybrid chatbot doesn't appear to be one in the most literal sense, the word "hybrid" has been adopted to indicate chatbots featuring an active agent handing over function. The agent hand-off is a fundamentally necessary chatbot functionality since, as demonstrated previously, chatbots won't compete with human brain power. When creating Zobot, our rule-centered automated systems medium, and Solution Bot, our artificial intelligence chatbot developer. which utilizes your expertise to respond to client issues, we continued to keep the aforementioned features in the mind.

**3. Voice bot:**

Voice bots comprise autonomous application initiatives that embrace voice methods for resembling interpersonal interaction between humans. Artificial intelligence (AI) and software for natural language processing (NLP) position of authority voice bots. By means of vocal associations, they are permitted to fully comprehend plus handle comments from users, produce pertinent replies, and carry out simple tasks.

Voice bots have the purpose of providing unparalleled client satisfaction for commercial organizations such as the finance sector, retail level and electronic shopping, and financial institutions to mention a few. Voice bots like the one created by Apple, Alexa, the virtual assistant from Amazon, and Google Assistant, among others, are prominent instances.

**4. machine learning:**

Service personnel employed exceedingly basic software schemes called chatbots that allowed individuals to plan solutions for particular inquiries. Individuals programmed chatbots that responded to certain of their most typical inquiries that clients may ask. As a result, every single time one of the queries popped up through the chatbot, it would inevitably search among the specified information and answer.

The chatbot's functionality was extremely restricted by those pre-written remarks. As a result, if an audience member asked an inquiry that couldn't be answered by the predetermined phases answers to frequently asked questions the chatbot was unable to respond, and the communications were transmitted directly to an actual consumer service staff.

**5. linguistic chatbot:**

Linguistic-based — frequently described as rules-based chatbots give users the optimized control as well as adaptability that machine-learning-based chatbots lack. The appropriate response to a query may be assessed in advance, and scripted tests can be generated to measure the device's reliability. Language instances are capable of being erected to examine terms, their order of occurrence, substitutes, and typical approaches to arrange a question, along with additional factors that will guarantee inquiries containing a comparable indicating acquire the same response. Human beings might change their surroundings if whatever is not completely consistent with their perception. Due to this incredibly laborious strategy, chatbots developed using an entirely linguistic framework could still be fragile and unsteady as they grow. Despite the truth that these automated devices employ Processing Natural Languages, their conversations with them have become extremely unique and constrained. These chatbots often seem like reactive FAQs while possessing simple functions.

**6. Menu chatbot:**

Perhaps the most simple and readily accessible sort of chatbots can be controlled using a menu or button design. These chatbots consistently have an inherent comprehension bookstore, and inquiries are given to the user as buttons. A button's bot engages in compliance with established standards. For the sake of getting a reaction, data submitted by the customer needs to comply with particular standards that have been established. A candid conversation with a bot of the sort is either impractical or extremely scarce because intelligence from machines is not employed here. Nevertheless, they are adequate to respond to commonly requested questions, which account for 80% of requests.

1.4. chatgpt:

The method of natural language processing is accomplished by ChatGPT, an artificially intelligent chatbot to produce a conversational discourse that sounds human. The computational simulation of language has the flexibility of reintroducing answers to queries and setting up a variety of pieces of writing, such as posts on blog sites, updates from social media, essays, code, as well as electronic mail. In any event, artificial intelligence (AI) methodologies are here for the remainder and have really improved significantly since they were revealed a handful of months prior to this. Major businesses are attempting experiments with it, especially when employing the computer's brain to develop marketing and advertising material. And the OpenAI initiative is spending a lot of money on it. GPT-4, an innovative language-writing model from OpenAI's research institutions, just gave ChatGPT a boost. GPT-4, which can write spontaneously and eloquently than the model that beforehand enabled ChatGPT, is accessible by those who are paying for ChatGPT. In conjunction with GPT-4, the OpenAI initiative has lately linked ChatGPT to the wider web alongside extensions for plugins that have been released now in alpha along with accessibility for users as well as those on the whitelist.

