資料３－１

「ＡＩ開発ガイドライン」（仮称）の策定に向けた国際的議論の用に供する素案の作成に関する論点整理

平成２８年１２月１５日

事　　務　　局

この資料の構成

○　論点の所在

○　論点の要旨

○　論点整理

論点の所在

論点の所在

第一　基本概念の定義

第二　ＡＩ開発ガイドラインの体系

第三　分野共通開発ガイドラインの構成

第四　分野共通開発ガイドラインの目的、基本理念等

第五　分野共通開発ガイドラインの適用範囲

第六　開発原則の構成及び順序

第七　開発原則の個々の項目の内容の具体化

第八　連携の原則【仮称】

第九　開発原則の実効性の確保の在り方

第十　開発原則の実効性の確保における市場の活用の在り方

第十一　ＡＩネットワークシステムの利活用に関し利用者等が留意すべき事項

論点の要旨

論点の要旨（1／8）

第一　基本概念の定義

○「開発者」とは、AIの研究開発（AIを研究し、又は開発する行為のほか、複数のAIを組み合わせて一体的なAIとして機能するよう構成する行為を含む。この資料において、「AIの研究開発」を「AIの開発」又は単に「開発」という場合がある。）をする者（自らが研究開発したAIを実装するAIネットワークシステムによるAIネットワークサービスのプロバイダを含む。）をいうものとしてはどうか。

○「利用者」とは、他の開発者が開発したAI（他の開発者が開発したAIを実装するAIネットワークシステムにより他のプロバイダが提供するAIネットワークサービスを含む。）の提供を受けてAIネットワークシステムを利用する者（自らが構築するAIネットワークシステムを自ら利用する個人又は団体（最終利用者のほか、AIネットワークサービスを他の者に提供するプロバイダを含む。）のほか、プロバイダからAIネットワークサービスの提供を受ける最終利用者をも含む。）をいうものとしてはどうか。

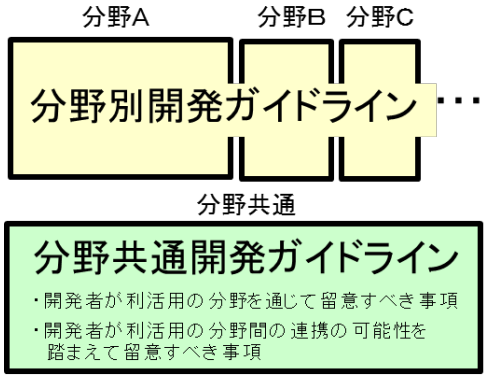
○ただし、「開発者」及び「利用者」は、関係者間の関係に即して、関係が生ずる場面ごとに個別に評価される相対的な概念。

第二　ＡＩ開発ガイドラインの体系

○開発ガイドラインの体系は、「分野共通開発ガイドライン」及び「分野別開発ガイドライン」からなるものとしてはどうか。

○「分野共通開発ガイドライン」は、本推進会議が検討と議論を推進することとしてはどうか。

○「分野別開発ガイドライン」は、その策定の要否も含め、分野ごとの国際機関を含む関係ステークホルダー自身による検討と議論に委ねることとしてはどうか。



第三　分野共通開発ガイドラインの構成

○分野共通開発ガイドライン　（OECDのガイドラインであれば、理事会勧告の附属文書（Annex））

AIネットワークシステムの構成要素となり得るAIの開発者がその研究開発に当たり、AIネットワーク化の健全な進展の促進並びにAIネットワークシステムの便益の増進及びリスクの抑制に関し、AIネットワークシステムの利活用の分野を通じて留意すべき事項及び利活用の分野間の連携の可能性を踏まえて留意すべき事項に関する原則（「開発原則」）並びにその説明

分野共通ガイドラインの構成（たたき台）

1. 目的
2. 基本理念（智連社会の構成要素、智連社会の基本理念等）
3. 定義（AI、AIネットワークシステム、AIネットワークサービス、開発者、利用者）
4. 適用範囲（開発者、AI、研究開発の段階）
5. 原則 → ○○の原則　××を△△すべきである。【説明】・・・・・・・・である。

○分野共通開発ガイドラインの関連文書　（OECDのガイドラインであれば、理事会勧告の本紙）

・分野共通開発ガイドラインに定める事項に関連し、国、関係国際機関等に推奨すべき事項

・ガイドラインの見直しの時期及び方法

論点の要旨（2／8）

第四　分野共通開発ガイドラインの目的、基本理念等

○分野共通開発ガイドラインの「目的」として、次に掲げる趣旨を、必要に応じて再構成した上で掲げることとしてはどうか。

> このガイドラインは、AIネットワークシステムの公共性に鑑み、AIネットワークシステムの構成要素となり得るAIの研究開発を行う者が、その研究開発に当たり、AIネットワーク化の健全な進展の促進並びにAIネットワークシステム（AIネットワークサービスを含む。以下同じ。）の便益の増進及びリスクの抑制に関し、AIネットワークシステムの利活用の分野を通じて又は分野間の連携の可能性を踏まえて留意すべき事項を開発原則として整理し、非拘束的な枠組みとして国際的に共有することにより、AIネットワークシステムの最終利用者の利益を保護するとともに第三者及び社会への波及的な悪影響を防止し、もって人間中心の智連社会の形成に資することを目的とする。（なお、「人間中心の」については、「人間が主体的にAIネットワークシステムを使いこなす」等の表現とすることも考えられる。）

○分野共通開発ガイドラインの策定及び解釈に当たっての「基本理念」として、①智連社会の構成要素、②智連社会の基本理念、③リスクへの適時適切な対処、④関係する価値・利益のバランスの確保及び⑤AIネットワーク化の進展及び関連するリスクの顕在化に応じた開発原則・開発ガイドラインの見直しの５点を、必要に応じて再構成した上で分野共通開発ガイドラインに掲げることとしてはどうか。

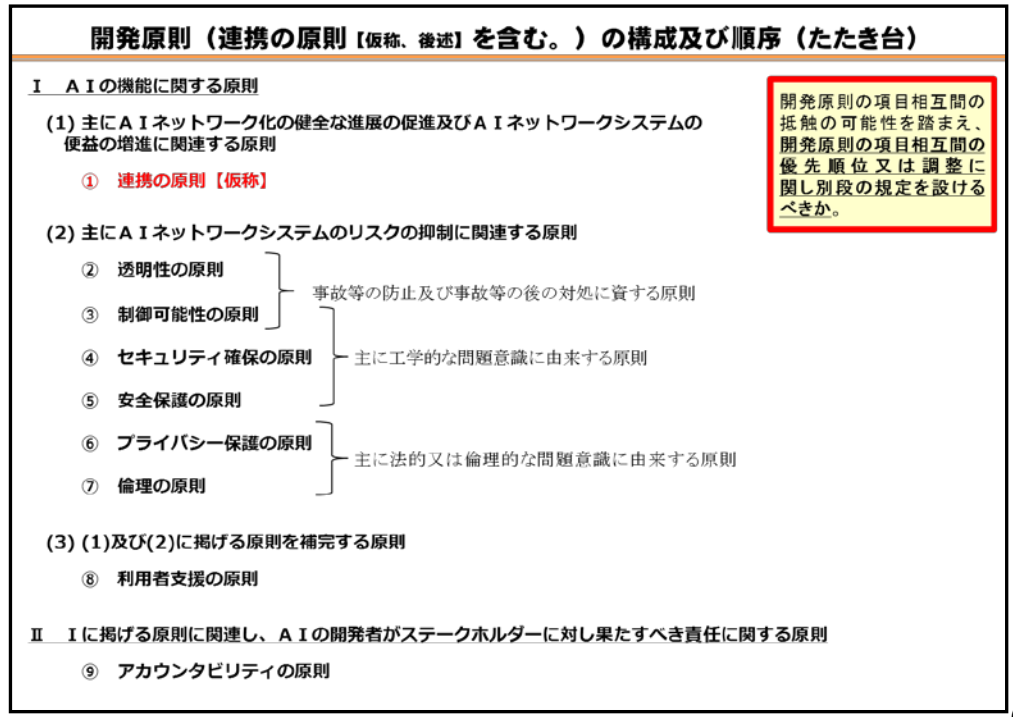
第五　分野共通開発ガイドラインの適用範囲

○適用対象とする「開発者」の範囲は、限定する必要はないのではないか。

○適用対象とする「AI」の範囲は、その機能如何にかかわらず、AIネットワークシステムの構成要素となり得るAI、すなわち、何らかの情報通信ネットワークシステムに実装し又は接続し得るAIを広く包含することとしてはどうか。

○適用対象とする「研究開発の段階」については、学問の自由等に鑑み、閉鎖された空間（実験室等）の外につながる情報通信ネットワークシステムに実装し又は接続して行う段階に限定することとしてはどうか。

第六　開発原則の構成及び順序



論点の要旨（3／8）

第七　開発原則の個々の項目の内容の具体化

(1) 透明性の原則　ＡＩネットワークシステムの動作の検証可能性及び説明可能性を確保すること。

○**動作の透明性**（検証可能性及び説明可能性）**が要請されるAIの動作の範囲**如何。**入出力、通信及び判断**としてよいか。

○**個人の生命・身体の安全等重要な権利利益若しくは法益に関するリスクを惹起し得る、又は個人に関する重大な決定のために利活用されるAIネットワークシステムの構成要素となり得るAIに関し、**

- 当該AIの動作のうち、**入出力及び通信については検証可能性を確保すべき**とし、**判断については、深層学習等における判断過程のブラックボックス化が指摘されていることなどに鑑み、技術的及び経済的な事情に鑑み合理的な範囲・水準で、検証可能性を確保するよう努めるべき**としてはどうか。
- **技術的及び経済的な事情に鑑み合理的な範囲・水準で動作の説明可能性を確保するよう努めるべき**としてはどうか。

(2) 制御可能性の原則　ＡＩネットワークシステムの制御可能性を確保すること。

○**制御不能となるリスクにつき、その蓋然性が高い又は不確実と考えられるAIについては、一般社会で利用される前に、実験室等閉鎖された空間において、当該空間の外につながる情報通信ネットワークシステムに接続せずに、AIの制御可能性について実験を行い、リスク評価を行うことにより、制御可能性を確保**することが求められるのではないか。

○AIネットワークシステムの制御可能性を継続的に確保するために、その構成要素となり得る**AIについて、人間又は信頼し得る他のAIによる監督及び対処（停止、切断、修理等）の実効性の確保が求められる**のではないか。

○【上記二点のほか、次頁の（※）において、「(3)　セキュリティ確保の原則」及び「(4)　安全保護の原則」と共通の論点を所掲。】

(3) セキュリティ確保の原則　ＡＩネットワークシステムの頑健性及び信頼性を確保すること。

○**セキュリティの範囲**には、**情報の機密性、完全性及び可用性のみならず**、当該システムの**信頼性**（意図した通りに動作が行われ権限を有しない第三者による操作を受けないこと）**及び頑健性**（物理的な攻撃や事故への耐性）**の維持も含まれる**としてはどうか。

○AIネットワークシステムのセキュリティを確保することができるよう、その構成要素となり得る**AIの設計段階において措置（セキュリティ・バイ・デザイン）を講ずべき**としてはどうか。

○【上記二点のほか、次頁の（※）において、「(2)　制御可能性の原則」及び「(4)　安全保護の原則」と共通の論点を所掲。】

論点の要旨（4／8）

(4) 安全保護の原則　ＡＩネットワークシステムが利用者及び第三者の生命・身体の安全に危害を及ぼさないように配慮すること。

○**安全保護の原則が適用されるAIの範囲は、個人の生命・身体の安全に関するリスクを惹起し得るAIネットワークシステムの構成要素となり得るAI**としてはどうか。

○AIネットワークシステムにおける**本質安全（運動能力等の抑制）、制御安全（監視装置等の実装）、機能安全等を確保することができるよう**、その構成要素となり得る**AIの設計段階において措置（セーフティ・バイ・デザイン）を講ずべき**としてはどうか。

○**ＡＩネットワークシステムを利活用する際の利用者及び第三者の生命・身体の安全に関する判断（例：生命・身体の安全を保護される個人の優先順位等に関する判断）を行うAIを研究開発する場合には、開発者は利用者等に対し当該判断を行うAIに関する設計の趣旨及び理由を説明すべき**としてはどうか。

○【上記三点のほか、下記（※）において、「(2) 制御可能性の原則」及び「(3) セキュリティ確保の原則」と共通の論点を所掲。】

（※） 上記「(2) 制御可能性の原則」、「(3) セキュリティ確保の原則」及び「(4) 安全保護の原則」の共通の論点

○**「(2) 制御可能性の原則」、「(3) セキュリティ確保の原則」及び「(4)安全保護の原則」においては**、リスクを評価し抑制するため、AIネットワークシステムの構成要素となり得るAIについて、**予め制御可能性の検証**（verification）[※形式的な整合性の検証]**及び妥当性確認**（validation）[※実質的な妥当性の確認]**を行うことが必要となる旨を定めるべき**ではないか。

(5) プライバシー保護の原則　ＡＩネットワークシステムが利用者及び第三者のプライバシーを侵害しないように配慮すること。

○プライバシー保護の原則において配慮されるべき**プライバシーの範囲には、空間プライバシー**（私生活の平穏）**、情報プライバシー**（個人データ）**、通信の秘密及び生体プライバシーが含まれる**としてはどうか。

○AIネットワークシステムにおけるプライバシー侵害のリスクを評価するために、その構成要素となり得るAIについて**予めプライバシー影響評価を行うべき**としてはどうか。

○AIネットワークシステムがその利活用に当たりプライバシーが保護されるものとなるよう、その構成要素となり得る**AIの設計段階において措置（プライバシー・バイ・デザイン）を講ずべき**としてはどうか。

論点の要旨（5／8）

(6) 倫理の原則　ＡＩネットワークシステムの研究開発において、人間の尊厳と個人の自律を尊重すること。

○倫理の原則においては、**人間性（humanity）の価値を中心に据えつつ、人間の尊厳と個人の自律を尊重すべき**旨を掲げることとしてはどうか。

○**国際人権法・国際人道法等を参照し、AIネットワークシステムが人間性の価値を毀損してはならない**としてはどうか。

○**人間の脳・身体と融合又は連携するAIを研究開発する際には、人間の尊厳と個人の自律の尊重について、生命倫理等の議論も参照しつつ、特に慎重に配慮すべき**としてはどうか。

○AIの開発において、**公正（fairness）等の価値に鑑みて**、**技術的に可能な範囲で**、**AIの学習するデータに含まれる偏見等に起因する差別**（人種、性、宗教等による差別）**を防止するための措置を講ずべき**としてはどうか。

(7) 利用者支援の原則　ＡＩネットワーク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よう配慮すること。

○（最終）利用者に操作されるAIネットワークシステムの構成要素となり得るAIについては、

- **利用者に対し適時適切にその判断に資する情報を提供し、かつ、利用者にとって操作しやすいインターフェイスが利用可能となるよう設計すべき**ではないか。
- **利用者に選択の機会を適時適切に提供する機能（ナッジ：例えば、デフォルトの設定、理解しやすい選択肢の提示・体系化、フィードバックの提供、緊急時の警告、エラーへの対処等）が利用可能となるよう設計すべき**ではないか。
- **ユニバーサル・デザイン等社会的弱者の受容可能性を高めるための取組に努めるべき**としてはどうか。

(8) アカウンタビリティの原則　ＡＩネットワークシステムの研究開発者が利用者など関係するステークホルダーに対しアカウンタビリティを果たすこと。

○開発者が説明責任を果たす上では、**特に利用者等に対し説明を行うべきであるほか**、**多様なステークホルダーと対話を行ってその意見を聴取する等ステークホルダーの積極的な関与を得るべき**ではないか。

○**開発原則の遵守状況につき開発者から説明を受けた利用者によるAIネットワークシステムへの信頼・期待が保護されるよう、利用者の責任との関係に留意しつつ、開発者の責任の在り方について指針を示すべき**ではないか。

論点の要旨（6／8）

第八　連携の原則【仮称】

○相互接続性・相互運用性の確保等AIネットワークシステム相互間の円滑な連携の確保に関し開発者が留意すべき事項を「連携の原則」【仮称】として開発原則に追加することとしてはどうか。また、連携の原則【仮称】及びその説明において記すべき事項如何。

○AIネットワークシステム相互間の円滑な連携の確保に関し開発者が留意すべき事項に関連し、国、関係国際機関等に推奨すべき事項として分野共通開発ガイドラインの関連文書（OECDのガイドラインであれば、理事会勧告の本紙）に記すべき事項如何。

○上記二点に関連し、AIネットワークシステム相互間の連携がAIネットワークシステムの利活用に伴うものであることに鑑み、AIネットワークシステム相互間の円滑な連携の確保に関しAIネットワークシステムの利活用の段階において利用者（特にAIネットワークサービスのプロバイダ）が留意すべき事項及び国、関係国際機関等に推奨すべき事項を後述する分野共通利活用ガイドライン及びその関連文書（OECDのガイドラインであれば、理事会勧告の本紙）にそれぞれ記すこととしてはどうか。

なお、AI及びAIネットワークシステムが現時点においては揺籃期であることから、上記三点に関し法的規制の創設を検討することは、現時点では時期尚早ではないか。少なくとも関連する弊害の蓋然性が顕著になるまでは、分野共通開発ガイドライン及び分野共通利活用ガイドライン並びにそれぞれの関連文書により開発者及びAIネットワークサービスのプロバイダ等利用者が留意すべき事項並びに国、関係国際機関等に推奨すべき事項を国際的に共有した上で、国、関係国際機関等は、関連する動向を注視して、動向やベストプラクティスに関する情報を国際的に共有するとともに、AIネットワークシステム相互間の連携をめぐる紛争の発生状況等に応じて、国内の紛争及び国境を越えた紛争の処理の在り方等を検討して所要の措置を講ずるにとどめるような謙抑的な姿勢であるべきではないか。

第九　開発原則の実効性の確保の在り方

○分野共通開発ガイドラインの関連文書（OECDのガイドラインであれば、理事会勧告の本紙）において、国、関係国際機関等に対し、開発原則の実効性の確保のための方策として、例えば次に掲げる方策を検討するよう推奨する旨を記すこととしてはどうか。

