情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

2018年12月18日
AIネットワーク社会推進会議
AIガバナンス検討会

AIとセキュリティ：政治参加の観点から

湯淺　墾道
（情報セキュリティ大学院大学）

AIと政治参加に関する論点

情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

- 福田雅樹・林秀弥・成原慧編『AIがつなげる社会』（弘文堂、2017年）湯淺墾道「AIネットワークと政治参加・政策決定」
  - Vote muchとAI、「AIペン」
  - マルウェア感染
- 山本龍彦編『AIと憲法』（日本経済出版社、2018年）工藤郁子「AIと選挙制度」
  - デモクラシーの機能不全、アドテク、ナッジ
  - 「支配されるという特権」→フェイクニュースを信じる自由(?)

- 統治の領域においては、長い歴史をへて民主主義が発達したが、それによって自然人だけが平等に政治に参加すべきであり、それ以外のものが政治には参加すべきではないという考えが確立
- 自然人よりもAIのほうが合理的・客観的な判断が行えるようになったとしても、統治に関する過程から非理性的な判断を排除することはできない
- 経済や社会がグローバル化した今日にあっても国内の政治を国際社会に開放しようとすることには反対の声、外国の技術によって開発されたAIやロボットが日本の統治に「参入」することへの障壁

- 国民と国家との中間にさまざまな団体（中間団体）が存在することは非民主的であるとするフランス革命以降の民主主義観
  - 政治に参加するのはあくまでも自然人であるべき（cf. 法人の選挙権と政治資金問題）
  - AIが自然人である有権者と国家との間に介在することは民主主義をゆがめる恐れ
  - AIによる判断に有権者が依存することは、市民が政治社会に能動的に参加するという市民社会の原理を否定する恐れ

■大衆社会論

- ●AIは衆愚政治や大衆迎合政治の危険を防止?
  - ◆大衆に理性を付加すれば、危険性は減少
  - ◆大衆が政治的判断を行う際に、AIという「理性」の助けを借りるようになれば、非合理的な判断や意思決定を行わないようになる可能性
- ●AIは衆愚政治や大衆迎合政治の危険を加速?
  - ◆大衆は欲望と権利意識のみを持った凡庸な人間の集合体(オルテガ)、せっかくAIが合理的・理性的な判断をしたとしても大衆はそれには従わない
  - ◆大衆の気質をAIが先取りして判断した結果、AIは、大衆の欲望と権利意識をくすぐる判断結果提示の恐れ

フェイクニュースとサイバーセキュリティ

- 谷脇康彦『サイバーセキュリティ』（岩波新書、2018年）148-149頁
  - 「ネット上の偽（フェイク）ニュースをめぐる議論がますます深刻化しています。（中略）こうした情報資産のCIAを確保するという視点からみると、情報の完全性が悪意をもって操作される、つまり偽情報が拡散される状況はサイバーセキュリティが確保されていない状況といえます。このため、情報の完全性を破り、偽情報を意図的に流通させる行為も広い意味でサイバー攻撃であり、その対策に関する議論もサイバーセキュリティ政策の射程に入ってくるものととらえることができます。」

フェイクニュースと政治過程

■ジュリアン・キング(Julian King)EU委員

- 2018年6月21日スピーチ
- 選挙に対する干渉：第1は制度、第2は有権者の行動に基づくもの
- 「私の見解では、これ（注：第2のカテゴリーのこと）には3つの形式があります。キャンペーン中の重要なポイントで情報を明らかにすることによって世論を変えるように設計されたハッキングとリーク。世論を動揺させ選挙結果に影響を与えるためのフェイクニュースの使用。そして、ユーザーの個人的な特性データから導き出された心理測定に基づいて特定ユーザーを標的としてメッセージを恣意的に送ること、ケンブリッジ・アナリティカのようなやり方です。これらの3つはサイバーを利用した選挙操作の別々の形式ですが、すべてが特定の方向に結果を歪めるように設計されています。」

- 「（有権者の）行動に基づく脅威に対抗するために、欧州委員会は4月に、ソーシャルメディアが民主主義に対する武器にならないようにするためインターネットプラットフォームに対して期待することを含めて、虚偽情報流布や行動の操作に対してさまざまな措置を提案しました。大規模な虚偽情報流布という武器は、現代の大量破壊兵器となりえます。」
- https://ec.europa.eu/commission/commissioners/2014-2019/king/announcements/commissioner-kings-speech-aspen-institute-protecting-western-democracies-manipulation-and_en

フェイクニュース・虚偽情報流布とAI

- 新米国機構(New America)
  - 選挙のサイバーセキュリティの3段階
    - NEW AMERICA, CYBERSECURITY INITIATIVE: HOW TO THINK ABOUT ELECTION CYBERSECURITY?, 6-7 (2018).
  - 有権者の民意形成への介入と世論誘導
  - 選挙管理機関へのサイバー攻撃や電子投票機へのサイバー攻撃等で直接的に選挙結果を操作
  - 選挙管理機関のウェブサイトへの攻撃や選挙に関するニュースサイトへの攻撃等によって選挙に混乱

