

令和 7 年度

一般選抜前期日程入学試験問題

総合問題

注意事項

- 1 試験開始の合図があるまで、この冊子を開いてはいけません。
- 2 問題冊子（17 ページ）には、解答用紙 3 枚と下書き用紙 2 枚が挟み込んであります。試験開始の合図があったら、直ちに中を確認、印刷や枚数の不備などがあった場合、監督者に申し出なさい。
- 3 問題冊子の間に挟み込んである解答用紙を取り出し、すべての解答用紙の所定欄に受験番号を記入しなさい。
- 4 解答はすべて、解答用紙に横書きで記入しなさい。間違って下書き用紙に記入しても、回収しません。
- 5 句読点は、一字と数えなさい。
- 6 試験室で配付された問題冊子及び下書き用紙は、退出時に持ち帰りなさい。

このページは白紙です。

このページは白紙です。

問 題

課題文①・②・③を読んで、次の問いに答えなさい。

問1 課題文①下線部(ア)に「心の《拡散》と《集中》」とあるが、どのようなことを指しているか、本文のことばを用いながら 45 字以内で答えなさい。

問2 課題文①下線部(イ)に「意識と世界を、情報という同じプロトコルで処理することが可能なのではないかという期待は、歴史的にも革新的なものであった」とある。筆者がそう考える理由を、「精神の歴史」・「情報技術」ということばを用いて、75 字以内で説明しなさい。

問3 課題文②下線部(ウ)「ヒトの脳の動作をモデル化したネットワークである」という A I についての説明に関して、筆者は「奇妙な説明」と述べている。筆者がそう考える理由を 30 字以内で答えなさい。

問4 課題文②下線部(エ)「A I の脆弱性^{ぜいじやく}と危険性」とは具体的にはどのようなことを言っているか、本文の内容に即して 150 字以内でまとめなさい。

問5 課題文③下線部(オ)に「このようなロボットを作るのはきわめて難事業」とある。筆者がそう考える理由を、「このようなロボット」が指す内容を明確にした上で、170 字以内で説明しなさい。

問6 課題文①の二重下線部「20世紀が機械によって人間の肉体や行為が凄まじい速度で代替されていった時代だったとすれば、21世紀は機械が同じような速度で人間の精神やコミュニケーションを代替していく時代になるだろう」という筆者の主張について、あなたはどうか考えるか、課題文①・②・③の内容を踏まえて、500字以内で述べなさい。その際、次の語群から3つ以上の語を選んで用い、用いた語はすべて答案上に□で囲みなさい。

〈語群〉

コスト アウトソーシング 脆弱性 危険性 目的意識 クオリア

課題文①

（ 次の文章は下西風澄『生成と消滅の精神史 終わらない心を生きる』の終章の一部である。）

カントは、答えることができないはずなのに、それでも問うてしまう人間の宿命的なジレンマこそが哲学の課題だと言った。私たちの心は世界のはじまりについても、世界の終わりについても考えることはできない。それにもかかわらず、私たちはいつも唯一の起源を求めてしまうし、また^{など}辿りつくべき終着点を求めてしまう。カントがすでに数百年前に警戒した心の誤ちは幾度となく人間を誘惑する。アルベール・カミュの書いたカリギュラは、月が欲しい、なぜならそれは不可能なものだからだ、と言った。不可能にむかって突進してしまうこと、それこそが人間の弱さであり愛すべき存在の理由である。

本書がこれまで辿ってきた精神の歴史は、（ア）心の《拡散》と《集中》の往復の歴史であると言いたい。人間の心はあるときは世界へと拡散し、あるときは自己へと集中した。

《拡散する心》は私の身体に囲われることはなく、樹々や水や夕陽や虫たちと、あるいは空の神々と、心を分散して共有した。心はネットワーク上に分散されており、私一人が意志を決定したり、感情を専有する必要はなかった。私が一人で世界全体について理解する必要もなく、鳥のことは鳥に任せ、海のことは海に任せ、心はそのようなネットワークに触れることでたしかめられた。世界を理解することは、世界に触れ、世界に成り、世界と交流することだった。

