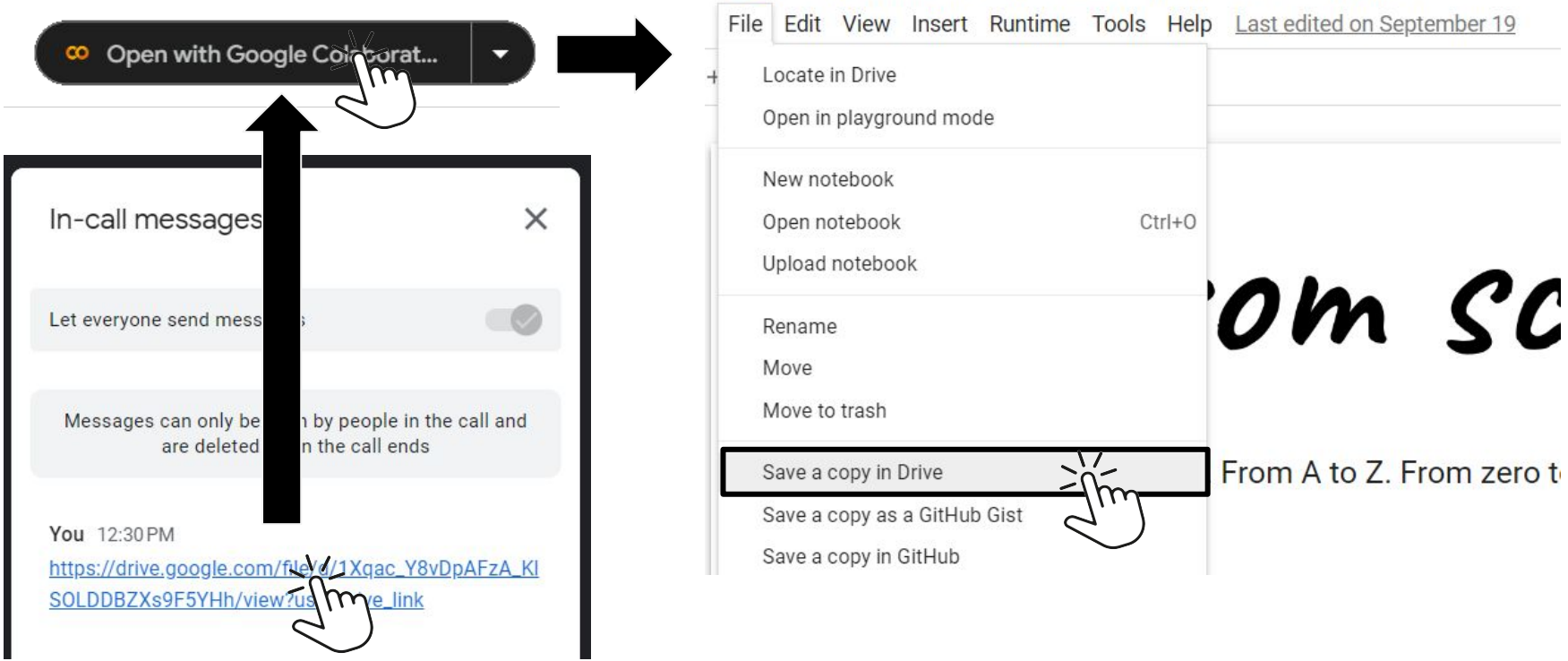
Before we get started:



NLP from scratch

Learn natural language processing. From A to Z. From zero to hero. Fast.

From Zero to NLP in 60

www.nlpfromscratch.com

Housekeeping

Camera on if comfortable doing so

This meeting will not be recorded

Stay muted unless speaking

Be professional

Who am I?

- Data Scientist
- Career consultant (SapientNitro, PwC, Accenture)
- Trainer
- Human


ChatGPT
Capabilities
Remembers what user said earlier in the conversation
Allows user to provide follow-up corrections
Trained to decline inappropriate requests

