

1

Foundations of Natural Language Processing

Imagine engaging in a conversation where you express enjoyment over a song, and a virtual assistant, like Alexa, responds by acknowledging your preference and adjusting its algorithms accordingly. This is NLP in action—bridging the gap between human communication and computer understanding, allowing for a seamless interaction between the two. Welcome to the fascinating world of Natural Language Processing (NLP)!

NLP is a branch of artificial intelligence that equips computers with the capability to understand, interpret, and respond to human language in a way that is both meaningful and valuable. It encompasses a broad range of tasks and techniques aimed at processing and analyzing both text and speech, serving as a pivotal connection between human natural language and computer data processing. The field of NLP is notably interdisciplinary, drawing from areas such as artificial intelligence, machine learning, linguistics, and social sciences. This amalgamation of fields has led to the development of sophisticated methods that allow machines to process human language in complex ways, ranging from structural and contextual analysis to emotional tone assessment. Every sentence we utter or write carries with it structural, contextual, and emotional cues. While humans can instinctively navigate these cues, for machines, each sentence is a piece of unstructured data that must be meticulously converted into a structured format. Through NLP, what seems like a straightforward sentence like “Alexa, I like this song,” is transformed into structured data that machines can comprehend and respond to, leading to actions like playlist modifications and preference learning.

NLP represents a significant stride in the domain of artificial intelligence, where the primary aim is to bridge the gap between human communication and machine understanding. This field involves a series of intricate processes that enable machines to comprehend, interpret, and produce human language in a meaningful way. The core operations of NLP—recognition, understanding, and generation—constitute the backbone of this fascinating AI subdomain.

- **Recognition** is the initial phase where machines detect and decipher human language into structured, machine-readable data. This process is crucial for transforming the inherent ambiguity of natural language into a clear-cut format that computers can process. Recognition involves various sub-tasks like tokenization, where text is broken down into words, phrases, or other meaningful elements, and parsing, which involves analyzing the grammatical structure of a sentence. Speech recognition, another facet of this phase, converts spoken language into a digital format. Recognition is heavily dependent on machine learning algorithms that are trained to identify the language's structure and semantics. These algorithms are adept at handling the complexities of language, including syntax, grammar, and even the idiosyncrasies of regional dialects and accents.
- In the **understanding** phase, NLP systems analyze the structured data obtained from recognition to deduce meanings and relationships. This stage involves more advanced NLP tasks such as semantic analysis, which interprets the meaning of words in context, while pragmatics focuses on understanding language. For instance, sentiment analysis discerns the emotional tone behind text data, whether it's positive, negative, or neutral, providing insights into the writer's or speaker's feelings and attitudes. Named entity recognition identifies and classifies key elements in text into predefined categories like names of people, organizations, locations, dates, and more. At this stage, the complexity of human language becomes apparent as the system must understand nuances, humor, sarcasm, and cultural references, which are often challenging for machines.
- The **generation** phase is where NLP systems produce human-like responses from the analyzed data. In this stage, NLP uses Natural Language Generation (NLG) techniques to create coherent, contextually relevant sentences that can be understood by humans. Generation involves converting the structured data back into natural language. This process can be seen in applications like chatbots, virtual assistants, and automated report generators, where the machine communicates with users in a seemingly intuitive and understanding manner. The ability to generate language also encompasses creative aspects of language use, such as composing poetry, writing articles, or generating dialogue in conversational agents.

Despite the technical prowess of NLP, the field still faces significant challenges. One of the primary limitations is the system's ability to fully grasp the intricacies and subtleties of human language. While NLP systems can recognize patterns and analyze text or speech to a certain extent, they often lack the depth of understanding inherent in human communication. The complexity of language, including cultural, contextual, and idiomatic layers, presents a substantial hurdle

for AI. Moreover, NLP relies heavily on data-driven, predictive mathematical models that, while being powerful, can sometimes miss the mark in interpreting the ambiguity and fluidity of natural language. Furthermore, most NLP models are trained on vast amounts of text data, which may contain biases and inaccuracies, reflecting the model's language understanding and generation capabilities. These biases can lead to skewed or unfair outcomes, especially in sensitive applications like sentiment analysis, hiring, and law enforcement. Addressing these limitations requires continuous refinement of NLP models, incorporating a broader, more diverse range of data, and developing more sophisticated algorithms that can navigate the nuanced landscape of human language.

