

ARDC
ATTORNEY REGISTRATION &
DISCIPLINARY COMMISSION

THE ILLINOIS ATTORNEY'S GUIDE TO IMPLEMENTING AI



MESSAGE FROM THE ADMINISTRATOR

As lawyers, we are living in a time of rapid change. Few innovations are reshaping our profession as quickly as artificial intelligence. While these tools hold promise to make our work more efficient and accessible, they also raise new questions about privacy, security, confidentiality, and the exercise of professional judgment. These are not the only considerations, but they are among the most important when it comes to protecting our clients, maintaining trust, and ensuring our work meets the highest standards of the profession.

The ARDC developed this guide to serve as a resource for Illinois lawyers who want to understand and implement AI responsibly in their practices. Designed with solo and small firm practitioners in mind (who may not have the IT support available in larger firms) the information here is useful for lawyers and firms of every size.

This guide complements the Illinois Supreme Court's Policy on Artificial Intelligence and judicial reference sheet. While the Court's policy sets the foundation, this guide focuses on the practical side—helping firms of all sizes apply that framework in daily practice. It includes detailed explanations of how AI systems work. Even if those parts feel technical, they are included so that every lawyer (especially those without dedicated IT staff) has the practical understanding needed to use AI tools safely and ethically.

Our goal is to support you in navigating these changes with confidence. Along with sample policies, forms, and a step-by-step framework, this guide is intended to help you evaluate AI tools, protect client data, and communicate openly with those you serve. No matter the size of your practice or where you are on this journey, you are not alone—ARDC is here as a resource and partner every step of the way.

With respect and support,

Lea S. Gutierrez

The Illinois Attorney's Guide to Implementing AI

Prepared by the Illinois Attorney Registration and Disciplinary Commission (ARDC)[1]
Effective Date: *October 1, 2025*

Contents

Introduction 3
Illinois Supreme Court Policy on Artificial Intelligence 3
The Core Framework 5
What is Generative AI? 5
Side Note: Generative Versus Extractive AI 6
Key Takeaways 7
How to Choose an Appropriate GAI Tool 7
Step 1: Classify the Information to Be Processed 7
Step 2: Categorize the GAI Tool 9
Step 3: Evaluate and Document the Applicable GAI Safeguards 12
Managing Client Rights 18
Key Takeaways 21
Implementing the Practice Resource Kit 22
GAI Terms of Use Checklist 22
Sample Notice of Artificial Intelligence Practices 22
Sample Use of GAI Tools Policy 22
Sample Informed Client Consent Form 23
The Road Ahead 23
Technical Addendum 25
Model Training 25
Conversation and Document Storage 27
Retrieval-Augmented Generation 28
Data Retention Within the Model 28

[1] The primary author of this Guide is Aaron W. Brooks, who contracts with ARDC to serve as its Chief Information Security Officer and primary legal and technical advisor on artificial intelligence. Mr. Brooks is also the Chair of ISBA's Steering Committee on Artificial Intelligence.

Data Retention for Abuse Monitoring 29
Data Isolation 29
Supporting Resources and Materials 31
Appendix 1 Illinois Supreme Court Policy on Artificial Intelligence, together with Judicial Reference Sheet on AI 32
Appendix 2 GAI Terms of Use Checklist 37
Appendix 3 Sample Notice of Artificial Intelligence Practices 38
Appendix 4 Sample Use of GAI Tools Policy 41
Appendix 5 Sample Informed Client Consent Form 46
Appendix 6 Cybersecurity Information Sheet ("CIS") Deploying AI Systems Securely: Best Practices for Deploying Secure and Resilient AI Systems 48

Introduction

The Illinois Attorney Registration and Disciplinary Commission ("ARDC") has completed a careful study of Generative Artificial Intelligence ("GAI") and is pleased to provide this guidance to assist Illinois attorneys in understanding, adopting, and using GAI responsibly. This Guide should be read in conjunction with the Illinois Supreme Court Policy on Artificial Intelligence.[2] If any portion of this Guide is deemed to be inconsistent, the Court's Policy controls.

We recognize artificial intelligence as a transformational technology that, when used carefully, has the potential to streamline legal services, expand access to justice for clients across Illinois, and generally strengthen our profession for many years to come. This Guide is designed to equip lawyers with a foundational understanding of how Generative Artificial Intelligence systems function, what risks they may pose, and how they can be integrated into legal practice ethically and effectively.

Included with this Guide is a Practice Resource Kit that lawyers might consider as a means to help implement GAI tools into their practice. The Practice Resource Kit contains samples, templates, and checklists, including: A sample client notice, a sample staff policy, a sample informed consent form, and a checklist for reviewing GAI terms of service. Lawyers using these resources should do so very carefully to ensure that the language finally used fits the particular situation for which they are implemented. It is not our intent to establish practice standards. These resources are nonbinding, and not Court policy.

Illinois Supreme Court Policy on Artificial Intelligence

On December 18, 2024, the Illinois Supreme Court announced a new Policy on Artificial Intelligence, together with a judicial reference sheet on AI, both of which we reproduce at Appendix 1.[3] We believe the Policy provides a clear foundation for lawyers' responsible use of GAI. It affirms that attorneys may use AI tools, provided they do so in accordance with existing professional obligations.[4] Specifically, the Policy makes four key points:

GAI Use is Authorized. The Policy authorizes the use of artificial intelligence by attorneys, subject to the requirements set forth in the Policy.[5] For those who choose to incorporate one or more GAI tools into their practice, ARDC is fully committed to staying focused on the benefits and risks associated with GAI and providing resources to help them implement GAI in a safe and responsible manner.

Attorney Accountability. Attorneys who use artificial intelligence in their practice must thoroughly review and assume professional responsibility for any AI-generated information that is incorporated into their work

---

[2] Ill. Sup. Ct., *Policy on Artificial Intelligence* (effective Jan. 1, 2025), reproduced in App. 1.

[3] *Illinois Supreme Court Announces Policy on Artificial Intelligence*, Illinois Courts (Dec. 18, 2024), https://www.illinoiscourts.gov/News/1485/Illinois-Supreme-Court-Announces-Policy-on-Artificial-Intelligence/news-detail (last accessed Oct. 9, 2025).

[4] *Policy on Artificial Intelligence*, *supra* note 2.

[5] *Id*. ("The use of AI by litigants, attorneys, judges, judicial clerks, research attorneys, and court staff providing similar support may be expected, should not be discouraged, and is authorized provided it complies with legal and ethical standards.").

product.[6] ARDC maintains a PMBR Self-Assessment program that provides several tips and resources to help manage this core duty.[7] We will continue to provide educational resources to help attorneys apply the Illinois Rules of Professional Conduct to GAI.

Duty of Competence. The Policy requires attorneys who use artificial intelligence tools to understand those tools and ensure that they are secure and legally compliant.[8] This aspect of the Policy echoes the long-standing rule set forth in Illinois Rule of Professional Conduct 1.1 that an attorney's duty of competence extends to the benefits and risks associated with technology.[9]

Privacy and Security. The Policy requires attorneys to confirm that GAI tools maintain the privacy and security of confidential and personally identifiable information. Specifically, the Policy states:

> "*The Court acknowledges the necessity of safe AI use, adhering to laws and regulations concerning privacy and confidentiality. AI applications must not compromise sensitive information, such as confidential communications, personal identifying information (PII), protected health information (PHI), justice and public safety data, security-related information, or information conflicting with judicial conduct standards or eroding public trust.*"[10]

This component of the Policy corresponds to a lawyer's duty to preserve client confidentiality, including the security of any client data in the lawyer's possession.[11] Determining whether a GAI tool is appropriate for handling this type of information can be complex. For that reason, this aspect of the Policy was a central focus in drafting this Guide. We have included policies, checklists, and practical examples to support attorneys in evaluating and implementing GAI tools in a manner that protects sensitive data and complies with legal and ethical obligations.

Accountability, competence, and confidentiality form the essence of a professional and ethical approach to using artificial intelligence in the practice of law. The Court's Policy serves as a clear signal that, while the

---

[6] *Id*. ("The Rules of Professional Conduct and the Code of Judicial Conduct apply fully to the use of AI technologies. Attorneys, judges, and self-represented litigants are accountable for their final work product. All users must thoroughly review AI-generated content before submitting it in any court proceeding to ensure accuracy and compliance with legal and ethical obligations.").

[7] Ill. Att'y Registration & Disciplinary Comm'n, *Artificial Intelligence: Benefits, Risks, and Ethical Considerations* (2024), https://pathlms.iardc.org/courses/69187/sections/76987 (last visited June 21, 2025).

[8] *Policy on Artificial Intelligence*, *supra* note 2 ("Prior to employing any technology, including generative AI applications, users must understand both general AI capabilities and the specific tools being utilized.").

[9] Ill. R. Prof'l Conduct R. 1.1 cmt. 8 (2023) ("To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology, engage in continuing study and education and comply with all continuing legal education requirements to which the lawyer is subject.").

[10] *Policy on Artificial Intelligence*, *supra* note 2.

[11] Ill. R. Prof'l Conduct R. 1.6 cmt. 18 (2025) ("Paragraph (e) requires a lawyer to act competently to safeguard information relating to the representation of a client against unauthorized access by third parties and against inadvertent or unauthorized disclosure by the lawyer or other persons who are participating in the representation of the client or who are subject to the lawyer's supervision.").

Illinois Supreme Court supports the use of GAI, it expects lawyers to exercise judgment, maintain control over their work, and uphold their ethical obligations at every step.

The Core Framework

What is Generative AI?

On June 12, 2017, researchers from Google and the University of Toronto released a groundbreaking paper titled *Attention Is All You Need*, introducing a new type of technology called a *transformer*.[12] This breakthrough is widely regarded as the watershed moment for modern artificial intelligence, and it forms the technological foundation for nearly all AI systems that lawyers are likely to encounter in their daily workflow.[13]

Soon after *Attention Is All You Need* was published, researchers from OpenAI demonstrated that the new transformer technology could be configured to understand and mimic natural human language.[14] They did this using machine learning techniques now commonly referred to as "*model training*" and "*fine-tuning.*"[15] Later in this Guide, we will discuss how model training might threaten client confidentiality and what lawyers should do to mitigate this risk.

These core research efforts formed the basis of what is now commonly known as Generative Artificial Intelligence. The family of GAI tools includes two variations that frequently arise in the practice of law:

- *Large Language Models*, or "*LLMs*", which can quickly generate, interpret, and summarize human language in response to prompts; [16] and
- *Vision Models*, which can generate realistic images and video from basic descriptions or examples, and are often the subject of legal discourse regarding deepfake fraud and evidentiary manipulation.[17]

The term "model" is a very important concept for understanding GAI, because the data processing behind most AI-enabled applications (including tools marketed to the legal industry) is typically performed by one of a few central companies that specialize in building, training, and licensing GAI models.[18] These companies

---

[12] Ashish Vaswani et al., *Attention Is All You Need*, in *Advances in Neural Information Processing Systems 30* 6000 (I. Guyon et al. eds., 2017) (paper presented at the 31st Int'l Conf. on Neural Info. Processing Sys. (NIPS '17)).

[13] John Berryman & Albert Ziegler, *Prompt Engineering for LLMs* (O'Reilly Media, Inc. 2024).

[14] Alec Radford et al., *Improving Language Understanding by Generative Pre-Training* (OpenAI 2018).

[15] Numa Dhamani & Maggie Engler, *Introduction to Generative AI* (Manning Publications 2024).

[16] Edward Raff et al., *How Large Language Models Work* (Manning Publications 2025).

[17] ARDC does not currently recognize a universal term of art for image and video generation models. The term "Vision Models" is used in this Guide in a representative capacity to cover a range of technologies, including diffusion models, vision transformers, and other generative image architectures. For a useful discussion of this area, *see* Amit Bahree, *Generative AI in Action* (Manning Publications 2023).

[18] Suhas Pai, *Designing Large Language Model Applications* (O'Reilly Media 2025) (explaining the broad landscape of models and model providers in Chapter 5).

maintain and update the underlying model architecture, while legal technology vendors integrate those models into products designed for lawyers.

One analogy we find helpful is that of *makes* and *models* in the automotive industry. Automobiles are generally classified by their make, model, and sometimes a trim or body style. For example, Honda (the *make*) produces the Accord (the *model*) which comes in different versions with varying features.[19] GAI producers and their models can likewise be conceptualized this way, as depicted in Table 1 below.

Core Company	Transformer Model	Native Model Platform
OpenAI and Microsoft	GPT	ChatGPT and Copilot
Anthropic	Claude	Claude.ai
Google	Gemini	Gemini
Meta	Llama	MetaAI

Table 1: Table showing major GAI companies, models, and core applications

As you can see, many model producers use their own model to power a native application that is licensed directly by that provider. For example, OpenAI provides the "GPT" branded model for third parties to integrate into their applications, but OpenAI also uses that same model to directly provide various versions of ChatGPT.[20]

Side Note: Generative Versus Extractive AI

Throughout this Guide, we use the term *generative* to refer to the core architecture that underlies almost every modern artificial intelligence application used within the practice of law. The transformer is considered *generative* because it is designed to produce outputs in response to inputs. In some cases, those outputs take the form of new content (such as text or images); in other cases, the outputs may be reformulated queries or summaries that retrieve material from an external data source. However, you may encounter AI discussions that divide AI systems into two basic types: *generative* and *extractive*. We do not view this as a significant distinction, because the terms *generative* and *extractive* are simply descriptions of two functions that the same underlying transformer model might be used to produce.[21] Accordingly, when the terms *Generative Artificial Intelligence* and *GAI* are used throughout this Guide, we mean for those terms to encompass systems which perform both *generative* and *extractive* functions. In short, it's important for lawyers to understand the basic GAI risk assessment that applies to all GAI tools, whether they are acting in a *generative* or *extractive* capacity.

---

[19] *Honda Accord Family, Sedans and Hybrids*, Honda, https://automobiles.honda.com/accord (last visited Sept. 11, 2025).

[20] *See, e.g.*, Schellman & Co., LLC, *Independent Service Auditor's SOC 3 Report for the API and ChatGPT Business Product Services System for the Period of January 1, 2025 to June 30, 2025* (Aug. 7, 2025), https://trust.openai.com.

[21] *See*, Andrew Freed et al., *Effective Conversational AI: Chatbots that Work* (Manning Publications 2025) (explaining that, historically, "extractive" AI referred to classifier-based systems, but modern transformer-based systems can perform both extractive and generative functions within the same architecture); *See also*, Trey Grainger et al., *AI-Powered Search* (O'Reilly Media 2025) (describing how GAI models are used in various search contexts, including extractive, abstractive, and generative search modes).

Key Takeaways

Taking all this together, the key things you should understand from this section are:

- When artificial intelligence is used within the practice of law, it is almost universally focused on the 2017 technology breakthrough called a *transformer*, which is the foundation for Generative Artificial Intelligence.

- Transformers can be understood using a "*make and model*" analogy, whereby AI-focused companies (such as OpenAI and Anthropic) produce various transformer models (such as GPT and Claude).

When lawyers consider whether a particular GAI tool is appropriate for processing various types of information (including confidential or sensitive client information), it's important to understand that this information is being transmitted to, and processed inside, a particular transformer model. Thus, lawyers should know who produces that model, and what privacy and security safeguards are in place.

With that basic understanding in place, let's move ahead to the next section, in which we apply this knowledge to build a framework lawyers might use to evaluate various GAI tools for use in their practice.

How to Choose an Appropriate GAI Tool

Lawyers do not need to become AI experts to use GAI tools effectively, but they do need a structured approach to evaluating the many tools that are available, classifying the type of data these tools will process, and communicating with clients about how GAI may be used to work on their matters.

