

Open Source Generation Models and LLMs

Pretrained Encoder/Decoder and Decoder-Only Models

- BART
 - Model: https://huggingface.co/docs/transformers/model_doc/bart
 - Paper: <https://arxiv.org/abs/1910.13461>
- T5
 - Model: https://huggingface.co/docs/transformers/model_doc/t5
 - Paper: <https://arxiv.org/abs/1910.10683>
- UL2
 - Model: <https://huggingface.co/google/ul2>
 - Paper: <https://arxiv.org/abs/2205.05131>
- Longformer Encoder-Decoder (LED)
 - Model: https://huggingface.co/docs/transformers/model_doc/led
 - Paper: <https://arxiv.org/abs/2004.05150>
- GPT-Neo/GPT-J
 - Model: <https://huggingface.co/EleutherAI/gpt-j-6b>,
<https://huggingface.co/EleutherAI/gpt-neo-1.3B>
- LLaMA
 - Model: https://huggingface.co/docs/transformers/main/model_doc/llama
 - *Need to request model weights, see above for details*
 - Paper: <https://arxiv.org/abs/2302.13971>

FLAN/Instruction-Tuned Models

- Flan-T5
 - Model: <https://huggingface.co/google/flan-t5-base>
 - Paper: <https://arxiv.org/abs/2210.11416>
- Flan-UL2
 - Model: <https://huggingface.co/google/flan-ul2>
 - Paper: <https://arxiv.org/abs/2205.05131>
- Alpaca
 - Model:
 - Paper: <https://crfm.stanford.edu/2023/03/13/alpaca.html>

Chat-Based Models

- Vicuna
 - Model: <https://huggingface.co/lmsys/vicuna-13b-v1.5-16k>
 - Description: <https://lmsys.org/blog/2023-03-30-vicuna/>
- LLaMA 2
 - Model: <https://ai.meta.com/llama/>

- *Need to request access, see above for details*
 - Paper: <https://arxiv.org/abs/2307.09288>
- GPT4All
 - Model/Description: <https://gpt4all.io/index.html>
 - Python package available here:
<https://github.com/nomic-ai/gpt4all/tree/main/gpt4all-bindings/python>
- Cerebras-GPT
 - Model: <https://huggingface.co/cerebras/Cerebras-GPT-13B>
 - Paper: <https://arxiv.org/abs/2304.03208>