

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）

ディープフェイクについて

2021.5.13

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

1. ディープフェイクとは

- 「ディープフェイク」は、「ディープラーニング」と「フェイク」を組み合わせた造語。
- 現在では人工知能を用いて、実際には存在しないリアルで高精細な人物の映像・動画を制作する行為や、それらで制作された映像・動画について指すことが多くなっている。
- 本来は人物の動画、画像、音声を人工的に合成するための処理技術の一種（マイクロソフト「Spot the Deepfake」より）。広大な撮影スタジオや、専用の撮影・編集機材を用いなくても、PCなどの環境で動画等を作ることができるため、映画制作等において貢献する。
- ディープフェイクが注目を集めたきっかけは、2017年、あるポルノ映像が作成されネット掲示板に投稿されたことによる。ハンドルネームが「ディープフェイクス（deepfakes）」であった。
- 2020年11月の米大統領選において「ディープフェイク」を用いた偽情報が制作され出回ることが懸念されたが、実際にはそのような状況にはならなかった（事例はゼロではなかった）。

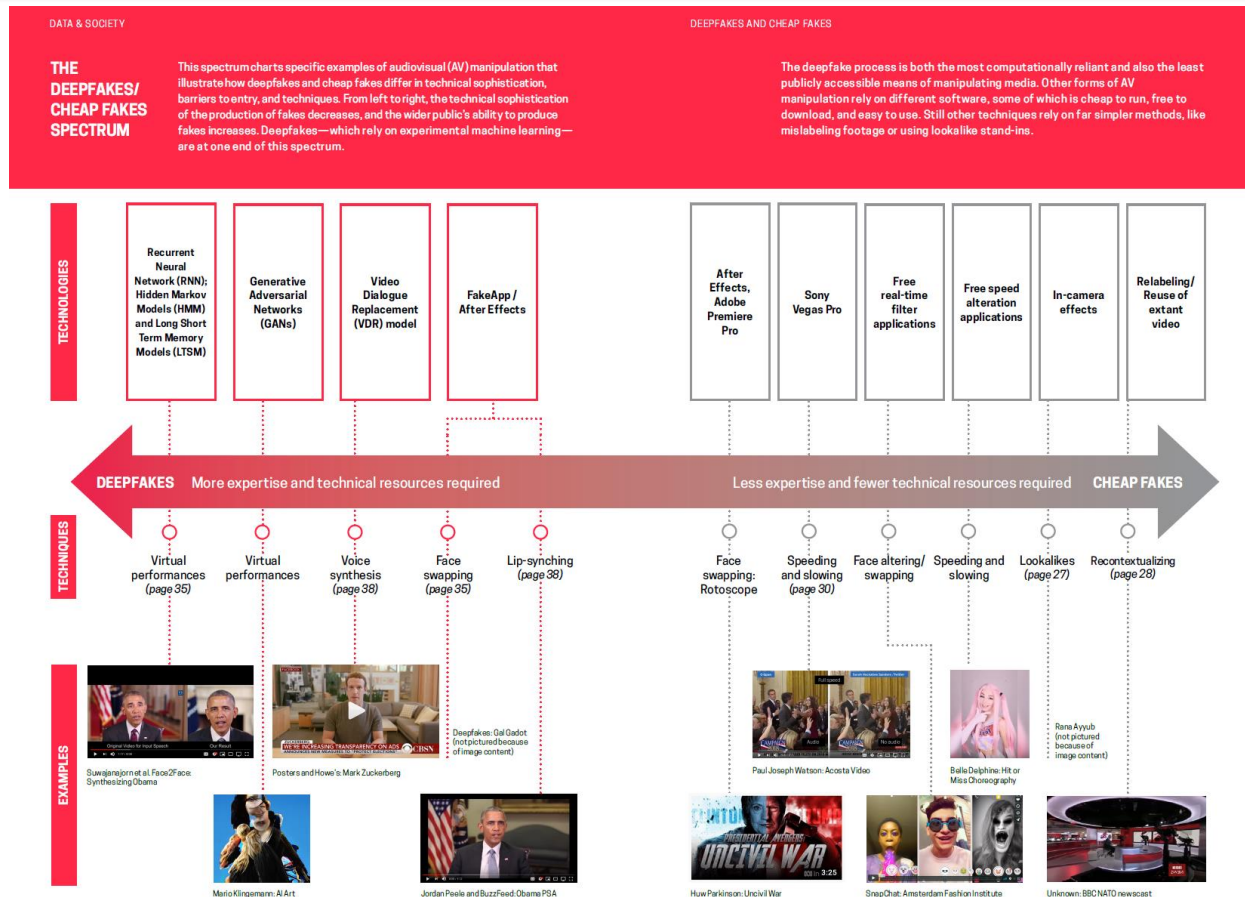
ディープフェイク動画例 1

俳優トム・クルーズ氏のディープフェイク動画

<https://vfxchrisume.com/>

1. ディープフェイクとは

- ❑ 「ディープフェイク」は、「チープフェイク」と対義して使われることもある。
- ❑ 「チープフェイク」は動画の再生速度を調整したり、画像編集ソフトを用いて画像の必要な部分を切り取る等の簡単な処理で作成する技術等を指す。
- ❑ なお、日本でも話題になった、米民主党の下院議長ナンシー・ペロシ氏が酔って話しているように見えた動画は、元の動画の再生速度を遅くして制作されており、「チープフェイク」に分類できる。



(出典) Data & Society「DEEPPFAKES AND CHEAP FAKES」

2. ディープフェイクの事案 2. 1 ディープフェイクの例

- これまでのディープフェイクに関連した主な事案について時系列で示した。
- ディープフェイクは、海外の事案が多いが、2020年10月には、わが国において自らの有料ウェブサイトや海外サイトにディープフェイクポルノ動画をアップロードしていたとして大学生とシステムエンジニアの男2人が逮捕された。

