

人工知能の倫理：何が問題なのか

久木田水生

2017年11月6日 AI ネットワーク社会推進会議資料

1. 序	1
2. 各国の動向	3
3. 文献紹介	12
4. 人工知能の何が倫理的に問題なのか	25

1. 序

近年、深層学習の成功や、チェスや将棋、クイズなどにおけるパフォーマンスの目覚ましい向上によって人工知能への注目が高まっており、「第三次人工知能ブーム」と呼ばれている。「第一次」（1950年代）、「第二次」（1980年代）のブームの後には、大きく膨らみすぎた期待に応えるほどの成果がなく、その後、人工知能研究は「冬の時代」を迎えた。しかし今回のブームは、実際に人工知能が様々な状況、用途で実用化されているという点で、以前の2回のブームとは大きく異なっているように思われる。人工知能の性能の飛躍的な向上にとともに、人工知能技術がコミュニケーション、交通・運輸、流通、販売、医療・ケア、エンターテインメント、芸術、教育、育児、警察、軍事など、およそありとあらゆる場面に応用されるようになってきているのである。

人工知能がIoT（Internet of Things；様々な機器、装置に情報通信機能が付加されることによって形成されるネットワーク）の中に組み込まれる、あるいはロボットという物理的身体の中に組み込まれると、現実の物理的世界とのダイレクトな相互作用が増大する。あるものは私たちの生活に密着した場面で活用され、あるものは私たちの財産や健康、あるいは生命にまで重要な影響をもたらしかねないような状況で利用されている。例えばソフトバンクの人間型ロボット「Pepper」は店頭やショッピングモールで買い物客と会話を交わす。iRobot社のRoombaは家庭で床の掃除をする。MITのシンシア・ブリージュールらの開発している「Jibo」のプロモーションビデオでは、Jiboが家庭で小さな子供の話し相手になったりお話を聞かせたりする場面が描かれる。IBMのWatsonは医師が患者の診察をするのを

助けている。昨年、Watson が膨大な医学論文を参照することで、ある患者が白血病である可能性を医師に示唆したというニュースが話題になった。人工知能を搭載した致死的自律型兵器を開発している国がある。また世界中の自動車メーカーや、Google のような IT 企業が自動走行車の開発を競っている。

このような状況の中で人工知能が社会や人間に与える影響についての懸念が非常に高まっている。例えばビル・ゲイツ、スティーヴン・ホーキング、イーロン・マスクといった著名人たちが人工知能の危険性を訴えている。ゲイツは「人工知能は核兵器ほど危険になりうる」と述べている。イーロン・マスクは人工知能の安全な発展のための研究を促進するために、Future of Life Institute に 1000 万ドルを出資している。現在、人工知能の開発と応用に関する法律や倫理的指針について、草の根からアカデミック、産業、政府、そして政府間レベルに至るまで、様々な議論が世界で同時並行的に始まっている。特にアメリカやヨーロッパではそういった議論をするための組織や会議が数多く形成されており、日本よりも進んでいるように思われる。

現在、人工知能にかけられている期待は非常に大きい。しかしながらその一方で人工知能に対する警戒感も大きい。人工知能の安全な発展を促進し、リスクを適切に管理する仕組みと、社会の人工知能に対する知識と理解を促進する仕組みが求められている。そういったことが適切に機能しなければ人工知能は危険なテクノロジーとして忌避されて、その有益な発展が阻害される可能性もあるのである。テクノロジーの発展と、法や制度などの社会システム、価値観や倫理などの人間的なファクターができるだけ激しく衝突しないようにしていくことが重要である。そしてこの点に関してアメリカやヨーロッパの取り組みに学ぶべきことがあるだろう。

本報告書では人工知能と法や倫理に関する世界の動向について報告する。上述のように人工知能に関する議論は様々な地域、レベル、形式で行われている。それらについて分類・整理したうえで特徴づけることは有用であるが、本報告書では大雑把な分類にとどまっている。地域の分類としては大きくアメリカ、ヨーロッパ、日本で分け、主体としては政府、NGO、アカデミック、企業に分けた。ただしこれらの分類は必ずしも厳密なものではない。

また報告書の後半には近年（大体 2014 年から 2017 年の間）発表された文献の紹介を行った。ここ数年は人工知能やロボットに関連する著作が非常に数多く出版されており、ここで紹介できたのはその一部であるが、できるだけ多様な観点、話題をカバーできるように選んだつもりである。

本報告書の最後では、人工知能と倫理に関してどのようなことが問題とされているのか、

あるいはされるべきなのかを簡単に論じる。

2. 各国の動向

2.1. アメリカの動き

2.1.1. 政府主導

PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE

ホワイトハウスが 2016 年 10 月に発表した報告書¹。アメリカ合衆国大統領行政府 (Executive Office of the President)、アメリカ合衆国科学技術政策局 (National Science and Technology Council)、科学技術委員会 (Committee on Technology) が作成したもの。特に機械学習に焦点を当てている。公共利益のための人工知能の応用、連邦政府における人工知能活用のための政策や指針、人工知能の規制、研究と活用のための人材育成、人工知能と経済、人工知能と公平性・安全性・ガバナンス、安全保障などが論じられている。多様性を配慮した設計、倫理性などが重要とされている。

Artificial Intelligence, Automation, and the Economy

ホワイトハウスが 2016 年 12 月 20 日に発表した報告書²。アメリカ合衆国大統領行政府 (Executive Office of the President) が作成。人工知能に駆動された自動化が経済に与える影響、それに関連する政府の政策について論じたもの。そのような影響の正確な予想は困難としながら、以下の点について政策決定者は準備をしておく必要があるとする。

- 生産性の向上を集積することへの貢献
- 高度にテクニカルなスキルへの需要の拡大など、労働市場において要求されるスキルの変化
 - 異なるセクター、賃金水準、教育水準、職種、地域の間での自動化の影響の不平等な分配。
 - 消える仕事と新しく作り出される仕事による労働市場の攪拌

1

https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

2

<https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF>

- 特定の労働者にとって短期的な仕事の消失、そして政策次第で長期的消失になりうる可能性。

2.1.2. NGO

Future of Life Institute

「命を守り、未来についての楽観的なビジョンを発展させるための研究とイニシアティブを促進し支援する」ことをミッションに掲げる団体³。創始者はジャン・ターリン (Jaan Tallin) (スカイプ共同創始者)、マックス・テグマーク (Max Tegmark) (宇宙物理学者)、Meia Chita-Tegmark (教育学者、哲学者)、Viktoriya Krakovna (DeepMindにおけるAIの安全性についての研究者)、Anthony Aguirre (理論宇宙物理学者)。Scientific Advisory BoardにはAlan Alda、ニック・ボストロム (Nick Bostrom)、エリック・ブリニョルフソン (Erik Brynjolfsson)、ジョージ・チャーチ (George Church)、モーガン・フリーマン (Morgan Freeman)、Alan Guth、スティーヴン・ホーキング (Steven Hawking)、Christof Koch、イーロン・マスク (Elon Musk)、Saul Perlmutter、Martin Rees、Francesca Rossi、スチュワート・ラッセル (Stuart Russell)、Frank Wilczekらが名を連ねる。

