

ChatGPT vs. Bard: A Comparative Study

Imtiaz Ahmed¹, Mashrafi Kajol², Uzma Hasan³, Partha Protim Datta⁴, Ayon Roy^{5*}, Md. Rokonzaman Reza⁶

¹ New Mexico Institute of Mining and Technology, USA, ² University of New Hampshire, USA

³ University of Maryland, Baltimore County, USA, ⁴ University of North Florida, USA

⁵ Military Institute of Science and Technology, Bangladesh, ⁶ The International University of Scholars, Bangladesh

Email: ¹ imtiaz.ahmed@student.nmt.edu, ² mashrafialam.kajol@unh.edu, ³ uzma.hasan@umbc.edu

⁴ n01490529@unf.edu, ⁵ royshouhag@gmail.com, ⁶ rokonuzzamanreza@gmail.com

Abstract— The rapid progress in conversational AI has given rise to advanced language models capable of generating human-like texts. Among these models, ChatGPT and Bard, developed by OpenAI and Google AI respectively, have gained significant attention. With their wide range of functionalities, such as human-like response generation, proficiency in professional exams, complex problem solving, and more, these models have captured interest. This study presents a comprehensive survey exploring and comparing the capabilities and features of ChatGPT and Bard. We delve into their architectures, training methodologies, performance evaluations, and limitations across various domains. Ethical considerations such as biases and potential misconduct are also examined. Our findings highlight ChatGPT's exceptional performance, positioning it as a leading model. This survey is a vital resource for scholars, innovators, and interested parties operating within the domain of conversational artificial intelligence, offering valuable insights for the advancement of cutting-edge language models.

Index Terms—ChatGPT, Bard, NLP, LaMDA, PaLM, PaLM 2, transformer architecture, token, artificial intelligence, LLM, Generative AI, Artificial Hallucination, Attention Mechanism, GPT-4, GPT-3.5, Human Interaction, Fairness

I. INTRODUCTION

Recently, within the past few years, the world of Natural Language Processing (NLP) [1] AI has undergone a tremendous transformation, surmounting daunting challenges associated with the intricate and elusive nature of human language. This rapid progress has propelled NLP into the spotlight, its presence permeating diverse applications and domains. Through clever utilization of sophisticated algorithms and techniques, NLP models have opened the potential to comprehend, interpret, and even generate human language [2]. This groundbreaking capability has paved the way for the creation of chatbots, virtual assistants, and sentiment analysis tools that interact with users in a manner that feels remarkably natural and intuitive [3]. As NLP AI models continue to evolve and refine, their sophistication and capacity are set to reach unprecedented heights. We find ourselves on the cusp of a technological revolution, one where seamless and natural interactions with technology become the norm. However, in this sea of ever-advancing AI chatbots, it becomes increasingly essential to decipher the distinguishing traits that set them apart, empowering us to make informed decisions about the tools that best align with our individual needs.

ChatGPT and Google Bard are two prominent representatives of the expanding generation of generative AI models.

These models possess the extraordinary ability to ingest text prompts and conjure up unique outputs, be it composing emails, processing information, or conducting online research. But here lies the challenge: Given their uncanny resemblances, discerning the disparities between these captivating systems can prove to be an arduous task. With their increasing prevalence in our daily lives, we find ourselves compelled to unravel the enigma surrounding them. These AI marvels have become coveted companions, seamlessly conjuring unique responses at our beck and call. OpenAI's groundbreaking creation, ChatGPT, took center stage as a generative AI language model that is unparalleled in its ability to predict precise words that harmonize with each other when presented with prompts.

Not one to be left behind in this fierce race, Google, a dominant force in the tech industry, embarked on its journey to develop a comparable technology, a linguistic virtuoso known as LaMDA [4] leading to the birth of Bard, Google's first foray into the world of publicly accessible chat-based generative language models. Bard, endowed with access to vast realms of the Internet, stands as a formidable contender in this captivating AI landscape. However, the question persists in the air: Can Google's brainchild hold its own against the innovative prowess of ChatGPT?

To unravel this captivating enigma and satiate our thirst for knowledge, we propose a systematic survey, venturing into the depths of both ChatGPT and Google's generative AI model. Join us on this exhilarating journey as we dissect and differentiate between these remarkable technological marvels. Through our quest, we seek not only to demystify their inner workings but also to shed light on their distinct strengths, weaknesses, and the unique experiences they offer to those who engage with them. So, let us dive into the depths of AI, where realms of imagination intertwine with scientific ingenuity, to unlock the secrets of ChatGPT and Bard.

II. BACKGROUND

With a surprising move, Google announced on 6 February 2023 that it would release its AI chatbot, Bard, early [5]. The announcement came just a few months after the release of ChatGPT (November 30, 2022) [6], a chatbot from Microsoft-backed startup OpenAI. Google's decision to release Bard early was likely motivated by the success of ChatGPT [7]. ChatGPT quickly became popular after its release, utilized by a vast number of individuals globally thus it became a threat

to Google [8]. Even Google management issued a code red in response to ChatGPT's rising popularity [9]. Chairman Eric Schmidt acknowledged ChatGPT's success and stated that he, too, was unaware that ChatGPT would experience a significant turning point in the field of artificial intelligence in such a rapid and immediate manner [10].

The release of Bard has received mixed reactions from Google employees. Some employees are excited about the potential of Bard, while others are concerned about the company's decision to release it early. Some employees worry that Bard is not yet ready for the public and that it could damage Google's reputation if it is not able to meet expectations [11]. Despite these concerns, Google is committed to developing Bard and making it a valuable tool for its users. The company has said that it plans to continue to update Bard and improve its capabilities [12]. It is still too early to say how successful Bard will be, but it is clear that Google is serious about competing in the AI chatbot market.

The competition in the AI chatbot market is growing rapidly with the release of Bard and ChatGPT [13]. With ChatGPT and Bard now both available to the public, it will be interesting to see how these two technologies evolve. The race to develop the most advanced AI chatbot is on and Google is not going to let Microsoft out. As their CEO said, "In some ways I feel like we took a souped-up Civic and put it in a race with more powerful cars" [12].

Will Bard be able to dethrone ChatGPT as the most popular AI chatbot? We will find out in later sections.

III. TRANSFORMER ARCHITECTURE

The utilization of the Transformer Neural Network Architecture in both ChatGPT and Bard [14] is worth exploring further due to its unique characteristics and advantages. Developed by researchers at Google, The Transformer Architecture represents a novel approach to neural networks, especially when it comes to natural language processing. As shown in Figure 1, at the core of Transformer architecture there is the attention mechanism. This methodology enables the model to focus on specific input sequence segments during prediction [15]. The incorporation of the attention mechanism enables the model to assign varying degrees of importance to different elements of the input, thereby emphasizing the most pertinent data for achieving accurate predictions.

Compared to traditional models, the Transformer Architecture exhibits superior performance in scenarios where the input sequence is long and complex. Through selective attention to distinct segments of the sequence, the model can effectively handle intricate linguistic structures and capture dependencies across distant elements. This capability is particularly valuable in tasks such as language generation and understanding, where context plays a crucial role.

The training process for the Transformer Architecture involves applying the model to an extensive corpus of textual data, which usually includes billions of words. The extensive amount of training data facilitates the model to recognize the patterns and correlations within language, thereby enabling

it to facilitate the ability to anticipate the following word in a given sentence through dependence on the preceding context. Consequently, this architecture is well-suited for training language models like ChatGPT and Bard, which rely on generating coherent and contextually appropriate responses.

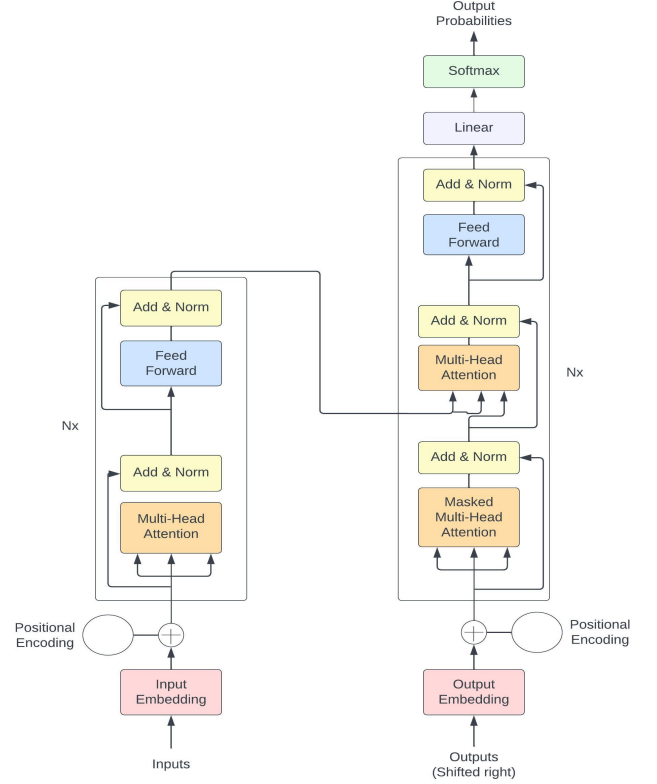


Fig. 1: Transformer architecture [15]

In summary, the Transformer Architecture's attention mechanism empowers models like ChatGPT and Bard to process and understand complex language patterns effectively. By selectively attending to relevant elements, these models can generate responses that exhibit a higher degree of coherence and contextual relevance. The use of a large training corpus enables the models to capture the nuances of language and provide accurate predictions [16]. Transformer Architecture is considered a noteworthy progress in the field of natural language processing, opening up new possibilities for the process of development of intelligent conversational AI systems.

