ＡＩ利活用ガイドライン

令和２年１月３１日

総務省情報通信政策研究所

主任研究官　高木　幸一

最近よく見る※ AIとは

1．基本的には「コンピュータ」（もしくは、コンピュータを使った技術）

- そもそも、「知性」や「知能」自体の定義が困難であることから、「人工知能」（AI）の定義も困難であるが、AIは基本的にはコンピュータである
- 計算式による統計的な分析などを基に判断を下す

2．エンジン（アルゴリズム／プログラム）及びデータにより成り立つ

- 「AIの開発」とは、良質なデータの収集・適用及びより効率的なエンジンの開発と同義である

3．学習により、自らの出力等を変化させる

- 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エンジンを変化させる機能を有する

4．自らは目的を持たない

- ex. 幸福になりたい、安全でいたい、等

※過去のAI及び将来のAIについて、上記と異なる性質を持つ可能性もあり得る

AIが得意なこと

１．ルール及びゴールが明確な作業

- 例：囲碁、将棋、チェスなどのボードゲーム
  cf. アルファ碁（AlphaGO）

２．データから導出される情報に基づく作業

- 医療画像データによる診断（の補助）
  （CT、MRI等の画像を元にした癌の検出など）
- 顔認証、監視カメラ画像を元にした保安
- 製造ラインやプラント等における異常検出
- 音声認識・翻訳

〔参考〕元々コンピュータが得意なこと

- 計算
- データ蓄積
- 検索

AIがあまり得意ではないこと

１．創造的な作業、未知の事例への対応

- AIのエンジンは（利用する前の過去の）データから作られているため、そこから何かを創造することはそれほど得意ではない。新しい事例への対応も同様。

２．完璧な作業

- AIのエンジンはデータから作られているため、（ルールに即して動作するエンジンと比較すると）意図しない動作をすることもあり得る。
- 作業を完璧にさせる（意図した動作をさせる）ために冗長構成を取ることなどが考えられるが、そのためにはコストを要する。

３．カスタマイズ（個別の事例への対応）

- AIのエンジンは大量かつ多様なデータから作られることが多いので、（多くのケースに）共通の特徴を持つエンジンは比較的容易に作成できるが、特定の個人や個別の特徴に対応したエンジンの作成は容易ではない（相応の量のデータが必要となることが多い）。
- 一方、少量のデータから精巧なエンジンを作り出すための技術の検討も行われている。

AIに対し・・・

○ AIの研究開発が進み、様々な分野においてAIの利活用が進展することが想定される。

○ その際、AIネットワーク化※の進展が想定される。

※ 「AIネットワーク化」とは、AIシステムがインターネットその他の情報通信ネットワークと接続され、AIシステム相互間又はAIシステムと他の種類のシステムとの間のネットワークが形成されるようになること。

多大な便益への期待 ＋ リスクへの懸念

便益の例	リスクの例
高齢者や過疎地などで公共交通機関の利用が困難な者が、自動運転車を使うと、病院への通院や買い物などに出掛け易くなる。	自動運転車がユーザの意図に反し特定の場所を通過するようルートを設定するなどのおそれがある。
過去の症例を参考に、AIが患者の病名を推定するとともに、適切な処置方法を提示する。	AIが不適切な推定を行ったり、不適切な処置方法を提示したりすることより、医療過誤が生ずるおそれがある。
道路や橋などに設置されたセンサーや衛星写真から得られる情報等から、AIが異常検知や故障予測を行う。異常を検知した場合には、ロボットが自動的に点検・修理を行う。	AIが異常を見逃すこと等により、道路の陥没や橋の崩落など事故が発生するおそれがある。

○ AIは様々な分野で利活用され、また、そのサービスはネットワークを通じて（場合によっては国境を越えて）提供されることが想定される。

- AIは人間・社会に多大な便益を広範にもたらすことが期待される一方、リスクの抑制も今から考えておくことが必要ではないか？
- イノベーション促進を図りつつ、AIを安全・安心に社会実装していくためにはどのような手法が有効か？

総務省 「AIネットワーク社会推進会議」概要

背景

○ AIの研究開発・利活用の進展、AIの相互連携・ネットワークの形成（AIネットワーク化）

○ 様々な分野におけるAI利活用、ネットワークを通じた（国境を越えた）サービス提供

○ 多大な便益を広範にもたらすことが期待されるとともに、リスクの抑制も図ることが重要

⇒
- AIの便益の増進、リスクの抑制のための取組について**中長期的な視点**で検討が必要
- **産学民官の幅広い関係者の参画を得て**、国際的にも議論することが重要

