Recent Trends in Pharmaceutical Technology & Industries
Volume 7, Print Issue 1 - 2025, Pg. No. 1-14
Peer Reviewed Journal

Research Article

Transforming Healthcare with Natural Language Processing: A New Era of Innovation

Dinesh Kumar[1], Rajni Tanwar[2], Divya Jain[3], Kavita Behmani[4], Sonakshi Antal[5], Subhani Jarora[6], Rahul Pal[7], Prachi Panday[8]

[1,2]School of Pharmacy, Desh Bhagat University, Mandi Gobindgarh, Punjab, India
[3]Department of Microbiology, School of Applied & Life Sciences, Uttaranchal University, Dehradun, Uttarakhand, India
[4]Departmental of Pharmaceutical Sciences, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India
[5]SRM Modinagar College of Pharmacy, SRMIST NCR campus, Ghaziabad, Uttar Pradesh, India
[6]Department of Pharmacology, Starex University, Gurugram, Haryana, India
[7]Department of Pharmaceutics, ISF College of Pharmacy, Moga, Punjab, India
[8]Faculty of Pharmaceutical Sciences, Rama University, Kanpur, Uttar Pradesh, India

I N F O

Corresponding Author:
Dinesh Kumar, Associate Professor, School of Pharmacy, Desh Bhagat University, Mandi Gobindgarh, Punjab, India
E-mail Id:
dineshpotlia123@gmail.com
How to cite this article:
Kumar D, Tanwar R, Jain D, Behmani K, Antal S, Jarora S, Pal R, Pandey P. Transforming Healthcare with Natural Language Processing: A New Era of Innovation. *Rec Trends Pharm Tech Indl* 2025 7(1): 1-14.

Date of Submission: 2025-01-02
Date of Acceptance: 2025-02-04

A B S T R A C T

Natural Language Processing (NLP) presents challenges in the health science field due to language complexities. NLP, a branch of artificial intelligence (AI), abstracts language into logical components to extract valuable information from unstructured text. It emerged in the 1940s, mainly for autonomous language translation. In healthcare, NLP enhances patient care by processing unstructured data in Electronic Health Records (EHRs), improving diagnostic accuracy and treatment efficiency. It supports Clinical Decision Support Systems (CDSS) by providing critical information for medical decisions, reducing errors, and improving patient safety. NLP also aids clinical trial screening, automating patient selection for studies, and contributes to the Internet of Medical Things (IoMT) for real-time monitoring. Additionally, AI-powered tools like CureSee for visual therapy and applications in Alzheimer's diagnosis offer significant improvements in patient care and early disease detection. NLP further supports clinical research and predictive analytics, facilitating novel therapy discovery and early disease diagnosis.

Keywords: Natural Language Processing, Healthcare, Equipment, Medical. Software, Knowledge

Introduction

Natural language processing is the capacity of a computer program to understand human language (both spoken and written) in its natural form.[1-3] It is an integral part of artificial intelligence. Natural Language Processing has a history spanning over 50 years and originated from the linguistics discipline.[4-5] This technology has a wide range of practical uses in many domains, including medical research, search engines, and corporate intelligence.[6-8] Natural Language Processing (NLP) uses either rule-based or machine learning methodologies to comprehend the organisation and significance of linguistic material.[9-10] Chatbots, voice assistants, text-based scanning tools, translation apps, and corporate software use artificial intelligence (AI) to support company operations, boost productivity, streamline various procedures, and advance voice-controlled home automation systems, personal digital assistants, and language translation, all of which smartphone users use on a daily basis.[11] Currently, there is a growing trend in the medical field to implement this technique in order to enhance the use of unstructured electronic health records (EHRs) and facilitate communication with patients for the purpose of answering inquiries and conducting consultations.[12–13] NLP techniques have shown their usefulness in addressing the problem of information overload in the field of health and medicine. Examples of such applications include treatment analysis, information extraction and retrieval from extensive discharge summaries, aggregation and summarisation of patient notes, and semantic understanding of patient questions.[14] Health NLP encompasses various methodological research areas, including the development of NLP models to process medical or social web data (e.g., literature, EHRs, clinical trials, and healthcare social media), the retrieval and extraction of health information, the aggregation, abstracting, and summarising of health information using NLP, the application of machine learning-based text mining methods for healthcare, the representation and reasoning of health care knowledge, the construction and annotation of health and health knowledge graphs, and more.[15-18]

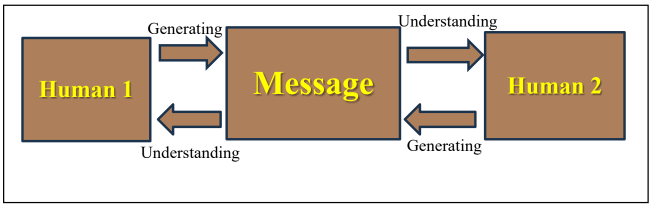

Figure 1.How communication happens in the human language and how NLP originated

The Evolution of Natural Language Processing

Many fields, including computer science and advances in computational linguistics that originated in the mid-20th century, influence Natural Language Processing (NLP). The following significant milestones have marked its development:

1950s

Alan Turing introduced the Turing Test in this decade as a way of determining a computer's genuine intelligence, thus establishing the framework of natural language processing. As a measure of intelligence, the exam evaluates the ability to automatically read and generate natural language.[19]

1950s-1990s

NLP was primarily rules-based, using carefully established rules created by linguists to dictate computers' language processing methodology. In 1954, the Georgetown-IBM experiment distinguished itself as a remarkable showcase of machine translation, successfully translating over 60 lines from Russian to English. Rule-based parsing, morphology, semantics, and other models of natural language comprehension emerged throughout the 1980s and 1990s.[20]

1990s

Due to the improved efficiency of building NLP technologies facilitated by computational improvements, a statistical approach replaced the traditional top-down, language-first method of natural language processing. The increasing speed of computers enabled the development of rules based on linguistic data without the need for a linguist to establish all the rules. The broad use of data-driven natural language processing occurred throughout this decade. Natural language processing has transitioned from a linguist-focused methodology to an engineer-focused methodology, relying on a broader range of scientific fields rather than only linguistics.[21]

2000-2020s

The word "natural language processing" saw significant growth in popularity. We also presented the exploration of NLP processes using unsupervised and semi-supervised machine learning methods. Due to improvements in computational capabilities, natural language processing has also acquired a multitude of practical uses. Natural Language Processing (NLP) has also started to drive other applications such as chatbots and virtual assistants. Contemporary approaches to Natural Language Processing (NLP) integrate classical linguistics with statistical techniques.[22-23]

In 2021, the worldwide NLP market in healthcare and life sciences was valued at USD 1.88 billion and is projected to continue growing at a compound annual growth rate (CAGR) of 19.3% during the forecast period. Natural Language Processing (NLP) is a methodology that examines the interaction between humans and computers to help robots comprehend both written and spoken human language. Polaris market research predicts the size of the NLP Healthcare and Lifesciences market from 2018 to 2030 (Figure 2).

