Unleashing the Power of Large Models: Exploring Human-Machine Conversations

Yuhan Liu[1], **Xiuying Chen**[2], **Rui Yan**[1*]
[1]Gaoling School of Artifical Intelligence, Renmin University of China, Beijing, China
[2]King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
yuhan.liu@ruc.edu.cn, xiuying.chen@kaust.edu.sa, ruiyan@ruc.edu.cn

Abstract

In recent years, large language models (LLMs) have garnered significant attention across various domains, resulting in profound impacts. In this paper, we aim to explore the potential of LLMs in the field of human-machine conversations. It begins by examining the rise and milestones of these models, tracing their origins from neural language models to the transformative impact of the Transformer architecture on conversation processing. Next, we discuss the emergence of large pre-training models and their utilization of contextual knowledge at a large scale, as well as the scaling to billion-parameter models that push the boundaries of language generation. We further highlight advancements in multi-modal conversations, showcasing how LLMs bridge the gap between language and vision. We also introduce various applications in human-machine conversations, such as intelligent assistant-style dialogues and emotionally supportive conversations, supported by successful case studies in diverse fields. Lastly, we explore the challenges faced by LLMs in this context and provide insights into future development directions and prospects. Overall, we offer a comprehensive overview of the potential and future development of LLMs in human-machine conversations, encompassing their milestones, applications, and the challenges ahead.

1 Introduction

In recent years, there has been a remarkable surge in the interest and impact of LLMs across diverse domains (Rodriguez, 2022; Khan et al., 2023). These models have revolutionized various fields, and the ability of LLMs to generate coherent and contextually relevant responses has opened up new possibilities for human-machine interaction (OpenAI, 2023). Within this expansive landscape, the realm of human-machine conversations has emerged as a particularly dynamic and rapidly evolving domain. The ability to engage in natural and meaningful dialogue with machines has long been a goal of AI research, and big models have played a pivotal role in making this aspiration a reality.

LLMs are sophisticated artificial intelligence systems that have the ability to process and understand human language at a remarkable scale. These models, such as GPT-3.5(Lin, 2023) and GPT-4 (OpenAI, 2023), are designed to generate text that is coherent and contextually relevant, making them valuable tools for a wide range of applications. In the context of human-machine conversations, LLMs excel at engaging in a natural and interactive dialogue with users. They can comprehend and respond to questions, provide information, offer suggestions, and even simulate human-like conversations. These models leverage vast amounts of pre-existing textual data to learn patterns and generate responses that mimic human conversation, enabling them to understand user input, adapt to different conversational styles, and provide meaningful and coherent answers. The characteristics of LLMs, including their immense size, computational power, and training on diverse datasets, contribute to their ability to generate accurate and contextually appropriate responses, making them valuable assets in enhancing human-machine interactions.

† Corresponding author.

To comprehend the significance of big models in human-machine conversations, it is essential to understand their background and evolution. We provide a diagram to help readers familiarize the overall structure (Figure 1). Section 2 gives a comprehensive overview of the development and milestones of LLMs, tracing their origins from neural language models to the transformative impact of the Transformer architecture on conversation processing. The rise of large pre-training models and their utilization of contextual knowledge at an unprecedented scale will also be explored. Furthermore, this section discusses the scaling to billion-parameter models, pushing the boundaries of language generation and paving the way for more advanced conversational capabilities. One key aspect that will be addressed is the advancement of LLMs in facilitating multi-modal conversations, bridging the gap between language and vision understanding. This opens up opportunities for more natural and immersive interactions between humans and machines. Section 3 focuses on two prominent areas: intelligent assistant-style dialogues and emotionally supportive conversations. Through successful case studies, we demonstrate how LLMs can assist users in various tasks and provide emotional support in sensitive contexts. Despite the promising potential, LLMs face challenges in the context of human-machine conversations as mentioned in Section 4. Ethical concerns, biases, and the need for interpretability are some of the key challenges that need to be addressed to ensure the responsible deployment of these models. Lastly, Section 5 highlights the future directions of development.

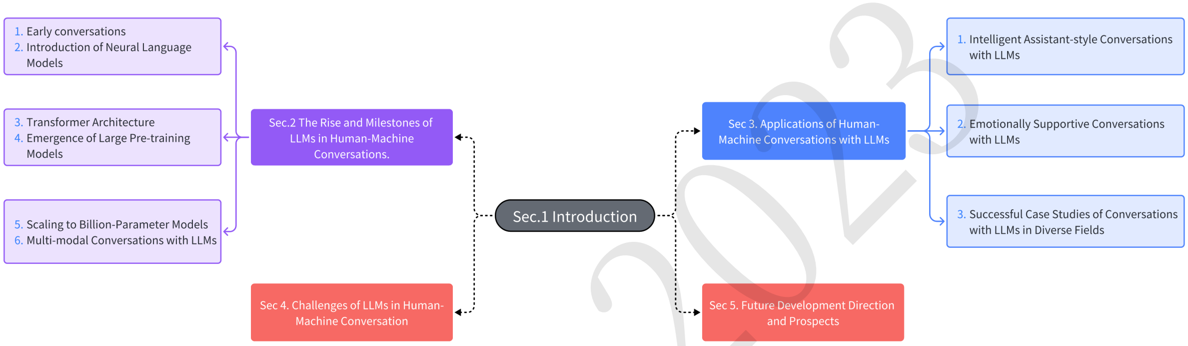

Figure 1: The overall diagram of this article

2 The Rise and Milestones of LLMs in Human-Machine Conversations

2.1 Early Conversations: Tracing the Roots of Conversational AI

Early conversations, such as ELIZA (Weizenbaum, 1966) and ALICE(Marietto et al., 2013), were pioneers in the field of human-machine conversations. However, these early systems had limitations. They could not truly understand the meaning of the user's input and relied heavily on pre-defined rules and patterns. Consequently, these systems often provided generic and impersonal responses.

Despite their shortcomings, early conversations paved the way for advancements in natural language processing and machine learning techniques. Researchers realized the need for more sophisticated models that could learn from data and context, leading to the development of modern LLMs like GPT-3.5.

