# シンギュラリティをめぐる不協和音

大東文化大学 経営学部

白井 康之

## 1. はじめに

Ray Kurzweil が，その著書「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」[1]の中で，AI における技術的特異点（シンギュラリティ）に言及したのは，2005 年のことである．シンギュラリティとは，「テクノロジーが急速に変化し，それにより甚大な影響がもたらされ，人間の生活が後戻りできないほどに変容してしまう」ことと定義されている．数理的な特異点本来の意味からすれば，「人類の進化が無限大に見えるような状態に達すること」とも理解できるが，より端的にいってしまえば「コンピュータが人間の知性を超えること」と解釈できる．Kurzweil によれば，このようなシンギュラリティ，すなわち，コンピュータ上に実際の脳の持つ機能が実装される時点は 2045 年とされている．

一般には，シンギュラリティは人類の存在自体を脅かす歴史的時点であるとして，否定的なニュアンスで言及されることも多いが，少なくとも Kurzweil 自身はシンギュラリティを否定的にはとらえていない．それからおよそ 14 年．AI をめぐる研究開発は，そのプレイヤーを徐々に変えつつも着実に進化しつつある．しかしながら，技術者を中心とした AI 開発を推進し，また期待する立場と AI の進展に警笛をならす立場が混在する状況は依然として存在している．本稿では，AI の研究開発の歴史を振り返りつつ，特に両者の視点の相違，また，AI 研究の現状と将来的な課題を整理する．

## 2. AI とシンギュラリティ

### 2.1 AI 研究開発の歴史

AI（人工知能）という言葉は，1956 年のいわゆるダートマス会議（The Dartmouth Summer Research Project on Artificial Intelligence）にて，John McCarthy によって提唱されたものである．この会議では，Allen Newell，Herbert A. Simon による初めての AI プログラムといわれる Logic Theorist が公開され，"Principia Mathematica" に含まれる 38 の定理を自動証明することに成功している．AI における「探索」や「ヒューリスティックス（発見的探索法）」といった現在でもつかわれている重要な概念がここで提示されることになるが，Logic Theorist で組み込まれていた探索手法や探索を効率化するための方法論（ヒューリスティックス）は現在でもなお AI 研究者にとって主要な研究テーマであることは注目に値す

る．これは，Allen Newell らのアプローチがきわめて先進的かつ革新的であったことと同時に，情報を探すこと，またこのために計算を効率化するという方法が知能の実現にあたって今なお本質的な難問であるということを意味する．ダートマス会議以降の AI ブームは第一次 AI ブームともいわれるが，ここでは，「コンピュータによる知性実現の可能性」が提示されたともいえる．

1980 年代には，エキスパートシステムを代表とする産業応用の観点から再び AI が脚光を浴びることとなる．特に，Feigenbaum らによる「第五世代コンピューター 日本の挑戦」（1984）[3] では，1982 年に日本で始まったいわゆる第五世代コンピュータプロジェクトが取り上げられ，自動車産業と同様に，日本が再び AI 開発の分野での世界を席巻するのではないかという危機感が世界に広まった．欧州や米国でも同様のプロジェクトが立ち上がり，また産業界の期待も相俟って，過剰な期待が形成されていた時期でもある．いわゆる第二次 AI ブームは，第五世代コンピュータプロジェクトでも主要なテーマであった論理推論と並列計算が技術的なベースであった．

ただし，前述したように，その主要な技術課題は「探索」と「ヒューリスティックス」であったことには変わりはない．第一次 AI ブームのころに比べれば，ハードウェア・ソフトウェアの進化に伴い，実現できる範囲は格段に広がったとはいえ，逆に実世界が直面する膨大な情報の取り扱いや，組み合わせ的に拡大する計算量の壁を強く認識することとなった．第二次 AI ブームでは，いわば，「コンピュータによる知性実現の限界」が提示されたともいえる．ただし，産業界においては，第五世代コンピュータプロジェクト以降のニューラルネットや遺伝的プログラミング等の応用が広く模索されるようになり，AI がより身近な存在として意識されるようになった契機であったともいえるだろう．

第二次 AI ブームにおける「壁」の認識は，そのまま AI 冬の時代へとつながっていく．現在は，第三次 AI ブームともいわれるが，実のところこの壁は現在でも何ら変わることはなく存在している．したがって，多くの AI 研究者は，第一次，第二次，第三次と歴史的に経てきてはいるものの，この間，本質的かつ革新的なパラダイムシフトは起こっていないと認識しているであろう．

