Intent detection in AI chatbots: a comprehensive review of techniques and the role of external knowledge

Jemimah Kandaraj[1], Rajkumar Kannan[1], Frederic Andres[2]
[1]Research Department of Computer Science, Bishop Heber College (Auto), Affiliated to Bharathidasan University, Trichy, India
[2]Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo, Japan

Article Info

Article history:

Received Jun 24, 2024
Revised Jul 7, 2025
Accepted Aug 6, 2025

Keywords:

Dialogue system
External knowledge
Intent detection
Natural language processing
Natural language understanding

ABSTRACT

Artificial intelligence (AI) chatbots have become essential across various industries, including customer service, healthcare, education, and entertainment, enabling seamless, and intelligent user interactions. A key component of chatbot functionality is intent detection, which determines the underlying purpose of user queries to provide relevant responses. Traditional intent detection methods, such as rule-based and statistical approaches, often struggle with adaptability, especially in complex, dynamic conversations. This review examines the evolution of intent detection techniques, from early methods to modern deep learning and knowledge-enriched models. It introduces the domain type-conversation turns-adaptivity-external knowledge (DCAD) classification, highlighting its significance in improving chatbot accuracy and contextual awareness. The paper categorizes existing intent detection models, analyzes their applications across various sectors, and discusses key challenges, including data integration, language ambiguity, and ethical concerns. By exploring emerging trends and future directions, this review underscores the critical role of external knowledge in enhancing chatbot performance and user experience.

Corresponding Author:

Rajkumar Kannan
Research Department of Computer Science, Bishop Heber College (Auto)
Affiliated to Bharathidasan University
Trichy, India
Email: rajkumar@bhc.edu.in

1. INTRODUCTION

In today's digital era, chatbots have emerged as a remarkable evolution in human-computer interaction, bridging the gap between technology and communication. These conversational agents, powered by artificial intelligence (AI), have transitioned from simple rule-based systems to sophisticated tools capable of understanding, learning, and responding to human queries with unprecedented precision. However, with advancements in natural language processing (NLP), machine learning (ML), and neural networks, chatbots have undergone a revolutionary transformation [1].

Figure 1 shows the origin, advancements, and the transformative impact of chatbots, tracing their journey from rudimentary dialogue systems to intelligent conversational agents reshaping the digital world. From aiding customer service to enhancing education and streamlining healthcare, chatbots are now a cornerstone of user engagement, transforming user interactions across various fields [2]. The evolution of intent detection in chatbots has marked a significant departure from early rule-based systems, shifting towards sophisticated, AI-driven models. Initially, chatbots operated on simple, fixed rules and relied on keyword matching, limiting their flexibility and depth of understanding [3]. However, advances in NLP and ML have revolutionized this process, enabling chatbots to interpret more complex and nuanced user queries with higher accuracy [4].

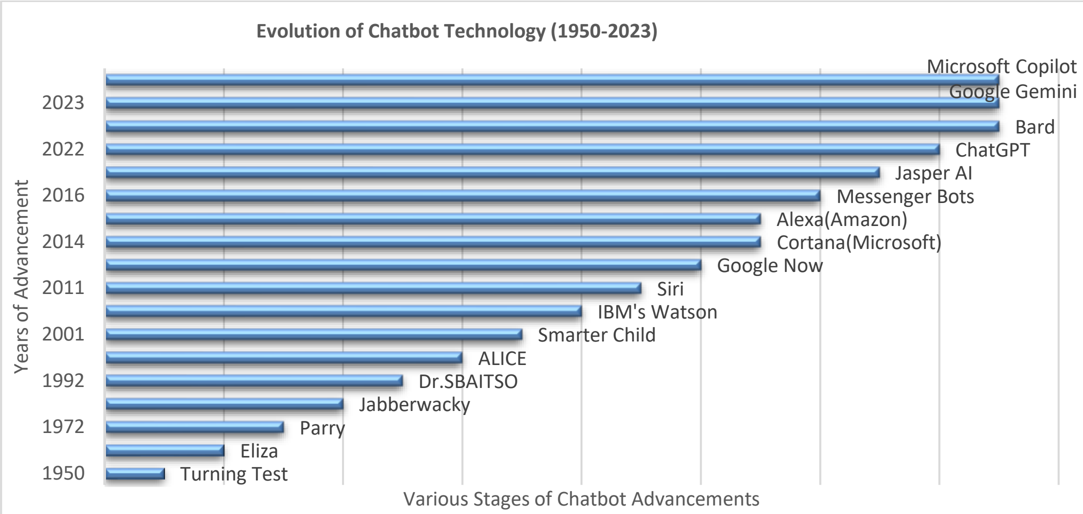

Figure 1. Evolution of chatbot technology

This review examines a range of intent detection techniques, from traditional approaches to the latest deep learning methods. Rule-based systems, although simple and easy to implement, often struggle with scalability and tend to falter when handling ambiguous queries. Statistical approaches such as bag-of-words (BoW) and term frequency-inverse document frequency (TF-IDF) represented a significant advancement by utilizing probabilistic techniques to enhance classification accuracy [5]. With the rise of ML methods like support vector machines (SVM) and decision trees, intent classification saw further enhancements. Deep learning techniques have brought the most substantial advances in intent detection. Neural network architectures, especially recurrent neural networks (RNNs) and their variants like long short-term memory (LSTM) networks [6], [7], have shown exceptional capability in handling sequential data and understanding context. Transformer-based models, such as bidirectional encoder representations from transformers (BERT) [8], [9], and generative pre-trained transformer (GPT), represent the leading edge in NLP, providing unparalleled accuracy and adaptability for intent detection tasks [10].