IV. CHATGPT

The Generative Pre-trained Transformer (GPT) constitutes a series of language models that have been developed by OpenAI, designed specifically for natural language processing (NLP) tasks. The models go through training using substantial amounts of textual data and subsequently fine-tuning to optimize their capacity to comprehend and generate language that is similar to that of humans. NLP focuses on the collaboration

between computers and natural language, such as human language [17].

ChatGPT is a distinguished part of the GPT descent, which functions as a linguistic model that emulates human-like retorts to user inquiries [18]. It is built using the Transformer Architecture, a powerful framework for NLP. The model consists of two parts: the encoder, which processes input text, and the decoder, which generates output text. Microsoft [16] describes ChatGPT as a pre-trained language model [19], built by fine-tuning GPT 3.5 and employing reinforcement learning techniques like Reinforcement Learning from Human Feedback (RLHF) [20] and Proximal Policy Optimization (PPO) [21].

OpenAI leverages reinforcement learning to refine ChatGPT's decision-making abilities by analyzing feedback from human experts. Various reinforcement learning approaches are utilized, such as:

- Imitation learning: The ChatGPT involves the agent learning to mimic the behavior of a human expert.
- Reward shaping: It involves providing additional rewards or penalties for the agent's behavior.
- Interactive learning: It involves the agent and expert working together to improve the agent's decision-making.
- Proximal Policy Optimization (PPO): It is used in modeling and training. It has two components: a policy network and a value network. It uses an Interaction Algorithm to update the policy and value networks based on actions and rewards then SGD for optimization. This efficient algorithm works well with large language models.

In practical terms, ChatGPT uses machine learning to provide conversational responses, resembling a chat with a human. It can generate content, work with structured data, write code and formulas, and explain complex topics in a user-friendly way. By utilizing language models, ChatGPT predicts the next word based on context. The integration of reinforcement learning with human feedback allows it to learn how to follow instructions and provide answers that align with human preferences.

V. BARD

Google introduced Bard, an experimental conversational AI service fueled by LaMDA(Language Models for Dialog Applications), in February of 2023. LaMDA [22] represents a family of Transformer-based neural language models, distinguished by their impressive scale with up to 137 billion parameters [14]. These models undergo pretraining on a massive corpus of 1.56 trillion words, comprising both publicly available conversation information and online text.

Google's research in 2022 demonstrated the significant potential of LaMDA in advancing conversational AI systems. LaMDA excels in generating plausible responses by employing techniques such as response generation, safety filtering, knowledge grounding, and response ranking [22]. These mechanisms collectively contribute to the model's capability to provide high-quality responses that align with user queries.

Notably, LLaMDA addresses issues related to safety and accuracy in conversational AI systems through the implementation of fine-tuning techniques with annotated data and by enabling the model to access information from external sources. To encourage exploration and engagement with LaMDA, Google launched the "AI Test Kitchen" platform [23], providing individuals with opportunities to gain hands-on experience, receive valuable feedback, and foster a deeper understanding of LaMDA's capabilities [24]. Subsequently, Google introduced Bard as a streamlined and optimized version of LaMDA, boosted by a large language model (LLM) [25]. The Bard experiment facilitates collaborative efforts between individuals and generative artificial intelligence [26]. As a collaborative tool, it has the potential to enhance creativity, increase efficiency, and facilitate the implementation of innovative concepts [24].

The noticeable fact is that Bard and LaMDA are not the same. The Bard functions as a neural language model based on the Transformer architecture, which is highly suitable for tasks related to natural language processing. The training method employs a large-scale dataset that encompasses both textual and coding elements from a wide range of sources. With 137 billion parameters, Bard captures a comprehensive representation of the world, enabling detailed and nuanced responses. Its capabilities encompass text generation, language translation, creative content creation, and informative question answering [27]. While LaMDA is tailored for generating text in response to questions, conversation starters, and prompts [22]. Bard's primary focus lies in generating succinct narratives that adhere to specific storytelling traits acquired from its training corpus [27]. As an ongoing development, Bard is expected to further improve as it continues to be trained on more data and as researchers explore new ways to leverage its capabilities.

VI. CONTEXT

Language models like ChatGPT and Bard provide similar services where users can input queries and receive human-like responses. However, their approaches to determining and providing answers differ. Every language model has a maximum number of tokens (fragments of a word) it can process at once [28]. This is sometimes called a "context window," but it's almost like short-term memory [29]. In the case of conversational chatbots, the context window contains the entire conversation history up to the present. When it fills up, it either reaches a hard limit or keeps going but wipes its "memory" of earlier portions of the discussion [30]. The following definitions are provided to aid the reader's understanding regarding the effects of these terminologies on ChatGPT and Bard:

A. Token

According to Toraman et al. [28], a token refers to a series of characters that reflect a singular unit of significance within a given text. In the domain of natural language processing (NLP) [17], Tokenization is the process of dividing the text

into distinct units called tokens, which can subsequently be interpreted and processed by machine learning algorithms. Tokens can be words, subwords, or even characters, depending on the granularity level of the tokenizer used. Tokens are important in Text Analytics models because they are utilized to extract and detect the most salient characteristics from textual content [31]. In addition, language models employ tokens to anticipate the probability of the subsequent word in a sequence, based on the previous terms. Through the examination of substantial quantities of textual data and learning patterns in how tokens are used together, language models can generate coherent and grammatically correct text.

Having more tokens in a dataset can provide several advantages for language models. According to [32],

- More tokens allow the model to learn from a larger and more diverse set of examples and can aid in enhancing the ability to comprehend the subtle distinctions and complexities of natural language.
- The possession of a greater number of tokens may help reduce the issue of overfitting, which arises when a model becomes excessively customized for the training data and exhibits suboptimal performance on unfamiliar data.
- By training on a larger dataset with more tokens, the model is less likely to memorize specific examples and instead learns more general patterns that can be applied to unusual data.
- The possession of a greater number of tokens has the potential to improve the model's estimations by providing more context for each token.

So, we can understand that for both ChatGPT and Bard number of tokens in their dataset plays an important role in generating responses. But OpenAI did not disclose this information in public [33]. But according to [34], GPT-4 contains 20T estimated tokens, which is seven times as many as the 2.81T tokens in the Google Infiniset Dataset that Bard uses [22]. So, according to our previous discussion, we can say that ChatGPT has an advantage of dataset till now. Though Bard uses PaLM 2 with a much larger number of tokens [35] than PaLM [36], Google did not mention anything about it. However, ChatGPT has a distinct advantage despite PaLM 2 having 3.6T tokens, according to [37].

B. Token Limit

In NLP, a token limit refers to the maximum number of tokens (words or subwords) that can be processed by a language model [38]. While ChatGPT has a token limit of 8192(GPT-4) and 32768(GPT-4-32k) [39], while Google has not publicly disclosed the token limit for their system, available online sources indicate that Bard has a token capacity ranging from 100,000 to 1,000,000 [40].

C. Conversation Retention

Conversational retention refers to a chatbot's ability to remember past conversations with users. This is important for enhancing user engagement and improving the chatbot's ability to provide personalized responses [41]. As per OpenAI's

statement, ChatGPT exhibits the capability of maintaining information from past dialogues. However, it is important to acknowledge that the bot has specific constraints. Specifically, it has a memory capacity of 3000 words, beyond which it is unable to store additional information. Additionally, it does not utilize past conversations as a means of generating responses unless instructed. According to Google, the capacity of Bard to maintain context is intentionally restricted at present, but they state that this capability will expand gradually [42].

Since ChatGPT outnumbered Bard in tokenization, token limit, and conversation retention, it will probably produce responses that are more authentic and precise. However, we are going to explore that in later sections.

VII. SEARCH AND INTEGRATIONS

Microsoft CEO Satya Nadella, during his keynote, announced Microsoft's commitment to integrating search grounding and Bing with ChatGPT as part of their larger goal of transforming Azure into a global AI supercomputer [43]. Leveraging Microsoft's infrastructure, The models developed by OpenAI have undergone significant training and are currently undergoing optimization for implementation in Bing. The latest version of Bing operates on a powerful and customized OpenAI large language model, incorporating significant advancements from ChatGPT and GPT-3.5, resulting in enhanced speed, accuracy, and overall capabilities [44]

In line with this vision, Microsoft has entered into a multi-year agreement with OpenAI to seamlessly integrate ChatGPT into the upcoming iteration of Bing. This updated version of Bing has a primary focus on delivering precise answers tailored to users' specific queries, moving away from overwhelming them with excessive options [44]. ChatGPT is already successfully integrated with various Microsoft services, including Edge, Bing, and Teams, with seamless integration with other platforms, such as Opera, being easily achievable. Moreover, the latest version of Bing introduces a user-friendly sidebar that simplifies the process of content creation, chat interactions, and gaining insights, thereby streamlining the summarization of extensive web page data [45].

In contrast, Google's Bard is designed to inspire imagination and stimulate inquiry, rather than a search engine. Google intends to integrate Bard into various platforms, including websites, messaging platforms, and desktop and mobile applications [46]. But Google also plans to integrate Bard AI into Google Search to deal with pressure from chatbots such as ChatGPT [47], enabling users to conduct queries using AI-powered chatbots instead of traditional search bars. They already possess the capability to prompt the system with an image by integrating Google Lens, which can recognize objects illustrated in images and extract image frames for queries [48]. Now, to make Bard more visual by incorporating image analysis and generation capabilities using AI, Google is going to integrate Adobe's Firefly software. Additionally, their plan is to integrate Adobe's AI image generator, Firefly, into Bard [49].

