

AIはなぜその答えを導き出したのか ～根拠を見える化する「説明可能なAI」～

FUJITSU JOURNAL / 2018年12月27日



現在、AI（人工知能）は第3次ブームの真っ只中にあります。そこから導き出される答えは右脳のような直感的なものであり、なぜその結論にたどり着いたのかという根拠の説明はできず、ブラックボックス化しています。今後、AIをより広範な社会実装が可能なアプリケーションに発展させていく上で不可欠なのが、「説明可能なAI」です。

AIの判断結果を論理的に説明することは可能か AIの第3次ブームをもたらしたディープラーニング

AIは1950年代から1970年代にかけて第1次ブーム、1980年代の第2次ブームを経て進化してきました。しかし、そこで追求してきた「人間（専門家）の知識を機械（コンピュータ）に教え込む」というアプローチは極めて困難なことから、思い描いていたような実用化には至らず、しばらく「冬の時代」が続きました。

そこにブレークスルーをもたらしたのが「ディープラーニング」と呼ばれる技術です。人間の脳を構成するニューロン（神経細胞）を模した機械学習モデルをコンピュータ上に実装し、実行するもので、大量のデータからコンピュータ自身が知識を獲得することを最大の特徴としています。

現在の第3次AIブームは、まさにこの技術によって巻き起こったものにほかなりません。実際、ディープラーニングをベースとした現在のAIは、囲碁の世界チャンピオンや将棋の名人も打ち破るなど、人間をはるかに凌駕する能力を示し、世の中に衝撃を与えました。

ブラックボックスのAIにはクリティカルな業務を委ねられない

ただ、このAIの技術を企業の様々な業務にすぐに活用できるかというと、なかなかそうはならないのが現実です。

囲碁や将棋などの世界で問われるのは、「勝つか負けるか」であり、AIが途中で「なぜその手を打ったのか」という理由は分からなくても、対戦相手に勝つことさえできれば、その結果をもって価値を評価することができます。

しかし、このようなブラックボックスのAIに、果たしてミッションクリティカルな業務を委ねることができるでしょうか。「AIがなぜその判断を下したのか」について、納得できる説明を得られなければ業務への適用は難しいといえます。

例えば、体の不調を感じて医師の診察を受けた際に、問診のほか血液検査やMRIなどの結果に基づいて、「あなたの脳には動脈瘤があり、なおかつ直径7mmに達しており、このままではくも膜下出血になる危険性が高いため、早期に手術が必要です」と説明されれば、その事実を受け入れることができます。それでも納得できない場合には、医師に過去の経験、論文や治療情報などを基にした論理的説明を求めます。

将来、高度なAIが医師を支援する時代になり、「AIによるとあなたは今すぐ手術が必要です」と何の根拠もなく医師に言われても納得することはできません。

高度な社会受容性を備えた、すなわち「信頼できるAI」でなければ、多くのビジネスや生活では通用しないのです。

そうしたこともあり、現時点でのAIの応用分野は、画像認識や一般会話レベルでの音声認識・翻訳などの限定的な範囲にとどまっています。これらはいずれも先の囲碁や将棋における「勝ち負け」と同様に、結果だけを見て「合っているか間違っているか」を比較的容易に判断できるアプリケーションです。社会実装で求められるアプリケーションとなるまでには、まだまだ高いハードルが立ちはだかっています。

現状のアプリケーション



社会実装で求められるアプリケーション



システム化して社会に根付かせることが必要



今後のAIに求められること

人間からの信頼を担保するために必須となる「説明可能なAI」

そもそもAIといえども、導き出す答えが必ずしも正しいわけではありません。今AIと呼ばれているもののほとんどは、まだ何も知らない「赤ちゃん」と同じです。うまく育てることができれば賢くなりますが、育て方を間違えると実用に堪えるものにはなりません。役立つ「大人」にするには、どんなテクノロジーをベースにしているかだけでなく、育て方や与える食事（データ）も非常に重要であることを認識しておく必要があります。これがAIの現実です。

例えばクレジットカードの与信をAIで行うシステムを構築したとします。そこで学習に用いるデータに偏りがあれば、AIが導き出す結果にもバイアスがかかり、特定の属性を持った人物に不利な審査結果を示すことになりかねません。

