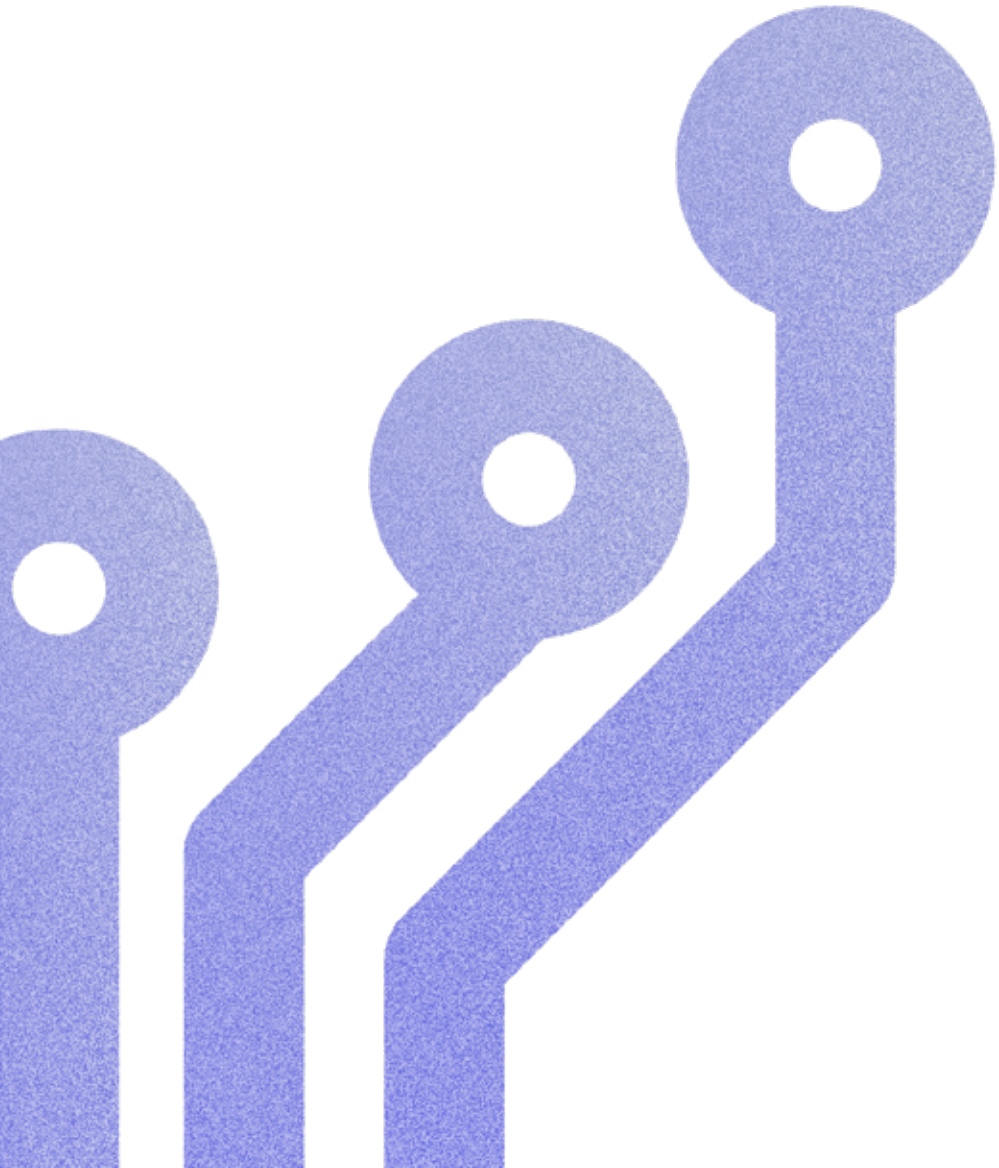
FS-ISAC

# Framework of an Acceptable Use Policy for External Generative AI

*Illustrative for Adaptation*



September 2023

# Contents

## Introduction

Trust underpins the financial services industry. Customers trust financial institutions with their financial data and transactions. Companies trust their employees to make decisions every day with the company's data, reputation, and more. Financial firms trust each other to honor commitments in trades, transfers, and other transactions.

To maintain this trust, financial firms design and implement policies and controls that enable employees to make good decisions and adhere to relevant regulations. One type of policy is known as Acceptable Use, which outlines good risk management and security practices on specific systems and technologies.

