生成AI活用のためのリスクマネジメント

PwC Japan有限責任監査法人
上席執行役員 リスクアシュアランス部長
パートナー　**綾部 泰二**

はじめに

生成AIの業務への活用を検討する企業が増えていますが、生産性の向上だけを目的に生成AIを導入することには多くのリスクが存在します。2023年には、G7でアジェンダとして取り上げられたり、著名な規格作成主体から生成AIのガイドラインが公表されるなど、世界が生成AIのリスクに目を向け始めています。

本稿では、2023年に発表された生成AIに関する制度やガイドラインの動向を解説するとともに、生成AIが企業経営にもたらす可能性のあるリスクや、リスク対応体制の必要性について考察します。

なお、文中の意見は筆者の私見であり、PwC Japan有限責任監査法人および所属部門の正式見解ではないことをお断りします。

1 生成AI活用におけるリスクマネジメントの必要性

まず、ビジネスに生成AIを取り込む際のリスクマネジメントの重要性について再確認します。

生成AIの応用分野には、画像生成、テキスト生成、動画生成、音声生成等、さまざまな種類があります。用途に応じて生成系AIを使い分けることで、多様な成果物を生み出すことができます。このように用途と成果物が多岐にわたる生成AIをいかに活用するかという点や、生成AIを活用した結果、発生することが想定されるリスクにどう対処するかが重要な経営課題となっています。また、活用により多数のリスクが想定される生成AIは、社会課題を生み出す要素のある技術であると言えます。

このような状況から、ルール化やガイドラインが必要であると言われて久しい状況です。

2 生成AIに関連する制度やガイドラインの動向

前述した状況への対応に向けた、2023年の生成AIに関連する制度やガイドラインの主な動向として、次の3点を紹介します。

(1) 広島AIプロセス

2023年5月に広島で開催されたG7広島サミットの結果を受けて、生成AIに関する国際的なルールの検討を行うため、「広島AIプロセス[※1]」が立ち上がりました。

この広島AIプロセスのアウトプットとして、「全てのAI関係者向けの広島プロセス国際指針」の12項目が公開されています。内容は次の通りです。

※1　総務省「広島AIプロセス」 https://www.soumu.go.jp/hiroshimaaiprocess/

「全てのAI関係者向けの広島プロセス国際指針」の12項目

① AIライフサイクル全体にわたるリスクを特定、評価、軽減するために、高度なAIシステムの開発全体を通じて、その導入前及び市場投入前も含め、適切な措置を講じる
② 市場投入を含む導入後、脆弱性、及び必要に応じて悪用されたインシデントやパターンを特定し、緩和する
③ 高度なAIシステムの能力、限界、適切・不適切な使用領域を公表し、十分な透明性の確保を支援することで、アカウンタビリティの向上に貢献する
④ 産業界、政府、市民社会、学界を含む、高度なAIシステムを開発する組織間での責任ある情報共有とインシデントの報告に向けて取り組む
⑤ 特に高度なAIシステム開発者に向けた、個人情報保護方針及び緩和策を含む、リスクベースのアプローチに基づくAIガバナンス及びリスク管理方針を策定し、実施し、開示する
⑥ AIのライフサイクル全体にわたり、物理的セキュリティ、サイバーセキュリティ、内部脅威に対する安全対策を含む、強固なセキュリティ管理に投資し、実施する
⑦ 技術的に可能な場合は、電子透かしやその他の技術等、ユーザーがAIが生成したコンテンツを識別できるようにするための、信頼できるコンテンツ認証及び来歴のメカニズムを開発し、導入する
⑧ 社会的、安全、セキュリティ上のリスクを軽減するための研究を優先し、効果的な軽減策への投資を優先する。
⑨ 世界の最大の課題、特に気候危機、世界保健、教育等（ただしこれらに限定されない）に対処するため、高度なAIシステムの開発を優先する
⑩ 国際的な技術規格の開発を推進し、適切な場合にはその採用を推進する
⑪ 適切なデータインプット対策を実施し、個人データ及び知的財産を保護する
⑫ 高度なAIシステムの信頼でき責任ある利用を促進し、貢献する。

