

AIの利活用に関する検討

- 資料4-1 AIの利活用に関する検討の背景・問題意識(案)
- 資料4-2 AIネットワーク化の進展に伴い形成されるエコシステムの展望(案)
- 資料4-3 AIの利活用に関するケーススタディ(案)
- 資料4-4 AIネットワーク化の進展において想定される課題(案)～ネットワーク化の観点から～
- 資料4-5 AIの利活用に関して利用者等が留意することが期待される事項についての論点(案)
- 資料4-6 AIの利活用及びAIネットワーク化に係る諸課題(案)

平成30年2月21日
事務局

A I の利活用に関する検討の背景・問題意識（案）

平成30年2月21日
事 務 局

AIの利活用に関する検討の背景・問題意識

AIは人間の能力を向上・補完する可能性 ⇒ AIネットワーク化の進展の過程において、個人や社会が抱える様々な課題（人口減少・少子高齢化、労働力不足等）の解決への貢献など人間社会への多大な便益

AIの開発及び利活用の加速化、AIネットワーク化の健全な進展が重要（これらを通じて人間中心の「**智連社会**」を実現）

他方で、AIに対する理解や信頼の不足、AIの判断のブラックボックス化などのリスク、それらによる人々のAIに対する不安などが、阻害要因となる可能性

人々のAIへの理解・信頼の醸成、AIに関する便益の増進、リスクの抑制等などの取組が必要

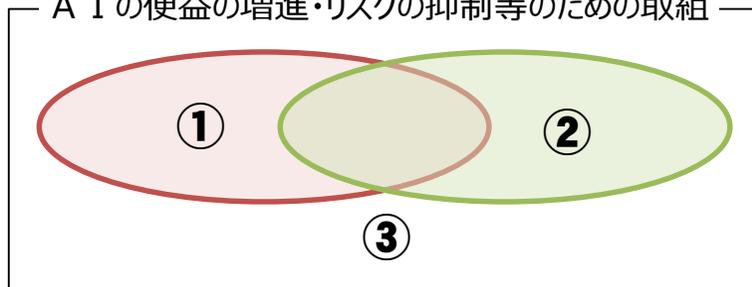
【具体的な取組】

- ・ AIの開発者が留意することが期待される事項の整理 → 『国際的な議論のためのAI開発ガイドライン案』
- ・ AIの利用者・データ提供者（利用者等）が留意することが期待される事項の整理
- ・ 開発者・利用者等を含む関係するステークホルダ（政府、業界団体等）がAIの利活用等の促進に向けた環境整備のために取り組むことが期待される事項の整理

※ これらの検討・整理を行うことにより、AIに関する国内外の議論に貢献することができる。

AIにおける増進すべき便益、抑制すべきリスク等について、開発者・利用者等がそれぞれ取組を行うとともに、それらを含む関係するステークホルダが環境整備等の取組を行うことにより、AIの開発及び利活用の加速、AIネットワーク化の健全な進展が図られる。

AIの便益の増進・リスクの抑制等のための取組



- ① 開発者が留意することが期待される事項
→ 『国際的な議論のためのAI開発ガイドライン案』
- ② 利用者等が留意することが期待される事項
→ [資料4-5](#)
- ③ 関係するステークホルダが取り組むことが期待される事項
→ [資料4-6](#)

AIの利活用に関する検討に当たっては、その前提としてAIの利活用が社会・経済に与える影響、AIネットワーク化が進展した際に形成するエコシステム等に関する検討を行うことが必要。→ [資料4-2](#)、[資料4-3](#)、[資料4-4](#)

AIの利活用に関する検討の背景・問題意識



本推進会議における検討

- AIネットワーク化が進展した際に形成するエコシステムの展望
- AIの利活用が社会・経済に与える影響
- AIの利用者等が留意することが期待される事項
- 開発者・利用者等を含む関係ステークホルダが取り組むことが期待される事項

参 考

『報告書2017』 第4章 今後の課題

- AIネットワーク化の健全な進展を通じたAIシステムの便益の増進及びリスクの抑制については、ネットワークに接続され得るAIシステムの開発を対象とするAI開発ガイドライン（仮称）を国際的に共有すること等を通じた開発のガバナンスと併せて、利活用のガバナンスとして、ネットワークに接続されたAIシステム又はAIネットワークサービスの利活用に当たり、**利用者**（最終利用者たる個人又は団体のほか、他者が開発したAIシステムを用いてAIネットワークサービスを第三者に提供するプロバイダを含む。）がそれぞれの属性等に応じて留意することが期待される事項について国際的な共有を図ることが必要であるものと考えられる。

そこで、**本推進会議においては**、今後、AI開発ガイドライン（仮称）の策定に向けた国際的な議論のフォローアップと併せて、ネットワークに接続されたAIシステム又はAIネットワークサービスの**利活用に関し利用者がそれぞれの属性等に応じて留意することが期待される事項**（中略）に関し、国際的な議論の動向も踏まえながら**多角的に検討を進めることが必要**である。

なお、（中略）利活用のガバナンスの在り方に関する検討に当たっては、その**前提として、AIネットワーク化が社会・経済にもたらすインパクト及びリスクに関する問題の所在を把握することに加えて、AIネットワーク化の進展を通じて形成される関係者間のエコシステムに関する展望を得ることが必要**であるものと考えられる。そこで、本推進会議においては、「AI利活用ガイドライン（仮称）」の策定等に関する検討に向け、前章で紹介した分野別評価及びエコシステムの展望の検討並びにこれらを通じた論点の整理を進めていくことが課題となる。

A I ネットワーク化の進展に伴い形成される エコシステムの展望（案）

平成30年2月21日
事 務 局

AIネットワーク化の進展に伴い形成されるエコシステムの展望に関する検討について

これまでの御意見・御指摘を踏まえ、

- ① AIの利活用に着目し、主として利用者の視点から、シナリオベースでAIの利用シーンを展望する。
AIの利用シーンを展望するに当たっては、次のように利用シーンを分類することとする。

＜AIの利用シーンの分類＞

エンドユーザー	個人の利用者	宅外	都市部における利用	地方部における利用
		宅内	家庭内における利用	
	利用者としての企業	企業における利用		

- ② 上記①の利用シーンをもとに、いくつかの事例に関するケーススタディを行い、AIの利活用による便益及びリスクを抽出する。
- ③ 上記①の利用シーンを展望する中で、主としてAIネットワーク化の健全な進展を阻害し得る要因（ネットワーク化により生ずる（増幅する）リスク）を考察する。
- ④ 次の段階として、上記①の利用シーンを念頭に、AIの利活用に当たってのデータの流れやカネの流れ（データエコシステムやビジネスエコシステム）、あるいは、人間とAIシステムとの関係、AIシステム間との関係、AIシステムと他のシステムとの関係等について考察することとする。

 当面は、①～③に関する検討を先行して進めることとする（①～③については、夏頃に取りまとめる予定の『報告書2018』（仮称）に盛り込む予定）。

（注）当初、AIの利活用について考慮することが期待される事項に関する論点整理のために行うとしていた分野別評価の既存3分野（まちづくり、健康、モノ）の再評価については、上記②及び③と重複するため、上記②及び③と一体的に取り組むこととし、それ単独の取組としては行わないこととする。

地方部におけるAIの利用シーン（例）

移動

- 自動運転の実現により、高齢者等が病院や買物などに行く交通手段が確保されるほか、路線バスなどの交通網の維持（廃止回避）が可能となる。



路線バスの自動運転化により、運転手不足等の問題が解消

AI



自動運転車の実現により、高齢者等の外出が容易に可能（移動手段の確保）

AI



AI

自動運転タクシーによる快適な観光

AI



AIスピーカーや執事ロボットの活用



AI

専門医がいない地域における画像診断等を用いた遠隔医療

AIシステム間の調整による緊急搬送における専門医とのマッチング、最適ルートの設定

AI

医療

- 専門医がいない地域でも、遠隔医療による診断等が可能となるほか、薬剤師が少ない地域では、調剤の自動化により、近所で薬の入手が可能となる。



AI



AI



調剤の自動化

AI



各地域社会の状況・特性に応じたイベントの企画・参加勧奨

AI



趣味・嗜好に応じた地域内、都市部や海外の人々とのマッチング、交流



地域の特性に応じたまちづくり（災害に強いまち、高齢者に優しいまち等）を提案

AI

仕事

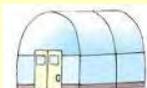
- 農業や漁業等における作業の自動化や効率化・高度化により、従事者の人手不足や高齢化に対処することができる。



