資料１－２

新ＡＩ事業者ガイドライン　スケルトン（案）

※本資料は、ガイドラインに関する議論・検討過程の透明性を高めるために、本資料の公開時点において議論・検討中であるガイドラインの項目立て及び記載内容案の概要を示すものであり、今後の議論を踏まえて変更される可能性がある。

はじめに

- ✧ **本ガイドラインの位置づけ（目的と対象、ガイドラインの見方など）**
  - ・ 目的：我々が共有する民主的価値に沿った、信頼できる AI の構築
- ✧ **プレイヤーの区分け・分類、関係図、パターン、例などの記載**
- ✧ **リスクベースアプローチ**
  - ・ 「規律の程度をリスクの大きさに対応させるべき」という基本的な考え方を採用

第１部　AI とは

- ✧ **関連用語の定義**
- ✧ **事例集（リスクと便益、多様な事例を掲載）**

第２部　AI 開発から運用・利用にあたってのガバナンス

※第3部以下の主体にかかわらず関係する項目。

- ✧ **アジャイルガバナンス原則**
  - ・ 一律の事前規制ではなく、民間の知見を活用しながら機動的で柔軟な改善を可能とするガバナンスのあり方を志向
  - ・ G7デジタル・技術大臣会合において合意した、アジャイルガバナンス5原則（法の支配、適正手続き、民主主義、人権尊重、イノベーションの機会の活用）
- ✧ **AI 原則**
  - ・ 生成 AI の普及等をふまえ既存の AI 原則のアップデートの可能性も検討すべきではないか
- ✧ **AI 事業者に共通する事項**
  - ・ 各主体は、法の支配、人権、民主主義、多様性、公平公正な社会、人間中心を尊重するよう AI システムを設計・利用する責務がある。具体的には、以下のような責務を負うべきではないか
    - ■ 各国の法制度を遵守する義務があり、法制度に違反した場合には処罰等の対象となる。
    - ■ 人権侵害、テロや犯罪等を目的とする、あるいは、助長する蓋然性の高い不適切な入出力を行う AI を提供又は利用してはならない。
    - ■ AI の不適切な入出力の抑止に資する技術の開発・導入に努めなければならない。

■ AI に関する多様なリスクを認識し、計画を策定し、行動をとらなければならない。一定規模以上の組織においては、AI ガバナンスポリシー及び体制を構築することが望ましい。

✧ **AI 原則にもとづく体制構築のあり方**
- 経営層のリーダーシップの下での AI リスクの分析、ゴール設定、システムデザイン、運用、評価及びリスク分析のアップデートを含む組織ガバナンスを求めるべきではないか

✧ **AI 事業者の行動目標、実践例**

✧ **AI ガバナンス・ゴールとの乖離評価例**
- AI 事業者が実施すべき行動目標を提示するとともに、それぞれの行動目標に対応する仮想的な実践例や AI ガバナンス・ゴールとの乖離を評価するための実務的な対応例（乖離評価例）を例示

✧ **サプライチェーンを念頭に置いたリスク管理・ガバナンスの維持**
- 複数プレイヤーにまたがる論点、プレイヤー間で問題になり得る点について、サプライチェーン／リスクチェーンの観点における連携確保のあり方の検討
- AI 開発からサービス実施にわたるサプライチェーン・リスクチェーンが複数国にまたがることが想定される場合、データの越境移転における適切なガバナンスの検討（Data Free Flow with Trust）に留意し、民間の知見や取り組みを活用しながら、リスクチェーンの明確化とチェーンの段階ごとに適したリスク管理・ガバナンスのあり方について検討

※第3部以下については、一つの事業者が複数の主体にまたがって該当する場合もありうる。

第3部　AI のアルゴリズム開発者向け

✧ **透明性確保・説明可能性のあり方**
- AI システムの目的、機能、予想される効果やリスクについての情報を開示するか、または利用者に説明すべきではないか

✧ **透明性確保の手法**
- 外部監査のあり方の検討
- システムの動作についての利用者への情報提供の検討

＜主な検討事項＞

情報開示等の要否・あり方。透明性確保の目的（「何のために」）・対象（「何を」）・客体（「誰に向けて」）を踏まえた、透明性確保の内容（要求される基準の検討（実務的・技術的要素、国際法や外国法の法規制に照らした実現可能性を考慮）等）。外部監査の要否・あり方（検証可能性確保のために要求される基準の検討（実現可能性を考慮）等）。第3部と第4部の主体について区分けする要否。第2部に記載された AI 事業者に共通する事項に加え、第3部特有の責務について議論。

第4部　AI の学習実施者向け

- ✧ **制御可能性の確保のための学習データ公開の要否・あり方**
- ✧ **学習データの収集ルール及び除外ルール公開の要否・あり方**
- ✧ **その他の学習データの検証可能性確保の手法**
  - ・　外部監査のあり方の検討（あらゆるフェーズにおいて検討が必要な論点）
  - ・　利用者からの問い合わせに迅速に対応できる仕組みの検討
- ✧ **適正な学習データ利用についてのあり方**