第三次 AI ブームを支えているのは，まず第一に圧倒的な計算資源の熟成である．次章で述べるように，計算機のハードウェアはムーアの法則に代表されるように指数関数的な成長をいまだ示しており，全く同じ手法であっても扱える問題の幅は格段に広がっている．加えて，多層のニューラルネットワークに対する深層学習の有効性が改めて示されたことにより，AI の進化に対する期待が従来以上に高まっているといえる．ただし，これは第二次 AI ブームからの課題である「探索」や「ヒューリスティックス」に対する技術的解答を与えているわけではない．あくまで，計算資源の高性能化と確率的なアルゴリズムにより，近似的に大規模な問題にアプローチできていると理解すべきであって，本質的に知能を実現できるようなパラダイムシフトが起こっているわけではない．

実際，現状の深層学習のアプローチは，画像や音声，自然言語などのパタン認識が主な応

用領域となっているが，人間の深層の知性はこうした確率的なパタン認識だけではとらえられないのは明らかである．第三次 AI ブームは，第二次 AI ブームで提示された「コンピュータによる知性実現の限界」を解消したわけではなく，指数関数的に増大する計算資源を背景に，異なるアプローチに基づき従来は扱うことができなかった応用領域を効率的に実現したものと見るべきである．

### 2.2 シンギュラリティ

Ray Kurzweil が 2005 年の著書「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」[1]の中で AI における技術的特異点（シンギュラリティ）に言及した主な要点は以下のとおりである．

- シンギュラリティとは，テクノロジーが急速に進化することで，人間の生活が後戻りできないほどの変容してしまうような未来のことを指す．
- テクノロジーの進化は，収穫加速の法則に基づく．すなわち，技術が改良される速度は，現時点での技術の達成度に比例するので，技術開発の速度は指数関数的になる．
- 一方，人間の生物学的な進化は極めてスピードが遅いことから，テクノロジーの進化が人間に追いつくのは必然である．
- 人類が扱う情報や知識は，指数関数的に成長している．一方，新しい情報を処理するための人間の帯域幅は非常に限られたものなので，コンピュータなしに人類の得た英知を未来に伝えていくことはもはやできない．
- 特異点以後の世界では，人間と機械，また物理的現実と VR との間には区別が存在しない．しかしだからといって，生物としての知能が終わりを迎えるということではない．
- シンギュラリティについて，多くの人が理解をしないのは，指数関数的な傾向も近視眼的に見れば，線形に見えるからである．
- 2020 年代までにナノボットにより，脳をスキャンすることが可能になる．人間の血球ほどの大きさ（7〜8 ミクロン）の数十億個のナノボットが脳内の毛細血管を駆け巡り，ニューロンの特徴をスキャンできる．これにより感情的知能も含めた人間の複雑な脳に匹敵する，あるいは凌駕する非生物的なシステムが創造されるだろう．
- 非生物的なシステムは人間の知能の拡大に比べ圧倒的なスピードで進化する．この結果として，非生物的な知能は人間の知能を凌駕する．
- いったん強い AI が完成すれば急速に知能が高性能化しはじめ，その現象には抑えがきかなくなる．

上記の根拠として挙げられているのは，主にハードウェア的な進化速度に関する外挿法に基づくものである．Kurzweil は，コンピュータで人間の知性を実現することに関して，

主にハードウェア的な側面から予測を行っている．本書が出版されてから 12 年以上が経過した 2019 年現在，Kurzweil の予想の一部を検証することができる．

まず，人間の情報処理能力は（感覚器官を含めて）$10^{16}$ Flops(10 PFlops）であると（保守的に）推定している．10 PFlops は，Karzweil の予想とほぼ同時期，2012 年には，当時最速のスーパーコンピュータにより達成されており，本書で述べられている 2010 年代初頭に達成するという予測と合致する．ちなみに，2016 年時点の最速のスーパーコンピュータでおよそ 100PFlops[1], 2018 年末時点の最速のスーパーコンピュータでおよそ 200PFlops[2] となっており，これらの進化速度も概ね予想通りである．

ただし，これは単純にコンピュータの情報処理能力の予測であって，これはムーアの法則が現在に至るまで変わらずに成り立つとすれば，いわば必然的な結果である．ムーアの法則には鈍化がみられるという主張もあるが，Kurzweil の論拠からすれば本質的な鈍化ではない．

さて，Kurzweil は，ハードウェアの指数関数的成長を強調し，これをシンギュラリティの主要な根拠としているが，他方で，入力デバイスの進化によるインプット情報の爆発，あるいはこれらの情報の組み合わせによる計算量の指数関数的爆発に正確に言及していないのは不思議なことである．シンギュラリティの実現におけるソフトウェアの観点からの疑問に対して，彼は，ソフトウェアの全体的な複雑性，生産性，効率性などの面で大きな進歩があり，ハードウェアのみならずソフトウェアも進化していると述べているが，どのように控えめにいっても，ハードウェアの進化に比べてソフトウェアの進化は遅い．

