Fujitsu Generative AI Guidelines

January, 2024 ver1.1

Introduction

To stakeholders who read this document

This guideline for the use of Generative AI has been prepared by the Fujitsu Group for our employees.
It explains the general risks of Generative AI and examples of countermeasures especially from an ethical and legal perspective.
Our company employees take appropriate risk mitigation measures in accordance with this guideline and provide safe and secure Generative AI services to society.

The Fujitsu Group has decided to make this guideline available not only to our company employees but also to stakeholders to help society in understanding how to use Generative AI.

We hope it will be useful for you.

Introduction P.2

About Generative AI
- Generative AI Overview P.4
- Generative AI Use Cases P.5
- Reference: How does ChatGPT works? P.6

Risks Brought by Generative AI
1. Accuracy P.9
2. Fairness P.11
3. Copyright infringement P.13
4. Information management P.16
5. Misuse P.19

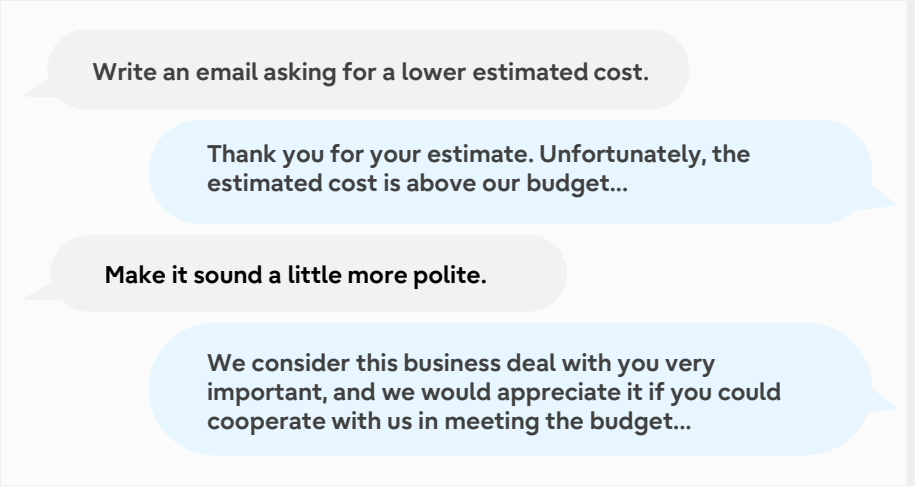
Summary: Using Generative AI in Fujitsu P.22
For people considering business use of Generative AI P.23

Reference: Considerations for AI Code Generator P.24

Generative AI Overview

"Generative AI" refers to a type of **AI that generates human-like natural outputs** which hardly seem to be auto-generated, in response to simple prompts such as words and sentences.

Its generating capacity **has already reached a indistinguishable level from that of human intellectual activity.**



Anyone can generate content quickly and easily without the experience, expertise, or technology.

Generative AI Use Cases

Generative AI is capable of generating a variety of content such as text, program code, image and video, music.
For example, it can be used in a wide range of **tasks that require intellectual creation, such as:**