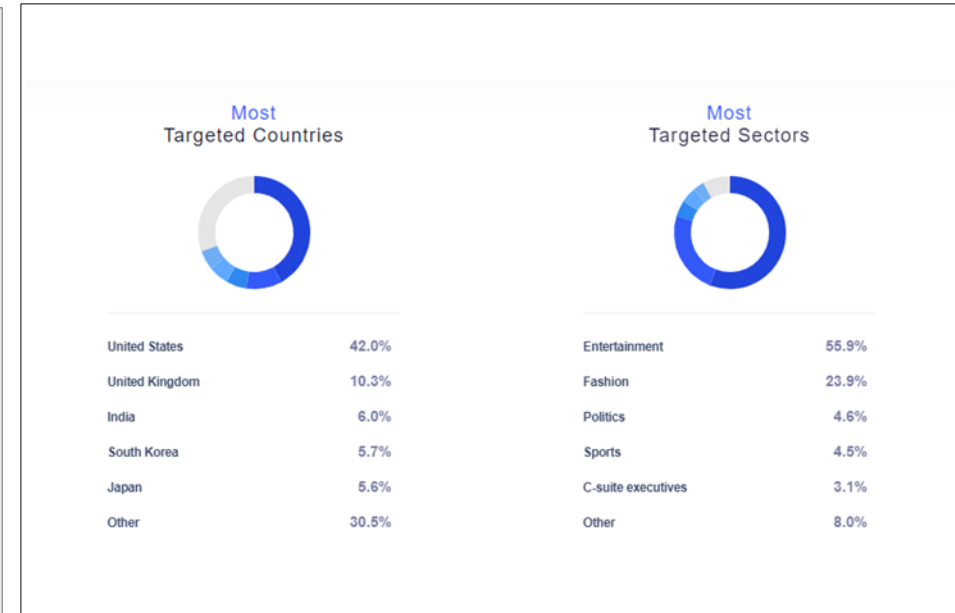
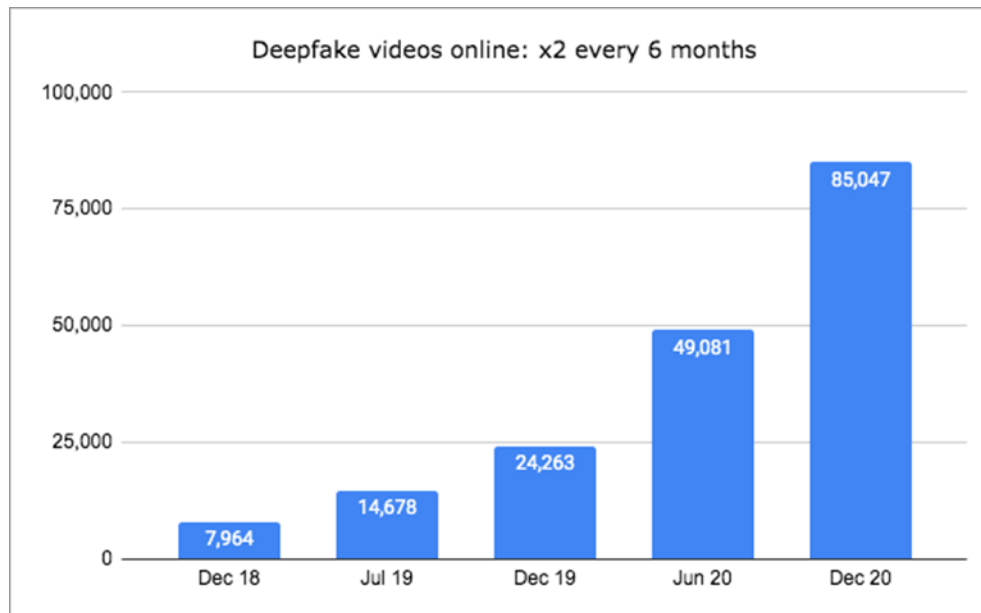
年	国	内容
2017	米国	• 匿名のRedditユーザーが、ポルノ動画を女性俳優の顔に差し替えて公開。ユーザー名が「ディープフェイクス（deepfakes）」。
2018	米国	• ハリウッド映画俳優の登場場面を集めたネット動画が公開。
	米国	• 映画監督が、ディープフェイクの問題点を訴えるため、バラク・オバマ元大統領が、実際に発言していない言葉を発する動画を制作。
2019	日本	• エイプリルフール向けに、日本人が作成したAmazonの飛行船のフェイク動画が話題に。
	欧州	• アーティストが、イギリスで開かれた芸術イベントの作品として、マーク・ザッカーバーグ氏の演説フェイク動画を制作。
	欧州	• 英国を拠点とするエネルギー企業のCEOが、フェイク音声により、ドイツの親会社のCEOと電話で話していると信じてしまう詐欺事件に巻き込まれる。被害額は22万ユーロ。
	中国	• 人気となったディープフェイクアプリ「ZAO」について、作成された生体画像をアプリベンダーに無償で永続的に提供する利用規約が問題に。
