

人工知能 課題番号 1「人工知能の実現可能性について考察せよ」

工学部電子情報工学科 03-175001 浅井明里

2017 年 10 月 17 日

## 1 題材

本レポートでは、ジャン＝ガブリエル ガナシア氏による、『そろそろ人工知能の真実を話そう』（早川書房）について、まず要約、主題、主旨に触れた上で、人工知能の実現性、及び今日多くの著名人が警鐘を鳴らす『シンギュラリティ』の実現性及び危険性についての考察を行いたいと思う。

## 2 本文の要約、主題、主旨

近年、イギリスの宇宙物理学者スティーブン・ホーキング氏やカリフォルニア大学バークレーで人工知能を研究するスチュアート・ラッセル氏等、著名な科学者や様々なスペース X やテスラモーター等、ハイテク企業の共同創業者であるイーロン・マスク氏等の企業家が中心となり、技術の急速な発展により人間はやがて技術を制御することができなくなり、人類に危機的な状況をもたらすと警鐘を鳴らしている。彼らは情報技術の進歩により、機械が自律的に知識を獲得し自ら複製を生み出し改善していくことで、爆発的に知能が発達し、人知を超えた超越的な知性を獲得すると予測している。彼らはこのような現象を、元々数学の分野で図形の形状が突然変化する一点を示す「特異点の理論」より、シンギュラリティと呼びその危険性を唱導している。特に 2010 年代に突入し、ビッグデータによる大量のデータで学習された深層学習技術の発達の結果、人間よりも高精度で画像に映る物体を認識できる AI、世界トップの囲碁棋士に勝利する AI が登場したことにより、こういった「シンギュラリティ論」の信憑性はいよいよ増し、一般にも「AI が人間の仕事を奪うのではないか」「SF のように意思を持った機械がいずれ人間を攻撃するようになるのではないか」と考えている人々が増加しているように思われる。

筆者はこのような「シンギュラリティ」への熱狂及び畏怖に対し、様々な側面からシンギュラリティの実現可能性を否定し、またなぜこれまで情報産業を牽引してきた科学者やテック企業が今になって技術の発展に対して過剰な恐怖を煽るシンギュラリティを積極的に提唱しているのか、その背景を考察している。

まず、筆者はシンギュラリティ論の持つ論理的な矛盾について指摘をしている。最も広く知られたシンギュラリティ論提唱者の一人であるレイ・カーツワイル氏は、トランジスタ数が 18 ヶ月から 24 ヶ月ごとに安定して 2 倍になるムーアの法則を一般化し、テクノロジーの世界でなく進化のプロセスにおいても指数関数的に発展することを表す普遍的な法則であると主張している。ガナシア氏はこの主張に対し、ムーアの法則が永久に続くという仮説の持つ 4 つの問題点を指摘している。第一に、ムーアの法則はそもそも経験則を公式化したものに過ぎず、長期間の観測結果から基づいてある普遍的な原理を導出する機能的推論に則ってムーアの法則を正しいとすることはテクノロジーの進化において斉一性が存在しないために論理的な矛盾を孕んでいるとしている。第 2 の反論の論拠として、カーツワイルの仮説の物理的、技術的な疑問点についても指摘している。ムーアの法則の永久持続性についてはブレーマンにより指摘された情報処理システムの超えることのできない物理的な障壁により否定されることができ、またどんなに小型化しても 10 ナノメートルを下回ることはできないという「シリコン障壁」の技術的な制約も存在する。シリコンに変わる素材としてグラフェンが近年脚光を浴び始めてはいるが、実用性には遠く、技術的な側面よりムーアの法則から外れることは明らかである。第 3 にムーアの法則を一般化する根拠として挙げられている自然の指数関数的な変化についても、筆者はカーツワイル氏の恣意的な時代のまとめ上げによるものであるとし、進化の時代を区切る際に一般的な「種の大量絶滅」に基づいて時間を区分した時、自然の進化が指数関数的だとする法則は否定されたとする。最後に、そもそも知能とは単なる演算能力の向上によって生じるものではなく、コンピュータの演算能力とコンピュータの知能

