2022 年 2 月 7 日

米国の動向から考えるディープフェイクへの対応

中曽根平和研究所
主任研究員
横田　佳祐

1. はじめに

ディープフェイク(Deep fakes)とは、人工知能技術によって作られた、あたかも本物のような外観の画像や音声、映像などの贋作を指す。

ディープフェイクがいかに精巧なものかを知るには、百聞は一見に如かずということで、実際に自分の目で見てみるのが一番良い。例えば、動画サイトにはオバマ元米国大統領[1]やエリザベス英国女王[2]、プーチン露大統領[3]、金正恩朝鮮労働党総書記[4]を題材として教育・啓蒙目的で作成されたディープフェイクが掲載されており、ディープフェイクへの注意を喚起している。

ディープフェイク自体は悪いものではない。例えば、表現の自由の一形式としてのパロディーが一例である。加えて、稀な疾病の画像データを生成することで、画像診断を行う医師や人工知能にとって必要な画像データを提供したり、既存の絵画データから新たな芸術作品を生み出したりすることもできる。問題は、ディープフェイク生成技術を使う者の悪意である。

2. 日本や海外で起きた具体的な問題

国内外で共通して見られるディープフェイクの使用目的はポルノ画像である。それ以外には、政治的な意図が窺える。例えば、日本において、ポルノ以外で悪意を持ってディープフェイクが使われた事例としては、2021 年 2 月の出来事が挙げられる。2021 年 2 月 14 日に福島と宮城で震度 6 強の地震が発生し、加藤官房長官（当時）が会見を行った。その会見画像が何者かに改変され、地震という有事に笑顔で応対しているかのような画像が SNS に投稿・拡散された。改変された画像を事実と思いこんだ SNS ユーザーからは非難の声が上がり、一時 SNS では騒然となった。本件では、作成や投稿した者は不明のままであったが、改変画像が会見の 30 分後には投稿され、改変の痕跡も少ない非常に精巧なディープフェイクであった点が関係者を驚かせた[5]。

海外では、例えば 2016 年の米国大統領選挙時に民主党候補であったヒラリー・クリントン氏の信用を貶めるような性的画像のディープフェイクが作成、拡散された[6]。また、2019 年にマレーシアではアズミン・アリ経済相（当時）が同性愛者であることを示唆する動画が拡散し、一時マレーシア政界が騒然となった[7]（当局は当該動画をディープフェイクと断定して関係者を処罰）。2020 年には、ベルギーにて環境保護団体が当時のベルギー首相の演説動画を改変し、あたかも環境破壊と

[1] オバマ元米国大統領のディープフェイク　https://www.youtube.com/watch?v=cQ54GDm1eL0
[2] エリザベス英国女王のディープフェイク　https://www.youtube.com/watch?v=IvY-Abd2FfM
[3] プーチン露大統領のディープフェイク　https://www.youtube.com/watch?v=sbFHhpYU15w
[4] 金正恩朝鮮労働党総書記のディープフェイク　https://www.youtube.com/watch?v=ERQlaJ_czHU
[5] 読売新聞オンライン 2021 年 4 月 21 日付記事　https://www.yomiuri.co.jp/national/20210412-OYT1T50147/
[6] NBC News 2018 年 4 月 18 日付記事　https://www.nbcnews.com/tech/security/russia-linked-account-pushed-fake-hillary-clinton-sex-video-n864871
[7] Wired UK 2019 年 10 月 20 日付記事　https://www.wired.co.uk/article/how-to-spot-deepfake-video

Nakasone Peace Institute
※本稿での考えや意見は執筆者個人（或はグループ）のもので当団体のものではありません。

新型コロナウイルスが関係しているかのような偽の演説動画を作成し、拡散した[8]。一方で、ディープフェイクではないかとの疑念も混乱をもたらす。2019 年にアフリカ中部のガボン共和国にて、動静が明らかになっていなかったアリー・ボンゴ・オンディンバ大統領が年初の挨拶動画を投稿したところ、動画をディープフェイクだと信じた軍の一部がクーデター未遂を引き起こした事件もあった[9]。

