## Preprocesado de data para hacer ajuste fino:

<https://github.com/bigscience-workshop/Megatron-DeepSpeed>

<https://github.com/bigscience-workshop/Megatron-DeepSpeed/blob/main/start_fast.md>

Crear dataset en HF

[Share a dataset to the Hub (huggingface.co)](https://huggingface.co/docs/datasets/upload_dataset)

## Fine Tune:

[Fine-tune a pretrained model (huggingface.co)](https://huggingface.co/docs/transformers/training)

Bloom Preprocesado de información:

[GitHub - bigscience-workshop/data-preparation: Code used for sourcing and cleaning the BigScience ROOTS corpus](https://github.com/bigscience-workshop/data-preparation)

## Bloom Fine tuning:

[GitHub - bigscience-workshop/petals: 🌸 Easy way to run 100B+ language models without high-end GPUs. Up to 10x faster than offloading](https://github.com/bigscience-workshop/petals)

[petals/prompt-tuning-sst2.ipynb at main · bigscience-workshop/petals · GitHub](https://github.com/bigscience-workshop/petals/blob/main/examples/prompt-tuning-sst2.ipynb)

[petals/prompt-tuning-personachat.ipynb at main · bigscience-workshop/petals · GitHub](https://github.com/bigscience-workshop/petals/blob/main/examples/prompt-tuning-personachat.ipynb)

## Kaggle Bloom fine tuning

[petals/prompt-tuning-personachat.ipynb at main · bigscience-workshop/petals · GitHub](https://github.com/bigscience-workshop/petals/blob/main/examples/prompt-tuning-personachat.ipynb)