2015年1月にイーロン・マスクがAIの安全な発展を促進する研究を助成するために1000万ドルを寄付した。

「An Open Letter: RESEARCH PRIORITIES FOR ROBUST AND BENEFICIAL ARTIFICIAL INTELLIGENCE」⁴と「AUTONOMOUS WEAPONS: AN OPEN LETTER FROM AI & ROBOTICS RESEARCHERS」⁵という公開書簡を出している。前者はAIの安全で有益な発展のための研究を呼び掛けている。後者はAI研究者たちに自律型兵の開発を進めないよう呼び掛けている。

2017年2月3日に人工知能の研究課題、倫理と価値、長期的な課題に関するガイドラインとして「Asilomar AI Principles」を発表した。Research Issues、Ethics and Values、Long-term Issuesの三分野に分かれて、全部で23の原則がある。Ethics and Valuesの分野では安全性、透明性、責任、価値との調和、プライバシー、自由、利益の共有、人間による制御、社会的市民的プロセスの尊重、AI軍拡競争などに関する原則が挙げられている。

³ <https://futureoflife.org/>

⁴ <https://futureoflife.org/ai-open-letter/>

⁵ <https://futureoflife.org/open-letter-autonomous-weapons/>

Artificial Intelligence Now

Kate Crawford (Microsoft Research, NYU) と Meredith Whittaker (Google Open Research) によって運営されているニューヨークに本拠地を置くリサーチ・イニシアティブ⁶。

2016年7月7日にホワイトハウスとニューヨーク大学の情報法研究所主催のシンポジウム「AI now public symposium」を開催。その報告書が公開されている⁷。ここでは医療、労働、不平等、倫理が特にフォーカスされた。倫理に関して言うと、AIによってもたらせる不平等やバイアスに対処する必要性が唱えられている。またAIに意思決定をゆだねる際に、人間が犯さないような失敗をAIが犯す可能性が指摘され、既存の倫理枠組みでは対処できないと主張されている。

The Future Society

Harvard Kennedy School 発の NGO。爆発的に発展するテクノロジーのもたらす深甚な帰結に対処することを目的にしている。この中に特に人工知能にフォーカスした The AI Initiative がある。

2.1.3. 企業

DeepMind

DeepMind はイギリスの人工知能研究所であったが 2014 年に Google に買収され、現在は Alphabet グループの傘下にある。買収される際に、条件として倫理委員会を設置することを Google に要求したと言われている。

2017年10月3日に DeepMind Ethics & Society という研究ユニットを立ち上げた。その目的は「現実世界への AI のインパクトを探求すること」とされており、Key Ethical Challenges として以下のものが挙げられている。

- Privacy, transparency, and fairness
- Impact on economy: inclusion and equality

⁶ <https://artificialintelligencenow.com/>

⁷ https://artificialintelligencenow.com/media/documents/AINowSummaryReport_3_RpmwKHu.pdf

- Governance and accountability
- Managing AI risk: misuse and unintended consequences
- AI morality and value
- AI and the world' s complex challenges

2.1.4. アカデミック

IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems

2016年12月、`Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems (AI/AS)''と題した文書を公表して、技術者が人工知能と自律的システムを作る際に倫理的考慮をすることを奨励するためのパブリックコメントを募っている。

The One Hundred Year Study on Artificial Intelligence

スタンフォード大学が主催している研究プロジェクト⁸。AI100とも表記される。100年という長期的視野で「人工知能の影響がどのように人々の働き方、生き方、遊び方のあらゆる側面に波及していくかを研究し予見する」ことを目的に設立されている。によって提唱された。中心になっているのは以下の研究者。

マイクロソフトのエリック・ホロヴィッツとスタンフォード大学のラス・アルトマンが提唱して設立した。

AI100は2016年9月にARTIFICIAL INTELLIGENCE AND LIFE IN 2030というタイトルの報告書を出した⁹。この報告書では運輸、サービス・ロボット、医療、教育、低資源コミュニティ、公共の安全と安全保障、雇用と労働環境、エンターテインメントなどの分野においてAIが最も影響を持ちうるとし、そしてそれぞれの分野の課題について論じている。大衆的なジャーナルで書かれるようにAIが近い将来に人類にとっての脅威になる可能性は少ないとしながら、経済と社会に対して新しい困難を生じさせるだろうと述べる。報告書によれば、

⁸ <https://ai100.stanford.edu/about>

⁹

https://ai100.stanford.edu/sites/default/files/ai100report10032016fnl_singles.pdf

社会が AI に不安と不信をもってアプローチするならば、AI の発展を遅滞させる失策が生じ、AI 研究は地下で行われ、安全で信頼性の高い AI 技術を保証できなくなる。それゆえオープンな気持ちでアプローチすることが重要である、と報告書は述べる。

この報告書は 2015 年に開かれた Study Panel に基づいており、そのメンバーにはピーター・ストーン (Peter Stone)、ロドニー・ブルックス (Rodney Brooks)、エリック・ブリニョルフソン (Erik Brynjolfsson)、ライアン・カロ (Ryan Calo) などが名を連ねている。

Machine Intelligence Research Institute

「人間より賢い機械知能が良いインパクトを持つことを保証する」という目的のために設立された研究機関¹⁰。略称は MIRI。基礎的な数学的研究を行う。

We Robot

Miami University School of Law が 2012 年から主催している学会。ロボットに関連する法的問題にフォーカスしている¹¹。

2.2. ヨーロッパの動き

2.2.1. 政府主導

Robolaw

2012 年から 2014 年まで実行された研究プロジェクト¹²。急激に発展するロボット技術の法的倫理的含意を理解することが目的。

2.2.2. NGO

¹⁰ <https://intelligence.org/>

¹¹ <http://robots.law.miami.edu/>

¹² <http://www.robolaw.eu/>

Responsible Robotics

Aimee van Wynsberghe (Twente University) とノエル・シャーキー (Noel Sharkey) (University of Sheffield) によって設立され、本拠地を Twente に置く NGO。「責任あるロボット工学」を目指し、人文社会学者と工学者が共同している。Executive Board には以下のような人々が名を連ねている。

- ジャンマルコ・ヴェルジオ (Gianmarco Verrugio) : イタリアのロボット工学者。ロボット工学に関する倫理的問題について、早くから取り組んでいた。`Roboethics` という言葉を作った。
- ガルリ・カスパロフ (Garry Kasparov) : ロシア出身の元チェスの世界チャンピオン。
- シェリー・タークル (Sherry Turkle) : MIT の心理学。`Second Self`、`Alone Together`、`Reclaiming Conversation` (邦訳は『一緒にいてもスマホ』) などの著作がある。
- デボラー・ジョンソン (Deborah Johnson) : University of Virginia。倫理学者。特にコンピューター倫理の専門家。
- パトリック・リン (Patrick Lin) : California Polytechnic State University, San Luis Obispo。哲学者。`Robot Ethics` の共編者。
- ウェンデル・ワラック (Wendell Wallach) : Yale University Interdisciplinary Center for Bioethics のコンサルタント、倫理学者。著書に`Dangerous Master` (邦訳は『人間 VS テクノロジー』)、Colin Allen との共著に`Moral Machines` がある。
- ピーター・アサーロ (Peter Asaro) : The New School。哲学者。自律型兵器についての論文などがある。
- ライアン・カロ (Ryan Calo) : Washington University。法学者。
- マシアス・シューツ (Matthias Scheutz) : Tufts University。ロボット工学者。近年、アメリカ海軍と共同で道徳的に行動するロボットの開発を推進している。
- ラファエル・カプーロ (Rafael Capurro) : Stuttgart Media University。哲学者。特に情報倫理に関する業績で知られる。

