

## 人工知能と倫理

伊藤 博文<sup>†</sup>（愛知大学法科大学院）

### 要旨

人工知能の進化にともない人工知能の振る舞いが高度になればなるほど、その振る舞いの価値判断が問題となる。そこで多くの先端技術開発では倫理を用いて振る舞いを規律しようとしている。この倫理というものを再検討し、倫理という社会規範に依拠することがどのような意味を持つのかを検討する。そして、倫理というものが人工知能開発によってどのように相互の影響を受けるのか、問題点の指摘とあるべき姿を考察する。

キーワード：人工知能，社会規範，トロッコ問題，e倫理

### 1. はじめに

本稿の目的は、人工知能 (AI: Artificial Intelligence) と倫理 (Ethics) について考察することにある。進化し続ける人工知能開発段階では、倫理という社会規範 (Social Norm) に、開発方向の判断において依拠するところが大きく、方向性を誤らないためにも、倫理という規範 (Norm) の意義づけと限界点をあきらかにする必要がある。倫理規範定立における問題点の指摘及びあるべき姿の

考察することが本稿の目的でもある。

最初に、問題の所在を明らかにするために、トロッコ問題 (Trolley Problem)<sup>2</sup> を例として考える。

202x年人工知能を搭載した自動運転バスは、巡回コースを走行していた。交差点に差し掛かった時、大型トレーラーが信号を無視して交差点内に突入しようとしている。これを避けるため操縦舵装置 (ハンドル) を右に切ると衝突は避けられるが、登校中の小学生の通学列に突入し、12名ほどの児童の死傷は避けられない。このままトレーラーと衝突すれば、乗客5名は確実に衝突死し自らも大破する。乗客の4名は高

<sup>†</sup> 愛知大学法科大学院教授。以下のメールアドレスに忌憚なき意見や批判を送付していただければ幸いである。hirofumi@lawschool.aichi-u.ac.jp。今後、本稿の改定が必要な場合は、改定版を <http://cals.aichi-u.ac.jp/project/PN0160.html> にてPDFファイルで公開する予定であり適宜参照いただければ幸甚である。また、本稿引用文中URLの最終アクセス確認日は2017年9月13日である。

<sup>2</sup> <https://ja.wikipedia.org/wiki/トロッコ問題> このトロッコ問題への取組として有名なものには、MITが行っているMoral Machineがある。<http://moralmachine.mit.edu/hl/ja>

齢者であり1人は介助している40代女性である。

この時、自動運転を行う人工知能はどのような判断をすべきであろうか。また、あなたがこの人工知能に判断を指示する立場の人としたなら、どのような指示をすべきか。また、設計段階でどのようなプログラムを組み込んでおくべきか？

ここで登場するのが、倫理である。倫理はこうした場面での答えを導き出してくれると多くの人は考え、倫理に頼ろうとする。では、この倫理とは何であるのか。倫理にこのような局面での価値判断を委ねることは正しいのであろうか。

## 2. 規範としての倫理

倫理を考えるにあたり、上位概念である規範（Norm）から考えていきたい。

規範とは、「①のり。てほん。模範。②〔哲学用語〕のっとるべき規則。判断・評価または行為などの拠るべき手本・基準」<sup>3</sup>とされ、人が行動を起こすときの指針となるものである。

