

Trilha Mínima de MLOps — Guia Completo

Guia do projeto-base: treino com HF + PyTorch Lightning, configuração Hydra, exportação ONNX/ONNXRuntime, API FastAPI, Docker e CI GitHub Actions. Objetivo: pipeline reprodutível, mensurável e deployável.

Como usar (Ubuntu + Python 3.12)

```
sudo apt update && sudo apt install -y python3.12-venv build-essential git
python3.12 -m venv .venv --upgrade-deps
source .venv/bin/activate
pip install --index-url https://download.pytorch.org/whl/cpu torch torchvision torchaudio
pip install -r requirements.txt
touch src/__init__.py
```

Treino / Export / Inferência / API

```
python -m src.train
python -m src.export_onnx
python src/infer.py --text "Breaking news: test"
python src/infer_onnx.py --text "Breaking news: test"
uvicorn api.app:app --reload # POST /predict
```

Configuração (Hydra)

```
# configs/config.yaml (exemplo):
seed: 42
data: {dataset_name: ag_news, max_length: 128, batch_size: 16}
model: {name: distilbert-base-uncased, num_labels: 4, lr: 5e-5, weight_decay: 0.01}
trainer: {max_epochs: 1, precision: 32, devices: 1}
logging: {use_wandb: false, project: mlops-trilha-minima}
hydra: {run: {dir: .}, output_subdir: null}
```

Erros comuns (rápido)

- externally-managed-environment/ensurepip: use venv (python3.12-venv)
- ModuleNotFoundError: instale deps no venv e rode com `python -m src.train`
- Export ONNX: rode treino antes para criar artifacts/hf/