国外の検討動向

米国科学技術政策局 「人工知能の未来に備えて」 (2016)
- 人工知能がもたらす利益とリスクに関する一般市民にも公開されたワークショップとそれらに基づき政策提言を行う報告書。

〇米国 人工知能研究開発戦略 (2016)
- 人工知能の研究を推進すべき産業領域や悪影響の最小化しつつ利益を得るための開発の戦略を策定

〇米国 人工知能、自動化、そして経済 (2016)
- 人工知能や自動化が経済に及ぼす影響を検討し、政策提言を行う報告書。

〇 Stanford AI100 (2014-)
- アメリカ人工知能学会の検討部会の流れをくむ人工知能がもたらす未来を検討する組織。

〇 The IEEE Global Initiative for Ethical Considerations in AI/AS (2016-)
- 倫理標準を目指してEthically Aligned Designを公開

〇 Future of Life Institute (2014-)
- 人工知能がもたらす危機を検討し、その研究を支援する自主組織。

〇 Future of Humanity Institute (2005-)
- 英国オックスフォード大学の技術がもたらす未来、特に危機を研究する組織。

〇 Centre for the Study of Existential Risks (2012~)
- 英国ケンブリッジ大学の人工知能や科学技術の発展がもたらす世界的な破滅的危機を検討する組織。

〇 Partnership on AI (2016~)
- 企業が連合して、人工知能が社会に及ぼす影響について検討し、一般市民に正しい情報を伝えるための活動を行う。

〇 Machine Intelligence Research Institute (2000~)
- 人を超えるような人工知能がもたらす影響の予測や良い影響を与える人工知能についての研究組織。

〇 Open AI (2015~)
- 人によって有益な人工知能を開発するためのオープンな組織。

〇 RoboLaw (2012~2014)
- EUのFP7によるロボット法に関する研究プロジェクト。

〇OECD Seizing the benefits of digitalization for growth and well-being (2017-2018)
- デジタル化によって生じる恩恵を把握し，必要な政策のための問題点を検討。

米国 科学技術政策局「人工知能の未来に備えて」

- Whitehouse Office of Science and Technology Policy (OSTP) Preparing for the Future of Artificial Intelligence
- **2016年5月3日、ホワイトハウスのウエブサイトで科学技術政策局次長がAIがもたらす利益とリスクを深く知るための活動を開始すると発表。**
- **アカデミア、NPOとワークショップを共催し、AIと機械学習についての議論を促し、それらの課題とチャンスを明らかにする。**
  1. **5月24日: AIと法・統治の関わり**
  2. **6月7日: 社会的利益のためのAI**
  3. **6月28日: AIの安全性と制御**
  4. **7月7日: AI技術と社会・経済の関わり**
- **AIと機械学習を活用して行政サービスの提供を改善するため、政府機関横断型のワーキンググループを設置。**
- https://obamawhitehouse.archives.gov/blog/2016/05/03/preparing-future-artificial-intelligence

報告書「人工知能の未来に備えて」

- 10月12日、国家科学技術会議 (NSTC) と科学技術政策局 (OSTP) が中心となってまとめた報告書を公表。23の提言を含む。
  - https://www.whitehouse.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

- **提言1**：民間及び公的機関は、AI及び機械学習を、社会に便益をもたらすように責任をもって活用できるのかどうか、如何にしてそれが可能かを検討すべきである。
- **提言2**：連邦政府関係機関は、AIのためのオープンな学習データやオープンなデータ標準に関する取組みに優先的に取組むべきである。
- **提言3**：連邦政府は、重要な政府関係機関のAI活用能力を向上させる方法を模索すべきである。
- **提言4**：ホワイトハウス 国家科学技術会議 機械学習・人工知能小委員会は政府のAI担当者のための実践コミュニティを作るべきである。
- **提言5**：政府関係機関は、AIによりもたらされた製品に対する規制にあたっては、シニアレベルの適切な専門家の意見を参考にすべきである。
- **提言6**：政府関係機関は、技術の現状についてのより多様な視点を持った連邦政府機関職員を育成するため、あらゆる人員配置・交換モデル (例：第一人者の雇用) を活用すべきである。
- **提言7**：運輸省は、安全や研究、又はその他の目的のためのデータ共有の促進に、産業界や研究者とともに取り組むべきである。
- **提言8**：合衆国政府は、高度な拡張性を備え、自動と手動両方の航空機等に完全に対応可能な、先進的で自動化された航空交通管理システムの開発と社会実装に投資すべきである。
- **提言9**：運輸省は、新しい車両設計を含む、完全自動走行車とドローンを交通システムに安全に統合するための発展途上の規制枠組みの構築を継続すべきである。
- **提言10**：ホワイトハウス 国家科学技術会議 機械学習・人工知能小委員会は、AIの進展を注視し、AIの現状、とりわけ画期的出来事について定期的に上級の政府指導者に報告すべきである。
- **提言11**：政府は他国のAIの現状、とりわけ画期的出来事を注視すべきである。