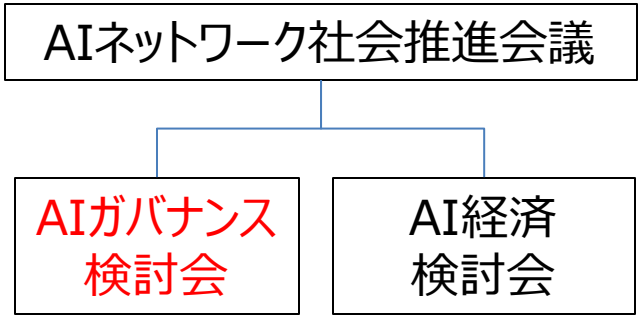
AIネットワーク社会推進会議

目的・検討事項

AIネットワーク化に関して、社会的・経済的・倫理的・法的課題に関する事項を検討。具体的には、以下について検討。

○ AI開発ガイドライン・AI利活用ガイドライン

○ AIに関する経済政策　　等



検討体制

【議　長】　須藤修（東京大学大学院情報学環教授・東京大学総合教育研究センター長）

【構成員】　産学民の有識者（関係学会の会長経験者、関係企業の会長又は社長等）

【オブザーバ】関係行政機関、関係国立研究開発法人 等

AI利活用ガイドラインとは

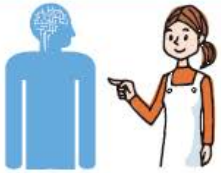
Q	どのようなものか？
A	AIの利用者（AIを利用してサービスを提供する者を含む）が利活用段階において留意することが期待される事項を「原則」（全10原則）という形式でまとめ、その解説を記載したもの。
Q	なぜ作ったのか？
A	AIによる便益の増進とリスクの抑制を図り、AIに対する信頼を醸成することにより、AIの利活用や社会実装を促進するため。 政府全体で検討・決定した「人間中心のAI社会原則」において、「開発者・事業者それぞれにおいて、AI開発利用原則を策定することを期待」とある中で、事業者向けの解説書として。 我が国の考えを世界に共有するため（国際的な議論に資するため）。
Q	誰に対して作ったのか？
A	（事業者を中心に）AIを利用して事業を行う者。 消費者等の最終利用者に対しては「参考」として。
Q	どのような体制で作ったのか？
A	産業界、学術界、市民団体等によるマルチステークホルダによる会議で議論 （関係省庁もオブザーバ）。

AI利活用原則

AI利活用原則

出典:総務省情報通信政策研究所「AIネットワーク社会推進会議報告書2019 別紙1:AI利活用ガイドライン~AI利活用のためのプラクティカルガイダンス」より
http://www.soumu.go.jp/menu_news/s-news/01iicp01_02000081.html

適正利用の原則



適正な範囲及び方法でAIを利用

適正学習の原則

AIの学習等に用いるデータの質に留意

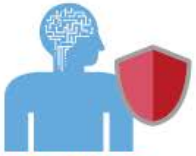
連携の原則

・AI相互間の連携に留意
・AI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

安全の原則

生命・身体・財産に危害を及ぼすことがないよう配慮

セキュリティの原則



AIのセキュリティに留意

プライバシーの原則

他者または自己のプライバシーが侵害されないよう配慮

尊厳・自律の原則

人間の尊厳と個人の自律を尊重

公平性の原則

AIの判断にバイアスが含まれる可能性があることに留意

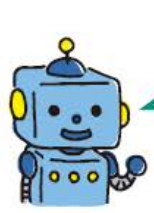
透明性の原則

AI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

アカウンタビリティの原則

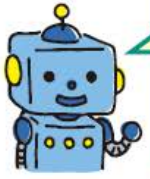
アカウンタビリティ※を果たすよう努力

※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、責任者を明確にしたうえで、判断に関する正当な意味・理由の説明、必要に応じた賠償・補償等の措置がとられること



そんなに窮屈に考える必要はありません。「守らないと…」と考えるより「守ったら得になる」と考えるのはどうでしょうか。提供する方はお客さまに安心してもらえる。使う方は自分が守られて安心につながる。そうやって利活用が進むとうれしいです。

