報告書２０１９概要

令　和　元　年　８　月　９　日
ＡＩネットワーク社会推進会議

報告書2019の構成

AIネットワーク社会推進会議 『報告書2019』

- はじめに
- 第１章　AIネットワーク化をめぐる最近の動向
  - １．国内の動向
  - ２．海外の動向
  - ３．国際的な議論の動向
- 第２章　AI利活用ガイドライン策定の考え方
  - １．背景・経緯
  - ２．AI利活用ガイドラインの位置づけ
  - ３．AI利活用ガイドラインの概要
  - ４．今後の展開
- 第３章　今後の課題
- 結び

- 【別紙】 AI利活用ガイドライン
  - 目的・基本理念
  - 主体の整理
  - AI利活用原則
  - 一般的なAI利活用の流れ
  - AI利活用原則の解説
  - AI利活用原則を考慮すべきタイミング

第１章　AIネットワーク化をめぐる最近の動向

国内

○「人間中心のAI社会原則」公表（平成31年３月29日、統合イノベーション戦略推進会議決定）

AIをよりよい形で社会実装し共有するための基本原則を策定し、国際的な議論に供するため、政府は、AI戦略実行会議の下、産学民官による「人間中心のAI社会原則会議」を設置し、「人間中心のAI社会原則」を公表。社会原則は、人間中心、教育・リテラシー、プライバシー確保、セキュリティ確保、公正競争確保、公平性・説明責任及び透明性、イノベーションの７原則から構成。

海外

○米国電気電子学会（IEEE）「倫理的に調整された設計　第１エディション」公表（平成31年３月25日）

AIの倫理的な設計、開発及び実装において参照されるべき一般原則として、人権、幸福、データ仲介、効能、透明性、アカウンタビリティ、悪用への警戒、技能を掲記。また、倫理課題から技術への橋渡しを目的とした標準であるIEEE P7000™シリーズや用語集の作成など「原則から実行へ」を意識。

○欧州委員会「Ethics Guidelines for Trustworthy AI」公表（平成31年４月８日）

欧州委員会は、選定した52名の専門家グループにより作成された「信頼できるAIのための倫理ガイドライン」を公表。同ガイドラインでは、信頼できるAIのためには合法的、倫理的、及び、頑健であるべきとし、その上で基本的人権に基づき尊重すべき４つの倫理原則（人間の自律性の尊重、危害の防止、公平性、説明可能性）、および７つの要求条件（人間の営みと監視、技術的な頑健性と安全性、プライバシーとデータガバナンス、透明性、多様性・無差別・公平性、環境及び社会の幸福、アカウンタビリティ）を掲げ、さらにそれらを評価するためのリスト（Assessment list）を列挙。

国際的な議論

○G7マルチステークホルダ会合（カナダ、平成30年12月6日）

G7各国が、『①社会のためのAI』、『②イノベーションの解放』、『③AIにおけるアカウンタビリティ』、『④仕事の未来』のうち一つのテーマを担当し、ディスカッションペーパーを作成の上マルチステークホルダによる議論を実施。我が国は、カナダとともに『③AIにおけるアカウンタビリティ』を担当。

○OECD理事会勧告公表（令和元年５月）

閣僚理事会は、平成30年９月から４度開催されたAIに関する専門家会合（AIGO: AI expert Group at the OECD）での検討を踏まえて作成された理事会勧告を承認。同勧告は、AIの開発者・運用者等に対する「信頼できるAIのための責任あるスチュワードシップに関する原則」（包摂的な成長・持続可能な開発及び幸福、人間中心の価値及び公平性、透明性及び説明可能性、頑健性・セキュリティ及び安全性、アカウンタビリティ）と国に対する「信頼できるAIのための国内政策と国際協力」から構成。なお、勧告の記載は原則とその概括的な説明にとどまっており、具体的に講じるべき措置等については勧告策定後のCDEP会合で別途検討。

