SNSをはじめとしたデジタル空間の健全化への取組み

株式会社エルテス　リスクコンサルティング本部
マーケティング部 マーケティングGr マネジャー　奥村（おくむら）　高大（たかひろ）

1. コロナ禍での「人間×デジタル空間」の変化と問題

コロナ禍を契機にデジタルトランスフォーメーション（DX）が加速している。ビジネスでは、物理的な距離を問題としないオンライン会議やオンライン接客など、非対面コミュニケーションが増加。リアルからデジタル空間へのコミュニケーション代替が進行している。B2Bのみならず、B2Cマーケティングもオンラインの活用が著しく、今後ますますデジタル空間でのレピュテーション（評判）の重要度が高まることが予想される。そのため、デジタル空間を支配している論調を把握することは、企業経営の意思決定にとって不可欠である。

しかし、デジタル空間における論調を正確に把握することは、極めて難易度が高い。検索エンジンでの情報収集だけであれば、論調の把握と検証はある程度可能だ。だが情報が猛烈な勢いで流れていくSNSでは、正確な論調をつかむのは困難だ。なぜなら、発信者の意図と異なる曲解が積み重なり、事実と異なる情報が拡散するケースが後を絶たないからだ。それにより、デジタル空間特有のネット炎上や風評被害というリスクにつながってしまう。

新型コロナウイルスの感染拡大においても、インフォデミック（情報の氾濫）によって、多くの人々が不安や恐怖にさらされたことは記憶に新しい。例えば、日本では「マスクの材料に回される」「中国から原材料を輸入できなくなる」といった臆測がSNSで拡散され、トイレットペーパーが品薄になった。また、イランでは「度数の高いアルコールを飲めば、体内のウイルスが死滅する」というSNS上のデマを信じた人々が、メタノールを混ぜた酒を手にしてしまったことで、メタノール中毒に陥り多数の死者が出た。インフォデミックへの注意喚起として、WHOでは新型コロナウイルスの特設ページに「迷信や不安に対するアドバイス」というコンテンツを設け、不確かな情報に惑わされないためのアドバイスを掲載した。

また、拡散されたデマのなかには、論文や他者の投稿の一部を引用しながらも、フレーズを切り取って再構成する際に投稿者のバイアスがかけられていると思しき投稿も散見された。例えば、「緑茶に含まれるエピガロカテキンガレートは抗ウイルス作用が高い」という情報が、「新型コロナウイルスの治療に有効」と変質し、拡散されてしまった例がある。これは、表現の切り取りや言い回しの変化によって情報の本質が損なわれたと考えられる事例といえるだろう。

SNS上のコミュニケーションは、短い文章でのやり取りとなる。そのため、言葉足らずで本質が伝わらなかったり、連続したツイートの一部を切り取ることで意味が変質してしまったりと、伝達の過程で情報に歪みが生じる傾向が強い。また、読み手も膨大に流れてくるタイムラインを流し読みすることで、慎重に情報を吟味せず誤った解釈をしている可能性がある。膨大な情報にさらされる現代人は、要約された情報を消費することに慣れており、流れた情報を鵜呑みにしてしまうケースが少なくない。

エルテスでは、新型コロナウイルスに関連する情報の分析を実施。すると、一度SNSで拡散された不確かな情報が、メディアのファクトチェック記事などによって鎮静化した後に、再び掲示板や別のメディアなどで広まる例が散見されることが分かった。多様化するデジタル空間の中において、一度ファクトチェックを実施しただけでは、不確かな情報が是正されず、再燃する可能性があるといえる。こうした分析から、不確かな情報が一度拡散されると、デジタル空間で次々に変質し、バリエーションを増やすという問題点が浮き彫りになった。

発信者の文章力不足、受け手の読解力不足により、情報が曲解される可能性が高いばかりか、デジタル空間では不確かな情報が場所を変えて何度も再発するケースが多い。こうしたことから、一度拡散されてしまった情報をリアルタイムに把握するのは、ますます困難になっているのだ。

2. デジタル空間の歪みがリアルの世界に影響を及ぼす危険

世界中のデジタル、モバイル、そしてソーシャルメディア上での人々の動向や傾向を分析したレポート「DIGITAL 2020：GLOBAL DIGITAL OVERVIEW」では、2020年の年初には、45億人以上がインターネットを利用、ソーシャルメディアの利用者は38億人の大台を突破したと報告して

いる。世界人口の60％近くがインターネットを利用していることになり、2019年と比較すると7％増加している。また、平均的なインターネットユーザーは毎日6時間43分をオンラインで過ごしており、睡眠時間を1日8時間とすると、起床時の40％以上に及ぶ時間をインターネットの利用に費やしていることになる。

