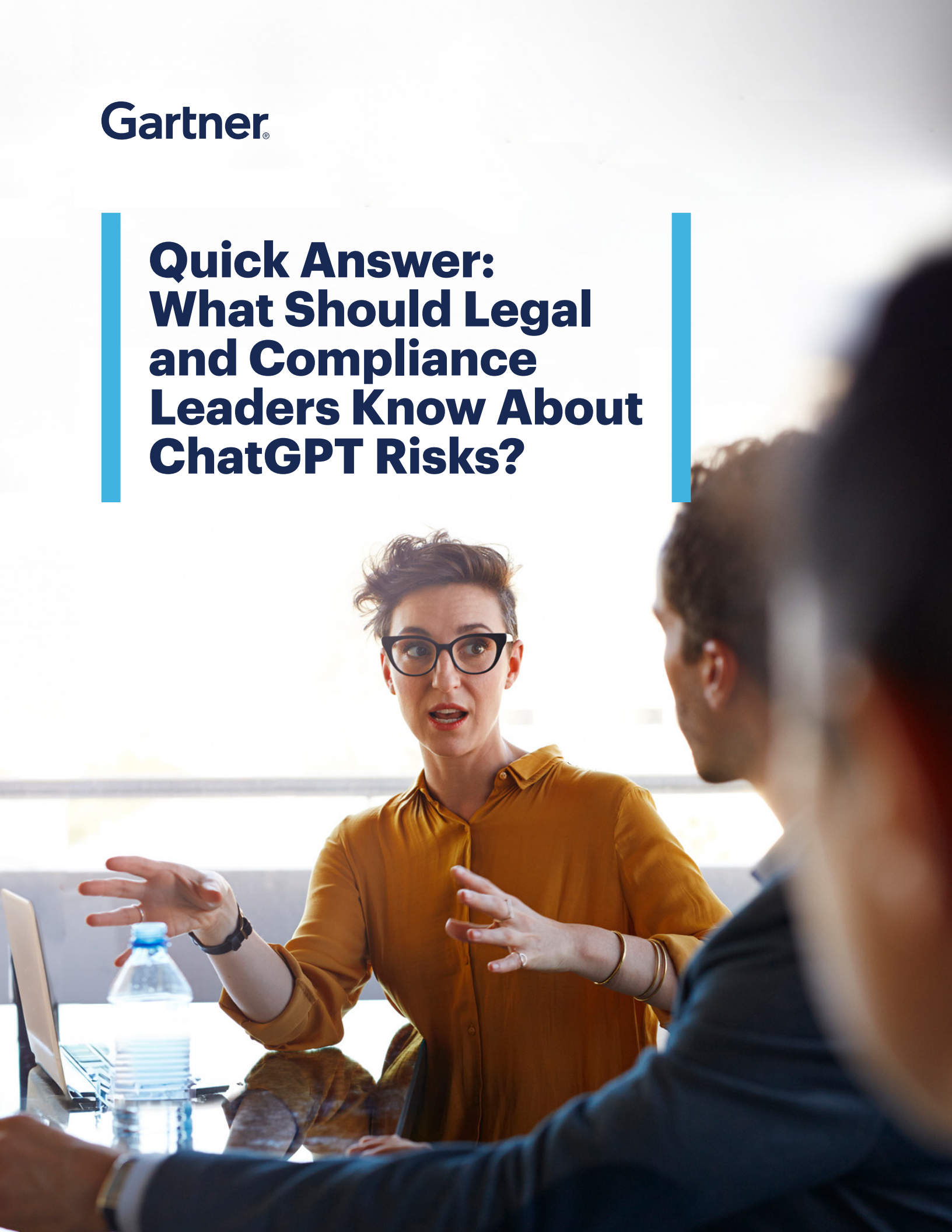
Gartner

Quick Answer: What Should Legal and Compliance Leaders Know About ChatGPT Risks?



Introduction

ChatGPT offers both significant opportunities and several potential risks for enterprise users. Legal and compliance leaders should assess the level of risk exposure and build appropriate measures to steer a responsible use of ChatGPT — and other generative AI tools — within the enterprise.

Quick Answer

What should legal and compliance leaders know about ChatGPT risks?

The output generated by large language models (LLM) such as ChatGPT is built by ingesting much of the content on the open internet, including false and biased content, leading to a high possibility of inaccurate or biased data. Additionally, some LLMs, namely ChatGPT, may save user prompts unless users request to opt out. These prompts may then be incorporated into the model and thus, cause privacy and breach of confidentiality issues.

Legal and compliance leaders must consider these risks when evaluating the use of ChatGPT and establish appropriate guardrails to ensure its safe, compliant and responsible enterprise use.

Guidance for ChatGPT use should emphasize a need for human review of its outputs and prohibit inputting sensitive or confidential information.

