# Title – NLP Transformers (Option #2)

Natural Language Processing (NLP) bridges the gap between machines and humans it enables machines to comprehend, decode and interpret human language. NLP models can be trained to understand the data in text format, pdf, voice commands etc. As the processors and chips are becoming more and more powerful and less costly business can create complex NLP models giving them the edge over their competitors.

As depicted in the figure below, we had models that processed fewer than 100 million parameters in 2018 (which is impressive already). NLP models can now interpret more than 100 billion parameters, which indicates development of more than 1,000 fold. Although not all organizations use such big models (and they should not, it is not a smart investment for lots of businesses), improvements in chip technology positively affect the general capacity of NLP models.

LLM’s size over years

A graph with numbers and lines

Description automatically generated

## Future applications of NLP

* **Conversational AI** – chatbots and virtual assistants will become more intuitive and context aware.
* **Language barrier and cross-cultural communication** – NLP will continue to contribute to the language translation tools more so in real time.
* **Content creation** – Content like news articles and reports will be automated and it can even be tailored to suit individual preferences.
* **Improved accessibility** – NLP will make the digital content more reachable to individuals with disabilities. Screen readers, speech recognition and voice assistants with empower people with disability.
* **Healthcare** – with NLP medical staff can interact with systems in their language without having to worry about the coding and can use the power of machines for research, pattern matching, image scanning and classifications.
* **Education sector** – with the power of NLP we will see a lot of interactive learning with no limits on asking questions and getting into details of any problem.
* **Responsible AI** - With great power comes great responsibility. As NLP becomes more integrated into our lives, we must grapple with ethical questions surrounding privacy, data security, and algorithmic bias. Striking the right balance between convenience and safeguarding user rights will be an ongoing challenge.

## How might NLP affect other areas of the AI field?

* **Improved data analysis and mining** – NLP can unlock insights from vast amounts of unstructured text data like social media posts, customer reviews, and documents. By extracting meaning and sentiment from this data, AI can inform better decision-making in areas like marketing, finance, and social good initiatives.
* **Better data collection** – NLP can enhance the data preparation step for ML algorithms by identifying key concepts and relationship with the data leading to more robust and accurate AI models.
* **Robotics** - NLP can equip robots with the ability to understand human commands and respond accordingly. This can revolutionize fields like manufacturing, healthcare, and disaster response by enabling robots to better collaborate with humans.
* **Retraining AI models** - NLP can automate the process of extracting knowledge from text sources like scientific papers and news articles. This can accelerate scientific discovery and keep AI systems up to date with the latest information.

## What kinds of ethical and social problems might a convincing NLP agent cause?

* **Misinformation and Bias:** NLP agents trained on massive amounts of text data can inherit biases and factual inaccuracies present in that data. A convincing agent could spread misinformation or present biased views as truth, manipulating people's opinions or actions.
* **Deepfakes and Social Engineering**: NLP could be used to create highly realistic audio or video forgeries (deepfakes) that impersonate real people. These deepfakes could be used to damage reputations, sow discord, or manipulate elections. Additionally, convincing NLP agents could be used for social engineering, tricking people into revealing personal information or taking unwanted actions.
* **Loss of Trust and Human Connection**: Over-reliance on convincing NLP agents could erode trust in real human communication. People might become accustomed to perfectly tailored and manipulative interactions, leading to skepticism of genuine human interactions.
* **Job Displacement and Automation**: As NLP agents become more adept at tasks currently done by humans, they could automate jobs in areas like customer service, writing, and even creative fields. This raises concerns about job displacement and the need for retraining and social safety nets.
* **Existential Questions and Control:** Highly convincing NLP agents could blur the line between human and machine intelligence. This could lead to complex philosophical questions about sentience and consciousness and raise concerns about who controls these powerful agents and how they are used.

## What machine learning techniques are used in NLP and what techniques/technologies have promise for the future?

* **Sentiment Analysis** – this involved analyzing the sentiments hidden behind the data whether it’s positive, negative, or neutral. By using sentiment analysis, we can convert customer feedback from large volumes of data into actionable insights which can be used to improve the product in every iteration.
* **Named Entity Recognition (NER)** – also known by the name of entity extraction/identification is a NLP technique that automatically identifies named entities in a text and classifies them into predefined categories. Entities can be names of people, organizations, locations, times, quantities, monetary values, percentages, and more.





With NER it becomes easy to put context on the incoming text data.

* **Text Summary** – At times it’s a daunting task to go through the entire text and fetch meaningful information because it can be written by an expert in a medical, technical or any other field of study so it becomes critical to get the summary out of that text. BY applying basic noun-verb linking algorithms text summary can quickly summarize this text

We can try this out here - <https://monkeylearn.com/text-summarizer-online/>

* **Topic Modeling** - Topic Modeling is an unsupervised Natural Language Processing technique that utilizes artificial intelligence programs to tag and group text clusters that share common topics.

