Task-based & Social Conversational Agents

Bilgehan Emiral
TECHNICAL UNIVERSITY OF MUNICH
bilgehan.emiral@tum.de

Abstract

This paper examines the evolution and current state of dialogue systems, focusing on task-oriented and social dialogue agents and the role of large language models (LLMs) as conversational agents. We explore the fundamental principles of human conversation, such as grounding, speech acts, and empathy, which are essential for developing natural and effective dialogue systems. The study discusses task-oriented dialogue systems, dialogue management techniques, and evaluation metrics, highlighting advancements and ongoing challenges. Social dialogue agents are investigated for their ability to engage in open-ended conversations and provide emotional support. The development of chatbots from early rule-based systems to advanced deep learning-powered AI, such as GPT-3, is also covered. Despite significant progress, challenges remain in managing privacy risks, addressing ethical concerns, and ensuring system robustness.

1 Introduction

Today, the increasing importance of human-machine interactions in the digital world is undergoing a significant transformation with the development of natural language processing (NLP) and artificial intelligence (AI) technologies. In this context, speech-based interfaces and dialogue systems have become one of the most effective ways to interact with users and provide services to them. Dialogue systems can be divided into two main categories: task-oriented dialogue systems and social dialogue systems.

Task-oriented dialogue systems are designed to help users complete specific tasks. These systems allow users to perform certain tasks such as booking flight tickets, ordering food, or contacting customer service. Task-oriented systems typically rely on structured dialogue flows and interactions aimed at achieving a specific goal.

On the other hand, social dialogue systems aim to establish more natural and human-like interactions with users. These systems are developed to chat with users, provide emotional support, or engage in conversations for entertainment purposes. Social dialogue systems are generally designed to capture users' interest and provide long-term interactions. Such systems can exhibit personality traits and possess emotional AI.

Chatbots combine features of both task-oriented and social dialogue systems. Modern chatbots are AI-based software that can answer users' questions, provide information, and perform specific tasks. One of the first chatbots, ELIZA, was developed in 1966 by Weizenbaum (1966) and could interact with users using simple pattern matching techniques. Later, ALICE (Artificial Linguistic Internet Computer Entity), developed in the 1990s, offered more sophisticated interactions using NLP rules (Wallace, 2009). Modern chatbots have significantly improved with the use of deep learning and NLP techniques. For example, GPT-3, developed by OpenAI (Brown et al., 2020), is a massive language model with millions of parameters and can produce highly fluent and meaningful texts. Chatbots are widely used in various fields such as customer service, healthcare, education, and entertainment.

Both types of dialogue systems use various methods and techniques to respond to different usage scenarios and user needs. While task-oriented systems generally focus on completing a specific task, social systems are more open-ended and interaction-oriented. This study aims to reveal the differences and similarities between these two approaches by examining the basic features, methods, and application areas of task-oriented and social dialogue systems. Additionally, it will discuss current developments and future research trends in this field by including important studies in the existing literature. In this context, the development

Feature	Task-Oriented Dialogue Systems	Social Dialogue Systems	LLMs as Dialogue Agents
Primary Goal	Complete specific tasks	Establish natural, human-like interactions	Both task-oriented and social interactions
Examples	Booking flights, ordering food	Chatting, emotional support	Personal assistants, customer support
Dialogue Management	Structured dialogue flows	Open-ended interactions	Advanced dialogue management using neural networks
Evaluation Metrics	Task success rate, slot error rate, user satisfaction	User satisfaction, interaction duration, emotional response accuracy	Task completion rate, response quality, user satisfaction
Key Techniques	Frame-based models, reinforcement learning	Emotional intelligence, personality traits	Instruction tuning, reinforcement learning
Datasets	MultiWOZ (Budzianowski et al., 2018)	EmpatheticDialogues (Rashkin et al., 2018)	Various large-scale datasets

Table 1: Comparison of Different Dialogue Systems and Models

of dialogue systems and the expansion of their usage areas are of great importance in understanding the evolution of human-machine interactions in the digital age.

