Contents lists available at ScienceDirect

Natural Language Processing Journal

journal homepage: www.elsevier.com/locate/nlp

Deep learning-based natural language processing in human–agent interaction: Applications, advancements and challenges

Nafiz Ahmed [a], Anik Kumar Saha [a], Md. Abdullah Al Noman [a], Jamin Rahman Jim [a,b], M.F. Mridha [a,b,*], Md Mohsin Kabir [a,c]

[a] *Department of Computer Science and Engineering, American International University-Bangladesh, Dhaka 1229, Bangladesh*
[b] *Advanced Machine Intelligence Research Lab, Dhaka 1207, Bangladesh*
[c] *Faculty of Informatics, Eötvös Loránd University, H-1117 Budapest, Hungary*

ARTICLE INFO

Keywords:
Human–Agent Interaction
Deep learning
Natural language processing
Neural networks
Machine learning

ABSTRACT

Human–Agent Interaction is at the forefront of rapid development, with integrating deep learning techniques into natural language processing representing significant potential. This research addresses the complicated dynamics of Human–Agent Interaction and highlights the central role of Deep Learning in shaping the communication between humans and agents. In contrast to a narrow focus on sentiment analysis, this study encompasses various Human–Agent Interaction facets, including dialogue systems, language understanding and contextual communication. This study systematically examines applications, algorithms and models that define the current landscape of deep learning-based natural language processing in Human–Agent Interaction. It also presents common pre-processing techniques, datasets and customized evaluation metrics. Insights into the benefits and challenges of machine learning and Deep Learning algorithms in Human–Agent Interaction are provided, complemented by a comprehensive overview of the current state-of-the-art. The manuscript concludes with a comprehensive discussion of specific Human–Agent Interaction challenges and suggests thoughtful research directions. This study aims to provide a balanced understanding of models, applications, challenges and research directions in deep learning-based natural language processing in Human–Agent Interaction, focusing on recent contributions to the field.

1. Introduction

Human–Agent Interaction (HAI) is a multidisciplinary discipline that studies and designs interfaces for effective communication between humans and agents, as stated by Lewis (1998). This has become a prominent and trending study area, reflecting an increasing interest in understanding the dynamics between humans and agents. This increased attention is underscored by the growing importance of Deep Learning (DL) and Natural Language Processing (NLP) in the design of such interactions. The research article by Otter et al. (2020) stated that deep learning-based Natural Language Processing emerges as a critical technical enabler in this paradigm, providing sophisticated algorithms that allow robots and agents to understand intricate verbal nuances ranging from environmental clues to subtle alterations in attitude and meaning. Integrating DL-based NLP techniques can play a central role in improving the quality and nuances of HAI, and this synergy is increasingly gaining recognition for its transformative impact. At the same time, Findings by Kopp et al. (2021) provided empirical evidence of critical success factors in industrial Human–Agent Interaction and emphasize the need for technological and human-centric considerations for the effective integration of collaborative robots and agents in production environments. The use of deep learning-based NLP in HAI can create more successful collaboration between humans and agents, enhancing the capabilities and acceptance of robotic systems in various sectors, including healthcare and industrial automation. This work makes a valuable contribution to the wider discourse on HAI and highlights the ongoing evolution of this technology, which promises exciting developments in this dynamic field.

The term ‘Agent’ has become a buzzword in popular computing and artificial intelligence (AI) discourse, functioning as both a technical concept and a metaphor. This widespread use, however, has led to confusion reminiscent of challenges faced by other sensationalized terms like ’artificial intelligence’ itself (Nwana, 1996). In the realm of AI, we define an agent as an autonomous computational system that interacts with its environment to achieve specific goals (Russell and Norvig,

[*] Corresponding author at: Department of Computer Science and Engineering, American International University-Bangladesh, Dhaka 1229, Bangladesh.
E-mail addresses: nafizahmed273273@gmail.com (N. Ahmed), anikkumarsaha08@gmail.com (A.K. Saha), alnoman41330@gmail.com (M.A. Al Noman), jaminjim.aiub@gmail.com (J.R. Jim), firoz.mridha@aiub.edu (M.F. Mridha), mdmkabi@gmail.com (M.M. Kabir).

https://doi.org/10.1016/j.nlp.2024.100112
Received 19 March 2024; Received in revised form 3 October 2024; Accepted 3 October 2024

2016). This encompasses a spectrum from software programs to physical robots (Siciliano, 2008), all characterized by autonomy, reactivity, pro-activeness, and social ability (Wooldridge, 2009). In our research, we use “agent” and “robot” interchangeably, acknowledging their presence in both virtual and physical domains. This inclusive perspective allows us to explore how these intelligent entities can tackle complex challenges across various environments. By clarifying this definition, we aim to navigate the dual nature of ‘agent’ as both a precise technical term and a broader conceptual metaphor, enabling rich exploration while maintaining academic rigor. The Human–Agent Interaction revolution, spurred by robotics and artificial intelligence advancements, has experienced exponential growth due to the convergence of advanced algorithms and deep learning models such as BERT and GPT. As highlighted by Zhang et al. (2023a), BERT enhances contextual comprehension, whereas GPT contributes to natural and contextually relevant responses, significantly improving agent conversational intelligence. The fusion of Human–Agent Interaction and deep learning-based natural language processing drives transformative progress across various domains. A study by Li et al. (2023c) emphasized the integration of NLP in robotic assembly lines, enhancing communication and coordination between humans and agents, thereby increasing efficiency. Smith and Williams (2023) showcased virtual assistants equipped with NLP, revolutionizing customer interactions through an intuitive understanding and personalized responses. Education benefits from interactive learning experiences facilitated by robots with NLP, as exemplified by Brown and Lee (2022), which offers dynamic and personalized educational content. Healthcare applications, such as Mabu by Fadhil and others (2019), leverage NLP-based Deep Learning to enhance communication between patients and robots or agents while improving healthcare assistance. Additionally, in smart homes and retail environments, the integration of humanoid robots with advanced NLP capabilities, as exemplified by Amazon’s Alexa, Google Assistant, and Pepper (Luger and Sellen, 2016) and Hoffmann et al. (2019), underscores the trend towards interactive and intuitive communication in intelligent environments. These examples highlight the widespread impact and prevalence of combining HAI with NLP-based Deep Learning across diverse sectors. Despite the growing importance of deep learning-based NLP models in improving communication between humans and agents, the present literature lacks a focused study of the numerous applications and implications of these models in HAI. A comprehensive investigation of the application of DL-based NLP in HAI is required, including a thorough exploration of DL algorithms, preprocessing methods, datasets, diverse applications, related challenges, and future research directions. This survey is required to aggregate and synthesize disparate achievements across multiple sectors. The lack of a thorough survey limits the ability of the academic community to identify general trends, gaps, and future research objectives. Therefore, we conducted a systematic review focusing exclusively on recent cutting-edge research articles to gain insights into the latest developments in DL-based NLP in HAI. This study seeks to fill this gap by providing a comprehensive overview of the impactful integration of deep learning-based NLP in HAI, allowing for a more holistic understanding of the current state and prospects at this transformative intersection of technology and human–agent collaboration.

Motivation: This review paper addresses the growing need for a comprehensive overview of integrating deep learning-based NLP into HAI. Despite the significant impact of this integration in various sectors, there is a lack of consolidated insight into its applications, challenges and future directions. Through a systematic review of recent research, our study attempts to address this gap by providing valuable insights into methods, datasets, emerging trends and potential further developments. With this review, we aim to provide researchers and practitioners with a holistic understanding of the current state and prospects of DL-based NLP in HAI to facilitate informed decision-making and drive further innovation in this dynamic field.

Comparison with the closest survey currently available: In the absence of a direct study addressing the integration of DL-based NLP into HAI, this study conducts a comparative analysis with the most closely related studies in relevant fields. Although existing studies have addressed aspects of NLP and HAI separately, our study seeks to bridge this gap by offering a comprehensive examination of their overlap. By synthesizing insights from the literature on NLP, robotics, and HAI, our paper provides a unique perspective on the applications, methods, challenges, and future developments of DL-based NLP in the context of HAI. This comparative approach contributes to a more nuanced understanding of technological advances in Human–Agent Interaction, thereby fostering interdisciplinary research and innovation. Table 1 presents a comparison with the closest existing surveys done.

The main contributions of this study are:

- This study comprehensively explores the different application domains where Deep Learning-based Natural Language Processing and Human–Agent Interaction intersect. The analysis addresses the nuanced advances and innovations within this intersection and provides insights into various real-world applications that utilize the synergy of DL-based NLP in HAI.
- This study offers an in-depth exploration that provides valuable insights into data preprocessing methods for optimizing the model training. It also highlights commonly used datasets that are important for research and benchmarking in the context of DL-based NLP in HAI.
- A detailed examination of the predominant DL-based NLP algorithms used in HAI is presented, highlighting the individual strengths and limitations of each algorithm. This thorough investigation improves our understanding of the applications of DL-based NLP in the complex dynamics of Human–Agent Interaction.
- A thorough analysis that thoroughly explores and examines recent advances and contributions by researchers and highlights the latest experimental results shaping the evolving landscape of DL-based NLP in HAI.
- A comprehensive discussion discusses the challenges in the field of DL-based NLP in HAI and identifies future research opportunities to overcome these obstacles.

The key sections are systematically discussed throughout the remainder of this paper, each of which contributes to a comprehensive understanding of the study. Section 2 explains the systematic approach used to investigate the interplay between DL-based NLP and HAI. Section 3 describes the key data sources for subsequent analyses. The examination in Section 4 provides detailed insight into the various Deep Learning models used to improve natural language understanding in the context of HAI. The preprocessing methods are described in detail in Section 3.2. Section 5 describes the practical implementation of these models in various real-world scenarios. Section 6 rigorously breaks down the results and provides nuanced insights into this study’s findings. Section 7 critically discusses changes and advances in the field to ensure that the study is up-to-date. Finally, Section 8 captures the essence of the study by summarizing the main findings, reflecting on their implications, and suggesting possible avenues for future research in the field of DL-based NLP in HAI.

2. Survey methodology

In this study, a systematic literature review (SLR) was conducted based on the frameworks proposed by Keele et al. (2007) and Kitchenham et al. (2009). This review focuses exclusively on high-quality academic articles from reputable databases such as ScienceDirect, SpringerLink, ACM Digital Library, IEEE Xplore, and well-known conferences such as ICML, CVPR, and ICCV. The Identification of critical resources was based on the recommended guidelines shown in Fig. 1, which reflects the Preferred Reporting Items for Systematic

Table 1
Comparative Analysis of Recent closest DP-based NP on HAI surveys.

Ref.	Coverage				Contribution
	Deep Learning	NLP	HAI	DL based NLP	
Soori et al. (2023)	✓	✗	✓	✗	Provided a comprehensive overview of AI, ML and DL applications in advanced robotics, which include autonomous navigation, object recognition and more. It highlights their impact on various industries such as manufacturing, aviation and transportation and underlines their potential to increase productivity. It also identifies research gaps and provides insights into future investigations, contributing significantly to the understanding and advancement of the field of AI in robotics.
Mohammed and Hassan (2020)	✓	✗	✓	✗	Provided a comprehensive overview of the challenges in emotion recognition for Human–Agent Interaction and identifies sensor channels for emotion recognition. It provides a thorough literature review, outlines existing problems and gives recommendations for future work. It also discusses state-of-the-art advances and trends to support further research into the field of emotional HAI systems.
Patel and Patel (2021)	✓	✓	✗	✓	Gave an overview of deep learning architectures (CNN, RNN, Attention) in NLP and emphasizes the value of external knowledge in applying rules to dialogues. Syntactic and semantic analyses are covered, with a focus on word order and relationships. Successful uses of deep learning in tasks such as spam detection and machine translation are highlighted, providing a comprehensive overview of its use in various NLP applications.
Giachos et al. (2020)	✗	✗	✓	✗	Provided an overview of Human–Agent interfaces and highlights a gap in the ability of agents to understand commands based on timing cues. It discusses speech understanding, dialogue and decision making, but notes a lack of research in the area of speech generation. It also examines natural language interfaces in collaborative robotic systems and assisted living environments and provides insights into current advances and challenges in this field.
Mukherjee et al. (2022)	✓	✗	✓	✗	Provided a novel taxonomy for Human–Agent Interaction levels and explored machine learning methods applicable to adaptive collaborative agent in industrial environments. It addresses the gap in machine learning-based research on human-agent collaboration and examines communication modes, safety measures, motion prediction and manipulation techniques in this context. Overall, it provides a comprehensive overview that closes important knowledge gaps and enables advances in industrial human–agent collaboration.
Károly et al. (2020)	✓	✗	✓	✗	Summarized the challenges, model selection guidance and adaptability of deep learning in robotics. It discusses successful solutions, identifies suitable tasks for deep learning and explores high-level perception and training strategies, providing valuable insights for practitioners and researchers.
Caldera et al. (2018)	✓	✗	✓	✗	Evaluated deep learning techniques for robotic grasping, describes their improvements in grasp recognition and argues for one-shot recognition for real-time applications. It addresses data scarcity through transfer learning and discusses trends such as CNNs, DCNNs, custom neural networks, multimodal predictors and custom architectures. It offers insights into current challenges and future research perspectives in this evolving field.
This paper	✓	✓	✓	✓	Provided a comprehensive analysis of DP based NLP on HAI and a systematic review of all mentioned areas of DL, NLP, HAI focusing on the latest research articles

Reviews and Meta-Analyses (PRISMA) model. The criteria for inclusion or exclusion of papers in the review are clearly outlined in Table 2, thereby providing a clear and systematic approach to the selection process.

At the outset of this study, 592 articles were selected for review. The keywords used to select articles in the different databases are Deep Learning-based Natural Language Processing in Human–Agent Interaction, Deep Learning-based NLP in Human–Robot Interaction, Human–Computer Interaction, Sentiment Analysis, Emotion Recognition, NLP in Robotics, Dialogue Systems in Robotics. This selection was made in four distinct periods: February 2023 to March 2023, October 10, 2023 to October 25, 2023, November 15, 2023, to November 29, 2023, and a final round from December 21 2023 to January 10 2024. After a comprehensive assessment, 175 articles that explicitly dealt with application advances, user authentication, and authorization were included in the review. The deliberate focus on more recent articles

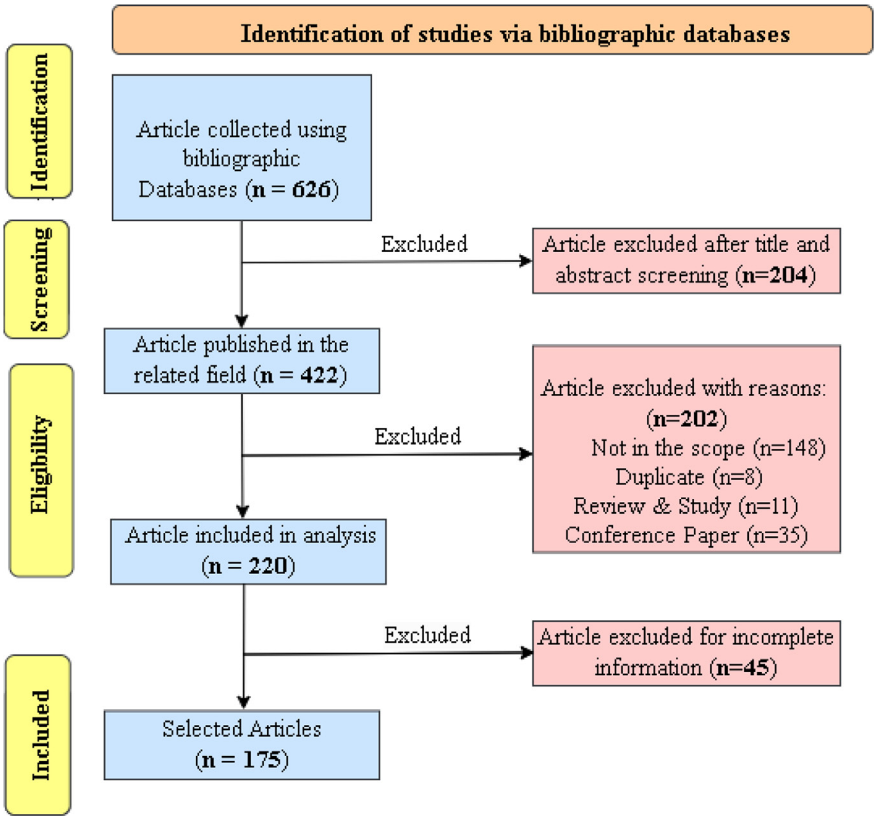
Identification of studies via bibliographic databases
Identification
Screening
Eligibility
Included
Article collected using bibliographic Databases (n = 626)
Excluded
Article excluded after title and abstract screening (n=204)
Article published in the related field (n = 422)
Excluded
Article excluded with reasons: (n=202)
Not in the scope (n=148)
Duplicate (n=8)
Review & Study (n=11)
Conference Paper (n=35)
Article included in analysis (n = 220)
Excluded
Article excluded for incomplete information (n=45)
Selected Articles (n = 175)

Fig. 1. PRISMA flow diagram of the article selection procedure.

Table 2
The inclusion and exclusion criteria used to choose articles are described in the table..

	Inclusion criteria	Exclusion criteria
Types of study	Original and review articles.	Thesis, white papers, communication letters, reports and editorials.
Source	Scientific articles published exclusively in academic journals (few articles from conferences)	Articles with inadequate information and reviews.
Publication year	2019–2024 (For Application, Result Analysis and State-of-art)	Irrelevant Articles
Language	English-language research articles	non-English-language research articles
Region	Not restricted to a particular region	–
Intervention	DL and NLP methods.	Traditional and statistical methods.
Settings	Deep learning based NLP	not related to Deep learning based NLP

Table 3
The table discusses the keywords used for article selection in different databases and the adjusted paper count for respective keywords.

