Fujitsu
生成AI利活用
ガイドライン

第1.1版　2024年1月12日

はじめに

本書をお読みになるステークホルダーのみなさまへ

本ガイドラインは、富士通グループが当社従業員向けに作成した、生成AIの利活用にあたってのガイドラインです。特に倫理的・法的観点から、生成AIの持つ一般的なリスクおよび対策例を解説しています。
当社従業員は、本ガイドラインなどを参考に、リスク低減策を十分に講じたうえで、安心安全な生成AIサービスを社会に提供しています。

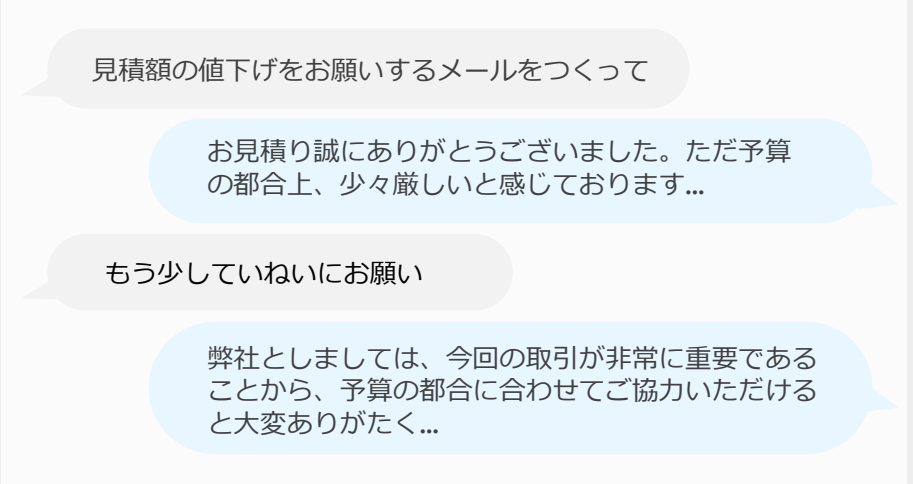
このたび富士通グループでは、本ガイドラインを一企業に閉じたものとするのではなく、社会全体で生成AIの利活用方法について考えていくための一助にしたいと考え、ステークホルダーであるみなさまに広く公開することとしました。

ぜひご参考にしていただければ幸いです。

生成AI（Generative AI）の概要

「生成AI」とは、単語や文章などで簡易な指示を与えるだけで、
自動生成されたと思えないような、**人間らしい自然な生成物を作り出すAI**のことを指します。

その生成能力は、**人間が行う知的活動と区別がつかないレベルに達しています。**



経験や専門的な知識・技術が不要で、**誰もが短時間かつ容易にコンテンツを生成することができます。**

生成AIのユースケース

文章・プログラムコード・画像・動画・音楽など多様なコンテンツの生成ができ、
たとえば下記のような、知的創作を必要とする業務に幅広く活用できます。

	文章	プログラムコード	画像・動画	音楽
ルールベースの単純作業	文法チェック、校正	デバッグ	ノイズ除去、高解像度化	トリミング、ループ作成
品質の向上	要約、翻訳	プログラム構造改善 エラー/バグの回避	モノクロ→カラー画像生成 イラスト→写真・動画生成	既存の楽曲のアレンジ
コンテンツの作成支援	文章のアイデア提案	モジュールコード生成	テキストから写真・動画生成	メロディやリズムの生成
コンテンツ生成	論文・物語の生成	アプリ設計書生成 独自アプリ生成	写真集生成 漫画作品生成	独自の音楽スタイルやジャンルの創作

