

Wendell Wallach and Colin Allen, *Moral Machine: Teaching Robots Right from Wrong*,
Introduction 要旨

久木田水生

段落 1 ~ (“In the Affective...”)

近年、人間の表情を読むロボット、金融取引を行う世界的なコンピュータのネットワーク、高齢者の介護をするロボット、武器を運ぶロボットなどが開発・実用化されている。一方で人間の活動を促進し、モニターし、分析をするコンピュータチップや「ボット」があらゆる機器、あらゆるバーチャルな環境に配備されている。これらのロ-ボットによって集められた情報が商業的、政治的、医療的目的に利用されている。

これらの発展は人間の直接的な監視から離れた自立的なロ-ボット、そして人間の福祉に影響を与える可能性があるロ-ボットの創造に収束している。

今日、ロボット工学に従事するエンジニアたちは、彼らの作るシステムが人間にとって有益であり、人々に危害を与えることがないことを保証するという課題に直面している。

段落 4 ~ (“Within the next...”)

数年以内に、人間の監視から独立に意思決定を行うコンピューター・システムによって深刻な事件が起こるだろうと著者たちは予測している。

安全性と社会的利益はこれまでも工学の最優先課題だったが、今日のシステムは余りに複雑で、システム自体が道徳的決定を下さなければならないレベルに達している。これは道徳的行為者の枠を人間を超えて人工知能にまで広げなければならないことを意味する。このような行為者を著者たちは人工道徳的行為者(artificial moral agents; AMAs)と呼ぶ。

段落 6 ~ (“We don’t know...”)

破滅的な事件の可能なシナリオの一つは、金融取引システム、安全保障システム、警備システムなどの誤作動によって生じる事故、混乱である。それらは何百人の人命の損失、何十億ドルの損失を引き起こしうる。

段落 14 (“Time may prove...”)

著者たちの目標は AMA を設計するという工学的課題を建設的にガイドするような仕方で、議論の枠組みを与えることである。上述した予言の目的は恐怖をあおるためではなく、道徳的機械の構築にいま取り組む必要があるということに注意を促すことである。

段落 15 (“The field of machine morality...”)

機械道徳の分野はコンピュータ倫理の分野を、人がコンピューターを使って何をするかという関心から、コンピューター自身が何をするかという疑問へと拡張する。著者たちはコンピューターを明示的に道徳的な推論を行うもの(explicit moral reasoner)に変えることに関する技術的な問題について議論する。人工知能が自律的行為者の範囲を拡張するにつれて、人間が道徳的行為者に要求する価値や法の広範な体系が人工的行為者に適用されるようにそれらをどのように設計するかという課題は、ますます緊急のものになっている。

段落 16 (“Does humanity...”)

私たちは道徳的な判断をコンピューターに任せるべきではなく、重大な危険を引き起こしかねないテクノロジーを発展させることをやめるべきだという意見もある。しかし市場や政治の圧力がそのようなテクノロジー

によってもたらされる利益を要求するだろうと著者たちは考える。従ってこのテクノロジーに關与する人は誰でも、コンピューター、ロボット、そしてコンピューター・ネットワーク上のバーチャル「ボット」に道徳的決定を実装する課題に正面から取り組まなければならない。

著者たちの見解では、自律的システムへの依存の増大が人々の基本的な人間性を劣化させることにはならないし、高度なロボットが人間を支配することにもならない。人間は常にテクノロジーの生産物に適應してきたし、自立的機械を身の周りに置くことの利益はそのコストを上回るだろう。しかし悪い自律的人工的行為者を生み出してしまうことを避けるためには、人工的行為者を良くするためには何が必要かについて十分真剣に考える心構えをしておかなければならない。

段落 19 (“In proposing...”)

著者たちは、AMA の夜明けを予言しているわけではないし、人間がロボットに道徳的助言を求めるようになるだろうと言っている訳でもない。著者たちの関心は、道徳的意思決定の能力の必要性を示唆する現在のテクノロジーからの前進的ステップに向けられている。完全な人工知能には及ばずとも、エンジニアが直面する現実的な問題があり、そしてそれはエンジニアだけでは対処することができない。

段落 20 (“Is it too early...”)

この話題を持ち出すのは早すぎるだろうか？ そうではない。産業用ロボットによる事故はすでに起こっている。2010 年には家庭用ロボットとサービスロボットの市場は産業用ロボットの市場の 2 倍に、2020 年には 4 倍になるだろう。家庭用ロボットとサービスロボットの登場で、ロボットはもはや、訓練を受けた労働者だけがロボットと接触する制御された産業的環境に限定されない。ペットロボット、荷物運びロボット、家事ロボット、介護ロボット、金融取引プログラム、等。SF に描かれるような人工知能は当分実現しないだろうが、人工的なシステムが人間の生活に影響を与える決定、そして倫理的な選択肢を持つ決定を下す地点まで私たちはすでに到達している。人間に等しい道徳的能力を持つ意識的なシステムは永久に作られないかもしれないが、より限定的なシステムはすぐに作られるだろう。そのようなシステムは彼らの行動の倫理的な選択肢を評価する能力を持つだろう。

段落 23 ~ (“The task of...”)

