

AI のリスクアセスメントガイドライン

生成 AI 編

特定非営利活動法人 リスク共生社会推進センター

先進技術リスク検討ユニット

第 1 版

2026 年 8 月 3 日

改訂履歴

版数	発行日	改訂内容
第 1 版	2026 年 7 月	初版

目 次

1. 本ガイドラインのスコープ	5
1.1. 本ガイドラインの目的.....	5
1.2. 対象とする AI.....	5
1.3. 本ガイドラインの利用対象者	5
2. 用語定義（参考 JIS Q 31000）	6
3. AI のリスクアセスメントの基本.....	6
3.1. AI のリスクアセスメントの基本ステップ	6
3.2. 既存の工学システムのリスクアセスメント手法を適用する上での課題.....	7
3.3. リスク特定の前提となる目的の設定について.....	7
4. リスク特定	7
4.1. リスク特定の基本ステップ	8
4.2. リスクの捉え方の整理.....	8
4.3. 目的の設定(目的の視点からリスクを特定する場合のリスク特定の考え方).....	9
4.4. AI 活用者の目的の視点からリスク特定する例	10
4.4.1. 目的の体系的整理例.....	10
4.4.2. 目的毎のリスク特定例	11
4.5. AI 開発者の目的の視点からリスク特定する例	12
4.5.1. 目的の体系的整理例.....	12
4.5.2. 目的毎のリスク特定例	12
5. リスク分析	13
5.1. リスク分析の基本ステップ	13
5.2. リスク分析の前提の整理	14
5.2.1. リスク分析の前提となる組織・個人の内外の状況確定	14
5.2.2. リスク分析の対象となる AI の特徴.....	14
5.2.3. 組織・個人の内部状況.....	14
5.2.4. 組織・個人の外部状況.....	15
5.3. リスク分析の適用範囲の設定	15
5.4. リスク基準の設定.....	15
5.5. 対象リスクの性質の整理	16
5.6. リスク分析の手法の選定	16
5.7. リスク分析の実施.....	17
5.7.1. リスク分析の実施における留意事項.....	17
5.8. リスク分析の結果の整理	17
6. リスク評価	18

6.1. リスク評価の基本ステップ	18
6.2. リスク分析の結果の整理	18
6.3. 目的毎の評価	18
6.4. リスク評価による決定事項の整理	19
7. 今後の科学技術社会における AI の活用に向けて	20

1. 本ガイドラインのスコープ

1.1. 本ガイドラインの目的

本ガイドラインは、AIに関するリスクアセスメントを行うに際して、生成 AI を題材とした場合のリスクアセスメントを行うためのガイドラインである。

本ガイドラインは、AI のリスクアセスメントにおける各ステップである「リスク特定」「リスク分析」「リスク評価」について、その方法論を中心に記載している。

本ガイドラインを活用する者は、その活用目的に添って、本ガイドラインに従って必要なリスクアセスメントを実施されたい。

また、本ガイドラインの活用に際して、本ガイドラインはあくまで AI のリスクアセスメントの実施方法を述べたものであり、リスクアセスメントの結果を示すものではないことに注意されたい。

1.2. 対象とする AI

本ガイドラインで対象とする AI を「生成 AI」とする。ここでの生成 AI とは、既存データの分析や分類、予測によって新しいコンテンツ創造を行う機能を持つ AI のことを指す。

AI には、直接的に機器をコントロールするものや、生成 AI のように個人や組織に知識を提供するもの等がある。本ガイドラインのリスクアセスメントの対象として「生成 AI」を選定したのは、AI の中でも生成 AI が与える影響は、これまでの工学システムにおいて考慮されてきた物理的影響とは異なり、個人の知識や認識に加えて、個人の成長への影響などの特有の性質を持つためである。そのため、生成 AI のリスクは、その活用者によって受けるリスクが大きく異なるという特徴を持つ。

本ガイドラインではこの特徴に着目して、「生成 AI」を対象としてリスクアセスメントガイドラインを作成した。以下、本ガイドラインでは、生成 AI を「AI」と記す。

本ガイドラインでは、AI のレベルやバージョン等を限定せずに AI の持つ「アウトプットとして情報を作成・提供する」という機能を対象としてリスクアセスメントの考え方を記載する。