このレポートからも、デジタル空間への接続時間は年々増加傾向にあることが見てとれる。さらに、コロナ禍が拍車を掛け、対面の機会が絞り込まれたことにより、オンラインでの情報収集活動はより活発になっていると考えられる。しかし、デジタル化は人々の生活を便利にしている一方で、新たなリスクも生み出している。事実、接続時間の増加に比例するかのように、ネット上での誹謗中傷も拡大傾向にある。投稿内容の悪質さにより、逮捕や訴訟に発展するケースも多々見受けられるようになった。また、匿名による心無い誹謗中傷が人命を奪ったと考えられる事件も続発。デジタル空間のモラルを問う声が、日に日に高まっている。

デジタル空間がリアル社会と別物であるという考え方は今や昔。デジタル空間における人格と、それに対するレピュテーションは、良くも悪くも本人の実存にとって切り離せないものになっている。こうしたことから、デジタル空間に歪みが生じた場合、リアルの世界にも影響を及ぼす危険が高まっていると考えられる。

これは、個人の問題にとどまらず、企業についても同じことが当てはまる。デジタル空間特有の歪みのリスクに対し、企業経営者は然るべき備えをしておく必要がある。そこで重要なのが、意思決定の基になる情報を加工、分析したインテリジェンスを持つことだ。ネット上で誹謗中傷や炎上が発生したときに、多くの人はそれがあたかも世の中の論調だと捉えがちである。しかし、本当に全体の論調なのか、一部の人だけが騒いでいるのかを正しく判断しなければ、的確な手立てを講じることができない。ちなみに、日本の戦国時代には、戦いにおいて兵力を多く見せるため、人形を用いた戦術があったといわれている。これと同じように、SNSでも裏アカウントなどを使って、同一人物が誹謗中傷を展開しているだけで、実態は一部の人だけが騒いでいるケースが少なくない。風評被害は自然災害のようなもので、発生を未然に防ぐことは不可能である。しかし、事象を正しく捉えることで、被害の最小化を図ることは可能だ。リアルの世界に悪しき影響を広めないためにも、デジタル空間のモニタリングをはじめとするインテリジェンスが、あらゆる企業に求められている時代だといえよう。

3. ネット炎上・風評被害対策領域でのエルテスの実績

エルテスは、「デジタルリスクと戦い続ける。」というポリシーを掲げ、デジタルリスクマネジメントの専門家集団として、多様なデジタルリスクを解決するためのソリューションを開発している。情報通信インフラ技術とデジタルデバイスの発展に伴い普及した検索エンジン、SNS、オンラインバンキングなど、社会のDXが進む過程において発生するデジタルリスクマネジメントを支援。例えば、事業環境の変化とともに顕在的なリスクとなった企業のソーシャルメディアの運用に関するリスクマネジメントソリューションを包括的に提供している。これまで、NTTドコモやマツダ、サントリーなど上場企業をはじめ、1000社以上のデジタルリスクマネジメントに関連するサービスを提供してきた。

デジタルリスクマネジメントは、大きく2つに分類される。1つは「ソーシャルリスクマネジメント」である。SNS利用者の増加に伴い“事故”が頻発するようになり、ネット炎上の発生件数は2011年以降、毎年増加傾向にある。その事故を防止する打ち手として、ソーシャルリスクマネジメントでは、潜在リスクや業務改善の把握、ルール策定を行う「調査分析/体制構築」、リスク顕在化時の早期検知と初動対応を素早く行う「運用」、リスク低減支援にあたる「対策」の3つのフェーズに分けて支援を展開している。

まず、「調査分析/体制構築」は、企業や商品・サービスのインターネット上の情報を収集・分析し、レポートとして納品するサービスだ。競合比較を通じたマーケティング分析、海外での情報漏洩や事件・事故情報の収集も提供している。まずは、クライアント企業に関連するWeb上の記事を網羅的に収集。ポジティブ、ネガティブなどの事前に定めた要件に基づき記事を分類する。さらに、製品の細かな部分の評判や、潜在リスクを洗い出し、今後の取組みを分析し、レポートにまとめて提出する。また、公式SNSを新たに導入する際に必要な運用規定やマニュアルの策定、策定してから数年が経過したものの現在の状況に適合できていないソーシャルメディアポリシーなどの改訂を支援し、SNS運用の体制構築をフォローする。

続いて「運用」では、Webリスクモニタリングをメインに実施している。企業や商品・サービスに係る風評や自社従業員による情報漏洩リスク、不正広告や薬事法、景品表示法に係るリスクなどの特定事象に関するインターネット上の情報を24時間365日監視。万が一、緊急性の高い情報が検知された場合は、緊急通知と対応手法に関するコンサル

