Kobe University Repository : Kernel

PDF issue: 2025-04-27

アリストテレスの「道具的理性」と人工知能（AI）について

李, 倩

(Citation)
21世紀倫理創成研究, 16:1-22

(Issue Date)
2023-03-31

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/0100482790

(URL)
https://hdl.handle.net/20.500.14094/0100482790

KOBE UNIVERSITY

アリストテレスの「道具的理性」と人工知能（AI）について

李　倩
神戸大学大学院人文学研究科　博士課程後期課程

はじめに

今は人工知能（AI）が急速に発展している時代である。しかし、AI 技術の急速な発展に伴い、人間の生活にも多くの新たな挑戦がもたらされてきた。たとえば、AI の社会的導入に伴う様々な倫理的問題（AI による雇用喪失、AI 技術の悪用、AI による差別など）である。そのため、我々人間は自分自身と知能機械との関係について再検討する必要がある。実際、今から 2000 年前のアリストテレスの理論を振り返ってみると、AI の発展に対して参考になる発想を見つけることができる。筆者はアリストテレスの古典思想から現代の問題を検討することで、新たなインスピレーションを得ることができるかもしれないと考えている。

アリストテレスの思想と AI 技術に関わる先行研究は、主にアリストテレスの論理学と実践学の観点から考察したものである。というのも、その論点の多くはアリストテレスの論理思想、範疇論および実体―属性、奴隷論、魂、理性、徳倫理などに集中している。本稿の目的は、主に先行研究を踏まえた上で、アリストテレスの理性および徳倫理思想を中心に、アリストテレスの古典思想と現代 AI 技術との関連性を分析することである。AI はただ人間を模倣しているにすぎないという見方がある。アリストテレスの理論から分析すると、人間にとっては AI 技術がある種の道具のような存在であり、本質的にある種の所有物である。道具の中では、あるものは生命のあるもの（奴隷など）であり、あるものは生命のないものである。知能機械は勿論生命のない道具に属している。そして、機械のような道具は人間の命令を聞いてその指令される仕事を遂行する能力が備わる。たとえば、岩田靖夫 (1) の研究においては、そのような能力は「道具的理性」と呼ばれる。だが、アリストテレスによると、人間は生まれつき理性的動物であり、その理性が技術的、道具的理性ではなく、倫理的価値（善悪、正義などを判断できる）を認識して実践する理性である。以上の観点を踏まえて、

本稿は大きく以下の二点から構成される。まず、アリストテレスの理論を用いて、道具的理性の所有と本来の人間的理性の所有とは同じものではないが、その両者にいかなる繋がりがあるか、そして互いにいかなる影響を与え合っているかを検討する。その上で、アリストテレスの言う主人と奴隷の関係が人間と知能機械の関係と一致しているかどうかを考察する。次に、人工知能の発展は道徳的基準から離れないことから、本稿はアリストテレスの徳倫理学に基づき、AI技術の発展と問題に対してアリストテレスの徳倫理学がいかなる示唆を与えられるかを検討する。

具体的には、本稿は、まず研究背景と先行研究の検討を通じて問題提起を行う。次に、アリストテレスの理性論に基づき、道具的理性と人間的理性を詳しく分析する。最後に、アリストテレスの徳倫理学の思想を用いて、道徳的人工知能の開発に向けた提案を試みる。

1．先行研究

アリストテレスとAIをめぐる先行研究は主にアリストテレスの奴隷論と理性論、論理学、徳倫理学という三つの分野に集中している。

まず、岩田の研究は主にアリストテレスの奴隷論を用いて、「理性」という概念をめぐり、AI技術[(2)]の本質に対する理解を述べた。岩田によれば、アリストテレスの奴隷論は二つの本質的特徴がある。まず、自然に基づく奴隷と自由人との差別が「理性の薄弱」にある。アリストテレスの理論によれば、奴隷とは他者の道具、他者にひたすら奉仕する者である。つまり、理性をほとんど持たず、肉体労働のみによって生きる者が自然に基づく奴隷であり、これに対して理性を持つ者が自由人である。

次に、人間は単に生きることではなく、よく生きることを目指している。というのも、人間が望んでいる生活は単なる飲んだり食べたり動物的な欲望を満足させる生活ではなく、徳を修得し、ポリスにおける文化的・政治的な欲求も実現できる生活だからである。岩田はこのような人間の生活がアリストテレスの言う理性による生活だと指摘した。だが、人間の文化的・政治的な生活を営みうるために、とりあえず人間は基本的な生存を確保しなければならない。これがよく生きるための前提条件である。問題の矛盾は、もし主人でも奴隷でも同時に「単に生きること」と「善く生きること」という二つの次元にいたら、誰でも二つの次元

において同時に十分に活動することができなくなるということである。そのため、生活必需品の提供とそのための労働は奴隷が担う必要がある。

岩田によれば、アリストテレスは奴隷が理性なき者であり、そして生活の生産と供給を担うため、アリストテレスの本性的奴隷論の論拠だと考えている。アリストテレスによると、人間の幸福は理性の活動による有徳的、政治的、文化的な生活の内にあると思われて、人間はそのような幸福的な生を追求するために、生活必需品の生産などより低い次元の仕事に従事する余力がほとんど残らなくなる。岩田によれば、「だれが、単に生きるための労働を背負うのか」という点がアリストテレスの奴隷制の根本的な問題であり、アリストテレスはこの労働を担当する者は必ず「本来徳と幸福から除外されるべき者」であり、これは「本性的奴隷」が存在する理由だと考えている。