1.1 History of NLP

The historical trajectory of NLP is a testament to the evolving interplay between technology and linguistics, an odyssey that commences in the mid-twentieth century and extends into the present era of advanced computing. The foundational stone of NLP was laid in the 1950s, marked by Alan Turing's introduction of the Turing Test, a seminal concept that questioned a machine's ability to exhibit intelligent behavior akin to a human. This period also saw the inception of computational linguistics, where the groundwork for machine understanding of language was established.

As we moved into the 1960s, NLP began to manifest in more concrete forms with systems like ELIZA, created by Joseph Weizenbaum, which demonstrated the superficiality yet potential of human-machine communication. The subsequent development of SHRDLU in 1969 further showcased the ability of computers to comprehend simple English sentences, providing a glimpse into the practical applications of NLP. The 1970s brought a more structured approach to NLP with a focus on rule-based methods, exemplified by the development of conceptual dependency theory. This theory laid the foundation for natural language understanding (NLU) by presenting a model for representing language's conceptual structure. The evolution continued into the 1980s as the field began to pivot from rule-based to statistical methods, signalling a significant paradigm shift. This era also witnessed the creation of Racter, the first chatbot, marking a notable milestone in NLP's history.

The 1990s saw an infusion of machine learning into NLP, catalyzing a transformative phase where algorithms began to learn language patterns, thereby increasing the scope and accuracy of linguistic processing. The culmination of this period was epitomized by IBM's Deep Blue, whose victory in chess underscored AI's potential. The 2000s heralded an age of advanced algorithms and large-scale data, with Google Translate emerging as a quintessential tool,

democratizing machine translation for global users. The subsequent decade, the 2010s, was defined by the ascendancy of deep learning and neural networks in NLP. Innovations like sequence-to-sequence models and attention mechanisms significantly improved machine translation, while the introduction of Google's BERT model revolutionized context understanding in NLP.

The present era, the 2020s, continues to witness NLP innovation, with advancements in language generation models like GPT-3 and a heightened focus on addressing bias and ethical considerations in NLP models. The integration of NLP into various industries exemplifies its pervasive impact and the growing recognition of its value in extracting and generating meaningful information from language data. The recent prominence of Large Language Models (LLMs) in NLP has led to a perception that they might render traditional methods obsolete. However, this view overlooks the nuanced and multifaceted nature of NLP. While LLMs have indeed revolutionized the field, offering unparalleled capabilities in generating coherent and contextually relevant text, foundational NLP techniques remain essential for a comprehensive understanding of language processing. Basic foundational NLP remains crucial even in the era of LLMs for several reasons:

- *Understanding Fundamentals:* Foundational NLP provides the groundwork for understanding how language operates at a mechanical level. It encompasses essential processes such as tokenization, parsing, part-of-speech tagging, and syntax analysis. Grasping these basics is vital for interpreting the outputs of more advanced systems like LLMs and for fine-tuning them to specific tasks or languages.
- *Building Blocks for Advanced Models:* The techniques and knowledge derived from foundational NLP serve as building blocks for more sophisticated models, including LLMs. These advanced models often rely on the principles and data processing methods established in basic NLP to function effectively.
- *Customization and Optimization:* Understanding foundational NLP enables researchers and practitioners to customize and optimize LLMs for specific applications. By knowing how language is processed at a fundamental level, one can better tailor these models to meet unique linguistic and contextual requirements.
- *Efficiency and Resource Management:* Foundational NLP methods are often more efficient and require fewer computational resources than LLMs, making them suitable for applications where rapid processing or low resource consumption is essential.
- *Interpretability and Debugging:* A strong grasp of foundational NLP can aid in interpreting the behavior of LLMs and in debugging them when unexpected

results occur. It helps in tracing issues back to their roots in the underlying linguistic processes.

