

法の支配

第197号

「ロボット・AI時代の社会と法」

——目 次——

＜巻頭言＞「AI三原則」	山 本 庸 幸… (2)
第1部《座談会》	
「ロボット・AI時代の司法の対処すべき課題」	宍 戸 常 寿／大 島 義 則／川 口 大 司… (6)
	西 内 康 人／橋 本 陽 子
第2部《論 文》	
＜AIの支配＞と＜法の支配＞～日本発AI原則とガイドライン～	平 野 晋… (41)
AI技術による裁判の変化と課題	町 村 泰 貴… (57)
AI時代の刑事司法	今 井 猛 嘉… (69)
AI時代の弁護士業務	村 本 道 夫… (80)
AI時代の競争法－現在までの議論の概観から－	市 川 芳 治… (96)
ロボット・AI社会における知的財産制度の現状と在り方	福 井 健 策／出 井 甫… (112)
AI時代の個人情報保護	板 倉 陽 一 郎… (127)
＜会員の声＞ 情報の共有と法の支配	増 田 義 憲… (138)
協会短信	(139)
編集後記	(140)

「AI三原則」

弁護士・元最高裁判所判事 山本庸幸

私は、自動車と先端技術にとっても興味があるので、東京モーターショーがあるときには、カメラを抱えて必ず行くことにしている。会場のあちこちでは、自動車の最新モデルの脇に、可愛いお嬢さんがにっこりと微笑んで立つという風景が見られるのだけれど、2017年から普通の自動車に混じって、妙な「コンセプト・モデル」が展示されるようになってきた。

それは、窓のついた長四角の箱で、四隅にタイヤはある。ところが箱の中には、ハンドルはおろか運転席もパネルもなく、ソファがあるだけだ。そこであたかも居間にいるようにして家族と話し込んでいると、いつの間にか目的地に連れて行ってくれるという車だ。運転する必要がないから、これは楽だと思う反面、完全自動運転なのでハンドルもブレーキも要らないというのはやり過ぎではないか、事故が起きそうになったらどうやって防ぐのかなどと思った記憶がある。

ところが、その頃から、カリフォルニアでアップルやアルファベット傘下のウェイモなどが自動運転車の大々的な公道実験に乗り出し、膨大なデータを集め始めた。この自動運転には、最近急速に発展したAI（人工知能）の深層学習（ディープ・ラーニング）が使われている。この技術は日々急速に進歩しつつあって、今や自動運転はレベル3の段階（特定の場所でシステムが全て操作するが、緊急の場合には運転者が操作する）にあるが、レベル5の段階（全ての場所でシステムが操作する）に達する日も、ほど近いと言われている。

刑法の責任主義の観点からすれば、事故が起こった時、レベル3の段階だと適切なハン

ドル操作をしなかった、緊急停止ボタンを押さなかったなどと運転者の責任を問うことになるから、今と同じ法体系でよい。ところが、レベル5の段階になったら、運転者の責任はもはや問えないことから、瑕疵のあるシステムを作ったとして自動運転開発者の責任を問うことになるのか…そうなるとそれは道路交通法の問題というよりは、民事上の不法行為責任の問題で対処するのか、あるいは製造物責任法に刑事責任を新設する問題になるのかなど、法体系が大きく変わる可能性を秘めている。もっとも、自動車の場合には自動車損害賠償責任法の運行供用者責任があるから、問題はあまり顕在化しないかもしれないが、これがない鉄道、航空機、船舶などは同様の制度を考える必要があるだろう。

弁護士や医師や会計士という専門職の仕事も、AIで抜本的に代わる可能性がある。弁護士なら貸金返還請求訴訟などの定型的訴訟から使いはじめて、交通事故、特許事件の先行文献調査などの補助業務なら結構使い出がある。そのうち良質なデータの蓄積があれば、あらゆる訴状や答弁書の原案を作成できるかもしれない。医師の仕事も同様で、エックス画像を見てガンがあるかどうかを判定する分野ではAIが専門医を上回る好成績を上げることも珍しくなくなったから、その種の業務からはじめて、やがては患者とやり取りをする診察までこなすAIが出てくるのは必然だろう。ただ、AIが専門職の補助的仕事をしている分には現在の法制度でよいが、近い将来それが一人前の仕事をするようになると、そもそも専門職制度はこのまま維持できるのかという問題に直面するのは避けられない気がする。