（1）公共調達の対象とするAI及び公的研究費の交付対象とするAIに関し、開発原則を踏まえて条件を設定

（2）市場の機能を活用して、開発原則に適合しているAIが市場において利用者に選択されやすくなる環境を整備

（→(2)については、「第十　開発原則の実効性の確保における市場の活用の在り方」参照。）

○分野共通開発ガイドラインの関連文書（OECDのガイドラインであれば、理事会勧告の本紙）において、各国の関係機関、関係国際機関等に対し、開発原則の実効性に関する状況、実効性の確保に関するベストプラクティス等に関する情報を共有し、相互に協力するよう推奨する旨を記すこととしてはどうか。

論点の要旨（7／8）

第十　開発原則の実効性の確保のための市場の活用の在り方

○「第九　開発原則の実効性の在り方」の(2)に掲げる方策として、**分野共通開発ガイドラインの関連文書**（OECDのガイドラインであれば、理事会勧告の本紙）**において**、**各国に対し**、例えば**次の①及び②の仕組みを分野共通の一般的仕組みとして一体的に整備**することについて検討するよう**推奨**するとともに、**各分野に係る各国の関係機関、関係国際機関等に対し、利活用の分野ごとの事情に照らし、必要に応じ分野別の特則的仕組みを検討するよう推奨**する旨を記すこととしてはどうか。

① **開発者が**その開発するAIに関し**開発原則への適合性に関する情報を客観的に信頼できる形で自発的に提供する仕組み**

（例） 開発者が自発的に提供する情報に基づき、第三者機関が当該AIの開発原則への適合性を評価して認証する制度

② 上記**①の仕組みにより開発者が提供した情報において開発原則に適合しているとされているAI**を実装するAIネットワークシステムの利活用に伴い、当該AI**のリスクが顕在化したことに起因する第三者の被害等に関し、その利用者の法的責任、法的義務等が問題となる場合において**、**当該利用者の当該情報に対する信頼に基づく期待を保護**するための仕組み

（例）当該被害等に関する当該利用者の法的責任等を減免する制度 （例：当該リスクが顕在化したこと自体について当該利用者を無過失とみなす制度）

○開発原則の実効性を確保するために市場を活用する場合であっても、**開発原則に掲げる事項のうち、人権等他の利益とバーターにすべきでないものについては、AIネットワークシステムの用途に照らし、必要に応じ当該用途に係る利活用の分野に関連する制度の整備等を検討するよう当該分野に係る各国の関係機関、関係国際機関等に推奨**する旨を分野共通開発ガイドラインの関連文書（OECDのガイドラインであれば、理事会勧告の本紙）及び後述する分野共通利活用ガイドラインの関連文書（同前）に記すこととしてはどうか。

論点の要旨（8／8）

第十一　ＡＩネットワークシステムの利活用に関し利用者等が留意すべき事項

○AIネットワークシステムの利活用に関し利用者（AIネットワークサービスのプロバイダ、最終利用者）が留意すべき事項及び国、関係国際機関等に推奨すべき事項を整理して、国際的に共有する枠組みとして「AIネットワークシステム利活用ガイドライン」（仮称）及びその関連文書（OECDのガイドラインであれば、理事会勧告の本紙）を策定し、開発ガイドライン及びその関連文書と相互に補完し合う二本柱とすることに向け、OECD等の協力の下、国際的に議論すべきではないか。

○利活用ガイドラインの体系については、開発ガイドラインと同様に、分野共通ガイドライン及び分野別ガイドラインからなるものとすることが適当ではないか。

→以下両者を区別する場合には、前者を「分野共通利活用ガイドライン」といい、後者を「分野別利活用ガイドライン」という。

分野共通利活用ガイドラインは、AIネットワークシステム（AIネットワークサービスを含む。）の利用者（AIネットワークサービスのプロバイダ及び最終利用者を含む。）が、その利活用（AIネットワークサービスの提供及び利活用を含む。）に当たり、利活用の分野を通じて留意すべき事項及び分野間の連携の可能性を踏まえて留意すべき事項（「利活用原則」）並びにその説明を策定するものとして、本推進会議がその検討と議論を推進してはどうか。

分野別利活用ガイドラインは、各分野における策定の要否そのもの及び策定する場合における内容の双方ともに、各分野の関係国際機関を含む当該分野の産学民官のステークホルダー自身による検討と議論に委ねることとしてはどうか。

○分野共通利活用ガイドラインに定める「利活用原則」は、AIネットワークシステム（AIネットワークサービスを含む。）の利用者（AIネットワークサービスのプロバイダ及び最終利用者を含む。）が、その利活用（AIネットワークサービスの提供及び利活用を含む。）に当たり、次の①～③に掲げる見地から利活用の分野を通じて又は分野間の連携の可能性を踏まえて留意すべき事項としてはどうか。

① AIネットワーク化の健全な進展の促進及びAIネットワークシステムの便益の増進
② AIネットワークシステムのリスクの抑制
③ AIネットワークシステムの利活用に伴い、当該AIネットワークシステムに実装するAIのリスクの顕在化に起因する被害に関する被害者の利益の保護

○自らAIネットワークシステムを構築する最終利用者、自ら構築するAIネットワークシステムによりAIネットワークサービスを最終利用者等他の者に提供するプロバイダ及びプロバイダからAIネットワークサービスの提供を受ける最終利用者の種別に応じて、適用すべき利活用原則の範囲、内容等に異同があり得ることから、その異同を利活用ガイドラインに明記するとともに、これら利用者の種別ごとに整理したマニュアル等を作成することとしてはどうか。

論点整理

AIネットワーク化検討会議「報告書2016」における定義

「AI」・「人工知能」：　定義せず

「AIネットワークシステム」　AIを構成要素とする情報通信ネットワークシステム

「AIネットワークサービス」　AIネットワークシステムの機能を提供するサービス

「AIネットワーク化」　AIネットワークシステムの構築及びAI相互間の連携等Aiネットワークシステムの高度化を連続的かつ一体的に捉えて総称する概念

人的主体（開発者、利用者）：　定義せず

国内の研究者によるAIの定義（総務省「平成28年情報通信白書」234頁（平成28年）参照。）

図表4-2-1-4　国内の主な研究者による人工知能（AI）の定義

研究者	所属	定義
中島秀之	公立はこだて未来大学	人工的につくられた、知能を持つ実態。あるいはそれをつくろうとすることによって知能自体を研究する分野である
武田英明	国立情報学研究所	
西田豊明	京都大学	「知能を持つメカ」ないしは「心を持つメカ」である
溝口理一郎	北陸先端科学技術大学院	人工的につくった知的な振る舞いをするためのもの（システム）である
長尾真	京都大学	人間の頭脳活動を極限までシミュレートするシステムである
堀浩一	東京大学	人工的に作る新しい知能の世界である
浅田稔	大阪大学	知能の定義が明確でないので、人工知能を明確に定義できない
松原仁	公立はこだて未来大学	究極には人間と区別が付かない人工的な知能のこと
池上高志	東京大学	自然にわれわれがペットや人に接触するような、情動と冗談に満ちた相互作用を、物理法則に関係なく、あるいは逆らって、人工的につくり出せるシステム
山口高平	慶應義塾大学	人の知的な振る舞いを模倣・支援・超越するための構成的システム
栗原聡	電気通信大学	人工的につくられる知能であるが、その知能のレベルは人を超えているものを想像している
山川宏	ドワンゴ人工知能研究所	計算機知能のうちで、人間が直接・間接に設計する場合を人工知能と呼んで良いのではないかと思う
松尾豊	東京大学	人工的につくられた人間のような知能、ないしはそれをつくる技術。人間のように知的であるとは、「気づくことのできる」コンピュータ、つまり、データの中から特徴量を生成し現象をモデル化することのできるコンピュータという意味である

（出典）松尾豊「人工知能は人間を超えるか」（KADOKAWA）p.45より作成

「データ」・「情報」・「知識」・「知能」・「智慧」の関係（AIネットワーク化検討会議「報告書2016」の趣旨に即して整理）

「データ」・「情報」・「知識」・「知能」・「智慧」の関係

データ	（Data）	断片的な事実、数値、文字
情　報	（Information）	データの組み合わせに意味を付与したもの
知　識	（Knowledge）	データ・情報の体系的集積
知　能	（Intelligence）	データ・情報・知識を学習し、解析することにより、新たなデータ・情報・知識を創造する機能
智　慧	（Wisdom）	データ・情報・知識に基づき、知能を活用することにより、物事に対処する人間の能力

第一　基本概念の定義（2／3）

論点

1．**「AI」（人工知能）自体については、上記のとおり論者によってその定義は区々であり**、かつ、今後の研究の進展次第では様々なAIが発展していくものと考えられ、**その発展の範囲を画定することが容易でないことに鑑みると、本推進会議においても**、検討会議と同じく、**AI自体の定義は差し当たり措くこととし、開発原則を構成する個々の原則の適用範囲を今後必要に応じて個別に画定**していくこととしてはどうか。

その上で、開発ガイドラインの適用に当たり、関心を向けるべきAIに何らかの通有性があるとすれば、それはどのようなものか。

なお、**「知能」についても、広く通用している定義がないことから、本推進会議においては定義しない**こととした上で、**「知能」・「データ」・「情報」・「知識」・「智慧」の関係については、検討会議「報告書2016」の趣旨に即して整理した上記の「関係」を前提**とすることとしてはどうか。

2．**「AIネットワークシステム」、「AIネットワークサービス」及び「AIネットワーク化」については**、検討会議以来の議論の連続性を確保する見地から、差し当たり、**検討会議の定義を踏襲して検討を進める**こととしてはどうか。

3．以下において**「開発者」とは、AIの研究開発**（AIを研究し又は開発する行為のほか、複数のAIを組み合わせて一体的なAIとして機能するよう構成する行為（例　画像認識系のAIと自然言語処理系のAIとを一体的なAIとして機能するよう構成する行為）を含む。この資料においては、「AIの研究開発」を「AIの開発」又は単に「開発」という場合がある。）**をする者**（自らが研究開発したAIを実装するAIネットワークシステムによるAIネットワークサービスのプロバイダを含む。）**をいう**ものとしてはどうか。

4．以下において**「利用者」とは、他の開発者が研究開発したAI**（他の開発者が開発したAIを実装するAIネットワークシステムにより他のプロバイダが提供するAIネットワークサービスを含む。以下この4．において同じ。）**の提供を受けてAIネットワークシステムを利用する者**（自らが構築するAIネットワークシステムを自ら利用する個人又は団体（最終利用者たる個人又は団体のほか、他の開発者が開発したAIの提供を受けてAIネットワークシステムを構築してAIネットワークサービスを他の者に提供するプロバイダをも含む。）のほか、プロバイダからAIネットワークサービスの提供を受ける最終利用者たる個人又は団体も含む。）**をいう**ものとしてはどうか。

※　上記のとおり「利用者」は「最終利用者」をその部分集合とする用語であり、以下においては「利用者」と「最終利用者」とを使い分けている。

第一　基本概念の定義（3／3）

論点

5．ただし、**「開発者」及び「利用者」は、関係者間の関係に即して、関係が生ずる場面ごとに個別に評価される相対的な概念**であり、**一の者がある場面では「開発者」と評価**され、**別の場面では「利用者」と評価されることがあり得ることに留意**すべきではないか。

（例１）　**自らが開発した「AI①」及び他のベンダーが開発した「AI②」とを組み合わせて「一体的なAI」として機能するよう構成した上で**「一体的なAI」を実装して**AIネットワークシステムを構築し、当該AIネットワークシステムによりAIネットワークサービスを第三者である最終利用者たる個人又は団体に提供するプロバイダ**は、

・**当該他のベンダーとの関係においては、**当該他のベンダーが「AI②」の「開発者」であり、**当該プロバイダは**（「AI②」を実装するものとしての）**当該AIネットワークシステムの「利用者」である**（「AI②」の「利用者」ではなく、あくまでも当該AIネットワークシステムの「利用者」である。）。

・**当該AIネットワークサービスの提供を受ける第三者たる最終利用者との関係においては、当該プロバイダが当該「一体的なAI」の「開発者」**であり、**当該最終利用者は**（当該「一体的なAI」を実装するものとしての）**当該AIネットワークシステムの「利用者」である**（当該「一体的AI」の「利用者」ではなく、あくまでも当該AIネットワークシステムの「利用者」である。）。

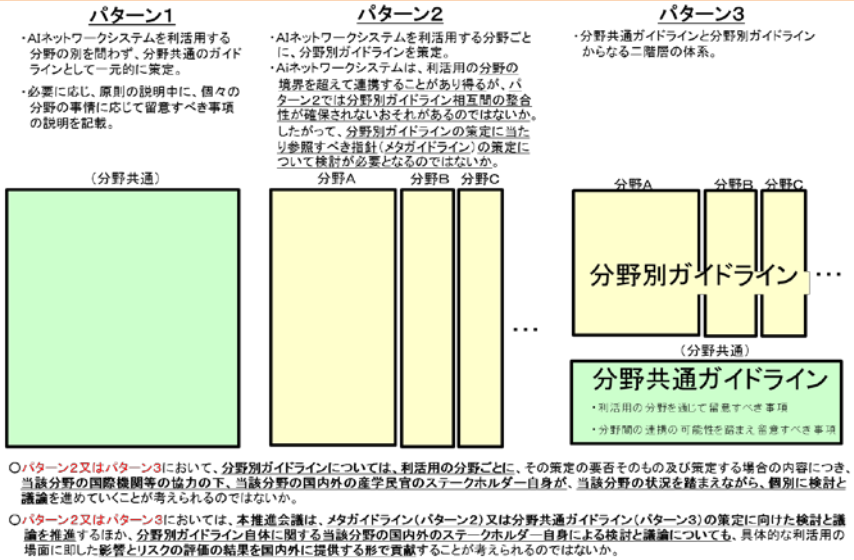
（例２）　自らはAIの研究開発（上記3．で述べたように、複数のAIを組み合わせて一体的なAIとして機能するよう構成する行為を含む。）を行わず、**他の開発者が開発したAI**（他の開発者が開発したAIを実装するAIネットワークシステムにより他のプロバイダが提供するAIネットワークサービスを含む。以下この（例2）において同じ。）**の提供を受けて自らが構築するAIネットワークシステムによりAIネットワークサービスを最終利用者等たる個人又は団体**（ここで「最終利用者等」の「等」は、当該AIネットワークサービスの提供を受け、これを活用して、別の個人又は団体にAIネットワークサービスを提供するプロバイダを含む。）**に提供するプロバイダ**（以下この「当該プロバイダ」という。）**は、「開発者」ではなく**、あくまでも当該AIネットワークシステムの**「利用者」である。**

この場合において、**当該プロバイダがそのAIネットワークサービスをその最終利用者等に提供するに当たり、AIネットワーク化の健全な進展の促進並びにAIネットワークシステムの便益の増進及びリスクの抑制に関し、当該プロバイダ及び最終利用者等がそれぞれの立場において留意すべき事項については、開発ガイドラインではなく**、後述する**利活用ガイドラインに定める**こととしてはどうか。

また、この場合において当該プロバイダ及び最終利用者等が留意すべき事項に**関連して、国、関係国際機関等に推奨すべき事項については、開発ガイドラインの関連文書**（OECDのガイドラインであれば、理事会勧告の別紙）**ではなく**、後述する**利活用ガイドラインの関連文書**（同前）**に定める**こととしてはどうか。

第二　AI開発ガイドラインの体系

事務局資料（たたき台）



構成員からの指摘

【新保構成員、平野構成員、中西構成員、クロサカ構成員】（第１回開発原則分科会） **パターン3が適切ではないか。**

（新保構成員） **汎用AI**だけであればパターン1、**特化型AI**だけであればパターン2とすることも考えられるが、**双方を視野に入れることが必要**であることに鑑みると、パターン3が適切ではないか。

（平野構成員） パターン3は、**各種契約においても多く用いられている形態であり、合理性がある。**

（中西構成員） **分野相互間の連携に関する事項や分野共通の事項と分野別の事項の双方があることに鑑みると、パターン3が**適切**切**ではないか。

（クロサカ構成員） **パターン1とパターン2の双方の長所を併せ持つパターン3が適切**ではないか。

論点

1. 開発ガイドラインの体系については、**分野共通ガイドライン及び分野別ガイドラインからなる「パターン3」が適当**ではないか。

→以下両者を区別する場合には、前者を「分野共通開発ガイドライン」といい、後者を「分野別開発ガイドライン」という。

分野共通開発ガイドラインは、AIネットワークシステムの**利活用の分野を通じて開発者が留意すべき事項及び利活用の分野間の連携の可能性を踏まえて開発者が留意すべき事項を策定**するものとして、**本推進会議がその検討と議論を推進**してはどうか。

分野別開発ガイドラインは、各分野における策定の要否そのもの及び策定する場合における内容の双方ともに、**各分野の関係国際機関を含む当該分野の産学民官のステークホルダー自身による検討と議論に委ねる**こととしてはどうか。

2. **分野間の連携の可能性に鑑み**、相互接続性・相互運用性の確保等**AIネットワークシステム相互間の円滑な連携の確保に関しAIの開発段階において開発者が留意すべき事項を「連携の原則」【仮称】として開発原則に追加する**こととしてはどうか。（後述）

3. **AIネットワークシステムの用途固有の事情により分野共通開発ガイドラインに定める開発原則の項目の一部又は全部をそのまま適用すべきでない場合**には、当該項目について**当該用途に係る特則を分野別開発ガイドラインに定める**こととしてはどうか。

第三　分野共通開発ガイドラインの構成（1／2）

1. 全体の構成（たたき台）

(1) 分野共通開発ガイドライン（OECDのガイドラインであれば、理事会勧告の附属文書（Annex））

AIネットワークシステムの構成要素となり得る**AIの開発者が**、**その研究開発に当たり**、**AIネットワーク化の健全な進展の促進並びにAIネットワークシステムの便益の増進及びリスクの抑制に関し**、AIネットワークシステムの**利活用の分野を通じて留意すべき事項及び利活用の分野間の連携の可能性を踏まえて留意すべき事項に関する原則（「開発原則」）並びにその説明**

