How truly open is "open-source" AI in reality?

NLUUG Spring Conference 2025

Luuk van Drunen
luukvandrunen@gmail.com

Who am I?

- 19-years-old
- Second year AI student
- Self-acclaimed AI-critic
- Tech-savy

What is this?

- My second talk ever at the NLUUG
- Surface-level
- My own unfiltered opinions
- Lowering the average age
- Interactive

The current AI/LLM landscape

- OpenAI/ChatGPT
- Meta/LLama
- Deepseek/r1, v3
- Google/Gemma
- Anthropic/Claude Sonnet
- xAI/Grok
- Alibaba Cloud/Qwen
- In my eyes none of these are **truly** open

The current AI/LLM landscape

- In this day and age what does an “open” AI model mean

Hugging Face

What is Hugging Face?

- Github for ML
- “Open-source” models
- Hosting for datasets
- Transformers library

What is the Hugging Face Hub?

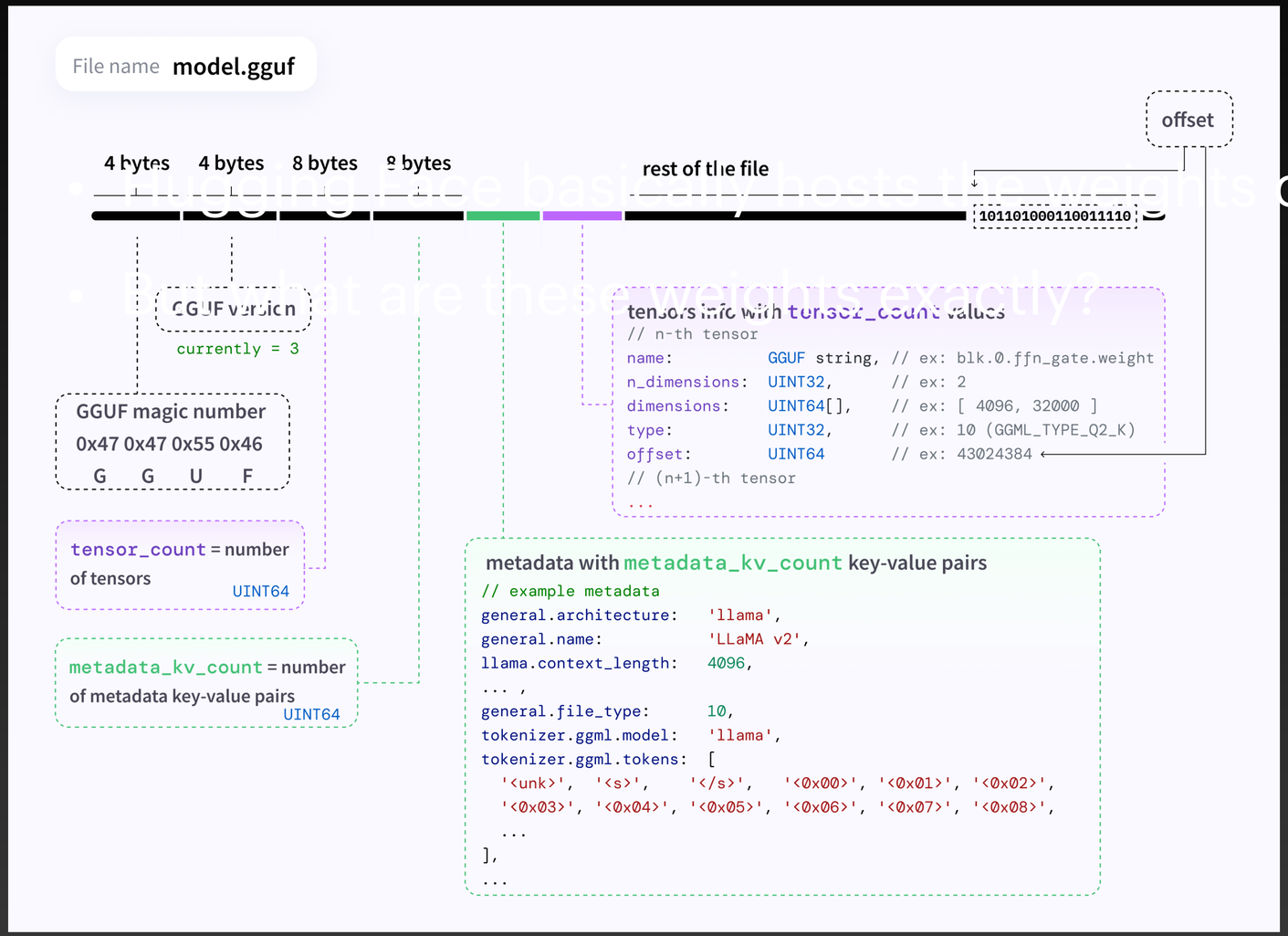
The Hugging Face Hub is a collaboration platform that hosts a huge collection of open-source models and datasets for machine learning, think of it being like Github for ML. The hub facilitates sharing and collaborating by making it easy for you to discover, learn, and interact with useful ML assets from the open-source community. The hub integrates with, and is used in conjunction with the Transformers library, as models deployed using the Transformers library are downloaded from the hub.