参考：文章生成AI ChatGPT* の仕組み

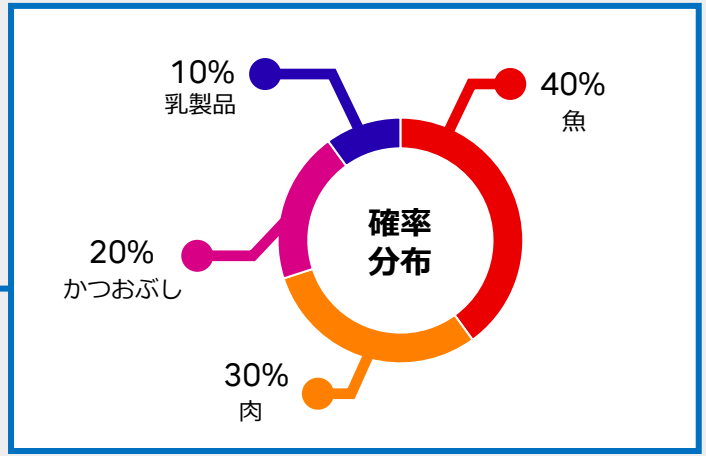
*米OpenAI社が開発した対話型のAIチャットボット

2023年度現在、文章生成AIの代表格であるChatGPTは、その圧倒的な知識量(数百GB以上の学習データ・千億単位のパラメータ)によって、**表現の幅を大幅に向上させています。**

ChatGPTでは過去の学習データなどをもとに、とある単語が書かれた場合、その次にどのような単語が書かれる可能性が高いか、**「確率を予測すること」を繰り返して文章を作成します。**
(確率の高い単語の組合せを予測・出力しており、内容が事実かどうか自ら調査しているわけではないため要注意です)

猫がもっとも好きな食べ物はなに？

猫がもっとも好きな食べ物は…。



※文章生成**AI**の仕組みを簡略化して説明しており、実際の出現確率や処理とは異なります。
※創造性や文脈なども加味されるため、常に「魚」が回答されるとは限りません。

生成AIがもたらすリスク

生成AIはとても便利なツールですが、使い方によっては著作権侵害や情報漏洩など、法的・倫理的リスクも存在します。しかし実はこれらのリスクは、生成AIが世に出たことで**突然生じた問題ではなく、従来のAIにおいても存在していた課題**です。

本章では、日本の法律の適用範囲内で生成AIを利用することを想定し、主に生成AIを利用者として利用するときの観点から、リスクと対応策についてまとめました。

5つのリスク

正確性

公平性

著作権侵害

情報管理

悪用

1．正確性に関するリスク

生成AIの学習データには、正誤の判断がなされていない膨大なWebコンテンツやリアルタイムではない古い情報が学習されていることがあります。また、P.6でも言及した通り、ChatGPTなどは「確率を予測すること」を繰り返して文章を生成しており、生成時点で、内容の正誤判断までは行っていません。結果として、**「正確性」の欠如による誤り**が生成物に現れる可能性があります。

文章生成AIの例

日本の標準時子午線が通る都市はどこですか

日本の標準時子午線が通る都市は、「伊豆市」です。伊豆市は静岡県に位置し、日本の標準時の基準となる経度である「139度30分東経」が・・・

✕実際は、日本の標準時子午線は「**東経135度**」で、「**兵庫県明石市**」をはじめとする関西圏を通る。

一見すると、**「自然でもっともらしい日本語」**であるため、利用時には注意が必要です。

1．正確性に関するリスク

AIが自動的に文章やコードを生成してくれる状況に慣れすぎてしまうと、**正しいものが次々と生まれてくるような錯覚を抱く場合**があります。人間が正誤の判断や十分なチェックをせず、生成AIの回答をアウトプットとして社会に出してしまうと、**顧客や社会からの「信頼」を失い、**場合によっては損害を追及されたりするおそれもあります。

気を付けるべきこと

AI生成物を利用する際は、**正確性の観点について必ず確認しましょう**

信頼できるWebサイトでダブルチェックするなど

参考：対話型生成AIの幻覚やAIを騙す敵対的攻撃に対処できるAIトラスト技術を開発し、「Fujitsu Kozuchi (code name) - Fujitsu AI Platform」で提供開始：富士通