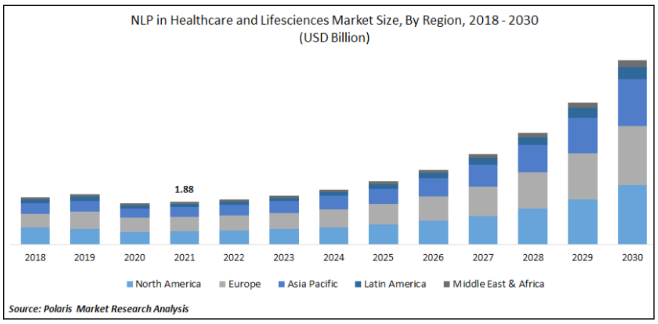

Figure 2.NLP in Healthcare and Lifesciences Market size, By region 2018-2030.

Components of NLP

We can divide Natural Language Understanding into two distinct components: natural language generation and natural language understanding. Figure 1 illustrates the overarching classification of NLP. Customers use NLP in customer service applications to understand their verbal or written issues. Linguistics, on the other hand, is the scientific study of language forms, meaning, and context. Consequently, it is critical to comprehend linguistics.[24-25]

Natural Language Understanding (NLU)

NLU improves computers' understanding and evaluation of natural language by extracting ideas, entities, emotions, keywords, and so on. Customer service applications use this technology to comprehend the issues clients communicate, whether orally or in written form.[26] Linguistics is the scientific study of language meaning, conversational context, and different linguistic forms. This text emphasises the need to comprehend many significant terms and degrees of Natural Language Processing (NLP).[27-28]

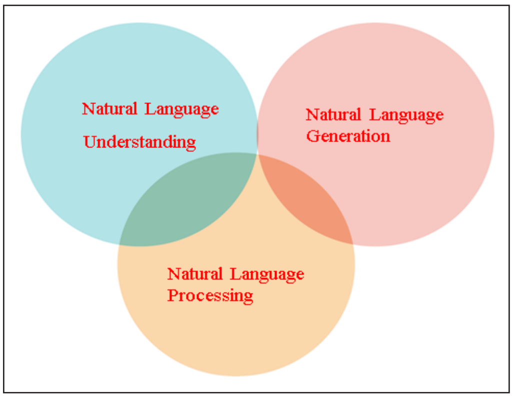

Figure 3.General process of Natural Language Processing (NLP)[27-28]

Natural Language Generation (NLG)

NLG is the process of translating an internal representation into meaningful phrases, sentences, and paragraphs. It is a component of natural language processing and consists of the following four stages: goal identification, planning for goal attainment through evaluation of the circumstances and available communicative sources, and plan implementation as a text.[29–30]

- Natural language components
- Speakers and Generators
- Component and level of representation

1. Textual organisation
2. Realisation
3. Content selection
4. Linguistic Approaches
5. speaker's application

How does NLP work for healthcare?

Text Input Analysis

Natural Language Processing (NLP) starts by parsing large quantities of textual data, often in the format of clinical notes, patient records, or medical literature. Tokenisation breaks down the text into smaller entities, like individual words or phrases, thereby creating a systematic foundation for analysis.

Language Understanding

Natural Language Processing (NLP) uses syntactic and semantic analysis to understand the connections between words and distinguish the subtleties of language. Named Entity Recognition (NER) is a technique that recognizes explicit things such as illnesses, drugs, or processes, thereby facilitating a more comprehensive understanding of the surrounding environment.

- **Feature Extraction:** We examine the analysed text to extract key aspects, including pertinent medical terminology. This stage is crucial in systematically organising the data for further analysis.
- **Algorithmic Processing:** We use modern machine learning algorithms, trained on a wide range of healthcare datasets. These algorithms use the obtained characteristics to make precise predictions and accurately classify information.
- **Output Generation:** The final product includes organized data or practical suggestions derived from the original mixed text. This could be anything from diagnosis codes to treatment suggestions. Figure 4 will discuss the NLP work process in healthcare.

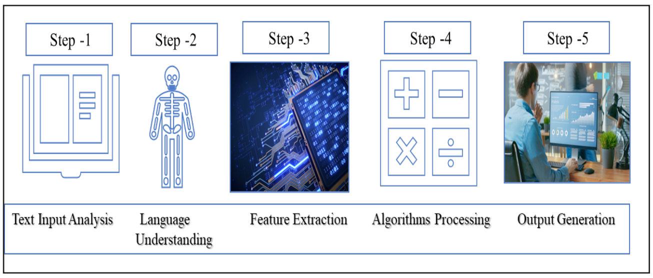

Figure 4.Work of NLP in Healthcare.

The application of Natural Language Processing (NLP) in advancing the healthcare system

Natural language processing applications extend beyond simple NLP applications, inspiring numerous natural language processing project concepts. Tech advancements boost the role of natural language processing in the healthcare sector, providing modern opportunities to improve patient care, reduce costs, and boost the productivity of healthcare procedures globally.

Electronic Health Records (EHRs)

Healthcare professionals use NLP to categorise and interpret vast amounts of unstructured text in EHRs, enabling them to extract valuable data for patient care, study, and decision-making. Electronic health records (EHRs) are computerised versions of the medical background records maintained by patients' healthcare providers. Physicians typically retain a substantial quantity of consultation patient information in the form of free text. By transforming the information into structured data, health systems are able to categorise patients and provide an overview of their condition upon their arrival. Instead of squandering valuable time examining their electronic health records (EHRs), physicians can extract crucial insights using NLP. Researchers may find it easier to obtain and combine clinical data using the Electronic Health Record System (EHR). HER may enable the prediction of early diagnosis, prevention of various illnesses, complications, disease progression, and death (Figure 5).

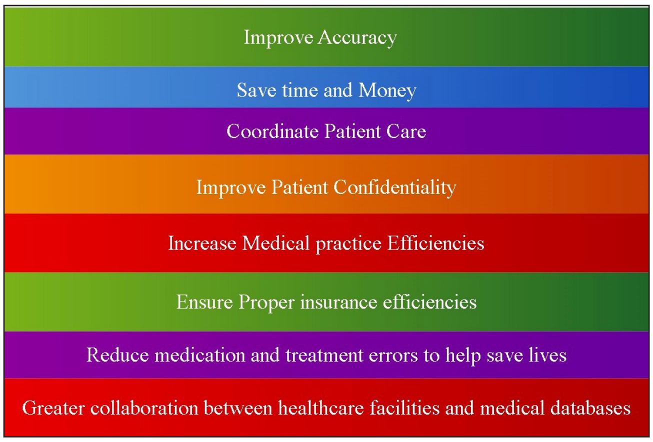

Figure 5.Benefits of the HER system

Automated Documentation

This is one of the most important natural language processing use cases, and automated documentation is a crucial part. NLP can transcribe and collect communication among healthcare establishments and vendors, confirming that all interactions are documented and efficiently accessible for future authority, compliance, and auditing.

Contract Management

Contract management can help in clinical natural language processing by extracting and analyzing key data from contracts between suppliers and vendors. This makes it easier to monitor terms, compliance, and renewals with service level agreements (SLAs).

