IDENTIFYING AND MITIGATING GENERATIVE AI SECURITY AND RELIABILITY ISSUES FOR LEGAL PROFESSIONALS

ANDREW EDGE, *Austin*
McGinnis Lochridge

State Bar of Texas
36TH ANNUAL
ADVANCED GOVERNMENT LAW
July 25-26, 2024
San Antonio

CHAPTER 18

ANDREW EDGE
McGinnis Lochridge

Andrew Edge is a partner in the litigation section at McGinnis Lochridge, LLP. His practice encompasses a wide range of civil litigation matters, serving a diverse clientele of individuals, companies, governmental entities, and nonprofit organizations.

Andrew is the firm's attorney lead for developing policies, strategies, and vendor engagements related to generative artificial intelligence, and a member of the firm's technology committee. Andrew is a longtime legal technology enthusiast who frequently writes and speaks on the topic of generative AI, educating both attorneys and businesses about its potential in enterprise settings as well as associated risks and legal issues.

Andrew also authors the monthly newsletter "AI, Esq.," which explores the intersection of generative AI and legal services. He hopes to serve as a resource for those navigating the complexities of AI implementation in legal practice and to build community and facilitate discussion on this topic.

If you are interested in discussing these topics more, please contact the author at aedge@mcginnislaw.com

TABLE OF CONTENTS

IDENTIFYING AND MITIGATING GENERATIVE AI SECURITY AND RELIABILITY ISSUES FOR LEGAL PROFESSIONALS

Abstract: This paper examines key risks in using generative AI for legal practice, focusing on data security and accuracy. It analyzes these risks and proposes mitigation strategies such as vendor evaluation, limiting input of confidential information, output verification, and implementation of comprehensive acceptable use policies. The paper also briefly addresses additional concerns, specifically copyright infringement and bias.

I. OVERVIEW OF GENERATIVE AI

Generative Artificial Intelligence ("genAI") entered the national conversation following the release of OpenAI's enormously popular ChatGPT. ChatGPT was the fastest-growing consumer software application of all time, reaching over 100 million users less than three months after its release. In the interim, several other genAI tools have been released, and ChatGPT and its ilk have been continually expanded and updated, with the notable recent releases of GPT-4o, Claude Sonnet 3.5, and Gemini 1.5 Pro.

These tools have infiltrated workplaces, ranging from employees' informal use of free ChatGPT versions to custom-developed enterprise genAI software. Increasingly, the software tools that workers use every day have genAI functionality, either as an add-on or even as the default option. This availability includes all of the main Microsoft Office applications,[1] as well as the Google equivalents.[2]

Not only is the use of genAI increasingly widespread in business enterprises, it is being used by government agencies as well. The federal ai.gov website collects the wide variety of AI uses being employed by many federal agencies to support their work, including genAI use cases.[3] And, according to the Texas Department of Information Resources, at least a third of Texas state agencies are already making use of genAI.[4]

The legal sector has not been immune. A large number of attorneys have at least experimented with using genAI tools to support their work.[5] And a plethora of legal-specific genAI offerings have entered the market over the past 18 months—these include offerings from Westlaw, Lexis, and Casetext, genAI addons from established software providers such as Relativity, Disco, iManage, as well as scores of startups. As of January 2024, one survey found that over half of law firms had purchased at least one genAI tool.[6]

GenAI tools, whether general-purpose or legal-specific have the numerous potential applications in legal practice. These include various ways to support or even replace attorney work in researching, analyzing, and drafting. These tools may also assist other business functions involved in the practice of law such as billing, marketing, recruitment, and others.

However, most legal organizations are moving cautiously in adopting these tools, cognizant of the associated risks. This paper focuses on two primary risks of genAI use in legal practice: data security and reliability.

II. KEY ISSUES IN USING GENERATIVE AI

A. Data Security

Data security is paramount when using genAI tools in legal practice. Using these tools involves inputting text queries and often analyzing documents or other data. In legal work, many tempting use cases would involve inputting client-confidential or other protected information, which raises significant security risks. Some of these risks are specific to genAI; others are more general data security risks where genAI use just serves as an additional potential risk vector. There are several strategies that legal organizations can use to mitigate, if not eliminate, the data security risks posed by genAI use.