例えば，扱う情報量が N 倍になった場合，逐次的な処理であれば，N 倍の計算資源を持てば同じ時間内に取り扱うことができるが，もし組み合わせ的な処理が必要になるのであれは，指数関数的に計算量は増大する．扱う情報量の増大による組み合わせ的な計算量の増大速度はハードウェアの進化をはるかに凌駕する．このような計算量の増大に対して，既存のソフトウェア技術は何ら本質的な解決ができていない．

無論，人間であっても，膨大な組み合わせ計算を脳内で処理しているわけではないが，逆に，近似的な方法やフォーカシングを行うことで，バランスよく処理を行っている．もしコンピュータが人間の処理をまねるというアプローチをとるのであれば，人間の思考方法の解明が必要だが，現在のところ，形式的に模倣できるというレベルではほとんどできていないといってよいし，研究も進んではいない．後述するように，第三次 AI ブームを牽引している深層学習は，そもそも人間の脳をモデルとしたものではない．

また，古典的かつ本質的な AI の難問としてフレーム問題が知られているが，人間が常識的にこなしている処理をコンピュータ内で処理するには大きな壁がある．より正確にいうと，AI で判断するには，知識としての膨大な常識や知識を扱うためのハンドリング技術が

[1] https://www.top500.org/lists/2016/11/

[2] https://www.top500.org/lists/2018/11/

必要であり，また，時間の経過とともにこれらの知識も変わりうる．ハードウェアの計算能力は人間の知能を実現するための必要条件ではありが，決して十分条件ではない．

また，Kurzweil は次のようにも述べている[1]．「人間の脳にリバースエンジニアリングによって人間の知能の並列的で自己組織的なカオス的アルゴリズムを，膨大な能力を持つコンピューティング基盤に適用できるだろう．するとこの知能は今度は自分自身の設計をハードウェアもソフトウェアも含めて改良する立場になり，それが急速に加速しながら繰り返される」．この見込みは，あまりに楽観的と指摘する向きも多い．一般的に困難とされるリバースエンジニアリングは，とりわけ人間の知能に適用するには大変な困難であるし，ソフトウェアの自動生成も，純粋なアルゴリズム的な部分ではそれなりの進化もあるが，外的インタフェースなども含めた部分はほとんど技術的には進化していないからである．

Kurzweil は著名な未来学者であり，また発明家でもあるが，一方でテクノロジー原理主義者であると考えれば理解できるような極端な主張も多い．例えば「将来的には，人体内にさまざまな IT 機器が入り込み，情報を収集し，記憶力を高め，我々の健康や生活をサポートしていく」，「人間とサイボーグの区分は現在でも曖昧なもので過剰に反応する必要はない．むしろ非生物的知能に対する考え方は根本的に変わっていくだろう」といった予測を行っている．さらに，Kurzweil は心を別の媒体にコピーすることで，個々の人間が獲得した知識や能力を半永久的に残す可能性にまで言及している．

## 3. 研究者の立場

以上のような AI 開発の歴史と，近年における AI ブームを踏まえ，研究者の立場からの AI 開発の現状をまとめる．

### 3.1 第三次 AI ブームの現状

1950～1960 年代の第一次 AI ブーム，1980～1990 年代の第二次 AI ブームに続いて，現在は第三次 AI ブームであるといわれている．前述したように，現在の AI ブームのキーファクタは，計算資源の拡大，ネットワーク化が本質的であり，技術的には深層学習が大きなインパクトを与えている．特に，2012 年の画像認識のコンテスト ILSVRC（ImageNet large scale visual recognition challenge）において，トロント大学の Geoffrey E. Hinton らによって発表された深層学習を用いた SuperVision[6]は，前年の優勝記録の誤り率を 10% 以上削減し，画像認識技術に対して大きな衝撃を与えた．また，同年に発表されたスタンフォード大学と Google のグループによる YouTube の画像認識に対する深層学習の応用[11] は，旧来の技術に比べて高い精度を達成するとともに，教師なし学習により自動的に概念を学習するという意味において，また，大規模データを利用することにより，従来のややもすれば Toy Problem と揶揄されてきた機械学習が現実的に応用可能であるとの期待を抱かせ

る結果であった．この研究に先立つ 2009 年には Google の Alon Halevy らが，"The Unreasonable Effectiveness of Data（データの不合理な有効性）"[13]の中で，「いかに乱雑なデータであっても数が多い方が機械学習上は有効に作用する」と主張し，ビッグデータ活用の可能性に言及している．