他方で《集中する心》は世界に分散された意識の切片をこの一個の身体のなかに凝縮し、すべてを一人の心に束ね上げた。意志はこの私という小さな箱の中で独立し、感情は何者とも分かち難いアイデンティティを構成した。鳥のことも海のことも、私の表象／対象としてはじめて理解可能なものであって、心は複雑なネットワークを切断して世界を俯瞰^{みかん}することでたしかめられた。世界を理解することは、世界を対象化し、記号化し、再配列することだった。

心のモデルは、この拡散と集中をくり返してきた。その往復、あるいは緊張、あるいは^{そうこく}相克の歴史こそ本書が捉えたかった精神の歴史である。あるいはこれを運動として捉えれば、《発展（develop）》と《内包（envelop）》をくり返してきたと言ってもよい。「velop（包み）」を開いて（de）、世界へと伸び拡がってゆく

運動と、世界を心の「velop（包み）」の内（en）へと内包してしまう運動。

春の雲ひとつない青空へと融け入る心、夏の鬱蒼とした木立に風が吹くように澄み渡る心。また、秋の夕暮れに名前を与えて包み込んでしまう心、冬の凍てつく空気の孤独の中で星を繋いで記号する心。私たちの心は世界へと融け入って伸び広がる性質と、世界を可能な限り把握して内部へと収めてしまうという性質の両方を持っている。

ホメロスの心は風のように世界へと《拡散》していたし、ソクラテスはこれを《集中》させて自己の内に収めた。デカルトはこの《集中》する心の中心に「我」という原点と「考える」という機能を与え、パスカルは《拡散（develop）》と《集中（envelop）》を同時に実現しようとして引き裂かれた。カントは世界と心を分離し、《拡散》と《集中》を分割統治する方法を考案した。フッサールは《集中》の束を一つずつ手に取りながら緩やかにほどこき、ハイデガーは《拡散》する心の可能性を持ちながら、その裏側で極度に《集中》する心を論理化してしまった。

初期の認知科学は《集中》する心を記号と推論によって強化させたが、「情報」という世界との交通路を開き、媒介することで容易に心を世界へと《拡散》する可能性も加速させた。自然現象はコンピュータによってシミュレーション可能な存在となり、同時に意識もコンピュータ上でシミュレーション可能な現象に思われた。(4)意識と世界を、情報という同じプロトコルで処理することが可能なのではないかという期待は、歴史的にも革新的なものであった。すなわち情報技術は《拡散／集中》の方向性そのものを無化するような《媒介》という新たな心の可能性を開いた。一方で私たちは情報を通じて世界へと《拡散》することも容易であり、他方で世界を情報を通じて回収して《集中》することも容易になった。情報技術を媒介にした私たちの心は、むしろ世界へと拡散しているのかあるいは集中しているのか、そもそもが区別不可能な媒介的な接続状態を創り出した。

現代の情報技術が無際限に世界へのアクセスを切り開いたことで、意識と世界の交流はなだらかで終わらない手続きであることがリアルなものとなった。その意味で、はじめから出発点も終着点もないメルロ＝ポンティの思想の重要性は際立ってくるように思える。

世界とは、私が思惟しているものではなくて私が生きているものであって、私は世界へと開かれ、世界と疑いようもなく交流しているけれども、しかし

私は世界を所有しているわけではなく、世界はいつまでも汲みつくし得ないものなのだ。

メルロ＝ポンティは、心と世界の《拡散》と《集中》の緊張関係が拮抗しながら運動するプロセスを描いた。この終わりのなき運動がメルロ＝ポンティにとっての哲学の出発点であり、終着点はなかった。ヴァレラもそれを強く自覚していたからこそ生命という次元にまで遡ったし、逆に夏目漱石は、この意識と世界の分断と拮抗を生き抜くにはあまりに繊細で過敏すぎた。世界に心を《拡散》すれば私は消えてしまうし、世界を心に《集中》させようとすれば、すべてを渦のように絡め取ってしまおうとするほど激烈であった。

汲み尽くしえない世界。それは豊かさであると同時に恐怖でもある。人類の歴史は、この豊かさと恐怖のなかで心を構築してきた。私たちの心は、この汲み尽くしえない世界に対してどのように生きるのかということに賭けられている。この恐怖と豊かさを同時に与える世界に描く「心」という人間の自画像、決して完成することのない自画像にどう向き合うのか、それだけが重要な問いであると本書は考えている。

