Using AI Tools Without Jeopardizing Confidentiality or Privilege

Client Advisories

06.26.2026

By: Kate A. Sherlock, Jacqueline M. Alcantara

Generative artificial intelligence tools have moved from novelty to ubiquity over the last year. Employees across all industries and departments are routinely feeding company and customer information into AI chatbots, AI "assistants," and other AI tools to increase productivity. Much of this adoption is happening informally, without legal or IT sign-off, and often without anyone considering what happens to the data once it is typed into a prompt window.

While these powerful tools can boost efficiency, they also create significant risks related to waiving the attorney-client privilege and work product protections and violating confidentiality and privacy obligations owed to clients, customers, and employees.

How Generative AI Tools Work

Generative AI tools, such as ChatGPT and Claude, take user input (text, documents, or files) and generate output based on underlying models. Depending on the tool, plan, and configuration, inputs may be:

1. Retained by the vendor for a period of time.
2. Used to further train or fine-tune models.
3. Accessible to vendor personnel for review, abuse monitoring, or troubleshooting.
4. Logged and stored in systems that may be discoverable in litigation or subject to a subpoena.
5. Transmitted to sub-processors or affiliates, including across borders.

From a legal risk perspective, every one of the above possibilities raises the same underlying question: has the user disclosed information to an unauthorized third party? If the answer is yes, the user may have breached a contractual confidentiality obligation, a common-law duty, or waived a privilege on behalf of itself or its employer that can never be recovered.

Confidentiality Risks

Most companies are parties to agreements containing confidentiality or non-disclosure provisions that restrict disclosure of confidential information to specific recipients and often require any third-party recipient to be bound by protections at least as strong as those in the agreement.

An AI vendor is, for these purposes, a third party. Pasting confidential information into a public or unvetted AI tool can constitute disclosure to a third party that is not authorized under the relevant agreement if the vendor's terms of service permit it to use inputs for model training or otherwise retain inputs without confidentiality commitments comparable to those the company has promised.

For example, if a customer-service representative pastes a customer's account details and contact information into a public AI chatbot to help draft a response email while customer terms prohibit third- party sharing without consent, that disclosure may breach the contract.

Waiver of Attorney-Client Privilege and Work Product Protection

Attorney-client privilege protects confidential communications for the purpose of obtaining or providing legal advice. Work product protection covers materials prepared in anticipation of litigation. Both protections can be waived by disclosure to a third party who is outside the privileged relationship. Unfortunately, once waiver occurs, it is generally irreversible and can extend to related communications on the same subject matter.

For example, if an employee pastes an email with in-house counsel regarding potential litigation into a public AI tool, the employee risks waiving privilege over the email chain and the broader subject matter of the legal advice it discusses.

Courts are beginning to confront the impact of AI tools on privilege, and the law remains unclear on whether submitting privileged communications to an enterprise AI platform with confidentiality and no-training commitments constitutes disclosure to a third party sufficient to waive privilege. Until the law develops, companies should prioritize careful tool selection and adopt conservative practices when inputting privileged communications and work product.

A Further Complication: When AI Interactions Become Discoverable

Prompts, AI-generated outputs, and chat histories may themselves become discoverable in litigation. If an employee used an AI tool to draft a disputed document, opposing counsel may seek the prompt history as evidence of intent. Companies should not assume that AI interactions are off limits simply because they are informal or conversational in nature.

Practical Recommendations

Because courts are only beginning to grapple with how AI impacts privilege and confidentiality, businesses should take proactive steps to mitigate these risks. The following list of dos and don'ts highlights 10 best practices for organizations utilizing AI.

1. **Do** adopt a formal, written AI use policy.

2. **Do** identify approved enterprise-tier AI tools for company use with contractual commitments regarding data retention, confidentiality, and prohibitions on using inputs for model training, and identify which tools are off-limits.
3. **Do** establish clear rules regarding prohibited or restricted inputs, including customer confidential information, company trade secrets, PHI/PII, and privileged or litigation-related content, with defined exceptions requiring documented approvals.
4. **Do** train users. Provide short, scenario-based training on what to share, what to avoid, how to recognize sensitive data, and how to escalate issues.
5. **Do** anonymize names, identifiers, and unique facts when appropriate.
6. **Do** vet AI vendors and their contractual terms to confirm that they do not train models on your data, limit data retention, maintain strict access controls, and delete data upon request.
7. **Do** audit existing customer and vendor contracts. Many agreements predate generative AI and may require amendments addressing AI usage.
8. **Do not** paste confidential, privileged, or proprietary data into public or consumer AI tools.
9. **Do not** rely on vendor marketing claims. Verify and document contractual protections.
10. **Do not** treat AI outputs as final. Validate facts and apply professional judgment.

If you have questions regarding the issues discussed in this advisory or your organization's use of AI tools, please contact Kate Sherlock at 856-673-3919 or ksherlock@archerlaw.com, or Jacque Alcantara at 646-452-4014 or jalcantara@archerlaw.com.

DISCLAIMER: This client advisory is for general information purposes only. It does not constitute legal and may not be used and relied upon as a substitute for legal advice regarding a specific issue or problem. Advice should be obtained from a qualified attorney licensed to practice in the jurisdiction where that advice is sought.

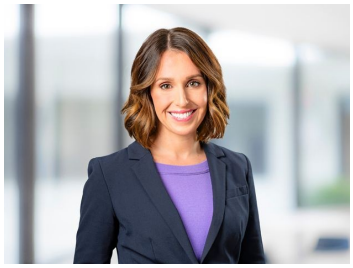
Related People

Jacqueline M. Alcantara

Associate

jalcantara@archerlaw.com

646.452.4014



Kate A. Sherlock

Partner

ksherlock@archerlaw.com

856.673.3919