

Understanding Chinchilla Scaling Laws

Instructor

Sourab Mangrulkar

Machine Learning Engineer at

Creator of PEFT



Introduction



Introduction



GPT-3.5

300B tokens

Introduction



PaLM 2

3.6T tokens

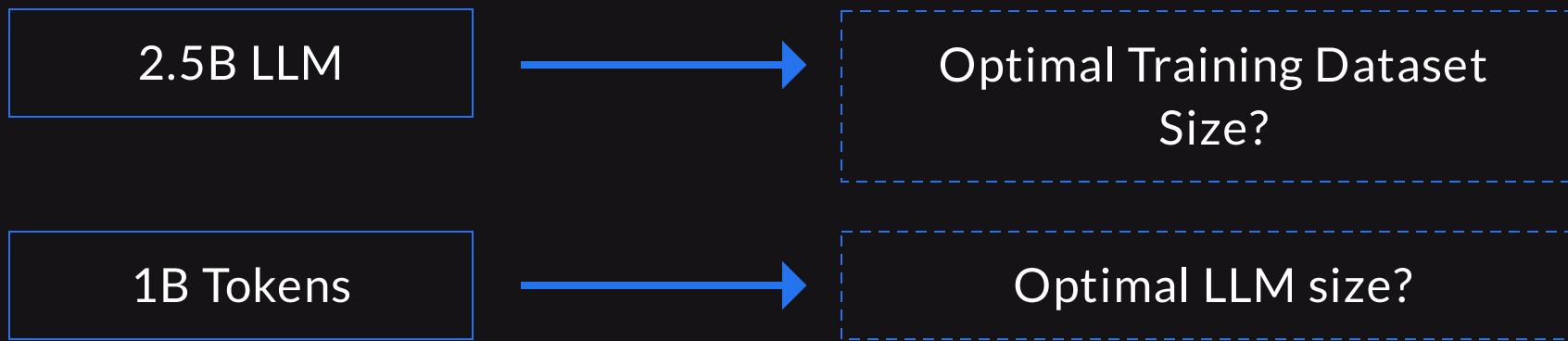
Introduction



2T tokens

How do we determine the optimal model size and size of the training dataset?

Introduction



What are Scaling Laws?

Training data size vs Model size

OpenAI Scaling Laws

Increasing model size was more important than scaling data size.

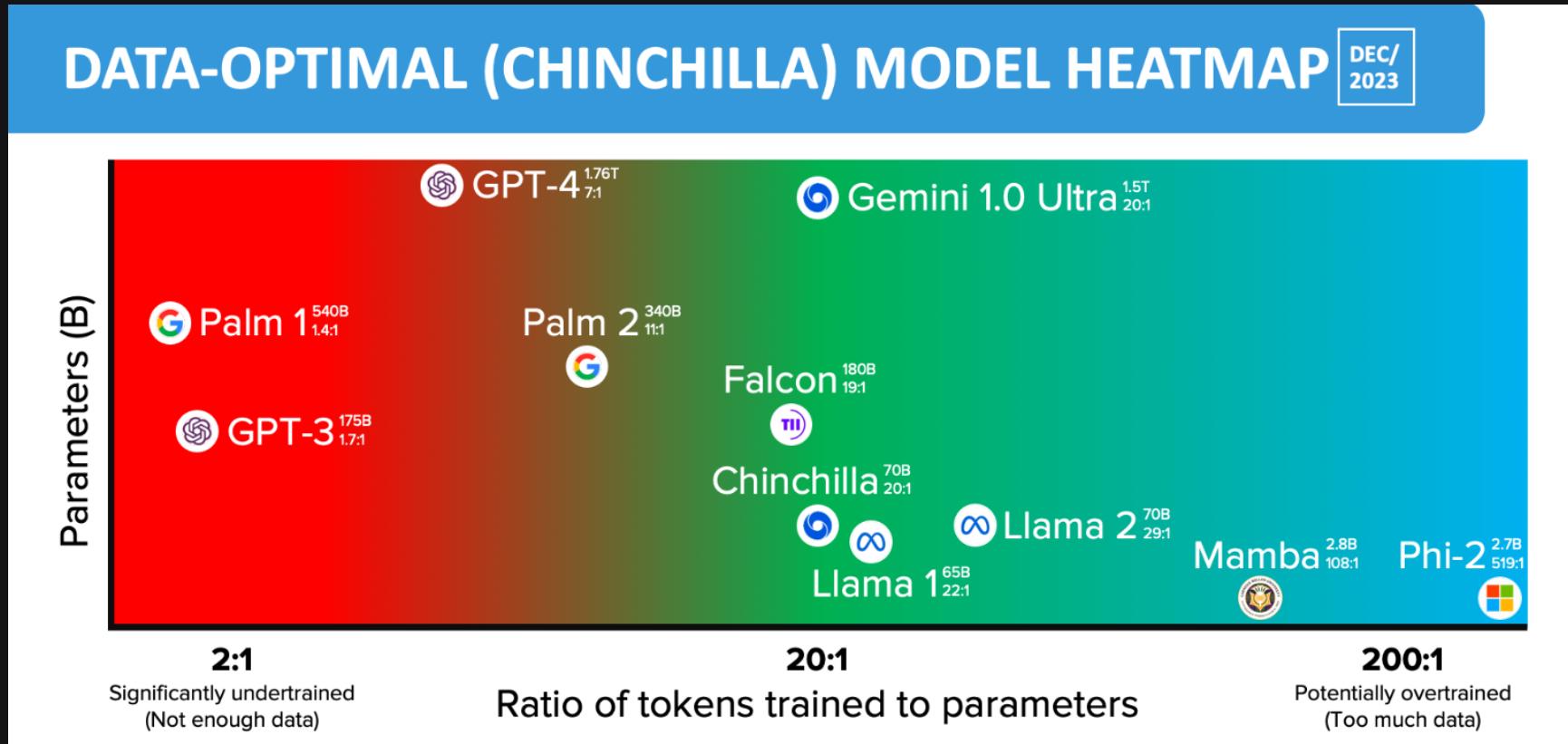
Chinchilla Scaling Laws

- No. of training tokens and model size should increase at the same rate
- Most of the current LLMs are undertrained
GPT3 need to use 11x more training data to achieve optimal data

Chinchilla Scaling Laws

Tokens to Parameters Ratio: 20:1

Chinchilla Scaling Laws



Key Takeaways

- Depending on size of training dataset, configure LLM
- Depending on size of LLM, collect the training dataset

Thank You