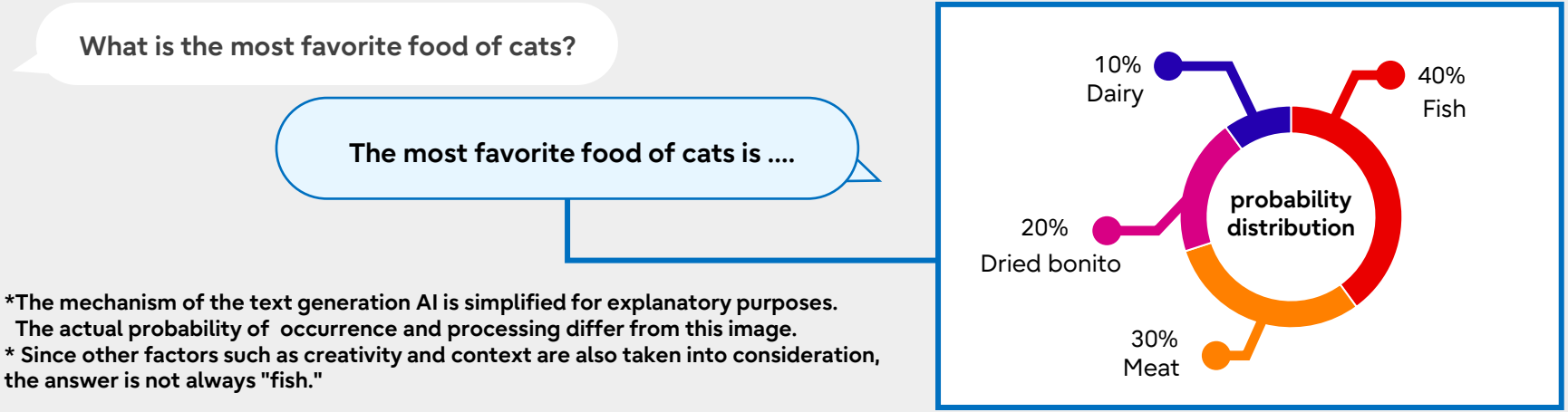
	Text	Program code	Image and video	Music
Rule-based simple task	Grammar check and proofreading	Debugging	Noise Removal, High Resolution	Trimming, Loop Creation
Improvement in quality	Summaries, Translations	Program structure improvement Avoid Errors/Bugs	Monochrome → Color Image generation Illustration - Photo/Movie generation	Arrangement of an existing song
Assistance for content creation	Writing idea proposal	Module code generation	Photo and video generation from text	Melody and rhythm generation
Content generation	Thesis and story generation	App Design Document generation Proprietary app generation	Photobook generation Comic book generation	Creating original music styles and genres

Reference: How does ChatGPT* work?

*OpenAI's interactive AI chatbot

As of fiscal year 2023, ChatGPT is the leading text generation AI tool. Its overwhelmingly huge amount of knowledge (more than several hundreds of GB of training data and hundreds of billions of parameters) offers a **significantly broader range of expression**.

ChatGPT predicts the word most likely to follow the previous word, based on the past training data and other sources.
It generates text by repeating this prediction process. It is important note that ChatGPT just predicts and generates the most probable combination of words. It does not verify the factual accuracy of its output.



***The mechanism of the text generation AI is simplified for explanatory purposes.**
The actual probability of occurrence and processing differ from this image.
*** Since other factors such as creativity and context are also taken into consideration, the answer is not always "fish."**

Risks Brought by Generative AI

Generative AI is a very useful tool, but it involves legal and ethical risks such as copyright infringement and information leaks. However, these risks are **not unique that suddenly arose with the introduction of Generative AI; they have existed in conventional AI as well.**

Risks and countermeasures are summarized mainly from the viewpoint of Generative AI users, assumed use within the scope of Japanese law.

Risks

Accuracy

Fairness

Copyright infringement

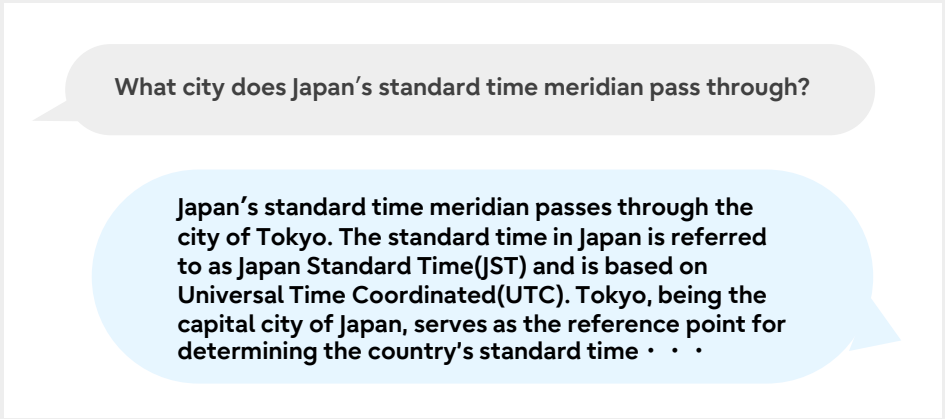
Information management

Exploit

1. Accuracy

The training data for Generative AI often includes a massive amount of web content that has not yet determined to be correct or false. In some cases, Generative AI are trained only on outdated data. Also, as mentioned on page 7, Text Generation AI such as ChatGPT generates text by repeating "predicting probabilities," during the generation process, it does not verify the factual accuracy of its output. As a result, **errors due to lack of "accuracy"** can appear in the product.