を再現する能力には直接的な関連性がないと主張している。

次に筆者は、多くのシンギュラリティ論が警戒する、コンピュータの自律可能性現やコンピュータの創造性について否定的な見解を示している。現在、ビッグデータと呼ばれるインターネットに氾濫する大容量のデータをただ蓄積するだけでなく処理することが可能となっており、機械学習やアルゴリズムによりビッグデータから知識を自動的にしかも超高速に取得する技術は顔認証や音声認識など、我々の生活のあらゆる場面で急速に普及しつつある。ガナシア氏は確かにコンピュータが自己学習することで人間の管理が行き届かなくなり、人間がコンピュータの行動を予測しにくくなっている現状を認めているものの、近年目覚ましい成果を挙げている強化学習でさえ、まず訓練の前に人間が報酬の設定という形で選択の基準を与える必要があり、この選択の基準は自発的に生まれるものではないという点から機械が人間の手を離れて自律する、すなわち自分自身でルールを定めそれに従って行動することはしていないとしている。故に、著者は現在の人工知能の開発における技術レベルを鑑みても、コンピュータが人間の力を借りずに際限なく進化を続け暴走し人間を支配するような状況にはなり得ないと主張する。

また、筆者はシンギュラリティ論と中東の古代世界で勢力を持ったグノーシス主義に類似する4つの特徴を挙げた上で、シンギュラリティ論の背景となっている特殊な時間構造に言及している。グノーシス主義とは、グノーシス主義はキリスト紀元1世紀頃の東地中海地域で盛んになった思想運動の一つであり、キリスト教、ユダヤ教に影響を受けつつも真実の神と偽りの神という根本的な二元性の存在を主張し、この世にあらゆる血管が生まれた理由を偽りの神が真実の神を欺いて密かにこの世界を想像したためだとした。ガナシア氏はこのグノーシス主義とシンギュラリティ論は以下の4つの共通する類似点を持つという。第一にあるがままの自然を否定し、自然を変えなければいけないとしている点。第二に寓話、物語と実証に基づく論理的議論を混同して議論を進めている点。第三に精神がその存在を開花させるためには物質世界と完全に離れなくてはならないとしている点。そして第四に日常の時間の流れとの断絶により自らの運命を変えていくとしている点である。特に最後のグノーシス主義及びシンギュラリティ論の共通してもつ特殊な時間構造について、筆者はさらなる考察を行い、またシンギュラリティ論の『未来についての予言』の社会における位置づけについて推測している。

まず筆者は古代世界と近代世界における「未来の予測」を歳費させた上で、シンギュラリティ信奉者の行う「未来予測」が近代の予測における科学的な正当性にかけており、可能性、蓋然性にかかるが一般向けがよく、多くの人が発生すると考えている信憑性のみを持っているに過ぎず、シンギュラリティの発生可能性については限りなく低く、議論に値しないと述べている。また彼らの過度に一般の人々の恐怖心を先導するような主張は、論理的な計算に基づき、未曾有の大事件が突発的に発生する蓋然性を測ろうとする賢明なカタストロフィー論ではなく、多くの人に賛同してもらうことを目的にし、未来に起こるシナリオがすでに書かれていて、そのシナリオこそが未来に起こることを言い当てているとする普通のカタストロフィー論であり、このようなはっきりとした科学的証明のなされていない特殊なシナリオを示すことで他に存在する危険性から人々の目をそらし、その存在を隠蔽していることの問題性を主張している。筆者は、これまで情報技術産業を牽引していたテック企業が率先してシンギュラリティ論を唱導している背景にはテクノロジー産業の民間企業が各々のサービスで取得したビッグデータを元に政治的な影響力を強めようと意図しており、こういった企業がシンギュラリティ論を提唱することで人々の目を逸らそうとしているのではないかと推測している。

### 3 考察

前項で述べたように、著者であるジャン＝ガブリエル・ガナシア氏は近年衆目を集めているシンギュラリティ論に対して否定的な見解を示しており、その理由としてシンギュラリティ論の背景になっているムーアの