○ G20茨城つくば貿易・デジタル経済大臣会合（令和元年６月８日・９日）

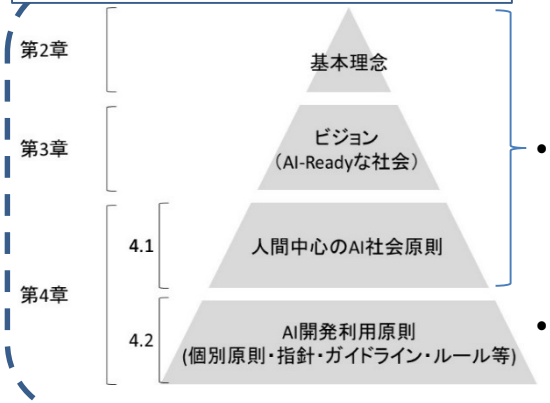
AIの開発や利活用の促進に向け、G20ではじめて「人間中心」の考えを踏まえたAI原則（「G20 AI原則」）に対し賛同が得られ、その内容を含む閣僚声明が採択。同原則は、OECD理事会勧告を引用して作成されたものであり、閣僚声明の附属文書として合意。

（注）主なものを記載

第２章　AI利活用ガイドラインの位置づけ

■民間等で原則等を策定する際の参照

「人間中心のAI社会原則」より抜粋



・・・「社会（特に、国などの立法・行政機関）が留意すべき『AI社会原則』」

・・・「開発者及び事業者において、基本理念及びAI社会原則を踏まえた
AI開発利用原則を定め、遵守するべき」

開発者・事業者それぞれにおいて、AI開発利用原則を策定することを期待

そのために参照すべき具体的な解説書が必要

■国際的な議論への貢献

AI原則の項目については、国際的にほぼコンセンサスが得られつつあり、今後は原則の実効性を確保するための具体的手段についての議論に移行。これらの議論に貢献し、認識の共有を図る。

（例）

- 欧州委員会：「信頼できるAIのための倫理ガイドライン」におけるAssessment list
  →今後レビューを行い、2020年に取りまとめる予定
- OECD：「理事会勧告」を実現するために具体的に講じるべき措置等
  →本年７月以降、CDEP会合で別途検討

第２章　AI利活用ガイドライン（全体構成）

パート１：AI利活用原則の考え方

1. 目的

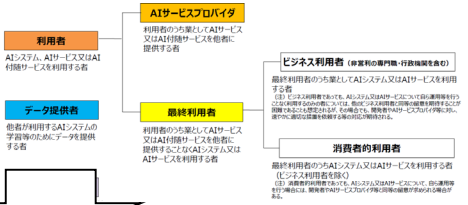
AIの利活用・社会実装の促進

2. 基本理念

- 人間中心の社会の実現
- AI利活用における多様性の尊重・包摂
- AIネットワーク化による持続可能な社会の実現
- 便益とリスクの適正なバランスの確保
- ステークホルダ間の知識・能力相応の役割分担
- 指針やベストプラクティスの国際的な共有
- 不断の見直し・柔軟な改定