実際、あるITベンダーが、ユーザーと会話しながら学習していくAI型のチャットボットをインターネットに公開したところ、人種差別や性差別、暴力表現など悪意を持った会話が繰り返され、結果として意図しない方向に学習が進み、極めて不適切な発言を連発するようになりました。わずか半日ほどで停止に追い込まれたケースもありました。

社会実装で求められるレベルのAIのアプリケーションを構築するためには、その判断（思考）のプロセスが、倫理規定やコンプライアンスにも沿った形で行われていることを実証できなくてはなりません。どのようなロジック、あるいはどういった経緯から結果が導き出されたのか、人間がプロセスを理解し、管理し、間違っていたなら修正できるようにする必要があります。

こうした人間の信頼を担保できるのが「説明可能なAI」であり、AI自身で「説明責任を果たせる」ことがその大前提の条件となります。これがあって初めて、人間はAIが導き出した新しい発

見や洞察に対しても信頼を置くことができ、ひいてはそれがAIの提供範囲の拡大や高度利用など、さらなる改善（発展）につながっていきます。

ディープラーニングはAIのブレークスルー技術
しかし、ブラックボックスのため社会実装できる領域が極めて限定的



「説明可能なAI」ができること

「説明可能なAI」に向けた2つのアプローチ

具体的に「説明可能なAI」はどのようなアプローチによって実現することができるのでしょうか。現在、世界において研究が進められているのは大きく次の2つの方法です。

1つは、これまで学習能力は非常に高いが理由の説明は苦手とされてきたブラックボックスAI、すなわちディープラーニングに説明機能を持たせるものです。例えば画像データの認識や分類を行う際に、学習のプロセスから特に大きな重みづけが行われた部分や領域をニューラルネットから抽出し、判別の理由として示します。

もう1つは、ホワイトボックスAIです。文字どおり判断の仕組みが分かる学習モデルを備えたAIであり、結論を導き出すまでのプロセス（計算過程）をトレースして示すことで、人間にとっても比較的理解しやすい理由を説明することができます。

