2023 年 5 月

企業が生成 AI の奔流を乗り越えるためのアジャイルリスク管理

はじめに

急速な進化を遂げた生成 AI 技術は、多くの分野へ大きな影響を与えることが予想され、注目が集まっている。OpenAI 社の ChatGPT や Stability AI 社の Stable Diffusion は、近年の研究に基づいた生成 AI モデルに広く一般からのアクセスを可能としたもので、衝撃をもって迎えられた。周知の通り、現在ではより多くの機関や企業が生成 AI に関係するシステムやサービスを次々と発表している。一方では、生成 AI の学習過程におけるデータの取扱いへの批判や、悪意をもった生成 AI の利用から生じる弊害についての懸念も広がっている。

生成 AI は間違いなく大きな可能性のある技術で、現在そのポテンシャルが発揮されようとする時期にある。その強力さゆえ、積極的な利用や新サービスの情報、慎重論や予防的な規制について様々な視点からの議論が交わされている。生成 AI を積極的に活用することを選んだ場合も、その逆の場合でさえも、ビジネス上のインパクトは大きなものとなるだろう。各企業は「生成 AI とどう向き合うか」という命題を突き付けられている。氾濫する情報の中でより良い方針を決定しなければならない。

本稿の課題意識はこの点にあり、企業リスクの視点から[1]日々更新される生成 AI に関する情報を整理分析するためのフレームワークを提案するものである。急速に変化する対象のリスク管理に適したアジャイル・ガバナンスの考え方に基づき、生成 AI リスクの全体像を整理した上で適応的な管理を実践するための手法を論じる。想定読者には、主に企業のリスク責任者（CRO、全社リスク統括）やデジタルリスク責任者（CTO、CIO、CISO など）を想定するが、生成 AI の利用に関連する幅広い層にご活用いただきたい。

構成は以下の通りである。

まず 1 章では、近年生成 AI に起こった進化の核心部分を振りかえり、今後特に着目すべきポイントを整理する。

次に 2 章では、現在の生成 AI を巡る議論のスピードに適するアジャイル・ガバナンスの考え方に沿ったリスク統制の方策を概観する。

最後に 3 章では、生成 AI のエコシステムを分析し、企業がアジャイル・ガバナンスを実践するためのツールとして、JCIC が独自に作成したリスク管理マトリクスを紹介する。生成 AI に関連する課題の整理と対応リーダーの明確化を速やかに行う助けとしていただきたい。

本稿が、生成 AI の特性をよりよく理解し、効率的な情報分析を行う助けとなれば幸いである。

[1] ここで、リスクは将来的に生じる不確実な結果とし、損失と利益の双方を含むものとする。特に生成 AI は指摘される課題も大きい反面、広範なビジネスメリットも見込まれるため、両者の視点を強調することは重要である。

1. 生成 AI に起こった進化と着目すべきポイント

生成 AI に関連するリスクを正しく捉えるため、まず生成 AI の基本的な特徴と現在までの進展を振り返る。そのうえで、今後起こる様々な事象を効率的に分析するために着目すべきポイントをまとめる。

① そもそも生成 AI とは何か

生成 AI（Generative AI）とは、文字通り文章や画像などを生成する機能をもった AI の総称である。これまで主に実用化が進んでいた識別 AI（Discriminative AI）は、推薦スコアの算出や画像識別、不正検出のような、所定のフォーマットに従う数値予測やクラス分類のタスクを得意としてきた。

これに対して生成 AI は、入力データの構造を解析し、関連性の高い訓練データをもとに新しいコンテンツを生成するというタスクを実行することから性質が異なる。文章や画像の作成というタスクは汎用的が高く、幅広い活用の可能性が期待されつつも、実用化は遅れていた。ヒトが行う作業を代替するためには、ヒトをベンチマークとした評価で十分な性能が求められることになるが、これは簡単なことではなかった。

② 生成 AI の急激な進化

2015 年頃から生成 AI にとって急激な進化の時期が始まった。一つ目のきっかけは、強力なモデル表現の獲得である。Diffusion Model[2] や Transformer[3] といった新たなモデルの発見により、文章をはじめとする系列データの学習効率や画像や音声の生成品質の向上にブレイクスルーが生じた。