2 Understanding Human Conversation

Human-machine interactions have made significant progress with the development of NLP and AI technologies to understand and model human conversation. Human conversation is a complex process that typically involves factors such as context, intent, and emotional state (Clark, 1996). In this context, speech-based AI systems have been developed to understand the fundamental features and dynamics of human conversation and use them in digital interactions.

Clark (1996), in his work "Using Language," states that conversation is a process of mutual understanding between two people. In this process, the speaker and listener continuously provide feedback to understand and confirm each other's expressions. This process, called "grounding," is critical for successful communication and is a fundamental principle that should be applied in dialogue systems. Grounding can occur in two ways: explicit grounding, where the listener directly confirms or repeats the provided information, and implicit grounding, where understanding is acknowledged indirectly without restating the content.

Explicit Grounding:	***User:*** *"I need to pick up my order at 5 PM."* ***System:*** *"Got it, you need to pick up your order at 5 PM."*
Implicit Grounding:	***User:*** *"I need to pick up my order at 5 PM."* ***System:*** *"Okay, see you then."*

Table 2: Examples for Explicit and Implicit Grounding

Bach and Harnish (1979), in their studies on speech act theory, argue that conversations are actions that serve specific purposes. Speech acts can be categorized into informing, directing, committing, and acknowledging. These categories can be used for dialogue systems to understand user intentions and produce appropriate responses. For example, when a user is arranging a meeting, the system needs to recognize and respond to these intentions correctly.

Constatives:	***User:*** *"The meeting is at 3 PM today."*
Directives:	***User:*** *"Could you send me the report by tomorrow?"*
Commissives:	***User:*** *"I will complete the project by the end of the week."*
Acknowledge.:	***User:*** *"Thank you for your help with the project."*

Table 3: Examples of Speech Acts

The GUS system (Bobrow et al., 1977), an important step in the development of dialogue systems, uses a frame-based approach to handle user requests in a structured manner. This system uses predefined frames to understand the user's intentions and context. The GUS system has formed the foundation of modern dialogue systems and pioneered the development of task-oriented dialogue systems.

Empathy is an important component of human conversation and is crucial for AI systems to establish meaningful interactions with humans. Cuff et al. (2016) emphasize that empathy comprises emotional and cognitive components. Emotional empathy involves feeling others' emotions, while cognitive empathy involves understanding these emotions. Empathetic dialogue systems have been developed to detect emotional cues in the user's language and provide supportive, validating responses that mirror and acknowledge those emotions (Shum et al., 2018).

To illustrate how empathetic dialogue systems work, consider the responses provided by Woebot, an AI-based chatbot designed to offer emotional support:

Sadness:	**User:** "I'm feeling really down today." **Woebot:** *"I'm sorry to hear that you're feeling this way. It's okay to feel sad sometimes. Do you want to talk about what's on your mind?"*
Anxiety:	**User:** "I'm so anxious about my upcoming exams." **Woebot:** *"Exams can be really stressful. Let's take a moment to breathe together. You got this!"*

Table 4: Examples of Woebot's Responses to Basic Emotions

In conclusion, the dynamics and characteristics of human conversation play a critical role in the development of speech-based AI systems. These systems should apply fundamental principles such as grounding, speech acts, and empathy to establish successful communication. Future research should focus on further developing these principles and making human-machine interactions more natural and meaningful.

3 Task-Based Dialogue Systems

Task-based dialogue systems play a crucial role in helping users accomplish specific tasks through natural language interactions. These systems are designed to assist users in various tasks such as booking flight tickets, ordering food, or providing customer support. The foundation of these systems relies on their ability to understand user intent, manage dialogue, and generate appropriate responses.