Scuola di Robotica

イタリアで、ロボット工学者のジャンマルコ・ヴェルジオ (Gianmarco Veruggio)、哲学

者の Fiorella Operto (現会長) によって設立された NPO¹³。学生、教師、公衆に対してロボット工学の知識を普及させることを目的としている。

2.2.3. アカデミック

Future of Humanity Institute

オックスフォード大学に設立された異分野融合研究センター。人工知能とマクロ戦略を二つの主要なテーマに挙げている。

AI に関しては、発展する AI の能力についての理解を得ること、技術的な研究をすること、付随する問題をコントロールすること、先進的な AI の政策に対する含意を研究することが目的として挙げられている。

FHI には下部組織として Strategic Artificial Intelligence Research Centre が置かれている。これはオックスフォード大学とケンブリッジ大学のジョイント・イニシアティブである。AI が安全で有益でありつづけるための戦略とツールを開発することを目的とする。

Robo-Philosophy

Aarhus University で 2014 年、2016 年に開催された学会。ロボットに関する哲学的な問題に取り組んでいる。オーガナイザーの Johanna Seibt は Responsible Robotics のメンバーでもある。2018 年はウィーン大学で開催される。

2.3. 日本の動き

2.3.1. 政府主導

AI ネットワーク化検討会議／AI ネットワーク社会推進検討会議

総務省情報通信政策研究所が開催した会議。「AI ネットワーク化検討会議」は、2016 年 2 月から 6 月まで開催され、「2040 年代を見据え、AI ネットワーク化の進展を通じて目指す

¹³ <http://www.scuoladirobotica.it/en/homesdr.html>

べき社会像や基本理念の整理」、「AI ネットワーク化が社会・経済にもたらす影響及びリスクの基礎的な評価や検討すべき課題の整理」を進めた¹⁴。

同年6月には報告書「AI ネットワーク化の影響とリスクー智連社会 (WINS ウインズ) の実現に向けた課題ー」¹⁵が提出された。

2016年10月には「AI ネットワーク化検討会議」を改組して「AI ネットワーク社会推進会議」が設立された。「AI 開発ガイドライン」の策定を目指すとともに、AI ネットワーク化が社会や経済にもたらす影響とリスクの検討を進めた。

2017年3月13-14日には国内外の研究者を集めて「AI ネットワーク社会推進フォーラム」を開催した。

2017年7月には報告書が提出された¹⁶。

人工知能と人間社会に関する懇談会

内閣府が主催。2016年5月から開催されている。メンバーには原山優子（総合科学技術・イノベーション会議 議員）、新井紀子（国立情報学研究所 教授・社会共有知研究センター長）、江間有沙（東京大学教養教育高度化機構 特任講師）、大内伸哉（神戸大学大学院法学研究科 教授）、新保史生（慶應義塾大学総合政策学部 教授）、鈴木晶子（京都大学大学院教育学研究科 教授）、西川徹（株式会社 Preferred Networks 代表取締役社長・最高経営責任者）、橋本和夫（早稲田大学研究戦略センター 教授）、林いづみ（桜坂法律事務所 弁護士）、松尾豊（東京大学大学院工学系研究科 特任准教授）、柳川範之（東京大学大学院経済学研究科 教授）、若田部昌澄（早稲田大学政治経済学術院 教授）がいる。

2017年3月24日には報告書が発表された¹⁷。

¹⁴ http://www.soumu.go.jp/main_sosiki/kenkyu/iict/

¹⁵ http://www.soumu.go.jp/menu_news/s-news/01iicp01_02000050.html

¹⁶ http://www.soumu.go.jp/menu_news/s-news/01iicp01_02000067.html

¹⁷ http://www8.cao.go.jp/cstp/tyousakai/ai/summary/aisociety_jp.pdf

2.3.2. アカデミック

人と情報のエコシステム

2016年から設立されたJST/RISTEXの研究開発領域。人工知能、ロボット、IoTなどの情報技術の急速な発展の社会的影響を考え、人間と情報技術の「なじみ」のとれた社会を実現するための、分野横断的な共同研究のプラットフォームの構築を目指す。領域総括は国領二郎（慶應義塾大学総合政策学部教授）。

ロボット法研究会

情報ネットワーク法学会の分科会。「人とロボットが共生する社会を実現するための制度上の課題」を研究している。新保史生（慶應義塾大学）、赤坂亮太（慶應義塾大学）らが中心となっている。

AIR：人工知能が浸透する社会を考える

科学技術社会論研究者、社会学者、人工知能研究者、哲学者、倫理学者などからなる、人工知能の社会的影響について考える研究会。2014年に設立。江間有沙（東京大学、STS）、秋谷直矩（山口大学、社会学）、大澤博隆（筑波大学、人工知能）、服部宏充（立命館大学、人工知能）らが中心メンバー。

ロボットの応用哲学研究会

哲学者、倫理学者からなる研究会。ロボットや人工知能に関する哲学的・倫理的問題について研究している。2010年に設立。メンバーに神崎宣次（南山大学、倫理学）、本田康二郎（金沢医科大学、哲学・倫理学）、佐々木拓（金沢大学、倫理学）、久木田水生（名古屋大学、哲学）、西條玲奈（北海道大学、哲学・倫理学）、大家真也（神戸大学、哲学・倫理学）、岡本慎平（尾道市立大学、哲学・倫理学）らがいる。

AI 社会論研究会

人工知能が社会に与える影響について考える研究会。哲学 (Humanity)、経済学 (Economics)、法学 (Law)、政治学 (Politics)、社会学 (Sociology) といった多様な観点からのアプローチを目指し、これらを「HELPS」と総称している。井上智弘 (駒澤大学、経済学)、高橋恒一 (理化学研究所、人工知能) が共同発起人。

『人工知能』32 巻 5 号 (2017 年 9 月) で「AI 社会論特集」の企画を組んでいる。

人工知能学会倫理委員会

人工知能学会の倫理委員会。2017 年 3 月に人工知能研究者のための「倫理指針」を公表した。倫理指針は全部で 9 条からなり、「人類への貢献」、「法規制の遵守」、「他者のプライバシーの尊重」、「公正性」、「安全性」、「誠実な振る舞い」、「社会に対する責任」、「社会との対話と自己研鑽」、「人工知能への倫理遵守の要請」について定めている。

2.4. コメント

アメリカやヨーロッパでは企業や政府による人工知能への巨額の投資が行われている一方で、人工知能の危険性に関しての懸念も大きく、政府、アカデミア、NGO などでもかなり大規模に対策が取られ初めている。一方、日本では人工知能への投資は増えてきたが、まだアメリカには及ばない。また安全面については政府の主導による議論が中心となっており、アカデミア、NGO ではそれほど大規模なものはない。