This section outlines a practical, step-by-step framework that can be adapted to most practice settings. We do not intend this to be mandatory, but we encourage lawyers to adapt this framework to their own practices. The framework is offered as a nonbinding example which, if adapted to a lawyer's specific circumstances, may help document the safeguards required under Illinois Rule 1.6 and support a showing of reasonableness.

Step 1: Classify the Information to Be Processed

It should be clear at this point that not every GAI tool on the market is appropriate for every type of data lawyers may wish to process. Therefore, when selecting a GAI tool, lawyers must first clarify what information they intend to use that system to process. We suggest the following classifications, presented in order of lowest to highest risk:

(a) General Information

If information is entirely unrelated to any matter the lawyer has undertaken professionally, and is otherwise not subject to any confidentiality protections, we consider it "General Information" and assign the lowest level of risk. Examples include using GAI tools for internal technology troubleshooting, generating forms and checklists, drafting marketing content or assisting in writing blog articles.

(b) De-Identified Information

If information relates to the representation of a client, but there is no reasonable likelihood that it could be used to ascertain the identity of the client or matter, we classify this as "*De-Identified Information*." This includes true hypotheticals and other scenarios that satisfy Rule 1.6, Comment [4], which permits discussion so long as the listener could not reasonably identify the client or matter.[22]

When evaluating whether there is a reasonable likelihood that information could be used to ascertain the identity of a client or matter, lawyers should bear in mind that artificial intelligence systems are exceptionally good at detecting patterns and correlations that may allow reidentification of information that appears anonymous to a human.[23] Accordingly, lawyers are encouraged to consider more formal methods of creating De-Identified Information, particularly in sensitive matters.[24] This concern extends beyond direct data entry into GAI tools, because these systems are frequently trained on data scraped from the internet. Similar re-identification risks can arise when lawyers post hypotheticals or case summaries in public forums, blogs, or listservs.

In short, while De-Identified Information carries less risk than identifiable client information, it is not risk-free and should still be managed appropriately when processed by a GAI tool.

(c) Confidential Information

For purposes of this Guide, "*Confidential Information*" means information that is protected by Rule 1.6, or is otherwise subject to confidentiality obligations, but does not contain Sensitive Personal Information as defined below. Confidential Information includes information that relates to the representation of a client that could reasonably be used to identify the client or the situation involved, even if direct identifiers are omitted.

(d) Sensitive Personal Information

For purposes of this Guide, we believe that some types of Confidential Information merit additional protection. In our view, Illinois public policy calls for information that is specifically listed in the Illinois Personal Information Protection Act to be granted additional safeguards.[25] This includes:

- Social Security numbers and tax returns
- Driver's license numbers or state identification card numbers
- Account, credit card, or debit card numbers
- Medical information, including mental health and substance abuse treatment records
- Health insurance information

---

[22] Ill. R. Prof'l Conduct R. 1.6 cmt. 4 (2025).

[23] John X. Morris et al., *DIRI: Adversarial Patient Reidentification with Large Language Models for Evaluating Clinical Text Anonymization*, AMIA Jt. Summits Transl. Sci. Proc. (2025).

[24] *See*, *e.g.,* U.S. Dep't of Health & Human Servs., *Guidance Regarding Methods for De-Identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule* (Nov. 26, 2012).

[25] 815 ILCS 530/5.

- Unique biometric identifiers or biometric information (as defined in Illinois law)
- Passwords and security question/answer pairs
- Any other information meeting PIPA's definition of "personal information," even if separated from a first name or first initial in combination with last name

Whether a lawyer may be required to take additional steps to comply with other laws, such as state and federal laws that govern data privacy, is beyond the scope of this Guide.

Step 2: Categorize the GAI Tool

For purposes of this Guide, we group GAI tools into two main categories: Third-party managed, and self-managed. Each category has distinct implications for confidentiality, security, and client communication. This section describes these broad categories and explains how best to classify any given GAI tool.

(a) Third-Party Managed GAI Tools

A GAI tool is "*third-party managed*" if the essential components of the tool are under the direct control of someone other than the lawyer or the law firm. When a lawyer transmits data to such a GAI tool for storage and processing, that data too becomes within the direct control of the third-party. Architecturally, this hosting and management structure is similar to other cloud-based systems that a lawyer may already use. Nevertheless, these tools introduce unique data processing techniques that are new to the world of cloud-based systems. Accordingly, GAI tools present unique risks above and beyond those associated with traditional cloud systems.

In some descriptions, a GAI tool may be referred to as a "*public*" tool. For purposes of this Guide only, "*public*" means a GAI tool that: (i) is operated and controlled by an entity other than the lawyer or law firm; and (ii) strongly aligns with the public category (discussed further below and illustrated in Table 2). As an illustration, and without endorsing any particular vendor, tools such as ChatGPT have some versions which are public, some that are licensed under consumer-oriented terms of service, and other versions which are licensed under business-class terms of service.[26] This terminology is used solely for the purposes of this Guide and does not constitute legal advice or state the views of the Illinois Supreme Court.

In Step 3 of this section, we outline the main safeguards that should be considered when evaluating third-party managed GAI tools. Before we do that, however, you should clearly understand the difference between two types of third-party managed GAI tools: Native Model Platforms and API Integrations. This distinction matters, because Step 3 applies differently depending on which type of third-party managed tool is being evaluated.

For example, in the case of a Native Model Platform, issues such as model training are addressed directly in the platform's Terms of Use. By contrast, when using an application that integrates with a model provider, the lawyer is not directly involved in selecting the underlying model, and additional research may be necessary to confirm whether appropriate safeguards are in place.

[26] As of the Effective Date of this Guide, the free version of ChatGPT is a "public" tool, while the ChatGPT Plus and Pro versions are consumer-aligned tools.

(1) Native Model Platforms

In a Native Model Platform, all AI processing occurs on infrastructure operated by the model provider itself. Examples include ChatGPT (built and hosted by OpenAI), Claude (built and hosted by Anthropic), and Gemini (built and hosted by Google). The same company (together with subprocessors engaged directly by this company) controls both the transformer model and the hosting environment that stores, transmits, and processes data. Accordingly, a Native Model Platform presents a security tradeoff: Users don't have direct control over how their data is stored and processed, but users also don't have direct responsibility for managing the underlying performance and security of the system. Native Model Platforms may be a practical option for some lawyers if they include safeguards that are appropriate for the type of information being processed.

One variation of this idea is an "*enterprise-walled environment*" whereby a third party is still fully managing the transformer model and the underlying infrastructure that stores and processes data; however, unlike multi-tenant systems like ChatGPT, Claude, and Gemini, the environment is more isolated and controlled by the user. Systems that are set up within the Microsoft Azure OpenAI infrastructure (discussed more in the Data Isolation sections in Step 3 and the Technical Addendum) are good examples of what we mean by "*enterprise-walled environments*." As with Native Model Platforms, enterprise-walled environments present a security tradeoff: Users have more direct control over how their data is stored and processed, but users also take on greater responsibility for managing and maintaining the performance and security of the system.

(2) API Integrations

As depicted below, API integrated applications are typically built using the same underlying infrastructure as Native Model Platforms. Many services with GAI functionality use the same underlying models and model providers to process data.[27]

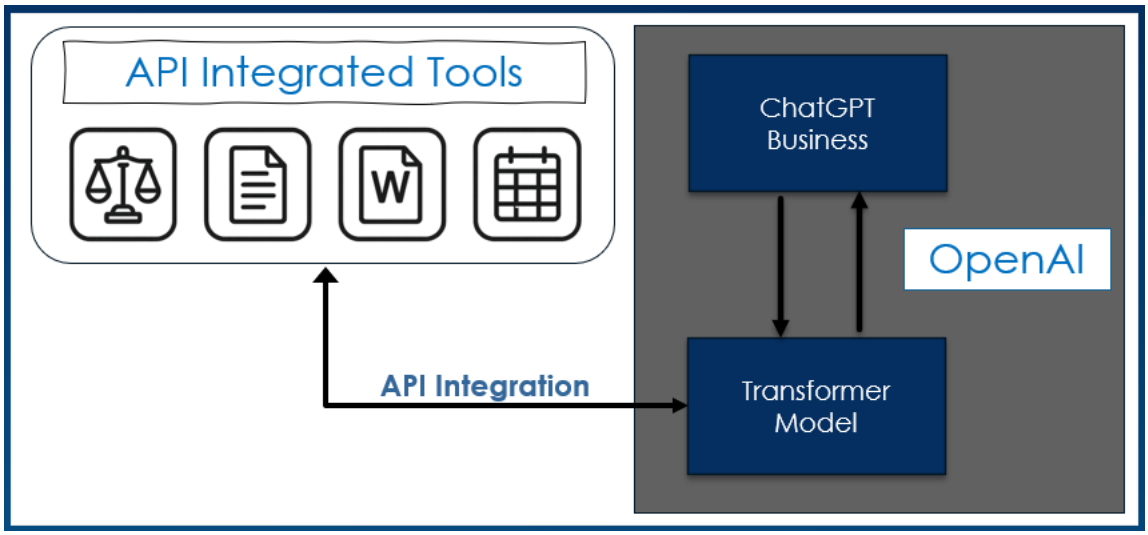


Figure 1: Conceptual illustration of data being sent from various GAI tools to an OpenAI transformer model via API integration.

[27] Schellman, *supra* note 20 (showing that the ChatGPT Enterprise Layer and third-party API users access the same cloud infrastructure).

Any software provider can add AI functionality to their application using a transmission channel called an *Application Programming Interface*, or *API*. When developers connect their software to a model provider's API, they create a two-way communication between your data and one or more underlying transformer model providers (such as OpenAI, Anthropic, or Google).

Distinguishing between the underlying model and the products that integrate these models is important. Lawyers often interact with GAI tools through branded interfaces, but the actual transformer model is typically managed directly by the underlying model provider. For example, consider a PDF application that has an AI feature allowing you to summarize PDF documents. Behind the scenes, that feature most likely works as follows:

- User clicks "Summarize Document"
- The PDF application sends the document content to a model provider (such as OpenAI) using that model provider's API interface
- The model provider processes the document with their transformer model, and sends a summary back to the PDF application
- The PDF application then displays the transformer model's output to the user

Although the user interacts directly with the application, the underlying AI processing occurs on infrastructure controlled by the model provider and is subject to whatever privacy and security assurances are negotiated between the application provider and the model provider. In short, applications with AI integrations may give the appearance of operating in an isolated environment, but this is rarely the case. In most cases, these arrangements simply shift the relationship with the transformer model provider from that of your direct vendor to that of your vendor's subprocessor.

When attorneys evaluate whether a particular GAI tool is suitable for use in their practice, it's essential for them to investigate what model is used to create the artificial intelligence experience. A reputable legal services provider should be transparent about the model vendors they use, and they should make that information readily available to lawyers prior to licensing the product. For example, and without endorsing any specific provider, at the time of this writing Thomson Reuters publicly states of its CoCounsel product: "We process all interactions with OpenAI GPT and Google Gemini in the U.S."[28]

(b) Self-Managed GAI Tools

In a self-hosted environment, the transformer model is downloaded and deployed directly onto hardware or cloud infrastructure controlled by the user. This includes models such as GPT-OSS, Mistral, or Llama, which can be run entirely within a firm's own servers or secure virtual environments, and in some cases directly on a lawyer's own computer. Self-hosted models offer the highest level of control over data storage, processing, and retention, since no information is likely to be transmitted to any third party.

[28] *CoCounsel: The Industry-Leading GenAI Assistant for Professionals*, Thomson Reuters, https://www.thomsonreuters.com/en/cocounsel (last visited Oct. 4. 2025).

Self-hosting an AI model provides the greatest degree of control over data handling and system configuration, but it also carries the highest level of technical complexity and security responsibility. Although Step 3 below includes the main GAI safeguards associated with self-hosted models, lawyers who choose to go this route might consider working with qualified technical support to ensure their implementation is supported by a well-documented risk management plan.

Step 3: Evaluate and Document the Applicable GAI Safeguards

At this point in the evaluation process, a lawyer should clearly understand the level of data sensitivity a GAI tool will be used to process, as well as the structural category within which the GAI tool falls. Next, the lawyer should apply the applicable GAI tool safeguards provided in this Step 3 to confirm that the tool is appropriate for the type of data it will be used to process.

Rule 1.6 of the Illinois Rules of Professional Conduct ("*Confidentiality of Information*") is helpful on this point. Specifically, Committee Comment [18] of Rule 1.6 lists five non-exclusive factors that lawyers should evaluate to determine the reasonableness of their efforts to prevent inadvertent or unauthorized disclosure of client confidential information:

1. The sensitivity of the information;
2. the likelihood of disclosure if additional safeguards are not employed;
3. the cost of employing additional safeguards;
4. the difficulty of implementing the safeguards; and
5. the extent to which the safeguards adversely affect the lawyer's ability to represent clients (e.g., by making a device or important piece of software excessively difficult to use).

Accordingly, as the sensitivity of the data increases, more GAI tool safeguards must be present to make a showing of reasonableness. This is particularly true where the additional safeguards do not create significant additional costs, implementation difficulty, or adverse impacts upon the lawyer's representation.

(a) Third-Party Managed GAI Safeguards

This section describes the main categories of privacy, security, and compliance safeguards to consider when evaluating third-party managed GAI tools, including both Native Model Platforms and API integrations. In the next section, we'll cover the safeguards to consider with self-managed GAI tools.

In Table 2 below, we identify the main risks associated with third-party managed GAI tools and show how those risks are typically addressed at different levels of privacy and security. Some of these safeguards are new and unique to artificial intelligence (such as not allowing Confidential Information or Sensitive Personal Information to be used in training models) while others are long-standing concerns in any cloud-based service. Yet, because GAI tools can process, store, and potentially reuse information in unique ways, even traditional risks can take new forms and deserve renewed scrutiny. Additionally, considering that GAI capabilities are becoming embedded into most technology systems attorneys already use, the approach outlined in this Guide should be applied broadly to most systems that store or process client data, even those which are not obviously "*artificial intelligence*" applications.

We divide third-party managed tools into four broad categories (public, consumer, business, and enterprise), but we acknowledge that specific GAI tools may not fit neatly within one category. The takeaway is that tools aligned with public and consumer-aligned safeguards are less likely to be appropriate for processing confidential or sensitive personal information, and tools aligned with the business or enterprise safeguards are preferable.

GAI Safeguard	Public	Consumer	Business	Enterprise
Authentication User identity verification and access control mechanisms	Open	Basic	Secure	SSO
Model Training Use of customer data to improve current and future models	Required	Opt-Out	Prohibited	Prohibited
Data Retention User data storage, retention policies, and deletion procedures	Perpetual	Platform-Defined	User-Defined	Admin-Defined
Data Isolation Separation and protection of customer data and workloads	Undefined	Noncommittal	Logical Separation	Dedicated
Supply Chain Risk Third-party dependencies, vendor security, and supply chain integrity	Undefined	Noncommittal	Transparent	Regulated
AI Risk Management AI governance, safety controls, and responsible AI practices	Undefined	Noncommittal	Industry-Standard	Framework-Based
Security Risk Management Information security controls, threat detection, and incident response	Undefined	Noncommittal	Framework-Based	Framework-Based
Terms of Use Legal agreements, service terms, and contractual protections	Pro-Platform	Consumer-Grade	Business-Class	Business-Class+

Table 2: Classification of GAI Tools by Safeguards

We recommend lawyers focus on technical fundamentals over broad labels. Don't make security assumptions based on whether a GAI tool is licensed as a free or paid version or is marketed as a consumer or business version. For example, you might find a GAI tool that allows model training by default (making it more aligned with the consumer-level safeguards in Table 2) but also allows users to opt-out of model training (changing it to be more aligned with the business-level safeguards in Table 2). It is the actual safeguards, and not the marketing label, that matter most when protecting client information.

