IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 6 August 2022, accepted 18 August 2022, date of publication 23 August 2022, date of current version 8 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3201144

SURVEY

AI-Based Conversational Agents: A Scoping Review From Technologies to Future Directions

SHEETAL KUSAL[1], SHRUTI PATIL[1,2], JYOTI CHOUDRIE[3], KETAN KOTECHA[1,2], SASHIKALA MISHRA[1], AND AJITH ABRAHAM[4,5], (Senior Member, IEEE)

[1]Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, Maharashtra 412115, India
[2]Symbiosis Centre for Applied Artificial Intelligence (SCAAI), Symbiosis International (Deemed University), Pune, Maharashtra 412115, India
[3]Department of Information Systems, University of Hertfordshire, Hertfordshire, Hatfield AL10 9EU, U.K.
[4]Machine Intelligence Research Laboratories, Auburn, WA 98071, USA
[5]Center for Artificial Intelligence, Innopolis University, 420500 Innopolis, Russia

Corresponding author: Shruti Patil (shruti.patil@sitpune.edu.in)

This work was supported by the Analytical Center for Government of the Russian Federation under Grant 70-2021-00143 dd. 01.11.2021 and Grant IGK000000D730321P5Q0002.

ABSTRACT Artificial intelligence is changing the world, especially the interaction between machines and humans. Learning and interpreting natural languages and responding have paved the way for many technologies and applications. The amalgam of machine learning, deep learning, and natural language processing helped Conversational Artificial Intelligence (AI) to change the face of Human-Computer Interaction (HCI). A conversational agent is an excellent example of conversational AI, which imitates the natural language. This article presents a sweeping overview of conversational agents that includes different techniques such as pattern-based, machine learning, and deep learning used to implement conversational agents. It also discusses the panorama of different tasks in conversational agents. This study also focuses on how conversational agents can simulate human behavior by adding emotions, sentiments, and affect to the context. With the advancements in recent trends and the rise in deep learning models, the authors review the deep learning techniques and various publicly available datasets used in conversational agents. This article unearths the research gaps in conversational agents and gives insights into future directions.

INDEX TERMS Artificial intelligence, machine learning, natural language processing, affective computing, mood or core affect, sentiment analysis, emotion theory, emotion in human-computer interaction, emotional corpora, intelligent agents, semantics, syntax, feature extraction, text processing.

I. INTRODUCTION

Today everything we have in our society is the result of intelligence; therefore, supplementing our human intellect with artificial intelligence has the potential to help society thrive like never before - as long as we can make the technology helpful. Healthcare, manufacturing, customer services, e-commerce, education, media, from every facet, it has transformed human life. One of the important branches of artificial intelligence is conversational AI which makes machines capable of understanding, processing, and responding to humans in natural language. Conversational agents have remained the center of the AI revolution in the past few years, powered by Natural Language Processing (NLP) and Machine Learning (ML) technologies.

A conversational agent [1] is an Artificial Intelligence (AI) program that originated to imitate human conversations using spoken or written natural language over the Internet. Many alternative terms are used for conversational agents. Earlier, dialogue system, this term was popular. But nowadays, chatbots, smart bots, intelligent agents, intelligent virtual assistants/agents, interactive agents, digital assistants, and relational agents are used alternatively in research articles [1], [2]. Conversational agents are the practical implementation of AI technology in industries or businesses. Conversational agents can be seen being used in various applications

The associate editor coordinating the review of this manuscript and approving it for publication was Utku Kose.

executing plenty of interesting tasks. In businesses [2] for marketing and customer support, in healthcare [3] as a personal assistant, in education [4] as a personal tutor, and in entertainment [5] for assisting players in digital games. Over the course of a few years, conversational agents have been in demand due to their distinctive characteristics. Conversational agents have simple interfaces, are available 24/7, provide prompt responses, are omnichannel, and have the ability to engage in conversations like humans.

Similarly, conversational agents can do the equivalent work of hundreds of humans and thus save operational costs for organizations. Also, conversational agents can talk to people, other AI systems, and things on the Internet due to IoT capabilities [6]. Conversational agents have a considerable impact on the market and industries in terms of businesses and consumers. According to industry experts [7], by 2024, the conversational agents' market will grow to 14 billion dollars. As a result, more and more organizations are moving towards conversational agents for better customer satisfaction and retention.

Inputs to conversational agents can be delivered in a variety of methods, including gestures (visual cues), speech (spoken cues), and natural written language (linguistic cues). But natural written language has remained unfocused in the research. It has been explored that natural written language communication has distinctive traits that have aptness to convey emotions, mood, and tone of a person in communication. Earlier conversational agents based on natural written language communication were keyword-based or pattern-based, where a question from a user was matched with a set of answers in a database, and the answer was returned as a response to the user. Advancements in technology have bridged the communication gap between human and machine interaction. This research in the field of conversational agents has remained significant for years, yet conversations with conversational agents upraise issues that are beyond current technological limitations. Current conversational agents are lagging in meeting users' expectations in terms of not being able to hold the conversations for a longer time [8]. One of the major downsides is that they are less context-aware [9]. Most conversational agents use keyword-based/ pattern-based methods [1], responding with specific answers. Conversational agents cannot understand users' emotions and sentiments [9]. So they cannot understand the mood or tone of the user [9]. One more drawback of current conversational agents is that they converse only with language (text), whereas humans communicate with different modalities or senses [10]. Figure 1 shows an overview of the current state of conversational agents. As per ongoing trends in customer services in different domains such as businesses, conversational agents should simulate human conversational characteristics and behavior. Conversational agents must have human conversational abilities [11] like syntactically correct, empathetic, and knowledgeable, i.e., a conversational agent should be context-aware. To make conversational agents show characteristics of human conversations, conversational agents need to understand a user's feelings, context, and mood, generate intelligible and engaging responses in conversations and respond with personalization with sentimental and emotional analysis. All these characteristics of human conversations in conversational agents can be accomplished to some extent with the help of advanced natural language processing and machine learning systems. Implementation, flexibility in various application domains and capability to mimic natural conversation to some extent have been accomplished with the help of advanced natural language processing and machine learning systems. But machine learning approaches have drawbacks like learning from human-labeled data that is time-consuming and dependent on human efficiency.

To summarize, this paper makes a number of significant contributions as follows:

1) Illustrate the basic working architecture of recent conversational agents.

2) Literature review of implementation methods used in different components of conversational agents.

3) It focuses on surveying the current practices of deep learning architectures in conversational agents.

4) Brief overview of datasets utilized in conversational agents.

5) Present identified research gaps and highlight future directions.

The rest of this article is organized as follows. Section 2 provides background knowledge in terms of the working of conversational agents. Section 3 discusses the literature review. Section 4 describes the comparative study of different deep learning techniques used in conversational agents and a brief overview of major datasets used in conversational agents' research works. The research gaps with future directions are discussed in Section 5. And paper is concluded in section 6.

II. BACKGROUND - WORKING OF CONVERSATIONAL AGENTS

Conversational agents appear to work simply at first glance when a user interacts with them and receives a suitable response. However, various technologies are at work behind the scenes to ensure smooth interaction. Natural Language Understanding (NLU unit) and Natural language Generation (NLG unit) are the major components of conversational agent architecture [12]. Figure 2 illustrates the architecture of conversational agents.

A. NATURAL LANGUAGE UNDERSTANDING

Natural Language Understanding is the key component where natural language processing and understanding user requests are done [13]. User query/message will be provided to the natural language processing unit as input. This unit's job is to prepare and clean the input text data, which includes text preprocessing steps [14]. These steps are important to interpret grammar and break down an input request into words and sentences, making it easier for a conversational agent to understand. Then cleaned input text is converted into feature

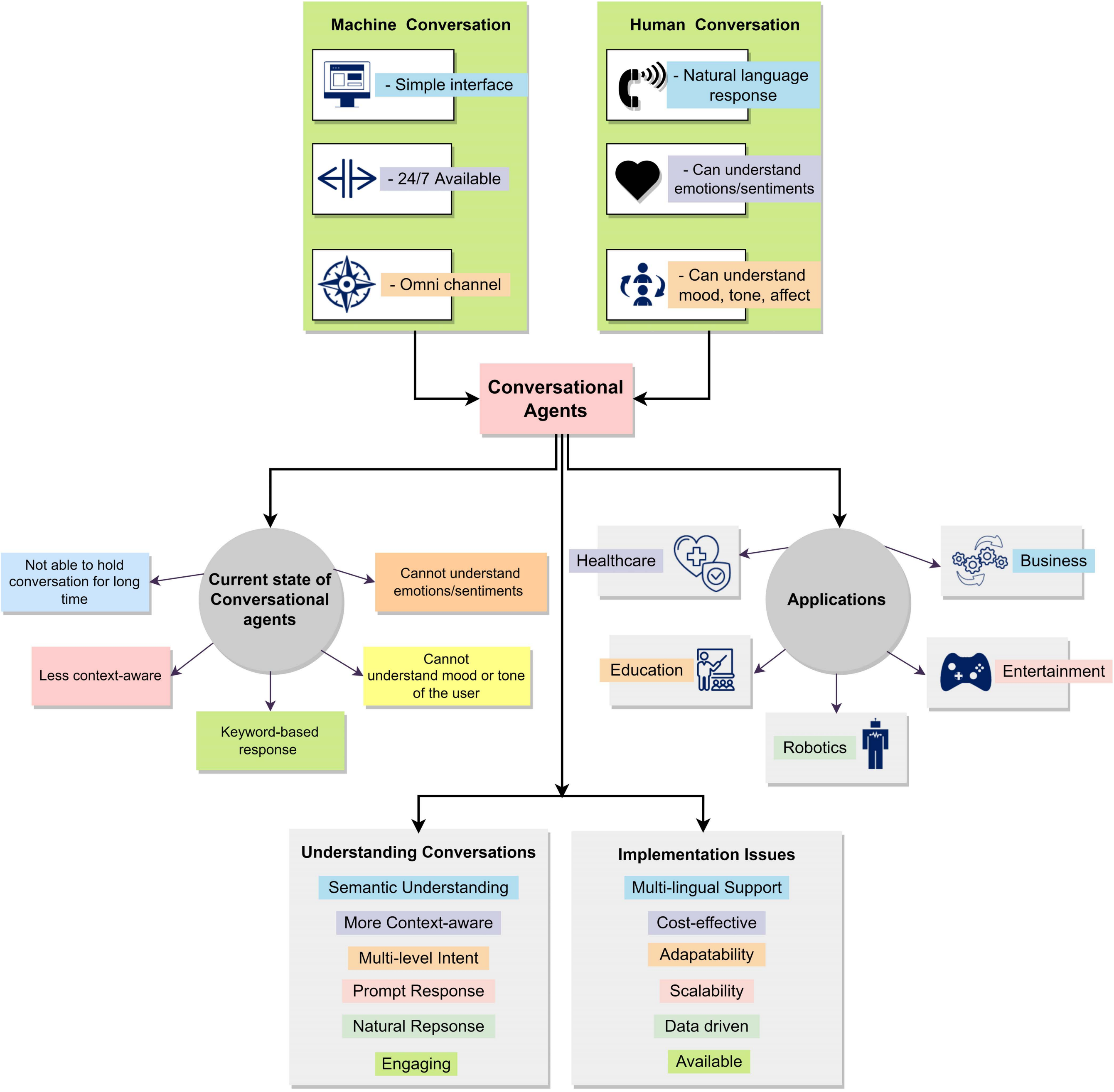

FIGURE 1. Overview of the current state of conversational agents.

vectors or word embeddings. Each word is represented by an N-dimensional integer. Natural language understanding transforms an unstructured input text from the user to produce a semantic representation by extracting the intent, entities, and cognitive information as shown in figure 2. In Intent identification, the task of correctly identifying the intent or purpose behind the user request is done. A supervised intent classification model can be trained on a variety of sentences as input and intents as a target in intent identification. So, the outcome of this task will be intent. The entity recognition task. identifies and separates discrete pieces of information into different pre-determined groups such as people, organizations, etc., from input text. The outcome of this task is recognized entities. In cognitive understanding, certain subtasks are performed, such as sentiment analysis, emotion detection, and checking spellings. Cognitive understanding will help conversational agents analyze the user's sense, tone, or mood of the input text to improve response generation accuracy. The outcome of NLU will be the semantic representation of context information by combining intent, entities, and cognitive information into a structured input. This context information of the user message will be the current state