Generative AI is one such technology that has the potential to revolutionize every industry. It is a powerful driver of optimization, efficiency, and cost reduction as well as the basis for new business lines and products. It will be integrated into our companies at all levels. But there are many risks that come with it, and financial firms must be proactive in managing internal adoption and use of generative AI.

This framework is a guide for firms to design their own Acceptable Use policy for external generative AI. Given the rapid development and adoption of generative AI, we hope this guide serves as a helpful tool for firms to upgrade their security and risk management policies to incorporate safe and responsible AI use into their security programs and beyond.

Some argue that financial institutions (FIs) should take a stringent approach and block external generative AI systems, as these are still nascent, untested, and unvetted. Others believe that employees may find workarounds to blocking these systems, and so it would be more productive and indeed, more secure, to educate employees on how to safely use them. This framework offers policy guidance on both permissive and stringent approaches, allowing firms to decide the right balance for themselves.

What follows is a short explanatory text followed by sample policy text (labeled "Policy Guidance") that firms can adapt for their own use as they see fit.

## Policy Guidance

### ▸ Introduction

This policy defines requirements for the acceptable use of external generative AI services. This policy describes management's directive to:

> Ensure protection of the company's intellectual property

> Ensure that use of these systems reflects the culture and ethics of the company, as well as the regulatory, privacy, and legal obligations of the company and its employees

> Establish a baseline of proper use of these systems for all employees

> Ensure compliance with all applicable laws, rules, and regulations, including privacy requirements such as General Data Protection Regulation (GDPR)

> Ensure compliance with the terms of use of many generative AI systems (especially attribution and lack of copyright protection)

The policy assumes as an overarching principle that data loss risks are present in the use of generative AI systems, like any other third-party system.

This document applies in all manners of consumption, such as via API, UI, or any other interface, as well as in all manners of access, including both corporate and personal devices.

## Policy Guidance

### ▸ Defining External Generative AI

This policy is for generative AI services running external to the firm (as opposed to services hosted within a firm), irrespective of whether the firm is using the paid or free versions of the services. Examples of external generative AI include, but not limited to, all versions of:

- Text (and related) generation, such as large language models (LLMs), including:

  > OpenAI's ChatGPT

  > Microsoft's Bing with GPT integration

  > Google's Bard

- Microsoft O365 integration with OpenAI CoPilot
- Microsoft GitHub CoPilot
- BloombergGPT

- Image generation, such as latent diffusion models (LDMs), including:
  - Stable Diffusion
  - Midjourney AI
  - OpenAI's DALL-E
- Other multimedia creation tools (e.g., "deepfake" tools)

## Data Provided to Generative AI Systems

### Confidentiality

Currently, most external generative AI systems use queries for future training (although some systems allow opt-outs). The service may preserve the queries, which hackers could breach and release. Queries with sensitive data put companies at significant risk.

| Policy Guidance | |
|---|---|
| ▶ Permissive | ▶ Stringent |
| Employees must take care to preserve the confidentiality of the company's work products above any uses of generative AI systems. Employees must adhere to this and all company policies and standards.<br><br>Employees must consider:<br>> The impact, importance, or uniqueness of the company's intellectual property (IP) before sharing it. Consider generalizing and stripping any company | Employees must respect the confidentiality of the company's work products above any uses of generative AI systems. Employees must adhere to this and all company policies and standards.<br><br>Employees must not share:<br>> The company's internal IP, especially proprietary, copyrighted, or related IP<br>> Any data having Personally Identifiable |

references, especially if the data is proprietary or copyrighted

- Limiting or obfuscating the data if it has Personally Identifiable Information (PII), Nonpublic Personal Information (NPI), or other data under the purview of reg-ulators or similar entities
- Repurposing or reducing extraneous information that if exposed publicly:
- Might tarnish the reputation of the company
- Could open the company to regulatory or legal action
- May allow reverse engineering
- May give access to IP or systems
- A reminder to not share (in Generative AI or any form externally) any information that, if exposed publicly, could reveal the company's strategy and/or would be in violation of safe harbor statements

Information, Nonpublic Personal Information, or other data under the purview of regulators or similar entities.[1]

