



A review of natural language processing in contact centre automation

Shariq Shah¹ · Hossein Ghomeshi² · Edlira Vakaj² · Emmett Cooper² · Shereen Fouad³

Received: 22 August 2022 / Accepted: 14 June 2023

© The Author(s) 2023

Abstract

Contact centres have been highly valued by organizations for a long time. However, the COVID-19 pandemic has highlighted their critical importance in ensuring business continuity, economic activity, and quality customer support. The pandemic has led to an increase in customer inquiries related to payment extensions, cancellations, and stock inquiries, each with varying degrees of urgency. To address this challenge, organizations have taken the opportunity to re-evaluate the function of contact centres and explore innovative solutions. Next-generation platforms that incorporate machine learning techniques and natural language processing, such as self-service voice portals and chatbots, are being implemented to enhance customer service. These platforms offer robust features that equip customer agents with the necessary tools to provide exceptional customer support. Through an extensive review of existing literature, this paper aims to uncover research gaps and explore the advantages of transitioning to a contact centre that utilizes natural language solutions as the norm. Additionally, we will examine the major challenges faced by contact centre organizations and offer recommendations for overcoming them, ultimately expediting the pace of contact centre automation.

Keywords Contact centres (CCs) · Natural language processing (NLP) · Automation · Machine learning (ML)

1 Introduction

Modern contact centres (CCs) are designed to manage all customer interactions through multiple channels, including telephone, email, web forms, and online live chat. Their

primary goal is to provide customers with a seamless and efficient service while also tracking customer engagement and interaction for an enhanced customer experience. However, CCs face several challenges due to the rising number of customer demands and the enormous volume of data they generate. To overcome these challenges, innovative and smart technologies have become critical success factors for CCs. These technologies help CCs meet the evolving expectations of customers and effectively handle the vast amount of data they produce.

The growing impact of information and communications technologies (ICT) evolution has led to a rapid application of recent scientific advances in new ubiquitous and personalized products and processes, as well as a shift to more knowledge-intensive industries and services [1]. In recent years, CC organizations have been busy laying out strategies to adapt advanced technology imminently, from multi-channel CC capabilities, deploying off-premise cloud services and remote working to adopting advanced data-driven platforms [2].

Basic data and analytics tools are becoming standard practice in most current CCs. While that is a solid first step, most organizations are likely not taking full advantage of the technology. According to [3], merely 37% of organizations

✉ Shariq Shah
shariq.shah@bcu.ac.uk

Hossein Ghomeshi
h.ghomeshi@gmail.com

Edlira Vakaj
edlira.vakaj@bcu.ac.uk

Emmett Cooper
emmett.cooper@bcu.ac.uk

Shereen Fouad
s.fouad@aston.ac.uk

¹ Research, Innovation, Enterprise, Employability (RIEE), Birmingham City University, 15 Bartholomew Row, Birmingham B5 5JU, UK

² School of Computing and Digital Technology, Birmingham City University, 15 Bartholomew Row, Birmingham B5 5JU, UK

³ School of Informatics and Digital Engineering, Aston University, Aston St, Birmingham B4 7ET, UK

believe that they are using advanced analytics to create value, thus revealing significant missed opportunities. In the past few years, data analytic and artificial intelligence (AI) technologies have advanced rapidly and CC organizations now have more choices than ever before. Unlike earlier classic data analytics solutions, which helped companies understand what is currently happening within their CCs, advanced analytics can help them generate actionable insights about what will happen next, through both internal and customer-facing applications [3]. This can result in reduced costs, increased revenue, and most importantly higher customer satisfaction. But to fully reap the benefits of advanced analytics, organizations must have the right foundations in place to make the most of their rapidly proliferating data [3].

The continuous growth in computing power, recent breakthroughs in natural language processing (NLP) have further increased the potential of generating valuable insights and radically improving the range of CC tasks. NLP is the AI domain of computer science that understands, learns, and generates natural language data. In other words, a computational technique that deconstructs human language into smaller chunks, analyses relationships, and investigates how they join together to create a meaningful content [4]. The technology combines data science and linguistics to understand language in a similar way to humans. Recently, many CC organizations have moved from the traditional interactive voice response (IVR) system to the NLP technology [5]. The deployment of NLP can help businesses remove day-to-day frustrations that customers face with IVR systems [6], and therefore provide a better customer experience. It can also help organizations collect valuable insights from customer data for a better understanding of customers' demands.

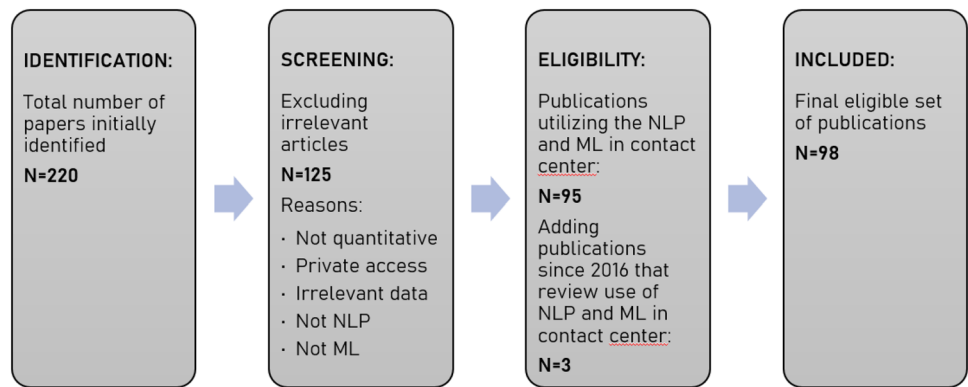
Despite the importance of this topic (i.e. the use of NLP technology in CCs), empirical evidence suggests that there have not been enough studies reviewing this field of research. This finding highlights the motivation and significance of the proposed review paper. Only a few relevant survey researches papers were found in this area (e.g. [7–9]); however, their focus did not completely address the use of NLP within the CC domain. The research in [8] compiles eighteen definitions of CC from the reviewed literature and proposes an updated definition. The authors review 90 papers and classify them into 2 categories; “analytical” and “managerial” studies. The former category contains the majority of studies that implement text-mining techniques for customer satisfaction, sentiment detection, troublesome call detection and segmentation, all with the aim of monitoring calls. Further, CC administration tasks such as logging telephone calls and email routing are included. Contrarily, the managerial category discloses studies on CC performance, customer service representatives (CSRs), and outsourcing the CC. The authors identified two existing research gaps, which supports our view as well, i.e. lack of

studies on CC in current literature and lack of data integrity in CCs. The authors recommend using big data analytical techniques to extract insights from high volumes of unstructured CC data to enhance CC performance. However, it does not solve the issue of data integrity entirely, which demands process changes, from the stage of when the data arrives, how and where it is stored. The limitation of this paper is that it does not thoroughly discuss major analytical problems of CCs and particularizes them with call monitoring.

Another review research in [9] identified four gaps in the existing CC domain research. The authors emphasize on big data plays a key role in the development of the next generation of intelligent CCs. The four gaps are lack of mechanisms for cleansing customers' duplicate profiles, lack of interactive CC for recognition of customers with common names, lack of decision support system (DSS) for CSRs, and lack of studies using advanced techniques to show how CCs could decrease the high CSR churn rate. In their literature analysis, two different techniques, i.e. text mining and data mining, were discussed. Other than data issues within CC which were also mentioned in this paper, the authors recommend incorporating ML and NLP can assist in the development of DSS, helping CSRs in completing CC tasks efficiently. However, the authors examined the literature and highlighted that there is a lack of studies in the development of such DSSs. The other research gaps were specific to particular factors of the CC that are concerning mainly duplication and commonality of CC data and measurement of CSR churn rate. Another recent review article in [7] studies the ethical issues and related considerations of using NLP and ML techniques in CC systems, which is beyond the scope of our research.

The focus of our paper is to conduct a literature review on advanced NLP methods and their important applications in CCs. We firstly discuss popular existing technologies used in CC automation while highlighting their main benefits and limitations to the CC business. We then review the state-of-the-art NLP methodologies and their main applications, challenges, and solutions within CCs. The outcome of this paper will help CC experts better understand the future opportunities of NLP technology, which will facilitate the development of the next generation of CCs, that is well suited for today's evolving competitive world. To our latest knowledge, we are the first to publish a detailed review of using NLP and ML for automating CC tasks. This paper is structured as follows: Sect. 2 provides details on the methodology used to conduct the first-ever systematic literature review of using NLP in CC. Section 3 explains the main highlights in CC automation and Sect. 4 presents a brief overview of NLP. In Sect. 5, current applications of NLP in CC are discussed. In Sect. 6, the results from an experimental study are presented. Finally, we conclude and draw some perspectives in Sects. 7, 8 and 9.

Fig. 1 Prismatically representing screening and study identification process



2 Literature review methodology

The papers we collected are from various sources like Google Scholar, ScieneDirect, Emerald Insight, IEEE Xplore, ACL Anthology, arXiv, and AAAI. The authors resorted to the papers, which were published in the period between 2003 and 2023. While the interest in research on the intersection of natural language processing (NLP) and contact centres (CCs) began in the late 90 s, the majority of papers in this domain were published during 2000 and onwards. The keywords used were “NLP”, “contact centre”, “call centre”, “deep learning”, “natural language processing”, and “transformer”. Although a lot of work has been published in NLP and ML, however, the number of publications falls when searching explicitly for, NLP and ML within the contact centre or call centre domain. The total number of relevant papers identified was 220 and after the removal of duplicate papers and papers that were not related to CC, the number came down to 125. Finally, we included the most relevant 98 following a manual review of all the remaining papers (Fig. 1).

We analysed various studies to determine their relevance to CC, specifically focusing on whether the data used was retrieved from a CC or not. Three reviewers assessed each study’s eligibility and any discrepancies were discussed extensively to ensure thoroughness and high quality. Many studies used a range of methods, algorithms, systems, and evaluation strategies. Multiple modelling techniques were commonly utilized, while only a handful of studies applied a single modelling technique to the data.

3 Main highlights in CC automation

3.1 Customer contact channels

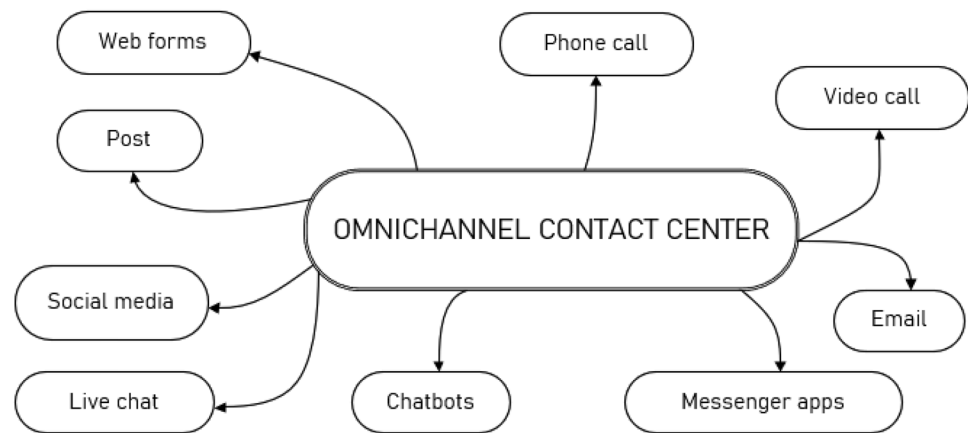
CCs traditionally known as call centres are among the most important contributing factors to customer relationship management (CRM) and serve as the primary interface

between organizations and their customers. Today CCs are referred to as worksites where CSRs interact with customers over omnichannel platform that integrates channels such as phone, email, fax, letter, website, live chat, and social media [8]. Typically, customers use three different communication devices to communicate with an organization’s CC: traditional phone, computer (laptop or desktop), and smartphones. In the early call centres, the only communication channel was a voice but now because people are more tech-savvy and interactions are dominantly widespread in the personal communication/social media market, CC has become omnichannel. Omni-channel CC is a progression from the multi-channel model where various channels of communication are supported and integrated, such as voice chat, video chat, emails, SMS, webchat, and social media messaging [10]. Voice (phone) channel is the most used communication channel in a CC and is usually categorized into two operational modes: inbound and outbound. Inbound is when customers call into CC and outbound is when CSRs call customers. Figure 2 provides an example of various channels offered by a modern contact centre today.

