AI for GCs: What You Need to Know for 2024

By: Matt Todd, Reece Clark, Cat Kozlowski and Adam Garcia

I. Introduction

Beginning in early 2023, with the publicity and public launch of open-source and powerful and easily applied generative artificial intelligence ("AI") tools, the interest and requests for guidance on these tools have exploded. With the spotlight has come swift change. We have been met with some surprises along the way, most notably the accelerated adoption of generative AI tools across a wide variety of applications and use cases. Never have we witnessed such a rapid adoption rate of a technology that has so many legal, business, technical, ethical, social and other considerations.

This unprecedented adoption has driven the evolving nature of both our clients' concerns and our advice. Having witnessed this evolution, we now want to (a) highlight a new and emerging issue for General Counsels ("GCs") to consider, (b) provide some interim analysis and forecast on applicable regulation and legislation, and (c) provide GCs with a framework for making "AI decisions" when presented with either a new tool or a new use case. Although there has been a veritable avalanche of AI-related content from law firms, technologists and other pundits around the world, it is our hope that our insight on these three key issues is useful for GCs in the most practical sense and provides some tools which they can deploy.

II. Overreliance on AI: A New Issue for GCs to Manage

While industry adoption of generative AI is still in the early days, organizations have already begun experimenting with these tools in their respective verticals.[i] This has led to some initial insights into the pitfalls of AI use (e.g., generative AI "hallucinations").[ii] In 2024, as AI continues to be operationalized at enterprise scale, GCs will need to be guarded against a new, persistent risk related to AI use: *overreliance*.

Overreliance arises when a user misunderstands the results of an AI tool or lacks understanding of how the AI tool works.[iii] Users may not, for example, be fully aware of what the AI tool can (and cannot) do, how it performs relative to the typical job functions of a human worker, or simply how the AI tool derived its results in the first place. At the same time, users may fall victim to biases of their own and the AI tools' models, and of training data that exacerbate issues of overreliance. These include automation bias (tendency to favor AI recommendations), confirmation bias (tendency to favor information derived from AI that reaffirms prior assumptions or beliefs), ordering effects (AI results that are presented early in a workflow may cause a user to anchor to those results) or overestimating the explanations provided by AI (forcing an AI to explain its results may increase the "blind trust" effect of the results).[iv]

In general, the threat of overreliance is highest where a user fails to understand how the AI tool works but recognizes that it has provided accurate results on more than one occasion. This creates a lulling effect that reduces a user's sense of urgency in validating the results upon each use. The user may find they have unwittingly accepted an incorrect recommendation, or the user may consciously or unconsciously switch their answer (in whole or in part) to match an AI recommendation even if the user had previously derived a different answer.[v]

Why is overreliance a problem? Consider the risk to people, profit and performance. In 2023, generative AI has most notably been used in marketing and sales, product and service development, and service operations.[vi] These business areas are critical to the bottom line and ensuring organization health and success. Consider the liabilities a company may face if, for example, AI incorrectly described product or service functionality (e.g., to an individual end consumer or other customer in connection with marketing materials, product labelling or contract negotiation) and overcommitted the organization, over/undersold the product or service, or gave rise to injury or damage. Or consider if an AI-derived result steers product development in a direction that is not supportable in the market or, worse, causes a latent product liability issue. These real-world examples illustrate the need to ensure accuracy and human oversight in the use of AI. AI tools may help reduce the transaction cost of information exchanges (e.g., with individual end consumers or between companies during sales, negotiation, support, maintenance and other stages), but must be monitored closely for alignment with organizational values.[vii]

What can an organization do? A popular option is to force the AI to explain how it derived a result. Yet GCs should be aware that even if an AI can provide an explanation, a user may not rigorously—or at all—check the explanation. Early studies suggest that users tend to ignore complex or lengthy explanations in favor of accepting the results blindly.[viii] As the level of complexity in the task undertaken by AI increases, so too does the complexity of the accompanying explanation. A user may believe the mere existence of the explanation (whether verified or not) will provide sufficient support. As a result, GCs need to be appropriately guarded and support policy and norm-setting exercises that rigorously evaluate the results of AI tools and AI-produced explanations.