- Any information that if exposed publicly might tarnish the reputation of the company, or information that would open the company to regulatory or legal action
- Any information that someone may use to reverse engineer or give access to IP or systems
- Any information that if exposed publicly could reveal the company's strategy and/or would be in violation of safe harbor statements

*1. This can be adjusted if, through contractual agreements, the firm has the keys to their data vault and can protect the data held at the service.*

## Responsibility

Although generative AI has the power to increase cyber crime through convincing mimicking of existing communications, excellent translation capabilities, deep fake images, audio, and video, ease of finding code vulnerabilities, and much more, the kinds of threats still are the same as when using other third-party systems. For instance, phishing emails may become more personalized with generative AI, but the way these enter the organization still is the same. The responsibility to safeguard against malicious activity when using generative AI is therefore same as general internet usage to protect IP and other concerns such as compliance.

### Policy Guidance

#### ▶ Permissive

Generative AI systems are no different from any third-party system. They bring massive benefits, if used responsibly. They can do harm too, so employees need to take care of the queries and inputs to these systems.

Employees should consider the following as part of that responsible use:

- That the accuracy of these systems is still suspect; they have been known to "hallucinate" (see "Accuracy" section below)
- Remove references to IP, NPI, PII
- Whether the firm's name needs mentioning
- If sharing code, genericize any variable names
- Genericize the request, making it harder to attribute back to the firm
- Not share any of the aspects mentioned in the confidentiality section

#### ▶ Stringent

Acceptable use of these systems is still unclear, given the accuracy of the output (see "Accuracy" section below). To use them responsibly, at a minimum, employees must genericize the queries or inputs and remove any reference to the company or its IP, akin to asking a question on communities such as Stack Overflow or Discord.[2]

If allowed to access these systems in a limited way, employees should:

- Remove any references to IP, NPI, PII
- Remove any reference to the firm
- If sharing code, genericize any variable names
- Genericize the request to be as abstract as possible
- Not share any of the aspects mentioned in the confidentiality section

*2. Other policies may prohibit the use of Stack Overflow and related systems though, and if so, those prohibited use cases should also apply to generative AI systems.*

## Access

This section elucidates the range of options for controlling access to generative AI systems, including guidance for which staff and from what devices they can access, as well as whether to use corporate identities when using them, given that it is possible that such queries become public. Also, firms should publish a list of approved generative AI systems, based on its assessment of the vendor's compliance with laws, rules, and regulations. (FS-ISAC AI Risk Vendor Risk Subgroup is working on a questionnaire to aid this assessment.)

| Policy Guidance | |
|---|---|
| ▶ Permissive | ▶ Stringent |
| All employees can access approved external generative AI systems from desktops, mobile phones, and similar corporate devices. Firms must limit access to generative AI systems from production servers when not part of a service or application workflow.<br><br>Employees can use their corporate identity (email or related) as a login to these systems. | Employees must not access generative AI services from any corporate device or server.<br><br>Employees must not use their corporate identity (email or related) as a login to these systems. |

### Monitoring

Monitoring is a common section in Acceptable Use Policies, trusting but verifying compliance with the policy. Firms need to consider what they intercept, monitor, limit, etc.

| Policy Guidance | |
|---|---|
| ▶ Permissive | ▶ Stringent |
| The company has the right to monitor the use of these systems per applicable laws and regulations. Management may occasionally verify that the use of these systems adheres to the directives outlined in this document.<br><br>Users should inform their management at once if the use of these systems may have infringed on the standards in this policy for acceptable use. | The company will monitor the use of these systems per applicable laws and regulations. Management will ensure that the use of these systems adheres to the directives outlined in this document.<br><br>Users should inform their management at once if the use of these systems may have infringed on the standards in this policy for acceptable use. |

## Data Received from Generative AI Systems

### Accuracy

Generative AI systems are far from perfect, especially with fact-based output. These systems may "hallucinate" when their answers appear convincing but are completely wrong. Users should not rely on their accuracy.

Incorrect answers can cause severe issues for companies. For example, inaccurate instructions for IT administrators may lead to data loss or other system damage. Distortion of statistics or other facts may lead to employees making public claims or key decisions based on erroneous information.

Other considerations:

- Since generative AI may not be current on legislation, answers on accounting, tax, or other legal considerations may be inaccurate
- Generative AI cannot know organizational culture or intricacies of industries, leading to suggestions that may do more harm than good.

