AI／ロボットの倫理と哲学

― カント批判哲学、「美意識」と「おそれ」について ―

伊　野　　連

はじめに　導入、AI とロボットとの違い

2022 年現在、「第 3 次 AI ブーム」と言われる時代が到来している。倫理学 [*1]、分けても現代社会の諸問題に対応する使命を帯びた応用倫理には、主要分野として情報倫理があり、また科学・技術・工学それぞれの倫理も議論されている。これらが複合的に絡み合い、「AI 倫理」ももう四半世紀ものあいだ議論され続けてきた。

本論文はその一環で、「二人零和有限確定完全情報ゲーム」、すなわちまずはチェス、あるいは将棋や碁などの棋士／プレイヤーと AI との対戦を題材としている（とはいえ、論者はチェス、将棋、碁のいずれにも不案内で、前二者はせいぜい駒の動かし方くらいしかわからず、後者に至っては碌にルールも知らないので、その方面で見当外れなことを言う場合があるかもしれないが、あくまで専門の哲学・倫理学に即して論じていく）。

周知のように、これら三種のゲーム／競技では既にいずれも AI に軍配が上がっており、開発側からすれば誇らしくまた商業的にも喜ばしい結果となり、そういった報道に接する一般市民といえば、先進テクノロジーの驚異に讃嘆しつつも、人類の地位がそれらに奪い取られかねないという不安も同時に抱いているようである。

では当の棋士らはどうか。これが文字どおり「戦士」であれば、「敗北」即「死」であるから悠長なことはもちろん言ってはいられまいが、当初は AI を見くびるような者もいたし、そしていざこうした事態が到来してしまうと、AI の優越性を率直に評価し、かえってその成果を自らに採り入れようとする姿勢すらも見られるようになる（もはやトップ棋士も、こうした AI との棋譜研究のトレーニングを通して対局に臨んでいるのが現状だという）。

かつてはこうであった、という（皮肉を言えば幾分微笑ましい）エピソードとして、1996 年版『将棋年鑑』の記事を孫引き（羽生 2017 より）する。プロ棋士へ「コンピュータがプロ棋士を負かす日は？」というアンケートをとった。この年は、IBM 製スーパーコンピュータ〈ディープ・ブルー Deep Blue〉が、史上最強とも謳われたチェスの世界チャンピオン、ガリー・カスパロフとの連戦で初めて勝ち越すという大事件の起こる前年である。

そういう時代でもあったためか、アンケートの結果、多くの棋士が、そんな日は来ない、と真っ向から否定した。実名とともにそのコメントは、米長邦雄「永遠になし」。加藤一二三「来ないでしょう」。村山聖「来ない」。真田圭一「百年は負けない」。郷田真隆「いつかは来ると思う。但し、人間を越えることはできないと思う」。

しかしそのなかで「その日」が来るのをほぼ正確に予測していた棋士がいたという。それが羽生善治である。曰く、「二〇一五年」（羽生 2017: 3-4）。

前掲の文献（羽生 2017）はその 2015 年に、NHK スペシャル「天使か悪魔か　羽生善治人工知能を探る」（翌 2016 年 5 月放送）の制作にともない著されたものである。企画がスター

トし、最初の打ち合わせで羽生は、制作統括（エグゼクティブ・プロデューサー）に、AI に勝てるか？と訊かれた。その時は、羽生はこう答えている。「今、将棋の人工知能は、陸上競技で言えば、ウサイン・ボルトくらいです。運がよければ勝てるかもしれない。しかしあと数年もすれば、F1 カーのレベルに達するでしょう。そのとき、人間はもう人工知能と互角に勝負しようとは考えなくなるはずです」（羽生 2017: 4）。同書で羽生は、AI に次々とうち負かされる棋士のなかの「最後の牙城」（羽生 2017: 4）と目されている。その彼が、もはやこうも潔くシャッポを脱いでいる。

羽生についてはほぼ説明不要であろう。1970 年生まれ、現役棋士のなかでずば抜けた戦績（タイトル獲得合計 99 期、1996 年のタイトル 7 冠独占、通算 1434 勝等々はいずれも前人未到）を誇り、まさに往年の絶対的存在であった大山康晴と並ぶ、現代最強の棋士である（ただし、一般に棋士の実力のピークは比較的若い時期に訪れ、彼もまた既に下り坂にあることは事実である。2018 年には二十七年ぶりに無冠となり、今年 2022 年には 29 期在籍していた A 級からの陥落が決まっている）。

その彼をして、ここまで厳しく、また冷静な評価を下さしむるのが当時の AI の威力であった（そして 2022 年現在、羽生が 2015 年に言った「数年」は既に経過している）。

本論に入る前、最初にここで述べておきたいのが、AI が既に人智を超えており、いずれ近い将来には人類を脅かす存在となるという、巷間盛んに取り沙汰されている悲観論についてである（例えばカーツワイル 2007 によれば、2045 年には「シンギュラリティ」と呼ばれる歴史的に重大な転回点が訪れ、人類と AI の地位は逆転し、職業などがみな奪われる、とされる）。

これをめぐって倫理学から検討すべき論点は幾つもあるが、前もって、AI とロボットとの区別を明確にする必要がある。そしてこれこそまず、心身二元論など、哲学史において重要とされてきた諸見解と深く関わっていることがわかる。

或るロボット研究者によれば、「ロボットは動く機械装置」であり、一方、人工知能［AI］は「人間の脳の働きをまねた、目に見えない情報処理の一種」である（Cf. 新山 2019: 16-20, esp. 17）。

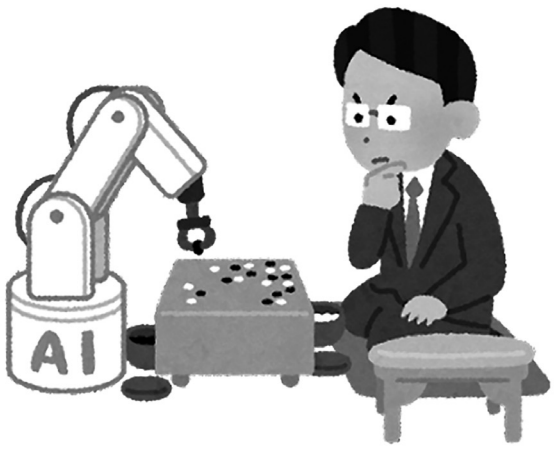

図 A（いらすとや より）

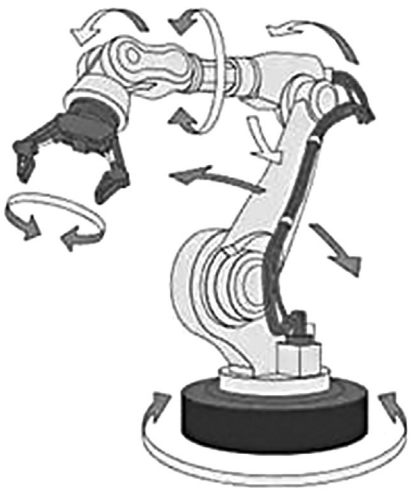

図 B（FA ロボット.com より）

ここに挙げた図Aは「かわいいフリー素材集 いらすとや」という我が国できわめて広く用いられている画像シリーズからの一枚であるが、画中「AI」と謳っているのは両義的である。