ティングを実施している。

そして、最後のフェーズにあたる「対策」については、Webリスクモニタリングを提供しているクライアントに対して、リスクが検知された際に専任のコンサルタントが当該事案のリスクアセスメントをシームレスに実施。危機が顕在化したコンテンツの信憑性、影響力を分析し、情報発信者のプロファイリングを行う。さらに、リスク検知後のエスカレーションフローに関しても、プレスリリース作成や記者会見トレーニングなど、必要に応じた危機管理広報対応の支援を実施し、その後の対応に関するアドバイスを行っている。また、クライシスに発展した場合は、危機管理広報対応のコンサルティングサービスを提供している。さらに、検索エンジン評判対策として、ユーザーのブランド体験やレピュテーションを形成する大きな主要因である、企業や商品・サービスに関する検索エンジン上の見え方に関する課題を抽出。課題解決と目的達成を実現するための手段をプランニングし、KPIを設計する。

ここで、Webリスクモニタリングを導入された企業の事例を紹介したい。食品業のA社では、2016年に食品業界内でSNSの炎上が話題となったことを受けて、「対岸の火事ではない」と危惧。そこで、「24時間365日、休日問わずリスク検知できる体制であること」「専任スタッフとのスピーディーな連携が可能であること」を評価され、エルテスのWebリスクモニタリングを選定。現在までに効率的なリスクモニタリングが実現でき、さらにはA社に関わる日々の投稿に対する肌感覚が意思決定時の直観力につながっているとのこと。また、ネガティブな事象の発生時だけではなく、CM放映後の反響を知るための情報収集など、リスク以外のモニタリングサービスにも応用している。

Webリスクモニタリングは、自社の商品やプロモーション活動に対するSNSのモニタリングだけにとどまらない。サービス業のB社では、SNSに投稿されるリスク投稿の検知だけでなく、顧客から寄せられるサービス提供に関する称賛の声をSNSから拾い、社内表彰に活用している。

エルテスは、ここまで紹介してきた「ソーシャルリスクマネジメント」サービスに加えて、ログを横断的に分析し、行動分析から社内のリスク行動を検知する「インターナルリスクマネジメント」の2つのアプローチで、企業のデジタルリスク対策に取り組んでいる。

4. AI活用による高度なネット炎上・風評被害対策を実現

エルテスのWebリスクモニタリングサービスは、品質向上のためにAIを導入している。しかし、AIにすべてを任せるのではなく、AIと人間それぞれが持つ強みを融合させることで、サービスの品質、効率性を向上させていくことを目指した。そこでまず、ネット炎上・風評被害対策サービスであるWebリスクモニタリングでは、AIが、投稿をネガティブ、ニュートラル、ポジティブの3つで判別する仕組みの実装に取り組んだ。だが、この取組みには大きく3つの課題が存在した。

(1) 正確なAI判定には、大量かつ正確な教師データが必要
(2) 言葉は流行り廃りが存在し、アップデート（メンテナンス）が必要
(3) 文脈・文章の意味を正確に読み取ることの難易度が高い

最初に、(1) の課題を解決するべく、2011年のWebリスクモニタリングサービス開始から蓄積し続けたデータを基に、AIで投稿をネガティブ、ニュートラル、ポジティブの3つで判別する教師データの作成に取り掛かった。教師データの作成には、ネガティブ、ポジティブの判別に際して必要な「当たり前」を教え込む必要があるが、同一投稿であっても、業種、業界によって、ネガティブ、ニュートラル、ポジティブの判定が異なってしまう場合が存在するため、「当たり前」の選定が難しいことが分かった。ニュートラルと判定される「当たり前」の投稿を集めることは比較的容易で、ネガティブ、ポジティブな投稿を大量に準備することは、想定以上に困難を極めたが、最終的には、長年のサービス提供経験が質の高い教師データの作成に大きく貢献した。

また、(2) の言葉には流行り廃りがあるという課題については、人間の介在により解決を図っている。例えば、「頭がおかしい人」という意味の「あたおか」は、お笑い芸人の使用をきっかけに一般に広がり、Petrelが発表した2019年上半期のインスタ流行語大賞で1位を獲得した。「あたおか」は、数年前までは一般的な言葉ではなく、AIがネガティブと判定することはなかっただろう。このように、SNS上の言葉には、流行が存在するため、AIがネガティブ、ポジティブと判定する教師データにも絶えず言葉の流行り廃りを反映する必要がある。これに対し、エルテスでは、言葉が持つ意味や感覚を、人間の感性で捉え、的確に教師データに反映している。

