



【最新号】談 no.127 web版

自動化のジレンマ

表紙:池田晃将

本文ポートレート撮影:すべて坂本政十賜

Gallery:小林健太

自動化のコノテーション…… A I 研究の進展と自動化が意味するもの



杉本 舞

すぎもと・まい

京都大学大学院文学研究科科学哲学科学史専修博士後期課程指導認定退学。博士（文学）。専門は、科学技術史（20世紀）、コンピューティング史（1930年代～50年代における計算機科学前史、黎明期人工知能研究）、科学史教育の歴史。関西大学社会学部社会学科社会システムデザイン専攻准教授。著書に『「人工知能」前夜：コンピュータと脳は似ているか』（青土社 2018）、監訳書に『コンピューティング史：人間は情報をいかに取り扱ってきたか（原著第三版）』（共立出版 2021）他がある。

そもそもノイマン型コンピュータは、プログラムを入れたらどんな用途にも使える汎用機ですから、プログラムによってコンピュータのする仕事の性格もまったく違ってきます。コンピュータは結局道具、人間がやりたいことをやるための道具であって、そこには必ず人間の意思や制御が介在します。漠然と何かを生んでくれるものではない以上、やはり重要なのは、人間との協働のありようだと思います。

自動化は自律化をもたらすのか



鈴木貴之

すずき・たかゆき

東京大学大学院総合文化研究科単位取得退学。博士（学術）。専門は心の哲学。近年は、JST/RISTEXの研究開発プロジェクトなどで人工知能の哲学に取り組んでいる。東京大学大学院総合文化研究科 広域科学専攻 相関基礎科学系（科学史・科学哲学研究室）教授。著書に『人口知能とどうつきあうか』（勁草書房 近刊）、『100年後の世界』（化学同人 2018）、『ぼくらが原子の集まりなら、なぜ痛みや悲しみを感じるのだろうか：意識のハード・プロブレムに挑む』（勁草書房 2015）、編著書に『実験哲学入門』（勁草書房 2020）他。

生物の知能の基盤には、「これはおいしい」とか、「これは痛い」「ここは寒い」というような、自分にとっての利害関心があり、生物はそれを手がかりとして学習していくことができるわけです。しかも私たちは、学習課題を誰から与えられるわけでもなく、世界を自由に動きながらさまざまなことを学習していきます。人間と同様の自律的な知能をもつためには、同じように「勝手に学ぶ」ためのメカニズムが必要になるでしょう。

自動運転とトロリー問題……自動化・人工知能・倫理



笠木雅史

かさき・まさし

名古屋大学大学院情報学研究科准教授／専門は分析哲学、ロボット哲学、実験哲学。Doctor of Philosophy (Philosophy)(University of Calgary)。著書に『モビリティ・イノベーションの社会的受容：技術から人へ、人から技術へ』分担執筆（北大路書房 2022）、『実験哲学入門』分担執筆（勁草書房 2020）、『信頼を考える：リヴァイアサンから人工知能へ』分担執筆（勁草書房 2018）他。

自動運転車の倫理に関する問題は大きく二点に絞られています。一つは、どのようなアルゴリズムを自動運転車に実装するのかということで、これ自身が「トロリー問題」と同一視されます。もう一つは、事故が起きた時に誰が責任を取るのかという問題です。自動運転について倫理学者が論じることは、ほぼこの二つのどちらかです。

editor's note

自動化の台頭、問われる人間の知恵 (wisdom)

自動運転するAI

現在、急速にAIの開発が進んでいます。AI（Artificial Intelligence）すなわち人工知能は私たちの生活を便利に、そして豊かにする技術だと期待されています。社会生活にも活用され始めていて、たとえばクルマの自動運転は、その典型例でしょう。人による操作が一切不要な自動運転車の実現普及したあかつきには、交通事故の減少や渋滞の緩和につながり、運転が難しくなった人も引き続きクルマを利用できると期待されています。