また、リスク分析を実施する際は、対象とする AI の対象製品を限定する場合もある。

1.3. 本ガイドラインの利用対象者

本ガイドラインの利用対象者は、「何らかの目的を持って AI を活用しようとする個人または組織」とする。ここでの「個人」とは、AI を活用する目的と一定の活用技術を持つ者のことを言う。また、ここでの「組織」とは、AI を活用する目的と一定の活用技術を持つ二人以上の集団のことを言う。対象とする個人または組織の捉え方はリスク分析の対象や適用範囲によって異なるため、リスク分析の前提条件として設定する必要がある。

なお、本ガイドラインは「既存の工学システムのリスクアセスメントに関する一定の知見と経験を有し、これから AI のリスクアセスメントを実施しようとしている方」として、既存

の工学システムのリスクアセスメントとの違いを捉えやすい記載となるように努めている。

2. 用語説明（参考 JIS Q 31000）

用語	定義
AI 開発者	AI プログラム・システムを開発する者。
AI 活用者	目的をもって AI を利用、活用する者。
生成 AI	大量のデータを学習した AI が、新しい文章・画像・音声・プログラムコードなどのコンテンツを自動的に作り出す技術またはサービスの総称。本ガイドラインにおいて AI は生成 AI をいう。
目的	組織や個人の活動の目的。
リスク	目的に対する不確かさの影響。
リスクアセスメント	リスク特定、リスク分析及びリスク評価を網羅するプロセス全体を指す。
リスク基準	リスクの重大性を評価するための目安とする条件。
リスク対応	リスクを修正するプロセス。
リスク特定	リスクを発見、認識および記述するプロセス。
リスク評価	リスク及び／またはその大きさが受容可能か又は許容可能かを決定するために、リスク分析の結果をリスク基準と比較するプロセス。
リスク分析	リスクの特徴を理解し、リスクレベルを決定するプロセス。
リスクマネジメント	リスクについて、組織を指揮統制するための調整された活動。
リスクレベル	結果とその起こりやすさとの組み合わせとして表される、リスクまたは組み合わせったリスクの大きさ。

3. AI のリスクアセスメントの基本

3.1. AI のリスクアセスメントの基本ステップ

本ガイドラインにおけるリスクアセスメントとは、ISO31000 のリスクマネジメントにおけるリスクアセスメントとして設定されている、「リスク特定」、「リスク分析」、「リスク評価」にて構成される一連のステップをさす。

ステップ 1 リスク特定 …4. 章

AI の活用・非活用により発生する可能性のある影響の中で、何をリスクと考えるかを判断する。

ステップ 2 リスク分析 …5. 章

ステップ1で特定したリスクのうち、分析対象としたリスクに関して、そのリスクの性質を明らかにする。このとき明らかにすべきリスクの性質としては、リスクシナリオ、リスクシナリオの進展のキー要素、リスクシナリオの発生確率、リスクシナリオの影響の種類や大きさ等がある。

ステップ3 リスク評価 ……6. 章

ステップ2で分析したリスク情報を元にリスクへの対応方針を判断する。

3.2. 既存の工学システムのリスクアセスメント手法を適用する上での課題

既存の工学システムのリスクアセスメントにおいては、すでに明示されているハザードを初期事象とした分析によりリスクを明らかにするという方法がとられる場合が多い。例えば、化学プロセスにおける事故や災害では、取り扱う化学物質の可燃性や爆発性などのハザードはすでに知られており、それによって顕在化するリスクを分析/アセスメントすることが必要であることは自明である。しかし、AIのような先進技術は未だ活用事例が少なく、そもそも何をリスクと考えるべきかが明示的でない場合が多い。したがって、AIのリスクアセスメントにおいては「リスク特定」のステップが重要であり、まずはAIのリスクを特定するための考え方を理解することが重要である。

3.3. リスク特定の前提となる目的の設定について

リスク特定においては、リスクアセスメントと同様に目的を設定した上で実施することが重要である。ただし、本ガイドラインにおいては、リスク特定における目的の意味内容自体の是非については言及しない。

また、ある目的に対して実施されたリスク特定の結果は、あくまでもその単一の目的に対して特定されたリスクを示したものであり、それが当該リスクアセスメントにおいて対象とした個人や組織におけるすべてのリスクを示しているわけではないことにも注意が必要である。個人や組織は、特定の単一の目的のために活動を行っているわけではなく、目的を複数持つ場合もある。したがって、個人もしくは組織全体のあり方を総合的に判断するためには、その個人や組織の持つ目的を体系的に設定し、それぞれの目的に対して特定・分析されたリスクを議論の俎上に載せた上で、総合的に評価および意思決定する必要がある。

