

ロボット倫理はどこに向かうのか

平成30年度グローバル社会文化研究センター
研究プロジェクトC 研究成果報告

研究期間：平成30年4月1日～令和2年3月31日

根 村 直 美

寺 沢 幹 雄

March, 2022

ロボット倫理はどこに向かうのか

共同研究者

研究代表者 根村 直美¹

副代表者 寺沢 幹雄²

目 次

序文	根村 直美	1
ロボット倫理はどこに向かうのか —哲学・倫理学の立場から—	根村 直美	5
親近感に基づくロボットの倫理的責任	寺沢 幹雄	21
法的分野における責任の新たな類型 —「厳密な意味での行為者」としてのロボット—	根村 直美	31
Abstract (英文要旨)		38

¹ 日本大学経済学部教授 nemura.naomi@nihon-u.ac.jp

² 日本大学経済学部教授 terasawa.mikio@nihon-u.ac.jp

※日本大学経済学部 〒101-8360 東京都千代田区神田三崎町 1-3-2

The future of Robot Ethics

Member

Principal Investigator	Naomi Nemura ¹
Deputy Principal Investigator	Mikio Terasawa ²

Contents

Preface	Naomi Nemura	1
A Study of Philosophical and Ethical Perspectives on Robot Ethics	Naomi Nemura	5
Ethical Responsibility of Robots Based on Intimacy	Mikio Terasawa	21
A New Type of Responsibility in the Legal Field : Robots as Strict Agents	Naomi Nemura	31
Abstract		38

¹ Professor, Faculty of Economics, Nihon University nemura.naomi@nihon-u.ac.jp

² Professor, Faculty of Economics, Nihon University terasawa.mikio@nihon-u.ac.jp

※Faculty of Economics, Nihon University 1-3-2 Kandamisakicho, Chiyoda-ku, Tokyo 101-8360

序文

Preface

根村 直美

Naomi Nemura

本ワーキングペーパーは、日本大学経済学部グローバル研究センターより助成を受けて行われたプロジェクト研究 (C)「ロボット倫理はどこに向かうのか」から得られた知見を報告するものである。本プロジェクト研究は、研究代表者・根村直美 (日本大学経済学部, 倫理学) と副代表者・寺沢幹雄 (日本大学経済学部, 情報学) により 2018 年度～2019 年度にかけて行われた。さらに、2020 年度にワークショップ (2020 年 12 月 3 日) を開催して議論を深め、このワーキングペーパーの知見へと至っている。

*

応用倫理学の文脈では、2000 年代後半にすでにロボット倫理に関する議論の必要性が意識されはじめていたが、そうした議論はいささか SF 的な思考実験と捉えられてきた (久木田, 2009)。しかしながら、近年、限定的ではあるが自律性をもつロボットが社会の様々な領域で活躍するようになってきている。加えて、シンギュラリティ (技術的特異点) という概念が現実味をもって受けとめられるようになってきている。本研究プロジェクトは、こうした状況を背景として進展しつつあるロボット倫理の議論の現状を明らかにするとともに、今後どのように考えていけばよいのかを主として倫理学、および、情報学の立場から探ることを目的とした。

ロボット倫理を考える際に射程となる問いには、大きく言って 2 つの方向がある (久木田, 2009)。1 つは、ロボットに課されるべき倫理的制約、つまり、「ロボットが守るべき倫理」とは何かという問いを設定することができる。もう 1 つは、人間はロボットに対してどのような道徳的配慮をなすべきなのかを問うことができる。すなわち、「ロボットに対する倫理」とはどのようなものなのかを問題とすることができるのである。そして、「ロボットが守るべき倫理」を問題にするためには、ロボットが「道徳的行為者」となりうることを前提とされる必要がある。その一方で、「ロボットに対する倫理」が問題とされるためには、ロボットが道徳的な配慮を受けなければならない対象である「道徳的受容者」となりうることを前提となる。

倫理学の立場から考察を行った根村論文「ロボット倫理はどこに向かうのか」では、どのような存在が「道徳的行為者」、そして、「道徳的受容者」となりうるのかを考察している (以下、この論文は「根村論文 A」と表記する)。その考察においては、そもそも、人間社会が想定してきた「道徳的行為者」「道徳的受容者」という概念は、先だって存在している何らかの <実体> に拠って立つものではなく、<有機体も非有機体も含む諸々の存在が相互に作用し合う具体的状況において構築されるもの> と考えられることを示している。そして、そうであるとすれば、ロボットを「道徳的行為者」として

見なすべきかという問い、あるいは、「道徳的受容者」として見なすべきかという問いに対する答えは、ロボットが人間社会を含む周囲の環境とどのように関わっているかに依存すると結論づけている。

こうした立場からすると、ロボット倫理を考えるための第一歩としては、今現在、ロボットが人間社会を含む周囲の環境とどのように関わっているかを把握することが必要となる。現状分析に際して、根村論文Aでは、2018年の『情報処理』第59巻8号が、近年のAIブームの中で誕生している製品が人間とのコミュニケーションを重視し家庭や社会など暮らしの中で活躍する製品であることを背景としつつ、人と「コミュニケーションをとる機械」としてのロボットに焦点を当てた特集を組んだことに注目する（真部・細野・今井，2018: 684-685）。また、『ロボット白書2014—社会を変えようとするとき、そこにロボット技術がある！—』が、ロボットの定義は「ロボットの役割や受け取られ方」によって「変遷してきている」と論じていることに目を向ける（平井・水川・林原，2014: 1）。そして、ロボットが人と「コミュニケーションをとる機械」と定義されるようになっているのは、「コミュニケーション」を中核とするようなロボットと人間との関係が生じつつあることを示唆するという認識を提示している。

寺沢論文「親近感に基づくロボットの倫理的責任」は情報学の立場からの考察が行われている。寺沢論文では、ロボットは、「自律的に作動し、人とのインタラクションによって知的に行動する機械」と定義されており、やはりコミュニケーションを取ることを前提とし生活や社会の中で人間と関わりあうロボットに焦点を当てて議論が展開されている。

寺沢論文では、「産業用ロボット」「家庭用ロボット」という一般的なロボットの分類を紹介しつつも、この分類は、人間がロボットを道徳的主体と感ずるかどうかなどは無関係であるとする。そして、倫理的側面が問題になるのはロボットが人間と同等の機能、形態を有する場合であることが多いことに基づき、人間らしい親しみやすさ、すなわち、「親近感」というハードウェアの分類基準を提示している。

産業用ロボットは、組み立て、加工、検査などの生産工程で利用される汎用な自動機械であり、形態的な特徴をもたない。その一方、サービスロボット、ペットロボット、コミュニケーションロボットなどの家庭用ロボットは、知能や形態において親近感をもちうるような特徴をもつ。寺沢論文では、従来の一般的な分類を横断する形で、「外形(形状、大きさ、装飾)」、「利用目的(利用者数、専門性)」、「複雑性(センサ・駆動系の自由度、入出力の関係性)」の観点から親近感の定式化を試みている。

また、寺沢論文では、ロボットの動作を規定する主体の責任についての考察を行っている。その考察では、ロボットの責任主体としては設計者、育成者、利用者に分けられ、動作設定としては、組み込みプログラム、ティーチング、学習に分けられる。そして、組み込みプログラムで稼働するロボットでは設計者の責任が最も重くなり、ティーチングでは育成者の責任が最も重くなること、利用者の責任は、設計者および育成者の責任を除いた補集合となるが、基本的には製造物責任が問われるためその割合は小さいことが示されている。

寺沢論文はさらに、社会の中で知識を学習するロボットが満たすべき行動規範と、行動規範に基づく責任問題について考察している。そして、ロボットがビッグデータによって倫理的知識を獲得する際には、身近さ、権威、時間を要因とする道徳的価値観が重要であることが示されている。また、社会的制約に基づくロボットの行動規範と責任についても論じられており、親近感が大きければ大きいほど、許容範囲の狭い制約となることが示唆されている。

こうした責任をめぐる寺沢論文の議論は、根村論文「法的責任の新たな類型：『厳密な意味での行為者』としてのロボット」に引き継がれている（以下、この論文は根村論文 B と表記する）。根村論文 B では、ロボットの法的責任について新たな類型を導入する必要性を説くロボット法研究の第一人者 Ugo Pagallo⁽¹⁾がロボットの法的責任に関してどのような議論を行っているのかを明らかにしている。

Pagallo の考えでは、ロボットに、憲法上の権利義務を有する法人格を認めるかどうかは論理的というよりもプラグマティックな問題であるが、ロボットに法人格を認めることが今日の法制度の機能の改善につながる可能性は低い。その一方で、Pagallo は、契約法や不法行為など民事法分野においては、ロボットを「厳密な意味での行為者」(Pagallo: 2013)⁽²⁾として認めることは必要だと論じる。Pagallo がそのような<行為者性>を認めることの必要性を説くのは、ロボットが下す判断を人間が予測することの不可能性が増す中、国連が言う「人類の幸福にとって有益なサービス」(Pagallo: 2013, 82)⁽³⁾を提供するロボットの利用または製造を妨げるような立法を回避するためである。

根村論文 B は、Pagallo がそうした<行為者性>を認めようとするのは、ロボットの判断の予測不可能性に加え、我々がすでに個人あるいは家庭内で利用するロボットとの相互作用の中で生活しはじめていることとも深く関連しているという分析を提示する。すなわち、ロボットの<行為者性>を認めるのはプラグマティックな観点からは是認されるアプローチであるが、そのプラグマティックな観点とは、我々との生活において相互に作用し合うようになっているロボットが提供され続けることを目指すものであると結論づけている。

Pagallo の議論の妥当性については、今後、法学者などとの議論を通じて検証されていく必要がある。しかし、Pagallo の議論が日常生活において人間とロボットの相互作用が浸透しつつあることを背景としていることは確かであり、そうした認識は、我々のプロジェクトにおいても共有されているものである。そして、何より、意識、自由意思、意図といった条件を欠いていても「道徳的行為者」と認めることができると考える点において立場を等しくしている。

人間とコミュニケーションをとり相互作用するロボットは今後ますます人間にとって不可欠なものとなっていくと考えられる。それに伴い、本プロジェクト研究が示したような知見は、近い将来、倫理学や情報学のみならず、法学をはじめとする関連の諸分野で新たな潮流となっていくと予想される。その意味では、本研究プロジェクトは先駆的なものと言えるであろう。また、本研究プロジェクトの示した知見は、現在我々の生活に浸透しつつあるロボットとの相互作用を途絶えさせることなく、人間とロボットのよりよい関係性を築いていくことにつながるであろう。

しかしながら、本研究プロジェクトには残された課題も少なからずある。最初に述べたロボット倫理の 2 つの方向の問いのうち、本プロジェクト研究では、「ロボットが守るべき倫理」に関する議論が中心となっており、「ロボットに対する倫理」については十分に考察を加えることができなかった。また、「ロボットのロボットへの倫理」についても問うことができるにも関わらず、その問いに関しては寺沢論文において今後の課題として指摘されるにとどまっている。「ロボットに対する倫理」、そして、「ロボットのロボットへの倫理」について問うことがロボット倫理をより一層深化させることは疑う余地はないであろう、それらの問いに対する考察は今後取り組むべき課題としたいと思う。

*

最後になるが、「ロボット倫理」という様々な研究分野に関わる複合的な新領域に取り組む本研究は、日本大学経済学部グローバル研究センターの「プロジェクト研究」という形を取ることではじめ

て可能となったものである。グローバル研究センター，そして，委員長をはじめ委員の先生方にはあらためて感謝の意を表したい。

注

- (1) イタリアのトリノ大学法学部において法律学の教授を務める。着任前は弁護士として実務に携わる（ウゴ・パガロ『ロボット法』松尾剛行・工藤郁子・赤坂亮太訳，勁草書房，2018年の著者紹介より）。
- (2) この語の日本語訳は上述の訳書に従った。
- (3) この日本語訳についても上述の訳書に従った。邦訳では，p. 92.