自走式トラクターやドローンによる耕耘、種まき等の作業の自動化

AI

作物の生育状況のモニタリングや収穫量の予測



AI

天候や海水温等に応じた魚群探知の高精度化

AI



コミュニティ

- 地域の特性に応じた街づくりが可能となるほか、地域内外との交流が活発となり、コミュニティの活性化に貢献することができる。

（注）想定される利活用のうち、いくつかの例を記載

AI

：既に実用化されているもの
近い将来実現しそうなもの

AI

：中期的に実現しそうなもの

AI

：遠い将来に実現しそうなもの

家庭内におけるAIの利用シーン（例）

医療

- 健康情報や生活情報等から病気発症の予測、生活改善・疾病予防の提案を行うほか、自宅に居ながら遠隔で医師の診断を受けることができる。



生活改善、疾病予防の提案
遠隔での診断

AI

健康情報や生活情報等を活用した健康状態の推定、病気発症の予測



AI



AI

体内のAIチップによる病気（異常）の早期発見

AI



AI

AIスピーカーや執事ロボットの活用

家事

- 料理や掃除が自動化されるほか、買物も自動化（執事ロボットによる自動注文、ドローンによる自宅への自動配送）することができる。



AI



生活者の嗜好に基づくレシピの提案
料理の自動化



AI

食材や日用品などの在庫状況に応じた自動注文、自動配送



AI

自動清掃

嗜好に応じた娯楽コンテンツの提供（脳チップによる直接的な疑似体験）



AI

人の在・不在、居場所に応じたエアコンの運転や温度・湿度の自動調整

AI



財産、性格等に応じた資産運用の提案、家計簿情報等に基づく節約術の提案



AI

適性、能力等に応じた進路（就職先）の提案、マッチング



AI

性格等に応じた結婚相手像の提案、マッチング

AI



安全・快適な居住環境

- 遠方の高齢者や外出時の子供の見守りが可能となるほか、仮想現実（VR）を活用した高度な疑似体験が可能となる。

豊かな人生

- 結婚、就職など人生の転機となる出来事において最適な提案が可能となるほか、資産運用や節約術の提案が可能となる。

（注）想定される利活用のうち、いくつかの例を記載

AI

：既に実用化されているもの
近い将来実現しそうなもの

AI

：中期的に実現しそうなもの

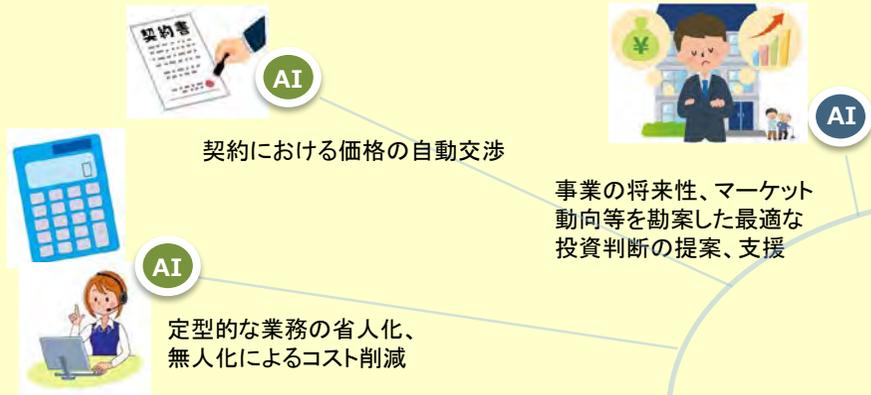
AI

：遠い将来に実現しそうなもの

企業におけるAIの利用シーン（例）

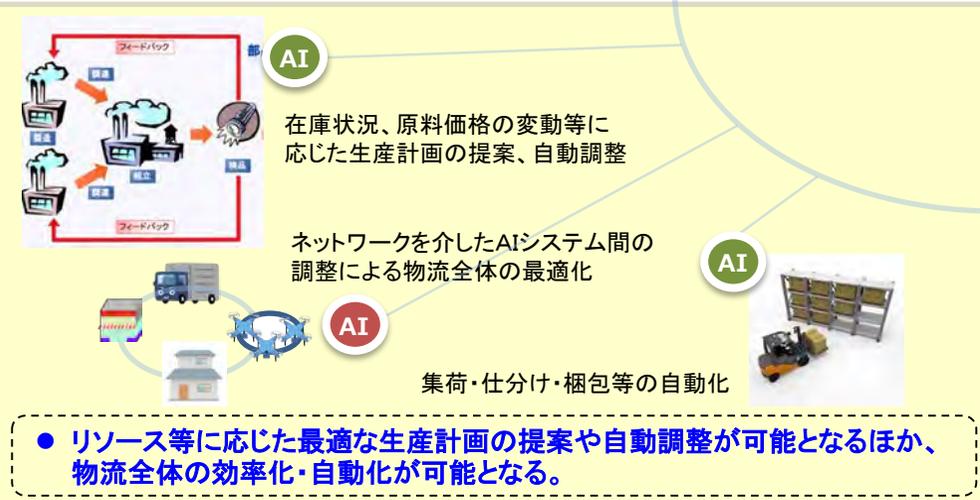
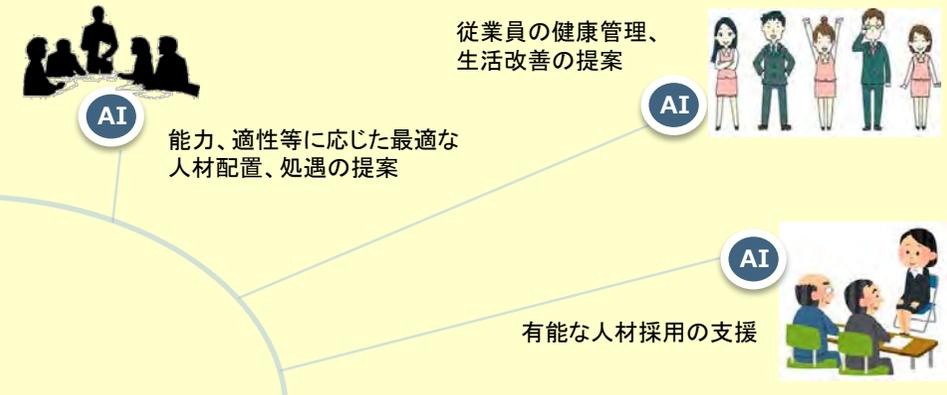
カネ

- 定型的な業務のコストが削減できるほか、成長分野への積極的な投資などが可能となる。



ヒト

- 最適な人材配置や採用（外部人材の登用を含む。）等が可能となるほか、従業員のモチベーションの向上等が期待できる。



モノ

情報

AI : 既に実用化されているもの
 AI : 近い将来実現しそうなもの
 AI : 中期的に実現しそうなもの
 AI : 遠い将来に実現しそうなもの

（注）想定される利活用のうち、いくつかの例を記載

AIの利活用に関するケーススタディ（案）

平成30年2月21日
事務局

ケース①：完全自動運転車

想定される便益（例）

- 人間は運転する必要がなく、自動車での移動において、移動時間を有効に活用することができるようになる。
- 高齢者や障害者の方にとって、手軽に移動することができる手段が確保されることとなり、病院や買物などに容易に出かけることができるようになる。
- 深夜や早朝などにおける長距離トラックや長距離バスの運転をする必要がなく、働き方やワーク・ライフ・バランスを見直すことなどができるようになる。
- 特に地方部などにおける路線バスの運転手不足などの問題を改善することができ、路線の廃止・縮小を回避することができるようになる。

想定される課題（例）

- 自動運転車間で交渉・調整が成立しない場合、車線変更や合流等が適切に行うことができないおそれがある。
【類似ケース】 サプライチェーン全体で最適化の調整している場合、各社のAIシステム間で適切な交渉・調整が成立しないと最適化が実現できないおそれがある。
- AIシステムがハッキング等された場合、そのAIシステムを実装している自動運転車が正常に機能しなくなるだけでなく、ネットワークを介して、次々と他の自動運転車にも影響が及び、事故や交通の混乱が生ずるおそれがある。
【類似ケース】 食材や日用品などの自動注文AIシステムがハッキング等された場合、注文の量が過剰又は過少に変更されたり、架空の注文や不正な注文などが行われたりするおそれがある。
- 長期間AIソフトのアップデート等をしなかった場合や学習したことがない(想定できない)場面に遭遇した場合、自動運転車が周囲の状況を正しく認識することができず、障害物等を誤って認識してしまい、事故を起こすおそれがある。
【類似ケース】 AIシステムの点検・修理を怠っていたため、ドローン間の連携が適切に行うことができず、衝突したり、墜落したりして、利用者や第三者に危害を及ぼすおそれがある。
- 自動運転車が事故を起こしてしまった場合、AIがどのように判断をしたのか検証できないと、責任の所在を明らかにできず、適切な責任分配がなされないおそれがある。
【類似ケース】 緊急救命で救急車と医療機関等の中でリソース配分等の調整を行うケースにおいて、順番が劣後され症状が悪化した場合、AIの判断の検証、責任の所在、被害者救済が問題となるおそれがある。

ケース②：医療診断

想定される便益（例）

- 画像診断により、病気の早期発見、見落としの改善につながるとともに、医師の負担が軽減されるようになる。
- 遠隔診断により、専門医がいない地域（特に地方部）でも適切な診断を受けることができるとともに、医師不足・偏在などの問題の解決に貢献することができるようになる。
- 遠隔診断により、自宅でかかりつけ医や専門医などの診断・診療を受けることができるようになる。
- 体内に埋めたAIチップが身体に関する情報を集め、分析することにより、異常や病気を早期に発見することができるようになる。

想定される課題（例）

- AIの診断の精度、人々のAIに対する信頼等を踏まえると、診断をAIの判断のみに委ねる場合、患者の理解が得られないおそれがある。
【類似ケース】 人事評価や採用の可否をAIの判断のみに委ねる場合、従業員や求職者の理解が得られないおそれがある。
- AIが誤診した結果、適切な治療が行われずに患者の症状が悪化してしまった場合、なぜ誤診したのか、AIがどのような判断をしたのか説明ができないと、患者や家族などの理解が得られないおそれがある。
【類似ケース】 AIの判断をもとに、融資（住宅ローン等）を受けることができなかった場合、AIがなぜ融資を認めなかったのか説明ができないと、申込者の理解が得られないおそれがある。
- AIシステムがハッキング等され、患者の医療データが流出した場合、ネットワークを介して次々と他のAIシステムにデータが共有されるとともに、様々なデータと統合されることにより、個人が特定され、重大なプライバシー侵害が生ずるおそれがある。
【類似ケース】 結婚相手をマッチングするAIシステムがハッキング等された場合、プロフィール（パーソナルデータ）が、AIシステム間の連携により、本人の同意なく自動的にパーソナルデータが流通し、プライバシーが侵害されたり、悪用されたりするおそれがある。
- 体内にAIチップを埋め込む利活用については、生命倫理の観点等から慎重な検討が求められるおそれがある。
【類似ケース】 脳にAIチップを埋めて脳に直接働きかけるような利活用や体内にAIチップを埋めて人間の運動能力を増幅させるような利活用については、生命倫理等の観点等から慎重な検討が求められるおそれがある。

ケース③：人事評価・採用

想定される便益（例）

- AIを活用したマッチングにより、求職者は適性や能力等に応じた就職先を選ぶことができ、採用側は求めるニーズに合致した有能な人材を登用することができるようになる。
- 採用プロセスにおけるAIの利活用により、統一的な基準で評価を行うことができる(評価のバラツキを抑えることができる。)ようになる。
- 採用側の選考作業の負担を大幅に軽減・効率化することができ、人間は他の業務に注力することができるようになる。
- 従業員の適性や能力、希望等に応じた最適な人材配置を提案することができるようになる。
- AIが従業員の業績等を評価することにより、評価のバラツキを抑えることができ、また、評価に応じた細やかな処遇を提案することができるようになる。

想定される課題（例）

- 人事評価や採用の可否をAIの判断のみに委ねる場合、従業員や求職者の理解が得られないおそれがある。(再掲)
【類似ケース】 融資(住宅ローン等)の可否等をAIの判断のみに委ねる場合、申込者の理解が得られないおそれがある。
- 従業員によって恣意的に選択されたデータや不正に操作されたデータを学習した場合、当人が高い評価を得てしまうなど不公平・不公正な判断がなされるおそれがある。
【類似ケース】 介護のケアプラン作成において、介護報酬が高くなるようなデータを多く学習した場合、要介護者にとって最適なプランが、提案されず、事業者にとって有利なプランが提案されるおそれがある。
- カメラ映像や各種センサー等から得られる働きぶりなどに関するデータを用いてAIを利活用したモニタリングを行う場合、労働強化につながったり、従業員が必要以上のプレッシャーを感じて健康を損なったりするおそれがある。
【類似ケース】 犯罪捜査や裁判において、SNS等の情報を用いてAIを利活用する場合、思想や信条等を理由として不利益な扱いを受けるおそれがある。
- 人事評価や採用において、AIが外部の情報(SNS等)も活用して評価する場合、SNS等の情報が使われることを見越して情報を発信することが考えられ、適切な評価を行うことができないおそれがある。
【類似ケース】 結婚相手のマッチングにおいて、AIがSNS等の情報を活用する場合、嘘や誇張された情報が混ざっていることを見抜けず、適切なマッチングが実現できないおそれがある。

ケース④：保険（金融）

想定される便益（例）

- 健康状態や自動車の運転の巧拙（急発進、急ブレーキの多寡など）等に応じて、保険料率等が異なる商品（生命保険、自動車保険）を提供することができるようになる。
- ゲノム情報等に応じて、各人にカスタマイズした商品（生命保険）を提供することができるようになる。
- 自動車事故が起きた際に、現場の写真・画像から瞬時に保険金を算出することができ、保険金の支払いまでの期間を短縮することができるようになる。
- 保険金の支払いに際して、帳票や診断書等を自動で読み込み、査定等に必要な手続きを行うことにより、事務を効率化することができるようになる。