By exploring the applications of intent detection, this review underscores the broad impact of chatbots across different sectors. In customer service, chatbots enhance user experience by offering timely and accurate responses. In healthcare, they assist in preliminary diagnostics and patient interaction [11]. In education, chatbots facilitate personalized learning experiences, while in entertainment, they engage users with interactive storytelling and customized recommendations [12]. Despite these advancements, intent detection faces persistent challenges. Issues such as ambiguity in user queries, diverse language variations, and the need for ongoing adaptation present significant obstacles [13]–[15].

While various techniques for intent detection have been developed, a unified perspective that categorizes these methods based on essential features-such as domain specificity, conversational turns, adaptivity, and knowledge integration-is still lacking. This review introduces the domain type-conversation turns-adaptivity-external knowledge (DCAD) classification model, a framework that organizes chatbot models by these core features. DCAD aims to offer a structured approach to evaluating and categorizing intent detection techniques across diverse applications. This framework not only simplifies the process of comparing models but also highlights patterns and emerging trends, which are essential for researchers and practitioners looking to develop or refine chatbot systems.

The rest of the paper is structured as follows. In section 2, the review methodology for intent detection based DCAD classification of chatbots is discussed. In section 3, various techniques that are used in chatbot models are explained with research papers. In section 4, the results and discussion section compare the results of the existing methodologies and also discusses the challenges, advantages, limitations and implications in intent detection, followed by the dataset that is used in the research papers were discussed. Finally, in section 5, the conclusion is made.

2. REVIEW METHODOLOGY

Intent detection in chatbots can be analyzed across four key aspects: domain type, conversational type, adaptivity, and knowledge type. Collectively referred to as the DCAD classification, these aspects are essential to the intent detection process within chatbot technology. As depicted in Figure 2, the classification tree outlines 16 categories based on these feature combinations, facilitating easy categorization of chatbot models. Primarily used as a review approach, the DCAD classification provides readers with insight into how existing models are structured around specific features. Each of these features will be examined in detail in the following sections.

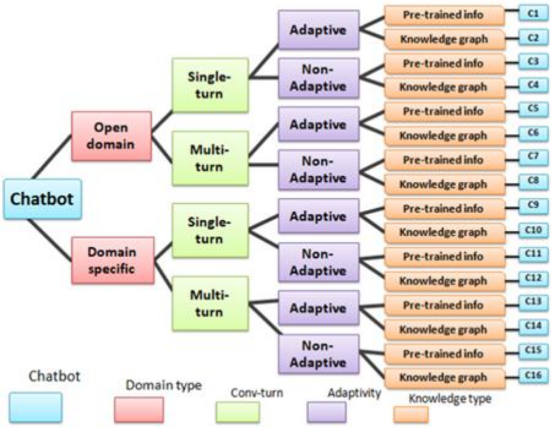

Figure 2. Intent detection based DCAD classification tree

2.1. Domain type in DCAD classification

Chatbots come under two types of domains, one is open domain and another is closed domain or domain specific. Most of the early research was mainly based on closed domain which became a drawback for the chatbot community. This feature will help to find chatbots which falls into the open domain and domain specific categories.

2.2. Conversation type in DCAD classification

Many personalized chatbots provide a multi-turn conversational experience to the user. Whereas, a minority of bots are used as information retrieval agents which give a single-turn conversational experience to the user. Recent advancement in chatbot technology has focued on this feature of multi-turn conversation.

2.3. Adaptivity type in DCAD classification

An adaptive bot is something that understands and works according to the users' needs and keeps track of the history context of the users' request. Only personalized bots can inherit adaptivity. A non-adaptive bot works on its own path and does not deviate according to users' request which is a major flaw in a personalized chatbot. User experience is a very crucial role in which adaptivity plays an important role.

2.4. Knowledge type in DCAD classification

Some chatbots are supported with an information base or knowledge base which contains information based on specific topics. Thus, the information retrieval happens based on the knowledge provided. In contrast, some bots do not use a separate knowledge base; instead, they work with a pre-trained data model which classifies the request within the trained features.

3. TECHNIQUES IN INTENT DETECTION

Intent detection is traditionally approached as a sentence-level classification problem, aiming to identify the underlying intent behind a user's utterance within the realm of NLP. A total of 56 research articles on intent detection were studied and reviewed. This section presents a summary of selected works, covering both traditional approaches and recent advancements in intent detection.

3.1. Overview of intent detection approaches

Before 2020, intent detection research primarily relied on traditional ML techniques, including k-nearest neighbors (KNN), SVM, and random forest (RF). In contrast, this article emphasizes recent advancements that utilize deep learning and transformer-based models for improved performance. It also addresses multilingual intent detection which uses multilingual text data [16]. Notably, category C14, which includes domain-specific combinational features for intent detection, shows a high number of successful models, as shown in Figure 3. This indicates that domain-specific features play a crucial role in enhancing user experience for intent detection tasks in many of the latest chatbot models.

3.1.1. Attention based intent detection

A unified CNN-based parallel architecture with cross-fusion and masking improved joint intent and slot prediction. Early approaches to intent classification for dialogue utterances employed a BoW model combined with naïve Bayes, followed by continuous-BoW integrated with SVM. The final classification was performed using LSTM networks [17]. To enhance natural language interpretation for customized digital assistants, adaptive intent detection in multi-domain environments (AidMe) was introduced. This system employed a semantic similarity evaluation model built on intent data to differentiate between known intents and autonomously identify new variations of existing intents [18].

A multimodal model for intent detection and slot filling was proposed, leveraging historical context and external knowledge within the framework of spoken language understanding (SLU). This model catered to specific task categories [19]. An open-domain chatbot was trained end-to-end, mapping input tags to optimized output sequence tags. This chatbot demonstrated the ability to perform common-sense reasoning during human interactions [20].