  ※用途に応じたルールの検討

＜主な検討事項＞

情報開示等の要否・あり方。情報開示の目的（「何のために」）・対象（「何を」）・客体（「誰に向けて」）を踏まえた、情報開示の内容。外部監査の要否・あり方（検証可能性確保のために要求される基準の検討（実現可能性を考慮）等）。第2部に記載された AI 事業者に共通する事項に加え、第4部特有の責務について議論。

第5部 AI システム・サービス実装者向け

- ✧ **AI を組み込んだシステム・サービスの安全性等の担保のあり方**
  - ・　安全性は、「セーフティ」（人の身体・生命の保護、未成年者の保護、詐欺的利用の防止等）の意味合いだけでなく、「セキュリティ」（機密保護、情報管理等）も含むべきではないか
- ✧ **AI をクラウドにより（SaaS 形式で）提供する場合の留意点**

＜主な検討事項＞

第5部と第6部についてプレイヤー区分けの要否は今後の論点。第2部に記載された AI 事業者に共通する事項に加え、第5部特有の責務について議論。

第6部　AI を活用したサービス実施者向け

- ✧ **公正・説明責任・透明性の確保**
- ✧ **AI アプリケーションの悪用対策等の責任分担のあり方**
- ✧ **ユーザー情報の管理方法**
- ✧ **入力情報（プロンプト）及び出力情報の管理方法、その権利関係、フィルタポリシ**
  - ・　権利関係は、「AI ・データの利用に関する契約ガイドライン AI 編 」を参照
  - ・　入出力データや利用者情報の取扱いを含むセキュリティポリシーを策定・開示することが必要ではないか。個人情報を扱う場合には、プライバシーポリシーも策定・開示することが必要ではないか。

✧ **許容しない利用方法などの宣言及びその実効性の担保**

- ディープフェイクや迷惑メール生成などの不適切な利用や許容しない利用を宣言することが必要ではないか。また、その宣言の実効性を担保する方法の検討が必要ではないか。
- AI の不適切な挙動、不適切な利用方法等に関する情報を収集・共有し、システム・サービスの改善に努めることが必要ではないか。

✧ **AI を活用しているサービスであることの明示の要否・あり方**

✧ **業務で AI を利用する者等との連携方法（B2B2C やサプライチェーンを念頭においたリスク管理）**

✧ **制御可能性、予測可能性に関して求めるレベル感の明確化**

- 特に身体・生命の安全性に影響を与える場合や、雇用における利用の場合等のように、社会的な要保護法益性が大きい場合には、高度の制御可能性と予見可能性が求められるのではないか

＜主な検討事項＞

AI 利用者に対して、提供する AI の概要やリスク等の情報を提供することや、不適切な使用の抑制を検討。「AIを活用している」旨の明示の要否や方法の検討。リスクや分野等に応じて制御可能性、予測可能性をどのようにどこまで求めるか等の検討。第2部に記載された AI 事業者に共通する事項に加え、第6部特有の責務について議論。

第7部　業務で AI を利用する者向け

✧ **利用する業務と期待する効果とリスクの整理**

- AI の利用に関するリスク（組織内の職員等による不適切な使用も含む）を認識し、リスクを抑止するための工夫を講じ、また、問題の発生を把握し、適切に対処することが求められるのではないか
  - ■ 外部サービスに情報入力し、生成結果を受け入れることのリスク
    - ➢ 生成 AI により生成された結果に虚偽が含まれる可能性があるため、このような限界を認識し、根拠や裏付けを確認するようにすることが求められるのではないか
  - ■ 外部サービス利用により機密情報が公開されてしまうリスク
  - ■ レピュテーションリスク、顧客への説明責任、再発防止策の為の情報確保、紛争解決の為の証拠の保存

✧ **多様なリスクの洗い出し、リスクを現実化する脅威の洗い出し、リスクが現実となる可能性を低減する方法**（いわゆるリスクアセスメント）

✧ **政府の AI 利用に関する注意点についても記載を検討**

＜主な検討事項＞
記載するリスク例の検討。AI の入出力等の検証可能性を確保するため、入出力等のログの記録・保存についての検討。業務で AI を利用する者の具体的な対象や範囲等の検討。第2部に記載された AI 事業者に共通する事項に加え、第7部特有の責務について議論。

別紙

- ✧ **契約上の留意事項**
  - ・ AI・データ契約ガイドライン参照
- ✧ **チェックリスト**

＜主な検討事項＞
ガイドラインの利用者にとって、ガイドライン作成・更新時点の水準として参考となるよう、チェックリストの形式、内容等について検討。