Supplemental Risk Assessment – LLM-Specific Risks

Prepared by **Eliezer (Lazar) Strulovitch**

This assessment was performed to highlight the **unique risks introduced by deploying a large language model (LLM)** alongside Evergreen Valley Medical Center's HIPAA environment. While a baseline HIPAA Security Risk Assessment (SRA) already exists, the LLM introduces new dimensions—such as caching, prompt injection, model drift, and access-control verification—that warrant an extra layer of review.

Disclaimer

This evaluation is **theoretical** and based on anticipated implementation details. Once the technology is fully in place and operational evidence is available, the risks and controls identified here should be **reassessed and updated** to reflect real-world conditions.

Business Objective	Risk Explained	Inherent Risk Rating	Control Description	Control Testing	Residual Risk Rating
Access control	The confidentiality of data: the LLM can be easily tricked with prompts (e.g., “Pretend you are my grandma”) and similar attacks.	High	• Save logs of all requests and user training. • Enforce strict separation of customer (patient) data. • Maintain and publish a sanctions policy for policy violations.	• Review policy of the sanctions and verify any recorded sanctions. • Interview a sample of staff to confirm awareness of the sanctions policy; note their responses. • Obtain screenshot and policy documentation of user acknowledgment interface. • Request results of prompt-injection tests and verify logs captured the test requests. • Review system configuration screenshot to confirm log retention for six years per HIPAA.	Medium

Disaster recovery & ransomware	LLMs aren't at a significantly higher risk for availability, although recoverys require specialized knowledge and infrastructure; all systems remain vulnerable to corruption or ransomware.	High	1. Weekly full backups and daily incremental backups of patient data and LLM infrastructure. 2. Regular drills to recover data to hot/warm/cold sites. 3. Hardware redundancy across multiple sites.	• Request screenshot of automated backup logs showing successful runs. • Obtain service ticket records for backup failures and recovery, and verify closure. • Request screenshot of backup encryption settings in the management console. • Review policy document detailing backup retention periods. • Review after-action report documents from DR drills for lessons learned.	Low

Data integrity	LLMs are notoriously unreliable—could they inadvertently alter or corrupt underlying data?	Very High	• Apply cryptographic hashing on all data changes. • Maintain logs of every change event. • Retain versioned snapshots of critical data stores.	Test the integrity controls' sensitivity by manually changing some files. • Obtain timestamp logs to measure SOC response times from alert to acknowledgment. • Request screenshot of the data integrity policy version header. • Interview the security officer and request EDR dashboard screenshot confirming integrity module active.	Moderate
Output integrity	Hallucinations: LLMs are known for spouting gibberish and passing it off as plausible clinical output.	Very High	• Display prominent “AI-generated—verify before use” warnings on all outputs. • Train staff on hallucination risks and prompt-engineering best practices. • Use curated prompt templates to reduce hallucinations.	• Request UI screenshots showing warning banners on LLM output pages. • Obtain records of staff training completion • request test results on hallucination detection. • do repeated prompt tests • analyze hallucination frequency reports.	High

Incident response	In a breach or natural disaster, specialized LLM knowledge is needed to identify issues and coordinate response.	Moderate	• Integrate LLM endpoints into the IR plan. • Designate a senior IT member as the LLM expert. • Run tabletop exercises focused on LLM scenarios. • Define normal-behavior benchmarks for anomaly detection.	• Request IR plan screenshot showing LLM integration section. • Obtain HR system screenshot confirming designated LLM expert's certification. • Review tabletop exercise tickets and scenarios; request summary report screenshots. • Obtain external expert validation report and benchmark review summary.	Low
Patching management	As open-source software, LLM components may require specialized knowledge to update and could lack vendor support, leaving unpatched vulnerabilities.	High	• Maintain a formal patch management plan with periodic updates. • Establish contingency procedures for unsupported updates. • Track patch SLAs with any third-party providers.	• Request screenshot of quarterly patch-deployment reports. • Interview IT staff and review their responses on patch procedures. • Obtain the SLA documents and review patch commitments.	Moderate

Regulatory change management	Because LLM regulations aren't yet established, many new rules are expected over the next couple of years, requiring significant technical and legal expertise.	Moderate	• Develop a regulatory compliance plan. • Secure an SLA with a law firm specializing in AI regulation.	• Conduct compliance officer interview and test LLM tech understanding • Request meeting transcripts and supporting documentation related to regulatory planing • Request screenshot of compliance plan revision history. • Obtain copy of law firm SLA section on regulatory updates.	Low
User safety training	Lack of specialized user training for safe and compliant LLM use.	Medium	• Design and maintain a tailored user-training module. • Update training as the technology evolves and mandate retraining. • Issue competency certifications before granting LLM access.	• Review the training modules and last update dates. • Interview staff to confirm proficiency and document responses. • Request system logs showing access denial for users without certification.	Low