Keyword	Paper count
Deep Learning based Natural Language Processing in Human–Agent Interaction	137
Deep Learning based NLP in Human–Agent Interaction	106
Deep Learning based NLP in Agent	90
Deep Learning based NLP in Computer	77
Dialogue Systems	62
Sentiment Analysis	50
Emotion Recognition	40
NLP in Robot	30
Total	**592**

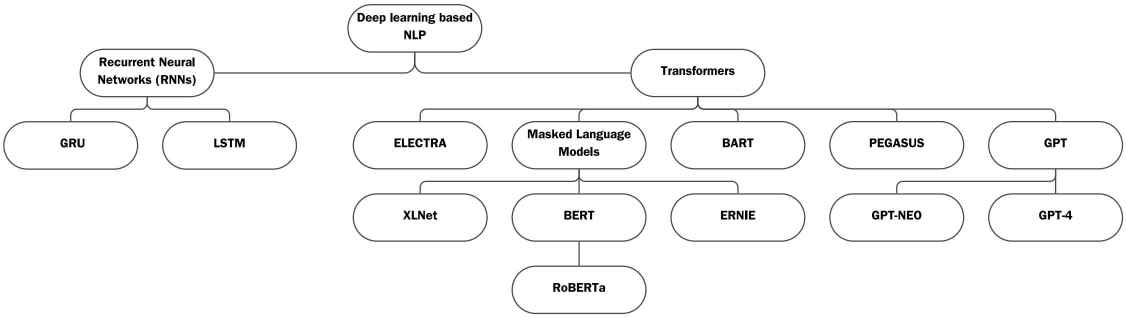

Fig. 2. Diverse Models of DP based NLP.

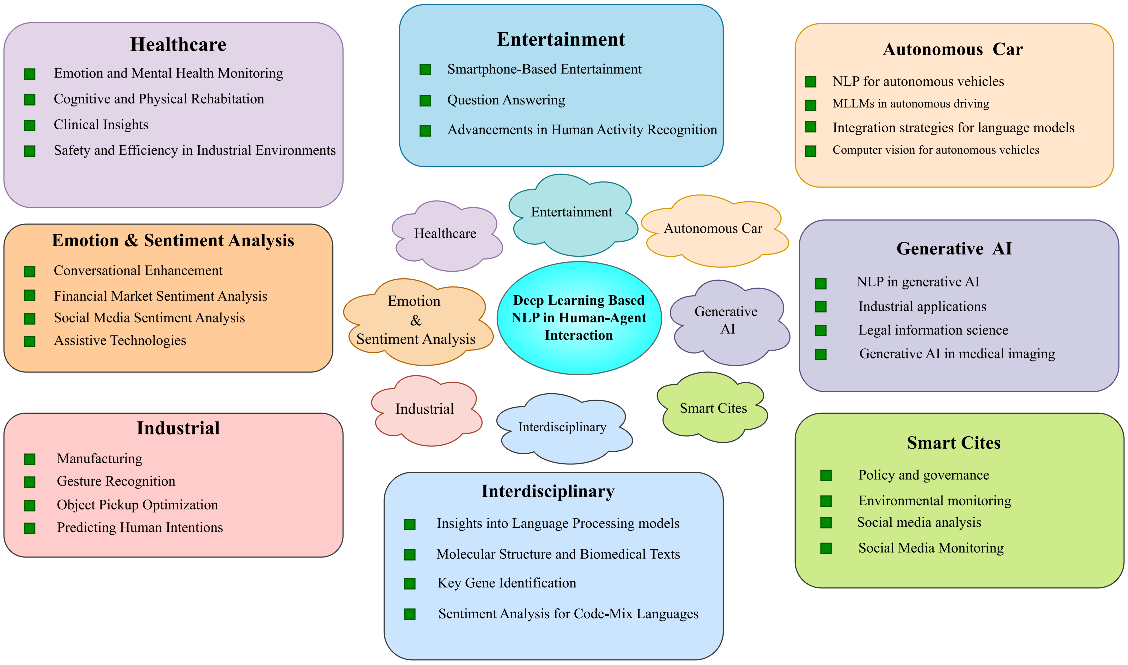
Healthcare
Emotion and Mental Health Monitoring
Cognitive and Physical Rehabilitation
Clinical Insights
Safety and Efficiency in Industrial Environments
Entertainment
Smartphone-Based Entertainment
Question Answering
Advancements in Human Activity Recognition
Autonomous Car
NLP for autonomous vehicles
MLLMs in autonomous driving
Integration strategies for language models
Computer vision for autonomous vehicles
Emotion & Sentiment Analysis
Conversational Enhancement
Financial Market Sentiment Analysis
Social Media Sentiment Analysis
Assistive Technologies
Healthcare
Entertainment
Autonomous Car
Emotion & Sentiment Analysis
Deep Learning Based NLP in Human-Agent Interaction
Generative AI
Industrial
Interdisciplinary
Smart Cites
Generative AI
NLP in generative AI
Industrial applications
Legal information science
Generative AI in medical imaging
Industrial
Manufacturing
Gesture Recognition
Object Pickup Optimization
Predicting Human Intentions
Interdisciplinary
Insights into Language Processing models
Molecular Structure and Biomedical Texts
Key Gene Identification
Sentiment Analysis for Code-Mix Languages
Smart Cites
Policy and governance
Environmental monitoring
Social media analysis
Social Media Monitoring

Fig. 3. The diverse applications of DP based NLP in HAI.

underlines our aim of offering an up-to-date and forward-looking review (see Table 3).

3. Datasets, pre-processing methods

This section examines the latest trends in DL-based NLP in HAI. It begins by exploring datasets tailored to HAI contexts and highlighting the challenges and opportunities that arise in this unique environment. It then discusses preprocessing strategies designed explicitly for DL-based NLP, emphasizing the delicate balance between preserving privacy and maximizing the utility of information. In addition, this section examines the current models used in HAI for their adaptability and effectiveness in improving Natural Language Processing in robotic systems. By presenting the findings from these three interrelated components, this section provides a concise overview of recent advances in Deep Learning for NLP in the context of HAI and provides valuable perspectives on the evolving landscape of HAI.

3.1. Dataset

The undeniable requirement for datasets in any given field cannot be overstated. Table 4 lists the essential datasets used in Deep Learning-based Natural Language Processing for Human–Agent Interaction, underlining their importance in driving research and innovation. It is noteworthy that the datasets compiled in this study represent a subset of the most frequently referenced datasets in the literature analyzed. Their widespread adoption and recurrence across research papers underscore their significance in advancing HAI research and applications.

3.2. Pre-processing methods

Pre-processing is vital in deep learning-based Natural Language Processing for HAI, converting raw text into machine-readable formats using data cleansing and tokenization techniques. Similar to sentiment analysis, these methods refine the input data and enhance model performance. Rigorous pre-processing ensures data integrity and contributes significantly to scientific advancements in optimizing NLP applications in HAI. Table 5 below elucidates prevalent pre-processing methods in this field.

4. State-of-the-art deep learning based NLP models

In the academic field of HAI using Deep Learning-based Natural Language Processing, this resource is a valuable guide for navigating

Table 4
Frequently used datasets & usage in recent research articles in Deep Learning-based NLP in Human–Agent Interaction..

Dataset Paper	Year	Name	Type	Domain	Feature	References
Lin et al. (2014)	2014	MS COCO dataset	Image	Interdisciplinary	A dataset of 328,000 photos of 91 easily identifiable objects for four-year-olds with 2.5 million labeled instances was subjected to detailed statistical analysis and performance evaluation, and differences from other datasets were identified	Golech et al. (2022) Amirian et al. (2021) Hwang et al. (2022) Kesavan et al. (2019)
shahhaard47 (2020)	2020	IMSDb	Text	Entertainment	A fine-tuned GPT-2 and BART models that incorporate specific genre tags to produce scripts tailored to each genre.	Dharaniya et al. (2023) Kim et al. (2018) Gross et al. (2021) Arikatla and Chinnapottu (2021)
Li (2023)	2022	Virtual-Assistant-Max	CSV	Industrial	The data is divided into 2 folders, They contain 10 and 6 intents respectively, each with 40 labels, which are used to train robust models for different user interactions in different service environments.	Li et al. (2023a)
Kunchukuttan et al. (2018)	2018	IIT-Bombay Eng-Hin Corpus	Text	Emotion & Sentiment Analysis	The dataset used for translations since 2016 includes 49K sentence pairs from various sources, with 1.65M segments from GNOME, KDE4 and TED presentations in the training set.	Vashistha et al. (2022)
Chen et al. (2015)	2015	ELDERLY-AT-HOME corpus.	Video	Emotion & Sentiment Analysis	This multimodal data set comprises 1516 utterances and 6593 words and represents a valuable resource for studying complex interactions in elderly care.	Shervedani et al. (2023)
Shen et al. (2018)	2018	Chinese KBQA Dataset	Text	Interdisciplinary	Largest Chinese KBQA dataset with 43 million subject–predicate–object triples and 6 million entities. It contains 14.6K training pairs and 9,870 test-question-answer pairs provided by Microsoft researchers and obtained from Baidu Encyclopedia and Fobox.	Su et al. (2019)
Poria et al. (2018a)	2018	MELD dataset	Text	Emotion & Sentiment Analysis	This dataset from the TV series "Friends" contains over 13K annotated utterances spread across around 1.5K dialogues with multiple speakers. It also plays a central role in improving human–computer interaction by enabling systems to better understand and respond to users' emotional states.	Peng et al. (2023) Poria et al. (2018a) Mohammad et al. (2023) Thakur et al. (2023) Verina et al. (2023)
dat (2023c)	2010	MSVD Corpus	Text	Entertainment	The data set comprises around 120K sentences collected in the summer of 2010. Mechanical Turk employees were paid to watch short video clips and summarize the action in a single sentence. This resulted in a data set with roughly parallel descriptions for over 2K video snippets.	Varma and Peter (2022) ¨Ozer et al. (2020) Phuc et al. (2022)
dat (2023b)	2020	MSR-VTT	Text	Entertainment	Consisting of 10K video clips in 20 categories, each annotated with 20 English sentences from Amazon Mechanical Turks. The dataset contains approximately 29K unique words in subtitles and uses 6K clips for training, 497 for validation and 2,990 for testing in the standard split.	Varma and Peter (2022) Rafiq et al. (2023) Islam et al. (2021)
Rajpurkar et al. (2016)	2016	SQuAD Dataset	Text	Interdisciplinary	Merging 100K questions from SQuAD1.1 with 50K opponent-generated unanswerable questions designed by crowd workers to be very similar to answerable questions.	Budiharto et al. (2020)
dat (2023a)	2014	n2c2 dataset	Text	Healthcare	The dataset consists of 1237 discharge reports from the Partners HealthCare Research Patient Data Repository.	Kumar et al. (2020)

(*continued on next page*)

Table 4 (*continued*).

Dataset Paper	Year	Name	Type	Domain	Feature	References
Rohrbach et al. (2015)	2015	MPII Human Pose Dataset	Image	Entertainment	The dataset comprises approximately 25K images, of which 15K are for training, 3,000 for validation and 7K for testing (with labels withheld). The dataset was extracted from YouTube videos and includes 410 human activities, each annotated with up to 16 body joints.	Cascianelli et al. (2018)
Deruyttere et al. (2019)	2019	Talk2Car	Image	Industrial	Task2Car contains 8349 training examples, 1163 validation examples and 2447 test examples.	Dong et al. (2023) Rufus et al. (2021)
Alomari et al. (2017)	2017	Leeds Robotic Commands(LRC)	Image	Industrial	dataset comprises 204 videos with around 17K images. It contains 1024 commands, an average of five per video. A variety of 51 objects are manipulated in the videos, including basic block shapes, fruit, cutlery and office supplies.	Alomari et al. (2022)

the complicated landscape of Deep Learning and Natural Language Processing, especially for beginners, and facilitates informed model selection. Given the paramount importance of DL-based NLP in HAI, these are elucidated below the comprehensive overview of key models in the Fig. 2 & Table 6 contributes to a holistic understanding of the diverse applications in this specialized field.

5. Applications

Human–Agent Interaction is a crucial milestone in robotics development and underlines its importance in our increasingly connected world. It goes beyond mere functionality, enabling robots to understand commands and the emotions & intentions contained in human language. The profound impact of HAI is felt in many areas, from healthcare, where robots can provide companionship and support, to education, where they can enhance learning experiences. In customer service, robots can better answer inquiries, whereas in manufacturing, they can work seamlessly with human workers to improve production efficiency. DL-based NLP is the linchpin of this transformation, enabling robots to recognize nuances, respond empathetically, and generate contextual responses. This advanced NLP technology streamlines interactions and improves accessibility, and thus, the user experience. Consequently, it accelerates the integration of robots into daily life and industry, bringing us closer to a future where HAI are efficient and remarkably human-centric, fostering an environment of trust and acceptance. For a comprehensive understanding of the diverse applications of HAI, Fig. 3 describes the significant impact of this interaction in various sectors such as healthcare, education, customer service, and manufacturing.

5.1. Autonomous car

Autonomous vehicles will transform HAI by improving safety and efficiency of transportation. Deep learning-based NLP enables these vehicles to understand and respond to verbal commands and cues, improving communication with passengers and pedestrians. By processing large amounts of textual data, NLP helps autonomous systems to adapt effectively to dynamic conditions. The integration of NLP increases the functionality and safety of autonomous vehicles and facilitates seamless interactions between humans and intelligent machines used for transportation. Table 7 presents some of the recent applications of HAI in emotion and sentiment analysis.

5.2. Emotion and sentiment analysis

The profound impact of HAI in the realm of emotion and sentiment analysis is pivotal, endowing robots with the capability to comprehend and respond effectively to human emotions. This integration not only elevates the user experience but also empowers robots to modify their behavior in response to the emotional cues of their human counterparts. Numerous applied research papers in this domain underscore the active role of HAI in furthering the progress in emotion and sentiment analysis, demonstrating the potential for the development of more emotionally intelligent and responsive robotic systems. Table 8 presents some of the recent applications of HAI in emotion and sentiment analysis.

5.3. Generative AI

Generative AI is a promising approach to HAI that improves the user experience by enabling robots to interact naturally. DL-based NLP improves communication between humans and agents. NLP models trained on extensive datasets help robots understand and produce human-like responses, which promotes trust. In addition, NLP advances enable robots to adapt to different social contexts, improving interactions in areas such as customer service and healthcare. Table 9 presents some of the recent applications of HAI in Generative AI.

5.4. Healthcare

In healthcare, Human–Agent Interaction is proving to be a transformative force. HAI has the potential to revolutionize patient care and encompasses a wide range of applications, from assisting medical staff during surgeries to providing emotional support to patients with mental illness. This review paper explores the multifaceted landscape of HAI in healthcare and examines its critical role in improving medical services, patient outcomes, and healthcare challenges. We address the various applications and innovations that demonstrate the growing importance of HAI and pave the way for a comprehensive understanding of its evolving role in the healthcare ecosystem. Table 10 spotlights recent use cases of HAI in the healthcare sector.

5.5. Entertainment

In the entertainment industry, interaction between humans and agents opens up new dimensions of user experience. Interactive robots create immersive and engaging entertainment and foster unique creative expression and shared social experiences. This integration reflects technological advancements and shapes the future of entertainment by providing audiences with engaging and interactive forms of enjoyment. Table 11 highlights some of the recent applications of HAI in entertainment.

5.6. Industrial

The importance of Human–Agent Interaction in industrial contexts is underlined by its transformative impact on efficiency, safety, and

Table 5
Common conventional pre-processing techniques and their application in current research on HAI..