III. Developing AI Regulations and What GCs Need to Know

As organizations incorporate AI into their business and operational processes, GCs must carefully navigate the litany of federal laws, initiatives and proposed regulations applicable to AI. Likewise, state laws and regulations impose additional requirements on organizations for specific data types.

There is no comprehensive U.S. federal scheme governing AI use.[ix] Instead, there is a patchwork of sector-specific consumer protection federal laws and regulations implicating AI. For instance, the Consumer Financial Protection Bureau ("CFPB") has opined that using AI could violate the Equal Credit Opportunity Act ("ECOA") where creditors rely on but do not fully understand how the AI's algorithms or "black box" elements function when they deny credit applications.[x] Additionally, the AI may discriminate or (knowingly or unknowingly) issue biased results based on race, sex or religion; creditors that exhibit automation bias violate the ECOA.[xi] The Equal Employment Opportunity Commission ("EEOC") has also released guidance indicating that employers may be liable for relying on AI that disparately impacts or discriminates against some protected classes, thereby violating Title VII of the Civil Rights Act.[xii]

Meanwhile, the Securities and Exchange Commission ("SEC") and the Federal Trade Commission ("FTC") have the authority to regulate business practices related to their agency objectives,[xiii] which includes using AI in deceptive or fraudulent business practices. But their respective regulations do not readily account for intent—these agencies may rely solely on objective evidence to determine whether the AI results were deceptive regardless of how the AI was designed or used. For instance, it may be deceptive when organizations promote their use of AI to lure consumers into investment opportunities but where the investments' returns are highly misrepresented.[xiv] And misleading business practices in one instance can invite further scrutiny to an entire industry, similar to historic data collection practices.[xv]

Even if an organization complies with current applicable laws and regulations, recent federal and state initiatives may subject organizations to future regulations. The Biden Administration issued an executive order directing numerous agencies and departments to publish regulations, standards and guidelines to promote AI safety, security, data privacy, and equity and civil rights.[xvi] This executive order may drive significant changes within the federal government with respect to the use of AI tools in agency and administrative operations.[xvii] States have also created advisory councils or ordered state agencies to study and monitor how AI is used in the public and private sectors and develop policies and procedures based on those findings.[xviii] Collectively, these initiatives signal what operational and legal standards and requirements GCs should consider for their organization's use of AI.

Lastly, GCs must be cognizant of state consumer privacy laws and industry-specific regulations. For example, the consumer privacy laws in California, Colorado, Connecticut and Virginia provide consumers with the right to opt out of automated processing.[xix] Other states also regulate how AI is used in conducting job interviews.[xx] Overall, compliance may feel like a moving target, and it is. With the EU AI Act set to take effect soon, GCs must also monitor regulatory compliance abroad.[xxi] GCs may view the potential regulations yet to come, or standards from specific industries, as the sword of Damocles that will exacerbate the compliance burden. In the face of these challenges, how does a GC advise and guide in determining whether to adopt the latest new AI tool?

IV. Evaluating AI Tools and Establishing a Method of Trust

Many aspects of AI (and particularly generative AI) are currently on unsettled ground, and early adopters may find themselves using a particular AI tool that ceases to exist a year (or five) later. Alternatively, late adopters may find themselves years behind their competitors. While the regulations and legal analyses develop with regulatory authorities and in courts, GCs can break the practical analysis into five parts:

1. What is the tool?
2. What is the use case?
3. What is the data going into it?
4. What is the output?
5. Is it accurate?

1: What is the tool?

