HARSHIL TANNA

**Abstract**

Natural Language Processing (NLP) is the umbrella term for the various processes and techniques needed to interpret human language. This technology has become an integral part of our lives that shapes the way we function.

Initially, the cognitive abilities of NLP models were restricted, resulting in numerous misinterpretations when given commands outside a specified word set. Similarly, initial iterations of voice-controlled assistants, such as Siri, required users to articulate commands multiple times to be understood.

These challenges were common in the early NLP systems as they struggled to handle the complexities of the human language. Over time, advances in artificial intelligence and machine learning have considerably improved a machine's ability to understand and process natural language to expand the possibility of smoother interactions. For example, the repetitive process of articulating commands facilitated Siri's acquisition of unique speech patterns and accents; thereby improving her capability to comprehend user requests.

This article aims to discuss this fast-growing technology that is seemingly becoming part of our lives more than we could have imagined.

**The Humble Beginnings**

The origins of NLP began long before the birth of personal voice assistants such as Siri—all the way back in the 1950s. The idea of computers understanding and responding to human language was just taking root during this time. One of the earliest breakthroughs was the Georgetown-IBM experiment in 1954, which successfully translated 60 Russian sentences into English. While this achievement sparked excitement, the model's limitations were clear; it could only handle a predefined set of words and grammar syntax[1].

During the 1970s and 1980s, significant advancements were seen in speech-to-text systems. These systems relied still heavily on set rules and manually developed language comprehension[2], but instead employed *statistical models* for processing spoken language. For example, a basic model could decompose a sentence into smaller components such as words or phonemes (the fundamental units of sound).

An earlier system called Harpy was developed by Carnegie Mellon University. It expanded on this to recognise more than 1,000 words by associating phonemes with words through *Hidden Markov Models* (HMMs).

These models worked by examining the sequence of phonemes and predicting the next word or phrase that is likely to occur next. For instance:

**Phoneme Prediction Using HMMs:** - Input phonemes: $/k/, /æ/, /t/$ - Prediction: "cat"

The system would then calculate the probability of each new word depending on the previous word.

**Example Sentence:** "The cat sat on the mat."

**Probability of Words:**

$$P(\text{"sat"} \mid \text{"cat"}) = 0.7$$

$$P(\text{"dog"} \mid \text{"cat"}) = 0.2$$

Here, *P* is the probability of a word (*'sat'*/*'dog'*) occurring when given a reference word such as 'cat'. These probabilities were determined through training on large data sets such as books and transcripts which contained a general idea of the patterns within the English language.

Although these models were groundbreaking during the time of their release, they showed notable weaknesses. They encountered difficulties with variations in

accents and speech patterns due to the abstract nature of natural language and the rapid evolution of individual speaking habits. Nevertheless, the statistical model shown above was crucial in progressing NLP.

**From Babble to Brilliance**

Siri wasn't always Apple's masterpiece. She started as a project by a startup founded in 2007, Siri Inc. Through research funded by DARPA, the start-up had a mission of designing an intelligent AI assistant that could understand human language and perform tasks with ease. Originally, Siri was a standalone app on the App Store and stole the centre stage for 2 months[3]. She was especially proficient at understanding natural language and integrating with external APIs like OpenTable and Yelp during the early days of voice assistant technology.

In 2010, Apple acquired Siri Inc. and used their platform to propel Siri from a simple app to an exclusive feature of the iPhone 4s backed by the massive processing power of the iOS ecosystem[4].

But this isn't about Apple's success.

The impact that Siri's 'boom' had on the rest of the world was unprecedented. Her introduction marked a pivotal moment not just for Apple, but for the entire field of NLP. Almost overnight, Siri transformed NLP from a niche area of research at the time into a mainstream field that was exceptionally high in demand. Companies across the world were racing to develop their own voice assistants, and universities saw a surge of interest in AI and NLP courses.

**How Does Modern NLP Work?**

After the initial 'boom', various NLP models were developed in quick succession with one another. Today, these models have reached heights previously never imagined before.

Modern Natural Language Processing utilises deep learning techniques developed throughout the years to accurately break down and analyse human language. The methods below describe a few brief ideas in modern NLP.

1. <u>Tokenization</u>

Before any words or grammar syntax can be processed by a machine, the sentence must be broken up into 'tokens'.
For instance:

$sentence\ =\ \textit{'the quick brown fox jumped over the lazy cat'}$

$$tokenized = [\text{'}the\text{'}, \text{'}quick\text{'}, \text{'}brown\text{'}, \text{'}fox\text{'}, \text{'}jumped\text{'}, \text{'}over\text{'}, \text{'}the\text{'}, \text{'}lazy\text{'}, \text{'}cat\text{'}]$$

Notice how the example above created two tokens ($\text{'}the\text{'}$) even though they are the same word. A method many modern NLP models employ to optimise large word sets is an algorithm known as *Byte-Pair Encoding*. This is left for the reader to explore further.