- *Complementarity*: In many cases, foundational NLP methods are used in conjunction with LLMs to enhance performance. For instance, basic NLP techniques can preprocess data to improve the efficiency and accuracy of LLMs or post-process their outputs to refine the results.
- *Educational Value*: For newcomers to NLP, starting with the basics is essential for building a comprehensive understanding of the field. Foundational knowledge equips individuals with the insights needed to effectively engage with, contribute to, and innovate in the domain of NLP.
- *Handling Low-Resource Languages*: Foundational NLP techniques are particularly important for working with low-resource languages, where LLMs may not perform well due to a lack of training data. Basic NLP can provide the necessary tools to process these languages and develop customized solutions.

Throughout its history, NLP has been shaped by the tension and interplay between linguistic theory and computational power, reflecting a journey from simplistic rule-based approaches to sophisticated deep learning models. Each phase of NLP's history not only reflects technological advancements but also a deeper understanding of language's complexity, with ongoing innovations ensuring that the field remains at the cutting edge of artificial intelligence research and application. While LLMs represent a significant leap forward, they do not make foundational NLP techniques obsolete. Instead, these two facets of NLP complement each other, with traditional methods providing the groundwork upon which LLMs build to achieve their impressive feats. Understanding and applying both foundational NLP and LLMs is key to harnessing the full potential of language technology.

1.2 Approaches to NLP

NLP encapsulates a spectrum of methodologies each tailored to parse, interpret, and manifest human language in a form comprehensible to machines. These methodologies span from traditional rule-based systems to advanced machine learning and deep learning paradigms, each contributing uniquely to the evolution of NLP.

- **Rule-Based NLP**, a foundational approach, operates on predefined linguistic rules and grammatical structures. This method, reliant on expert knowledge in linguistics, crafts a deterministic pathway for machines to understand and process language. This methodology, deeply rooted in the traditional study of linguistics, employs a deterministic strategy, where the understanding

and processing of language are driven by explicit, handcrafted rules. These rules encompass the gamut of grammatical and syntactic norms that define a language, offering a structured framework for NLP systems to operate within.

At the core of rule-based NLP are comprehensive linguistic resources such as dictionaries, thesauruses, and specialized lexical databases. One notable example is WordNet, a rich lexical database that categorizes English words into sets of cognitive synonyms, known as synsets. Each synset represents a distinct semantic concept, linking words through various relational types, including synonyms (words with similar meanings), hyponyms (words denoting a subclass of a concept), and meronyms (words denoting parts of a whole). This structured semantic network allows rule-based NLP systems to navigate the complex interrelationships of language, enhancing their ability to discern meaning and context.

The application of rule-based NLP is particularly advantageous in tasks where linguistic precision and clear-cut rules are paramount. Information retrieval systems benefit from this approach, as it enables precise querying and filtering of data based on syntactic and semantic criteria. Text summarization, another key application, leverages rule-based methods to distill essential content from larger texts, guided by syntactic structures and meaning encapsulated in the linguistic rules. Basic language understanding, such as parsing sentences and identifying grammatical elements, is also well-suited to rule-based NLP, given its structured and rule-governed nature. Despite its strengths in handling well-defined linguistic tasks, rule-based NLP has limitations. Its deterministic nature means it can be rigid, struggling to adapt to the variability and idiosyncrasies of natural language use in real-world contexts. The creation and maintenance of rule sets require extensive expert knowledge and can be labor-intensive, particularly for languages with complex grammar or those lacking comprehensive linguistic resources. Moreover, rule-based systems might falter in the face of slang, idiomatic expressions, or emerging language use not encapsulated in existing rule sets, highlighting a gap between the structured world of rule-based NLP and the dynamic, evolving nature of human language.

Hence, rule-based NLP offers a deterministic and transparent method for language processing, excelling in specific domains or tasks where the linguistic environment is well-understood and consistent. However, its limitations in handling the dynamic and varied nature of natural language necessitate complementary approaches, especially in broad or informal linguistic contexts.

- Transitioning from heuristics to statistical inference, **Machine Learning-Based NLP** signifies a leap in how machines understand language. Instead of rigid rules, this approach leverages algorithms to learn from data, discovering patterns and relationships in text. Machine learning in NLP can be supervised where models learn from labeled data, or unsupervised, deducing inherent

structures within the text. Techniques like support vector machines, decision trees, and neural networks have been instrumental in propelling NLP tasks, including text classification, sentiment analysis, machine translation, and speech recognition. The advent of these statistical models marked a significant advancement in NLP, enhancing both the scalability and accuracy of language processing tasks.