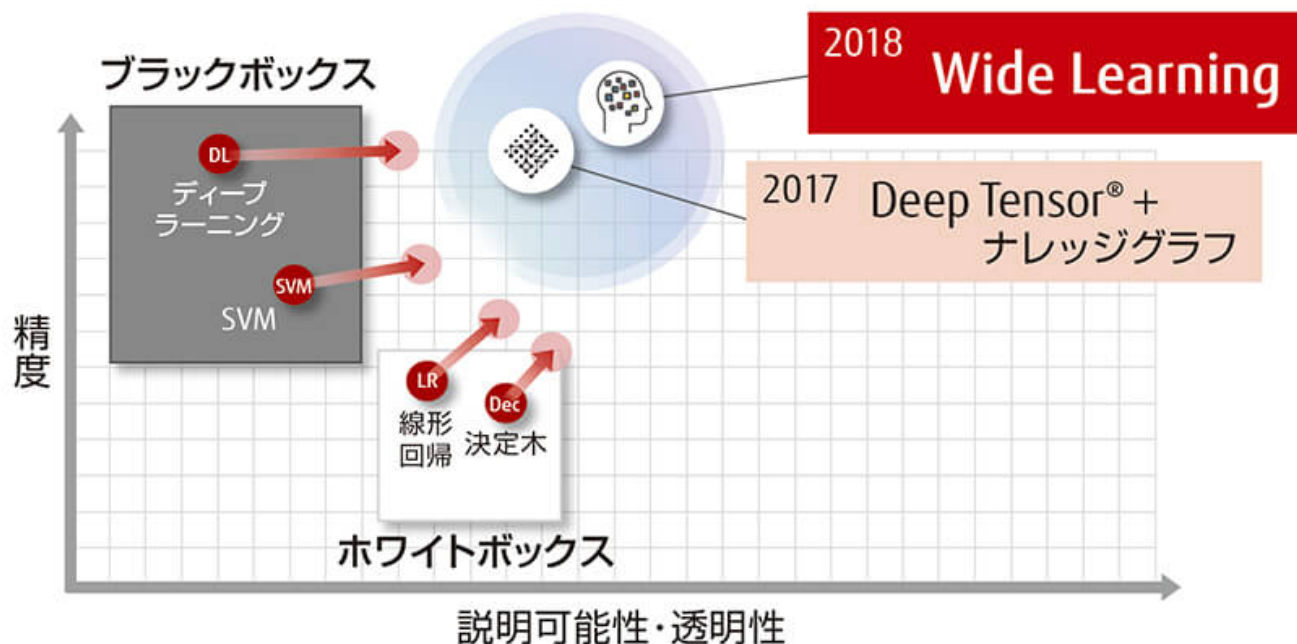
ブラックボックス

学習能力は高いが
理由の説明は苦手

VS

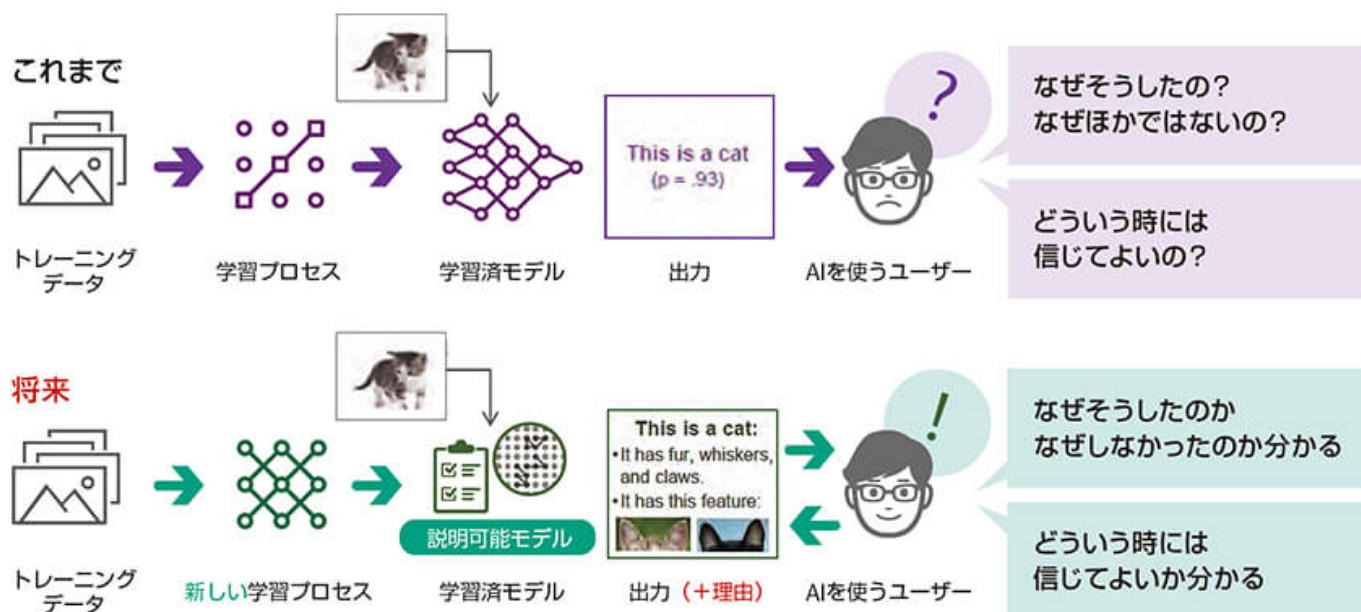
ホワイトボックス

判断の仕組みが分かるモデル、
理由の説明が可能



「説明可能なAI」の要素

また、より人間に理解しやすい「説明可能なAI」を実現すべく、例えば動物の画像を読み込ませた際に「毛皮とひげ、爪を持ち、さらに耳の特徴から"猫"と判断した」といった言葉（自然言語）による説明を行えるようにする機能も検討されています。こちらの研究はまだ始まったばかりであり、今後の成果が期待されるところです。



「説明可能なAI」により「なぜそうしたのか」が分かるようになる

推定に至った「理由」「根拠」を示すディープテンソル+ナレッジグラフ

富士通はディープラーニングをベースとしたAIに関して世界でも最先端の研究に取り組み、豊富な成果を有しています。そして、実用段階にある画像・音声・手書き文字などの認識系のみならず、時系列データや事象の「つながり方」や「関係の特徴」に着目して高精度な学習を行うグラフデータにまでディープラーニングの適用領域を広げています。