In the banking sector, a real-time chatbot was implemented to promote sustainable development by enhancing customer satisfaction. This system identified users' continuance intentions based on satisfaction, trust, and perceived usefulness, with trust exerting the strongest influence [21]. A contrastive learning-based task adaptation model was proposed for few-shot intent recognition, using a contrastive loss to enhance category separation and extract task-specific features within meta-tasks [22].

The chunk-level intent detection (CLID) framework uses a sliding window-based self-attention (SWSA) mechanism to detect chunk-level intents and intent transitions, segmenting utterances and predicting sub-utterance intents effectively [23]. A multi-turn dialog intent detection model was developed, incorporating historical information. Semantic data from both users and systems were combined using a self-attention network, while an LSTM captured historical context [24]. A meta-learning framework for zero-shot intent classification was presented, utilizing a hybrid attention mechanism that merges distributional signature attention with multi-layer perceptron (MLP)-based attention to effectively capture patterns of key word occurrences [25]. Finally, a multitask learning approach was proposed for multilingual intent detection and slot filling, handling inputs in English, Bengali, and Hindi using multilingual word embeddings (MWE) to jointly model both tasks across languages [16].

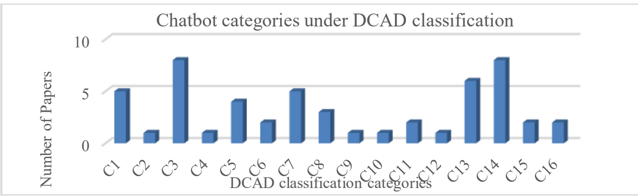

Figure 3. Chatbot categories with combinational features

3.1.2. Bi-directional models used in intent detection

A convolutional neural network-bidirectional gated recurrent unit (CNN-BGRU) model leveraging word embeddings was proposed for effective intent classification in dialogue systems [30]. Emotion and intention are two major factors which influence peoples' communication in their everyday life. A hierarchical variational model was developed to predict the emotions and intents in sequence and is used to convey the response according to the predicted emotion in an open domain system [31]. A two-stage triplet training framework using ratio-memory cellular nonlinear network (RMCNN) for structural features and a Siamese BERT encoder for intent detection improved multiclass classification [32]. A multi-task SLU model with bidirectional long short-term memory network (BiLSTM)/GRU and attention mechanisms enhanced intent detection, slot filling, and dialogue act classification with statistically validated results [33].

3.1.3. Memory network based intent detection

The memory network with patterns (MNP) model improves upon traditional memory networks by incorporating two distinct memory components. One dedicated to capturing contextual information from earlier dialogue turns, and another designed to store patterns identified by regular expressions. This dual-memory architecture allows the model to represent intent patterns more effectively through its specialized pattern memory, enhancing its overall capability to determine user intent [34].

3.1.4. Graph network based intent detection

A deep learning model using hierarchical BiLSTM with attention was proposed for text representation and similarity, enhanced by an external knowledge graph to retrieve information relevant to user queries [8],

[35]. A reinforcement learning-based approach was introduced to interpret intent in multi-turn dialogues by identifying paths in a knowledge graph, using a Markov decision process to generate interpretable intent trajectories [36]. A co-occurrence-enhanced ERNIE network was proposed for complex intent recognition in medical queries, leveraging concept co-occurrence patterns mined using the Apriori algorithm to improve contextual understanding [37]. A symptomatic assessment chatbot (SAC) was developed to raise awareness of breast cancer symptoms in women and to facilitate formal medical consultations, utilizing a hospital-based knowledge base from Taiwan [38]. Multi-scale dynamic convolutional network (M-DCN), a multi-scale dynamic convolutional network, enhances knowledge graph embedding by modeling complex relations using dynamic filters and alternating entity-relation compositions, outperforming baseline models across five link prediction benchmark datasets [39].

3.1.5. Low-resource based methods used in intent detection

A two-stage intent detection model was designed for handling complex user utterances, integrating sentence compression and intent classification within a multi-task learning framework [40]. An intent detection and entity extraction system was proposed for healthcare queries in Indic languages. The primary contribution includes the creation of a multilingual healthcare dataset comprising real-world hospital queries in English and six Indic languages (Hindi, Bengali, Tamil, Telugu, Marathi, and Gujarati), annotated with both query intents and entities [41].

A zero and few-shot intent identification was proposed using large language models (LLMs). It demonstrated four approaches namely domain adaptation, data augmentation, zero-shot prediction with prompting and parameter-efficient fine-tuning. Zero-shot and few-shot intent prediction completely remove or substantially reduce the work to provide intent-utterances, respectively [42]. An intent and entity detection approach were developed for a mental health virtual assistant chatbot, incorporating data augmentation to automatically generate training data. The model jointly addressed intent and entity detection by leveraging their interdependencies, rather than treating them as separate classification tasks [43].

A CK-Keras model using Word2Vec embeddings achieved the best F1-score for word-level language identification in Kannada-English code-mixed YouTube comments. This study explores Kannada-English word-level language identification using character- and word-based features in ML and deep learning models [44]. To address broader code-mixing issues, a hierarchical transformer (HIT) model with multi-headed self-attention and outer product attention was introduced, outperforming existing multilingual models across multiple languages and tasks [45], [46]. Furthermore, a Hindi-English parallel corpus and enhanced translation pipeline demonstrated improved machine translation performance for code-mixed texts, offering scalability to other language pairs [47].

