

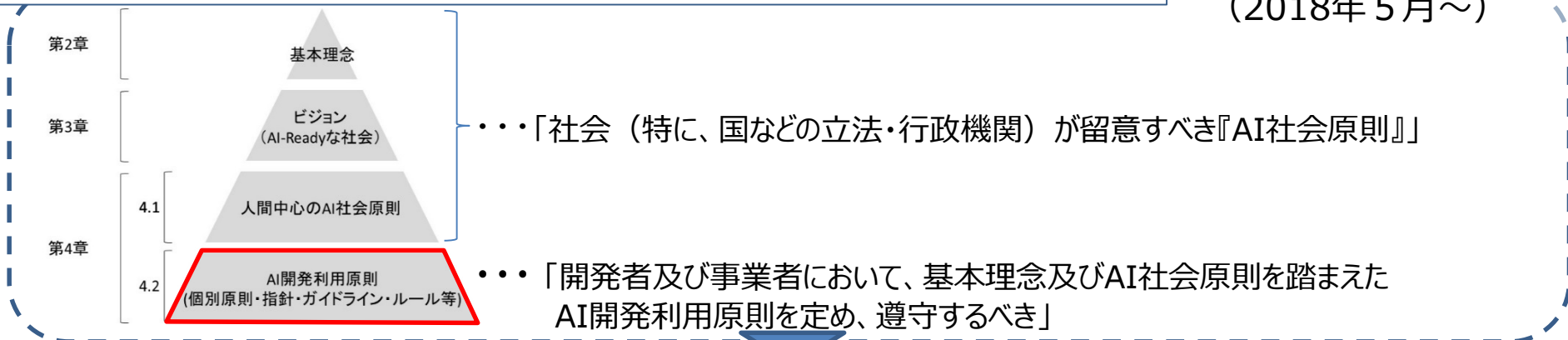
安心・安全で信頼性のあるA Iの社会実装に向けて

令和2年6月
令事務局

「人間中心のAI社会原則」と「AI利活用ガイドライン」の関係

「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～)



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

総務省の取組

AIネットワーク社会推進会議
(2016年2月～)

AI開発ガイドライン
開発者が留意すべき事項と解説

AI利活用ガイドライン
事業者が留意すべき事項と解説

2017年7月とりまとめ

2019年8月とりまとめ

国際的な議論への貢献(OECD等)

今後は、目指すべき社会モデルを検討し、社会実装を加速

民間企業によるAI原則策定が進展 →総務省検討内容は分野共通、企業による原則検討の参考に

民間企業により策定・公表されたAI原則等の例：

Microsoft AI principles (2018/2)

公平性、信頼性&安全性、
プライバシー、セキュリティ、
包摂性、透明性、
アカウントビリティ など

IBM's Principles for Trust and Transparency (2018/5/30)

アカウントビリティ、価値の整合、
説明可能性、解釈可能性、
公平性、ユーザデータの権利

AI at Google: our principles (2018/6)

社会的価値、公平性、安全性、
プライバシー&透明性、
アカウントビリティ など

ソニーグループ AI倫理ガイドライン (2018/9/25)

豊かな生活とよりよい社会、
ステークホルダとの対話、安心、
プライバシー保護、公平性尊重、
透明性、AI発展・人材育成

富士通グループ AIコミットメント (2019/3/13)

AIによる価値の提供、人間中心、
持続可能性、人の意思の尊重、
透明性と説明責任

NECグループ AIと人権に関するポリシー (2019/4/2)

公平性、プライバシー、透明性、
説明責任、適正利用、
AIの発展と人材育成、
マルチステークホルダとの対話

NTTデータグループ AI指針 (2019/5/29)

持続可能な幸福社会、
共創による新しいAI価値の創出、公正で信頼
できる説明可能なAI、安心安全なデータの流通、
AIを健全に普及させる活動の推進

OKIグループ AI原則 (2019/9/30)

人権の尊重、説明と透明性、
対話と協調、安全およびデータの取扱い、
人財育成

NRIグループ AI倫理ガイドライン(2019/10/8)

ステークホルダーとの対話・共創の推進、
AIの発展と人材の育成、
公平性の尊重、
安全性・セキュリティの担保、
データとプライバシーの保護、
透明性の確保

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」という形式でまとめ、その解説を記載

原則	
適正利用	適正な範囲及び方法でAIを利用
適正学習	AIの学習等に用いるデータの質に留意
連携	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	生命・身体・財産に危害を及ぼすことがないように配慮
セキュリティ	AIのセキュリティに留意
プライバシー	他者又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	人間の尊厳と個人の自律を尊重
公平性	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 個人及び集団が不当に差別されないよう配慮
透明性	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
アカウントビリティ	アカウントビリティを果たすよう努める