「ガイドライン」ってことは、守らないと……。



いいえ、用途によって気にした方がいい原則やその重みも変わってくると思います。例えば前のページに出てきた「お掃除ロボット」や「自動運転」などでは「安全性」等が重視されるでしょうし、スマートスピーカーとの会話では「公平性」等が重視されるのではないでしょうか。このような形で、自分が携わっている事業ではどうか、自分の利用シーンではどうかを一緒に考えていきましょう。

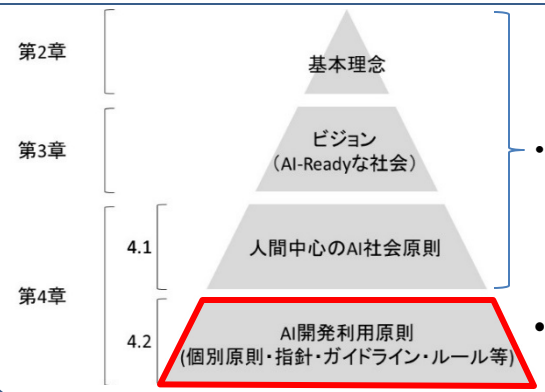
これ全部気にしないといけないの?

<引用>総務省広報誌 令和元年12月号 https://www.soumu.go.jp/main_content/000656829.pdf

「AI利活用ガイドライン」の位置づけ

「人間中心のAI社会原則」（2019年３月統合イノベーション戦略推進会議決定）より抜粋

人間中心のAI社会原則会議
（2018年５月～）



・・・「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

そのための参考となるガイドラインが必要

AIネットワーク社会推進会議
（2016年２月～）

総務省の取組

AI開発ガイドライン
開発者が留意すべき事項と解説

2017年７月とりまとめ

AI利活用ガイドライン
事業者が留意すべき事項と解説

2019年８月とりまとめ

関係省庁に共有の上、開発者・事業者に提供。自主的対応を支援。

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により自らの権利・利益に影響を受ける者

AIを利用して事業を行う者

「AI利活用ガイドライン」は利用者の中でもこちらの者が留意することが期待される事項を中心に記載

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に対し、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。

