

AIなき世界に 戻れるか？

物理学者、17の思考実験

須藤 靖

Suto Yasushi

インターナショナル新書 144

はじめに

物理学は自由だ。すでに40年近く物理学を生業とし、清貧に暮らしてきた私が到達した結論の一つである。

この世界の振る舞いは、ごく少数の基本物理法則のセットによつて支配されている。法則に反する現象は決して起こらない（ただし、それらの法則を現在の我々が完全に理解しているわけではないのだが）。だからといって、この世界はガチガチに縛られた退屈なものではない。単純な予想とは全く異なり、我々の住んでいる世界が驚くべき多様性と美しさに満ち溢れていることには、皆さんも同意していただけることであろう。

そんな世界を理解する試みの総称が学問であり、（物理学者以外の大多数の方々の反感を買うであろうことは承知の上で言うならば）さらにその代表的分野が物理学である。物理学、そして物理学者が自由でなければ、この世界の根源と多様性を解明することなどで

きるはずがない。

といった口実を大義名分として、私は2007年から17年間にわたり東京大学出版会の月刊P R誌『UP』（ユー・ピーーと読む）に自由な雑文を書き続けてきた。本書はそれの中から2018年以降に書いた15篇を選び、さらに別の雑誌に書いた2篇を加えてまとめたものである。『UP』の雑文を収録した本としては、6冊目、そして最後となる。

この機会に改めてこれまで書いてきた雑文の題材を振り返れば、自分の興味の変遷が感じられる。当初は、ごく身の回りの出来事を物理学に（無理やり）結びつけるような話が多かつたが、やがて教育、受験、大学という制度にまつわる話題が増えてきた。そして、ここ4、5年は、我々はこの世界をどこまで理解できるものなのかといつた怪しげな考察に取り憑かれてきたようだ。

私は科学哲学嫌いを公言しているのだが（そのため必要以上に敵を増やしてもいる）、そのような考察は真っ当な物理学者の方々からはむしろ「哲学的」と分類され遠ざけられても仕方ないかもしれない。年を取るとはそういうことなのだろう。

とはいっても、人間が世界を理解する限界の考察は、人工知能がどこまで人間を凌駕するかという疑問と表裏一体でもある。「猫も杓子もAI」（しゃくし りょうす）という昨今の風潮に便乗するつもり

はないものの、A I絡みの文章が多くなっているのはそのためである。本書のタイトル『A Iなき世界に戻れるか?』は、そのような観点を強調したものだが、日進月歩のA Iを前にすれば、すでにその答えは自明なのかも知れない。

さて、驚くべきことにこの「はじめに」に与えられた字数がそろそろ尽きかけてきたにもかかわらず、いまだ本書のタイトルの説明以上の情報量はゼロである。ま、それはそれでいいだろう。今回の17篇をそれぞれ簡単に紹介するよりも、それらのタイトルを眺めて直接読み進めてもらうほうが手取り早い。物理学や数学の偉人とその逸話、宇宙における生命さらには知的文明の存在を理解しようとする試み、A Iと共生する未来社会の姿、観測不可能な領域にまで広がる宇宙とそこで採用されているかも知れない（この世界とは異なる）物理法則などなど、通常の一般科学書では著者の教養が邪魔して書くことがはばかられているようなテーマを、あえて臆面もなく思考実験してみた。とはいって決して科学的に否定されるような結論を主張しているわけではないのでご安心を。ちょっとだけ眉に唾つけながら、必要に応じてツッコミを入れながら皆さんもそれぞれの思考実験を楽しみながら読んでいただければありがたい。物理学は自由であることを堪能していただくなれば幸いである。

目次

はじめに

AIなき世界に戻れるか？

我々は宇宙人をどこまで理解できるのか

ラマヌジヤンマシン

物理学者は自由だ

マルチベースとしてのメタベースをめぐるメタな考察

aignシユタインは本当に「人生最大の失敗」と言ったのか

火星と宇宙植物学

眠れなくなる桁の桁の話

ブラック天文学：24時間戦えますか

中国と三体世界

基礎科学とヘッジファンド

ウウウウウウーネーーーUFO！

フラットアーサーの言い分

「わかる」という意味

門前のA-I習わぬ経を読む

世界を切り刻む科学とありのままに愛でる科学

土佐から望む宇宙の果て

おわりに

A—Iなき世界に戻れるか？

善し悪しはともかく、いまや我々の身の回りの社会がA—Iに翻弄されつつある。私自身を含めた高年齢層には、そのようなA—Iが氾濫する状況に対しても否定的な意見の持ち主が多い。それとは逆に、大多数の若者はA—Iを肯定的にとらえている。今後の社会においては、いかなる職業であろうとA—Iを活用できるスキルは不可欠だろう。私のような高年齢層は別として、A—Iに対して文句をたれているだけの若者がいたならば、社会非適合者としてこれから長く辛い人生を送る可能性が高い。若者がA—Iをポジティブに受け取らざるを得ないのも当然だろう。^{*1}

