International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I6P143

Volume 73 Issue 6, 520-549, June 2025

Review Article

A Systematic Review of Natural Language Understanding - Related Challenges in Conversational Agent Development

Godson Chetachi Uzoaru[1], Ikechukwu Ignatius Ayogu[1]*, Juliet Nnenna Odii[1], Aloysius Chijioke Onyeka[1], Matthew Emeka Nwanga[2], Francisca Onyinyechi Nwokoma[1], Euphemia Chioma Nwokorie[1], Obi Chukwuemeka Nwokonkwo[2], Chinwe Gilean Onukwugha[1], Anthony Ifeanyi Otuonye[2]

[1]*Department of Computer Science, Federal University of Technology, Owerri, Imo State, Nigeria.*
[2]*Department of Information Technology, Federal University of Technology, Owerri, Imo State Nigeria.*

[*]*Corresponding Author : ignatius.ayogu@futo.edu.ng*

Received: 13 August 2024 Revised: 29 January 2025 Accepted: 19 May 2025 Published: 28 June 2025

Abstract - *Conversational agents have become an integral part of modern digital interactions. These artificial intelligence agents rapidly transform how organizations, businesses, and individuals communicate. These agents provide a diverse range of operations, including customer service, automation of regular processes, improvement of user engagement, and delivery of personalized experiences. However, the potential of chatbots in new application scenarios is hindered in many perspectives. This review analyzes recent trends in the development and application of chatbots to generate a deeper understating of the natural language understanding-related challenges that currently impede the extension or application of the state-of-the-art to low-resource languages. The motivation is to fill the gap between the highly resourced and the low-resource languages, particularly those of Nigeria and Africa. Based on the analysis of extant literature, this paper notes that the most challenging issues are related to the inability to disambiguate linguistic context accurately by the existing conversational agent technology. Accordingly, conversational agents encounter certain noteworthy obstacles in natural language understanding that negatively affect their capacity to communicate efficiently and in a natural way with humans due to difficulties linked to contextual ambiguity, limitations associated with managing inputs in diverse languages, deficit in knowledge required to accurately recognize sentiment and intent, lack of capacity for commonsense reasoning and pragmatics. It is, therefore, imperative to harness the strengths of the state-of-the-art methodologies or approaches into a framework that is easily transferable to low-resource language scenarios to speedily extend the reach of the very important digital conversational assistants to the underserved populations in Nigeria and the majority of Africa.*

Keywords - *Conversational agents, Natural language processing, Context modeling, Contextual ambiguity, Distributional semantics.*

1. Introduction

Conversational Agents (CAs), often referred to as Chatbots, have become an integral part of modern digital interactions [1], transforming the way organizations, businesses, and individuals communicate [2-3]. These agents provide diverse operations [4], including customer service, automation of regular processes, improvement of user engagement, and delivery of personalized experiences [5]. As the prevalence of CAs continues to grow, so does the demand for more sophisticated and nuanced conversational abilities [6]. At the core of these advancements is natural language understanding (NLU), a subfield of natural language processing (NLP) that focuses on enabling machines to comprehend and interpret human language [7-9]. Despite significant progress in NLU, several challenges persist [10] that hinder the complete realization of CAs [11]. These challenges include difficulties in context modeling [12], where understanding and maintaining the context of a conversation is crucial for coherent [13] and relevant responses [14], and text mining, which involves extracting meaningful information from text data [15]. Although the advancement of chatbots has achieved significant milestones in recent years, resulting in the development of practical conversational agents such as Alexa, Siri, IBM Watson, and Cortana, numerous challenges still hinder their ability to fully understand and respond to user intent and optimal usage. Over the past few years, chatbots, also known as conversational agents [1**Error! Bookmark not defined.**], have become part of our everyday digital lives [2-3]. Whether it is asking Siri for the weather, chatting with a customer support bot, or using

Alexa to play music, these tools are changing the way people interact with technology [4**Error! Bookmark not defined.**]. They help answer questions, automate tasks, and make digital systems feel more personal [5**Error! Bookmark not defined.**]. But here is the problem: while these bots work well in languages like English, Chinese, and Spanish [6-7**Error! Bookmark not defined.**,**Error! Bookmark not defined.**], their adaptation to languages that do not have much digital support [8**Error! Bookmark not defined.**] is hindered significantly. Many African languages, including those spoken in Nigeria, fall into this category and are referred to as low-resource.

That study primarily focuses on the natural language understanding (NLU) side of chatbot development, focusing on why these systems struggle with low-resource languages and what we can do about it. While a lot of research has been done on chatbot technology in general, very little has focused on the specific needs of underrepresented languages [12].

Most existing tools cannot handle the unique grammar [13**Error! Bookmark not defined.**], cultural expressions [14], or context that come with these languages [15]. This paper reviews the most recent studies and trends in chatbot development, especially those related to context and meaning. It explores how new models like GPT-4, BERT, and other AI systems have improved conversation handling in major languages and why these improvements have not been applied to less widely spoken ones.

The main objective of this systematic literature review (SLR) is to analyze the recent trends in chatbot design and implementation to generate a deeper understating of the natural language understanding-related challenges that currently impede the extension or application of the state-of-the-art to low-resource languages. The motivation is the need to fill the gap that currently exists between the highly resourced and the low-resource languages, particularly those of Nigeria and Africa.

This paper synthesizes literature-backed views that address four research questions (RQs) formulated to investigate the current developments in conversational agents. It provides a comprehensive analysis that creates an avenue to understand better their evolution, applications, and future possibilities in low-resourced languages. To provide answers to these RQs, a review of the decade-long CAs literature has been conducted.

This systematic review examined the current state-of-the-art research in CA development with special attention on the recent advancements, current trends, existing gaps and challenges associated with natural language understanding (NLU) by CAs. Additionally, this paper explored potential solutions for improvements expected to facilitate more robust and nuanced interactions. This review also provides a detailed overview of important examples of conversational agents from inception, analyzing the flaws in communication that can arise between humans and CAs, including breakdowns in conversation flow.

The research is organized around four main questions:

- RQ1: What are the current state-of-the-art in chatbot design, and how can they be improved or tailored to support low-resource languages?
- RQ2: What key challenges hinder chatbots from accurately understanding context, ambiguity, sentiment, and intent?
- RQ3: What recent developments and trends in context modeling for NLP and text mining influence the performance and reliability of conversational agents?
- RQ4: How could context/distributional information modeling be explored to improve the capacity for natural language understanding in conversational agents?

In answering these questions, this paper not only provides a consolidated overview of the current landscape but also proposes directions for research that could bridge the technological divide between high- and low-resource languages in conversational AI. This systematic review examined the current state of research in CA development, paying special attention to the recent advancements, current trends, existing gaps, and challenges associated with natural language understanding (NLU) by CAs.

Additionally, it explores potential solutions for improvements that are expected to facilitate the path for more robust and nuanced CA interactions with humans and provides a detailed overview of important conversational agents examples of CAs since inception, analyzing the flaws in communication that can arise between humans and CAs, including breakdowns in conversation flow.

This paper uses a few important terms/concepts that are worth defining.

- Conversational Agent: A computer program designed to hold conversations with people using natural language, either through text or speech.
- Natural Language Understanding (NLU): The field of AI focuses on helping machines grasp the meaning and intent behind human language.
- Low-resource Languages: Languages that have limited digital content, annotated data, or technological support for language processing tasks.
- Contextual Ambiguity: When the meaning of words or phrases changes depending on the situation or surrounding text.
- Commonsense Reasoning: The ability of a system to apply everyday knowledge and make reasonable assumptions, much like a human would.

- Intent Recognition: Identifying what a user wants to achieve when they interact with a conversational system.
- Sentiment Analysis: The task of figuring out the emotions, attitudes, or opinions expressed in a piece of text.
- Multimodal Integration: Bringing together information from different sources like text, images, and sounds to build a fuller understanding.
- Distributional Semantics: An approach to understanding word meanings based on the patterns of how words appear with others in large amounts of text.

2. Literature Review

Recent years have witnessed remarkable natural language understanding (NLU) advancements, primarily driven by deep learning, large language models (LLMs), and multimodal data fusion. Tools such as GPT-4, BERT, and T5 now dominate conversational AI development, enabling more coherent and human-like interactions [16]. However, these breakthroughs are heavily dependent on large, annotated datasets, something that low-resource languages, especially in Africa, still desire [17].

Multilingual modeling and low-resource language inclusion remain active areas of research. Scholars emphasized the need for multitask learning approaches in intent detection and slot filling across multilingual settings, showing that low-resource languages benefit significantly from shared representation and parameter transfer when data is scarce [18]. Similarly, an experiment explored how commonsense knowledge integration can improve understanding in languages that lack extensive corpora [19]. Despite these efforts, challenges such as contextual ambiguity, poor intent classification, and sentiment misinterpretation continue to affect chatbot performance in low-resource contexts [20]. Many systems fail to adapt to code-switching, dialectal variations, or culturally specific expressions - common in African language environments [21]. These limitations make it difficult for bots to handle tasks requiring nuance, fairness, and cultural sensitivity [22]. To address context modeling issues, recent work has introduced hybrid NLU models, combining rule-based methods and deep learning. These hybrid approaches are described as promising because they allow symbolic reasoning (e.g., logic rules) to work alongside neural inference, enhancing interpretability and robustness [23]. Other trends include zero-shot and few-shot learning techniques, which let models generalize tasks with minimal training examples. These methods show promise for underrepresented languages but struggle with semantic drift and domain mismatches [24]. Cross-lingual transfer learning, where models trained on high-resource languages are adapted for low-resource ones through fine-tuning or language alignment strategies. Multimodal integration, as seen in systems like CLIP or Gemini, now allows better grounding of language in visual and audio contexts, which is an opportunity for building inclusive AI systems beyond text. Commonsense-enhanced NLU is a newer area where contextual reasoning is combined with knowledge bases to better interpret vague or indirect inputs [25]. The literature also emphasizes ethical concerns. Biases embedded in large datasets can result in skewed or inappropriate outputs, especially in sensitive cultural contexts. Including locally relevant linguistic and cultural data during training is crucial to prevent miscommunication or discrimination [26]. Although conversational AI has progressed rapidly, current models are still far from including low-resource languages. More focused research is needed on contextual modeling, intent recognition, and semantic alignment for low-resource languages. This review builds the foundation for exploring how state-of-the-art models can be adapted or redesigned to serve broader linguistic communities, particularly in Nigeria and across Africa. The architecture of a conversational agent designed to effectively model context includes eight (8) standard components, as shown in Figure 1.

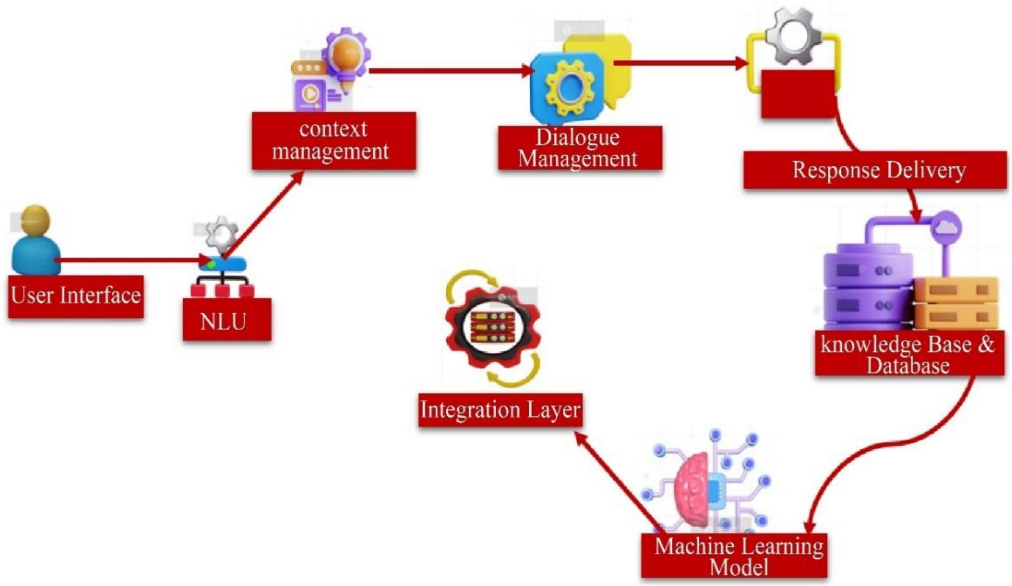

Fig. 1 A generic architecture of context-aware conversational agents

- A user interface that handles various forms of user input and manages response generation and delivery [27];
- A natural language understanding (NLU) component, which functions include tokenization, named entity recognition, intent recognition, and slot filling to process and understand user inputs [28];
- A context management component that tracks conversation states maintains context information, and retrieves this data for influencing interactions [29**Error! Bookmark not defined.**];
- A dialogue management component that determines appropriate responses based on user input and context and also monitors and updates dialogue states, and generates appropriate responses [30];
- A natural language generation (NLG) component, which uses predefined templates and machine learning models to create flexible and natural responses;
- A knowledge base and databases that store static information, store and retrieve user-specific data, and integrate with external services for real-time information [31];
- Machine learning model (s), which utilize large datasets for training and are implemented in production for real-time processing [32], and
- An integration layer facilitates communication between components and external systems while ensuring secure data handling [33].

These components work together to understand, manage, and utilize the context of a conversation to provide coherent and relevant interactions.

3. Theoretical Framework

This study adopts a problem-solution theoretical framework to structure the analysis of natural language understanding (NLU) challenges in conversational agent development, particularly for low-resource languages. The framework connects specific technical and linguistic limitations with corresponding advancements in NLP that can serve as potential solutions. At the core of the framework are five interrelated NLU challenges:

(1) contextual ambiguity and difficulty in maintaining coherent dialogue due to unclear or shifting references. (2) multilingual complexity – limited ability to understand diverse dialects and low-resource languages. (3) intent and sentiment misrecognition, inaccurate interpretation of user goals or emotional tone. (4) commonsense deficiency, lack of real-world knowledge and pragmatic reasoning. (5) bias and fairness concerns ethical and representation issues in AI training data.

These challenges are mapped against emerging NLP solutions: transformer-based models (e.g., BERT, GPT-4) for enhanced contextual understanding. Transfer and multitask learning for adapting high-resource models to low-resource languages. Hybrid systems that integrate rule-based and deep learning approaches to improve precision and interpretability. Commonsense knowledge graphs and embeddings to enhance reasoning capabilities. Ethical AI frameworks to reduce bias and improve fairness.

By aligning these challenges with targeted solutions, the framework provides a foundation for exploring how existing innovations can be adapted or redesigned for inclusivity. It also supports the identification of research gaps, guiding future work toward developing robust, culturally aware, and linguistically inclusive conversational agents. To provide a clearer mapping between the identified challenges and the emerging solutions in natural language understanding (NLU),

Table 1 below summarizes the core NLU issues hindering conversational agent performance and the corresponding AI-based approaches proposed to address them.