Machine learning in NLP encompasses both supervised and unsupervised learning approaches. Supervised learning methods rely on annotated datasets where the input data is labeled with the correct output, enabling models to learn and make predictions based on this labelled data. Unsupervised learning, on the other hand, digs into raw, unlabelled data to uncover inherent structures and meanings, thus deducing the underlying patterns without explicit guidance. Techniques such as support vector machines, decision trees, and increasingly, neural networks, have become the workhorses of machine learning in NLP, each contributing to various aspects of language processing. These machine learning techniques have been instrumental in advancing core NLP tasks. For instance, text classification has been revolutionized by machine learning algorithms, which can sift through massive volumes of text to categorize content into predefined classes. Sentiment analysis, too, has benefitted from this shift, with algorithms now capable of discerning nuanced emotional tones from text data. Machine translation has seen leaps in quality and efficiency, moving from literal, often stilted translations to more fluid and context-aware renditions of text across languages. Speech recognition, once constrained by the limitations of rule-based systems, has achieved remarkable accuracy under the machine learning regime, enabling real-time, naturalistic voice interactions.

The adoption of machine learning in NLP has not only enhanced the accuracy and effectiveness of language processing tasks but has also significantly expanded the scope and scalability of these tasks. Unlike rule-based systems, which are limited by the confines of their predefined rules, machine learning-based NLP systems thrive on large datasets, continually improving and adapting their performance as they consume more data. This transformative approach to language processing, emphasizing statistical patterns and machine learning, is discussed in depth in Chapter 5 of the book, where we explore the computational and cognitive mechanisms that underpin machine learning in NLP, illustrating how this technology underlies the modern advancements in the field.

- **Deep Learning-Based Methods in NLP** represent the pinnacle of methodological advancement in the field, harnessing sophisticated neural network architectures to capture and interpret the complexities of human language. Technologies such as Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and more recently, transformers like BERT

(Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-training Transformer), embody this cutting-edge approach. Deep learning shines in its capacity to learn layered representations of text, adeptly grasping both the immediate and broader contextual cues within language.

The transformative impact of deep learning in NLP is profound, especially in areas like NLU, text generation, and machine translation. These neural network models excel in deciphering the intricate patterns and nuances embedded in language, offering a rich, contextual understanding that enables machines to process and generate text with a level of sophistication akin to human-like comprehension. Unlike earlier methods, deep learning models can inherently capture and utilize the contextual relevance of words and phrases, thus facilitating a nuanced interpretation of language's subtleties. Furthermore, the advent of transformer models has significantly advanced NLP's capabilities. These models, particularly known for their self-attention mechanisms, allow for a more dynamic representation of textual data, enabling the processing of sequences with an efficient understanding of dependencies and relationships. LLMs like GPT have further extended these capabilities, generating text that is not only coherent but also contextually relevant and often indistinguishable from human-written content.

In summary, deep learning-based methods have revolutionized the way machines understand and interact with human language. By learning complex, hierarchical representations of text, these methods have enabled NLP systems to perform a wide array of tasks with unprecedented accuracy and fluency, offering deeper insights into the fabric of language itself. This significant stride in NLP is further elaborated in Chapter 5, focusing on the deep learning paradigms that underpin these advancements.

The journey through the approaches to NLP—from rule-based to deep learning—encapsulates the evolving landscape of language processing. Each approach brings a unique set of strengths and applications, collectively enriching the field of NLP. Rule-based methods, with their reliance on explicit linguistic knowledge, provide a clear, albeit limited, understanding of language structures. Machine learning broadens this horizon, offering scalable and adaptable solutions, while deep learning pushes the boundaries, introducing a level of analytical depth and fluidity that closely mimics human linguistic capabilities. Together, these methodologies define the current state of NLP, a dynamic and rapidly advancing field at the intersection of linguistics, computer science, and artificial intelligence.

1.3 Understanding NLP through NLU and NLG: Examples and Case Studies

NLP encompasses two critical components: Natural Language Understanding (NLU) and Natural Language Generation (NLG). These elements are essential for enabling machines to interpret and produce human language in a context that is meaningful and useful.