One of the pioneering works in this field is the GUS system developed by Bobrow et al. (1977). GUS presented a frame-based approach where each task is represented by a frame consisting of slots that need to be filled with information provided by the user. This approach enables the system to manage and track the dialogue state, ensuring all necessary information is collected. The frame-based model has become a standard in the design of task-based dialogue systems, offering a structured method for handling complex user interactions.

Dialogue management is a critical component of task-based systems. It involves maintaining the context of the conversation, monitoring the dialogue state, and determining the next action. Systems like the neural belief tracker (NBT), developed by Mrkšić et al. (2017), have advanced dialogue state tracking by using neural networks to predict and update the dialogue state based on user inputs. This method enhances the system's ability to understand and manage conversations, making interactions more efficient and accurate.

The training of task-based dialogue systems is typically carried out through supervised learning on labeled datasets. These datasets contain dialogue examples labeled with intents, slots, and dialogue states. Models learn to map user inputs to appropriate actions based on this training data. The MultiWOZ dataset provided by Budzianowski et al. (2018) is a large-scale, multi-domain dataset widely used in the training and evaluation of dialogue systems.

The evaluation of task-based dialogue systems is conducted using various metrics such as task success rate, slot error rate, and user satisfaction (Deriu et al., 2021). The task success rate measures how often the system completes the user's task. The slot error rate assesses the accuracy of slot filling, which is critical for task completion. User satisfaction surveys are frequently used to gather feedback on the system's performance and usability. The slot error rate (SER) assesses the accuracy of slot filling, which is critical for task completion. SER is calculated using the following formula:

Domain	Categorical slots	Non-categorical slots	Intents
Taxi	-	destination, departure, arriveby, leaveat, phone, type	book
Train	destination, departure, day, bookpeople	arriveby, leaveat, trainid, ref, price, duration	find, book
Bus	day	departure, destination, leaveat	find

Table 5: Example of Categorical and Non-categorical Slots and Intents Across Different Domains (MultiWOZ 2.2 dataset)

$$SER = \frac{N_{incorrect} + N_{missed} + N_{spurious}}{N_{total}} \quad (1)$$

Slot Error Rate (SER) formula, where $N_{spurious}$ is the number of extra slots.

Recent advancements in deep learning and NLP have further improved task-based dialogue systems (Mrkšić et al., 2017; Budzianowski et al., 2018).Techniques such as reinforcement learning and end-to-end learning have been applied to enhance dialogue management and response generation. These advancements have led to more sophisticated and capable dialogue systems that can handle a wider range of tasks with higher accuracy and efficiency. However, these systems face challenges in handling complex and diverse user interactions, which can limit their overall effectiveness.

In conclusion, task-based dialogue systems play a vital role in automating and facilitating various tasks through natural language interactions. The continuous development driven by innovative approaches and robust datasets in this field promises to make these systems more efficient and user-friendly. Future research will focus on integrating these systems more seamlessly into everyday applications, aiming to increase their benefits and enhance user experience.

4 Social Agents

Social agents are AI-based systems designed to establish natural and human-like interactions with people. These agents are created to chat with users, provide emotional support, or engage in conversations for entertainment purposes. The primary goal of social agents is to build long-term and meaningful relationships with users, meet their emotional needs, and enrich the user experience.

Shum et al. (2018) discussed the development and challenges of social chatbots, exploring how such systems can offer human-like interactions. Social agents exhibit personality traits and can possess emotional intelligence to attract users' attention and interact with them. These features allow users to have a more intimate and satisfying interaction with the agents.

Empathy, as previously discussed, enables social agents to effectively understand and respond to users' emotional states. This capability allows these agents to establish deeper and more meaningful connections with users, ultimately enhancing user satisfaction.

The datasets used in developing social agents are also of great importance. The "EmpatheticDialogues" dataset is widely used in training such agents. This dataset contains dialogue examples designed to respond to users' emotional expressions and helps social agents develop emotional intelligence capabilities (Rashkin et al., 2018).

