**Types of Models**

During this research all the model interactions were with the LLM hosted on the Hugging Face Hub. The models, have different uses, types, methods of interaction, memory (parameters) and data used to train them.

Types of Models used based on architecture:

* Encoder-only models: Good for tasks that require understanding of the input, such as sentence classification and named entity recognition.
* Decoder-only models: Good for generative tasks such as text generation.
* Encoder-decoder models or sequence-to-sequence models: Good for generative tasks that require an input, such as translation or summarization.

The main types of models used during this research according to the type of task they perform:

* Text Classification
* Zero-Shot Classification
* Token Classification
* Summarization
* Text Generation
* Text2Text Generation

The size of the model I used during the research goes from 248 million parameters (Google/Flan-t5-base) up to 7 billion.

Models used are a combination of pretrained to finetuned versions of the original. Finetuned models where chosen because of a medical fine tuning they had.