JPR&C Research&Consulting
JPR&C アナリスト・レポート
JPR&C Analyst Report
Vol. 23

2025 年 9 月 8 日

信頼に対するあらたな定義

―生成 AI 時代の偽情報リスクと、企業が築くべき戦略的レジリエンス―

JPR&C アナリスト　大成　昂

はじめに　見過ごされたパラダイムシフト

生成 AI の活用は、もはや未来への備えではありません。現代の国際競争を勝ち抜くための「必須戦略」となりつつあります。

しかし日本の現状を見ると、欧米企業の AI 活用率が 3〜6 割に達する一方、国内では 1〜3 割程度に留まり、その差は歴然です。この差は、今後「生成 AI デバイド」として、より深刻な形で社会に現れるでしょう。

特に、世代間の構造変化は深刻です。すでに 10 代の過半数が AI を日常的に使う「AI ネイティブ世代」が、数年後には労働市場の主役となります。しかし、多くの企業では生成 AI に対する認識が追いついておらず、組織内に深刻なジェネレーションギャップが生まれることは想像に難くありません。

その結果、個人の AI 活用能力がキャリアを直接左右する時代が到来し、スキルを持つ者と持たない者との格差は、もはや覆すことのできないレベルにまで拡大していくと予測されます。文章、映像、画像から自動化まで、生成 AI に関する驚異的なアップデートについては多くの人たちに驚きと発見を見せる一方で、その裏側で静かに、しかし急速に進行している、より構造的で深刻なパラダイムシフトにも目を向ける必要があります。それは、AI を使いこなす側が、そうでない人々の目に触れる情報や、得るべき機会をコントロールされるリスクが起こりうるということです。

AI によって巧妙に生成されたフェイクニュースは、個人を知らず知らずのうちに、偏った情報だけが心地よく響く「エコーチャンバー」へと誘い込みます。そこで形成された意見は、果たして本当に自分自身のものだと言えるのでしょうか。AI を理解しているつもりだとしても、その影響を洞察できないことは、「自らの判断や選択の自由を静かに明け渡すことと、同義になりかねない」のです。

本稿は、AI の脅威を解説しつつ、今後にふえていくであろう偽情報によって「信頼」が陳腐化し、「事実」の定義が揺らぐ新時代において、企業や組織が事業の存続をかけて構築

すべき、新たなインテリジェンス戦略の必要性を提示するものです。

1 ケーススタディ：「証拠」が捏造される時代

「百聞は一見に如かず」ということわざが、その意味を失いました。この変化こそ、現代の偽情報問題を理解する上での出発点です。

これまでの偽情報が、主に虚偽の「物語」や「記事」を作成することであったのに対し、生成 AI は根本的な変革をもたらしました。それは、捏造された「一次情報」そのものを生み出す能力です。つまり、加工された写真、音声クリップ、動画など、一見すると直接的な「証拠」に見えるものを、ゼロから創り出せるようになったのです。これにより、私たちは単に物語の信憑性を疑うだけでなく、自らの目や耳といった感覚そのものを疑わなければならない時代に突入しました。これは受け手にとって遥かに高い認知的負荷を強いるものであり、証拠に基づいて真偽を判断するという、従来のファクトチェックを著しく困難にしています。

岸田元首相が日本テレビのニュース番組（日本テレビ）で卑猥な発言をしているかのようなフェイク動画がネット上で拡散された。日テレは注意喚起し、制作者を名乗る人物は X 上で謝罪をした騒動[1]。

この中心的技術が、ディープフェイクです。「ディープラーニング」と「フェイク」を組み合わせたこの造語は、極めてリアルでありながら捏造された動画や音声記録を作成する技術を指します。これは、偽物を作る AI（生成者）と、それを見破る AI（識別者）が互い

[1] https://youtu.be/0jP2oLiIUIw?feature=shared

に競い合い、訓練を重ねることで、生成者が次第に本物と見分けがつかないほど精巧な偽物を作れるようになる相互に研磨していくことでできあがるシステムです。

この技術は、かつての不鮮明で不自然さが残る偽物から、現在ではフォトリアルな動画や、わずか数秒の音声サンプルから本人と区別がつかない声を複製できるレベルにまで急速に進化しました。この変化が意味するのは、ビジネスや社会の根源的なプロセス、すなわち「証拠に基づく事実認定」が根本から揺るがされているという現実です。これこそ、我々が直面する脅威の新たな本質なのです。

