**6. Decoder models**

Decoder models use only the decoder of a Transformer model. At each stage, for a given word the attention layers can only access the words positioned before it in the sentence. These models are often called *auto-regressive models*.

The pretraining of decoder models usually revolves around predicting the next word in the sentence.

These models are best suited for tasks involving text generation.

Representatives of this family of models include:

* [CTRL](https://huggingface.co/transformers/model_doc/ctrl)
* [GPT](https://huggingface.co/docs/transformers/model_doc/openai-gpt)
* [GPT-2](https://huggingface.co/transformers/model_doc/gpt2)
* [Transformer XL](https://huggingface.co/transformers/model_doc/transfo-xl)