もう一つのきっかけは、大量データの学習と大規模モデルの構築である。自然言語処理を例にとると、Tranformer モデルの採用によってこれまでより効率的に大量データの学習が実現することになった。BERT[4]や GPT[5]といった大規模言語モデル（LLM）は、数十億語・数十テラバイトという膨大なデータを学習している。さらに、大規模言語モデルはその構成パラメータ数が一定以上に増加すると急激にその性能が向上する[6]ことが明らかになった。当初 BERT や GPT-1 が 1 億強のパラメータ数だったことに対し、近年主流の PaLM や GPT-3.5 は数千億のパラメータ数[7]を誇る。

これらの要因によって劇的な性能向上を果たしたモデルが、現在私たちが目にしている生成 AI による変革の基盤となっている。

[2] Sohl-Dickstein, et al., “Deep Unsupervised Learning using Nonequilibrium Thermodynamics", http://proceedings.mlr.press/v37/sohl-dickstein15.pdf (Jun. 2015)
[3] Jakob Uszkoreit, “Transformer: A Novel Neural Network Architecture for Language Understanding”, https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html (Aug. 2017)
[4] Jacob Devlin, et al., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, https://arxiv.org/abs/1810.04805v2 (Oct. 2018)
[5] OpenAI, “Improving language understanding with unsupervised learning”, https://openai.com/research/language-unsupervised
[6] Jason Wei and Yi Tay, “Characterizing Emergent Phenomena in Large Language Models”, https://ai.googleblog.com/2022/11/characterizing-emergent-phenomena-in.html (Nov. 2022)
[7] GPT-4 のパラメータ数は現在非公開。

③　普及と応用のフェーズへ

急激な進化を経た生成 AI は実用レベルのアウトプットの閾値を超え、普及と応用のフェーズに差し掛かった。直接 AI 開発に携わる企業以外にも多くの企業が生成 AI に触れ、事業との関係性を深めていくことになるだろう。生成 AI の応用を考える際に鍵となるいくつかの概念を交えつつ、着目すべきポイントを解説する。

AI システムの構造は、AI モデルと AI システム・AI サービスを区別して考えることで理解しやすくなる。AI モデルは、データの学習によって訓練されたパラメータによって構成されるソフトウェアを指し、AI モデルを構成要素として具体的な用途に応じたアプリケーションに組み込まれたシステムの全体を AI システムまたは AI サービスという。

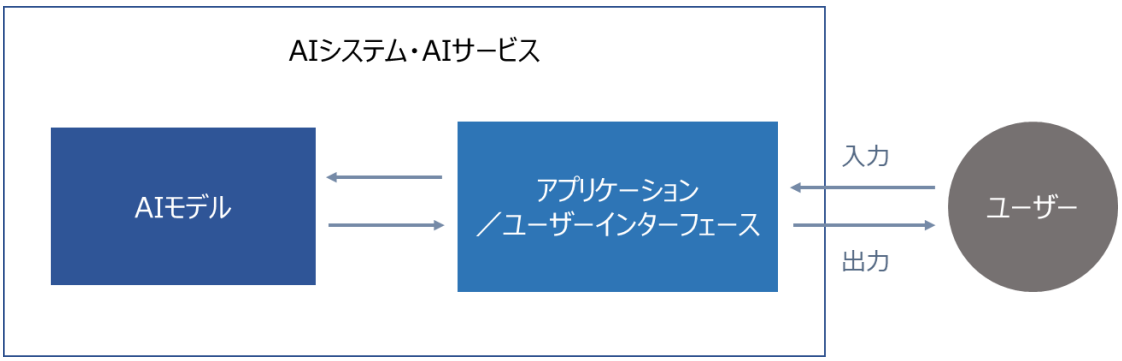


図 1 AI モデルと AI システム・AI サービスの関係

チャットアプリ形式の生成 AI を例にとると、OpenAI の ChatGPT であれば GPT-3.5 または GPT-4 をベースにした AI モデルがあり、ウェブブラウザやモバイルアプリ上の対話インターフェースやセーフガード機能などを有するアプリケーションに組み込まれることでサービスとしての ChatGPT が提供されている。同様に、Google の Bard であれば、LaMDA や PaLM をベースにしたモデルを対話型アプリケーションに組み込んだものとなる。

