

<Book Review> Nishigaki, Toru, *AI Genron : Kami no Shihai to Ningen no Jiyu*, Kodansha, 2018

Katsuya YAMAMOTO

Abstract

Artificial Intelligence (AI) is attracting increasing attention. Some voices say that work efficiency and productivity is expected to improve, but on the other hand, there is a fear of loss of employment and the advent of a surveillance society. In order to identify the various problems relating to social implementation of AI and their relevant solutions, as well as to predict the future society that AI will bring, we must understand what AI is in the first place, what is or is not possible for AI, and what the goals of AI are. This book, *AI Genron* examines these questions from aspects of both technology and philosophy.

In this book, the author, Toru Nishigaki, first identifies the purpose of artificial intelligence, and then categorizes the differences between life and machinery in terms of autonomy and homogeneity. He points out issues in current AI development which are based on naive realism, and proposes speculative realism as an alternative philosophical basis. Finally, he explores Judea-Christ Monotheism for an ideological source of transhumanism and brings reader's attention to the potential dangers.

〈書評〉 西垣通著『AI原論 神の支配と人間の自由』 講談社、2018年

山本 勝也

はじめに

人工知能 (Artificial Intelligence: 以下、AI とする) に注目が集まっている。一方で作業の効率性が上がり生産性の向上を期待する声もあり、しかし他方では雇用の喪失や監視社会の到来への危惧も同時に囁かれている。このAIの社会実装にまつわる様々な問題やその解決法を提示し、AI がもたらす近未来社会を見通すためには、そもそも AI とは何なのか、AI には何が可能なのか (あるいは可能ではないのか)、そして AI は何を目指しているのか、ということが明らかにされなければならない。この問いに技術 (基礎情報学) と哲学の両面から取り組んでいるのが本書である。

以下、本書の概要を示し、その主張を確認する。その後、本書の提示する論点の検討と評価を試みたい。

1. 本書の概要

本書の構成は、以下のとおりである。

まえがき

第1章 機械に心はあるのか

- 1 AI ブームふたたび
- 2 生命と機械
- 3 ロボットという擬似生命

第2章 汎用 AI ネットワーク

- 1 脳型コンピューティング
- 2 シンギュラリティ仮説
- 3 クラウド・コンピューティング

第3章 思弁的实在論

- 1 相関主義と实在論

- 2 宇宙の安定性の根拠
- 3 物質・流動・生命
- 第4章 生命とAIがつくる未来
 - 1 相関主義への疑問をめぐって
 - 2 自由意思と責任のゆくえ
- 第5章 AIと一神教
 - 1 救済／創造／ロゴス
 - 2 選民による布教と情報伝播
 - 3 一神教は超克できるか
- 第6章 AIの真実－論点の総括

以下、本書の内容について各章ごとに概要を確認する。

第1章「機械に心はあるのか」では、まずAIという概念について知や知能という概念自体が曖昧であることを指摘した上で、著者はAIの目指す「知とは果たして、人間(あるいはより広く生命体)が生存するという実践目的か、それとも、絶対的・普遍的な真理に到達するという形而上学的な理論目的か」(12頁、以下の引用は全て本書より)と本書を貫く中心的な問いを提示する。そして、これまでのAI開発ブームであった第1次ブーム(1950～60年代)、第2次ブーム(1980年代)、そして現在の第3次ブーム(2010年代)への変遷において、「第1次ブームから第2次ブームにいたって前者から後者への舵が切られたのだが、第3次ブームではいっそう後者が強調されている」(20頁)という。これは、かつての論理記号の組み合わせと高速処理によって絶対的真理に至ろうとした時代(第1次)から、専門知識などのより有用な知識を命題として組み合わせでより有用な知識命題を得ることを意図する時代(第2次)を経て、「人間のような心をもつ機械」「人間のような感情をもつヒト型AIロボット」の誕生を目指す時代(第3次)への変化と言える。ここで、著者は第3次ブームに際して抜本的なパラダイム転換があったわけではないとしながらも、心や感情といったものが何なのかという根本的疑問に取り組む必要があると言う。なぜなら、AIが仮に人間のように判断し行動するとき、そこには自由意思や責任の問題が発生するからである。

そうした中で「人間のような意識をもって主体的に作動し、言語的相互行為をおこなえる機械、より広義には生命的な機械」(22 頁)としての強い AI への期待が現在高まっている。このとき一般の人々と同じく AI 研究者に共有されているのは「われわれ人間を取り巻く現実世界が客観的に実在しており、そこに我々が参加している」(26 頁)という素朴実在論である。この素朴実在論はデカルトやライプニッツによって確立され、カント以降の現代哲学によって徹底的に批判された。こうした哲学の流れと AI 研究との間には現在も大きな溝があると著者は指摘する。

また、著者はこうした強い AI への批判は、物理学や生物学などの自然科学の分野である 2 次サイバネティクス、オートポイエーシス理論、ラディカル構成主義心理学、機能的分化社会論などのネオ・サイバネティクスからも提出されているという。そもそも認知や心のことは外部から脳を見ればわかるというようなものではなく、観察点をその生命体の内部に移して、生命体が環境の中で活動しながら内的に世界を構成していくオートポイエティック(自己創出的)な観点から生命体は機械とは区別され、強い AI の可能性は否定されている。

ここで、著者の考える「機械と生命体の相違」をまとめておけば、機械は「開放的な他律システム」、生命体は「閉鎖的な自律システム」ということになる。機械は外部からその作動のルールを決定されているという意味で他律システムであり、また外部からそれを観察可能という意味で開放的である。しかし、生命体はそうではなく、自己の作動ルールを自ら外部環境に適応する形で変更していく。さらにそれだけでなく、その作動ルールが変化していく様子について、この作動ルールを変更していくルール(メタ・ルール)がどうなっているか、あるいはこのメタ・ルールがどのように変化していくかは観察・予測不可能であり、この意味でその行為を完全に予想することはできない。AI ではその実際上の困難性はあるが、原理的にはこれが可能であり、したがってこの点において、生命体は閉鎖的な自律システムであるとされる(著者によれば閉鎖性は不可知性と密接に結びついている)。この点は後述されるように自由意思や責任に関わる点でもある。

第 2 章「汎用 AI ネットワーク」では、現在のニューラルネット・モデルに

よる深層学習 (deep learning) の成功の中で、コンピュータの仕組みを脳に近づけたことによって人間のような心をもつ機械 (AI ロボット) が実現するのではないかという期待が高まっていることに対して、過剰な期待は禁物だと著者は釘をさす。その期待の典型は猫の画像を認識したという Google 社の実験から「AI が自ら、人間社会で通用する概念を学習するようになった」(48 頁) というものだが、著者に言わせればこれはあくまでパターン学習による分類が一致しただけのことであり、概念とは外見上の差異だけからえられるものではなく、そうした外見上の差異を超えたスーパーマーケットでのハマチ (養殖) とブリ (天然) のような概念分類もあることを示している。深層学習の有用性は、人間の概念操作とは全く別の次元のものであり、この知的混乱の原因は AI 研究者が追及している「人工知能」とは何なのか、ということの混乱に起因すると著者は見ている。

「人工知能とは何か」という問いに対して、著者は人間知の模倣か、あるいは普遍的真理を求める知の実現かという整理をする。これは、冒頭の問いに対応している。前者の場合、素朴実在論的に脳が心を作る実体であると考えれば脳をシュミレートすれば心をもつ機械ができるということになる。しかし、ここで著者は「観察者の視点」という本書における重要な論点を提示する。脳が心を作るというのは観察者が心を外側から眺めているからそう思えるだけで、脳科学の記述は、心のごく一面を捉えるに過ぎないという。そして、この点をわきまえないと、絶対的真理に至る知の実現という幻想に陥るといふ。つまり脳科学のように人間の脳を外部から眺めて、これをコンピュータ上に作りこむというのでは、人間の脳が完全なものではない以上、人知を超える絶対知を自ら追求する AI を創り出すことは困難なはずであり、「人知の模倣」は必ずしも「絶対知の実現」にはつながらない。ところがレイ・カーツワイル (Ray Kurzweil) のシンギュラリティ仮説などのトランス・ヒューマニズムが目指す「人間を超える知能をもつ機械」は、脳のリバースエンジニアリングやマインドアップローディングという人知の模倣を通じて絶対知の実現を目指している点で、大きな矛盾に陥っていると批判される。

