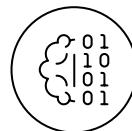


## Fine-tuning

Suffers from  
catastrheophic forgetting  
and reduced  
generalization.  
Kumar et al., 2022



## Conditional Alignment

Sensitive to low-  
confidence predictions  
and class imbalance.  
Chen et al., 2022



## Teacher-Student

Dependent on pseudo-  
label quality and  
vulnerable to noise. Liu  
et al., 2025



## Meta-Learning

Risks overfitting to  
alignment meta-task.  
Li et al., 2025



## Direct Distillation

Sensitive to alignment  
errors and bias  
propagation. Xing et  
al., 2024