こちらの者へのメッセージは「参考」の位置づけ

（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。

AI利活用ガイドラインとりまとめの経緯

課題の抽出

- ケーススタディ等
- 課題の整理

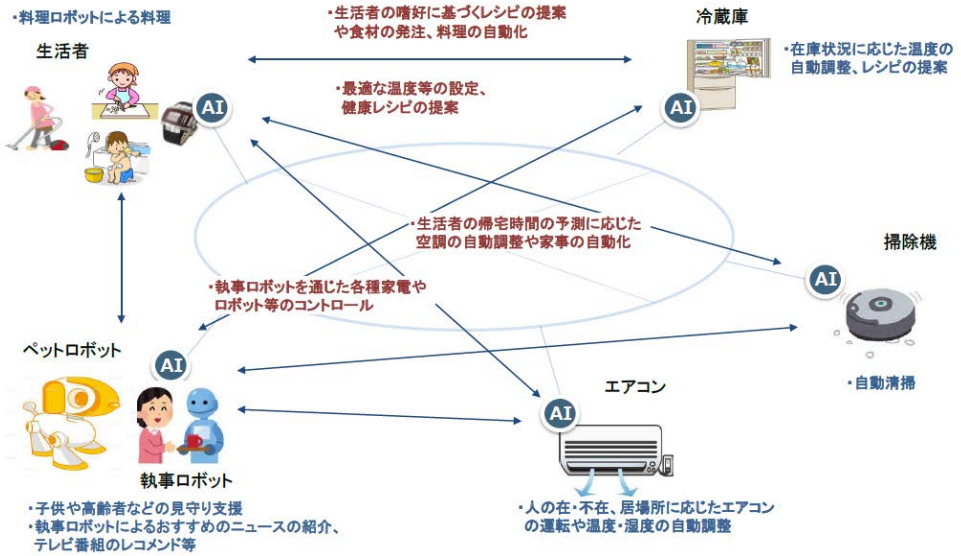
原則化

- 原則化（AI利活用原則案）
- 原則ごとの論点の抽出

具体化

- 論点ごとのヒアリング（構成員以外を中心に）
- 論点ごとの詳説作成

ケーススタディ例（家庭内での生活）



＜引用＞AIネットワーク社会推進会議報告書2018 別紙１　を元に加工

想定される利活用	想定される便益	想定されるリスク
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	・快適な空間で生活可能に。 ・電力消費の削減やピークカット/ピークシフトが実現。	・在宅・不在や生活習慣等に関する情報が明らかになりプライバシーが侵害されるおそれ。**[プライバシー]**
子供や高齢者などの見守り支援	・安心して外出可能に。 ・就労・地域活動等への参加が容易に。	・ハッキング等により、ペットロボットの制御が失われ（暴走し）、子供等が怪我をしたり、家具が壊れるおそれ。**[安全]**
執事ロボットによるおすすめのニュースの紹介、テレビ番組のレコメンド等	・手が離せない状況などにおいてニュース等の検索が容易に。 ・見たいテレビ番組を視聴可能。	・生活者の趣味・趣向を誤って判断し、必要な情報・関心の高い情報を提供できないおそれ。　**[正当性・公平性、セキュリティ]**
生活者の帰宅時間の予測に応じた空調の自動調整や家事の自動化	・より快適な空間で生活可能に。 ・家事の負担を軽減可能に。	・料理ロボットや掃除ロボットやペットロボットなど異なるロボット間の連携・調整が十分でなく、ロボット同士が衝突したり、破損するおそれ。**[連携、安全]** ・家事をAIに依存しすぎると、災害発生等でAIが利用できなくなった場合、生活が困難になるおそれ。**[役割分担、連携]**
生活者の嗜好に基づくレシピの提案や食材の発注、料理の自動化	・家事の負担を軽減。 ・嗜好に合わせた料理や健康に良い料理を（容易に）選択可能に。	
執事ロボットを通じた各種家電やロボット等のコントロール	・執事ロボットだけで家庭内の複数のロボットや家電等をコントロール可能に。	

課題の整理→原則化

【主として生命・身体の安全、権利・利益等を守るための課題】

○ 生命・身体・財産の安全に関する課題（事故の防止など） ⇒ ①適正利用の原則、④安全の原則

→ どのように事故が発生しないようにするか、また、事故が生じた場合にどのように対応すべきか（責任の在り方を含む。）について検討が必要ではないか。

○ AIによる判断の正当性や公平性に関する課題（差別、生命倫理との関係など）⇒ ②適正学習の原則、⑦尊厳・自律の原則、⑧公平性の原則

→ どのようにAIによる判断の正当性や公平性を確保し差別的な取扱いがなされないようにするか、データの適正性・正確性や人間の介在の在り方を含めて検討が必要ではないか。

○ プライバシーに関する課題（プライバシーの尊重、プロファイリングなど）⇒ ⑥プライバシーの原則

→ どのようにプライバシーを尊重するのか、本人同意の在り方やプロファイリングの在り方などを含めて検討が必要ではないか。

【主として人間とAIとの関係等に関する課題】

○ 人間とAIとの役割分担等に関する課題（人間の判断の介在、関係者間の協力など）⇒ ①適正利用の原則、③連携の原則

→ どのような場合に人間の判断を介在させるべきか、その介在の要否の基準を含めて検討が必要ではないか。また、
安心して安全にAIを利活用するために、どのように関係者が協力すべきかについて検討が必要ではないか。

○ AIに対する受容性に関する課題（利用者に対する説明責任など） ⇒ ⑩アカウンタビリティの原則

→ どのように利用者・社会のAIの信頼性を醸成すべきかについて検討が必要ではないか。

【主として技術的な観点からの解決が求められる課題】

○ AIの判断のブラックボックス化に関する課題（事故が発生した場合の原因究明など）⇒ ⑨透明性の原則

→ どのような場合に、どの程度AIの判断の根拠・理由を明らかにすべきかについて検討が必要ではないか。

○ セキュリティに関する課題（ハッキング対策など） ⇒ ⑤セキュリティの原則

→ どのようにセキュリティを確保すべきかについて検討が必要ではないか。

○ AI間の連携に関する課題（AI間の交渉・調整など） ⇒ ③連携の原則

→ どのようにAI間の円滑な交渉・調整を実現するか、データ形式やプロトコル等の観点も含めて検討が必要ではないか。

【主としてデータに関する課題】

○ AIが学習するデータに関する課題（データの正確性など）⇒ ②適正学習の原則

→ どのようにデータの適正性・正確性を担保するか、また、どのように適切なデータを確保するかについて検討が必要ではないか。

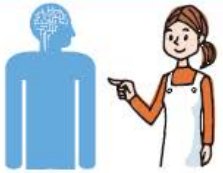
＜引用＞AIネットワーク社会推進会議報告書2018 別紙２－１ AIの利活用の展望 を元に加工