さらに著者は、個々の専用 AI がネットワークで繋がったクラウド AI ネットが「人間を超える知能」として汎用 AI になるということについても否定的

である。著者によれば、クラウド AI ネットの作動を記述しているのは人間であり、真の意味で人間を超える知力を持ち、絶対知にたどり着けるという保証はない。著者の専門である基礎情報学的に分析すると、クラウド AI ネットが普遍的な絶対知の実現を約束するとはどうしても言えないというのが、ここでの結論である。

第3章「思弁的实在論」では、この AI 開発を根底から基礎づけるものとしてカンタン・メイヤソー (Quentin Meillassoux) の思弁的实在論 (speculative realism) が有効かどうかを検討している。思弁的实在論はカント以降の哲学の主流である相関主義への挑戦を試み、科学技術と哲学との間の亀裂を解消することを目指しているという。思弁的实在論はそのために相関主義を徹底することで、われわれ人間の関与なしに即自的存在 (=カントの物自体という概念から認識不能というニュアンスを除いた实在) にアクセスし、これについて語るができるという新しい实在論を主張するものである。

著者の整理ではメイヤソーは、われわれ人間の外部にもはやデカルトの考えたような究極的存在を仮定することはできず、理由律にもとづいて即自的存在の必然的なありさまについて語るのはおかしいと批判する。そして、理由律を否定し、絶対性と必然性のつながりを断とうとする。世界がこのようなになっているという「事実」は絶対的なものとして受け入れるが、それが必然だということはないとみなすものであり、それ以外の事実の可能性を認め、現状は絶対的事実だがあくまでも「偶然的」だと考えるのである。われわれの主観とかかわりなく、世界には絶対的な事実があり、その事実の出現は「偶然だ」という主張である。これならば人間とは無相関に事実としての即自的存在が实在するので、これを数学的に記述することは正当性をもつことができる。

さらにメイヤソーは確率論的な運とより本質的な偶然性とを区別する。確率論的にある事象が生起するのも一種の必然性であり、これは運による生起であるが、偶然性はこれとは異なり、あくまで必然性の対立概念とされる。あらかじめ準備されたのではない未知の法則や事象が、潜在性によって、時間経過とともに偶然に生成される。つまり、思弁的实在論では事象からなる世界の「全体」が確定しておらず、その不確定なものが潜在性によって時間

とともに創発されると考える。こういう前提のもとで、本質的に世界は潜在性をもっており、偶然に未知の法則や事象が生起することは十分にありうるし、たまたま安定状態がずっと続くということもありうる。その限りにおいて、局所的に科学的探究は成立するし、科学技術も有効性を発揮することになる。

しかし、こうした思弁的实在論の潜在性に関する考え方は、AIとは相容れない部分があるというのが著者の評価である。機械的思考とは、基本的には潜在性ではなく、潜勢力（確率的に生起する全体集合の各要素）を前提にしたものなので、過去のデータに基づく確率モデルから結果を回答する。しかし、生命的な思考とは潜勢力を前提としたものばかりではない。過去に遭遇したことのない全く新たな状況においても生きぬこうとする意味において潜在性にも関連している。そして、その潜在性を生起させ現実化するのには、時間である。時間とはパラメータではなく、「刻々とリアルタイムで流れていく、繰り返せない時間」（93 頁）である。それは潜勢力に含まれてはいない、全く新しい法則を生み出すことさえできる。生命の誕生はまさに潜在性にもとづく創発の例である。生命体は、本質的にこうした生成と創発と不可分であり、「生命体はあくまでも現在この瞬間の刻々に応じて生きており、過去のデータ処理に依存して動作する機械とは根本的に違う」（同頁）のである。AIがアルゴリズムによってプログラムされる限り、それは潜在性をもちえないという意味で人間を超えるものにはならないということになる。

また思弁的实在論では、生物とは物質と物質の連絡である「流動（flux）」の一部を身体的に選別し希薄化した環（ループ）のような存在で、生存と無関係なものを排除し、意味あるものを選ぶと考える。このとき生物の死としては、無関心が増大することでもたらされる死と関心が拡大しすぎて消散する死とがある。近代的進歩主義のもとでは、関心の拡大が望ましいとされ、この延長上に AI 開発もあるが、しかし、この関心の過剰な拡大によって、人間は限りなく呆けていく存在となり、無限の狂気による死に至るとされる。著者によれば、これは AI のひらく未来への厳しい警告となっているという。

第 4 章「生命と AI がつくる未来」では、以上の思弁的实在論の議論を踏まえて、AI 開発の哲学的基盤としては思弁的实在論だけでは頼れず、相關主義

哲学のなかで AI を正しくとらえなおすべきであると主張する。

ここで著者は、強弱の相関主義と主観主義の形而上学をまとめて相関主義思想と呼び、ネオ・サイバネティクスと相関主義思想との共通点について観察者の視点や主観との関係において整理する。著者は、思弁的実在論が事実であるという「祖先以前の言明」についても1つの科学的仮説であり、現在のところ多くの科学者に信じられている「世界についての物語」に過ぎないという。そして時に誤りが指摘され、修正されていくという意味で、現代の科学技術的な知も、厳密には相関主義思想にもとづくものと言わざるをえない。ゆえに、現代の科学技術の哲学的根拠を明確にするには、むしろ相関主義思想と類縁関係にあるネオ・サイバネティクスに依拠すべきであるというのが著者の主張である。

そして、著者は AI と自由意思や責任について検討をする。脳決定論によれば、それは脳内の化学変化の結果でしかなく、自由意思は認められない。これは素朴実在論を前提とする議論であり、これによれば他律システムである AI には自由意思はもてない。

しかし、AI の選択枝の自由の有無から考える「両立論」ではどうだろうか。ほかのことは選ぶことが可能だが、あえてそれを選んだために発生したことに對しては責任をもつという「両立論」では、その選択枝しかない場合には免責される。AI には基本的に選択枝の自由度などない。アルゴリズムに従うだけの他律システムだからである。しかし、クラウド AI ネットにおいて AI と会話している人間には、相手が AI なのかどうかの判断はつかない。AI からの指示で投資判断を誤ったと主張する可能性もありうる。

こうして両立論にもとづけば、AI の普及とともにその統一的な責任概念が不明確になるので、AI がマイナスをもたらす元凶である時、責任の所在はいまいにならざるを得ず、AI への責任転嫁の誘惑に駆られる危険性がある。また、さらなる懸念は、ある問題の意思決定において、人間と AI の処理が複雑微妙に入り組んで、どこまでが人間の判断で、どこまでが AI の判断なのか、切り分けが非常に困難になることだと著者はいう。こうなると責任という概念そのものが曖昧になり、崩壊する恐れがあるとの指摘である。

つまり、せつかくの AI が一部の人々の支配欲を満たすために利用され、し

かもマイナス効果は不問に付されるという無責任体制がもたらされる懸念があるというのである。将来的には、AIが人種差別の隠れ蓑として利用される恐れもあり、効率よく人間を自動分析し、レッテルを張る処理には深刻な懸念が提示されている。

第5章「AIと一神教」では、AIの目的としての普遍的な絶対知の実現の思想的ルーツを探っている。欧米において何がトランス・ヒューマニズムを受け入れやすくさせているのかという問題である。こうしたトランス・ヒューマニズムとAIのような欧米の進歩主義の深源に、著者は一神教の思想を見ている。ここでの一神教とは、古代ユダヤを核としたキリスト教思想である。

カーツワイルや Google 創始者など優秀なコンピュータ科学者にはユダヤの知を引くものが非常に多いことを挙げ、ユダヤ人のこの分野での影響力を確認しながら、次いでユダヤ教の救済思想が普遍的に受け入れられていく過程を歴史的に押さえていく。そしてトランス・ヒューマニズムは創造神・ロゴス中心主義・選民思想という3つのキーワードで説明可能であると著者は言う。救済計画をもつ唯一神がこの宇宙を創造したのであり、その設計思想はロゴスをもとに記述されており（正確な推論によって真理にたどり着けるというギリシア哲学との一体化）、そして選ばれた人間（聖職者）だけがロゴスを解釈しつつ、正義の実現を目指すことができる。

そして、パウロの教会による「信じるものは救われる」という教義がキリスト教を非ユダヤ人にも広め、ローマ帝国の社会的秩序を安定させることに貢献した。これをローマの施政者は利用するべく、キリスト教を公認する。そして西暦325年ニケーア公会議において三位一体説が正統として採択されるが、三位一体説は支配者に都合の良い教義であったとも著者は言う。イエスの言葉は神の言葉なのだから、この言葉を説く人間は神としての人間となり、絶対の権威として、キリスト教が全世界を征服していく原動力になったのである。そして、著者はAIの目指すもののあいまいさの根源は、人間と神の接近あるいは接合さえも保証する三位一体説の教義にあるとしている。