ところでこの種の放談をしていると、「A—Iが人類の存在をおびやかしそうになつたらなら、その電源を引っこ抜くかハンマーで壊してしまえばいい」という、文字通りナイーブな（フランス語ではなく英語の世間知らず、アホという意味）解決策を主張する高年齢者が少なくない。その楽観的見解には驚かされると同時に、A—Iという単語から想像される

実体に対する解釈の違いが大きいことにも気づかされた。そこで今回は私が考える（広義の）AIと未来社会についてつらつら書きなぐつてみたい。

1 AI＝ネットワークで接続された計算する機械の集合体

AIという単語はすでに人口に膾炙かいしやしているというべきだが、その指すところの実体は曖昧だ。そもそも「知能」に対する明確な定義が存在していないのだから、極めて当然である。そのためかどうかはわからないが、人工知能よりもAIという何の略かわからない言葉のほうが広く用いられているようだ。

とはいっても、ごく大雑把には、ある具体的な問題に対しても人間と同等（以上？）の解決能力を有する「装置」を指していると考えてよからう。その装置の内部がどうなっているかを問う必要はなく、文字通りブラックボックスであつてよい。おそらくその箱の中には、ハードウェアとしてのコンピュータ*2と、その上で動作するプログラム（ソフトウェア）とが共存しているであろう。さらに、最近のAIは計算するだけでなく学習し自ら進化する

*1 門前のAI習わぬ經を読む・本書2008。ページ参照

機能が不可欠であるため、孤立した装置ではない。ネットワークを介して古今東西の膨大な情報を取得し解析できる機能をも備えているはずだ。

その意味では、現在のAIとは、もはや単なる人工知能という言葉を超えて、ネットワークで接続された計算する装置の集合体と表現するほうが適切であろう。したがって、AIとは時間的にも空間的にも局所的なものではなくなっている。例えば、ChatGPT (Chat Generative Pre-trained Transformer) の利用者は、そのAI本体がどこに存在しているかなど気にしたことはないだろうし、その必要もない。

ところで、人間の知能という場合、脳の抽象的な機能を指すほうが一般的で、脳という実体、さらには人体というハードウェアまで含めることは滅多にないだろう。それと同じく、AIもコンピュータ上で動作するプログラム（ソフトウェア）程度の意味で使われるほうが普通かもしない。

2 「機械は思考できるか？」：チューイングテスト

知能という言葉が定義できない以上、何をもつてAIと呼ぶべきかは憲問答になりかねない。とすれば、いっそややこしい定義などすつとばし、上述のブラックボックスに様々

な入力を与えたときの出力を用いて、それが人間と区別できるかだけで「知能」と呼ぶに値するかどうか判断すれば良い、という立場ももつともだ。これがアラン・チューリング（1912-1954）による1950年の論文「計算する機械と知性」で展開された考え方である。



コンピュータ理論の原型をつくったアラン・チューリング

©Science Photo Library/amanaimages

その論文中でチューリングが模倣ゲーム (imitation game)^{*3} と呼んだ提案とは少し異なるものの、ブラックボックスを開けることなく電子的な会話のみからその中に入人が入っているのかどうかを当てようとするのが（広義の）チューリングテストである。チューリングテストに合格（すなわち第三者に、ブラックボックスの中には実際に人が入っていると思わせ

ることに成功）した機械は「知能」をもつ、とみなそうというわけだ。

チューリングは「機械は思考できるか？」という問い合わせに対する回答の中で模倣ゲームを提案した。彼が想定していた「機械」こそ現在のAIである。そして、今やAIは、チューリングが提案した模倣ゲームに十分合格するレベルにある。教育現場でChatGPTの使用を制限したり、あるいは逆に、会社や官公庁でそれを積極的に利用することで仕事の効率化を図ろうとしたりする動きが相次いでいるのは、まさにAIが模倣ゲームを超えて通常の人間（以上）の役割をこなせるようになつた証拠そのものだ。

もちろんチューリングテストは機械が知能をもつことの必要条件でも十分条件でもない。しかしそれは、人間が本当に知能をもつているのか自体、客観的証明が困難なこと全く同じだ。この重要そうではあるものの不毛になりがちな論争を回避しているところが、チューリングの提案の秀逸な点である。