引用・参考文献一覧（アルファベット順）

- 平井成興・水川真・林原靖男(2014)「ロボットについて」『ロボット白書 2014—社会を変えようとするとき，そこにロボット技術がある！—』新エネルギー・産業技術総合開発機構，第1章。
- 久木田水生(2009)「ロボット倫理学の可能性」『Prospectus（京都大学文学部哲学研究室紀要）』京都大学，<http://hdl.handle.net/2433/71114>。最終アクセス日：2020年6月15日
- 真部雄介・細野繁・今井晴基(2018)「編集にあたって」『情報処理』（情報処理学会），vol. 59 No. 8, pp. 684-685,
- Pagallo, Ugo (2013) *The Laws of Robots: Crimes, Contracts, and Torts*. Dordrecht: Springer
〔邦訳：ウゴ・パガロ（松尾剛行・工藤郁子・赤坂亮太訳）『ロボット法』勁草書房，2018年〕

ロボット倫理はどこに向かうのか

—哲学・倫理学の立場からの考察—

A Study of Philosophical and Ethical Perspectives on Robot Ethics

根村 直美

Naomi Nemura

〔概要〕

本稿では、＜人間とコミュニケーションをとる機械＞としてのロボットが生活の中に入り込みつつある状況を背景に、ロボットを「言葉や時にはジェスチャーを用いて人とコミュニケーションできる人工物全般」と定義した。

一方、本稿では、Luciano Floridi の情報倫理など現代哲学・倫理学の議論を概観し、人間社会が想定してきた「道徳的行為者」「道徳的受容者」という概念は、何らかの＜実体＞の存在に拠って立つものではなく、＜有機体も非有機体も含む諸々の存在が相互に作用し合う具体的状況において構築されるもの＞という考えを示した。そして、ロボットを「道徳的行為者」として見なすべきかという問いにせよ、また、「道徳的受容者」として見なすべきかという問いにせよ、その答えは、ロボットが人間社会を含む周囲の環境とどのように関わっているかに依存すると論じた。

こうした考察を受け、本稿は、＜人間とコミュニケーションをとる機械＞としての存在感を高めつつある現在の状況においては、すでにロボットを道徳的受容者と見なすことが要請されはじめているという考えを示した。

キーワード：ロボット倫理、道徳的行為者、道徳的受容者、Luciano Floridi

Keywords: robot ethics, moral agent, moral patient, Luciano Floridi

1. はじめに

応用倫理学の文脈では、2000年代後半にすでに、ロボット倫理に関する議論の必要性が意識されはじめていた。とはいうものの、そうした議論はいささかSF的な思考実験と捉えられてきた。

近年、限定的ではあるが自律性をもつロボットが社会の様々な領域で活躍するようになってきている。加えて、シンギュラリティ(技術的特異点)という概念が現実味をもって受けとめられるようになってきている。本稿は、こうした状況を背景として進展しつつあるロボット倫理の議論の現状を明らかにするとともに、今後どのように考えていけばよいのかを哲学・倫理学の立場から探ることを目的とする。

2. ロボットの定義

『ロボット白書 2014—社会を変えようとするとき、そこにロボット技術がある！—』は、ロボットについて完全に一般性をもった定義というのは実は存在しないとしつつ、ロボットの学術的な定義として、日本ロボット学会が編纂した『ロボット工学ハンドブック』（1990, 2005）の定義をいくつか紹介している（平井・水川・林原, 2014: 1）。1967年に開催された第1回ロボットシンポジウムでは、森政弘と合田周平により「移動性、個性性、知能性、汎用性、半機械半人間性、自動性、奴隷性の七つの特性をもつ柔らかい機械」という定義が提唱された。また、人間型ロボット研究開発の草分けであった加藤一郎は「①脳と手と足の3要素をもつ個体、②遠隔受容、接触受容器をもつ、③平衡覚、固有覚をもつ、これらの3条件を備える機械」としている。

スタンフォード大学のB. Rothの定義も紹介されている。Rothは「ロボットとは人間や他の動物あるいは機械と連携して仕事をする機械であって、自動型と半自動型がある。ロボットとは他の自動機械との区別はあまりはっきりしていなくて、かなり気まぐれ的で商業主義的なところがあり、また時間的にも意味は移り変わっている」と述べているという（平井・水川・林原, 2014: 1）。

一方で、実用的な「産業用ロボット」については、日本工業規格（JIS）の「産業用マニピュレーティングロボット-用語」（JIS B 0134-1998）の用語の1100番に「自動制御によるマニピュレーション機能または移動機能を持ち、各種の作業をプログラムによって実行できる、産業に使用される機械」と規定されている（平井・水川・林原, 2014: 1）。

『ロボット白書 2014』によれば、ロボットの定義は「時と場合による」のであり、その定義が流動的な理由は、「そもそも時代とともに科学技術はもとより、産業構造、社会制度、文化なども変化し、ロボットの役割や受け取られ方も変遷してきている」（平井・水川・林原, 2014: 1-2）からである。

ところで、2018年の『情報処理』第59巻8号では、近年のAIブームの中で誕生している製品が人間とのコミュニケーションを重視し、家庭や社会など暮らしの中で活躍する製品であることを背景としつつ、「コミュニケーションをとる機械」としてのロボットに焦点を当てた特集が組まれた（真部・細野・今井, 2018: 684）。『ロボット白書 2014』が言うようにロボットの定義は多種多様であるが、本稿がプロジェクト研究「ロボット倫理はどこへ向かうのか」の一環であり、そのプロジェクト研究が倫理学と情報学の立場からロボットを考えるものであることに鑑み、本稿では、『情報処理』の特集と同様のアプローチを取りたいと思う。すなわち、本稿では、ロボットを、必ずしも「機械の体」をもつものと捉えず、コンピュータグラフィックなども含め、「自律的に判断して動き、言葉や時にはジェスチャーを用いて人とコミュニケーションできる人工物全般」（今井, 2018: 692）を指すこととした。

3. ロボット倫理の基礎理論—哲学・倫理学の立場から—

哲学・倫理学の立場からロボット倫理を考える際に射程となる問いには、大きく言って2つの方向がある。まず、ロボットが人間生活に組み込まれるようになりロボットの人間生活への影響が大きくなるにつれて、ロボットの振る舞いに一定の倫理的制約を課す必要が出てくるであろう。かくして、ロボットに課されるべき倫理的制約とは何か、つまり、「ロボットが守るべき倫理」（久木田, 2009）とは何か問いとして設定されることになる。また、ロボットと生活の中で深く関わるようになったことは、人間はロボットに対してどのような道徳的配慮をなすべきなのかという問いを呼び起こすであ

ろう。すなわち、「ロボットに対する倫理」(久木田, 2009)とはどのようなものなのかを問題とすることができるのである。

「ロボットの守るべき倫理」を問題にするためには、ロボットが「道徳的行為者」になりうることを前提とする必要がある。そこで、いわゆる AMA (Artificial Moral Agent) の概念が議論となってくるのである。その一方で、「ロボットに対する倫理」が問題とされるためには、ロボットが道徳的な配慮を受けなければならない対象である「道徳的受容者」、「道徳的被行為者」となりうるものが前提となる。

2. 1. 道徳的行為者をめぐって

2. 1. 1. 機械倫理の試み

機械倫理の主導者の1人 Susan L. Anderson がそうであるように (Anderson, 2012), 人工的な道徳的行為者をつくる試みは、これまでのところ「人間が従っている倫理的規則を明らかにし、それを機械にプログラムして実行させる」というパラダイムに従って進められている (久木田, 2017: 12)。

Anderson らが開発しているシステムは、何らかの規則ベースの推論システムを前提としているが、あらかじめすべての規則が与えられているわけではない。様々な事例において専門家が下す判断から、機械がその背後にある規則を学習するようになっており、そうした学習をして得た規則に基づき新しい状況においても適切に判断できるようになることを目指している (久木田, 2017: 12)。

このような発想の根底にあるのは、道徳とは理性に基づくものであり、行為者が正しい一般的基準と自分の置かれた状況を適切に理解しており、かつ、十分な思考能力を有しているならば、その行為者は適切に道徳的判断を下すことができる、という「合理主義」(久木田, 2017: 10) と呼ばれる思考である。この思考形式は西洋の倫理学においては1つの主流をなしてきており、機械倫理は西洋の倫理学の伝統を受け継ぐ試みと言えよう。

2. 1. 2. 人工的道徳行為者は本当に「道徳的」と言えるのか

一方で、哲学・倫理学者の間では、ロボットが道徳的規則に従った判断を下し、その規則に従った振る舞いをしたとして、果たして、そのロボットを「道徳的」と評価しうるのかということが問われてきている。

この問題は、人工知能をめぐる「思考をしているかのように振る舞う人工知能は本当に思考をしていると言えるのか？」(久木田, 2017: 17), 「人工知能は心をもつのか？」といった問題と関連させて考えることができるものである。

Alan Turing は、人間の思考のプロセスは観察できないのであるから、外部から観察できる入力と出力の対応だけを思考能力のテストとしようと考えた。そのテストとは次のようなものである (人工知能学会, 2017)。2台のディスプレイの前にテストをする人がいる。1台のディスプレイには隠れている別の人が、もう1台は人間をまねるように作られたコンピュータが受け答えした結果がそれぞれ出てくる。テストをする人はどんな質問をしてもよいとする。詩を作らせたり、音楽の感想を尋ねたりする。また、コンピュータも人間をまねる努力をする。例えば、わざと計算に時間をかけたり、間違えたりする。こうして、テストをする人が、どちらが人間でどちらがコンピュータか分からなければ、このコンピュータには思考能力、あるいは、知性や心があるとするのがチューリングテストであ

る。

これに反論したのが哲学者 John Searle である。Searle は「中国語の部屋」という次のような思考実験を考え出した(人工知能学会, 2017)。英語しか分からない人が部屋にいる。その部屋には、中国語が分からなくても、中国語の文字を書いてあるとおりに置き換えると、中国語の受け答えができてしまう完璧な説明書がある。つまり、この部屋の人には、英語しか分からないが、中国語の質問に中国語で答えることができる。ということは、中国語の受け答えができるだけでは、中国語が分かるとは限らないことになる。同様に、まるで思考能力、あるいは、知性や心があるような受け答えができていうだけでは、本当に思考能力、あるいは、知性や心があるかは分からないという反論である。

道徳的な行為についても同様の議論が可能であろう。道徳的問題に対して対応表を参照しながら答えを出している人がいるとする。外部から見ると、その人は道徳的判断をしているように見える。しかし、その人は道徳的な判断能力をもっているとは限らないのではなかろうか。

確かに、哲学・倫理学では、ある人の判断が道徳的かどうかを問うとき、その人がどう行動するかを問うばかりではなく、どのような意図をもってその行動がなされたのかを問題にしている。例えば、Immanuel Kant は、実践理性の無条件に善い行為をせよと命じる命法に従うことを道徳的な振る舞いとし、義務に従おうとする善意志を重視した(Kant, 1994 [1785])。

こうした意図や動機など道徳的行為者の〈心的な態度〉といったものを問題にする立場からすれば、機械倫理の試みに対して、「人工的道徳行為者は、本当の意味では、“道徳的”とは言えない」という批判が向けられることになるのである。

2. 2. Floridi の道徳的行為者

情報倫理学の分野では、人工的道徳行為者の意図や動機が問題にならないように道徳的行為者の概念を再構成しようとする動きがある。その論者の 1 人が、イタリアの哲学者であるが、オックスフォード大学のスタッフとして主として英米圏で活躍する Luciano Floridi である。

Floridi は、「情報倫理」への統合的アプローチを提唱し、マクロ倫理としての情報倫理の中に道徳的行為者を位置づける(Floridi, 2010: 52-55) ⁽¹⁾。

Floridi は、情報的存在物がつくる情報環境を「情報圏 (infosphere)」と呼ぶ。Floridi の言う「情報的存在物」とは、情報的に理解されるすべての存在物であり、「一群のデータつまり情報オブジェクト」として記述することができるものである。

Floridi は、人工的行為者を道徳的行為者として認めることの問題を認識している。道徳性はふつう、責任(responsibility)を含意する。それゆえ、「しばしば、人工的行為者は自らの行為に責任を取ることができないため道徳的行為者となりえない」と議論されることになる。しかしながら、Floridi は、規範言説の中には、責任の所在を問う問題とは独立のものがいくらかでもあるとして、搜索救助犬の例を挙げる。搜索救助犬は行方不明者の後を探索するよう訓練されている。搜索救助犬はしばしば人命救助に貢献し、飼い主や救助した人から褒められ、報酬を受ける。だがこれは肝心の点ではない。人間は動物に感謝の念をもちうるが、犬にとってはゲームなのであり、その行為について道徳的な責任があると考えすることはできない。大切なのは、犬が道徳と関係のあるゲームに主要なプレイヤーとして参加した点であり、それゆえに、それらが起こした事柄の善悪について結果責任を負いうる(accountable) ⁽²⁾道徳的行為者として認められるのは正当であるという点なのである。

Floridi によれば、道徳的責任を問えなくとも、道徳的行為者は存在しうる。すなわち、道徳に関連する行為について結果責任 (accountability) と能力のみがあり道徳的責任を問えない場合にも、規範的行為をするように促すことができるのである。Floridi は、道徳的責任には意図や自覚、その他種々の心的態度が必要であるが、道徳的責任と道徳に関連する「結果責任」とは区別できると考える。そして、「心をもたない道徳」というアプローチを探究するのである。

こうした Floridi の人工的道徳行為者をめぐる議論は、「自由意志に基づいて行動する理性ある人間」(西垣, 2007) を、道徳的責任を負う道徳的行為者とみなす西洋の倫理学の伝統を問い直すものとなっている。そして、そうした倫理学の伝統を問い直す動きは、現代哲学のうちに表れた動向とも呼応するものである。例えば、「心の哲学」の領域においては、「行動主義」といった立場が出現しており (ウォーバートン, 2010: 216-217), Turing のように、観察できる入力と出力のみに注目しようとする考え方がうまれてきているのである。

2. 3. クリティカルなポスト・ヒューマニズムの立場からの考察

上述のような、「心をもたない道徳」、あるいは、外部から観察される道徳的な振る舞いのみを問題にしようとするアプローチとまではいかなくとも、人間の自律や主体性という概念の見直しは、現代哲学・倫理学の重要な課題の 1 つとなっていることは確かである。

<自律的な主体>の存在を前提とする西洋の<啓蒙主義的ヒューマニズム>の見直しの端緒を開いた 1 人が、Michel Foucault である (Foucault, 1966)。Foucault のそうした考えは、人間の認識過程を知の産出の中心に置くことを拒否した。すなわち、Foucault が注意を喚起しているのは、人間主体が意味を発展させる言説的な構造である。主体は、言説に先んじて存在するのではない。言説によって確立された様式の中にのみ存在する。すべての主体は、言説の中で形成されるのである。

啓蒙主義以来、<自分の理性に従って考え判断し、行動することができる自律的な主体>としての自己という観念こそが、ヒューマニズムを支えてきた。<啓蒙主義的ヒューマニズム>は、自己が自由な意志をもち自律的な主体であることに道徳性の根源を見出し、その自由な意志を尊重することこそがその人の尊重となると考えてきたのである。

しかしながら、Foucault は、個人を、自律し自らの意志を強くもつ主体ではなく、言説の効果と捉えることで、<啓蒙主義的ヒューマニズム>の基盤を掘り崩した。そして、近年、クリティカルなポスト・ヒューマニズムを展開する論者たちは、このような Foucault のヒューマニズム批判を受けた<自律的主体としての人間>概念の再構成を試みている。

例えば、Pranav Nayar のクリティカルなポスト・ヒューマニズムは、自律的な主体に代わり、統一的でない主体像を提示しようとする (Nayar, 2014: 8)。すなわち、そのポスト・ヒューマニズムは、機械と有機的な身体、人間と他の生の形式が、多かれ少なかれ、境界が分からないようにつなぎ合わされており、相互に依存し、ともに進化する仕方に注意を払う。哲学的アプローチとしての Nayar のクリティカルなポスト・ヒューマニズムは、人間の主体性を、共進化する機械と人間の寄せ集めと見るため、主体という観念の捉え直しを迫るのである。