2.2 Neural Language Models: Opening the Doors to Language Understanding

Neural Language Models (NLMs) have revolutionized the field of human-machine conversations, enabling more dynamic and contextually aware conversations, which leverage deep learning techniques, such as recurrent neural networks (RNNs), to process and understand human language (Sutskever et al., 2014). By training on large-scale datasets, these models learn the statistical patterns and semantic relationships within the language, allowing them to generate more natural and contextually relevant responses. The integration of NLMs into human-machine conversations has significantly improved the quality and naturalness of conversations. These models can consider the context of the conversation, understand nuances, and generate coherent and contextually appropriate responses. Furthermore, they

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 16-29, Harbin, China, August 3 - 5, 2023.
Volume 2: Frotier Forum

can also exhibit a sense of personality, empathy, and adaptability, which enhances user engagement and satisfaction.

While there are still challenges to overcome, such as handling ambiguous queries and maintaining privacy, NLM-based conversations have become invaluable tools for various applications in natural language understanding and interaction. Their ability to generate human-like responses and engage in meaningful conversations opens up new possibilities for human-machine conversations.

2.3 Transformer Architecture: Revolutionizing Conversation Processing

The Transformer (Vaswani et al., 2017) architecture has emerged as a breakthrough in the field of artificial intelligence, particularly in the realm of human-machine conversations. The Transformer model revolutionized the way neural networks process and generate human language. Unlike earlier recurrent neural networks (RNNs) that relied on sequential processing (Sutskever et al., 2014), the Transformer introduced a novel attention mechanism that allowed for parallel processing of words in a sentence. In addition, this architectural innovation overcame the limitations of sequential models, enabling the Transformer to capture long-range dependencies and contextual relationships more effectively. In the context of human-machine conversations, the Transformer architecture has proven highly effective. It excels at understanding and generating coherent responses, exhibiting a level of contextual awareness that makes conversations feel more natural and engaging. Furthermore, the Transformer's architecture allows for parallel processing, making it highly efficient for large-scale training and inference. This scalability has played a pivotal role in training LLMs, such as GPT-3.5, which have pushed the boundaries of human-machine conversations by generating human-like responses across a wide range of topics.

The transformer has significantly improved the quality and coherence of responses, allowing dialogue models to engage in more interactive and contextually aware conversations. With its scalability and versatility, the Transformer architecture continues to drive advancements in natural language understanding and conversational AI systems.

2.4 Emergence of Large Pre-training Models: Harnessing Contextual Knowledge at Scale

Large pre-training models have emerged as game-changer in the field of artificial intelligence, particularly in the domain of human-machine conversations. In the context of human-machine conversations, large pre-training models have shown tremendous potential, which possesses the ability to engage in natural and interactive dialogues with users, simulating human-like conversations.

Model	Publishing Agency	#Parameters	Architecture
BERT (Devlin et al., 2019)	Google AI	110M/340M	
RoBERTa (Liu et al., 2019)	Facebook	123M/354M	
SpanBERT (Joshi et al., 2020)	Stanford	110M/340M	
ERNIE (Sun et al., 2019)	Baidu	110M	
ERNIE-2.0 (Sun et al., 2020)	Baidu	110M/340M	Encoder
ALBERT (Lan et al., 2020)	Google	12M-235M	
DistilBERT (Sanh et al., 2019)	Hugging Face	66M	
ELECTRA (Clark et al., 2020)	Google	14M/110M	
SqueezeBERT (Iandola et al., 2020)	Hugging Face	62M	
GPT (Radford et al., 2018)	OpenAI	117M	Decoder
XLNet (Yang et al., 2019)	CMU & Google	110M/340M	Encoder/Decoder
UniLM (Dong et al., 2019)	Microsoft	340M	
BART (Lewis et al., 2020)	Facebook	140M,406M	Encoder-Decoder
PEGASUS (Zhang et al., 2020)	Google	223M,568M	

Table 1: Overview of Large Pre-training Models

In large pre-training models, as shown in Table 1, we can categorize them into three types based on their model architectures: Encoder-Only, Encoder-Decoder, and Decoder-Only. The Encoder-Only models primarily focus on encoding input data, which transforms textual or other forms of input data into semantic vector representations, where the encoder is responsible for encoding the input information into high-dimensional representations. The Encoder-Decoders model combines the functionalities of both an encoder and a decoder. The encoder encodes the input data into high-dimensional vector representations, while the decoder generates output based on the semantic information provided by the encoder. The Decoder-Only model specializes in generating task-related outputs. This model generates appropriate output sequences by utilizing a decoder based on given conditions or context.

- **Encoder-Only**: BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019) pioneered this trend by leveraging Transformer-based pre-training and utilizing masked language model (MLM) to generate deep bidirectional language representations for comprehensive contextual understanding. Subsequently, RoBERTa (Liu et al., 2019) improved upon BERT by enlarging the dataset, increasing model parameters and batch size, as well as removing the next sentence prediction (NSP) task, leading to enhanced performance through improved text encoding and dynamic masking. SpanBERT (Joshi et al., 2020) introduced novel pre-training objectives specifically designed to better represent the context and establish long-distance dependencies.

  In parallel, Baidu introduced two pre-training models, ERNIE (Sun et al., 2019) and ERNIE 2.0 (Sun et al., 2020), which leveraged large-scale Chinese corpora such as Baidu Baike, Baidu Search, Baidu Zhidao, along with the English Wikipedia, for pre-training. ERNIE 2.0 (Sun et al., 2020) further incorporated a continuous learning semantic understanding framework that continuously learns from massive data and knowledge using techniques like deep neural networks and multi-task learning. ALBERT (Lan et al., 2020), on the other hand, is a lightweight version of BERT that reduces model size while maintaining high performance through parameter factorization of word embeddings and cross-layer parameter sharing.

  The development of these pre-training language models is closely related to the advancements in conversation models. By better understanding context, effectively representing language, and establishing long-distance dependencies, these models provide a foundation and inspiration for conversation construction and optimization. From the lightweight model DistilBERT (Sanh et al., 2019) to the adversarial training-based ELECTRA (Clark et al., 2020) and the smaller and faster SqueezeBERT, these models have not only achieved breakthroughs in performance but also significantly reduced model size and computational costs. They have made important contributions to both academic research and practical applications in the field of human-machine conversation.