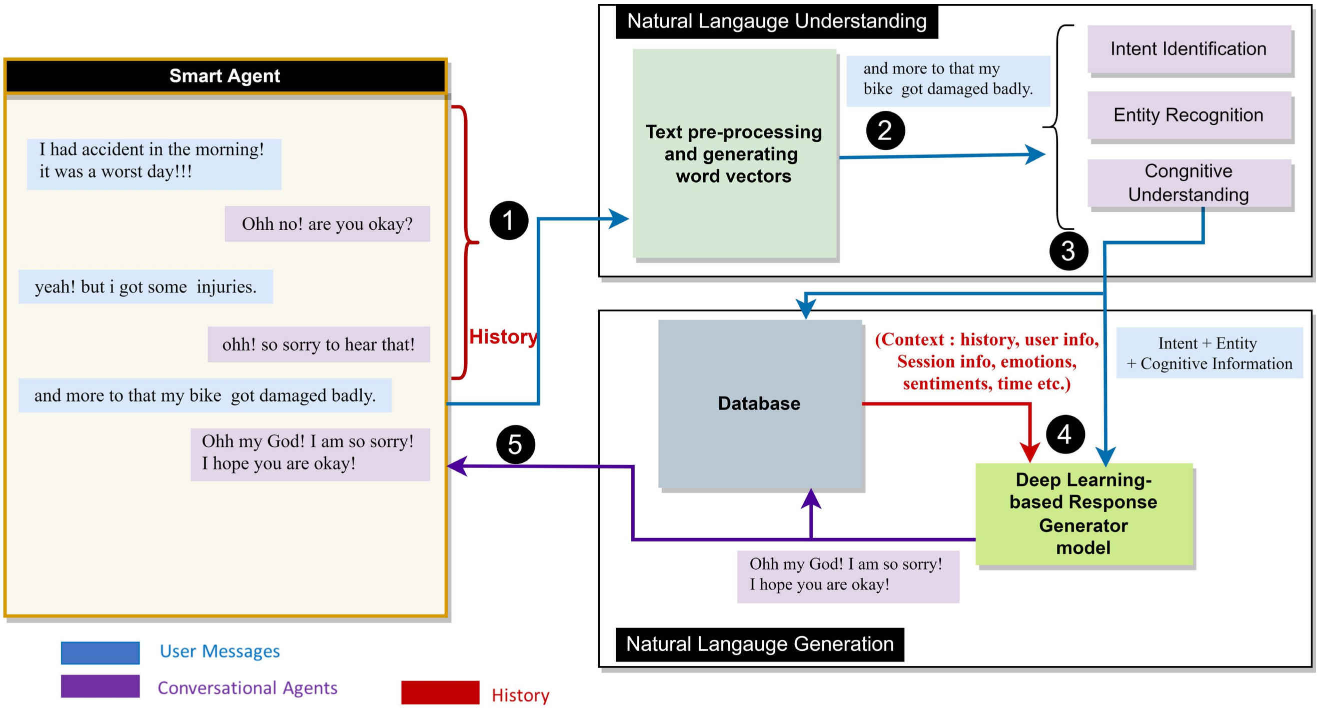
Smart Agent
I had accident in the morning! it was a worst day!!!
Ohh no! are you okay?
yeah! but i got some injuries.
ohh! so sorry to hear that!
and more to that my bike got damaged badly.
Ohh my God! I am so sorry! I hope you are okay!
History
1
Natural Langauge Understanding
Text pre-processing and generating word vectors
and more to that my bike got damaged badly.
2
Intent Identification
Entity Recognition
Congnitive Understanding
3
Database
(Context : history, user info, Session info, emotions, sentiments, time etc.)
Intent + Entity + Cognitive Information
4
Deep Learning-based Response Generator model
Ohh my God! I am so sorry! I hope you are okay!
5
Natural Langauge Generation
User Messages
Conversational Agents
History

FIGURE 2. **The architecture of conversational agents.**

of the conversation and will be provided to the response generation unit and stored in the database for future reference. Fig. 2 shows the working of conversational agents.

B. NATURAL LANGUAGE GENERATION

NLG receives the semantic representation in the form of context from the natural language understanding unit and generates a matching textual response [13]. The goal of the Natural language generation unit is to produce natural language sentences given a semantic. Natural language generation handles the actual context of user conversation. After understanding user requests, it must decide on its own set of actions to effectively continue the conversation. However, it is possible that the agent does not have all of the knowledge necessary to make decisions on the next steps. Based on this, Natural language generation maintains the states of previous conversations in the database, such as history, session information, user information, etc. And depending upon the state, it decides the next action. This unit helps pass the results and current state to the user in an understandable format. Natural language generation converts structured data into user-understandable representation. In language generation, retrieval-based and generative-based methods are used. Using specified templates, a retrieval-based approach maps a non-linguistic structured input question straight to natural language representation. Other hand, generative methods can generate new dialogues based on large amounts of conversational training data. Then this generated reply from the response generator will be stored again in the database as history and returned as a reply to the user.

III. RELATED WORK - LITERATURE REVIEW

On the topic of conversational agents' inconsiderable literature reviews have been written. Table 1 presents an overview of different survey papers in conversational agents. The presented surveys lack thorough analyses of datasets, affective components, multimodality studies, and performance metrics. In the field of conversational agents, different methods and techniques have been used for implementation. This study aims to examine various approaches and methods in conversational agents that can be used as a foundation for future empirical research. We've concentrated on crucial research areas, such as technical obstacles, datasets, the methodologies proposed in each study, and their performance metrics and application fields. Advanced deep learning models (pre-trained models) have been increasingly used in conversational agents. In this section, the key findings of conversational agents and related literature work are discussed from different perspectives, such as approaches used in conversational agents for the implementation of different tasks, the current trend of making empathic and context-aware conversational agents, deep learning approaches used in conversational agents, multimodality, datasets and application areas are given in this section.

A. APPROACHES USED IN CONVERSATIONAL AGENTS FOR THE IMPLEMENTATION OF DIFFERENT TASKS

In conversational agents, different tasks are performed to understand the user input and generate a response according to that input. Figure 3 shows an overview of techniques used in different tasks of conversational agents. Different

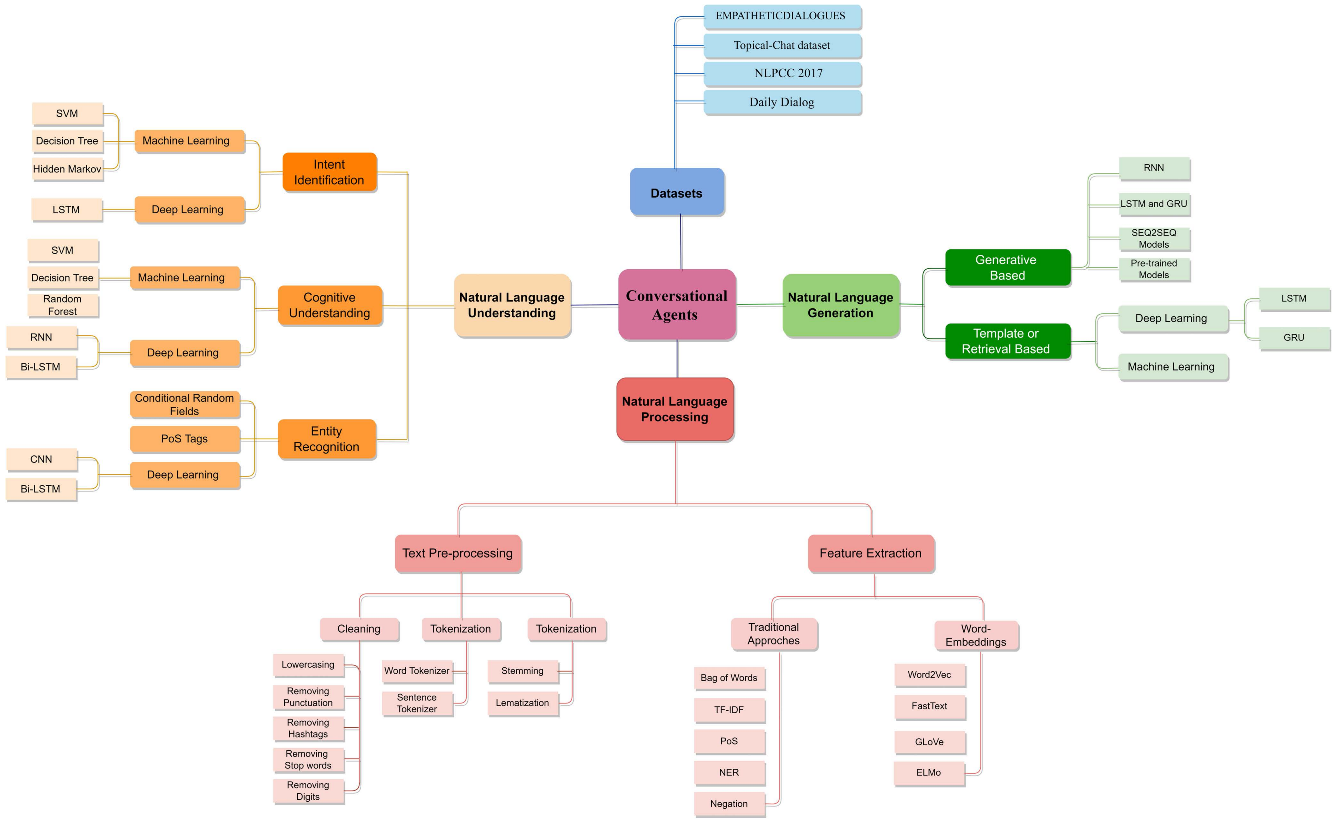

FIGURE 3. Overview of techniques used in different tasks of conversational agents.

methodologies are used in different components of conversational agents. It comprises two major components, each of which is subdivided into basic preliminary elements. Components of a conversational agent include Natural Language Understanding (NLU) and Natural Language Generation (NLG). Initially, text pre-processing and feature extraction steps are considered natural language processing (NLP).

1) NATURAL LANGUAGE PROCESSING (NLP)

When a user enters the text or query, the first step of a conversational agent is to prepare the data in the appropriate form to be passed to the natural language understanding unit. User input may have emojis, short text, informal words, incomplete words, etc., that make pre-processing a prerequisite. As well, NLU and NLG units employ machine learning or deep learning algorithms to perform the different tasks. These algorithms utilize statistical data to execute any sort of regression or classification task. Therefore, text pre-processing and feature extraction steps become important in NLP applications. Text pre-processing includes cleaning, tokenization and normalization.

- Data cleaning – It includes converting text to lowercase, removing punctuation marks from text, removing digits, stop words, hashtags & HTML tags from text etc.
- Tokenization/Segmentation – Tokenization separates sentences, words, and characters. Basically, it includes splitting text/strings into tokens representing words. Different methods are used for tokenization: whitespace tokenizer, word tokenizer, sentence tokenizer, etc.
- Normalization - Stemming and lemmatization comprise trimming a word to its origin form.

Figure 4 shows the different steps and tools used for text pre-processing.

2) NATURAL LANGUAGE UNDERSTANDING (NLU)

This component has three language comprehension tasks: intent classification, entity identification, and cognitive understanding. Intent classification comprehends the purpose of the input. Entity identification finds the distinct pieces of information. So, entities combined with the intent, allow the agent to fully understand the user's input. Conversational agents must comprehend the user intent and perform the required actions. Intent classification is to understand the why of the input [15], and entity identification is to understand the what of the input [16]. Cognitive understanding has become a significant step in conversational agents to understand the input and understand the user.

a: ENTITY IDENTIFICATION

The name-Entity Recognition (NER) task identifies and separates the named entities of input sentences into different pre-determined classes. Earlier regular expressions have been explored for named entity recognition. However, CRFs

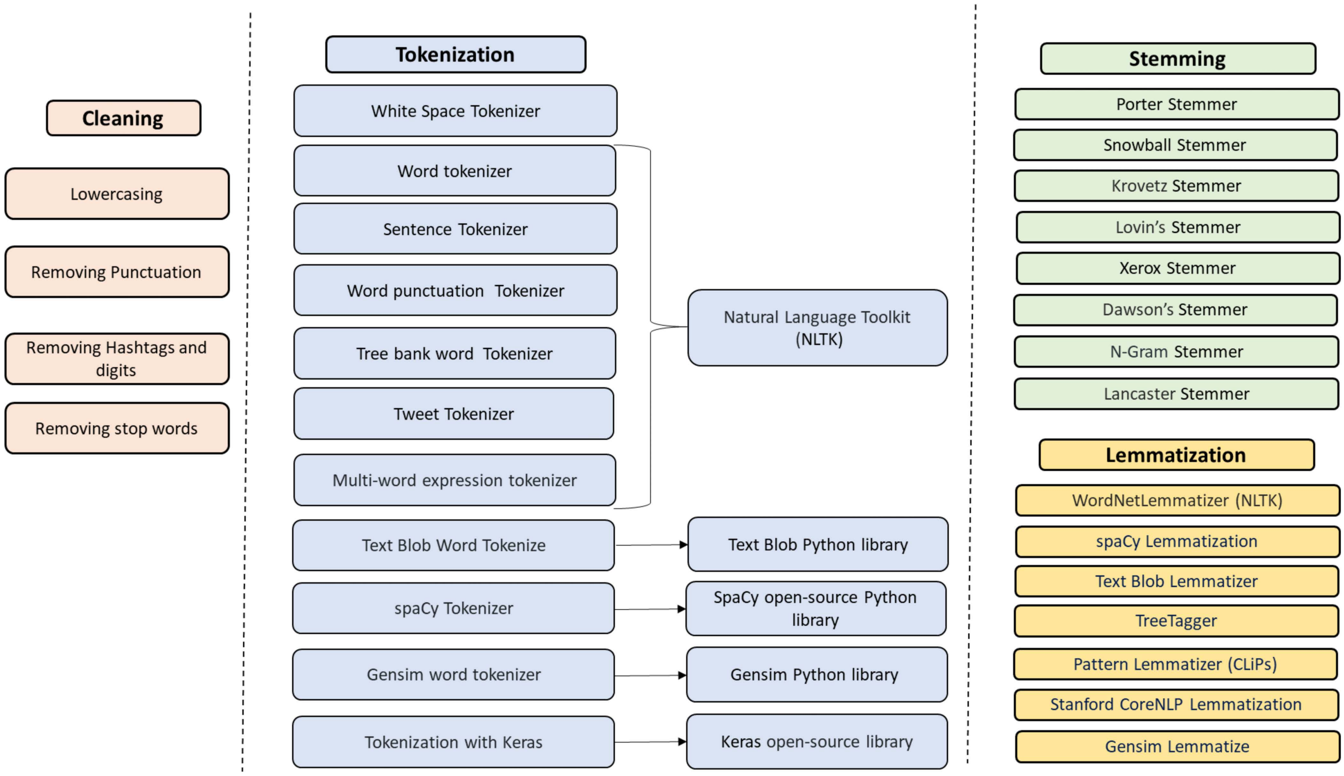

FIGURE 4. Different steps and tools used for text pre-processing.