- Depending on how users ask the questions, generative AI may not apply GDPR or other applicable laws to the answers given.
- Employees can use these systems for nefarious purposes, such as proxy avoidance, unapproved network uses, or manipulating financial services transactions for personal gain

## Policy Guidance

### ▶ Permissive

Given that the accuracy of generative AI is imperfect, the firm strongly encourages employees to manually verify the output before use. Employees should ensure the output does not lead to:

- Fraudulent, destructive, and/or inappropriate system usage
- Poor, biased, and/or unethical business practices or decisions
- Reputational and/or financial harm of the company

The firm also recommends employees:

- Ensure compliance with applicable regulations, laws, and acceptable fair use of others' copyright protections
- Avoid circumventing or bypassing this or other company policies

Akin to any other third-party solution, if the generative AI system results seem suspect, it is best to verify and consider alternative sources.

### ▶ Stringent

Employees must assume the output accuracy of generative AI systems are imperfect, and therefore must manually check the accuracy of the output (especially with regards to fact-based queries). Employees must check that the output does not lead to:

- Fraudulent, destructive, and/or inappropriate system usage
- Poor, biased, and/or unethical business practices or decisions
- Reputational and/or financial harm of the company

Employees must also:

- Ensure compliance with applicable regulations, laws, and acceptable fair use of others' copyright protections.
- Not circumvent or bypass this or other company policies

When there is doubt around the accuracy, completeness, or copyright protections of the output from a generative AI system, the employee must not use the output and, instead, rely on other sources of information.

## Representation as the Employee or Company

External generative AI systems often require users to give attribution to the system for any output, for example, in the Open AI's Terms of Use. Knowing this stipulation, it could:

- Invalidate some uses of the output for proprietary or copyrighted work
- Open users to copyright violations
- Open users to conflicts with other firms potentially using the same or similar output

As generative AI becomes increasingly powerful, organizations will need to reevaluate what they consider "acceptable" use of such systems. No policy document, permissive or stringent, can cover all situations. For example, a manager short on time uses a generative AI system to write an employee appraisal. That employee does not like their appraisal and files a complaint. Since an actual human did not write the appraisal, there is good reason to believe that the employee's grievances would be upheld.

The onus is on the user to attribute the output correctly.

| Policy Guidance | |
|---|---|
| ▶ Permissive | ▶ Stringent |
| There are considerations for employees before they use generative AI output as their own:<br>> In internal documents on behalf of the employee (emails, HR documents, et cetera.), employees must remember that the output stands for them. Employees must take care to remove sensitive, proprietary, or confidential/secret company information. Employees should also check whether the output is in the desired tone and language<br>> Employees must note that there is greater attribution, legal, ethical, and copyright concerns in external communications and perform stronger reviews in those uses | Employees must not use the output of generative AI systems to stand for activity they undertook. Employees must specifically not use the output of generative AI to:<br>> Write any internal documents on behalf of the employee (emails, HR documents, et cetera.), especially those having sensitive, proprietary, or confidential/secret company information<br>> Write any external communication |

## Attribution

Several commercial generative AI tools require attribution in their Terms of Use. Given that users often do not attribute output to AI, firms may be liable for violations of these terms. Consumers of these systems cannot (typically) treat the output as their proprietary IP, so use of this output can only advise and guide users.

Specific for programming code uses of generative AI, the input to the base of many of these systems is open-source software (OSS) code found in common repositories. Some firms limit the open-source licenses used in their products and need to verify that the output comes from an OSS license they approved.

While there are tools that try to detect if a generative AI system developed a specific output, they are not (yet) dependable enough to find whether an employee's work is original or not. This document assumes you cannot use such tools, but that may change.

| Policy Guidance | |
|---|---|
| ▶ Permissive | ▶ Stringent |
| Any output from generative AI systems must follow the policies of the various external systems, specifically where they require supplying attribution to those systems.<br><br>Employees should take care that the use of generative AI systems does not compromise the firm's external reputation, even when supplying attribution. Additionally, employees must respect and honor any applicable open-source licenses. | Any output from generative AI systems must follow the policies of the various external systems, specifically where they require supplying attribution to those systems. The firm will block or limit the use of systems that require attribution.<br><br>Employees must not engage in any use of generative AI systems that could compromise the firm's external reputation, even when supplying attribution. Additionally, employees must strictly follow any applicable open-source licenses. |