

Space Japan Book Review

衛星通信研究者が見た From a satcom researcher point of view

Reviewer: 飯田尚志, 編集顧問 Takashi Iida, Editorial Adviser

ニック・ボストロム, 倉骨彰訳: "スーパーインテリジェンス 超絶 AI と人類の命運", 日本経済新聞社, 2017.

Nick Bostrom: "Superintelligence: Paths, Dangers, Strategies", Oxford Univ. Press, 2014.

本欄でこれまで扱った人工知能 (AI) 関連の書評[1][2][3]を読み返して、私は AI の議論には次の 3 つの論点があるのではないかと考えた。第 1 は AI の能力が人間を越えれば人類が滅亡させられるのではないかという論点、第 2 は経済学の観点から人間の労働が奪われるのではないかという論点、そして、第 3 は技術的研究の観点から AI が人間の能力を超えることはないという論点である。今回取り上げる書籍は、第 1 の論点に立つものと思われるが、その書評[4]によると、AI が本当に危険になり得るシナリオはないのかを徹底的に考えたものということである。ただ、本書の内容に評者自身は賛同したくないという。AI がいかにも人間の不安を煽るような表現が度々出てくると、本書の著者はよく勉強してはいるが、現在の技術を正確に理解しているとも思えない論理展開が頻出するからということである。私は 700 ページを超える本書から書評[4]と同じような観点が読み取れるかどうか確かめたいと思い、本書を読んでみた。しかし、私にはかなり難解の本であったので、私が理解できた範囲のみを以下に記述することになってしまったことをご容赦願いたい。

著者のニック・ボストロム (Nick Bostrom) は、1973 年生まれのスウェーデン人哲学者で、オックスフォード大学哲学科教授、同大学の人類の未来研究所所長、戦略的人工知能研究センター所長である。分析哲学のほかに、物理学、計算論的神経科学、数理論理学の研究を行い、宇宙の構造の理由を人間の存在に求める考え方をとる人間原理に関する業績で知られるという[5]。哲学、数学、芸術などの人間・自然科学分野での優れた研究者に授与されるユージン・R・ギャノン賞を受賞。米フォーリン・ポリシー誌の「世界の頭脳 100 人」に 2 度選出されているという。

本書は、次のような大胆かつ緻密な思考実験である。もし、近未来にスーパーインテリジェンス(超絶知能)が出現することになれば、われわれ人類は滅亡のリスクに直面する可能性がある。スーパーインテリジェンスとは、汎用的な能力においても、思考能力においても、専門的な知識・能力においても、人類の叡智を結集した知力よりも遥かにすぐれたスーパー知能を意味する。私は、それは究極のコンピューターのようなものかと思ったのだが、まず示されたのは、進化が極めて速い人工生物が誕生したとすると瞬く間に人間の頭脳を飛び越えるスーパーインテリジェンスに進化するとしていることに驚いた。このスーパーインテリジェンスは人間の知能を一瞬で飛び越えてしまう。その活動は、地球外宇宙資源を獲得し、宇宙の資源を支配するまでに拡大する。こうなると現在の人類はとても対抗できるものではなく、スーパーインテリジェンスに人類が滅亡させられるという脅威に晒されることになる。

本書では、このようなリスクを回避するためにはコントロール問題を解決する必要があるとしている。コントロール問題とは、集団的かつ総合的な知能において、自分の判断で行動する超絶知能主体に遥かに劣る人類が、いかにすればそのような超絶知能の振る舞いをコントロールできるかということである。これについては非常に広範な議論が行なわれている。ただ、私にはこれ以上はうまく記述できそうもない。詳しくは本書を読んで頂ければと思うが、以下では、本書に書かれている興味深い情報について摘み食的にトピックスとして記述することでその議論の一旦が見えれば幸いである。

- 本書の第 1 章に、ホモ・サピエンスの登場により脳容量が増大し、認知能力の大躍進に繋がったと同様に、知能爆発からスーパーインテリジェンスが実現される時代がやがて到来するのではないかという議論が紹介されている。これを読みながら私は、スーパーインテリジェンスに人類が征服されるというのは全く未経験のことが起きるように考えるけれども、ホモ・サピエンスが現れて抜群の知能の高さを背景に地球を征服して行った歴史を人類は既に経験済みではないかと考えた。
- スーパーインテリジェンスといっても身近には感じられないと思っていたが、「巨大和」というアプローチがあるということである。つまり、人間同士、人間と各種コンピューターや環境センサー、様々なロボット、こうしたものすべてをネットワーク経由で接続し、全体を組織化し、情報を統合、処理することにより達成される「集成的超絶知能」が考えられるというものである。
- スーパーインテリジェンスを人間に役立つようにコントロールするためにソフトウェアで教える必要があるということだが、難しい問題もある。例えば、「快樂と苦痛を量的に勘定し、快樂の期待効用を最大化せよ」という意味のことをコンピューターに実装可能なコードで表すとすると多く