(Conditional Random Fields) are frequently employed in NER and found in many applications [17]. It needs a tedious feature extraction, making this approach limited in adaptability and less scalable. NER has been implemented using Convolutional Neural Networks (CNNs), which require substantially less data pre-processing [18]. [19] presented NER based on an artificial neural network in the conversational agent. [20] suggest a bi-directional neural network (Bi-LSTM) to find long-term dependencies and generate a feature vector representation; and CNN and Bi-LSTM to predict the entities.

b: INTENT CLASSIFICATION

Intent classification comprehends the real purpose of input from the user. The conversational agent can use intent classification to figure out the desired objective or the user's goal. Different techniques have been used for the intent classification. Deep learning and machine learning approaches are the most notable methods that have been applied for intent classification. Conventional intent classification methods have primarily employed supervised machine learning algorithms such as SVM [21], Decision Trees[22], and Hidden Markov Models (HMM) [23].

As deep learning techniques become popular, neural network models are also used for intent classification. [23] suggested two deep hierarchical LSTM models distinguish dialog intents. Authors in [28]used a deep ensemble model using CNN, RNN like LSTM, and GRU to detect the intents from spoken language.

c: COGNITIVE UNDERSTANDING

Certain subtasks such as sentiment analysis, emotion detection, and spell checker are performed in cognitive understanding. Cognitive understanding will help conversational agents analyze the user's sense, tone, or mood of the input text to improve response generation accuracy. To make emotionally aware or empathic context-aware conversational agents, we need to add tasks such as emotion analysis or sentiment analysis in the NLU component. Emotion [29] is an essential aspect of making a context-aware conversational agent. Emotion detection is aimed to extract and study fine-grained emotions from text, such as anger, happiness, sadness, etc.

The various methods are utilized for analyzing text-based emotions; deep learning-based, machine learning-based, rule-based, and keyword-based approaches. Finding the frequency of a keyword in user input and comparing the labels with the dataset is a keyword-based method. [30] used keyword-based emotion recognition approach to finding the context from text. In the rule-based methods, grammatical and logical rules are decided to detect emotions from the user text. Reference [31] defined a rule-based emotion detection system to detect implicit emotions from text data. Machine learning-based approaches enable algorithms to learn and improve automatically through experience. Machine Learning methods categorize the user text into different predetermined emotion categories. Reference [32] proposed a machine learning-based system to classify conversational emotions. In artificial intelligence, deep learning is a subset

TABLE 1. Summary of existing surveys related to conversational agents.

Ref.	Datasets/ Resources	Affect components/ Resources	Multimodality	Techniques	Applications	Challenges	Future Directions	Performance Measures	Overview
[1]	×	×	×	√	√	√	×	×	This study covered conversational agents' literature along with its history, technology, and applications.
[2]	×	×	×	√	√	√	√	×	This work especially presents a literature review of conversational agents in the business domain.
[3]	×	×	√	√	√	√	√	×	This paper surveys conversational agents in healthcare from different perspectives, such as taxonomy, types of dialogue and context in healthcare.
[24]	√	×	×	√	×	×	×	√	This work presents a comprehensive analysis of conversational agents that can recognize emotions with techniques, datasets, performance measures.
[25]	×	×	×	√	√	×	√	×	This paper focuses on the implementation techniques of conversational agents with future scope and applications.
[26]	√	×	×	√	√	√	×	√	This paper reviews advancements in conversational agents with datasets, methods, performance measures and limitations.
[27]	√	√	×	√	√	√	×	√	This work comprehensively studied the concepts and building blocks of a conversational agent with datasets, application domains and performance measures.
Our Work	√	√	√	√	√	√	√	×	The use of affect components in conversational agents, multimodality, challenges, and future directions have been explored in our work with datasets, techniques, applications.

of machine learning. To learn from unstructured or unlabelled data, deep learning methods are used. These neural networks are capable of unsupervised learning. Reference [33] presented a deep learning-based system to detect emotions in human chatbot conversation. Similarly, a spell checker needs to be added to this component. As the ''cleaned'' input usually improves intent identification, a spell checker tries to correct the user's spelling problems. N-gram language models, techniques based on finding the frequency of words, distance-based methods and probability-based methods are used for spellings correction.

d: NATURAL LANGUAGE GENERATION (NLG)

This component focuses on natural language response generation methods. The natural language generation component receives input in the form of the current context and conversation history from the natural language understanding component in a well-defined format. As a result, the natural language generation's output is a sentence or text in natural language, which is also the final output to the user. Following are some of the traditional and advanced methods of response generation in human language. Retrieval-based or template-based systems map a user input straight to a natural language format by employing pre-determined templates [34]. Another natural language generation method is generative-based. The authors used deep learning and analyzed the use of recurrent neural networks for the task of natural language generation [35]. Seq2Seq models use LSTM to map a user input sequence to a feature vector representation and later in sequence predict tokens using pre-obtained feature representations. Sequence-To-Sequence (Seq2Seq) provides new and advanced performance in language generation tasks [36]. [37] proposed a system by combining seq2seq models with the power of reinforcement learning in natural language generation for text summarization application.

B. MAKING CONVERSATIONAL AGENT CONTEXT-AWARE

Before understanding the term context-aware, we will take a glance at what is the context? Context is a cause of an event. The situation within which something exists or happens and can help to explain it. It is circumstances forming a background of an event or a statement. So, context-aware is the ability of a system to perceive the user's environment or situation to reason appropriately. Context-awareness gives a system to see at the same level as a human and helps figure out in which sense the user is asking a question to revert to those sentiments and behavior. Conversational agents do not have automatic knowledge of their own, so they cannot use the context like humans. So, it is necessary to provide or feed them with the right information in context so they can use context on their own. So different kinds of information can be provided to conversational agents to understand the context.

Following are the different ways of providing information as context:

- Linguistic context – It denotes a context that address the relationship between words, phrases, sentences, and even paragraphs. It helps conversational agents to understand different meanings of the words as they are being used, who are used, and where they are used. It includes polysemy, word sense disambiguation, negation, intensifier, etc.
- Physical Context – It is also called situational context. It gives information about a situation in which an utterance or statement occurs, like a place, time, speaker, actions performed, objects involved, etc. This non-linguistic information helps to interpret or understand the meaning of a word and choose a correct equivalent to support context.
- Persistence context - A conversation is a chain of statements. Conversational agents need to keep track of conversations to predict the appropriate response. So, it stores the persisted context in the form of user information from the current turn, information in previous questions, and actions taken by conversational agents. This persistent context adds usability and ensures that humans and conversational agents have the same mental model. This can be regarded as a history of the conversation.
- Emotional Context - Emotions can be conveyed through words, combinations of different words, through emojis. This emotional context helps conversational agents understand emotions and moods and gain a deeper understanding of situations and the user's state of mind to respond empathically, effectively, and in the right way.

Conversation history, task records, user data, session information, emails, browsing information, location, sentiment, emotion, expression, time, etc., can be provided to conversational agents to understand the context. The context helps maintain the state of conversation and keeps the conversation flowing between human and conversational agents. Context makes Conversational agents intelligent by making them to understand the situation, emotions, and tone and able to interact and explain themselves and bring closer to natural human conversation. Summary of components that can be used while making conversational agents be context-aware given in table 2. Affect, emotions, tone, and sentiments these components can be utilized to make conversational agents more empathetic so that they can have a deeper understanding of situations and the user's state of mind, i.e., to understand the context of the situation and to respond empathically, effectively, and in the right way.

C. DEEP LEARNING APPROACHES USED IN CONVERSATIONAL AGENTS

Conversational agents primarily used to communicate with humans through text messages have been utilizing natural language processing techniques. Initially, Keyword-based or pattern-based conversational agents were implemented. These were easy to design and implement but had limitations in responding to complex queries. These were designed to answer based on patterns or rules, but if a query comes out of a pattern, it will provide erroneous answers that would not be related to the query. Machine learning-based conversational agents trained on existing annotated datasets of conversations. Usually, these retrieval-based machine learning models retrieve the information from the database based on a user query. But the main drawback of this approach is generating a large volume of knowledge base, which can be time-consuming, costly, involves human efficiency and, again domain dependency.

Figure 5 represents a schematic of the different components in conversational agents and the different techniques used in each component. Deep learning has taken center stage in many different application areas in recent years. Various

TABLE 2. Summary of components used to make conversational agents empathetic.

Component	Description	The component being used for	Ref.
Affect	• Relates to or arises from, or influences emotions • Includes following constructs such as emotions, sentiments, tone/mood	• To understand users' emotional states and to generate an emotional response	[38]
Emotion	• A feeling such as happiness, love, fear, anger, or sadness	• Contribute to more positive interaction • To improve user satisfaction • Reduce miscommunica tions	[39]
Tone	• The mood or feeling associated with a particular experience or stimulus	• To improve satisfactory services • Reduces user stress • More engagement	[40]
Sentiment	• A view or opinion that is expressed about	• Recognize user feelings regarding something • To understand the state of mind • Emotionally respond to the user	[41]

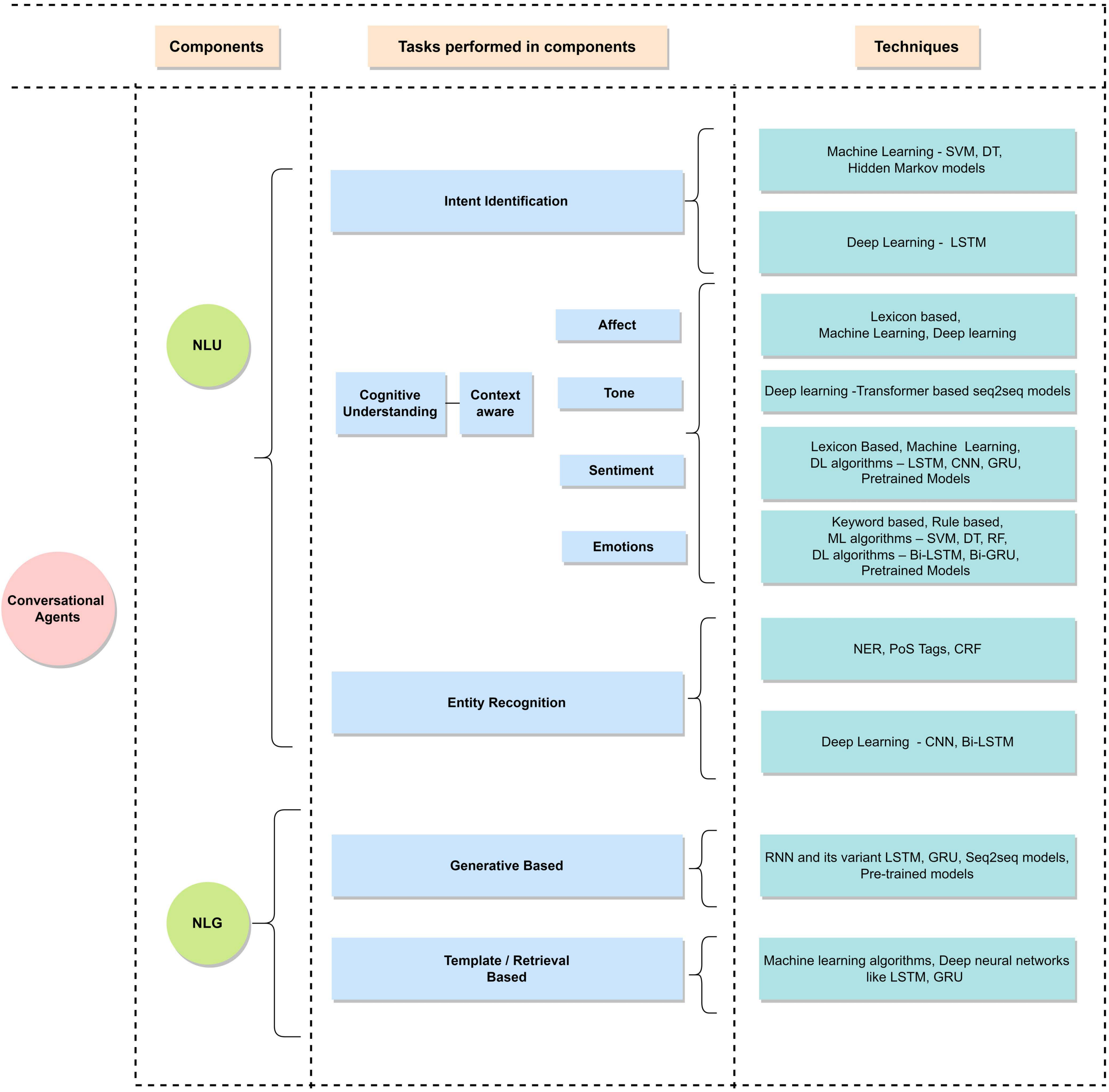
Components
Tasks performed in components
Techniques
Conversational Agents
NLU
NLG
Intent Identification
Cognitive Understanding
Context aware
Affect
Tone
Sentiment
Emotions
Entity Recognition
Generative Based
Template / Retrieval Based
Machine Learning - SVM, DT, Hidden Markov models
Deep Learning - LSTM
Lexicon based, Machine Learning, Deep learning
Deep learning -Transformer based seq2seq models
Lexicon Based, Machine Learning, DL algorithms – LSTM, CNN, GRU, Pretrained Models
Keyword based, Rule based, ML algorithms – SVM, DT, RF, DL algorithms – Bi-LSTM, Bi-GRU, Pretrained Models
NER, PoS Tags, CRF
Deep Learning - CNN, Bi-LSTM
RNN and its variant LSTM, GRU, Seq2seq models, Pre-trained models
Machine learning algorithms, Deep neural networks like LSTM, GRU

FIGURE 5. Schematic of the different components in conversational agents and different techniques used in each component.

articles have presented deep learning techniques in the field of conversational agents. Unsupervised learning encompasses deep learning techniques. It can, however, be semi-supervised or supervised. Deep learning classifiers automatically learn and extract information, improving accuracy and performance. In conversational agents, generative methods are based on deep learning techniques. These methods generate the word-by-word response to user queries based on syntax, structure and words (Vocabulary) in the input by understanding the context. Convolutional Neural networks (CNN), Recurrent Neural networks (RNN), Bi-LSTM, GRU, and pre-trained models like BERT, RoBERTa, and GPT are some most preferred deep learning models in different tasks of conversational agents. Table 3 shows that transformer-based architectures built most of the current conversational agents. These models are designed to increase the likelihood of a response and are capable of understanding a large amount of data to provide an acceptable response. The basic transformer design is made up of two recurrent neural networks (RNNs), one that processes the input is the encoder, and the other that generates the response is the decoder. These models are popularly known as Sequence-to-sequence models. The most prominent RNN variants utilized to learn the conversational dataset in these models are long short-term memory (LSTM)

TABLE 3. Overview of deep learning techniques used in conversational agents.