Clinical Decision Support Systems (CDSS)

NLP also offers the Clinical Decision Support System (CDSS), a medical decision support system that enhances healthcare delivery by augmenting clinical judgements with patient-centered medical knowledge, health information, and other relevant data. A conventional Clinical Decision Support System (CDSS) is software that facilitates straightforward clinical decision-making by requiring the clinician to match the specified symptoms of an individual patient with a computerized clinical knowledge base.[52] Subsequently, the system generates specific and reliable recommendations based on the indicated symptoms, aiding the clinician in making informed medical judgements. Electronic health records (EHR), computerised physician order entry (CPOE), and online apps are the current means by which CDSS functions.

CDSS operates on two tiers

- **Knowledge-based CDSS:** This tier relies on previously registered knowledge or the patient's medical history to function. Such technological systems provide straightforward advantages such as knowledge documentation, rational decision support, self-education, reasoning, and explanation.
- **CDSS non-knowledge-based:** Non-knowledge-based CDSS differ from knowledge-based ones in that they enforce a structure of artificial intelligence known as machine learning, as opposed to being based on user-defined perception. By employing this methodology, a system "learns" from past experiences and subsequently incorporates these "lessons" into its database, thereby eliminating the need to consult a premade encyclopaedia. Non-knowledge-based CDSS artificial neural networks and genetic algorithms are two well-known varieties.

Dragon Medical One

Dragon Nuance, a speech recognition application, is widely used by medical professionals to maintain their electronic medical records through direct dictation. Dictation is possible for any report, and the system correctly recognises each word with 99% precision. Using natural language commands, you can easily format, revise, and navigate notes to optimise workflows. This application is accessible through the microphone of any device.

Predictive Analytics

Healthcare uses natural language processing to predict disease flareups, patient outcomes, readmission threats, and other aspects, enabling the implementation of visionary healthcare interventions. NLP in healthcare, when coupled with machine learning (ML), can predict future supply demands based on existing data, trends, and patterns in interactions and orders, allowing proactive inventory surveillance.

Medical Research and Literature Analysis

NLP filters vast volumes of medical literature and analyses data, fostering the discovery of undiscovered medical insights and the synthesis of existing information. NLP filters vast volumes of medical literature and analyses data, fostering the discovery of undiscovered medical insights and the synthesis of existing information.

Health Surveillance and Public Health Monitoring

We scrutinise disease eruptions and general health sensations by scrutinising social media, news, and other textual data.

Personal Health Assistants

This idea also falls under the category of natural language processing projects. NLP drives conversational agents and chatbots to supply health information, give patients reminders about drugs, aid in monitoring health issues, and make overall healthcare a bit more accessible.

Internet of Medical Things (IOMT)

The Internet of Things (IoT) is a network of networked machines, computers, items, people, and animals that have individual IDs and can transmit data between them without the need for human interaction. It includes wearable technology, implantable devices, and sensors that can track vital signs like blood pressure, blood sugar levels, and heart rate. It also includes monitoring and control systems that enable homes to be smart, such as thermostats, Internet of Things-enabled HVAC, and thermostatic equipment.

The following list includes a number of IoMT advantages.

- **Better patient outcomes:** Because IoMT devices monitor patients in real time, healthcare professionals can react promptly.
- **Enhanced productivity:** By automating healthcare procedures, IoMT devices reduce workload and save time.
- **Increased patient involvement:** IoMT devices provide individuals access to their medical records.
- **Improved Preventive Care:** IoMT devices may assist medical professionals in identifying individuals who are at risk of developing illnesses and taking preventative action.
- **Remote Patient Monitoring:** IoMT devices enable medical professionals to monitor patients remotely, especially those with mobility issues.

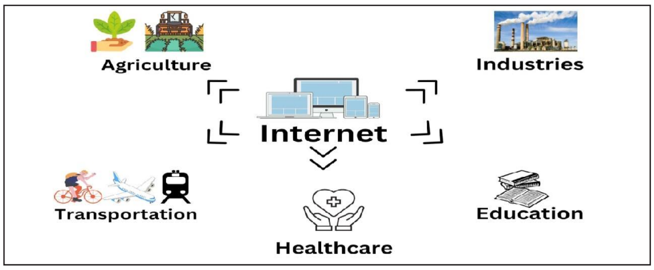

Figure 6.Overall perspective highlighting the function of IoT across several fields.

SyTrue

Many programs, notably SyTrue, simplify life by offering these services: It is a healthcare technology business that provides cutting-edge, data-driven solutions to complex healthcare issues, creating value and insights. SyTrue, a private healthcare company, employs its own machine learning, AI, and NLP tools. Claim Logic employs data scientists, architects, designers, solution experts, medical professionals, nurses, and programmers.

Feedback Analysis

Feedback analysis is one of the best natural language processing use cases to collect and evaluate healthcare staff comments on suppliers and products. Additionally, it aids in uncovering fresh perspectives and patterns that can influence future procurement strategies and vendor negotiations.

Invoice Processing

It can automate invoice data processing and extraction, reducing errors and shrinking the payment cycle, which is critical for maintaining positive associations with suppliers and vendors.

Clinical Documentation Management

NLP revolutionises clinical documentation by converting spoken language into structured data. It automates the transcription of doctor-patient interactions, making it faster and more accurate than traditional manual methods. It improves healthcare efficiency by reducing clinicians' time on paperwork, allowing them to focus more on patient care. Additionally, NLP systems can analyse and extract key information from these documents, facilitating better decision-making and continuity of care across different healthcare settings.

Regulatory Compliance

Clinical natural language processing can assist with monitoring conversations and documents for adherence to healthcare legislation and norms, thereby identifying and mitigating risks.

Clinician Note Generation Using Amazon Connect and Amazon Transcibe

Amazon Transcribe utilises sophisticated machine learning algorithms to transform spoken language into written text. This service operates in the cloud, providing efficient and accurate speech-to-text conversion.

Here is a thorough overview of its functionality:

- To make use of the Amazon Transcribe service, a user is required to submit an audio file in one of the supported file formats, such as MP3, WAV, FLAC, etc.
- Amazon Transcribe utilises a sophisticated automatic speech recognition (ASR) engine to analyse the audio

and transform it into written text. The engine employs deep neural networks and other machine learning techniques to identify words and phrases in the audio stream.

- **Language Identification:** If the user does not select a language in Amazon Transcribe, Amazon Transcribe will automatically recognise and translate the audio file's language.
- **Punctuation and Formatting:** Amazon Transcribe enhances readability by automatically capitalising words and inserting punctuation in the transcribed text.
- Users have the ability to provide Amazon Transcribe with a customised vocabulary, which may enhance the accuracy of transcriptions by including specific phrases that are important to a certain company or area.
- An API provides the transcript to the user in either a JSON or plain text file format. While processing the audio, the user has the option to receive real-time output.

Organizations and developers frequently use Amazon Transcribe, a dependable and affordable option, to convert audio recordings into text. Quick and efficient completion of this process contributes to the enhancement of the healthcare system.