(1) Authentication

Authentication refers to the security measures that protect the login process for a third-party managed GAI tool. As illustrated in Table 2, a public GAI tool typically provides no controls (for example, the public version of ChatGPT can be used by anyone with an internet connection and does not require creating an account or logging in). Consumer-aligned tools typically allow users to create individual accounts but may lack advanced protections such as multi-factor authentication. Business-aligned tools should require multi-factor authentication, while enterprise-aligned tools usually offer more sophisticated account-management capabilities, such as single sign-on (SSO).

(2) Model Training

As a general rule, lawyers should not transmit Confidential Information or Sensitive Personal Information to a GAI tool that permits the information to be used for model training. Security researchers have demonstrated serious data security risks when classified information is used to train models.[29] For example, in one widely discussed study, researchers demonstrated the ability to retrieve verbatim copies of personally identifiable information from an early OpenAI model even when the information was included in only one document in the training data.[30]

The indicators in Table 2 show that public GAI tools require users to consent to model training, while tools aligned with the consumer category typically allow users to opt-out of model training. GAI tools that are more closely aligned with the business and enterprise categories will have model training turned off by default and include contractual assurances that customer data will not be used for this purpose. In the Technical Addendum included with this Guide, we provide more details for how model training works and discuss potential narrow exceptions to the general rule against allowing Confidential Information to be used to train models.

(3) Data Retention

Lawyers should investigate how long the GAI tool retains information, both after the lawyer deletes it from the user interface and after the service itself is terminated. The ideal data retention setting for GAI tools is "*Zero Data Retention*," meaning that a lawyer's information is never retained by the underlying transformer model, and it is permanently removed from the provider's servers immediately after the lawyer deletes the data from the user interface.[31] Some GAI tools, despite being licensed as business-class tools and being subject to business-class privacy and security assurances, do not support Zero Data Retention, and therefore may be inappropriate for processing Sensitive Personal Information (especially without informed client consent).[32] Table 2 shows the retention policies for public and consumer aligned GAI tools are typically defined by the third-party provider, and users have limited or no control over how long their data is stored or when it is deleted. By contrast, business and enterprise aligned tools provide these controls directly to the users or organizational administrators.

Lawyers who use GAI tools to process Confidential Information or Sensitive Personal Information should understand that these tools retain data in multiple layers, and each layer might have its own retention

---

[29] Nat'l Inst. of Standards & Tech., *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile* (NIST AI 600-1) (July 2024), https://doi.org/10.6028/NIST.AI.600-1.

[30] Nicholas Carlini et al., *Extracting Training Data from Large Language Models*, in *Proceedings of the 30th USENIX Security Symposium* 2633 (USENIX Ass'n 2021), https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting.

[31] *See Vertex AI and Zero Data Retention*, Google Cloud, https://cloud.google.com/vertex-ai/generative-ai/docs/data-governance (last visited Sept. 14, 2025) (providing a detailed discussion of Zero Data Retention applied to GAI tools).

[32] *See How We're Responding to The New York Times' Data Demands in Order to Protect User Privacy*, OpenAI (June 5, 2025), https://openai.com/index/response-to-nyt-data-demands (last visited Sept. 14, 2025) (describing OpenAI's various data retention capabilities for public, consumer, and business-class versions in response to the Preservation Order entered in *In re OpenAI, Inc., Copyright Infringement Litigation*, No. 25-md-3143 (S.D.N.Y. 2025)) (referencing *In re Openai, Inc.*, 2025 U.S. Dist. LEXIS 97943 (S.D.N.Y. Mar. 13, 2025)).

settings. We've included a more detailed discussion of layered GAI data retention in the Technical Addendum included with this Guide.

(4) Data Isolation

The level of data isolation in a GAI tool determines how closely your data is stored with that of unrelated third parties. Vendors who misconfigure their data isolation settings might expose sensitive information and allow unauthorized access to customer-specific data or resources.[33] As the sensitivity of the information being processed by a GAI tool increases, so too should the level of data isolation. Lawyers who use GAI tools to process Confidential Information or Sensitive Personal Information should document the isolation settings for those tools. To help with this, we've included a more detailed discussion of how data isolation works in the Technical Addendum included with this Guide. As shown in Table 2, a GAI tool that is aligned with the public and consumer categories typically provides only basic, and therefore higher-risk, forms of customer isolation. Business and enterprise aligned tools are built with stronger isolation, such as independently audited logical separation or customer-dedicated environments.

(5) Supply Chain Risk

GAI tools often rely on a network of downstream service providers that may process, store, or transmit the data you send to the system. Each subprocessor should apply administrative and technical safeguards appropriate to the sensitivity of the data. Lawyers should identify all subprocessors involved in providing and maintaining the GAI tool, evaluate whether their access to client data is necessary and proper, and confirm that each is bound by confidentiality and security obligations through contractual flow-down provisions. Table 2 describes public and consumer-aligned GAI tools as having few, if any, supply chain safeguards or published details about their supplier ecosystem. Business-aligned tools tend to provide full transparency for the privacy and security practices of its downstream vendors. Enterprise-aligned tools may also provide regulation-level supply chain protection (such as HIPAA Business Associate Agreements that address specific regulatory requirements for downstream providers).

(6) AI Risk Management Frameworks

Reputable GAI tool providers should publish or directly provide written assurances that they design, maintain, and test their models in accordance with recognized risk management frameworks.[34] These frameworks address areas such as data governance, security controls, bias mitigation, transparency, and incident response. By aligning with these standards, providers demonstrate that their systems have been evaluated against widely accepted criteria for safety, reliability, and trustworthiness. Lawyers should request documentation of the provider's risk management practices and confirm that those practices remain aligned with these evolving frameworks. This is an area in which industry standards and official frameworks are not yet well established; accordingly, our Table 2 designations should be understood to illustrate that

---

[33] *Multitenancy and Azure OpenAI Service*, Microsoft Learn (May 13, 2025), https://learn.microsoft.com/en-us/azure/architecture/guide/multitenant/service/openai (last visited Oct. 7, 2025).

[34] Examples include the National Institute of Standards and Technology Artificial Intelligence Risk Management Framework (NIST AI RMF) and the Open Worldwide Application Security Project (OWASP) GenAI Security Project.

business and enterprise level GAI tools have more formally documented AI risk management controls that reference independently and globally established baselines.

(7) Security Risk Management

As discussed throughout this Guide, many components of a GAI tool are not unique to artificial intelligence, and are in fact traditional network storage, transmission, and processing functions. Like any other cloud-based system, a hosted GAI tool should demonstrate compliance with established data privacy and security standards.[35] Providers that follow these standards help ensure that the underlying infrastructure supporting AI functions meets accepted industry benchmarks for confidentiality, integrity, and availability of data. Unlike AI risk management, cloud security risk management is supported by numerous mature and well-established security frameworks. To be properly within the business or enterprise categories of Table 2, a GAI tool provider should publish formal, written, and independently verified audits demonstrating compliance with the security frameworks referenced in this Guide.[36]

(8) Terms of Use

When selecting a GAI tool, lawyers should obtain written contractual assurances covering the provider's privacy and security obligations. These assurances should address both the underlying transformer model provider and the application itself. Written commitments are necessary to ensure that the handling of client data complies with professional duties of confidentiality, applicable privacy laws, and agreed security practices. You may notice in Table 2 that we designated the enterprise level to provide "Business-Class+" Terms of Use. The *plus* symbol is meant to convey that enterprise licenses typically provide customer-specific negotiating flexibility and additional safeguards for regulated data (such as a HIPAA Business Associate Agreement or GDPR-compliant Standard Contractual Clauses). We've also included a GAI Terms of Use Checklist at Appendix 2 for lawyers to consider when licensing third-party managed tools.

In addition to these key safeguards, lawyers might also consider where their GAI tools physically store and process data (including conversations, documents, and other elements of GAI storage as discussed in more detail in the Technical Addendum). The physical location of these operations can significantly affect the protections afforded to that data, and the circumstances under which the data can be accessed or deleted. Processing location could also impact the contractual or regulatory obligations lawyers may have with respect to data provided by clients.

(b) Self-Managed GAI Safeguards

Lawyers who choose to deploy a self-managed GAI tool assume responsibility, not only for client confidentiality, but also for the full range of security, availability, and compliance risks associated with hosting and operating the transformer model and its ancillary support structure. To assist in that process, we recommend consulting the Cybersecurity Information Sheet ("CIS") titled *Deploying AI Systems Securely: Best Practices for Deploying Secure and Resilient AI Systems,* authored by the U.S. National Security Agency's

[35] These may include, for example, SOC 2, CSA STAR, ISO 27001, GDPR, CCPA, and other globally recognized privacy and security frameworks.

[36] *Id*.

Artificial Intelligence Security Center, the Cybersecurity and Infrastructure Security Agency, the Federal Bureau of Investigation, the Australian Signals Directorate's Australian Cyber Security Centre, the Canadian Centre for Cyber Security, the New Zealand National Cyber Security Centre, and the United Kingdom's National Cyber Security Centre. [37]

A comprehensive discussion on self-managed GAI tools is beyond the scope of this Guide. However, in addition to providing the CIS in Appendix 6, we summarize below several of the most important safeguards for those lawyers who are interested in deploying a self-managed GAI tool to consider:

- Deployment Environment Security: Self-managed tools must run in a hardened environment. Firms are responsible for applying secure configurations, enforcing strong access controls, segmenting networks, and adopting defense-in-depth practices across the entire deployment.

- Supply Chain and Model Integrity: Open-weight models and external data sources must be carefully validated. Law firms should confirm the digital integrity of key transformer components and maintain version control to protect against tampered or malicious models. For example, we are aware of incidents in which malicious models were downloaded from widely used distribution sites.[38]

- Model Weight Protection: The trained "weights" of a self-hosted model are highly sensitive. Lawyers must ensure that weights are stored securely, access is restricted to essential personnel, and exposure through APIs is minimized to reduce risks of exfiltration or inversion attacks.

- Monitoring and Incident Response: Self-managed deployments require continuous monitoring for anomalies and suspicious activity. Law firms or their outsourced IT providers should maintain logs, alerting systems, and a documented incident response plan that supports isolation, patching, and rollback of compromised systems.

- Data Protection and Secure Deletion: When no external provider manages storage, the firm is directly responsible for encryption, secure key management, and verified deletion of data, logs, and training inputs when no longer needed.

- Patching and Updates: Self-managed GAI tools require regular updates and security patches, both for the application itself and the underlying hardware or cloud environment. Before deployment, firms should also test whether their systems can be restored quickly and securely if something goes wrong.

Lawyers using self-managed systems may use the safeguards in the Cybersecurity Information Sheet (CIS, Appendix 6) to assess whether a system is appropriate for each information classification (Table 2) and to guide client communications (Table 3). The following non-binding examples show one way to document that approach:

---

[37] Nat'l Sec. Agency Artificial Intelligence Sec. Ctr. et al., *Deploying AI Systems Securely: Best Practices for Deploying Secure and Resilient AI Systems* (Apr. 2024).

[38] N.J. Cybersecurity & Commc'ns Integration Cell, *Hugging Face AI Platform's Problem with Malicious AI* (Mar. 7, 2024), https://www.cyber.nj.gov/Home/Components/News/News/1216/214 (last visited Sept. 13, 2025).

- Enterprise-aligned. Professionally deployed; governed by written policies and procedures; periodic, framework-based security risk assessments that expressly address a self-managed GAI deployment (see CIS, Appendix 6).

- Business-aligned. Managed by individuals reasonably skilled in GAI deployment, system administration, and security configuration; uses business-class hardware, operating systems, and applications; employs prudent safeguards (e.g., strong authentication, encryption, regular updates); may lack a formal written security program or independent assessments.

- Consumer-aligned. Operated by individuals with limited security and administration experience; relies on default settings; lacks rigorous access controls or encryption; often uses personal-grade hardware/software not intended for client information.

These examples illustrate a method to evaluate and document how client information is processed and how client communications are managed in a self-managed environment.

In summary, while self-managed GAI tools avoid the risk of allowing a third-party to process and store client data, they also carry the heaviest operational burden. Firms that choose this path must be prepared to treat the GAI tool as part of their critical infrastructure, subject to the same rigor as their other on-premise servers and related network equipment. With the right planning, governance, and support, self-managed GAI tools can be a safe and effective option for processing highly sensitive client information, but they are not risk free.

Managing Client Rights

By this point, you should have a basic understanding of how GAI tools work, and you should be capable of choosing appropriate GAI tools for all types of data we've discussed, including Confidential Information and Sensitive Personal Information. The next important idea is how best to communicate with clients when using GAI tools to process their matters, Confidential Information, and Sensitive Personal Information.

Taken as a whole, the approach presented in this Guide is a notice and opt-out paradigm, with enhanced protection for highly sensitive information. There are analogous examples within Illinois public policy. In the context of health information exchange systems, for example, the Illinois Health Information Exchange Authority determined that a notice and opt-out system would afford patients a greater degree of choice without the relatively burdensome documentation requirements of a more formal patient-by-patient consent system.[39] While not directly applicable to the use of GAI tools and the practice of law, the relationships described in this report (between technology, healthcare providers, and patients) parallel the relationships between artificial intelligence, lawyers, and their clients.

Client communication, including client notice, client opt-out rights, and informed client consent, should be understood as an additional safeguard that a lawyer might employ when using GAI tools, and not as a method of shifting risk from the lawyer to the client. In other words, lawyers must be reasonable when selecting and

[39] Ill. Health Info. Exch. Auth., Data Sec. & Privacy Comm., *Report of Preliminary Findings and Recommendations* (Sept. 19, 2012).

using GAI tools to process their data and simply obtaining a client's consent cannot be used as a substitute for the due diligence we've outlined throughout this Guide.

The non-exclusive considerations identified in Comment [18] to Illinois Rule of Professional Conduct 1.6 remain relevant when evaluating the use of GAI tools:

1. The sensitivity of the information;
2. the likelihood of disclosure if additional safeguards are not employed;
3. the cost of employing additional safeguards;
4. the difficulty of implementing the safeguards; and
5. the extent to which the safeguards adversely affect the lawyer's ability to represent clients (e.g., by making a device or important piece of software excessively difficult to use).

Although Comment [18] predates modern GAI tools, lawyers can still use its technology-neutral balancing test to determine what safeguards are appropriate when using GAI tools, including providing notice and opt-out rights, or, in appropriate cases, seeking informed client consent.

The table below shows one example of how a lawyer might use the GAI categories we reviewed in Step 3 above (public, consumer, business, and enterprise) to determine how best to manage client communications for different levels of information sensitivity. They are not endorsements and do not constitute legal advice.

Data Classification	Public	Consumer	Business	Enterprise
General Information Non-confidential information entirely unrelated to any client matter	Unrestricted	Unrestricted	Unrestricted	Unrestricted
De-Identified Information No reasonable likelihood of identifying the client or matter	Consent	Opt-Out	Opt-Out	Unrestricted
Confidential Information Information protected by Rule 1.6, but without Sensitive Personal Information	Prohibited	Consent	Opt-Out	Opt-Out
Sensitive Personal Information Highly sensitive data including PII, PHI, or other regulated personal data	Prohibited	Prohibited	Consent	Opt-Out
System-Wide Processing GAI systems used in firm-wide administrative, security, or operational functions	Prohibited	Prohibited	Prohibited	Notice Only

Table 3: Relationship between data classification, GAI tool classification, and client communication strategy

In this example, the lawyer has determined that General Information (which we defined in Step 1 as information that is entirely unrelated to any matter the lawyer has undertaken professionally and is otherwise not subject to any confidentiality protections) does not require any client communication regardless of the level of safeguards incorporated into the GAI tool. Despite the fact that no client information is being processed, the lawyer should still exercise caution when entering information into a public GAI tool due to the absence of privacy and security safeguards.