（2）AIのマネジメントシステム「ISO/IEC 42001」

AIマネジメントシステムの国際規格がISO/IEC 42001として公開されました。経済産業省はWebサイトにおいて次のように解説しています[※2]。

2023年12月18日に発行された国際規格「AIマネジメントシステム（ISO/IEC 42001）」は、AIを開発、提供または使用する組織を対象として、AIシステムを適切に開発、提供、および使用するために必要なマネジメントシステムを構築するに際して遵守すべき要求事項について、リスクベースアプローチによって規定しています。

また、AIの開発、提供または使用に際して、①信頼性、②透明性、③説明責任が必要になりますが、①から③までのリスクを特定してマネジメントすることを求め、公平性や個人のプライバシーについてもその考慮を求めています。

（3）NIST「AIリスクマネジメントフレームワーク」

AIのリスク管理手法として、米国商務省の国立標準技術研究所（NIST）から「Artificial Intelligence Risk ManagementFramework（AI RMF 1.0）」（以下、AIリスクマネジメントフレームワーク）が2023年1月に公開されました[※3][※4]。同フレームワークでは7つのリスクで構成される「信頼できるAIシステムの特徴」として、**図表1**のようにリスク要素を特定しています。

ISO/IECやNISTという欧州や米国で著名な規格作成主体からのガイドラインの公表、およびG7でアジェンダとして取り上げられている点を鑑みると、生成AIが世界的に対応すべき技術であることは言うまでもない状況です。またこれらの3つの内容において一貫して活用されているキーワードは、リスクベースアプローチという言葉に基づく「リスクマネジメント」です。

皆さんの会社において、生成AIにおける各種リスクへの対応体制は整っているでしょうか。

リスク対応体制の整備は、昨年の主要な生成AIに関連する制度やガイドラインの公表状況からも急務であると言えます。

3 生成AIがもたらすリスクの具体例

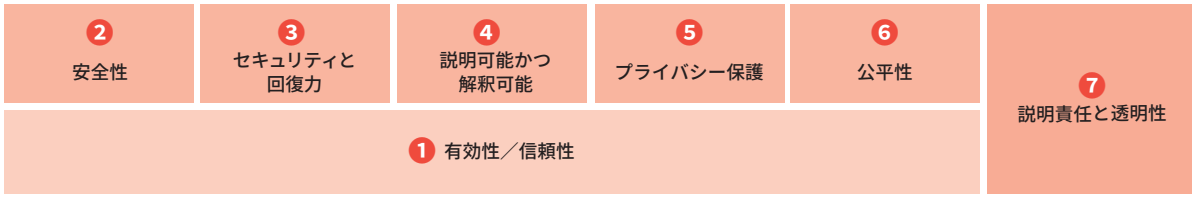
企業として生成AIのリスクマネジメントにおいてどのような対応が求められるか、上述した各種規格の内容も踏まえつつ具体的事例を考察していきます。

※2 経済産業省「AIマネジメントシステムの国際規格が発行されました」2024年1月15日
https://www.meti.go.jp/press/2023/01/20240115001/20240115001.html

※3 National Institute of Standards and Technology（NIST）「Artificial Intelligence Risk Management Framework（AI RMF 1.0）」2023年1月
https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

※4 PwCコンサルティング合同会社「NIST『AIリスクマネジメントフレームワーク（AI RMF）』の解説」2023年7月12日
https://www.pwc.com/jp/ja/knowledge/column/awareness-cyber-security/generative-ai-regulation04.html