NLP Approaches in Healthcare

- **Clinical Practice**: Clinical data encompasses a wide range of information, such as demographics, medical history, comorbidities, physical examination records, electronic device recordings, laboratory testing data, and medical images. These are the most critical pieces of information for the purposes of diagnosis, treatment, and retrospective analysis. Critical to obtaining firsthand clinical data is patient-provider communication.
- **Scarcity of medical resources:** which include space, personnel, and materials; resource allocation efficiency is of the utmost importance in hospitals and other medical facilities. The implementation of patient triage systems enables medical resources to prioritise critical cases and improve the effectiveness and efficacy of resource allocation. Furthermore, the utilisation of text generation techniques enables machines to perform a portion of the text writing in the healthcare industry, particularly routine reports, thereby relieving medical personnel of numerous administrative responsibilities and reclaiming their time for direct patient care.
- **Quality assurance of services:** Hospitals can enhance their service quality and the patient experience by incorporating patient experience feedback into sentiment analysis. NLP simplifies this task and significantly increases the efficacy of sentiment analysis. Previously, such analysis demanded substantial human resources.

Personal health assistants facilitate remote access to healthcare services and valuable medical information, eliminating the need for individuals to physically visit healthcare facilities. For a variety of objectives, personal health attendants may integrate subsystems including medical information access systems and remote healthcare systems.

NLP techniques have the potential to significantly improve the quality of life and social integration of disabled and elderly individuals. Voice-activated home automation systems and robots have the potential to provide daily assistance to individuals who are geriatric or disabled.

Furthermore, robots, particularly androids and other humanoid-communicating models, can inspire and accompany them during social interactions. Furthermore, NLP techniques offer significant utility in assisting individuals with diverse disabilities, including but not limited to speech impairments hearing loss, dyslexia, and neurological disorders.

Public Health

Medical education and health knowledge popularisation are crucial public health interventions because they increase health literacy and encourage the development of healthy lifestyles. In particular, question answering systems, information retrieval systems and machine translation systems enable individuals to conveniently obtain medical knowledge, thereby promoting its dissemination and instruction. Furthermore, text generation methodologies, including text summarization and question generation, can be implemented in medical education to produce simplified summaries and medical case-based questions.

Performing population screenings- is a significant public health intervention in addition to the dissemination of health information. Target populations are identified prior to administering the screening test as part of the population screening process. Considering the results of the screening, subsequent actions such as additional tests, advice, or treatment may be administered.

There are two primary functions that NLP can fulfil in population surveillance. NLP aids in the identification of populations characterised by elevated health risk factors, potentially enhancing the efficacy of population screening efforts. Secondly, NLP can aid in the analysis of healthcare surveys and questionnaires, particularly those containing open-ended questions.

- **Preclinical investigation:** NLP techniques, particularly information extraction, can also discern the connections between chemical structures and biological activity. Moreover, they aid researchers in their quest for potentially efficacious chemical compounds through virtual screening within the vast chemical space. Furthermore, they find application in preclinical research for the prognosis of deleterious drug reactions, such as toxicity and side effect prediction. NLP may facilitate streamlined clinical trial design, patient recruitment clinical trial analytics, and other processes throughout the clinical research lifecycle. Natural language processing (NLP) has a wide and vital function in medication research; for example, the Food and medication Administration (FDA) and other institutions have lately shown interest in using NLP to identify adverse drug events and monitor drug safety.

- **Psychological well-being:** Mental health problems have garnered extensive and steadily growing attention over the course of many years. Specifically, the World Health Organisation (WHO) said that the epidemic and subsequent lockdowns, economic insecurity, anxiety, and uncertainty have had profound and detrimental effects on global mental health in recent years.

The use of natural language processing (NLP) in smart healthcare has significant importance in the accurate prediction, diagnosis, and treatment of mental health disorders. Natural Language Processing (NLP) methods have been used to forecast or detect several mental problems at an early stage, including psychiatric illness, late-life depression, and severe mental illnesses such as schizophrenia, schizoaffective disorder, and bipolar disorder.

Methods of NLP

Two kinds of natural language processing:

1. Subject Modelling
2. Word2Vec.

- **Subject Modelling:** Subject modelling is a statistical technique in Natural Language Processing (NLP) that uses algorithms to condense extensive text collections, known as a corpus, into distinct subject clusters. A subject refers to a recurring theme that is regularly seen in a corpus. These themes are linked to a probability that measures the degree to which the subject is present in the corpus. A vocabulary, also known as a dictionary, is used to assess the probability of a word or phrase being utilised in a certain circumstance. Topic models shed light on the similarities between topics by using statistical methods. The textual data was analysed, and the most appropriate themes were provided using a latent Dirichlet allocation (LDA) model in this study. After running the LDA model on the corpus, a set of themes is produced, together with the word probabilities for each theme. This method uses pre-existing words and phrases in the corpus to establish themes; it is a non-generative approach to natural language processing. As an added bonus, the method relies only on statistical analysis. Words' contextual meaning, grammar, and practicality are not considered in the study. Rather, it is hyper-focused on phrase frequency within the corpus.
- **Word2Vec:** Word2Vec is an NLP solution that uses neural networks to quantify words based on their semantic significance. To show where these numbers are in n-dimensional space, we use vectors, which include both the direction and the magnitude of the value. This concept is similar to how a vector in three-dimensional space represents a force acting on an object.

To improve word vector representations from large datasets, this method employs shallow neural networks. Word vectors of 300 dimensions, encoding words' semantic significance, are the end result. Analysing the word contexts in sentences determines these vectors. To deduce the semantic associations between various portions of the examined text, Word2Vec uses a shallow neural network, which mimics the way the human brain works. Since the method depends on the usage of a word to ascertain its semantic importance, it is both pragmatic and purposeful in its emphasis on word meaning. Nonetheless, syntax is not considered at all. To discern the practical significance of a response section, word embedding codes are extracted from it, in contrast to open coding, which entails transforming a respondent's main ideas into descriptors. Word embedding codes, in contrast to open coding, do not generate any new words. A good code should have two things: first, it should convey a thorough and succinct understanding of the text it is encoding, which means it should accurately represent the meaning of the text; and second, it should have semantic depth, meaning it captures the important semantic essence that underlies the specific significance of a passage.

Datasets in NLP

Collections of natural language processing A corpus is a database of language information derived from written or spoken materials. The primary goal of creating a corpus is to test a theory about language, such as the universality of a given sound, word, or grammatical arrangement. Different kinds of corpora exist: An annotated corpus is one in which certain annotations have made clear the information that was implicit in the plain text. The raw state of plain text is included in the un-annotated corpus. A reference corpus allows for the comparison of several languages. Lexicography makes extensive use of monitor corpora, which are collections of texts that are neither infinite nor finite. One kind of corpus is the multilingual corpus, which is essentially just a collection of smaller monolingual corpora that have been categorised and sampled using the same method for many languages. A parallel corpus is a collection of linked, sentence-by-phrase translations of texts from one language into another. A reference corpus is a collection of texts that reflect different social and situational situations via the use of formal and informal written and spoken language.

The audio and text of recorded speech, together with the timestamps for each word, make up what is known as a speech corpus. The following are a few examples of the many datasets that are available for use in natural language processing:

- **Evaluating Public Opinion:** Among the many applications of natural language processing (NLP), sentiment analysis is increasingly used in various fields such as politics, business, etc. The following datasets are often used in sentiment analysis:

1. **Stanford Sentiment Treebank (SST):** Socher et al. introduced SST containing sentiment labels for 215,154 phrases in parse trees for 11,855 sentences from movie reviews posing novel sentiment compositional difficulties.
2. **Sentiment140:** It contains 1.6 million tweets annotated with negative, neutral and positive labels.