さらに、個人や組織の目的と、それらが帰属している上位環境としての社会等の目的とが一致しないことにより課題が顕在化する場合もある。その場合には、個人や組織が、当該上位環境の視点で目的を設定した場合のリスクについて検討し、本ガイドラインで特定する個人や組織のリスク分析の結果と合わせて判断する必要がある。

4. リスク特定

AIのリスク特定とは、AIの活用・非活用により発生する可能性のある影響の中で何をリ

スクと考えるかを判断する事である。

4.1. リスク特定の基本ステップ

リスク特定とは、リスク分析の対象とするリスクを定める事である。既存の工学システムのリスクアセスメントにおいてこのステップがあまり必要とされてこなかった理由の1つは、リスクを分析する必要があると経験的に認識されている事故やトラブルを主なリスク分析の対象としてきたからである。一方、AIのような先進技術においてはその経験が十分に蓄積されておらず、その重要な課題も事前に明らかではないため、従来の経験的認識のみを元にリスク分析を行うことは、一連のリスクアセスメントにおいて分析すべきリスクの見落としにつながる可能性がある。したがって、先進技術システムをリスクアセスメントの対象とする場合は、「何をリスクとして捉えるか」を意味するリスク特定が特に重要になる。

ステップ 1 リスクの捉え方の整理 ……4.2. 節

ステップ 2 目的の設定 ……4.3. 節

ステップ 3 目的ごとのリスク特定

AI 活用者の目的の視点からリスク特定する例 ……4.4. 節

AI 開発者の目的の視点からリスク特定する例 ……4.5. 節

4.2. リスクの捉え方の整理

AI のリスク特定を考える際には、以下のように「AI 開発者の目的」と「AI 活用者の目的」の双方からの視点でリスクを特定することが必要である。なぜなら、1. 2. 節で述べたように、AI のリスクは、特に AI 活用者の特性や性質の違いによって受けるリスクが大きく異なるという特徴を持つからである。

- AI 開発者の目的の視点からのリスク特定

この視点は、「影響を与える側の視点」「AI の持つ機能の視点」からのリスク特定である。

- AI 活用者の目的の視点からのリスク特定

この視点は、「影響を受ける側の視点」「個人や組織が持つ機能の視点」からのリスク特定である。

この双方の視点を切り分けてそれぞれリスクを特定することにより、「影響を与える側の視点」に着目した実施のみに留まる傾向があった従来のリスクアセスメントに対して、リスク特定のステップをより意識的に取り込んだ形として実施することができる。

また、本ガイドラインではリスクの定義として、ISO31000 における定義である「目的に対する不確かさの影響」を用いる。

AI リスクにおいて検討すべき影響には「好ましくない影響」だけでなく、「好ましい影響」も含まれる。当該定義における「影響」は「好ましい影響」と「好ましくない影響」の双方の要素を含むものであり、AI リスクにおいて検討すべき両者の影響を「リスク」というかたちで取りま

とめて捉えることができる。

一方で、影響を「好ましくない影響」のみで捉える方法もある。その場合は、

- AI 開発者の目的の視点からは
 - 「AI を開発することによる(好ましくない影響を持つ)リスク」
 - 「AI を開発しないことによる(好ましくない影響を持つ)リスク」
- AI 活用者の目的の視点からは
 - 「AI を活用することによる(好ましくない影響を持つ)リスク」
 - 「AI を活用しないことによる(好ましくない影響を持つ)リスク」

として切り分けることで、検討すべき影響をそれぞれ捉えることができる。

例えば、AIの導入・活用を意図したある組織において「業務効率化する」ことを目的とした場合のAIのリスクは、以下のように捉えることができる。

- 影響として「好ましい影響」と「好ましくない影響」の双方を取りまとめて捉える場合
 - 「業務効率化リスク(業務効率化ができる/できない場合の双方を含む概念)」
- 影響として「好ましくない影響」のみを捉える場合
 - AI 活用時を想定した「目的とした業務効率化ができないリスク」
 - AI 非活用時を想定した「業務効率化ができないリスク」

なお、本ガイドラインにおけるリスク特定では、わかりやすさのために、後者の「影響として『好ましくない影響』のみを考える場合のリスクの捉え方」を採用する。このとき上記の通り、実際にはAI活用時だけでなくAI非活用時を想定したリスク特定も実施されるべきであるが、ここではAIの導入・活用を前提している組織・個人にとってのリスクを検討する観点から、AI非活用時のリスク特定については記載を省略する。