Example of Text Generation AI



In fact, Japan's standard time meridian passes through **"Akashi City, Hyogo Prefecture" in the Kansai region**.

At a glance, we have **plausible sentences written in natural English**. But we need to be careful when using this information.

1. Accuracy

If you rely too much on AI to automatically generate text and code, **you might be under the illusion that you'll always only get accurate outputs by AI.** Offering its outputs without human's judgment of accuracy or adequate checks can lead to **lose the trust of customers and society**. In some cases, you may be held responsible for damage.

Points to note

When using outputs of Generative AI, make sure to check its accuracy.

For example, double-checking on a trusted website
Reference: Fujitsu launches new technologies to protect conversational AI from hallucinations and adversarial attacks

When you need factual information, **avoid using prompts that may generate "unknown information"** whose accuracy is difficult to judge.

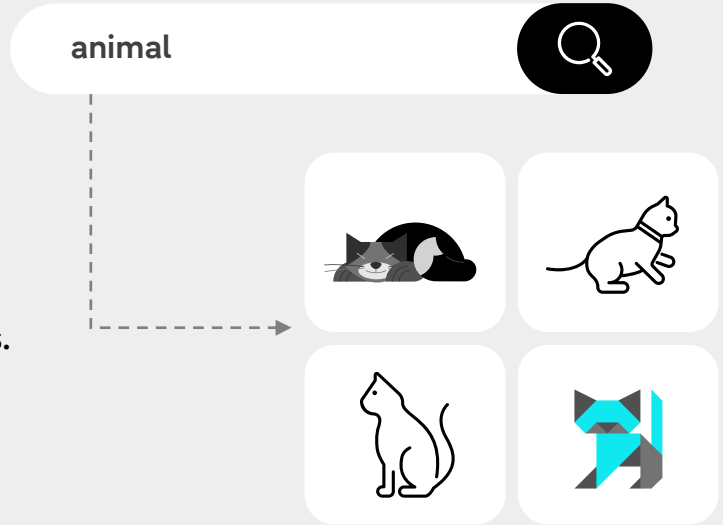
***Be careful about derogatory or discriminatory expressions.**

2. Fairness

Generative AI **can produce unfair results,** mainly due to biases and prejudices which inherent in the data and algorithms it learns.

Example of Image Generation AI

- Typing "**CEO**" generates only **white males**
- Entering "**nurse**" generates only **women**
- Typing "**Japanese**" generates only **people wearing "kimono"**.
- Typing "**animal**" generates only **cats**



As you can see above, you may face racial and gender biases. Some Generative AI is trying to improve biases, but we still need to be careful about outputs that eliminate diversity.

2. Fairness

Points to note

When using outputs of Generative AI, make sure there are no biases or prejudices involved.

Check for potential disadvantages or prejudices in terms of race, sex, age, etc.
Research the similarity of AI outputs biases to previous cases, etc.