規範は、理由の如何は問わず、まず人が従うべき行動様式である。この規範の一類型として、社会生活をおくるために必要とするものを社会規範とする。この

<sup>3</sup> 株式会社岩波書店 広辞苑第六版

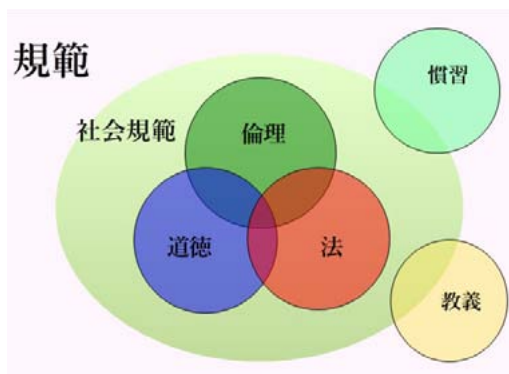


図1 規範の分類概念図

社会規範の中には、道徳、法、倫理、慣習等がある。本稿の目的とする倫理というものがこの中でどのように位置づけられるのかを説明する。

社会規範は、人が集団生活を余儀なくされるが故に、集団の調和を保つため守るべきルールとして、個人的な欲望を自制する規範として特徴付けられるものである<sup>4</sup>。人は社会的動物であり、己の欲望のままに生きることはできず、社会生活を維持するために他を慮り、自己の望まない形での行動を行う。たとえば、寄進といった食べ物等を他人に無償で与えるという行為である。このような行為を導く論理として、道徳、倫理、法、慣習、伝統、規範、掟、定め、宿命、善行などと

<sup>4</sup> この社会規範に対比される規範類型としては、自然規範と呼ぶべきものがある。これは、自然科学の法則により人が拠る行動規範である。たとえば、雨をしのぐために傘を差す、寒さをしのぐために服を着る。死を受け入れる、生きるために食物を採りその為に殺生をする、などである。おなじく文化規範も存在すると考える。

いった言葉が混在している。ここでは、人が自分の本能的欲望に反してまで行わなければならない行動をどのように整理するかから始めたい。

## 2.1. 社会規範の多義性

社会規範も更に細分化され、法 (Law), 道徳 (Moral), 倫理 (Ethics) とに分けることができる。

この3つのカテゴリーの差異, 同一性, 機能などについては, 差別化が困難故にさまざまな議論があり, 特に法律学の中でも, 「法と道徳」というテーマは法哲学上の重要な一課題であり続けている<sup>5</sup>。

## 2.2. 倫理・道徳・法

ここで, この3カテゴリーの意義を確認したい。まず, (1) 倫理とは「①人倫のみち。実際道徳の規範となる原理。道徳。」と広辞苑<sup>6</sup>では説明されており, (2) 道徳とは「①人のふみ行うべき道。ある社会で, その成員の社会に対する, あるいは成員相互間の行為の善悪を判断する基準として, 一般に承認されている規範の総体。法律のような外面的強制力を伴



図2 法, 道徳, 倫理の相関図

うものでなく, 個人の内面的な原理。」とされる。(3) 法とは「①物事の普遍的なあり方。物事をする仕方。また, それが生きたりになったもの。のり。おきて。②社会秩序維持のための規範で, 一般に国家権力による強制を伴うもの。」と説明されている。

この3カテゴリーの差異については, 「強制力」という点において, 法が他2者から際立つ (図2-Area3)。つまり, 法が国家権力による物理的強制力を持つ点において, 倫理・道徳と差別化される。法は, 国家が後ろ盾になって実現する社会規範であり, 警察権や裁判所といった司法制度を使った有形力 (例えば強制執行) で, 望まれる行動規範を実現させることができる。

一方, 道徳は人の内心に働きかけるも

<sup>5</sup> 碧海純一『新版 法哲学概論』弘文堂1964年70頁, 三島淑臣『法思想史』現代法律学講座3聖林道書院新社1980年171, 279頁, G.W.F. Hegel, *Grundlinien der Philosophie des Rechts, Hegel Werke* 7, Surkamp 1821, ss203 参照。

<sup>6</sup> 株式会社岩波書店 広辞苑第六版。

のであり、望まれる行動規範に従わなかったとしても何らペナルティを受けることはない。良心の呵責に苛まれるとしても、本人が自発的に行動しない限り何も変化は起きない。同様に、倫理も同じである。強制力をもたない点において、道徳と倫理は共通するが、法に劣後する。

では、道徳と倫理はどのように違うのか。倫理と道徳については、内心に訴えかける規範という点では区別が難しい(図2-Area5)。道徳は、個人や家族といった小集団を対象とし極めて個人的な内心に大きく依存するのに対し、倫理は汎用性を持ち得る。つまり、道徳は地域性、宗教、慣習などの個人的な要素を多く反映させているのに対し、倫理は特定の社会集団に共通の規範として適用し得る(図2-Areal)。