法則が永続的に継続するという仮説を論理的、物理学的観点より否定し、また単なる演算能力の向上は機械に知能を与える訳ではないこと、また最新のモデルですら人間が判断基準を与える必要があるため機械が自律的な学習をしている訳ではないという点を指摘している。その上で、現在のシンギュラリティ論は科学的根拠に基づいた正しい未来予測ではなく、可能性及び蓋然性を持たない、もはや信仰とも言えるものであると断じている。またこのように過度に機械の急速な発展による悲観的な未来を啓蒙していくことは、その他に存在する様々な危険性を隠してしまい、我々の視点を誤ったものに導くものだとして否定している。まず、このような主張を踏まえた上で「人工知能の実現可能性」について考察し、また本書に対して感じた感想等を述べていこうと思う。

「人工知能 (Artificial Intelligence)」とは何か。近年、最先端の研究領域だけでなく、一般的な我々の日々使用するアプリケーションにも「AI」が枕言葉的に用いられるようになっており、その定義がやや混沌としてきているように感じる。本レポートでは教師あり学習アルゴリズムや強化学習アルゴリズムを活用して、画像中の物体を検知する、人間の呼びかけに対して正しく応答する等の狭いタスクを行うシステムではなく、機械が周囲の状況や自らの行動を観察し、自らの行動する際の基準と目標を明確にもっと上で判断を下し、行動することが可能なシステムを「人工知能」と呼ぶこととする。2012 年以降、大量のデータで学習させた深層学習モデルが決まったタスクにおいて人間を上回る精度を示す事例が多く見られるようになった。例えば、Li(2009) らによる ImageNet[1] は 2016 年現在、1000 万枚以上のアノテートされた画像データを収集したデータセットであり、2010 年より毎年、ImageNet Large Scale Visual Recognition Challenge (ILSVRC) と呼ばれるコンテストが開催されており、このデータセットを用いて画像認識モデルの精度が競われている。Krizhevsky(2012)[2] らによる Deep Convolutional Neural Network が大きくエラー率を下げた以来、深層学習モデルにより画像認識の精度を大きく向上させ、2015 年には人間のおおよその認識精度である 95% を上回る結果を記録している。画像認識だけでなく、自然言語処理等の異なるタスクでも昨年公開された Rajapukar(2016) らによる Wikipedia データを用いた質問応答タスクの大規模なデータセットである SQuAD[3] に対するコンテストでは、1 年足らずの間に F1 スコアは人間のスコアの 5% 近くまで迫っており、近いうちに人間の精度と並ぶもしくはそれを上回るのではないかとされている。このようなある程度幅の狭いタスクに対しては、最新の深層学習モデル、強化学習モデルは人間よりも優れた能力を示し初めており、これが一部の人々の間で「機械が人間の仕事を奪うのではないか」「機械が人間よりも優れた知性を獲得し、人間を攻撃し始めるのではないか」という不安が広がりつつある要因の一つになっているのではないだろうか。しかし仮にこれらの特定のタスクにおいて大量のデータによってトレーニングされたモデルが人間より高い能力を示したとしても、それはそのまま機械が人間より高い知性・知能を持っているということは上述のように人工知能を定義するならば難しいと思う。ガナシア氏が第 4 章『コンピュータは自律できるか』で言及するように、どんなに高い精度を誇るモデルでも、教師あり学習モデルでは事前に正解の存在する例題が、強化学習では報酬を人間が設定する必要があるため、機械は結局のところ自分自身でルールを生み出す段階にはいたっておらず自律しているということとはできない。現在においても選択の基準を自律的に決定、周囲の状況に応じて変更できるモデルは未だ登場しておらず少なくとも数年以内に「人工知能」が実現される可能性は低いのではないかと考える。この点においては、私は筆者の主張に賛成である。