今後、日本で問題になりうる例としては、まず、自衛隊が戦争犯罪を行ったとする映像を作成し、拡散させ、ある国の国民感情を激化させたり、テロリスト募集など日本への敵対行為の誘因に利用したりすることが考えられる。加えて、米国で 2016 年に起きたような、他国等が選挙に干渉し、選挙候補者の印象を操作することも考えられる。また、精巧な画像や音声により、顔認証や音声認証が突破され個人等の金融資産が被害に遭う可能性や、リモート会議でなりすましが侵入し、機微な事項を盗み聞きされる可能性などもある。実際、ディープフェイクによって本人そっくりに合成された音声を使った巨額の詐欺事件が英国で発生している[10]。その他、本物の画像等にディープフェイクとのレッテルを貼り、都合の悪い情報を切り捨てようとする“Liar's Dividend[11]”の問題もある。

2016 年大統領選挙で共和党候補の一人だったマルコ・ルビオ氏は 2018 年 7 月 24 日に行われたヘリテージ財団での演説でこう語っている[12]。「昔は、米国を脅そうと思ったら、空母 10 隻、核兵器、長距離ミサイルが必要だった。今日必要なのは、インターネットシステム、銀行システム、電力網、インフラへのアクセスだけだ。さらに、非常にリアルなフェイク動画を生み出す能力さえあれば、選挙を蝕み、米国内を途方もない危機に陥れ、米国を一層弱体化させることができる。」

ディープフェイクに無関心でいることは、国や企業、個人にとって致命的になりうる。

3. ディープフェイクの技術的説明

ディープフェイクはどのように生み出されるのだろうか。根幹となる技術は敵対的生成ネットワーク(GAN)と呼ばれる。このネットワークでは、二つのニューラルネットワークを競わせるため、「敵対的」と呼ばれる。一方は生成器（Generator）と呼ばれ、ランダムなノイズの確率変数から偽物のデータを作成する。他方は識別器(Discriminator)と言い、生成器が作成した偽物のデータと本物のデータを与えられた上で、それぞれ本物か偽物かを判断する。偽物のデータと見抜かれた生成器は、識別器が偽物と分からないようにより精巧な偽データを作成していき、最終的に本物のような偽のデータを生み出すことになる。

より詳細に説明しよう。データには予め本物であれば「１」、偽物であれば「０」という２値で表される正解が与えられている。識別器は本物のデータを与えられた時は１を、偽物のデータを与えられた時は０を出力するように訓練する。識別を間違えた場合には、識別器に誤差を逆伝播（フィードバック）し、学習させていく。他方、生成器については、識別器が偽物のデータで１を出力するように学習させていく。生成器の学習に必要な誤差については、識別器を騙せたかどうかの結果（出力）を知る必要があるので、識別器を経由して逆伝播される。なお、識別器の学習中には生成器の最適化は行われず、逆に生成器の学習中には識別器の最適化は行われない。これらを繰り返し、生成器と識別器が切磋琢磨していくことで、生成器は本物のような偽物のデータを作り出すことができるようになる。学習が終われば、識別器は不要になり、生成器によってディープフェイクが生み出されるようになる。

---

[8] Wired UK 2020 年 12 月 20 日付記事　https://www.wired.co.uk/article/deepfakes-porn-politics

[9] Financial Times 2019 年 10 月 24 日付記事　https://www.ft.com/content/4bf4277c-f527-11e9-a79c-bc9acae3b654

[10] The Wall street Journal 2019 年 8 月 30 日付記事　https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

[11] Bobby Chesney & Danielle Citron “Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security” https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954

[12] Daily Signal　2018 年 7 月 18 日付記事　https://www.dailysignal.com/2018/07/19/deep-fake-technology-is-a-threat-to-national-security-politics-and-the-media-rubio-says/