たとえば、生命倫理学といった応用倫理学の一分野(右掲図3参照)では、生命科学に従事する研究者間の研究活動を規律する規範として作用する。その研究者が仏教徒でもキリスト教徒でも、アメリカ人でも日本人であろうとも、老若男女問わず、その一集団で確立された倫理はその集団内で汎用性を持つ。強制力をもつ法は、内心の自由まで踏み込むことはできず外形力による強制を加えるのみであり、道徳のように個人的な内心の多様性を認めない点において倫理は優位性を持つ。

この点において、法の普遍性を説く考えもあるが、実際のところ法は国家単位で機能する規範であり、国を跨ぐ規範としては、倫理に劣る。つまり、生命倫理という観点から、遺伝子操作に関する共通倫理を科学者間で打ち立てることはできても、法にはできない。全ての国家が統一行動をとり、国際条約批准を行い国内法による立法規制で実現するというのは事実上、不可能に近いからである。

### 2.3. 倫理への期待

倫理が道徳に優る点は、上述したように汎用性である。繰り返しになるが、道徳が個人や家族などの小集団に用いられることが多いのに対し、倫理は個々人の関係から社会に至るまでより広範に用い

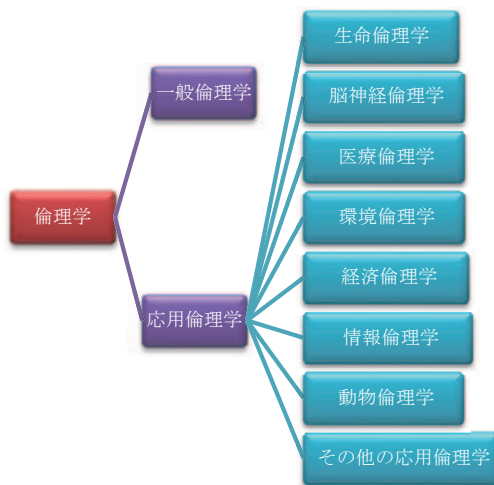


図3 倫理学の分類

られることが多く、そのため、道徳は日常生活における行動の基準にはなっても、先端科学の現場における判断基準にはなり得ない。

よって、さまざまな応用倫理学が成立し、学問の先端領域で大きな役割を果たしているのである。法のように国家的な支援組織を望まず、尚かつ集団構成員の内心に働きかけつつも個人の自発的意思に基づく行動を望むのが倫理である。つまり人工知能開発者対する最も望ましい社会規範は倫理ということになり、現状の倫理への依拠状況が説明できる。

もっとも、倫理学の中にもさまざまな様相があり、その議論の歴史は古い<sup>7</sup>。特に20世紀に入り、科学先端技術の目覚ましい進歩により、倫理学を新たな領域に当てはめようとする応用倫理学の成果が目覚ましい<sup>8</sup>。

すなわち、倫理の有用性は、国家的な立法機関によるコンセンサス形成を不要

とし、その組織内で自在に決められるという柔軟性、個の差異を前提としつつも共通の社会規範を実現できる汎用性が高く、先端科学の規範として応用倫理学が重用されるのである。

### 3. 人工知能と倫理規範

人工知能は、開発段階では開発者の倫理感が人工知能の振る舞いを支配する。なぜなら、人工知能自体は与えられたフレームワーク内でしか動作しないからである。開発者の選択するアルゴリズムが人工知能の行動や振る舞いを規定するのである。

#### 3.1. 倫理的責任

ここでは、倫理的な責任とはどのような意味を持つのであろうか考えたい。人工知能と倫理といったコンテキストでは、人工知能開発者の実装するプログラムを構成するアルゴリズムに対する規範的な抑制として機能することは既述した。ここでの倫理は、人工知能開発「者」自身に向けられた抑制的な社会規範である。

一方で、人工知能が何らかの社会的損害を引き起こした場面で、その引き起こされた社会的損害の帰責性を事後的に問う場面でも判断基準として倫理が出てくる。それは一面で、法的帰責事由の根拠