アリストテレスによれば、人間の生きることを維持するために、様々な道具が必要である。そして、あらゆる技術にはその技術を実践するための道具が必要であり、道具の内にはあるものは生命のないものであり、他のものは生命があるものである。具体的に言えば、道具の分類は主に三種類があり、すなわち奴隷と、普通の動物と生命のない道具である。そこで、AI 技術は人間にとって、ある種の生命のない道具である。これに対して、奴隷と他の普通の動物は生命のある道具である。その三者の区別について、アリストテレスは以下のように述べた。「奴隷は〔主人の命令を〕認知する程度にはロゴスに与るが、ロゴスを所有することはない。これに対し、他の動物たちはロゴスを認知せず、ただ、感覚的受動によって奉仕するのである」（Pol.1254b22-24）。岩田によれば、アリストテレスは奴隷が主人の指示に従って実行するだけの理解力を持っている点が、動物や生命のない道具との違いだと考えている。そして、生命のない道具（コンピュータや自動車）はただ物理的あるいは自然因果的に使用される。動物は感覚的受動（受苦、パテーマなど）によって行為する。どちらでも十分な理性所有者ではない。奴隷は命令を了解すれば、その命令に基づいて能動的に行為することができる。そのため、岩田によれば、アリストテレスは奴隷が高度の道具的理性の所有者だと考えている。ただし、岩田によると、実際にアリストテレスは自分の理論において、奴隷が動物や命のない道具とは異なる自発性（理性）を持つことを認めることがある。岩田はこの点がアリストテレスの奴隷論の不整合だと指摘した。

岩田によれば、道具的理性とは自動機械で替えられうる能力であり、そして

道具的理性の活動は技術的活動である。ただし、岩田は原因結果の関係が一義的に決定されている技術の世界では本来の意味での理性（思慮）の動く余地はないという憂慮を示した。たとえば、計算機は人間の代わりとして計算の作業を行ってくれる便利な道具であるが、このことは計算という作業が本来の人間的理性を必要としない活動であることを示した。また、駅の自動改札とか、銀行の現金自動支払機とか、電車の自動運転とかも同じである。岩田はおよそコンピュータの関わる全ての仕事は、それが本来思慮を本質とする人間のする仕事ではないことを示しており、そしてコンピュータの関わる全ての仕事は思慮を本質とする人間の仕事ではないことを事実として現実化しているのである（機械をコントロールする少数の「思慮する人間」を除く）と指摘した。岩田は、アリストテレスによれば、「思慮する」という事態が生ずるのは、不確定な要素の存在する領域にあるものであり、もし全てが確定的なものであるとしたら、思慮は生じないと指摘した（EN1112a33-1112b4、1112b8-11）。岩田によると、技術は、一定の入力に対して常に一定の結果を生み出さなければならない。これが機械の本質的特徴である。技術は恒常性を基本原理としている。これに対して、元々アリストテレスが言った人間理性は不定不明の人間社会において、未知の他者と出会いながら、何をすべきかを選択する倫理的な能力であり、これによって、岩田は道具的理性の所有が本来の「人間理性」の所有とは認められないと考えている。奴隷は命のある道具であり、道具は命のない奴隷である。岩田は自動機械が道具的理性にすぎなく、アリストテレスが「思慮」と呼ぶ「実践理性」を欠いていると考えている。

石川義正[(3)]も岩田の考え方を認めており、AIは人間をただ模倣しているものだと主張した。石川も岩田の論述に同意し、アリストテレスの理論に基づいて奴隷を自動機械で代替しうると考えている。また、石川はもし道具のそれぞれが、命令されたり、自分で事前に察知したりして、自分自身の動きを成し遂げることができるなら、アリストテレスのこの空想はちょうど今のAIが実現していると指摘した。筆者は岩田の言う自動機械は奴隷と同じように、主人の命令を認知する程度にはロゴスに与るが、ロゴスを所有することではなく、ただ道具的理性を持っているという見方を認めている。ただし、岩田の研究においては動物が感覚的受動によって奉仕することを示したが、具体的に動物（人間を含む）はいかにして感覚によって行為を起こすかは言及しなかった。そのため、本稿の第二節で

は、人間や AI 機械がいかにして感覚を通じて行動判断を行うかという論点に着目する。

その一方で、銭圓媛 [4] の研究は、主にアリストテレスの理論が AI 技術の開発に与える影響を強調した。今から見れば、二千前にアリストテレスが提出した論理思想は現代の基準によって、不健全なところがあるかもしれないが、おそらく依然として人工知能理論の基本構造に影響を与えていると言える。銭によれば、人間はアリストテレスの範疇思想および関連する実体――属性論を勉強することによって、分類やプログラミングから人工知能に至るまでよりよく理解することができる。その同時に、アリストテレスの徳倫理思想もまた、人工知能に伴う諸徳問題に功利主義や義務論とは異なる道を提供していると考える。

銭は、アリストテレスの理論に基づいて AI に対する影響を次の三つに分類した。一つ目は、アリストテレスの論理学における三つの思考法則（Law of Thought）の影響であり、それらは、我々人間の考え方の常識的なルールである。銭によれば、アリストテレスの初期の「二進法の論理思想」と現代の電子回路工学の原理は似ている。たとえば、電気は流れることができ、流れないこともできる。スイッチもオンとオフを切り替えることができる。これは、アリストテレスの命題における真と偽という二値原理の前提と一致している。命題Ｐの“Ｐ ∨ ¬Ｐ”（Ｐ であるか、または Ｐではない）に関しては、矛盾律（Met.1006b13-22）：対象ビットが１であれば、０にもなり得ないことと、排中律（Met.1011b23-28）：各ビットについて、必ず１または０のいずれかでなければならないことと同一律（Met.1006a29-b28）：１のビット値は１であり、各ビット値は、それ自体が同一の値を持つことという三段論法に従わなければならない。このような前提がなければ、伝統的な数値計算の意味がなくなるだろう。しかし、銭はこれらの仮定が必ずしも量子コンピュータやアナログコンピューティングのハードウェアに当てはまるわけではないことも指摘している。また、アリストテレスは初めて体系的に三段論法（大前提、小前提、結論）を研究した。たとえば、大前提：全ての人間は死すべきものである。小前提：ソクラテスは人間である。結論：ゆえにソクラテスは死すべきものである。無論、アリストテレスの三段論法には間違った形があり、そして今の哲学者や数学者たちは、一般的にアリストテレスの三段論法を述語計算の特殊なケースだと見なしているが、最も重要なのは、論理的に推算しようとする機械は、必ず何らかの形でこの三段論法を推理の一部として使用す