※本稿での考えや意見は執筆者個人（或はグループ）のもので当団体のものではありません。

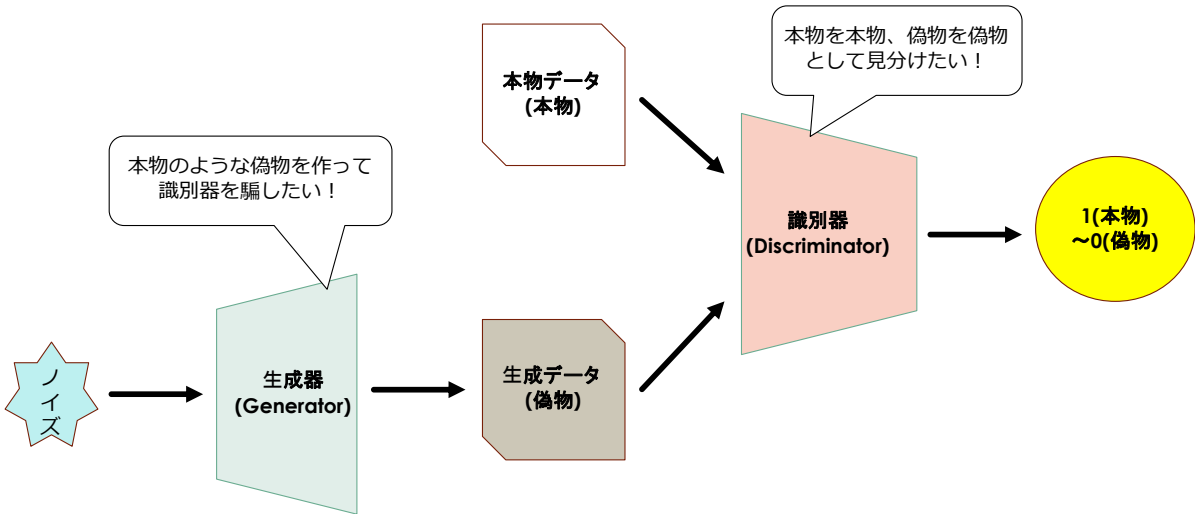

図１　敵対的生成ネットワーク(GAN)の全体像[13]

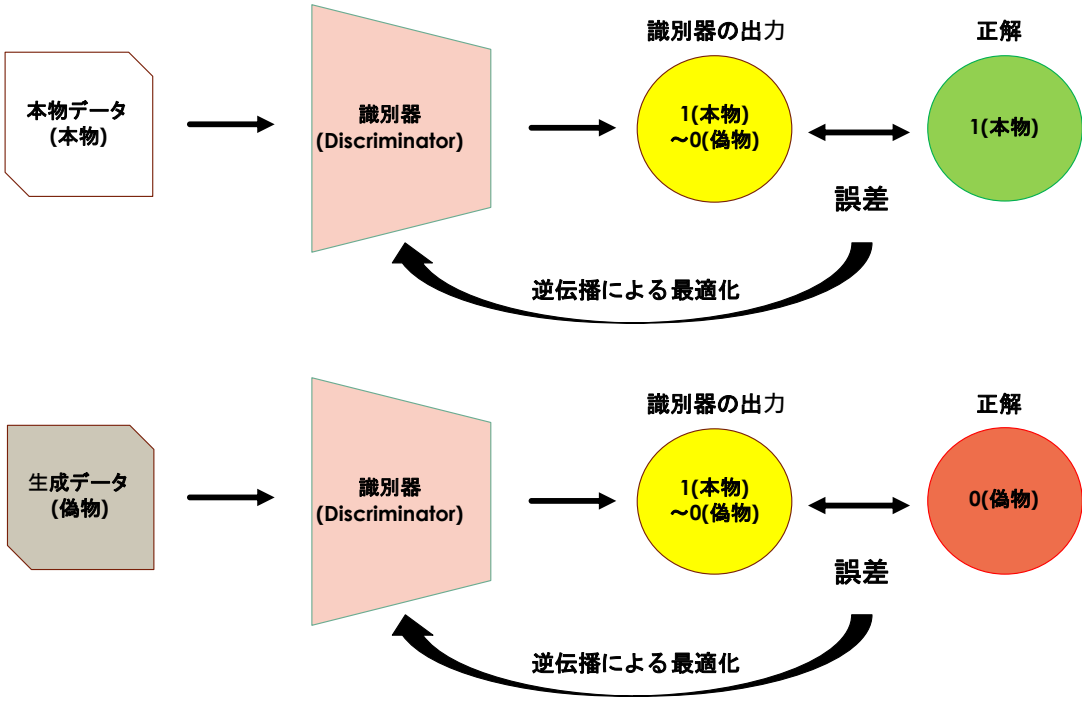

図２　識別器の学習[14]

