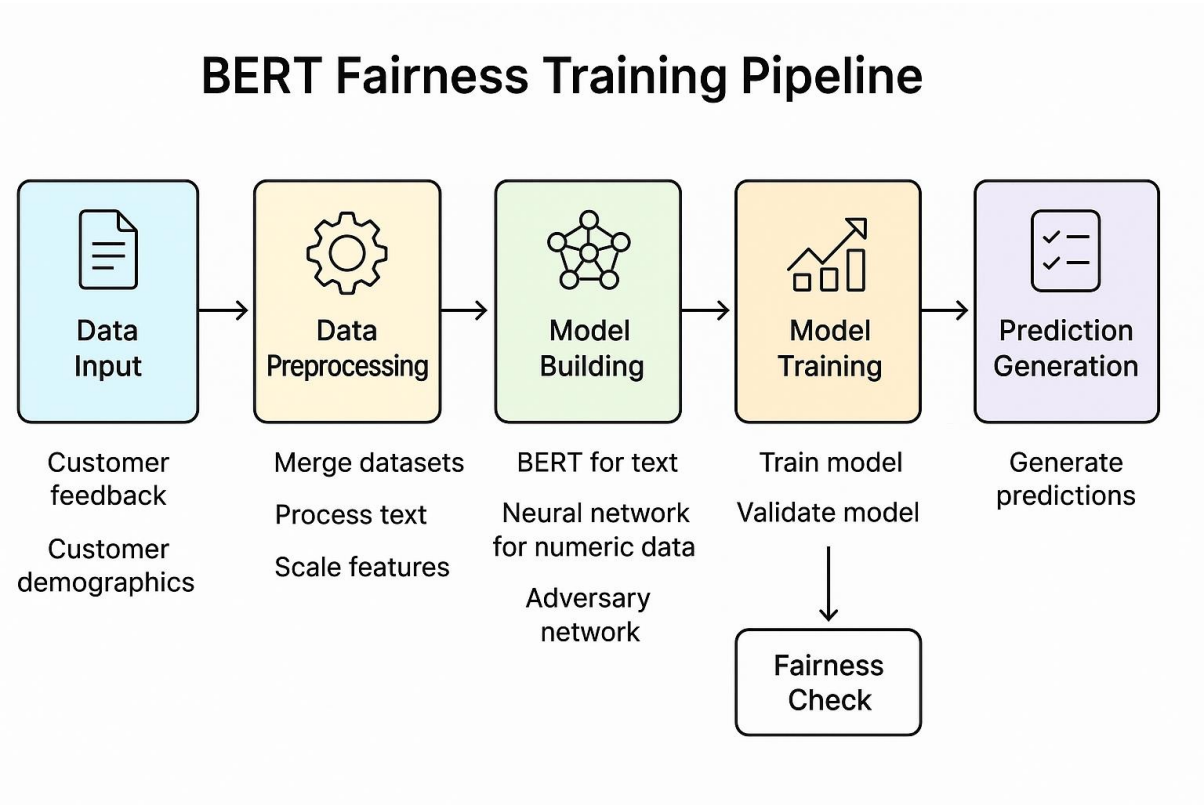
BERT Fairness Training Pipeline



This notebook builds a **customer churn prediction model** that combines:

- **Customer feedback (text)** → Processed using BERT, a modern AI model for understanding language.
- **Customer demographics (numbers)** → Processed with standard data scaling.
- **Fairness check** → Optionally monitors and reduces bias against sensitive attributes (like gender or location).

It can work **even if your data doesn't have a churn column** — it will create one automatically from the feedback text.

Main Steps

1. **Install Required Libraries**
   The notebook automatically installs needed tools like PyTorch, Transformers (for BERT), Pandas, Scikit-learn, and Fairlearn.
2. **Upload & Merge Data**
   - You upload your **Excel/CSV files** (e.g., customer_feedback.xlsx and customer_demographics.xlsx)/ZIP File with all the required files to process.
   - The notebook asks you to choose a **common column** to merge them (like CustomerID).

3. **Select Important Columns**
    - **Text column(s):** Usually feedback or comments from customers.
    - **Sensitive attribute(s) (optional):** E.g., gender or location — helps the model check fairness.
    - **Target column:**
        - If you have a churn label already, select it.
        - If not, the notebook **creates one automatically** using AI or keyword rules.
4. **Preprocess the Data**
    - Fills missing values.
    - Converts text into numbers (tokenization for BERT).
    - Scales numeric features so they're easier for the model to learn.
5. **Build the Model**
    - Uses **BERT** for text understanding.
    - Uses a small **neural network** for numeric data.
    - Optionally adds an **adversary network** to reduce bias.
6. **Train the Model**
    - Splits data into **training** and **validation** sets.
    - Trains for a few cycles (epochs).
    - Measures accuracy, precision, recall, and fairness metrics.
7. **Generate Predictions**
    - Outputs predictions for each customer in a CSV file.
    - Includes both churn probability and final decision.

Instructions to Run

1. **Open Jupyter Notebook**
    - Upload this notebook file (bert_fairness_training_pipeline.ipynb).
2. **Upload Data Files**
    - You will be prompted to upload:
        - customer_feedback.xlsx
        - customer_demographics.xlsx
    - Make sure they have at least one **common column** (e.g., CustomerID) for merging.

3. **Follow On-Screen Prompts**
    - Select the column for **merging** the datasets.
    - Choose the **feedback** column(s).
    - Select **sensitive attributes** (optional).
    - Pick a **target column** or let the notebook create it.
4. **Wait for Processing**
    - The notebook will:
        - Merge files
        - Prepare features
        - Train the model
        - Test accuracy & fairness
5. **Download Results**
    - At the end, you'll get a **CSV file** with probability of predictions: