研究論文

科学の哲学と技術工学の倫理
－ AI に関するホーキングの提言をめぐって －

伊　野　　連

Philosophy of science and ethics of technological engineering: Concerning Hawking's recommendations on AI

INO Ren

〈Abstract〉

In this paper, in order to consider the philosophical and ethical issues surrounding science and technology, I first set up two categories, "philosophy of science" and "ethics of engineering", and then discuss each of them.

Second, I discuss philosophy and ethics from the viewpoint of international relations, especially language. Today the meaning of philosophy and ethics of science is becoming more and more important. This is because the benefits of science are so universal, regardless of race or nationality.

As a specific point of contact between the two, this paper considers the philosophy and ethics of AI. Stephen Hawking's critical remarks are widely known. Hawking is one of the biggest beneficiaries of advanced technology, including AI, and his statements are not the often-misunderstood blind rejection, but rather optimistic; if anything, his optimism rather must even be criticized.

キーワード：シンギュラリティ、ネオ・サイバネティックス、オートポイエーシス、不可知論、楽観主義

はじめに

論者の主な関心は、現代の科学技術をめぐる哲学・倫理学的問題の研究および教育にある。それを具体的に掘り下げて考えると、「科学の哲学」と「技術工学 technological engineering の倫理［倫理学］」という二つの面から展開できると考えられる。そして自然科学の諸分野（医学、工学等）に直結した「応用倫理」諸学が展開していることの重要性についても言及する（第 1 節）。

第 2 節では、もう一つの主題である国際交流という観点から、哲学／倫理学研究の系譜を辿り、現代における科学（その代表例が AI）の哲学／倫理学のあるべき姿勢について言及する。

第 3 節では、哲学的考察の対象である科学と、倫理学的考察の対象である技術工学との、

結節点にしてまさしく集大成ともいえる、現代の人工知能（Artificial Intelligence：AI）を例に挙げ、第 1 節で述べた「科学の哲学」および「技術工学の倫理」という二つの側面[1)]からの考察を実際に施してみたい。

最後の第 4 節では、おそらく 20 世紀ではアインシュタインに次いで最も著名な物理学者であるスティーヴン・ホーキング（英 1942-2018）が AI について述べた「警告」に関して論ずる。これは巷間よくある「AI による人類の支配」（最も極端には「AI による人類の侵略・殲滅」）の例に、名声と権威ある世界的科学者の言としてよく引かれるものである[2)]。

しかし実際には、ホーキングの提言はけっして AI を敵視したり、また、盲目的に強大な脅威の対象とみなしたりしたものではない、ごくごく穏当なもの、むしろ部分的にはその楽観主義が論者には咎められるべきと思われるほどのものである。それについて幾つか述べたい。

そして「終わりに」で、まさしく AI についての（哲学的考察と）倫理学的考察（との双方）に基づく、現代および将来の人類のこの先進技術との関わり方について本論文の観点から、とにかく徹底的に「慎重であれ」という提言を示したい。

第 1 節　「科学の哲学」と「技術工学の倫理」

科学 science が技術 technology と区別されて考えられるという機会は、日常でもよく目にすることである。この場合、「科学」は先端科学やあるいは基礎科学などが主に念頭に置かれ、一方「技術」はというと、元はそうした科学から生み出され、実用化されて実践的に利用（すなわち活用）されているもの、という対比が見られる。

それでは「工学 engineering」はといえば、先に述べた「基礎科学」である数学や化学や物理学などを、工業生産に応用したもの、と目されていて、例えば「数学と自然科学を基礎とし、時には人文自然科学の知見を用いて、公共の安全、健康、福祉のための有用な事物や快適な環境を構築することを目的とする学問」[3)] などと定義されている。

そしてこれはあくまで私見ではあるが、「科学」は哲学的考察の対象ではあれ、その倫理学的な考察は第二義的なものとなると考えられる。それは、科学そのものは中立的で、倫理的価値観（善－悪）といった拡がりないし方向性［ベクトル］を科学は有さないからだと考えられる。

科学に関して哲学的／倫理学的考察を施すことを使命とする者が仮にもこう述べるのは少々勇気が要り、それを聞く者からはややもすれば責任転嫁と批難されるか、あるいは呆れた楽観主義と罵倒されかねぬだろう。

また当然ながら、世に「科学倫理」と題する著書は幾つも存在するし、その中にはたしかに良本も含まれている。手っ取り早くそれらの著者の科学倫理の方途を採り入れればよいのだが、それではやはり模倣の域を出まいと憚られる。

したがって論者自身について言えば、科学そのものを倫理学的吟味にかける、それも新たな方途はいまだ模索中である。それゆえこうした、例えば AI のような、哲学と倫理とに横断的に関わる題材を考察することを通じて、その方途を確立したいと考えている。