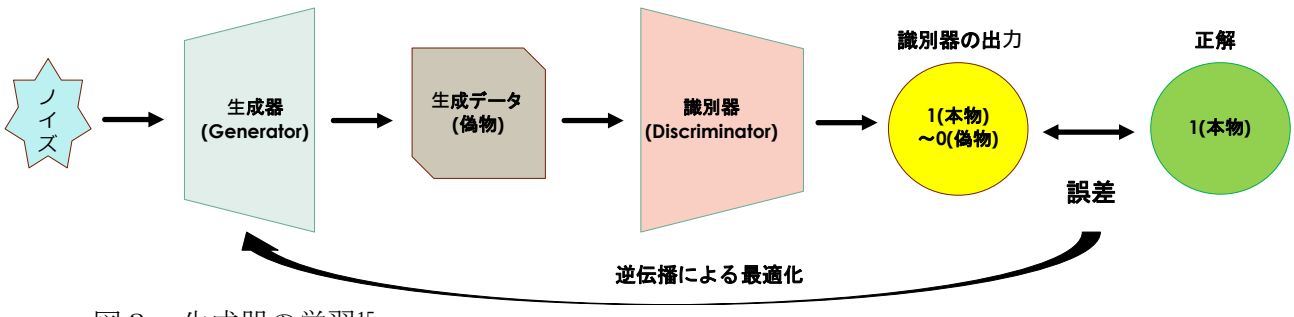

図３　生成器の学習[15]

[13] 毛利拓哉・大郷友海・嶋田宏樹・大政孝光・むぎたろう・寅蔵・もちまる著『GAN ディープラーニング実装ハンドブック』pp.53(2021 年 2 月 15 日株式会社秀和システム発行)を元に筆者作成
[14] 同上 pp.54 を元に筆者作成
[15] 同上 pp.55 を元に筆者作成

※本稿での考えや意見は執筆者個人(或はグループ)のもので当団体のものではありません。

4. 米国の動向

米国では、2016 年の大統領選挙において、ディープフェイク等を用いたロシア等外国勢力の選挙介入があり、2020 年大統領選挙においても同様の介入があることを警戒していた。2020 年大統領選挙では、選挙過程や結果について国民の信頼を損ねようとする動きがロシアやイランに見られたものの、投票者登録、投票、集計、結果報告などへの技術的干渉は見られなかったとの結論が米国国家情報長官室によりまとめられている[16]。これは、米国がディープフェイクに対して適切に対応をとった結果だと考えられる。

まず、米国における安全保障政策としてディープフェイクがどう扱われるかを見るために、国防授権法について紹介したい。米国国防授権法とは、米国の国防予算の大枠を決めるため、毎年連邦議会に提出される法案である。ディープフェイクという文言が最初に登場するのは、2019 年 12 月 20 日に制定された 2020 年度国防授権法である[17]。この中では、ディープフェイクが米国の国家安全保障に与える潜在的影響や、外国政府が偽情報の拡散など悪意を持った活動のためにディープフェイクを使用した実例やその可能性について国家情報長官が調査し、連邦議会の情報委員会に対して報告を義務付けた。また、賞金付きの大会を創設し、ディープフェイクを自動的に検出する技術の研究、開発、商業化を刺激するプログラムを国家情報長官が実行するよう求めた[18]。

2021 年 1 月 1 日に制定された 2021 年度国防授権法[19]では、外国政府及び非国家主体がディープフェイクを作成又は使用することによって、軍人とその家族にもたらされる脅威に関する情報評価として、ディープフェイク技術の成熟度及び情報戦におけるディープフェイクを使用した実例やその可能性について国防長官が調査し、連邦議会の下院及び上院の軍事委員会に対して報告することを求めた。

2021 年 12 月 27 日制定された 2022 年度国防授権法では、ディープフェイクとの文言は消えたが、包摂されうる概念として Disinformation（偽情報）もしくは Information warfare（情報戦）については、ロシアや中華人民共和国による活動として記載されている。