Name	Description	Studies
Image Crop	Image cropping is a pre-processing technique that involves removing a specific part of an image, typically to focus on a region of interest or to resize it for further analysis or display.	Zhang et al. (2023b) Kushol et al. (2023) Fanjie et al. (2023)
Tokenization	Breaking down a text into smaller linguistic units such as words, phrases, or symbols to facilitate further analysis.	Halawani et al. (2023) Baghaei et al. (2023) Aldunate et al. (2022) Zhou et al. (2023b) Pandey et al. (2022)
Padding	Commonly used in sequence-based models such as Recurrent neural networks (RNNs) and transformers to ensure that all input sequences are the same length, allowing for efficient batch processing.	Ashraf et al. (2023) Inamdar et al. (2023) Jianan et al. (2023) Karasoy and Ballı (2022)
Lowercasing	Helpful for standardizing text to reduce vocabulary complexity and prevent the model from treating words with different capitalization as different entities.	Zaheer et al. (2023) Duong et al. (2023) Alshahrani et al. (2023)
Lemmatization & Stemming	Techniques for reducing words to their basic form in order to deal with word variations and thus reduce the size of the vocabulary and improve the generalization capability of the model.	Budiharto et al. (2021) Ayanouz et al. (2020) Guazzo et al. (2023) Wang et al. (2023b)
Normalization	Converting text data to a standard format by removing accents, special characters, or diacritics to enable consistent and uniform representation.	Abdalla et al. (2023) Matti and Yousif (2023) Kumar et al. (2022)
Noise removal	An important step in improving the quality of text data by removing irrelevant information such as special characters, symbols, or irrelevant words that do not contribute to the overall meaning.	Amaar et al. (2022) Khera (2023) Merdivan et al. (2019)
Feature-Extraction	Essential for capturing the most important information from the text data and creating meaningful representations that can be effectively used by the model to learn patterns and make predictions.	Johnston et al. (2023), Yohanes et al. (2023), Zhou et al. (2023a)
Word Embedding	A technique used to represent words as dense vectors in a multidimensional space, preserving semantic relationships between words and improving the model's ability to understand context and meaning.	Wan (2023) Chang and Ghamisi (2023) Wang et al. (2023a)
Stop-word removal	Removes frequently occurring words (e.g., 'and', 'the', 'is') that do not contain important information and are often ignored during analysis to reduce noise and increase processing speed.	Balouch and Hussain (2023) Mithun et al. (2023) Das and Saha (2022) Nijhawan et al. (2022)
Filter Alphanumeric	Helps clean up the text by removing non-alphabetic characters and numeric digits to ensure that the data focuses on the textual information relevant to the analysis.	Olthof et al. (2021) Niţoi et al. (2023) Das et al. (2023)
Vectorization	The process of converting text data into numeric vectors makes it suitable for various machine learning models requiring numeric input for processing.	Gupta et al. (2023) Xavier and Chen (2022) Marulli et al. (2021)
Part-of-Speech Tagging	This involves assigning grammatical tags to words in a sentence, allowing the model to understand the role of each word and its relationship to other words in the text.	Eppe et al. (2016) Pandy et al. (2023) Villa-Pérez et al. (2023)
Handling Contractions	Expanding contractions to their full form helps standardize text data and avoid ambiguity, especially in cases where the contraction may have a different meaning.	Chai et al. (2023) Mahimaidoss and Sathianesan (2023)
Named Entity Recognition (NER)	A process identifies and classifies named entities in text, such as names, places, dates, and numeric values, so the model can recognize specific entities and their context.	Ahmed and Wang (2023) Jang et al. (2022)
Punctuation removal	Removing punctuation from text data helps to simplify the text and ensures that the model focuses on the context of the text, improving the accuracy of the analysis and predictions.	Agarwal et al. (2023) Motyka et al. (2023) Ashfaque et al. (2023)

Table 6
The table analyzes the commonly used Deep Learning models in NLP-based Human–Agent Interaction..

Model	Description	Advantages	Limitations	Studies
Recurrent neural networks (RNN)	RNNs are neural networks with cyclic connections that enable them to process and remember sequences of data, making them suitable for tasks involving sequential or temporal context, such as NLP in HAI. (Rumelhart et al., 1986).	**Sequential Processing:** Handles sequential and time-related data processing. **Adaptable Lengths:** Flexible handling of input sequences of different lengths. **Memory Retention:** Remembers previous inputs and improves context-dependent calculation.	**Vanishing Gradient:** The gradient decreases in RNNs, making long-term learning more difficult. **Limited Long-Term Dependency:** Difficulties in capturing distant dependencies.	Hinkka et al. (2019), Sharfuddin et al. (2018), Pitsilis et al. (2018), Sari et al. (2020)
Long Short Term Memory (LSTM)	LSTMs, a type of RNN, enhance memory and context understanding in NLP, enabling more accurate and effective HAI. (Hochreiter and Schmidhuber, 1997).	**Extended Memory Span:** Excels in retaining information over long sequences. **Gradient Stability:** Ensures stable gradients through gating mechanisms in training. **Versatile Adaptability:** Adapts effectively to diverse tasks, processing complex patterns in data.	**Complexity:** More difficult to understand due to complexity. **Calculation Requirements:** Requires more resources for calculation. **Risk of Overfitting:** More prone to overfitting, especially with limited data.	Yao et al. (2014), Gandhi et al. (2021), Wen et al. (2015)
Gated Recurrent Unit (GRU)	GRU, another RNN, utilizes gating mechanisms for efficient information flow regulation, addressing challenges like vanishing gradients in sequential data processing (Gers et al., 2000).	**Efficient Training:** GRUs ensure faster convergence in recurrent neural networks. **Reduced Parameters:** GRUs have fewer parameters, increasing the calculations' efficiency. **Parallelization Advantage:** GRUs support faster training through improved parallelization.	**Short Memory Range:** GRUs may have difficulty capturing long-term patterns. **Hyperparameter Sensitivity:** GRU performance is sensitive to hyperparameter settings. **Limited Expressive Power:** GRUs may be less powerful at complex tasks than architectures such as LSTMs.	Zulqarnain et al. (2019), Santur (2019), Zulqarnain et al. (2020)
Bidirectional Encoder Representations from Transformers (BERT)	BERT is an NLP algorithm that employs bidirectional context comprehension through pre-training on extensive unlabeled text data, enabling nuanced contextual understanding and substantial performance improvements across diverse downstream NLP tasks (Devlin et al., 2018).	**Precision:** Excellent understanding of words in context. **Flexibility:** Adapts to tasks with minimal data through versatile pre-training. **Contextual Awareness:** Utilizes bidirectional attention for comprehensive understanding.	**High Compute Requirements:** Intensive calculations limit use in resource-constrained environments. **Sequential Limits:** Difficulties capturing sequential nuances affect task performance. **Memory Requirements:** Large memory requirements pose a challenge for use on devices with limited memory.	Moon et al. (2022), Devlin et al. (2018), Rahman et al. (2020), Kenton and Toutanova (2019)
GPT-4 (Generative Pre-trained Transformer 4)	GPT-4 is an advanced language model developed by OpenAI that excels in understanding and generating human-like text, and it enhances HAI by enabling more natural, coherent, and contextually aware communication between humans and agents. Achiam et al. (2023).	**Precision in Context:** Excellent understanding of words in sentence context. **Flexible Pre-training:** Adapts to tasks with minimal task-specific data through versatile pre-training. **Contextual Awareness:** Utilizes bidirectional attention to comprehensively understand surrounding words.	**Bias & Fairness Issues:** GPT-4 may still generate biased or inappropriate responses based on its training data, which may lead to unfair or offensive interactions. **Reliability:** In some cases, GPT-4 may generate incorrect, misleading or nonsensical responses, which may confuse users or affect the quality of the interaction. **Ethical issues:** The use of extensive training data raises concerns about privacy and the ethical handling of information, especially in sensitive applications such as healthcare.	Vrins et al. (2024)

(*continued on next page*)

productivity. As technology advances, these benefits are further enhanced by integrating Deep Learning-based Natural Language Processing, enabling more nuanced and adaptive interactions between humans and agents. Below are numerous examples of such applications in the industry that demonstrate how the synergy of Human–Agent Interaction and Deep Learning-based Natural Language Processing contributes to a more efficient, responsive and effective operating environment. Table 12 highlights some of the recent applications of HAI in industrial.

5.7. *Interdisciplinary*

Human–Agent Interaction is crucial in interdisciplinary research as it combines insights from robotics, psychology, computer science,

Table 6 (*continued*).

Model	Description	Advantages	Limitations	Studies
RoBERTa (A Robustly Optimized BERT Pretraining Approach)	RoBERTa is an optimized version of BERT designed to improve natural language understanding performance through robust pretraining techniques. Liu et al. (2019).	**Contextual Precision:** Excellent understanding of words in sentence context. **Flexible pre-training:** Adapts to tasks with minimal task-specific data through versatile pre-training. **Contextual awareness:** Uses bidirectional attention to fully understand surrounding words.	**Computational requirements:** RoBERTa's advanced architecture demands significant computational resources, limiting its use in resource-constrained robotic systems. **Potential Biases:** Despite efforts, RoBERTa may exhibit biases, raising ethical concerns in HAI. **Language Dependency:** RoBERTa's effectiveness in HAI heavily depends on accurate language comprehension; errors could undermine user trust.	Bird et al. (2023)
XLNet	XLNet is a state-of-the-art language model that achieves outstanding performance by using a permutation-based training approach to effectively capture bidirectional context. Yang et al. (2019).	**Enhanced Understanding:** XLNet's advanced architecture improves language comprehension, enhancing communication in HAI. **Contextual Relevance:** XLNet captures bidirectional context for more relevant robot responses in HAI. **Versatility:** XLNet's robustness makes it suitable for diverse HAI tasks like personal assistance and customer service.	**Computational Complexity:** The advanced architecture of the model demands significant computational resources, potentially limiting its application in resource-constrained robotic systems or real-time HAI. **Training Data Bias:** Despite mitigation efforts, the model may still be influenced by bias in training data, raising fairness concerns in HAI. **Integration Complexity:** Integrating the model into HAI systems can be challenging due to its complex architecture and training procedures, requiring expertise from developers and researchers.	Chen et al. (2024), Li and Yang (2023), Trueman et al. (2021)
ERNIE (Enhanced Representation through Knowledge Integration)	ERNIE is a language representation model developed by Baidu that aims to integrate structured knowledge into text representations to improve the understanding and generation of texts. Zhang et al. (2019).	**Knowledge Integration:** The ability to integrate structured knowledge into text representations improves the agent's understanding of complex concepts and contextually relevant information, enhancing its performance in HAI tasks. **Semantic Understanding:** Leveraging external knowledge sources, the model better captures the semantics of user commands and queries, leading to more accurate interpretation and generation of responses in HAI scenarios. **Adaptability:** The architecture enables fine-tuning to domain-specific data, allowing the robot to adapt and specialize to specific HAI applications, achieving better performance.	**Source Dependency:** The performance in HAI relies heavily on the quality and relevance of integrated external knowledge sources, potentially limiting effectiveness in domains with incomplete coverage. **Increased Complexity:** Integrating into HAI systems may introduce additional complexity in data processing, deployment, and computational requirements, posing challenges for system development and maintenance. **Ethical Concerns:** Integrating external knowledge raises ethical concerns regarding accuracy, bias, and privacy of sources, impacting fairness and trustworthiness in HAI.	Sun et al. (2019)

(*continued on next page*)

and design to integrate robotic technologies seamlessly. This collaborative approach requires a sophisticated understanding of human behavior and preferences. The interdisciplinary nature of HAI is critical to developing sophisticated, socially aware robotic systems that can work seamlessly with humans. The following commentary examines specific contributions of relevant work that underscore the importance of HAI as a catalyst for interdisciplinary exploration and innovation. Table 13 highlights some of the recent applications of HAI in interdisciplinary.

5.8. *Smart cites*

Smart cities leverage advanced technologies to enhance urban efficiency and livability. Within this framework, HAI addresses diverse challenges from transportation to healthcare. Deep learning-based Natural Language Processing enables seamless communication between humans and agents, improving interaction efficiency and fostering trust in robotic technologies. This integration enhances smart city services, contributing to their success. Table 14 presents some of the recent applications of HAI in Smart cites.

5.9. *Others*

Below are several noteworthy studies that hold considerable value concerning their application in HAI. These papers contribute significantly to the understanding and advancement of HAI dynamics, and offering valuable insights into various aspects of the field. Table 15 highlights some recent applications of HAI in certain sectors.

Table 6 (*continued*).

Model	Description	Advantages	Limitations	Studies
ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately)	ELECTRA is a language model developed to improve efficiency and accuracy by training on token-level replacements and offers a promising approach for various natural language processing tasks. Clark et al. (2020).	**Efficiency:** Focus on token-level substitutions enables more efficient learning compared to traditional masked language models, suitable for real-time HAI applications with limited computational resources. **Accuracy:** Training at the token level achieves higher accuracy in speech understanding and generation tasks, leading to more reliable and contextually relevant interactions between humans and agents. **Versatility:** Efficient learning approach and improved performance make it applicable to various HAI scenarios, improving the robot's ability to understand and respond effectively to user queries.	**Dependence on training data:** The performance of ELECTRA in HAI can be highly dependent on the quality and variety of training data, potentially leading to biases or limitations in understanding and generating human-like responses. **Complexity of integration:** Integrating ELECTRA into HAI systems may require significant effort and expertise due to its unique training approach and architecture, which could be challenging for developers and researchers in the field. **Ethical considerations:** As with any AI model, there are ethical considerations when using ELECTRA in HAI, including concerns about privacy, fairness, and transparency, which need to be carefully considered to ensure responsible adoption and use.	Baek et al. (2023)
DeBERTa (Decoding-enhanced BERT with Disentangled Attention)	DeBERTa is an advanced language model that extends the BERT architecture by incorporating decoding mechanisms and disentangled attention to improve its performance on various natural language understanding tasks. He et al. (2020).	**Improved User Intent Understanding:** Decoding-enhanced architecture and decoupled attention mechanisms enable better understanding of user requests and commands, leading to more accurate interpretation of user intent in HAI scenarios. **Enhanced Contextual Responses:** Incorporating decoding mechanisms allows for more contextualized responses, improving interaction quality in various HAI tasks such as conversation and information retrieval. **Reduced Computational Complexity:** Designed to be computationally efficient compared to traditional transformer models, resulting in faster inference and reduced resource requirements, beneficial for real-time HAI applications.	**Integration Complexity:** Integrating into HAI systems may require significant effort and expertise due to its advanced architecture and mechanisms, posing challenges for developers and researchers. **Risk of Overfitting:** The decoding-enhanced architecture could be prone to overfitting, especially with limited or biased datasets, potentially reducing generalization performance in real-world HAI. **Ethical Concerns:** Ethical considerations include privacy, fairness, and transparency, which must be carefully addressed for responsible deployment and use.	Li et al. (2022a)
BART (Bidirectional and Auto-Regressive Transformers)	BART is a versatile speech generation model that combines bidirectional pre-training with autoregressive decoding and can perform various natural language processing tasks such as text summarization, translation and question answering. Chipman et al. (2010).	**Summary Generation:** Creating concise summaries for HAI scenarios. **Language Translation:** Facilitating multilingual communication in HAI. **Question Answering:** Providing contextual responses in HAI tasks.	**Computational Complexity:** Decoding processes may be resource-intensive, limiting application in resource-constrained robotic systems. Fine-tuning Challenges: Adapting for specific tasks requires expertise, posing challenges for developers. **Ethical Considerations:** Raises privacy, fairness, and transparency concerns, especially in sensitive information scenarios.	Hanoch et al. (2021), Ren et al. (2023), Hromei et al. (2023), Shrestha et al. (2024)

(*continued on next page*)

6. Results analysis

The results and analysis derived from the aggregated table of papers in this survey, focusing on using DL-based Natural Language Processing in HAI, provide insightful perspectives on the current research landscape. Examining these works helps identify trends, challenges, and innovative areas within the field and provides researchers with a nuanced understanding of advances and existing gaps in the literature. Detailed findings on the prevalence and performance of various models are outlined in Table 16, contributing to a comprehensive understanding of the most advanced applications in the field.

The overview of deep learning-based NLP in HAI highlights a multifaceted landscape characterized by various models and applications in interdisciplinary, industrial, medical and entertainment domains. Outstanding models such as BERT, Bi-LSTM, Meshed Memory Transformers

Table 6 (*continued*).

Model	Description	Advantages	Limitations	Studies
PEGASUS (Pre-training with Extracted Gap-sentences for Abstractive Summarization Sequence-to-sequence models)	PEGASUS is a state-of-the-art abstract summarization model that uses gap-sentences extracted during pre-training to create concise and informative summaries of text documents. Zhang et al. (2020).	**Text Summarization:** PEGASUS creates concise summaries, aiding user comprehension and efficiency in HAI. **Translation:** Facilitates multilingual communication, enhancing inclusivity and accessibility in HAI. **Question Answering:** Provides accurate responses, improving user satisfaction and engagement in HAI tasks.	**Computational Complexity:** Advanced techniques may require significant computational resources, limiting use in resource-constrained robotic systems or real-time HAI. **Fine-tuning Complexity:** Adapting for specific tasks involves complex optimization processes, hindering widespread adoption. **Ethical Considerations:** Raises concerns about privacy, fairness, and transparency, especially in sensitive information scenarios.	Dinerstein et al. (2008)
GPT-Neo	GPT-Neo is a comprehensive language model developed by EleutherAI that is more accessible and less expensive than previous versions, but still delivers top performance on a wide range of natural language processing tasks. Black et al. (2021).	**Language Understanding:** GPT-Neo's extensive language modeling capabilities enhance natural conversations in HAI. **Task Customization:** Its open-source nature allows customization for various HAI tasks, improving effectiveness. **Accessibility:** Open-source release and low-cost training make it accessible for broader experimentation and innovation in HAI.	**Misinformation Risk:** Like other large-scale language models, GPT-Neo may unintentionally perpetuate biases or generate misinformation, requiring bias mitigation and content validation in HAI. **Contextual Understanding Limitations:** Despite impressive speech generation, GPT-Neo may struggle with nuanced contextual understanding, impacting interaction quality. **Ethical Concerns:** Raises privacy, security, and transparency issues, requiring adherence to ethical guidelines for responsible use.	Li et al. (2022b)

Table 7
Summary of applications of HAI in Autonomous Car..

Citation	Contribution
Lei et al. (2023)	Focused on improving HAI in CAVs by utilizing ChatGPT to improve voice assistants, environmental perception, and driving judgments. It also addressed integration difficulties such as data privacy and GPT-4 deployment on cloud servers, which affect the timeliness of V2X communication.
Cui et al. (2024)	Explored the challenges and opportunities of Multimodal Large Language Models (MLLMs) in autonomous driving, highlighting their potential in perception, motion planning, and industry applications.
Huang et al. (2023)	Investigated strategies for autonomous driving that used foundation models and large language models (LLMs), including simulation, world modeling, data annotation, and planning. It proposed using LLMs to remodel autonomous driving systems by combining human knowledge, common sense, and reasoning, and using rapid engineering to change pre-trained LMs for task-specific predictions without extra training.
Kim (2024)	Described a method for recognizing forward-driving automobiles using a YOLO deep learning network based on a pre-trained ResNet-50 and DashCam pictures. It obtained significant accuracies in vehicle recognition, type, and color categorization, indicating potential uses in improving autonomous vehicle safety.

and GPTs highlight the adaptability and efficiency of deep learning methods in addressing complex challenges in HAI scenarios. The choice of evaluation metrics, which include the F1-score, accuracy, BLEU score and domain-specific metrics, depends on the nuanced requirements of each task. Although interdisciplinary applications predominate, indicating the need for versatile models in practice, challenges remain regarding task specificity, bias in the training data and model interpretability. This review shows a recognizable trend towards using pre-trained models for language representations, which represents significant progress in improving the capabilities of Human–Agent Interactions. Sustained efforts are essential to improve the existing challenges and strengthen the robustness of deep learning-based NLP models in the dynamic environment of HAI.