https://huggingface.co/blog/noob_intro_transformers#what-is-the-hugging-face-hub

What exactly is hosted on Hugging Face?

- Datasets
  - Needed for model training
- Models
  - What does that actually mean??

What does a model hosted on Hugging Face look like?



- Hugging Face basically hosts the weights of t
- But what are these weights exactly?

File

model-00001-of-00012.safetensors 4.85 GB

2/13 Download View all files

Tensors	Shape	Precision
language_model.model		
language_model.model.embed_tokens.weight	[262 208, 5 376]	BF16
language_model.model.layers.0.input_layernorm.weight	[5 376]	BF16
language_model.model.layers.0.mlp.down_proj.weight	[5 376, 21 504]	BF16
language_model.model.layers.0.mlp.gate_proj.weight	[21 504, 5 376]	BF16
language_model.model.layers.0.mlp.up_proj.weight	[21 504, 5 376]	BF16
language_model.model.layers.0.post_attention_layernorm.weight	[5 376]	BF16
language_model.model.layers.0.post_feedforward_layernorm.weight	[5 376]	BF16
language_model.model.layers.0.pre_feedforward_layernorm.weight	[5 376]	BF16
language_model.model.layers.0.self_attn.k_norm.weight	[128]	BF16
language_model.model.layers.0.self_attn.k_proj.weight	[2 048, 5 376]	BF16
language_model.model.layers.0.self_attn.o_proj.weight	[5 376, 4 096]	BF16
language_model.model.layers.0.self_attn.q_norm.weight	[128]	BF16
language_model.model.layers.0.self_attn.q_proj.weight	[4 096, 5 376]	BF16
language_model.model.layers.0.self_attn.v_proj.weight	[2 048, 5 376]	BF16
language_model.model.layers.1.mlp.gate_proj.weight	[21 504, 5 376]	BF16
language_model.model.layers.1.self_attn.k_norm.weight	[128]	BF16
language_model.model.layers.1.self_attn.k_proj.weight	[2 048, 5 376]	BF16
language_model.model.layers.1.self_attn.o_proj.weight	[5 376, 4 096]	BF16
language_model.model.layers.1.self_attn.q_norm.weight	[128]	BF16
language_model.model.layers.1.self_attn.q_proj.weight	[4 096, 5 376]	BF16
language_model.model.layers.1.self_attn.v_proj.weight	[2 048, 5 376]	BF16
multi_modal_projector		
multi_modal_projector.mm_input_projection_weight	[1 152, 5 376]	BF16
multi_modal_projector.mm_soft_emb_norm.weight	[1 152]	BF16
vision_tower.vision_model		
vision_tower.vision_model.embeddings.patch_embedding.bias	[1 152]	BF16
vision_tower.vision_model.embeddings.patch_embedding.weight	[1 152, 3, 14, 14]	BF16
vision_tower.vision_model.embeddings.position_embedding.weight	[4 096, 1 152]	BF16