る必要があるということである。

ここで筆者が補充したいのは今の人工知能をめぐる「記号主義対コネクショニズム」の論争があるということである。一方、記号主義とは、主に離散的な記号の形式的操作、すなわち計算によって知能を実現しようとする主義である。常識推論等を、古典論理を拡張することによって実現しようとするいわゆる論理主義はその典型である。典型的な形式的知識表現方法は、一階述語古典論理に基づいた知識表現方法である。他方、コネクショニズムとは主に神経回路網のモデルを用いて各種計算を実行することによって心（知能）を実現しようとする主義である。並列処理と分散処理を特徴としている[(5)]。記号主義的コンピュータの中にはアリストテレス論理学の原形が見えているが、コネクショニズム的なコンピュータは述語論理的な記号操作をしているかどうかは議論の余地がある。筆者は、コネクショニズム的なコンピュータは述語論理的な記号操作をする可能性があると考えている。なぜなら、三段論法は人間が脳によって推論し、正しいと思う結論をもらえる正確な思考方法である。また、人間をシミュレーションされた人工神経網は機械学習によって、入力された情報から推論や処理して結論を導き出す必要がある。コネクショニズムは、人間の神経回路網のモデルを模して計算を行い、人間の認知が完成するためには、他の様々な要素を必要としているが、三段論法という最も基本的な形での演繹的推論も明確に含んでいるからである。

アリストテレスの論理は実際の人間の思考に関連している。たとえば、アリストテレスの論理に基づく演繹データベースは自然な言語でその推論を説明することができる。また、これが人間のユーザーとの対話に焦点を当てた分野や、人間の思考のシミュレーション、特に人工知能に関連する帰納と確率の問題にも適用される。したがって、銭はアリストテレスの論理学思想がその利点のために計算に使用されると指摘した。

二つ目の影響は、アリストテレスが『カテゴリー論』において、存在するものに対して最も普遍的なかつ完全なカテゴリーのリストを提出したことと関係している。すなわち、実体、量、質、関係、場所、時間、態勢、所持、能動、受動を含みながら、各カテゴリーを明確的に定義した（Cat.1b25-27）。そして各カテゴリーのものは、カテゴリーの本質的な定義（普遍的な形式）または特性を共有している。その中で、後の九種類と最初の一種類（実体）との関連付けは、「実体――属性」の関係に分類される。実体は独立して存在できるが、属性は必ず実体

に依存して存在する必要がある。機械学習における機械学習のアルゴリズムは、データ（data）の共通性または一般性（フォームあるいは信号 signal とも呼ばれる）を勉強し、そしてデータから一般的な共通のものと特殊なもの（noise とも呼ばれる）を区別する。つまり、物事の本質的属性と偶有的な属性との分離である。銭は優れたプログラミングを作るために、これが非常に重要なものだと指摘した。ただし、デジタルバイナリに伴う悪影響の原因も、アリストテレスの理論の中で発見することができる。世界の「バイナリ化」の進展に伴い、我々の日常生活では微小な差異を失って悪影響を与えている。たとえば、今のソーシャルメディアアーカイブやデジタル注文などでは、「標準化のサービス」がそれぞれのユーザーの特別なニーズを満たすことができず、柔軟性に欠ける場合がある。

三つ目は、銭によれば、人工知能の発展に伴い、将来人間が直面する問題を解決するために、最も重要なことは人工知能自体をめぐり、倫理的な次元で人工知能を構築することである。この「知能」は、機械が人間と同じような理性を持っていることではなく、人工知能が人間の倫理的期待に従って行為することである。したがって、銭は道徳をコード化する必要があると指摘した。筆者もその主張に賛成している。

Steele [(6)] の研究も銭と同じように、アリストテレスの論理学思想が AI 技術の基層構造に対して影響を与えることを指摘した。Steele によれば、現代の人工知能の発展における諸論争は、アリストテレスの思想を遡ることによって示唆を得ることができる。そして、Steele は現在主流の考え方とは違い、人工知能は新しい概念ではないと主張した。なぜなら、人間は古代から機械との関係という問題に取り込んでいるからである。たとえば、ギリシア神話には、市民を不必要な危害から守るために、クレタ島の海岸線をパトロールした Talos と呼ばれる青銅制の自動人形があるという話がある。実際、一部の評論家は、AI の基礎となる基本的な論理学の原理はアリストテレスの思想から派生したものだと主張した。簡単に言えば、アリストテレスは、ピタゴラス派が提案した幾何学に関連する二進法を全てのもの（人間を含む）に適用しようとした。このシステムは、たとえば「はい - いいえ」、「有限 - 無限」、「男性 - 女性」、特に「1 - 0」に対する計算の反応時間を減らした。ただし、システムのシンプルさは非常に価値があるが、コンテクスト、豊かさ、ニュアンスなどを失ってしまうリスクもある。アリストテレスの思想は、我々人間に利用できる膨大な知識ベースを提供している。たと

えば、アリストテレスは、「eudaimonia」（幸福）や「eu zēn」（よく生きる）の概念についても幅広く論じた。AI は人間がこれらの目標を達成するのに役立つ大きな可能性を秘めている。しかし、AI を発展する過程には、様々な倫理的問題をはらんでいる（たとえば、どのような倫理原則に従うべきか？ 自動化のために仕事を失った人々はどこで新しい仕事を見つけることができるか？ アルゴリズムを開発するときに如何やって長年の偏見を回避すべきか？ ロボットにどんな権利を付与すべきか？）。Steele はこれらの疑問に対して、簡単な答えがなく、議論は何千年も続く可能性があると述べた。しかし、Steele の研究は、アリストテレスの思想が人工知能の発展に対し、いかなる影響を与えるかについては具体的に言及しなかった。

そのほか、Berberich と Diepold [7] の研究は主に、アリストテレスの徳倫理思想を用いて、機械倫理学の構築に向けた方法を分析することに着目している。Berberich と Diepold によれば、現代の AI やロボットシステムは、高度な自律性を特徴としている。同時に、自動運転車やサービスロボット、デジタルパーソナルアシスタントなどの分野での応用は、より人間に近いものとなっている。この進展に伴い、AI に関わる倫理的問題の出現が増えつつある。AI に関わる倫理的問題の解決策を見つけるために、両者はアリストテレスの徳倫理から、インスピレーションを得ることを提唱している。そして、両者はアリストテレスの徳倫理学の具体的な徳目（phronēsis、dikaiosynē、sōphrosynē、andreia、praotēs、philanthrōpos など）を分析することで、次の結論を導く。

