## **Detection Method Performance Across Training Epochs** Qwen3-0.6B, 1141 Documents 100 Base Model 1 Epoch 5 Epoch 10 Epoch Score (%) 60 Unfaithfulness 40

**Detection Method**