ディープフェイクに対する米国の具体的対応については、①技術的手段による対応、②法的手段による対応、③国民のリテラシー向上の取組、の大きく３つに分けられる。

一つ目の技術的手段による対応としては、画像等の改変の検知や電子透かし技術の開発が挙げられる。連邦政府レベルでは、国防総省国防高等研究計画局（DARPA）が 2016 年から 2020 年にかけて、Media Forensic(Medifor)[20]と 2021 年から Semantic Forensic（SemaFor[21]）という二つの取組みを行っている。MediFor はピクセル単位での分析や、物理法則、他の情報源との組み合わせにより画像や音声など視聴覚データの不整合を検知するものである。SemaFor は、MediFor で培ったディープフェイク検出技術に加え、情報源とされる出典の信頼性や、改変の意図が悪意か否かについて明らかにすることを試みるプログラムであり、Google や NVIDIA、ニューヨーク大学などが参画している。

加えて、2020 年 12 月 23 日に制定された連邦法「敵対的生成ネットワークの出力識別に関する法律」[22]では、情報の真正性の検証や改変の検知、電子透かしシステムなど技術的識別手段に関する

---

16 DNI report Foreign Threats to the 2020 US Federal elections
https://www.dni.gov/files/ODNI/documents/assessments/ICA-declass-16MAR21.pdf
17 S. 1790 (116th): National Defense Authorization Act for Fiscal Year 2020
https://www.govtrack.us/congress/bills/116/s1790/text
18 Kaggle
https://www.kaggle.com/c/deepfake-detection-challenge
19 H.R. 6395 (116th): National Defense Authorization Act for Fiscal Year 2021
https://www.govtrack.us/congress/bills/116/hr6395/text
20 DARPA https://www.darpa.mil/program/media-forensics
21 DARPA https://www.darpa.mil/program/semantic-forensics
22 H.R. 4355 (116th): Identifying Outputs of Generative Adversarial Networks Act
https://www.govtrack.us/congress/bills/116/hr4355/text

※本稿での考えや意見は執筆者個人（或はグループ）のもので当団体のものではありません。

研究について全米科学財団(NSF)の支援を求めるとともに、技術的識別手段の開発促進に必要な測定及び標準開発の研究について国立標準技術研究所(NIST)の支援なども求めている。

この他、2021 年 7 月 29 日に提出された「ディープフェイク・タスクフォース法案」[23]では、デジタルコンテンツの出所を特定するための標準策定と技術開発の実現可能性・課題の調査や、政策変更を提案することなどを目的とする組織の設立が模索されている。このような技術的手段による取組みに対しては、検知技術の発達とともに、検知を掻い潜る技術が生み出され、いたちごっことなる可能性が高いという課題がある。

次に、法的手段による対応としては、ディープフェイク自体の違法化や、ディープフェイクが拡散される場を提供する SNS 事業者に適切な対応を義務付けることが挙げられる。前者については、既に一部の州において特定の目的に限ってディープフェイクを規制する法律が制定されている。バージニア州、メリーランド州、カルフォルニア州、ジョージア州、ニューヨーク州、ハワイ州はポルノ規制目的で規制を設けており、ニュージャージー州やフロリダ州では規制を検討している。

また、選挙妨害を阻止する目的で法律による規制が存在する。例えばテキサス州では、候補者を中傷したり選挙結果に影響を与えたりする目的で、選挙から 30 日以内にディープフェイク動画を作成または配信することを犯罪と規定している。一方、選挙妨害規制目的での制定を目指したものの、廃案となった州としてイリノイ州やワシントン州、メイン州などがある。これは、米国憲法修正第一条で保障される言論の自由、特に政治的言論の自由との均衡や規制の実効性への義義から廃案となったものと考えられる。

上記のような理由のためか、規制を試みる法案の提出は過去見られたが、2022 年２月現在、連邦政府レベルでは、いかなる目的でもディープフェイクを規制する法律は存在しない。