3.関係する主体の整理

「利用者」の分類



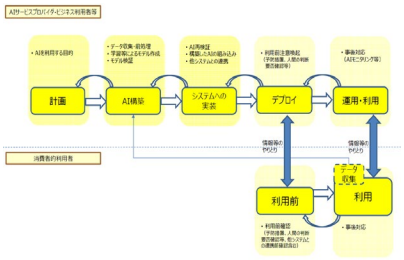
4.AI利活用原則

10原則
適正利用、適正学習、連携、安全、セキュリティ、プライバシー、尊厳・自律、公平性、透明性、アカウンタビリティ

パート２：AI利活用原則の解説

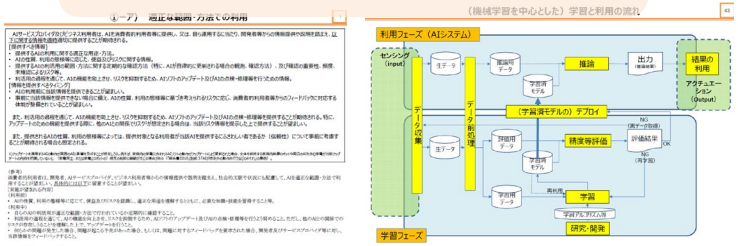
5. 一般的なAI利活用の流れ

AI利活用の一般的な流れ



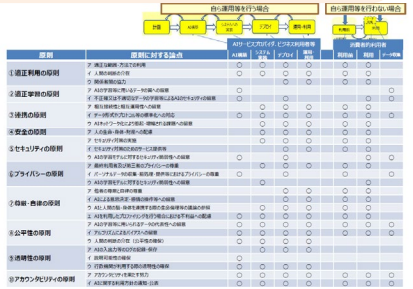
6.AI利活用原則の解説

10原則の各論点に関する解説
（およびその詳説（附属資料））



7.AI利活用原則を考慮すべきタイミング

各原則各論点を
考慮すべきタイミング



第２章　AI利活用ガイドライン（目的･基本理念）

目的

AIネットワーク化の健全な進展を通じて、AIの便益の増進とリスク※の抑制を図り、AIに対する信頼を醸成することにより、AIの利活用や社会実装を促進する。

基本理念

- 人間がAIネットワークと共生することにより、その恵沢がすべての人によってあまねく享受され、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**すること
- **AIの利活用において**利用者の**多様性を尊重し**、多様な背景と価値観、考え方を持つ人々を**包摂する**こと
- **AIネットワーク化により**個人、地域社会、各国、国際社会が抱える様々な課題の解決を図り、**持続可能な社会を実現**すること
- AIネットワーク化による便益を増進するとともに、民主主義社会の価値を最大限尊重しつつ、権利利益が侵害されるリスクを抑制するため、**便益とリスクの適正なバランスを確保**すること
- AIに関して有していると期待される**能力や知識等に応じ、ステークホルダ間における適切な役割分担を実現**すること
- AIの利活用の在り方について、非拘束的なソフトローたる**指針やベストプラクティスを国際的に共有**すること
- AIネットワーク化の進展等を踏まえ、国際的な議論を通じて、本ガイドラインを**不断に見直し**、必要に応じて**柔軟に改定する**こと

※「リスク」とは「損害等の不利益をもたらす可能性」を意味する。

第２章　AI利活用ガイドライン（関係する主体の整理）

開発者

AIシステムの研究開発を行う者

利用者

AIシステム、AIサービス又はAI付随サービスを利用する者

データ提供者

他者が利用するAIシステムの学習等のためにデータを提供する者

第三者

他者の利用するAIシステム又はAIサービスにより自らの権利・利益に影響を受ける者

AIサービスプロバイダ

利用者のうち業としてAIサービス又はAI付随サービスを他者に提供する者

最終利用者

利用者のうち業としてAIサービス又はAI付随サービスを他者に提供することなくAIシステム又はAIサービスを利用する者

ビジネス利用者（非営利の専門職・行政機関を含む）

最終利用者のうち業としてAIシステム又はAIサービスを利用する者

（注）ビジネス利用者であっても、AIシステム又はAIサービスについて自ら運用等を行うことなく利用するのみの者については、他のビジネス利用者と同等の留意を期待することが困難であることも想定されるが、その場合でも、開発者やAIサービスプロバイダ等に対し、速やかに適切な措置を依頼する等の対応が期待される。

消費者的利用者

最終利用者のうちAIシステム又はAIサービスを利用する者（ビジネス利用者を除く）

（注）消費者的利用者であっても、AIシステム又はAIサービスについて、自ら運用等を行う場合には、開発者やAIサービスプロバイダ等と同等の留意が求められる場合がある。

・AIシステム　：　AIソフトを構成要素として含むシステム
・AIサービス　：　AIシステムの機能を提供するサービス
・AI付随サービス：　AIシステムのアップデート又は追加的な学習等に係るサービス

