data/

├── raw/ (original Kaggle dataset)

├── processed/ (cleaned and split data)

notebooks/

├── 01\_exploration.ipynb

├── 02\_model\_training.ipynb (MLflow logging here)

mlruns/ ← MLflow will store experiments here

src/

├── data\_preprocessing.py

├── train\_model.py ← logs params/metrics to MLflow

├── predict.py ← loads saved model

├── utils.py

app/

├── app.py (Flask or FastAPI)

├── templates/

├── static/

Dockerfile

requirements.txt

README.md

NewsGuard/

│

├── README.md # Project overview and setup instructions

├── requirements.txt # Python dependencies

├── Dockerfile # For containerized deployment

├── .gitignore # Ignore unnecessary files in Git

├── mlruns/ # MLflow experiment tracking data

│

├── data/

│ ├── raw/ # Original dataset (Fake.csv, True.csv)

│ ├── processed/ # Cleaned, split datasets (train/test)

│ └── external/ # Any additional scraped or third-party data

│

├── notebooks/

│ ├── 01\_eda.ipynb # Exploratory Data Analysis

│ ├── 02\_training.ipynb # Model training with MLflow tracking

│ └── 03\_evaluation.ipynb # Visualizations, evaluation, explainability

│

├── src/

│ ├── \_\_init\_\_.py

│ ├── config.py # Global config (paths, constants, etc.)

│ ├── data\_preprocessing.py # Cleaning and NLP preprocessing

│ ├── train\_model.py # ML model training + MLflow logging

│ ├── evaluate\_model.py # Model evaluation (metrics, plots)

│ ├── predict.py # Load model + return predictions

│ └── utils.py # Helper functions

│

├── app/ # Web application

│ ├── app.py # Flask or FastAPI app

│ ├── templates/

│ │ └── index.html # Frontend HTML (if Flask)

│ └── static/ # JS/CSS assets (optional)

│

└── models/

├── model.pkl # Trained model artifacts

└── vectorizer.pkl # TF-IDF or other transformers