The evaluation of social agents is done using metrics such as user satisfaction, interaction duration, and accuracy of emotional responses (Rashkin et al., 2018). These agents are assessed based on their ability to respond to users' emotional needs and the naturalness and fluency of their interactions. User satisfaction surveys and interaction analyses are common methods used to evaluate the performance of social agents.

In recent years, advancements in deep learning and NLP have significantly enhanced the capabilities of social agents (Shum et al., 2018). In particular, multi-modal learning and emotional response generation techniques have enabled social agents to produce more human-like and emotionally satisfying responses. These developments allow social agents to meet a wider range of user needs. Despite their advancements, social dialogue agents often struggle with maintaining consistent personality traits and handling sensitive topics ethically.

In conclusion, social agents are important AI systems that aim to establish more natural and meaningful interactions with humans. The development and evaluation of these systems are continuously progressing to increase user satisfaction and interaction quality. Future research will aim to further enhance the emotional intelligence capabilities of social agents, striving to create deeper and more satisfying interactions with users.

5 Chatbots

Chatbots are AI-based software that can interact with users in natural language, making them a crucial and revolutionary tool in modern technology. Their ability to simulate human-like conversations has transformed how businesses and services operate. The importance of chatbots lies in their widespread application across various sectors such as customer service, healthcare, education, and entertainment. By providing instant responses and automating tasks, chatbots have not only improved efficiency but also enhanced user experience, leading to significant changes in these industries. Whether answering users' questions, providing information, or performing specific tasks, chatbots have become indispensable tools in our increasingly digital world.

One of the first chatbots, ELIZA, was developed in 1966 by Weizenbaum (1966). ELIZA was a system that could interact with users using simple pattern-matching techniques. Later, in the 1990s, Richard Wallace developed ALICE (Artificial Linguistic Internet Computer Entity), which offered more sophisticated interactions using NLP rules. ALICE was able to engage in meaningful dialogues with users using a special language called AIML (AI Markup Language) (Wallace, 2003).

Modern chatbots have significantly improved with the use of deep learning and NLP techniques (Brown et al., 2020). For example, GPT-3, developed by OpenAI, is a massive language model with millions of parameters and can produce highly fluent and meaningful texts. GPT-3 uses extensive language knowledge and world knowledge to understand user inputs and generate appropriate responses (Brown et al., 2020).

The training of chatbots is typically carried out using supervised learning methods on large datasets. These datasets contain various user inputs and the responses given to them. Chatbots are trained on these datasets to gain the ability to understand users' intentions and provide appropriate responses. Datasets such as EmpatheticDialogues and MultiWOZ are widely used to develop chatbots' emotional intelligence and task completion capabilities (Rashkin et al., 2018; Budzianowski et al., 2018).

The evaluation of chatbots is done using various metrics such as user satisfaction, response quality, interaction duration, and task completion rate (Roller et al., 2021). User satisfaction measures how effectively and satisfactorily chatbots respond. Response quality focuses on the accuracy and appropriateness of the responses given by chatbots. Interaction duration shows how long users interact with chatbots, and the task completion rate measures how well chatbots help users complete specific tasks.

In conclusion, chatbots play an important role in various sectors, and their ability to establish natural and meaningful interactions with users is continuously improving. Advancements in deep learning and NLP are making chatbots smarter, more responsive, and user-friendly. However, modern chatbots still face issues related to privacy, security, and the ethical use of user data. Future research aims to further develop chatbots, optimizing them to meet a wider range of user needs and provide more satisfying interactions. social interactions.