一方で、それに代替するという意義からも、科学の具体化・実用化に相当する技術工学について、倫理学的考察に従事する営為もまた不可欠だといえる。

以上のような次第で、本論文では AI を論題として採り上げる。同時に AI がそれ単独では十分な機能を果たせず、その「意図」（AI が人類の頭脳を模倣して開発され、その人類の何に相当するかと仮定して）が実現されるためには、必ず「手」に相当する物が代わりに機能することが求められる、という点も重要な意味を持つ。それはすなわち、AI＝目的、ロボット（・アーム等）＝手段という構図である。

換言すれば以下のとおりである。AI は人類の「脳」（きわめて人口に膾炙したデカルト哲学でいえば「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そのような働きではなく、単なる「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そうした動きの主体だということになる（または、やはりデカルトの所説に従えば、前者が「考えるもの res cogitans」であり、後者が「延長 extentum、あるいは端的に corpus」に相当する）。

すなわちここにも「科学の哲学」がはっきりと認められ、同じように、AI を搭載したロボット等の工業生産物（そしてこの product は同時に他の生産物 products を生み出す生産主体でもある）をめぐる「技術工学の倫理」が問われる、という構図も認められるのである。

第 2 節　国際交流に関する（科学をめぐる）哲学／倫理学

論者は西洋哲学史を専攻とする者であるから、西欧中世における学術的公用語としてのラテン語という歴史的背景を十分に理解しているつもりである。ラテン語を「母語」とする者は一人も存在せぬという状況下で、西欧はもちろん、遠く辺境からも多くの俊才がパリ大学をはじめとする西欧中枢に集い、すべての者が後天的に学んだラテン語を用いて議論を戦わせていたというのは、やはり学問の理想郷であったと思える。

産業革命以降の英仏を中心とする全世界支配によりいわゆるパラダイムは一新された。1600 年頃、イギリスは未だ文化の面では後進国、せいぜい途上国にすぎなかった。17 世

紀にイギリスが主導となって大々的に展開された科学革命、それに続く産業革命（Cf. 伊野 2016 B：54-57）、いわゆる「パックス・ブリタニカ」が 1919 年まで展開されたことは（この三世紀の間に米・加・豪、旧大英帝国から幾つもの大国が独立）、文明≒英をはじめとする西洋文明という構図を決定的に確立させた（ただし 18 世紀末のカント登場によるドイツ哲学の優勢がナチス政権成立まで続いた）。

さて、第一次世界大戦後の英仏の没落と米国の覇権奪取によって、英語圏による文明支配は決定的となった。英から米への継承となったためである。「パックス・アメリカーナ」の成立と展開は第二次世界大戦を経てさらに堅固なものとなる。戦後は半世紀足らず冷戦構造はあったが、東欧崩壊後ほどなく到来したインターネット（元は米国防総省が開発）時代は、ウィンドウズ 95 発売によりほぼ全世界に浸透し、今度こそ英語の支配をほぼ半永久的に決定づけた。「千年王国」の樹立である。このパラダイムの次なる転換は想像を絶するが、例えば文字どおり「メディア」を介さぬ、直接的な意思同士のコミュニケーション手段が確立でもすれば、その時には英語という媒体もまた必要と無くなる事態となるかもしれない。だが AI による高性能翻訳装置の方が現実的である。「十分に熟達した、非母語話者レベルの英語力」を、AI で実現すればよいだけである。ただし AI は言語運用に際しての帰納的推理についてはまだ開発されて日が浅い。

このように、17 世紀末から半永久的に継続していく英語による支配は、必ずしも学問的な理由ばかりで出来したわけではなかった（学術語としての仏・独語、古典語としての希・羅語の存在意義も今日なお保たれている）。

・哲学研究の高水準国日本

このへんで、国際交流としての哲学／倫理学研究における使用言語の話題に戻ろう。

一、中世期のラテン語はまさに国際学術語の名に十二分に値するものであった

二、産業革命後のイギリスの支配は比較にならぬほど決定的だった。ただし、学術語としてはやはりフランス語、そして新興国のドイツ語が秀でていた。

三、1930 年代からは政治的理由による英仏独の衰退ないし混乱（亡命等を含む）のため、および自己の成長により米が伸長した。それにともない米が文化的にも世界の拠点となっていった。

四、アングロサクソン系による言語支配は、むしろインターネット時代で決定的となる。英語がシステムの標準言語であるため、英語以外のすべての言語は「文字化け」する事態となる。