心は目に見えぬものであるからこそ、そのイメージは自らを制約する。洞窟に壁画を描いた人類は、いつからか心という不可視の存在をイメージし、その存在の在り方を身に纏ってきた。生命の哲学者ハンス・ヨナスは人間の自画像についてこう語っている。

人間は人間という理念を「生きている」—— それに同意するにせよ、反対するにせよ […] 時として幸福な動物へ戻りたいという願望を人間のうちに呼び起こすとしても、人間の像は人間を見放しはしない。

人間は自らのイメージにどこまでも囚われた存在である。しかし同時に、自分とは何者なのか？ 私の心とは何なのか？ そう考えようとした瞬間に、そこに見える心や自分は、そのような問いを発した最初の自己からはすでに離れてしまっている。見ようとしているそのものが手のひらにある。終わりのなき循環、この永遠に届かない、すれ違い続ける無限の鏡。自己に直面することで不可避的に出会う裂け目。これこそが人間に絶望をもたらし、また高揚をもたらず。悲しみも

喜びも、この人間の心が抱く自らのイメージとの裂け目に生じるのだ。本書が追跡してきた心のメタファー、心のイメージとは、人間が自らを捉えようとして描く自画像にほかならない。あるときは人間は自らに神を覗き込んだし、またあるときは自らにコンピュータを見出した。最後に、おそらくは現代の人類の最も象徴的な自画像である「A I」について触れて終わろう。

Artificial Intelligence。人工知能は現代における最も象徴的な私たちの心の自画像であると言っていいだろう。カントは心を形式化することによって、ある種の全能な機能を持つソフトウェアに仕立てるとともに、それ自身は何ものでもなく単なるデータ処理のアーキテクチャであるという空虚な存在として、パスカルの矛盾を両義性のうちに調停した。その意味で、本当にA Iを思想的に用意したのはカントかもしれない。A Iこそ、心の不安、心の暴走を鎮めるために私たち人類が造った心の鎮魂歌だったのだ。

ネグリとハートが『〈帝国〉』で早々に指摘したように、現代では労働は「非物質化」（知識・情報・情動・コミュニケーション）され、グローバルな帝国は領土的で物理的な支配よりもむしろ、アメーバ状に張り巡らされたネットワークを生活の隅々にまで浸透させ、非物質的な情動やコミュニケーションを通じて、微分化された^{コマンド}指令群として私たちの脳と身体を支配する。情報技術が包括する私たちの精神は今や常にネットに接続している。あらゆる意思決定はインターネットを経由して広告やマーケティングに代替され、コンテンツは最適化されたアルゴリズムをもとに提供され、自らの精神で考えることは燃費の悪い車に乗って走ることのようなのだ。あなたが自分で考える必要はない。思考はアルゴリズムに委ねましょう。そうした網の目が、いま無数に私たちの精神のノードを奪いつつある。精神はあなたの身体にある資産ではない。精神はネットワーク上の変数として、アルゴリズムにデータを与える変数としてはじめて価値を持つ。

こうした現代の問題の核心は「心を持つことのコスト」に起因している。自分で自分の意思決定をすることのストレス。A Iもインターネットも、宗教も、資本主義も、あるいはマスメディアと連動した民主主義さえも、これら現代を覆う巨大なシステムたちは、いかに心のコストを代替する装置としてうまく機能するかという覇権争いをしている。しかしこのような事態はまったく驚くに値しない。なぜなら本書で見てきたように、かつて私たちは心を一個の個体で所有していたのではなく、自然や神々という膨大なネットワーク上で分有してきたのだ。一個

の身体だけで処理できる心のキャパシティをそう簡単に増やすことはできず、その仕事量が増大すれば、そして自然や神々と分有することができないのであれば、A Iやインターネットや資本主義を通じた匿名的なシステムと分有することになることは必然である。現代では、一個の人間が抱えきれなくなった心の仕事を、技術が、制度が、市場が、様々なかたちでその役割を引き受けようとしている。

