

# 人工知能が作り変える社会とその倫理的課題

久木田水生

2018年3月25日

市民公開シンポジウム

「情報倫理と人工知能・ロボット——その課題を考える」

京都女子大学

## 1 はじめに

人工知能は現在、急速に発展し、実用化されつつある。それは様々な点で社会を変えるだろう。そのことはまた様々な倫理的含意を持つ。この発表では人工知能がもたらしうる社会の変化とそれがもたらす倫理的課題について、いくつかの具体的な例を取り上げながら論じる。

## 2 テクノロジーと人間

### 2.1 媒介としてのテクノロジー

テクノロジーは人間と世界の「媒介」、あるいは「インターフェイス」であるという考え方がある。つまり私たちはテクノロジーを通じて世界と相互作用するということである。この相互作用には二つの方向性がある。それは「認知」と「行動」である。私たちはテクノロジーを通じて世界についての情報を獲得する（認知）一方で、テクノロジーを通じて世界に対して働きかける（行動）。この認知と行動の間をつなぐのが私たちの思考と判断、意思決定である。私たちは世界について獲得した情報に基づいて思考、判断、意思決定し、そしてそれに従って行動する。よりよい情報が得られれば、私たちは効率的な行動、すなわち少ないコストとリスクで大きな利益をもたらすような行動をとることができる。

しかしながら人工知能やロボットの発展はこのような「媒介」としてのテクノロジーという見方に大きな変更を要求する。というのも人工知能やロボットは自動的に情報処理を行うことによって、人間から思考という作業の労を取り除くからである。人工知能やロボットに頼ることによって私たちはもはや自分で考えることなく、この世界において効率的な行動を取る（あるいはロボットに代わりに取ってもらう）ことができる。それゆえに私たち自身が世界について（あるいはそこに含まれる他者について）の情報を手に入れる必要も少なくなってくる。このことは私たちの世界に対する態度、世界との付き合い方に大きな影響を与える可能性がある。

## 2.2 テクノロジーの影響を予想することの難しさ

例えば鉄条網について考えよう。もともとはアメリカのビジネスマンが家畜を囲うための道具として考案したものだ。安価で設置が容易だったため多くの農家がこれを用い、そのことはアメリカにおいて広大な土地の砂漠化を引き起こした。さらに鉄条網は本来の目的のためだけではなく、多くの人々を隔離・弾圧するために使われた。第一次世界大戦では塹壕戦を長く悲惨なものにすることに貢献した。このような影響を開発者や初期のユーザーは予想できただろうか？

テクノロジーの影響というものは予想するのが難しい。テクノロジー自体は些細なものでも、それが社会という複雑なシステムの中に組み込まれたときに、どのような使われ方をして、どのような長期的な影響を持つかということは、誰にも分らない。

## 2.3 テクノロジーは中立か？

しばしば「テクノロジーそれ自体は価値中立的なもので、それを良いものにするのも悪いものにするのも人間次第だ」という意見を耳にすることがある。確かに人間と全く関わることがなければ、いかなるテクノロジーも善でも悪でもないことは確かである。しかしそれが人間と相互作用するとき、あるテクノロジーは他のテクノロジーよりも悪用・濫用されやすいバイアスを持っているということもまた確かである。例えば麻薬は使い方によっては薬になるということは事実だが、しかし圧倒的に悪用・濫用されやすいバイアスを持っている。そういったテクノロジーの中には、人間の心理や生理、あるいは社会のシステムの脆弱性をハッキングすることによって、人間にほとんど利益を与えることなく、むしろ多くの害悪を与えているにも関わらず、社会に蔓延しているものがある\*1。

