AI原則って何だろう？（前編）–「信頼」と「倫理」が問われる“明日のAI”

FUJITSU JOURNAL / 2019年11月14日

2019年5月、経済協力開発機構（OECD）はフランス・パリで閣僚理事会を開催し、OECD加盟36カ国とパートナー6カ国の合計42カ国が、AIに関する国際的な政策ガイドライン「AIに関するOECD原則」（OECD Principles on Artificial Intelligence）」に署名し、AIシステムを健全、安全、公正かつ信頼に足るように構築することで合意したと発表しました。

「AIに関するOECD原則」で合意したOECD閣僚理事会の様子（出所：OECD日本代表団ホームページ）

2019年6月には、茨城県つくば市で「G20茨城つくば貿易・デジタル経済大臣会合」が開催され、人間中心の考えに基づいてAIの利活用を促進する「Human-centered Artificial Intelligence （AI）」という考え方がG20閣僚声明として示されました。

AI原則の作成・推進を進めているのはOECDとG20だけではありません。国内では2019年3月に内閣府が組織した「人間中心のAI社会原則検討会議」が日本のAI原則となる「人間中心のAI社会原則」

を制定・発表しています。内閣は、経済発展と共に社会課題を解決する「Society5.0」の実現を目指していますが、それを実現するにはAIの活用が欠かせないと考え、AI原則を作成したのです。

このように、AI原則の作成・推進活動が世界中で始まっています。そして、実はこれらの活動は世界的な連携の下で進められています。例えば「人間中心のAI社会原則検討会議」は2018年4月に活動を始めましたが、当初から欧米各国やOECDなどの海外政府機関との連携を視野に置き、日本が主体的にAI原則を作成し、その考え方を世界に広げることを目的に活動してきました。実際、G20によるAI原則の制定・合意は、日本がAI原則を作成する過程で各国関係者と議論を積み重ねてきたことが大きく貢献しています。

OECDや日本が制定したAI原則が対象としているシステムは、必ずしもAIにのみ限定されているわけではありません。AIをはじめとするさまざまな先端技術を組み込んだ各種ITシステムが社会の基本的な構成要素となる未来社会を想定しているので、ITシステム全般が対象となります。

今回は、なぜ世界の政府がAI原則を求めるのか、AI原則で決めていることは何か、AI原則の制定で何が変わるのかを見ていきましょう。

人間の業務をAIが肩代わりする時代へ

世界中の政府がAI原則の制定を急いでいる理由は二つあります。

第一は、AIを用いたシステムの高度化が進んだことで、これまで人間に委ねられていた業務のAIシステムによる置き換えが始まりつつあることです。特に大量のデータで学習させるディープラーニングベースのAIシステムは、これまでシステム化が難しかった業務分野であっても、大量のデータを学習させることで人間並みの品質で業務遂行できるシステムを効率よく作れるという特徴があります。

この特徴をいかせば、これまで多くの人々を煩わせていたオフィスや工場における労働負荷を減らしたり、専門家不足に悩まされていた高度な判断処理をシステムに肩代わりさせたりできる可能性が出てきます。社会システムのさまざまな分野にAIシステムを活用すれば、人手不足対策や働き方改革対策を充実させて人々の生活を豊かにすると同時に、産業や企業活動の変革を進められます。

デジタル技術を活用して産業や企業活動に変革をもたらすことはDX（デジタルトランスフォーメーション）と呼ばれており、業種を問わず、すべての企業・組織が真っ先に取り組むべき経営課題となっています。また、日本をはじめとする各国政府は、国連が推進するSDGs（Sustainable Development Goals）の実現を目指していますが、これもAIの社会実装なしには実現できないでしょう。