（注）同一の個人・事業者が複数の主体に該当する場合がある。

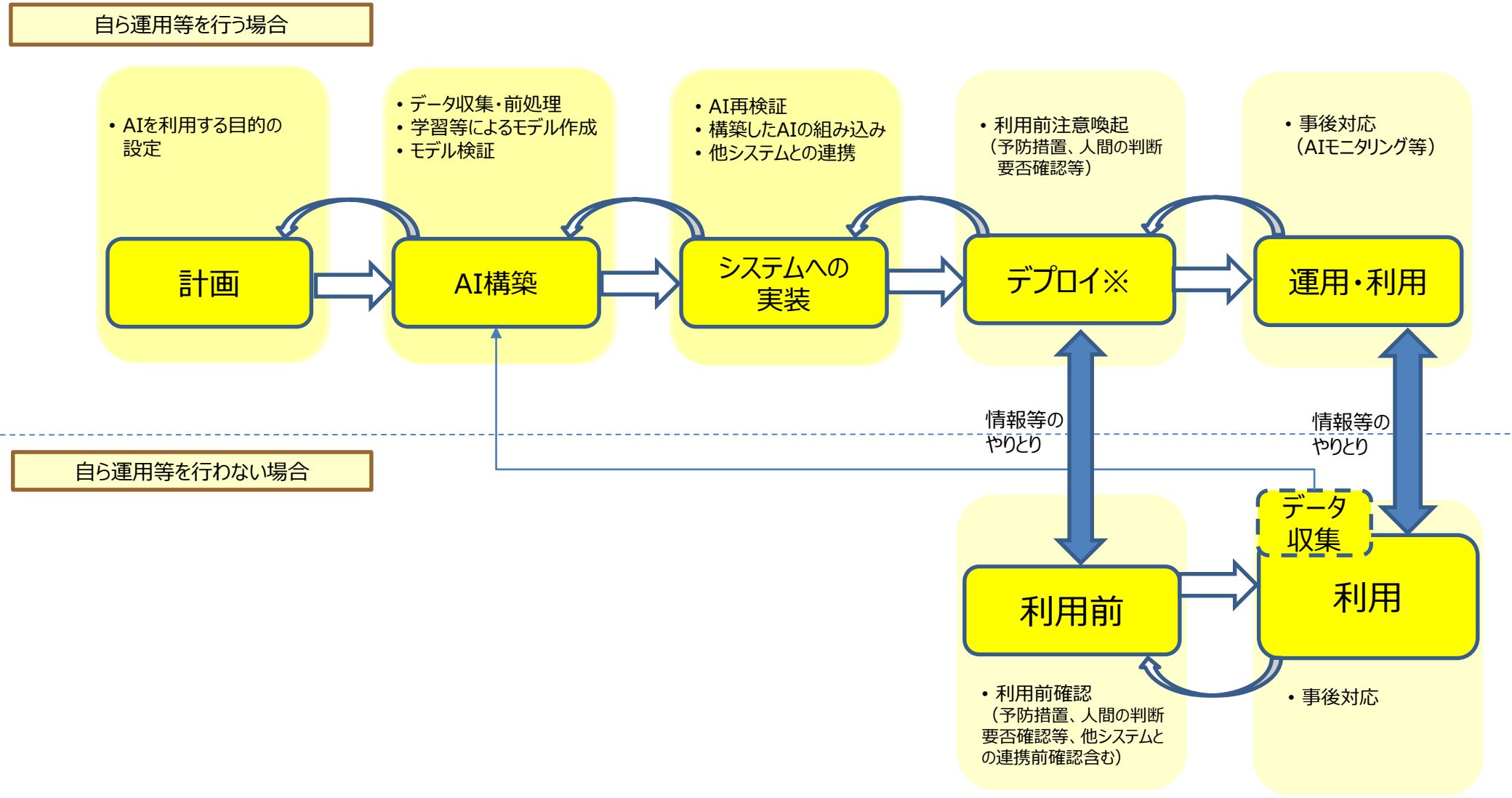
第２章　AI利活用ガイドライン（AI利活用原則）

AIサービスプロバイダやビジネス利用者等が**自主的に参照**するものとして、また**国際的な認識の共有**を図るものとして取りまとめ

原則	説明
①適正利用の原則	利用者は、人間とAIシステムとの間及び利用者間における適切な役割分担のもと、適正な範囲及び方法でAIシステム又はAIサービスを利用するよう努める。
②適正学習の原則	利用者及びデータ提供者は、AIシステムの学習等に用いるデータの質に留意する。
③連携の原則	AIサービスプロバイダ、ビジネス利用者及びデータ提供者は、AIシステム又はAIサービス相互間の連携に留意する。また、利用者は、AIシステムがネットワーク化することによってリスクが惹起・増幅される可能性があることに留意する。
④安全の原則	利用者は、AIシステム又はAIサービスの利活用により、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する。
⑤セキュリティの原則	利用者及びデータ提供者は、AIシステム又はAIサービスのセキュリティに留意する。
⑥プライバシーの原則	利用者及びデータ提供者は、AIシステム又はAIサービスの利活用において、他者又は自己のプライバシーが侵害されないよう配慮する。
⑦尊厳・自律の原則	利用者は、AIシステム又はAIサービスの利活用において、人間の尊厳と個人の自律を尊重する。