確かに，大量のデータを使い，また近年の大規模な計算資源を用いることで，旧来の確率的アルゴリズムが大きなブレークスルーを可能にした点は否定できない．実際に上記の SuperVision や Google の画像認識技術は研究者にとっても大きな衝撃であった．しかしながら，確かにこの認識技術は大きな話題となったとはいえ，その技法は人間の処理に比べれば，明らかに本質的な点で作為的である．例えば，画像認識を行う際のフレームの定義方法やフレームの重ね合わせ方など，おそらくは最も重要である認識部分は自動的に機械が学習したものではなく，アプリオリに与えられている．したがって，これらの技術が人間の知性実現に対する端緒となるかどうかは，少なくとも筆者は否定的である．深層学習による画像認識技術の進展が，シンギュラリティの論拠のひとつとなるとは考えられない．

上記のような深層学習によるアプローチは，いわば大量データをもとにした確率的アプローチといえるものだが，一方，第三次 AI ブームでは，こうした確率的アプローチだけではなく，いわば旧来の AI 技術をベースにしたものも非常に多いことにも注意すべきである．

例えば，囲碁では，DeepMind 社によって開発された AlphaGo が，2015 年 10 月に人間のプロトップ棋士であるイ・セドルとの 5 番勝負で 3 勝を挙げた． AlphaGo は強化学習と呼ばれる手法に基づくが，強化学習自体はかなり古い歴史を持つ機械学習の手法である．また，将棋では，2010 年にコンピュータ将棋プログラム「あから 2010」が当時の清水市代女流王将と対戦を行い勝利をおさめたほか，その後の対戦でも，コンピュータ将棋が優勢な結果を残している． 将棋プログラムがベースとする技術は，$\alpha\beta$ 法やモンテカルロ法などの旧来の探索技術がメインである．将棋プログラムの進化は，その多くは，探索のためのヒューリスティックスや計算資源の拡大による性能進化が主たる要因であり，少なくとも人間の知性に接近できるようなレベルで，アルゴリズムの革新的なブレークスルーがあったとはいえない．

以上のように，第三次 AI ブームにおいては，確かに深層学習などの技術的進化はあったものの，その多くは，旧来の技術の線形的な進化に基づくものが多く，したがって，ハードウェアの進化のように指数関数的に性能が向上しているとはいえないのが実情である．

実際，第二次 AI ブームに比べれば，ネットワーク化により大量のデータが取得できるようになったこともあり，また膨大な計算資源が活用できるといった変化はあるものの，人間の知性に近づくような本質的な変化があったとは思えない．ゲームや画像認識など，特定のきわめて限られた場面において性能を向上させることは可能になっているが，依然として人間の日常生活の判断とは圧倒的に次元が異なっていることに留意すべきである．

### 3.2 AI に対する倫理的課題

シンギュラリティの実現の可否は別としても，AI 研究者の中では以下のような懸念も表明されている．

- 自動化された AI の行動による責任は誰がとるのか？
  例えば，自動運転の車が事故を起こした場合の責任の所在はどこにあるのか？また，製造業やサービス業において導入されつつあるロボットやソフトウェアに大きなミスや事故があった場合，誰が責任をとるのか？
- AI は倫理的あるいは政治的に問題のある行動を獲得しないか？
  マイクロソフトの Tay が有名だが，ユーザと会話することで学習するチャットボットにおいては，問題のある思想もしくは危険な思想を植え付けることはさほど難しいことではない． 実際 Tay はネット上に公開されるや否や，差別や陰謀論に染まってしまい，一日で公開停止となっている．また，金融機関における融資審査では，以前には住所や性別（あるいは人種）による差別が公然と行われ来ていたが，こうした偏見のある基準を用いないだろうか？
- AI による社会的格差拡大
  AI により仕事が代替されるようになっても，過去の技術革新と同様失業者が増えることにはつながらない．しかし，仕事の形態は大きく変わってくるはずである．特に，クリエイティブな仕事が増えるのだとすると，能力のある人とそうでない人との格差が拡大する可能性がある．また，富の集中が進む恐れもある．

現状の AI のレベルでは以上のような課題は必ずしも顕在化しているわけではない．ただし，シンギュラリティの可否は別にしても，より高度な AI が今後直面するであろう課題でもあり，AI 開発者としては意識しておく必要がある．

### 3.3 汎用人工知能へのチャンレンジ

前述した囲碁，将棋などのゲームの AI では，特定の場面を学習しているだけで，人間の日常生活の判断とは次元が異なる．現在ではほとんどの AI が特定の目的に特化され，チューニングされたものであり，特化型 AI とも呼ばれる．

一方，汎用人工知能は，さまざまな種類のタスクを都度学習によって実行できるようになるような一般的な仕組みを指す．しばしば誤解されることも多いが，脳をコンピュータ上に実現するという試みでは必ずしもない．とはいえ，汎用人工知能はいまだ実現は困難であるとされている．特に，知の身体性，すなわち，知性とは単に脳で実現されているものではなく，五感等の身体を含めてはじめて実現されているものであり，脳だけで人は考え，行動することはできないという観点からすれば，汎用人工知能への道のりはいまだ遠い．