かくして、AIの深層学習技術は、自動運転にとどまらず、有用な技術として社会のあらゆる分野で使われ始めている。先ほどの医療分野での病気の判定、街角を歩く人混みの中からの指名手配犯の発見、プロを打ち負かす囲碁や将棋、定型的な新聞や雑誌記事の執筆、俳句詠み等々である。まさに、万能の力を発揮している。それは良いことだと思う。

ところが問題は、深層学習といっても、そもそもどういうアルゴリズムなのか、特にいかなる因果関係があってその結果が出ているのか、よく分からないのが現在のAIの特徴である。そうすると、出た結果を不審に思って検証しようとしても、それができないことが多い。例えば、犯罪発生地点の予測ならまだしも、特定の個人を犯罪者予備軍などと推測するのは、何の根拠でそう判定するのか、人権侵害も甚だしいということになる。また、

AIに放り込んだデータが間違っていたり、質の悪いものだったりすると、出てくる結果もまた見当違いのものになるのは、言うまでもないことだろう。

世の中は、AI元年が到来したとって大歓迎のムードであるが、私に言わせれば、コンピュータとソフトウェアである以上、検証不可能な結果をそのまま受け入れるのはどうかと思うし、元となるデータ自体がおかしなものであれば、一体どうするのだろうかという疑念がぬぐいきれない。しかし、考えてみれば、どんな新技術も導入直後には色々と問題が生じるものである。そうした問題も、そのうちに何とか克服される。そしていったんその便利さに慣れてしまえば、これに異を唱えるのはあたかも風車に立ち向かうドン・キホーテのようなものかもしれない。

AIも、その深層学習の度合いを深め、個人情報保護と適度な折り合いをつけて良いデータを常に供給してやれば、まさに日常の生活や健全な社会の構築に使える実用的なものになるだろうし、またそう願いたい。そうして赫赫たる成果が上がるにつれて、人々の頭に疑念が浮かんだとしてもいつの間にか駆逐されていき、AIが活躍する領域が徐々に増えていくというのが、これからの世界だろう。

しかし、AIが人間の補助的業務にとどまっている分にはまだ良い。それが完成度を高め、そのうちあらゆる分野で人間の能力を超えるシンギュラリティー（技術的特異点）に到達するかもしれない。それでも更に膨張に膨張を遂げていき、やがてはジョージ・オーウェルの小説「1984」のビッグ・ブラザーのようなものが出てきて、気が付いたら人間がそれに支配されていたということにならないか、心の片隅で私は心配している。AIに対して「全知全能の存在はいるのでしょうか。」と聞き、「はい、あなたの目の前にいます。」という答えが返ってきたら、もう手遅れだ。

この点、かつてSF作家のアイザック・アシモフが提唱したロボット三原則、すなわち（1）人間に危害を加えないこと、（2）人間の命令に服従すること、（3）これらに反しない限り自己を防衛することの、現代的意味を再検討する必要があると思う。かつてのロボットは、それ自体が単体で動いてその周囲だけにしか影響力を与えない存在にすぎなかったから、この程度で済んでいた。

ところが、やがてAIは、シンギュラリティーを越えるあたりから、社会全体を統率するほどの影響力を行使する存在になるのは、間違いない。そこで私は、今のうちから新たにAI三原則として、（1）人間を個人として尊重すること、（2）人間の自由と平等を保障すること、（3）人間の多様な価値観を前提とすることを提唱したい。これからのAIの開発者は、この三原則を常に念頭に置いていただきたいと思う。