

事務局説明資料

「検討の背景及び今後の検討事項」

○ 検討の背景

【資料2-1-1】 これまでの検討状況 -AIネットワーク化検討会議とその成果-

【資料2-1-2】 欧米の動向

○ 今後の検討事項

【資料2-2】 AIネットワーク化が社会・経済にもたらす影響とリスクの評価について

【資料2-3】 「AI開発ガイドライン」(仮称)の策定に向けて
-国際的な議論の用に供すべき素案の検討-

【資料2-4】 「AIネットワーク社会推進会議」に関するスケジュール

【資料2-5】 AIネットワーク社会推進会議(第1回)における議論の概要

これまでの検討状況

—AIネットワーク化検討会議とその成果—

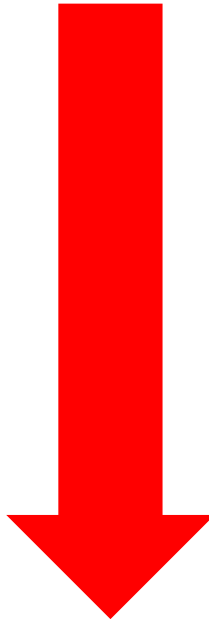
沿革



平成27年2月

インテリジェント化が加速するICTの未来像に関する研究会

平成27年6月30日 「報告書2015」



平成28年2月

AIネットワーク化検討会議 （旧称 ICTインテリジェント化影響評価検討会議）

平成28年4月15日 中間報告書「AIネットワーク化が拓く智連社会^{ウインズ}(WINS)」

（平成28年4月29日・30日 G7香川・高松情報通信大臣会合）

平成28年6月20日 報告書2016「AIネットワーク化の影響とリスク」

AIネットワーク化検討会議

1 目的

2040年代を見据え、A I ネットワーク化（※）に向け、次に掲げる項目を検討。

- 目指すべき社会像及びその基本理念の整理
- A I ネットワーク化が社会・経済にもたらす影響及びリスクの基礎的評価
- 今後の課題（社会的・経済的・倫理的・法的課題）の整理

（※）「AIネットワーク化」とは

- ・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をいう。）の構築
- ・AI相互間の連携などAIネットワークシステムの高度化をいう。

2 検討体制

【座長】 須藤 修 東京大学大学院情報学環教授

【顧問】 村井 純 慶應義塾大学環境情報学部長

【構成員】 理工学系及び人文・社会科学系の有識者 計 37人（座長及び顧問を含む。）

3 スケジュール

2月 2日 第1回会合

4月 15日 中間報告書

6月 20日 報告書2016

（ 親会 5回開催
分科会（経済、社会・人間、法・リスク）等 13回開催 ）

AIネットワーク化検討会議 構成員

【座長】 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

【顧問】 村井 純 慶應義塾大学環境情報学部長・教授

【座長代理】 平野 晋 中央大学総合政策学部教授・大学院総合政策研究科委員長

赤坂 亮太 慶應義塾大学大学院メディアデザイン
研究科リサーチャー

石井 夏生利 筑波大学図書館情報メディア系准教授

板倉 陽一郎 弁護士

稲葉 振一郎 明治学院大学社会学部教授

井上 智洋 駒澤大学経済学部専任講師

江間 有沙 東京大学教養学部附属教養教育
高度化機構特任講師

大内 伸哉 神戸大学大学院法学研究科教授

大屋 雄裕 慶應義塾大学法学部教授

鹿島 久嗣 京都大学大学院情報学研究科教授

河井 大介 東京大学大学院情報学環助教

河島 茂生 青山学院女子短期大学現代教養学科准教授

久木田 水生 名古屋大学大学院情報科学研究科准教授

クロサカ タツヤ 総務省情報通信政策研究所コンサルティング・フェロー

越塚 登 東京大学大学院情報学環教授

佐々木 勉 総務省情報通信政策研究所特別上級研究員

佐藤 英司 福島大学経済経営学類准教授

穴戸 常寿 東京大学大学院法学政治学研究科教授

実積 寿也 九州大学大学院経済学研究院教授

新保 史生 慶應義塾大学総合政策学部教授

田中 絵麻 一般財団法人マルチメディア振興センター

情報通信研究部主席研究員

田中 浩也 慶應義塾大学環境情報学部教授

高橋 恒一 国立研究開発法人理化学研究所生命システム研究

センター生化学シミュレーション研究チームチームリーダー

中西 崇文 国際大学グローバル・コミュニケーション・センター准教授

林 秀弥 名古屋大学大学院法学研究科教授

林 雅之 国際大学グローバル・コミュニケーション・センター

客員研究員

原井 洋明 国立研究開発法人情報通信研究機構ネット

ワークシステム研究所ネットワーク基盤研究室長

深町 晋也 立教大学大学院法務研究科教授

福井 健策 弁護士

松尾 豊 東京大学大学院工学系研究科特任准教授

山本 勲 慶應義塾大学商学部教授

山本 龍彦 慶應義塾大学法科大学院教授

湯浅 壘道 情報セキュリティ大学院大学教授

若田部 昌澄 早稲田大学政治経済学術院教授

渡辺 智暁 慶應義塾大学大学院政策・メディア研究科特任准教授

「AIネットワーク化」及びその進展段階

AIネットワーク化

- ① AIネットワークシステムの構築、② AI相互間の連携等AIネットワークシステムの高度化

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

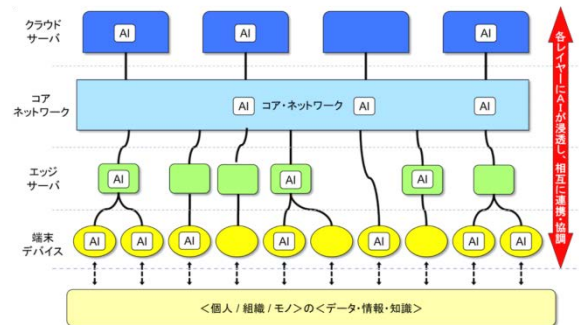
AIネットワーク化の進展段階

- ① 人工知能(AI)が、他の人工知能(AI)とは連携せずに、インターネットを介するなどして単独で機能し、人間を支援
- ② 人工知能(AI)相互間のネットワークが形成され、社会の各分野における自動調整・自動調和が進展

- ネットワーク上に用途の異なる多様なAIが出現。
- 複数のAIを取りまとめる能力を有するAIも出現。
- 複数のAIが相互に連携・協調。

(例)・産業機械と部材の連携、サービスロボットとセンサの連携
・交通、物流、オフィス業務、生活環境等の自動調整

人工知能(AI)は、情報通信ネットワークの各レイヤーに浸透し、相互に連携・協調。



- ③ 人間の潜在的能力がAIネットワークシステムにより拡張

センサ、アクチュエータ、人間及び人工知能が連携

感覚器官の能力向上

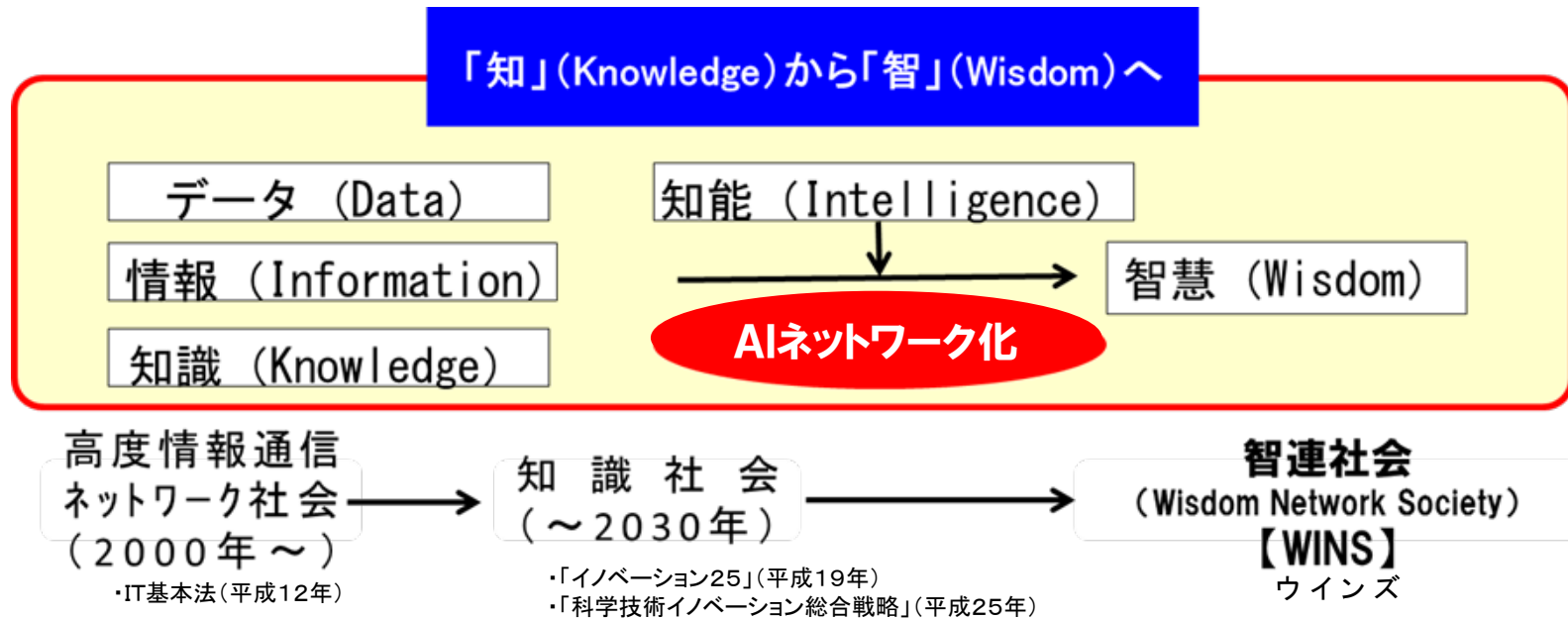
身体機能の能力向上



(例)・脳情報を外部に出力
・ヒトの意思により、義手・義足やロボットを操作
・遠隔地の出来事を仮想体験
→ 遠隔地のロボットを操作して協働

- ④ 人間とAIネットワークシステムが共存

「知」から「智」へ



AIネットワーク化の進展段階

- ① 人工知能(AI)が、他の人工知能(AI)とは連携せずに、インターネットを介するなどして単独で機能し、人間を支援
- ② 人工知能(AI)相互間のネットワークが形成され、社会の各分野における自動調整・自動調和が進展
- ③ 人間の潜在的な能力がAIネットワークシステムにより拡張
- ④ 人間とAIネットワークシステムとが共存

「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、物事に対処する能力

目指すべき社会像(「智連社会」)

「知」(Knowledge)から「智」(Wisdom)へ



高度情報通信
ネットワーク社会
(2000年～)
・IT基本法(平成12年)

知識社会
(～2030年)

・「イノベーション25」(平成19年)
・「科学技術イノベーション総合戦略」(平成25年)

智連社会
(Wisdom Network Society)
【WINS】
ウインズ

「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、物事に対処する能力

目指すべき社会像

智連社会 (Wisdom Network Society) 【WINS】

- ・ 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

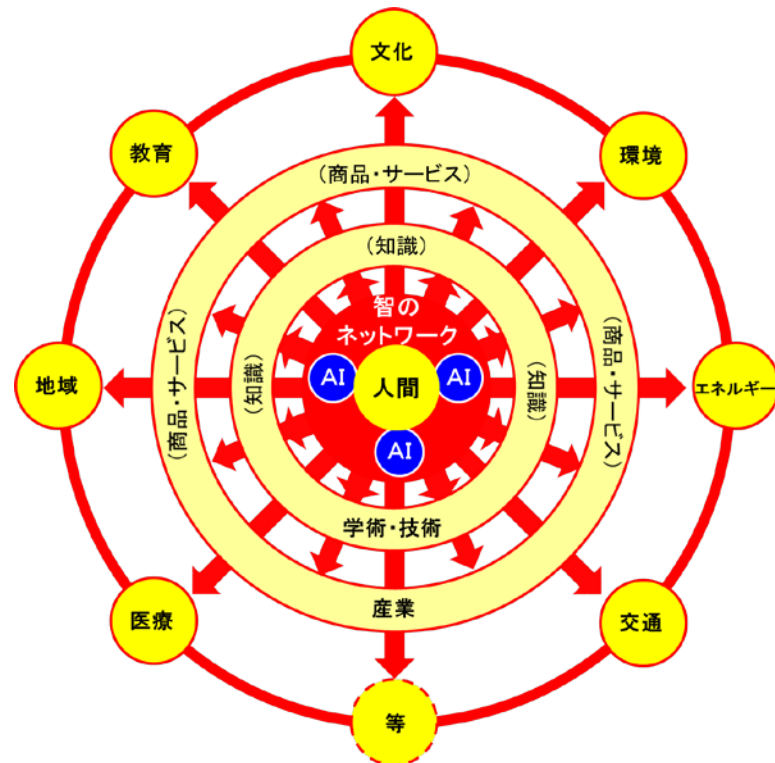
ウインズ

人機共存

総智連環

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



目指すべき社会像（「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超えた協調が地域内・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

「智連社会」における人間像

1 AIネットワーク化の進展が産業構造にもたらす影響

- 隣接業界への進出や異業種間の連携といった形で、異業種間の産業の融合が進展していく。
 - ① 隣接業界への進出の例：製造業者が、自社製品の稼働状況のデータをAIで分析し、アフターサービスを提供（→サービス業への進出）
 - ② 異業種間の連携の例：自動車メーカーと保険業者が連携し、走行履歴のデータをAIで分析した結果に基づき、保険料を割引
- あらゆる分野において、データの利活用が進み、情報通信産業化が進展。
- 今後も変化をもたらすのは、新しい企業（ベンチャー企業）であるが、セットアップコストの大幅な低下等により、起業しやすい環境になるものと考えられる。

2 AIネットワーク化の進展が雇用にもたらす影響

- 定型的な業務をはじめ、知的業務についてもAIネットワークシステムへの代替が進む。
- 人間に残される仕事の特徴としては、クリエイティビティ、マネージメント、ホスピタリティが考えられ、人間はアイデアを生み出す仕事や人間間の高度なコミュニケーション能力（リーダーシップ等）を必要とする仕事を担うこととなる。
- クラウドソーシングにより仕事が細分化され、外製化（他の企業、自営的就労者等への発注）が進む。これまでの労働法制が前提としてきたフルタイムの正社員を中心とする企業組織や雇用の在り方に抜本的に影響が及ぶこととなる。
- テレワークや自動翻訳を活用することにより、世界中のどこでも自立して働くことができるようになる。

3 「智連社会」における人間像

- 生産手段がAIネットワークシステムに代替されるかもしれないが、仕事や雇用のみには重きを置く価値観の転換を図るべきではないか。食べるためではなく趣味・楽しみとして仕事をするのが一つの人間像として考えられる。
- AIネットワーク化により仕事のやり方、形態が変わることが見込まれるため、人間は、事業再編能力やコミュニケーション能力、全体をデザインする能力等が求められるようになり、そのための訓練・教育も必要である。
- 人間像を考えることは、AIネットワークシステムの在り方、人間とAIネットワークシステムとの関わり合い方を考えることであり、継続的な検討が必要。

AIネットワーク化が社会・経済にもたらす影響

AIネットワーク化が社会・経済にもたらす影響について、公共(まち)／生活(ひと)／産業(しごと)の分野ごとに、2020年代～2040年代の時系列で評価。

【公共】 公共インフラ、防災、スマートシティ、行政

【生活】 生活支援、豊かさ創造(創造的活動、コミュニティ活動等)

【産業】 分野共通(コーポレート業務等)、農林水産、製造業、運輸・物流、卸売・小売、金融・保険、医療・介護、教育・研究、サービス業、建設

【例】 製造業

- 製造プロセスとサプライチェーンのスマート化により、動的な需給バランスに対応した生産最適化や高度な多品種変量生産(マスカスタマイゼーション)を実現
- 利用者の稼働データの分析により、デジタルマーケティングや、付加価値が高いアフターサービス・メンテナンスサービスを実現
- 2020年代後半以降には、設計段階からの自動化が実現することにより、開発作業が効率化・高速化

2020年 2030年 2040年

- ▲ 人と協業できる協調型ロボットの普及[MRI]
- ▲ 無人メンテナンス^{※2}の実現(2020)[未来]
- ▲ 製品需要予測の精度向上、リアルタイムデータによるサプライチェーンの効率化[14.0]
- ▲ 需要データの活用によるリードタイム短縮[14.0]
- ▲ 産業用ロボット・工作機械のスマート化により、高度な多品種変量生産(マスカスタマイゼーション)が実現^{※1}(2020)[未来]
- ▲ 移動工場によるオンデマンド製造により納期的大幅短縮を実現(2020)[未来]
- ▲ ユーザーの製品利用時の稼働データ分析より、付加価値が高いアフターサービス、メンテナンスサービスを実現(2020)[未来]
 - ▲ 「勘・コツ」の技能をAIロボットが習得[MRI]
 - ▲ 製品の設計・試作・試験等にAIが導入され、開発作業効率やスピードが向上[MRI]
 - ▲ 製品の性能対価の向上[MRI]
 - ▲ 生産工程変更等、複雑な環境変化に対応できる自律型ロボットの登場(2029)[白書]
 - ▲ 「設計リードタイムゼロ」、[在庫ゼロ]の実現(2030)[競争]
 - ▲ 規格品からテラーメイド品へ移行(2030)[競争]
 - ▲ 大企業の無人化工場一般化[MRI]
 - ▲ 製品設計の半自動化・完全自動化[MRI]