話題を明治中期の日本の西洋哲学研究に移そう。

哲学研究者以外のために説明すると、我が国の哲学研究の水準は、欧米圏を除けばずば

抜けて高いどころか、本場・現地（例えばカントであればドイツ本国）に次ぐ、分野によっては本場さえも凌駕している場合すらままある。

もっとも、これには以下のような若干の断わりが必要である。

1. 特定の哲学者についての研究にとどまればそれは哲学「学」であり、本来の哲学は、或る哲学者の出身国や使用言語等を超えて普遍的なものたるべきである。或る国や言語（もっぱらドイツのそれ）に過度に依存する傾向は、我が国の哲学研究が創造性あるいは独創性に欠けることを説明する顕著な例としてよく挙げられる。
2. 世界トップクラスの研究水準で、本国すら凌駕するというのになぜ国際的にさしたる評価を得ておらぬかといえば、ほとんどそれは発表言語が日本語だからという理由による。我が国の定評ある先行研究がもっぱら英語（仏・独語でもなく）で著されていたら事態は一変しているであろう。
3. 科学の哲学／倫理学は、とりわけ国際社会における普遍性が必要である。それは文字どおり科学の普遍性のゆえで、バイオエシックス（医療における技術やアクセス権の格差）や環境倫理（例えば米中はじめ各国の主張による非協力態勢）といった不統一とも次元を異にし、研究開発は大国が主導となっているが、その恩恵は国境や肌の色を問わない。

だが最後に、「負」の国際交流ともいうべき、国際紛争について補足しよう。まさしく「AI の軍事利用」である。「軍産複合体」、「科学技術社会論」、「産官学」さらに「産官学民」、「産学連携」等々、喧喧囂囂たる議論に飛び交うキーワードが多数存在する。

タテマエを言えば「AI を兵器に用いるとは言語道断」であろうが、現実に紛争がおこなわれているどころかたった一日として無くならないことを鑑みれば、AI と戦略兵器および軍機能の代替について考えぬことは逆に不誠実である。例えば小野 2019 は AI 開発史を概観し、21 世紀における軍の知的労働の代替可能性も緻密に論じている（同氏は論者と同じく羽生善治の AI 将棋論についても強い関心を示しつつ、紛争における戦局は前提（将棋盤は 9×9 升）すら時々刻々と変化する点を指摘している。Cf. 小野 2019：14）。

さらに、宇宙開発を頂点とする、きわめて苛酷な状況下（極地、深海、高山、密林、砂漠等）での AI 搭載ロボットによる作業に際しては、AI に自律能力を持たせる必要が出て来る可能性があり、それは当然ながら「科学の自重」をめぐる議論を喚起する。もちろん、その成果が軍事転用される危険性は火を見るより明らかであろう。問題は、哲学／倫理学に携わる我々の提言が、現実的には脆弱であること（「既成事実」に直面して、それを無かったこととしたり、過去に戻したりすることはほぼ不可能である点、「衆寡敵せず」ではないが、輿論の圧倒的支持に沈黙しないまでも、それを覆す説得力を発揮するのはき

わめて困難である点、など。Cf. 伊野 2016 B：22; 28; 69-70）である。我々はそれを十分に自覚しており、だからこそ新情勢に絶えず関心を向け、より説得力ある提言をなすべく努めている。

第 3 節　「AI が人類を支配するかもしれない」という危惧

未来論者レイ・カーツワイルの著書で話題になった「シンギュラリティ［技術的特異点］」をめぐり、AI と人類との地位の逆転、さらには前者による後者の支配などが広く話題に上るようになっている。

例えば宮家 2018 では AI の急速な普及とそれがもたらす爆発的な影響力拡大について、政治／経済／軍事（同書の題名「（新・）地政学」がそれを物語る）等、哲学や倫理学（応用倫理学は別として）がややもすれば不得手としがちな、現代の現実社会を舞台に国際的に展開している各方面の関心に応え論じており、特に同書のキーワードとして現実社会の「ダークサイド」が挙げられている。AI はそうした不安材料を生み出す要因として消極的な評価が下されている。

さらに栗原 2019 ではさらに、ずばり「AI 兵器」が題材となっている。SF 映画の古典《2001 年宇宙の旅》における叛逆コンピュータ「HAL」どころか、もはや映画《ターミネーター》シリーズをも髣髴させる。だが、2022 年 10 月現在、戦闘のハイテク化は凄まじい様相を呈してはいるものの、世界の何処かで〈AI 兵器〉なるものが戦闘を繰り広げているという事実は無い（そもそも厳密な意味で人類の知能を完全に模した「AI」すらも、いまだ世界には生み出されていないのである）。

