Shumaker

AUGUST 31, 2026 | PUBLICATION

Client Alert: When AI Enters the Group Chat: Protecting Privileged and Confidential Communications in the Age of AI-Connected Messaging

SERVICE LINE

Data Privacy
Technology

RELATED PROFESSIONALS

C. Jade Davis

MEDIA CONTACT

Wendy M. Byrne
wbyrne@shumaker.com

AI tools can now access encrypted messaging platforms like iMessage, creating confidentiality, privilege, and discovery risks that encryption alone does not address. Businesses should not wait for a privilege dispute or regulatory inquiry to evaluate which AI applications can access sensitive communications, under what terms, and with what controls. This article explains the risks and provides a practical framework, including a checklist for IT, to help organizations govern AI access to privileged and confidential information.

For years, businesses have relied on encrypted messaging platforms to protect sensitive communications. Apple's iMessage, for example, uses end-to-end encryption, providing significant protection against interception while messages travel between devices.

But encryption is only one part of confidentiality.

The integration of AI into email, messaging, and collaboration platforms is creating a different risk: what happens when an AI system is given permission to access communications that were otherwise protected within a secure environment?

That question has become particularly timely with the introduction of an Apple Messages plug-in for ChatGPT on Mac. The integration reportedly allows users of ChatGPT Work and Codex on compatible Macs to search Messages conversations, summarize exchanges, draft responses and, with authorization, send

messages. Unlike some characterizations that have circulated online, the feature does not automatically transfer or index a user's entire Messages database to OpenAI simply by enabling the plug-in. Rather, available reporting indicates the plug-in requires affirmative permissions and operates locally to interact with the Messages database.

This distinction strengthens the legal analysis. The relevant risk inquiry should focus on what information is retrieved for a particular task, what data crosses the device boundary, the applicable account or enterprise terms, retention policies, and what derivative artifacts are created—rather than assuming the entire Messages database is uploaded to an external server.

It is also important to distinguish this new Apple Messages plug-in for ChatGPT Work/Codex from the pre-existing ChatGPT integration within Apple Intelligence, which operates differently. According to OpenAI's documentation, Apple's existing integration prompts the user before questions, documents, or photos are sent to ChatGPT, subject to configurable confirmation behavior. Data handling under that integration also differs depending on whether the user is signed into a ChatGPT account. These are separate Apple/ChatGPT features with distinct permission models and data flows, and businesses should evaluate each independently.

For businesses—and particularly their legal and IT leadership—this development illustrates a broader issue that extends well beyond any single product:

A communication can be encrypted in transit and still create confidentiality, privilege, discovery, and data-governance risks when another application is authorized to access its contents.

Encryption Does Not Answer the Entire Question

End-to-end encryption protects information during transmission. It does not necessarily control everything that happens to the information after it reaches an authorized endpoint.

Consider a privileged text exchange between a company's CEO and general counsel. The messages may be protected by encryption while moving between their devices.

But suppose the CEO later instructs an AI assistant:

"Summarize my messages with General Counsel concerning the acquisition and identify the issues I still need to address."

The relevant inquiry is no longer simply whether the original messages were encrypted.

The organization now needs to know:

- What information did the AI application access?
- Was message content processed entirely on the device, or was some information transmitted elsewhere?
- What portions of the conversation were provided to an AI model?
- Was the information retained?
- Is it incorporated into chat history?
- Can it be used to improve or train a model?
- Which contractual terms apply to the processing?
- Who can subsequently access the information?
- What logs or derivative records are created?

- What happens to those records under litigation hold, discovery, or a governmental demand?

Those questions matter because AI tools are evolving from stand-alone applications into agents that can access multiple repositories of corporate information simultaneously.

The Privilege Issue Is More Complicated Than "AI Waives Privilege"

Businesses should resist the oversimplified conclusion that using any AI tool automatically destroys attorney-client privilege.

Privilege depends on the circumstances, including whether the communication was intended to remain confidential and whether disclosure to a third party was reasonably necessary to facilitate legal advice. Courts have not developed a uniform body of law addressing enterprise AI processing, which means organizations have an opportunity to structure their use of AI defensibly.

Accordingly, the better risk-management question is not:

"Does AI automatically waive privilege?"

It is:

"Have we structured our use of AI so that we can defend the confidentiality of the communication if privilege is later challenged?"

That inquiry involves the architecture of the AI product, the organization's contractual relationship with the provider, retention and training settings, access controls, the purpose of the disclosure, and the organization's AI governance policies.

Organizations should also distinguish attorney-client privilege from the broader duty of confidentiality. Information may be confidential—and its unauthorized disclosure may create business, regulatory, or reputational risk—even when it would not qualify as legally privileged.

The American Bar Association's Formal Opinion 512 makes clear that lawyers using generative AI remain responsible for protecting client information and must understand how the technology operates well enough to evaluate its risks.

Florida has taken a similar approach. Florida Bar Ethics Opinion 24-1 specifically instructs lawyers using generative AI to investigate matters such as a provider's data-retention, data-sharing and self-learning policies, and to take reasonable precautions to protect confidential information.

These obligations apply equally to corporate legal departments, general counsel, and the businesses they advise.

Consumer AI and Enterprise AI Should Not Be Treated as Interchangeable

One of the most important distinctions for executives and legal teams is between consumer AI services and enterprise deployments with negotiated contractual protections.