Topic modeling could be used to identify the topics of a set of customer reviews by detecting patterns and recurring words. Let’s look at how an ‘unsupervised’ technique would group the below review for Eventbrite, for example:

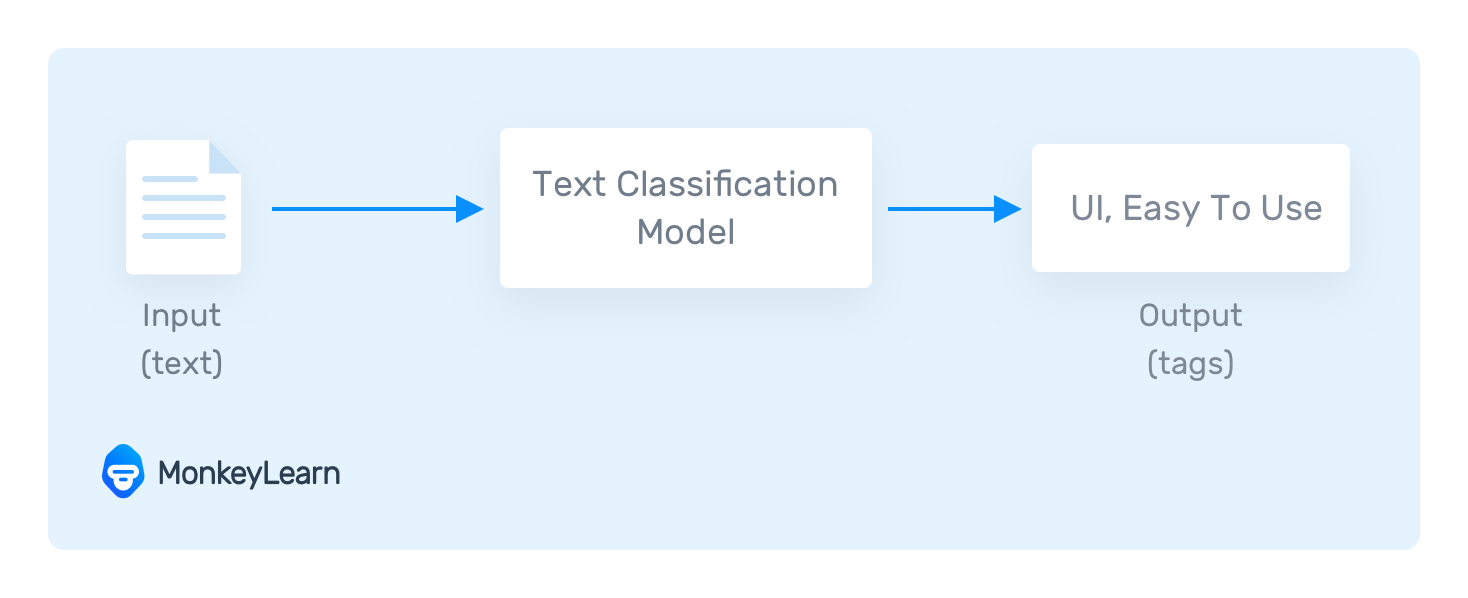
“The nice thing about Eventbrite is that it's free to use as long as you're not charging for the event. There is a fee if you are charging for the event – 2.5% plus a $0.99 transaction fee.”

By identifying words and expressions such as free to use, fee, charging, 2.5% plus 99 cents transaction fee, topic modeling can group this review with other reviews that talk about similar things (these may or may not be about pricing).

* **Text Classification -** Text classification is the organizing of large amounts of unstructured text (meaning the raw text data we are receiving from customers).

Text classification takes the text dataset then structures it for further analysis. It is often used to mine helpful data from customer reviews as well as customer service logs.

E.g. -

A [text classifier](https://monkeylearn.com/text-classifiers/) can take this phrase as an input, analyze its content, and then automatically assign relevant tags, such as *UI* and *Easy To Use*. 

* **Keyword Extraction -**  Keyword extraction is the automated process of extracting the most relevant information from text using AI and machine learning algorithms.

Try it out here - <https://monkeylearn.com/keyword-extractor-online/>

* **Lemmatization and Stemming - L**emmatization and stemming refers to the breakdown, tagging, and restructuring of text data based on either root stem or definition. That might seem like saying the same thing twice, but both sorting processes can lend different valuable data. This involves cleaning up the input data and removing extra noise from it like stop words and repeating words so that it can be staged for further processing.

## Promising Technique for future

I think here are the 2 areas of further exploration which needs to be worked upon

* Explainable AI (XAI): Making NLP models more transparent and interpretable is crucial for building trust and understanding how these models arrive at their decisions.
* Neuro-Symbolic AI: This emerging field aims to combine the strengths of symbolic AI (rule-based) with deep learning for more robust and commonsense reasoning capabilities in NLP tasks.

# Ref –

<https://research.aimultiple.com/future-of-nlp/>

<https://huggingface.co/learn/nlp-course/en/chapter1/4>

<https://www.sciencedirect.com/science/article/pii/S1551741123002802>

<https://www.geeksforgeeks.org/ethical-considerations-in-natural-language-processing-bias-fairness-and-privacy/>

<https://monkeylearn.com/blog/natural-language-processing-techniques/>