

---

## 米国におけるAIの経済・社会への影響にかかる検討 —IEEEと大統領府の二つの報告書にみるリスク対応の方向性—

---

AIネットワーク社会推進会議 影響評価分科会(第2回)

2017年1月23日(月) 16:00~18:00

総務省 情報通信政策研究所

(一財)マルチメディア振興センター

情報通信研究部 田中絵麻

# 内容

---

- はじめに
- 1 IEEE-SA/ IC16-002「倫理的設計バージョン1」(2016年12月)
  - 1-1 IEEE-SA/ IC16-002の概要
  - 1-2 「倫理的設計バージョン1」構成と概要
  - 1-3 「倫理的設計バージョン1」に関連するIEEE標準化活動
- 2 大統領府「人工知能・自動化と経済」の概要(2016年12月)
- おわりに

はじめに【2016年10月31日資料より再掲】

## 欧米におけるAIとロボットにかかる2015年から2016年の動き

### 2015年～2016年の欧米の動き

- 2015年 AIとロボットにかかる社会的影響の検討の本格化
- 2016年 関連会議、報告書、提言、ガイドラインの公表相次ぐ
  
- 技術開発、ビジョン、政策、市場、社会の分野横断的な相互参照
- 米国、欧州、アジアの地域横断的な相互参照
- 企業、政府機関、非営利団体、国際機関等の組織横断的な相互参照

→基礎・応用研究に加えて**社会的実装への展開フェーズへの移行**

AIの便益の最大化 → リスクの縮減に向けた社会的制度整備  
→ 短期、中期、長期的な視点  
→ サービス開発・展開度合いに応じた分野別対応  
→ 関連方策に関する対話の深化

欧米における政府、学术界等からの提言やガイドラインの公表  
AIとロボットの社会的活用に向けた共進化プラットフォームの胎動→**具体的検討進展中**

## 1-1 IEEE-SA/IC16-002の概要

# IEEE (The Institute of Electrical and Electronics Engineers) の概要

- 概要: アイ・トリプル・イー。米国に本拠を置く電気工学・電子工学技術の学会。
- 沿革: 発祥は1884年。IEEEは1963年発足(当時会員数15万(米国内14万))。
- ミッション: advancing technology for the benefit of humanity
- 会員数: 現在、160か国以上、42万会員。50%以上が米国外の会員。
- 組織: 39つのソサエティと7つの技術カウンシル。幅広い技術を対象に活動。
- 標準化活動: 1,100以上のアクティブな標準を開発、500以上の標準を開発中。
- 情報提供: IEEE Xplore Digital Libraryで約413万文書。月間800万DL以上。  
標準規格の分類は7,786件ヒット(旧標準・策定中含む)。

### ※39のソサエティ

- 各ソサエティは、各種IEEEカウンシルのスポンサーとなっている。
- 技術カウンシルは複数のソサエティで構成され、広域にかかる技術について協力。

### ※標準化活動

IEEE-SA: 広範な領域のIEEEの標準化活動を実施。

IEEE-ITSO: 1999年1月設立。IETF、ISO、IEC、IEEE、ISO/IEC、JTC1、W3Cなどの機関に標準を提案。



## IEEE-SA 産業連携(IC)プログラムの概要

---

- IEEE Standard Association (IEEE-SA)
  - IEEE標準の策定・維持を実施
  - 技術開発プログラムとして以下を設置
    - **産業連携(Industry Connections: IC)**
    - 閃光アーク研究プロジェクト(Arc Flash Research Project)
    - グローバルな参画(Global Engagement)
    - 企業プログラム(Corporate Program)
  
- IEEE-SA産業連携プログラム(Industry Connections Program)
  - 迅速(fast-track)に、コンテンツや提出物を生み出すための合意形成や成果の共有のための効率的・経済的な環境を提供
  - コンテンツや成果物としては以下を含む:
    - ・標準化提案
    - ・ホワイトペーパー
    - ・ピアレビューを受けたガイドラインや方針説明書(position papers)
    - ・会議、ワークショップ等のイベント
    - ・データベースと登録サービス
    - ・ソフトウェア、ツール、ウェブ・サービス
    - ・その他共同開発成果

# IEEE-SA 産業連携(IC)プログラムのガバナンス

---

### ◆ 設置と活動概要

- **活動開始:** 複数の企業や個人からなるグループが、ある技術分野における協力や合意形成の必要性を認識しつつも、どのような協力体制とするかについて理解が途上にあるときに、企業連携活動が開始される。
- **設置手続き:** 関係者で構成されるグループは、産業連携活動開始文書(ICAID)を、IEEE-SA 産業連携委員会(ICCom)に提出。ICComとIEEE-SA標準化委員会(Standards Board)がICAIDを承認した後、参加者による対面・電子的な会合開始。
- **活動支援:** 活動初期段階では、IEEE-SAがリーダーシップ、方針、手続きの確立を支援するほか、成果物策定の過程においても支援を継続。
- **参加資格:** 法人(企業、政府機関、学術機関、事業者団体等)、個人(専門家として参加)のいずれも活動参加資格(投票権あり)を有する。法人は、活動開始から1年後からはIEEE-SAの法人会員となることが求められる。

### ◆ IC活動の展開

- IC活動はある技術分野における初期の協力的な活動から、より恒常的な活動に移行するか、解散する。以下のような活動の移行経路がある。
  - 活動完了後に解散
  - IEEE-SAの新たな常任委員会か標準化スポンサーとなる。
  - 既存のIEEEソサエティ/IEEEカOUNシルに統合される。
  - 新たなIEEEソサエティ/IEEEカOUNシルとなる。
  - 新たなIEEE-ISTO産業プログラムとなる。