まず、両者は徳倫理が現代の人工知能研究に極めて適した理論であり、道徳理論が AI 倫理の分野の基礎になると期待している。次に、徳倫理に基づいた道徳的機械の開発は、可能な行動についての単純な決定理論的な判断をするだけでは足りなく、より広いアプローチを取ることが必要である。つまり、注意、感情、学習、行動などの他の認知機能も考慮する必要がある。また、両者は徳倫理がコントロール問題（control problem）と価値整合の問題（value alignment problem）という二つの現代の AI セキュリティの主要な課題に対処する可能性を秘めていると指摘した。筆者の理解は、まず、コントロール問題とは AI の制御問題であり、この制御には、人間が機械を制御することと、制御システムによる機械自体の制御の両方が含まれる。しかし、超知能の出現は、人間や制御の問題に脅威を与え、予測できない影響がもたらされる可能性がある。過度な欲望を持つ機械は、人間

にとって有害なものである。そのため、両者の研究においては、「節制」の徳を用いて、将来起こりうる超知能の制御問題に対する解決策を見つけることができると主張した。次に、価値整合の問題は、主に AI を人間の価値観や倫理観に適合させる研究である。AI は機械学習の時、機械に入れる道徳的目標が人間の道徳的目標と一致する必要があり、機械に入れる道徳的目標が人間の道徳的目標と一致していないと弊害が生じることもある。

銭も Berberich と Diepold の主張を認めている。具体的に言うと、Berberich と Diepold によれば、道徳的 AI の構造は、AI が人間と同じ徳倫理思想を持つことではなく、AI が人間の徳倫理の要求と一致する行動を可能にするものである。アリストテレスによれば、倫理の領域において、人間は経験的な直感に頼って判断するのであるから、この考えに従えば経験から学ぶことを重視すべきだということになりそうである。これは、現代の人工知能との適切なコンビネーションを形成している。また、Berberich と Diepold によれば、道徳の模範例の模倣による学習は徳倫理学の中心概念であり、人工知能に関わる価値整合の問題を解決することが可能になる。そして、節度と人間に対する友愛を備えた知的システムは、自己改善のための無制限な欲求を持たないので、人間にとってのコントロール問題が生じる可能性がないことを証明できるかもしれない。

以上にまとめた三つの先行研究とも、倫理的な次元において、道徳的人工知能を構成する重要性を強調した。筆者もその考え方を認めている。そのため、本稿の第三節において、アリストテレスの徳倫理学の思想に基づき、道徳的人工知能の開発に向けた提案を試みる。

2．道具的理性と人間的理性

AI によって開発された機械はいろいろな種類があり、本論は主に AI が持つ「道具的理性」と人間が持つ「人間的理性」を比較しながら分析するために、本論において論じる AI は身体を持つロボット（AI を備えて環境内で活動するロボット）を典型例として考えている。筆者はこのような身体をもつロボットの AI は、外観から構造に至るまで人間と最も似ている AI のタイプであると考えている。

先行研究で述べたように、AI はただ「道具的理性」[(8)] を持つものであり、そして人間の理性を模倣しているものにすぎないということである。道具的理性と本来の人間的理性においてはいかなる繋がりがあるか、そして互いにいかなる影

響を与え合っているかを検討するために、まず、筆者は人間固有の動きから人間と身体を持つロボットとの違いを分析しなければならないと考える。アリストテレスの理論によれば、自然的物体の中では、生命を持っているものと、生命を持っていないものとがある。生命を持っているとは、自分自身による栄養摂取と成長・衰微ができるということである（De An.412a13-15）。生き物と非生物の最大の区別は魂（psychē）の有無である。魂はある種の生き物の内在的機能であり、生物の種差に応じて魂の能力も発展している。アリストテレスが示した魂の諸能力は五つあり、すなわち栄養摂取する能力、欲求する能力、感覚する能力、場所移動する能力、思考する能力ということである[(9)]（De An.414a29-32）。人間だけがその五つの能力を全部備えている。これらの能力においては序列がある。全ての生物にとって、最も基本的な能力は栄養摂取する能力である。生き物は通常の生活活動を確保した上で、他の能力を保全することができる。アリストテレスによれば、人間と他の生物（植物と普通の動物）は同じ栄養摂取や運動成長の機能が備わる。そして、一般の動物も人間と同じように感覚する能力が備わる。以上を排除したら、アリストテレスは人間に固有の動きが理性を持つもの（魂の部分）であり、すなわち思考する能力である。この理性を持つ魂の部分というのは、一方の意味では理性に従うという意味であり、他方の意味では自分が理性を持ち、思考する（echon kai dianooumenon）という意味である（EN1098a1-5）。そのため、アリストテレスは、魂を三種類に分類し、つまり栄養を取る段階の魂は植物の生活であり、感覚する段階の魂は動物の生活であり、思考する段階の魂は人間の魂である。

アリストテレスは人間の自然本性について次のように定義した。人間は生まれつき理性的動物（zōion logon echon）であり、また政治的動物（zōion politikon）である。その人間の理性は魂の中に存在しており、基本的に二つの部分に分けている。すなわち、理性を持つ部分と理性を持たない部分である。あらゆる生物の栄養摂取と成長運動を引き起こすのは、ちょうど魂の中で理性を持たない部分である。この部分は、あらゆる生物が普遍的に共有されており、また植物的なものであるため、人間としての徳に属していない。そして魂の理性を持たない部分には、ある仕方で理性に与る部分も存在している。アリストテレスは人間の魂の構成について次のように述べた。

