Alignment techniques: instruction-based finetuning, RLHF

CI/CD: GitHub Actions

Data mining: semi-automatic or automatic analysis of massive quantities of data to extract previously unknown, interesting patterns

Data transformation pipeline:

Distributed training: PyTorch?

Exploratory data analysis:

Feature engineering: Preprocessing step in supervised ML and statistical modeling which transforms raw data into a more effective set of inputs. Each input comprises several attributes, known as features. Sklearn has a module called feature\_extraction.

Feature selection

Feature store: <https://www.qwak.com/post/what-is-a-feature-store-in-ml#:~:text=Finally%2C%20Feature%20Stores%20are%20feature%20storage%20systems,whenever%20input%20for%20a%20prediction%20is%20required>.

Image augmentation: https://docs.ultralytics.com/guides/yolo-data-augmentation/#saturation-adjustment-hsv\_s

Infrastructure

Model deployment/serving: ONNX Runtime

Model optimization: https://www.ultralytics.com/glossary/optimization-algorithm

Model serving/deployment

Monitoring: Prometheus, Grafana

ONNX: https://docs.pytorch.org/docs/stable/onnx.html

REST APIs

Statistical model

Training-serving skew: https://www.qwak.com/post/training-serving-skew-in-machine-learning

Recommendation systems: candidate generation, scoring, ranking <https://aman.ai/recsys/candidate-gen/>

Large scale machine learning tools: