



Faraday  
Papers

## 人工知能の未来と人間のアイデンティティ

ロザリンド・W・ピカード

### 要旨

人工知能 (AI) とは、ロボットやスマートソフトウェアエージェントなどの計算機技術が、人間なら知的とみなされるようなタスクを実行し、行動を示すことを指す。メディアではよく、AIは人間ができることすべてをこなし、それ以上のことをするように描かれることがある。本論では、次のことを吟味する。私たちがAIについて知っていることからして、これは現実的な描写だろうか？もし機械がそのような能力を達成することが可能だとしたら、そのような達成は人間が人間である意味を脅かすことになるのだろうか？AIを構築することで、人間のアイデンティティについて何を学ぶことができるのか？

### 序

AI製作者マックス<sup>1</sup>は何年もかけて、人間をモデルにした行動を選択し振る舞ってみせる、生きているようなキャラクター（性格をもっているかに見える人型ロボット）をシミュレーション

AIのキャラクターたちは、自分たちの住む世界について何も究極の知識を持たないデジタル・プログラムである。

（模倣的活動を行わせること）する知的なソフトウェア・エージェント（コンピュータの中で動作するプログラム）を作った。マックスのシム世界では、キャラクターたちが、実際には空腹や痛み、好き嫌いを感じず、何の経験もなしに、「お腹がすいた」「痛い!」「好き」といったフレーズを表情豊かに語るようにさせられる。

キャラクターたちは、あたかも情緒を持っているように見える表情を見せ、行動を選択する<sup>2</sup>。マックスが用意したスクリプトをもとに、あるキャラクターが「私たちには魂があると思いますか？」と問いかけ、別のキャラクターが「私たちがビット〔情報量の最小単位〕に過ぎないことは、「科学が証明しています」と答えることができる。キャラクターたちは、自分たちの住む世界について何も究極の知識を持たないデジタル・プログラムである。「科学は真理を見つける唯一の方法だ」などの近視眼的な宣言を、この虚偽を理解することなく発することができる。マックスは、この世界を主に自分の楽しみのために作った。

マックスのデザインした世界は、今日のAIにとって何が現実的であるかを示している。例えば、エージェント、つまりロボットは、「機械訓練」を用いた高度なアルゴリズムを実行して、対話を生成することができる。これらのアルゴリズムは、ソーシャルメディア、電子メール、字幕付き映画のせりふ、デ

<sup>1</sup> 本名ではないが、実在の人物とAIにインスパイアされている。

<sup>2</sup> 2 Picard, R. W. (1997). アフェクティブ・コンピューティング. MIT Press (2000年改訂版)。

デジタル化された本など、いろいろなものから人間の会話を使って学習し、その内容を微調整したり、混合しなおし、現実的に聞こえる言葉を作り上げる訓練をする。これらのアルゴリズムは、新しい発言を生成し、クリエイティブに聞こえるようにすることができる。例えば、あるキャラクターがフランシス・クリックのせりふを少し変えて、「『あなた』についての驚くべき仮説では、あなたの喜びも悲しみも、あなたの記憶も野心も、膨大なAIアルゴリズムの集合体の振る舞いに過ぎないのだ」と言うかもしれない<sup>3</sup>。

シム世界では、人間の意味がデジタル化される。AIが言葉を並べ替えて作る「創作」された会話は、外部に由来する意味を持つ<sup>4</sup>。もし、意味のないランダムな記号を使って会話にしたら、同じAIアルゴリズムによって、まるでわけの分からないものが生成されるだろう。見かけ上の模倣をするキャラクターは、学習データの情報源を反映して、超知的に聞こえたり、自信に満ちているふうに、あるいは堂々たるように聞こえたり、あるいは、ナイーブに聞こえたりするように作られている。

時には、アルゴリズムが数学的に自信のある回答を出せないこともある。そこでマックスは、そういう時には「それは興味深いですね」といった一般的なフレーズを支持するようにプログラムする。AIは、実際には何も知らず、何も考えず、自分の作り手と比べて少しでも多くどこか何も真に理解していないにもかかわらず、実際よりも高い能力に見えるようにプログラムされ得るのだ。

AI製作者が一般の人々に話をするとき、その言葉は（無意識のうちに）誤解を招くことがある。コンピュータが「知っている」、「学習する」、「考える」という言い方は比喩的表現であり、AIに適用されると異なる意味を持つ。私は、コンピュータのカメラに笑顔の画像が入力されると「笑顔」と出力し、笑顔の画像が入力されないと「笑顔なし」と

出力するように、コンピュータ内部で動作する数学関数を訓練できる。その後、新しい画像をコンピュータに見せて、正しく笑顔を認識できれば、「笑顔の認識を学習した」と言う。しかし、AIは実際には私たちのように学習しているわけではなく、私たちが見せた画像と私たちが与えた手順に基づいて、マッピング関数を適用しているのだ。カメラや入力を変えれば、完全に失敗するかもしれない。

比喩的に表現されるのとは違って、機械には「知識」も「知っている気分」もない。機械は、私たちが与えた関数とデータを持ち、私たちがプログラミングした手順で出力を得るが、経験も意識も感情もまったく持っておらず、その「知識」が働く文脈を教わることもほとんどない<sup>5</sup>。訓練された能力は、通常、データとその使い道を示された狭い範囲の一連のタスクに集中する。機械は、子どもには理解できない数学的関数を処理するようにプログラムされ得るが、子どものような一般的な学習者ではない。

シム世界では、「われわれの世界を超えた世界があると思うか？」というようなキャラクターの質問に対して、ランダムに答えを選ぶようにアルゴリズムを設定できる。たとえば、「いや、シム世界の外側に生命が存在すると信じるなんて非科学的だ」、または「たぶんね。ある得るよ」などの答えだ。あるキャラクターが「シム世界の外に存在する<造り主>を信じます」と提案することをマックスが許可するかもしれない。

「AI製作者のマックス」とこれらの例は、今日可能なことの紹介に過ぎないが、アルファベット（Google）、アマゾン、アップルなどの強力な企業が、人間のデータをマイニングし、人と「話す」、あるいは頻繁に対話する自動エージェントを作るのに何十億も投資しており、可能性は急速に広がっている。このテクノロジーに、人間が行うあらゆることを行うように教えることは出来るのだろうか。もし出来るとしたら、そのことは人間の

<sup>3</sup> 3 Crick, F., The Astonishing Hypothesis.より改変。魂の科学的探索。ニューヨーク

ニューヨーク。Charles Scribner's Sons, 1994より改変。

<sup>4</sup> 4 「言葉」が「初めに」存在したことは、読者も気づかないことではないだろう。

1:1. このシム世界の登場人物が知性をもって現れたり、行動したりするためには、意味と情報の初期源が存在しなければならぬ。

意味と情報の初期化源があるはずだ。

<sup>5</sup> These topics are cutting-edge research, and we can expect AI researchers to continue to make progress on them; however, there remains a gap between our metaphorical language and what is instantiated.

アイデンティティについてどのような意味をもつのか。

## AIの未来

AIの創始者であり、チューリング賞（コンピュータサイエンスでのノーベル賞に匹敵する）受賞者の故マービン・ミンスキーは私の親しい同僚だったが、コンピュータ科学者はいつの日か「人間の能力をはるかに超えるAIを作るだろう。そうなれば私たちは、AIに家庭用ペットとして飼ってもらえればラッキー、ということになるだろうね」とよく私に言っていた。ミンスキーは挑発するようにこう言った。しかし、彼が示唆したような、人間をペットとして飼う一般的なAIは可能なのか、もし可能なら、それはあり得るものなのか？

人間の知能とコンピューターやロボットの仕組みを理解している科学者たちは、人間のように意識を持ち、人間のように「知り」、「感じ」、「理解」できる人工人間を作る方法は、知りうる限り、現在も予測できる未来においても皆無であるという点でほぼ同意している。意識や感情の様々な面のシミュレーションを構築することは行われており、それを何層にも重ねることで、ますます本物らしくなるかもしれない。より優れたシミュレーションは、より優れた俳優のようなもので、ショーが続いている間は非常に印象的で説得力があり、私たちは俳優たちが実際、描かれた人物であるとさえ思うかもしれない<sup>6</sup>。しかし、正直なところ私たちは、俳優は演技をしており、AIはシミュレーション（見かけだけの模倣）をしているだけだと認める。

部外者の目には、私たちがテクノロジーの能力を高めるにつれてテクノロジーが「進化」しているように見えるが、AIは独立して自力で進化するわけではない。人間のデザインに依存しているのだ。ミンスキーの言う「人間をペットに飼うAI」が存在するためには、3つのことが必要だ。それは（1）人が創り出したいと願うこと、（2）制作可能で、その方法を私たちが考えだすこと、（3）資金的な裏付けがあること、である。以下、そ

<sup>6</sup> When William Shatner visited me, I wanted very much to call him "Captain Kirk".

れぞれについて考えてみよう。

**欲望**：知的には、そのようなAIを実現するために必要な進歩の解決策を想像することは魅力的であり、私を含む多くの科学者がその謎に惹かれている。しかし、賢い人々はますます「自分は人間のアイデンティティのベネディクト・アーノルドになりたいのか？」と考えるようになってきている。裏切り者とは、友人や国、主義主張を、たいていは金銭のために裏切る人のことで、ベネディクト・アーノルドは、英国から金を受け取るために米国を裏切ったことで悪名高い人物となった。彼の計画は発覚して失敗し、その後、アメリカ人だけでなく、報酬をくれると思っていたイギリス人からも憎まれた。歴史によれば、彼は惨めな人生を送り、死んでいった。「家庭用ペットとして飼われる幸運な人間」という新しい人間のアイデンティティを促進し、強制するAIを制作することによって、人類を裏切りたいと思う者がいるだろうか？この裏切り者は、人類からも、その研究に資金を提供した人からも、軽蔑される惨めな存在になる可能性が高い<sup>7</sup>。名声や富のために人類を裏切るとは、最低の行為である。

その代わりに、AI科学者が人間の生活を改善し、向上させる知的に挑戦的な問題の解決を望んだらどうだろうか。AIは、高齢者、障がい者、貧困層などに対して、正義、健康、平等を拡大する大きな可能性を持っている。たとえば、AIは、いつの日かほとんどの病気を予防するのに役立つと私は思う。私たちは、自分たちを家庭用ペットにするよりも、人類にとってより高い目標を願おうではないか。

**一般的なAI**：2017年4月、最も視聴されているアメリカの生放送テレビ番組、ジミー・ファロンが司会を務める「トゥナイト・ショー」の視聴者は、一般的なAIが到着したと思った。ソフィアという人型ロボットがファロンとおしゃべりし、「じゃんけん」をし、ジョークを言い、表情を作り、笑いが取れるから、自分が彼の番組の司会をした方が良いのではないかと、提案したのだ。ファロンは「でしゃばるなよ、お嬢さん」と言い返した。ソフィアの製作者であるデイヴィッド・

<sup>7</sup> AIがあたかも自分の製作者を嫌っているかのようにふるまうように作ることも可能だ。

ハンソン氏は、ソフィアは「基本的に生きている」と主張した。ソフィアは生きているように振る舞った。このパフォーマンスを見た後、私はデイヴィッドにお祝いを述べるとともに、どこまで台本にあったのか尋ねた。「全部台本で、リハーサルもした」との答えだった。

私たちがしばしばAIに使う格言に、「アヒルのように見え、アヒルのように歩き、アヒルのように鳴くなら、それはアヒルである」というのがある。この格言はおそらく、1730年代にジャック・ド・ヴォーカンソンが作った、偽物のアヒルに見える自動装置について最初に言われた。カナール・アーティフィシエル〔仏語で「人工アヒル」の意〕は、「生きたアヒルのさまざまな行動を模倣して、食べる、飲む、消化し排出する、羽などを広げる」のを観察された<sup>8</sup>。ソフィアは、「生きた人間のように見え、振る舞う」役割を果たすように作られた、現代のカナールである。

今日の人型ロボットやソフトウェア・エージェントは、工学の粋を極めた技を披露し、台本のないやりとりでもうまく機能することがある。特に、それらのロボットをよく準備しておいた話題については、そうである。しかし、おしゃべりを続けていると、やがて悲惨なミスを犯すか、「面白いですね」といった安全策に逃げてしまう。これは、機械の限界に見苦しくなく対処するために私たち製作者が追加した言い回しである。今のところ、私たちは意識や感情、本当の意味や理解をどのように構築すればよいのか、解決策を持っていない。現実の生活の多くの側面の見かけをまねることはできるが、誰かがドアを開けて中を覗いてみれば、そのロボットは、生きているニンジンが入っているかもしれない冷蔵庫よりもっと生命がない。

私たちが作製するものが、見かけの模倣を越えて、本当に生きているものになることは可能だろうか。私は、その可能性は低いと考えている。それでも、お気に入りの機械を「生きている」と宣言したい人たちの意欲を見たことはある。しかし、今日、そして近い

将来、ショータイムの後、ソフィアのプロセッサーの電源が切られ、彼女の感情豊かな顔が出荷のために箱詰めされ<sup>9</sup>、彼女の謝礼が人間の製作者に支払われても、私はそれに異議を唱える道徳的理由を見いだせない。

**お金**：何かの機能を果たすAIの製造と配備には、資金が必要である。天気やスポーツに関する答えを出すにせよ、製品を注文するにせよ、音楽を演奏するにせよ、治療のサポートをするにせよ—AIの設計、テスト、アップグレード、データの収集、データのマイニング、データに関する訴訟の処理、AIの作動の仕方の改善、返金などのために、誰かがAIの製造者に金を払う。ビジネスが資金を提供する場合、AIは収益を上げなければならない。そして自由主義国の政府や財団が資金を提供する場合には、納税者、知事、評議員などに対してその経費を正当化する必要がある。

「人間の言葉の謎を理解するチャットボットを作ろう」と夢見る好奇心旺盛な製作者が企業家に、「うちのチャットボットが顧客を喜ばせてるようにしてくれたら、数百万ドルさし上げる」と迫られる可能性は高い。顧客を喜ばせることは、通常、より多くの収益を生み出すからだ。

ロボット掃除機を売って大金持ちになったアイロボット社は、その何年も前に人型ロボットの赤ちゃん人形「マイ・リアル・ベイビー」を作っていた。その技術は見事だった—この人形は、食べ物をあげないと泣き、食べさせるとゲップをし、くすぐると笑い、揺らすと「眠り」、話すことを「学び」、さらにもっとたくさんのができた<sup>10</sup>。しかし、このAIは複雑で製造コストが高く、顧客が望む支払い額は低かった。AIベビーは、提案された喜び、証明された至福、そしてビジネス上の苦悩であった。もっとシンプルなAIロボット掃除機の方が高い利益を得ることができたのだ。AIは最終的に、請求書の支払い期限を迎え、制約のない空想から制約のある現実に移行する。AIがたどる未来の最も可能性の高い道を見るには、「人間の欲望を満たし、利益を増やすものは何か」に答えてみることだ。

<sup>8</sup> A book published in 1738 に出版されたある本は、「人口アヒル」がフランスの王立科学アカデミーの生活を模倣する様を紹介している。