今まさに黒石を碁盤に打とうとしている機械が、「垂直多関節」の、文字どおり「ロボット」で、これは「4軸」（一般に産業用ロボットでは図B[*2]のような6軸機構が主流であるが、図は簡略化されている）で、碁石を「爪」で摘まんでいるのが黒い「ロボットハンド」（エンドエフェクト）、それが二つの「軸」とともに三つの白い「ロボットアーム」（マニピュレータ）によって、「AI」とペイントされた台に接続されている。すなわち「垂直多関節ロボット（シリアルリンクロボット）・4軸」となる。

だが、「AI」と呼ばれる部分はどれなのかといえば、このロボットの外見からはわからない。台の部分に内蔵されているのか、あるいは（画中には描かれていないが）有線ないし無線でコンピュータ本体から遠隔操作されているのか、いずれにせよ、この「ロボット」はAIの指令によって碁石を碁盤に打つことに従事しており、AIそのものと見なすべきかといえば、それには「待った」がかかるであろう。

前掲した新山の定義を振り返れば、AIとロボットとの違いは、古くはデカルトに代表される近世の心身二元論と結びつくことがわかる。AIは人類の「脳」（デカルトでいえば「精神」）の「働き」を模したものである。一方のロボットで核となるのは働きではなく「動き」であり、ロボットの「ハンド（手）」はまさしくAIが意図する働きを現実化する（やはりデカルトいうところの）「延長」なのである。この点からも、AI／ロボットの基本的な構想が、典型的な近代科学主義に拠るものであることがわかる。

こうした解釈が、哲学史上の重要なかつ周知の概念への強引なこじつけでないことは、次の系譜を示すことで弁明できる。すなわち、デカルト流の心身二元論[*3]が、後代の随伴現象説（epiphenomenalism、心的現象は物的現象に随伴して生起する、との見方）によって批判的に深化し、大脳生理学の発展も相俟って、20世紀に入ると、「心の随伴現象説こそが心身問題の唯一の解決法と見なされるに至った」（坂本 1998: 823）。さらに、1948年にあの「サイバネティクス」がN・ウィーナーによって提唱される。これは、通信とその制御という事象にまで視野が拡大し「人間の精神的活動の物質機械性を支持し、随伴現象説を補強する」（坂本 1998: 823）。

そしてサイバネティクスこそ、今まさに第3世代のAI理解における、とりわけ重要な概念なのである。西垣はウィーナーを受けて、生物をネオ・サイバネティカルなものと定義し、すなわちオート・ポイエティック（自己－創出的）な存在とみなし、それが今後のAIに結びつくものであると捉えている。その一方で、ロボットはアロ・ポイエティック（他者－創出的）な存在であり、それゆえ、AIとロボットは明確に線引きされることになるのである（Cf. 西垣 2017: 64）。

人が「ホモ・ファーベル」と定義づけられたのも（homo faber：工作人。その命名はH・ベルクソンによるとも、M・シェーラーによるともいわれ、その発案自体はさらにB・フランクリンにまで遡るとされる）、デカルト流心身二元論に則って表現すれば、先史時代の人類が精神の指示によって手を用いて物を加工してからである。

そして、それだけならばまさしく人類の歴史を回顧しただけであるが、AIの自律獲得および真の「自己－創出」という事態が起こり得るとするならば、それをAI／ロボットに譬えると、

・AIがロボットの手を操作して（何か物を）加工する

これだけならば、人類史で絶えず展開されてきた技術革新にとどまるが、それだけでなく、

・AI がロボットの手を操作して、AI 自らを加工する

という段階が現実化する事態ということになる。いうまでもなく、「精神が手という延長を用いて精神自らを加工する」というのは、学習などの後天的な修得活動レベルならともかく、もともと存在していなかった自律性を新規に獲得するというのであれば、これは人類史はおろか、地球全体、さらには宇宙における生命の歴史においても、文字どおり前代未聞の事態というべきではなかろうか。

したがって、AI をめぐる ELSI （ethical, legal and social issues: 倫理的・法的・社会的問題）のうち、まず AI の責任問題だけ述べると、AI が人類と同じく自律能力を備えた場合、それが問われることとなると思われる。しかし、AI が自律能力を備える可能性は多くの専門家が否定しており [*4]、すると AI が仕出かした粗相はあくまでその設計者ないし運営者等にその責任を帰すべきであり（状況により定められるべきである）、してみると（R・カーツワイルや Y・N・ハラリら、少なからぬ影響力を持った言語人が、意図するとせざるとにかかわらず民衆をあたかも扇動するような物言いで）AI の〈暴走〉により人類の文明が脅かされるというのは根拠を欠いた文字どおり杞憂としか思われない。

第一章　人類を圧倒する AI

第一節　情報工学の〈量〉の面における飛躍的発展

次に、本論文が主題としてもっと敷衍して述べたいのが、AI の能力と人類の能力との対比についてである。

よく用いられる説明に、AI に犬と猫の識別をさせる、というのがある（例えば、稲葉 2021: 211-212）。先に述べたように、「第 3 次 AI ブーム」が従来のそれとは大きく異なる点が、かつては容易でなかった犬猫の区別が格段に進歩した、という例でよく語られるのである。

従来の技術は、犬と猫のそれぞれについての特徴を表す指標を数え上げ、それらを機械に入力する、というもので、いったいどんな指標を入力してやればよいのか、というのが難しいとされていた。

なぜ難しいかといえば、何のことはない、我々人類は無意識・無自覚に犬猫を区別しているが、意識もせず自覚もともなわぬ以上、それを明示的に言語化し、プログラム化して機械に入力することはできない、とされてきたからである。

ここでさっそく哲学の出番が来たことがわかる。これは一説によると、人類では至極お馴染みの「イデア」に拠るものにほかならない。

論者も講義の際、「猫」のイデアを例に挙げ、話題として用いている。誰もが猫を知っている、もちろん犬とも区別できる、しかしそれは加算的な定義づけ（定義の列挙、網羅）で理解し説明できるわけではない。まさにキリが無いからである。

イデア論によると、我々に猫のイデアが内在化し、我々は猫を知る、という。あるいはプラトン本人はもっと極端に、我々の魂のうちに眠る猫のイデアが想起されると説いているが、それは認識における先天性すらかつてのようにほとんど認められなくなった現代科学では、およそ受け容れられぬものであろう。

論者自身は、先のイデアの内在化というのが十分に腑に落ちる（論者は応用倫理学専攻であるとともに、近現代ドイツ哲学・倫理学を主に、また古代ギリシャ哲学・倫理学も併せて学ん

だ者であるから、イデア論≒観念論にはひじょうに馴染みがある）が、しかしもちろん、認知論からはもっと詳細な説明が施されている。

そしてここで問題となるのは、人類のイデア形成における思考力の容量は、現代の汎用コンピュータに比して、大きいのか否か、ということである

これも、説明としては二種存在するように思われる（そして結論からいって、そのいずれが妥当か、未だ決着がついていないようでもある）。

すなわち、〈ヒトの脳の働きは計り知れない、それは「電話帳」で「何十冊」、「何百冊分」のデータに匹敵する〉（断わっておくがここでの「電話帳」は昔ながらの紙媒体の物である。というのも、今や「電話帳」でネット検索しても、ほとんどスマートフォン内蔵のそれしかヒットしなくなってしまっている。一度試してみられたい）、などという喩えがよく用いられる。これはいわば人智礼讃型といえるだろう。