Linguistic-adaptive retrieval-augmentation (LARA) a framework addresses multi-turn intent classification by integrating a fine-tuned model with retrieval-augmented LLMs, enabling context-aware predictions across six languages without extensive retraining [48]. To overcome data scarcity in mental health dialogue systems, the single-turn to multi-turn inclusive language expansion (SMILE) technique generates multi-turn conversations from single-turn data using ChatGPT. This led to the creation of the SMILECHAT dataset and the development of the MeChat chatbot, validated with a real-life counseling dataset [49].

3.2. Issues addressed and solutions proposed in intent detection

3.2.1. Data scarcity

A generalized zero-shot (GZS) classification model was introduced to address the problem of data scarcity for unseen intents by incorporating commonsense knowledge into the intent detection framework [13]. This approach enables the model to better understand and classify user intents, even when it has not encountered specific examples during training. Additionally, an automatic intent-slot induction method was proposed as a domain-independent technique that uses role labeling, concept mining, and pattern mining to identify both broad intent roles and more specific abstract concepts, making it adaptable to new domains [14]. These advancements help improve the flexibility and robustness of intent detection systems across a variety of application areas.

3.2.2. Emerging intent

Contextualized Deep SemSpace was introduced for intent detection by generating synset vectors with WordNet 3.1 and converting words from intent datasets into contextualized semantic vectors. This approach functions in a manner similar to deep contextual embeddings such as BERT, embeddings from language model (ELMo), and GPT-3, allowing for a richer representation of word meanings in context [10]. Additionally, an incremental intent detection method was proposed for the medical domain using Contrastive Replay Networks, which enable the model to learn new intents from user queries while effectively handling data imbalance and rare medical terminology [15]. These innovations enhance the adaptability to emerging intents and performance of intent detection systems, particularly in specialized domains.

3.2.3. Imbalanced data

Class lifelong learning-based intent detection (CLL-ID) enables continual learning of new intent classes while mitigating catastrophic forgetting, employing structure-based retrospection and contrastive knowledge distillation to address class imbalance [50]. Specialized BERT models trained on medical corpora outperformed general BERT in biomedical tasks. The models were evaluated on biomedical datasets, including informatics for integrating biology and the bedside, and drug-drug interaction datasets [51].

Table 1 describes the issues addressed and the methodology used to overcome the issues from various contributions under each DCAD classification with respect to various categories from 1 to 16. The table highlights only the latest contributions made in each categoriey, which reflects the most reliable model based on the features in DCAD classification to design and deploy a successful chatbot. The most recent works are under category 1, 10 and 13 which directly shows the chatbot models' recent works are based on adaptability as one of its important features.

Table 1. Methodology and addressed issues on each category

Paper	Addressed issues	Approaches
C1-open domain, single-turn, adaptive and pre-trained information		
Orhan *et al.* [10]	incrementally inferring new intent from utterances outside the existing set	Deep contrastive clustering algorithm
C2-open domain, single-turn, adaptive and knowledge graph		
Siddique *et al.* [13]	training data scarcity of unseen intents	Relationship meta-feature computation and integrates commonsense knowledge
C3-open domain, single-turn, non-adaptive and pre-trained info		
Abro *et al.* [9]	reduce the need for massive supervised data	Weighted finite state transducer (WFST)-BERT
C5-open domain, multi-turn, adaptive and pre-trained info		
Lair *et al.* [18]	learning new intents in customized assistants	Sentence similarity method
C6-open domain, multi-turn, adaptive and knowledge graph		
Wang *et al.* [19]	Limited ability of SLU models trained on large-scale corpus to understand multi-turn dialogue	History dialogue encoder and knowledge attention
C7-open domain, multi-turn, non-adaptive and pre-trained info		
Ren and Xu [32]	Enhance the ability of detecting intent from SLU using Triplet training and fusion strategy	RMCNN as Siamese encoders and BERT, ELMo and GPT for utterance representation
C8-open domain, multi-turn, non-adaptive and knowledge graph		
Yang *et al.* [36]	Lacks transparency in how intents are predicted in multi-turn dialogues	Policy-guided reinforcement learning and Markov Decision Process
C10- domain specific, single-turn, adaptive and knowledge graph		
Zhang *et al.* [37]	Detecting single-intent, multi-intent and implicit intent in patients' utterance	ERNIE with Apriori algorithm and Node2Vec
C11-domain specific, single-turn, non-adaptive and pre-trained info		
Bao *et al.* [35]	Answer complex medical questions	Hierarchical BiLSTM attention model (HBAM) with knowledge graph
C13-domain specific, multi-turn, adaptive and pre-trained info		
Parikh *et al.* [42]	Intent detection in low resource setting	Zero-shot learning with fine-tuning. Uses domain adaptation and data augmentation
C14-domain specific, multi-turn, adaptive and knowledge graph		
Xu *et al.* [6]	automatic medical diagnosis incorporating medical knowledge	Knowledge routed Deep Q-network

4. RESULTS AND DISCUSSION

The DCAD classification framework offers a structured approach to analyzing intent detection across diverse chatbot models. It categorizes models based on four key dimensions: domain specificity, conversational turns, adaptivity, and knowledge integration. This framework facilitates a nuanced evaluation of how these features impact chatbot performance and effectiveness. The classification results are visually represented in Figure 2, illustrating 16 distinct categories, each corresponding to a unique combination of the DCAD features.

4.1. Frequency analysis of feature combinations

The graphical representation indicates that certain feature combinations, particularly domain-specific models with integrated knowledge and adaptive capabilities, are more frequently used in recent research. For instance, category C14, which includes domain-specific features with enhanced knowledge integration, accounts for a significant portion of successful models. This highlights the importance of domain relevance and enriched knowledge sources in delivering accurate and context-sensitive responses for intent detection tasks.