- **提言12**：産業界は、近々起こりそうな画期的出来事を含め、産業界におけるAIの最新動向を政府と共有すべく、政府と協働すべきである。
- **提言13**：連邦政府は基礎的で長期的なAI研究に重きを置くべきである。
- **提言14**：ホワイトハウス 国家科学技術会議 機械学習・人工知能小委員会とネットワーキング・情報技術研究開発小委員会は、科学・技術・工学・教育委員会とともに、AIの研究者、専門家、ユーザを含む人材の規模、質、多様性を適切に高度化するための対策を取るべく、AI人材パイプラインの研究に着手すべきである。
- **提言15**：大統領府は、アメリカ労働市場に対するAIとオートメーションの影響についてさらに調査し、推奨される政策の輪郭を描くべく、年末までに追加の報告書を公表すべきである。
- **提言16**：個人に関する重大な意思決定を支援するAIシステムを利用する連邦政府機関は、エビデンスに基づく検証と実証に基づいて、そうしたシステムの効果と公平性の確保にとりわけ配慮すべきである。
- **提言17**：個人に関する重大な意思決定のためのAIシステムの利用を支援すべく州政府や地方政府に助成している連邦政府機関は、連邦補助金で購入したAI製品やAIサービスが透明性の高い結果を生み、効果と公平性に関するエビデンスに支持されていることを保証するため、助成に関する条項を再検討すべきである。
- **提言18**：学校と大学は、AI、機械学習、コンピュータ科学、データ科学のカリキュラムの重要項目として、倫理や、セキュリティ、プライバシー、安全についての関連トピックを加えるべきである。
- **提言19**：AIや安全性についての専門家集団や学界は、AI安全性工学の成熟に向けて協働すべきである。
- **提言20**：合衆国政府は、AIに関する国際的関与についての政府を挙げての戦略を立て、国際的な関与とモニタリングを要するAIのトピックのリストを作成すべきである。
- **提言21**：合衆国政府は、情報交換とAI研究開発の協働の促進のため、外国政府や国際機関、産業界、学術界等を含む重要な国際的ステークホルダーとの関わりを深めるべきである。
- **提言22**：連邦政府機関の計画や戦略は、AIのサイバーセキュリティに対する影響と、サイバーセキュリティのAIに対する影響に対して説明責任を負うべきである。
- **提言23**：合衆国政府は、自動、半自動の兵器に関して、国際的人道法に沿った、政府を挙げての唯一の方針を打ち立てるべきである。

米国人工知能研究開発戦略

- The National Artificial Intelligence Research and Development Strategic Plan
- **ホワイトハウス 国家科学技術会議** (The National Science and Technology Council) **の機械学習・人工知能小委員会の依頼により、同会議 ネットワーキング・情報技術研究開発小委員会が** 2016**年**10**月**13**日付で策定。**
- **産業界の取組みが期待しにくい領域を特に意識**
- **究極の目的は、悪影響を最小化しつつ、多岐に渡る社会的便益をもたらすAIに関する新たな知識と技術を生み出すこと。**

参考URL:
https://obamawhitehouse.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf

- **戦略1：AI研究への長期的投資。**
- **戦略2：人間とAIの効果的協働の方法の開発。**
- **戦略3：AIの倫理、法、社会的意味合いの理解と対処。**
- **戦略4：AIシステムの安全・安心の保障。**
- **戦略5：AIの学習とテストのための公に共有可能なデータセットと環境の整備。**
- **戦略6：技術的な標準やベンチマークによるAI技術の測定と評価。**
- **戦略7：国家的なAI研究開発人材の需要についてのより深い理解。**

人工知能、自動化、そして経済
Artificial Intelligence, Automation, and the Economy

- **米国大統領行政府** (the Executive Office of the President (EOP)) **による人工知能や自動化が経済に及ぼす影響についての報告書** (2016年12月20日)。
- **同年10月のOSTPレポート「人工知能の未来に備えて」を受けて検討。**
- **人工知能による自動化が経済や労働市場、雇用に及ぼす影響を調べ、それに対する適切な政策を明らかにする。**

参考URL:
https://obamawhitehouse.archives.gov/blog/2016/12/20/artificial-intelligence-automation-and-economy

- 戦略1：人工知能への投資と開発の推進
- 戦略2：将来の仕事に対応するための国民への教育と訓練の整備
- 戦略3：労働者の移動、業務移行を可能とし強化するための政策、それに付随するセーフティネットの整備

Stanford AI100

One hundred year study of Artificial Intelligence

- **組織**
  - **主催者** Eric Horvitz (MS**ディレクター，コンピュータ科学**)
  - **常任委員会**7**名，**5**年毎評価**
    - **コンピュータ科学，生物計算学，自然言語処理，情報セキュリティ，ゲーム理論，機械による言語学習，**AI**応用**
  - **調査団** (study panel) 17**名**(2015)
  - **多くが**AI**に近い自然科学者，工学者。大学，研究者。**
  - **組織的活動。**

