

第3章 AI倫理の基底

1 人間機械論批判

AIの倫理をめぐる論考は多々あるが、生物と機械との差異に基づきながら検討した研究は見当たらなかった。二〇〇〇年以降、AIもしくはロボットをめぐる倫理は大きく分けて二方向で議論が進められている。一点目は、ロボエシックス (robethics) と呼ばれ、社会生活に組み込まれるAIと人間との関わりを探索しようとする方向性である。⁽¹⁾たとえばダニエル・デネット (Daniel Clement Dennett) (1997 = 1997) やデイヴィッド・レヴィ (David Levy) (2007)、西條玲奈 (2013) らの研究が挙げられる。二点目は、マシン・エシックスと呼ばれ、いかにしてAIに倫理的な側面を実装して道徳的な振る舞いをするAMAsを設計するかという方向性である。たとえば関口海良・堀浩一 (2008) やウエンデル・ウォラック (Wendell Wallach) & コリン・アレン (Colin Allen) (2009 = 2019)、マイケル・アンダースン (Michael Anderson) & スーザン・アンダースン (Susan Anderson) (2011)、久木田水生 (2012)、岡本慎平 (2012) らの研究が位置づけられる。これらの二方向の研究は、AIをめ

(1) 「ロボエシックス」は、二〇〇二年にジャンマルコ・ヴェルジオ (Gianmarco Veruggio) によって作られた造語である (Veruggio, 2006)。

ぐる倫理を考えるうえで示唆に富むものの、いずれも知能や感情に焦点が合わされており、生命と機械との相違についてはほとんど問題にしてこなかった。

第2章の前半で述べたように人間と機械との同質性を主張する声が一九四〇年代以降徐々に強まっている。AIの開発は、細かい違いはあるにせよ、「人間のように思考するコンピュータ」が目指されている面があり、人間と機械との境界を乗り越えようとする思想が見出せる。そうした傾向に合わせるように、情報哲学の分野でも自然物と人工物との境界を設けず、情報の存在として生命／非生命を連続的に捉える理論も提示されている(Florida, 2011)。ケイト・ダーリング (Kate Darling) (2012) のように、基礎的な議論を経ないまま財産権を超えるたロボットの保護について論じている研究者も現れている。しかし倫理的な問題において、生物と機械との差異について検討せず両者を同一線上に位置づけてよいのだろうか。

この相違を議論せずに倫理的な問題を検討していけば、後で述べる数々の課題が生じる。人間の尊厳の根源を破壊するほどの大きな倫理的問題を引き起こす可能性がある。むしろ生命と機械との差異を踏まえたうえで、これまでのAI倫理の議論を整理・再編することが望ましいのではないだろうか。生命と機械との差異は、倫理的議論の大きな分かれ目である。それゆえ、第2章の後半では生物と機械との違いを議論の俎上に載せた。本章では、その立論をもとにAI倫理の基底的な領域を検討することとしたい。倫理的責任 (moral responsibility) の所在や倫理的配慮の必要性、道徳的共同体への包摂の可否、擬人化の問題についても扱っていく。

なお本書で扱う責任は、法的責任 (liability) ではなく、あくまで倫理的責任である。法的責任として扱われることが多い議題も倫理的観点から論じる。周知のように責任 (responsibility) という語は多義語であるが、本書では過去の行為の賞罰に関わる責任という意味でも、現時点ならびに未来の行為の義務や責務に関わる責任という意味でも用いる。

人間と機械との区分は、AIの倫理を考えるうえでの基盤をなす。人間を含む生物と機械との間には埋めがたい差がある。最先端技術の結集であるAIでも、それは変わらない。仮に人間と機械との違いがないのであれば、機械自体に権利をもたらす方向で議論が展開できるのと同様、逆に人間を機械扱いする立論にも容易に結びついてしまうからである。こうした事態は、情報倫理の淵源としても挙げられるウィーナーがもつとも危惧したことである (Bynum, 2005; 西垣, 2010)。

(権力者の支配のもとで) 人間は、或る高級な神経系をもつ有機体といわれるものの行動器官のレベルに引き下げられてしまった。私は本書を、人間のこのような非人間的利用 (inhuman use of human being) に対する抗議に捧げたいのである。(中略) 人間の機械化は彼らの野望を実現する一つのかんたんな道である。思うに、権力へのこういう容易な道は、実は、人類にとって道徳的価値があると私が考える一切のものの廃棄であるばかりでなく、人類の今後かなり長期にわたる存続のための今やはまだ細くなった道の廃棄をも意味する (Wiener, 1954 = 1979: 23-24)