また、リスク特定の方法は本ガイドラインで解説する方法以外にも存在するが、それらについては今後別途整理を行う。

4.3. 目的の設定(目的の視点からリスクを特定する場合のリスク特定の考え方)

目的の視点からリスクを特定する場合は、以下の2つの視点によるリスク特定をセットで考えることが望ましい。

- (1) 目的が達成されないリスク
- (2) 目的の達成を目指すことで発生する可能性のあるリスク

(2)の視点は、前節4.2.で記述した複数の目的設定によって異なるリスクに対応する上での課題を事前に検討の対象とするものでもある。これは、「可能性のある目的を体系的に捉えることが望ましい」という考え方に基づくものであり、AIがもたらす社会への影響も踏まえて考えた場合には、個々の目的の視点においてリスクを特定するだけでは多様なリスクの一面しか捉えられない可能性があることを反映している。

AIのリスク特定は、設定された各目的の達成を妨害、低下、または遅延するかもしれない事象を検討するものである。その際、以下の事項を考慮することが望ましい。

- 有形及び無形のリスク源、原因及び事象
- 脅威及び機会、脆弱性及び能力
- 外部及び内部の状況の変化
- 新たに発生するリスクの指標
- 資産及び使用できる資源の性質及び価値
- 結果及び結果が目的に与える影響

なお、設定された各目的を達成するために追加されるリスク対応によって新たに発生するリスクについては、リスク評価実施後のリスク対応(リスクマネジメントの範疇のため、本ガイドラインではスコープ外)の検討の中で明らかになる場合もあることに注意されたい。

※参考:ISO31000 におけるリスクの考え方

ISO31000 では、リスクを以下のように定義している。

リスク:目的に対する不確かさの影響

注記 1 影響とは、期待されていることから乖離することをいう。影響には、好ましいもの、好ましくないもの、又はその両方の場合があり得る。

以降略

また、リスク特定に関しては、「設定された各目標の達成を促進、妨害、低下、または遅延するかもしれない事象を検討する」としており、リスク特定の際の考慮事項として以下の事項を示している。

- 有形及び無形のリスク源、原因及び事象
- 脅威及び機会、脆弱性及び能力
- 外部及び内部の状況の変化
- 新たに発生するリスクの指標
- 資産及び組織の資源の性質及び価値
- 結果及び結果が目的に与える影響
- 知識の限界及び情報の信頼性 等

4.4. AI 活用者の目的の視点からリスク特定する例

ここでは、AI 活用者の目的の視点からのリスク特定に関して説明する。

4.4.1. 目的の体系的整理例

本ガイドラインでいう目的とは、組織や個人が AI を活用する目的をいう。この目的をどの

ように設定するかによって特定されるリスクが異なるために、目的の体系的な整理が必要である。以下に AI 活用者の目的の設定例を示す。なお、下記はあくまで複数例を示しているのみであり、体系的な整理はなされていないことに注意されたい。

- (1) 作業を効率化すること
- (2) 競争力を強化すること
- (3) 人材を有効活用すること
- (4) 自分で実施しても十分な成果を期待できないことを実現すること
- (5) 人間にとって時間がかかる判断を素早く実施すること
- (6) 複雑なデータを短時間で分析・整理すること 等

4.4.2.目的毎のリスク特定例

以下、ある個別の目的を設定した場合のリスク特定例を示す。

【目的例 1】作業を効率化すること

- (1) 作業を効率化することができないリスク
 - (ア) 効率化自体が不十分(質・時間)となるリスク
 - (イ) 生成した成果物の確認に要するリソースが増加するリスク
- (2) 作業を効率化することを目指すことで発生する可能性のあるリスク
 - (ア) 質の悪い成果物が発生するリスク
 - ① 品質が低下するリスク
 - ② 法律に違反する成果物が生成されるリスク(成果物作成者が法律違反、社会的犠牲者の発生)
 - (イ) 機密情報が漏洩するリスク
 - (ウ) AI 利用のための費用が増大するリスク

【目的例 2】競争力を強化すること

- (1) 競争力を強化することができないリスク
 - (ア) 品質が低下するリスク
 - (イ) 他社との類似製品の開発が頻発するリスク
- (2) 競争力を強化することを目指すことで発生する可能性のあるリスク
 - (ア) 法律違反が発生するリスク(成果を急ぐため、経験のない領域での活動のため)
 - (イ) 成果が出る活動への投資が集中するリスク