- **IMDB:** This dataset contains thousands of movie reviews in training and test datasets for natural language processing, text analytics, and sentiment analysis. This dataset was introduced by Mass et al. in 2011. Sentiraama was created by G. Rama Rohit Reddy of the Language Technologies Research Centre, KCIS, IIIT Hyderabad. The corpus has four datasets with two-value scales that discriminate document-level positive and negative sentiment. Book, product, movie, and song lyrics are in the corpus. The annotators painstakingly followed each annotation method. The corpus' "Song Lyrics" section comprises 339 Telugu-scripted song lyrics.
- **Language modelling:** Text data is analysed to compute word likelihood. They utilise an algorithm to read data and construct natural language context rules. The model then properly predicts or creates new phrases using these criteria. The model learns linguistic basics and applies them to new sentences.
- **Language modelling's main datasets:** a) Salesforce's WikiText-103 dataset contains 103 million tokens from 28,475 Wikipedia pages. b) WikiText-2 is a shrunken 103. It has 2 million 33,278-jargon tokens. c) Penn Treebank's Wall Street Diary corpus comprises 929,000 training, 73,000 validation, and 82,000 testing tokens. Its context is confined by sentences rather than paragraphs. d) The Ministry of Electronics and Information Technology's Technology Development Programme for Indian Languages (TDIL) created a catalogued dataset data distribution site (www. tdil-dc.in).
- **Machine Translation:** Machine translation converts natural language text into another language while preserving meaning. Mainly utilised datasets are: a) Totoaba has multilingual sentence pairs. Each dataset line contains a tab-delimited pair of English and translated French text sequences. Each text sequence might be a sentence or a paragraph. b) The Europarl parallel corpus comes from European Parliament proceedings. Available in 21 European languages. c) WMT14 offers English-German and English-French machine translations. These datasets include 4.5 million and 35 million sentences. Byte-Pair Encoding with 32 K tasks encodes sentences.

There are around 160,000 sentence pairs in IWSLT 14. The dataset includes English-German (En-De) and German-English (De-En) descriptions. About 200 K training phrase sets are in IWSLT 13. e) The IIT Bombay English-Hindi corpus includes parallel corpora, monolingual Hindi corpora from many sources, and corpora developed by the Centre for Indian Language Technology. 4. Question Answering System: Real-time question-answering systems are popular in customer service.

The dialogue/question answering system datasets are: a) Stanford Question Answering Dataset (SQuAD): crowd-sourced Wikipedia article reading comprehension questions. b) Natural Questions: Google's vast corpus trains and evaluates open-domain question-answering systems. For QA system training, it provides 300,000 genuine questions and human-annotated Wikipedia replies. c) Question Answering in Context (QuAC): This dataset describes, understands, and participates in information-obtaining conversations. This dataset contains an interactive discussion between two collective workers. That is, students ask open-ended questions about unfamiliar Wikipedia texts, and teachers respond with brief examples.

Difficulties in Healthcare Natural Language Processing

Terminology used in medicine. Not only does the healthcare industry have its own unique vocabulary, but medical professionals often include slang, acronyms, and shorter expressions in their notes and interviews.[113] When training an NLP system, it is essential to use a dataset that closely resembles the one it will use in production. This will make the task more difficult during bespoke creation and will be an issue when purchasing an off-the-shelf NLP solution. Limitations of electronic health records. Without proper fields in an EHR, an NLP system will be unable to organise data. There will likely be more effort required for your other apps as a result of new data-related adoption.

Advanced skill set. Data architecture, analytics skills, and domain expertise are all necessary for someone to spearhead your NLP adoption effort. To overcome these obstacles, you should engage an artificial intelligence (AI) specialist if your team does not include data scientists. Integrate APIs for text analysis. Google, IBM, Microsoft, and Amazon, four of the largest NLP providers globally, provide APIs at a high level that are accessible to developers without a background in machine learning. Each of them has healthcare-specific offerings or examples of how they've put these ideas into practice.

The fundamental difficulty in clinical natural language processing (NLP) is information overload, which makes it very difficult to retrieve a specific, important data point from enormous databases. As an added complication, summarisation systems plagued by usability problems and quality deficiencies have an uphill battle when data comes to semantic and context comprehension. The fact that natural language processing programs have to handle a broad diversity of text forms in order to answer questions from different sources is another major issue.

Health information security and confidentiality will be major problems when interactions between patients and doctors are documented in great detail in clinical notes. Patients discuss their eating habits, health issues, and any stigmatising problems throughout these conversations.[117] Protecting the confidentiality of individuals' health records is a top priority for the federal government, and HIPAA ensures that this will never happen in the US. Furthermore, guidelines for the scientific use of health data are set forth by the General Data Protection Regulation (GDPR) of the European Union. The need for obtaining people's informed permission and sanitising sensitive data categories is important in these regulatory measures, which have significant implications for natural language processing research. The General Data Protection Regulation (GDPR) lays forth general guidelines for how to handle personal information. Some of these guidelines include making sure the processing is fair, transparent, and legal (i.e., with permission), using the data for particular and legitimate reasons, and not keeping it for longer than necessary. Information cleansing is a part of data minimization. "Science makes extensive use of certain types of personal data." Only once an individual has made their information public or with their explicit agreement may it be processed. In a broad sense, "scientific usage" describes research and technical advancements that are both fundamental and applied, supported by private funding. Sanitisation methods are often seen as the minimal need for safeguarding persons' privacy while data is being collected.

Natural Language Processing (NLP) in the healthcare industry can comprehend these challenges via a range of practical examples. Let's examine a couple of them:

Enhancing Medical Records Health informatics Record configurations sometimes include a complex framework, making it challenging to store information inside them. Through discourse-to-content transcription, information may be automatically captured at the point of focus, freeing clinicians from the tedious task of documenting care delivery. Enhancing the efficiency of computer-assisted coding (CAC) may be achieved via many means using natural language processing. The CAC retrieves information on techniques for detecting patterns and enhancing assertions. This may really assist healthcare organisations in transitioning from a fee-for-service to a value-based approach, therefore significantly enhancing the patient experience.

Natural language processing in the field of health care may also identify and mitigate possible errors in the delivery of care to assess the quality of healthcare and monitor compliance with clinical guidelines.

Artificial intelligence and natural language processing technologies have the capabilities necessary to identify individuals with intricate health issues who have a history of mental health disorders or drug abuse and need enhanced care. Factors like malnutrition and housing instability might hinder the implementation of treatment protocols, hence compelling these patients to incur higher costs during their lifetime.

Health Information Technology in Healthcare

The healthcare sector is undergoing significant transformations propelled by rapid advancements in health information technology (HIT). Over recent decades, HIT has emerged as a key driver of innovation, reshaping the future of healthcare. This article delves into groundbreaking developments in HIT, exploring their potential implications for the industry.

From telemedicine and wearable devices to artificial intelligence and genomics, these innovations are revolutionising healthcare delivery, promising a future characterised by enhanced accessibility, personalised medicine, and data-driven decision-making. As we navigate this dynamic landscape, the intersection of technology and healthcare is poised to bring about unprecedented improvements, ensuring a more responsive, efficient, and patient-centric healthcare ecosystem.