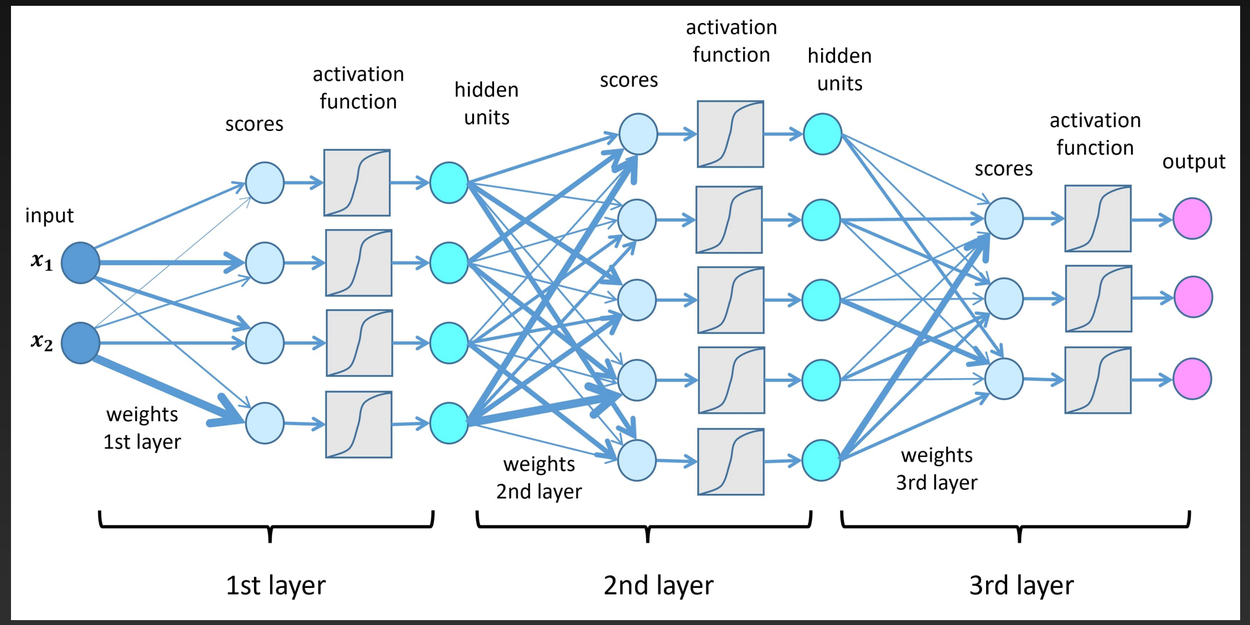
What are these weights exactly?

Without getting to technical

- Representation of the model
- Basically a lot of numbers
  - Tensors to be exact
- These tensors represent the 'weight' of each parameter
- These weights are what make an AI model tick

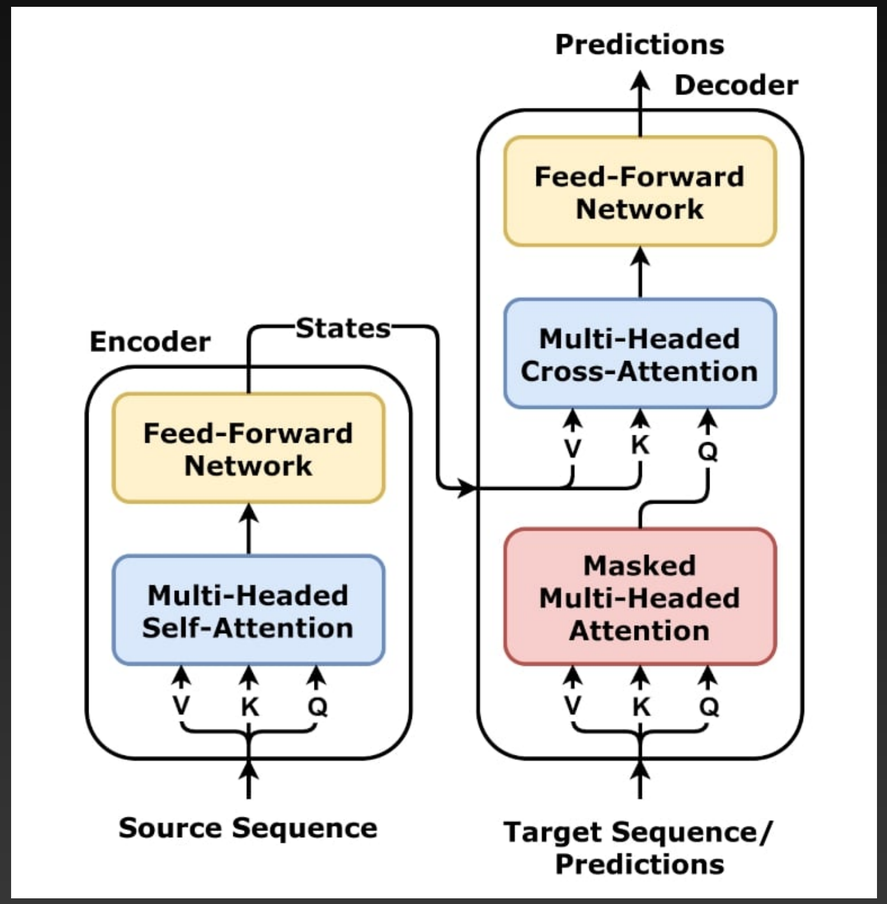
But how do these models work?
Again without getting TO technical