AI利活用原則（再掲）

AI利活用原則

出典：総務省情報通信政策研究所「AIネットワーク社会推進会議報告書2019 別紙1：AI利活用ガイドライン～AI利活用のためのプラクティカルガイダンス」より
http://www.soumu.go.jp/menu_news/s-news/01iicp01_02000081.html

適正利用の原則



適正な範囲及び方法でAIを利用

適正学習の原則

AIの学習等に用いるデータの質に留意

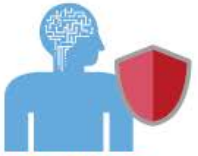
連携の原則

・AI相互間の連携に留意
・AI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

安全の原則

生命・身体・財産に危害を及ぼすことがないよう配慮

セキュリティの原則



AIのセキュリティに留意

プライバシーの原則

他者または自己のプライバシーが侵害されないよう配慮

尊厳・自律の原則

人間の尊厳と個人の自律を尊重

公平性の原則

AIの判断にバイアスが含まれる可能性があることに留意

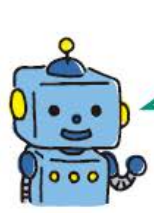
透明性の原則

AI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

アカウンタビリティの原則

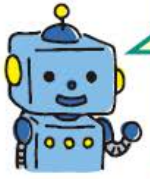
アカウンタビリティ※を果たすよう努力

※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、責任者を明確にしたうえで、判断に関する正当な意味・理由の説明、必要に応じた賠償・補償等の措置がとられること



そんなに窮屈に考える必要はありません。「守らないと…」と考えるより「守ったら得になる」と考えるのはどうでしょうか。提供する方はお客さまに安心してもらえる。使う方は自分が守られて安心につながる。そうやって利活用が進むとうれしいです。

「ガイドライン」ってことは、守らないと……。



いいえ、用途によって気にした方がいい原則やその重みも変わってくると思います。例えば前のページに出てきた「お掃除ロボット」や「自動運転」などでは「安全性」等が重視されるでしょうし、スマートスピーカーとの会話では「公平性」等が重視されるのではないでしょうか。このような形で、自分が携わっている事業ではどうか、自分の利用シーンではどうかを一緒に考えていきましょう。

これ全部気にしないといけないの？

<引用>総務省広報誌 令和元年12月号 https://www.soumu.go.jp/main_content/000656829.pdf

原則ごとの論点と消費者的利用者に望まれること（例）

原則	原則に対する論点
①適正利用	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	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	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	ア　人の生命・身体・財産への配慮
⑤セキュリティ	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	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	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	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	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	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

最終判断をすることが適当とされる場合には、適切に判断ができるよう必要な能力及び知識を習得しておくこと

AIが不正確または不適切なデータを学習することにより脆弱性が生じるリスクに留意すること

事業者等※の情報をもとに、必要に応じて点検・アップデート及びセキュリティ対策を行うこと

過度に感情移入すること等により、特に秘匿性の高い情報をむやみにAIに与えることのないように

自らの情報が正しく利用されているかを意識し、必要に応じ事業者等※に確認すること

AIの判断結果に疑義を生じた場合には、必要に応じ事業者等※に問い合わせること

※AIサービスプロバイダ、ビジネス利用者をまとめて「事業者等」と記載。

「消費者的利用者向けのハンドブックが欲しい」「具体化、明確化が必要」との声。

まとめ と 本WGへの期待

「AI利活用ガイドライン」は

- AIの利活用や社会実装を促進するため、AIの利用者に参照して頂くためのものとして作成。
- 特にAIを利用して事業を行う者が留意することが期待される事項を中心に記載。
- 消費者的利用者向けには「望まれる事項」を記載しているが、 あくまで「参考」の形。具体化・明確化を期待する声あり。