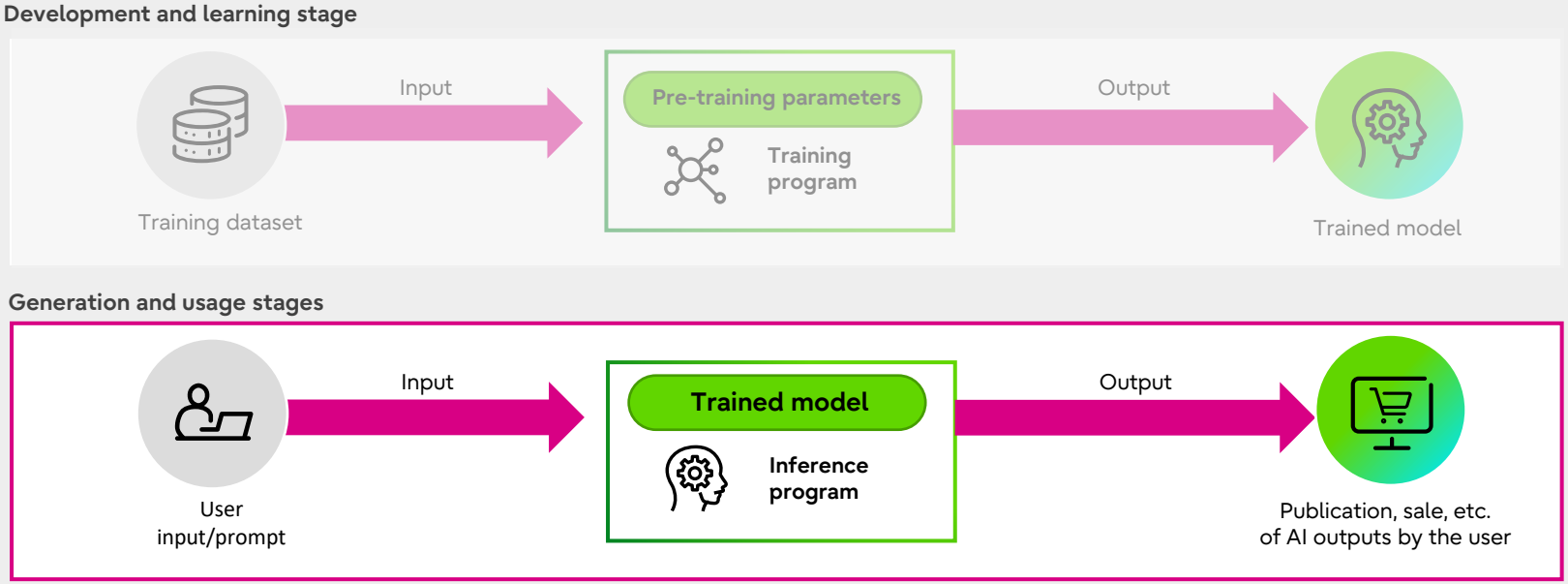
Use equity-conscious AI tools and services.

Check that developers are working to ensure fairness and mitigate biases, etc.

3. Copyright infringement

Generative AI generates new contents by learning other people's copyrighted materials on the Internet, such as text and images. As a result, its outputs may infringe the copyright of a third party. On considering the relationship between Generative AI and copyrights, it is desirable to separate the "development and learning stages" and "generation and usage stages" because the point to discuss differs in these stages.

This document deals with the risks in the "generation and usage stages."



3. Copyright infringement

In the stages of using Generative AI, outputs may be considered to violate a copyright under Japan's Copyright Act if it meets the following two requirements.

Similarity The output of is **"identical" or "similar" to an existing copyrighted work**.

Reliance The output is generated based on an **existing copyrighted work**.

There is not yet a definite social consensus on whether the results produced by Generative AI can be thought to "rely" on its training data.
Given this lack of consensus on the "reliance," **it is important to take every precaution to avoid the risk of "similarity."**

*** Even if these two requirements are met, the output may be considered not to infringe a copyright in some cases (when the user has obtained permission from the copyright owner, when the term of copyright protection has expired, etc.).**

3. Copyright infringement

Points to note

Avoid using prompts that involve copyrighted works whose reproduction or modification is prohibited.

* For example, avoid including the name of an author, the title of a book, or the like in a prompt, as in "Create an image that looks like (an existing character)

