

Medi tron 70B
for LLM

Medical Pretraining

[Medical learn source]
closed source (GPT-4 or PaLM)

limited or state $\leq 13B$ Parameters

MEDITRON - built on open source LLM's
[7B and 70B] builds on Llama 2

Through the adaptation of
Mistral's Megatron LLM distributed Trainers

and trained on Medical corpus

such as
[PubMed Articles, abstracts and More]

MEDITRON

Reasoning

Evaluated on four Medical
Benchmarks. using both on

[in context learning
prompt within
context window] and Task
specific
fine-tuning.

Engineering Behind Medi TRON 70B

To harness the power of Large Language
Model Parameter size and pretraining
token Count

MEGATRON LLM DISTRIBUTED
TRAINING LIBRARY

Extended from Nvidia's Megatron LM
to support three open source LLM's
(Llama, Falcon and Llama2)

supports ① Data Parallelism (DP)
② Pipeline Parallelism (PP)
③ Tensor Parallelism (TP)

Megatron LM \Rightarrow Natively support
GPT like Architecture

The Researcher of Meditron 70B
extended to support (Llama, Llama2,
Falcon)

Meditron 7B

Context length
2048

Meditron 70B

Context length
4096

They followed open AI's ChatML format
to format the instruction data.

ChatML: series of Messages starting with
<|im-start|> followed by the role
(user / Assistant)

example <|im-start|> System
<|im-start|> question etc.