A Iへの欲望や期待というのはつまるところ、精神を持つことのコストに耐えきれない人間たちの精神のアウトソーシング（外注）なのだ。歴史的に見ればA Iは長い夢であり希望であったが、現実に精神への負担の速度が加速していくとき、それはもはや希望ではなく逃避である。人間は自分たちの有機体という母体のなかに、心というひとつの情報処理機構を開発し、それによって世界をミラーリングして再生産し、かつそのシミュレーションによって世界それ自体を創り変えていくという循環システムを作ることで、知的労働を最も効率のよい生産メソッドとして定着させた。しかし、^{ぼうぼう}茫漠とした巨大な世界の情勢をこの小さく制限された有機体のなかで処理するには様々な副作用があることを本書で見てきた。もし仮に人間に代わってA Iが世界を処理して再生産してくれれば、人間はその精神（知的）労働から解放される。かつて人間が産業革命において肉体労働を機械労働に代替させることで解放されたことと、全く同じ歴史の要請がいま繰り広げられている。

20世紀が機械によって人間の肉体や行為が^{すさ}壊まじい速度で代替されていった時代だったとすれば、21世紀は機械が同じような速度で人間の精神やコミュニケーションを代替していく時代になるだろう。

下西風澄『生成と消滅の精神史 終わらない心を生きる』（文藝春秋、2022年）による。一部改変。

課題文②

脳は人が考え得る精密機械とは異なっているが、脳を真似している、あるいは脳に近づいている機械があるという。それはA I（人工知能）と称されるシステムであり、高性能の機械（コンピュータ）で動作するプログラムである。従来、ヒトが行ってきた作業のうち、言語を介したコミュニケーションをA Iが担当す

る場面が目立っているが、その他、医療分野における画像解析や車の自動運転などでも活用されつつある。ヒトの身体が行ってきた力仕事を動力機械が担当するように、ヒトの脳が行ってきた作業のいくつかをA Iが担当するようになりつつある。このことから、A Iはいずれ脳と同じように働くようになり、将来、ヒトの脳を超えた人工脳に発展するという予測も出てきた。

しかし、A Iが脳と同じように働くようになれば、A Iも心をもつということになるが、それはあり得るのだろうか？ そもそも、現在流布している「脳を真似たA I」というのは、本当に真似しているのだろうか？

A Iとは、人の作業をサポートしたり、人の代わりを務めたりするために開発されたコンピュータプログラムである。最近、その性能は急速に向上しており、将棋ではプロ棋士を打ち負かし、医療現場では人が見逃してしまうような病変を見つけ出し、自動車の自動運転も実現に近づいているらしい。人が苦手とする作業、あるいは人ができない作業を担当するシステムとして、今後も活用されていくことはまちがいない。なぜここまでA Iが実用化されるようになったのか、その主な理由は二つある。一つは、従来から続いているコンピュータの性能（演算速度と記憶容量）の飛躍的向上であり、もう一つは、やはり従来からあるニューラルネットワークを発展させたディープラーニング（深層学習）という新たな計算法（アルゴリズム）の開発である。

ニューラルネットワークとは、ニューロンが繋がった神経回路の動作を真似た数学的モデル、つまり計算式である。その真似ている神経回路の動作とは、一つのニューロンに複数のニューロンからの入力があると、そのニューロンが発火し、それが次のニューロンに伝わるという一連の動作である。またそれに、ニューロンが発火すると、その入力部分つまりシナプスが変化することでニューロンの感受性（ニューラルネットワークでは重みづけと呼ぶ）が増し、次の入力の効果がより大きくなる、という動作も加わる。このように振る舞うネットワーク（計算式）が、入力層、中間層（隠れ層ともいう）、出力層に分かれて存在し、入力された情報を入力層→中間層→出力層と処理していくことで、最適な答を出していく。ディープラーニングとは、この中間層の数を増やしたニューラルネットワークであり、その結果、答の精度を飛躍的に向上させることに成功した。この成功には、新たなアルゴリズムの開発はもちろんであるが、多くの層にまたがる膨大な計算量を短時間でこなせる高性能コンピュータの登場が大きく寄与して

いる。

しかし、このようなディープラーニングにより本格的な実用化が始まったAIについて、奇妙な説明を目にすることが多い。それが「(a)ヒトの脳の動作をモデル化したネットワークである」という説明である。たしかにディープラーニングの基礎にあるニューラルネットワークは、ニューロン間の信号伝達を真似ているが、実際の脳では一つのニューロンに数千の入力部分、すなわちシナプスがあり、そこでは非常に不確実で確率的な信号伝達が行われている。しかも、入力を担当する樹状突起上で生じる信号の逆方向^{でんぱ}伝播、変化することがない直接結合である電気シナプス、軸索を覆うミエリンの変化による信号の伝達速度の調整、電気信号以外に軸索上を機械的に伝わる圧縮波などが、脳の信号伝達には関係している可能性がある。さらには、シナプスを介さない細胞外空間における神経修飾物質の拡散や、そこを電場とした信号伝達も関わっているらしい。つまり、脳の動作はまだ未解明であり、それをモデル化することは不可能なはずである。