6 LLMs as Agents

The quality of dialogue and reasoning in today's large language models (LLMs) enables research on simulating human agents. For example, Argyle et al. (2023) demonstrate that LLMs can mimic the responses of different test group demographics using appropriate prompts. To further this research, Liu et al. (2022) developed a dialogue user-simulator, a tool designed to generate synthetic user interactions with dialogue systems, which helps in training and evaluating LLMs in various conversational contexts. In another study, Liu et al. (2023) placed LLM agents equipped with memory into a grid world, encouraging them to discuss sensitive topics and aiding in the emergence of social norms. Additionally, Park et al. (2023a) were able to simulate social dynamics in a simulated village by having LLMs act with different personas and retrieve relevant memories (see Figure 1).

LLMs such as GPT-3 (Brown et al., 2020), has significantly advanced the capabilities of conversational agents. These models have demonstrated remarkable skill in producing human-like texts and have been used in various applications from personal assistants to customer support. The following literature review highlights recent developments

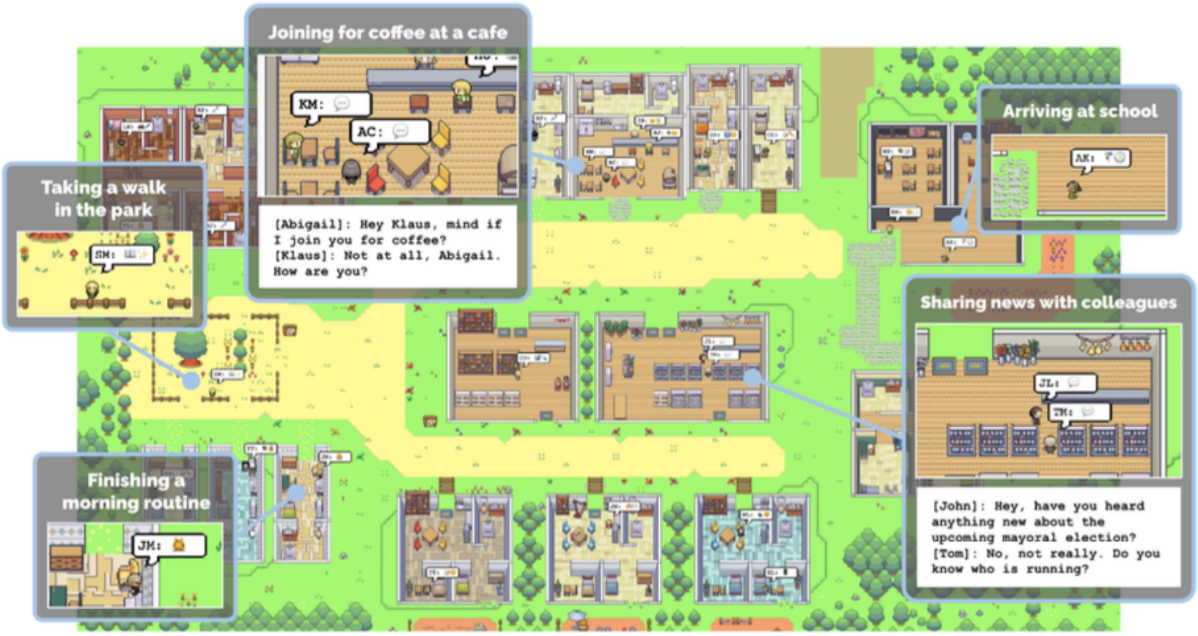

Figure 1: Generative Agents: Interactive Simulacra of Human Behavior (Park et al., 2023b).

and state-of-the-art techniques in using LLMs as agents.

LLMs have emerged with their ability to produce coherent and contextually appropriate texts. Their capacity to understand and generate human language has allowed them to be used as conversational agents in task-oriented and socially interactive conversation settings. These models are trained on large amounts of data and use complex neural network architectures based on the Transformer model introduced by Vaswani et al. (2017). This fundamental architecture has enabled LLMs to perform superiorly in sequential data processing and contextual understanding.