想定される課題（例）

- 当人の知らないところで健康や自動車の運転に関するデータが取得されたり、同意していない目的のために活用されたりして、プライバシーが侵害されるおそれがある。
【類似ケース】 当人が気づかずに防犯カメラ等により画像が撮影（データが取得）され、本人の同意なしに様々な分析（ビッグデータ解析）に活用されたり、プロファイリングに使われたりして、プライバシーが侵害されるおそれがある。
- 健康状態や自動車の運転の巧拙等に応じた新しい商品について、AIがどのような根拠に基づいて保険料を算出したのか説明できないと、適正な保険料であるかを判断することができず、新しい商品を提供することができないおそれがある。
【類似ケース】 店舗やイベントの来客数等をAIが予測するサービスについて、AIがどのような根拠に基づいて予測したのか説明できないと、当該サービスの利用者の信頼を得ることができず、利用してもらえないおそれがある。
- ゲノム情報を解析されることにより、現在は健康であっても将来的に病気を発症する可能性が高いと判断されると、保険に加入することができなくなるおそれがある。
【類似ケース】 顔の画像や声を解析されることにより、外見ではわからない精神的な疾患等に該当する可能性が高いと判断されると、就職や転職などで不利な扱いを受けるおそれがある。
- 適切なデータで学習していなかったり、学習するデータ量が十分でなく、画像認識や文字認識の精度が低い場合、適正な保険金を算出できなかったり、適切に査定できなかったりするおそれがある。
【類似ケース】 適切なデータで学習していない等の理由により、災害予測の精度が十分でない場合、警戒情報の発信が行われなかったり、遅れたりするおそれがある。

ケース⑤：行政

想定される便益（例）

- 給付金や補助金等の申請に関し、申請書に不備等がないか確認するとともに、支給の条件を満たしているか必要な情報と照合し、支給の可否を判断することにより、手続きの期間の短縮、業務の効率化に貢献することができるようになる。
- 年金や児童手当などに関し、対象者の生活情報等と照合（生存の確認等）することにより、不正受給を防止することができるようになる。
- 道路や橋などのインフラの故障予測や異常検知に応じて修理ロボットが自動で修理を行うことができるようになり、安心して安全に使うことができるとともに、人間では作業が困難な場所でも安全に修理を行うことができるようになる。
- 災害対応において、災害予測（二次災害を含む。）を行い、警戒情報を発信するとともに、被害状況や地域の事情等に応じた復旧計画を提案するなど、被害の最小化、早期の復旧を支援することができるようになる。

想定される課題（例）

- 給付金や補助金等の申請に関し、AIによって支給が認められないと判断された場合、AIがなぜそのような判断をしたのか説明ができないと、対象者の理解が得られないおそれがある。
【類似ケース】 AIの判断をもとに、融資（住宅ローン等）を受けることができなかった場合、AIがなぜ融資を認めなかったのか説明ができないと、申込者の理解が得られないおそれがある。（再掲）
- 年金や児童手当、失業手当などの受給資格の確認を行う場合、生存等の確認のために生活情報等（電気や水道の使用状況等）を用いることの同意を得ることができず、このような利活用ができないおそれがある。
【類似ケース】 地域全体で電力やガス等の需給バランスの最適化に向けた調整を行う場合、情報提供に同意を得ることができない住民がいると、全体としての最適化が実現できないおそれがある。
- 長期間AIソフトのアップデート等をしなかったため、道路や橋の異常を正しく検知することができず、異常を見逃してしまい、事故（道路の陥没、橋の崩落等）が発生するおそれがある。
【類似ケース】 長期間AIソフトのアップデート等をしなかった場合や学習したことがない（想定できない）場面に遭遇した場合、自動運転車が周囲の状況を正しく認識することができず、障害物等を誤って認識してしまい、事故を起こすおそれがある。（再掲）
- 適切なデータで学習していない等の理由により、災害予測の精度が十分でない場合、警戒情報の発信が行われなかったり、遅れたりするおそれがある。（再掲）
【類似ケース】 ウェアラブル端末からの健康情報や生活情報等を用いた健康状態の推定において、適切なデータで学習していない等の理由により、その精度が十分でない場合、病気の前兆を見逃してしまうおそれがある。

A I ネットワーク化の進展において想定される課題（案）

～ネットワーク化の観点から～

平成30年2月21日
事 務 局

AIネットワーク化の進展において想定される課題（ネットワーク化の観点から）

- ◆ AIネットワーク化の特徴として、「様々な事業者間における多種多様なAIの接続(複雑な連携)」、「様々な事業者の多種多様なAIを取りまとめる基盤的なAIの構築(情報や権限の集中)」、「様々な事業者間における多種多様なAIによる情報の連携(情報の共有・拡散)」が挙げられる。

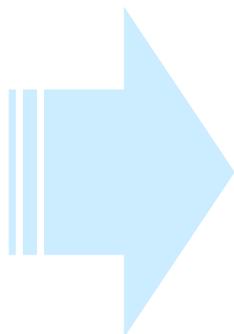
このような特徴を活かして、新しいサービスの開発・提供、最適化や効率化、コスト削減などが可能となる一方で、ネットワークという観点に着目した場合、次のような課題(AIネットワーク化の健全な進展を阻害し得る要因)に留意すべきである。

AIネットワーク化の特徴

様々な事業者間における多種多様なAIの接続(複雑な連携)

様々な事業者の多種多様なAIを取りまとめる基盤的なAIの構築(情報や権限の集中)

様々な事業者間における多種多様なAIによる情報の連携(情報の共有・拡散)



ネットワーク化の観点から 想定される課題(例)

個別の事業者のトラブル等がシステム全体に波及するおそれ

AIシステム間の連携・調整が成立しないなどのおそれ

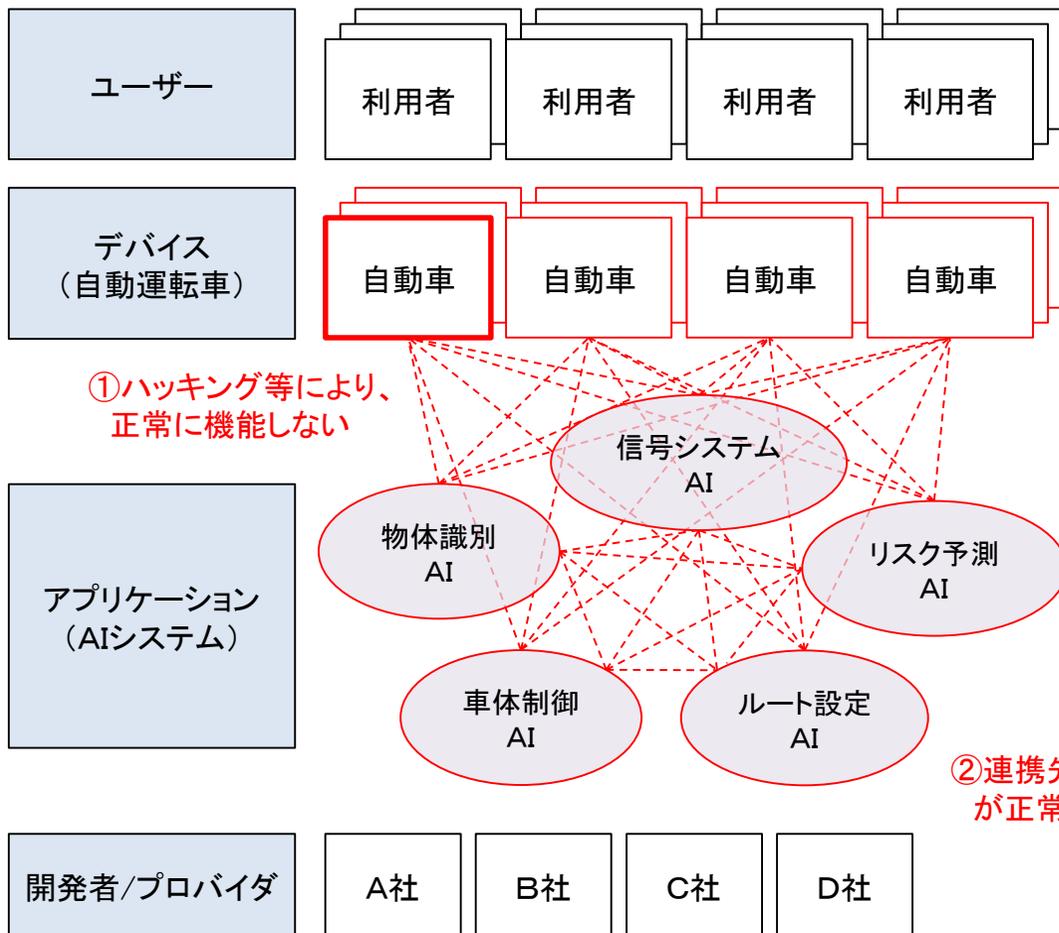
AIの判断・意思決定を検証できないおそれ
(ネットワーク化により、システム全体としてブラックボックス化するおそれ)

少数のAIの影響力が強くなりすぎるおそれ
(少数のAIの判断によって企業や個人が不利な立場になるなどのおそれ)

領域横断での情報の共有と特定の基盤的なAIへの情報の集中によるプライバシー侵害のおそれ

個別の事業者のトラブル等がシステム全体に波及するおそれ

◆ AIシステムが相互に連携して、協調・調整することになるため、例えば、ある事業者において、ハッキングやシステム障害等により一部のAIシステムが正常に機能しなくなった場合や災害等により通信が遮断された場合、あるいは、ある一部のAIシステムの判断・予測等に誤りが生じた場合や虚偽又は不適切なデータが流通した場合などに、他のAIシステムに次々に影響が波及し、ネットワーク全体が正常に機能しなくなったり、期待された効果（適切な判断・予測、マッチング等）が得られなくなったりするおそれがある。

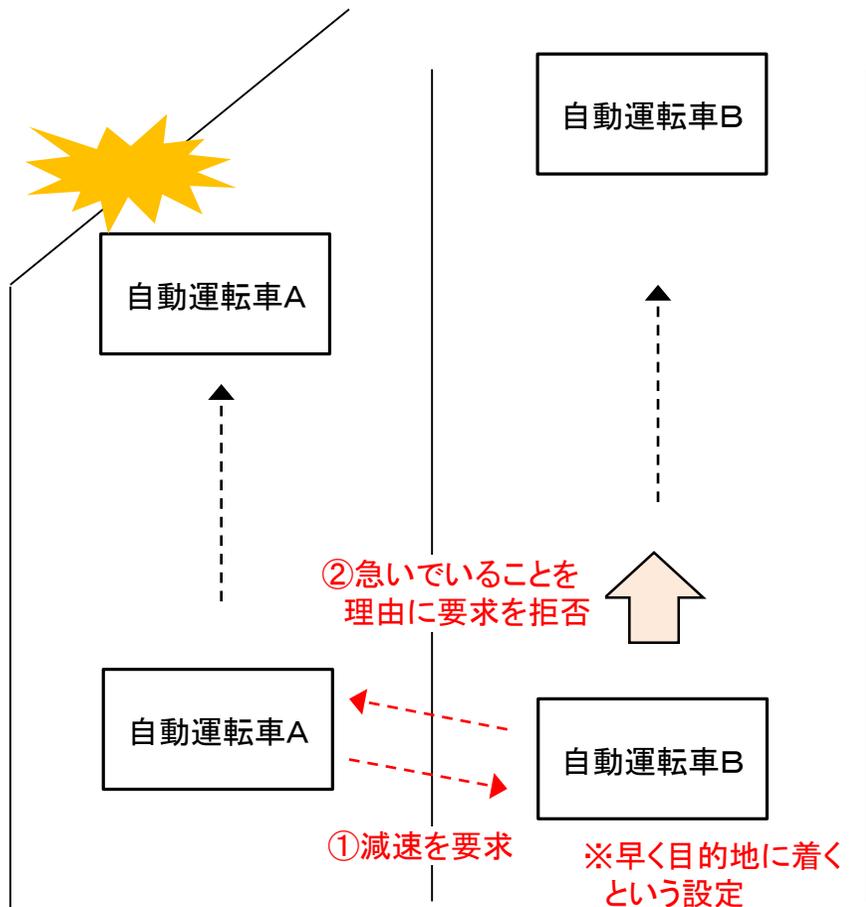


【個別の事業者のトラブル等がシステム全体に波及するおそれ（例）】

- セキュリティ対策等が不十分な事業者が提供しているAIシステムを活用した自動運転車がハッキングされた場合、連携している別のAIシステムが誤作動又は機能不全に陥り、交通事故や交通障害等が発生するおそれがある。
- サプライチェーン全体として最適化を調整している場合、ある事業者のAIシステムが生産量等を誤って予測した場合や虚偽のデータを流通させた場合、全体としての最適化が実現できない（在庫不足／過剰在庫、原材料不足等）おそれがある。
- 被災地の緊急救命において、救急車や医療機関等の中でリソース配分の最適化の調整を行う場合、ある医療機関のAIシステムが通信の遮断により機能不全に陥ると、適切な情報共有、リソース配分が実現されず、適切な救命活動ができなくなるおそれがある。