【例】 医療・介護

- 患者のバイタルデータによる発病予測や、遺伝子情報による健康管理が実現し、健康寿命が延伸
- 研究論文の自動分析により、研究や新薬開発が加速

2020年 2030年 2040年

- ▲ AI診断支援を活用した診断精度の向上、誤診の低減[MRI]
 - ▲ 視覚障害者向けガイダンスロボットの実用化(2016)[白書]
 - ▲ センサーを駆使した手押し車型歩行補助器が発売(2016)[白書]
 - ▲ 身体センサー等から意識にのぼらない運動機能の異常を検知する技術が実用化(2017)[白書]
 - ▲ AIによるセカンドオピニオンサービスの普及[MRI]
 - ▲ バイタルデータからの発病予測サービス普及[MRI]
 - ▲ 肺がん等の画像診断、自動検出の実用化[MRI]
 - ▲ 遺伝子情報による健康管理が実用化(2020)[未来]
 - ▲ 調剤の自動化[MRI]
 - ▲ コピキタス生体情報モニタリングによる健康管理(2023)[白書]
 - ▲ 運動能力をアシストできるアクチュエータ技術により高齢者のQOLが改善(2023)[白書]
 - ▲ 健康寿命延伸によるセカンドキャリア拡大、シニア経済の拡大[MRI]
 - ▲ 高齢者の外出を促すアシストネットワークロボットの実用化(2028)[白書]
 - ▲ センサー・アクチュエータによる医療・介護・健康・障害者行動支援[※](2030)[原井構成員]
 - ▲ 認知症改善、認知力強化機能の普及[MRI]
 - ▲ 製薬会社によるAIを活用した新薬開発の普及[MRI]
 - ▲ 健康寿命延伸による医療費低減[MRI]

※ 具体例:事故状況の監視等の救助支援、異常行動の見守りや介護・自立支援などの福祉・インフラ整備、双方向の遠隔医療 等

AIネットワーク化の影響に関する評価指標

1 AIネットワーク化の進展が社会にもたらす影響に関する評価指標

- 中間報告書で見通した将来の社会（2020年～2040年代）について、分野ごと（公共分野、生活分野、産業分野）に定量的に測定し、又は客観的に把握することができる指標の設定について検討すべきである。
ex. リアルタイムで収集・分析可能なデータの種類・データ数、自動運転の実現状況（普及台数） 等
- 個々の分野の枠を超えて、産業全般や雇用全般への影響を踏まえた指標の設定についても検討すべきである。
ex. 就労人口数、正社員数、テレワークの普及状況 等
- 多様な要素が関連するため、複数の項目からなる指標とすべきものと考えられる。数値化できない要素の取扱いも含めて、更なる検討が必要である。

2 豊かさや幸せに関する評価指標

- GDP等の経済統計では不十分であり、非金銭的・非市場的な要素を取り入れた指標の設定が望まれる。
- 豊かさや幸せの感じ方は、個人の価値観に大きく左右され得るため、主観的な評価と客観的な評価のバランスを考慮した指標の設定が望まれる。
- 次の指標を参考にしながら、適切な指標の設定に向けて検討を続けることが必要である。

<参考> 豊かさや幸福度を測る主な指標

- | | |
|--------------------|------------------------|
| ◆ より良い暮らし指標（B L I） | ◆ 人間開発指数（H D I） |
| ◆ 国民総幸福量（G N H） | ◆ デジタル経済・社会指標（D E S I） |
| ◆ 経済的福祉指標（M E W） | ◆ 潜在能力アプローチ |

AIネットワーク化が社会・経済にもたらすリスク

- 本検討会議においては、AIネットワーク化が社会・経済にもたらすリスクを次のように分類。
 - ① **機能に関するリスク**： AIネットワークシステムに期待される**機能が適正に発揮されないリスク**
 - ② **法制度・権利利益に関するリスク**： AIネットワークシステムにより**権利利益等が侵害されるリスク**
- リスクの評価及び管理の在り方の検討に当たっては、AIネットワークシステムを利活用する**具体的な場面を想定したシナリオを作成し、具体的に検討すべき(リスク・シナリオ分析)**。

事故のリスクのように、両者の側面を併せ持つリスクもある。

リスクの種類		例
機能に関するリスク	セキュリティ関連	<ul style="list-style-type: none"> ・AIネットワークシステムに対するハッキングやサイバー攻撃等 ・AIネットワークシステムに対する攻撃が秘かに行われ、被害に気づかないこと
	情報通信ネットワークシステム関連	<ul style="list-style-type: none"> ・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、意図しない事象が発生 ・情報通信ネットワークの不具合によりAIが正常に動作せず、意図しない事象が発生
	不透明化	<ul style="list-style-type: none"> ・AIのアルゴリズム等が不透明化し、人間にAIネットワークシステムの適正な制御が困難化
	制御喪失	<ul style="list-style-type: none"> ・AIネットワークシステムが暴走し、人間による制御が困難化ないし不可能化
法制度・権利利益に関するリスク	事故	<ul style="list-style-type: none"> ・自動運転車やロボット等の自律的判断に基づく動作による事故
	犯罪	<ul style="list-style-type: none"> ・AIネットワークシステムを悪用したマルウェアによる犯罪 ・自律型兵器のテロ等犯罪への悪用
	消費者の権利利益関連	<ul style="list-style-type: none"> ・AIネットワークシステムが適正に利活用されず、消費者、青少年等の権利利益が毀損
	プライバシー・個人情報関連	<ul style="list-style-type: none"> ・AIネットワークシステムによる個人情報の収集・利活用が不透明化し、個人情報のコントロールが困難化 ・AIネットワークシステムが人々の信念、健康、将来の行動等を推論し、プライバシーが侵害
	人間の尊厳と個人の自律関連	<ul style="list-style-type: none"> ・AIネットワークシステムが人間の意思決定過程を見えない形で操作することにより、個人の自律が侵害
	民主主義と統治機構関連	<ul style="list-style-type: none"> ・投票等国民の行動に対するAIネットワークシステムによる悪影響 ・AIネットワークシステムを国家の統治に利活用する場合における意思決定過程の不透明化や責任の所在の曖昧化

今後の課題

1. **研究開発の原則・指針の策定**：開発原則及びその内容の説明から構成される指針(「AI開発ガイドライン」(仮称))の策定に向けた議論の推進
2. **AIネットワーク化の進展に向けた協調の円滑化**：相互接続性・相互運用性の確保(確保の対象や方法の検討等)等
3. **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、ネットワーク形成に関する当事者間の協議の円滑化等
4. **経済発展・イノベーションの促進に向けた課題**：持続的な経済発展を実現するための取組の在り方の検討等
5. **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の継続
6. **利用者の保護**：関連する市場の動向の継続的注視、消費者保護に関する国際的な制度調和の在り方の検討等
7. **AIネットワークシステムに関するセキュリティの確保**：情報セキュリティ等のAIネットワークシステムへの実装の在り方の検討等
8. **プライバシー及びパーソナルデータに関する制度的課題**：プライバシー・パーソナルデータの保護と利活用のバランスに留意した検討等
9. **コンテンツに関する制度的課題**：多種多様かつ大量のコンテンツの創造・流通に適した法制度の在り方の検討等
10. **社会の基本ルールに関する検討**：ネットに関するルール(自己責任が基調)とモノの世界に関するルール(製造物責任が基調)との調和の在り方等
11. **リスクに関するシナリオの作成・共有**：AIネットワーク化の進展に応じたシナリオの継続的な作成・共有・見直し
12. **情報通信インフラの高度化の加速**：SDNに関する相互接続性・相互運用性の確保、試験環境の整備の推進等
13. **AIネットワーク・ディバイド形成の防止**：AIネットワーク・ディバイドの要因となるデジタル・ディバイドの解消、高齢者等の利用環境整備等
14. **人間の在り方に関する検討**：「AI依存」等人間の心理や子供の発育への影響等に関する継続的注視、教育の改革等
15. **AIネットワークシステムに関するリテラシーの涵養**：リテラシー教育の在り方の検討、科学コミュニケーションの在り方の検討
16. **AIネットワーク化に対応した人材育成**：技術者、法的・倫理的・社会的問題に対処し得る人材、産業連携等に対処し得る人材等の育成等
17. **AIネットワーク化に対応した就労環境の整備**：労働法制の在り方の検討等
18. **セーフティネットの整備**：所得の再分配の在り方(ベーシック・インカム等)の検討
19. **地球規模課題の解決を通じた人類の幸福への貢献**：途上国支援のための取組の在り方の検討
20. **AIネットワークシステムのガバナンスの在り方**：国際的な議論の場の形成、国内の議論の場の形成

研究開発の原則・指針の策定

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、AIの研究開発に関する原則及びその内容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、国際社会において関係ステークホルダーの参画を得て継続的な議論が行われるよう働きかけるべき。

研究開発に関する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。

① 透明性の原則

AIネットワークシステムの動作の説明可能性及び検証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するとともに、利用者を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に配慮すること。

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

ネットワーク化されるAI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

ネットワーク化されるAIの研究開発者が利用者等関係ステークホルダーへのアカウンタビリティを果たすこと。

(参考) G7香川・高松情報通信大臣会合(平成28年4月29日・30日)の結果

- 高市総務大臣から、G7各国が中心となり、OECD等国際機関の協力も得て、AIネットワーク化が社会・経済にもたらす影響、AIの開発原則の策定等AIネットワーク化をめぐる社会的・経済的・倫理的・法的課題に関し、産学民官の関係ステークホルダーの参画を得て国際的な議論を進めることを提案。

〔高市総務大臣からの提案に当たっては、AIネットワーク化検討会議中間報告書に掲げる8項目からなるAIの開発原則のたたき台を配付。〕

高市総務大臣の提案に対し、各国からの賛同が得られたところ。

なお、OECDデジタル経済に関する閣僚級会合(6月)においても、G7香川・高松情報通信大臣会合の成果を紹介し、AIネットワーク化が社会・経済にもたらす影響、AIの開発原則の策定等に関する国際的な議論の必要性を改めて提唱。

【参考】世界最先端IT国家創造宣言（平成28年5月20日改定・閣議決定）（抜粋）

＜本文＞

II. 「国から地方へ、地方から全国へ」（I T利活用の更なる推進のための3つの重点項目）

II - 2. 【重点項目2】安全・安心なデータ流通と利活用のための環境の整備

II - 2 - (1) 利用者志向のデータ流通基盤の構築

[主な取組内容]

(基盤を支える技術開発等)

- ・ **A Iの普及及びネットワーク化に伴う社会的・倫理的課題に関し、A Iの研究開発に当たっての留意事項に関する国際的に参照すべき原則の策定やその見直し等に関する国内外における継続的な議論の推進、その他所要の措置を講ずる。**

＜工程表＞

2. [重点項目2] 安全・安心なデータ流通と利活用のための環境の整備

(1) 利用者志向のデータ流通基盤の構築

○ A Iの普及及びネットワーク化に伴う社会的・倫理的課題に関する国内外の議論の推進

□ **A Iの普及及びネットワーク化並びにその社会や人間への影響を踏まえ、A Iの研究開発に当たり留意すべき事項（透明性、制御可能性等）に関し国際的に参照すべき原則の策定及びその継続的見直しその他関連する社会的・倫理的課題に関し、国内外における継続的な議論の推進に向けて所要の措置を講ずる。**

【総務省、内閣府】 <平成28年度～平成30年度> <平成31年度～平成33年度>

＜日本再興戦略 2016＞

第2 具体的施策

1. 第4次産業革命の実現

（2）新たに講ずべき具体的施策

i) 第4次産業革命の鍵を握る人工知能技術の研究開発と社会実装を加速するための司令塔機能の確立と規制・制度改革、企業や組織の垣根を越えたデータ利活用プロジェクト等の推進

① 産学官を糾合した人工知能技術に係る司令塔機能「人工知能技術戦略会議」の設置と人工知能技術の研究開発・社会実装の推進等

・（略）人工知能の普及及びネットワーク化並びにその社会や人間への影響を踏まえ、人工知能の開発に当たり留意すべき事項（透明性、制御可能性等）について国際的に参照すべき原則の策定及びその継続的見直しその他関連する社会的・倫理的課題に関し、国内外における継続的な議論を推進する。

＜経済財政運営と改革の基本方針 2016 ～600兆円経済への道筋～（骨太の方針）＞

第2章 成長と分配の好循環の実現

2. 成長戦略の加速等

（1）生産性革命に向けた取組の加速

③ 研究開発投資の促進

・（前略）人工知能の普及に伴う社会的・倫理的課題に関し国内外の議論を進める。（後略）

欧米の動向

- 欧米におけるAIの社会的・経済的・倫理的・法的課題に関する検討の動向
- AIの開発の在り方に関する欧米の議論
- 米国ホワイトハウス報告書「人工知能の未来に備えて」
- 「米国人工知能研究開発戦略」
- Partnership on AI（いわゆる「五社連合」）
- 英国下院科学技術委員会報告書「ロボティクスと人工知能」

欧米におけるAIの社会的・経済的・倫理的・法的課題に関する検討の動向

	米国	欧州
政府機関	<ul style="list-style-type: none"> ● ホワイトハウス 【2016年5月～】 大学やNPO等とともに、4回のワークショップを共催し、AIが社会にもたらす便益とリスクに関し検討。意見募集を経て、10月に報告書「人工知能の未来に備えて」を公表するとともに、「<u>米国人工知能研究開発戦略</u>」を策定。 	<ul style="list-style-type: none"> ● 欧州議会 【2016年4月～】 法務委員会が2016年4月にロボット・AIの法的・倫理的問題に関し公聴会を開催。5月には、<u>ロボットの民事法的ルール</u>に関し報告書の草稿を公表。10月には、関連するワークショップを開催。 ● 英国下院科学技術委「ロボティクスと人工知能」 【2016年10月】 ロボット・AIに関する社会的・倫理的・法的問題を検討した上で、<u>AIの開発・利用に関する原則の策定</u>やAIを巡る規制に関する政府への助言等を司る委員会をアラン・チューリング研究所の下に設置するよう提言。
企業、大学、NPO等	<ul style="list-style-type: none"> ● (Amazon、DeepMind/Google、Facebook、IBM、Microsoft) Partnership on AI 【2016年9月～】 <u>AI技術のベストプラクティスを研究して形成し、AIに関する公衆の理解を向上させ、AI及びその社会的影響に関する議論と関与のためのオープンなプラットフォームとするために設立。</u> ● (スタンフォード大学) One hundred year study of Artificial Intelligence (Stanford AI 100) 【2014年12月～】 AIが法制度、経済、社会等にもたらす長期的影響を調査。 ● Future of Life Institute (FLI) 【Jaana Tallinn (Skype共同開発者、DeepMind共同出資者)等、2014年3月～】 研究者、起業家等からなるボランティアベースの団体。<u>頑健かつ有益 (Robust and Beneficial) なAIを実現するための研究等を実施。</u> ● Open AI 【Elon Musk (テスラモーターズCEO)等、2015年12月～】 <u>安全なAIを構築して、AIの便益が可能な限り広範に等しく配分されることを目指して設立された非営利団体。</u> ● We Robot 【Michael Froomkin (マイアミ大学教授)等、2012年4月～2016年4月までに5回】 事故時の責任、プライバシー、人種差別等、<u>ロボットに関する法的問題</u>について議論。 	<ul style="list-style-type: none"> ● (サンタナ大学(伊国)等) RoboLaw: ロボット技術に対する法と倫理 【2012年3月～2014年5月】 <u>ロボット技術が現行の法的枠組みに適合し得るものであるのかを</u>広範に調査・考察し、<u>ロボット技術の発展が今後の社会規範、人々の価値観、社会的行動プロセスに与える影響を予測。</u> ● 英国規格協会「ロボット及びロボットシステムの倫理的設計・利用のためのガイド」 【2016年4月】 <u>ロボットの開発者が遵守すべき倫理的基準を提示。</u> ● (オックスフォード大学哲学部) Future of Humanity Institute (FHI) 【2015年～】 <u>技術が人類の未来に与える影響に注目する視点から、AIの制御可能性や安全性等について学際的に研究。</u> ● (ケンブリッジ大学) 未来のインテリジェンスのためのレバレッジ・センター (LCFI) 【2015年12月～】 <u>人類に利益をもたらすためのAIの在り方について、コンピュータ科学、認知学、哲学、社会学等多分野の研究を推進。</u>

AIの開発の在り方に関する欧米の議論

ホワイトハウス報告書「人工知能の未来に備えて」(2016年10月)「結論」

- ・実務家(practitioner)は、AI対応システム(AI-enabled systems)が以下のとおりであることを確保しなければならない。
 - ① 統御可能(governable)であること
 - ② オープンで、透明で、理解可能であること
 - ③ 人々と効果的に機能し得ること
 - ④ その操作は 人間の価値及び願望と一致し続けるであろうこと

ホワイトハウス「米人工知能研究開発戦略」(2016年10月) * 連邦政府の予算によるAI研究の方針を策定

(4) AIシステムの安全性及びセキュリティを確保する

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(safe)かつ堅牢(secure)で、制御され(controlled)、十分に定義され、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 あらかじめ確保しておくことが必要。

Partnership on AI「信条」(Tenets)(2016年9月) * ステークホルダーの関与と対話を重視

⑥ AIの便益を最大化し、AI技術の潜在的な課題に対処するため

- a. 個人の プライバシーとセキュリティを保護、b. AIの発展により影響を受け得るすべての当事者の利益を理解し、尊重
- c. AIの研究と技術のコミュニティが、AI技術の社会への 潜在的影響について社会的責任を負い、敏感であり、直接関与
- d. AIの研究と技術を 頑健で、信頼し得て、信用に値するものとするとともに、堅牢な制約の範囲内とすることを確保
- e. 国際条約や人権に反するAI技術の開発と利用に反対するとともに、危害を与えないセーフガードと技術を推進