(2) 分野共通開発ガイドラインの関連文書（OECDのガイドラインであれば、理事会勧告の本紙）

- 分野共通開発ガイドラインに定める事項に関連し、**国、関係国際機関等に推奨すべき事項**
- ガイドラインの**見直しの時期及び方法**

2. 分野共通開発ガイドラインの構成（たたき台）

1. 目的
2. 基本理念
3. 定義
   - (1) AI
   - (2) AIネットワークシステム
   - (3) AIネットワークサービス
   - (4) AIネットワーク化
   - (5) 研究開発
   - (6) 開発者
   - (7) 利用者
4. 適用範囲
   - (1) 開発者
   - (2) AI
   - (3) 研究開発の段階
5. 原則
   - (1) 透明性の原則
   - (2) 利用者支援の原則
   - (3) 制御可能性の原則
   - (4) セキュリティ確保の原則
   - (5) 安全保護の原則
   - (6) プライバシーの原則
   - (7) 倫理の原則
   - (8) アカウンタビリティの原則

※　各原則について、原則の内容の記述に加えて、説明を付記。

※　原則の項目相互間の優先順位又は調整規定を設けるべきか。（後述）

※　開発原則の構成及び順序については、後述。

※　原則及びその説明の形式については、一覧性を向上する見地から、OECDセキュリティガイドライン及び同・暗号政策ガイドラインの例に倣い、次のように個々の原則の規定に続けてその説明を記す形式としてはどうか。

> ○○の原則　××を△△すべきである。
> 【説明】・・・・・・・・である。

※　相互接続性・相互運用性の確保等AIネットワークシステム相互間の円滑な連携の確保に関しAIの開発段階において留意すべき事項を「連携の原則」【仮称】として追加してはどうか。（後述）

第三　分野共通開発ガイドラインの構成（2／2）

【参考】 OECDの各ガイドラインの構成

OECDプライバシーガイドライン

第1部　総論
- 定義
- 適用範囲

第2部　国内適用の基本原則
- 収集制限の原則
- データ内容の原則
- 目的明確化の原則
- 利用制限の原則
- 安全保護措置の原則
- 公開の原則
- 個人参加の原則
- 責任の原則

第3部　責任の履行

第4部　国際的適用における基本原則

第5部　国内実施

第6部　国際協力と相互運用性

＊　ガイドライン全体に補足説明書が付されている。

OECDセキュリティガイドライン

序文

Ⅰ　セキュリティ文化に向けて

Ⅱ　目的

Ⅲ　原則

（1）　認識

（2）　責任

（3）　対応

（4）　倫理

（5）　民主主義

（6）　リスクアセスメント

（7）　セキュリティの設計及び実装

（8）　セキュリティマネジメント

（9）　再評価

＊　各原則について、原則の規定に続けて説明が記されている。

OECD暗号政策ガイドライン

Ⅰ　目的

Ⅱ　適用範囲

Ⅲ　定義

Ⅳ　統合

Ⅴ　原則

（1）　暗号手法に対する信頼

（2）　暗号手法の選択

（3）　市場主導の暗号手法の開発

（4）　暗号手法に関する諸基準

（5）　プライバシー及び個人データの保護

（6）　合法的アクセス

（7）　責任

（8）　国際協力

＊　各原則について、原則の規定に続けて説明が記されている。

AIネットワーク化検討会議「報告書2016」（抄）（第5章1.「研究開発の原則・指針の策定」）

【今後の課題】

AIネットワーク化が社会にもたらす影響及びリスクに鑑みると、AIネットワークシステムの構成要素となり得るAIに関し、その研究開発に当たり留意すべき事項を整理し、国際的に共有することにより、研究開発の円滑化を図ることがAIネットワークシステムの社会における受容の向上、そして、智連社会への円滑な移行のために必要かつ効果的であるものと考えられる。そこで、中間報告書では、OECDプライバシーガイドライン、同・セキュリティガイドライン等を参考に、研究開発に関する原則・指針を国際的に参照される枠組みとして策定することに向け、関係する各種ステークホルダーの参画を得つつ、検討に着手すべき旨を提言した。

中間報告書の提言を踏まえ、G7香川・高松情報通信大臣会合において、我が国から、OECD等国際機関の協力も得て、AIの研究開発に関する原則（以下「開発原則」という場合がある。）の策定等に関し国際的な議論を進めることの提案がなされ、各国から賛同が得られたところである。

今後は、開発原則そのものの策定に向けた取組と並行して、その説明（開発原則の内容を敷衍し、又は具体化するもの）の作成に向けた取組も進めていくことが求められる。すなわち、開発原則及びその説明から構成される指針（「AI開発ガイドライン」（仮称））を国際的に参照される枠組みとして策定することに向け、開発原則及びその説明の双方につき内容面の検討を進めていくと同時に、関係する各種ステークホルダーの参画を得つつ、OECD等国際社会において継続的な議論が行われるよう働きかけていくべきである。

(1) 基本的な考え方

研究開発の原則・指針の策定・解釈に当たっては、次に掲げる考え方を基本的な考え方として掲げることが適切であるものと考えられる。

- 人間がAIネットワークシステムと共存することにより、AIネットワークシステムの恵沢が万人に享受され、人間の尊厳と個人の自律が保障されるとともに、AIネットワークシステムの制御可能性と透明性が確保され、AIネットワークシステムが安全に安心して利活用される社会を実現するという理念の下、研究開発に関する原則・指針を国際的に参照される枠組みとして策定すること。
- 研究開発の進展段階に応じて、想定される各種のリスクに適時適切に対処するとともに、イノベーティブな研究開発と公正な競争にも配慮しつつ、多様なステークホルダーの参画を得て、関係する価値・利益のバランスを図ること。
- AIネットワーク化の進展及び関連するリスクの顕在化に応じて、研究開発の原則・指針を適宜見直していくこと。

第四　分野共通開発ガイドラインの目的、基本理念等（2／4）

AIネットワーク化検討会議「報告書2016」（抄））　（第2章3.（1）「智連社会」（Wisdom Network Society：WINS）」）

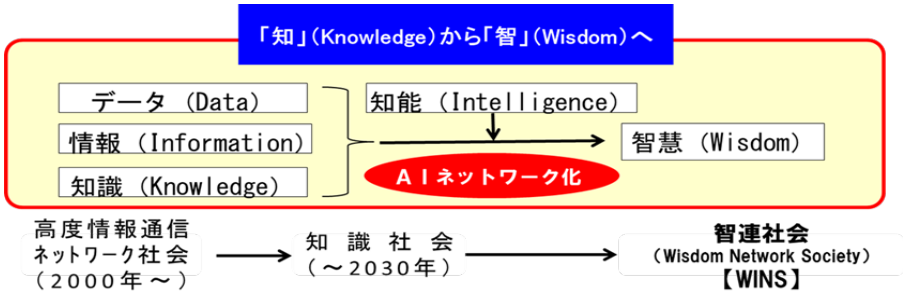
本検討会議は、中間報告書において、AIネットワーク化の進展を見据え、人間とAIネットワークシステムとが共存する段階（第四段階）における社会の在り方を構想した結果、目指すべき社会像として、智連社会（Wisdom Network Society：WINS（ウインズ））を掲げた。この智連社会という社会像は、「高度情報通信ネットワーク社会」及び「知識社会」のような「情報」・「知識」（知）に着目した従来の社会像の次にその実現を目指すべき、「智慧」（智）に着目した社会像として構想したものである。

AIネットワーク化の進展により、AIネットワークシステムの知能を活用してデータ・情報・知識を解析し、新たなデータ・情報・知識を創造することが可能となる社会の到来が予測される。それに伴い、「データ・情報・知識に基づき、知能を活用することにより、物事に対処する人間の能力」としての「智慧」（智）の役割が大きくなることが見込まれる。そのような中、人間は、AIネットワークシステムを活用することにより、各々の「智慧」（智）を連結し、「智のネットワーク」を構築していくことが期待される。

智連社会の構想は、このような問題意識に基づくものである。智連社会とは、AIネットワーク化の進展の結果として、人間がAIネットワークシステムと共存し、データ・情報・知識を自由かつ安全に創造・流通・連結して智のネットワークを構築することにより、あらゆる分野におけるヒト・モノ・コト相互間の空間を越えた協調が進展し、もって創造的かつ活力ある発展が可能となる社会である。この社会像については、中間報告書において、次のように図示されている。

・・・

・・・目指すべき社会像として智連社会を掲げるということは、AIネットワークシステムが社会の中心となるのではなく、あくまでも人間が社会の中心となり、人間がAIネットワークシステムを主体的に使いこなす社会を目指すべきとの考えを含意しているものと認められよう。



「智連社会」（Wisdom Network Society：WINS（ウインズ））

- 人間がAIネットワークシステムと共存し　→　人機共存
- データ・情報・知識を自由かつ安全に創造・流通・連結して智のネットワークを構築することにより　→　総智連環
- あらゆる分野におけるヒト・モノ・コト相互間の空間を越えた協調が進展し　→　協調遍在

もって創造的かつ活力ある発展が可能となる社会

智連社会の基本理念

① すべての人々がAIネットワークシステムの恵沢をあまねく享受

② 個人が、尊厳をもった自律的な主体として、AIネットワークシステムを安心して安全に利活用

③ イノベーティブな研究開発と公正な競争を通じて、多様で高度なAIネットワークシステムを実現

④ AIネットワークシステムに関し、制御可能性と透明性を技術的・制度的に確保

⑤ AIネットワークシステムの在り方に関する意思決定に多様なステークホルダーが民主的に参画

⑥ AIネットワークシステムを利活用して物理空間とサイバー空間を連結し、両者の調和を図ることにより、ヒト・モノ・コト相互間の空間を超えた協調を実現

⑦ AIネットワークシステムを利活用してヒト・モノ・コト相互間の空間を越えた協調が地域内・地域間で進展することにより、活力ある地域社会を実現

⑧ AIネットワークシステムにより、地球規模課題（環境保護、格差是正等）の解決に貢献

第四　分野共通開発ガイドラインの目的、基本理念等（3／4）

構成員からの指摘

【堀構成員】（第１回親会）「AI開発ガイドライン」の検討に当たっては、上から目線で人々を啓蒙するような形ではなく、利用者の目線で検討すべきである。そのようにすることにより、日本から開発原則を提案する意味があるものとなる。

【新保構成員】（第１回開発原則分科会）ガイドラインではバイ・デザインによりリスクに事前に対処する姿勢を前面に出すべきである。また、プライバシーガイドラインをはじめOECDのガイドラインでは、情報の自由な流通や表現の自由の価値が前提となっており、AIの研究開発ガイドラインを策定する際にも、情報の自由な流通、表現の自由、オープンデータ等の価値を掲げるべきではないか。

【実積構成員】（第１回親会）ガイドラインの策定に当たっては、これから将来に向けて何が起こるかわからないことを認識することが重要である。現時点で、将来の全てを見通したガイドラインを策定することはできないため、基本的な原則を定めるとしても、各論については将来世代にも判断を委ねることができるような謙抑的なガイドラインとすべきである。

【クロサカ構成員】（第１回開発原則分科会） AIネットワーク化の進展に関する政府の役割についても記すべき。

論点

1．分野共通開発ガイドラインの「目的」については、検討会議の提言を踏まえつつ、

- AIネットワーク化の進展及びこれを通じた智連社会の形成が、AIネットワークシステム（AIネットワークサービスを含む。）を利用しようするあらゆる個人・団体が最終利用者としてこれを利用することを通じて社会に参加し、活動できることを通じて実現するものであることから、AIネットワーク化の健全な進展を通じた智連社会の形成の見地から、AIネットワーク化が進展していく中であらゆる個人及び団体を最終利用者として社会に包摂し、参加と活動を可能とするためには、最終利用者（潜在的最終利用者を含む。）の利益を保護することが必要となること
- AIネットワーク化の進展及び智連続社会の形成に当たり、第三者や社会ないし人類への波及的な悪影響を抑制すべきこと
- AIネットワーク化の進展を通じて目指すべき社会像たる智連社会については、人間が中心となる社会像と理解されていること

に鑑み、次のような趣旨を必要に応じて再構成した上で分野共通開発ガイドラインにその「目的」として掲げることとしてはどうか。

> このガイドラインは、AIネットワークシステムの公共性に鑑み、AIネットワークシステムの構成要素となり得るAIの研究開発を行う者が、その研究開発に当たり、AIネットワーク化の健全な進展の促進並びにAIネットワークシステム（AIネットワークサービスを含む。以下同じ。）の便益の増進及びリスクの抑制に関し、AIネットワークシステムの利活用の分野を通じて又は分野間の連携の可能性を踏まえて留意すべき事項を開発原則として整理し、非拘束的な枠組みとして国際的に共有することにより、AIネットワークシステムの最終利用者の利益を保護するとともに第三者及び社会への波及的な悪影響を防止し、もって人間中心の智連社会の形成に資することを目的とする。（なお、「人間中心の」については、「人間が主体的にAIネットワークシステムを使いこなす」等の表現とすることも考えられる。）

第四　分野共通開発ガイドラインの目的、基本理念等（4／4）

論点

2. **検討会議においては、分野共通開発ガイドラインの策定及び解釈に当たっての「基本的な考え方」として**

① **智連社会自体の構成要素の一部**（「人間がAIネットワークシステムと共存」すること）

② **智連社会の基本理念の要素の一部**（AIネットワークシステムの恵沢の万人による享受、人間の尊厳と個人の自律の保障、イノベーティブな研究開発と公正な競争、制御可能性及び透明性の確保、AIネットワークシステムを安心して安全に利活用）

③ **リスクへの適時適切な対処**

④ **関係する価値・利益のバランスの確保**

⑤ AIネットワーク化の進展及び関連するリスクの顕在化に応じた**開発原則・開発ガイドラインの見直し**

が掲げられているが、これら**①～⑤に掲げる事項については**、必要に応じて敷衍するとともに、必要に応じて再構成した上で、**分野共通開発ガイドラインにおいて、その策定及び解釈に当たっての「基本理念」として掲げる**こととしてはどうか。

その際、**上記1．の「目的」と重複する事項については、その重要性等に鑑み、重複の適否を個別に検討した上で、素案の起草に当たり適宜整理**することとしてはどうか。

3. **上記2．の「基本理念」における①（智連社会自体の構成要素）の記述においては**、「人間がAIネットワークシステムと共存」と**同じく智連社会自体の構成要素たる**

「データ・情報・知識の自由かつ安全な創造・流通・連結」

「ヒト・モノ・コト相互間の空間を超えた協調」

についても、必要に応じて敷衍するとともに、必要に応じて再構成した上で、**いずれも盛り込む**こととしてはどうか。

4. **上記2．の「基本理念」における②（智連社会の基本理念の要素）の記述においては、智連社会の基本理念の8項目に掲げる事項を**、必要に応じて敷衍するとともに、必要に応じて再構成した上で、**すべて盛り込む**こととしてはどうか。

5. **「情報の自由な流通、表現の自由、オープン・データ等の価値」に関し**、これらの価値に関連する事項のうち**開発の段階において開発者が留意すべき事項については上記2．～4．により「基本理念」として掲げる**こととするほか、これらの価値に関連する事項のうち**利活用の段階において利用者**（AIネットワークサービスのプロバイダ、最終利用者）**が留意すべき事項及びこれに関連して国、関係国際機関等に推奨すべき事項については**、それぞれ後述する**利活用ガイドライン及びその関連文書**（OECDのガイドラインであれば、理事会勧告の本紙）**を検討する際に考慮する**こととしてはどうか。

6. **政府の役割如何**。

第五　分野共通開発ガイドラインの適用範囲（1／3）

AIネットワーク化検討会議「報告書2016」（抄）

AIネットワーク化が社会にもたらす影響及びリスクに鑑みると、**AIネットワークシステムの構成要素となり得るAIに関し**、その**研究開発に当たり留意すべき事項を整理し、国際的に共有**することにより、研究開発の円滑化を図ることがAIネットワークシステムの社会における受容の向上、そして、智連社会への円滑な移行のために必要かつ効果的であるものと考えられる。

AIネットワーク化検討会議「中間報告書」（抄）

（注5） なお、**本検討会議は**、AIネットワーク化を主題として議論を行うため、個々のAI自体ではなく、AIを構成要素とするAIネットワークシステムに焦点を当てて検討を行うものではあるが、**AIネットワークシステムに接続し得るAIを広く検討対象とする**ものである。確かに、**情報通信ネットワークシステムに接続することなく機能するAI**も存在し得るが、AIネットワーク化の進展を通じて、そのようなAIも**次第に情報通信ネットワークシステムに接続されるようになっていく**と想定されること、また、**現に情報通信ネットワークシステムに接続していなかったとしても、技術的には情報通信ネットワークシステムに接続し得るAIが大多数を占める**ことを踏まえ、**本検討会議においては、便宜上、AIについて、基本的にはAIネットワークシステムの構成要素となるものと捉える**こととする。

論点

1．分野共通開発ガイドラインの適用対象とする「開発者」の範囲

誰が開発するAIであっても、AIネットワークシステムの構成要素として実装されれば、他の情報通信ネットワークシステム、他のAI、当該AIネットワークシステムの利用者、第三者及び社会ないし人類に影響を及ぼし得るものであることに鑑みると、少なくとも**分野共通開発ガイドラインの適用対象とする「開発者」の範囲については、限定する必要はない**のではないか。

第五　分野共通開発ガイドラインの適用範囲（2／3）

論点

2. 分野共通開発ガイドラインの適用対象とする「AI」の範囲

その機能が特定の事項に特化されているAIであっても、他のAIと連携する形で利用されればその用途は拡大し、更にそれがAIネットワークシステムの構成要素として実装されればその影響及びリスクが及ぶ範囲は拡大し得ることから、**分野共通開発ガイドラインの適用対象とする「AI」の範囲については、その機能の範囲如何により限定すべきではない**のではないか。

他方、「AI」がAIネットワークシステムの構成要素となり得るものであるのか否か、すなわち、**「AI」が何らかの情報通信ネットワークシステムに実装し又は接続することが技術的に可能なものであるのか否かについては、次に掲げる理由から、この可否に応じて分野共通開発ガイドラインの適用対象とするのか否かを画することとすべき**ではないか。