たしかにこの後で詳述するホーキングも、彼の最後の著書『ビッグ・クエスチョン』で AI 兵器には危惧の念を表明している。「世界中の軍部が、ターゲットを自ら選んで抹殺する自律兵器システムの軍拡競争に乗り出そうとしている。国連ではそんな兵器を禁止するための条約が議論されている……」。あるいは、「どんどん高度になる AI システムを、長期的に支配できるかどうかも定かではないことを思えば、AI を武装させ、AI に私たちの守備を任せてしまってよいのだろうか？……自律兵器の軍拡競争をストップさせるなら、最善のタイミングは今だ」（ホーキング 2018：203-204）。

そして、ここでホーキングが言及している「自律 autonomy」は、倫理学における最重要キーワードの一つなのである。それは人類の人格が人格たる証であり、仮にもその自律能力が AI に備わってしまうならば（それは人類の生物的進化とは決定的に異なり、あくまで開発者による意図的な改変に拠らざるを得ない）、AI 開発の規制に対する容認論を吹き飛ばしかねない。なぜなら AI に関する根本理解として、AI はネオ・サイバネティカ

ル[4]でオートポイエーシス的（＝オートポイエスティック。自己－創出的ないし自己－制作的）な科学技術の産物ではあるものの、しかし完全なオートポイエーシスではなく、むしろはっきりとアロポイエスティック［＝他者－創出的］な存在である（Cf. 西垣 2017：64）はずであり、また断じてそうでなければならないからである。この「核心」をなす点について詳説しよう。

論者は夙に、人類の人類たる所以は（当然ながら無数に存在すれど）、例えば「自己言及」および「再帰」にあると考えていた。そして西垣ら AI 理論の専門家もまた、AI と人類との比較論考において、これらの点に言及している（例えば、西垣 2017）。

そしてそれはホーキングも同様である。彼は「進化ということから考えて、ミミズの脳と人類の脳との間には定性的、すなわち質的な違いはあるはずがない」と述べ、また「原理的には、コンピュータは人類の知性を真似 emulate できるし、人類の知性よりも優れたものさえエミュレートできる」とも述べる（Cf. ホーキング 2018：200）。

しかし（そしてだからこそ）ホーキングは「何らかの人工知能 AI が、人類よりも上手に AI を設計し、人類の力を借りずに自らを再帰的に改良できるようになれば、人類がカタツムリよりも頭が良いというレベルを超えて、機械が我々よりも賢くなる「知能の爆発」に直面するかもしれない」（ホーキング 2018：200-201。強調引用者）と危機感を表明するのである。

すなわち、自己言及／自己創出／自己制作と再帰性こそが、AI 批判（この「批判 Kritik」はカント哲学における意味、すなわちその可能性と限界との見極めの意）の「核心」なのである。論者がホーキングからの引用文中に強調した箇所に着目すれば、論者と西垣とホーキングが、期せずして同じ観点から AI の可能性と限界とを見据えていることがわかるはずである。限界とは、AI は単独ではきわめて高精度で圧倒的な処理能力を有する電算機（ただし AI を組成するコンピュータは元来たんなる計算機械ではなく、人類の正確な思考活動を再現する理想的な機械と見なされている。Cf. 西垣 2012：203）ではあるものの、それは外部からの命令によってのみ統御され、また、自己や他の AI に手を加えることが可能となるロボットアームを有していないということである（Cf. 伊野 2023 A）。

したがって、西垣 2017 等も強調するように、この一線を越えてしまえば、まさしくホーキングも危惧する「知能の爆発」（ホーキング 2018：200-201）によって、AI が人類を圧倒的に凌駕しついには征服するという半ば都市伝説は、文字どおり「我々が犯す最悪の過ち」（ホーキング 2018：201；バラット 2015）として現実化しかねないのである。

このように、本節および次節でホーキングの AI 観についてなお詳しく採りあげるが、

それらはもっぱら倫理に関すること、すなわち「技術工学」倫理の問題である。20 世紀を代表する理論物理学者である彼は、AI 開発の当事者にはもちろんなく、「車椅子の宇宙物理学者」として世界的にもよく知られていたように、ALS［筋萎縮性側索硬化症］という重度の身体障碍を抱える、きわめて脆弱な一個人にすぎなかった。そんな彼は当然ながら、AI に対する大いなる期待（先端科学技術は彼の文字どおり生命線を支えていた）とともに、ひとたび情勢が激変すれば、自らはその圧倒的な威力にまったく翻弄されてしまうだろうという不安にも苛まれていたのである。

すなわち AI は、当然ながらホーキングがその成立にも開発に関与しているわけでもなく、またホーキング自身も自らの専門である宇宙物理学からの考察対象とも考えていなかったはずである（もちろん AI は宇宙物理学の発展にも大いに貢献するが、それは別問題である）。なぜなら、AI 論はホーキングがまず自らを関連づけ、そして社会全般をも関連づけるかぎりでは、第一義的にはあくまでその倫理的な是非を問う文脈においてのものであり、論者が別に考察しているような哲学的関心ないしは論理の哲学的基礎づけからのものではないからなのである。

