**5. Encoder models**

Encoder models use only the encoder of a Transformer model. At each stage, the attention layers can access all the words in the initial sentence. These models are often characterized as having “bi-directional” attention, and are often called *auto-encoding models*.

A diagram of a number

Description automatically generated

The numerical representation is called either a feature vector or feature tensor. Each word in the initial sequence affects every words representation.

A screenshot of a pink screen

Description automatically generated

Why would you use an encoder?A white background with blue text

Description automatically generated

The pretraining of these models usually revolves around somehow corrupting a given sentence (for instance, by masking random words in it) and tasking the model with finding or reconstructing the initial sentence.

Encoder models are best suited for tasks requiring an understanding of the full sentence, such as sentence classification, named entity recognition (and more generally word classification), and extractive question answering.

Representatives of this family of models include:

* [ALBERT](https://huggingface.co/docs/transformers/model_doc/albert)
* [BERT](https://huggingface.co/docs/transformers/model_doc/bert)
* [DistilBERT](https://huggingface.co/docs/transformers/model_doc/distilbert)
* [ELECTRA](https://huggingface.co/docs/transformers/model_doc/electra)
* [RoBERTa](https://huggingface.co/docs/transformers/model_doc/roberta)

Encoders with their bi-directional context are good at guessing words in the middle of a sequence.

A screenshot of a computer

Description automatically generated

This requires a semantic understanding as well as a syntactic understanding.

A screenshot of a message

Description automatically generated

Encoders are good at obtaining an understanding of sequences; and the relationship/interdependence between words.

We use the model to compute a prediction and to classify the sequences among these two classes (positive and negative). Whilst the two sequences are very similar, containing the same words, the meaning is entirely different and the encoder is able to grasp that difference

To understand what happens inside the Transformer network on a deeper level, we recommend the following blogposts by Jay Alammar: - The Illustrated Transformer: [https://jalammar.github.io/illustrate...](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbVJNeVZiN0lmZ25TSGljaWc3cWptbHNlaWJwd3xBQ3Jtc0tsM0NBbzJxTXd2TGlDQ0J6LVBkenZJQlZvLVhBS092YkFwenhqNVU5V0Jja0RpN3BQck8xN29GYlBZUlppZmE3czVmaW1WMTlEVmJmTDRuVzFBSzBTLTVZWUtmZTFscmpPb2xwUk1qaWtrd0RyU19fWQ&q=https%3A%2F%2Fjalammar.github.io%2Fillustrated-transformer%2F&v=MUqNwgPjJvQ) - The Illustrated GPT-2: [https://jalammar.github.io/illustrate...](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqa3UyMGdWR0pLWVBxQTBWZHpvcDZpV3VEdmp2d3xBQ3Jtc0ttMnJEWUlnMmw0WkRBOS04NjdEVWJYaFc3ZlBVR1hTT3p3T3RZWElaRDF0M0RCQml0bUdHbDFEbE1UT3pmUXhiaS1rUW52VlZ6Z2Q5Wl9UNmxIbkhpRzIwMVpIVHdqUVFXd2dlZnJlaVVQdVhiODNwRQ&q=https%3A%2F%2Fjalammar.github.io%2Fillustrated-gpt2%2F&v=MUqNwgPjJvQ) - Understanding Attention: [https://jalammar.github.io/visualizin...](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbnVEUEJNNWpseXAwU21mbXFDSncwMHRGMUdPd3xBQ3Jtc0tsb2FPSXJEajg4ckR3QjE4R3J1RWkwSFdKOXZSYnZiV1dsdDVTWnBKOVRiTVQ3T2Z6bEVpUFdrVkczaGNOX1JLb2tfREZiaUtFdEhPS19pTU1NZnh0dm5XSU9ENExwbjVrenhKalRGOXFnZW1NMU5Gaw&q=https%3A%2F%2Fjalammar.github.io%2Fvisualizing-neural-machine-translation-mechanics-of-seq2seq-models-with-attention%2F&v=MUqNwgPjJvQ)

Furthermore, for a code-oriented perspective, we recommend taking a look at the following post: - The Annotated Transformer, by Harvard NLP [https://nlp.seas.harvard.edu/2018/04/...](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqa1dMNlY3X0d2MGM5X2Nzc2p5UllwVkZUc3N0UXxBQ3Jtc0ttTlRMcmEwdG1IbWFLQVYydllzdDlQZXItYjluRFJMeDdmZm43SmVRS1BNQ0dkODNTUDFGeHJTWTFpc0dIbnpJcDVKYUFLVmhkWXNTQ3p5eHY5OThoTGJmNWdFaXJURTdRazBfdWFBM0t1VVhUZ2dCZw&q=https%3A%2F%2Fnlp.seas.harvard.edu%2F2018%2F04%2F03%2Fattention.html&v=MUqNwgPjJvQ)