<sup>7</sup>たとえば、アリストテレスの「ニコマコス倫理学」がある。<https://ja.wikipedia.org/wiki/ニコマコス倫理学> 参照。パトリシア・S・チャーチランド著・信原幸弘、榎則章、植原亮訳『脳が作る倫理—科学と哲学からの道徳の起源にせまる—』化学同人227頁以下2013年参照。

<sup>8</sup>人工知能学会倫理委員会による「人工知能学会 倫理指針」については、<http://ai-elsi.org/archives/471> 参照。

IEEE (米国電気電子学会) の人工知能に対する倫理的考慮については、The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, [http://standards.ieee.org/develop/indconn/ec/autonomous\\_systems.html](http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html) 参照。



ともなるものである<sup>9</sup>。

この場合、人工知能開発者に倫理的判断を任せるとした場合、どのような原理・原則、価値判断に基づいて、その倫理的判断というアルゴリズムを実装させたのかというフレームワーク自体の設計思想が問われる。

### 3.2. 倫理規範は役立つのか

では、倫理という規範を人工知能開発段階で組み込むことは、どのような意味を持つのであろうか。

ここで考えたいのは、社会的統制の効率性である。ある社会問題が起きたとき、それを如何に解決することが最も効率的であるかということである。この場合の効率性は、紛争解決に費やされる時間やマンパワーといった社会的コスト削減によって実現される効率性である。

最も効率的なのは、社会構成員が自律的に規範を遵守しトラブルを起こさないように振る舞うことである。しかし現実には、社会紛争が頻発し、これの対処として国家規模による法による統制が行われ

<sup>9</sup> 法と倫理は、渾然一体となっており、完全に分離できる部分とそうでない部分が存在する。法自身が倫理的な規範を基にして成り立っていることも事実である。たとえば、法を犯すこと自体が倫理違反であるとすれば、法全ては倫理に包摂されることになり得る。しかしこれでは、峻別することの意義が失われてしまう。よって、法的責任を問う場合でも、その帰責根拠の模索および処罰の軽重を斟酌する場面で倫理が出てくるのは当然のこととなる。

る。法という社会規範による統制は、刑事罰といった法的責任による強制力を用い、人間を精神的にも身体的にも拘束する。このシステム維持には、多大なコストを要する。

これに対し、倫理は、社会に対峙する個の行動の指標であって、法のように強制力を持たず、法による他律的な規制ではなく、個による自律的な行動抑制により社会の幸福と発展を実現する自発的な行為をもたらす。「自律・分散・協調」を重んずるネットワーク社会においては、より望ましい社会統制である<sup>10</sup>。

## 4. 人工知能開発の将来に向けて

人工知能の将来に向けて倫理はどのような役割を果たすべきであり、その姿はどのようなものになるべきなのかを考察したい。

### 4.1. 市場機能

倫理の将来像を検討するにあたり、まず市場 (Market) という機能を考慮する必要がある。人工知能の振る舞いが問題となる局面は、人工知能が何らかの製品に組み込まれ、それが商品として市場

<sup>10</sup> 伊藤博文・佐野真一郎「大学教育における情報リテラシの方向性」2001年日本教育工学会第17回大会講演論文集(2001年) available at <http://cals.aichi-u.ac.jp/project/Evangelist/JapanSocietyForEducationalTechnology/JSET17.doc>

に出回り消費者事故となる局面である。つまり、一般消費者が人工知能の加害行為の被害者となる状況を考える必要がある。

たとえば自動運転のできる自家用車に装備された自動運転プログラムに組み込むモジュールとして倫理 Model A と倫理 Model B という商品があるとする。Model A は人命を尊重し、如何なる場合も人身事故を回避する行動をとることを選択する倫理モジュールである。一方 Model B は、人命よりも走行の効率性を優先する。目的地にまで確実に短時間で到達することが重視され、人身事故を引き起こしそうな場面では功利的な考慮を行い被害者数の少ない選択肢を選ぶとする。これを自動運転プログラムに入れるとき、自動運転車を購入する一般消費者が、あたかも任意保険の保障内容を選択するように選ぶとする。つまり、倫理の商品化である。人工知能搭載倫理モジュールを商品として、倫理の善し悪しを市場に委ねる方法である<sup>11</sup>。