アメリカやヨーロッパでは人工知能が危険だという認識が現実的なものであるように思う。そのことは次節で紹介する文献 (例えば Bostrom, *Superintelligence* やバラッド『人工知能——人類最悪にして最後の発明』) にも見て取れる。

3. 文献紹介

この節では、人工知能やロボット、あるいは一般的に先端的な情報技術に関連する倫理的・法的・社会的問題を取り扱った近年の文献を紹介する。

Wendell Wallach and Colin Allen, *Moral Machines: Teaching Robots Right from*

Wrong, Oxford University Press, 2009

人口道徳的行為者 (artificial moral agents) を作るべきであるという提言、およびそのためのアプローチの提案。著者たちは倫理学者・哲学者であり、具体的な工学の話ではなく、倫理学の観点からどのような倫理理論が工学的な実装に向いているかという話をする。著者たちが推薦するのは義務論や帰結主義ではなく徳倫理、また固定された規則を与えるトップダウンのアプローチと学習に基づくボトムアップのアプローチのハイブリッド・アプローチである。

David J. Gunkel, *The Machine Question: Critical Perspective on AI, Robots, and Ethics*, The MIT Press, 2012

「動物の問題 The Animal Question」は動物倫理の文脈で使われる言葉で、動物の道徳的な立場をどのように考えるべきかという問題。著者はそれと類比的な、「機械の問題 (The Machine Question)」について考えることを提案する。明確な答えはない。答えを与えることではなく問題を与えることが哲学にとって重要なのだ、と著者は主張する。

西洋の伝統的な倫理学・道徳哲学の文脈では、道徳は「道徳的行為者 moral agent」と「道徳的被行為者 moral patient」(道徳的行為の影響を被る対象、道徳的配慮を必要とする対象) という二項構造の枠組みにおいて考えられてきた。そしてもっとも標準的な理論においては道徳的行為者たりうるのは意識や理性、自由意志を持ち、それを十分に行使できる存在だけとされ、また道徳的被行為者となる資格をもつのは道徳的行為者のみとされた。本書の著者はその構図に疑問を提示する。第一章では機械が道徳的行為者となる可能性を、人工知能(の哲学)の歴史と現状を丁寧に解説しながら探っていく。第二章では機械が道徳的被行為者となる可能性を、動物倫理・環境倫理などの動向を参照しながら探る。第三章では道徳的行為者-道徳的被行為者という二項構造に対して疑問を投げかけ、そのような構図を「脱構築」することを提案する。

Sherry Turkle, *Reclaiming Conversation: The Power of Talk in the Digital Age*, The MIT Press, 2015 (邦訳: 日暮雅通訳『一緒にいてもスマホ —SNS と FTF—』、青土社、2017)

著者はかつて *Second Self* (1984年、邦訳は『インティメイト・マシーン』) という著作でコンピューターの持つ大きな可能性を称揚したが、前作 *Alone Together* および本作では、

新しい情報技術のマイナスの側面にフォーカスしている。端的に言うと、情報技術は私たちの関心を目の前の人や状況よりも、コンピューターやタブレット、スマートフォンなどのデバイスへと向けさせる。スマートフォンの普及によって私たちは面と向かったリアルタイムの対話よりも、デバイスを介したテキストベースのコミュニケーションを選ぶようにさせている。そしてそのことは他者に共感する能力の発達を阻害し、人間関係に重大な影響を与える可能性がある、と著者は警告する。

スマートフォンなどのデバイスは、他者と向き合う能力のみならず、自分自身と向き合う能力にも影響を与える。私たちは退屈や時間を持て余すことを厭うため、手持ち無沙汰になるとすぐにスマートフォンに手を伸ばし、退屈しのぎのためのアプリを起動する。しかしながら私たちには、外側からの刺激によって注意をそらされることなしにじっくりと自分の内側の声を聞くための時間が必要なのだ。そのような時間は自分について知ることを可能にし、そして他者と向き合うための基盤を形成するとともに、創造性の源泉ともなる。

本書は心理学や社会学などの実証的な研究を豊富に参照しているだけでなく、著者自身が行った膨大なインタビューに基づいて、現代の人々がどのように情報技術を使い、そしてそれによって影響を受けているかを報告している。

Ezio Di Nucci and Filippo Santoni de Sio (eds.), *Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on Remotely Controlled Weapons*, Routledge, 2016

遠隔操作される兵器、いわゆる「ドローン」に関する「責任」に関する論考を集めた論文集。全体で4部からなり、第一部が「ドローンと法的責任」、第二部が「国家の責任とドローンの使用」、第三部が「設計と社会技術的観点」、第四部が「ドローンと道徳的責任」となっている。

遠隔操作されるドローンに関して、あるいはより進んだ自律型兵器に関しては、それが民間人を殺傷するなどの戦争犯罪に関わったときに誰も責任をとれなくなるという批判が、早い段階から提出されていた。そしてこの批判はいまでもドローンや自律型兵器に関する議論の中心に置かれている。しかしドローンと責任に関する議論はより広いパースペクティブから語られるべきだ、と本書の編者たちは主張する。具体的には、ドローンの利用はそれを使用する国家の責任、テクノロジーを推進させる社会の責任という観点まで含まなければいけない、と編者たちは言う。

そのような関心のもとで、本書には法学、政治、社会学、社会技術論、哲学といった多様な分野の専門家がそれぞれの観点からドローンと責任について論考を展開している。

Patrick Lin, Keith Abney and George A. Bekey (eds.), *Robot Ethics: The Ethical and Social Implications of Robotics*, The MIT Press, 2011

現時点で「ロボット倫理学」について最も包括的な（というよりむしろほぼ唯一の）論文集。全体として8部に分かれており、第一部が「序論」、第二部が「設計とプログラミング」、第三部が「軍事」、第四部が「法」、第五部が「心理と性」、第六部が「医療とケア」、第七部が「権利と倫理」、第八部が「エピローグ」になっている。

Michael Anderson and Susan L. Anderson (eds.), *Machine Ethics*, Cambridge University Press, 2011

編者たちが提唱する「機械倫理 (machine ethics)」とは、倫理的に判断し行為する能力を機械に実装しようという試み。マイケル・アンダーソンとスーザン・アンダーソンは専門家が実際に直面する倫理的なジレンマ状況において従っているルールを、「帰納論理プログラミング」と呼ばれる機械学習の手法によって学習させ、そのルールに従って新しい状況に対処できる人工知能を開発しようとしている。

本書においてはそのほかにも様々な機械倫理へのアプローチ、並びに機械倫理の原理的・倫理的なチャレンジについての論文が収められている。

Luciano Floridi, *The 4th Revolution: How the Infosphere Is Shaping Human Reality*, Oxford University Press, 2014 (邦訳: 春木良且・犬束敦史監訳, 先端社会科学技術研究所訳『第4の革命—情報圏(インフォスフィア)が現実をつくりかえる』、新曜社、2017)