- **Encoder-Decoder**: XLNet (Yang et al., 2019), UniLM (Dong et al., 2019), BART (Lewis et al., 2020), and PEGASUS (Zhang et al., 2020) are pre-training models closely related to the development of human-machine conversation. XLNet (Yang et al., 2019) significantly improves text understanding by freely capturing contextual information through a ”permutation-based training” prediction approach. UniLM (Dong et al., 2019), combining the BERT encoder structure with diverse pre-training tasks, demonstrates excellent performance across various natural language processing tasks, making it highly applicable to human-machine conversation research. BART, with its combination of bidirectional encoders and autoregressive decoders, possesses broad adaptability and efficiency, effectively addressing the generation models in conversational systems. PEGASUS (Zhang et al., 2020), utilizing the Transformer architecture and employing the Gap Sentences Generation pre-training objective, comprehends context by generating missing sentences and leverages the Fine-tuning with an Easy Data Selection method for performance enhancement. The introduction of these models has provided new insights and techniques for the development of human-machine conversations, leading to improved performance and efficiency in dialogue modeling, including enhanced contextual processing and generation capabilities.

- **Decoder-Only**: The GPT series (Radford et al., 2018), proposed by OpenAI, is a powerful pre-training language model that achieves remarkable performance in complex NLP tasks without requiring supervised fine-tuning. By increasing the scale of training data and the number of network parameters, the GPT series continually improves its model capacity, thus demonstrating the effectiveness of continuously enhancing model capacity and corpus size.

As a result, large pre-training models have revolutionized human-machine conversations by providing a powerful tool for generating high-quality and contextually appropriate responses.

2.5 Scaling to Billion-Parameter Models: Pushing the Limits of Language Generation

In this paper, we refer to large pre-training models with billion-parameter parameters as LLMs. Billion-parameter models represent a significant milestone in the development of LLMs and have ushered in a new era of human-machine conversations. These models are characterized by their immense size and computational power, pushing the boundaries of what was previously thought possible. It are built upon the foundations of their predecessors, such as GPT-3.5, but with significantly increased capacity. They are trained on vast amounts of textual data from diverse sources, allowing them to capture a wide range of linguistic patterns and semantic relationships. The sheer scale of these models grants them a deeper understanding of human language, resulting in more accurate and contextually appropriate responses.

In the domain of human-machine conversations, billion-parameter models have demonstrated remarkable capabilities. They can engage in natural and interactive dialogues, understand complex queries, and generate highly coherent and contextually relevant responses. These models have the potential to provide users with more personalized and tailored experiences, as they can adapt to different conversational styles and preferences.

Similar to the pre-training models mentioned in the subsection 2.4, as shown in Table 2, billion-scale language models are primarily divided into Encoder-Decoder architectures and Decoder-only architectures. Moreover, as the model size increases, the model structures tend to become more standardized.

- **Encoder-Decoder**: The large-scale language models, namely T5 (Raffel et al., 2020), ERNIE-3.0 (Sun et al., 2021), ERNIE-3.0 Titan (Wang et al., 2021), PaLM-2 (Google, 2023), and GLM-130B (Zeng et al., 2022), which adopt the Encoder-Decoder architecture, play a significant role in the human-machine conversation. Google's T5 (Raffel et al., 2020) approaches all NLP tasks as "text-to-text" problems, which grants it excellent adaptability when dealing with human-machine conversation tasks. Baidu's ERNIE-3.0 (Sun et al., 2021) and ERNIE-3.0 Titan (Wang et al., 2021) demonstrate outstanding performance in knowledge enhancement and self-supervised learning, making them particularly effective in handling knowledge-driven human-machine conversations. Google's PaLM-2 (Google, 2023), with its advanced reasoning capabilities, is especially well-suited for handling complex human-machine conversation scenarios. On the other hand, Tsinghua University's GLM-130B model (Zeng et al., 2022), as a bilingual model, is particularly suitable for addressing cross-lingual human-machine conversation tasks. The development and application of these models have greatly enhanced the capabilities of human-machine conversation in understanding, reasoning, and generating dialogue content, thereby significantly improving the performance and user experience of such systems.

- **Decoder-Only**: Large-scale pre-training models such as GPT2 (Brown et al., 2020), GPT3 (Ye et al., 2023), GPT3.5(Lin, 2023), FLAN (Wei et al., 2022), InstructionGPT (Ouyang et al., 2022), PaLM (Chowdhery et al., 2022), OPT (Zhang et al., 2022), Bloom (Scao et al., 2022), FLAN-PaLM (Chung et al., 2022), and LLaMA (Touvron et al., 2023) have played a crucial role in various domains, including human-machine conversation, natural language understanding, and generation. The GPT series models from OpenAI (Brown et al., 2020; Ye et al., 2023; Lin, 2023), along with OPT (Zhang et al., 2022) and LLaMA(Touvron et al., 2023) from Meta AI, leverage their extensive parameters and complex model structures to provide robust semantic understanding and response generation capabilities for human-machine conversation. Google's FLAN and FLAN-PaLM enhance the model's handling of unknown questions and generalization abilities in human-machine conversation through instruction fine-tuning techniques. The InstructionGPT (Ouyang et al., 2022) optimizes GPT-3 to address toxic language and misinformation issues that may arise in human-machine conversation. The PaLM (Chowdhery et al., 2022), trained by Google using large-scale datasets and a distributed training architecture, enables the handling of complex human-machine conversation tasks. Bloom (Scao et al., 2022), as an open-source model with extensive parameters that supports multiple languages, offers powerful support for multilingual human-machine conversation scenarios.