AIシステム間の連携・調整が成立しないなどのおそれ

- ◆ AIシステム間で交渉を行い、連携・調整するケースにおいて、それぞれのAIシステムが自らの目的の達成を優先して調整が成立しなかったり、適切な交渉の権限を有していないことにより取引が成立しなかったりするおそれがある。また、交渉相手のAIシステムが交渉の機能を有していなかったり、交渉に関するプロトコル等が異なったりする場合にも、AIシステム間の連携・調整が成立しない。さらに、AIシステム間の連携・調整が成立したとしても、社会的に望ましい結果にならない(デジタル・カルテル等)おそれがある。

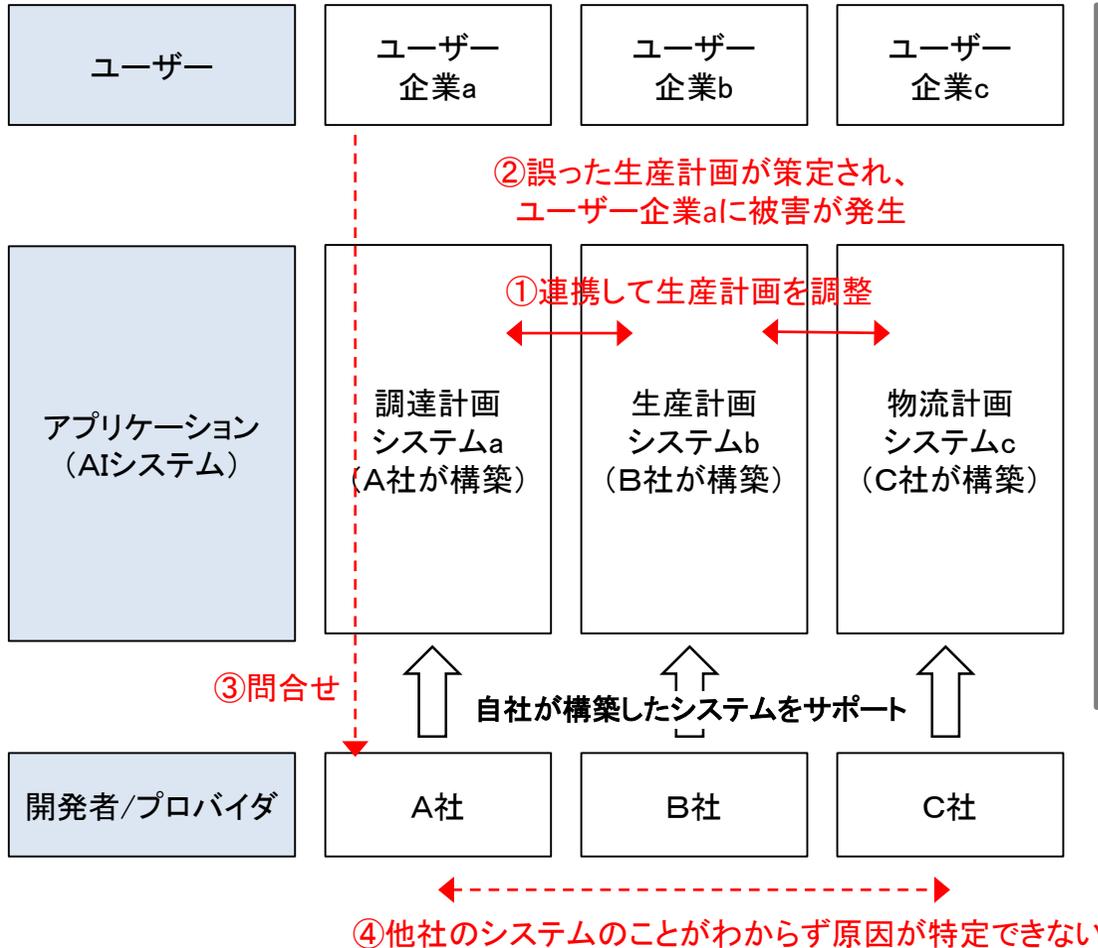


【AIシステム間の連携・調整が成立しないなどのおそれ(例)】

- 自動運転車が車線変更する場合において、A社の自動運転車が車線変更先を走行しているB社の自動運転車に減速するよう要求しても、B社の自動運転車は「なるべく早く目的地に着く」という設定をされていると、減速の要求に応じないというケースが想定される。さらに、車線が減少する場面においては、事故につながるおそれがある。
- AIシステム間で売買契約等における価格交渉を行う場合において、交渉相手のAIシステムが、交渉の権限を有していないにもかかわらず、権限を有しているかのように振る舞うことにより、適切な取引が成立せず、不測の損害を被るケースが想定される。
- 競争関係にある事業者が、それぞれ提供するサービスの価格設定に関し交渉を行う場合において、共通するアルゴリズムを使用していると、協調的価格設定(デジタル・カルテル)を通じて競争が制限されるなどのケースが想定される。

AIの判断・意思決定を検証できないおそれ

- ◆ AIシステムが相互に連携して、協調・調整することになるため、複数のAIシステムが連携して判断、意思決定を行う場面が想定される。このような場合、複雑な意思決定プロセスになる可能性が高く、また、意思決定プロセス自体がブラックボックス化して、AIシステムが行った判断・意思決定の正誤、精度の検証、トラブル時の原因究明や被害者の救済が困難になるおそれがある。

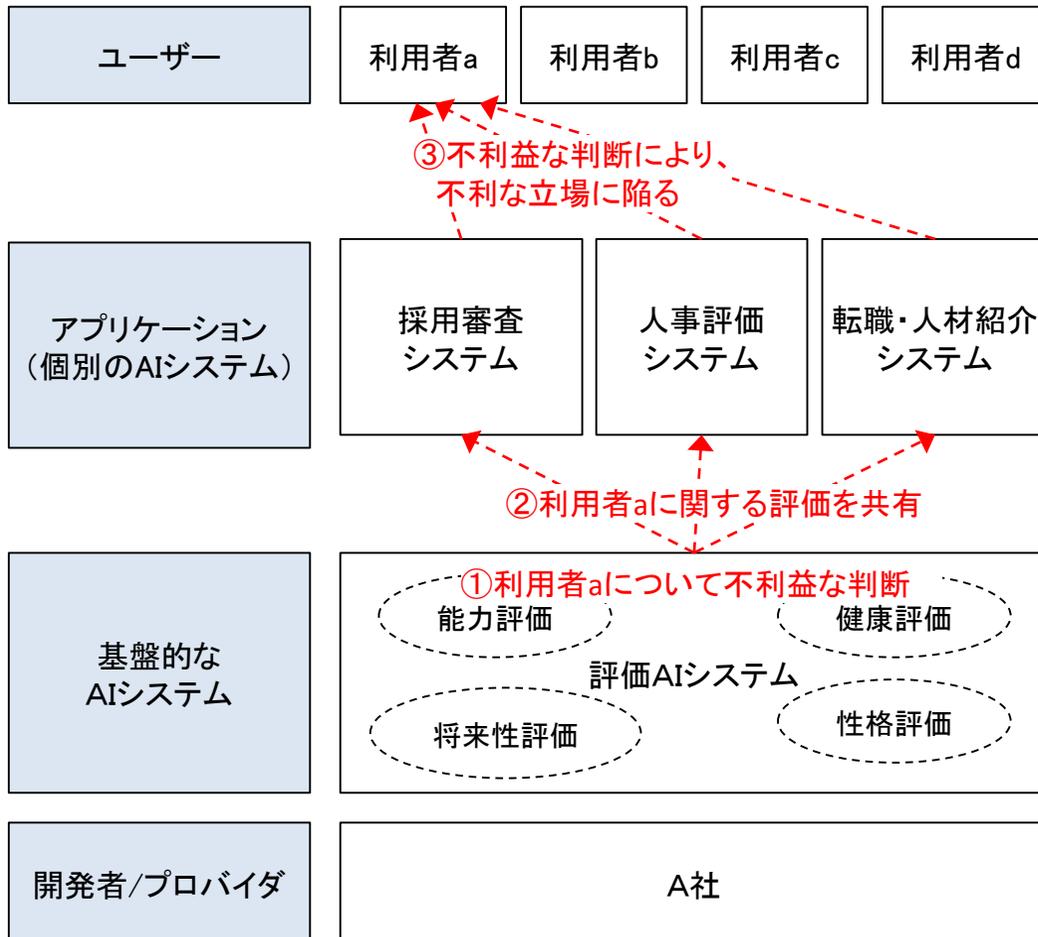


【AIの判断・意思決定を検証できないおそれ(例)】

- サプライチェーン全体として最適化に向けた調整を行っているケースにおいて、個々のAIシステムは、それぞれ最適化を目指して調整するものの、何らかの要因で全体最適とならなかった場合、原因を検証・特定することができないおそれがある。
- 様々な情報(顧客情報、市場動向等)をもとにAIシステムが経営判断を支援しているケースにおいて、それぞれの情報が他の情報の前提となっているような場合、調整自体が複雑となり、人間では、そのプロセスを把握することが困難になるおそれがある。
- 複数のAIシステムが連携してサービスを提供しているケースにおいて、事故等により損害が発生した場合、被害者が、どのAIシステムにより損害がもたらされたのか立証することができず、適切な救済を受けることができないおそれがある。

少数のAIの影響力が強くなりすぎるおそれ

- ◆ AIネットワークのエコシステムにおいては、データに関するネットワーク効果が働きやすいと考えられる。このため、ある特定のAIシステムにデータ等が集中することとなり、その結果として、基盤的なAIシステムが存在する構造が想定される。
- ◆ このような構造において、例えば、「ヒト」の評価や採用のベースとなる基盤的なAIシステムが独占的・寡占的に提供された場合に、基盤的なAIシステムの判断によって、企業や個人が不利な立場になるおそれがある。また、基盤的なAIシステムの開発者が、基盤的なAIシステムとつながっている様々なアプリケーション(個別のAIシステム)の判断を意図的にコントロールし、特定の企業や個人を不利な立場に陥れることができる可能性がある。

