

# Overview of NLP

## Defining NLP

This is anything relating to processing natural languages, specifically algorithmically. This could be interpreting spoken word or text, generating spoken word or text. Through NLP we give computers the ability to use and understand our language.

## AI vs NLP

NLP is a branch of AI. AI is "Artificial Intelligence" and means broadly achieving a simulation of human intelligence or capabilities. NLP serves as a tool, like Machine Learning, to achieve that end.

## NL Understanding vs NL Generation

Both natural language understanding and natural language generation deal with with an algorithm understanding the concept of language, just like humans do. But NLU strictly creates/finds meaning in text, while NLG deals with creating meaning *out* of language.

It's worth noting reading text or hearing audio (and producing audio) are problems that are separate from comprehension, but closely intertwined.

## Modern NLP

It be a bit redundant to talk about the wildly popular chat bots of today (ChatGPT). But that is the main way I interact with NLP bots. However, I think the most common NLP interface are the auto suggestions when typing on modern phones. Autocomplete on android and iOS devices is quite complicated nowadays, and tech like that is reaching into code completion (like github copilot). My favorite use of NLP is Google Lens automatically translating text in images (on the image), it makes shopping in asian markets simple.

## 3 Main Approaches to NLP

How might we build NLP algorithms that learn from text?

### Rules-Based

There are some simple rules in natural language that we might use to understand/process language. While natural languages are often too complex for this to work completely, it can get us very far. A good example is translating the language of Pig Latin to English. A programmer

would have to tokenize syllables, remove the 'ay' at the end, and other stuff, but it is possible with just rules.

## Statistical and Probabilistic

Language is more complicated than that however, and often there is nuance like tone, or complex dialects that can't be read with just rule. A simulation could emulate things like the common emotional meaning of a word, through statistics. Machine Learning methods could find the meaning of a corpus (group of text) just by interpreting the amounts of different words in the text. If you were to apply statistical methods to analyze the entirety of what I've spoke, you would find I'm usually quite informal!

## Deep Learning

Given a large enough dataset, we create models that are beyond our understanding of simply "this is more probable". Language models may grow a complex (and transferable) understanding of language by itself, and applied to different use cases. This is seen in how a lot of OpenAI's work is very powerful in not only it's accuracy in emulating human language, but it's applications in smaller projects. I enjoy **ai dungeons** that utilize GPT-3 to create roleplaying worlds to explore.

## Why am I Here?

I've been on a bit of a quest to find what I am interested in. Machine Learning, NLP, and AI are all fields that seem to be the future of computing to me, and that is partly why I am interested in them. However, I also hope to create my own impact on the computing world with *something* that promotes environmentally friendly smart tech usage. This starts in minimizing our constant reliance on remote networks for AI solutions. I want to move processing like NLP in smart assistants closer to the user. NLP is the field of AI that most people understanding the impact of, so that is the place to start *reducing* the impact.