他にも「脳内のネットワークが徐々に学習を重ねることで画像や音声を認識するという方法を真似ている」あるいは「大脳皮質の神経回路構造を真似て」などの説明もよく見るが、どちらも実際の脳ではよくわかっていない。ヒトの認識のメカニズムはいまだ不明であり、それが徐々に学習を重ねることで可能となるとは、必ずしもいえない。動物の幼体やヒトの乳幼児は、かなり初期の段階から人の顔を他の物体と区別できるからである。

また、大脳皮質の回路構造は当初の予想よりはるかに複雑で、その解明は困難をきわめている。ヒト脳の神経回路の構造を完全に解明することを目指すヒトコネクトーム計画が始まったのは2009年であった。しかしこれまでに、302個のニューロンと7800カ所のシナプスしかもたない全長1ミリの線虫（C・エレガンス）の神経回路構造を解明しただけである。この脳をもたない小さな虫を相手に10年以上の歳月が必要であった。一方、ヒトの脳には1000億のニューロンと500兆のシナプスがある。コネクトーム計画は現在、ハエの神経回路構造に取り組んでいるが、それが解明された後、ようやく哺乳動物のマウスに取り掛かるらしい。

もちろん、このような批判は、決してAIの価値を否定しているわけではない。また、言葉尻をとらえたり揚げ足取りをしたりしているわけでもない。AIの研究者が「脳を真似て」というとき、それは「現時点でわかっている脳の概略から

ヒントを得て」というぐらいの意味であり、脳が完全にわかっているとは誰も思っていないであろう。しかし、脳についてこれまでわかってきた断片的な知見や、教科書に載っているような概略こそが本質であるとは、現時点ではとてもいえないであろう。現在、次々と見つかった詳細な構造や機能が、あるいは、まだ見つからない構造や機能が、きわめて重要である可能性も高い。

AIが脳を真似ているという言動をしっかりと批判しなければならない理由がもう一つある。それは、AIの研究者の一部が（少数ではあるが）、AIの研究が、特にディープラーニングの研究が、逆に脳の神経回路で起きていることを明らかにすると本気で考えているからである。AIが人と同じように画像を認識できると、脳はAIのような方法で画像を認識していると考え。あるいは、メタ記憶（覚えていることを覚えていること）のような高次な機能も、AIで同じように実現できることがわかっているが、そこから脳はメタ記憶をAIのような方法でつくっていると結論するのである。しかし、これはさすがにまったくまちがっている。空を飛ぶ飛行機をつくっても、それが、鳥が空を飛ぶメカニズムを明らかにしたわけではないのと同じである。

（中略）

このようにコンピュータと脳は、構造についても、またそこを流れる信号についても、まったく性質が異なっている。コンピュータ上のプログラムであるAIの動作が、脳の動作を解明する上でまったく参考にならないことは明らかである。

たしかにAIは多くの場面で人にとって代わり、また人の能力を凌駕^{りょうが}しつつある。しかしそのことは、すべての面で人の能力を超えていくことを意味しない。AIは特定の場面でのみ使える能力しかもたないからである。当時、世界最強の囲碁棋士であったイ・セドル九段を破ったAI「アルファ碁」は、囲碁以外は何もできない。一方、イ・セドルは、言語を自由に使うことはもちろん、料理もし、小説も読み、映画も楽しめる。AIの高性能化は、専用システムとしての高性能化にすぎない。そして高性能化とはうらはらに、(エ)AIの脆弱性^{ぜいじやく}と危険性が次第に指摘されるようになってきた。

現在のAIが得意とする分野の一つが画像認識（パターン認識）である。ディープラーニングとコンピュータの高速化が相まって、過去の膨大なデータを検索し比較することで、人が見逃してしまうような些細^{きさい}な画像の変化を迅速に検出できるようになった。しかし同時に、人の認識にはまったく影響しない些細なノイ