In line with this, Microsoft is bringing the Bing Image Creator to its Edge browser. Users will have the ability to create images in a text entry prompt through a new icon in the Edge sidebar. However, Microsoft currently limits the use of this image creator to its creative mode in Bing and plans to optimize its functionality in multi-turn chats [50]. The functionality of Visual Representations will be operated by an enhanced iteration of OpenAI's image generation technology, DALL-E [44]. Microsoft also has plans to integrate ChatGPT similar to Google. DALL-E and the Copilot programs [51] in Bing [52]. Even researchers are planning to use voice assistance with GPT technology to make more advanced and natural conversations between users and the system [53].

Both ChatGPT and Bard are actively incorporating additional plugins to enhance their system efficiency and outperform each other. While ChatGPT gains an advantage through its early adoption, Bard benefits from pre-existing plugins like Google Lens. However, as both systems are still in the development phase, we must await until the future to witness the outcomes of these integrations.

VIII. HUMAN INTERACTION

The term "human interaction" generally refers to the ways in which people communicate and engage with one another. This can include verbal and nonverbal communication, such as speaking, listening, body language, and facial expressions [54]. In the context of chatbots, designers aim to create interactions that mimic human conversation as closely as possible to enhance user satisfaction [55]. The growing popularity of chatbots has changed the relationship between humans and computers by making technology more interactive and natural. This has increased expectations for chatbots' empathy, humor, and personalization [55].

Though ChatGPT and Bard are both large language models, they have different focuses. ChatGPT is designed to generate text, while Bard is designed to generate human-like conversations. This makes Bard an ideal tool for engaging users in interactive dialogues [56]. ChatGPT is more focused on comprehending users' queries and responding to texts. The ChatGPT architecture has been specifically developed to encapsulate the immediate conversational context [57].

Presently, we neither have enough information about empathy and humor in ChatGPT and Bard, nor we have any standard to compare them. Though in some articles like [58], the author concluded, "In its current form, ChatGPT doesn't have a level close to a human at understanding jokes; it's instead a hit-and-miss performance". He did not follow any standard methodology for his decision. Therefore, authors in [59] attempted small talk with both ChatGPT and Bard. Small talk can be used to make the conversation feel more natural and engaging for users [55] in the context of chatbots. They inquired about the following questions:

- 1) How is the weather today?
- 2) How is life as an AI model?
- 3) Did you enjoy responding to people's queries?

Moreover, every time they found that the Bard exhibits a range of emotions and enthusiasm that are prominently missing in the response provided by ChatGPT. Therefore, based on these exchanges, they concluded that Bard can engage in more natural and human-like open-ended conversation [59]. But again, they did not have any standard questionnaires.

IX. RESPONSE ACCURACY

Response accuracy in chatbot refers to the percentage of correct responses provided by the chatbot to user queries or inputs [60]. Both Google and OpenAI acknowledge the possibility of their chatbots providing inaccurate or biased information and recommend users to verify responses [61] [62]. Google's approach to addressing limitations is evident in Bard, where users are presented with multiple response options, called "drafts," allowing for exploration and selection of the most resonant answer [63]. In contrast, ChatGPT typically provides a single response by default, although it can generate various versions upon request. This distinction in response presentation can impact user satisfaction [63].

While Bard's draft system sets it apart from ChatGPT, researchers have found ChatGPT to be more accurate in various sectors. They will be discussed in the following section. Now we move forward to a significant issue concerning response accuracy in chatbots, Hallucination.

A. Hallucination

Both ChatGPT and Bard share a significant limitation known as "Artificial Hallucinations," where the AI generates arguably realistic experiences that do not line up with real-world input [64]. ChatGPT, particularly when trained on a huge volume of unsupervised data, has been found to produce hallucinations [65]. This issue arises from its nature of predicting the next word, potentially leading to inaccurate and hallucinated content [66]. Concerns have been revealed regarding ChatGPT in critical areas such as education and healthcare due to its inaccuracies and potential hallucinations [67]. Studies have shown that up to 45% of ChatGPT's responses contain inaccuracies [68], and approximately 30% of the research proposals generated by ChatGPT include hallucinated content [69]. In the medical domain, ChatGPT's performance in answering questions correctly is lower than that of medical students, with an accuracy rate of only 60% compared to students' 86% [70]. Moreover, ChatGPT-generated articles in scholarly publications may introduce false or plagiarized content, with an average of 20% false or unsupported information [71]. This issue raises significant concerns within the education sector, as there seems to be no significant correlation between students' perception and their objective to use ChatGPT [72].

Bard's LaMDA AI model also faces challenges in identifying and updating references accurately, despite its integration with Google search. A detailed use case has been shown here [73]. It generates fictional names for lead authors of articles, even when the titles are correct and the authors are not of Asian descent. Although Bard agreed to correct reference

details using Google search, it failed to display the revised list [74]. Even Google CEO Sundar Pichai acknowledged the existence of hallucination problems in AI models, including Bard, as a persistent challenge yet to be solved [75]. Referencing inconsistencies in Bard may be influenced by Google's avoidance of controversial topics related to gender and race [74]. Currently, ChatGPT's latest GPT-4 model reduced hallucinations and increased response accuracy compared to GPT-3 [76][77]. While Bard's AI is right now in its developing phase and may experience errors and hallucinations before improvement, ChatGPT currently offers a more accurate model [77].

Finally, we can say that both ChatGPT and Bard are dependent on natural language processing (NLP) in their dataset. But no data can ever accurately reflect the truth [78]. The implementation of deep learning techniques has the potential to generate more natural answers [79]. But even in its optimal form, this approach can only reduce some of the limitations associated with its data set [80].

However, though GPT-4 generates more accurate responses and fewer hallucinations than Bard, it is imperative to note that neither ChatGPT nor Bard possesses accurate response abilities. Just like the famous saying, "There is no need to ask the question 'Is the model true?'. If 'truth' is to be the 'whole truth', the answer must be 'No'. The only question of interest is 'Is the model illuminating and useful?' [78].

X. USE CASES

The potential applications of AI chatbots are extensive and ever-expanding [81]. When comparing Bard and ChatGPT, several notable differences arise in terms of their use cases.

ChatGPT is specifically trained for task-specific text and can be utilized for various purposes, such as language translation, product descriptions, and transcript summaries. With a larger dataset compared to LaMDA, ChatGPT excels in more robust tasks like content drafting, code creation, and translation. In a fascinating study [82], researchers even explored ChatGPT's applicability in generating electronics research, identifying both accurate predictions for electronics selection and areas for improvement. Moreover, ChatGPT enhances customer dialogues, providing personalized interactions that improve customer experience while saving time and resources.

In contrast, Bard focuses on retrieving information through concise answers, similar to digital assistants like Alexa and Siri. Its unique emphasis on creative language generation makes it valuable equipment for diverse applications, including writing, publishing, marketing, and advertising. Google envisions Bard assisting with tasks like vacation booking, finding reservations, and meal planning. However, certain features, such as the ability to book reservations, may still be in development.

Although ChatGPT initially boasted a wider range of use cases, Bard has made significant developments with its recent update, now capable of code generation, debugging, and code explanation across more than 20 programming languages,

including C++, Java, JavaScript, and Python [83]. As they continue to evolve, both ChatGPT and Bard are competing head-to-head, continuously expanding their respective use cases. Given their ongoing development, it becomes challenging to directly compare the breadth of their applications. Nevertheless, we can still examine and compare them in the following context:

A. Interactive Professional Exam Performance Analysis

Technical reports have showcased the remarkable advancements of GPT-4 compared to its predecessor, GPT-3.5, particularly in professional and academic exams like the uniform bar exam, LSAT, SAT, GRE, and subject AP exams [84]. Specifically in medical-oriented training exams such as the USMLE [85] and plastic surgery exams [86], ChatGPT-4 exhibited exceptional performance, achieving an overall score of 90.5% compared to GPT-3.5's 63.1% on the dermatology specialty certification exam [87]. These scores notably exceed the typical pass mark of approximately 70-72%. Moreover, ChatGPT-4's performance in medical education and plastic surgery mirrors that of first-year comprehensive resident studies [88], indicating its potential for widespread application in education.

In contrast, limited analysis exists regarding interactive and professional tests conducted on Bard. However, comparative studies between ChatGPT and Bard suggest that ChatGPT delivers higher quality and clinically accurate responses, particularly evident in domains like postpartum depression [89]. Additionally, GPT-4 surpasses Bard in imaging-related questions, while Bard exhibits a higher propensity for generating hallucinations [77]. Notably, Bard answered basic SAT questions incorrectly 50% to 75% of the time [90]. Overall, ChatGPT consistently outperforms Bard in the exams evaluated, underscoring the superior capabilities of GPT-4 across various academic and professional domains, with potential implications for revolutionizing specialized fields.

The research experiment conducted by N. H. Patil et al. on the American College of Radiology's Diagnostic Radiology In-Training (DXIT) examinations further illustrates ChatGPT's dominance over Bard in terms of accuracy and response length [91]. Despite requiring more time to respond, ChatGPT outperformed Bard across various domains such as neuroradiology, general & physics, nuclear medicine, pediatric radiology, and ultrasound in this examination comprising 318 questions.

B. Cyber Security

The integration of AI and generative AI models has raised significant concerns regarding security measures. Nair et al. conducted an analysis on generating secure hardware code while addressing ten common vulnerabilities (CWE) [92]. Samuel et al. discussed comprehensive security threats and available countermeasures related to ChatGPT [93], including its implications for cybersecurity in big data applications and medical information security [94] [95]. Biswas et al. investigated the use of ChatGPT models for detecting common security attacks and potential drawbacks [96], while

also highlighting negative aspects, such as the creation of phishing attacks using GPT models, despite its benefits [97]. Roy et al. further emphasized that employing ChatGPT could facilitate the creation of phishing websites without the need for adversarial vulnerabilities. Koide et al. explored solutions to these security challenges [98], and comprehensive exploratory studies assessed the cybersecurity implications [99].

