経営リスクとしての誤/偽情報
―マルチステークホルダー連携への期待―

三井物産戦略研究所
産業社会情報部産業調査室
菊池一真

Summary

- 生成AIの進化による精巧な誤/偽情報の氾濫は、企業の株価や信用を脅かす重大な経営リスクである。世界経済フォーラムも最大級のリスクと警告しており、経営レベルでの対応が不可欠となっている。
- 誤/偽情報は真実性と害意の強さに応じ、誤情報（Misinformation）、偽情報（Disinformation）、悪意ある情報（Malinformation）の3つに分類される。ただし、これらの境界はあいまい化している。
- 個人のリテラシー向上や個社での対策には限界があり、官民の枠組みを超えたマルチステークホルダーによる対策が国際的な潮流だ。有識者は、この枠組みの社会実装の主導役として、強力な海外網を持つ総合商社などの企業に期待を寄せている。

1. 緊要性を増す誤/偽情報への対策

誤情報や偽情報（以下、誤/偽情報）への対応が、企業経営において緊要性を増している。生成AIの目覚ましい進化に伴い、真偽の見分けがつかない映像や文章が容易に生成、拡散されるようになった。これらの中には、特定の意図をもとに受け手を誘導し、企業に看過できない経済的損失をもたらす事例もある。

世界経済フォーラム（WEF）は「グローバルリスク報告書」の中で、誤/偽情報を2024、2025年の2年連続で今後2年間の最大リスクに位置付けた。最新の2026年版でも地経学的対立に次ぐ2位となり、依然として世界の政財界のリーダーの深刻な懸念事項であることを物語る[1]。また、米調査会社ガートナーは、2028年までに、企業の誤/偽情報対策支出が300億ドルを超えると推計する。同社は「虚偽の情報が組織に重大な財務リスクやレピュテーションリスクをもたらす」と警鐘を鳴らしている[2]。もはや、誤/偽情報は、企業の広報部門やリスク管理部門の課題にとどまらず、株価や信用を損なう重大な経営リスクとなっているといえるだろう。

[1] World Economic Forum(2026年1月14日)Global risks report 2026。回答者の年代別では、40歳未満は誤/偽情報を最も重大なリスクと見なしている（p.18、figure15）。
[2] ガートナージャパン(2025年10月30日)Gartner、2028年までに誤情報／偽情報対策への企業支出が300億ドルを超えるとの展望を発表。