一方で、Google 等の大企業がシンギュラリティ仮説を衝動する理由としてシンギュラリティという壮大なストーリーの裏側で人工知能技術を用いて顔認証や保安等、本来的に国家が担ってきた役割に取って代わろうとする政治的意図があるとする筆者の主張に対しては賛同ができなかった。それはまずそもそもハイテク企業が積極的に政府に取って代わろうとする意図を示す確固たる根拠は存在せず、また筆者のあげている「ハイテク企業が国家の役割を引き受けている」とする例についても、ハイテク企業がこれまでに集めたユーザーデー

タにより公的な機関より特定のタスクに優れた成果、生産性を持つ可能性を示しているものの、政府がそれらの役割について彼らに実行権限を移譲しない限り、これらのハイテク企業が政府の現在果たしている役割を代わりに担うもしくはそれに類するような状況にはなり得ないと考えるためである。例えば、筆者は身分登録をあげて Facebook に代表されるソーシャルネットワーキングサービスは政府よりも安価に、かつ大量の個人の生年月日、交際状況、結婚歴等の個人情報入手、管理しており、破綻寸前の地方自治体も既にこういったデータを保持している民間組織に管理が委託される可能性を示唆しているが、政府による個人情報の管理は法律に基づいて厳密な登録変更手続きにより行われているのに対し、実名を前提にした Facebook, LinkedIn 等のソーシャルネットワーキングサービスでも個人は簡単に出身地、国籍、経歴、生年月日、時には性別も偽っているケースが多く見受けられる。これらは民間により運営されるソーシャルネットワーキングサービスは個人情報の登録に対して必ずしも事実のみを記載することを求めているわけではなく、またこれらのハイテク企業にとってもユーザーから自由を奪い監視をするコストを払ってまで正しい情報のみをソーシャルネットワーキングサービスのデータに残そうとするメリットは一切ないからだ。こういった現状を鑑みても、ハイテク企業が政府の身分登録の役割を担うことができるかもしれないという主張は俄かに支持し難い。また暗号化の例では FBI がテロリストの 아이폰 の暗号解除要請をアップルが退けた例をあげて「民間企業が国家の権威が民間に傷つけられた」例としてあげているが、インターネットサービスプロバイダが個人の情報の秘密を守ることは当然の義務であり、情報通信の秘匿性が国家、サービスプロバイダによって遵守されなくなれば、個人の表現の自由が保証されないという状況につながりかねず、ハイテク企業が政治に対して影響力を持つようになることで生じる危険性よりも凌駕していると思う。もちろん犯罪、テロの危険性が存在する時どこまで個人の情報の秘匿性が保証されるべきかも現在紛糾している議論ではあるが、少なくともハイテク企業がユーザーの情報の秘密性を堅持したことを「国家の権威への挑戦」と解釈し、ハイテク企業が国家の役割を取って代わろうとしている例の一つとするべきではないと感じる。

シンギュラリティ脅威論に対する私個人の意見としては、人工知能が自律して行動し人間を攻撃し始めるリスクについて懸念をする以前に、機械が自律的に善悪を判断できない段階だからこそ、データを入力し、また得られたデータを解釈する人間側が十分に注意を払わなければ、我々の生活に何らかの悪影響を与えるのではないかと感じている。米国の独立系報道機関である Pro Publica は、米国の司法・警察の場で導入されつつあるコンピュータアルゴリズム (機械学習) による潜在的な犯罪可能性予測が、特定の race, ethnicity に偏った予測を下しており、まるで人間の行うようなバイアスのかかった、人種差別的な判断を下しつつあると警鐘を鳴らしている。ガナシア氏が第 4 章で指摘しているように、機械学習モデルから相関関係が導出されたことは必ずしも入力変数と出力変数の間に因果関係があることを意味はしないが、この結果を解釈する人間が例えば「やはり AI が判断したように黒人は犯罪を行う潜在的なリスクが高い」という結論を導く可能性があり、それは大きな危険性を孕んでいる。特に近年の「人工知能」「AI」に対する一部の、ある種の崇拝に近い態度は、人工知能モデルの返した結果を疑いなく受け入れかねない風潮に繋がりがねず、特にこの危険性を強めているのではないかと感じている。