かくして、Nayar は次のように言う (Nayar, 2014: 10)。クリティカルなポスト・ヒューマニズムは、人間の自律を、環境へ反応するシステムの自己制御として見ている。つまり、人間の内的な複雑さを、自己充足的な自律の印ではなく、むしろ、自身を制御し、組み込まれた環境から自分自身とそ

の作用を切り離そうとするシステムの試みと見る。そして、そのポスト・ヒューマニズムは、人間の主体を環境の中で生きられた経験によって生み出されるものと見るとともに、生きられた経験を相互的な関係性を通じ主体によって形成されるものと見るのである。

こうしたクリティカルなポスト・ヒューマニズムは、実体的な道徳的行為者たる〈自律的な自己〉の存在を否定していると見ることができる。その一方で、〈自己〉やその〈心的な態度〉が言説システムを通じて生み出されることを否定してはいない。クリティカルなポスト・ヒューマニズムは、道徳的行為者を社会的な相互作用、とくに〈言語〉を媒介とする相互作用の結果として生み出されるものと見るのである。そうした立場に拠って立つならば、「自律的に判断して動き、言葉や時にはジェスチャーを用いて人とコミュニケーションできる人工物全般」（今井，2018: 692）としてのロボットに対し、道徳的行為者ではないと断言してしまうことはできないであろう。

4. 道徳的受容者をめぐって

さて一方、「ロボットに対する倫理」が問題とされるためには、先述したように、そもそもロボットが道徳的な配慮を受けなければならない「道徳的受容者」、「道徳的被行為者」でありうるものが前提となる。続いて、この問題について考察してみたい。

4. 1. Floridi の道徳的受容者

まず前節で取り上げた Floridi が道徳的受容者についてどのように論じているか、その議論を見てみよう。Floridi は、情報的存在物はすべて固有の道徳的価値をもち、何らかの尊重に値すると論じる (Floridi, 2010: 56-57)⁽³⁾。

そして、「最小限の道徳的尊重」とは、「利害によらず認められ配慮される」ということである (Floridi, 2010: 56)⁽⁴⁾。Floridi によれば、非生物的存在である情報的存在も道徳的配慮を受けるべき存在であり、道徳的受容者にほかならない。Floridi の情報倫理は、「存在物中心的、受容者指向的で、エコロジカルなマクロ倫理」なのである (Floridi, 2010: 47)⁽⁵⁾。

Floridi はまた、情報的存在、あるいは、情報圏一般にとって何が善なのかという、情報倫理によってなされる基本的な道徳的問いについて、次のように論じる (Floridi, 2010: 59)⁽⁶⁾。いかなる情報的存在物も何らかの基本的な倫理的要求の中心と認められる。そういった要求は承認に値するものであり、それと関連するあらゆる情報のプロセスの実行を規制するのに資するべきものである。したがって、道徳的行為者 A の決定や行動を受け入れるか否かは、A の行動が情報圏の善き在り方にいかなる影響をおよぼすかに基づくべきだということになる。換言すると、情報圏を豊かにするか貧しくするかにかかっているのである。道徳的行為者の義務とは、情報圏全体の持続可能な繁栄に対する貢献という点から評価されるべきである。そしてまた、情報圏全体に対して負の影響を与えるプロセスや行為や出来事が悪の事例と見なされるのである。

4. 2. 倫理学における道徳的受容者

続いて、道徳的受容者に関して、倫理学ではどのように考えられてきたかについてまとめる。

近代以降の西洋倫理学においては、いわゆる〈自律的な人間〉というのが道徳的受容者と考えられてきた。この考えは、Kant の「君は、みずからの人格と他のすべての人格のうちに存する人間性を、

いつでも、同時に目的として使用しなければならず、いかなる場合にもたんに手段として使用してはならない」(Kant, 1994[1785]: 52)⁽⁷⁾という議論と結びつけられてきた。Kantの言う「目的」として扱うということは、その相手の意志を尊重することである。Kantの考えに従えば、自由な意志を備えた相手として扱うということが、その人を尊重すること、つまり、その人の「尊厳」を守ることの基盤になっている。

自由な意志を備えた存在であることを道徳的受容者の要件とする考えは、道徳的行為者と道徳的受容者を表裏一体に捉える考え方である。この考えは、近代以降の我々の社会において一定の役割を果たしてきた。また、現在でもそうであると言ってよいであろう。しかしながら、20世紀後半以降の応用倫理学の分野では、我々が配慮すべき存在の範囲はより広く考えられるようになってきている。

例えば、生命倫理学においては、厳密な意味では<自律的な人間>とは言いえない人間存在に対してどのように振る舞うべきかがしばしば重要なテーマとなる。意思決定能力を失った患者や十分な意思決定能力をもつにいたっていない子どもの治療停止に関する議論はその1つの例である。人間社会には、倫理的な関係性を通じて、道徳的行為者とは言えないが、道徳的受容者として位置づけられている存在があるのである。

生命倫理学の分野では、この道徳的受容者には、<個体として生存している人間>と言うことはできない「胎児」、および、「ヒト受精卵」も含まれるようになってきている。日本におけるそうした存在の位置づけを例に取って見てみよう。まず、出生前の胎児については、墮胎罪の規定によって、出生後の人と同程度ではないが、刑法上の保護の対象となっている。また、総合科学技術会議は、2004年に「ヒト胚の取扱いに関する基本的考え方」を取りまとめ、「ヒト受精卵」を次のように位置づけた(総合科学技術会議, 2004: 5)。すなわち、「これまでの社会実態を踏まえて定められた我々の社会規範の中核である現行法体系は、ヒト受精卵を『人』として扱っていない。ヒト受精卵を『人』として扱う考え方を採用することは、この現行法体系を大幅に変更し、受精卵を損なうことを殺人と同義に位置づけることを意味するが、人工妊娠中絶手術が行なわれ、また生殖補助医療において余剰胚等の一部の受精卵を廃棄せざるを得ない現在の社会実態を踏まえれば、そのような制度変更は現実的とは考えられない。また、そのような制度変更について社会的合意を得る見通しもない」とする一方で、ヒト受精卵は「人」へと成長し得る「人の生命の萌芽」として位置づけられ、「通常のヒトの組織、細胞とは異なり、特に尊重されるべき存在」とした。ヒト受精卵は人そのものではないが、人間の尊厳という社会の基本的価値を維持するために特別に尊重されるべき存在とされたのである。

こうした総合科学技術会議の位置づけの背景にあるのは、2003年の生命倫理専門調査会の中間報告書で示された考えである。そこでは、「胚は人ではないとしても『人の生命の萌芽』としての尊重を軽視すれば、それが成長した存在である人に『人の尊厳』があるとする我々の理念の維持は難しくなる」(総合科学技術会議, 2003: 1)と論じている。

さらに、生命倫理学以外の応用倫理学の分野では、道徳的受容者の範囲は「人間」を超えて広がっている(神崎, 2017: 92-94)。

例えば、Peter Singerは、少なくとも人間以外の動物の一部は道徳的受容者であり、それらの動物の利害は人間と等しく配慮されるべきと主張した(Singer, 2002 [1975])。

功利主義者であるSingerにとって、道徳的に重要なのは快樂と苦痛である。それゆえ、道徳的に配慮される道徳的受容者であるための条件は、快樂や苦痛を感じる能力をもつかどうかである。そして、

少なくとも一部の動物は、快樂や苦痛を感じる能力をもっており、その意味で利害をもっている。かくして、それらの動物は人間と等しくその利害を配慮されるべきなのである。Singer は、生物学的に人間に属している個体のみを道徳的な配慮の対象とし、その以外の動物の個体を道徳的な配慮の対象に含めないのは、人種差別と同じく非難されるべき態度であるとし、そのような態度を「種差別」と呼んだ (Singer, 2002 [1975]: 6) ⁽⁸⁾。

一方、環境倫理学もまた、人間だけを道徳受容者であるとする事へ異議を唱えてきている。例えば、Aldo Leopold の「土地倫理」は、生態系を道徳的な配慮の対象とした (Leopold, 1968 [1949])。Leopold の言う「土地」とは、ある空間や地域にすむ動物や植物だけではなく、土や岩や水などの無機物なども含めてすべてのものによって構成される「共同体」としての生態系のことである。Leopold はこの意味での土地こそが最も重要な道徳受容者であるとする立場を表明している。そして、「物事は、生物共同体の全体性、安定性、美観を保つものであれば妥当だし、そうでない場合は間違っているのだ」 (Leopold, 1968 [1949]: 224-225) ⁽⁹⁾と論じている。

4. 3. クリティカルなポスト・ヒューマニズムからの考察

次に、筆者がこれまで注目してきたクリティカルなポスト・ヒューマニズムの立場から、道徳受容者について考えてみたい。

例えば、Pramod Nayar のクリティカルなポスト・ヒューマニズムによれば、人間の主体や身体は、置かれた環境において機械や他の生命との相互的な関係性によって生み出されるものであるが、その一方でまた機械や他の生命との相互作用は形成された主体や身体によって生み出されていく (Nayar, 2014)。つまり、主体と身体は機械や他の生命と複雑に関わり合うことにより<生きられた経験>を通じて生み出されるとともに、<生きられた経験>は機械や他の生命と相互に作用し合う主体と身体によって作りだされるのである。

Nayar のクリティカルなポスト・ヒューマニズム (Nayar, 2014) は、人間の主体と身体を上述のように捉え直すことにより、<啓蒙主義的なヒューマニズム>の<人間>の概念は排他的なものであり、その概念は、ある種のグループに属する人、機械、動物などその概念から排除されてきた者を下位におき抑圧してきたと見なす。

こうした考えに基づき、Nayar は、生物、非生物いずれもを包含するアプローチとしての‘Species Cosmopolitanism’という概念を提示する (Nayar, 2014: 150-156)。Nayar によれば、‘Species Cosmopolitanism’は、ジェンダーであれ (男性/女性)、生の形式であれ (動物/人間、人間/植物)、構成要素であれ (有機体/非有機体)、安易な二項対立を拒絶しているため、クリティカルなポスト・ヒューマニズムの究極の形である。クリティカルなポスト・ヒューマニズムは、あらゆる生の形式を含むマルチな市民権を提唱している。

Nayar などクリティカルなポスト・ヒューマニズムの論者は、サイエンス・フィクションの分析を通じてその思考を鍛え上げてきている。そうした動向を受け、本稿筆者は、アニメーション映画『イノセンス』(2004年、監督:押井守、制作:プロダクション I.G)にもそうしたクリティカルなポスト・ヒューマニズム的思考が認められることを明らかにしてきた (根村, 2016)。

『イノセンス』の、バトーと草薙素子、そして、少女の次の会話を見てみよう。

バトー「犠牲者ができることは考えなかったのか、人間のことじゃねえ・・・魂を吹き込まれた人形がどうなるかは考えなかったのか」

少女「だ・・・だって・・・だって、私は人形になりたくなかったんだもの」

草薙「鳥の血に悲しめど魚の血に悲しまず、声あるものは幸いなり。人形たちにも声があれば、人間になりたくなかったと叫んだでしょうね」

バトーの言葉や草薙の言葉に見て取れるのは、＜啓蒙主義的なヒューマニズム＞が提示する＜人間＞の概念とは異質な存在を、そのままの形で敬おうとする＜敬意＞とも呼ぶべきものである。『イノセンス』における人間以外の存在への配慮は、同質なものへの共感ではなく、異質なものへの＜敬意＞に基づくものと考えられるのである。

『イノセンス』に見られる＜人間性＞を中心とした価値づけを回避しようとする姿勢、その姿勢に基づく異質な他者への＜敬意＞は、Nayarの言うクリティカルなポスト・ヒューマニズムとしての「倫理的プロジェクト」(Nayar, 2014: 30-31)にほかならないと言えよう。

4. 4. 範囲は定められるのか

これまでの考察で、道徳的受容者の範囲と道徳的行為者の範囲は必ずしも一致するものと考えられてはこなかったことは明らかであろう。したがって、ロボットを道徳的行為者と見ることができるかどうかという問題について一定の見解に達することができていなくとも、ロボットを道徳的受容者として見ることができるかどうかという問題について考えることは十分に可能である。

では、道徳的受容者に関してどう考えるべきなのか。神崎宣次は、W. H. Huntが「単なるモノは道徳的に考慮されるか」という論文(Hunt, 1980)でこの問いに対する究極的な回答を与えたと論じている(神崎, 2017: 96-97)。Huntは、Singerが「生物学的に人間である」ことを基準とする立場を種差別としたことに対して、「苦痛を感じる能力」も恣意的に選ばれた基準にすぎず、本当に恣意的ではないのは「生きている」ことなのだと主張したKenneth Goodpasterを批判する。すなわち、Huntは、そのGoodpasterの主張もまた同様に恣意的なものに過ぎず、「存在する」を基準にしてもかまわない、と指摘する。つまり、これまで提案されてきた様々な基準はそれぞれの支持者が恣意的に選んだ条件を主張しているという点で大差はないと言う。そして、特定の条件を道徳的に配慮すべき範囲の境界線を引く基準とするには哲学的正当化が必要だということを主張するのである。

しかしながら、神崎は、そのような哲学的な正当化によって唯一の基準を示すことができるかについて懐疑的である。少なくとも、現状の議論を見る限り、唯一の基準を示すことに成功した倫理学者はいないとする。そして、異なる角度から問題を見ることを提案する。すなわち、「理論的に考えれば道徳的被行為者の範囲を定める条件はこうであるから、これは道徳的被行為者に含まれるというのが話の順序なのではなく、ある特定の対象を道徳的被行為者に含めざるをえなくするような倫理的問題が登場した状況において、その状況の改善につながる条件が基準として主張されるのだと」(神崎, 2017: 97)。

こうした見方は、上述のような主張をうんできた応用倫理学分野の成り立ちを考えるならば頷けるものである。すなわち、人間による環境破壊が深刻になった時期に環境倫理学が登場し、生態系は道徳的受容者であると主張されるようになった。また、動物が苦痛に満ちた扱いをされている状況を改