本稿は、誤/偽情報による被害事例を概観し、現時点の対策の潮流を整理することを目的としている。

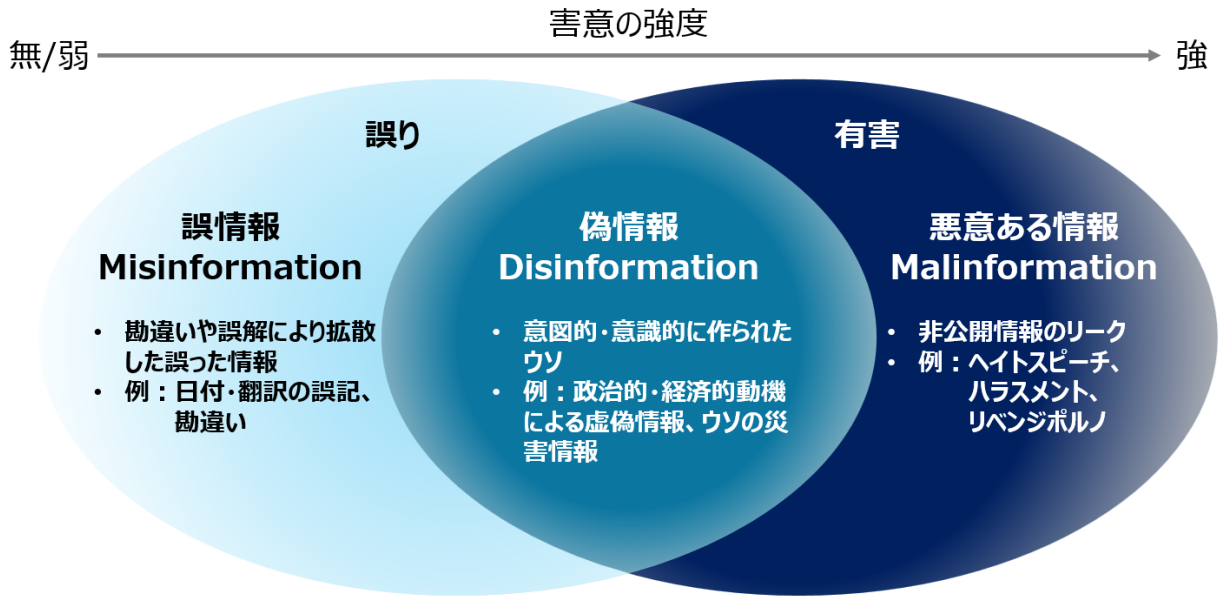
2. 誤/偽情報の定義と被害事例

2-1. 誤/偽情報とは

一口に誤/偽情報といっても、その情報の文脈や背景、受け手次第で真偽の判断が変わりうる。従って、専門家間でも「定義や用語の使い方について意見が一致しておらず、統一した用語体系が存在しない」[3]とされる。そのなかで、本稿では、情報の真実性と害意の強さに着目して誤/偽情報を整理したワードルらの定義を援用する[4]。企業が誤/偽情報に直面した際、事実誤認への訂正対応で済むのか、あるいは組織的な攻撃として法的措置を含む全社的な対応が必要なのかを判断するにあたり、この分類が有用であると考えるからだ。ワードルらは誤/偽情報を以下のように3分類する。

- **誤情報（Misinformation）**…虚偽であるが、害を与える意図で作成されたものではない情報。
- **偽情報（Disinformation）**…虚偽であり、個人、社会集団、組織、または国に害を与える目的で意図的に作り出された情報。
- **悪意ある情報（Malinformation）**…事実に基づくこともあるが、個人、組織または国家に損害を与える目的で利用される情報。しばしば、非公開情報のリークによりもたらされる。

図表１　誤/偽情報の類型と例



注：誤情報、偽情報、悪意ある情報の境界があいまい化していることを踏まえ、グラデーションで表現した
出所：Wardle & Derakhshan(2017)、山口真一（2022)『ソーシャルメディア解体全書―フェイクニュース・ネット炎上・情報の偏り―』、総務省（2025)「インターネットとの向き合い方～ニセ・誤情報にだまされないために～第2版」から三井物産戦略研究所作成

[3] ジョン・ルーゼンビーク ＆ サンダー・ヴァン・ダー・リンダン(2025年)『現代誤情報学入門』、加納安彦訳、日本評論社、p. 2。
[4] Wardle, C. & Derakhshan, H (2017年) Information disorder: Toward an interdisciplinary framework for research and policy making、pp. 20-21。

図表1は、ワードルらの分類に基づき、「誤情報」、「偽情報」、「悪意ある情報」を害意の強度順に整理したものである。各分類をグラデーションで表現しているのは、AIの進化により情報の出所や作成意図が判別しにくくなり、これらの境界があいまいになっている現状を反映するためである。

直近では、生成AIが誤/偽情報の脅威を質・量ともに増幅させている。ネット情報の評価機関である米ニュースガードによると、人の編集を介さないAI生成のニュースサイトが、2025年10月時点で英語、中国語、アラビア語など16言語・2,089件存在する[5]。ディープフェイク技術も高度化しており、日本政府も「これまで以上に注意深く情報に接することが重要」であると注意を促す[6]。実際、同年11月に公開されたGoogleの画像生成AI「Nano Banana Pro」は、政治的なメッセージや有名ブランドに対する虚偽の拡散を目的としたプロンプトを拒否せず、指示通りに精度の高い偽画像を生成したとの報告もある[7]。