- **時期**
  - 2014**年発足**
  - 2008-2009**年の**Advancement of Artificial Intelligence (AAAI) **の**Long-term AI Futures**を継承**

- **内容**
  - **長期的にAIが社会にもたらす影響の調査・推定**
  - **まず，良い面・可能性を検討，次に懸念事項。**
  - **骨子には18トピック**
    - **技術動向，AIの主要な可能性，社会実装，プライバシー，民主主義，法律，倫理，経済，戦争，犯罪，AIと人の協働，認知，安全性，制御不可能問題，心理，社会受容，神経科学，哲学**

- **ELSIとの関係**
  - **プライバシー：何をどこまで推定されても許容できるか**
  - **民主主義：人の信念や投票への影響が既にある**
  - **法律：AIの自動性によって生じる責任法**
  - **倫理：非倫理的なAI，人のようなAIとの対話時の倫理，AIは自らがAIであることを常に明示するべきか。**
  - **経済：人の雇用・職種変化，AIによる金融市場操作の危険性**
  - **戦争：兵器システムへのAI利用は現実的問題。**
  - **犯罪：犯罪的AI，悪意のあるAIは秘密裏に学習し続け，秘密裏に犯罪行為を行う。**
  - **人と機械の協働：人からAI，AIから人への権限委譲をどう行うか。AI推論の可視化・説明可能性（透明性）をどう確保するか。**

- ELSIとの関係
  - 安全性：新しい状況での自律性，ロボット三原則的なものが必要か。
  - 制御不能問題：singularityへの不安・恐れは，検討し，確率を示し，先見的な努力やガイドラインを示すことで低減可能。
  - 心理：人を超える知性に対する不安，Oz法などによる事前の研究が必要
- 関連URL
  - https://ai100.stanford.edu/reflections-and-framing
  - https://ai100.stanford.edu/sites/default/files/ai100_framing_memo_0.pdf
  - http://www.aaai.org/Organization/presidential-panel.php
  - http://www.aaai.org/Organization/Panel/panel-note.pdf
  - http://www.aaai.org/Organization/asilomar-study.pdf

Ethically Aligned Design:

A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems

- **組織**
  - **米国電気電子学会** (IEEE) 、The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems
  - 100**人以上の**AI、**倫理、法などの有識者の意見に基づき最初の報告書を**2016**年**12**月に公開**
  - **自律的な知能システムの全ての開発者に倫理的な視点や考慮すべき重要課題を提案すること、および倫理指針の標準化への提言**
  - http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

- **時期**
  - 2016**年**12**月、最初の報告書を公表。**
  - **広く意見を集め、**2017**年**4**月にその結果を公表予定**
  - 2017**年**6**月に公開シンポジウムを開催予定**
  - 2017**年秋に第**2**版を公開予定**

- 第1章：AIと自律システムに関する高次な倫理原則
- 第2章：AIS（自律化した知能システム）に価値（規範）を実装することについて
- 第3章：倫理的な研究と設計に導く方法について
- 第4章：AGI（汎用人工知能）の安全性と便益について
- 第5章：個人情報保護と情報アクセスについて
- 第6章：自律兵器システムについて
- 第7章：経済的問題や人道的問題について
- 第8章：法的問題について

Future of Life Institute (FLI)

- **組織**
  - **ボストンの自主組織**
  - **創立者** Jaan Tallinn (Skype**共同創始者**)，Max Tegmark (MIT，**宇宙科学**) **など**5**名。博士学生も**2**名。**
  - **科学アドバイザリーボード**13**名**
    - Nick Bostrom (Oxford**哲学**)，Morgan Freeman (**俳優**)，Stephen Hawking，Christof Kock，Elon Musk (SpaceX, Tesla)，Stuart Russell (UCBarkley，AI)
  - **自然科学，哲学，俳優，学生，起業家など幅広い。**
- **時期**
  - 2014**年**3**月発足**
  - 2014**年**5**月**24**日** “The Future of Technology: Benefits and Risks” **パネルディスカッション**
  - 2015**年**1**月**2-5**日** “The Future of AI: Opportunities and Challenges” **カンファレンス**

- **内容**
  - AI**の発展を中心に，バイオテクノロジー，核兵器，気候変動の危険性** (Existential Risk) **を扱う。**
  - Web**サイトはブログ形式。ニュースの寄稿。**
  - **寄付者から集めた資金で，研究助成。**
  - **議論やカンファレンスの開催。**
  - **啓蒙活動** (open letter (**頑健で有益なAI開発への署名活動**) , news letter)
  - General AI (AGI, strong AI) **の将来的危険性と**AI Safety
- ELSI**との関係**
  - 2015**年1月2-5日** “The Future of AI: Opportunities and Challenges” **カンファレンス** (**スライド多数**) **，** 18 talks
  - **経済**：Erik Brynjolfsson, MIT **人はIT の進化に遅れている** (https://www.ted.com/talks/erik_brynjolfsson_the_key_to_growth_race_em_with_em_the_machines?language=ja)