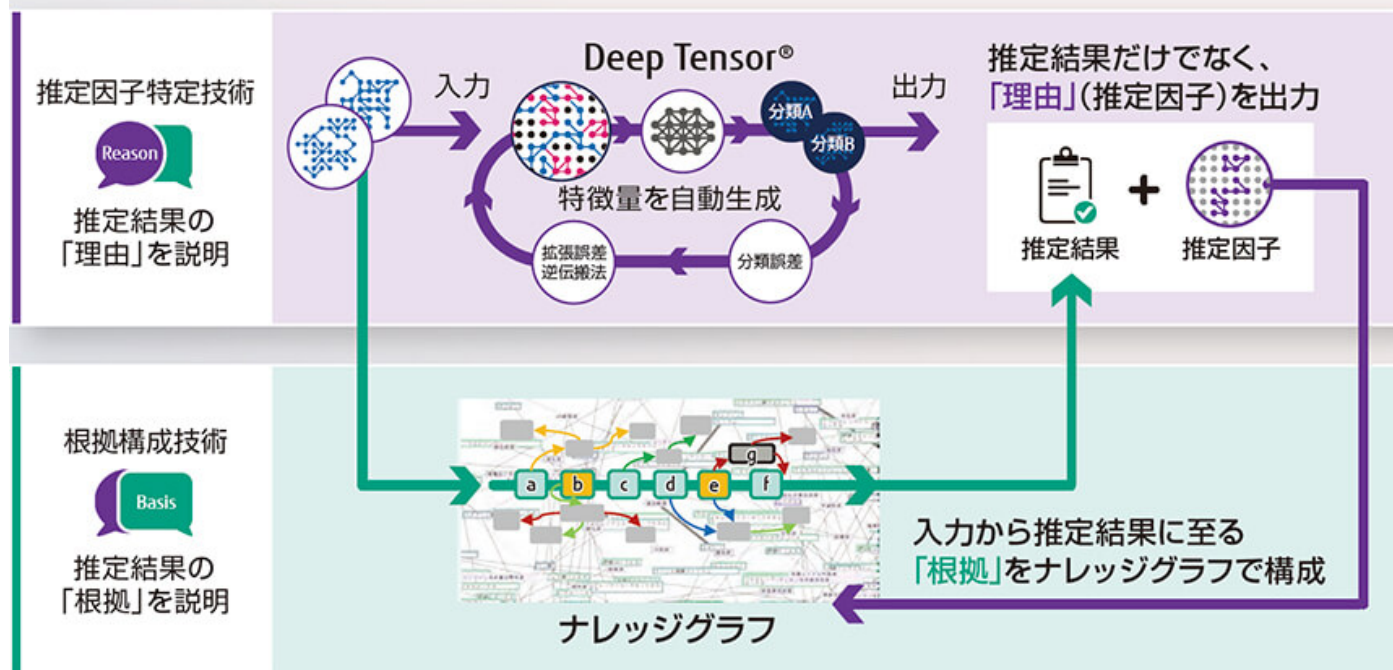
そうした中で、富士通は「説明可能なAI」の実用化に向け、独自技術の研究開発を大きく前進させています。

その1つが、「DeepTensor[®]（ディープテンソル）」ならびに「ナレッジグラフ」という2つの新技術を組み合わせた「人が信頼・理解・管理できるAI」です。

ディープテンソルとは、従来のディープラーニングにテンソル（行列やベクトルなどの概念を一般化した多次元の配列で表現したデータ）という表現を加えて進化させたもので、人やモノのつながりを表すグラフ構造のデータから新たな知見を導き出します。具体的には、ディープラーニングの学習効率を飛躍的に向上するとともに、結果を導き出すまでに至った「理由」も示します。

人が信頼・理解・管理できるAI

2つの説明を行う全く新しいAI: 結果の「理由」と「根拠」を説明



説明可能なAI : ディープテンソル+ナレッジグラフ

例えばある1日の通信ログデータをAIに読み込ませたところ、「何らかのサイバー攻撃があった」という結論が出た場合、ログのどこを見て攻撃と判断したのか、IPアドレスやポート番号などの推定因子を抽出し、「理由」として提示します。

一方の「ナレッジグラフ」は、学術文献などの様々な情報源から収集した情報を意味付けしてつなぎ合わせたグラフ構造の知識ベースです。

ディープテンソルから導き出された結果とナレッジグラフに蓄積された知識をひもづけることで、推定に至った「理由」と併せて、その裏付けとなる「根拠」を示すことが可能となりました。これにより人間の専門家はAIが導き出した結果が信頼に値するものかどうかを確認するとともに、そこから新たな洞察（気づき）を得ることができます。専門家がAIと協調しながら問題を解決するという、これまでにない形のAI活用が実現されるのです。

