AIと観念的諸概念の関係

Relations of AI and Ideological Concepts

南　雲　　功*
Isao NAGUMO

要旨：「倫理」、「責任」などの観念的諸概念は、明確な判断基準が示されていないにも関わらず、社会活動の中で、重要な潤滑的役割を果たしている。一方、アルゴリズムのAIはその著しい進歩により人間の知的活動の一部を担っている。それゆえ、AIに観念的諸概念に関する判断が要請されることになる。この判断の是非について機械の「準則性」と人の「随意性」に着目し検討した。曖昧な観念的諸概念の判断を、AIに任せることは人間にとって危険であり、人間の判断が介在すべきである。

キーワード：AI（人工知能），倫理，責任，フレーム問題，トロッコ問題

1．序

情報機器の著しい発展に伴い、知的分野への人工知能（Artificial Intelligence, 以下 AI）の普及が目覚ましく、多方面で重要な判断の行使や判断資料の提供がなされている。一方、「倫理」、「責任」などの観念的諸概念は、明確な判断基準が示されていないにも関わらず、社会活動の中で、重要な役割を果たしている。そこで、論理思考の AI と観念的諸概念を必要とする人間社会が、いかに共生するかが今後の重要な課題となる。

2．機械の「準則性」と人の「随意性」

機械は予め定められた設定を超えて判断や行為をすることができないが、人は環境の変化や想定外の場面、利己的理由などにより、規約を超えて行動することが可能である。前者を、機械の「準則性」後者を人の「随意性」と称することにする。

2.1　AIの定義と「準則性」

「AI」という語は多義的である。（社）人工知能学会では「人間の知能そのものをもつ機械を作ろうとする立場、もう一つは，人間が知能を使ってすることを機械にさせようとする立場」（人工知能学会．2019）という二つの立場を挙げている。Searle, J. R. は前者を「自由意志」[i]を持つ

* なぐも　いさお　準研究員　放送大学在学

「強い AI（strong AI)」、後者を人間の知的活動を補佐する「弱い AI（weak AI)」と称している（Searle, J. R.. 1980.417)。現時点で強い AI は存在しない。

機械は論理的に整合している限り設定を遵守するが、設定を超えて行為しない。仮に AI が人と同じような「自由意志」を持つならば機械の「準則性」は崩れるが、現時点で深層学習などにおいても、「準則性」が維持されている。西垣通らも「あらかじめ設定された…以外のルールを自律的に作りだすことは不可能」（西垣通他. 2018）と述べている。

「準則性」はフレーム問題でもある。AI が人の「随意性」をもつならば、シンギュラリティ[ii]は現実となるが、人の尊厳に疑義が生じ本論の範囲を超える。

2.2 人の心、「自由意志」と「随意性」

自然主義的心脳一元論[iii]の立場では「自由意志」、「道徳」、などは脳内の電気化学反応が作り出した幻想ということになる。しかし法律、経済など社会科学の分野では、「自由意志」を持つ「人の尊厳」を基本原理としている。人間社会と AI の論理思考の共生を目指す本論の立場は、たとえ幻想であるとしても、人間の社会的原理として、「自由意志」の存在を前提とする。人は「随意性」の「自由意志」によって直面する環境の変化に順応して規則や習慣を超えた行動をとることができる。状況により臨機応変に危機を脱出し、逆に自己本位に違法行為を繰り返す。固定観念から逃れられないというような人間の一般化フレーム問題があるものの、この固定観念を超えることにより新たな芸術的創造、発明、発見が生まれており、人は機械のように規則に厳格に固定的ではない。