Model	Publishing Agency	#Parameters	Architecture
T5 (Raffel et al., 2020)	Google Brain	220M-11B	Encoder-Decoder
ERNIE-3.0 (Sun et al., 2021)	Baidu	10B	
ERNIE-3.0 Titan (Wang et al., 2021)	Baidu	260B	
PaLM-2 (Google, 2023)	Google	1.04B-2.7B	
GLM-130B (Zeng et al., 2022)	Zhipu.AI	100M-515M	
GPT-2 (Brown et al., 2020)	OpenAI	1.5B	Decoder
GPT-3 (Ye et al., 2023)		2.6B-200B	
GPT-3.5 (Lin, 2023)		-	
FLAN (Wei et al., 2022)	Google	137B	
InstructGPT (Ouyang et al., 2022)	OpenAI	1.3B-175B	
PaLM (Chowdhery et al., 2022)	Google	8B-540B	
OPT (Zhang et al., 2022)	Meta AI	6.7B-175B	
Bloom (Scao et al., 2022)	HuggingFace	560M-176B	
FLAN-PaLM (Chung et al., 2022)	THUNLP	250M-11B	
LLaMA (Touvron et al., 2023)	Stanford	780M-65B	

Table 2: Overview of Large Laguage Models

Nevertheless, billion-parameter models hold tremendous promise for the future of human-machine conversations. As they continue to evolve, they have the potential to revolutionize various domains, including customer support, virtual assistants, education, creative writing, and more. Their ability to generate human-like responses and engage in meaningful interactions opens up new possibilities for enhancing user experiences and pushing the boundaries of conversational artificial intelligence.

2.6 Multi-modal Conversations with LLMs: Bridging Language and Vision

Multi-modal conversation in LLMs represents an exciting frontier in the field of artificial intelligence, particularly in the context of human-machine conversations. Traditionally, language models have focused primarily on text-based interactions. However, with advancements in computer vision and multi-modal learning, there is a growing interest in incorporating visual and other modalities into conversations.

In recent years, a major focus in the field of artificial intelligence has been on multi-modal large-scale pre-training models, as shown in Table 3, which goal is to enable machines to understand and generate various modalities of human conversations, including text, images, and sound. These models have played a crucial role in making human-machine conversations more natural, rich, and intelligent. For example, the PaLM-E (Driess et al., 2023), jointly developed by Google and the TUB, is an embodied vision and language model. It is a generative model that takes multi-modal sentences as input to generate text, providing natural and coherent responses for human-machine conversations. OpenAI's CLIP (Radford et al., 2021), on the other hand, employs contrastive learning to enable machines to understand the relationship between images and text, providing a powerful tool for understanding user-provided image inputs and generating relevant descriptions.

DeepMind's Flamingo (Alayrac et al., 2022) and Google's CoCa (Yu et al., 2022) establish connections between visual and language modalities. They are capable of processing and understanding both visual and textual data, providing support for image understanding and description in human-machine conversations. The Flamingo, in particular, can handle arbitrary interleaved sequences of visual and

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 16-29, Harbin, China, August 3 - 5, 2023.
Volume 2: Frotier Forum

textual data, seamlessly processing image or video inputs. Google's PaLI (Chen et al., 2022a) and Alibaba DAMO Academy's OFA (Wang et al., 2022a), both with multilingual and multi-modal capabilities, support multiple languages and understand inputs from various modalities, allowing them to adapt to different human-computer dialogue environments and requirements.

Microsoft's BEiT-3 (Wang et al., 2022b) and Salesforce Research's BLIP (Li et al., 2022) and BLIP-2 (Li et al., 2023) establish deeper connections between visual and language modalities. Through their deep understanding and generation of images and text, these models provide richer and more accurate responses. The KOSMOS-1 (Huang et al., 2023b), a multi-modal large-scale language model, has a Transformer-based causal language model as its backbone. It can integrate inputs from language, vision, and other modalities, enabling it to consider information from multiple modalities when understanding user input and generating responses.

Finally, OpenAI's GPT-4 (OpenAI, 2023) is a novel language model that has been improved in terms of creativity, visual input, and longer contexts, which allows it to generate more natural, coherent, and relevant responses. Overall, these multi-modal large-scale pre-training models have their unique advantages and characteristics, and they have all contributed to the advancement of human-machine conversations to varying degrees.

Model	Publishing Agency	#Parameters
PALM-E (Driess et al., 2023)	Google & TUB	562B
CLIP (Radford et al., 2021)	OpenAI	428M
Flamingo (Alayrac et al., 2022)	DeepMind	3B-80B
CoCa (Yu et al., 2022)	Google	383M-2.1B
PaLI (Chen et al., 2022a)	Google	3B-17B
OFA (Wang et al., 2022a)	DAMO Academy, Alibaba	33M-930M
BEiT-3 (Wang et al., 2022b)	Microsoft	1.9B
BLIP (Li et al., 2022)	Salesforce	446M
BLIP-2 (Li et al., 2023)	Salesforce	474M-1.2B
KOSMOS-1 (Huang et al., 2023b)	Microsoft	1.6B
GPT-4 (OpenAI, 2023)	OpenAI	-

Table 3: Overview of Multimodal LLMs

Multi-modal conversation in LLMs also holds promise for applications such as virtual assistants, interactive storytelling, and social chatbots. For instance, a virtual assistant equipped with multi-modal capabilities can process both text and images to provide more accurate and contextually relevant responses. Multi-modal dialogue in LLMs has the potential to reshape the landscape of human-machine conversations, creating more immersive and context-aware interactions that better align with human communication modalities. The integration of multi-modal capabilities in LLMs enables them to comprehend not just the text but also the contextual visual information, allowing for more contextually appropriate responses. This opens up new possibilities for more dynamic and engaging human-machine interactions.

3 Applications of Human-Machine Conversations with LLMs

The application of LLMs for human-machine conversations is revolutionizing various industries by harnessing the capabilities of intelligent assistant systems and emotionally supportive conversations. In the field of intelligent assistant-based human-machine conversation, LLMs have significantly improved user experiences through natural language interactions and personalized recommendations. Another notable application of human-machine conversation with LLMs is emotional support conversations. These systems aim to establish empathy with users, provide emotional support, and engage in meaningful conversations. By analyzing user inputs and offering appropriate responses, emotional support dialogues can help individuals cope with stress, anxiety, or loneliness. Such systems have shown promising results in supporting mental health by providing users with a safe and confidential environment to express their feelings and receive guidance. Successful cases of LLMs in human-machine conversation have emerged across various domains, demonstrating their transformative impact.