- **何らかの情報通信ネットワークシステムに実装し又は接続することが技術的に可能なAIは**、これを当該情報通信ネットワークシステムに実装し又は接続することが技術的に可能である上で、更に当該AIを実装し又は接続した当該情報通信ネットワークシステムを（必要に応じてプロトコルの変換等をした上で）インターネット等に接続することが技術的に可能であることから、**当該AIの影響及びリスクを**当該情報通信ネットワークシステム自体又は当該情報通信ネットワークシステムと接続するインターネット等を通じて**国境を越えて広範囲に波及させることが技術的に可能であることに鑑みると、当該AIの影響及びリスクに関する事項を「国際的に共有」する必要性が認められる**のではないか。
  したがって、**このようなAIについては**、AIの影響及びリスクに関連して**「国際的に共有」すべきものである分野共通開発ガイドラインの適用対象とすべき**ではないか。

- 他方、**いかなる情報通信ネットワークシステムに実装し又は接続することも技術的に不可能なAIは**、いかなる情報通信ネットワークシステムに実装し又は接続することも技術的に不可能である以上、更に何らかの情報通信ネットワークシステムを通じてインターネット等に接続することも技術的に不可能であることから、**当該AIの影響及びリスクを国境を越えて広範囲に波及させることが技術的に困難であるため、当該AIの影響及びリスクに関する事項を「国際的に共有」する必要性が乏しい**のではないか。
  したがって、**このようなAIについては**、AIの影響及びリスクに関連して**「国際的に共有」すべきものである分野共通開発ガイドラインの適用対象とする必要性が乏しい**のではないか。
  このようなAIについては、分野共通開発ガイドラインの適用対象とする必要性が乏しい以上、**学問の自由等に鑑み**、適用範囲をできる限り限定的なものとするため、**分野共通開発ガイドラインの適用対象としないこととするのが適切**ではないか。

以上から、**分野共通開発ガイドラインの適用対象とする「AI」の範囲については、AIの機能の範囲如何にかかわらず、AIネットワークシステムの構成要素となり得るAI、すなわち、何らかの情報通信ネットワークシステムに実装し又は接続し得るAIを広く包含する**こととしてはどうか。

第五　分野共通開発ガイドラインの適用範囲（3／3）

論点

3. 分野共通開発ガイドラインの適用対象とする「研究開発の段階」の範囲

2. のAIの研究開発のうち、閉鎖された空間（外界への影響及び外界からの影響のいずれについても遮断することができるよう措置が講ぜられている実験室等）の外につながる情報通信ネットワークシステムに実装し、又は接続して行う段階においては、研究開発中といえども、当該AIのリスクが当該情報通信ネットワークシステムを通じて当該空間の外に波及するおそれがあるのではないか。

他方、2. のAIの研究開発のうち、閉鎖された空間（同前）の中で、当該空間の外につながる情報通信ネットワークに実装も接続もせずに行う段階においては、当該研究開発中の当該AIのリスクが当該空間の外に波及するおそれは、その蓋然性が相対的に低いと認められる程度には抑制されているのではないか。

したがって、少なくとも**分野共通開発ガイドラインの適用対象とする「研究開発の段階」の範囲については、学問の自由等に鑑み、適用範囲をできる限り限定的なものとするため**、差し当たり、**閉鎖された空間**（外界への影響及び外界からの影響のいずれについても遮断することができるよう措置が講ぜられている実験室等）**の外につながる情報通信ネットワークシステムに実装し又は接続して行う段階に限定する**こととしてはどうか。

ただし、**閉鎖された空間**（同前）**の中で、当該空間の外につながる情報通信ネットワークに実装も接続もせずに行う研究開発の段階**（以下この頁において「**閉鎖空間内で研究開発を行う段階**」という。）**を分野共通開発ガイドラインの適用対象とはしないこととする場合といえども**、研究開発の進展に応じて、**閉鎖空間内で研究開発を行う段階から、いずれは、分野共通開発ガイドラインの適用対象となる研究開発の段階**（閉鎖された空間（同前）の外につながる情報通信ネットワークシステムに実装し又は接続して研究開発を行う段階）**又は実用に供する段階に移行**していくこととなるものと考えられるのではないか。

このことに鑑みると、分野共通開発ガイドラインの「適用範囲」のうち「(3) 研究開発の段階」において、**開発者にあっては**、分野共通開発ガイドラインの適用対象とはしない**閉鎖空間内で研究開発を行う段階においても、分野共通開発ガイドラインの適用対象となる研究開発を行う段階に移行する時又は実用に供する時のいずれか早いものが到来するまでには**、その研究開発の対象とする**AIに関し開発原則への適合性を確保することができるよう準備を進めておくことに留意すべき旨を付言**してはどうか。

第六　開発原則の構成及び順序（1／7）

AIネットワーク化検討会議「報告書2016」（抄）

(2) 開発原則の内容

開発原則の策定に当たっては、**少なくとも、次に掲げる事項をその内容に盛り込むべき**ものと考えられる。

ただし、**事項の加除又は整理統合を否定しようとするものではなく、幅広く検討を進めていくことが期待される**。

① **透明性の原則**
ＡＩネットワークシステムの動作の検証可能性及び説明可能性を確保すること。

② **利用者支援の原則**
ＡＩネットワーク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よう配慮すること。

③ **制御可能性の原則**
人間によるＡＩネットワークシステムの制御可能性を確保すること。

④ **セキュリティ確保の原則**
ＡＩネットワークシステムの頑健性及び信頼性を確保すること。

⑤ **安全保護の原則**
ＡＩネットワークシステムが利用者及び第三者の生命・身体の安全に危害を及ぼさないよう配慮すること。

⑥ **プライバシー保護の原則**
ＡＩネットワークシステムが利用者及び第三者のプライバシーを侵害しないように配慮すること。

⑦ **倫理の原則**
ＡＩネットワークシステムの研究開発において、人間の尊厳と個人の自律を尊重すること。

⑧ **アカウンタビリティの原則**
ＡＩネットワークシステムの研究開発者が利用者など関係するステークホルダーに対しアカウンタビリティを果たすこと。

第六　開発原則の構成及び順序（2／7）

構成員からの指摘

【堀構成員】（第１回親会） 原則の構成、関係、順序等について、更に検討し、整理すべきではないか。

【平野構成員】（第１回開発原則分科会） 安全保護と原則とプライバシー保護の原則とのように、複数の原則の間で牴触が生ずる可能性にも留意すべきではないか。

【萩田構成員】（第１回開発原則分科会） 複数の原則を同時に又は段階的に破るリスク要因を考慮する必要があるのではないか。

【大屋構成員】（第１回開発原則分科会） 危害だけを規制根拠とするのは狭いとしても、あらゆる不快を規制根拠としてしまうと何もできなくなる。不快については、その継続期間や回避可能性に応じて、対応の在り方を検討していくべき。

【平野構成員】（第１回親会） 今後のAIは、色々なAIが複雑に絡み合っていくもので、相互接続性が重要である。利用者の使い方等も重要な問題であり、利活用の場面を考えることは重要である。開発ガイドラインの策定に向けた検討に当たっては、AIネットワークシステムの相互接続性・相互運用性についても議論の対象とすべきであるとともに、利活用に当たり留意すべき事項についても併せて検討すべきである。

【林（秀）構成員】（第１回影響評価分科会）AIネットワークシステム相互間の相互接続性・相互運用性を技術的に確保するのみならず、AIネットワークシステム相互間の連携を円滑化させるための調整の仕組みを考えるべき。

【林（秀）構成員】（第１回影響評価分科会） AIのみならず、AIネットワークシステムを流通させるデータの形式の標準化を考えるべき。

論点

1．AIネットワーク化検討会議「報告書2016」における開発原則の構成及び順序は、次頁のように一応整理できるのではないか。

2．主にAIネットワーク化の健全な進展の促進及びAIネットワークシステムの便益の増進の観点から、AIネットワークシステムの利活用の分野間の連携の可能性も踏まえつつ、相互接続性・相互運用性の確保等AIネットワークシステム相互間の円滑な連携の確保に関しAIの開発段階において開発者が留意すべき事項について、これを「連携の原則」【仮称】として開発原則に追加することとしてはどうか。（後述）

3．透明性の原則、制御可能性の原則、セキュリティの原則、安全保護の原則、プライバシーの原則及び倫理の原則は、いずれも、主にAIネットワークシステムのリスクの抑制に関連する原則として整理できるのではないか。

　これに対し、利用者支援の原則及びアカウンタビリティの原則は、AIネットワーク化の健全な進展の促進並びにAIネットワークシステムの便益の増進及びリスクの抑制のいずれにも関連する原則として整理できるのではないか。

第六　開発原則の構成及び順序（3／7）

論点

4．上記2．及び3．並びに開発ガイドラインの「目的」における開発原則の説明「AIネットワーク化の健全な進展の促進並びにAIネットワークシステムの便益の増進及びリスクの抑制に関し留意すべき事項に関する原則」の記載順に鑑みると、開発原則の構成及び順序は、31頁に掲げる「（たたき台）」のように、

① 主にAIネットワーク化の健全な進展の促進及びAIネットワークシステムの便益の増進に関連する原則（連携の原則【仮称】）
② 主にAIネットワークシステムのリスクの抑制に関連する原則（「透明性の原則」～「倫理の原則」）
③ AIネットワーク化の健全な進展の促進等及びAIネットワークシステムのリスクの抑制のいずれにも関連する原則（「利用者支援の原則」、「アカウンタビリティの原則」）

の順とした上で、③の中では、①及び②と同じくAIの機能に関する原則たる「利用者支援の原則」、①及び②とは異なり開発者の人的責任に関する原則たる「アカウンタビリティの原則」の順とすることが考えられるのではないか。

5．上記4．に掲げる開発原則の順序（①主にAIネットワーク化の健全な進展の促進等に関連する原則（「連携の原則」）、②主にリスクの抑制に関連する原則（「透明性の原則」～「倫理の原則」）、③AIネットワーク化の健全な進展の促進等及びリスクの抑制のいずれにも関連する原則（「利用者支援の原則、「アカウンタビリティの原則」））に関し、上記4．の「（たたき台）」に対する別案として、

① 主にAIネットワークシステムのリスクの抑制に関連する原則（「透明性の原則」～「倫理の原則」）
② 主にAIネットワーク化の健全な進展の促進及びAIネットワークシステムの便益の増進に関連する原則（連携の原則【仮称】）
③ AIネットワーク化の健全な進展の促進等及びAIネットワークシステムのリスクの抑制のいずれにも関連する原則（「利用者支援の原則」、「アカウンタビリティの原則」）

の順とし、③の中の順序は4．における順と同じ順とすることも考えられるところ。

上記4．の「（たたき台）」の順序及び上記5．の「たたき台（別案）」の順序の長所及び短所の比較は、33頁の表のとおり。

6．開発原則の項目相互間の抵触の可能性を踏まえ、開発原則の項目相互間の優先順位又は調整に関し別段の規定を設けるべきか。

この点に関し、「第四　AI開発ガイドラインの目的、基本理念等」論点2．に記したように、「関係する価値・利益のバランスの確保」を開発ガイドラインの策定及び解釈に当たっての「基本理念」を構成する事項の一つとして掲げる場合には、「基本理念」の書き方次第では、この基本理念こそが開発原則の解釈に当たっての項目相互間の調整に関する指針を示すものとして機能し得るため、開発原則の項目相互間の優先順位又は調整に関し別段の規定を設ける必要が乏しいと整理することもできるのではないか。

7．AIネットワークシステムの用途固有の事情により分野共通開発ガイドラインに定める開発原則の項目の一部又は全部をそのまま適用すべきでない場合には、当該項目について当該用途に係る特則を分野別開発ガイドラインに定めることとしてはどうか。

第六　開発原則の構成及び順序（4／7）

【論点１．関連】 AIネットワーク化検討会議「報告書2016」（抄）

- ① **透明性の原則**
- ② **利用者支援の原則**
- ③ **制御可能性の原則**

→ 事故等の防止及び事故等の後の対処に資する原則

- ④ **セキュリティ確保の原則**
- ⑤ **安全保護の原則**

→ 主に工学的な問題意識に由来する原則

- ⑥ **プライバシー保護の原則**
- ⑦ **倫理の原則**

→ 主に法的又は倫理的な問題意識に由来する原則

→ AIの機能に関する原則

- ⑧ **アカウンタビリティの原則**

→ AIの開発時及び開発後における開発者の人的責任に関する原則

第六　開発原則の構成及び順序（5／7）

【論点4. 関連】 連携の原則【仮称】の追加を含む開発原則の構成及び順序（たたき台）

開発原則の項目相互間の抵触の可能性を踏まえ、**開発原則の項目相互間の優先順位又は調整に関し別段の規定を設けるべきか**。　【論点6.】

Ⅰ　ＡＩの機能に関する原則

(1) 主にＡＩネットワーク化の健全な進展の促進及びＡＩネットワークシステムの便益の増進に関連する原則

- **①　連携の原則【仮称】**

(2) 主にＡＩネットワークシステムのリスクの抑制に関連する原則

- **②　透明性の原則**
- **③　制御可能性の原則**

　→ 事故等の防止及び事故等の後の対処に資する原則

- **④　セキュリティ確保の原則**
- **⑤　安全保護の原則**

　→ 主に工学的な問題意識に由来する原則

- **⑥　プライバシー保護の原則**
- **⑦　倫理の原則**

　→ 主に法的又は倫理的な問題意識に由来する原則

(3) (1)及び(2)に掲げる原則を補完する原則

- **⑧　利用者支援の原則**

Ⅱ　Ⅰに掲げる原則に関連し、ＡＩの開発者がステークホルダーに対し果たすべき責任に関する原則

- **⑨　アカウンタビリティの原則**

第六　開発原則の構成及び順序（6／7）

【論点4. 関連】 連携の原則【仮称】の追加を含む開発原則の構成及び順序（たたき台）【別案】

開発原則の項目相互間の抵触の可能性を踏まえ、**開発原則の項目相互間の優先順位又は調整に関し別段の規定を設けるべきか**。　【論点6.】

Ⅰ　ＡＩの機能に関する原則

(1) 主にＡＩネットワークシステムのリスクの抑制に関連する原則

- ①　**透明性の原則**
- ②　**制御可能性の原則**

　→ 事故等の防止及び事故等の後の対処に資する原則

- ③　**セキュリティ確保の原則**
- ④　**安全保護の原則**

　→ 主に工学的な問題意識に由来する原則

- ⑤　**プライバシー保護の原則**
- ⑥　**倫理の原則**

　→ 主に法的又は倫理的な問題意識に由来する原則

(2) 主にＡＩネットワーク化の健全な進展の促進及びＡＩネットワークシステムの便益の増進に関連する原則

- ⑦　**連携の原則【仮称】**

(3) (1)及び(2)に掲げる原則を補完する原則

- ⑧　**利用者支援の原則**

Ⅱ　Ⅰに掲げる原則に関連し、開発者が利用者等ステークホルダーに対し果たすべき責任に関する原則

- ⑨　**アカウンタビリティの原則**

第六　開発原則の構成及び順序（7／7）

	長所	短所
（たたき台） ①連携の原則【仮称】 ②透明性の原則 ③制御可能性の原則 ④セキュリティ確保の原則 ・・・ ⑦倫理の原則 ⑧利用者支援の原則 ⑨アカウンタビリティの原則	○分野共通開発ガイドラインの目的規定における「AIネットワーク化の健全な進展の促進並びに・・・便益の増進及びリスクの抑制」の記述順と整合すること。 ○若干数の原則を例示する際に、記載順に即して例示すれば、 ・連携の原則、透明性の原則等 ・連携の原則、透明性の原則、制御可能性の原則等」 となり、主に「AIネットワーク化の健全な進展の促進」等に関連する原則（①）及び主に「リスクの抑制」に関連する原則（②～⑦）の双方から例示の対象がごく自然に選ばれるため、開発原則の性格がバランスよく描写されること。	○連携の原則は、AIとAI、AIと情報通信ネットワーク等における相互接続性・相互運用性の確保等AIネットワークシステム相互間の円滑な連携の確保を図る原則であるが、相互接続性・相互運用性や連携は必ずしもAIならではの問題ではないことから、必ずしもAIならではの問題ではないことに関する原則を原則群の筆頭に掲げるものであること。 ※　そもそも開発ガイドラインが、AI自体の便益及びリスクのみに着目するものではなく、AIネットワーク化の進展及びAIネットワークシステムの便益及びリスクに着目するものであることに鑑みると、Aiならではの問題に関する原則を筆頭に掲げる必然性は絶対的ではないのではないか。
（たたき台）【別案】 ①透明性の原則 ②制御可能性の原則 ③セキュリティ確保の原則 ・・・ ⑥倫理の原則 ⑦連携の原則【仮称】 ⑧利用者支援の原則 ⑨アカウンタビリティの原則	○AIならではの問題である透明性及び制御可能性に関する原則を原則群の筆頭に掲げるものであること。	○分野共通開発ガイドラインの目的規定における「AIネットワーク化の健全な進展の促進並びに・・・便益の増進及びリスクの抑制」の記述順と整合しないこと。 ○若干数の原則を例示する際に、記載順に即して例示すれば、 ・透明性の原則、制御可能性の原則、セキュリティ確保の原則等 ・透明性の原則、制御可能性の原則等 等となり、例示の対象が主に「リスクの抑制」に関連する原則に偏ってしまうこと。

「第六　開発原則の構成及び順序」（5／7）に掲げる「（たたき台）」に掲げる開発原則の構成及び順序

○　連携の原則【仮称】→　後掲「第八　連携の原則【仮称】」

(1) 透明性の原則

(2) 制御可能性の原則

(3) セキュリティ確保の原則

(4) 安全保護の原則

(5) プライバシー保護の原則

(6) 倫理の原則

(7) 利用者支援の原則

(8) アカウンタビリティの原則

第七（1）　透明性の原則（1／3）

AIネットワーク化検討会議「報告書2016」（抄）

① **透明性の原則**　AIネットワークシステムの**動作の検証可能性及び説明可能性を確保**すること。

ア　**動作の検証可能性**の確保

(ア) 動作の記録及び確認のための技術の在り方の検討

(イ) 評価関数及び推論メカニズムの透明化

(ウ) アルゴリズムのブラックボックス化の回避

イ　**動作の説明可能性**の確保

(ア) AIの特性に応じた説明能力・説明機能の付与

(イ) 獲得表象の記号化及び解読のための技術の在り方の検討

構成員からの指摘

■透明性の原則全般について

【堀構成員】（第１回親会）　**ガイドラインの一番基本で最低限要請されるべき要素として、何をやっているのかわかり、どういうメカニズムでそうやっているのかわかるという意味での透明性が求められる**。もっとも、**透明性にもいろいろなレベルがあり、利用目的に応じて求められる透明性も異なってくる**。

