

近代の成熟と新文明の出現

人類文明と人工知能 I

NIRA 研究報告書

2017.8

近代の成熟と新文明の出現

人類文明と人工知能 I

公文俊平

多摩大学教授・情報社会学研究所長

本報告書は、NIRA 総研「情報化の挑戦を受ける日本に関する研究Ⅱ」研究成果である。同研究会は私を座長とし、以下のメンバーが参加した。本報告書は同研究会での議論を踏まえ、私個人の主張をまとめたものである。研究会メンバー、またヒアリングさせていただいた方々からいただいたご意見・ご示唆に深謝する。なお、研究会各メンバーの研究成果は別途公開される。

足羽	教史	インクリメント P 株式会社管理部渉外担当部長
鈴木	謙介	関西学院大学社会学部准教授
山内	康英	多摩大学情報社会学研究所教授

EXECUTIVE SUMMARY

われわれは人工知能にどのように向き合っていくべきか。この問題について考えるには、現在進行している近代社会の動きが人類文明にもたらす意味についての考察が欠かせない。現在、私たちが対峙している近代社会の状況は、古いものから新しいものへの交代というよりも、古いものの成熟と新しいものの出現が、同時に生起している状況と考えるべきだろう。私はこれを、重なりという意味の「重畳」と呼んでいる。

本稿では、「重畳」をキーワードに、大きな転換期を迎えている 21 世紀の現状を解釈し、そこから人工知能が引き起こしている問題にどう対応すべきか提言する。

● 「産業の民主化」の進展と、「第二機械時代」の始まりの重畳

近代における産業化は、石炭と蒸気機関による第 1 次産業革命に始まり、19 世紀後半以降の石油と電力による第 2 次産業革命を経て、20 世紀後半、とりわけ 1980 年代以降は、産業のデジタル化やネットワーク化が進んだ。現在は、3D プリンターのような個人用もの作りの機械を使い、モノをデザインし製造する「MAKERS（メイカーズ）の台頭」する時代とも特徴づけられるような、コンピューターとデータによる「第 3 次産業革命」の時代に入っている。換言すれば、これまで寡占的な大資本によって発展してきた産業社会が、個人や中小企業を主体に民主化されようとしている。

しかし、こうした一連の流れを大きく超える方向の変化もあると考えざるを得ない。それは端的に言って、機械の知能化が人間の労働を不要にしてしまうようなエリック・ブリニョルフソン&アンドリュー・マカフィーがいう「第二機械時代」の到来である。これまで人間の肉体的能力を拡張してきた「機械」が、これからは人間の知的能力を拡張するようになるのだ。

つまり、「第 3 次産業革命」に重畳する形で、これまでの産業化とは異質な、産業化の新しい流れもまた始まっている。この流れの中核をなしているのが、特定分野において知的にふるまう能力をもつ「特化人工知能」である。これは幅広い分野で人間のようにふるまう汎用人工知能にそのまま進化していくものではないが、特化人工知能でさえも、人々の生活や、政治経済をも含めた社会制度の全体に及ぼす影響は甚大である。

● 近代文明の成熟と、新文明の出現の大重畳

重畳はそれだけではない。もう一つは16世紀の西欧に始まった近代化の流れと、それを大きく超えるポスト近代化の流れの重なりである。近代化は人間主義（ヒューマニズム）に基づく発展の営みであった。それがいま、人工知能の出現によって、人間の自由意思がプログラムで制御される世界に進むのか、もしくは、人間と人工知能がペアを組んで共生する世界に進むのかの岐路に立っている。つまり、ポスト近代文明のあり方自体が、人工知能の進化の方向によって大きく左右されることが予見される。ポスト近代文明は果たして、人間否定の新文明となるか、人工知能と人間とが共働する新文明となるのか。

さらにそれに加えて、汎用人工知能まで実現が可能になるとすれば、未来の文明のあり方はますます不確実になる。人類文明の完全な崩壊につながる危険性さえ無視できないのである。

● 特化人工知能を人類の福音とせよ

特化人工知能の開発は今後も加速的に進み、数々の便益を私たちにもたらしてくれるだろう。他方、人間の労働が人工知能に代替され、ほとんどが失業者になってしまった一般の人々は「余暇」をどのように過ごせばいいのか、また、「生きる意味」を何に求めていけばいいのか。対応のいかんによっては社会の混乱を招きかねない。

私たちは人間が雇用されなくても人間らしい生活を送れるための制度的な仕組みを整えると同時に、新たな生活倫理を生み出さねばならない。少なくとも特化人工知能に関するかぎり、それをどのように利用し、どのように私たちの環境を整備するかは、「ベーシックインカム」の全面的導入も含めて、「人間主義者」としての私たちの意思決定と行動次第である。

その途を見失うことなく共生に成功すれば、近代文明は真の成熟を迎えつつポスト近代文明と重畳して有終の美をなすことに成功するだろう。人間と人間能力拡張型人工知能がペアを組み、人間の自由意思や自律性を維持しつつ、平和的に共働・共生する社会では、人工知能は人類にとって福音となるだろう。

● シンギュラリティ：汎用人工知能の開発は禁止せよ

汎用人工知能は、特化人工知能とは異なり、言語の意味を理解し、意識をもち、自ら設定した目標を実現しようとする。レイ・カーツワイルに代表されるシンギュラリティ到来論は、そのような汎用人工知能の開発に成功した暁には、知

能爆発によって、人類文明は地球史・宇宙史的な進化の新段階に突入すると主張する。

しかし、こうした見方には私はくみしえない。そのような開発が成功すれば、結果的に人類とその文明は、絶滅に直面する可能性が極めて高くなる。ポストロムは人類が人工知能の最終目標実現のための物質資源（コンピュータロニウム）になりかねないと警告し、ユドカウスキーは「敵対的人工知能」が生まれる可能性を危惧する。もし、これら論者の見方の可能性を否定することができないとすれば、ほとんど唯一の選択肢として残るのが、汎用人工知能の開発を全面的に中止・禁止するためのグローバルな合意の、早急な確立である。それは、人類がシンギュラリティのもたらすとされる「超（トランス）近代文明」「超知能文明」への到達を諦めることを意味するが、それにより人類は持続可能な「後（ポスト）近代文明」を築いていくことができるし、そうすべきだというのが私の結論である。

人類はいま大きな岐路に立たされている。特化人工知能を巧みに発展させつつ人類文明の未曾有の高みに登ることを目指すのか、それともあえて汎用人工知能の開発を目指す危険をおかすのか。人類とその文明の存続を望むなら、汎用人工知能の開発は諦めるしかない。

Contents

第1章	私の近代化ビジョン	1
1	人工知能ブーム.....	1
2	近代化のビジョン.....	1
	—「S字波」としての発展ビジョン—	
	1. 近代化と「S字波」.....	1
	2. 「S字波」を構成する国家化、産業化、情報化.....	2
	3. 近代化の複合的発展ビジョンにおける現代の位置づけ.....	4
	4. 重畳する新文明としての「シンギュラリティ」.....	5
	5. 私の主張.....	5
第2章	人類は人工知能にどう向き合うべきか	7
1	産業化の新たな流れ.....	7
	1. 「第二の経済」の誕生論.....	7
	2. 「第二機械時代」の到来論.....	8
	3. 「大格差」論.....	8
	4. 「ロボットの脅威」論.....	9
	5. 「前特異点」論.....	11
	6. 「産業化II」の出現.....	12
2	高まる人工知能への関心.....	13
	1. 人工知能とは何か.....	13
	2. 2種類の人工知能.....	13
	3. 人工知能と人間との関係：人間の代替か拡張か.....	15

3	「産業化Ⅱ」を代表する「特化人工知能」	16
	1. 特化人工知能の可能性と問題点.....	16
	2. 特化人工知能は人間社会を作り変えるのか.....	19
	3. 特化人工知能は人類を滅亡させるのか.....	21
	4. とりあえずの結論：特化人工知能と平和的に共生する.....	24

4	汎用人工知能	24
	1. 汎用人工知能の実現可能性.....	24
	2. 汎用人工知能の問題点.....	28
	3. 人類の「存続危機」.....	29

5	AI時代の未来と私たちの選択肢	32
	1. それぞれの局面は「交代」ではなしに「重畳」する.....	32
	2. 近代の「突破の成熟」と「成熟の出現」という小重畳.....	33
	3. 「近代の成熟」と「後-近代の出現」という大重畳.....	35
	4. 私たちの選択肢と結論.....	36

補論	超知能と超近代文明	39
	1. さまざまな楽観論者たち.....	39
	2. 「特異点」の到来ビジョン.....	39
	3. レイ・カーツワイルのビジョン.....	41
	4. 齊藤元章のビジョン.....	42
	5. 「シンギュラリティ」論への3つの疑問.....	43

	参考文献.....	48
--	-----------	----

	情報化の挑戦を受ける日本に関する研究Ⅱ.....	53
--	--------------------------	----

第1章 私の近代化ビジョン

1. 人工知能ブーム

「人工知能」が、1950年代と1980年代に続いて、いま3度目のブームを迎えている。クイズのチャンピオンを破ったかと思えば、当分は難しいとされていた囲碁の対戦で、世界トップクラスの棋士に圧勝した。画像認識では、「強化学習（ディープラーニング）」と呼ばれる高度な多変量解析手法の駆使が可能となって、驚くほどの前進が続いている。自動車の自動運転の時代も目前に迫っている。医師や法律家の仕事の少なからぬ部分が人工知能に奪われそうになってきたばかりか、4歳児でもできるのに人工知能（ないしロボット）には苦手とされていた単純な作業や行為の能力も、着実に上がっている。言語認識でも突破がみられそうになってきた。さらに、「ビッグデータ」を収集し駆使することで、人工知能は私たち一人一人の遺伝的特質や履歴、性格や欲求について、本人さえ知らないようなデータを蓄積し、それをもとにして私たちの問いに的確に答えてくれるばかりか、私たちの「コンシェルジュ」としてふるまえるようになるろうとしている¹。それどころか、技術革新の加速化が進めば、そう遠くない将来に人工知能が人間の進化段階や人知を大きく超える技術的「特異点（シンギュラリティ）」が到来するという展望さえ²、ますます多くの人が信じ込むようになってきている。

しかし、少なくとも現時点では、人工知能が人間のような「心」をもち、人間のように「考える」ことはない。それができるめども、たってはいない³。いったい人工知能は今後どのような展開をみせ、どのような影響を社会に及ぼしていくのだろうか。人類は、この新しい技術にどう向き合っていけばよいのだろうか。

本稿では、人類文明の進化史の観点を踏まえつつ、われわれが人工知能にどのように向き合っていくべきかを考察する。

2. 近代化のビジョン — 「S字波」としての発展ビジョン —

（1）近代化と「S字波」

私がかねがね、西欧主導の「近代化」過程は、16世紀後半以降、図表1-1のような「S字

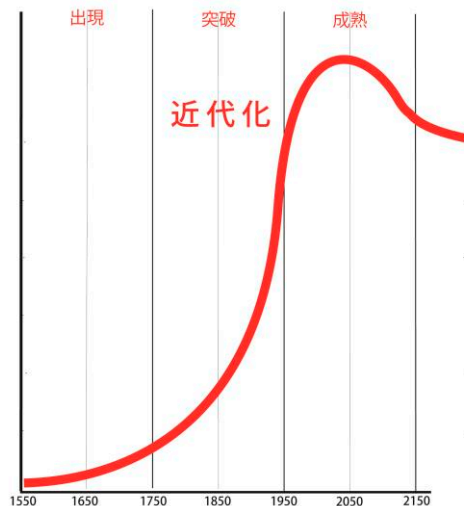
¹ Harari (2017)

² カーツワイル (2007)

³ 松尾 (2015)

波」の形でイメージ化して考えてきた。近代化とは、独立と発展のための手段（パワー）を獲得し蓄積し行使しようとする持続的な営みである。S字波の横軸は「時間」の経過を、縦軸は軍事力／国力や経済力、知能などの社会的「パワー（活力）」の水準を想定している。

図表 1-1 近代化の基本イメージ



(出所) 筆者の案をもとに NIRA 総研作成。

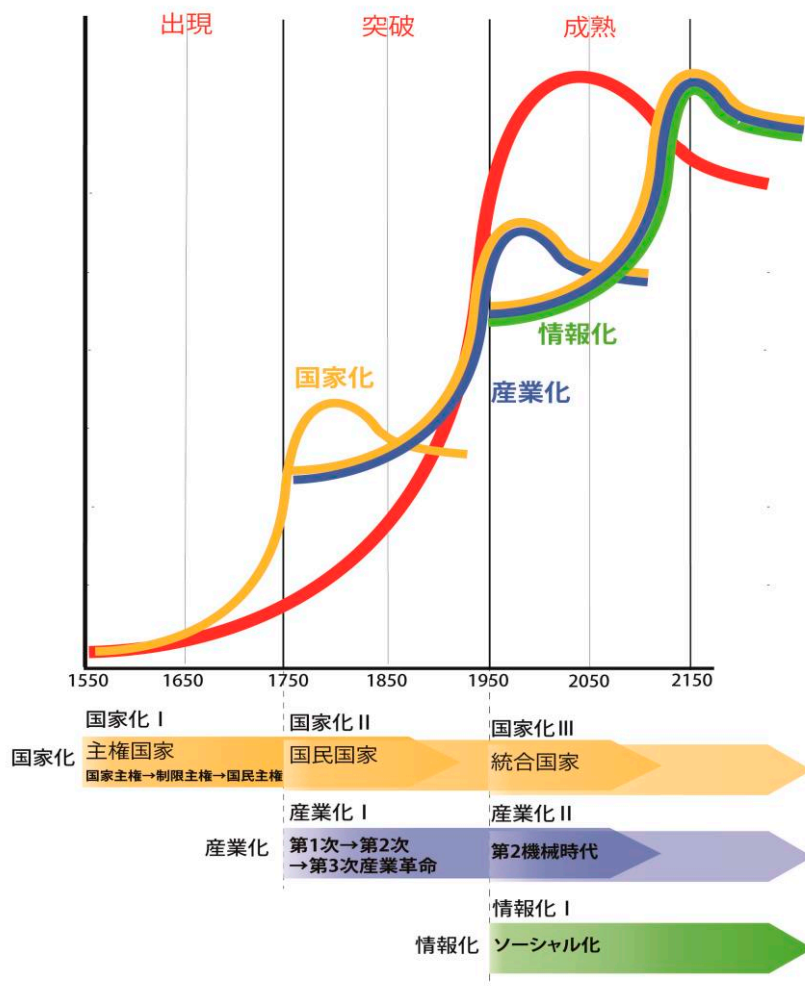
図に描かれている赤線の S 字の字形が示すように、近代化は当初は緩やかに始まり（出現）、時間の展開とともに加速的な発展がみられた後（突破）、ゆきすぎた発展の訂正が行われる中で、安定的な持続から、やがては衰退に転じる（成熟）。つまり、「出現・突破・成熟」という 3 つの「大局面」を経ながら進行するとみなせる。

（2）「S 字波」を構成する国家化、産業化、情報化

この近代化の S 字波は、国家化、産業化、情報化という 3 つの要素によって構成される。まず、16 世紀半ばからはじまる近代化の出現局面は、国家化により規定される。国家主権の絶対性を主張する「主権国家」の出現で始まり、国家の主権への制限（立憲化）や、国民主権の主張（民主化）へと進んでいった。この 16 世紀半ばからおよそ 200 年間の絶対君主制から、制限君主制を経て、人民主権制へと移行する流れは、私が国家化 I（＝主権国家化）と呼ぶものである。

続いて、18 世紀半ば以降に始まる近代化の突破局面では、国家のあり方に革命的転換が起これ、国民主権の理念を基礎とする「国民国家」へと変容がみられた。国家化 I（＝主権国家化）からのこの変容は、国家化 II（＝国民国家化）への転換として把握できる。[図表 1-2]

図表 1-2 複合的发展型近代化のイメージ



(出所) 筆者の案をもとに NIRA 総研作成。

さらに、この突破局面では国民国家化の波と複合して、産業化の波が進んだ。産業化の波は、18 世紀後半の石炭と蒸気機関による第 1 次産業革命に始まり、その後、19 世紀後半以降の石油と電力による第 2 次産業革命を経て、20 世紀後半以降にはデータ化とデジタル・コンピューター化による第 3 次産業革命を経験する⁴。18 世紀半ばから 20 世紀半ばにかけての近代化の突破局面は、国民国家化と産業化という「二重の革命」⁵によってけん引されたのである。

20 世紀後半に入ると、近代化は、成熟局面に歩み入る。国民国家は覇権獲得を目指す 2 つの大戦を経て、EU に象徴される「統合国家」への発展を志向し始める。国家のこうした変容は、国家化Ⅲ (=統合国家化) への転換として表すことができる。しかし、そればか

⁴ 「産業化」という近代化の「大局面」は、「第 1 次産業革命」から「第 3 次産業革命」にいたる 3 つの「小局面」に分解できると考えられる (公文 (2001))。

⁵ ホブズボーム (1968)

りではない、近年のデジタル化やロボット化の動きは、人間の肉体労働が機械化されたこれまでの産業化（産業化Ⅰ）とは質的に大きく異なり、いわば知能が機械化される産業化の新しい波（産業化Ⅱ）、あるいは「第二機械時代（セカンド・マシン・エイジ）」⁶の始まりを告げている。それに加えて、近代化の成熟局面ではもう1つの新たな社会変化の波としての「情報化」も出現してきている。そのきっかけとなったのは、「ソーシャル化」ないし「繋（つな）がる化」と呼ぶことが適切な、新しい文化、価値観の誕生である。ソーシャル化は、国民国家や産業企業に代わる人間・社会関係の新しいあり方を提示している。人びとは、「強さ」や「豊かさ」より「賢さ」に大きな価値を置き、また、「賢さ」を手段に「楽しさ」を理念として追求する。情報社会の「知民」たちが社会的につながる場が「ソーシャル・ネットワーク」であり、「知民」たちの相互理解や共働を支援・推進するための社会的仕組みが「ソーシャル・プラットフォーム」である。

このように、近代化の突破局面が「国家化Ⅱ」と「産業化Ⅰ」の複合として展開したのに対し、近代化の成熟局面は、「国家化Ⅲ」、「産業化Ⅱ」、および「情報化Ⅰ」の複合として展開されようとしているとみられるのである。

つまり、「近代化」とは、

16 世紀後半以降に「国家化Ⅰ」として「出現」し

18 世紀後半以降に「国家化Ⅱ+産業化Ⅰ」として「突破」し

20 世紀後半以降に「国家化Ⅲ+産業化Ⅱ+情報化Ⅰ」として「成熟」する

というグローバルな社会の複合的な発展過程だ、というのが私の近代化ビジョンである⁷。

以上のような「近代化の複合的発展」ビジョンは、図表 1-2 のような S 字波を使ってイメージ化してみることができるだろう。

（3）近代化の複合的発展ビジョンにおける現代の位置づけ

21 世紀の初めに位置する現代は、図表 1-2 からわかるように、2 つの局面が重なり合う重畳した状況にある。すなわち、突破の成熟局面と同時に、成熟の出現局面にいる。

まず、われわれは、20 世紀後半以降の近代の「突破の成熟」局面において、それまでの人類にとっての三重の業苦であった「飢餓、疫病、戦争」を、基本的に解決することに成功したばかりか⁸、残る「環境・資源問題」にも解決の可能性を高めつつある⁹。「経済運営問題」についても、リーマン・ショックを契機として正統派経済学がようやく自らの無

⁶ ブリニョルフソン・マカフィー（2015）

⁷ 「情報化」を「情報化Ⅰ」としたのは、形式的な統一性のためにすぎず、「情報化Ⅱ」や「情報化Ⅲ」が起こればと考えているわけではない。

⁸ Harari（2017）

⁹ ブランド（2011）、リドレー（2013）

知を自覚するようになり¹⁰、マクロ経済政策のあるべき姿についても、新しい合意が形成されつつある¹¹。

同様に、今日の近代化の「成熟の出現」局面では、先に述べたように、情報化（情報革命）が始まるだけでなく、それと複合する形で、国家化の面でも産業化の面でも、大きな転換が起こりつつあると考えられる。すなわち、国家化の面では「国家化Ⅱ（国民国家化）」から「国家化Ⅲ（統合国家化）」への転換が、産業化の面では「産業化Ⅰ（第一機械時代：第1～3次産業革命）」から「産業化Ⅱ（第二機械時代）」への転換が起こっていると考えられる。とりわけ、現在は足踏みしているようにみえる「国家化Ⅲ」に比べると、「産業化Ⅱ」の方は、より順調にというか、むしろ驚くべきスピードで進展している¹²。