人間が機械であれば、冷蔵庫のように三六五日二四時間働かせても構わない。コンピュータは三年から五年ほどしたら処理スピードが遅くなるため不要物として捨てられるが、人間がそれと同じような扱われ方をされてしまいかねない。機械とは、用立てるものであり、なにかに役に立つために作られる。アロポイエティック・システムであり、人間によって設計・製造・維持されるものであり、人間の指示通り動くように要請されている。そのため、故障して目的の機能を果たせなければ廃棄されても仕方がない。

けれども、人間はそうではないのか。マルティン・ハイデッガー (Martin Heidegger) がいうように、

人間はそもそもなかに役立てるために作られているわけではない (Haidtger, 1962 = 1965)。人間は、細胞が次々と分化して生成し、その内部を常に作り続けているオートポイエティック・システムの集合体であり、一人ひとりとは唯一無二の存在である。みずから内部を存立させ外部との境界を作り出すがゆえ「主観」なるものが生成する。また個々のオートポイエティック・システムは、内的メカニズムに沿って環境を認知するが、同じ時空間を占めるほかのシステムがない以上、個別に環境を生み出す。すなわち、オートポイエティック・システムの内的メカニズムも唯一無二であり、それに伴いシステムが接する環境も唯一無二となる。他者との厳密な交換はきかない。私たちは、あらゆる物事を含みこむ単一の絶対的・客観的世界に生きているわけではなく、集合的にいえば多元的現実 (many realities, a multiversa)、すなわち複数の現実⁽²⁾に生きているのであって、一人ひとり固有の現実⁽²⁾に生きている (Maturana 1988)。したがって、人間を「役立つ／役立たない」の尺度だけで見るべきではないし、たとえ役立たなくとも社会から排除すべきではない。

人間と機械を同じ線分上に連続的に位置づける議論が広まりつつあるなかであえて紙幅を割いて異質性を主張するのは、私たちはいとも簡単に人間を非人間化することがあるからである。ルイス・マンフォード (Lewis Mumford) は次のように述べている。

西欧の人々が機械に頼るようになるずっと以前から、社会生活の一要素としてのメカニズムはもう出現していたし、発明家が人間力に代るエネルギーをつくり出す以前から、群集の指導者は多数の人間を訓練し、編成していた。つまり、彼等は人間を機械に貶しめる方法を発見していたのである。鞭の音にリズムを合わせ、ピラミッドの石を運搬した奴隷や農民たち、あるいは各員が座に鎖で縛りつけられ、かぎられた機械的運動のほかにもできないローマのガリー船を漕いだ奴隷たち、マケドニアの方陣の整列と行軍と攻撃隊形

——こうしたものはすべて、機械的現象なのである (Mumford, 1934 = 1972: 57)。

たとえばロボット兵器は、攻撃する相手が生身の人間ではないかのように錯覚する危険性を秘めている。ピーター・シンガー (Peter Warren Singer) は、ドローン等を使って遠く離れた場所から相手を攻撃する戦闘員の体験について述べるなかで、P T S D (心的外傷後ストレス障害) に苦しむ人もいるものの、カタルにある部屋からイラク戦争を戦った人物が「テレビゲーム感覚だ。ちょっと残酷になる場合がある」と語ったことに言及している (Singer, 2009 = 2010: 481)。*war* に兵器の自動化が進んで完全に自動化されたシステムが選別・判定・攻撃を行い、戦闘員が相手の映像さえ見ることがないようになれば、人が確認するのは単なる数字にすぎなくなり戦闘相手が人間であることを忘れてしまいかねない。

逆にいえば、実際には機械が人間とは異質であるがゆえに、さまざまなことが可能となっている面がある。技術的人工物である機械だからこそ、壊れたら捨てることができる。人と同じであれば、勝手に機械の電源を切ったり内部の配線を変えたりすることがためらわれる。もし本当の人間であるかのように、もしくはそれに近い権利を与えてロボットを扱うのであれば、損壊しても捨てられず、OSやアプリケーションのインストールにあたってはロボット自身に許可をとらなければならない。あるいは、勝手にデータを読み込ませることも倫理的に認められない行為になるだろう。それは、いわば個人の内面を無断で操作する行為であり、不可侵な領域の侵犯にあたるおみなされてもおかしくはない。原子炉内などの人間にとって過酷な環境にロボットを送り込みにくくなるだろう。現在のところ、ロボットが人間のような自律性を備えていないからこそ、無許可でこうした行為をし

(2) とはいえ、第2章で論じたように「なんでもあり」というわけではない。

でも倫理的問題に問われないのである。