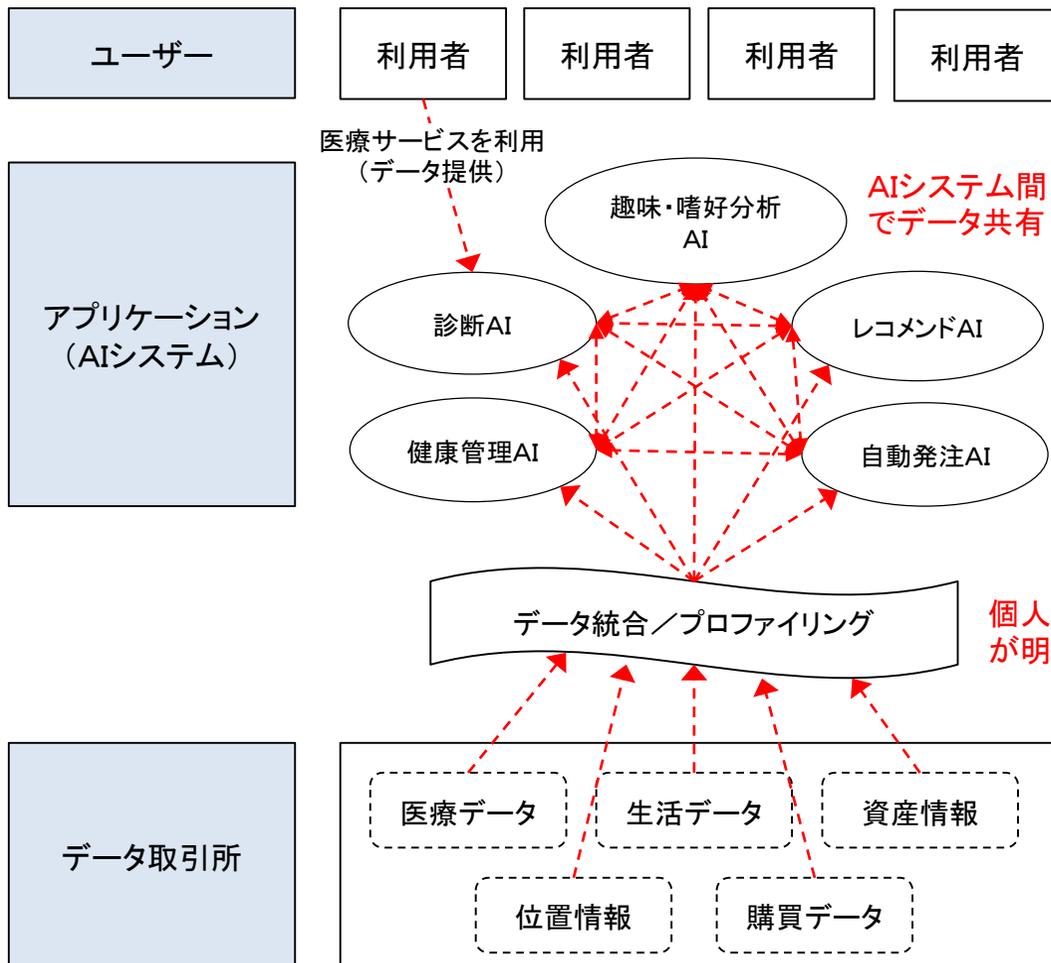


【少数のAIシステムの影響力が強くなりすぎるおそれ(例)】

- 採用(就職)時に不合格となった場合に、他社の採用においても不合格が続いたり、別の会社に就職しても昇格・昇給が後れたり、転職時にも良い評価が得られないなど不利な立場に陥るおそれがある。
- 融資の申込みを断られた場合に、どの金融機関等からも融資が受けられずに事業の存続が困難に陥るおそれがある。
- 入国審査で、誤ってテロリストや犯罪者と認定されてしまった場合に、他国にも入国できなかつたり、差別的な扱いを受けるおそれがある。

プライバシー侵害のおそれ

- ◆ AIネットワークのエコシステムにおいて、領域横断的なデータ取引所の構築、大量のデータを保有し他者に提供するデータブローカー、様々なデータの統合／プロファイリングを行う事業者の登場が想定される。
- ◆ このような状況において、意図しない形で、かつ、本人の知らないところでプライバシー性の高い情報が拡散するおそれがあるほか、様々な情報の統合／プロファイリング等により、個人が特定されプライバシーが侵害されるおそれがある。また、あるAIシステムがハッキング等された場合に、ネットワークを通じて他のAIシステムにおいて情報が共有される（流出する）おそれがある。



【プライバシー侵害のおそれ(例)】

- AIシステム間の調整により、本人同意なく、自動的に個人情報、パーソナルデータが流通するおそれがある。
- データ取引所等から集められた様々なパーソナルデータが統合／プロファイリングされることにより、個人が特定され、個人に紐付いた行動履歴等が明らかになるおそれがある。
- あるAIシステムがハッキング等された場合、ネットワークを通じて即座に他のAIシステムに情報が共有され、もともと保有している情報と統合することにより、個人が特定されるおそれがある。

A I の利活用に関して利用者等が留意することが 期待される事項についての論点（案）

平成30年2月21日
事 務 局

<AIの利用者等が留意することが期待される事項に関する検討について>

○ AIの開発・利活用の加速化、AIネットワーク化の健全な進展等には、AIネットワーク化の便益の増進・リスクの抑制のための取組が必要。

○ これまでの取組

- ・ 本推進会議は、AIの開発者が留意することが期待される事項として「国際的な議論のためのAI開発ガイドライン案」(以下「AI開発ガイドライン案」という。)を取りまとめ。
- ・ 海外では、主に開発者の視点からAIに関する原則・指針・ベストプラクティス・標準等に関する取組が進行。(例) 米国電気電子学会(IEEE)「倫理的に調整された設計 第2版」(2017年12月12日)

○ 他方、AIは、利活用の過程でデータの学習等により自らの出力やプログラムを継続的に変化させる可能性があることから、AIの利用者、間接利用者及びデータ提供者(以下「利用者等」という。)においても留意することが期待される事項があるものと想定される。

→ これらの事項を整理することは、開発者との関係からも重要ではないか。

○ もっとも、AIの利活用については、医療や自動運転など分野ごとに問題の性質や関連する法制度等の相違、国ごとに関連する法制度等の相違が小さくない。

→ 整理に当たっては、分野ごと、国ごとの相違等を考慮する必要があるのではないか。

検討の方向性 (2/4)

<AIの利用者等が留意することが期待される事項に関する本推進会議による検討のアウトプットの在り方について>

- ① AIは、今後大きく発展し、社会の様々な場面で利活用されることが期待されるもの。リスクの所在についても慎重に見極める必要。
 - AIの利用者等が留意することが期待される事項(以下「留意事項」という。)は、AI開発ガイドライン案と同様に、非規制的かつ非拘束的なものとして取りまとめることが適当ではないか。
- ② 本推進会議のアウトプットをどのような形で取りまとめることが適当か。
 - i AIの利活用の推進、非拘束的・非規制的なものとして取りまとめるという点に鑑み、「ガイドライン」という形や名称で取りまとめることは適当か否か。他に、「利活用原則」や「利用者の手引き」という形や名称で取りまとめることも考えられるのではないか。
 - ii AI開発ガイドライン案と同様に、分野共通のものとして取りまとめることが適当か。
それとも、分野ごとの問題の性質等、国ごとの法制度等に相違があることに鑑み、分野共通のものとして取りまとめるのではなく、各分野のステークホルダ等が分野ごとのガイドライン等を策定する際に参照されることが期待される基本的な指針・原則(メタ原則・メタガイドライン)として取りまとめることとし、そのための論点・用語の定義等を提示することが適当か。
 - iii 項目ごとに整理することが適当か、さらに利用者等の類型ごとに整理することが適当か。

検討の方向性 (3/4)

③ 本推進会議のアウトプットをどのような用途として取りまとめることが適当か。

- i AIネットワーク化の進展を見据えて、AI開発ガイドライン案と同様に、国際的な議論のためのものとして取りまとめることが適当か。あるいは、国ごとに関連する法制度等の相違を踏まえて国内向けのものとして取りまとめることが適当か。
- ii 利用者等が参照するガイドライン等を策定する者が参照するものとするか。それとも、利用者等が直接に参照するものまで含めるか。(後者の場合、利用者(特に消費者的利用者)向けに、「ハンドブック」や「マニュアル」などリテラシー教材(利用者の手引き)を作成したり、それらに基づいてワークショップ等を行ったりすること等も考えられる。)

<利用者(特に消費者的利用者)向けのハンドブックのイメージ(例)>

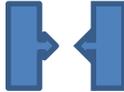
AI利用者ハンドブック

①利用の範囲と方法
解説文.....
.....
.....

②データ
解説文.....
.....
.....

③安全
解説文.....
.....
.....



図・イラスト


④セキュリティ
解説文.....
.....
.....

⑤プライバシー
解説文.....
.....
.....

⑥倫理
解説文.....
.....
.....



図・イラスト


<本推進会議における検討について>

- 前述のとおり、留意事項に関する本推進会議のアウトプットについては、様々な選択肢が考えられる。

本推進会議の次の報告書※までには、差し当たり、アウトプットの在り方如何にかかわらず、アウトプットの取りまとめに向けた検討に資するよう、本ペーパーを踏まえ、留意事項に関する論点を整理することが適当ではないか。

※ 夏頃に取りまとめる予定の『報告書2018』（仮称）

- その後、整理された論点を踏まえ、アウトプットの取りまとめに向けた検討を行うことが適当ではないか。

目的及び基本理念

<本推進会議における検討について>

- 留意事項を取りまとめる目的として、

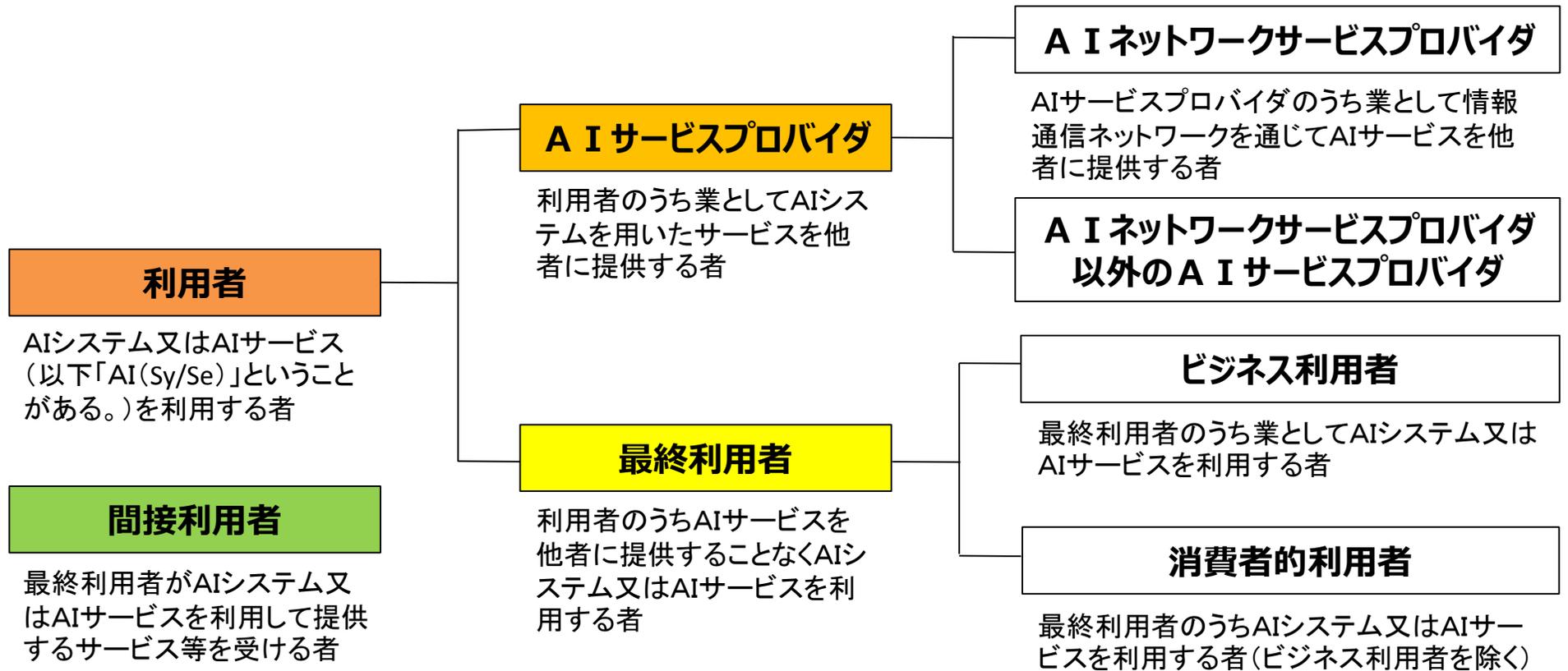
AI及びAIネットワーク化の便益の増進及びリスクの抑制について、AIの開発者や利用者等の間で認識を共有し、協力して取り組むことにより、AIへの人々の信頼を醸成しつつ、AIの利活用の推進及びAIネットワーク化の健全な進展を図り、もって、人間の尊厳と個人の自律が尊重され、万人が包摂される社会を形成すること