第 4 節　人類の叡智の本質的な限界と楽観主義

ここまででも既にたびたびにホーキングについて言及し、その文章も引いてきたが、この最終節ではさらに詳しく彼の所説を見てみる。

第 3 節で既に引用したように、たしかにホーキングは AI のあまりにも急激な進化に危機感を抱いてはいる。その根拠となるのは、やはりよく知られている「ムーアの法則[5]」（p=2^n/2）である。人類の進化が数百万年単位で観測される周期であるのに対して、コンピュータのそれはあまりにも速く、その差は歴然である。と同時に、ムーアの法則の限界も既に指摘されている。2010 年代後半には、半導体の開発ペースが鈍化し始めたからである。

しかしホーキングは（ムーアの法則の限界を否定する意図から言及したわけではなかろうが）、さらなる懸念材料をも指摘している。それは「量子計算」である。「量子計算は、計算速度を指数関数的にスピードアップさせることで、人工知能に革命を起こすだろう」（ホーキング 2018：211）。

ただし量子計算の開発はテクノロジーの進化がもたらす驚くべき成果である点で、元来は喜ばしいことなのである。したがってここには文明の逆説という、典型的な図式がある。すなわち、人類の科学技術が進化すればしたで、もちろん我々の多くはその恩恵に与れるのではあるが、しかし同時に危惧すべきことが霧消するわけではけっしてないという、冷

厳なる事実である。一難去ってまた一難、或る苦境をようやく脱したかと思えば、次の新たな局面でやはり我々人類を追い詰めるような難題がまた現れるのである。

それは科学と人類との間の本質的なジレンマであるようにも思える。レスピレーター［人工呼吸器］の発明と実用化が 1967 年 12 月の世界初の心臓移植を機に脳死という問題を突きつけたように、また、中世の黒死病［ペスト］がもたらした暗黒時代はもはや過去のものとなり、20 世紀半ばには長らく人類を苦しめ続けた結核という不治の病を克服した、天然痘も根絶したと思ったら、今度は現代人に癌というさらなる強敵が立ちはだかるなど、そんな具合にである。

カント批判哲学が示したような、人類の叡智のやはり本質的な限界（それも、思索が深遠になればなるほど、どこまでもなくなることのない永久に続くアポリア［袋小路］）を人類はいつまでも見せつけられ続ける。

ノーベル物理学賞に輝く米国の実験物理学者レーダーマンは物理学の限界について、シェイクスピアの名戯曲に登場する世界で最も有名な主人公を引き合いに出し、従来の難題をようやくすべて解決し終えたかと思うと「ハムレットがぴょんっと現れて Hamlet pops up」、この世にはまだまだ謎があるんだよ、と思い知らせてくれるのだと歎息しつつ語っている（Cf. Lederman 2013: 125; Ino 2022: 127. ハムレットは友ホレイショーに、永遠に続く不可知について語っている）。

さて、二十代の若さで ALS という不治の病を宣告され、一度は絶望の淵に叩き落されながらも不屈の精神（そしてそれは優れた自然科学者特有の資質[6]である「楽観［楽天］主義 optimism」であるはずなのだ）で先も述べたように「車椅子の宇宙物理学者」として世界にその名を知らぬ者はおらぬほど有名になったホーキングこそ、テクノロジーの恩恵抜きにしては、一日として生きること能わざる存在であった。健常者が酸素を必要とするように、彼は AI に象徴されるような科学技術・技術工学を必要としていた。「世間では私は、人類という種の未来について楽天家 optimist だとされているが、私自身はそうとも思えない」（ホーキング 2018：203）と彼は弁明する。しかし本論文でこの後見ていくように、そんな彼は彼が自覚するよりもっと根源的には「楽天家」である。

ホーキングは理論物理学者で、実験物理学者戸塚洋二（注 6 参照）やその師小柴正俊とは立場が大きく異なる（彼ら実験物理学者の普段の研究活動はほぼ肉体労働の従事なのである）とはいえ、ホーキングが四肢を動かすことはもちろん、発語することも出来ぬほど悪化し、コンピュータ音声によってようやく当時の妻子や周囲の人々とかなり円滑なコミュニケーションが可能となった事実などを鑑みれば、むしろ楽天的でなければ生きることそのものにおいてすべての望みを絶たれていたであろう（そのため神の恩寵を否定する彼は、

バチカンで時のヨハネ・パウロ二世教皇から、天地創造説に親和であるビッグバン理論への貢献が表彰される段になって、当惑を隠せずにいた)。

したがって、ホーキングは AI（そしてそれに象徴される先進テクノロジー）に自分を筆頭とする全人類への恩恵を期待しつつ、かつその暴走への倫理的な懸念をけっして忘れずにいたと判断できるだろう。