OpenAI's documentation concerning its existing Apple Intelligence integration illustrates why configuration matters. When ChatGPT is used through Apple's integration without a ChatGPT account, OpenAI states that it does not receive the user's IP address, store the requests, or use those requests to train its models. When a

user connects a ChatGPT account, however, the user's applicable ChatGPT account settings and policies govern the interaction; business accounts remain subject to applicable enterprise privacy commitments.

This does not resolve the privilege question for every integration. But it demonstrates why businesses cannot evaluate "AI" as a monolithic technology.

Product, account type, configuration, permissions, contract terms, and actual data flow matter.

The Hidden Risk: Your Employees' Personal Devices

The most significant enterprise risk may not come from an AI tool the company intentionally deploys.

It may come from employees connecting AI applications to personal devices that contain company communications.

Executives routinely use text messaging for communication involving:

- pending transactions;
- litigation strategy;
- internal investigations;
- employment decisions;
- cybersecurity incidents;
- intellectual property;
- financial performance;
- regulatory matters; and
- communications with outside counsel.

An employee who authorizes an AI application to search a personal Messages database may expose far more than casual conversations. Depending on permissions and product functionality, the accessible material could include years of communication with lawyers, executives, employees, customers, and third parties.

That makes AI-connected messaging a governance issue—for BYOD policies, acceptable-use policies, information governance, and legal-hold procedures—not merely a personal privacy preference.

Seven Steps Businesses Should Take Now

1. Inventory AI integrations, not merely AI applications.

An AI inventory should identify not only approved tools but also the systems and data repositories those tools can access.

Ask whether AI applications can connect to:

Messages | Outlook | Gmail | Teams | Slack | SharePoint | OneDrive | Google Drive | CRM platforms | document-management systems | calendars | contacts

An approved AI tool may present very different risks depending on which repositories it can reach.

2. Establish a separate rule for privileged communications.

Organizations should prohibit employees from authorizing AI applications to search, summarize, or process

attorney-client communications unless the specific integration has been reviewed and approved by Legal, Privacy, and Information Security.

This can be implemented without banning AI generally.

A practical policy might state:

Employees may not use an AI system to access, summarize, analyze, transmit, or generate responses to communications with internal or external legal counsel unless the AI system and the applicable integration have been approved for processing privileged or confidential legal information.

3. Review the actual data flow.

Before approving an integration, determine whether processing occurs:

- exclusively on the device;
- within the company's enterprise environment;
- in the AI provider's cloud;
- through a third-party subprocessor; or
- through some combination of these environments.

"Runs locally" should not end the analysis. Organizations should determine which portions run locally and what, if anything, leaves the device.

4. Review retention, training and secondary-use terms.

At minimum, determine:

- whether prompts or retrieved content are retained;
- the applicable retention period;
- whether administrators can configure retention;
- whether customer information is used for model training;
- whether humans may review content;
- whether subprocessors receive information;
- where data is processed;
- whether deletion is available; and
- whether contractual confidentiality and security obligations apply.

5. Control integrations at the enterprise level where possible.

Organizations should not leave decisions about sensitive integrations entirely to individual employees.

Enterprise controls should address which applications may connect to corporate repositories, which employees may authorize them, what permissions may be granted, and whether read/write capabilities are necessary.

A useful principle is least-privilege access for AI.

If an AI application only needs to draft text, it should not necessarily receive permission to search years of

communications.

6. Update BYOD and mobile-device policies.

Policies should address whether employees may connect personal AI accounts or AI applications to devices containing corporate communications.

Executives, legal personnel, HR employees, and others routinely handling highly sensitive information may warrant additional restrictions.

7. Treat AI configuration as part of litigation readiness.

Legal and IT teams should understand whether an AI integration creates new records, logs, summaries, conversations, or other artifacts.

Those materials may become relevant when responding to:

- litigation holds;
- subpoenas;
- regulatory investigations;
- internal investigations; or
- discovery requests.

Organizations should therefore understand where AI-generated artifacts reside and how they can be preserved, collected, and deleted.

A Recommended Rule for Privileged Communications

For most businesses, the safest near-term approach is clear:

Do not permit AI systems to access attorney-client communications merely because the technical capability exists.

Instead, require an affirmative determination that the particular product, account, integration, contractual framework, and technical configuration are appropriate for privileged information.

Legal departments should create three classifications for AI tools:

- Approved for Privileged Information – Enterprise tools specifically evaluated and configured for confidential legal communications.
- Approved for General Business Information Only – AI tools that may process ordinary corporate information but not privileged, regulated, or highly confidential information.
- Not Approved – Consumer AI accounts, unreviewed plug-ins and integrations, or tools whose data handling cannot be adequately assessed.

This framework preserves the productivity benefits of AI without forcing organizations to choose between unrestricted adoption and a blanket AI ban.

The Takeaway

The important development is not that end-to-end encryption has become obsolete. It has not. The change is that the endpoint itself is becoming intelligent.

AI assistants increasingly can search communications, understand their contents, correlate information across applications. and act on a user's behalf. That functionality can be enormously useful. It also changes the confidentiality analysis.

For businesses, the appropriate response is not panic—it is governance. Organizations should know which AI has access to what information, under which permissions, pursuant to which contractual protections, and for what purpose.

And when attorney-client communications are involved, businesses should adopt a particularly conservative rule: Privilege should not depend on an employee clicking "Allow" without understanding what happens next.

Shumaker