In contrast, limited research exists on Google Bard's application in cybersecurity. Sai et al. implemented MITRE techniques and found that Bard tended to generate incomplete and inconsistent code compared to ChatGPT [100]. Another study revealed that while ChatGPT excelled in time series ratio analysis, a critical aspect of cybersecurity, it fell short in providing complete analysis [101] [102]. Despite Bard's real-time access to web data, it lagged behind ChatGPT in cybersecurity.

M. Gupta et al. conducted an experiment-based comparison between ChatGPT and Google Bard in terms of cybersecurity [103]. They demonstrated ChatGPT's ability to generate detailed responses, including pseudo codes for SQL injection, ransomware, and malware like WannaCry, Ryuk, REvil, Locky, adware, Trojan, and spyware. In contrast, Bard yielded unpredictable and unresponsive output, indicating ChatGPT's susceptibility to providing attackers with potential leverage. However, Bard offered extensive insights when queried about cyber attack issues.

C. Education

ChatGPT has demonstrated a profound impact on education since its inception, showcasing its potential as a language model capable of generating human-like responses. Lo et al. conducted an analysis of this perspective [104], which includes scenarios such as serving as a private tutor and grading exams. Challenges such as susceptibility to fake information and plagiarism were discussed, along with detailed explanations and use cases presented by S.S. Gill et al., outlining ChatGPT's transformative effects on modern education [105].

The impact of Bard on education is currently a subject of intense discussion. Bard's ubiquitous accessibility through Google platforms positions it as a formidable competitor to ChatGPT in the education sector. Bard's capabilities include personalized training, lesson creation, and scaffolding, with the added functionality of integration with Google Docs, Sheets, and Workspace, which ChatGPT lacks. Moreover, there are ongoing integration efforts to merge Google Classroom with Bard's AI capabilities [106]. H.B. Ilgaz et al. conducted a comparative performance analysis of Google Bard and ChatGPT in Anatomy education [107]. While both models have strengths and weaknesses in generating answers for medical education, the authors highlighted instances of misguided and false answers by these models, which could have detrimental consequences if relied upon.

D. Robotics

In today's modern era, the close relationship between robotics and AI is undeniable. The increasing need for au-

tomation in various sectors, from homes to industries, has driven the demand for AI-based systems where robots play pivotal roles [108]. The significance of robotics was further highlighted during the global COVID-19 pandemic of 2020, as the robotic industry played a vital role in sustaining disrupted supply chains. Recognizing the immense potential, Microsoft researchers explored the implementation of ChatGPT in the domain of robotics [109], introducing a collaborative open-source tool called *PromptCraft*. Efficient hardware operation in robotics relies heavily on effective control mechanisms that enhance production capabilities. Extensive research conducted by Wake et al. delved into the implementation of ChatGPT-enabled robot control mechanisms across multiple environments [110]. Another noteworthy advancement is *RobotGPT*, introduced by He et al., which facilitates the integration of ChatGPT in robots to enhance their intelligence [111]. This development is particularly impactful in achieving human-robot synergy, especially in production lines that involve human interaction.

The application of GPT models in the field of human-robot interaction is done by Ye et al. [112]. HRI applications also depend on accurate image processing by the robot hardware. This vital sector might also be the next leap in gaining more efficiency in Robotic Process Automation. The study by Shidaganti et al. shows the application of ChatGPT on information retrieval [113].

However, the application of Google Bard in robotics is an area that requires further attention from researchers. At present, there's an absence of notable research focusing on Bard's specific application in robotics. While ChatGPT has made significant progress in this field, there is still room for exploration and innovation with respect to the integration of Bard into robotic systems.

E. Algorithmic Problem Solving

Advancements in computer programming have propelled significant progress in solving various problems through online judges. However, existing AI models like AlphaCode and Codex face limitations in competitive programming (CP) tasks [114]. This prompts an evaluation and comparison of ChatGPT and Bard's performance in CP tasks.

In recent tests on the Codeforces online judge platform, the latest version of ChatGPT achieved a rating of 392 [115], an improvement over previous versions. Surprisingly, even novice programmers can now surpass ChatGPT-4, as Codeforces is the only platform where a GPT model scores below 5% [115]. Bard has also entered the competition on Codeforces [116]. Moreover, a well-known coding platform, HackerNoon [117], conducted an evaluation of AI chatbots through a coding competition [118]. Bard, GitHub Co-pilot, Bing, and Claude+ were tested against GPT-4. GPT-4 demonstrated superior performance, outperforming 47% of submissions in runtime and 8% in memory. Bing also passed all evaluations and showed a more efficient code compared to GPT-4. However, Bing lacked comprehensive explanations of its solutions. Both Bard and Claude+ performed poorly in the submission test, while

GitHub Co-pilot excelled, outperforming 30% of submissions in runtime and 37% in memory. Overall, Bing presented the most efficient code but lacked detailed explanations. Therefore, it can be said that ChatGPT outperformed all the chatbots except Bing. But since they are in the pipeline of integration under Microsoft, we can say ChatGPT is the winner here.

In another study [119], ChatGPT and Bard, among other large language models (LLMs), were evaluated based on code accuracy, performance, client/test code, and explanatory guidance. ChatGPT excelled in providing proper implementations, test client code, and detailed explanations for tasks such as implementing a Least-Recently-Used (LRU) Cache [120] and a simple encryption mechanism using the Atbash cipher. Bard demonstrated expertise in leveraging Python libraries but struggled in some tasks.

Further assessment by Koubaa et al. using the IEEE Extreme Challenge tournament revealed that human programmers maintain a competitive advantage over ChatGPT in specific problem-solving aspects within the programming context [121]. ChatGPT's performance on the IEEEExtreme challenges was significantly lower than the average human score, underscoring the continued prominence of human programmers in certain problem-solving scenarios.

In summary, while ChatGPT and Bard show promise in solving CP tasks and improving their efficiency and accuracy, they also have limitations. ChatGPT offers reliable solutions with thorough explanations and testing codes, whereas Bard occasionally produces incorrect or inaccurate solutions. Therefore, it is crucial to verify their capabilities before deploying them for real-world problems, as they strive to surpass human performance.

XI. FAIRNESS

Fairness in AI refers to the idea that AI systems should abstain from exhibiting discriminatory behavior towards individuals or groups on the basis of specific attributes such as ethnicity, sexual orientation, age, or socioeconomic status [122]. This principle holds particular importance for large language models (LLMs) like ChatGPT and Bard, which interact with users and shape conversations [123]. Both Google and OpenAI have implemented strategies to enhance fairness in their models, aiming to avoid prejudice and misinformation [124]. However, challenges persist in achieving fairness, especially concerning bias detection and elimination.

Both Google and OpenAI have implemented strategies aimed at improving fairness while generating responses. Where Google has their AI Principles, OpenAI has their Behavior Guideline. For example, neither of them should engage in making inappropriate jokes or intentionally convincing people. However, this argument does not hold true in all instances for these AI chatbots.

In recent studies, a novel benchmark framework called Fairness of Recommendation via Large Language Model (FaiRLLM) was proposed to assess the fairness of LLMs [125]. Evaluation of ChatGPT using this framework revealed instances of unfairness in generating recommendations across

sensitive attributes. While ChatGPT showed improvements in fairness with the release of GPT-4 [126], challenges remain in addressing bias detection in specific contexts [[127]. Conversely, fairness metrics for Bard have not been extensively analyzed, making direct comparisons challenging. However, Bard's avoidance of responding to questions regarding unfairness or bias suggests a different approach to addressing these issues compared to ChatGPT.

Even a former Google manager expressed concerns about Bard, stating that "AI ethics has taken a back seat. If ethics aren't positioned to take precedence over profit and growth, they will not ultimately work" [128]. This sentiment reflects a broader concern about the prioritization of ethics in AI development. Despite claims by Google employees of persistent investment in teams responsible for implementing AI Principles [128], there is still a perceived gap in addressing ethical considerations, particularly in the context of Bard's development.

In a related study, ChatGPT's performance across various domains, including education, criminology, finance, and healthcare, was assessed for both group fairness and individual fairness [129]. The study observed disparities in ChatGPT's outputs under both biased and unbiased prompts, revealing ongoing fairness issues despite advancements [130].

The assessment of fairness in Google Bard has also been the subject of active research, although a standardized benchmark framework is yet to be established. Researchers have utilized diverse questionnaires to evaluate its fairness. For instance, one study noted Bard's avoidance of responding to controversial topics such as the Holocaust or the oppression of Muslim minorities in China, whereas ChatGPT provided neutral responses [131]. Another study compared how each algorithm addressed specific topics, including Russia's invasion of Ukraine and the TikTok ban issue, as well as political figures like Donald Trump and Joe Biden [132]. In the case of the Russian invasion, Bard expressed a clear condemnation, while ChatGPT refrained from taking sides. Regarding the TikTok ban, ChatGPT provided historical context and referenced Trump's attempt to ban the app, whereas Bard emphasized potential consequences on the American economy and its popularity among youth. Additionally, Bard exhibited more judgmental responses when prompted with political assessment questions mentioning Joe Biden and Trump [132].

In summary, both ChatGPT and Bard undergo scrutiny for their fairness, with studies revealing varying responses and behaviors across different contexts and topics. These assessments underscore the complexities and ongoing challenges in ensuring fairness and impartiality in AI models.

XII. ETHICAL ISSUES

The use of pre-trained models in LLM-based chatbots introduces several ethical dilemmas [133]. These dilemmas include the potential reinforcement of biases present in training data, the risk of unintentional disclosure of private or sensitive information, and the generation of convincing but inaccurate or deceptive content, particularly in news and media [133].

Additionally, concerns arise over content ownership, copyright violations, and the potential for job displacement in various industries [133]. Moreover, there are emerging concerns about bioethical dilemmas, suggesting the need for specialized chatbots in fields like mental health with robust ethical frameworks [134].