⑧公平性[1]の原則	AIサービスプロバイダ、ビジネス利用者及びデータ提供者は、AIシステム又はAIサービスの判断にバイアスが含まれる可能性があることに留意し、また、AIシステム又はAIサービスの判断によって個人及び集団が不当に差別されないよう配慮する。
⑨透明性の原則[2]	AIサービスプロバイダ及びビジネス利用者は、AIシステム又はAIサービスの入出力等の検証可能性及び判断結果の説明可能性に留意する。
⑩アカウンタビリティ[3]の原則	利用者は、ステークホルダに対しアカウンタビリティを果たすよう努める。

1) 「公平性」には複数の基準があることに留意する必要がある。
2) 本原則は、アルゴリズム、ソースコード、学習データの開示を想定するものではない。また、本原則の解釈に当たっては、プライバシーや営業秘密への配慮も求められる。
3) アカウンタビリティ：　判断の結果についてその判断により影響を受ける者の理解を得るため、責任者を明示した上で、判断に関する正当な意味・理由の説明、必要に応じた賠償・補償等の措置がとれること。

第２章　AI利活用ガイドライン（一般的なAI利活用の流れ）



※デプロイ：　（AIソフト／システムを）利用可能な状態にすること

（注）上記のAI利活用の一般的な流れは、利活用のどのフェーズでAI利活用原則のどの論点を考慮すべきか（後述）についての整理を行うために典型的な事例を記載。他方、AIの利活用は本図にあるとおり各フェーズが時系列で整理される場合だけではなく、DevOpsのように開発と運用が一体のものとして検討される場合など、多様なケースが存在。

