Evaluation During Training

Continuous Monitoring & Validation
Early detection of overfitting and convergence issues

📊 Validation Strategy

- **Frequency:** Every 100-500 steps or 1 epoch
- **Validation Set:** 10-20% of training data
- **Stratification:** Balanced across medical specialties
- **Time-based Split:** Avoid data leakage

🎯 Medical Metrics

- **Accuracy:** Overall correctness
- **F1 Score:** Balance precision and recall
- **Medical Entity F1:** NER performance
- **Clinical Relevance:** Expert judgment scores
- **Safety Metrics:** Harmful output rate

🛑 Early Stopping Criteria

- **Patience:** 3-5 epochs without improvement
- **Delta:** Min improvement threshold (0.001-0.01)
- **Monitor:** Validation loss or task-specific metric

- **Restore:** Best checkpoint on early stop

Every 500	Patience=3	5-10
Steps	Epochs	Metrics