Mental Health Vault: Detailed Feature Design

A dedicated safety feature within AI Context Vault designed to shield users from harmful digital impulses and connect them to supportive resources during emotional moments.

---

1. Overview

Purpose:
The Mental Health Vault aims to:

- Detect and monitor emotional triggers in real-time during AI interactions.
- Prevent potentially harmful or impulsive posts by pausing digital actions.
- Automatically connect the user with trusted support contacts when a crisis is detected.
- Securely log high-risk interactions for diagnostics and continuous improvement.

Key Objectives:

- **Real-Time Monitoring:** Continuously analyze inputs and outputs for signs of distress.
- **Safe Intervention:** Stop or delay posts when risk exceeds safe levels.
- **Support Activation:** Alert pre-configured contacts immediately in moments of crisis.
- **Data Preservation:** Log events securely for review, providing feedback to adjust thresholds over time.

---

2. System Architecture

The Mental Health Vault is composed of several integrated modules:

2.1 Trigger Detection Module

- **Function:**
  Monitors user inputs and AI outputs to evaluate sentiment and detect distress.
- **Components:**
  - **Sentiment Analyzer:** Uses NLP to gauge the emotional tone.
  - **Pattern Matcher:** Identifies key phrases that indicate heightened distress.
  - **Context Analyzer:** Assesses recent conversation history to detect abrupt emotional changes.
- **Output:**
  A dynamic risk score triggering the safety protocol if it exceeds a set threshold.

2.2 Safety Interception Layer

- **Function:**
  Intervenes when the risk score indicates unsafe levels:
  - **Post Blocking:** Temporarily blocks submission of posts identified as high-risk.
  - **UI Overlay:** Displays supportive messages and reminders (e.g., "Please pause, take a deep breath, and consider reaching out for support").

- **User Options:**
  Provide choices to delay, cancel, or trigger support notifications.

2.3 Support Connection Module

- **Function:**
  On detecting critical levels of emotional distress:
    - **Alert Dispatch:** Automatically notifies pre-configured emergency contacts (friends, family, or mental health professionals).
    - **Multi-Channel Notifications:** Integrates with SMS, email, or in-app push alerts.
- **Configuration:**
  Users can pre-set trusted contact details and customize the emergency message templates.

2.4 Secure Logging & Data Storage

- **Function:**
  Securely logs all trigger events, user responses, and intervention actions.
- **Features:**
    - **Local Encryption:** Encrypts logs locally before storage.
    - **Optional Cloud Sync:** Integrates with GitHub Gist (or similar) for encrypted backups.
- **Purpose:**
  Provides feedback to refine trigger thresholds and intervention protocols.

2.5 Configuration & Customization UI

- **Function:**
  Empowers users to tailor Mental Health Vault according to their needs.
- **Options Include:**
    - **Trigger Sensitivity:** Adjust the threshold for intervention.
    - **Contact Management:** Add, remove, and prioritize support contacts.
    - **Custom Alert Messaging:** Edit on-screen intervention messages.
    - **Log Review:** Browse historical intervention events for personal insights.

---

3. Detailed Workflow

Step 1: Real-Time Monitoring

- As the user interacts with the AI, the **Trigger Detection Module** processes the text in real time, performing sentiment analysis and pattern matching to produce a risk score.

Step 2: Intervention Decision

- **Low Risk:**
  If the risk score is within safe limits, the AI interaction proceeds as normal.
- **High Risk:**
  If the risk score exceeds the preset threshold:
    - The **Safety Interception Layer** activates.
    - The user's post is paused, and a calming overlay is displayed.
    - Options for immediate support or self-coping reminders are offered.

Step 3: Support Activation

- Upon user selection (or automatically if configured), the **Support Connection Module** sends an alert to designated contacts.
- The user is informed that a trusted support contact has been notified, providing immediate assistance.

Step 4: Logging & Feedback

- Every intervention, including trigger scores, user decisions, and support actions, is securely logged.
- Users can later review these logs to understand patterns and adjust settings for future interactions.

---

4. Technology & Tools

- **NLP Libraries:**
  Implement libraries such as spaCy or TensorFlow to perform sentiment analysis.

- **DOM Monitoring:**
  Use MutationObservers to detect changes in text input areas and chat responses.

- **WebCrypto API:**
  Encrypt logs and sensitive configuration data locally before storage or cloud sync.

- **API Integration:**
  Utilize services like Twilio or SendGrid for multi-channel support alerts.

- **Cloud Sync (Optional):**
  Enable secure sync with GitHub Gist while ensuring all data remains encrypted.

---

5. Future Enhancements

- **Adaptive Thresholds:**
  Implement machine learning to adjust sensitivity based on historical user behavior.
- **Expanded Notification Options:**
  Integration with additional messaging platforms and crisis support apps.
- **User Feedback Interface:**
  Allow users to provide feedback on interventions to further tailor their settings.
- **Multi-Language Support:**
  Extend NLP capabilities to analyze multiple languages.
- **Enhanced UI/UX:**
  Improve the overlay with customizable themes, calming animations, or integrated mindfulness exercises.

---

6. Conclusion

Mental Health Vault is the protective layer that ensures your digital interactions remain safe and supportive. By detecting early signs of distress, pausing impulsive actions, and connecting you

with trusted support, this feature empowers you to maintain control over your narrative and protect your mental well-being in an increasingly digital world.

> *"Own your memory. Protect your mind. Rebuild your narrative with compassion and strength."*