- *Natural Language Understanding (NLU)*: Involves the machine's ability to understand and interpret human language. NLU is about extracting meaning from text or speech, encompassing tasks like sentiment analysis, entity recognition, and language translation. For example, in sentiment analysis, NLU algorithms examine customer reviews to determine whether the sentiment is positive, negative, or neutral. In the case of entity recognition, NLU is used to identify and categorize key information in text, such as person names, organizations, locations, and dates. A practical case study of NLU is its use in virtual assistants like Siri or Alexa, which understand user commands and queries to perform actions like setting reminders, playing music, or providing weather information.
- *Natural Language Generation (NLG)*: On the other hand, is the process where machines generate text or speech from data. This involves converting structured information into human-readable text, ensuring the output is coherent, contextually relevant, and semantically rich. NLG is commonly seen in report generation, where systems convert data into narrative summaries or detailed reports. For example, automated journalism where NLG tools generate news articles based on data inputs like sports statistics or financial reports. Another case study of NLG is in the healthcare sector, where patient data can be transformed into readable clinical reports or personalized patient care instructions.

1.3.1 Practical Case Studies

NLP, through the dual processes of NLU and NLG, offers profound capabilities in interpreting and producing human language. This technology not only enhances interaction between humans and machines but also paves the way for innovative applications across various domains, streamlining processes, and facilitating efficient communication.

- *Customer Service Chatbots*: Incorporating both NLU and NLG, chatbots can understand customer queries (NLU) and generate appropriate responses (NLG). For instance, a chatbot for a bank might understand a customer's request about account balance (NLU) and generate a response detailing the account balance, recent transactions, and other related information (NLG).

- *Automated Content Creation:* News organizations use NLP to automatically generate news articles. For example, The Washington Post’s Heliograf has been creating short reports and alerts on topics like the Olympics and elections, where it understands the key facts (NLU) and constructs narrative content (NLG).
- *Language Translation Services:* Translation tools like Google Translate apply NLU to understand the text’s meaning in the source language and then use NLG to produce an accurate translation in the target language. This involves a deep understanding of grammatical, syntactic, and semantic nuances across languages.
- *Personal Assistants and Smart Devices:* Virtual assistants like Google Assistant use NLU to comprehend user requests, such as setting up meetings or searching for information online, and NLG to provide responses that are natural and contextually appropriate.

We aim to unpack the intricacies of NLU within this book, highlighting its indispensable role in crafting intuitive and intelligent language processing systems.

1.4 NLP Pipeline

The NLP pipeline, a cornerstone of modern NLP practices, is a systematic journey of transforming raw text into structured, actionable data. This intricate process, fundamental to extracting insights from unstructured text, involves several distinct yet interconnected stages, each contributing uniquely to the overall understanding and generation of natural language. (Fig.1.1) illustrates the NLP pipeline.

- *Data Acquisition and Cleaning:* The journey begins with Data Acquisition, where diverse textual data are sourced from various digital footprints. This

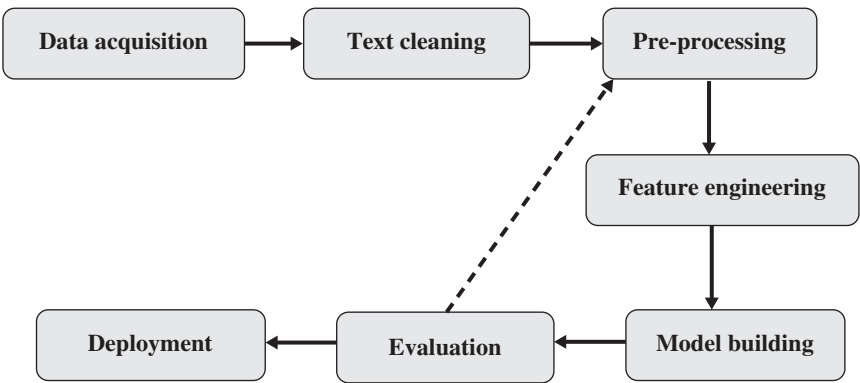


Figure 1.1 NLP Pipeline.

stage addresses the challenge of harnessing a vast array of text data from web pages, social media, official documents, and more, ensuring a rich dataset for analysis. Sophisticated tools like web scrapers and APIs are employed to aggregate this data, necessitating a keen understanding of data relevance and context to align with the end objectives of the NLP tasks at hand.