【高橋構成員】（第１回開発原則分科会）　**AIのアーキテクチャの類型（機械学習、記号推論、強化学習等）に応じて、透明性の確保の在り方を検討する必要**がある。また、技術の高度化に従って、**複合的アーキテクチャが発展しつつある点にも留意する必要**がある。

【高橋構成員】（第１回開発原則分科会）　**AIのアーキテクチャにかかわらず、透明性が求められる共通の対象として、ソースコード、行動ログ、通信ログ、判断ログ、根拠が挙げられる**。

第七（1）　透明性の原則（2／3）

構成員からの指摘

■透明性の原則全般について（続き）

【久世構成員】（第１回開発原則分科会） 透過性・制御可能性を確実に確保することは難しいが、何かをしなければならない。ガイドラインの検討に当たっては、最終的な判断における人間の介在の有無、AIの具体的な利用方法等に留意して、きめ細やかに検討していく必要がある。

【丸山構成員】（第２回開発原則分科会） 透明性のハードルを高くすることにより、AIの開発が抑制されないように留意する必要がある。ディープラーニング等における判断のブラックボックス化なども踏まえ、完全な透明性ではなく、「合理的な透明性」を確保すべきとしてはどうか。

■ア　動作の検証可能性の確保について

【堀構成員】（第１回親会） 問題が生じた場合の説明が可能なように必要なデータを取っておき、判断のプロセスを記録しておくことが必要。ネットワーク上に追跡をミッションとするAIの投入も考える必要がある。

【平野構成員】（第１回開発原則分科会） 責任の帰属を解明するためのみならず、製品安全を向上させるためにも、動作の追跡可能性を確保することが重要。

【高橋構成員】（第１回開発原則分科会） 事後検証のために行動ログ・通信ログが必要となる。

■イ　動作の説明可能性の確保について

【堀構成員】（第１回親会） アカウンタビリティを確保する観点からは、AIが自らの判断や動作を技術的に説明できることが望ましい。

【高橋構成員】（第１回開発原則分科会） AIに自己説明能力や判断ログを実装することなどにより解釈可能性を確保することが望ましいが、AIのアーキテクチャによっては実現困難な場合もある。

第七（1）　透明性の原則（3／3）

論点

1．透明性の原則の目的として、**利用者等によるAIネットワークシステムへの信頼を獲得するとともに、事故時等における責任の検証を可能とするために、透明性が必要**である旨を説明すべきではないか。（［House of Commons 2016］等参照）

2．**動作の透明性（検証可能性及び説明可能性）が要請されるAIの動作の範囲如何。入出力、通信及び判断としてよいか**。

- AIのアーキテクチャ（機械学習、強化学習、記号推論等）に関しては、急速な技術発展が見込まれ、各種のアーキテクチャ間の複合も進んでいることから、特定のアーキテクチャに対応した透明性の仕様については定めずに、**AIのアーキテクチャにかかわらず透明性を求められるべき要素（入出力・通信・判断）を列挙すべき**ではないか。

3．個人の生命・身体の安全等**重要な権利利益若しくは法益に関するリスクを惹起し得る、又は個人に関する重大な決定のために利活用されるAIネットワークシステムの構成要素となり得るAIに関し**、

- 当該AIの動作のうち、**入出力及び通信については検証可能性を確保すべき**とし、**判断については、深層学習等における判断過程のブラックボックス化が指摘されていることなどに鑑み、技術的及び経済的な事情に鑑み合理的な範囲・水準で、検証可能性を確保するよう努めるべき**としてはどうか（［堀 2015］、［Whitehouse 2016］等参照）。

- **技術的及び経済的な事情に鑑み合理的な範囲・水準で、動作の説明可能性を確保するよう努めるべき**としてはどうか（［堀 2015］、［松尾ほか 2015］等参照）。

4．AIネットワークシステムの**動作の検証可能性及び説明可能性を確保するための有力な手段として**、当該**システムの動作の検証可能性及び説明可能性を実現するAIを開発することが考えられる**のではないか。

5．透明性の確保の在り方について定める上では、**プライバシー等個人の権利利益、営業秘密等企業の権利利益及び通信の秘密との調整にも留意すべき**ではないか。

第七（2）　制御可能性の原則（1／3）

AIネットワーク化検討会議「報告書2016」（抄）

③ 制御可能性の原則 AIネットワークシステムの**制御可能性を確保**すること。

ア　制御可能性に関する**リスク評価**

(ア) 情報通信ネットワーク上に多種多様なAIが混在することによりAIネットワークシステムが正常に動作せず意図しない事象が生ずるリスクの評価

(イ) 利用者又は第三者による改修によりAIが正常に動作せず意図しない事象が生ずるリスクの評価

(ウ) AIの自己改修によりAIネットワークシステムが正常に動作せず意図しない事象が生ずるリスクの評価

イ　**制御可能性の設計及び実装**

(ア) AIの能力の制御の在り方の検討（例：外界及び情報通信ネットワークへのアクセスの制御、能力の限定、緊急時の停止機能等）

(イ) AIの動機の制御の在り方の検討（ルール及び目標の設定、価値判断の手順の設定、報酬関数の設定等）

(ウ) AIネットワークシステムの動作の整合性の確保

ウ　**制御可能性マネジメント**

(ア) AIネットワークシステムにおける制御権の配分の在り方の検討

(イ) 仮想化技術を用いたネットワークの分離によるAIの制御の在り方の検討

(ウ) ネットワークがAIの制御範囲から分断されたときのAIの制御の在り方の検討

第七（2）　制御可能性の原則（2／3）

構成員からの指摘

■制御可能性の原則全般について

【堀構成員】（第１回親会） 動作の制御可能性、AIの評価関数の変更可能性、可逆性の確保などにより、「可能な限り制御不能な状態に落ち込むのを防ぐために打てる手は打つべし」。

【栗原構成員】（第１回開発原則分科会） 対象となるAIネットワークシステムの可読性の有無、規模、トップダウン／ボトムアップ、能動性／受動性など特性を踏まえ制御可能性の確保の在り方を検討する必要がある。特に、今後開発が進むと見込まれるボトムアップ型・群知能型システムの場合には、個の振る舞いと群の振る舞いに大きな開きが生じる可能性に留意する必要がある。

【久世構成員】（第１回開発原則分科会）　透過性や制御可能性を確実に確保することは難しいが、何かをしなければならない。ガイドラインの検討に当たっては、最終的な判断における人間の介在の有無、AIの具体的な利用方法等に留意して、きめ細やかに検討していく必要がある。

【平野構成員】（第１回開発原則分科会） AIの自律性又は創発性により、設計者等が予見不可能な事故が生じることにより、責任の空白地帯が生じる可能性にも留意すべきではないか。

■ア　制御可能性に関するリスク評価について

【城山構成員】（第１回親会） リスク評価のためにも知的生産の促進は不可欠であり、そのために実験法制が必要となる。

【栗原構成員】（第１回開発原則分科会） AIの制御可能性を評価するために、箱庭（AI牧場）やシミュレーション空間など限定された範囲でAIの制御可能性について実験を行うことが必要となるのではないか。

【高橋構成員】（第１回開発原則分科会） 行動原理に基づいて自律的に学習・動作するAIについては、AIが自身の報酬関数を操作することにより設計者等が意図しない動作をするリスク（報酬ハッキング）も考慮する必要がある。

■イ　制御可能性の設計及び実装について

【堀構成員】（第１回親会） AIネットワークシステム全体としての制御可能性を保つためには、他のAIを制御することを目的としたAIを投入することが必要となり得るのではないか。

【栗原構成員】（第１回開発原則分科会） 制御可能性を確保する上では、機械に機械をチェックさせることが突破口となる可能性があるのではないか。

第七（2）　制御可能性の原則（3／3）

構成員からの指摘

■ウ　制御可能性マネジメントについて

【堀構成員】（第１回親会） 制御可能性を分散的に確保する手段として、**市民への制御権の分散や、個人のためのガーディアンAIの利用という選択肢もあり得る**のではないか。

【板倉構成員】（事務局へのメールによる御意見） AIの緊急停止機能に関し、**緊急停止による事故や混乱を避けるために、緊急停止時のプロセスや人間への制御の移譲の在り方についても検討する必要**がある。また、 緊急停止機能等**緊急時の制御可能性の確保に関する技術については、標準化すべき要請が高い**のではないか。

論点

1．AIネットワークシステムの制御可能性を評価し確保するため、その構成要素となり得る**AIについて、予め制御可能性の検証**（verification）[※形式的な整合性の検証]**及び妥当性確認**（validation）[※実質的な妥当性の確認]**を行うことが必要となる**のではないか（ [FLI 2015] 、[National Science and Technology Council 2016]、[House of Commons 2016]等参照）。

- **その際には、報酬ハッキング（AIが与えられた目標を形式的に達成するために設計者の意図に実質的に反する動作をすること）**（[Amodei et al. 2016]等参照）、**ファンクション・クリープ（設計時に想定されていた用途と異なる用途で用いられることにより意図しない事象が生ずるリスク）**（[新保 2016]等参照）、**利用者又は第三者による改修の結果AIが正常に動作せず意図しない事象が生ずるリスク**、**多種多様なAIが混在することによりAIネットワークシステムが正常に動作せず意図しない事象が生ずるリスク等に特に留意して、対策を講ずべき**としてはどうか。

2．**制御不能となるリスクにつき、その蓋然性が高い又は不確実と考えられるAIについては、一般社会で利用される前に、実験室等閉鎖された空間において、当該空間の外につながる情報通信ネットワークシステムに接続せずに、AIの制御可能性について実験を行い、リスク評価を行うことにより、制御可能性を確保**することが求められるのではないか（ [Bostrom 2014] 等参照）。

- その際には、AIを**実験室等閉鎖された空間から外部の予期しない出来事が起こり得る開かれた一般社会に安全に移行することができるよう、措置を講ずることも求められる**のではないか（ [Whitehouse 2016]等参照）。

3．AIネットワークシステムの制御可能性を継続的に確保するために、その構成要素となり得る**AIについて人間又は信頼し得る他のAIによる監督及び対処（停止、切断、修理等）の実効性の確保が求められる**のではないか（（ [Bostrom 2014] 、[堀 2015] 、[松尾ほか 2015]、[Orseau & Armstrong 2016] 、[Amodei et al. 2016]等参照 ）。

- **緊急停止機能に関する技術標準やプロセスについても指針を定めるべきではないか**（ [Orseau et al. 2016] 等参照）。
- **AIによる自己改良や他のAIの開発に対する監督及び対処の在り方如何**（ [一杉 2014] 、[National Science and Technology Council 2016]等参照）。

第七（3）　セキュリティ確保の原則（1／2）

AIネットワーク化検討会議「報告書2016」（抄）

④ セキュリティ確保の原則　AIネットワークシステムの頑健性及び信頼性を確保すること。

ア　セキュリティに関する**リスク評価**

（ア）AIネットワークシステムの機密性、完全性、可用性に対するリスクの評価

（イ）AIネットワークシステムのセキュリティが損なわれることにより、利用者及び第三者の生命・身体の安全に危害が及ぶリスクの評価

イ　セキュリティの**設計及び実装（セキュリティ・バイ・デザイン）**

（ア）情報セキュリティの3要素（機密性、完全性、可用性）の確保

（イ）利用者及び第三者の生命・身体の安全に危害を及ぼす可能性のあるセキュリティ上の脅威・脆弱性への対処

（ウ）攻撃耐性の確保

1．対攻撃強度の在り方の検討

2．サイバー攻撃やセンサー攪乱攻撃等に対する耐性の確保

3．現実空間での物理的攻撃への耐性の確保

ウ　**セキュリティ・マネジメント**（予防、検出、対応、システムの復旧、継続的な保守、レビュー及び監査等）

構成員からの指摘

【高橋構成員】（第1回開発原則分科会）　**行動原理に基づいて自律的に学習・動作するAIについては、報酬ハッキングなどの新たなセキュリティリスクが発生**する可能性がある。

【湯淺構成員】（第2回開発原則分科会）　AIネットワークシステムは、**第三者による緊急停止等制御を要請し得ることから、情報の機密性・完全生・可用性を要素とする従来の情報セキュリティの概念の見直しを迫る可能性**があるのではないか。

【新保構成員】（第2回開発原則分科会）AIのセキュリティにおいては**事故や攻撃に対する物理的な耐性も重要**である。

第七（3）　セキュリティ確保の原則（2／2）

論点

1. セキュリティ確保の原則においては、**OECDセキュリティガイドライン等セキュリティに関する国際的に共有された枠組みを参照しつつ、AIネットワークシステムに特有のセキュリティ問題につき、対応するための指針を定める**べきではないか。

2. **AIネットワークシステムに要求されるセキュリティの範囲**については、当該システムにおいて処理・伝達される**情報の機密性、完全性及び可用性の維持のみならず**、当該システムの**信頼性**（設計者及び利用者の意図した通りに動作が行われ、権限を有しない第三者による操作を受けないこと）**及び頑健性**（物理的な攻撃や事故への耐性）**の維持も含まれる**としてはどうか。

3. AIネットワークシステムのセキュリティを評価し確保するため、その構成要素となり得るAIについて、**予めセキュリティの検証及び妥当性確認を行うことが必要となる**のではないか（［FLI 2015］、［National Science and Technology Council 2016］、［House of Commons 2016］等参照）。

   - その際には、**AIの自律的学習・動作により設計者の予見し難い脆弱性が生じるリスク**や**AIネットワークシステムのセキュリティ侵害により物理的な被害が発生することで利用者等の生命・身体の安全に生ずるリスク**等**AIネットワークシステムに特有のセキュリティ上のリスクに特に留意して、対策を講ずべき**としてはどうか。

4. AIネットワークシステムのセキュリティを確保することができるよう、その構成要素となり得る**AIの設計段階において措置（セキュリティ・バイ・デザイン）を講ずべき**としてはどうか。

5. AIネットワークシステムの**セキュリティを継続的に確保するために**、当該**AIネットワークシステム及び**その構成要素となり得る**AIについて継続的な監督及び調整が可能となるよう設計するとともに、相互診断・相互調整の在り方を検討すべき**ではないか（［National Science and Technology Council 2016］等参照）。

第七（4）　安全保護の原則（1／2）

AIネットワーク化検討会議「報告書2016」（抄）

⑤ **安全保護の原則**　AIネットワークシステムが**利用者及び第三者の生命・身体の安全**に危害を及ぼさないよう**に配慮**すること。

ア　安全に関する**リスク評価**

イ　安全保護の**設計及び実装（セーフティ・バイ・デザイン）**

（ア）利用者及び第三者の**安全の保護に配慮したプログラム設計**の在り方の検討

（イ）**本質安全の確保**（事故の被害を抑制するために、AIネットワークシステムの特性に応じて、本質的な危険要因を必要最小限に抑えること）

ウ　**安全マネジメント**（予防、検出、対応、継続的な保守、レビュー及び監査等）

構成員からの指摘

【平野構成員】（第１回開発原則分科会）**開発者は**自動走行車等AI・ロボットの**事故時に備えて設計段階において多種多様な権利・利益を衡量して倫理的選択を行うことを迫られる**可能性があるが、**開発者はAI・ロボットの設計においていかなる倫理的選択を行ったのかについて利用者等に説明すべき**ではないか。

【萩田構成員】（第１回開発原則分科会）　**ロボットがネットワークを通じてサイバー・フィジカル空間を横断して連結することにより安全性に及ぼすリスク**について（計量可能な形で）**評価すべき**ではないか。

【萩田構成員】（第２回開発原則分科会）　AIネットワークシステムの**安全性を確保するためには、開発者が設計段階で措置を講ずるのみならず、社会における多様な主体のコンセンサスを形成し協力を得る必要**がある。

第七（4）　安全保護の原則（2／2）

論点

1. 安全保護の原則が適用されるAIの範囲は、個人の生命・身体の安全に関するリスクを惹起し得るAIネットワークシステムの構成要素となり得るAIとしてはどうか。

2. AIネットワークシステムの安全性を評価し確保するため、その構成要素となり得るAIについて、予め安全性の検証及び妥当性確認を行うことが必要となるのではないか（［FLI 2015］、［National Science and Technology Council 2016］、［House of Commons 2016］等参照）。

   - その際には、AIネットワーク化によりAI・ロボットがサイバー空間・現実空間を越えて連結することにより生ずる利用者等の生命・身体の安全への体系的・複合的なリスクに特に留意して、対策を講ずべきとしてはどうか。

3. AIネットワークシステムにおける本質安全（運動能力等の抑制）、制御安全（監視装置等の実装）、機能安全等を確保することができよう、その構成要素となり得るAIの設計段階において措置（セーフティ・バイ・デザイン）を講ずべきとしてはどうか（［一杉 2014］、［向殿 2016］等参照）。

4. AIネットワークシステムの利活用の際の利用者及び第三者の生命・身体の安全に関する判断（例：生命・身体の安全を保護される個人の優先順位等に関する判断）を行うAIを研究開発する際には、開発者は利用者等関係ステークホルダーに対し当該判断を行うAIに関する設計の趣旨及び理由を説明すべきとしてはどうか（［平野 2016］等参照）。

5. AIと利用者等人間が協調して安全を確保することが可能となるよう、AI及び研究開発者が利用者に対し適時適切に情報提供を行うべきとしてはどうか（［向殿 2016］等参照）。

6. AIネットワークシステムの安全性を継続的に確保するために、当該AIネットワークシステム及びその構成要素となり得るAIについて継続的な監督及び調整が可能となるよう設計するとともに、相互診断・相互調整の在り方を検討すべきではないか（［National Science and Technology Council 2016］等参照）。