著者は現在の情報技術がもたらす人間の自己意識における変革を、「第4の革命」と特徴づける。それはコペルニクスの地動説、ダーウィンの進化論、フロイトの精神分析学について、「人間とは何か」という問いに対する答えを大きく変容させる。

情報技術が発展して、人工物が単に人間と世界の間での媒介ではなく、人工物同士が自

律的にインタラクションをするようになったとき、人間と世界は大きな変貌を遂げる。人工物の中を情報が駆け巡り、人間もまたその中に取り込まれるとき、私たちにとっての現実 reality は、物理的な世界とバーチャルな世界を包含した「infosphere 情報圏」と化す。情報圏においては人間はロボットや人工知能と同じ、情報的存在者となるのである。

情報圏が豊かに繁栄するためには、その中でいかに「摩擦なく」情報を流通させるかが鍵である。ルチアーノ・フロリディはそのためにプライバシーや自己といった、人間存在の本質にかかわる既存の概念を大胆に変容させようとする。しかしながら彼はそれをできるだけ衝突の少ない形で実行しようとしているように思える。そのあたりが非常に戦略的に感じられる。

Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, 2014

人間をはるかに凌駕する超知性について、それがいかにして生まれるか、いかなる形をとりうるか、そしてそれが人類にとってどれほどの脅威になりうるかを細かく論じている。著者によれば、超知性が誕生すれば、それは世界の方向性を決定する単一の意思決定システム（これを著者は「シングルトン」と呼ぶ）になる見込みが高い。そして一つの超知性の持つ最終的な目的は人間や他の知性が価値を見出す目的とは全く異なるもの（たとえば円周率を計算する、紙クリップをできるだけたくさん生産するなど）になる見込みが高い一方で、最終的な目的を達成するための道具的な目的はたとえば認知能力の向上や資源の獲得などのようなものに収束するだろう。こういったことから超知性の誕生は人類にとって生存のリスクになりうる、と著者は予言する。

また著者は超知性が人類の生存を脅かすことがないようにコントロールするためにどのような手段を取りうるかについても細かく考察をしている。コントロールの仕方としては、能力を制限する方法と、動機を選択的に与える方法の二種類がていあんされるが、しかしそれぞれに困難があることが論じられている。

ジェイムズ・バラッド著、水谷淳訳『人工知能一人類最悪にして最後の発明』、ダイヤモンド社、2016

著者はドキュメンタリーフィルムの製作者。人工知能にかかわっている多くの人々に取材して、人工知能の危険性を訴えている。本書で前提としているのは、人間を超える人工知能が実現するということであり、そのような知的存在は人間とは相いれず、やがて人間を滅

ぼすだろう、というのが主要な論点である。

西垣通『ビッグデータと人工知能』、文春新書、2016

人工知能の基礎、原理を簡潔に解説し、「超知能」への信奉や、それに付随する楽観論や悲観論が間違いであると断じ、「人工知能ブーム」の過熱に待ったをかける。機械と違って、生命は「自己を作る」、「オートポイエティック」なシステムであり、人間によって作られ、人間の指令に従うだけの機械に人間と同様の知性を持つことはできない、というのがその主要な理由である。機械にできるのは人間の知能の作用の一部を模倣することであり、機械が感情や目的を持つこと、記号の意味を理解すること、状況において適切な判断をすることはできない、と著者は言う。

人工知能が人間を超える知能を持つという幻想の根源は、人間も単なる物理的な機械の一種と考える誤謬であり、そしてその根源には西洋文明に特有の世界観、人間観があると主張する。

久木田水生・神崎宣次・佐々木拓『ロボットからの倫理学入門』、名古屋大学出版会、2017

二部からなり、一部はロボットに「道徳性」を持たせることができるかという問題についての考察を通じて、人間の道徳性について従来とは異なる角度から光を当てている。

二部はロボットや人工知能が社会に実装されるにつれて実際に生じてくる新しい問題、あるいはより深刻化する既存の問題に焦点を当てる。二部で取り上げられているのは「プライバシー」、「ソーシャルロボット」、「軍事」、「労働」である。

稲葉振一郎『宇宙倫理学入門—人工知能はスペース・コロニーの夢を見るか?』、ナカニシヤ出版、2016

有人宇宙飛行、あるいは宇宙植民がテーマであり、人工知能やロボットが中心的なテーマではないが、ロボット倫理学についても一章を割いて論じている。

生身の人間が恒星間飛行をするのは困難であるため、「有人」恒星間飛行を実行するためには、「人間」の定義を変え、人間並みの知性を備えた自律的かつ人格的ロボットを宇宙に

送り出すことが合理的であるという議論に続いて、そのようなロボットを作ることや、それを宇宙へのミッションに送り出すことの倫理的な含意について論じている。

また「補論」においてはニック・ポストロムの『*Superintelligence: Paths, Dangers, Strategies*』についての短い論考を収録している。そこでは超知能による「シングルトン」の形成に対するポストロムの危惧が、社会やコミュニケーションの喪失、それによるアーレント的な意味での「人間」の消失にあるのという可能性が考察されている。また宇宙への拡散する自己増殖ロボットに関しては、それは物理的限界からシングルトンと切り離されたコミュニティーを形成することになると論じ、ポストロムの予想を論駁している。

ウェンデル・ウォラック著、大槻敦子訳『人間 VS テクノロジー：人は先端科学の暴走を止められるのか』、原書房、2016

テクノロジーのリスクにどのように対処するべきかを、様々な先端テクノロジーの事例を取り上げながら論じている。テクノロジーの発展の道筋は変えられないという悲観論を退け、人類の未来に大きな影響を与えることができる「変曲点」と呼ぶべき時機があり、そしてサイボーグ技術、遺伝子工学、人工知能とロボットなどのテクノロジーに関して、現在が重要な変曲点であると著者は言う。

人工知能とロボットに関して一章を費やして論じており、そこでは自律型兵器、自動走行車、ソーシャルロボットなどが取り上げられている。

ケヴィン・ケリー著、服部桂訳『＜インターネット＞の次に来るもの——未来を決める12の法則』、NHK出版、2016

原題は *The Inevitables* (不可避なもの)。コンピューター、インターネットといった情報技術の持つ固有のバイアスによって定められる技術発展のトレンドを見定め、そこからこの先のテクノロジーの行方を予測する。著者はこの先、物理的な「プロダクト」に代わって、手に触れることのできないサービスとプロセスが生み出されるものの主流になっていくだろうと予測する。

キャス・サンステイーン著、伊達尚美訳『選択しないという選択——ビッグデータで変わ

る「自由」のかたち』、勁草書房、2017

個人の自由な選択を尊重しつつ、社会のシステムの中に個人がよりよい選択を自動的にできるようなデフォルトを仕込むことを提唱する。自由や自律を否定するのではなく、個人が選択を要求されるあらゆる局面において、熟慮の末に選択をすることは不可能なのだから、せめてデフォルトとしてより良い選択を用意しておく、そしてその際に現在の情報技術のアドバンテージを活用する、という点がポイントである。