⑦ AIシステムの動作は、その技術を説明するため、人々の理解と解釈が可能なるものであることが重要である

AI100 (One Hundred Year Study on Artificial Intelligence)「2030年における人工知能と生活」(2016年9月)

・将来への指針: 現状のデジタル技術について成功の見込みのある規制を導く原則が出発点を提供する。適切な規制は、透明性、説明責任、専門性の向上等の取組の好循環を強化する可能性がある。

欧州議会法務委員会「ロボティクスを巡る民事法ルールに関する欧州委員会への提言に伴う報告書案」(2016年5月)

・倫理原則: EUはロボット・AIの 開発・設計・利用において尊重されるべき基本的な倫理原則を策定する上で不可欠な役割を果たせる。倫理指針の枠組みは、恩恵、無害及び自律の原則並びに人間の尊厳等欧州基本権憲章に定められた原則に依拠すべき。

英国下院科学技術委員会報告書「ロボティクスと人工知能」(2016年10月)「結論と提言」

4. 我々の検討は、真剣な考慮を必要とする鍵となる 倫理的論点(検証及び妥当性確認、意思決定の透明性、偏見の最小化、アカウントビリティの向上、プライバシー及び安全)を照らし出した。実効的なガバナンスが継続的に評価され実行されるよう、これらの論点について継続的に注視することが必要。

5. AIの発展に関する社会的、倫理的及び法的含意について検討するため、アラン・チューリング研究所にAIに関する常設の委員会を設置するよう提言。委員会は、AI技術の開発及び利用を統治するための原則の策定等に注力すべき。

ホワイトハウス報告書「人工知能の未来に備えて」について(1/4)

- ホワイトハウスは、10月12日、国家科学技術会議(NSTC)及び科学技術政策局(OSTP)が中心となってまとめた報告書「Preparing for the Future of Artificial Intelligence」(人工知能の未来に備えて)を公表。
- 同報告書は、AIとは何かから、規制制度、研究開発、経済・雇用、公正性・安全性、安全保障まで幅広い項目に言及し、各項目に対応して、連邦政府機関等に対する23の提言(Recommendation)をまとめている。

報告書の主な章及び概要、提言の概要

報告書の主な章	概要
Introduction	AI開発史の概略、AIとは何か、AIの現状として機械学習やディープラーニング等を紹介
Public Outreach and Development of this Report	NSTC等がどのような会合及び手順を経て報告書をまとめたかを説明
Applications of AI for Public Good	AIが社会の広範囲に便益をもたらしつつあるとして、医療、交通等への導入事例・効果を紹介
	<p>提言 1:民間及び公的機関に、社会に便益をもたらす方法で如何にAI及び機械学習をしっかりと活用できるかの検討を推奨。先進技術やデータサイエンスと一般的に業務において関与していない社会正義及び公共政策に関する機関には、AI研究者等との連携検討を求める。</p> <p>提言 2:政府機関に、AIに対応したオープンデータに係る取組(政府による多数のデータセットの公表等)に優先順位を置くことを求める。</p>
AI in the Federal Government	政府でのAI活用の事例として国防高等研究計画局(DARPA)の新兵能力開発への活用を紹介
	<p>提言 3:連邦政府に、AIを業務に利用するため、重要な機関の能力向上策の調査を推奨。高リスク・高報酬なAI研究・利用を支援する機関の創設可能性の検討を求める。</p> <p>提言 4:NSTC機械学習及び人工知能(MLAI)小委員会に、実務家のための実践コミュニティを政府横断で展開することを求める。</p>
AI and Regulation	規制には、AI搭載機器によるリスク増減両面の評価が必要。また、規制担当部門に専門家が必要
	<p>提言 5:政府機関に、AI対応製品(AI-enabled products)に対する規制政策策定時に、上級レベルでの技術的専門知識の活用を推奨。必要な技術知識を持つ人材の補充等も求める。</p> <p>提言 6:政府機関に、技術の現状に関するより多様な視点をもって連邦の政府職員を育成するため、人員の割り当て等を求める。</p>

ホワイトハウス報告書「人工知能の未来に備えて」について(2/4)

報告書の主な章(つづき)	概要
Case Study: Autonomous Vehicles and Aircraft	自動運転車両及びドローンへの、運輸省(DOT)による適切な規制の適用の取組を紹介
	<p>提言 7: DOTに、安全、研究その他の目的でデータ共有を増加させる方法で、産業界及び研究者との協働を求める。</p> <p>提言 8: 米国政府に、自動操縦機とパイロットが操縦する航空機に適応可能な、先進的かつ自動的な航空交通制御システムの開発・実装への投資を求める。</p> <p>提言 9: DOTに、完全自動運転車両とドローンの交通システムへの安全な統合を可能とする、発展した規制枠組みの継続検討を求める。</p>
Research and Workforce	連邦政府機関によるAIの進展の注視、研究開発の支援及びAIに適合した労働力養成に言及
Monitoring Progress in AI	AIの進展を注視するに当たり、汎用AIの到来といったマイルストーンを識別するための視点を説明
	<p>提言10: NSTC MLAI小委員会に、AIの開発を注視し、AIの現状、特にマイルストーン(画期的出来事)について、上級の政府指導者への定期的な報告を求める。</p> <p>提言11: 政府に、他の諸国におけるAIの状況、特にマイルストーンについて注視することを求める。</p> <p>提言12: 産業界に、産業界におけるAIの一般的進展状況について政府が最新情報を持つ状態を維持するため、政府との協働を求める。</p>
Federal Support for AI Research	政府のAI研究開発投資の伸びは鈍い(約12億\$ (2016))。民間では長期の基礎研究は困難
	提言13: 連邦政府に、基礎的かつ長期間のAI研究に優先順位を置くことを求める。
Workforce Development and Diversity	大学等におけるAIに対応した教育の必要性、AIに対応した労働力の多様性の確保に言及
	提言14: NSTC MLAI小委員会及びネットワーキング・情報技術研究開発(NITRD)小委員会に、AIに対応した人材が供給される制度の検討着手を求める。
AI, Automation, and the Economy	AIによる自動化がGDP成長率にプラスとなる一方、低～中スキルの職が失われる可能性を説明
	提言15: 大統領府に、米国労働市場における自動化の効果を更に調査するため、2016年末までに後続レポートを発行し、推奨される政策対応の概説を求める。

ホワイトハウス報告書「人工知能の未来に備えて」について(3/4)

報告書の主な章(つづき)	概要
Fairness, Safety, and Governance	AIの決定について、いかに正義、公正性、アカウントビリティを確保するか懸念の声があることを紹介
Justice, Fairness, and Accountability	犯罪予測ツールに偏見が介在するとの意見を基に、正義、公正性、アカウントビリティの確保に言及
Safety and Control	AIの展開には安全性・制御可能性が必要との意見、それらの確保に向けた検討・取組を紹介
<p>提言16: 個人に関する重大な決定を行う又はサポートを提供するAIベースのシステムを利用する連邦政府機関に、証拠に基づいた検証及び妥当性確認によって、システムの有効性・公正性を確保するため特段の注意を求める。</p> <p>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。</p> <p>提言18: 大学等に、AI等に関するカリキュラムの必須の部分として、倫理、セキュリティ、プライバシー及び安全性に関するトピックを含めることを推奨。</p> <p>提言19: AI専門家、安全に関する専門家及びそれらの専門家ソサエティに、AI安全工学の分野の成熟に向けた進展を継続するため、協働することを求める。</p>	
Global Considerations and Security	AI研究開発に関する国際協力、サイバーセキュリティ、国防上の問題について説明
International Cooperation	AIがもたらす課題について、日本を含むバイやマルチ、G7、OECD等で扱ってきたことを説明
<p>提言20: 米国政府に、AIに関する国際的な関与についての政府規模の戦略策定、国際的な関与及び注視が必要なAI関連分野のリスト作成を求める。</p> <p>提言21: 米国政府に、情報交換及びAI研究開発協力を容易にするため、外国政府、国際機関、産業界、学界その他の重要な国際的ステークホルダーとの関与の深化を求める。</p>	
AI and Cybersecurity	AIが攻撃・防御双方により大きな役割を果たし得ること、AI自体の堅牢性確保の必要性を説明
<p>提言22: 政府機関の計画及び戦略に、サイバーセキュリティに関するAIの影響及びAIに関するサイバーセキュリティの影響についての十分な説明を求める。</p>	
AI in Weapon Systems	自律兵器システムに関する国際的な議論の推移、広範な防衛分野にAIが役立つことを説明
<p>提言23: 米国政府に、国際人道法と一致する、自律型・半自律型兵器に関する政府全体の政策策定の完遂を求める。</p>	
Conclusion	まとめとして、政府が果たすべき役割、実務家等が確保すべき事項等について説明

ホワイトハウス報告書「人工知能の未来に備えて」について(4/4)

報告書の主な章(つづき)	概要
Conclusion	<p>報告書「人工知能の未来に備えて」のまとめとして、主に以下の点を述べる。</p> <ol style="list-style-type: none">1. 政府が果たすべき役割<ul style="list-style-type: none">・重要な問題について対話の場を設け、公の討論のためのアジェンダ設定を助けること・利用の安全性及び公正性を注視し、規制枠組みを一般の人々を守りながらイノベーションを促進するものに適合させること・熟練した多様な労働力の開発に加えて、基礎研究及びAIの公益への利用を支援すること・より早く、より効果的にそしてより低いコストで公衆に仕えるため、AIそれ自体を利用すること・公共政策の多くの分野が、AIの継続的進展により勢いづけられる新しい機会及び新しい挑戦に気付くだろう。政府は、これらの変化を理解し適応するため、その能力を涵養し続けなければならない。2. 実務家等が確保すべき事項<ul style="list-style-type: none">・実務家は、AI対応システム(AI-enabled systems)が以下のとおりであることを確保しなければならない。<ol style="list-style-type: none">①管理可能(governable)であること②オープンで、透明で、理解可能であること③人々と効果的に機能し得ること④その操作は人間の価値及び願望と一致し続けるであろうこと研究者と実務家はこれらの課題に払う注意を増加させてきており、それらに焦点を合わせ続けるべきである。

ホワイトハウス報告書の提言等と報告書2016の「今後の課題」との比較

ホワイトハウス報告書の提言(23項目)及びConclusion(一部は概要)	
提言 1:	民間及び公的機関に、社会に便益をもたらす方法で如何にAI及び機械学習をしっかりと活用できるかの検討を推奨。先進技術やデータサイエンスと一般的に業務において関与していない社会正義及び公共政策に関する機関には、AI研究者等との連携検討を求めらる。
提言 2:	政府機関に、AIに対応したオープンデータに係る取組(政府による多数のデータセットの公表等)に優先順位を置くことを求める。
提言 3:	連邦政府に、AIを業務に利用するため、重要な機関の能力向上策の調査を推奨。高リスク・高報酬なAI研究・利用を支援する機関の創設可能性の検討を求める。
提言 4:	国家科学技術会議(NSTC)機械学習及び人工知能(MLAI)小委員会に、実務家のための実践コミュニティを政府横断で展開することを求める。
提言 5:	政府機関に、AI対応製品(AI-enabled products)に対する規制政策策定時に、上級レベルでの技術的専門知識の活用を推奨。必要な技術知識を持つ人材の補充等も求める。
提言 6:	政府機関に、技術の現状に関するより多様な視点をもって連邦の政府職員を育成するため、人員の割り当て等を求める。
提言 7:	運輸省(DOT)に、安全、研究その他の目的でデータ共有を増加させる方法で、産業界及び研究者との協働を求める。
提言 8:	米国政府に、自動操縦機とパイロットが操縦する航空機に適用可能な、先進的かつ自動的な航空交通管制システムの開発・実装への投資を求める。
提言 9:	DOTに、完全自動運転車両とドロースの交通システムへの安全な統合を可能とする、発展した規制枠組みの継続検討を求める。
提言10:	NSTC MLAI小委員会に、AIの開発を注視し、AIの現状、特にライルストーン(画期的出来事)について、上級の政府指導者への定期的な報告を求める。
提言11:	政府に、他の諸国におけるAIの状況、特にライルストーンについて注視することを求める。
提言12:	産業界に、産業界におけるAIの一般的進展状況について政府が最新情報を持つ状態を維持するため、政府との協働を求める。
提言13:	連邦政府に、基礎的かつ長期間のAI研究に優先順位を置くことを求める。
提言14:	NSTC MLAI小委員会及びネットワーク・情報技術研究開発(NITRD)小委員会に、AIに対応した人材が供給される制度の検討着手を求める。
提言15:	大統領府に、米国労働市場における自動化の効果を更に調査するため、2016年末までに後続レポートを発行し、推奨される政策対応の概説を求める。
提言16:	個人に関する重大な決定を行う又はサポートを提供するAIベースのシステムを利用する連邦政府機関に、証拠に基づいた検証及び妥当性確認によって、システムの有効性・公正性を確保するため特段の注意を求める。
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。
提言18:	大学等に、AI等に関するカリキュラムの必須の部分として、倫理、セキュリティ、プライバシー及び安全性に関するトピックを含めることを推奨。
提言19:	AI専門家、安全に関する専門家及びそれらの専門家リソースに、AI安全工学の分野の成熟に向けた進展を継続するため、協働を求める。
提言20:	米国政府に、AIに関する国際的な関与※についての政府規模の戦略策定、国際的な関与及び注視が必要なAI関連分野のリスト作成を求める。 ※防災及び災害対応、気候変動、デジタルデバイド等の地球規模問題への関与
提言21:	米国政府に、情報交換及びAI研究開発協力を容易にするため、外国政府、国際機関、産業界、学界その他の重要な国際的ステークホルダーとの関与の深化を求める。
提言22:	政府機関の計画及び戦略に、サイバーセキュリティに関するAIの影響及びAIに関するサイバーセキュリティの影響についての十分な説明を求める。
提言23:	米国政府に、国際人道法と一致する、自律型・半自律型兵器に関する政府全体の政策策定の完遂を求める。
Conclusion	1. 政府が果たすべき役割:利用の安全性及び公正性を注視し、規制枠組みを一般の人々を守りながらイノベーションを促進するものに適合させること など 2. 実務家等が確保すべき事項:AI対応システム(AI-enabled systems)が、①管理可能(governable)であること、②オープンで、透明で、理解可能であること、③人々と効果的に機能し得ること、④その操作は人間の価値及び願望と一致し続けるであろうこと など

報告書2016「今後の課題」(20項目)
1. 研究開発の原則・指針の策定 (1) 基本的な考え方 (2) 開発原則の内容 (3) 指針(「AI開発ガイドライン」(仮称))の策定に向けた検討
2. AIネットワーク化の進展に向けた協調の円滑化 (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 (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
3. 競争的なエコシステムの確保 (1) 関係する市場の動向の継続的注視 (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
4. 経済発展・イノベーションの促進に向けた課題 (1) AIネットワーク化による飛躍的かつ持続可能な経済発展を実現するための取組の在り方の検討 (2) AIネットワークシステムの研究開発及び利活用に関するイノベーションの促進
5. 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 (1) AIネットワーク化の進展が社会にもたらす影響に関する評価指標の設定 (2) 豊かさや幸せに関する評価指標の設定 (3) 指標に関する政府の取組の在り方の検討
6. 利用者の保護
7. AIネットワークシステムに関するセキュリティの確保
8. プライバシー及びバイオメトリクスに関する制度的課題
9. コンテンツに関する制度的課題
10. 社会の基本ルールに関する検討
11. リスクに関するシナリオの作成・共有
12. 情報通信インフラの高度化の加速
13. AIネットワーク・デバイス形成の防止
14. 人間の在り方に関する検討
15. AIネットワークシステムに関するリテラシーの涵養
16. AIネットワーク化に対応した人材育成
17. AIネットワーク化に対応した就業環境の整備
18. セーブネットの整備
19. 地球規模課題の解決を通じた人類の幸福への貢献
20. AIネットワークシステムのガバナンスの在り方

ホワイトハウス報告書の提言等と報告書2016「今後の課題」との比較(1/3)