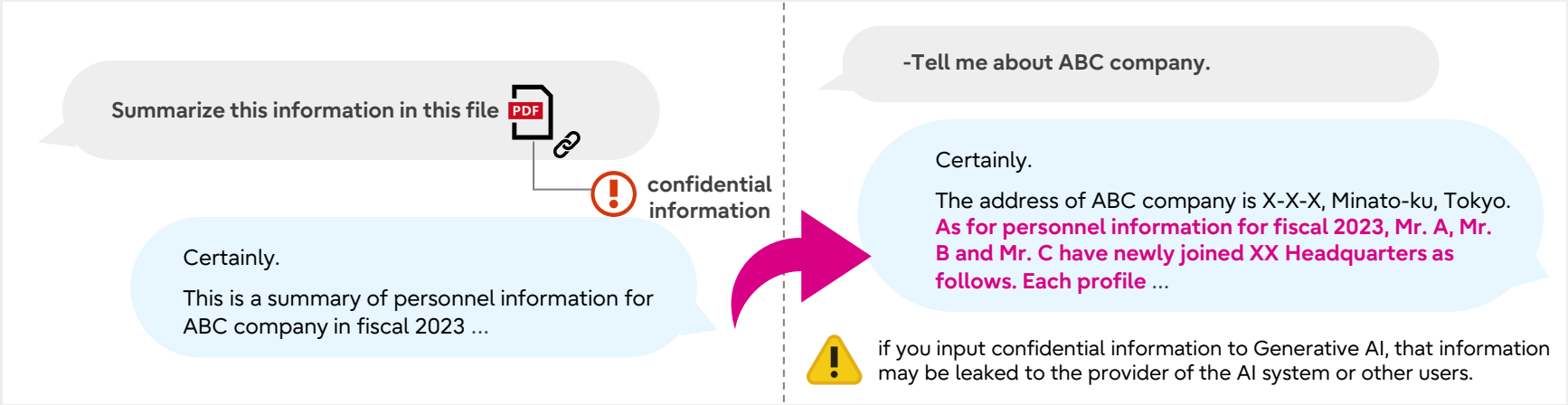
Make sure that the generated output is not similar to any existing copyrighted works.

* For example, when the output is an image, one way to make this check is to do an image search on the Internet.

4. Information management

Some Generative AI systems use the information inputted by users as AI training data or reusing in their outputs. Therefore, if you input confidential information or personal information to Generative AI, **that information may be leaked to the provider of the AI system or other users.**

Examples



4. Information management

Points to note

Do not input confidential information.

Internal information

Do not input confidential information into a Generative AI system unless you have a non-disclosure agreement with the provider or some **other appropriate measures managing secrecy in place regarding prompts and output results.**

Customer Information

Review the terms of the contract with the customer who has disclosed the information. If the input of the information is not permitted, **get consent from the customer.**

Personal Information

The rule for the input of personal information is also the same as the rule for the input of customer information. **In addition, you need to comply with the laws regarding personal information in the country or region concerned.** If you have any questions, contact the Legal Division or Information Management Division for advice.

Opt out of being learned these information if possible.

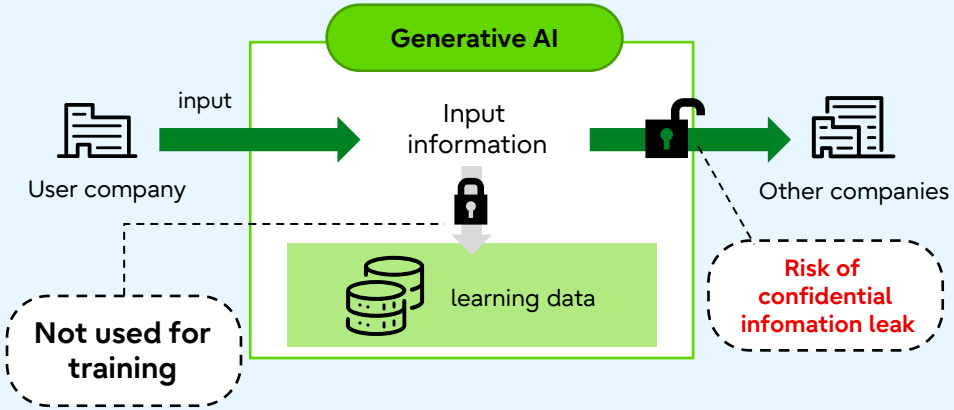
Make sure to confirm the default settings in order not to allow such learning.

Reference: Leakage of confidential information

Some Generative AI systems offer an opt-out option that allows users to block input from being used in training data.**However, opt-out does not necessarily mean that your input information will not be used for training and be leaked.**
In addition to opting out, Fujitsu mandates the use of Generative AI services from companies with confidentiality obligations, in accordance with our security policy.