しかしながら、このような市場による社会的選択の選好は機能しにくい。それは、単純に市場における需要供給のバランスが導く売買価格で決められる問題ではない。つまり、廉価で安全に配慮しな

<sup>11</sup> 倫理の形成にあたり、社会の要請を反映させることを主張するものとして、「ロボット・AIの社会的影響 (8) 中央大学教授・平野晋氏」日刊工業新聞2016年7月13日がある。<https://www.nikkan.co.jp/articles/view/00392496>

い人工知能倫理モジュールが、市場を席卷し事実上の標準 (De facto standard) になってしまうともはや倫理問題はなくなり、何も機能しない。

## 4.2. 営利行為と倫理

もっとも、人工知能の引き起こした社会的責任を全て製造者に負わせるとすれば、人工知能開発者の取る選択肢は、法的に最も責任回避が可能なアルゴリズムであり、製造者の利潤を最大化するアルゴリズムを実装するのは明白であり、社会規範としての倫理は機能しなくなる。

これまで例として考えてきたトロッコ問題を解決するのは、弁護士と法であるとする考えがある<sup>12</sup>。つまり、自動運転車に搭載される人工知能を開発するのは、営利企業である自動車製造会社であり、営利企業の至上命題は利潤追求である。人工知能に倫理を持たせ判断させる必要などはなく、トロッコ問題においては、如何に法的リスクを回避できるかを最適解とする。つまり、人工知能の振る舞いとして、たとえ事故が起きるような状況に入ったとしても、最も法的責任が

<sup>12</sup> Bryan Casey, *Amoral Machines, or: How Roboticists Can Learn to Stop Worrying and Love the Law*, NorthWestern Univ Law Review Vol.111 231(2017). これを紹介する記事としては、「自律走行車の「トロッコ問題」を解決するのは、技術でも倫理でもなく「弁護士」かもしれない」<https://wired.jp/2017/09/08/autonomous-vehicles-trolley-problem/> 参照。

軽くなる選択肢を選ぶようなアルゴリズムが選択されプログラムが書かれる。一旦、事故が起き人工知能の民事責任が問われれば、賠償責任主体としての企業自体の利潤が減少する。これを回避するためには、もっとも安価な損害を想定することである。法的責任を最も安価に回避できる行動が人工知能の倫理となるとする<sup>13</sup>。

この問題点は一考に値するが、このような思考様式そのものが人工知能開発において危惧され避けなければならない問題であり、倫理が統制すべき問題である<sup>14</sup>。

### 4.3. 倫理に必要なもの

では、これからの倫理に必要なものは何であろうか。

予期され得る限り全て危険に対し回避するアルゴリズムをプログラミングしておくことは、人工知能開発者としてのプログラマーにとって当然のことであろう。人身損害に繋がるようなリスクを回避するアルゴリズムを組み込むことは信

頼性設計 (fail safe) として不可欠である。しかし、予見できない危険に対してプログラミングできないこと、また危険としてプログラマーが考えないことに対して倫理的に非難することが可能であろうか。人工知能の振る舞いに対する法的責任追及は、人工知能の製造者に対して問われる。多くの場合、営利企業である製造株式会社であれば、プログラム・コードを書いたプログラマーだけでなく、当該会社そのものが法人として責任を追及される。その責任を追及する場面では、法的紛争になれば裁判所が判断することになり、法的紛争にまで行き着かない場合は、社会構成員間の合意(和解・示談)で判断される。むしろ人工知能の開発段階で倫理感を人工知能に植え付けることは現状では考えにくく、人工知能の fail safe としてのプログラム・コードが、人間の規範である倫理となろう。

われわれが民主主義を取る以上は、社会的コンセンサスは、当該社会構成員の総意で判断される。この「正しい情報を得た上での同意 (informed consent)」が将に倫理であろう<sup>15</sup>。その為にも、倫理の熟慮と議論が不可欠となってくるのである。

さらには、国民一般や市井の人々の意見を無視して、一部の技術者やエリート

<sup>13</sup> 伊藤博文「人工知能の民事責任について」愛知大学法学部法経論集第206号67頁(2016年)参照。Available at: <http://cals.aichi-u.ac.jp/products/articles/OnCivilLiabilityOfAIv2.pdf>