汎用人工知能開発のモチベーションは，現状では，主に研究者の「技術的挑戦」であると

もいえるが，実際にさまざまなタスクに応用可能であれば，とりわけ有益なシステムであるともいえる．ただ，現実のタスクはまさに多様であり，共通したモデルで実現可能かどうかも厳しい．より未来的な話をすれば，マレー・シャナハンは全脳エミュレーション[12]として，(1) 脳のニューロンを模倣するマッピング，(2) シミュレーション，(3) 身体化の三段階により，脳の機能を人工的に実現するための障壁はないと述べている．確かに物理的な構造そのものは単独では模倣できる可能性もあるが，環境とのインタラクションを含めた系としての実現は極めて困難であるといえるだろう．また，人間の模倣による高度な AI の実現は，著名な物理学者である Roger Penrose が指摘[10]しているように，論理的推論の不完全性や量子理論的な不確定性（ゆらぎ）により，形式的に模倣することは理論的にも不可能である．Penrose の議論は，人間の脳をそのまま実現するのではなく，つまり鳥を作るのではなく飛行機を作るのだと考えれば，その論拠は揺らぐ．しかし，いずれにしても，現在のコンピュータがベースとする論理演算によって，人間が行っているような複雑かつ曖昧な処理を根源的に実現できると考えることには，現時点ではかなりの距離感があるといえるだろう．

人間の知識は完全でもなければ，健全（無矛盾）でもないのは明らかだが，逆にいえば，不完全性や不健全性により，厳密な計算を回避し，必要に応じて妥当な解を導き出すことができる．

### 3.3 コンピュータによる知性実現の根本的課題

Elaine Rich [9]によれば，「AI とは，領域知識を用いて，指数関数的に難しくなる問題を多項式時間で解決していく技術の研究」である．この定義自体はかなり古いものではあるが，筆者はこの定義に深く同意する．仮に計算資源が無限にあり，また計算時間も無限にあるとすれば，フレーム問題として指摘されているようなあらゆる情報を取り込み，あらゆる可能性を考慮し，その時点で最適な解を導き出すことも可能である．しかしながら，そのような仮定は全く現実的ではない．だとすれば，いかにして必要な情報のみを取捨選択して近似的に解くかというのが知性の根源であるともいえる．

より具体的には以下のような技術的課題や問題点が指摘されている．

- AI は自分自身のプログラムを修復できるか？
  数学的には，メタ機能として考えられるこの機能は，形式的に実現することは極めて困難である（また一部は不可能でもある）．機械が自身の想定と異なる事態に遭遇した際に，どのレベルで，どの程度プログラムを修復すればよいのか，またその安定性はどのように担保されるのか，現状では全く解決の糸口すらない．
- AI に人間的な視点が備わるか？
  AI を人間的な視点でとらえるというのは，ユーザイリュージョンと呼ばれる一種の錯

誤であり，むしろ特別な能力を持っている別個の存在として向き合ったほうが AI とのより良い関係性が築けるのではないかとの議論もある．

- AI は原因と結果を解釈できるか？
  一般的に，相関と因果の識別は非常に難しい．相関関係はデータから定量化することができるが，因果関係を把握するためには，膨大な量の背景知識が必要である．現状の AI ではこの違いを理解できないが，実社会における判断として，相関と因果関係の取り違えはしばしば致命的である．

現在の AI の論文では，深層学習に基づくニューラルネットワークにより，うまくパラメータを調整したら「このようなことができた」という結果論的な論文が非常に多い．深層学習もニューラルネットワークも概念そのものは 1980 年代からあったものである．パラメータや構造を工夫することにより，それぞれのタスクにおける精度が向上することはむろん好ましいことではあるが，理論的なモデル化はほとんど進んでいないといってよい．

シンギュラリティという言葉は，コンピュータの知能が人間を超えるときとも解釈されるが，歴史的に見れば，我々はさまざまな技術やツールによるシンギュラリティを経験してきたともいえる．例えば，ユヴァル・ノア・ハラリによるベストセラー「サピエンス全史」では，言語と文字という記録媒体の獲得を「認知革命」であるとし，実はそれ以前には人間は現在とは全く違う時間間隔を生きていたのではないかと指摘している．結果的には言葉を生み出すことにより，虚構を作り出すことができるようになり，人間の思考の抽象度を上げることができた．また，言語の文法構造が人間の世界観を作っているという仮説もある．サピア・ウォーフは言語が思考を規定する（言語的相対論）と考え，テクノロジーによって人間の世界観は大きく変わると考えている．Ray Kurzweil は人類自体はほぼこの数千年ほとんど変わらない一方で，コンピュータの進化が指数関数的なので，いずれコンピュータが人間に追いつくといった議論を展開しているが，実は，周辺技術の進化により人間の思考能力自体もまた大きく変わってきているという見方もできるのである．