3.1 Intelligent Assistant-style Conversations with LLMs

Intelligent Assistant-Style Human-Machine conversations system, exemplified by popular platforms such as Siri, Alexa, and ChatGPT, has revolutionized the way users interact with technology. These systems are primarily designed to address information needs and provide personalized assistance to users. With a deep understanding of various business processes, they can offer comprehensive responses and fulfill a wide range of user inquiries. Their rich knowledge base allows them to handle tasks such as answering questions, providing recommendations, and assisting with navigation, making them invaluable tools for traditional customer service products.

Modern Human-Machine conversations system are rapidly advancing, giving rise to a series of remarkable models. OpenAI's GPT-4 (OpenAI, 2023) is a large-scale multi-modal model that accepts both image and text inputs and produces text outputs. While its capabilities still fall short of humans in certain real-world scenarios, GPT-4 has demonstrated human-level performance on many professional and academic benchmarks, particularly in the domains of creativity, visual input processing, and understanding longer contexts. Furthermore, OpenAI has developed an AI chatbot called ChatGPT (Lin, 2023), based on GPT-3.5 and GPT-4 architectures. It engages in text-based interactions and leverages reinforcement learning techniques to provide useful outputs.

In contrast, Google's Bard (bar, 2023) is a chatbot built on the large language model LaMDA. Its lightweight version extends to a broader user base while collecting and applying user feedback to continuously improve model performance. Claude (Bai et al., 2022), developed by Anthropic, is another large-scale language model designed to detect and avoid pitfalls such as logical errors and inappropriate content that ChatGPT may encounter. The model emphasizes usefulness and harmlessness, employing the RLAIF algorithm.

For the Chinese-English bilingual environment, ChatGLM-6B (Du et al., 2022) from Tsinghua University is a language model with billions of parameters optimized specifically for Chinese. It supports local deployment on consumer-grade graphics cards. Baidu's ERNIE Bot (Sun et al., 2021) is a generative dialogue product built on the ERNIE model series, leveraging the power of the large language model ERNIE 3.0-Titan, showcasing excellent text understanding and generation capabilities. Finally, MOSS (mos, 2023) from Fudan University, as the first large-scale language model in China similar to ChatGPT, offers enhanced functionalities through plugins, such as support for search engines, image generation, calculators, equation solvers, and more, providing a richer interactive experience.

3.2 Emotionally Supportive Conversations with LLMs

Emotionally Supportive Human-Machine conversations have revolutionized the field of human-machine interaction by focusing on emotions and social interaction. These systems aim to provide users with information, emotional support, and engaging conversations. With rich emotions, knowledge, and personality as their main characteristics, these conversations can empathize with users, understand their emotional states, and respond accordingly.

In the field of emotion-aware human-machine conversations, leading technology giants and academic institutions such as Google, Meta, and Baidu have developed various outstanding models. Google's Meena (Adiwardana et al., 2020), a chatbot developed with 2.6 billion parameters and trained on 341GB of social media conversation text, demonstrates human-level coherence and specificity in its responses. Meta's BlenderBot (Roller et al., 2021), on the other hand, is an open-domain dialogue bot with capabilities for online searching and long-term memory. It is built upon deep learning models and is trained to engage in interactive and responsive conversations. Another conversational application model by Google, LaMDA (Thoppilan et al., 2022), can learn discussions on various topics and exhibits impressive coherence and specificity in its responses after training and fine-tuning.

Baidu's PLATO (Bao et al., 2020) is a large-scale open-domain dialogue generation network that models background knowledge using discrete latent variables. In the Chinese dialogue model domain, EVA (Zhou et al., 2021) and OPD(opd, 2023) have demonstrated notable performance. EVA is a large-scale

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 16-29, Harbin, China, August 3 - 5, 2023.
Volume 2: Frotier Forum

Chinese open-domain dialogue pre-training model that surpasses other Chinese pre-training dialogue models in both automatic and human evaluation metrics. Its subsequent version, EVA2.0, has been optimized in various aspects, and the 300M-parameter EVA2.0 (Gu et al., 2023) achieves the performance of the 2.8B-parameter EVA1.0. OPD (opd, 2023), on the other hand, is currently the world's largest open-source Chinese dialogue pre-training model with 6.3 billion parameters. It exhibits excellent chit-chat capability and knowledge question-answering ability, enabling in-depth multi-turn dialogue interactions with users.

3.3 Successful Case Studies of Human-Machine Conversations with LLMs in Diverse Fields

Successful case studies of LLM in human-machine conversations have demonstrated their effectiveness and impact across various fields. These systems have streamlined tasks such as content generation (Alkaissi and McFarlane, 2023; Rodriguez, 2022), disease diagnosis and treatment(Duong and Solomon, 2023; Khan et al., 2023; Rao et al., 2023a; Rao et al., 2023b), and assisted software development(Amos, 2023; Castelvecchi, 2022; Surameery and Shakor, 2023), enhancing user experiences and improving productivity. In addition, The integration of ChatGPT into the realm of data processing has the potential to revolutionize the landscape of scientific research (Macdonald et al., 2023).

In the field of biomedicine, LLMs such as LLaMa (Touvron et al., 2023) and ChatGLM (Du et al., 2022) often underperform due to a lack of specialized medical knowledge. To address this issue, HuaTuo (Wang et al., 2023) has developed a Chinese medical instruction dataset using a combination of medical knowledge graph and the GPT3.5 API. Additionally, leveraging the same medical data, this project also trained a healthcare-oriented version of the ChatGLM model: ChatGLM-6B-Med.Bloomberg has released BloombergGPT (Wu et al., 2023a), which is specifically trained on various financial data to comprehensively support natural language processing tasks in the financial domain.

Overall, the successful case studies of LLM human-machine conversations in various fields highlight their potential to transform industries, optimize processes, and enhance human-machine interactions.

4 Challenges of LLMs in Human-Machine Conversation

The use of LLMs in human-machine conversations presents several challenges that researchers and developers need to address:

- **Data Bias and Ethical Issues**: LLMs are trained on vast amounts of data, which may inadvertently reflect biases present in the data. This can lead to biased responses or perpetuation of stereotypes (Azaria, 2023). It is crucial to identify and mitigate these biases to ensure fair and inclusive interactions. Additionally, the ethical implications of deploying powerful conversations should be carefully considered, such as the potential for misuse or manipulation of information (Liebrenz et al., 2023).