このように、生成AIの普及はコンテンツの大量生成という量的側面と、表現の精巧化という質的側面の両面で誤/偽情報のリスクを増幅させているといえる。

2-2. 多様化する被害事例

前述のような背景から、誤/偽情報による被害はその手法や影響範囲が多様化している（図表2）。

まず、SNS上の「なりすまし」が株価を直撃したのが、2022年11月の米製薬大手イーライリリーのケースだ。Twitter（現X）上で公式を装って投稿された「インスリンが無料になったことをお知らせできること

図表２　誤/偽情報の主な被害事例

発生年	対象企業・業界	誤/偽情報の内容	主な影響	毀損された経営資産
2016年	**ニューバランス** （米・スポーツ用品）	• 同社幹部の通商政策に関する発言が、**文脈を離れて「トランプ大統領支持」として切り取られ拡散** • さらに**極右系サイトが「白人の公式シューズ」と勝手に位置付け**	不買運動、若年層を中心とした支持離反	ブランド価値・顧客基盤・売上
2022年	**イーライリリー** （米・製薬）	• 公式アカウントを装う**偽アカウントが「インスリンが無料になった」と投稿**し、SNSに偽情報が拡散した	株価 ▲4.37% 下落	株主価値・時価総額
2023年	**ターゲット** （米・小売）	• **「悪魔崇拝の服を販売している」とする偽画像がSNSで拡散**。従業員への脅迫などが発生した	売上下落、店舗運営コスト増（警備・対応）	売上・人的資本
2024年	**アラップ** （英・エンジニアリング）	• **ビデオ通話を用いたディープフェイク詐欺。同社幹部を精巧に模倣したAI映像**を用い、信じ込ませた従業員へ多額の送金を指示	2,500万ドルの送金被害	財務資産 ・ガバナンス（内部統制）
2025年	**日本の** **インバウンド業界**	• **「2025年7月に日本で巨大地震が起きる」という予言が香港などアジア圏で拡散** • **出典は1999年出版のマンガ**とされる	消費機会損失の可能性	売上

出所：各種報道から三井物産戦略研究所作成

[5] NewsGuard(2025年10月24日)Tracking AI-enabled misinformation: Over 2000 undisclosed AI-generated news websites (and counting), plus the top false narratives generated by artificial intelligence tools。
[6] 内閣官房「偽情報にだまされないために」外国による偽情報等に関するポータルサイト。
[7] NewsGuard(2025年12月3日)Google’s new AI image generator is a misinformation superspreader。

を嬉しく思う」という偽情報は、インスリン価格高騰への社会的不満を背景に拡散。その後、同社の株価が4.37%急落した[8]。また、2024年には、英エンジニアリング大手アラップの香港事務所が、ディープフェイク技術を用いた詐欺により2,500万ドルをだまし取られた[9]。犯人は、ビデオ通話上で同社のCFOに精巧になりすまし、同事務所の従業員に送金を指示したという。

さらに、業界全体が風評被害を受けたのが、日本のインバウンド業界の事例である。「2025年7月に日本で巨大地震が起きる」という科学的根拠のない流言がアジア圏のSNSで広がった。訪日旅行のキャンセルによる消費機会の損失は、約5,600億円に上るとの試算もある[10]。

一方、誤情報への迅速な対応により、損失を回避した事例もある。2024年11月、「チロルチョコに虫が入っていた」という動画付きのSNS投稿の拡散に対し、チロルチョコの公式アカウントが事実関係の調査と説明を迅速に行った。虫の侵入が製造段階ではなく、投稿者の保管中に起きたものであることを示して誤情報の打ち消しに成功した[11]。

