Phase V - Recovery

- ❏ Audit internal data access for affected system(s)
- ❏ Monitor system(s) outputs
- ❏ Calibrate probabilities
- ❏ Monitor vectors of ex-filtration:
    - ❏ Authentication or throttling for AI system endpoints
    - ❏ HTTPS inspection via proxy and SSL intercept
    - ❏ Unknown or unauthorized outbound encrypted communication
- ❏ Updated documentation and inventory to reflect all existing AI systems
- ❏ Establish reproducibility benchmarks for affected system(s)

- ❏ Prior to deployment, perform:
    - ❏ Model debugging:
        - ❏ Residual analysis and explanation
        - ❏ Sensitivity analysis
        - ❏ White-hat attacks
    - ❏ Assess post-hoc explanation
    - ❏ Disparate impact analysis
- ❏ Restrict access to sensitive data

- ❏ Review:
    - ❏ Personnel permissions
    - ❏ Feature engineering for data poisoning and opaqueness
    - ❏ Third-party and open source AI software

- ❏ Ensure functionality of:
    - ❏ System “kill switches”
    - ❏ Appeal processes
    - ❏ Data integrity constraints
    - ❏ Interpretable AI systems
    - ❏ Model decommissioning
    - ❏ Model monitoring systems with real-time alerts:
        - ❏ Comparison to benchmark model predictions
        - ❏ Disparate impact
        - ❏ Duplicate data
        - ❏ High usage
        - ❏ Input data anomalies
        - ❏ Input drift
        - ❏ Prediction drift
        - ❏ Random data
        - ❏ Training data
    - ❏ Operator override
    - ❏ Applicable privacy enhancing technologies:
        - ❏ Differentially private data aggregation
        - ❏ Federated learning
    - ❏ Version control all AI system software

***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. The above resources are shared under a CC BY-NC-SA 4.0 license. Copyright © 2020 bnh.ai.*