- **Explainability and Transparency**: LLMs operate as complex black boxes, making it difficult to understand their decision-making processes. Users and stakeholders may have concerns about how the models arrive at their responses or recommendations(Larsson and Heintz, 2020). Ensuring transparency and providing explanations for the system's behavior are essential to build trust and accountability (Wischmeyer, 2020; OpenAI, 2023).

- **Security and Malicious Use**: As LLMs become more powerful, there is an increased risk of them being exploited for malicious purposes, such as generating deceptive or harmful content(Ali and others, 2023). Protecting the integrity of conversations and preventing malicious use is a significant concern that requires robust security measures and monitoring (Hargreaves, 2023; Kasneci et al., 2023).

- **Incorrect, Long-term Memory and Persistence**: Despite advancements, conversation models can still produce inaccurate or nonsensical responses. Ensuring the systems have reliable mechanisms for validation and error correction is essential. Additionally, conversations should be able to maintain a coherent context and memory over extended conversations, as well as recognize and address inconsistencies in their responses. (Blog, 2023; Borji, 2023; Zhuo et al., 2023; OpenAI, 2023).

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 16-29, Harbin, China, August 3 - 5, 2023.
Volume 2: Frotier Forum

Addressing these challenges is essential for the responsible and effective use of large models in human-machine conversations. Researchers and practitioners need to collaborate and innovate to ensure ethical considerations, transparency, security, reliability, and resource efficiency in the development and deployment of these powerful conversations.

5 Future Development Direction and Prospects

This section explores the future development direction and prospects, outlining the potential pathways and opportunities that lie ahead in the field:

- **Support and Research for Low-Resource Languages**: LLMs have demonstrated remarkable performance in high-resource languages, resources and data are scarce available for low-resource languages (Huang et al., 2023a). It is crucial to invest in research and develop techniques to make these models more accessible and effective in low-resource language settings, enabling users from diverse linguistic backgrounds to benefit from conversations (Mohtashami et al., 2023).

- **Model Explainability and Transparency**: Enhancing model explainability and transparency is another significant prospect. Ensuring that these models provide interpretable and transparent responses is essential to build user trust and understand how the system arrives at its conclusions (Wu et al., 2023b). Explainability and transparency is an ongoing area of research in the field.

- **Personalized Conversations and Intelligent Assistants**: LLMs have the potential to offer personalized experiences, but there are challenges in understanding and adapting to individual user preferences, needs, and contexts. Designing conversations that can accurately capture and incorporate user feedback, dynamically adapt to user preferences, and provide personalized recommendations is a complex task that requires further research and development (Chen et al., 2022b).

- **Social Applications and Industrial Adoption**: There is a need to explore social applications and promote the industrial adoption of large models in human-machine conversations. Integrating conversations into social platforms and applications can enhance user experiences, facilitate social interactions, and offer new opportunities for information access and engagement (Zhao et al., 2023). Encouraging the adoption of large models in various industries, such as healthcare, finance, and entertainment, can lead to significant advancements and real-world impact in these domains.

These prospects will contribute to the advancement and responsible deployment of large models in human-machine conversations. Continued research, collaboration, and innovation are necessary to overcome these obstacles and unlock the full potential of large models in transforming the way humans interact with machines.

6 Summary

In this paper, we investigates the role of LLMs in facilitating human-machine dialogue. It examines the rise and development of these models, explores their applications in various domains, and discusses the challenges associated with their deployment. Furthermore concludes by outlining future directions and prospects, highlighting the need for ongoing research and addressing ethical considerations. It serves as a valuable resource for researchers and practitioners interested in leveraging the potential of large models in human-machine conversations.

References

Daniel Adiwardana, Minh-Thang Luong, David R So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, et al. 2020. Towards a human-like open-domain chatbot. *arXiv preprint arXiv:2001.09977.*

Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, et al. 2022. Flamingo: a visual language model for few-shot learning. *Proc. of NeurIPS*, pages 23716–23736.

Faizan Ali et al. 2023. Let the devil speak for itself: Should chatgpt be allowed or banned in hospitality and tourism schools? *Journal of Global Hospitality and Tourism*, pages 1–6.

Hussam Alkaissi and Samy I McFarlane. 2023. Artificial hallucinations in chatgpt: implications in scientific writing. *Cureus*.

David Amos. 2023. Chatgpt is an extra-ordinary python programmer.

Amos Azaria. 2023. Chatgpt: More human-like than computer-like, but not necessarily in a good way.

Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, Carol Chen, Catherine Olsson, Christopher Olah, Danny Hernandez, Dawn Drain, Deep Ganguli, Dustin Li, Eli Tran-Johnson, Ethan Perez, Jamie Kerr, Jared Mueller, Jeffrey Ladish, Joshua Landau, Kamal Ndousse, Kamile Lukosuite, Liane Lovitt, Michael Sellitto, Nelson Elhage, Nicholas Schiefer, Noemi Mercado, Nova DasSarma, Robert Lasenby, Robin Larson, Sam Ringer, Scott Johnston, Shauna Kravec, Sheer El Showk, Stanislav Fort, Tamera Lanham, Timothy Telleen-Lawton, Tom Conerly, Tom Henighan, Tristan Hume, Samuel R. Bowman, Zac Hatfield-Dodds, Ben Mann, Dario Amodei, Nicholas Joseph, Sam McCandlish, Tom Brown, and Jared Kaplan. 2022. Constitutional ai: Harmlessness from ai feedback.

Siqi Bao, Huang He, Fan Wang, Hua Wu, and Haifeng Wang. 2020. PLATO: Pre-trained dialogue generation model with discrete latent variable. In *Proc. of ACL*, pages 85–96.

2023. Bard Google. https://bard.google.com.

Back To Blog. 2023. Ai and academic integrity: How ai technology might influence the future of scholarly publishing.

Ali Borji. 2023. A categorical archive of chatgpt failures. *arXiv preprint arXiv:2302.03494*.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Proc. of NeurIPS*, pages 1877–1901.