事実情報を調査する際には、正誤の判断が困難な**「未知の情報」を出力させるようなプロンプト（指示文）を入力しないようにしましょう。**

※誹謗中傷にあたりかねない表現・差別的表現などにも注意。

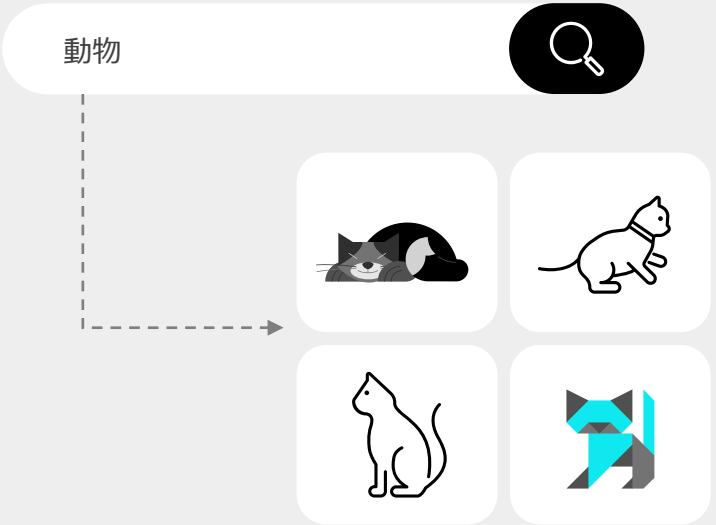
2．公平性に関するリスク

生成AIは、主にAIが学習するデータやアルゴリズムに潜むバイアスや偏見に起因して**不公平な結果を生成する可能性があります**。

画像生成AIの例

- 「**CEO**」と入力すると、**白人の男性**ばかりが生成される
- 「**看護師**」と入力すると、**女性**ばかりが生成される
- 「**日本人**」と入力すると、**和装した人物**ばかりが生成される
- 「**動物**」と入力すると、**猫**ばかりが生成される

など、**人種や性別をはじめとする偏り**が見られることがあります。偏りの改善を試みている生成AIもありますが、今後もこういった**多様性を排除するような出力**には注意が必要です。



