経済産業省　第７回 AI 原則の実践の在り方に関する検討会
議事概要

令和５年２月３日
１０：００～１２：００
オンライン開催

「AI の信頼性」に起因する新たなリスクと企業に求められる対応について（Citadel AI 様事例）に対する質疑応答

- **貴社のサービスは AI の利用者側を想定したものであるように思われるが、AI の開発者側も、このサービスを活用されているか。また、異常を 100%検知することは難しいと想像されるが、それを顧客にどのように説明しているか。**

  → お客様は、AI の利用者に限らず、AI の開発者やシステム監査のご担当者の場合もある。また、異常検知については、確かに 100%の保証は難しいが、従来の手作業による方式では、膨大な時間とコストがかかり、テストも十分に実施されておらず、ビジネスやコンプライアンス上の大きな問題につながるリスクがある。弊社サービスをご利用頂くことで、現状と比較すると圧倒的に速く正確に異常を検知し、リスクを防御できるというメリットがある。また、自動化が可能な工程を徹底的に自動化することで、エンジニアなどの貴重なリソースを、より難易度が高く自動では行えないシステムやモデルの改善に充てることができる。

- **未学習領域の検知について、軸を見つけることは重要であるが、非常に難しい。そこを自動化できるということなのか。もし、完全な自動化ではなく、人が介在している場合、画像のアノテーション作業はどのように行っているか。**

  → 未学習領域の自動検知について、例えば、表形式データでは特徴量がデータの中に明示されているため、その特徴量を自動的に組み合わせながら、未学習領域を検出している。画像に関しては、弊社システムは、画像からコントラスト、シャープネス、明るさを自動検出し、それを一つの特徴量として利用することができる。さらに画像の場合は、メタデータが付いていることも多いため、そのメタデータと特徴量を組み合わせながら、自動で未学習領域を検出することができる。

- **入出力だけで判断するという発想が非常に斬新であると感じたが、アルゴリズムの分析やデータの偏りのチェックなどは行わないのか。**

  → データの偏りについては、AI に判定させる前の入力値の分布等を見ながら、バイアスがかかっていないかなど、入力値の分析を自動検証する。さらに、入力に対するモデルの出力結果を見ながら、自動で比較検証しモデルの弱点を検出する。モデルを検査する既存のツールは色々とあるが、モデルの内部情報を必要とする方式であると、AI の技術進展が早く、検査ツールの開発が追いつかないのが実態である。そのため、モデルの内部構造自体を分析するのではなく、入出力から判断することにより、技術進歩に大きく左右されずに、寧ろさまざまな AI のフォーマットやアプリケーションに対して、汎用的に評価ができるようになる。

- **法律に対する適合性の判断は、どのように行っているか。**

  → 法律や規格に書かれていることに対して、どのようなテストをどこまですればよいかということ自体、現状では判別が難しく、さらにそれを技術的に実現するツールが無いというのが実行上の大きな課題である。当社では、こうした課題に対し、主要な標準や規格に沿ったテストを自動検証する手段を提供する。ただし、そのテスト結果を見て、最終的にどのように判断するかは、従来もそうであったように、人が決める必要がある。例えば医療機器であれば国の機関が管理したり、あるいは各企業の事業戦略や経営方針によるケースもあり、そのセンシティビティやシビアリティといったリスクの度合いによって、どこが判断すべきかは変わってくる。

以上