【目的例 3】人材を有効活用すること

- (1) 人材を有効活用することができないリスク

- (ア) 活用できる人材が不足するリスク
- (イ) AI の能力不足により人材の有効活用ができないリスク
- (ウ) 余剰人材が他の業務への転換ができないリスク
- (2) 人材を有効活用することを目指すことで発生する可能性のあるリスク
 - (ア) AI 活用者の経験不足による能力・技術力不足が発生するリスク
 - (イ) AI 活用により、他社と同様の状況になり競争力に差が付かないリスク

4.5. AI 開発者の目的の視点からリスク特定する例

ここでは、AI 開発者の目的の視点からのリスク特定に関して説明する。

4.5.1. 目的の体系的整理例

4.4.1. で述べたのと同様に、ここでも目的の体系的な整理が必要である。以下に AI 開発者の目的の設定例を示す。なお、下記はあくまで複数例を示しているのみであり、体系的な整理はなされていないことに注意されたい。

- (1) 既存の作業を効率的に行うこと
- (2) 新たなコンテンツ作成を簡単に行うこと
- (3) 複雑なデータを処理して、精度良く分析すること
- (4) AI の処理速度を上げること
- (5) 消費電力を小さくすること
- (6) 多様な国や組織の需要に対応すること
- (7) 多様な用途のシステムに活用すること 等

4.5.2. 目的毎のリスク特定例

以下、ある個別の目的を設定した場合のリスク特定例を示す。

【目的例 1】既存の作業を効率的に行うこと

- (1) 既存の作業を効率的に行うことができないリスク
 - (ア) 品質が劣化するリスク
 - (イ) 作成時間・コスト増大するリスク
 - (ウ) AI 製品が販売不振に陥るリスク
- (2) 既存の作業を効率的に行うことを目指すことで発生する可能性のあるリスク
 - (ア) 既存の作業者が失職するリスク
 - ① これまでの業務が消失するリスク
 - ② 他の職種への転換が失敗するリスク
 - (イ) 勉強・経験の機会が喪失するリスク
 - (ウ) 品質が劣化するリスク

(エ) 業務価値が低下するリスク

【目的例 2】新たなコンテンツ作成を簡単に行うこと

- (1) 新たなコンテンツ作成を簡単に行うことができないリスク
 - (ア) 品質が劣化するリスク
 - (イ) 作成時間・コストが増大するリスク
 - (ウ) 事業が失敗するリスク
 - (エ) AI 製品が販売不振に陥るリスク
- (2) 新たなコンテンツ作成を簡単に行うことを目指すことで発生する可能性のあるリスク
 - (ア) 類似製品サービスが多発するリスク
 - (イ) 開発者が失職するリスク
 - (ウ) 業務価値が低下するリスク
 - (エ) 不正製品が多発するリスク
 - ① 作成者が法律違反を犯すリスク
 - ② 著作権侵害を受ける者が増加するリスク
 - ③ 社会的混乱(世論操作、プライバシー侵害等)が生じるリスク

5. リスク分析

5.1. リスク分析の基本ステップ

リスク分析は、リスク特定の段階で特定されたリスクを対象に、リスクのレベル、リスクの性質及び特徴を理解し、リスクの評価・判断に資する情報を提供するための活動である。リスクの性質及び特徴に関しては、不確かさ、リスク源、結果、起こりやすさ、事象、シナリオ、管理策及び管理策の有効性の詳細な検討等が含まれる。

ここで注意が必要なのは、**AIに関するリスク分析では、リスク情報が起りやすさと影響の大きさの組み合わせで表現できるとは限らない**ということである。したがって、これまで既存の工学システムのリスクアセスメント等で行われてきた起りやすさと影響の分析以外にも、リスク評価・判断へ有効な情報を提供する観点で、リスク対応の必要性及び方法やその判断の分岐をもたらす要素、最適なリスク対応の戦略及び方法の決定のための情報等も含めて分析する必要がある。

リスク分析の結果として、どのような情報が必要となるかは、AI のリスクに関する判断をどのレベルで実施するかによっても異なってくるため、リスクアセスメントの目的に合わせて実施するレベルを決定することが望ましい。

- | | | |
|--------|---------------|----------|
| ステップ 1 | リスク分析の前提の整理 | ……5.2. 節 |
| ステップ 2 | リスク分析の適用範囲の設定 | ……5.3. 節 |

ステップ 3	リスク基準の設定	……5.4. 節
ステップ 4	対象リスクの性質の整理	……5.5. 節
ステップ 5	リスク分析の手法の選定	……5.6. 節
ステップ 6	リスク分析の実施	……5.7. 節
ステップ 7	リスク分析の結果の整理	……5.8. 節