Clear differences can be observed when examining specific model performances in the studied works. Variants of BERT, such as Mask-RCNN and CNN–LSTM networks, exhibit remarkable F1 results in interdisciplinary tasks, whereas industrial applications, especially the BERT-based model by Li et al. (2023a) achieved remarkable accuracy. The Generic User Simulator Model (GUS Model) shines in emotion and sentiment analysis, and the Bi-LSTM-CRF and GRU by Su et al. (2019) showed exemplary accuracy in Chinese KBQA. Nonetheless, challenges are also evident, ranging from scalability issues with Dong et al. (2023) model of BERT to limitations in the multimodal capabilities of the model of Peng et al. (2023) of CNN and Bi-LSTM. Overall, the observed successes and the identified areas for improvement emphasize the differentiated use of DL-based NLP models in the heterogeneous spectrum of HAI scenarios.

Recent endeavours listed in 5 are from diverse fields such as emotion and sentiment analysis, entertainment, healthcare, industrial processes, and interdisciplinary studies have achieved significant performance gains through rigorous methodological refinements and technological innovations. These advances have been driven in particular

Table 8
Summary of applications of HAI in Emotion and Sentiment Analysis..

Citation	Contribution
Giocondo et al. (2022)	Demonstrated how emotions influence motor responses to objects. This can be integrated into HAI by using NLP based on deep learning to recognize and respond to human emotions. This improves the contextual understanding and adaptive behavior of robots and makes interactions more intuitive and emotion-aware.
Tohma et al. (2023)	Proposed a hybrid sentiment analysis approach optimized for Turkish question-and-answer systems, which combines preprocessed text and polarity vectors to improve emotional understanding. It achieved 91.05% accuracy using machine learning techniques such as Linear SVM, Logistic Regression, Decision Trees, and Random Forest, paving the way for more authentic and sympathetic HAI in Turkish contexts.
Machová et al. (2020)	Suggested a lexicon-based sentiment analysis method augmented by the Particle Swarm Optimisation (PSO) algorithm to solve subjective human labeling difficulties. They used machine learning to identify messages without using vocabulary terms, obtaining over 99% classification accuracy and demonstrating the potential for emotion analysis in Human–Agent Interactions. Furthermore, in 'Context 2', Naive Bayes was utilized as a baseline for text classification and then applied to word labeling in a vocabulary.
Atzeni and Reforgiato Recupero (2018a)	Presented ongoing research integrating semantic technologies, deep learning, and NLP in Human–Agent Interaction, including a sentiment analysis system that uses RNNs with LSTM and GloVe word embeddings. The system, written in Java and using Stanford CoreNLP, was integrated into Zora, a humanoid robot, demonstrating the effectiveness of sentiment analysis and deep learning in Human–Agent Interaction.
Lakomkin et al. (2018)	Evaluated neural acoustic emotion detection models' durability in Human–Agent Interaction scenarios and proposed ways to bridge performance gaps between training and real-world testing. They dramatically improved model performance on the IEMOCAP dataset for both categorical and dimensional labels by conducting tests on the iCub platform and using data augmentation approaches such as noise overlaying.
Eom et al. (2022)	Investigated how Deep Learning NLP could be used to predict the vaccination sentiments of South Korean Twitter users during the Omicron variant outbreak, hence improving awareness of public opinion and informing responsive HAI.
Tejaswini et al. (2022)	Developed a novel Deep Learning hybrid model, "Fasttext CNN with LSTM and Fully Connected Layer (FCL)," which achieves extremely accurate depression identification from social media texts, hence helping to more effective and sensitive HAI in mental health apps.
Chakraborty et al. (2021)	Proposed deep learning NLP models (ANN and RNN-LSTM) to analyze visual focus and attention levels, enabling the robotic system to interact verbally with the user based on the detected attention levels.
Arumugam et al. (2017)	Proposed a model for interpreting language at different specificity levels in HAI, which improves accuracy by rooting commands in a hierarchical planning framework. Leveraging hierarchy increased efficiency and enabled rapid task response, as demonstrated on a physical robot. They used AMDPs for hierarchical planning and approached task grounding as a machine translation problem, using the IBM Model 2 to score reward functions.

Table 9
Summary of applications of HAI in Generative AI.

Citation	Contribution
Gamieldien (2023)	Utilized NLP and LLMs to improve the qualitative analysis, revealing self-regulated learning (SRL) mechanisms using exam wrapper responses. It investigated SRL disparities between students in an engineering physics course based on exam performance. The study offered zero-shot learning (ZSL) for building and measuring codebook accuracy, as well as potential solutions to biases and resource limits in AI-assisted research, with the goal of increasing the adaptability and fairness of Human–Agent Interactions in educational contexts.
Fezari et al. (2023)	Explored the possibilities of Generative AI in several industries, focusing on models such as autoencoders, VAEs, GANs, and transformers. It covered applications in advertising, gaming, healthcare, and HAI, with a focus on improving efficiency and personalized experiences, hence enhancing HAI's capabilities and efficacy across different domains.
Brynjolfsson et al. (2023)	Investigated the impact of generative AI-based conversational assistants on customer support agents, demonstrating productivity improvement, especially for novice workers. Discussed implications for worker experience and compensation for AI system training data.
Bautista et al. (2023)	Explored the application of LLMs in healthcare to address health disparities and enhance communication in human-doctor interactions. Compared domain-specific LLMs like SciBERT with multi-purpose LLMs like BERT. Discussed challenges and potential solutions for equitable model development in healthcare.
Baidoo-Anu and Ansah (2023)	Focused on incorporating ChatGPT into educational teaching and learning processes, with a focus on personalized learning experiences, formative assessment prompts, and continuing feedback. It emphasized collaborative efforts to ensure the safe and constructive use of generative AI technologies in education, resulting in improved learning outcomes and positive Human–Agent Interactions in educational contexts.

Table 10
Summary of applications of HAI in Healthcare.

Citation	Contribution
Ilyas et al. (2021)	Described a deep transfer learning technique for emotion detection in traumatic brain injury (TBI) patients, which used CNN and CNN–LSTM to improve classification performance on difficult datasets by using temporal data and transfer learning. Furthermore, it contributes to a robotic rehabilitation framework by demonstrating the efficacy of emotion monitoring using the SoftBank Pepper robot, advancing the use of robotics in healthcare, and encouraging more empathetic Human–Agent Interactions in rehabilitation settings.
Kim et al. (2021)	Described a new artificial diagnostic technique that combines the Symptom2Vec and AMoRSD models to improve real-time symptom-based diagnosis accuracy by analyzing patient responses and emotional expressions. This novel method enables rapid user symptom collecting and diagnosis creation, providing a solid foundation for clinical settings and encouraging more effective Human–Agent Interactions in healthcare scenarios.
Tejaswini et al. (2022)	Contributed in depression detection using FCL, a hybrid deep learning model that combines Fasttext Embedding, CNN, and LSTM. This method improves the early detection of depression in social media texts, offering useful information for preemptive intervention and encouraging more supportive Human–Agent Interactions in mental health contexts.
Mariani et al. (2022)	Explored five-year trends in language and NLP using the NLP4NLP+5 corpus, with a particular emphasis on AI, neural networks, reinforcement learning, and word embedding. This investigation gives essential scientific insights into the changing landscape of language and NLP technologies, motivating future research and supporting more sophisticated Human–Agent Interactions in the field of artificial intelligence.
Hollenstein et al. (2021)	Investigated EEG signal integration in NLP models and proposed a multi-modal architecture for sentiment analysis and relationship detection tasks. It proved the efficacy of filtering EEG signals into frequency bands, boosting sentiment classification, and relation recognition performance using contextualized BERT embeddings, demonstrating the potential for augmenting text input with EEG data.
Orsag et al. (2023)	Used deep learning, specifically LSTM networks, to recognize the spatiotemporal activity of human workers, achieving a training accuracy of 91.365% based on the InHARD dataset, improving human–robot collaboration through contextual knowledge
Kim et al. (2020)	Developed a BERT-based deep learning model for extracting keywords from pathology reports and demonstrated its performance by comparing it to existing methods on labeled reports as well as analyzing recovered keywords from unlabeled ones. This supervised technique, fine-tuned specifically for pathologic keywords, was useful in obtaining critical data. It used a classification layer for token classification and trained with cross-entropy loss, which helped to automate data extraction and improve Human–Agent Interactions in medical situations.
Sarraju et al. (2022)	Developed and evaluated clinical BERT-based NLP models that achieved high accuracy in classifying statin non-use and causes for non-use in patients with atherosclerotic cardiovascular disease using electronic health information. This result provides vital insights for targeted interventions, which will improve personalized healthcare strategies and create more successful Human–Agent Interactions in clinical settings.

Table 11
Summary of applications of HAI in Entertainment.

Citation	Contribution
Atzeni and Reforgiato Recupero (2018b)	Demonstrated the use of semantic technologies, deep learning, and NLP in Human–Agent Interaction, with a focus on allowing the Zora humanoid robot to interact with humans using natural language. Using external services such as Sentiment Analysis and a Generative Conversational Agent, the project used deep learning techniques such as LSTM-based sentiment analysis to demonstrate their effectiveness in increasing interaction with the robot.
Russo et al. (2022)	Utilized GPT-2 in deep learning-based NLP to extract significant neural coding from functional MRI data collected during narrative listening. GPT-2 successfully revealed unexpected and salient neural responses in language-related brain regions, demonstrating the power of deep learning models in unraveling complicated neural mechanisms involved in human language comprehension. This study advances our understanding of how the brain interprets language and lays the road for more advanced Human–Agent Interactions guided by neuroscience principles.

by the introduction of sophisticated pre-processing techniques such as tokenization, text cleansing, data augmentation and normalization. For example, the application of these techniques has led to measurable improvements in performance metrics, with DA accuracy reaching 79.3% and action accuracy reaching 80.13% in sentiment analysis (Shervedani et al., 2023). In addition, integrating state-of-the-art model architectures, such as bi-directional long term memory networks (Bi-LSTMs) and transformer-based models such as BERT, has played a critical role in improving performance. These architectural advances have led to exceptional levels of accuracy, for example in answering healthcare questions, reaching up to 99.27%. In addition, the strategic integration of multimodal inputs, comprising textual, auditory and visual data, in conjunction with diverse datasets has facilitated the extraction of richer semantic representations. This integration was underpinned by quantitative metrics such as Symptom2Vec similarity of 0.983 and an AMoRSD area under the curve (AUC) of 0.99% (Kim and Joe,

Table 12
Summary of applications of HAI in Industrial..

Citation	Contribution
Liu et al. (2023)	Presented an NLP model trained on a leap motion dataset for hand motion recognition in human–robot collaboration. Three dense layers and dropout regularization achieved high training accuracy (0.98) and low loss (0.20). The tests confirmed its robustness with a high test accuracy of 0.96 and a low loss of 0.12, underlining the effectiveness of NLP in improving Human–Agent Interaction.
Ahn et al. (2018)	Presented the Interactive Text2Pickup (IT2P) network, which uses deep learning to interpret ambiguous commands in human–robot collaboration when picking objects. Experiments demonstrated its effectiveness in clarifying instructions and improving task accuracy, which is promising for improving Human–Agent Interaction in the real world.
Keshinro et al. (2022)	Proposed NLP techniques (ConvLSTM and Long-Term Recurrent Convolutional Network (LRCN)) provide a valuable opportunity to improve HAI by enabling more intuitive and natural communication between humans and agents, which can lead to improved collaboration and task performance.
Lu et al. (2022)	Demonstrated a novel lip-speech decoding system that uses low-cost triboelectric sensors and an improved recurrent neural network to achieve high accuracy. This groundbreaking method shows promise for people with vocal cord injuries, with potential uses in voice translation systems. This breakthrough broadens the spectrum of assistive technologies, opening up new opportunities for greater communication and interaction for people with speech difficulties.

Table 13
Summary of applications of HAI in Interdisciplinary..

Citation	Contribution
Wahab et al. (2021)	Presented 4mCNLP-Deep, a superior computational model for N4 methylcytosine site identification that utilizes deep learning and word embedding. The importance of DNA methylation was emphasized, and the effectiveness of deep learning in classifying genomic data was demonstrated which mitigates human labor for classification and provides easy-to-use interface for new achieved data.
Ruffolo et al. (2023)	Presented IgFold, a quick deep learning method for antibody structure prediction, not only provides profound insights into a large number of paired antibody sequences but also emphasizes the need of exact structure prediction in understanding adaptive immune responses. This development has the potential to improve HAI by allowing for more precise and targeted therapeutic interventions, resulting in better health outcomes and well-being in collaborative healthcare settings.
Liu et al. (2023)	Demonstrated OPED, a deep learning-based optimization algorithm for priming-editing guide RNA designs, has demonstrated greater accuracy, efficiency, and versatility in genome editing applications and is revolutionizing the sector. The introduction of the OPEDVar database expands access to optimized designs, providing robots with powerful tools for precision genome editing jobs and fostering significant collaboration across fields.
Sanchez-Fernandez et al. (2023)	Presented CLOOME a multimodal contrastive learning system that significantly improved the search for chemical structures in bioimaging databases, showed remarkable transferability to various drug discovery tasks, and outperformed existing methods in predicting the mechanism of action.
Zeng et al. (2022)	Introduced KV-PLM, a BERT-based machine reading system that combines molecular structure with biological literature to boost drug discovery and research support. With 12 stacked Transformer layers and 110M parameters, KV-PLM outperformed human professionals in understanding molecular characteristics, as proven by extensive trials against powerful baseline models such as RXNFP and BERTwo. This achievement has the potential to improve human–machine collaboration in drug development processes and stimulate creativity in biomedical research.
Mao et al. (2023)	Introduced IKGM, a unique deep-learning algorithm that employs attention mechanisms to discover critical genes in macroevolution. IKGM has been successfully used to diurnal butterflies and nocturnal moths, revealing insights into genomic-level macroevolutionary mechanisms, furthering evolutionary biology, and encouraging collaborative study between humans and machines.
Shanmugavadivel et al. (2022)	Introduced a method for sentiment analysis and offensive language detection on Tamil–English code-mixed data that uses transformer-based language models such as BERT, RoBERTa, and adapter-BERT. Using adapter-BERT, the system improved its accuracy and investigated various pre-trained models for sentiment analysis and objectionable language identification. These findings help to construct models for analyzing code-mixed data, which improves cross-cultural communication and enables more effective Human–Agent Interactions in multilingual situations.
Diviya and Karmel (2023)	Introduced a deep neural network that generates graphics from Tamil text using the BASEGAN and HSRGAN algorithms. The model used TBERTBASECASE for text embedding and evaluated its performance using measures such as F1 Score, FID, and IS. This suggestion emphasizes the importance of image synthesis for regional languages such as Tamil, which promotes cultural representation and advances communication in a variety of linguistic contexts.

Table 14
Summary of applications of HAI in Smart Cites.

Citation	Contribution
Dong and Liu (2023)	Utilized Latent Dirichlet Allocation (LDA), the study investigated smart cities and AI governance, revealing critical concerns such as "service transformation" and "privacy management." It emphasized the importance of human-centered, sustainable smart city governance that aligns with social ideals. This study helps to better understand the obstacles and prospects for promoting urban sustainability using smart technologies and AI, opening the path for more inclusive and ethical advances in urban development.
Ullah et al. (2024)	Enhanced ICT procedures with deep learning and huge language models promote innovation in smart city environments, hence increasing Human–Agent Interaction. Federated learning improves urban security by streamlining activities like facial recognition and surveillance, while blockchain and natural language processing provide smooth data-driven forecasts in smart city infrastructures. These developments improve urban security and operational efficiency, fostering innovative and secure HAI as smart cities evolve.
Ahmad and Khan (2023)	Developed bilingual information extraction and propagation recognition systems with machine translation and transformer-based models such as RoBERTa. The researchers improved word-level categorization models such as BiLSTM, CRF, and BiLSTM-CRF. They used Nvidia RTX 3090 Ti graphics cards for training and achieved faster training times with mixed-precision models. This development improves HAI by allowing for efficient and accurate bilingual communication and information processing.
Li et al. (2023b)	Created deep learning models such as ResNet and XLNet for data-centric content categorization in smart city residential services. Using word embedding, context categorization, and data preprocessing, the researchers improved learning performance. XLNet achieved higher accuracy and F1 performances, whereas ResNet improved. These developments improve HAI in smart city settings.

Table 15
Summary of applications of HAI in other various places..