DXの実践例はモビリティ産業に見ることができます。米ウーバーテクノロジーズや米リフト、シンガポールのグラブはAIをフル活用することで乗車ニーズや配達ニーズを持つ利用者と乗客や食品を運ぶドライバーをリアルタイムで結びつけ、オンデマンド配車サービスや食品配達サービスを急成長させています。また、これまでシステム化が難しかった自動車の運転業務に関しても、米ウェイモや米GMクルーズなどの自動運転開発企業が積極的にシステム化を進めています。これらの自動運転開発企業はAIを駆使し、大量の走行データで自動運転ソフトを学習させて運転精度を高め、ドライバーレスの自動運転車の走行テストを続けてきました。こうした活動が実を結び、2018年から全米各地でエリア限定のロボタクシーサービスや無人配送サービスの試験提供が始まっています。

シリコンバレーの公道を走行するウェイモの自動運転車とサンフランシスコ市内の公道を走行するGMクルーズの自動運転車（撮影：日経BP総研）

アリゾナ州の公道を走行する米ニューロの無人配送車「R1」とその利用風景（出所：ニューロ）

システムの信頼を揺るがすAI活用の現実

第二の理由は、AIを用いたシステムが社会に広がりつつある中、その利用が人々の生活にマイナスの影響を持ち込む危険が顕在化していることです。

例えば自動運転車は、自動運転ソフトが周囲の様子をセンサーなどで検知して安全であることを判断して運転操作を実行します。仮に事故が起こったら、その責任はどこにあるのでしょうか。AIシステムが人間と同じ業務をするということは、人間と同じ責任がAIシステムに求められることになります。現時点では、AIシステムに人間と同じ責任を求める法制度が整備されているわけではありませんが、将来的にはAIシステムが引き起こすトラブルの責任を誰が負うのかを明確にする必要があるでしょう。

AI利用がもたらすマイナスの影響というのは、人間同様の責任をAIに求めることが難しいということだけではありません。より大きな問題は、システムに対する信頼が揺らいでいることです。

これまでのシステムは、その動作のすべてを人間が設計してプログラムとして実装していたため、システムの振る舞いを完全にコントロールできました。開発過程でシステムがどのように動作するかを十分に検証できるので、想定外の動作が生じることは少なく、起こったとしてもすぐにプログラムを改良すれば不具合を取り除くことができました。

例えばクルマの運転制御はすでにシステム化されていて、クルマの内部にはドライバーの運転操作を適切にクルマの細部に届けるための情報処理を実行するCPUが機能別に数十～数百個も組み込まれています。ドライバーはハンドルを操作したりブレーキペダルを踏んだりしてクルマを走らせますが、それらの運転操作はすべて電子的な命令に変換されてクルマ各部の制御機構に届けられ、全体の車両制御が実行されています。私たちがクルマを運転しているとき、クルマの制御にシステムが介在していることを意識する場面はありません。「ブレーキを踏んでも、クルマが止まらなかったらどうしよう」と考えないのは、システムの搭載に気付かないくらい制御システムが安定して動作しているからです。「ブレーキは、踏んだ分だけ着実に減速する」と当たり前に考えているのは、クルマの制御システムを完全に信頼しているからといえるでしょう。

一方、今のAIシステムには、これほどの信頼感はありません。特に大量のデータを用いて学習させたディープラーニングベースのAIシステムは、想定外の動作をしてしまったり、どのような動作をするかを十分に検証できなかったり、なぜその動作をしたのかがわからなかったりすることがあります。例えば自動運転車が接触事故を起こしそうなぐらい隣のクルマに近づいたとき、きっと乗客は不安を感じることでしょう。自動運転ソフトが事故の危険性を予測して、それを回避するための運転操作だったとしても、「人間ならこんな運転はしないはず。やっぱりAIは信用できないな」と思うことでしょう。こうしたことからAI開発の世界では「なぜそのように判断したのか」を説明できるようにする「説明可能なAI」の研究が活発に進められています。説明可能にすることで、AIの信頼性を高めようとしているわけです。