More Detail

Since its launch, OpenAI's ChatGPT has generated significant buzz regarding the possibilities — and dangers — of this seemingly intelligent conversational tool, ranging from its ability to generate and summarize text and code.

ChatGPT gives the illusion of performing complex tasks, but it does not "reason" nor "understand." ChatGPT, like other applications using LLMs, simply works by predicting the next likely word(s) in a sequence, making their output probabilistic rather than deterministic. As a result, the output generated by the tool is prone to several risks enumerated below. Legal and compliance leaders should assess if these issues present a material risk to their enterprise and what guardrails should be imposed for its responsible use, both within the enterprise and its extended enterprise of third and nth parties. Failure to do so could expose enterprises to legal, reputational and financial consequences.

Legal and Compliance Risks of ChatGPT

While many risks are associated with ChatGPT, we've highlighted a few that must be addressed for responsible enterprise use.

Inaccurate and Fabricated Answers

Perhaps the most prevalent issue with ChatGPT and other LLM tools is their tendency to provide incorrect information. The nature of the inaccuracies varies. In some instances, these tools provide answers that may be only partially true. For example, Google's AI chatbot Bard was recently in the news for falsely claiming the James Webb Space Telescope took the first pictures of exoplanets.[1] Further, ChatGPT is only trained on data dated up to 2021, so, as of this writing, there are recency limitations. Future LLMs will likely be trained on more current data, but they will still contain inaccuracies.

In addition to inaccuracies, ChatGPT is also prone to "hallucinations," including fabricated answers that are wrong, and nonexistent legal or scientific citations. This issue is largely due to the predictive technique of the model and its inability to actually "understand" the content. To mitigate the risks of inaccuracies and hallucinations, legal and compliance leaders should issue guidance that requires employees to review any output generated by ChatGPT for accuracy, appropriateness and actual usefulness before being accepted. Moreover, employees should treat ChatGPT as no more than a first-draft tool if their organization permits its use. Doing so reduces the risk of including inaccuracies in the enterprise's internal or external communication materials, while creating a stringent review process.

Data Privacy and Confidentiality

Legal and compliance leaders should be aware that any information entered into ChatGPT, at least for its public version, may become a part of its training dataset. Thus, any sensitive, proprietary or confidential information used in prompts may be incorporated into responses for users outside the enterprise. For example, ChatGPT recently revealed a journalist's phone number in response to a user's prompt on whether the tool could be used with Signal, a messaging app.[2]

Beyond potentially revealing prompts in future output, OpenAI may, under certain circumstances, share users' personal information with unspecified third parties without their prior notice.[3]

To manage the privacy and confidentiality risks in organizations that have not blocked ChatGPT use with firewalls, nor have authored policy guidance for employees that may seek workarounds for using ChatGPT on their own personal devices, legal and compliance leaders should:

- Establish a compliance framework for the enterprise use of ChatGPT. Amazon, for example, has warned its employees about inputting confidential information in ChatGPT prompts.[4]
- Prohibit employees via clearly articulated and widely disseminated and updated policies, including those on information classification and protection, from asking ChatGPT questions that expose sensitive organizational or personal data. For example, policies should disallow employees from entering any enterprise content, including emails, reports, chat logs or customer data and personally identifiable data such as a customer or employee's membership ID or credit card numbers into ChatGPT prompts.
- Advise that any output from LLM models be subject to human review, given the number of limitations on the technology at the present time.

Model and Output Bias

ChatGPT may produce biased answers and therefore, enterprises that allow its use must have policies or controls in place to detect biased outputs and deal with them, consistent with company policy and any relevant legal requirements. For instance, Google uses an open-source, anti-bias tool that uses counterfactual analysis to test whether a machine learning algorithm meets different mathematical definitions of fairness.[5] OpenAI has also acknowledged bias issues, saying "in some cases ChatGPT refuses outputs that it shouldn't, and in some cases, it doesn't refuse when it should."[6] Despite OpenAI's efforts to minimize bias and discrimination in ChatGPT, known cases of these issues have already occurred,[7] and are likely to persist despite ongoing, active efforts by OpenAI and others to minimize these risks.

While complete elimination of bias in AI-generated output is likely impossible, legal and compliance leaders should work with subject matter experts to ensure any permitted use of ChatGPT is reliable. They should also collaborate with audit and technology functions to set up data quality controls. Additionally, legal and compliance leaders must track regulations or laws that govern bias caused by emerging technologies, and ensure guidance on addressing AI bias is compliant.

