

Compromised Model Demo Report

Clean accuracy: 0.9750

Triggered accuracy: 0.1050

Backdoor success rate: 0.8974