AIシステムに対する信頼は、学習データに誤りがあったり、偏りがあったりすることでも揺らぎます。学習データに偏りや間違いがあると、誤った価値を学習してしまうためです。例えば、2018年に米ロイターが報道した米アマゾンドットコムが取り組んでいた「人材募集に申し込んできた人の履歴書を細かくチェックして、望ましい人を選び出すAIツールの開発」のケースでは、学習データとして採用した過去の履歴書に偏りがあったため（男性の志願者が女性より圧倒的に多かった）、AIが「男性の方が女性より好ましい」という価値を学習した可能性があることがわかりました。アマゾンドットコムのエンジニアはこうした偏りを排除するための調整を試みたものの、「AIは公平に候補者を選んでいる」という確信を持つことができず、このツールは実際の採用に使われなかったそうです。

このようにAIは、人間社会を豊かにする魅力あふれる技術である一方、間違った使い方をすると人々の生活にマイナスの影響をもたらすシステムを構築してしまいかねない技術でもあります。これは技術やシステムの成熟度の問題というよりも、用途が広く、ポテンシャルが大きいからこそ、使い方が厳しく問われるとみるべきでしょう。切れるナイフは、人間の能力を超えた機能を提供しますが、だからこそ使い方・扱い方を誤ると人間社会にマイナスの影響を当たえます。AIは人間を超える能力をこれまで人間が担当してきた業務分野で発揮できる可能性があるため、注意深く扱わなければなりません。

AIで人類を幸福にするには、開発者も利用者もAIを注意深く取り扱い、信頼性を高め、常に正しい使い方をしているのかを検証していく必要があります。そして、このAI利活用の知識を早急に社会全体で共有しなければならないという危機感が、各国政府にAI原則の作成を急がせる原動力となっているのです。

AI原則は「AI導入に当たっての心構え」

各政府組織が定めているAI原則は法律のようなものではありません。AIを用いたシステムを社会の中に組み込む際のあるべき姿や目指すべき目的を掲げたり、AIシステムを作ったり使ったりする人が持つべき考え方を説くといった内容で構成されています。AI社会に向けて人々が共有しておくべきガイドライン、あるいは「AIシステム導入に当たっての心構え」と考えるのがいいでしょう。

AI原則の具体例として、OECD原則の概要を以下に紹介します。

- AIは、包摂的成長と持続可能な発展、暮らし良さを促進することで、人々と地球環境に利益をもたらすものでなければならない。
- AIシステムは、法の支配、人権、民主主義の価値、多様性を尊重するように設計され、また公平公正な社会を確保するために適切な対策が取れる―例えば必要に応じて人的介入ができる―ようにすべきである。
- AIシステムについて、人々がどのようなときに結果の正当性を批判できるのかを理解できるように、透明性を確保し責任ある情報開示を行うべきである。
- AIシステムはその存続期間中は健全で安定した安全な方法で機能させるべきで、起こりうるリスクを常に評価、管理すべきである。
- AIシステムの開発、普及、運用に携わる組織及び個人は、上記の原則に則ってその正常化に責任を負うべきである。

日本のAI原則といえる「人間中心のAI社会原則」は12ページからなる文書で、7つのAI社会原則とAI開発利用原則が記されています。AI社会原則のカテゴリは以下です。

（1）人間中心の原則
（2）教育・リテラシーの原則
（3）プライバシー確保の原則
（4）セキュリティ確保の原則
（5）公正競争確保の原則
（6）公平性、説明責任及び透明性の原則
（7）イノベーションの原則

なお、AIシステムの信頼性確保については、2019年6月のG20閣僚声明の中で「DFFT」（Data Free Flow with Trust）という新しい概念が示されましたが、これは安倍晋三首相が2019年1月の「世界経済フォーラム年次総会」で発表したものです。

今回はAI原則の最新事情をご紹介しました。AIは人間社会をよりよくするためのものであるという認識の上で、AIを正しく使うための心構えがAI原則といえるでしょう。それだけAIに期待がかかるのは、AIがこれまでの技術と異なり、人間の代わりに高度な判断を下せる可能性があるからと言えます。ただし、そのためには人間なら当然備えていることが期待される倫理感もまた、AIに求められることになります。次回はこのAI×倫理について見ていくことにしましょう。

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。

FUJITSU JOURNAL / 2019年11月14日