

事務局説明資料

- 欧米の動向－開発原則に関連する議論を中心に－
 - ・ 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
 - ・ 米国ホワイトハウス報告書「人工知能の未来に備えて」
 - ・ 「米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<u>ロボット・AIの法的・倫理的問題</u>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<u>法制度、経済、社会等にもたらす長期的影響</u>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】 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～】 技術が人類の未来に与える影響に注目する視点から、<u>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>

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、自動操縦機とドローン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。
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. プライバシー及びパーソナルデータ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労環境の整備
18. セーフトイレットの整備
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
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。	— (政府全体のAI研究開発に係る方針であるため)
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。	— (政府全体の人員等、リソースに関することであるため)
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。	17.AIネットワーク化に対応した就労環境の整備
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。	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
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。	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
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。	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
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。	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
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	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

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

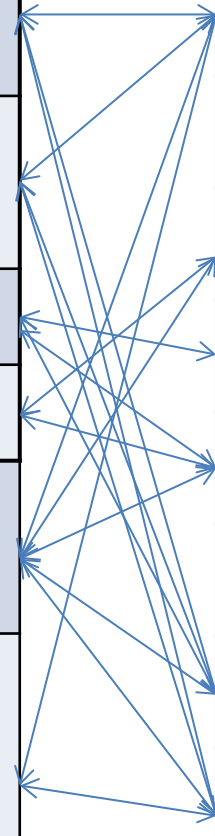
④セキュリティ確保の原則

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

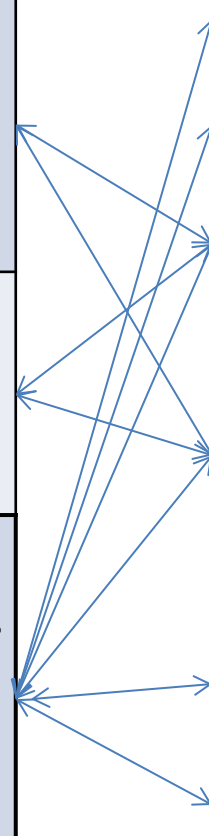
④セキュリティ確保の原則

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

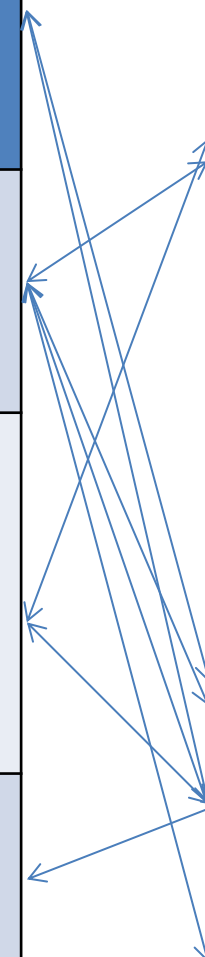
④セキュリティ確保の原則

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

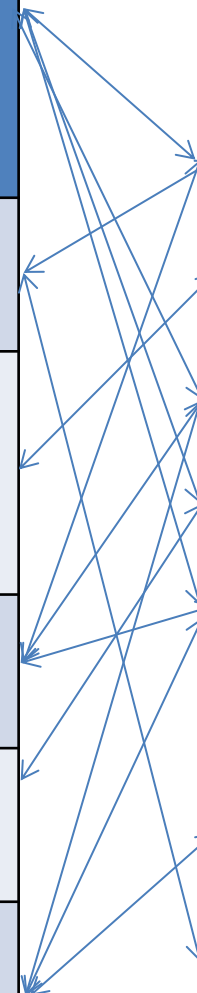
④セキュリティ確保の原則

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

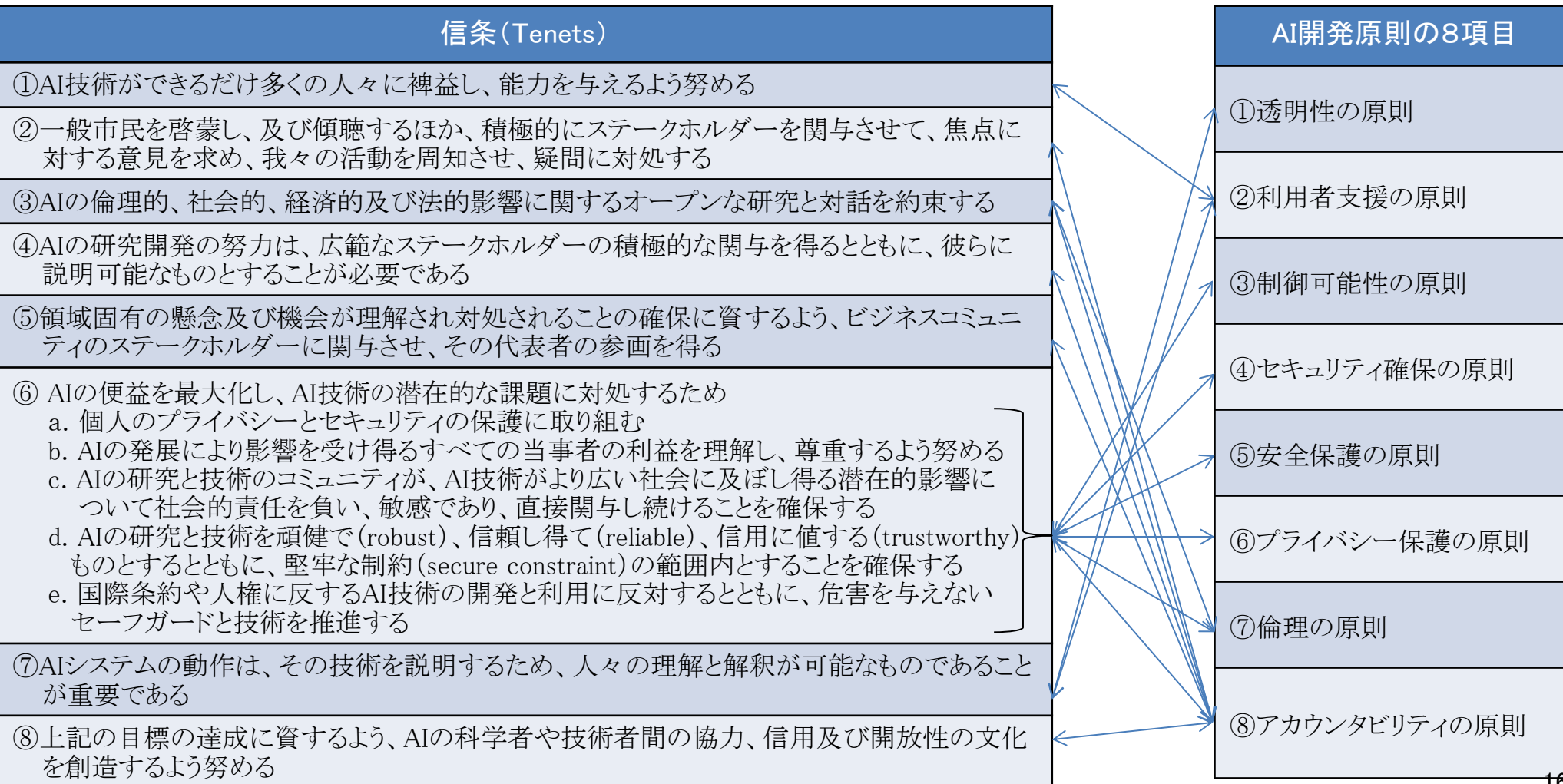
設立: 2016年9月28日(水)

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・法的問題を検討した上で、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

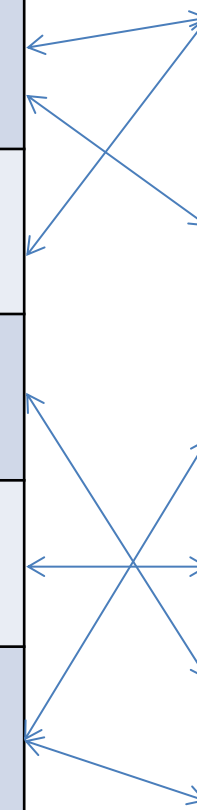
④セキュリティ確保の原則

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)とも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

