AIとデジタル政策

2023年7月12日

谷脇　康彦

人間中心のAI社会原則(2019年3月）

原則論の議論は既にかなり進んでいる。原則を社会において担保するための手段の議論が必要。

1. 人間中心の原則
2. 教育・リテラシーの原則
3. プライバシー確保の原則
4. **セキュリティ確保の原則**
5. **公正競争確保の原則**
6. **公平性、説明責任及び透明性の原則**
7. イノベーションの原則

▫**特定の企業にAIに関する資源が集中した場合**においても、**その支配的な地位を利用した不当なデータの収集や不公正な競争が行われる社会であってはならない**。

▫AIの設計思想の下において、人々がその人種、性別、国籍、年齢、政治的信念、宗教等の多様なバックグラウンドを理由に不当な差別をされることなく、**全ての人々が公平に扱われなければならない**。

▫AIを利用しているという事実、AIに利用されるデータの取得方法や使用方法、AIの動作結果の適切性を担保する仕組みなど、**用途や状況に応じた適切な説明**が得られなければならない。

▫AIを効率的かつ安心して社会実装するため、（中略）**倫理的側面、経済的側面など幅広い学問の確立及び発展が推進されなければならない**。

（注）AI開発利用原則については「現在、多くの国、団体、企業等において議論されていることから、我々は早急にオープンな議論を通じて国際的なコンセンサスを醸成し、非規制的で非拘束的な枠組みとして国際的に共有されることが重要である」と記載。

（出典）総合イノベーション戦略推進会議「人間中心のAI社会原則」（2019年3月）

AIとデジタル政策(1/3)

利用者保護の視点

AIの開発運用原則を踏まえ、原則を現実にどう適用するか。

- 例えば、「公平性、説明責任及び透明性の原則」に照らし、AIのブラックボックス化※を回避する手法の明確化が必要。
  ※ AIを使っているという事実、アルゴリズムの透明性、学習データの適法性（個人情報や著作権）等
- ブラックボックス化を回避する手段を何によって担保するか（例：法律、ガイドライン、技術実装等）。

個別化が進展する中、区別と差別を切り分けるアルゴリズムが必要。

- 個別化医療やロードプライシングには制度の合理性あり。通信の世界でも学割や家族割が存在。
- 「電気通信事業者は、電気通信役務の提供について、不当差別的な取り扱いをしてはならない。」（事業法第6条）
- 他社乗り換えの可能性が高い加入者の料金を低くすることはどうか。

（参考文献）OFCOM “Personalized Pricing for Communications : Making Data Work for Consumers” (Aug 2020)

AIを使わないというオプトアウトの仕組みをどう確保するか。

- AIが組み込まれた行政サービスにおいて「AIを利用しない」という選択肢はあり得るのか。
- 上記のサービスに個人情報を入力することで、別の時点で出力される可能性はないのか。
- 上記のような懸念を払拭するためんはどういう手段があり得るのか。

AIとデジタル政策(2/3)

情報通信政策の視点

AIの活用は情報の不確かさをもたらし、サイバー空間の信頼性を損ねることになるのか。

- Chat GPTの文章の生成は単語と単語の確率的つながりによるため、そこで生じる誤りがサイバー空間の信頼性を損なわないか。
- AIが学習するデータの確からしさを確保する仕組みが必要か。例えば、データサプライチェーンのような仕組みは必要か。

AIの普及はインテリジェンスの分散化をもたらすか。

- SDN/NFVのネットワーク実装化が進む中、通信リソースの分散管理をAIベースのオーケストレーションで行う状況へ進む。
- 「自動化された」通信ネットワークの信頼性におい、AIの信頼性や確実履行性などを担保することが必要になる。
- ハードソフト一体を基本とする法的枠組みの見直しについて検討が必要。

産業政策として、日本独自の大規模言語モデル（LLM）の整備が必要。

- 日本独自モデルが必要だとすると、そのメリットは何か。
- 上記のメリットをどのように活用していくことが可能か。

AIとデジタル政策(3/3)