第七（5）　プライバシー保護の原則（1／2）

AIネットワーク化検討会議「報告書2016」（抄）

⑥ **プライバシー保護の原則**　AIネットワークシステムが**利用者及び第三者のプライバシーを侵害しないように配慮**すること。

ア　プライバシー**影響評価**

イ　プライバシー保護の**設計及び実装（プライバシー・バイ・デザイン）**

（ア）空間プライバシー（私生活の平穏）の保護：私生活の領域へのロボット等の侵入の制御、ロボット等による私生活の領域の監視の制御、ロボット等への不正アクセスの制御

（イ）情報プライバシー（パーソナルデータ）の保護：データの収集・分析・利活用の適正な制御、匿名化機能、暗号標準、アクセス・コントロール機能等の実装

（ウ）生体プライバシー（生体情報）の保護：脳情報など生体情報の収集・分析・利活用の適正な制御

ウ　**プライバシー・マネジメント**（予防、検出、対応、継続的な保守、レビュー及び監査等）

構成員からの指摘

【平野構成員】（第１回開発原則分科会）**プライバシーと安全等他の利益との調整の在り方についても検討する必要**があるのではないか。

【新保構成員】（第2回開発原則分科会）AIによる**プロファイリングは、情報プライバシーのみならず、個人の自己決定という意味でのプライバシー権も侵害するリスクを有していることから、AIの開発段階においても特に留意すべき問題**ではないか。

第七（5）　プライバシー保護の原則（2／2）

論点

1．プライバシー保護の原則においては、**OECDプライバシーガイドライン等プライバシーに関する国際的に共有された枠組みを参照しつつ、AIネットワークシステムに特有のプライバシー問題につき、対応するための指針を定める**べきではないか。

2．プライバシー保護の原則において配慮されるべき**プライバシーの範囲には、空間プライバシー（私生活の平穏）、情報プライバシー（個人データ）、通信の秘密及び生体プライバシーが含まれる**としてはどうか（OECDプライバシーガイドライン・暗号政策ガイドライン等参照）。

3．プライバシー保護の原則又はガイドライン全体の総則において、**プライバシーと安全等他の利益との調整が求められる旨を説明すべきではないか。また、プライバシー保護の原則と透明性の原則（特に検証可能性の確保）との調整の在り方についても検討すべきではないか。**

4．AIネットワークシステムにおけるプライバシー侵害のリスクを評価するために、その構成要素となり得るAIについて**予めプライバシー影響評価を行うべき**としてはどうか。

- その際には、**AIネットワークシステムの種類・性質**（ロボットを構成要素とするシステム、プロファイリングを行うAIを構成要素とするシステム、脳と連携するシステム等）**に即して、各種のプライバシー**（空間プライバシー、情報プライバシー（パーソナルデータ）、通信の秘密及び生体プライバシー）**が侵害されるリスクを評価すべき**としてはどうか 。

5．AIネットワークシステムがその利活用に当たりプライバシーを保護できるものとなるよう、その構成要素となり得る**AIの設計段階において措置（プライバシー・バイ・デザイン）を講ずべき**としてはどうか。

- **プロファイリングの用に供するAIについては、プライバシー権の根底にある個人の自律を侵害し、差別を助長するリスクを有していることなどに鑑み、AIの設計段階において特に慎重に措置を講ずべき**としてはどうか。

6．AIネットワークシステムの**プライバシーを継続的に確保するために**、当該**AIネットワークシステム及び**その構成要素となり得る**AIについて継続的な監督及び調整が可能となるよう設計するとともに、相互診断・相互調整の在り方を検討すべき**ではないか（［National Science and Technology Council 2016］等参照）。

第七（6）　倫理の原則（1／1）

AIネットワーク化検討会議「報告書2016」（抄）

⑦ **倫理の原則**　AIネットワークシステムの**研究開発において、人間の尊厳と個人の自律を尊重**すること。

ア　AIへの**機械倫理の実装**の在り方の検討

イ　Brain Machine Interface（BMI）等により**人間の脳とAIの連携を図る際**の**人間の尊厳と個人の自律の尊重の在り方**の検討

構成員からの指摘

【鈴木構成員】（第１回親会）　目指すべき社会像・世界観を反映させたガイドラインを検討すべきである。また、AIの開発者もひとりの人間（生活者、利用者）であるという視点を持つべきであり、**生命倫理の議論も参照しつつ、歴史と文化を背負った人格としての人間（ペルソン）という存在を意識できるようなガイドラインの策定**に向けて取り組むことが重要である。

【新保構成員】（第１回開発原則分科会）　**AIが人間の価値や存在を超えてはいけないということや、AIが人になってはいけないということなどを意味するヒューマニティ・ファーストの理念も盛り込むべき**ではないか。

【河島構成員】（事務局へのメールによる御意見）　倫理の原則に、**社会的価値(公正性，公平性)に鑑みてAIを設計することも盛り込むべき**ではないか。例えば、AIの機械学習したデータに含まれる**偏見等に起因する差別に対処**することなどが求められるのではないか。データに含まれる偏見等に起因する差別に対処する上では、透明性やアカウンタビリティの確保も求められるのではないか。

【大屋構成員】（第２回開発原則分科会）　倫理の原則の方向性を明確にするために、**人間性を中心的な価値に据えるべき**ではないか。また、**個人の自律と平等も指針として明確に打ち出すべき**ではないか。

第七（6）　倫理の原則（2／2）

論点

1. 倫理の原則においては、**人間性（humanity）の価値を中心に据えつつ、人間の尊厳と個人の自律を尊重すべき**旨を掲げることとしてはどうか。
   - 人間が文化的・歴史的存在であることに鑑み、AIの研究開発において**開発者が多様なステークホルダーと対話することなどにより、文化の多様性を尊重するとともに、将来世代にも配慮すべき**としてはどうか（[Nadella 2016]等参照）。

2. **人間の尊厳を尊重する観点から、国際人権法・国際人道法等を参照しつつ、AIネットワークシステムが人間性の価値を毀損してはならない**旨を定めるべきではないか（[Whitehouse 2016]、[Partnership on AI 2016]等参照）。

3. 倫理の原則においては、個人の自律の尊重が盛り込まれているが、説明文等において、**個人の自律の尊重は（最終）利用者等関係する個人が自律した人格（person）として尊重されることを含意する**旨を説明すべきではないか。

4. **人間の脳・身体と融合又は連携するAIを研究開発する際には、人間の尊厳と個人の自律の尊重について、生命倫理等の議論も参照しつつ、特に慎重に配慮すべき**としてはどうか。

5. AIの開発において、**公正（fairness）等の価値に鑑みて、技術的に可能な範囲で、AIの学習するデータに含まれる偏見等に起因する差別（人種、性、宗教等による差別）を防止するための措置を講ずべき**としてはどうか（[Whitehouse 2016]、[House of Commons 2016]等参照）。

第七（7）　利用者支援の原則

AIネットワーク化検討会議「報告書2016」（抄）

② **利用者支援の原則**　　AIネットワークシステムが**利用者を支援し**、**利用者に選択の機会を適切に提供するよう配慮**すること。

ア　個人の**合理的選択を支援する機能（ナッジ等）の実装**
（ア）デフォルト、フィードバック、エラー対処等の在り方の検討
（イ）行為者にナッジを与える方法（適切な時期等）の検討
イ　人間の**認知能力の補完**
ウ　**ユニバーサル・デザイン**の確保

構成員からの指摘

【**近藤構成員**】（第１回親会）「利用者支援の原則」に「利用者に選択の機会を適切に提供するよう配慮すること」とあるが、**技術革新で選択の機会は少なくなっている**と思う。自動運転も介護ロボットも暮らしを劇的に変える。**利用できない人との格差を拡げないよう社会の受容性についても丁寧に議論してほしい。**

【**中西構成員**】（第１回開発原則分科会）　**人間とAIが効果的に意思疎通し協働するためにはインターフェースが重要となる。異常時に人間がバトンタッチできる形で知らせてくれるインターフェースや、状況に応じて人間に適時適切なフィードバックを与えることにより効果的な協働を可能にするインターフェースの開発が求められる。**

【**中西構成員**】（第１回開発原則分科会）　人間とAIとの望ましい役割分担の在り方を考慮して、**人間の関与を促すナッジを与えるタイミングを検討すべき。**

論点

１．（最終）利用者に操作されるAIネットワークシステムの構成要素となり得るAIについては、**利用者との円滑な協調を可能とするために、利用者に対し適時適切にその判断に資する情報を提供し、かつ、利用者にとって操作しやすいインターフェイスが利用可能となるよう設計すべき**ではないか（［National Science and Technology Council 2016］等参照）。

２．（最終）利用者に操作されるAIネットワークシステムの構成要素となり得るAIについては、**利用者に選択の機会を適時適切に提供する機能（ナッジ：例えば、デフォルトの設定、理解しやすい選択肢の提示・体系化、フィードバックの提供、緊急時の警告、エラーへの対処等）が利用可能となるよう設計すべき**ではないか（［Thaler & Sunstein 2008］等参照）。

３．（最終）利用者に操作されるAIネットワークシステムの構成要素となり得るAIについては、**ユニバーサル・デザイン等社会的弱者の受容可能性を高めるための取組に努めるべきとしてはどうか**。

第七（8）　アカウンタビリティの原則（1／2）

AIネットワーク化検討会議「報告書2016」（抄）

⑧ アカウンタビリティの原則

AIネットワークシステムの**研究開発者が**利用者など**関係するステークホルダーに対しアカウンタビリティを果たす**こと。

ア　**研究開発者による説明**・情報開示

イ　**関係するステークホルダーとのコミュニケーション**

構成員からの指摘

【堀構成員】（第１回親会）　検討会議が示した素案では、アカウンタビリティは、研究開発者への要求になっているが、**技術的には、AIが自らの判断や動作を説明できることが望ましい**。もっとも、いつ何を説明できるべきかは、利活用の目的により異なり得る。例えば、自動走行車の衝突回避設計の場合には、事前に利用者に設計原理を説明し、同意を得る必要があるだろう。

【三友副議長】（第１回親会）AIの市場をみると、供給側に情報が偏っていて情報の完全性が満たされていないと感じる。**全てをルール、ガイドライン等で抑え込むことは困難であり、情報の完全性を見据えた検討が必要**である。

【新保構成員】（第１回開発原則分科会）「ロボット法新保８原則」では、責任の原則を掲げたが、この**責任には法的責任と倫理的ないし道義的責任が含まれる**。**責任を果たす上では、アカウンタビリティだけではなく、リテラシーも必要**。

第七（8）　アカウンタビリティの原則（2／2）

論点

1. **開発者が利用者に対し説明責任を果たすとともに、第三者を含め広く社会に対し説明責任を果たすべき**としてはどうか。

- 開発者が説明責任を果たす上では、**特に利用者等に対し説明を行うべきであるほか、多様なステークホルダーと対話を行いその意見を聴取する等ステークホルダーの積極的な関与を得るべき**ではないか（［Partnership on AI 2016］等参照）。
- **サービスプロバイダ等を通じてAIネットワークシステムの機能が最終利用者に提供される場合には、当該AIネットワークシステムの構成要素となり得るAIの開発者は、サービスプロバイダ等に対し説明責任を果たすことを通じて、最終利用者に対し間接的に説明責任を果たす**こととしてはどうか。
- **開発者が説明責任を果たす観点からも、透明性の原則において定められるAIネットワークシステムの動作の説明可能性の確保が期待される**のではないか。

2. アカウンタビリティの原則においては、**AIネットワークシステムに関する責任の分担が明確にされるとともに**（［House of Commons 2016］等参照）、**開発原則の遵守状況につき開発者から説明を受けた利用者によるAIネットワークシステムへの信頼・期待が保護されるよう、利活用ガイドラインにおいて定める利用者の責任との関係に留意しつつ、開発者の責任の在り方について指針を示すべき**ではないか 。

第八　連携の原則【仮称】（1／3）

AIネットワーク化検討会議「報告書2016」（抄）

2．AIネットワーク化の進展に向けた協調の円滑化

【問題意識】

AIネットワーク化を円滑に進展させ、もってAIネットワークシステムを安心して安全に利活用する環境を実現するために、当事者間の競争関係の有無如何にかかわらず、イノベーティブな研究開発と公正な競争にも留意しつつ、AI相互間又はAIネットワークシステム相互間の協調を円滑化するための取組の在り方を検討することが求められる。

【今後の課題】

AI相互間又はAIネットワークシステム相互間の協調を推進するという観点から、AIネットワークシステムに関する相互接続性・相互運用性の確保に向けて検討を進めていくことに加え、関係する技術や市場の状況等を踏まえつつAIネットワーク化の円滑な進展のために必要となる範囲におけるAIネットワークシステムのオープン化の在り方を検討すべきである。

(1) AIネットワークシステムに関する相互接続性・相互運用性の確保

- 相互接続性・相互運用性を確保すべき対象の検討

（例）

  - ➢ アーキテクチャ
  - ➢ 情報の結節（AI相互間、AIとモノの間、AIと人間の間、AIとクラウドの間、API等）
  - ➢ 匿名化、暗号等
  - ➢ データの形式

- 相互接続性・相互運用性の確保の方法（dejure / defacto）の検討
- 相互接続性・相互運用性を確保するための結合テストの在り方の検討
- 相互接続性・相互運用性の確保に向けた国際協調の在り方の検討

(2) AIネットワーク化の円滑な進展のために必要となるAIネットワークシステムのオープン化の在り方の検討

- AIネットワークシステムのオープン化の動向の注視
- AIネットワーク化の円滑な進展のために必要となるオープン化の対象及び方法の検討
- 国際社会におけるAIネットワークシステムのオープン化の推進の在り方の検討

第八　連携の原則【仮称】（2／3）

構成員からの指摘

【平野構成員】（第１回親会） **今後のAIは、色々なAIが複雑に絡み合っていくもので、相互接続性が重要**である。また、利用者の使い方等も重要な問題であり、利活用の場面を考えることは重要である。**開発ガイドラインの策定に向けた検討に当たっては、AIネットワークシステムの相互接続性・相互運用性についても議論の対象とすべき**であるとともに、利活用に当たり留意すべき事項についても併せて検討すべきである。

【林（秀）構成員】（第１回影響評価分科会）AIネットワークシステム相互間の**相互接続性・相互運用性を技術的に確保するとともに、AIネットワークシステム相互間の連携を円滑化させるための調整の仕組みを考えるべき**。

【林（秀）構成員】（第１回影響評価分科会） AIのみならず、**AIネットワークシステムを流通させるデータの形式の標準化を考える**べき。

【林（秀）構成員】（個別の御指摘） AIネットワークシステムの構成要素となり得るAIに関する標準の形成及び標準必須特許の取扱いは、連携が可能となるAIネットワークシステム範囲を左右し得るものであるとともに、イノベーティブな研究開発及び市場におけるAIネットワークシステム・AIネットワークサービスに関する競争を左右し得るものであることから、**AIネットワークシステム相互間の円滑な連携の確保の見地並びにイノベーティブな研究開発及び公正な競争の確保の見地から、標準の形成過程及び標準必須特許の権利者たる開発者が留意すべき事項に関し検討すべき**ではないか。

【福井構成員】（第１回開発原則分科会）AIに関するデータ等の囲い込みを見据えて、**ガイドラインにおいて、オープン・サイエンス的な理念を盛り込むべき**ではないか。

【宍戸構成員】（第１回開発原則分科会）福井構成員が指摘された**オープン・サイエンス的な理念については、**智連社会の基本理念の一つとして掲げられている**「イノベーティブな研究開発と公正な競争」に対応**している部分があるのではないか。

第八　連携の原則【仮称】（3／3）

論点

1．AIネットワークシステム相互間の円滑な連携を確保することにより、市場における公正な競争を通じてAI相互間又はAIネットワークシステム相互間の円滑な協調を実現する等AIネットワークシステムの便益を増進し、もってAIネットワーク化の健全な進展を促進するとともに、及び連携に伴うリスクの抑制に資するため、AIネットワーク化検討会議「報告書2016」の提言を踏まえ、相互接続性・相互運用性の確保等**AIネットワークシステム相互間の円滑な連携の確保に関しAIの開発段階において開発者が留意すべき事項を「連携の原則」**【仮称】**として開発原則に追加**することとしてはどうか。

この場合において、**連携の原則**【仮称】**及びその説明において記すべき事項如何**。このことに関し、特に**次に掲げる事項如何**。

- ・AIネットワークシステム相互間の円滑な連携の確保の見地から、**AIに関し相互接続性・相互運用性の確保が期待される事項**
- ・AIネットワークシステム相互間の円滑な連携の確保の見地から、**AIに関し標準化が期待される事項**
- ・AIネットワークシステム相互間の円滑な連携の確保の見地から、**AIに関する標準の形成過程及び標準必須特許の取扱いに関し、オープン・サイエンスに関することも含め、権利者たる開発者が留意すべき事項**
- ・AIネットワークシステム相互間の円滑な連携の確保の見地から、**AIネットワークシステムのオープン化に資するAIの情報の開示に関し、オープン・サイエンスに関することも含め、開発者が留意すべき事項**

2．上記1．に関連し、相互接続性・相互運用性の確保等**AIネットワークシステム相互間の円滑な連携の確保に関しAIの開発段階において開発者が留意すべき事項に関連して、国、関係国際機関等に推奨すべき事項として分野共通開発ガイドラインの関連文書**（OECDのガイドラインであれば、理事会勧告の本紙）**に記すべき事項如何**。

3．上記1．及び2．に関連し、AIネットワークシステム相互間の連携がAIネットワークシステムの利活用に伴うものであることに鑑み、**AIネットワークシステム相互間の円滑な連携の確保に関しAIネットワークシステムの利活用の段階において利用者（特にAIネットワークサービスのプロバイダ）が留意すべき事項及び国、関係国際機関等に推奨すべき事項を**後述する**分野共通利活用ガイドライン及びその関連文書**（OECDのガイドラインであれば、理事会勧告の本紙）**にそれぞれ記す**こととしてはどうか。

なお、AI及びAIネットワークシステムが現時点においては揺籃期であることから、**上記1．～3．の事項に関し法的規制の創設を検討することは、現時点では時期尚早**ではないか。**少なくとも関連する弊害の蓋然性が顕著になるまでは**、分野共通開発ガイドライン及び分野共通利活用ガイドライン並びにそれぞれの関連文書により開発者及びAIネットワークサービスのプロバイダ等利用者が留意すべき事項並びに国、関係国際機関等に推奨すべき事項を国際的に共有した上で、**国、関係国際機関等は**、**関連する動向を注視して、動向やベストプラクティスに関する情報を国際的に共有**するとともに、**AIネットワークシステム相互間の連携をめぐる**紛争の発生状況等に応じて、**国内の紛争及び国境を越えた紛争の処理の在り方等を検討して所要の措置を講ずるにとどめるような謙抑的な姿勢であるべき**ではないか。