※ 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

AIの開発者が研究・開発段階において留意することが期待される事項を「原則」という形式でまとめ、その解説を記載

原則	
連携	AIシステムの相互接続性と相互運用性に留意
透明性	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
制御可能性	AIシステムの制御可能性に留意
安全	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に配慮
セキュリティ	AIシステムのセキュリティに留意
プライバシー	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
倫理	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
利用者支援	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
アカウントビリティ	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（説明責任）を果たすよう努める

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

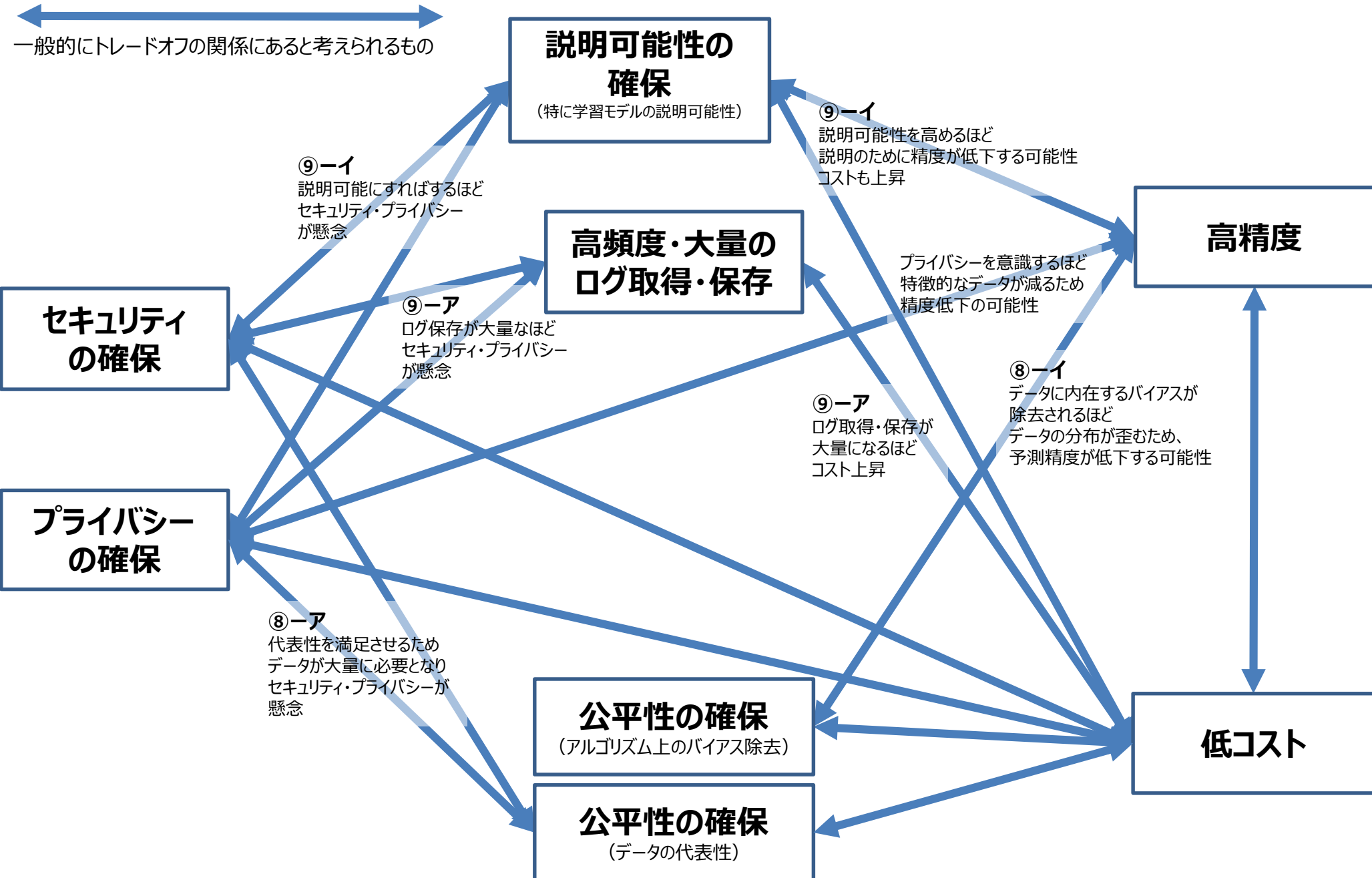
最終利用者のうちAIシステム又はAIサービスを利用する者（ビジネス利用者を除く）