A significant development in customizing LLMs for specific tasks is the instruction tuning method, as demonstrated by Ouyang et al. (2022). This method involves training models on human-generated instructions and example responses, enhancing their ability to follow specific directives. However, this method can be resource-intensive as it requires extensive labeled data. To address this issue, Liu et al. (2022) and Argyle et al. (2023) have explored more efficient data collection methods using self-talk techniques where LLMs engage in conversations in various roles. This approach allows for the creation of high-quality training sets while reducing manual annotation effort.

Reinforcement learning (RL) has also been used to improve the performance of LLMs in conversational contexts. The self-play technique used in AlphaGo (Silver et al., 2016) has inspired similar methods in language models. Liu et al. (2023) placed LLM agents equipped with memory into simulated environments, encouraging them to discuss sensitive topics and helping to emerge social norms. This technique allows models to learn from their interactions and develop more natural and contextually aware dialogues.

Recent research has also explored the dynamics of multi-agent systems where multiple LLMs interact in a controlled environment. Park et al. (2023a) were able to simulate social dynamics in a virtual village where LLMs act with different personas and retrieve relevant memories. This approach not only increases the realism of interactions but also provides a rich dataset for training models on complex social behaviors.

To refine the training process, automatic metrics measuring the success of dialogues have been introduced (Deriu et al., 2021; Mehri and Eskenazi, 2020). These metrics are used to filter generated conversation data and ensure that only high-quality interactions are used for supervised fine-tuning. This iterative process increases the robustness and reliability of LLMs in real-world applications (Roller et al., 2021; Liu et al., 2021).

LLMs have been used as agents in various practical applications such as personal assistants and customer support. These applications leverage the models' ability to adapt to new tasks and user requirements quickly. However, challenges remain in the ethical use of these models, managing fine-tuning costs, and ensuring consistency of responses.

The integration of LLMs as conversational agents has enabled new user-centric applications using advanced techniques such as instruction tuning, self-talk, and reinforcement learning. These methods have significantly improved LLMs' abilities to perform specific tasks and establish meaningful social interactions. Nevertheless, LLMs can propagate biases present in training data and require significant resources for fine-tuning and adaptation. Future research will focus on optimizing these processes, addressing ethical issues, and expanding the applicability of LLMs in various conversational settings.

7 Discussion

This paper has explored the evolution and current state of dialogue systems, with a particular focus on task-oriented and social dialogue agents, and the emerging role of LLMs as conversational agents. By examining the fundamental principles of human conversation, such as grounding, speech acts, and empathy, we have highlighted the essential elements that contribute to the development of more natural and effective dialogue systems.

7.1 Challenges and Limitations

Despite the advancements in dialogue systems, several challenges and limitations need to be addressed to further improve their effectiveness and usability.

7.1.1 Task-Oriented Dialogue Systems

While task-oriented dialogue systems have shown significant advancements, their ability to adapt to diverse user needs and scenarios remains a challenge. These systems often rely on structured dialogue flows, which can limit their flexibility. Future research should focus on developing more adaptable architectures that can handle a wider range of tasks and user interactions.

The effectiveness of dialogue management techniques, such as the NBT, depends heavily on the quality and quantity of training data. Ensuring that these systems can maintain context and manage dialogue states accurately in real-time interactions is an ongoing challenge. Exploring unsupervised or semi-supervised learning methods could help mitigate the dependency on large labeled datasets.

Current evaluation metrics, such as task success rate and slot error rate, provide valuable insights but may not fully capture the user experience. Developing more comprehensive evaluation frameworks that consider user satisfaction, interaction naturalness, and long-term engagement will be essential for advancing task-oriented dialogue systems.

7.1.2 Social Dialogue Agents

Social dialogue agents aim to establish natural and human-like interactions by incorporating emotional intelligence. However, accurately detecting and responding to users' emotional states remains a complex task. Improving the emotional intelligence of these agents through advanced multi-modal learning and emotional response generation techniques is necessary.