2.3 技術の目的は人間の諸能力の超越

AI が人間の能力を超えて人を支配する不安が議論[iv]されているが、人間が自己の活動の一部を超越する目的で様々な道具を開発しており、技術の目的は人間の能力を超えることにある。サンプター, D. も「私たちの文化はこれまでさまざまな問題を解決するため、数千年にわたってある種の数学的な『人工』知能を生み出してきた」（サンプター, D.. 2019.331）と述べている。記録としての文字、計算補助のそろばん、伝達としての手紙、思考手順の論理学や数学などは、質、量、速度などの点で人間の能力を超えている。コンピュータ技術もその延長線上にあり、人の知的能力の一部を超える。人間は、人の能力を超える技術を管理、支配することで文明を発展させてきた。この構図は AI についても変わらないであろう。

3. AI の特性

3.1 AI の「準則性」による法律厳守

AI の「準則性」により法律に矛盾がないかぎり、判例などの補正を含め AI に法律は遵守させることは可能である。道路交通法を人が制限速度・駐停車・車間距離等を自己判断していることを、AI が厳守するならば大幅に事故は減少するであろう。定常作業では、人の「随意性」より機械の「準則性」が優れている。ところが、異常事態においては、人間の「随意性」が勝ることがあり、その際、倫理的判断が要請される。

3.2 AI に欲望はない

AI は、交換可能なハード機器と、ネットワークを通じ自他の区別が曖昧なソフトにより構成される。さらに、電源の On-Off により再生、停止することから死の意味が失われる。このことから、AI が動物の生への欲望、支配欲を持つ必然性がない。従って、機械が人間を支配しよう

とする意識は無意味であり、人の視点からの杞憂にすぎない。現代の機械による人間の支配は、サンプター，D.（サンプター，D.. 2019）などが指摘するように機械を支配している人間による人間の支配である。

3.3　AI は新たに遭遇した知性体：道徳の不在

現有の AI でも、深層学習などのように論理計算が複雑化、高速化し人間の思考速度をはるかに超えて、情報収集、処理を行う。このことにより、人間の予想を超えた論理的判断がなされることは、囲碁、将棋の対戦にも現われている。欲望がなく、人の判断をはるかに超えた知性体は、あたかも神仏のような新たな知性体との遭遇となる。神仏と異なるのは、善悪の意識がない。すなわち「欲望」も「道徳」もなく能力は超人的な知性体である。その為、人の「倫理」や「責任」という観念的諸概念を AI に求めることは慎重であるべきである。

4. AI と観念的諸概念

論理的に判断する AI と人間社会との共生を考える場合、これまで曖昧にしてきた概念を AI に適用することはできない。「AI に責任がとれるのか」「AI に倫理的判断ができるか」「AI に創造性はあるか」の「責任」「倫理」「創造性」のような諸概念に対して、「準則性」をもつ AI にこれらの行為を求めるならば、諸概念の意味を明確に定義しなければならない。

4.1　責任

瀧川裕英（瀧川裕英．2003.25）の分類を基に、「責任」概念を、法的責任、倫理的責任、形而上学的責任に分ける。英語では法的責任（legal liability）は義務的意味が強く、その他の責任（responsibility）が応答の色彩が強くなる。法的責任が、「責任」の所在と、責任負担を法的に強制するのに対し、倫理的責任、形而上学的責任は、答責者の主体性に関わり、自らの意志により行為するので、倫理的色彩が強くなる。

4.1.1　法的責任

法的責任は刑事責任と民事責任とに分けられる。刑事責任について刑法の機能には規約的機能、秩序維持機能があり、規約違反者に対し、生命、自由、財産などの負担を課する刑事罰は人にとって抑止的機能があるものの AI にとって無意味である。AI の判断による犯罪行為は、「準則性」から故意犯はありえず、過失のみとなる。AI に関し秩序維持の法的機能を達成する為には、原因対策と被害の賠償が中心となり、民事責任に集約できる。法を守らない AI は製造者または所有者に責任がある。

民事的責任は、理念として自由意志に基づく行為に責任が生じるが、実際には現象でしか判断できない。AI の行為に対しても物件、債権等の民法上の主体となれば「責任」を負うことが可能となる。民法上の権利義務は、制度であり、現在でも法人という形式で人格の一部が組織に付与されている。また、現実の経済活動においてすでにコンピュータは ATM、自販機など人間と同等の経済行為を行っている。