5.2. リスク分析の前提の整理

リスク分析では、のちの意思決定や判断を行うために必要な情報を体系的に整理する必要があるが、その分析条件は多様であり、一度に全ての状況に対する分析を行うことは難しい。したがって、リスク分析を行う際はその前提を明確にすることで、そのときに考慮している条件および考慮していない条件を明らかにする必要がある。

5.2.1. リスク分析の前提となる組織・個人の内外の状況確定

AI のリスク分析を行う際の前提条件を設定するためには、その組織・個人の内外における状況を調査し設定する必要がある。AI のリスクは、その AI を誰が何の目的で活用するかによって、その影響が大きく変化する可能性がある。したがって、AI のリスク分析においては AI そのものに関する工学的知識と共に、それを活用する組織や人間に関する知識や見識が必要となる。

5.2.2. リスク分析の対象となる AI の特徴

- (1) 使用目的(悪意も含む)によって有効性は異なる
- (2) 前提とするデータによって回答が異なる
- (3) 複数視点をカバーしているわけではない
- (4) 間違った情報を生成する可能性もある
- (5) 生成する回答が社会的倫理、法規等を満足する保証は無い
- (6) 質問のしかたによって回答が異なる
- (7) 使用するデータ・手法・成果に法的な問題が生じる可能性がある 等

5.2.3. 組織・個人の内部状況

- (1) 組織の価値観及び文化・風土
- (2) AI を適用しようとする業務内容
- (3) 業務内容や専門知識そして AI に関するリソースや知識の観点からみた状況・可能性(例:資本金、人材、能力、プロセス、システム、技術力等)
- (4) マネジメントの仕組み(経営方針、ガバナンス、人材育成方針、権利・責任体制等)
- (5) 達成しようとしている目的及びその戦略
- (6) 内部ユニットのステークホルダの位置づけ

- (7) 組織の AI 活用に関する規定 等

5.2.4.組織・個人の外部状況

- (1) 対象となる AI の進化状況(2.1.1. 参照)
- (2) 社会における AI の活用状況・予測
- (3) 社会の文化・教育、政治、法律、規制、経済及び環境(海外、国内、地域的なものに係らない)
- (4) 外部のステークホルダの認知度と評価
 - (ア) 法的制約
 - (イ) 事業内容
 - (ウ) AI
- (5) 組織の目的に影響を与える主な要因及び動向
 - (ア) 市民の価値観
 - (イ) 競争相手の状況 等

5.3. リスク分析の適用範囲の設定

リスクを分析する際は、その時の分析範囲を明確にして、その条件下で体系的に分析を行う必要がある。

マネジメントプロセスは、検討される適用範囲、検討される関連目的、並びにそれらと組織の目的との整合を明確にする。

- (1) 対象とする AI の機能範囲、レベル
- (2) 分析の対象とする期間、組織
- (3) 分析の対象とする AI の活用範囲
- (4) 目的、及び下す必要のある決定
- (5) AI のリスクに対して取られる対策によって期待される結末
- (6) AI の活用に必要とされる資源、責任、及び残すべき記録
- (7) 他のプロジェクト、プロセス及び活動との関係

5.4. リスク基準の設定

リスク基準は、組織の価値観、目的及び資源、リスクマネジメント方針を反映することが望ましい。また、リスク基準は、組織の義務及びステークホルダの見解を考慮に入れて規定することが望ましい。

リスク基準は、リスクアセスメントプロセスの開始時に確定することが望ましい。また、リスク基準は動的なものであるため、継続的にレビューを行い、必要に応じて修正することが望ましい。

- (1) AI に関するリスク基準を設定する際の考慮事項

- 特定したリスクに影響を与える不確かさの特質及び種類
- 結果及び起こりやすさをどのように定め測定(一貫性)するか
- 経時変化に関連する要素
- リスクレベルの決定法
- 複数のリスクの組合せ及び順序をどのように考慮に入れるか

(2) AI のリスク基準例

- 例 1 目的とする機能を平均的に満足する
- 例 2 目的とする機能を高い水準で満足する
- 例 3 目的とする機能を平均的に満足し、大きな好ましくない影響を与えない
- 例 4 目的とする機能を高い水準で満足し、大きな好ましくない影響を与えない
- 例 5 例 4 に加え、新たな気づきを与えてくれる

