

2. Robust backdoor fine-tuning

Initial model

Finetune model w/ WP