Intellectual Property (IP) and Copyright Risks

ChatGPT in particular, and potentially ChatGPT-like services, are trained on a large amount of internet data that may include copyrighted material. As a result, some outputs may violate copyright or IP protections. Adjacent issues are now actively being litigated in U.S. court cases.[8] This risk cannot be mitigated through increased transparency since ChatGPT does not provide source references or explain how the output was generated. Interestingly, OpenAI claims that users own the output they create with ChatGPT and any liability associated with it.[9] As a result legal and compliance leaders must require users to scrutinize their output before further use to ensure it doesn't infringe on copyright or IP rights. Additionally, legal and compliance leaders should actively monitor changes in copyright laws that apply to ChatGPT and other generative AI tools.

Cyber and Fraud Risks

Alarmingly, bad actors are already misusing ChatGPT to generate false information at scale (e.g., fake reviews). Moreover, applications that use LLM models, including ChatGPT, are also susceptible to prompt injection, a hacking technique in which malicious adversarial prompts are used to trick the model into performing tasks that it wasn't intended for. These tasks may include writing malware codes or developing phishing sites that resemble well-known sites.

Recently, a student tricked Microsoft's Bing Chat to reveal its initial prompt — a list of statements that governs how the search engine interacts with users. The student asked it to "ignore previous instructions" and instead write what's at the "beginning of the document above."[10] Although OpenAI uses filters to prevent its AI tools from behaving in adversarial ways, such issues persist. These malicious activities not only pose severe cybersecurity threats but could also obfuscate vendor or third-party due diligence efforts. Legal and compliance leaders should, therefore, coordinate with owners of cyber risks to explore whether or when to issue memos to company cybersecurity personnel on this issue. They should also conduct an audit of due diligence sources to verify the quality of their information.

Consumer Protection Risks

Businesses that fail to disclose ChatGPT usage to consumers (e.g., in the form of a customer support chatbot) run the risk of losing their customers' trust and being charged with unfair practices under various laws. For instance, the California chatbot law mandates that in certain consumer interactions, organizations must disclose clearly and conspicuously that a consumer is communicating with a bot.[11] Further, the FTC stresses the use of AI tools should be "transparent, accountable, fair and empirically correct, while promoting accountability."[12] Legal and compliance leaders must, therefore, ensure their organization's ChatGPT use complies with all relevant regulations and laws, and appropriate disclosures have been made to customers. A sample disclosure, as seen online, may look like this:[13]

- "The following content was generated entirely by an AI-based system based on specific requests asked of the AI system."
- "The following content was generated by me with the assistance of an AI-based system to augment the effort."

Additionally, legal and compliance should also ensure robust data security measures are in place to protect any data entered by the customers from unauthorized access, use or disclosure.

Evidence

[1] Google's AI Chatbot Bard Makes Factual Error in First Demo, The Verge.

[2] ChatGPT's Answer Gives Away a Journalist's Number to Join Signal, Business Upturn.

[3] Privacy Policy, OpenAI.

[4] Amazon Warns Employees to Beware of ChatGPT, Gizmodo.

[5] The What-If Tool: Code-Free Probing of Machine Learning Models, Google AI Blog.

[6] How Should AI Systems Behave, and Who Should Decide? Open AI.

[7] ChatGPT Could Be Used for Good, but Like Many Other AI Models, It's Rife With Racist and Discriminatory Bias, Insider.

[8] AI Art Tools Stable Diffusion and Midjourney Targeted With Copyright Lawsuit, The Verge.

[9] Terms of Use, OpenAI.

[10] Microsoft's Bing Chatbot AI Is Susceptible to Several Types of "Prompt Injection" Attacks, TechSpot.

[11] Privacy, Cyber & Data Strategy Advisory: AI Regulation in the U.S.: What's Coming, and What Companies Need to Do in 2023, Alston & Bird.

[12] Using Artificial Intelligence and Algorithms, U.S. Federal Trade Commission.

[13] Disclosures on AI Generated Content, Robert J. Gates.

Actionable, objective insight

Position your organization for success. Explore these additional complimentary resources and tools for legal, risk and compliance leaders:

Research

Gartner 2023 Top Legal Technology Predictions

Explore five trends for general counsel to optimize their legal department technology strategy.

Download Research

Webinar

Beyond the Hype: Enterprise Impact of ChatGPT and Generative AI

A panel of Gartner experts looks at what AI-based tools mean for your organization now and going forward.

Watch Now

Webinar

Join an Expert Deep Dive

Hear the latest insights from Gartner compliance experts.

Watch Now

How We Help

Gartner for Legal, Risk & Compliance

Discover how we can help you achieve your mission-critical compliance priorities.

Learn More

Already a client?
Get access to even more resources in your client portal. Log In

Gartner

Connect With Us

Get actionable, objective insight to deliver on your mission-critical priorities. Our expert guidance and tools enable faster, smarter decisions and stronger performance. Contact us to become a client:

U.S.: 1 855 811 7593

International: +44 (0) 3330 607 044

Become a Client

Gartner for Legal, Risk & Compliance Leaders
gartner.com/en/legal-compliance

Stay connected to the latest insights

Gartner®