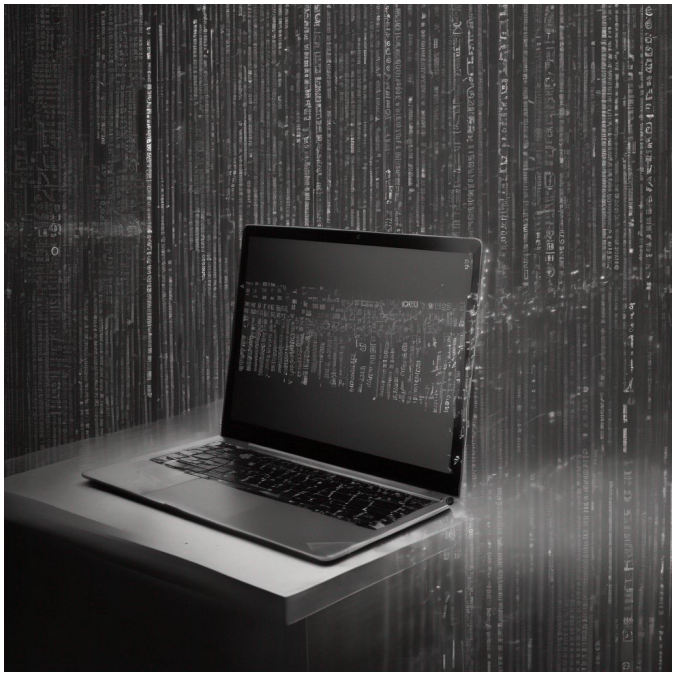
What is Natural Language Processing?

Natural language processing (NLP) is situated at the intersection of the fields of computational linguistics, computer science, and artificial intelligence.

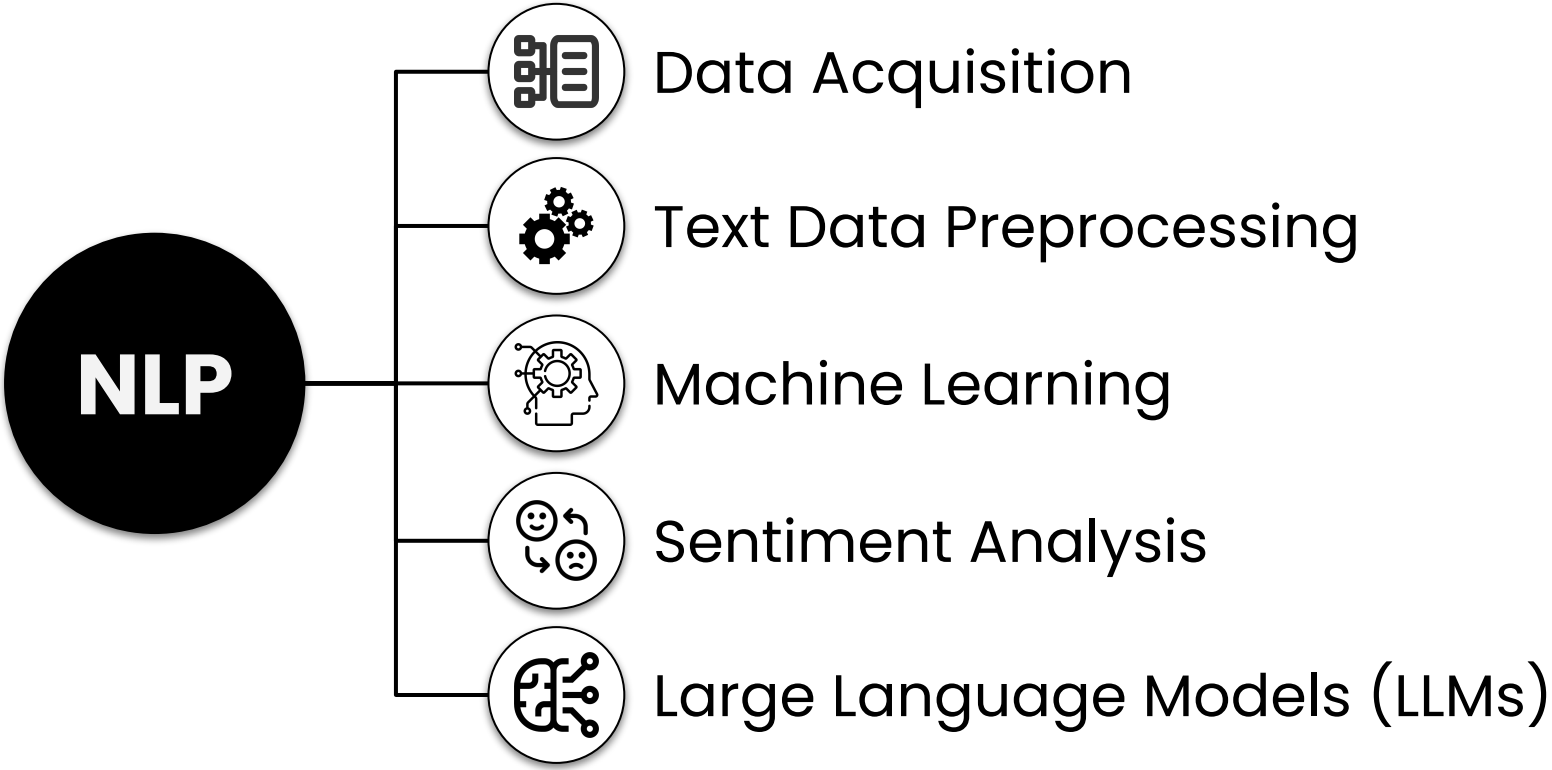
Within the context of data science and AI, NLP aims to enable computers to work with - and potentially even understand - human language for various tasks.