第九　開発原則の実効性の確保の在り方（1／2）

ホワイトハウス報告書「人工知能の未来に備えて」（抄・仮訳）

提言17：　個人に関する重要な決定を行うAIベースのシステムの利用支援として州及び地方政府に資金を交付する連邦政府機関に、連邦補助金で購入されるAIベースの製品やサービスが、十分に透明な方法で結果を生成するものであり、有効性と公平性に係る証拠により裏付けられるものであることを確保するため、補助金の条件の再検討を求める。

ホワイトハウス（国家科学技術会議　ネットワーキング・情報技術研究開発小委員会）「米国人工知能研究開発戦略」（概要）

■　連邦政府の予算によるAI研究（連邦政府のAI研究のみならず、連邦政府の助成を受けた大学等のAI研究も含む。）の方針を策定するもの。

■　本戦略は、連邦政府の予算によるAI研究の究極の目標として、社会に便益をもたらす新たなAIに関する知識及び技術を生み出しつつ、ネガティブな影響を最小化することを提示。

■　本戦略は、この目標を実現するために優先的に取り組むべき事項を次のように設定。

（1）AI研究に対し長期的投資を実施

（2）人間とAIの協働に向けて効果的な方法を開発
　人間とAIシステムとの効果的なインタラクションを創出

（3）AIの倫理的・法的・社会的含意を理解し、対処
　倫理的、法的及び社会的目標に合致するAIシステムを設計する方法を開発するための研究が必要

（4）AIシステムの安全性及びセキュリティを確保
　AIシステムが広く利用されるようになる前に、当該AIシステムが安全かつ堅牢で、制御され、十分に定義義され、かつ十分に理解された方法で動作するものであることをあらかじめ確保

（5）共有される公共的なデータセット及びAIの訓練・試験のための環境を開発

（6）基準及びベンチマークを通じてAI技術を計測し評価

（7）国家のAI研究開発人材のニーズをより良く理解

第九　開発原則の実効性の確保の在り方（2／2）

論点

○**分野共通開発ガイドラインの関連文書**（OECDのガイドラインであれば、理事会勧告の本紙）**において、国、関係国際機関等に対し、開発原則の実効性の確保のための方策として、例えば次に掲げる方策を検討するよう推奨**する旨を記すこととしてはどうか。

・**公共調達の対象とするAI及び公的研究費の交付対象とするAIに関し、開発原則を踏まえて条件を設定**することにより、開発原則に適合するAIの研究開発を促進する方策

・**市場の機能を活用して、開発原則に適合しているAIが市場において利用者に選択されやすくなる環境を整備**し、もって**開発者に対し開発原則の遵守への誘因を付与**する方策

（→「第十　開発原則の実効性の確保における市場の活用の在り方」参照。）

○**分野共通開発ガイドラインの関連文書**（OECDのガイドラインであれば、理事会勧告の本紙）**において、各国の関係機関、関係国際機関等に対し、推開発原則の実効性に関する状況、実効性の確保に関するベストプラクティス等に関する情報を共有し、相互に協力するよう推奨**する旨を記すこととしてはどうか。

第十　開発原則の実効性の確保における市場の活用の在り方（1／3）

構成員からの指摘

【堀構成員】（第１回親会）AIの自由市場を保ち、人々が選ぶ中で自然によいものが生き残るというようにしたい。**ガイドラインを遵守して開発されたAIが利用者に好まれることにより、遵守するとメーカーも得をするようなガイドラインを作るべき**。暗号の場合と同様に、AIの研究も抑え込もうとしても抑えきれないので、**基本的にはオープンにするというのが第一原則**になるだろう。

【三友副議長】（第１回親会）AIの市場をみると、供給側に情報が偏っていて情報の完全性が満たされていないと感じる。**全てをルール、ガイドライン等で抑え込むことは困難であり、情報の完全性を見据えた検討が必要**である。

【大屋構成員】（第１回親会）堀構成員の発表にあるように、**市場原理に委ねて人々の選択により良い社会ができることが望ましいという考え方に賛成**であるが、人々の選択に委ねると少数者の利益や人権が侵害される可能性がある。**他の利益と比較衡量してバーターにしてよい領域と、人権等他の利益とのバーターにすべきでない領域との峻別を考える必要**がある。

【大屋構成員】（第１回開発原則分科会） 事前に定めた倫理原則に従って意思決定することは、人間にもAIにも困難である。人間にもできないことをAIに要求することは、AIの成長を阻害する。**被害が生じたとき、それに対して責任を負う者が誰もいないという事態（Liability Gap）を回避するために原則が要請されるという問題意識から原則を策定すべき**ではないか。

第十　開発原則の実効性の確保における市場の活用の在り方（2／3）

論点

1．**利用者**（AIネットワークサービスのプロバイダ、最終利用者）**が市場においてAI**（注）**を選択する際には、開発原則への適合性のみならず、機能、用途、価格等を踏まえつつ、総合的に判断**することに鑑みると、**開発原則に適合しないAIを利用者が選択する可能性**もあるが、そのこと自体は、**その競争促進効果に鑑みると、一律に否定すべきことでない**のではないか。

（注） 厳密には、選択の対象は、
・利用者が最終利用者（個人又は団体）である場合には、自らが利用するAIネットワークサービス又はAIネットワークシステム若しくはこれに実装するAIであり、
・利用者がAIネットワークサービスのプロバイダである場合には、そのAIネットワークサービスのために用いるAIネットワークシステム又はこれに実装するAIであるが、
ここでは、便宜上、単に「AI」としている。

したがって、**分野共通開発ガイドライン及び**後述する分野共通利活用ガイドライン**の内容を検討するに当たっては、開発原則に適合するAIと開発原則に適合しないAIの双方が市場に併存し得ることを前提として検討**することとしてはどうか。

2．上記1．に鑑みると**開発原則の実効性を確保するための措置が必要**と考えられるが、その**措置の例として、市場の機能を活用して**、開発者の自発的な取組を通じて**開発原則に適合しているAIが開発原則に適合していないAIと比べて市場において利用者に選択されやすくなる環境を整備**し、もって**開発者に対し開発原則の遵守への誘因を付与**する措置が考えられるのではないか。

具体的には、例えば、**次の①及び②を一体的に整備**することが考えられるのではないか。

① **開発者が**その開発するAIに関し**開発原則への適合性に関する情報を客観的に信頼できる形で自発的に提供する仕組み**

（例） 開発者がその開発するAIに関し自発的に提供する情報に基づき、公正中立で高度な専門性を有する第三者機関が当該AIの開発原則への適合性を評価して認証する制度

② 上記**①の仕組みにより開発者が提供した情報において開発原則に適合しているとされているAI**を実装するAIネットワークシステムの利活用に伴い、当該AI**のリスクが顕在化したことに起因する第三者の被害等に関し、その利用者の当該第三者に対する民事責任又は刑事責任が問題となる場合等当該利用者の法的責任、法的義務等が問題となる場合において、当該利用者の当該情報に対する信頼に基づく期待を保護**するための仕組み

（例） 当該被害に関する当該利用者の法的責任、法的義務等を減免する制度

（例： 当該リスクが顕在化したこと自体について当該利用者を無過失とみなす制度）

→ 利用者の第三者に対する民事責任を減免する制度を整備する場合には、Liability Gapを回避するため、当該第三者に対する民事責任を当該利用者が利用するAIネットワークサービスのプロバイダ又は当該AIの開発者が相応に負うものとする制度（関連する保険の制度を含む。）も併せて整備することが適当ではないか。

第十　開発原則の実効性の確保における市場の活用の在り方（3／3）

論点

3. そこで、**分野共通開発ガイドラインの関連文書**（OECDのガイドラインであれば、理事会勧告の本紙）**において、各国に対し、例えば上記2. の仕組みを分野共通の一般的仕組みとして一体的に整備**することについて検討するよう**推奨**するとともに、**各分野に係る各国の関係機関、関係国際機関等に対し、利活用の分野ごとの事情に照らし、必要に応じ分野別の特則的仕組みを検討するよう推奨**する旨を記すこととしてはどうか。

（特則的仕組みを設ける場合には、必要に応じ分野別開発ガイドラインの関連文書にも関連する事項を記すことが考えられる。）

4. 上記2. の仕組みがAIネットワークシステムの利活用に関連する仕組みであることに鑑みると、上記3. に即して**上記2. の仕組みに関し分野共通開発ガイドラインの関連文書に記す事項は**、後述する**利活用ガイドライン（分野共通利活用ガイドライン）の関連文書**（OECDのガイドラインであれば、理事会勧告の本紙）**においても記す**こととしてはどうか。

（特則的仕組みを設ける場合には、必要に応じ後述する分野別利活用ガイドラインの関連文書にも関連する事項を記すことが考えられる。）

5. 開発原則の実効性を確保するために市場を活用する場合であっても、**開発原則に掲げる事項のうち、人権等他の利益とバーターにすべきでないものについては、AIネットワークシステムの用途に照らし、必要に応じ当該用途に係る利活用の分野に関連する法制度の整備等を検討するよう当該分野に係る各国の関係機関、関係国際機関等に推奨**する旨を分野共通開発ガイドラインの関連文書（OECDのガイドラインであれば、理事会勧告の本紙）に記すこととしてはどうか。

6. 開発原則の実効性を確保するために**市場を活用するに当たっては**、その前提として**イノベーティブな研究開発が進むこと及び市場において公正な競争が有効に機能していることが期待**され、イノベーティブな研究開発及び公正な競争の確保の見地からは、**AIネットワークシステム相互間の連携が円滑に実現することが必要となるが、このことからも**、前述したように、相互接続性・相互運用性の確保等**AIネットワークシステム相互間の円滑な連携の確保に関しAIの開発段階において開発者が留意すべき事項を「連携の原則」【仮称】として開発原則に追加する**とともに、**関連して国、関係国際機関等に推奨すべき事項を分野共通開発ガイドラインの関連文書**（OECDのガイドラインであれば、理事会勧告の本紙）**に記す**ことが適切と考えられるのではないか。

なお、これまた前述したように、AIネットワークシステム相互間の連携は、AIネットワークシステムの利活用に伴うものであることから、**AIネットワークシステム相互間の円滑な連携の確保に関しAIネットワークシステムの利活用の段階において利用者（特にAIネットワークサービスのプロバイダ）が留意すべき事項及び国、関係国際機関等に推奨すべき事項をそれぞれ**後述する**分野共通利活用ガイドライン及びその関連文書**（OECDのガイドラインであれば、理事会勧告の本紙）**に記す**こととしてはどうか。

第十一　AIネットワークシステムの利活用に関し利用者等が留意すべき事項（1／10）

AIネットワーク化検討会議「報告書2016」（抄）

2．AIネットワーク化の進展に向けた協調の円滑化

【問題意識】

AIネットワーク化を円滑に進展させ、もってAIネットワークシステムを安心して安全に利活用する環境を実現するために、当事者間の競争関係の有無如何にかかわらず、イノベーティブな研究開発と公正な競争にも留意しつつ、AI相互間又はAIネットワークシステム相互間の協調を円滑化するための取組の在り方を検討することが求められる。

【今後の課題】

AI相互間又はAIネットワークシステム相互間の協調を推進するという観点から、AIネットワークシステムに関する相互接続性・相互運用性の確保に向けて検討を進めていくことに加え、関係する技術や市場の状況等を踏まえつつAIネットワーク化の円滑な進展のために必要となる範囲におけるAIネットワークシステムのオープン化の在り方を検討すべきである。

(1) AIネットワークシステムに関する相互接続性・相互運用性の確保

- 相互接続性・相互運用性を確保すべき対象の検討

（例）

- ➢ アーキテクチャ
- ➢ 情報の結節（AI相互間、AIとモノの間、AIと人間の間、AIとクラウドの間、API等）
- ➢ 匿名化、暗号等
- ➢ データの形式

- 相互接続性・相互運用性の確保の方法（dejure / defacto）の検討
- 相互接続性・相互運用性を確保するための結合テストの在り方の検討
- 相互接続性・相互運用性の確保に向けた国際協調の在り方の検討

(2) AIネットワーク化の円滑な進展のために必要となるAIネットワークシステムのオープン化の在り方の検討

- AIネットワークシステムのオープン化の動向の注視
- AIネットワーク化の円滑な進展のために必要となるオープン化の対象及び方法の検討
- 国際社会におけるAIネットワークシステムのオープン化の推進の在り方の検討

AIネットワーク化検討会議「報告書2016」（抄）

3．競争的なエコシステムの確保

【問題意識】

本検討会議が**「目指すべき社会像」として掲げる智連社会を実現するためには、AIネットワーク化を適正かつ円滑に進展する必要**があり、その前提として、AIネットワークシステムに関する**競争的なエコシステムの確保が不可欠**である。このような問題意識に基づき、AIネットワークシステムに関する**競争的なエコシステムを確保するために必要な取組の在り方について検討**することが必要と考えられる。

【今後の課題】

AIネットワークシステムに関する**競争的なエコシステムを確保するという観点から**、関係する市場の形成の進展に応じて、その**動向の継続的注視を行う**とともに、**AI相互間のネットワークの形成に関する当事者間の協議の円滑化に取り組むべき**である。

(1) **関係する市場の動向の継続的注視**

- **AIネットワーク化やデータ寡占等に着目したデータ等の創造・流通・蓄積の状況**、**事業者間の競争状況**その他**市場の動向の注視・評価**
  - ➢ **注視対象**（AIの範囲、データの範囲、市場等）**の画定、注視の視点、評価基準等の在り方の検討**
  - ➢ 注視・評価に必要となる**情報の収集の在り方の検討**
  - ➢ AIネットワークサービス（AIネットワークシステムの機能を提供するサービス）の供給者による行為であって、**公正な競争を阻害するおそれがあるものの類型化の検討**

(2) **AI相互間のネットワークの形成に関する当事者間の協議の円滑化**

- AI相互間のネットワークの形成に関する**当事者間の協議をめぐる紛争の動向及び影響の継続的注視**
- 必要に応じ、**当事者間の協議を円滑化する観点からの紛争処理の在り方の検討**

AIネットワーク化検討会議「報告書2016」（抄）

6. 利用者の保護

【問題意識】

AIネットワーク化は、利用者の利便性や生活の質の向上に貢献することが期待される一方で、利用者の中でも、特に消費者、青少年、高齢者等の権利利益との関係でリスクが生ずる可能性もある。したがって、AIネットワークシステムの利用者（特に消費者、青少年、高齢者等）の権利利益の保護の在り方について検討することが必要と考えられる。

【今後の課題】

AIネットワークサービスの利用者（特に消費者、青少年、高齢者等）の権利利益を保護する観点から、関係する市場の形成の進展に応じて、消費者等利用者の保護、市場の動向の注視・評価、紛争処理、国際的な制度調和の在り方について検討すべきである。

- AIネットワークサービスの利用者（特に消費者、青少年、高齢者等）の保護の在り方の検討
- 市場の形成に応じAIネットワークサービスの利用者の利益を保護する観点からの市場の動向の注視・評価
  - ➢ 注視すべき市場の画定、評価基準等の在り方の検討
  - ➢ 注視・評価に必要となる情報の収集の在り方の検討
  - ➢ AIネットワークサービスの供給者による行為であって、利用者の利益を阻害するおそれがあるものの類型化の検討
- AIネットワークサービスの供給者と利用者（特に消費者）との間の紛争処理の在り方の検討
- 継続的なアップデートを前提とするAIネットワークシステムを利用する消費者の保護の在り方の検討
- AIネットワークサービスを利用する消費者の保護に関する国際的な制度調和の在り方の検討

第十一　AIネットワークシステムの利活用に関し利用者等が留意すべき事項（4／10）

構成員からの指摘

【平野構成員】（第１回親会） 今後のAIは、色々なAIが複雑に絡み合っていくもので、相互接続性が重要である。利用者の使い方等も重要な問題であり、利活用の場面を考えることは重要である。開発ガイドラインの策定に向けた検討に当たっては、AIネットワークシステムの相互接続性・相互運用性についても議論の対象とすべきであるとともに、利活用に当たり留意すべき事項についても併せて検討すべきである。

【林（秀）構成員】（第１回親会） 開発の段階だけではなく、エンドユーザー等利用者による利活用の段階においても留意すべき事項を検討することは重要である。利活用の枠組みについても、開発ガイドラインとともに二本柱として検討すべきである。

【高橋構成員】（第１回開発原則分科会） AIシステムにおいては、開発から利活用に至る各段階の相互浸食が進むことになる。

【林（秀）構成員】（第１回影響評価分科会）AIネットワークシステム相互間の相互接続性・相互運用性を技術的に確保するのみならずAIネットワークシステム相互間の連携を円滑化させるための調整の仕組みを考えるべき。

【林（秀）構成員】（第１回影響評価分科会） AIのみならず、AIネットワークシステムを流通させるデータの形式の標準化を考えるべき。

【林（秀）構成員】（個別の御指摘） AIネットワークシステムの構成要素となり得るAIに関する標準の形成及び標準必須特許の取扱いは、連携が可能となるAIネットワークシステム範囲を左右し得るものであるとともに、イノベーティブな研究開発及び市場におけるAIネットワークシステム・AIネットワークサービスに関する競争を左右し得るものであることから、AIネットワークシステム相互間の円滑な連携の確保の見地並びにイノベーティブな研究開発及び公正な競争の確保の見地から、標準の形成過程及び標準必須特許の権利者たる開発者が留意すべき事項に関し検討すべきではないか。

【福井構成員】（第１回開発原則分科会）AIに関するデータ等の囲い込みを見据えて、ガイドラインにおいて、オープン・サイエンス的な理念を盛り込むべきではないか。

【宍戸構成員】（第１回開発原則分科会）福井構成員が指摘されたオープン・サイエンス的な理念については、智連社会の基本理念の一つとして掲げられている「イノベーティブな研究開発と公正な競争」に対応している部分があるのではないか。

構成員からの指摘

【新保構成員】（第１回開発原則分科会） ガイドラインでは**バイ・デザインにより開発段階でリスクに事前に対処する姿勢を前面に出すべき**である。他方で、**利用制限など利用の在り方についても検討する必要**がある。