2　攻撃者の「民主化」と「嘘つきの配当」という新たな戦場

私が特に注目するのは、この脅威の中核にある「アクセシビリティ（利用の容易さ）」です。かつてはハリウッドレベルの専門知識と数億円単位の予算を必要とした映像合成が、今やスマートフォンアプリやウェブサービスを使い、無料または安価に誰でも行えるようになりました。

この偽情報作成能力の「民主化」は、潜在的な攻撃者の数を爆発的に増大させました。もはや脅威は、国家や大規模な組織に限定されません。企業の内部情報を知る不満を抱いた従業員、競合他社を貶めようとする者、金銭目的の犯罪者、社会を混乱させたい活動家、あるいは単なる愉快犯、そして一般の個人までもが、強力な偽情報兵器を手にしているのです。

さらに深刻なのが、この技術進化がもたらす「嘘つきの配当 Verantwortung（Liar's Dividend）」と呼ばれる現象です。人々が「どんな映像や音声も偽造可能だ」と広く認識するようになると、今度は本物の証拠までもが信じられなくなるという問題が発生します。例えば、ある企業の不正行為を捉えた本物の内部告発映像が公開されても、「あれは捏造されたものだ」の一言で片付けられてしまうかもしれません。この現象は、真実を否定するための便利な口実を提供することで、不正を働いた側に利する結果となります。それは、政治、司法、ビジネスのあらゆる場面で、検証可能な証拠という概念そのものを侵食し、社会の説明責任の文化を根底から揺るがす深刻な脅威なのです。

この問題に拍車をかけるのが、作成と検出の間に存在する深刻な「非対称性」です。説得力のあるディープフェイクを「作成」する方が、それを「検出」するよりも指数関数的に容易かつ安価になっています。検出ツールは、生成ツールとの終わりのない「イタチごっこ」を続けており、恒久的な解決策とはなり得ません。安易な技術的解決策への期待は、もはや現実的ではないのです。

3　静かなる侵食 - 経営を揺るがす偽情報のリスクシナリオ

この脅威は、決して抽象的な社会問題に留まりません。既に、具体的かつ甚大な損害として確認できる事例が報道されていますのでここでいくつか紹介させていただきます。

■　**経済的損害と金融詐欺：**

企業詐欺:香港で多国籍企業に勤務する会計担当者が、ビデオ会議で最高財務責任者（CFO）を装った相手にだまされて、計 2 億香港ドル（約 38 億円）を詐欺グループに送金する事件。詐欺グループは、巧妙な手口で会計担当者をだましてビデオ会議に出席させ、出席者全員が同僚だと思い込ませましたしかし実際は全員が、動画や写真を加工して本人のように見せかけるディープフェイクで作り出した偽者でした。（2024 年 2 月に発生）

市場操作： ハッキングされた AP 通信の X アカウントから「ホワイトハウスで爆発があり、オバマ大統領が負傷した」という偽の速報が流れた際、株式市場は一時的に約 1390 億ドルもの価値を失いました。AI が生成したとされる国防総省（ペンタゴン）付近での偽の爆発画像も、小規模ながら市場に影響を与えていたとされます。（2013 年 4 月）

消費者詐欺： SNS や WEB 上の公告で、著名人が投資を推奨しているかのような偽の広告動画や、AI が生成した魅力的なプロフィールで相手を騙す高度なロマンス詐欺も増加しています。2024 年 6 月には「ホリエモン（実業家の堀江貴文氏）が優良株を教える」とかたるフェイスブック（FB）上の広告から誘導され、投資詐欺被害に遭った男性が、広告業者が詐欺の首謀者に送金用口座を提供し加担したとして、約 1 千万円の損害賠償などを求めた訴訟がありました。

■　**社会・政治的操作：**

選挙介入： ロシア政府とつながる影響工作ネットワーク「ドッペルゲンガー」が、2024 年の米国大統領選挙を標的とし、バイデン大統領の偽音声を使って予備選挙での投票を妨害しようとした事件や、台湾やトルコの選挙でも、候補者攻撃にディープフェイクが使用され印象操作が行われました。同団体は今でも暗躍しています。

