

(OECDが人工知能(AI)の開発や運用に関する基本指針を採択)

経済協力開発機構(OECD)は2019年5月22日、OECD加盟36か国及びパートナー6諸国(アルゼンチン、ブラジル、コロンビア、コスタリカ、ペルー、ルーマニア)が人工知能(AI)の開発や運用に関する基本指針を採択したと発表した。人にとって本当に役立つAIを前提に考える「人間中心」のほか、差別をなくす「公平性」や説明責任など、AIの開発や利用にあたって重視すべき原則を打ち出している。AIは様々な分野で人の生活を便利にする技術として期待される一方、普及にはリスクもある。今回採択された基本指針には、AIの負の側面を最小限にし、信頼できるAIをつくっていくため、AIや開発する人の暴走を防ぐ防波堤の役割も期待されている。

AIの「深層学習(ディープラーニング)」と呼ばれる技術については、大量に学習したデータから導き出された結論がなぜそうなったのかの過程が外部からわかりにくいという「ブラックボックス」化が懸念されており、AIが下した判断の理由を説明できる「透明性」の確保が求められている。

(参考) AIに関するOECD原則概要

1. 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。
2. 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。
3. AIシステムについて、人々がどのようなときにそれと関わり結果の正当性を批判できるのかを理解できるようにするために、透明性を確保し責任ある情報開示を行うべきである。
4. 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。
5. 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。

(注) OECD公表資料(2019.5.22)による。

(汎用人工知能(AGI)と特定型人工知能(NAI))

AIには、汎用人工知能AGI(Artificial General Intelligence)と特定型人工知能(Narrow Artificial Intelligence)の二つがある。前者は、人間のように考え、計画し、反応できるだけでなく、人間を超えた超能力を備えうると考えられている。一方、後者は自動走行、音声認識、画像を用いた医療診断など個別の目的に資するであり、テクノロジーを規制する法律、制度等が存在するので、それらを積極的に活用し、安全装置を強化していくことが可能であるが、前者にはそれが働かない可能性がありAIのパフォーマンスと透明性確保をどう調和させるかという問題に直面する。

（「FOREIN AFFAIRS REPORT」の「人工知能への備えはできているか」に示された有識者所見）

これについて、2019年8月号の「FOREIN AFFAIRS REPORT」における「人工知能への備えはできているか」というケネス・クキエル、エコノミスト誌シニアエディターのレポートの中で、①アンドリュー・エン、スタンフォード大学の研究者の「現時点での AGI は依然初歩的なもので、やっと人の顔を認識し始め、スピーチを解釈でき始めたところであり、AGI のことを心配するのは、火星が人口過剰になることを心配するようなものだ」として、「AI の機能を抑えることを今から考えるべきではなく、まずは本来の機能の開花に努力すべきである」とする AGI 開発優先論ともいうべきものと、②アメリカの心理学者スティーブンビンカーの、「制御不能になった AGI が人類にダメージを与えるという考えは科学的思考だけでなく、憶測にも依存しているために、憶測がらみの厄介なシナリオを回避するために大きな資源を注ぎ込むのは間違っている」との考えを引用しながら、「(AI に対し) テクノロジーに関する規範、法律、制度の積極的運用による安全装置を導入しつつ、インターネットを作り上げたときに強固なセキュリティ機能を埋め込む必要性を見過ごした誤りに学び、AI の設計者は最初からセキュリティ機能を AI に埋め込んでおくことが重要である」とのケネス・クキエル自身の開発と規制のバランスに配慮した条件付き開発論が記述されている。

（人工知能に対する寄稿者の全体所見）

最後に、本レポートの寄稿者ケネス・クキエル氏の所見が、寄稿冒頭に以下のように要約がなされているので紹介しておこう。

AI は良くもあり、悪くもある。賢いが、鈍い部分もある。文明の救世主であるとともに、世界の破壊者でもある。実際、どのようにして決断を導き出しているか分からないし、それを人間が解明することもできない。特に、汎用人工知能（AGI）については、「独自に進化し、人間が管理できなくなるのではないかと懸念され、特定型 AI についても「デザイナー（である人間）がその意図を完全に伝えられず、壊滅的な結果が引き起こされるのではないかと心配されている。一方で、AGI のことを心配するのは、火星が人口過剰になることを心配するようなもので、先ず、(AGI に関して) 想定されていることが実現する必要があると主張する専門家もいる。イノベーションの進化ペースをどの程度「警戒」し、(機械のメカニズムに関する) 説明の「正確さ」をどこまで求め、(個人データを利用することによる) パフォーマンス強化と「プライバシー」のバランスをどこに求めるか。社会がこれらのバランスをどうみなすかで、人間が AI とどのような関係を築いていくかが左右される。

（荒井 俊行）