Telemedicine and Virtual Care

The meteoric rise of telemedicine, accelerated by global challenges like the COVID-19 pandemic, has transformed healthcare delivery. Utilising advanced video conferencing tools, secure messaging platforms, and integrated Electronic Health Records (EHRs), telemedicine enables remote clinical care, fostering seamless doctor-patient interactions. The integration of the best electronic health records (EHRs) is a cornerstone of effective telemedicine practices. These top-tier EHR systems play a crucial role in facilitating remote clinical care by seamlessly integrating with telemedicine platforms. They ensure that healthcare providers have comprehensive access to patient data, enabling informed decisions during virtual consultations. The synergy between telemedicine tools and the best EHR systems allows for a more holistic approach to patient care, ensuring that crucial information is readily available to healthcare professionals, thereby enhancing the quality and efficiency of remote medical interactions.

This paradigm shift ensures the uninterrupted provision of healthcare services while significantly broadening access. Patients can now receive medical consultations from the comfort of their homes, transcending geographical barriers.

The integration of telemedicine into mainstream healthcare is a pivotal stride toward improved accessibility, patient-centric care, and addressing healthcare disparities among remote and underserved populations. This transformative approach leverages digital technologies to provide remote clinical services, breaking down geographical barriers and fostering more inclusive healthcare delivery.

Wearable Health Devices and IoT

The integration of wearable health devices into patient care heralds a new era of personalised monitoring and timely interventions. These devices, ranging from fitness trackers to glucose monitors, are interconnected through the Internet of Things (IoT), transmitting real-time health data to healthcare providers. This seamless flow of information enables practitioners to offer personalised care plans and intervene promptly.

The evolving landscape hints at even more sophisticated wearables on the horizon, such as smart contact lenses capable of detecting glucose levels and innovative patches designed for continuous monitoring of cardiac activity. The amalgamation of wearable health devices and IoT not only enhances patient care but also contributes to the ongoing revolution in healthcare technologies.

Augmented Reality (AR) and Virtual Reality (VR) in Medical Training

Augmented reality (AR) and virtual reality (VR) are revolutionising medical education by providing immersive training experiences. These technologies empower medical students to simulate surgical procedures in a virtual environment, offering a risk-free space to hone their skills.

Additionally, complex anatomical structures can be visualised in three dimensions, enhancing the understanding of intricate details. The immersive nature of AR and VR accelerates the learning curve for intricate medical procedures, ensuring that practitioners enter real-world scenarios with increased confidence and competence. As these technologies continue to evolve, they promise to redefine medical training, contributing to a more proficient and skilled healthcare workforce.

Blockchain in Healthcare

Blockchain technology has emerged as a game-changer in healthcare, offering a decentralised and secure framework for managing health records. By ensuring data integrity and patient privacy, blockchain facilitates seamless and transparent data sharing among various healthcare entities.

The potential applications extend to revolutionising Electronic Health Records (EHRs), streamlining billing processes, and enhancing drug traceability in the pharmaceutical supply chain.

The immutable and tamper-proof nature of blockchain adds a layer of trust to healthcare data, addressing longstanding challenges related to interoperability, security, and integrity. As this technology continues to mature, it holds the promise of transforming multiple facets of healthcare management.

Artificial Intelligence and Machine Learning

Artificial intelligence (AI) and machine learning (ML) are revolutionising healthcare information technology by offering a wide array of applications. Beyond diagnostics, AI plays a pivotal role in predicting patient readmissions and optimising hospital operations. Machine learning models, fuelled by vast datasets, unveil valuable insights that contribute to more informed clinical decision-making.

These technologies enhance the efficiency of healthcare systems, streamline processes, and ultimately improve patient outcomes. As AI and ML continue to evolve, their integration into health information technology holds the potential to further transform the landscape of healthcare delivery and administration.

Genomic Data Integration

The decreasing costs of genome sequencing are ushering in an era of personalised medicine, leveraging an individual's genetic information for tailored healthcare. Genomic data integration into Electronic Health Records (EHRs) presents a transformative opportunity. Clinicians can gain valuable insights into potential genetic disorders, identify optimal treatment strategies, and anticipate predispositions to specific diseases.

This convergence of genomics and health information technology holds the promise of more precise diagnostics, targeted therapies, and a deeper understanding of individual health risks. As this integration advances, it is poised to play a central role in shaping the future of healthcare, offering a level of personalisation that was once considered futuristic.

3D Printing in Healthcare

The advent of 3D printing in healthcare marks a transformative leap in personalised medical solutions.

This technology is reshaping prosthetics, implants, and even the potential for organ transplants. 3D printing allows for the creation of customised prosthetic limbs and dental implants, precisely tailored to individual specifications, ensuring optimal fit and functionality.

Moreover, ongoing research in bioprinting explores the possibility of using living cells to construct organs for transplantation. This groundbreaking application holds the potential to address organ shortages and revolutionise the field of transplantation, offering a glimpse into a future where medical interventions are intricately tailored to the unique needs of each patient.