Davide Castelvecchi. 2022. Are chatgpt and alphacode going to replace programmers? *Nature*.

Xi Chen, Xiao Wang, Soravit Changpinyo, AJ Piergiovanni, Piotr Padlewski, Daniel Salz, Sebastian Goodman, Adam Grycner, Basil Mustafa, Lucas Beyer, et al. 2022a. Pali: A jointly-scaled multilingual language-image model. *arXiv preprint arXiv:2209.06794*.

Yirong Chen, Weiquan Fan, Xiaofen Xing, Jianxin Pang, Minlie Huang, Wenjing Han, Qianfeng Tie, and Xiangmin Xu. 2022b. Cped: A large-scale chinese personalized and emotional dialogue dataset for conversational ai.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. 2022. Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311*.

Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Eric Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, et al. 2022. Scaling instruction-finetuned language models. *arXiv preprint arXiv:2210.11416*.

Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. ELECTRA: pre-training text encoders as discriminators rather than generators. In *Proc. of ICLR*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proc. of NAACL*, pages 4171–4186.

Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, and Hsiao-Wuen Hon. 2019. Unified language model pre-training for natural language understanding and generation. *Proc. of NeurIPS*.

Danny Driess, Fei Xia, Mehdi SM Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, et al. 2023. Palm-e: An embodied multimodal language model. *arXiv preprint arXiv:2303.03378*.

Zhengxiao Du, Yujie Qian, Xiao Liu, Ming Ding, Jiezhong Qiu, Zhilin Yang, and Jie Tang. 2022. GLM: General language model pretraining with autoregressive blank infilling. In *Proc. of ACL*, pages 320–335.

Dat Duong and Benjamin D Solomon. 2023. Analysis of large-language model versus human performance for genetics questions. *medRxiv*, pages 2023–01.

Google. 2023. Palm 2.

Yuxian Gu, Jiaxin Wen, Hao Sun, Yi Song, Pei Ke, Chujie Zheng, Zheng Zhang, Jianzhu Yao, Lei Liu, Xiaoyan Zhu, and Minlie Huang. 2023. EVA2.0: investigating open-domain chinese dialogue systems with large-scale pre-training. *Mach. Intell. Res.*, pages 207–219.

Stuart Hargreaves. 2023. 'words are flowing out like endless rain into a paper cup': Chatgpt & law school assessments. *The Chinese University of Hong Kong Faculty of Law Research Paper*.

Haoyang Huang, Tianyi Tang, Dongdong Zhang, Wayne Xin Zhao, Ting Song, Yan Xia, and Furu Wei. 2023a. Not all languages are created equal in llms: Improving multilingual capability by cross-lingual-thought prompting.

Shaohan Huang, Li Dong, Wenhui Wang, Yaru Hao, Saksham Singhal, Shuming Ma, Tengchao Lv, Lei Cui, Owais Khan Mohammed, Qiang Liu, et al. 2023b. Language is not all you need: Aligning perception with language models. *arXiv preprint arXiv:2302.14045*.

Forrest N. Iandola, Albert E. Shaw, Ravi Krishna, and Kurt Keutzer. 2020. Squeezebert: What can computer vision teach NLP about efficient neural networks? In *Proceedings of SustaiNLP: Workshop on Simple and Efficient Natural Language Processing, SustaiNLP@EMNLP 2020, Online, November 20, 2020*, pages 124–135.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S Weld, Luke Zettlemoyer, and Omer Levy. 2020. Spanbert: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics*, pages 64–77.

Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. 2023. Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and Individual Differences*, page 102274.

Rehan Ahmed Khan, Masood Jawaid, Aymen Rehan Khan, and Madiha Sajjad. 2023. Chatgpt-reshaping medical education and clinical management. *Pakistan Journal of Medical Sciences*, page 605.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2020. ALBERT: A lite BERT for self-supervised learning of language representations. In *Proc. of ICLR*.

Stefan Larsson and Fredrik Heintz. 2020. Transparency in artificial intelligence. *Internet Policy Review*.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proc. of ACL*, pages 7871–7880.

Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. 2022. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *Proc. of ICML*, pages 12888–12900.

Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. *arXiv preprint arXiv:2301.12597*.

Michael Liebrenz, Roman Schleifer, Anna Buadze, Dinesh Bhugra, and Alexander Smith. 2023. Generating scholarly content with chatgpt: ethical challenges for medical publishing. *The Lancet Digital Health*, pages e105–e106.

Zhicheng Lin. 2023. Why and how to embrace ai such as chatgpt in your academic life.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Calum Macdonald, Davies Adeloye, Aziz Sheikh, and Igor Rudan. 2023. Can chatgpt draft a research article? an example of population-level vaccine effectiveness analysis. *Journal of global health*.

Maria das Graças Bruno Marietto, Rafael Varago de Aguiar, Gislene de Oliveira Barbosa, Wagner Tanaka Botelho, Edson Pimentel, Robson dos Santos França, and Vera Lúcia da Silva. 2013. Artificial intelligence markup language: a brief tutorial. *arXiv preprint arXiv:1307.3091*.

Amirkeivan Mohtashami, Mauro Verzetti, and Paul K. Rubenstein. 2023. Learning translation quality evaluation on low resource languages from large language models.

2023. MOSS: Openlmlab's repository. https://github.com/OpenLMLab/MOSS.

2023. OPD: Open-domain pre-trained dialogue model. https://github.com/thu-coai/OPD.

OpenAI. 2023. Gpt-4 technical report.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Proc. of NeurIPS*, pages 27730–27744.

Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. 2018. Improving language understanding by generative pre-training.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *Proc. of ICML*, pages 8748–8763.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, pages 5485–5551.

Arya Rao, John Kim, Meghana Kamineni, Michael Pang, Winston Lie, and Marc D Succi. 2023a. Evaluating chatgpt as an adjunct for radiologic decision-making. *medRxiv*, pages 2023–02.

Arya S Rao, Michael Pang, John Kim, Meghana Kamineni, Winston Lie, Anoop K Prasad, Adam Landman, Keith Dryer, and Marc D Succi. 2023b. Assessing the utility of chatgpt throughout the entire clinical workflow. *medRxiv*, pages 2023–02.

