

# 博士論文(要約)

論文題目

ソーシャルロボットの倫理の基礎づけ

——道徳的行為者性の虚構的解釈による人間中心的枠組みの構築——

氏 名 水上 拓哉

## 論文の概要

本論文で私は、ソーシャルロボットの倫理を考えるための哲学的基礎を提供するために、ソーシャルロボットの道徳的重要性をよりよく理解するための理論的枠組みを構築することを試みた。特にここで目指したのは、ソーシャルロボットのエージェンシー（あたかも行為主体であるかのようにユーザに現れること）を、ポストヒューマニズム的な行為者性解釈に頼らず、むしろ人間の相互行為を中心に据えて説明することである。

現在、ソーシャルロボットの社会浸透について様々な倫理的課題が指摘されている。技術が「他者」として現れることをどのように受け止めるべきかについては、ユーザ側だけでなく設計者側も混乱の中にあるといえる。

他方、技術哲学においては、技術の役割や本質を理解するための理論的枠組みの構築が様々な角度から試みられてきた。現代の技術哲学は、技術倫理に接近しつつ、技術の本質の説明（記述）だけでなくその倫理的設計のための道筋の提案（規範）も重視するようになってきている。また、特に最近では日進月歩で進歩する技術の影響力の増大を背景に、人間中心で語られる倫理学を批判するポストヒューマニズム的な立場が影響力をもっている。

ソーシャルロボットは「他者」として現れ社会的関係を代替する技術であるため、その倫理も一見このようなポストヒューマニズム的な技術哲学で考えることができるように思われる。しかし、本論文で見ていくように、人間中心主義を脱却する形で技術の道徳的地位を再解釈することが、かえってソーシャルロボットの道徳的影響力の理解を妨げてしまうのだ。ひいてはこれが、ソーシャルロボットの倫理的設計のために誰がどのように行動していくべきかも曖昧なものにしてしまうのである。

私はソーシャルロボットの倫理を考えるための理論的基礎を提供するため、ポストヒューマニズム的な行為者性解釈の問題点を批判的に継承しつつ、記述的研究と規範的研究が（見かけだけではなく実態として）両立するようなパースペクティブを提案した。ここで私が提示したのは、ソーシャルロボットの道徳的地位を虚構的真理を生み出す「小道具」として理解する道であった。あるソーシャルロボットの道徳的行為者性は、そのロボットについてユーザが見出す虚構的真理のひとつとして考えることができる。ソーシャルロボットの道徳的行為者性が関係的な性質をもつことは確かだが、それは（道徳的責任の帰属の基礎となるような）客観的な性質は備えておらず、むしろユーザの想像力に基づいた動的・一時的に見いだされる虚構的真理なのである。本論文の最後では、このパースペクティブに基づき、ソーシャルロボットの設計者が倫理的評価をどのような形であれば遂行できるのかを提案し、設計者の道徳的責任の範囲と内実を明らかにした。「小道具説」の提供するパースペクティブによって、設計の各ステークホルダーがロボットの道徳的影響力を「適度に」怖がりつつ、その社会浸透を前向きに目指すことが可能になる。

## 各章の要約

本論文は3部構成（結論を含めると4部構成）となっている。第1部では現在の技術哲学および技術倫理において影響力をもっている関係的転回の、ソーシャルロボットの倫理を考える上での可能性と限界について検討した。関係的アプローチは、技術の役割を技術単体の能力ではなくそれがユーザとの

関係性において果たしている機能に基づいて説明しようとするものである。

第 2 章では技術的人工物の道徳的行為者性を論じる先行研究を、これらに共通して見られる構造を描出することで再整理し、その整理に基づいてソーシャルロボットの行為者性を記述するためには関係的なアプローチが少なくとも視座としては重要であることを述べた。私はここで、それぞれの議論が概ね、1.道徳的行為者性に必要な性質のラインナップ、2.そのそれぞれの性質の内実、3.道徳的行為者性の位置づけの 3 つの論点について何かしらの前提をもっていることを明らかにした。本章の後半では、前半の論点整理に基づき、ソーシャルロボットの行為者性を問う際に出てくる論点を洗い出しつつ、その道徳的影響力が相互行為的・関係的に構築される可能性を示唆した。

続く第 3 章では、ソーシャルロボットの行為の記述に关系的視点が要請されることを受け、それを道徳的行為者性の解釈そのものにまで適用するべきなのかについて検討した。ここでは Coeckelbergh の「仮想的道徳的行為者性」概念と Verbeek の媒介理論に着目し、本論文の目的から考えた妥当性を検討した。私はここで、彼らのようなフラットな行為者性解釈がもたらす道徳的責任についての概念的混乱が、技術の倫理的設計の道を提示することを難しくさせている点を指摘した。

