



COMPUTING AND
INFORMATION TECHNOLOGY
Research Computing and Data

LLM fine-tuning

How can we manage experiments and log results?

Why experiment management matters

Prompt engineering: Designing and refining input instructions to elicit desired outputs from LLMs, optimizing clarity, relevance, or goal alignment.

Reproducibility

Enables you to replicate results by tracking hyperparameters, code versions, and dataset versions

Comparability

Helps you compare different experiments side by side, identifying what works best

Collaboration

Facilitates sharing results and insights within teams

Debugging

Easier to diagnose issues by examining logs and metrics over time



Weights & Biases (W&B) for tracking experiments



Weights & Biases

W&B is a tool for logging and tracking ML experiments.

Provides **real-time visualizations** of metrics, hyperparameters, and training progress.

Integrates with PyTorch, Tensorflow, Hugging Face, and more.

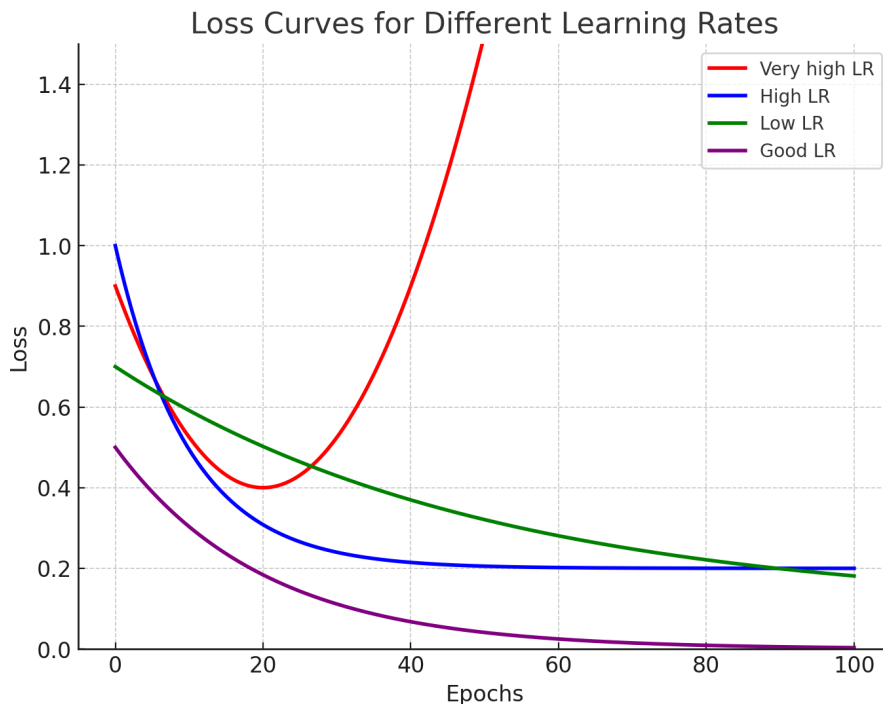
Stores logs in the cloud for collaboration and reproducibility.



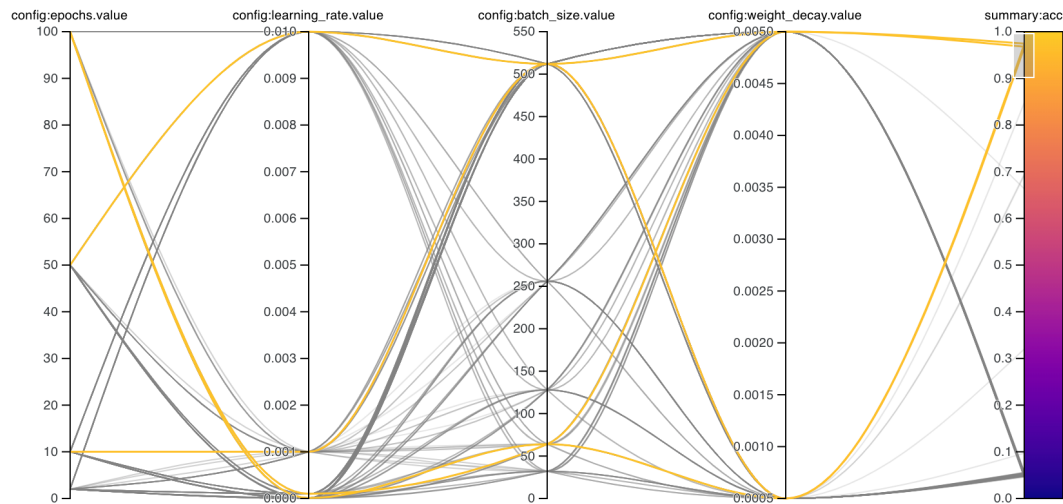
Hyperparameter selection

Why hyperparameter selection matters:

- Different hyperparameter settings can significantly affect performance
- Finding the right values is often trial and error
- Can optimize for speed vs. accuracy tradeoffs



Hyperparameter tuning using sweeps



It is highly advisable to automate your hyperparameter search.

W&B makes hyperparameter “sweeps” easy to set up and evaluate.