## 2. Embedding

Tokens are converted into numerical representations called embeddings which capture the meaning of words.

For instance:

### Numerical Representation:

$$\text{king} = [1.2, 3.4, -0.5, \ldots]$$

$$\text{queen} = [1.1, 3.3, -0.4, \ldots]$$

Interestingly, these numerical representations also allow for analogies such as:

### Word Embeddings for Analogy:

$$\text{king} - \text{man} + \text{woman} \approx \text{queen}$$

## 3. Contextual Understanding

One of the most significant breakthroughs in NLP was the introduction of transformer models such as *BERT* (Bidirectional Encoder Representations from Transformers) and *GPT* (Generative Pre-trained Transformer). GPT is specifically used in generative text and may be familiar to the reader due to its use in *ChatGPT*. Transformers use a feature called *self-attention* to understand the relationships between words in context.

For instance:

## Contextual Understanding Example:

**Sentence 1:** "I went to the bank to deposit money."
**Sentence 2:** "I sat by the bank of the river."
Transformers use surrounding context to infer whether "bank" refers to a financial institution or the edge of a river.

Self-attention lets a model look at the whole sentence at once and figure out which words are most relevant to each other.

For example, in the sentence *'The cat sat on the mat because it was tired,'* self-attention helps the model identify that *'it'* refers to *'the cat'*. It does this by assigning higher importance (or *attention*) to related words. For further research, the reader is advised to study the paper *'All You Need is Attention'*.[5]

   4. **Sequence Modelling**

After Tokenization and Embedding, the model must determine the order of words in the input as this can change the meaning of a prompt dramatically. This is achieved through *positional encodings*, which are added to the embeddings once calculated.

Furthermore, models may be asked to complete sentences with high grammatical accuracy. These words are predicted using probabilities from datasets.

## Predicting the next word:

$$P(\text{"on"} \mid \text{"The cat sat"}) = 0.8$$

$$P(\text{"under"} \mid \text{"The cat sat"}) = 0.2$$

Given these probabilities, the model predicts the sentence:
"The cat sat on the ..."

   5. **Pre-Training and Fine-Tuning**

Even without extensive knowledge of machine learning, it's fairly obvious that learning cannot occur if someone repeatedly uses the same examples in an attempt to understand a concept. Some augmentation is needed to ensure that learning can be applied outside of set examples. This concept applies fairly similarly to machines too. The model must be trained on a large data set with augmentation

to avoid overfitting, which in layman's terms simply means 'getting used to only one problem.'

## Next Time?

As NLP evolves further, a future with personal AI assistants for every human on Earth doesn't seem so distant. Although this may raise ethical concerns about integrating such invasive technology into our lives, it is undeniably valuable, especially to those with disabilities that may restrict them from the online world.

Circling back to the title of this paper, Siri herself has recently received major upgrades to her software after more than a decade, now running on the new Apple Intelligence model. This highlights just how vast the future of NLP is—not just as a tool, but as a bridge between humans and machines. Siri's story is a testament to how far we have already come, and maybe, just maybe, she'll finally answer the question, *'Hey Siri, what's the meaning of life?.'*

Until then, we'll keep asking, and NLP will keep evolving.

[1] *The invention of voice recognition, this century's phenomenon* (2019) *CIO*. Available at: https://www.cio.com/article/220152/the-invention-of-voice-recognition-this-centurys-phenomenon.html (Accessed: 19 November 2024).

[2] *Natural language processing* (2024) *Encyclopædia Britannica*. Available at: https://www.britannica.com/technology/natural-language-processing-computer-science (Accessed: 19 November 2024).

[3] *History of Siri* (2018) *YouTube*. Available at: https://www.youtube.com/watch?v=4ryQTkDWmBg (Accessed: 19 November 2024).

[4] Johnson, B. (2013) *How siri works*, *HowStuffWorks*. Available at: https://electronics.howstuffworks.com/gadgets/high-tech-gadgets/siri.htm (Accessed: 19 November 2024).

[5] Vaswani, A. *et al.* (2023) *Attention is all you need*, *arXiv.org*. Available at: https://arxiv.org/abs/1706.03762 (Accessed: 19 November 2024).