3.2 Interactive voice responses (IVRs)–benefits and limitations

The two main goals of CCs are improving customer satisfaction and reducing operating costs, essentially providing efficient service at a reasonable cost. There are trade-offs in achieving these two goals concurrently as they are perceived as incompatible with each other [11]. It is estimated that 70% of all company interactions are from the CCs [12]. Another report highlights that it costs organizations\1.3 trillion every year on 265 billion customer service calls globally [13]. Thus, automating even a fraction of the interactions handled by the CSRs can generate tremendous cost savings. Organizations have reduced operational costs by focusing mainly on automating critical processes such as automatic call distribution using touch-tone interactive voice responses (IVRs) or outsourcing CCs to other countries with

Fig. 2 Various types of communication channels used in a modern CC



lower labour costs, which accounts for 60–80% of total CC expenditure [14]. However, this has jeopardized customer satisfaction and only resulted in high customer churning and employee attrition rates. Outsourcing has several issues including language and culture, time zone, geographical and legal, and political instability [15]. CCs historically have aimed of achieving the lowest cost of customer service delivery. This is why most businesses relocated their CCs to countries that are inexpensive concerning operations costs. CSRs are not valued which leads to high agent churn, ultimately adding cost for organizations. Key performance indicators (KPs) focused solely on metrics related to cost. Businesses are after the shortest possible average handling time (AHT) and customers are treated less as individuals by subjecting them to generic “scripts” and keeping them on hold for longer periods. Repeat calls have become common as CSR’s focus on minimizing time rather than fixing the issue, meaning issues are often not resolved. Further, customers shift to competitors providing better services, and replacing lost customers with new ones becomes far more expensive than retaining the existing ones.

In addition, touch-tone IVRs have led to problems like complicated menus, homogeneous service, and poor design of user interfaces, and most importantly, customers feel neglected [16]. A survey conducted reports that customers felt frustrated and angry due to the widespread adoption of IVR systems by the CC [17]. As a result, customers seek CSR assistance at the first opportunity, thereby increasing call-waiting time. While touch-tone IVRs are widespread, speech-enabled IVRs have made substantial headway at replacing them. A study reports that customers prefer natural language-based call routing over usual touch-tone cumbersome menus, therefore delivering significant cost savings and meeting customer expectations [18]. The study also shows that about 20% of the callers who opted for touch-tone-based IVR system routing do not get routed correctly to the service department, ensuing in transferred calls subsequently.

3.3 Call routing techniques

In the paper from [19], it is emphasized that skill-based routing is an important but understudied research area. Wrong routing decisions lead to customers being transferred to the wrong department which is a major concern for both customers and businesses. The study from [20] shows that in an outbound call centre context, their proposed algorithm for call scheduling improves the Right Party Call (RPC) rate by 10–15%, which could mean huge savings on cost for a large CC. A study from Bain & Company reports that for most organizations, a 5% increase in retaining customers could mean a 25% to 95% increase in profit [21]. However, unlike costs or productivity, it is difficult to measure customer satisfaction. Most CCs conduct a manual survey with a small group of customers, typically via a telephone interview or mail-in form. As manual surveys are costly to be conducted on all customers, only 1–5% of customers end up being surveyed weeks after their interactions [22]. A study has also found that for decades response rates have been falling across all types of survey research [23]. Hence, conclusions drawn from manual surveys are not very reliable and do not reflect the correct picture of overall customer satisfaction.

3.4 The need for smarter CCs

Organizations have now realized that by focusing too much on minimizing the direct cost of running CCs, they failed to factor in the opportunity costs. Thus, resulting in frustrated customers, falling customer loyalty, loss of valuable cross-sell and up-sell opportunities, and the squandering of customer feedback by treating CCs as an afterthought or as a silo that is measured outside the range of corporate goals. It has become paramount to roll out efficient ways by which the expectations of customers and CSRs are realized. It is not sufficient to just have a skilled-based group of agents in the CC; the total customer experience at every point of contact has to be addressed to create a sustainable experience [24].

Therefore, organizations that recognize the changing customer needs and market have already begun the process of applying advanced NLP techniques into their CCs. Not only it can provide a strong and engaging customer experience and a better understanding of their intent but it also offers cost-effective ways to add value to current customer service offerings, decrease churning rates, and increase sales.

The COVID-19 pandemic has accelerated many trends that were due to happen soon. Remote working agents, digital or social media self-service, messenger bots, and ML have started to replace previous business processes. For end customers, it means a well-crafted service, boosting their experience to a level closer to their expectations. Customers can move swiftly between channels and pick up any error-free, frictionless channels. New technologies such as chatbots are rapidly becoming the norm—which orchestrate interactions in an automated way without human intervention. It is no longer about effectively managing telephone contacts at a lower cost but more about delivering end-to-end experience, using advanced technology to stimulate advocacy and loyalty. However, there is a range of challenges that can slow that acceleration.

Many scholars have recognized the lack of data integrity [25], lack of conjoint between CRM and CC data [26], and complexity of CC's back-end operations [27] as the main challenges of CC. Another issue is the work and effort required to program on the back end that is not fine-tuned and well-structured [27]. As a result, the majority of the data remains in an unstructured format, thus reinforcing the significance of adopting modern techniques that can efficiently analyse unstructured data. One possible way of addressing this is using Big Data tools and technologies and the work from [28] is a good example where they propose an automated system to measure call centre performance. However, the main challenge mentioned by them was the lack of call record corpus. Although existing literature holds practical methods and examples for mining semi-structured and unstructured datasets, the issues of unclean data and heterogeneity within the CC domain remain unaddressed and a paucity of studies remains prevalent. In addition, enhanced NLP applications have progressed significantly and taken the market by storm but there are still challenges that need to be addressed [29]. Organizations need to address these limitations and put in place processes that bridge the gap towards CC automation.

4 Natural language processing (NLP)

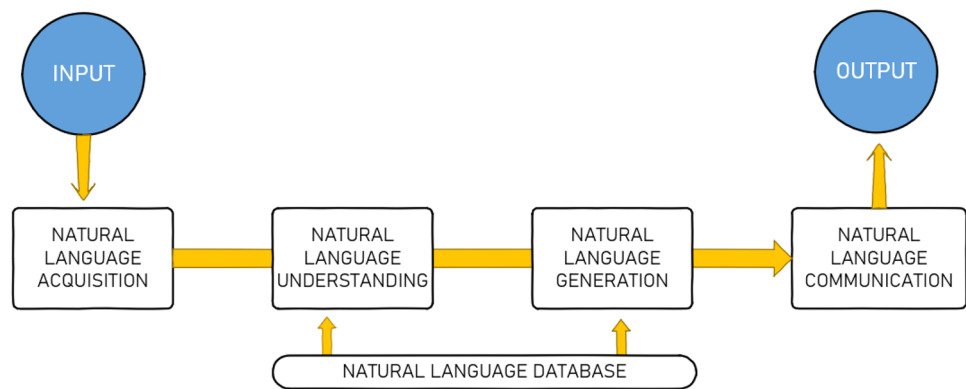
Natural language processing (NLP) is the subset of AI and can be described as an approach based on both a set of theories and a set of technologies that computationally manipulates natural language data (text, speech, or video)

[30]. NLP is a very active research field area and there is not a single definition commonly agreed upon yet. For instance, IBM's Watson is designed to answer questions using a vast amount of data sources and Google Translate is developed for language translation. The field of NLP is deep and diverse and contains a collection of techniques to extract grammatical structure and meaning from natural language. NLP systems can be based on different approaches, i.e. linguistics-focused, statistics-focused, acoustics-focused, or hybrid that combines all approaches. NLP system can often be explained as a system that processes levels of language such as Phonology (deals with the interpretation of speech sounds), Morphology (deals with systematically describing words), Semantics (deals with collecting vital information such as objects and actions from a sentence), and Pragmatics (analysis of the real meaning by disambiguating and contextualizing) [31]. NLP systems are also developed considering various task-oriented tasks like Translation, Categorization, Question-Answering, Dialogue Systems, Summarization, Sentiment Analysis, Recommendation Systems, Named-Entity Recognition (NER), Chatbots, Human-Computer Interface (HCI), and Point of Speech (PoS) Tagging [32]. There is no single approach yet that performs all tasks satisfactorily. It depends on the task and data availability to build a high-performing NLP system.

4.1 A brief history of NLP

The history of NLP goes back to the late 1940s when the term was not even in existence; however, work on machine translation had started. Weaver and Booth started one of the earliest Machine Translation projects in 1946 based on expertise in breaking enemy codes in World War II [33]. It was their idea of using cryptography and information theory for language translation that inspired many projects. It was not until the early 1980s computational grammar theory became a prominent research field, which concentrated on understanding logic, meaning, and extracting beliefs and intentions [34]. By the end of the 1990s, powerful all-purpose sentence processors such as SRI's Core Language Engine [35] and Discourse Representation Theory [36] came into existence, offering practical resources, grammars, tools, and parsers for analysing natural language. The use of statistics became a major theme in the 90s, involving automatic summarization and information extraction and efforts from cross-disciplines became necessary to properly address the challenges of NLP [37, 38]. Until 1990, the progress was slow due to computational and power limitations and research work was mainly in the development of NLP concepts and machine translation. Subsequently, other NLP application areas started emerging and are now significantly researched such as speech recognition [39]. Recent NLP research has evolved majorly with the use of advanced

Fig. 3 Representation of various stages in a typical NLP pipeline



ML algorithms gaining a lot of prominence, especially complex deep learning techniques [40–42]. Current NLP work is dominated by recently proposed NLP models by Google, OpenAI, Toyota, Facebook, and Carnegie Mellon University such as TransformerXL, GPT versions, BERT, XLNet, ALBERT, RoBERTa, and Wav2vec 2.0. They have proven superlative when compared with traditional models. This has also opened many new opportunities for businesses and the open-source community. The reason for their success is due to their fast processing speed and completeness in representing the language.

4.2 NLP pipeline steps

NLP helps in organizing natural language and solving a wide range of problems—Machine Translation, Text Summarization, Named-Entity Recognition (NER), Topic Modeling and Topic Segmentation, Sentiment Analysis, Speech Extraction, Semantic Parsing, Question and Answering (Q &A), Relationship Extraction, etc. In solving the above-mentioned problems, a pipeline needs to be built that follows a methodical workflow.

A typical NLP architecture is a pipeline of distinctive components that may start from either input speech or text data, followed by exploratory data analysis, pre-processing steps such as data cleaning, parsing, and feature engineering techniques whose purpose is to extract meaningful features that help in the task of prediction. There are various steps involved in a pipeline such as for text data—it involves segmentation, tokenization, lemmatization, stop words removal, dependency parsing, noun phrases, NER, etc. However, steps can be skipped or re-arranged depending on the NLP problem. Figure 3 shows a representation of components of a typical NLP system, starting from injecting natural language into the system.

Following that, the data passes through the natural language understanding stage, which performs various tasks of understanding the intent from speech, text, or both. In this stage, speech data may undergo transcription if necessary,

otherwise known as speech-to-text (STT). Depending on the problem, deployed modelling and pattern mining produce outputs in this stage.

In the next stage, i.e. natural language generation, the output of the previous stage helps in generating a response with support from the back-end information source (service management databases, CRM systems, etc.). Following that, natural language communication helps in synthesizing a response into speech, otherwise called text-to-speech (TTS). Combining all the components results in a loop, which repeats each time new data is loaded into the system.

5 NLP applications and methods in CCs

Given NLP and ML algorithms widespread applications in various fields such as translation, spam classification, and question answering, as shown in Fig. 4, organizations have been successfully able to extract customer trends, behaviour, detect associations, and predict best actions. CC's too have the potential to become more customer-driven by adopting advanced NLP and ML algorithms since it generates tremendous amounts of data from distinct channels. Due to NLP and ML attaining high levels of maturity, it is increasingly receiving attention from organizations to help them capture customers' voices, optimize their communication channels, and make better-informed decisions. The main benefit of NLP in CC is in the time savings associated with the automation of various tasks. Automating various tasks with NLP and ML can help CC to shift away from rules-based processes and redundant labour tasks to seamless and personalized processes. Ultimately, this will significantly increase productivity, customer experience, and satisfaction and reduce costs. Research has shown that customer satisfaction strongly correlates with profitability and customer loyalty [43], and drives customer retention [44]. Although the benefits are many, few empirical studies have applied NLP and ML approaches for automating CC tasks. Most of the studies attempted to perform customer satisfaction

Fig. 4 Key capabilities and applications of NLP

Key NLP capabilities					
CC Applications	Recommendation	Customer Feedback Mining	Conversational Agent	Language Generation	Routing
	Intent Analysis	Sentiment Analysis	Agent Monitoring	Voice Recognition	AI Agents
NLP Techniques	Search Engines	Content-based Recommendation	Topic Recognition	Understanding of Commodities	Text Comprehension
	Social Media Analysis	Spam Classification	Discourse Analysis	Image-Recognition	Jeopardy
	Parsing	Speech Synthesis	Knowledge Reasoning	Summarization	Keyword Extraction
NLP Core Features	Constituent Syntax	Syntax Representation	Named Entity	Part-of-speech	
	Dependency Grammar	Word Sense Disambiguation	Sorting	Spelling Check	
	Tokenization	Semantic Role Labelling	Topic Model	Segmentation	

analysis in [22, 45–48], reshaping IVR systems in [16, 18], and sentiment analysis in [49–51]. Numerous studies have used either traditional ML or statistical methods with only a handful exploring deep learning models or state-of-the-art models in the field of NLP.

In the next section of this paper, a review of studies specific to their application field is presented. This is to ensure each key element of CC where NLP has the potential or has already been successfully applied is addressed.