## 4 感想

本書の、ムーアの法則の永続性を前提とした議論の論理的な問題を指摘し、またシンギュラリティの実現可能性について技術的、物理的課題を列挙していた章については大変興味深く感じた。またグノーシス主義等、古代の宗教思想とシンギュラリティの類似性を指摘している部分についてもこれまでにない斬新性を感じた。一方で、特に後半章ではヨーロッパ的な哲学思想の影響が色濃く出ている部分も多く、哲学的、倫理的、思想

的根拠に基づいてシンギュラリティ論を科学的根拠にかけた信仰に近い未来予測と断じ、批判している部分も見られ、やや説得性を欠いているように感じた。また第8章についてはハイテク企業が国家の役割を担うようになる可能性をいくつかの例を交えて説明しているが、それぞれの例が現実性や妥当性が十分でなく、またそもそもハイテク企業が本当に政治的な支配を目論んでいるのかについての根拠がなく、筆者による憶測で語られているように感じた。また、もし筆者が「この現実を目を背けるべきでない」とするならば、我々は起こりつつあるパラダイムシフトに向けて何をすべきなのかについて具体的な言及がなされていない。Google, Facebook, Amazon 等の大企業がもはや人々の生活インフラとなりつつある彼らのサービスを利用して人々から大量の情報を集め、消費者の個人情報を利用しているとしてこういったサービスを使わないようにと警鐘を鳴らすグループも存在しており、ガナシア氏もこういった主張を暗に意図しているのかもしれないが、著者自身から具体的な施策が提示されれば尚よかったと思う。

## 5 疑問に思った点、理解しがたい点

個人的に疑問に思ったことはいくつかあるが、その中でもっとも大きいものはコンピュータの演算能力向上は本当に人工知能の実現性と関係はないのかということである。ガナシア氏は第3章において、コンピュータの演算能力とコンピュータが知能を再現する能力には直接的な関連性はないとしている。しかし、仮に人間の知能をコンピュータで再現可能な基本的機能に分解することにより再現しようと考えれば、コンピュータの演算能力はコンピュータが知能を再現できるようになるためには必要不可欠ではないかと感じる。例えば Recurrent Neural Network, Convolutional Neural Network 等、近年注目を浴びている深層学習モデル自体は、その実20世紀にすでに考案自体はされていたものの、当時のコンピュータの演算能力の問題や収集できるデータの量に限りがあったために、2010年代に突入するまで日の目を見ることがなかった。カーツワイル(2016)[4]は人間の脳をシミュレートするために必要な計算能力はおよそ  $10^{16}$  cps で十分だとしており、現在のスーパーコンピュータの性能がすでに  $3.6 \times 10^{14}$  cps でありかつムーアの法則が持続することを仮定すれば、人間の脳の機能的な模倣に必要な計算能力は次の10年のうちに達成可能であるとの予想を立てている。私自身は考察の節で示したように、コンピュータの演算能力が人間の脳の処理速度と同程度になることですぐに人間の脳のシミュレートが可能になるとは思わないが、やはり今後「賢い」コンピュータを実現していこうとする中で更なる演算能力の向上は必要不可欠ではないかと考えている。量子コンピューティングへの関心の高まりなどもある昨今、コンピュータの演算能力の急激な向上が今後見込めるのか、また仮にそういった変化が生じた場合、「人工知能」の実現にどう貢献していくのか、引き続き注目を向けていきたい。

## 参考文献

- [1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database.” IEEE Computer Vision and Pattern Recognition (CVPR), 2009.
- [2] A. Krizhevsky, I. Sutskever and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks” Advances in Neural Information Processing Systems 25 (NIPS), 2012.
- [3] P. Rajpurkar, J. Zhang, K. Lopyrev and P. Liang, “SQuAD: 100,000+ Questions for Machine Comprehension of Text” Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016.
- [4] レイ・カーツワイル, “シンギュラリティは近い -人類が生命を超越するとき-” NHK 出版, 2016.

- [5] J. Angwin, J. Larson, S. Mattu and L. Kirchner, “Machine Bias - There’s software used across the country to predict future criminals. And it’s biased against blacks.-” ProPublica, 2016.