近年大幅な性能向上を果たしたのは AI モデルの部分にあたる。チャットアプリの例では、大規模言語モデル（LLM）が、テキスト生成という汎用的なタスクを極めて広い範囲で高精度に行うことを可能にした。そして、チャットというアクセシビリティが高いインターフェースを持ったアプリケーションの形で一般に公開されたことが現在のトレンドの起点となったと理解できる。そして、オンライン検索との統合やコード生成といった形でのアプリケーションの進化が続いている。

更に応用を推し進める構図として、AI モデルの第三者利用がある。AI モデルの開発者とアプリケーションの開発者は同一である必要はない。実際に LLM を開発した各社は API を通じたモデル利用のサービスを提供、または計画している。また、Meta 社の LLaMA[8]をはじめ、LLM をオープンソース公開する動きも複数見られている。アプリケーションの開発者は複数の候補から事前学習済の AI モデルを選択し、独自の様々なサービスのコンポーネントとして利用することが出来る。

[8] Meta AI, “Introducing LLaMA: A foundational, 65-billion-parameter large language model”, https://ai.facebook.com/blog/large-language-model-llama-meta-ai/ (Feb. 2023)

図 2 AI モデルの 第三者利用

また、ファインチューニングと呼ばれる手法を用いることで、大規模データによって事前学習されたモデルを基盤とし、特定のタスクに適した追加データを学習させることでパラメータの微調整を行うことも可能である。

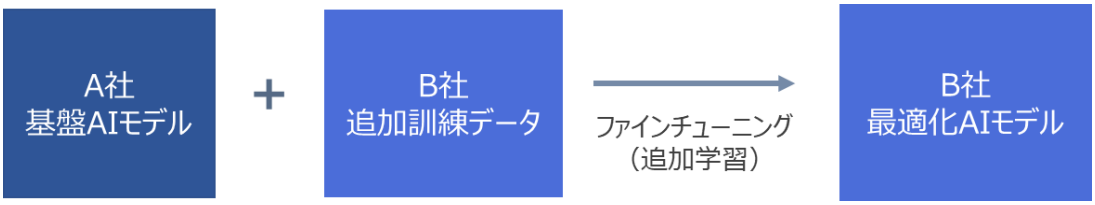


図 3 ファインチューニング

アプリケーションの観点からみると、生成 AI は現在そのポテンシャルのごく一部を発揮しているに過ぎず、今後様々なステークホルダーが様々な用途に適したサービスを開発、提供していくことになる。同時に、リスクの面でもこれまで主に論じられてきた一般化された AI のリスクに留まらず、より個別具体的な課題が表出化することが予想される。

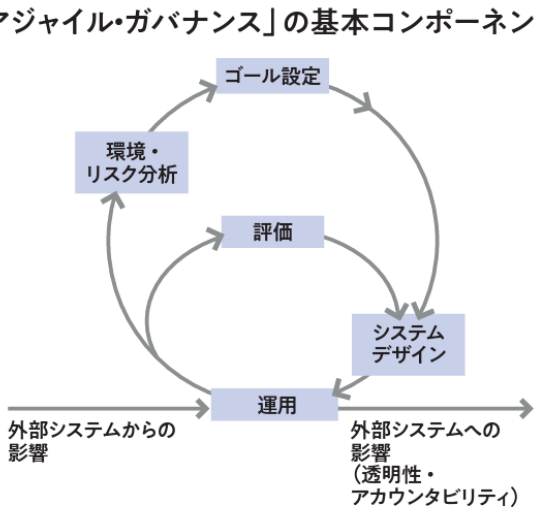
AI モデル自身についても更なる改善は見込まれるが、学習データの枯渇問題やスケールアップによる性能向上の限界などの問題から基本性能の向上はペースダウンし、コスト改善等が中心になるという見通しもある[9]。このことからも、生成 AI の今後の展開として、最も着目すべきは普及と応用の進展であり、企業は生成 AI によって生じる事業環境の変化を見極め、機会およびリスクへの取り組みに最善をつくすことが重要になる。

2. 頻繁な変化に対応するためのアジャイル・ガバナンス

ここまで生成 AI の進化の経緯と今後より実社会への普及と応用が進む見通しについて述べてきた。現在も生成 AI を巡り日々刻々と新しい変化が起きているが、より広い範囲でより踏み込んだ議論が行われることになるだろう。企業の視点からは、生成 AI の応用範囲は広く、コンプライアンス、倫理、プライバシー、セキュリティといった様々な観点からのガバナンスの実践が期待されることになる。