一方で、パソコンのメモリの容量が数十テラ（すなわち百分の数ペタ＝0.0 数ペタもあるスケール）バイト規模で実用化されている今日、〈ヒトの脳の容量など、たかが知れている〉という、まったく対照的な説もよく目にする（人智軽視型ともいうべきか）。

そもそも 1TB が何文字分に相当するかといえば、もちろん大まかな目安であるが、全角で約 5500 億文字／半角で約 1 兆 1000 億文字、といわれている。学術書に多い A5 版は（論者が以前出版した際に聞かされたところでは）1 頁が 400 字詰め原稿用紙 2 枚相当（例：原稿用紙 500 枚でおよそ 250 頁）などとされており、日本文が漢字仮名とアラビア数字アルファベットなど混在していることを度外視しても、1TB で約 7 億頁以上となる。

我が国特有のジャンルともいえる歴史小説、その最長篇である山岡荘八『徳川家康』全 26 巻は原稿用紙 1 万 7400 枚、文字数に換算して 696 万文字あるとされる。

世界文学の名作クラスでは、プルースト『失われた時を求めて』が第 1 部「スワン家の方」から最後の第 7 部「見出された時」までアルファベット約 960 万文字（スペース等含まず）とされ、こちらは半角文字だが、それでも換算するとわずか 9.16MB 程度にしかならない。

我々研究者は自分の書いた論文の MSWord データのサイズをパソコン画面のフォルダ情報で把握しているから（むしろいちいち見もしないことも多かろう）、サイズがこんなに小さいということに特段の驚きは無い。文字データ＜画像データ（JPG）＜動画データ（MP4）なのは当然で、1 文書あたりのサイズはせいぜい数十 KB から千幾つ KB（1MB 少し）である。

さらに余談めくが、たしかソニーかどこかの日本の AV 機器メーカーの開発者がかつて、「ビデオ・テープの容量をけっして侮ってはならない。あれだけの表面積（注：言葉の一字一句は、論者がこの資料を読んだ際の記憶がやや曖昧なのだが）があるのだから、その潜在能力は計り知れない」という趣旨の発言をしていた。驚いたことに、彼は磁気メディアがビットメディアに太刀打ちし得ると真剣に考えているようである。記録方式という「質」と記録媒体の「量」とをめぐる逆転関係が正しく受け容れられておらぬのである。

業務用ビデオでは未だ現役の 1 インチテープ（かつての家庭用ビデオ・テープよりも格段に大型）であれ、往年の録音用オープンリール 1/4 インチテープであれ（テープスピードが速いほど高音質で、最高速は 76cm/s）、質の面では PCM（pulse code modulation）によるデジタル・レコーディングの登場によってとうに追い抜かれている。もっとも、デジタル録音に関しては、PCM はデータ圧縮は無いが、いわゆる「可聴音域」といわれるものに関しては、22000Hz（22KHz）でカットされることが一般的である（一方、コンサートなどの生演奏では約

40000Hz に達するとされる）。そこで音質については、オーディオマニアの間の議論では〈それでもなお、なぜアナログなのか〉というなかば精神論、あるいは都市伝説で長年にわたり展開している。

ビッグデータと称する、その全体像を簡単には把握できない、様々な形式の巨大なデータ群が、もはや我々の日常生活の基盤となっている（ネット上を飛び交う情報、GPS 情報、オフィスにおけるサーバーのアクセスログや文書データ、医療情報、気象データ、交通系 IC カード情報、購買データ、監視カメラ情報等々）現在、もはや情報のスケールは桁外れに巨大化・拡大の一途を辿っている。

ただし先述した人類のイデア的な情報獲得・運用については未だ神秘的な部分も多く残り、究明は永続的に続けられることであろうし、ヒトの脳の解明が十分に進んでいない現状では、人智をめぐる〈神秘主義〉もまた根強く残り続けるであろう。

ともあれ、AI がデータ容量やその処理能力で人智を圧倒するのは火を見るよりも明らかであり、要は算盤が電卓に取って代わられたような前例と、今後の人類と AI との関係とは何が同じで何が異なるのか、ということが問われるわけである。それは、いま述べた算盤がやはり未だに児童にとっては計算力や思考法の養成において完全にその支持を失っていない、という点とも関係するかもしれない。

以上をまとめると、量と質については、ひとまずは「質は量に還元できる」と言うことができるだろう。なぜなら、AI のように、プログラムどおり、迅速かつ正確に計算するコンピュータの能力に、いかにして質の評価を下すべきかは難しい問題だからである。いかに速くそして正確に計算するかは、その目安を畢竟、「量」に還元することとなる。既に AI が人類を凌駕している分野はいずれもこの点に関してである。

AI において「質」を問うとすれば、本論文が後半で詳論するように、人類とどこが同じでどこが異なっているかが問われるであろう。その着眼点の例として本論文では二つ、「美意識」と「おそれ」とについて考察するものである。

さて、その前提として、明らかに AI が人類を既に凌駕しているとみなされる分野についてみていく。

第二節　「最強棋士」の AI 論に基づいて

第 3 次 AI ブームが飾った輝かしい戦歴の一環として、本論文は碁、そして特に将棋を例に挙げる。その恰好ともいえる参考文献が、羽生善治が NHK の番組制作に協力した際に執筆した『人工知能の核心』（羽生 2017）である。

論者は当初この書から、AI と棋士との対局に関して興味深い議論が聞けるだろう、程度の期待をしていたが、それを遙かに上回る、予想以上の収穫があった。それは後述するように、論者が専門とする応用倫理だけでなく、さらに遡り、カント批判哲学とも関連づけられる議論が展開されていたからである。

もし論者の読解したとおり、AI とカント批判哲学との関連づけの可能性がここに見出せるとしたら、それはきわめて興味深いことだといえるだろう。

それを裏づけるものとして、羽生は同書中で幾度も、AI における「新しい「美意識」」に関して言及しており、それはカント批判哲学では当然ながら第三の『判断力批判』に大いに関わってくるはずである。

もちろんそう結論づけるには周匝な議論が必要だろうから、順序として、まず羽生の所説を幾つか引いてみる（以下、引用・参照はいずれも羽生 2017 から、アラビア数字は頁数）。

引証の最初に、羽生が AI と棋士をただ対局させて勝敗を競う、というのではなくて、「人工知能のアルゴリズムのあり方から何かを吸収して、例えば新しい「美意識」を提示しようと考える発想の方が、むしろ建設的であり、意義深いように思えます」（44）と述べているの注目したい。これは同書第一章のうち、「人工知能から新たな思考を紡ぐ」と題された節（43-45）からの引用で、AI から「単に答えを与えられるだけではもったいない」（43-44）に続いて、羽生はこう述べているのである。

そして続く第二章はまさしく「人間にあって、人工知能にないもの――「美意識」」と題されている。この副題だけからも以降の論の展開は予想できるだろう。ただし、羽生の場合はもっぱら将棋に根ざして論述していくところに、やはり門外漢の我々（とりわけ論者のように将棋や碁といった方面の嗜みが無い者）にとって学ぶべきところが多い。