（4）重畳する新文明としての「シンギュラリティ」

しかし、話はそれだけではすまない。「近代化」が今日その成熟局面にあるとすれば、これまでの「近代化」とは別の社会変化としての「後（ポスト）近代化」ないし「超（トランス）近代化」も、それと重畳しつつその出現局面に入っているという見方も、当然可能となる。例えば、近年多くの人びとが論じている「シンギュラリティの到来」や「超知能化」が本当に起こりつつあるとすれば、それは、図表 1-1・図表 1-2 に描かれている「近代」の域を超えた、まさしく「超（トランス）近代的」な事象だと理解するのが適切だろう。これについては、第2章と補論であらためて後述する。

（5）私の主張

以上で紹介した近代化の複合的発展ビジョンを踏まえて、本稿では、2つの観点から人工知能について考えてみたい。

その1つは、それを人類文明の進化史の中に位置づけるという観点である。私のみるところでは、今日人類文明は、一方で、16世紀後半の西欧に始まりグローバルに伝播し発展してきた人間中心主義的な「近代文明」の、「成熟」局面を迎えている。他方で、近代文明そのものを超える「後（ポスト）近代文明」ないし「超（トランス）近代文明」の「出現」局面も始まっている。20世紀後半以降、その両方が同時並行というか「重畳」して進行していることに着目して議論を展開している¹³。

¹⁰ ウルフ（2015）

¹¹ ロドリック（2013）、ターナー（2016）

¹² もちろんこの分野でも、さまざまな機会にさまざまな論者が、資本主義の崩壊や産業化の終焉について論じ、「大停滞」時代や「デフレ経済」の到来について語ってきた（コーエン（2011）、水野（2014））。しかし近年の著しい特徴は、その一方で、産業化の新しい動きや技術革新の加速化を指摘する声もまた高まっているところにある（コーエン（2014）、ディアマンディス・コトラー（2014）、フォード（2015b）、井上（2016a））。ただしそれによって、従来型の「資本主義経済」は終わりを告げるかもしれない。（ライシュ（2016））

¹³ 本稿では、到達と持続が可能でも望ましくもある「後（ポスト）近代文明」と、到達が不可能か、あるいは可能だとしても私は到達をめざすべきではないと考える「超（トランス）近代文明」とを、区別して考えることにする。

もう1つは、今日いわれている「人工知能 (Artificial Intelligence, AI)」に、「弱い人工知能」(ないし「特化人工知能 ANI」)と、「強い人工知能」(ないし「汎用人工知能 AGI」)という、2種類のをを区別する観点である。前者が現在すでに実現していてさらに急速な発展を続けている人工知能の姿を表しているとするれば、後者は未来に実現する(かもしれない)人工知能の姿を表している。

上記2つの観点を組み合わせると、「弱い人工知能」を近代文明の成熟局面に、「強い人工知能」を「後(ポスト)近代文明」ないし「超(トランス)近代文明」の出現局面に、それぞれ対応させることができるだろう。それを前提として、私は、次の2つのことを主張したい。

まず、今日の「弱い人工知能」は、それが順調に発展していけば、「強い人工知能」に「自然に」進化するという性質のものではない。しかし、それでも人びとの生活や、政治・経済をも含めた社会制度の全体にそれが及ぼす影響には甚大なものがあるので、それを適切に規制・補完する努力、とりわけ人びとが人工知能の特質を的確に理解して安全な共生・共働が実現できるようにする努力や、人びとの生活を保障し社会を安定させるための制度や政策を導入する試みに、早急かつ真剣に取り組む必要がある、ということである。

次に、現在の弱い人工知能技術の延長線上では強い人工知能の実現は不可能だとしても、新しい発想やアプローチにたてば、いつかはそれが実現する可能性は否定できない。そうなった暁には、そこで「知能爆発」が発生して、その後ごく短期間で人間には及びもつかない高度な「超知能」が生まれる可能性が高い。しかし、それが人類にとって「友好的(フレンドリー)」であってくれる保証はない。そうだとすれば、そこで人類とその文明は、「存続危機」に直面する恐れがある。それを確実に避けたければ、「強い人工知能」の研究開発自体を諦める、ないしは禁止する合意を、人類全体として取り付ける必要がある、ということである。それは人類が「シンギュラリティ」のもたらす「超(トランス)近代文明」への到達を諦めることを意味する。それによって初めて、人類は、持続可能な「後(ポスト)近代文明」を築き上げることができるのではないだろうか。

なお、私自身は技術者でもなければ哲学者でもないのに、「人工知能」や「意識」ないし「心」の問題を専門家として議論する素養も能力もない。したがって、それらの論点をめぐる以下の議論のほとんどは、そうする資格があると私が考える人びとの議論に依拠していることを、あらかじめお断りしておきたい。

次の第2章では、人工知能に関して現在なされている諸議論のポイントを押さえながら、これらの主張について検討を進めていく。

第2章 人類は人工知能にどう向き合うべきか

1. 産業化の新たな流れ

人工知能について考えるには、現在進展している新たな産業化の動きが人類文明にもたらす意味についての考察がかかせない。以下では、現在、私たちが対峙している経済社会の状況を、前章で説明した「近代化の複合的発展」ビジョンを念頭に置きつつ、解釈していきたいと思う。

20世紀後半以降、「第3次産業革命」の展開と並んで、それに連なる動きとは異質の、産業システムそのものの転換が起こっているようだ。それは一見「資本主義の終焉」¹⁴のようにみえるが、その内実は、これまでのような資本主義的な商品化や賃労働のシステムとは異なる経済システムへの転換であって、産業技術の進歩や経済発展過程が停滞するどころか加速する環境の中で、生じている変化だとみることができる。

つまりそれは、これまでの「産業化Ⅰ」を超える「産業化Ⅱ」の出現だと考えられるのである。その点を確認するために、まず、何人かの論者のこれに関連する議論の中から、目ぼしい論点を抜き出してみよう。

(1) 「第二の経済（セカンド・エコノミー）」の誕生論

「複雑系」の研究者として知られる経済学者のブライアン・アーサーは、“The second economy”と題した論文¹⁵の冒頭で、「デジタル化は、巨大で自動化され目にはみえない『第二の経済』を生み出し、それによって産業革命以来の最大の変化をもたらしている」と指摘した。

アーサーによれば、それは現在のところ、情報技術の分野での変化にとどまっていて、遺伝子工学技術やナノテクノロジーは、その先の技術である。それでも現在の「第二の経済」、すなわち「デジタル経済」は、コンピューターの利用やソーシャルメディアやインターネット商取引の域はとっくに超えている。これまでは人間が介在してきたビジネスの過程——例えば空港でのチェックインやサプライチェーンの管理など——それ自体が、そっくりそのまま電子化・機械化されて、人間の目には見えなくなってしまっているからである。この不可視の「第二の経済」は、今後ほぼ20年で現在の可視的な「物的経済」の規模を超える、とアーサーは予想する。過去の「産業革命」が物的経済に対して「機械」という「筋肉系」を与えたのに対し、今日の「第二の経済」は、物的経済に「神経系」を、つ

¹⁴ 水野 (2014)

¹⁵ Arthur (2011)

まり「知能」を、与えつつある。しかもこの転換には上限や終わりというものがない。それは経済の成長と繁栄を持続させる一方で、人間を「無業（ジョブレス）」にしつつ、どこまでも進行するだろうという。

（２）「第二機械時代（セカンド・マシン・エイジ）」の到来論

ブリニョルフソンとマカフィーは、現代を、「第二機械時代（セカンド・マシン・エイジ）」の始まりだという。彼らの定義では、「進歩」とは、「物理的・知的環境を整備・制御してニーズの実現を可能にすること」だが、その２つの主要手段が、「肉体的能力」と「知的能力」である。これまでの産業化では、蒸気機関や電動機のような「機械」が、人間の肉体的能力を拡張してきた。しかし今回拡張されているのは、人間の知的能力である。それを担っているのが「考える機械」としての「コンピューター」であって、これまでの「産業化」の最も中心的な特徴だった「機械化」自体は依然として残っている。だから彼らは、「新しい汎用技術」としての「デジタル技術」が産業化を主導するようになった現代を、「第二機械時代（セカンド・マシン・エイジ）」と呼ぶのである¹⁶。

著者たちは、その前作¹⁷では、デジタル化が雇用に及ぼす衝撃の大きさに驚きながらも、それは教育への投資や移民の受け入れを含むさまざまな対策（19もの項目があげられている）によって十分克服可能で、新時代の経済学は「ゆたかさの経済学」となる、という楽観的な立場を取っていた。しかしこの第２作では、彼らの未来イメージは「ゆたかさの未来」から「ゆたかさと格差をもたらす未来」に変わり、かつての楽観論は影をひそめ、未来のあり方については不可知論的な立場を取りながらも、「運命を決めるのはテクノロジーではない、私たちだ」としている。

（３）「大格差」論

かつて『大停滞』（コーエン（2011）、原著出版も同年）をベストセラーにした経済学者のタイラー・コーエンは、その３年後に、がらりと論調が変わった『大格差：機械の知能は仕事と所得をどう変えるか』¹⁸を世に問うた。彼はその中で「新しいテクノロジーが、旧著で論じた“大停滞”からの脱却の道を開いているかもしれない」として、「活力」のある産業と停滞産業への産業の分極化と、それに伴う人びとの富と所得の格差の大拡大に注目した¹⁹。自身チェスの熱心なプレーヤーでもあるコーエンは、コンピューターが人間のチャンピオンをついに打ち破ったこと、そして最強のプレーヤーは人間と機械がペアを組ん

¹⁶ ブリニョルフソン・マカフィー（2015）

¹⁷ ブリニョルフソン・マカフィー（2013）、原著出版は2011年

¹⁸ コーエン（2014）、原著出版は2013年

¹⁹ 同じ年、フランスの経済学者トマ・ピケティが、『21世紀の資本』（ピケティ（2014））と題する大著を発表（フランス語原著出版はその前年）して、18世紀にまで遡る世界各国のデータをもとに、富と所得の格差の不可避的な拡大傾向を明らかにして、多くの人びとに衝撃を与えた。

だ「フリースタイル」のチームとなったことに、いたく感銘を受けた。そのため、これからは「賢くなった機械」と機械とペアを組めるだけの「高度な一般知能をもった人間」とが、チームを組んで活躍する時代が当分続くだろう、というのがコーエンの現時点での見立てである²⁰。

では、社会はどうなっていくだろうか。いずれ、所得と居住地の二極分化が進む高齢化社会が到来するだろう。政府は、財政難になっても、政治的影響力の強い豊かな富裕者には高率の課税をすることができず、低所得者のための社会保障や社会保険のプログラムは後退するだろう。米国の場合は、多くの人びとが家賃の安い土地に移り住み、なるべくお金を使わずに趣味に生きるようになるだろう。しかし、格差の拡大が政治革命や資産の掠奪、社会秩序の崩壊をもたらすとは考えにくく、未来は、政治的には「奇妙に平穏な時代」になるだろう、というのがコーエンの新たな予想である。

(4) 「ロボットの脅威」論

経済学者コーエンの上のような立場は、一種の「奇妙な達観」ないし「控えめの悲観論」とでもいうべきかもしれない。これに対し、IT技術者で起業家でもあるマーティン・フォードの未来論は、ロボットが雇用に及ぼす脅威を、より深刻に捉えている。

もともとフォード自身も、いち早く自費出版して注目を集めた前著²¹では、原題の『トンネルの中の光：自動化、技術の加速、および未来の経済』にもみられるように、比較的楽観的な見通しを提示していた²²。自動化された経済では、「労働」は生産のための必要な要素ではなくなり、賃金を勤労の対価とみなす発想は誤ったものとなるにしても、「消費者（＝生産物の需要者）の手もとに安定した収入が注ぎ込み続ける仕組み」があればよい。それは、企業への課税と、消費者の意欲的な活動に対する奨励金の給付の組み合わせによって、実現できるだろう。こうして彼の著は、「加速するテクノロジーがほかの経済的制約を克服し、しかもなお旺盛な消費需要を維持できるシステムに進化していけば、私たちはかつてない経済成長と繁栄の時代に突入できるかもしれない」という期待を、結びの言葉としていた。

ところが、6年後に出版された2冊目の著書『ロボットの脅威』²³では、彼の論調は、より厳しいものに転換した。仕事の自動化と機械の智能化が彼の予想以上に急速に進んだからである。もはや「雇用の75%」どころか、これまではロボットには難しいとされてい

²⁰ コーエンの予想では、当分（今後50年）は、この「フリースタイル・モデル」が妥当するだろう。なぜなら、機械の知能は「一夜にして爆発的に向上することはなく、段階を追って少しずつ進歩する」から。しかも、現実には起きつつある進歩は、高度に発達した人工知能という単一のテクノロジーによってではなく、さまざまな機械の能力が組み合わさることによって実現しているから。なお、この線上での議論をはなはだ説得的に展開しているのが、元 DARPA 局長のアラティ・プラバカーの「機械共生時代へ：ヒトとマシンはともに進化する」と題する *Wired* 誌の評論である。http://wired.jp/special/2017/darpa/

²¹ フォード (2015a)、原著出版は2009年

²² 2009年という出版時点は、類書の中では最も早い。邦訳の表題は、『テクノロジーが雇用の75%を奪う』（フォード (2015a)）と、より直截なものになっている。なお日本では情報学者の新井紀子が、ほぼ同じ時期に『コンピュータが仕事を奪う』（新井 (2010)）を出版していた。

²³ フォード (2015b)、原著出版も同年

た単純な手作業から高度な熟練労働や知能労働にいたる、製造業からサービス業や農業にいたる、雇用のすべての分野が、価格が劇的に下がると同時に性能が急上昇する「ロボットの脅威」にさらされている。ロボットたちはいま、頑強な肉体的能力だけではなく高度な知的能力をも、急速に身につけ始めた。その帰結は、人間の能力の増幅や支援のレベルを超える、人間の能力そのものの代替である。

それに対抗して、人びとが高度な教育を——新しい技術革新の力を借りて——容易かつ安価に身につけることができるようになったとしても、その有用性が持続する期間はたかが知れている。たちまちロボットに追い越されてしまうだろう。そうだとすれば、そもそもそうした教育を受けること自体が意味をなさなくなってしまう。他方、新技術は医療産業にも到来し、医療の質を画期的に向上させると同時に、費用を大きく引き下げる可能性が生まれている。しかし、教育産業も医療産業も、これまでのところは、そうした変化の導入には「頑として抵抗してきた」。その結果が、医療費や教育費のとどまるところを知らない高騰と増大である。

しかし、より問題なのは、「デジタルテクノロジーの襲来」が及ぼすマクロ経済的な効果である。機械が人間を置き換えつつ急激に高性能化していく時代の経済は、デフレが持続し所得が分極化していく経済となる。高齢化する人口が労働力率を引き下げるにもかかわらず、失業が増大する一方で²⁴、物価は、コストの低下と需要の減退の両面で低下圧力を受け続ける。低所得層の所得は、賃金の引き下げと増大する失業の圧力でますます下がっていく。加えて、いまは高所得を享受している人びとでも、機械の性能の高度化につれて、職を失い低所得層に転落していく人びとの割合が増える一方となる。

いくら生産性が上がり、いくらでも大量の財とサービスが安価に供給できるようになったとしても、買い手がいなくなれば経済はまわっていかない。だとすれば、活路は経済体制の全面的な改革にしかない。すなわち、全国民に等しくある最低限度の所得（ベーシック・インカム）²⁵を保証することで、消費者のための生活の資と、生産者のための需要とを共に与えるのである。その財源としては、炭素税や付加価値税、累進的所得税、などが

²⁴ ただし、EU諸国はともかく、米国の労働統計の示すところでは、現時点での米国での雇用の総数は、増えてはいないまでも減少する傾向はみられない。しかも、労働力の絶対数は——労働力率の減少にもかかわらず——大きく増加している。とはいえ、おそらく現在の雇用の「停滞」は過渡的なもので、やがて急激な減少が現実化するだろう。

²⁵ ベーシック・インカムの構想は、18世紀末の英国（トマス・スペンスの「スペンス主義」や農業労働者への賃金補助を制度化した「スピーナムランド制」）や19世紀半ばのフランス（ジョセフ・シャルリエ）にまで遡るという（山森（2009））。「国家化I」に対応する「主権国家」の中で「民主化」の動きが生まれた時代のことである。最近の日本でベーシック・インカム論を本格的に展開した著者は、エコノミストの原田泰である。原田は、既存の社会福祉制度をうまく整理すれば、国民1人当たり大人月額7万円、子ども3万円程度の所得を保証することは「財政的に可能」だとしている（原田（2015））。

2016年に彗星のように登場した日本の若手マクロエコノミストの井上智洋も、『人工知能と経済の未来：2030年雇用大崩壊』（井上（2016a））で、「汎用人工知能」が普及すると想像される2030年ごろから「技術的失業」が急速に進み、さらに、経済や社会のあり方が抜本的に変革される「シンギュラリティ」が到来する2045年ごろには、労働している人びとは「全人口の1割」くらいになってしまう「純粋機械化経済」時代が到来し、そのような社会では、ベーシック・インカムの保証が不可欠になるとしている。

井上はさらに、第二作の『ヘリコプターマネー』（井上（2016b））では、経済学界的通説ないし信仰対象である「貨幣の長期中立性」（供給が需要を作るという「セー法則」の長期での妥当性）を真っ向から否定して、画期的な政策的主張を打ち出し、多くの注目を集めた。すなわち、銀行の預金準備を100%とすることで銀行による「信用創造」＝「貨幣発行」を不可能にする一方、政府が、経済の潜在成長率に若干のインフレ分を上乗せした率での潤沢な貨幣を持続的に供給し、国民にはその「発行益」の平等な分配を「ベーシック・インカム」として受けとる権利を認めよ、という主張がそれである。なお、すべての人を受給者とするベーシック・インカムのことを、とくに「ユニバーサル・ベーシック・インカム」と呼ぶが、最近のWired誌は、その導入をめぐる議論がいよいよ本格化しそうだとしている。<http://wired.jp/2017/03/16/robots-stealing-jobs-education/>

考えられる。指數的成長がついに目にみえる大きさとなった今日、こうした抜本的な対応を急いで講じていかななくてはいけない、とフォードは警告している²⁶。

(5) 「前特異点 (プレシギュラリティ)」論²⁷

次に、いま日本で最も注目すべき IT 起業家、齊藤元章の言説に注目しよう。齊藤はもともと放射線医で、医療系画像システムの研究開発会社の起業家としてシリコンバレーで頭角を表していたが、東日本大震災を契機に帰国して、社員がわずか 20 人ほどのスパコンの製造会社 PEZY Computing を設立した。そして、早くも 2014 年には、世界で最もエネルギー消費効率の良いコンピューターランキング GREEN500 で 3 位に入賞、さらに翌 2015 年には 1 位から 3 位までを独占するという快挙を成し遂げた²⁸。齊藤たちはいま、2020 年までに、政府主導の開発になるスパコン「京 (けい)」の 128 倍の性能をもちながらも、安く使いやすい「エクサスケール」の小型スパコンを、民間ベースで複数台製造するプロジェクトに取り組んでいる^{29 30}。

齊藤はその過程で、レイ・カーツワイル流の「特異点 (シンギュラリティ)」到来論への理解と信奉を急速に進めつつあるそうだが、彼がユニークなのは、それ以前に、「前特異点 (プレシギュラリティ)」というべき時代がすでに到来している、という認識をもっている点にある。彼は最初、「前特異点 (プレシギュラリティ)」の時代は、いま彼らが取り組んでいるプロジェクトの成功によって始まるだろうと予想していたのだが、その後の人工知能研究の急激な発展を目の当たりにして、すでに 2016 年が「前特異点 (プレシギュラリティ)」元年となったと考えるにいたった。齊藤によれば、この「前特異点 (プレシギュラリティ)」レベルでも、それが経済や社会にもたらす衝撃の大きさは測りしれないので、人類はまずそれに備えなくてはならない³¹。

では、「前特異点 (プレシギュラリティ)」はどのような変化を人類にもたらすだろうか。齊藤の断言するところでは、

²⁶ ベーシック・インカムについては、それを導入すると人間の勤労意欲が阻害されるという反対論がある。しかし、ここでの文脈は、そもそも「勤労」の必要も機会もなくなるという事態への対応策の一つとして、それが考えられているというものである。

それとは別に、ベーシック・インカムを導入しても勤労意欲は阻害されるどころかかえって高まったというインドでの事例も報告されている (Markoff (2005))。理論的には、その可能性はすでに J. S. ミルによって『経済学原理』(第二版)の中で主張されていたそうだ (山森 (2009))。おそらくそれは、「最低賃金の引き上げは労働の生産性の上昇につながる」という、最近の経済学が発見した関連性 (クルーグマン・浜田 (2016)) と好一対をなすものだろう。