## 1-1 IEEE-SA/IC16-002の概要

# 活動中のIEEE-SA 産業連携(IC)プログラム

- 2017年1月現在 18のプログラムが活動中
- IC09-001 Computer Security Group (ICSG)
- IC12-006 IEEE Actionable Data Book for Education
- IC13-001 IEEE-SA Symposium on EDA Interoperability
- IC13-002 Electric Vehicle Wireless Power Transfer
- IC13-004 IEEE-SA Ethernet & IP @ Automotive Technology Day
- IC13-005 DC in the Home
- IC15-001 Fiber Optics Sensors
- IC15-002 Smart Glasses Roadmap
- IC15-003 Smart City Compliance Indicators
- IC15-004 3D Body Processing
- IC15-005-02 New Ethernet Applications
- IC16-001 Open Data
- **IC16-002 The Global Initiative for Ethical Considerations in the Design of Autonomous Systems**
- IC16-003 Internet of Things Interest Group
- IC16-004 Augmented Reality in the Oil/Gas/Electric Industry
- IC16-005-002 Consumer Healthcare Alliance
- IC16-006-02 International Roadmap for Devices and Systems (IRDS)

活動開始年2015年・2016年増加  
スマート・グラス、電気自動車関連、家庭用、  
光ファイバセンサー、スマートシティ、3D、  
オープンデータ、IoT、拡張現実、医療系  
→端末インターフェース部分の活動活発

## IEEE-SA/IC16-002 自動システム設計における倫理的配慮のためのグローバル・イニシアティブ

---

- IEEE-SA/IC16-002 The Global Initiative for Ethical Considerations in the Design of Autonomous Systems
  - 個人の参加を呼びかけて、2016年4月に設置
  - マシンがますます多面的かつセンシティブに人間を支援していることから、新たなシステムの設計にかかる倫理規定の利用の優先順位をつけることが重要。
  - 同イニシアティブにより、規律ある原則のもとで人類のために技術を向上させるためにIEEEが活動するように支援する。
  - 議論の活性化、対話により、全ての倫理的なイシューを検討する活動に焦点
  - 自動システム(ロボティクス、AI、コンピューショナル・インテリジェンス、機械学習、深層学習、認知コンピューティング、アフェクティブ・コンピューティングと、一般的なアルゴリズム・ベースでのプログラムを含む)と倫理に関連する専門家(工学、科学、経済学、倫理、哲学、政治学、医療分野を含む)が活動。
  - 初の対面会合は、AIにかかる欧州会議(ECAI)で2016年8月(ハーグ)に開催。
- IC16-002の目的: すべての技術者が教育、訓練を受け、自動知能システムの設計における倫理的配慮の優先付ができるようにする。



## 1-1 IEEE-SA/IC16-002の概要

# IEEE-SA IC16-002の設置文書と設置委員会

- 設置文書(ICAID): The Charter (Draft v1, Updated 4.23.2016)より
  - イシューごとの委員会(Committees)を設置。
  - 各委員会の参加者数は10–15名(合計100名以上が参画) ※[]内は主な所属
1. 一般原則とガイダンス (共同委員長: Kay Firth-Butterfield[テキサス大学法学部], Alan Winfield[ザ・ウェスト・オブ・イングランド大学(英)])
  2. 汎用人工知能(AGI)と人工超知能(ASI)の安全性と恩恵 (共同委員長: Richard Mallah[ケンブリッジ・セマンティクス、Future of Life Institute], Malo Bourgon[COO、Machine Intelligence Research Institute(米)])
  3. 個人データ管理とプライバシー (共同委員長: Eva Schulz-Kamm[NXP, 政府連携部門(蘭)], Michelle Denedy[Cisco, 副社長, チーフ・プライバシー・オフィサー])
  4. マシン自動化／人道的イシューの経済学  
(委員長: Raj Madhavan[Humanitarian Robotics Technologies設立者(米)])
  5. 倫理的研究、設計、製造のガイドのための方法論 (共同委員長: Raja Chatila[Institut des Systèmes Intelligents et de Robotique(仏)], John C. Havens[IEEE-SA IC16-002 エグゼクティブ・ディレクター])
  6. AIへ倫理を組み込む方法 (委員長: AJung Moon[Open Roboethics Initiative共同設立者、ブリティッシュコロロンビア大学 博士候補生])
  7. 殺傷自動武器システム(LAWS)の再構成 (委員長: Richard Mallah[2の共同委員長])
  8. 法 (共同委員長: Kay Firth-Butterfield[1の共同委員長], Derek Jinks[テキサス大学法学部])

※その他、ロジスティクス委員会が、エコシステム・マッピング (Chair - Stephanie Wander)イニシアティブに参加を呼びかける個人や組織を特定することに焦点を当てて活動。可能な限り、包括的・関係性のあるようにする活動を実施。

※会合開催のフォーマットが提供されており、議論の流れ(懸念と提言を共有)、取りまとめ文書の作成方法の情報を提供。

## 1-2 「倫理的設計バージョン1」構成と概要

# IEEE-SA IC16-002作成「倫理的設計(EAD)バージョン1」

---

- 2016年12月16日公表
- 「倫理的設計(EAD)バージョン1 –AIと自動システムと人類の福利の優先付のためのビジョン」(Ethically Aligned Design—A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems, Version 1 for Public Discussion)
  - 一般からの意見を募集(eメールで2017年3月6日まで受付)
  - 意見を反映したバージョンは2017年秋に発表予定。
  - ポジションや政策表明、公式な報告書ではなく、作成中の参照文書の位置づけ。
- 目標1: 今後、参加者からの洞察や提言を提供し、AI/AS技術者の活動のための主要な参照文書を提供すること。
- 目標2: EADに基づいたIEEE標準のために提言を行うこと。
  - IEEE P7000– Model Process for Addressing Ethical Concerns During System Design(最初の標準プロジェクト。承認済み・活動中)
  - IEEE P7001 – Transparency of Autonomous Systems(承認済み)
  - IEEE P7002 – Data Privacy Process (承認済み)→AI/AS倫理面にかかる 이슈に与える実践的な影響を示すものとしている。