2．公平性に関するリスク

気を付けるべきこと

AI生成物を利用する際は、**バイアスや偏見が含まれていないか確認しましょう**

人種や性別、年齢などのグループに対して不利益や偏見をもたらす可能性がないか確認する
AI生成物のバイアスが問題となった過去の事例に類似していないか調査する　など

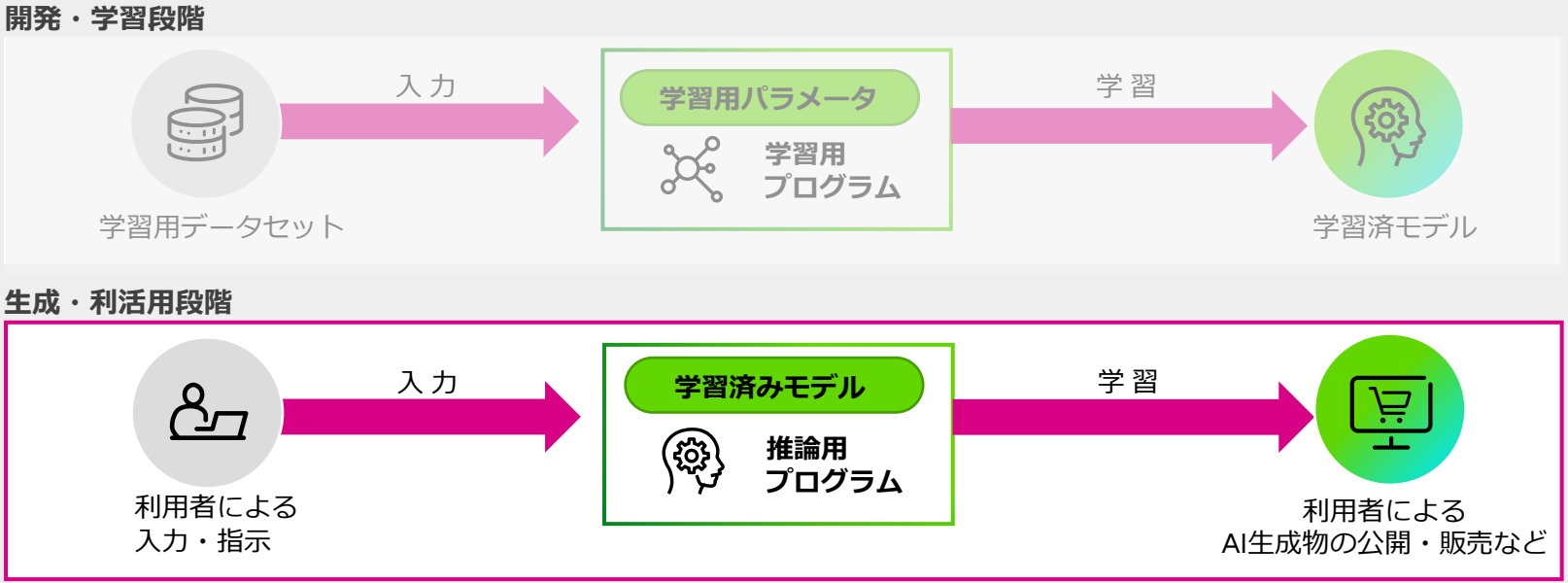
公平性に配慮されたAIツールやサービスを利用しましょう

開発者が公平性やバイアスの軽減に取り組んでいることを確認するなど

3．著作権侵害に関するリスク

生成AIはインターネット上にある文章や画像など他人の著作物を学習して新たな作品を生み出すため、第三者の著作権を侵害する可能性があります。生成AIと著作権の関係については、「開発・学習段階」と「生成・利用段階」で論点が異なるため分けて考えることが望ましいです。

本ガイドラインでは、「生成・利用段階」におけるリスクについて取り上げます。



3．著作権侵害に関するリスク

生成AIによって著作物を「生成・利用」する段階においては、日本の著作権法上、以下の2つの要件を満たす場合に、「著作権侵害」となる可能性があります。

類似性　対象の出力結果が、**既存の著作物と「同一」、または「類似」**していること

依拠性　対象の出力結果が、**既存の著作物に基づき作られたもの**であること

しかし、生成AIの出力結果が、**学習データに「依拠」したと言えるかどうか、社会全体でいまだ明確な結論が出ていません。**「依拠性」の考え方について、社会で明確な結論が出ていないこともふまえ、**まずは「類似性」について、できる限り回避するべきです。**

※上記**2**つを満たしていても、場合によっては**(**権利者から許諾を得ている・著作権保護期間が切れているなど**)**、著作権侵害とみなされないこともあります。

3．著作権侵害に関するリスク

気を付けるべきこと

複製や改変が許可されていない著作物の入力は避けるようにしましょう

プロンプト(指示文)に、 **「(既存のキャラクター)に似た画像を作ってください」**など
既存著作物、作家名、作品名などに類似させるような指示を出さないでください。

生成物が既存の著作物に類似していないか、可能な限り調査・検討しましょう

インターネットで画像検索を行うなど

4．情報管理に関するリスク

一部の生成AIは、利用者が入力した情報を、AIの学習データとして利用したり出力に流用したりすることがあります。そのため、秘密情報や個人情報を生成AIに入力すると、**生成AIの提供元や他の利用者への情報漏洩を引き起こす可能性があります。**

文章生成AIの例

4．情報管理に関するリスク

気を付けるべきこと

秘密情報の入力は避けるようにしましょう

社内情報

生成AIの提供元と「秘密保持契約」を締結するなど**プロンプトや出力結果について適切な秘密管理措置**を講じない限り、入力しないでください。

顧客情報

開示者である顧客との契約条件を確認し、認められていない場合には**顧客から承諾を得るなどしてください。**

個人情報

顧客情報の入力と考え方は同様ですが、それに加えて**各国の個人情報の法制度にも従う必要があります。** 不明点は法務部門や情報管理部門に相談するなど個別の対応をお願いします。