- AIによるフェイクニュースの生成・発信
- AIによるフェイクニュースの効果的な流布（流布先・流布方法分析を含む）
- AIによる個人の政治的意思形成や投票行動の代理・代行

- AIによるフェイクニュースの検知
- AIによるフェイクニュースのアトリビューション、流布対象者の検知と通知
- AIによって誘導された政治的意思や投票行動の検知

- Transcript of Mark Zuckerberg's hearing, Commerce and Judiciary committees U.S. Senate, April 10, 2018.
  - FLAKE: There are obviously limits, you know, native speakers that you can hire or people that have eyes on the page. Artificial intelligence is going to have to take the bulk of this. How — how much are you investing in working on — on that tool to — to do what, really, we don't have or can't hire enough people to do?
  - ZUCKERBERG: Senator, I think you're absolutely right that over the long term**, building A.I. tools is going to be the scalable way to identify and root out most of this harmful content**. We're investing a lot in doing that, as well as scaling up the number of people who are doing content review.

- THUNE: Well — As we discussed in my office yesterday, the line between legitimate political discourse and hate speech can sometimes be hard to identify, and especially when you're relying on artificial intelligence and other technologies for the initial discovery.
- ZUCKERBERG: So, from the beginning of the company in 2004 — I started in my dorm room; it was me and my roommate. We didn't have A.I. technology that could look at the content that people were sharing. So — so we basically had to enforce our content policies reactively.
  People could share what they wanted, and then, if someone in the community found it to be offensive or against our policies, they'd flag it for us, and we'd look at it reactively. Now, increasingly, **we're developing A.I. tools that can identify certain classes of bad activity proactively** and flag it for our team at Facebook.

EUにおける規制の試み

■2018年4月26日　虚偽情報に対する「多元的な対応(multi-dimensional approach)」提案

◆http://europa.eu/rapid/press-release_IP-18-3370_en.htm

- 虚偽情報に関する行動規範
- 各ファクトチェッカーの独立ネットワークの構築
- 虚偽情報に対するセキュアなヨーロッパのオンライン・プラットフォーム
- メディアリテラシーの強化
- 加盟国に対する選挙のレジリエンス強化支援
- 質の高い多様な情報の支援
- 戦略的なコミュニケーション政策の調整

■共通の行動規範策定を要求

1　スポンサーがついているコンテンツ、特に政治広告についての透明性を確保すること、また政治広告のターゲティングオプションを制限し、虚偽情報の提供者の利得を削減すること。

2　アルゴリズムの機能と第三者による検証を可能にすることについて、明確に説明すること。

3　他の視点を代表する異なるニュースソースをユーザーが発見してアクセスしやすいようにすること。

4　フェイクアカウントの特定と閉鎖対策、自動ボットの問題への取組を開始すること。

5　ファクトチェッカー、研究者、および公的機関がオンラインの虚偽情報を継続的に監視できるようにすること。

- EU Code of Practice on Disinformation (2018/9/26)
  - The signatories recognise the ongoing legislative work to develop standards for transparency about the main parameters of ranking included in the draft Platform to Business Regulation as well as the work being carried out by the EU Artificial Intelligence Expert Group as well as the EU consumer acquis.

■High-Level Expert Group on Artificial Intelligence (AI HLEG)

●52名の専門家任命

- ◆Advise the Commission on next steps addressing AI-related mid to long-term challenges and opportunities through recommendations which will feed into the policy development process, the legislative evaluation process and the development of a next-generation digital strategy.
- ◆Propose to the Commission draft AI ethics guidelines, covering issues such as fairness, safety, transparency, the future of work, democracy and more broadly the impact on the application of the Charter of Fundamental Rights, including privacy and personal data protection, dignity, consumer protection and non-discrimination
- ◆Support the Commission on further engagement and outreach mechanisms to interact with a broader set of stakeholders in the context of the AI Alliance, share information and gather their input on the group's and the Commission's work.