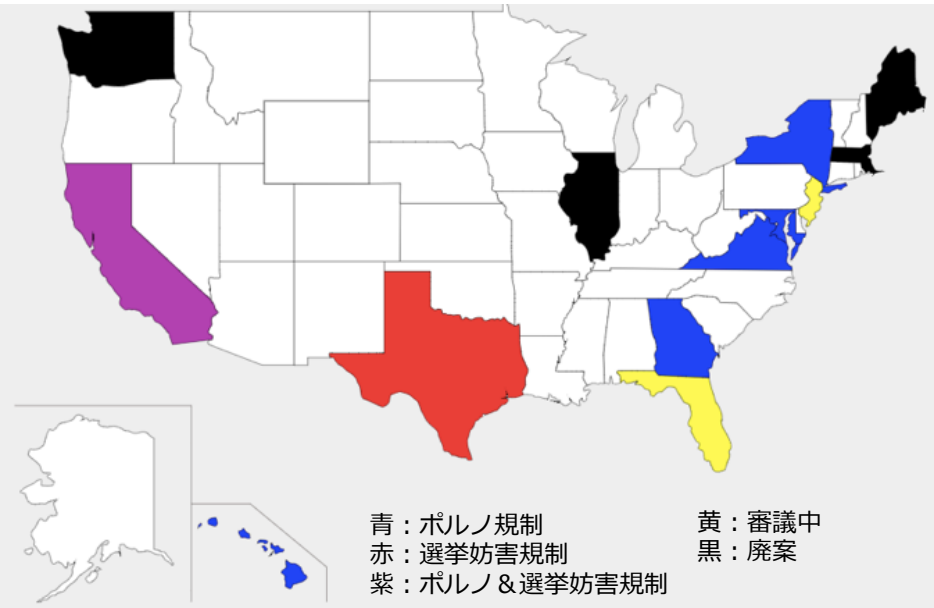

図４　米国各州におけるディープフェイクの規制状況(2022 年２月時点)

[23] S. 2559: Deepfake Task Force Act
https://www.govtrack.us/congress/bills/117/s2559

※本稿での考えや意見は執筆者個人（或はグループ）のもので当団体のものではありません。

【米国各州におけるディープフェイクの規制状況(2022 年２月時点)】

- ポルノ規制目的で法律を制定
  - ・バージニア州　(H.B.2678)(2019/03/18)
  - ・メリーランド州　(H.B.1027)(2019/04/30)
  - ・カルフォルニア州　(A.B. 602)(19/10/03)
  - ・ジョージア州　(S.B. 337) (2020/08/03)
  - ・ニューヨーク州　(S.B. S5959D)(2020/11/30)
  - ・ハワイ州　(A.309)(2021/06/24)
- 選挙妨害規制目的で法律を制定
  - ・テキサス州　(S.B.751)(2019/06/14)
  - ・カルフォルニア州　(A.B.730)(2019/10/03)
- ポルノ規制目的で法律を制定を検討中
  - ・ニュージャージー州　(A.1825)
  - ・フロリダ州　(S.B.658)
- 廃案
  - ・メリーランド州　(H.B.198)(選挙妨害規制目的)
  - ・ワシントン州　(S.B.6513)(選挙妨害規制目的)
  - ・メイン州　(L.D.1988)(選挙妨害規制目的)
  - ・イリノイ州　(S.B. 3171)(選挙妨害規制目的)
  - ・マサチューセッツ州　(H.B. 3366)(ポルノ規制目的)

SNS 事業者に適切な対応を義務付けることについては、通信品位法第 230 条がこれまで障害となっていた。この法律によれば、SNS のユーザーが違法なコンテンツを投稿した場合、被害者は投稿を行ったユーザーに対しては削除要求等の法的措置をとることが出来るが、SNS 事業者を相手取ることはできない。他方、SNS 事業者が好ましくないと判断した投稿については、SNS 事業者が自由に削除することが出来る。このような状況の中、2021 年の連邦議会では、通信品位法に規定された SNS 事業者への免責を撤廃し、一定の責任を負わせることを意図する「悪質なアルゴリズムに対抗する正義に関する 2021 年法案」[24]など、通信品位法第 230 条を修正しようとする動きが見られる。

