Alberto Andrés Valdés González.
Degree: Mathematical Engineer.
Work position: Data Scientist.
Mail: anvaldes@uc.cl/alberto.valdes.gonzalez.96@gmail.com
Location: Santiago, Chile.

Overview LLM

Main problems of NLP:

1. Completation: This is when we try complete one text. For example:

```
INPUT = What is your

OUTPUT = What is your name?
```

2. QA: This is very common for chat bots. For example:

```
INPUT = Who is Cristiano Ronaldo?

OUTPUT = Cristiano Ronaldo is a portuguese footballer who is known
for his good skills and awards
```

3.Intruction: This is when we try to indicate to do something to the AI. For example:

```
INPUT = Create a poem

OUTPUT = POEM
```

4. Classification: This is when we want to classify some text. For example:

```
INPUT = This is a good day!

OUTPUT = Positive
```

Main libraries of LLM:

- Hugging Face
- Scikit LLM
- Langchain

Versions of LLMs: Base vs Instruct vs Chat.

Base: These models are trained on a wide variety of different texts, so they make minimal assumptions about the structure of the text they're completing.

Instruct: This version of the model has been fine-tuned to be able to follow prompted instructions. These models 'expect' to be asked to do something.

Chat: The 'instruct' Models with the 'chat' suffix have been fine-tuned to work in chatbots. These models 'expect' to be involved in a conversation with different actors.

Meaning of abreviatures:

GO: Start of sentence.
EOS: End of sentence.
PAD: Padding.
UNK: Unknown.

How to use a LLM to other tasks?

Is important to provide context. For example:

```
[INST]Analyze the sentiment of the news headline enclosed in square brackets,
determine if it is positive, neutral, or negative, and return the answer as
the corresponding sentiment label "positive" or "neutral" or "negative"

[/INST] [{data_point["text"]}] = {data_point["sentiment"]}
```

For inference you have to provide the next text and the LLM will complete all the text:

```
[INST]Analyze the sentiment of the news headline enclosed in square brackets,
determine if it is positive, neutral, or negative, and return the answer as
the corresponding sentiment label "positive" or "neutral" or "negative"

[/INST] [{data_point["text"]}] =
```