	中国	• 政府が、オンラインコンテンツプロバイダーに対して、AIを利用して作成したコンテンツにその旨を明記することを義務付け。

年	国	内容
2020	日本	• 自らの有料サイトや海外サイトにディープフェイクポルノ動画をアップロードしていたとして、大学生とシステムエンジニアの2人が名誉棄損や著作権法違反で逮捕。
	米国	• ディープフェイクの可能性を警告するため、フロリダ州の民主党下院議員候補フィル・エアーが、対立候補の共和党議員マット・ゲイツのディープフェイク版の選挙広告を制作。
	日本	• 女性タレントの顔に差し替えたディープフェイクポルノ動画を、まとめサイトで紹介していたとして、名誉毀損や著作権法違反の疑いで警視庁などが容疑者を逮捕。
	米国	• ジョー・バイデン大統領候補がフロリダ州の集会にやって来た支持者に対して、ミネソタ州の支持者であると間違えてあいさつしたように見せるフェイク動画が公開。
2021	日本	• 加藤官房長官が地震に関する記者会見で笑みを浮かべる画像がTwitterに掲載。
	米国	• アーティストとのまね芸人が、俳優トム・クルーズ氏のディープフェイク動画をTikTokで公開。
	米国	• 娘が所属するチアリーディングのチームメイトをチームから追い出すため、母親がディープフェイク技術を使い、チームメイトのわいせつな画像や動画を作成したとして、逮捕。