国家間の情報戦： ロシアによるウクライナ侵攻では、ゼレンスキー大統領が戦いをやめようと降伏を呼びかけるディープフェイク動画が拡散されるなど、情報戦の実験場と化しています。（2022 年 3 月）

■　**個人・企業の評判毀損：**

名誉毀損： 有名人や政治家をターゲットにしたポルノグラフィックなディープフェイクは深刻なハラスメントであり、日本でも作成者が名誉毀損容疑で逮捕される事例が発生しています。同様の手法が、企業の経営陣や従業員に向けられるリスクは計り知れません。

偽情報攻撃： 競合他社や悪意ある第三者が、ある企業の製品に関する虚偽の噂を流布したり、公式アカウントになりすまして偽情報を発信したり、偽のプレゼントキャンペーンの当選 DM を送ることで、詐欺や外部のサービスケースの登録に誘導するなど乗っ取り被害も確認されています。AI は、特定の個人や団体を攻撃するための誹謗中傷コメントを大量に自動生成するためにも悪用されかねません。

SNS 上で表示された回数（インプレッション）に応じて広告費が支払われるようになると、世界中で無作為に投稿にコメントをするアカウントが激増しました。当初は絵文字だけ利用したものでまだ言語の壁がありましたが、生成 AI の登場によりあたかも一個人のような文章を投稿するようなりました

https://x.com/hkazano/status/1758141777343844804

これらの事例は、偽情報がもはや単なる「噂」ではなく、直接的かつ大規模な経営リスク

へと変貌を遂げた冷徹な現実を示しています。

4　対症療法から戦略的レジリエンスへ - 「多層防御」という思考

ディープフェイクや偽情報を見破る方法は徐々に困難になっています。そのため偽情報の拡散後にそれを追いかけるファクトチェックやコンテンツ削除は、拡散速度が真実を上回る現代においては、常に後手に回る「事後対応型」のアプローチです。それは、絶え間ない「消火活動」に他なりません。

今、真に求められているのは、根本的な思考の転換であり、これまでの「問題が起きてから対処する」事後対応型ではなく、「問題の発生そのものを未然に防ぐ」予防型のアプローチへと移行しなければなりません。その実現には、複数の防御策が相互に連携して機能する、強靭な「多層防御エコシステム」の構築が重要です。

これは、セキュリティやリスクに対する高い意識と実践を日々から行い、その認識が社会全体に習慣として根付かせる試みをしていくことです。

■　人間（Human Cognition）- 第一の防衛線：

組織における最大の脆弱性は、スパムメールや嘘のメッセージをなんの疑いなく開いてしまうような社員です。しかし、これは裏を返せば、個人が偽情報に対する第一の防衛線になり得ることを意味します。重要なのは、情報を鵜呑みにせず、発信元（信頼できる報道機関か、匿名の SNS か）、一次情報（元の論文や公式発表は存在するか）、複数の情報源（他のメディアも報じているか）、そして情報の日付を確認するなど、物事を多角的に段階的にチェックする習慣です。客観的な事実と個人の意見を区別し、強い感情を煽る情報にこそ注意を払う。そして最も重要なルールは、怪しいとおもったら一度止まる、相談をするとことです。人は信じられないことや驚くようなニュースを目にすると思わず人にシェア（感情を共有したい）したい心理が働きやすく、これが、個人が加担者になることを防ぐ最後の砦となります。

■　組織プロセス（Organizational Process）- 組織的防御壁：

企業にとって、偽情報は経営リスクそのものです。従業員への定期的なメディアリテラシー研修は、経営層も含めて実施すべき、最も費用対効果の高い投資の一つです。同時に、自社のブランド名や役員に関する言及を監視する「ソーシャルリスニング」を導入し、脅威を早期に発見する体制が求められます。そして、危機が発生する「前」に、迅速かつ透明性の高い情報発信を定めた「危機管理広報計画」を策定しておくことが極めて重要です。沈黙や対応の見極めは非常に難しく、ほとんどの場合は状態を悪化させる可能性がたかいです。CEO 詐欺のような脅威に対しては、多額の送金指示に多要素認証を義務付け、ディープフェイクが疑われる場合は別の手段で真偽を確認する「内部セキュリティ規程」の徹底が不可