ピーター＝ポール・フェルベーク著、鈴木俊洋訳『技術の道德化』、法政大学出版局、2016

テクノロジーは人間の認識や行為の媒介であり、人がどのように世界を認識し、どのように世界に対して働きかけるかに影響を与える。それゆえに私たちは意図的に人の良い行為を促すようにテクノロジーをデザインすることができるし、そうするべきであると著者は主張する。

Luís Moniz Pereira and Ari Saptawijaya, *Programming Machine Ethics*, Springer, 2016

「機械倫理」における著者たちの研究の報告。個人的観点と集団的観点の両方を取り扱っているところが特徴的である。

個人的な観点からは道徳的認知のダイナミクスをモデリングするために、論理プログラミングの枠組みを使う。倫理学における「ダブルエフェクト」理論、意思決定の「二重過程」理論、「反事実的推論」などを考察している。

Jai Galloiett, *Military Robots: Mapping the Moral Landscape*, Ashgate, 2015

軍事ロボットについての詳細な論考。軍事ロボットを端的に良い、悪いと論じるのではなく、どのような条件であれば受け入れ可能であるかを、リスク分析や正戦論の枠組みの中で論じている。またよく議論される自律型兵器に関する責任の分配問題についても論じられている。無人機は戦争犯罪の責任の帰属を不可能にするから、倫理的に悪いという Sparrow などの議論に対して、筆者は人間及び非人間の行為者の間で分配される責任、という新しい責任概念を提案する。

ジョン・マルコフ『人工知能は敵か味方か』、日経 BP 社、2016

コンピューター、情報技術の発展を、AI (Artificial Intelligence) と IA (Intelligence Augmentation) という二つの大局的なパラダイムの枠組みの中で考察し、どちらがより有用なテクノロジーを生むだろうかと著者は問いかける。AI は「人間を機械によって置き換える」という思想に基づく。それはジョン・マッカーシーやマーヴィン・ミンスキーが提唱したコンピューターの未来である。一方で IA は「人間の能力を機械によって拡張する」、「人間と機械が協働する」という思想に基づく。このアプローチを提唱したのはダグラス・エンゲルバートやテリー・ウィノグラードである。

著者は人工知能によって人間を置き換えるのではなく、あくまでも人間の同僚として人間の能力を拡張するために人工知能を使うことを提唱する。

井上智洋『人工知能と経済の未来——2030 年雇用大崩壊』、文春文庫、2016

一部の人が主張するように人工知能が人間を超えるような知能を獲得するとすれば、人工知能は人間の仕事を奪うだろう。そのとき経済に何が起こるのか、そしてそのとき私たちはどのような経済システムを選べばよいのか。おそらく人工知能は経済をこれまでないほどの規模とスピードで成長させる。その一方で社会は多くの人々の失業という問題に直面しなければならない。そこで著者が提案するのが、すべての国民に一定の収入を無条件で付与するという「ベーシック・インカム」の制度である。人工知能とベーシック・インカムによって人々はお金を稼ぐための労働から解放され、より純粋に楽しみのために時間を過ごすことが可能になる、と著者は述べる。

エリック・ブリニョルフソン、アンドリュー・マカフィー著、村井章子訳『機械との競争』、日経 BP 社、2013

人工知能・ロボットの経済に対する影響についての論考。アメリカはリーマンショックから立ち直って、設備投資が回復したにも関わらず失業率は低下しなかった。著者たちはその原因が IT の発展にあると考える。企業は高額な情報技術に投資したにも関わらず、労働者を雇うことをしなくなった。IT による失業は単調な頭脳労働を行うオフィスワーカーにとって最も脅威になるため、この傾向は高所得者と低所得者の間のギャップをより鮮明にす

ることになるだろうと予想する。重要なのは IT を利用して創造的な仕事をできるようにすることで、そのカギは教育であると著者たちはいう。

エリック・プリニョルフソン、アンドリュー・マカフィー著、村井章子訳『ザ・セカンド・マシン・エイジ』、日経 BP 社、2015

『機械との競争』の著者たちによる、同じテーマについてのより詳細な論考。一般に大きな技術革新が起こると、それによって人間の労働力が不要になり失業が増えるという主張と、技術革新は商品の価格を下げ、需要を増やすために労働力は不要にならないという主張が対立する。「労働塊」すなわち世の中の仕事の量は一定しかないという考えは、「労働塊の誤謬」と呼ばれ、主流派経済学者たちからは否定される。なぜなら人間の欲望はいくらでも増大させられるからである。人間の欲望が増大する限り、完全雇用は実現可能である、と彼らは考える。

情報技術の発展に際しても同じことがいえるだろうか、と著者たちは疑問を投げかける。情報技術の発展は労働のありとあらゆる場面を自動化してしまい、いずれは人間の労働を全く不要にしてしまうかもしれない、そのとき人間がどれだけ欲望を増大させたとしてもやはり失業は避けられない。例えば人間の移動に対する需要は 20 世紀の間に圧倒的に増大したが、しかし馬の労働力は完全に機械に置き換わった。そして労働力として存在する馬の頭数は激減した。同じことが情報技術と人間の間にかかる可能性は十分にあると著者たちはいう。そしてその時、社会は格差と貧困によって苦しめられるだろう。

一方でしかし人間と馬には重要な違いがある。それは馬が自分たちの将来を主体的に選択できなかったのに対して、人間は民主的な政治システムを通じて自分たちが生きる社会の在り方を決定することができる、ということである。著者たちはここで「セカンド・マシン・エイジ」と彼らが呼ぶ時代において、その悪影響を緩和させるために政治がやるべきことを具体的に考察し、提言している。

セカンド・マシン・エイジにおいても人間の労働力が一気にすべて機械に置き換わるわけではない。人間の労働力がすっかり（あるいはほとんど）置き換わるまでにはまだ何十年もかかるだろう。したがって現在の失業対策として有効な手段はセカンド・マシン・エイジにおいても当分は有効である。具体的には基本的に経済成長を促すのが正しい政策だということである。そのために著者たちは以下のような提言をする・

- 初等・中等教育を改善する：テクノロジーが急速に発展するとき、教育がテクノロ

ジーに追いつかないと、格差が拡大する。また初等・中等教育と経済成長には密接な関係があることも分かっている。したがって教育の改善は重要である。教育におけるデジタル技術の利用や、教師の報酬を高く、責任を大きくすることは有効だろう。

● 起業環境を整備する：新しい挑戦的な企業は、新しい雇用を創出する。したがって起業しやすい環境を作ることは重要である。

● 求人と求職のマッチングを強化する：より良いマッチング方法があれば、求人側と求職側の摩擦やすれ違いが減り、就職活動に伴う不必要な苦勞の負担も減る。求人と求職を結びつけるデータベースが地域、国、グローバルなレベルで整備されれば、双方に有益である。

● 科学者を支援する：基礎研究は極めて大きな正の外部性をもたらす。2005年からアメリカは基礎研究への予算を減らしているが、これを逆転させて一層の支援を続けるべきだ。また知財制度、ソフトウェア特許と著作権の期間にも改革が必要。知財の保護はイノベーションの動機でもあるが、過度の保護は新しいイノベーションを阻害する。バランスが重要。