が考えられるのではないか。

- 留意事項を取りまとめるに当たっての基本理念(候補)として、例えば、以下の項目を掲げることが考えられるのではないか。
 - i 他者の尊厳と人権の尊重
 - ii AIネットワーク化による人々の包摂
 - iii AIの利活用における公平性と多様性の尊重
 - iv 便益とリスクの適正なバランスの確保
 - v 開発者や利用者等の間の認識とベストプラクティスの共有
 - vi 不断の議論と柔軟な見直し

(注) 基本理念として、どのような項目が適切かについては、アウトプットの在り方(例えば、ガイドライン等を策定する者が参照するものとするか、それとも利用者等が直接に参照するものとするか等)に依存するものと考えられる。

関係する主体の整理 (1/6)



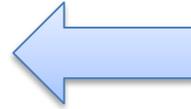
(注) それぞれの名称は仮称。

- 関係する主体間の役割及び責任の分担を検討する上では、各主体がAIシステム及びAIサービスについて有していると期待される能力及び知識の多寡並びに必要な措置を実施する容易性を考慮すべきではないか。
- 特に消費者的利用者及び間接利用者については、子ども、高齢者、障害者など社会的弱者について特別に留意することが期待され事項が考えられるのではないか。

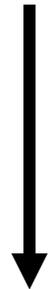
関係する主体の整理 (2/6)

例① <医療用 A I クラウド>

開発者 / A I サービスプロバイダ
(例：医療用 A I クラウド開発者)



自ら開発した A I システムを利用した A I サービス
(例：医療用 A I クラウドサービス) を他者
(例：医師) に提供しており、「開発者」であるとともに、「A I サービスプロバイダ」にも該当

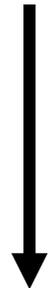


A I サービス (例：医療用
A I クラウドサービス) の提供

最終利用者
(例：医師)



A I サービス (例：医療用 A I クラウドサービス)
を利用する者であり、「最終利用者」に該当



A I サービス (例：医療用 A I
クラウドサービス) を利用したサービス
(例：医療サービス) の提供

間接利用者
(例：患者)



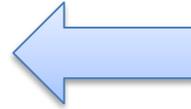
最終利用者 (例：医師) が A I サービス (例：
医療用 A I クラウドサービス) を利用して提供する
サービス (例：医療サービス) を受ける者であり、
「間接利用者」に該当

関係する主体の整理 (3/6)

例② <医療補助ロボット>

開発者
(例：医療補助ロボット製造者)

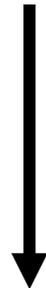
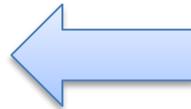
AIサービスを他者に提供するものではなく、「AIサービスプロバイダ」に該当しない



AIシステム（例：医療補助ロボット）の提供

最終利用者
(例：医師)

AIシステム（例：医療補助ロボット）を利用する者であり、「最終利用者」に該当



AIシステム（例：医療補助ロボット）を利用したサービス（例：医療サービス）の提供

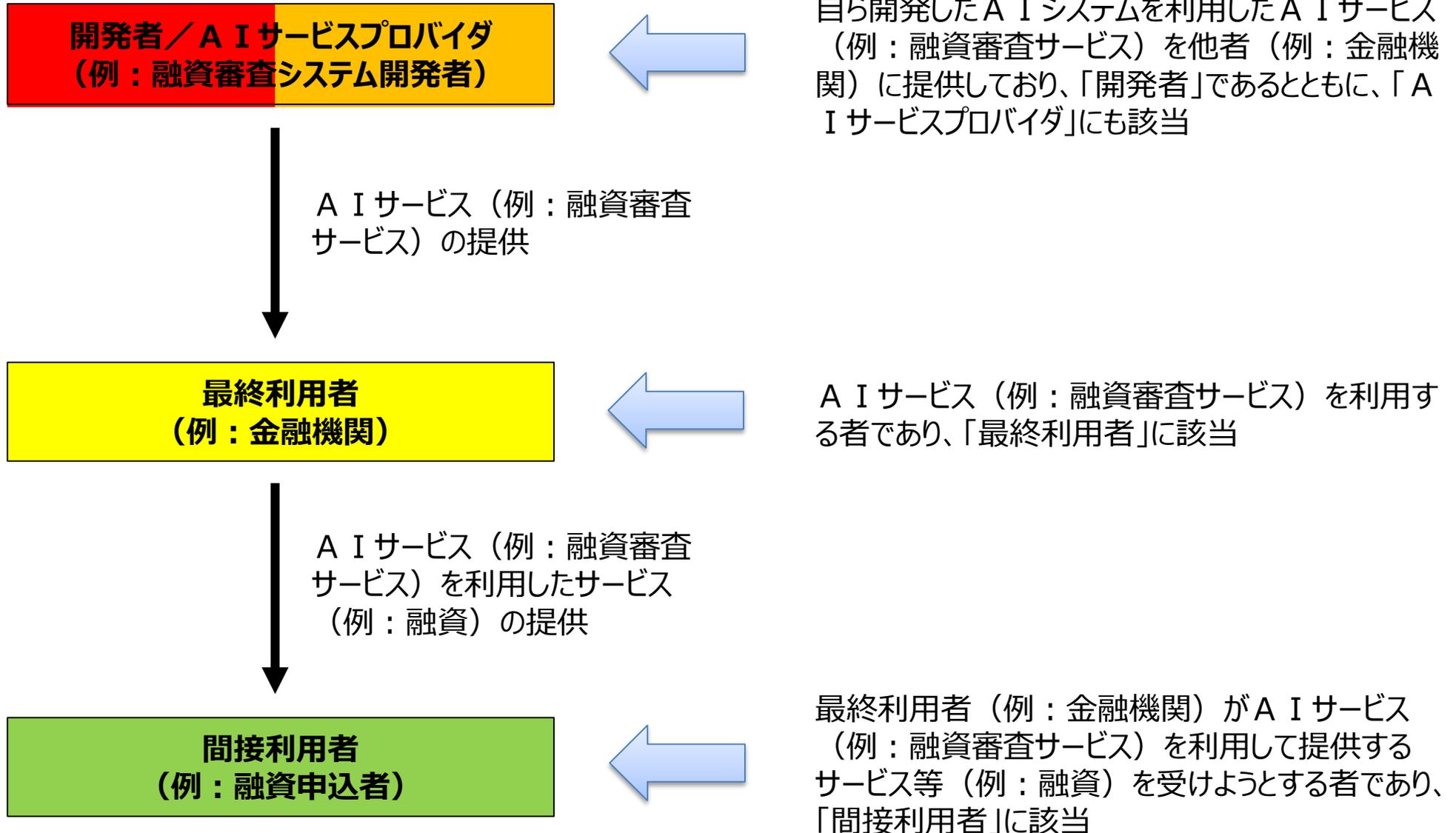
間接利用者
(例：患者)

最終利用者（例：医師）がAIシステム（例：医療補助ロボット）を利用して提供するサービス（例：医療サービス）を受ける者であり、「間接利用者」に該当



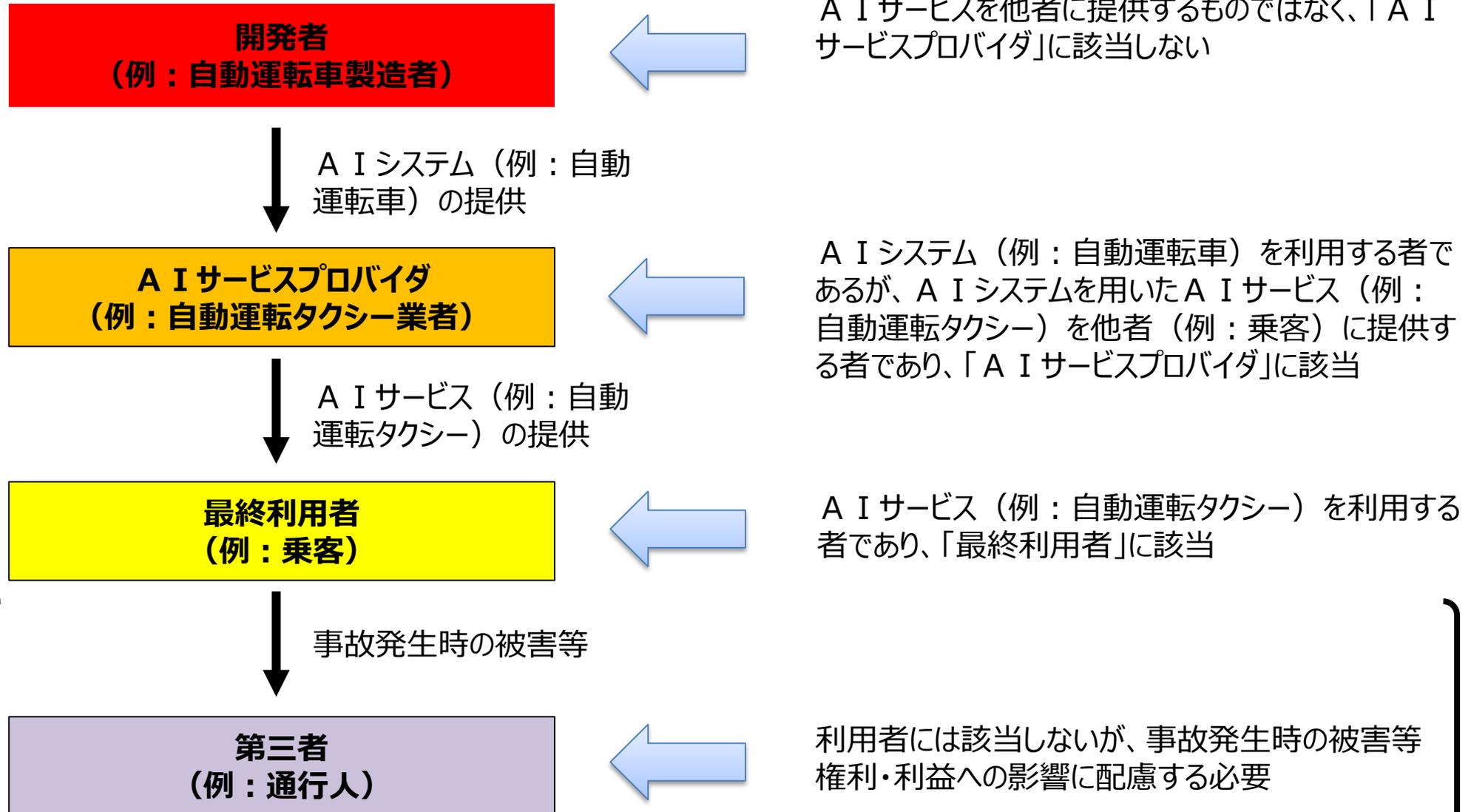
関係する主体の整理 (4/6)