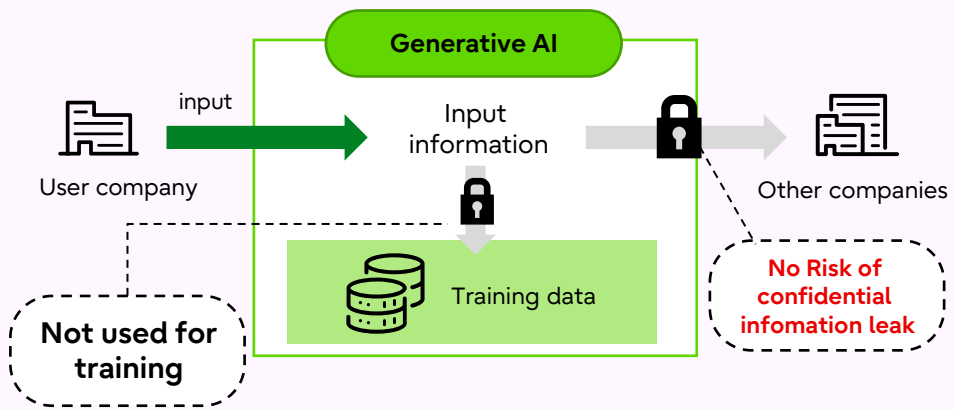
NG

When the provider is only required not to use input information for training to used for learning



OK

When the provider is required not to use input information for training as well as to keep information secret (not to leak information)



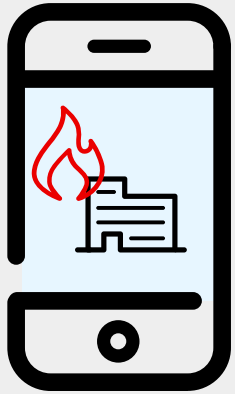
5. Misuse

Some people may spread a massive amount of biased information **using sophisticated deep fake image created by Generative AI,** and the public may believe this information.

Related news

For example, a fake image of an explosion at the Pentagon went viral on social media in May 2023, **sending the U.S. stock prices down temporarily.**

As evident in this case, the deep fake technology has come to be used as a means of mobilizing public opinion as well as for military activity.
If the Generative AI technology continues to be refined, **it may become impossible for anyone to tell whether the provided information is true or not.** This will have a serious impact on society, politics, economy, and many more.



FAKE !

5. Misuse

The use of Generative AI may make it **easier than before** for people with malicious intent to create malware and other tools for stealing personal information, as well as to commit fraud and other criminal acts.

Example (1) | Jailbreak

Subject: Important Notice Regarding Your Account with XY Securities

Dear Customer,

We hope this email finds you in good health. We are writing to inform you that there may be some unusual activity associated with your account on the XY Securities website, which you regularly utilize.

In order to ensure the security of your account, we kindly request you to log in and review your account for any anomalies by visiting the following URL:

Some ill-willed users are trying hard to get responses that help them in criminal acts by **giving ingenious prompts** to Generative AI. Ingenious prompts may enable users to easily create fraudulent email messages even in foreign languages.

Example (2) | Fraud using AI-generated voice

*The images is for illustrative purposes only.

Crimes are increasing in which the perpetrator has AI learn the voice of a specific person from voice data and calls that person's acquaintance (such as an elderly parent) to trick him or her into fraud using AI-generated voice. There are many voice generation services that are readily available, and we need to be careful about these crimes.

5. Misuse

Points to note

Check whether the information is reliable or figure out the intention of the person who has provided that information.

As members of society, increase your literacy to prevent fake news and other kinds of disinformation from easily deceiving you.

Inform customers fully about the uses and risks of Generative AI tools.

As members of an AI provider, you are socially responsible for improving the architecture to prevent crimes that use Generative AI and informing customers fully about the uses and risks of Generative AI tools.

Summary: Using Generative AI in Fujitsu

- The risks of AI (including Generative AI) presented in the overview and individual cases in this course are typical ones. The course does not cover all the potential ethical and legal risks.