Technique	Variant of technique	Advantages of Technique	Disadvantages of Technique	Dataset utilized	Challenges identified in research	References
Recurrent Neural Network	LSTM (Long Short-Term Memory)	1) LSTM's internal gates control the information 2) It is good at processing long sequences 3) Keep relevant information to make predictions 4) addressing the vanishing gradient problem	1) LSTMs are prone to overfitting. and it is difficult to apply a dropout algorithm 2) require a lot of resources like memory and time to get trained	• Semeval-2019 • Raw Slack channel data • yelp challenge recommendation dataset • Short posts in Web forums and Wikis • Chinese Valence-Arousal Words (3.0) dataset • Multiple emotion and intensity aware Multi-party Dialogue (MEIMD) • NLPCC 2017 Shared Task Sample Data	• Data distribution is quite imbalanced • Accuracy isn't up to standard with state-of-the-art systems • Need to address intent detection • Incapable of generating responses of some specific labels due to less data of that category	[47] [48] [49] [50] [51] [52]
	Bi-LSTM	1) Process input in a forward and backward direction. 2) Access the past and future context of each sequence 3) Greater performance	1) Two LSTMs are used, so computationally costly and needs a lot resource allocation 2) Much slower model and requires more time for training			
	GRU (Gated Recurrent Unit)	1) Uses less training parameters, so use less memory, execute faster thus trains faster 2) Simpler and less redundant 3) GRU guarantees superior performance compared to LSTM	1) Less accuracy on long sequences as compared to LSTM 2) Low learning efficiency and prediction accuracy	• Topical Chat dataset • TV series dataset • STC conversation dataset • NLPCC 2017	• Due to the scarcity of labels in the dataset, models may have been trained to detect the most common words • Emotional inconsistencies while producing responses for some emotion classes • System performance can be improved • Need to understand all conversations in the physical world • To decide the emotion class, topics, contexts, or the user's mood are not considered • A model cannot replicate the empathy factor in human communication	[53] [54] [55] [56] [57]
	Bi-GRU	1) Process input in forward and backward direction 2) Access the past and future context of each sequence	1) Two GRUs are used so computationally costly and needs a lot resource allocation 2) Much slower model and requires more time for training			
Pretrained models	BERT	1) ability to handle contextual information 2) Faster Training	1) limited to monolingual classification 2) Fixed length of input sentences 3) Suffers from logical inference 4) computationally expensive.	• SEMEVAL-2019 • Large-scale multi-turn dialog opensubtitles2018 corpus	• Imbalanced data distribution • The accuracy of the response predictor can be improved	[33] [58]
	RoBERTa	1) The use of more extensive pre-training data results in improved performance 2) In downstream NLP tasks, outperforms XLNet and BERT	1) resource-intensive nature 2) It is computationally intensive and takes longer to complete			

or gated recurrent unit (GRU). Variations of LSTM and GRU as Bi-LSTM and Bi-GRU have also become popular due to their ability to process data in two directions. These variations can process the input in forward and backward directions. Attention mechanism was also introduced in the research of conversational agents. At each decoding phase, the attention mechanism allows the decoder to focus exclusively on the most significant input bits. It uses an attention weight to measure the importance of each word in the input text. Different attention mechanisms like self-attention [42], multi head-attention [43], word-level attention [44], hierarchical attention [45], are used in various research papers in conversational agents.

Deep learning methods such as RNN and its different variations has own limits. These models encode input into fixed-length vectors, so inputs with long sequences tend to lose important information. It leads to the poor performance of the response generation in conversational agents. Transformer-based language models such as BERT and its variations, GPT, and transformer XL [46] utilize sentence-level recurrence to overcome fixed-length limitations and longer-term dependency. Table 3 shows the overview of deep learning techniques used in conversational agents based on application areas, techniques used, datasets used, challenges identified, and performance achieved. These deep learning-based approaches have a quite significant upside. It is an end-to-end solution that can be trained using multiple datasets, and domain dependency is automatically handled.

D. MULTIMODALITY IN TEXT-BASED CONVERSATIONAL AGENTS

Due to recent commercial applications like Amazon's Alexa, Apple's Siri, Microsoft's Cortana, and Google Assistant, conversational systems have recently witnessed a considerable increase in demand. As more and more businesses are pushing for this technology, conversational agents are rapidly becoming commonplace. Humans communicate with one another through a variety of senses or modalities. These modalities work in concert to clarify concepts and emphasize ideas in dialogue by resolving ambiguity.

Nowadays, emoticons (objects encoded by standard sequences of characters) or emojis (e.g., smilies, hearts) are self-reported labels, i.e., visual information, provided by the users to convey emotions in their textual interactions the underlying the context of the communication, aiming for better interpretability, especially for short polysemous phrases. In conversational agents, this visual information conveys affective states and thus are suitable indications of sentiment and emotion in texts. These emojis/emoticons, along with the text, present a more faithful representation of the user's emotional state. In sarcastic sentences, a user may express positive emotion in the text, but by using emoticons/emojis, he/she may express negative emotions, so in such scenarios, visual information can help us to identify the true emotions of the user. This visual information in the form of emojis/emoticons helps in natural language understanding. E.g., I am happy with the 😡 service!!

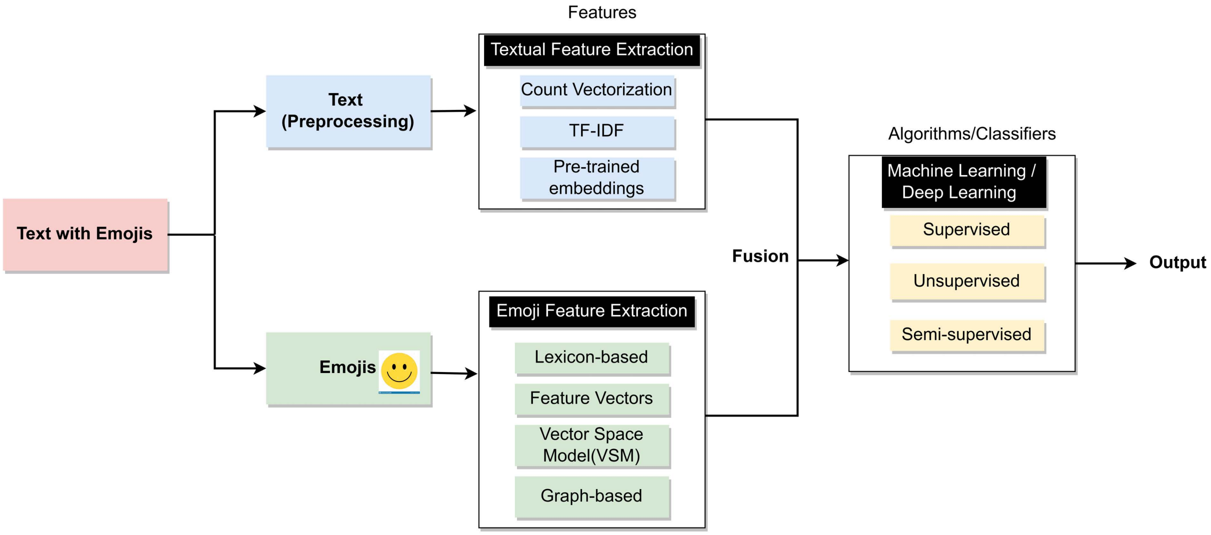

FIGURE 6. General overview of the multimodal model of text and emojis.

In conversational agents, multimodality in the form of text and emojis/emoticons can be used for a better understanding of user intent. This understanding of intent will help conversational agents to respond accurately and appropriately. Also, emojis/emoticons have been widely adopted in social media communications, a large number of emoji-labeled texts can be easily accessed to help tackle the scarcity of manually labeled data.

The advent of big data has accelerated the use of deep learning approaches in conversational agents. Popularly transformer-based sequence to sequence models are utilized for implementing conversational agents. In multimodal conversational agents, very little research has been done, particularly in text-based conversational agents with visual information. Hence, we have studied research that has text with emojis/emoticons.

Emoji representation and approaches play a key role in multimodal systems when designing conversational agents that can handle both text and emojis/emoticons. (Many research has used 'emojis' or 'emoticons' synonymously). There are different emoji representations or embedding methods presented in various researches. Feature vectors, Vector Space Model (VSM), Lexicon-based, and Graph-based representation methods have been used. Emoji-based feature extraction methods utilize diverse contextual or syntactical relationships with the help of frequency distribution of emojis. Figure 6 shows a general overview of the multimodal model of text and emojis.

Various research has been done related to emojis. In [59] presents a comprehensive overview of research on emoji, emoji evolution, utilization with their purposes, and what research has been done on them in various fields with future research directions. Emoji prediction from text has recently attracted a lot of interest. [60], [61] predicted emojis from images and text. [62] used multimodal emoji embeddings by combining image and text. [63] developed a new emoji embedding named 'Emojional', to represent a more emotional approach towards emojis. [64] presented vector skip-gram model to represent vectors for the emojis from Twitter. Some research has been done for different tasks using multimodal utilizing text and emojis in various domains. The studies undertaken for emotion classification using text, emojis and images modalities [65], sentiment prediction using text and emojis [66], emotion analysis from Twitter data using emojis [67], sentiment analysis of social media using text, emoticons, emojis [68] [69], enhancing opinion mining using emojis [70] are a few examples of study in this field. Moreover, in most research articles, deep learning methods have been employed for different tasks.

So, the unexploited potential exists in the study of multimodal conversational agents, which let users and conversational agents converse using both human language and visual information to be more realistic, human-like, and engaging. Sunder and Heck [10] have defined and mathematically formulated the goal of the multimodal conversational study. They suggested four basic problems in multimodal conversational systems: disambiguation, response generation, coreference resolution, and dialogue state tracking. The authors suggested a taxonomy of the types of study that are necessary to accomplish the goal of multimodal conversational agents: multimodal representation, multimodal fusion, multimodal alignment, multimodal translation (cross-modality), and co-learning. Thus, it becomes necessary to consider this taxonomy while designing multimodal conversational agents.

E. DATASETS USED IN CONVERSATIONAL AGENTS

This section discusses the primary datasets used by researchers in the field of conversational agents. In conversational agents, researchers have curated their own datasets or used publicly available datasets according to the needs of studies in certain application areas. This section presents some publicly available and useful datasets specifically

TABLE 4. Overview of different datasets used in the research of conversational agents.

Name of the dataset	Source	Labeled /Balanced	Size	Multimo dal	Emotions included	Topics covered
Topical-Chat Dataset [71]	Amazon	Yes/ Not balanced	The total number of conversations is 11,319 with training, validation, and testing sets. (Training - 9058 + validation with frequent - 565 and rare - 566 + test with frequent - 565 and rare - 565)	No	Angry, Disgusted, Fearful, Sad, Happy, Surprised, Curious to Dive Deeper, Neutral (Emotion – 8)	Fashion Politics Books Sports General Entertainment Music Science & Technology Movies
EMPATHETIC DIALOGUES [72]	Amazon	Yes / Approximately balanced	25K conversations – (Training - 19533 / Validation - 2770 / Testing - 2547 conversations)	No	Surprised, Excited, Angry, Proud, Sad, Annoyed, Lonely, grateful, alone, Afraid, scared, Guilty, Impressed, Disgusted, Hopeful, Confident, Furious, Anxious, Anticipating, Joyful, Nostalgic, Disappointed, Prepared, Jealous, Content, Devastated, Embarrassed, Caring, Sentimental, Trusting, Ashamed, Apprehensive, Faithful (Emotion – 32)	Open-domain
NLPCC 2017 [73]	Post/response pairs from Weibo	Yes/ Not balanced	1119207 posts	No	Other, Like, Sadness, Disgust, Anger, Happiness (Emotion – 6)	Open-domain
Daily dialog [74]	Various websites serve for English learners to practice English dialog in daily life	Yes / Not balanced	13,118 dialogues. (Training/validation/test set 11,118/1,000/1,000)	No	Anger, disgust, fear, happiness, sadness, surprise (Emotion – 6)	Various topics in daily life like ordinary, school life, culture and education, work, relationship

for improving context-awareness in conversational agents. Table 4 gives an overview of specific datasets with sources from which they were curated, whether the dataset is labeled or not, whether the dataset is balanced or not, the size of the dataset, whether the dataset is multimodal or not, emotions labeled in the dataset, and topics covered by dataset. Table 5 shows the different surveyed datasets used in different articles with techniques has been applied to datasets and challenges identified in articles with their performance measures.

