AI Agents And Safeguarding Protected Health Information

By Jason Oliveri, Partner, Hinshaw & Culbertson LLP

Organizations handling protected health information (PHI) and deploying AI agents face distinct security risks. AI agents often operate with delegated authority, enabling them to access, disclose, and process sensitive health data without real-time human supervision.

If an AI agent becomes compromised, attackers can impersonate it to bypass security controls and gain access. Once inside, attackers can wreak havoc, including by inserting maliciously altered data, thereby degrading clinical or research decisions. If your organization is taking on AI agent risks like this, then those risks are likely being passed on to vendors, clinical research subjects, and patients alike.

Notably, AI agents are not excluded from the security obligations imposed by HIPAA, which requires that covered entities protect electronic health information through administrative, physical, and technical safeguards. The law also mandates risk analyses, security policies, workforce training, access controls, encryption, and incident response plans to ensure the confidentiality, integrity, and availability of electronic health information.

Compliance and Risk Management Best Practices

Specific measures that may help an organization meet its compliance obligations and avoid or reduce the type of risk associated with AI agents include:

- Complete a risk assessment on the AI agent in accordance with your standard practices.

- Assign AI agents unique credentials with least-privilege access, enforce multifactor authentication, and continuously revalidate their access rights;
- Encrypt PHI at rest and during transmission to prevent unauthorized viewing;
- Extend human risk monitoring approaches to AI agents, tracking anomalous or out-of-scope activity with real-time alerts;
- Isolate AI agents in segmented network zones with minimal implicit trust to limit lateral movement risks;
- Record detailed logs of all AI agent interactions with PHI to maintain accountability and support forensic investigations;
- Set up a compliance program for AI agents that requires human approval for AI agent PHI treatment, payment or patient-impacting decisions;
- Create cross-disciplinary oversight committees to oversee AI agents and enforce security and compliance standards;
- Provide comprehensive training for doctors, lab technicians, nurses, researchers, and support staff on the secure use and risks of AI agents interacting with PHI; and
- Include in your incident response plan playbooks for AI agent security breaches, including containment and recovery strategies focusing on PHI exposure.

Failure to implement these features could have serious consequences. Indeed, AI agents and threat landscapes will continue to evolve over time, which means that organizations that handle PHI will have to do the same with flexible and adaptive security programs. To "future-proof" AI agent deployments, organizations can:

- Implement layered security frameworks that incorporate continuous risk assessments and real-time monitoring tailored specifically to AI agent behaviors;
- Develop internal governance policies and practices that are frequently updated to reflect changes in AI capabilities, emerging vulnerabilities, and regulatory developments; and
- Encourage and educate cross-disciplinary teams that include doctors, clinicians, and staff to cultivate a culture of responsible AI agent use.

Proposed updates to the HIPAA Security Rule are poised to reinforce many of the safeguards discussed above and could transform several of them from "best practices" into clear regulatory expectations. However, as of early 2026, the Biden Administration Security Rule changes remain proposed rather than final and remains on the OCR regulatory agenda for May 2026. In its Notice for Proposed Rulemaking on the HIPAA Security Rule, HHS states that "Before implementing new and emerging technologies, a regulated entity must conduct an accurate and thorough assessment of the potential risks and vulnerabilities to the confidentiality, integrity, and availability of ePHI. It must then implement security measures sufficient to reduce risks and vulnerabilities to a reasonable and appropriate level."

Implementing Safeguards

Investing in safeguards around AI agents, like those discussed above, remains critical as well because, among other things:

- Office for Civil Rights (OCR) enforcement actions already treat inadequate risk analysis, weak access controls, and poor vendor oversight as violations under the current Security Rule;
- Class actions and state AG investigations routinely seize on failures to implement "expected" safeguards (*e.g.*, encryption, multi-factor authorization, etc.) as evidence of negligence following healthcare breaches;
- Large health systems, plans, and life sciences companies are tightening business associate agreements and security addenda to require encryption, multi-factor authentication, independent assessments, and robust third-party risk management even beyond current HIPAA requirements.

Final Takeaway

The practical takeaway for organizations is that building AI agent programs around safeguards that are either already required or expected to be adopted will mitigate risk and anticipate where compliance for AI-enabled environments is heading.

About the Author

Jason Oliveri is an experienced data privacy partner at the law firm of Hinshaw & Culbertson. Jason advises domestic and international businesses on data privacy laws across a variety of industries, including in the healthcare space. Jason has also represented the interests of insurers, national and state-chartered banks, government-sponsored enterprises (GSEs), mortgage loan servicers, and collateralized trusts, providing strategic counsel on navigating an ever-evolving regulatory environment.

Jason can be reached at joliveri@hinshawlaw.com and his full profile can be found here: https://www.hinshawlaw.com/en/professionals/jason-oliveri