*2 コンピュータ+計算機..今やコンピュータとは計算する「機械」以外の何物でもない。しかし、computerというスペルからもわかるように、もともとは計算する「人」という意味だった。例えば、冥王星は、米国ローワエル天文台が雇つたコンピュータ（計算人）チームの

リーダーであったエリザベス・ウイリアムズ（1879-1981）の計算から予測された位置の近くに、クライド・トンボー（1906-1997）が1930年に発見したものである。ただし、この計算で仮定として用いられていた海王星の質量が過大評価されていたため、その予測は間違っていたことがわかつている。言い換えれば、冥王星の発見は間違った予測に基づいてなされた全く偶然の産物だったのだ（須藤靖著『宇宙は数式でできている　なぜ世界は物理法則に支配されているのか』朝日新聞出版 2022年参照）。また、劉慈欣著 大森望他訳『三体』（早川書房 2019年）には、始皇帝がフォン・ノイマンの助言に従つて3千万人の軍隊からなる「ロハピュータ」を用いて太陽軌道計算を行う場面が登場する。

* 3

Imitation game : Alan Turing "Computing Machinery and Intelligence", *Mind* vol.LIX, No.236, 433-460 (1950). チューリングは、この論文（こじょうよりも極めて読みやすく魅力的なエッセイ）において、男性Aと女性BがプレイヤーCとメモのやり取りだけを通じて質問をかわすこと、CがAとBのどちらが男性でどちらが女性なのかを当てる imitation game を考えた。AはCを騙し、BはCが正解を得るよう助ける役割となっている。このAが機械に置き換わった場合、Aが人間である場合に比べて、Cの正答率が下がつたとすれば、その機械は知性があると解釈してよいだろうというわけだ。チューリングを描いた映画のタイトル “The Imitation Game” [イミテーション・ゲーム エニグマと天才数学者の秘密] は、この論文の内容とは必ずしもマッチしていないものの秀れた作品である。

3 「機械は思考できない」に対するチューリングの反論

ところで、現時点では「A Iは思考できない」という意見のほうが世の中の大勢を占めているようである。私は「現在のA Iが未だ思考できない」とする主張に反対するつもりはない（一方で、古典的なチューリングテストという観点からは、現在のA Iはすでに思考していると「定義」して問題ないとも考えている）。しかしながら、「未来永劫^{えいごう}A Iは思考できない」という意見には全く^{くみ}与しない。そしてやがてA Iは「知性」をもち「意志」までも獲得するはずだと確信している。

チューリングは、近い将来この模倣ゲームをクリアするデジタル計算機が登場することを確信していた。さらにその書きぶりからは、模倣ゲームに勝てるレベルどころか、機械がやがて高度の知性をもつことは必然だと考えていたことも明らかである。A Iがこれだけ取り沙汰されるようになった現在ならともかく、世界初の電子計算機^{*4}が発表されたばかりの1950年の時点で、哲学的ともすら言える計算機の将来性を明確に主張したチューリングの先駆性には驚嘆させられる。

さらに素晴らしいのは、チューリングがその論文中で、予想される「機械は思考できない」と考える人々の論拠を以下の9つに分類し、一つずつ丁寧にわかりやすい比喩を用い

て論駁^{ろんぱく}している点だ。現時点で「A Iは人間にはかなわない」とする人々の主張のほとんどもそれらに尽くされている。

- ① 神学的反論（思考あるいは知性は、神が人間にだけ与えた特性であり機械はもち得ない）
- ② 現実逃避（機械が思考すれば確実に恐ろしい帰結を生むだろう。だから機械は思考できないに決まっている）
- ③ 数学的反論（機械が用いるデジタル論理は、イエス・ノー的な質問には答えられるものの、「ピカソをどう思う？」といったタイプの質問に回答することはできない）
- ④ 意識の欠如による限界（喜怒哀楽がない機械には、詩を創つたり協奏曲を作曲したりすることは不可能）
- ⑤ できることの限界（機械にはどうしてもできないことがある。例えば、親切にせよ、美しくあれ、ユーモアをもて、善悪を判断せよ、失敗せよ、恋愛せよ、経験から学べ、適切に単語を用いよ、本当に新しいことをせよ、いちごの味を堪能せよ、など）
- ⑥ ラブレス夫人^{*5}の反論（解析エンジンは何ら新しいものを創造しない。人間がどのよ