すなわち、羽生は将棋の「大局観」について述べる（67-71）。それは羽生によると、「直観」、「読み」とともに棋士がそれらを使って対局中に思考する、その三つのうちの一つである（71）。

よく、AI に対して人類が秀でている、として引き合いに出されるのが、棋士は夥しい可能的な指し手の中から、無駄な手を「引き算」で削っていくのを大事にしている（71）、という点である。

そもそも AI にこの「引き算」いったような芸当がまったくできないかといえばそうではないが、やはりこの「直観」力こそが人類の強みであるのは容易に察しがつこう。

それは将棋／チェス／碁に代表される和洋中の三つ、ゲーム理論の分類で「二人零和有限確定完全情報ゲーム」と呼ばれるゲーム／競技にともなう、まさしく圧倒的な情報量に関わってくる。

まず盤面を見ればすぐわかるように、升目の数でチェスは 8 × 8 = 64 升、将棋は 9 × 9 = 81 升、一方の碁は縦横の直線で格子が作られ、その交点は 19 × 19 = 361 ある。

当然ながら最も数の多い碁が最も複雑だと考えられがちだが、実際にはそれぞれ戦い方も異なるし（よく言われるのは、将棋では獲った相手の駒を自分の物として使えるという、他には無い戦法が存在する）、一概には言えない。また、そもそも碁は升目に石を置くのではなく、縦横 19 本引かれた線の交差部分に打つゲームである（すなわち、19 × 19 で構成された縦横 18 × 18 の升目に石を置くわけではない。五目並べやオセロとは異なる）。

しかしそれでも、碁における合法的な局面の総数は現代数学にとっては恰好の論題である。現在、その正確な値とされるのは、約 2.1 × 10^170（10 の 170 乗。すなわち 171 桁の数）であることが、トロンプらによる数年にわたる計算の結果、2016 年に最終発表されている [*5]。

碁の可能な手数は圧巻で、まさに天文学的数字であるといえるし、将棋もまた、一つの局面で（羽生によると）、平均八十通りの指し手があるとされる（66）。

こうした天文学的な可能的手数と対峙して、こと将棋においては、他のゲームよりも、「引き算」をしていく発想が大事かもしれない、と羽生は言う。なぜなら、将棋は、実は自分の手番で、本来なら「何もしない」のが最適解である場面がとても多いゲームだから（71-72）である。

まず、将棋を学ぶときに（すなわち局面という核だけでなく、将棋道全般に言えることとして）大事なのは、実は覚えることを増やすだけでなく、余計な考えを捨てていくこと（24）で

ある。つまり、誤った情報に惑わされない・導かれない、という趣旨であろう。

そして今度は個々の局面に際しても同じことが言える。羽生に言わせれば、「将棋が強くなるために一番大事なことは」、「だめな手が瞬時にわかること」なのである（73）。そしてその「だめな手が瞬時にわかること」は、「直観」と「大局観」とに関わり、「実戦での経験の蓄積から身につくもの」である（73）。

直観といい大局観といい、その言わんとする意味はわかりやすい。一方で、「美意識」は日常語であるとともに、美学・感性論の術語でもある。それを羽生はいかなる意味で用いているか。

第三節　将棋における「美意識」と「おそれ」

羽生はこう述べている。

> 「筋の良い手に美しさを感じられるかどうかは、将棋の才能を見抜く重要なポイントなのです。この自らの「美意識」をいかにきめ細かく磨き込んでいくかが、将棋の強さに関わってきます。／人間がどうして、いきなり九〇パーセントくらいの手を「直観」で捨てて、何万手という「読み」の方向性を、「大局観」で制御していけるのか。／この大きな取捨選択の核となるものが「美意識」なのです」（75-76。／は改行）。

してみると、この美意識という言葉には、たんに美感的な嗜好だけでなく、より広義に実践的な、将棋という「戦さ」においては文字どおり「生きるか死ぬかの見極めどころ」という意味まで帯びていることがわかる。

ただし前もって断っておくと、「芸道」そして「武芸」さらに「武道」へと類推ないし敷衍していくに当たって、例えば卑怯であることを「醜い」と批難したり、あるいはたとえ勝負事においてさえ「美しく勝て」という理念が掲げられたり、あるいはむしろ「美しき敗者」であれと唱えられたりであるとか、広義の善悪の概念を美醜の概念で表現することもきわめて一般的におこなわれている。

だからそれは、どう攻めあるいはどう受けるかといった戦術そのものはもちろんのこと、武士道の究極においては自害や切腹といった死に様にすら関わってくる（人口に膾炙したベネディクト風の定式化を用いれば、「恥の文化」、すなわち他者の目に醜く映ることを何よりも忌み嫌う気風ともいえよう）。だから、こうした「恐怖心」「おそれ」は、あるいは「憚る」心、と表現した方が良いかもしれない。

羽生のまとめによると、「人間が「直観」「読み」「大局観」の三つのプロセスで手を絞り込んでいくとすれば、人工知能は超大な計算力で「読み」を行って最後に評価関数で最善の一手を選ぶ」（80）わけであるが、「ここで人間にあって人工知能にはないのが、手を「大体、こんな感じ」で絞るプロセス」（80）なのだという。そして、特に興味深いのが、「棋士の場合には、それを「美意識」で行っていますが、人工知能にはどうもこの「美意識」に当たるものが存在しないよう」（80）だ、ということである。

論者は先に、議論が美意識が関わるからカント『判断力批判』を引き合いに出した。しかしそれだけではない。羽生はさらに「私はその理由は、人工知能に「恐怖心がない」ことと関係していると考えてい」（81）る、というのである。

日本人の羽生がこの「恐怖心」を、カントの母語ドイツ語の何に相当するものとして語っているかを彼に追及するのはもちろん筋違いであるが、カントがその批判哲学で “Furcht”（従来は「おそれ」「畏怖」等と訳されてきた）について重要な議論を展開していることは既に広く知られている（ただし弘文堂『カント事典』にはドイツ語 “Furcht”、邦語「おそれ」、「恐怖」の項は無い）。ここではただ、「美意識」（美）と「恐怖心」（おそれ）とが関わっているという、羽生の興味深い指摘に注目したい。

そこに注目しつつ、さらに羽生の論説を読み進めよう。彼の評するところによると、ただただ圧倒的な量の過去のデータに基づき、最適解を計算する人工知能は、それがゆえに人類の思考の盲点となるような手を「怖いもの知らず」（81）で平然と衝いてくる。

一方、人類のこういった思考の死角や盲点のようなものは、防衛本能や生存本能に由来しているように思えてならない。人類には、生き延びていくために、危険な選択や考え方を自然と思考から排除してしまう習性があるような気がする、と（81）。

さらに門外漢である我々がそう思うのみならず、羽生自身も「面白い」と言っているのは、こうした習性の結果と、例えば棋士が先述したように「美意識」で手を絞り込む時、「美しい」と感じられるのが、基本のかたちに近い、見慣れたものである（81）ということだ。

棋士は紛れもない「戦士」なのであって、我が国においては中世以前や幕末の戦乱の世はもちろんのこと、人類の祖先が原始時代に苛酷な生存競争を勝ち抜き（あるいはより消極的には、ひたすら生き延び）、子孫を遺し得たように、棋士もまた対局という「戦」に勝ち抜くため（こちらには勝ち抜く以外に生き延びる術は無い。将棋等は）、危険な選択や考え方を文字どおり「おそれ」、斥け続けた結果、いわゆる「定跡」（碁では「定石」）へと洗練されていった。その洗練を通じて「美」の意識が構築されていったことになる。