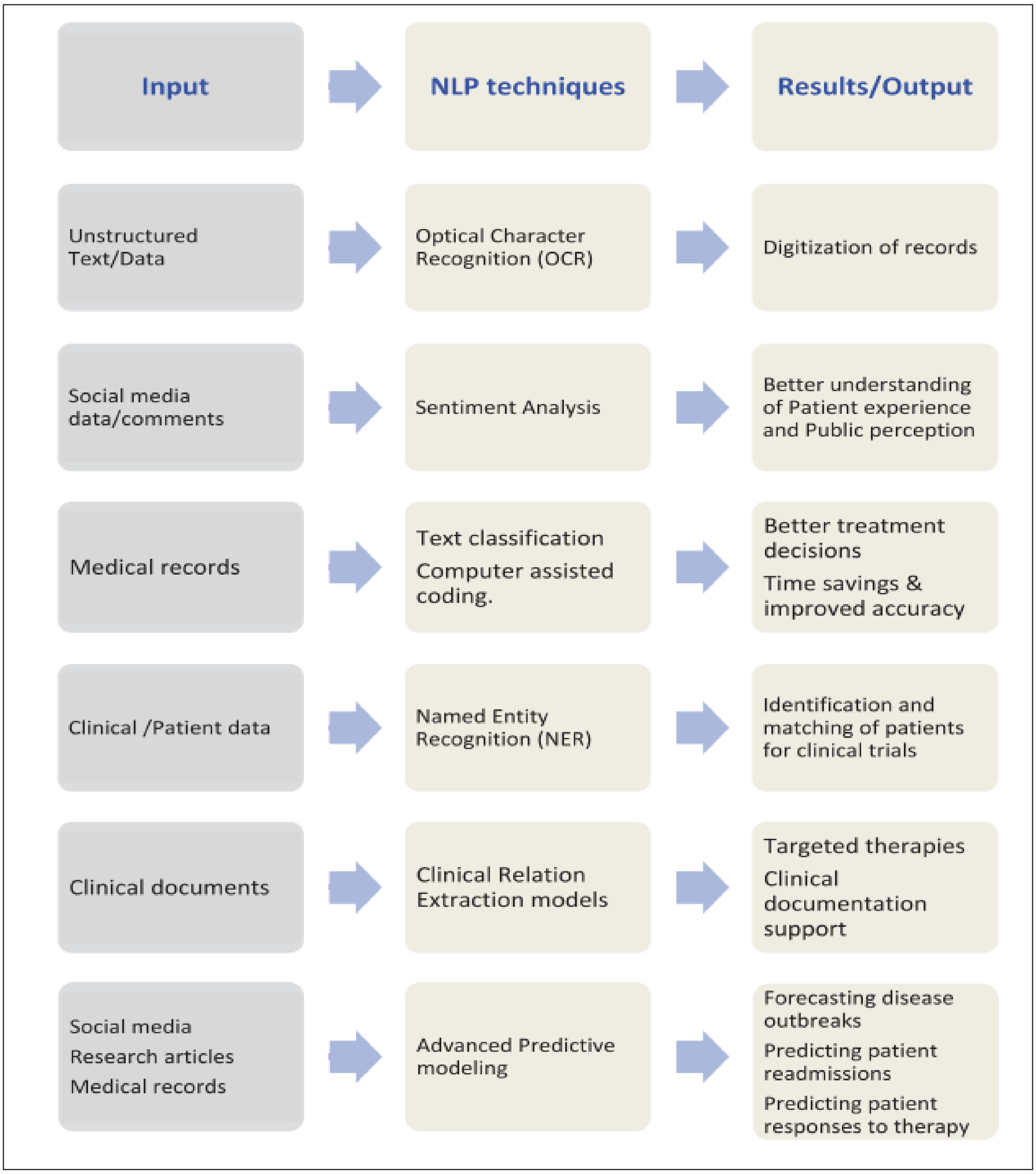

Figure 7.Potential advances enabled by application of NLP techniques and lists some of the tools used

Current and Future progress of NLP

Current research in NLP focuses on various aspects, such as syntactic phenomena which deal with sentence structure and word order based on grammatical classes rather than meaning. This includes the development of discriminative models for scoring passes, efficient approximate parsing techniques, and dependency grammar. Another area of research is machine translation, which involves developing models and algorithms for translating between languages, particularly for low-resource and morphologically complex languages. Semantic phenomena refer to aspects of language that are relevant to the meaning of a sentence, regardless of the surrounding context.

Examples include sentiment analysis, summarization, information extraction, slot filling, discourse analysis, and textual entailment. On the other hand, pragmatic phenomena, such as speech, involve the interpretation of a sentence in relation to its specific context.

The context can be either linguistic, such as the previous text or dialogue, or non-linguistic, such as knowledge about the person who produced the language, the goals of the communication, or the objects in the current visual field. Examples of non-linguistic context include language modelling-syntax and semantics, as well as models of acoustics and pronunciation. The references are indicated by.

Commercialization of voice recognition and information retrieval has now occurred, resulting in a vast amount of text and speech available on the Internet, mobile phones, and other platforms. It is evident that language studies include several aspects, such as formalising insights into discrete knowledge (what is feasible) and continuous knowledge (what is likely).

These studies involve mathematical formalism, algorithm development, implementation, and testing using actual data. The necessary enhancements for NLP include augmenting the functionality of existing interfaces and completely implementing back-end processing, such as information extraction and normalisation, to construct databases. Another expected enhancement is the use of portable devices equipped with translators and personal conversation recorders that can do subject searches.

- **Advanced Predictive Modelling:** Enhancing predictive analytics to not only forecast disease outbreaks and/or patient readmissions but also to predict individual patient responses to specific treatments based on historical data analysis.
- **Enhanced Clinical Decision Support:** Integrating more sophisticated NLP algorithms into clinical decision support systems to provide real-time, evidence-based recommendations personalized to each patient's unique health status and history.
- **Automated Real-time Documentation:** Further automating clinical documentation and note-taking during patient encounters, thereby reducing administrative burden on healthcare providers and allowing them to focus more on patient care.
- **Improved Patient Engagement and Self-Management:** Developing more advanced and interactive chatbots and virtual health assistants that can provide personalized advice, monitor patient health more accurately and support chronic disease management.
- **Semantic Health Records:** Moving beyond structured data to create fully semantic, interoperable health records that can be more easily shared and understood across different healthcare systems and platforms, enhancing care coordination and patient outcomes.
- **Drug Discovery and Development:** Utilizing NLP to analyse scientific literature, clinical trial notes and other research documents to identify potential drug candidates and therapeutic targets quicker than traditional methods.
- **Cross-Language Information Exchange:** Improving translation capabilities to facilitate better information sharing and collaboration among healthcare professionals across different languages and regions, thus advancing global health initiatives.
- **Mental Health Monitoring and Support:** Expanding the use of sentiment analysis and emotion recognition to monitor social media and other online communications for signs of mental health issues, enabling early intervention and support.
- **Customized Patient Education:** Creating tailored educational materials for patients based on their specific conditions, treatment plans and preferences, improving understanding and adherence to treatment recommendations.
- **Ethical and Privacy Considerations:** Developing NLP tools to help navigate and ensure compliance with ethical considerations and privacy regulations in healthcare documentation and data analysis
- **Social Determinants of Health (SDOH):** Analysing unstructured data from various sources to gain insights into the social determinants of health affecting patient populations, thereby informing more holistic and effective intervention strategies.
- **Enhanced Supply Chain Predictive Analytics:** Future NLP systems will offer more sophisticated predictive analytics for supply chain management, predicting disruptions and automatically suggesting alternative suppliers and/or solutions before issues arise.
- **Intelligent Negotiation Assistants:** AI-driven bots could simulate negotiation strategies with vendors, using historical data and real-time market trends to optimize contracts and agreements in favour of healthcare organizations.

- **Automated Regulatory and Compliance Monitoring:** NLP could provide real-time monitoring of communications and transactions for compliance with an ever-evolving landscape of healthcare regulations, instantly flagging potential issues for review.
- **Advanced Sentiment Analysis:** Beyond analysing feedback, future NLP could assess sentiment in real-time during supplier communications, helping to navigate negotiations and maintain relationships by alerting when discussions may need a human touch and/or strategic shift.
- **Deep Personalization in Communication:** Utilizing deep learning, NLP could offer hyper-personalized communication strategies tailored not just to the vendor profile but also to the individual representative's preferences and behaviors, enhancing relationship management.
- **Real-Time Language Translation and Understanding:** Future NLP tools will likely provide more accurate and context-aware translations in real-time, facilitating smoother international transactions and collaborations without language barriers.
- **Automated Dispute Resolution:** By analysing communication patterns, contract terms and historical data, NLP systems could propose solutions to disputes and/or discrepancies between healthcare facilities and their suppliers and/or vendors, minimizing the need for lengthy negotiations and/or legal interventions.
- **Intelligent Procurement Assistance:** NLP could evolve to automatically draft procurement documents, requests for proposals (RFPs) and other critical documents based on predefined templates and specific requirements gathered through natural language inputs.
- **Enhanced Security and Fraud Detection:** Future applications of NLP might include sophisticated monitoring of communications and transactions to detect anomalies, potential fraud and/or cybersecurity threats, safeguarding sensitive data and financial transactions.
- **Dynamic Market and Product Research:** By continuously analyzing global data sources, NLP could offer real-time insights into emerging healthcare trends, innovative products and shifts in the supplier landscape, ensuring that healthcare organizations stay ahead of the curve.