Citation	Contribution
Kasmaiee et al. (2023)	Utilized both rule-based and deep learning techniques to create a Persian text spelling correction system. The deep learning method used a deep encoder–decoder network with LSTM, FastText word embeddings, convolutional, and capsule layers to achieve an accuracy of 87%. It used capsule networks to preserve hierarchical relationships. In contrast, the rule-based solution used 112 rules and a database of 700 misspellings. These methods help to improve Human–Agent Interaction (HAI) by improving natural language processing capabilities in Persian-language scenarios.
Slack et al. (2023)	Developed TalkToModel was an advanced conversational system that outperformed traditional point-and-click explanation systems, allowing users to interactively understand and explain machine learning models, especially in critical domains such as healthcare.
Nandini and Schmid (2023)	Implemented BERT for hate speech prediction and LIME for explanation generation, and proposed an approach for identifying hate speech on social media that is interpretable using feature vectors. Investigated LIME's model-agnostic approach and classified interpretability approaches, then performed error analysis to assess model performance. These techniques improve openness and accountability in automated content moderation, hence increasing Human–Agent Interaction on social media platforms.

2023). In addition, innovative feature selection approaches, embedding techniques and hybrid model designs have further improved performance results. In particular, these advances have led to remarkable F1 scores of 97.01% and 93.34% for medical text summarization (Wang et al., 2015). Overall, these tangible advances underscore the concerted efforts within the academic research community to push the boundaries of artificial intelligence applications and thereby achieve significant performance gains across a wide range of domains.

7. Challenges and future directions

This survey provides a broad examination of the challenges associated with the fusion of deep learning-based NLP in the field of HAI. Through a comprehensive review of the relevant literature, this section addresses the recurring challenges identified in various studies and provides nuanced insights into the prevailing research landscape in this interdisciplinary field. Furthermore, potential avenues for future research efforts are outlined to advance the development of deep learning-based NLP applications in the dynamic context of Human–Agent Interaction.

7.1. Data collection

Collecting data for deep learning models in HAI is challenging due to the different requirements. Relying on freely available online resources, particularly in healthcare, can be insufficiently specific. The nuances of patient communication require different data sets to avoid misinterpretations. For example, different expressions of symptoms, such as chest pain, require customized data. A chatbot unfamiliar with these variations may be unable to adequately answer the nuanced health questions of older users. Targeted data collection from conversations with caregivers, elderly support forums and dialogues between patients and physicians is critical. This will ensure that the model is trained on a diverse and relevant data set, promoting effective and empathetic Human–Agent Interaction in healthcare for older people.

Researchers can use strategic solutions to overcome the complex challenges of data collection in NLP. Pre-trained models are central to leveraging existing knowledge to reduce reliance on extensive labeled data and enable more efficient training on task-specific datasets. Qiu et al. (2020) conducted a comprehensive investigation focusing on language representation learning and downstream task adaptation, thereby providing valuable insights into future research directions.

Table 16
The table dissects the experimental outcomes of research articles on HAI published between 2022 and 2023.

Ref.	Application	Dataset	Pre-Processing	Model	Results	Limitations
Shervedani et al. (2023)	Emotion and Sentiment Analysis	ELDERLY-AT-HOME corpus	Tokenization Text Cleaning Vectorization Normalization Feature Selection	Generic User Simulator Model (GUS Model)	**DA-Accuracy**:79.3% **Action-Accuracy**:80.13%	**Task specificity**: As the simulator is tailored to the "find" task, the applicability of the simulator beyond similar scenarios needs to be verified. **Bias of the corpus**: Using the ELDERLY-AT-HOME corpus can lead to bias and limit the adaptability to scenarios not covered. **Realism of the demonstrations**: The effectiveness of the simulator depends on realistic demonstrations, potentially affecting its adaptability.
Peng et al. (2023)	Emotion and Sentiment Analysis	MELD dataset EmoryNLP IEMOCAP	Tokenization Data Augmentation Lemmatization and Stemming Normalization	Bi-LSTM	**Accuracy:** 64.03%	**Limited multimodal capability**: Focuses on the text and ignores the potential benefit of audio or video input. **Dataset-specific assessment**: Performance is assessed against specific datasets, which questions adaptability. **Challenges distinguishing similar emotions**: Problems with closely related emotions, as seen in error analysis.
Prottasha et al. (2022)	Emotion and Sentiment Analysis	Self-collected from social media	Feature Extraction Data Augmentation	Bangla-BERT Hybrid of BERT and CNN–LSTM	**Accuracy:** 94.15%	**Scarcity of labeled data**: Limited Bangla data restricts model training and generalization. **Binary focus**: precludes multi-class or aspect-based sentiment analysis and limits the scope. **Embedding analysis**: comparative word embedding lacks a nuanced examination of strengths and weaknesses.
Martins et al. (2018)	Entertainment	RoboCup's generator GPSR FBM3	Tokenization Lowercasing Data Augmentation	RNNs LSTM	GPSR-dataset: Action-detection Acc: 0.895 Slot-filling Acc: 0.873 FBM3-dataset: Action-detection Acc: 0.687 Slot-filling Acc: 0.637	**Outdated ERL 2017 Model**: Relies on an outdated model in the ERL 2017 competition, impacting competitiveness and outcomes. **Limited Metric Exploration**: Relies solely on accuracy, neglecting other relevant metrics for comprehensive model evaluation. **Small Dataset Size**: Utilizes relatively small datasets (100 instructions for GPSR, 180 for FBM3), potentially limiting the model's generalization ability.
Dharaniya et al. (2023)	Entertainment	Internet Movie Script Database (IMSDb) movie dataset	Data Augmentation Feature selection HTML tags	DBN Bi-LSTM GPT3 GPT Neo X models	**BLUE:** 73.77% **CHRF:** 51.23% **GLEU:** 48.23% **METEOR:** 51.18% **NIST:** 52.81% **ROGUE:** 64.60%	**SHRDLU System Challenges**: Implementing SHRDLU for English interaction and object creation in scenes poses difficulties. **Animation Complexity**: Translating text into animations is challenging, highlighting complexities in the generation process. **NLP Nuance Difficulty**: Recognizes challenges in understanding intricate language nuances within natural language processing.
Kumar et al. (2020)	Healthcare	Harvard Medical School's N2C2 NLP dataset.	Feature Selection Vectorization Lemmatization Data Augmentation Tokenization	BiLSTM Word2Vec	**F1** : 99.27 **Reduced deviation:** 2.35 to 0.27	**KB limitation**: Chinese KB evaluation may limit the adaptability to different questions. **Entity ambiguity**: Common entity names in the KB lead to question ambiguity and make it difficult to find accurate answers. **Coverage limitation**: A limited KB may hinder answering questions about unrepresented entities.
Kim and Joe (2023)	Healthcare	WebMD Dictionary NHS inform Snomed Ct Cleveland Clinic AMoRSD	Word embedding	BERT	**Symptom2Vec similarity**: 0.983 **AMoRSD AUC:** 0.99%	**Symptom bias**: Over-emphasis on symptoms and neglect of other important aspects of the medical history. **Limited data set**: relies on a small data set (2,000 symptoms, 526 diseases), resulting in a risk of outdated information and low completeness. **Validation for rare diseases**: The performance of the rare disease model has not been validated, leading to uncertainties.

(*continued on next page*)

Table 16 (*continued*).

Ref.	Application	Dataset	Pre-Processing	Model	Results	Limitations
Budiharto et al. (2020)	Industrial	Stanford Question Answering Dataset (SQuAD) 100 dimensions of Global Vectors (GloVe)	Stemming Tokenization Named Entities Disambiguation (NED)	RNN	**F1**: 82.43%	**Restricted Encoder Exploration**: Does not explore encoder options beyond RNN and CNN. **GPU and Training Duration Dependency**: Relies on specific GPU and has a lengthy training duration. **Metric Selection Restriction**:Focuses solely on EM and F1 scores, overlooking other metrics.
Khodadadi et al. (2022)	Industrial	10 years of customer service calls, 100K recorded calls.	Lemmatization Data Augmentation Feature Selection Word Embedding POS tagging Tokenization	Bi-LSTM CNN	**Accuracy**: 0.84 Precision: 0.92	**Limited applicability**: A company's data set of 100,000 service calls may limit the model's ability to generalize to different real-world scenarios. **Dependence on historical data**: The model's dependence on past failure patterns may limit its ability to adapt to evolving trends. **Sensitivity of hyperparameters**: Insufficiently tuned hyperparameters may affect the optimal performance of the model.
Zhang et al. (2021)	Industrial	CoNLL-2005 CoNLL-2012	Syntax Reduction Feature Selection Labeling Statistical Learning Action Sequence Transformation	Bi-LSTM	**Action-sequence correctness**: 0.865	**Limited Generalization**: LD3PA might struggle with diverse robot platforms, hindering seamless adaptation. **Manual FL Labeling Dependency**: Manual Functional Labels (FLs) introduce subjectivity, impacting performance. **Subjective Evaluation Metrics**: FL-based metrics depend on subjective assignment, leading to varied results.
Li et al. (2023a)	Industrial	Virtual-Assistant-Max	Tokenization Embeddings	BERT	**Intent accuracy**:0.977 F1-score:0.968	**Dealing with unexpected requests**: Lack of discussion about Max's ability to adapt to unexpected requests. **Impact of environmental noise**: High production noise affects the accuracy of the virtual assistant, increases the error rate and poses a challenge to communication. **Integration challenges**: Brief mention of MES/ERP integration challenges without addressing compatibility issues.
Su et al. (2019)	Industrial	NLPCC-ICCPOL 2016's KBQA eval task's Chinese KB	Word Embedding	Bi-LSTM-CRF GRU	**Accuracy**: 99.14%	**Algorithm limitation**: The dependency on Weka's CML limits the results to particular algorithms. **Healthcare focus**: The focus on comorbidity neglects important aspects of healthcare. **Dataset representatio**n: n2c2 clinical notes may not fully capture real-world data diversity.
Larisch et al. (2023)	Industrial	Blue Gene/L (BGL) and Spirit	Tokenization Padding Windowing Data Augmentation	Compact Convolutional Transformer	**Precision:** 94.52 **Recall:** 57.87 **F1-Score:** 74.27	**Hyperparameter exploration gap**: Insufficient exploration of CCT hyperparameters affects potential performance. **Lack of analysis of computational efficiency**: the lack of specific metrics makes it difficult to understand the efficiency of CCT for practical comparisons. **Generalizability insufficient**: Limited discussion of CCT's adaptability to different protocol data limits insights into its versatility.
Tan et al. (2020)	Industrial	TC-QA dataset	Tokenization Part-of-speech tagging Named entity recognition Data Augmentation	Hybrid DL methods and Symbolic reasoning	**Text-score**:0.6193 **Image-score**:0.5719 **Video-retrieval-score**:0.5875 **Video-mIoU-score**:0.4557	**Limited task scope**: The focus on gearbox assembly limits generalizability. **Small data set**: The small size of the data set (991 QA pairs) can hinder effective model training. **Subjectivity of question types**: Subjective categorization can overlook variations of real-world scenarios.

(*continued on next page*)

Table 16 (*continued*).

Ref.	Application	Dataset	Pre-Processing	Model	Results	Limitations
Huang et al. (2022)	Industrial	Command Analysis(CA) dataset Perspective Disambiguation(PD) dataset	Word Embedding Tokenization Position embeddings Data Augmentation	LD3PA	**CA task:** 0.996 **PD task:** 0.995	**Incomplete Limitation Disclosure**: Lacks clear discussion of any specific drawbacks associated with the LD3PA. **Ambiguity in Object Applicability**: Whether the algorithm can handle a diverse range of objects is uncertain. **Limited Generalization**: The paper does not explore how well LD3PA can be adapted to other robots with different kinematics and capabilities.
Wang et al. (2015)	Interdisciplinary	COCO	Tokenization POS Lemmatization Stemming	Mask-RCNN, CNN–LSTM network	**F1-score(NIH):** 97.01% ,93.34% **F1-score(JRST):** 97.50%, 95.78%	**Limited vocabulary**: Dependence on a fixed dictionary can hinder understanding of commands outside its scope. **Parser limitations**: Reliance on parsers can lead to misinterpretation as they cannot capture complex semantic relationships. **Task complexity**: Limited exploration of the framework's ability to adapt to complex tasks.
Dong et al. (2023)	Interdisciplinary	Talk2Car	Tokenization Data Augmentation Normalization Vectorization	BERT	**AP50:**76.74	**Scalability**: The paper does not discuss the scalability of HuBo-VLM to larger datasets or complex tasks, which affects its practical utility. **Real-world deployment**: the challenges of deploying HuBo-VLM in real-world scenarios, including hardware limitations, are not addressed. **Incomplete related work**: The introduction lacks details on the differences between HuBo-VLM and existing models.
Takano (2020)	Interdisciplinary	Not applicable	Feature Extraction Data Encoding Word Sequences	Probabilistic Graphical Model GANs RNNs	**Accuracy:**0.578	**Sensor dependency**: Limits adaptability due to dependency on specific IMU sensors, limiting compatibility with different sensor technologies. **Choice of hidden states**: Lack of justification for the fixed 4000 hidden states, no analysis of alternative values. **Scoring shortcomings**: May overlook crucial aspects of sentence quality as it relies on BLEU scores and subjective judgments.
Golech et al. (2022)	Interdisciplinary	MS COCO dataset	Tokenization Data Augmentation Handling Contractions	Meshed-Memory-Transformers	**Bleu-1**: 0.72.	**Generalizability limitation**: The Turkish focus limits the applicability; the adaptability to different datasets is unclear. **Dataset-specific evaluation**: The exclusive use of the Turkish MS COCO raises concerns about the versatility of other datasets. **Lack of comparative analysis**: the lack of model comparisons compromises the clarity of the performance evaluation.
Wang et al. (2015)	Interdisciplinary	CITIC Institute and JST parallel corpus	POS CHUNK Word Embedding	LSTM	**BLEU:** 39.52.	**Toolkit dependency**: Limits generalizability by relying on the specific CURRENT toolkit. **Default setting**: Using default settings can affect performance and robustness. **Data Preprocessing**: Replacing digits with "#" lost valuable sequential information,both words "Tel192" and "Tel6" are converted into "Tel#".
Vashistha et al. (2022)	Interdisciplinary	IIT Bombay Eng-Hin Parallel Corpus	Lowercasing Data Augmentation Tokenization Normalization	Neural Machine Translation RNN	**BLEU:** 24.54.	**Limited Metric Exploration**: Focused on BLEU and perplexity, missing a broader metric assessment. **Interpretability Gap**: Fails to discuss interpretability, crucial for understanding model decisions. **Shallow Training Loss Analysis**: Incomplete exploration of training loss patterns, hindering holistic understanding.

Data augmentation techniques such as text paraphrasing or image manipulation provide a practical means of overcoming data scarcity by artificially augmenting the training dataset. Wei and Zou (2019) and Shorten and Khoshgoftaar (2019) provided insightful overviews of various data augmentation strategies. Synthetic data generation is an innovative solution where artificial instances are created to augment or replace real-world datasets to improve model generalization and mitigate biases in NLP tasks. Overall, these techniques help reduce the challenges associated with data acquisition for deep learning-based NLP models.

In addition, Data democratization (Lefebvre et al., 2021) plays a central role in addressing these challenges by enabling broader access to different datasets. By facilitating access to data, data democratization promotes collaboration and increases the robustness of AI models.

7.2. Human–robot communication gap

The fundamental challenge in the field of HAI lies in the mismatch between human communication patterns and the ability of robots to understand implicit cues, non-verbal communication and contextual language. Humans inherently incorporated these nuanced elements into their communication, making it difficult for robots to decode and respond appropriately. A vivid scenario illustrating this challenge is a human instructing a robot to "grab the red object on the table". While humans can easily grasp the contextual and implicit cues contained in this instruction, a robot may have difficulty recognizing the subtleties of the instructions, which can lead to errors in execution. To make matters worse, contemporary language trends bring dynamism to human communication. The emergence of new idioms, linguistic conventions and patterns of expression over time presents additional obstacles for robots as they strive to understand and adapt to the evolving subtleties of language. Furthermore, variations in tone of voice present an additional layer of complexity that can significantly alter the semantic interpretation of a given sentence, making it even more difficult for robots to decode human communication accurately. Bridging this communication gap is essential to improve the efficiency and naturalness of Human–Agent Interaction.

The challenges posed by the gap between humans and agents can only be comprehensively addressed. Human-in-the-loop strategies that integrate human feedback throughout the training and deployment phases of NLP models facilitate continuous refinement based on real-world interactions. The study by Budd et al. (2021) demonstrated the importance of human-in-the-loop computing to ensure the safety and efficacy of deep learning applications for medical image analysis in clinical practice. Future researchers can explore Crowdsourcing, XLNet and ELECTRA to bridge the gap in human–robot communication and improve contextual understanding and responsiveness in various interaction scenarios.Vemprala et al. (2024) explored the potential of ChatGPT for robotics tasks, presented a pipeline for its application and introduced PromptCraft, an open-source platform for collaborative prompting strategies to bridge the communication gap between humans and agents. Mubin et al. (2014) investigated the influence of language, microphone type and robot head movement on speech recognition accuracy during robot interactions. No difference was found between ROILA and English, but the significant influence of the microphone type and the robot's head movement was highlighted. Krishna et al. (2022) presented a framework for socially situated artificial intelligence in which agents learn to ask informative questions through reinforcement learning while interacting with people in a photo-sharing social network, leading to improved visual intelligence.

In addition, initiatives such as the Humane AI Pin and Rabbit R1 are critical to the development and use of AI technologies in HAI. By emphasizing ethical principles, transparency and human-centered design, these initiatives ensure that AI systems prioritize human well-being, respect ethical norms and promote responsible use of artificial intelligence in Human–Agent Interaction scenarios. Adherence to the principles outlined in Humane AI Pin and Rabbit R1 can help address ethical concerns and promote trust between humans and agents in HAI contexts.