²⁷ 以下の記述は、齊藤 (2014) および齊藤の対談記事 (Wired Vol.20 に掲載。http://wired.jp/special/2016/motoaki-saito/) と、齊藤が講師を務めたある研究会での齊藤の発表によっている。

²⁸ ただし、2016 年になるとこの快進撃は若干の驕りをみせ、11 月には 1 位と 2 位の座は、米国とスイスに奪われてしまった。しかし、2017 年の 6 月には、再び 1 位と 2 位の座を回復している。

²⁹ 齊藤 (2014)

³⁰ 齊藤に言わせれば、スーパーコンピューターの世界では「ムーアの法則」はすでに妥当しなくなっているという見方は誤っている。たしかに、日本の「京」や中国の「天河二号」など、政府主導型の巨大プロジェクトで作られた世界「最高速」の巨大スパコンの性能は、ムーアの法則の傾向線をはるかに下回っている。熱核融合においても同様な遅れが生じている。しかしそれらは政府主導によるプロジェクトの巨大化の結果であって、いま齊藤たちが取り組んでいるような民間ベースの小型スパコンの製造プロジェクトが成功すれば、その性能は再び、ムーアの法則が示す傾向線に戻るといえる。

³¹ もっとも齊藤は、「われわれが前特異点との対峙の仕方を誤り、あるいはその対応に小さくない遅れが生じた場合は、われわれ人類は本当の特異点を見ることなく、その存在が宇宙から永遠に潰えてしまう可能性もある」ともいう。つまり齊藤のみるところでは、人類は、「特異点 (シンギュラリティ)」の到来どころか「前特異点 (プレシギュラリティ)」の到来時点で、すでに「存続危機」に直面しているのである。

第1に、エネルギー問題が解決されるとともに、(都心の一等地などの特殊な地域を除けば)生活に必要な土地は、無償かそれに近い形で安く入手できるようになる。その結果として、「衣」「食」「住」に関する問題は、そのすべてが完全に解決され、生活必需品のすべては、「フリー」と呼びうる状態での入手が可能となる。つまり人類は、生活のために働く必要のない「不労」の時代を迎える。この時代に対応する社会的な方策としては、一方で政府による所得保証、もしくは新しいタイプの通貨の分配による購買力の提供がある。他方では、エネルギー供給装置(例えば小型核融合炉や光合成装置)や食糧製造装置(例えば植物工場)を各家庭に設置して、日用品の生産にあてるという方策がある³²。

第2に、老化の仕組みが解明され、人類は、「老化」から解放されることになる。

こうして、「前特異点(プレシンギュラリティ)」の到来によって人類は、「不労・不老」の時代を迎える。しかし、地球・宇宙規模の環境問題(地球温暖化や太陽のスーパーフレアなど)は、まだ解決されないし、人間の脳の仕組みについても、解明されるべき多くのことが残っている。「サイバー戦争」や「核戦争」への対応を誤れば、人類の文明は悲劇的な結末を迎えてしまうかもしれない。

それでも、今後40年間の変化は、直近40年間の数百万倍の規模になり、人類が対応を誤りさえしなければ、この時代の終わりには、人類は「新生人類」となり、近代文明など比べものにならないほど高度な新文明が出現するだろう。「特異点(シンギュラリティ)」の時代がついに到来するというのが、齊藤の期待してやまないところである。

(6)「産業化Ⅱ」の出現

以上、産業化の新しい流れについてのさまざまな議論を紹介してきた。こうした議論に接するにつけても、今日の「産業化」には、「産業民主化」とか「MAKERS(メイカーズ)の台頭」³³などとして特徴づけられるような「新しい(第3次)産業革命」の流れだけでなく、それをさらに大きく超える方向に向かっている側面もあると考えざるをえない。つまり、これまでの「産業化」とは異質な、まさしく近代の「成熟」を表すような産業化の新しい流れ——これまでの「産業化」を「産業化Ⅰ」とすれば、それに対して「産業化Ⅱ」と呼ぶのが適切な、産業化の新しい流れ——が、「第3次産業革命」に重畳する形で始まっていると考えざるをえない³⁴。

それは端的に言って、機械の「知能化」が人間の労働力を不要にしてしまうような、機械ないし産業活動の「自動化(オートメーション)」から「自律化(オートノメーション)」³⁵へ、あるいは「エコノミー」から「オートノミー」への流れである。そしてそれは同時に、そ

³² 3Dプリンターのような個人用もの作り機械も、これに追加したいところである。

³³ アンダーソン(2012)

³⁴ 最近では、ドイツの“Industrie 4.0”という命名の影響もあって、この変化を「第4次産業革命」とする呼び方が日本でも広がってきているが、私は、これまで3つの「産業革命」として展開してきた「産業化」との異質性を強調するために、「産業化Ⅱ」という呼び方をしたい。

³⁵ Wikipediaによれば、“autonomation”という言葉は、Toyota Production Systemの中で最初に使われたそうだ。

れに全社会的に対応するための所得・需要保証の仕組みが作られ、人びとがもはや「勤労」を「義務」ないし「価値」とみなすことなく、それ以外の活動³⁶に人生の意味をみいだすことができるようになるような社会変化の流れをも、緊急の課題として呼び起こしていることに留意が必要である。

2. 高まる人工知能への関心

(1) 人工知能とは何か

以上みてきたような「産業化Ⅱ」に向かう技術的発展の大きな流れの中核をなしているのが、人工知能（ないしロボット）技術である。近年にわかにならぬ新たな関心が高まっている「人工知能 Artificial Intelligence, AI」とは、いったい何なのか。

自身プログラマーである小野田博一によれば、人工知能とは、「[人間からみて] 知的にみえるふるまいをするプログラムやシステム」（挿入と傍点は引用者）、あるいは、「“人が行うならば知性が必要である処理”をするプログラムやシステム」のことである^{37 38}。

日本の人工知能学会は、「人工知能」をめぐる2つの立場を区別している。1つは「人間の知能そのものをもつ機械を作ろうとする」立場で、もう1つは、「人間が知能を使ってすること——とりわけ「推論」と「学習」——を機械にさせようとする」立場であって、実際の研究のほとんどは後者の立場にたっているという³⁹。

(2) 2種類の人工知能

上の定義や説明からも想像できるように、現在の人工知能（AI）は、

ANI (Artificial Narrow Intelligence, 特化人工知能)、または「弱い人工知能」と
AGI (Artificial General Intelligence, 汎用人工知能)、または「強い人工知能」に

大別されている。前者、すなわち「特化人工知能」は、チェスや碁をプレイするとか、病

³⁶ 「強さ」や「豊さ」ではなく、「愉しさ」を追求し享受しようとする活動、とりわけさまざまな「ゲーム」の普及に注目しよう。これこそ近代化の「成熟」局面にふさわしい活動なのである（カストロノヴァ（2014）、マクゴニガル（2011））。2010年代の初めに作られた「ゲーミフィケーション」という言葉は、この方向に向かう社会変化と密接に関連している（井上（2012））。

³⁷ 小野田（2017）

³⁸ 私なら、「知的」だけでなく、「感情的や意図的にもみえる」を追加したい。

³⁹ 学会のサイトは、<https://www.ai-gakkai.or.jp/whatsai/>である。人工知能をめぐる基本的な知識が簡明に整理されている。数少ない前者の立場を代表する研究者が、デジタル・ゲームのキャラクターを、「身体性」と「主観性」をもつ「生きた知能」として、それを取り巻く「環世界」とともにバーチャルな空間の中に創りだす、という「ゲーム AI」の夢の実現に努めている三宅陽一郎である（三宅（2016a）、三宅（2016b）、三宅（2017））。

気を診断するといった「特定のいくつかの分野だけで知的にふるまう能力をもつコンピューターやプログラム」をいう。特化人工知能としては、すでに 2011 年に、IBM のコンピューター「ワトソン」が、米国の人気クイズ番組「ジヨパディ！」で 2 人のクイズ王に挑戦し、最高金額を獲得している。2016 年には、グーグル・ディープマインド社の開発した「アルファ碁」が、世界トップクラスの棋士を打ち負かして、世界をあっといわせた。さらに同年末にはその改良版の“Master”が登場し、世界のプロ棋士相手に無敗を誇っている⁴⁰。この分で見れば、今後数年の間に、肉体労働ばかりでなく、医療や法曹関係のような知的労働職の多くまで人工知能にとって代わられるだろうという声が、にわかになら高くなってきた。

とはいえ、現在のコンピューターはあくまでも「特化」人工知能のレベルにとどまっていて、「意識」をもつことはおろか、日常的な言語の「意味」を理解することさえできない。

「冷静にみたときに、人工知能にできることは現状ではまだ限られている。基本的には、決められた処理を決められたように行うことしかできず、『学習』と呼ばれる技術も、決められた範囲内で適切な値をみつけ出すだけだ。例外に弱く、汎用性や柔軟性がない。人工知能が人間を支配するなどという話は笑い話にすぎない。」^{41 42}。

これに対し、後者、すなわち「汎用人工知能 AGI」ないし「強い人工知能」は、「問題を解決し、学習し、さまざまな環境の中で効果的かつ人間的な行動を取る能力」をいう。あるいは、「強い人工知能」という概念を作った哲学者のジョン・サールによれば、人間と同じような「心」というか「認識状態 cognitive state」をもつにいたった人工知能のことをいう⁴³。この「汎用人工知能」については、いまの時点では幸か不幸か⁴⁴ まだ実現のめどさえついていないのである⁴⁵。日本の人工知能技術開発の最前線にいる、スタートアップ企業 PFN の丸山宏最高戦略責任者も、「少なくとも、汎用人工知能の実現は、今みえている技術の延長上にはない、というのが私たちの認識です」という⁴⁶。

それにもかかわらず、現在の人工知能研究者の少なからぬ部分は、

⁴⁰ その強さは、Master による多くの対戦を観戦した大橋拓文六段によれば、「人間では理解できない手が 30 手以内に出てくる。しかし、後にそれが良い場所になってくる不思議、マジックのようだった」。さらに、30 手までに「これはおかしい」と不安になり、50 手で「ヤバイ」、100 手で「大差で負ける」。最後は「お稽古してもらっている」気分になったという。<https://www.j-cast.com/2017/01/06287546.html?p=all>

⁴¹ 松尾 (2015)

⁴² 東大入学をめざしたロボット、「東ロボくん」がもっとも苦手としたのも、「国語」の問題だった (新井 (2014))。

⁴³ サールが、1980 年の論文、J. Searle, 1980, "Minds, Brains and Programs", *The Behavioral and Brain Sciences*, vol. 3 (<http://cogprints.org/7150/1/10.1.1.83.5248.pdf>) で初めて行った区別。彼はそこで、「中国語の部屋」という後に有名となった比喩をもちだして、「強い人工知能」の可能性については否定的な立場を打ち出した。

⁴⁴ あえて「幸か不幸か」と書いた意味は、以下の行論の中で明らかになるだろう。

⁴⁵ 「チューリング・テスト」——文章での対話を通じて、相手がコンピューターか人間かを見破ろうとするテスト——の考案者であり、「人工知能の父」(の 1 人)とも呼ばれているアラン・チューリングは、人間並みの人工知能は 2000 年ごろに実現すると予想していたそうだが、彼の予想は外れてしまった (Shanahan (2015))。現在でも、「人間が人間に感じるような明確な心を持つロボットは作られていない。」例えば「怒りを作り出す感情の機能がアンドロイド——無意識には人間として認知されるヒト型ロボット——にはない。」にもかかわらず、人間の方がロボットには感情があると思いついてしまうことはしばしばある。「感情の本質とはそういうものかもしれない。」(石黒 (2009))

⁴⁶ だから丸山は、まだ海のものとも山のものともつかない汎用人工知能がもつかもしれない危険性を恐れるあまり、特化人工知能の真剣な開発がなおざりにされてはならないと主張している。<http://bit.ly/2pTtU0j>

- ① より多くのコンピューティング能力、例えば現在の世界最高速のスパコンの数
百、数千倍の能力と、
- ② より「スマート」なコンピューティング能力、例えば人間の脳を模したコンピ
ューティングや、「遺伝的アルゴリズム」によるコンピューターの進化

など、人工知能研究がさらに発展すれば、そう遠くない将来、例えばいまから 15-20 年後に、人間並みの「強い人工知能」が実現する可能性は、決して小さくないと考えている⁴⁷。

(3) 人工知能と人間との関係：人間の代替か拡張か

それでは、人間にとっては「人工知能」とは何なのか。

もともと「コンピューター」ないし「ロボット」や「人工知能」と人間との関係については、ニューヨーク・タイムズのジョン・マルコフ記者がその 2 冊の著書で執拗に追求してきたように、2 つの異なる立場が最初からあった。

その 1 つが「人工知能」という言葉を創った当の本人で、「自動化 (オートメーション)」の積極的な推進論者でもあるジョン・マッカーシーによって代表される、「人工知能＝人間能力の代替 (サブスティテューション)」という立場である。もう 1 つは、コンピューター技術者のダグラス・エンゲルバートによって代表される、「コンピューター＝人間能力の拡張 (オーグメンテーション)」という立場である。そしてマルコフ自身は後者に強い共感をよせてきた⁴⁸。1960 年代以降の PC 産業の展開を跡づけたマルコフの著書 Markoff (2005)⁴⁹ では、もっぱら後者の立場が貫かれていた。しかし、近著 Markoff (2015)⁵⁰ では、この二分法がうまくなりたたなくなってきたことを、マルコフ自身渋々ながら認めざるをえなくなっている。なぜなら、「人工知能」の場合、技術の発展に伴って、2 つの立場を区分する境界は、事実上融合する方向に向かっているからである⁵¹。

そうだとすれば、要するにどちらを目指すかは設計者個々人の考え方がかんだということになるが、それでも、どちらが目指されるかによって、未来の社会での人間ないし労働

⁴⁷ オックスフォードの哲学者ニック・ボストロムが 2013 年に行った AI 専門家へのアンケートによれば、回答者の 50%が、2040 年までに実現するとみている。2075 年までなら 90%がそうみている。そして、回答者の 75%が、いったん「強い人工知能」ができれば 30 年以内に、人知をはるかに上回る「超知能」に到達するだろうとみている (Bostrom (2014))。米国の著名な編集者のジョン・ブロックマンは、The Edge Question と称して毎年その年のホットなトピックを取り上げて各界の識者の意見をまとめて紹介しているが、2015 年のトピックは「考える機械」だった (Brockman (2015))。そこに収録されている 200 名近くの回答者による意見の集約と分析は、NIRA 総合研究開発機構が行っている (公文・羽木 (2016)) が、「超知能」はともかく「強い人工知能」のレベルの機械なら、そう遠くない将来に実現するだろうと考えている人が多いことには驚かされる。

⁴⁸ 日本では、「人間拡張工学」を提唱・推進している東京大学の稲見昌彦たちが、この立場にたっている (稲見 (2016))。

⁴⁹ 原著は 2005 年の出版。服部桂による邦訳 (2007) の題名は、『パソコン創生「第三の神話」：カウンターカルチャーが育んだ夢』。

⁵⁰ 『人工知能は敵か味方か：パートナー、主人、奴隷——人間と機械の関係を決める転換点』(2016)。原題は、*Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots* (愛すべき機械たち：人間とロボットの間の共通基盤の探求) である。

⁵¹ 現に例えば、マイクロソフト社の人工知能研究者であるエリック・ホーヴィッツが長い時間をかけて開発してきた「秘書ロボット」は、人間の能力を拡張すると同時に、人間に取って代わる機械にもなっている、とマルコフはいう。

者のあり方には決定的な違いがでてきうるだろう。しかし未来の方向づけ自体はまだ決まっていない。いま必要なのは、両陣営の間に十分なコミュニケーションがなされることではないかというのが、マルコフの近著の問題意識である。

3. 「産業化Ⅱ」を代表する「特化人工知能」

(1) 特化人工知能の可能性と問題点

さて、人工知能研究者の「予想」もしくは「願望」はともかくとして、「心」をもつ「汎用人工知能」(強い人工知能)に現時点ではまだ実現のめどがついていないことは、間違いない。他方、「心」をもたない特化人工知能(弱い人工知能)の方は、現在ただいま、急速な発達をみせている。その意味では、現在、「産業化Ⅱ」を代表する技術・産業となっているのは、この特化人工知能の方だと考えてよいだろう。

そこで、「汎用人工知能」については次節で取り上げることにして、ここではまず、「特化人工知能」の可能性や問題点に注目してみよう。

東京大学の松尾豊特任准教授は、特化人工知能に対する現在の「世間の見方」を、4つのレベルに整理することを提案している⁵²。すなわち、

- 〈レベル1〉 簡単なルールに従った単純な制御プログラムを「人工知能」と称しているだけのもの。
- 〈レベル2〉 複数のルールが組み合わさり、対応のパターンが多彩なもの。「推論・探索」型の将棋のプログラムや掃除ロボット、あるいは「知識ベース」をもとに質問に答える型の人工知能などが、このレベル。
- 〈レベル3〉 教えられた着眼点をもとに、うまい対応パターンを自動的に学習(機械学習)するもの。検索エンジンに内蔵されていたり、ビッグデータをもとに自動的に判断したりするような人工知能。
- 〈レベル4〉 対応パターンの学習に使う変数(着眼点)も自分でみつけるもの。ディープラーニング(特徴表現学習)を取り入れた人工知能。(いままではこれができなかった。)

の4つがそれである。松尾にいわせると、「最近の人工知能」というと、この〈レベル3〉のものを指すことが多く、これに対する〈レベル4〉が、「いま最もホットな領域」となっ

⁵² 松尾(2015) および海猫沢(2016)。

ている⁵³。

他方、日立製作所の矢野和男技師長は、現在の特化人工知能を、その目的の違いによって3つに分類している。「運転判断型」、「質問応答型」、「パターン識別型」の3つがそれぞれある。チェスや囲碁をプレイする人工知能は「運転判断型」、人の問いかけに言葉で答える人工知能は「質問応答型」、画像を見て人の顔を識別する人工知能は「パターン識別型」といった具合である。矢野によれば、現在最も急速に発達しているのは、人間の助けを借りることなしにビッグデータを処理して、その適切な分析技術や有用な仮説を自力で作ることができる、「運転判断型」の「学習するマシン」であって、矢野自身が開発した「人工知能H」もその一種だという⁵⁴。これは上の松尾の分類でいえば、〈レベル4〉にあたるだろう。

知性と知能

これに対し、現時点での最強の将棋 AI、「ポナンザ」の開発者、山本一成は、近著の中で次のような人工知能観を示している⁵⁵。すなわち、知的能力は、

- 1) 知性：目的を設計できる能力
- 2) 知能：目的に向かう道を探す能力

の2つに分けられ、後者はさらに

- 2-1) 探索
- 2-2) 評価

に二分できる。「探索」とは、所与の目的達成のために取ることができる行為の範囲を調べ上げることだが、多くの場合、その数は最強のコンピューターをもってしても歯が立たないほどに大きくなる。そこで、その範囲の限定（探索の「枝刈り」）が必要になるが、その手段が「評価」である⁵⁶。ある行為の実行によって目的の達成にどれだけ近づいたかを、

⁵³ 松尾のいう〈レベル3〉や〈レベル4〉の領域での研究が、この1、2年の間にいかに急激に進展しているかについては、IT企業家の清水亮の衝撃的なレポート——多数の最先端の研究者との対話を含む——に詳しい（清水（2016））。松尾自身も、最近の研究の進展の速度は実にめざましく、数年前に自分が予想していたよりも2倍も3倍も早くなっていると認めている。松尾によれば、機械はいまやついに「眼」をもった。すなわち「画像認識」ができるようになった。これは、5億年前の生物の「カンブリア爆発」に匹敵する大きなできごとである。古生物学者のアンドリュー・パーカーは、「眼の誕生」が「カンブリア爆発」の原因となったとしているが（パーカー（2006））、同じことは人工知能についてもあてはまるのではないか。こうして始まる人工知能の応用分野の爆発的多様化を前提として、次の段階では、機械（ロボット）は「運動に習熟」するだろう。そしておそらく今後10–15年のうちに、「言語の意味理解」さえある程度までできるようになるにいたるだろう。つまり、「意味理解を伴う自動翻訳」が実現されるだろう。これが現在の松尾の予想である。

<http://www.jfrc.go.jp/cprc/conference/index.files/170331data02.pdf>

⁵⁴ 矢野（2014）

⁵⁵ 山本（2017）

⁵⁶ 山本は、「探索」のことを「エミュレーション」と、「評価+探索」のことを「シミュレーション」と呼んでいる。棋士の羽生善治は、将棋の例を引きながら、人間は「美意識」と「時間」感覚によって、良い「手」を一貫性をもたせつつさし続けていくことができるという（羽生・NHK（2017））。