(出典) 各種ウェブサイトより、みずほリサーチ&テクノロジーズが作成

2. ディープフェイクの事案 2. 2 発生件数

- ディープフェイクで作成された動画は増加傾向にある。
- オランダの情報セキュリティ調査会社の「Sensity（旧Deeprtrace）」は、約500の情報源を対象に監視活動を通じてディープフェイク動画の検出を行っている。**2020年12月には8.5万件の動画を検出した。**同社が活動を始めた2018年以降、**6か月ごとに約2倍のペースでディープフェイク動画が検出されている。**
- ディープフェイクの脅威に晒されている国は、米国が42%と最も高いが、日本は5.6%と5番目にランクされている。
- また、ディープフェイク動画の分野は、「エンタテインメント」55.9%、「ファッション」23.9%、「政治」4.6%の順で高くなっている。
- 動画の内容はポルノが多いと言われており、対象者のプライバシーや肖像権保護の観点から問題視されている。



(出典) Sensityウェブサイト

3. ディープフェイク対策例 3. 1 対策ツール等

- ディープフェイクで作成された動画を検出する技術・ツールが開発されている。Microsoft「Video Authenticator」、Sensityの「DEEPFAKE DETECTION」等、実用化したサービスもある。また、トルコの情報セキュリティ企業の「deepware」は、対象動画がディープフェイクの可能性のあるかどうかを判断できるツール「DEEPWARE SCANNER (β版)」を開発し公開している。
- ディープフェイクの脅威に対抗するための専門組織も立ち上げられている。例えば、米国の情報セキュリティ企業の「マカフィー」ではAIを活用しディープフェイクの検出を行う「ディープフェイクラボ」を2020年10月に設立した。
- しかし、ディープフェイク制作技術も同時に進歩しており、完全に抑え込めた状況にはなく、検出のための研究開発も進められている。

リアルタイムで動画の信頼性が表示される。赤枠はディープフェイク。



検出対象となるのは、Youtube, Facebook, Twitter上にあるディープフェイク動画。PCのウェブブラウザ、Androidのアプリで利用できる。

以下はYoutubeにアップされたトム・クルーズのディープフェイク動画をかけた結果、ディープフェイク(メーターが赤)と検出された。

ディープフェイク動画例 2

deepware「DEEPWARE SCANNER」

<https://scanner.deepware.ai/result/8fccb073eee38f5cbf40bcb0be2787bc28ba42c3-1615060363/>

(出典) Microsoft「Video Authenticator」

<https://news.microsoft.com/ja-jp/2020/09/07/200907-disinformation-deepfakes-newsguard-video-authenticator/>