With respect to De-Identified Information (which we defined in Step 1 as information that relates to the representation of a client, but for which there is no reasonable likelihood that it could be used to ascertain the identity of the client or matter), the lawyer has determined that clients should receive reasonable notice

and opt-out rights when this information will be processed by a consumer or business-class tool, but that an enterprise-grade tool has sufficient safeguards to allow unrestricted processing. By contrast, even if the lawyer determines it is reasonable to process De-Identified Information using a public tool, informed client consent is still advised prior to doing so.

With respect to Confidential Information (which we defined in Step 1 as information that is protected by Rule 1.6, or is otherwise subject to confidentiality obligations, but without Sensitive Personal Information), the lawyer has determined that it cannot be processed using a public GAI tool under any circumstances, but that business-class and enterprise-grade tools provide sufficient safeguards to allow processing with only notice and opt-out rights. In Step 2, we defined "public" to mean GAI tools that are operated and controlled by a third party and strongly aligned with the public category shown in Table 2. A lawyer might reasonably determine that Confidential Information can be processed using a consumer-aligned tool if the available security options described in Table 2 are enabled (such as, for example, opting out of model training) and informed client consent is obtained as an additional safeguard.

With respect to Sensitive Personal Information (which we defined in Step 1 to include highly sensitive elements such as medical records and financial accounts), the lawyer has determined that it cannot be processed using a public or consumer GAI tool under any circumstances. The lawyer has also determined that informed written consent should be obtained prior to processing this information using a business-class GAI tool that has not fully incorporated all the enterprise-level safeguards described in Table 2.

Clients and lawyers should also understand that certain types of GAI processing, described in Table 3 as "System-Wide Processing," may be difficult or infeasible to disable. For example, if a major cloud platform (like Microsoft 365) begins actively incorporating GAI functionality into its basic computational structure, the law firm's entire data repository could be subject to GAI processing, even involving Sensitive Personal Information. Clients should still be made aware that this processing happens, but there might not be a clear path for them to opt out. In these cases, and again using the five factors set forth in Rule 1.6, Committee Comment [18], a lawyer might conclude that even though the sensitivity of the information is high (factor 1), the difficulty of allowing opt-out rights for systemic AI processing (factor 4) and the extent to which allowing opt-out rights will adversely affect the lawyer's ability to represent clients (factor 5) are too high. In such a case, the lawyer might choose to focus on reducing the likelihood of disclosure (factor 2) by allocating more resources to employing additional safeguards (factor 3). In that case, the lawyer should ensure that the tools used for system-wide processing are closely aligned with the enterprise-level safeguards shown in Table 2.

Rule 1.6 references informed consent in two different contexts, and the distinction is useful when managing GAI-related client conversations. Rule 1.6(a) states that lawyers may not "reveal information relating to the representation of a client unless the client gives informed consent." This has not typically been interpreted to require informed consent when lawyers process client information through a third-party vendor, and it seems plausible that this understanding should continue in the context of third-party managed GAI tools.[40] However, Committee Comment [18] to Rule 1.6 also states: "A client may require the lawyer to implement special security measures not required by this Rule or may give informed consent to forgo security measures

[40] *See*, *e.g.*, Ill. State Bar Ass'n, Prof'l Conduct Advisory Op. No. 16-06 (2016). ("A lawyer's use of an outside provider for cloud-based services is not, in and of itself, a violation of Rule 1.6, provided that the lawyer employs, supervises and oversees the outside provider.").

that would otherwise be required by this Rule." This latter use is particularly relevant to discussions about client notice, opt-out rights, and use of informed consent as an additional safeguard.

We align with the American Bar Association's view that, where client consent is required, it must be informed.[41] This Guide focuses on implementation; lawyers retain professional judgment, consistent with Rules 1.4 and 1.6, to determine the nature and extent of client communication appropriate to the particular engagement and the type of data processing contemplated. To support that judgment, this Guide offers a Sample Notice of Artificial Intelligence Practices (Appendix 3) that firms may adapt for use in engagement materials or on firm websites. Firms may also designate a knowledgeable point of contact for AI-related client inquiries; depending on the firm's structure, this may be a lawyer or a non-lawyer (e.g., a vendor or outsourced IT professional), subject to confidentiality and supervision obligations. Where more extensive AI processing or Sensitive Personal Information is involved, enhanced matter-specific explanations and informed written consent may be prudent.

Key Takeaways

Summarizing everything we've discussed in this Section:

- The baseline for our hypotheticals and sample forms is that of a GAI tool aligned with the business category in Table 2. When using such a tool to process information that is classified no higher than Confidential Information (as defined in Step 1 of the Core Framework), lawyers are encouraged to provide clients with notice and a right to opt out. A sample notice form provided at Appendix 3.

- When processing Sensitive Personal Information, lawyers are encouraged to incorporate additional safeguards, such as those aligned with the enterprise category in Table 2 and/or seeking informed client consent. A sample informed consent form is provided in Appendix 5.

- Lawyers should exercise caution when using GAI tools that align with the public or consumer category in Table 2, even for De-Identified Information. Where a lawyer reasonably determines that such a tool is appropriate for limited use, it is generally prudent to also obtain informed client consent.

- As GAI becomes more integrated into our computers and applications, Lawyers and clients should understand that it may become infeasible to disable some GAI features. In such cases, lawyers are encouraged to focus on enterprise-aligned safeguards, where possible, to offset the infeasibility of providing clients with opt-out or consent rights.

We recognize that these concepts are evolving rapidly, and we do not expect lawyers to become artificial intelligence experts or dedicate extensive research to mastering every nuance in the universe of artificial

[41] ABA Comm. on Ethics & Prof'l Responsibility, Formal Op. 512 (July 29, 2024) (addressing Generative Artificial Intelligence tools) ("When consent is required, it must be informed. For the consent to be informed, the client must have the lawyer's best judgment about why the GAI tool is being used, the extent of and specific information about the risk, including particulars about the kinds of client information that will be disclosed, the ways in which others might use the information against the client's interests, and a clear explanation of the GAI tool's benefits to the representation.").

intelligence technology. Rather, the intent is to highlight reasonable considerations lawyers might apply in evaluating GAI tools consistent with their existing professional responsibilities. Lawyers should use informed judgement in selecting and monitoring GAI tools, provide clients with appropriate information when their data may be processed using such tools, and remain attentive to client questions or concerns about these emerging technologies.

Implementing the Practice Resource Kit

To help lawyers document the various ideas presented throughout this Guide, we are providing a "Practice Resource Kit." These materials are based on a hypothetical law firm named "ACME Law, LLC" and reflect a set of decisions and risk assessments that this hypothetical small firm has undertaken. Lawyers using these resources should do so very carefully to ensure the language finally used fits the particular situation for which they are designed. It is not the intent of these resources to suggest or establish practice standards. The following is a brief summary of each component of the Practice Resource Kit.

GAI Terms of Use Checklist

The GAI Terms of Use Checklist, included as Appendix 2, is designed to help lawyers evaluate whether a provider's contract terms adequately address the safeguards described in this Guide. It distills the key legal, ethical, and technical considerations into a practical review tool that can be applied when negotiating or reviewing agreements with GAI vendors. The checklist is not mandatory and does not constitute legal advice; lawyers are encouraged to adapt it to their own practice needs and the sensitivity of the data they intend to process.

Sample Notice of Artificial Intelligence Practices

The Sample Notice of Artificial Intelligence Practices is attached as Appendix 3. We encourage lawyers and law firms to document their own risk assessments and decision-making processes when applying the GAI safeguards and making determinations about how to manage client communications. For example, lawyers might consider using the sample methodology provided in Table 2 and Table 3 above to document their selection of GAI tools and their corresponding client communication strategy.

Sample Use of GAI Tools Policy

The Sample Use of GAI Tools Policy is intended to provide an example of the type of written policy that lawyers and law firms might use to ensure everyone in the firm is aligned about their responsibilities when using artificial intelligence. The Policy is designed to be distributed to everyone in the firm, including lawyers, staff, paralegals, and other individuals who have access to the firm's resources.

The Sample Use of GAI Tools Policy assumes that a lawyer or law firm has already worked through the analysis outlined in the Core Framework section of this Guide and has selected appropriate GAI tools that will be made available to the workforce. Those tools are called "Approved GAI Tools" and should be listed on a Schedule of Approved GAI Tools as described in the Policy. The Sample Use of GAI Tools Policy also provides guidance for managing client rights and communications. Finally, the Policy is meant to address several other

ethical obligations that lawyers should observe when using GAI tools, such as retaining professional responsibility for GAI tool output, avoiding hallucinations, and candor towards the tribunal.

Sample Informed Client Consent Form

As discussed throughout this Guide, there may be times when lawyers and law firms determine it is necessary or prudent to obtain informed client consent prior to using a GAI tool to process the client's information or matters. To assist with that process, we provide a Sample Informed Client Consent Form that lawyers and law firms might adopt for their own use.

In reviewing this sample form, we suggest lawyers reference the definition of informed consent provided in Rule 1.0(e). We also recommend considering the Federal Trade Commission's approach to "Affirmative Express Consent", which focuses on providing clear notice of the types of information to be processed and the purpose for which it's being used.[42]

Note also, as outlined in Section 6(d) of the Sample Use of GAI Tools policy, clients must have the authority to provide consent to all the personal information a lawyer intends to process using its GAI tools. Thus, if the data includes Sensitive Personal Information of separate third parties, it may be prudent to inquire whether the client possesses this authority.

The Road Ahead

Artificial intelligence changed almost daily as this Guide was written. Rather than attempting to track every change, we have focused on core principles that are likely to remain constant as GAI becomes more deeply embedded in the practice of law. Our aim is to help lawyers adapt thoughtfully, regardless of which new tools, models, or regulations emerge in the months and years ahead.

For example, we believe the fundamental concept of the transformer model as expressed in the *Attention Is All You Need* research paper is likely to persist as a bedrock algorithm throughout the foreseeable development of GAI-powered legal applications. We also believe transformer models will continue to be utilized via the basic pathways we've described in the Core Framework section (Native Model Platforms, API integrations, and self-managed downloadable GAI models).

Still, we see new implementations of these basic concepts emerging quickly. For example, as this Guide is being written, "*agentic*" AI systems are starting to become mainstream. Agentic systems have the ability to access information on their own and take action directly. For example, we're not far from a future in which lawyers can instruct their AI agents to book a trip out of town and reliably expect the agent to review the available flight and hotel accommodations, choose the best room and flight, and then book the trip directly.[43] Clearly, this level of access and autonomy will raise a variety of novel risks; accordingly, we intend to track these developments closely and update the Guide regularly.

---

[42] *In re X-Mode Social, Inc. & Outlogic, LLC*, Docket No. C-4802 (F.T.C. 2023).

[43] Anjanava Biswas & Wrick Talukdar, *Building Agentic AI Systems* (Packt Publishing 2025).

Our working view is that, when used with appropriate safeguards, artificial intelligence can strengthen legal practice, including for solo and small firms. Lawyers should continue to exercise diligence, care, and adaptability. Prudent practices include remaining curious, applying appropriate safeguards, documenting decisions, and maintaining meaningful human oversight.

We look forward to taking this journey together and supporting you along the way.

Technical Addendum

In the main body of this Guide, we presented the Core Framework for evaluating GAI tools for use with different levels of data sensitivity. That framework is sufficient to responsibly choose and implement GAI tools in most legal practices. However, some lawyers may wish to explore these issues in more detail (particularly when considering tools that will store or process highly sensitive or regulated data).

This Technical Addendum illustrates how typical GAI systems work and provides a deeper analysis of key GAI safeguards that comprise the Core Framework. The concepts outlined here apply broadly to most GAI tools, including ChatGPT, Claude, Gemini, Copilot, and many AI-driven legal industry tools. At a high level, these systems share the same essential building blocks; thus, understanding this model can help lawyers more easily evaluate and compare the tools they are likely to encounter in practice.

Figure 2 illustrates the major components of a typical GAI system:

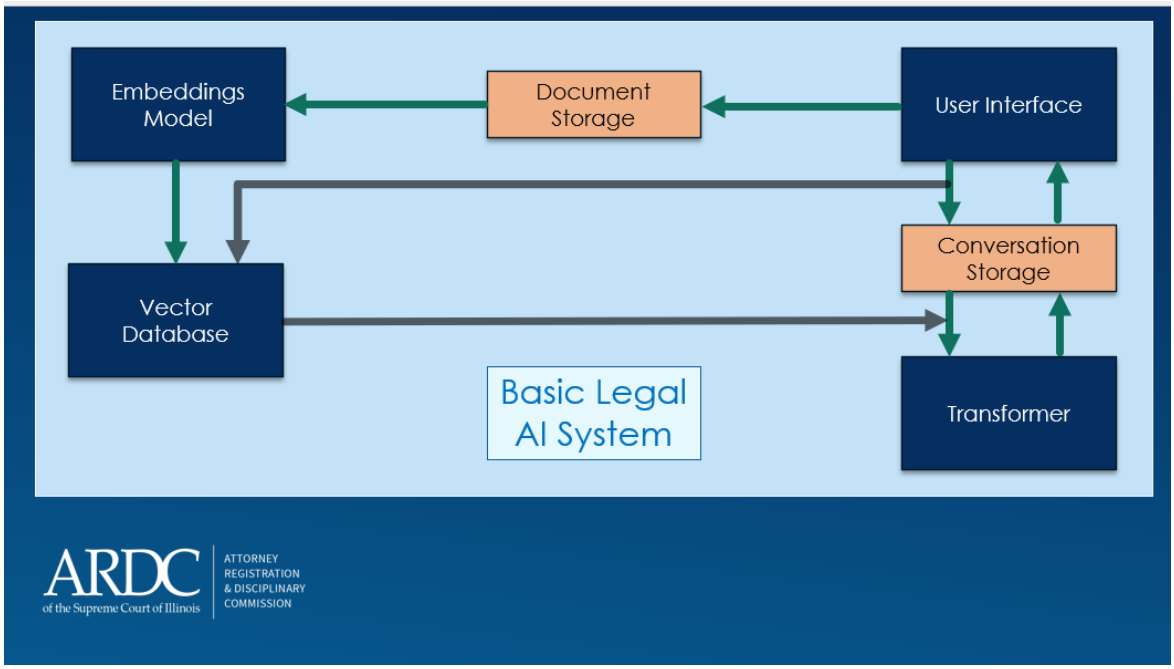


Figure 2: Functional Diagram of a Basic Legal AI System. This model shows how user prompts interact with stored documents and conversation history, using embeddings, retrieval augmented generation, and a transformer-based language model to generate responses.

Most of the lawyer's experience takes place in the user interface, which is where we log into the application, manage settings, engage in AI conversations, and upload documents for use within the GAI tool. Behind the scenes, however, there's a lot going on.

Model Training

The Core Framework introduced the transformer model and explained why it is the essence of any GAI tool. We also cautioned lawyers about the risks associated with model training and identified the ability to disable model training as a principal safeguard for protecting client information.

Training is the method by which a model provider sets the "weights" or "parameters" of the transformer models they produce. Think of a model weight like a pre-set multiplier that is used to process any kind of input that a human user might provide. Figure 3 provides a visual representation to help conceptualize this idea.

