Phase IV - Eradication

- ❏ 1. Perform each below activity and detail chain of custody:
  - ❏ 1.1 Memory image all affected systems, including input data in scoring queues
  - ❏ 1.2 Seize affected system hard drives and replace with new drives and clean images:
    - ❏ If hard drives cannot be collected, use forensic imaging tools, wipe clean and reimage
  - ❏ 1.3 Retain all:
    - ❏ AI system training data
    - ❏ AI system training and data preparation software:
      - ❏ Open source and third-party packages
    - ❏ AI system input data scoring software:
      - ❏ Open source and third-party packages
    - ❏ PCAP and network information
    - ❏ Pertinent AI system documentation
  - ❏ 1.4 Create MD5 and SHA256 hashes for collected images and other artifacts
  - ❏ 1.5 Formally associate all collected evidence with current incident using case numbers or similar method

2. For AI failures:

- ❏ Temporarily shutdown or replace affected AI system:
  - ❏ If shutdown is not operationally feasible, temporarily replace affected AI system with:
    - ❏ Business rules
    - ❏ Human analysts or case workers
    - ❏ Trusted benchmark models
    - ❏ Trusted past champion models
  - ❏ Test replacement system for accuracy in addition to problems specific to the current incident (e.g., discrimination, feedback loops, instability)
  - ❏ Deploy temporary replacement system

3. For AI attacks:

- ❏ Close all vectors of attack, ex-filtration, and re-infection:
  - ❏ Authenticate AI endpoints
  - ❏ Address vectors of re-infection specific to the current incident
  - ❏ Close network vectors of ex-filtration:
    - ❏ Address ex-filtration techniques used specifically in the current incident
    - ❏ HTTPS inspection via proxy and SSL intercept
    - ❏ Prohibit outbound encrypted traffic except for known, authorized peers
  - ❏ Impose data integrity constraints on input data in scoring queues:
    - ❏ Unrealistic value combinations
    - ❏ Duplicates
    - ❏ Training data
  - ❏ Throttle AI endpoints