最後に、(3) の「文脈・文章の意味を正確に読み取ること」がAIは苦手であるという課題について。文章を細かな単位に区分して意味を取り出す形態素解析では、助詞によって大きく意味が変わる文意をくみ取る難易度が高い。例えば、「AよりBが好き」といった“他社商品との比較”に関する投稿があった場合、ネガティブ、ニュートラル、ポジティブの判定が難しい傾向にある。また、「ずっと行きたかったお店に行けず残念」のように、ポジティブな単語を交えたネガティブ投稿や、普段使用されない表現での投稿は、AIでの判定が現状では困難と考え、人間の目視で判別することとした。

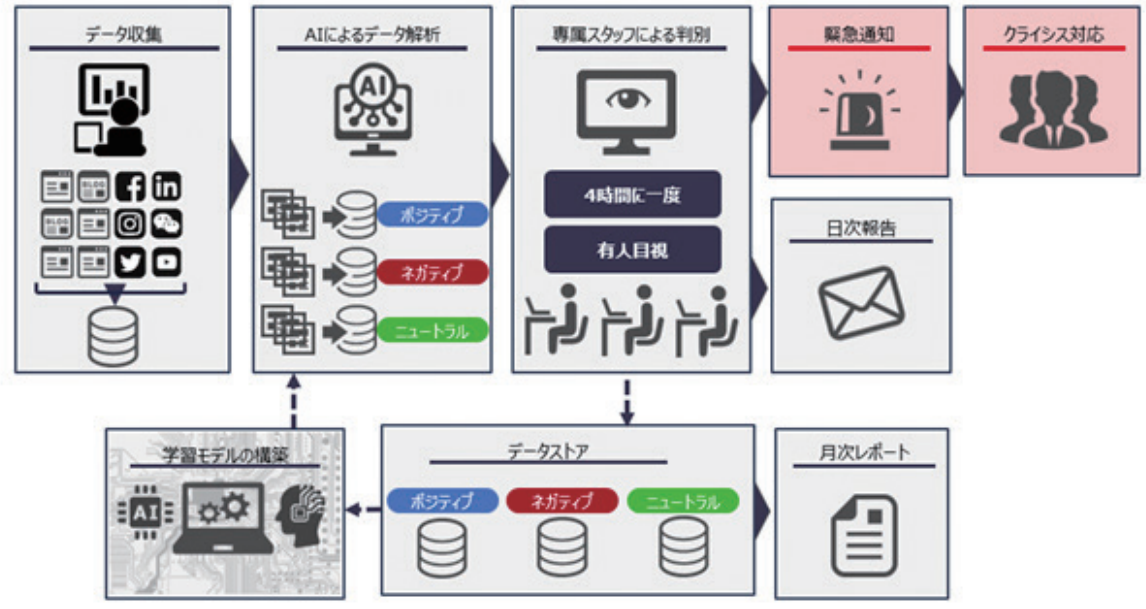
5. エルテスがたどり着いたデジタルリスク対策の最適解

Webリスクモニタリングサービスの価値は、早期にリスク投稿を検知することにある。言い方を変えれば、ネガティブ投稿を見逃すことはあってはならない。AI導入以前の目視のみの判別においても、80 ～ 90%の投稿はニュートラルに分類されるものの、「(3) 文脈・文章の意味を正確に読み取ることの難易度が高い」という課題が残っている以上、AIにWebリスクモニタリングを完全に任せきるのは難しいと判断した。そこで、AIにはニュートラルな投稿をスクリーニングさせる役割を任せ、AIが苦手な「文脈・文章の意味を正確に読み取ることが難しい、ポジティブやネガティブを意味する単語が入り交じる投稿」などネガティブである可能性のある投稿はニュートラルに含めず、専任担当者が目視で判定を行うフローを構築した。

結果として、企業のネット炎上・風評被害につながり得る機微な投稿にのみ、集中的に人の目を入れることで、ヒューマンエラーの防止、速やかな炎上検知、素早い緊急通知を行える環境を構築。結果として、サービスの品質向上を実現している。また、人の目で判定されたネガティブ、ポジティブな情報をAIに学習させ、言葉の流行り廃りを反映させた上で、AI判定の精度の向上につなげている。

また、同一の投稿内容が行われても、対象企業、社会トレンドによって、企業のレピュテーションへの影響は千差万別である。エルテスから企業へのリスク投稿の緊急通知を行うだけでなく、専任のコンサルタントが初期対応のサポートを行っている。

デジタル空間に存在する表面的な数字や文字だけで、リアル社会に与える影響を安易に判断することは難しい。人によって生み出されるデジタル空間の歪みだからこそ、エルテスではAIだけでなく、人の強みを上手く組み合わせ、デジタルリスクに立ち向かっている。そして、これからも絶えず変容を遂げるデジタルリスクと戦い続けていく。



■図. エルテス社の「Webリスクモニタリングサービス」の概要
様々な事象に関する投稿を24時間365日監視、危険度の高い内容を含む投稿が確認された際は、緊急通知や対応コンサルティングまで実施