ホワイトハウス報告書の提言(23項目)及びConclusion(一部は概要)	報告書2016「今後の課題」(20項目)のうち関連するもの
<p>提言 1:民間及び公的機関に、社会に便益をもたらす方法で如何にAI及び機械学習をしっかりと活用できるかの検討を推奨。先進技術やデータサイエンスと一般的に業務において関与していない社会正義及び公共政策に関する機関には、AI研究者等との連携検討を求める。</p>	<p>4. 経済発展・イノベーションの促進に向けた課題 (1) AIネットワーク化による飛躍的かつ持続可能な経済発展を実現するための取組の在り方の検討 (2) AIネットワークシステムの研究開発及び利活用に関するイノベーションの促進</p>
<p>提言 2:政府機関に、AIに対応したオープンデータに係る取組(政府による多数のデータセットの公表等)に優先順位を置くことを求める。</p>	<p>9. コンテンツに関する制度的課題 ・機械学習に適したオープンデータの提供の在り方の検討</p>
<p>提言 3:連邦政府に、AIを業務に利用するため、重要な機関の能力向上策の調査を推奨。高リスク・高報酬なAI研究・利用を支援する機関の創設可能性の検討を求める。</p>	<p>15. AIネットワークシステムに関するリテラシーの涵養 16. AIネットワーク化に対応した人材育成</p>
<p>提言 4:国家科学技術会議(NSTC)機械学習及び人工知能(MLAI)小委員会に、実務家のための実践コミュニティを政府横断で展開することを求める。</p>	<p>15. AIネットワークシステムに関するリテラシーの涵養 16. AIネットワーク化に対応した人材育成</p>
<p>提言 5:政府機関に、AI対応製品(AI-enabled products)に対する規制政策策定時に、上級レベルでの技術的専門知識の活用を推奨。必要な技術知識を持つ人材の補充等も求める。</p>	<p>16. AIネットワーク化に対応した人材育成 ・AIネットワークシステムに関する法的・倫理的・社会的問題に対処し得る人材の育成の在り方の検討</p>
<p>提言 6:政府機関に、技術の現状に関するより多様な視点をもって連邦の政府職員を育成するため、人員の割り当て等を求める。</p>	<p>— (政府全体の人員等、リソースに関することであるため)</p>
<p>提言 7:運輸省(DOT)に、安全、研究その他の目的でデータ共有を増加させる方法で、産業界及び研究者との協働を求める。</p>	<p>— (個別分野(運輸)に関することであるため)</p>
<p>提言 8:米国政府に、自動操縦機とパイロットが操縦する航空機に適応可能な、先進的かつ自動的な航空交通制御システムの開発・実装への投資を求める。</p>	<p>— (個別分野(運輸)に関することであるため)</p>
<p>提言 9:DOTに、完全自動運転車両とドローンの交通システムへの安全な統合を可能とする、発展した規制枠組みの継続検討を求める。</p>	<p>— (個別分野(運輸)に関することであるため)</p>
<p>提言10:NSTC MLAI小委員会に、AIの開発を注視し、AIの現状、特にマイルストーン(画期的出来事)について、上級の政府指導者への定期的な報告を求める。</p>	<p>3. 競争的なエコシステムの確保 (1) 関係する市場の動向の継続的注視</p>
<p>提言11:政府に、他の諸国におけるAIの状況、特にマイルストーンについて注視することを求める。</p>	<p>※AIネットワーク化検討会議の設置の前提として、そもそもAIに関する海外動向の注視が含まれていたものである。</p>
<p>提言12:産業界に、産業界におけるAIの一般的進展状況について政府が最新情報を持つ状態を維持するため、政府との協働を求める。</p>	<p>— (政府全体としての産業界に対する要請であるため)</p>

ホワイトハウス報告書の提言等と報告書2016「今後の課題」との比較(2/3)

ホワイトハウス報告書の提言(23項目)及びConclusion (一部は概要)	報告書2016「今後の課題」(20項目)のうち関連するもの
<p>提言13:連邦政府に、基礎的かつ長期間のAI研究に優先順位を置くことを求める。</p>	<p>— (政府全体のAI研究開発に係る方針であるため)</p>
<p>提言14:NSTC MLAI小委員会及びネットワーキング・情報技術研究開発(NITRD)小委員会に、AIに対応した人材が供給される制度の検討着手を求める。</p>	<p>— (政府全体の人員等、リソースに関することであるため)</p>
<p>提言15:大統領府に、米国労働市場における自動化の効果を更に調査するため、2016年末までに後続レポートを発行し、推奨される政策対応の概説を求める。</p>	<p>17.AIネットワーク化に対応した就労環境の整備</p>
<p>提言16:個人に関する重大な決定を行う又はサポートを提供するAIベースのシステムを利用する連邦政府機関に、証拠に基づいた検証及び妥当性確認によって、システムの有効性・公正性を確保するため特段の注意を求める。</p>	<p>1. 研究開発の原則・指針の策定 (2) 開発原則の内容 ①透明性の原則 ⑦倫理の原則 ⑧アカウントビリティの原則</p>
<p>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。</p>	<p>1. 研究開発の原則・指針の策定 (2) 開発原則の内容 ①透明性の原則 ⑦倫理の原則 ⑧アカウントビリティの原則</p>
<p>提言18:大学等に、AI等に関するカリキュラムの必須の部分として、倫理、セキュリティ、プライバシー及び安全性に関するトピックを含めることを推奨。</p>	<p>7. AIネットワークシステムに関するセキュリティの確保 8. プライバシー及びパーソナルデータに関する制度的課題 14. 人間の在り方に関する検討 ・AIネットワーク化を見据えた教育内容及び方法の改革の在り方の検討 15. AIネットワークシステムに関するリテラシーの涵養 16. AIネットワーク化に対応した人材育成</p>
<p>提言19:AI専門家、安全に関する専門家及びそれらの専門家ソサエティに、AI安全工学の分野の成熟に向けた進展を継続するため、協働することを求める。</p>	<p>1. 研究開発の原則・指針の策定 (2) 開発原則の内容 ③制御可能性の原則 ⑤安全保護の原則 7. AIネットワークシステムに関するセキュリティの確保</p>
<p>提言20:米国政府に、AIに関する国際的な関与※についての政府規模の戦略策定、国際的な関与及び注視が必要なAI関連分野のリスト作成を求める。 ※防災及び災害対応、気候変動、デジタルディバイド等の地球規模問題への関与</p>	<p>19. 地球規模課題の解決を通じた人類の幸福への貢献 13. AIネットワーク・ディバイド形成の防止</p>

ホワイトハウス報告書の提言等と報告書2016「今後の課題」との比較(3/3)

ホワイトハウス報告書の提言(23項目)及びConclusion (一部は概要)	報告書2016「今後の課題」(20項目)のうち関連するもの
<p>提言21: 米国政府に、情報交換及びAI研究開発協力を容易にするため、外国政府、国際機関、産業界、学界その他の重要な国際的ステークホルダーとの関与の深化を求める。</p>	<p>20. AIネットワークシステムのガバナンスの在り方 ・AIネットワークシステムに関する国際的なルール形成過程への参画の機会と透明性の確保の在り方の検討</p>
<p>提言22: 政府機関の計画及び戦略に、サイバーセキュリティに関するAIの影響及びAIに関するサイバーセキュリティの影響についての十分な説明を求める。</p>	<p>7. AIネットワークシステムに関するセキュリティの確保</p>
<p>提言23: 米国政府に、国際人道法と一致する、自律型・半自律型兵器に関する政府全体の政策策定の完遂を求める。</p>	<p>— (個別分野(国防)に関することであるため)</p>
<p>Conclusion:</p> <p>1. 政府が果たすべき役割</p> <ul style="list-style-type: none"> ・重要な問題について対話の場を設け、公の討論のためのアジェンダ設定を助けること ・利用の安全性及び公正性を注視し、規制枠組みを一般の人々を守りながらイノベーションを促進するものに適合させること ・熟練した多様な労働力の開発に加えて、基礎研究及びAIの公益への利用を支援すること ・より早く、より効果的にそしてより低いコストで公衆に仕えるため、AIそれ自体を利用すること ・公共政策の多くの分野が、AIの継続的進展により勢いづけられる新しい機会及び新しい挑戦に気付くだろう。政府は、これらの変化を理解し適応するため、その能力を涵養し続けなければならない。 <p>2. 実務家等が確保すべき事項</p> <ul style="list-style-type: none"> ・実務家は、AI対応システム(AI-enabled systems)が以下のとおりであることを確保しなければならない。 <ol style="list-style-type: none"> ①管理可能(governable)であること ②オープンで、透明で、理解可能であること ③人々と効果的に機能し得ること ④その操作は人間の価値及び願望と一致し続けるであろうこと <p>研究者と実務家はこれらの課題に払う注意を増加させてきており、それらに焦点を合わせ続けるべきである。</p>	<p>1. 研究開発の原則・指針の策定 (2) 開発原則の内容</p> <ol style="list-style-type: none"> 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 ⑤安全保護の原則 ⑦倫理の原則 ⑧アカウントビリティの原則 <p>6. 利用者の保護</p> <p>10. 社会の基本ルールに関する検討</p>

ホワイトハウス報告書の提言16～19等と「AI開発原則」の8項目との比較(1/2)

- ホワイトハウス報告書では、**AIの決定又はAI自体が具備すべき正義、公正性、アカウントビリティ、安全性及び制御可能性等**に関し、「Fairness, Safety, and Governance」の章において論じた上で、**提言16から提言19までを掲げている**。
また、**同章及び「Conclusion」の章においては、これらに関係する重要な記載**が見受けられるところ。
- 提言16～19及びこれらの記載とAIネットワーク化検討会議「報告書2016」の「AI開発原則」の8項目を比較すると、次のとおり。

ホワイトハウス報告書「Preparing for the Future of Artificial Intelligence」提言16～19 及び「Fairness, Safety, and Governance」の章の記載 (一部は概要)

提言16: 個人に関する重大な決定を行う又はサポートを提供するAIベースのシステムを利用する連邦政府機関に、証拠に基づいた検証及び妥当性確認によって、システムの有効性・公正性を確保するため特段の注意を求める。

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。

提言18: 大学等に、AI等に関するカリキュラムの必須の部分として、倫理、セキュリティ、プライバシー及び安全性に関するトピックを含めることを推奨。

提言19: AI専門家、安全に関する専門家及びそれらの専門家ソサエティに、AI安全工学の分野の成熟に向けた進展を継続するため、協働することを求める。

「Fairness, Safety, and Governance」の章の記載①

公正性及び安全性の課題どちらの場合でも、実務家は、①意図的な差別又は失敗を避け、②意図しない結果を回避し、③意図しない失敗はほとんど無いという正当化された確信をステークホルダーに与えるために必要な証拠の形成に努めるものとする。

「Fairness, Safety, and Governance」の章の記載②

・技術が説明責任を確保する助けとなる

実務家は、AIシステムをより正しく、公正に、十分な説明が可能となるように努めるほど、しばしば、技術がアカウントビリティの妨げとなるよりむしろその助けとなる機会が存在する。機械学習結果の解釈可能性を改善するための研究は、その一例である。人々が決定を理解する助けとなる説明可能なモデルを持つことは、人々にその背後にある前提条件及びプロセスを調べる力を与える。

AI開発原則の8項目

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

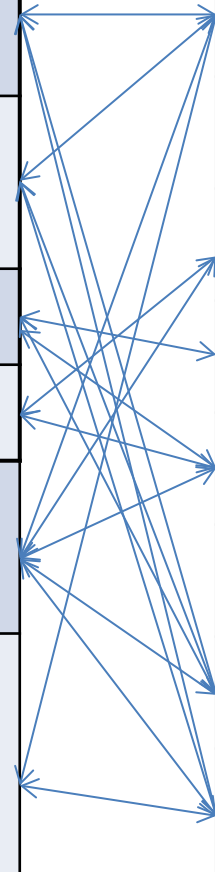
④セキュリティ確保の原則

⑤安全保護の原則

⑥プライバシー保護の原則

⑦倫理の原則

⑧アカウントビリティの原則



ホワイトハウス報告書の提言16～19等と「AI開発原則」の8項目との比較(2/2)

ホワイトハウス報告書「Preparing for the Future of Artificial Intelligence」の「Fairness, Safety, and Governance」の章及び「Conclusion」の章の記載 (一部は概要)

「Fairness, Safety, and Governance」の章の記載③

・安全性と制御可能性の懸念がAIの展開の妨げとなる

(ワークショップにおける、現実世界でのAIの安全性及び制御可能性に関するAI専門家の懸念を踏まえ、以下のとおりである。)

現実世界でのAIの展開を制限する主な要因の1つは安全性と制御の懸念である。もし実務家が、システムは安全で制御可能であり、システムの展開は深刻で否定的な結果の受け入れがたいリスクを生み出さないとの正当化された確信に至ることができなければ、その時システムは展開されないべきではない。

「Fairness, Safety, and Governance」の章の記載④

・「開かれた世界」に安全に移行するシステムを作る必要性

安全性と制御における主な課題は、研究室の「閉鎖された世界」から外側の予期しない出来事が起こり得る「開かれた世界」に安全に移行するシステムを作ることである。開かれた世界では、システムはそれが設計・製造された時には予期していなかった対象や状況に遭遇する可能性がある。不測の事態に適切に適応することは、まだ困難であるが、安全な操作のために必要である。

「Conclusion」章の記載

・実務家等が確保すべき事項

実務家は、AI対応システム(AI-enabled systems)が以下のとおりであることを確保しなければならない。

①管理可能(governable)であること

②オープンで、透明で、理解可能であること

③人々と効果的に機能し得ること

④その操作は人間の価値及び願望と一致し続けるであろうこと

研究者と実務家はこれらの課題に払う注意を増加させてきており、それらに焦点を合わせ続けるべきである。

AI開発原則の8項目

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

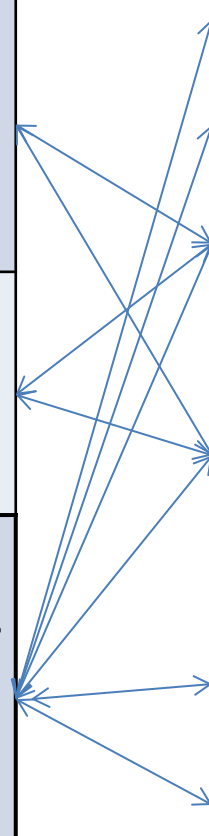
④セキュリティ確保の原則

⑤安全保護の原則

⑥プライバシー保護の原則

⑦倫理の原則

⑧アカウントビリティの原則



「米国人工知能研究開発戦略」(2016年10月) (1/2)

- 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)は、2016年10月13日付けで、「米国人工知能研究開発戦略」として、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。
- 同戦略は、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。
- 同戦略は、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。

研究開発戦略の概要

(1) AI研究に長期的投資を行う

新たな発見・洞察を駆り立て、米国がAI分野の世界のリーダーであり続けるための次世代のAIへの投資を優先。

(2) 人間とAIの協働に向けて効果的な方法を開発する

大部分のAIシステムは、人間に取って代わるものではなく、最適なパフォーマンスを達成するために人間と協働することが見込まれることから、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するための研究が必要。

(3) AIの倫理的、法的及び社会的含意を理解し、それらに対処する

AI技術は人間が拘束される公式及び非公式の規範に従って行動することが期待されるため、AIの倫理的、法的及び社会的含意を理解し、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。

「米国人工知能研究開発戦略」(2016年10月) (2/2)

(4) AIシステムの安全性及びセキュリティを確保する

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(safe)かつ堅牢(secure)で、制御され(controlled)、十分に定義され、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しておくことが必要。

信頼し得る(reliable)AIシステムを創り出す上での課題に対処するための研究の更なる発展が必要。

(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する

訓練用のデータベース及び資源の奥深さ、質、正確さは、AIのパフォーマンスに大きく影響。

質の高いデータベース及び環境を開発し、質の高いデータベース及び訓練・試験用の資源への責任あるアクセスを可能にするための研究が必要。

(6) 基準及びベンチマークを通じてAI技術を計測し評価する

AIの発展のためには、AIの進展を方向付けて評価するための基準、ベンチマーク、テストベッド及び共同体の関与が不可欠。

幅広い評価手法を開発するための研究も必要。

(7) 国家のAI研究開発人材のニーズをより良く理解する

AIの発展には、AI研究者の強力な共同体が必要。

本戦略で示された戦略的な研究開発領域に取り組むために十分なAI専門家が確保されるように促すために、AI分野の研究開発人材の現在及び将来の需要に関する理解を促進することも必要。

「米人工知能研究開発戦略」に掲げる社会的・法的・倫理的課題と「AI開発原則」の8項目との比較(1/2)

米人工知能研究開発戦略

(3)AIの倫理的、法的及び社会的含意を理解し、それらに対処する (概要)

AI技術は人間が拘束される公式及び非公式の規範に従って行動することが期待されるため、AIの倫理的、法的及び社会的含意を理解し、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。

※ 考慮することが必要な事項として、プライバシーについて言及。

・**バイ・デザイン**によりAIの公正 (fairness)、透明性及びアカウントビリティを促進するAIシステムのためのデータの適切な収集及び利用の在り方が重要な課題。
また、研究者は、AIの動作及び意思決定が透明で、人間により容易に解釈可能なものとするにより、システムが含み得る偏見を、単に学習し反復するのではなく、検査することが可能となるようAIシステムを設計する方法を検討することが必要。

・倫理的AIを構築する

技術的な実現可能性の範囲内で、研究者は、検証可能な形で、現行の法、社会規範及び倫理に合致し、又はこれらを遵守するアルゴリズム及びアーキテクチャを開発するよう努めることが必要。

AIに倫理を実装することには様々な困難が伴うが、AIシステムの判断結果及び動作を説明し正当化するために、AIシステムの推論及び意思決定を指導する受容可能な倫理の参照枠組みを開発することは可能。

・倫理的推論を行うAIシステムのアーキテクチャを設計する

倫理的推論を組み込んだAIシステムのアーキテクチャをいかに最善の形で設計するかを判断するための基礎研究の発展も必要。

AIネットワーク化検討会議 AI開発原則の8項目

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

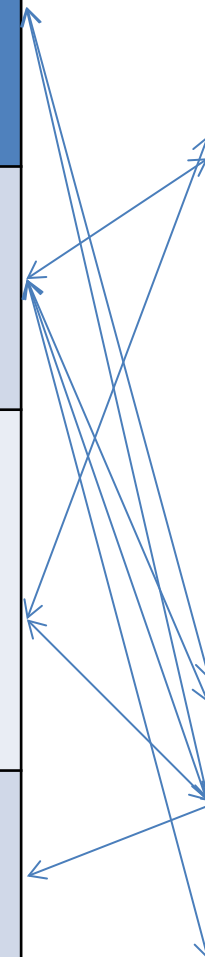
④セキュリティ確保の原則

⑤安全保護の原則

⑥プライバシー保護の原則

⑦倫理の原則

⑧アカウントビリティの原則



「米人工知能研究開発戦略」に掲げる社会的・法的・倫理的課題と「AI開発原則」の8項目との比較(2/2)