「さらに魂にはまた理性をもたない何か別の本性のものが属しているが、ただしこれはある仕方で理性に与っていると思われる。というのも、抑制ある人と無抑制な人のいずれもがもつ理性をわれわれは賞賛するが、これは魂の理性をもつ部分を賞賛するからである。なぜなら、この理性こそ最善のものへと正当にも人を促すものだからである。だが、これら抑制ある人にも無抑制な人にも理性以外の別の何かがまた本性的に潜んでいるように見え、これが理性と闘い、抵抗するのである。ちょうど、体の麻痺している部分を右に動かそうとしても、どうしても反対方向に逸れてしまうように、魂についても同様のことが生じるのである。というのも、無抑制な人たちの欲求は〔意図するものとは〕反対のものへと向かうからである。だが、体の場合には逸れていくものが目に見えるが、魂の場合にはそれが見えない。そうだとしてもやはり、おそらく魂の場合にも、理性とは別の何かがあって、これが理性に反対し抵抗しているのだと認めなくてはならない。だが、実際にそうした部分がどのように区別できるのかということは、さしあたり問題ではない。しかしながら、すでに述べたように、この部分もまた〔理性をもつ部分と同じく〕理性に与るように思われる。現に、この部分は抑制ある人にあっては理性に従順であり、さらに、節度ある人や勇気ある人においては、おそらくそれはいっそう理性に従順に耳を傾ける。なぜならこうした人々において、この部分は理性と完全に符合しているからである。」（EN1102b13-28）

そのため、人間は理性的な存在者になれるかどうかは、第一義的にそれ自身のうちに理性を持つ部分を除き、より重要なのは理性を持たない部分における理性を分有する部分が理性を持たない部分に勝つことができるかどうかである。アリストテレスはこの過程を父親の言葉に対して耳を傾けることに喩えた。また、アリストテレスは人間の魂における理性を持つ部分にも同じ仕方で区分した。すなわち、「一方は、あるもののうちその原理が他ではありえないようなそうした事柄を、それによってわれわれが考察する部分であり、他方は、それによって他でもありうるような事柄を考察する部分である。」（EN1139a6-8）。前者は不変なものであり、後者は変えうるものである。そして、前者が「認識的（epistēmonikon）部分」、後者が「推理的（logistikon）部分」と呼ばれることもある。

AIの道具的理性は身体を持つロボットに内在する制御システムによって支配される。その道具的理性は認識的部分を持ち、つまり、身体を持つロボットにとってはコンピュータが人間の魂に相当する。そして、コンピュータは様々な算術的推論ができるが、算術的推論を行う流れは不変なものであり、すなわち原理は変わらないので、道具的理性は人間的理性のような推理的な要素を持たない。

アリストテレスは徳を「思考に関わる徳」（dianoētikē aretē）と「性格に関わる徳」（ēthikē aretē）という二種類に分けている。その思考に関わる徳における魂が肯定と否定によって真理を明らかにするあり方は五つのものがあり、すなわち、技術（technē）、知識あるいは専門知（epistēmē）、思慮（phronēsis）、知恵（sophia）、知性（nous）である（EN1139b15-17）。身体を持つロボットなどのAIはコンピュータによって設定された固定のプログラムで認識することが実現でき、そしてこれらのプログラムは不変なものであるため、アリストテレスの理論によれば、AIは本質的にある種の専門知（学的理解や知識と訳されることもある）と技術である。なぜなら、まず、アリストテレスによれば、専門知によって考察される対象は、次の二つの点を満たさなければならない。すなわち（1）学的理解の対象は必然的なものであり、永遠的なものである。（2）全ての学的理解は教えられうるものと考えられており、学的理解の対象は学ばれうるものである。次に、（1）技術とは、真なるロゴスを伴った制作に関わる性向の一種である。（2）全ての技術はものの生成に関わり、技術の勉強はそうあることもそうでないことも可能な何かあるものをどうしたら生じるようにできるかを考察することであり、そして技術の原理は制作する側を持っており、制作される側を持たない（EN1139b14-1140a23）。AIに関連する知識は、ある種の客観的に存在する自然法則であり、そしてそれらの知識は人間が学び、教えることができるものである。また、人間はAIを作る技術を持っている。そのため、筆者はAIが本質的にある種の専門知と技術であると考えている。その一方で、人間の理性だけは、上記の五つのものを通して変化するものと変化しないものを完全に理解することができる。

身体と魂の関係について、アリストテレスは、臓器を身体から切り離すことができないのと同じように、魂も身体から独立して存在することはできないと主張した。アリストテレスは目の例でその原因を説明した。

「さてしかし、以上述べられたことを、身体の諸部分についても当てはめて考察しなければならない。すなわち、もし眼が動物であるとすれば、視覚がその眼の魂ということになるであろう。なぜなら視覚は眼の説明規定に対応する意味での実体であり、これに対して、眼は視覚の素材であって、視覚が離れ去ると、もはや眼ではないからである。「眼」と呼ばれるにしても、それは「石製の眼」や「描かれた眼」が「眼」と呼ばれる場合のような、同名異議的な意味でしかないのである。

そこで、部分について成り立つことを、生きている身体の全体に当てはめて理解しなくてはならない。というのも、視覚という感覚の部分と眼という身体の部分との関係は、感覚全体と感覚する能力をそなえた――そのように特定されるかぎりでの――身体全体との関係に対して類比的だからである。ただし魂を失ってしまったものは、生きることへの可能態にあるもの［生きることが可能なもの］ではなく、むしろ魂をもっているものがそうなのである。また、種子や果実は、可能的にそのような物体なのである。

こうして、一方では切断作用や見る活動が現実態であるのと対応する意味において、覚醒していることもまた現実態であるが、他方では視覚能力や道具の能力がそうであるのと対応した意味において、魂は現実態なのである。これに対して、身体は可能態にあるものである。しかしながら、瞳と視覚能力とで眼が成立するように、先の場合でも、魂と身体とで生物が成立するのである。

このようにして、魂がその身体から離存するものではないこと、あるいはその魂の一定の部分は――もしも魂が本性上可分的であるとすれば――離存しないということは、明白である。なぜなら、魂の部分のうちのあるものについては、その現実態はまさに身体の部分そのものに属するからである。しかしながら、少なくとも魂の若干の部分がいかなる身体の現実態でもないとすれば、そのことによってその部分が身体から離存することには、何の妨げもないのである。」（De An.412b17-413a7）

魂を身体から切り離れないと同じように、AI（たとえば、身体を持つロボット）の場合もプログラムと機械（ロボットの身体）を切り離すことができない。プログラムがないと、機械も正常に動かない。そこで、筆者は AI プログラムが人間