5.1 Customer sentiment analysis and customer satisfaction

Sentiment analysis is to identify, extract, and quantify customers’ emotions and intentions, and translate them into data in real-time. Sentiment analysis tools have been widely used to analyse human feedback and monitor the level of satisfaction in various NLP applications, including social media content (e.g. [52, 53]) as well as in CCs platforms.

Earlier efforts focused on developing an integrated approach where CC data can be utilized for enabling business intelligence, text classification, and interactive text labelling for capturing customer satisfaction [54]. Later, [22] proposed a model that estimated customer satisfaction categorized as satisfied, neutral, and dissatisfied using a 5-point classification scheme, comprising of Naïve bayes, decision tree, support vector machines (SVMs), and logistic regression models. In relation to sentiment analysis, it has been widely studied and some studies have notably used

CC data [49–51, 55–57]. In the last few years, sentiment analysis has gained major research interest, mainly because of its potential application in dialogue systems to produce sentiment-aware and considerate dialogues [58]. However, studies using real-life data extracted from CCs are scarce.

In the study conducted by [46], a method proposed predicted the emotional states (anger or neutral) of the users. Their method employs combining features with N-gram, sentiment words, and domain-specific words. Their study informs on ways in which features can be combined statistically to predict user sentiments. The result is enhanced user satisfaction in a call centre. The dataset that they used was of China mobile call centre. A combination of acoustic and linguistic rules applied supported the development of a multi-dimension model. The classifiers selected were SVMs, Maxent entropy, and traditional Bayesian. The main contribution of their work lies in how they incorporated the results from each of the individual classifiers they used in their work and added acoustic and language rules to it as well. An evaluation of experiments conducted highlighted that their fused system’s F1 measurement result improved to 69.1%, outperforming the baseline SVM model whose F1 measurement was 65.4% (Table 1).

Much attention has been directed to studying the emotional content using speech signals and many systems have been proposed. In [83], authors survey speech-led emotion classification which addresses three crucial aspects; suitable features for speech representation, design of a system, and preparation of a database. Numerous other works have also

Table 1 Summary of studies on customer sentiment analysis and customer satisfaction

Research	Year	Model	Corpus	Access
[59]	2003	HMM	Burmese and Mandarin non-professional actors	Private
[60]	2004	Unigram	French banking CC	Private
[61]	2004	SVM	Global support services survey and knowledge base survey	Private
[57]	2005	SVM and logistic model tree (LMT)	CEMO	Private
[55]	2006	Several classifiers and classification strategies	CEMO—French medical emergency CC	Private
[62]	2007	Trigram language model and term frequency–inverse document frequency (TF–IDF)	Linguistic data consortium (LDC) and real-life CC calls	Public & Private
[49]	2007	SVM, multilayer perceptron–artificial neural networks (MLP–ANN), K-nearest neighbour (KNN), K* instance-based classifier, random forest (RF)	Electrical company CC and Burmese and Mandarin non-professional actors	Private
[63]	2007	Continuous density HMM with Gaussian mixtures and SVM Gaussian	CEMO—French medical emergency CC	Private
[54]	2008	Bayesian and SVM	Telecom CC	Private
[64]	2008	Bayesian and SVM	CC calls	Private
[22]	2009	Decision trees (DT), Naïve bayes (NB), logistic regression (LR) and support vector machines (SVMs)	Automotive company CC	Private
[65]	2010	SVMs	CEMO and EmoVox—French energy company CC	Private
[66]	2011	SVM	CC calls	Private
[67]	2011	BOW, term frequency (TF), TF-IDF and self-referential information (SRI), SVMs and MLP	German IVR, US IVR and German WoZ AIBO	Private
[68]	2011	LR	Greek telecom CC	Private
[69]	2012	SVM, Bag-of-Words (BoW), TF-IDF	JEMO and EmoVox	Private
[70]	2013	SVM	Greek telecom CC	Private
[71]	2013	Hidden Markov model (HMM)	Berlin Emotional Data Corpus (BEDC)	Public
[56]	2013	Boosting and SVM	Emails	Private
[50]	2015	Enhanced pseudo-standard deviation scoring (ESDS), adjective priority scoring (APS) with hybrid evaluation method (HEM)	Telecom CC	Private
[72]	2015	Gaussian mixture models (GMM) and ANN-based classifier	Berlin emotional database (BEDC) and speech-enabled railway enquiry system (SERES)	Public
[46]	2016	SVMs	Chinese Telecom CC	Private
[73]	2016	SVMs	Recorded CC calls	Private
[74]	2016	SVM, ANN, and k-NN	BEDC and CC calls	Public
[75]	2016	Extreme learning machines (ELM), deep neural network (DNN) and convolutional neural network (CNN)	Chinese CC	Private
[76]	2016	CNN	SSPNet Conflict Corpus (SC2)—French televised political debates and Spanish Latin American CC	Private & Public
[45]	2017	BoW, principal component analysis (PCA), XGBoost and CNN	Spanish telecom CC	Private
[77]	2017	Rank score and isotonic regression	US insurance company CC	Private
[78]	2017	NN	eINTERFACE'05 and RML	Public
[51]	2018	NB, Max Entropy and Boosted Trees	Recorded CC calls	Private
[58]	2018	Lexicon-based, recurrent neural network (RNN), LSTM, Bi-LSTM	NLPCC dataset from Weibo	Public
[48]	2020	LSTM-RNNs	Japanese CC (acted and real)	Private
[79]	2020	RoBERTa, Wa2Vec and FABNET	MELD, CMU-MOSEI, CMU-MOSI and IEMOCAP	Public
[47]	2021	Graph neural network (GNN)	US Company CC and Amazon Product Reviews	Private & Public
[80]	2021	Wav2Vec 2.0	IEMOCAP and RAVDESS	Public
[81]	2022	NLP, ML and deep learning classification methods	CC calls	Private
[82]	2022	Unsupervised representation learning techniques	CC calls	Private

investigated the estimation of emotion classification and customer satisfaction at call level using acoustic features such as pitch, duration, energy, intensity, log frequency power coefficients (LFPC), and Mel-frequency cepstral coefficients (MFCCs) [22, 59, 62, 71–73]. Subsequently, Bag-of-Words (BoW) and N-gram are also used in several studies to extract sentiment-related phrases [22, 61, 62, 64, 73, 77]. In the case of [77], features like call dominance or call–turn overlap that reflects customer emotions were exploited. In the work of [72], customer dialogue features like answer repetition were used. Historical events data on customer interactions and in-queue waiting or hold time found in the metadata of calls were used in the work of [22, 77]. SVMs have been mostly used in the above-mentioned works. In the study of [74], a method similar to call level has been utilized for emotion recognition, estimating customer satisfaction during the call using information from the start to the present call time. Features used at call level have proven to be effective [66] including call user’s gender as a feature [70]. Some studies have also proven the use of linguistic event features such as laughing to be also effective [60] as well as the use of visual features when it comes to video-based customer interactions [78]. A recent study in [81] proposed a framework for recognizing interlocutors’ emotions that are specifically designed for CC systems. This approach detects the emotional state of clients as well as agents using text and audio interactions. The study utilizes actual discussions that occurred during the operation of a big commercial CC. They used a wide range of NLP approaches including vectorization, word embedding, transcription methods, dictionaries of emotional expressions as well as multiple machine learning and deep learning classification methods for emotion detection. The detection accuracy obtained for the textual interactions was 70% for agent utterances and up to 60% for client utterances. Whereas, the detection accuracy obtained for the combined interactions (textual as well as audio) exceeds 68%. This method was utilized in [84] to develop an emotion detection method for CC conversations taking into account a wide range of emotions including, anger, fear, happiness, sadness, and neutral. The obtained results were in line with the previously achieved results for both textual and audio channels.

Since call-level customer satisfaction captures the global characteristics of calls, it often becomes too complex for it to work accurately on some real CC calls. For instance, some calls could contain both positive and negative customer reactions as the customer could be dissatisfied with the service at first and then might be either neutral or satisfied at the end of the call [48]. Another method where much attention has been given is an estimation of customer satisfaction and emotion recognition at turn level. Turn level can be explained as several unique segments by the speaker in a given call. It is detectable by identifying each customer turn from other turns between channels. Acoustic and linguistic

features at the lexical level are most commonly applied in the turn-level task [49, 63, 65, 67–69, 74, 78]

A study in [45] assessed the significance of acoustic features from customer-agent interactions to predict customer satisfaction using deep neural architecture. They investigated whether speech prosodic features can be complementary to speech transcriptions. Convolutional neural networks (CNNs) were trained on an amalgamation of acoustic features and word embedding for the binary classification task of “high” and “low” satisfaction. The real call centre dataset of a large Spanish corporation was used. A range of experiments conducted using various modelling approaches BoW, principal component analysis (PCA), XGBoost, and CNN were used. Their study first highlighted the point that linguistic features more accurately predict satisfaction than low-level prosodic and conversational descriptors such as fundamental frequency (F0), loudness and articulation rate. Secondly, turn-level features generally outperform call-level features. Lastly, on the application of fused linguistic and prosodic features using CNN, they reported the best performance of F-score 73.3% compared without prosodic which stood at 60.05%. Other similar works using CNNs also incorporate low-level acoustic features or Automatic Speech Recognizer (ASR) metadata as part of training data for their chosen models [75, 76]. In the study of [76] convolutional neural networks (CNNs) were used on audio frequencies to automatically learn valuable features and predict self-reported customer satisfaction from Spanish CC data.

Another study of [48] employs both turn and call-level features for estimating customer satisfaction. For turn level, they utilized prosodic, lexical and interactive features. They proposed a method that utilizes long-range sequential information and jointly optimizes them to assess the relationship between call–turn-level customer satisfaction. Long short-term memory recurrent neural networks (LSTM-RNNs) were used on call and turn levels to capture long-range sequential call contexts. Both were stacked hierarchically such that turn-level outputs can be utilized for call-level estimation directly. Three experiments highlighted that their proposed framework outperforms SVM and fully connected neural network (NN)-based classifiers for both turn level and call level. More recently, graph neural networks (GNN) was proposed to predict customer satisfaction in a real-life US corporate call centre that takes into account the relative satisfaction scores during training. Their experiments proved more accurate compared with standard regression or classification models [47].

The study from [80] used pre-trained Wav2vec 2.0 embeddings to detect emotions. The authors reported superior performance compared to the result in the literature for two open-source datasets. The authors proved that the Wav2vec 2.0 model performs better when Wav2vec features are combined with a set of prosodic features. Also, the work

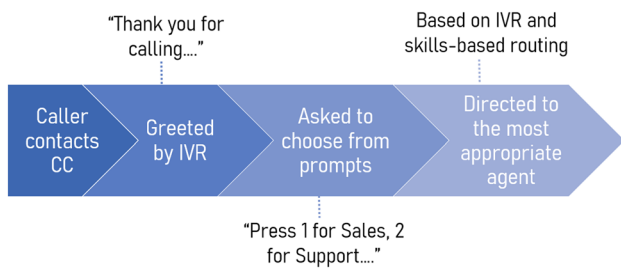


Fig. 5 Typical call routing based on IVR and skills-based rules

from [79] focused on a prominent research direction in representation learning, i.e. using pre-trained self-supervised learning (SSL) models as feature extractors to improve the task of emotion recognition. To achieve this, a transformer-based multimodal fusion mechanism was employed. Their results suggest that SSL features can be effectively used from pre-trained models and the SSL algorithms allow to leverage the potential within largely accessible unsupervised data. Upon evaluation, their approach outperforms the state-of-the-art models on four datasets.

Despite recent advancements in the automatic detection of customer satisfaction, it remains a challenging task due to the scarcity of labelled training data. Collecting large amounts of CC interaction data with customer satisfaction annotations is costly and time-consuming. Recently, authors in [82] have addressed this problem by proposing a customer satisfaction estimation method using unsupervised representation learning techniques. The method demonstrated its effectiveness using real-life CC data interactions.

5.2 Call routing

Call routing also referred to as an automatic call distribution (ACD) can be explained as the process of placing live calls in a queue and distributing them to the relevant departments or agents based on pre-established rules and criteria as shown in Fig. 5. The rules can be based on both customer and agent behaviour, including common routing factors like the reason for the customer's call or the amount of time an agent has gone without speaking to a caller. Intelligent call routing involving various routing strategies such as skills-based, longest available agent, and first available agent allows to instantly connect the caller to a specific phone line or extension without placing the caller on hold. Call routing impacts customer experience significantly as it can benefit in faster resolution, reduced wait time, decreased call abandonment rate, and a more balanced agent workload.

Several works have been published previously on routing calls using natural language call processing. Among many methods and approaches proposed were those using a boosting-based system [85], a vector-based information retrieval

technique [86–88], and a probabilistic model with salient phrases [89]. In [19], various CC functions are reviewed including call routing, skill-based routing, and networking. The authors outline important unaddressed problems and provide promising future research directions.