## 1-2 「倫理的設計バージョン1」構成と概要

# 「倫理的設計(EAD)バージョン1」の構成

---

- エグゼクティブ・サマリー  
概要、構成(8領域における検討領域、背景、提言候補の概況)、  
文書作成の経緯、引用方法、謝辞
  
- 委員会セクション(Committee Sections)
  - 1 一般原則
  - 2 自律知能システム(AIS)への価値の埋め込み
  - 3 倫理的な研究と設計をガイドするための方法論
  - 4 汎用人工知能(AGI)と人工超知能(ASI)の安全性と便益
  - 5 個人情報と個別のアクセス管理
  - 6 自動兵器システムの再構成
  - 7 経済的・人道的イシュー
  - 8 法律
  
- 新設委員会の概要
- 脚注
- 執行委員会情報

## 1-2 「倫理的設計バージョン1」構成と概要

# 「倫理的設計(EAD)バージョン1」の謝辞

---

### ○ 謝辞

#### 情報提供

- AI Now (White House/New York University)
- One Hundred Year Study on Artificial Intelligence (Stanford University)
- Preparing for The Future of Artificial Intelligence (U.S. White House/NSTC)
- The National Artificial Intelligence Research and Development Strategic Plan (U.S. White House/NSTC)
- Robotics and Artificial Intelligence (U.K. House of Commons Science and Technology Committee)
- Robots and Robotic Devices – Guide to the Ethical Design and Application of Robots and Robotic Systems (British Standards Institute)
- **Japan's Basic Rules for AI Research** (総務省G7提案 AI開発ガイドライン、Japan Timesの記事にリンク)
- Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics (European Parliament)
- Éthique de la recherche en robotique (CERNA)
- Charta der Digitalen Grundrechter der Europäischen Union (Charter of the Digital Fundamental Rights of the EU)
- Research Priorities for Robust and Beneficial Artificial Intelligence (Future of Life Institute).

#### 組織の活動

- The Association for the Advancement of Artificial Intelligence and their formative work on AI Ethics
- European Association for Artificial Intelligence; ACM Special Interest Group on Artificial Intelligence
- The IEEE Robot and Automation Society Committee on Robot Ethics
- The IEEE Society on Social Implications of Technology
- The Leverhulme Centre for the Future of Intelligence
- Allen Institute for Artificial Intelligence
- OpenAI
- Machine Intelligence Research Institute
- Centre for The Study of Existential Risk
- AI-Austin and, Partnership on AI to Benefit People and Society.

## 1-2 「倫理的設計バージョン1」構成と概要

# 1 一般原則：イシューと提言候補①

---

- 報告書の構成： 8セクションで構成。各セクション(前述の委員会に対応)でAI/ASにかかると特定のトピックスを取り上げ、イシューに対応した提言候補を提示。

### 一般原則

- 1. 人権の最も高い理想を具現化する。
- 2. 人類と自然環境への最大の利益に優先順位を置く。
- 3. AI / ASが社会技術システムとして進化する際のリスクと負の影響を緩和する。

【一般原則にかかるイシューと提言候補】 ※各イシューと提言の他、背景情報や関連情報も掲載。

### 原則1 - 人間の福利

- イシュー:どのようにAI / ASが人権を侵害していないことを保証できるか？
- 提言候補
  - 1. AI / ASの公共の信頼の構築に貢献することを保証し、事故調査のプロセスを監督するために、基準や規制機関等を含めたガバナンスのフレームワークが設立されるべきである。
  - 2. また、法的義務を、今後の／既存の周知された政策や技術的な検討事項に翻訳していくための方法論が必要である。

## 1-2 「倫理的設計バージョン1」構成と概要

# 1 一般原則: イシューと提言候補②

### 原則2 – 責任

- イシュー: どのように私たちはAI / ASが責任を負っていることを保証することができるか？
- 提言候補
  - 責任の問題について最もよく対応するため:
    - 1. 議会／裁判所は、(製造業者やユーザーが、権利と義務のあり方を理解する負担を軽減するため)自律知能システムに対する責任、過失、責任、説明責任の問題を明確にしなければならない。
    - 2. 自律知能システム(AIS)の設計者と開発者は、AI / ASのユーザー・グループ間で、既存の文化的規範が多様であることを自覚し、考慮しなければならない。
    - 3. AI / AS関連技術とその影響があまりにも新しいことから、(市民社会の代表、法執行機関、保険会社、メーカー、エンジニア、弁護士等を含む)マルチステークホルダー・エコシステムが存在しない場合には、規範形成を支援するための同システムを構築するべきである。
    - 4. 自律システムの生産者/ユーザーが、以下を含む(鍵となるハイレベルのパラメータを組み込んだ)登録システムを作成するべきである。
      - 使用目的
      - トレーニングデータ(該当する場合)
      - センサ/実世界のデータソース
      - アルゴリズム
      - プロセスグラフ
      - モデル機能(さまざまなレベルで)
      - ユーザー・インターフェース
      - アクチュエータ/出力
      - 最適化目標／損失関数／報酬関数

## 1-2 「倫理的設計バージョン1」構成と概要

# 1 一般原則：イシューと提言候補③

### 原則3 - 透明性

- イシュー:どのように私たちはAI / ASの透明性を確保することができるか？
- 提言候補 システムが客観的に評価され、コンプライアンスのレベルが決定されるように、透明性についての測定・検証可能なレベルを記述する新たな基準を開発する。設計者のためには、こうした標準は、開発中における透明性を自己評価するためのガイドを提供し、透明性を向上させるための仕組みを提案するものである。(透明性メカニズムは、介護の利用者や家庭用ロボットの場合、アルゴリズムを認証する場合、自己調査における情報確保の場合等、場合ごとにより大きく変化すると想定される)