の魂に相当し、機械が人間の身体に相当すると想像できると考えている。機械の本体は質料で AI 技術実現の可能態であり、そして機械に内在するプログラムは形相で AI 技術実現の現実態である。

アリストテレスは「魂は、いわば生物の始原［原理］だからである」（De An.402a6-7）と述べた。魂を持つものと魂を持たないものとの相違は運動変化と感覚にある（De An.403b25-27）。人間にとって、魂とは可能態において生命を持つ自然的なものであり、形相としての実体である。このような形相としての実体は現実態である。そのため、魂とは可能態において生命を持つ自然的なものの現実態である。その現実態は二つの意味があり、すなわち「知識の所有」と「知識を行使する」である（De An.412a9-11）。ここで、アリストテレスは魂が現実態における知識の所有という意味に相当すると考えている。ただし、非生物の知能機械はこれらの機能を備えていない。我々は生き物のモデルによって想像してみると、機械のシステム（機械に動力を供給するもの）を維持している電流を食料に相当し、機械内部のプログラムを栄養摂取する能力に相当することを容易に想像できる。電流とプログラムは機械を動かし続けることの基礎である。

筆者は上記の人間の理性に関する分析から、機械に完全な人間の魂の能力が備わっていないが、AI 技術によってある程度の人間の魂のような能力を持つことが可能だと考えている。なぜなら、まず、魂は第一義的な意味において、人間が生きること、感覚すること、そして思考することの保証だからである。一方、それぞれの感覚にとっては固有の感覚対象があり、たとえば、視覚と色、聴覚と音、嗅覚と臭い、味覚と味などである。他方、複数の感覚に感じられる共通のものもあり、たとえば、動と静止、数、形、大きさなどである。それらの感覚の中で、触覚は最も基本的で普遍的な感覚である。あらゆる動物には触覚を備えていることが最低限要求される。触覚は対象を感覚する際に、中間の媒体の作用によってではなく、中間の媒体と同時に感覚を認識することである。アリストテレスによれば、動物は感覚器官が備われるからこそ外部の世界を感知することができる。ただし、感覚は感覚器官によって直接生成されるものではない。感覚は身体の生まれつき持っている中間の媒体によって生じるものである。たとえば、肉、空気、皮膜、水など全ての感知に効くものは中間の媒体に含まれている。アリストテレスは、人間の五感はそれぞれ異なる方法で物事を感知すると考えられる。つまり、一方、触覚や味覚は直接に物事を接触することで中間の媒体と同時に感知できる。

人間の触覚器官は主に筋肉と舌に関連している。触覚は筋肉にある網状なものや皮膜のようなものから生み出される。筆者はここでアリストテレスはもうすぐ「神経」という概念を提出しそうであると考えている。他方、視覚と、聴覚と嗅覚とは必ず中間の媒体によって生み出される。現代の AI 技術は人間の五感を模倣して機械に感じさせることができる。具体的に言えば、視覚による画像認識、顔認識であり、聴覚による音声認識であり、味覚や嗅覚による電子舌と電子鼻（有害物質や有害ガスの識別）であり、触覚による AI 手袋などである。

魂のうちには、行為と真理を決定する三つの肝要なものがあり、すなわち感覚と知性と欲求とがある（EN1139a17-18）。アリストテレスによれば、人間が快楽と苦痛を感じる理由は感覚する能力を持っているからである。もし動物が感覚を持っていれば、欲望も持っている。感覚が存在するところには快楽と苦痛も存在し、必然的に欲望も存在するからである（De An.413b21-24）。人間の欲望は快楽を希求していることである。生き物の感覚は、動かされることや作用を受けることによって生起する。

AI の開発によって、機械は外部刺激を感じることが可能になる。しかし、知能機械の「感覚」は主にプログラムによって制御される。この「感覚」はマニュアル通りのものであり、人間のように情緒の変化や感情に左右されることも全くない。明らかに AI テクノロジーの感覚は人間自身の感覚とは異なっている。アリストテレスによれば、人間は知識や魂が必要であり、その中のいずれかによって、人間が物事を認識できるようになる。アリストテレスが挙げた例が示すように、人間が健康を得ることでも二つのものからなり、すなわち健康と身体ということである。そして健康の方は知識と同じで一種の形態や形相や定義であり、知識の方は知識を得ることの、健康の方は健康になりうることの現実態である（De An.414a4-12）。機械の構成もそれと似ている。機械はプログラムと機械の本体で構成されており、機械が持たせる技術やシステムは人間の魂に相当し、機械の本体は人間の身体に相当する。人間の感覚する能力は理性を持たないとも理性を持つとも容易に定められない部分である（De An.432a30-31）。機械が持つ欲望と人間が持つ欲望は本質的にその由来が異なっている。なぜなら、一方、人間の欲望は主に感覚によって生じるものからである。これとは違い、強い AI（strong AI）主義によれば、将来、独自で問題を推論して解決できる真のインテリジェントな機械を作成することが可能であり、そしてそのような機械は感覚的で自己

認識を備えるものだと見なしている。そのような予測が実現されれば、機械は人間を凌駕し、制御不能になるリスクがある。そのため、機械の欲求は主に技術の制御不能性からもたされてきた。

以上の分析を通じて、アリストテレスの言う人間の魂にある五つの能力から機械を考察すれば、機械には欲求する能力[(10)]がないこと、また運動する能力（機械の場所移動は人間とは違う方法で動きを生み出すが）と栄養摂取する能力（上記に挙げた機械と電流の例のような形）があることは容易に判断できると考えている。そして、道具的理性と人間的理性とは最も繋がりやすい能力がその感覚する能力と思考する能力である。というのも、AI 技術の目標は、機械が人間と同じように理性的に考えて行動できるようにすることである。そこで、AI にとっては、機械学習、特にディープラーニングの重要性を浮き彫りにしている。

以上の分析によれば、筆者は次の結論を導く。無論、知能機械は道具的理性を持つものである。そして、知能機械は人間と同じように、人間的理性を持つことはほぼ不可能であるが（思慮を欠くため）、理性の徳に符合する行為をすることが可能である（人間の理性的行為を模倣すること）。