- ELSIとの関係
  - **哲学**：Stuart Russell (UC Berkeley). Beneficial AIの提案。AI倫理ではなく，AIの常識common senseが大切。AIの開発の方向を再考すべき。
  - **雇用**：Michael Osborne (Oxford), 雇用・職種分布の変化が必然
  - **人レベルの**AI：Hassabis (DeepMind)など
  - **知性爆発**：Bostrom (Oxford)，Superintelligenceの著者。FHI主宰。いずれsuperintelligenceは来るので，人間性に有益に，倫理的に開発されるべき。
  - **モラル**：Joshua Greene (Harvard)。研究者，技術者のためだけでなく全ての人についてのモラルを考える必要
  - **法，倫理，自動兵器**：Heather Roff Perkins (Univ. Denver)，殺人ロボットの標的を人にしてよいか。USポリシー (DoD) では自動兵器は人を標的にすることを禁止，ただし半自動であれば許される。しかしこの境界は曖昧である。

- 2017年1月3-7日　Beneficial AI 2017 conference
  - **前回の会議 (2015年) を引き継ぎ、開催**
  - 1/3-4**は、**FLI**の研究助成の成果発表ワークショップ**
  - 1/5-7**は、**Beneficial AI **に関する学術会議**
  - **会議の成果として、**Asilomar AI Principles **を公表**
    https://futureoflife.org/ai-principles/
    - AI**の研究、助成、政策、安全性、透明性、バリューアラインメントなど倫理法社会的問題についての原則**

- **関連**URL
  - http://futureoflife.org/team/ **メンバー**
  - http://futureoflife.org/wp-content/uploads/2016/02/FLI-2015-Annual-Report.pdf
  - http://futureoflife.org/ai-open-letter/ **オープンレター**
  - http://futureoflife.org/data/documents/research_priorities.pdf
  - http://futureoflife.org/2015/10/12/ai-safety-conference-in-puerto-rico/ **カンファレンス**
  - http://futureoflife.org/background/benefits-risks-of-artificial-intelligence **多くの**AI safety**リソース，リンク**
  - http://futureoflife.org/2016/01/25/ai-open-letter-2/ **オープンレター改訂**
  - https://futureoflife.org/background/benefits-risks-artificial-intelligence-japanese/ **日本語での人工知能の利益と危険性に関する説明**
  - https://futureoflife.org/bai-2017/ Beneficial AI conference

Future of Humanity Institute (FHI)

- **組織**
  - Oxford**大学，**
  - **創立者** Nick Bostrom，「Superintelligence: Paths, Dangers, Strategies」**の著者**
  - **メンバー**：12**名以下の研究者。論文**200**本，引用**5,500**以上。**
  - **多くが哲学者，倫理学者など人文系。＋自然科学，**AI。
  - The Strategic Artificial Intelligence Research Centre，AI**開発の政策提言を目指す。**Oxford**と**Cambridge**の共同。**
- **時期**
  - 2005**年**3**月発足**

- **内容**
  - Existential risk**全般を扱う。**
  - **現実的な倫理学：ある選択肢の中でどちらがより良いかを検討。**
  - AI**の安全性：いかにして危険性を排除しつつ，高度な**AI**を利活用するか。**
  - **技術予測と評価：将来の複数のリスクとチャンスの間で検討する優先順位を決め，技術の間の相互作用を見極め，人類の未来のための実行可能な介入の仕方について判断。**
  - **産業と政策：様々な行政，産業組織とのコラボレーション。**
- **関連**URL
  - http://www.fhi.ox.ac.uk/
  - https://www.fhi.ox.ac.uk/about/the-team/ **メンバー**
  - https://www.fhi.ox.ac.uk/research/research-areas/strategic-centre-for-artificial-intelligence-policy/

Centre for the Study of Existential Risks (CSER)

- **組織**
  - **ケンブリッジ大学**
  - **主宰：** Huw Price (**哲学**) , Lord Martin Rees (**宇宙学，宇宙物理**) , Jaan Tallinn (**スカイプ，コンピュータ科学**)
- **時期**
  - 2012**年発足**
- **内容**
  - **特に**General-AI**を中心として科学技術の発展で起きる**global catastrophic risks**の予測，影響の評価**
  - **最終的なゴールは，**AI**の安全性やリスクについての研究を前進させ，産業界のリーダーや政策立案者に，**AI**のもたらす利益が安全に現実のものとなるよう，適切な戦略や規制に関する情報を提供することである。**