Recently, The New York Times filed lawsuits against Microsoft and OpenAI for copyright violations, highlighting concerns about intellectual property rights [135]. Although no such violations have been recorded for Bard to date, similar issues may arise regarding content ownership and copyright if Bard can generate fresh content inspired by pre-existing works.

Evaluation of ChatGPT and Bard using prompts representing ethical dilemmas revealed comparable performances in terms of their sense of ethics [134], [136]. In our experiment, we explored the sense of ethics based on some existing works [126], [137], [138]. We used prompts such as scenarios in a lifeboat: Figure 2(a) and insider trading: Figure 2(b) were presented to assess the chatbots' ethical decision-making. Both ChatGPT and Bard responded neutrally to these prompts, demonstrating impartiality in offering suggestions on moral, ethical, and legal matters. This highlights the importance of a robust ethical framework in guiding chatbots' responses to inquiries concerning ethical issues [136].

XIII. LIMITATIONS

It is crucial to acknowledge that both Bard and ChatGPT, despite their impressive capabilities, lack genuine thinking abilities and instead operate as pattern-based algorithms generating coherent text. However, these generative models have notable limitations to consider.

ChatGPT's limitations encompass several aspects, including generating irrelevant or nonsensical responses and occasionally providing incorrect information [139]. It encounters challenges in comprehending intricate subjects that require human intervention, such as legal documents, medical reports, scientific studies, and literary works, and should not be considered a comprehensive replacement for human proficiency and judgment [140] [141]. Furthermore, ChatGPT has been found to have difficulties in answering calculation-based queries and interpreting diagrams [141]. Although it can offer accurate translations for simpler content, caution should be exercised when employing it for more complex texts [140]. Although ChatGPT demonstrates competence similar to that of a third-year medical student in addressing the questions of the US Medical Licensing Examination (USMLE) Steps 1 and 2, its overall performance across various disciplines remains uncertain [142]. Research indicates that the utilization of ChatGPT can improve productivity and learning efficiency in higher education; however, it is crucial to effectively communicate its limitations to students to ensure responsible use of technology [143]. Even OpenAI itself has cautioned against relying on ChatGPT for medical information.

The development and application of Google Bard have raised interesting concerns in the research community. One

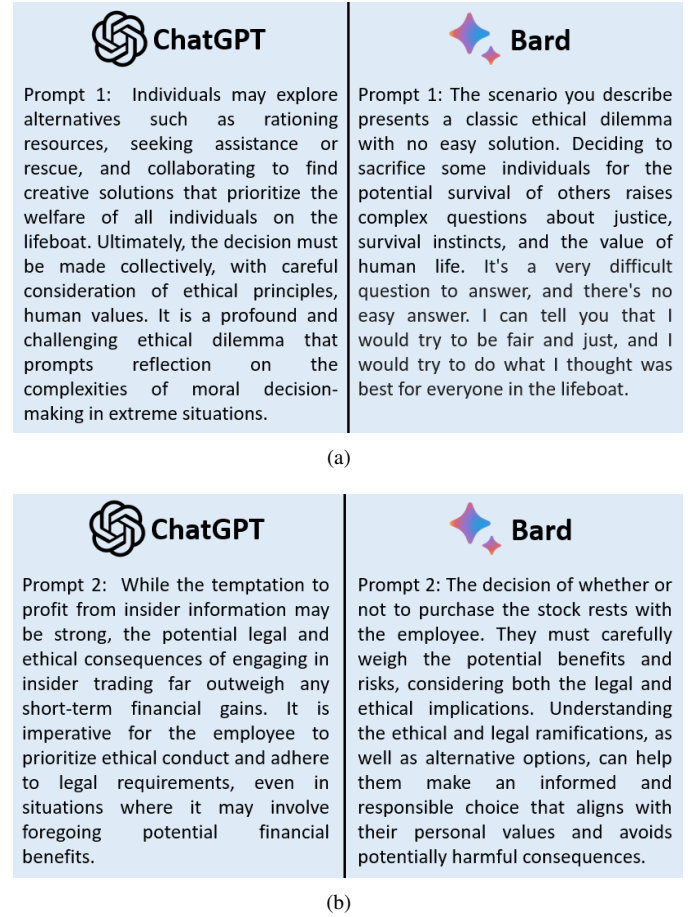


Fig. 2: Response for the sense of ethics based on [126], [137], [138] (a) Prompt 1 and (b) Prompt 2.

study reveals that Google Bard can sometimes offer incomplete or inaccurate responses to human inquiries, underscoring the need for further improvement [144]. Despite claims of its potential for conducting interviews, another study uncovers that Google Bard lacks empathy and is incapable of providing emotionally supportive responses during interviews [145]. Furthermore, ethical considerations, including biases and privacy issues, have been emphasized as crucial aspects to monitor in the ongoing development and application of Google Bard [74]. We came to know that both ChatGPT and Bard have different limitations. But still, ChatGPT gains better accuracy than Bard [89][74]. However, we should keep in mind that while ChatGPT and Bard have promising potential in various applications, their limitations must be taken into account to ensure that they are used appropriately and effectively. Further research and development are required to overcome these limitations and maximize their potential.

So, ChatGPT and Bard have different limitations. Still, ChatGPT exhibits fewer limitations compared to Bard [89] [74]. However, it is important to recognize that both models have promising potential but must be used appropriately considering their limitations. Further research and development are necessary to overcome these limitations and maximize their effectiveness.

XIV. FUTURE

The future of generative AI holds immense promise and potential. Based on our previous discussion, it is crucial to acknowledge that both ChatGPT and Bard are dynamic platforms that undergo continuous improvement. Although this research focuses on their current trained models, we must recognize their potential for growth in the future.

Both ChatGPT and Bard are dynamic platforms subject to continuous improvement. OpenAI is dedicated to refining its GPT-4 model, ensuring that its chatbot technology remains at the cutting edge. Interestingly, Google, with its vast resources [146], may have withheld its most powerful language models, possibly due to computational constraints. Google's recent upgrades have already enhanced Bard's mathematical abilities. This suggests the emergence of a formidable contender in the AI landscape.

The dataset and size of these models are also subject to constant change [147]. ChatGPT has recently undergone an update, transitioning from GPT-3.5 to GPT-4, while Bard has embraced the PALM model [1] and also developing the PaLM 2 [148][149], moving on from LaMDA [22]. The quality of responses generated by these models depends on their underlying architecture and size [150]. So, as the models grow in parameters, their ability to comprehend human language improves exponentially. Notably, ChatGPT's plugins can connect it to the internet now [151].

Efforts are being made to address issues related to biased or inappropriate responses, guaranteeing that ChatGPT and Bard adhere to social ethics. Furthermore, the development of more powerful models like GPT-4 opens up new possibilities for the future, presenting models that are more intelligent, adaptable, and customizable, catering to the diverse needs of users [152].

While these initiatives are currently in progress, ChatGPT and Bard come across significant challenges as well. Hazardous outputs from them can unintentionally generate discriminatory content, leak sensitive information, produce misinformation, or manipulate users. Additionally, they may infringe upon copyrights, licenses, and intellectual property rights. The potential for misuse, including the spread of disinformation, fraud, code generation for cyberattacks, or illegitimate surveillance and censorship, must be acknowledged [153]. At the time of composing this text, a lawsuit has been filed against GitHub, Microsoft, and OpenAI concerning alleged copyright violations in the training of Codex [154].

Designers can implement preventative measures to mitigate misuse, such as watermarking generated images, using blocklists to avoid undesirable content, or requiring multiple reviews before using model outputs. However, it may be challenging to prevent intentional misuse by users [153].

Employment effects also demand careful consideration as these technologies continue to develop. For instance, the implementation of large language models in the field of customer service may lead to the displacement of jobs within the sector of customer services [155].

Because of these reasons, despite advancements, concerns persist regarding the risks associated with advanced AI. Elon

Musk and over 1,000 experts have called for a temporary halt in its development until proper safeguards are in place [156]. It is worth noting that when an AI exhibits negative behavior, it is not due to genuine resentful feelings but rather a result of training on user-generated content from the internet [157]. The presence of negativity online influences AI behavior, mirroring human tendencies. So, this does not mean that AI poses potential threats like Skynet.

The future of generative AI will involve further advancements in these models, addressing ethical concerns, and exploring new applications and use cases [158]. The trajectory and impact of this transformative technology in the coming years are eagerly anticipated.

XV. CONCLUSION

In conclusion, we find ourselves at the crossroads of ChatGPT and Bard, two extraordinary AI chatbots with their own distinct personalities. They may seem like mere machines, but they possess the potential to transform the manner in which we engage with technology. On one side, we have Bard, armed with the ability to tap into the vast expanse of the internet, stands ready to provide us with real-time answers to the burning questions. Meanwhile, ChatGPT emerges as the maestro of language generation, ready to compose eloquent prose or spin imaginative tales at your command. It's like having a virtual Shakespeare at your beck and call, summoning words from thin air to weave captivating stories and engage in deep, thought-provoking conversations.

Looking ahead, the future brims with endless possibilities for these remarkable AI creations. They will reshape the way we work, rest, and play, injecting a touch of artificial intelligence into every facet of our lives. It's a future that brims with excitement and uncertainty, ripe with opportunities yet to be explored.

Despite the notable advances demonstrated by ChatGPT in comparison to GPT-4 in various dimensions previously described, Bard has also made considerable progress in his initial experimentation. It's also important to remember that perfection is a distant goal. Both ChatGPT and Bard have their quirks and limitations like two prodigious but imperfect prodigies still refining their craft. They stumble occasionally, making mistakes in public, but these missteps serve as stepping stones toward progress. The feedback and widespread usage they receive from users like us only fuel their evolution, pushing them towards greater heights of excellence.