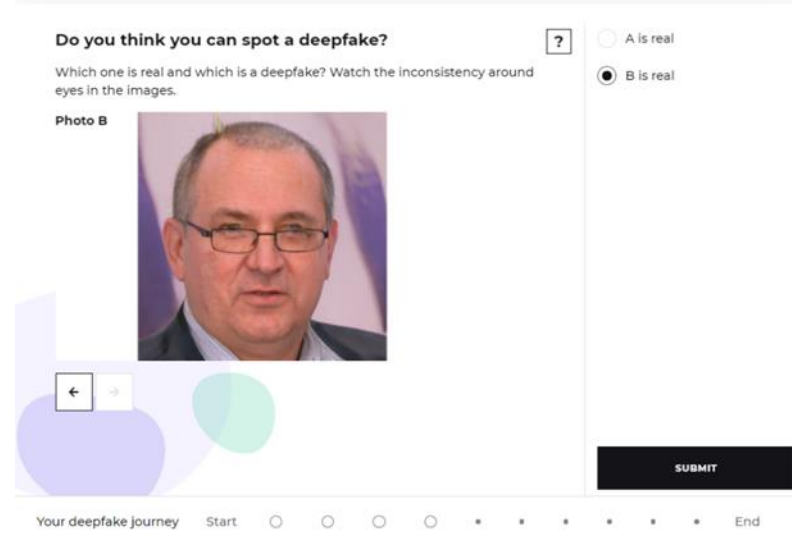
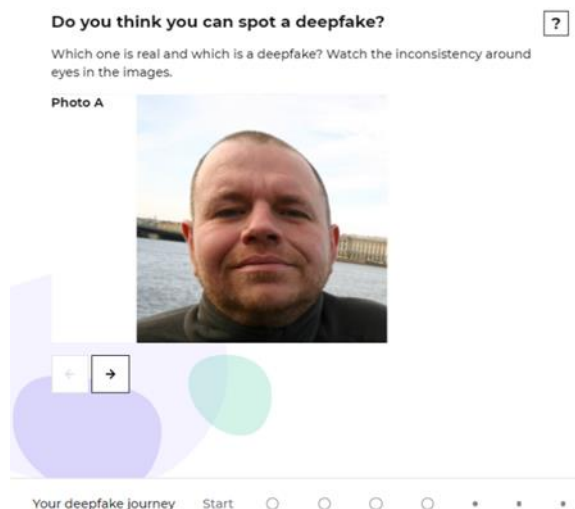
3. ディープフェイク対策例 3. 2 自習用教材 (Spot the Deepfake)

- ❑ 一般の人がディープフェイクについて自己学習できるオンライン教材も存在する。
- ❑ Microsoft、ワシントン大学 (UW)のCenter for an Informed Public、USA Today、Sensityは共同で「Spot the Deepfake(仮訳：ディープフェイクを見つけよう)」をウェブ上で公表した。
- ❑ 学習者は10問のクイズに答えながらディープフェイクの特徴を学ぶことができる。

本物の画像とディープフェイクで制作された静止画像を見分ける質問

質問横の「?」ボタンを押すとヒントが表示される。

見分けるために毛や眼鏡に注目するよう説明されている。



(出典) Microsoft他「Spot the Deepfake」

<https://www.spotdeepfakes.org/en-US/quiz>

3. ディープフェイク対策例 3. 3 研究開発(deepfake detection challenge (DFDC))

- ディープフェイク検出技術の公募コンテスト。Facebook, Microsoft, ミュンヘン工科大学, フェデリコ2世ナポリ大学, コーネル工科大学, MIT, オックスフォード大学, UCバークレー, メリーランド大学カレッジパーク校, ニューヨーク州立大学オルバニー校, Partnership on AI(※)が立ち上げた。
- **ディープフェイク動画と本物の動画を識別する**もので、2019年12月から2020年5月まで開催。優勝賞金は総額100万ドル、1位には50万ドルが授与された。2,114チームが参加した。

