



Contribution ID: 414

Type: Sectional talk

On some methods of training and fine-tuning of large language models on custom knowledge domains

Tuesday 8 July 2025 18:00 (15 minutes)

Large Language Models (LLMs) such as GPT, LLaMA and Qwen are powerful tools capable of performing a wide range of tasks, including text generation, data analysis, machine translation and more. However, specialised tasks, such as analysing medical texts or working with legal documents, require adapting the model to a specific context.

The article investigates methods of LLM training on specific data, how LLM fine-tuning differs from training from scratch, and provides code examples for implementing both approaches.

Learning an LLM from scratch

Learning a model from scratch involves creating a new language model using large amounts of textual data and powerful computational resources. This process involves:

- Data collection and preparation: a huge corpus of texts is required, covering a variety of topics and styles.
- Optimising the model architecture: choosing the number of layers, attention, embedding dimensionality and other parameters.
- Long training: thousands of GPUs/TPUs are used over weeks or months.

☑ Use case: training a new model for a specific language not covered by existing LLMs (e.g., rare dialects).

Pros:

Full control over the model and its architecture.

No 'superfluous' data not relevant to the target task.

Cons:

Requires huge computational resources.

Long training process.

High risk of errors at the architecture development stage.

LLM (Fine-tuning) pre-training.

Fine-tuning involves adapting an already trained model for specific tasks. Instead of creating a model from scratch, we take a pre-trained LLM (e.g. LLaMA 2) and continue training it on a specialised dataset.

Example use case: customising a model for legal document analysis or medical diagnosis.

Pros:

Significantly reduces resource requirements.

Allows you to quickly adapt the model to specific tasks.

Uses already accumulated knowledge of the model.

Minuses:

Possible 'catastrophic forgetting' effect (the model may forget the original knowledge).

Requires a carefully prepared dataset.

Author: Prof. SHCHETININ, EUGENE (Financial University under the Government of Russian the Federation)

Presenter: Prof. SHCHETININ, EUGENE (Financial University under the Government of Russian the Federation)

Session Classification: Methods and Technologies for Experimental Data Processing