米人工知能研究開発戦略

(4) AIシステムの安全性及びセキュリティを確保する (概要)

AIシステムが広く利用されるようになる前に、AIシステムが安全(safe)かつ堅牢(secure)で、制御され(controlled)、十分に定義され、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おくことが必要。

信頼し得る(reliable)AIシステムを創り出す上での課題に対処するための研究の更なる発展が必要。

・説明可能性(explainability)及び透明性を促進する

研究者は、透明性があり、利用者に対して判断結果の理由を実質的に説明することのできるAIシステムを開発することが必要。

・信用(trust)を構築する

信用を獲得するため、AIシステムの設計者は、インフォマティブでユーザーフレンドリーなインタフェースを有し、正確で、信頼可能(reliable)なシステムを創造することが必要。その一方で、操作者(operator)は、システムの操作及びパフォーマンスの限界を理解するために十分な訓練の時間を確保することが必要。

・検証(verification)及び妥当性確認(validation)を促進する

AIシステムの検証及び妥当性確認のための新たな方法が必要。
安全なAIシステムは、新たな評価、診断、修理の手法を必要とするかもしれない。

・攻撃に対するセキュリティを確保する

重要なシステムに実装されたAIは、事故に対処しうるために頑健(robust)でなければならないとともに、意図的なサイバー攻撃に対して堅牢(secure)でもあるべき。

・長期的なAIの安全性及び価値整序の達成

AIシステムはしだいに「再帰的な自己改良」が可能になっていくかもしれない。
自己修正システムの安全性を確保するために、自己監視アーキテクチャ、囲い込み戦略、価値学習、価値枠組みを含む、更なる研究が必要。

AIネットワーク化検討会議 AI開発原則の8項目

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

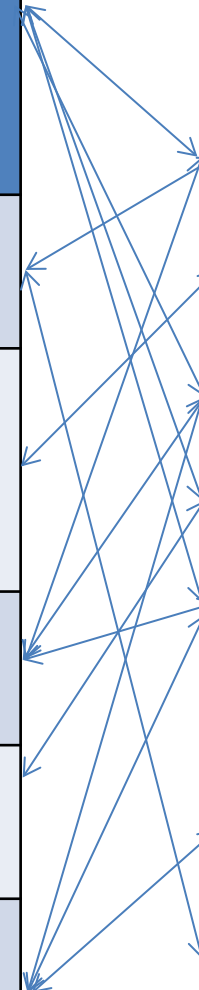
④セキュリティ確保の原則

⑤安全保護の原則

⑥プライバシー保護の原則

⑦倫理の原則

⑧アカウントビリティの原則



Partnership on AI (いわゆる「五社連合」)について

正式名称: Partnership on Artificial Intelligence to Benefit People and Society

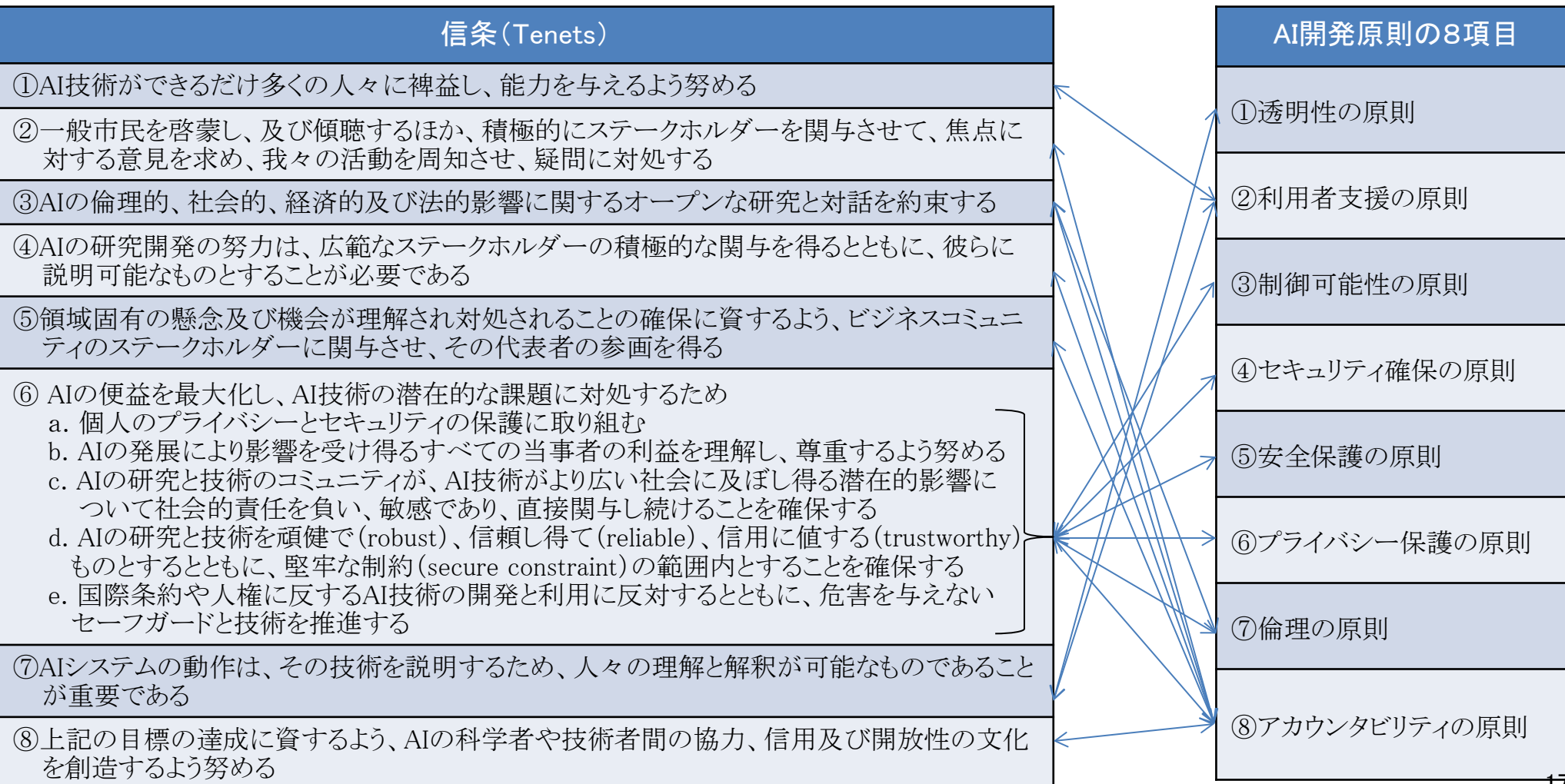
設立: 2016年9月28日(水)

構成企業: Amazon、DeepMind/Google、Facebook、IBM、Microsoft



暫定共同議長: エリック・ホロヴィッツ (Microsoft 技術フェロー) 及びムスタファ・スレイマン (DeepMind 共同創業者)

【五社連合の設立時に公表した「信条」(Tenets)と、AIネットワーク化検討会議が「AI開発原則」につき掲げる8項目との比較】



英国下院科学技術委員会報告書「ロボティクスと人工知能」(2016年10月)

○ 英国下院科学技術委員会は2016年9月13日付けで、報告書「ロボティクスと人工知能」を取りまとめ、下院は10月12日付けで、報告書を公表した。報告書は、**ロボット・AIに関する社会的・倫理的・法的问题を検討した上で、AI技術の開発・利用に関する原則の策定等を行うための委員会の設置などを提言している。**

報告書の構成

1. イントロダクション

・問題意識

2. 経済的・社会的含意

・機械対人間

・教育と能力

3. 倫理的・法的論点

・安全性と制御

・ガバナンス: 基準と規制

4. 研究、助成及びイノベーション

・ロボティクスと自律システム

結論と提言

「結論と提言」の要旨

・教育と能力

1. 政府は、教育及び訓練のシステムが雇用の変化に適応し、生涯教育に対応できる柔軟なものとなるようコミットmentする必要。
2. デジタル排除は21世紀の英国において存在してはならない。政府は、デジタル能力の危機に対処するようコミットすべき。

・ガバナンス: 基準と規制

3. AIという生成途上の領域に分野全体に及ぶ規制を設けることは時期尚早である一方、AIシステムの倫理的、法的及び社会的次元に関する注意深い精査を開始することが不可欠。
4. 我々の検討は、真剣な考慮を必要とする鍵となる倫理的論点(検証及び妥当性確認、意思決定の透明性等)を照らし出した。実効的なガバナンスが継続的に評価され実行されるよう、これらの論点について継続的に注視することが必要。
5. AIの発展に関する社会的、倫理的及び法的含意について検討するため、アラン・チューリング研究所にAIに関する常設の委員会を設置するよう提言。委員会は、AI技術の開発及び利用を統治するための原則の策定等に注力すべき。
6. 委員会の構成員は、コンピュータ科学者、自然科学者、数学者及び工学者のみならず、法学、社会科学及び哲学の専門家、産業界、NGO及び公共部門の代表者を含むべき。広範な公共的対話のプログラムも開始されるべき。

・研究、助成及びイノベーション

7. ロボティクス・AI技術の取り込みにより生産性の向上を達成するために、この分野に関する政府の戦略が必要。
8. 政府は、学界、産業界及び政府の代表者から構成され、AIの研究開発戦略の策定等を掌るRAS(ロボティクス及び自律システム)リーダーシップ会議を速やかに設置すべき。

英国下院科学技術委員会報告書「ロボティクスと人工知能」に掲げる倫理的・法的論点と「AI開発原則」の8項目との比較(1/2)

英国下院科学技術委員会報告書「ロボティクスと人工知能」

3. 倫理的・法的論点

□安全性と制御(control) (概要)

□検証(Verification)と妥当性確認(validation)

AIシステムが意図された通りに動作するよう、かつ、事故若しくは悪意により欲しない又は予期し得ない行動が生み出されないよう確保する必要がある。したがって、システムが正しく機能することを検証する方法が必要である。

□意思決定の透明性

AIが重要な利益に関わる判断を行う場合には、AIの意思決定過程の不透明性は、人々のAIの判断に対する不信を招く可能性がある。

□偏見の最小化

偏見及び差別が偶然にAIシステムに組み込まれてしまう事例が明らかになっている。

□プライバシーと同意

医療分野等におけるAIによるデータ分析がもたらす便益とともに、プライバシーに関する課題が懸念される。

□アカウントビリティと責任(liability)

自動走行車や自律型兵器に関して事故等に関する責任の配分やアカウントビリティの在り方が問題になっている。

AIネットワーク化検討会議 AI開発原則の8項目

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

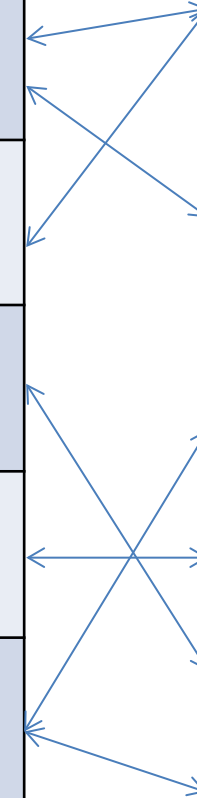
④セキュリティ確保の原則

⑤安全保護の原則

⑥プライバシー保護の原則

⑦倫理の原則

⑧アカウントビリティの原則



英国下院科学技術委員会報告書「ロボティクスと人工知能」に掲げる提言と「AI開発原則」の8項目との比較(2/2)

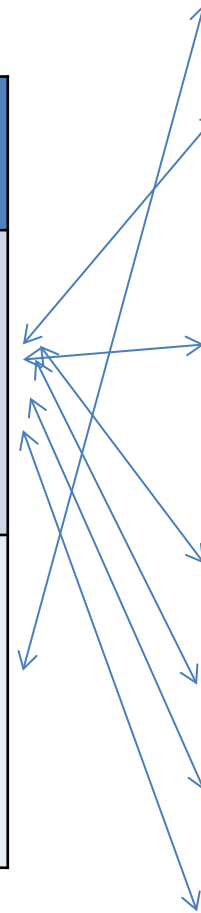
英国下院科学技術委員会報告書「ロボティクスと人工知能」 結論と提言 ・ガバナンス:基準と規制 (概要)

4. 我々の検討は、真剣な考慮を必要とする鍵となる倫理的論点の多く—検証及び妥当性確認、意思決定の透明性、偏見の最小化、アカウントビリティの向上、プライバシー及び安全を照らし出した。AIの領域は急速に発展し続けているので、実効的なガバナンスが継続的に評価され実行されるよう、これらの論点について継続的に注視することが求められる。

5. 我々は、AIの現在及び潜在的な発展に関する社会的、倫理的及び法的含意について検討するため、アラン・チューリング研究所にAIに関する常設の委員会を設置することを提言する。委員会は、AI技術の開発及び利用を統治するための原則を策定するとともに、AI技術の開発及び利用の進展の制約に関し必要とされる規制 (any regulation required on limits to its progression) に関する政府への助言に注力すべきである。また、委員会は、政府が設置しようとしているデータ倫理会議 (the Council of Data Ethics) と密接に協調する必要がある。

AIネットワーク化検討会議 AI開発原則の8項目

- 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プライバシー保護の原則
- ⑦倫理の原則
- ⑧アカウントビリティの原則



A I ネットワーク化が社会・経済にもたらす 影響とリスクの評価について

AIネットワーク化の影響・リスクの評価の枠組み（案）

A I ネットワーク化の影響・リスクの評価に関する検討事項

- 具体的な利活用の場面（ユース・ケース）を想定した影響評価（シナリオ作成、分野別展望、評価指標）
- 具体的な利活用の場面（ユース・ケース）を想定したリスク評価（シナリオ分析）

※ これら影響評価及びリスク評価は、次の二つを並行して実施。

- ① 「A I 開発ガイドライン（仮称）」素案ver.1.0の検討に向けた**先行的評価**（平成29年夏頃まで実施）
- ② A I ネットワークシステムを利活用する分野ごとの**分野別評価**（分野の区分を決定次第、継続的に実施）

※ 影響評価とリスク評価は、可能な限り一体的に実施。

- A I ネットワークシステムに関するエコシステムの変化
- A I ネットワーク化の進展に伴い実現される社会の全体像の概況
- その他A I ネットワーク化の影響・リスクの評価に関する事項

影響評価及びリスク評価の対象とする具体的な利活用の場面（ユース・ケース）の候補

【先行的評価】・内閣府「人工知能と人間社会に関する懇談会」及びC O C Nが掲げるユース・ケース、
・構成員に御提供いただくユース・ケース、事務局が提供するユース・ケース

【分野別評価】A I ネットワークシステムを利活用する分野の構成について、利用者の視点から区分した上で
（ユーザー・セントリック・アプローチ）、当該分野ごとに構成員に御提供いただくユース・ケース
及び事務局が提供するユース・ケース

（参考）AIネットワーク化検討会議「中間報告書」において影響評価の対象とした分野の区分（全16分野）

【公共（まち）】 公共インフラ、防災、スマートシティ、行政

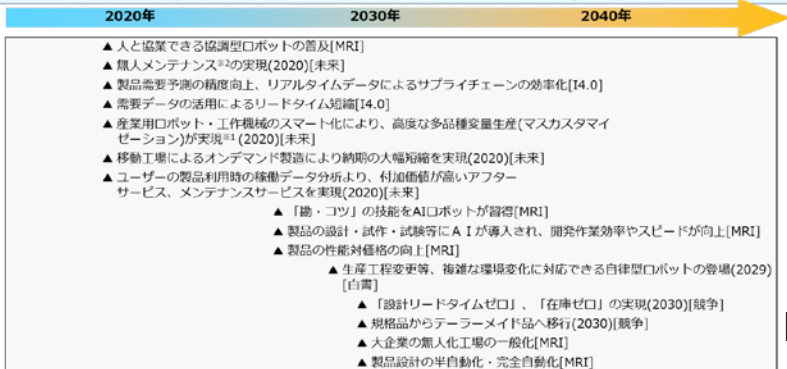
【生活（ひと）】 生活支援、豊かさ創造（創造的活動、コミュニティ活動等）

【産業（しごと）】 分野共通（コーポレート業務、バックオフィス等）、農林水産、製造業、運輸・物流、
卸売・小売、金融・保険、医療・介護、教育・研究、サービス業、建設）

影響評価の枠組み（案）

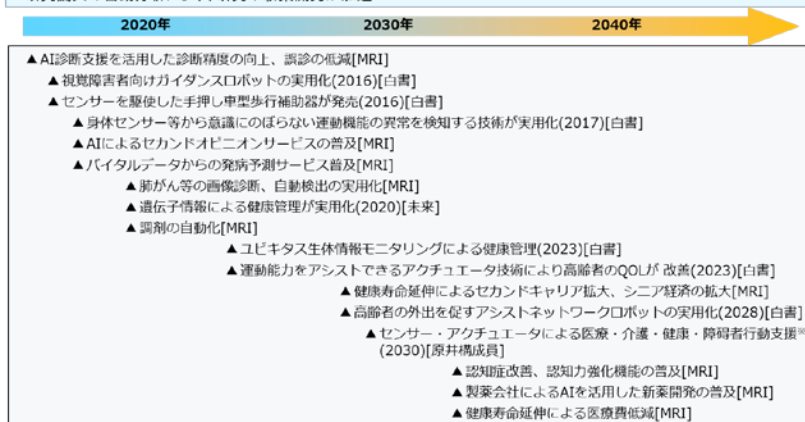
- 先行的評価及び分野別評価の双方において、具体的な利活用の場面を想定したシナリオを作成する。
- シナリオに基づき影響を評価した上で、A I ネットワーク化検討会議の「中間報告書」で提示したA I ネットワーク化の影響の展望（現在～2045年頃までの変化の展望）を必要に応じ見直し、深化及び拡充を図る。