第２章　AI利活用ガイドライン（主な論点と考慮すべきタイミング）

原則	原則に対する論点	AI構築	システム実装	デプロイ	運用・利用		利用前	利用	データ収集
①適正利用の原則	ア　適正な範囲・方法での利用	○	○	○	○		○	○	
	イ　人間の判断の介在	○	○	○	○		○	○	
	ウ　関係者間の協力			○	○		○	○	
②適正学習の原則	ア　AIの学習等に用いるデータの質への留意	○							○
	イ　不正確又は不適切なデータの学習等によるAIのセキュリティの留意	○		○				○	○
③連携の原則	ア　相互接続性と相互運用性への留意		○	○	○		○	○	
	イ　データ形式やプロトコル等の標準化への対応	○	○	○	○		○	○	○
	ウ　AIネットワーク化により惹起・増幅される課題への留意		○	○	○		○	○	
④安全の原則	ア　人の生命・身体・財産への配慮		○	○	○		○	○	
⑤セキュリティの原則	ア　セキュリティ対策の実施		○	○	○		○	○	
	イ　セキュリティ対策のためのサービス提供等			○	○		○	○	
	ウ　AIの学習モデルに対するセキュリティ脆弱性への留意	○		○				○	○
⑥プライバシーの原則	ア　最終利用者及び第三者のプライバシーの尊重		○	○	○		○	○	
	イ　パーソナルデータの収集・前処理・提供等におけるプライバシーの尊重	○		○			○		○
	ウ　自己等のプライバシー侵害への留意及びパーソナルデータ流出の防止		○					○	
⑦尊厳・自律の原則	ア　他者の尊厳と自律の尊重			○	○		○	○	
	イ　AIによる意思決定・感情の操作等への留意			○	○		○	○	
	ウ　AI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		○	○	○		○	○	
	エ　AIを利用したプロファイリングを行う場合における不利益への配慮	○	○	○	○		○	○	
⑧公平性の原則	ア　AIの学習等に用いられるデータの代表性への留意	○	○	○	○		○	○	○
	イ　学習アルゴリズムによるバイアスへの留意	○	○	○	○		○	○	○
	ウ　人間の判断の介在（公平性の確保）	○	○	○	○				
⑨透明性の原則	ア　AIの入出力等のログの記録・保存		○	○	○				
	イ　説明可能性の確保	○							
	ウ　行政機関が利用する際の透明性の確保	○	○	○	○				
⑩アカウンタビリティの原則	ア　アカウンタビリティを果たす努力	○	○	○	○		○	○	○
	イ　AIに関する利用方針の通知・公表	○	○	○	○		○	○	○

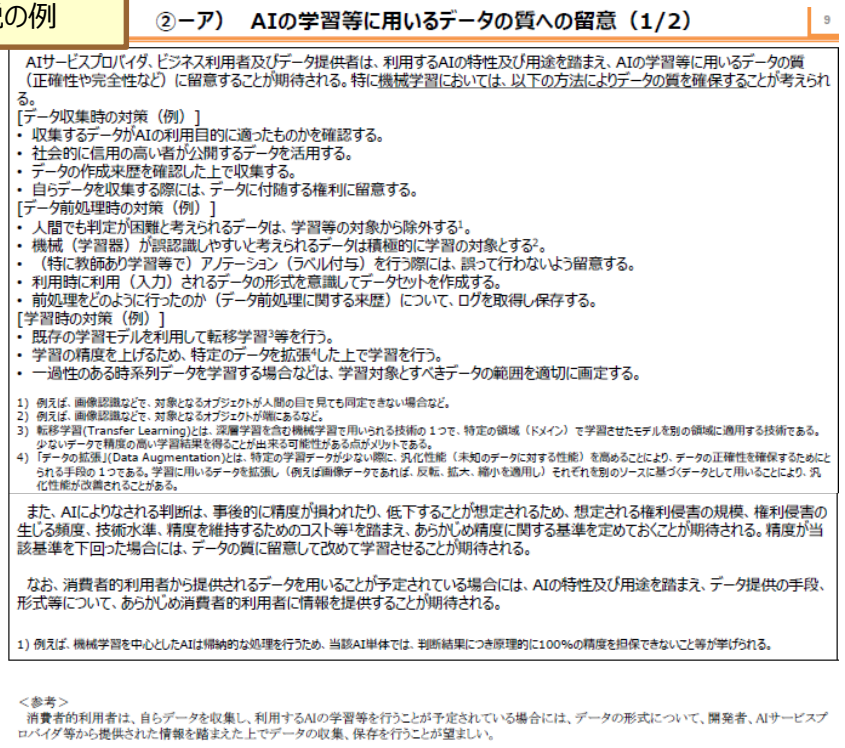
※　上記の表は、AIサービスプロバイダ、ビジネス利用者等については、自らAIの運用等を行う場合を想定し、また、消費者的利用者については、自ら運用等を行わない場合を想定して作成

第２章　AI利活用ガイドライン（各論点の詳説）

詳説の整理に当たっての考え方

- 各原則の各論点に対する詳説を、以下の両者の立場で留意すべき事項として整理し、併記（左下図参照）：
  - AIサービスプロバイダ、ビジネス利用者及びデータ提供者
  - 消費者的利用者
- 定期的に見直すことを前提に、近年AI技術として利活用が進められている「機械学習」の概念図を用いて解説（右下図参照）