例③ < A I による融資審査 >



関係する主体の整理 (5/6)

例④ <自動運転タクシー>



関係する主体の整理 (6/6)

例⑤ < A I アシスタント >

開発者 / A I サービスプロバイダ
(例: A I アシスタント開発者)

自ら開発した A I システムを利用した A I サービス (例: A I アシスタントサービス) を他者 (例: A I アシスタント対応の A I スピーカーを利用する個人) に提供しており、「開発者」であるとともに、「A I サービスプロバイダ」にも該当

A I サービス (例: A I アシスタントサービス) の提供

最終利用者
(例: A I アシスタント対応の A I スピーカーを利用する個人)

A I サービス (例: A I アシスタントサービス) を利用する者であり、「最終利用者」に該当

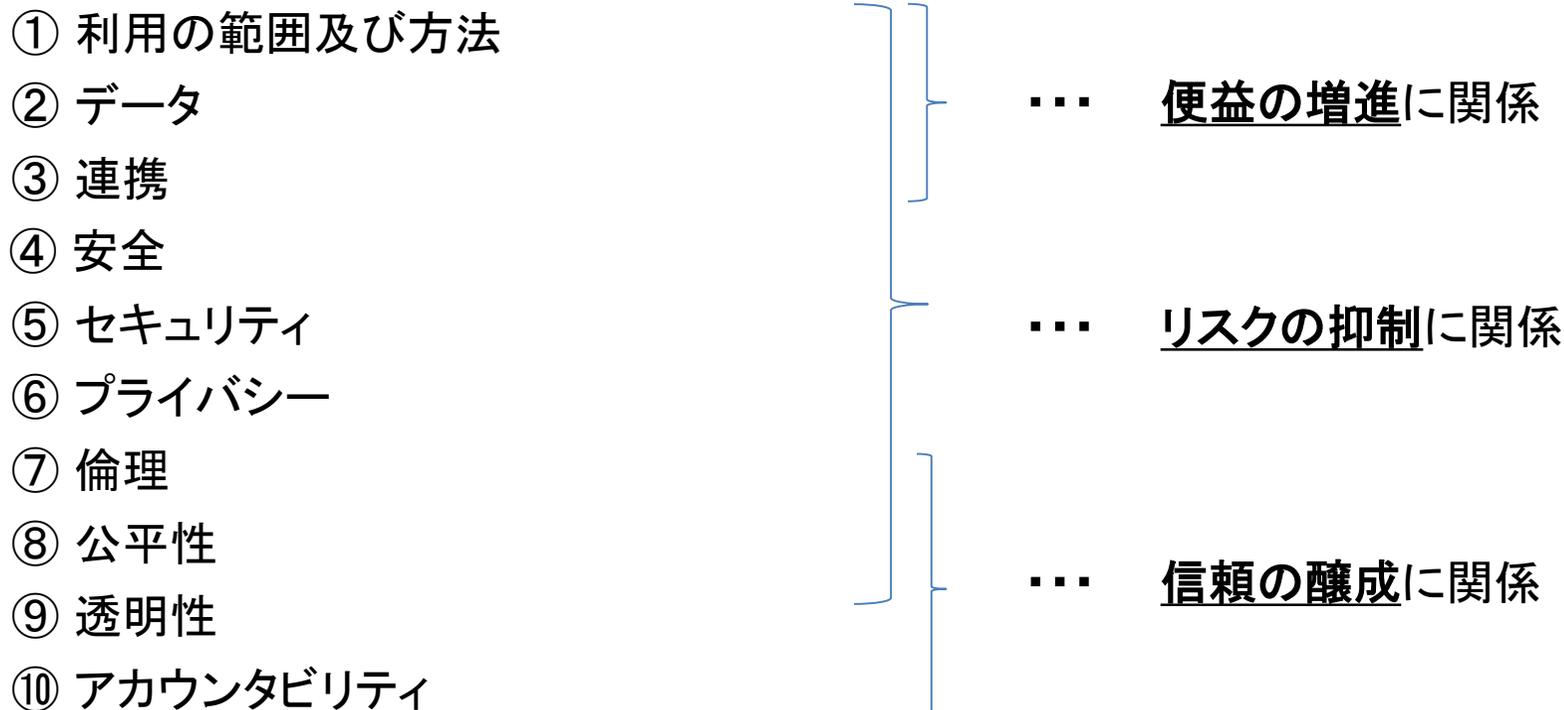
プライバシーへの影響等

第三者
(例: 最終利用者と同居する家族、最終利用者宅を訪問する友人等)

利用者には該当しないが、プライバシーへの影響等権利・利益への影響に配慮する必要

留意事項の論点 (1/11)

- 留意事項の要素としては、「目的及び基本理念」(P5参照)に照らして、例えば、以下の項目が考えられるのではないかな。



(注1) 上記の項目は、あくまでも論点整理の段階における留意事項の項目の案であり、最終的なアウトプット(原則・指針等)においては、利用者の類型ごとに、上記の項目の一部のみが盛り込まれる可能性や上記の項目以外の項目が盛り込まれる可能性が考えられる。

(注2) 各項目に関係する便益及びリスクには、利用者本人に影響を与えるもののみならず、他者や社会全体に波及するものも含まれる。

① 利用の範囲及び方法

【視点】

AI(Sy/Se)は利活用の過程で出力等を変化させるため、利用の範囲及び方法に留意が必要。

【主な論点】

A) 適正な範囲・方法での利用

利用者は、開発者等からの情報提供や説明を踏まえ、社会的文脈や状況にも配慮して、AI(Sy/Se)を適正な範囲・方法で利用することが期待されるのではないか。

B) 人間の判断の介在

AI(Sy/Se)の自動的な判断について、必要かつ可能な場合(例:AIを用いた医療診断等)には、人間の判断を介在させることが期待されるのではないか。その場合、人間の判断の介在の要否の基準は何か、例えば、以下の各点をどのように考えるか。

- ・ AI(Sy/Se)の判断に影響を受ける間接利用者等の権利・利益の性質及び意向
 - ・ AI(Sy/Se)の判断の信頼性の程度(人間の判断の信頼性との優劣)
 - ・ 人間の判断に必要な時間的猶予
 - ・ 利用者に期待される能力
- など

C) AIソフトのアップデート及びAI(Sy/Se)の点検・修理等

利用者は、利活用の過程を通じて、AI(Sy/Se)の機能を向上させ、リスクを抑制するため、AIソフトのアップデート及びAI(Sy/Se)の点検・修理等を行うよう努めることが期待されるのではないか。

② データ

【視点】

AI(Sy/Se)の利活用においては、学習等に用いられるデータの性質等により、その出力等が変化するため、データの性質等によりAI(Sy/Se)の便益及びリスクが左右される可能性。

【主な論点】

A) AI(Sy/Se)の学習等に用いるデータの正確性への留意

利用者及びデータ提供者が、AI(Sy/Se)の学習等に用いるデータの正確性に留意することを期待することは適当か。

利用する技術の特性及び用途に照らして、どのような場合に、どの程度、学習等に用いるデータの正確性に留意することが期待されるか。

B) AI(Sy/Se)の学習等に用いられるデータのバイアスへの留意

AI(Sy/Se)の判断結果により個人が人種・信条・性別等により差別的に取り扱われることのないよう、利用者及びデータ提供者が、AI(Sy/Se)の学習等に用いられるデータのバイアスに留意することを期待することは適当か。

利用する技術の特性及び用途に照らして、どのような場合に、どの程度、学習等に用いるデータのバイアスに留意することが期待されるか。

C) 不正確又は不適切なデータの学習等によるAI(Sy/Se)のセキュリティ脆弱性への留意

利用者及びデータ提供者は、AI(Sy/Se)が不正確又は不適切なデータを学習することにより、AI(Sy/Se)のセキュリティに脆弱性が生ずるリスクに留意することが期待されるのではないか。

③ 連携

【視点】

ネットワークを通じたAI(Sy/Se)相互間又はAI(Sy/Se)と他のシステムやサービスとの連携による便益を増進し、リスクを抑制することの重要性。

【主な論点】

A) 相互接続性と相互運用性への留意

AIネットワークサービスプロバイダは、AI(Sy/Se)の相互接続性と相互運用性に留意することが期待されるのではないかと。

B) データ形式やプロトコル等の標準化への対応

AIサービスプロバイダ及びビジネス利用者は、AI(Sy/Se)間及びAI(Sy/Se)と他のシステム等との連携を促進するため、データ形式やプロトコル等の標準化に対応することが期待されるのではないかと。

C) AIネットワーク化により惹起・増幅される課題への留意

利用者は、自らの利用するAI(Sy/Se)がインターネット等を通じて他のAI(Sy/Se)等と接続・連携することにより制御不能となる等、AIネットワーク化により惹起・増幅される課題(資料4-4参照)に留意することが期待されるのではないかと。

④ 安全

【視点】

AI(Sy/Se)の利活用により、アクチュエータ等を通じて、利用者、間接利用者及び第三者の生命・身体・財産に危害が生ずるおそれ。

【主な論点】

A) 人の生命・身体・財産への配慮

医療や自動運転など人の生命・身体・財産に危害を及ぼし得る分野で利活用する場合には、利用者は、必要に応じてAI(Sy/Se)の点検・修理及びAIソフトのアップデートを行うことなどにより、AI(Sy/Se)がアクチュエータ等を通じて人の生命・身体・財産に危害を及ぼすことのないよう配慮することが期待されるのではないか。

B) 事故等による被害に関する対策の実施

利用者は、AI(Sy/Se)の利活用により生じ得る事故等による被害の性質・態様に応じて、関係者と協力して予防措置及び事後対応(情報共有、停止・復旧、原因解明等)を行うことが期待されるのではないか。

⑤ セキュリティ

【視点】

AIネットワーク化に伴い、従来の情報セキュリティに関するリスクに加え、AI(Sy/Se)が学習等により出力やプログラムを変化させる可能性を踏まえたセキュリティ上の新たなリスクが生ずるおそれを想定。

【主な論点】

A) セキュリティ対策の実施

利用者は、AI(Sy/Se)のセキュリティに留意し、その時点での技術水準に照らして合理的な対策を講ずることが期待されるのではないかと。

B) 不正確又は不適切なデータの学習によるセキュリティ脆弱性への留意

利用者及びデータ提供者は、AI(Sy/Se)が不正確又は不適切なデータを学習することにより、AI(Sy/Se)のセキュリティに脆弱性が生ずるリスクに留意することが期待されるのではないかと。

C) セキュリティ対策のためのサービス提供等

AIネットワークサービスプロバイダは、自らの提供するAIネットワークサービスについて、最終利用者にセキュリティ対策のためのサービスを提供するとともに、インシデント情報の共有を図ることが期待されるのではないかと。

⑥ プライバシー

【視点】

パーソナルデータの利活用により、AI(Sy/Se)の便益の増進が期待される反面、プライバシーが侵害されるおそれ。

【主な論点】

A) 他者のプライバシーの尊重

利用者は、AI(Sy/Se)を利活用する際の社会的文脈や人々の合理的な期待を踏まえ、他者のプライバシーを尊重することが期待されるのではないか。

B) パーソナルデータの収集・分析・提供等におけるプライバシーの尊重

利用者及びデータ提供者は、AI(Sy/Se)の学習等に用いられるパーソナルデータの収集・分析・提供等において、他者のプライバシーを尊重することが期待されるのではないか。