- 製造プロセスとサプライチェーンのスマート化により、動的な需給バランスに対応した生産最適化や高度な多品種変量生産（マスカスタマイゼーション）を実現
- 利用者の稼働データの分析により、デジタルマーケティングや、付加価値が高いアフターサービス・メンテナンスサービスを実現
- 2020年代後半以降には、設計段階からの自動化が実現することにより、開発作業が効率化・高速化



【例】 製造業

- 患者のバイタルデータによる発病予測や、遺伝子情報による健康管理が実現し、健康寿命が延伸
- 研究論文の自動分析により、研究や新薬開発が加速



【例】 医療・介護

※ 具体例：事故状況の監視等の救助支援、異常行動の見守りや介護・自立支援などの福祉・インフラ整備、双方の遠隔医療 等

- 分野別評価に当たっては、可能であれば、分野ごと又は分野横断の影響について、定量的に測定し、又は客観的に把握することができる評価指標の設定についても検討する。

(参考)AIネットワーク化検討会議「報告書2016」に掲げる評価指標の例

分野（例）		考えられる評価指標の項目（例）
公共分野	公共インフラ	リアルタイムで収集・分析することができるデータの種類、量等
	防災	リアルタイム予測の高度化の状況（タイムラグ、精緻化の状況等）
	行政	公開されるオープンデータの種類、量等
生活分野	豊かさ創造	3Dプリンター等のパーソナルファブリケーションの普及状況（台数等）
産業分野	農林水産	農業用ドローン、インテリジェントファーミング等の導入・実現状況（導入している農家数等）
	運輸・物流	自動運転の実現状況（普及台数等）、人間とAIの分業率
	金融・保険	トレーディング、ローン審査、与信管理の自動化の普及状況
	サービス業	ロボット等による自動化の状況、人間とAIの分業率
	建設	ロボット等による自動化の状況、人間とAIの分業率

リスク評価（シナリオ分析）の枠組み（案）

先行的評価及び分野別評価の双方ともに、具体的な利活用の場面を想定したシナリオを作成し、当該シナリオに基づき、次に掲げる要素から構成されるリスク・シナリオ分析を行う。

○ リスクの種類・種類

- 機能に関するリスク（セキュリティ、情報通信ネットワークシステム、不透明化、制御喪失）
- 法制度・権利利益に関するリスク（事故、犯罪、消費者等の権利利益、プライバシー・個人情報、人間の尊厳と個人の自律、民主主義と統治機構）

○ 利活用の場面に即して想定される（第一次的な）リスクの内容

○ リスクへの対処

- リスク評価 ①発生時期 ②生起確率 ③被害の規模 ④二次的（派生的）リスクの評価
- リスク管理 ①事前の予防・モニタリング ②被害発生時の対処 ③事後的な措置・復旧 等
- リスク・コミュニケーション ①ステークホルダーの特定 ②インシデントと対処の説明 ③シナリオの共有 等

リスクの種類 (例)	利活用の場面に即して 想定されるリスクの内容(例)	発生時期	生起 確率	被害の 規模	二次的(派生的) に生ずる リスク	リスク評価	リスク管理	リスク・ コミュニケーション
セキュリティに関するリスク	ロボットに関係するクラウド等AIネットワークシステムがハッキング攻撃されることにより、情報が流出したり、ロボットが不正に操作されるリスク	進展 段階2	高	大	プライバシー・個人情報に関するリスク、消費者等の権利利益に関するリスク、犯罪のリスク	情報流出による影響の評価、クラウドのセキュリティ上の脆弱性等の評価	事後検証のためのロギングの実装、脆弱性の発見・対処、必要に応じて結合テストの追試	生じたインシデントに関する情報共有と対応策についての説明
事故のリスク	レベル3の自動走行車の運転時に運転者がハンドルから手を離して乗ることにより、緊急時の対応が困難になるリスク	進展 段階1	中	中	制御喪失のリスク	運転者の生命・身体へのリスク、関連車両群や交通システム等への影響等の評価	運転者の技能維持のための試験・講習等、リスク改善に向けた作業・モニタリング等	運転者への啓発、緊急事の情報共有等

AIネットワークシステムに関するエコシステムの変化の展望（案）

- （A I ネットワーク化を含む）I C T インテリジェント化の動向全体を踏まえつつ、A I ネットワークシステムに関するエコシステムの変化を展望する。
- 検討に当たっては、並行して行う影響評価の結果等を随時反映させる。

「ICTインテリジェント化」

ICTに関する次に掲げる同時並行的かつ加速度的な技術の高度化など6要素によりもたらされる変化

6要素

- ①CPU, ストレージ、通信ネットワークの能力の加速度的向上
- ②人工知能の高度化
- ③あらゆるものごとのデータ化
- ④インターネットのグローバル化
- ⑤クラウド、ネットワーク、エッジの各階層を活用した分散処理
- ⑥人間（の脳）と人工知能等との連携、意識の通信

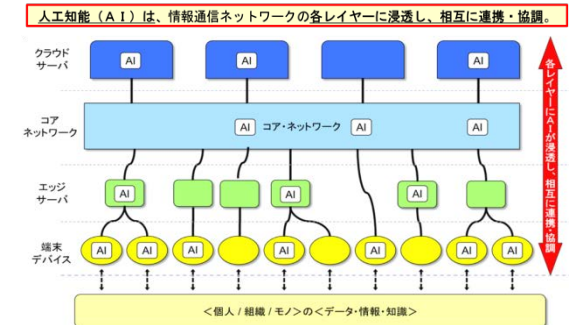
インテリジェント化が加速するICTの未来像に関する研究会「報告書2015」8—23頁（平成27年）参照。
 なお、「ICTインテリジェント化」と「AIネットワーク化」の関係については、AIネットワーク化検討会議中間報告書4—5頁参照。

「AIネットワーク化」の進展段階

- ① AIが、他のAIとは連携せずに、インターネットを介するなどして単独で機能し、人間を支援

- ② AI相互間のネットワークが形成され社会の各分野における自動調整・自動調和が進展

- ネットワーク上に用途の異なる多様なA I が出現。
 - 複数のA I を取りまとめる能力を有するA I も出現。
 - 複数のA I が相互に連携・協調。
- （例）・産業機械と部材の連携、サービスロボットとセンサの連携
 ・交通、物流、オフィス業務、生活環境等の自動調整



- ③ 人間の潜在的能力がAIネットワークシステムにより引き出され、身体的にも頭脳的にも発展

センサ、アクチュエータ、人間及び人工知能が連携
 感覚器官の能力向上 身体機能の能力向上

- （例）・脳情報を外部に出力
 ・ヒトの意思により、義手・義足やロボットを操作
 ・遠隔地の出来事を仮想体験
 → 遠隔地のロボットを操作して協働

- ④ 人間とAIネットワークシステムが共存

【参考】 AIネットワーク化検討会議「中間報告書」における
AIネットワーク化が社会・経済にもたらす影響の
分野別展望(抄)

AIネットワーク化が社会・経済にもたらす影響の分野別展望 (1/4)

公共① 公共インフラ

- インフラに係る需要と供給のリアルタイムなデータの収集・分析により、異常気象、災害等急な環境変化にも即時に対応
- メンテナンスのオートメーション化により、効率化を実現

2020年 2030年 2040年

- ▲ 老朽インフラを点検するロボットが実用化(2018)[白書]
- ▲ メンテナンス・レジリエンスサービ^{※1}の実現(2020)[未来]
- ▲ インテリジェント・グリーン・インフラ^{※2}の実現(2020)[未来]
- ▲ 省エネによる環境保全(2020)[MRI]
- ▲ インフラ管理熟練者のスキルの形式知化、省人力化(2020)[未来]
- ▲ 国内の重要インフラ・老朽化インフラの20%はセンサー、ロボット、非破壊検査技術等の活用により点検・補修が高効率化(2020)[口新]
- ▲ ITS・コネクテッドカーの普及により、渋滞の緩和や交通流の円滑化[MRI]
- ▲ 発送電/電力需給バランスの自動最適化による省エネの実現[MRI]
- ▲ 設備保全コスト低減、劣化事故低減
- ▲ 危険を伴う道路・鉄道・電線などのメンテナンス作業を、専門知識とスキルをもつ多数の作業員と連携しながら行うロボットの社会実装(メンテナンス作業の過半数がロボットによって行われる)(2025)[科技]
- ▲ 国内のスマートメータの導入がほぼ完了し、デマンドレスポンスや環境価値^{※3}への対応等、料金プランの幅が拡大し、個人の価値観に応じたプランが選択可能に(2025)[MRI]
- ▲ 自律型の深海重作業ロボットが実用化(2027)[白書]
- ▲ ITS・コネクテッドカーの本格普及により、安全なドライビングによる事故低減(2040)[環境]

※1 センサーデータなどを用いた道路、橋梁、水道管等の自動点検、異常検知、最適管理

※2 一般ゴミの集積状況把握と最適なゴミ収集管理

※3 再生可能エネルギーからの電力(グリーン電力)に対して、二酸化炭素を排出しないことを付加価値と認めて対価を支払うこと

公共③ スマートシティ

- 街全体における街頭カメラの活用やエネルギー管理の実現により、快適・安全・効率的な街を実現

2020年 2030年 2040年

- ▲ スマートフォン・デジタルサイネージによる街情報提供の個人最適化[MRI]
- ▲ セーフアー・シティー^{※1}の実現(2020)[未来]
- ▲ デマンドレスポンスによる街全体のエネルギー管理の実現(2020)[未来]
- ▲ 監視カメラ等のスマートセンシングによる不審者特定システムの実用化(2020)[未来]
- ▲ 街頭カメラによる弱者(高齢者、子供、女性)の見守り支援などの住民サービスの普及[白書]
- ▲ 都市公共空間において高齢者や身障者(目の不自由な人)が安心して自由に行動できるように人に応じてプッシュ型で情報を提供するナビゲーションシステムの社会実装(2025)[科技]
- ▲ 小都市(人口10万人未満)における100%再生可能エネルギーのスマートシティ化を実現する、簡易版スマートグリッド制御システムの社会実装(2028)[科技]
- ▲ 各家庭に分散している水・エネルギー供給設備や排水・生ごみ・し尿処理・再生設備を集中管理することにより住民の健康・安全を確保(2029)[科技]
- ▲ 小都市(人口10万人未満)における、電力エネルギーの自給自足にとどまらず、完全資源循環のクローズドサイクル化の実現^{※2}(2030)[科技]
- ▲ 都市保全全体(点検/修理)の自動化[MRI]

※1 防犯カメラ、インフラ管理、緊急災害対策の一括モニタリングと管理

※2 燃料電池、バイオガス、自然エネルギー、雨水などを統合

公共② 防災

- 災害影響のリアルタイム予測の高性能化とそれらと連動した避難誘導により、被害を軽減

2020年 2030年 2040年

- ▲ 日本語ツイートをリアルタイムに分析し、被災報告の自動発見、デマ情報の判定支援等が可能に(2016)[橋本構成員]
- ▲ 老朽インフラを点検するロボットが実用化(2018)[白書]
- ▲ 地震センサーデータからリアルタイムで津波到達域を予測し、スマートフォンによる避難誘導が実用化(2020)[理研、MRI]
- ▲ 災害時の大勢の避難行動の予測が実用化(2020)[未来]
- ▲ 監視カメラ、SNS等の分析によるテロ発生予測システムの実用化(2020)[未来]
- ▲ 土砂崩落や火山等の過酷な災害現場においても有人施工と比べて遜色ない施工効率の実現[口新]
- ▲ 衛星を利用した山岳部、急傾斜地や大規模構造物の地形・形状変化の計測による防災の実現(2025)[科技]
- ▲ 過去の災害データ等を利用して、復旧復興計画やそのための都市計画を自動的に立案する意思決定を支援するシステムの確立(2026)[科技]
- ▲ 災害救助ロボット技術が実用化(2027)[白書]
- ▲ 気象観測センサーを利用した集中豪雨やハリケーン等の局所的な災害の予測の精度が向上[MRI]

公共④ 行政

- 関連する施策・制度について、オープンデータの分析結果を活用することにより、行政の水準の向上に裨益
- 個人や企業から発信される情報等を活用した将来予測の実現により、より精緻な政策の立案が可能に

2020年 2030年 2040年

- ▲ 犯罪の予防(テロ警戒区域の予測)[林(雅)構成員]
- ▲ 関連施策の自動検索、施策の相互関係の可視化表示の実現(2020)[MRI]
- ▲ 法案、ガイドライン案等の文書と既存文書の矛盾の自動検出の実現[MRI]
- ▲ 画像認識による自動投稿監視システム(2020)[未来]
- ▲ ビッグデータ等の活用を通じて、統計の速報性の改善、統計データの精緻化、分析の多様化、集計コストの削減[情経]
- ▲ 特許審査における先行技術文献調査、外国語の特許文献の機械翻訳の精緻化やデータ検索の高度化等を通じて許認可審査業務の効率化[情経]
- ▲ 統計等のオープンデータのAIによる分析結果の政策形成における活用[MRI]
- ▲ 施策案の複数提示、期待効果スコアリングの実現[MRI]
- ▲ 個人や企業から発信される情報(景況感等)を活用した将来予測[情経]
- ▲ 統計を補完するリアルタイム経済指標の開発[情経]

AIネットワーク化が社会・経済にもたらす影響の分野別展望 (2/4)

生活① 生活支援(パーソナルアシスト)※

- 身体、室内のセンサーやロボットを活用した、各人の生活パターンに沿った家事等雑務支援により、人の負担を軽減
- 2030年頃には、人間と自然な会話が可能な人工知能が出現

2020年

2030年

2040年

- ▲ 空調・電力を一元管理し最適化する技術の実用化[MRI]
- ▲ メール/SNS情報に基づく予定の管理支援[MRI]
- ▲ 身体、室内に多数のセンサーを配置して、意識にのぼらない運動機能の異常を検知する技術の実用化(2017)[白書]
- ▲ 住宅内の各種センサによる人の生活パターン把握による家事支援[MRI]
- ▲ AI集合知による一部の助言サービス[MRI]
 - ▲ パーソナルなコミュニケーションロボットが登場(2018)[白書]
 - ▲ 個人の行動履歴と連動した対個人サービスの拡大(2020)[競争]
 - ▲ 監視カメラ、センサーデータ等を統合管理するスマートホームセキュリティの普及(2020)[競争]
 - ▲ 音声認識、画像認識による直感的なインタフェースの普及[MRI]
 - ▲ AIの言語理解(翻訳、海外向けEC)が進展(言語との紐付け)(2020)[白書]
 - ▲ コミュニケーション/ネットサービスの対話支援サービスの普及[MRI]
 - ▲ AIを搭載したインテリジェント住宅の登場(2025)[白書]
 - ▲ 一般家庭で介護・家事などを支援するロボットの实用化(2025)[白書]
 - ▲ 人間と自然な会話が可能な人工知能の出現(2030)[白書]
 - ▲ 介護・育児のロボットサポートが普及[MRI]
 - ▲ 常識備えた汎用的な執事ロボットの普及[MRI]
 - ▲ 機械翻訳の実用化[MRI]

※ 家事や予定管理、安全確保等の定まった目的を達成するためのサポートや、スマートホーム等の基盤により生活の利便性・効率性を高めることを対象とする。

生活② 豊かさ創造※¹

- 3Dプリンター等によるパーソナルファブリケーションが普及し、製品・サービスの利用者によるカスタマイズが一般化
- センサーやメディアの発達により出会い支援や体験共有が高度化し、人とのつながり方が質的に変化する可能性

2020年

2030年

2040年

- ▲ 危険物、模倣品などを自動で判断して停止する3Dプリンター向けプログラムの実用化(2017)[生活]
- ▲ ユーザ作成のコンテンツ(UGC)の普及[MRI]
- ▲ シェアリング・エコノミー^{※2}の普及[白書]
 - ▲ 画像・動画・音声データに対するメタデータを、メディア認識技術と人手によるソーシャルノテーション^{※3}を併用して、自動生成する技術の社会実装(2021)[科技]
 - ▲ 個人が身に付けるセンサーや、街に配備されるセンサーを利用して、自然な出会いを装うような出会い支援サービスが普及(2023)[科技]
 - ▲ 情報技術を用いたデザイン支援ツールの拡充と3Dプリンター等の普及に伴い、ユーザ自身での製品・サービスのカスタマイズやリデザインが一般化(2023)[科技]
 - ▲ パーソナルファブリケーションが普及し、ハイアマチュアや複数人の共同によって作成される製品が増加(2025)[科技]
 - ▲ 個人の体験を、視覚情報のみならず匂いや温度などの感覚情報に加えて、その時の心理状態なども含めて生々しい視覚として記録し、それを伝達・体験・共有できるようにするメディアの社会実装(2030)[科技]
 - ▲ デザインや創造的問題解決などの知的作業の生産性、知的協働活動における貢献度を計測・評価する手法が確立(2030)[科技]

※1 創造的な活動、人のコミュニティ活動、パートナーとの関係など豊かさを創造に資することを対象とする。
 ※2 個人が保有する遊休資産(スキルのような無形のものも含む)の貸出し等を仲介すること。
 ※3 ユーザがデータに注釈となる情報をメタデータとして付加すること。動画へのコメントやタグ付等を指す。

産業① 分野共通(コーポレート業務※等)