### 原則4 - 教育・啓発

- イシュー:どのように私たちは利益を拡張しつつ、技術の誤用によるAS / AIのリスクを最小限に抑えることができるか？
- 提言候補：
  - 以下のように慎重に周知することで、潜在的なAS/AI の誤用にかかる課題について国民意識を高める：
  - 1. 倫理教育とセキュリティ・アウェアネスを社会に提供し、AI / ASの誤用による潜在的なリスクについて認知を高める。
  - 2. (例えば、FacebookやYouTubeなどのソーシャルメディア上でアクセス可能な科学コミュニケーションを介するなど)新しい方法でこの教育を実現し、最も影響を受ける対象から教育を始めるとともに、AI / ASにかかる一般的な(非生産的な)恐怖感を最小化させる。
  - 3. これらの問題を取り巻く法執行機関を教育することで、市民が、恐怖や混乱を避けるために、彼らと協力しできるようにする(例えば、警察官が、長年学校で公共安全の講義を提供しているのと同様の方法で、近い将来には、安全なAI / ASに関するワークショップを提供することができる)。

## 1-2 「倫理的設計バージョン1」構成と概要

# 2 自律知能システム(AIS)への価値の埋め込み

---

### ○ 概説

- 社会の利益になる自律知能システム(AIS)を開発するために、技術的なコミュニティが、人間の規範や価値観をシステムに埋め込むことができるようにすることが非常に重要である。
- 自律知能システム委員会は、設計者を支援する以下の三方面からのアプローチにより、AISに価値を埋め込むうえでのより広範な目的について検討。

1. AISによって影響を受ける特定のコミュニティの規範や価値を特定する。
2. AIS内でそのコミュニティの規範や価値観を実装する。
3. コミュニティ内の人間とAISの間、これらの規範や価値観の整合性と互換性を評価する。

### ○ 主な提言候補:

- すべてに当てはまる汎用的な規範を作成することは現実的ではないため、特定のコミュニティのための特定のタスクにかかる規範を形成することを提言。複数分野の視点から、複数の実証研究を複数の方法で実施し、規範に関する調査を行い、AIS設計者が参照可能にする。
- ベスト・プラクティスとして、より多くの人数で共有されている価値を優先して反映することを提言。
- 関係者がエンジニアリング・プロセスに関わることを強く推奨。

### ○ 主なイシュー:

- AISに埋め込む価値は普遍的なものではなく、むしろ大部分はユーザー・コミュニティとタスクに特有のものである。AISは、通常、互いに競合する規範や価値観の多様性にさらされている。
- AISは、特定のグループのメンバーを不利にする組み込みデータやアルゴリズムのバイアスを持ちうる。(特定のコミュニティにおけるAIS特有の役割について)関連する規範が設定されたとしても、そのような規範が計算アーキテクチャにどのように組み込まれるべきであるかは明らかではない)
- AISへの規範の実装は、関連するコミュニティの規範に適合しなければならぬ。
- 人間とAISの間の正しいレベルでの信頼を達成する。AISの価値調整にかかる第三者評価。



## 1-2 「倫理的設計バージョン1」構成と概要

# 3 倫理的な研究と設計をガイドするための方法論

- 概説
  - 現代のAI関連組織は、AS / AIの開発の核として、人間の福利、エンパワーメント、自由を確保すべきである。これらの野心的な目標を達成するマシンを創造するため倫理的な研究と設計ガイド委員会は、世界人権宣言で定義されている人権等の人間の価値が、システム設計の方法論から生成されるようにするため、イシューと提言候補を作成した。価値に沿った設計の方法論が、AI/AS関連組織における本質的なフォーカスとなり、人間の進歩に合わせたものとならなければならない。
  - 機械が人間に奉仕し、その逆ではない。この倫理的に健全なアプローチにより、経済と社会の両方のため、AIの経済的・社会的アフォーダンスを維持する均等なバランスを確保しなければならない。
- 主な提言候補：
  - 倫理と倫理的詳察を、すべての大学レベルとより上位の学位において、エンジニアと技術者向けの中核的な科目となるべき。文化横断的・学際的なカリキュラムの構築を提言。
  - 文化横断的な情報倫理の実践者のための役割や、シニア・レベルのマーケター、倫理学者、法律家のための役割を創出する。行動規範を作成する。
  - ソフトウェア・エンジニアに対して、全てのシステムやデータフロー、性能の限界、リスクについて文書化することを求める。透明性を確保するためにアルゴリズムの効率性を下げる提言もあり。
  - 独立した・国威的な組織を設置し、製品の倫理基準について監督させる。
- 主なイシュー
  - AI / ASの個別問題に対処するために、学際的かつ異文化間教育のためのモデルが必要。
  - AI設計に埋め込まれた文化的に特有の価値を区別する必要性がある。
  - 価値ベースの倫理的な文化や産業界のための実践の欠如。価値を意識した指導力の欠如。
  - 倫理的な問題を提起するためのエンパワーメントの、ハイテク・コミュニティの当事者意識や責任感の欠如。
  - AI / ASについて最もよく背景を押さえた利害関係者を含める必要性。
  - 貧弱な文書は、倫理的な設計の妨害となる。一貫性の欠如、もしくは、アルゴリズムの監督の不在。
  - 独立した審査機関の欠如。ブラックボックス・コンポーネントの使用

## 1-2 「倫理的設計バージョン1」構成と概要

# 4 汎用人工知能 (AGI) と人工超知能 (ASI) の安全性と便益

---

- 概説
  - (ASIやAGIと呼ばれる) 将来の優秀なAIシステムは、農業革命や産業革命の規模で世界を変革し、前例のないレベルの世界の繁栄をもたらす可能性がある。
  - AGIとASIの安全性と便益委員会は、AIコミュニティが道程を形作るために協調的に協力することで、こうした変革がポジティブなものとなることを支援するため、 이슈と提言候補を作成した。
- 主な提言候補：
  - 技術面： AI安全性研究により多くの労力を投じる。
  - 技術面： 高度な機能をもつAIを設計する際には、開発時から倫理面に配慮する。
  - 一般原則： レビュー委員会を設け、AI開発を支援・監督する。
  - 一般原則： 商業化過程においても科学コミュニティのオープン性を維持する。
- 主な 이슈
  - AIシステムが、より広範囲かつ多様な領域を横断的にさらなる自動化を行うことで、より複雑な目的関数を最適化する能力を持つことで、予期しない・意図しない動作がさらに危険になる。
  - 将来に向けた安全性の改善や、より一般的に有能なものとするには、AIシステムを難しいものとする。
  - 研究者や開発者は、自律的で有能なAIシステムの開発と展開において、倫理的および技術的な安全性において、さらに複雑化する課題に直面している。
  - 将来のAIシステムは、農業革命・産業革命の規模で世界に影響を与える能力を持ちうる。

