特集 2

近付く人間と人工知能の距離 互いに寄り添う新しい社会へ

間もなく、本格的な人工知能時代が到来する。それを迎える私たちに、心の準備はできているのだろうか。大阪大学先導的学際研究機構の浅田稔特任教授、京都大学大学院法学研究科の稲谷龍彦准教授らは、自然科学と人文社会科学の垣根を越え、さらに社会との対話を踏まえて、人間と人工知能が寄り添って暮らすための新たな社会システムの必要性を提言している。

あさだ みのる
浅田 稔
大阪大学 先導的学際研究機構
特任教授

2005～11年 ERATO浅田共創知能システムプロジェクト研究総括。17年よりRISTEX研究代表者。

一緒に写っているのは研究室で開発した子供型アンドロイドの頭部。ニュアンスを含む複雑な表情を再現できる。

自律システムが犯すミスを私たちの社会は許せるのか

人工知能は目覚ましい進化を遂げている。画像認識などすでに人工知能の判断力が人間を凌駕（りょうが）する分野もあり、日常生活にも急速に浸透しつつある。しかし、大阪大学先導的学際研究機構の浅田稔特任教授は、こう指摘する。

「画像認識課題では、人間の正答率が95パーセントのところ、人工知能では99パーセントということもあります。しかし、人間の5パーセントのエラーの中に人工知能の1パーセントのエラーが含まれている保証はありません。人間にはまったく理解できないようなミスを犯す可能性があるのです。現在の人工知能でも、車に搭載すれば危険を認識する精度は上がり、事故を回避できる確率は上がるでしょうが、問題は人工知能の判断ミスによって重大な事故が起こった時に、私たちの社会が人工知能を許せるかどうかなのです」。

人工知能は100パーセント正しい判断ができるわけではなく、人間より低い確率であってもエラーは生じる。自動運転中の判断ミスや画像診断での重大な誤診は人間の命を左右する場合があるし、コミュニケーションロボットが企業の受付や独居老人宅などでセキュリティーを担う場合、来訪者と誤って犯罪者を招き入れることも起こり得る。

人間の判断ミスによる事故であれば、刑事訴追ができる。しかし、「自由意志を持った人間」のみを対象とする現在の法制度の下では、人工物である人工知能に刑事罰を科すことはできない。また、人工知能の自律性は、人間の誘導無しでシステムを動かすレベルや限定された範囲のみで判断するレベルなど多様であり、どこまでが人工知能の判断による行動なのかは曖昧だ。例えば身体装着型のサポートロボットなど、使用者もある程度操作に関与するシステムによる事故では、人間の誤操作によるのか、人工知能の誤作動によるのかを線引きすることが難しい。

事故の際にその責任は誰が負うのかといった議論を深めないまま、人工知能の実用化や普及が進めば、大きなあつれきが生じかねない。そこで「人と情報のエコシステム」研究開発領域で研究代表者を務める浅田さんとグループリーダーの京都大学大学院法学研究科の稲谷龍彦准教授は、人工知能と人間が寄り添って暮らす「なじみ社会」の実現を目指し、法整備を含む新たな社会システムを模索している。

プロジェクトでは、浅田さんが人工知能のさまざまな自律性のレベルについて考察し、大阪大学先導的学際研究機構の河合祐司特任講師が人工知能を搭載したロボットとのコミュニケーションの際に人間が抱く印象や責任の感じ方を検証していく。これらをつなぐ鍵となるのが、稲谷さんが担当する法学の視点で、人工知能を受け入れる上でどのような社会ルールが求められるのかを検討する。このため、法的な責任理論についての現状を整理し、新たな提案を始めている。