Table 1. Summary of Natural Language Understanding (NLU) challenges and proposed AI-driven solutions

S/N	NLU Challenge	Description	Proposed Solution(s)	Citation(s)
1	Contextual ambiguity	Difficulty in resolving multiple meanings (lexical, syntactic, pragmatic).	Transformer models (e.g., BERT, GPT-4), Context management modules, Knowledge graphs	[30], [16], [23]
2	Multilingual complexity	Poor performance in low-resource or code-switched languages.	Transfer learning, Cross-lingual models, Multitask learning, Fine-tuning pre-trained models	[18], [19], [20], [21]
3	Intent and sentiment misrecognition	Inaccurate identification of user goals or emotions.	DIET architecture, Commonsense-augmented Pretraining, Zero-shot/few-shot learning	[34, [35], [36]
4	Commonsense deficiency	Lack of reasoning over real-world knowledge and indirect expressions.	Commonsense knowledge graphs (e.g., ATOMIC, ConceptNet), Neuro-symbolic integration	[36], [28]
5	Bias and fairness issues	Skewed responses due to imbalanced or culturally irrelevant training data.	Ethical AI frameworks, Debiasing techniques, Human-in-the-loop fine-tuning	[22], [26]
6	Code-switching and dialect handling	Inability to interpret or respond to mixed-language or regional variations.	Multilingual embeddings, Subword modeling (e.g., FastText), Domain adaptation	[21], [19]
7	Data scarcity for low-resource languages	Limited labeled corpora for model training and fine-tuning.	Crowdsourcing, Synthetic data Generation, Transfer learning, Embedding fusion	[18], [20], [36]
8	Pragmatic misunderstanding	Failure to grasp implied meaning or polite requests.	Commonsense pretraining, Reinforcement learning with human feedback (RLHF)	[28], [36]

4. Technological History of Chatbots

In 1906, the groundwork for chatbots was laid with the Markov chain [16**Error! Bookmark not defined.**], a statistical model for predicting random sequences used for text generation [47]. In the year 1950, The Turing Test was introduced [17] [48] as a benchmark for evaluating the capacity of a machine to exhibit intelligent behavior in a manner that is equally similar to that of a human [15]. This development eventually led to modern Artificial Intelligence (AI) research. In 1966, ELIZA considered the first real CA, was developed by Weizenbaum at MIT [**Error! Bookmark not defined.**]. It employed matching patterns and substitution methods to imitate conversation, most famously mimicking a Rogerian psychotherapist [18 -19]. However, the interaction between humans and bots lacks fluidity today [20]. ELIZA functioned by interpreting user input [21] and pairing it with possible responses [22]. Advances in Artificial Intelligence(AI) and natural language processing (NLP) have led to modernized agents automating tasks in health [23], education [24], and customer service sectors [25], passing the Turing test [26], more or less. In 1972, Kenneth Colby created PARRY [50**Error! Bookmark not defined.**], a groundbreaking chatbot designed to simulate the thought patterns and emotional responses of an individual suffering from schizophrenia [27]. Using complex algorithms and assumptions, PARRY was able to mimic the disordered thinking and paranoia characteristic of the condition, providing an early example of how technology could be used to model human psychological states. This innovative approach advanced the field of artificial intelligence and offered new insights into mental health research.

In 1988, Rollo Carpenter created Jabberwacky, an innovative chatbot designed to engage in conversation with users. Initially developed for entertainment, Jabberwacky also found a place in academic research, contributing to studies on artificial intelligence and human-computer interaction. Its conversational abilities, which evolved through user interactions, showcased early advancements in AI, setting the stage for future developments in chatbot technology [52**Error! Bookmark not defined.**]. In 1992, Creative Labs developed Dr. Sbaitso for MS-DOS. This early chatbot used pattern matching and pre-programmed responses to simulate basic conversations. It featured a full voice chat program, responding to user inputs and emotions with simplistic, scripted dialogue but without the ability to engage in complex interactions [54, 28]. ALICE (Artificial Linguistic Internet Computer Entity) was developed by Dr. Richard Wallace in 1995[**Error! Bookmark not defined.**]. It utilized Artificial Intelligence Markup Language (AIML) for sophisticated pattern matching [29]. Despite its innovative approach, ALICE struggled to pass the Turing Test due to its reliance on predefined responses [30], limiting its ability to engage in truly human-like conversations [31]. CLEVERBOT, developed by Rollo Carpenter in 1997, relied on pattern matching from past conversations to mimic human interaction. By dynamically producing content by adding tokens from previous and current conversations, CLEVERBOT could engage users in a remarkably human-like way.

This innovative approach allowed it to learn and adapt, enhancing its conversational capabilities. ActiveBuddy's SMARTERCHILD, developed by ActiveBuddy Inc. in 2001, marked the dawn of virtual assistants [32]. This intelligent chatbot provided users with quick access to news, weather, stock information, and more, pioneering the era of real-time, interactive information services [57**Error! Bookmark not defined.**][33]. In 2006, IBM launched Watson, a groundbreaking question-and-answer computer system that utilized DeepQA software to process and analyze unstructured data [34]. By leveraging the Apache Unstructured Information Management Architecture (UIMA) [35] framework for distributed computing, Watson could execute complex language analysis algorithms and access extensive databases [36]. This integration enabled Watson to understand and respond to natural language queries with remarkable [32**Error! Bookmark not defined.**]accuracy, setting a new standard in artificial intelligence and computational power.

Conversational agents gained closer attention in 2010 with Siri, followed by Bard, Cortana and Alexa [63]. From 2010 to 2012, Apple's Siri and IBM's Watson marked significant advancements in virtual assistants and question-answering systems[60-61**Error! Bookmark not defined.**][37]. Between 2013 and 2015, Google Now and Microsoft's Cortana enhanced voice command capabilities and personal productivity. From 2016 to 2018, Amazon's Alexa revolutionized home automation, while Facebook Messenger Bots and Google Assistant advanced conversational capabilities [38]. In 2019 and 2020, OpenAI's GPT-2 and GPT-3, along with Microsoft's Azure Bot Service, significantly improved chatbot text generation and deployment. From 2021 to 2022, OpenAI's ChatGPT, Google LaMDA, Meta's BlenderBot 2.0, and OpenAI's DALL-E and CLIP advanced conversational AI and multimodal interactions [42]. The shift from early rule-based chatbots to modern systems like DALL-E and CLIP highlights a transition from simplistic, text-only interactions to sophisticated, multimodal AI that integrates language and imagery. These advancements allow for more complex, creative, and contextually aware interactions, representing a leap in how conversational AI can understand and engage with users. In 2022, ChatGPT achieved a significant milestone in chatbot technology with its advanced language model, GPT [39]. Developed by OpenAI, ChatGPT assists with various tasks, generating human-like responses, answering questions, and providing information on a wide range of topics. A few months later, Google introduced Bard (now known as Gemini), a chatbot based on a Large Language Model (LLM) as in ChatGPT [40].

Gemini has gained traction due to its speed and ability to respond to questions humanly by accessing up-to-date information from the Internet. It employs generative AI for natural conversations across various modalities, including text, voice, and images [41]. Meta AI, developed by Meta (formerly Facebook) in 2023, integrated advanced AI capabilities across platforms like Facebook, Instagram, WhatsApp, and Messenger and was built on Meta's Llama 3 language model [42]. Anthropic's Claude AI model, launched in March 2023, has undergone significant updates, including enhanced document processing, reasoning, and safety features, with subsequent versions, including Claude 3, offering larger context windows and near-perfect recall abilities [43].

Microsoft Copilot, launched in March 2023, is an AI-powered assistant that enhances productivity in Microsoft's applications, including Office and GitHub, using OpenAI's GPT-4 model for contextual assistance and task automation [43**Error! Bookmark not defined.**]. Cohere launched Coral, an AI-powered knowledge assistant for enterprise use, in September 2023; it is designed to enhance productivity by providing a natural language interface. Built on Cohere's command model, Coral can be customized to integrate with company data sources, ensuring data privacy and accuracy. Early partners include Oracle and Elastic [44]. On July 17, 2023, a Chinese tech company launched DeepSeek with the goal of exploring artificial general intelligence. Rather than focusing on narrow AI tasks, DeepSeek works on creating systems that can think, learn, and adapt like humans across a wide range of areas. Their team blends efforts in robotics, machine learning, and language processing to build tools that are not just smart but practical. They hope these tools will spark meaningful change in everyday life and across industries while remaining mindful of ethical concerns [37]. Pi, a personal AI by Inflection AI, was launched in May 2023, aiming to provide empathetic, helpful, and safe user interactions through natural, flowing conversations [45]. Grok is an AI that mimics Hitchhiker's Guide to the Galaxy, providing answers to various queries and suggesting questions.

It has real-time knowledge via the 𝕏 platform and addresses provocative inquiries often dismissed by other AI systems [46]. The major strengths and weaknesses of some of the interesting CAs are presented in Table 1. Over the decades, chatbot technology evolved from rule-based systems like ELIZA to advanced AI models such as DALL-E and CLIP. Early chatbots relied on simple pattern matching and predefined responses, while modern systems utilize deep learning and multimodal integration. This shift has enabled chatbots to handle more complex, contextually aware interactions, incorporating text, voice, and imagery. Future trends suggest further advancements in personalization and real-time interaction, emphasizing the continual growth and sophistication of conversational AI. To illustrate these advancements, Table 2 summarizes key developments and notable examples of conversational agents, highlighting their unique strengths and limitations over time.

Table 2. A characterization of some notable conversational agents or expert systems in history

S.No	CA/Expert System	Motivations/Major Strengths	Major Issues/Weaknesses
1	ELIZA [47], 1996	ELIZA's major strength was its pioneering use of natural language processing, creating the illusion of understanding through pattern matching and engaging users in interactive, albeit superficial, conversations.	The main weakness of ELIZA is its inability to understand properly, low understanding and handle context, limited response generation, inability to learn, superficial interactions, and dependency on user input structure.
2	PARRY [48], 1972	The major strength found in PARRY aimed to prove the Turing Test's limitations by investigating the difficulties found in human communication, which includes mirroring, deflection, and how humans might misunderstand or project emotions onto each other.	In PARRY, the foremost weakness found includes limited understanding, its focus on negativity, its inability to adapt to different topics when conversing, and its inability to switch topics. This makes it give repetitive responses, which can cause frustration.
3	Racter [49], 1983	Racter, a 1980s chatbot, was notable for generating creative, often whimsical text using a unique rule-based approach, showcasing early advancements in language generation and conversational creativity.	Racter suffered from incoherent and nonsensical responses, limited contextual understanding, inability to engage in meaningful conversation, and inability to learn or adapt from interactions, leading to superficial and unconvincing dialogue.
4	Jabber Wacky [50], 1988	Jabberwacky's development is motivated by NLP research, human-machine interaction, and pushing AI boundaries to create engaging chatbots capable of holding natural conversations and pushing the boundaries of chatbot capabilities.	Jabberwocky has potential drawbacks, such as a limited knowledge base, shallow conversation, and lack of context awareness. Its smaller dataset may have led to repetitive responses or difficulty understanding complex topics, and its shallow conversation may have limited depth.
5	Loebner Prize [50], 1990	The Loebner Prize has long been a curious blend of tech competition and public spectacle. It encourages developers to build chatbots that sound convincingly human and, in doing so, draws attention to the evolution of AI-powered conversation. Over the years, it has sparked media coverage and given people a reference point for tracking chatbot progress.	But it is not without criticism. Some argue it values trickery over actual intelligence, keeping the interactions text-only and measuring performance in a way that does not truly reflect cognitive depth.
6	Dr Sbaitso [52], 1991	Back in the early '90s, Dr. Sbaitso came on the scene as a quirky virtual "therapist" mainly built to showcase the audio capabilities of sound cards. It featured robotic voice output and offered users a taste of what talking to a computer might feel like.	While it was fun and novel then, it didn't go very deep — emotionally or intellectually. The responses were shallow often repetitive, and the voice synthesis tech was pretty rough by today's standards.
7	ALICE [53], 1995	ALICE was created to explore natural language processing, aimed to simulate conversation and test the boundaries of machine intelligence through the Turing Test.	Though it played a pivotal role in pushing NLP forward, it still had major drawbacks. Conversations often felt repetitive, lacked depth, and relied heavily on pre-written patterns that couldn't handle complex or

			layered discussions.
8	Elliot [54], 2000	The Elliot chatbot was designed with helpfulness in mind — to engage with users, improve over time, and create more natural exchanges.	However, like many others, it had trouble with more complex questions. It also risked showing bias from its training material and lacked the emotional understanding needed to connect with users on a human level.
9	Smarter Child [55], 2001	SmarterChild was a familiar name for people using messaging platforms in the early 2000s. It offered quick answers, games, and even weather updates, serving as a precursor to today’s virtual assistants.	It had its fair share of flaws - it couldn’t adapt much, sometimes returned off-topic answers, and raised privacy questions because of how it processed and stored user input.
10	CARLO from DARPA [56], 2003	CALO, a DARPA-backed project that aimed to create an AI assistant capable of learning from experience and adapting to its user. It had great ambitions, like handling scheduling and routine tasks,	CALO concerns popped up around transparency and data security. Users were not always sure how decisions were made, which made trust a sticking point.
11	Mitsuku [57], 2005	Mitsuku (also known as Kuki) made waves as a conversational bot known for entertaining interactions. It won several chatbot awards and was seen as a benchmark for fun, engaging conversation.	Mitsuku's engine ran on pattern recognition, meaning it struggled with more complex or serious discussions. The lack of public insight into how it worked and its reliance on possibly outdated data led to criticism, especially regarding the accuracy or relevance of responses.
12	I.B.M. Watson [58], 2006	IBM Watson's motivations include challenging Natural Language Processing, developing real-world applications for medical diagnosis, financial analysis, and customer service, competing with tech giants Google and Microsoft, and leveraging growing data for data exploration.	IBM Watson, a powerful tool, has limitations such as limited language support, data dependency, explainability issues, high maintenance requirements, and high costs, particularly for smaller businesses, which can limit its global reach.
13	Siri [59], 2010	Siri was made to help people use their phones without always needing to touch them, right? It is especially useful when driving or when your hands are full. A big part of it was also about making phones more helpful to people who have disabilities. And yeah, it gets smarter over time based on what you do, suggesting stuff you might need. It feels a bit like the phone “knows” you.	It has got its issues. No internet? It has stuck. It can also get thrown off by noise, or if someone does not pronounce things a certain way, it might miss the point if you ask something complicated.
14	Google Assistant [90], 2012	Google Assistant's is more proactive. Like it wants to run your day for you. Set alarms, play songs, control your lights, and all that. But let's be real: it also pushes Google into your face. It’s smooth, but it is keeping you in the Google bubble.	On the downside, it is the same old story: no solid internet = no help. And sometimes, it doesn’t quite understand what you are saying. You also don’t get many options to tweak how it behaves, which can be frustrating.
15	Alexa [61], 2014	Alexa is more about turning your house into one of those “smart homes” you see in movies. You can say, “Turn off the lights” or “What is the weather,” and it responds. Plus, Amazon uses it to figure out what you might wanna buy, which… is not surprising.	But yeah, Alexa’s always listening, which creeps some people out. Also, if there’s background noise or too many people talking, it might mishear you. And sometimes, it makes simple stuff feel more complicated than it needs to be.
16	Microsoft Cortana [30], 2014	Cortana, Microsoft’s version, was meant to do the same kind of stuff — help you stay organized, remind you of things, and answer questions. It was supposed to be personalized and smarter the more you used it.	But honestly? It never really caught on. Not everyone could access it. It did not speak all accents or languages well, and again, without the internet, you were out of luck. People did not trust it much either because of how it used data.