4.2. Limitations and future directions

While the DCAD framework offers a comprehensive classification approach, there are limitations to consider. Some feature combinations, particularly those involving high adaptability or extensive knowledge integration, may require significant computational resources and advanced infrastructure. Furthermore, the framework currently relies on a static categorization, which may not capture the evolving nature of conversational AI as new techniques and hybrid models emerge. Future research should explore dynamic classification models that account for these evolving factors and investigate the applicability of underutilized feature combinations in different contexts. From the Table 2, BERT and variants of BERT performs better than the previous models but the combination of features matters while performing NLP task. As we categorized various combination of features category C10 and C13 has recent models in domain specific and C1 in open domain which shows the interests of building chatbot with these features plays major role in successful user interaction.

Table 2. Performance comparison table based on DCAD framework categories

Papers	Year	Model name	Accuracy
Siddique *et al.* [13]	2021	LSTM+Intent Encoders+ConceptNet	92.54
Xue and Ren [52]	2021	IE-BERTCapsNet	97.52
Kumar and Baghel [53]	2021	BERT+ELMo+BiLSTM	97.77
Sun *et al.* [24]	2022	Self-attention+LSTM	96.7
Abro *et al.* [9]	2022	WFST+BERT	98.12
Zhang *et al.* [22]	2022	BERT+Self Attn+ProtoNet	93.3
Huang *et al.* [23]	2022	Sliding Window+Self-Attn	94.7
Orhan *et al.* [10]	2023	Generalized SemSpace+BiLSTM	99.78
Firdaus *et al.* [16]	2023	mBERT+Attn+CRF	99.18
Zhang *et al.* [37]	2023	ERNIE+CRF+Attn	98.13
Babu and Boddu [11]	2024	BERT Fine-tuned	98
Huang *et al.* [54]	2024	DistillBERT+FAN	97.9
Wu *et al.* [55]	2024	CBLMA-B (CNN+BiLSTM+Multi-head Att+BERT)	94.9

4.3. Dataset

Table 3 presents a categorized overview of the datasets used in this study, offering valuable insights into their distribution and diversity across different intent detection scenarios. This categorization supports a comprehensive evaluation of the models, ensuring robust benchmarking. By detailing these datasets, the table underscores the relevance and applicability of the selected data to real-world intent detection tasks.

Table 3. List of datasets used in various categories

S. No	Datasets	Domain	Languages	Categories
1	ATIS, TRAINS, and FRAMES	Travel and Hotel	English	C3 and C7.
2	Askubuntu and WebApplications (Q&A)	Software Support	English	C3
3	SNIPS	Restaurant and Entertainment	English	C3, C6
4	ConceptNet	Commonsence	Multilingual	C6
5	Cornell Movie Dialogue Dataset	Movie Dialogue	English	C6
6	DailyDialogues Dataset	Daily Dialogue	English	C5 and C7
7	DX Medical Dialogue and MZ Dataset	Medical	English	C14
8	Facebook Multilingual Dialogue	Multi-Domains	Multilingual	C3 and C7.
9	KVRET (Key Value Retrieval)	In-Car Assistant	English	C6
10	MIT Dataset	Movie and Restaurant	English	C6
11	MSDialogue and Ubuntu Dialogue Corpus	Microsoft Products	English	C8
12	Mturk (Amazon)	Multi-Domain	Enlish	C5
13	MultiWOZ Dataset	Multi-Domain	English	C6, C8
14	SMP	Social Media	English	C3 & C7.
15	STS SemEval Dataset	News	English	C5
16	Movie Ticket Booking	Movie Booking	English	C13
17	Restaurant Reservation	Restaurant	English	C13
18	Taxi Ordering	Travel	English	C13
19	CMedQ	Medical Query	Chinese	C10
20	MedDialogue	Medical Dialogue	Chinese & English	C10