5.5. 対象リスクの性質の整理

AI のリスク分析を行うにあたって踏まえるべき、いくつかのAIのリスクの性質がある。

- (1) AIのリスクには、同様な AI の活用をした場合に AI 活用者によって影響が大きく変わらない(同様の影響を及ぼす)リスクもあるが、AI 活用者によって影響が大きく異なるリスクもある。したがって、このようなリスクに関しては、影響を受ける対象となる組織や個人の性質についての分析が重要になる。
- (2) AI のリスクには、その影響がリスクであることを多くの AI 活用者が共有するまでに長期間を要するリスクもある。したがって、どの時点および期間におけるリスクを分析すると設定するか、が重要な要件となる。
- (3) AIのリスクには、「AIを活用することによるリスク」と「AIを活用しないことによるリスク」がある。したがって、AIの活用に関する総合的判断を行うためには、その双方の検討が重要である。
- (4) AIのリスクには、その影響の大きさや発生確率の定量的な分析が可能なリスクとそうでないリスクがある。したがって、必ずしも従来の工学システムにおけるリスク分析において用いられてきた考え方や分析手法が適用できない可能性がある。

5.6. リスク分析の手法の選定

使用するリスク分析手法は、リスクアセスメントの目的によって様々な手法が使用できるが、以下にその例を示す。

- (1) 使用者の目的とする価値が明確で、その毀損されるリスクを分析したい場合
 - Fault Tree (FT)手法の活用

頂上事象に、「確保したい価値が毀損される」という事象を設定し、その原因を演繹的に分析する
- (2) AI の機能がもたらす多様なリスクを分析したい場合

- Event Tree (ET)手法の活用

初期事象に「対象とする AI の機能の課題(特徴)」を設定し、多様な分岐事項(法律に違反しない活用をする等の事象)を設定し、の成否の組み合わせによって、どの様な事象が発生するかを分析する。なお、分岐事象の失敗シナリオの展開は、FT 等を活用すると有効に分析できる。

5.7. リスク分析の実施

AI のリスク分析においては、整理した内外の状況特定情報を元に、設定した適用範囲に留意しつつ、リスク分析を行う。その際、AI のリスク分析内容は、設定したリスク基準と比較することを前提として定める。

5.7.1. リスク分析の実施における留意事項

AI 活用のリスク分析に関しては、以下の手順がある。

手順① 影響のみに着目し、検討の優先順位の決定

- 優先順位 1・・・影響が大きいと想定するリスク
- 優先順位 2・・・影響の可能性があると想定するリスク
- 優先順位 3・・・影響の有無を確認する必要があるリスク

優先順位に従い実施する AI のリスク分析の検討を行う。

手順② 事象の起こりやすさ及び結果の分析

- AI の活用における事象の起こりやすさや影響は、対象事業や活用者、事前に行う教育等によって異なることに留意すべきである。

手順③ 結果の性質及び大きさ、複雑さ及び結合性の分析

- 起こりやすさや影響の種類・大きさは、一定の分散に収まるとは限らず、また正規分布のような形になるとも限らない。

手順④ 時間に関する要素及び変動性の分析

- AI 活用に関する事前教育のあり方や導入のステップ等で、リスクは大きく変化する可能性がある。
- また、長期間活用した後に、影響が顕在化する場合もあることに注意が必要である。

手順⑤ 既存の管理策の有効性、機微性及び機密レベルの分析

- これまでの事業活動に関するリスク対策の有効性は、改めて検討する必要がある。

5.8. リスク分析の結果の整理

リスク分析は、リスク評価のための情報の作成である。以下の影響を検討し、文書化し、意

思決定者に伝達することが望ましい。

- 意見の相違、先入観、リスク認知及び判断によって影響される事項
- 使用される情報の質、加えられた前提及び除外、手法の限界、並びに 実行方法

6. リスク評価

6.1. リスク評価の基本ステップ

リスク評価は、分析したリスク情報を元にリスクへの対応方針を判断する活動である。リスク評価はリスクに関する決定を裏付けるものであるが、リスク評価の結果をもって自動的に対応が決定するわけでは無い。それは、実施したリスク分析には必ずその前提があり、対応の決定に必要な前提が設定されているとは限らないからである。実際に意思決定を行う際には、リスクがおかれているさらに広い範囲の状況について考慮し、法律、規制、ならびにその他の要求事項を満足する必要がある。

リスク評価は、リスク基準とリスク分析の結果を比較できる場合は、その比較した結果をもとに行う。一方、リスク基準とリスク分析の結果の直接的な比較が難しい場合は、組織・個人の目的と分析したリスクの特徴を照らし合わせて判断を行う。

