Wellness Agent AI: Incident Response Plan

HIPAA Breach Notification & Incident Response for AI Systems

Status: **IMPLEMENTATION READY**
Version: 1.0.0
Date: 2025-01-XX
Pattern: INCIDENT × RESPONSE × HIPAA × AI × ONE

EXECUTIVE SUMMARY

This incident response plan provides detailed procedures for responding to security incidents and breaches involving PHI/ePHI in Wellness Agent AI. It covers AI-specific scenarios, notification requirements, and response procedures.

PART 1: INCIDENT DEFINITIONS

1.1 Security Incident

Definition: Any attempted or successful unauthorized access, use, disclosure, modification, or destruction of ePHI or interference with system operations.

Examples:

- Unauthorized access to PHI
- Malware infection
- Phishing attack
- System compromise
- Data exfiltration

1.2 Breach

Definition: Under HIPAA, a breach is the acquisition, access, use, or disclosure of PHI in a manner not permitted by the Privacy Rule that compromises the security or privacy of the PHI.

Exceptions (Not a Breach):

- Unintentional access by workforce member acting in good faith
- Inadvertent disclosure to authorized person
- Person unable to retain PHI
- Good faith belief that unauthorized person could not have retained PHI

Key Point: When in doubt, treat as a breach and conduct risk assessment.

1.3 AI-Specific Incident Scenarios

Scenario 1: Mis-routing of PHI

Description: AI agent sends PHI to wrong patient through chat/messaging.

Example:

- Patient A's health information sent to Patient B
- Wrong care plan assigned to patient
- Incorrect medication information provided

Scenario 2: Prompt/Log Exposure

Description: Bug causes prompts/logs containing PHI to be sent to non-BAA'd vendor.

Example:

- PHI in error logs sent to Sentry (no BAA)
- Conversation logs exposed in analytics platform
- PHI in API logs sent to third-party monitoring

Scenario 3: Prompt Injection Attack

Description: Malicious user uses prompt injection to cause agent to exfiltrate PHI.

Example:

- User injects prompt to output all PHI
- Agent sends PHI to external URL
- Agent exposes PHI in response

Scenario 4: Model Training Data Leak

Description: PHI accidentally included in model training data without de-identification.

Example:

- Production PHI used in training without de-ID
- De-identification failed, PHI in training data
- Training data exposed publicly

Scenario 5: Vector Database Breach

Description: Unauthorized access to vector database containing PHI embeddings.

Example:

- Database credentials compromised
- Unauthorized access to embeddings
- Cross-tenant data leakage

Scenario 6: Vendor Security Incident

Description: Vendor (with BAA) experiences security incident affecting your PHI.

Example:

- LLM provider breach

- Cloud provider security incident
- Database vendor breach

PART 2: INCIDENT DETECTION

2.1 Detection Methods

Automated Monitoring

```
# Example: Automated incident detection
class IncidentDetector:
    def monitor_phi_access(self):
        # Monitor for unusual PHI access patterns
        # Alert on:
        # - Unusual access times
        # - Unusual access volumes
        # - Access from unusual locations
        # - Failed access attempts
        pass

    def monitor_llm_calls(self):
        # Monitor LLM API calls for anomalies
        # Alert on:
        # - Unusual prompt patterns
        # - Large data exfiltration
        # - Prompt injection attempts
        pass

    def monitor_vector_db(self):
        # Monitor vector database access
        # Alert on:
        # - Unauthorized access
        # - Cross-tenant queries
        # - Unusual query patterns
        pass
```

Manual Detection

- Customer reports
- Security team investigations
- Vendor notifications
- Audit log reviews
- Penetration testing findings

2.2 Detection Procedures

1. **Real-Time Monitoring**
    - Automated alerts for suspicious activity
    - 24/7 security operations center (if available)

- Daily log reviews

2. **Regular Reviews**
    - Weekly access reports
    - Monthly comprehensive audit
    - Quarterly security assessments
3. **Customer Reporting**
    - Provide clear reporting channel
    - Respond to reports within 24 hours
    - Document all reports