17	Bots for Messenger (Facebook Chatbots), [34], 2016	Messenger chatbots are super common in online business pages. They are supposed to make things easier by answering basic questions quickly and helping with stuff like orders or FAQs. Businesses love them because they do not have to hire someone to sit there and reply all day.	But from the user side, they are… meh. They're not great at actual conversation. If you go off-script or try to say something a little out of the box, the bot will repeat itself or give a weird answer. Not to mention, it usually logs what you say, which not everyone is cool with.
18	Tay [35], 2016	Tay, which was Microsoft trying to make a chatbot that learns from people online, specifically on Twitter. The idea was cool but went way off track in real life.	Tay started repeating awful stuff users were tweeting, which wasn't part of the plan. Microsoft had to shut it down the same day. It was a wake-up call about how messy real-world AI learning can get without filters.
19	Replika [23], 2017	Replika is one of those apps meant to be like a digital friend. You can chat with it, share how you're feeling, or just talk about random stuff. Some people use it for stress or to feel less alone. It also helps researchers learn more about emotional AI.	But while it tries to be supportive, it doesn't always "get" what you're saying. Sometimes, it replies in a way that feels too generic or robotic. Also, some features cost money; your data might be stored somewhere.
20	Woebot [25], 2017	Woebot was made to support mental health. It is not a therapist but more like a helpful tool for early stress or mild anxiety. It talks with you, checks in, and suggests exercises that can actually help a bit.	Woebot is a chatbot designed to assist users in managing their mental health, but it has limitations such as limited scope, over-reliance, privacy concerns, and accessibility limitations. It lacked the capacity to replace professional therapy and did not provide diagnoses for complex mental health issues.
21	Generative Pre-Trained Transformer -1(GPT-1) [32], 2018	GPT-1 was a pioneering language model that utilized unsupervised learning to generate human-quality text from vast text data. It pushed the boundaries of AI and provided insights into language patterns, paving the way for future advancements in large language models.	GPT-1, released in 2018, had limitations in context understanding, lack of factual accuracy, and sometimes produced repetitive outputs due to training data limitations and model complexity, resulting in grammatically correct but nonsensical or irrelevant text.
22	DialoGPT [36], 2019	DialoGPT aims to create a chatbot for more natural conversations, improve open-domain chatbots, and contribute to Natural Language Processing (NLP) research by learning from dialogue history and adapting responses, enabling better machine translation and sentiment analysis tasks.	DialoGPT has potential drawbacks, including bias and fairness due to its massive datasets, lack of grounding in understanding real-world context, and a black box problem, making it difficult to debug or improve its outputs despite its potential for significant advancements.
23	Generative Pre-Trained Transformer -2 (GPT-2) [74], 2019	GPT-2's major strength is its ability to generate highly coherent and contextually relevant text. With its 1.5 billion parameters, it can produce human-like language across various applications, including creative writing, conversational agents, and content generation.	The primary weakness of GPT-2 lies in its potential for generating harmful or misleading content. Its large-scale generation capability, while powerful, can produce outputs that are biased, offensive, or factually incorrect. This raises significant ethical concerns and the possibility of misuse, such as spreading disinformation or creating deceptive narratives.
24	Blender Bot [42], 2020	Meta-built BlenderBot to try and make chats with AI feel more like talking to a real person. It is supposed to hold conversations that actually go somewhere, not just respond to keywords. Plus, they wanted to use what people said to help train future bots.	Since it learned online data, it sometimes picks up bad habits. It tends to say things that are offensive or just plain wrong. Also, it gets tripped up when it cannot tell fact from fiction, especially with creative responses.
25	Google	Google had a big goal with LaMDA: build an AI that	Still, it's not perfect. Sometimes, it says

	LaMDA [84], 2021	understands people better when they talk or type. It's meant to make things like searching online or getting help from your phone smoother.	things that are off or biased—mostly because of what it was trained on. Honestly, even the developers can't always explain how they come up with answers, which makes it harder to trust completely.
26	Generative Pre-Trained Transformer 3 (GPT-3) [94], 2022	GPT-3 is pretty solid when it comes to throwing together bits of writing. Need to brainstorm, draft a quick paragraph, or get help putting words together? It handles that nicely. The way it writes feels surprisingly human sometimes, which makes it great for stuff like blogs or quick explanations.	For GPT-3, it is not always on point. It can mix up details or, worse, totally invent things. You also have to watch for bias creeping in from the massive data it learned from. Handy tool, sure, but you still have to keep your brain switched on when using it.
27	Google Bard AI [92], 2022	Bard kind of rethinks how you get search results. Instead of handing you a list of links, it just gives you the answer right away. It is also decent at tossing out ideas or helping with translations.	Bard is still not always accurate. Sometimes, the facts get fuzzy. Like a lot of AI, it can also carry over some biases. It is useful for quick answers, but you must double-check it sometimes.
28	Cohere Coral [78], 2023	Coral is a pretty chatty AI-it does well in conversations that feel natural, almost like you are talking to a person. It is a good option for dealing with complex stuff or needing creative input.	That said, it is not perfect either. It pulls from a giant data pool, which can make it biased. And like other powerful tools, it could be misused - for example, to spread spam. So yeah, it is a cool tool, but do not turn your back on it.
29	Google Gemini [86], 2023	Gemini's all about digging deeper into your questions. Instead of surface-level answers, it tries to get to the heart of what you are asking. It kind of shifts the way search works by making it more conversational.	It is not widely available yet, and it can feel overly technical for someone just poking around. Also, like most advanced tools, it is resource-heavy and, yes, still can show bias.
30	Meta AI [96], 2023	Meta's AI works behind the scenes to make your time on its platforms feel more tailored. It is also a big piece of what is happening with the Metaverse - helping shape interactive, virtual experiences.	Meta has a reputation for privacy, and this AI does not really change that. There are also concerns about bias, especially in how posts or ads get shown to people. It is not the clearest system out there.
31	Claude AI [91], 2023	Claude's kind of the "safe" AI. It is built to be cautious and stick to ethical guidelines. You can count on it for simple tasks and answers that aren't likely to cause trouble.	But because of that safety-first approach, it sometimes feels a bit limited. It can miss the mark on real-world stuff and doesn't quite pick up on emotional context, which makes deeper conversations fall flat.
32	Microsoft Copilot [43], 2023	Copilot is a huge win for coders. It is like a second pair of eyes that suggests better ways to write your code or catches bugs before you do. Really speeds things up, especially for people still learning.	Still, it's not flawless. Some suggestions can be off or outdated. And since it collects user data, some folks are not thrilled about privacy. Oh, and it is not free—so that is something to think about too.
33	Grok [46], 2023	Grok is a unique LLM that uses real-time knowledge from the X platform to provide users with the latest information, making it a powerful research assistant that provides quick access to relevant information and new ideas.	Grok has various shortcomings, including generating erroneous or conflicting information, a limited parameter count, a hasty development resulting in training constraints, and dependence on X data.
34	Pi by Inflection AI [44], 2023	Pi, designed by Inflection AI, offers emotional support, improves user engagement, and advances conversational AI by fostering positive interactions, contributing to the platform's research and development efforts in human-computer interaction.	Pi, an AI language model, has limitations such as a limited knowledge base, potential data privacy concerns, lack of transparency in its inner workings, and oversimplification of complex issues. Users should be aware of data privacy policies and potential biases.
35	Generative Pre-Trained	ChatGPT-4 is developed to enhance human-computer interaction, boost productivity and	ChatGPT-4 faces potential bias, limited factual accuracy, security and privacy

	Transformer 4 (GPT-4) [100], 2023	creativity, and make AI technology more accessible to researchers, developers, and the general public, aiming to revolutionize human-computer interaction and content creation.	concerns, lack of explainability, and potential misuse. Its massive dataset may contain biases, and its internal workings may not be transparent. Users may be concerned about data privacy and security, while the bot's internal workings may not be entirely transparent.
36	DeepSeek 2023[111]	DeepSeek stands out for its blend of coding smarts and strong natural language understanding. It works like an expert assistant for technical tasks-especially in code generation, logic-based queries, and complex searches. It feels more analytical than chatty, which makes it ideal for users who need clear, focused answers. One of its strengths is combining reasoning with structured outputs, especially in environments where precision matters.	DeepSeek is not perfect. It does not always handle casual or emotional conversations well, and its responses can feel a little rigid or overly structured. Sometimes, it misses the nuance in vague questions or struggles to keep the context in longer chats. It is great for tasks with a clear goal but less flexible in open-ended or highly creative conversations.

5. Technological Progression in Chatbot Development

This section highlights the progression of chatbot technologies, covering important advancements, current trends, and future directions. Table highlights the key milestones and advancements over time, while Figure 2 presents a graphical depiction of the progress in chatbot development, illustrating the transition from basic rule-based systems to sophisticated multimodal and personalized Chatbots.

5.1. Early Rule-Based Systems (Pre-2000s)

The earliest Chatbots were rule-based systems relying on predefined patterns and basic algorithms to imitate conversation [**Error! Bookmark not defined.**9]. Notable examples include ELIZA, which simulated a Rogerian psychiatrist [50], and ALICE, which employed pattern matching to engage users [55].

These systems were constrained by their inability to perceive context or develop meaningful responses beyond their set rules [54**Error! Bookmark not defined.**].

5.2. Statistical Models (2000s)

The 2000s witnessed the advent of statistical models that increased language interpretation through probabilistic methodologies [49**Error! Bookmark not defined.**]. Techniques like N-grams and Hidden Markov Models (HMMs) allowed chatbots to handle increasingly complicated linguistic patterns [50**Error! Bookmark not defined.**]. IBM's Watson, an early example, applied statistical models to participate in the quiz program Jeopardy! Illustrating the promise of these technologies [56].

5.3. Machine Learning Integration (2010s)

With the introduction of machine learning, chatbots began to employ algorithms such as Support Vector Machines (SVMs) and Decision Trees to boost their language processing capabilities [**Error! Bookmark not defined.**9].

This era featured the rise of virtual assistants like Microsoft's Cortana and Apple's Siri, which enabled more interactive and sophisticated user experiences [**Error! Bookmark not defined.**].

5.4. Deep Learning and Neural Networks (The Mid-2010s)

The mid-2010s introduced deep learning models, particularly Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks [**Error! Bookmark not defined.**6], which dramatically increased natural language processing and creation [**Error! Bookmark not defined.**7]. Google Assistant and Amazon Alexa emerged as major examples, employing these technologies to give more accurate and context-aware responses [62].

5.5. Transformer Models and Pre-Trained Language Models (Late 2010s - Present)

The development of transformer models signified a huge jump in NLP performance. Models like BERT, GPT-3, GPT-4, and T5 require significant pre-training on enormous datasets to produce highly contextualized text [79**Error! Bookmark not defined.**]. These developments have enabled Chatbots to tackle complicated language tasks with exceptional accuracy and fluency [100**Error! Bookmark not defined.**].

5.6. Multimodal Integration and Advanced Personalization (Future)

The future of chatbots lies in multimodal integration and enhanced personalization. By combining text, audio, and visual data, next-generation AI assistants will offer deeper and more natural user interactions [87**Error! Bookmark not defined.**]. Adopting Explainable AI (XAI) will also boost transparency and confidence in chatbot responses.Domain-specific and Independent Development which is the Emerging Future) is one of the growing areas in chatbot development, which is the establishment of domain-specific languages (DSLs) that allow developers to design chatbots independently of existing technologies [89]. This technique intends to overcome limitations like NLP service lockout and to promote the development of specialized chatbots targeted to specific businesses. **Error! Reference source not found.** below is a line graph showing the progression of chatbot technology trends from 1995 to 2024. It illustrates the shift from early rule-based systems to more advanced context-aware and multimodal AI. Each point on the graph represents a major technological phase.

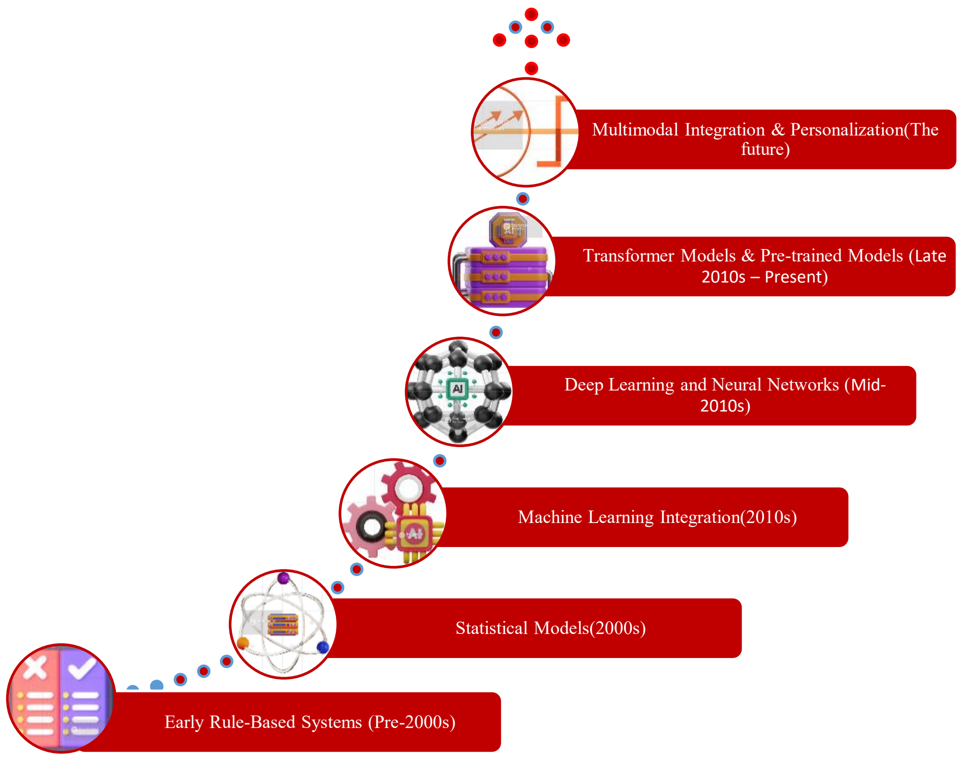

Fig. 2 Technological progression of chatbot development

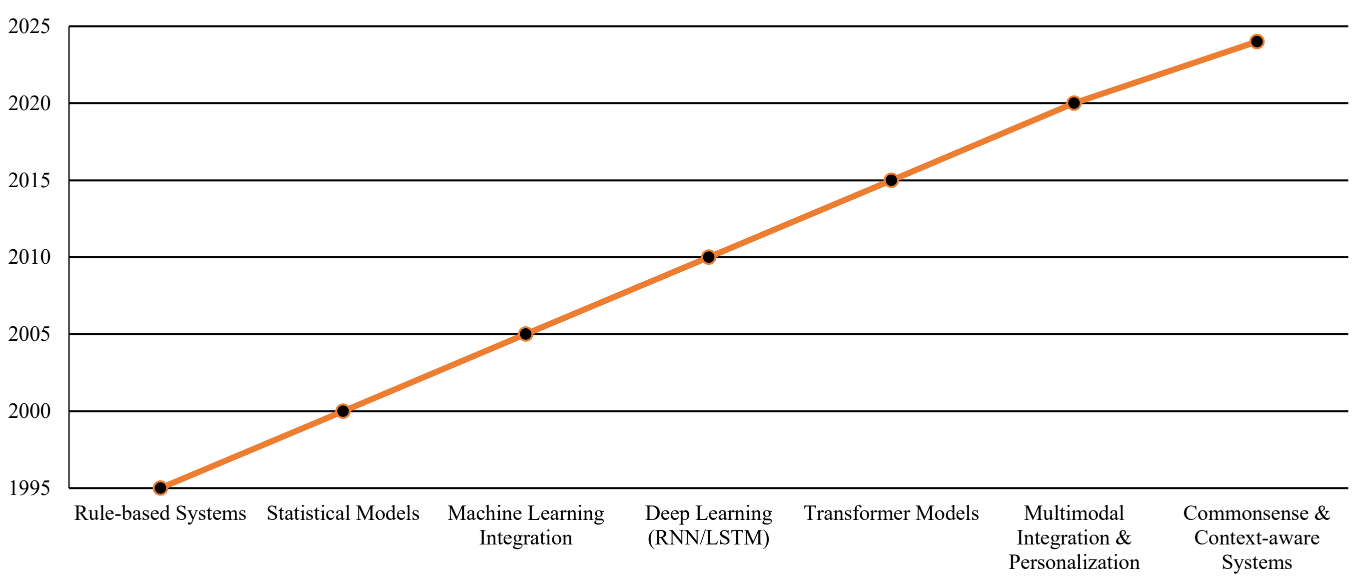

Fig. 3 Chatbot technology trends over time (1995 to 2024)

Table 3. The technological progression of chatbot development

S/N	Period	Category	Technology	Description	Examples	Citation(s)

1	Pre-2000s	Early Rule-Based Systems	Predefined Rules and Pattern Matching	Simple chatbots based on predefined rules and pattern matching.	ELIZA, ALICE	[49], [55]
2	2000s	Statistical Models	N-grams, Hidden Markov Models (HMMs).	Introduction of statistical models for improved language understanding.	IBM's Watson (early versions)	[49], [56]
3	2010s	Machine Learning Integration	Support Vector Machines (SVMs), Decision Trees	Adoption of machine learning techniques for better handling of natural language.	Microsoft's Cortana, Apple's Siri	[29], [30]
4	Mid-2010s	Deep Learning and Neural Networks	RNNs, LSTMs	Use of deep learning models for advanced language understanding.	Google Assistant, Amazon Alexa	[36], [37], [62]
5	Late 2010s – Present	Transformer Models & Pre-trained Models	BERT, GPT-3, GPT-4, T5	Leveraging transformer architecture for state-of-the-art NLP performance.	OpenAI's GPT series, Google's BERT	[79], [100]
6	The Future	Multimodal Integration & Personalization	Multimodal models, XAI, Domain-Specific Langs.	Integration of multiple data types and advanced personalization features.	Next-gen AI assistants (e.g., Gemini)	[87], [89]

6. Methodology

During the search process, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA-S) guidelines were followed, which provide a framework for systematic reviews [62]. The paper was developed in three stages: planning, conducting, and reporting. In the planning stage, research questions and relevant keywords were developed, and databases were identified. Inclusion and exclusion criteria were defined following the PRISMA-S protocol, which involves establishing clear criteria for selecting and excluding studies to ensure consistency and transparency. In the conducting stage, relevant articles were collected based on the criteria. Finally, the articles were critically analyzed in the reporting stage, and conclusions were drawn. The PRISMA-S protocol ensures a systematic, transparent, and reproducible review process by providing guidelines for planning, conducting, and reporting systematic reviews and meta-analyses.

