

## AI原則って何だろう？（後編）-人間中心のAI原則で広がる「AI×倫理」

FUJITSU JOURNAL / 2019年11月14日



今回は[前回](#)に続き、世界中の政府機関が作成と推進を進めているAI原則を取り上げたいと思います。

前回はOECDやG20、日本政府などがそれぞれの立場でAI原則を制定していることをご紹介しました。それぞれのAI原則を見ていて興味深いのは、どのAI原則も「AIシステムは人間中心でなければならない」という人間尊重の考え方を強く主張していることです。このことは、AIシステムに人間同様の倫理を求めることにつながります。

AIシステムが人間の代わりに判断する用途で使われるようになると、法律を犯しているわけではないけれど、倫理的に望ましくない判断をしてしまうことで、人々が不快になったり、精神的な被害を受けてしまったりするかもしれません。そうなってしまえば、人間中心のシステムとはいえません。例えば、AIを活用した顧客対応ツールとして導入が進んでいるチャットボットの利活

用では、ユーザーからの質問に対する回答が「倫理的に適切な回答といえるのか」と論議の対象となるケースがしばしば発生しています。倫理的に誤っているという指摘があるなら、人間中心の観点からすれば改善すべきシステムとなります。つまり、人間中心の社会においてAIシステムが社会の信頼を勝ち取るには、人間同様の倫理を備えなければならないのです。

こうした考え方を受けて、AI原則と同じようなAI倫理原則を制定する動きもあります。例えば、欧州連合（EU）の欧州委員会は2019年4月、「信頼されるAI（Trustworthy AI）のための倫理ガイドライン」（Ethics guidelines for trustworthy AI）として、7つの守るべき指針を発表しました。以下に示します。

- 人間の尊重
- 堅牢性と安全性
- プライバシーとデータガバナンス
- 透明性
- 多様性、非差別、公平性
- 社会と環境の幸福
- 説明責任

AIシステムに倫理が求められるのは、AIシステムの信頼を高めるためでもあります。信頼のかけないシステムは社会に受け入れられません。AIシステムが信頼されるためにも、倫理的であることが期待されるのです。実際、欧州委員会のガイドラインはTrustworthy AIに求められる三つの要件として、法令遵守（Lawful）、倫理的（Ethical）、頑健または堅牢（Robust）を挙げています。

ただし、倫理は法律と異なり、明文化されていないことが多く、地域や宗教、時代による相違も法制度より大きいといえます。倫理は人々が幸せに社会生活を営むためのルールですから、コミュニティごとに形作られた文化の影響を強く受けます。国や地域が異なれば、そこで求められる倫理も違ってきます。こうしたことから、AIシステムが信頼されるものになるには、データの信頼性を確保することに加え、AIシステムが守るべき倫理ガイドラインの詳細はどうあるべきかなどについて、それぞれの地域、コミュニティごとに監視を続け、あるべき倫理とは何かを議論し、共有する努力が欠かせないでしょう。

## 製品開発の前段階でAI×倫理を実践

実際の社会システムの中にAIを実装する場合は、対象となるコミュニティが求めるAI倫理の姿を追求する必要があります。こうした取り組みの例としては、Fujitsu Laboratories of EuropeとFujitsu EMEIAが2015年6月にスペインのサンカルロス臨床病院生物医学研究財団（FIBHCSC）と交わした戦略的研究協力があります。このプロジェクトはメンタルヘルス障害の患者を対象に、AI

ベースのシステムを用いて患者の行動の手がかりとなる有用な情報を抽出して視覚化することを目的としたものでした。実証実験にはサンカルロス医療病院の精神科の専門医が参加して6か月以上にわたって進められ、匿名化された3万6000名以上の患者の情報を収集し、患者のこれまでの診断結果や合併症、自殺の潜在的リスク、薬物依存、アルコール依存などの未解決問題の関連性についてAI解析したところ、臨床における既存の問題点を特定できるなど、非常に高い精度のリスクアセスメントが実現できました。その結果、自殺・アルコール依存・薬物依存のリスクを85%以上の精度で算出できたそうです。専門医が納得できるAI倫理を実装した医療情報システムだからこそ実現できた成果といえるでしょう。

動画：[サンカルロス医療研究所様：AIで精神病患者の命を救う！～医師の診断時間半減に成功～](#)

サンカルロス医療研究所様：AIで精神病患者の命を救う！～医師の診断時間半減に成功～

AI倫理をテーマとした研究は、欧米のさまざまな大学や団体で進められていますが、AIベンダー自らが主催するケースもあります。例えば富士通は、2018年10月には米サンタクララで、2019年7月には英ロンドンでそれぞれ開催した自社イベントにおいて、識者によるAI×倫理のパネルディスカッションを実施しています。



Fujitsu Innovation Gathering 2019（2019年7月）におけるAI×倫理をテーマにしたパネルディスカッション「Towards Trustworthy AI: transparency and auditability」の様子

皆さんも機会があれば、AI倫理の議論に参加することをお勧めします。AI倫理は、ITとAIの未来に関連したテーマですが、そこでの議論される内容は「人間社会としてあるべき姿」であり、その

実現にAIをどう活用するのか、AIの活用がマイナスになることはないのか、といった基本的かつ根源的なものです。普段はあまり考えることのない普遍的なテーマですが、今後の社会のデジタル化を考える上で重要なヒントを得られる機会となることでしょう。

こうした議論を聞いていて感じるのは、AIシステムの活用は、これまで難しかった人間社会の課題解決に大きな貢献ができるのではないかという期待です。AIシステムの社会実装に関しては、人間社会へのマイナスの影響をゼロにするという視点だけでなく、今の社会に存在する差別などの課題を解決するにはAIをどう活用すればいいか、という議論も進められています。もしかしたら、社会に存在する差別がAIシステムによって是正される時代がくるかもしれません。

