信頼の構築：AI イノベーションにおける説明可能性と倫理

Fujitsu Laboratories Advanced Technology Symposium 2018

アナリストによるイベント概要および考察：Jessica Groopman

序文

人工知能とマシンの学習能力はデジタル時代の新たなる章の始まりを意味する。すなわち、大量の非構造データやソフトウェアの可能性に新たな息吹を吹き込み、インターフェースや顧客体験に大いなる変化をもたらす。どの点から見ても、我々はまだ AI 時代の初期段階にいるが、既にあらゆる場所で AI は使用されている。メディア・レコメンデーションからナビゲーションアプリ、音声や顔認識からサイバー攻撃の脅威分析まで、AI は 世界中の顧客、企業、政府関連市場にまたがる何百ものアプリケーションを動かしている。

しかしながら、人間の能力をマシンに与え、自動化が高まるに連れ、倫理問題、自己チェック、説明責任も生じてくる。どのように結果や決定の説明をするか、マシンの乱用を避け、社会に役立てるにはどうしたらよいか、マシンで置き換えられたり、コントロールを奪われたりすることなく、どのようにマシンをうまく使い、人間の能力を高めるかについて考えなければならない。さらに最も重要な点は、人間、企業、マシン間の信頼関係をどのようにして築くかである。

個人データの使用に関する懸念、アルゴリズムの偏りや差別、戦争における AI の使用、個人、ビジネス、社会の健全性や安全性に対する脅威が高まる中、各組織は積極的に前例のないレベルで疑惑や倫理問題を予測評価しなければならない。

イベント概観

上記のような、今日における最も差し迫った課題に対処するため、Fujitsu Laboratories Advanced Technology Symposium 2018 (FLATS 2018) では、説明可能な AI、AI の倫理問題、ビジネスや産業環境における影響や適用方法に焦点を合わせ、忌憚ない議論が交わされた。イベントは、カリフォルニア州サンタクララに 400 人以上が出席し終日に渡り開催され、AI 研究の第一人者であるマサチューセッツ工科大学 Dept. of Brain & Cognitive Sciences 教授 Tomaso Poggio 博士および、カリフォルニア大学バークレー校 Center for Human-Compatible Artificial Intelligence のエグゼクティブディレクターである Mark Nitzberg 博士両氏の基調講演を始め、産業界、学界の多様なリーダーによる研究成果、意見、成功事例の発表・交換を行う、非常に興味深い会議であった。

AI の急速な台頭はさまざまな障害および社会問題を提起

今日、組織の約 80 パーセントが何らかの形で生産に AI を活用していると回答しているが、会社や従業員たちは一点型アプリケーションの域を超える広範な価値の実現に奮闘している。課題は内外から提起されている。たとえば、データサイエンティストたちはデータの質 (および量) の必要性を力説している。セキュリティ管理者たちは かつてないほど動的なネットワークトポロジー (接続形態) を構築しようと競い合っているし、IT 部門は急ピッチでインフラを更新し、データガバナンス基準を適用しようとしている。重役たちはビジネスの結果を求めているし、法務部門や規制当局は監査可能性を要求している。従業員たちは自分たちへの風当たりや配置転換などを懸念しながら、新しいツールを模索している。

これらの背景には AI ブーム再来に対する誇大報道が原因している。その結果、ベンダー、学界、政府は断片的で無秩序な努力をすることとなり、一貫性のない規制制度を気にかけず、メ

ディアや社会の混乱を招き、倫理問題を引き起こしている。新しい技術は常に懐疑心を生じさせるが、AI の場合、人間の認知の理解と再現という根本的に独特な基盤に根ざしており、過大な期待と人間の脆弱性の影響を受けやすい。

大規模人工知能には広範囲に渡る説明可能性と倫理性が必要

AI の発展により企業は、自動化において説明責任を犠牲にすることができなくなった。AI における説明責任という主題は非常に広大だが、究極的には各企業が取り組まなければならない二つの重要な領域、すなわち、「説明可能性と倫理性」という問題に至る。

説明可能性：マシン (特にディープ) ラーニングネットワーク内部をチェックし、どの要因、層、次元、ノードが意思決定に重要な影響を与えているかを調べることはもちろん、どのような処理を経て、結論に至ったかを理解するための機能についてはまだ不透明で、よく理解されていない。DARPA (アメリカ国防高等研究計画局) Information Innovation Office のプログラムマネジャーで本イベントのパネリストである David Gunning 氏は「私たちは結論を得られるが、どうしてその結論に至ったのか、なぜ他の結論に至らなかったかを問うことができない。」と発言している。