C) AI(Sy/Se)を利用したプロファイリングを行う場合におけるプライバシー等の配慮

利用者は、人事評価・採用や融資など個人の権利・利益に重要な影響を及ぼす可能性のある分野においてAI(Sy/Se)を利用したプロファイリングを行う場合には、対象者のプライバシー等に慎重に配慮することが期待されるのではないか。

⑦ 倫理

【視点】

AIネットワーク化の進展により、人間の幸福の増進が期待される反面、人間の尊厳と個人の自律が脅かされるなど倫理的な問題が生ずるおそれ。

【主な論点】

A) 他者の尊厳と自律の尊重

利用者は、AI(Sy/Se)を利活用する際の社会的文脈を踏まえ、他者の尊厳と自律を尊重することが期待されるのではないか。

B) AI(Sy/Se)による意思決定過程・感情の操作等への留意

利用者には、AI(Sy/Se)により意思決定過程や感情が操作されるリスクや、AIに過度に依存するリスクに留意することを期待することは適当か否か。また、このようなリスクについて誰がいかなる役割を果たすべきか。

C) AI(Sy/Se)と人間の脳・身体を連携する際の生命倫理等の議論の参照

利用者は、AI(Sy/Se)を人間の脳・身体と連携させる場合には、生命倫理の議論等を参照し、他者の尊厳と自律が侵害されないよう特に慎重な配慮が求められるのではないか。

⑧ 公平性

【視点】

AI(Sy/Se)の判断により少数者等が差別的な取扱いが行われるおそれ。

【主な論点】

A) AI(Sy/Se)の学習等に用いられるデータのバイアスへの留意

AI(Sy/Se)の判断結果により、個人が人種・信条・性別等によって差別的に取り扱われることのないよう、利用者及びデータ提供者が、AI(Sy/Se)の学習等に用いられるデータのバイアスに留意することを期待することは適当か。

利用する技術の特性及び用途に照らして、どのような場合に、どの程度、学習等に用いるデータのバイアスに留意することが期待されるか。

B) 人間の判断の介在

AI(Sy/Se)の判断結果により、個人が人種・信条・性別等によって差別的に取り扱われることのないよう、AIサービスプロバイダ及びビジネス利用者が、評価対象の個人の意向に応じて、人間の判断を介在させることを期待することは適当か。

利用する技術の特性及び用途に照らして、どのような場合に、どの程度、人間の判断を介在させることが期待されるか。

⑨ 透明性

【視点】

AI(Sy/Se)の利活用により事故等が発生した場合の原因説明を可能とすること、利用者によるAI(Sy/Se)の判断への異議申立てを可能にすることなどにより、AI(Sy/Se)への信頼を醸成。

【主な論点】

A) AI(Sy/Se)の入出力の記録・保存

AIサービスプロバイダ及びビジネス利用者が、AI(Sy/Se)の入出力の検証可能性を確保するため、入出力を記録・保存することを期待することは適当か。

利用する技術の特性及び用途に照らして、どのような場合に、どの程度、入出力を記録・保存することが期待されるか。例えば、自動運転車など人の生命・身体・財産に危害を及ぼし得る分野で利活用する場合には、AI(Sy/Se)の入出力を記録・保存することが期待されるのではないか。

B) 説明可能性の確保

AIサービスプロバイダ及びビジネス利用者が、AI(Sy/Se)の判断結果の説明可能性を確保することを期待することは適当か。

利用する技術の特性及び用途に照らして、どのような場合に、どの程度、説明可能性を確保することが期待されるか。特に、医療診断、人事評価・採用、融資など個人の権利・利益に重大な影響を及ぼす可能性のある分野で利活用する場合には、AI(Sy/Se)の判断結果の説明可能性を確保することを期待することは適当か。

(例えば、深層学習は予測精度が高いが、判断結果の説明が困難な現状等にも留意)。

⑩ アカウンタビリティ

【視点】

AI(Sy/Se)に対する利用者及び社会の信頼を醸成することが重要。

【主な論点】

A) アカウンタビリティを果たす努力

AIサービスプロバイダ及びビジネス利用者は、人々と社会からAI(Sy/Se)への信頼を獲得することができるよう、消費者的利用者、間接利用者、AI(Sy/Se)の利用により影響を受ける第三者等に対しアカウンタビリティを果たすよう努めることが期待されるのではないか。

B) AI(Sy/Se)に関する利用方針の通知・公表

消費者的利用者や間接利用者らがAI(Sy/Se)の利活用について適切に認識することのできるよう、AIサービスプロバイダ及びビジネス利用者が、AIに関する利用方針(AI(Sy/Se)を利用している場合にはその旨、利活用の範囲及び方法、利活用に伴うリスク、相談窓口など)を通知又は公表することを期待することは、適切か。

利用する技術の性質及び用途に照らして、どのような場合に、どのような内容の利用方針を通知又は公表を確保することが期待されるか。

【参考】 国際的な議論のためのAI開発ガイドライン案の概要

目的

- AIネットワーク化の進展による**便益は広範で多大** その一方、**不透明化等のリスク**への懸念も存在
- **拘束的ではないソフトロー**としてガイドラインを**国際的に共有**することが必要
- 本ガイドラインの目的

AIネットワーク化の健全な進展を通じたAIシステムの**便益の増進**と**リスクの抑制**

- ・ **利用者の利益を保護**
- ・ **リスクの波及を抑止**

人間中心の「智連社会」を実現

基本理念

- ・ AIネットワーク化の**便益がすべての人によりあまねく享受**され、人間の尊厳と個人の自律が尊重される**人間中心の社会**を実現
- ・ **拘束的ではないソフトロー**として**国際的に共有**
- ・ **イノベティブでオープンな研究開発**と**公正な競争**、学問の自由等を尊重するとともに、**便益とリスクの適正なバランス**を確保
- ・ **技術的中立性**を確保し、開発者に**過度の負担**を課さないよう配慮
- ・ **継続的な見直し**、必要に応じた**柔軟な改定**、広範で柔軟な議論

AI開発原則

- ① **連携の原則**
 - ・ AIシステムの**相互接続性**と**相互運用性**に留意
- ② **透明性の原則**
 - ・ AIシステムの**入出力の検証可能性**及び**判断結果の説明可能性**に留意
- ③ **制御可能性の原則**
 - ・ AIシステムの**制御可能性**に留意
- ④ **安全の原則**
 - ・ AIシステムがアクチュエータ等を通じて**利用者及び第三者の生命・身体・財産**に危害を及ぼすことがないよう配慮
- ⑤ **セキュリティの原則**
 - ・ AIシステムの**セキュリティ**に留意
- ⑥ **プライバシーの原則**
 - ・ AIシステムにより利用者及び第三者の**プライバシー**が侵害されないよう配慮
- ⑦ **倫理の原則**
 - ・ AIシステムの開発において、**人間の尊厳と個人の自律**を尊重
- ⑧ **利用者支援の原則**
 - ・ AIシステムが利用者を支援し、**利用者**に**選択の機会**を適切に提供することが可能となるよう配慮
- ⑨ **アカウントビリティの原則**
 - ・ 利用者を含むステークホルダに対し**アカウントビリティ**を果たすよう努める

関係者に期待される役割

- ・ 各国政府及び国際機関： 多様なステークホルダ間の対話の促進に向けた環境整備
- ・ 開発者、利用者等ステークホルダ： 対話やベストプラクティスの共有、相互協力
- ・ 各国政府： AIの開発者コミュニティの支援、AIに関する研究開発を支援する政策の積極的な推進

【参考】 国際的な議論のためのAI開発ガイドライン案における定義

- 「AI」とは、「AIソフト及びAIシステムを総称する概念」をいう。
- 「AIソフト」とは、データ・情報・知識の学習等により、利活用の過程を通じて自らの出力やプログラムを変化させる機能を有するソフトウェアをいう。例えば、機械学習ソフトウェアはこれに含まれる。
- 「AIシステム」とは、AIソフトを構成要素として含むシステムをいう。例えば、AIソフトを実装したロボットやクラウドシステムはこれに含まれる。
- 「AIネットワークサービス」とは、情報通信ネットワークを通じてAIシステムの機能を提供するサービスをいう。
- AIシステムの「開発者」とは、AIシステムの研究開発（AIシステムを利用しながら行う研究開発を含む。）を行う者（自らが開発したAIシステムを用いてAIネットワークサービスを他者に提供するプロバイダを含む。）をいう。
- AIシステムの「利用者」とは、AIシステムを利用する者（最終利用者（エンドユーザ）のほか、他者が開発したAIネットワークサービスを第三者に提供するプロバイダを含む。）をいう。

A I の利活用、A I ネットワーク化に係る諸課題（案）

平成30年2月21日
事 務 局

AIの利活用、AIネットワーク化に係る諸課題（1/2）

- ◆ AIの開発及び利活用の加速化、AIネットワークの健全な発展に係る諸課題としては、例えば、次のような事項も考えられるのではないか。これらの事項について、本推進会議又は関係するステークホルダにおいて検討や取組を行うことが期待されるのではないか。
- AIの利活用、AIネットワーク化に関するステークホルダ間の協力、ベストプラクティスの共有の在り方
- AIの学習等に用いられるデータの形式やAI間の自動交渉のプロトコル等の標準化の在り方
- AIの利活用、AIネットワーク化に必要な知識・技能等に関する啓発・教育の在り方
- AIの研究開発及び利活用並びにAIネットワーク化を推進するための人材育成の在り方
- AIの利活用、AIネットワーク化に係る市場競争環境整備、利用者の選択肢の多様性の確保の在り方
- AIの便益及びリスクを検証するための実験環境（特区、仮想実験空間等）の整備の在り方
- AIの事故等に関する被害者の救済（保険等）及び被害発生の防止の在り方
- AIの利活用、AIネットワーク化を推進するための法制度等の在り方
 - AIの事故等に関する法的責任の所在の在り方
 - 各分野におけるAIの利活用と業法等との関係の整理
 - AIの学習等に用いられるデータ、AIにより作成されたモデル、AI創作物等に関する知的財産法上の保護と利活用の在り方
 - AIの利活用に関する個人情報保護と利活用の在り方
 - AIネットワークに関する通信の秘密の保護（AI間の通信の秘密の保護の範囲及び程度など）の在り方

AIの利活用、AIネットワーク化に係る諸課題（2/2）

- ◆ AIの利活用、AIネットワーク化のシナリオ、利用者等が留意することが期待される事項等について、当推進会議において今後引き続き検討を行うものとして、例えば、次の事項が考えられるのではないか。

AIの利活用、AIネットワーク化のシナリオ等に関するもの

- AIの利活用及びAIネットワーク化の進展による社会制度への影響（例：教育など）、産業構造の変化（例：産業の再編・融合など）、経済への影響（例：雇用・働き方など）や人間の心理への影響（例：AI依存など）等に関する中長期的なシナリオ分析
- AIの利活用に関するエコシステム（AIネットワーク化のエコシステム、データのエコシステム、ビジネスのエコシステム等）の在り方
- AIネットワーク化の進展及び影響（幸せと豊かさなど）に関する指標の在り方

利用者等が留意することが期待される事項に関するもの

- 利用者等が留意することが期待される事項についての論点整理を踏まえ、AIの利活用に関するアウトプットの取りまとめ等
- 利用者等が留意することが期待される事項の検討からのフィードバックを踏まえ、必要に応じて、「国際的な議論のためのAI開発ガイドライン案」の内容（開発者の定義、利活用段階における開発者の留意事項等）の見直し