Partnership on AI

- **組織**
  - Amazon、DeepMind/Google、Facebook、IBM、Microsoft、Apple**から構成される。正式名称は、**Partnership on Artificial Intelligence to Benefit People and Society**。**
- **時期**
  - 2016**年発足**
- **内容**
  - **人工知能技術についてのベストプラクティスの研究と定式化**
  - **一般市民のAIについての理解の増進**
  - **AIとその人間、社会に対する影響についての議論と活動のためのプラットフォームとしての貢献。**
- **関連URL**
  - https://www.partnershiponai.org/

Machine Intelligence Research Institute (MIRI)

- 組織
  - 独立研究所，アメリカ・バークレー
- 時期
  - 2000年
- 内容
  - 人を超えるAIの創造がポジティブなインパクトを与えるための研究活動。
  - 研究テーマ
    - 未来のAI技術について予測できる（できない）こと
    - AIの高信頼性とエラー許容性
    - いかにしてAIに人間の価値観を教えるか
- 関連URL
  - https://intelligence.org/

Open AI

- **組織**
  - **主宰**Elon Musk (**テスラ**)
  - **独立の非営利研究組織，アメリカ・サンフランシスコ**
- **時期**
  - 2015**年**
- **内容**
  - **金銭的な利益を上げる必要性に縛られることなく，人類全体の利益になるように**AI**を前進させる。**
  - **特許や研究をオープンにすることで，他の研究機関や研究者と自由にコラボレーション。**
- **関連**URL
  - https://openai.com/blog/introducing-openai/

RoboLaw

- 組織
  - EUの研究プロジェクトFP7
  - 独立の非営利研究組織，アメリカ・サンフランシスコ
- 時期
  - 2012年03月から2014年05月
- 内容
  - 応用領域：自動運転，ロボット人工装具（義手義足，外骨格），手術ロボット，コミュニケーションロボット
  - 各領域について，技術分析，関連する問題の列挙と検討，政策立案
  - 法令として必要なルール，インセンティブなどの検討
- 関連URL
  - http://www.robolaw.eu/

OECD Seizing the benefits of digitalization for growth and well-being

- **組織**
  - OECD (経済協力開発機構)・**CDEP (The Committee on Digital Economy Policy)**
- **時期**
  - 2017-2018**年**
- **内容**
  - **デジタル化によって生じる恩恵を把握し，必要な政策のための問題点を洗い出す。**
    - **知性と制御：　ローカル化・個人化が進み，脱中央・中枢化する。**
    - **情報の流れ：　情報が国境やセクターを超えて流れる。中央制御は困難となり新しい仕組みが必要。**
    - **資本の性質：　モノ的な資産から知識的・非モノ的資産へ**
    - **スケール効果：　デジタル生産物は流通コストがほぼゼロ。少人数で大規模資産が動かせ，少数の勝者がすべてを得る。**
    - **市場と関係：　グローバル経済が標準であり，低コストの新しい流通システムとそれに対応する政策が必要。**
    - **意思決定：　従来からの階層的意思決定システムではなくなる。**Bigdata やAI**が活用される意思決定システムに。**

- 内容
  - なすべき施策・課題
    - 経済と社会の非中央化・中枢化への対応
    - デジタル化によって生じる職業の変化
    - 将来必要となるスキル・知識
    - デジタルイノベーション，規制，政治経済の変革
    - 生産性と包括的成長（誰もが平等に）
    - デジタル化で生じる社会的・環境的課題
    - 実証的で評価の伴う政策への移行

国内の検討動向

○ **総務省 AIネットワーク化検討会議** (2016)
○ **総務省 AIネットワーク社会推進会議** (2016~)
- ネットワークの観点から、ＡＩ開発原則、社会的・経済的・倫理的・法的課題を総合的に検討。

○ **JST-RISTEX 人と情報のエコシステム**(2016~)
- 社会の理解のもとに技術と制度を協調的に設計していくための研究開発を推進

○ **内閣府 新たな情報財検討委員会**(2016~)
- 社会の理解のもとに技術と制度を協調的に設計していくための研究開発を推進

○ **NEDO 次世代人工知能技術社会実装ビジョン** (2015)
- AI関連技術・社会実装の予測。5年後，15年後，20年後を予測。

○ **経産省 ＣＰＳによるデータ駆動型社会の到来を見据えた変革** (2015)
- サイバーフィジカルシステムの課題の検討。

○ **JST 知のコンピューティングとELSI/SSH** (2014)
- 知のコンピューティングのELSIに関するワークショップ。

○ **人工知能学会 倫理委員会** (2014~)
- 倫理指針 公表（2017/02）。

○ **Acceptable Intelligence with Responsibility** (2014~)
- 人間・ＡＩ融合時代の社会の制度や倫理等について議論する、有志の研究者からなる組織。