An article by [90] described a Markov queueing model with three groups of specialized agents and two customer classes. The authors believe that skills-based routing with priority-based rules produces both performance measures and steady-state probabilities. In the work of [86], a routing matrix was trained on statistics of word sequences and the occurrence of words in a training corpus following morphological and stop-word filtering. New user requests represented as feature vectors were routed based on the cosine similarity score with the model destination vectors encoded in the routing matrix. The performance of the above-explained routing system often depends on the routing matrix quality. In the work of [91], the use of discriminative training on the routing matrix was also proposed to improve accuracy and robustness. Instead of simply counting in conventional max likelihood training as shown in the work of [86], they use the min classification error (MCE) criterion in discriminative training of the routing matrix parameters. Discriminative training proved an effective technique when experiments were conducted, outperforming max likelihood classifiers by reducing error rate and increasing robustness. For evaluation, USAA call routing task consisting of 4000 calls belonging to a banking domain and QASIS task involving calls to the UK's British Telecom (BT) operators were used.

Automating call routing has been a challenging task and complexity comes in combining several classifiers to optimize the process as well as when the process scales and involves many different classes (or decisions). This has been a complex problem that has only received little attention as discussed by [85] and [92]. The work of [93] provides a substantial solution to this problem by proposing a global optimization process based on an optimal channel communication model allowing for a combination of heterogeneous binary classifiers. The approach adopted was inspired by Markov modelling in which computational feasibility is achieved through simplifications and easy-to-interpret independent assumptions. The experiments showed call-type classification error rate decreased in a natural language dialogue system by 50%.

The discriminative term selection method has been explored in which the discriminative power of the term is measured. This is calculated by measuring the average entropy variation on the topic when the term is either absent or present. This helps in assigning a numeric value indicating its importance as shown in the work of [94]. The work from [95] highlights the benefits of improving a single classifier's functionality by applying automated relevance

feedback, boosting as well as discriminative training. The study aimed to construct a more accurate classifier. Their proposed algorithm performs by studying each iteration and using the one which is more accurate to minimize training errors. Results were compared to the baseline classifiers and 41–50% improvement in the classification error rate (CER) was observed. More importantly, synergised outputs of discriminative training on the boosting algorithm were also demonstrated and reduced the CER of re-weighted trained classifiers by an average of 72%.

A study from [96], experimented with four models—generalized linear model (GLM), NN, SVMs, random forest. Their study evaluated all four models' performance and NN and SVMs were reported as better performers than the rest for the task routing calls. Similarly, the work from [97] used seven models to predict the most appropriate call operator for the customers. Their results highlight LightGBM as the best model and authors point out that using large amounts of business data can further improve the performance when using innovative algorithms. The work from [98] applied seven various term weighting techniques for feature selection tasks based on a self-adaptive genetic algorithm (GA). k-NN, linear SVM, and NN methods were used as classification models. Experiments demonstrated that the most effective term weighting is term relevance ratio (TRR) and the classification model is NN. Selecting features with self-adaptive GA proves highly effective for classification and dimensionality reduction.

In most natural language-based routing systems, the main purpose of an ASR is to transcribe a user's request in a speech-to-text (STT) so that analysis on the transcription can be performed to determine the most appropriate service destination (agent). Given the level of uncertainty in accurately recognizing words by an ASR, the call can often be incorrectly transcribed, thus raising the possibility of calls being routed to the wrong agent. To tackle this issue, the study from [99] proposes a technique for using confidence scores that an ASR metadata contains to reweigh query vectors in a latent semantic indexing (LSI) classifier. Their results show that it can reduce the number of wrongly routed calls by a significant margin.

More recently, the study from [100] presents an intelligent call routing system that integrates text processing and speech processing. Their system route calls to the most suitable agent using routing rules built by the text classifier. It includes various components: telephone communication network, speech recognition, text classifier, and speech synthesizer. When evaluating the system in the real-world environment, the system proves its accuracy by achieving more than 95%. In call routing problems, understanding the context of customer requests or customer intention holds high importance and any context not understood well could potentially lead to problems. In a study conducted by [101],

context analysis in call routing was investigated and an adaptive neuro-fuzzy inference system and HMM was proposed for solving this problem. Their system can be implemented in any language call routing domain since there are no syntactic or lexical features used in the classification task. Their proposed system reduces errors and increases accuracy to 93% on their dataset.

Yang et al. [102] proposed an automated call routing system that monitors all active live chat conversations in real-time to identify unsatisfied clients who wish to escalate their issues before they end their calls. The intention is to automatically direct their calls to a specialized agent who can help them address their issue before they end the interaction with the original agent. They use a hybrid model by integrating recurrent neural networks with manually engineered features. Experiments show that this method outperforms competitive baselines improving customer service.

The work from [103] proposed an automated triage design that reduces transfer rates and improves routing accuracy in a live chat using combined results from five ML algorithms (SVM, neural network, random forest, Naïve bayes, and adaptive boosting) and text analytics. For evaluation, a real-world large-scale dataset was used and it is noted that routing performance improved by 14%. However, many possible real-world scenarios such as customers with multiple questions that are handled by different CC service categories were ignored as stated by the authors (Table 2).

5.3 Optimizing customer-agent interactions via data analysis

Several works have been completed on analysing customer interactions data that help automate different CC tasks. For instance, areas where customer interactions data has been analysed, include call-type classification for categorizing calls [104], acquiring call logs summaries [105], monitoring and assisting CC agents [28, 106], and development of domain models [107]. Identifying and filtering controversial dialogs from the automatic speech recognizer has also been explored [108–110].

Another area well studied is insight mining patterns in databases where associations are made through structured dimensions [111]. For textual data, many ML-based approaches to mining and classification have been studied [112, 113]. In the research of [114], a method has been proposed to automate the process of extracting knowledge from emails. Their paper reviewed four generations of building systems and their challenges. Their approach used NLP techniques and the results were encouraging; however, they argue user intervention is still required for the system to be accurate enough in providing substantial results. Topic unigram language model has also been explored on counting the word occurrences for each topic as well as storing all words

Table 2 Summary of studies on call routing

Research	Year	Model	Corpus	Access
[89]	1997	Neural network (NN)	AT &T calls	Private
[86]	1998	n-gram and IDF	Financial services CC	Private
[85]	2000	Adaptive Boosting—AdaBoost.MH and AdaBoost.MR	Reuters-21450, AP Titles, UseNet	Public
[87]	2000	n-gram and IDF	US insurance company CC	Private
[92]	2000	SVM and AdaBoost, LR and decision trees	UCI	Public
[88]	2001	Discriminative training (DT)	OASIS	Private
[94]	2002	Discriminative training (DT)	USAA Banking	Private
[91]	2003	Discriminative training (DT)	USAA Banking and OASIS	Private
[93]	2003	Boostexter and SVM	AT &T calls	Private
[95]	2003	Automatic relevance feedback (ARF), boosting and DT	OASIS	Private
[90]	2004	Markov	Numerical experiments	Private
[99]	2004	Latent semantic indexing (LSI)	Recorded calls	Private
[96]	2011	Generalized linear model (GLM), NN, SVMs, and RF	Axion's customer data integration (CDI), Census and CSR surveys	Private
[100]	2016	SVM and Definite Clause Grammar (DCG)	Recorded calls	Private
[98]	2017	k-NN, SVM, and ANN	Speech Cycle	Private
[101]	2018	ANFIS and HMM	Azerbaijani Education company	Private
[102]	2019	Hybrid model by integrating recurrent neural networks with manually engineered features	Recorded CC calls	Private
[97]	2020	LR, NB, RF, NN, AdaBoost (AB), XGBoost (XGB), and LightGBM (LGBM)	Portugese Telecom CC	Private
[103]	2020	SVM, RF, NN, AB, and NB	S &P 500 services company	Private
[103]	2022	Unsupervised extractive summarisation tools	Recorded calls	Private

for each topic. The probability of the query in every topic is calculated and the optimal and most resembling is selected [115, 116]. The study performed [117] an analysis on agent entered call summaries of customers by extracting words based on domain-specified standpoint. In another analysis, insights were extracted based on the usage frequency of the dialogue patterns within customer interactions [118] and [119] analysed and attempted mining from a collection of complete interactions (recorded calls data) from a rental car reservation office to predict whether a customer intends of making a booking or not. Their work identified accurate standpoints and nominated expressions for every standpoint, thus resulting in the chance discovery of valuable insights.

Alternatively, the study from [28] proposed a system that automatically analyses a large number of CC conversations to provide an interface to CC managers measuring CC agent performance. Similarly, the study of [120] assessed the performance of call centre agents like time management or quality by adopting a variety of decision trees, neural networks, and statistical techniques. Also, the study from [121] developed a continuous-time Markov chain model that optimizes the call centre queuing process, thus promising to reduce hold time.

A recent call summarisation study for CC platforms was proposed in [122]. The study applies and compares the summarisation performance of various extractive summarisation

methods. These techniques work by selecting key/important sentences from a given text and present them in the summary verbatim. Unlike abstractive summarisation techniques, extractive summarisation tools are unsupervised methods; hence, they are easy to develop and deploy as they do not require labelled data for training. The paper conducted a comparative analysis of such methods by comparing the summarisation performances of CC calls using subjective and objective evaluation measures. The study reveals that TopicSum and Lead-N methods outperform the baseline summarisation methods as they can produce meaningful summaries of CC interactions.

Although text and audio mining of call centre data have been researched, sequential analysis of the same has not been thoroughly explored. Sequential models have distinct applications but rarely do they appear to be focused on business intelligence. Their most common applications are within telecommunication systems, game strategies, inventory management, and maintenance problems as discussed by [123]. The model proves effective for decisions where outputs are partially controlled and random, thus helping to depict problems and compare strategies objectively. The study from [120] and [121] although adopt sequential techniques they focus precisely on staffing instead of an evaluation of CSR strategies that facilitate conversational flow and outputs. The study from [124] adopted distributed

Table 3 Summary of studies on optimizing customer–agent interactions via data analysis

Research	Year	Model	Corpus	Access
[115]	1994	HMM, contingency table and multinomial	Switchboard	Public
[116]	1997	HMM with Estimate-Maximize (EM)	Broadcast News (available on CDROM)	Public
[117]	2001	Text analysis and knowledge mining (TAKMI)	US PC CC	Private
[108]	2002	Classification and regression trees (CART)	DARPA Communicator dialogues	Public
[109]	2002	RIPPER—incremental reduced error pruning (IREP) algorithm	HMIHY—AT &T CC	Private
[104]	2003	n-gram, Naïve bayes, TF-IDF and SVMs	University of Colorado CC	Private
[120]	2004	Linear NN, MLP, probabilistic NN, CART, decision tree-ANN, SVM	Insurance company CC	Private
[105]	2005	BoosTexter	AT &T CC	Private
[106]	2005	Juru Information Retrieval engine	British National Corpus (BNC) and IBM internal CC	Public and Private
[107]	2006	n-gram	IBM internal CC	Private
[118]	2007	K-Means and AprioriAll	Internal IT CC	Private
[121]	2007	Continuous-time Markov chain (CTMC)	Bell Canada CC	Private
[110]	2008	GMMs, context-dependent HMM, n-gram and Insight Discoverer Clusterer (IDC)	EDF—French Energy company CC	Private
[111]	2008	Typical patterns mining (TPM)	German Financial services CC	Private
[112]	2008	Information granulation and latent semantic indexing (LSI)	UCI Machine learning repository	Public
[113]	2008	Latent semantic analysis (LSA) and probabilistic latent semantic analysis (PLSA)	BizBlogs07	Private
[119]	2009	SVM	Car rental CC	Private
[114]	2012	Email knowledge extraction process (EKE)	Academic emails, UK &I company and Enron email	Private & Public
[28]	2016	Hadoop MapReduce—Cosine and n-gram	Turkish CC	Private
[124]	2017	Latent dirichlet allocation (LDA)	China Central Television CC	Private
[125]	2019	Markov decision process (MDP)	Customer experience management company	Private

computing in the development of topic models from call centre conversations. Although the NLP technique used produced high-level insights, it did not help identify the sequential insights and proved insufficient for turn-level process improvement. In contrast, the work from [125] took into account the sequential nature of agent–customer conversations and used a Markov decision process (MDP) to identify customer states and agent actions. This helped them to identify the most frequent sequence from successful conversations and estimate outcomes when an agent performs a particular action for a customer in that given state. This helps in process improvement and training agents as ideal outcomes can be often used to direct customer conversation flow such that it concludes positively, thereby providing an overall better experience to customers.