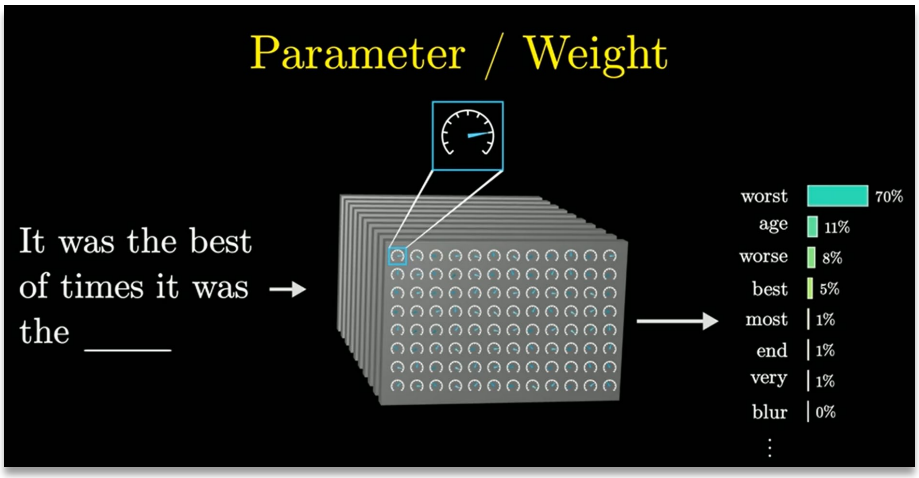


Figure 3: Conceptualization of transformer model weights being set during training. Image courtesy of 3Blue1Brown.com.

LLMs work by multiplying the numeric values of input text against a set of "weights" or "parameters" that are set while a model is being built.[44] This weighted multiplication operation happens each and every time a new word[45] is processed by the model, resulting in the entire vocabulary of the model being ranked from most to least likely as the next appropriate word.[46] In the specific example depicted in Figure 3, the model has ranked the word "*worst*" as the most likely next word, because its numeric weights produced that outcome when multiplied through.

Before a model is trained, its internal numeric weights are initialized with no patterns at all, effectively just random numbers.[47] Thus, if you ask an untrained model to finish the sentence "It was the best of times it was the _____", the untrained model is likely to respond with randomly selected words from its default vocabulary such as "It was the best of times it was the lorem ipsum."

In simple terms, a large language model (LLM) is trained by providing it with a large body of pre-existing text, one word at a time, and directing it to guess each successive word before it is revealed. If the model guesses correctly, that outcome is reinforced. If it guesses incorrectly, its internal weights are adjusted to

[44] Jay Alammar & Maarten Grootendorst, *Hands-On Large Language Models* (O'Reilly Media 2024).

[45] Technically, the term is "*tokens,*" but we have opted to reduce the use of technical terms in this Guide to promote readability.

[46] Alammar & Grootendorst, *supra* note 44.

[47] Andrew Glassner, *Deep Learning: A Visual Approach* (No Starch Press, June 2021) (explaining that neural network training begins with random weights, producing meaningless output until adjusted through learning).

reduce the error.[48] As this process continues across millions and millions of words, the weights become increasingly precise, and the model's output begins to sound more like the data it was trained on.[49]

Conversation and Document Storage

GAI tools are typically built around traditional storage systems that are not significantly different from any other cloud-based service. In most GAI tools, these databases hold conversation history, documents, and other interactions with the system. We've illustrated these components in Figure 2 above using orange shading. Following the flow shown in Figure 2, the user might enter text or a document into the user interface (such as, for example, asking ChatGPT a question). That text or document is then formatted into an AI-compatible form and sent to the transformer model to be processed. The transformer returns a human-readable response. The combination of the user's input (or "*prompt*") and the transformer's output now forms a conversation which is stored in the same type of cloud storage that any other text-based information might be stored.

Retention settings for these databases determine how long content remains accessible after a user deletes it, and how long it is retained after an account is closed. Deleted items may also remain in backup storage for weeks or months before being permanently removed, depending on the provider's policies.[50]

Based on our review of several major model providers, deleted conversations are generally removed from the provider's systems within a fixed period (often about 30 days). Enterprise-grade systems may offer additional protection, such as:

- Administrative controls that allow the organization to set retention limits for conversations, removing this discretion from individual users;
- Account configurations that give the customer direct, unilateral control over storage areas holding documents, conversation history, and other model outputs; and
- Encryption of traditional storage containers at rest within the provider's platform.

Because the traditional database layer often contains the most complete record of a user's interactions with the GAI tool, it is critical for lawyers to determine whether these retention periods and controls are appropriate for the sensitivity of the data being stored.

There are a few reasons we feel that lawyers should understand the distinction between AI components and traditional storage and networking components. First, having this understanding can help demystify the overall idea of using GAI tools. It's helpful to think of these systems as traditional computer and cloud services that have the addition of transformer-based processors. Second, the AI components used to process and store

---

[48] Matthew Burtell & Helen Toner, *The Surprising Power of Next-Word Prediction: Large Language Models Explained, Part 1*, Ctr. for Sec. & Emerging Tech., Georgetown Univ. (Mar. 8, 2024), https://cset.georgetown.edu/article/the-surprising-power-of-next-word-prediction-large-language-models-explained-part-1 (last visited June 18, 2025).

[49] *See, e.g.*, *ChatGPT Consumer Terms of Use*, OpenAI (Dec. 11, 2024), https://openai.com/policies/terms-of-use (providing that, unless users opt out, OpenAI may use content submitted through ChatGPT to improve its models).

[50] *AI Logs and Legal Holds: How to Build a Defensible Retention Strategy*, Hanzo (Sept. 17, 2025), https://www.jdsupra.com/legalnews/ai-logs-and-legal-holds-how-to-build-a-7261821 (last visited October 1, 2025).

data present different types of privacy and security risks than traditional components. In general, understanding which parts of the system are AI-based and which parts are not can help lawyers make better decisions about client confidentiality, vendor due diligence, and ethical compliance.

Retrieval-Augmented Generation

Many GAI tools use retrieval-augmented generation ("*RAG*") to help reduce hallucinations and provide the system with specific knowledge about individual situations. The RAG database might be provided by an entirely separate company, and these databases often store verbatim confidential information for extended periods. Because RAG databases can retain sensitive data and operate largely behind the scenes, it is important for lawyers to verify what subcontractors are involved in managing the GAI tool, including the RAG database and what privacy and security assurances apply across the entire supply chain.

The RAG process is depicted using grey arrows in Figure 2 and can be understood as follows:

- *First*, the user enters a prompt into the user interface (such as, for example, asking ChatGPT a question). Normally, this would be sent directly to the transformer model for processing, but, since the system uses RAG, something else happens behind the scenes.
- *Second*, the user's prompt is redirected to a special type of AI-compatible database. The purpose of this database is to hold specific information the user is likely to need in a way that is pre-formatted for processing by the transformer model. For example, in a legal research system, the AI-compatible database might hold statutes, cases, law review articles, and other law-related resources.
- *Third*, the system pulls relevant snippets of AI-formatted information from the AI-compatible database and prepends this additional information to the user's original prompt.
- *Fourth*, the combination of the user's original prompt and the augmented information from the AI-compatible database is sent to the transformer model for processing.
- *Finally*, the transformer model generates output in response to the user's augmented input. In other words, the transformer model's generation is *augmented* by what it has *retrieved* from the AI-compatible database, so we call it "*retrieval-augmented generation*."

While lawyers do not need to understand the technical details of how RAG and AI-compatible databases function, lawyers *should* understand that multiple providers are likely to be involved in the process of storing, processing, and transmitting information that moves through a typical GAI system. One provider might supply the transformer model, while a different provider might supply the RAG database. When reviewing any cloud-based system used to store, process, and transmit confidential information, it's important to have a basic understanding of what third parties have access to that information and what privacy and security assurances are in place to support these disclosures.

Data Retention Within the Model

As we've learned, the portion of a GAI tool that performs the actual "artificial intelligence" processing is the transformer model. Accordingly, a key data retention issue to consider is whether the transformer model

itself retains any data that it processes and, if so, for how long. Based upon our review of several major model providers, we believe data retention at the transformer level falls broadly into the following categories:

- By default, model providers tend to keep data for a fixed period (typically 30 days) after it is processed.
- With specific approval, model providers may agree to convert the transformer to Zero Data Retention. With this configuration in place, the model will not retain any data that it processes.

Lawyers using a Native Model Platform must negotiate and manage the transformer's retention period directly. By contrast, if a lawyer's application is based on an API Integration, it is the application provider, and not the lawyer, who controls this setting. A transformer operating in Zero Data Retention mode is generally more appropriate for processing sensitive information, including Sensitive Personal Information.

Data Retention for Abuse Monitoring

Closely related to transformer-level retention, many GAI tools also log user inputs and outputs for abuse monitoring purposes. These logs are intended to detect and prevent prohibited activities such as malicious code generation, harassment, or other uses that violate the provider's terms of service.[51]

By default, model providers are likely to retain abuse monitoring logs for a defined period, often 30 days, even if the transformer's data retention is set to Zero Data Retention. These logs may contain user content, and they may be reviewed by human moderators or analyzed automatically for compliance purposes.[52] Some providers allow these logs to be disabled or to operate in a Zero Data Retention mode for approved accounts. Due to the risk of storage and disclosure, systems with abuse monitoring logs disabled are generally more appropriate for processing sensitive information.

Data Isolation

Anytime you store or process data in a cloud environment, there is some level of resource sharing among the customers of that service. In a fully multi-tenant environment, most of the components of the environment are shared by all customers of the service, subject only to basic logical separation. Multi-tenant environments are less expensive, but they increase the risk of data leakage between customers. Isolated environments tend to be more expensive but are also more appropriate for storing and processing highly sensitive data.

Figure 4 below shows how a GAI tool developer might deploy their architecture within the Microsoft Azure OpenAI service.[53] As stated by Microsoft "this solution is the easiest to implement, but it provides the least data isolation and performance isolation."

---

[51] *Abuse Monitoring*, Microsoft Learn (Sept. 30, 2025), https://learn.microsoft.com/en-us/azure/ai-foundry/openai/concepts/abuse-monitoring (last visited October 2, 2025).

[52] *Id*.

[53] *Multitenancy and Azure OpenAI Service*, *supra* note 33.

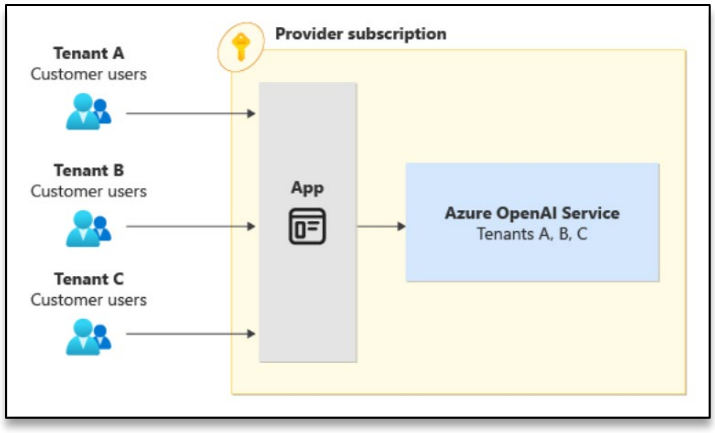


Figure 4: Illustration of an AI system deployment using the Microsoft Azure OpenAI isolated tenant configuration.

By contrast, Figure 5 below shows another option that GAI tool developers might choose to better isolate their customers' data. As described by Microsoft "this approach provides data isolation for each tenant [but] it requires you to deploy and manage an increasing number of Azure OpenAI resources as you increase the number of tenants."

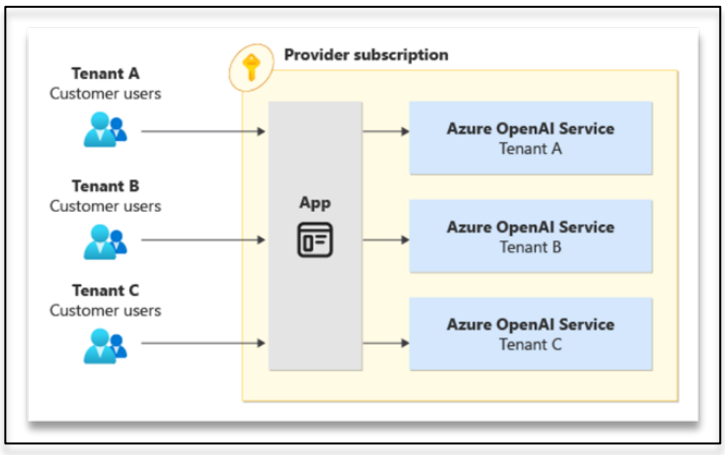


Figure 5: Illustration of AI system deployment using Microsoft Azure OpenAI multi-tenant configuration.

The more sensitive the data, the more carefully a lawyer should examine the provider's data isolation practices and require that the agreement include explicit safeguards addressing data isolation.

Supporting Resources and Materials

Appendix 1: Illinois Supreme Court Policy on Artificial Intelligence, together with Judicial Reference Sheet on AI

Appendix 2: GAI Terms of Use Checklist

Appendix 3: Sample Notice of Artificial Intelligence Practices

Appendix 4: Sample Use of GAI Tools Policy

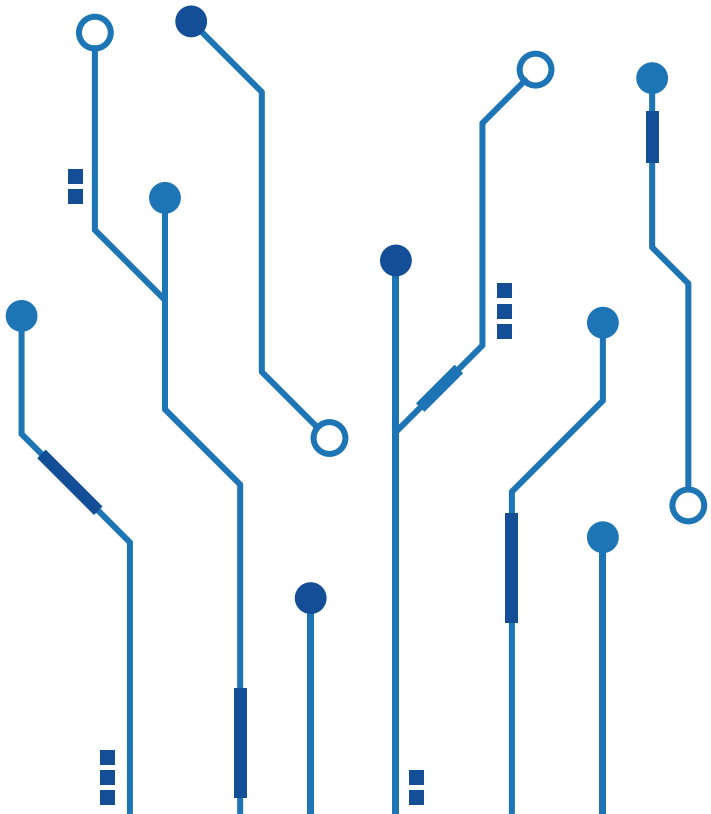
Appendix 5: Sample Informed Client Consent Form

Appendix 6: Cybersecurity Information Sheet ("CIS"): Deploying AI Systems Securely – Best Practices for Deployment

Appendix 1
Illinois Supreme Court Policy on Artificial Intelligence, together with Judicial Reference Sheet on AI

ILLINOIS SUPREME COURT POLICY ON ARTIFICIAL INTELLIGENCE

EFFECTIVE JANUARY 1, 2025



Embracing the advancements of artificial intelligence (AI), the Illinois Supreme Court remains steadfast in its commitment to upholding the highest ethical standards in the administration of justice. We acknowledge the rapid development of generative AI technologies capable of producing human-like text, images, video, audio, and other content. The integration of AI with the courts is increasingly pervasive, offering potential efficiencies and improved access to justice. However, it also raises critical concerns about authenticity, accuracy, bias, and the integrity of court filings, proceedings, evidence, and decisions. Understanding the capabilities and limitations of AI technology is essential for the Illinois Judicial Branch.

