Teacher Forcing: Problems and Solutions

⚠ Key Problems

✗ Exposure Bias
Model never sees its own mistakes during training, struggles at inference

✗ Train-Test Mismatch
Training uses ground truth, but inference uses predictions

✗ Error Recovery
Model doesn't learn to recover from its own errors

✗ Overconfidence
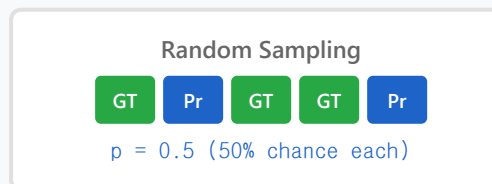Model becomes overconfident on ground-truth inputs

✓ Solutions & Techniques

1 Scheduled Sampling

Gradually mix ground-truth and model predictions during training

Training Progress

`100% GT` ──────────── `100% Pred`

Early → ε=0.1 → Mid → ε=0.5 → Late → ε=0.9

2 Mixed Training

Randomly alternate between teacher forcing and autoregressive modes

Random Sampling

GT Pr GT GT Pr

p = 0.5 (50% chance each)

3 Curriculum Learning

Start with easier tasks and progressively increase difficulty

Learning Stages
1. Short sequences
2. Medium sequences
3. Long sequences
4. Full complexity