AI への対応を検討する場合には、そのリスク分析が前提とする事項を良く理解することが望ましい。そのため、リスク評価の結果を記録し、伝達し、更に検証することが望ましい。

- | | | |
|--------|-----------------|---------|
| ステップ 1 | リスク分析の結果の整理 | …6.2. 節 |
| ステップ 2 | 目的毎の評価 | …6.3. 節 |
| ステップ 3 | リスク評価による決定事項の整理 | …6.4. 節 |

6.2. リスク分析の結果の整理

リスク評価を実施する者の立場で、リスク分析結果の再整理を行う。リスク分析により、影響や起こりやすさが一定の定量性をもって分析できている場合は、リスクマトリクスによる整理も有効である。また、影響や起こりやすさに大きな変化を及ぼすトリガーとなる事項等が明らかになった場合は、それも有効な判断基準となる。さらに、影響や起こりやすさ等の不確かさの大きさも、評価の重要な要素となる。

6.3. 目的毎の評価

AI に関するリスク評価は、複数存在する AI 開発者および AI 活用者の目的によってそれぞれ結果が異なる場合がある。総合評価を行う際には、目的ごとにリスク分析の結果を整理することが望ましい。

例：AI 活用による、目的毎のリスク評価

- | | |
|-------------|---|
| 省力化したい | ○ |
| 社員の能力を開発したい | △ |

独自性を題したい

×

6.4. リスク評価による決定事項の整理

リスク評価による決定事項の例を、AI の活用に関する対応判断を題材に説明する。

(1) 更なる活動は行わない

- 好ましくない影響を対策で必要なレベルに低下できない場合、リスクを生じさせる活動を開始又は継続しないと決定することによってリスクを回避する
 - ◇ AI の活用に関する判断例:対象とする AI を活用しないまたは活用を継続しない

(2) リスク対応の選択肢を検討する

- リスク分析の結果とリスク基準の関係で、対応方針を決定する
- リスクを生じさせる活動を開始又は継続しないと決定することによってリスクを回避する
- ある機会を追求するために、そのリスクを取る又は増加させる
- リスク源を除去する
- 起こりやすさを変える
- 結果を変える
- そのリスクを共有する(例:契約、保険購入)
- 情報に基づいた意思決定によって、そのリスクを保有する
 - ◇ AI の活用に関する判断例:対象とする AI に対してリスクに影響をもたらす対策を行う

(3) AI のリスクをより深く理解するために、更なる分析に着手する

- 判断に使用できないレベルの分析で無理に分析を行わず、分析のレベルを向上させる
 - ◇ AI の活用に関する判断例:対象とする AI に関するリスク分析をさらに詳細に実施する

(4) 既存のリスク対応策を維持する

- リスク分析の結果把握されたリスクを保有する
 - ◇ AI の活用に関する判断例:現状の AI に対する方針を維持する。

(5) 目的を再考する

- リスク分析の結果が重要なリスク基準を満たさない場合、または有効なリスク対応を実施したとしても重要なリスク基準を満たさない場合等に、組織・個人の活動の目的そのものを見直す
 - ◇ AI の活用に関する判断例:対象とする AI の活用の前提とする組織・個人の目的を再考する。

7. 今後の科学技術社会における AI の活用に向けて

AI は、個人、組織、社会を大きく変えるポテンシャルを有しており、その活用は、好ましい影響、好ましくない影響を共にもたらす可能性がある。今後の科学技術社会において、AI の活用は進むと想定されるが、そのリスクの大きさを十分に認識した上で活用していく必要がある。

AI のリスクに関して、本ガイドラインが AI 開発者や AI 活用者を含む、多様な目的に向かって AI を活用しようとする人々にとって、そのリスクを事前に把握し、適切な対応を取るための一助として活用されれば幸いである。

以上

AI のリスクアセスメントガイドライン 第1版

初版 2026 年 7 月

■著作権者

特定非営利活動法人 リスク共生社会推進センター

〒240-8501 神奈川県横浜市保土ヶ谷区常盤台 79-5

横浜国立大学 総合学術高等研究院 次世代工学システムの安全科学研究ユニット内

特定非営利活動法人 リスク共生社会推進センター 事務局

■利用に関する注意

本書の複製・転載、および記載内容に関するご意見・ご要望は、

特定非営利活動法人 リスク共生社会推進センターまでお問い合わせください。

<問い合わせ先> Email: risk.npo@riskkyosei.org