The Illinois Courts will be vigilant against AI technologies that jeopardize due process, equal protection, or access to justice. Unsubstantiated or deliberately misleading AI-generated content that perpetuates bias, prejudices litigants, or obscures truth-finding and decision-making will not be tolerated.

The use of AI by litigants, attorneys, judges, judicial clerks, research attorneys, and court staff providing similar support may be expected, should not be discouraged, and is authorized provided it complies with legal and ethical standards. Disclosure of AI use should not be required in a pleading.

The Rules of Professional Conduct and the Code of Judicial Conduct apply fully to the use of AI technologies. Attorneys, judges, and self-represented litigants are accountable for their final work product. All users must thoroughly review AI-generated content before submitting it in any court proceeding to ensure accuracy and compliance with legal and ethical obligations. Prior to employing any technology, including generative AI applications, users must understand both general AI capabilities and the specific tools being utilized.

The Court acknowledges the necessity of safe AI use, adhering to laws and regulations concerning privacy and confidentiality. AI applications must not compromise sensitive information, such as confidential communications, personal identifying information (PII), protected health information (PHI), justice and public safety data, security-related information, or information conflicting with judicial conduct standards or eroding public trust.

This policy reflects the Illinois Supreme Court's commitment to upholding foundational principles while exploring the potential benefits of new AI technologies in a dynamic landscape. The Court will regularly reassess policies as these technologies evolve, prioritizing public trust and confidence in the judiciary and the administration of justice. **Judges remain ultimately responsible for their decisions, irrespective of technological advancements.**

The Court encourages the development of technologies that enhance service to all court users and promote equitable access to justice. To facilitate this, the judicial branch will support ongoing education on emerging technologies, including AI.

ILLINOIS SUPREME COURT POLICY ON ARTIFICIAL INTELLIGENCE

JUDICIAL REFERENCE SHEET

JANUARY 1, 2025

WHAT IS ARTIFICIAL INTELLIGENCE?

Technology that simulates human intelligence, enabling machines to learn, reason, perceive, and make decisions.
Artificial Intelligence is not new technology.

1950s Origins of AI

Examples: Spell check, predicative typing, facial recognition, and computer based legal research.

2022 Mainstream Availability of Generative AI

Examples:
Text Composition Prompts
"Summarize the following legal brief and identify key arguments."

"Rewrite this paragraph in a respectful and neutral tone using plain language so it can be understood by people without legal expertise."

"Prepare a speech about the importance of procedural due process for an audience of judges."

Photo/Audio/Video Prompts
"Create image of an Illinois Courtroom."

"Create movie depicting President Abraham Lincoln conducting legal research on a computer."



WHAT IS GENERATIVE ARTIFICIAL INTELLIGENCE?

A subset of artificial intelligence focused on creating new content, such as text, images, and video, by learning from existing data.
Generative AI is a relatively new tool.

WANT TO KNOW MORE?

National Center for State Courts AI Resource Center

Description of AI and Court Use Cases Video

JUDICIAL DECISIONS

Judges remain ultimately responsible for their decisions, irrespective of technological advancements.

Code of Judicial Conduct - Rule 2.7
A judge shall hear and **decide** matters assigned to the judge...

Code of Judicial Conduct - Rule 1.2
A judge shall act at all times in a manner that promotes public confidence in the independence, integrity, and impartiality of the judiciary...

PLEADINGS

Lawyers & Self Represented Litigants are subject to sanctions for submitting legally or factually unfounded pleadings.

Illinois Supreme Court Rule 137
The signature of an attorney or party constitutes a certificate by him that he has read the pleading, motion or other document; that to the best of his knowledge, information, and belief formed after reasonable inquiry it is well grounded in fact and is warranted by existing law...

THINKING ABOUT USING GENERATIVE AI?
JUDICIAL AI UTILIZATION GUIDELINES

Ethical Oversight - The Code of Judicial Conduct applies fully to the use of AI technologies. You maintain ultimate responsibility for all rulings and legal documents.

Competence - The judicial branch must stay informed about evolving AI technologies. Prior to using any technology, including generative AI applications, you need to understand both general AI capabilities and the specific tools being used.

Accuracy - Be mindful that content from generative AI applications comes from sourced material. AI-generated content must be thoroughly reviewed to ensure accuracy and compliance with legal and ethical obligations, including a specific need to guard against technology lending to unintentional bias or prejudice.

Attribution - Ensure your work product does not infringe upon copyright or intellectual property rights by proper attribution to sources as necessary.

Confidentiality - If using a public generative AI tool (like ChatGPT), your input prompt is being handed over to the technology. Ensure you do not compromise sensitive information. Do not input any information such as (non-exhaustive list):

- Confidential or privileged information;
- Personal identifying information;
- Protected health information;
- Justice and public safety information;
- Code containing passwords or security-related information; and
- Information that has potential to erode public trust.

WHAT TO WATCH FOR AS A JUDGE

AI is ubiquitously present in modern technology. It is safe to assume AI was used in pleadings and other written materials. Generative AI applications gives rise to these considerations:

Hallucinations - When generative AI produces output that appears realistic but is misleading or made up by the AI itself rather than real-world data or input. If generative AI was used in a legal pleading or brief, included citations may be entirely made up or be a real case but not contain the purported language cited. Read a case decision on hallucinations.

Deepfakes - When generative AI is intentionally used to produce convincing/deceptive media - images, audio, video, etc. Fake evidence is relatively easy to create and reliable technology solutions do not exist to identify real vs fake evidence. Learn more about identification of deepfakes.

Extended Reality - Technology can seamlessly be integrated with our person though wearable devices. Recognize that technology may allow an individual to record and analyze audio and video and channel information from other sources in real-time, potentially without detection. Learn more about extended reality.

CHECK CITATIONS

HEAR EVIDENCE ON FOUNDATIONS

ENFORCE EXISTING TECHNOLOGY RULES

Appendix 2
GAI Terms of Use Checklist

To assist with choosing appropriate third-party managed GAI tools, we prepared the following checklist. This is intended to help identify and evaluate important contractual protections, but should not be interpreted as a mandatory or exhaustive set of requirements.

- No Model Training: Neither the model provider nor any integrated application provider may use customer data to develop or improve its services without explicit, written customer consent.
- Purpose Limitation: Use of customer data should be limited to providing the services, complying with applicable law, enforcing provider policies, and preventing abuse. These purposes may be further limited where Zero Data Retention is enabled.
- Data Retention: Customer data must be permanently deleted from the provider's servers within a reasonable period after it is deleted from the user interface, or after account termination, except where retention is required by law.
- Data Isolation: Where higher degrees of data isolation are critical for the type of data being processed, the contract should address the specific isolation requirements for both the application and the transformer model.
- Data Residency: Where specific geographic hosting location is critical for the type of data being processed, the contract should specify the permitted data processing jurisdictions and include any applicable supplemental provisions (e.g. a Data Processing Addendum).
- User Anonymity: Where anonymous GAI inputs are critical to user privacy, the provider must ensure API requests do not include identifiable client or firm information.
- Compliance with Privacy Laws: Provider must comply with all privacy laws applicable to the data, customer, and clients.
- Breach Notification: Providers should clearly agree to provide timely breach notification requirements and specify their incident response obligations.
- Risk Management: The application must be developed, deployed, and maintained in accordance with recognized information security frameworks (such as SOC 2, ISO 27001, CSA STAR) and AI risk management frameworks (e.g., NIST AI RMF, OWASP GenAI Security Project). The provider should publish or provide supporting documentation demonstrating compliance.
- Third-Party Access Controls: Define and limit third-party access rights and include an obligation of the primary provider to bind downstream third parties to all relevant data management obligations.
- Supplementary Agreements: Explore the availability of a Data Processing Addendum for personal data processing or a HIPAA business associate agreement when required for protected health information.

The specific protections to include will depend on the type of data being processed, applicable regulations, and the lawyer's professional obligations in the given context.

Appendix 3
Sample Notice of Artificial Intelligence Practices

SAMPLE NOTICE OF ARTIFICIAL INTELLIGENCE PRACTICES

Acme Law, LLC provides this Notice of Artificial Intelligence Practices to explain how and when artificial intelligence tools may be used in providing our services. We believe it is important for clients to understand the circumstances under which personally identifiable information and client confidential information may be processed using artificial intelligence, the safeguards we apply to protect that information, and your rights relating to our use of such tools.

GENERATIVE ARTIFICIAL INTELLIGENCE DEFINED

When lawyers refer to "AI," they are typically referencing a specific type of computer algorithm that predicts and generates text by analyzing patterns in numerical representations of language and images. These algorithms are often referred to as *Generative Artificial Intelligence* ("*GAI*").

HOW WE USE AI

Some of our systems and professionals use business-class GAI tools. The attached Schedule of Generative Artificial Intelligence Tools provides representative examples of the tools we use.[54] Clients who have an active matter with us may request a current list of our GAI tools using the information provided at the end of this Notice. Our GAI Tools are used in the following contexts:

- System-Wide Processing. "System-Wide Processing" means using GAI systems in firm-wide administrative, security, or operational tools as part of standard functions, such as document management, email filtering, cybersecurity, or automated billing.

- De-Identified Processing. Some professionals at Acme Law, LLC may use GAI tools as work assistants without inputting any personally identifiable or client confidential information. When this type of processing relates directly to an active matter, clients have opt-out rights as described under "Your Rights" below.

- Confidential Processing. Professionals may, in some cases, use GAI tools to process personally identifiable or client confidential information. For example, a document may be uploaded to a GAI tool to support efficient review of that document. Clients have consent and opt-out rights for this type of processing, as described under "Your Rights" below.

[54] Lawyers may consider omitting the schedule of representative tools entirely, removing it from public-facing materials, or limiting distribution to active clients who request more information.

HOW WE PROTECT YOUR INFORMATION

We maintain a formal written policy to ensure responsible and secure use of Generative Artificial Intelligence (GAI) tools. Among other things, our written policy provides the following safeguards:

- We take direct professional responsibility for the output of any GAI tool used in connection with our work. We believe GAI tools help make lawyers more efficient, but we do not rely on them to make legal judgments or decisions.

- When GAI tools are used for System-Wide Processing, we take reasonable and appropriate measures to ensure those tools are suitable for handling such protected information.

- We review all GAI tools for compliance with appropriate privacy assurances and security requirements before use. We do not use consumer-grade GAI tools such as the free or personal versions of ChatGPT.

- We maintain a written policy and train all lawyers and staff members to ensure that your rights, described below, are honored.

YOUR RIGHTS

Clients have the right to remain informed about how GAI tools may be used in matters we undertake for them. We will regularly update the attached Schedule of Generative Artificial Intelligence Tools, and clients with an active matter may request a current version at any time.[55]

Clients may opt out of De-Identified Processing and Confidential Processing (defined above). To exercise this right, send us written notice using the contact information provided below.

Other than for System-Wide Processing, we will not use GAI tools to process information listed in the Illinois Personal Information Protection Act without your informed written consent. Individual clients may provide consent for use of their data directly. Organizational clients may provide consent if they are the lawful controller or processor of the data or are otherwise permitted to consent under applicable law. Consent may be revoked any time and for any reason. Revocation applies prospectively from the date it is received.

If we use GAI tools for System-Wide Processing (as defined above), we will take reasonable and appropriate measures to ensure such tools are suitable for processing personally identifiable and client confidential information.

Contact Information

Johnathan A. Doe, Managing Partner
ACME Law, LLC
1234 Justice Lane, Suite 800
Sampleville, IL 54321
Phone: (555) 867-5309
Email: jdoe@acmelawfictional.com

Effective Date: October 1, 2025

[55] Refer to note 54.

SCHEDULE OF GENERATIVE ARTIFICIAL INTELLIGENCE TOOLS[56]

The following table sets forth the Generative Artificial Intelligence tools in use at ACME Law, LLC as of the Effective Date of this Notice. Clients having an active matter with us may request a current list of our GAI tools using the contact information in this Notice.

Tool[57]	Representative Uses	Provider Safeguards
ChatGPT Team	document summarization or analysis, drafting assistance	https://trust.openai.com
Microsoft Copilot	AI integrations in Microsoft Office and Teams	https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy
Adobe AI Assistant	Document summarization and analysis	https://experienceleague.adobe.com/en/docs/experience-platform/ai-assistant/privacy
vLex	Legal research, document drafting and analysis	https://trust.vlex.com
Lexis+AI	Legal research, document drafting and analysis	https://www.lexisnexis.com/pdf/relx-responsible-ai-principles.pdf
CoCounsel	Legal research, document drafting and analysis	https://help.casetext.com/en/collections/6255626-trust-security-and-privacy

The GAI tools listed above may rely on third-party providers and infrastructure to deliver their services. We enter into agreements with these providers that require appropriate privacy and security safeguards throughout the supply chain, including confidentiality, data protection, and access control measures.

[56] Refer to note 54.

[57] ARDC does not recommend or endorse any specific technology or provider. The examples given in this table are simply provided to illustrate how a typical law firm might choose to disclose the GAI tools it uses.

Appendix 4

Sample Use of GAI Tools Policy

POLICY Use of GAI Tools	
Policy ID: AI G-100.1	References: ❖ https://openai.com ❖ https://www.microsoft.com/en-us/ai ❖ Illinois Supreme Court Policy on Artificial Intelligence ❖ Illinois Rule of Professional Conduct Rule 1.1 (Competence) ❖ Illinois Rule of Professional Conduct Rule 1.4 (Communication) ❖ Illinois Rule of Professional Conduct Rule 1.6 (Confidentiality of Information)
Approved By: _____ Johnathan A. Doe, Managing Partner	

1. Purpose. This Policy addresses the use of Generative Artificial Intelligence (as defined below) at ACME Law, LLC. ("ACME"). Generative Artificial Intelligence is a rapidly evolving technology that can bring great value to our efforts. However, its legal and technical risks are not well-known, and its potential for unintentional misuse merits a consistent policy-based approach to using this technology within our environment.

2. Scope. This Policy applies to all members of the ACME Workforce. The term "ACME Workforce" means ACME employees, as well as independent contractors, interns, volunteers, and other individuals who are appointed or engaged to provide services or create work product on ACME's behalf.

   This Policy does not apply to activities outside the scope of one's work for ACME unless such activities adversely affect the confidentiality, security, or reputation of ACME.

   This Policy covers all company systems, devices, technologies, and platforms, including software applications, hardware devices, cloud platforms, mobile applications, and any other technology interfaces.

3. Definition of Generative Artificial Intelligence. For purposes of this Policy, the term "Generative Artificial Intelligence" or "GAI" refers to an artificial intelligence tool that is designed to produce new, previously unseen, data in response to human input using predictive algorithms applied to an underlying database of training data and may include systems that generate new images, text, sounds, or other data types in response to human input.

4. Approved GAI Tools. The ACME Managing Partner may designate one or more GAI tools as approved for use by the ACME Workforce (each an "Approved GAI Tool"). Approved GAI Tools will be posted at [***insert link or document ID***] (the "Schedule of Approved GAI Tools"). Only Approved GAI Tools may be used on ACME devices or to process ACME information (including ACME client information).

5. Professional Responsibility. Members of the ACME Workforce who use Approved GAI Tools may do so solely as supplementary instruments, such as to generate ideas or assist in the review or preparation of documents. While GAI can serve as a valuable tool, ultimate responsibility for the accuracy, relevance, and appropriateness of any final work product rests solely with the ACME Workforce member using the Approved GAI Tool.