## 4. 利用者の懸念

### 4.1 危険な AI の利用

AI の進展に伴い，AI 利用に関する懸念も広がりつつある．ただし，一部の AI に対する拒絶反応は，AI というよりはむしろ，ビッグデータやネットワーク化された社会に対する懸念から生じているともいえる．これらの懸念は，確かに技術の本質を理解しないことによる誤解も多いものの，今後我々が AI やデータ利用する局面で留意しなければならない重要な視点も含まれている．以下では，AI を含め，現在の，また今後進みうる情報化社会に対する懸念をみていきたい．

(1) 選別された情報しか得られないことの弊害

Facebook や Google などのアプリケーションにおけるパーソナライズ機能は，ユーザ自身が意識しない状況で，提供される情報が偏ってしまうことがありうる．Facebook や Google はユーザにどのような情報を提供するかは，いわばビジネス的な観点でとらえ，また最適化されているため，本来接しなければならない情報からむしろ遠ざけられてしまう可能性もある．また，仮にビジネス的観点以外の基準を用いたとしても，いずれにしても情報の偏りは避けることができない．むろん既存の新聞やテレビなどの媒体においても，こうした情報の偏りは重大な懸念であるが，コンピュータネットワーク上のアプリケーションによる情報の偏りはこれらの旧来的なメディアとは比較にならないほどの弊害をもたらす懸念がある．このような状況をパリサーはその著書「閉じこもるインターネット」[15]の中で，「自分自身の情報皮膜の中で知的孤立に陥る」状態と指摘している．

(2) 本来の社会機能を逸脱した AI の利用

AI は一体何のために存在するのか．現在でも，米国の株式市場においては，人間の判断ではなく，AI によるアルゴリズム取引が主流になってきている．ミリ秒単位での自動取引となっているため，もはや人間が直接取引に関与する余地はない．しかし，本来株取引の経済的意味を考えるとき，単なる金儲け以外の重要な役割があるはずである．AI がその目的を「高速に売買を繰り返すことで収益をあげること」と考えているとすれば（実際，そのようになっている），我々はそのような知性を歓迎すべきだろうか？

(3) 人間の精神性の崩壊

知性を持ったコンピュータを全面的に信頼することによって，人間としての自主性や本来持っている精神性が損なわれる可能性がある．ただ，これは実は高度な知能を持ったコンピュータに限定される話ではない．例えば，ELIZA とよばれるシステムは，MIT のジョセフ・ワイゼンバウムによって，1960 年代に開発されたソフトウェアであるが，その機能は単純な構文解析と所定のパタンによる回答を返すだけの今の基準でいえば極めて簡素なものである．しかし，一部のユーザはこの ELIZA に真剣にのめりこみ，自身の深刻な悩みを相談したり，ELIZA の応答を真剣に受け止めたと報告されている．ELIZA はいわゆる人工無脳の起源ともいえるソフトウェアであるが，現在では発話や表情を理解したり，マルチモーダルなインタフェースを備えたより高度なシステムが開発されている．

以上のような弊害はいずれも深刻なものであり，何らコントロールのない状態でコンピュータの知性を高めてしまうことの危険性も示している．

4.2 機械によるモデル化と人間によるモデル化

筆者は以前にクレジットカードのスコアリングシステムの開発に携わったことがある．申込者の属性や借入状況を考慮し，クレジットカードの発行の可否や利用限度額を自動設定するシステムである．計算機上での自動化を行うにあたり，現場の審査担当者からさまざまな観点からのヒアリングを行った際，ある審査担当者から以下のような指摘を受けた．「申込書に押されるハンコの大きさは審査の上で大変重要である．経験上，ハンコが大きい人は返済が滞ることが多いので，ハンコの大きさをコンピュータで自動認識して審査基準に反映してもらえないだろうか？」．ハンコの大きい人の事故率が高いからといって，ハンコの大きい人が不当に扱われる理由は何もないのは明らかである．

では次のような例はどうだろうか？「特定の地域に居住する人を郵便番号で識別し，マイナスの評価を与えたい」．実はこのような基準は旧来の金融機関には公然と存在した審査基準である．例えば，米国のアメックスでは，ある特定の店舗で買い物をした人は支払いを滞納する可能性が高いという事実から，その店舗で買い物をした人のスコアを下げていた．現在では，居住地域や人種により審査に影響を与えることは表向きはなくなっているが，人間が判断を行う際にこのような偏見が存在していることは否定できない．