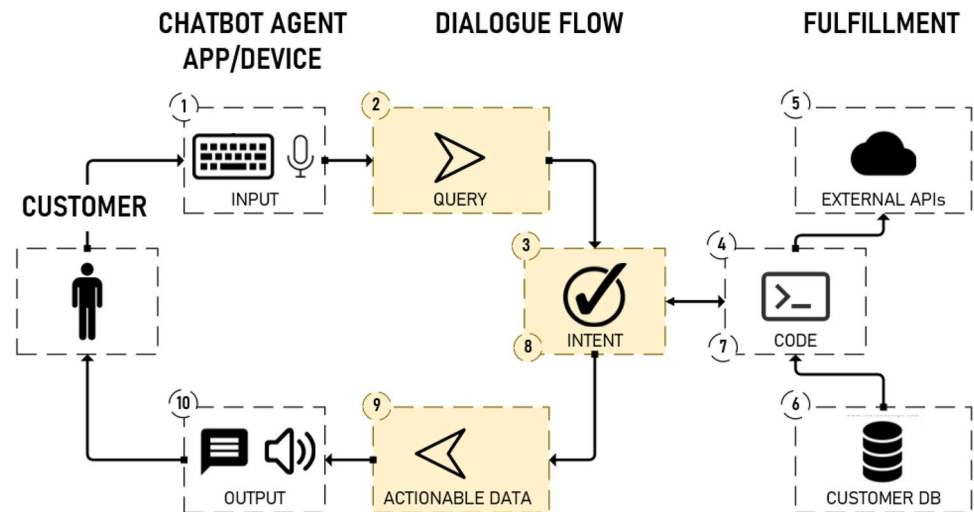
Concerning call-type classification, the work from [126] put forward a method enabling automatic identification of calls that were problematic and required managerial evaluation for call centres. In the work of [106], a call centre monitoring system was proposed which facilitates text analytics and information gathering. Their system analysed the content of call centre data and detected the main issues pointed

out in the data. In [110], a system was presented which could recognize speech and apply text-mining techniques for French call centre data. Whereas, the work of [126] shows an interactive mining tool built on pragmatic analysis and applied to a data corpus containing manually transcribed call centre interactions within the banking domain. Meanwhile, the author mentioned the limitations of the transcription process as not accurate and incapable of identifying phrases that accompany emotions such as gratitude or sarcasm (Table 3).

5.4 Customer service chatbots

Another area of research interest in the domain of CC has been the use of chatbots or virtual agents and speech-enabled IVRs. Chatbots are essentially part of a system with dedicated components such as a dialogue manager, responsible for communicative goals, which is interfaced with a task manager that knows the underlying goals of the communication. Regardless, both are responsible for natural language generation to produce meaningful language utterances which fit the circumstances and specific goals are achieved by following appropriate courses of exchanges. Such a system is

Fig. 6 Typical chatbot workflow



often part of a large spoken dialogue system as well such as speech-enabled IVRs in CCs. The workflow of a typical chatbot is illustrated in Fig. 6.

In an early study of [127], technical innovation within AT & T's eContact space focused on voice-enabled CC automation highlights VoiceTone, an intelligent virtual agent that uses speech and language technology. It acts as a replacement for an existing IVR system and converses naturally to complete customer requests. It emphasizes replacing a cumbersome, menu-based interaction with a more natural and flexible user experience. For the development of a conversational agent, the MDP framework has often been applied. Another early study in [128] proposed a learning dialogue system that used stochastic MDP for an Airlines information system. While the model could successfully reveal optimal strategies, it was not used on the human-human dialogue system but a man-machine system that has less variability than the former.

Over the last ten years, there has been a growing interest around chatbots in CC systems (e.g. [129, 130]). Chatbot technologies gained further attention following the COVID-19 pandemic, which transformed the model of interpersonal communication. A chatbot implementation in [131] was proposed to improve virtual communication with people and provide them with answers about the COVID-19 disease. Another recent work in [132] developed chatbot tool to help with the daily screening of healthcare workers to prevent the spread of COVID-19 in the healthcare setting.

One of the key challenges in modern chatbot systems is to design accurate automatic models for customer intent detection. Early work in [133] proposed a hidden Markov model (HMM) system to model the intention of a sentence using the Viterbi algorithm. The model not only considered the phrase frequency but the syntactic and semantic structure of a phrase frequency. It is substantiated that an accurate determination of the caller's intention helps significantly in

conversing functionality. The experiment results showed a correct response rate of 80.3%. A method that combines two different approaches (Hidden Markov and neuro-fuzzy models) has also been suggested which automatically identifies user intention in a dialogue. The results show that the overall performance of a human-computer dialogue system improved [134]. Other approaches have also been suggested [135–137]. The work from [138] surveys several past and present computational approaches to natural language that generate utterances by using speech acts or words as particular types of actions in solving a problem.

In contrast to other approaches, reinforcement learning (RL) is suited particularly for such tasks where the best strategy to achieve a goal is unknown and the system tries to automatically find an optimal policy from interactions with the user and the environment. An interesting study is from [139] in which hierarchical reinforcement learning (HRL) is used for jointly optimizing spatial behaviours and dialogue behaviours. The proposed method learns to provide navigation instructions by making use of the customer's prior knowledge into account. To improve AHT or response times, CCs need to build systems that can categorize user requests, complaints, and questions and filter them by priority keywords. Also, an automated process that works like a search engine and recommends possible solutions to CC agents. The automated process must have the capability to surface content quickly and offer insights by identifying the relevant patterns from the data. One such publication presents a novel approach in which HRL is utilized for natural language generation in a dialogue system that learns the optimal utterance through reward function [140]. The proposed method optimizes content selection, utterance planning, and surface realization decisions in a joint fashion, otherwise strictly interdependent. Results show that their combined approach outperforms baselines that followed the independent optimization approach. More recently, [141] conducted a study

Table 4 Summary of studies on customer service chatbots

Research	Year	Model	Corpus	Access
[133]	1998	HMM and Viterbi algorithm	ATIS	Public
[128]	2000	Markov decision process (MDP) and reinforcement learning (RL)	Air travel information system (ATIS)	Public
[137]	2001	Text mining and Fuzzy logic	News reports and movie reviews	Public
[127]	2005	n-gram and HMM, GMM, continuous-density HMM, SVM and augmented transition network (ATN)	AT &T CC	Private
[136]	2007	Fuzzy theory	Knoma simulation	Private
[135]	2009	LDA and HMM	Simulated ASR-Channel: Tourist Information (SACTI)	Private
[139]	2011	Hierarchical RL	Wayfinding domain dialogue system	Private
[134]	2012	HMM, LDA, and hybrid neuro-fuzzy	Azerbaijan Education company	Private
[140]	2015	Hierarchical RL	Giving Instructions in Virtual Environment (GIVE)	Public
[141]	2021	Markov process	Numerical assessment	Private
[129]	2022	Inference rules	Polish CC platforms	Private
[142]	2022	Evidence-based machine learning, deep neural network	Seven hours of recorded calls from a real-estate CC	Private

in which a Markov process describing a model function was constructed. The numerical assessment of their model highlights a positive effect of chatbot usage particularly when CC is experiencing an overload of customer queries.

Modern CC systems are increasingly using intent recognition systems in their chatbots systems to improve the quality of their virtual assistance. Recent studies have focused more on this direction by proposing more accurate and robust models for recognizing customer intent. For example, [129] proposed an intent recognition system in CC platforms that takes into account certain human emotions in customer-agent interaction. They used inference rules to detect human emotions regarding the actual intentions of the customer using recorded CC calls given in the Polish language. Another work in [142] introduced an evidence-based machine learning framework for the automatic detection of subjective calls. They used deep neural network to assess a corpus of seven hours of recorded calls from a real-estate CC and achieved an accuracy of 75% for subjectivity detection (Table 4).

6 Sentiment analysis experiments

This section aims to outline the sentiment analysis experiments conducted on the publicly available dataset that resonates with the structure and form of the CC data, demonstrating the effectiveness of well-known algorithms. The code has been uploaded on GitHub¹ and can be used for reproducing the experiments.

¹ <https://github.com/SShah30-hue/sentiment-analysis-review>.

6.1 Dataset description

A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations (MELD)—an enhanced and extended EmotionLines dataset has been selected for this experiment [143, 144]. MELD contains similar dialogue instances that are available in EmotionLines, but it encompasses audio and visual modality along with text also. MELD contains about 13,000 utterances from 1,400 dialogues from the TV series called 'Friends'. The textual part of the dataset included two label columns. The column 'Emotion' contained seven labels: neutral, joy, sadness, anger, surprise, fear, and disgust labels. The column 'Sentiment' contained three labels (positive, negative, and neutral) which are what were used in our experimental study. The audio part of the dataset was retrieved from converting MPEG-4 Part 14 files into a WAVE format. Our experiments used 9988, 1108, and 2608 audio files and textual utterances for training, development, and testing. The data was passed as a CSV file with columns for text and sentiment label and an audio file directory path. The statistics of the MELD dataset are presented in Table 5.

Table 5 Description of the MELD dataset used in the experiments

Data type	Description
Sentiment classes	Positive, neutral, negative
#training set	9988/8.72 h
#development set	1108/0.96 h
#testing set	2608/2.31 h
Elicitation	Play-acted

Table 6 Experimentation results of MELD audio dataset

Model	Loss	Weighted accuracy (%)	F1-score
2D CNN MFCC	1.04	50.3	0.46
2D CNN Mel spectrogram	1.02	51.3	0.42
Wav2vec 2.0 base	1.02	49.0	0.46
Wav2vec 2.0 large	1.02	50.0	0.47

Table 7 Experimentation results of MELD text dataset

Model	Loss	Weighted accuracy (%)	F1-score
BERT	0.84	70.0	0.70
ALBERT	0.85	69.1	0.69
RoBERTa	0.84	71.0	0.71

6.2 Results

Different models were experimented individually on two different data formats, i.e. audio and text. For audio—2D CNN and Wav2vec models were experimented, whereas for text—ALBERT, BERT, and RoBERTa models were experimented with. This eventually helped in comparing the model results of the MELD dataset and shortlisting the best-performing model for the fusion experiment. For audio data, 2D CNN was trained and deployed with MFCC and spectrogram features. The outperforming model was the one trained on Wav2vec 2.0 large, followed by Wav2vec 2.0 base and 2D CNN MFCC as shown in Table 6. For text, RoBERTa’s performance was notably better than the rest of the models as shown in Table 7. The RoBERTa model performed even better when the output of its last four hidden layers was concatenated and used for predictions as opposed to using the last layer only, i.e. pooler output. Following individual model training and deployment, the audio and text embeddings from best-performing models were fused and then loaded into the RoBERTa model. Despite our initial expectations, the fused embeddings (audio and text) loaded RoBERTa model did not exhibit any improvement over the text-only RoBERTa model, as demonstrated by the results in 8. We evaluated all models using multiple metrics such as weighted accuracy, loss, and F1-score. The number of training epochs used was set to 10, the learning rate was set to 2e-5 and the batch size was set to 16 for pre-trained models—text input and 20 training epochs, 0.001 learning rate, and 16 batch size for 2D CNN and Wav2vec 2.0—audio input.

Table 8 Experimentation results of fusion of audio and text MELD dataset

Model + Configuration	Loss	Weighted accuracy (%)	F1-score
RoBERTa-1 layer	0.92	62.9	0.60
RoBERTa-2 layer	0.87	66.9	0.66
RoBERTa-3 layer	0.86	68.2	0.67
RoBERTa-4 layer	0.86	68.4	0.68
RoBERTa-5 layer	0.86	68.2	0.67
RoBERTa-last four hidden layers-3 layer	0.85	68.7	0.68
RoBERTa-last four hidden layers-4 layer	0.85	68.4	0.68
RoBERTa-last four hidden layers-5 layer	0.86	68.3	0.68

Table 9 Comparison table of all models experimented

Model	Weighted accuracy (%)
2D CNN MFCC	50.3
2D CNN Mel spectrogram	51.3
Wav2vec 2.0 base	49.0
Wav2vec 2.0 large	50.0
BERT	70.0
ALBERT	69.1
RoBERTa	71.0
RoBERTa-last four hidden layers (audio+text)	68.7

6.3 Discussion

The primary objective of this experiment was to showcase the advantages of using pre-trained transformer models for sentiment analysis. We first used audio and text data separately and then combined their features to evaluate the performance of different models as shown in Table 9. The results presented in this study clearly demonstrate the potential of utilizing the latest NLP techniques to achieve better results. This experimental study could serve as a guiding framework for developing sentiment analysis systems in the future. It can also help CC organizations in driving innovation by leveraging the latest models. However, it should be noted that this experiment is a proof of concept, and more research is needed to develop a production-ready sentiment analysis system.

The results of this study indicate that transformer models perform better than classical ML and language models, particularly for textual data. Hence, we recommend the use of transformer models for sentiment analysis tasks. However,

the performance of the audio data was not satisfactory, indicating a need to explore a wider range of features in future studies. Some of the limitations of this experiment include the small dataset size, transcription quality, and lack of better audio-quality data. Future studies could focus on addressing these limitations to further improve the accuracy of sentiment analysis systems. Also, the focus should be given to implementing newer advanced transformer-based language models. Overall, this study has shown the potential of transformer models in the field of sentiment analysis and offers valuable insights for future research.

7 Challenge and solutions of NLP in CCs

In this study, a systematic review of a wide variety of NLP techniques applied in CCs was completed. Our findings indicate that NLP methods have been applied more on a few key precise tasks of the CC operations. The outcome of this study does not apply necessarily to all other NLP-related studies but to those studies that have been shortlisted for this review paper. Despite the continuous and rapid improvement in NLP technology, its application in the CCs domain is still limited. In this section, we discuss various challenges in integrating NLP in CCs and highlight some potential solutions.