○ **ロボット法研究会** (2016~)
- 「ロボットをとりまく人間の保護と責任」として法的課題を概観。

○ **AI社会論研究会** (2015~)
- 人工知能が社会に与える影響について議論する有志による組織。

○ **社会におけるAI研究会** (2006~)
- AIの研究成果の社会実装促進，安全性や生活の利便性向上を議論する人工知能学会の分科会。

○**NIRA わたしの構想 人工知能の近未来**(2015)
- AIの近未来についての有識者へのテーマインタビュー。

○ **NIRA AIをどう見るか “Edge Question”から探るAIイメージ**(2016)
- 欧米の有識者のAIに対する意見集の紹介と分析。

AIネットワーク化検討会議

- 組織
  - 主催者 総務省
  - 座長：須藤修（東大情報学環），顧問：村井純（慶應大）
  - 理工系，人文社会系有識者含め，計37名
- 時期
  - 2016年2月発足
- 内容
  - 目指す社会像・基本理念
    - AIの恵沢をあまねく享受，個人の尊厳と安心安全，イノベーティブな研究開発と公正な競争，制御可能性と透明性，地球規模課題の解決に貢献

AI開発原則

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

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、利用者に選択の機会を適切に提供するよう配慮

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

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

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

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･第三者のプライバシーを侵害しないよう配慮

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

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に対するアカウ ンタビリ ティを遂行

AIネットワーク社会推進会議

- 総務省 AIネットワーク化検討会議の後継（2016年から）。
- 検討事項
  1. 「AI開発ガイドライン」（仮称）の策定に向けた国際的な議論の用に供すべき素案の検討
  2. AIネットワーク化が社会・経済の各分野にもたらす影響とリスクの評価
  3. 1.および2.に掲げる事項のほか、社会全体におけるAIネットワーク化の推進に向けた社会的・経済的・倫理的・法的課題に関連する事項
- 「開発原則分科会」と「影響評価分科会」の2つの分科会を持つ。

人と情報のエコシステム

- **組織**
  - JST-RISTEX (**社会技術研究開発センター**) **による研究開発事業**
    - 領域総括：國領二郎（慶應義塾大学総合政策学部 教授）
- **目的**
  - **ビッグデータを活用した人工知能、ロボット、IoTなどの情報技術の急速な進歩の一方で、情報技術は様々な問題をもたらしうるとの指摘もなされている。それらの問題に適切に対処していくために、情報技術を人間を中心とした観点で捉え直し、社会の理解のもとに技術と制度を協調的に設計していくための研究開発を推進し、情報技術と人間のなじみがとれた社会の実現を目指す。**
- **時期**
  - 2016**年秋発足**
- **関連**URL
  - http://ristex.jst.go.jp/hite/

- 2016年度採択プロジェクト
  - 多様な価値への気づきを支援するシステムとその研究体制の構築
    - 江間 有沙　(東京大学 教養学部附属教養教育高度化機　特任講師)。
  - 日本的Wellbeingを促進する情報技術のためのガイドラインの策定と普及
    - 安藤 英由樹　(大阪大学 大学院情報科学研究科　准教授)
  - 「内省と対話によって変容し続ける自己」に関するヘルスケアからの提案
    - 尾藤 誠司　(独立行政法人国立病院機構 東京医療センター 臨床研究センター 政策医療企画研究部臨床疫学研究室　室長)
  - 未来洞察手法を用いた情報社会技術問題のシナリオ化
    - 鷲田 祐一　(一橋大学 大学院商学研究科　教授)
  - 法・経済・経営とAI・ロボット技術の対話による将来の社会制度の共創
    - 新保 史生　(慶應義塾大学 総合政策学部　教授)

新たな情報財検討委員会

- 組織
  - 内閣府知的財産戦略本部
- 目的
  - 総論：新たな情報財の保護・利活用について、我が国として目指すべき基本的な方向性を検討
  - 論点①：データの保護・利活用の在り方
  - 論点②：AI※の作成・保護・利活用の在り方
  - ②－1　学習用データ
  - ②－2　AIのプログラム
  - ②－3　学習済みモデル
  - ②－4　AI生成物（AI創作物とサービスにつながる出力等を含む）
    - ※　本検討委員会では、産業競争力強化の観点から具体的に知財制度上の検討が必要と考えられる「いわゆる弱いAI」（人間が知能を使ってすることを機械にさせようとする立場からのAI）のうち、特に深層学習を含む機械学習を念頭に検討。
- 時期
  - 2016年発足
- 関連URL
  - http://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/

次世代人工知能技術社会実装ビジョン

- 組織
  - 次世代人工知能技術社会実装ビジョン作成検討会
  - NEDO
  - 麻生英樹（産総研人工知能研究センター，副センター長）
  - 川上登福（株式会社共創基盤，マネージングディレクター）
  - 松尾豊（東京大学准教授）
- 時期
    - 2016年4月21日公表

- 内容
  - 人工知能技術の進展予測
  - 時間軸を設定
    - 2020年,
    - 2030年,
    - 2040年
  - 出口分野を設定
    - ものづくり
    - モビリティ
    - 医療･健康, 介護
    - 流通･小売り, 物流
  - 時間軸3x出口分野4のマトリックスで分析