[9] https://www.wired.com/story/openai-ceo-sam-altman-the-age-of-giant-ai-models-is-already-over/

しかし、生成 AI のような急進的なイノベーションによる変化は一般に予測が困難であり、既存の法や制度に基づく固定的なゴールに対してルールや手順を設定していく従来のアプローチによる統制は適さない。こうしたケースに適した統制手法としてアジャイル・ガバナンスがある。アジャイル・ガバナンスでは、常に変化する環境とゴールを踏まえ、最適な解決策の見直しを続けるという方針を採用する。経済産業省の定義では、*“「環境・リスク分析」「ゴール設定」「システムデザイン」「運用」「評価」「改善」といったサイクルを、マルチステークホルダーで継続的かつ高速に回転させていくガバナンスモデル”*とされる（図 4 参照）。



（出所）経済産業省, “GOVERNANCE INNOVATION Ver.2: アジャイル・ガバナンスのデザインと実装に向けて」報告書”, 2021 年 7 月

図 4「アジャイル・ガバナンス」の基本構成要素

そして企業の視点では、*“その（ガバナンスシステムの）運用やモニタリングを行い、問題があればその改善を行うと共に、環境の変化等を踏まえ、その都度「ゴール」を見直していく”*ことが使命にあたる。

この一連のサイクルは、生成 AI のリスク統制においても有効といえる。特に次々と状況が更新される現況においては、生成 AI を取り巻く環境の変化を素早く捉え、統制目標を適正に維持し続けることが重要となる。生成 AI のエコシステムを分析し、生成 AI に関連するリスクを継続的にコントロールすることで、イノベーションから得られる利益を最大化することを目指す。

3. 生成 AI のリスク管理マトリクス

これまでの議論を踏まえ、企業が生成 AI をアジャイル・ガバナンスのアプローチによって最大限活用することを支援するためのツールとして、JCIC が独自に作成したリスク管理マトリクスを紹介する。このリスク管理マトリクスは、生成 AI に関連する情報が極めて多岐にわたり、なおかつ高頻度で更新されるため、企業が行う事業環境とリスクの分析を逼迫させているという課題の解消を目的とするものである。

基本的なアイデアは、まず生成 AI に関連する機会・リスクと生成 AI のエコシステムを構成するステークホルダーの軸からマトリクスを作成し、生成 AI に関連する各情報が企業の統制目標のどの部分に影響を与えるかを可視化するとい

うものである。ここで目標は、リスクへの対応内容を詳細に検討することではなく、必要な専門性を特定し、リスクへの対応内容の検討をリードすべき者と関与すべき者を明確にすることである。個別の検討によって対応内容の詳細が定まった後は、必要に応じてマトリクスの内容へフィードバックを行う。

以後、具体的なマトリクスの作成例を紹介する。各種資料を参考に、生成 AI に関連する機会とリスクおよび生成 AI のエコシステムを構成するステークホルダーを企業の統制目標に馴染む形で整理することを考えると、例えば次のようにまとめることが出来る[10]。

表 1 生成 AI のエコシステムおよび機会とリスクの整理例

生成 AI に関連する機会とリスク	生成 AI エコシステムのステークホルダー
– イノベーション – 倫理的・社会的リスク、説明責任 – 著作権リスク（学習データ、アウトプット） – プライバシーリスク – 安全性とセキュリティのリスク など	– AI モデル、AI システム・AI サービスの開発者 – 自社、関連企業 – 自社の顧客、関係者 – 政府・規制当局、コミュニティ – 競合他社、外部組織 など

生成 AI に関連する機会とリスクは、AI リスクとして代表的にあげられる要因を素地としつつ、個別の論点では生成 AI 固有のリスクの視点を加えるべきである。現在、生成 AI エコシステムの中心には AI モデルや AI システム・AI サービスの開発者がいる。一方で、自社の生成 AI 活用におけるリスクを分析するにあたって、自社と周囲のステークホルダーの関係性が重要となる。自社内で生成 AI サービスを利用するという基本的なユースケースの他に、自社が生成 AI モデルを組み込んだサービスを提供するケース、政府や当局の規制への対応、外部の第三者による生成 AI の利用が自社に与える影響といった観点から事業環境とリスクの分析を行うことが望ましい。