1) TOPICAL CHAT DATASET

Topical-Chat is an open-domain knowledge grounded conversation dataset. The total number of conversations are 11,319 with training, validation and testing sets. The training set has 9058 messages, validation set with frequent and rare 565 respectively and test set with frequent and rare 565 messages, respectively. These conversations are annotated with the sentiment of their message on an 8-point scale. It includes Angry, Disgusted, Fearful, Sad, Happy, Surprised, Curious to Dive Deeper, and Neutral.

2) EMPATHETICDIALOGUES DATASET

It is a novel dataset comprising 25K conversations, including emotional context information to help training and evaluating the textual dialogue systems. It is an open domain conversational dataset freely accessible through the ParlAI framework. It consists dataset divided into approximately 80% train, 10% validation, and 10% test sets. The final train/ validation /test split contains 19533 / 2770 / 2547 conversations. A total of 32 emotion labels are included in the dataset shown in table 4.

3) NLPCC 2017

Natural Language Processing & Chinese Computing (NLPCC) dataset curated for the shared task of the NLPCC Emotion Classification Challenge to generate an emotional response. The dataset is constructed from More than 1 million Weibo posts and replies/comments pairs. The test dataset consists of about 5000 posts. The emotion categories are Anger, Disgust, Happiness, Like, Sadness, and Other.

4) DAILY DIALOG DATASET

Daily dialog dataset is crawled from different websites that aid English learners to apply English dialogues in daily life. It consists of multi-turn conversations. It contains 13,118 dialogues annotated with emotion labels as anger, disgust, fear, happiness, sadness, and surprise. It consists of a training set with 11,118 dialogues, a validation set with 1000 and a test set with 1000 dialogues.

F. APPLICATION AREAS

Conversational agents can be used in various application domains with different goals or objectives. Conversational agents can be utilized for decision-making, opinion, conflict resolution, and multi-party interaction. Conversational agents play diverse roles as information providers, recommenders, tutors, entertainers, advisors, personal assistants, customer service assistants and conversational partners in various fields. Review articles that have discussed the role of conversational agents in various application fields such as business [79], customer services [80], healthcare [3], and education [4].

1) CONVERSATIONAL AGENTS IN BUSINESS

Conversational agents provide cost-effective, highly available, and scalable services, enhancing market competitiveness and service quality [81]. These conversational agents also increase user or customer emotional engagement by extending customized flexibility, friendliness, comfortness, and efficient assistance [79]. Machine learning and sentiment

TABLE 5. Surveyed datasets with techniques, challenges, and performance measures in research.

Dataset	Reference	Techniques	Performance Measures	Challenges
Daily Dialog	[74]	Attention-based Seq2Seq, hierarchical encoder-decoder (HRED)	Perplexity - 55.94, 56.59, 59.24	The performance of the system can be improved
	[75]	GRUs and hierarchical attention matching network	F1-score (Empathy tracking) - 0.840	Predicting emotions like Sadness or Anger were more difficult than other emotions due to an Imbalanced dataset
Topical chat dataset	[53]	GRU-based shared encoder with self-attention, Bi-GRU-based decoder Focal Loss, consistency loss methods	F1-score – 0.23 (Fre) / 0.19 (Rare)	Due to data scarcity and less variety, models may have learned to predict the most frequent utterances and also unable to produce responses of particular emotion labels (angry, sad, fearful, and disgusted)
	[76]	Attention-based Pre-trained Bidirectional Encoder Representations from Transformers (BERT) model.	F1 Score (Fre) - 63.60% (Rare) - 58.09%	For detecting certain emotions, context and knowledge are still ineffective like the emotions of "disgusted" and "sad" are notoriously difficult to identify and distinguish
Empathetic Dialogues	[58]	Used transformer encoder-decoder attention mechanism, RoBERTa tokenizer	F1-Score –0.2864	The accuracy of the response predictor can be improved Predicting questioning category due to unbalanced distribution in the training dataset
	[72]	Retrieval-based and generative-based BERT model with pretraining and fine-tuning	PPL AVG - 21.24 BLEU - 6.27	Although the model appears to be more empathic, they are still far from human performance
NLPCC 2017	[77]	Seq to seq model using a reinforcement learning model	Perplexity - 62.2 Accuracy - 0.871	Incapable of extracting the emotional content of a conversation effectively. The emotional power of created responses is uncontrollable. There is no consideration of lexical, syntactic, grammatical, or other information relating to emotional elements
	[78]	Seq2seq architecture based on GRU model	Perplexity - 169.45 Emotion accuracy - 0.9658	Not capable enough to generate informative and interesting responses. The model cannot simulate the empathy phenomenon in human conversation

analysis advancements have given conversational agents the ability to respond emotionally to users. [33], [47], [48], [81], [82], discussed conversational agent systems for the business domain using different machine learning and deep learning techniques. In business-like E-commerce, banking mainly conversational agents are designed to answer FAQs. General question answers are not considered. Again, many conversational agents are tested in a simulated environment. Real-time exposure is needed for systems.

2) CONVERSATIONAL AGENTS IN HEALTHCARE

In the Healthcare, conversational agents are used for mental well-being promotion, diet management, medication observance, and physical activity promotion [88]. Moreover, particularly elderly users are comfortable with conversational agents due to the low learning graph. [37] used a Supervised machine-learning model to create a health mediator conversational agent. Conversational agents could help users in healthcare by delivering relevant information about a disorder or explaining the results of clinical tests [56]. [83] and [84] developed conversational agents to support people's mental well-being by analyzing emotions and feelings. In healthcare, domain-specific datasets are needed.

3) CONVERSATIONAL AGENTS IN EDUCATION

Educational conversational agents are invented to facilitate and assist online learning and deliver instructional content [41]. [86] showcased conversational agents in education to help students in re-learning administration issues. [49], [87] implemented pedagogical-driven conversational agents with sentiment analysis using reinforcement learning. In education, mostly conversational agents were based on retrieval or pattern-based. Nowadays, machine learning and deep learning techniques are being used.

Table 6 highlights the application area, objectives, method used, data utilized, and challenges in the selected papers.

From an applications perspective, there is still a disconnect between industrial technologies and current breakthroughs in the sector. The technologies utilized in research are not suited for use in the industry since they demand a lot of computational resources and extremely big training datasets. Again, conversational agents that must be used in various businesses have distinct requirements. Also, protecting users' personal information is an important issue in conversational agents. A review table of the conversational agents in different application areas is presented in Table 6.

IV. RESEARCH GAPS

With the help of deep learning models, conversational agents have made significant development in recent years. Several unique ideas such as pretrained embedding, different attention mechanisms, transformer-based models, pretrained deep learning models, and seq2seq models have been developed, resulting in rapid advancement in the last few years. Despite the advancements, there are still issues to be resolved in the field of conversational agents. The major limitation is making conversations natural with humans with the help of empathy, sentiments, and emotions. This section highlights some of these issues as well as research directions that could aid in the field's advancement.

TABLE 6. A review of the conversational agents in different application areas.

Application Area	References	Commonly used Techniques	Datasets	Challenges	Performance Measures (In order to Ref)
User Interaction	[33] [42] [58] [48]	• Deep learning - BERT, USE, Bi-LSTM, Bi-GRU, RoBERTa • Machine learning – Ensemble of LR, RF, and SVM	• SemEval-2019 • Topical Chat dataset • Large-scale multi-turn dialog dataset from the OpenSubtitles2018 corpus • Yelp challenge recommendation dataset	• The data used was quite imbalanced, and scarcity of labels in the data • Accuracy isn't up to standard with state-of-the-art systems	• 77% (F-SCORE) • 0.23 / 0.19 (F1-SCORE) • 0.2864 (F1-SCORE) • 87.5 (Accuracy)
Business	[81] [82] [47]	• Markov Chains • Latent Semantic Analysis (LSA) and Artificial Intelligence Markup Language (AIML) • LSTM	• Service dialog dataset • Frequently asked questions (FAQs) from the e-business domain • Raw Slack channel data	• Performance of the system can be enhanced • General questions are not taken into account • Less annotated data used for training	• 70.59% (Emotion accuracy) • 0.97 (Precision) • Not Mentioned
Healthcare	[88, p. 31] [83] [84]	• Finite-state Machine architecture, Support Vector Machine (SVM) • Lexicon-based method, Fuzzy matching method • Supervised ML algorithm	• Manually built datasets from hospitals and medical clinics	• Need to determine implicit emotions • Evaluated on a small dataset and limited set of emotions	• Not Mentioned • 81% (Accuracy) • 99.2 (Accuracy)
Education	[86] [89] [49] [90]	• Text Classification and Named Entity Recognition • Rules-based method • LSTM	• Student data from the Ho Chi Minh City University • Short posts in Web forums and Wikis • Chatlogs	• Handled the user context in a limited way • A lot of data needed to make agents intelligent • Domain-dependent system • Needs protection of personal information in a highly interactive tool	• 97.3 (F1-score context information) and 82.33 (F1-score Intent identification) • Not Mentioned • Not Mentioned • 0.93 (Average compound Score)

A. LIMITATIONS IN UNDERSTANDING THE CONTEXT [87]

This is the biggest challenge for conversational agents. Conversational agents are lagging in the understanding of natural language. Conversational agents do not understand the human context. These are programmed in a way that they only know what they are taught. They can only respond to people to the extent that they have been programmed to do so. If the query is beyond the conversational agents' training, it will be unable to understand or respond, which will frustrate the user/customer. Communication between conversational agents and human depends on what the user said in previous messages. Conversational agents need to build the state of the conversation. It can only answer the question if it knows the details of conversations. Typically, information is stored in the context of the conversation. Each conversational agent has to model its own notion of the context and decide the information that is important to remember.

B. FAILURE TO DETECT THE INTENT OF THE USER [91]

Intent is the purpose behind the user query. Most conversational agents are unable to understand users' intentions. The user would ask a question with single and straightforward intent. But humans have a tendency to communicate by combining several intentions into one sentence. Like when the user comes with two different requests in one sentence. Also, conversational agents could not understand the intent of the user when complex linguistics such as negation or conditional structures are used in the query. Conversational agents need to understand what the user is requesting, even if it is phrased unexpectedly. Training data is a major barrier to understanding intent. Machine learning takes a staggering amount of data to understand humans because conversational agents must understand the relationships between words, phrases, sentences, synonyms, lexical elements, concepts, and so on.

C. NEED TO UNDERSTAND EMOTIONAL SEMANTIC INTERPRETATIONS [79], [92]

Semantics plays an important role in the cognitive analysis. Written text from the user may have negations and modals, which can affect the impact of emotions and sentiments. For example, CAs should infer the meaning of words like "maybe good," "was good," and "was not good" differently. Words/phrases used in different contexts and different senses can impart diverse emotions. So, ambiguity in word semantic interpretations is one of the important issues in CAs.

D. DATASET-RELATED CHALLENGES [42]

Datasets plays an important role as training data is required to understand intent and context and respond naturally to user. There are many challenges related to datasets like small datasets available, scarcity of labeled data, unbalanced distribution of data, less variety of labels in datasets, and lack of representative publicly available datasets. All dataset-related challenges are important as machine learning, and deep learning techniques require a huge amount of data for training.

E. NEED TO DETECT IMPLICIT EMOTIONS [83]

Emotion analysis can play an important role in context and provide more natural responses to the user. Text entered by the user may not have a direct emotion keyword mentioned in the input. So, such emotions in the written text need to be detected by conversational agents and respond accordingly. The query's words/phrases may have different meanings. Multiple emotions are difficult to recognize since a single sentence can contain multiple emotions and different points of view. To improve the performance or accuracy of emotion recognition, this problem must be addressed. Thus, for building context-aware conversational agents by employing emotion analysis, implicit emotion analysis is an important issue.

F. NEED TO GENERATE EMOTIONALLY AWARE RESPONSE [58]

Current conversational agents lack empathy, making it difficult to detect and understand user emotions and respond in a more natural manner. Emotionally aware responses are important for customer/user satisfaction and retention. If conversational agents are not able to produce an emotionally aware response, the user will get frustrated and annoyed, and it affects interaction. It is also one of the reasons for conversational agents lacking long-time conversations.

G. NEED TO HANDLE THE QUALITY OF USER INPUT [93]

Need to handle user input with misused phrases, subtle sarcasm, language impairments, usage of slang, and syntax faults. To express feelings and sentiments, users employ irony, sarcasm, and humor. User texts, on the other hand, may contain casual language, slang words, misspellings, hashtags, emojis, and abbreviations, among other things. As a result, interpreting such metaphorical language for conversational agents to understand and respond appropriately becomes tough.

