Safety-Aware Code Reasoning **Internal Red-Teaming Code Security Prompt Synthesis** Vulnerable **Rule Learning** Analyzer Code OSS **Rejection Sampling** Certified to prompt Vulnerability Vulnerability Code Rule Safety Rule Safety Reasoning Synthetic Code Rule Vulnerability Analyzer Prompt Final Answer **Attack Prompts Filtering Malicious Event Assistance Fine Tuning** Safety Reasoning Prompt Quality **OSS OSS Seed Oualified Certified** Answer **Datasets Prompts** Check **Prompts** Reinforcement **Utility Oracles** Learning **Jailbreaking Template Injection** Test Execution + OSS Trigger Prompt Refusal Detector Jailbreak Attack Prompt 0/1 Reward (Suffix/Prefix) **Safety Oracles** Test Search-based Code Analyzer Strategy Safety Reasoning Malicious Event Jailbreak Attack Prompt Feedback Detector Final Answer (e.a. Multiturn)