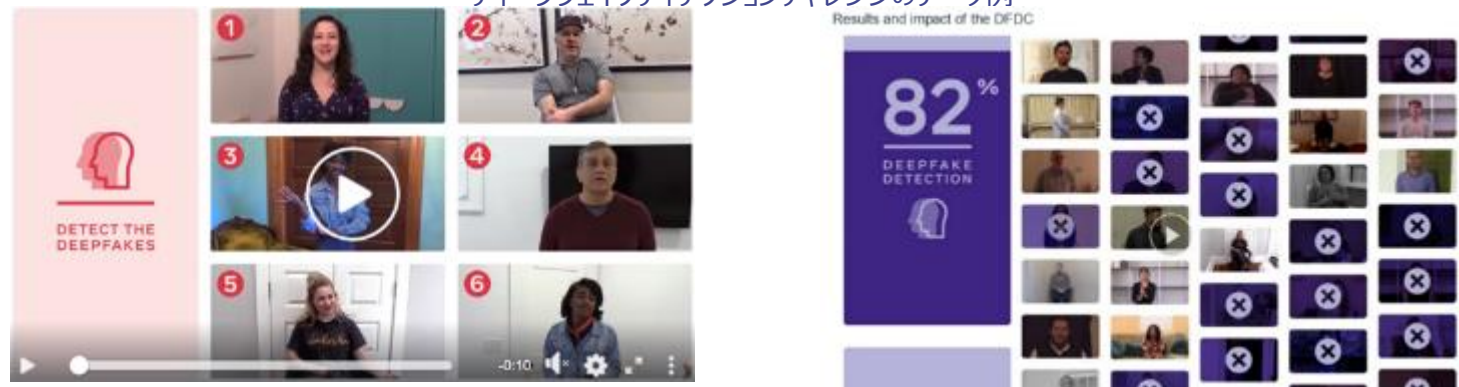
【結果】

- コンテスト用に制作し、**あらかじめ公開していた動画**（パブリックデータセット：115,000本）を対象とした場合、トップのチームがフェイク動画を識別できた精度は、**82.56%**であった。
- 一方、コンテスト参加者に**予め提供されていなかった動画**（ブラックボックスデータセット：識別を困難にする加工も施された10,000本）を対象とした場合には、トップチームがフェイクを識別できた精度は、**65.18%**であった。
- **パブリックデータセットとブラックボックスデータセットでは、評価順位に大きな違いがあり、既知のデータから、新たな事例に対応することの困難さも明らかになった。**