皆様と共に、消費者的利用者に「望まれる事項」を、わかりやすく、かつ納得していただける形にしていきたい。

- AIの利用にあたり
  消費者的利用者にどう考えてほしいか。
  消費者的利用者に当たり前に思っていただきたいことは何か。

以下参考

総務省における主な取組と統合イノベーション戦略推進会議における位置づけ

2016

G7情報通信大臣会合（高松、2016年4月）
AIの開発に着目したルール作りに向け、国際的な議論を進めるよう提案（我が国から原則のたたき台を紹介）

2017

「AI開発ガイドライン」策定
（2017年7月公表）

OECD、G7等における検討の場に日本からも有識者を派遣
（例：OECD/総務省共催AIカンファレンス（パリ、2017年10月）、
OECD理事会勧告に向けたAI専門家会合（2018年9月～2019年2月）
G7マルチステークホルダ会合（モントリオール、2018年12月）　等）

2018

統合イノベーション戦略推進会議
＜議長：官房長官＞

人間中心のAI社会原則（2019年3月）

AI戦略 2019 ～人・産業・地域・政府全てにAI～
（2019年6月）

2019

OECD閣僚理事会（2019年5月）
AIに関する理事会勧告を採択

G20貿易・デジタル経済大臣会合
（つくば、2019年6月）
「G20AI原則」を採択

「AI利活用ガイドライン」策定
（2019年8月公表）

AI中核センター改革・
次世代AI基盤技術推進　等

国際的な議論への貢献（OECD等）

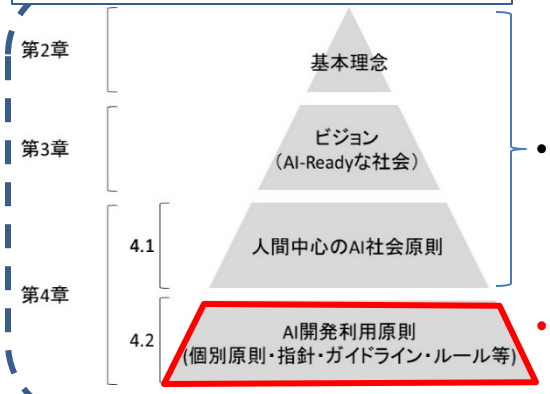
日本型モデル構築・創発研究の推進

今後は、目指すべき社会モデルを検討し、社会実装を加速

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年にとりまとめる予定
- 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
  →CDEP（デジタル経済政策委員会）会合で検討中

「AIネットワーク社会推進会議」（親会）構成員

2019年７月現在

議　長　須藤 修　（東京大学大学院情報学環教授・東京大学総合教育研究センター長）

副議長　三友 仁志　（早稲田大学国際学術院大学院アジア太平洋研究科教授）

構成員

【研究者（社会・人文系）】

大橋 弘　（東京大学大学院公共政策大学院・経済学研究科教授）
大屋 雄裕　（慶應義塾大学法学部教授）
小塚 荘一郎　（学習院大学法学部法学科教授）
宍戸 常寿　（東京大学大学院法学政治学研究科教授）
実積 寿也　（中央大学総合政策学部教授）
城山 英明　（東京大学大学院法学政治学研究科教授）
新保 史生　（慶應義塾大学総合政策学部教授）
鈴木 晶子　（京都大学大学院教育学研究科教授）
橋元 良明　（東京大学大学院情報学環教授）
林 秀弥　（名古屋大学大学院法学研究科教授）
平野 晋　（中央大学国際情報学部教授・学部長）
福田 雅樹　（大阪大学大学院法学研究科教授）
柳川 範之　（東京大学大学院経済学研究科教授）
山本 勲　（慶應義塾大学商学部教授）

【研究者（技術系）】

大田 佳宏　（東京大学大学院数理科学研究科特任教授、Arithmer代表取締役社長兼CEO）
喜連川 優　（国立情報学研究所所長、東京大学生産技術研究所教授）
杉山 将　（理化学研究所革新知能統合研究センター長、東京大学新領域創成科学研究科教授）
高橋 恒一　（理化学研究所生命機能科学研究センターチームリーダー）
中川 裕志　（理化学研究所革新知能統合研究センターグループディレクター）
中西 崇文　（武蔵野大学データサイエンス学部データサイエンス学科長・准教授）
西田 豊明　（京都大学大学院情報学研究科教授）
萩田 紀博　（大阪芸術大学アートサイエンス学科長・教授、株式会社国際電気通信基礎技術研究所招聘研究員）
堀 浩一　（東京大学大学院工学系研究科教授）
松尾 豊　（東京大学大学院工学系研究科教授）
村井 純　（慶應義塾大学環境情報学部教授）
森川 博之　（東京大学大学院工学系研究科教授）
山川 宏　（全脳アーキテクチャ・イニシアティブ代表）