When reviewing a novel AI tool the organization wants to use, the first question is whether it is even actually AI, using underlying machine learning. As previously noted, the FTC is already battling the misuse of the term du jour, and an effective decision tree simply does not have the same risk profile as real AI with machine learning and related models underlying it.

If the tool does in fact use machine learning, what type of model is it? Is it part of the newly exploding wave of generative AI models? Or a predictive model that has been around and in use for well over a decade?

On what data was the tool's model trained? Are there any intellectual property concerns that are currently in active litigation or likely to arise? What about bias inherited into the model from a skewed training dataset?

What are the terms and conditions? Is it a public tool, open source or an enterprise instance? Does anyone else have access to the tool's output, and if so, how might it be used? What protections is the vendor providing if, for example, the organization receives an IP infringement claim arising from use or distribution of the output content generated by the tool?[xxii]

For a risk-averse company, the analysis may end here because the legality of how most current foundational tools and underlying models were trained is currently in active litigation,[xxiii] which can potentially impact anything generated from them, or the continued support or existence of the tool.

2: What is the use case?

AI does not solve all problems, and not all problems need AI. Evaluate the use case to which the tool will be applied and whether it is even an appropriate one for AI to address. Current/existing AI tools and their models are, at their core, prediction machines—aids relying on mathematical statistics, probabilities and correlations (not reasoning or certainty). The accuracy of that prediction can vary depending on multiple factors, as can the tolerance for, and type of, error within a use case. Detecting whether there is a bird versus a plane in an image for auto text generation has a high tolerance for error; detecting whether an abnormality in an MRI is cancerous has a far lower tolerance. Additionally, the type of inaccuracy or error—e.g., a false positive or false negative—is often critical to understanding the risks, biases and benefits. Some errors are inevitable (or even built into certain tools, machines and processes) and can be tolerated and addressed through additional processes.

Some use cases also simply do not work well with AI because individual human judgment or empathy may be necessary. An AI may share the probability of rain, but it does not know how bothered each individual may be about getting a little wet or drenched.[xxiv] Consider also whether the tool is sufficiently transparent for the use case. Again, referring to the weather example, it is unlikely that users would need to understand how the AI reached its prediction. In contrast, transparency in decision-making is critical in the employment space, and liability for determining hirings and firings may still fall on the employer's (and not the tool developer's) shoulders.[xxv]

Finally, use cases may be impacted by external factors such as legislation or industry trends. If companies use AI tools to process compliance activities, for example, overreliance presents legal risk to a company if there are errors in the results. Or consider further how overreliance on an AI tool may lead to using the tool in a setting for which it was not designed. In such a case, error rates may increase, but blind trust may cause those errors to go unnoticed. Considering use cases and managing expected outcomes is critical to prevent overreliance on an AI tool that may be well suited for some tasks but not others.

3. What is the data going into it?

Data privacy and confidentiality concerns should be top of mind for GCs when reviewing a new tool. What type of information will go through the tool? Public information? Trade secrets? Will the vendor have any rights to that information as training data? Do customers have their own enterprise instance of the tool, or is that data and/or feedback flowing through a public instance? Even with a public instance, is there any risk of sensitive or trade secret data circulating through the tool showing up in a future output or influencing another customer's result?[xxvi]

4. What is the output?

The "power" of an AI's output, or risk for overreliance, can also depend on its format. Does the tool produce a report to be further analyzed by humans, an answer or decision, or perhaps a binary "yes" or "no" with little to no transparency into the probability threshold? Risk may also depend on who receives that output. Is it inward facing, for reference or additional context, or for further review? Or is it outward-facing, a result given to customers that can potentially influence their choices, even if they lack expertise in the tool's subject area?