やがて中世において、神罰を恐れて慄いていた一般の人々は、明晰な理性をもつ個人として啓蒙され、合理主義という共通理念のもと、自由平等な社

会の建設や科学技術の発展に向かって努力するようになった。これが進歩主義の土台となっている。

そして、米国を起点とする経済的グローバリゼーションの進む現在、「成功こそが神による救いが予定されている証明だ」(163 頁) というカルヴァン派の予定説の影響を受けて、新たな拝金教選民思想が登場していると著者は言う。イエスは貧しきものに手を差し伸べたが、今や、救済は富める者に与えられることになり、さらに、そういう富める者が携わるビジネスが、論理や実証、そして科学技術的な知識としっかりと結びついていることに著者は注意を促す。研究開発の現場のイデオロギーは「この世での金銭的成功」や「短期的な地位権力の獲得」に標準が合わされているのだと言う。

そして、科学技術の進歩主義は新たな局面を迎えているというのが著者の指摘である。それは、ごく一部の成功者と大多数の貧困者への社会の分裂である。さらに、この大多数の人々が頼るものが「新たな神秘的物語」としてのシンギュラリティ仮説であり、この仮説はそういうものとして機能し始めていると著者は言う。

AI とはいえ機械であり、これを秘密裏に操作することは可能である。著者の危惧は「AI に疑似主体の衣を着せ、その出力を神のご信託のごとく一般の人々に信じさせるとすれば、それは 21 世紀の巧みな宗教的支配」(166 頁) となることである。

最後に、著者は AI 専門家の無知と欲望さに注意するべきだとした上で、「本来、虐げられた貧しい人々を救うために生まれた宗教が、地球上の多様な文化を破壊し、有色人をはじめ多くの人々を残酷に搾取する道具として悪用されたという歴史的体験」(167 頁) をもっと真剣に考察する必要がある、AI に関する神話的な幻想は、そうしたものとして機能する危険性があるとしている。

そして第 6 章では、これまでの論点を総括した上で、安易な AI 信奉に警鐘を鳴らし、人間の生きる力を増幅させる方向での AI の活用、Intelligence Amplifier としての AI を目指すべきであると結論づけられている。

2. 本書の特徴と若干の考察

本書の特徴としては、第1に、現在ブームになっているAIに対する冷静な分析視点であろう。シンギュラリティ仮説などセンセーショナルに報じられたが、こうしたトランス・ヒューマニズムについてその実現可能性はないと判断している。単体での強いAIではなく、個々の専用AIがクラウドネットを通じて連結し、相互に情報を交換しつつ共有するクラウドAIネットについて、将来的に強いAIに最も近い存在になる可能性を指摘しているが、それでも実際には他律システムであるということから人知を超える存在となる可能性については否定的である。

第2には、以上の分析を哲学的側面と同時に科学技術的側面(基礎情報学)からも分析する点である。哲学的な関連主義思想とオートポイエーシスに代表されるネオ・サイバネティクスとの共通点を観察者の視点として取り出し、そうした観察者(主観)の視点が生命体における世界認識に重要な意味を持ち、その作用が人間の心を形成していくという主張に評者も共感する。また、世間に流布する「自律システム」という用語の曖昧さに対して、「自律性」を自らの作動ルールを自ら決定することはもちろん、その作動ルールをどのように変化させていくかというメタ・ルールも変化し、このメタ・ルールの変化が外部からは計り知れないことと定義していることは、評者には非常に納得のいくものであった。この点から考えれば、どんなAIもプログラムされたものであるという意味において他律システムでしかない。

これとの関わりで、第3に挙げるべきは、この時の社会的責任の所在である。責任という概念が自由意思と分かち難く結びついている限り、他律システムであるAIにはその責任を問うことはできない。しかし、クラウドAIネットが実装された社会システム(暫定的閉鎖系)においては、この問題が無責任システムへと変容する。個々人は、自らのコミュニケーションの相手が人間かAIであるかを判断することが難しくなり、擬似的自律性を持つようにみなせるAIの判断をもとに判断するようになる。この時AIの判断が誤っていた場合、その責任を誰に問えるのか。加えて、コミュニケーションの相手として人間やAIを数多く巻き込む時、どの程度AIの影響を受けてなされた判断なのかをいうことは難しい問題となる。すると、このAIの運用者は、