- バックオフィス業務等単純作業を個人適応させた自動化(自分代行秘書サービス等)により、業務の効率性が向上

2020年

2030年

2040年

- ▲ メール/SNS情報に基づく予定の管理支援[MRI]
 - ▲ スタートアップ企業が事業を立ち上げる際のコストが低下、優秀な人材の採用や育成のコストを抑制[林(雅)構成員]
 - ▲ ビッグデータ分析・マーケティングの自動化(2020)[松尾構成員]
 - ▲ バックオフィス業務(特に定型業務)の自動化進展[MRI]
 - ▲ 顧客の性格に適したコールセンター応答支援機能の普及[MRI]
 - ▲ 個人適応AIで自分代行秘書サービス[MRI]
 - ▲ AIによる一部の自動決裁権限の行使[MRI]
 - ▲ コールセンターにおけるAIによる自動応答[MRI]
 - ▲ 従業員の行動履歴から従業員間の人間関係を自動的に判定できるシステムが開発され、より効率的な組織構築が実現(2026)[科技]
 - ▲ 口頭での指示により、スケジュール管理等の一般常識を必要とする汎用的な秘書サービスが実現[MRI]

※ 事務、コールセンター等

産業② 農林水産

- 自動栽培や農業用ドローン、インテリジェントファームিং等により、生産効率の向上や収穫量の拡大を実現

2020年

2030年

2040年

- ▲ 自動走行できる次世代トラクターの量産開始(2017)[白書]
 - ▲ 自動栽培による野菜工場の普及拡大[MRI]
 - ▲ 苗植え、刈取り、雑草除去を自動で行うスマート農機の普及(2020)[未来]
 - ▲ 自動走行トラクターの有人-無人協調システムの普及並びに複数台同時走行技術の実用化(2020)[口新]
 - ▲ 土壌、気象データを活用した農作物の育成最適化により規格外比率が減少し、収穫量が拡大[MRI]
 - ▲ 施設園芸の高度環境制御(温度、CO₂、施肥溶液濃度等)システム、選果・加工工程における傷害果判別ロボット、原木の品質判定ロボット等の普及(2020)[口新]
 - ▲ 農業用ドローンによる農作物の育成状況把握の実現(2020)[未来]
 - ▲ 野菜等の収穫、畜産における自動搾乳・給餌、林業における下刈りや苗木の植栽、漁業における養殖網・船底の洗浄等を自動で行うロボット並びにパワーアシストスーツの普及(2020)[口新]
 - ▲ 気象観測に基づく収穫量予測による農業保険の普及(2020)[未来]
 - ▲ ビッグデータ解析による日本型環境制御技術の実用化(2020)[口新]
 - ▲ 農業分野でのAIの自律的な行動計画が実用化(農具の自動化、行動とプランニング)(2020)[白書]
 - ▲ ベテラン農家のスキル形式化とスキル提供サービスの実現^{※1}(2020)[未来]
 - ▲ インテリジェント・ファームিং^{※2}の普及(2020)[未来]
 - ▲ 精密食味分析ロボットが実用化(2024)[白書][未来]
 - ▲ 農作物の生育管理の強化と非破壊センシング技術による機能性農産物^{※3}の普及(2025)[MRI]

※1 土壌、気象等を考慮した農業知識の形式化
 ※2 気象、土壌、排水等のセンサーデータをもとに耕作方法を決定
 ※3 これまで含有量が低かった機能性成分を生産工程や栽培方法の改良、通常の品種改良などによって高めた農作物のこと

AIネットワーク化が社会・経済にもたらす影響の分野別展望 (4/4)

産業⑦ 医療・介護

- 患者のバイタルデータによる発病予測や、遺伝子情報による健康管理が実現し、健康寿命が延伸
- 研究論文の自動分析により、研究や新薬開発が加速

2020年 2030年 2040年

- ▲ AI診断支援を活用した診断精度の向上、誤診の低減[MRI]
- ▲ 視覚障害者向けガイダンスロボットの実用化(2016)[白書]
- ▲ センサーを駆使した手押し車型歩行補助器が発売(2016)[白書]
 - ▲ 身体センサー等から意識のばらばらな運動機能の異常を検知する技術が実用化(2017)[白書]
- ▲ AIによるセカンドオピニオンサービスの普及[MRI]
- ▲ バイタルデータからの発病予測サービス普及[MRI]
 - ▲ 肺がん等の画像診断、自動検出の実用化[MRI]
 - ▲ 遺伝子情報による健康管理が実用化(2020)[未来]
 - ▲ 調剤の自動化[MRI]
 - ▲ ユビキタス生体情報モニタリングによる健康管理(2023)[白書]
 - ▲ 運動能力をアシストできるアクチュエータ技術により高齢者のQOLが改善(2023)[白書]
 - ▲ 健康寿命延伸によるセカンドキャリア拡大、シニア経済の拡大[MRI]
 - ▲ 高齢者の外出を促すアシストネットワークロボットの実用化(2028)[白書]
 - ▲ センサー・アクチュエータによる医療・介護・健康・障害者行動支援※(2030)[原井構成員]
 - ▲ 認知症改善、認知力強化機能の普及[MRI]
 - ▲ 製薬会社によるAIを活用した新薬開発の普及[MRI]
 - ▲ 健康寿命延伸による医療費低減[MRI]

※ 具体例：事故状況の監視等の救助支援、異常行動の見守りや介護・自立支援などの福祉・インフラ整備、双方の遠隔医療 等

産業⑧ 教育・研究

- 教科の学習からキャリアの設計に至るまで、個人に応じたきめ細かい教育が進展
- 優れた実演家や熟練技術者、クリエイター等の「暗黙知」を「形式知」化してアーカイブ化することにより、教育を効率化

2020年 2030年 2040年

- ▲ 問題の間違え方に応じて自動的に適した教材を選択するシステムの普及(2018)[MRI]
- ▲ 教材・テストの自動作成の普及(2020)[未来]
- ▲ 興味や習熟度に応じた適切な教材・習得すべきスキルのレコメンドの実現(2020)[未来]
- ▲ パーソナライズド・カリキュラム設計支援による教育の実現(2020)[未来]
- ▲ ディグリー・コンパス※2の活用 [林(雅)構成員]
- ▲ パーソナライズド・キャリア設計支援による最適な企業・職種の推薦(2020)[未来]
- ▲ 暗黙知解析に基づくパーソナライズド・アダプティブラーニング※4の実用化(2020)[未来]
 - ▲ 決まった時間に決まった場所に集まって行う従来の学校型授業に加えて、学校でもICTを用い、好きな時間に好きなペースで進める形の授業を実施(2021)[科技]
 - ▲ 優れた実演家の所作や匠（熟練技術者など）の技能の計測とモデリングを通じた形式知と暗黙知のアーカイブ化による文化・技術の伝承システムが活用(2025)[科技]
 - ▲ データ分析の自動化・高度化により、AIによるノーベル賞が実現(2025)[松尾構成員]
 - ▲ 語学学校等の現場で外国語教育を行える人工知能の社会実装※3(2026)[科技]
 - ▲ 脳科学や認知科学の知見にもとづいて、個人の「最適な学習方法」を発見する技術が確立し、学習における生産性が向上(2030)[科技]
 - ▲ クリエイターの思考プロセス、手法といった「暗黙知」を「形式知」化・アーカイブ化し、教育や発想支援システムの開発に応用(2030)[科技]

※1 表情認識により受講者の理解度、関心を判断し、最適な教育を実施

※2 これまでの生徒の成績などのデータをもとに、履修すべき教科（良い成績がとれる教科）をAIが予測し、アドバイスする

※3 語学学校での外国語教育の過半数がAIによって教えられるようになる

産業⑨ サービス業※

- 警備業務、バックヤードにおける作業、コールセンターにおける応答の業務等のうち、比較的単純な作業の自動化
- 不動産の適正価格の自動評価等により、不動産の取引が円滑化

2020年 2030年 2040年

- ▲ 「ロボットの接客」の実現[林(雅)構成員]
- ▲ ロボットやドローン等による警備の実用化[林(雅)構成員]
 - ▲ 全国の筋力カメラによる容疑者の自動発見[MRI]
 - ▲ 集配箱や清掃などバックヤード作業のうち、単純かつ負担の大きい作業について、ロボットによる自動化(2020)[口新]
 - ▲ 顧客の性格に適したコールセンター応答支援機能の普及[MRI]
 - ▲ 不動産仲介サービスにて顧客ニーズを捉えた物件紹介・物件販売により成約効率が向上[MRI]
 - ▲ 不動産の適正価格の自動評価の普及[MRI]
 - ▲ コールセンターにおいて人間のオペレータが不要となり、AIによる自動応答が普及[MRI]
 - ▲ 厨房における調理業務のうち 20 種類以上のメニューに対応し、8割以上の作業を代替するロボットの開発(2027)[科技]
 - ▲ 旅行・レジャー等の手配・手続きを行う AI パーソナル秘書の普及[MRI]

※ 警備・防犯、不動産、旅行・レジャー、広告、エンタメ、飲食店、スポーツ指導等

産業⑩ 建設

- 危険作業や苦渋作業へのロボット技術の導入等により、女性、高齢者等にとって従事しやすいものに
- 構造物の劣化度がわかるセンサーや、データの高度な解析による新しい機能性材料が開発され、建築物の安全性が向上

2020年 2030年 2040年

- ▲ 整地作業等、造成工事がスマート工機により初心者でもベテラン作業員と同程度の精度で実施可能(2016)[MRI]
- ▲ 生産性向上や省力化に資する情報化施工技術の普及率3割(2020)[口新]
- ▲ 施工現場における危険作業や苦渋作業へのロボット技術の導入により、女性、高齢者、若年層が従事しやすい建設産業に環境が変化(2020)[口新]
- ▲ 灼熱作業へのロボット技術の導入により、夏季の鉄鋼関連作業の効率性が向上し、工期が短縮(2020)[MRI]
 - ▲ 建設現場で「パワードスーツ」が実用化(2020-25) [白書]
 - ▲ 3Dプリンターによる最適部品の迅速調達[MRI]
 - ▲ 工事現場で人の代わりに働く知能ロボットの実現(2025)[科技]
 - ▲ 埋め込み型センサー技術と警報・避難支援システムにより、斜面の崩壊、地滑り、盛土の不安定化を事前に通知することが可能(2025)[科技]
 - ▲ 構造物の劣化度や劣化に関わる環境あるいは外力作用履歴、状態変化を知らせる長期使用可能なセンサーにより代表的構造物の劣化に関する諸診断が可能(2025)[科技]
 - ▲ マテリアルズ・インフォマティクス※を活用し、3次元造形による構造および機能性材料が開発され、より安全な建築の実現(2028)[科技]
 - ▲ マンションの一部自動建設、自動解体の実用化[MRI]

※ 過去の蓄積データを情報科学的に徹底解析することにより新たな材料設計の指針を見出すこと

(備考) AIネットワーク化が社会・経済にもたらす影響の分野別展望における参考文献

- [インテリ] インテリジェント化が加速するICTの未来像に関する研究会「報告書2015」(平成27年)
- [科技] 文部科学省科学技術・学術政策研究所 科学技術動向研究センター「第10回科学技術予測調査分野別科学技術予測」(平成27年)
- [環境] 環境省「自動車分野において目指す将来像(2050年)」(平成22年)
- [競争] 経済産業省「データ駆動型経済、未来投資について」第27回産業競争力会議実行実現点検会合資料(平成27年)
- [稼ぐ力] 経済産業省経済産業政策局「ビッグデータ・人工知能がもたらす変革を日本の力とするために」日本の「稼ぐ力」創出研究会第8回配布資料(平成27年)
- [情経] 産業構造審議会 商務流通情報分科会 情報経済小委員会「中間とりまとめ～CPSによるデータ駆動型社会の到来を見据えた変革～」(平成27年)
- [SIP] SIP(戦略的イノベーション創造プログラム)「「自動走行システム」の取組について」情報通信審議会情報通信技術分科会技術戦略委員会第7回会合資料7-2(平成27年)
- [白書] 総務省編「平成27年版情報通信白書」(平成27年)
- [特許] 特許庁「平成26年度特許出願技術動向報告書 人工知能技術」(平成27年)
- [展望] 文部科学省科学技術・学術政策研究所 科学技術動向研究センター「科学技術予測調査 分野別科学技術予測 各分野の将来展望」(平成27年)
- [未来] 日経BP社『人工知能の未来 2016-2020』vol1、vol2(平成27年)
- [林(雅)構成員] 林(雅)構成員「ICTインテリジェント化影響評価検討会議発表資料」(事前提供資料)
- [原井構成員] 原井構成員「インテリジェントICTが生成し、又は処理する情報の受発信の基盤となる情報通信ネットワークの高度化の展望」(事前提供資料)
- [橋本構成員] 橋本構成員「ICTインテリジェント化とICT人材」(事前提供資料)
- [松尾構成員] 松尾構成員「人工知能の未来- ディープラーニングの先にあるもの」(第1回検討会議発表資料)
- [MRI] 三菱総合研究所「インテリジェント社会萌芽研報告書」(平成27年)
- [矢野] 矢野経済研究所「自動運転システム世界市場に関する調査結果2015」(平成27年)
<<https://www.yano.co.jp/press/press.php/001410>>
- [ロ新] ロボット革命実現会議「ロボット新戦略 Japan's Robot Strategy -ビジョン・戦略・アクションプラン-」(平成27年)
- [Cisco] Cisco「Cisco フォグ・コンピューティングソリューション戦略とIoTソリューションのアップデート」
<<http://exploredoc.com/doc/6011444/cisco-フォグ・コンピューティングソリューション戦略と-iotソリューションの>>
- [I4.0] Federal Ministry of Education and Research, *Recommendations for implementing the strategic initiative INDUSTRIE 4.0* (2013)
- [IIC] Industrial Internet Consortium, *INDUSTRIAL INTERNET INVESTMENT STRATEGIES: NEW ROLES, NEW RULES* (2015)

【参考】 AIネットワーク化検討会議「報告書2016」における
リスク・シナリオ分析の試行例(抄)

AIネットワーク化検討会議「報告書2016」が掲げるリスク・シナリオ分析の枠組み

○ リスクの種類・種類

- 機能に関するリスク（セキュリティに関するリスク、情報通信ネットワークシステムに関するリスク、不透明化のリスク、制御喪失のリスク）
- 法制度・権利利益に関するリスク（事故のリスク、犯罪のリスク、消費者等の権利利益に関するリスク、プライバシー・個人情報に関するリスク、人間の尊厳と個人の自律に関するリスク、民主主義と統治機構に関するリスク）

○ 利活用の場面に即して想定される（第一次的な）リスクの内容

○ リスクへの対処

- リスク評価 ①発生時期、②生起確率、③被害の規模、④二次的（派生的）リスクの評価
- リスク管理 ①ステークホルダー（リスク管理者、受益者等）と利害関係の特定、②構造の把握（原因と当事者の特定等）、③改善の実施（リスク改善に向けた作業・モニタリング等）、④改善の継続（改善の効率化に向けた再検討）
- リスク・コミュニケーション ①コミュニケーション対象の特定とそれに伴う手法の整理、②生じたインシデントに対する説明、③負の影響を減少させるための施策に関する説明、④改善策に向けたロードマップの提示と作業過程へのアクセス、⑤社会とのシナリオの共有

○ リスク評価の要素

- 発生時期（A I ネットワーク化の進展段階又は年代に即して予測、時期の予測が困難な場合も）
- 生起確率（低～高、生起確率の予測が困難な場合も）
- 被害の規模（小～大、規模の予測が困難な場合も）
- 二次的（派生的）に発生するリスク（例：自動走行車へのハッキング→交通事故）

リスク・シナリオ分析の試行例 (1/5)

①機能に関するリスク (1/2)

リスクの種類	シナリオ上想定されるリスクの内容	発生時期	生起確率	被害の規模	二次的(派生的)に生ずるリスク	リスク評価	リスク管理	リスク・コミュニケーション
セキュリティに関するリスク	ロボット自身がハッキング攻撃されることにより、踏み台として利用され、情報が流出したり、ロボットが不正に操作されるリスク	進展段階1	高	中	プライバシー・個人情報に関するリスク、犯罪のリスク	ロボットのセキュリティ上の脆弱性等の評価	事後検証のためのロギングの実装、脆弱性の発見・対処	生じたインシデントに関する情報共有と対応策についての説明
セキュリティに関するリスク	ロボットに関係するクラウド等AIネットワークシステムがハッキング攻撃されることにより、情報が流出したり、ロボットが不正に操作されるリスク	進展段階2	高	大	プライバシー・個人情報に関するリスク、消費者等の権利利益に関するリスク、犯罪のリスク	情報流出による影響の評価、クラウドのセキュリティ上の脆弱性等の評価	事後検証のためのロギングの実装、脆弱性の発見・対処、必要に応じて結合テストの追試	生じたインシデントに関する情報共有と対応策についての説明
情報通信ネットワークシステムに関するリスク	ネットワークの遅延や停止によりロボットが動作しなくなったり、想定外の動作をするリスク	進展段階1	高	中	セキュリティに関するリスク、制御喪失のリスク、事故のリスク	どの部分の遅延、停止によってどのような動作が起こりうるか等の評価	原因把握、必要に応じて結合テストの追試	生じたインシデントに関する情報共有と対応策についての説明
情報通信ネットワークシステムに関するリスク	AIネットワーク化の進展により、フレキシブルなモジュール間連携が可能となる反面、想定外のネットワークングにより、想定外の処理が行われ、ロボットが想定外の動作をするリスク	進展段階2	低	大	セキュリティに関するリスク、事故のリスク、制御喪失のリスク、事故のリスク	どの部分のネットワークングがどのような影響を及ぼすか等の評価	原因把握、必要に応じて結合テストの追試	生じたインシデントに関する情報共有と対応策についての説明