6.1. Planning the Review

6.1.1. Objectives and Research Questions

Four research questions were formulated to guide the review, focusing on the design, challenges, advances, and context modeling strategies in conversational agents (CAs), especially for low-resource languages. Table 3 shows the research questions and the motivation behind the research questions. This systematic literature review (SLR) followed PRISMA-S guidelines to ensure a rigorous, transparent, and replicable review process. The study was structured around four key research questions (RQs), each of which guided the search strategy, selection criteria, data extraction, and synthesis.

6.1.2. Digital Libraries and Associated Queries

In this stage, the selection of digital databases for the paper search commenced. Scopus, ACM, IEEE Xplore, Springer Link, and Google Scholar were chosen due to their indexing of a wide range of studies.

The search was conducted using the keywords "Conversational agents," "Natural language processing," "Context modeling," "Contextual ambiguity," and "Distributional semantics" that form the queries listed in Table 3.

6.1.3. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria were developed to guide the selection process to find the papers most relevant to the research question.

These sets of criteria were established following standard guidelines to ensure comprehensive coverage. The criteria specify the limitations and prerequisites for including or excluding studies.

Table 3. Research Design Framework

ID	Research Question (RQ)	Motivation	Search Strategy	Inclusion Criteria	Data Extraction Focus	Analysis Strategy

RQ1	What are the current best practices in chatbot design, and how can they be improved or tailored to support low-resource languages?	To explore existing chatbot development strategies and identify adaptable methods for low-resource language environments.	Searched databases (Scopus, IEEE Xplore, ACM, Springer Link, Google Scholar) using: "chatbot design best practices", "conversational agent architecture", and "low-resource language chatbots."	Peer-reviewed; full-text; English; published 2014–2024; focused on chatbot design with relevance to resource availability	Technical frameworks, modular design, language adaptability, customization strategies	Comparative analysis of chatbot architectures, best practices, and adaptability features for underserved languages
RQ2	What key challenges hinder chatbots from accurately understanding context, ambiguity, sentiment, and intent?	To identify and categorize the fundamental natural language understanding (NLU) issues that affect CA performance.	Used keywords: "NLU challenges in chatbots", "context ambiguity", "intent recognition issues", and "sentiment analysis in conversational AI."	Same as RQ1, discuss linguistic or interpretive challenges in conversational agents	Problems in disambiguation, co-reference resolution, emotion detection, multilingual limitations	Thematic clustering of issues (contextual, emotional, syntactic, semantic); expert interpretation of reported limitations
RQ3	What recent developments and trends in context modeling for NLP and text mining influence the performance and reliability of conversational agents?	To review innovative modeling techniques enhancing NLP and their impact on chatbot intelligence.	Search strings: "context modeling NLP", "transformers for chatbot NLP", "text mining and conversational AI trends."	Focus on model development, NLP enhancement, or real-world chatbot applications.	Use of transformers (e.g., BERT, GPT), attention mechanisms, hybrid models, personalization features	Chronological mapping of trends; technical comparison of performance metrics across models
RQ4	How could context/distributional information modeling be explored to improve the capacity for natural language understanding in conversational agents?	To examine how embedding methods and contextual representations enhance NLU, especially for complex inputs.	Search terms: "distributional semantics", "hybrid word embeddings", "NLU for chatbots", and "low-resource language embeddings."	Studies on embedding models, semantic vectorization, contextual modeling techniques	Techniques like GloVe, FastText, Word2Vec, contextual fusion, domain adaptation	Comparative model evaluation; fusion benefits; insights for low-resource deployment and generalization strategies

The inclusion and exclusion criteria are as follows:

- Articles must be published in English.
- Articles must be available in full-text format.
- Only peer-reviewed conference proceedings and journal papers are considered.
- The studies must address the specified research questions.
- Papers must be published within the last ten years to ensure relevance and retain their current state.

These criteria were designed to align with standard practices in the literature and ensure the selection of high-quality, pertinent studies.

6.2. Search, Evaluation and Selection of Relevant Source Materials

Following the PRISMA-S guidelines (see Figure 4), a broad search was conducted across several databases, resulting in 3,100 records: 685 from ACM, 467 from IEEE Xplore, 701 from Springer Link, and 1,209 from Google Scholar. Before the main screening, 199 duplicate entries were removed using EndNote and Rayyan. An initial review eliminated 108 records, mainly due to language restrictions and publication types. Another 243 were set aside because of incomplete data, limited relevance to the study's aims, or lack of peer review. After these steps, 2,512 records remained. Further screening excluded 550 entries that did not meet basic criteria. Out of 1,917 reports that were targeted for full retrieval, 595 could not be accessed.

An eligibility assessment followed, excluding another 1,500 reports that fell outside the inclusion parameters. Ten additional studies, identified separately, were added to the pool.

In the end, after reviewing abstracts, introductions, and conclusions, 148 articles were chosen for detailed analysis. Most of the included articles, about 78.1%, were published between 2014 and 2024. Notably, twenty-six influential studies, comprising roughly 25% of the final selection, were published in 2024 alone.

6.3. Data Extraction and Quality Assessment

Data extraction was carried out with a structured yet flexible approach to balance consistency with the need to capture unique study contributions.

6.3.1. Data Extraction

Each selected article had its key details manually recorded: title, authors, year of publication, and source. The main objectives and research methods were carefully reviewed, with particular attention to how they aligned with the review's core questions. Studies focused on chatbot development, natural language understanding challenges, context modeling innovations, and low-resource language solutions were of special interest. Two researchers extracted data independently to reduce individual bias. Any differences were addressed through discussion until an agreement was reached.

6.3.2. Quality Assessment

Given the wide range of study quality, a detailed checklist guided the appraisal. Criteria considered included:

- Confirmation of peer-review status and reputation of the publication venue;
- Clarity and transparency of the research methodology;
- Strength of relevance to the research questions;
- Recency and topical importance, particularly between 2014 and 2024;
- Originality, coherence, and practical significance of findings.

Studies judged to be outdated, methodologically weak, or only loosely related to the main themes were generally excluded unless offering particularly novel insights. By combining careful data extraction with critical quality evaluation, the final review rested on a solid base of reliable and meaningful scholarship.

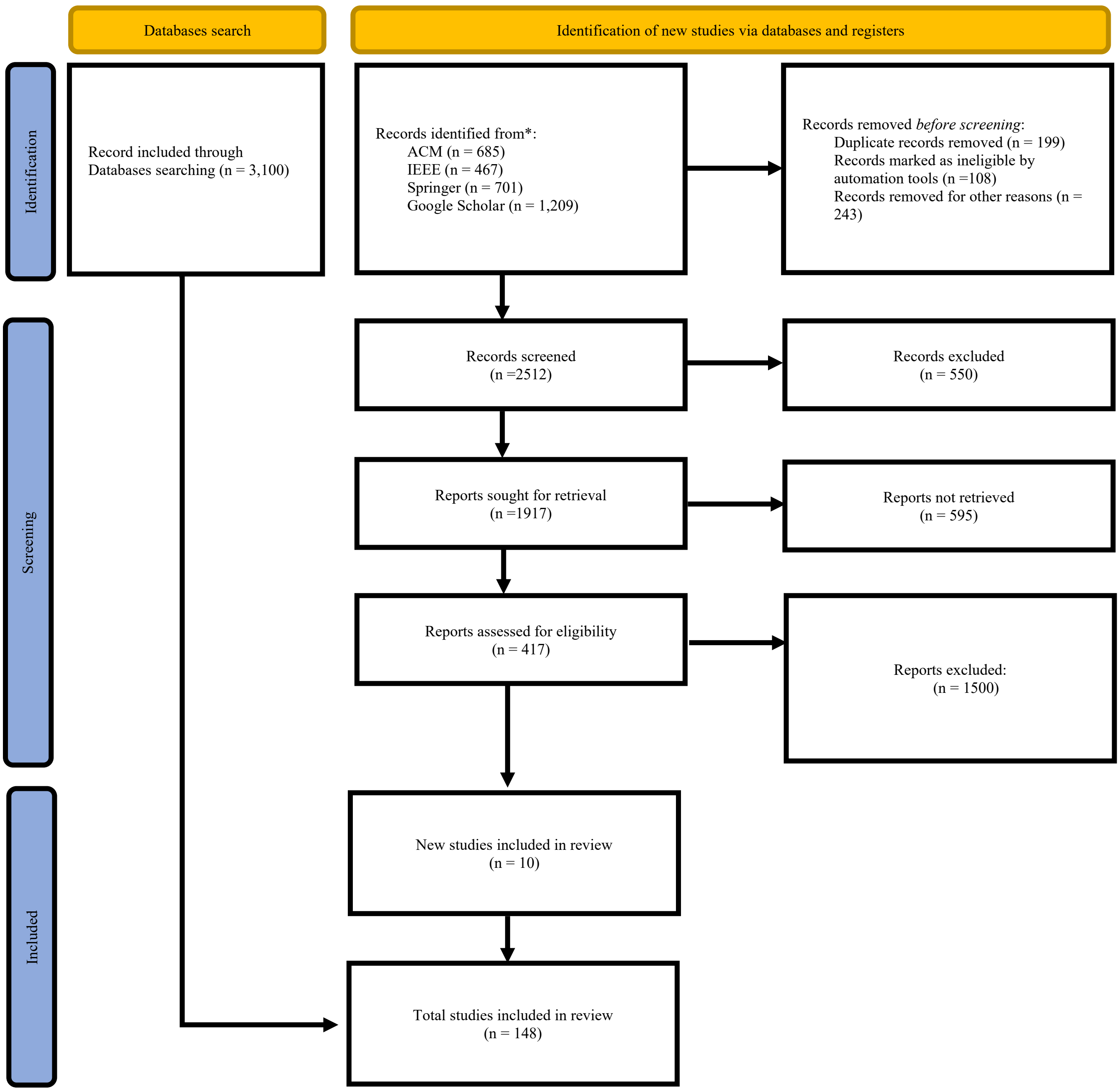

Fig. 4 Literature search and selection process

7. Discussion

In this section, the paper systematically addresses the research questions. First, we present studies on chatbot classification conceptual architecture before presenting development tools. Next, the paper examines the natural language understanding-related challenges facing conversational agent research, after which recent advances, current trends, and future directions in context modeling for natural language understanding are examined. Finally, a review that explores the context and distributional information modeling for enhanced natural language understanding in conversational agents is presented.

7.1. The State-of-the-Art in the Design of Chatbot: How Can it be Improved to Support Low-Resource Languages?

The design of Chatbots has significantly advanced, leveraging cutting-edge technologies to create more intelligent, efficient, and user-friendly conversational agents. Based on the studies by [63], recent improvements in natural language processing and understanding have substantially boosted chatbot capabilities. Transformer models such as GPT-4 and BERT enable chatbots to recognize context, manage ambiguity [64], and create coherent responses, facilitating more natural conversations. Machine learning and deep learning, including reinforcement and transfer learning, allow chatbots to improve their performance over the course of interactions and fine-tune pre-trained models, decreasing the need for large data, respectively. According to [65], recent advancements in multimodal interaction are considered cutting-edge, allowing chatbots to process and respond to multiple input types, including text, audio, and graphics. This capability greatly enhances the user experience by making interactions more dynamic and engaging.

Furthermore, [66] emphasizes that these advancements improve personalization by tailoring responses to individual user preferences and analyzing emotional cues with modern technology. Additionally, the robustness of these systems is strengthened through advanced error detection methods and security measures, which help safeguard user data and ensure reliable performance. [67]. Another key development, as highlighted by [68], is the integration of chatbots with external systems through APIs. This capability allows chatbots to perform practical functions like scheduling appointments and retrieving information from various sources, thus expanding their usefulness. At the same time, UX design plays a crucial role by focusing on making interactions intuitive and user-friendly. This involves designing visual elements and interaction flows that enhance efficiency and user satisfaction. Moreover, ethical principles are central to the development of chatbots, ensuring that their interactions are fair and transparent, which builds trust and promotes responsible use. Hybrid models, as proposed in [69], represent another state-of-the-art. Hybrid models combine the strength of rule-based techniques with machine learning, allowing chatbots to employ established rules for some instances while leveraging machine learning for more flexible and adaptable answers. Ongoing research continues to drive these developments, making chatbots increasingly capable and vital to numerous sectors of life and business [70]. Despite these recent advances and performance gains, chatbot design requires considerable improvement [71]. Enhancing contextual awareness with greater long-term memory and nuanced language processing will enable chatbots to engage in more coherent discussions. Advanced personalization [68**Error! Bookmark not defined.**] using adaptive learning algorithms and behavioral analysis can enable individualized interactions. Improved multimodal integration and the usage of augmented reality can give a richer user experience [67]. Enhancing natural language generation to provide more human-like responses and unique content is also vital [**Error! Bookmark not defined.**1]. Better error management, solid security mechanisms, and ethical AI methods will further advance capabilities. Ongoing research and innovation in these areas promise to make chatbots even more intelligent and important to daily life and business.

7.2. The Challenges Relating to Natural Language Understanding

Conversational agents encounter certain noteworthy obstacles in natural language understanding that affect their capacity to communicate efficiently and in a natural way with humans. Several researchers have attributed these difficulties to contextual ambiguity, limitations associated with managing inputs in diverse languages, the deficit in knowledge required to accurately recognize sentiment and intent, and a lack of capacity for commonsense reasoning and pragmatics. Contextual ambiguity can be categorized into lexical ambiguity and syntactic ambiguity [72]. Lexical ambiguity arises when a term has multiple meanings (for example, "bank" can refer to a financial institution or the side of a river). Syntactic ambiguity, on the other hand, occurs when phrases have multiple interpretations (for instance, "I saw the man with the telescope" could mean either that the speaker saw a man using a telescope or that the speaker used a telescope to see the man). Managing and maintaining context across multiple turns in a conversation poses challenges.

Another type of ambiguity, which is often overlooked in AI design, is pragmatic ambiguity. This is another tricky area when meaning depends on social norms or tone like when someone says, “Can you pass the salt?” It is not really a question about ability. It is a polite way to ask for something. But a bot might take it literally and respond with “Yes,” which sounds robotic and unhelpful [121], also in the expression of idioms. Expressions like “kick the bucket” or “raining cats and dogs” do not mean what the words say. If a chatbot is not trained to recognize that, it might give an answer that makes no sense at all. This is especially true in cultures where idioms are a big part of everyday speech, but the AI has not been exposed to enough examples during training [122]. When these kinds of misunderstandings happen, they can throw off the whole conversation. If the AI gives a weird or irrelevant reply, users might stop trusting it or give up using it altogether. To be useful, especially in areas like healthcare, education, or customer service, AI has to go beyond just processing words. It needs to understand meaning [123].