膨大な仮説の組み合わせを漏れなく検証するワイドラーニング

ただ、「ディープテンソル+ナレッジグラフ」によるAIも、基本的にはディープラーニングの延長線上の技術であり、大量のデータを用いて学習させなければ精度が上がらない、ひいては信頼性の高い説明ができないという課題があります。

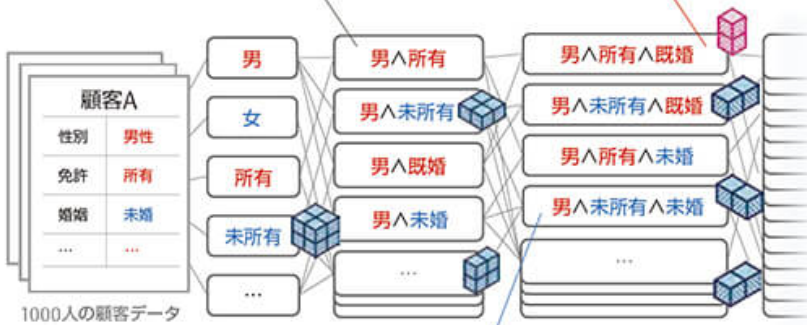
そこで富士通は、十分な学習用データがなくても高い精度で判定できる新たな機械学習技術を開発しました。「WideLearning（ワイドラーニング）」と呼ぶホワイトボックスAIの技術がそれで、従来のディープラーニングのように1つのデータモデルを深掘りして学習を行うのではなく、複数のデータ項目を網羅的に組み合わせで判断することで、データ量が少なくても判定精度を高めることを可能としました。

1 データ項目を組み合わせて大量のナレッジチャンクを抽出

膨大な組み合わせパターンからヒット率の高い仮説(ナレッジチャンク)を高速に抽出

仮説: 男性・免許所有の人は購入
ヒット率60% (=100人のうち60人が購入)

仮説: 男性・免許所有・既婚の人は購入
ヒット率90% (=10人のうち9人が購入)

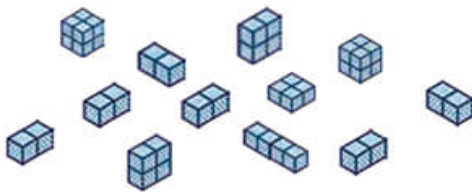


仮説: 男性・免許未所有・未婚の人は未購入
ヒット率90% (=20人のうち18人が未購入)