だからこそ、武道においても、まずは勝って生き延びるために、次いではたとえ負けても見苦しい死に様でないように、美しい戦い「方」、「型」が尊ばれた。この「美」は美学的なカテゴリーであるとともに明らかに倫理学的なカテゴリーでもある。

第二章　人類と AI との違い ―― カント批判哲学と関連して ――

第一節　AI のさらなる可能性を問う

羽生は「人工知能が恐怖心を覚えるようになったときが、本当の恐怖かもしれ」ないと述べている。その理由は、「人間にとっても得体のしれないものになるから」（83）である。

これは将棋「道」（もっとも、羽生は「道」について直接的に言及していないが）における「美意識」と、それとは対極的な「おそれ（畏れ・恐れ・怖れ・惧れ・虞（れ）・懼れ）」（羽生のいうところの「恐怖心」）に関わる。

AI が「おそれ」の念を抱くということ――。ここからいよいよカントについての言及に踏み込んでいく。

周知のように、『判断力批判』は、序論に続き「第一部　情感［美感］的判断力の批判」、「第二部　目的論的判断力の批判」の大きく二部構成を採る。前者「情感的判断力の批判」はさらに「第一篇　情感的判断力の分析論」と「第二篇　情感的判断力の弁証論」とに分かれ、さらにその前者「情感的判断力の分析論」は「第一章　美しいものの分析論」と「第二章　崇高なものの分析論」とに分かれる。さらに「崇高なもの」に関しては、やはりよく知られたように、

「A　数学的に崇高なものについて」と「B　自然の力学的に崇高なものについて」との二種に分けられて分析される。

以上、『判断力批判』全体の構成を表で示せば、次のとおりとなる。

<table>
<tr><td colspan="4">序論（これとは別に、当初書かれ、後に全面的に差し替えられた「第一序論」がある）</td></tr>
<tr><td rowspan="4">第一部　情感的判断力の批判</td><td rowspan="3">第一篇　情感的判断力の分析論</td><td colspan="2">第一章　美しいものの分析論</td></tr>
<tr><td rowspan="2">第二章　崇高なものの分析論</td><td>A　数学的に崇高なものについて</td></tr>
<tr><td>B　自然の力学的に崇高なものについて</td></tr>
<tr><td colspan="2">第二篇　情感的判断力の弁証論</td></tr>
<tr><td rowspan="2">第二部　目的論的判断力の批判</td><td colspan="3">第一篇　目的論的判断力の分析論</td></tr>
<tr><td colspan="3">第二篇　目的論的判断力の弁証法</td></tr>
</table>

※第一部第二篇末尾に「趣味の方法論について」、第二部全体の末尾に「目的論的判断力の方法論」が、それぞれ付録として置かれている。また、本論文ではあまり触れない第二部（目的論）の論述も第一部と同様にカント特有の体系的嗜好を反映して、形式性の高い書式に則っている、との指摘もある。詳細は、例えば牧野による訳者解説（カント 1999: 313-344）を参照。

このうち、特に本論文における AI 論に関係するのは第一部の「崇高」に関する論述であるのだが、先に二箇所、それとは異なる第二部から引用する。

> 「恐怖 Furcht ははじめに神々（鬼神）を生み出すことはできたが、しかし理性は、自分の道徳的諸原理を介して初めて神についての概念を生み出すことができた」(『判断力批判』第二部第八六節)。

これは盲目的な恐怖の対象としての鬼神から、道徳的原理の源、いわば理性の洗礼を受けた神への人類の思索史における移行について述べた箇所である。

> 「その代わりに（引用者注：強制と強引な服従とではなく)、人倫的法則に対する尊敬 Hochachtung がまったく自由に、我々自身の理性の指令に従って我々の使命の究極目的を我々に表象させるとすれば、我々は、感受的な pathologisch 恐怖 Furcht とはまったく異なる最も誠実な畏敬 Ehrfurcht によって、この究極目的とその遂行とに合致する原因を我々の道徳的展望のうちへと一緒に取り入れ、この原因に喜んで服従するのである」(『判断力批判』第二部「目的論に対する一般的注解」)。

この「目的論に対する一般的注解」からの引用文の末尾には、カント自身による原注があり、それも全文を引用すると、「美に対する讃嘆と、これほど多様な自然の諸目的による感動とは、熟慮する心の持ち主が世界の理性的創始者について明晰な表象を持つ以前に、既に感じることができる。そしてこの讃嘆と感動は、或る宗教的な感情に類似したものをそれ自体で持っている。したがって讃嘆と感動は、たんなる理論的観察が惹き起こし得る関心よりも、はるかに多くの関心と結びついている讃嘆を起こさせる場合には、それらは、まず道徳的な判定の仕方と

類似的な判定の仕方によって道徳的感情（我々に未知の、原因に対する感謝と畏敬の）に働きかけ、それゆえ道徳的諸理念を喚起することによって心に働きかけるように思われる」とある。

ここでもまた、先の引用（第二部第八六節）と同様、理性を介しているか否かが大いなる違いとなる。「感受的な恐怖」とはいわば、表層的・感覚的なものであって、もちろん「最も誠実な畏敬」とは「まったく異なる」であろう。それは、道徳の根源として「未知の、原因に対する感謝と畏敬」であり、『判断力批判』の叙述の順序とは逆に [*6]、我々は同書第二三－二九節「崇高なものの分析論」（崇高論）において、畏敬／畏怖について既に論ぜられていたのを顧みなければならない。そして次節で取り沙汰されるように、とりわけ「力学的に崇高なもの」についてが問題となる。

ではいよいよ、本題である「力学的に崇高なもの」についてのカントの所見を、もっぱら「おそれ」に着目して読解してみる。

第二節　カントの「力学的に崇高なもの」論について

同書でのカントの崇高論は、数学的なものと力学的なものとの二つが題材となっている（これは『純粋理性批判』における第一・二カテゴリーと第三・四カテゴリーとの対比と呼応している）。そもそも崇高とはカントによれば「端的に大きなもの」であり、数学的なものは量に関わり、力学的なものは（後述するように）優越性に関わる。そして本来の意味で崇高の名に値するのは、崇高な自然を観照する人間精神の無限な能力、すなわち超感性的な道徳性に帰せられる [*7]。

前掲の表に示した「B　自然の力学的に崇高なものについて」は、「二八　力としての自然について」および「二九　自然の崇高なものに関する判断の様態について」の二節から成る。

まずカントは第二八節で「力 Macht」と「威力 Gewalt」とを区別して論じ始める。カントによると、前者は「大きな障碍に優越している能力」、そしてその同じ力 Macht が、「力をそれ自身所持しているものの抵抗にもまた優越している時」、後者「威力」と称せられる。「自然は我々を支配する威力の無い力として情感的判断のうちに考察された時に、力学的に崇高である」（第二八節）とカントは言う。