7.3. Cost and resource allocation

The challenge of NLP poses based on deep learning in the context of Human–Agent Interaction is the considerable computational effort associated with these models. The high computational requirements for training and subsequent integration into robotic systems significantly increase the costs. The choice of a particular deep learning-based NLP model exacerbates this challenge, as the varying computational costs depend on the complexity of the chosen model. This financial burden hinders the seamless integration of these models into robots and limits their widespread adoption. Integrating sophisticated NLP models into robotic systems requires additional components and software to enhance the robot's user interface. However, this enhancement incurs additional costs, further complicating the economic viability of integrating DL-based NLP models into HAI scenarios.

Effective cost management of NLP models based on deep learning requires strategic approaches such as active learning, where informative instances are selectively annotated to minimize the need for labeled data and reduce costs. Simulators provide virtual training environments that accelerate development and reduce the requirement for extensive physical hardware. The GPU integration demonstrated by Peng et al. (2020) also optimizes memory usage. The combination of active learning, simulators and GPU integration in the context of an NLP model for HAI ensures cost-efficient implementation by mitigating the computational challenges. The Transformers library developed by Wolf et al. (2020) increases efficiency through state-of-the-art transformer architectures and pre-trained models and provides accessibility, extensibility, and a unified API for optimized use in natural language processing. Shafahi et al. (2019) introduced an algorithm that minimizes the cost of generating adversarial examples by reusing gradient information during model parameter updates. These integrated solutions enable researchers to effectively manage cost variations by integrating NLP based on deep learning.

7.4. Handling failure

In HAI, predicting obstacle intentions using probabilistic occupancy models poses a challenge, as they can only consider unforeseen behavior caused by human free will to a limited extent. Machine predictions and statistical AI models based on event probabilities are unreliable for industrial applications. Simplified 3D models improve the prediction accuracy but increase the computing time. The challenge also extends to potential malfunctions owing to the loss of control or obstructed positions. To solve this problem, it is crucial to develop reliable algorithms with low computational effort and consider the natural movement tendencies of obstacles is crucial. Failure to accurately predict human intentions in industrial environments can jeopardize safety protocols and system operations, highlighting the need to overcome these challenges for the smooth operation of robotic systems in dynamic environments.

Overcoming challenges in dealing with errors requires strategic solutions. Reinforcement Learning from Human Feedback refines model predictions with human input, as shown by Bai et al. (2022), whereas Adversarial Training increases robustness against unexpected human actions. Input Sanitization ensures valid data processing and reduces the risk of erroneous prediction. Failover mechanisms to ensure system stability through redundancy and mitigate the impact of unforeseen human behavior. Feedback loops enable continuous learning and promote the development of resilient Human–Agent Interaction systems for reliability and safety in dynamic environments.

7.5. Data security

In the area of HAI, the key considerations of safety and trust coincide with the critical need to ensure data security. The prospect that a robot could be stolen is a tangible risk that raises concerns about protecting sensitive information in the robotic system. Furthermore, the vulnerability also extends to the server infrastructure, which could potentially be compromised, exposing confidential user data and jeopardizing the integrity of the HAI. This scenario poses a major challenge, especially in applications where personal or sensitive data is processed, as it raises fears of privacy breaches and unauthorized access to sensitive data (Pritee et al., 2024). In healthcare, for example, where a robot is used in patient care, the theft of the robot could compromise patient data, violating privacy regulations and undermining trust in the use of robotic systems for sensitive tasks.

Overcoming the challenges of safety, trust and data security in HAI requires comprehensive solutions. Simulators provide controlled test environments and minimize risks. Encryption ensures the security of data at rest and in transit. Anonymization (Majeed and Lee, 2020) and differential privacy strengthen privacy by obscuring individual data or adding noise. Homomorphic encryption enables secure computation without decryption, and tokenization reduces the risk of unauthorized access. SSL/TLS protocols secure data during transmission. Zero-knowledge proofs (Goldwasser et al., 2019) improve data protection, and secure multi-party computation enables collaboration without exposing raw data. Privacy-enhancing technologies and trusted execution environments protect user privacy and critical processes. Integrating these solutions creates a robust framework for using deep learning-based NLP models in sensitive applications, ensuring security, trust and data safety.

7.6. Execution time

In HAI, NLP models based on deep learning pose a major challenge in terms of execution time owing to their inherent complexity and extensive training on large data sets. This complexity, combined with the large amounts of data, leads to long execution times, which is a critical problem in time-critical scenarios such as emergencies. This challenge is particularly pronounced in critical areas such as industrial or medical applications, where even slight delays in processing natural language input can have serious consequences. An NLP model that analyzes diagnostic information in the medical field can struggle to provide timely insights in emergencies, thereby jeopardizing patient outcomes. These examples highlight the urgency of addressing the time constraints of running sophisticated NLP models in real-time scenarios.

Effectively addressing time constraints in executing deep learning-based NLP models involves optimizing algorithms and architectures for efficiency. Exploring lightweight architectures and integrating hardware accelerators such as GPUs or TPUs significantly reduces the execution time. These strategic solutions ensure systems' prompt and efficient operation, which is particularly crucial in time-sensitive scenarios, such as industrial robotics reliant on real-time NLP interpretation.

7.7. World model

A world model, as described by researchers from DeepMind, is "an internal simulation of the external environment that allows an agent to predict future states, plan actions, and learn from interactions without constant real-world feedback"(Ha and Schmidhuber, 2018). The integration of world models into deep learning-based NLP greatly enhances the possibilities of HAI by providing a comprehensive context for understanding and responding to human speech. World models enable robots to obtain an internal representation of their environment, including spatial arrangements, object positions and dynamic changes. This enhanced context allows robots to interpret verbal commands more accurately and perform tasks more efficiently. For example, when a user asks a robot to find a specific object in a cluttered room, the world model helps the robot understand the layout of the room, recognize objects and navigate effectively to find the object. In addition, world models facilitate better dialogue management by maintaining situational awareness and allowing robots to seamlessly switch between tasks and maintain context in conversations with multiple interlocutors.

Future research should focus on improving the integration of world models with deep learning-based NLP to improve Human–Agent Interaction further. One focus should be on developing sophisticated and scalable world models that can handle complex and dynamic environments in real time. Researchers should also explore ways to improve the personalization of world models so that robots can adapt to individual user preferences and previous interactions. In addition, addressing challenges related to privacy and security is critical, especially as robots are increasingly used in personal and sensitive environments such as homes and healthcare facilities. By taking these considerations into account, future advances in world models can lead to more intuitive, efficient and safer interactions between humans and agents, paving the way for wider acceptance and more practical applications of robotics in everyday life.

8. Conclusions

In summary, the maturation of Deep Learning methods with Natural Language Processing in the context of HAI represents a transformative paradigm in this rapidly evolving field. This study deftly navigates the complicated dynamics of HAI and emphasizes the central role that Deep Learning plays in shaping the communication between humans and agents. Departing from conventional sentiment analysis, our investigation encompasses a spectrum of HAI facets, including dialogue systems, language understanding, and contextual communication. The study systematically scrutinizes the applications, algorithms and models that describe the current landscape of Deep Learning-based NLP in HAI. It also provides valuable insights into common pre-processing techniques, datasets and specific evaluation metrics. By revealing the benefits and challenges that machine learning and Deep Learning algorithms bring to HAI and providing a comprehensive overview of current state-of-the-art experiments, this review serves as a navigation aid for the Field and a catalyst for future advances. The concluding discussion of specific challenges in the field of HAI sets the stage for future research and ensures a nuanced understanding of models, applications, challenges, and the trajectory of Deep Learning-based NLP research in the field of HAI.

CRediT authorship contribution statement

Nafiz Ahmed: Writing – original draft, Data curation, Conceptualization. **Anik Kumar Saha:** Writing – original draft, Methodology, Investigation, Formal analysis. **Md. Abdullah Al Noman:** Writing – review & editing, Validation, Resources, Methodology. **Jamin Rahman Jim:** Visualization, Validation, Methodology. **M.F. Mridha:** Writing – review & editing, Validation, Supervision. **Md Mohsin Kabir:** Writing – review & editing, Validation, Investigation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: M. F. Mridha reports was provided by American International University Bangladesh. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We gratefully acknowledge the Advanced Machine Intelligence Research Lab (AMIR Lab) for their invaluable support and guidance throughout this project.

References

Abdalla, M.H., Majidpour, J., Rasul, R.A., Alsewari, A.A., Rashid, T.A., Ahmed, A.M., Hassan, B.A., Tayfor, N.B., Qader, S.M., Salih, S.Q., 2023. Sentiment analysis based on hybrid neural network techniques using binary coordinate ascent algorithm. IEEE Access 11, 134087–134099.

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al., 2023. Gpt-4 technical report. arXiv preprint arXiv:2303.08774.

Agarwal, A., Nikitha, P., Ramkumar, S., Sinha, A., Maheshwari, P., Saini, A.S., 2023. DeepGram: Combining language transformer and N-gram based ML models for YouTube spam comment detection. J. Data Sci. Intell. Syst.

Ahmad, P.N., Khan, K., 2023. Propaganda detection and challenges managing smart cities information on social media. EAI Endorsed Trans. Smart Cities 7 (2), e2.

Ahmed, Z., Wang, J., 2023. A fine-grained deep learning model using embedded-CNN with BiLSTM for exploiting product sentiments. Alex. Eng. J. 65, 731–747.

Ahn, H., Choi, S., Kim, N., Cha, G., Oh, S., 2018. Interactive text2pickup networks for natural language-based human–robot collaboration. IEEE Robot. Autom. Lett. 3 (4), 3308–3315.

Aldunate, Á., Maldonado, S., Vairetti, C., Armelini, G., 2022. Understanding customer satisfaction via deep learning and natural language processing. Expert Syst. Appl. 209, 118309.

Alomari, M., Hogg, D.C., Cohn, A.G., 2017. Leeds Robotic Commands. University of Leeds.

Alomari, M., Li, F., Hogg, D.C., Cohn, A.G., 2022. Online perceptual learning and natural language acquisition for autonomous robots. Artificial Intelligence 303, 103637.

Alshahrani, H.J., Hassan, A.Q., Almalki, N.S., Alnfiai, M.M., Salama, A.S., Hamza, M.A., 2023. Applied linguistics with red-tailed hawk optimizer-based ensemble learning strategy in natural language processing. IEEE Access 11, 132448–132456.

Amaar, A., Aljedaani, W., Rustam, F., Ullah, S., Rupapara, V., Ludi, S., 2022. Detection of fake job postings by utilizing machine learning and natural language processing approaches. Neural Process. Lett. 1–29.

Amirian, S., Rasheed, K., Taha, T.R., Arabnia, H.R., 2021. Automatic generation of descriptive titles for video clips using deep learning. In: Advances in Artificial Intelligence and Applied Cognitive Computing: Proceedings from ICAI'20 and ACC'20. Springer, pp. 17–28.

Anon, 2023a. N2c2 NLP Research Data Sets — portal.dbmi.hms.harvard.edu. https://portal.dbmi.hms.harvard.edu/projects/n2c2-nlp/. (Accessed 23 November 2023).

Anon, 2023b. Papers with Code - MSR-VTT Dataset — paperswithcode.com. https://paperswithcode.com/dataset/msr-vtt. (Accessed 23 November 2023).

Anon, 2023c. Papers with Code - MSVD Dataset — paperswithcode.com. https://paperswithcode.com/dataset/msvd. (Accessed 23 November 2023).

Arikatla, G., Chinnapottu, B., 2021. Movie prediction based on movie scriptsusing natural language processing and machine learning algorithms.

Arumugam, D., Karamcheti, S., Gopalan, N., Wong, L.L., Tellex, S., 2017. Accurately and efficiently interpreting human-robot instructions of varying granularities. arXiv preprint arXiv:1704.06616.

Ashfaque, M.W., Malik, S.I., Kayte, C.N., Banu, S.S., Balobaid, A.S., Hannan, S.A., 2023. Design and implementation: Deep learning-based intelligent chatbot. In: 2023 3rd International Conference on Computing and Information Technology. ICCIT, IEEE, pp. 84–89.

Ashraf, M.R., Jana, Y., Umer, Q., Jaffar, M.A., Chung, S., Ramay, W.Y., 2023. BERT based sentiment analysis for low-resourced languages: A case study of Urdu language. IEEE Access.

Atzeni, M., Reforgiato Recupero, D., 2018a. Deep learning and sentiment analysis for human-robot interaction. In: The Semantic Web: ESWC 2018 Satellite Events: ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers 15. Springer, pp. 14–18.

Atzeni, M., Reforgiato Recupero, D., 2018b. Deep learning and sentiment analysis for human-robot interaction. In: The Semantic Web: ESWC 2018 Satellite Events: ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers 15. Springer, pp. 14–18.

Ayanouz, S., Abdelhakim, B.A., Benhmed, M., 2020. A smart chatbot architecture based NLP and machine learning for health care assistance. In: Proceedings of the 3rd International Conference on Networking, Information Systems & Security. pp. 1–6.

Baek, S., Kim, J., Lee, J., Lee, M., 2023. Implementation of a virtual assistant system based on deep multi-modal data integration. J. Signal Process. Syst. 1–11.

Baghaei, K.T., Payandeh, A., Fayyazsanavi, P., Chen, Z., Ramezani, S.B., Rahimi, S., 2023. Deep representation learning: Fundamentals, technologies, applications, and open challenges. IEEE Access.

Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., Drain, D., Fort, S., Ganguli, D., Henighan, T., et al., 2022. Training a helpful and harmless assistant with reinforcement learning from human feedback. arXiv preprint arXiv:2204.05862.

Baidoo-Anu, D., Ansah, L.O., 2023. Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. J. AI 7 (1), 52–62.

Balouch, B.A.K., Hussain, F., 2023. A transformer based approach for abstractive text summarization of radiology reports. In: International Conference on Applied Engineering and Natural Sciences. Konya, Turkey.

Bautista, Y.J.P., Theran, C., Aló, R., Lima, V., 2023. Health disparities through generative AI models: A comparison study using a domain specific large language model. In: Proceedings of the Future Technologies Conference. Springer, pp. 220–232.

Bird, J.J., Ekárt, A., Faria, D.R., 2023. Chatbot interaction with artificial intelligence: human data augmentation with T5 and language transformer ensemble for text classification. J. Ambient Intell. Humaniz. Comput. 14 (4), 3129–3144.

Black, S., Gao, L., Wang, P., Leahy, C., Biderman, S., 2021. Gpt-neo: Large scale autoregressive language modeling with mesh-tensorflow. If you use this software, please cite it using these metadata 58, 2.

Brown, S., Lee, H., 2022. Interactive learning experiences: Integrating natural language processing in educational robots. J. Educat. Technol. 45 (3), 189–204.

Brynjolfsson, E., Li, D., Raymond, L.R., 2023. Generative AI at Work. Technical Report, National Bureau of Economic Research.

Budd, S., Robinson, E.C., Kainz, B., 2021. A survey on active learning and human-in-the-loop deep learning for medical image analysis. Med. Image Anal. 71, 102062.

Budiharto, W., Andreas, V., Gunawan, A.A.S., 2020. Deep learning-based question answering system for intelligent humanoid robot. J. Big Data 7, 1–10.

Budiharto, W., Andreas, V., Gunawan, A.A.S., 2021. A novel model and implementation of humanoid robot with facial expression and natural language processing (NLP). ICIC Express Lett. B: Appl. 12 (3), 275–281.

Caldera, S., Rassau, A., Chai, D., 2018. Review of deep learning methods in robotic grasp detection. Multimodal Technol. Interact. 2 (3), 57.

Cascianelli, S., Costante, G., Ciarfuglia, T.A., Valigi, P., Fravolini, M.L., 2018. Full-GRU natural language video description for service robotics applications. IEEE Robotics Autom. Lett. 3 (2), 841–848.

Chai, Y., Kakkar, D., Palacios, J., Zheng, S., 2023. Twitter sentiment geographical index dataset. Sci. Data 10 (1), 684.

Chakraborty, P., Ahmed, S., Yousuf, M.A., Azad, A., Alyami, S.A., Moni, M.A., 2021. A human-robot interaction system calculating visual focus of human's attention level. IEEE Access 9, 93409–93421.

Chang, S., Ghamisi, P., 2023. Changes to captions: An attentive network for remote sensing change captioning. arXiv preprint arXiv:2304.01091.

Chen, L., Javaid, M., Di Eugenio, B., Žefran, M., 2015. The roles and recognition of haptic-ostensive actions in collaborative multimodal human–human dialogues. Comput. Speech Lang. 34 (1), 201–231.

Chen, X., Kang, J., Hu, C., 2024. Design of artificial intelligence companion chatbot. J. New Media 6.

Chipman, H.A., George, E.I., McCulloch, R.E., 2010. BART: Bayesian additive regression trees. Ann. Appl. Stat. 4 (1), 266–298.

Clark, K., Luong, M.T., Le, Q.V., Manning, C.D., 2020. Electra: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv:2003.10555.

Cui, C., Ma, Y., Cao, X., Ye, W., Zhou, Y., Liang, K., Chen, J., Lu, J., Yang, Z., Liao, K.-D., et al., 2024. A survey on multimodal large language models for autonomous driving. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 958–979.

Das, R.K., Islam, M., Hasan, M.M., Razia, S., Hassan, M., Khushbu, S.A., 2023. Sentiment analysis in multilingual context: Comparative analysis of machine learning and hybrid deep learning models. Heliyon 9 (9).

Das, A., Saha, D., 2022. Deep learning based Bengali question answering system using semantic textual similarity. Multimedia Tools Appl. 1–25.