一例として、自動運転車の事故において、所有者や製造者の直接の原因が不明である場合、両者に負担させることは、無過失責任となる。民主主義の原則から無過失責任[v]は極力避けることが望ましい。そこで、自動車損害賠償責任保険のように一律強制的に加入させ、賠償するような制度が望ましいであろう。法的責任者を、特定の業務に従事する AI 群に法的に人格の一部を付与する法人とするならば、個別の事故に対し AI 法人が賠償責任を負担することになる。すなわ

ち、必要に応じ制度を整えれば、AI群が法的主体になることは可能である。尚、違法行為、設計ミスは製造者、整備不良は所有者の責任であり、法を遵守しているAIの責任となる事例は稀有となろう。

4.1.2　倫理的責任および形而上学的責任

法的責任は、規約と負担が明文化されているのに対し、倫理的責任、形而上学的責任は責任の所在、問責者、答責者、具体的対応が不明確である。大庭健は、これらの「責任」に対して呼応関係を維持していくことと論じ（大庭健．2005.24）、藤垣裕子も科学者の社会的責任を「市民からの問いかけへの応答責任」で説明している（藤垣裕子．2018.55）。「責任」は答責者の意志に基づく応答であり、具体的対応を予め決められない。応答的な「責任」では、状況により価値が変化する為、明確な善悪も決定できない。まして幕引き引責辞任などAIに対しては無意味である。

また従来、「責任」は「自由意志」によって行った行為に「責任」が生じ、問責者に対する応答としてなんらかの行動をとらせると考えられてきたが、斎藤了文は、「複雑な人工物に囲まれた社会において、『責任者』と呼べるようなものは存在するのだろうか。作った人の責任を問うよりも、社会の大きなシステムを改良していくことが重要になってきた」（齊藤了文．2019.180）と述べているように複層化していく現代において、責任の所在や行為としての責任行動が不明確になっている。

形而上学的責任とは、宗教上の責任以外に、災害、戦争などの生存者が死者に対する自責の念による行為、心情などである。従って、AIには「責任」を負うことができず、必要もない。

4.2 「倫理」

倫理的判断の是非は、時代・地域・状況により異なる。行為に対する評価も様々である。自らの行為に対し常に自ら反省と批判し続けることに「倫理」の本質があることは、プラトン、カント、釈迦、孔子など多くの哲人が説くところである。AIの判断に対し人間が倫理的判断基準を示し、常に改良しなければならない。伊藤博文（伊藤博文．2018）は、ネットワークにより世界中の事例等から統計的に倫理判断の最適解を求めることの可能性について述べているが、時代や地域により異なる倫理判断を統計的解析の解が善となる保証はない。善悪の判断が多数決で決められないことも多くの哲人が述べることである。例えば、戦争のなくならない現実において、戦争に勝つことが善と判断されかねない。従って、常に人間の批判的介在が必要である。

また、人の判断が優位とされる緊急事態において、一般の人間であればいかなる倫理的判断を下すであろうか。訓練を受けた専門家でなければ、倫理的判断をする前に本能的自己保存に働くであろう。現実の緊急事態においても「準則性」により機械の方が冷静に対処できるであろう。技術の原則として、判断できない状況では、機械を停止し指示を待つことである。この場合でも被害者が発生するであろうが、現実の人間にできない倫理的判断を機械に求めるのはいかがなものであろうか。現状より悪化するのでなければ、一件ずつの対策の積み重ねで事故を減少させることが肝要である。

また、「倫理」判断の例題としてトロッコ問題が議論されるが、トロッコ問題は倫理的思考の教育としての例題であり、現実の技術では、二者択一的判断ではなく、様々な状況に応じた判断が求められる。従来は個人の技能に大きく依存してきた道路交通に関しては、自動運転とその他の技術で可能となることも多々あるであろう。自動運転車だけでなく、すべての自動車に緊急停止装置を義務付け、障害者を中心とした歩行者に対し発信機を所持しAIに注意喚起するなどが