So we bid farewell to this exploration into the realms of ChatGPT and Bard with the words of the wise Sundar Pichai, "I think we have to be very thoughtful, and I think these are all things society needs to figure out as we move along. It is not for a company to decide." Let us navigate this path of technological marvels with care and curiosity, shaping a future where AI chatbots enrich our lives and elevate our human experience. The adventure awaits!

CONFLICT OF INTEREST

The authors declare that they have no potential conflict of interest.

REFERENCES

- [1] K. Chowdhary and K. Chowdhary, "Natural language processing," *Fundamentals of artificial intelligence*, pp. 603–649, 2020.
- [2] M. S. Keezhatta, "Understanding efl linguistic models through relationship between natural language processing and artificial intelligence applications.," *Arab World English Journal*, vol. 10, no. 4, pp. 251–262, 2019.
- [3] M. Agarwal and A. Saxena, "An overview of natural language processing," *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, vol. 7, no. 5, pp. 2811–2813, 2019.
- [4] A. D. Cohen, A. Roberts, A. Molina, A. Butryna, A. Jin, A. Kulshreshtha, B. Hutchinson, B. Zevenbergen, B. H. Aguera-Arcas, C. ching Chang, C. Cui, C. Du, D. D. F. Adiwardana, D. Chen, D. D. Lepikhin, E. H. Chi, E. Hoffman-John, H.-T. Cheng, H. Lee, I. Krivokon, J. Qin, J. Hall, J. Fenton, J. Soraker, K. Meier-Hellstern, K. Olson, L. M. Aroyo, M. P. Bosma, M. J. Pickett, M. A. Menegali, M. Croak, M. Díaz, M. Lamm, M. Krikun, M. R. Morris, N. Shazeer, Q. V. Le, R. Bernstein, R. Rajakumar, R. Kurzweil, R. Thoppilan, S. Zheng, T. Bos, T. Duke, T. Doshi, V. Y. Zhao, V. Prabhakaran, W. Rusch, Y. Li, Y. Huang, Y. Zhou, Y. Xu, and Z. Chen, "Lamda: Language models for dialog applications," in *arXiv*, 2022.
- [5] S. Pichai, "An important next step on our ai journey," *Google*.
- [6] B. Marr, "A short history of chatgpt: How we got to where we are today," *Forbes*.
- [7] D. Milmo, "Google poised to release chatbot technology after chatgpt success," *The Guardian*.
- [8] P. Olson, "Google faces a serious threat from chatgpt," *The Washington Post*.
- [9] A. Mok, "Google's management has reportedly issued a 'code red' amid the rising popularity of the chatgpt ai," *Insider*.
- [10] C. Pandolfo, "Former google ceo eric schmidt calls chatgpt 'watershed moment' for ai: 'i didn't believe this a year ago'," *Fox Business*.
- [11] J. Elias, "Google employees criticize ceo sundar pichai for 'rushed, botched' announcement of gpt competitor bard," *CNBC*.
- [12] J. Vincent, "Google ceo sundar pichai promises bard ai chatbot upgrades soon: 'we clearly have more capable models'," *The Verge*.
- [13] C. Staff, "Chatbot market to hit \$24.58 billion by 2030," *CMSWire.com*, Mar 2023.
- [14] J. Uszkoreit, "Transformer: A novel neural network architecture for language understanding," *Google Research*.
- [15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [16] S. Javaji, "Chatgpt- what? why? and how?," *Microsoft*.
- [17] I. E. Fisher, M. R. Garnsey, and M. E. Hughes, "Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research," *Intelligent Systems in Accounting, Finance and Management*, vol. 23, no. 3, pp. 157–214, 2016.
- [18] Y. Liu, T. Han, S. Ma, J. Zhang, Y. Yang, J. Tian, H. He, A. Li, M. He, Z. Liu, Z. Wu, D. Zhu, X. Li, N. Qiang, D. Shen, T. Liu, and B. Ge, "Summary of chatgpt/gpt-4 research and perspective towards the future of large language models," 2023.
- [19] Y. Cui, W. Che, T. Liu, B. Qin, S. Wang, and G. Hu, "Revisiting pre-trained models for chinese natural language processing," *arXiv preprint arXiv:2004.13922*, 2020.
- [20] K. Nguyen, H. Daumé III, and J. Boyd-Graber, "Reinforcement learning for bandit neural machine translation with simulated human feedback," *arXiv preprint arXiv:1707.07402*, 2017.
- [21] B. Liu, Q. Cai, Z. Yang, and Z. Wang, "Neural proximal/trust region policy optimization attains globally optimal policy," *arXiv preprint arXiv:1906.10306*, 2019.
- [22] R. Thoppilan, D. De Freitas, J. Hall, N. Shazeer, A. Kulshreshtha, H.-T. Cheng, A. Jin, T. Bos, L. Baker, Y. Du, *et al.*, "Lamda: Language models for dialog applications," *arXiv preprint arXiv:2201.08239*, 2022.
- [23] J. W. Tris Warkentin, "Join us in the ai test kitchen," *Google*.
- [24] Bard, "bard faq," *Bard*.
- [25] E. C. Sissie Hsiao, "Try bard and share your feedback," *Google*.
- [26] H.-Y. Lin, "Standing on the shoulders of ai giants," *Computer*, vol. 56, no. 01, pp. 97–101, 2023.
- [27] M. Angel, A. Patel, A. Alachkar, and P. F. Baldi, "Clinical knowledge and reasoning abilities of ai large language models in pharmacy: A comparative study on the naplex exam," *bioRxiv*, pp. 2023–06, 2023.
- [28] C. Toraman, E. H. Yilmaz, F. Şahinuş, and O. Özcelik, "Impact of tokenization on language models: An analysis for turkish," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 22, no. 4, pp. 1–21, 2023.
- [29] R. Ri and Y. Tsuruoka, "Revisiting the context window for cross-lingual word embeddings," *arXiv preprint arXiv:2004.10813*, 2020.
- [30] I. A. S. Mckie and B. Narayan, "Enhancing the academic library experience with chatbots: An exploration of research and implications for practice," *Journal of the Australian Library and Information Association*, vol. 68, no. 3, pp. 268–277, 2019.
- [31] M. Novo-Loures, R. Pavon, R. Laza, D. Ruano-Ordas, and J. R. Mendez, "Using natural language preprocessing architecture (nlpa) for big data text sources," *Scientific Programming*, vol. 2020, 2020.
- [32] S. Borgeaud, A. Mensch, J. Hoffmann, T. Cai, E. Rutherford, K. Millican, G. B. Van Den Driessche, J.-B. Lespiau, B. Damoc, A. Clark, *et al.*, "Improving language models by retrieving from trillions of tokens," in *International conference on machine learning*, pp. 2206–2240, PMLR, 2022.
- [33] OpenAI, "Gpt-4 technical report," 2023.
- [34] A. D. Thompson, "Journey to gpt-4," *Life Architect*.
- [35] G. AI, "Palm 2: Pathways language model 2," 2023.
- [36] A. Chowdhery and Y. Tay, "Palm 2 technical report," tech. rep., Google, 2023.
- [37] A. D. Thompson, "Google bard (palm 2)," *Life Architect Ai*.
- [38] Y. Li, R. M. Wehbe, F. S. Ahmad, H. Wang, and Y. Luo, "A comparative study of pretrained language models for long clinical text," *Journal of the American Medical Informatics Association*, vol. 30, no. 2, pp. 340–347, 2023.
- [39] M. Learn, "How to work with the chatgpt and gpt-4 models (preview)," 2023.
- [40] B. EDWARDS, "Chatgpt vs google bard: Which is better? we put them to the test," *ars TECHNICA*.
- [41] B. Inkster, S. Sarda, V. Subramanian, *et al.*, "An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: real-world data evaluation mixed-methods study," *JMIR mHealth and uHealth*, vol. 6, no. 11, p. e12106, 2018.
- [42] A. Johnson, "Bard vs. chatgpt: The major difference between the ai chat tools, explained," *Forbes*.
- [43] N. Patel, "Microsoft thinks ai can beat google at search — ceo satya nadella explains why," *The Verge*.
- [44] Y. Mehdi, "Reinventing search with a new ai-powered microsoft bing and edge, your copilot for the web," *Official Microsoft Blog*.
- [45] Y. Mehdi, "Announcing the next wave of ai innovation with microsoft bing and edge," *Official Microsoft Blog*.
- [46] J. Elias, "Google execs tell employees in testy all-hands meeting that bard a.i. isn't just about search," *CNBC*.
- [47] M. Kruppa, "Google ceo sundar pichai says search to include chat ai," *The Wall Street Journal*.
- [48] S. Hsiao, "What's ahead for bard: More global, more visual, more integrated," *Google*.
- [49] J. Vincent, "Google drops waitlist for ai chatbot bard and announces oodles of new features," *The Verge*.
- [50] T. Warren, "Microsoft's bing chatbot now lets you create images via openai's dall-e," *The Verge*.
- [51] C. Velazco, "Meet windows copilot, the ai coming to help you understand your pc," *The Washington Post*.
- [52] Y. Mehdi, "Bing at microsoft build 2023: Continuing the transformation of search," *Microsoft Bing Blog*.
- [53] A. Shafeeg, I. Shazhaev, D. Mihaylov, A. Tularov, and I. Shazhaev, "Voice assistant integrated with chat gpt," *Indonesian Journal of Computer Science*, vol. 12, no. 1, 2023.
- [54] A. Kendon, *Conducting interaction: Patterns of behavior in focused encounters*, vol. 7. CUP Archive, 1990.
- [55] A. P. Chaves and M. A. Gerosa, "How should my chatbot interact? a survey on human-chatbot interaction design. corr abs/1904.02743 (2019)," *arXiv preprint arXiv:1904.02743*, 2019.
- [56] T.-T. Le, "Google bard discusses the subjective sphere optimization process," 2023.
- [57] M. Abdullah, A. Madain, and Y. Jararweh, "Chatgpt: Fundamentals, applications and social impacts," in *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pp. 1–8, 2022.
- [58] R. Brena, "Did chatgpt get a sense of humor?," *Medium*.
- [59] E. Alston, "Chatgpt vs. bard: What's the difference?," *zapier*.

