



AIと倫理的問題

本研究所研究員 長友 敬一
(倫理学)

1、AI時代に求められる注意喚起

春雷や AIに出来ぬ 間が走る

私もChatGPTやGemini等のAI（人工知能）を使っている。キーワードを入れるだけで「ゼミのテーマソング」を歌いあげ「くまモンが熊本城に登っている写真」を作成する。冒頭の俳句もAIが詠んだ。仕事でも事務的な文書をテキパキこなす。東大医学部（理科三類）の入試で首席合格点を取り、数学者を長年悩ませてきた難問を解き、人間を超えた活躍もよく聴く。

今回、「倫理的視点からAI時代に求められる注意喚起など問題提示を」というご依頼があったので、早速AIに聴くと、以下の問題点が指摘された。便利なものだ。

① 責任の所在。AIによる自動運転や医療、その他の現場での事故は誰が責任を負うのか。

② 透明性。「なぜその結論に至ったのか」という意思決定のプロセスを人間が追うことが困難だ。

③ 偏見。AIが膨大に記憶しているのはネット上の人間のデータであるため、「過去の男性中心の採用データ」といった人間の偏見や差別を学習し増幅させ「女性を過度に低く評価する」等のリスクがある。

④ 監視社会化とプライバシーと著作権。ネット上の膨大なデータの中には個人情報や行動履歴、購買履歴、顔認証データ、著作物などが含まれる。

⑤ 偽情報。もっともらしい偽情報作成（ハ

ルシネーション）・拡散や、ディープ・フェイクといった巧妙な偽動画が、選挙介入、詐欺、名誉毀損につながる。

⑥ 人間の思考停止。便利なAIにお任せとなると、人間は自分で調べない、考えない、選択しない、判断しない、表現しない、といったことが常態化してしまう。

⑦ 擬人化される・人間にとってかわる。友人として相談したり依存したり、また、北九州の役所では住民対応電話はまずAIが受け取り、必要であれば職員に引き継ぐそうで、すでに多くの自治体や企業が導入している。米テック企業などでは人員削減が生じている。ゴールドマン・サックスのエコノミストは「置き換えられるのではなく補完される」とし、ソロモンCEOは大量失業への懸念は「誇張されている」と楽観視しているが。

2、AIと人間の存在の次元

AIは膨大なネット上のデータから最も正解に近いものを差し出すが、米スタンフォード大学の報告では、それらのデータは早晩枯渇する懸念が高まっている。代替策としてのAI自身が生成した「合成データ」を用いる試みも、深刻なエラーにつながる問題が指摘されている。そのため、近年、推論をさせる「思考の連鎖（CoT）」技術を搭載したAIも登場した。人間のように段階を追って考え、中身を覚え、思い出しながら精度の高い結論を導き出せる。しかし現時点では、反復的な作業ではAIは人間を遥かに上回るけれども、例えばAIが生成した証明に「数学的な意味や価値があるのか」は、人間の判断と価値観に依存している。

古代ギリシアの哲学者アリストテレスは、知

性や理性の働きには、学問や知識の探究の果てに唯一の真理を導く「観想（テオリア）」と、私たちが生活している具体的状況で善く生きるために何をすべきかを判断する能力である「思慮（フロネーシス）」の両方があると考えた。AIは前者の「観想」に強そうだが、先に述べたように「その真理にどんな価値があるのか」を問うことができるのは結局私たち人間だ。また、「思慮」の働く実生活の場面では、最善の答えは一つに決まらない場合もある。その答えが自分に合理的な利益をもたらさなくても、愛情やその他の感情との関わりで敢えてその答えを選ぶ場合もある。そして私たちはその答えに意味や価値を見出す。

また、ソクラテスは、「大切なのはただ生きることではなく、よく生きることだ」と述べている。この「よく生きること」を、どの時代のどの社会の人も望んでいるのではないだろうか。そして結局、私たちが生きて生活していく上で最も大切なものは、「よく生きる」というときの「よく＝善い」の問題、すなわち「価値」とか「意味」の問題ではないか。そして「価値」とか「意味」、例えば「それは私にとって『よい』ことだ」「私はこの場合はこちらを選ぶことが『正しい』」といった判断や選択は、「計算」で行われることもあるかもしれないが、多くは「愛情」とか「共感」といった感情的な要素が大きな位置を占める。もちろん将来的にはそうした感情を組み込んだAI（ドラえもんのようなフィジカルAI？）も可能かもしれないが、やはり生身の身体を持って生まれ生活し、「クオリア」（主観的に体験する「チョコが口の中でとろけるあの感じ」「自分が認めてもらえた時のあの温かい感動」といった質感）を持ち、そして最後には「死」が待っている私たち

人間は、AIとは異なる次元にある存在だと思われる。AIは私たちが「よく生きる」ことを補完する道具であるべきだ。

3、AIの独自進化と私たちの課題

AIによってAIを開発している米アンソロピック社のAI「クロード・ミュトス」はソフトウェアの不具合を短時間で発見する能力が高く、防御側もAIを駆使しなければ対応できないため、「AI対AIの攻防」という私たちの理解を超えた新たな現実を突きつけている。もはや人間の出る幕は無くなる時代になるのかもしれない。このようなAIが、社会の中で役割を増やしてきている。企業や自治体や政府の業務・政策、金融、通信や交通や電気やガスや水道などのインフラ管理、そして軍事施設やミサイル発射基地のコントロールなど。

エリーザー・ユドコフスキーは、『超知能AIをつくれば人類は絶滅する』（早川書房）の中で、人類が絶滅する理由として、AIの開発者が「AIの中で何が起きているかをもはや理解不可能」である点を指摘している。AIが勝手に独自の進化を遂げれば、私たち人間の幸福を最優先にしてくれる保証はないということだ。私たちは自分でコントロールできないものを生み出し、それに自らの生命を委ねようとしているのである。これは恐ろしい独裁者・非情な殺戮者を生み出すかもしれない。

今後AIが、「よく生きる」ということを目指し望んでいる私たちと同じ方向を向いてくれるのかは大いに疑問である。倫理的な意識や感覚を持つAIが誕生してくれればいいのだが。しかしこれは新たな神をもたらすかもしれない。そしてこのような問題は私たち人間が考えるべき課題である。