AMA の設計という課題は、倫理学理論に真剣に取り組むことを要求する。宗教的哲学的価値を機械に応用するのは容易ではない。規則ベースの倫理体系は比較的容易にコンピューターに埋め込めるように思われるかもしれないが、アシモフの小説に見られるように、非常に少数の単純な法則でも多くの倫理的ジレンマを引き起こしうる。規則よりも人格を重視し、良い人格を発達させることを目的とするアリストテレスの徳倫理学については、なおさらコンピュータに応用することが難しい。これらの工学的困難に向き合うためには、進化、学習と発達、神経心理学、哲学の分野で研究されている、人間の道徳性の起源に目を向けることが必要である。

機械道徳はまた人間的な意思決定についての問題でもある。AMA を作ることについて考え、そのための実験を行うことで、私たちは、人間がどのように機能しているか、人間のどの能力が機械に実装できるか、そしてどのような性質が人間と動物、あるいは人間が創造する新しい形態の知性とを隔てるのか、といった問題について考えざるを得ない。AI が心の哲学において新しい探求の方針を喚起したのと同様に、機械道徳は倫理学において新しい探求の方針を喚起する可能性がある。

段落 25 ~ (“Three questions...”)

ここまでの議論から三つの疑問が生じる。

1. 世界は AMA を必要としているのか？
2. 人々はコンピューターが道徳的意思決定を下すことを望むのか？
3. 道徳的な意思決定を下すコンピューターが必要だと人々が信じているとしたら、エンジニアと哲学者はどのようにして AMA を設計するべきか？

第1章と第2章は第一の疑問に関わる。著者たちはそこで AMA が不可避であること、そして道徳決定の能力が単純なものでも現実的な困難を提示することを論じる。そのようなシステムの設計者がどのような目的を設定するべきかを特定することが、かなり重要な問題である。

第2章では、AMA の洗練の段階的發展を理解する枠組みを提示する。その際に、自律性と、道徳的に重要な事実に対する感受性という二つの次元が強調される。これらの次元において程度の低いシステムは「操作的道徳」を持つと言われる。機械が洗練されるにつれてそれはある種の「機能的道徳性」持つようになる。

第3章では、上の第2の疑問が扱われる。特に、AMA は人々が道徳的責任を放棄して機械に押し付けることにならないか、という懸念は切実である。また破滅的な事故の可能性と新しいテクノロジーの提供する利益を勘案するリスク評価の問題も未解決である。

第4章では、意識や感情を欠いた「単なる」機械が道徳的行為者になれるのか、という問題を扱う。操作的道徳と「真の」道徳の間には大きな空間がある。これが2章で論じた機能的道徳のニッチである。第4章の目的は AI における現在の成果が、様々な応用のための AMA を作るのに必要な特徴を特定するのに適しているか、という問題を論じることである。

一般的な人工知能の問題を扱った後、特定の実装の問題へと目を転じる。第5章ではエンジニアと哲学者がお互いに何を提供すべきかということ概観し、また AMA の設計に対するトップダウンアプローチとボトムアップ（あるいは発達の）アプローチを描写する。第6、7章はそれぞれ義務論理、帰結主義にパラレルなトップダウンアプローチと、学習的、発達の、進化的方法を応用するボトムアップアプローチを詳しく描写する。

第6、7章での議論から、トップダウンとボトムアップのアプローチの区別が単純過ぎることが明らかになる。エンジニアはこの二つのアプローチを組み合わせる必要がある。トップダウンアプローチの困難は徳倫理について議論することを動機づける。第8章では AMA のための徳倫理が論じられる。

第9章ではコンピュータによる道徳的意思決定の開発のために利用されているソフトウェアツールをサーベイする。

トップダウン、ボトムアップアプローチは倫理における推論能力の重要性を強調する。しかし近年の道徳心理学における研究は、人間の道徳的意思決定における感情、社会性、意味論的理解、意識など、合理性以外の能力を強調する。第10章では、コンピュータやロボットに合理性以外のこれらの能力を提供する近年の研究を論じる。第11章では一つの機械の中に合理的能力と非合理的能力を組み合わせるための特定の枠組みを提示する。

第12章では上記の第2の疑問に立ち戻る。ここでは公共政策や社会的経営的責任 (liability) のメカニズムを通じてどのように危険をモニターし、管理すればよいかということについての提案をすることをもくろむ。

エピローグでは、AMA を設計するというプロジェクトがどのようにして、道徳的行為者としての人間自身についての理解に、そして道徳的理論そのものの本質についての理解にフィードバックされるか、ということ簡単に論じる。