の問題がある。「快樂」とか「苦痛」といった単語が何を意味するかをこと細かに示し、プログラミング言語の表現に翻訳する必要があるからである。しかも、人間の価値観は時代とともに変化するものであるので、解は一定ではないかもしれない。例えば、本書によれば、中世ヨーロッパでは、政治犯が拷問死する様子が鑑賞され、それは品のよい娯楽であったということであるし、猫を火炙りにする猫焼きは、16世紀においてもまだフランスのパリではポピュラーな娯楽であったという。また、奴隷制がアメリカ南部の各地で実践され、それが法律においても完全に支持されていたのは、ほんの150年前の話であるからである。

- スーパーインテリジェンスは自己価値の長期的な累積を最大にするため、スーパーコンピューター施設、宇宙探査ロケット発射施設といった巨大な建造物など地球上のありとあらゆる施設を専有し、AIの目的実現に有用な情報を有していると判断すれば、人間の脳からも情報を吸い取る可能性があるとしている。さらに、ナノアセンブラマシンに接続された駆動装置を持ち、それを自分の手足として自在に動かすことができるスーパーインテリジェンスは、永遠に生き長らえる上でのいかなる障壁をも克服できる能力を有するという。したがって、自己増殖型探査機であるフォン・ノイマン探査機を宇宙に向けて送り出し、宇宙をコロニー化するという終わりのないプロセスを始めることができる。本書で、宇宙との関連での記述例は以上のようなものであるが、ナノアセンブラマシンもフォン・ノイマン探査機もその実現可能性はどのようなのであろうか。

本書で使用されている言葉に私の浅学のため馴染みのないものが多くあった。例えば、「マリグナント」、「ビニン」、「インフラ灌流」、「エソテリック」、「馴化性」、「不可謬」などである。また、衛星通信研究者からみると間違った記述も見られる：「高帯域」は「広帯域」、「低帯域」は「狭帯域」と書くべきであろう。その他、長距離電話に絡む話で「使用帯域幅の関係で長距離電話のほうが安上がり」との記述があったが、その意味は私には良くわからなかった。さらに、「全脳スキャン」という用語も良くわからなかった。

音声認識質疑応答システム「オラクル」が述べられている。スーパーインテリジェンスとなればこれは聞いたことを何でも叶えてくれるアラジンの魔法のランプのようなものであろう。このような話は日本の昔話にも出てくるが、私の孫の保育園児がスマートホンの音声質問システムに「おぼけはいますか」と質問したら、「いますよ。あなたの後ろに」という回答が返ってきたということである。

本書で記述されている議論は実に詳細であらゆる場合の議論を尽くしているようにみえる。ただ、語られている内容は概念的なものが多く具体的でないもので、私には難しくずっと理解できないものが多かったように思う。それに、人間と相対するスーパーインテリジェンスがあたかも人のように記述されていてどうもよく分からなかったし、本書の文章は正直に言って読みにくいものであった。読んでいてもどうもストーンと頭に入らないという意味で。私には、何だか大学1年のときの哲学概論の講義を聞いているような錯覚を覚えた。気晴らしに芥川龍之介「Hankechi」[6]を読んだ。大正5年(1916年)作で旧仮名づかいにも拘わらずさすがにストーンと頭に入った。

参考文献

- [1] 飯田尚志: "Space Japan Book Review -衛星通信研究者が見た ジェイムズ・バラット, 水谷淳訳: "人工知能 人類最悪にして最後の発明", ダイヤモンド社, 2015.", Space Japan Review, No.92, Spring, 2016, <http://satcom.jp/92/spacejapanbookreview1.pdf>
- [2] 飯田尚志: "Space Japan Book Review -衛星通信研究者が見た トーマス・H・ダベンポート, ジュリア・カービー, 山田美明訳, 石崎雅之解説: "AI時代の勝者と敗者 機械に奪われる仕事, 生き残る仕事", 日経BP社, 2016.", Space Japan Review, No.97, Summer, 2017, <http://satcom.jp/97/spacejapanbookreview1j.pdf>
- [3] 飯田尚志: "Space Japan Book Review -衛星通信研究者が見た ジョン・マルコフ, 滝口範子訳: "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, 2016.", Space Japan Review, No.97, Summer, 2017, <http://satcom.jp/97/spacejapanbookreview2j.pdf>
- [4] 松尾 豊: "スーパーインテリジェンス ニック・ポストロム著 AIは本当に「反乱」しないか", 日本経済新聞, Feb.24, 2018, <https://www.nikkei.com/article/DGXXZO27309070T20C18A2MY7000/>
- [5] <https://ja.wikipedia.org/wiki/ニック・ポストロム>
- [6] 芥川龍之介: "hankechi", 43-ruby_547, 1916, <https://www.aozora.gr.jp/cards/000879/card43.html>