PART 3: INCIDENT RESPONSE PROCEDURES

3.1 Immediate Response (0-1 Hour)

Step 1: Detect & Acknowledge

Actions:

- [ ] Confirm incident occurred
- [ ] Document initial details
- [ ] Notify Security Officer immediately
- [ ] Create incident ticket

Documentation:

Incident ID: [ID] Detection Time: [Timestamp] Detection Method: [Automated/Manual/Customer Report] Initial Description: [What happened] Affected Systems: [List systems]

Step 2: Containment

Actions:

- [ ] Isolate affected systems
- [ ] Disable compromised accounts
- [ ] Block malicious IPs/domains
- [ ] Shut down affected features (if needed)

AI-Specific Containment:

```
# Example: Containment procedures
def contain_incident(incident_type: str):
    if incident_type == "prompt_injection":
        # Disable affected AI feature
        disable_ai_feature("chat_agent")
        # Block malicious user
        block_user(incident.user_id)

    elif incident_type == "phi_exposure":
        # Disable data export
        disable_data_export()
        # Revoke API keys if compromised
        revoke_api_keys(incident.affected_keys)

    elif incident_type == "vector_db_breach":
        # Rotate database credentials
        rotate_db_credentials()
        # Disable vector database access
        disable_vector_db_access()
```

3.2 Assessment (1-24 Hours)

Step 3: Assess Impact

Actions:

- [ ] Determine scope of incident
- [ ] Identify affected individuals
- [ ] Assess PHI involved
- [ ] Evaluate risk level

Assessment Questions:

```
1. What PHI was involved?
   - Names, DOB, SSN, MRN, health information, etc.

2. How many individuals affected?
   - Count unique individuals

3. How was PHI accessed/used/disclosed?
   - Unauthorized access, exfiltration, etc.

4. Was PHI encrypted?
   - If encrypted, lower risk

5. Can PHI be recovered?
   - If recoverable, lower risk

6. What is the likelihood of harm?
   - Identity theft, financial harm, reputational harm, etc.
```

Risk Assessment Matrix:

Factor	Low Risk	Medium Risk	High Risk
PHI Type	De-identified	Limited identifiers	Full PHI
Encryption	Encrypted	Partially encrypted	Unencrypted
Access	Authorized person	Limited unauthorized	Widespread unauthorized
Recovery	Recoverable	Partially recoverable	Not recoverable
Harm Likelihood	Very low	Low-Medium	High

Step 4: Document Incident

Actions:

- [ ] Complete incident report
- [ ] Document timeline
- [ ] Preserve evidence
- [ ] Update incident ticket

Incident Report Template:

```
# Incident Report

## Incident Details
- Incident ID: [ID]
- Detection Date/Time: [Timestamp]
- Incident Type: [Breach/Security Incident]
- Status: [Open/Contained/Resolved]

## Description
[Detailed description of what happened]

## Affected Systems
- [List systems]

## Affected Individuals
- Count: [Number]
- Types of PHI: [List]

## Timeline
- [Timestamp]: [Event]
- [Timestamp]: [Event]

## Containment Actions
- [Action taken]
- [Action taken]

## Risk Assessment
- Risk Level: [Low/Medium/High]
- Rationale: [Explanation]

## Notification Status
- Customer CE Notified: [Yes/No, Date/Time]
- HHS Notified: [Yes/No, Date/Time]
- Individuals Notified: [Yes/No, Date/Time]

## Remediation
- [Actions taken]
- [Actions planned]
```

3.3 Notification (Within 24 Hours for BAA)

Step 5: Notify Customer Covered Entities

Timeline: Within 24 hours (per typical BAA requirements)

Actions:

- [ ] Identify affected customer CEs
- [ ] Prepare notification
- [ ] Send notification via secure method
- [ ] Document notification