ヨーロッパ各国の他、アメリカや南アメリカ、中国、シンガポールなどさまざまな地域で、現在、自動運転の実証実験が行われています。アメリカのSAE（Society of Automotive Engineers）が2016年に策定した自動運転のレベル分けによれば、レベル0（すべての環境において人が運転する）からレベル5（すべての環境においてクルマが自動運転する）まで6段階に分けられていて、レベルが高いほど自動化の度合いが高くなります。たとえば、レベル3までは、自動運転機能が使えなくなった時には、人が運転に復帰しなければならないなど、基本的に人が運転席に座り、自動運転の状況を見守る必要があります。人が運転に復帰する必要がないのはレベル4以上で、人が運転に一切かかわらない完全自動運転はレベル5になります。

日本政府が今後の自動運転の発展や普及の見通しをまとめました。2014年から改訂を続けてきた「官民ITS構想・ロードマップ」ですが、今後の自動運転実現に向けたシナリオです。レベル4の自動運転は、限定された地域のバスやタクシーとしてまずは実現し、その応用として限定地域での無人の配送サービスの実現化を目指しています。日本では、2023年4月からレベル4の自動運転機能を公道で使えるようになりましたが、自家用車での自動運転は、技術的に難易度が低い高速道路から先に実現していく見込みです⁽¹⁾。究極的には、どんな場所でも走れるレベル5を目指していますが、初めて通る道でも正しく車線や標識などを認識して走れるのが絶対条件。それをクリアできるまでには、まだ相当の年月が必要のようです。

クルマの運転は、一般的に認知、判断、操作の繰り返しだといわれています。運転手が、たとえば信号の色や歩行者などを認知し、クルマの速度を上げるか下げるか、左右に曲がるかなどを判断します。そして、その判断にもとづいて実際にハンドルやアクセル、ブレーキを操作するわけです。クルマの運転はこの繰り返しですが、この一連の作業は誰でも簡単に習得できるというものではありません。それを十全に行うためには、それなりの訓練が必要です。

たとえば、運転者は、歩行者や信号、標識、道路に書かれた白線（区画線）など、さまざまなものを認識する必要があります。基本的にこれらをすべて正確に認知する必要があり、どれか一つでも見逃せば、重大な事故につながる危険があります。人はこれらを目で見て、脳でそれが何かを識別し、身体動作につなげます。現実の道路沿いには、道路標識とは無関係な看板、さまざまな形の建物や樹木などがあります。道路上で工事が行われて通行止めになっていることもあるでしょうし、信号や標識の一部が街路などで隠れていることもあるかもしれません。道路の白線が薄くなっていたり、ほとんど見えなくなっている場所もあるでしょう。人は常識や経験から、まぎらわしい看板があっても標識と区別できるし、白線が消えていても推測で正しい位置を走らせることができます。そこで活躍が期待されるのがAIです。画像中の物体を識別する「画像認識」は、AIの得意分野の一つです⁽²⁾。このAIを搭載することで、一般道でも初めて自動運転が可能になるのです。

脳をまねるAIの登場

1956年に開催された「ダートマス会議」は学問分野としてのAI研究の出発点とみなされていますが、その後二度の黄金期（第1次、第2次）と二度の冬の時代を経験し、現在、AIは三度目の黄金期（第3次）をむかえているといわれています。そのAIブームのきっかけをつくったのが深層学習（ディープラーニング）です。

深層学習は、簡単にいうとコンピュータに大量のデータを読み込ませて、そのデータのなかから何らかの法則性をコンピュータ自らが見出していく手法です。たとえば、手書きの「4」の画像をたくさん読み込ませることで、数字の「4」がどんな形をしているか（判断基準）コンピュータが自ら学習します。「4」がどんな形の数字なのか人が教える（プログラムで指定する）必要はありません。学習の結果、コンピュータは初めて見る手書きの「4」を、多少のくせがあっても、数字の「4」であると正しく認識できるようになるわけです⁽³⁾。

第2次ブームまでのAIでは、主に人がコンピュータに問題を解くためのルールを与えたり、知識を詰め込んだりすることで、知能（知識）を再現しようとしてきました。それに対して、現在のAI研究で主流となっているのが「機械学習」です。機械学習とは、機械（コンピュータ）が自ら「学習」するためのしくみです。機械学習にはいくつかの手法がありますが、その一つが脳のしくみをまねて情報を処理する仕組みで、ニューラルネットワークと呼ばれています。このニューラルネットワークの手法を発展させたものが深層学習です⁽⁴⁾。