図表1：信頼できるAIシステムの特徴



❶ 有効性／信頼性	システムに不具合がなく、意図したとおりに動作している状態
❷ 安全性	人間の生命、健康、財産、環境が危険にさらされていない状態
❸ セキュリティと回復力	不正アクセスや外部からの攻撃に耐えられる状態、予期せぬ事態から回復できる状態
❹ 説明可能かつ解釈可能	AIシステムの動作の根底にあるメカニズムを言語化でき、アウトプットが解釈できる状態
❺ プライバシー保護	個人情報やプライバシーが保護されている状態
❻ 公平性（有害なバイアス管理）	偏見や差別といった問題が起きず、平等性や公平性に配慮されている状態
❼ 説明責任と透明性	AIのライフサイクルに応じた適切なレベルの情報へのアクセスを提供できる状態

出所：NIST「AIリスクマネジメントフレームワーク」をもとにPwC作成

(1) 企業の信頼性

例えば、会計監査で求められるエビデンス類が生成AIで作成される可能性を想像した場合にはどのような影響があるでしょうか。

エビデンスを提出する会社側も監査する監査法人側も、それぞれの立場において、対象となるエビデンスは生成AIで捏造されたものではないと立証する責任があります。どのようにしてエビデンスの信頼性を担保すればよいでしょうか。現状において、簡易な方法でそれを立証する術はないと言えます。

会計監査におけるエビデンスの信頼性が担保されないということは、企業内開示制度の根幹が崩れることになり、企業が公表する財務諸表の信頼性が担保されないことになります。

(2) 企業の成長

生成AIを活用することで、今までの仕事の流れにおける「調査・検討」、「資料作成」、「意思決定」といった要素が、劇的に効率化が図られることが想定されます。

すなわち、「調査・検討」については不明な点について生成AIに問い合わせればよく、また「資料作成」についてもほとんどが生成AIを活用することによって行われ、「意思決定」についても生成AIからのレコメンデーションによって時間の短縮が図られることが想定されます。

このような、生成AIにより生産性が向上する業務を長年担当していた方であれば、そのメリットを享受することが現状では可能と言えるでしょう。しかし将来に目を移すと、生成AIにより効率化された業務は人が担当することはなくなっていくことが想定されます。業務担当者にノウハウを身に付ける術がなくなることが想定され、次のような点がリスクとなり、企業の成長阻害要因になると考えられます。

① 人に当該業務のノウハウがなくなる。
② ①から生成AIのアウトプットの正確性等の検証ができなくなる。
③ 行き過ぎると①と②が相まって、会社としてマネジメント可能な業務ではなくなる。

このように、単に生産性の向上だけを目的として積極的に生成AIを導入していくことは、企業活動のさまざまな面でリスク要因を発生させることになるため、避けるべきです。

また、生成AIを活用しようとする部署だけのリスクマネジメントでは対応不可能なリスク要因であることは(1)と(2)の具体例からイメージを持って頂ければと思います。

4 全社的な生成AIのリスクマネジメントからガバナンス体制確立へ

生成AIの活用は企業にとって不可避であり、企業の成長に大きく寄与する可能性を持っていることは言うまでもありません。

しかしながら、上述のように生成AIの活用においてそのリ

スクを認識しない場合には、企業活動の信頼性や成長性が損なわれるリスクがあることも事実です。

生成AIを活用するためにも、改めて生成AIに対するリスクマネジメントおよび当該リスクマネジメントの実効性を担保するガバナンスの構築を全社的に行うこと、すなわち、生成AIという成長のためのアクセルを安心して踏めるガードレールの構築が必要であると言えます。

本稿が生成AIを活用した企業成長の一助となれば幸いです。

綾部 泰二（あやべ たいじ）

PwC Japan有限責任監査法人
上席執行役員 リスクアシュアランス部長　パートナー
2001年入所、2006年CISA（公認情報システム監査人）。以後、セキュリティやITガバナンス等のリスクマネジメント業務に多数従事。2019年7月よりPwC Japanグループのサイバーセキュリティ Co-Leaderを務める。共著に『クラウド・リスク・マネジメント』（同文舘出版）、『経営監査へのアプローチ――企業価値向上のための総合的内部監査10の視点』（清文社）がある。
メールアドレス：taiji.t.ayabe@pwc.com