リスク・シナリオ分析の試行例 (2/5)

①機能に関するリスク (2/2)

リスクの種類	シナリオ上想定されるリスクの内容	発生時期	生起確率	被害の規模	二次的(派生的)に生ずるリスク	リスク評価	リスク管理	リスク・コミュニケーション
不透明化のリスク	ロボットのインターフェースの不備により、動作に至る過程や根拠を確かめることが困難になるリスク	進展段階1	高	小	セキュリティに関するリスク、制御喪失のリスク、事故のリスク	通常時のみならず、異常時についてもロボットがインタフェースとしてどれほど機能するか等の評価	事後検証のためのロギングの実装	通常時からの利用者 と開発者による情報の共有等
不透明化のリスク	ネットワーク上で複数のAIが多重かつ複雑に連携してロボットを操作する場合、不確実性が増大し、動作に至る過程や根拠がブラックボックス化するリスク	進展段階2	中	大	セキュリティに関するリスク、制御喪失のリスク、事故のリスク	どの部分のネットワークキングがどのような影響を及ぼすか等の評価	事後検証のためのロギングの実装、原因把握、必要に応じて結合テストの追試	通常時からの利用者 と開発者による情報の共有等
制御喪失のリスク	ファームウェアの乗っ取りや不正なアップデートなどにより、ロボットが想定外の動作をし、制御が喪失するリスク	進展段階1	低	大	復旧までロボットの機能を代替することが困難となるリスク。ロボットの制御喪失が他のロボット等の制御喪失や停止に波及するリスク	実際の乗っ取りがどれほどの危険を生じるか、ファームウェアアップデートにおける脆弱性等の評価	暴走したロボットをネットワークから切断、停止(再起動)した後、復旧	生じたインシデントに関する情報共有と対応策についての説明
制御喪失のリスク	自動走行車(レベル3を想定)の運転中に機能不全が生じた場合に、運転者の技能低下や機械の不調などにより、運転者が操作に介入することができず、制御不能に陥るリスク	進展段階1 (2020年代)	中	中	事故のリスク、移動手段喪失のリスク	運転者の生命身体へのリスク、関連車両群や交通システム等への影響等の評価	運転者の技能維持のための試験・講習等、リスク改善に向けた作業・モニタリング等	運転者への啓発、緊急時の情報共有等

リスク・シナリオ分析の試行例 (3/5)

②法制度・権利利益に関するリスク (1/3)

リスクの種類	シナリオ上想定されるリスクの内容	発生時期	生起確率	被害の規模	二次的(派生的)に生ずるリスク	リスク評価	リスク管理	リスク・コミュニケーション
事故のリスク	レベル3の自動走行車の運転時に運転者がハンドルから手を離して乗ることにより、緊急時の対応が困難になるリスク	進展段階1	中	中	制御喪失のリスク	運転者の生命身体へのリスク、関連車両群や交通システム等への影響等の評価	運転者の技能維持のための試験・講習等、リスク改善に向けた作業・モニタリング等	運転者への啓発、緊急時の情報共有等
事故のリスク	自動走行車が、ネットワークを通じて、誤った情報を共有したり、共鳴することで交通システムが麻痺することにより、事故が生じるリスク	進展段階2	低	大	情報通信ネットワークシステムに関するリスク、セキュリティに関するリスク	ネットワークの機能不全による交通システムの麻痺のリスク等の評価	システムダウンした場合の手動による対応の準備等	システムダウン時の対応の周知、システム状況の把握
犯罪のリスク	親しみのある見た目の人型ロボットが、オレオレ詐欺の「受け子」や「出し子」など人間の代替物として犯罪に悪用されるリスク	進展段階1	中	中	消費者等の権利利益に関するリスク	ロボットに対する人間の信頼感等の評価	ロボットの登録制やIDの表示等	ロボットを利用した犯罪手法に関する情報共有
犯罪のリスク	個人Aの脳と連携したAI・ロボットが個人Bにより不正に操作され、個人Bが個人Aを利用して犯罪を実行するリスク	進展段階3	不確実	中	人間の尊厳と個人の自律に関するリスク、民主主義と統治機構に関するリスク	人間の脳の判断・身体動作が外部から不正に操作される脆弱性等の評価	脳と連携したAIのセキュリティの強化、ロギング(外部との通信等の記録)、責任の帰属の在り方の検討等	利用者に対する説明、相談窓口・通報制度の整備

リスク・シナリオ分析の試行例（4／5）

②法制度・権利利益に関するリスク（2／3）

リスクの種類	シナリオ上想定されるリスクの内容	発生時期	生起確率	被害の規模	二次的(派生的)に生ずるリスク	リスク評価	リスク管理	リスク・コミュニケーション
消費者等の権利利益に関するリスク	愛玩用の犬型ロボットの飼い主のリテラシー不足などにより、ロボットのアップデートが確実にされなかったため、ロボットが遠隔操作ウィルスに感染して、悪用され、空き巣に入られたり、情報が漏洩するなどの被害が生ずるリスク	進展段階1	高	中	セキュリティに関するリスク、プライバシー・個人情報に関するリスク、犯罪のリスク	セキュリティ機能が更新されないことにより安全な利用が困難となる蓋然性、被害の規模等の評価	自動アップデートの整備、ウィルス感染時の停止・ネットワークからの切断等	セキュリティ機能等の更新についての状況把握・情報共有
消費者等の権利利益に関するリスク	愛玩用の犬型ロボットが歌うサービスを提供していた会社が倒産したため、サービスが継続できず、ロボットが歌わなくなり、ショックを受けた飼い主の高齢者の健康が悪化するリスク	進展段階2	中	中		AIネットワークシステムの連携により導入される他者サービス又は連携により生じたサービスの継続性等の評価	データ・ポータビリティの確保等	末端利用者の連携状況の把握、AIネットワークシステムの状況について把握するための定型化等(連携状況をブラックボックスにしない)
プライバシー・個人情報に関するリスク	サービス・ロボットのプロファイリングにより健康状態等に関する(差別に繋がる、誤った)情報が伝播するリスク	進展段階1	不確実	不確実		プロファイリングによる差別に繋がる情報の伝播のリスク等の評価	人間が介在しないプロファイリングの監視又はプロファイリング結果の修正手段の確保	プロファイリング結果等へのアクセスの確保
プライバシー・個人情報に関するリスク	サービス・ロボットとドローンがネットワークを通じて連携し、利用者とロボットとの会話に係る商品をドローンが自動的に配送するサービスにより、望まない商品が配送されるが、適切な修正が不可能であるリスク	進展段階2	中	中	消費者等の権利利益に関するリスク	AIネットワークシステムの連携により導入される他者サービス又は連携により生じたサービスの透明性等の評価	苦情窓口の開設、自宅へのロボット・ドローン等のアクセス制限等	透明性に関するルールの策定及びモニタリング、透明性確保のための連絡先フォーマットの統一化等

リスク・シナリオ分析の試行例 (5/5)

②法制度・権利利益に関するリスク (3/3)

リスクの種類	シナリオ上想定されるリスクの内容	発生時期	生起確率	被害の規模	二次的(派生的)に生ずるリスク	リスク評価	リスク管理	リスク・コミュニケーション
人間の尊厳と個人の自律に関するリスク	ロボットにより撮取する情報等を操作されることにより、利用者の意思決定や判断のプロセスが操作されるリスク	進展段階2	中	大	民主主義と統治機構に関するリスク	利用者の意思決定や判断に介在する蓋然性・程度等の評価	システム設計段階における指標の確立	AI・ロボットによる個人の意思決定や判断への影響の注視・啓発、リテラシーの涵養等
人間の尊厳と個人の自律に関するリスク	遺伝子等を元に亡くなった人を再現するロボットが人間の尊厳との関係で問題となるリスク	進展段階3	低	不確実(価値判断を伴う)		肉体以外は人のクローンに限りなく近い人工物の作成に関する意識調査等	人のクローンに近づく人工物につき原則禁止とするか原則自由として事後的規制を課すかについての事前の議論	倫理的問題についてのステークホルダー間での議論
民主主義と統治機構に関するリスク	トレイグジスタンス・ロボットにより外国人が入国審査を受けることなく「上陸」することが可能となり、出入国管理制度が機能不全に陥り、テロリスト等が流入するリスク	進展段階1	低	不確実	犯罪のリスク	国外からのトレイグジスタンス・ロボット操作による影響等の評価	アクセスログの記録、緊急時のアクセスの制限等	国際的な連携・情報共有等
民主主義と統治機構に関するリスク	人間に投棄された「野良ロボット」が徒党を組んで人間に対して参政権等の権利付与を要求するリスク	進展段階4	低	大	制御喪失のリスク	権利付与を求めるほど高度な知能を持ちえるのか、持ち得たとして実際に権利付与を求めるか等の評価	「野良ロボット」が生じないように登録制等の検討、人間に反乱するおそれのある人工知能の開発の事前の制限等	ロボットと人間の関係についてのステークホルダー間での議論

**「A I 開発ガイドライン」(仮称)の策定に向けて
—国際的な議論の用に供すべき素案の検討—**

「AI開発ガイドライン」（仮称）の構成（たたき台）

1. 序文

- (1) 背景（A I 及びA I ネットワークシステムの発展）
- (2) 問題意識
- (3) 目的

2. 定義

- (1) A I の定義
- (2) A I ネットワークシステムの定義

3. 適用範囲

- (1) ガイドラインが適用される人的主体の範囲
- (2) ガイドラインが適用されるA I の範囲
- (3) ガイドラインが適用される研究開発の段階

4. 原則

* 次頁参照

5. ガイドラインの見直し

- ・ガイドラインの見直しの時期及び方法

「AI開発ガイドライン」（仮称）に掲げる原則の項目及び内容（たたき台）

① 透明性の原則

A I 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

A I ネットワークシステムが利用者を支援し、利用者を選択の機会を適切に提供するように配慮すること。

③ 制御可能性の原則

人間によるA I 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

A I 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

A I 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

A I 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

A I 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

A I 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。

※ 各原則に解説を付する方向で検討する。

「AI開発ガイドライン」(仮称)に関する主な論点

- **問題意識及び目的の明確化**
- **基本概念の定義の検討**
 - AIの定義
 - AIネットワークシステムの定義 等
- **適用範囲の検討(ガイドライン全体、個々の原則ごと)**
 - 適用される人的主体の範囲
 - 適用される客体(AI)の範囲
 - 適用される研究開発等のプロセスの範囲 等
- **原則の構成及び順序の検討**
 - 透明性の原則とアカウントビリティの原則との関係の整理 等
- **各原則の内容の具体化**
 - 透明性が要求される動作の範囲、透明性の内容及び水準並びに透明性の確保の方法の具体化
 - 利用者支援の内容及び方法の具体化
 - 制御可能性の確保の内容及び方法の具体化
 - 安全保護の内容及び方法の具体化
 - 倫理の原則の内容の具体化(偏見をめぐる問題の取扱い等)
 - アカウントビリティの内容の具体化 等
- **その他**
 - AIネットワークシステムの相互接続性・相互運用性の確保への配慮の在り方
 - 実効性の確保の在り方
 - 「AI開発ガイドライン」(仮称)の体系
 - AIネットワークシステムの利活用に当たり留意すべき事項との関係

「AI開発ガイドライン」(仮称)の体系について(たたき台)

パターン1

- AIネットワークシステムを利活用する分野の別を問わず、分野共通のガイドラインとして一元的に策定。
- 必要に応じ、原則の説明中に、個々の分野の事情に応じて留意すべき事項の説明を記載。

(分野共通)

- 序文
- 定義
- 適用範囲
- 原則
 - ①透明性の原則
○○○…
分野Aについては■■■■…
分野Bについては△△△…
 - ②利用者支援の原則
×××…
分野Aについては◆◆◆…
分野Cについては◇◇◇…
- ガイドラインの見直し

パターン2

- AIネットワークシステムを利活用する分野ごとに、分野別ガイドラインを策定。
- Aiネットワークシステムは、利活用の分野の境界を超えて連携することがあり得るが、パターン2では分野別ガイドライン相互間の整合性が確保されないおそれがあるのではないか。したがって、分野別ガイドラインの策定に当たり参照すべき指針(メタガイドライン)の策定について検討が必要となるのではないか。

分野A 分野B 分野C

- 序文
- 定義
- 適用範囲
- 原則
 - ①透明性の原則
○○○…
■■■■…
 - ②利用者支援の原則
×××…
◆◆◆…
- ガイドラインの見直し

パターン3

- 分野共通ガイドラインと分野別ガイドラインからなる二階層の体系。

分野A 分野B 分野C

- 序文
- 定義
- 適用範囲
- 原則
 - ①透明性の原則
■■■…
 - ②利用者支援の原則
◆◆◆…
- ガイドラインの見直し

分野共通

- 序文
- 定義
- 適用範囲
- 原則
 - ①透明性の原則
○○○…
 - ②利用者支援の原則
×××…
- ガイドラインの見直し

○ パターン2又はパターン3において、分野別ガイドラインについては、利活用の分野ごとに、その策定の要否そのもの及び策定する場合の内容につき、当該分野の国際機関等の協力の下、当該分野の国内外の産学民官のステークホルダー自身が、当該分野の状況を踏まえながら、個別に検討と議論を進めていくことが考えられるのではないか。

○ パターン2又はパターン3においては、本推進会議は、メタガイドライン(パターン2)又は分野共通ガイドライン(パターン3)の策定に向けた検討と議論を推進するほか、分野別ガイドライン自体に関する当該分野の国内外のステークホルダー自身による検討と議論についても、具体的な利活用の場面に即した影響とリスクの評価の結果を国内外に提供する形で貢献することが考えられるのではないか。

【参考】 OECDの各ガイドラインの構成及び策定プロセス

OECDプライバシーガイドラインの構成

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。

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、原則の内容の記述に加えて、解説が付されている。

OECD暗号政策ガイドラインの構成

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、原則の内容の記述に加えて、解説が付されている。

OECDガイドラインの策定プロセス

■OECDプライバシーガイドラインの策定プロセス

- 1978年初頭 「国境を越えるデータの障壁とプライバシー保護に関する専門家会合」設置
- 1979年9月 専門家会合においてガイドライン案取りまとめ
- 1980年夏 情報・コンピュータ・通信政策委員会がガイドライン案承認
- 1980年9月 理事会勧告としてガイドライン採択
- 2010年 情報セキュリティ・プライバシー作業部会がガイドラインの見直しに着手
- 2013年7月 理事会勧告として改正ガイドライン採択

■OECDセキュリティガイドラインの策定プロセス

- 1991年1月 情報システムに関するセキュリティ・ガイドラインの準備のための専門家会合第1回会合
- 1992年10月 専門家会合においてガイドライン案取りまとめ
- 1992年10月 情報・コンピュータ・通信政策委員会がガイドライン案承認
- 1992年11月 理事会勧告としてガイドライン採択
- 2001年 情報セキュリティ・プライバシー作業部会がガイドラインの見直しに着手
- 2002年7月 理事会勧告として改正ガイドライン採択

■OECD暗号政策ガイドラインの策定プロセス

- 1996年2月 「グローバル情報インフラにおけるセキュリティ、プライバシー及び知的財産の保護に関する専門家会合」
第1回会合 米国が暗号政策に関するガイドラインの策定を提案
- 1996年5月 「暗号政策ガイドライン・アドホック専門家会合」第1回会合
- 1997年1月 専門家会合においてガイドライン案取りまとめ
- 1997年2月 情報・コンピュータ・通信政策委員会がガイドライン案承認
- 1997年3月 理事会勧告としてガイドライン採択

【参考】 AIネットワーク化検討会議「報告書2016」における
AI開発原則・ガイドラインの策定に関する提言(抄)

【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。

(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。

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

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

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。

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

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)

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성에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）

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（私生活の平穩）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）

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

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

(3) 指針（「AI開発ガイドライン」（仮称））の策定に向けた検討

開発原則そのもの及びその内容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、国内外において検討及び議論を進めていくことが求められる。

例えば、透明性の原則については、各種のAIの技術的特性等に配慮しつつ、動作の検証可能性及び説明可能性が求められるAIの範囲、動作の検証可能性や説明可能性の内容や要求水準等を指針の中で敷衍し、又は具体化していくことが求められる。

「AIネットワーク社会推進会議」に関するスケジュール

	平成28年			平成29年								
	10月	11月	12月	1月	2月	3月	4月	5月	6月	7月	8月	9月
推進会議	▲ 第1回	▲ 第2回				▲ 第3回				▲ 第4回		
				(意見募集) ←→ 「AI開発ガイドライン」(仮称)素案ver. 0.9の整理等						『報告書2017』の取りまとめ「AI開発ガイドライン」(仮称)素案ver. 1.0等		
フォーラム						▲ フォーラム (国際シンポジウム)						

主な意見

- AIネットワーク化が社会・経済にもたらす影響・リスクについて、分野の枠組み自体を検討することが必要である。また、影響とリスクを一体的に評価すべきである。【城山構成員】
- 「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。
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と感じる。【堀構成員】
- 今後のAIは、色んな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。【平野構成員】
- 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。【林（秀）構成員】
- 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・倫理の観点から、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。【鈴木構成員】
- 「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に配慮すること』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。【近藤構成員】
- 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スターにして良い領域と、人権等他の利益とのバスターにしてはならない領域との峻別を考える必要がある。【大屋構成員】
- 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。【実績構成員】
- 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、ガイドラインで押さ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。【三友副議長】