シンポジウムでは、FICO、PwC、スタンフォード大学および、富士通の代表が産業における説明可能な AI の要件について発言した。機械による処理過程の自己チェック機能の欠如は、企業の立場から見ると、説明責任、法規制遵守、反差別、消費者保護、およびモデルの誤りに関するチェック能力の低さという重大な問題を引き起こす。モデルの修正、調整、デバッグが困難になる上、当該組織が公然あるいは不注意により、不正あるいは無責任な行為を行ったかについて知りたい外部組織からすると、これは非常に問題である。

この「ブラックボックス」問題は動的な問題でもある。なぜなら、データ、ユーザー、測定基準、法規則、およびセキュリティ要件は常に進化し続けているからである。パネリストたちはさらに、AI の全ての決定に関し説明する場合の (計算、公表競争、不正確さにより生じる) コ

ストや、潜在的な性能とのトレードオフについても議論した。Kyndi の CEO で本イベントのパネリストである Ryan Welsh 氏は、Fortune 500 に名を連ねる 100 社以上の CEO に対し彼の行ったインタビューを振り返りながら、「説明可能性 は AI を展開する上で一番の課題である」と述べている。

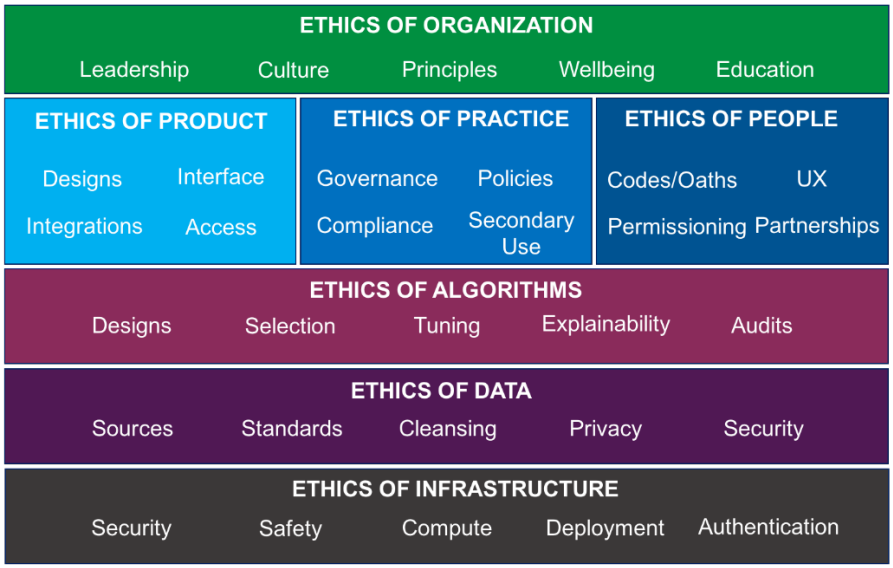
今日、説明可能性は金融、医療、その他規制の厳しい産業において深刻な問題である。この問題に対し、我々は調査を通して、説明可能な AI は全産業にとって、意思決定や説明責任レベル向上、新しい考え方の発見、顧客満足度向上に有益であるという結論に至った。デジタル世界で成功するには、各企業は従業員が AI と協業し、AI を管理できるようにしなければならない。説明可能性研究を行っている米国富士通研究所 Digital Life Lab ディレクターの Ajay Chandler は「**説明可能性は、**[さまざまなタイプの] ユーザーとマシン間の**信頼関係構築に必要不可欠である。**」と述べている。

倫理： 倫理は我々が決定を下す際の道徳基準である。AI はマシンがこれまで人間が行っていたタスクをうまく実行する能力の基礎となるため、きわめて多方面に渡る倫理問題が生じ、これらの問題は人間の主体性、自由、アイデンティティ、アクセス、公衆衛生、不正行為に関連する広範囲な社会問題に発展する。