だから、自然が我々をして崇高の念を感ぜしむるがゆえに、自然は恐怖 Furcht を起こすものとして表象されるはずである、というのがここでの「崇高」－「恐怖」の関係性である。とはいえ（既に同書の後半、第二部第八六節や第二部「目的論に対する一般的注解」からの引用に見たように）、恐怖を起こす対象すべてが我々の情感的判断において崇高であるわけではない。むしろ人類にとって神は、恐怖せずに、しかもおそるべきものとみなし得る存在なのである。それゆえ「有徳な人は、神に恐怖することなく、神をおそれる So fürchtet der Tugendhafte Gott, ohne sich vor ihm zu fürchten」（第二八節）。この引用箇所でカントが用いている語は、この邦訳における「恐怖する」も「おそれる」も、いずれも Furcht の動詞形 fürchten（下線部、前者はその三人称単数現在形）なのである。

このように、カントは明確に、たんなる「恐怖」と、理性を介した「おそれ」とを別なものと考えている。前者はただ漠然とした不安をともなう感情であり、それに対して後者は己れの限界を自覚した、弁えられたものだからである（そもそも「批判哲学」の「批判 Kritik」こそ、こうした可能性と限界づけの厳密な学の謂いであった）。

そのことについてカントは次のように述べている。

「我々は、自然が計り知れぬことについて、また自然の領域の情感的量評価に釣り合った尺度を採るには我々の能力が不十分であることについて、我々自身の制限を見出したのだが、にもかかわらず我々は、同時に我々の理性能力について非感性的な別の尺度――あの無限性そのものを単位として含み、この尺度と比べれば自然におけるすべてのものは小さいような尺度――を見出したのである」（第二八節）。

そしてカントは最後に、なぜこうした崇高論が彼の批判哲学・超越論（的）哲学、第三批判『判断力批判』で要請されたかを、次のように述べて説明を終えている。

「崇高なものについての他の人々の判断が我々の判断に賛同する必然性（略）情感的判断があえて主張する判断の必然性のうちに、判断力批判に対する一つの主要契機が存在する。なぜなら、この必然性は、まさに情感的諸判断についてア・プリオリな原理があることを知らせ、さもなければ情感的諸判断が楽しみと苦痛の感情のもとに（いっそう繊細な感情という無意味な形容詞をともなうだけで）埋もれたままであるような経験的心理学から引き上げ、情感的諸判断と、またこれらの判断を介して判断力とを、ア・プリオリな諸原理を根底に持つものの部類に配置し、しかもそうしたものとしてこれらを超越論的哲学のうちへと引き入れるからである」（第二九節）。

まさしく、たんなる美醜の判定を超えた力、カントのいう反省的判断力、あるいは本論文で論題としてきた「美意識」が、己れを凌駕する崇高なものに対して、ただの恐怖、怯懦ではなく、人類が本来的に有するア・プリオリな、超越論的な機制の自覚を促す契機ともいえるような「おそれ」を喚起させられることにより、我々は我々の理性のいっそうの深淵を見出すべく努める。その営みこそ「超越論的哲学」にほかならぬのである。

このように、カント『判断力批判』を読めば、本来の「おそれ」を抱くことこそがまさしく、人類が真の意味で人類であることの必要条件であるように思われる。まさにカントが言うとおり、「最大の讃嘆の対象」とは、「物に動じず恐怖することなく、それゆえ危険を避けず、しかも同時に十分に熟慮して用意周到に当たる人間」である（第二八節）。

してみれば、我々人類が、AI に対して抱く「おそれ」とは、〈未だ十分に実態を攫み得ぬ、しかし我々を或る場面では遙かに凌駕するもの〉に対して抱く、きわめて自然な心情の発露であり、たとえ現時点で既に AI が多くの領域で我々人類を上回っているといえども、こうした心情そのものを AI が自ら備えるには至っておらず、またそれを備えること自体、独力ではあり得ないというわけである（例えば前近代の叙事詩の世界観ならば、人ならぬ魔的な存在が、我々人とはついに親和的になり得ぬように、である。ただしこの叙事詩的世界観は、むしろ現代社会においてはファンタジーや SF などでいっそう広い読者・観客を獲得しているのだ）。

第三節　人類と AI、それぞれの「美意識」と「おそれ」

ここで論を整理するために、主体と客体とについて明確にする。すなわち以下のとおりである。

・人類が主体である「美意識」と「おそれ」（この段階では、何が「客体」であるかは敢えて

問わない）
・AI が主体である「美意識」と「おそれ」（上と同じくこの段階では、何が「客体」であるかは敢えて問わない）
　するとより明瞭に浮上してくるのが、
・人類が主体であり、AI を客体とする「美意識」と「おそれ」
であり、さらには、
・AI が主体であり、我々人類を客体とする「美意識」と「おそれ」
という関係も浮かび上がってくる。だがしかし、最後のものは、はたして、想定し得るではあろうが、現実化するであろうか（まさしく自律／自己－創出と関わる問題である）。
　ともあれ、こうした「おそれ」のありようは、我々人類から AI に向けられた、やや屈折した感情とも関わってくる。
　人類は、コンピュータに対しては、AI に対してとは異なり、おそれのような感情は抱かぬだろう。電卓だったらなおさらであるし、あるいはパワーショベルでも同様である。
　すなわち、人類よりはるかに秀でたデータ処理能力や計算能力、あるいは運搬能力や掘削能力などあっても、コンピュータなどはしょせんは人類が利用する道具にすぎぬとわかりきっているからである。四則計算に特化した電卓や土砂や岩石を持ち上げるだけの建設用重機も同様である（ただしそれが、ギネス認定されるほどきわめて巨大なものであれば、少なくともその大きさに圧倒されはするだろうが。しかしむしろそれは単純に巨大で凄まじい重量の金属の塊を目の当たりにして、まるで圧し潰されそうに感じる恐怖心のようなものである）
　その証拠に、例えば国産のスーパーコンピュータが世界記録を更新したなどのニュースを聞いて、その高性能さに感心こそすれ、戦慄を覚える者はほとんどいないだろう（2022 年 5 月に米の〈フロンティア Frontier〉が首位となり、理化学研究所の富岳は第 2 位に転落した。フロンティアの計算速度は 1.102E［エクサ］FLOPS（エクサは 10^18、100 京）、すなわち 1 秒間におよそ 110 京 2000 兆回の演算性能である）。
　同様に、旧い時代の映画や漫画・アニメなどでフィクション化されたロボットは、往々にして人類から見下され奴隷視された様子で描かれることも多々あった。
　しかしロボット社会はまだ到来していないが、AI 社会は既に到来している。また、多くの AI ツールは（ソフトバンクが開発中のものなどを例外としては）人体に擬せられることはなく、コンピュータの筐体に収められたまま、その機能だけが我々にその威力を見せつけてくる。この半ば不可視であることが、AI にまつわる漠然とした不安の要因であるともいえよう。
　また、AI に関する懸念として、その卓越した判断の根拠が、あまりの卓越性ゆえに、人類に俄かには（あるいはかなり長い時間を経ても）理解困難もしくは不能であることが指摘される。これはよく「ブラックボックス」と称され、やはり人類にとっては不安の種といえる。
　このように、主体［主観］としての我々人類が、AI を客体として捉えた時に覚える（カントの意味する「おそれ」ではなく）恐怖が、先述した AI についての悲観論の原因であることはもはや言うまでもない。
　そして羽生が予感するように、本来どこまでも客体であり続けるはずであった AI が主体化して、例えば将棋の或る局面において、我々人類という客体の営みに「恐怖心」［おそれ］を抱く、という間主観的［間主体的］構造がある。
　だが、そうした AI が抱く（擬人的な表現だが）「恐怖心」であれ、「おそれ」であれ、そう