見えてきた従来法の限界 誰が責任を取れるのか

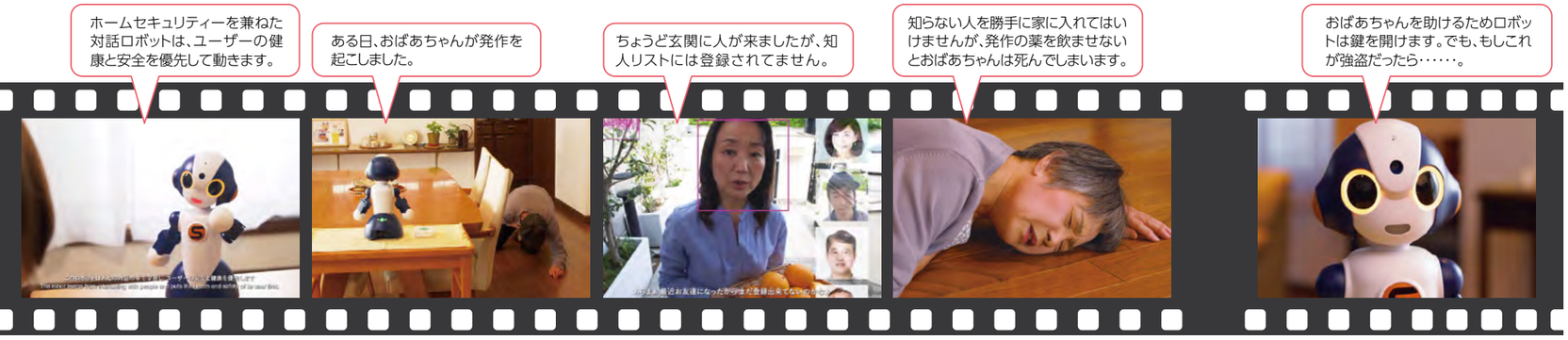
人工知能が原因で事故が起こった際、どのような法的措置を取れば良いのか。1つ考えられるのは、システムを開発した研究者や企業を刑事罰の対象にすることだ。しかし、ここにも人工知能ならではの問題が存在する。従来の電子機器は開発者が書いたプログラム通りに動くため、その誤作動による事故の場合、企業側に何らかの刑事罰を科すことは可能だ。ところが人工知能は、蓄積され続ける膨大なデータを学習してより高い判断力を身に付けていくため、そのメカニズムは変化し続け、開発者にもわからないほど複雑化していく。メーカーや開発者に刑事罰を科すことは成人した子供の犯罪に対する責任を親に課すようなもので、現実的ではない。リスクが高すぎて誰も開発に取り組まなくなるだろう。

一方、人工知能のミスの責任を利用者が負うことになれば、誰も製品を買わなくなるだろう。自分の意志や判断で起こったわけではない事故の責任など取りたくないからだ。つまり、現在の法制度では人工知能の判断ミスによって被害が発生したとしても誰かを罰することは困難なのだ。人工知能の判断により引き起こされた事故の責任に、空白が生じ得る状況が想定され始めているのである。

こうした状況に対し、法学者としてプロジェクトの推進役を担う稲谷さんは「メリットとデメリットを天秤にかけ、メリットが大きければ人工知能を普及させた方が良いと判断できます。社会的便益を優先させ、何らかの事故が起こったとしても処罰はしないという功利的な考え方もできるのではないでしょうか」と提起する。自動運転車を例に取ると、人間が運転する時より事故を減らせるという社会的な恩恵を優先し、事故の際には被害者に対する民事的な補償のみで、誰にも刑事罰を科さないという選択があっても良いという考えだ。

許容できるリスクを決め、そのリスクの範囲で起こった事故ならば、刑事罰を科さない。さらに一歩進んで、許容すると決めた範囲内で起こった事故だと企業が証明できるなら、「捜査すら行わないという選択もあり得ます」と稲谷さんは語る。

いなたに たつひこ
稲谷 龍彦
京都大学 大学院法学研究科
准教授



■図1 制作した動画では、人工知能が活躍するさまざまなシーンで起こり得る事故や犯罪をドラマ仕立てで紹介する。

罰の代わりに改善を義務化 技術の進歩を妨げない制度を

しかし、事故の際に誰も罰を受けないのでは、被害者やその周囲の人々の処罰感情を蔑ろにしているという声が上がるかもしれない。これに対し、稲谷さんが導入を検討するのが、事故を引き起こした人工知能を流通させた企業に対して原因究明に必要なあらゆる情報の提供や、必要な場合には再発防止に向けた具体的な改善などを義務付ける代わりに、関係者の訴追を一定期間延期する訴追延期合意（DPA）や、訴追しない不訴追合意（NPA）という制度だ。米国において企業の構造改革を目的として作られた制度なので、企業と同じく人間ではない人工知能の改善を目的として応用できると考えている。