数千万、数億個のナレッジチャンク

2 影響度を調整して高精度な分類モデルを構築

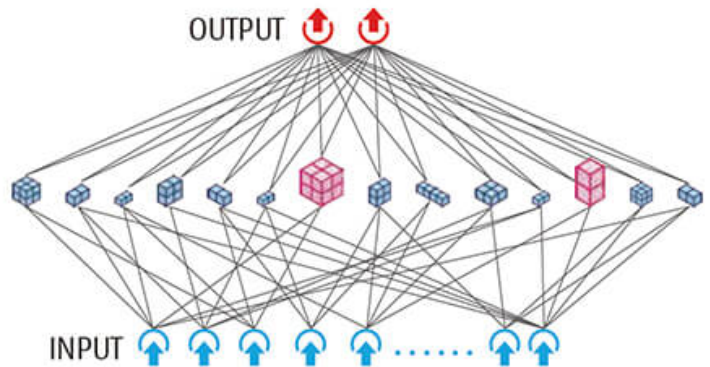
未購入のナレッジチャンクが多数



購入のナレッジチャンクが少数

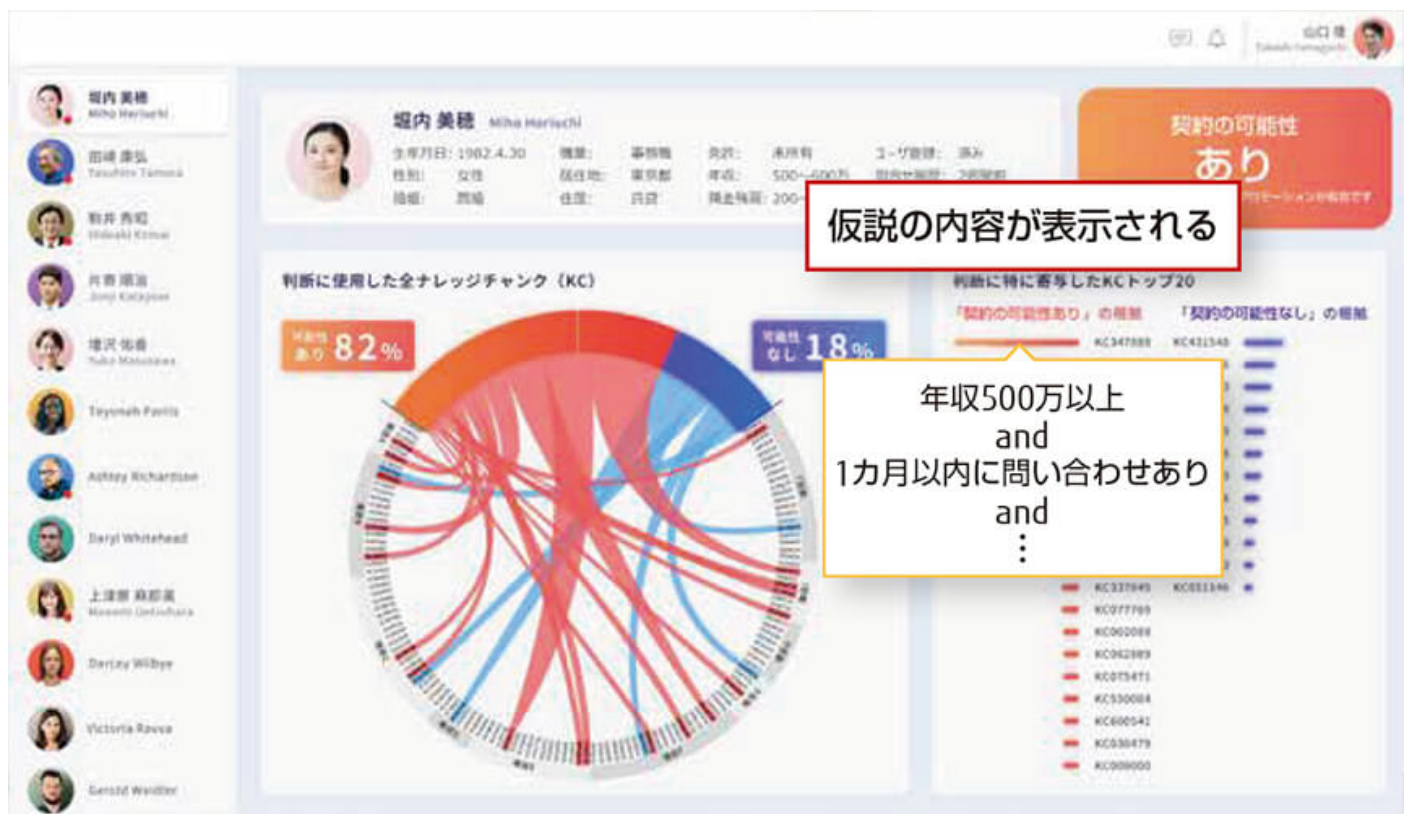


影響度を調整



ワイドラーニングの独自技術

例えば、デジタルマーケティングにおける商品ごとの購入傾向分析にワイドラーニングを適用した場合、「性別」「免許保有」「婚姻」「年齢」などのデータ項目の膨大な組み合わせのパターンを「仮説」とし、実際の商品購入者のデータとどのくらいヒットするか検証します。検証の結果、ヒット率の高い仮説(ナレッジチャンク)を重要な仮説として取り上げ、それらに基づいた購入分類モデルを構築します。顧客が購入するかどうかの判別に用いられたナレッジチャンクは列挙され、そのリストから任意の仮説を選択することで、「年収500万円以上」「1カ月以内に問い合わせがあり」といった中身(=説明)を確認することができます。