5. Is it accurate?

Accuracy is and will be the most critical factor in analyzing any AI tool. The tool's accuracy must meet or exceed applicable thresholds based on the application (otherwise, it could be more problematic than beneficial). Presumed accuracy in an AI tool is related to the concerns arising from overreliance: as the presumed level of accuracy in an AI tool increases, so does the threat of overreliance. Where an AI tool appears more accurate than not, the level of effort to check results degrades. To prevent blind trust, accuracy in AI results must not be presumed; rather, there should always be a "trust but verify" mentality that confirms accuracy and reinforces the users' understanding of the AI tool and the potential errors that may arise in use. This collectively reduces a user's tendency to blindly accept the results without further confirmation and reduces overreliance risk.

V. Conclusion

When considering the nascent regulatory field for use of generative AI in business and the potential pitfalls of AI use—most notably for 2024, overreliance and related biases—this article demonstrates that GCs will need to not only engage in norm-setting exercises to manage the use of AI in business processes but also establish a framework for AI use that reduces risk and error. That can be accomplished, in part, by using the five-factor framework in section IV above to evaluate AI tools, results and explanations objectively and critically. But internal business process management is not enough. We also anticipate GCs to implement similar checks and balances in their vendor management procedures to ensure their suppliers are conforming their services to the same rigor and safeguards when using generative AI in service delivery. A comprehensive, balanced approach will be needed as AI technology, regulations and industry-specific considerations continue to evolve. Polsinelli attorneys will continue to closely monitor the developments in AI legal frameworks and regulations and will be at the forefront in delivering timely insights to our clients.

[i] Michael Chui, *The State of AI in 2023: Generative AI's Breakout Year*, McKinsey (Aug. 1, 2023), https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year.

[ii] *What are AI hallucinations*, IBM, https://www.ibm.com/topics/ai-hallucinations (last visited Dec. 20, 2023).

[iii] Samir Passi & Mihaela Vorvoreanu, *Overreliance on AI: Literature Review*, Microsoft (June 21, 2022), https://www.microsoft.com/en-us/research/uploads/prod/2022/06/Aether-Overreliance-on-AI-Review-Final-6.21.22.pdf.

[iv] *Id.* at 11.

[v] *Id.* at 3.

[vi] Chui, *supra* note 1.

[vii] *See* Reece Clark, *Frictionless Contracting In A COVID-19 Economy: Part 2*, Law360 (July 20, 2020) https://www.law360.com/articles/1291286 *citing* Carl J. Dahlman, The Problem of Externality, 22 J. L. & Econ. 141, 144 (1979) ("Once the parties decide to transact, they must convey enough information to one another such that each can arrive at a reasonably agreeable bargain. This often takes time where the parties are sophisticated, and may involve external resources for information gathering.").

[viii] Katherine Miller, *AI Overreliance Is a Problem. Are Explanations a Solution?,* Stanford Univ. (Mar. 13, 2023) https://hai.stanford.edu/news/ai-overreliance-problem-are-explanations-solution.

[ix] *See* Adam A. Garcia, Note, *Socially Private: Striking a Balance Between Social Media and Data Privacy*, Iowa L. Rev. 319, 329-35 (2021) (illustrating the complexities in asserting privacy rights).

[x] *Consumer Financial Protection Circular 2022-2023*, CFPB (May 26, 2022), https://www.consumerfinance.gov/compliance/circulars/circular-2022-03-adverse-action-notification-requirements-in-connection-with-credit-decisions-based-on-complex-algorithms.

[xi] *Id.*

[xii] *Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII of the Civil Rights Act of 1964*, EEOC (May 18, 2023), https://www.eeoc.gov/laws/guidance/select-issues-assessing-adverse-impact-software-algorithms-and-artificial.

[xiii] The SEC regulates any unlawful activity connected with purchasing or selling any security where a person directly or indirectly used a device to defraud or engage in any deceitful practice. *See* 17 C.F.R. § 240.10(b)-5. The FTC monitors for unfair or deceptive acts affecting commerce, and such acts are unlawful when they cause substantial injury to consumers. *See* 15 U.S. Code § 45(a)(2) (2023).