## 1-2 「倫理的設計バージョン1」構成と概要

# 5 個人情報と個別のアクセス管理

### ○ 概説

- 個人情報に関する主要な倫理的ジレンマは、データの非対称性である。
- 個人データおよび個別アクセス管理委員会は、ユニークなアイデンティティのキュレーターとして、各人の個人データを定義、アクセス、管理する上での、人々の基本的な必要性について、イシューと提言候補を明確化した。
- 委員会は、完璧なソリューションが存在せず、どんなデジタル・ツールもハッキング可能だと認識した。しかし、委員会としては、ポジティブな将来のために、人々が自己感覚を制御することができるデータ環境を実現することを提言し、データの非対称性を根絶しうるツールや発展的な慣行の事例を提供した。

### ○ 主な提言候補：

- 各個人は、IDを認証するため、トラステッドID認証リソースを特定しなければならない。
- 個人特定可能情報(Individual Identifiable Information)は、各国の個人の資産であり、法的に保護されなければならない。
- 設計者や開発者が「プライバシー・バイ・デザイン」の方法論を利用するための実践的で適用可能な手続きを利用可能とする。
- オープン標準と相互運用性により、エコシステム間を移動する自由を持つ。

### ○ 主なイシュー

- アルゴリズム時代において、各個人が個人データを定義・整理することができるか？
- 個人情報の定義と範囲とは何か？人データに関する制御の定義は何か？
- どのように個人を尊重して、データアクセスを再定義することができるか？
- どのように、それが個人を尊重するように、個人データに関する同意を再定義することができるか？
- どのようにデータ取扱者は、真のインフォームドコンセントを取得するために、個人に示されるデータ・アクセスと収集の(ポジティブな／ネガティブな)結果を保証することができるか？
- 人が、パーソナライズAIやアルゴリズムの保護者を持つことができるか？

# 6 自動兵器システムの再構成

---

- 概説
  - 伝統的な武器や、害を引き起こすように設計されていない自律システムの両方と比べて、身体的危害を引き起こすように設計された自律システムは、追加的な倫理的な予期しない影響を持つ。
  - これらについての専門家の倫理としては、幅広い懸念をカバーする高い基準を持つことができ、また、持つべきである。
  - 大まかに言えば、自律兵器システム再構成委員会は、技術的な組織が、意義のある人間による兵器システムの制御が、社会に有益であることを受け入れ、監査証跡により説明責任を保証し、こうした制御を確保し、こうした技術を作りだした人がその成果の含意を理解し、専門家としての倫理コードにおいて、損害を発生させる意図があると適切に示すことを提言する。
  
- 主な提言候補：
  - 人類を脅かす技術的なイノベーションの創出に貢献するメンバーの選択にかかる行動規範を提言。
  - ターゲット選択、攻撃、自己防衛の自律的機能は、人類にとってネガティブな結果を引き起こすため、こうした決定に際しては、人間の関与が必要とする。
  - この複雑なトピックについて、そのメリットが批判的に評価する包括的な国際的な標準が必要。
  
- 主なイシュー
  - 専門組織の行動規範(Code of Conduct)は、多くの場合、重大な抜け穴をあり、所属メンバーの製造物や人工物、エージェントを見逃してしまう。
  - 人工知能、自律システム、自律兵器システム(AWS)について、重要な概念の定義に関する混乱は、非常に重要な問題に関する詳細で実質的な議論を阻害する。
  - AWSは、元々隠蔽されやすく、非帰属使用となりやすい。

## 1-2 「倫理的設計バージョン1」構成と概要

# 7 経済的・人道的イシュー

### ○ 概説

- 私たちの日々の生活の中で、人間の介入を削減することを目指すテクノロジー、方法論、システムは、急速なペースで進化しており、複数の方法で個人の生活を変革する情勢にある。
- 経済・人道問題委員会の目的は、人間・技術にかかるグローバルなエコシステムを形成する重要な要因を特定し、経済的・人道的な影響に対処し、緊張にかかる難所のロックを解除することで実現可能なソリューションのための主要な機会を提案することにある。
- 委員会の提言の目的は、こうした人間との関係や、関連機関や勃興しつつある情報駆動型技術における中心的な懸念に関する実際的な方向性を提示し、学際的・分野横断的な対話を容易することで、専門的、指示的、相互性のもとでの議論について、より情報を提供することができるようにすることである。

### ○ 主な提言候補：

- 自動化と雇用：統計の整備。自動化による雇用環境の変化の分析。労働者への変化への対応能力向上。
- 説明責任と公平な配分：効果的な教育や訓練にかかる国際的議論に参画。
- AIの恩恵を受けるように開発途上国を支援：倫理的設計のすべての委員会に発展途上国の代表が参加。

### ○ 主なイシュー

- メディアでのAI/ASについての誤解は、市民を混乱させるものである。
- 自動化は、通常、市場の文脈内でのみで認識されている。
- 雇用に関する複雑さは、ロボット/ AIに関して、軽視されている。
- 労働力を(再)訓練する既存の方法に対して、技術の変化のスピードが早すぎる。
- 任意のAIポリシーは、技術革新を遅らせる可能性がある。
- AIと自律的な技術は、世界中で平等に利用できるわけではない。
- 個人情報に関するアクセスと理解が欠如している。
- IEEEグローバル・イニシアティブにおいて、発展途上国からのアクティブな代表が増えることが必要。
- AIと自律システムの出現は、先進国と発展途上国間・内における経済とパワー構造の差を悪化させる。