組織の立場から見ると、これらの問題は偏見、差別、プライバシー、透明性、同意、コンプライアンス、顧客満足度、信頼関係崩壊などにつながる多様で予測困難なリスクである。まさに今年、世界中の多くの AI 研究者たちが、フェイスブック／ケンブリッジ・アナリティカ問題から Uber の自動運転による死亡事故のような倫理違反問題に巻き込まれた。大批判は世界中に広がり、法規制に対する関心や議論、従業員の反発を呼び、たとえば Google では従業員 4000 人が戦争関連技術開発の禁止請願書に署名した。

この広大な課題に対処するにはいわゆるビジネス・スタック全体にわたる緩和策や解決策を探る必要がある。

フレームワーク (基本構成)：倫理をビジネス・テクノロジー・スタックに統合



Source: Architecting Trust: Explainability & Ethics in AI Innovation Whitepaper

KALEIDO INSIGHTS

AI はビジネスの観点から見ると、決して過ちを犯してはならない重要機能である。ほとんどの産業の将来 (例えば、収益) にかかわる AI の重要性がこれらの取り組みの背後にある原動力である。組織が倫理的な AI 設計に失敗すればこの問題はますます深刻になる。

AI ベンダーたちはさまざまな方法で AI の説明可能性と倫理性を構築しようとしている

AI ブームは同時に、その説明可能性と倫理性確立を目指すさまざまな商業活動につながるのは当然である。過去 12 ヶ月間、世界の AI 技術関連大企業は倫理関連プログラムや基本方針、製品を完成しようと競い合ってきた。

プログラム：おそらく、それらの企業が共通して採った手段は産業コンソーシアムに加入することである。例えば、Partnership on AI は、社会における AI 関連研究、成功事例構築、公開講演を目的とする学際的ワーキンググループである。Google、Amazon、Microsoft、

Facebook、IBM、Apple などの企業が皆、そのようなグループに参加する一方、社内で Chief Ethics Officer や倫理委員会を設置している企業もある。Facebook、Microsoft、および法執行機関関連製品を製造している Axon は、専用チームを立ち上げて製品に関わる AI 倫理問題に取り組んでいる。

基本方針： ほとんどの企業が使命や指導的な価値を主張しているが、実際にはどの企業も AI に関する基本方針を構築していない。2018 年 6 月に初めて、この 10 年近く世界の AI 開発を先導している企業の一つである Google が「我々の研究や製品開発を積極的に管理し事業決定を左右すると思われる 7 つの基本方針」を発表した。それ以来、Microsoft、Uber、GE その他の企業が皆、独自の AI 基本方針を発表した。基本方針は重要不可欠な出発点ではあるが、それに対する責任、実行プロセス、および意欲の再確認をしなければ大きな効果は得られない。

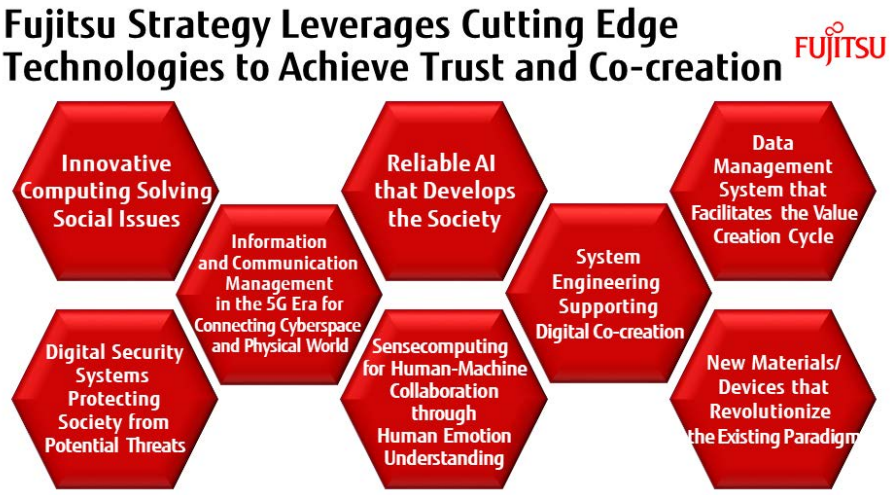
製品：多くの企業が製品そのものに関する倫理性や説明可能性確立にも取り組んでいる。Accenture の Fairness Tool は統計手法を採用し、性能および予測される同等性がどのように (バイアスに関して) センシティブな変数に関連するか分析することにより、アルゴリズムによる特定集団の人々に対する不公平な処理を検知できるようにした。Facebook や Microsoft も最近、AI システムをトレーニングするデータを分析し、特定グループの人々に対する偏見がないかを チェックする新しいツールを発表した。一方、IBM は世界最大の注釈つきデータセットをリリースすることにより各コミュニティーの認知バイアス (偏見や先入観による判断) 問題への取り組みを支援している。このデータセットはジオタグ (地理情報) 付で、肌の色、性別、年齢を等しい割合でトレーニングデータに分布し、ソース素材のバランスをとり、AI による誤った標本抽出を減少させることを目指している。