善しようとする意識が高まり、動物解放の主張がなされるようになった。つまり、道徳的受容者の範囲をめぐる問いは実践的関心に基づいて提起されてきているのである（神崎，2017: 97-98）。

そう考えるとすれば、ロボットを道徳的受容者と見なすべきかどうかという問題は、ロボットを道徳的受容者として考えることが要請されるような状況がどれだけ認めうるかということに依存することになるであろう。神崎は、将来においてロボットが道徳的受容者としてみなされる可能性がないことを意味しないとしつつも、現時点ではそのような状況にはないと論じている（神崎，2017: 98）。しかしながら、ロボットは、「自律的に判断して動き、言葉や時にはジェスチャーを用いて人とコミュニケーションできる人工物全般」（今井，2018: 692）と定義されるようになっていく状況を考えるならば、現時点でも、すでにロボットを道徳的受容者と見なすことが要請される状況が生じつつあると言えるのではなかろうか。

5. ロボット倫理はどこに向かうのか

5. 1. ロボット倫理に向けての第一歩

ここまでの考察をまとめるならば、そもそも、人間社会が想定してきた「道徳的行為者」「道徳的受容者」という概念は、先だって存在している何らかの＜実体＞に拠って立つものではなく、＜有機体も非有機体も含む諸々の存在が相互に作用し合う具体的状況において構築されるもの＞と考えることができるであろう。そして、そう考えることができるならば、ロボットを「道徳的行為者と見なすべきか」という問いにせよ、また、「道徳的受容者と見なすべきか」という問いにせよ、その答えは、ロボットが人間社会を含む周囲の環境とどのように関わっているかに依存すると考えることになるであろう。

その意味では、ロボット倫理を考えるための第一歩は、今現在、ロボットが人間社会を含む周囲の環境とどのように関わっているかをできるだけ具体的に把握することだと言えよう。そして、この第一歩を踏み出した際に目を向けざるをえないのは、やはり、ロボットを「自律的に判断して動き、言葉や時にはジェスチャーを用いて人とコミュニケーションできる人工物全般」（今井，2018: 692）と定義する動きが出てきていることである。

これまでも述べてきたように、人間とのコミュニケーションを主眼とするロボットの捉え方に拠って立つならば、ロボットを道徳的行為者と想定する可能性が開かれることになる。しかしながら、啓蒙主義的な人権思想に基づいて諸制度が設計されている社会においては、一定の条件を満たした人間のみを道徳的行為者と見なしているというのが現状であろう。

それに対し、啓蒙主義的な人権思想に基づく社会においても、人間とコミュニケーションができる動物を配慮すべき存在と見る見方が浸透しつつあり、道徳的受容者に関しては、その範囲が広がりはじめていくと見ることができる。とすれば、人とのコミュニケーションを主眼とするロボットの捉え方は、すでに、既存の人間以外の道徳的受容者と照らし合わせつつロボットを道徳的に位置づけることを要請しはじめていくと考えられるのである。

このように見てくると、Floridi が、人間と機械が融合するシステムがますます増大する情報社会における倫理の出発点として、道徳的受容者に着目したこと（Floridi: 2010）は、妥当性のあるアプローチと言えるであろう。

5. 2. 無自覚の人間中心主義の回避

さて、ロボット倫理をより具体化していくに際して、今一度向き合わなければならないのは道徳的受容者の範囲を定めようとする議論が含みもつ問題点である。これまで道徳的受容者の範囲を定めるための唯一の基準を提示することに成功した者はいないことについては先に論じた通りであるが、そもそも、そのような唯一の基準を設定しようとするには大きな問題が内在するのではなかろうか。

西垣通もこの問題を取り上げる(西垣, 2007: 18-23)。西垣は、「情報圏においては、さまざまな存在物に対応する情報オブジェクトを具体的にいかにかに定めるかが決定的に重要である」が、実際には「人間の分類した概念構造が用いられている」にもかかわらず、この点に関して、Floridi の「反省的な洞察はかならずしも十分ではないように思われる」と論じている(西垣, 2007: 21)。そして、Floridi の情報倫理は、「理性ある人間はとくに選ばれ、神からの管理者としての使命を与えられている。したがって、生物も含めて万物のあり方を正しく保護管理し、調和ある繁栄をもたらさなければならない」と考える西洋流エコロジー思想をデジタル情報圏に展開したものにほかならないとしている(西垣, 2007: 20)。つまり、人間が唯一の基準を設定しようとすることは、人間を「神から使命を与えられた特別な代理者」の地位に置くものであり、人間中心主義を超克しきれてはいないとするのである(西垣, 2007: 23)。

これまでの考察では、倫理学者らの議論の現状を受けて唯一の基準を提示することに誰も成功していないと論じてきた。しかしながら、西垣の言うように、無自覚の人間中心主義という陥穽に陥ってしまうことを避けるためには、特定の研究者が唯一不変の基準を設定しようとする事自体を避けねばならないと言うべきであろう。そして、道徳的受容者の範囲に関する問いが一義的な解答に行きつくことがないことを前提とするならば、その問いは常に問われ続けるものとならざるをえないであろう。このことはまた、「どういう存在を道徳的行為者と見なすべきか」という問いに関しても同様である。特定の研究者が唯一不変の基準を設定しようとする事は避けねばならない。その答えが一義的な解答に行きつくことはないがゆえ、その問いは問われ続けなければならないのである。

ところで、無自覚の人間中心主義を避けるためとして、西垣は伝統的・人文主義的な西洋哲学とは異なる視点から倫理を構想しようとする(西垣, 2007: 25-30)。そこで提示されるのが、人間が生物の一員だという自然科学的観点をより徹底させるオートポイエーシス理論である。ここで大きな疑問がうまれてくる。このオートポイエーシス理論に基づいて、生物と機械を峻別できるとすることは、この理論に拠って立つ者を無自覚に神の視座に置くことになっているのではなかろうか。そして、ある範囲の設定をすることにより神の視座に立つことに対して無自覚なことはまさに西垣自身が批判しようとしていることではないのか。

西垣の議論を見ていると、オートポイエーシス理論を無意識の人間中心主義から免れているものとするのは、それが「自然科学的観点」に基づいたものであるからのように思われる。

しかしながら、科学論者の中には、何らかの自然科学言説が<ありのままの事実>を映し出している、すなわち、<真理>を言い当てていると考えることに異を唱え、自然科学も時代や地域といった社会・文化背景から切り離すことはできないという認識を提示する者が少なからずいる(Haraway, 1999, 2000 ; Hacking, 2006 ; Gergen, 2009, 2015)。

このような自然科学に対する考え方は、社会構成主義と言われる立場に属している。その立場は、

Keneth J. Gergen を援用するならば、次のように論じることができるであろう。すなわち、その考え方は、何かを命名し記述する自然科学者の言葉には存在する何かを写し出す特権があり、それゆえ、自然科学の言葉が私たちに唯一の現実を教えてくれると考えることに対して危惧を示し（Gergen, 2015: 231-232）⁽¹⁰⁾、自然科学が「それ以外のあらゆる声を沈黙させる独裁者」（Gergen, 2015: 22）⁽¹¹⁾とならないようにと警鐘を鳴らす。そして、自然科学者が自らの説明を〈ありのままの事実〉、あるいは、〈真理〉と考えることにより、その説明を反省的に問い直すことができなくなっていることを批判するのである。

本稿は先に、実践的関心がどのようなところにあるのかを明らかにすることを第一歩とすると述べたが、無自覚の人間中心主義を回避するためにも、特定の研究者が唯一不変の絶対的な境界線を設定することを避け、常に変化する現実の実践的関心を探ることに基づいて暫定的な境界線を設定するというアプローチを取るのがロボット倫理の構築にはふさわしいのではなかろうか。

このようなアプローチは、結局は人間がその関心に基づいて境界設定をしているという意味では、「人間中心主義」を完全に超克しているとは言えないかもしれない。しかしながら、このアプローチは人間の実践的関心を探ることによって基準を設定せざるをえないということを深く自覚するアプローチである。それは、人間の生み出す知はどれも絶対的基準を設定しえないと認識するがゆえに、人間を特別な地位に置くことが出来ないと考えるアプローチであるという意味では「人間中心主義」を回避しえていると言えるであろう。

5. 3. 実践的関心をいかにして探るか

それでは、実践的関心を探るには、どのような方法があるのだろうか。その1つとして挙げられるのは、本稿筆者がこれまでサイバー・カルチャー研究において取ってきた現象学的方法である。この方法は、フッサール現象学の、普遍的なものである本質を把握する方法としての現象学的方法ではなく、経験の事実的な構造を明らかにしようとする方法である（根村, 2017）。〈本質〉を取り出そうという態度は現象を何らかの世界の現れとして捉えることであり、現象をその現象たらしめている起源を想定することである。しかしながら、経験の事実的構造を明らかにしようとする現象学的方法は、そうした起源も基礎も想定しない。現象を何らかの世界の現れとしてではなく、現れそのものとして捉えようとする試みは、「現象学のラディカル化」、あるいは、「現象学に対するラディカルな批判」と見なされてもいる（Colebrook, 2002: 6）⁽¹²⁾。

また、現象学的方法は、多くの場合「解釈学的方法」を要求する（リー, 2009: 29-30）。経験は、その研究に携わる者が自己反省を通して直接的に把握しうるものと、内省を通じて直接には把握できないものの2種類に分けられるからである。第1の種類の経験の場合は、内省という方法により研究に携わる者に直接的に把握されるため、解釈学的方法が要請されるということはない。一方、第2の種類の経験は、研究に携わる者が直接には把握できない他者の経験や自身の経験であっても過去のものとなっている経験などがそれに当たる。第2の種類の経験は直接的に内省によって把握されるのではなく、解釈によって間接的にのみ把握されるのであり、解釈学的方法が必要となる。この場合、解釈学的方法は現象学的方法にとって不可欠な要素となる。解釈学的方法が現象学的研究を遂行するために必要とされるのである。

この方法に関連して、知能研究の新しい方法として「一人称研究」が提案されていることにも触れ

ておきたい。その方法は、「数少ない被験者の個別具体的状況に（一人称的な N=1 の事象に）面白い現象を見出し、それを基に知の面白い側面について仮説を立て、その仮説を検索クエリーにして、同じ現象が他者にも成りたちはしないかと探す」（諏訪, 2015: vi）ことによって普遍を探究しようとするものである。

人工知能研究の新たな潮流である「一人称研究」という方法と本稿筆者が提示する方法とは探究のプロセスにおいて相通じるところが多いと言えよう。本稿が提示する方法も、ごく個別的経験を基に、ロボットと人間のコミュニケーションのうちどのような倫理が立ち現れつつあるのかを探り、そこから仮説を立てていくことになるであろう。この仮説は唯一で普遍的な基準となりえないものである。本稿筆者の方法論においては、それは状況依存的なものにならざるをえず、どのような倫理が立ち現れているのかという問いは常に問われ続けなければならないものとなるであろう。諏訪正樹は、「一人称研究」においては、「個から普遍をみる」とは普遍的な知見にたどり着くということそのものではなく、個々の研究者が自分なりの仮説を立て普遍的知見を指向する探究のプロセスを意味するのかもしれないと論じているが（諏訪, 2015: 41-42）、本稿筆者の方法においてもまた、仮説そのものにおいてではなく、その探究のプロセスにこそ普遍を見ることができのかもしれない。

6. 結び

本稿では、〈人間とコミュニケーションをとる機械〉としてのロボットが社会の中に入り込みつつある状況を背景に、ロボットを、「自律的に判断して動き、言葉や時にはジェスチャーを用いて人とコミュニケーションできる人工物全般」（今井, 2018: 692）と定義し、議論をはじめた。

一方、本稿では、現代哲学・倫理学の議論を参照しつつ、人間社会が想定してきた「道徳的行為者」「道徳的受容者」という概念は、何らかの〈実体〉の存在に拠って立つものではなく、〈有機体も非有機体も含む諸々の存在が相互に作用し合う具体的状況において構築されるもの〉と考えるに至った。そして、そうした考えに基づき、ロボットを「道徳的行為者」として見なすべきかという問い、「道徳的受容者」として見なすべきかという問いのいずれも、その答えは、ロボットが人間社会を含む周囲の環境とどのように関わっているかに依っていると結論づけた。

かくして、ロボットが〈人間とコミュニケーションをとる機械〉としての存在感を高めている状況を考慮するとき、すでにロボットを道徳的受容者と見なすことが要請される状況にあると考えるのが本稿の立場である。人間はロボットに対してどのような道徳的配慮をなすべきなのかを、経験の構造を探る現象学的方法を用いて探究するのは今後の課題としたい。

注

- (1) 日本語訳は、西垣通の邦訳「情報倫理の本質と範囲」（西垣通・竹之内禎編著訳『情報倫理の思想』所収、NTT 出版、2007年、pp. 48-98）を参考とした。邦訳では、pp. 69-77.
- (2) 邦訳では accountable は「説明できる(説明義務がある)」と訳されているが、本稿では「結果責任がある」という訳を当てている。accountability も同様に「結果責任」という訳を当てた。
- (3) 邦訳では、pp. 77-81.
- (4) 元々は R. Hepburn の議論である(Hepburn, 1984)。邦訳では、p. 77.
- (5) 邦訳では、p. 59.

- (6) 邦訳では, pp. 83-84.
- (7) 日本語訳は, 中山元の邦訳『道徳形而上学の基礎づけ(1785)』光文社〔光文社古典新訳文庫〕, 2012年)を参考にした。邦訳では, p. 136.
- (8) 日本語訳は, 戸田清の邦訳『動物の解放』改訂版(人文書院, 2011年)を参考にした。邦訳では, p. 27.
- (9) 日本語訳は, 新島昭の邦訳『野生のうたが聞こえる』(講談社〔講談社学術文庫〕, 1997年)を参考にした。邦訳では, p. 349.
- (10) 日本語訳は, 東村知子の1999年版の邦訳『あなたへの社会構成主義』(ナカニシヤ出版, 2004年)を参考にした。邦訳では, pp. 349-350.
- (11) 邦訳では, p. 77.
- (12) 日本語訳は, 國分功一郎の邦訳『ジル・ドゥルーズ』(青土社, 2006年)を参考にした。邦訳では, pp. 20-21.