Ensuring that social dialogue agents exhibit consistent personality traits and maintain coherent interactions over long periods is challenging. Future research should explore methods for better personality modeling and consistency maintenance in social interactions.

Social dialogue agents must handle sensitive topics with care and avoid causing harm to users. Addressing ethical concerns, such as bias, privacy, and user manipulation, is crucial. Developing guidelines and frameworks for ethical AI interactions will be essential.

7.1.3 Chatbots and LLMs

Handling sensitive information with LLMs in chatbots introduces significant privacy and security risks. Implementing privacy-preserving techniques, such as differential privacy and homomorphic encryption, can mitigate these risks, but they often come with trade-offs in performance and cost.

LLMs can inadvertently perpetuate biases present in the training data, leading to biased or inappropriate responses. Ensuring ethical use and addressing bias in LLMs is a critical challenge. Developing robust bias detection and mitigation techniques will be necessary to ensure fair and responsible AI interactions.

While LLMs have demonstrated remarkable capabilities, fine-tuning them for specific tasks and ensuring adaptability to new domains remains resource-intensive. Exploring more efficient fine-tuning methods and self-improvement techniques, such as iterative refinement and external knowledge integration, will be crucial for enhancing the performance and scalability of LLMs.

7.2 Future Work

The integration of self-improvement techniques into task-oriented dialogue systems represents a significant step forward. By leveraging methods such as iterative refinement and external knowledge integration, researchers can enhance the performance and adaptability of dialogue systems. Future research should focus on refining these techniques, exploring their application across different domains, and addressing the challenges related to ethical use and data privacy.

Developing more comprehensive evaluation frameworks that consider user satisfaction, interaction naturalness, and long-term engagement will be essential for advancing both task-oriented and social dialogue systems. Future research should aim to create standardized evaluation metrics that can be widely adopted across the field.

Addressing ethical concerns, such as bias, privacy, and user manipulation, is crucial for the responsible deployment of dialogue systems. Future research should focus on developing guidelines and frameworks for ethical AI interactions, ensuring that dialogue systems are transparent, fair, and aligned with user expectations.

Improving the emotional intelligence of social dialogue agents through advanced multi-modal learning and emotional response generation techniques is necessary. Future research should explore innovative approaches to enhance the emotional capabilities of these agents, enabling them to establish deeper and more meaningful interactions with users.

Developing more scalable and cost-effective dialogue systems that can be deployed in various real-world applications is a key research direction. Future work should focus on reducing the reliance on human intervention and manual fine-tuning, leveraging self-improvement techniques and efficient training methods to create more autonomous and self-sufficient systems.

In conclusion, while significant progress has been made in the field of dialogue systems, addressing the challenges and limitations discussed above will be crucial for future advancements. By focusing on these areas, researchers and practitioners can continue to enhance human-machine interactions, making them more natural, meaningful, and beneficial for users.

8 Conclusion

This paper has explored the evolution and current state of dialogue systems, focusing on task-oriented and social dialogue agents, and the transformative role of LLMs as conversational agents. By examining the fundamental principles of human conversation, such as grounding, speech acts, and empathy, we have highlighted the essential elements for developing more natural and effective dialogue systems.

Our study examined dialogue management techniques, and evaluation metrics of task-oriented systems, and investigated social dialogue agents, showcasing their ability to engage in open-ended conversations and provide emotional support.

The development of chatbots from early rule-based systems to advanced deep learning-powered AI, exemplified by models like GPT-3, has significantly enhanced automated interactions, making them indispensable in various sectors.

We explored cutting-edge applications of LLMs as dialogue agents, discussing advancements in instruction tuning, self-talk techniques, and reinforcement learning approaches. These innovations have enabled LLMs to perform specific tasks with greater accuracy and establish meaningful social interactions.

Despite these advancements, the field faces ongoing challenges, including managing privacy risks, addressing ethical concerns, and ensuring system robustness and reliability. The development of privacy-preserving techniques is crucial for mitigating these risks while balancing model utility and privacy protection.