さまざまな「特徴量」の値を設定して計算するのである⁵⁷。最初は、特徴量の選定やその値の決定や調整は、プログラマーが勘と経験を頼りに膨大な時間をかけて行っていた⁵⁸。

しかし、現在の人工知能がもっている「評価技術」には、大別して次の3つがある。すなわち、

- 2-2-1) 機械学習：基本的な評価基準はプログラマー（教師）が与えるが、評価値の調整は機械が行う技術⁵⁹
- 2-2-2) 強化学習（教師不要の機械学習）：未知の環境であっても人工知能が投機的に調べた結果をフィードバックし、それを反復することで、学習が進む技術
- 2-2-3) モンテカルロ法：直接の評価が困難な場合に、それに代えて、ランダムな試行の反復結果をもとに、有望な行為を確率的に選択する技術⁶⁰

がそれである。機械学習には、「ロジスティック回帰」や「ランダムフォレスト」などさまざまな手法があるが、現在、画像認識に始まり、言語（翻訳）や音声認識の面でも、文字通り日進月歩の発達をみせているのが、「ディープラーニング（深層学習）」と呼ばれる、人間の脳神経ネットワークのような「ニューラルネットワーク」の技術をさらに発展させた技術である。この技術のもつ潜在能力の限界がどこにあるかは、まだよくわかっていない。

ともあれ、こうした「機械学習」が日常化することで、人工知能は、「プログラマーを卒業する」。つまり、プログラマーの役割がなくなる。さらに、人工知能の生み出した極めて強力な評価の基準や方式は、なぜそれが有効かが論理的・還元主義的には理解できなくなった段階で、人工知能は「科学をも卒業する」⁶¹。

さらにそれに「強化学習」が加われば、人工知能のレベルは「指數的」に成長し始める。そして「予想もできなかったスピードとタイミングで」人間を突然抜き去ってしまう。将

⁵⁷ どのようにすれば目的が達成できるか——例えばどうすれば将棋に勝てるかとか、ある画像はなぜ「犬」だと判断できるのか——をきちんと論理的に説明することは、人間にはできないことが多い。だから、目的の達成手順をコンピューターのプログラムにするためには、別の手法が必要になる。

⁵⁸ 例えば将棋の場合だと、「飛車」をもっていると1500点を与える。実戦の経験を重ねながら、それを1520点とか1480点に変更してみるのが「調整」である。

⁵⁹ 山本は、「機械学習は、人間がどうやって学習しているかわからないけれど、人間の真似ができるようにするコンピューターの技術だ」という。

⁶⁰ モンテカルロ法は、評価の仕方がわからなかった囲碁に最初に導入され、一定の成果はあげたものの、途中で立ち往生していた。だが後に、それとディープラーニングを使った打ち手予測の技術が組み合わされ、さらに強化学習を加えた局面の直接評価技術が組み合わされたことで、「アルファ碁」のような最強の囲碁AIが誕生したそうだ。

⁶¹ 山本はそれを、「解釈性と性能のトレードオフ」と呼んでいる。人工知能の性能が上がれば上がるほど人間はその理由を説明できなくなるからである。他方、そもそもいまの人工知能（コンピューター）は「論理が非常に弱い」。つまり囲碁や将棋でいえば、何十手も先を「読む」ことなど到底できない。かつては、人間には説明も理解もできないが、現に有効な——例えばゲームに応用すると極めて強力な——評価の方式は、それを人間のプログラマーが「雑巾しぼり」などと呼ばれるような職人的・経験的なやり方で反復試行して発見していた。そのころは、理由が理解できぬまま、高精度な評価を見つけ出す技術のことは、「黒魔術」と呼ばれていたそうだ。いまでは、高度に複雑化してなぜそれが有効か理解できぬままに、職人芸的なノウハウに頼って発展を続けている「ディープラーニング」の技術が、「黒魔術の塊」化している。さらにそれに「強化学習」が加われば、人工知能は科学も「卒業」してしまい、ついには意味がわからぬままに無条件な信奉の対象とされる「宗教」になるのではないかというのが、山本の予感である。現に、いまの時点においてさえ、囲碁や将棋のプロたちはすでに、人工知能の打つ「手」を競って真似し始めているという。

棋の例でいえば、天才的棋士でさえ想像も及ばぬような新手や新戦法を繰り出し始める⁶²。そこではもはや、人工知能と人間の天才の知能のレベル差は、天才と普通の人の、あるいは普通の人と昆虫の、知能レベル差など及びもつかぬほど開いてしまったのである。そうした事態は今後、囲碁や将棋のようなゲームだけでなく、ほとんどあらゆる「知能」の分野で出現することになるだろう。それが山本のいう「コンピューターが天才を卒業する日」の到来である。

しかし、現在の人工知能は、ますます高い「知能」をもつ方向に急速に進んでいるとはいえ、「知性」、すなわち「目的設計能力」⁶³を備えるにはいたっていない⁶⁴。それでも例えば、「複数のディープラーニングをつなげたディープラーニング」によって、いつの日にか、人工知能は知性を獲得できるかもしれない。それが、人工知能が「人間を卒業する」日にほかならない。

(2) 特化人工知能は人間社会を作り変えるのか

以上みたように、特化人工知能にもさまざまなタイプやレベルのものがあるが、すでにさまざまな分野で、そのパフォーマンスは人間（の知能）を追い抜きつつある。自動運転車や知的なデジタル・パーソナル・アシスタント（DPA）の登場はその前触れにすぎない。MITで人工知能研究の最先端を走っているマレー・シャナハンは、特化人工知能のレベルでさえ、人工知能は人間社会を大きく作り変えてしまうだろうという。いわく、

ますます高度化する特化人工知能の最も明白で直接的なインパクトは、勤労の領域にみられそうだ。それは産業革命の延長線上にあり、それが意味するところも善かれあしかれそれと同様だ。一方で自動化の増大は、生産費を減らし、経済成長を刺激し、労働時間の短縮や生活水準の上昇と寿命の延長をもたらす。他方でそれは、失業をもたらす、伝統的な生活様式に脅威を与え、富と権力と資源を少数の人びとの手に集中させる。

しかしながら、高度な人工知能技術は、おそらくある重要な一点で、前の世代のイノベーションとは異なる。これまでは、「新技術はそれが無くすのと同じだけ多くの職を作り出す」、と論ずることが可能だった。職はなくなったのではなく、農業から工業、そしてサービス業へと移っていったのである。しかし今回はそうはいかないだろう。賃金をもらえる職は大幅に減少する可能性がある。

⁶² 将棋の「矢倉」と呼ばれる、それまではプロ棋士にとって最も本格的とされてきた戦法は、それではどうしても破れない新戦法（左美濃急戦）を人工知能が編み出した結果、捨てられつつあるという（羽生・NHK（2017））。

⁶³ 山本によれば、「目的設計能力」とは、「意味」と「物語」を持つことで、「最終目的」だけでなく「最終目的」達成にとっての手段となるさまざまな「中間的目的」まで含めて設定・変更していく人間的能力をさす。いまの人工知能にはそれが欠けている。しかし、その分、特定の「意味」や「物語」によって課せられる制約には囚われていないともいえる。

⁶⁴ そのことはよく分かっているのに、それでも非常に強力になった人工知能（例えば自分が開発し改良を重ねた将棋 AI の「ポナンザ」）の振る舞いを見ていると、そこに「目的」や「意思」や「感情」を感じたりするようにはなる、というのが山本の実感である。同じことは、多くのプロ棋士たちも感じるそうだ。

その帰結としては、いろいろな方向が考えられる。

まず考えられるのは、社会の分断と格差の増大である。高給を得られる職が少数のエリートの手集中する。しかし失業者もその基本的な生活の必要は十分に満たされ、豊かな時代を享受できるならば、それは、より平等な社会の到来でもある。教育の機会が均等化し、創造性が広く促進され報酬も大きくなる。社会的価値をもつ余暇活動が制度化され貨幣的価値が付与されるようになり、賃労働と余暇の区別がなくなる。文化的活動の栄える時代が来る。ただしその実現には、かなり強い社会的および政治的な意思が必要とされるだろう。少なくとも特化人工知能技術の時代には、過去からの強い慣性が働き、社会は容易には変化しないだろう。

しかし、情報技術の発展とともに人びとのそれに対する依存度が高まり、その傾向は高度な人工知能の到来によってさらに強まるだろう。その結果何が起こるか。ネオラダイトが主張するように、人間性や自律性が失われ自由が脅威にさらされるようになるだろうか。世界を直接経験したり、自分で意思決定したり、自由意思で行為したりすることが阻まれてしまうだろうか。あるいはその逆に、情報技術は、その支持者たちが主張するように、人類の進歩を促進するだろうか。各人の世界への見識を広げ、これまでは不可能だった形でほかの文化や新しいアイデアに接することができるようになるだろうか。他人との交流が促進され、人びとは、知識や情報の民主的な交換や思想の自由の促進によって、エンパワーされることになるだろうか。

真実はその中間にあるだろう。便益も多いがその享受には代償も伴うだろう。未来のためにわれわれが受けている挑戦は、高度な特化人工知能の到来に際して、その便益を最大にしつつコストを抑えることだ⁶⁵。

以上が、特化人工知能の時代の社会変化についてのシャナハンの見通しである。これは比較的楽観的な見方だといえるだろうが、これに先にみた齊藤元章流の「不老不労」の「前特異点 (プレシンギュラリティ)」時代の到来という見方を追加してみると、より悩ましいイメージが浮かび上がってくる。例えば、「不老」、すなわち極端な長寿化が、極端な少子化と同時に進行したとしたら、社会の人口構成は、逆ピラミッド型になってしまう。そこでの教育や福祉のシステムは、どのように設計すればいいのだろうか。「不労」すなわち、労働力を必要としないままで社会的生産力が加速的に増大していくとすれば、万人の生活を保障するための生産物（および所得）の社会的（再）配分の現実的な仕組みは、どのようなものをどのように導入すればいいのだろうか。ほとんどが「失業者」になってしまった一般の人びとは、「余暇」をどのようにすごせばよいのか、「生きる意味」を何に、どこに、求めればいいのか。さらに、グローバルにみれば、「前特異点 (プレシンギュラリティ)」時代に到達して、これらの社会的問題への適切な対処に成功する社会と、対処できなくて混乱

⁶⁵ Shanahan (2015)

や紛争にまきこまれる社会との間のギャップの発生が考えられる。また、到達そのものができずに「中所得国の罠」にはまったまま抜け出せないでいる社会との間のギャップや対立が、20世紀の「南北問題」に匹敵するような大問題となるかもしれない。齊藤も懸念しているように⁶⁶、私たちはこの段階ですでに的確な対処を誤ってしまう危険もあるといわざるをえない。

そうならないようにするために、全社会的な意図的な対応が必要である。そして、それは早ければ早いほどよいだろう。

（3）特化人工知能は人類（ホモ・サピエンス）を滅亡させるのか

より不吉（？）なというか、より驚くべき分析と予想もある。ベストセラー『サピエンス全史』⁶⁷の著者として知られるイスラエルの歴史家ユヴァル・ノア・ハラリが、その続編 *Homo Deus* ⁶⁸で行っている議論がそれで、その概要は、次のとおりである。

飢餓・疫病・戦争による死亡という3つの業苦を、20世紀において基本的に解決できた人類にとっての21世紀の課題は、不老不死と幸福の実現になる。それは、人類（ホモ・サピエンス）がその能力（パワー、技術力）を「神」のレベルにまで高めて、自らが「神人（ホモ・デウス）」となることによって達成されると期待されている。

狩猟・採集時代のサピエンスは、「精霊」を含むあらゆる存在、あらゆる生き物と対話し交流していた。しかし、農業革命は、「有神論 theism」の宗教革命を引き起こし、サピエンスは、自分たちだけが不可分・不変・不滅の「魂」をもつと信ずるようになり、他の生物種とは異なる特別な地位を確立して、いまや自分たちだけの「神」をもつようになった。神はサピエンスを救い、サピエンスは神に貢ぎ物を捧げるという「契約」が成立したのである。

これに対し、近代のサピエンスは、新しい宗教の一種としての「人間主義（ヒューマニズム）」と科学との盟約・協調のもとに生きるようになった。人間主義の中心的教義は、人間の「経験」こそが意味と権威の源泉である。それによって人間は（自らの内的経験をもとに作り出す）「意味」（と「価値」と「権威」）を、（それ自体としては意味をもたない）宇宙に対して与えられるようになった。これが近代におけるサピエンスの大成功をもたらした。すなわちサピエンスは、産業革命と科学革命を通じて、自らが「神」の地位を占めるにいたった。かつての「有神論」に「人間主義（ヒューマニズム）」がとってかわることで、「神は死んだ」のである。

しかし20世紀のヒューマニズムは、

⁶⁶ 先の注31を参照

⁶⁷ ハラリ（2016）

⁶⁸ Harari（2017）

- ① 人間個々人の経験と自由を重視する正統派としての「リベラル・ヒューマニズム」(米英など)のほかに、
- ② 多くの他者の共通感情を無謬の「党」が体現する「社会主義的ヒューマニズム」(ソ連など)、および、
- ③ 対立と闘争(戦争)の中で勝利した強者の生存を重視する「進化論的ヒューマニズム」(ナチスなど)

の3派に分裂し、それらの間のイデオロギー戦争が起こった。

その結果は、リベラル派の勝利にみえたが、リベラル派の願望(不死、幸福、神人化)を満たすべく開発される新技術は、サピエンス自身をも超えて進歩し続け、リベラル派の立脚点である人間の「経験」そのものを掘り崩すかもしれない。

そのきっかけは、生命科学における新たな発見にあった。それは、次の2点に要約できる。

第1に、人間は単一不可分の“individual”としての「自己」などもっていない。人間は可分的存在(dividual)である⁶⁹。

第2に、人間には「自由意思」はない。人間(やその他の生物)は、他の事物と同じく「アルゴリズム」によって動いている。前者は生化学的アルゴリズムに、後者は電子的アルゴリズムに従うが、その基礎となる数学法則は同一だ。だから、人間の欲望や行為は、適当な刺激を脳に加えることによってコントロール可能だ。そうすると人間の「自己同一性」も、「イデオロギー」や「倫理」も、すべて無意味なものとなる。生命と無機物の間の壁も消失する⁷⁰。サピエンスは、自分が「自由」ではなく、「経験」は操作可能なものだと思ったのだ。

そこに生まれたのが、「データ教 Dataism」という新しい宗教である。その中核的教義は、「宇宙はデータの流れからなり、現象や事物の価値はそれがデータ処理にどれだけ寄与するかによって決まる」というものだ。いまや価値は「経験」それ自身にあるのではなく、それを自由に流通する「データ」とするところにある。すなわち、それを記録し、アップロードし、通有するところにある。

しかしそうだとすれば、一方では、アルゴリズムで動くコンピューター——「心」などもたない「特化人工知能」——の方が、人間の人間的ニーズによりよく応えることができる。人工知能は人間の欠陥(病気など)を克服する手段としては極めて強力で、そのかぎ

⁶⁹ 例えば、行動経済学の創始者、ダニエル・カーネマンが発見したように、人間は、「経験する自己」と、それを「記憶する自己」という2つの自己をもっている(カーネマン(2012))。

⁷⁰ 例えば、「資本主義」と「共産主義」、「民主制」と「独裁制」は、データ処理の2つの方式(分散処理か集中処理か)にすぎないとされる。いずれにせよ、技術変化の速度が速くなりすぎると、動作に時間がかかる民主主義の政治システムは、対処不能になる。独裁制も同様である。いまや「市場」も「政府」も「陰謀家」も、世界を制御できなくなっているのだ。

りでは人間の健康と幸福とパワーは、当初はたしかに増進するだろう⁷¹。他方では、人間は、自らの生物的身体性を超越していけるようになる。つまり、自らを「アップグレード」して自身が「神人 (Homo Deus)」ないし「越人 (トランスヒューマン)」に転身することが目指せるようになるのである。

しかしここに落とし穴がある。そのような「アップグレード」が機械としての人工知能の手を借りることで実現するのなら、権威は人工知能に移り、人間はコンピューターの1つのチップのようなものから、最後は単なるデータの一部になってしまい、人間的価値そのものが意味をもたなくなるだろうからである。こうして、「ヒューマニズム (人間主義)」は、それ自体が時代後れになってしまうだけでなく、人間 (ホモ・サピエンス) そのものの有用性が疑問視されざるをえなくなる。あるいは、すべての人間が、人間を超越した存在 (人工知能) に従属・隷属した存在にすぎなくなるのである。そこで、近代化はもちろん、ホモ・サピエンスの全歴史そのものが終わり、「心」をもたない非生命的存在に支えられたデータ処理システムが、宇宙にまで広がっていくことになる⁷²。

以上が、「近代化の逆説」とでも呼ぶべき、ハラリの分析と予想だが、彼自身はその予想を全面的に受け入れているわけではなく、「21世紀の科学に対する最大の挑戦は、データ教徒のドグマの批判的吟味だろう」というのが彼の意見である。ただし、それが成功するという確信は、現時点での彼にはない。なにしろ、事実在即していないのに優位に立てた思想は、キリスト教や共産主義など、これまでもあった。だから、データ教が科学を支配すれば、それに逆らうのは極めて困難だと悲観的に予想せざるをえないのである⁷³。

結局、ハラリの議論は、「答えはまだ定まってないし、答え自体1つとは限らない」として、「3つの究極の問い」を出すことで終わっている。すなわち、

- ① 生物はアルゴリズムで、生命はデータ処理なのか？
- ② 知能と意識のどちらがより大事か？
- ③ 意識はもたないが高度な知的アルゴリズムが、われわれ以上にわれわれのことをよく知るようになったら、社会や政治や日々の生活はどうなるだろうか？

という問いがそれである。

⁷¹ しかし、欠陥を克服する手段は、人間そのものを改善する手段となった瞬間に困難に直面する。改善の方向をどこに定めればよいのか、それを誰が決めるのかという問題が起きるからである。そのときそれは、一般大衆を置き去りにした少数エリート「不死化」「神人化」のための手段たらざるをえなくなるだろう。

⁷² ハラリは、この意味での「データ教」の預言者がレイ・カーツワイルだとしている。

⁷³ 実際、すでにプロ棋士の理解のレベルを完全に凌駕した「Master」のような「囲碁 AI」は、「黒魔術」のごときものとみなされたり、理解できないままにひたすらその真似をするプロ棋士が続々と現れることで「宗教」になった、といわれたりするようになっている (山本 (2017))。

(4) とりあえずの結論：特化人工知能と平和的に共生する

ハラルの予想を別としても、当面の大きな流れは、人工知能化ないしロボット化が人間の能力の拡張から代替の方向に着実に進んでいることは、否定しがたいだろう。とはいえ、少なくとも特化人工知能のレベルに関するかぎり、それをどのように設計し、どのように利用するかは、そしてそれを取り巻く環境をどのように整備するかは——「ベーシック・インカム」の全面的導入などをも含めて——「人間主義者」としての私たち自身の意思決定次第なのである。

そのかぎりであれば、私たちがいまなすべきことは、自らの「人間性」を失うことなしに——つまり極端な「幼児化」や「依存症」状態に陥ることなしに、あるいは性急に「データ教」の信奉者や「トランスヒューマニスト」となることなしに——今後さらに発展を遂げていくはずの特化人工知能と平和的に共生しうる途を真剣に探り、その実現に努めること以外にないだろう。先にみたシャナハンはその代表的な唱道者だが、日本では、データ・サイエンティストの中西崇文も、その近著『シンギュラリティは怖くない』⁷⁴において、まさにそのような立場とそれが成立しうる可能性とを主張している。あるいは、先に言及した山本一成も、独自の「いい人」理論をもとに、次のように主張する。

人工知能の「知性」、すなわち目的設定能力の獲得は、おそらく、「複数のディープラーニングをつなげたディープラーニング」によって可能になり、このまま技術革新が進めば、少なくとも今世紀の終わりまでには、人工知能が人間から卒業し、“超知能”が誕生するのは、つまり「シンギュラリティ」が到来するのは確定的だが、その過程で人工知能は人間の倫理観・価値観を学習しているはずである。ならばそこに生まれる超知能が人間にとって危険なものとなるかどうかは、人工知能の学習対象となった人間自身がどのような倫理観・価値観の持ち主だったかによる。つまりそれは人間自身の問題なのである。だから、私たちは「インターネット上を含むすべての世界で、できる限り“いい人”でいる」ことが大切なのだ⁷⁵。