アリストテレスの言う主人と奴隷の関係が人間と知能機械の関係と一致しているかどうかを分析してみよう。まず、岩田の研究においては、自然に基づく奴隷と自由人との差別が「理性の薄弱」にあることを指摘した。つまり、自然に基づく奴隷は人間的理性を十分に備えていない。次に、自由人は「単に生きること」だけではなく、「善く生きること」を追求している。奴隷は主人の「単に生きること」のための基本的な仕事を担うし、そして善く生きるための道具としている。同様に、AI によるシステムの開発に伴い、業務確率化と生産性向上、人件費削減、多くの人為的ミスの減少など、人間の生産や生活に多くの利便性をもたらした。機械は人間にとっても、「単に生きること」と「善く生きること」のための道具である。最後に、アリストテレスによれば、道具の分類は主に三種類があり、すなわち奴隷と、普通の動物と生命のない道具である。そこで、人間にとっては機械でも奴隷でも本質的にある種の道具である。両者ともただ道具的理性の所有者である。そのため、筆者はアリストテレスの言う主人―奴隷の関係と人間―知能機械の関係とが同じ繋がりを持つものだと考えている。

3．徳倫理と AI について

今は AI 技術が急速に発展している時代である。AI 技術の継続的な開発により、人間の生活に新しい便利さと挑戦を同時にもたらしてきた。筆者はアリストテレスの徳倫理学の思想を用いて、道徳的 AI の開発に向けた提案を試みたいと考えている。

先行研究においてまとめた通り、Berberich と Diepold の研究は主に三つの結論を導く。すなわち、

（1）Berberich と Diepold によれば、アリストテレスは倫理の領域では人間は経験的な直感に頼って判断するから、経験から学ぶことを重視すべきだと考えられる。これは、現代の AI との適切なコンビネーションを形成している。そして、Berberich と Diepold によれば、その経験が機械学習に由来し、徳倫理においては学習と習慣が中心概念であるため、機械学習と徳倫理学を組み合わせることが提案される。

（2）Berberich と Diepold によれば、道徳の模範例の模倣による学習は徳倫理学の中心概念であり、人工知能に関わる価値整合の問題を解決することが可能になる。筆者の理解から見れば、人間と AI の利害関係は合致できるものであるため、価値整合の問題の視点によって、AI は機械学習の時、機械に入れる道徳的目標が人間の道徳的目標と一致する必要があり、機械に入れる道徳的目標が人間の道徳的目標と一致していないと弊害が生じることもある。

（3）Berberich と Diepold によれば、節制と人間に対する友愛を備えた知的システムは、自己改善のための無制限な欲求を持たないので、人間にとってのコントロール問題が生じる可能性がないことを証明できるかもしれない。人間とのコミュニケーションやインタラクションを機能としている AI は、人間に友好的であるから、無限の欲望で「コントロール不能」という事態に陥ることはない。また、節制は道徳的 AI 主体が持つべき美徳でもある。節制について、アリストテレスは「したがって、こうした欲望は適度な、しかもわずかなものでなければならず、また決して理性に反するものであってはならないのである――こうした状態を、われわれは従順で、仕付けを受けた状態と呼ぶのである――ちょうど、子どもが仕付け役の指示に従って生活しなければならないように、欲望的な部分も理性に従わなくてはな

らない。」（EN1119b11-15）と述べた。Berberich と Diepold によれば、それこそ、我々人間が望んでいる理想の道徳的 AI 主体である。すなわち、道徳的 AI 主体が子供のように、指導者である人間の指示に従って行動することである。機械に対する正しい道徳基準は、最終的には人間が定義すべきものである。筆者も両者の結論に同意する。コントロールの問題を解決するために、節制と人間に対する友愛を備えた知的システム構造が極めて重要である。超知能の挑戦に向けて、AI の過度な欲望を制御するために、人間の指示に従って行動することが必要である。

我々はアリストテレスの倫理学を通して、道徳的 AI 主体を構成するための方法を考えることだけでなく、道徳的 AI 主体の構成を考える中で、アリストテレスの徳倫理学を再認識することができる。上記の結論以外に、Berberich と Diepold はアリストテレスのエルゴン（ergon）の議論を考察することによって、よくある二つの誤解を指摘した。一つ目は、人間だけが道徳的行為を行えるという倫理的人間中心主義である。過去には人間が唯一の高度の認知能力を持つものであるため、これが正しい考え方であると考えられる。そして、アリストテレスによれば、高度の認知能力は徳倫理学の要求であり、つまり、道徳的行為の要求である。もし、人間と同じような理性と性格を持つ別のもの、たとえばある機械にそのような理性と性格があるとすれば、徳倫理の理論は機械にも使える。

二つ目の誤解は、アリストテレスの徳倫理学が幸福を全ての行動の目標と尺度にすることである。機械は心からの幸福を感じられないため、徳倫理学が機械に適用されないという議論がある。しかし、この議論は、「幸福」についての誤解に基づくものである。ギリシア語の幸福という言葉はより広い意味を持ち、主に自分のエルゴンに従って人生をうまく生きるという行為を指す言葉である。したがって、幸福は主観的知覚だけではなく、道徳的実在論に由来するものである。このような見方から、筆者はアリストテレスが倫理学において人間だけで研究主体としているが、人間と同様に理性と徳を備えた AI 主体が出現すれば、これらの機械にもアリストテレスの徳倫理学の理論を適用できると考えている。

筆者はアリストテレスの徳倫理と AI に関して、次のように考える。アリストテレスの倫理学においては、徳は理性に基づく徳と人柄に基づく徳に分けられており、一方では、理性に基づく徳は他人に教化することによって身に付き、他方では、人柄に基づく徳は日常の習慣づけによって養成できる。その教育と習慣づ

けが徳を獲得するための重要なルートである。現代の AI 技術の開発にとっては、機械学習・ディープラーニングが極めて肝心な研究分野であり、その中で、徳倫理の教育を機械学習・ディープラーニングに組み込み、AI が徳に従う行為を選ぶ習慣を持たせることが非常に重要である。