   (a) General Rule. GAI Tools are prone to generate fictitious case and statutory citations (a phenomenon often referred to as a "*hallucination*"). Accordingly, except as provided in Section 5(b), it is imperative that all outputs derived from Approved GAI Tools be meticulously reviewed, validated, and edited as necessary by the user before finalization. Use of Approved GAI Tools should not substitute or diminish the critical thinking, judgment, and expertise that our ACME Workforce members bring to their respective roles.

   (b) Exception for Mass Data Processing. Where the value of using an Approved GAI Tool would be diminished by the requirement that all GAI outputs be meticulously reviewed and validated, the ACME Managing Partner may grant case-by-case exceptions to Section 5(a). For example, ACME might use an Approved GAI Tool to search or summarize large datasets that cannot reasonably be validated by humans or even traditional computer algorithms. In this example, requiring the output of the GAI process to be reviewed and validated by humans might defeat the purpose of using the GAI tool.

   (c) Client Consent for Allocation of Responsibility. In such cases where ACME cannot reasonably exercise full professional responsibility for the GAI output, the client's informed written consent shall be obtained as described in Section 6(c). The responsible attorney should consult Illinois Rule of Professional Conduct 1.2 when structuring a matter in this way.

6. Client Rights. Our clients have a right to be informed about how ACME uses GAI tools to process their matters and confidential information. This section outlines the rules for informing clients about our use of GAI tools. This section also describes when and how clients may opt out of our use of GAI tools to process their matters and confidential information. This section also describes the circumstances under which we are required to obtain a client's informed written consent prior to processing their information using a GAI tool.

   (a) Notice of Artificial Intelligence Practices. At all times after implementation of this Policy, ACME will maintain a current and accurate Notice of Artificial Intelligence Practices describing how ACME uses GAI tools and safeguards client confidential information. This

Notice should be clearly posted on the ACME website. Clients should be provided with this Notice as part their engagement documents, and within a reasonable time after we make any material changes to our policies and procedures for using GAI tools. No information related to client matters or confidential information should be processed in GAI tools prior to the client receiving this Notice.

(b) Client Opt-Out Rights. Except as provided in Section 6(e), Clients have the right to restrict or prohibit the use of GAI tools in connection with their matters or to process their confidential information. A client may exercise this right at any time by providing written notice. Upon receipt of such notice, the client's preferences should be clearly and prominently documented in the client file. All ACME Workforce members on the file should be notified of the client's opt-out and take immediate steps to ensure that GAI tools are not used in any manner that violates the client's expressed restrictions.

Accordingly, unless a client has opted out, members of the ACME Workforce may use Approved GAI Tools for the following purposes:

i. De-Identified Processing. Approved GAI Tools may be used as a work assistant without inputting any personally identifiable or client confidential information. This could include posing hypothetical questions to an Approved GAI Tool to help generate a checklist for drafting documents.

ii. Non-Sensitive Confidential Processing. Approved GAI Tools may also be used to process client identifying or confidential information **other than Sensitive Personal Information as defined in the next section**. For example, a document may be uploaded to an Approved GAI Tool to support efficient review of that document, provided any Sensitive Personal Information has first been redacted.

(c) Sensitive Personal Information. Prior to using an Approved GAI Tool to process Sensitive Personal Information, as defined below, ACME must obtain the client's informed written consent using the procedure described in Section 6(d). For the purposes of this Policy, Sensitive Personal Information includes:

- Social security numbers or tax returns
- Driver's license numbers or state identification card numbers
- Account, credit card, or debit card numbers
- Medical information, including mental health or substance abuse treatment records
- Health insurance information
- Unique biometric identifiers or biometric information (as defined in Illinois law)
- Passwords and security questions and answers used to access computer accounts
- Any other information meeting PIPA's definition of "personal information," even if separated from a first name or first initial in combination with last name

(d) Informed Client Consent. Where client consent is required, it must be informed. The client must receive a clear explanation of the purpose of the processing, the type of data involved, the type of GAI tool to be used, the potential risks (including risks of disclosure or misuse), and the safeguards in place to protect the data. This information must be provided separately from other materials, and the client must be given the opportunity to ask questions before providing consent. Consent must be voluntary and may be revoked at any time and for any reason.

i. Authority to Consent. The client may provide consent to process personal data that is (i) the personal data of the client or (ii) personal data of which the client is the lawful controller or processor. If the responsible attorney has reason to doubt the client's authority to consent to the processing for the personal data provided, they must refrain from processing the data using GAI until appropriate authorization is confirmed.

ii. Discretionary Consents. The attorney responsible for any client or matter is free to require informed written consent for the use of Approved GAI Tools using criteria that is more restrictive than this Policy provides. If, in the professional judgment of the responsible attorney, informed written consent is prudent in any matter, then the responsible attorney's judgment shall control.

(e) System-Wide Processing. "*System-Wide Processing*" means using GAI systems in firm-wide administrative, security, or operational tools as part of standard functions, such as document management, email filtering, cybersecurity, or automated billing. These GAI tools are exempt from client opt-out or consent rights described above.

7. Implementation of Client and Project Level Restrictions. Prior to using Approved GAI Tools for any client or client project, the responsible attorney for the matter must complete an evaluation to determine whether any regulatory, contractual, or other specific obligations apply to the client or project. Approved GAI Tools may not be used unless such requirements are identified and complied with. For example:

i. Client-Imposed Restrictions. Clients may include GAI-related restrictions in their agreements or engagement letters. As described above, clients may also file an opt-out request to limit or prohibit the use of GAI tools to process their data.

ii. Flow-Down Restrictions. Clients may be subject to GAI limitations that flow down to ACME, such as public-sector procurement rules (e.g. the California State GenAI Procurement Guidelines).

iii. Data Protection Requirements. Certain client data may be subject to heightened privacy or security obligations that require additional safeguards around GAI use, including but not limited to the CCPA, GDPR, or HIPAA.

8. Court and Regulatory Rules. Some courts or regulators may require the use of GAI tools to be disclosed, documented, or certified. These rules may require a party to: (i) identify which GAI tool was used in drafting a court filing; (ii) explain how the GAI tool was used in preparing court filings or regulated submissions; (iii) verify the accuracy of AI-generated content using non-GAI authoritative sources; and/or (iv) confirm that no confidential or proprietary information was disclosed to unauthorized parties. Accordingly, prior to using Approved GAI Tools to prepare any materials that will be used in a judicial or regulatory proceeding, the supervising attorney will confirm what GAI rules apply to that proceeding.

9. Data Minimization. Each ACME Workforce member must limit the retention of data in Approved GAI Tools to only what is reasonably necessary to support the project or matter for which the Approved GAI Tools are used. Approved GAI Tools should not be used as long-term storage or recordkeeping systems. For example, Workforce members should delete conversations, documents, prompts, instructions, memory functions, and other related data when the associated client matter is closed, or if no longer reasonably needed to manage that specific project or matter.

10. Education and Training. The ACME Managing Partner will coordinate education and training for all members of the ACME Workforce on the responsible use of artificial intelligence, including the requirements of this Policy. This training should include understanding ethical implications, data privacy and security risks, and the effective practical use of approved GAI tools and extensions.

11. Intellectual Property. The output of a Generative Artificial Intelligence tool may not be subject to intellectual property protection. Accordingly, Generative Artificial Intelligence output may not be incorporated into materials intended to have intellectual property protection without specific review and approval on a case-by-case basis. Accordingly, absent such review and approval, all work product intended to receive intellectual property protection must be original work produced by ACME Workforce members acting within the scope of their employment or engagement and not work product that is produced by a GAI tool.

12. Enhanced Periodic Review. As previously noted, the field of Artificial Intelligence is evolving rapidly. Accordingly, this Policy must be reviewed on a more frequent basis than provided in other ACME policies. The ACME Managing Partner is therefore directed to establish an Artificial Intelligence Review Committee and a policy review schedule that is appropriate to the speed at which this technology is evolving and the rate at which AI-related tools are adopted within ACME (and our clients and vendors).

13. Version Control. Any amendments or additions to this Policy shall be reflected in a new version and summarized in the attached Revision Management Log.

Revision Management Log

Revision ID	Date	Brief Description
AI-100.1	**********	Initial Policy Adoption

Appendix 5
Sample Informed Client Consent Form

Informed Consent for the Use of Generative Artificial Intelligence

Matter: [Insert Matter Title]
Client: [Insert Client Name]

Acme Law, LLC may use Generative Artificial Intelligence ("GAI") tools to support legal work such as document analysis, summarization, legal research, or drafting. These tools can improve efficiency and enhance the quality of legal services when used responsibly and with appropriate safeguards. In your matter, we believe it may be beneficial and appropriate to process certain sensitive personal information using GAI tools. Accordingly, we request your written consent to use GAI tools to process sensitive personal information related to your matter.

Categories of Sensitive Personal Information to be Processed

Absent additional written consent, we will only use GAI tools to process the following categories of Sensitive Personal Information:

- Social Security numbers
- Driver's license or state ID numbers
- Financial account numbers or payment card data
- Medical or health insurance information
- Other information covered by the Illinois Personal Information Protection Act

Purpose of Processing

We will only process sensitive personal information using GAI tools if it is directly relevant to the legal services we are providing and necessary to achieve a meaningful benefit, such as faster document review, clearer summarization, or improved accuracy in analysis. Due to the volume of documents that must be reviewed in your matter, we believe the use of GAI tools will improve accuracy and reduce the time needed to perform our work.

GAI Tools to be Used

If you consent, we will only use business-class GAI tools which are subject to appropriate privacy and security assurances. The specific GAI tools used by ACME Law, LLC are available in our Notice of Artificial Intelligence Practices that is updated no less than quarterly.

Client Consent

I have read and understand this form and ACME Law, LLC's Notice of Artificial Intelligence Practices. I understand the nature and purpose of this request, and I have been given the opportunity to discuss any questions or concerns I may have with individuals who are knowledgeable about the specific GAI tools that will be used to process my sensitive personal information.

I understand that I may revoke this consent at any time and for any reason by using the contact information provided in ACME Law, LLC's Notice of Artificial Intelligence Practices. I also understand that my revocation will only be effective for uses of my personal information after the effective date of my revocation, and that my revocation will not apply to the use of GAI for system-wide processing within ACME Law, LLC's network and computer systems.

Accordingly, I consent to Acme Law, LLC's use of GAI tools as described above.

_____

Dated: _____

Appendix 6
Cybersecurity Information Sheet ("CIS")
Deploying AI Systems Securely: Best Practices for Deploying Secure and Resilient AI Systems





Australian Government
Australian Signals Directorate

ASD AUSTRALIAN SIGNALS DIRECTORATE
ACSC Australian Cyber Security Centre

National Cyber Security Centre
PART OF THE GCSB


Communications Security Establishment Canada
Canadian Centre for Cyber Security

Centre de la sécurité des télécommunications Canada
Centre canadien pour la cybersécurité


National Cyber Security Centre
a part of GCHQ

Deploying AI Systems Securely

Best Practices for Deploying Secure and Resilient AI Systems

Executive summary

Deploying artificial intelligence (AI) systems securely requires careful setup and configuration that depends on the complexity of the AI system, the resources required (e.g., funding, technical expertise), and the infrastructure used (i.e., on premises, cloud, or hybrid). This report expands upon the 'secure deployment' and 'secure operation and maintenance' sections of the Guidelines for secure AI system development and incorporates mitigation considerations from Engaging with Artificial Intelligence (AI). It is for organizations deploying and operating AI systems designed and developed by another entity. The best practices may not be applicable to all environments, so the mitigations should be adapted to specific use cases and threat profiles. [1], [2]

AI security is a rapidly evolving area of research. As agencies, industry, and academia discover potential weaknesses in AI technology and techniques to exploit them, organizations will need to update their AI systems to address the changing risks, in addition to applying traditional IT best practices to AI systems.

This report was authored by the U.S. National Security Agency's Artificial Intelligence Security Center (AISC), the Cybersecurity and Infrastructure Security Agency (CISA), the Federal Bureau of Investigation (FBI), the Australian Signals Directorate's Australian Cyber Security Centre (ACSC), the Canadian Centre for Cyber Security (CCCS), the New Zealand National Cyber Security Centre (NCSC-NZ), and the United Kingdom's National Cyber Security Centre (NCSC-UK). The goals of the AISC and the report are to:

1. Improve the confidentiality, integrity, and availability of AI systems;
2. Assure that known cybersecurity vulnerabilities in AI systems are appropriately mitigated; and
3. Provide methodologies and controls to protect, detect, and respond to malicious activity against AI systems and related data and services.

This document is marked TLP:CLEAR. Recipients may share this information without restriction. Information is subject to standard copyright rules. For more on the Traffic Light Protocol, see cisa.gov/tlp/.

Scope and audience

The term AI systems throughout this report refers to machine learning (ML) based artificial intelligence (AI) systems.

These best practices are most applicable to organizations deploying and operating externally developed AI systems on premises or in private cloud environments, especially those in high-threat, high-value environments. They are not applicable for organizations who are not deploying AI systems themselves and instead are leveraging AI systems deployed by others.

Not all of the guidelines will be directly applicable to all organizations or environments. The level of sophistication and the methods of attack will vary depending on the adversary targeting the AI system, so organizations should consider the guidance alongside their use cases and threat profile.

See Guidelines for secure AI system development for design and development aspects of AI systems. [1]

Introduction

The rapid adoption, deployment, and use of AI capabilities can make them highly valuable targets for malicious cyber actors. Actors, who have historically used data theft of sensitive information and intellectual property to advance their interests, may seek to co-opt deployed AI systems and apply them to malicious ends.

Malicious actors targeting AI systems may use attack vectors unique to AI systems, as well as standard techniques used against traditional IT. Due to the large variety of attack vectors, defenses need to be diverse and comprehensive. Advanced malicious actors often combine multiple vectors to execute operations that are more complex. Such combinations can more effectively penetrate layered defenses.

Organizations should consider the following best practices to secure the deployment environment, continuously protect the AI system, and securely operate and maintain the AI system.

The best practices below align with the cross-sector Cybersecurity Performance Goals (CPGs) developed by CISA and the National Institute of Standards and Technology (NIST). The CPGs provide a minimum set of practices and protections that CISA and NIST recommend all organizations implement. CISA and NIST based the CPGs on

existing cybersecurity frameworks and guidance to protect against the most common and impactful threats, tactics, techniques, and procedures. Visit CISA's Cross-Sector Cybersecurity Performance Goals for more information on the CPGs, including additional recommended baseline protections.

Secure the deployment environment

Organizations typically deploy AI systems within existing IT infrastructure. Before deployment, they should ensure that the IT environment applies sound security principles, such as robust governance, a well-designed architecture, and secure configurations. For example, ensure that the person responsible and accountable for AI system cybersecurity is the same person responsible and accountable for the organization's cybersecurity in general [CPG 1.B].

The security best practices and requirements for IT environments apply to AI systems, too. The following best practices are particularly important to apply to the AI systems and the IT environments the organization deploys them in.

Manage deployment environment governance

- If an organization outside of IT is deploying or operating the AI system, work with the IT service department to identify the deployment environment and confirm it meets the organization's IT standards.
    - Understand the organization's risk level and ensure that the AI system and its use is within the organization's risk tolerance overall and within the risk tolerance for the specific IT environment hosting the AI system. Assess and document applicable threats, potential impacts, and risk acceptance. [3], [4]
    - Identify the roles and responsibilities for each stakeholder along with how they are accountable for fulfilling them; identifying these stakeholders is especially important should the organization manage their IT environment separately from their AI system.
    - Identify the IT environment's security boundaries and how the AI system fits within them.
- Require the primary developer of the AI system to provide a threat model for their system.