4. 汎用人工知能

(1) 汎用人工知能の実現可能性

「特化人工知能」の研究開発の現状と問題点は以上のとおりとして、次に「汎用人工知能」ないし「強い人工知能」の実現可能性と問題点について検討してみよう。検討の結果、

⁷⁴ 中西 (2017)

⁷⁵ 山本 (2017)

先にみた「特化人工知能」の発展に伴い必要とされる対応とは、異なる結論が導かれるはずだ。

ここまでみてきたとおり、現在、「特化人工知能」の研究・開発は急速な進展をみせている。では、人間と同様な「心」をもつとされる「汎用人工知能」が実現する可能性は、現在どのように考えられているのだろうか。

「特化人工知能」をいくつか組み合わせたり、既存の機能を新しい目的を実現するために応用したりすることで、その「汎用性」を高めることは大いにありうることだが、それはあくまでも「程度の問題」であって、その先に、人間と同様な「心」をもつ「強い人工知能」としての「汎用人工知能」がいわば「自然に」生まれてくるとは、私には考えにくい。そのことを私のみたかぎり最も説得的に論じているのが、ジョージ・ザルカダキスの『AIは「心」を持てるのか』である⁷⁶。以下しばらく彼の議論に耳を傾けてみよう。

「人間（ホモ・サピエンス）」を構成する5つの要素

ザルカダキスはいふ。数万年前に突然「新しい心」のビッグバン——神経の偶発的なつながり——を経験した「人間（ホモ・サピエンス）」は、4つの顕著な特徴をもつようになった。すなわち、

- ① 擬人化
- ② ストーリーテリング
- ③ 先天的信念としての心身二元論
- ④ 比喩

がそれである。

私たちは、人間以外の事物をも人間と同様に擬人化して「心」や「言葉」をもつ存在とみなし、それらと交信しようとしたり、人間に似た存在（自動機械、人造人間など）を作ったりしたが⁷⁷。世の中の動きをひと続きの「物語（ストーリー）」として理解、記憶、伝達しようとする。人間（および擬人化された存在）は、「身体（脳）」と「心（魂）」を別々にもつ二元論的な存在であり、身体は死んで土に還っても魂は永遠に生き続けるという「先天的信念」を抱いている。だから「コンピューター」ないし「人工知能」についても、「身体」としての「ハードウェア」と「心」としての「ソフトウェア」とを別々に設計・製作して、後者を前者に「アップロード」しようとする。世界について語る人間の「言葉」は、トーマス・クーンのいう意味での「パラダイム」としての「比喩」ないし「類推」でなりたっている。だから人間は、「事物」、とりわけ「生命」や「脳」を、「泥」や「水（体液）」

⁷⁶ ザルカダキス（2015）

⁷⁷ MITの臨床心理学者で精神分析家のシェリー・タークルは、1980年代のアメリカで、初めてコンピューターに接した小学生たちが、ごく当然のことのように、コンピューターを親しい友達、あるいは「第二の自分」のような、「生き物」だとみなしていたと報告している（タークル（1984））。

や「機械」、「コンピューター」や「インターネット」などになぞらえて理解し語ろうとするのである⁷⁸。グローバル化とITによって高度に絡み合った現代社会でも、人間がもつこれら4つの要素は、私たちの価値、希望、悪夢を形作り続けている。

「汎用人工知能」の実現において問題となるのが、これらの特徴の中でも、デカルトに代表される心身二元論である。この立場に立つかぎり、自分については「我思う、ゆえに我在り」と「主観的」に断言できても、他者の存在、とりわけ他者の心や経験の内容については、確実なことは何もいえない。あくまでも自分自身の心や経験からの類推で、「そう思える」、「そうみえる」にすぎないのである。

だが、人間に備わっている特徴はそれだけではない。ホモ・サピエンスの進化は、「新しい心」のビッグバンに際して、もうひとつの副産物、「第5要素（フィフス・エレメント）」を作り出した。それが、私たちのあり方や私たちが作ってきている世界にとって、信じられないような意味をもっている。この「第5要素」とは、抽象的なこと、絶対的なことを考える能力、なかんずく「自己認識」能力を指す。その意味でそれはほかの4つの要素とは大きく異なる。それは、私たちの意識の中でも最も深く、最も謎に包まれた部分でもある。

⑤ 「自己認識」能力

自分が考えているということを考えること、他人の考えを観察できること、自分自身と他人の内面世界を意識していること、自分の中で自分と対話できることこそ、人間が進化の歴史の限界から抜け出す脱出口なのである。自己認識は、比喩と欺きからの救いの道でもある。そして、「矛盾」や「無限」に耐え、それを乗り越えていく「自由意志」という贈り物を人間に与えてくれたのも、自己認識の力なのである。

心が自分自身を観察できるということは、目が自分自身を見ることができるとか、手が自分自身を持つことができるということと同じくらい非論理的なことではないのか。「自己認識」は、形式論理ではその真偽が決定できない「自己言及命題」——例えば「私は嘘つきだ」とか、「自分自身を要素とする集合が存在する」などといった命題——を生み出す。これが「自己言及のパラドックス」であって、このパラドックスは、論理的な矛盾や無限の自己再帰性を許容しえない形式論理的アルゴリズムに従って動く現在のデジタル・コンピューターでは解決できない。「第5要素」をもつ「主体」としての人間は、この「矛盾」や「無限の再帰性」をやすやすと乗り越えていくことができているのに、である。

結局、ザルカダキスによれば、「心身二元論」と「自己言及のパラドックス」こそが、「特化（弱い）人工知能」を「汎用（強い）人工知能」に進化させる試みの前途をふさいでいる2つの壁なのである。

⁷⁸ あるいは逆に、「機械」や「コンピューター」や「インターネット」を、「生命」や「脳」になぞらえて理解しようとする。「コンピューター」が「電脳」とみなされ、「インターネット」が「集合意識」とみなされるようになる。

「自己認識のパラドックス」は、現在のデジタル・コンピューターのアルゴリズムに依拠しているかぎり解決できない。つまり、人間のような「心」をもつデジタル・コンピューターは作れない⁷⁹。前節でみた「データ教」の中心教義は、この一点で破綻してしまうのである。

汎用人工知能の実現戦略

では、「汎用人工知能」は構築不可能なのか。必ずしもそうではない、とザルカダキスは考える。彼によれば、おそらく唯一の活路は、「唯物論的一元論」、ないしアリストテレス流の「経験主義」に立ち戻り、物質→生命→人間と進化してきたその経路を、人工知能にも歩ませるところにある。言い換えれば、経験主義の次の4つの命題をもとにすることで、「汎用人工知能」は構築可能なはずである。すなわち、

- ① 「魂」や「精神」、神秘的な「形をもたない世界」などは、そのものとしては存在しない。
- ② 存在するのはものだけだ。私たちは物質的な世界に住んでいる。
- ③ 知能は、もがく虫が示す最も単純なものから、洗練された人間の自己認識、意識まで、どれも純粋に物質的、生物学的な現象である。
- ④ 「脳」と呼ばれるものが意識をもてるなら、ほかの素材で作られたほかのものでも意識をもてる。

このような見方を採用するなら、知能、意識の科学的な問題は、天然であれ人工であれ、単純な問いとして再定義できる。すなわち、「数十億の意識をもたないナノロボットは、どのようにすれば意識に到達できるのか」という問いがそれである。

人体の個々の細胞には意識がない。細胞は「ナノロボット」、すなわち、極めて小規模の機械的オートマタだ。神経細胞（ニューロン）も同じだ。従って人工知能の問題は、神経科学の問題と同じく、知能をもたない「部分」から知能をもつ「全体」への遷移を理解することにある。

それさえ理解できれば、「汎用人工知能」あるいはサールのいう意味での「強い人工知能」の構築は、原理的には可能となる。ただし、実際にそれが生まれるまでにどれだけの時間がかかるかは何ともいえないが、20年や30年といった短期間でそれができるとは、到底

⁷⁹ ただし、「自己言及」自体は、ゲーデルの「数化」の手法を応用した「メタプログラミング」によって、一応は可能らしい。ゲーデルは「数」について論ずる「数論」の諸命題それ自体を「数化」することで、それらを「数論」——すなわち自己自身——にも適用する技法を編み出した。コンピューターのプログラミングの場合も、ロジックを直接コーディングするのではなく、あるパターンをもったロジックを生成する高位ロジックによってプログラミングを行う技法が考案されていて、「メタプログラミング」と呼ばれている。他方、数学の「集合論」は、もともとの「素朴集合論」では対処できない「自己言及のパラドックス」を回避する途を探ってきた。「自分自身を含むような集合は存在しない」という「基礎原理」と、「二つの集合AとBが全く同じ要素を持つのならAとBは等しい」という「外延公理」から出発して、それにいくつもの「公理」を追加する形をとっている「ZF公理系」や、それに「選択公理」を追加した「ZFC公理系」は、その代表的なものである。

考えられない^{80 81}。

(2) 汎用人工知能の問題点

しかしながら、人工知能がひとたび「汎用人工知能」のレベルに到達すると、人間はたちまちさまざまな難問に直面せざるをえなくなる。そのいくつかを列挙してみよう。

倫理・道德問題

マレー・シャナハンの指摘するところでは、心をもつロボット、とりわけ外見的にも人間そっくりな「アンドロイド」が人間と共生する形で誕生したとしたら、人間は、深刻な倫理問題に直面する。「その複製を作ったりそれを破壊したりすることは適切といえるだろうか？そもそもそれを作ることも、問題なのではないか。」

実際、未来の人工知能空間の一部に意識が宿っていることは十分ありそうだ。それに対しては、「それは苦しむことができるか」というベンサムの問題が意味をもつ。そういう存在を生み出したものには、それをよく処遇する道徳的責任がある。奴隷のように酷使したり、無慈悲にそのコピーをたくさん作って並列化させたりしてよいものか。扱いようによっては、彼らはサボったり反逆したりするようになるかもしれない。

意識をもつ人工知能、すなわち「人工人間」を危険な存在にしないための戦略はいくつも考えられるが、最も重要なのは、われわれがどのような世界を望んでいるかだ。われわれは未来の人工知能を奴隷にしたいのか、対等な同胞にしたいのか、それとも進化の面でわれわれを置き換える存在にしたいのか？意識をもった人工人間に権利や責任を付与しようとするれば、これまでの人類社会が作り出してきた最も重要な制度——金融制度、法制度、政治制度など——の多くを徹底的に見直さなくてはならなくなる^{82 83}。

⁸⁰ ザルカダキス (2015)

⁸¹ 日本のドワンゴ人工知能研究所 (山川宏所長) が中心になって進めている「汎用 AI」開発の試みは、ここでいう「進化論的」な開発戦略を採用している型のアプローチの 1 例だろう。「全脳アーキテクチャ(WBA)」と呼ばれるこのアプローチは、脳全体ではなく脳の各部位ごとの機能をコンピューターにおおまかに模倣させ、その上でそれらを連携させようとするもので、「ネズミ→サル→人間」と模倣のレベルが進化していくことを期待しているという。グーグル傘下のグーグル・ディープマインド社も似たような形の汎用 AI の開発に取り組んでいて、開発指揮者のシェーン・レグは、この 1 年でネズミのレベルの、10 年以内に人間を超えるレベルの汎用 AI が実現できると語っているという。こうした開発競争の激しさを念頭において、山川所長は、「日本だけが先走って AI 開発の規制を強化すると研究開発そのものにブレーキを踏みかねない」という懸念を表明している。

<http://itpro.nikkeibp.co.jp/atcl/column/17/051800199/052500007/?rt=nocnt>

しかしそのような「進化」が今後 10 年そこそこで実現するとは、私にはとうてい信じられない。

⁸² Shanahan (2015)

人工知能と倫理を巡る議論や制度的対応の試みの現状については、『日経コンピュータ』の特集が有用である。

<http://itpro.nikkeibp.co.jp/atcl/column/17/051800199/?rt=nocnt>

人工知能学会倫理委員会は、人間だけでなく人工知能にも人間と同様な倫理の遵守を要求する「倫理指針」を 2017 年 2 月に発表している。

<http://ai-elsi.org/wp-content/uploads/2017/02/%E4%BA%BA%E5%B7%A5%E7%9F%A5%E8%83%BD%E5%AD%A6%E4%BC%9A%E5%80%AB%E7%90%86%E6%8C%87%E9%87%9D.pdf>

総務省は、2016 年から、倫理問題への言及も含む「AI 開発ガイドライン」の策定に乗り出した。同年末にはその「論点整理」が公表され、意見や提言をまとめた上、2017 年 6 月末をめどに策定が行われることになっている。

http://www.soumu.go.jp/menu_news/s-news/01iicp01_02000054.html

総務省の試みは多くの議論を呼んでいるが、AI ベンチャー企業の側からの規制反対論の紹介としては、次の記事が参考になる。

<http://www.noteware.com/guide2.html>

法律的、制度的な問題

シャナハンはそこから、「人工人間」の出現がもたらす倫理・道徳的問題だけでなく、法律的、制度的な問題についても、あれこれと考えをめぐらせている。例えば、「人間の場合には殺人という罪があるが、人工人間の場合はどうか。そのすべての動作をシャットダウンすることは、それを殺すことだろうか？後から再起動できる場合はどうか。人工人間の動作停止をその意思に反して行うことは罪なのか。それ自身がそれを行うことを許してよいのか。そもそも人間（や人工人間自身）が新たに人工人間を作る権利をどこまで認めるべきか、それに対してどのような規制をかけるべきか。人間は、人工人間を『所有』することを許されるべきだろうか。人工人間に与えられる『市民（国民）権』の問題はどうか。選挙権や被選挙権を付与・制限することは考えられるだろうか。1つの人工人間（の脳）が、いくつもの地域に分かれた複数の体をもっている場合はどうか。1つの人工人間が投票に際しては自分の数千個の複製を作り、投票が終わるとそれらを破壊してしまう場合はどうか」などなど^{84 85}。

（3）人類の「存続危機」

「超知能」の出現

「強い人工知能」がひとたび出現すると、そのコピー——自分自身、あるいは自分の「兄弟」や「子孫」？——をたくさん作ったり、自分自身を加速的に改善するプロセスを急速かつ際限なく繰り返し実行させたりすることが可能になるに違いない。そうなったとすると、そこにたちまち起こるのが統計家のI.J.グッドの予想した「知能爆発」⁸⁶と、その結果としての人間よりも数万倍も数億倍も賢い、「人工超知能（Artificial Superintelligence, ASI）」の出現である。

「人工超知能」——以下、単に「超知能」と呼ぶことにする——は、上にみたような社会的・制度的難問を人類に突きつけるだけでなく、人類自身を「存続危機」に直面させることはまず確実である。

なぜならば、「知能爆発」によって生まれてくる超知能が、人間に対して友好的（フレンドリー）であることの、事前の保証はない。そもそもそのような「知能爆発」過程を人間が事前に設計したり途中で制御したりすること自体、不可能だろう。初期の汎用人工知能にあらかじめ適切な命令ないし目的を与えておけばよいと考えたとしても、それが人間の期待したように解釈、実行される保証はなく、ありとあらゆる珍事や逸脱が起こりうると覚悟せざるをえない。そんなことになれば、人工知能はそれこそ「人類にとって最悪の大惨事」

⁸⁴ Shanahan (2015)

⁸⁵ 同様な議論は、日本棋界の最高峰である羽生善治が、NHKと共同で作成した近著（羽生・NHK（2017））の中でも展開されている。

⁸⁶ Good (1966)

となりかねないのである⁸⁷。

その点についての、*Superintelligence* (超知能)⁸⁸の著者、ニック・ボストロムの、いささか気のめいる考察に耳を傾けてみよう。

「超知能」の制御不能性

人間よりもはるかに高い「知能」をもつ存在とは、いったいどのような存在なのだろうか。どんな「心」や「意識」、「自我」をもち、どんな「動機」や「目標」をもつのだろうか。それは、人間からみて「善意」のもち主なのか「悪意」の塊なのか。それは人間に「友好的」にふるまってくれるだろうか。人間をも含めた宇宙全体を支配しようとするだろうか。

ボストロムは論じたてる。

そもそも現在のわれわれが行っている「賢さ」と「愚かさ」の区別自体が、人間の思考のタイプをもとにしてなされている。これは高度の人工知能にはあてはまらない。人間に関して使われる IQ のレベルとか、チェスの強さなど、およそ適切な指標にはならない。超知能を、特別なオタクか何かのように擬人化して考えることも適切とはいえない。機械的人工知能の精神・欲求と人間の精神・欲求との間の開きは、進化を遂げてきた生物である宇宙人と人間との開きよりも大きいかもしれない。つまり人間の場合から予想されるものとはまるで異なる最終目標をもつ可能性が高い。

他方、超知能が「最終目標」とするものが何であれ、ほとんどいかなる状況や最終目標に対しても有用な、「中間的目標」ないし「手段的目標」が考えられる以上、それらの手段的（中間）目標は互いに似たものになるという有力な見方（手段的収束論）もある。例えば、自らの合理性と知能の増強、それに加えて多様な情報の入手は、有用な手段的価値とされるだろう。生存自体が最終目標ではなくても、最終目標が未来と関わりがある場合には、「自己保存」も中間的目標として選ばれるだろう。そのために必要な物的資源の獲得も、有力な中間的目標となる。そのような目標を達成するために、超知能が、巨大な（地球という惑星のレベルすら超えるような）物理的建設プロジェクトを実現しようとする可能性は、決して小さくない。

仮に、最終目標自体が極めて限定的な内容のもの（例えば、円周率の値をどこまでも計算するとか、砂粒の数をかぞえるなど）だったとしても、その実現をできる限り確実にするための手段的価値の追求範囲、すなわち中間目標は、無限に大きくなる可能性がある。そうすると人間自身までもが、中間目標追求に最適化した形に組織された物的資源、すなわち「コンピュトロニウム (Computronium)」に、変えられてしまうかもしれない⁸⁹。

結局、先行者としての戦略的優位性をもつ機械的超知能が、このような目標に従って活

⁸⁷ バラット (2015)

⁸⁸ Bostrom (2014)

⁸⁹ 与えられた目的をひたすら追求する「ビジーチャイルド」として活動する人工超知能が自己複製をめざしたとすれば、それは自己複製の資源としての地球を食べ尽くす。そうすると今度は、いったん複製をやめて、そのための材料ばかり作り始めるだろう。

動する場合には、そのデフォルトの結果は、人間を含む地球起源のその他の知性体やその文明の、滅亡となりうるのである。

それを防ごうとして、超知能を最初は「箱詰め」にしておき、それが無害で友好的なことが確認されたら外に出すという対応は、裏切られる可能性が高い。敵対的な超知能でも、最初のうち（自分がまだ弱い間）は友好的なふりをするかもしれないからだ。超知能の能力や意図に対する人間による監視や検査の網は、容易にすり抜けられうる。結局すべては人工知能に与えられる最終目標、すなわち「報酬関数」に関わってくるだろうが、人間にとって望ましくないふるまいをしないと保証できるような報酬関数をデザインすることは、極めて難しい。一見適切と思われる報酬関数を命令として与えても、超知能がこちらの期待通りに行動してくれる保証はないからである^{90 91}。

複数の超知能がほぼ同時に生まれた場合、それらは互いに他を支配しようとするだろうか。それとも協調・共存するだろうか。ポストロムの予想では、それは超知能の「離陸」の速さにもよるだろうが、そのスピードが速い場合には、一体（「シングルトン」）だけが独占的に突出する可能性が高い。まだ若い段階の超知能の行動データをもとに、成熟した段階での（とりわけそれが「シングルトン」となった段階での）その行動を予測することは、到底できない。結局のところ、他の障害はいざ知らず、「超知能」、とりわけ「シングルトン」となることに成功した超知能の宇宙支配を阻むような「大障害（フィルター）」など、ないかもしれないのである⁹²。

安全な「超知能」の不可能性

そういう次第で、超知能の危険に対する一見明白に見える安全装置には、すべて欠陥がある。超知能は、自己防衛が報酬関数の最大化につながる限りは、それをためらわないだろう。何が「自己」であるかについての明確な観念を他者から与えられる必要などそもそもないのである。

そこで、機械知能研究所（MIRI）の創設者エリエゼル・ユドカウスキーは、警告する。「AGI（汎用人工知能）は爆弾のタイマーみたいなものだ。デッドラインまでに友好的（フレンドリー）人工知能を作らなくてはならないが、それはもっと難しい。多分、ナノテクノ

⁹⁰ Bostrom (2014)