次項のマトリクスは、上述の観点を軸にとり、代表的な機会とリスク要因を記入した一例である。横軸には機会とリスクが対応し、縦軸にはステークホルダーが対応する。ただし、AI モデル、AI システム・AI サービスの開発者については、すべてのリスク要因に関係するため、専用の枠を設けることはしていない。

[10] 主に参考にした資料は以下の通り。
- 総務省, “AI 利活用ガイドライン”, https://www.soumu.go.jp/iicp/research/results/ai-network.html (2019 年 8 月)
- 経済産業省, “AI ガバナンスの相互運用性を促進等するためのアクションプラン”, https://www.meti.go.jp/press/2023/04/20230430001/20230430001.html (2023 年 4 月)
- カテライ アメリア, 井出 和希, 岸本 充生, ”生成 AI（Generative AI）の倫理的・法的・社会的課題（ELSI）論点の概観：2023 年 3 月版”, https://elsi.osaka-u.ac.jp/research/2120 (2023 年 4 月)
- NIST, “NIST Risk Management Framework Aims to Improve Trustworthiness of Artificial Intelligence”, https://www.nist.gov/news-events/news/2023/01/nist-risk-management-framework-aims-improve-trustworthiness-artificial (Jan. 2023)
- Council of the EU, “Artificial Intelligence Act: Council calls for promoting safe AI that respects fundamental rights”, https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/ (Dec. 2022)
- Europol, “The criminal use of ChatGPT – a cautionary tale about large language models”, https://www.europol.europa.eu/media-press/newsroom/news/criminal-use-of-chatgpt-cautionary-tale-about-large-language-models (Mar. 2023)

表 2 生成 AI リスク管理マトリクス（例）

カテゴリ	自社内の利用*1	自社サービスの提供*2	政府・当局による規制	第三者からの影響
イノベーション ビジネス （機会）	生産性の向上 従業員満足度向上 セキュリティ対策等への生成 AI 活用	品質向上、新規性 劣位な市場の逆転	補助金等の可能性 DX 銘柄等	協業先の品質向上 競合の品質向上 優位な市場の逆転
倫理・社会 説明責任	不正確な出力による混乱、ミス[11] 不適切なコンテンツの作成 説明可能性の欠如 デジタル格差	不正確な出力による混乱、損害 社会的バイアス、不公平の増強 不適切なコンテンツの公開 学習データの透明性	過剰な発展への懸念 不適切なコンテンツの学習・生成に関する規制 ポリティカルコレクトネス 国・地域間における方針のギャップ	社会的バイアス、不公平の増強 他社起点の社会的混乱への巻き添え
著作権 ライセンス	他者の著作物を不作為に利用 自社の著作物が外部の AI モデルに学習される ソフトウェアライセンス違反	他者の著作物を不作為に外部提供 ユーザーの著作物を意図せず学習・外部提供 ソフトウェアライセンス違反	生成 AI の特性を踏まえた著作権法等の対応 著作権に関する海外規制対応	自社の著作物が第三者の AI モデルに利用される 利用している AI モデルに著作権法違反が発覚
プライバシー	他者の個人情報を不作為に利用 自社の個人情報が外部の AI モデルに学習される	他者の個人情報を不作為に漏えい ユーザーの個人情報を意図せず学習、漏えい	生成 AI の特性を踏まえた個人情報保護等の対応 GDPR 等海外規制対応	自社の個人情報が第三者の AI モデルから漏えい 利用している AI モデルにプライバシー侵害が発覚
セキュリティ 安全性	自社の機密情報が外部の AI モデルに学習される	自社の生成 AI サービスの脆弱性を侵害される 自社の生成 AI サービスが悪意ある利用をされる 他者の機密情報を不作為に外部提供	生成 AI の特性を踏まえたセキュリティ規制対応 海外規制対応	サイバー攻撃者が生成 AI 技術を悪用 偽情報の拡散・ディープフェイク 自社の機密情報が第三者の AI モデルから漏えい 利用している AI モデルに脆弱性が発覚