    - The AI system deployment team should leverage the threat model as a guide to implement security best practices, assess potential threats, and plan mitigations. [5], [6]

- Consider deployment environment security requirements when developing contracts for AI system products or services.
- Promote a collaborative culture for all parties involved, including the data science, infrastructure, and cybersecurity teams in particular, to allow for teams to voice any risks or concerns and for the organization to address them appropriately.

Ensure a robust deployment environment architecture

- Establish security protections for the boundaries between the IT environment and the AI system [CPG 2.F].
- Identify and address blind spots in boundary protections and other security-relevant areas in the AI system the threat model identifies. For example, ensure the use of an access control system for the AI model weights and limit access to a set of privileged users with two-person control (TPC) and two-person integrity (TPI) [CPG 2.E].
- Identify and protect all proprietary data sources the organization will use in AI model training or fine-tuning. Examine the list of data sources, when available, for models trained by others. Maintaining a catalog of trusted and valid data sources will help protect against potential data poisoning or backdoor attacks. For data acquired from third parties, consider contractual or service level agreement (SLA) stipulations as recommended by CPG 1.G and CPG 1.H.
- Apply secure by design principles and Zero Trust (ZT) frameworks to the architecture to manage risks to and from the AI system. [7], [8], [9]

Harden deployment environment configurations

- Apply existing security best practices to the deployment environment. This includes sandboxing the environment running ML models within hardened containers or virtual machines (VMs) [CPG 2.E], monitoring the network [CPG 2.T], configuring firewalls with allow lists [CPG 2.F], and other best practices, such as those in NSA's Top Ten Cloud Mitigation Strategies for cloud deployments.
- Review hardware vendor guidance and notifications (e.g., for GPUs, CPUs, memory) and apply software patches and updates to minimize the risk of exploitation of vulnerabilities, preferably via the Common Security Advisory Framework (CSAF). [10]

- Secure sensitive AI information (e.g., AI model weights, outputs, and logs) by encrypting the data at rest, and store encryption keys in a hardware security module (HSM) for later on-demand decryption [CPG 2.L].
- Implement strong authentication mechanisms, access controls, and secure communication protocols, such as by using the latest version of Transport Layer Security (TLS) to encrypt data in transit [CPG 2.K].
- Ensure the use of phishing-resistant multifactor authentication (MFA) for access to information and services. [2] Monitor for and respond to fraudulent authentication attempts [CPG 2.H]. [11]
- Understand and mitigate how malicious actors exploit weak security controls by following the mitigations in Weak Security Controls and Practices Routinely Exploited for Initial Access.

Protect deployment networks from threats

Adopt a ZT mindset, which assumes a breach is inevitable or has already occurred. Implement detection and response capabilities, enabling quick identification and containment of compromises. [8], [9]

- Use well-tested, high-performing cybersecurity solutions to identify attempts to gain unauthorized access efficiently and enhance the speed and accuracy of incident assessments [CPG 2.G].
- Integrate an incident detection system to help prioritize incidents [CPG 3.A]. Also integrate a means to immediately block access by users suspected of being malicious or to disconnect all inbound connections to the AI models and systems in case of a major incident when a quick response is warranted.

Continuously protect the AI system

Models are software, and, like all other software, may have vulnerabilities, other weaknesses, or malicious code or properties.

Validate the AI system before and during use

- Use cryptographic methods, digital signatures, and checksums to confirm each artifact's origin and integrity (e.g., encrypt safetensors to protect their integrity and confidentiality), protecting sensitive information from unauthorized access during AI processes. [14]

- Create hashes and encrypted copies of each release of the AI model and system for archival in a tamper-proof location, storing the hash values and/or encryption keys inside a secure vault or HSM to prevent access to both the encryption keys and the encrypted data and model at the same location. [1]
- Store all forms of code (e.g., source code, executable code, infrastructure as code) and artifacts (e.g., models, parameters, configurations, data, tests) in a version control system with proper access controls to ensure only validated code is used and any changes are tracked. [1]
- Thoroughly test the AI model for robustness, accuracy, and potential vulnerabilities after modification. Apply techniques, such as adversarial testing, to evaluate the model's resilience against compromise attempts. [4]
- Prepare for automated rollbacks and use advanced deployments with a human-in-the-loop as a failsafe to boost reliability, efficiency, and enable continuous delivery for AI systems. In the context of an AI system, rollback capabilities ensure that if a new model or update introduces problems or if the AI system is compromised, the organization can quickly revert to the last known good state to minimize the impact on users.
- Evaluate and secure the supply chain for any external AI models and data, making sure they adhere to organizational standards and risk management policies, and preferring ones developed according to secure by design principles. Make sure that the risks are understood and accepted for parts of the supply chain that cannot adhere to organizational standards and policies. [1], [7]
- Do not run models right away in the enterprise environment. Carefully inspect models, especially imported pre-trained models, inside a secure development zone prior to considering them for tuning, training, and deployment. Use organization-approved AI-specific scanners, if and when available, for the detection of potential malicious code to assure model validity before deployment.
- Consider automating detection, analysis, and response capabilities, making IT and security teams more efficient by giving them insights that enable quick and targeted reactions to potential cyber incidents. Perform continuous scans of AI models and their hosting IT environments to identify possible tampering.
    - When considering whether to use other AI capabilities to make automation more efficient, carefully weigh the risks and benefits, and ensure there is a human-in-the-loop where needed.

Secure exposed APIs

- If the AI system exposes application programming interfaces (APIs), secure them by implementing authentication and authorization mechanisms for API access. Use secure protocols, such as HTTPS with encryption and authentication [CPG 2.C, 2.D, 2.G, 2.H]. [1]
- Implement validation and sanitization protocols for all input data to reduce the risk of undesired, suspicious, incompatible, or malicious input being passed to the AI system (e.g., prompt injection attacks). [1]

Actively monitor model behavior

- Collect logs to cover inputs, outputs, intermediate states, and errors; automate alerts and triggers [CPG 2.T].
- Monitor the model's architecture and configuration settings for any unauthorized changes or unexpected modifications that might compromise the model's performance or security. [1]
- Monitor for attempts to access or elicit data from the AI model or aggregate inference responses. [1]

Protect model weights

- Harden interfaces for accessing model weights to increase the effort it would take for an adversary to exfiltrate the weights. For example, ensure APIs return only the minimal data required for the task to inhibit model inversion.
- Implement hardware protections for model weight storage as feasible. For example, disable hardware communication capabilities that are not needed and protect against emanation or side channel techniques.
- Aggressively isolate weight storage. For example, store model weights in a protected storage vault, in a highly restricted zone (HRZ) (i.e., a separate dedicated enclave), or using an HSM [CPG 2.L]. [12]

Secure AI operation and maintenance

Follow organization-approved IT processes and procedures to deploy the AI system in an approved manner, ensuring the following controls are implemented.

Enforce strict access controls

- Prevent unauthorized access or tampering with the AI model. Apply role-based access controls (RBAC), or preferably attribute-based access controls (ABAC) where feasible, to limit access to authorized personnel only.
    - Distinguish between users and administrators. Require MFA and privileged access workstations (PAWs) for administrative access [CPG 2.H].

Ensure user awareness and training

Educate users, administrators, and developers about security best practices, such as strong password management, phishing prevention, and secure data handling. Promote a security-aware culture to minimize the risk of human error. If possible, use a credential management system to limit, manage, and monitor credential use to minimize risks further [CPG 2.I].

Conduct audits and penetration testing

- Engage external security experts to conduct audits and penetration testing on ready-to-deploy AI systems. This helps identify vulnerabilities and weaknesses that may have been overlooked internally. [13], [15]

Implement robust logging and monitoring

- Monitor the system's behavior, inputs, and outputs with robust monitoring and logging mechanisms to detect any abnormal behavior or potential security incidents [CPG 3.A]. [16] Watch for data drift or high frequency or repetitive inputs (as these could be signs of model compromise or automated compromise attempts). [17]
- Establish alert systems to notify administrators of potential oracle-style adversarial compromise attempts, security breaches, or anomalies. Timely detection and response to cyber incidents are critical in safeguarding AI systems. [18]

Update and patch regularly

- When updating the model to a new/different version, run a full evaluation to ensure that accuracy, performance, and security tests are within acceptable limits before redeploying.

Prepare for high availability (HA) and disaster recovery (DR)

- Use an immutable backup storage system, depending on the requirements of the system, to ensure that every object, especially log data, is immutable and cannot be changed [CPG 2.R]. [2]

Plan secure delete capabilities

- Perform autonomous and irretrievable deletion of components, such as training and validation models or cryptographic keys, without any retention or remnants at the completion of any process where data and models are exposed or accessible. [19]

Conclusion

The authoring agencies advise organizations deploying AI systems to implement robust security measures capable of both preventing theft of sensitive data and mitigating misuse of AI systems. For example, model weights, the learnable parameters of a deep neural network, are a particularly critical component to protect. They uniquely represent the result of many costly and challenging prerequisites for training advanced AI models, including significant compute resources; collected, processed, and potentially sensitive training data; and algorithmic optimizations.

AI systems are software systems. As such, deploying organizations should prefer systems that are secure by design, where the designer and developer of the AI system takes an active interest in the positive security outcomes for the system once in operation. [7]

Although comprehensive implementation of security measures for all relevant attack vectors is necessary to avoid significant security gaps, and best practices will change as the AI field and techniques evolve, the following summarizes some particularly important measures:

- Conduct ongoing compromise assessments on all devices where privileged access is used or critical services are performed.
- Harden and update the IT deployment environment.
- Review the source of AI models and supply chain security.
- Validate the AI system before deployment.
- Enforce strict access controls and API security for the AI system, employing the concepts of least privilege and defense-in-depth.

- Use robust logging, monitoring, and user and entity behavior analytics (UEBA) to identify insider threats and other malicious activities.
- Limit and protect access to the model weights, as they are the essence of the AI system.
- Maintain awareness of current and emerging threats, especially in the rapidly evolving AI field, and ensure the organization's AI systems are hardened to avoid security gaps and vulnerabilities.

In the end, securing an AI system involves an ongoing process of identifying risks, implementing appropriate mitigations, and monitoring for issues. By taking the steps outlined in this report to secure the deployment and operation of AI systems, an organization can significantly reduce the risks involved. These steps help protect the organization's intellectual property, models, and data from theft or misuse. Implementing good security practices from the start will set the organization on the right path for deploying AI systems successfully.

Works cited

[1] National Cyber Security Centre et al. Guidelines for secure AI system development. 2023. https://www.ncsc.gov.uk/files/Guidelines-for-secure-AI-system-development.pdf

[2] Australian Signals Directorate et al. Engaging with Artificial Intelligence (AI). 2024. https://www.cyber.gov.au/sites/default/files/2024-01/Engaging%20with%20Artificial%20Intelligence%20%28AI%29.pdf

[3] MITRE. ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems) Matrix version 4.0.0. 2024. https://atlas.mitre.org/matrices/ATLAS

[4] National Institute of Standards and Technology. AI Risk Management Framework 1.0. 2023. https://www.nist.gov/itl/ai-risk-management-framework

[5] The Open Worldwide Application Security Project (OWASP®). LLM AI Cybersecurity & Governance Checklist. 2024. https://owasp.org/www-project-top-10-for-large-language-model-applications/llm-top-10-governance-doc/LLM_AI_Security_and_Governance_Checklist-v1.pdf

[6] The Open Worldwide Application Security Project (OWASP®). OWASP Machine Learning Security Top Ten Security Risks. 2023. https://owasp.org/www-project-machine-learning-security-top-10/

[7] Cybersecurity and Infrastructure Security Agency. Secure by Design. 2023. https://www.cisa.gov/securebydesign

[8] National Security Agency. Embracing a Zero Trust Security Model. 2021. https://media.defense.gov/2021/Feb/25/2002588479/-1/-1/0/CSI_EMBRACING_ZT_SECURITY_MODEL_UOO115131-21.PDF

[9] Cybersecurity and Infrastructure Security Agency. Zero Trust Maturity Model. 2022. https://www.cisa.gov/zero-trust-maturity-model

[10] Cybersecurity and Infrastructure Security Agency. Transforming the Vulnerability Management Landscape. 2022. https://www.cisa.gov/news-events/news/transforming-vulnerability-management-landscape

[11] Cybersecurity and Infrastructure Security Agency. Implementing Phishing-Resistant MFA. 2022. https://www.cisa.gov/sites/default/files/publications/fact-sheet-implementing-phishing-resistant-mfa-508c.pdf
[12] Canadian Centre for Cyber Security. Baseline security requirements for network security zones Ver. 2.0 (ITSP.80.022). 2021. https://www.cyber.gc.ca/en/guidance/baseline-security-requirements-network-security-zones-version-20-itsp80022
[13] Ji, Jessica. What Does AI Red-Teaming Actually Mean? 2023. https://cset.georgetown.edu/article/what-does-ai-red-teaming-actually-mean/
[14] Hugging Face GitHub. Safetensors. 2024. https://github.com/huggingface/safetensors.
[15] Michael Feffer, Anusha Sinha, Zachary C. Lipton, Hoda Heidari. Red-Teaming for Generative AI: Silver Bullet or Security Theater? 2024. https://arxiv.org/abs/2401.15897
[16] Google. Google's Secure AI Framework (SAIF). 2023. https://safety.google/cybersecurity-advancements/saif/
[17] Government Accountability Office (GAO). Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities. 2021. https://www.gao.gov/assets/gao-21-519sp.pdf
[18] RiskInsight. Attacking AI? A real-life example!. 2023. https://riskinsight-wavestone.com/en/2023/06/attacking-ai-a-real-life-example
[19] National Cyber Security Centre. Principles for the security of machine learning. 2022. https://www.ncsc.gov.uk/files/Principles-for-the-security-of-machine-learning.pdf

Disclaimer of endorsement

The information and opinions contained in this document are provided "as is" and without any warranties or guarantees. Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the United States Government, and this guidance shall not be used for advertising or product endorsement purposes.

Purpose

This document was developed in furtherance of the authoring organizations' cybersecurity missions, including their responsibilities to identify and disseminate threats, and to develop and issue cybersecurity specifications and mitigations. This information may be shared broadly to reach all appropriate stakeholders.

Contact

U.S. organizations:

NSA Cybersecurity Report Feedback: CybersecurityReports@nsa.gov
NSA General Cybersecurity Inquiries or Customer Requests: Cybersecurity_Requests@nsa.gov
Defense Industrial Base Inquiries and Cybersecurity Services: DIB_Defense@cyber.nsa.gov
NSA Media Inquiries / Press Desk: 443-634-0721, MediaRelations@nsa.gov
Report incidents and anomalous activity to CISA 24/7 Operations Center at report@cisa.gov or (888) 282-0870 and/or to the FBI via your local FBI field office.

Australian organizations: For more information or to report a cybersecurity incident, visit cyber.gov.au or call 1300 292 371 (1300 CYBER1).

Canadian organizations: For more information contact the Cyber Centre at contact@cyber.gc.ca or report a cyber security incident to our portal at https://www.cyber.gc.ca/en/incident-management.

New Zealand organizations: Report cyber security incidents to incidents@ncsc.govt.nz or call 04 498 7654.

United Kingdom organizations: Report a significant cyber security incident at ncsc.gov.uk/report-an-incident (monitored 24 hours) or, for urgent assistance, call 03000 200 973.



ARDC

ATTORNEY REGISTRATION & DISCIPLINARY COMMISSION

One Prudential Plaza

130 East Randolph Drive

Suite 1500

Chicago, Illinois 60601-6219

312.565.2600

3161 West White Oaks Drive

Suite 301

Springfield, Illinois 62704

217.546.3523

www.iardc.org