安全保障政策の視点

サイバー攻撃にAIを活用しないとするノーム（規範）は成立するか。

- 脆弱性発見やマルウェア開発にAIを使うことで「自動戦争」につながるのではないか。
- 安全保障の観点からみて、サイバー空間における国際ルールについての合意は成立していない。
- 有志国によるノームに関する合意を先行させることの意味は何か。---抑止戦略の一環か。
- 「国家の関与が疑われる攻撃」には効果がないか。

武器としてのAI活用の是非に関する議論が必要。

- 米国国務省による「人工知能及び自律性の責任ある軍事利用に関する政治宣言」は単なる「政治宣言」なのか。
- 有志国によるノームの確認（共有、合意）はどこまで効果があるのか。

認知戦におけるAIの活用についてどう考えるか。

- 防衛力整備計画（2022年12月閣議決定）では、「人工知能(AI)を活用した公開情報の自動収集・分析機能の整備、各国等による情報発信の真偽を見極めるためのSNS上の情報等を自動収集する機能の整備、（中略）を行う」と明示。
- 年内に開始される「G7広島プロセス」では「偽情報を含む外国からの情報操作への対応」が検討課題として例示。
- 日本としての基本方針を速やかに定める必要。

人工知能及び自律性の責任ある軍事利用に関する政治宣言

(US DoS, “Political Declaration on Responsible Use of Artificial Intelligence and Autonomy”, Feb 2023)

”軍事領域における責任あるAIに関する会議[REAIMSummit]”(2023年2月＠ハーグ）において、米国務省が提案。軍事分野におけるAIの開発・配備・使用に際し、自主的に遵守し、そのコミットメントをオープンにすることを提案。

- 軍事AI(military AI capabilties)が国際法（特に国際人道法）の義務に合致した形でのみ使用されることを保証するため、法的審査 (legal reviews)などの効果的な措置を実施
- 核兵器の使用に関する重要な情報付与や判断は、人間による管理と関与を維持
- 兵器システムを含むすべての軍事AIの開発・配備は、高官(senior offioials）が監督
- 軍事AIの責任ある設計・開発・配備・使用に関する原則を採択・公表
- 軍事AIの開発・配備・使用は、適切なレベルの職員が判断
- 軍事AIの意図しない偏り(unintended bias)を最小化する対策を実施
- 監査可能(auditable)な方法やデータソース、設計手順、文書によって軍事AIを開発
- 軍事AIを使用する職員、及び使用を承認する職員は、その能力と限界を十分に理解し、その使用について適切な判断を行うことができるよう訓練。
- 軍事AIの安全性・セキュリティ・有効性についは、ライフサイクル全体にわたって厳格なテストと保証の対象とし、自己学習による軍事AIは、重要な安全機能が低下していないことを確認する監視プロセスに従うこと。
- 意図しない結果を検出・回避・解除できるよう設計する等のセーフガードを導入。
- 軍事AIの開発・配備・使用に関する議論を継続し、他の適切なコミットメントを見出すよう努力。

(出典) https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/

G7によるAI原則の検討
（2023年5月20日、広島G7首脳コミュニケ）

✔国や文化を超えてますます顕著になっているAIの機会及び課題について**直ちに評価する必要性**を認識し、人工知能グローバルパートナーシップ（GPAI）
が実践的なプロジェクトを実施することを奨励。

✔生成AIに関する議論のために、包摂的な方法で、OECD及びGPAIと協力しつつ、**G7の作業部会を通じた、広島AIプロセスを年内に創設**する。

✔これらの議論は、**ガバナンス、著作権を含む知的財産の保護、透明性の促進、偽情報を含む外国からの情報操作への対応、これらの技術の責任ある活用**といったテーマを含み得る。

AIを巡る議論において基本となる問題意識

- ✔「まず使う」が議論の前提
- ✔人権保護や安全保障に関する議論は何よりも重要
- ✔規制論は、AIがmoving targetであることに注意
- ✔ハードローありきではなく、ソフトローや技術実装を含めて柔軟に考える必要
- ✔価値観に関わる問題もあり、広範なステークホルダーを巻き込む議論を