脳は、たくさんの神経細胞（ニューロン）からできていて、神経細胞どうしが互いにつながってネットワークを形成しています。一つの神経細胞は、シナプスと呼ばれる接続部分を通じて、他の多数の神経細胞から信号を受け取ります。そして受け取った信号の強さに応じて、その神経細胞は他の神経細胞へ信号を伝えます。脳はこうして神経細胞が次々と信号を伝えていくことで、情報を処理しているのです。ニューラルネットワークでは、脳の神経細胞の働きをコンピュータ上のプログラムで、人工的なニューロン（人工ニューロン）として再現します。人工ニューロンは、複数の数値（入力値）を受け取って、その入力値に応じた別の数値（出力値）を出力する「関数」です。一つの人工ニューロンは、複数の入力値を受け取ります。そして、それらの入力値にいくつかの計算をほどこした数値を出力するように設計されています。つまり、ニューラルネットワークでは、多数の人工ニューロンを複数の層に分けてつなげ、初めに入力値（データ）を次々と変換していくことで、情報を処理していくわけです。

ニューラルネットワークのなかで、人工ニューロンの層を多層化した（深くした）ものが深層学習です。深層学習が、高い性能を発揮したことで、第3次AIブームが起きました。こうしたブレークスルーはなぜ生じたのでしょうか。それは、深層学習によって、ものがもっているさまざまな特徴を、AIが自ら見つけ出すことができるようになったからです。

深層学習が登場する前の機械学習では、たとえば、チューリップとヒマワリを見分けるために「色と花びらのかたち」に着目しなさい」というように、着目すべき特徴を人が指示しました。AIは、人に教えられた特徴をもとに、「赤みがこの程度で、花びらがこのような形ならチューリップ」などのように学習していたのです。しかし、どのような特徴に着目するかによって、AIの性能（答えを予測する精度）は大きく変わります。深層学習では、大量の画像を読み込ませるだけで、画像に含まれる着目すべき特徴をAIが自分で抽出できるようになったのです。しかもAIが抽出する特徴には、画素と画素の複雑な関係のような、人が明確な言葉で表せないものや、人では捉えきれないものも

含まれています。そうしたことから、深層学習を使うことで、人が特徴を教えるよりもずっと精度の高い予測ができるようになったのです⁽⁵⁾。クルマの完全自動運転が可能になりつつあるのも、AIの深層学習の発展があったからに他なりません。

自動化の倫理学

深層学習の発展を後押ししたものに1990年代後半からのインターネットの発展・普及があります。従来のデータ処理ソフトウェアではとても扱うことのできない巨大な規模のデータ、いわゆるビッグデータがインターネット上に流通するようになり、そうした大量のデータを統計的手法やパターン認識の手法を用いて分析し、そこから有用な知識やパターンを発見するデータマイニングが急速に普及しました⁽⁶⁾。

深層学習は、人間の脳の神経細胞のつながりを模してつくられたシステムです。そもそもニューラルネットワークは、非常に複雑なシステムです。それゆえコンピュータが習得した独自の判断基準は、人はほぼ理解できないといわれています。先ほどの例でいうと、なぜコンピュータが数字の「4」を判断できるようになったのか、判断基準を示すデータを見ても理解できません。人は、AIの「頭の中」が理解できないため、たとえばAIのふるまいが途中からおかしくなったとしても、人がデータを直接書き換えて修正するのは困難です⁽⁷⁾。

じつは、このようなAIの特徴は、自動運転に利用する時に、問題となります。自動運転技術を語る際に避けて通れないのが一般的に「トロリー（トロック）問題」と呼ばれる倫理的な問題です。もともと哲学の応用倫理学の分野で話題になっていた問題ですが、それがクルマの自動運転技術においても同様の問題として、注目されるようになりました。

倫理学における「トロリー問題」は、およそ次のようなものです。「止まれなくなったトロリーがあり、そのまま進むと線路上で作業をしている5人をひいてしまう。その手前には分岐点があり、トロリーの進路を切り換えれば、5人の命を助けることができる。ただし、切り換えた線路にも作業員が一人いる。今度は、そちらの一人が犠牲になる。5人を救うために一人を犠牲にするか、一人を救うために5人を犠牲にするかあなたならどうする？」