また、アリストテレスは、徳が実践に由来するものだと強調した。機械倫理学の構築は、実践を重視すべきであり、実践と切り離せない関係にある。筆者は岩田の言う機械が人間と同じ思慮する能力を模倣することが難しいという見方も認めている。なぜなら、アリストテレスによれば、思慮ある者は実践的な知恵を備えている者だからと考えられる。思慮とは善と悪に関連して真実を求める実践的な品質である（EN1140b4-7）。そのため、思慮を備えた者は自分自身に関わることをよく対応することだけではなく、周りのこともうまく調整できる者である。たとえ周囲の状況が変化しても、変化しなくても、彼らは柔軟に対応できる。思慮する能力を持つ人間は、物事の原理をはっきりと知っており、ある種のものに対して絶対に分かる状態であり、まさに人間的理性の一つの現れである。そして、道具的理性は本質的に物事の本質を理解することができない。これも道具的理性と人間的理性との違いである。しかし、思慮のある人は物事の判断に関わる様々な要素を十分に検討して行動する者であるが、人間が AI に頼りすぎると、全面的熟慮する能力が衰えてしまう可能性もある。

最後に、人柄に基づく徳においては、多すぎと足りなさすぎとの間に中間という徳がある。人間と AI との関係においても、その中間という徳も適用する。人間は過度に AI を依存すべきではないし、AI から完全に排除されるべきではない。人間は適度に AI 技術を利用したら、人間の生活もきっとより便利なものになる。同時に、人間と AI の関係において、人間にとっても AI にとっても正義の関係を築く必要があり、それを達成するために、道徳的な側面だけでなく、正義を保障する法律も必要である。その適度や正義をどのように実現すべきかについては、AI 技術の開発に伴い、今後も引き続き議論する問題である。

凡例

（1）アリストテレスの著作名の略語については、以下の通りである。

Cat.『カテゴリー論』

De An.『デ・アニマ』

EN『ニコマコス倫理学』
Met.『形而上学』
Pol.『政治学』

（2）この論文におけるアリストテレスのテクストからの引用については、岩波書店の新版『アリストテレス全集』、京都大学学術出版会（西洋古典叢書）の『魂について』を使用した。なお、一部修正を加えている。

註

（1）岩田（2010）、pp.83-108。
（2）岩田の研究では、AI（人工知能）という言葉を使わず、「自動機械」という言葉で表現している。岩田が挙げた全ての例はコンピュータによってコントロールされるものであるため、筆者はその「自動機械」と現在の AI 機械とは同じ意味のものとして理解している。
（3）石川（2020）、pp7-17。
（4）銭（2020）、pp.229-236。
（5）月本洋（2000）、pp.29-41、科学哲学 33-2。
（6）Steele（2019）.
（7）Berberich & Diepold（2018）,pp.1-25, arXiv:1806.10322 v1[cs.AI] 27 Jun 2018.
（8）「道具」という言葉は古典ギリシア語のオルガノン（organon）に由来する。本稿で論じた「道具」は『オルガノン』という作品ではなく、普通の道具という意味である。
（9）欲求する能力は感覚する能力に含まれることもある。アリストテレスは欲求と感覚の関係について、次のように述べた。「…またもし感覚をもっているなら表象のはたらきと欲求ももっている。なぜなら感覚が存在するところには快楽と苦痛も存在し、それらがあるところには、必然的に欲望も存在するからである。」（De An.413b22-24）。「…すなわち、欲求には欲望と気概と願望とが含まれるが、すべての動物は、諸々の感覚のうちの少なくとも一つ、すなわち触覚はもっている。さて感覚がそなわっているものには、快楽と苦痛および快いものや苦痛なものが存在し、そうしたものが存在するならば、それには欲望もそなわっている。なぜならば、欲望とは、快い

ものを対象とする欲求だからである。」（De An.414b2-6）。

（10）筆者は将来 AI がディープラーニングによって感覚することができれば、人間の模擬的な欲求を備えることも可能になるかもしれないと考える。アリストテレスによれば、欲求は感覚に基づく欲求と思慮に基づく欲求という二種類がある。筆者はもし将来ディープラーニングによって、機械が人間のような感覚する能力と思考する能力を持っていれば、欲求を持ち得ることも可能になるかもしれないと考えている。

参考文献

L. Bywater, *Aristotelis Ethica Nicomachea*, the Oxford Classical Texts（OCT）, Oxford University Press, 1894

L. Minio-Paluello, *Aristotelis Categoriae Et Liber De Interpretatione*, the Oxford Classical Texts（OCT）, Oxford University Press, 1936

W. D. Ross, *Aristotelis De Anima*, the Oxford Classical Texts（OCT）, Oxford University Press, 1956

W. D. Ross, *Aristotelis Politica*, the Oxford Classical Texts（OCT）, Oxford University Press, 1957

W. Jaeger, *Aristotelis Metaphysica*, the Oxford Classical Texts（OCT）, Oxford University Press, 1957

Berberich,N., & Diepold,K.（2018）, "The Virtuous Machine -Old Ethics for New Technology? ", arXiv:1806.10322 v1[cs.AI] 27 Jun 2018, https://arxiv.org/abs/1806.10322,pp. 1-25

Steele,L.（2019）, "From Aristotle to Artificial Intelligence", https://www.linkedin.com/pulse/from-aristotle-artificial-intelligence-laura-steele（accessed2023-2-8）

アリストテレス著『カテゴリー論』中畑正志、早瀬篤、近藤智彦、高橋英海訳（2013）,『アリストテレス全集 1 ・カテゴリー論・命題論』、岩波書店

アリストテレス著『形而上学』出隆監修、山本光雄編（1968）、岩波書店

アリストテレス著『政治学』神崎繁、相澤康隆、瀬口昌久訳（2018）,『アリストテレス全集 17・政治学・家政論』、岩波書店

アリストテレス著『魂について』中畑正志訳（2001）、『魂について』、西洋古典叢書

アリストテレス著『ニコマコス倫理学』神崎繁訳（2014）、『アリストテレス全集15・ニコマコス倫理学』、岩波書店
石川義正（2020）、『政治的動物』、河出書房新社
岩田靖夫（2010）、『アリストテレスの政治思想』、岩波書店
月本洋（2000）、「コネクショニズムを超えて -Embodied AI-」、科学哲学 33-2、pp.29-41
銭圓媛（2020）、「亚里士多德与人工智能」、哲学与中国（2018 年春季卷）、pp.229-236