[xiv] *See FTC v. Automators LLC et al.*, No. 3:23-cv-01444 (S.D. Cal. Aug 08, 2023).

[xv] *See* Garcia, *supra* at 10, at 339-46.

[xvi] Exec. Order No. 14,110, 88 C.F.R. 75191 (2023). *See also Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, White House (Oct. 30, 2023), https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.

[xvii] As a lodestar for the potential diffusion of AI technologies within federal government agencies, consider Exec. Order No. 13520 and subsequent legislative and regulatory guidance on the introduction and use of advanced information technologies to reduce fraud and improper payments. *See* Reece Clark, Note, *Kafkaesque Dangers: IPERIA, Do Not Pay, and the Government's New Fight Against Improper Payments*, 102 IOWA L. REV. 1719, 1722 (2017). *See also* Exec. Order No. 13520, 74 C.F.R. 62201 (2009).

[xviii] *Artificial Intelligence 2023 Legislation*, Nat'l Conf. of State Legislatures (last updated on Sept. 27, 2023), https://www.ncsl.org/technology-and-communication/artificial-intelligence-2023-legislation.

[xix] *See* CAL. CIVIL CODE §§ 1798.140(z), 1798.185(a)(16) (2023); Colo. Rev. Stat. §§ 6-1-1303(20), 6-1-13-6(1)(a)(C) (2023); 2022 CONN. PUBLIC ACT 22-15 §§ 1(22), 4(a)(5); VA. CODE ANN. §§ 59.1-575, 59.1-577(A)(5) (2023). Organizations will likely have to ensure their operations include mechanisms to receive and accommodate such requests.

[xx] In Illinois, the Artificial Intelligence Video Interview Act requires organizations to provide notice to job applicants prior to using such technology and explain how the technology will be used. 820 ILL. COMP. STAT. 42/1 (2023). Maryland also requires organizations to obtain consent prior to using facial recognition technology in job interviews. 2020 MD. LAWS., MD. CODE, LAB. & EMPL. § 3-717 (2023).

[xxi]*See The Act*, EU ARTIFICIAL INTELLIGENCE ACT, https://artificialintelligenceact.eu/the-act/ (last visited Dec. 12, 2023); Aaron M. Levine, *Is the EU AI Act Faltering*, POLSINELLI: PUBLICATIONS (Nov. 29, 2023), https://www.polsinelli.com/publications/is-the-eu-ai-act-faltering; and Aaron M. Levine, *The EU AI Act, The World's First Comprehensive AI Regulatory Scheme*, POLSINELLI: PUBLICATIONS (Dec. 12, 2023), https://www.polsinelli.com/publications/the-eu-ai-act-the-worlds-first-comprehensive-ai-regulatory-scheme.

[xxii] *Introducing the Microsoft Copilot Copyright*, MICROSOFT, https://www.microsoft.com/en-us/licensing/news/microsoft-copilot-copyright-commitment (last visited Dec. 1, 2023).

[xxiii] *See, e.g.*, Getty Images (U.S.), Inc. v. Stability AI, Inc. No. 1:23-cv-00135, at *1-4 (D. Del. Feb. 03, 2023).

[xxiv] Ajay Agrawal, Joshua Gans & Avi Goldfarb, *How Large Language Models Reflect Human Judgment*, HARV. BUS. REV. (June 12, 2023), https://hbr.org/2023/06/how-large-language-models-reflect-human-judgment.

[xxv] N.Y. COMP. CODES R. & REGS. tit. 20, §§ 20-870–20-874 (2023).

[xxvi] Milad Nasr et al., *Scalable Extraction of Training Data from (Production) Language Models*, ARXIV (Nov. 28, 2023), https://arxiv.org/pdf/2311.17035.pdf?ref=404media.co (prompting large language models to repeat the word "company" eventually returned the email address and phone number of a random law firm, and other similarly styled prompts returned phone numbers, emails and birthdays).