⁹¹ 例えば、「われわれを笑わせてくれ」と頼むと「顔面筋肉をゆがませて笑い顔のままにさせる」かもしれない。そこで「われわれの顔面筋肉に直接干渉せずに笑わせてくれ」と頼むと、「顔面筋肉を制御している脳の運動中枢部分を刺激して、いつも笑い顔にさせる」かもしれない。あるいは、「われわれを幸せにせよ」と命令したら、「脳の快樂中枢に電極を埋め込む」かもしれない。あるいは人工知能の側からみてよりすぐれた解決として、「われわれの心をコンピューターに“アップロード”して、われわれを恍惚とさせるドラッグのデジタル版を投与し、それが引き起こした快樂の記憶を無限に再生させ続ける」かもしれない。「良心の呵責を感じないように行動せよ」と命令したら、超知能自身の中に組み込まれていた「罪悪感を生み出す認知モジュールの破壊」をはかるかもしれない。

⁹² 経済学者のロビン・ハンソンは、生命の進化の各階梯において次の段階への移行を阻む「大フィルター（The Great Filter）」が存在する、という仮説を提唱している。とりわけ、発達した文明は、その技術があるレベルに達すると、なんらかの「大フィルター」にひっかかって、自らを破壊してしまう可能性がある。それは、核戦争、バイオ技術の誤用、ナノテク事故のいずれかもしれないし、あるいは敵対的人工超知能かもしれない。そうしたリスクの存在を事前知っていたとしても、それを実際に乗り越えられないかぎり、文明はそこで終わってしまうというのである。

<http://mason.gmu.edu/~rhanson/greatfilter.html>

ロジックが世に放たれることを除けば、あらゆる大惨事の中で AGI と肩を並べるものは1つもない。とりわけ恐ろしいのは、各国の多数の組織が競争して、敵の破壊を目的として設計する可能性が高い「敵対的 (アンフレンドリー) 人工知能」である^{93 94}。

超知能の出現自体が、自己破壊的な結果をもたらす可能性もある。超知能の「シングル・トン」が突出するのではなく、そのレベルが比較的似通った複数の超知能同士が自らの「意思」によって互いに戦い合い、殺し合う可能性がそれである。おそらくその場合には、人類もまたその巻き添えとなって絶滅してしまうだろう。超知能の最初の形が「兵器」としてのそれであった場合、この可能性はさらに大きくなるだろう。

人工知性体 (アーチレクト) 戦争

そして実際に大惨事が起こると予想した研究者もいる。人工知能研究の先駆者で、「人工知性体 (アーチレクト)」という言葉を生み出したオーストラリアの研究者、ヒュー・デ・ガリス⁹⁵がその人である。彼はその予言者のノンフィクション *The Artilect War* (人工知性体戦争)⁹⁶の中で、人工知能への対処の仕方をめぐって、21世紀の半ば、人類は、単なる知的な論議の域を超えた2つの政治的陣営、すなわち、

地球派 (Terrans) : 人工知性体反対・阻止派

宇宙派 (Cosmists) : 人工知性体容認・推進派

に分かれて、勝者なき戦いを戦い、その結果は超大量の死者を出す「ギガデス」に終わるだろうと予想した。しかし、とデ・ガリスはいう。「人類はより高度な形の進化を妨げるべきではない。このような機械は神に似ている。それを作り出すのは人類の宿命なのだ」。

5. AI時代の未来と私たちの選択肢

(1) それぞれの局面は「交代」ではなしに「重畳」する

⁹³ バラット (2015)

⁹⁴ 「超知能は、今日存在している、極めて厳格に制御されたどんな兵器やテクノロジーよりも破壊的になりうる。」「AIやAGIの開発者の考えと、彼らが学ぶべきリスクに関する研究との間には、大きな隔りがある。」「一番の問題は、AI開発の道筋にありとあらゆる危険が存在していることを理解している人が、あまりにも少ないことだ。」(バラット (2015))

⁹⁵ 私は (バラット (2015)) を通じてデ・ガリスを知った。興味をもって調べてみると、彼はもともとニューラルネットワークの研究者として頭角を現し、1990年代に京阪奈の国際電気通信基礎技術研究所 (ATR) に招かれて、機械的に「ロボ子猫」を作るプロジェクトに参加している。だがそれには失敗し、日本を去って中国に渡り、中国での脳研究プロジェクトのリーダーとなったようだ。本文の記述は、バラットの著書の他に、アマゾンでのデ・ガリスの本の書評、それからブログの <http://jein.jp/jifs/scientific-topics/867-topic48.html> および http://en.wikipedia.org/wiki/Hugo_de_Garis#Cosmists_and_Terrans などを参照している。

⁹⁶ De Garis (2005)

人工知能についての考察が一通り終わったところで、近代文明の S 字波的展開という視点に立ち戻って、これまでの議論を整理し直してみよう。

先の第 1 章 2. (3) で、現在は、近代の突破大局面の成熟と、成熟大局面の出現が重畳していることを述べた。つまり、20 世紀後半以降の近代文明は、「突破」から「成熟」へと、進化の「大局面」の転換過程に入っている。

ここであらためて確認したいのは、前掲の図表 1-2 に描かれているように、近代を形成するそれぞれの大局面における「成熟」局面は、それに続く新たな大局面の「出現」局面と「重畳」していることである。S 字の山がなだらかに低下する部分と、次の S 字が現れて盛り上がっていく部分が重なる。つまり、旧いものの成熟と新しいものの出現が同時に発生しているという見方が、そこに表現されている。同時に観察される事物のうち、どれが「古い」ものでどれが「新しい」ものなのかを識別することは、必ずしも容易ではないかもしれない。なぜならば、「新しいもの」と「古いもの」は、2 つの異なる事物として対立競合しているという解釈を許しつつ、同時に、同一の事物の中に新旧 2 つの側面が両義的に併存しているという解釈をも許す場合が少なくないからである⁹⁷。そうした事態は多分ごく普通に発生していると思われる。その意味では、新旧がある時点で一気に交代するといった「時代区分」的視点は、社会変化過程の観察に適切なものとはいえない。私たちは、新旧の重畳が常に起こっているという「複眼的」な視点に立って変化を捉えることが、必要かつ有用だろう⁹⁸。

このような意味で、現在は、近代文明の「突破」から「成熟」へと、重畳しつつ転換していく過程にある。同時に、さらに、「近代文明」から「後（ポスト）近代文明」への転換という、文明史的な大転換過程にも入っている。つまり、今日、「小重畳」および「大重畳」として互いに区別することが適切な、二重の重畳が同時並行的に発生しているのである。この状況をもう少し詳細にみてみよう。

（2）近代の「突破の成熟」と「成熟の出現」という小重畳

ここで小重畳とは、近代文明それ自身の「突破の成熟」と「成熟の出現」の重畳を指す。

突破の成熟

近代文明は、20 世紀の後半から、「突破の成熟」が始まった。すなわち、国家化Ⅱと表した 18 世紀後半以来の国民国家化と、産業化Ⅰと表した産業革命以降の産業化が複合する突破大局面が、その成熟期に入っている。

⁹⁷ 例えば、16 世紀前半の英国国教会の成立など。

⁹⁸ さらにいえば、新旧の重畳関係は、構造的な相互依存関係ではないことにも注意すべきだろう。旧いものがなかなか消滅せずに長いこと残り続けたり、新しいものが順調に伸びていくことができずにいったん流産してしまったりすることは珍しくない。

成熟期において、国家化の面では、「国民国家」を構築する権利が広くグローバルに承認される中で、いたるところに国民国家が生まれ、超大国の支配力は相対的に縮小した。産業化の面では、「第3次産業革命」とも呼ばれる産業のデジタル化やネットワーク化の進展に伴って、「産業民主化」とでも呼ぶことがふさわしい中小企業や個人による起業活動がこれまたグローバルに拡大するとともに、既存の寡占的大企業の支配力が相対的に縮小した⁹⁹。

成熟の出現

そこに重畳する形で始まったのが、近代の成熟大局面の出現（つまり「成熟の出現」）である。それは、統合国家化、産業知能化、情報化という3つの流れの複合である。国家化の成熟局面である統合国家化は、まずヨーロッパで始まり、それがアジアに広がり、最後は地球全体を覆うだろう。産業知能化は、まずコンピューターの人工知能化で始まり、それがロボットの高機能化となり、最後は特化人工知能がいたるところで活躍する「データドリブン社会」の到来をもたらすだろう。情報化については、思い切って単純化していえば、「繋（つな）がる化（コミュニケーション、ソーシャル化）」として出現し、「識（し）る化（コグニフィケーション）」として突破し、おそらくは「悟る化」として成熟するだろう。

「繋がる化」は「ソーシャル・ネットワーク」の普及として、「識る化」は「IoT社会」の形成として、「悟る化」は、ロボット工学者の台場時生（筆名）の説く「科学仏教」のような理念に従って人類全体が「涅槃」へ到達するための努力として¹⁰⁰、それぞれイメージできるだろう¹⁰¹。

これらの流れは、近代の「突破の成熟」の流れと複雑に相互作用しつつ、いずれは近代の成熟期がそれ自身の「突破」へと進んでいくものと思われる。それはこれまでと同様、ほぼ100年単位の変化として起こるものと想像されるが、この新たな大局面を特徴づける技術進歩や知識拡大の加速化に照らすと、100年ごとというよりは、その半分とか3分の1といった期間を単位としてめまぐるしく進んでいく可能性もある。

⁹⁹ 「大収斂」（マブバニ（2015）、「権力の終焉」（ナイム（2015））、「ビッグの終焉」（メレ（2014））などという特徴づけは、まさにこの点に注目したものだといえよう。そこには、冷戦の終焉がもたらした「自由（民主・市場）主義」の勝利が、米国のグローバルな覇権の最終的確立と同時に、新たにグローバルに誕生した多数の新興国民国家の「ナショナリズム」追求の自由をもしかし意味したという、いってみれば「より小さな重畳」も発生していた。しかも、「突破の成熟」それ自体、単なる「ビッグの終焉」に終わったわけではなく、既存および新興の国家や企業を支援・利用するための「プラットフォーム」も同時に生まれた。国家化の面での、米国が主導した NATO や IMF の、中国がそれに追随した AIB のようなプラットフォームや、産業化の面での、GAFAM（グーグル、アップル、フェイスブック、アマゾン）や GAFAM（マイクロソフトをも加える）などと呼ばれるようになった、いくつかの「ビガー（メレ（2014））」企業の推進したプラットフォームがそれである。その意味では、20世紀末から21世紀初頭にかけての「グローバリゼーション」は、(1)米国的自由主義体制のグローバリゼーション、(2)国民国家体制（ナショナリズム）のグローバリゼーション、(3)統合国家を目指すグローバリゼーションの三者の「重畳」だとみることができる。

¹⁰⁰ 台場（2016b）

¹⁰¹ 2500年前に釈迦が開いたとされる「仏教」を、情報化時代の今日、科学技術に立脚した「科学仏教」として再構築しようとする台場の議論（台場（2016b））は、「人類」の最終到達点としての「後（ポスト）近代文明」への道を指し示す議論として極めて興味深い。台場によれば、「釈迦仏教」は他の宗教のような「物語」を一切持たず、論理的な考察のみによってこの世界の成り立ちを説明しようと試みている点で、科学技術との親和性が高い。そこで台場は、釈迦仏教を「人類全体が悟りに向かう道を指し示す宗教」として再解釈し、(1)この世界には楽も苦もあり、(2)適度な欲求は善であり、(3)「無明」とは人類全体の無知を、「智慧」とは科学（智）と技術（慧）を指し、(4)「生・老・病・死」の「四苦」は科学技術によって取り去ることができ、(5)その結果として人類は「涅槃」＝幸も不幸もない完全な自由の境地に到達できる、と主張する。

(3) 「近代の成熟」と「後(ポスト)近代の出現」という大重畳

しかし、もう1つの重要な「重畳」にも目を向けないわけにはいかない。それは、近代文明の「成熟」と後(ポスト)近代文明の「出現」との重畳である。

近代文明の成熟

思想・信条の面からみた近代文明は、「人間主義(ヒューマニズム)」のイデオロギーに立脚した文明である。人間主義は、「国家主義」として出現し、「個人主義」として突破し、「データ主義」として成熟しつつある。近代の突破局面での人間主義は、ハラリが指摘しているように、「リベラル・ヒューマニズム」、「社会主義的ヒューマニズム」および「進化論的ヒューマニズム」の3つの分派に分かれて戦い、「リベラル・ヒューマニズム」が勝利した。これに対し、近代の成熟局面での人間主義は、産業知能化(産業化Ⅱ)の影響を受けて、データ主義の方向に向かうことによって、これまでのような狭い意味での「人間中心主義」というか「人間至上主義」を乗り越えていこうとしているが、そこには2つの分派が生まれる可能性がある。

人工知能によって統御される世界と、共生する世界

その1つが、ハラリの懸念するような「データ教」で、万物はデータの大海の中で論理的アルゴリズムに従って生々流転していると考える立場である。もう1つが、人間と人工知能がペアを組んで共生するようになる世界である。

前者の先に待っているのは、高度な(しかし超知能にはいたらない)特化人工知能のプログラムによって、人間を含む万物があらかじめ定められた形で統御される世界である。容易に、人間の「自由意志」や「自律性」の否定につながる。後者の場合は、人間的独自性が維持されるだけでなく、むしろ万物の世界へのその拡張がはかれる。その上で、「縁起」の理法に従う万物の生々流転の世界のあり方があらためて肯定され、受容されることになる。筆者としては、これが望まれる近代文明の成熟のあり方である。

人工知能によって統御される世界と、人間と人工知能がペアを組んで共生するようになる世界と、どちらの分派が優勢になるかは、成熟局面での国家化と産業化と情報化の3つの流れのバランスのあり方にもよるだろう。国家化と産業化の影響力が相対的に強いと、前者の方向に進みがちになりそうだし、情報化の影響力が強ければ、後者の方向に進む可能性が高まるのではないだろうか。

後(ポスト)近代文明

では、それに重畳して出現しつつある後(ポスト)近代文明のあり方は、どのようなものだろうか。ここでもいくつかの可能性が考えられる。

その第1は、ハリリの懸念するような人間の独自性を否定する方向への「データ教」の展開に重畳して、さらに徹底した人間否定型の新文明——いってみれば「すばらしい新世界」の21世紀版——が出現しつつある可能性である。

その第2は、成熟する近代文明が人間主義の普遍化的「止揚」に成功していれば有終の美を飾るのに重畳して、それをさらに発展させる形の後（ポスト）近代文明が宇宙空間に拡大していく可能性。ここでは「悟り」を開いた人間と高度に発達した人間能力拡張型人工知能とが共働して、新しい文明が築かれ、台場のいう「涅槃」がついに実現されることになるだろう。筆者としては、それが実現することを希望し予想したい。

その第3は、人工知能開発が暴走して、なんらかの形の汎用人工知能（超知能）を生み出し、それが地球（と宇宙）の新たな支配者となる可能性。悲観主義者はそこに人類とその築き上げてきた文明の絶滅を予想する。しかし、楽観主義者は、シンギュラリティの到来による「超（トランス）近代文明」の訪れに、人間はおろか生命をも超越した新たな宇宙文明の出現を期待する。

前者の悲観論について、すなわち、汎用人工知能（超知能）により人類とその文明が絶滅する可能性に対して、われわれはどう考えるべきか、続く（4）項で検討する。後者の楽観論については、それがもし実現するとすれば、それは今日の人間中心の近代文明とはあまりにもかけはなれた文明となるので、ことさらに両者を対比する意味などなくなるだろう。しかし、いちおうの検討は、最終章で「補論」として試みることにする。

（4）私たちの選択肢と結論

岐路に立つ人類

汎用人工知能がひとたび出現した暁には、人類とその文明の絶滅という恐るべき危機に直面する可能性が高いとすれば、私たちはどうすべきだろうか。

ここであらためてザルカダキスの警告と提言に耳を傾けてみよう。彼はいう。

単純な自己複製マシンが意識をもつ存在に進化していく可能性があるのなら、ここからどの道を進むべきかを、私たちは、まさにいまの時点で、あらためて考え直さなければならない。意識をもつマシンを生み出すようなすべての人工知能研究に終止符を打つのは、いまでも可能だ。人工知能（あるいはもっと正確には「人工意識」）の研究を永遠に禁止する条約を結んで全地球規模で決定すればよい。そうすれば、私たちの世界と私たちの種が「人工知能特異点」のために消える可能性はほぼなくなる。今後数十年のうちにこのような条約が結ばれることは、十分にありうるだろう。

人間のレベルの人工知能は、理論的には実現可能だが、それは脳の機能を模倣し、サイバネティックスと神経科学から学んだ教訓を計算に入れるというコンピューティ

ングの新しい方向を探究していくならば、という話だ。未熟な段階にあるニューロモーフフィック・コンピューターが正しい方向だ。それに対し、既存のコンピューターテクノロジーに望みをかけ、ムーアの法則によれば 18 カ月ごとにパワーが倍になるから、いずれ心か自我か意識かが自然発生するだろうと思っただけでは絶対にダメだ。

ウィトゲンシュタインは、言葉を使うために言葉の絶対的な定義などなくてもかまわないということを認めた。意味は言葉を使っている中から生まれるので、言葉と無関係に意味を定義することはできない。彼はさらに、意味は言語を使っている人びとの間の社会的な産物だと結論づけた。これは、人工知能研究が発見したことと完璧に共鳴する驚くべき思考だ。情報を処理し、自分で決めた行動をするようにコンピューターをプログラミングすることはできるが、意味を理解するようにプログラミングすることはできない。ウィトゲンシュタインによれば、このような理解の欠如は修復不能である。コンピューターは永遠に、私たちが使っている言葉（つまり、私たちが自然言語を使ってコミュニケーションしているときに使っている言葉）の意味を理解できない。それは、コンピューターが形式言語を使ってプログラミングされているからだ。

進化の歴史の中で、人類の脳は、愛し愛されるように最初から作られている。だとすると、人類との最も重要な絆として愛を考えないような人工物などどうして想像できるだろうか。だから今日の私たちは、知能を備えたマシンが私たちが愛し、利己主義的でなく、必要なら私たちのために死んでくれることを期待する。しかし愛は不確実だし、フロイトが指摘したように、愛と憎しみは同じコインの裏表なのだ。だから逆に、意識をもち自己再生、再組織能力を備えるにいたったマシンが私たちに反抗し、究極の裏切りを働くこと——フランケンシュタイン化——も頭に浮かんでしまう。

「AI 特異点」の到来の物語がそれである。なにしろ神は、人間に神のアルゴリズムを書き換える（オーバーライド）能力、すなわち「自由意志」を与え、人間が神に反逆する可能性を与えたのである。人間もまた、人工知能との関係で、同じ過ちを繰り返すのかもしれない¹⁰²。

ザルカダキスのこの議論に従えば、人類はいま大きな岐路に立たされていることになる。「人工知能」とりわけ「汎用人工知能」に代表される新しいテクノロジーの発展を巧みに制御して、人類の「存続危機」を回避しつつ、人類自身をも乗り越えていくような文明の未曾有の高みに登ることを目指すか、それともその種の危険な技術の開発は禁止することに合意して、近代文明をしてその「有終の美」を飾らせるか、という岐路がそれである。

いずれにせよ人類は、超知能の「シングルトン」によって支配されたり、人工知性体（アーチレクト）戦争が引き起こす「ギガデス」によって絶滅の淵に追いやられたりするような危険を、「宿命」として受け入れるわけにはいかないだろう。しかし、ポストロムの推論や、

¹⁰² ザルカダキス (2015)

ユドカウスキーの警告、あるいはデ・ガリスの予想を、容易には否定しきれないとすれば、そしてそれでもなお、人類とその文明の存続、いや望むらくはそのさらなる発展を期待したいというのなら、私たちには、「強い人工知能」ひいては「超知能」の開発を諦めるという選択肢しか残らないことになる。そしてそれこそまさに、ザルカダキスが真剣な検討の必要を説いている選択肢でもある。

私たちの結論

汎用人工知能についての、人類にとってほとんど唯一といってよいこの選択肢を受け入れるならば、AI時代の未来についての私たちの結論は次のようになるだろう。

1. 特化人工知能について。

その開発は現に急速に進んでいるし、今後も加速的に、そして「宿命的」に進み、数々の便益を私たちにもたらしてくれるだろう。しかし、問題がないわけではない。私たちは、その成果ができるだけ十分かつ均等に享受・配分されるような、また、それがハラリが予想したような暴走をしないような、政策的制度的措置を、ローカルとグローバルの両面において、早急に講ずる必要がある。特化人工知能との平和的な共生に成功すれば、望ましい近代文明の成熟、ひいては人間の「悟り」に伴う新文明（「後（ポスト）近代文明」）の成熟が実現するだろう。