興味深いことに、これらの取り組みは成長著しいスタートアップ (Kyndi、Cognitive Scale、DarwinAI) や、企業・機関の活動 (例えば LIME、Generating Visual Explanations、DARPA’s XAI) に影響を与え、上述の活動で不足している倫理や説明可能性に関する欠陥解消に取り組んでいる。

もちろん、抽象的な倫理に関する声明、新規採用、一点型ソリューションなどはある程度魅力的ではある。しかし必要なのは、さまざまなアプリケーションコンテキスト (実行環境) 、ビジネスモデル、ユーザタイプの説明に関する基本的な枠組みと方法である。たとえば従業員の選別や常習的犯行チェックに AI を使用する際の倫理問題は自律型兵器あるいは医療診断に使用する場合の倫理問題とは異なる。また、これらの領域における技術への理解と信頼を得るためには、顧客から重役、データサイエンティストから医者まで、さまざまなユーザーモデル (ペルソナ) を作る必要もある。

富士通はアーキテクチャに重点を置いた AI 戦略を推進

ここで富士通が他企業や組織と異なる点は、テクノロジースタックの枠を超え、さまざまな種類のユーザーや顧客との非常に多くの革新的な技術開発に投資し、協業を推進することによりこれらの問題に取り組んでいる点である。



イベントで、富士通は 15 事例の展示を行った。これらの展示は単に先進コンピュータ技術を紹介するものではなく、知識共有、迅速な意思決定、説明可能性、および人間と AI の相補関係によって、技術がいかに信頼関係確立と協業を可能にするかを示すものだ。以下に際立った 2 つの技術について特筆する。

「ワイドラーニング」は性能を落とすことなく透明性を維持する。この画期的な技術 wide learning は 2 つの重大な問題に対処することを目指している。一点目はトレーニングデータが少ない場合でもそれらを学習、マッピングし、仮説を重要度に従って抽出・評価することにより、データ不足を補う。二点目は、結論に至った考え方の経緯や説明が他の技術と比べ卓越していることだ。たとえば、「25 才から 35 才で収入が 5 万ドル以上の女性が購入するだろう」という仮説は論理式として記録されるため、判断理由はデータサイエンスの単位を取得していなくても理解できる。

「アクセス可能な Deep Tensor」は GUI を使用した可視化により説明責任とアクセシビリティを向上させる。この技術は富士通の独自技術である Deep Tensor の重要な拡張機能であり、「アクセス可能な」Deep Tensor は、統合型グラフィカルインターフェース (GUI) の採用により、エンジニアがモデルやシステムのフィルタリング、可視化、構成を相互連結スクリプトの編集なしに実現することを可能にする。ダッシュボード機能は徐々に向上しているが、ユーザーモデル定義型インターフェースは AI の説明可能性にとって非常に重要な要件だ。実際、展示ではこの他に、営業職や保険のエキスパートに特化した事例や Deep Tensor と、学術論文やゲノム医療データで構成されたナレッジグラフを組み合わせ、医療診断を平易な言葉で行う将来有望な技術の紹介もあった。

主な技術的利点は以下のとおりである

- さまざまなユーザーモデルをターゲットとした説明機能 (データアナリストだけでなく)
- [大量の] トレーニングデータを対象とするため、AI の民主化に役立つ (データを大量に保持する組織だけを対象とするのではなく)
- 革新的手法がエコシステムを発展させる。(富士通の社内用途としてだけでなく)
- 開発段階から信頼性を最優先した技術開発 (生産段階における事後措置でなく)
- 編集・構成ツールの向上を図り、より強固な AI 管理・制御を実現