● インフラを整える：教育・研究と同じくインフラへの投資も正の外部性をもたらす。インフラへの投資が景気を浮揚させるかどうかについては議論があるが、良いインフラは暮らしやすい環境、事業経営のしやすい生産的な環境を作ることにつながる。移民を積極的に受け入れることも経済に好影響を与える。移民は本国の労働者の仕事を奪うという考えもあるが、実際には移民を受け入れても本国の失業は増えなかった例がある。特に教育水準の高い移民は、起業を活発化して雇用の創出につながる。

● 賢く課税する：公害に課税することは、環境の汚染を食い止め、クリーンな技術への投資を促し、政府の収入を増やす。セカンド・マシン・エイジの技術はこのような課税の実効性を高めることができる。また渋滞を緩和するために情報技術を活用して、柔軟な通行料を徴収することができる。課税に対して日弾力的な財、例えば土地などには高額な税をかけるとよい。また高額報酬を得ている人からはより多くの税金を徴収すればよい。最高税率と総合的な経済成長の間にはほとんど相関関係は見られないことが知られている。

以上はセカンド・マシン・エイジが雇用にそれほど劇的な影響を与えていない段階での政策提言であるが、長期的にはまた異なる政策が必要だろう。機械が人間の労働をほとんど代替できるようになった時には私たちは何をすべきなのか。著者たちは次のような提言を行う。

- テクノロジーの進歩を阻むような政策は望ましくない。
- 資本主義に代わる体制を探すことも望ましくない。
- ベーシック・インカムは望ましくない。なぜなら労働は単にお金を稼ぐ手段という

だけではなく、もっと価値のある多くのものを手にする手段だからである。それはプライドや自信、仲間、情熱を注ぐ対象、健全な価値観、地位や尊厳といったものである。

● 負の所得税は望ましい。負の所得税とは税金がかからない所得金額を境に、それを上回った人からは上回った金額に応じて税金を徴収し、それを下回った人には下回った金額に応じて税金を支払う、という仕組みである。この制度のメリットは最低所得を補償すると同時に就労のインセンティブを与えるということである。また既存の納税・還付税の制度を利用できるので運用しやすいというメリットもある。

ジェリー・カプラン著、安原和見訳『人間さまお断り』、三省堂、2016

人工知能の発展によって起こる問題、特に労働の問題に対して、類書の多くが問題の指摘に終わっているのに対して、本書は非常に具体的で現実的ないくつかの提案をしている。

本書はまず現在のテクノロジーについての概説から始まる。著者はロボットの頭脳となる部分、一般に人工知能と呼ばれるものを含む部分を「合成頭脳」と呼ぶ。また機械の身体を「労働機械」と呼ぶ。これらを制御するテクノロジーの進歩が、社会に重大な変化を急速に与えつつあるのが現在である。

従来のロボットは決まった動きしかできないので人間がロボットが働くための環境を整え、機械の邪魔をしないように気を付けなければならなかった。しかし現在のロボットは人間の邪魔にならないように、人間に危害を与えないように、それでいて与えられた仕事を適切にこなすことができるようになってきている。このことを可能にしたのは次の4つの技術的進歩のおかげである。すなわち、コンピュータの処理能力が大幅に向上したこと、機械学習技術が進歩したこと、産業用ロボットの制御技術や素材が進歩したこと、そして何より機械知覚の性能が向上したこと。これらの技術の進歩によって、やがて人間が手作業で行う仕事の多くを機械が遂行できるようになるだろうと著者は予言する。また著者はこの発展の生み出すロボットは従来のようなロボットの概念には当てはまらないものになるという。というのも私たちが思い描くロボットはC3P0のような一か所の身体に機能を集めた個体である一方、おそらく現実のロボットは様々な機能をあちこちに分散させたものになる可能性が高い。また著者はこれからのロボットは様々な機能を備えた汎用的なものになっていくだろうと予想する。

人工知能が重大な負の影響を与えた現実の事件として、2010年の「瞬間暴落（フラッシュ・クラッシュ）」がある。わずか数分のうちに株価が9パーセント下落するという事件が起こった。この株価の下落は金額にすると1兆ドルを超す資産価値に相当する。幸いこの時

は、しばらくして株価は急落の前の状態に復帰した。しかし同じようなことが再び起きないという保証はない。この原因は株を売買するコンピュータ・プログラムがコンフリクトを起こしたことだった。現在の株式市場においては一秒間に 100 万回の取引を行うことができるコンピュータが用いられている。このような取引は超高速取引 (high-frequency trading; HFT) と呼ばれている。著者はこのような取引で儲けを得ることは、通りにロボットを歩かせて人が落としたお金を素早く拾わせるようなものという。そしてそのようなもののおかげで市場が大きなリスクを抱えなければいけないことを問題視している。これに対して著者は二つの対策を提案する。一つは情報のリクエスト (「気配値」¹⁸のリクエスト) に対してわずかな額を課金するという方法である。もう一つはあらゆる取引 (人間によるものもコンピュータによるものも) に対して、一秒遅延させるという方法である。この遅延は人間による取引においてはそれほど大きなリスクをもたらさないが、HFT を行うコンピュータにとっては大きなリスクに膨れ上がる。またこの方法は待ち時間を短縮するための熾烈な「軍拡競争」に対する歯止めになるかもしれない。

ロボットの責任の問題についてもかなり具体的な提案をしているのも本書の特徴の一つである。著者はロボットが自律的に行動するようになり、また学習によって適切な行動を覚えるようになるだろうということを予想し、その場合には法人のように責任の主体になりうるだろうと論じている。ロボットが人間に危害を与えたときには、ロボットがその「目的」を遂行するのが困難になるような「罰」を与えればよい、そしてまたロボットが「財産」を持つことも認めると良い、と著者は提案する。

著者が最も懸念しているのは労働の問題である。テクノロジーの進歩は人間の仕事を奪う。長期的には人間はこの変化に対処できるだろう。しかし問題はその変化のスピードによって引き起こされる悲劇である。失業の問題は避けられない。また新しいテクノロジーが新しい仕事を創出したとしても、多くの人はその仕事をするのに必要な知識や技術を身に着けることができない。そこで教育と訓練がカギになる。重要なのは社会人になってからも新しい技術や知識が必要になったら繰り返し教育を受けることが可能でなければならない。しかしそのための費用は誰が持つのか。ここで著者が提案するのが、「職業訓練ローン」あるいは「就活ローン」と著者が呼ぶ制度である。これは次のような制度である。まずローンに対して「見込み雇用主」という保証人がいる。雇用主は申込者を雇うことを約束するわけではないが、その技能をもつ労働者を必要としているということを書き添えるものになる。したがって一人の雇用主が保証人になれる申込者の数は、その雇用主が実際に雇う気がある労働者の人数に制限される。教育を受けた後、保証人が申込者を雇った場合には税の優遇が受けられる。このローンの返済は賃金のみからなされる。