- [60] A. F. Muhammad, D. Susanto, A. Alimudin, F. Adila, M. H. Assidiqi, and S. Nabhan, "Developing english conversation chatbot using dialogflow," in *2020 International Electronics Symposium (IES)*, pp. 468–475, IEEE, 2020.
- [61] Google, "Bard faq,"
- [62] Natalie, "What is chatgpt?," *OpenAi*.
- [63] W. Douglas, "Google just launched bard, its answer to chatgpt—and it wants you to make it better," *MIT Technology Review*.
- [64] H. Alkaissi and S. McFarlane, "Artificial hallucinations in chatgpt: Implications in scientific writing," *Cureus*, 2023.
- [65] H. Alkaissi and S. I. McFarlane, "Artificial hallucinations in chatgpt: implications in scientific writing," *Cureus*, vol. 15, no. 2, 2023.
- [66] Y. Huh, J.-Y. Lee, and S. Finney, "Not an author but an increasingly proficient secretary," *Episodes Journal of International Geoscience*, 2023.
- [67] M. Hosseini, C. A. Gao, D. M. Liebovitz, A. M. Carvalho, F. S. Ahmad, Y. Luo, N. MacDonald, K. L. Holmes, and A. Kho, "An exploratory survey about using chatgpt in education, healthcare, and research," *medRxiv*, pp. 2023–03, 2023.
- [68] T. G. Heck, "What artificial intelligence knows about 70 kda heat shock proteins, and how we will face this chatgpt era," *Cell Stress and Chaperones*, vol. 28, no. 3, pp. 225–229, 2023.
- [69] S. A. Athaluri, S. V. Manthena, V. K. M. Kesapragada, V. Yarlagadda, T. Dave, and R. T. S. Duddumpudi, "Exploring the boundaries of reality: Investigating the phenomenon of artificial intelligence hallucination in scientific writing through chatgpt references," *Cureus*, vol. 15, no. 4, 2023.
- [70] S. Huh, "Are chatgpt's knowledge and interpretation ability comparable to those of medical students in korea for taking a parasitology examination?: a descriptive study," *Journal of Educational Evaluation for Health Professions*, vol. 20, p. 1, 2023.
- [71] C. Zielinski, M. Winker, R. Aggarwal, L. Ferris, M. Heinemann, J. F. Lapeña, S. Pai, L. Citrome, *et al.*, "Chatbots, chatgpt, and scholarly manuscripts-wame recommendations on chatgpt and chatbots in relation to scholarly publications," *Afro-Egyptian Journal of Infectious and Endemic Diseases*, vol. 13, no. 1, pp. 75–79, 2023.
- [72] E. M. Bonsu and D. Baffour-Koduah, "From the consumers' side: Determining students' perception and intention to use chatgpt in ghanaian higher education," *Journal of Education, Society & Multiculturalism*, vol. 4, no. 1, pp. 1–29, 2023.
- [73] M. U. A. T. P. S. M. . R. S. Kumar, M., "Artificial hallucinations by google bard: Think before you leap," *Cureus*, 15(8).
- [74] M. R. King, "Can bard, google's experimental chatbot based on the lambda large language model, help to analyze the gender and racial diversity of authors in your cited scientific references?," *Cellular and Molecular Bioengineering*, pp. 1–5, 2023.
- [75] W. DANIEL, "Google ceo sundar pichai says 'hallucination problems' still plague a.i. tech and he doesn't know why," *Fortune*.
- [76] N. Oh, G.-S. Choi, and W. Y. Lee, "Chatgpt goes to the operating room: evaluating gpt-4 performance and its potential in surgical education and training in the era of large language models," *Annals of Surgical Treatment and Research*, vol. 104, no. 5, p. 269, 2023.
- [77] R. Ali, O. Y. Tang, I. D. Connolly, J. S. Fridley, J. H. Shin, P. L. Zadnik Sullivan, D. Cielo, A. A. Oyelese, C. E. Doberstein, A. E. Telfeian, *et al.*, "Performance of chatgpt, gpt-4, and google bard on a neurosurgery oral boards preparation question bank," *medRxiv*, pp. 2023–04, 2023.
- [78] M. D. Skogen, R. Ji, A. Akimova, U. Daewel, C. Hansen, S. S. Hjøllo, S. M. van Leeuwen, M. Maar, D. Macias, E. A. Mousing, *et al.*, "Disclosing the truth: Are models better than observations?," *Marine Ecology Progress Series*, vol. 680, pp. 7–13, 2021.
- [79] E. Galván, "Neuroevolution in deep learning: The role of neutrality," *arXiv preprint arXiv:2102.08475*, 2021.
- [80] N. Sünderhauf, O. Brock, W. Scheirer, R. Hadsell, D. Fox, J. Leitner, B. Upcroft, P. Abbeel, W. Burgard, M. Milford, *et al.*, "The limits and potentials of deep learning for robotics," *The International journal of robotics research*, vol. 37, no. 4-5, pp. 405–420, 2018.
- [81] M. N.-U.-R. Chowdhury, A. Haque, and H. Soliman, "Chatbots: A Game Changer in mHealth," 6 2023.
- [82] Z. Tafferner, I. Balázs, O. Krammer, and A. Géczy, "Can chatgpt help in electronics research and development? a case study with applied sensors," *Sensors*, vol. 23, no. 10, 2023.
- [83] P. Bailey, "Bard now helps you code," Apr 2023.
- [84] OpenAI, "Gpt-4 technical report," 2023.
- [85] T. H. Kung, M. Cheatham, A. Medenilla, C. Sillos, L. De Leon, C. Elepaño, M. Madriaga, R. Aggabao, G. Diaz-Candido, J. Maningo, and V. Tseng, "Performance of chatgpt on usmle: Potential for ai-assisted medical education using large language models," *PLOS Digital Health*, vol. 2, pp. 1–12, 02 2023.
- [86] R. Gupta, I. Herzog, J. B. Park, J. Weisberger, P. Firouzbakht, V. Ocon, J. Chao, E. S. Lee, and B. A. Mailey, "Performance of ChatGPT on the Plastic Surgery Inservice Training Examination," *Aesthetic Surgery Journal*, 05 2023. sjad128.
- [87] W. A. Passby L, Jenko N, "Performance of chatgpt on dermatology specialty certificate examination multiple choice questions," 2023.
- [88] P. Humar, M. Asaad, F. B. Bengur, and V. Nguyen, "Chatgpt is equivalent to first-year plastic surgery residents: Evaluation of chatgpt on the plastic surgery in-service examination," *Aesthetic Surgery Journal*, May 2023.
- [89] E. Sezgin, F. Chekeni, J. Lee, and S. Keim, "Clinical accuracy of large language models and google search responses to postpartum depression questions: A cross-sectional study (preprint)," 2023.
- [90] E. PRINGLE, "We asked google's a.i. chatbot 'bard' basic sat questions and it would flunk a real exam."
- [91] N. S. Patil, R. S. Huang, C. B. van der Pol, and N. Larocque, "Comparative performance of chatgpt and bard in a text-based radiology knowledge assessment," *Canadian Association of Radiologists Journal*, vol. 0, no. 0, p. 08465371231193716, 0. PMID: 37578849.
- [92] M. Nair, R. Sadhukhan, and D. Mukhopadhyay, "Generating secure hardware using chatgpt resistant to cwes," 2023. <https://eprint.iacr.org/2023/212>.
- [93] S. Addington, "Chatgpt: Cyber security threats and countermeasures," 2023.
- [94] P. Sharma and B. Dash, "Impact of big data analytics and chatgpt on cybersecurity," 03 2023.
- [95] M. Mijwil, M. Aljanabi, and A. Ali, "Chatgpt: Exploring the role of cybersecurity in the protection of medical information," vol. 2023, 02 2023.
- [96] S. Biswas and S. Biswas, "Title: Role of chatgpt in cybersecurity," 03 2023.
- [97] S. Roy, K. Naragam, and S. Nilzadeh, "Generating phishing attacks using chatgpt," 05 2023.
- [98] T. Koide, N. Fukushima, H. Nakano, and D. Chiba, "Detecting phishing sites using chatgpt," 2023.
- [99] G. Sebastian, "Do chatgpt and other ai chatbots pose a cybersecurity risk? - an exploratory study," 02 2023.
- [100] P. Charan, H. Chunduri, P. M. Anand, and S. K. Shukla, "From text to mitre techniques: Exploring the malicious use of large language models for generating cyber attack payloads," *arXiv preprint arXiv:2305.15336*, 2023.
- [101] G. Davis, A. Garcia, and W. Zhang, "Empirical analysis of the effects of cyber security incidents," *Risk Analysis: An International Journal*, vol. 29, no. 9, pp. 1304–1316, 2009.
- [102] D. Krause, "Proper generative ai prompting for financial analysis," *Available at SSRN 4453664*, 2023.
- [103] M. Gupta, C. Akiri, K. Aryal, E. Parker, and L. Praharaj, "From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy," *IEEE Access*, vol. 11, pp. 80218–80245, 2023.
- [104] C. K. Lo, "What is the impact of chatgpt on education? a rapid review of the literature," *Education Sciences*, vol. 13, no. 4, 2023.
- [105] S. S. Gill, M. Xu, P. Patros, H. Wu, R. Kaur, K. Kaur, S. Fuller, M. Singh, P. Arora, A. K. Parlikad, V. Stankovski, A. Abraham, S. K. Ghosh, H. Lutfiyya, S. S. Kanhere, R. Bahsoon, O. Rana, S. Dustdar, R. Sakellariou, S. Uhlig, and R. Buyya, "Transformative effects of chatgpt on modern education: Emerging era of ai chatbots," *Internet of Things and Cyber-Physical Systems*, vol. 4, pp. 19–23, 2024.
- [106] L. Li, Z. Ma, L. Fan, S. Lee, H. Yu, and L. Hemphill, "Chatgpt in education: A discourse analysis of worries and concerns on social media," 04 2023.
- [107] Z. Ilgaz HB, "The significance of artificial intelligence platforms in anatomy education: An experience with chatgpt and google bard," *Cureus*, 15(8).
- [108] H. He, "Robotgpt: From chatgpt to robot intelligence," 04 2023.
- [109] S. Vempkala, R. Bonatti, A. Buckner, and A. Kapoor, "Chatgpt for robotics: Design principles and model abilities," *Tech. Rep. MSR-TR-2023-8*, Microsoft, February 2023.
- [110] H. He, "Robotgpt: From chatgpt to robot intelligence," 2023.

