

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

Chapter 5 “Philosophers, engineers, and the design of AMAs”

要旨

久木田水生

2012 年 7 月 6 日

Two Scenarios

P. 73– (“Scenario A:.....”)

哲学者は非常に抽象的な原理でものを考える傾向があるのに対して、工学者は実際の設計課題を達成しなければならない。この両者の間には大きな溝がある。しかし AMAs の開発において哲学者が果たす役割はある。一般的な原理は設計に指針を与える（ただし一般原理だけでは十分ではないが）。

訓練を受けた倫理学者と哲学者が AMAs のデザインに対して、どのような貢献ができるだろうか？ 倫理的原則、理論、枠組みは、ある程度自律的に行動する計算的システムの設計に指針を与えるのに有用だろうか？あるいは倫理学者の貢献は主にその課題の複雑さを強調することで、工学者には役に立たないように思えるものだろうか？

Grounding the Collaboration

P. 74– (“Over the past half century.....”)

過去半世紀の間、哲学者の中には AI を熱狂的に支持するものもいれば、楽観主義的な見方を批判するものもいた。AI に批判的な哲学者の方が工学者の直面する問題に対してより良い判断をしていたことが分かった。「強い AI」の実現に対する計算的なアプローチの限界についての哲学者の批判は今でも有効である。しかしこれは私たちが「機能的道德性」と呼んだものを高いレベルで持つシステムを開発するという、「弱い AI」の課題にとっては関係がない。そこで本章とその後の 4 つの章では哲学的な問題は置いておいて、現在利用可能なプラットフォームに、どのようにして倫理的な考察が導入されるかに焦点を当てよう。

倫理的に行動するロ-ボットにおいて、倫理理論はどのような役割を果たしうるだろうか？工学者は、道德的意思決定をするシステムに対する情報的要件がどのようなものであるかを決定する必要もまたある。

このとき倫理学者は実際にどのように役立ちうるだろうか？良く訓練された倫理学者は道德的ジレンマの複雑性を認識するよう教わっており、AMAs が直面するだろう課題をカバーすることを意図された、どの一つのアプローチも十分でないことを敏感に理解するだろう。一方で工学者は、道德的考慮に対して敏感なシステムにしたいという倫理学者の願望は、ただでさえ困難な仕事をますます難しくする、と懸念するだろう。倫理的ジレンマの複雑性と扱いづらさについての理論的な議論は役に立つとは見なされないだろう。

P. 75- (“Here it is useful.....”)

ここで工学倫理の分野を参考にするのが有用だろう。そこでは哲学者はしばしば自分たちの学問領域が工学者にとって関係がありかつ利用できるものにするという問題について考えているからだ。科学者と工学者に倫理学を教えるためのアプローチを概説する際に、Caroline Whitbeck は Stuart Hampshire が立てた「判事の視点 a judge perspective」と「行為者の視点 an agent perspective」という区別を持ちだしている。Whitbeck は判事の視点を伝統的な倫理学への哲学的アプローチと同一視する。すなわち抽象的な原理を特定の事例にあてはめるというアプローチである。ここでは倫理的な課題 ethical challenges は、対立する陣営あるいは原理の対立として説明され、倫理的な判断は互いに排他的でしばしばどちらも満足のいかない選択肢の間で強いられる。対照的に「行為者の視点」においては、倫理的な課題 ethical challenges は文脈における行為者 an actor in a context の視点から扱われる。彼は問題に対する解決を見つけなければならない。工学者はケース・スタディを通じて工学的問題にアプローチすることに慣れており、これは Hampshire の行為者アプローチと同列に並べられる。Whitbeck は「倫理的道德的問題 (problem) は対立する (通常二つの) 立場あるいは対立する原理の間の衝突としてしばしば表現されるが、すべて同時には満たすことができない複数の (倫理的) 制約を含むものとして理解した方が良くある」と書いている。このことは、倫理的な課題は、設計問題と同様のものとして捉えられるべきであり、そしてそのようなものとして扱われるべきである、ということ Whittbeck に示唆する。

AMAs の設計に貢献しようと思う倫理学者にとって、この行為者の視点を心にとどめておくことは二つのレベルで有用である。第一に、行為者の視点は工学者の理解する問題解決アプローチに似ている。第二に、ロボットや計算的システムは、特定の文脈において倫理的制約の中で行為する方法を探す、単純な精神をもった (simple minded) 行為者として理解するのが最も良い。倫理的制約の中で働く計算的システムを設計する際には、実践的なアプローチが強調されなければならない。

P. 76- (“This emphasis.....”)

哲学者にはこのように実践的な側面を強調することは倫理学の過度の単純化であるように思われるかもしれない。複雑性の認識は、計算的システムをより洗練させる方法を示唆する限りにおいて有用である。AMAs を作るプロジェクトを拒けるために複雑性を強調することは有用ではない。倫理学を複雑にしている要素は様々あり、AMAs にそれらの問題すべてを扱うことを期待するのは実際的でない。著者たちの基本的な立場は、ロボットにおける、道德的考察に対する感受性に向けた歩みはすべて、どれほど単純なものであっても、正しい方向への一歩だ、というものである。

自律的なソフトウェア行為者の道德的能力を強化するというタスクは、科学者と工学者に、倫理的意思決定を構成部分に分解すること、どの種の決定がコード化でき、機械的なシステムによって制御できるかを認識すること、そして曖昧さと観念の衝突を制御する能力を持つ認知的情動的 (cognitive and affective) システムを設計するにはどうしたらよいかを学ぶことを強要するだろう。このプロジェクトは、人間の道德的意思決定を、まだ知られていない詳細まで分析することを要求するだろう。