Deruyttere, T., Vandenhende, S., Grujicic, D., Van Gool, L., Moens, M.F., 2019. Talk2Car: Taking control of your self-driving car. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. EMNLP-IJCNLP, pp. 2088–2098.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Dharaniya, R., Indumathi, J., Kaliraj, V., 2023. A design of movie script generation based on natural language processing by optimized ensemble deep learning with heuristic algorithm. Data Knowl. Eng. 146, 102150.

Dinerstein, J., Egbert, P.K., Ventura, D.A., 2008. Learning Policies for Embodied Virtual Agents Through Demonstration. IJCAI.

Diviya, M., Karmel, A., 2023. Deep neural architecture for natural language image synthesis for Tamil text using BASEGAN and hybrid super resolution GAN (HSRGAN). Sci. Rep. 13 (1), 14455.

Dong, L., Liu, Y., 2023. Frontiers of policy and governance research in a smart city and artificial intelligence: an advanced review based on natural language processing. Front. Sustain. Cities 5, 1199041.

Dong, Z., Zhang, W., Huang, X., Ji, H., Zhan, X., Chen, J., 2023. Hubo-VLM: Unified vision-language model designed for human robot interaction tasks. arXiv preprint arXiv:2308.12537.

Duong, M., Nguyen, L., Vuong, Y., Le, T., Nguyen, H.T., 2023. A deep learning-based system for automatic case summarization. arXiv preprint arXiv:2312.07824.

Eom, G., Yun, S., Byeon, H., 2022. Predicting the sentiment of South Korean Twitter users toward vaccination after the emergence of COVID-19 Omicron variant using deep learning-based natural language processing. Front. Med. 9, 948917.

Eppe, M., Trott, S., Feldman, J., 2016. Exploiting deep semantics and compositionality of natural language for human-robot-interaction. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 731–738.

Fadhil, A., et al., 2019. Catalia health's mabu: A novel AI-powered telemedicine chatbot for patient engagement and chronic disease management. J. Med. Internet Res. 21 (7), e13364.

Fanjie, K., Yaqi, L., Miaomiao, X., Silamu, W., Yanbing, L., 2023. SUST and RUST: Two datasets for uyghur scene text recognition. IEEE Access 11, 126209–126220.

Fezari, M., Al-Dahoud, A., Al-Dahoud, A., 2023. Augmanting reality: The power of generative AI. University Badji Mokhtar Annaba, Annaba, Algeria.

Gamieldien, Y., 2023. Innovating the Study of Self-Regulated Learning: An Exploration through NLP, Generative AI, and LLMs. Virginia Tech.

Gandhi, U.D., Malarvizhi Kumar, P., Chandra Babu, G., Karthick, G., 2021. Sentiment analysis on twitter data by using convolutional neural network (CNN) and long short term memory (LSTM). Wirel. Pers. Commun. 1–10.

Gers, F.A., Schmidhuber, J., Cummins, F., 2000. Learning to forget: Continual prediction with LSTM. Neural Comput. 12 (10), 2451–2471.

Giachos, I., Piromalis, D., Papoutsidakis, M., Kaminaris, S., Papakitsos, E.C., 2020. A contemporary survey on intelligent human-robot interfaces focused on natural language processing. Int. J. Res. Comput. Appl. Robotics 8 (7), 1–20.

Giocondo, F., Borghi, A.M., Baldassarre, G., Caligiore, D., 2022. Emotions modulate affordances-related motor responses: a priming experiment. Front. Psychol. 13, 701714.

Goldwasser, S., Micali, S., Rackoff, C., 2019. The knowledge complexity of interactive proof-systems. In: Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali. pp. 203–225.

Golech, S.B., Karacan, S.B., S""onmez, E.B., Ayral, H., 2022. A complete human verified turkish caption dataset for MS COCO and performance evaluation with well-known image caption models trained against it. In: 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering. ICECCME, IEEE, pp. 1–6.

Gross, J.A., Roberson, W.C., Foley-Cox, J.B., 2021. Cs 230: film success prediction using nlp techniques.

Guazzo, A., Longato, E., Fadini, G.P., Morieri, M.L., Sparacino, G., Di Camillo, B., 2023. Deep-learning-based natural-language-processing models to identify cardiovascular disease hospitalisations of patients with diabetes from routine visits' text. Sci. Rep. 13 (1), 19132.

Gupta, A., Carpenter, D., Min, W., Rowe, J., Azevedo, R., Lester, J., 2023. Detecting and mitigating encoded bias in deep learning-based stealth assessment models for reflection-enriched game-based learning environments. Int. J. Artif. Intell. Educat. 1–28.

Ha, D., Schmidhuber, J., 2018. World models. arXiv preprint arXiv:1803.10122.

Halawani, H.T., Mashraqi, A.M., Badr, S.K., Alkhalaf, S., 2023. Automated sentiment analysis in social media using harris hawks optimisation and deep learning techniques. Alex. Eng. J. 80, 433–443.

Hanoch, Y., Arvizzigno, F., Hernandez García, D., Denham, S., Belpaeme, T., Gummerum, M., 2021. The robot made me do it: Human–robot interaction and risk-taking behavior. Cyberpsychol. Behav. Soc. Netw. 24 (5), 337–342.

He, P., Liu, X., Gao, J., Chen, W., 2020. Deberta: Decoding-enhanced bert with disentangled attention. arXiv preprint arXiv:2006.03654.

Hinkka, M., Lehto, T., Heljanko, K., Jung, A., 2019. Classifying process instances using recurrent neural networks. In: Business Process Management Workshops: BPM 2018 International Workshops, Sydney, NSW, Australia, September 9-14, 2018, Revised Papers 16. Springer, pp. 313–324.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Hoffmann, L., et al., 2019. Pepper in reality: Selecting and applying a humanoid robot for social human-robot interaction in retail. In: Proceedings of the 2019 ACM/IEEE International Conference on Human-Robot Interaction. pp. 495–503.

Hollenstein, N., Renggli, C., Glaus, B., Barrett, M., Troendle, M., Langer, N., Zhang, C., 2021. Decoding EEG brain activity for multi-modal natural language processing. Front. Human Neurosci. 378.

Hromei, C.D., Margiotta, D., Croce, D., Basili, R., 2023. An end-to-end transformer-based model for interactive grounded language understanding. In: Proceedings of the Seventh Workshop on Natural Language for Artificial Intelligence (NL4AI 2023) Co-Located with 22th International Conference of the Italian Association for Artificial Intelligence. AI* IA 2023.

Huang, Y., Chen, Y., Li, Z., 2023. Applications of large scale foundation models for autonomous driving. arXiv preprint arXiv:2311.12144.

Huang, K., Han, Y., Wu, J., Qiu, F., Tang, Q., 2022. Language-driven robot manipulation with perspective disambiguation and placement optimization. IEEE Robot. Autom. Lett. 7 (2), 4188–4195.

Hwang, B., Lee, S., Han, H., 2022. LNFCOS: Efficient object detection through deep learning based on lnblock. Electronics 11 (17), 2783.

Ilyas, C.M.A., Rehm, M., Nasrollahi, K., Madadi, Y., Moeslund, T.B., Seydi, V., 2021. Deep transfer learning in human–robot interaction for cognitive and physical rehabilitation purposes. Pattern Anal. Appl. 1–25.

Inamdar, S., Chapekar, R., Gite, S., Pradhan, B., 2023. Machine learning driven mental stress detection on reddit posts using natural language processing. Human-Centric Intell. Syst. 3 (2), 80–91.

Islam, S., Dash, A., Seum, A., Raj, A.H., Hossain, T., Shah, F.M., 2021. Exploring video captioning techniques: A comprehensive survey on deep learning methods. SN Comput. Sci. 2 (2), 1–28.

Jang, B.-S., Park, A.J., Kim, I.A., 2022. Exploration of biomedical knowledge for recurrent glioblastoma using natural language processing deep learning models. BMC Med. Inform. Decis. Mak. 22 (1), 267.

Jianan, G., Kehao, R., Binwei, G., 2023. Deep learning-based text knowledge classification for whole-process engineering consulting standards. J. Eng. Res.

Johnston, P., Nogueira, K., Swingler, K., 2023. NS-IL: Neuro-symbolic visual question answering using incrementally learnt, independent probabilistic models for small sample sizes. IEEE Access.

Karasoy, O., Ballı, S., 2022. Spam SMS detection for Turkish language with deep text analysis and deep learning methods. Arab. J. Sci. Eng. 47 (8), 9361–9377.

Károly, A.I., Galambos, P., Kuti, J., Rudas, I.J., 2020. Deep learning in robotics: Survey on model structures and training strategies. IEEE Trans. Syst. Man Cybernet. Syst. 51 (1), 266–279.

Kasmaiee, S., Kasmaiee, S., Homayounpour, M., 2023. Correcting spelling mistakes in Persian texts with rules and deep learning methods. Sci. Rep. 13 (1), 19945.

Keele, S., et al., 2007. Guidelines for Performing Systematic Literature Reviews in Software Engineering. Technical report, ver. 2.3 ebse technical report. ebse.

Kenton, J.D.M.-W.C., Toutanova, L.K., 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NaacL-HLT, Vol. 1.

Kesavan, V., Muley, V., Kolhekar, M., 2019. Deep learning based automatic image caption generation. In: 2019 Global Conference for Advancement in Technology. GCAT, IEEE, pp. 1–6.

Keshinro, B., Seong, Y., Yi, S., 2022. Deep learning-based human activity recognition using RGB images in human-robot collaboration. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 66, No. 1. SAGE Publications Sage CA: Los Angeles, CA, pp. 1548–1553.

Khera, D., 2023. Leveraging the convolutional neural network (CNN) based on deep learning to classify and caption Images1.

Khodadadi, A., Ghandiparsi, S., Chuah, C.N., 2022. A natural language processing and deep learning based model for automated vehicle diagnostics using free-text customer service reports. Mach. Learn. Appl. 10, 100424.

Kim, J., 2024. Deep learning-based vehicle type and color classification to support safe autonomous driving. Appl. Sci. 14 (4), 1600.

Kim, M., Joe, I., 2023. Automatic diagnosis of medical conditions using deep learning with Symptom2Vec. IEEE Access.

Kim, D.-M., Lee, S.-H., Cheong, Y.G., 2018. Predicting emotion in movie scripts using deep learning. In: 2018 IEEE International Conference on Big Data and Smart Computing. BigComp, IEEE, pp. 530–532.

Kim, S., Lee, C.k., Choi, Y., Baek, E.S., Choi, J.E., Lim, J.S., Kang, J., Shin, S.J., 2021. Deep-learning-based natural language processing of serial free-text radiological reports for predicting rectal cancer patient survival. Front. Oncol. 11, 747250.

Kim, Y., Lee, J.H., Choi, S., Lee, J.M., Kim, J.H., Seok, J., Joo, H.J., 2020. Validation of deep learning natural language processing algorithm for keyword extraction from pathology reports in electronic health records. Sci. Rep. 10 (1), 20265.

Kitchenham, B., Brereton, O.P., Budgen, D., Turner, M., Bailey, J., Linkman, S., 2009. Systematic literature reviews in software engineering–a systematic literature review. Inf. Softw. Technol. 51 (1), 7–15.

Kopp, T., Baumgartner, M., Kinkel, S., 2021. Success factors for introducing industrial human-robot interaction in practice: an empirically driven framework. Int. J. Adv. Manuf. Technol. 112, 685–704.

Krishna, R., Lee, D., Fei-Fei, L., Bernstein, M.S., 2022. Socially situated artificial intelligence enables learning from human interaction. Proc. Natl. Acad. Sci. 119 (39), e2115730119.

Kumar, S., Rajesh, D.D., Pranesh, S., Kollipara, V.H., Agrawal, G.K., Anbarasi, M., Valarmathi, J., 2022. Classification of Indian media titles using deep learning techniques. Int. J. Cogn. Comput. Eng. 3, 114–123.

Kumar, V., Recupero, D.R., Riboni, D., Helaoui, R., 2020. Ensembling classical machine learning and deep learning approaches for morbidity identification from clinical notes. IEEE Access 9, 7107–7126.

Kunchukuttan, A., Mehta, P., Bhattacharyya, P., 2018. The IIT bombay english-hindi parallel corpus. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation. LREC 2018, European Language Resources Association (ELRA), Miyazaki, Japan, URL https://aclanthology.org/L18-1548.

Kushol, R., Parnianpour, P., Wilman, A.H., Kalra, S., Yang, Y.-H., 2023. Effects of MRI scanner manufacturers in classification tasks with deep learning models. Sci. Rep. 13 (1), 16791.

Lakomkin, E., Zamani, M.A., Weber, C., Magg, S., Wermter, S., 2018. On the robustness of speech emotion recognition for human-robot interaction with deep neural networks. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 854–860.

Larisch, R., Vitay, J., Hamker, F.H., 2023. Detecting anomalies in system logs with a compact convolutional transformer. IEEE Access.

Lefebvre, H., Legner, C., Fadler, M., 2021. Data democratization: toward a deeper understanding. In: ICIS.

Lei, L., Zhang, H., Yang, S.X., 2023. ChatGPT in connected and autonomous vehicles: benefits and challenges. Intell. Robot 3 (2), 145–148.

Lewis, M., 1998. Designing for human-agent interaction. AI Mag. 19 (2), 67.

Li, C., 2023. GitHub - lcroy/virtual-assistant-max: This is an open-source project, industrial virtual assistant, which uses NVIDIA Jetson. — github.com. https://github.com/lcroy/Virtual-Assistant-Max. (Accessed 6 November 2023).

Li, C., Chrysostomou, D., Yang, H., 2023a. A speech-enabled virtual assistant for efficient human–robot interaction in industrial environments. J. Syst. Softw. 205, 111818.

Li, D., Sun, J., Liu, Y., Du, W., 2023b. Data-centric content classification of smart city residential services. Authorea Preprints.

Li, B., Weng, Y., Ma, Z., Sun, B., Li, S., 2022a. Scene-aware prompt for multi-modal dialogue understanding and generation. In: CCF International Conference on Natural Language Processing and Chinese Computing. Springer, pp. 179–191.

Li, S., Yang, B., 2023. Personalized education resource recommendation method based on deep learning in intelligent educational robot environments. Int. J. Inf. Technol. Syst. Approach (IJITSA) 16 (3), 1–15.

Li, C., Zhang, X., Chrysostomou, D., Yang, H., 2022b. Tod4ir: A humanised task-oriented dialogue system for industrial robots. IEEE Access 10, 91631–91649.

Li, J., et al., 2023c. Advancing human-robot collaboration in manufacturing through deep learning-based natural language processing. J. Manuf. Syst. 58, 210–225.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: Common objects in context. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer, pp. 740–755.

Liu, F., Huang, S., Hu, J., Chen, X., Song, Z., Dong, J., Liu, Y., Huang, X., Wang, S., Wang, X., et al., 2023. Design of prime-editing guide RNAs with deep transfer learning. Nat. Mach. Intell. 1–14.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

Lu, Y., Tian, H., Cheng, J., Zhu, F., Liu, B., Wei, S., Ji, L., Wang, Z.L., 2022. Decoding lip language using triboelectric sensors with deep learning. Nature Commun. 13 (1), 1401.

Luger, E., Sellen, A., 2016. Like having a really bad PA: The gulf between user expectation and experience of conversational agents. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. pp. 5286–5297.

Machová, K., Mikula, M., Gao, X., Mach, M., 2020. Lexicon-based sentiment analysis using the particle swarm optimization. Electronics 9 (8), 1317.

Mahimaidoss, N.K., Sathianesan, G.W., 2023. Emotion identification in Twitter using deep learning based methodology. J. Electr. Eng. Technol. 1–18.

Majeed, A., Lee, S., 2020. Anonymization techniques for privacy preserving data publishing: A comprehensive survey. IEEE Access 9, 8512–8545.

Mao, J., Cao, Y., Zhang, Y., Huang, B., Zhao, Y., 2023. A novel method for identifying key genes in macroevolution based on deep learning with attention mechanism. Sci. Rep. 13 (1), 19727.

Mariani, J., Francopoulo, G., Paroubek, P., Vernier, F., 2022. NLP4NLP+ 5: The Deep (R) evolution in Speech and Language Processing. Front. Res. Metrics Anal. 7, 863126.

Martins, P.H., Cust'odio, L., Ventura, R., 2018. A deep learning approach for understanding natural language commands for mobile service robots. arXiv preprint arXiv:1807.03053.

Marulli, F., Verde, L., Campanile, L., 2021. Exploring data and model poisoning attacks to deep learning-based NLP systems. Procedia Comput. Sci. 192, 3570–3579.

Matti, R., Yousif, S., 2023. AutoKeras for fake news identification in arabic: Leveraging deep learning with an extensive dataset. Al-Nahrain J. Sci. 26 (3), 60–66.

Merdivan, E., Singh, D., Hanke, S., Holzinger, A., 2019. Dialogue systems for intelligent human computer interactions. Electron. Notes Theor. Comput. Sci. 343, 57–71.

Mithun, S., Jha, A.K., Sherkhane, U.B., Jaiswar, V., Purandare, N.C., Rangarajan, V., Dekker, A., Puts, S., Bermejo, I., Wee, L., 2023. Development and validation of deep learning and BERT models for classification of lung cancer radiology reports. Inform. Med. Unlocked 101294.

Mohammad, F., Khan, M., Marwat, S.N.K., Jan, N., Gohar, N., Bilal, M., Al-Rasheed, A., 2023. Text augmentation-based model for emotion recognition using transformers. Comput. Mater. Continua 76 (3).