---

[1] https://www.microsoft.com/en-us/microsoft-365/enterprise/copilot-for-microsoft-365

[2] https://workspace.google.com/solutions/ai/

[3] https://ai.gov/ai-use-cases/

[4] https://dir.texas.gov/sites/default/files/2024-06/Accenture_DIR%20Presentation.pdf

[5] According to data published in August 2023, only 15% of lawyers had used genAI tools to support their legal work. https://www.lexisnexis.com/pdf/lexisplus/international-legal-generative-ai-report.pdf. A more recent survey found that 47% had done so. https://www.expertise.com/resources/research/small-business/study-half-lawyers-used-ai-20-percent-dont-fact-check

[6] https://www.lexisnexis.com/community/pressroom/b/news/posts/new-survey-data-from-lexisnexis-points-to-seismic-shifts-in-law-firm-business-models-and-corporate-legal-expectations-due-to-generative-ai

One thing that users must be aware of when using genAI tools is the effect of entering confidential information into these tools. For example, under the default settings, the information a user enters into ChatGPT is both saved on OpenAI servers as part of the user's chat history, and the data may also be incorporated into the training data for future models. This highlights two different data security risks posed by inputting confidential data.

In the first place, the incorporation of confidential information into the model's training data may be considered a disclosure of such information. For example, in April 2023, Samsung employees entered sensitive technical data into ChatGPT, not realizing that the data would be stored. In one instance, an employee pasted confidential source code into the chat to check for errors, while another shared code for optimization. Samsung reacted by banning use of ChatGPT across the organization. While some AI commentators and vendors like OpenAI themselves stress that incorporation of information into training data is not the same as such information being "saved" in the model,[7] this is a hotly contested issue. At least in certain circumstances, certain portions of training data may be retrieved from a model.[8] Therefore, there is a risk that any confidential information used to train a model could later be disclosed to third parties, either inadvertently, or resulting from a targeted extraction.

The safest approach is to avoid inputting any client confidential or other sensitive information into genAI tools. However, in practice this may be less simple than it appears. In the first place, managers in a legal organization are not merely concerned with their own use of genAI, they need to ensure that good data security practices are followed throughout the organization. Also, while some categories of confidential information, like financial data or trade secrets, are self-evidently sensitive, users must be aware of all non-public information they have access to and not only refrain only uploading confidential documents and the like, but also incorporating any of this non-public information in the queries and prompts they enter into these tools. Finally, individual users and organizations must consider whether there is any information that, while technically public, is nonetheless sensitive enough that transmission to genAI vendors or other third parties should be limited.

Another data security risk that must be considered is the risk that input information stored by a genAI vendor might be exposed in a data breach. Data breaches are a distressingly common feature of modern life, with thousands of breaches affecting over 300 million individuals occurring in 2023 alone.[9] Clearly, the risk of data breach is not specific to genAI. However, to the extent that genAI tools store users' confidential information, this category of data is subject to breach. Indeed, an OpenAI data breach in 2023 exposed elements of users chat history, albeit for a brief period.[10] Like the more AI-specific risks associated with incorporation of confidential information into training data, an organization may manage the risks associated with a vendor data breach by limiting the confidential information inputted into a genAI tool and carefully vetting such tools to ensure that their vendors employ sufficient security practices.

B. Mitigating Data Security Risks

To mitigate data security risks associated with generative AI tools, legal professionals should adopt a multi-faceted approach that includes the following strategies:

1. **Check Your Vendors' Security Credentials**: While even the largest and most sophisticated organizations have experienced data breaches, an organization's policies and practices make a huge difference in the risk of a data breach disclosing user information. Your information technology department will be a key resource in evaluating these policies and practices, which should be set out clearly in writing by any reputable vendor. Also, certain certifications, such as SOC-2, may provide a general indication of the robustness of the vendor's security measures.[11] Also, legal organizations should ensure that all vendor agreements contain sufficient warranties and duties concerning data security.
2. **Limit Input of Sensitive Information**: One of the simplest yet most effective strategies is to avoid inputting sensitive or confidential information into AI models. Not only attorneys, but also any support staff using