同様の倫理的問題が自動運転でも考えられます。自動運転車の目の前に自転車が飛び出してきたとしましょう。クルマがそのまま進むと自転車は確実にクルマと衝突します。しかし、自転車を避けようとハンドルを切ると、対向車にぶつかり、今度は自分のクルマと対向車の乗員の命が危険にさらされる。自動運転車はどのような操作をするのが正解か。運転者（人）がいれば、運転者の責任を追求することができますが、もとより自動運転車ですから、運転者はいません⁽⁷⁾。

私たちは、ある人間に責任を問う時、問題の原因となった「行為」がその人自身によるものかどうかを考えます。哲学では、行為を生み出す能力をそなえた人を「行為者」といいます。つまり、責任を問う時は、その人間がそもそも行為者たり得るかを問う必要があるのです。

旧来、行為者は人間のみだと考えられてきました。なぜならば、行為者になるためには自由意志が必要と信じられていたからです。たとえば哲学者のカントは、自由意志にもとづいて自律的に行動できる人間を「人格」として、人格を他の事物と区別しました。この考えにのっとると、単に規則だけに従い、自ら動き自ら考えるAIは、行為者にはなり得ません。したがって、AIに責任を問うことは、原理的にできないのです⁽⁸⁾。

AIの開発と応用が急速に進んでいます。そして、さまざまな倫理的問題が生じています。その問題には、自動化というより根本的な問題が控えています。今号は、この自動化という問題を掘り下げようと思います。

まず、AI研究の進展と自動化が意味するものについて、科学技術史が専門の杉本舞先生にコンピューティングの歴史的検証をとおして考察していただきます。次に、自ら考え判断するAIは、自律する主体といえるのか。心の哲学が専門の鈴木貴之先生に、深層学習、ニューラルネットワークを手がかりに考察していただきます。最後は、自動運転の倫理的問題として注目されるトロリー問題について、笠木雅史先生に言及していただきます。（佐藤真）

引用・参考文献

- (1) newton別冊『ゼロからわかる人工知能完全版 ついにはじまったAI時代 社会や暮らしは急速に変化する』（ニュートンプレス 2022）P74-75
- (2) 同上 P66-67
- (3) 同上 P73-75
- (4) 同上 P16-17
- (5) 同上 P24-29
- (6) 雑誌『RIKEHAKU』特集「人工知能の現代史」（国立歴史民俗博物館 2022）P54-57
- (7) newton別冊『ゼロからわかる人工知能完全版 ついにはじまったAI時代 社会や暮らしは急速に変化する』（ニュートンプレス 2022）P73
- (8) 同上 P74



書物のフィールドワーク（発行年順）

◎オートメーションとAI

イラストで学ぶ 世界を変えたコンピュータの歴史 R・イグノトフスキー 杉本舞訳 創元社 2023

雑誌 REKIHAKU 特集・人工知能の現代史 国立歴史民俗博物館 2022

オートメーションと労働の未来 A・ベナナフ 佐々木隆治監訳 岩橋誠他訳 堀之内出版 2022

コンピューティング史：人間は情報をいかに取り扱ってきたか M・キャンベル＝ケリー、W・アスプレイ 他 杉本舞監訳 喜多千草他訳 共立出版 2021

人工知能のアーキテクトたち AIを築き上げた人々の語るその真実 M・フォード 松尾豊監訳 水原文訳 オライリージャパン 2020

IT全史 情報技術の250年を読む 中野明 祥伝社黄金文庫 2020

人類の歴史とAIの未来 B・リース 古谷美央訳 ディスカヴァー・トゥエンティワン 2019

「人工知能」前夜 コンピュータと脳は似ているか 杉本舞 青土社 2018

コンピュータ、どうやってつくったんですか？ 川添愛 東京書籍 2018
復刊 計算機の歴史 パスカルからノイマンまで H・H・ゴールドスタイン 末包良太他訳 共立出版 2016
コンピュータ開発のはてしない物語 起源から驚きの近未来まで 小田徹 技術評論社 2015
コンピュータ理論の起源 [第1巻] チューリング 伊藤和行編 佐野勝彦、杉本舞訳 近代科学社 2014
オートメーション・バカ 先端技術がわたしたちにしていること N・G・カー 篠儀直子訳 青土社 2014
実物でたどるコンピュータの歴史 石ころからリングへ 東京理科大学出版センター編 竹内伸 東京書籍 2012
コンピュータ開発史 歴史の誤りをたずねる「最初の計算機」をたずねる旅 大駒誠一 共立出版 2005