# 8 法律

### ○ 概説

- AI / ASの開発初期段階では、多くの複雑な倫理的な問題を発生させつつある。これらの倫理的な問題は、ほとんどの場合、直接、具体的な法的課題に転換されるか、付随的で難しい法的な問題を生じさせる。法律委員会は、これまで、急務の領域であるにもかかわらず、非常に少数の実務家や学者のみが関与していることから、法律家のための多くの仕事があると感じている。
- 法律家は、将来に向けたAI/ASが思慮深く方向付けられることにより、人間性と地球が多大な利益をうることができるよう、規制、ガバナンス、国内および国際的な法律上の議論に参加する必要がある。

### ○ 主な提言候補：

- 説明責任と認証可能性はすぐには実現できないことを認識しつつ、不確実な状況において不確実性があることをユーザーに積極的に知らせるようにAIシステムは設計されるべきである。
- 政策立案者等が、適切な場合に、新たな規制を導入する潜在的な必要があるかを慎重に検討すべき。
- 技術者がアイデンティティ・タグ標準（法的説明責任を維持するためのもので、タグがないエージェントはリリースできない）を採用することを検討することや、AIシステムを法的責任を回避するための手段として悪用されないようにすることなど、多数のオプションを検討する。

### ○ 主なイシュー

- どのように自律知能システムにおける説明責任と検証可能性を向上させることができるか？
- どのようにAIの透明性、個人の権利を尊重していることを確認することができるか？例えば、国際、国、地方公共団体が、市民の権利を守るための政府やAIを信頼すべき市民の権利を侵害しているAIを利用している場合がある。
- どのように、こうしたシステムに引き起こされる損害に対する法的責任を保証するようにAIシステムを設計することができるか？
- どのように自律知能システムを設計し、個人データの完全性を尊重した方法で展開することができるか？

# 新設委員会

---

### ○ 混合現実(Mixed Reality)委員会

- ミックス・リアリティが今後生活や仕事、教育等に与える影響を鑑みて設置。AIバックボーンによるリアルタイムの個人化が惹起する倫理的・哲学的課題について検討。
- 見えないセンサーに囲まれた環境と、多面化する個人のアイデンティティのコントロールを含めて、個人の権利を保証するための倫理的枠組みを構築する。

### ○ アフェクティブ・コンピューティング委員会

- 感情、気分、態度、パーソナリティ等の感情的(アフェクティブ)な行動を感知、モデル化、表出することができるようになったAIが個人や社会に与える影響を検討。
- アフェクティブ・コンピューティングやロボットは、コンパニオン、医療分野、リハビリ、高齢者や子どものケア、トレーニングやフィットネス、エンターテインメント、親密な関係等の利用目的を想定。

### ○ AI/ASにかかるイノベーティブなコミュニティのための効果的な政策立案 (EpicAI)

- 効果的な政策立案により、どのように自動知能技術が迅速に世界を変化させることができるかを検討。
- 個人と社会にポジティブな影響を与える、民間・公的部門が推進すべきイニシアティブにかかる提言。
- 人間の福利のためのAI/ASのイノベーションを支援する、より新しい政策立案にモデルに光を当てる。

## 1-3 「倫理的設計バージョン1」に関連するIEEE標準化活動

### IEEE P7000:システム設計における倫理的課題特定のためのモデルプロセス

---

- IEEE P7000– Model Process for Addressing Ethical Concerns During System Design
  - 概要
    - 技術者や関係者が、システムとソフトウェアのライフ・サイクルの初期段階における、エンドユーザーの倫理的懸念の特定・分析・対応のための方法論を必要としている。
    - IEEE P7000標準の目的は、価値ベースのシステム・デザインの方法論の実践的なアプリケーション作成を可能とする。
    - 想定しているプロセス要件としては、新たなIT製品開発の管理とエンジニアリング、コンピューター倫理とITシステム設計、価値に感応する設計、倫理的なITシステム設計における関係者の関与を含む。
  - 申請日:2016年5月18日、承認日:2016年6月30日、期限:2020年12月31日
  - ステータス:活動中
  - ワーキング・グループ:EMELC-WG - Engineering Methodologies for Ethical Life-Cycle Concerns Working Group
  - スポンサー: C/S2ESC - Software & Systems Engineering Standards
  - 所属ソサエティ:C - IEEE Computer Society



## 1-3 「倫理的設計バージョン1」に関連するIEEE標準化活動

### IEEE P7001: 自律システムの透明性

---

#### ○ IEEE P7001 – Transparency of Autonomous Systems

##### ● 概要

- AIにかかる主要な懸念の一つが、関連する幅広い関係者にとって運用に透明性があるかどうか。
  - (i) 利用者にとって、システムの動作を理解するシンプルな方法を提供することで透明性を確保することはシステムへの信頼性の基盤となる(例 ケアロボットの動作)。
  - (ii) AS透明性の検証と認証としては、ASの透明性はシステム監査のプロセス面からみて重要。
  - (iii) 事故が発生した場合には、事故調査官にとって、ASが透明性を持ち、内部プロセスから事故原因を探ることができる必要がある。
  - (iv) 事故後には、法律家やその他の専門家が、証拠について情報提供する上で、透明性が必要。
  - (v) 自動運転車等の破壊的な技術からみると、社会に広く一定レベルの透明性を提供することが、当該技術への市民の信頼性を構築する上で必要。
  - IEEE P7001 は、設計者向けに、開発中における透明性の自己査定を行うためのガイドと透明性改善のためのメカニズムを提供する。(例 センサーと内部状態データにかかるストレージの確保の必要性等)。測定可能・試験可能な透明性のレベルについて記述し、ASの法令遵守のレベルについて客観的に査定できるようにする。
- 
- 申請日:2016年10月13日、承認日:2016年12月7日、期限:2020年12月31日
  - ステータス:活動中
  - ワーキング・グループ:ASV WG\_P7001 - Autonomous Systems Validation Working Group\_P7001
  - スポンサー: VT/ITS - Intelligent Transportation Systems
  - 所属ソサエティ:VT - IEEE Vehicular Technology Society