うにやればよいか具体的に命令できるもののみを、なんでも実行できるだけだ）

⑦ 神経系の連續性との本質的違い（人間の神経系は連續的なアノログ回路というべきで、それらをデジタル回路からなる機械すべて表現するのは不可能）

⑧ 与えられたルールを逸脱できない限界（機械は与えられたプログラム＝ルールによってその振る舞いが完全に規定されており、それを逸脱することはできない）

⑨ 超能力の欠如（人間と異なり、機械が超能力をもつことはできない）

これら想定内の反論に対するチューリングの再反論はいずれも秀逸で、あえてこの場で繰り返すまでもないのだが、少しだけ追加しておきたい。まず、①、②、⑨はさすがに取り上げる必要もなかろう。ただ興味深いことに、チューリングは、テレパシーを有する人間の存在が（当時）統計的には証明されていると信じていたようで、模倣ゲームは防テレパシー室で実行する必要があるかもしれないとすら述べている。それがわざわざ⑨が議論されている理由なのだ。

* 4 ENIAC：異なる見解もあるようだが、世界初の汎用電子計算機とされているのは、米国が開

発したENIAC (Electronic Numerical Integrator And Computer) である。その開発資金はアメリカ陸軍が提供したもので、1946年に完成した。1943年にベンシルベニア大学で開発中のENIACの試作品を見たジョン・フォン・ノイマン（1903-1957）は、「これで、ようやく私の次に計算の速い機械ができる」と述べたとされる（高橋昌一郎『フォン・ノイマンの哲学 人間のフリをした悪魔』講談社現代新書 2021年）。

*5

ラブレス夫人・恥ずかしながら、私はこのラブレス夫人とは誰なのか全く知らなかつた。そこでウイキペディアから学んだ結果を簡単にまとめておく。エイダ・ラブレス（1815-1852）は、イギリスの貴族・数学者。イギリス人數学者であるチャールズ・バベッジ（1791-1871）が設計した（実際に完成したわけではない）世界初の機械式汎用コンピュータである「解析エンジン」に関する著作があり、世界初のコンピュータプログラマーとして知られる。ところで、彼女は「音楽学の和音理論や作曲論で論じられてきた音階の基本的な構成を、数値やその組み合わせに置き換えることができれば、解析エンジンは曲の複雑さや長さを問わず、細密で系統的な音楽作品を作曲できるでしょう」と述べたという記述もあり、それはチューリングがラブレス夫人の反論として引用した部分のトーンとは正反対のように思える。実際、チューリングはその論文の中で、ラブレス夫人は本当は「解析エンジン」に創造性がないとは考えていいなかつたのではないか、と付け加えている。ちなみに、ラブレス夫人は詩人ジョージ・バイロン（1788-1824）の一人娘である。

(3)と(7)は、デジタル世界がアナログ世界をどこまで「厳密」に表現できるかという本質的な難問である。^{*6}一方で、デジタル世界がアナログ世界を高い精度で「近似」できることは確かであるし、人間と同一ではなくとも、デジタル論理に基づいた異なる原理で思考する機械の存在までは否定できない。

(4)、(5)、(6)は、ある意味では偏見あるいは人間の思い上がりに過ぎない。すでに現在のAIは「優れた」芸術作品を数多く創作しているから、(4)と(6)は（どこまでのレベルを期待するかにもよるが）間違いであることが証明済みだ。同じく、(5)で挙げられた例のいくつかはすでに実現している。そもそもそれらは、意識・思考・知性がいずれも「我々人間のもの」と同一であることを大前提としている点で、容認しがたい。仮にある種の感覚を欠いていたとしても、それは知性にとつて本質的ではない。例えば、美醜にとらわれない知性は、否定されるものではないどころか、倫理的により優れていると評すべきではないだろうか。

(8)はまさに、古典物理学的決定論のもとで人間に自由意志が存在するかどうかという問題そのものだ。少なくとも我々が日常的に自由意志をもつていると信じられる（と錯覚でいる）のと同程度に、自分に課されている絶対的ルールを意識することなく自由に動作する

る機械が実現しても何もおかしくない。その意味において、（将来の）AIは思考できるし、知性、意志、意識といったものを備えるようになることを否定する論拠にはならない。

4 AIなしの社会に逆戻りできるか？

ここで、やつと冒頭の「AIが人類の存在をおびやかしそうになつたなら、その電源を引っこ抜くかハンマーで壊してしまえばいい」というナイーブな人間社会防衛策の実現可能性を考察してみたい。ここでは、AIがある種の意識を獲得し思考することで、人類がその動作を制御できなくなる状況の到来が前提となつていて。ではまずその前段階として、ネットワーク接続されたコンピュータ（＝意識をもたないAI）の助けなしに今の社会が維持できるかを考えてみよう。