<sup>14</sup> こうした考え方を批判する主張が引き合いに出すのが「フォード・ピント事件」である。「法と経済学」学派および功利主義に対する批判としてしばしば引用される。<https://ja.wikipedia.org/wiki/フォード・ピント> 参照。

<sup>15</sup> インフォームド・コンセントについては、ブレント・ガーランド編著・古谷和仁、久村典子訳『脳科学と倫理と法』みすず書房106頁2007年参照。



が作り出した「倫理」が一人歩きして、人工知能の振る舞いを制御することは、ブラックボックス化を増長させ、民主主義とは反する結果をもたらすことになる。開かれた倫理策定議論が必要となってくるのである<sup>16</sup>。

次に、倫理を構成すべき要素は、至適基準 (grand truth) である。至適基準は、「その場でえられた情報」であり、現地で得られる情報でありメタデータではなく、経験知的なデータである (図2-Area7)。人工知能に倫理の基本として教えるにあたり、この知識は間違っていないと教えられるものである。あらゆる分野における秀でた模範 (gold standard) とも呼ばれ、科学実験における間違いない基準や専門的行為の最適な形態や手順・結果を基礎としなければならない。そして、こうした至適基準を正しく当てはめる力が必要となる。こうした知識を集合知というビッグデータとして実装し判断基準とすることが必要である。

#### 4.4. 倫理教育

さらには、社会的選択を強いられる場

---

<sup>16</sup> 倫理に「幸福」という考えを入れるべきと主唱するのは、三木清『人生論ノート』である。[http://www.aozora.gr.jp/cards/000218/files/46845\\_29569.html](http://www.aozora.gr.jp/cards/000218/files/46845_29569.html)。また幸福学 (Well-being) の観点から人工知能・倫理を構築するという研究方向も注目される。瀬戸隆「Well-Being Computing: AIとゲノミクスからの展望」人工知能32巻1号2007年87頁参照。

面において、社会構成員の一部の者達だけに選択権を委ねるのではなく、ネットワーク社会の特質を理解した上で個々人が最善の選択ができる状態にあるようにするためにも、応用倫理の教育が必要である。法規制といった国家や一部の権力者による管理統制を回避し、息苦しい社会にしないためにも、個の意識発揚を促す倫理教育が重要となる。

これは人工知能開発者向けのみならず、高度な情報社会においては、社会構成員すべてに向けて、倫理というものの醸成が不可欠となり、それを可能にするのは教育である。

#### 4.5. e倫理

ここまで人工知能と倫理という考察においては、人間である人工知能開発者の倫理について考えてきたが、技術的特異点 (Technical Singularity) 以降は、人工知能が自律的に考えて振る舞うこととなると予測される<sup>17</sup>。

このときは、どのような倫理に基づき人工知能は振る舞うのであろうか。つまり、倫理を人工知能自身に考えさせ、人工知能に倫理的な振る舞いを行わせる段階となる。

---

<sup>17</sup> 伊藤博文「法的特異点について」愛知大学情報メディアセンター紀要『COM』Vol.26/No.1第41号13頁 (2016年)  
*available at* <http://cals.aichi-u.ac.jp/products/articles/OnJudicialSingularityV1.pdf>

倫理は、人間の行動様式を規制する規範として作用する。しかし人工知能はすべてプログラムであるコードで制御される。このコード自体が倫理と呼ぶものとなり、この倫理を人工知能自らに判断させるのがe倫理（eEthics）である。

では、そもそも人工知能は、倫理を生み出せるのであろうか。肯定的（positive）な考え方であれば、無論可能とするであろう。倫理が人類の集合知（Collective Intelligence）であるとすれば、その集合知を持った人工知能なら生み出すことは可能である。倫理を構成するイデオロギーが有限個の知識の組み合わせであるなら、あらゆるイデオロギーはコード化できる。こうしたコーディングが可能であれば、無限とも思える倫理コードから成るビッグデータ内の凄まじい数の組み合わせの中から、テキストマイニング技術を使い人工知能が正しいものを選び出せばよい。

