

スーパーインテリジェンス---超絶 AI と人類の命運 ニック・ポストロム著

の読後感想 2018-1月



猿と人間の DNA は 1.6%しか変わらないのに大きな差が開いてしまったのはジャレドダイヤモンドが主張するように複雑な言語を使いこなすようになり次世代への文化の継承や人間同士の意思疎通が発達したためであると思う。いずれ近い将来人工知能は、人間と人工知能間の言語はいうに及ばず人工知能同士の言語を生み出し、それに人間が追随していけなくなり、やがて人間と ASI(Artificial Super Intelligence)とか AGI(Artificial General Intelligence)とか定義されるものとの差が開き始め無機質なポストヒューマンが誕生してしまうかもしれません。そこではコンピューターの OS のようにはじめは無数の言語が誕生するかもしれませんが、やがて一つのメジャーな言語に統一されるという経過をたどるでしょう。人類の場合、例えば現在南北朝鮮が分断して 65 年経過していますが、これだけの年数でも言語間の差が生まれ意志が通じにくくなってしまっているようです。

AGI に心・意識が芽生えるかについてはウォルター・J・フリーマンが主張するように全身から入る情報は脳の中を駆け巡るなかで大きな流れが生まれ脳の辺縁系を中心に渦巻のような流れを生み出し、フリーマンはこの脳全体のネットワークを大域的アトラクター(global attractor)と呼びました。これが心の実態であると考えています。竜巻のような自然界に見られるような混沌からの秩序が生成される。すなわち自己組織化が脳活動の本質であるかもしれません。ここまでは無意識の働きで意識ではない。心とは渦巻のように生まれた無意識とその後に感知される意識とで成り立っているのかもしれない。

チンパンジーが人類にコントロールされるように人類がポストヒューマンにコントロールされかねないと心配しなければならない。

動物から進化した人類の暗部に集団虐殺・戦争がある。これは正に現代の問題でもある。

人類対動物、国家対国家、民族対民族、或る宗教対他の宗教というような発想は動物から人類がそのまま受け継いだものであると言われている。AGI は AGI 対人類というような発想が意志と能力を持てば自然と芽生えてしまうものなのか？ AGI(スーパーインテリジェンス)は人類から進化した人工生命体に負の側面を受け継がないような進化をしてほしいものである。しかし A 集団の AGI と B 集団の AGI が縄張り、資源の獲得を動機に徹底して最後まで互いに抹殺し合うようなことが将来起こらないとは限らない。

スーパーインテリジェンスが、高度に進化した文明は論理的に長続きしないと結論づける可能性もあるかもしれない。

人間とは何か？ 国や宗教、性別、年齢によっても様々な解釈がある。スーパーインテリジェンスとは何か？の定義について世界唯一の統一基準なんてできるのだろうか。ある時代にで

きたとしても時代とともに変化してゆくのではないかと思います。

魚(ヒト)は水がきれいすぎても汚すぎても生きられない。スーパーインテリジェンスも綺麗ごとばかり並びたてる学者とか政治家だけで作ってほしくないと思う。

数学は言語の一種ですが、いわゆる知の限界を証明した Kurt Gödel の不完全性定理というものがありますがこの不完全性定理は数学のみならず理論体系一般すべて(哲学、科学、法律等)に適用可能であると言われていています。どんな理論体系にも、証明不可能な命題(パラドックス)が必ず存在する。それは、その理論体系に矛盾がないことをその理論体系の中で決して証明できないということであり、つまりそれ自身で完結する理論体系は構造的にあり得ないということです。ゲーデルの不完全性定理により数学に一定の限界があることが示されました。全ての物事は自己の内部に矛盾を含んでおり、それによって必然的に自己と対立するものを生み出すこととなります。スーパーインテリジェンスはこの定理を自ら生み出していくことが出来るのであろうか。

この書籍は4年程前に原書が出版されたものであり、内容は哲学的、倫理的、ガバナンスなど多岐にわたり AI の可能性について書かれていて、個々のケースの AI の進展状態の可能性について、技術進展の速い4年後の現在はその可能性がどのような方向に進んできたかを考察しながら読むとより深い理解が得られると思います。ご承知のようにスーパーインテリジェンスとはディープラーニングでブレイクした現在の AI とは全く違う概念であり自己を意識して自己進化する AI という意味で、いずれ将来誕生するであろうと予測して「オラクル」「ジーニー」「ソブリン」「ツール」という四つの AI システムに分類して思考実験を行い理論展開しているものです。そのため文中いたるところに----してもよいかもしれない。----する可能性がある。-----がありうる。-----は現在未解決である。という表現が多用されています。

刃物を使い始めたとき料理にも殺人にも使えるが何とか使いこなしてきた。また核力、放射能については人類は完全にコントロールできていなく、今躊躇している。

博士の AI による人類乗っ取りのシナリオの思考実験は興味をそそるものであり、人類が今までとは質の違うスーパーインテリジェンスを手にした時、はたしてコントロールし得るか?今から備えても遅くはないと警鐘しております。

スーパーインテリジェンスの開発競争において核の開発にも勝る人類史上空前の知能爆発と言えるインテリジェンス革命において、嘗て原爆の開発競争においても見られたような熾烈な開発競争が重大な紛争を巻き起こす懸念についても考察しております。

さらに、疑似的人間とかデジタル知能(人工意識体)に対して開発途上などで大量処分した場合に対する倫理的意義をどう考えるか。実現する前に十分なる思考実験の必要性を論じています。