Recently NLP has very much come to the forefront of AI within the popular consciousness, giving the popularity of generative text applications such as ChatGPT.

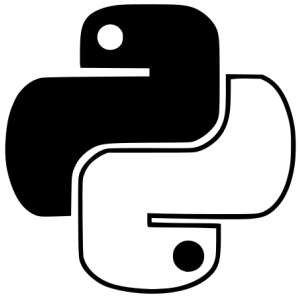
Today we will explore some of what is possible in natural language processing with building from a simple example of data acquisition and processing to fitting a machine learning model.



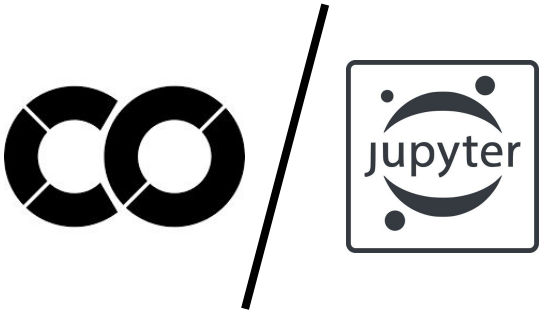
Today's Coverage



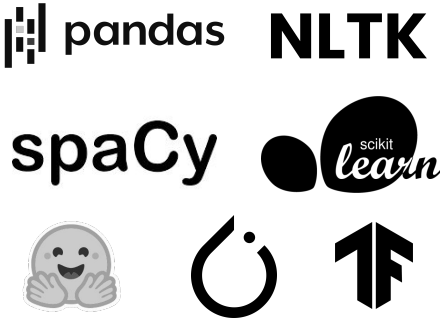
Tools of the Trade



Python 3



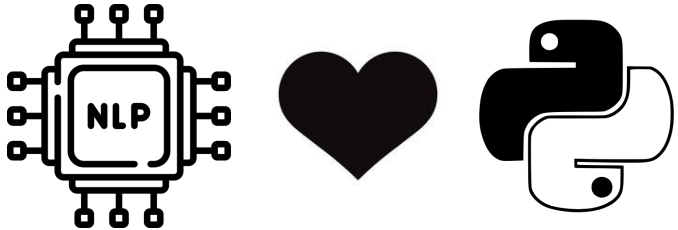
Google Colab
/ Jupyter



NLP and DS Libraries

Python Fundamentals

- Python is a powerful programming language and has become the de facto standard for doing data science work (and a majority of NLP)
- It is easy to learn even for the non-technical or those without prior programming experience
- Working with text at a fundamental level is built into base python and modules in the **pandas library**
- For natural language processing, there is a wide array of free, open source libraries that cover a wide variety of use cases and natural language processing tasks