ＣＰＳによるデータ駆動型社会の到来を見据えた変革

- 組織
  - 経済産業省･産業構造審議会　商務流通情報分科会　情報経済小委員会
- 時期
    - 2015年4月　中間とりまとめ公表
- 内容
  - サイバーフィジカルシステムCPSの実装における課題，必要な施策，具体的な取り組みの調査
- 倫理･人間社会との関係
  - 「ＣＰＳが産業や社会にもたらす影響」
    - CPSの自律化が進むと人間の果たす役割の代替が進む
    - 個々の情報がより他者に晒され，情報コントロール権の喪失や，データの由来者が特定される可能性
    - セキュリティ･リスク，コンプライアンス･リスクの増大

JST CRDS「知のコンピューティングとELSI/SSH」

- **組織**
  - JST **研究開発戦略センター** (CRDS)
- **時期**
  - 2014**年**9**月**8**日開催**
- **内容**
  - **スーパーコンピューター「京」がネット上の言語をくまなく分析した上で，国民の多数は憲法改正に賛成しているとか，反対しているとかいうことスーパーコンピューターが発表したら，多くの日本人はそれが正しいと思わずにはいられないのではないか。これはある意味民主主義に対する挑戦とも言える。**
  - **高度な**BMI (Brain Machine Interface)**によって，全身麻痺の患者に繋がれたモニターから何らかの文章が表示されたとして，それが法律上その人の意思とみなして良いかどうか。これは法律の意思表示理論の根幹に関わる重大な問題。**
- **関連**URL
  - http://www.jst.go.jp/crds/pdf/2014/WR/CRDS-FY2014-WR-09.pdf

人工知能学会倫理委員会

- 組織
  - 主催者 松尾豊 (東京大学准教授)
  - 西田豊明 (京都大)，堀浩一 (東大)，武田英明 (NII)，長谷敏司 (SF作家)，塩野誠 (経営共創基盤)，服部宏充 (立命館)，江間有沙 (東大)，長倉克枝 (科学ライター)
  - オブザーバー：栗原聡 (電通大)，山川宏 (ドワンゴ)，松原仁 (はこだて未来大)
- 時期
  - 2014年12月発足

- 内容
  - AIが人々の幸せや夢の実現に寄与し，その発展が社会全体から歓迎されるように，研究者と社会の正しい関係を作る
  - 技術の現状認識
    - 研究の影響を予見。犯罪的AIや軍事AI，中毒性AIを予測
  - 指針
    - 万人のための人工知能
    - 人間の尊厳を守る
    - AIを作る・使う側の倫理
    - 社会がコントロールする仕組み：透明性，説明性，制御権
  - 職業
    - 短期的影響に配慮

  - 心を持つように見えるAIを作ってもよいのか？

人工知能学会 倫理指針　2017年2月28日公表

1. 人類への貢献 … 人類の平和、安全、福祉、公共の利益に貢献。
2. 法規制の遵守 … 研究開発に関わる法規制を遵守し、知的財産、他者との契約や合意を尊重。
3. 他者のプライバシーの尊重 … 他者のプライバシーを尊重し、関連する法規に則って個人情報を適正に取り扱う。
4. 公正性 … 常に公正さを持ち差別を行わない。
5. 安全性　… 安全性と制御可能性、機密性に留意する。
6. 誠実な振る舞い … 社会に対して誠実に信頼されるように振る舞う。
7. 社会に対する責任 … 人工知能がもたらす結果を検証し、潜在的危険性について警鐘をならす。悪用を防止する。
8. 社会との対話と自己研鑽 … 社会の多様性を理解し、高度な専門家として絶え間ない自己研鑽に努める。
9. 人工知能への倫理順守の要請 … 人工知能に倫理を順守させる。

AIR: Acceptable Intelligence with Responsibility
人工知能が浸透する社会を考える

- 組織
  - 主催者 江間有沙 (えま・ありさ, 科学技術社会論) 東京大学講師, 服部宏充 (AI) 立命館大学准教授ら
  - 科学よりの人文社会系を専門 (科学技術論, 技術哲学など) とする若手研究者らによる研究会。
    - 人文・社会科学者による①ELSI調査グループ
    - 人工知能研究者による②AI社会応用調査グループ
    - 両者を有機的に結びつける科学技術社会論や科学コミュニケーションを専門とする③対話基盤設計グループ
- 時期
  - 2014年秋発足
  - 2014年9月, 11月, 2015年2月, 7月, 9月, 11月 (公開)「人工知能が浸透する社会を考える」ワークショップ
- 関連URL
  - http://sig-air.org/

- **内容**
  - **人工知能と社会に関する研究の全体の動きを俯瞰して知見を蓄積**
  - **技術社会学，科学哲学者やAI研究者による研究会（勉強会，情報共有と議論）**
  - **「政府による干渉や産業による利益誘導に左右されない異分野間の対話・交流を促すための媒体や基盤をボトムアップで構築すること」**
  - **「対話を通して，人工知能の目指すべき共通アジェンダや社会の未来ビジョンを設計し，技術開発・実装時の新設計基準や規範・倫理・制度に関する価値観を提案すること」**

  - **参考論文**： Arisa EMA, Naonori AKIYA, Hirotaka OSAWA, Hiromitsu HATTORI, Shinya OIE, Ryutaro ICHISE, Nobutsugu KANZAKI, Minao KUKITA, Reina SAIJO, Otani TAKUSHI, Naoki MIYANO, and Yoshimi YASHIRO. Future Relations between Humans and Artificial Intelligence – A Stakeholder Opinion Survey in Japan -. *IEEE Technology and Society Magazine*, Vol. 35, No. 4, pp. 68-75, 2016.