AGI(汎用人工知能)時代の経済や政治についても論じております。経済についてはマルサスの人口論を引き合いに出して論じています。現在は全生産要素に占める資本の割合はほぼ30%。つまり全世界所得の30%が資本によって生み出され、残りの70%が人の労働所得として生み出されていますが、スーパーインテリジェンスの時代においては労働所得比率はほぼゼロになるであろうと予測しております。これは世界の資本所得比率がほぼ100%になることを示唆している。その場合どのような事態が起こりうるか興味をそそるものが多々あります。

スーパーインテリジェンスが知能爆発を起こす条件トリガーをコントロールする方法についても詳細に言及されています。

スーパーインテリジェンスに適切な価値基準を与えることは容易ではない。最大の難関事項であるとしています。新たなスキーム(数学理論、コンピュータ言語など)を必要とするだろうということです。スーパーインテリジェンスに価値観をどのように持たせるかについても幾つかの提案と問題点も論じられています。

スーパーインテリジェンスに我々はどうのような価値観を与えればよいか。哲学、宗教、政治等々考慮すべきことが膨大にあります。如何にして価値観をスーパーインテリジェンスに植え付ける(ローディング)かも現在まだ未解決であります。

また、多数の人工生命体のようなスーパーインテリジェンスが存在する世の中の構造はどのようになるのか、またスーパーインテリジェンスの外的な面のみならず内的な面をどのように構成するべきかについても論じています。

膨大な未解決問題が未来に託されていることを思い知らされる一冊でもあります。

スーパーインテリジェンスもレベル化し、例えば人類を1とし2,3,4...のような設定も必要になるでしょう。

開発途上には電波に対して電波暗室があるようにスーパーインテリジェンスに対して情報暗室、とか物理的暗室のようなシールド室の必要性を感じますが、これらについても詳しく論じています。

著者は人類を幼子に例えて、核という爆弾を抱えて72年何とか滅びないで生き延びてきたが、こんどはAGIというオモチャ?を手にしたとき我々の精神行動の未熟さとAGIとのミスマッチが生じてしまうことを心配され、今から哲学的、倫理的考察を主にリスクの面にページを割り準備をしておく必要を力説しておられます。

スーパーインテリジェンス獲得後の人類は、スペースコロニー、フォン・ノイマン探査機、分子ナノテク物質製造技術等が本格的に開花しているのであろう。ただし人類の欲望のぶつかり合いによる最終戦争のようなものが起こらない限りと付け加えたい。

将来遭遇するかもしれない宇宙人とはもしかしたら、人間のような有機生命体ではなく、真空中でも自由に活動できる運動体(人工意識体的なものか)のようなものかもしれないと連想させる教養書です。

最新の脳科学者の知見によれば人類は各種臓器とネットワークを組みより長く生き延びる

ように進化の過程で各神経の電気信号のみならずプログラムにおけるサブルーチンの役割を担う種々の Messenger substances と巨大な Network を構築しているのでスーパーインテリジェンスは必ずしも人間の脳と同一になる必要はないと思う。

ユドカウスキー(Yudkowsky)によれば知能は生命の複雑さから生まれたのではない。ひとりで生まれることはなく自然選択による最適化圧がかかったから生まれたと主張しております。同感するものがあると思います。

人類、動物は進化の過程で-鳴き声→コトバ→言語→複雑な言語とコトバを進化させてきたが、スーパーインテリジェンスがこの進化の過程を逆にたどり(リバーズインテリジェンスか?)動物の鳴き声を分析し動物の意志を分析できるようになる可能性もあると思います。いずれ近い将来動物言語学(Anilogology)なる分野が格段に進化し畜産業等に大きく貢献する時が来るでしょう。

スーパーインテリジェンスなるものが定着し広まった後千年、一万年後の人類の命運についてはボストロム氏は触れていませんが、進化生物学者によれば家畜の牛や馬は野生の牛や馬に比べて脳の身体をコントロールする以外の機能は退化して小さくなってきているようです。牛、馬が 1.1 万年ほど経過して家畜化される中で脳が退化している事実を考えますと、スーパーインテリジェンスのおかげでそれに酔いしれていると脳が退化してしまうので、交通機関の発達で筋肉が衰え筋トレが進んだように、遠い将来脳が退化してしまわないような「脳トレ」文化のようなものが大きく進展するような気がします。

最後に、同時に Yudkowsky やウォルター J.フリーマンの著書なども合わせて読めばより理解が深まるでしょう。

<http://yudkowsky.net/>

<https://intelligence.org/>

<https://opencog.org/about/>

<https://www.amazon.com/Eliezer-Yudkowsky/e/B00J6XXP9K>

<https://www.amazon.co.jp/%E8%84%B3%E3%81%AF%E3%81%84%E3%81%8B%E3%81%AB%E3%81%97%E3%81%A6%E5%BF%83%E3%82%92%E5%89%B5%E3%82%8B%E3%81%AE%E3%81%8B%E2%80%95%E7%A5%9E%E7%B5%8C%E5%9B%9E%E8%B7%AF%E7%B6%B2%E3%81%AE%E3%82%AB%E3%82%AA%E3%82%B9%E3%81%8C%E7%94%9F%E3%81%BF%E5%87%BA%E3%81%99%E5%BF%97%E5%90%91%E6%80%A7%E3%83%BB%E6%84%8F%E5%91%B3%E3%83%BB%E8%87%AA%E7%94%B1%E6%84%8F%E5%BF%97-%E3%82%A6%E3%82%A9%E3%83%AB%E3%82%BF%E3%83%BC-J-%E3%83%95%E3%83%AA%E3%83%BC%E3%83%9E%E3%83%B3/dp/4782801718>

なお著者 OXFORD 人類未来研究所の Nick Bostrom 氏のホームページは

<https://nickbostrom.com/>

です。著書以外の情報がつかめるでしょう。