5. CONCLUSION

Chatbots are a great companion for people, providing satisfying responses to the user with a wealth of information. Intent detection and entity recognition is the core task which should be performed accurately. According to the intent the information is searched from the external knowledge and the appropriate response is sent. In this paper, some important intent detection tasks were discussed, followed by a brief history on chatbots' development. Four important features were studied which classifies the chatbot into various categories. Intent detection in chatbot usecases were also discussed which gave an idea on how intent detection plays a major role in real-time applications. DCAD classification show a clear picture of how a successful chatbot can be built with added features and also shows how recent models perform better whith respective features. This paper will help the readers to analyze the recent trends and areas in which intent detection can be made possible in real-time applications. It also helps the readers to know more about feature combinations in a chatbot design. In future, using the DCAD classification analysis, new intent detection models will be built in areas which are yet to come. Moreover, further features will also be added in the DCAD classification which will give some clarity for the budding researchers.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Jemimah Kandaraj	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Rajkumar Kannan		✓		✓	✓	✓				✓		✓	✓	
Frederic Andres	✓			✓		✓	✓		✓	✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] H. Abdulla, A. M. Eltahir, S. Alwahaishi, K. Saghair, J. Platos, and V. Snasel, “Chatbots development using natural language processing: a review,” in *2022 26th International Conference on Circuits, Systems, Communications and Computers (CSCC)*, 2022, pp. 122–128, doi: 10.1109/CSCC55931.2022.00030.
[2] V. Gupta, V. Joshi, A. Jain, and I. Garg, “Chatbot for mental health support using NLP,” in *2023 4th International Conference for Emerging Technology*, 2023, pp. 1–6, doi: 10.1109/INCET57972.2023.10170573.
[3] A. Mccallum and K. Nigam, “A comparison of event models for naive Bayes text classification,” in *Proceedings in Workshop on Learning for Text Categorization, AAAI’98, AAAI*, 1998, pp. 41–48.
[4] J. Weizenbaum, “ELIZA—a computer program for the study of natural language communication between man and machine,” *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, Jan. 1966, doi: 10.1145/365153.365168.
[5] J. Schuurmans and F. Frasincar, “Intent classification for dialogue utterances,” *IEEE Intelligent Systems*, vol. 35, no. 1, pp. 82–88, Jan. 2020, doi: 10.1109/MIS.2019.2954966.
[6] L. Xu, Q. Zhou, K. Gong, X. Liang, J. Tang, and L. Lin, “End-to-end knowledge-routed relational dialogue system for automatic diagnosis,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 7346–7353, Jul. 2019, doi: 10.1609/aaai.v33i01.33017346.
[7] N. Vedula, N. Lipka, P. Maneriker, and S. Parthasarathy, “Towards open intent discovery for conversational text,” *arXiv*, Apr. 2019.
[8] Z. Zhang, Z. Zhang, H. Chen, and Z. Zhang, “A joint learning framework with BERT for spoken language understanding,” *IEEE Access*, vol. 7, pp. 168849–168858, 2019, doi: 10.1109/ACCESS.2019.2954766.
[9] W. A. Abro, G. Qi, M. Aamir, and Z. Ali, “Joint intent detection and slot filling using weighted finite state transducer and BERT,” *Applied Intelligence*, vol. 52, no. 15, pp. 17356–17370, Dec. 2022, doi: 10.1007/s10489-022-03295-9.
[10] U. Orhan, E. G. Tosun, and O. Ozkaya, “Intent detection using contextualized deep SemSpace,” *Arabian Journal for Science and Engineering*, vol. 48, no. 2, pp. 2009–2020, Feb. 2023, doi: 10.1007/s13369-022-07016-9.
[11] A. Babu and S. B. Boddu, “BERT-based medical chatbot: enhancing healthcare communication through natural language understanding,” *Exploratory Research in Clinical and Social Pharmacy*, vol. 13, 2024, doi: 10.1016/j.rcsop.2024.100419.
[12] T. A. I. T. Baha, M. El Hajji, Y. Es-Saady, and H. Fadili, “Towards highly adaptive Edu-Chatbot,” *Procedia Computer Science*, vol. 198, pp. 397–403, 2022, doi: 10.1016/j.procs.2021.12.260.
[13] A. B. Siddique, F. Jamour, L. Xu, and V. Hristidis, “Generalized zero-shot intent detection via commonsense knowledge,” in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021, pp. 1925–1929, doi: 10.1145/3404835.3462985.
[14] Z. Zeng, D. Ma, H. Yang, Z. Gou, and J. Shen, “Automatic intent-slot induction for dialogue systems,” in *Proceedings of the Web Conference 2021*, Apr. 2021, pp. 2578–2589, doi: 10.1145/3442381.3450026.
[15] G. Bai, S. He, K. Liu, and J. Zhao, “Incremental intent detection for medical domain with contrast replay networks,” in *Findings of the Association for Computational Linguistics: ACL 2022*, 2022, pp. 3549–3556, doi: 10.18653/v1/2022.findings-acl.280.