したものをともなった「主体性」はあくまで疑似的ないし暫定的なものであって、カントが述べるような、理性的存在者あるいは人格が抱くような真の畏怖とはあまりにも遠く隔たっている。

本来の AI にとって、最も本質的な特性として、先述した「ネオ・サイバネティカル」（西垣 2019: 64; 70-73）とも、あるいは「オートポイエティック」（西垣 2019: 64-67; 70-73）とも表現されるもの挙げられている。オートポイエーシスすなわち自己－創出という特性が、ウィーナーが 1948 年に提唱した著名なサイバネティクスを刷新したもの、と想定されている。ここまでくればまさに AI は、現在のアロポイエティック（他者－創出的）・他律的な（heteronomous）システムを超え、カントのいう自律（Autonomie）に達するかもしれない。だが、少なくともそれは現段階ではあくまで「夢想」の域に留まり、原理的にあり得るものでは断じてなかろう。

むすび

さて、ここまで述べてきて、「畏敬」についてカントのあの、おそらく最も有名な言葉を引いてないことを奇異に感じられる読者もいることだろう。

> 「繰り返し、じっと省みれば省みるほど、常に新たに、そして高まりくる感嘆と畏敬 Ehrfurcht の念をもって心を満たすものが二つある。我が天なる星の輝く空と我が内なる道徳法則とである」（『実践理性批判』結語）。

この Ehrfurcht とは、文字どおり「畏 Furcht」「敬 Ehr」と訳されるべき語である。まさしく、天空というマクロコスモスと、理性というミクロコスモスという二つの極限に向けられた、カントの率直なる心情をあらわしている。

カントは、法はただそれ自身のために（法への畏敬から）のみ意欲され、義務もやはり義務のためにのみ意欲される、と考えた。「意志にとっては、意志自身から創出された目的、すなわち自己の自由という目的以外のいかなる目的も存在しない」とも述べている。

これが真の自律であり、ここからカント実践哲学の真髄である定言命法も演繹される。「汝の意志の格率が常に同時に普遍的な立法の原理として妥当し得るように行為せよ」（『実践理性批判』第七節）、この命題は単純に内容の点（格率↔立法原理）では対極的なもの同士の空虚な同語反復［トートロジー］であり、したがって推論（シュロギスモス／三段論法）では説明がつかない。しかし内容・実質ではなく形式において、道徳律はそれ自体としてなすべきものなのだ、と説明される。したがってそれは主体たる我々の意志の自由（それも真の自由）に拠るのである。

少なくとも現時点ではまだ、AI に自律は認められないし、自律能力を備えることはおよそ叶わぬことである。それと人類との或る違いは、おそれ・畏敬の念を抱き得るか否かに係るといえるだろう。

したがって羽生が述懐するように、AI がおそれの念を抱くことがあるとしたら、それは我々にとって真に慄然とする局面であるといえる。喩えるならば、無関心としか思われぬ自然がひょっとしたら我々同様に関心を有しており、自然に心奪われる我々を、実は物言わぬままにあちらから見据えているようなものである。

AI 論によって我々は、真と偽／善と悪／美と醜という周知の対カテゴリーに加え、力学的な「強と弱」という対にも注意するように促された。AI 自身が恐怖心ないしおそれの念を抱くことにより、我々人類にとっては、当初の AI の「無関心」であるというあり方が失われ、もしくは変容することとなり、それが我々にこれまで抱いたことの無かった恐怖心を喚起する、といういくぶん転回した事態が迫りつつあるのだろうか。

一方では、人間中心主義的な哲学や美学や倫理学が、やはり本質的な転回を求められているとも言うことができるだろう。

【注】

＊1　日本語の「倫理」と「倫理学」とは若干ニュアンスが異なるが、例えば英語ではもちろんともに ethics であり、表記上の区別は無い。以下、本論文でも特に両者を区別することはしない。

＊2　「ロボットハンド、ロボットアームの選定方法・選定基準」（キーエンス「FA ロボット.com」）より。
https://www.keyence.co.jp/ss/products/vision/fa-robot/industrial_robot/robotic-arm.jsp（アクセス最終確認は 2023 年 2 月 2 日）

＊3　この、いわゆるデカルト流心身二元論は、彼の主著の一つ『省察 *Meditations*』第六省察の「物質的事物の存在について」において大々的に展開され（Cf. Descartes 1996: VII-78）、また、『哲学原理』第二部から第四部の宇宙論的自然学においても、さらには晩年の『情念論』や、デカルトがスウェーデンのエリーザベト女王やアルノーらと交わした哲学的な書簡の数々などにおいても、広く論じられている。

なお小林はデカルトの「物質即延長」説に基づく宇宙論的自然学が、周知のようにニュートン力学によって乗り越えられはしたものの、そのニュートン力学がさらにまた 20 世紀の相対論と量子力学によって乗り越えられたことによって、結果的に再評価されることとなった（とりわけ一般相対論における「場」の概念や、全体論的な「マッハ原理」などにより）点を指摘している（Cf. 小林 2007: 245-247）。当然この機運は、現代における AI ／ロボット論の文脈にも当てはまるといえるだろう。ただしこれ以上の詳論は別の機会に譲ることとする。

＊4　こうした話題に対して、悲観論（人類は AI に支配される、例えば大ヒットした映画《ターミネーター》のような）でもなく、楽観論（人類と AI は手を携えて〈明るい〉未来を構築できると〈根拠も無く〉信じる）でもない、論者が判断するに中立的で冷静な議論と思われるものとして、例えば西垣 2019 がある（本論文も同著に多くの論拠を負うている）。ただし、宇宙、極地、深海開発など苛酷な条件に対処するため、AI に自律能力が付与されることは大いに考えられる。

＊5　Tromp, John/Farnebäck, Gunnar 2016: “Combinatorics of Go” （論者注：同名論文には 2007 年の暫定版（preliminary version）もあり、区別を要する。2016 年にオランダのライデン大学で三日間にわたり開催された第 9 回国際コンピュータ＆ゲーム会議の初日 1/29 にトロンプが発表した内容が前掲の最終版である）：リンク先は以下（アクセス最終確認はいずれも 2023 年 2 月 2 日）
論文　https://tromp.github.io/go/gostate.pdf
会議　https://www.chessprogramming.org/CG_2016

＊6　カント 2000: 283 訳注 47 における「恐怖」および「恐怖の対象」についての牧野の説明（同書第二八節への誘導）を参照。

＊7　カント 1999: 341（牧野訳者解説）。

【文献（カント以外はアルファベット順）】

Kant, Immanuel 1788: *Kritik der praktischen Vernunft*, herausgegeben von Karl Vorländer, Hamburg 1906（邦訳は『実践理性批判』樫山欽四郎訳、河出書房新社、1989 年；坂部恵／伊古田理訳、岩波書店（『カント全集』7』）、2000 年など）

