資料３

偽・誤情報検知等を目的に研究開発されたICTツール例

2023.5.25

みずほリサーチ＆テクノロジーズ株式会社
デジタルコンサルティング部

One MIZUHO
未来へ。お客さまとともに

目次

1.偽・誤情報検知等を目的に研究開発されたICTツール例

- 諸外国（一部日本含む）における**偽・誤情報検知等を目的に研究開発が行われたICTツール**例を収集した。
- 目的から分類すると、**「コンテンツの検証」、「リソースの信頼性検証」、「フェイクニュース検出」、「コンテンツの検証作業支援」「その他」**にできる。
- なお、日本でも話題となったGPT-3、4やLaMDA、PaLM 2、LLaMAなどの大規模言語モデル（LLM）は、偽情報を含む文章生成ツールに用いられる可能性があるが、逆に偽情報を検知するためのツールに用いられる可能性もある。**日本発のLLMも対象とした。**

（本資料の対象としていないもの）

- 別目的で開発されたツールを偽誤情報の検知等目的で応用できる可能性もある。
- メディア・情報リテラシーの向上を目的としたゲーミフィケーション教材や動画教材も広義の対策ツールと考えられる。

No.	ツール名	目的	内容
1	**SYNTHETIQ VISION（日本）**	• コンテンツ検証（動画、画像）	• AIにより生成されたフェイク顔映像を自動判定するプログラムを開発。 • 国立情報学研究所シンセティックメディア国際研究センター長の越前功氏と副センター長の山岸順一氏のグループが開発。サイバーエージェントが採用した。
2	**Microsoft Video Authenticator**	• コンテンツ検証（動画、画像）	• ディープフェイク動画、画像を検出する技術。 • リアルタイムで動画の信頼性が表示される。ディープフェイク部分を赤枠で示される。 • マイクロソフトが開発。
3	**Reality Defender**	• コンテンツ検証（動画、画像）	• ディープフェイクやGenerative AIで生成されたメディアを検出することができる。 • NATO、米国防総省、米国土安全保障省などの政府機関や、米国やアジアの放送メディア等が真偽検証用に利用。 • Reality Defenderが開発。
4	**SBIs(Self-Blended Images)（日本）**	• コンテンツ検証（動画、画像）	• ディープフェイク検出AIを開発。 • 東京大学情報理工学系研究科電子情報学　准教授山崎俊彦氏と院生による研究。
5	**FakeCatcher**	• コンテンツ検証（動画）	• 人間の生物学的信号とデータを使用してディープフェイクを96%の精度で識別および分類するツールを開発。 • インテルとビンガムトン大学のグラフィックス&イメージコンピューティング研究所の共同研究。
6	**deepware scanner**	• コンテンツ検証（動画）	• YouTube, Facebook, Twitter上の動画の信頼性をリアルタイムで表示する。 • ディープフェイク動画を検知するとツール内のメーターが赤を示す。 • Deepwareが開発。PCのウェブブラウザ、Androidのアプリで利用できる。
7	**TinEye**	• コンテンツ検証（画像）	• リバースイメージ検索ツール。「Idée Inc.（カナダ）」が提供する。
8	**The News Provenance Project**	• コンテンツ検証（画像）	• 偽・誤情報を防止するために、ブロックチェーン技術を利用してニュース画像の信頼性を確認できる基盤を構築する。IBMのHyper Ledger Fabric上に構築される。 • New York Timesが2019年に計画を公表した。

1.偽・誤情報検知等を目的に研究開発されたICTツール例

No.	ツール名	目的	内容
9	**Sphere**	• コンテンツ検証（テキスト）	• 記事の真正性を評価する偽情報検出AIツール。開発にあたってはWikipedia上のデータを使用した。 • Metaが開発。
10	**WeVerify/InVID**	• コンテンツ検証（テキスト、動画、画像） • リソースの信頼性検証（作成者）	• EUのHorizon 2020の資金拠出を受けて開発された動画・画像検証ツール。 • Ontotext ADがプロジェクト代表機関。 • 機械学習技術を用いて、ソーシャルメディア上のテキスト、画像、動画を分析しコンテンツの出典を明らかにし、また、大量データから偽情報を作成・宣伝するアカウントを判別する。 • 検証をおこなった偽情報の画像や動画を蓄積した分散型データベースを作成して公開する（ブロックチェーンを使用）。
11	**NewsGuard**	• リソースの信頼性検証	• ニュースメディアの信頼性評価を行う。AIではなく人間が評価する点が特徴。
12	**Disinformation Index（GDI）**	• リソースの信頼性検証	• 特定のメディアに関する偽情報の可能性を評価、国単位での評価も実施。
13	**TruthNest**	• リソースの信頼性検証（作成者）	• Twitter上の情報を対象として、リソースの信頼性検証を行うための支援ツール。 • アカウントの投稿傾向を分析して、Botや偽アカウントの検出を行うことができる。
14	**SocialTruth Project**	• リソースの信頼性検証（コンテンツ、作成者） • フェイクニュース検出	• AI、ブロックチェーンを活用し、①SNS上に流通するコンテンツや発信者(著者)の信頼性検証、②偽情報の増加検出を行う。 • 画像、ビデオ、テキストが対象。
15	**Grover**	• フェイクニュース検出	• 英語の文章を対象に人間かAIが書いたものか否かを92%以上の精度で判別できる（2019年の公表時点において）。 • アレン人工知能研究所の研究。
16	**SummarizeBot API「Fake News Detection」**	• フェイクニュース検出	• 文章の要約サービス。フェイクニュース検出も可能。URL先のニュース記事文章をAIが分析し、事実/偽の確率を表示する。 • API形式で提供。対象文章のURLを入力すると、判定結果が表示される。