しかしホーキングはこうも言っている。「火を使い始めた人類は、何度も痛い目を見た後に消火器を発明した。核兵器や合成生物学、強い人工知能といった、もっと強力なテクノロジーについては、あらかじめ計画を立てて最初からうまくいくようにしなければならない。なぜならそれは一度きりのチャンスになるかもしれないからだ。我々の未来は、増大するテクノロジーの力と、それを利用する知恵との競争だ。知恵が確実に勝つようにしようではないか」（ホーキング 2018：213)。この段落は同書で AI について論じた章の結びである。

だが、この見解はあまりにも楽観的すぎるのではないか。たしかに、原始時代と現代の高度科学技術文明とでは、火の扱い方の違いには雲泥の差があるだろう。しかし、人類はまだ身の周りの火事からすら解放されてはいないではないか。大袈裟でなく、我々は火一つとっても、完璧にコントロールできてはいないのである。

個人的な回顧を述べたい。論者がまだ二十歳になる前の頃、当初は原子力発電所の開発に従事し、後にそれを理論的専門家の視点から厳しく批難し、反原発運動の理論面における指導者ともなった高木仁三郎（1938-2000）の講演を聴きに行ったことがある。講演後、個別に質問に行った論者の「先生は、原子力開発の失敗は、もはや「成功の母」とはならぬ、人類にとって取り返しのつかないものとなるとお考えですか」との問いに、氏ははっきりと「そうだと思います」と答えた。

家が燃えたらなんとか消火はできるが、事によると被害も甚大である。大規模山火事など、現代の消火技術をもってしても容易に消せない火事は枚挙に暇が無い。21 世紀の高度文明社会においてもこうした被害はなおも絶えず（米国で毎年ハリケーンによる大きな被害が生じることを現代文明をもってしても防ぎきれぬように)、そのたびに我々は自身の無力さを痛感する。

ましてや今、我々の前に立ちはだかる最難の火は原子力のそれである。プロメテウスの火が人類に文明を齎しはしたが、その「第三の火」は人類にとって永久に手に余る物なのかもしれず、失敗を重ねてもいつの日か核融合（「第四の火」）を実現させる、などというのは自然科学者のあまりにも甘い見通しにほかならぬのではないか。もはや「失敗は成功の母」とはならぬのではなかろうか。

だが科学の進歩は本来むやみに妨げるべきではないはずである。科学の可能性と限界とを知り尽くすよう努めること。それが科学倫理の抱える最も大きな課題の一つなのである。

しかしそれをホーキングのように「知恵が確実に勝つようにしよう」（ホーキング 2018：213）と口で言うのはあまりにもたやす過ぎる。

応用倫理学ではリスクとベネフィットについて相関関係（relative で「比較」「比量」という意味合いもある）を問う。ホーキングは次のように言う（特に彼のような重症者ならではの実感こもった発言だろう）。「我々の頭脳が AI で増幅されたら何ができるようになるかは、予測もつかない」（ホーキング 2018：210）、あるいは「コミュニケーションの未来は、脳とコンピュータとのインターフェイスにあると私は信じている」（ホーキング 2018：211）。しかしそこには、文字どおり「予測もつかない」分、不測（想定外、さらにはもっと極端には「意想外」）の事態すら起こりかねない。

これも私は戸塚の著書から教わったことで（Cf. 伊野 2023 B）、彼が好んで引用する、米を代表する理論物理学者で、ハッブル宇宙望遠鏡プロジェクト推進の中心人物として貢献した（戸塚にとってはまた、スーパーカミオカンデの観測データをもとにきわめて有益な提案を何度もしてくれた、長年頼りにしてきた人物である）ジョーン・バコール（1934-2005）が次のようなことを述べていたという。

よく「宇宙望遠鏡で何が観測できるのか」が問題とされるが、バコールによれば、できそうだと期待されたものを見つけるだけなら「ちっとも面白くない anticlimactic」。最も重要な発見は、想いもよらなかったもの、「どういう風に尋ねたらよいかわからない we do not yet know how to ask」もの、「これまでに想像もしたことのない we have not yet imagined」もの、そんな発見なのだ、とバコールは言うのである。

だが基礎研究はこういう立場で進めねばならない、そしてこれこそが研究の醍醐味である、それが戸塚の科学観・科学理念なのである（Cf. 伊野 2023 B）。

しかしこれは科学によって齎される想いもよらぬ展開として、正の面だけでなく負の面についても当然当てはまるはずなのである。それこそ想定外どころか「意想外」である。カオス理論を念頭に置けばすぐわかるように、我々人類が自らの所業の結果をすべて想定内に置くなどは無理もいいところであり、多くの場合我々は泥縄で仕出かしてしまった事態に必死で取り組まねばならぬのが宿命［さだめ］なのである。