（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。

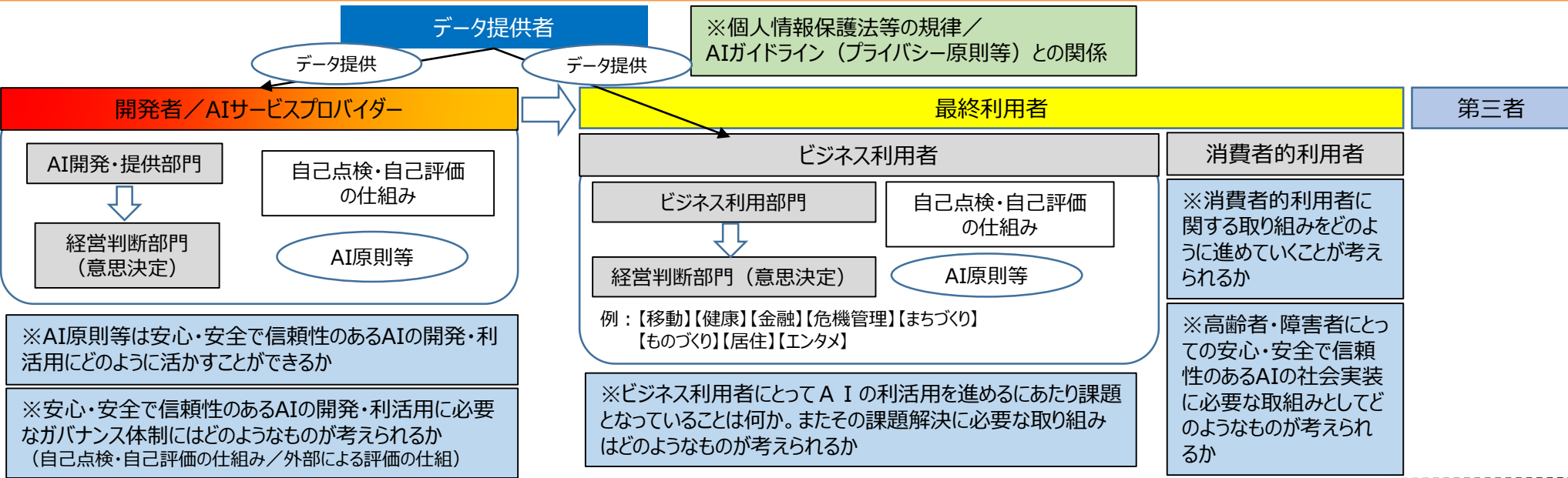
- ・AIシステム : AIソフトを構成要素として含むシステム
- ・AIサービス : AIシステムの機能を提供するサービス
- ・AI付随サービス : AIシステムのアップデート又は追加的な学習等に係るサービス

（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。

トレードオフの例



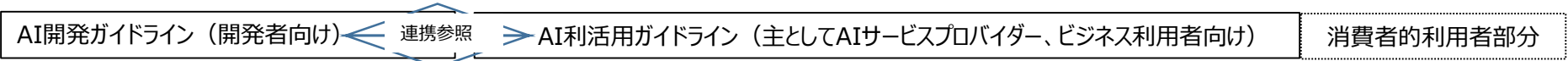
議長ヒアリング全体像～安心・安全で信頼性のあるAIの社会実装に向けて～（案）



安心・安全で信頼性のあるAIの社会実装のための環境整備等

- 将来像→AIの利活用の進展等を踏まえ、近い将来（2025年頃）、中期的な将来（2035年頃）を視野にAIの利活用の将来像としてどのようなことが考えられるか
- 技術面→AIと情報セキュリティ：開発・利活用等の各フェーズで何が課題でどのような取り組みが必要か
他、Certification、品質の確保等
- セーフティネット→保険の仕組みにはどのようなものが考えられるか
- 責任のあり方：契約責任（私的自治の原則）、不法行為責任（過失責任主義）

AIネットワーク社会推進会議におけるガイドラインの策定



海外における安心・安全で信頼性のある社会実装促進の環境整備としてのガイドラインの策定（例）

- OECD：プラクティカルガイダンス（オブザーバトリ（oecd.ai）に掲載・継続的に更新）
- EU：Trustworthy AI Assessment List（2019年12月までPilot Phaseを試行、2020年内を目処にとりまとめ予定）、AI白書
- 米国：Guidance for Regulation of Artificial Intelligence Applications（民間AI向け規制検討のベース、2020年3月まで意見募集）