引用・参考文献一覧

- Anderson, Suzan Leigh (2012) “Machine Metaethics.” In Michael Anderson and Suzan Leigh Anderson (eds.), *Machine Ethics*. Cambridge: Cambridge University Press, pp. 21-27.
- Colebrook, Claire (2002) *Gilles Deleuze*. London: Routledge.
〔邦訳:クレア・コールブルック(國分功一郎訳)『ジル・ドゥルーズ』青土社, 2006年〕
- Floridi, Luciano (2010) “Information Ethics: Its nature and Scope.” In Jeroen van den Hoven and John Weckert (eds.), *Information Technology and Moral Philosophy*. Cambridge: Cambridge University Press, pp. 40-65.
〔邦訳:ルチアーノ・フロリディ(西垣通訳)「情報倫理の本質と範囲」西垣通・竹之内禎編著訳『情報倫理の思想』, NTT出版, 2007年, pp. 48-98〕
- Foucault, Michel (1966) *Les mots et les choses. Une archeologie des sciences humaines*. Paris: Gallimard.
〔邦訳:ミシェル・フーコー(渡辺一民・佐々木明訳)『言葉と物—人文科学の考古学—』新潮社, 1974年〕
- Gergen, Kenneth J. (1999) *An Invitation to Social Construction*. London: Sage.
〔1999年版の邦訳:ケネス・ガーゲン(東村知子訳)『あなたへの社会構成主義』ナカニシヤ出版, 2004年〕
- (2009) *An Invitation to Social Construction*. Second edition. Los Angeles: Sage.
- (2015) *An Invitation to Social Construction*. Third edition. Los Angeles: Sage.
- Hacking, Ian (1999) *The Social Construction of What?* Cambridge, Massachusetts: Harvard University Press.
〔邦訳:イアン・ハッキング(出口康夫・久米暁訳)『何が社会的に構成されるのか』岩波書店, 2006年〕

- Haraway, Donna, J. (1991) *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge.
- [邦訳:ダナ・ハラウェイ (高橋さきの訳)『猿と女とサイボーグー自然の再発明ー』青土社, 2000年]
- (2000) *How Like a Leaf: An Interview With Thyrza Nichols Goodeve*. New York: Routledge.
- [邦訳:ダナ・ハラウェイ+シルザ・ニコルズ・グッドイヴ (高橋透・北村有紀子訳)『サイボーグ・ダイアローグズ』水声社, 2007年]
- Hepburn, R. (1984) *Wonder and the Other Essays: Eight Students in Aesthetics and Neighboring Fields*. Edinburgh: Edinburgh University Press.
- 平井成興・水川真・林原靖男(2014)「ロボットについて」『ロボット白書 2014—社会を変えようとするとき, そこにロボット技術がある!—』新エネルギー・産業技術総合開発機構.
- Hunt, W. H. (1980) “Are Mere Things Morally Considerable?” *Environmental Ethics* 2.1, pp. 59-65.
- 今井倫太 (2018) 「なぜロボットを使うの?」『情報処理』(情報処理学会), vol. 59 No. 8, pp. 692-697.
- 人工知能学会 (2017) 「チューリングテストと中国語の部屋」 <https://www.ai-gakkai.or.jp/whatsai/AItopics3.html>, 最終アクセス日: 2019年3月27日.
- Kant, Immanuel (1994) *Grundlegung Zur Metaphysik Der Sitten* [1785]. Hamburg: Meiner.
- [邦訳:イマヌエル・カント (中山元訳)『道徳形而上学の基礎づけ(1785)』光文社 (光文社古典新訳文庫), 2012年]
- 神崎宣次 (2017) 「この映画の撮影で虐待されたロボットはいません—道徳的被行為者について考える—」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, pp. 83-101.
- 久木田水生(2009)「ロボット倫理学の可能性」『Prospectus (京都大学文学部哲学研究室紀要)』(第11巻, <http://hdl.handle.net/2433/71114>. 最終アクセス日: 2020年6月15日.
- 久木田水生(2017)「機械の中の道徳—道徳的であるとはそもそもどういうことかを考える—」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, pp. 3-21.
- リー, ナミン(2009)「現象学と質的研究の方法」『死生学研究』第12号, 東京大学大学院人文社会系研究科, 吉田聡訳, pp. 8-34.
- Leopold, Aldo (1968) *A Sand County Almanac and Sketches Here and There: With Other Essays on Conservation from Round River* [1949]. New York: Oxford University Press.
- [邦訳:アルド・レオポルド (新島昭訳)『野生のうたが聞こえる』講談社 (講談社学術文庫), 1997年]
- Nayar, Pramod K. (2014) *Posthumanism*. Cambridge: Polity Press.
- 根村直美 (2016) 「『イノセンス』に見るポスト・ヒューマニズムと<身体>の構築主義」『社会情報学』(社会情報学会), 第5巻1号, pp. 73-88.

- (2017) 『ポストヒューマン・エシックス序説—サイバー・カルチャーの身体を問う—』, 青弓社.
- 西垣通 (2007) 「普遍倫理への模索」西垣通・竹之内禎編著訳『情報倫理の思想』NTT 出版, pp. 7-45.
- 真部雄介・細野繁・今井晴基(2018)「編集にあたって」『情報処理』(情報処理学会). vol. 59 No. 8, pp. 684-685,
- Singer, Peter (2009) *Animal Liberation* [1975]. New York : Harper Perennial Modern Classics.
〔邦訳:ピーター・シンガー (戸田清訳)『動物の解放』改訂版, 人文書院, 2011年〕
- 総合科学技術会議・生命倫理専門調査会 (2004) 『ヒト胚の取扱いに関する基本的考え方(中間報告書)』, 内閣府.
- 総合科学技術会議 (2004) 『ヒト胚の取扱いに関する基本的考え方』, 内閣府.
- 諏訪正樹 (2015) 『一人称研究のすすめ—知能研究の新しい潮流—』近代科学社.
- ウォーバートン, ナイジェル (2010) 『哲学の基礎』(栗原泉訳), 講談社.

【謝辞】

副代表者・寺沢幹雄教授には、本稿の執筆に当たって重要な資料、および、情報を提供していただいた。心より感謝を申し上げたい。また、安田静教授には、ワークショップにおいて貴重な指摘をしていただいた。同様に感謝の意を表したい。

〔追記〕本稿は、『研究紀要』第 89 号 (日本大学経済学部, 2020 年 9 月 20 日発行) に掲載された論文「ロボット倫理はどこに向かうのか」に修正を加えたものである。

親近感に基づくロボットの倫理的責任

Ethical Responsibility of Robots Based on Intimacy

寺沢 幹雄

Mikio Terasawa

〔概要〕

本研究は、人とロボットが良好な関係を築くことをめざし、親近感の観点からロボットを分類し、制御手法に応じた倫理的責任を定量化することを目的とする。また、ロボット制御のための AI 技術の知識学習における倫理的課題を明確にする。親近感を表す指標として、ロボットの外形、利用目的、複雑性要因を定義することでロボットの倫理的関与の度合いを明確にした。ロボットの責任主体として設計者、育成者、利用者に分け、組み込みプログラム、ティーチング、学習の動作設計による過失割合の定量化をはかった。また、ロボットの設計者やロボット単体の中で閉じた挙動ではなく、社会の中で知識を学習するロボットが満たすべき行動規範と、行動規範に基づく責任問題について論じ、親近感により考慮すべき点が異なることを明らかにした。ロボットがビッグデータによる倫理的知識を学習する際に、身近さ、権威、時間を要因とする道徳的価値観が重要であることを明確にした。本研究では、ロボットから人への倫理問題に注目したが、人からロボットへ、ロボットからロボットへの倫理問題については今後の検討課題である。

キーワード：ロボット倫理，責任能力，親近感，知識学習

Keywords: robot ethics, responsibility, intimacy, knowledge learning

1. はじめに

1. 1. 人工知能，IoT 技術の発展とロボットの普及

2010 年代に第 3 次 AI ブームが到来し、ビッグデータ解析をはじめとした実用的な問題解決に有効な手法として人工知能が注目されるようになった。産業応用のみならず、日常生活の中でも Siri や Google Assistant をはじめとした人工知能による対話技術が検索やパーソナルアシスタントとして広く利用されている。いっぽう、インターネットとセンサ技術に基づく IoT も、自動制御をはじめとした産業応用のみならず、高齢者や子供たちの見守りをはじめとしたホームコンピューティング技術として身近な存在となっている。

人工知能と IoT 技術が日常生活に浸透することにより、知能とセンサの基本技術にアクチュエータを統合したものとして発展してきたロボットの一般消費者への普及も期待される。現在、家庭内におけるロボットは、Aibo や Robi などのペットロボットが主流であり、一部の愛好家を対象とした趣味の領域を出ていないが、ロボット掃除機、見守りポット、スマートスピーカなどのロボット家電が広く認知され、違和感なく受け入れられていることから、今後の家庭用ロボットには市場の拡大が見込まれる。また、店頭の販促用に

Pepper が使われたり、ロボットが受け付け業務をするホテルが現れたり、日常生活でもロボットに接する機会が増えている。

ロボットが人の暮らしの中に浸透すると、人とロボットのつきあい方において、今以上に考慮すべき課題に直面することとなる。ロボットは単純な機械と異なり、人間らしさが本質と考えられる。したがって、人間らしい機械への接し方、すなわち倫理的側面に関する考察が重要になる。

1. 2. ロボット三原則

ロボットの倫理的側面に関して、古くは SF 作家のアイザック・アシモフによって、以下のようなロボット三原則が提唱され、広く知られている(アシモフ, 1983)。

第一条 ロボットは人間に危害を加えてはならない。また、その危険を看過することによって、人間に危害を及ぼしてはならない。

第二条 ロボットは人間にあたえられた命令に服従しなければならない。ただし、あたえられた命令が、第一条に反する場合は、この限りでない。

第三条 ロボットは、前掲第一条および第二条に反するおそれのないかぎり、自己をまもらなければならない。

ロボット三原則については、矛盾した場合の対処法や、適用順序などの厳密性を欠くとのことで、後に様々な修正が加えられている。当時のロボットは、与えられたプログラムで動く機械であり、静的なアルゴリズムに基づいていた。したがって、開発段階における行動規範の策定が重要な意味を持っていた。しかし、現在のロボットは、AI 技術を導入したものが多く、後天的な知識の獲得により挙動が動的に決められる。また、ビッグデータから知識を得ることが多く、年齢、性別、地域、民族、宗教などの多様性が尊重され、状況がめまぐるしく変化する現代では、固定した倫理観に基づくロボット制御はいっそう困難となっている。

1. 3. 従来のロボット倫理研究

従来のロボット倫理研究は、ロボットが道徳的行為者であるか否かに論点が集中していた。すなわち、「ロボットは快樂や苦痛を感じるか」、「ロボットは善悪の判断能力があるか」、「ロボットは責任を負うことができるか」という議論である(久木田・神崎・佐々木, 2017)(フロリディ・カプーロ・エス, 2007)。しかし、ロボットは知的な機械という認識はあるものの、大前提であるロボット自体についての定義、分類のないままに議論が進展している。

一般的に、制御系と駆動系を持つ機械を自動機械と呼ぶが、ロボットと自動機械との境界は必ずしも明確ではない。場合によっては、ロボットコンテストにあるような単一機能の機械をロボットと呼ぶ場合もある。また、産業用ロボットのように、決められたアルゴリズムに基づくプログラムで動作するものや、オペレータによる変数設定の自由度だけを持つものもある。主に産業応用の観点から、ロボットは、用途、関節の自由度、ティーチングの有無などに応じて分類されていた。産業応用の分類は、倫理問題を論じる上で適切な類型化とはいえない。また、AI 技術の知識学習は、教師あり学習、教師なし学習、強化学習をベースとしているが、その特性に基づいた倫理的責任については論じられない。

1. 4. 本研究の目的

本研究では、倫理的問題を論じるのに適切なロボットの分類をハードウェアとソフトウェアに分けて提案する。ハードウェアについては、倫理問題を論じる上で人間らしさの尺度によって倫理観も大きく変わることから、人間らしさ、すなわち親近感に基づいてロボットを分類・整理する。また、ソフトウェアについては、動作設定の主体を明確にすることで、責任の所在を明確にする。

本研究は、人とロボットが良好な関係を築くことをめざし、親近感の観点からロボットを分類し、倫理的課題を明らかにするとともに、動的な倫理知識学習の注意点を探ることを目的とする。

1. 5. 本稿の構成

2章では、従来からの一般的なロボットの分類を整理する。3章では、親近感に基づいてロボットハードウェアを分類する。4章では、責任の主体と動作設定手法を分類する。5章では、各要因とロボットの責任の関係を明らかにする。6章では、倫理的ロボット実現のための知識学習における注意点を考察する。7章では、社会的責任とロボットの行動規範について論じる。

2. 一般的なロボットの分類

「NEDO ロボット白書 2014」によると、ロボットは、センサ、知能・制御系、駆動系の3つの要素技術を持つ機械システムとされている(NEDO, 2014)。この定義に基づくと、単純な自動機械を含む多くの機械がロボットの範疇に属する。一般にロボット倫理で想定される知能ロボット以外にも、多種多様なロボットが存在することを示すために、代表的なロボットの一例を産業用と家庭用に分けて列挙する。