[7] https://www.merlin.tech/llm-security/

[8] Carlini et al., Extracting Training Data from Large Language Models, https://arxiv.org/abs/2012.07805; https://www.technologyreview.com/2023/02/03/1067786/ai-models-spit-out-photos-of-real-people-and-copyrighted-images/

[9] https://www.idtheftcenter.org/publication/2023-data-breach-report/

[10] https://www.securityweek.com/chatgpt-data-breach-confirmed-as-security-firm-warns-of-vulnerable-component-exploitation/; https://openai.com/index/march-20-chatgpt-outage/

[11] https://secureframe.com/hub/soc-2/what-is-soc-2

genAI tools must be trained to recognize the types of data that should not be shared with genAI tools. Any exceptions to this overall policy must be carefully considered.

3. **Use of Moderation Layers and Audit Logs**: Implementing moderation layers that screen inputs for sensitive information before they are processed by the AI can help prevent data leaks.[12] Additionally, maintaining audit logs of all interactions with AI tools can help track and monitor the use of these tools, ensuring compliance with data security policies.[13]
4. **Redacting Inputs**: Before inputting any data into AI models, sensitive information may be redacted. This can include removing client names, case details, and other identifiers that could compromise confidentiality or replacing these with pseudonyms. However, the redaction strategy has limitations. In the first place, to be effective, such redaction must be comprehensive and executed without any errors. A single piece of identifying information that goes unredacted may result in the entire document or input being identifiable. Also, even if the redaction is performed flawlessly, in some cases enough contextual clues will remain in the unredacted information that the subject and other key information might be identified.[14]
5. **Enterprise Solutions for Secure Access**: Utilizing enterprise-grade AI solutions that offer secure access may materially reduce security concerns. For example, unlike consumer products, the user agreements for enterprise tools like Azure OpenAI and Microsoft Copilot for Microsoft 365, as well Lexis and Westlaw's genAI offerings, all represent that user input data will not be used for model training.[15] These products also generally limit the amount of time user data is stored on the vendor servers, or offer options to limit this storage. Finally, some enterprise solutions do not involve sending data to a third party at all, but rather implement "on-premise" large language models. Legal organizations must carefully research these offerings to determine appropriate products that meet their data security needs.
6. **Opting Out**: Some genAI tools allow users to limit data storage ensuring that the information input into the model is not retained beyond a certain point. For example, ChatGPT users may use the "temporary chats" feature to limit storage of input information and chat history to 30 days.[16] Also, some genAI tools, including ChatGPT, that default to using data inputs as training data allow users to opt out of this use.[17]
7. **Policies and Training**: As discussed below, a well-considered acceptable use policy, supported with mandatory and ongoing training is a key strategy for managing risks associated with genAI. Such policies and training should identify the accepted genAI tools, use cases, and input data needed to meet the legal organization's data security needs.

C. Accuracy

One of the most important concerns of any attorney using genAI tools is the accuracy of the outputs. Many of these tools currently operate through a "chatbot" interface, where users ask questions and the tool provides a response. Such a format is likely to lead inexperienced users to assume that the answer generated by the tool is accurate. Unfortunately, this is often not the case. The term "hallucination" is used to refer to instances where genAI models produce outputs that are factually incorrect, unresponsive, or even nonsensical. And such hallucinations commonly occur when genAI tools are asked legal questions, especially those that are complex or jurisdiction-dependent. A famous example involves a New York lawyer who was sanctioned for submitting a brief containing fabricated case citations generated by ChatGPT.[18] The lawyer had used the AI tool to conduct legal research on bankruptcy issues that were outside his area of specialty. ChatGPT provided him with analysis and several apparent supporting citations. Unfortunately, he incorporated the analysis and citations into a filing without verifying their accuracy. Opposing counsel discovered that the citations were hallucinated, the cases simply did not exist. As a result the lawyer was subject to sanctions. While this is the most well-known instances, scores of similar situations have popped up in courts around

[12] https://www.cyberhaven.com/blog/introducing-cyberhaven-for-ai

[13] https://www.credal.ai/blog/the-benefits-of-ai-audit-logs-for-maximizing-security-and-enterprise-value

[14] https://www.wired.com/story/redact-pdf-online-privacy/

[15] https://www.merlin.tech/llm-security/

[16] https://help.openai.com/en/articles/8914046-temporary-chat-faq

[17] https://help.openai.com/en/articles/7730893-data-controls-faq

[18] https://abovethelaw.com/2023/05/chatgpt-bad-lawyering/; https://www.nytimes.com/2023/06/08/nyregion/lawyer-chatgpt-sanctions.html

the country. This not only demonstrates the risk that genAI outputs may be inaccurate, but also that they are often *persuasively inaccurate*.