ChatGPT could have linked retroactively into the formation of GPT (Generative Pre-training Transformer), an artificial intelligence language abstraction pioneered by the OpenAI organization in 2018. GPT was created in order to determine the following word in a specific combination of sentences while learning pertaining to the syntax and conventions of human-generated written communication from a sizable collection of text. A wide range of programs, encompassing language creation, computerized interchange, and possibly music composition, quickly began using this methodology because of its speedy accomplishment. With its technology fused with AI-driven machine learning for language processing and GPT, CHATGPT is the inaugural chatbot medium that delivers precise and realistic responses. Additionally, it possesses the capability to acquire knowledge and comprehend intricate dialogues, rendering it an influential asset for corporations seeking to mechanize their customer service operations. Expanding upon the triumph of GPT, the term at OPENAI endeavored to develop a conversational bot that could engage in organic exchanges with humans. The outcome of this was the creation of ChatGpt, which became available at the start of 2020.ChatGpt has been subject to several revisions and enhancements, eventually emerging as one of the most sophisticated chatbots currently accessible

1.5. how does it work?

The operational paradigm employed by ChatGPT involves the interpretation of the given query, followed by the provision of a series of lexicon elements that are projected to effectively address the specific inquiry. This process is dependent on the informational corpus used for the system's original conditioning.

A comprehensive discourse pertaining to the training is deemed necessary. The aforementioned procedure encompasses endowing the nascent Artificial Intelligence (AI) with certain fundamental principles, subsequently either subjecting it to diverse scenarios or furnishing it with sizeable quantities of input data, thereby facilitating the evolution of customized algorithms. ChatGPT is a modern language model that has been developed by OpenAI, featuring impressive rise improvements over its predecessor, GPT-3. When compared to several other Large Language Models, ChatGPT possesses the ability to generate text for various purposes and in different styles. Nevertheless, what sets it apart from its counterparts is its exceptional accuracy, attention to detail, and coherence among features. The present model signifies the progression in OpenAI's continuum of comprehensive language models, developed with a noteworthy emphasis on receptive dialogues. By using approximately 500 billion "tokens" in its training process, GPT-3 has acquired a skillful aptitude to recognize significance within the written language and accurately predict forthcoming semantic components. While longer or intricate words may split into several tokens, there are still numerous words that are equivalent to singular tokens. Typically, tokens comprise approximately four characters. In order to improve the performance of ChatGPT, the developers employed a blend of Supervised and Reinforcement Learning methodologies. Despite all, it is the singular element of Reinforcement Learning that confers unprecedented characteristics to ChatGPT. By incorporating human input into the training process through a distinctive approach called RLHF, the authors aim to minimize unfavorable, deceitful, or inadequate outputs. In the domain of machine learning, the term capability pertains to the efficacy of a model in executing a particular task or a collection of tasks. The performance of a model is commonly assessed based on its proficiency in optimizing its objective function. This objective function is presented as a phrase in mathematics. that outlines the overall aim of the model. As an illustration, a mathematical model purposed to anticipate stock prices could incorporate an objective function that evaluates the level of precision demonstrated by the model's estimations. Should the model exhibit a capacity to precisely anticipate the trends in stock prices across a given span, it would be deemed to possess a significantly elevated level of proficiency in this realm.

The step-by-step method can be categorized into three fundamental phases:

1. Supervised fine-tuning: The supervised fine-tuning method involves the refinement of an already existing language model using a small set of data for performance that is carefully selected and engraved by mankind annotators. By following this process, one can acquire a monitored strategy referred to as the SFT paradigm, which facilitates the production of outcomes from a carefully selected array of cues. This indicates the fundamental prototype.

2. Mimic human preference: Those who evaluate are tasked with the duty of deliberating on a significant quantity of information generated by the SFT framework, culminating in the development of a fresh dataset to be used for evaluation. The current set of data has been utilized to instruct an original and pioneering framework. In academic literature, the construct that is being mentioned is commonly identified as the reward model (RM).

3. Proximal policy optimization: The compensation display is employed to refine and enhance the SFT model. The outcome of this phase is known as the display layout.

The initial step is undertaken singularly, whilst the subsequent two phases may be repeated indefinitely. Additional comparative data is acquired with respect to the present optimum solution, which is subsequently utilized in generating an updated recompense model and a novel arrangement.

A detailed explanation of the step in the diagram below:

STEP1: The Supervised Fine-Turning(SFT) model:

To begin with, the initial step involves gathering informative data in order to form a structured demonstration approach known as the SFT demonstration.

Data collection: To collect data, a set of cues is chosen and a team of human evaluators are tasked with documenting the anticipated response outcome. ChatGPT has utilized two distinct prompt avenues including ones created by labelers or developers and others obtained from OpenAI's API requests. as acquired from their clientele using GPT-3 technology. Due to the gradual and costly nature of this procedure, the outcome is a small but refined dataset (expectedly containing 12-15k observation points) that will serve as the basis for improving an already established language model.