P. 76- (“Different specialists are likely.....”)

AMAs の実装に対するアプローチは専門家によって異なるだろう。工学者と計算機科学者にとっての自然なアプローチは、倫理を単なる付加的な制約として扱うことだろう。この見方では道德推論には特別なことはない。しかしその制約がどんなものであるべきかについては議論の余地が残される。道德的なロボットを作るということは、適切な制約の組み合わせと、衝突を解決する適切な方式を見つけるということの問題になるだ

ろう。従って AMAs を開発するという問題は、知的システムの制御アーキテクチャーの中に抽象的な価値を実装する方法を見つけることとして理解されるかもしれない。

そのような制約はどこから来るのだろうか？ 哲学者はおそらく特定の倫理理論をソフトウェアにコード化するというトップダウンアプローチを提案するだろう。しかし計算可能性に関して言えば、哲学者の提案する道徳的原理は多くの点で不十分である。それはしばしば相いれない複数の行動を提案する、あるいはいかなる行動も推奨しない事がある。また行為がもたらしうる結果には本質的に限りがないので、鍵となる倫理的原則は計算論的にいって手に負えないように思われる。

しかし倫理理論がコンピューター・プログラムとして実装できないのであれば、そのような理論が人間の行動の方針としての役割を果たしているのかが疑問に思われるかもしれない。機械ができることとできないことについて考えることは計算的領域における倫理理論の限界についてのより深い反省をもたらすかもしれない。この観点から見ると、AMAs の問題は、どのようにして抽象的理論的知識を AMAs に与えるかではなく、世界において適切に反応する傾向をどのように身体化するかということの問題である。これは道徳的心理学の問題である。

心理学者は人間の子供の中に道徳的感覚が発達する仕方に注目する可能性が高い。発達のアプローチは機械道徳に向かう最も実的なルートかもしれない。しかし道徳的な人間を発達させるこの過程が信頼できないということが知られていることを考えれば、ロボットが AMAs になるよう訓練する試みがどれだけ信頼できるかと問うことは正当である。そしてここでもまた複雑性は大きな困難である。

Whose Morality or What Morality?

P. 78— (“There’s another kind of reaction.....”)

AMAs を作るプロジェクトは、誰の道徳的基準が実装されるのか、どのような倫理的サブルーチンを作るのか、という難しい問題に直面する。工学者は良く特定された課題のためのシステムを作ることに長けている。しかし道徳的振る舞いには課題が明瞭に特定されていない (there is no clear task specification for moral behavior)。道徳的な問題に関しては人々の間には大きな意見の不一致がある。

様々な行為の道徳性についての意見の不一致は、人工的システムの行動に道徳性を帰属させるための規準を決定することのむずかしさを強調する。行為の道徳性に関する観点の多様さを考えれば、AI に誰の道徳性あるいはどの道徳性を実装するかという問題には一つの答えはないのかもしれない。すべての計算システムが同じ行動規範に従わなければいけない理由もない。AMAs の設計にどの倫理規範が優勢になるかに関わらず、そのシステムはそれが道徳的行為者としてうまく機能しているかどうかに関わらずに決めた規準を満たさなければならないだろう。

Top-Down and Bottom-Up Approaches

P. 79— (“The study of ethics.....”)

倫理学の研究は一般的に、トップダウン的な理論的アプローチに焦点を当てる。すなわち倫理的議論は典型的には道徳性の広い基準を特定の事例に適用することに注目する。このアプローチでは規準、規範、原理が行為の道徳性を評価する基礎である。

「トップダウン」という言葉は工学においては異なる意味で使われる。彼らの言うトップダウン分析とは、課題をより単純な部分的な課題に分解することである。構成要素はモジュールに組み立てられ、それらは個別

的により小さな部分課題を実装する．それからモジュールが階層的に配置され，元々のプロジェクトによって特定されていた目標を達成する．

著者たちの議論において，「トップダウン」はこの二つの異なる意味を組み合わせで使われる．AMAs の設計に対するトップダウンアプローチとは，特定された倫理理論を取り上げ，その理論を実装することのできるアルゴリズムとサブシステムの設計に指針を与えるために，その理論の計算的要件を分析するアプローチである．

機械道徳へのボトムアップ的アプローチにおいては，行為者が行動の選択肢を検討し，道徳的に賞賛される振る舞いを学習し，そのような振る舞いによって報酬を受けるような環境を作り上げることに重点が置かれる．このアプローチには子供の発達や進化などのモデルがある．トップダウンの倫理理論と異なり，ボトムアップ的アプローチでは，いかなる倫理的原理も発見され構成されるものである．

実践においては工学者やロボット工学者は典型的にはトップダウンとボトムアップの両方のアプローチを使って複雑なシステムを作る．プロジェクトの目標を達成する方法は通常は複数ある．そしてプロジェクトの構造の分析と，その目標を達成するために設計されたシステムの試験の間にはダイナミックな相互作用（interplay）がある．

トップダウン／ボトムアップの二分法は多くの複雑な工学的課題にとって単純すぎる．そして AMAs の設計もその例外ではないと考えるべきだろう．にもかかわらず，課題の分析についてのこの二つの概念は AMAs の設計にとっての倫理理論の二つの可能な役割を強調する．