Mohammed, S.N., Hassan, A.K.A., 2020. A survey on emotion recognition for human robot interaction. J. Comput. Inf. Technol. 28 (2), 125–146.

Moon, S., Chi, S., Im, S.B., 2022. Automated detection of contractual risk clauses from construction specifications using bidirectional encoder representations from transformers (BERT). Autom. Constr. 142, 104465.

Motyka, V., Vysotska, V., Chyrun, L., Vlasenko, O., Holoshchuk, R., Nagachevska, O., 2023. Information technology of transcribing Ukrainian-language content based on deep learning. In: 2023 IEEE 18th International Conference on Computer Science and Information Technologies. CSIT, IEEE, pp. 1–6.

Mubin, O., Henderson, J., Bartneck, C., 2014. You just do not understand me! speech recognition in human robot interaction. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication. IEEE, pp. 637–642.

Mukherjee, D., Gupta, K., Chang, L.H., Najjaran, H., 2022. A survey of robot learning strategies for human-robot collaboration in industrial settings. Robot. Comput.-Integr. Manuf. 73, 102231.

Nandini, D., Schmid, U., 2023. Explaining hate speech classification with model agnostic methods. arXiv preprint arXiv:2306.00021.

Nijhawan, T., Attigeri, G., Ananthakrishna, T., 2022. Stress detection using natural language processing and machine learning over social interactions. J. Big Data 9 (1), 1–24.

Niţoi, M., Pochea, M.M., Radu, Ş.C., 2023. Unveiling the sentiment behind central bank narratives: A novel deep learning index. J. Behav. Exp. Finance 38, 100809.

Nwana, H.S., 1996. Software agents: An overview. Knowl Eng. Rev. 11 (3), 205–244.

Olthof, A.W., van Ooijen, P.M., Cornelissen, L.J., 2021. Deep learning-based natural language processing in radiology: The impact of report complexity, disease prevalence, dataset size, and algorithm type on model performance. J. Med. Syst. 45, 1–16.

Orsag, L., Stipancic, T., Koren, L., 2023. Towards a safe human–robot collaboration using information on human worker activity. Sensors 23 (3), 1283.

Otter, D.W., Medina, J.R., Kalita, J.K., 2020. A survey of the usages of deep learning for natural language processing. IEEE Trans. Neural Netw. Learn. Syst. 32 (2), 604–624.

̈Ozer, E.G., Karapinar, I.N., Basbug, S., Turan, S., Utku, A., Akcayol, M.A., 2020. Deep learning based, a new model for video captioning. Int. J. Adv. Comput. Sci. Appl. 11, URL https://api.semanticscholar.org/CorpusID:214698333.

Pandey, S., Sharma, S., Wazir, S., 2022. Mental healthcare chatbot based on natural language processing and deep learning approaches: ted the therapist. Int. J. Inf. Technol. 14 (7), 3757–3766.

Pandy, A., Harangi, B., Hajdu, A., 2023. Extracting drug names from medical reports. In: 2023 IEEE 18th International Conference on Computer Science and Information Technologies. CSIT, IEEE, pp. 1–4.

Patel, R., Patel, S., 2021. Deep learning for natural language processing. In: Information and Communication Technology for Competitive Strategies (ICTCS 2020) Intelligent Strategies for ICT. Springer, pp. 523–533.

Peng, X., Shi, X., Dai, H., Jin, H., Ma, W., Xiong, Q., Yang, F., Qian, X., 2020. Capuchin: Tensor-based gpu memory management for deep learning. In: Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems. pp. 891–905.

Peng, S., Zeng, R., Liu, H., Cao, L., Wang, G., Xie, J., 2023. Deep broad learning for emotion classification in textual conversations. Tsinghua Sci. Technol. 29 (2), 481–491.

Phuc, D.T., Tran, Q.T., Van Tinh, N., Dau, S.H., 2022. Video captioning in Vietnamese using deep learning. Int. J. Electr. Comput. Eng. 12 (3), 3092.

Pitsilis, G.K., Ramampiaro, H., Langseth, H., 2018. Effective hate-speech detection in Twitter data using recurrent neural networks. Appl. Intell. 48, 4730–4742.

Poria, S., Hazarika, D., Majumder, N., Naik, G., Cambria, E., Mihalcea, R., 2018a. Meld: A multimodal multi-party dataset for emotion recognition in conversations. arXiv preprint arXiv:1810.02508.

Pritee, Z.T., Anik, M.H., Alam, S.B., Jim, J.R., Kabir, M.M., Mridha, M., 2024. Machine learning and deep learning for user authentication and authorization in cybersecurity: A state-of-the-art review. Comput. Secur. 103747.

Prottasha, N.J., Sami, A.A., Kowsher, M., Murad, S.A., Bairagi, A.K., Masud, M., Baz, M., 2022. Transfer learning for sentiment analysis using BERT based supervised fine-tuning. Sensors 22 (11), 4157.

Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., Huang, X., 2020. Pre-trained models for natural language processing: A survey. Sci. China Technol. Sci. 63 (10), 1872–1897.

Rafiq, G., Rafiq, M., Choi, G.S., 2023. Video description: A comprehensive survey of deep learning approaches. Artif. Intell. Rev. 1–80.

Rahman, M.M., Pramanik, M.A., Sadik, R., Roy, M., Chakraborty, P., 2020. Bangla documents classification using transformer based deep learning models. In: 2020 2nd International Conference on Sustainable Technologies for Industry 4.0. STI, IEEE, pp. 1–5.

Rajpurkar, P., Zhang, J., Lopyrev, K., Liang, P., 2016. Squad: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250.

Ren, Q., Hou, Y., Botteldooren, D., Belpaeme, T., 2023. Behavioural models of risk-taking in human–robot tactile interactions. Sensors 23 (10), 4786.

Rohrbach, A., Rohrbach, M., Tandon, N., Schiele, B., 2015. A dataset for movie description. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3202–3212.

Ruffolo, J.A., Chu, L.S., Mahajan, S.P., Gray, J.J., 2023. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. Nature Commun. 14 (1), 2389.

Rufus, N., Jain, K., Nair, U.K.R., Gandhi, V., Krishna, K.M., 2021. Grounding linguistic commands to navigable regions. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 8593–8600.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. nature 323 (6088), 533–536.

Russell, S.J., Norvig, P., 2016. Artificial Intelligence: A Modern Approach. Pearson.

Russo, A.G., Ciarlo, A., Ponticorvo, S., Di Salle, F., Tedeschi, G., Esposito, F., 2022. Explaining neural activity in human listeners with deep learning via natural language processing of narrative text. Sci. Rep. 12 (1), 17838.

Sanchez-Fernandez, A., Rumetshofer, E., Hochreiter, S., Klambauer, G., 2023. CLOOME: contrastive learning unlocks bioimaging databases for queries with chemical structures. Nature Commun. 14 (1), 7339.

Santur, Y., 2019. Sentiment analysis based on gated recurrent unit. In: 2019 International Artificial Intelligence and Data Processing Symposium. IDAP, IEEE, pp. 1–5.

Sari, W.K., Rini, D.P., Malik, R.F., Azhar, I.S.B., 2020. Sequential models for text classification using recurrent neural network. In: Sriwijaya International Conference on Information Technology and Its Applications. SICONIAN 2019, Atlantis Press, pp. 333–340.

Sarraju, A., Coquet, J., Zammit, A., Chan, A., Ngo, S., Hernandez-Boussard, T., Rodriguez, F., 2022. Using deep learning-based natural language processing to identify reasons for statin nonuse in patients with atherosclerotic cardiovascular disease. Commun. Med. 2 (1), 88.

Shafahi, A., Najibi, M., Ghiasi, M.A., Xu, Z., Dickerson, J., Studer, C., Davis, L.S., Taylor, G., Goldstein, T., 2019. Adversarial training for free!. Adv. Neural Inf. Process. Syst. 32.

shahhaard47, 2020. Shahhaard47/script-generation: Generating movie scripts by genre using CTRL framework and GPT-2 github. URL https://github.com/shahhaard47/Script-Generation.

Shanmugavadivel, K., Sathishkumar, V., Raja, S., Lingaiah, T.B., Neelakandan, S., Subramanian, M., 2022. Deep learning based sentiment analysis and offensive language identification on multilingual code-mixed data. Sci. Rep. 12 (1), 21557.

Sharfuddin, A.A., Tihami, M.N., Islam, M.S., 2018. A deep recurrent neural network with bilstm model for sentiment classification. In: 2018 International Conference on Bangla Speech and Language Processing. ICBSLP, IEEE, pp. 1–4.

Shen, C., Huang, T., Liang, X., Li, F., Fu, K., 2018. Chinese knowledge base question answering by attention-based multi-granularity model. Information 9 (4), 98.

Shervedani, A.M., Li, S., Monaikul, N., Abbasi, B., Di Eugenio, B., Zefran, M., 2023. An end-to-end human simulator for task-oriented multimodal human-robot collaboration. arXiv preprint arXiv:2304.00584.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6 (1), 1–48.

Shrestha, S., Zha, Y., Banagiri, S., Gao, G., Aloimonos, Y., Fermuller, C., 2024. NatSGD: A dataset with speech, gestures, and demonstrations for robot learning in natural human-robot interaction. arXiv preprint arXiv:2403.02274.

Siciliano, B., 2008. Springer handbook of robotics. Springer-Verlag google schola 2, 15–35.

Slack, D., Krishna, S., Lakkaraju, H., Singh, S., 2023. Explaining machine learning models with interactive natural language conversations using TalkToModel. Nat. Mach. Intell. 5 (8), 873–883.

Smith, M., Williams, R., 2023. Enhancing customer service robots with deep learning-based NLP for improved dialogue management. Int. J. Hum.-Comput. Interact. 37 (4), 321–336.

Soori, M., Arezoo, B., Dastres, R., 2023. Artificial intelligence, machine learning and deep learning in advanced robotics, a review. Cogn. Robotics.

Su, L., He, T., Fan, Z., Zhang, Y., Guizani, M., 2019. Answer acquisition for knowledge base question answering systems based on dynamic memory network. IEEE Access 7, 161329–161339.

Sun, Y., Wang, S., Li, Y.-K., Feng, S., Chen, X., Zhang, H., Tian, X., Zhu, D., Tian, H., Wu, H., 2019. ERNIE: enhanced representation through knowledge integration. CoRR abs/1904.09223 (2019). arXiv preprint arXiv:1904.09223.

Takano, W., 2020. Annotation generation from IMU-based human whole-body motions in daily life behavior. IEEE Trans. Hum.-Mach. Syst. 50 (1), 13–21.

Tan, H.L., Leong, M.C., Xu, Q., Li, L., Fang, F., Cheng, Y., Gauthier, N., Sun, Y., Lim, J.H., 2020. Task-oriented multi-modal question answering for collaborative applications. In: 2020 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 1426–1430.

Tejaswini, V., Babu, K.S., Sahoo, B., 2022. Depression detection from social media text analysis using natural language processing techniques and hybrid deep learning model. ACM Trans. Asian Low-Res. Lang. Inform. Process..

Thakur, P., Shahu, R., Gupta, V., 2023. Audio and text-based emotion recognition system using deep learning. In: Advances in Signal Processing, Embedded Systems and IoT: Proceedings of Seventh ICMEET-2022. Springer, pp. 447–459.

Tohma, K., Okur, H.I., Kutlu, Y., Sertbas, A., 2023. Sentiment analysis in Turkish question answering systems: An application of human-robot interaction. IEEE Access.

Trueman, T.E., Gopi, K., Kumar, A., 2021. Online text-based humor detection. In: 2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications, Vol. 1. CENTCON, IEEE, pp. 313–316.

Ullah, A., Qi, G., Hussain, S., Ullah, I., Ali, Z., 2024. The role of LLMs in sustainable smart cities: Applications, challenges, and future directions. arXiv preprint arXiv:2402.14596.

Varma, S., Peter, J.D., 2022. Deep learning-based video captioning technique using transformer. In: 2022 8th International Conference on Advanced Computing and Communication Systems, Vol. 1. ICACCS, IEEE, pp. 847–850.

Vashistha, N., Singh, K., Shakya, R., 2022. Active learning for neural machine translation. arXiv preprint arXiv:2301.00688.

Vemprala, S.H., Bonatti, R., Bucker, A., Kapoor, A., 2024. Chatgpt for robotics: Design principles and model abilities. IEEE Access.

Verina, Y., Tolstoukhov, D., Egorov, D., Kravchenko, O., Sunchalina, A., 2023. Hybrid model for sentiment analysis based on both text and audio data trained on MELD. In: AIP Conference Proceedings, Vol. 2819, No. 1. AIP Publishing.

Villa-Pérez, M.E., Trejo, L.A., Moin, M.B., Stroulia, E., 2023. Extracting mental health indicators from English and Spanish social media: A machine learning approach. IEEE Access 11, 128135–128152.

Vrins, A., Pruss, E., Ceccato, C., Prinsen, J., De Rooij, A., Alimardani, M., De Wit, J., 2024. Wizard-of-oz vs. GPT-4: A comparative study of perceived social intelligence in HRI brainstorming. In: Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction. pp. 1090–1094.

Wahab, A., Tayara, H., Xuan, Z., Chong, K.T., 2021. DNA sequences performs as natural language processing by exploiting deep learning algorithm for the identification of N4-methylcytosine. Sci. Rep. 11 (1), 212.

Wan, Z., 2023. Text classification: A perspective of deep learning methods. arXiv preprint arXiv:2309.13761.

Wang, P., Qian, Y., Soong, F.K., He, L., Zhao, H., 2015. A unified tagging solution: Bidirectional lstm recurrent neural network with word embedding. arXiv preprint arXiv:1511.00215.

Wang, Z., Wang, W., Chen, Q., Wang, Q., Nguyen, A., 2023a. Generating valid and natural adversarial examples with large language models. arXiv preprint arXiv:2311.11861.

Wang, S., Wang, L., Li, F., Bai, F., 2023b. DeepSA: a deep-learning driven predictor of compound synthesis accessibility. J. Cheminform. 15 (1), 103.

Wei, J., Zou, K., 2019. Eda: Easy data augmentation techniques for boosting performance on text classification tasks. arXiv preprint arXiv:1901.11196.

Wen, T.H., Gasic, M., Mrksic, N., Su, P.H., Vandyke, D., Young, S., 2015. Semantically conditioned lstm-based natural language generation for spoken dialogue systems. arXiv preprint arXiv:1508.01745.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., et al., 2020. Transformers: State-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. pp. 38–45.

Wooldridge, M., 2009. An Introduction to Multiagent Systems. John wiley & sons.

Xavier, B.A., Chen, P.H., 2022. Natural language processing for imaging protocol assignment: Machine learning for multiclass classification of abdominal CT protocols using indication text data. J. Dig. Imag. 35 (5), 1120–1130.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R.R., Le, Q.V., 2019. Xlnet: Generalized autoregressive pretraining for language understanding. Adv. Neural Inf. Process. Syst. 32.

Yao, K., Peng, B., Zhang, Y., Yu, D., Zweig, G., Shi, Y., 2014. Spoken language understanding using long short-term memory neural networks. In: 2014 IEEE Spoken Language Technology Workshop. SLT, IEEE, pp. 189–194.

Yohanes, D., Putra, J.S., Filbert, K., Suryaningrum, K.M., Saputri, H.A., 2023. Emotion detection in textual data using deep learning. Procedia Comput. Sci. 227, 464–473.

Zaheer, K., Talib, M.R., Hanif, M.K., Sarwar, M.U., 2023. A multi-kernel optimized convolutional neural network with Urdu word embedding to detect fake news. IEEE Access.

Zeng, Z., Yao, Y., Liu, Z., Sun, M., 2022. A deep-learning system bridging molecule structure and biomedical text with comprehension comparable to human professionals. Nature Commun. 13 (1), 862.

Zhang, C., Chen, J., Li, J., Peng, Y., Mao, Z., 2023a. Large language models for human-robot interaction: A review. Biomim. Intell. Robotics 100131.

Zhang, W., Ding, Y., Zhang, M., Zhang, Y., Cao, L., Huang, Z., Wang, J., 2023b. TCPCNet: a transformer-CNN parallel cooperative network for low-light image enhancement. Multimedia Tools Appl. 1–16.

Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M., Liu, Q., 2019. ERNIE: Enhanced language representation with informative entities. arXiv preprint arXiv:1905.07129.

Zhang, M., Tian, G., Zhang, Y., Duan, P., 2021. Service skill improvement for home robots: Autonomous generation of action sequence based on reinforcement learning. Knowl.-Based Syst. 212, 106605.

Zhang, J., Zhao, Y., Saleh, M., Liu, P., 2020. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In: International Conference on Machine Learning. PMLR, pp. 11328–11339.

Zhou, S., Han, Y., Chen, N., Huang, S., Igorevich, K.K., Luo, J., Zhang, P., 2023a. Transformer-based discriminative and strong representation deep hashing for cross-modal retrieval. IEEE Access 11, 140041–140055. http://dx.doi.org/10.1109/ACCESS.2023.3339581.

Zhou, H., Tang, S., Huang, W., Zhao, X., 2023b. Generating risk response measures for subway construction by fusion of knowledge and deep learning. Autom. Constr. 152, 104951.

Zulqarnain, M., Ghazali, R., Ghouse, M.G., Mushtaq, M.F., 2019. Efficient processing of GRU based on word embedding for text classification. JOIV: Int. J. Inform. Visual. 3 (4), 377–383.

Zulqarnain, M., Ghazali, R., Hassim, Y.M., Rehan, M., 2020. Text classification based on gated recurrent unit combines with support vector machine. Int. J. Electr. Comput. Eng. 10 (4), 3734–3742.