Firstly, multiple publications [8, 9, 54, 64] have cited the challenge of using massive amounts of CC data. This review paper conforms with those publications as it is a critical gap that needs to be addressed to steer CC automation. Specifically, CC data face labelling issues and thus require an organizational policy to be enacted and an efficient method to be utilized that automatically labels the data. The availability of labelled data is extremely scarce. Even when labelled data is available, it is either acted out, which may sound different than genuine emotions, or labelled independently, which is highly time-consuming and/or subjective. While there may be different databases for each interaction type, there are no studies that have shown a method in which data can be merged with their associated customer survey results and agent monitoring scores from CC supervisors to overcome the labelling issue. One of the most reoccurring themes identified in publications is that there is no unified database for CCs wherein all important data variables for each type of customer interaction are stored.

Second, there is a lack of data sharing and insufficient interoperability capabilities that has limited NLP and ML automation. Further, the existence of the data protection policy has made it difficult for organizations to share private data with 3rd parties including research institutions. Organizations store CC data mostly to aid them in case of legal lawsuits and other litigation fronts [9]. The demand for using the same data for enabling automation, personalisation

of services, and gaining a competitive advantage has grown in the last few decades only. Most organizations are still unclear on how to shift from their previous data storage and processing policies to new policies that essentially aid NLP and ML development [3].

Third, the issue of data quality also restricts the production of outputs from the NLP system such as transcription or audio processing [141]. Therefore, a number of techniques have been proposed in the open-source community related to enhancing the quality of standard telephony audio calls. However, it remains an issue hindering achieving high performance and is simply just not good enough, particularly when it comes to audio processing. Industry-wide efforts are needed to recognize this challenge and promote the use of tools and systems that can generate and store quality data.

Validating externally is crucial to ensuring model accuracy but it was not conducted in all studies reviewed in this paper. There could be many reasons but it is suspected it is mostly down to the unavailability of suitable datasets or unawareness of the gravity of external validation. The publications covered in this review paper have resorted to either private or publicly available data corpus mostly. The publicly available data corpus is mostly either acted data, i.e. actors who have recorded sentences and scripts from movies, news, or TV shows. A resemblance can be drawn in a few of them as their nature correlates with the CC domain generated data, i.e. conversational nature. In our study, we did not evaluate the quality of the real-life dataset used in some publications to build, assess, or test their proposed models. While not exactly related to this review, it must be noted that all real-life data limitations apply despite the approach employed. Nonetheless, when such data is used for ML-based research, how dependent proposed methods are on the data availability and structure must be known and a comprehensive evaluation of a data source helps in ensuring its appropriateness for the ML work. Similarly, it is recommended that all data variables present in the databases should be completely understood, including those variables that might possess predictive/prognostic value.

Beyond data complexities, there are a number of modelling strategies proposed that have been employed given specific CC tasks. The range of strategies that have been identified in the review papers implies there are many approaches, each proving beneficial to an extent. It has also been long known that there is no single algorithm that can produce desired results, instead of utilizing only one algorithm can often lead to uncertainty and variability. Also, due to the growth of multimodal data generated from CCs, it has become necessary to set a standard where multiple algorithms are considered while prototyping. While in some cases—depending on the CC task, one model may be enough to overcome data fitting issues as well as produce a more accurate output, the surety of that one model can be made through its novelty. Until more and more

advanced models are introduced in the future, the best practice would be to assess the quality of each language and machine learning model and evaluate their performance as well as when combined. Also, as NLP and ML development within the CC domain extends, the need to externally validate becomes more important. It would be otherwise difficult to generalize models without their application on CC domain data precisely.

Due to the nature of language, it keeps evolving and a set of rules-based inputs assigned to CC tasks have proven to be leading towards customer dissatisfaction [24]. On the other hand, it is now vastly demonstrated that NLP and ML algorithms can help to switch towards more cognitive-based systems that allow for more intelligent prediction and early reaction to customer needs [31]. However, the notion of NLP and ML completely replacing a human CSR team is still a long way off, especially until the CC data challenges are solved. Also, the attitudes of many towards AI in customer service are not widely favoured yet. For instance, 9/10 people have stated that chatbots should have the option to transfer to a human agent in the CC [145]. This means that there is still a need for human intervention. Having said that, there is no denying that NLP and ML have the potential to significantly improve the CC customer service capabilities but to truly fulfil its potential, cross-domain efforts are needed wherein experts from different core disciplines collaboratively solve its challenges and integrate NLP and ML models based on sophisticated linguistic and acoustic processing that is closer or even better than human agent [146]. This will help in minimizing the flaws in its implementation, ensure risks are efficiently managed, and deliver services efficiently.

Having reviewed papers that are directly related to the CC, it has become clear that significant research efforts are severely needed to precisely tackle the areas where recent breakthrough NLP and ML models can add value and at the same time suggest solutions for the above-mentioned issues. The challenges that have been mentioned above should be at the forefront while developing new strategies. While at the designing stage, state-of-the-art NLP and ML methods should be adopted that allow flexibility in integration. To ensure high-performance of those methods, new CC management policies and processes, especially regarding CC data labelling and conjoint, must become a frequent practice within CC, particularly when it comes to back-end processes.

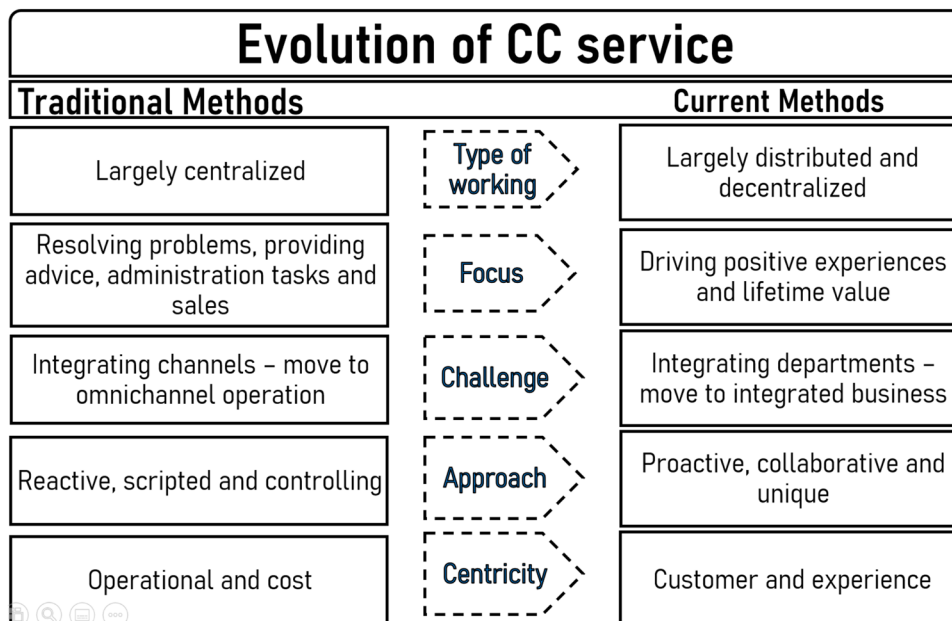
8 Future directions for CCs

Organizations are constantly challenged to keep pace with the changing needs and expectations of their customers. Among all departments, customer service has had to adapt and evolve the fastest in response to the new era of customer

requirements, the use of multiple communication channels, and the challenges posed by younger (“millennial” and “generation z”) employees. As the bridge between employees and customers, the customer service department plays a crucial role in continuously improving service delivery. Today’s customer service centres are modern and have progressed from voice-only channels to multi-channel and omnichannel platforms, from simple to multi-skilled workforce management, and from random to interaction-based analytics that captures the voice of the customer (VoC). The introduction of performance management, desktop guidance, automation of traditional customer service tasks, real-time authentication, bots, and customer journey analytics offer a range of solutions for the efficient functioning of call centres in today’s market [2]. Most organizations now offer cloud services, while providing distributed models of operation, allowing greater flexibility and silos opportunities within the business. Gartner forecasts that by 2024, there will be more cloud contact centre agents (9.2M) than premises-based agents (7.2M) [147]. While so many changes have emerged over the years, customer needs keep constantly changing. Therefore, continuous innovation is required from the CC organizations to help advance towards CCs that can provide idiosyncratic and cutting-edge customer service. The following points are worth considering when envisaging future CCs:

- In tomorrow’s customer service landscape, automation, analytics, workflow technology, and bots will play a significant role. However, organizations must not rely on assumptions but instead gather and utilize data effectively to stay updated and understand their customers’ perceptions [3]. To provide proactive support and personalized services, both historical and real-time data from various sources must be utilized. While smart bots may eventually provide optimal support, human agents with a wide range of skills will remain as valuable problem-solvers for situations that bots are not capable of resolving [29]. Consequently, future customer service will combine human and machine efforts, including automation and machine learning, with the option of escalation to human agents if necessary.
- Organizations must also understand the new demands from the next generation of agents who prefer decentralized operations [19]. It becomes paramount to recruit and retain the best agents and provide sufficient training, especially technical support in handling an array of channels while fulfilling customer needs. Therefore, agents will effectively play a defining role in the next era of CCs.
- CC data holds invaluable information, which can support organizations to build a connected enterprise and drive operations. CCs in the future will no longer solely focus

Fig. 7 Evolution of CC service



on problem resolution or campaign-based selling but more focused on promoting interactive experience hub, which can have profound effects on customer experiences [19, 148] (see Fig. 7). CC data can be both an opportunity and a threat. This means if the organization lacks the ability to analyse infinite volume, variety, and velocity of CC data for operational improvements and business performance, it could become difficult to strengthen its position in the market.

- Like most publicly accessible IT systems, call centres (CCs) are highly susceptible to cyber-attacks. Criminal enterprises find customer personal information particularly attractive, making CCs a prime target. This is mainly due to the various customer-account-related issues that call centres need to handle, which often require access to sensitive information, particularly financial data like billing details linked to a customer’s account. As a result, CCs are vulnerable to both internal and external security threats, including denial of service (DoS) attacks, hacking and data breaches, social engineering, and inappropriate access by internal CC staff [149, 150]. Shockingly, 30% of agents have access to customer payment information, even when not on the phone with them, and 42% of agents do not report data breaches [151]. For this reason, businesses need to improve their data privacy protocols. To prevent these threats, effective measures such as organizational practices, staff training, cultural changes, and secure technological solutions are essential [152].
- Just as the CC has evolved, NLP and ML in parallel have also significantly progressed. The recent advancements have brought a wide range of capabilities to CCs such

as ASR-based IVR systems have evolved to route calls with good accuracy. Newly proposed NLP models have demonstrated state-of-the-art results and are continuously being researched and implemented. Going forward, these models and more advanced models of the future will provide a real opportunity to precisely understand language and mine customer data [141]. Early adoption of these models into the CCs will help organizations in coping with the changing demands, delivering unique services, assimilating knowledge when employing new technologies, and supporting the transfer of efforts from people to intelligent systems, thus leading towards efficient automation of human tasks.

9 Conclusion

The purpose of this paper is to present a detailed study on the utilization of NLP and ML techniques in the CC domain. To the best of our knowledge, this is the first effort made towards achieving this goal. The paper aims to assist researchers and practitioners in comprehending the current gaps, overcoming challenges, and obtaining direction for developing an intelligent NLP system for CC. We have explored a range of models, techniques, and strategies employed in the application of ML and NLP. Additionally, we have assessed the effectiveness of the latest language models on the MELD dataset. Although NLP and ML are becoming standard practices for future CCs, they must tackle various issues outlined in Sect. 8. Furthermore, extensive research efforts are required to ensure that potential solutions are experimented with using CC domain data since this area

remains mostly unexplored. CC is on track to become the interaction hub for the digital enterprise, managing support, interaction, and data gathering in an increasingly complex and connected world. Organizations need to make structural reforms and address all complex issues to ensure the successful implementation of CC automation.

Data availability No new data were created. Also, the dataset used in the experiments can be downloaded from <https://affective-meld.github.io/>. The code for all the above experiments can be accessed from <https://github.com/SShah30-hue/sentiment-analysis-review>.