In part due to this accuracy issue, many courts have entered standing orders restricting the use of genAI in filings or mandating disclosure where such tools are used.[19] Also, some state bar associations have issued ethical opinions or guidances regarding genAI use, with accuracy (along with confidentiality) being a principal concern.[20] Of course, many commentators have observed that existing ethical rules cover many or all of the accuracy and other concerns posed by genAI, including the duties of competence and diligence.[21] Also, the ABA and others consider that the duty of competence includes a duty to stay reasonably abreast of new technologies affecting the practice of law.[22]

There is widespread confusion concerning the nature of gen AI tools and the mechanisms by which their underlying large language models operate. An in-depth discussion of these topics is beyond the scope of this paper. But in short, hallucination occurs because genAI tools like ChatGPT generate responses based on patterns and probabilities derived from their training data, rather than true understanding or reasoning. While these models can produce highly convincing and coherent text, it does so by predicting what text is "plausible," i.e. what sort of response might be expected to such a query. In many situations, what the model determines is a plausible response is also factually accurate; where these two concepts diverge, there is hallucination. Another issue is that these models are trained on gargantuan datasets, including a vast array of text from the Internet and other sources. Since some of this data contains inaccuracies or incomplete information, the model can generate hallucinated outputs based on these flawed patterns.[23]

Regardless of the source of the hallucination problem, it is significant. Many observers, including myself have noted that the problem seems especially acute when genAI tools are used to research complex legal topics. In part, the value proposition of legal-specific genAI tools like those offered by Westlaw and Lexis is that the models, through incorporation of these vendors' legal databases, would exhibit extremely low rates of hallucination when used to perform legal research.[24] However, a recent study by Stanford's center for Human-Centered Artificial Intelligence found that these tools still exhibited quite significant accuracy problems.[25] When tested on 200 legal research queries, hallucination rates of 17%-33% were observed.[26] While the hallucination rate of general-purpose tools such as ChatGPT were even higher for these queries, these results are unlikely to give attorneys much confidence in the accuracy of genAI tools. While the study has been quite controversial, none of the clarifications or criticisms offered by the vendors or other commentators appear to contradict the study's core finding of serious accuracy problems.[27]

In the face of such research, attorneys might question what use genAI tools could have in supporting their practice. Though it is apparent that these tools are not ready to supplant attorneys' overall legal research efforts, that does not mean that there are not many uses for this technology. Responsible use generally involves using genAI to support more limited tasks, such as generating keywords for a search, initial identification of overall legal doctrines or concepts, or summarizing or synthesizing specific documents. Many of these rely on including a "human in the loop" where trained professionals use genAI thoughtfully and in a targeted manner, verifying and information and outputs before incorporating them into final work product.

---

[19] https://www.ropesgray.com/en/sites/artificial-intelligence-court-order-tracker

[20] https://www.floridabar.org/etopinions/opinion-24-1/; https://www.calbar.ca.gov/Portals/0/documents/ethics/Generative-AI-Practical-Guidance.pdf

[21] https://www.americanbar.org/groups/business_law/resources/business-lawyer/2024-winter/ethics-attorney-responsibility-in-the-age-of-generative-ai/

[22]https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_1_1_competence/comment_on_rule_1_1/

[23] In addition to inaccuracy in the training data, the nature of legal knowledge creates another potential hallucination issue. The answer to a particular legal question often varies significantly across jurisdictions or may have had a different answer at different times. The nature of large language model training is likely to blur these important but subtle distinctions. Therefore, even if a user requests up to date law from a specific jurisdiction, there is a significant risk that out-of-date law from other jurisdictions incorporated in the training data may infect the genAI tool's output.