例えばアメリカ社会における銃について考えてみよう。アメリカでは年間1万人以上の国民が銃によって命を落としており、その中には幼児による発砲事故も20件ほども含まれている\*2。これだけの被害を引き起こしているにもかかわらず、銃が規制されないのには様々な要因があるが、その一つとして人間が相互に持っている他者への不信と警戒心がある。銃を規制してもあくどい人間たちはどうにかして銃を持ち続けるだろう。そうなれば正直に銃を手放した人間が不利な状況に置かれる。したがって自分は銃を手放すわけにはいかない、というロジックである。多くの人がこのように考えている限り、ひとたび銃が蔓延してしまった社会から銃を取り除くことは困難である。これはゲーム理論における「囚人のジレンマ」的な状況であり、一人一人が合理的に思考して意思決定した結果として、全体として利益の少ない選択をしてしまう。これは人間の心理の脆弱性にハッキングしてテクノロジーが蔓延している例である。

より直接的なハッキングの例としては麻薬やたばこのようなものを考えれば良い。これらは人間の生理に作用して、ユーザーがそれに依存するようにはさせる。ユーザーはそれを使用することで健康に大きな害を被るにも関わらず、それをやめることができない。

テクノロジーについて考える際には、このような人間の心理・生理の脆弱性を考慮せずに、「テクノロジーは中立」、「テクノロジーによって生じる危害は人間の責任」と断じることは、テクノロジーに向き合う適切な姿勢ではない。

\*1 「ハッキングする」という言葉はテクノロジーを擬人化した表現であるが、もちろんテクノロジーに意図や目的があってそういうことをしているということの意味しているのではない。この点についてより詳しくは久木田 [6] を参照せよ。

\*2 <http://www.euronews.com/2017/01/31/armed-toddlers-kill-twice-as-many-americans-each-year-than-terrorists>

以下、いくつかの例を参照して、人工知能・ロボットが人間や社会システムの脆弱性、セキュリティホールをどのようにハッキングしうるかということを見る。具体的には、ソーシャルロボット、軍事ロボット、人工知能によるプロファイリングを取り上げる。

### 3 ソーシャルロボット

人間にとって、互いに協力し合える他者を見つけるということは文字通り死活を分かち重要な課題であり、そのため人間は、他者が協力的であるかどうか、相手はこちらに対して好意的であるかどうかのサインに敏感である。例えば言語習得前の幼児でも、協力的なエージェントと非協力的なエージェントを区別して、前者に対する選好を示すということが分かっている (Cf. Hamlin, Wynn and Bloom [1], Kanakogi et al. [2])。興味深いのはこれらの研究に用いられたのが、丸や四角の幾何学的な形に目を付けたキャラクターであるということである。本物の人間とはかけ離れた見た目でありながら、幼児たちはそこに意図や目的を読み込み、そして協力的であるかないかを区別していたのである。

ソーシャルロボットは、ユーザーに対して好意的であるような言動を発することで、ユーザーから愛着を持たれるように設計されている。心理学者のシェリー・タークルはこのことを「ダーウィンのボタンを押す」と表現する (Turkle [5])。倫理学者のロバート・スパローは、ソーシャルロボットはユーザーに「ロボットが自分のことを気にかけている」という幻想を持たせることによって効果を上げることを意図したものであり、それは欺瞞であり非倫理的であると批判する (Sparrow [4])。久木田・神崎・佐々木 [7]、第5章はこれに対して、スパローのような批判は「嘘をつくべきではない」という義務を前提としたものであり、例えば功利主義的な立場からはこのような批判は必ずしも妥当ではないことを論じている。久木田らによれば、ソーシャルロボット一般が非倫理的と言えるような根拠はない。

このことはしかしながらソーシャルロボットの危険性を無視して良いということの意味しない。スマートフォンやソーシャルメディアのようなテクノロジーが既に私たちのコミュニケーションの実践と規範に大きな変化をもたらしており、その影響は甚大である。ソーシャルロボットはそのような傾向にさらに拍車をかけるだろう。しかしながらソーシャルロボットは普及し、私たちの生活の一部になるだろう。特に高齢化の進む日本ではソーシャルロボットの技術は医療・介護の分野で必須になる。ソーシャルロボットを求める人々がいて、ソフトバンクや Google のような大企業がそれを売りたいがり、そして政府がそれを成長戦略の一部として推進するだろう。だからこそ私たちはその影響について慎重に検討していかなければならない。