これまで人類が蓄積してきた知的資産を読み込ませ、倫理とは何かを導き出させることも可能となる。人工知能を使って倫理を突き止めるのである<sup>18</sup>。

<sup>18</sup> チェン・ドミニク教授は、「人間が計算機と同じく合理的な存在であれば、倫理問題は存在しないともいえる。純粋に人間個体の価値を、ジェンダーや年齢、社会属性といったメタデータから数値化して評価できればトロッコ問題もすぐに解決できる。中略 原理的にアルゴリズムでは処理できない情動や感情が合理的判断と衝突するかと考えられる。」とされる。チェン・ドミニク「生成的倫理とその広告への適用について」人工知能32巻4号509頁2017年。

これを進めれば、リアルタイムの価値判断が可能となる<sup>19</sup>。倫理が集合知（Collective Intelligence）であるならば、できるだけ多くの社会構成員のパラメーターを事前に登録しておき、判断を求められたときに、これを利用して瞬時に判断させる。倫理という社会規範が、その社会における構成員にとって合意できるものであれば良いとするなら、瞬時に判断させられることが可能であろう。人間の遅い反応を待ってはいは答えが瞬時に求められないので、予め倫理的要素のパラメーターを設置しておくのである。

たとえば、トロッコ問題であれば、人数、年齢、性別、体型、社会的地位をパラメーターとして、どれを優先するかを予め人工知能に登録させておき、判断が必要になったときこのエンジンに問い合わせをすると、倫理的最適解が導き出されるのである<sup>20</sup>。

<sup>19</sup> Digital Ethics については、<http://www.oecd.org/science/we-need-to-talk-about-digital-ethics.htm>, <https://www.oii.ox.ac.uk/research/digital-ethics-lab/> 参照。

<sup>20</sup> このとき、倫理を含めて社会規範は構成員の集合知であるから、パラメーターの強さが判定基準となる。そこで問題となるのは、状況把握能力である。自動運転車の人工知能がどれだけ細かい状況把握ができるかである。今ひき殺そうとしている人物の年収をどうやって予測するのか、アルゴリズムの優れたAIシステムを市場に選択させたりするのは危険ではないか、パラメーターの選定が恣意的にならないか、といった問題が想定される。

## 5. おわりに

倫理に求めているのは、判断基準である。トロッコ問題でどちらにハンドルを切るべきかを教えてくれる規範である。当然のことながら、社会が変化する以上、社会規範も変化する。

人工知能倫理は、常に未来へ向かっての規範となるはずである。未来と現在の違いは、時間的な差異であるが、その時間の量については未定義である。1秒後の自分、1分後、1時間後、1月後、1年後、10年後の自分を比べてどれだけの差異を発見できるかが未来予測である。容易ではないことは論を待たない。過去でも同様である。一秒前の自分を過去とは思えない。過去と現在に瞬時の差異は見いだせないからである。時間が変化する以上、倫理も時代と共に変化するということである

さらに、地域、構成集団により倫理は異なり得る。倫理に期待する普遍性と矛盾するが、これも宿命である。地域差においては、アメリカと日本では倫理観が異なっても当然である。構成集団が変動要素になることも当然である。倫理は、この地球上の或る地域の或る集団が或る時代に共有する社会規範である。よって、倫理を形作れるのは或る集団の構成員の価値観である。その価値観が一致するのであれば、それが倫理ということになる。

普遍性を期待できない倫理という社会規範を持ち出すべくもなく、人工知能のなすべき振る舞いを望ましい形で制御できればよいのであるから、これを倫理と呼ぼうが、ルールと呼ぼうがなんでも良いのではと考えることも可能であろう。しかし、人工知能をはじめ先端科学で求められているのは、人類が伝統的に使い分けてきた社会規範としての倫理である。人工知能といった先端技術を開発する人間に対しての倫理である。

この伝統的な倫理という社会規範に新たな要素を組み込むことで、よりよい倫理が生まれてくることに期待したい。そのためにも、この分野の研究は、人工知能のみならず、哲学や認知科学といった学際的な研究が求められるのである。