2. 1. 産業用ロボット

産業用ロボットは、加工、組み立て、検査、搬送の各工程で使われる汎用機械であり、利用者は主に専門的な技術者である。代表的な産業用ロボットを以下に挙げる。

① スカラロボット

関節の自由度が少なく、高速な組み立てが可能な、機械に近いロボット

② 産業用5軸ロボット

人の腕のようなアームで複雑な組み立てや塗装が可能なロボット

③ 災害救助ロボット

災害救助犬の代替として、災害地で遠隔操作が可能なロボット

④ 医療ロボット

複数の腕を持ち、遠隔操作で細かい手術を行うことができるロボット

⑤ 清掃ロボット

動物のように多関節で、細い通路などを点検・検査・清掃できるロボット

⑥ 人型ロボット

全身が人と同様な関節で構成された、主に実験的なロボット

2. 2. 家庭用ロボット

家庭用ロボットは、家庭内や店頭などで一般消費者と接するロボットであり、実用性と同様に、いやし効果が求められることが多い。代表的な家庭用ロボットを以下に挙げる。

- ① ロボット掃除機
Roomba などの環境獲得と最適な経路設定機能をもつ掃除用ロボット
- ② ロボット型携帯電話
ロボホンなどのペットロボットの機能を持つ電話機
- ③ スマートスピーカ
Amazon Echo, Google Home, LINE Clova などの高度な対話機能を持つスピーカ
- ④ 対話ロボット
Robi などの関節数は少ないが人に似た形状を持ち、人の話を聞き取り応答するロボット
- ⑤ ペットロボット
Aibo などのペットとしての機能に特化した多関節の動物型ロボット
- ⑥ アシスタントロボット
Pepper などの対話機能と豊かな表情を持ち、簡単な動きのできる人型ロボット

3. ロボットハードウェアの分類

3. 1. 親近感

2 章で述べた一般的なロボットの分類は機能や応用に基づくものであり、同じ分類に属するロボットであっても人間が道徳的主体と感ずるかどうかとは無関係である。倫理的側面が問題になるのはロボットが人間と同等の機能、形態を有する場合であることが多いことから、本研究では、人間らしい親しみやすさ、すなわち「親近感」をハードウェアの分類基準として用いる。

3. 2. 親近感によるハードウェア分類

本研究では、ロボットを自律的に作動し、人とのインタラクションによって知的に行動する機械と定義する。広い意味では、AI スピーカのようなソフトウェアもこの範疇に属する。

産業用ロボットは、組み立て、加工、検査などの生産工程で利用される汎用な自動機械であり、通常は形態的な特徴を持たない。いっぽう、サービスロボット、ペットロボット、コミュニケーションロボットなどの家庭用ロボットは、知能や形態において、親近感を持つことが特徴的である。親近感については、相手が目に見えない場合、チューリングテストが一つの基準となりうる。すなわち、見えない対話相手を機械と判別できるかどうかを調べる方法である。しかし、必ずしもチューリングテストほど厳しい基準を満たさなくても、親近感を感じることはある。たとえば、人形やぬいぐるみをはじめ、極端な場合には縫い針などの日用品にも魂を感じる文化もある。

親近感を持つか持たないかは、人として守るべき行為としての倫理問題を考える上で大きな要因となる。本研究では、外形、利用目的、複雑性の観点から親近感の定式化をはかる。

3. 2. 1 外形による指標

外形からみたロボットに対する親近感は、形状、大きさ、装飾によって影響を受ける。

(1) 形状

形状については、人間や動物に近い生命感をもつほうが、機能的・機械的なものよりも親近感を感じられる。たとえば、直線的な骨格だけで機能を果たせるロボットであっても、曲線的な形状とすることで親しみやすくなる。

(2) 大きさ

大きさについては、人間よりも小さいほうが、人間を威圧するような大きさのものよりも親近感を感じられる。たとえば、ペットや子供のような大きさのものは親しみやすさが感じられ、巨大なロボットは脅威を感じさせる。

(3) 装飾

デザイン的については、自然に近い装飾のほうが、無機質なものよりも親近感を感じられる。たとえば、同じ産業用ロボットでも、無機質なデザインよりも、顔や模様などの装飾を施した方が愛着を得られやすいとされる。

3. 2. 2 利用目的による指標

本来、ロボットは、人間が行う作業を代替する機械としての役割を持ち、作業内容が機能的であるか、娯乐的であるかによって親近感が異なる。

(1) 利用者数

利用者数については、利用者が多いほど一般的となり親しみを感じられる。たとえば、利用者の多い低価格なロボットほど多くの人になじみがある。

(2) 専門性

専門性については、汎用的なロボットほど用途が専門的なものよりも親近感を感じられる。たとえば、産業応用を目的として専門家のみが扱う産業用ロボットより、誰でもが使える家庭用娯楽ロボットのほうが親近感を抱きやすい。

3. 2. 3 複雑性による指標

動作が単純な機械はすぐに飽きてしまうが、動作が複雑に変化する機械であれば長く利用することにより親近感がわく。複雑さの要因としては、センサ・駆動系の自由度と、入出力の関係性が親近感に影響する。

(1) センサ・駆動系の自由度

入力であるセンサの個数が多く、出力である駆動系が多いほど、飽きることなく親近感がわく。たとえば、産業用ロボットのように決められた組み立て、溶接、検査などの単一動作を繰り返すロボットよりも、ペットロボットのように人間の動作に応じて会話や反応を行うロボットが親近感を感じられる。

(2) 入出力の関係性

入力と出力の関係が複雑なほうが飽きることなく親近感がわく。たとえば、入力が多数で、結果が予想不可能なものであるほど複雑であり、複雑なものには興味が持続し、結果として親近感を感じられる。

3. 3. 親近感による分類

3.2.1 から 3.2.3 節で述べた外形, 利用目的, 複雑性による指標をそれぞれ, F_D , F_P , F_C とおき, $0 \leq F_D \leq 1$, $0 \leq F_P \leq 1$, $0 \leq F_C \leq 1$ と数値化する. F_D は, 形状, 大きさ, 装飾を, F_P は, 利用者数, 専門性を, F_C は, センサ・駆動系の自由度, 入出力の関係性をそれぞれパラメータとする. それぞれの指標が 1 に近いほど親近感が高く人間的であり, 0 に近いほど親近感が低く機械的であるものとする. このとき, 親近感の指標 F をそれぞれの指標に重みをつけて線形に結合し, $F = w_D F_D + w_P F_P + w_C F_C$ と定義する. ロボットの倫理感が問題になるのは, ロボットが人間に近いとき, すなわち, F が大きく親近感が高いときである.

外形, 利用目的, 複雑性の各パラメータには, 客観的な数値で表される量的変数と, 数値で表現できない質的変数がある. 数値が一意に決まらない場合でも, 実体を比較することによって順序尺度を与えることができる. 順序尺度の数値化にあたっては, ロジックモデルで分布関数をあてはめる.

ビッグデータ等で収集したデータを利用してロボットの実体についてパラメータをあてはめ, 外形, 利用目的, 複雑性の親近感の指標を 3 変数とする 3 次元空間でクラスタリングを行うことによりロボットの分類を行う.

4. 動作設定によるソフトウェア分類

4. 1. ロボット動作を規定する主体

「ロボットは快楽や苦痛を感じるか」, 「ロボットは善悪の判断能力があるか」, 「ロボットは責任を負うことができるか」という議論では, ロボットが自律した動作主体であることが前提となっている. しかし, ロボットは機械であり, 基本的には人の意思が介在する. ロボットの善悪の判断能力, 責任能力を論ずるにあたっては, ロボットの動作を規定する主体を明確にする必要がある.

ロボットの動作を規定する主体は大きく分けると, 以下の通りである. なお, 設計者と育成者, 育成者と利用者が同一である場合もある.

(1) 設計者

設計者とは, ロボット挙動のアルゴリズムを設計者だけでなく, 広くロボットの製造に携わる者を表す.

(2) 育成者

育成者とは, 完成したロボットに制御変数を設定する者を表す.

(3) 利用者

利用者とは, 受動的に使用しながらも, 結果的にロボットの動作に影響を与える者を表す.

4. 2. ロボットの動作設定

ロボットの動作はプログラムによって規定されるが, 設定した手続きがいつい変化しない静的なプログラムもあれば, 利用開始後に動的に変化するプログラムもある. したがって, 動作設定アルゴリズムによって責任の所在が大きく異なる.

ロボットのプログラム実現方法は, 大きく分けると以下の通りである.

(1) 組み込みプログラム

組み込みプログラムは, 設計段階で指定した通りに静的な挙動を繰り返し, その後の利用によって変化しない. ただし, 設計段階で考慮した環境情報の入力に対しては動的に反応する.

(2) ティーチング

ティーチングとは、主に産業用ロボットのならい運転のことで、人間のエキスパートの挙動に準じて動作する。通常は、エキスパートが指定した後に動作は変化しない。

(3) 学習

近年は、機械学習、深層学習などの AI 技術でロボットを制御することが多くなっている。あらかじめ学習したものを利用者に提供する場合もあれば、利用段階で動的に学習を進める場合もある。学習としては、以下に分かれる。

(3-1) 教師あり学習

ロボットの教師あり学習は、入力と挙動の組を検証データとして用いることで動作モデルを作り、新たな入力に対する挙動を決定する。

(3-2) 教師なし学習

ロボットの教師なし学習は、多数のデータを利用することで挙動を決定するものであり、主に画像処理や協調動作に使われる。

(3-3) 強化学習

ロボットの強化学習は、各駆動系の検証データはないが、目的とする評価値が明確な場合に有効で、多関節のロボットなどで使われる。

5. ロボットの善悪判断能力と責任

製品であるロボットの動作による結果については、善悪の判断を含めて製造物責任法により製造者に責任がある。しかし、故障などではなく、意図した動作の場合、設計者の責任については議論が残る。特に、複雑性による指標が高いロボットの場合は、あらかじめ作られたプログラムよりも、後から得られた知識に問題があることが考えられる。

ロボットの善悪判断、および責任は、最終的に設計者、育成者、利用者のいずれかに起因する。設計者、育成者、利用者の過失割合をそれぞれ r_D , r_I , r_U とする。ただし、 $0 \leq r_D \leq 1$, $0 \leq r_I \leq 1$, $0 \leq r_U \leq 1$ であり、 $r_D + r_I + r_U = 1$ とする。ロボットの責任の総量を R とすると、設計者、育成者、利用者の責任は、それぞれ $r_D R$, $r_I R$, $r_U R$ となる。なお、ロボットの責任の総量 R は、親近感 F に比例するものとする。

ここで、ロボットの動作設定プログラムのうち、組み込みプログラム、ティーチング、教師あり学習、教師なし学習、強化学習の占める割合をそれぞれ a_E , a_T , a_S , a_U , a_R とする。ただし、 $0 \leq a_E \leq 1$, $0 \leq a_T \leq 1$, $0 \leq a_S \leq 1$, $0 \leq a_U \leq 1$, $0 \leq a_R \leq 1$ であり、 $a_E + a_T + a_S + a_U + a_R = 1$ とする。

このとき、設計者の過失割合 r_D を、それぞれの重みをつけて $r_D = w_{DE}a_E + w_{DT}a_T + w_{DS}a_S + w_{DU}a_U + w_{DR}a_R$ のように線形結合とする。このとき、 $w_{DE} > w_{DS} > w_{DR} > w_{DU} > w_{DT}$ のような重みの順序関係が成り立つものと考えられる。すなわち、組み込みプログラムで稼働するロボットでは設計者の責任が最も重くなり、ティーチングでは設計者の責任は最も軽くなる。

育成者の過失割合 r_I を、それぞれの重みをつけて $r_I = w_{IE}a_E + w_{IT}a_T + w_{IS}a_S + w_{IU}a_U + w_{IR}a_R$ のように線形結合とする。このとき、設計者の過失とは逆に、 $w_{IT} > w_{IU} > w_{IR} > w_{IS} > w_{IE}$ のような重みの順序関係が成り立つものと考えられる。すなわち、ティーチングでは育成者の責任は最も重くなり、組み込みプログラムで稼働するロボットでは育成者の責任が最も軽くなる。

利用者の責任は、設計者および育成者の責任を除いた補集合となるが、基本的には製造物責任が問われ

ることになるので、利用者の過失割合は小さい。

6. 倫理的ロボットのための知識学習

6. 1. ビッグデータによる知識学習

5章では、ロボットの責任を設計者、育成者、利用者の責任の線形結合と定義した。このうち、設計者の責任については、さらに考察が必要である。AI手法を用いた学習を利用する際に設計者は、モデルの選択、パラメータの設定、検証データの選択、外れ値の除外、前処理などについて責任がある。しかし、教師あり学習、教師なし学習、強化学習の、いずれも、多くのデータを入力、出力として利用することで行動を決定するため、学習データの質を無視することはできない。学習データに問題がある場合には、必ずしも設計者の責任とは言い切れない。

人工知能の学習では、データとしてインターネット上のビッグデータを利用することが多い。Web2.0によりSNSなどの双方向の情報が増えると同時に、ビッグデータの正当性を評価することが難しくなった。深い洞察や知識に基づく意見と、思いつきの意見とが同じ重みを持つSNSの世界では、道徳的な正当性を評価することは困難である。

6. 2. 道徳的価値観の指標

ロボットの知識学習にビッグデータを利用するためには、前提としてデータの重要性に重みをつける必要がある。道徳的価値観に重みをつけるための要因を以下に挙げる。

6. 2. 1 身近さの要因

道徳的価値観の影響を表す第1の指標として身近さの要因がある。幼少期の道徳的価値観は主に身近な人の意見で醸成される。一般的には、身近な人からの意見ほど影響を受けやすく、関係が遠ざかるにつれ影響が小さくなると考えられる。たとえば、親、兄弟、先輩、友達、マスコミ、SNSの意見の順に、道徳的価値観の影響は小さくなる。

6. 2. 2 権威の要因

道徳的価値観の影響を表す第2の指標として権威の要因がある。道徳的価値観に限らず、多くの価値観は権威によって影響を受ける。一般的には、権威のある人の意見ほど世間に影響を与えやすく、個人的な意見は影響が小さくなる。たとえば、偉人、インフルエンサー、有名人、上司、先生の意見の順に、道徳的価値観の影響は小さくなると考えられる。

