

Agenda Open Source

② \$\$\$ in compute

Base Model

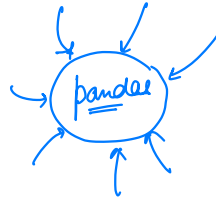
① Enough corpus
very large

③ Days
of training

Pretrained
language
Model

Benefits

- ① Cost effective
- ② Data ownership & control
- ③ Avoid vendor lock-in
- ④ Security & privacy
- ⑤ Customization & specialization



① VMware

↳ open language model → Huggingface Starcoder

② Brave → Mixtral → Mistral AI → Conversational Assistant.

↓
leo

Klausur 2

③ IBM

ASKHR

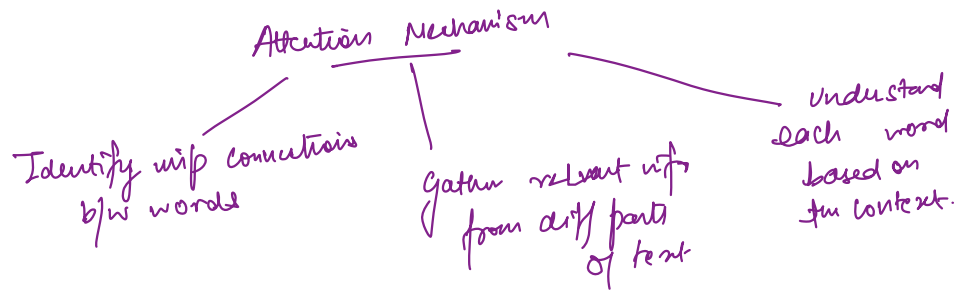
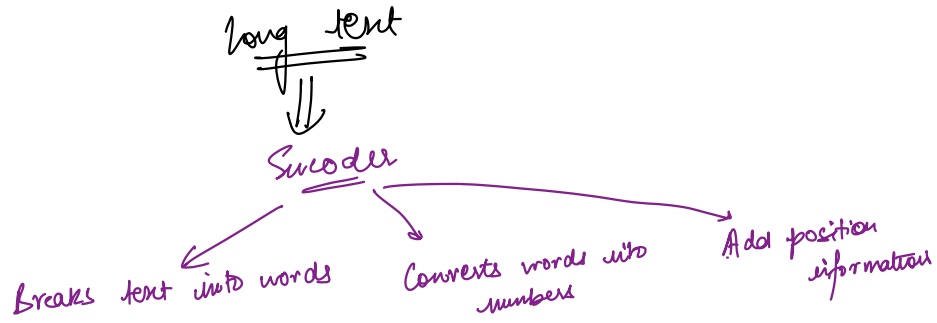
Huggingface

Raw text → embeddings ⇒ logits ⇒ prediction
" " [- - - - -]

downloaded → cached ↙ ↘ ↗ ↖

⊗ Sentiment analysis

- ① Feature extraction
- ② question - answering
- ③ summarization
- ④ translation
- ⑤ text-generation
- ⑥ Zero-shot-classification



Rich Understanding of Text



Decoder

Choose the imp info
to include

logically organise

generates
the concise
text.

Output → Summarized text.