さらに，以下のような例はどうだろう？「大企業の部長には高い限度額を与え，中小企業の課長以下には限度額を抑える」．このような基準は果たして偏見といえるか，それとも統計的解釈に基づく妥当な判断といえるだろうか．

著者らが開発した自動審査システムでは，明確に偏見と思われるものは説明変数から除外された．しかし，それでも自動化システムは本当に偏見はないといえるのだろうか？自動化システムの説明変数は，年齢，職業や地位，年収，家族構成などの属性情報が含まれている．このうち，年齢，職業や地位，年収，家族構成などを説明変数とすることの意味は何か？これもレベルの違いはあるにせよ，偏見のひとつであるともいえる．

いうまでもなく，統計的解釈と偏見の違いは極めて難しい．実際の申し込み者の素性がわからない以上，ある程度統計的解釈に基づかざるを得ない面がある．ではどこまでが統計的解釈として許され，どこからが偏見になるのだろうか？ このような微妙な判断，また時代によって異なる解釈を機械が果たして吸収できるのだろうか．

当然ながら，機械による評価に不当なバイアスがかかる可能性があるからといって，人間の評価が公平・公正であるわけではない．むしろ，前述のハンコの例のように人間の評価もまた同じように強いバイアスがかかっている．ただし，機械による評価との大きな違いは，人間の評価にはフィードバックがかかりうるということである．人間が評価を誤ったとき，その責任の所在は明確なので，過ちを正すこともできる．しかし，機械の評価にはフィードバックがかかりにくい．

AI に限った話ではないが，効率性やスピードに重きを置くがゆえに，誤った方向にバイアスが効いてしまうようモデルを Cathy O'Neil [2] は「数学破壊兵器」と呼んでいる．「数学破壊兵器」の主な特徴は，不透明である，規模拡大が可能である，有害であるの三点であ

る．一般に，人間の行動や能力，潜在能力をアルゴリズムに落とし込むというのはきわめて困難であるにも関わらず，融資判断，ニュース記事の選別，優良顧客や優良生徒の選別，業務管理，人事評価，健康管理，また政治にまで「設計に不備のある数理モデル」によって遂行されている．一般に，機械による判断は，中身が不透明で責任の所在も曖昧であるという特徴を持つ．

モデルとは本来抽象的なものであり，これ自体は否定されるものではない．しかし，何か新しい発見や新しい解釈があればフィードバックとして反映させ，モデルの健全化が行われなければならない．実際，人間の思考モデルでは，誤った際のフィードバックがあり，これこそが人間社会を進化させてきた原動力でもある．しかし，機械におけるモデルではこのようなフィードバックがききにくい．

Cathy O'Neil [2] の「数学破壊兵器」では，他にも以下のようなケースが挙げられている．

- 大学ランキング
  大学ランキングは，大学の一部の側面のみを取り上げてランキングしたものであるが，その影響力は絶大である．多くの大学では，ランキングをあげるために必要な投資を行うために授業料が高騰している．授業料自体は評価項目ではないので，授業料を高くして施設を充実させるという方向に向かうのは必然である．
- オンライン広告
  個人の行動により選好が暴かれ，また企業側の都合により広告が選別されている．また広告だけでなく，提示されるニュース記事やさらには検索結果までバイアスがかかっている．
- 犯罪予測モデル
  これは良く知られた例だが，アメリカでは地域や時間帯により犯罪の発生を予測し，実際に警察のシステムとして組み込まれている．検挙率は上がるため，効果的に活用されているようにも見えるが，一部の地域においては，不当に逮捕されるなどの弊害も指摘されている．
- 人事評価モデル
  人間が行う従業員雇用や人事評価は多分に偏見に基づいている．このため，機械による人事評価モデルは公平で勘や推測に頼る必要がないとされている．これはサポートツールとしては有用だが，人間にはわからないバイアスが組み込まれていないか注意しなければならない．
- 保険料の算定
  生命保険では，健康状態や嗜好品，食事や生活のパタンなどについて，多くの情報が得られるようになり，より正確な保険料の算定ができるようになった．保険会社の視点では，統計的観点に基づく判断は経営上当然のことではあるが，これにより不当な評価を得ているケースも多々ある．

疑似科学と呼ばれる（方位学，ホメオパシーなどの）分野は，もともと幾分かの根拠はあったにせよ，現在では科学的にはほぼ意味のないである．これらはもともとは何らかの根拠（もしくは事象）があったはずであるが，検証がなかったことが問題である．ビッグデータや AI による判断は，一見すると確かな科学的根拠に基づくものに見えるが，検証が困難であるという意味ではむしろ危険であり，また責任の所在も曖昧である．