また、CRISPR/Cas 9［クリスパー・キャス 9］に関連する遺伝子改変についても「遺伝子操作の最良の意図は、遺伝子を修正して、起こってしまった突然変異を元に戻すことにより、科学者が遺伝病を治療できるようにすること」（ホーキング 2018：211-212）と述べており、あくまで「元に戻すこと」が前提で、飛躍的向上（いわゆるエンハンスメン

ト）などを望んでいるわけではない。ましてや「元に戻すこと」すら叶わぬ事態に陥った場合は、いったいどうなるのであろうか。

「知能とは、変化に適応する能力と特徴づけることができる」（ホーキング 2018：212）、そして「変化を恐れてはならない。必要なのは、その変化を我々に役立つものにすることだ」（ホーキング 2018：212）とも、さらに「我々は今、素晴らしき新世界の入り口に立っている。危険な面もあるにせよ、それは胸躍る世界であり、我々はその世界の開拓者なのだ」（ホーキング 2018：212）とも彼は力強く語ってはいる。ここには彼の祖国の先人オルダス・ハックスリー（1894-1963）の有名なディストピア小説『すばらしい新世界』（1932）を連想させるものがある（「すばらしい新世界」は邦題で、原題は Brave New World である）。

だが人類が自らにそう言い聞かせるには、それに応じた熟慮、およびそこから導き出された根拠とそれに応える実践とが必要である。遺憾ながら、ホーキング生前最後に近くなされたこれらの提言にはそれらが脆弱であり、十分な慎重さや「悲観的な」姿勢に欠けるきらいがあると言わざるを得ないのである。

終わりに　「失敗から学ぶこと」、「失敗して消え去ること」

ホーキングは「キルスイッチ」（ホーキング 2018：207）についても言及している。先に「叛逆コンピュータ」の例として挙げた《2001 年宇宙の旅》の「ハル」を「機能不全を起こしたロボティック・コンピュータ」として引き合いに出し、「ハルとともに宇宙船に乗り込んだ科学者たちにとって、キルスイッチは役に立たなかった」とまで述べている（Cf. ホーキング 2018：208）にもかかわらず、「AI のことを、なぜそれほど心配しなければならないのでしょう？／人類はいつでも好きな時に、AI のプラグを引き抜くことができるのでは？」（ホーキング 2018：214）と述べている。これはやはり、いくぶん根拠を欠いた楽観主義と言わざるを得ない。

キューブリックはこの映画（1968）の前に製作した《博士の異常な愛情 Dr. Strangelove》（1964）で、次作のクライマックスを連想させるような、しかしあくまで地球上における、米ソの全面核戦争をコミカルに描いていた。

キューブリックは戯画的に描いたつもりだったろうが、現実世界はもっと怖しく戦慄的である。

それはもちろん他愛もない妄想の産物ではなかった。1983 年 9 月 26 日 0 時 40 分、当時ソ連の防空軍中佐であったスタニスラフ・ペトロフ（1939-2017）は、米国から彼の祖国へ向けて発射された一発のミサイルを感知する計器の警告を受け、熟慮の末（しかしそ

れはなんという切羽詰まった、それも短時間での決断であったろうか)、監視衛星の警報システムにコンピュータ・エラーが発生したための誤報と断じた。

果たしてそれは重大なシステム上の欠陥であったことが調査の末明らかとなり、軍上層部を慄然とさせることとなったのであるが、しかしペトロフは報告義務違反の咎で左遷され、さらには神経衰弱を患う羽目にまでなった。彼が報復攻撃による世界全面核戦争から人類滅亡を救った英雄と讃えられるようになったのは、ソ連は疾に崩壊し、晩年に年金生活を送っていた 1998 年のことである。これは失敗学ならぬ成功学だが、我々が後学と出来るような普遍的原則は遺憾ながら導出困難である。

論者は科学技術の倫理をめぐる、ホーキング書からの引用と考察を締めるにあたり、かつて拙著に綴った次の一節を引き、ささやかな反論の証として擱筆したい。「これらの歴史を知ることで、人間がもっぱら失敗からしか学べないことがわかる。倫理学は、そうした人間の性向を承知した上で、失敗を未然に防ぐという困難な課題を背負っているのである」(伊野 2016 B：22)。