ズが混ざるだけで、非常に不可解な回答を出してしまうことが、しばしば報告されている。

たとえば、交通標識を認識する際、「止まれ」と書かれた標識の一部に小さなシールを貼るだけで、A Iは「時速45キロメートル制限」というまったく異なる標識と判断した。あるいは、道路を横切る人と、風で飛ばされてきたポリ袋を区別できなかったこともあるという。人はまちがえるとはいえ、これらはさすがにあり得ないミスであり、A Iによる自動運転の車など、まだとても怖くて乗れない。また、画像全体に非常に薄いノイズを重ねるだけで、まったく違う画像であると回答することもわかっている。パンダの画像に、人にはほとんどわからないようなノイズを重ねるだけで、雄羊であると回答し、全体の色調を変えると、今度はテディベアと回答した。これではA Iによる病理診断など、まだとても信用できない。

このようなミス、つまり誤認自体、大きな問題であるが、さらに大きな問題は、A Iがなぜ誤認したのかという理由が、なかなかわからないことである。計算方法もプログラムも人が開発したものであるから、どこに原因があるのかすぐにわかると思いがちであるが、そうではない。たしかにプログラムにより実行されるプロセスを丹念に追うことは可能である。しかし、コンピュータの性能の向上を最大限活用し、膨大なデータを膨大な計算で処理することができるようになった結果、そこで計算され処理される膨大な量の数字を人が追えなくなっており、どここの計算結果が悪かったのかわからないのである。そしてこのことは、A Iのミス、つまり、もとは人がつくったプログラムのミスを回避する方法もわからないことを意味する。

現在、インターネット上で共有されているプログラムを利用しリモートでA Iを騙す方法（アルゴリズム）がいろいろと考案されており、実際、多方面でA Iを混乱させている。このような、いわゆる「敵対的攻撃」は、これからもますます増えるであろうが、A Iのどこがどのように混乱したのかわからないことが多い以上、防衛は苦戦を強いられている。

これまで述べてきた事実からわかるように、A Iが認識する方法は、人が認識する方法とまったく異なっている。たしかにA Iの機能は向上したが、決して人に、つまり脳に近づいてはいない。そもそも、脳とはまったく違う人工的材料からつくられた構造物（コンピュータ）の中で動作し、脳とはまったく違う方法で

信号を伝達し処理しているのであるから、同じようなシステムであるはずがない。もちろん脳にはA Iに必要な起動ボタンやプログラム開始の命令も必要ない。構造、動作、働き方のすべての面でまったく違うものである以上、脳が生み出す心をA Iも生み出すと考えることには、当然、無理がある。A Iが自由意志をもつことはあり得ず、常に与えられた課題や目的に向かって動くだけである。

櫻井芳雄『まちがえる脳』（岩波新書、2023年）による。一部改変。

課題文③

コンピュータに心を持たせることができるだろうか。もっと具体的に、人工知能が心の機能を作り出し、心を持ったロボットが活躍する社会がくるだろうか。

人の心は、長年の進化の過程を経て生まれた。そこに神秘的なものは何もない。脳という物質の上に情報を乗せ、自らの機能を高めるように進化が進み、自然に発生した。だから、ニューロンという生物細胞ではなくて、シリコンで作られた情報機械の上に意識が生じて心が宿っても、それ自体は何の不思議もない。

意識は、多数の情報の統合の上に生ずる。これは、脳という超並列のコンピュータが環境を生き抜く必要上から発生した。ニューロンの動作のスピードは、コンピュータと比べると非常に遅い。ミリ秒のオーダーである。この遅い装置を使って素早い情報処理を行うためには、並列のシステムが好都合である。そこでは、多くの情報処理の課題を専用のサブシステムに任せ、ゾンビシステムとして無意識で素早く働かせる。

一方で、ゾンビシステムでは扱えない重要な案件を処理するのに、意識というシステムを作り上げた。ここでは多種類の情報が統合され、ポストディクション（後付け）の過程が機能する。これが「意識」である。意識は、実行前なら自己の決定を覆せる。また、これを反省の糧として、学習を進めることができる。

意識とは、自分がいま何をしようとしているかを自分で知っていることである。こう考えれば、コンピュータに意識を植え付けることは容易であろう。

コンピュータのプログラムが走っているときに、これは何を計算しているのか、その計算はどのような意義を持ち、どの程度重要なのか、いまどこまで計算が進

み、次はどこへ行くのか、などを理解し監視するプログラムを付加しておけばよい。これは後付けの機能のように見える。このプログラムが問題ありと判定すれば、元のプログラムに介入し、その決定を変えることができる。