[24] *See* https://www.lexisnexis.com/community/insights/legal/b/product-features/posts/how-lexis-ai-delivers-hallucination-free-linked-legal-citations; https://legal.thomsonreuters.com/blog/legal-research-meets-generative-ai/

[25] https://hai.stanford.edu/news/ai-trial-legal-models-hallucinate-1-out-6-or-more-benchmarking-queries

[26] *Id.*

[27] https://www.legalcurrent.com/our-commitment-to-our-customers

Also, it is expected that the accuracy of these tools will continue to improve over time. Measures of hallucination tend to show increasing accuracy from more recent and larger models.[28] Thus, legal organizations should periodically reassess the capabilities and accuracy of genAI tools to determine whether their use may be responsibly expanded to additional tasks.

D. Mitigating the Risks of AI Hallucination

To minimize the risks associated with AI hallucination, law departments may employ some of the following strategies:

1. **Verification of AI Outputs**: Always verify the information generated by genAI tools against reliable and authoritative sources. This includes cross-checking legal citations, case summaries, and other critical details with traditional legal research methods and databases.
2. **Ongoing Conversation with AI**: Take advantage of the popular chatbot interface used by many genAI tools by engaging in a back and forth "conversation," allowing the tool to clarify and refine its outputs. If an AI-generated response appears suspect or incomplete, ask follow-up questions to gather more information and context, which can help in identifying hallucinated content.
3. **Retrieval-Augmented Generation**: Use retrieval-augmented generation (RAG) techniques to enhance the accuracy of AI outputs. RAG involves supplementing the AI's responses with information retrieved from reliable sources. For example, tools like Harvey, Westlaw and Lexis integrate databases of specific legal information into the genAI's processes. While this does not eliminate accuracy issues, it does provide a marked reduction of a certain type of hallucination common in the more general-sue tools—the wholesale invention of legal citations that do not exist.
4. **Understanding AI Limitations**: Legal professionals should have a clear understanding of the limitations of AI tools. Recognize that these tools are not infallible and that their outputs should be treated as a starting point rather than definitive answers. This understanding can help in setting appropriate expectations and ensuring careful review of AI-generated content. Also, a more realistic assessment of genAI capabilities allows legal organizations to find more limited ways for these tools to support attorney or staff workflows, rather than trying to completely supplant existing research, analysis, or drafting processes.
5. **Training and Awareness**: Provide training for legal professionals on the risks of AI hallucination and best practices for mitigating these risks. This includes educating attorneys on how to effectively use AI tools, recognize potential hallucinations, and implement verification processes.

E. Additional Risks in Brief: Copyright, Bias, and Compliance

While data security and accuracy are two critical concerns for attorneys in using genAI, these are far from the only risks that need to be considered. While a comprehensive risk analysis is outside the scope of this paper, two other key risks are briefly highlighted below. Please refer to the linked resources for more in-depth information on these risks and mitigation strategies.

1. Copyright Issues

Generative AI tools, such as ChatGPT, are generally trained on large datasets that include large amounts of copyrighted material. This raises significant concerns about potential copyright infringement, as the AI's outputs could inadvertently reproduce protected content without proper authorization.[29] Legal professionals must be aware of these issues to avoid liability.

A key concern is that using AI-generated content based on copyrighted material without permission could result in legal challenges. For example, if an AI tool generates a document that closely mirrors a copyrighted text, the creator of the original content could seek legal recourse against the user for infringement. While there have not been fully litigated instances of such end-user infringement claims to date, many commentators have identified this as a potential risk, and the magnitude of potential liability is unclear.

To address these concerns, many companies have made commitments to assume responsibility for copyright claims related to their AI tools. Microsoft's copyright commitment is a representative example; it summarizes the

[28] *See* https://github.com/vectara/hallucination-leaderboard

[29] Whether the process of model training itself constitutes copyright infringement is a separate legal issue. While this issue has generated a large amount of commentary and litigation, since the liability risk is borne by the model's creator rather than the end-user, it is outside the scope of this paper.

indemnification offered to users as follows, "As customers ask whether they can use Microsoft's Copilot services and the output they generate without worrying about copyright claims, we are providing a straightforward answer: yes, you can, and if you are challenged on copyright grounds, we will assume responsibility for the potential legal risks involved."[30] However, commentators have noted several limitations in these sort of indemnification clauses.[31] Therefore attorneys must still exercise caution in using AI-generated content. The willingness of genAI companies to provide broad indemnification to millions of users for infringement liability does provide some indication that these companies and their attorneys do not believe that there is a significant amount of end-user infringement liability resulting from the use of these tools, at least where users are not specifically seeking to use the tools to reproduce copyrighted works.