*1 グループ会社等での利用を含むことも想定される　*2 他社向けの受託サービス等を含むことも想定される

■：機会　■：リスク

[11] 特に生成 AI がもっともらしい不正確な出力を生成することをハルシネーション（幻覚）という。

表 2 中の機会・リスクは生成 AI の普及初期といえる現在に指摘されているものを、概略レベルでまとめた例である。本表の活用方法として、主に以下の 3 点が期待される。

✓ リスクの全体像の整理

企業のリスク管理者は、本表を活用して生成 AI の事業環境とリスク分析の全体像をアップデートしていくことができる。新しい情報を随時収集し、どのリスクに該当するか、不要な重複や抜け漏れがないか、新規のリスクやカテゴリを追加する必要があるかといった視点から整理を行う。整理した情報に基づき、アジャイル・ガバナンスにおける環境・リスク分析を実施し、ゴールの設定や評価・運用の見直しといった後続の検討につなげることが可能となる。

✓ 検討責任者・検討に参加すべき者の確認

リスク管理者が新しく捉えたリスクに対し、誰が・どの枠組みの中で対応すべきかを検討する際に、本表のカテゴリ別整理を参照することが考えられる。個別の情報を対象にした詳細リスク分析や管理策の検討には、カテゴリの属性に適した責任者、所管部門をアサインすることが効果的であると考えられる。各論点には、AI の専門性に限らず複数の専門性が要求されることから、横断的な体制を構築することが望ましい。

例えば、リスクがプライバシーのカテゴリ上に該当し、自社サービスの提供に影響を与える可能性がある場合、プライバシーの責任者のリードのもと、自社サービスの各責任者が検討に加わり、対応を協議するような枠組みを設定することが可能である。

✓ カテゴリ軸に対応した分析、施策の検証（妥当性・網羅性）

各リスクの詳細分析を担当する責任者は、本表を分析観点の妥当性、網羅性の確認に活用することが出来る。例えば、事業戦略の策定を担当する経営企画の責任者は、第三者による生成 AI の活用が自社事業環境に与える影響を、イノベーションの機会に加え、倫理・社会的影響、著作権、プライバシー、セキュリティの関連から網羅的に確認し、作成する事業戦略の妥当性を検証することが出来る。

整理したリスクの全体像に基づき、検討に必要な要員と施策の分析を行うためには、例えば以下のように基本的な責任者、所管部門をあらかじめ可視化することも有効である。

表 3 必要な要員の検討（著作権に関する政府・当局による規制を扱う場合の例）

所管部門		総務・人事 情報システム	事業・サービス 情報システム	**経営企画、政策渉外、総務・人事**	経営企画、総務・人事、アライアンス
	カテゴリ	自社内の利用	自社サービスの提供	**政府・当局による規制**	第三者からの影響
事業企画 R&D	イノベーション ビジネス （機会）	生産性の向上 従業員満足度向上 セキュリティ対策等への生成 AI 活用	品質向上、新規性 市場の逆転	補助金等の可能性 DX 銘柄等	協業先の品質向上 競合の品質向上 優位な市場の逆転
経営企画 CSR 広報 リスク	倫理・社会 説明責任	不正確な出力による混乱、ミス 不適切なコンテンツの作成 説明可能性の欠如 デジタル格差	不正確な出力による混乱、損害 社会的バイアス、不公平の増強 不適切なコンテンツの公開 学習データの透明性	過剰な発展への懸念 不適切なコンテンツの学習・生成に関する規制 ポリティカルコレクトネス 国・地域間における方針のギャップ	社会的バイアス、不公平の増強 他社起点の社会的混乱への巻き添え
法務 リスク	**著作権 ライセンス**	他者の著作物を不作為に利用 自社の著作物が外部の AI モデルに学習される ソフトウェアライセンス違反	他者の著作物を不作為に外部提供 ユーザーの著作物を意図せず学習・外部提供 ソフトウェアライセンス違反	**生成 AI の特性を踏まえた著作権法等の対応 著作権に関する海外規制対応**	自社の著作物が第三者の AI モデルに利用される 利用している AI モデルに著作権法違反が発覚
法務 セキュリティ リスク	プライバシー	他者の個人情報を不作為に利用 自社の個人情報が外部の AI モデルに学習される	他者の個人情報を不作為に漏えい ユーザーの個人情報を意図せず学習、漏えい	生成 AI の特性を踏まえた個人情報保護等の対応 GDPR 等海外規制対応	自社の個人情報が第三者の AI モデルから漏えい 利用している AI モデルにプライバシー侵害が発覚
セキュリティ リスク	セキュリティ 安全性	自社の機密情報が外部の AI モデルに学習される	自社の生成 AI サービスの脆弱性を侵害される 自社の生成 AI サービスが悪意ある利用をされる 他者の機密情報を不作為に外部提供	生成 AI の特性を踏まえたセキュリティ規制対応 海外規制対応	サイバー攻撃者が生成 AI 技術を悪用 偽情報の拡散・ディープフェイク 自社の機密情報が第三者の AI モデルから漏えい 利用している AI モデルに脆弱性が発覚