- ENISA (= European Network and Information Security Agency)オピニオン・ペーパー（2018年4月）
  - https://www.enisa.europa.eu/publications/enisa-position-papers-and-opinions/fake-news/
  - ネットワーク及び情報セキュリティに関する一般的な対策として、選挙に関するシステムや機器類を重要インフラに含めることを提案
  - オンライン虚偽情報流布と、選挙に関係するシステムについて勧告
  - 経済的対策と技術的対策を平行実施するべき

■A.　人工知能

- ●AIアルゴリズムの利用を、オンライン詐称キャンペーンやスクラップやスパムなどのオンラインプラットフォームの誤用の検出を支援するために、導入するべきである。
  これらのアルゴリズムの出力は、何らかの措置を講じる前に、人間が確認する必要がある。
  - ◆※詳細については湯淺墾道「2019年欧州議会議員選挙とインターネット・SNS(2)」選挙2018年9月号1～4頁

AIセキュリティの法規制の困難性

情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

- PL法・ネットワーク経由（リモート）の場合
- カリフォルニア州IoTセキュリティ法
  - 2018年9月28日知事署名、2020年1月1日施行
  - ユーザーの選択によって追加される無連携のサードパーティーのソフトウェアまたはアプリケーションに関連する接続される装置の製造者に義務を課すものとは解釈されない
  - ソフトウェアもしくはアプリケーションの購入もしくはダウンロード手段の提供者に対し、本法の遵守の審査または執行を義務づけるものとは解釈されない

■カリフォルニア州IoTセキュリティ法仮訳

【接続される機器のセキュリティ】

1798.91.04条

(a) 接続される機器の製造者は、当該機器に合理的なセキュリティ機能または以下のすべてを備えているものとする。

(1) 機器の性質及び機能に適するもの

(2) 収集し、包有し、又は発信することができる情報にふさわしいもの

(3) 機器および機器に含まれる情報を、不正アクセス、破壊、使用、改変または開示から保護するように設計したもの

(b)接続される装置がローカルエリアネットワークの外部に認証手段を備えている場合、(a)項の要件を全て満たすことを条件として、以下のいずれかの要件が満たされている場合は小項目(a)に基づく合理的なセキュリティ機能とみなされるものとする。

(1) あらかじめプログラムされたパスワードは、製造された各機器に固有のものであること

(2) 当該機器は、初回アクセスが許可される前にユーザーが新しい認証手段を生成しなければならないセキュリティ機能を備えていること

1798.91.05条

本法の目的に照らして、次の用語は、次の意味を有するものとする。

(a)「認証」とは、情報システム内のリソースにアクセスするユーザー、プロセスまたは装置の権限を検証する方法を意味するものとする。

(b)「接続機器」とは、直接又は間接にインターネットに接続することができ、かつ、インターネットプロトコルアドレス又はブルートゥースアドレスを割り当てられた機器その他の物理オブジェクトをいうものとする。

(c)「製造業者」とは、カリフォルニアにおいて販売または販売の申し出がなされている接続機器を製造する者、または他人と契約して当該他人のために製造する者を意味するものとする。本項の目的に照らし、他人に代わって製造することに係る他人との契約は、接続される装置の購入、または接続される装置の購入およびブランド設定のみの契約を含まない。

(d)「セキュリティ機能」とは、装置に対してセキュリティを提供するように設計された装置の特徴を意味するものとする。

(e)「不正アクセス、破壊、使用、変更又は開示」とは、消費者が許可していないアクセス、破壊、使用、変更又は開示をいうものとする。

1798.91.06条

(a) 本法は、接続される装置にユーザーの選択によって追加される無連携のサードパーティーのソフトウェアまたはアプリケーションに関連する接続される装置の製造者に義務を課すものとは解釈されないものとする。

(b) 本法は、電子ストア、ゲートウェイ、市場またはその他のソフトウェアもしくはアプリケーションの購入もしくはダウンロード手段の提供者に対し、本法の遵守の審査または執行を義務づけるものとは解釈されないものとする。

(c) 本法は、接続される装置の製造者に対し、ユーザーの裁量で装置上で動作するソフトウェアまたはファームウェアを修正する能力を含めて、ユーザーが接続される装置に対して完全な制御を行うのを防止する義務を課すものとは解釈されないものとする。

(d) 本法は、その機能性が、その執行権限に従って連邦政府機関により公布された連邦法、規則またはガイダンスに基づくセキュリティ要件の対象となる接続機器には、適用されない。

(e) 本法は、民事訴訟を提起する権利を付与するものと解釈してはならない。司法長官、市法務官、郡法務官又は地方法務官は、この法律を執行する排他的権限を有する。

(f) 本法により課される義務および義務は、他の法律に基づいて課されるその他の義務または義務に重複するものであり、いずれの当事者も免他の法律に基づいて課される義務から除されるとは解釈されないものとする。

(g) 本法は、法律または管轄裁判所の命令により権限を付与された製造業者から接続機器情報を取得する法執行機関の権限を制限するものと解釈されないものとする。

(h) 1996年連邦医療保険のポータビリティと説明責任に関する法律(HIPAA法)または医療情報の機密保持法(第1章2.6条(第56項以降)の適用対象となる企業、医療提供者、ビジネスアソシエイト、医療サービス計画、請負業者、雇用主、またはその他の個人は、これらの法律により規制される活動に関して、本法の適用を受けない。

(i) 本法は、2020年1月1日に発効する。