There have been many useful articles published related to copyright and genAI. The law journal article Building and Using Generative Models Under US Copyright Law provides a very in-depth review of the copyright issues raised by genAI, including a detailed description of how the technical aspects of large language models relate to copyright arguments.[32] The Congressional Research Service's article Generative Artificial Intelligence and Copyright Law provides a more succinct but still informative primer.[33] Please refer to these resources for a more in-depth analysis of the issue.

2. Potential Bias

Since genAI tools are built using large language models trained on vast datasets, any inherent biases contained in information in these datasets may be incorporated into the models. These biases can lead to the AI generating outputs that reflect societal prejudices or discriminatory viewpoints.

In legal practice, bias in AI tools can result in unintentional discrimination, inaccurate legal analysis, and ethical breaches. For instance, an AI might generate biased legal advice or draft documents that unfairly favor certain groups, undermining the fairness and impartiality crucial in legal work. This not only affects the quality of legal services but also erodes trust in AI tools among legal professionals. Of course, this concern is even more heightened with respect to work product being used by government agencies, given the stringent Constitutional, statutory, and regulatory requirements for governmental entities to act in a non-discriminatory manner.

A recent joint report by Thompson Reuters and Duke University entitled Addressing Bias In AI is a critical resource for users of AI technology, including genAI tools, to increase their awareness of potential bias in the use of these tools.[34] The Harvard Business Review article What Do We Do About Biases In AI provides a great analysis of the problem of bias in machine learning and artificial intelligence more generally.[35] These are good materials to review to further explore how your legal organization may mitigate the risks of bias in using genAI.

III. ENTERPRISE ACCEPTABLE USE POLICIES FOR GENERATIVE AI

Several of the strategies set out above to mitigate risks associated with genAI use require coordination across the legal organization. Any legal organization's strategy for safer and more effective genAI use should include the careful consideration of formal policies related to genAI. The most fundamental of such policies is an acceptable use policy governing the organization's use of genAI.

It is important for legal organizations to carefully consider the specifics of a potential acceptable use policy for genAI to craft a policy that is appropriate for their organizational needs, clients, and reflects their risk tolerance. However, there are some general components that organizations should consider, as follows:

1. **Guidelines for Data Input and Tool Use**: Clearly outline what types of data can be input into AI tools and emphasize the prohibition of sensitive or confidential information. Organizations should also consider a process to evaluate and approve specific genAI tools and either allow only those tools to be used, or specify that those tools must be used for higher risk applications, e.g. uses that will create work product will involve the use of more sensitive data.

---

[30] https://blogs.microsoft.com/on-the-issues/2023/09/07/copilot-copyright-commitment-ai-legal-concerns/

[31] https://www.law.com/legaltechnews/2024/02/09/gen-ai-providers-offer-ip-indemnity-heres-why-its-not-fool-proof/

[32] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4464001

[33] https://crsreports.congress.gov/product/pdf/LSB/LSB10922

[34] https://www.thomsonreuters.com/en-us/posts/wp-content/uploads/sites/20/2023/08/Addressing-Bias-in-AI-Report.pdf

[35] https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai

2. **Training Requirements**: Mandate training for employees on the safe and effective use of genAI tools, including data security practices and the importance of verifying AI outputs. Completing this training should be a required precondition before an employee is authorized to use genAI tools in the workplace.
3. **Audit and Monitoring**: Establish procedures for auditing and monitoring the use of AI tools to ensure compliance with organizational policies and identify any potential misuse.
4. **Responsibility and Accountability**: Define the roles and responsibilities of employees and management in using AI tools, including who is accountable for verifying the accuracy of AI-generated outputs. In a legal organization, this responsibility may be shared by the information technology department, the loss management partner or organization's inside counsel, and the organization's technology committee, if applicable.
5. **Legal and Ethical Considerations**: Incorporate legal and ethical guidelines to ensure that the use of AI tools aligns with the organization's values, client requirements, professional responsibility standards, including bar requirements, and relevant laws and regulations (e.g. HIPAA).