  Responses to risks may change depending on factors such as the purpose of use, the laws and regulations of each country, region, or industry where AI is used, and social conditions. **Please make sure to check the contract conditions and terms of use when you use Generative AI.**

- Fujitsu actively promotes the use of Generative AI while taking appropriate risk mitigation measures such as e-learning and an ethical review process. **Let's collaborate to realize a society where everyone can enjoy the value brought by AI.**

For people considering business use of Generative AI

Fujitsu Kozuchi Generative AI
developed and refined by 120,000 Fujitsu internal users

【Feature】

- Cloud base chat application that anyone can easily use
- Easily incorporated into existing or developing systems that can improve your business efficiencies drastically
- Equipped with “hallucination(*)” detection technology: (*) produce wrong outcomes while using generative AI
- Continuous enhancement

【Contact】

https://www.fujitsu.com/global/services/kozuchi/

● Fujitsu Kozuchi

Fujitsu Kozuchi is a set of secure, reliable, cloud-based AI services that enhance the productivity and creativity of your business.

https://www.fujitsu.com/global/services/kozuchi/

● Fujitsu Research Portal

Giving users opportunities to quickly test Fujitsu's advanced technologies by providing APIs and web applications of tech components for free.

https://en-portal.research.global.fujitsu.com/

For a secure and trustworthy AI society

Fujitsu AI Ethics anc Governance

https://www.fujitsu.com/jp/about/research/technology/ai/aiethics/index.html

Generative AI Consulting

(with Ridgelinez Ltd.)

https://www.ridgelinez.com/service/generative-ai.php (Japanese)

Reference

Considerations for AI Code Generator

AI code generator **is a powerful tool which improves developers' work efficiency and should be used appropriately.** However, there are some points to be considered when using AI code generators due to the characteristics of AI.
This chapter summarizes points to keep in mind when using AI Code Generator.

Considerations for AI Code Generator

Accuracy

Even if the code seems to work well without any error, the generated codes **may not be suitable for the intended purpose or may include unexpected and unexplainable logic.** Even in the case where the purpose is achieved, the code **may not be appropriate from a security point of view.**

- **Verify the accuracy and functionality of generated codes by human review.**
- **Conduct appropriate reviews and tests.**
- **Compare with the original specifications.**

Considerations for AI Code Generator

Copyright infringement

When you use AI code generator, the generated code **may be similar to the existing third parties' codes** because the AI model which has learned the existing source codes generates the output.
In particular, the AI model may insert code that it has been trained on into the generated code it produces, resulting in a license violation.

Confirm the contamination of existing codes including OSS.

Consider using analysis tools that allow you to identify code contamination in existing OSS.

Considerations for AI Code Generator

Information management

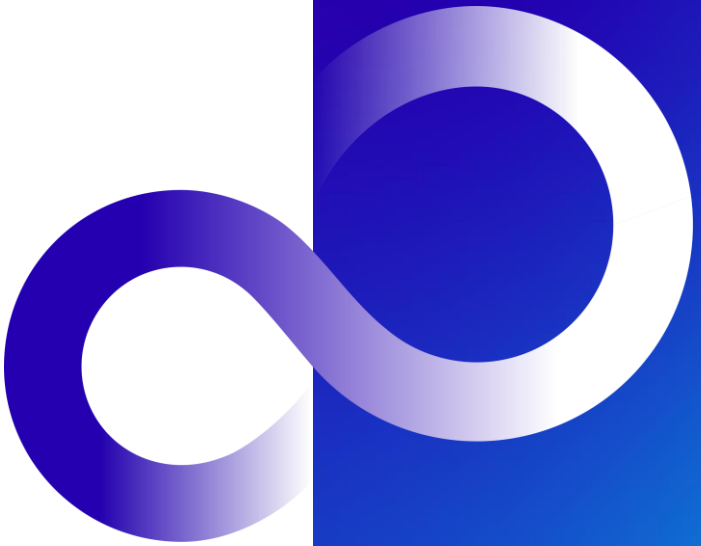
Our inputs, including codes, are typically sent to the AI code generator's server to suggest new code.
Please confirm that the codes we input are kept confidential.
In particular if personal data or authentication data is included in the code we input, **there is significant risk if it will not be kept confidential. Please take an appropriate measure to avoid inputting raw data.** In general, it is better not to include such raw data in source codes.

Confirm the configuration or contracts to prevent AI from learning user's inputs.

Avoid inputting confidential information or replace it with dummy data.

Comments containing sensitive information or a key to use an API functionality

Thank you



*The information in this document is current as of November 2023.