¹⁸ 気配値とは、買い方が買いたいと希望する値段のうち最も高いもの (買い気配値)、あるいは売り方が売りたいと希望する最も安いもの (売り気配値) のこと。

格差の問題に対する著者の提言は非常にユニークである。彼はまず企業などがどれだけ公共的な性格を持っているかを示す「公益指数 (public benefit index; PBI)」という概念を考案する。たとえばその株がどれだけの人にどれだけ均等に所有されているかということによって PBI を定義することができる。ここではジニ係数のような計算を用いることができる。そして PBI の高い企業には減税をするというのである。そのことによって企業はできるだけたくさんの人に株を持ってもらうように努力をするだろう。そして人々はその株の配当を受けることができる。この仕組みがうまくいけば格差は縮小するはずだという。

4. 人工知能の何が倫理的に問題なのか

4.1. 人工知能とそのネットワーク化に特有の倫理的問題はあるのか

本節では人工知能の倫理的問題について簡単にまとめる。初めに倫理的問題の射程について説明しておく。倫理が問題にするのは「価値」である。しかしそれは経済学や工学が扱う価値（経済的繁栄、効率など）には限られない。倫理学が中心的に扱うのはしばしばそれらよりも捉えどころのない、かといって無視するわけにもいかない普遍性を持った価値である。典型的なのは幸福、快と苦、公平さ、正義、道徳、真正さ、尊厳、自律性、権利、平等、共同体、伝統などである。

従って人工知能の倫理的問題とは、これらの価値が人工知能によって侵される危険性に関連している。本報告の中で紹介された倫理指針や倫理原則などもこういったものに関わっている。論点は様々あるが、一般化して要約すると倫理的課題とは「人類全体の繁栄、幸福を促進しつつ、犠牲をできるだけ少なくする」ということである。

しかしこのように一般化すれば、あらゆる科学技術について同じことが言える。人工知能において特殊な論点は存在するのだろうか？ 人工知能、そのネットワーク化に特殊な性質、それに特有な倫理的問題は何か？

Jerry Kaplan は *Artificial Intelligence: What Everyone Needs to Know* の中で、人工知能とは何かを論じる際、シンプルにそれは「絶えまない自動化の進歩」なのだと主張している。産業革命は単純な物理的労働を自動化した。コンピューター革命は単純な知的労働を自動化した。現在の人工知能革命、ロボット革命、IoT 革命においてはますます複雑な物理的労働、知的労働を自動化することに成功している。しかしこの新しい技術革命が意味することは単に遂行できるタスクの複雑化・高度化だけではない。

一つにはパーソナライズされた選択の自動化が挙げられる。「機械的な作業」ということには、決まったルールに従っているということ、それゆえ誰がやっても同じ結果になるということが含まれている。しかし現在、自動化が進行しているのはそのようないわゆる「機械的な作業」ばかりではない。人工知能のネットワーク化は、私たちのあらゆる行動からデータを採集し、そのことは個人にカスタマイズされた選択や意思決定を機械がサポートすることを可能にする。しかしながら技術には（意図されたものであれ、意図的ではないものであれ）一定のバイアスがある。そのようなバイアスが不適切なもの（例えば不当な差別を反映したようなもの）である場合には問題であることももちろんだが、そうでない場合でも、バイアスの存在そのものが個人の意思決定の自律性という価値を脅かすものでありうるということには注意が必要である。ゲーテはウェルテルに「僕の知識なんか誰にでも持てるものだ。僕の心、これだけが僕にしか持てないものなのだ」と語らせているが、このような個性の核をなす心というものが機械という外部からの介入者の持つバイアスに操られること自体が、近代以降の社会が重要視してきた価値に大きな影響を与えるものであることは強調されるべきだろう。

上の論点と関連することとして、第二に行為や意思決定の責任の問題がある。社会心理学者の小坂井敏明が『責任という虚構』の中で指摘するように、今日私たちが持っている「責任」の概念は近代以降の「自律的な個人」という概念と密接に結びついている。しかし上での述べたように現在、機械化・自動化の波は個人の意思決定にも大きな影響を与えるようになってきている。その中には単なる好みの問題というだけではない、道徳的な意思決定も含まれている。前節で紹介したサンスティーン『選択しないという選択』やフェルベーク『技術の道徳化』、Anderson and Anderson, *Machine Ethics*、Wallach and Allen, *Moral Machines*などはテクノロジーに道徳的価値を埋め込むことや、人工知能に明示的な道徳的推論、道徳的行為の能力を持たせることを提唱する。このような試みは、責任についての私たちの考えや実践の見直しと並行しなければならない。なぜならば自律性を減じられた個人に対して重い責任を帰すのは道徳的に問題があるからである。

第三に、現在の機械化の波は物理的労働、知的労働から、感情労働へとその射程を広げている。感情労働とは「自らの感情をコントロールすることを要求される労働」のことであり、接客業、教育、ケアなどがその典型である。この波は職業としての感情労働だけでなく、友人や恋人や家族の間での「感情労働」にも広がっていくだろう。おそらくこれは人間同士の感情的なつながりに大きな影響を与えるだろう。

4.2. テクノロジーと倫理に関する間違っただ考え方

以上、人工知能に関連する倫理的問題について述べたが、最後に技術と倫理の関連についてしばしば述べられる間違っただ考え方について注意を促しておこう。

4.2.1. 長い目で見れば・・・

しばしば次のような意見を聞くことがある。「科学技術の発展は未知のリスクや既存の価値観との衝突をもたらすことは確かである。しかしこれまでも科学技術は様々な問題を引き起こしながら、長い目でみれば人類により多くの利益をもたらしてきた。人工知能に何らかのリスクがあるとはいえ、それゆえに将来の利益を放棄することは間違っている」

この考えが正しいとすれば、私たちが現在享受している便利さは、例えば公害病に苦しんだ人々が受けた危害を補償するのだろうか？ もちろん将来世代のことを考える世代間倫理というものもあるが、それは将来の世代が少なくとも私たちと同じレベルの生活を送れるようにということを考えることであって、将来の人たちが私たちより快適便利な生活を送るために、現在の私たちが苦しまなければならないということではない。

4.2.2. テクノロジーは中立だから・・・

「科学やテクノロジーは価値中立的であって、それ自体では善でも悪でもない」というのもまたよく聞く意見である。この見解は全米ライフル協会のスローガン「銃は人を殺さない、人が人を殺す (Guns don't kill people, people kill people)」に顕著に反映されている。しかしこの意見はいくつかの点で間違いである。

第一に科学やテクノロジーそれ自体に固有の価値観がある。例えば科学においては「一般性」、「再現性」、「定量的であること」、「シンプルさ」などが良い科学の規範とされる。テクノロジーにおいてはさらに「効率性」、「制御可能性」、「コストとベネフィットの差し引き」などが規範的価値になっている。科学やテクノロジーを推進することはこういった価値を社会に浸透させる一方で、それに当てはまらない価値（「個別性」、「一回性」、「定量化できないこと」、「単純化できないこと」、「非効率性」、「制御できないこと」、「コストとベネフィットの観点から扱えないこと」など）をマージナライズする。

第二に、ある種のテクノロジー（の産物）は悪用に傾きやすい固有のバイアスを持っている。例えば、麻薬のようなものは、確かに薬としても使えるということは事実だが、圧倒的

に悪用・濫用されやすいバイアスを持っている。そのような個々のテクノロジーの持つ固有のバイアスを考えることはエンジニア、ユーザー、政策決定者の責務である。