詳説の例



②ーア）　AIの学習等に用いるデータの質への留意（1/2）　9

AIサービスプロバイダ、ビジネス利用者及びデータ提供者は、利用するAIの特性及び用途を踏まえ、AIの学習等に用いるデータの質（正確性や完全性など）に留意することが期待される。特に機械学習においては、以下の方法によりデータの質を確保することが考えられる。

[データ収集時の対策（例）]
- 収集するデータがAIの利用目的に適ったものかを確認する。
- 社会的に信用の高い者が公開するデータを活用する。
- データの作成来歴を確認した上で収集する。
- 自らデータを収集する際には、データに付随する権利に留意する。

[データ前処理時の対策（例）]
- 人間でも判定が困難と考えられるデータは、学習等の対象から除外する[1]。
- 機械（学習器）が誤認識しやすいと考えられるデータは積極的に学習の対象とする[2]。
- （特に教師あり学習等で）アノテーション（ラベル付与）を行う際には、誤って行わないよう留意する。
- 利用時に利用（入力）されるデータの形式を意識してデータセットを作成する。
- 前処理をどのように行ったのか（データ前処理に関する来歴）について、ログを取得し保存する。

[学習時の対策（例）]
- 既存の学習モデルを利用して転移学習[3]等を行う。
- 学習の精度を上げるため、特定のデータを拡張[4]した上で学習を行う。
- 一過性のある時系列データを学習する場合などは、学習対象とすべきデータの範囲を適切に画定する。

1）例えば、画像認識などで、対象となるオブジェクトが人間の目で見ても同定できない場合など。
2）例えば、画像認識などで、対象となるオブジェクトが端にあるなど。
3）転移学習(Transfer Learning)とは、深層学習を含む機械学習で用いられる技術の１つで、特定の領域（ドメイン）で学習させたモデルを別の領域に適用する技術である。少ないデータで精度の高い学習結果を得ることが出来る可能性がある点がメリットである。
4）「データの拡張」(Data Augmentation)とは、特定の学習データが少ない際に、汎化性能（未知のデータに対する性能）を高めることにより、データの正確性を確保するためにとられる手段の１つである。学習に用いるデータを拡張し（例えば画像データであれば、反転、拡大、縮小を適用し）それぞれを別のソースに基づくデータとして用いることにより、汎化性能が改善されることがある。

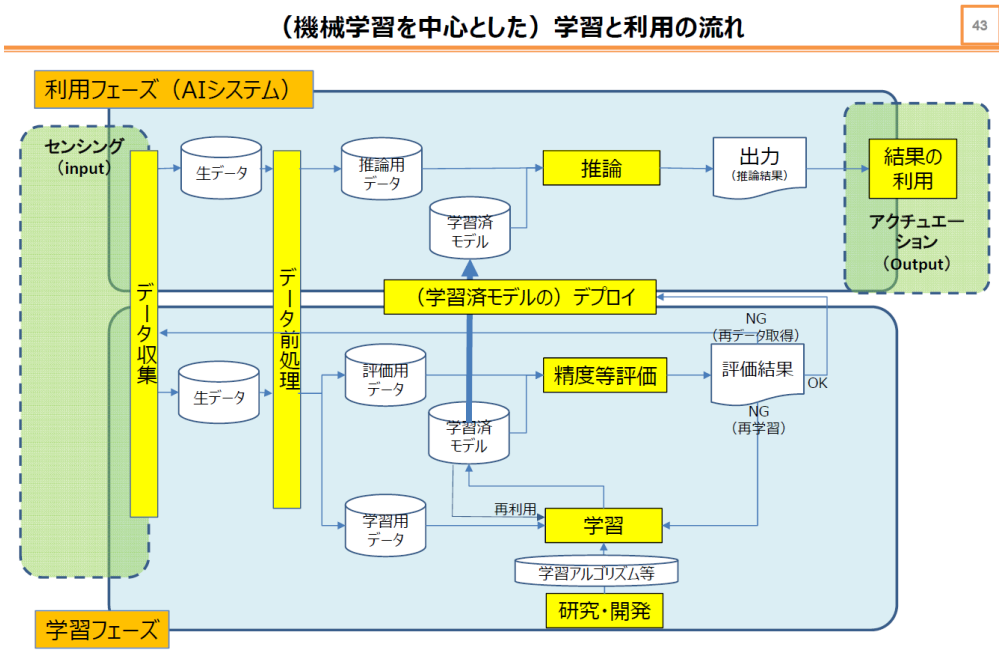
また、AIによりなされる判断は、事後的に精度が損われたり、低下することが想定されるため、想定される権利侵害の規模、権利侵害の生じる頻度、技術水準、精度を維持するためのコスト等[1]を踏まえ、あらかじめ精度に関する基準を定めておくことが期待される。精度が当該基準を下回った場合には、データの質に留意して改めて学習させることが期待される。