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Larson D, Chang V (2016) A review and future direction of agile, business intelligence, analytics and data science. *Int J Inf Manag* 36(5):700–710
- Roscow E, Moore R, Singh S (2020) Contact centre transformation-bring the future forward. [Accenture.com](https://www.accenture.com)
- Benjamin G, Berg J, Das AC, Gupta V (2019) How advanced analytics can help contact centers put the customer first. [mckinsey.com](https://www.mckinsey.com)
- Wong A, Plasek JM, Montecalvo SP, Zhou L (2018) Natural language processing and its implications for the future of medication safety: a narrative review of recent advances and challenges. *Pharmacother J Hum Pharmacol Drug Ther* 38(8):822–841
- Mocanu B-C, Filip I-D, Ungureanu R-D, Negru C, Dascalu M, Toma S-A, Balan T-C, Bica I, Pop F (2022) Odin ivr-interactive solution for emergency calls handling. *Appl Sci* 12(21):10844
- Wang L, Huang N, Hong Y, Liu L, Guo X, Chen G (2020) Effects of voice-based ai in customer service: evidence from a natural experiment
- Binza L, Budree A (2022) Towards a balanced natural language processing: a systematic literature review for the contact centre. In: *International conference on social implications of computers in developing countries*, pp 397–420. Springer
- Saberi M, Hussain OK, Chang E (2017) Past, present and future of contact centers: a literature review. *Bus Process Manag J* 2:58
- Saberi M, Karduck A, Hussain OK, Chang E (2016) Challenges in efficient customer recognition in contact centre: state-of-the-art survey by focusing on big data techniques applicability. In: *2016 international conference on intelligent networking and collaborative systems INCoS*, pp 548–554. IEEE
- Fernandes S (2021) Omnichannel contact center: a guide for 2021. [Lifesize](https://www.lifesize.com)
- Anderson EW, Fornell C, Rust RT (1997) Customer satisfaction, productivity, and profitability: differences between goods and services. *Mark Sci* 16(2):129–145
- Dhesi A, Gupta P, Kumar A, Parija GR, Roy S (2011) Contact center scheduling with strict resource requirements. In: *International conference on integer programming and combinatorial optimization*, pp 156–169. Springer
- Reddy T (2017) How chatbots can help reduce customer service costs by 30%. In: *The analytics maturity model IT best kept secret is optimization*
- Armony M, Maglaras C (2004) On customer contact centers with a call-back option: customer decisions, routing rules, and system design. *Oper Res* 52(2):271–292
- Owens AR (2014) Exploring the benefits of contact centre offshoring: a study of trends and practices for the Australian business sector. *Int J Hum Resource Manag* 25(4):571–587
- Soujanya M, Kumar S (2010) Personalized ivr system in contact center. In: *2010 international conference on electronics and information engineering*, vol 1, pp 1–453. IEEE
- Buesing E, Gupta V, Kleinstein B, Mukhopadhyay S (2019) Getting the best customer service from your ivr: Fresh eyes on an old problem. [mckinsey.com](https://www.mckinsey.com)
- Suhm B, Bers J, McCarthy D, Freeman B, Getty D, Godfrey K, Peterson P (2002) A comparative study of speech in the call center: natural language call routing vs. touch-tone menus. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, pp 283–290
- Gans N, Koole G, Mandelbaum A (2003) Telephone call centers: tutorial, review, and research prospects. *Manuf Serv Oper Manag* 5(2):79–141
- Bollapragada S, Nair SK (2010) Improving right party contact rates at outbound call centers. *Prod Oper Manag* 19(6):769–779
- Reichheld FF, Reichheld FR (2001) *Loyalty rules! How today's leaders build lasting relationships*. Harvard Business Press, Boston
- Park Y, Gates SC (2009) Towards real-time measurement of customer satisfaction using automatically generated call transcripts. In: *Proceedings of the 18th ACM conference on information and knowledge management*, pp 1387–1396
- Brennan M, Benson S, Kearns Z (2005) The effect of introductions on telephone survey participation rates. *Int J Mark Res* 47(1):65–74
- Millard N (2006) Learning from the 'wow' factor-how to engage customers through the design of effective affective customer experiences. *BT Technol J* 24(1):11–16
- Parameswaran AG (2013) *Human-powered data management*. Stanford University, California
- Awasthi P, Sangle PS (2012) Adoption of crm technology in multichannel environment: a review 2006–2010. *Bus Process Manag J* 2:579
- Kirkpatrick K (2017) Ai in contact centers. *Commun ACM* 60(8):18–19
- Karakus B, Aydin G (2016) Call center performance evaluation using big data analytics. In: *2016 international symposium on networks, computers and communications ISNCC*, pp 1–6. IEEE
- Quarteroni S (2018) Natural language processing for industrial applications. *Spektrum* 41:105
- Hirschberg J, Manning CD (2015) Advances in natural language processing. *Science* 349(6245):261–266
- Reshamwala A, Mishra D, Pawar P (2013) Review on natural language processing. *IRACST Eng Sci Technol Int J ESTIJ* 3(1):113–116

32. Kalyanathaya KP, Akila D, Rajesh P (2019) Advances in natural language processing—a survey of current research trends, development tools and industry applications. *Int J Recent Technol Eng* 7:199–202
33. Joseph SR, Hlomani H, Letsholo K, Kaniwa F, Sedimo K (2016) Natural language processing: A review. *Nat Lang Process* 6:207–210
34. Khurana D, Koli A, Khatter K, Singh S (2017) Natural language processing: state of the art, current trends and challenges. arXiv preprint [arXiv:1708.05148](https://arxiv.org/abs/1708.05148)
35. Alshawi H (1992) *The core language engine*. MIT press, London
36. Kamp H, Reyle U (2013) *From discourse to logic: introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory*, vol 42. Springer, Dordrecht
37. Mani I, Maybury MT (1999) *Advances in automatic text summarization*, vol 293. Camb MA
38. Yi J, Nasukawa T, Bunescu R, Niblack W (2003) Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques. In: *Third IEEE international conference on data mining*, pp 427–434. IEEE
39. Liddy ED (2001) *Natural language processing*. Marcel Decker, Inc., New York
40. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
41. Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61:85–117
42. Fayek HM, Lech M, Cavendon L (2017) Evaluating deep learning architectures for speech emotion recognition. *Neural Netw* 92:60–68
43. Hallowell R (1996) The relationships of customer satisfaction, customer loyalty, and profitability: an empirical study. *Int J Serv Ind Manag* 5:214
44. Ranaweera C, Prabhu J (2003) The influence of satisfaction, trust and switching barriers on customer retention in a continuous purchasing setting. *Int J Serv Ind Manag* 5:68
45. Luque J, Segura C, Sanchez A, Umbert M, Galindo LA (2017) The role of linguistic and prosodic cues on the prediction of self-reported satisfaction in contact centre phone calls. In: *INTER-SPEECH*, pp 2346–2350
46. Sun J, Xu W, Yan Y, Wang C, Ren Z, Cong P, Wang H, Feng J (2016) Information fusion in automatic user satisfaction analysis in call center. In: *2016 8th international conference on intelligent human-machine systems and cybernetics IHMSC*, vol 1, pp 425–428. IEEE
47. Kanchinadam T, Meng Z, Bockhorst J, Singh V, Fung G (2021) Graph neural networks to predict customer satisfaction following interactions with a corporate call center. arXiv preprint [arXiv:2102.00420](https://arxiv.org/abs/2102.00420)
48. Ando A, Masumura R, Kamiyama H, Kobashikawa S, Aono Y, Toda T (2020) Customer satisfaction estimation in contact center calls based on a hierarchical multi-task model. *IEEE/ACM Trans Audio Speech Lang Process* 28:715–728
49. Morrison D, Wang R, De Silva LC (2007) Ensemble methods for spoken emotion recognition in call-centres. *Speech Commun* 49(2):98–112
50. Priyadarshana Y, Gunathunga K, Perera KNN, Ranathunga L, Karunaratne P, Thanthriwatta T (2015) Sentiment analysis: measuring sentiment strength of call centre conversations. In: *2015 IEEE international conference on electrical, computer and communication technologies ICECCT*, pp 1–9. IEEE
51. Sehgal RR, Agarwal S, Raj G (2018) Interactive voice response using sentiment analysis in automatic speech recognition systems. In: *2018 international conference on advances in computing and communication engineering ICACCE*, pp 213–218. IEEE
52. Palicki S-K, Fouad S, Adedoyin-Olowe M, Abdallah ZS (2021) Transfer learning approach for detecting psychological distress in brexit tweets. In: *Proceedings of the 36th annual ACM symposium on applied computing*, pp 967–975
53. Fouad S, Alkoheji E (2023) Sentiment analysis for women in stem using twitter and transfer learning models. In: *2023 IEEE 17th international conference on semantic computing (ICSC)*, pp 227–234. IEEE
54. Godbole S, Roy S (2008) Text to intelligence: building and deploying a text mining solution in the services industry for customer satisfaction analysis. In: *2008 IEEE international conference on services computing*, vol 2, pp 441–448. IEEE
55. Devillers L, Vidrascu L (2006) Real-life emotions detection with lexical and paralinguistic cues on human-human call center dialogs. In: *Ninth international conference on spoken language processing*
56. Gupta N, Gilbert M, Fabbriozzo GD (2013) Emotion detection in email customer care. *Comput Intell* 29(3):489–505
57. Vidrascu L, Devillers L (2005) Detection of real life emotions in call centers. In: *Ninth European conference on speech communication and technology*
58. Zhou H, Huang M, Zhang T, Zhu X, Liu B (2018) Emotional chatting machine: emotional conversation generation with internal and external memory. In: *Thirty-second AAAI conference on artificial intelligence*
59. Nwe TL, Foo SW, De Silva LC (2003) Speech emotion recognition using hidden markov models. *Speech Commun* 41(4):603–623
60. Devillers L, Vasilescu I (2004) Reliability of lexical and prosodic cues in two real-life spoken dialog corpora. In: *LREC*
61. Gamon M (2004) Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In: *COLING 2004: Proceedings of the 20th international conference on computational linguistics*, pp 841–847
62. Gupta P, Rajput N (2007) Two-stream emotion recognition for call center monitoring. In: *Eighth annual conference of the international speech communication association*. Citeseer
63. Vidrascu L, Devillers L (2007) Five emotion classes detection in real-world call center data: the use of various types of paralinguistic features. In: *Proceedings of international workshop on paralinguistic speech between models and data, ParaLing*
64. Godbole S, Roy S (2008) Text classification, business intelligence, and interactivity: automating c-sat analysis for services industry. In: *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining*, pp 911–919
65. Devillers L, Vaudable C, Chastagnol C (2010) Real-life emotion-related states detection in call centers: a cross-corpora study. In: *Eleventh annual conference of the international speech communication association*
66. Nomoto N, Tamoto M, Masataki H, Yoshioka O, Takahashi S (2011) Anger recognition in spoken dialog using linguistic and para-linguistic information. In: *Twelfth annual conference of the international speech communication association*
67. Polzehl T, Schmitt A, Metz F, Wagner M (2011) Anger recognition in speech using acoustic and linguistic cues. *Speech Commun* 53(9–10):1198–1209
68. Erden M, Arslan LM (2011) Automatic detection of anger in human-human call center dialogs. In: *Twelfth annual conference of the international speech communication association*
69. Vaudable C, Devillers L (2012) Negative emotions detection as an indicator of dialogs quality in call centers. In: *2012 IEEE international conference on acoustics, speech and signal processing ICASSP*, pp 5109–5112. IEEE
70. Galanis D, Karabetsos S, Koutsombogera M, Papageorgiou H, Esposito A, Riviello M-T (2013) Classification of emotional speech units in call centre interactions. In: *2013 IEEE*