ただ、生成AIにより説得力ある偽情報が大量生成できるようになった現在、人手による常時監視と都度の「初期消火」では、コストとスピードの両面で限界があるだろう。

3. 対策の潮流

3-1. 誤/偽情報の「予防」と「対処」

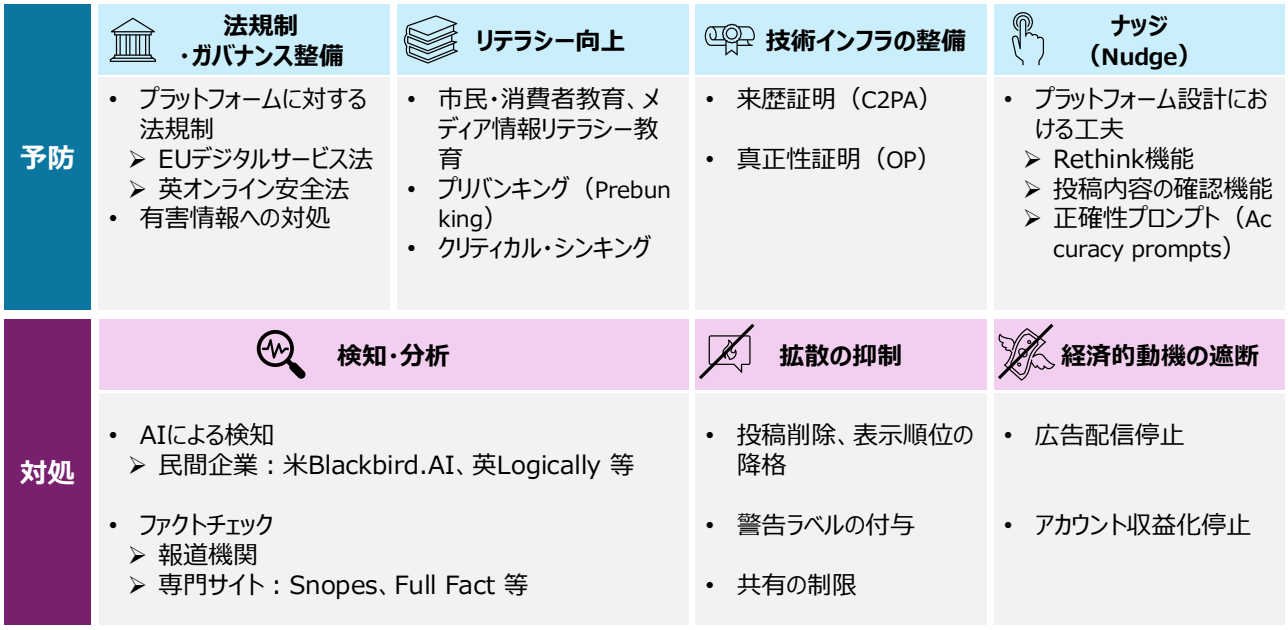
誤/偽情報への対応は、発生前の「予防」と、発生後の拡散を抑え込む「対処」に大別できる（図表3）。

まず、予防の柱は四つに集約される。第一に、法規制・ガバナンスの整備である。EUのデジタルサービス法（DSA）や各国のSNSに対する規制などが該当する。第二に、リテラシー教育やプリバンキング（Prebunking）といった、受け手の抵抗力を高める方策である。プリバンキングとは、誤/偽情報に接する前に、その事例の学習や事前の警告を行い、受け手に心理的な「免疫」を獲得させる手法である。第三に、コンテンツの来歴や発信者の真正性を技術で証明するといった、技術インフラの整備である。後述する「C2PA（Coalition for Content Provenance and Authenticity）」や「OP（Originator Profile）」が該当する。第四に、行動経済学の知見であるナッジ（Nudge）の活用である。ナッジとは、小さなきっかけを与えることで、人々の行動変容を（良い方向に）促す手法のことである。例えば、SNS投稿時に、「本当にこの内容で投稿してよいか」などと確認をとったり、再考を促したりするプラットフォーム設計の工夫が挙げられる。