- 内容
  - 議論，検討例
    - 人工知能に話しかけられて心理的に傷つくという被害を無くそうというルールを作ることが，「人の心を動かす・揺さぶる人工知能」を作る研究を放棄することになりかねない。
    - 嘘をつくコンピュータは，倫理的にどのように判断されるか。
    - 査読システムに倫理的審査規定を入れる可能性の検討
    - 人工知能の責任問題　Johnson & Noorman (応用倫理学者)
      - 原則　：人工的エージェントの行動の責任は人間が負うべき
      - 勧告1：人工的エージェントは人工物と社会的実践の構築物であって，それらの相互作用を通じて特定の課題を達成すべく編成されていると理解すべき
      - 勧告2：責任問題は人工的エージェント技術の開発初期段階で扱うべき
      - 勧告3：人工的エージェントが自律的であるという主張は，法的，明示的に特定されるべき
      - 勧告4：人工的エージェントの責任問題は責任の「実践」という観点から扱うのが一番良い

ロボット法研究会

- 組織
  - 情報ネットワーク学会
  - 主催者 新保史生(慶應義塾大学教授)
- 時期
  - 2016年5月発足，5-6月で3回のシンポジウム開催
- 内容
  - ロボット法原則
    - 人間第一の原則，命令服従の原則，秘密保持の原則，利用制限の原則，安全保護の原則，公開･透明性の原則，責任の原則

AI社会論研究会

- 組織
  - 井上　智洋 (駒澤大学 講師, 経済学)
  - 高橋　恒一 (全脳アーキテクチャ・イニシアチブ理事・副代表, 理化学研究所チームリーダー, 慶應義塾大学特任准教授, 生化学シミュレーション, バイオインフォマティクス)
  - 人文社会, 自然科学研究者, ビジネスマン, 官僚, 芸術家など多様。
- 時期
  - 2015年12月発足
- 内容
  - 人工知能が社会に与える影響
    - 哲学, 経済学, 法学, 政治学, 社会学
- 関連URL
  - http://aisocietymeeting.wix.com/ethics-of-ai

社会におけるAI研究会

- 社会システムと情報技術研究ウィーク (2016年3月1日(火)-4日(金))

  ■主　催■

  人工知能学会 知識ベースシステム研究会
  人工知能学会 社会におけるAI研究会
  情報処理学会 知能システム研究会
  電子情報通信学会 人工知能と知識処理研究会
  人工知能学会 データ指向構成マイニングとシミュレーション研究会

- 知能や社会・経済システムのモデル化・データマイニング・シミュレーション・ネットワーク分析，複雑系の解明と利用，環境・福祉・金融・デジタルコンテンツなどに関する社会システムの諸問題と情報技術など
- 関連URL
  - http://www.ni.is.uec.ac.jp/wssit16/

わたしの構想　人工知能の近未来
NIRA（総合研究開発機構）テーマインタビュー

- 組織
  - NIRA（総合研究開発機構）
- 時期
  - 2015年8月
- 内容
  - ロボットに代替されるホワイトカラー　新井紀子
  - 巨大プロジェクトより個人の能力を養え　小林雅一
  - AI開発競争で日本にも勝算ある　松尾豊
  - ウェアラブルを通じて人間の知能の強化を　塚本昌彦
  - 既に逆転は起きている　佐倉統
- 関連URL
  - http://www.nira.or.jp/outgoing/vision/entry/n150805_783.html

AIをどう見るか
“Edge Question”から探るAIイメージ
NIRA（総合研究開発機構）モノグラフ

- 組織
  - NIRA（総合研究開発機構）
- 時期
  - 2016年7月
- 内容
  - 公文俊平（NIRA総研上席客員研究員／多摩大学情報社会学研究所所長）、羽木千晴（NIRA総研研究コーディネーター・研究員）著
  - 世界的に著名な科学技術ウェブサイト“Edge.org”の寄稿誌である“Edge Question（エッジ・クエスチョン）”を取り上げ、そこで掲載されている192名の人々の人工知能についての論稿を分析し、世界におけるAIの議論の動向を把握し、今後日本においても議論すべき論点を明らかにする。
- 関連URL
  - http://www.nira.or.jp/outgoing/monograph/entry/n160726_820.html