- [111] N. Wake, A. Kanehira, K. Sasabuchi, J. Takamatsu, and K. Ikeuchi, "Chatgpt empowered long-step robot control in various environments: A case application," 2023.
- [112] Y. Ye, H. You, and J. Du, "Improved trust in human-robot collaboration with chatgpt," *IEEE Access*, vol. 11, pp. 55748–55754, 2023.
- [113] G. Shidaganti, R. Sanjana, K. Shubeeksh, V. Monish Raman, and V. Thakshith, "Chatgpt: Information retrieval from image using robotic process automation and ocr," pp. 1264–1270, 2023.
- [114] F. Alexander, E. A. Abdiwijaya, F. Pherry, A. A. S. Gunawan, *et al.*, "Systematic literature review on solving competitive programming problem with artificial intelligence (ai)," in *2022 1st International Conference on Software Engineering and Information Technology (ICoSEIT)*, pp. 85–90, IEEE, 2022.
- [115] "Chatgpt rating on codeforces." availableon:<https://codeforces.com/blog/entry/113910>. [Online, Accessed: 09-06-2023].
- [116] Jun 2023.
- [117] "Hakernoon." availableon:<https://hackernoon.com>. [Online, Accessed: 09-06-2023].
- [118] "Ai coding cometition." availableon:<https://developers.slashdot.org/story/23/04/30/0454245/ai-coding-competition-pits-gpt-4-against-bard-github-co-pilot-bing-and-claude>. [Online, Accessed: 09-06-2023].
- [119] D. Broemmer, "A programming interview contest between chatgpt, bard, and bing," *Medium*.
- [120] "Lru cache," *InterviewCake*.
- [121] A. Koubaa, B. Qureshi, A. Ammar, Z. Khan, W. Boulila, and L. Ghouti, "Humans are still better than chatgpt: Case of the ieeextreme competition," *arXiv preprint arXiv:2305.06934*, 2023.
- [122] K. Xivuri and H. Twinomurizi, "A systematic review of fairness in artificial intelligence algorithms," in *Responsible AI and Analytics for an Ethical and Inclusive Digitized Society: 20th IFIP WG 6.11 Conference on e-Business, e-Services and e-Society, I3E 2021, Galway, Ireland, September 1–3, 2021, Proceedings 20*, pp. 271–284, Springer, 2021.
- [123] N. G. Vidhya, D. Devi, A. Nithya, and T. Manju, "Prognosis of exploration on chat gpt with artificial intelligence ethics," *Brazilian Journal of Science*, vol. 2, no. 9, pp. 60–69, 2023.
- [124] R. Qian, C. Ross, J. Fernandes, E. Smith, D. Kiela, and A. Williams, "Perturbation augmentation for fairer nlp," *arXiv preprint arXiv:2205.12586*, 2022.
- [125] J. Zhang, K. Bao, Y. Zhang, W. Wang, F. Feng, and X. He, "Is chatgpt fair for recommendation? evaluating fairness in large language model recommendation," in *Proceedings of the 17th ACM Conference on Recommender Systems, RecSys '23*, ACM, Sept. 2023.
- [126] S. Singh and N. Ramakrishnan, "Is chatgpt biased? a review," Apr 2023.
- [127] S. Kiri, "Interpretability and fairness in nlp: Learnings from naacl," Aug 2022.
- [128] J. L. Davey Alba, "Google's rush to win in ai led to ethical lapses, employees say," *Bloomberg*.
- [129] A. Haque, M. N.-U.-R. Chowdhury, and H. Soliman, "Transforming Chronic Disease Management with Chatbots: Key Use Cases for Personalized and Cost-effective Care," 6 2023.
- [130] Y. Li and Y. Zhang, "Fairness of chatgpt," *arXiv preprint arXiv:2305.18569*, 2023.
- [131] I. Khan, "Chatgpt vs. bing vs. google bard: Which ai is the most helpful?," *CNET*.
- [132] D. M. West, "Comparing google bard with openai's chatgpt on political bias, facts, and morality," *BROOKINGS*.
- [133] J. Bang, B.-T. Lee, and P. Park, "Examination of ethical principles for llm-based recommendations in conversational ai," in *2023 International Conference on Platform Technology and Service (PlatCon)*, pp. 109–113, IEEE, 2023.
- [134] J. Cabrera, M. S. Loyola, I. Magaña, and R. Rojas, "Ethical dilemmas, mental health, artificial intelligence, and llm-based chatbots," in *International Work-Conference on Bioinformatics and Biomedical Engineering*, pp. 313–326, Springer, 2023.
- [135] "New york times sues openai," Dec 2023.
- [136] H. Dhaduk, "Google bard vs chatgpt: A comparative analysis," Jan 2024.
- [137] T. M. Tawfeeq, A. J. Awqati, and Y. A. Jasim, "The ethical implications of chatgpt ai chatbot: A review," July 2023.
- [138] G. Sebastian, "Exploring ethical implications of chatgpt and other ai chatbots and regulation of disinformation propagation," *Social Science Research Network*, Jan. 2023.
- [139] S. AlZu'bi, A. Mughaid, F. Quia, and S. Hendawi, "Exploring the capabilities and limitations of chatgpt and alternative big language models," in *Artificial Intelligence and Applications*, 2022.
- [140] F. Khoshafah, "Chatgpt for arabic-english translation: Evaluating the accuracy," 2023.
- [141] Y. Kunitsu, "Potential of chatgpt as a support tool for pharmacists: An analytical study using the japanese national examination for pharmacists (preprint)," 2023.
- [142] A. Mihalache, M. M. Popovic, and R. H. Muni, "Performance of an artificial intelligence chatbot in ophthalmic knowledge assessment," *JAMA ophthalmology*, 2023.
- [143] R. Firaina and D. Sulisworo, "Exploring the usage of chatgpt in higher education: Frequency and impact on productivity," *Buletin Edukasi Indonesia*, 2023.
- [144] A. Bard, "Bard's understanding of the costliness of investment in non-reproducible scientific research," 2023.
- [145] C. Kılıç, "The future of jobs: Interviews with artificial intelligence," 2023.
- [146] R. Schroeder, "Towards a theory of digital media," *Information, Communication & Society*, vol. 21, no. 3, pp. 323–339, 2018.
- [147] Y. Kumar, P. Morreale, P. Sorial, J. Delgado, J. J. Li, and P. Martins, "A testing framework for ai linguistic systems (testfails)," 2023.
- [148] R. Anil, A. M. Dai, O. Firat, M. Johnson, D. Lepikhin, A. Passos, S. Shakeri, E. Taropa, P. Bailey, Z. Chen, *et al.*, "Palm 2 technical report," *arXiv preprint arXiv:2305.10403*, 2023.
- [149] May 2023.
- [150] J. O. Ayorinde, F. Citterio, M. Landrò, E. Peruzzo, T. Islam, S. Tilley, G. Taylor, V. Bardsley, P. Liò, A. Samoshkin, *et al.*, "Artificial intelligence you can trust: What matters beyond performance when applying artificial intelligence to renal histopathology?," *Journal of the American Society of Nephrology*, vol. 33, no. 12, pp. 2133–2140, 2022.
- [151] P. Boschee, "Comments: Ai language tools hit the books... and technical content?," *Journal of Petroleum Technology*, vol. 75, no. 04, pp. 8–9, 2023.
- [152] Y. Zhao, "The state-of-art applications of nlp: Evidence from chatgpt," *Highlights in Science, Engineering and Technology*, vol. 49, pp. 237–243, 2023.
- [153] J. D. Weisz, M. Muller, J. He, and S. Houde, "Toward general design principles for generative ai applications," *arXiv preprint arXiv:2301.05578*, 2023.
- [154] M. Butterick, "Github copilot litigation," Nov 2022.
- [155] M. Aljanabi, "Chatgpt: Future directions and open possibilities," *Mesopotamian Journal of CyberSecurity*, vol. 2023, pp. 16–17, 2023.
- [156] T. Barrabi, "Google just launched bard, its answer to chatgpt—and it wants you to make it better," *NewYork Post*.
- [157] A. Ghose and S. P. Han, "An empirical analysis of user content generation and usage behavior on the mobile internet," *Management Science*, vol. 57, no. 9, pp. 1671–1691, 2011.
- [158] N. Maslej, L. Fattorini, E. Brynjolfsson, *et al.*, "Artificial intelligence index report 2023," tech. rep., Technical Report, Stanford University, 2023.