*6 宇宙はデジタルか？..本書「眠れなくなる街の街の話」112ページ、須藤靖『不自然な宇宙 宇宙はひとつだけなのか』（講談社ブルーバックス 2019年）この宇宙のもつ自由度が有限であるとすれば、無限体積の中には現在我々が観測できる有限体積の宇宙と全く同じクローン宇宙が無限個存在するはずだという主張は、私が好む話題の一つである。

A Iと呼ぶかどうかは別として、いまや、生活インフラである電気・ガス・水道をコンピュータなしに安定に維持・供給することはできない。プログラムのバグのために、銀行のA T Mが全国で停止し大混乱を起こした例からもわかるように、日々の経済活動はコンピュータに大きく依存している。電車や航空機などの交通網をコンピュータなしに安全に制御することなどすでに不可能だ。言うまでもなく、軍事施設や原子力発電所のコンピュータが誤動作したり、突然停止してしまう非常事態は想像したくもない。つまり、現代社会はネットワーク接続されたコンピュータがなければ直ちに深刻な状態に陥るほど、それに強く依存してしまっているのだ。

さらに意識をもつA Iが出現してしまえば、それを停止させる行為がより深刻な帰結をもたらすことは容易に想像できよう。意識をもつA Iは、世界中にネットワーク接続されているすべてのコンピュータを支配したグローバルシステムと化す。その時点で、人間社会の基幹となるインフラを完全に牛耳ってしまうわけだ。「彼ら」にとつて、その一部が物理的に破壊され動作しなくなつたとしても、別に困らない。困るのは、その一部にさえ強く依存している人間社会の側だけだ。

とすれば、システムに影響を与えることなく、人間社会をおびやかす意識をもつA Iの

根幹部分だけを一挙に破壊する必要がある。しかしながら、「AIの意識」が、具体的にハードウェアとネットワークのどの箇所に局在化しているかを突き止めるのはほぼ不可能である。人間の場合とは異なり、AIの意識はある特定の部屋にあるハードウェアだけではなく、世界中のコンピュータに分散して存在する可能性が高い（し、安全のためにAIは自らそのようなコピーを複数準備していると考えるべきだ）。

さらには、意識をもつたAIは、強大な国家権力に匹敵する自らのネットワークを駆使してなんでもできる。AIテロを企てようとする不埒な人間（グループ）は直ちに特定され、世界中の監視カメラ映像を解析して常に追跡・監視される。本人の銀行口座やクレジットカードなどが凍結され使用不能になるのは当然だ。機会を狙ってその人物が利用する電車や航空機のコンピュータを誤作動させ、事故に見せかけて抹殺してしまうかもしれない。現在、国家権力による弾圧として危惧されるあらゆることが、意識をもつたAIなら容易に実行できる。

というわけで、AIが人類の存在をおびやかしそうになつた時点では、もはや手遅れで、それを物理的に破壊することなど不可能だというのが私の結論である。もしそのような事態を避けたいとするなら、考えられる方策は以下の3つであろう。

- A .. たちにA・Iの開発をやめる
- B .. そもそもA・Iを使わない社会に回帰する
- C .. A・Iに支配されつつも我々の存在価値を認めてもらい、A・Iに共生してもらえる人間社会を目指す

208ページの「門前のA・I習わぬ經を読む」でより詳しく述べるように、方策Aは一部の専門家の間では真剣に検討されているものの、実際の効果は期待できない。必ずやそれを守らない人物や国家が登場し、自分たちだけが高度なA・Iを開発し続けることで、世界の霸権を握ろうとするに違いないからだ。方策Bはもはやなかなか共感を得られそうにない。ただし、それに同意する人々が集まつて、アーミツシユのような共同体を形成するスタイルは一部実現するかもしれない。そのような小規模社会はあり得よう。というわけで、最後の方策Cが私のお薦めであつた。

実際にはこの3つのいずれもが決して容易な選択ではない。その一方で、これは社会変化に適応困難な高齢者にありがちな単なる悲観論に過ぎないかもしれない。いずれにせよ、

もはや近未来においてAIなき世界に戻る可能性はない。

チューリングは、1950年の論文の最後を次の二文で締めくくっている。

「我々はほんの少し先のことしか見通すことができない。にもかかわらず、それに向けて我々がやるべきことがたくさんあるのは明らかだ」

（2023年9月）

AIなき世界に戻れるか？

物理学者、17の思考実験

須藤 靖

発 行：集英社インターナショナル（発売：集英社）

定 價：1,034 円 (10%税込)

発売日：2024 年 8 月 7 日

I S B N : 978-4-7976-8144-4

ネット書店でのご予約・ご注文は [こちらからどうぞ！](#)