欠です。

■　**テクノロジー（Technology）- 技術的基盤：**

テクノロジーは脅威であると同時に、対抗手段も提供します。日本では日本ファクトチェックセンター（JFC）[2]のような専門機関が活動しており、その検証結果は重要な判断材料となります。また、偽情報対策における極めて重要な新興技術が、「コンテンツの来歴証明」です。Adobe、Microsoft などが参加する業界標準「C2PA（Coalition for Content Provenance and Authenticity）」は、画像や動画が「いつ、どこで、どのように」作成・編集されたかをファイル自体に埋め込む技術で、いわばコンテンツの「デジタル出生証明書」として機能します。これは完璧な検出器ではありませんが、信頼できる情報を見分けるための強力な手がかりを提供します。

これらの機関に関して検証の結果が速報ベースでででることは難しくリアルタイムで判断することはできないものの、これまでの事例を確認することで、過去の教訓や、事案を参考にすることで、一度立ち止まることにつながることになります。

5　結論　「共有された現実」の崩壊後、企業は未来を選択できるか

生成 AI の進化はまだ序盤であり、本稿で提示した脅威やまた新しいリスクが登場する可能性は十分にあります。我々が予測する 2030 年に向けた未来の脅威は、さらに深刻な様相を呈します。

- **ハイパー・パーソナライゼーション：** 将来の生成 AI は、個人の心理的特性や過去の行動履歴に合わせて偽情報をカスタマイズし、その説得力を飛躍的に高めるようになると予測されています。
- **偽情報の自動化と大規模化：** AI によって、偽情報キャンペーンの作成と展開が、人間のモデレーターでは到底対応不可能な規模と速度で自動的に行われるようになるでしょう。
- **現実と合成の境界の消失：** 技術が向上し続けることで、人間の目では本物と偽物の区別が事実上不可能になります。

これらの脅威が行き着く先は、深刻な「共有された現実の崩壊」です。アルゴリズムが生成する情報によって人々がそれぞれ全く異なる「現実」を生きるようになれば、社会は分断されてしまいます。そうなれば、「何が現実か」という共通の土台が失われ、気候変動から経済政策に至るまで、社会が直面するいかなる重要課題も解決できなくなるでしょう。これ

[2] https://www.factcheckcenter.jp/

は、民主的な議論や社会的な合意形成の基盤そのものが失われることに他なりません。

しかし、物事やトレンドにも周期があるように、この高度なテクノロジーが引き起こす問題の解決策が、最終的には「人間中心のスキル」へと回帰するという可能性も考えられます。それは、批判的思考、情報源の確認、自らのバイアスの自覚といった、古くからある知性です。私たちの思考を欺くように設計されたテクノロジーへの最良の防御策は、私たち自身の「思考の質」を高めること、物事を一歩引いて俯瞰する、すなわち「どのように考えるかを考える力（メタ認知）」を養うことに尽きます。

この新たな現実を前に、すべての企業とビジネスリーダーは、自らの役割と責任を再認識する必要があります。未来の不確実性に備え、事業を守るためには、表面的な情報を追うだけでは不十分です。今こそ、無数の情報の中から真のシグナルを捉え、地政学リスクや市場動向と結びつけて未来を洞察する、高度なインテリジェンスが不可欠です。それは、様々な情報を多角的に精査し、点と点をつないで本質を見抜く力と言えるでしょう。

こうした状況を踏まえ、生成 AI の目まぐるしい進化に乗り遅れないよう、企業には継続的な学習と適応が求められます。また、必要に応じて専門家のアドバイスを迅速に得られる体制を整え続けること。それこそが、これからの時代を勝ち抜くための、企業の永遠の課題となるに違いありません。

《著者略歴》
大成 昂（おおなる あきら）　株主や取引先のキーパーソンの把握など、個人を対象としたバックグランド調査に従業。2020 年から Open Source Intelligence を利用した調査を開始。国内外に限らず、SNS のソフトウェアの仕様（通称：API）を通じて取得できる、対象の写真や投稿内容から読み取れる言動や思考、SNS 上のやり取りから判別できる風評収集や人的交流の精査に強みを持つ。生成 AI の応用能力を証明する「Google　AI　Essential 資格」保有者。