[16] M. Firdaus, A. Ekbal, and E. Cambria, “Multitask learning for multilingual intent detection and slot filling in dialogue systems,” *Information Fusion*, vol. 91, pp. 299–315, 2023, doi: 10.1016/j.inffus.2022.09.029.
[17] A. Bhasin, B. Natarajan, G. Mathur, J. H. Jeon, and J.-S. Kim, “Unified parallel intent and slot prediction with cross fusion and slot masking,” *Natural Language Processing and Information Systems (NLDB 2019)*, Cham, Switzerland: Springer, 2019, pp. 277–285, doi: 10.1007/978-3-030-23281-8_22.
[18] N. Lair, C. Delgrange, D. Mugisha, J.-M. Dussoux, P.-Y. Oudeyer, and P. F. Dominey, “User-in-the-loop adaptive intent detection for instructable digital assistant,” in *Proceedings of the 25th International Conference on Intelligent User Interfaces*, Mar. 2020, pp. 116–127, doi: 10.1145/3377325.3377490.
[19] Y. Wang, T. He, R. Fan, W. Zhou, and X. Tu, “Effective utilization of external knowledge and history context in multi-turn spoken language understanding model,” in *2019 IEEE International Conference on Big Data (Big Data)*, Dec. 2019, pp. 960–967, doi: 10.1109/BigData47090.2019.9006162.
[20] S. S. Abdullahi, S. Yiming, A. Abdullahi, and U. Aliyu, “Open domain chatbot based on attentive end-to-end Seq2Seq mechanism,” in *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*, Dec. 2019, pp. 339–344, doi: 10.1145/3377713.3377773.
[21] D. M. Nguyen, Y.-T. H. Chiu, and H. D. Le, “Determinants of continuance intention towards Banks’ chatbot services in Vietnam: a necessity for sustainable development,” *Sustainability*, vol. 13, no. 14, p. 7625, Jul. 2021, doi: 10.3390/su13147625.
[22] X. Zhang, F. Cai, X. Hu, J. Zheng, and H. Chen, “A contrastive learning-based task adaptation model for few-shot intent recognition,” *Information Processing & Management*, vol. 59, no. 3, p. 102863, May 2022, doi: 10.1016/j.ipm.2021.102863.
[23] H. Huang, P. Huang, Z. Zhu, J. Li, and P. Lin, “CLID: a chunk-level intent detection framework for multiple intent spoken language understanding,” *IEEE Signal Processing Letters*, vol. 29, pp. 2123–2127, 2022, doi: 10.1109/LSP.2022.3211156.
[24] X. Sun, H. Zheng, and Z. Tang, “Historical information-based intent detection for multiturn dialogue,” in *Proceedings of the 8th International Conference on Computing and Artificial Intelligence*, Mar. 2022, pp. 566–572, doi: 10.1145/3532213.3532299.
[25] H. Liu *et al.*, “A simple meta-learning paradigm for zero-shot intent classification with mixture attention mechanism,” in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Jul. 2022, pp. 2047–2052, doi: 10.1145/3477495.3531803.
[26] M. Firdaus, A. Kumar, A. Ekbal, and P. Bhattacharyya, “A multi-task hierarchical approach for intent detection and slot filling,” *Knowledge-Based Systems*, vol. 183, 2019, doi: 10.1016/j.knosys.2019.07.017.
[27] T.-E. Lin and H. Xu, “A post-processing method for detecting unknown intent of dialogue system via pre-trained deep neural network classifier,” *Knowledge-Based Systems*, vol. 186, p. 104979, Dec. 2019, doi: 10.1016/j.knosys.2019.104979.
[28] H.-M. Chuang and D.-W. Cheng, “Conversational AI over military scenarios using intent detection and response generation,” *Applied Sciences*, vol. 12, no. 5, p. 2494, Feb. 2022, doi: 10.3390/app12052494.
[29] M. H. Maqbool, M. Fereidouni, A. B. Siddique, and H. Foroosh, “Model-agnostic zero-shot intent detection via contrastive transfer learning,” *International Journal of Semantic Computing*, vol. 18, no. 01, pp. 5–24, Mar. 2024, doi: 10.1142/S1793351X24410010.
[30] Y. Wang, J. Huang, T. He, and X. Tu, “Dialogue intent classification with character-CNN-BGRU networks,” *Multimedia Tools and Applications*, vol. 79, no. 7–8, pp. 4553–4572, Feb. 2020, doi: 10.1007/s11042-019-7678-1.
[31] Y. Li, R. Zhang, W. Li, and Z. Cao, “Hierarchical prediction and adversarial learning for conditional response generation,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, pp. 314–327, Jan. 2022, doi: 10.1109/TKDE.2020.2977637.
[32] F. Ren and S. Xue, “Intention detection based on Siamese neural network with triplet loss,” *IEEE Access*, vol. 8, pp. 82242–82254, 2020, doi: 10.1109/ACCESS.2020.2991484.
[33] M. Firdaus, H. Golchha, A. Ekbal, and P. Bhattacharyya, “A deep multi-task model for dialogue act classification, intent detection and slot filling,” *Cognitive Computation*, vol. 13, no. 3, pp. 626–645, May 2021, doi: 10.1007/s12559-020-09718-4.
[34] W. A. Abro, G. Qi, H. Gao, M. A. Khan, and Z. Ali, “Multi-turn intent determination for goal-oriented dialogue systems,” in *2019 International Joint Conference on Neural Networks*, Jul. 2019, pp. 1–8, doi: 10.1109/IJCNN.2019.8852246.
[35] Q. Bao, L. Ni, and J. Liu, “HHH: an online medical chatbot system based on knowledge graph and hierarchical bi-directional attention,” in *Proceedings of the Australasian Computer Science Week Multiconference*, Feb. 2020, pp. 1–10, doi: 10.1145/3373017.3373049.
[36] K. Yang, X. Kong, Y. Wang, J. Zhang, and G. De Melo, “Reinforcement learning over knowledge graphs for explainable dialogue intent mining,” *IEEE Access*, vol. 8, pp. 85348–85358, 2020, doi: 10.1109/ACCESS.2020.2991257.
[37] B. Zhang, Z. Tu, S. Hang, D. Chu, and X. Xu, “Conco-ERNIE: complex user intent detect model for smart healthcare cognitive bot,” *ACM Transactions on Internet Technology*, vol. 23, no. 1, pp. 1–24, Feb. 2023, doi: 10.1145/3574135.
[38] M. D. I.-Manzano, N. V. López, N. A. González, and C. C. Rodríguez, “Implementation of chatbot in online commerce, and open innovation,” *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 7, no. 2, p. 125, Jun. 2021, doi: 10.3390/joitmc7020125.
[39] Z. Zhang, Z. Li, H. Liu, and N. N. Xiong, “Multi-scale dynamic convolutional network for knowledge graph embedding,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 5, pp. 2335–2347, May 2022, doi: 10.1109/TKDE.2020.3005952.
[40] S. Song, X. Chen, C. Wang, X. Yu, J. Wang, and X. He, “A two-stage user intent detection model on complicated utterances with multi-task learning,” in *Companion Proceedings of the Web Conference 2022*, Apr. 2022, pp. 197–200, doi: 10.1145/3487553.3524232.
[41] A. Mullick, I. Mondal, S. Ray, R. R, G. Chaitanya, and P. Goyal, “Intent identification and entity extraction for healthcare queries in Indic languages,” in *Findings of the Association for Computational Linguistics: EACL 2023*, 2023, pp. 1870–1881, doi: 10.18653/v1/2023.findings-eacl.140.
[42] S. Parikh, M. Tiwari, P. Tumbade, and Q. Vohra, “Exploring zero and few-shot techniques for intent classification,” in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 5: Industry Track)*, 2023, pp. 744–751, doi: 10.18653/v1/2023.acl-industry.71.
[43] A. Zamani *et al.*, “Intent and entity detection with data augmentation for a mental health virtual assistant chatbot,” in *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*, Sep. 2023, pp. 1–4, doi: 10.1145/3570945.3607324.
[44] O. E. Ojo, A. Gelbukh, H. Calvo, and A. Feldman, “Language identification at the word level in code-mixed texts using character sequence and word embedding,” in *Proceedings of the 19th International Conference on Natural Language Processing*, 2022, pp. 1–6.
[45] T. Suresh, A. Sengupta, M. S. Akhtar, and T. Chakraborty, “A comprehensive understanding of code-mixed language semantics using hierarchical transformer,” *IEEE Transactions on Computational Social Systems*, vol. 11, no. 3, pp. 4139–4148, Jun. 2024, doi: 10.1109/TCSS.2024.3360378.
[46] A. Sengupta, S. K. Bhattacharjee, T. Chakraborty, and M. S. Akhtar, “HIT - a hierarchically fused deep attention network for robust code-mixed language representation,” in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2021, pp. 4625–4639, doi: 10.18653/v1/2021.findings-acl.407.