コンピュータのスピードはナノ秒の世界であり、ニューロンの基本動作の100万倍も速い。もちろんこれを並列化して、単純な作業はゾンビシステムに任せることはできる。

しかし、このスピードならば、直列のままで時間をスライスして、時分割であるときにある仕事をし、次の微小時間でまた別の仕事をするということで、並列化しなくても十分に間に合う。

重要で困難な案件に対しては、コンピュータもこの課題に全能力を割り当てなければならない。とくに、いろいろな領域からの情報をもれなく用い、統合する必要がある。これがまさに「意識」の仕組みであった。これもコンピュータで行うことができよう。

人は、意識の機能をもとに心を持った。さらに、他人も自分と同じ心を持つことを知り、その動きを推測する。同時に他人も、私の心を推測して行動していることを知っている。

ロボットに、この心を持たせることができるだろうか。将来、人とロボットが共生する社会がくれば、ロボットには人の心の働きを理解して対応して欲しいと私は思う。

ロボットが人の心の働きを理解し、これをシミュレートすることは、そう難しいことではあるまい。そうなれば、ロボットも心を持ち、その命ずるところに従って行動しているように、我々には見えるかもしれない。人は容易に感情移入する。小説、映画、テレビ、ゲームでは主人公に感情移入し、その人のつもりになって応援する。そうならば、ロボットにも心があるように感じ、そこに共感することも当然あるだろう。

しかし、私たちの人生は一回限りの掛け替えのないものである。人は誕生し、成長し、そして亡くなる。一回限りの人生を生きる中に、喜びもあり悲しみもある。自分を律してよりよくしようと努力する。人は信念を持ち、自分に誇りを感じる。

ロボットにはこれがない。ロボットは、故障すれば修理が利くので死亡することもない。代替が可能で、取り替えても記憶を移植すれば同じことをする。一回

限りの掛け替えのない人生という感慨は生じないから、ロボットはしがらみに縛られた人間よりもはるかに合理的な判断を下すことができるかもしれない。

人間は進化の過程で、自己の生命の保持とともに種の保存に利する仕組みを内蔵した。これが発展して、人は文明を作るに至った。人間は多くの弱点を抱えている。これを理解したうえで、一回限りの個々人の人生が輝くことで、社会全体が輝くような文明社会が実現したらよい。残念なことに、現代文明は理想にはまだ遠い。

それでは、ロボットに目的意識を植え付け、ロボット集団が全体として進化するような仕組みを備えさせることはできるだろうか。ロボットが自らの心を持ち、社会を作り、進化することは原理の問題としては可能であろう。しかし私は、これが自動的に実現するには、人類進化と同じような長い年月が必要であると考え

る。

人間が、目的意識をもち進化するロボットを作るだろうか。もしそれが現実になれば、ロボットは人間の愚かさに愛想をつかし、人間に代わって自ら決定を下し、しまいには人間を駆逐するかもしれない。これがSFの描く世界である。

しかし、(※)このようなロボットを作るのはきわめて難事業で、私には事態がこのように進むとは思えない。「心を持つように見える」ロボットは生まれるだろう。「心を持っている」といってもよいかもしれない。ロボットが心と個性を持つように見えれば、人はこれに愛着を覚える。ロボットでも、状況に応じて喜びや悲しみを表現し、これを人と共有できる。だから、人はロボットを家族のように感じることもできるかもしれない。

でも、ロボットが喜びや悲しみを表現しても、これだけではロボット自身が喜び、また悲しんだとはいえない。喜びや悲しみの状況の認識は、喜ぶこと、悲しむこと自体とは違う。これはクオリア（質感、しみじみとした感覚）の問題といってもよく、個人の長い経験の蓄積の末に生じる。ロボット自身は一回限りの人生をいとおしみながら終えていくということはないのだから、すべての経験がそのまま役に立ち、クオリアのようなものが生ずる必要がない。人間が作るロボットは、あくまで人と協調し、人を助けるものに留まるだろう。

甘利俊一『脳・心・人工知能 数理で脳を解き明かす』（ブルーバックス、2016年）による。一部改変。