Acceptable use policy templates and associated guidance are widely available.[36] These may provide a good starting point for developing a policy.

While a carefully considered and robust acceptable use policy is a good start, legal organizations must take further steps for such a policy to be effective in practice. As mentioned above, a policy should be supported by a fully-developed training program and potentially monitoring and audit systems. Also, a clear governance structure and decision-making authority will allow for more flexible implementation of the policy; an attractive feature given the quickly-developing nature of this technology.

In crafting and implementing a policy, the organization must consider not only its own goals, but also client policies and its duties under client representation agreements, as well as relevant court rules, bar guidance, and applicable laws and regulations. Given that all of these outside factors will change over time, it is essential for the organization to keep abreast of these changes and periodically revise its policy accordingly.

The U.S. Office of Personnel Management has released a guidance for federal employees entitled Responsible Use of Generative Artificial Intelligence for the Federal Workforce.[37] This guidance is relatively general, and for the most part defers to individual government agencies to set genAI acceptable use policies. Some Texas governmental agencies are also developing acceptable use policies as well as genAI implementation strategies.[38] Even where governmental clients do not have explicit requirements for their attorneys concerning genAI use, government attorneys may look to the use policies of governmental unit or agency clients as guidance for determining appropriate genAI use in supporting these clients.

IV. CONCLUSION

GenAI tools hold great promise for supporting legal practice by automating routine tasks and enhancing efficiency. However, the adoption of these tools must be approached with caution due to potential risks, including significant risks related to data security and accuracy. By implementing robust data security measures, developing comprehensive acceptable use policies, and implementing robust training programs, legal organizations can harness the benefits of genAI while mitigating its risks. As the technology continues to evolve, staying informed and adaptable will be key to leveraging these technologies effectively and responsibly in the legal field.

[36] https://www.aihr.com/blog/generative-ai-policy/; https://legalnodes.com/template/generative-ai-company-use-policy; https://fpf.org/wp-content/uploads/2023/07/Generative-AI-Checklist.pdf

[37] https://www.opm.gov/data/resources/ai-guidance/

[38] https://www.kxan.com/news/texas-politics/4-texas-agencies-detail-how-theyre-already-using-artificial-intelligence/

APPENDIX A: GLOSSARY

Acceptable Use Policy: A set of rules and guidelines that outline how an organization's network, website, or systems may be used.

Artificial Intelligence (AI): Computer systems designed to perform tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and language translation.

Audit Log: A chronological record of system activities to enable the reconstruction and examination of the sequence of events and changes in an organization's digital environment.

Bias (in AI context): Systematic errors in AI outputs that can result in unfair or discriminatory outcomes, often reflecting societal prejudices present in training data.

ChatGPT: A popular generative AI chatbot developed by OpenAI, capable of engaging in human-like conversations and performing various text-based tasks.

Data Breach: An incident where confidential, sensitive, or protected information is accessed, stolen, or exposed by an unauthorized party.

Data Security: Protective measures implemented to secure data against unauthorized access, corruption, or theft throughout its lifecycle.

Enterprise Software: Application software designed to meet the needs of an entire organization rather than individual users.

Generative AI (genAI): A type of AI that can create new content, including text, images, or code, based on patterns learned from existing data.

Hallucination (in AI context): When an AI model generates false or nonsensical information that appears plausible but has no basis in its training data or factual reality.

Large Language Model (LLM): An AI model trained on vast amounts of text data, capable of understanding and generating human-like text.

Machine Learning: A subset of AI that focuses on the development of algorithms and statistical models that enable computer systems to improve their performance on a specific task through experience.

Moderation Layer: A system or process that reviews and filters content to ensure it meets specified guidelines before it's processed or published.

On-premise: Referring to software or hardware that is installed and run on computers on the premises of the organization using the software, rather than at a remote facility.

Retrieval-Augmented Generation (RAG): An AI technique that enhances language model outputs by incorporating information retrieved from external sources.

SOC-2: A voluntary compliance standard for service organizations, developed by the American Institute of CPAs (AICPA), which specifies how organizations should manage customer data.

Training Data: The dataset used to teach an AI model patterns and information, forming the basis of its knowledge and capabilities.