以上から、新規の情報を詳細分析する課程において、次のように実施手順をまとめることができる。

新規情報の詳細分析手順例

1. マトリクスを参考に新規の情報のカテゴリを分析する
2. カテゴリに対応する責任者、所管部門に詳細分析の実施担当を割り当てる
3. 担当による詳細分析を実施する
4. 影響範囲を特定し、関連する計画や文書、継続検討タスク等の具体化を行う
5. 必要に応じてマトリクスを更新する

表中に用いた例でもある「EU が生成 AI の学習に使用したデータの開示を企業に義務づける法案を検討中」[12]という情報を新規に入手した場合の検討手順を示す。

1. マトリクス上の「著作権に関する政府・当局による規制」に分類可能と判断する
2. 法務部、リスク管理部、経営企画、政策渉外、総務・人事部から必要な人材を分析担当に割り当てる
3. 情報の詳細分析を行う

   *この場合、「文書生成 AI や画像生成 AI が主に対象であること」「EU の AI 法案へ盛り込む形での実現が検討されていること」「著作権コンテンツの学習自体を禁じる形ではなくデータの表示義務として検討されていること」「AI の生成物に AI によって生成されたコンテンツであることを義務付けることが検討されていること」「GDPR との関連でデータ利用の拒否権が議論されていること」などが着目要素として考えられる。
4. 法案の影響を受ける可能性がある規則、社内文書、海外事業計画などを確認する

   *この時点では、法案の内容が検討段階にあるため、影響範囲の特定と当件に関連する情報収集を継続することまでを決定し、具体的な変更内容の検討は保留するといった判断も考えられる。
5. マトリクス上で「著作権に関する政府・当局による規制」に関係するリスクの内容の見直しや新規カテゴリ追加の要否などを検討する

   *特に EU 事業が社内で重要である場合など、マトリクス上で「EU の規制」を明示的にテーマとして分離するような措置も考えられる。

この他にも「日本ディープラーニング協会が生成 AI の利用ガイドラインを公開[13]」という情報を入手した際、以下のような形で複数のカテゴリに関連する検討を行うことも考えられる。

1. 「自社内の利用」のカテゴリ全体に対応すると判断する
2. 総務・人事部や情報システム部の所管する社内規定の参考になる可能性があるため担当として割り当てる
3. ガイドラインの内容と自社で作成したマトリクスの内容を比較・分析する
4. 自社規定の作成、改訂に活用する
5. ガイドラインの情報から自社のマトリクスにないリスクが発見された場合、必要に応じてマトリクスを更新する

[12] https://www.reuters.com/technology/eu-lawmakers-committee-reaches-deal-artificial-intelligence-act-2023-04-27/
[13] https://www.jdla.org/document/#ai-guideline

4. まとめ

本レポートでは、生成 AI の急速な普及に対応するためのリスク整理の方法論を整理した。生成 AI のもたらす強力なイノベーションの可能性については周知の通りであり、大きな機会とリスクが眼前にある。生成 AI を積極的に活用することを考えるほど、リスクのコントロールは不可分の課題となる。

デジタル技術の進化は素早く、頻繁に進展する。司法や行政による規制は後追いにならざるをえない。ゴールを固定せず事業環境の分析と運用の改善を繰り返すアジャイル・ガバナンスは、こうした状況を乗り越えるために適した方法論となりえる。企業が生成 AI による変革で先んじるためには、不確実性に対応する迅速なリスクマネジメントのサイクルを構築するほかない。