[47] V. Srivastava and M. Singh, "PHINC: a parallel Hinglish social media code-mixed corpus for machine translation," in *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, 2020, pp. 41–49, doi: 10.18653/v1/2020.wnut-1.7.
[48] J. Liu, T. Y. Keat, B. Fu, and K. H. Lim, "LARA: linguistic-adaptive retrieval-augmentation for multi-turn intent classification," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, 2024, pp. 1096–1106, doi: 10.18653/v1/2024.emnlp-industry.82.
[49] H. Qiu, H. He, S. Zhang, A. Li, and Z. Lan, "SMILE: single-turn to multi-turn inclusive language expansion via ChatGPT for mental health support," in *Findings of the Association for Computational Linguistics: EMNLP 2024*, 2024, pp. 615–636, doi: 10.18653/v1/2024.findings-emnlp.34.
[50] Q. Liu *et al.*, "Class lifelong learning for intent detection via structure consolidation networks," in *Findings of the Association for Computational Linguistics: ACL 2023*, 2023, pp. 293–306, doi: 10.18653/v1/2023.findings-acl.20.
[51] S. Ayanouz, B. A. Abdelhakim, and M. B. Ahmed, "Using natural language processing to evaluate the impact of specialized transformers models on medical domain tasks," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 2, p. 1732, Jun. 2024, doi: 10.11591/ijai.v13.i2.pp1732-1740.
[52] S. Xue and F. Ren, "Intent-enhanced attentive BERT capsule network for zero-shot intention detection," *Neurocomputing*, vol. 458, pp. 1–13, Oct. 2021, doi: 10.1016/j.neucom.2021.05.085.
[53] N. Kumar and B. K. Baghel, "Smart stacking of deep learning models for granular joint intent-slot extraction for multi-intent SLU," *IEEE Access*, vol. 9, pp. 97582–97590, 2021, doi: 10.1109/ACCESS.2021.3095416.
[54] L. Huang, S. Liang, F. Ye, and N. Gao, "A fast attention network for joint intent detection and slot filling on edge devices," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 2, pp. 530–540, Feb. 2024, doi: 10.1109/TAI.2023.3309272.
[55] T. Wu, M. Wang, Y. Xi, and Z. Zhao, "Intent recognition model based on sequential information and sentence features," *Neurocomputing*, vol. 566, p. 127054, Jan. 2024, doi: 10.1016/j.neucom.2023.127054.

BIOGRAPHIES OF AUTHORS

Jemimah Kandaraj pursuing Ph.D. in Computer Science at Bishop Heber College (Auto), Affiliated to Bharathidasan University, Trichy, India. She holds a Master of Philosophy Degree from Bharathidasan University, Trichy. She received her M. Sc. and B.Sc. in Computer Science degrees from Bishop Heber College (Auto). Her research interest includes natural language processing, chatbot technologies and low resource languages. She has attended many international conferences and workshops organized by ACM India. She can be contacted at email: jemimahk.cs.res@bhc.edu.in.

Rajkumar Kannan holds a Ph.D. in Computer Applications from the National Institute of Technology, Tiruchirappalli, India (2007). He is currently the Controller of Examinations and Professor in the Department of Computer Science (Shift-I) at Bishop Heber College (Autonomous), India. Additionally, he leads the Artificial Intelligence Research and Technology Development Lab at the institution. His research focuses on information retrieval, multimedia semantics, data science, and collaborative and collective intelligence. Previously, His was associated with King Faisal University, Kingdom of Saudi Arabia, where he taught ABET-accredited undergraduate and postgraduate programs in Computer Science and Information Systems. He is the author of the widely used textbook Java Programming (Pearson) and has edited volumes for Springer LNCS and IGI Global (USA). His is the Program Chair of the International Conference on Data Engineering and Management series and serves on editorial boards of several international journals. He is a senior member of ACM (USA), life member of CSI, ISTE, and IFIP WG 8.2, and a former member of IEEE and BCS (UK). He can be contacted at email: rajkumar@bhc.edu.in.

Frederic Andres received his Ph.D. in Information Systems from the University of Paris VI "Pierre et Marie Curie" in 1993, followed by his Dr. Habilitation in Informatics from the University of Nantes in 2000. Since 2000, he has been an associate professor at the Digital Content and Media Sciences Research Division, NII, Tokyo, Japan, and since 2002, with Sokendai (Graduate University for Advanced Studies), Tokyo. He has authored over 170 publications across international journals, books, and conferences. His research interests collective intelligence-based community management, early warning systems, mulsemedia, cooking recipe computing, carbon footprint optimization in multi-dish planning, and water crystal computing. He and his team earned the 1st prize in the IEEE Brain Data Bank Challenges at COMPSAC 2018. His serves as co-editor-in-chief of the International Journal of Agricultural and Environmental Information Systems alongside colleagues from AIT and JAIST. He co-organizes the Decor Workshop series at ICDE conferences globally. As a member of Japan's ISO/IEC JTC 1/SC 36 Committee, he has contributed as editor to key standards on student data privacy and semantic interoperability in e-learning.. He can be contacted at email: andres@nii.ac.jp.