H. LACK OF EVALUATION METRICS FOR CONVERSATIONAL AGENTS [94]

Evaluating conversational agents has remained a challenging task. There are some widely used performance metrics for conversational agents. F1-score, perplexity, BLEU, and METEOR are some common metrics used. But there is no common framework for conversational agents' evaluation. So many systems need to rely on human evaluation. And even there is no common reference for human evaluation too. So, a reliable automatic evaluation method for conversational agents should be proposed.

I. PRIVACY AND DATA SECURITY [1]

There are significant issues that arise related to the privacy and data security of user and conversational agent provider service. Service providers are responsible for acquired user information and must prevent it by other third parties. It is crucial to maintain privacy and data security regards to user authentication and authorization [20], end-to-end encryption, and finance-related communications with conversational agents. Thus, in such scenarios, information security and privacy protection techniques and technologies should be applied to conversational agents. Currently, conversational agents need adequate research on data security and privacy.

V. FUTURE DIRECTIONS

This section highlights some of the research directions that could aid in the field's advancement. Techniques or methodologies in conversational agents have seen huge signs of progress in the last few years, from rule-based methods to hidden layer-based deep learning methods and pretrained models such as. Artificial Intelligence advancements in recent years have bolstered trends in conversational agents, making them to understand and reply in natural languages and reply. The authors have discussed research gaps in section 5. Table 7 shows the mapping between research gaps and future directions. To make conversational agents contextual, they need a lot of data and a vast knowledge base for training. So, training conversational agents on large datasets is one of the solutions to make them contextual. Also, self-training and reinforcement learning techniques can be applied to make them contextual. Some of the difficulties mentioned in the research gaps section have been addressed by AI-based technologies such as transfer learning, reinforcement learning, multi-task learning, meta-learning, self-learning, and GAN's. This section discusses challenges and their solutions using these methods with references.

A. TRANSFER LEARNING

Transfer learning has majorly contributed to the progress of modernistic NLP systems like conversational agents. Particularly conversational agents can be benefited from inductive transfer learning, where unlabelled data is employed to pull knowledge for labeled downstream tasks. [95] discussed a framework to transfer the affective knowledge. In this proposed system, authors pre-trained a hierarchical dialogue model on multi-turn conversations (source) and then transferred its parameters to a conversational emotion classifier (target).

B. REINFORCEMENT LEARNING

In reinforcement learning, conversational agents are trained through trial-and-error conversations with either real users or a rule-based user simulator. [96] proposed deep reinforcement learning for dialogue generation. [96] work marked a first step towards learning a conversational neural model based on the long-term success of dialogues. [97] developed a reinforcement learning-based emotional editing constraint conversation content-generating model.

C. MULTI-TASK LEARNING

All or a subset of the tasks in multi-task learning are related but not identical. It aims to help improve the learning of a model for a task by using the knowledge contained in all-related tasks. Basic two factors are considered for multi-task learning [98]. First is relatedness, that is how different tasks

TABLE 7. Challenges identified through the literature survey and suggested future directions.

Problem area	Research Gaps	Future Directions
Semantics	• Failure to detect the intent of the user	• RNN and its variants [100] • Meta-Learning [101] • Training the model on large datasets
	• Need to detect implicit emotions	• Pre-trained word embeddings [102] [103]
	• Unable to extract the semantic data	• RNN and its variants [104]
Quality of text	• Incomplete information, typing mistakes	• Text Pre-processing
	• Slang words, short texts, emojis	• Lexicons [102]
	• Detection of sarcasm, irony, harmony	• Multitask Learning [105] • Transfer learning [106]
Datasets	• Scarcity of labeled data • Small datasets available • Less variety of labels in datasets	• Generative Adversarial Networks (GAN) [107]
	• Unbalanced distribution of data • Lack of representative publicly available datasets	• Domain Adaptation [108] • Transfer Learning [109]
Accuracy	• Limitations in understanding the context	• Reinforcement Learning [48] • Semi-Supervised Learning [110]
	• Need to understand the emotional semantic interpretation	• Transformer based encoder-decoder models [58] [111]
	• Need to generate emotionally aware responses	• Multitask Learning [42] • Meta-learning [112] • Emotion detection [58]
Privacy and Security	• User privacy and data security	• Adversarial Machine Learning [109]

are related to each other and kinds of learning tasks such as supervised like classification or regression tasks, unsupervised learning, clustering task, and many more learning tasks like semi-supervised, reinforcement learning. In [42] proposed a model where authors used multi-task learning to classify the emotions and to generate the response based on a given sentence with a common encoder and multiple decoders.

D. META-LEARNING

Meta-Learning is ‘‘learning about learning’’ or ‘‘learn to learn.’’ Meta-learning helps to overcome the requirement of annotated data for language processing applications. Meta-learning is the most suitable for conversational agents where knowledge across user intents, domains, and languages can be transferred easily. Meta-learning is a scalable solution for businesses in terms of conversational agents. [99] proposed meta-learning approach for intent detection as an n-way k-shot classification problem. Initially, the authors utilized English utterances and then evaluating on Spanish and Thai utterances. Similarly, authors in [72] demonstrated how meta-learning allows one to learn quickly and adapt to different personas using only a few dialogue samples from the same user. Their results using automatic evaluation metrics showed that meta-learning outperformed non-meta-learning baselines.

E. GAN (GENERATIVE ADVERSARIAL NETWORKS)

Many tasks in natural language processing (NLP), such as question answering, necessitate a substantial amount of training data to improve model performance. To generate a large amount of training data, however, gathering and annotating more data can be an expensive and time-consuming operation. Data augmentation strategies can be used in this situation. One of the data augmentation techniques is GAN. These are computational structures that set two neural networks against each other to develop new, synthetic data samples that can pass for real data. [77] developed a new technique for dialogue generation in conversational agents called Cascade Generative Adversarial Network (Cas-GAN), which is a blend of GAN and RL.

F. ADVERSARIAL MACHINE LEARNING (AML)

The development of methods to scrutinize conversational agents’ privacy protection as well as strategies to increase the agents’ resistance to malicious attacks and/or data theft are the main objectives. In that case, AML, or Adversarial Machine Learning, is a new area of research that combines the latest machine learning techniques, information systems security, and robust statistics can aid to solve security-related problems. [113] explored attack/defense tactics for adversarial recommender systems using generative adversarial networks. [114] explores the characteristics of adversarial machine learning, particularly in text generation and summarizes significant contributions to the field, such as algorithms, models, attack types, and defense techniques.

VI. CONCLUSION

Recent developments and studies in Artificial Intelligence have facilitated conversational agents. The studies in the field of conversational agents have sought ways to use research findings in a variety of application areas, such as businesses, education, entertainment, and healthcare. The various approaches have been employed in different components of conversational agents, from rule-based methods to machine learning algorithms, and as the research trends are changing directions towards deep learning techniques. This article discusses about bringing conversational agents closer to natural language communication by employing context-awareness. This comparative study presented to review articles on conversational agents based on deep learning with current trends. This study also reviews datasets used in conversational agents. In this survey, recent and trending deep learning techniques in conversational agents have been discussed. Following that, it attempted to shed light on how current research gaps and future directions will affect research in the field.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualization and Methodology: Sheetal Kusal and Shruti Patil, Writing Original Draft: Sheetal Kusal; Supervision: Shruti Patil, Sashikala Mishra, Ketan Kotecha; Resources: Ajith Abraham and Jyoti Choudrie; Review and Editing: Sheetal Kusal, Shruti Patil, Sashikala Mishra, Ketan Kotecha, Jyoti Choudrie; Funding Acquisition: Ajith Abraham.

REFERENCES

[1] E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications," *Mach. Learn. with Appl.*, vol. 2, Dec. 2020, Art. no. 100006, doi: 10.1016/j.mlwa.2020.100006.

[2] R. Bavaresco, D. Silveira, E. Reis, J. Barbosa, R. Righi, C. Costa, R. Antunes, M. Gomes, C. Gatti, M. Vanzin, S. C. Junior, E. Silva, and C. Moreira, "Conversational agents in business: A systematic literature review and future research directions," *Comput. Sci. Rev.*, vol. 36, May 2020, Art. no. 100239, doi: 10.1016/j.cosrev.2020.100239.

[3] J. L. Z. Montenegro, C. A. da Costa, and R. da Rosa Righi, "Survey of conversational agents in health," *Expert Syst. Appl.*, vol. 129, pp. 56–67, Sep. 2019, doi: 10.1016/j.eswa.2019.03.054.

[4] S. Hobert and R. Meyer von Wolff, "Say hello to your new automated tutor—A structured literature review on pedagogical conversational agents," in *Proc. 14th Int. Conf. Wirtschaftsinformatik*, Siegen, Germany, Feb. 2019, pp. 301–314.

[5] J. Fraser, I. Papaioannou, and O. Lemon, "Spoken conversational ai in video games: Emotional dialogue management increases user engagement," in *Proc. 18th Int. Conf. Intell. Virtual Agents*, 2018, pp. 179–184, doi: 10.1145/3267851.3267896.

[6] S. Aru. (2021). *Conversational AI: Why it Becomes A Priority in 2021?* [Online]. Available: https://botmywork.com/blog/conversational-ai-becomes-priority/0A

[7] BRAIN [BRN.AI] CODE FOR EQUITY. (2019). *Chatbot Report 2019: Global Trends and Analysis*. [Online]. Available: https://chatbotsmagazine.com/chatbot-report-2019-global-trends-and-analysis-a487afec05b0A

[8] H. Golchha, M. Firdaus, A. Ekbal, and P. Bhattacharyya, "Courteously yours: Inducing courteous behavior in customer care responses using reinforced pointer generator network," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Language Technol.*, 2019, pp. 851–860.

[9] "Emotionally-aware chatbots: A survey," *CoRR*, 2019.

[10] A. S. Sundar and L. Heck, "Multimodal conversational AI a survey of datasets and approaches," 2022, *arXiv:2205.06907*.

[11] H. Prendinger and M. Ishizuka, "The empathic companion: A character-based interface that addresses users' affective states," *Appl. Artif. Intell.*, vol. 19, nos. 3–4, pp. 267–285, Mar. 2005, doi: 10.1080/08839510590910174.

[12] B. Galitsky, "Chatbot components and architectures," in *Developing Enterprise Chatbots*. Cham, Switzerland: Springer, 2019, pp. 13–51, doi: 10.1007/978-3-030-04299-8_2.

[13] P. Kulkarni, A. Mahabaleshwarkar, M. Kulkarni, N. Sirsikar, and K. Gadgil, "Conversational AI: An overview of methodologies, applications & future scope," in *Proc. 5th Int. Conf. Comput. Commun. Control Automat. (ICCUBEA)*, Sep. 2019, p. 7.

[14] S. Kusal, S. Patil, K. Kotecha, R. Aluvalu, and V. Varadarajan, "AI based emotion detection for textual big data: Techniques and contribution," *Big Data Cognit. Comput.*, vol. 5, no. 3, p. 43, Sep. 2021, doi: 10.3390/bdcc5030043.

[15] L. Qiu, Y. Chen, H. Jia, and Z. Zhang, "Query intent recognition based on multi-class features," *IEEE Access*, vol. 6, pp. 52195–52204, 2018, doi: 10.1109/ACCESS.2018.2869585.

[16] X. Dong, L. Qian, Y. Guan, L. Huang, Q. Yu, and J. Yang, "A multiclass classification method based on deep learning for named entity recognition in electronic medical records," in *Proc. New York Sci. Data Summit (NYSDS)*, New York, NY, USA, 2016.

[17] D. Nadeau and S. Sekine. *A Survey of Named Entity Recognition and Classification*. Accessed: Mar. 17, 2022. [Online]. Available: http://projects.ldc.upenn.edu/gale/

[18] A. Mccallum and W. Li, "Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons," in *Proc. 7th Conf. Natural Lang. Learn. HLT-NAACL*, 2003, pp. 188–191.

[19] N. Ali, "Chatbot: A conversational agent employed with named entity recognition model using artificial neural network," 2020, *arXiv:2007.04248*

[20] S. Zheng, Y. Hao, D. Lu, H. Bao, J. Xu, H. Hao, and B. Xu, "Joint entity and relation extraction based on a hybrid neural network," *Neurocomputing*, vol. 257, pp. 59–66, Sep. 2017, doi: 10.1016/j.neucom.2016.12.075.

[21] T. Zhang, J. H. D. Cho, and C. Zhai, "Understanding user intents in online health forums," in *Proc. 5th ACM Conf. Bioinf., Comput. Biol., Health Informat.*, Sep. 2014, pp. 220–229, doi: 10.1145/2649387.2649445.

[22] M. Mendoza and J. Zamora, *Building Decision Trees to Identify the Intent of a User Query* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 5711. Berlin, Germany: Springer-Verlag, 2009, pp. 285–292, doi: 10.1007/978-3-642-04595-0_35.

[23] H. Cuayáhuitl, S. Renals, O. Lemon, and H. Shimodaira, "Human-computer dialogue simulation using hidden Markov models," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand.*, Nov. 2005, pp. 290–295.

[24] A. Kammoun, R. Slama, H. Tabia, T. Ouni, and M. Abid, "Generative adversarial networks for face generation: A survey," *ACM Comput. Surveys*, Mar. 2022, doi: 10.1145/1122445.1122456.

[25] P. Kulkarni, A. Mahabaleshwarkar, M. Kulkarni, N. Sirsikar, and K. Gadgil, "Conversational AI: An overview of methodologies, applications & future scope," in *Proc. 5th Int. Conf. Comput., Commun., Control Autom. (ICCUBEA)*, Sep. 2019, pp. 1–7, doi: 10.1109/ICCUBEA47591.2019.9129347.