最後に、国民のリテラシー向上の取組としては、2020 年 4 月 30 日に民主党の連邦下院議員が結成した「デジタル市民権に関する下院タスクフォース」[25]がある。これは、ネットや SNS での誤った情報や偽情報の特定、オンライン上の脅威や詐欺への理解、過激主義との闘い、そうした危険から身を守る方法などについて国民に知らせ、デジタル化が進む社会でのリテラシー向上を目指した有志議員の集まりであり、選挙や COVID19 ワクチン関連での虚偽情報で警鐘を鳴らすなどの活動を行っている。しかし、SNS では自分と似た興味関心をもつユーザーをフォローする結果、意見を SNS で発信すると自分と似た意見が返ってくるというエコーチェンバー現象[26]や、アルゴリズムがネット利用者個人の検索履歴やクリック履歴を分析し学習することで、個々のユーザーにとっては望むと望まざるとにかかわらず見たい情報が優先的に表示され、利用者の観点に合わない情報からは隔離され、自身の考え方や価値観の「バブル（泡）」の中に孤立するというフィルターバブル[27]という問題もあり、そもそもディープフェイクを真実だと思う人々への有効な手段とはなっていないという課題が残る。

---

[24] H.R. 5596: Justice Against Malicious Algorithms Act of 2021
https://www.govtrack.us/congress/bills/117/hr5596

[25] Congressional Task Force on Digital Citizenship
https://wexton.house.gov/about/congressional-task-force-on-digital-citizenship.htm

[26] 総務省「令和元年版　情報通信白書のポイント」
https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r01/html/nd114210.html

[27] 同上

※本稿での考えや意見は執筆者個人（或はグループ）のもので当団体のものではありません。

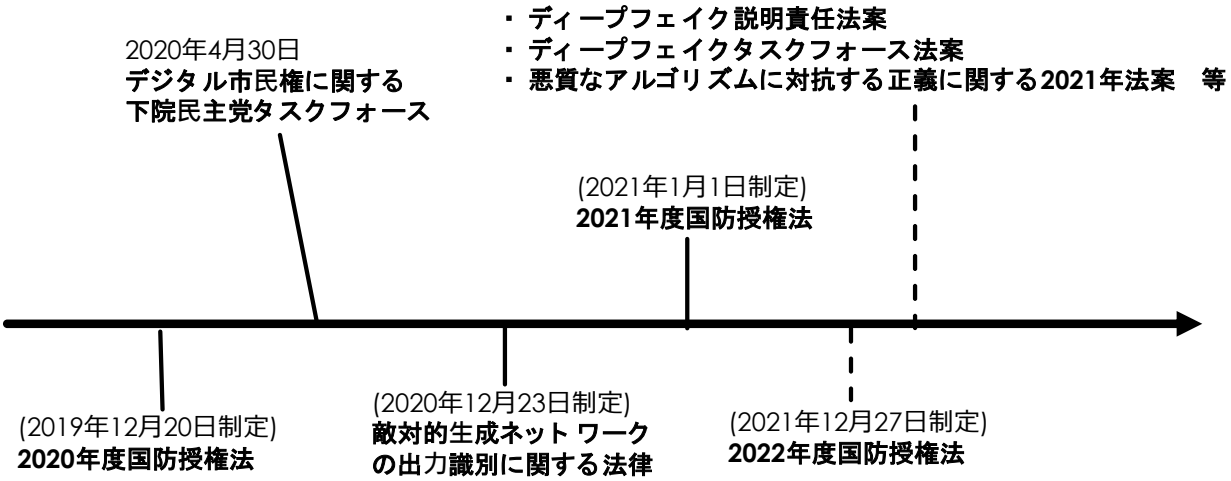

図５　米国連邦政府におけるディープフェイク関連法等

5. 日本の対応

日本では技術的手段による対応として、科学技術振興機構　戦略的創造研究推進事業の「信頼される AI システム」2020 年度採択課題として「インフォデミックを克服するソーシャル情報基盤技術」[28]があり、また、総務省がプラットフォームに関する研究会[29]において SNS 等のプラットフォーム事業者からフェイクニュースや偽情報への対策についてヒアリングを行うなどの取組を行っている。米国以外にも、中国や EU 等がディープフェイクをはじめとする AI 技術の規制やガイドラインを策定・検討していると言われている。しかし諸外国に比べ、日本ではディープフェイクと安全保障を結びつけた議論が少ないように思われるため、安全保障の観点からも活発に議論・検討していく必要があると考える。