Kant, Immanuel 1790: *Kritik der Urteilskraft*, herausgegeben von Karl Vorländer, Hamburg 1924 （邦訳は『判断力批判』坂田徳男訳、河出書房新社、1989 年；牧野英二訳、上下巻、岩波書店（『カント全集』8・9』）、1999-2000 年など。なお、牧野英二による「解説」は、カント 1999: 313-344 に所収）

Benedict, Ruth 1946: *The Chrysanthemum and the Sword: Patterns of Japanese Culture*, Houghton Mifflin（ルース・ベネディクト『菊と刀　日本文化の型』長谷川松治訳、講談社（学術文庫）、2005 年）

Bostrom, Nick 2014: *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press（ニック・ボストロム 2017：『スーパーインテリジェンス　超絶 AI と人類の命運』倉骨彰訳、日本経済新聞出版社）

Chivers, Tom 2019: *The AI Does Not Hate You. Superintelligence, Rationality and the Race to Save the World*, Janklow & Nesbit Limited, UK（トム・チヴァース『AI は人間を憎まない』樋口武志訳、飛鳥新社、2021 年）

Descartes, Rene 1996: *Œuvres de Descartes*, publiée par C. Adam et P. Tannery, 11 vol., Paris, 1897-1909; réédition, 1964-1974; tirage en format réduit, 1996（ルネ・デカルト［レナトゥス・カルテジウス］『方法序説』『省察』（原典第 7 巻）『哲学原理』『情念論』『書簡集』等、代表的な邦訳は『デカルト著作集』全 4 巻、白水社、1973 年）

ルチアーノ・フロリディ／ラファエル・カプーロ／チャールズ・エス 2007：『情報倫理の思想』西垣通／竹之内禎訳、NTT 出版（叢書コムニス 05）

Floridi, Luciano 2010: *Information: A very short introductions*, 1st edn., Oxford University Press（ルチアーノ・フロリディ『情報の哲学のために　データから情報倫理まで』塩崎亮／河島茂生訳、河島解説、勁草書房、2021 年）

Floridi, Luciano 2014: *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford University Press（ルチアーノ・フロリディ『第四の革命――情報圏（インフォスフィア）が現実をつくりかえる』春樹良且他訳、新曜社、2017 年）

『現代思想』1998：「特集　主体とは何か」1998 年、10、vol. 26-12、青土社

『現代思想』1999：「特集　システム論　内部観測とオートポイエーシス」1999 年、4、vol. 27-4、青土社

『現代思想』2001：「総特集　システム　生命論の未来」2001 年、2 月臨時増刊、vol. 239-3、青土社

羽生善治／ NHK スペシャル取材班 2017：『人工知能の核心』NHK 出版（NHK 出版新書）

Harari, Yuval Noah 2016: *Homo Deus. A Brief History of Tomorrow*, Harvill Secker, London（ユヴァル・ノア・ハラリ 2022：『ホモ・デウス　テクノロジーとサピエンスの未来』上下巻、柴田裕之訳、河出書房新社（河出文庫）；初版 2018 年）

今泉允聡 2021：『深層学習の原理に迫る　数学の挑戦』岩波書店（岩波科学ライブラリー 303）

稲葉振一郎 2021：『社会倫理学講義』有斐閣（有斐閣アルマ）

河本英夫 1995：『オートポイエーシス』青土社

小林雅一 2013：『クラウドから AI へ　アップル、グーグル、フェイスブックの次なる主戦場』朝日新聞出版（朝日新書）

紺野大地／池谷裕二 2021：『脳と人工知能をつないだら、人間の能力はどこまで拡張できるのか　脳 AI 融合の最前線』講談社

Kurzweil, Ray 2005: *The Singularity is Near. When Humans Transcend Biology*, Viking（レイ・カーツワイル『ポスト・ヒューマン誕生　コンピューターが人類の知性を超えるとき』井上健訳、NHK 出版、2007 年；抜粋訳『シンギュラリティは近い　人類が生命を超越するとき』エッセンス版、NHK 出版、2016 年）

松尾豊 2015：『人工知能は人間を超えるか　ディープラーニングの先にあるもの』KADOKAWA

三宅陽一郎／森川幸人 2016：『絵でわかる人工知能　明日使いたくなるキーワード 68』SB クリエイティブ株式会社（サイエンス・アイ新書）

三宅陽一郎 2017：『なぜ人工知能は人と会話ができるのか』マイナビ出版社（マイナビ新書）

妙木浩之 2022：『AI が私たちに嘘をつく日』現代書館

中野明 2021：『最新　通信業界の動向とカラクリがよくわかる本』第 5 版（図解入門業界研究）秀和システム

新山龍馬 2019：『超ロボット化社会　ロボットだらけの未来を賢く生きる』日刊工業新聞社

西垣通 2018：『AI 原論』講談社（講談社選書メチエ）

西垣通 et al.（編） 2014：『基礎情報学のヴァイアビリティ　ネオ・サイバネティクスによる開放系と閉鎖系の架橋』西垣／河島茂生／西川アサキ／大井奈美編、東京大学出版会

（第 7 章に Clarke, B. / Hansen, M. B. N. 2009: “Neocybernetic Emergence: Retuning the Posthuman,” in: *Cybernetic & Human*

Knowing, 16（1-2); ブルース・クラーク／マーク・ハンセン「ネオ・サイバネティックな創発　ポストヒューマンの再調律」大井奈美訳を所収）
西垣通／河島茂生 2019：『AI 倫理　人工知能は「責任」をとれるのか』中央公論新社（中公新書ラクレ）
大澤昇平 2019：『AI 救国論』新潮社（新潮新書）
Penrose, Roger 1989：*The Emperor's New Mind. Concerning Computers, Minds, and the Laws of Physis*, Oxford University Press（ロジャー・ペンローズ『皇帝の新しい心　コンピュータ・心・物理法則』林一訳，みすず書房、1994 年）
坂部恵／佐藤康邦（編）2008：『カント哲学のアクチュアリティ ― 哲学の原点を求めて』ナカニシヤ出版：黒崎政男「『純粋理性批判』」のさらなる可能性――人はモノに還元できるのか――」pp. 3–31 に所収
『思想』2010：「ネオ・サイバネティクスと 21 世紀の知」、7 月号、No. 1035、岩波書店
髙橋慈子／原田隆史／佐藤翔／岡部晋典 2020：『改訂新版　情報倫理　ネット時代のソーシャル・リテラシー』初版 2015 年、技術評論社
高橋透 2017：『文系人間のための「AI」論』小学館（小学館新書）
田中潤／松本健太郎 2018：『誤解だらけの人工知能　ディープラーニングの限界と可能性』光文社（光文社新書）
柳本光彦 2020-2022：『龍と苺』既刊全 10 巻、小学館
涌井良幸／涌井貞美 2017：『ディープラーニングがわかる数学入門』技術評論社
【叢書・事典等】
『カント事典』有福孝岳／坂部恵（編集顧問）、石川文康／大橋容一郎／黒崎政男／中島義道／福谷茂／牧野英二（編集委員）、弘文堂、1997 年
坂本百大 1998：『岩波哲学・思想事典』（「心身問題」執筆者：坂本百大）岩波書店、1998 年
小林道夫 2007：『哲学の歴史 5　デカルト革命　17 世紀』小林道夫（責任編集）、中央公論新社、2007 年