また、生成 AI のイノベーションの影響は極めて広い範囲に及ぶことが見込まれ、企業には横断的な対応が求められる。本レポートで示した「生成 AI のリスク整理マトリクス」は、生成 AI の特性を踏まえた上で、検討すべきテーマと検討を担当すべき者を決定するための補助ツールとしての活用を期待するものである。企業のリスク管理担当者が、氾濫する情報を整理し、必要な人材と必要な分析検討を行うサイクルを構築する上で役立てていただきたい。

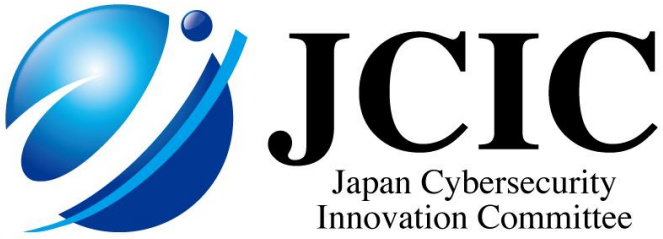
現状分析の中で述べたように、いま話題となっている生成 AI のポテンシャルは一部のアプリケーションが示したものに過ぎない。今後様々なステークホルダーによる活用が進む中で、文中で挙げたリスクの例などは早晩古めかしいものになってしまうだろう。重要なのは、リスク整理のフレームワークと改善のサイクルである。本レポートの内容が、様々な企業が生成 AI のリスクに向き合いながらも、そのポテンシャルを引き出す挑戦の一助となれば幸いである。

以　上

参考資料　生成 AI リスク管理マトリクス作成例

所管部門		総務・人事 情報システム	事業・サービス 情報システム	経営企画、政策渉外、総務・人事	経営企画、総務・人事、アライアンス
	カテゴリ	自社内の利用*1	自社サービスの提供*2	政府・当局による規制	第三者からの影響
事業企画 R&D	イノベーション ビジネス （機会）	生産性の向上 従業員満足度向上 セキュリティ対策等への生成 AI 活用	品質向上、新規性 市場の逆転	補助金等の可能性 DX 銘柄等	協業先の品質向上 競合の品質向上 優位な市場の逆転
経営企画 CSR 広報 リスク	倫理・社会 説明責任	不正確な出力による混乱、ミス 不適切なコンテンツの作成 説明可能性の欠如 デジタル格差	不正確な出力による混乱、損害 社会的バイアス、不公平の増強 不適切なコンテンツの公開 学習データの透明性	過剰な発展への懸念 不適切なコンテンツの学習・生成に関する規制 ポリティカルコレクトネス 国・地域間における方針のギャップ	社会的バイアス、不公平の増強 他社起点の社会的混乱への巻き添え
法務 リスク	著作権	他者の著作物を不作為に利用 自社の著作物が外部の AI モデルに学習される ソフトウェアライセンス違反	他者の著作物を不作為に外部提供 ユーザーの著作物を意図せず学習・外部提供 ソフトウェアライセンス違反	生成 AI の特性を踏まえた著作権法等の対応 著作権に関する海外規制対応	自社の著作物が第三者の AI モデルに利用される 利用している AI モデルに著作権法違反が発覚
法務 セキュリティ リスク	プライバシー	他者の個人情報を不作為に利用 自社の個人情報が外部の AI モデルに学習される	他者の個人情報を不作為に漏えい ユーザーの個人情報を意図せず学習、漏えい	生成 AI の特性を踏まえた個人情報保護等の対応 GDPR 等海外規制対応	自社の個人情報が第三者の AI モデルから漏えい 利用している AI モデルにプライバシー侵害が発覚
セキュリティ リスク	セキュリティ 安全性	自社の機密情報が外部の AI モデルに学習される	自社の生成 AI サービスの脆弱性を侵害される 自社の生成 AI サービスが悪意ある利用をされる 他者の機密情報を不作為に外部提供	生成 AI の特性を踏まえたセキュリティ規制対応 海外規制対応	サイバー攻撃者が生成 AI 技術を悪用する 偽情報の拡散・ディープフェイク 自社の機密情報が第三者の AI モデルから漏えい 利用している AI モデルに脆弱性が発覚

*1 グループ会社等での利用を含むことも想定される　*2 他社向けの受託サービス等を含むことも想定される



[本調査に関する照会先]

客員研究員　古澤一憲　furusawa@j-cic.com

JCIC 事務局　info@j-cic.com