- Neural networks are the base
  - And a huge buzzword
- Modeled after human synapses
- Deep learning
  - Multi-layer NN
  - Supervised vs un-supervised
- LLMs are a bit trickier than this
  - Transformer model used for training



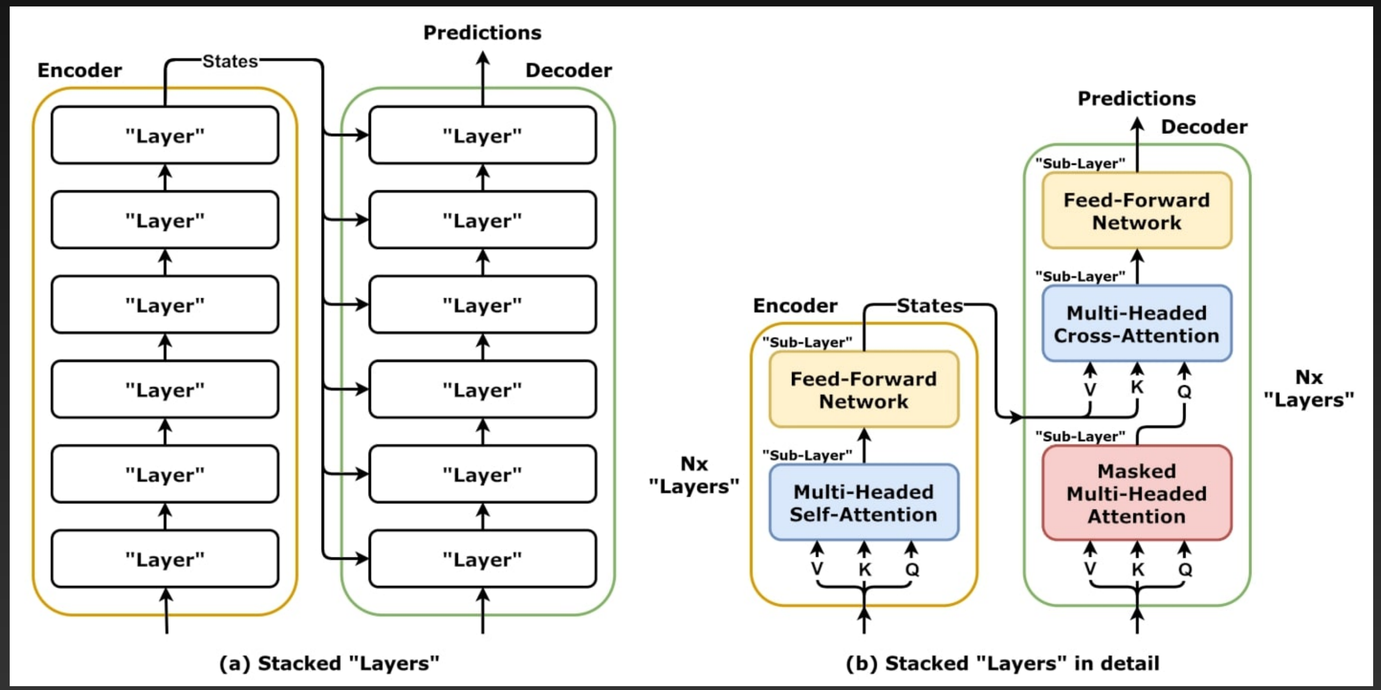
Transformers

- Developed by Google (2017)
  - Attention is all you need
- Tokenization
- Attention
  - Contextualizing
  - Amplifying key tokens
  - Diminish less important tokens



Transformers
Encoder-decoder

- Stacked encoder-decoder layers
- Encoder
  - Contextualizes tokens
  - Self-attention
- Decoder
  - Cross-attention
  - "Self"-attention



Time for the demo

So what's the problem of non-open AI models?

- Nothing can be checked or peer-reviewed
    - DeepSeek paper is an exception
- Keeping the black-box problem alive
    - Input --> black-box --> output
- No (easy) way for right to explanation
    - Kate Vredenburgh (2022)
- Data opacity

What should an truly open model look like?
In my opinion

- Open-weights is a good starting point
  - For tinkerers
- Code used to train the model
  - Replicability
- The data used to train the model
  - The most important in my opinion
  - Contrary to OSI
  - Privacy
  - Copyright

In conclusion

Any questions?

luukvandrunen@gmail.com