Jesus Rodriguez. 2022. How to create diagrams with chatgpt.

Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. 2021. Recipes for building an open-domain chatbot. In *Proc. of EACL*, pages 300–325.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, et al. 2022. Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100*.

Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, and Hua Wu. 2019. Ernie: Enhanced representation through knowledge integration. *arXiv preprint arXiv:1904.09223*.

Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Hao Tian, Hua Wu, and Haifeng Wang. 2020. Ernie 2.0: A continual pre-training framework for language understanding. In *Proc. of AAAI*, pages 8968–8975.

Yu Sun, Shuohuan Wang, Shikun Feng, Siyu Ding, Chao Pang, Junyuan Shang, Jiaxiang Liu, Xuyi Chen, Yanbin Zhao, Yuxiang Lu, et al. 2021. Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation. *arXiv preprint arXiv:2107.02137*.

Nigar M Shafiq Surameery and Mohammed Y Shakor. 2023. Use chat gpt to solve programming bugs. *International Journal of Information Technology & Computer Engineering (IJITC) ISSN: 2455-5290*, pages 17–22.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. *Proc. of NeurIPS*.

Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, YaGuang Li, Hongrae Lee, Huaixiu Steven Zheng, Amin Ghafouri, Marcelo Menegali, Yanping Huang, Maxim Krikun, Dmitry Lepikhin, James Qin, Dehao Chen, Yuanzhong Xu, Zhifeng Chen, Adam Roberts, Maarten Bosma, Vincent Zhao, Yanqi Zhou, Chung-Ching Chang, Igor Krivokon, Will Rusch, Marc Pickett, Pranesh Srinivasan, Laichee Man, Kathleen Meier-Hellstern, Meredith Ringel Morris, Tulsee Doshi, Renelito Delos Santos, Toju Duke, Johnny Soraker, Ben Zevenbergen, Vinodkumar Prabhakaran, Mark Diaz, Ben Hutchinson, Kristen Olson, Alejandra Molina, Erin Hoffman-John, Josh Lee, Lora Aroyo, Ravi Rajakumar, Alena Butryna, Matthew Lamm, Viktoriya Kuzmina, Joe Fenton, Aaron Cohen, Rachel Bernstein, Ray Kurzweil, Blaise Aguera-Arcas, Claire Cui, Marian Croak, Ed Chi, and Quoc Le. 2022. Lamda: Language models for dialog applications.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Proc. of NeurIPS*.

Shuohuan Wang, Yu Sun, Yang Xiang, Zhihua Wu, Siyu Ding, Weibao Gong, Shikun Feng, Junyuan Shang, Yanbin Zhao, Chao Pang, et al. 2021. Ernie 3.0 titan: Exploring larger-scale knowledge enhanced pre-training for language understanding and generation. *arXiv preprint arXiv:2112.12731*.

Peng Wang, An Yang, Rui Men, Junyang Lin, Shuai Bai, Zhikang Li, Jianxin Ma, Chang Zhou, Jingren Zhou, and Hongxia Yang. 2022a. Ofa: Unifying architectures, tasks, and modalities through a simple sequence-to-sequence learning framework. In *Proc. of ICML*, pages 23318–23340.

Wenhui Wang, Hangbo Bao, Li Dong, Johan Bjorck, Zhiliang Peng, Qiang Liu, Kriti Aggarwal, Owais Khan Mohammed, Saksham Singhal, Subhojit Som, et al. 2022b. Image as a foreign language: Beit pretraining for all vision and vision-language tasks. *arXiv preprint arXiv:2208.10442*.

Haochun Wang, Chi Liu, Nuwa Xi, Zewen Qiang, Sendong Zhao, Bing Qin, and Ting Liu. 2023. Huatuo: Tuning llama model with chinese medical knowledge.

Jason Wei, Maarten Bosma, Vincent Y. Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M. Dai, and Quoc V. Le. 2022. Finetuned language models are zero-shot learners. In *Proc. of ICLR*.

Joseph Weizenbaum. 1966. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, pages 36–45.

Thomas Wischmeyer. 2020. Artificial intelligence and transparency: opening the black box. *Regulating artificial intelligence*, pages 75–101.

Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023a. Bloomberggpt: A large language model for finance.

Zhaofeng Wu, William Merrill, Hao Peng, Iz Beltagy, and Noah A. Smith. 2023b. Transparency helps reveal when language models learn meaning.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *Proc. of NeurIPS*.

Junjie Ye, Xuanting Chen, Nuo Xu, Can Zu, Zekai Shao, Shichun Liu, Yuhan Cui, Zeyang Zhou, Chao Gong, Yang Shen, et al. 2023. A comprehensive capability analysis of gpt-3 and gpt-3.5 series models. *arXiv preprint arXiv:2303.10420*.

Jiahui Yu, Zirui Wang, Vijay Vasudevan, Legg Yeung, Mojtaba Seyedhosseini, and Yonghui Wu. 2022. Coca: Contrastive captioners are image-text foundation models. *Trans. Mach. Learn. Res.*

Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, Weng Lam Tam, Zixuan Ma, Yufei Xue, Jidong Zhai, Wenguang Chen, Peng Zhang, Yuxiao Dong, and Jie Tang. 2022. Glm-130b: An open bilingual pre-trained model.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter Liu. 2020. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In *Proc. of ICML*, pages 11328–11339.

Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, Mona Diab, Xian Li, Xi Victoria Lin, et al. 2022. Opt: Open pre-trained transformer language models. *arXiv preprint arXiv:2205.01068*.

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. 2023. A survey of large language models.

Hao Zhou, Pei Ke, Zheng Zhang, Yuxian Gu, Yinhe Zheng, Chujie Zheng, Yida Wang, Chen Henry Wu, Hao Sun, Xiaocong Yang, Bosi Wen, Xiaoyan Zhu, Minlie Huang, and Jie Tang. 2021. Eva: An open-domain chinese dialogue system with large-scale generative pre-training.

Terry Yue Zhuo, Yujin Huang, Chunyang Chen, and Zhenchang Xing. 2023. Exploring ai ethics of chatgpt: A diagnostic analysis. *arXiv preprint arXiv:2301.12867*.