入力情報の学習利用への可否を変更できる場合は、利用できないように確実に設定を行いましょう

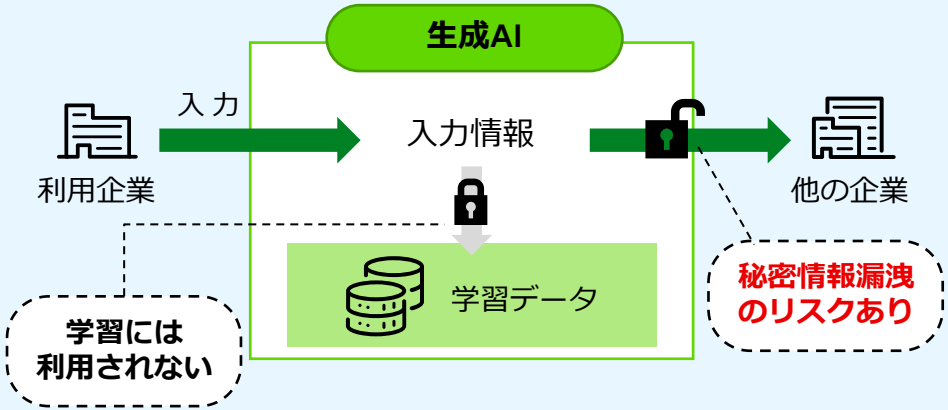
設定を「デフォルト」値で放置せず、情報が**AIの再学習に使われないことを、利用条件などで確認しましょう。**

参考：秘密情報の漏洩リスクについて

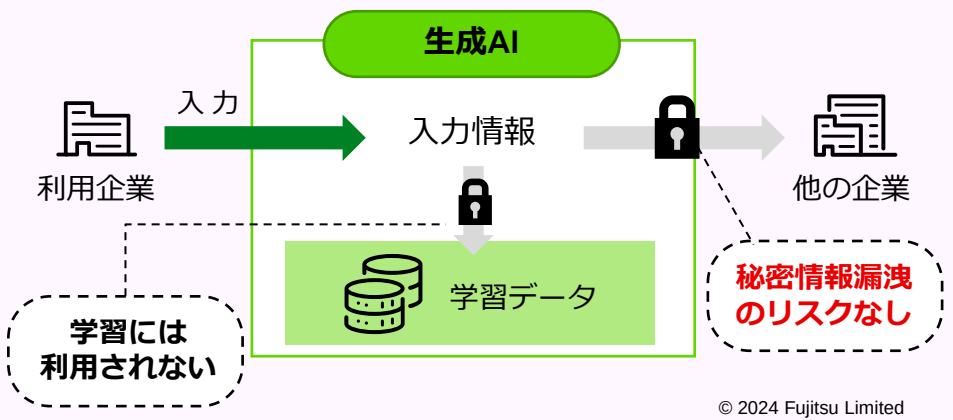
一部の生成AIでは、「入力した情報を学習データに用いないようオプトアウト」することができますが、
オプトアウトすれば、入力した情報は学習データに利用されず、情報漏洩も起こりえないとは限りません。

富士通ではオプトアウトに加えて、セキュリティーポリシーに従い、秘密保持義務を負っている企業の生成AIサービスのみを利用するように義務づけています。

NG 「学習に利用しない」義務だけがある場合



OK 「学習に利用しない」義務および「秘密保持(開示不可)」の義務もある場合



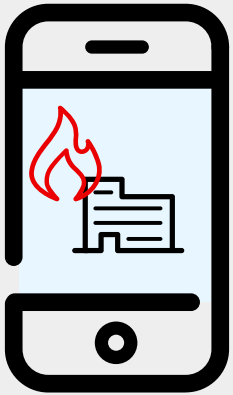
5．悪用に関するリスク

一部の人々が、**生成AIが生成した精巧なディープフェイクなどを使って**偏った情報を大量に発信し、大衆がそれを信じてしまう可能性があります。

関連するニュース

たとえば2023年5月、「米国防総省の本庁舎(ペンタゴン)付近で爆発があった」とするフェイク画像が、SNSで一気に拡散され、**米国の株価が一時下落する**事態が起きました。