Choice of model: In lieu of meticulously optimizing the initial GPT-3 exhibition, the developers behind ChatGPT instead opted for a preexisting demonstration housed within the GPT-3.5 architecture. Evidently, the prevailing exhibit employed is the latest iteration, namely text-DaVinci-003, which is a GPT-3 model that has been substantially refined for processing programming codes.

The designers' avid interest in developing a conversational agent akin to ChatGPT led them to opt for the refinement of a "code model" as opposed to a pristine content display.

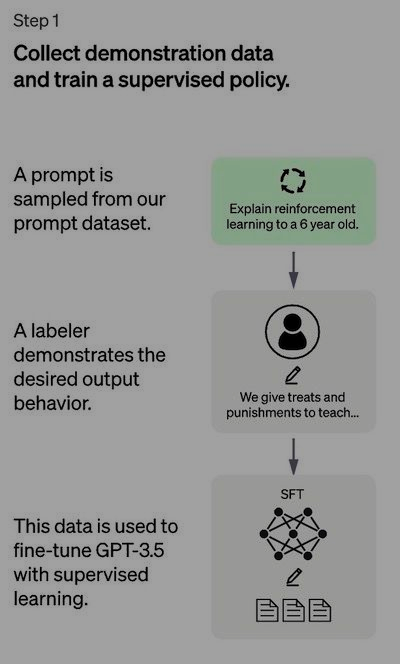


Fig1.

Owing to the limited corpus for this organization, it is conceivable that the Comprehensive Utilitarian Hypothesis (SFT) demonstrates that produced post this strategy may surrender content that's insufficiently mindful to the client, and is inclined to misalignment, as illustrated within the preceding segments. The display challenge lies within the reality that the directed learning stage experiences noteworthy challenges in terms of adaptability costs.

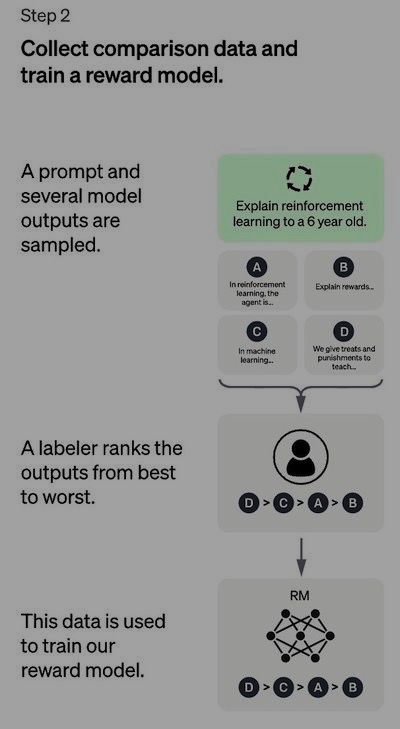
To address this predicament, the approach embraced is moving absent from the strenuous and time-consuming assignment of enrolling human labelers to produce an essentially bigger curated dataset. Or maybe, the current suggested technique involves enrolling labelers to rank different yields of the SFT demonstration, thereby creating a remunerate demonstration. This prepare might be explained broadly within the following segment.

### Step 2: The reward model (RM):

The goal is to commit to memory of a particular job's impartial (the pay presentation) solely based on the data provided. The purpose of this task is to assign a rating to the SFT exhibition's outputs based on their attractiveness to individuals. This will indisputably mirror the distinct preferences of the designated group of individuals assigned to labeling, as well as the agreed-upon general guidelines they are expected to adhere to. Afterward, this strategy will derive an automated structure from the gathered data that mimics human tendencies.

Typically the way it works:

* The SFT demonstration produces an extent of yields, ordinarily yielding around 4 to 9 diverse reactions, in reaction to a set of chosen prompts.
* The labelers organize the comes about in arrange of quality from most noteworthy to most reduced. A recently labeled dataset has been accomplished with rankings being allocated as the names. The greatness of this dataset is generally 10 times more noteworthy than the custom-fitted dataset utilized for the SFT demonstration.
* The later data is utilized in teaching a compensate demonstrate (RM) cleverly. The strategy includes utilizing a determination of SFT show comes about and organizing them agreeing to inclination.