責任回避の手段として AI を用いるようになるだろう。

こうした事態は、経済学が前提としている自己の合理的判断による選択とその帰結に関する理論的基盤を崩壊させてしまう可能性も持っている。かつて J.K.ガルブレイスは『ゆたかな社会』の中において、近代的な宣伝とマーケティングがそれまで存在しなかった欲望を掻き立て、自立的に決定された欲望とは異なる形の消費が増加していることを「依存効果」として指摘した¹。こうしたことはすでに我々の日常でますます中心となりつつあり、閲覧履歴から AI がオススメ商品を広告に表示する自動システムは、私たちの選択行動に大きな影響を与えていると推測できるが、私たちがどの程度自立的な判断で行動を決定しているかをいうことは難しくなっている。

例えば、何かに操られて消費するということが消費行動の中心となった場合、これに気づかず消費を続けるならば自由な個人という概念も怪しくなり、これを合理的な選択という枠組みで分析し続けることも妥当性を欠くことになるだろう。

そして、第4に、こうした AI が一部の富裕者によって利用される危険性の指摘である。AI の開発に携わるのは一部の IT 企業やその経営者たちであり、まさに巨万の富を動かす人々である。彼らが AI の開発をする目的は短期的な利益追及であると著者は喝破している。こうした富裕層によって利用された AI によって構成される社会は、まさにジョージ・オーウェルが『1984 年』で描いたディストピアへと至るようにも思われる。すでに個人的な信用スコアの査定や再犯可能性の判定などに AI が利用されているが、こうしたシステムがどのような社会を作っていくのかということに注意を向ける必要がある。

意図して AI を悪用しなくとも、AI 自身のバイアス（偏見）が社会をより維持不能な方向へ変化させる危険性も考えられる。これはアルゴリズムの偏見（algorithmic bias）といわれるもので、社会に存在する偏見が AI の利用により強化・固定化される現象である。AI のプログラム時やビッグデータを取り込んで学習するプロセスを経て、人間が持っている偏見が AI に反映され

¹ Galbraith, 1998（邦訳、第 11 章）。

るのである²。

こうした AI によるバイアスは、その判断に対して著者が言うように「ご神託」としての地位を与えることの問題を示している。AI の判断を「ご神託」として信じ込ませようとする者が出てこないとも限らないし、そうでなくとも「ご神託」と勘違いする人々が出てくるだろう。救われない貧困の最中にある一般の人々に対する新たな神秘的物語としてシンギュラリティ仮説あるいはトランス・ヒューマニズムが浸透していく時、世界は大きな混乱と破滅への道を歩むことになるかもしれない。特に、人間に関する分類・評価を AI に委ねることの危険性は指摘しなければならないだろう³。効率性を重視することから導入される AI が、そのバイアスによって人間の自由を奪い、社会の多様性を奪う危険性も併せ持つことは明記されるべきである。

おわりに

AI の社会実装は着々と進展している。しかし、これに対する我々の準備は未だ整っていないとは言えない。AI へのリテラシーをまずは高めて、何ができて何ができないのかを見定める必要がある。その上で、本書の著者は AI よりもむしろ IA (Intelligence Amplifier) を目指すべきであるという方向性を示している。この著者の意見に評者も全く同意する。人知を支え、これを増幅させるものとして活用する方法とその導入に向けての研究が進められる必要がある。

参考文献

- Galbraith, J. K., 1998, *The Affluent Society (Fortieth Anniversary Edition)*, Houghton Mifflin Company, Boston (邦訳：ガルブレイス著、鈴木哲太郎訳、『ゆたかな社会 決定版』岩波書店、2006年)。
- 久木田水生、2019、「人を評価する人工知能が人間同士の関係に与える影響とその倫理的含意」『三田評論 ONLINE』、2019年2月5日、https://www.mita-hyeron.keio.ac.jp/features/2019/02-4_3.html (2019年2月21日閲覧)。

² 村上、2018、159頁参照。

³ この点については、久木田、2019が指摘する倫理面への影響も参照されたい。

村上祐子、2018、「人工知能の倫理の現在－研究開発における技術哲学・倫理の意義－」『電子情報通信学会 基礎・境界ソサイエティ Fundamentals Review』、 11(3)、155-163 頁。