6. 2. 3 時間の要因

道徳的価値観の影響を表す第3の指標として時間の要因がある。SNSでは新しい情報を得ることができるが、ほとんどの場合、一過性の情報であり、普遍的な価値観として残ることは少ない。いっぽう、長年にわたり伝わる情報は道徳的価値観として影響を与えることが多い。

6. 3. 道徳的価値観の影響

6.2.1 から 6.2.3 節で示した身近さ、権威、時間の要因をそれぞれ、 M_F 、 M_A 、 M_T とおき、 $0 \leq M_F \leq 1$ 、

$0 \leq M_A \leq 1$, $0 \leq M_T \leq 1$ と数値化する。それぞれの指標が 0 に近いほど、道徳的価値観の影響が小さく、指標が 1 に近いほど影響が大きいものとする。

道徳的価値観の指標 M を重み w_{MF} , w_{MA} , w_{MT} による線形結合 $M = w_{MF}M_F + w_{MA}M_A + w_{MT}M_T$ と定義する。道徳的価値観の影響を受けるのは、 M が大きいときである。

7. 社会的制約に基づくロボットの行動規範と責任

単純な機械であれば、与えられた目的を達成するための機能のみを必要とするが、親近感 F の値が大きくなり、倫理感が問題視されるロボットでは、人間らしいふるまいとして適切かどうかの社会的制約に基づく動作が必要となる。

社会的制約には、法律、条例、道徳、慣習などがあり、条文として規制されているものから、時と場合や主観によって変化するものまで制約の許容範囲は様々である。たとえば、現実社会でも、SNS の投稿に見られるように、投稿者はいたずら程度の認識であっても、他者からは法律違反や慣習になじまないと見られて炎上することは多く見られる。ロボットの社会的制約は、年齢、性別、地域、民族、宗教などの多様性により、必ずしも組み込みプログラムで実現可能なものばかりではない。

本研究では、ロボットが行動規範として満たす必要性を社会的制約 S と定義し、 $0 \leq S \leq 1$ と数値化する。 S が 1 に近いほど、制約を満たす必要性が高いことを表す。法律、条例、道徳、慣習による制約を、それぞれ S_L , S_R , S_E , S_C とすると、一般的には、 $S_L > S_R > S_E > S_C$ のような順序関係が成立する。

社会的制約は、親近感が大きければ大きいほど、許容範囲の狭い制約となることから、ロボットの行動に対する制約 C ($0 \leq C \leq 1$) は $C = FS$ で定義できる。

8. 結論と課題

本研究では、日常生活において人とロボットが良好な関係を築くことをめざし、親近感の観点からロボットを分類し、倫理的責任の所在を明確にするとともに、動的な倫理知識学習の問題点を明確にした。親近感を表す指標として、ロボットの外形(形状、大きさ、装飾)、利用目的(利用者数、専門性)、複雑性(センサ・駆動系の自由度、入出力の関係性)の要因を定義することで、倫理問題が重要になるロボットハードウェアの属性を明確にした。ロボットの責任については主体を設計者、育成者、利用者に分けて、組み込みプログラム、ティーチング、学習の動作設定を利用した場合の責任の所在を明確にした。また、ロボットの設計者やロボット単体の中で閉じた挙動ではなく、社会の中で知識を学習するロボットが満たすべき行動規範と、行動規範に基づく責任問題について論じ、親近感により考慮すべき点が異なることを明らかにした。ロボットがビッグデータによる倫理的知識を獲得する際に、身近さ、権威、時間を要因とする道徳的価値観が重要であることを示し、道徳的価値観の要因によって重み付けをした上で知識を学習することが、倫理的ロボット実現には重要であることを明らかにした。分類を明確にするすることで、これまで議論にすれ違いが多かった倫理問題の論点を整理することができる。

本研究では、ロボットから人への倫理的課題を明確にすることを目的としたが、人からロボット、ロボットからロボットに対しても倫理的問題は存在する。親近感の要因を考慮して、これらの問題を明確にすることは今後の課題である。

引用・参考文献一覧

アイザック・アシモフ著, 小尾美佐訳(1983)『われはロボット』, ハヤカワ文庫.
久木田水生, 神崎宣次, 佐々木拓著(2017)『ロボットからの倫理学入門』, 名古屋大学出版会.
ルチアーノ・フロリディ, ラファエル・カプーロ, チャールズ・エス著, 西垣通, 竹之内禎編集訳(2007)『情報倫理の思想』, NTT 出版.
NEDO 編(2014)『ロボット白書 2014』, 新エネルギー・産業技術総合開発機構.

【謝辞】

本研究は, 日本大学経済学部グローバル社会文化研究センターの平成 30 年度研究プロジェクト C 研究として実施された. 日本大学経済学部の根村直美教授には, 本研究の契機を与えていただき, 多くの資料と議論で課題を明確にいただいた. また, 令和 2 年 12 月 3 日開催のワークショップ「ロボット倫理はどこに向かうのか」において, 日本大学経済学部の安田静教授には建設的な議論に参加していただいた. ここに厚く謝意を表す.

法的分野における責任の新たな類型
—「厳密な意味での行為者」としてのロボット—
A New Type of Responsibility in the Legal Field: Robots as Strict Agents

根村 直美

Naomi Nemura

〔概要〕

本稿では、ロボット法の第一人者 Ugo Pagallo がロボットの法的責任に関してどのような議論を行っているのかを明らかにするとともに、その議論が我々のこれまでの考察に対してどのような示唆を与えるのかを探った。

Pagallo は、当分の間は、ロボットが法的人格と位置づけられること、すなわち、権利義務を有する存在と認められることはないと考えている。しかしながら、民法の分野においてロボットを「厳密な意味での行為者」と認めることの必要性を唱える。Pagallo が行為者性を認めるようとするのは、その判断が人間に予見不可能なロボットが現れてきているからであるが、我々が玩具ロボットなどの、個人、あるいは、家庭内で利用するサービスロボットとの相互作用の中で生活しはじめていることとも深く関連していると考えられる。こうした Pagallo の議論は、＜人とのコミュニケーション＞が定義の中核に置かれるようなロボットに注目する本プロジェクトの議論とも呼応するものである。

キーワード：ロボット法, Ugo Pagallo, 責任, 厳密な意味での行為者

Keywords: laws of robots, Ugo Pagallo, responsibility, strict agent

1. はじめに

本プロジェクトでは、これまで、人工的道德行為者をめぐる議論は、「自由意志に基づいて行動する理性ある人間」(西垣, 2007) を道徳的責任を負う道徳的行為者とみなす西洋の倫理学の伝統を問い直すことで、ロボットに責任を負わせ道徳的行為者として位置づける可能性が開かれることを考察してきた。

ところで、法学分野においては、ロボットの法的責任について新たな類型を導入して考える必要性があることを説く論者がいる。その1人が、Ugo Pagallo である。本稿では、Pagallo がロボットの法的責任に関してどのような議論を行っているのかを明らかにするとともに、その議論が我々のこれまでの考察に対してどのような示唆を与えるのかを探ってみたい。

2. 法的責任の新たな類型

2. 1 法人格としてのロボットの可能性

Pagallo はまず、ロボットを「法人格」として認めるべきかどうかについて検討する (Pagallo: 2013,

155-166) ⁽¹⁾.

「法人格」の概念は、啓蒙主義とあいまってはじめて平等と権利享有という考え方と結びついた。その考えは、「すべての人間は、生まれながら自由で、尊厳と権利について平等である」という「世界人権宣言」へと至る。同様に、刑事手続きの改革と法典の体系化を通じて法の構造を合理的に説明しようとする試みは啓蒙主義の遺産である。

さて、Pagalloによれば、Lawrence Solum は、アメリカ合衆国修正第 13 条を、人工的行為者に対して正当に拡大適用することが出来るかどうかを検討している。Solum は、法制度がロボットに対して独立した法人格を与えるべきかを判断するために、弁証法的な方法で考察を進め、人工知能に権利を認めるという考えに対して提示されうる 3 つの反論を考慮に入れる。

その 1 つ目は、「人工知能は人間ではない」というものである。すなわち、「啓蒙主義の発想を参考にして、現在の法規制では、中世の偏見や迷信を克服し、法の領域における唯一の考えうるアクターとして最終的に人間だけを残した。そうだとすれば、どうして法制度が人間中心の立場を放棄するべきなのだろうか。ロボットに完全な法的人格を付与することの利益は何であろうか」と反論するのである。

2 つ目の反論は、「何かが欠けている」というものである。その反論は、「ロボットは、意識、意思、関心などの人格性にとって重要な要素の一部を欠く。現在の最先端の水準に照らし、結果としてロボットは、一連の刑法分野において帰責するための前提条件を欠く。(中略)人間は、深刻な精神疾患または感情的および知的未熟さを理由に、責任のない権利を享有することがある。これを基礎に、我々は、ロボットの人格性を子どもや精神障害者の権利になぞられるべきだろうか」と論じる。

3 つ目の反論は、「人工知能は財産であるべき」というものである。その反論によれば、「ロックの私的財産論に従えば、ロボットは人間の労働の成果であるため、製造者はロボットを所有する権利を有することになる」。かくして、その反論は、「ロボットを現代の奴隷と考えることができるのであれば、ロボットを解放する必要があるのだろうか」と問うのである。

これに対して、Solum は、ロボットの人格性を否定することができる法的理由や概念的原動力はなく、法は、合理的な選択と実証的証拠を根拠として人格性を付与すべきという。Solum は、先のような反論に対して、以下のように再反論をする。

まず、「人工知能は人間ではない」という反論については、Solum は、ロボットの行為に関する新しい形態の答責性は、その自律性のレベルにおいて、刑事裁判でロボットに有罪判決を言い渡すには不十分であることを認める。しかし、契約法分野においては関連する効果を生じさせる上では十分であろうから、今日の法制度における人間中心主義とも両立しうる、と論じるのである。

また、「何かが欠けている」という主張については、Solum は、もし普通の生活の中で、人が AI に直面する際、意図を有するシステムとして扱うことが実際的なのであれば、John Searle がいう「中国語の部屋」によって生じる直観は法的にはあまり意味がない、と述べる。Solum は、民法分野について考えてみる。すなわち、ロボットが、例えば意図的に損失を出さないという制約条件の下で、入札や均一分布の中から無作為に選択して注文する場合、ロボットは自分が何をしているのか本当は理解していない。その意味で、Searle の言うことは正しいかもしれない。しかし、法的な観点からみれば、重要なのはロボットの自意識ではなく、そうした機械で人間より優れた結果を残すことができるかどうかなのである。

さらに、Solum は、「欠けている」と言われている行為または属性を示すロボットを十分に想定することができるため、人間の独自性に基づいて「何かが欠けている」と主張することは単純に根拠に欠けると言う。自由意思が法的人格の前提条件であると主張する者がいる。しかしながら、Solum は、人間の自由意思に関して最もありうる物語は、意識的な理由づけと熟考という適切な行動を通じて生じた行動こそが自由であるというものであるが、この意味で、AI も自由意思を有する可能性がある、と論じるのである。

その他、「AI は意識をもてない」という議論に対しては、Solum は、ほとんどの学者が、自意識が法的人格の重要な前提条件であるとしているため、確かにこの点は重要であると言う。しかし、現在の人工的行為者がその能力をもたないことが事実であるとしても、将来のロボットがその能力をどの程度まで達成しうるのかは誰もわからない、と論じる。

Pagallo によれば、ロボットに法人格を与えるべきかどうかに関する Solum の考え方は、論理的というよりもプラグマティックな問題を提起する。つまり、Samir Chopra と Laurence White も言うように、「人工的行為者を法的人格と見なすことは、概して、発見というよりもむしろ決定の問題である。それは、人工的行為者の法的人格を否定するまたは付与するにあたっての最善の議論は、概念的ではなくプラグマティックな観点から行われるからである」。

そこで、Pagallo はロボットを法人格として認めるプラグマティックな理由の検討を行う。

まず単なる奴隷として扱われるロボットに対する倫理的逸脱を防ぐことを目的とする論者がいる。しかしながら、Pagallo は、そうした奴隷制を防ぐ手段として法的人格をロボットに認める必要はないとする。すなわちロボットに不正に損害を与えたり損壊したりする人間に対し、それらの法的人格に関わらず、法制度によって新たな犯罪類型を策定することができるのである。動物虐待事案のために確立した法制度と同様に、ロボットを悪用する人間を訴追可能にする立法は 1 つの解決策である。これは、ロボットに苦痛を感じる能力があることや感情を抱いていることを意味するものではない。むしろ、ロボットは「情報客体」として尊重や保護に値する道徳的価値の担い手と見なされるべきである。

また、ロボットの法的人格を認めることで、今日の法的枠組みの一貫性が向上すると主張する論者もいる。Pagallo は、人間のような自由意思、自律性、道徳的感覚が与えられた新世代ロボットが現実化すれば、ロボットに憲法上の人格権を認めることを宣言することも視野に入れて取り組むことは合理的ということになると認める。

しかしながら、意識、自由意思、人間のような意図といった、当事者に帰責するための一連の前提条件を欠いており刑事法廷においては責任を負わない (Pagallo: 2013, 48) ⁽²⁾、つまり、その犯罪行為に関する答責性を追わないロボットの場合においては、ロボットに憲法上の権利を認めることにより、今日の法制度の機能をいかにして改善できるのか疑問を呈する。プラグマティックな観点からするならば、近い将来ロボットの独立した法的人格が法的な課題になる可能性は低いと見るのである。

2. 2. 厳密な意味での行為者としてのロボット

2. 2. 1 契約

Pagallo は、当分の間、ロボットが独立した法的人格、すなわち、自身の権利義務を有するものとして認められることはほとんどありえないであろうが、ロボットの「厳密な意味での行為者性」を認め

ることの必要性を論じる (Pagallo: 2013, 166)⁽³⁾。契約法における答責性は、ロボットと安全に取引や相互行為を望む相手方の関心と、ロボットの自律性の拡大や行為予測不可能性によって損害を受けたくないという利用者および所有者の主張との間でバランスを取る方法として支持しうるのである (Pagallo: 2013, 166)⁽⁴⁾。