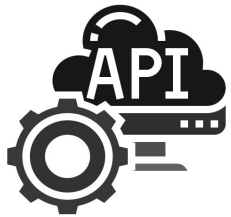


Pandas Library

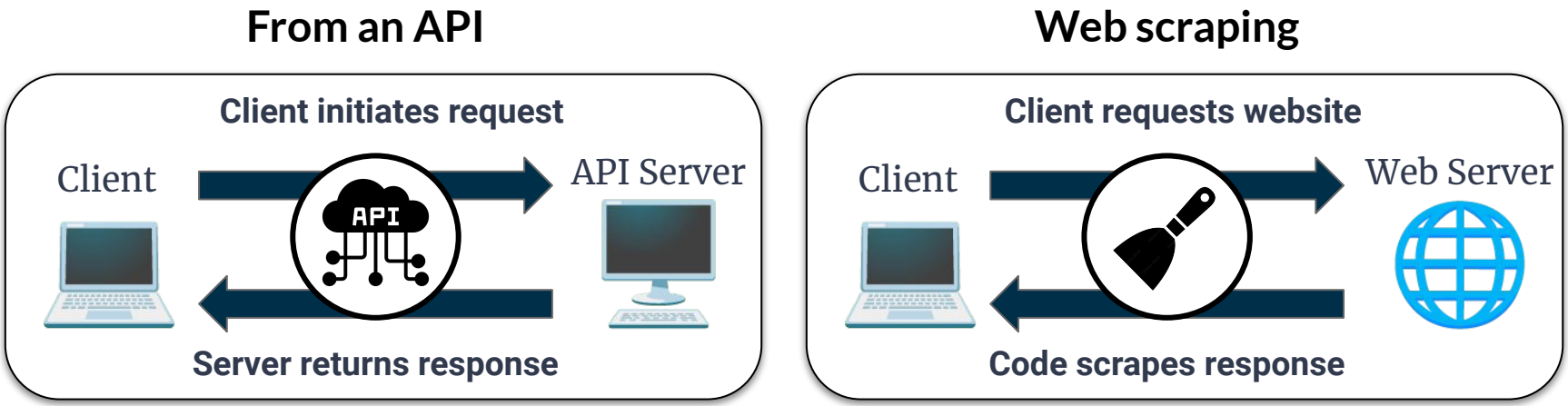
- Pandas is the core library for data manipulation in python and part of the “data science stack”
- Stores data in abstractions of columns (*Series*) and tables of data (*DataFrames*)
- Fast and optimized for working with datasets row-wise with array operations
- For NLP, built-in string accessors for doing text data manipulation on columns easily

Data Acquisition

- We can acquire data either from **requesting it directly from an API**, or extracting that locked in websites by doing **web scraping.**
- This usually requires writing code or using software or a third-party service designed for this task
- A **web service** is an application running on a computer which can provide data or perform transactions when you interact with over the wire
- The machine hosting the service is referred to as a **server** and a machine interacting with it a **client**

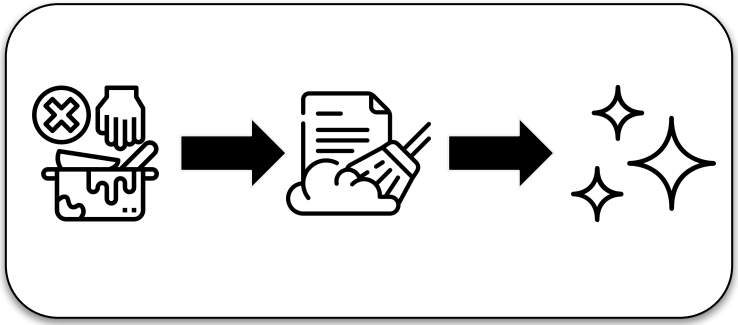


Data Acquisition - API vs. Web Scraping

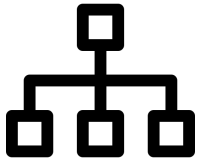


Text Preprocessing

- Refers to cleaning and transforming the original text data into **structured data** to make it suitable for machine learning (or other uses).
- While some preprocessing steps may need to be considered carefully as they may be specific to your use case, there are developed **standard approaches** that work well
- This kind of preprocessing is for traditional NLP approaches; **cutting-edge machine learning methods have their own preprocessing methods** which are more advanced (*e.g.* deep learning, LLMs)



Text Preprocessing Steps



Tokenization

Break free-form text documents down into tokens: constituent units of language (usually words)

Normalization

Apply techniques to reduce the noise and variance in the language data and standardize

Vectorization

Convert text data to numeric features: structured data suitable for machine learning or analytics

NLTK and scikit-learn

- NLTK is the *natural language toolkit*, a free open source python library for working with NLP
- Originally developed at University of Pennsylvania for teaching purposes but now standard in NLP workflows
- Scikit-learn (sklearn for short) is the standard open source library for machine learning in Python
- Covers the complete gamut of ML and also has modules and classes for text-specific tasks (datasets, vectorization, metrics, etc.)