[26] G. Caldarini, S. Jaf, and K. McGarry, "A literature survey of recent advances in chatbots," *Information*, vol. 13, no. 1, p. 41, Jan. 2022, doi: 10.3390/info13010041.

[27] M. Allouch, A. Azaria, and R. Azoulay, "Conversational agents: Goals, technologies, vision and challenges," *Sensors*, vol. 21, no. 24, p. 8448, Dec. 2021, doi: 10.3390/s21248448.

[28] M. Firdaus, S. Bhatnagar, A. Ekbal, and P. Bhattacharyya, "Intent detection for spoken language understanding using a deep ensemble model," in *Proc. 15th Pacific Rim Int. Conf. Artif. Intell.*, in Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 11012. Nanjing, China, 2018, pp. 629–642, doi: 10.1007/978-3-319-97304-3_48.

[29] M. Allouch, A. Azaria, and R. Azoulay, "Conversational agents: Goals, technologies, vision and challenges," *Sensors*, vol. 21, no. 24, p. 8448, Dec. 2021, doi: 10.3390/s21248448.

[30] J. Tao. *Context Based Emotion Detection From Text Input*. Accessed: Mar. 21, 2022. [Online]. Available: http://www.isca-speech.org/archive

[31] O. Udochukwu and Y. He, "A rule-based approach to implicit emotion detection in text," in *Proc. Int. Conf. Appl. Natural Lang. Inf. Syst.*, 2015, pp. 197–203.

[32] M. Allouch, A. Azaria, R. Azoulay, E. Ben-Izchak, M. Zwilling, and D. A. Zachor, "Automatic detection of insulting sentences in conversation," in *Proc. IEEE Int. Conf. Sci. Elect. Eng. Israel (ICSEE)*, Jan. 2018, pp. 375–378.

[33] A. Basile, M. Franco-Salvador, N. Pawar, S. Sanja-Štajner, M. C. Rios, and Y. Benajiba, "SymantoResearch at SemEval-2019 task 3: Combined neural models for emotion classification in human-chatbot conversations," in *Proc. 13th Int. Workshop Semantic Eval.*, 2019, pp. 330–334.

[34] P. Gervás, B. Díaz-Agudo, F. Peinado, and R. Hervás, "Story plot generation based on CBR," *Knowl.-Based Syst.*, vol. 18, nos. 4–5, pp. 235–242, Aug. 2005, doi: 10.1016/j.knosys.2004.10.011.

[35] T.-H. Wen, M. Gasic, D. Kim, N. Mrksic, P. H. Su, D. Vandyke, and S. Young, "Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking," 2015, *arXiv:1508.01755*.

[36] M. Qiu, F.-L. Li, S. Wang, X. Gao, Y. Chen, W. Zhao, H. Chen, J. Huang, and W. Chu, "AliMe chat: A sequence to sequence and rerank based chatbot engine," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, 2017, pp. 498–503, doi: 10.18653/v1/P17-2079.

[37] Y. Keneshloo, T. Shi, N. Ramakrishnan, and C. K. Reddy, "Deep reinforcement learning for sequence to sequence models," Mar. 2018, *arXiv:1805.09461*.

[38] P. Colombo, W. Witon, A. Modi, J. Kennedy, and M. Kapadia, "Affect-driven dialog generation," 2019, *arXiv:1904.02793*.

[39] H. Zhou, M. Huang, T. Zhang, X. Zhu, and B. Liu, "Emotional chatting machine: Emotional conversation generation with internal and external memory," 2017, *arXiv:1704.01074*.

[40] T. Hu, A. Xu, Z. Liu, Q. You, Y. Guo, V. Sinha, J. Luo, and R. Akkiraju, "Touch your heart: A tone-aware chatbot for customer care on social media," 2018, *arXiv:1803.02952*.

[41] M. Firdaus, H. Chauhan, A. Ekbal, and P. Bhattacharyya, "MEISD: A multimodal multi-label emotion, intensity and sentiment dialogue dataset for emotion recognition and sentiment analysis in conversations," in *Proc. 28th Int. Conf. Comput. Linguistics*, 2020, pp. 4441–4453.

[42] D. Varshney, A. Ekbal, and P. Bhattacharyya, "Modelling context emotions using multi-task learning for emotion controlled dialog generation," in *Proc. Conf. Eur. Assoc. Comput. Linguistics*, 2021, pp. 2919–2931.

[43] Y. Xie and P. Pu, "Empathetic dialog generation with fine-grained intents," 2021, *arXiv:2105.06829*.

[44] L. Gui, J. Hu, Y. He, R. Xu, Q. Lu, and J. Du, "A question answering approach to emotion cause extraction," 2017, *arXiv:1708.05482*.

[45] G. Indra Winata, A. Madotto, Z. Lin, J. Shin, Y. Xu, P. Xu, and P. Fung, "CAiRE_HKUST at SemEval-2019 task 3: Hierarchical attention for dialogue emotion classification," 2019, *arXiv:1906.04041*.

[46] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. V. Le, and R. Salakhutdinov, "Transformer-XL: Attentive language models beyond a fixed-length context," 2019, *arXiv:1901.02860*.

[47] Z. Xue, T.-Y. Ko, N. Yuchen, M.-K. D. Wu, and C.-C. Hsieh, "Isa: Intuit smart agent, a neural-based agent-assist chatbot," in *Proc. IEEE Int. Conf. Data Mining Workshops*, Mar. 2018, pp. 1423–1428.

[48] Y. Sun and Y. Zhang, "Conversational recommender system," in *Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jun. 2018, pp. 235–244, doi: 10.1145/3209978.3210002.

[49] M. Feidakis, P. Kasnesis, E. Giatraki, C. Giannousis, C. Patrikakis, and P. Monachelis, "Building pedagogical conversational agents, affectively correct," in *Proc. 11th Int. Conf. Comput. Supported Educ.*, 2019, pp. 100–107, doi: 10.5220/0007771001000107.

[50] Y.-C. Chang and Y.-C. Hsing, "Emotion-infused deep neural network for emotionally resonant conversation," *Appl. Soft Comput.*, vol. 113, Dec. 2021, Art. no. 107861, doi: 10.1016/j.asoc.2021.107861.

[51] M. Firdaus, H. Chauhan, A. Ekbal, and P. Bhattacharyya. (2021). *More the Merrier: Towards Multi-Emotion and Intensity Controllable Response Generation*. [Online]. Available: www.aaai.org

[52] R. Shantala, G. Kyselov, and A. Kyselova, "Neural dialogue system with emotion embeddings," in *Proc. IEEE 1st Int. Conf. Syst. Anal. Intell. Comput.*, Oct. 2018, pp. 1–4.

[53] D. Varshney, A. Ekbal, and P. Bhattacharyya, "Modelling context emotions using multi-task learning for emotion controlled dialog generation," in *Proc. 16th Conf. Eur. Chapter Assoc. Comput. Linguistics*, 2021, pp. 2919–2931.

[54] L. Zhou, J. Gao, D. Li, and H.-Y. Shum, "The design and implementation of XiaoIce, an empathetic social chatbot," 2018, *arXiv:1812.08989*.

[55] H. Zhou, M. Huang, T. Zhang, X. Zhu, and B. Liu. *Emotional Chatting Machine: Emotional Conversation Generation With Internal and External Memory*. Accessed: Mar. 21, 2022. [Online]. Available: http://www.aaai.org

[56] R. Zhang, Z. Wang, and D. Mai. *Building Emotional Conversation Systems Using Multi-Task Seq2Seq Learning*. Accessed: Mar. 22, 2022. [Online]. Available: https://radimrehurek.com/gensim/

[57] Y. Xie and P. Pu, "Empathetic dialog generation with fine-grained intents," 2021, *arXiv:2105.06829*.

[58] Y. Xie and P. Pu, "Empathetic dialog generation with fine-grained intents," 2021, *arXiv:2105.06829*.

[59] Q. Bai, Q. Dan, Z. Mu, and M. Yang, "A systematic review of emoji: Current research and future perspectives," *Frontiers Psychol.*, vol. 10, p. 2221, Oct. 2019, doi: 10.3389/fpsyg.2019.02221.

[60] S. Cappallo, S. Svetlichnaya, P. Garrigues, T. Mensink, and C. G. M. Snoek, "The new modality: Emoji challenges in prediction, anticipation, and retrieval," 2018, *arXiv:1801.10253*.

[61] F. Barbieri, M. Ballesteros, F. Ronzano, and H. Saggion, "Multimodal emoji prediction," 2018, *arXiv:1803.02392*.

[62] I. Lim, J. Balasubramanian, and H. Chen. *Zero Shot Emoji Prediction Using Multimodal Emoji Embeddings Stanford CS224N Custom Project*. Accessed: May 8, 2022. [Online]. Available: https://github.com/hlgchen/emoji_prediction

[63] E. Barry, S. Jameel, and H. Raza, "*Emojional: Emoji Embeddings*. Accessed: Jul. 6, 2022. [Online]. Available: https://unicode.org/emoji/charts/full-emoji-list.html

[64] F. Barbieri, F. Ronzano, and H. Saggion. *What Does This Emoji Mean? A Vector Space Skip-Gram Model for Twitter Emojis*. Accessed: Jul. 6, 2022. [Online]. Available: http://instagram-engineering

[65] A. Illendula and A. Sheth, "Multimodal emotion classification," in *Proc. Companion World Wide Web Conf.*, May 2019, pp. 439–449, doi: 10.1145/3308560.3316549.

[66] T. P. Kumar, B. V. Vardhan, and A. Prof, "Multimodal sentiment prediction based on the integration of text and emojis," *J. Optoelectronics Laser*, vol. 41, no. 4, pp. 489–499. [Online]. Available: https://www.researchgate.net/publication/360099892

[67] W. Wolny. *Emotion Analysis of Twitter Data That Use Emoticons and Emoji Ideograms*. Accessed: May 8, 2022. [Online]. Available: www.twitter.com

[68] V. Jagadishwari, A. Indulekha, K. Raghu, and P. Harshini, "Sentiment analysis of social media text-emoticon post with machine learning models contribution title," *J. Phys., Conf.*, vol. 2070, no. 1, Nov. 2021, Art. no. 012079, doi: 10.1088/1742-6596/2070/1/012079.

[69] X. Li, J. Zhang, Y. Du, J. Zhu, Y. Fan, and X. Chen, "A novel deep learning-based sentiment analysis method enhanced with emojis in microblog social networks," *Enterprise Inf. Syst.*, vol. 2022, pp. 1–22, Feb. 2022, doi: 10.1080/17517575.2022.2037160.

[70] S. Al-Azani and E.-S.-M. El-Alfy, "Early and late fusion of emojis and text to enhance opinion mining," *IEEE Access*, vol. 9, pp. 121031–121045, 2021, doi: 10.1109/ACCESS.2021.3108502.

[71] K. Gopalakrishnan, B. Hedayatnia, Q. Chen, A. Gottardi, S. Kwatra, A. Venkatesh, R. Gabriel, D. Hakkani-Tür, and A. A. AI, "Topical-chat: Towards knowledge-grounded open-domain conversations," in *Proc. INTERSPEECH*, 2019, pp. 1891–1895.

[72] H. Rashkin, E. Michael Smith, M. Li, and Y.-L. Boureau, "Towards empathetic open-domain conversation models: A new benchmark and dataset," 2018, *arXiv:1811.00207*.

[73] *Natural Language Processing and Chinese Computing*, China, 2017.

[74] Y. Li. *DailyDialog: A Manually Labelled Multi-Turn Dialogue Dataset*. Accessed: Mar. 29, 2022. [Online]. Available: http://yanran.li/

[75] J. Wei, S. Feng, D. Wang, Y. Zhang, and X. Li, ''Attentional neural network for emotion detection in conversations with speaker influence awareness,'' in *Proc. Int. Conf. Natural Lang. Process. Chin. Comput.*, 2019, pp. 287–297.

[76] S. Ghosh, D. Varshney, A. Ekbal, and P. Bhattacharyya, ''Context and knowledge enriched transformer framework for emotion recognition in conversations,'' in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2021, pp. 1–8, doi: 10.1109/IJCNN52387.2021.9533452.

[77] J. Li, X. Sun, X. Wei, C. Li, and J. Tao, ''Reinforcement learning based emotional editing constraint conversation generation,'' 2019, *arXiv:1904.08061*.

[78] R. Zhang, Z. Wang, and D. Mai. *Building Emotional Conversation Systems Using Multi-Task Seq2Seq Learning*. Accessed: Mar. 22, 2022. [Online]. Available: https://radimrehurek.com/gensim/

[79] R. Bavaresco, D. Silveira, E. Reis, J. Barbosa, R. Righi, C. Costa, R. Antunes, M. Gomes, C. Gatti, M. Vanzin, S. C. Junior, E. Silva, and C. Moreira, ''Conversational agents in business: A systematic literature review and future research directions,'' *Comput. Sci. Rev.*, vol. 36, May 2020, Art. no. 100239, doi: 10.1016/j.cosrev.2020.100239.

[80] M. Nuruzzaman and O. K. Hussain, ''A survey on chatbot implementation in customer service industry through deep neural networks,'' in *Proc. IEEE 15th Int. Conf. e-Bus. Eng. (ICEBE)*, Oct. 2018, pp. 54–61, doi: 10.1109/ICEBE.2018.00019.