### 4 軍事ロボット

軍事テクノロジーの歴史は、できるだけ離れた距離から効果的に敵にダメージを与える方法の開発の歴史だった。素手から、剣、槍、投石、弓矢、銃、大砲、ミサイルと、テクノロジーの発展とともに、人々はより遠くからより破壊的なダメージを敵に与えることが可能になった。地球の裏側からでも遠隔操作できるドローンはこの発展の一つの到達点を示す画期的なテクノロジーである。そして戦争の歴史が示すように、画期的な軍事テクノロジーは戦争の戦われ方、命の扱われ方を変えてしまう。

2015年に「The Intercept」というアメリカのニュースサイトが、アメリカによるドローンを使ったテロリスト暗殺の実態を明らかにする文書（「ドローン文書」と呼ばれている）を内部告発者から入手し、公開し

た\*3。それ以前からドローンを使った爆撃によって、無関係の市民が巻き添えになっていることが問題視されていたが、この文書によれば例えば 2012 年の 5 月から 9 月に行われた作戦においては、ドローンによって殺害された人間の 8 割以上が本来のターゲット以外の人間だった。

生身の人間がテロリストを襲撃する場合と比較して、ドローンを使った作戦では空高くから破壊力のあるミサイルをターゲットをめぐって撃ち込むことになる。それゆえに必然的にターゲット以外の人間を巻き添えにすること、「付随的損害」が増えるのである。このような問題にも関わらず、アメリカはますますドローンに頼るようになってきている。アメリカ空軍航空戦闘軍団のカーライル司令官はドローンを「戦場におけるもっとも貴重な財産の一つ」と呼ぶ\*4。味方が殺されることなく、相手を殺傷できるドローンは使う側にとっては理想的な兵器である。そのような兵器を使うために、戦場における実践と規範が変容している。オバマ政権はドローンによる攻撃を「合法的」、「倫理的」、「賢明」、かつ「人道的」と称している\*5。

戦場の主役が遠隔操作されるドローンからより自律性の高い兵器に代わる時、戦争の戦われ方により劇的な変化が起きるかもしれない。アメリカの NPO、Future of Life Institute は自律型兵器に反対する公開書簡の中で、自律型兵器をめぐって新たな軍拡競争が起こる危険性や、自律型兵器がテロリストに利用され、未来の「カラシニコフ」になる危険性を指摘している\*6。自律型兵器は人間の兵士と違って感情によって判断を狂わされることがないために、戦場をより「倫理的にする」という意見もあるが、しかし私たちは新しい軍事テクノロジーが戦場をよりましな場所にしたことは歴史上かつてなかったことを知っている。

## 5 プロファイリング

キャシー・オニール [3] は、様々なデータに基づく数学的なモデルによって人間を分類することが、社会に有害な影響を及ぼしていることに警鐘を鳴らしている。彼女は不公平で、大勢の人間に害を与える、不透明な数学的モデルを「weapons of math destruction (WMDs)」と呼び、批判する。

彼女が挙げている例の一つは予測警察である。過去の犯罪のデータに基づいて、ある時期に犯罪の起こる可能性の高い地域を予測する数学的モデルがアメリカでは実際に使われている\*7。しかしこのようなモデルの実用化には問題がある。犯罪が発生する確率が高いと予想された地域に多くの警察官を配備すれば、その場所での検挙数は当然上がるだろう。そのことはこのモデルが正しかったという確証を与える。そしてその場所に住む住人はますます厳しい監視の目にさらされ、軽微な犯罪でも警察に捕まることになる。一方で、この予測から免れた地域の人々は軽微な犯罪ならば捕まらない可能性が高くなる。犯罪が発生するのは貧しい人々が暮らす地域であり、したがってこのようなシステムは貧しい人々をさらに苦しめることになりかねない。