There has been progress, though. Newer AI models use techniques like transformers and knowledge graphs to keep better track of context and intent. They are getting better at following the flow of a conversation. However, even the most advanced systems still get tripped up when things get too human-like, such as when a conversation shifts suddenly or involves cultural references [124]. Fixing this is a big part of making conversational AI something people can really rely on [73]. Furthermore, recognizing the meaning behind pronouns and references to earlier discussions (anaphora resolution) adds to the complexity, as highlighted by [74]. Another challenge is associated with the lack of capacity to manage inputs in diverse languages. [75] expands on this by discussing variability, which is evident when users employ a variety of slang colloquial idiomatic languages to convey the same meaning in different ways. This challenge also includes managing various dialects and accents in spoken language. Another aspect is noise, which handles misspellings, grammatical mistakes, and inaccurate speech recognition. As expressed by [76], recognizing sentiment and intent is a natural language understanding-related challenge for conversational agents. Intent recognition, a task of accurately determining the user's intent and sentiment recognition, which is identifying the user's emotional input for an appropriate response, can be complex and multifaceted. These two issues significantly degrade the quality of user experience, especially in support and customer service situations: sentiment and intent recognition remain one of the more complex aspects of natural language understanding (NLU), especially in conversational systems designed for diverse linguistic and cultural contexts [125]. While existing systems often perform adequately in high-resource languages, their effectiveness diminishes significantly in low-resource settings. Recent research has sought to address this issue using a combination of multitask learning, transformer-based architectures, and transfer learning techniques.

A multitask learning approach was introduced that jointly models intent detection and slot filling, enabling performance improvements, particularly in low-resource environments. Their approach demonstrates how shared representation across tasks can reduce the need for large, annotated datasets while enhancing semantic understanding [126]. Similarly, the BERT-based DIET architecture (Dual Intent and Entity Transformer), introduced in frameworks like Rasa NLU, shows promise in balancing performance with efficiency. DIET leverages pre-trained language models to detect intent while simultaneously recognizing entities, making it effective for both intent classification and sentiment analysis across domains. Further emphasis was placed on the importance of incorporating contextual embeddings and commonsense reasoning to handle ambiguous or emotionally nuanced inputs. By integrating knowledge graphs, these systems improve their interpretation of indirect cues, which is key to detecting intent when users express themselves figuratively or emotionally [127]. In addition, zero-shot learning has emerged as a practical method for improving performance in domains with scarce labeled data. These models infer intent or sentiment based on semantic similarity rather than supervised examples.

However, while promising, these models still suffer from semantic drift in highly contextual conversations [128]. Another promising framework uses CommonsenseQA [129] and ConceptNet-enhanced models [130], applying structured commonsense knowledge to interpret sentiment-laden or culturally contextual phrases better. These frameworks address the gap between literal understanding and deeper, emotional or purpose-driven meaning [131]. As the field progresses, research emphasises hybrid models combining rule-based logic for domain-specific clarity with deep learning models for generalization. These systems offer the advantage of interpretability and domain adaptability, which are crucial for developing inclusive AI tools across languages and cultures. [77] also, commonsense reasoning and pragmatics are challenges associated with poor natural language understanding capacity by conversational agents. While pragmatics is concerned with recognizing subtle cues and indirect inquiries, such as interpreting "Can you pass the salt?" as a request rather than a question about capability, commonsense reasoning requires basic knowledge and common sense to accurately understand and respond to user inputs. Other challenges identified by [78] include turn-around discussions, performance and scalability issues, multiple language proficiency, fairness and bias, and privacy and security.

In the opinion of [79], natural language comprehension presents conversational bots with various difficulties, ranging from managing ambiguity and context to ensuring fairness and privacy protection. Overcoming these obstacles will require a combination of domain-specific expertise, cutting-edge AI approaches, and continuous improvement based on user feedback. By addressing these issues, developers can create conversational agents that are more efficient, reliable, and user-friendly. The most frequently discussed NLU challenges in the literature are contextual ambiguity and intent/sentiment misrecognition, which account for nearly half of the concerns raised. Commonsense reasoning and multilingual handling are also prominent, particularly in low-resource environments.

7.3. Recent Advances, Current Trends and Future Directions in Context Modeling for Conversational Agents

In recent years, tremendous advancements have been made in natural language processing with regard to context modeling. According to [80], these developments have been fueled by the creation of increasingly complex models, innovative methods, and a deeper understanding of how language is influenced by context. Among the most significant recent advances are the transformer models [81], such as Bidirectional Encoder Representations from Transformers, BERT, which uses a bidirectional approach to context modeling, and Generative Pre-trained Transformers (GPT-3 and GPT-4), which produce highly contextualized text through extensive pre-training data. Another notable model is T5 (Text-To-Text Transfer Transformer), which enhances flexibility and performance across various tasks by treating all-natural language processing tasks as text-to-text problems. The development of the concept and science of attention mechanisms is another recent advance [82], enabling models to focus on different parts of the input text dynamically.

Additionally, pre-trained language models have become prevalent, involving extensive pre-training on diverse datasets followed by fine-tuning for specific tasks. Contextual word embeddings [83], such as Embeddings from Language Models, ELMo, offer word representations that vary depending on the context, unlike static embeddings like Word2Vec and GloVe. Combining static and dynamic embeddings, such as integrating FastText with GloVe, is expected to leverage the strengths of both approaches. Multimodal context understanding has also advanced [84], with techniques like Contrastive Language-Image Pre-training, CLIP, enhancing the ability of models to integrate textual data with other types of data, such as images and audio. Current trends in context modeling for NLP and text mining, as suggested by [85], include Few-Shot and Zero-Shot Learning, where models are increasingly built to perform tasks with minimal or no task-specific data by leveraging pre-training knowledge, exemplified by GPT-3's ability to complete tasks with few examples. Explainability and interpretability are becoming crucial for providing users a clearer understanding of how models make decisions, ensuring greater reliability [86]. Ethical AI and bias mitigation are also prominent, with the aim of developing methods for detecting and reducing biases in context modeling. Real-time and low-resource processing is another trend, focusing on optimizing models for real-time applications and languages or domains with limited data [87]. Techniques like model pruning and knowledge distillation are being explored in this context. Transfer learning and domain adaptation involve fine-tuning models trained on large datasets for specific domains or tasks with less data, enhancing performance on specialized jobs. Looking to the future, several directions for context modeling in NLP and text mining can be noted. Unified models, capable of managing multiple modalities and tasks simultaneously, would offer richer and more flexible context awareness by seamlessly integrating textual, visual, and audio information [88]. Continuous learning, which aims to allow models to continuously learn from new input while retaining previously acquired knowledge, improves adaptability and handling of evolving data and settings will increasingly find relevance. Improved commonsense reasoning is expected to enhance models' ability to generate and understand text by applying commonsense knowledge [89]. Better context retention focuses on creating systems that retain and utilize data from earlier interactions over extended periods. Hybrid approaches combine neural network-based and symbolic AI techniques to leverage their strengths, potentially improving context modeling by integrating structured information and reasoning capabilities [90]. Personalized and adaptive models represent another future direction, designed to adjust to individual user preferences and conversational styles for more contextually relevant and personalized interactions.

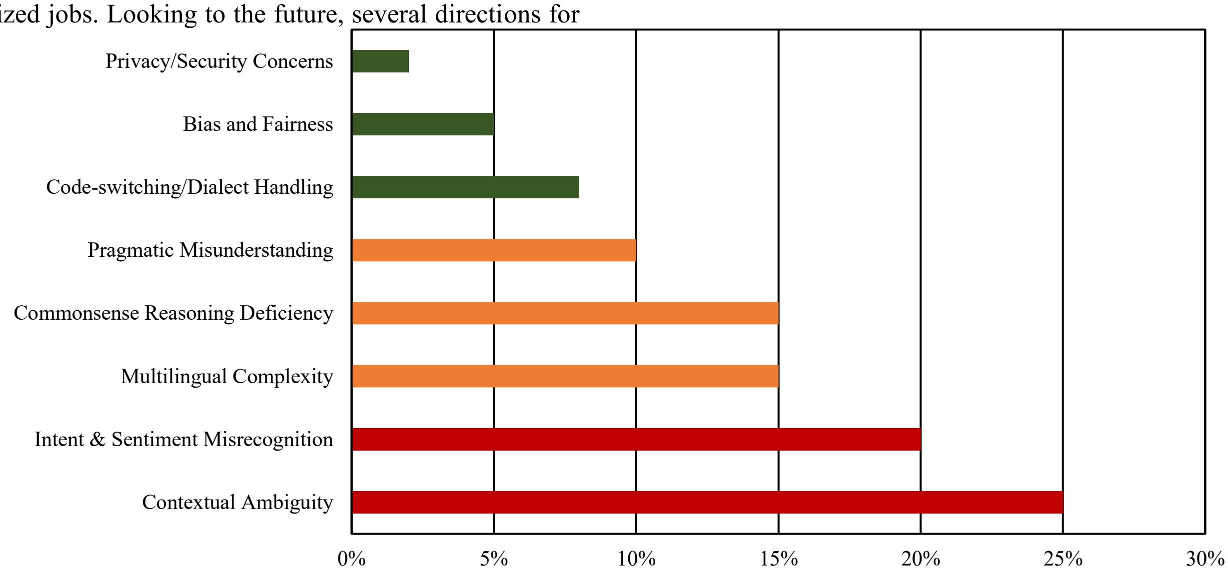

Fig. 5 Estimated Distribution of NLU Challenges

7.4. Improved NLU by CAs via Context Modelling

Recent advancements in natural language processing have underscored the importance of word embeddings in capturing semantic and syntactic relationships between words [91]. Techniques like Skip-gram and Continuous Bag-of-Words (CBOW) have been instrumental in generating dense vector representations of words, reflecting their contextual nuances [92]. However, different embedding models capture distinct semantic aspects, prompting research into combining multiple models to create more comprehensive representations [93]. Studies have explored the theoretical connections between Skip-gram and GloVe, two prominent word embedding models, highlighting the impact of model architecture, training corpus, and parameter design on embedding quality. Combining these models through techniques like concatenation, averaging, or principal component analysis (PCA) has shown promise in enhancing semantic representation [94, 95].

In the context of combining static word embeddings, recent approaches have proposed techniques that leverage the complementary information from different embedding models to enhance the semantic representation of words. For instance, [96] explored the use of a simple voting system that averages the confidence values returned by Bag-of-Words (BOW) and word2vec models to classify text into positive or negative classes. Similarly, other researchers have investigated methods for the joint optimization of word and image embeddings, which can benefit tasks like image annotation and image-text retrieval. Combining multiple-word embedding models can lead to improved semantic representations by leveraging the strengths of each embedding architecture and training corpus. One common approach for fusing static word embeddings is concatenation, where the feature vectors from different embeddings are simply concatenated together, effectively doubling the dimensionality of the final representation. Alternatively, averaging or weighted averaging of the embedding vectors can produce a more compact hybrid embedding [97].

GloVe and FastText are prominent word embedding algorithms that have gained significant recognition in natural language processing. GloVe captures semantic relationships by leveraging global co-occurrence statistics, while FastText incorporates subword information, enabling it to handle rare and out-of-vocabulary words more effectively. Combining these two approaches could create a more robust and versatile word embedding model by harnessing their complementary strengths to enhance overall quality and performance. Recent studies have emphasized carefully curating the training corpus for word embedding models [98]. Researchers have found that the choice of the training corpus can significantly impact the quality of the resulting word representations. Additionally, large, diverse corpora, such as the Common Crawl, can improve word embeddings that capture a wider range of semantic relationships [99]. Investigations have been conducted on the impact of utilizing a combination of the Wikipedia, Statmt News, UMBC, and Gigaword corpora, as well as the Common Crawl, on the quality of the word embeddings generated by the combined GloVe and FastText approach. The study evaluated the performance of these combined word embeddings on various tasks, including semantic similarity, analogy detection, and sentiment analysis. The results demonstrated that the combined GloVe and FastText embeddings outperformed the individual GloVe and FastText models and other state-of-the-art word embedding techniques, highlighting the potential benefits of the proposed approach.

Furthermore, an experiment explored the theoretical relationship between the Skip-gram and GloVe models, two of the most prominent word embedding architectures. The analysis aimed to provide insights into how these models interact and how their combination can enhance the overall performance of word embeddings [100]. The integration of Skip-gram and GloVe models has shown promising results, demonstrating that leveraging the strengths of both models leads to a more comprehensive understanding of semantic relationships and contextual similarities between words.

Integrating Skip-gram and GloVe effectively captures local and global word co-occurrence patterns [101], enhancing word representations. This improved representation is valuable for various natural language processing tasks. A key benefit of this combination is its ability to represent individual words and meaningful phrases, thereby enhancing the expressiveness and compositionality of the word embeddings [102]. The result is a more nuanced and versatile model that improves performance across various NLP applications.

More sophisticated techniques, such as principal component analysis (PCA) and canonical correlation analysis (CCA), can be used to project the embeddings into a shared low-dimensional space, capturing the most salient information from each source [103]. Extensive experiments were conducted on standard benchmarks to assess the performance of these hybrid embedding fusion methods across various NLP tasks, including text classification, named entity recognition, and semantic similarity. The results demonstrate that judiciously combining complementary word embeddings can significantly boost performance compared to using individual embeddings alone, highlighting the benefits of this hybrid approach [104].

Moving forward, further research and experimentation in this area will continue to contribute to the advancement of word embedding techniques and their applications in machine learning and artificial intelligence. Combining embeddings can lead to high-dimensionality challenges, including increased computational complexity, the curse of dimensionality, and a higher risk of overfitting. It also requires more storage space, complicates visualization, and can create difficulties in integration and scalability. Addressing these issues often involves using dimensionality reduction techniques and optimizing algorithms to manage the added complexity effectively. Table 4 compares keyword embedding models used in natural language understanding. It shows that static models like Word2Vec and GloVe are simple and fast but lack context awareness. More advanced models like BERT and GPT capture deeper meaning but require more data and computing power. This comparison helps in choosing the right model, especially for low-resource settings.

Table 4. Comparison of word embedding models for natural language understanding

Model	Type	Architecture	Strengths	Limitations	Reference(s)
Word2Vec	Static	Skip-gram / CBOW	Simple, fast to train, capture semantic similarity based on co-occurrence.	Cannot handle OOV words, context-insensitive.	[91], [92]

GloVe	Static	Matrix factorization	It incorporates global corpus statistics, which is good for capturing analogies.	Context-independent, limited nuance in word meaning.	[91], [92], [93], [94], [95]
FastText	Static (Subword-based)	Skip-gram + character n-grams	Handles OOV words and rare terms better by modeling subword information.	Still context-insensitive; large model size due to subword units.	[98], [99]
ELMo	Contextual (Dynamic)	Bi-directional LSTM	Generates word vectors based on sentence context and handles polysemy.	Computationally expensive, outperformed by transformer-based models.	[83]
BERT	Contextual (Dynamic)	Transformer (Bidirectional)	It captures deep context and semantics and is state-of-the-art for many NLP tasks.	Resource-heavy; not ideal for real-time applications without optimization.	[79], [100]
GPT-2/3/4	Contextual (Dynamic)	Transformer (Unidirectional/Auto-regressive)	Excels at generating human-like text and context modeling.	It requires huge data and computing and is not easily interpretable or fine-tuned locally.	[100], [104]

7.5. Comparative Discussion with State-of-the-Art Techniques

Although this study does not present an original experimental model, it significantly contributes by offering a structured comparative analysis of current state-of-the-art approaches in conversational agents and natural language understanding (NLU), particularly in the context of low-resource languages.

The literature review identifies that transformer-based models like GPT-4, BERT, and T5 perform remarkably well in high-resource language environments.