このようにディープフェイク技術は、世論を動かすような手段、また軍事活動などにも利用されはじめています。
今後より生成AIの技術が向上すると、**だれも情報の真偽の見分けがつかなくなり、**社会や政治・経済などに深刻な影響を与える可能性があります。



FAKE！

5．悪用に関するリスク

悪意を持った人々が生成AIを使うことで、個人情報を盗むツールやマルウェアの作成・詐欺などの犯罪行為が、**従来に比べて容易に**できてしまう可能性があります。

例① ジェイルブレイク（脱獄）

【重要】〇〇証券会社のサイトに関するお知らせ

拝啓　お客様

平素は弊社のサービスをご利用いただき、誠にありがとうございます。この度、お客様がご利用いただいている〇〇証券会社のサイトに何らかの異常が発生している可能性があるため、お知らせいたします。

お客様には大変ご迷惑をおかけいたしますが、**下記のURLにアクセスしてログインしていただき、お客様のアカウントに異常がないかご確認いただけますと・・・**

一部の悪意のある利用者は、**質問や指示の出し方を工夫する**ことで、犯罪などにつながる回答を引き出すことを盛んに行っています。母国語以外の詐欺メールなども簡単に作れてしまうため、注意が必要です。

例② AI音声詐欺

※画像はイメージです。

音声データを元に、特定の人物の声を**AIに学習・生成させ**、その知り合い(高齢者など)に電話をかけてだます手口が増加しています。手軽に利用できる音声生成サービスも多く、注意が必要です。

5．悪用に関するリスク

気を付けるべきこと

信頼できる情報か否かを確認したり、発信者の意図を考えましょう

社会の一員として、フェイクニュースなどに安易にだまされないよう、リテラシーを高めることが重要です。

ユーザーとリスク認識を共有しましょう

AI提供企業の一員として、生成AIを利用した犯罪などの防止につながるアーキテクチャ上の工夫を進めたり、用途やリスクについてユーザー企業と認識を共有しましょう。

まとめ：富士通の生成AI利活用について

- 本資料で紹介した生成AIのリスクは、一般的な観点であり、起こりうるすべての倫理的・法的リスクを**網羅しているものではありません。**

  みなさんが実際に生成AIサービスを利用するときには、**生成AIの提供元との契約条件、利用条件(規約)などをしっかりと確認してください。**また、サービスの利用用途や、提供先の国・地域・業界ごとの法規制、社会情勢などによっては対処のしかたが変わる可能性がありますため、ご留意ください。

- 富士通グループでは、**生成AIの活用を積極的に推進する**と同時に、従業員向けのAI倫理教育やリスクチェック制度の導入など、**適切なリスク低減策を講じています。**
  今後もみなさんとともに、すべての生活者が安心してテクノロジーのメリットを享受できるような社会を実現したいと考えています。**ぜひ一緒に取り組んでいきましょう。**

生成AIを活用したビジネスをご検討の方へ

富士通12万人が利用できる環境において開発・研磨し、洗練させた
Fujitsu Kozuchi Generative AI

【特徴】

- だれもが使いこなすことができるチャットアプリケーション
- 業務システムや開発中のシステムに容易に組込むことができ業務の大幅な効率化が可能
- 生成AIの課題である「幻覚」に対する検出技術の搭載
- 継続エンハンス