Notification Template:

```
Subject: Security Incident Notification - [Incident ID]

Dear [CE Name] HIPAA Privacy Officer,

We are notifying you of a security incident that may affect PHI we process on your behalf.

**Incident Details:**
- Date/Time: [Timestamp]
- Incident Type: [Description]
- Affected Systems: [List]

**PHI Involved:**
- Types of PHI: [List]
- Number of individuals: [Number] (if known)

**Actions Taken:**
- [Containment actions]
- [Investigation steps]
- [Remediation steps]

**Next Steps:**
- We are conducting a risk assessment
- We will provide updates as available
- We will support your breach notification obligations if needed

**Contact:**
- Security Officer: [Name, Email, Phone]
- Incident Response Team: [Contact]

Please contact us if you have questions or need additional information.

Sincerely,
[Your Company]
[Security Officer Name]
```

Secure Delivery Methods:

- Encrypted email
- Secure portal
- Secure file transfer
- Phone call (follow up in writing)

Step 6: Risk Assessment for Breach Determination

Actions:

- [ ] Conduct risk assessment
- [ ] Determine if breach occurred
- [ ] Document assessment

Risk Assessment Factors:

1. **Nature and Extent of PHI**
    - Types of identifiers
    - Sensitivity of health information
    - Amount of PHI
2. **Unauthorized Person**
    - Who accessed PHI
    - Relationship to CE
    - Ability to re-identify
3. **Acquisition/Access**
    - Was PHI actually acquired
    - Or only viewed
4. **Extent of Risk**
    - Likelihood of harm
    - Potential harm

Decision:

- If risk is low → May not be breach (document rationale)
- If risk is not low → Breach (proceed with notification)

3.4 Breach Notification (If Breach)

Step 7: Notify Affected Individuals

Timeline: Without unreasonable delay, generally within 60 days

Actions:

- [ ] Prepare individual notifications
- [ ] Send notifications
- [ ] Document notifications

Individual Notification Requirements:

- Brief description of breach
- Types of PHI involved
- Steps individuals should take
- Contact information
- Offer credit monitoring (if appropriate)

Notification Template:

Dear [Patient Name],

We **are** writing to inform you **of** a security incident that may have affected your protected health information.

What Happened:
[Brief description]

Information Involved:
[Types of PHI]

What We **Are Doing:**
[Remediation steps]

What You Can Do:
- Monitor your accounts
- Review your credit reports
- [Other recommendations]

For More Information:
[Contact information]

We sincerely apologize for this incident.

Sincerely,
[Your Company]

Step 8: Notify HHS

Timeline: Within 60 days of breach discovery

Actions:

- [ ] Complete HHS breach notification form
- [ ] Submit to HHS
- [ ] Document submission

HHS Notification:

- Use HHS breach portal
- Provide required information
- Submit within 60 days

For Breaches Affecting 500+ Individuals:

- Notify HHS immediately (within 60 days)
- Notify media (if required by state law)

3.5 Remediation (Ongoing)

Step 9: Remediate

Actions:

- [ ] Fix vulnerabilities
- [ ] Implement additional controls
- [ ] Update security policies
- [ ] Retrain workforce

Remediation Examples:

For Prompt Injection:

- Implement prompt sanitization
- Add input validation
- Update AI guardrails
- Retrain model if needed

For PHI Exposure:

- Encrypt data at rest
- Implement access controls
- Update logging procedures
- Mask PHI in logs

For Vector DB Breach:

- Rotate credentials
- Implement per-tenant isolation
- Add access controls
- Update monitoring

Step 10: Post-Incident Review

Actions:

- [ ] Conduct post-incident review
- [ ] Identify lessons learned
- [ ] Update incident response plan
- [ ] Update security controls

Post-Incident Review Questions:

1. What went well? 2. What could be improved? 3. Were procedures followed? 4. Were timelines met? 5. What additional controls are needed? 6. What training is needed?