人間による意思決定は，欠陥も多いが進化しうるという大きな長所もある．他方，自動化されたシステムは，エンジニアが変更を加えるまで立ち止まったままになる．また，その誤りにも気が付きにくい．

プリンストン大学では，Princeton Web Transparency & Accountability Project として，データの透明性のある正しい管理や説明責任を監視・計測するプロジェクトを立ち上げている[5]. AI やビッグデータを動かすアルゴリズムは決して万能ではなく，むしろ欠陥だらけのまま使われている面も否定できない. 誤った偏見に基づく批判は無用だが, それでも AI の進化を厳しく監視・計測していくことは必須であろう．

## 5. まとめ

ほとんどの AI 研究者は，衝撃的な意味でのシンギュラリティの到来は信じていない．したがって，人間の存在自体が脅かされるのではないかといった過度な反応は不要である．しかし，仮に AI がさらに高度化し，人間に近い仕事をこなすようになったとき，本稿で指摘したようにさまざまな問題が生じうるのも確かである．

アイザック・アシモフは有名なロボット工学三原則（人間に危害を加えない，与えられた命令に服従する，以上 2 点に反するおそれのない限り自己を防衛する）を提唱しているが，これに対して，著名な AI 研究者である Stuart Russell は，以下のような「安全な AI のための三原則」を提示している[4].

- AI は利他的であること
  自己を防衛するのではなく，人間にとって価値あることが最大限に実現できるようにすること．
- AI はおせっかいをしないこと
  人間行動の価値や目的がなんであるか推測しないこと．
- AI は自身の行動の価値や目的を人間の行動からのみから知ること
  何のためにそれをしているのか，理由が明確でない限り真似をしないこと．

加えて，我々ユーザ側の AI リテラシーについても最後に指摘しておきたい．AI を全くのブラックボックスとして畏怖するのではなく，仮に細部に至るまで理解することはなく

とも，どのような原理で判断をしているのか，またその限界は何かを感覚的に理解しておくことが必要である．たとえば，機械学習というプロセスはどのように機能しているのか，また何を学習して何を学習していないのか．それを理解している人と理解していない人で AI を使いこなせる度合いが変わってくる．仮にどのように高度な知性が実現されたとしても，Russell の原則のとおり，コンピュータはあくまで道具である．人間の価値基準を無視して縦横無尽に動き回る AI の姿はそこにはない．

参考文献

[1] レイ・カーツワイル,「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，2007．原著は，Ray Kurzweil, "The Singularity Is Near: When Humans Transcend Biology", Penguin Books, 2006

[2] Cathy O'Neil (著)，久保尚子（翻訳)，「あなたを支配し、社会を破壊する、AI・ビッグデータの罠」，インターシフト，2018

[3] エドワード・ファイゲンバウム，パメラ・マコーダック,「第五世代コンピューター 日本の挑戦」, 阪急コミュニケーションズ, 1983. 原著は, Edward A. Feigenbaum, Pamela McCorduck, "The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World", Macmillan, 1983

[4] 3 principles for creating safer AI | Stuart Russell
https://www.english-video.net/v/ja/2781
https://headlines.yahoo.co.jp/ted?a=20170724-00002781-ted

[5] Princeton Web Transparency & Accountability Project,
https://webtap.princeton.edu

[6] Alex Krizhevsky, Ilya Sutskeve, Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", In Proc. of the 25th International Conference on Neural Information Processing Systems (Vol.1, pp. 1097-1105)

[7] 中島秀之，ドミニク・チェン,”人工知能革命の真実 シンギュラリティの世界”，WAC BUNKO 271，2018

[8] ジョン・L・キャスティ（著)，佐々木光俊／小林傳司／杉山滋郎（訳)，"パラダイムの迷宮 AI・生命の起源・ET・言語…未解決の謎をめぐる科学の法廷"，白揚社，1997

[9] Elaine Rich, Kevin Knight, "Artificial Intelligence", McGraw-Hill College, 1990

[10] Roger Penrose (著)，林一（翻訳)，”皇帝の新しい心―コンピュータ・心・物理法則”，みすず書房（1994/12/20)

[11] Quoc V. Le, et al., "Building High-level Features Using Large Scale Unsupervised Learning", In Proc. of the 29 th ICML, 2012

[12] マレー・シャナハン（著)，ドミニク・チェン（監修，翻訳)，"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，2016
[13] Alon Halevy, et al., "The Unreasonable Effectiveness of Data", IEEE INTELLIGENT SYSTEMS, March/April 2009
[14] AI 白書 2019（独立行政法人 情報処理推進機構）, 2019
[15] Eli Pariser (著), 井口 耕二 (翻訳), "閉じこもるインターネット―グーグル・パーソナライズ・民主主義"，早川書房, 2012