1.偽・誤情報検知等を目的に研究開発されたICTツール例

No.	ツール名	目的	内容
17	**ファクトチェック支援システム（日本）**	• コンテンツの検証作業支援	• 疑似言説収集システム（FCC）＋疑似言説データベース（Claim Monitor）＋ファクトチェックナビから構成される。 • 認定NPO法人ファクトチェック・イニシアティブ（FIJ）が提供する。 • 11メディア・団体（新聞社・放送局4社含む）、2つの教育機関が利用。
18	**Truly Media**	• コンテンツの検証作業支援 • コンテンツ検証（画像）	• Twitter、Facebook、YouTube等上から集めたソーシャルメディア上のコンテンツの検証を支援するためのツール。 • 画像分析アルゴリズムにより画像の加工状況を分析する。
19	**ClaimBuster**	• コンテンツの検証作業支援 • フェイクニュース検出	• Webベースの自動化されたライブファクトチェックツール。事実と虚偽の情報を特定する。 • テキサス大学アーリントン校が開発。
20	**Full Fact AI「Automated Fact Checking」**	• コンテンツの検証作業支援 • コンテンツ検証（テキスト）	• 偽誤情報の研究機関である「FULLFACT」がGoogleの言語モデル「BERT」を利用したAI分析ツールを開発し、実運用している。有料ライセンスで外部への提供もおこなう。 • BERTは膨大なテキスト化された文章データの中からクレーム部分を検出して分類化（原因と結果、数量など）する。
20	**FactChat**	• ライブファクトチェック結果共有	• 2020年米大統領選期間中、メッセージングアプリの「WhatsApp」上に、米国内の10のファクトチェッカーがFactChatに、ファクトチェック結果を英語とスペイン語で登録公開した。 • Poynter InstituteのIFCN（米国）が開発・提供した。
21	**フェイクアラート（日本）**	• フェイクニュースへのアラート	• ネット上で、意見の分断が起きている、少数の人が発信しているニュース記事に対してアラートを出す仕組みを開発し実証を行った（2022年）。 • TDAI Lab、東京大学鳥海教授、ＮＨＫの共同研究、実証。

※ 日本においては、インターネットの「信頼」には早期から取り組んでおり、TrustedWeb構想や、最近では、情報的健康の共同宣言、Originator Profile技術の取組が行われている。

2.日本発の大規模言語モデル（LLM）

- **2023年に入り日本語に特化したLLMの開発が増えている。**

■日本発の大規模言語モデル（LLM）

No.	ツール名	企業名	内容
1	**HyperCLOVA**	ワークスモバイルジャパン株式会社	• 日本語に特化した汎用型AIをLINEとNAVERで共同開発（2020年～）。大規模な日本語データを学習データ（1,750億以上のパラメーター、100億ページ以上の日本語）とすることで日本語におけるAIの水準を向上させるために取り組む。2022年12月には第12回対話システムシンポジウムにおいて2部門で1位となった。 • 2023年4月1日よりLINEのAI事業は「ワークスモバイルジャパン」に移管された。 • ソフトバンクはLINEとHyperCLOVA技術をベースに日本版のGPTの開発に取り組むと説明（2023年5月10日のソフトバンクの決算会見）
2	**LHTM-2**	株式会社オルツ	• 日本発のGPT-3と同水準のパラメータ数の大規模言語処理モデルを公表（2023年2月14日）。同社の初期バージョン「LHTM」を発展させた。
3	**大規模言語モデル**	NTT株式会社	• 2023年度中に独自の大規模言語モデルの開発、生成AIの商品化予定（2023年5月12日の決算会見）。
4	**日本語LLM**	株式会社サイバーエージェント	• 最大68億パラメータ日本語大規模言語モデルを一般公開した（2023年5月17日）。 • 学習ソースはWikipediaおよびCommon Crawlのオープンな日本語データを使用。 • 利用は無料。商用利用可能なCC BY-SA 4.0ライセンスで提供。
5	**汎用GPT言語モデル** **対話GPT言語モデル**	rinna株式会社	• 日本語に特化した汎用言語モデル（36億パラメータ）と対話言語モデルを公開した（2023年5月17日）。 • オープンソース。商用利用可能なMIT Licenseとなっている。 • 学習ソースは汎用言語モデルは日本語Wikipedia、C4、CC-100のオープンソースデータ。対話言語モデルはHH-RLHF、SHP、FLANの一部を日本語に翻訳した。

本資料の内容は2023年５月作成時点のものとなります。

みずほリサーチ＆テクノロジーズ株式会社

デジタルコンサルティング部　上席主任コンサルタント

中　志津馬（なか　しづま）