Looking ahead, integrating self-improvement techniques into task-oriented dialogue systems represents a significant step forward. Leveraging methods such as iterative refinement and external knowledge integration can enhance the performance and adaptability of dialogue systems.

Ultimately, the field of dialogue systems is an ever-evolving landscape, with advancements in AI and NLP driving the creation of more sophisticated and human-like conversational agents.With the emergence of generative models, in recent years, we have witnessed the rise of the concept of chatbots replacing the clearly distinguished concepts of task-based and social conversation agents. In these years, the tendency to consider these two tasks as inseparable will be much higher. Future research should focus on refining these technologies, addressing ethical and privacy concerns, and expanding their applicability across diverse domains to improve human-machine interactions, making them more natural, meaningful, and beneficial for users.

References

Liam Argyle, Yohan Bang, and Julia Stoyanovich. 2023. Simulating human agents using llms: Methods and applications. *Journal of Artificial Intelligence Research*.

Kent Bach and Robert M Harnish. 1979. *Linguistic communication and speech acts*. MIT Press.

Daniel G Bobrow, Ronald M Kaplan, Martin Kay, Donald A Norman, Henry Thompson, and Terry Winograd. 1977. Gus, a frame-driven dialog system. *Artificial Intelligence*, 8(2):155–173.

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. *Advances in neural information processing systems*, 33:1877–1901.

Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. 2018. Multiwoz-a large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5016–5026.

Herbert H Clark. 1996. *Using language*. Cambridge University Press.

Benjamin MP Cuff, Sarah J Brown, Laura Taylor, and Douglas J Howat. 2016. Empathy: A review of the concept. *Emotion Review*, 8(2):144–153.

Jan Deriu, Aitor Rodrigo, Arantxa Otegi, Guillermo Echegoyen, Sophie Rosset, Eneko Agirre, and Mark Cieliebak. 2021. Survey on evaluation methods for dialogue systems. *Artificial Intelligence Review*, 54(1):755–810.

Liang Liu, Jian Tang, and Fei Wu. 2022. Dialogue user simulator: A new training framework for task-oriented dialogue systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Liang Liu, Jian Tang, and Fei Wu. 2023. Memory-augmented llm agents for simulating social interactions. *arXiv preprint arXiv:2303.12814*.

Yang Liu, Ryan Lowe, Iulian Serban, Michael Noseworthy, Laurent Charlin, and Joelle Pineau. 2021. A survey of dialogue evaluation methods. *arXiv preprint arXiv:1909.06031*.

Shikib Mehri and Maxine Eskenazi. 2020. Unsupervised evaluation of interactive dialog with DiScore. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 681–707.

Nikola Mrkšić, Diarmuid Ó Séaghdha, Tsung-Hsien Wen, Blaise Thomson, and Steve Young. 2017. Neural belief tracker: Data-driven dialogue state tracking. *arXiv preprint arXiv:1606.03777*.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. *arXiv preprint arXiv:2203.02155*.

Hyunjin Park, Kyungmin Lee, and Yoonjin Choi. 2023a. Simulating social dynamics in a virtual village using llms. *IEEE Transactions on Neural Networks and Learning Systems*.

Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023b. Generative agents: Interactive simulacra of human behavior. Stanford University and Google Research.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2018. Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381.

Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric Michael Smith, et al. 2021. Recipes for building an open-domain chatbot. *arXiv preprint arXiv:2004.13637*.

Heung-Yeung Shum, Xiao-dong He, and Di Li. 2018. From eliza to xiaoice: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, 19(1):10–26.

David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. 2016. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Richard S Wallace. 2003. The elements of aiml style. *Alice AI Foundation*.

Richard S Wallace. 2009. *The anatomy of A.L.I.C.E.* Springer.

Joseph Weizenbaum. 1966. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45.