【大屋構成員】（第１回開発原則分科会） 事前に定めた倫理原則に従って意思決定することは、人間にもAIにも困難である。人間にもできないことをAIに要求することは、AIの成長を阻害する。**被害が生じたとき、それに対して責任を負う者が誰もいないという事態（Liability Gap）を回避するために原則が要請されるという問題意識から原則を策定すべき**ではないか。
　中長期的には、ある種の実験環境の下で、責任の配分を考えていくべき。

【久世構成員】（第１回開発原則分科会） **ガイドラインの検討に当たっては、最終的な判断における人間の介在の有無、AIの具体的な利用方法等に留意して、きめ細やかに検討していく必要**がある。

【丸山構成員】（第１回開発原則分科会） **機械学習においては開発と利活用は不可分の関係**にあり、**ガイドラインの策定時にも開発と利活用の在り方を併せて検討する必要**がある。

第十一　AIネットワークシステムの利活用に関し利用者等が留意すべき事項（6／10）

論点

1．次の①～③に鑑み、**AIネットワークシステムの利活用に関し利用者**（AIネットワークサービスのプロバイダ、最終利用者）**が留意すべき事項及び国、関係国際機関等に推奨すべき事項を整理して、国際的に共有する枠組みとして「AIネットワークシステム利活用ガイドライン」**（仮称）**及びその関連文書**（OECDのガイドラインであれば、理事会勧告の本紙）**を策定し、開発ガイドライン及びその関連文書と相互に補完し合う二本柱とすることに向け、OECD等の協力の下、国際的に議論すべき**ではないか。

① **AIネットワーク化の健全な進展の促進及びAIネットワークシステムの便益の増進の見地から、AIネットワークシステムの利活用に関し**、例えば次に掲げる事項のように**利用者（特にAIネットワークサービスのプロバイダ）が留意すべき事項及び国、関係国際機関等に推奨すべき事項がある**と考えられること。

（例１） AIネットワーク化の進展を通じた**智連社会の形成は、AIネットワークシステム（特にAIネットワークサービス）を利用しようとするあらゆる個人・団体が最終利用者としてこれを利用することを通じて社会に参加して活動し、AIネットワーク化が健全に進展することにより実現**するもの。

→ **最終利用者としてAIネットワークシステム**（特にAIネットワークサービス）**を利用しようとする個人・団体の公平な利用の確保の在り方については**、主にAIネットワーク化の健全な進展の促進の見地から、**AIネットワークサービスのプロバイダが留意すべき事項がある**のではないか。また、**国、関係国際機関等に対し、関連する動向やベストプラクティスに関する情報を国際的に共有する**ほか、関連する弊害の蓋然性が顕著となった場合には、その対応の在り方に関し、イノベーティブな研究開発及び公正な競争への影響、最終利用者の利益への影響等を踏まえつつ、慎重かつ多角的に検討する**よう推奨**すべきではないか。

（例２） **AIネットワークシステム相互間の連携の在り方は**、AIネットワークシステムと連携して利活用できるAIの範囲を左右することから、AIネットワークシステムに関するイノベーション及びAIネットワークサービスに関する競争を左右するとともに、AIネットワークシステムの利活用の範囲を左右することから、**AIネットワーク化の進展及び利用者（特に最終利用者）や社会にとってのAIネットワークシステムの便益を左右**するもの。

→ **AIネットワークシステム相互間の連携の在り方については**、主にAIネットワーク化の健全な進展及びAIネットワークシステムの便益の増進の見地から、**AIネットワークサービスのプロバイダが留意すべき事項がある**のではないか。また、**国、関係国際機関等に対し、関連する動向やベストプラクティスに関する情報を国際的に共**有するほか、関連する弊害の蓋然性が顕著となった場合には、その対応の在り方に関し、イノベーティブな研究開発及び公正な競争への影響、最終利用者の利益への影響等を踏まえつつ、慎重かつ多角的に検討する**よう推奨**すべきではないか。

② **開発原則に適合するAIを実装するAIネットワークシステムといえども、その利活用に伴い**利用者、他のAIネットワークシステム、第三者及び社会・人類に**悪影響を及ぼすリスクが皆無である訳ではない**ことに加えて、**開発原則に適合しないAIを実装するAIネットワークシステムとも共存することに鑑みると、AIネットワークシステムの利活用の段階についても**、これに伴う**リスクの抑制の見地から、利用者が留意すべき事項があるとともに、国、関係国際機関等に推奨すべき事項がある**と考えられること。

【開発原則に適合するAIを実装するAIネットワークシステムの**利活用に伴うリスクの例】**

（例1）　開発原則に適合するAIを実装するAIネットワークシステムが**誤ったデータを学習し、その利活用に伴い第三者の利益が害されるリスク**

（例2）　開発原則に適合するAIの**開発者が、そのリスクに鑑み、その用途を限定すべき旨をあらかじめ明示していたにもかかわらず**、当該AIを実装するAIネットワークシステムを**その利用者が限定された用途以外の用途のために利活用することに伴い、当該リスクが顕在化するリスク**

第十一　AIネットワークシステムの利活用に関し利用者等が留意すべき事項（7／10）

論点

③ AIネットワークシステムは国境を越えて相互に連携して利活用されていくものであるため、**AIネットワークシステム相互間の連携をめぐる紛争や、利用者又は第三者とAIネットワークサービスのプロバイダ等との間の紛争が国境を問わず発生**し得ることに鑑み、**これら国内の紛争及び国境を越えた紛争の処理の在り方に関し**、AIネットワーク化の健全な進展の促進並びにAIネットワークシステムの便益の増進及びリスクの抑制の見地から、**国、関係国際機関等に対し、紛争の動向を注視して、動向やベストプラクティスに関する情報を国際的に共有するとともに**、紛争の発生状況等に応じて**国内の紛争及び国境を越えた紛争の処理の在り方を検討して、所要の措置を講ずるよう推奨すべき**と考えられること。

2. **利活用ガイドラインの体系については**、開発ガイドラインと同様に、**分野共通ガイドライン及び分野別ガイドラインからなるものとすることが適当**ではないか。

→以下両者を区別する場合には、前者を「分野共通利活用ガイドライン」といい、後者を「分野別利活用ガイドライン」という。

分野共通利活用ガイドラインは、AIネットワークシステム（AIネットワークサービスを含む。）の**利活用の分野を通じて利用者**（AIネットワークサービスのプロバイダ、最終利用者）**が留意すべき事項や利活用の分野間の連携の可能性を踏まえて利用者**（同前）**が留意すべき事項を策定**するものとして、**本推進会議がその検討と議論を推進**してはどうか。

分野別利活用ガイドラインは、各分野における策定の要否そのもの及び策定する場合における内容の双方ともに、**各分野の関係国際機関を含む当該分野の産学民官のステークホルダー自身による検討と議論に委ねる**こととしてはどうか。

3. **分野共通利活用ガイドラインの構成については**、分野共通開発ガイドラインと同様に、**次に掲げる構成をたたき台**としてはどうか。

(1) 分野共通利活用ガイドライン　（OECDのガイドラインであれば、理事会勧告の附属文書（Annex））

AIネットワークシステム（AIネットワークサービスを含む。）の**利用者**（AIネットワークサービスのプロバイダ及び最終利用者を含む）**が、その利活用**（AIネットワークサービスの提供及び利活用を含む。）**に当たり**、AIネットワークシステムの**利活用の分野を通じて留意すべき事項及び利活用の分野間の連携の可能性を踏まえて留意すべき事項に関する原則（「利活用原則」）並びにその説明**

(2) 分野共通利活用ガイドラインの関連文書　（OECDのガイドラインであれば、理事会勧告の本紙）

- 分野共通利活用ガイドラインに定める事項に関連し、**国、関係国際機関等に推奨すべき事項**
- ガイドラインの**見直しの時期及び方法**

第十一　AIネットワークシステムの利活用に関し利用者等が留意すべき事項（8／10）

論点

4．分野共通利活用ガイドラインの「目的」については、分野共通開発ガイドラインの「目的」と同様の考え方により、次のような趣旨を必要に応じて再構成した上で分野共通利活用ガイドラインに掲げることとしてはどうか。

> このガイドラインは、AIネットワークシステム（AIネットワークサービスを含む。以下同じ。）の公共性に鑑み、その利用者（AIネットワークサービスのプロバイダ及び最終利用者を含む。）が、その利活用（AIネットワークサービスの提供及び利活用を含む。以下同じ。）に当たり、AIネットワーク化の健全な進展の促進並びにAIネットワークシステムの便益の増進及びリスクの抑制に関し、AIネットワークシステムの利活用の分野を通じて又は分野間の連携の可能性を踏まえて留意すべき事項を利活用原則として整理し、非拘束的な枠組みとして国際的に共有することにより、AIネットワークシステムの最終利用者の利益を保護するとともに第三者及び社会への波及的な悪影響を防止し、もって人間中心の智連社会の形成に資することを目的とする。
>
> （なお、「人間中心の」については、「人間が主体的にAIネットワークシステムを使いこなす」等の表現とすることも考えられる。）

5．分野共通利活用ガイドラインに定める「利活用原則」は、AIネットワークシステム（AIネットワークサービスを含む。）の利用者（AIネットワークサービスのプロバイダ及び最終利用者を含む。）が、その利活用（AIネットワークサービスの提供及び利活用を含む。）に当たり、次の①～③に掲げる見地から利活用の分野を通じて又は分野間の連携の可能性を踏まえて留意すべき事項としてはどうか。

① AIネットワーク化の健全な進展の促進及びAIネットワークシステムの便益の増進

（例）・AIネットワークサービスの公平な利用の確保に関する事項（例：AIネットワークサービスの提供に当たり不当な差別的取扱いをすべきでないこと）

・AIネットワークシステム相互間の円滑な連携の確保に関する事項（例：AIネットワークシステム相互間の連携に関し不当な差別的取扱いをすべきでないこと）

・イノベーティブな研究開発と公正な競争の確保に関する事項（例：AIに関する標準必須特許の許諾に関し不当な差別的取扱いをすべきでないこと）

② AIネットワークシステムのリスクの抑制

→ 開発原則に掲げる項目に対応するリスクに関連する事項のほかに、利活用固有のリスクの抑制に関し留意すべき事項如何。

③ AIネットワークシステムの利活用に伴い、当該AIネットワークシステムに実装するAIのリスクの顕在化に起因する被害に関する被害者の利益の保護

（例）・AIネットワークサービスのプロバイダがそのAIネットワークシステムに実装するAIのリスクの顕在化に起因する被害者の損害を賠償するために加入すべき保険に関する事項

第十一　AIネットワークシステムの利活用に関し利用者等が留意すべき事項（9／10）

論点

6. **分野共通利活用ガイドラインの関連文書**（OECDのガイドラインであれば、理事会勧告の本紙）**においては、次の④～⑥に掲げる見地から国、関係国際機関に推奨すべき事項並びに分野共通利活用ガイドラインの見直しの時期及び方法を掲げる**こととしてはどうか。

④ **上記5. ①に掲げる見地**（AIネットワーク化の健全な進展の促進及びAIネットワークシステムの便益の増進）

⑤ **上記5. ②に掲げる見地**（AIネットワークシステムのリスクの抑制）

⑥ **上記5. ③に掲げる見地**（AIネットワークシステムの利活用に伴い、当該AIネットワークシステムに実装するAIのリスクの顕在化に起因する被害に関する被害者の利益の保護）**→次項（７．）**

⑦ **AIネットワークシステム相互間の連携等をめぐるAIネットワークサービスのプロバイダ間における国内の紛争又は国境を越えた紛争の処理**の在り方

⑧ **利用者（特に最終利用者）又は第三者とAIネットワークサービスのプロバイダ等との間における国内の紛争又は国境を越えた紛争の処理**の在り方

【備考】⑦及び⑧に関連し、第三者機関によるADRの手続に応ずることを原則として拒んではならないことを制度化している例として、金融ADRの制度（銀行の例：銀行法第52条の67第2項第2号）参照。

なお、AI及びAIネットワークシステムが現時点においては揺籃期であることから、**上記6. ④に関し**、**AIネットワーク化の健全な進展の促進及びAIネットワークシステムの便益の増進の見地からの法的規制の創設を検討することは、現時点では時期尚早**ではないか。**少なくとも関連する弊害の蓋然性が顕著になるまでは**、分野共通開発ガイドライン及び分野共通利活用ガイドライン並びにそれぞれの関連文書により開発者及びAIネットワークサービスのプロバイダ等利用者が留意すべき事項並びに国、関係国際機関等に推奨すべき事項を国際的に共有した上で、**国、関係国際機関等は**、**関連する動向を注視し、動向やベストプラクティスに関する情報を国際的に共有**する（上記6. ④～⑧）ほか、紛争の発生状況等に応じて**国内の紛争及び国境を越えた紛争の処理の在り方等を検討して所要の措置を講ずる**（上記6. ⑦及び⑧）**にとどめるような謙抑的な姿勢であるべき**ではないか。

第十一　AIネットワークシステムの利活用に関し利用者等が留意すべき事項（10／10）

論点

7．AIネットワークシステムの利活用に伴うその**AIのリスクの顕在化に起因する利用者又は第三者の被害に関し**

① **利用者とその利用する**AIネットワークサービスの**プロバイダ又は**当該AIの**開発者との間の争訟**

② **第三者と**当該AIネットワークシステムによるAIネットワークサービスの**プロバイダ又は**当該AIの**開発者との間の争訟**

については、当該AIが開発原則に適合しているのか否かにかかわらず

ア　AIネットワーク化が進展していく社会において、個人・団体がAIのリスクに起因する自らの損害に対し不安を抱くことなく、安心してAIネットワークシステムを利用できるようにすることにより、社会に参加できるようにするとともに、第三者がAIネットワークシステムと安心して共存できるようにするためには、**利用者及び第三者の損害について、相応の賠償が確実になされるべき**こと

イ　**上記①及び②の争訟においては**、利用者又は第三者と当該プロバイダ又は当該開発者との間の**当該AIに関連する情報の非対称性に鑑みると**、情報の非対称性を否定する特段の事情がない場合には、**裁判手続における利用者又は第三者の当該リスクの顕在化に関する主張立証責任を軽減すべき**と考えられること

を踏まえ、**分野共通利活用ガイドラインの関連文書**（OECDのガイドラインであれば、理事会勧告の本紙）**において、例えば次の(1)・(2)を組み合わせて設けることについて検討するよう国、関係国際機関等に推奨**する旨を記すこととしてはどうか。

(1) **利用者若しくは第三者の損害又は**AIネットワークサービスの**プロバイダ若しくは**AIの**開発者の損害賠償に伴う損失に関する保険の仕組み又は制度**

（例1）　AIネットワークサービスのプロバイダ又はAIの開発者の損害賠償に伴う損失に関する保険の制度及び加入状況の公表

（例2）　利用者向けの保険の制度

※　（例1）・（例2）において、保険料については、AIの開発原則への適合性に応じて差異を設けることが考えられるのではないか。

(2) **利用者及び第三者の**当該リスクの顕在化に関する**主張立証責任をプロバイダ又は開発者に転換する仕組み又は制度**

（例1）　利用者及び第三者の当該リスクの顕在化に関する主張立証責任をプロバイダ又は開発者に転換する制度

（例2）　AIネットワークサービスの提供に関する契約により、利用者の当該リスクの顕在化に関する主張立証責任をプロバイダに転換することを定める場合における当該契約の条項のモデルの公表

第十一　AIネットワークシステムの利活用に関し利用者等が留意すべき事項（10／10）

論点

8．**自らAIネットワークシステムを構築する最終利用者**、自ら構築するAIネットワークシステムにより**AIネットワークサービスを最終利用者等他の者に提供するプロバイダ**及び**プロバイダからAIネットワークサービスの提供を受ける最終利用者の種別に応じて、適用すべき利活用原則の範囲、内容等**に異同があり得ることから、**その異同を利活用ガイドラインに明記**するとともに、これら**利用者の種別ごとに整理したマニュアル等を作成**することとしてはどうか。

主要参考文献

[Amodei et al. 2016] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman & Dan Mané, *Concrete Problems in AI Safety*, arXiv:1606.06565 [cs.AI] (2016)
[Bostrom 2014] Nick Bostrom, Super Intelligence: Paths, Dangers, Strategies (2014)
[FLI 2015] The Future of Life Institute (FLI), Research Priorities for Robust and Beneficial Artificial Intelligence (2015)
[House of Commons 2016] House of Commons Science and Technology Committee, Robotics and artificial intelligence, Fifth Report of Session 2016–17 (2016)
[Nadella 2016] Satya Nadella, The Partnership of the Future: Microsoft's CEO explores how humans and A.I. can work together to solve society's greatest challenges, SLATE (2016).
[National Science and Technology Council 2016] National Science and Technology Council, The National Artificial Intelligence Research and Development Strategic Plan (2016)
[Orseau & Armstrong 2016] Laurent Orseau & Stuart Armstrong, *Safely Interruptible Agents*, 32nd Conference on Uncertainty in Artificial Intelligence (2016)
[Partnership on AI 2016] Partnership on AI, Tenets (2016)
[Thaler & Sunstein 2008] Richard Thaler & Cass Sunstein, Nudge: Improving Decisions About Health, Wealth, and Happiness (2998)
[Whitehouse 2016] Whitehouse, Preparing for the Future of Artificial Intelligence (2016)
[一杉 2014]一杉裕志「ヒト型AIは人類にどのような影響を与え得るか」人工知能29巻3号509頁(2014)
[新保 2016]新保史生「ロボット法学の幕開け」Nextcom Vol.27(2016)
[平野 2016]平野晋「『ロボット法』と自動運転の『派生的トロッコ問題』—主要論点の整理と、AIネットワークシステム『研究開発8原則』」NBL1083号29頁以下(2016)
[堀 2015] 堀浩一「人工知能の研究開発をどう進めるか—技術的特異点(シンギュラリティ)を見据えて」情報管理58巻4号(2015)
[松尾ほか 2015] 松尾豊ほか「人工知能学会倫理委員会の取組み」人工知能30巻3号(2015)
[向殿 2016]向殿政男「IoT時代におけるものづくり安全の動向」情報通信学会誌Vol.34 No.1 (2016)