第 2 部では私は「小道具説」という立場を打ち出し、ソーシャルロボットのエージェンシーを人間中心で説明する枠組みの構築を行った。この立場の基本路線は、ソーシャルロボットの道徳的地位を「小道具」として理解した上で、ロボットとユーザの関係性の中で生まれる道徳的行為者性はあくまでそのロボットについての虚構的真理のひとつである、というものだ。

第 4 章では、「小道具説」を展開するための予備的検討を行った。私が目指すのは、关系的な行為者性解釈を批判的に継承しつつ、なお人間中心主義的な記述理論を作り上げることである。それを実現するための道として、本章の前半ではそのためにソーシャルロボットの倫理が技術倫理の中でどのような特殊性をもつのかを考察した。ソーシャルロボットは、私たちが独占していた(とされている)心的能力や自由意志、ひいては道徳的行為者性を自らも備えているかのように見せる技術であり、そこにソーシャルロボットの倫理を考える難しさがある。しかし、ソーシャルロボットに対して特定の道徳的地位や属性を見出しているのはあくまで私たち人間側のアクターであり、ソーシャルロボットが能動的にそうさせているわけではない。ソーシャルロボットがエージェンシーをもち道徳的重要性をもつ過程そのものは、基本的に人間の想像活動を中心に据えて説明することが可能なのだ。後半では、もうひとつの予備的検討として、技術の行為者性をめぐる論争における「道具説」の批判的検討を行った。本章の最後では技術的人工物をソーシャルロボットに置き換えるなど、Johnson の図式を修正を試みた。

第 5 章ではソーシャルロボットが「小道具(props)」である、という着想のもと、ソーシャルロボットのエージェンシーを「虚構的行為者性」という表現を用いて人間中心主義的に解釈する立場を展開した。私はここで、前章で用意した「三極の行為者性」のモデルの修正版をもとに枠組み構築を行った。「三極の行為者性」をソーシャルロボットの事例に当てはめるとき、もとの枠組みでは「技術的人工物」に相当するアクターは、実はロボットそのものではなく、相互行為において生み出されたそのロボットの虚構的な像なのだ。そして、この虚構的な像というものは、実際の道徳的責任の帰属の議論において参照できるような「一人前の」アクターではない。私はこの考え方を、Walton のフィクション論を手がかりに説明した。Walton のフィクション論を手がかりにソーシャルロボットの道徳的地位を小道具として特徴づけることで、

ソーシャルロボットの行為者性や道徳的重要性を、ユーザの心象やその心象の構成過程を中心に据えて分析することが可能になる。「小道具説」で描かれるのは、ソーシャルロボットの行為者性の、ユーザの想像力やそのベースとなる「モチーフ」によって動的に、かつ一時的に構成される性質である。この点が、「実質的」という表現を用いて道徳的行為者性を客観的・静的なものとして規定する Coeckelbergh の「仮想的行為者性」とは決定的に異なる。

最後の第3部では、第2部で展開したソーシャルロボットの倫理の記述的研究を踏まえ、ソーシャルロボットの倫理的設計を実践していく道について考察した。

6章では、私の小道具説に基づいて、ソーシャルロボットの設計の倫理における設計者の道徳的責任の範囲と内実について検討することで、ソーシャルロボットの倫理的設計の実践の道を探求した。私はここで Walton のフィクション論に基づき、あるソーシャルロボットについて「公認の遊び」と「非公式だが自然な遊び」の2つのタイプの遊びがあることを指摘し、設計者は自分が公認する遊びではなく、むしろ自然な遊びがどのようなものであるかについて倫理的評価を行う義務があることを示唆した。

最後の第7章では、前章の「パラダイムシフト」に基づいて、設計者が具体的にどのような形でソーシャルロボットの倫理的評価を行うことができるのかを検討した。ここでは、ソーシャルロボットの倫理的評価を、従来の工学的な評価スタイルから小道具の評価スタイルにシフトさせる方法について提案した。雑談対話システムのようなソーシャルな役割を担うプログラムの研究においては、しばしばその性能評価が他の自然言語処理の分野に見られるようなスタイルで行われていた。そこではしばしば、客観的で、何らかの数値において「最先端」で、設計者の開発意図の実現に志向した評価実践が行われていた。これに対して、小道具の評価実践のスタイルは工学的なそれとはちょうど対照的なスタイルで行われることが多い。私は小道具の評価スタイルを、おもちゃの評価実践と文学作品の評価実践の分析に基づいて特徴づけた上で、このスタイルこそが、ロボットそのものではなくそのロボットを媒介にした「遊び」を評価するための方法論として示唆的であることを指摘した。