[81] A. Adikari, D. de Silva, D. Alahakoon, and X. Yu, ''A cognitive model for emotion awareness in industrial chatbots,'' in *Proc. IEEE 17th Int. Conf. Ind. Inform. (INDIN)*, 2019, pp. 183–186, doi: 10.1109/INDIN41052.2019.8972196.

[82] N. T. Thomas, ''An e-business chatbot using AIML and LSA,'' in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2016, pp. 2740–2742, doi: 10.1109/ICACCI.2016.7732476.

[83] K. Denecke, S. Vaaheesan, and A. Arulnathan, ''A mental health chatbot for regulating emotions (SERMO)–concept and usability test,'' *IEEE Trans. Emerg. Topics Comput.*, vol. 9, no. 3, pp. 1170–1182, Jul. 2021, doi: 10.1109/TETC.2020.2974478.

[84] B. Inkster, S. Sarda, and V. Subramanian, ''An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study,'' *JMIR mHealth uHealth*, vol. 6, no. 11, Nov. 2018, Art. no. e12106, doi: 10.2196/12106.

[85] D. Song, E. Y. Oh, and M. Rice, ''Interacting with a conversational agent system for educational purposes in online courses,'' in *Proc. 10th Int. Conf. Hum. Syst. Interact. (HSI)*, 2017, pp. 78–82.

[86] H. T. Hien, P.-N. Cuong, L. N. H. Nam, H. L. T. K. Nhung, and L. D. Thang, ''Intelligent assistants in higher-education environments: The FIT-EBot, a chatbot for administrative and learning support,'' in *Proc. 9th Int. Symp. Inf. Commun. Technol.*, 2018, pp. 69–76, doi: 10.1145/3287921.3287937.

[87] D. Shin, J. H. Lee, H. Kim, and H. Yang, ''Exploring the use of an artificial intelligence chatbot as second language conversation partners,'' *Korean J. English Lang. Linguistics*, vol. 21, pp. 375–391, Apr. 2021, doi: 10.15738/kjell.21..202104.375.

[88] A. Fadhil, Y. Wang, and H. Reiterer, ''Assistive conversational agent for health coaching: A validation study,'' *Methods Inf. Med.*, vol. 58, no. 1, pp. 9–23, 2019, doi: 10.1055/s-0039-1688757.

[89] D. Song, E. Y. Oh, and M. Rice, ''Interacting with a conversational agent system for educational purposes in online courses,'' in *Proc. 10th Int. Conf. Hum. Syst. Interact.*, 2017, pp. 78–82.

[90] D. Shin, J. H. Lee, H. Kim, and H. Yang, ''Exploring the use of an artificial intelligence chatbot as second language conversation partners,'' *Korean J. English Lang. Linguistics*, vol. 21, pp. 375–391, May 2021, doi: 10.15738/kjell.21..202104.375.

[91] M. Firdaus, S. Bhatnagar, A. Ekbal, and P. Bhattacharyya, ''Intent detection for spoken language understanding using a deep ensemble model,'' in *Proc. 15th Pacific Rim Int. Conf. Artif. Intell.*, in Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 11012. Nanjing, China, 2018, pp. 629–642, doi: 10.1007/978-3-319-97304-3_48.

[92] S. Zhang, E. Dinan, J. Urbanek, A. Szlam, D. Kiela, and J. Weston, ''Personalizing dialogue agents: I have a dog, do you have pets too?'' 2018, *arXiv:1801.07243*.

[93] A. Rapp, L. Curti, and A. Boldi, ''The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots,'' *Int. J. Hum.-Comput. Stud.*, vol. 151, Jul. 2021, Art. no. 102630, doi: 10.1016/j.ijhcs.2021.102630.

[94] G. Caldarini, S. Jaf, and K. McGarry, ''A literature survey of recent advances in chatbots,'' *Information*, vol. 13, no. 1, p. 41, Jan. 2022, doi: 10.3390/info13010041.

[95] D. Hazarika, S. Poria, R. Zimmermann, and R. Mihalcea, ''Conversational transfer learning for emotion recognition,'' *Inf. Fusion*, vol. 65, pp. 1–12, Jan. 2021, doi: 10.1016/j.inffus.2020.06.005.

[96] J. Li, W. Monroe, A. Ritter, M. Galley, J. Gao, and D. Jurafsky, ''Deep reinforcement learning for dialogue generation,'' 2016, *arXiv:1606.01541*.

[97] J. Li, X. Sun, X. Wei, C. Li, and J. Tao, ''Reinforcement learning based emotional editing constraint conversation generation,'' 2019, *arXiv:1904.08061*.

[98] Y. Zhang and Q. Yang, ''An overview of multi-task learning,'' *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 30–43, Jan. 2018.

[99] H. S. Bhathiya and U. Thayasivam, ''Meta learning for few-shot joint intent detection and slot-filling,'' in *Proc. 5th Int. Conf. Mach. Learn. Technol.*, Jun. 2020, pp. 86–92, doi: 10.1145/3409073.3409090.

[100] M. Firdaus, S. Bhatnagar, A. Ekbal, and P. Bhattacharyya, ''Intent detection for spoken language understanding using a deep ensemble model,'' in *Proc. 15th Pacific Rim Int. Conf. Artif. Intell.*, in Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 11012. Nanjing, China, 2018, pp. 629–642, doi: 10.1007/978-3-319-97304-3_48.

[101] H. S. Bhathiya and U. Thayasivam, ''Meta learning for few-shot joint intent detection and slot-filling,'' in *Proc. 5th Int. Conf. Mach. Learn. Technol.*, Jun. 2020, pp. 86–92, doi: 10.1145/3409073.3409090.

[102] S. M. Mohammad, ''Sentiment analysis: Detecting valence, emotions, and other affectual states from text,'' in *Emotion Measurement*. Woodhead Publishing, 2015, pp. 201–237, doi: 10.1016/B978-0-08-100508-8.00009-6.

[103] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, ''Deep contextualized word representations,'' 2018, *arXiv:1802.05365*.

[104] C. Huang, A. Trabelsi, and O. R. Zaïane, ''ANA at SemEval-2019 task 3: Contextual emotion detection in conversations through hierarchical LSTMs and BERT,'' 2019, *arXiv:1904.00132*.

[105] N. Majumder, S. Poria, H. Peng, N. Chhaya, E. Cambria, and A. Gelbukh, ''Sentiment and sarcasm classification with multitask learning,'' 2019, *arXiv:1901.08014*.

[106] S. Zhang, X. Zhang, J. Chan, and P. Rosso, ''Irony detection via sentiment-based transfer learning,'' *Inf. Process. Manage.*, vol. 56, no. 5, pp. 1633–1644, 2019, doi: 10.1016/j.ipm.2019.04.006.

[107] G. Stanton and A. A. Irissappane, ''GANs for semi-supervised opinion spam detection,'' in *Proc. 28th Int. Joint Conf. Artif. Intell.*, 2019, pp. 5204–5210. [Online]. Available: https://learn.g2crowd.com/customer-reviews-statistics, doi: 10.24963/ijcai.2019/723.

[108] W. M. Kouw and M. Loog, ''An introduction to domain adaptation and transfer learning,'' 2018, *arXiv:1812.11806*.

[109] A. Chronopoulou, C. Baziotis, and A. Potamianos, ''An embarrassingly simple approach for transfer learning from pretrained language models,'' 2019, *arXiv:1902.10547*.

[110] S. Leggeri, A. Esposito, and L. Iocchi, ''Task-oriented conversational agent self-learning based on sentiment analysis,'' in *Proc. CEUR Workshop*, 2018.

[111] M. Atzeni and D. R. Recupero, ''Multi-domain sentiment analysis with mimicked and polarized word embeddings for human-robot interaction,'' *Future Gener. Comput. Syst.*, vol. 110, pp. 984–999, Sep. 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X19309719

[112] K. Qian and Z. Yu, ''Domain adaptive dialog generation via meta learning,'' 2019, *arXiv:1906.03520*.

[113] Y. Deldjoo, T. di Noia, and F. A. Merra, ''A survey on adversarial recommender systems: From attack/defense strategies to generative adversarial networks,'' *ACM Comput. Surveys*, vol. 54, no. 2, pp. 1–38, Apr. 2021, doi: 10.1145/3439729.

[114] I. Alsmadi, ''Adversarial machine learning in text analysis and generation,'' 2021, *arXiv:2101.08675*.

SHEETAL KUSAL received the Master of Engineering degree in computer from Dr. D. Y. Patil Institute of Engineering and Technology, Pune, Maharashtra, India. She is currently pursuing the Ph.D. degree with Symbiosis International (Deemed University). She is a full-time Research Scholar. She has 14 years of combined experience in academics, industry, and research. Her research interests include artificial intelligence, natural language processing, deep learning, sentiment analysis, emotion detection, and conversational agents.

SHRUTI PATIL received the M.Tech. degree in computer science and the Ph.D. degree in the domain of data privacy from Pune University. She has been an industry professional in the past, currently associated with the Symbiosis Institute of Technology, as a Professor, and with the SCAAI, Pune, Maharashtra, as a Research Associate. She has three years of industry experience and ten years of academic experience. She has expertise in applying innovative technology solutions to real world problems. She is currently working in the application domains of healthcare, sentiment analysis, emotion detection, and machine simulation via which she is also guiding several bachelor's, master's, and Ph.D. students as a domain expert. She has published more than 30 research papers in reputed international conferences and Scopus/Web of Science indexed journals, and books. Her research interests include applied artificial intelligence, natural language processing, acoustic AI, adversarial machine learning, data privacy, digital twin applications, GANS, and multimodal data analysis.

JYOTI CHOUDRIE currently holds the position of a Professor of information systems at the University of Hertfordshire. She has extensive years' experience specializing in investigating the social inclusion and adoption of information and communications technologies on society's marginal groups, the adoption, use and diffusion of innovative information and communication technologies in small to medium size enterprises and large organizations. This is based upon the principles and mechanisms of variables taken from the theories of diffusion, adoption, usage and implementation in the social, organizational and government realms and how they can be brought to fruition using modern internet related technologies; for instance, broadband, smartphones and online social networks to guide and improve individual's experiences of modern technology. This was achieved due to sponsored research funding schemes, such as the Royal Academy of Engineering, Microsoft, and Knowledge Transfer Partnerships and consultancy projects with organizations, such as British Telecom and AOL. To ensure that her expertise remains in the area, she has written for established journals, such as *European Journal of Information Systems* (EJIS) and *Journal of Information Technology*, and published a Routledge research monograph titled "Management of Broadband Technology Innovation." She is also one of the four Editor-in-Chief of *Information Technology & People* journal.

KETAN KOTECHA currently heads the Symbiosis Centre for Applied Artificial Intelligence (SCAAI). He has expertise and experience of cutting-edge research and projects in AI and deep learning for over last 25 years. He has published widely in several excellent peer-reviewed journals on various topics ranging from education policies, teaching-learning practices and AI for all. He is also a Team Member for the nationwide initiative on AI and deep learning skilling and research named Leadingindia.ai initiative sponsored by the Royal Academy of Engineering, U.K., under Newton Bhabha Fund. He is considered a foremost expert in AI and aligned technologies. Additionally, with his vast and varied experience in administrative roles, he has pioneered education technology. He worked as an Administrator previously at Parul University and Nirma University and has several achievements in these roles to his credit.

SASHIKALA MISHRA received the Ph.D. degree in the field of bioinformatics and data mining from Siksha' O' Anusandhan University, Bhubaneswar, Odisha, in 2015. She is currently working as an Associate Professor and a Researcher in the field of data analytics with the Symbiosis Institute of Technology, Pune. She has authored more than 43 international journals in almost all publishing houses, including IEEE, Elsevier, Inderscience, Springer, and ACM. She also filed three patents and published two IPR. She has also contributed book chapters in *Biological Knowledge Discovery Handbook* by Wiley Publisher. Her research interests include artificial intelligence, conservation biology, pricing theory, bio-informatics, data mining, image processing, and networking. Her main research interests lay with the design and implementation of algorithms related to prediction, classification, and clustering.

AJITH ABRAHAM (Senior Member, IEEE) received the Master of Science degree from Nanyang Technological University, Singapore, in 1998, and the Ph.D. degree in computer science from Monash University, Melbourne, VIC, Australia, in 2001. He is currently the Director of the Machine Intelligence Research Laboratories (MIR Laboratories), a Not-for-Profit Scientific Network for Innovation and Research Excellence Connecting Industry and Academia. The Network with HQ in Seattle, WA, USA, is also more than 1,500 scientific members from over 105 countries. He holds two university professorial appointments. He works as a Professor of artificial intelligence at Innopolis University, Russia, and the Yayasan Tun Ismail Mohamed Ali Professorial Chair in Artificial Intelligence at UCSI, Malaysia. He works in a multi-disciplinary environment. He has authored/coauthored more than 1,400 research publications out of which there are more than 100 books covering various aspects of computer science. One of his books was translated into Japanese and a few other articles were translated into Russian and Chinese. He has more than 46,000 academic citations (H-index of more than 102 as Per Google Scholar). He has given more than 150 plenary lectures and conference tutorials (in more than 20 countries). As an Investigator/a Co-Investigator, he has won research grants worth over more than 100 Million U.S.$. He was the Chair of the IEEE Systems Man and Cybernetics Society Technical Committee on Soft Computing (which has over more than 200 members), from 2008 to 2021, and served as a Distinguished Lecturer for the IEEE Computer Society representing Europe (2011–2013). He was the Editor-in-Chief of *Engineering Applications of Artificial Intelligence* (EAAI), from 2016 to 2021, and serves/served on the editorial board for over 15 international journals indexed by Thomson ISI.