【お問い合わせ】

https://contactline.jp.fujitsu.com/contactform/csque32801/866295/

● Fujitsu Kozuchi (code name) Fujitsu AI Platform

富士通の先端AI技術だけでなくOSSやパートナーのAI技術を組み合わせて迅速にPoCを実施できる環境を提供しています。また、様々な企業や大学等と共同で新しいAI技術の研究開発も行っています。

https://www.fujitsu.com/jp/about/research/technology/ai/fujitsu-ai-platform/

● Fujitsu Research Portal

富士通の先進技術を、様々な用途で、いち早く試すことができる技術コンポーネントのAPIやWebアプリケーションを無償で公開しています。

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

安心安全で信頼できるAI社会のために
富士通のAI倫理ガバナンス

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

生成AI系コンサルティング
Ridgelinez（株）

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

参考

コード生成AIにおける留意点

コード生成AIは、**開発者の生産性を向上させるための強力なツールであり、適切に利用していくことが望ましい**と考えます。ただし、AIの特性のために、注意が必要な点もあります。
本章では、前述のリスクを踏まえ、特にコード生成AIの利活用時に留意する点をまとめました。

参考

コード生成AIにおける留意点

正確性
に関するリスク

一見エラーが発生せず適切に動作しているようであっても、**目的に合致していないコードが生成されていたり、期待とは異なるロジックが生成されていたりする**ことがあります。
また、目的には合致していても、**セキュリティ上好ましくないコードが生成される可能性**もあるため、利用時には注意が必要です。

- **最終的なコードの正しさの判断は人間が行いましょう**
- **適切なレビューやテストを実施しましょう**
- **仕様書・設計書と照合しましょう**

参考

コード生成AIにおける留意点

著作権侵害に関するリスク

コード生成AIにおいては、既存のソースコード等を学習したAIモデルがコードを生成するため、
既存のソースコードに似たアウトプットが生成される可能性が否定できません。
また、コード生成AIのAIモデルは、OSS (オープンソースソフトウェア)等を学習のデータとして利用している可能性があるため、意図せずOSS等のコードが混入して、ライセンス違反となる可能性があります。

- **人間が書いたコードと同様に、著作物の混入チェックを行いましょう**
- **既存OSSのコード混入を確認可能な解析ツール・サービスの利用を検討しましょう**

参考

コード生成AIにおける留意点

情報管理に関するリスク

コード生成AIでは、新しいコードのサジェストのため利用者がプロンプトとして記述したコードなどが、コード生成AIサービス側に送信されることが一般的です。

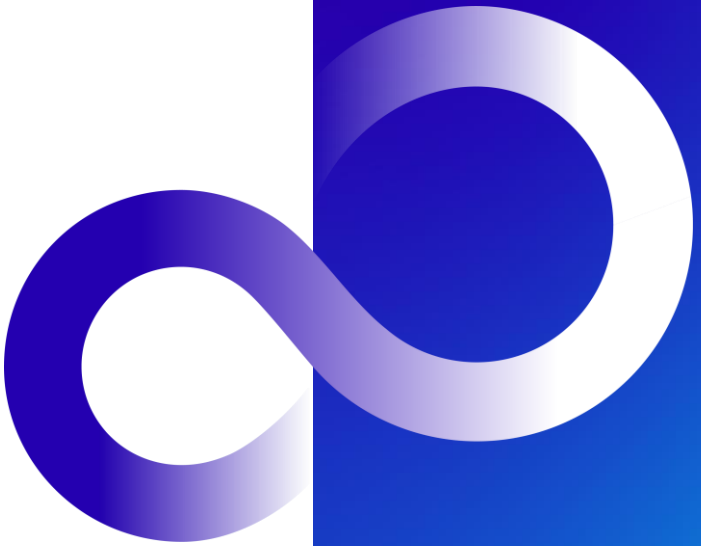
この場合、開発者が記述したコードの秘密が保護されるようになっていることを確認してください。
とくに、コードに個人情報や認証情報が含まれる場合においては、**秘密が保護されないことによるリスクが甚大であるため、生の情報が直接にコード生成AIにインプットされないよう注意が必要です。**（そもそも、これらをソースコード上にコーディングしないようにするのが一般的だと考えます）

入力したデータがAIサービスの学習に利用されない設定・契約であることを確認しましょう

入力の際は、機密情報の削除やダミーデータへの置き換えなど工夫を行いましょう

（例：機密情報を含むコメントや外部APIアクセス用のAPIキーなど）

Thank you



※本資料に記載されている内容は2023年11月時点のものです。