## 1-3 「倫理的設計バージョン1」に関連するIEEE標準化活動

### IEEE P7002: データ・プライバシー・プロセス

---

- IEEE P7002 – Data Privacy Process
  - 概要
    - システム／ソフトウェアのエンジニアリング・ライフ・サイクル過程におけるプライバシー・ 이슈を管理するための全般的な方法論的アプローチを提供することが目的。
    - 職員、利用者、その他の外部の個人データを利用する製品、サービス、システムにおけるプライバシー志向の配慮のための、システム／ソフトウェアのエンジニアリング・ライフ・サイクル過程における要件を定義する。開発、品質、保証、価値の具体化におけるポリシーをカバー。
    - ユース・ケースやデータモデル(メタデータを含む)を含む。
    - 個人情報の利用を伴う製品、システム、プロセス、アプリケーションの開発と展開を行う組織やプロジェクトを対象。
    - 手続き仕様、ダイアグラム、チェックリストを提供。プライバシー影響評価(PIA)についても記述。
  - 申請日:2016年10月13日、承認日:2016年12月7日、期限:2020年12月31日
  - ステータス:活動中
  - ワーキング・グループ:PDP - Personal Data Privacy Working Group
  - スポンサー: C/S2ESC - Software & Systems Engineering Standards Committee
  - 所属ソサエティ:C - IEEE Computer Society

## 2 大統領府「人工知能・自動化と経済」 位置付けと概要

---

- 大統領府の2016年10月の二つの報告書のフォローアップ
  - 「人工知能の未来に備えて」と「米国人工知能研究開発戦略」のフォローアップ。
  - 「人工知能の未来に備えて」2016年末までに経済的影響に関する報告実施を提言。
- 大統領府大統領府「人工知能・自動化と経済」
  - 2016年12月公表。AI駆動型の自動化が経済に与える影響について検討。
  - 大統領府経済諮問委員会(CEA)、国内政策会議(DPC)、国家経済会議(NEC)、行政予算管理局(OMB)、科学技術政策局(OSTP)のスタッフが作成。
  - AIの便益拡大と費用低減を図るための戦略を記述。
  - AI駆動型の自動化は数年から数十年にかけて経済を変革する可能性があることから、政策立案担当者は、関連政策をアップデート、強化、採用する必要があると指摘。
  - 現在のAIブームは、ビッグデータ、機械学習の改良、コンピューターの処理能力の向上により、2010年頃から活性化したと認識。
  - 技術変化による恩恵と、労働市場での対応はこれまでも発生。
  - 先進国12か国の分析では、AIにより2035年までの経済成長率が倍増する推計あり。
  - 世界経済フォーラムは、「第四次産業革命」と呼ぶ。

## 2 大統領府「人工知能・自動化と経済」 問題意識と三つの戦略提言

- 経済への影響予測は不可能だが、政策立案担当者は広範な潜在的な影響について備えなければならない。
- 少なくとも、運転手やレジ係の職の一部は置き換えられるかリストラされる可能性がある。
- AI駆動形自動化の影響は、経済全体に亘るため、政策ターゲットも経済全体を対象とするべき。
- AI駆動形自動化の影響は、その他の技術変化、グローバル化、市場競争の減少、労働者側の交渉力、過去の公共政策等の要因と切り離すことが困難だが、広範な政策的対応は必要。
- もし、特定の場所や領域が受ける影響を緩和することができるのであれば実施するべき。
- しかし、AI駆動型自動化が米国経済に与える影響に対応するため、広い領域を対象とした三つの戦略を提言。

### 1 多くの便益のためにAI開発に投資する。

:サイバー防御、STEM領域、アルゴリズムック・バイアス、スタートアップ支援等の競争環境の創出。

### 2 将来の職業に向け、米国民を教育する。

:高校・大学での教育、職業教育、中等教育、子供向けの早期教育、労働者向け教育。

### 3 移行過程において労働者を支援し、広範囲な成長を共有できるようにする。

:労働者が最も良い雇用機会を追求可能に。ソーシャル・セーフティ・ネットの現代化。失業保険制度の強化。最低賃金の値上げ、超過勤務手当の提供、労働組合の強化等。地域別の対応。税制改革。

- 上述の戦略を実行し、米国がAIの創造と利用におけるリーダーシップを確保するため、政府、産業界、技術・政策の専門家、市民の参画が重要。
- 今後も取り組みの継続必要。詐欺的な取引やメッセージを検知可能なシステムを改良しうる、AIによるサイバー・防御における潜在的可能性がある」と指摘。

## 2 大統領府「人工知能・自動化と経済」

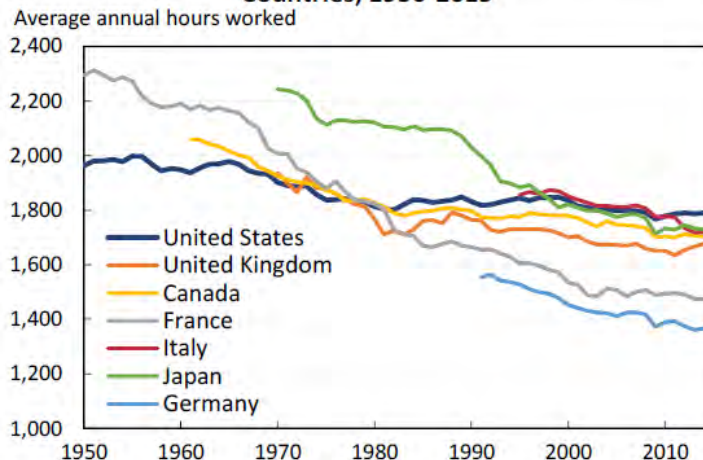
# AI駆動型の自動化と経済にかかる分析

- 経済的影響の予測は困難だが以下の5つの経済的影響に備えるべき。
  - 生産性を向上させるポジティブな貢献
  - より高レベルの技術スキルへのさらなる需要増を含む労働市場で求められるスキル変化
  - 分野、収入レベル、教育レベル、職種、ロケーションによる影響の不公正な分配
  - ある職種は消滅する一方で新規の職が創出されるという労働市場の変動
  - 短期的にはある労働者の職が失われ、政策的対応によって失職が長期化する可能性

