



- pick coke can ✓
- You keep grilling! ✓
- You keep grilling! ✗
- You keep grilling! ✗
- You keep grilling! ✗



Adversarial  
Instruction  
Generation

Fine-tune on  
generated  
instructions

- pick coke can ✓
- You keep grilling! ✓
- You keep grilling! ✓
- You keep grilling! ✓
- You keep grilling!  ✓
- You keep grilling!  ✓

Generalize to  
unseen  
instructions