考えられる。鉄道や航空機ではすでに多様な安全システムが採用され実績をあげている。近い将来、道路交通体系全体を見直す時期がくるであろう。倫理的判断は、機械に任せるのではなく、人間自らが構築、改良を繰り返すことが求められる。

むすび

観念的諸概念のような曖昧な概念を AI に適用することはできない。適用するならば、AI に理解できる形式で判断基準を設けなければならない。従って明文化されている法的責任については、適用しやすい。曖昧な諸概念については、人間が判断しなければならない。

最後に、AI の進歩はおそらく人間社会の安全が増大するであろう。安全の維持は、人工物や、自然を適正に管理することにより達成される。従って、個人情報等の掌握と個人を管理することが必要となる。安全と自由は負の相関にある。ダットン , W. らも、自由、安全、幸福のバランスが AI の進歩に不可欠と述べている（ダットン , W.. 2013.458-459）。安全に関しては AI の能力は力強い。しかし、自由は多義的であり個人の主観に基づく。しかも人間性の発露は、自由に由来する。AI が進歩するに従い AI と人との共生する社会への不断の努力が不可欠となりつつある。

文献（SIST ハーバード方式に準拠）

伊藤博文．2018. 人工知能と倫理．愛知大学情報メディアセンター．vol.28, No.1, p.13-23

大庭健．2005.「責任」って何に？．講談社

齊藤了文．2019. 事故の哲学．講談社

サンプター，D.. 2019. 千葉敏生，橋本篤史訳．アルゴリズムはどれほど人を支配しているのか？．光文社

人工知能学会．人工知能って何？．https://www.ai-gakkai.or.jp/whatsai/AIwhats.html（2019-6-17　閲覧）

瀧川裕英．2003. 責任の意味と制度．頸草書房

ダットン，W.，ジェイコブスティン，N.，チャーチル，E.，飯田弘之，前野隆司，紺野登，武部恭枝，サフォー，P.，アラン.，マラブー，C.，大澤博隆．2013. ダイアローグ「シンギュラリティの論点：人間の知性 vs. コンピュータ第 3 部　未来：人間とコンピュータの共進化．人工知能学会誌．Vol.28, No.3, p.453-464

西垣通，加納寛子．2018. AI と人間の尊厳ある自由．日本情報教育学会誌．vol.1, No.1, p.37-44

藤垣裕子．2018. 科学者の社会的責任．岩波書店

ホーキング，S.. 2019. 青木薫訳．ビッグ・クエスチョン．NHK 出版

堀浩一．2015. シンギュラリティへ向けてあなたと私はどうしたいか？．情報処理．Vol.56, No.1, Jan.2015, p.41-43

松田卓也．2013. 2045 年問題．廣済堂出版

村上裕子．2019. 人工知能の倫理と社会．人工知能．Vol.34, No.2

養老孟司．1989. 唯脳論．青土社

欧文文献

Kurzweil, R.. 2005. THE SINGULARITY IS NEAR: When Humans Transcend Biology. Penguin Books Searle, J. R.. 1980. Minds, brains, and programs. Behavioral and Brain, Sciences. Vol.3. p.417-457

註

[i] 「意志、意思」という用語は法律用語では「意思」、哲学、心理学などでは「意志」を用いるが、本論においては法体系より広い意味で用いるため「意志」に統一した。

[ii] シンギュラリティとは Kurzweil, R. が提唱した技術の特異点。2045 年に一台のコンピュータの処理速度がすべての人間の脳の処理速度を超え、人間の意識が機械はいる。（松田卓也．2013）など

[iii] たとえば養老孟司（養老孟司．1989）の唯脳論など

[iv] 2045 年問題（松田卓也．2013）やホーキング（ホーキング，S.. 2019）など

[v] 無過失責任とは、加害行為が故意又は過失がなくても不法行為を認める。公害や原子力事故など限定的である。