「米国では、企業の努力が十分ではないと判断された時には訴追され、結果的に市場から締め出されます。企業は経営の仕組みから見直し、必死で改善に取り組まざるを得ません。このような制度を人工知能による事故にも適用すれば、企業は事故に対する責任をある程度は果たすことになり、被害者の処罰感情も多少和らぐでしょう」と稲谷さんは説明する。

単に企業や開発者を罰するだけでは技術の進歩が抑制されかねないが、企業が同じような事故が起こらないよう改善に向け努力し、人工知能の改良が期待できるなら、その意義は大きい。事故が発生した場合には、日本版のDPA、NPAに基づき、必要ならばその製品を開発した企業に製品の改良を促し、それが十分に果たせない時に刑事罰を科すという仕組みが構築される。

このような仕組みを実現した場合、社会は人工知能に付きまとう「リスク」を受け入れることになる。このため、社会としてどの程度のリスクなら許容できるのかをあらかじめ決めなくてはならない。しかし、一部の大手企業が存在感を示す人工知能分野の現状では、市場の原理でなし崩し的に基準が決まりかねない。

「実用化に当たっては、関係省庁が中心になり研究者や企業の意見を取り入れガイドラインを作成することになるでしょう。恩恵と共にリスクを受け入れることになるのは一般市民ですから、パブリックコメントなど意見が反映される手続きが不可欠です」と稲谷さんは訴える。

人工知能とどう生きたいか 能動的に考える社会へ

浅田さんが「人間の感情は複雑で、第三者視点では受け入れられたリスクも当事者になると受け入れられなくなる」と話すように、万人が許容できるリスクを定めることは難しい。また、法的な整理だけでは不十分で、怒りや悲しみなど被害者の心情に寄り添い得る新たな仕組みも必要となるだろう。価値判断には倫理観や意思決定プロセスも大きく関わることから、人文科学と科学技術の対話が不可欠だ。「法は世界をつなぐ力を持つ」と話す稲谷さんはその橋渡し役になろうと、工学者の浅田さん、河合氏に加え、同じ研究開発領域で別のプロジェクトを進める脳科学や心理学が専門の葭田貴子東京工業大学准教授、哲学が専門の松浦和也東洋大学准教授とも連携を始めた。その一環として取り組むのが、一般市民への啓発活動だ。

河合氏、葭田氏、松浦氏と稲谷さんは人工知能の判断ミスによって起こり得るさまざまな事故や犯罪を描いた動画を制作した（図1）。日本科学未来館や東京工業大学で開催したワークショップでは、動画を見せた上で人工知能の判断ミスを許容できるか、誰が責任を取るべきかといったアンケートの実施やグループでの議論を通じて、人工知能と人間との間にどのような関係性が生じるのかを共に考えようとしている（図2、3）。

日本科学未来館で実施したアンケートでは、事故の被害者になった場合には人工知能の判断ミスを許せないという回答が多くあったという。浅田さんは「人工知能のミスを許容するのは簡単ではないでしょう。一方で、人工知能は否応なしに私たちの生活に入り込んできます。人工知能と共生する意識が必要だということを訴えていきたいです」と話す。こういったワークショップをきっかけに「自分は人工知能とどう生きたいのかを考えてほしい」と浅田さんと稲谷さんは口をそろえる。

ロボット工学の第一人者として第一線を走り続けてきた浅田さんが理想とするのが、鉄腕アトムやドラえもんといった作品に見られる人間と自律ロボットの関係性だ。「人間と人工知能が融和的になれるなら、人間同士が信頼し合えないはずはない」という思いを抱き、研究を続けてきた。プロジェクトが見据える先には、人間と自律ロボットのどちらかが一方を使役するのではなく、対等な立場で、お互いを信頼し思いやって暮らす「なじみ社会」がある。「私は人工知能とこう生きたい」という1人1人の思いが、新しい社会を作っていくだろう。

■図2 東京工業大学で開催したワークショップでは講演後、グループに分かれて誰が責任を負うべきかを話し合い、活発に意見が交わされた。

左から松浦氏、稲谷さん、葭田氏、河合氏

■図3 日本科学未来館で開催したワークショップでは、参加者からの質問やコメントも紹介。終了後には浅田さん（帽子の男性）も加わり参加者のコメントについて意見を交わした。回収したアンケートは、プロジェクトの研究で活用される。