AIが守るべき倫理を突き詰めていくと、必ず人間社会における倫理に行き当たります。あるべきAIを考えることは、あるべき人間の姿を考えることにほかなりません。AIにおける倫理の在り方については、AI黎明期から多くの先人たちが議論を重ねてきていますが、その見識を踏まえた上で、社会の変化に合わせて議論を続けていくべきなのでしょう。

## 作り手も使い手もAIコミットメントが求められる時代に

AI技術の進化は目覚ましく、これまで未開拓だった分野での利用が期待される新しいAI技術も登場しています。例えばFujitsu Laboratories of Europeが2019年7月に発表した「Label Gear」は、製造、インフラ保守、ヘルスケア分野などの分野で監視員の目視検査業務を自動化するアプリケーション構築に効果を発揮するAI技術です。新しいGUIとAI支援技術を組み合わせることで、大量の教師データ（入力データを評価するためにあらかじめ組み込んだ訓練用のデータ）がない状態でも、欠陥や異常状態を効率よく自動検出できるという特徴を持ちます。専門家が必要だった目視による検査や監視業務の自動化は社会システムの効率化に大きな貢献が期待できます。大量の教師データを作ることは大きな負荷となっていました。その負荷を軽減できるとなれば、これまで以上に多くの業務分野でのAI適用が進むことでしょう。ただし、このLabel Gearが開拓するソリューションもまた、人間が担当していた業務分野におけるAI活用となることから、そのAIソリューションの実装に当たっては、人間同様の倫理を備える必要が出てくるでしょう。

このような状況で、今後加速しそうなトレンドをお伝えしておきます。それは、それぞれの企業や組織におけるAI活用の理念を社会に約束する必要が出てくるかもしれないことです。セキュリティ問題が深刻化して、組織や企業がセキュリティポリシーを作成したのは、セキュリティポリシーを社会に発信することが組織や企業の社会的な責任であるという考えが広がったからです。同様に、AI原則が広く知られるようになれば、AIを導入する際には自らの組織や企業におけるAIポリシーやAIガイドラインを公表し、その順守を社会に約束するなどして、人間中心のAIシステムを実践していることを証明する姿勢が求められることになるでしょう。

サンカルロス病院でのシステム開発やLabel Gearの開発など、AIの先端技術に取り組んできた富士通は、2019年3月にAI開発者の立場で「富士通グループAIコミットメント」を発表しています。また、2019年9月にはAI倫理に関する外部の有識者で構成する「富士通グループAI倫理外部委員会」を設置しました。倫理外部委員会設置の目的は、AI倫理に関する各種テーマについて議論し、そこで得られた客観的な意見や考え方を「富士通グループAIコミットメント」にフィードバックするなどして、社会のステークホルダーとの対話を重ねることにあります。皆様の組織が「私たちのAIコミットメント」を作成するときは、これらの取り組みを参考にしてみてください。

**1.AIによってお客様と社会に価値を提供します：** 富士通はグローバルに展開するグループ会社をあげて、先端技術を活用した共創を通じ、お客様、社会の豊かな未来創造に取り組んでいます。継続的に開発が進むAIがエンドユーザーや社会に与える影響に配慮しつつ、お客様に寄り添い、お客様との共創によって新しい価値を提供します。

**2.人を中心に考えたAIを目指します：** 富士通は、Human Centric理念のもと、あらゆる人の可能性を広げ、それぞれが望む形で幸福を追求したり社会に貢献したりすることを支援するツールとしてAIを位置づけます。プライバシーを守り、人が差別を受けたり人が危害を受けたりすることのないよう公平や安産を配慮して実装することにより、人が安心して利用できるAIを目指します。

**3.AIで持続可能な社会を目指します：** 富士通は、持続可能な開発目標（SDGs）に積極的に取り組んでいます。AIの研究・開発・提供を通じて、様々な社会問題や地球規模の課題に挑戦し続け、より良い社会の構築とお客様の長期的な成功に貢献することを目指します。

**4.人の意思決定を尊重し支援するAIを目指します：** 富士通は、AIが導き出した提案や結果について、人が吟味して意思決定することが重要であると考えます。AIによる自動化を推進するにあっても、システム全体として、提案や結果の根拠を的確に示す仕組みを提供することを目指します。

**5.企業の社会的責任としてAIの透明性と説明責任を重視します：** 富士通は、社会の基盤としてあらゆる人に関わるものになっている情報通信技術が担う社会的責任の重さを自覚して、とくにAIについては、望ましくない、または不足の結果をもたらさないよう品質向上に努めるとともに、必要に応じて原因を究明できる仕組みを実装して、安心・安全な社会システムを目指します。

## 取材協力

Rouz Adel

欧州富士通研究所（Fujitsu Laboratories of Europe Limited） CEO



## 著者情報

林哲史

日経BP総研 主席研究員

1985年東北大学工学部卒業、同年日経BPに入社。通信/情報処理関連の先端技術、標準化/製品化動向を取材・執筆。2002年「日経バイト」編集長、2005年「日経NETWORK」編集長、2007年「日経コミュニケーション」編集長。その後、「ITpro」、「Tech-On!」、「日経エレクトロニクス」、「日経ものづくり」、「日経Automotive」等の発行人を経て、2014年1月に海外事業本部長。2015年9月より現職。2016年8月より日本経済新聞電子版にて連載コラム「自動運転が作る未来」を執筆中。2016年12月「世界自動運転開発プロジェクト総覧」、2017年12月「世界自動運転/コネクテッドカー開発総覧」、2018年6月「Q&A形式でスッキリわかる 完全理解 自動運転」を発行。2011年よりCEATECアワード審査委員。