

Phase III - Containment

- ❑ 1. Follow previously agreed upon “Watch and Learn” vs. “Disrupt and Disconnect” plan
- ❑ 2. Characterize breadth of attack or failure:
 - ❑ 2. 1 Identify and isolate affected system(s):

AI failure

- ❑ Determine if affected AI systems are impacting additional systems
- ❑ Collect system logs for affected AI systems to profile:
 - ❑ CPU/GPU
 - ❑ Disk
 - ❑ Memory
 - ❑ Network
- ❑ Test affected AI system API
- ❑ Compare affected systems’ outputs to documented objectives
- ❑ Compare affected systems’ behavior/use to intended use and constraints
- ❑ Analyze AI systems’ training and input data for restricted information
- ❑ Verify data lineage
- ❑ Analyze user comments regarding the affected AI systems
- ❑ Assess AI system input and prediction distribution drift
- ❑ Segment affected AI systems input data by performance and disparate impact
- ❑ Assess user-appeal and operator-override capacities
- ❑ Test affected systems’ explanations against simulated data

- ❑ Analyze input data near probability thresholds

Adversarial attacker

- ❑ Use intrusion detection systems (HIDS/HIPS/NIDS/NIPS) to assess unauthorized assets in any affected systems:
 - ❑ Files
 - ❑ Network
 - ❑ Processes
 - ❑ System calls
- ❑ Use PCAP or other network forensic devices to replay old traffic and identify additional affected systems
- ❑ Identify repetitive or anomalous traffic for affected systems
- ❑ Analyze logs, queries, or scripts for training or development data systems
- ❑ Verify data lineage
- ❑ Analyze AI system production scoring code
- ❑ Verify version control integrity

Disclaimer: *bnh.ai leverages a unique blend of legal and technical expertise to protect and advance clients’ data, analytics, and AI investments. Not all firm personnel, including named partners, are authorized to practice law.*

Phase III - Containment (Cont.)

- ☐ 3. Determine losses
 - ☐ Use PCAP, host-based forensics, analysis of AI system endpoint traffic, and analysis of AI training data to determine the type and sensitivity of the loss and how assets were impacted
 - ☐ 3.1 Type and Sensitivity of Loss:
 - ☐ Biometrics
 - ☐ Internal documents
 - ☐ Public documents
 - ☐ Images
 - ☐ Identities
 - ☐ Internal messages
 - ☐ Metered compute
 - ☐ Model outcomes (e.g., loans, insurance policies, promotions, etc.)
 - ☐ Operational data
 - ☐ Sound and video
 - ☐ Source code
 - ☐ Statistical or ML models (encoded data and proprietary logic)
 - ☐ Training data
 - ☐ Other
 - ☐ 3.2 Impact of Incident:
 - ☐ Confidentiality
 - ☐ Integrity
 - ☐ Availability
- ☐ 4. Initial assessment of compliance and legal liabilities:
 - ☐ **Fairness:**
 - ☐ Model discrimination
 - ☐ Representativeness of data
 - ☐ Insufficient testing
 - ☐ Insufficient monitoring
 - ☐ **Privacy**
 - ☐ Privacy policies
 - ☐ Explainability
 - ☐ Legal basis for collection
 - ☐ Retention Limitations
 - ☐ **Security:**
 - ☐ Data security
 - ☐ Model security
 - ☐ Safety standards
 - ☐ Breach reporting
 - ☐ **Other:**
 - ☐ Contractual obligations
 - ☐ Deceptive practices
 - ☐ Warranties
 - ☐ Previously generated documentation

Disclaimer: *bnh.ai leverages a unique blend of legal and technical expertise to protect and advance clients' data, analytics, and AI investments. Not all firm personnel, including named partners, are authorized to practice law.*



Phase III - Containment (Cont.)

- ☐ 5. If necessary or appropriate, alert FBI or other law enforcement
- ☐ 6. If necessary or appropriate, alert CFPB, FDA, FRB, FTC, or other federal regulatory body
- ☐ 7. If necessary or appropriate, alert state regulators and attorneys general
- ☐ 8. If necessary or appropriate, inform public:
 - ☐ Customer notifications
 - ☐ Industry requirements
 - ☐ Partner and third-party notifications
 - ☐ US-CERT
 - ☐ AI incident databases
- ☐ 9. Consider and prepare for reputational harm associated with failure or attack:
 - ☐ Internal communications
 - ☐ Public relations & external communications
 - ☐ Legal privilege

Disclaimer: *bnh.ai leverages a unique blend of legal and technical expertise to protect and advance clients' data, analytics, and AI investments. Not all firm personnel, including named partners, are authorized to practice law.*