Pagallo によれば、「今後の目標は人間の活動において用いられる単なる道具としてのロボットと民事法の分野において機能するロボットを区別することにある」 (Pagallo: 2013, 82)⁽⁵⁾。Pagallo は、人間の相互作用を補助する単なる道具としてのロボットと法制度における厳密な行為者としてのロボットとを区別すべきであると考えているのである。

ここで Pagallo が想定しているのは、ロボ・トレーダーの事例である。Pagallo によれば、ロボ・トレーダーの事例では機械自身が契約内容や条件を決定する能力を有することが示されている (Pagallo: 2013, 112)⁽⁶⁾。そうしたロボットは、オークションにおいて応答し、申し出を受け、見積りを要求し、取引を交渉し、さらに契約を履行することができる (Pagallo: 2013, 82)⁽⁷⁾。そのため、Pagallo は、自律のレベルは刑事責任能力を満たすには不十分としても、契約法における新たな形態の人工的行為者として認めるには十分であると考えているのである (Pagallo: 2013, 82)⁽⁸⁾。

しかしながら、現在のルールは、一定の事例においてロボットの法的行為者性を承認することを禁止しており (Pagallo: 2013, 82)⁽¹⁰⁾、このような機械の認知能力に起因する問題やこの種のロボットが原因になって生じた法的責任を適切に解決することができない (Pagallo: 2013, 112)⁽¹¹⁾。かくして、Pagallo は、「法的行為者性」は、人間がロボットに対して重要な認知作業を委任する際に重要になると言う (Pagallo: 2013, 82)⁽¹¹⁾。Pagallo は、その目標は、国連が言う「人類の幸福にとって有益なサービス」を提供するロボットの利用または製造を躊躇させるような立法の回避であると論じているのである (Pagallo: 2013, 82)⁽¹²⁾。

Pagallo によれば、ロボットが自らの動作に責任があるとする考えは、私有財産に関するローマ法に先例がある (Pagallo: 2013, 82, 103-106)⁽¹³⁾。ユスティニアヌス法典の「学説語彙」では、私有財産制度としての奴隷制度が認められているが、そこでは、私権の基礎たる法人格は与えられていないものの、不動産管理人、銀行家や商人として働くことを認めている (Pagallo: 2013, 82)⁽¹⁴⁾。そして、家長の法的責任は奴隷が保有する特有財産価値を限度とするが、契約の相手方はその交渉が当該奴隷の有する権限または経済的な自律性の範囲を超えるものか否かを確認できる (Pagallo: 2013, 104)⁽¹⁵⁾。Pagallo によれば、ロボットは、「デジタル特有財産」 (Pagallo: 2013, 103)⁽¹⁶⁾の理論を用いることで、ロボットの認知能力に起因する問題やこの種のロボットが原因になって生じた法的責任に適切に対応できるようになるのである (Pagallo: 2013, 112)⁽¹⁷⁾。

2. 2. 2 不法行為

Pagallo はまた、「契約外の責任」、つまり、「コモンローの法律家が不法行為と定義する分野」についても検討している (Pagallo: 2013, 115)⁽¹⁸⁾。

Pagallo が念頭に置くのは、社会生活の中で「活動を仲介するロボット」である (Pagallo: 2013, 112)⁽¹⁹⁾。そうしたロボットは、家事、娯楽、障害者支援、個人の移動など、家庭内の利用または個人的利用に関わるものであり、玩具ロボット、子守ロボット、AI 運転手などがその例である (Pagallo: 2013, 113)⁽²⁰⁾。

Pagallo は、この分野の法的責任は、子どもやペット、もしくは従業員のような他者の行為に関する伝統的な責任形態、すなわち、厳格責任と対比できるが、その伝統的な責任形態は新しい責任で補われる必要があると考える (Pagallo: 2013, 115)⁽²¹⁾。

Pagallo は、不法行為の分野においても、デジタル特有財産のような制限付きの法的責任を承認することの可能性について述べる (Pagallo: 2013, 137-138)⁽²²⁾。すなわち、その制度によって、ロボットに対して個人の答責性を付与すれば、第三者への危害および現実の損失損害について直接ロボットに法的責任を負わせることができると考える。Pagallo は、家庭内あるいは個人的に利用されるサービスロボットが第三者に対して危害および現実の損失を発生させたことについての法的責任をロボット自身に負わせることを提案するのである。

2. 3 保険システム

Pagallo は、新たな法的責任の種類としてロボットの行為者性を主張するが、その際には、保険モデルなどの政策を導入する必要性も説いている (Pagallo: 2013, 104)⁽²³⁾。

保険モデルとは、「ロボティクスに関する民事法準則についての委員会への勧告を附帯する 2017 年 2 月 16 日欧州議会決議」⁽²⁴⁾ で提案されたモデルを思いうかべればよいであろう。その決議では、「そのロボットによって引き起こされる可能性のある損害に保険をかけることをロボットの製造者又は所有者に要求する強制保険スキームを創設すること」が求められているのである。

3. Pagallo の議論をめぐって

Pagallo が日常生活において「行為者」と認めるべきとするのは、玩具ロボットや子守ロボットなど、家庭内や個人的な利用のためのサービスロボットである。Pagallo によれば、これらのロボットは、「愛情を与え、子どもや老人の世話をする」 (Pagallo: 2013, 5)⁽²⁵⁾。

そして、これらのロボットは、「産業用ロボットのように、最初から決められた行動だけを行い、予想外の行動をとらないほど良いとされるような従来型の機械とは異なる」 (平野, 2019: 187)。平野晋は、Watson を例にとる (平野, 2019: 187)。Watson のように、インターネット上の情報を収集・分析して回答を出すようなシステムは、インプットされる情報が刻々と変化し予見不可能であるから、その解答も予見不可能である。つまり、Watson はそもそも予見不可能に設計されているのである。

平野は、ロボットの〈考え・判断〉はヒトには予見不可能であるから、たとえその〈制御不可能性〉によってロボットが人間に危害を加える事態が不可避免的に生じても、その〈不透明性〉によって、設計者、開発者、または製造責任者に責任を課すことができなくなることを危惧する。

これに対して、Pagallo は、法が個人用および家庭用ロボットサービスの使用を妨げることのないように、家庭内や個人的な利用のためのサービスロボットを民事法分野において新たな行為者とみなすことを提言する。この主張は、サービスロボットと人間との相互作用が我々の社会に深く浸透しつつあることに拠って立っていると見ることができるであろう。

予見不可能性を前提としながらも、「親近感」の度合いが高いサービスロボットが存在感を増しつつある今日においては、Pagallo のような法的なアプローチは、ロボットの道徳的な「行為者性」、そして、「ロボットが守るべき倫理」を考えるにあたって重要な示唆を与えてくれると思われるのである。

4. 結び

Pagallo は、当分の間は、ロボットが法的人格と位置づけられること、すなわち、権利義務を有する存在と認められることはないと考えている。しかしながら、ロボットを「厳密な意味での行為者」と認めることの必要性を唱える。Pagallo は、民法分野において機能するロボットは、単なる道具としてロボットとは異なり、その自律のレベルについて人工的行為者として認めることができると考えるからである。

Pagallo が<行為者性>を認めるようとするのは、その判断が人間に予見不可能なロボットが現れてきているからであるが、我々が玩具ロボットなどの、個人、あるいは、家庭内で利用するサービスロボットとの相互作用の中で生活しはじめていることとも深く関連していると考えられる。Pagallo は、ロボットに対して<行為者性>を認めることにより、国連が言う「人類の幸福にとって有益なサービス」を提供するロボットの利用または製造を妨げるような立法を回避できると考えている。すなわち、ロボットの「行為者性」を認めるのは、プラグマティックな観点から是認されるアプローチであるが、そのプラグマティックな観点とは、すでに我々との生活において相互に作用し合うようになっているそれらのサービスロボットが提供され続けることを目指すものなのである。

こうした Pagallo の議論は、道徳的行為者をめぐる本プロジェクトの議論とも呼応するものである。Pagallo が特に注目するのは、日々の生活の中に入り込みつつあるようなロボットであるが、それは、本プロジェクトの議論が考察の中心に置いた「人とのコミュニケーション」が定義の中核に置かれるようなロボットと重なると言えよう。また、Pagallo は意識、自由意思、人間のような意図といった、当事者に帰責するための前提条件を欠いていても行為者性を認めることは可能だと考えるが、本プロジェクトの議論もまた同様である。

本プロジェクトが示したような知見は、法学分野など他の分野の動きと連動しつつ、今後ますます新たな潮流となっていくであろう。

注

- (1) 日本語訳は、松尾剛行・工藤郁子・赤坂亮太の邦訳『ロボット法』（勁草書房、2018年）を参考にした。邦訳では、p. 178-190.
- (2) 邦訳では、p. 54.
- (3) 邦訳では、p. 190.
- (4) 邦訳では、p. 191.
- (5) 邦訳では、p. 91.
- (6) 邦訳では、p. 131.
- (7) 邦訳では、p. 92.
- (8) 邦訳では、p. 92.
- (9) 邦訳では、pp. 91-92.
- (10) 邦訳では、p. 131.
- (11) 邦訳では、p. 92.
- (12) 邦訳では、p. 92.
- (13) 邦訳では、p. 92, pp. 120-123.

- (14) 邦訳では, p. 92.
- (15) 邦訳では, p. 121.
- (16) 邦訳では, p. 120.
- (17) 邦訳では, p. 131.
- (18) 邦訳では, p. 133.
- (19) 邦訳では, p. 131.
- (20) 邦訳では, p. 132.
- (21) 邦訳では, p. 133.
- (22) 邦訳では, p. 159.
- (23) 邦訳では, p. 120.
- (24) この決議については, 栗田昌裕「AI と人格」(山本龍彦編『AI と憲法』所収, 日本経済新聞社, 2018 年) を参考にした.
- (25) 邦訳では, p. 6.

引用・参考文献一覧

- 平野晋 (2019) 『ロボット法—AI とヒトの共生にむけて—〔増補版〕』弘文堂.
- 栗田昌裕 (2018) 「AI と人格」山本龍彦編『AI と憲法』日本経済新聞社, 2018 年, pp. 201-247.
- 西垣通 (2007) 「普遍倫理への模索」ルチアーノ・フロリディ『情報倫理の思想』(西垣通・竹之内禎編著訳), NTT 出版, pp. 7-45.
- Pagallo, Ugo (2013) *The Laws of Robots: Crimes, Contracts, and Torts*. Dordrecht: Springer
〔邦訳: ウゴ・パガロ (松尾剛行・工藤郁子・赤坂亮太訳) 『ロボット法』勁草書房, 2018 年〕

Abstract

In the context of applied ethics, the need for robot ethics discussions has already begun to be recognized in the latter half of the 2000s. However, such discussions have been considered a form of thought experiment. Recently, limited autonomous robots have started playing an important role in various areas of our societies. Furthermore, the concept of “singularity” has become more likely. This collaborative study clarifies the contemporary discussions on robot ethics, which are progressing in the abovementioned situation, and investigates how to think in the future mainly from the standpoints of ethics and informatics. The following three papers show the findings of our investigations

The first paper entitled “A Study of Philosophical and Ethical Perspectives on Robot Ethics” defines a robot as an artifact that can communicate with humans using language, gestures, or computer graphics. Such robots are already in use in numerous fields, and they are on their way to penetrating every aspect of human life. This paper also examines contemporary philosophical and ethical arguments, such as Luciano Floridi’s information ethics. This investigation reveals that the concepts of moral agent and moral patient, which human societies have assumed, are not based on the existence of a specific substance but rather constructed in the concrete situation through the interaction of all entities containing both non-organisms and organisms. Furthermore, two questions are asked here: “Is a robot a moral agent?” and “Is a robot a moral patient?” The answers to both questions would depend on how a robot relates to its environment, which includes humans and their societies. Thus, this paper demonstrates that a robot can be considered a moral patient because it is becoming increasingly adept at communicating with human beings.

The objective of the second paper entitled “Ethical Responsibility of Robots Based on Intimacy” is to classify robots in terms of intimacy and quantify ethical responsibility based on the classification to realize a better relationship between robots and humans. Particularly, ethical issues are discussed to control robots using machine learning. Shape, purpose, and complexity are defined as the factors of intimacy for determining the contribution of ethical responsibility. The weight of responsibility is determined by factors of embedded program, teaching, and learning according to designers, operators, and end users. The responsibility and behavior rules of robots that acquire knowledge in a social environment depend on the intense of intimacy. It is important to consider the ethical values caused by affinity, authority, and establishment when robots obtain knowledge through big data. Ethical issues of robots-to-human and robots-to-robots need to be researched further.

In the third paper entitled “A New Type of Responsibility in the Legal Field: Robots as Strict Agents,” the discussion by Ugo Pagallo, a leading expert in laws of robots, about the legal responsibility of robots and the suggestions regarding our considerations are explored. Pagallo believes that robots will not be considered legal persons for the time being, that is, they will not be recognized as having rights and obligations. However, he advocates the need to recognize robots as “strict agents” in the field of civil laws. Pagallo attempts to recognize the agency because humans are increasingly interacting with service

robots for personal and domestic use whose judgments are becoming unpredictable to humans. Pagallo's discussion is believed to be similar to our project's discussion, in which the robot's definition focuses on the "communication with humans."

Robots that communicate and interact with humans are expected to become increasingly indispensable to humans in the future. Furthermore, it is expected that the results of this study will become a new trend in the near future, not only in ethics and informatics but also in the related fields such as law. In that sense, this research can be considered a pioneering study. In addition, the results of this study will contribute to the development of a better interaction between humans and robots that are currently permeating in our lives.

日本大学経済学部グローバル社会文化研究センター
ワーキング・ペーパー・シリーズ No.2021-03

2022年3月1日 発行

発行元 日本大学経済学部グローバル社会文化研究センター
〒101-8360 東京都千代田区神田三崎町1-3-2

TEL 03-3219-3309 / FAX 03-3219-3329

URL : <https://www.eco.nihon-u.ac.jp/research/cgs/>