ワイドラーニングの活用例

人間とAIが協働する将来を見据え、さらなる研究開発を

こうした「説明可能なAI」は、AIの社会実装の可能性を大きく広げていきます。

「ディープテンソル+ナレッジグラフ」に注目すると、従業員の活動データから健康状態の変化を検知して働き方改革につなげていく「健康経営」、業績や経済指標などのKPIから企業の成長を予測する金融分野の「投資判断」など、すでに多くのミッションクリティカルなビジネスでの実践が始まっています。

例えばゲノム医療の分野では、ディープテンソルに18万件の疾患系の遺伝子変異データを投入した学習を行い、さらにナレッジグラフに1700万件の医療論文などから100億を超える知識を関連付けたAIシステムを構築しました。ここから得られた推定ロジックの流れを専門医師が再チェックすることで、分析からレポート提出までに要する時間を2週間からわずか1日に短縮するという絶大な効果を上げています。

一方のワイドラーニングもまた、従来のディープラーニングとの比較検証において、例えばデジタルマーケティングでは潜在顧客を見逃す確率を10～50%低減、医療判断支援では罹患者や疾患発生部位を見逃す確率を約20～30%低減するという効果を実証しています。

そして今後も、様々な機器や設備の重大障害を検知する予防保守、クレジットカード取引の不正使用の判断、製品故障の予兆検知、セールスプロモーションにおける潜在顧客の発見、商品コン

セプトの判断など、特にレアな事象の判断や高い透明性が求められる業務をターゲットとした新たな機械学習のアプローチとして、ワイドラーニングの早期の実用化を促進していく考えです。

想定するサービス → データ不十分な場合の判断や説明性が必要な業務



医療診断
疾患の早期発見



株式選定
上昇銘柄を判定



システム保守
重大障害を検知



クレジットカード取引
不正使用の判断



製品異常判定
故障の予兆検知



再犯予測
人物の特徴の予測



セールスプロモーション
潜在顧客の発見



新商品企画
商品コンセプトの判断

実用化が期待されている分野

さらに富士通は、こうした「説明可能なAI」が持つ無限の可能性に引き続き注目し、ユーザー企業における判断や意思決定支援の高度化、人間とAIの協働を含めた次世代のシステムの在り方などを見据えながら、グローバルで最先端を走るAIの研究開発を進めていきます。

富士通のAIビジネスを束ねるグローバルヘッドクォーター FUJITSU Intelligence Technologyが始動

富士通は、AIビジネスをグローバルに牽引する新会社「FUJITSU Intelligence Technology」をカナダのバンクーバーに設立し、2018年11月1日から事業を開始しました。富士通がこれまで日本を含む世界各地のリージョンごとに実施してきたAIビジネスを束ねる海外で初のヘッドクォーターとなり、AIの関連製品やサービス、データ、知見、ノウハウなどを集約。グローバル展開の戦略策定およびその実行を担っていきます。

具体的には、米国、EMEIA、アジア、オセアニアの各リージョンを、FUJITSU Intelligence Technologyを中核としたグローバルマトリックス体制の下、AIビジネスを推進していきます。シリ

コンバレーなどがある米国ではなく、なぜ今回カナダをAIビジネスの新たなヘッドクォーターに選んだかは、次のような理由があります。

1. カナダ政府およびブリティッシュコロンビア州政府は最先端のITを推進する施策を実施し、IT企業誘致に積極的である。
2. ブリティッシュコロンビア大学やトロント大学などAIや量子分野に強みを持つ大学・研究機関が地域に多数存在し、優秀な人材も豊富である。
3. 富士通がすでに出資している1QBInformationTechnologiesInc.をはじめ、最先端のITスタートアップ企業が多数起業している。
4. マイクロソフト、ボーイング、Amazonなど大手企業が進出しており、パートナーや顧客企業と共創するためのビジネス環境が整っている。
5. ほかの北米の主要都市と比べ、賃貸料や税金、労働費などがリーズナブルであり、企業運営にアドバンテージがある。
6. 北米は世界のAIビジネスの最大の市場であり、各企業のAI投資額も日本の17倍以上と非常に積極的である。

また、世界標準のスピードに対応するために、世界に通用する「商品開発の戦略企画」、自前主義から脱却した「エコシステムづくり」を2本柱として進めていきます。ビジネス規模としては、2018年度から2022年度までの5年間累計で4000億円を達成することを目指しています。

富士通は、共創×人材×テクノロジーによってデファクト・ソリューションをつくり上げ、ビジネスと研究の両輪を回し、AIビジネスを推進していきます。