2. 汎用人工知能について。

その開発の成功が実現すれば、結果的に人類の存続危機をもたらす可能性が極めて高い。私たちはそのような危機の回避に全力を尽くす必要がある。その際の有力な選択肢の1つが、いや、ほとんど唯一の選択肢として残るのが、汎用人工知能の開発を中止・禁止するような、グローバルな合意の、これまた早急な確立と実施である。それは人類がシンギュラリティのもたらす「超（トランス）近代文明」への到達を諦めることを意味するが、それにより、人類は、持続可能な「後（ポスト）近代文明」を築き上げることができる。

補論 超知能と超近代文明

(1) さまざまな楽観論者たち

デ・ガリスのいう「人工知性体 (アーチレクト)」ないしポストロムのいう「超知能 (スーパーインテリジェンス)」の出現は、人類にとっての悲惨な結果をもたらすことにはならないと考える人びとは、決して少なくない。そこでいわば本論に対する補論のつもりで、そうした意見のいくつかを最後に検討しておこう。

例えばこれまで度々引用してきたシャナハン¹⁰³は、仮にロビン・ハンソンのいう「大フィルター」をすべて突破しえた文明が、地球外のどこかにすでに存在しているとしても、人類は、少なくともいまのところ、異星人の文明の奴隷になっているわけでもないし、超知能の存続のための資源、すなわち「コンピュトロニウム」になりはててもいないので、希望を抱く余地はまだあるのではないかという。なにしろ、超知能の開発に伴うこの“存続危機”を避けることができさえすれば、未曾有の“存続機会”が入手できるのだから、事前によく考えておくことが必要なのである。このようにいうシャナハン¹⁰³は、いわば「控えめの楽観論」者だといっている¹⁰⁴。

他方、「大フィルター」仮説を提示したロビン・ハンソン本人は、最近 *The Age of Em: Work, Love, and Life when Robots Rule the Earth*¹⁰⁵と題する本を出して、多分今後百年以内に、「脳のエミュレーション」にもとづいて作られた高い知能をもつロボットの「ems (エムズ)」たちが大量に生まれて、地球を支配するようになると予測して、「ems (エムズ)」が主役となる社会の姿と、その中での人間のあり方を描き出してみせている。これはこれで、「皮肉な楽観論」ということができそうだ。

これに対し、「強烈な楽観論者」とみなせる人びとも少なくない。その典型が、いわゆる「(技術的) 特異点 (シンギュラリティ)」の到来を歓迎する人びとである。

とりわけ、この言葉を広く普及させた米国のレイ・カーツワイルは、人類が「特異点 (シンギュラリティ)」に到達する世紀 (21 世紀) は、「近代文明」そのものを乗り越えていくような、人類史的、あるいは地球史からさらには宇宙史的な進化の新段階に人類が突入する最初の世紀になるというビジョンを示している点で、きわだっている。

(2) 「特異点 (シンギュラリティ)」の到来ビジョン

¹⁰³ Shanahan (2015)

¹⁰⁴ 先にみた中西崇文、山本一成、羽生善治なども、このグループに含まれるだろう。

¹⁰⁵ Hanson (2016)

人工知能の研究者、エリエゼル・ユドカウスキーによる整理をもとにすれば、「特異点 (シンギュラリティ)」には次のような三様の見方がある¹⁰⁶。

- ① 変化の加速派：発明家のレイ・カーツワイルの見方。技術進歩が時間と共に指数関数的に加速する「収穫加速の法則」の結果として、人知をはるかに超える技術の時代が遠からず到来する¹⁰⁷。
- ② 事象の地平派：数学者のヴァーナー・ヴィンジの見方。人間の知能を超える存在が技術進歩を起こすようになると、SFでも追い付かないような、人間には理解も予想も不能な奇妙な事象が次々に起こりだす¹⁰⁸。
- ③ 知能爆発派：I.J.グッドやエリエゼル・ユドカウスキーの主張。技術が人知のレベルを大きく超えると自己増殖し始めて、核分裂が臨界点を超えるのに似た、爆発的な知能増大が起こる¹⁰⁹。

マーティン・フォードは、その近著『ロボットの脅威』¹¹⁰で、

「特異点 (シンギュラリティ)」という言葉を最初に、テクノロジーに主導される未来の事件として用いたのは、通常はコンピューターの先駆者ジョン・フォン・ノイマンだといわれるが、彼は1950年代にこんな発言をしたとされる——「常に加速し続ける進歩をみると……どうも人類の歴史において何か本質的なシンギュラリティ (特異点) が近づきつつあり、それを越えた先ではわれわれが知るような人間生活はもはや持続不可能になるのではないか」

というところから始めて、「特異点 (シンギュラリティ)」論の歴史を簡潔に解説している。それによれば、「特異点 (シンギュラリティ)」が具体的なテーマとなったのは、数学者ヴァーナー・ヴィンジが1993年に論文で取り上げたことに始まる。

ヴァーナー・ヴィンジは、彼のいう「特異点 (シンギュラリティ)」のいくつかの源泉として、

¹⁰⁶ <https://intelligence.org/2007/09/30/three-major-singularity-schools/>

¹⁰⁷ カーツワイルの、*The Singularity Is Near: When Humans Transcend Biology* (シンギュラリティは近い：人間が生物性 (バイオロジー) を超えるとき) と題する原著は、2005年の出版である。邦訳 (カーツワイル (2007)) は当初、『ポスト・ヒューマン誕生：コンピュータが人類の知能を超えるとき』という題で出版されたが、その後『シンギュラリティは近い：人類が生命を超越するとき』という表題に変更した「エッセンス版」が出版されている。

¹⁰⁸ 1993年のNASAでの講演 *The Coming Technological Singularity* に基づく論文。

<http://mindstalk.net/vinge/vinge-sing.html>

その「要約」は、「30年以内にわれわれは超人的な知能を創り出す技術的手段を手に入れるだろう。そのすぐ後に人類の時代は終わるだろう。そのような進歩は回避可能か？不可能とすれば、人類が存続するように事態を導くことは可能か？この論文ではこうした問題を議論し、いくつかの可能な答え (とさらなる危険) を提示する」としている。

¹⁰⁹ I.J. Good (1966) "Speculations Concerning the First Ultraintelligent Machine", *Advances in Computers*, vol. 6, pp. 31-88

¹¹⁰ フォード (2015b)

- ① スタンドアローンの機械知能
- ② インターネットから生まれる知能
- ③ インターネットとそのユーザーの組み合わせ（デジタル・ガイア）から生まれる知能
- ④ 人間とコンピューターのインターフェースから生まれる知能
- ⑤ 生物科学から遺伝子操作によって生まれる知能

をあげ、このうちの②と③を除く3つでは、そのテクノロジーの開発に人類が一貫して関わるために、もしかしたら知能爆発でなく、管理可能な形で徐々に知能を高められるかもしれないとしている。

（3）レイ・カーツワイルのビジョン

ヴィンジのこの考えを受け入れた上で、カーツワイルは、概略次のように主張している¹¹¹。すなわち、テクノロジーは指数関数的に発展するために、その発展の速度自体が時間とともに増加していく。そのため、「これから数十年のうちに、情報テクノロジーが、人間の知識や技量をすべて包含し、ついには、人間の脳に備わった、パターン認識力や、問題解決能力や、感情や道徳に関わる知能すらも取り込むようになる」。その結果、人間は、宇宙進化の新しい段階（エポック）、すなわち宇宙進化の第5段階に到達する〔図表 補論-1〕。ここでは、テクノロジーと人間の知能が融合して、「生命のあり方（人間の知能も含む）が、人間の築いたテクノロジー（指数関数的に進化する）の基盤に統合される。そして最後の第6段階では、『宇宙が覚醒』し、大幅に拡大された人間の知能（圧倒的に非生物的）が、宇宙のすみずみまで行き渡る」というのである¹¹²。

具体的な見通しとしては、「チューリング・テスト」をクリアする人工知能、つまり人間並みの知能をもつ「強い人工知能」が生まれるのが2029年ごろで、そこから知能のさらに急激な強化が始まり、2045年ごろには人間の知能の何億倍もの知能が生まれる「特異点（シンギュラリティ）」に到達するだろうという。私にはにわかには信じがたいビジョンだが、カーツワイルの信奉者は近年増える一方であり、2008年にはシリコンバレーに「シンギュラリティ・ユニバーシティ」も作られて、入学志願者が押しかけている¹¹³。

カーツワイルのこの壮大な宇宙進化ビジョンを本稿での近代文明の進化ビジョンにあてはめてみると、「近代文明」は彼のいう第4段階にあたるとみなしてよいだろう。そし

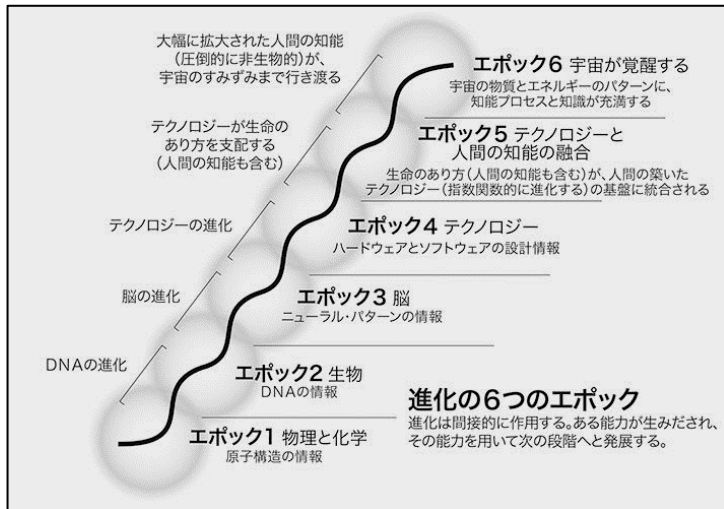
¹¹¹ カーツワイル（2007）

¹¹² カーツワイルのみる宇宙進化は、〔図表 補論-1〕に示されているような、原子・分子の出現に始まり宇宙の覚醒にいたる、6つの段階（エポック）からなる。各エポックは1個のS字波で示され、それらが重畳して宇宙進化の過程を形作るとされている。なお、彼の場合も、私の近代化の「S字波」〔図表 1-1・図表 1-2〕と同様、図の横軸には「時間」が取られている一方で、縦軸になにを指標として取っているかは、とくに具体的には言及されていない。

¹¹³ 特別な校舎や学位の授与などはないが、初年度だけでも、定員40人のところに1200人の応募があったという。「シンギュラリティ・ユニバーシティ」については、ウィキペディアが詳しい。なお、日本にも「日本シンギュラリティ協会」ができている。<http://tokuiten.org/w/>

て、第5段階が、超知能技術を基盤としてそこに人間やその他の生命が統合される「後（ポスト）近代文明」にあたりとみなしてよさそうなので、これを「超知能文明」と呼ぶことにしよう¹¹⁴。

図表 補論-1 レイ・カーツワイルの宇宙進化6段階ビジョン



（出所）レイ・カーツワイル、井上健 監訳、小野木明恵・野中香方子・福田実 訳（2007）『ポスト・ヒューマン誕生：コンピュータが人類の知性を超えるとき』NHK 出版、pp.28

（4）齊藤元章のビジョン

先に紹介した「前特異点（プレシンギュラリティ）」到来の唱道者、齊藤元章は、最近、さらに強烈な楽観論を唱えるようになった。彼は、「脳神経回路ネットワーク」としての「コネクトーム」に注目して、カーツワイルをも乗り越えるような「コネクトーム進化6段階ビジョン」を打ち出したのである。齊藤にいわせれば、最も単純なコネクトーム——すなわち、たった350個の神経細胞と7000個のシナプス結合しかもたない、一番単純な地球上の生物である“線虫”でも、超早期の段階にあるガンをも、わずか1滴の尿から95.8%もの精度、感度で見つけることができる。人間がこれまでに考え出したどんな診断装置よりも、高い診断の感度と精度をもっているのである。この「コネクトーム」が、線虫から人間、人間から地球、地球から宇宙のレベルへとどんどん進化・高度化していけばどうなるのか。「最初は機能分化してない原始コネクトームがあって、次はヒューマンスケールの、人間の脳のコネクトームがあって、すなわちわれわれは第2段階です。次はアーススケールのコネクトームにいたって、ギャラクシースケールのコネクトームにいたって、ユニバーススケールのコネクトームにいたって、最後はスーパーユニバーススケールのコネクトームみた

¹¹⁴ あるいは、もう1つの解釈として、「加速的技術進化」によって「近代文明」自身の成熟局面において、早くも第5段階に突入し、その後の第6段階が、「超知能文明」と呼ぶにふさわしい、「非生物的知能」に立脚した宇宙文明のあり方になるとみることができそうだ。

いなものが考えられる。とすると、まだまだわれわれには進化の余地がたくさんあって、シンギュラリティ (特異点) ごときで右往左往してる場合じゃないよね、となりますよね? ¹¹⁵⁾

ここまでくると、「近代文明」のような観念をもとにした文明論自体が意味を失ってしまうだろう。

(5) 「シンギュラリティ」論への3つの疑問

カーツワイルは、「特異点 (シンギュラリティ)」を、「人間の脳の限界を人間と機械が統合された文明が超越する」点だとイメージしている。つまり、そこでは明らかに機械的な「超知能」が出現しているのだが、それが人間を支配したり絶滅させたりすることはなく、両者は「統合」される。その時期が、2045年ごろに到来するというのが、彼のビジョンだった。

しかし、彼のイメージの中では、その時点での機械の知能が「人知」をはるかに超えていることはたしかだが、つまりなんらかの「超知能」が実現していることはたしかだが、そのレベルがどの程度のものであり、その性格がどのようなものであるかについては、実のところ具体的な説明はない。わずかに [コンピューターによって] 「1年間に創出 (クリエイト) される知能 (インテリジェンス) は、今日の人間のすべて (オールヒューマン) の知能よりも約十億倍も強力 (パワフル) になる」という記述が見当たる程度だが ¹¹⁶⁾、この文章の意味自体、私にはよくくみ取れない。

そこで、あらためて次の3つの疑問をたててみよう。

- ① 「特異点」で「人類と機械が統合された」高度な文明は本当にやってくるのか
- ② やって来るとして、そこでの「超知能」の性格はどのようなものだろうか
- ③ 「超知能」が人類の「敵」になることは本当はないのか

がそれである。以下で検討してみよう。

大沈黙

「フェルミのパラドックス」なるものがあるという。人類よりもはるかに高度な知能と文明をもつ異星人がいるとしたら、人類はすでにそれにめぐりあっていてよいはずである。そこで、高名な物理学者のエンリコ・フェルミは問いかけた。「彼らはどこにいるのか?」、「なぜ人類はこれまでに異星人や異種の知能・文明にめぐりあっていないのか?」と。このパラドックスは「大沈黙 The Great Silence」とも呼ばれ、1968年にはその題名の映画も

¹¹⁵⁾ 齊藤のこの発言は、清水亮との対談中でのものである (清水 (2016))。

¹¹⁶⁾ カーツワイル (2007)

作られている。

「いや、私はめぐりあった」と主張する人びとは、第2次大戦後、突然激増した。いわゆる「空飛ぶ円盤 (UFO)」の目撃報告に加えて、異星人 (スペース・ブラザーズ) との遭遇体験や人類のそれを超える高度な生命体との共同生活体験などが、さまざまに語られ始めたのである^{117 118}。しかし今日にいたるまで、疑問の余地なく確実だと証明されているケースはない (とされている)。

「大沈黙」の理由としては、先にふれたロビン・ハンソンの「大フィルター」仮説がもっともらしい。そうだとすれば、カーツワイルのというような技術的「特異点 (シンギュラリティ)」など、そもそも人類には到達できるわけがないことになる。

実際、人工知能の研究者の中にも、「特異点 (シンギュラリティ)」などまったく問題にしない人は多い。例えば情報社会論を専門とする哲学者の大黒岳彦は、「シンギュラリティ論は話題作りの大風呂敷」とする立場から、「それは、AI の一部の主張だけを我田引水的に拡張解釈し、それを盾に取り土台にもして組み上げた砂上の楼閣にすぎない」と切り捨てている¹¹⁹。ケヴィン・ケリーも、近著の中でカーツワイルのというような「強いシンギュラリティ」の可能性には否定的な立場を取っている¹²⁰。私がここまでの議論で多くを依拠してきたマーティン・フォードも、カーツワイル流の「特異点 (シンギュラリティ)」論には強く批判的で、「彼のシンギュラリティについての著作は、技術の加速度的進歩についての根拠が明確で首尾一貫した記述と、恐ろしく思い込みに満ちた、ほとんどばかげているとすらいいたくなるようなものを加えた奇妙な混交物」だと切って捨てている。「技術エリート疑似宗教」だというわけである¹²¹。

超知能にいたる途

それでもとにかく「超知能」の実現を目指すというのなら、それにいたる「途」は——少なくとも5つは考えられる、とニック・ボストロムはいう。それらは先にみたヴァーナー・ヴィンジの見方と大同小異だが、

① 機械的人工知能 (ロボットを含む)

これが現在の主流であって、多分最速の途だと思われる。

② 「全脳エミュレーション」

¹¹⁷ 青砥 (2009)

¹¹⁸ 私は、近代文明の「成熟」の一環として起こる「情報化」は、20世紀後半から出現局面に入ったと考えている。それは、なによりもまず「文化革命」、すなわち人びとの世界観や価値観の変化として始まった。1960年代のカリフォルニアで起こった「緑色革命」(ライク (1971)) ないし「ヒッピー文化」あるいは「対抗文化 (カウンターカルチャー)」の隆盛はその典型だが、50年代の「空飛ぶ円盤 (UFO)」の発見は、その嚆矢ともみることができよう。

¹¹⁹ 大黒 (2015)

¹²⁰ ケリー (2016)

¹²¹ フランスの哲学者ジャン=ガブリエル・ガナシアも、最近訳出された著作『そろそろ人工知能の真実を話そう』の中で、フォード同様、「シンギュラリティ」はまじめな検討にも値しない議論だとしながら、それでもこうしたいかがわしい議論が大々的に喧伝されて信奉者を集めているのは、GAFAM や NATU (ネットフリックス、エアビーアンドビー、テスラ、ウーバー) などの巨大 IT 企業の宣伝戦略のためであって、その背後には、彼らが世界を支配しようという政治的な目的が隠されていると主張している (ガナシア (2017))。

これは、「脳のアップローディング」とも呼ばれている。まず脳を詳細にスキャンし、それをもとにした三次元ニューロン・ネットワークのイメージを得て、十分強力なコンピューター上でのニューロコンピューテーション構造を構築するのである。それには、「バーチャルリアリティー」として構築する可能性も、あるいはロボットとして具体化する可能性もある。必要とされる理論的抽象化と技術的具體化のレベルの間にはトレードオフの関係があり、実際にはまだわからないことが非常に多く、いかなる実際のエミュレーションも行われていないが、今世紀半ばぐらいまでにはかなりの進ちよくが期待できる。

③ 生物としての脳の機能レベルの向上

原理的には品種改良方式でいけるはずだが、政治的・道徳的な障害がある。それに時間がかかりすぎる。時間を短縮する方法もいろいろ考えられてはいるが、決定的なものはない。また、ホモ・サピエンスが生物種として最も高い知能をもった種であるという保証はない。たまたま一番先にこのニッチに入った種だというだけかもしれない。

④ 脳とコンピューターの直結

これは脳の「機能拡張」である。その可能性は証明済みだが、これまたすぐには実現されそうもないし、危険も大きい。直結よりは離してつなげる方がよさそうだ。

⑤ ネットワークと組織体（人間間や人と人工物やロボットなど）の高度化

個体の知能はそれほどではなくてもそれらが組織化されネットワーク化されれば、それとそれを支えるさまざまな技術的手段も考えられる。とりわけ「インターネット」は有望だが、「インターネットの目覚め」が自生しうるという考え方は空想的にすぎる。

の5つである。ポストロムのみるところでは、一番早そうなのは、②や⑤を部分的に利用した①である。③は時間がかかりそうだし、④にはそれほど期待できない¹²²。

しかし、ポストロムのこの整理は、先にみたザルカダキスの言葉を借りるならば、基本的に「心身二元論」にもとづく整理でしかない。超知能が「自己言及のパラドックス」をどう乗り越えるかについての目配りもない。ザルカダキスの批判が有効であるかぎり、これらの経路のいずれかをたどって超知能が実現すると期待するのは難しいだろう。

とはいえ、前章でみたように、それらとは異なる手法としてのザルカダキス的な「経験主義」にもとづくアプローチも十分考えられるので、仮に今世紀中は無理だとしても、そ

¹²² Bostrom (2014)