※AI研究に関する非営利団体。Amazon、Google、Facebook、IBM、Microsoftを立ち上げメンバーとする。SONY等日本企業も参加。

<https://www.partnershiponai.org/partners/>

ディープフェイクディテクションチャレンジのデータ例



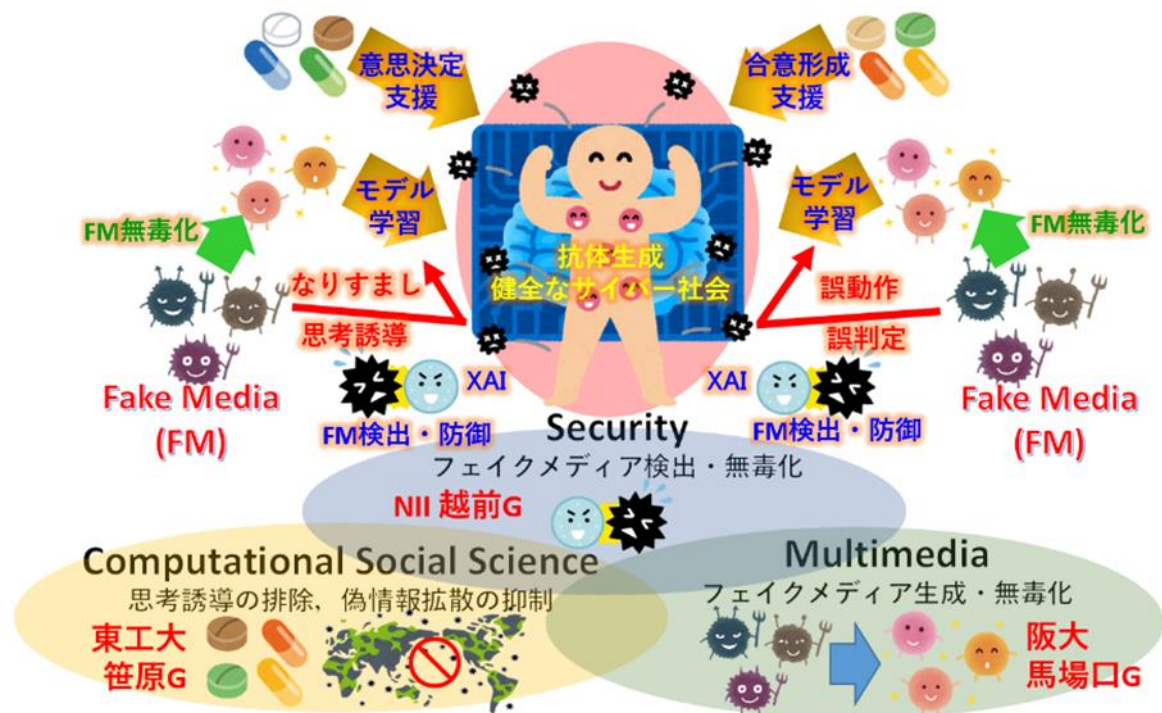
(出典) Facebook

<https://ai.facebook.com/blog/deepfake-detection-challenge-results-an-open-initiative-to-advance-ai/>

<https://www.kaggle.com/c/deepfake-detection-challenge>

3. ディープフェイク対策例 3. 3 研究開発(日本のディープフェイク検知の取組み)

- 日本では、文部科学省が定めた戦略目標「信頼されるAI」の下、2020年に科学技術振興機構（JST）の5つの戦略目標の中に、「信頼されるAI」も研究領域として位置づけられ、**公募の結果5件が採択された。**
- **国立情報学研究所（NII）越前 功教授**が研究代表者となり、「**インフォデミックを克服するソーシャル情報基盤技術**」は、2020年12月から5.5年間かけて、取り組まれる。
- 同研究は、「AIにより生成されたフェイク映像、フェイク音声、フェイク文書などの多様なモダリティによる**フェイクメディア（FM）**を用いた高度な攻撃を検出・防御する一方で、**信頼性の高い多様なメディアを積極的に取り込むことで人間の意思決定や合意形成を促し、サイバー空間における人間の免疫力を高めるソーシャル情報基盤技術**を確立する。」ことを目的としている。



(出典)文部科学省「令和2年度の戦略目標及び研究開発目標の決定について」
https://www.mext.go.jp/b_menu/houdou/2020/mext_00147.html
JST「信頼されるAIシステム」2020年度採択課題
https://www.jst.go.jp/kisoken/crest/project/1111109/1111109_2020.html
NII「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/research.html>

4. まとめ(本日時点)

①ディープフェイク動画は、まだたくさん世の中に出ている状況ではない

ディープフェイクは、専門環境が不要でPCで作成でき、従来よりもコストをかけずにリアルな動画等を制作できる技術。しかし、高精細な動画を制作しようとした場合には、現状では一定のスキルが求められ、かつ手間暇がかかる。前述したトム・クルーズのディープフェイク動画においては、制作者は、AIの画像学習に2か月、動画撮影に2日、編集に24時間かけている。

2020米国大統領選においても、事前に懸念されていたようなディープフェイク動画が多く出るような状況にはなかった。簡単で安価なツールを使い、画像やテキストなどのディープフェイク以外の効果的な偽情報を生成できたためと分析されている。

②ディープフェイク制作技術は進化し続けている

ディープフェイクの検知技術の開発を行った、国立情報学研究所の越前功教授によると「ディープフェイク動画は精度は上がり、人目では判別がつきにくくなってはいるが、時間軸で見たときにまだ違和感が出てくる状態。一方で、ディープフェイクで作った静止画像は人目では真偽の判別が難しい、よりリアルなレベルになっている。」と、画像・動画制作技術は進化し続けているとの指摘がある（ヒアリングによる）。

また、前述した「deepfake detection challenge」では、主催者側で、識別を困難にする加工が施され、予め提供されていなかった動画を用意し、それらを対象に検出を行った結果、トップチームが識別できた精度は65%であった。検出精度は完璧とはいえず状況である。

今後、新たなディープフェイク画像・動画等が出てくる可能性があり、検出技術の開発も進める必要がある。