Debevoise & Plimpton

Debevoise Update D&P

Does Your Company Need a ChatGPT Pilot Program? Probably.

March 21, 2023

Last month, we wrote about how many companies probably need a policy for Generative AI tools like ChatGPT, Bard and Claude (which we collectively refer to as "ChatGPT"). We discussed how employees were using ChatGPT for work (e.g., for fact-checking, first drafts, editing documents, generating ideas and coding) and the various risks of allowing all employees at a company to use ChatGPT without any restrictions (e.g., quality control, contractual, privacy, consumer protection, intellectual property, and vendor management risks). We then provided some suggestions for ways that companies could reduce these risks, including having a ChatGPT policy that organizes ChatGPT use cases into three categories: (1) uses that are prohibited; (2) uses that are permitted with some restrictions, such as labeling, training, and monitoring; and (3) uses that are generally permitted without any restrictions.

Since then, it has become clearer that many companies are not ready to implement a formal ChatGPT policy because (1) they have not yet adequately assessed which use cases should and should not be allowed (and if allowed, what restrictions, if any, should apply), and (2) they have not developed the governance that is needed to administer their desired ChatGPT policy. Some companies are therefore launching a Generative AI Pilot Program for several weeks that is effectively serving as an interim ChatGPT policy and that includes:

- **Scope**: Some companies are including only ChatGPT in their Pilot Program, while others are including all Generative AI tools or some other group of AI tools.

- **Technical Restrictions**: Some companies are blocking the use of ChatGPT for all employees for work that is not part of the Pilot Program until a list of generally permitted use cases is approved by an AI Governance Committee. Other companies are allowing all employees to use ChatGPT for certain limited low-risk purposes, but are using proxies to monitor what employees are inputting into ChatGPT and blocking access to ChatGPT for employees who are using it in a way that is not permitted as part of the Program (e.g., including confidential client information in their inputs).

- **Committee**: Creating a cross-functional AI Governance Committee to determine use cases that are prohibited and use cases that are permitted. For the permitted use cases, deciding which restrictions or guardrails, if any, should apply.

- **List of Uses**: Creating a regularly updated internal webpage that lists both the permitted and prohibited ChatGPT uses. That internal webpage could expressly indicate that if employees have a potential use case that does not appear on the permitted or prohibited list, they may submit a proposal to have that use case tested by a designated beta-tester (as described below) and, if approved by the AI Governance Committee, added to the List of Permitted Uses.

- **Testers**: Designating a group of beta-testers to receive requests for approval for potential uses cases and to test them. The testers may receive some training on the risks associated with using ChatGPT for work-related tasks (e.g., confidentiality, IP, quality control, reputational risks, etc.). After they have tested a proposed use case, they can:

    - approve that use case for the requester and recommend that the AI Committee add the use case to the List of Permitted Uses, either with or without guardrails;

    - deny the request and recommend that the AI Committee add the use case to the List of Prohibited Uses; or

    - allow the use case with conditions, such as requiring that all queries be run by beta testers and results being routed back to the requester.

- **Guardrails**: Restrictions that the AI Committee could place on approved uses cases may include:

    - **Authorized Users**: Only certain designated employees are approved for that specific use case.

    - **Trained Users**: Only employees that have received specific training on the risks associated with that use case are approved for that use case.

    - **Labeling**: Marking the outputs for that specific use case as having been generated, either in whole or in part, by ChatGPT.

    - **Human Review**: Requiring that the output of each approved use case be reviewed and approved for its designated use by someone with subject-matter expertise to ensure that the accuracy and quality of the output is fit for purpose.

- **Recording the Input**: Documenting and maintaining a record of the inputs for that specific use case that generated the corresponding outputs.

- **Routing through Beta-Testers**: Requiring non-beta-testers who want to use ChatGPT for a specific use case to have their proposed prompts sent to one of the beta-testers, who then will decide whether it is an acceptable prompt and, if approved, the beta-tester will send the output back to the requester.

- **Use Limitations**: Requiring that the output from the specific use case be used only for certain purposes (e.g., for internal training) but cannot be sent to a client or shared outside the company.

- **Licensing**: Exploring obtaining a company instance of the Generative AI tool from the company that developed it. For some of these tools, it is possible to secure an enterprise license that provides a closed-loop instance, whereby data that is input into the tool by the licensee is not shared with the licensor and is not added to the training set for the tool, which reduces many of the confidentiality risks associated with the use of the free public versions of some of these Generative AI tools.

- **Announcing**: Drafting a communication to employees announcing the Pilot Program.

There are many variations of these protocols. In order to simplify the Pilot Program and reduce the governance burden, some companies are collapsing the testers and the AI Governance Committee into a single group of individuals within the company. Those designated employees are authorized to review, test and approve use cases, both on an individual basis and categorically for their company, although they are encouraged (and in some cases required) to consult or reach consensus with the other testers before they can add a specific use case to the permitted or prohibited list.

It can be difficult to get the right balance among: (1) establishing an orderly review and approval process to make sure that ChatGPT is being used responsibly and is not creating significant risks for the company; (2) allowing employees to experiment with Generative AI tools to find low-risk/high-value use cases that significantly improve productivity and quality; and (3) creating a ChatGPT policy that is fair, understandable, actionable and does not require a burdensome compliance/governance structure. Running a ChatGPT Pilot Program for several weeks can help companies find the right balance for their organization.

To subscribe to the Data Blog, please click here.

The [Debevoise Artificial Intelligence Regulatory Tracker](#) ("DART") is now available for clients to help them quickly assess and comply with their current and anticipated AI-related legal obligations, including municipal, state, federal and international requirements.

The cover art used in this blog post was generated by DALL-E.

* * *

Please do not hesitate to contact us with any questions.

NEW YORK

Megan K. Bannigan
Partner
mkbannigan@debevoise.com

Avi Gesser
Partner
agesser@debevoise.com

Henry Lebowitz
Partner
hlebowitz@debevoise.com

Benjamin Leb
Associate
bjleb@debevoise.com

Jarrett Lewis
Associate
jxlewis@debevoise.com

Melissa Muse
Associate
mmuse@debevoise.com

Michael R. Roberts
Associate
mrroberts@debevoise.com

Lex Gaillard
Law Clerk
adgaillard@debevoise.com