注

1) 論者はこれまでにも AI の哲学および倫理学について考察している（Cf. 伊野 2023 A)。

2) 一例として、手元の新書の表紙がカラー刷りで、センセーショナルな黄色を配して大きく(「人工知能の真価は人類の終焉を意味する」)「ホーキング博士が警告し」と綴られている（Cf. 松田 2013)。こうした装幀・意匠から、広く読者が抱えているであろう「不安」を煽る構図がありありと看て取れる。そしてこうした例は枚挙に暇が無い。

3) 「工学とは？」 徳島大学工学部・先端技術科学教育部 HP（tokushima-u.ac.jp）による。

4) サイバネティクス cybernetics は広く知られているとおり、ウィナーの同名の著書（1948）によって提唱された、「人工頭脳学」(通信工学と制御工学とを融合し、生理学、機械工学、システム工学を統一的に扱う)。その語源「舵取り者 $\mathrm{k}\upsilon\beta\varepsilon\rho\nu\acute{\eta}\tau\eta\varsigma$ キュベルネーテース」から取られた英語「サイバー cyber」は、もはや現代科学文明の象徴となって日本語にも広く浸透している。

5) 後の米国インテル社創業者の一人ムーアが 1965 年に自著論文で示した見解。彼は最初、1965 年に、集積回路あたりの部品数が毎年 2 倍になると予測し、この成長率は少なくとも 10 年は続くと予測。そして実際に 1975 年には、次の 10 年を見据え、2 年ごとに 2 倍になるという修正予測した。その予測は 1975 年以降も維持され、それ以来「法則」として知られるようになった(なおホーキングは引用箇所では「一年半ごと」(ホーキング 2018：200-201）としている)。

6) 皮肉にもホーキングと同年生まれで、やはり同じくノーベル物理学賞を逃した実験物理学者戸塚洋二（2008 年直腸癌のため没）は、自らのような国際的プロジェクトのリーダーに必要な資質として「楽観主義」を挙げている（Cf. 伊野 2020；伊野 2023 B)。

参考文献（アルファベット順）

Barrat, James（2013）*Our Final Invention. Artificial Intelligence and the End of the Human Era,* William Clark, New York（ジェイムズ・バラット：『人工知能　人類最悪にして最後の発明』水谷淳訳、ダイヤモンド社，2015 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）

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 introduction,* 1 st 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）

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）

ホーキング、スティーヴン（1988）『ホーキング、宇宙を語る　ビッグバンからブラックホールまで』カール・セーガン序文、林一訳、早川書房

Hawking, Stephen（2018）*Brief Answers to the Big Questions,* Space Time Publications Ltd.（ホーキング、スティーヴン『ビッグ・クエスチョン　〈人類の難問〉に答えよう』青木薫訳、NHK 出版、2019 年。なお引用に際して、意味を変えないかぎりで一部の語彙を改めた箇所がある）

INO, Ren（2022）“A Debate Between Physicists and a Philosopher Over Kant. With an Introduction to Jaspers on Hamlet and Agnosticism”、『開智国際大学紀要』vol. 21-2、pp.127-138.

伊野連（2016 A）『哲学・倫理学の歴史』三恵社

伊野連（2016 B）『生命の倫理　入門篇』三恵社

伊野連（2020）「戸塚洋二の死生観 ― 自然科学者の自己客観視 ― 」関東医学哲学・倫理学会編『医療と倫理』第 12 号、pp. 21-36.

伊野連（2023 A）「AI とロボットに関する倫理と哲学 ― カント批判哲学、「美意識」と「おそれ」について ― 」三重大学人文学部編『人文論叢』第 30 号、2023 年、pp. 19-35.

伊野連（2023 B）「戸塚洋二、その科学観と死生観」『東洋大学大学院紀要』哲学篇、第 59 集、2023 年、pp. 1-17.

河本英夫（1995）『オートポイエーシス』青土社

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

栗原聡（2019）『AI 兵器と未来社会　キラーロボットの正体』朝日新聞出版（朝日新書）

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）

Lederman, Leon（2013）*Beyond the God Particle,* With Christopher Hill, Prometheus Books, New York

松田卓也（2013）『2045 年問題　コンピュータが人類を超える日』廣済堂（廣済堂新書）

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）

宮家邦彦（2018）『AI 時代の新・地政学』新潮社（新潮新書）

妙木浩之（2022）『AI が私たちに嘘をつく日』現代書館

西垣通（1990）『秘術としての AI 思考　太古と未来をつなぐ知』筑摩書房（ちくまライブラリー 34）

西垣通（2012）『生命と機械をつなぐ知　基礎情報学入門』高陵社書店

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）

小野圭司（2019）「人工知能（AI）による軍の知的労働の代替―AI と人間の共生の問題としての考察―」、防衛省防衛研究所（NIDS）編『防衛研究所紀要』21-2、2019 年、pp. 1-21.

『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

高橋透（2017）『文系人間のための「AI」論』小学館（小学館新書）