- 4th international conference on cognitive infocommunications CogInfoCom, pp 403–406. IEEE
71. Amarakeerthi S, Morikawa C, Nwe TL, De Silva LC, Cohen M (2013) Cascaded subband energy-based emotion classification. *IEEJ Trans Electron Inf Syst* 133(1):200–210
 72. Chakraborty R, Pandharipande M, Kopparapu S (2015) Event based emotion recognition for realistic non-acted speech. In: *TENCON 2015-2015 IEEE region 10 conference*, pp 1–5. IEEE
 73. Chowdhury SA, Stepanov EA, Riccardi G, et al. (2016) Predicting user satisfaction from turn-taking in spoken conversations. In: *Interspeech*, pp 2910–2914
 74. Chakraborty R, Pandharipande M, Kopparapu SK (2016) Mining call center conversations exhibiting similar affective states. In: *Proceedings of the 30th Pacific Asia conference on language, information and computation: posters*, pp 545–553
 75. Cong P, Wang C, Ren Z, Wang H, Wang Y, Feng J (2016) Unsatisfied customer call detection with deep learning. In: *2016 10th international symposium on chinese spoken language processing ISCSLP*, pp 1–5. IEEE
 76. Segura C, Balcells D, Umberto M, Arias J, Luque J (2016) Automatic speech feature learning for continuous prediction of customer satisfaction in contact center phone calls. In: *International conference on advances in speech and language technologies for Iberian languages*, pp 255–265. Springer
 77. Bockhorst J, Yu S, Polania L, Fung G (2017) Predicting self-reported customer satisfaction of interactions with a corporate call center. In: *Joint European conference on machine learning and knowledge discovery in databases*, pp 179–190. Springer
 78. Seng KP, Ang L-M (2017) Video analytics for customer emotion and satisfaction at contact centers. *IEEE Trans Hum-Mach Syst* 48(3):266–278
 79. Siriwardhana S, Kaluarachchi T, Billingham M, Nanayakkara S (2020) Multimodal emotion recognition with transformer-based self supervised feature fusion. *IEEE Access* 8:528
 80. Pepino L, Riera P, Ferrer L (2021) Emotion recognition from speech using wav2vec 2.0 embeddings. *arXiv preprint arXiv:2104.03502*
 81. Plaza M, Kazała R, Koruba Z, Kozłowski M, Lucińska M, Sitek K, Szyka J (2022) Emotion recognition method for call/contact centre systems. *Appl Sci* 12(21):10951
 82. Ando A, Murata Y, Masumura R, Suzuki S, Makishima N, Moriya T, Ashihara T, Sato H (2022) Customer satisfaction estimation using unsupervised representation learning with multi-format prediction loss. In: *ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp 8497–8501. IEEE
 83. El Ayadi M, Kamel MS, Karray F (2011) Survey on speech emotion recognition: features, classification schemes, and databases. *Pattern Recogn* 44(3):572–587
 84. Plaza M, Trusz S, Keczkowska J, Boksa E, Sadowski S, Koruba Z (2022) Machine learning algorithms for detection and classifications of emotions in contact center applications. *Sensors* 22(14):5311
 85. Schapire RE, Singer Y (2000) Boostexter: a boosting-based system for text categorization. *Mach Learn* 39(2):135–168
 86. Chu-Carroll J, Carpenter B (1998) Dialogue management in vector-based call routing. In: *COLING 1998 Volume 1: The 17th international conference on computational linguistics*
 87. Lee C-H, Carpenter B, Chou W, Chu-Carroll J, Reichl W, Saad A, Zhou Q (2000) On natural language call routing. *Speech Commun* 31(4):309–320
 88. Kuo H-KJ, Lee C-H (2001) A portability study on natural language call steering. In: *Seventh European conference on speech communication and technology*
 89. Wright JH, Gorin AL, Riccardi G (1997) Automatic acquisition of salient grammar fragments for call-type classification. In: *Fifth European conference on speech communication and technology*
 90. Stolletz R, Helber S (2004) Performance analysis of an inbound call center with skills-based routing. *OR Spectrum* 26(3):331–352
 91. Kuo H-K, Lee C-H (2003) Discriminative training of natural language call routers. *IEEE Trans Speech Audio Process* 11(1):24–35
 92. Allwein EL, Schapire RE, Singer Y (2000) Reducing multiclass to binary: a unifying approach for margin classifiers. *J Mach Learn Res* 5:113–141
 93. Haffner P, Tur G, Wright JH (2003) Optimizing svms for complex call classification. In: *2003 IEEE international conference on acoustics, speech, and signal processing, 2003. Proceedings. ICASSP'03, vol 1*, pp 1–3. IEEE
 94. Kuo H-KJ, Lee C-H, Zitouni I, Fosler-Lussier E, Ammicht E (2002) Discriminative training for call classification and routing. In: *Seventh international conference on spoken language processing*
 95. Zitouni I, Kuo H-KJ, Lee C-H (2003) Boosting and combination of classifiers for natural language call routing systems. *Speech Commun* 41(4):647–661
 96. Ali AR (2011) Intelligent call routing: optimizing contact center throughput. In: *Proceedings of the eleventh international workshop on multimedia data mining*, pp 1–9
 97. Jorge S, Pereira C, Novais P (2020) Intelligent call routing for telecommunications call-centers. In: *International conference on intelligent data engineering and automated learning*, pp 316–328. Springer
 98. Koromyslova A, Semenkina M, Sergienko R (2017) Feature selection for natural language call routing based on self-adaptive genetic algorithm. In: *IOP conference series: materials science and engineering*, vol 173. IOP Publishing
 99. Tyson N, Matula V (2004) Improved lsi-based natural language call routing using speech recognition confidence scores. In: *Second IEEE international conference on computational cybernetics, 2004. ICCS 2004*, pp 409–413. IEEE
 100. Tran TK, Pham DM, Van Huynh B (2016) Towards building an intelligent call routing system. *Int J Adv Comput Sci Appl* 7(1):528
 101. Rustamov S, Mustafayev E, Clements MA (2018) Context analysis of customer requests using a adaptive neuro fuzzy inference system and hidden Markov models in the natural language call routing problem. *Open Eng* 8(1):61–68
 102. Yang W, Tan L, Lu C, Cui A, Li H, Chen X, Xiong K, Wang M, Li M, Pei J, et al. (2019) Detecting customer complaint escalation with recurrent neural networks and manually-engineered features. In: *Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics: human language technologies, Vol 2 (Industry Papers)*, pp 56–63
 103. Ilk N, Shang G, Goes P (2020) Improving customer routing in contact centers: an automated triage design based on text analytics. *J Oper Manag* 66(5):553–577
 104. Tang M, Pellom B, Hacıoglu K (2003) Call-type classification and unsupervised training for the call center domain. In: *2003 IEEE workshop on automatic speech recognition and understanding IEEE Cat. No. 03EX721*, pp 204–208. IEEE
 105. Douglas S, Agarwal D, Alonso T, Bell RM, Gilbert M, Swayne DF, Volinsky C (2005) Mining customer care dialogs for daily news. *IEEE Trans Speech Audio Process* 13(5):652–660
 106. Mishne G, Carmel D, Hoory R, Roytman A, Soffer A (2005) Automatic analysis of call-center conversations. In: *Proceedings of the 14th ACM international conference on information and knowledge management*, pp 453–459
 107. Roy S, Subramaniam LV (2006) Automatic generation of domain models for call-centers from noisy transcriptions. In: *Proceedings*

- of the 21st international conference on computational linguistics and 44th annual meeting of the association for computational linguistics, pp 737–744
108. Hastie H, Prasad R, Walker M (2002) What's the trouble: automatically identifying problematic dialogues in darpa communicator dialogue systems. In: Proceedings of the 40th annual meeting of the association for computational linguistics, pp 384–391
 109. Walker MA, Langkilde-Geary I, Hastie HW, Wright J, Gorin A (2002) Automatically training a problematic dialogue predictor for a spoken dialogue system. *J Artif Intell Res* 16:293–319
 110. Garnier-Rizet M, Adda G, Cailliau F, Gauvain J-L, Guillemin-Lanne S, Lamel L, Vanni S, Waast-Richard C, et al. (2008) Call-surf: automatic transcription, indexing and structuration of call center conversational speech for knowledge extraction and query by content. In: LREC
 111. Hu H-L, Chen Y-L (2008) Mining typical patterns from databases. *Inf Sci* 178(19):3683–3696
 112. Chen M-C, Chen L-S, Hsu C-C, Zeng W-R (2008) An information granulation based data mining approach for classifying imbalanced data. *Inf Sci* 178(16):3214–3227
 113. Chen Y, Tsai FS, Chan KL (2008) Machine learning techniques for business blog search and mining. *Expert Syst Appl* 35(3):581–590
 114. Jackson TW, Tedmori S, Hinde CJ, Bani-Hani AI (2012) The boundaries of natural language processing techniques in extracting knowledge from emails. *J Emerg Technol Web Intell* 4(2):119–127
 115. McDonough J, Ng K, Jeanrenaud P, Gish H, Rohlicek JR (1994) Approaches to topic identification on the switchboard corpus. In: Proceedings of ICASSP'94. In: IEEE international conference on acoustics, speech and signal processing, vol 1, pp 1–385. IEEE
 116. Schwartz RM, Imai T, Kubala F, Nguyen L, Makhoul J (1997) A maximum likelihood model for topic classification of broadcast news. In: Eurospeech
 117. Nasukawa T, Nagano T (2001) Text analysis and knowledge mining system. *IBM Syst J* 40(4):967–984
 118. Padmanabhan D, Kumamuru K (2007) Mining conversational text for procedures with applications in contact centers. *Int J Doc Anal Recognit IJDAR* 10(3–4):227–238
 119. Takeuchi H, Subramaniam LV, Nasukawa T, Roy S (2009) Getting insights from the voices of customers: conversation mining at a contact center. *Inf Sci* 179(11):1584–1591
 120. Paprzycki M, Abraham A, Guo R, Mukkamala S (2004) Data mining approach for analyzing call center performance. In: International conference on industrial, engineering and other applications of applied intelligent systems, pp 1092–1101. Springer
 121. Deslauriers A, L'Ecuyer P, Pichitlamken J, Ingolfsson A, Avramidis AN (2007) Markov chain models of a telephone call center with call blending. *Comput Oper Res* 34(6):1616–1645
 122. Uma AN, Sityaev D (2022) Comparing methods for extractive summarization of call centre dialogue. Springer, Berlin
 123. Puterman ML (2014) Markov decision processes: discrete stochastic dynamic programming. Wiley, New Jersey
 124. Guo W, Liang L, Deng T (2017) Topic mining for call centers based on a-lda and distributed computing. *Concurr Comput* 29(3):245
 125. Lam S, Chen C, Kim K, Wilson G, Crews JH, Gerber MS (2019) Optimizing customer-agent interactions with natural language processing and machine learning. In: 2019 systems and information engineering design symposium SIEDS, pp 1–6. IEEE
 126. Koppurapu SK (2015) Non-linguistic analysis of call center conversations. Springer, Cham
 127. Gilbert M, Wilpon JG, Stern B, Di Fabrizio G (2005) Intelligent virtual agents for contact center automation. *IEEE Signal Process Mag* 22(5):32–41
 128. Levin E, Pieraccini R, Eckert W (2000) A stochastic model of human-machine interaction for learning dialog strategies. *IEEE Trans Speech Audio Process* 8(1):11–23
 129. Pawlik Ł, Plaza M, Deniziak S, Boksa E (2022) A method for improving bot effectiveness by recognising implicit customer intent in contact centre conversations. *Speech Commun* 143:33–45
 130. Matic R, Kabiljo M, Zivkovic M, Cabarkapa M (2021) Extensible chatbot architecture using metamodels of natural language understanding. *Electronics* 10(18):2300
 131. Amer E, Hazem A, Farouk O, Louca A, Mohamed Y, Ashraf M (2021) A proposed chatbot framework for Covid-19. In: 2021 international mobile, intelligent, and ubiquitous computing conference (MIUCC), pp 263–268. IEEE
 132. Judson TJ, Odisho AY, Young JJ, Bigazzi O, Steuer D, Gonzales R, Neinstein AB (2020) Implementation of a digital chatbot to screen health system employees during the Covid-19 pandemic. *J Am Med Inform Assoc* 27(9):1450–1455
 133. Wu C-H, Yan G-L, Lin C-L (1998) Spoken dialogue system using corpus-based hidden Markov model. In: Fifth international conference on spoken language processing
 134. Aida-zade K, Rustamov S, Mustafayev E, Aliyeva N (2012) Human-computer dialogue understanding hybrid system. In: 2012 international symposium on innovations in intelligent systems and applications, pp 1–5. IEEE
 135. Chinaei HR, Chaib-draa B, Lamontagne L (2009) Learning user intentions in spoken dialogue systems. In: ICAART, pp 107–114
 136. Salvador V, Andrade M, Kawamoto A (2007) Fuzzy theory applied on the user modeling in speech interface. In: IADIS international conference interfaces and human computer interaction, pp 201–205
 137. Subasic P, Huettner A (2001) Affect analysis of text using fuzzy semantic typing. *IEEE Trans Fuzzy Syst* 9(4):483–496
 138. Garoufi K (2014) Planning-based models of natural language generation. *Lang Linguist Compass* 8(1):1–10
 139. Cuayahuitl H, Dethlefs N (2011) Spatially-aware dialogue control using hierarchical reinforcement learning. *ACM Trans Speech Lang Process TSLP* 7(3):1–26
 140. Dethlefs N, Cuayahuitl H (2015) Hierarchical reinforcement learning for situated natural language generation. *Nat Lang Eng* 21(3):391–435
 141. Stepanov M, Muzata A, Zyuzin V, Kostina N, Shishkin M (2021) Estimation of contact center performance measures in case of overload and chatbot implementation. In: 2021 systems of signals generating and processing in the field of on board communications, pp 1–7. IEEE
 142. Ahmed A, Sivarajah U, Irani Z, Mahroof K, Charles V (2022) Data-driven subjective performance evaluation: an attentive deep neural networks model based on a call centre case. *Ann Oper Res* 5:1–32
 143. Poria S, Hazarika D, Majumder N, Naik G, Cambria E, Mihalcea R (2018) Meld: a multimodal multi-party dataset for emotion recognition in conversations. *arXiv preprint arXiv:1810.02508*
 144. Chen S-Y, Hsu C-C, Kuo C-C, Ku L-W, et al. (2018) Emotionlines: an emotion corpus of multi-party conversations. *arXiv preprint arXiv:1802.08379*
 145. Robyn: 12 top uses of artificial intelligence in the contact centre. *callcentrehelper.com* (2021)
 146. KS K SSS (2019) A survey of embeddings in clinical natural language processing. *arXiv preprint arXiv:1903.01039*
 147. Gartner: Forecast Analysis: Contact Center, Worldwide (2021). <https://www.gartner.com/en/documents/3995677>
 148. Andersen D (2021) The future of the call center: 6 predictions for 2022. *Invoca.com*
 149. Critchley T (2018) The threat on the end of the phone: the danger of contact centre agents. *Comput Fraud Secur* 2018(2):13–15
 150. Walter B (2020) Data security threats to call centers and compliance. <https://www.voicebase.com/data-security-threats-to-call-centers-and-compliance/>
 151. Sycurio: The state of data security in contact centres. *Sycurio Itd* (2022). <https://info.sycurio.com/download-state-security-contact-centres>

152. Sachs S (2021) Call center security best practices to protect customer data: TechTarget. TechTarget. <https://www.techtarget.com/searchcustomerexperience/tip/Call-center-security-best-practices-to-protect-customer-data>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.