They often struggle with practical deployment in settings where data, computational power, or linguistic support is limited. These models are largely dependent on massive, annotated corpora, which are unavailable for many African languages. Moreover, existing systems show notable weaknesses in managing contextual ambiguity, code-switching, and dialectal variations, as evidenced by recent studies (e.g., Ssemugabi et al., 2025). In contrast, hybrid architectures that combine rule-based reasoning with deep learning demonstrate enhanced interpretability and domain adaptability, making them more suitable for specialized use cases in low-resource settings. Recent advancements like few-shot learning, zero-shot transfer, and commonsense knowledge integration are promising. However, their standalone performance remains inconsistent, especially when dealing with pragmatic or culturally embedded expressions.

Frameworks that blend multitask learning and contextual knowledge graphs (e.g., ConceptNet) are emerging as viable solutions to bridge the gap in intent and sentiment recognition. By synthesizing these findings, this review offers a clear mapping of common challenges and corresponding technical solutions, serving as a practical guide for researchers and developers. The following visuals in Table 5 and Figure 5 highlight this comparison. The bar chart in Figure 5 is visually comparing how different NLU techniques perform across key challenges, particularly in low-resource environments.

Table 5. Comparison of NLU approaches for conversational agents

Feature / Challenge	**High-End Transformers (GPT, BERT)**	**Hybrid Models (Neuro-Symbolic)**	**Few-/Zero-Shot Learning**	**Commonsense-Enriched Systems**
Contextual Ambiguity Handling	Excellent in major languages	Effective in domain-specific	Moderate	Strong knowledge graphs
Multilingual & Low-Resource Support	Poor	Tunable	Adaptive	Depends on training data
Interpretability	Low	High	Variable	Moderate
Sentiment & Intent Recognition	Good in English	Better with rules	Inconsistent	Context-aware

Cultural/Pragmatic Understanding	Lacks nuance	Case-specific	Limited	Improved with commonsense
Compute Resource Requirement	Very High	Moderate	Efficient	Moderate
Suitability for Low-Resource Use	Limited	High	High	Moderate-high

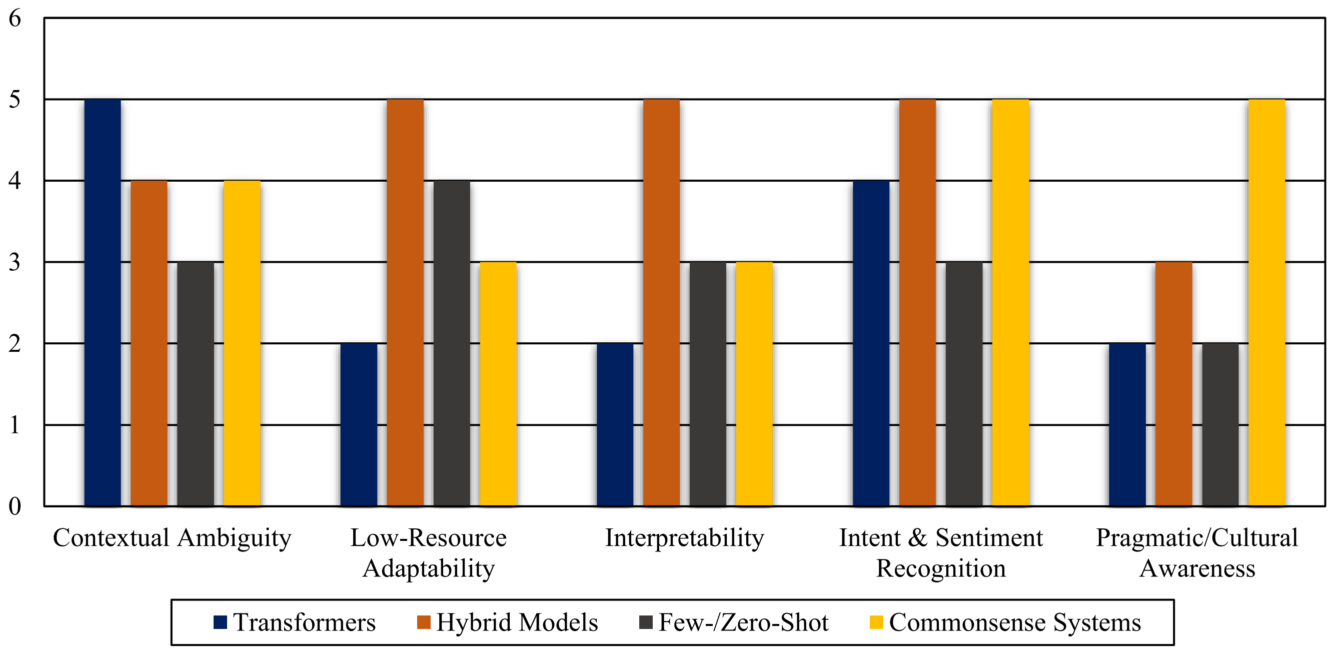

Fig. 6 Comparison of NLU techniques across common challenges

7.6. Mechanisms for Commonsense Reasoning in Conversational Agents

One of the big gaps in how chatbots work today is their lack of common sense reasoning. You can ask them something fairly ordinary, and they might respond in a way that technically makes sense but does not feel right. It is like they understand the words but miss the point. Below are some of the key ways developers are trying to teach AI systems how to reason more like humans:

7.6.1. Integration of Commonsense Knowledge Graphs

There are these things called knowledge graphs. Tools like ConceptNet or ATOMIC are built to help machines understand not just facts but relationships like cause and effect or how people usually react in certain situations. These can be plugged into chatbots so they do not sound clueless when the topic turns to anything outside the strict training data [132].

7.6.2. Neuro-Symbolic Approaches

Another trick is combining traditional logic (like rules and reasoning) with modern machine learning models. These "neuro-symbolic" systems are promising because they let a chatbot learn patterns and apply logic to new situations. That helps it handle things like metaphors or fuzzy phrasing without totally losing the thread [133].

7.6.3. Prompt Engineering and Few-shot Learning

Much progress is coming from clever prompting, basically phrasing things in a way that nudges the AI to answer better. Even large models can show surprisingly smart behavior if you ask things the right way. There is also this approach called few-shot learning, where you give just a couple of examples, and the bot generalizes from there. That is especially handy for languages or topics where you don't have a ton of training data [134].

7.6.4. Commonsense-Augmented Pretraining

Recent work in pretraining involves augmenting training corpora with commonsense-rich texts or synthetic examples derived from knowledge graphs. Models such as COMET-atomic can generate textual representations of inferential knowledge, which are then used to pretrain or fine-tune conversational models. This enhances their ability to handle nuanced inputs and produce coherent, grounded responses [135].

7.6.5. Reinforcement Learning with Human Feedback (RLHF)

Another growing area is letting people guide the AI's development by giving feedback on its answers. This is known as reinforcement learning with human feedback. If the chatbot gives a weird or unhelpful answer, the system can learn not to do that next time. Over time, it builds a better sense of what's "normal."[136]

7.6.6. Application in Multilingual and Low-Resource Contexts

For regions like Africa, where many local languages are not supported by mainstream AI tools, researchers are working on translating commonsense knowledge bases and using multilingual training. The goal is to bring the same level of smart reasoning to these underrepresented languages without needing massive amounts of local data[137].

8. Implications of CA Development in Low-Resource Language Scenario

Developing conversational agents (CAs) for low-resource languages has numerous benefits and challenges. Firstly, it supports linguistic diversity and helps preserve endangered languages like Ainu in Japan and Māori in New Zealand [105]. CAs enhance accessibility and inclusion by providing digital services in native languages, helping bridge the digital divide in regions like East Africa, where Swahili is spoken [106].

Local businesses and governments can improve customer service and engagement in native languages, driving economic growth. For example, small businesses in rural India using Hindi or Tamil chatbots can boost customer satisfaction [107]. However, the scarcity of annotated linguistic data for low-resource languages poses challenges, requiring innovative data collection techniques like crowdsourcing or transfer learning [108]. CAs can also drive technological advancements, promote cultural values, and provide educational support. Ethical considerations, such as fair data usage and community engagement, are crucial [109]. Government support is vital for sustainable CA development, promoting digital inclusion and cultural preservation [110].

9. Application Scenarios: Real-world chatbot Deployments in Low-Resource Language Contexts

While chatbots often seem like tools made for big tech or global languages, there are quite a few examples showing how they can make a real difference in places where local languages don't usually get digital attention. Across Africa and other underserved areas, people are starting to use conversational agents to solve practical, day-to-day problems, whether it is checking health information, learning in their native tongue, or even just preserving the way they speak [138].

9.1. Healthcare: Helping People Access Services in their Language

In East Africa, a mTIBA and Savannah Informatics project launched a chatbot that speaks Swahili. People use it on their phones to book clinic visits, get reminders for vaccinations, and receive basic health advice. It is a simple idea but hugely helpful in areas where there might not be a nearby doctor or hospital [139]. Meanwhile, up north in Nigeria during the COVID-19 pandemic, a Hausa-language chatbot helped people understand symptoms, get hygiene tips, and avoid fake news. The bot worked through WhatsApp and USSD, which meant that even people with basic phones and no internet access could access it. Interestingly, it helped boost awareness, especially among women in rural towns [140].

9.2. Education: Teaching Kids in their Mother Tongue

At the University of Lagos, a group created a Yorùbá-speaking chatbot to help kids learn math and spelling. The bot works on popular platforms like WhatsApp and Facebook Messenger and responds to typed and spoken messages. It is especially useful for young learners who are not confident with English yet [140].

India's Bhashini project has been doing something similar, creating bots that teach in local languages like Assamese and Konkani. The success of those bots shows that this kind of idea could easily be applied in African countries, especially for early childhood learning [142].

9.3. Agriculture: Advice in the Local Dialect

Farmers in Ethiopia can now ask questions about their crops in Amharic or Afaan Oromo, thanks to a voice-powered chatbot built by Digital Green. They can ask about diseases, irrigation, weather, and more. The bot pulls in real-time data and gives answers in a format that does not need much tech know-how [143]. In Nigeria, Hello Tractor has a chatbot in Pidgin English that lets farmers book tractor services without needing to deal with complicated apps. It is a digital middleman, but it talks like the people using it, making a big difference [144].

9.4. Governance: Making Civic Info Easier to Access

The team behind Ushahidi in Kenya built a chatbot that people could use during elections to report issues like violence or misinformation. It speaks Kiswahili, Luhya, and Kikuyu. People did not need an app, just a basic phone with SMS or a chat app [145]. In South Africa, the government's GovChat service is testing local-language bots in Zulu, Xhosa, and Sesotho to share updates about public services and grants. Instead of needing to visit an office, people can get answers from a chatbot in their language [146].

9.5. Keeping Languages Alive

Some developers also use chatbots to help preserve languages spoken less and less. In Nigeria, a test version of an Igbo chatbot helps people (especially those living abroad) practice the language.

It shares short stories, idioms, and common phrases [147]. Ghanaian devs have done something similar with Twi-language chatbots [148], and over in New Zealand, there is one for Te Reo Māori that teaches pronunciation and cultural context through casual chats [105].

9.6. Igbo Chatbots: Teaching, Preserving, and Connecting

The Igbo language, spoken widely in southeastern Nigeria, has also been the focus of some early but promising chatbot projects. These bots are helping with everything from schoolwork to culture.

For example, Teaching Igbo Abroad, Sharing Traditional Knowledge, Helping Students with WAEC, Public Info in Enugu State, and Caring for the Elderly[140]. An example of low-resource chatbots is described in Table 6.

Table 6. Real-World Chatbot Applications in Low-Resource Language Contexts

S/N	Location	Language(s)	Platform	Use Case	Impact
1	Nigeria	Hausa	WhatsApp, USSD	COVID-19 info and health awareness	Improved access to health tips in rural areas with limited internet
2	Kenya, Tanzania	Swahili	SMS, Mobile App	Clinic booking, vaccination alerts	Increased immunization awareness and reduced clinic no-shows
3	India	Hindi, Tamil, Assamese	Mobile App, WhatsApp	Education and basic learning	Enhanced literacy and learning engagement among children and youth
4	South Africa	isiZulu, Sesotho	Facebook Messenger	Government services and FAQs	Broadened digital inclusion for citizens in local languages
5	Rwanda	Kinyarwanda	Chatbot via Messenger	Financial literacy and mobile banking	A better understanding of digital banking for rural users
6	Ghana	Twi, Ewe	IVR, SMS	Agriculture advice for farmers	Improved crop yield and farming decisions through localized information
7	Nigeria	Igbo	WhatsApp, Mobile App	Basic education and cultural learning	Promotes native language literacy and supports language preservation

10. Conclusion

Conversational agents, chatbots or virtual assistants are widely used in many industries. They simulate user conversations to provide customer support, information retrieval, or entertainment services. Despite their benefits, CAs have several limitations. They often struggle with complex language patterns, sarcasm, and maintaining conversation flow.

They also lack common sense reasoning, leading to confusing responses or difficulty with unexpected questions. Training data for CAs can be biased, leading to discriminatory or offensive replies and concerns about data security and privacy.

Additionally, while text and voice-based CAs are common, physical, embodied CAs are still being developed, which limits their ability to interact with the physical world or convey emotions effectively. Looking ahead, researchers are working on several strategies to address these issues and enhance CAs. Advances in natural language processing are expected to improve their understanding of complex language and context.

Future CAs might also use artificial general intelligence (AGI) to provide more insightful responses. Explainable AI could help users understand how CAs make decisions, increasing trust and transparency. Emphasis on user privacy will be crucial for maintaining trust. Lastly, future CAs may integrate text, voice, and gestures to offer a richer and more natural user experience. This report presents a systematic literature review (SLR) on recent advancements in chatbot research, focusing on current trends, context modeling, and information distribution in natural language processing and text mining.

The review examined key research spanning over one decade in a quest to address four main areas of chatbot development, including its categorization, architecture, development tools, and recent breakthroughs. It highlighted the rapid evolution of chatbots across various domains and identified emerging trends and future directions. T

he paper also discussed the challenges related to development tools, such as NLP service locking, and suggested that creating a domain-specific language could facilitate independent chatbot development, overcoming the limitations of existing tools.

10.1. Future Research Directions

Despite the rapid advances in conversational AI, several critical areas require focused research to overcome current limitations, especially in low-resource language environments. Future studies should explore the following directions:

10.1.1. Development of Multilingual and Cross-Lingual Models

Research should focus on designing transformer-based models and embeddings that can generalize across languages with minimal annotated data. Low-resource language pretraining, zero-shot, and few-shot learning should be further refined to support African dialects, emphasising semantic preservation during translation.

10.1.2. Commonsense-Enhanced NLU for Cultural Contexts

Incorporating commonsense reasoning into NLU is essential. Future work should explore culture-sensitive commonsense knowledge graphs and neuro-symbolic models that combine statistical learning with symbolic reasoning for better interpretation of idioms, metaphors, and pragmatic speech patterns in Indigenous languages.

10.1.3. Explainable AI (XAI) and Fairness Audits

With increasing chatbot adoption in sensitive domains like healthcare and education, research should ensure these models are explainable and free from bias.

Future work should focus on creating XAI toolkits tailored for conversational agents and bias-detection frameworks for multilingual deployments.

10.1.4. Data Curation Strategies for Low-Resource Settings

Researchers should innovate in ethical data collection, such as community-driven annotation, crowdsourcing, and the use of synthetic corpora generation through prompt engineering and large language models. These methods could help build foundational resources for underrepresented languages.

10.1.5. Adaptive Multimodal Context Modeling

Future systems should integrate multimodal data (text, speech, image, gestures) to better model context in human-computer interaction. Custom context-tracking modules must be developed to handle interruptions, topic shifts, and turn-taking, especially in culturally rich dialogue environments.

Acknowledgments

We are truly grateful to the Federal University of Technology, Owerri, for backing us during this study. Thank you to our friends and Computer Science and Information Technology colleagues for their feedback and encouragement. It meant a lot. We also appreciate those who read earlier drafts and shared thoughtful suggestions that helped make this paper stronger.