◎ロボット・認知科学・深層学習

未来につながる！ロボットの技術:歴史からしくみ、人工知能との関係までよくわかる 日本ロボット学会監修 誠文堂新光社 2023
AI・ロボットと共存の倫理 西垣通編 河島茂生他 岩波書店 2022
AI 2041 人工知能が変える20年後の未来 カイファー・リー、チェン・チウファン 中原尚哉訳 文藝春秋 2022
文系のためのめっちゃやさしい人工知能 松原仁監修 ニュートンプレス 2022
ディープラーニング 学習する機械 ヤン・ルカン、人工知能を語る Y・ルカン 松尾豊監訳 小川浩一訳 講談社 2021
脳と人工知能をつないだら、人間の能力はどこまで拡張できるのか 脳AI融合の最前線 紺野大地、池谷裕二 講談社 2021
AIは人類を駆逐するのか？ 自律世界の到来 太田裕朗 幻冬舎 2020
人工知能 チューリング/ブルックス/ヒントン 名著精選 心の謎から心の科学へ 関一夫・中島秀之監修 岩波書店 2020
今日、僕の家にはロボットが来た。 未来に安心をもたらすロボット幸学との出会い 上出寛子他編著 北大路書房 2019
ロボット法 増補版 AIとヒトの共生にむけて 平野晋 弘文堂 2019
AIの時代と法 小塚荘一郎 岩波新書 2019
AIに心は宿るのか 松原仁 集英社インターナショナル新書 2018
ロボット・AIと法 弥永真生、宍戸常寿編 有斐閣 2018
ロボット工学と仏教 AI時代の科学の限界と可能性 森政弘、上出寛子 佼成出版社 2018
人工知能はどのようにして「名人」を超えたのか？ 最強の将棋AIポナンザの開発者が教える機械学習・深層学習・強化学習の本質 山本一成
ダイヤモンド社 2017
スーパーインテリジェンス 超絶AIと人類の命運 N・ポストロム 倉骨彰訳 日経BPマーケティング 2017
鉄腕アトムは実現できるか？ ロボカップが切り拓く未来 松原仁 河出書房新社 1999

◎AI・倫理・信頼

ニュートン別冊 ゼロからわかる人工知能 完全版 ニュートンプレス 2022
私たちはAIを信頼できるか 大澤真幸、川添愛他 文藝春秋 2022
AIと人間のジレンマ ヒトと社会を考えるAI時代の技術論 中尾悠里 千倉書房 2022
AI新世 人工知能と人類の行方 甘利俊一監修 小林亮太、篠本滋 文春新書 2022
AI・データ倫理の教科書 福岡真之介 弘文堂 2022
責任あるAI 「AI倫理」戦略ハンドブック 保科学世、鈴木博和 東洋経済新報社 2021
AI新生 人間互換の知能をつくる S・ラッセル 松井信彦訳 みすず書房 2021
知ってるつもり 無知の科学 S・スローマン、P・ファーンバック 土方奈美訳 ハヤカワ文庫 2021
AIの倫理学 M・クーケルバーク 直江清隆他訳 丸善出版 2020
人工知能が「生命」になるとき 三宅陽一郎 PLANETS/第二次惑星開発委員会 2020
AI倫理 人工知能は「責任」をとれるのか 西垣通、河島茂生 中公新書ラクレ 2019
裏側から視るAI 脅威・歴史・倫理 中川裕志 近代科学社 2019
虚妄のAI神話 「シンギュラリティ」を葬り去る J-G・ガナシア 伊藤直子他訳 ハヤカワ文庫 2019
信頼を考える リヴァイアサンから人工知能まで 小山虎編著 勁草書房 2018
高校生のための ゲームで考える人工知能 三宅陽一郎、山本貴光 ちくまプリマー新書 2018
人工知能のための哲学塾 三宅陽一郎 ビー・エヌ・エヌ新社 2016
人工知能の作り方 「おもしろい」ゲームAIはいかにして動くのか 三宅陽一郎 技術評論社 2016
人工知能は人間を超えるか ディープラーニングの先にあるもの 松尾豊 KADOKAWA/中経出版 2015