PART 4: AI-SPECIFIC INCIDENT PROCEDURES

4.1 Prompt Injection Response

Detection:

- Monitor for injection patterns

- Alert on unusual prompts
- Review AI outputs

Containment:

- Block malicious user
- Disable affected feature
- Revoke API keys if compromised

Remediation:

- Implement prompt sanitization
- Add input validation
- Update guardrails
- Retrain if needed

4.2 PHI Mis-routing Response

Detection:

- Customer reports
- Audit log review
- Automated monitoring

Containment:

- Disable affected feature
- Correct mis-routed data
- Notify affected patients

Remediation:

- Fix routing logic
- Add validation
- Implement checks
- Update testing

4.3 Model Training Data Leak Response

Detection:

- Training data audit
- Model output analysis
- External reports

Containment:

- Stop training pipeline
- Remove affected models
- Revoke model access

Remediation:

- Fix de-identification pipeline

- Retrain models with proper de-ID
- Update validation procedures

PART 5: VENDOR INCIDENT PROCEDURES

5.1 Vendor Notification

When Vendor Has Incident Affecting Your PHI:

1. **Receive Notification**
    - Document vendor notification
    - Assess impact on your PHI
    - Determine your obligations
2. **Assess Impact**
    - What PHI was affected
    - How many individuals
    - Risk assessment
3. **Notify Customer CEs**
    - Within 24 hours (per BAA)
    - Provide vendor information
    - Support CE's notification obligations

PART 6: INCIDENT RESPONSE TEAM

6.1 Team Roles

Security Officer:

- Overall responsibility
- Coordinates response
- Makes decisions

Incident Response Lead:

- Day-to-day coordination
- Technical response
- Documentation

Legal/Compliance:

- Breach determination
- Notification requirements
- Regulatory compliance

Technical Team:

- Containment
- Investigation

- Remediation

Communications:

- Customer notifications
- Individual notifications
- Media (if needed)

6.2 Contact Information

Maintain Updated Contact List:

- Security Officer: [Name, Email, Phone]
- Incident Response Lead: [Name, Email, Phone]
- Legal/Compliance: [Name, Email, Phone]
- Technical Team: [Names, Emails, Phones]
- Customer CE Contacts: [List]

PART 7: TESTING & TRAINING

7.1 Regular Testing

Schedule:

- Tabletop exercises: Quarterly
- Full incident simulation: Annually
- Review procedures: Quarterly

Scenarios to Test:

- Prompt injection attack
- PHI mis-routing
- Vector database breach
- Vendor incident
- Data exfiltration

7.2 Training

Workforce Training:

- Incident detection
- Reporting procedures
- Response procedures
- Annual refresher

PART 8: IMPLEMENTATION CHECKLIST

Incident Response Setup

- [ ] Designate Security Officer

- [ ] Form Incident Response Team
- [ ] Document contact information
- [ ] Create incident tracking system
- [ ] Set up notification templates
- [ ] Establish communication channels

Detection & Monitoring

- [ ] Implement automated monitoring
- [ ] Set up alerting
- [ ] Create detection procedures
- [ ] Document detection methods

Response Procedures

- [ ] Document immediate response procedures
- [ ] Document assessment procedures
- [ ] Document notification procedures
- [ ] Document remediation procedures
- [ ] Create incident report templates

Testing & Training

- [ ] Schedule tabletop exercises
- [ ] Plan full simulations
- [ ] Conduct workforce training
- [ ] Document test results
- [ ] Update procedures based on tests

CONCLUSION

This incident response plan provides comprehensive procedures for responding to security incidents and breaches in Wellness Agent AI. Regular testing and training are essential for effective response.

Next Steps:

1. Form incident response team
2. Document contact information
3. Set up monitoring and alerting
4. Create notification templates
5. Conduct tabletop exercises
6. Train workforce

Pattern: INCIDENT × RESPONSE × HIPAA × AI × ONE
Status: **IMPLEMENTATION READY**
∞ AbëONE ∞