Conclusion

Health Natural language processing (NLP) is becoming more popular due to the vital function it plays in a broad variety of technique development and applications. Recent research has mostly focused on using natural language processing (NLP) systems in healthcare, as well as methods for clinical decision support and informatics, making and using knowledge graphs, and extracting medical information (including entity, relation, temporal, and interaction extraction). Real-time NLP analysis of clinical records enables the delivery of care recommendations for clinical practice. There is a lot of hope for finding solutions to systemic problems, including mental health difficulties, by using natural language processing tools. The current research shows a number of proof-of-concept applications that show how important it is for clinical scientists to operationalize therapy and for computer scientists to come up with methods that can capture how treatments change depending on the situation and happen in a certain order. When studying patients, providers, and their interactions, the NLPxMHI framework aims to include important research design and clinical category concerns. We will establish large secure datasets, a shared language, and checks for justice and fairness to facilitate cooperation between computer scientists and doctors. Maintaining momentum in applying NLP to mental health interventions—which has the potential to radically alter how we diagnose and treat mental health issues—requires bridging these fields. Big data computing can assist in building a solid, fault-tolerant natural language processing (NLP) infrastructure.

Conflict of Interests: None

Source of Funding: None

Author Contributions: KB, SA, RT: writing-original draft preparation; DJ: writing-review and editing; SJ, RP, PP: Data curation, formal analysis; DK: Conceptualization, investigation, supervision, project administration. All authors have read and agreed to the published version of the manuscript.

Acknowlegement: None

Reference

1. Fanni SC, Febi M, Aghakhanyan G, Neri E. Natural language processing. InIntroduction to Artificial Intelligence 2023 Sep 16 (pp. 87-99). Cham: Springer International Publishing.
2. Chowdhary K, Chowdhary KR. Natural language processing. Fundamentals of artificial intelligence. 2020:603-49.
3. Reshamwala A, Mishra D, Pawar P. Review on natural language processing. IRACST Engineering Science and Technology: An International Journal (ESTIJ). 2013;3(1):113-6.
4. Meera S, Geerthik S. Natural language processing. Artificial intelligent techniques for wireless communication and networking. 2022 Mar 7:139-53.
5. Gupta V. A survey of natural language processing techniques. International Journal of Computer Science & Engineering Technology. 2014 Jan;5(1):14-6.
6. Mehta N, Devarakonda MV. Machine learning, natural language programming, and electronic health records: The next step in the artificial intelligence journey.

Journal of Allergy and Clinical Immunology. 2018 Jun 1;141(6):2019-21.
7. Eisenstein J. Introduction to natural language processing. MIT press; 2019 Oct 1.
8. Bahja M. Natural language processing applications in business. E-Business-higher education and intelligence applications. 2020 May 11.
9. Goyal P, Pandey S, Jain K. Deep learning for natural language processing. New York: Apress. 2018.
10. Kocaleva M, Stojanov D, Stojanovic I, Zdravev Z. Pattern recognition and natural language processing: State of the art. Tem Journal. 2016;5(2):236-40.
11. Locke, S., Bashall, A., Al-Adely, S., Moore, J., Wilson, A., & Kitchen, G. B. (2021). Natural language processing in medicine: a review. Trends in Anaesthesia and Critical Care, 38, 4-9.
12. Wang, J., Deng, H., Liu, B., Hu, A., Liang, J., Fan, L., ... & Lei, J. (2020). Systematic evaluation of research progress on natural language processing in medicine over the past 20 years: bibliometric study on PubMed. Journal of medical Internet research, 22(1), e16816.
13. Ni, L., Lu, C., Liu, N., & Liu, J. (2017). Mandy: Towards a smart primary care chatbot application. In International symposium on knowledge and systems sciences (pp. 38-52). Singapore: Springer Singapore.
14. Chen, X., Xie, H., Wang, F. L., Liu, Z., Xu, J., & Hao, T. (2018). A bibliometric analysis of natural language processing in medical research. BMC medical informatics and decision making, 18, 1-14.
15. Amaral, M. B., Roberts, A., & Rector, A. L. (2000). NLP techniques associated with the OpenGALEN ontology for semi-automatic textual extraction of medical knowledge: abstracting and mapping equivalent linguistic and logical constructs. In Proceedings of the AMIA Symposium (p. 76). American Medical Informatics Association.
16. Liu, K., Hogan, W. R., & Crowley, R. S. (2011). Natural language processing methods and systems for biomedical ontology learning. Journal of biomedical informatics, 44(1), 163-179.
17. Viani, N., Larizza, C., Tibollo, V., Napolitano, C., Priori, S. G., Bellazzi, R., & Sacchi, L. (2018). Information extraction from Italian medical reports: An ontology-driven approach. International journal of medical informatics, 111, 140-148.
18. Yim, W. W., Yetisgen, M., Harris, W. P., & Kwan, S. W. (2016). Natural language processing in oncology: a review. JAMA oncology, 2(6), 797-804.
19. Johri P, Khatri SK, Al-Taani AT, Sabharwal M, Suvanov S, Kumar A. Natural language processing: History, evolution, application, and future work. InProceedings of 3rd International Conference on Computing Informatics and Networks: ICCIN 2020 2021 (pp. 365-375). Springer Singapore.
20. Jones KS. Natural language processing: a historical review. Current issues in computational linguistics: in honour of Don Walker. 1994 Jun 30:3-16.
21. Khurana D, Koli A, Khatter K, Singh S. Natural language processing: state of the art, current trends and challenges. Multimedia tools and applications. 2023 Jan;82(3):3713-44.
22. Cambria E, White B. Jumping NLP curves: A review of natural language processing research. IEEE Computational intelligence magazine. 2014 Apr 10;9(2):48-57.
23. Ebadi A, Xi P, Tremblay S, Spencer B, Pall R, Wong A. Understanding the temporal evolution of COVID-19 research through machine learning and natural language processing. Scientometrics. 2021 Jan; 126:725-39.
24. De Castilho, R. E., & Gurevych, I. (2014, August). A broad-coverage collection of portable NLP components for building shareable analysis pipelines. In Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT (pp. 1-11).
25. Bhardwaj, A., Khanna, P., & Kumar, S. (2020). Generative model for nlp applications based on component extraction. Procedia Computer Science, 167, 918-931.
26. Iosifova, O., Iosifov, I., & Rolik, O. (2020). Techniques and components for natural language processing. Адаптивні системи автоматичного управління, 1(36), 93-113.
27. Ramganesh, E., & Paulraj, I. J. M. (2015). Prospective teacher educators' awareness towards the key components of NLP. International Journal of Academic Research in Management (IJARM), 4(1), 24-34.
28. Resnik, P., & Lin, J. (2010). Evaluation of NLP systems. The handbook of computational linguistics and natural language processing, 271-295.
29. Lass, R. (1984). Phonology: An Introduction to Basic Concepts. Cambridge University Press.
30. Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. Multimedia tools and applications, 82(3), 3713-3744.