【産業界】

岩本 敏男　（株式会社エヌ・ティ・ティ・データ相談役）
遠藤 信博　（日本電気株式会社取締役会長）
金井 良太　（株式会社アラヤ代表取締役ＣＥＯ）
北野 宏明　（株式会社ソニーコンピュータサイエンス研究所代表取締役社長）
エリー キーナン　（日本アイ・ビー・エム株式会社取締役会長）
谷崎 勝教　（株式会社三井住友銀行取締役専務執行役員グループＣＤＩＯ）

時田 隆仁　（富士通株式会社代表取締役社長）
東原 敏昭　（株式会社日立製作所代表執行役 執行役社長兼ＣＥＯ）
平野 拓也　（日本マイクロソフト株式会社代表取締役社長）
杉原 佳堯　（グーグル合同会社執行役員 公共政策担当）
村上 憲郎　（株式会社村上憲郎事務所代表取締役）

【消費者団体】

木村 たま代　（主婦連合会事務局長）
近藤 則子　（老テク研究会事務局長）

長田 三紀　（情報通信消費者ネットワーク）

顧　問　安西 祐一郎　（慶應義塾大学名誉教授）
長尾 真　（京都大学名誉教授）
西尾 章治郎　（大阪大学総長）
濱田 純一　（東京大学名誉教授）　（敬称略。五十音順）

オブザーバー　内閣府、内閣官房情報通信技術（ＩＴ）総合戦略室、個人情報保護委員会事務局、消費者庁、文部科学省、経済産業省、情報通信研究機構、科学技術振興機構、理化学研究所、産業技術総合研究所

①ーイ） 人間の判断の介在

AIサービスプロバイダ及びビジネス利用者は、AIによりなされた判断について、必要かつ可能な場合には、その判断を用いるか否か、あるいは、どのように用いるか等に関し、人間の判断を介在させることが期待される。その場合、人間の判断の介在の要否については、例えば以下の基準を踏まえ、利用する分野やその用途等に応じて検討することが期待される。

［人間の判断の介在の要否について、基準として考えられる観点（例）］

- AIの判断に影響を受ける最終利用者等の権利・利益の性質及び最終利用者等の意向
- AIの判断の信頼性の程度（人間による判断の信頼性との優劣）
- 人間の判断に必要な時間的猶予
- 判断を行う利用者に期待される能力
- 判断対象の要保護性（例えば、人間による個別申請への対応か、AIによる大量申請への対応か等）

また、AIによりなされた判断について人間が最終判断をすることが適当とされている場合に、人間がAIと異なる判断をすることが期待できなくなることも想定されることから、説明可能性を有するAIから得られる説明を前提として、人間が判断すべき項目を事前に明確化しておくこと等により、人間の判断の実効性を確保することが期待される[1]。

また、アクチュエータ等を通じて稼働するAIの利活用において、一定の条件に該当することにより人間による稼働に移行することが予定されている場合には、移行前、移行中、移行後等の各状態における責任の所在があらかじめ明確化されている必要がある。また、AIサービスプロバイダは、移行条件、移行方法等を最終利用者に事前に説明し、必要な訓練を実施するなど、人間による稼働に移行した場合に問題が生じないための事前対策を講じることが期待される。

1)加えて、人間が確認するAIの判断の適正性を確保するため、他のAIを利用したダブルチェック、AIへの入力を摂動させることよるAI動作の確認などの措置を検討することが望ましい。

＜参考＞

消費者的利用者は、AIの判断に対し、消費者的利用者が最終判断をすることが適当とされている場合には、適切に判断ができるよう必要な能力及び知識を習得しておくことが望ましい。

また、開発者及びAIサービスプロバイダにより人間の判断の実効性を確保するための対応が整理されている場合は、それに基づき適切に対応することが望ましい。

また、アクチュエータ等を通じて稼働するAIの利活用において、一定の条件に該当することにより人間による稼働に移行することが予定されている場合には、消費者的利用者は、移行前、移行中、移行後等の各状態における責任の所在を予め認識しておくことが望ましい。また、AIサービスプロバイダから、移行条件、移行方法等についての説明を受け、必要な能力及び知識を習得しておくことが望ましい。

＜引用＞ AIネットワーク社会推進会議報告書2019 別紙１（附属資料）