あるいは人事採用について考えてみよう。HireVue という面接プラットフォームは、候補者の身振りや語彙などを評価して、その会社のパフォーマンスの高い社員の身振りや語彙と近い順にランク付けする。そうしてランクの高い候補者にリソースを集中することで、面接にかかるコストを削減することができる、というのである。さらに HireVue のサイトでは、「偏見を減らし、公平な基準に基づいた採用を実現」とも謳われている\*8。しかし、そもそも身振りや語彙というパラメーターを評価の基準として選んだところに、このアルゴリ

---

\*3 <https://theintercept.com/drone-papers/>

\*4 <https://edition.cnn.com/2016/03/17/politics/air-force-pilot-shortage/index.html>

\*5 <https://www.ft.com/content/dae6552c-84c2-11e2-891d-00144feabdc0>

\*6 <https://futureoflife.org/open-letter-autonomous-weapons/>

\*7 日本でも神奈川県警が機械学習を利用した予測警察の実用化を検討している。<https://www.asahi.com/articles/ASL273VLQL27ULOB007.html>

\*8 <http://hr-odk.jp/hirevue/>

ズムの設計者のバイアスが入り込んでいる。さらにその会社では特定のジェンダー、人種、出身地などが優遇されているとすれば、同じようなバイアスを持った選考がアルゴリズムによってなされるだろう。ジェンダーや人種や出身地などを採用の基準にすることはアンフェアである。しかしこのようなアルゴリズムによって、明示的にそれらの要素に言及せず実質的な差別を、客観的で偏見がないという見せかけのもとに、実現することが可能になる。

IoT とビッグデータによってますます多様で大量のデータを利用することが可能になっている。そして人工知能はそこから私たちの行動や能力、犯罪的傾向、特定の病気にかかる可能性などなどについて、予測を立てる。データを活用できる企業は莫大な利益を上げることができる一方で、それは一部の人間を犠牲にしている。

## 6 終わりに

バリー・レヴィンソン監督の映画『ワグ・ザ・ドッグ』の冒頭では「犬が尻尾を振るのはなぜか。尻尾より賢いからだ。もし尻尾の方が賢ければ、尻尾が犬を振る」というジョークが掲げられる。このジョークをもじって私は次のように言いたい。「人が道具を使うのはなぜか。道具より賢いからだ。もし道具の方が賢ければ、道具が人を使う」。これはしかしジョークではない。ますますスマートになる道具によって人間が使われていると言ってもいいような状況を私たちはあちこちで目の当たりにしている。上述のキャシー・オニールは数学的モデルについて、それは「私たちの道具であるべきで、支配者であるべきではない」と述べている ([3], p. 207)。

私たちが道具に使われないためには何が必要なのか。道具よりも賢くなることである。その道具が何を目的として作られたのか、どのような仕組みで働いているのか、その道具によって誰が儲けているのか、だれが踏みつけにされているのか、どのような長期的な影響がありうるのか。そういったことに注意を向けておくことが重要である。

## 参考文献

- [1] J. K. Hamlin, K. Wynn, and P. Bloom. Social evaluation by preverbal infants. *Nature*, 450:415–424, 2007.
- [2] Y. Kanakogi, Y. Inoue, G. Matsuda, D. Butler, K. Hiraki, and M. Myowa-Yamakoshi. Preverbal infants affirm third-party interventions that protect victims from aggressors. *Nature Human Behaviour*, 2017.
- [3] C. O’Neil. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Penguin Books, 2017.
- [4] R. Sparrow. The march of the robot dogs. *Ethics and Information Technology*, 4(4):305–318, 2002.
- [5] S. Turkle. Authenticity in the age of digital companions. In M. Anderson and S. L. Anderson, editors, *Machine Ethics*, pages 62–76. Cambridge University Press, Cambridge, 2012.
- [6] 水生 久木田. 麦とペッパー — テクノロジーと人間の相互作用. *人工知能*, 32(5):653–659, 2017.
- [7] 水生 久木田, 宣次 神崎, and 拓 佐々木. *ロボットからの倫理学入門*. 名古屋大学出版会, 名古屋, 2017.