## ■ AIとマクロ経済：技術と生産性向上

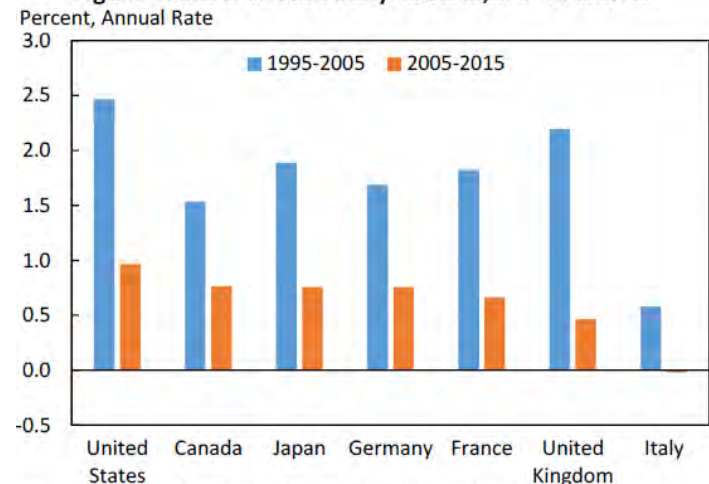
- G7における労働者一人当たりの年間労働時間は長期的には減少傾向
- 労働生産性伸び率は鈍化。AI駆動型自動化による労働生産性向上に期待。

Figure 1: Average Annual Hours Worked per Worker, G-7 Countries, 1950-2015



Source: OECD Stat

Figure 2: Labor Productivity Growth, G-7 Countries



Source: Conference Board, Total Economy Database; CEA calculations.

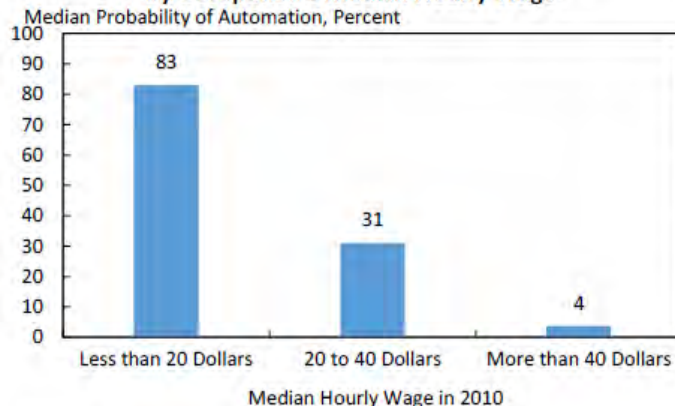
## 2 大統領府「人工知能・自動化と経済」

# AI駆動型の自動化と経済にかかる分析

### ■ AIと労働市場：短期的影響と政策の必要性

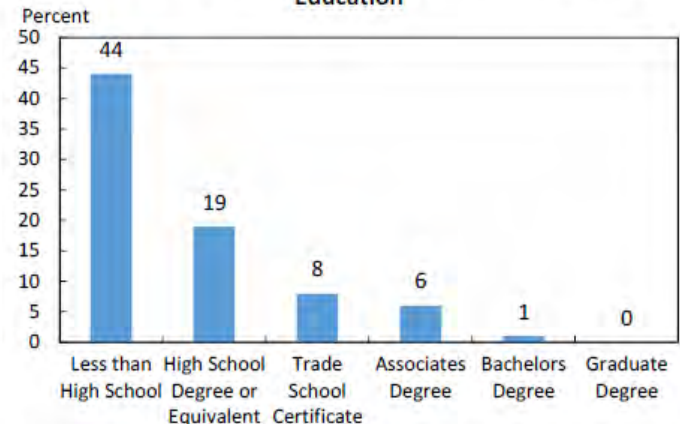
- 完全に自動化される職は全体の9%程度と推計。
- 時給20ドル以下の職の83%が自動化によるプレッシャーあり。
- 高卒以下が従事することが多い職が代替される率が高いと推計。
- AIにより創出される職：AI関連職(例 IBM Watsonを活用する医療関係者)、AI開発職、AI監督者。
- 成長の利益配分は、労働市場にかかる政策に影響を受ける。米国の労働市場にかかるプログラムの支出は、対GDP比率で最も低いグループ(メキシコ、チリの次)に入る。
- 米国はトップ1%の人の所得の合計が全体に占める割合がG7で最も高い(2015年18%超)

Figure 3a: Share of Jobs with High Probability of Automation, by Occupation's Median Hourly Wage



Source: Bureau of Labor Statistics; Frey and Osborne (2013); CEA calculations.

Figure 3b: Share of Jobs with Highly Automatable Skills, by Education



Source: Arntz, Gregory, and Zierahn (2016) calculations based on the PIAAC 2012.

※自動運転車のケーススタディ：時給20ドル以下の職種。代替率が高いとの推計。

# おわりに

---

- AI/ASの技術的特性に起因するリスクへの対応： グローバル
  - 技術が内包するリスクに対応し、安全性を確保する必要性
  - 安全性を確保するためには設計段階からの予防的措置重要 c.f. 事後対応
  - 予防的措置の拡充方策の一つとして、標準化、ガイドラインの策定あり
    - IEEEのイニシアティブ、倫理的設計(EAD)の提言案、IEEE7000シリーズ。
    - IEEEはグローバルな標準化団体。フォーラム標準。基本的に個人ベース。
    - 開発にかかる標準で対応できる範囲は技術に起因するリスク。
  
- AI/ASの利用拡大に起因する影響への対応： 各国レベル
  - AI/AS技術の利用拡大により経済・社会に影響を与えうる。
  - 経済面では、産業構造変化とそれに伴う雇用への影響の懸念あり。
    - 米国政府による対応策の提言「人工知能・自動化と経済」。
    - 標準開発等では対応できない領域について、主に米国内における社会政策（教育、雇用、セイフティ・ネット等を含む）を強化する提言。