---

[8] Forbes(2022年11月12日)Fake Eli Lilly Twitter account claims insulin is free, stock falls 4.37%。引用元の記事でも言及されているように、株価はさまざまな要因の影響を受けている点に留意する必要がある。
[9] CNN(2024年5月17日)British engineering giant Arup revealed as $25 million deepfake scam victim。
[10] 野村総合研究所(2025年5月29日)堅調なインバウンド需要に水を差す科学的根拠のない7月の大規模自然災害の憶測：5,600億円規模の経済損失試算も、木内登英のGlobal Economy & Policy Insight。
[11] 城戸譲(2024年11月8日)チロルチョコ「虫混入？」騒動対応が見事すぎた訳―迅速な対応と、消費者コミュニケーションの妙―、東洋経済オンライン。

図表３　主な誤/偽情報対策



出所：ルーゼンビーク & ヴァン・ダー・リンダン(2025)、山口(2022)、Kozyreva, A. et al.(2024) “Toolbox of individual-level interventions against online misinformation” などから三井物産戦略研究所作成

次に、すでに生成された誤/偽情報の拡散を抑え込む対処のフェーズは、主に三つのアプローチに分けられる。第一の検知・分析では、AIを活用して拡散の兆候を早期に捉え、ボットや詐欺組織などの発信元の特定を試みる。民間企業のAIを活用したサービスや、報道機関によるファクトチェックが代表例である。第二の拡散の抑制では、誤った内容の投稿の削除や表示順位の降格、警告ラベルの付与を行う。ここでは、表現の自由を尊重しつつ、透明性の高い手続きで過度な拡散を防ぐことが求められる。第三の経済的動機の遮断では、広告配信や決済の遮断を実施する。偽情報サイト運営者への資金流入を断つことで、経済的な動機を奪うことが目的である。

3-2. 企業の枠を超えた取り組みの重要性と今後の展望

前述の多様な対策の中でも、とりわけ技術インフラの整備への重要性が増している。生成AIによる誤/偽情報の大量生産に対し、個人のリテラシー向上や人手による事後的な対処では限界を迎えているからである。こうした背景から、コンテンツをいつ、誰が、どこで作ったかを示す来歴を技術的に証明し、デジタル空間の信頼を回復しようとする取り組みが、マルチステークホルダーの連携によって進められている[12]。

その代表例が、マイクロソフトやアドビ、インテル、BBCなど世界的なIT・メディア企業が主導するC2PAだ[13]。これは、デジタルコンテンツに対し、編集方法や場所、日時などの来歴情報を改ざん不可能なデジタ

[12] WEFの2026年版報告書でも、誤情報や社会の二極化を防ぐために、マルチステークホルダーの連携が重要であるとされている（p.39）。
[13] C2PA | Verifying Media Content Sources

ル署名として埋め込む技術規格である。特定のプラットフォームに依存しないオープンな規格であることが強みだ。

日本でも、新聞社や広告企業などが中心となり、OPの研究・開発が進められている。これは、ウェブサイト上のコンテンツに対し、「信頼できる発信元である」という証明書を付与する仕組みで、ネット上の記事や広告の信頼性を可視化することを目指している[14]。

もちろん、技術の導入だけですべての誤/偽情報の問題に対処できるわけではない。今後の対策のあり方について、情報経済論を専門とする国際大学グローバル・コミュニケーション・センターの山口真一教授は「健全な情報空間を作るためには、（C2PAやOPのような）単一技術の普及にとどまらず、多数のアクターが多面的に検討し、社会実装を進める体制が重要だ」と指摘する。その上で、「技術開発は技術系のプレイヤーが担う一方で、社会実装は海外に強力なネットワークがある総合商社などの企業が標準化に役割を果たせるのではないか」と期待を寄せる。こうした役割分担と連携を、国際的なルール形成や実証事業を通じて具体的な枠組みとして定着させられるかが今後の鍵だ。そのなかで、各国政府や企業、メディアなどを横断する高度な調整力を持つ総合商社は、旗振り役を担うことができるだろう。

何を信じてよいのかわからない――。そのような時代だからこそ、マルチステークホルダーの連携により、信頼の基盤となる技術を社会へ浸透させることが求められている。

[14] Originator Profile 技術について