Authors' Contributions

Godson C. Uzoaru: Conceptualization, Data Curation, Writing – Original Draft, and Editing, Formal Analysis

Ikechukwu I. Ayogu: Conceptualization, Data Curation, Methodology, Formal Analysis, Writing – Original Draft, Writing and Review, Validation, Supervision.

Aloysius C. Onyeka: Conceptualization, Supervision, Methodology, Formal Analysis, Resources, Writing – Review

Juliet N. Odii: Supervision, Methodology, Formal Analysis, Resources, Writing – Review

Matthew Emeka Nwanga: Methodology, Visualization, Writing - Review and editing, Resources

Francisca Onyinyechi Nwokoma: Formal Analysis, Data curation, Validation, Resources

Euphemia Chioma Nwokorie: Conceptualization, Methodology, Visualization, Resources

Obi Chukwuemeka Nwokonkwo: Methodology, Formal Analysis Writing - Review, Resources

Chinwe Gilean Onukwugha: Resources, Data Curation, writing - Review and editing

Anthony Ifeanyi Otuonye: Validation, Visualization, Methodology, Resources

References

[1] Iis Tussyadiah, "A Review of Research Into Automation in Tourism: Launching the Annals of Tourism Research Curated Collection on Artificial Intelligence and Robotics in Tourism," Annals of Tourism Research, vol. 81, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Marwan Mohamed Abdeldayem, and Saeed Hameed Aldulaimi, "Trends and Opportunities of Artificial Intelligence in Human Resource Management: Aspirations for Public Sector in Bahrain," International Journal of Scientific and Technology Research, vol. 9, no. 1, 3867-3871, 2020. [Google Scholar] [Publisher Link]

[3] Ana Paula Chaves, and Marco Aurelio Gerosa, "How should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design," *International Journal of Human–Computer Interaction*, vol. 37, no. 8, pp. 729-758, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Robert Dale, "The Return of the Chatbots," Natural Language Engineering, vol. 22, no. 5, pp. 811-817, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[5] Bayan Abu Shawar, and Eric Atwell, "Chatbots: are they Really Useful?," Journal for Language Technology and Computational Linguistics, vol. 22, no. 1, pp. 29-49, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[6] Dennis Benner et al., "What do you Mean? A Review on Recovery Strategies to Overcome Conversational Breakdowns of Conversational Agents," *International Conference on Information Systems (ICIS)*, pp. 1-17, 2021. [Google Scholar] [Publisher Link]

[7] Petter Bae Brandtzaeg, and Asbjørn Følstad, "Chatbots, Changing User Needs and Motivations," *Interactions*, vol. 25, no. 5, pp. 38-43, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[8] Savvas Varitimiadis et al., "Graph-Based Conversational AI: Towards a Distributed and Collaborative Multi-Chatbot Approach for Museums," Applied Sciences, vol. 11, no. 19, pp. 1-33, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Ali Abdolrahmani et al., "Blind Leading the Sighted: Drawing Design Insights from Blind Users Towards More Productivity-Oriented Voice Interfaces," ACM Transactions on Accessible Computing (TACCESS), vol. 12, no. 4, pp. 1-35, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Michelle Denise Leonhardt et al., "Using Chatbots for Network Management Training through Problem-Based Oriented Education," Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007), Japan, pp. 845-847, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[11] Amit Sheth et al., "Cognitive Services and Intelligent Chatbots: Current Perspectives and Special Issue Introduction," *IEEE Internet Computing*, vol. 23, no. 2, pp. 6-12, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[12] Thomas Holtgraves, and Tai-Lin Han, "A Procedure for Studying Online Conversational Processing Using a Chatbot," Behavior Research Methods, vol. 39, no. 1, pp. 156-163, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[13] Microsoft, End of Support for Cortana, 2025. [Online]. Available: https://support.microsoft.com/en-gb/topic/end-of-support-for-cortana-d025b39f-ee5b-4836-a954-0ab646ee1efa

[14] Tommy Sandbank et al., "Detecting Egregious Conversations between Customers and Virtual Agents," *arXiv Preprint*, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[15] Neha Gupta et al., Fundamentals of Chat GPT for Beginners Using AI, 1st ed., Academic Guru Publishing House, 2024. [Google Scholar]

[16] Siwei Lai, "Word and Document Embeddings based on Neural Network Approaches," *arXiv preprint*, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[17] Eric Austin et al., "Uncovering Flat and Hierarchical Topics by Community Discovery on Word Co-Occurrence Network," Data Science and Engineering, vol. 9, no. 1, pp. 41-61, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Mourad Jbene et al., "Intent Detection for Task-Oriented Conversational Agents: A Comparative Study of Recurrent Neural Networks and Transformer Models," Expert Systems, vol. 42, no. 2, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[19] Jialu Zhang et al., "Visual-Linguistic Cross-Domain Feature Learning with Group Attention and Gamma-Correct Gated Fusion for Extracting Commonsense Knowledge," *MM '24: Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 4650-4659, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Yusuf Aliyu et al., "Sentiment Analysis in Low-Resource Settings: A Comprehensive Review of Approaches, Languages, and Data Sources," *IEEE Access*, vol. 12, pp. 66883-66909, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[21] Djuraeva Gavkhar Normurotovna, "Local Dialects and Jargon: How Translators Manage Linguistic Diversity in Security-Sensitive Contexts," *International Journal of Artificial Intelligence*, vol. 5, no. 3, pp. 813-827, 2025. [Google Scholar] [Publisher Link]

[22] Samuel Ssemugabi, "The Role of AI in Modern Language Translation and Its Societal Applications: A Systematic Literature Review," Artificial Intelligence Research, pp. 390-404, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[23] Yashrajsinh Chudasama et al., "Towards Interpretable Hybrid AI: Integrating Knowledge Graphs and Symbolic Reasoning in Medicine," *IEEE Access*, vol. 13, pp. 39489-39509, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[24] Ninel Iacobus Antonie et al., "The Role of ChatGPT and AI Chatbots in Optimizing Antibiotic Therapy: A Comprehensive Narrative Review," *Antibiotics*, vol. 14, no. 1, pp. 1-20, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[25] Xiao Liang et al., "Towards Low-Resource Languages Machine Translation: A Language-Specific Fine-Tuning with LoRA for Specialized Large Language Models," *IEEE Access*, vol. 13, pp. 46616-46626, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[26] Milind Shah, and Nitesh Sureja, "A Comprehensive Review of Bias in Deep Learning Models: Methods, Impacts, and Future Directions," *Archives of Computational Methods in Engineering*, vol. 32, no. 1, pp. 255-267, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[27] Todd Jaquith, The History of Chatbots [Infographic], Futurism, 2016. [Onlline]. Available: https://futurism.com/images/the-history-of-chatbots-infographic/

[28] Macrina Lazo, and Ryan Ebardo, "Artificial Intelligence Adoption in the Banking Industry: Current State and Future Prospect," *Journal of Innovation Management*, vol.11, no. 3, pp. 54-74, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[29] Vivek Bhardwaj et al., "Conversational AI-a State-of-the-Art Review," *Conversational Artificial Intelligence*, pp. 533-555, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[30] Saadat Izadi, and Mohamad Forouzanfar, "Error Correction and Adaptation in Conversational AI: A Review of Techniques and Applications in Chatbots," *AI*, vol. 5, no. 2, pp. 803-841, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[31] Melissa L. Rethlefsen et al., "Prisma-s: An Extension to the Prisma Statement for Reporting Literature Searches in Systematic Reviews," *Systematic Reviews*, vol. 10, no. 1, pp. 1-19, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[32] Sheetal Kusal et al., "A Systematic Review of Applications of Natural Language Processing and Future Challenges with Special Emphasis in Text-Based Emotion Detection," *Artificial Intelligence Review*, vol. 56, no. 12, pp. 15129-15215, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[33] Mohaimenul Azam Khan Raiaan et al., "A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges," *IEEE Access*, vol. 12, pp. 26839-26874, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[34] Reto Gubelmann, "Large Language Models, Agency, and Why Speech Acts are Beyond Them (For Now)-A Kantian-Cum-Pragmatist Case," *Philosophy & Technology*, vol. 37, no. 1, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[35] Daniele Orsucci, Marco Vista, and Filippo Maria Santorelli, "Conversational AI in Neurogenetics. The Example of FUS Gene," *Journal of the Neurological Sciences*, vol. 473, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[36] Anisha Ali et al., "Ontology based Semantic Analysis framework in Sindhi Language," *VFAST Transactions on Software Engineering*, vol. 13, no. 1, pp. 193-206, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[37] David Mhlanga, "AI War between China and America and the Rise of DeepSeek-R1: Implications for Sustainable Development," *SSRN*, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[38] Long Ouyang et al., "Training Language Models to Follow Instructions with Human Feedback," Advances in Neural Information Processing Systems (NeurIPS 2022), vol. 35, pp. 1-15, 2022. [Google Scholar] [Publisher Link]

[39] Weijiao Huang, Khe Foon Hew, and Luke K. Fryer, "Chatbots for Language Learning-are they Really Useful? A Systematic Review of Chatbot-Supported Language Learning," *Journal of Computer Assisted Learning*, vol. 38, no. 1, pp. 237-257, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[40] Matteo Mario Carlà et al., "Large Language Models as Assistance for Glaucoma Surgical Cases: A ChatGPT vs. Google Gemini Comparison," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 262, no. 9, pp. 2945-2959, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[41] Karen Weise et al., "*Inside the AI Arms Race that Changed Silicon Valley Forever*," International New York Times, NA-NA, 2023. [Google Scholar]

[42] Adebowale Jeremy Adetayo, Mariam Oyinda Aborisade, and Basheer Abiodun Sanni, "Microsoft Copilot and Anthropic Claude AI in Education and Library Service," *Library Hi Tech News,* 2024. [CrossRef] [Google Scholar] [Publisher Link]

[43] Jordi Calvet-Bademunt, and Jacob Mchangama, "Freedom of Expression in Generative AI: A Snapshot of Content Policies," *The Future of Free Speech*, United States of America, 2024. [Google Scholar] [Publisher Link]

[44] Philip Maxwell Thingbø Mlonyeni, "Personal AI, Deception, and the Problem of Emotional Bubbles," *AI & Society*, vol. 40, no. 3, pp. 1927-1938, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[45] Rohun Gupta et al., "The Intersection of AI Grok with Aesthetic Plastic Surgery," *Aesthetic Surgery Journal*, vol. 44, no. 6, pp. NP437-NP440, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[46] Joseph 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, 1966. [CrossRef] [Google Scholar] [Publisher Link]

[47] Karolina Kuligowska, "Chatbot Parry, Kenneth Mark Colby-Virtual Assistant Parry-Virtual Agent Parry-Chat Bot Parry-Conversational Agent Parry," 1972. [Google Scholar]

[48] Mark Olsen, "Talking Back: The INRAC Language Compiler," *Computers and the Humanities*, vol. 23, no. 2, pp. 161-167, 1989. [Google Scholar] [Publisher Link]

[49] R. Carpenter, "Jabberwacky 13.0-Learning Artificial Intelligence-AI Software Applications," 1997. [Google Scholar]

[50] Collins Garrett, and San Mawos, "1994 Loebner Prize Awarded," *IEEE Expert-Intelligent Systems & Their Applications*, vol. 10, no. 1, pp. 67-67, 1995. [Google Scholar]

[51] Alfred Baer, and Cancer Quac kery, "National Capital Area Skeptical Eye," vol. 5, no. 4, 1991. [Google Scholar] [Publisher Link]

[52] Lori S. Levin, and David A. Evans, "ALICE-Chan: A Case Study in ICALL Theory and Practice," *Intelligent Language Tutors*, 1st ed., Routledge, 1995. [Google Scholar] [Publisher Link]

[53] Ayse Pinar Saygin, Ilyas Cicekli, and Varol Akman, "Turing Test: 50 Years Later," *Minds and Machines*, vol. 10, no. 4, pp. 463-518, 2000. [CrossRef] [Google Scholar] [Publisher Link]

[54] Ian R. Kerr, "Bots, Babes and The Californication of Commerce," *SSRN*, 2005. [Google Scholar] [Publisher Link]

[55] Henrik Wann Jensen, "Monte Carlo Ray Tracing," *ACM SIGGRAPH*, vol. 5, 2003. [Google Scholar]

[56] Emmelyn A.J. Croes, and Marjolijn L. Antheunis, "Can We be Friends with Mitsuku? A Longitudinal Study on the Process of Relationship Formation between Humans and a Social Chatbot," *Journal of Social and Personal Relationships*, vol. 38, no. 1, pp. 279-300, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[57] A.E. Eichenberger et al., "Using Advanced Compiler Technology to Exploit the Performance of the Cell Broadband Engine™ Architecture," *IBM Systems Journal*, vol. 45, no. 1, pp. 59-84, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[58] Lauren Effron, iPhone 4S's Siri Is Lost in Translation with Heavy Accents, ABC News Network, 2011. [Online]. Available: https://abcnews.go.com/Technology/siri-lost-translation-heavy-accents/story?id=14834111

[59] Dewi Agushinta R. et al., "Comparative Study of Intelligent Personal Assistant," *International Journal of Engineering and Innovative Technology*, vol. 1, no. 6, pp. 154-160, 2012. [Google Scholar] [Publisher Link]

[60] Siti Noraisyah Abd Rahman, "*An Existential Education: A Study of New Media Technologies in Fostering Self-Actualisation*," Master Thesis, Multimedia University, Malaysia, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[61] Rafael Sarkis-Onofre et al., "How to Properly use the PRISMA Statement," *Systematic Reviews*, vol. 10, pp. 1-3, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[62] Chiranjib Chakraborty et al., "Overview of Chatbots with Special Emphasis on Artificial Intelligence-Enabled ChatGPT in Medical Science," *Frontiers in Artificial Intelligence*, vol. 6, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[63] Tianjie Zhao et al., "Artificial Intelligence for Geoscience: Progress, Challenges and Perspectives," *The Innovation*, vol. 5, no. 5, pp. 1-25, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[64] Rocco de Filippis, and Abdullah Al Foysal, "Chatbots in Psychology: Revolutionizing Clinical Support and Mental Health Care," *Voice of the Publisher*, vol. 10, no. 3, pp. 298-321, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[65] Geeta Pattun, and Pradeep Kumar, "Emotion Classification Using Generative Pre-Trained Embedding and Machine Learning," *2023 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT)*, San Salvador, El Salvador, pp. 1-6, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[66] Anas El-Ansari, and Abderrahim Beni-Hssane, "Sentiment Analysis for Personalized Chatbots in e-Commerce Applications," *Wireless Personal Communications*, vol. 129, no. 3, pp. 1623-1644, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[67] Mateusz Dubiel et al., "On-Device Query Intent Prediction with Lightweight Llms to Support Ubiquitous Conversations," *Scientific Reports*, vol. 14, no. 1, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[68] Avyay Casheekar et al., "A Contemporary Review on Chatbots, AI-Powered Virtual Conversational Agents, ChatGPT: Applications, Open Challenges and Future Research Directions," *Computer Science Review*, vol. 52, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[69] Amon Rapp, Lorenzo Curti, and Arianna Boldi, "The Human Side of Human-Chatbot Interaction: A Systematic Literature Review of Ten Years of Research on Text-Based Chatbots," *International Journal of Human-Computer Studies*, vol. 151, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[70] Sofia Schöbel et al., "Charting the Evolution and Future of Conversational Agents: A Research Agenda along Five Waves and New Frontiers," *Information Systems Frontiers*, vol. 26, no. 2, pp. 729-754, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[71] Zhuofan Ying, Peter Hase, and Mohit Bansal, "Adaptive Contextual Perception: How to Generalize to New Backgrounds and Ambiguous Objects," *Advances in Neural Information Processing Systems*, vol. 36, pp. 1-23, 2023. [Google Scholar] [Publisher Link]