****

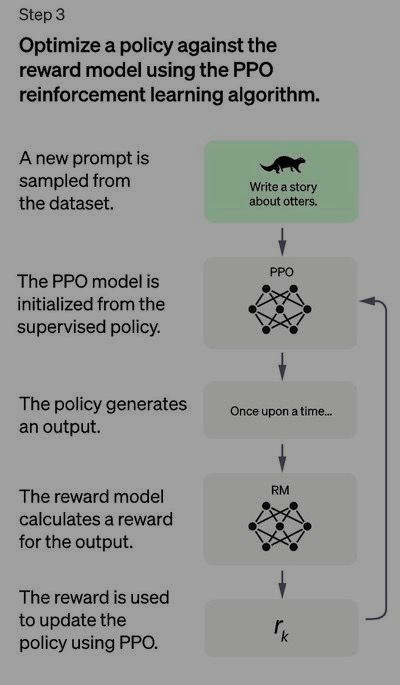
Regarding labelers, it is notably less complex to evaluate the outputs in comparison to producing them from the very beginning, thereby leading to a more effective scaling process. The dataset in question was derived from a deliberate selection of 30-40 thousand prompts, wherein multiple outputs, generated in varying quantities per prompt, were presented to each labeler during the assessment phase.

### Step 3: Fine-tuning the SFT model via Proximal Policy Optimization (PPO):

* The PPO methodology provides a crucial computational framework for training operators in the domain of reinforcement learning. The current method of learning in which the policy is directly upgraded from ongoing experiences is referred to as an "on-policy" calculation. This is in contrast to the "off-policy" calculations such as Deep Q-Network (DQN), which rely on past experiences for learning. This indicates that the PPO system is consistently adapting its current methodology in response to the actions taken by the expert and the benefits obtained.
* The utilization of a belief region optimization method by PPO in its operational practices is noteworthy, as it involves an approach that mandates that any alteration in the policy framework be restricted to a certain proximity range of the prior arrangement, thereby ensuring stability. This phenomenon is commonly observed in contrast to alternative strategies regarding slope, which may occasionally entail extensive overhauls of the system that could potentially impede the process of acquiring knowledge.
* PPO (Preferred Provider Organization) utilizes a valued method to evaluate the projected outcome of a particular condition or undertaking. The term "esteem work" is employed in order to calculate the work derived from benefits, which denotes the disparity between the expected yield and the prevailing yield. The term "advantage work" is commonly employed to improve the approach through a comparative analysis of the current course of action with the corresponding action that would have been taken under the prior strategy. This allows for PPO to make better-informed improvements to the system through the evaluation of the value of the undertaken actions.

In this procedural move, the Proximal Arrangement Optimization (PPO) demonstration is set up from the Secure Record Exchange (SFT) demonstration, whereas the esteemed work is initialized based on the compensation show. The environment in address is one characterized by its eccentric nature, whereby it issues a provoke of subjective nature and expects a comparing reaction. Upon receipt of a given incite and comparing reaction, a comparing remunerate is created, as decided by the remunerate demonstrate, coming full circle within the conclusion of the scene. The SFT demonstration forces a per-token KL punishment on the remunerate demonstration, subsequently easing the chance of over-optimization.

2. literature review:



References :

For introduction <https://link.springer.com/chapter/10.1007/978-3-030-49186-4_31>

29 may 2020 springer,cham

* <https://www.revechat.com/blog/chatbot-history/> Updated January 12th, 2023 juwel rana

for types of chatbot:

<https://tovie.ai/blog/types-of-chatbots-for-business> June 7 2022- diana kisling

<https://www.zoho.com/blog/salesiq/hybrid-chatbot.html> august 10,2022-Divya R

<https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-chatbot/> updated june 17,2022 -chiradeep

chatgpt: <https://www.techtarget.com/whatis/definition/ChatGPT> -amanda hetler

<https://techcrunch.com/2023/04/13/chatgpt-everything-you-need-to-know-about-the-ai-powered-chatbot/> Alyssa stringer- april13,2023

history: <https://www.electrodedigital.co.uk/chatgpt-history-and-future/>

#### <https://digitalscholar.in/history-of-chatgpt/> March 16, 2023 - Digital Scholar

#### how it work <https://www.atriainnovation.com/en/how-does-chat-gpt-work/#:~:text=These%20models%20are%20capable%20of,largest%20language%20model%20ever%20trained>.

<https://www.assemblyai.com/blog/how-chatgpt-actually-works/> - marco ramponi –dec 23,2023

<https://zapier.com/blog/how-does-chatgpt-work/>

image of steps from :

<https://www.assemblyai.com/blog/how-chatgpt-actually-works/> - marco ramponi –dec 23,2023