NLTK

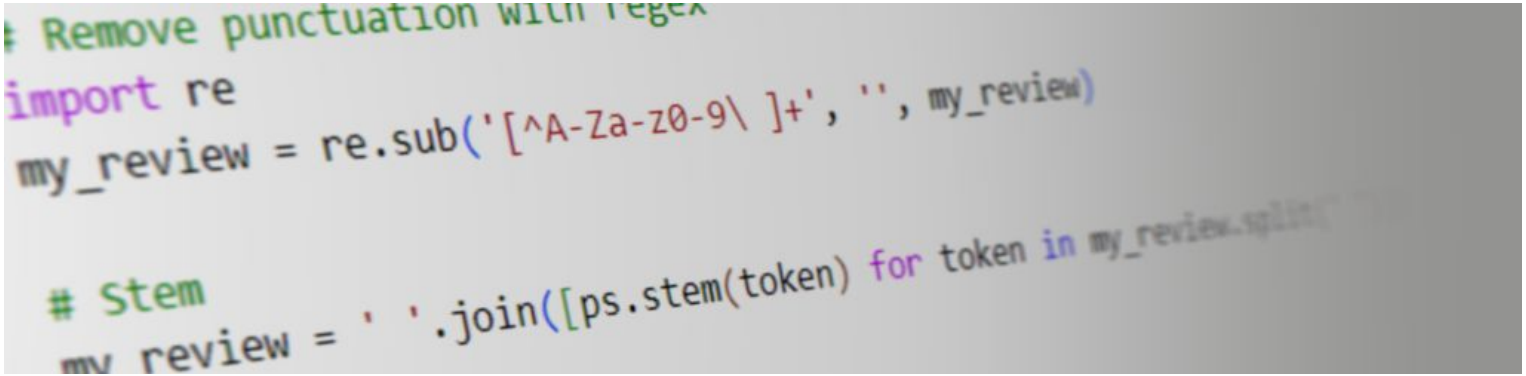
Machine Learning for Sentiment Analysis

- In **supervised learning**, we use our set of input features, X, and an associated set of data labels, y, to train a model to make predictions about unseen data.
- For natural language, a common use case is for that of **sentiment analysis** - predicting the emotional quality and strength of strings of text based on existing labelled data
- Though they are not explicitly trained for this task, given their generalizability, large language models (LLMs) **can perform sentiment analysis tasks** as a result of "zero-shot" learning.
- Here we will be testing Meta's LLaMA 2 chat model which was released in July 2023 and interacting with the through a Hugging Face space

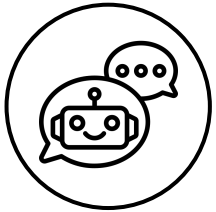
NLP4Free 🔠⚡🤖🧠😀

https://github.com/nlpfromscratch/nlp4free

A Free Natural Language Processing (NLP) microcourse, from basics to deep learning



LLM and Generative AI Workshops

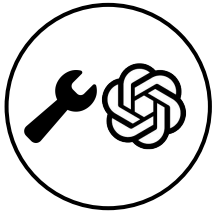


$199

Large Language Models and Generative Text

Dec 2023 **3 hours, 19:00-22:00 & 14:00-17:00 EST**

- Large Language Models (LLMs) and the transformer architecture
- Using LLMs for generative text and prediction tasks with Hugging Face
- Fine-tuning generative text models with custom data
- Performance efficient fine-tuning methods including quantization and LoRA



~~$199~~
$139

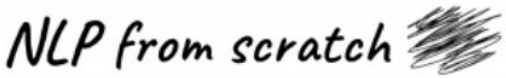
Discount Code:
INSIDER30 (30%off)

Building Generative AI Applications with OpenAI and GPT

Jan 2023 **3 hours, 19:00-22:00 & 14:00-17:00 EST**

- Introduction to Generative AI and text generation
- Working with the OpenAI API and models for text generation, chat, and more
- Getting set up with OpenAI and your generative AI development environment
- Building a streaming web application for chat using GPT
- Open source and offline alternatives to APIs

NLP from scratch



TRAINING CONSULTING FAQ A

Today ● December 2023 Filters

Upcoming eve
Dec 01 - Feb 2

Sign up at **nlpfromscratch.com/training!**

Sun	Mon	Tue	Wed	Thu	Fri	Sat
						2
3	4 ● What the Heck is a...	5 Deconstruct Care...	6	7 Zero to NLP in 60 ...	8	9
10	11 Large Language M...	12	13 What the Heck is a...	14	15 Large Language M...	16
17	18	19	20	21	22	23

DEC 04 What the LLM? - F

DEC 05 Deconstr Progress

DEC 07 Zero to N FREE Wo

DEC 11 Large La Models a

DEC 13 What the LLM? - F

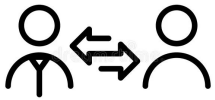
Manifesto

Knowledge is only valuable if it is useful.

The best way to learn is by doing.

Learning is a non-linear process.

Learning is exploration, not a journey.



Teaching and learning are complementary.

I would value your feedback.

NLP from scratch

Learn natural language processing. From A to Z. From zero to hero. Fast.

www.nlpfromscratch.com