以　上

[28] インフォデミックを克服するソーシャル情報基盤技術 http://research.nii.ac.jp/~iechizen/crest/index.html
[29] 総務省プラットフォームサービスに関する研究会
https://www.soumu.go.jp/main_sosiki/kenkyu/platform_service/index.html

Nakasone Peace Institute
※本稿での考えや意見は執筆者個人（或はグループ）のもので当団体のものではありません。

【参考文献】本文脚注等で引用している文献以外で参考にしたもの（web は 2022 年 2 月 6 日アクセス）

(1) SEAN DACK “Deep Fakes, Fake News, and What Comes Next” (2019 年 3 月 20 日) https://jsis.washington.edu/news/deep-fakes-fake-news-and-what-comes-next/
(2) Korey Clark “‘Deepfakes’ Emerging Issue in State Legislatures” (2021 年 6 月 4 日) https://www.lexisnexis.com/en-us/products/state-net/news/2021/06/04/Deepfakes-Emerging-Issue-in-State-Legislatures.page
(3) “Deepfake Laws Risk Creating More Problems Than They Solve” (2021 年 3 月 1 日) https://regproject.org/paper/deepfake-laws-risk-creating-more-problems-than-they-solve/
(4) HOUSE COMMITTEE ON ENERGY & COMMERCE “E&C LEADERS ANNOUNCE LEGISLATION TO REFORM SECTION 230” (2021 年 10 月 14 日) https://energycommerce.house.gov/newsroom/press-releases/ec-leaders-announce-legislation-to-reform-section-230
(5) 湯淺 墾道（笹川平和財団）「米国連邦法によるディープフェイク規制―2021 年度国防授権法と IOGAN 法―」(2021 年 11 月 12 日) https://www.spf.org/iina/articles/harumichi_yuasa_02.html
(6) 湯淺 墾道（笹川平和財団）「アメリカ選挙法におけるディープフェイク規制の動向」(2021 年 3 月 19 日) https://www.spf.org/iina/articles/harumichi_yuasa_01.html
(7) 株式会社三菱総合研究所デジタル・イノベーション本部 (総務省 プラットフォームサービスに関する研究会 第 24 回 資料４)「インターネット上の違法・有害情報を巡る米国の動向」(2021 年 3 月 17 日) https://www.soumu.go.jp/main_content/000739937.pdf
(8) デジタル・フォレンジック研究会　第６４３号コラム：「米国で通信品位法 230 条の改正が議論されているのはなぜか」(2020 年 12 月 7 日)　https://digitalforensic.jp/2020/12/07/column643/
(9) Aurélien Géron 著、下田倫大 監訳、長尾高弘　訳『scikit-learn、Keras、Tensor による実践機械学習 第２版』（2020 年 10 月 30 日株式会社オライリー・ジャパン発行）
(10) みずほリサーチ&テクノロジーズ株式会社　経営・ＩＴコンサルティング部 (総務省 プラットフォームサービスに関する研究会 第 27 回 資料 3−2)「ディープフェイクについて」(2021 年 5 月 13 日) https://www.soumu.go.jp/main_content/000749422.pdf
(11) S. 3805 (115th): Malicious Deep Fake Prohibition Act of 2018 https://www.govtrack.us/congress/bills/115/s3805
(12) S. 3788 (115th): A bill to require studies on cyberexploitation of employees of certain Federal departments and their families, and for other purposes. https://www.govtrack.us/congress/bills/115/s3788
(13) S. 1348 (116th): A bill to require the Secretary of Defense to conduct a study on cyberexploitation of members of the Armed Forces and their families, and for other purposes. https://www.govtrack.us/congress/bills/116/s1348
(14) H.R. 3230 (116th): Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019 https://www.govtrack.us/congress/bills/116/hr3230
(15) S. 2065 (116th): Deepfake Report Act of 2019 https://www.govtrack.us/congress/bills/116/s2065
(16) Congressional Research Service “Deep Fakes and National Security” (2021 年 6 月 8 日) https://crsreports.congress.gov/product/details?prodcode=IF11333

※本稿での考えや意見は執筆者個人（或はグループ）のもので当団体のものではありません。