なお、消費者的利用者から提供されるデータを用いることが予定されている場合には、AIの特性及び用途を踏まえ、データ提供の手段、形式等について、あらかじめ消費者的利用者に情報を提供することが期待される。

1）例えば、機械学習を中心としたAIは帰納的な処理を行うため、当該AI単体では、判断結果につき原理的に100%の精度を担保できないこと等が挙げられる。

＜参考＞
消費者的利用者は、自らデータを収集し、利用するAIの学習等を行うことが予定されている場合には、データの形式について、開発者、AIサービスプロバイダ等から提供された情報を踏まえた上でデータの収集、保存を行うことが望ましい。

詳説の例（AIサービスプロバイダ、ビジネス利用者及びデータ提供者に対する留意事項を中心に、参考として消費者的利用者に対する留意事項を記載）



流れ（フロー）に関する図解例

第３章　今後の課題

	課題	概要
1. AIネットワーク化の健全な進展に関する事項		
(1)	AI開発ガイドライン及びAI利活用ガイドラインの周知・展開	AI開発ガイドライン／AI利活用ガイドラインの周知のためのシンポジウムの開催、国際的な枠組みにおける原則を実現するための詳説の周知等
(2)	AIの開発及び利活用に関する原則・ガイドラインについての議論のフォローアップ	AI開発／利活用原則・ガイドラインに関する国際的な議論のフォローアップと継続的な見直し
(3)	関係するステークホルダが取り組む環境整備に関する課題	ステークホルダ間の協力・ベストプラクティスの共有、法制度等の在り方の検討等
(4)	AIシステム又はAIサービス相互間の円滑な連携の確保	関係ステークホルダ間で共有することが期待される関連情報の範囲等の検討
(5)	競争的なエコシステムの確保	関連する市場の動向の継続的注視
(6)	利用者の利益の保護	利用者に対する開発者等からの自発的な情報提供の在り方の検討、利用者を保護する仕組み（保険等）の在り方の検討等
2. AIネットワーク化が社会・経済にもたらす影響の評価に関する事項		
(1)	AIネットワーク化が社会・経済にもたらす影響に関するシナリオ分析	シナリオ分析の継続的な実施・国際的な共有等
(2)	AIネットワーク化の進展に伴う影響の評価指標及び豊かさや幸せに関する評価指標の設定	指標の設定に向けた検討
(3)	AIシステムの利活用に関する社会的受容性の醸成	社会におけるAIの利活用に関する受容度の継続的注視等
3. AIネットワーク化が進展する社会における人間をめぐる課題に関する事項		
(1)	人間とAIとの関係の在り方に関する検討	専門職（医師、弁護士、会計士等）とAIシステムとの役割分担の在り方等の検討
(2)	ステークホルダ間の関係の在り方に関する検討	AIのリスクが顕在化した場合の責任の分配の在り等の検討
(3)	セーフティネットの整備	労働市場の動向の継続的注視、AIネットワーク化の進展に伴う所得の再分配等格差防止の在り方の検討等