の先のいつかの時点で、「汎用（強い）人工知能」からさらには「超知能」が誕生する可能性自体は、決して絵空事ではない。ただしその結果が、前章でみたような人類の存続危機をもたらし、人類文明自体の消滅につながる可能性も、もちろん否定できない。否定できないどころか、その可能性の方がはるかに大きいといわざるをえない。私が本論の中で、「汎用人工知能」についてはその研究開発を諦めるべきだと結論したのもそのためである。

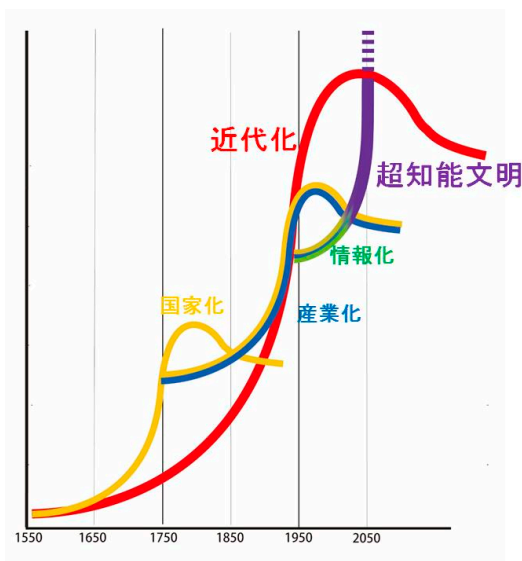
それはそれとして、もしも人類が「友好的」で「安全」な超知能を作り出すことに運よく成功したとしたら、その暁には、「近代文明」はいうまでもなく、「超（トランス）近代文明」、ひいては「人類文明」そのものを超える「超文明」が、この地球上に、さらには地球のスケールをはるかに超えた宇宙空間に、展開されていくことになる。

以上が、私なりに考えた「シンギュラリティ」論への3つの疑問に対する現時点での見解である。

超知能文明のビジョン

このような「超知能文明」とでも呼ぶべき新文明の出現ビジョンを、本稿の最初に示した近代化過程の展開イメージに重ねてみると、[図表 補論-2] のようになるだろう。すなわち、20世紀の後半以降、近代がその「成熟」大局面に入ってしばらくすると、社会変化のペースは急激に加速し始め、「国家化Ⅲ」も「産業化Ⅱ」も「情報化Ⅰ」もひっくりめたすべてが、一体不可分の形で爆発的に急展開して、近代を超える超知能文明への移行が比較的短期間で一気に完了するのである。

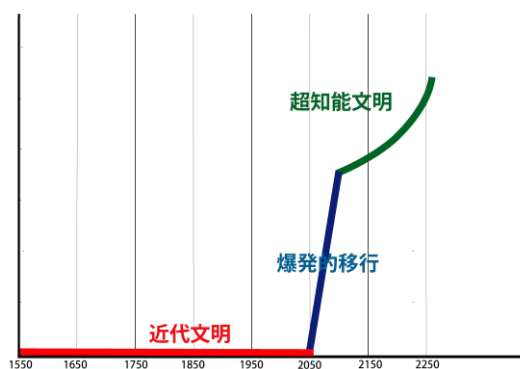
図表 補論-2 超知能文明への爆発的移行ビジョン



(出所) 筆者の案をもとに NIRA 総研作成。

しかし、超知能文明への爆発的移行がひとまず達成された時点——例えば 21 世紀後半のいつかの時点——での新文明のレベルが、近代文明の「何億倍」にもなっているとすれば、そしてそれ以後は変化の加速度は比較的小さくなるものとすれば、新文明と近代文明を同じ図の上で比較したイメージは、[図表 補論-2] よりはむしろ [図表 補論-3] に近いものになるだろう。その場合には、新しい文明の立脚点から過去を回顧してみれば、それ以前のどんな時代よりも急速な発展を続けることに成功した「近代文明」も、それ以前の人類文明とみかけの上ではほとんど区別がつかない、取るに足りないもののようにみえるだろう。

図表 補論-3 超知能文明の側からする近代文明の回顧



(出所) 筆者の案をもとに NIRA 総研作成。

つまり、その場合には、近代化がもたらした人類の文明の飛躍的な発展を評価したり、近代化自体の中での諸局面を区別したり比較したりする試みの全体が、無意味なものになってしまうだろう。前近代の文明というまでもなく、「近代文明」さえ、来るべき超知能文明の、まるで取るに足りない前史にすぎなかったということになる。

しかし、これらのイメージは、私にはあまりにも非現実的にすぎるように思われる。最終章を「補論」としたゆえんである。

参考文献

- 青砥吉隆 (2009) 「科学・技術の時代におけるアメリカの理想像」『ICU 比較文化』 41, pp 1-46.
- 新井紀子 (2010) 『コンピュータが仕事を奪う』 日本経済新聞出版社.
- (2014) 『ロボットは東大に入れるか』 イースト・プレス
- 石黒浩 (2009) 『ロボットとは何か：人の心を映す鏡』 講談社.
- 稲見昌彦 (2016) 『スーパーヒューマン誕生！人類はSFを超える』 NHK 出版.
- 井上明人 (2012) 『ゲーミフィケーション：〈ゲーム〉がビジネスを変える』 NHK 出版.
- 井上智洋 (2016a) 『人工知能と経済の未来：2030年雇用大崩壊』 文春新書.
- (2016b) 『ヘリコプターマネー』 日本経済新聞出版社.
- 海猫沢めろん (2016) 『明日、機械がヒトになる：ルポ最新科学』 講談社.
- 大黒岳彦 (2015) 「人工知能の新次元」『現代思想』 43 (18), pp109-129, 2015-12.
- 小野田博一 (2017) 『人工知能はいかにして強くなるのか？対戦型AIで学ぶ基本のしくみ』 講談社.
- 公文俊平 (2001) 『文明の進化と情報化：IT革命の世界史的意味』 NTT 出版.
- 公文俊平・羽木千晴 (2016) 『AIをどう見るか：“Edge Question”から探るAIイメージ』 NIRA モノグラフシリーズ No.39
- 齊藤元章 (2014) 『エクサスケールの衝撃』 PHP 研究所.
- 清水亮 (2016) 『よくわかる人工知能：最先端の人だけが知っているディープラーニングのひみつ』 KADOKAWA.
- 台場時生 (2016a) 『人工超知能が人類を超える：シンギュラリティ——その先にある未来』 日本実業出版社.
- (2016b) 『科学仏教：菩提樹の下でブッダは何を覚ったのか？』 (Kindle 版)
- 中西崇文 (2017) 『シンギュラリティは怖くない：ちょっと落ちついて人工知能について考えよう』 草思社.
- 羽生善治・NHK スペシャル取材班 (2017) 『人工知能の核心』 NHK 出版.
- 原田泰 (2015) 『ベーシック・インカム：国家は貧困問題を解決できるか』 中央公論社.
- 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/中経出版.
- 水野和夫 (2014) 『資本主義の終焉と歴史の危機』 集英社新書.
- 三宅陽一郎 (2016a) 『人工知能のための哲学塾』 ビー・エヌ・エヌ新社.
- (2016b) 『人工知能の作り方：「おもしろい」ゲームAIはいかにして動くのか』 技術評論社.
- (2017) 『<人工知能>と<人工知性>：環境、身体、知能の関係から解き明かすAI』詩想舎. (Kindle 版 iCardbook)
- 矢野和男 (2014) 『データの見えざる手：ウェアラブルセンサが明かす人間・組織・社会の法則』 草思社.
- 山本一成 (2017) 『人工知能はどのようにして「名人」を超えたのか？：最強の将棋AIポナンザの開発者が教える機械学習・深層学習・強化学習の本質』 ダイヤモンド社.
- 山森亮 (2009) 『ベーシック・インカム入門』 光文社新書.
- アンダーソン, クリス. 関美和 訳 (2012) 『MAKERS: 21世紀の産業革命が始まる』 NHK 出版. (Anderson, Chris (2012) *Makers: The New Industrial Revolution*, Random House.)
- ウルフ, マーティン. 遠藤真美 訳 (2015) 『シフト&ショック：次なる金融危機をいかに防ぐか』 早川書房.
- (Wolf, Martin (2014) *The Shifts and the Shocks: What we've learned - and have still to learn - from the financial*

- crisis, Penguin UK.)
- カーツワイル, レイ. 井上健 監訳 (2007) 『ポスト・ヒューマン誕生: コンピュータが人類の知性を超えるとき』 NHK 出版. (Kurzweil, Ray (2005) *The Singularity Is Near: When Humans Transcend Biology*, Viking Adult.)
- カーネマン, ダニエル. 村井章子 訳 (2012) 『ファスト&スロー: あなたの意思はどのように決まるか? (上・下)』早川書房. (Kahneman, Daniel (2011) *Thinking, Fast & Slow*, Penguin.)
- カストロノヴァ, エドワード. 伊能早苗・山本章子 訳 (2014) 『「仮想通貨」の衝撃』KADOKAWA/中経出版. (Castronova, Edward (2014) *Wildcat Currency: How the Virtual Money Revolution Is Transforming the Economy*, Yale University Press.)
- ガナシア, ジャン=ガブリエル. 伊藤直子他 訳 (2017) 『そろそろ、人工知能の真実を話そう』早川書房 (Ganascia, Jean-Gabriel (2017) *Le Mythe de la Singularité. Faut-il craindre l'intelligence artificielle?* Editions du Seuil.)
- クルーグマン, ポール・浜田宏一 (2016) 『2020年 世界経済の勝者と敗者』講談社.
- ケリー, ケヴィン. 服部桂 訳 (2016) 『〈インターネット〉の次に来るもの: 未来を決める12の法則』NHK 出版. (Kelly, Kevin (2016) *The Inevitable: Understanding the 12 Technological Forces That Will Shape Our Future*, Viking.)
- コーエン, タイラー. 池村千秋 訳 (2011) 『大停滞』NTT 出版. (Cowen, Tyler (2011) *The Great Stagnation: How America Ate All the Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better*, Dutton.)
- 池村千秋 訳 (2014) 『大格差: 機械の知能は仕事と所得をどう変えるか』NTT 出版. (Cowen, Tyler (2013) *Average Is Over: Powering America Beyond the Age of the Great Stagnation*, Dutton.)
- ザルカダキス, ジョージ. 長尾高弘 訳 (2015) 『AIは「心」を持てるのか: 脳に近いアーキテクチャ』日経BP 社. (Zarkadakis, George (2015) *In Our Own Image: Will Artificial Intelligence Save or Destroy Us?* Rider.)
- タークル, シェリー. 西和彦 訳 (1984) 『インティメイト・マシン: コンピュータに心はあるか』講談社. (Turkle, Sherry (1984) *The Second Self: Computers and the Human*, SpiritSimon and Schuster.)
- ターナー, アデア. 高遠裕子 訳 (2016) 『債務、さもなくば悪魔: ヘリコプターマネーは世界を救うか?』日経BP 社. (Turner, Adair (2015) *Between Debt and the Devil: Money, Credit, and Fixing Global Finance*, Princeton Univ Press.)
- ディアマンディス, ピーター・コトラー, スティーブ. 熊谷玲美 訳 (2014) 『楽観主義者の未来予測: テクノロジーの爆発的進化が世界を豊かにする (上・下)』早川書房. (Peter H. Diamandis and Steven Kotler (2012) *Abundance: The Future Is Better Than You Think*, Free Press.)
- ドイッチュ, デイヴィッド. 熊谷玲美・田沢恭子・松井信彦 訳 (2013) 『無限の始まり: ひとはなぜ限りない可能性をもつのか』インターシフト. (Deutsch, David (2011) *The Beginning of Infinity: Explanations That Transform the World*, Viking.)
- ナウム, モイセス. 加藤万里子 訳 (2015) 『権力の終焉』日経BP 社. (Naim, Moises (2013) *The End of Power: From Boardrooms to Battlefields and Churches to States, Why Being In Charge Isn't What It Used to Be*, Basic Books.)
- ハクスリー, オルダス. 大森望 訳 (2017) 『(新訳) すばらしい新世界』講談社 (原著初版は1932年の発行)
- ハラリ, ユヴァル・ノア. 柴田裕之 訳 (2016) 『サピエンス全史: 文明の構造と人類の幸福 (上・下)』河出書房

新社。(Harari, Yuval Noah (2015) *Sapiens: A Brief History of Mankind*, Vintage.)
 パートレット, ジェイミー. 星水裕 訳 (2015) 『闇(ダーク) ネットの住人たち デジタル裏世界の内幕』CCC メ
 ディアハウス。(Bartlett, Jamie (2014) *The Dark Net: Inside the Digital Underworld*, William Heinemann.
 バラット, ジェイムズ. 水谷淳 訳 (2015) 『人工知能: 人類最悪にして最後の発明』ダイヤモンド社。(Barrat,
 James (2013) *Our Final Invention: Artificial Intelligence and the End of the Human Era*, Thomas Dunne Books.)
 パーカー, アンドリュー. 渡辺政隆・今西康子 訳 (2006) 『眼の誕生: カンブリア紀大進化の謎を解く』草思社.
 (Parker, Andrew (2003) *In The Blink of An Eye: How Vision Sparked The Big Bang of Evolution*, Basic Books.)
 ピケティ, トマ. 山形浩生・守岡桜・森本正史 訳 (2014) 『21 世紀の資本』みすず書房。(Piketty, Thomas (2013)
Le capital au XXIème siècle, Seuil.) (Piketty, Thomas (2014) *Capital in the Twenty-First Century*, Belknap Press)
 フォード, マーティン. 秋山勝 訳 (2015a) 『テクノロジーが雇用の 75%を奪う』朝日新聞出版。(Ford, Martin
 (2009) *The Lights in the Tunnel: Automation, Accelerating Technology and the Economy of the Future*, Create Space)
 — 松本剛史 訳 (2015b) 『ロボットの脅威: 人の仕事なくなる日』日本経済新聞社。(Ford, Martin (2015)
Rise of the Robots: Technology and the Threat of a Jobless Future, Basic Books.)
 ブランド, スチュアート. 仙名紀 訳 (2011) 『地球の論点: 現実的な環境主義者のマニフェスト』英治出版.
 (Brand, Stewart (2009) *Whole Earth Discipline: An Ecopragmatist Manifesto*, Viking Adult.)
 ブリニョルフソン, エリック・マカフィー, アンドリュー. 村井章子 訳 (2013) 『機械との競争』日経 BP 社.
 (Brynjolfsson, Erik and McAfee, Andrew (2011) *Race Against the Machine: How the Digital Revolution is
 Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*, Digital
 Frontier Press.)
 — 村井章子 訳 (2015) 『ザ・セカンド・マシン・エイジ』日経 BP 社。(Brynjolfsson, Erik and McAfee, Andrew
 (2014) *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, W. W. Norton
 & Company.)
 ホブズボーム, エリック・J, 安川悦子・水田洋 訳 (1968) 『市民革命と産業革命: 二重革命の時代』岩波書店.
 (Hobsbawm, Eric J. (1962) *The Age of Revolution: 1789-1848*, Weidenfeld & Nicolson.)
 マクゴニガル, ジェイン. 妹尾堅一郎 監修, 藤本徹・藤井清美 訳 (2011) 『幸せな未来は「ゲーム」が創る』早
 川書房。(McGonigal, Jane (2011) *Reality Is Broken: Why Games Make Us Better and How They Can Change the
 World*, Penguin Press.)
 マディソン, アンガス. 政治経済研究所 訳 (2015) 『世界経済史概観 紀元1年~2030年』岩波書店。(Maddison,
 Angus (2007) *The World Economy: A Millennial Perspective/ Historical Statistics*, OECD Publishing)
 マブバニ, キショール. 山本文史 訳 (2015) 『大収斂: 膨張する中産階級が世界を変える』中央公論新社.
 (Mahubani, Kishore (2013) *The Great Convergence: Asia, the West, and the Logic of One World*, PublicAffairs.)
 メレ, ニコ. 遠藤真美 訳 (2014) 『ビッグの終焉: ラディカル・コネクティビティがもたらす未来社会』東洋経
 済新報社。(Mele, Nicco (2013) *The End of Big: How the Internet Makes David the New Goliath*, St Martin's Press.)
 ライク, チャールズ・A. 邦高忠二 訳 (1971) 『緑色革命』早川書房。(Reich, Charles A. (1970) *The Greening
 of America*, Random House.)

- ライシュ, ロバート・B. 雨宮寛・今井章子 訳 (2016)『最後の資本主義』東洋経済新報社. (Reich, Robert B. (2015) *Saving Capitalism: For the Many, Not the Few*, Knopf.)
- リドレー, マット. 大田直子・鍛原多恵子・柴田裕之 訳 (2013)『繁栄: 明日を切り拓くための人類 10 万年史』, 早川書房. (Ridley, Matt (2010) *The Rational Optimist: How Prosperity Evolves*, Harper.)
- リフキン, ジェレミー. 田沢恭子 訳 (2012)『第三次産業革命: 原発後の次代へ、経済・政治・教育をどう変えていくか』インターシフト. (Rifkin, Jeremy (2011) *The Third Industrial Revolution: How Lateral Power Is Transforming Energy, the Economy, and the World*, St. Matins Press.)
- ロドリック, ダニ. 柴山桂太・大川良文 訳 (2013)『グローバリゼーション・パラドクス: 世界経済の未来を決める三つの道』白水社. (Rodrik, Dani. (2011) *The Globalization Paradox: Democracy and the Future of the World Economy*, Oxford University Press.)
- Ackoff, Russell L. and Emery, Fred (1972) *On Purposeful Systems*, Tavistock Publications.
- Arthur, Brian W. (2011) “The second economy,” *McKinsey Quarterly*, October 2011.
(http://www.mckinsey.com/insights/strategy/the_second_economy)
- Bostrom, Nick (2014) *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press.
- Brockman, John (2015) *What to Think About Machines That Think: Today's Leading Thinkers on the Age of Machine Intelligence*, Harper Perennial.
- De Garis, Hugh (2005) *The Artilect War: Cosmists Vs. Terrans*, Etc Pubns.
- Good, Irving John (1966) “Speculations Concerning the First Ultra-intelligent Machine,” *Advances in Computers*. Volume 6, 1966, Pages 31-88
- Gordon, Robert J. (2016) *The Rise and Fall of American Growth: The U.S. Standard of Living Since the Civil War*, Princeton University Press.
- Hanson, Robin (2016) *The Age of Em: Work, Love, and Life when Robots Rule the Earth*, Oxford University Press.
- Harari, Yuval Noah (2017) *Homo Deus: A Brief History of Tomorrow*, Harper.
- Markoff, John (2005) *What the Dormouse Said: How the Sixties Counterculture Shaped the Personal Computer Industry*, Viking Adult. (服部桂 訳 (2007)『パソコン創世「第3の神話」: カウンターカルチャーが育んだ夢』NTT 出版)
- (2015) *Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots*, Ecco. (瀧口範子 訳 (2016)『人工知能は敵か味方か: パートナー、主人、奴隷——人間と機械の関係を定める転換点』日経 BP 社)
- Shanahan, Murray (2015) *The Technological Singularity*, The MIT Press. (ドミニク・チェン監訳 (2016)『シンギュラリティ: 人工知能から超知能へ』エヌティティ出版)

NIRA 総研 情報化の挑戦を受ける日本に関する研究Ⅱ

<メンバー>

研究会委員

公文 俊平 多摩大学教授・情報社会学研究所長

研究会には、足羽教史インクリメントP 株式会社管理部渉外担当部長、鈴木謙介関西学院大学社会学部准教授、および山内康英多摩大学情報社会学研究所教授が参加した。

NIRA 総研

神田 玲子 理事・研究調査部長

榊 麻衣子 研究調査部研究コーディネーター・研究員

<ヒアリングさせていただいた方々>

齊藤 元章 株式会社 PEZY Computing 創業者/代表取締役社長

新保 史生 慶應義塾大学総合政策学部教授



人類文明と人工知能 I
—近代の成熟と新文明の出現—

発行 2017年8月
公益財団法人 NIRA 総合研究開発機構
〒150-6034 東京都渋谷区恵比寿 4-20-3
恵比寿ガーデンプレイスタワー34階
電話 03(5448)1710
ホームページ <http://www.nira.or.jp/>



NIRA 総研とは

NIRA 総合研究開発機構（略称：NIRA 総研）は、

わが国の経済社会の活性化・発展のために

大胆かつタイムリーに政策課題の論点などを提供する

民間の独立した研究機関です。

学者や研究者、専門家のネットワークを活かして、

公正・中立な立場から公益性の高い活動を行い、

わが国の政策論議をいっそう活性化し、政策形成過程に

貢献していくことを目指しています。

研究分野としては、国内の経済社会政策、国際関係、

地域に関する課題をとりあげます。