上記の事例以外にも富士通はマサチューセッツ工科大学 Center for Brains, Minds, and Machines、INRIA、オックスフォード大学、スタンフォード大学のような世界有数の研究機関と提携し研究を進めている。

AI の採用：信頼できる AI 実現に向けた主な活動

AI が出現して数十年経つが、最近の AI ブームとその商業および公共部門での目覚しい採用は世界規模で信頼性が低下している中で高まっている。2018 Edelman Trust Barometer によると、政府やメディア各機関に対する信頼は過去数十年に比べ低下している。さらに、2018 年は、サイバーセキュリティ違反、選挙への干渉、公衆衛生・安全性に対する脅威などによってテクノロジー・カンパニーに対する信頼も低下している。社会として、我々は転換期を迎えている。今こそ、倫理設計が AI の継続的商業化を推進するか否かを決定づける時期である。

- **責任ある AI はビジネスにとって最重要課題である。**企業は単に存在感をアピールするためだけに説明可能性を必要としているわけではない。モデルに対する解釈、探求、監査、改良は直接、財政に影響する。たとえばコンプライアンス (法令遵守) には、多大なコストがかかる。なぜなら、過ちにより損害を与えた場合、多大な金額のかかる訴訟に発展し、ブランドへの厳しい批判、従業員の自然減少、顧客離れにつながり、これらを修復して元に戻すには何年もかかる。さらに、説明可能な AI は人間の説明可能性も高め、推論や直感を試し、新しいビジネスチャンスを発掘する重要なスキルとなる。

- **共創や協業はもはやオプションではない。**材料ベース経済から、情報が利益を生み出す資産となる情報ベース経済に移行するにあたり、企業は成長に向けたエコシステムに移行しなければならない。協業および「倫理、技術、産業面」での基準に対するコンセンサス、たとえばデータコントロールシステムのデザインや成功事例、技術手法に関するコンセンサスが極めて重要である。閉鎖的で、所有権のみに固執するモデルはデジタル時代においては時代遅れである。

- **ユーザーモデルが AI の説明可能性確立手段を左右する。**結果に至った理由を理解しやすく説明する場合、誰が解釈するかによって、あるいは彼らと AI モデル自体との関係によって、説明の中身は著しく異なる。データサイエンティストが求める説明可能性は、ソフトウェアエンジニア、エグゼクティブ、サービス部門、消費者、弁護士、規制当局が求める説明可能性とは異なる。また、説明可能性に対するニーズは、トレーニング中、編集中、あるいは微調整中、学習中かなど、AI ライフサイクルのどの段階を評価するかによっても異なる。

量、種類、速度ともに増え続けるデータの意味を理解する活力とニーズは人工知能ビジネスを推進する上での主要目標であり重要な原動力である。賢明な意思決定を行なうための補助機能から自動化機能へと AI が進化する際、データはまさに燃料である。しかし、透明性と信頼性はそれを推進するエンジンである。

本レポートについて： 本レポートは Kaleido Insights が富士通のご協力を得て作成しています。Kaleido Insights は Fujitsu Laboratories Advanced Technology Symposium 2018 に出席し、イベント分析を行ない、AI、デジタル倫理、その他関連分野におけるわが社の専門知識に基づき、幅広い市場観点から第三者としての論評をここに記載しています。本レポート作成にあたり、富士通役員、富士通研究所チーム、イベント参加者の方々にご協力いただき、皆さんのご意見を取り入れています。Kaleido Insights が技術提供者および採用者と行っているさまざまな活動の詳細あるいはその他のレポートをご覧になりたい方は次のホームページにアクセスするか、メールでお問い合わせください。https://www.kaleidoinsights.com/　メールによるお問い合わせ先：julie@kaleidoinsights.com

著者 Jessica Groopman について： Jessica Groopman は業界アナリストで Kaleido Insights の創立パートナーです。彼女は Kaleido における自動化技術のリーダーであり、AI、ブロックチェーン、IoT およびデジタル倫理の専門家として、これら分野の融合を推進しています。Jessica は新規技術の業界イベントに講演者として数多く参加し、ブログやメディアにも多数の記事を寄稿しています。また、Tractica、Harbor Research、および Altimeter の主席アナ

リストであり、International IoT Council、IEEE の Internet of Things Group や DigiGuru Network の contributing member（賛助会員）でもあります。また Onalytica の IoT における最も影響力のある 100 人にも選出されています。