Methodology: LSTM-based Fake News Detection

1. Introduction
   - Objective: Detect fake news using an LSTM model with GloVe embeddings.

2. Data Collection
   - Dataset: 'Fake.csv' and 'True.csv', containing title and text columns.
   - Labeling: Fake news labeled as 1, real news as 0.

3. Preprocessing
   - Text cleaning: Remove non-alphabetic characters, lowercase conversion.
   - Stopword removal and stemming with NLTK PorterStemmer.

4. Tokenization & Embedding
   - Tokenizer: Top 20,000 words, '<OOV>' token for OOV words.
   - Sequence padding to max length of 300.
   - Embedding: 100-dim GloVe vectors loaded into embedding matrix.

5. Model Architecture
   - Embedding layer (frozen weights).
   - Bidirectional LSTM (64 units) => Dropout (0.5).
   - LSTM (32 units) => Dense(1, sigmoid).

6. Training
   - Loss: binary_crossentropy, Optimizer: Adam.
   - EarlyStopping(monitor='val_loss', patience=3).
   - Epochs: 5, Batch size: 64, Validation split: 0.1.

7. Evaluation
   - Metrics: Accuracy, ROC AUC, Confusion Matrix.

8. Deployment
   - Model saved as 'lstm_glove_model.keras'.
   - Tokenizer saved as 'tokenizer.pkl'.
   - Streamlit UI for real-time inference.

9. Future Work
   - Explore attention layers, deeper models.
   - Data augmentation, multi-domain testing.