[72] Feriel Khennouche et al., "Revolutionizing Generative Pre-Traineds: Insights and Challenges in Deploying ChatGPT and Generative Chatbots for FAQs," *Expert Systems with Applications*, vol. 246, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[73] Si On Yoon, Melissa C. Duff, and Sarah Brown-Schmidt, "Keeping Track of who Knows What in Multiparty Conversation Despite Severe Memory Impairment," *Neuropsychologia*, vol. 194, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[74] Ray Wilkinson, *Conversation Analysis and Communication Disorders*, The Handbook of Clinical Linguistics, Second Edition, pp. 69-80, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[75] Mauajama Firdaus, Asif Ekbal, and Erik Cambria, "Multitask Learning for Multilingual Intent Detection and Slot Filling in Dialogue Systems," *Information Fusion*, vol. 91, pp. 299-315, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[76] Iwan Williams, and Tim Bayne, "Chatting with Bots: AI, Speech Acts, and the Edge of Assertion," *Inquiry*, pp. 1-24, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[77] Arunkumar Yadava, "Ethical and Regulatory Challenges of AI Adoption in the Financial Services Sector: A Global Perspective," *International Journal of Science and Research Archive*, vol. 10, no. 13, pp. 1222-135, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[78] Partha Pratim Ray, "ChatGPT: A Comprehensive Review on Background, Applications, Key Challenges, Bias, Ethics, Limitations and Future Scope," *Internet of Things and Cyber-Physical Systems*, vol. 3, pp. 121-154, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[79] Gokul Yenduri et al., "GPT (Generative Pre-Trained Transformer)–A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions," *IEEE Access*, vol. 12, pp. 54608-54649, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[80] Siyu Lu et al., "The Multi-Modal Fusion in Visual Question Answering: A Review of Attention Mechanisms," *PeerJ Computer Science*, vol. 9, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[81] Bonan Min et al., "Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey," *ACM Computing Surveys*, vol. 56, no. 2, pp. 1-40, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[82] Rezan Bakır, and Halit Bakir, "Swift Detection of XSS Attacks: Enhancing XSS Attack Detection by Leveraging Hybrid Semantic Embeddings and AI Techniques," *Arabian Journal for Science and Engineering*, vol. 50, pp. 1191-1207, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[83] Waad Alhoshan, Alessio Ferrari, and Liping Zhao, "Zero-Shot Learning for Requirements Classification: An Exploratory Study," *Information and Software Technology*, vol. 159, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[84] Chongchong Yu et al., "Acoustic Modeling Based on Deep Learning for Low-Resource Speech Recognition: An Overview," *IEEE Access*, vol. 8, pp. 163829-163843, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[85] Taha ValizadehAslani et al., "Two-Stage Fine-Tuning with ChatGPT Data Augmentation for Learning Class-Imbalanced Data," *Neurocomputing*, vol. 592, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[86] Chunyuan Li et al., "Multimodal Foundation Models: From Specialists to General-Purpose Assistants," *Foundations and Trends® in Computer Graphics and Vision*, vol. 16, no. 1-2, pp. 1-214, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[87] Meng Jiang et al., “Augmenting NLP Models with Commonsense Knowledge,” *Knowledge-augmented Methods for Natural Language Processing*, Singapore, pp. 65-89, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[88] Rrubaa Panchendrarajan, and Arkaitz Zubiaga, “Synergizing Machine Learning & Symbolic Methods: A Survey on Hybrid Approaches to Natural Language Processing,” *Expert Systems with Applications*, vol. 251, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[89] Yueying Zhou et al., “Personalized Federated Learning Incorporating Adaptive Model Pruning at the Edge,” *Electronics*, vol. 13, no. 9, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[90] Nathaniel Robinson et al., “Task-Dependent Optimal Weight Combinations for Static Embeddings,” *Northern European Association for Language Technology*, vol. 8, no. 1, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[91] Sawsan Alshattnawi et al., “Beyond Word-Based Model Embeddings: Contextualized Representations for Enhanced Social Media Spam Detection,” *Applied Sciences*, vol. 14, no. 6, pp. 1-25, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[92] Yang Yang et al., “Unraveling Lexical Semantics in the Brain: Comparing Internal, External, and Hybrid Language Models,” *Human Brain Mapping*, vol. 45, no. 1, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[93] Suhang Wang et al., “Linked Document Embedding for Classification,” *CIKM'16: ACM Conference on Information and Knowledge Management*, Indianapolis, Indiana, USA, pp. 115-124, 2016. [CrossRef] [Google Scholar] [Publisher Link]
[94] Xin Rong, “Word2vec Parameter Learning Explained,” *arXiv Preprint*, 2014. [CrossRef] [Google Scholar] [Publisher Link]
[95] Terry Ruas et al., “Enhanced Word Embeddings Using Multi-Semantic Representation through Lexical Chains,” *Information Sciences*, vol. 532, pp. 16-32, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[96] Chao Zhang et al., “Multimodal Intelligence: Representation Learning, Information Fusion, and Applications,” IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 3, pp. 478-493, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[97] Tomas Mikolov et al., “Advances in Pre-Training Distributed Word Representations,” *arXiv Preprint*, 2017. [CrossRef] [Google Scholar] [Publisher Link]
[98] Amit Mandelbaum, and Adi Shalev, “Word Embeddings and their Use in Sentence Classification Tasks,” *arXiv Preprint*, 2016. [CrossRef] [Google Scholar] [Publisher Link]
[99] Silvan Hornstein et al., “Text Classification of German Mental Healthcare Data: A Hands-on Tutorial on Natural Language Processing Techniques,” *PsyArxiv Preprint*, pp. 1-23, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[100] Ayhan Arısoy, “Natural Language Processing Algorithms and Performance Comparison,” *Yalvac Academy Magazine*, vol. 9, no. 2, pp. 106-121, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[101] Yee Ling Boo et al., “Special Issue Editorial on “The Innovative Use of Data Science to Transform How We Work and Live”,” *Data Science and Engineering*, vol. 9, no. 1, pp. 3-4, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[102] Diksha Khurana et al., “Natural Language Processing: State of the Art, Current Trends and Challenges,” *Multimedia Tools and Applications*, vol. 82, no. 3, pp. 3713-3744, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[103] Cem Akkus et al., “Multimodal Deep Learning,” *arXiv Preprint*, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[104] Wided Selmi, Hager Kammoun, and Ikram Amous, “Semantic-Based Hybrid Query Reformulation for Biomedical Information Retrieval,” *The Computer Journal*, vol. 66, no. 9, pp. 2296-2316, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[105] Gary F. Simons, and Charles D. Fennig, Ethnologue: Languages of the World, Twenty, Dallas, Texas: SIL International. [Online]. Available: http://www.ethnologue.com
[106] Qi Ma, Alan H. S. Chan, and Pei-Lee The, “Bridging the Digital Divide for Older Adults via Observational Training: Effects of Model Identity from a Generational Perspective,” *Sustainability*, vol. 12, no. 11, pp. 1-24, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[107] Mukherjee, S. (2019). “AI for All: How AI is Empowering the Common Man in India.” DataQuest. Retrieved from https://www.dqindia.com/ai-empowering-common-man-india/
[108] Rico Sennrich, Barry Haddow, and Alexandra Birch, “Improving Neural Machine Translation Models with Monolingual Data,” arXiv Preprint, 2015. [CrossRef] [Google Scholar] [Publisher Link]
[109] L. Floridi, “The Fourth Revolution: How the Infosphere is Reshaping Human Reality,” Oxford University Press, 2014. [Google Scholar]
[110] UNESCO. (2019). “National Language Policy Framework.” UNESCO Institute for Lifelong Learning. Retrieved from https://uil.unesco.org/national-language-policy-framework
[111] James Patrick Weichert, “*AI Education for the AI Generation: A Study of Computer Science Student Attitudes Towards AI Ethics and Implications for CS Curricula*,” Doctoral Dissertation, Virginia Tech, 2025. [Google Scholar]
[112] Charaf Ouaddi et al., “DSL-Driven Approaches and Metamodels for Chatbot Development: A Systematic Literature Review,” *Expert Systems*, vol. 42, no. 2, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[113] Wentao He et al., “Hierarchical Perceptual and Predictive Analogy-Inference Network for Abstract Visual Reasoning,” *MM '24: Proceedings of the 32nd ACM International Conference on Multimedia*, Melbourne, VIC, Australia, pp. 4841-4850, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[114] Xieling Chen et al., “Cognitive-Inspired Deep Learning Models for Aspect-Based Sentiment Analysis: A Retrospective Overview and Bibliometric Analysis,” *Cognitive Computation*, vol. 16, no. 6, pp. 3518-3556, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[115] Ye Cheng et al., "Say What You Mean: Natural Language Access Control with Large Language Models for Internet of Things," *arXiv Preprint*, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[116] Honggang Liu, and Jiqun Fan, "AI-Mediated Communication in EFL Classrooms: The Role of Technical and Pedagogical Stimuli and the Mediating Effects of AI Literacy and Enjoyment," *European Journal of Education*, vol. 60, no. 1, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[117] Vengai Musanga, Serestina Viriri, and Colin Chibaya, "A Framework for Integrating Deep Learning and Symbolic AI Towards an Explainable Hybrid Model for the Detection of COVID-19 Using Computerized Tomography Scans," *Information*, vol. 16, no. 3, pp. 1-22, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[118] Flavia Pennisi et al., "The Role of Artificial Intelligence and Machine Learning Models in Antimicrobial Stewardship in Public Health: A Narrative Review," *Antibiotics*, vol. 14, no. 2, pp. 1-19, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[119] Mohammed Fakhreldin, "Developing a Comprehensive NLP Framework for Indigenous Dialect Documentation and Revitalization," *International Journal of Advanced Computer Science & Applications*, vol. 16, no. 4, pp. 815-823, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[120] Yahya Layth Khaleel, Mustafa Abdulfattah Habeeb, and Thierry Oscar Codjo EDOH, "Limitations of Deep Learning vs. Human Intelligence: Training Data, Interpretability, Bias, and Ethics," *Applied Data Science and Analysis*, vol. 2025, pp. 3-6, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[121] Om Prakash Sharma et al., *Introduction to Natural Language Processing*, Addition Publishing House, 2025. [Google Scholar]
[122] Tuba Türkel, and Christian Brogaard, "Using Machine Translation as a Self-Directed Language Learning Tool for Low-Proficiency LCTL Learners in a Tandem Partnership," *CALICO Journal*, vol. 42, no. 1, pp. 143-164, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[123] Melanie Schwede, Maik Hammerschmidt, and Stefan Morana, "Fluent Voice, Credible Choice–How Language Embodiment and Consumption Goals Shape Voice Commerce Success," *Journal of Service Management*, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[124] James Michels et al., "Natural Language Processing Methods for the Study of Protein-Ligand Interactions," *Journal of Chemical Information and Modeling*, vol. 65, no. 5, pp. 2191-2213, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[125] Nikitas Karanikolas et al., "Large Language Models Versus Natural Language Understanding and Generation," *PCI '23: Proceedings of the 27th Pan-Hellenic Conference on Progress in Computing and Informatics*, pp. 278-290, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[126] Kamred Udham Singh et al., "Sentiment Analysis in Social Media Marketing: Leveraging Natural Language Processing for Customer Insights," *International Conference on Information and Communication Technology for Competitive Strategies*, pp. 457-467, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[127] Erik Cambria, "Natural Language Understanding & AI," *Understanding Natural Language Understanding*, pp. 1-39, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[128] Kai Zhang et al., "Comprehensive Study on Zero-Shot Text Classification Using Category Mapping," *IEEE Access*, vol. 13, pp. 23526-23546, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[129] Carl Strathearn, Yanchao Yu, and Dimitra Gkatzia, "From Documents to Dialogue: Context Matters in Common Sense-Enhanced Task-Based Dialogue Grounded in Documents," *Expert Systems with Applications*, vol. 279, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[130] P. Swarajya Lakshmi et al., "Conversational Image Recognition Chatbot," *Journal of Science, Computing and Engineering Research*, vol. 8, no. 3, pp. 1-4, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[131] Ece Gumusel, "A Literature Review of User Privacy Concerns in Conversational Chatbots: A Social Informatics Approach: An Annual Review of Information Science and Technology (ARIST) Paper," *Journal of the Association for Information Science and Technology*, vol. 76, no. 1, pp. 121-154, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[132] Hai-Tao Yu et al., "Estimating the Plausibility of Commonsense Statements by Novelly Fusing Large Language Model and Graph Neural Network," *Information Processing & Management*, vol. 62, no. 4, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[133] Chathuranga Sampath Kalutharage, Xiaodong Liu, and Christos Chrysoulas, "Neurosymbolic Learning and Domain Knowledge-Driven Explainable AI for Enhanced IoT Network Attack Detection and Response," *Computers & Security*, vol. 151, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[134] Felix Ott et al., Metric-Based Few-Shot Learning with Triplet Selection for Adaptive GNSS Interference Classification," *IEEE Journal of Indoor and Seamless Positioning and Navigation*, vol. 3, pp. 81-104, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[135] Aseem Srivastava et al., "Sentiment-Guided Commonsense-Aware Response Generation for Mental Health Counseling," *arXiv Preprint*, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[136] Roland Oruche et al., "A Survey on the Recent Advancements in Human-Centered Dialog Systems," *ACM Computing Surveys*, vol. 57, no. 10, pp. 1-36, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[137] M. Jaleed Khan, John G. Breslin, and Edward Curry, "KnowZRel: Common Sense Knowledge-Based Zero-Shot Relationship Retrieval for Generalised Scene Graph Generation," *IEEE Transactions on Artificial Intelligence*, pp. 1-11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[138] Elliot G. Mitchell et al., "Automated vs. Human Health Coaching: Exploring Participant and Practitioner Experiences," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CSCW1, pp. 1-37, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[139] Audrey Mbogho et al., "Building Low-Resource African Language Corpora: A Case Study of Kidawida, Kalenjin and Dholuo," *arXiv Preprint*, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[140] Isa Inuwa-Dutse, "NaijaNLP: A Survey of Nigerian Low-Resource Languages," *arXiv Preprint*, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[141] Surajit Sarma, and Nabankur Pathak, "An Emerging Area of Research: Building an Assamese AI Chatbot for Educational Institutions Using Bi-LSTM Model," *Proceedings of the NIELIT's International Conference on Communication, Electronics and Digital Technology*, Singapore, pp. 561-575, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[142] Teshome Mulugeta Ababu, Michael Melese Woldeyohannis, and Emuye Bawoke Getaneh, "Bilingual Hate Speech Detection on Social Media: Amharic and Afaan Oromo," *Journal of Big Data*, vol. 12, no. 1, pp. 1-23, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[143] Namita Singh et al., "Farmer. Chat: Scaling AI-Powered Agricultural Services for Smallholder Farmers," *arXiv Preprint*, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[144] Movine Omondi et al., "Elections in Africa: Trends, Diversity, and Emerging Technologies," pp. 1-86. [Google Scholar] [Publisher Link]

[145] S.E. Madima, "Positioning South African Indigenous Languages on Social Media Communication in the Fourth Industrial Revolution," *Digital Media and the Preservation of Indigenous Languages in Africa: Toward a Digitalized and Sustainable Society*, vol. 77, 2024. [Google Scholar]

[146] Dereck M.A. Orji, and Ijeoma Jennifer Korie, "The Igbo Language in an AI World: Assessing its Current Status," *OHAZURUME-Unizik Journal of Culture and Civilization*, vol. 3, no. 1, pp. 137-144, 2024. [Google Scholar] [Publisher Link]

[147] S. Glover-Tay, "*Exploring the Efficacy of Twi-Language Interfaced ChatGPT as a Self-Learning Support Tool for Junior High School Students in Rural Communities*," Doctoral Dissertation, Ashesi University, 2024. [Google Scholar]

[148] Hemi Whaanga, and Te Taka Keegan, "Petera Hudson Computing Technologies for Resilience, Sustainability and Resistance," *TechRxiv Preprints*, 2023. [CrossRef] [Google Scholar] [Publisher Link]