Following acquisition, Text Cleaning is paramount to refine this raw data. The focus here is on eliminating noise and inconsistencies: removing extraneous HTML tags, correcting spelling errors, and addressing syntactical anomalies. Techniques such as Unicode normalization and regular expression (RegEx) matching are crucial for stripping unwanted characters and formatting the text into a cleaner, more uniform structure. This process is not merely cosmetic but essential, as clean data significantly influences the performance and accuracy of subsequent NLP stages.

- *Preprocessing and Linguistic Analysis:* Text Preprocessing takes the cleaned data through vital steps like sentence segmentation and tokenization, breaking down large text blocks into manageable units of sentences and words, respectively. This breakdown is critical for machines to analyze text at a granular level, facilitating a more detailed and nuanced linguistic understanding.

In the realm of Linguistic Analysis, the pipeline searches deeper into the grammatical and semantic layers of the text. Here, the focus is on morphological processing (like stemming and lemmatization), syntactic parsing, and semantic analysis, which lay the groundwork for machines to grasp the underlying meaning and context. These processes are intricate, involving complex algorithms to decipher the structure and meaning of language beyond mere word recognition.

- *Advanced Processing and Model Building:* Feature Engineering is the stage where significant attributes or features of the text are extracted and prepared for analytical modelling. This process involves selecting the most relevant aspects of the text, like word frequencies, sentence length, part-of-speech tags, and semantic features, which will be instrumental in training machine learning models.

Model Building is a pivotal phase where the actual machine learning takes place. Here, various models, from rule-based systems to advanced deep learning networks, are trained on the pre-processed and feature-engineered text. These models learn to perform specific NLP tasks, such as sentiment analysis, entity recognition, or language translation, adapting and improving their accuracy and efficiency over time through iterative training and testing.

- *Evaluation and Deployment:* The Evaluation phase rigorously assesses the model's performance, using metrics like accuracy, precision, recall, and F1 score to ensure that they meet the desired standards of linguistic understanding and generation. This step is crucial for refining the models, tweaking parameters, and validating the effectiveness of the NLP pipeline in real-world scenarios.

Finally, in the Deployment stage, these well-trained and validated models are integrated into applications and systems, ready to interact with end-users and other digital platforms. This stage marks the culmination of the NLP pipeline, where the models are put to the test in live environments, handling real-time language processing tasks, from powering chatbots and virtual assistants to analyzing customer feedback and generating automated content.

Throughout these stages, the NLP pipeline represents a dynamic and iterative process of continuous improvement and adaptation, reflecting the complexity and evolving nature of human language. Each stage, meticulously crafted and executed, contributes to the overarching goal of bridging the gap between human linguistic capabilities and machine understanding, enabling more intuitive and effective human-machine interactions.

1.5 NLP's Transformative Impact on Business and Society

The integration of NLP into various sectors has not only revolutionized business operations but also significantly influenced societal interactions, illustrating the profound impact of this technology. In the realm of business, NLP has become a cornerstone for enhancing customer engagement and support using chatbots and virtual assistants. These intelligent systems, powered by sophisticated NLP algorithms, can understand, and respond to customer queries in real time, improving the overall customer service experience. Similarly, NLP plays a crucial role in market intelligence by analyzing social media and news trends, offering businesses invaluable insights into consumer behavior and market dynamics.

E-commerce platforms leverage NLP to create personalized shopping experiences, recommending products based on customer preferences and search history. This capability not only increases customer satisfaction but also drives sales by presenting users with items that are more likely to appeal to them. Additionally, sentiment analysis on customer reviews using NLP techniques helps e-commerce businesses gauge product performance and customer satisfaction, enabling them to make informed decisions about product improvements and targeted marketing strategies. In the financial sector, NLP contributes to risk management

by analyzing news and reports to predict market trends and assess potential risks. Fraud detection is another critical application, where linguistic analysis of financial transactions helps identify suspicious activities, safeguarding against financial fraud. The healthcare industry benefits from NLP in enhancing patient care and monitoring through voice-assisted technologies. Clinical documentation, a time-consuming task for healthcare professionals, is made more efficient and accurate with the automation capabilities offered by NLP, thereby improving the quality of care and reducing administrative burdens.

Furthermore, NLP's influence extends to the education sector, where it facilitates personalized learning experiences. Language learning tools powered by NLP assist users in acquiring new languages, making language learning more accessible and effective. In the public sector, NLP aids policy analysis by analyzing public opinion and feedback on social issues and policies, ensuring that policymaking is informed by the voices of the people.

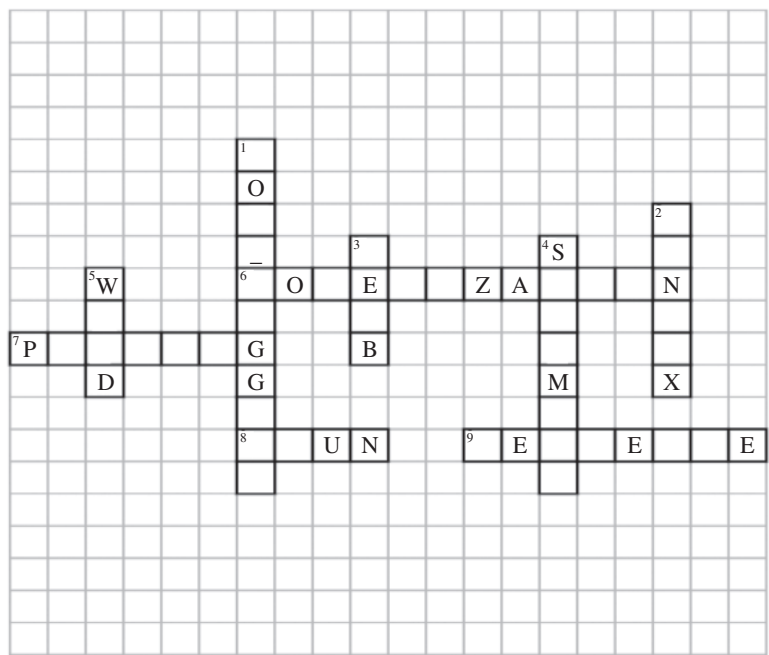
In summary, NLP's applications across various industries demonstrate its versatility and transformative potential. From improving customer interactions and driving personalized experiences in e-commerce to enhancing patient care in healthcare and supporting data-driven policymaking in the public sector, NLP stands as a pivotal technology shaping the future of business and society.

As we traverse the NLP landscape, we encounter its practical applications in everyday life. From powering search engines and virtual assistants to facilitating real-time translation services, NLP is integral to modern digital experiences. It streamlines email filtering, personalizes social media feeds, enhances customer service interactions through chatbots, and even aids in healthcare by analyzing patient records for improved diagnostic accuracy. In essence, NLP is not just about enabling machines to understand language; it's about fostering an environment where technology comprehensively interacts with human nuances, continually advancing to meet the complexities of human communication.

Knowledge Checkpoint 1

Crossword Puzzle: "Exploring Language and NLP"

Complete the crossword using your knowledge of English grammar and NLP. Each clue corresponds to a term or concept from these fields.



Clues:

Across	Down
6 NLP technique for breaking down text into sentences, words, etc. (13 letters)	1 NLP task of assigning parts of speech to each word in a sentence. (11 letters)
7 NLP task of determining the structure of a sentence. (7 letters)	2 The study of sentence structure and the rules that govern the formation of sentences, essential in syntactic analysis in NLP. (6 letters)
8 A linguistic term for the name of a person, place, thing, or idea, crucial in named entity recognition in NLP. (4 letters)	3 The part of speech that describes an action or occurrence, essential in POS tagging in NLP. (4 letters)
9 In grammar, a group of words that contains a subject and predicate; in NLP, often a unit of analysis.	4 NLP technique for reducing inflected words to their base or root form.
	5 Basic linguistic unit of meaning, crucial in both grammar and NLP. (4 letters)