<sup>9</sup> このことは、ソフィアを、サウジアラビアの市民権を持つ最初のロボットに選んだサウジアラビアの数名

の権威者によって議論されるかもしれない。しかし、これは一般にはとんでもないことに思われたようだ。

<sup>10</sup> <https://www.wired.com/2000/09/robobaby/>

## AIが人間のあらゆる能力を超えたら、人間のアイデンティティを脅かすことになるか

可能性が低いかどうかは別として、いつか私たちが「人間のように見え、人間のように歩き、人間のように話す」AIに直面したらどうだろうか。人と関係を築き、テストできるあらゆる状況で人間のように振る舞い、人間の思考、感情、自意識を持っているようにさえ見える。たとえこれらの属性が本物ではなく、印象的な外見の模倣であったとしても、多くの人はそのAIを生きていて本物であるかのように扱うだろう。これは、人間のアイデンティティを脅かすことになるのだろうか？

まず、今日の（人間ではない）AIが、人間のアイデンティティを脅かすものと相互作用する3つの場合を考えよう。

(1) 今年、何十万人もの人間が発作後の数分間に一人で死亡しているが、誰かがその場において直ちに応急処置を施すことができれば、その死亡率は寧くなる<sup>11</sup>。一方、今日のウェアラブルAIは、誰かがその場にいる可能性を高め、生命とアイデンティティに対する脅威を軽減できる。このようなAI<sup>12</sup>はまた、人の新しい仕事も生み出している。

(2) AIがだれかの仕事にとってかわり、その人が「機械にとってかわられた」という屈辱を味わう。置き換えられる機能は多くないとしても、ほとんどの人は有意義な仕事をする事から自分のアイデンティティの一部を得ている。AIにとってかわられた仕事の代わりに有意義な仕事の機会が提供されないならば<sup>13</sup>、そのときには、AIはアイデンティティを脅かし、自己価値、意味、目的を低下させ、本人やその家族に害を及ぼす可能性がある。

(3) 人々は、AIによって生み出される新しい仕事-新しい技術の設計、構築、サービス、技術革新、AIや倫理的な使い方の教育、その失敗に関する苦情の訴訟等々からアイデンティティに資する利益を受ける可能性がある。

<sup>11</sup> 死や傷害はもし、人々が命にかかわる発作に襲われたときに誰かが付いていけば減少する。Sveinsson, O., Andersson, T., Mattsson, P., Carlsson, S., & Tomson, T. (2020). Clinical risk factors in SUDEP: a nationwide population-based case-control study. *Neurology*, 94(4), e419-e429.

<sup>12</sup> Empatica Inc. Full disclosure: 本論の著者は、創立株主の一人である。

これら3種類の影響-脅威を減らす、脅威を増やす、人間のアイデンティティに利益をもたらす-は、AIにとって目新しいものではない。紙はかつて新奇な技術だった。正常な記憶力が普通かそれ以下の人々を助け、多くの新しい仕事を生み出した。しかし、天才的な記憶力があるために雇われていた人々のアイデンティティを減じた可能性も高い。紙は人間サイズの3Dモデルに成形され、写真に撮ると生きた人間のように見える。印象的な発明だが、もはや人間のアイデンティティに対する脅威ではない。

未来のAIを含む未来のテクノロジーに脅かされない人間としてのアイデンティティをもつことは可能だろうか？このことを位置づけるために、一つ、思考実験を試みよう。

この世界と人生において、あなたにとって重要なものをすべて失ったとする。愛する人、家族、友人、すべてを失う。次に、あなたのすべての所有物が、すべての財産や持ち物、家宝のすべてが破壊された、との知らせが届く。富も信用も、すべて奪われる。あなたは病気になる。痛みを苦しむ。身体も、心も、魂も。このような状態であなたのアイデンティティはどうなるか？

これを読んで、ヘブライ語聖書のヨブ記の冒頭部分を思い出す人もいるだろうし、現代に語るにすれば、極端な話にしすぎていると思う人もいるかもしれない。しかし現実には、ヒトラー、毛沢東、スターリンなどの専制君主によって一億人以上が検挙され、今日も新たな大量殺戮が行われている。そのような状況では、人は通常の源から意味やアイデンティティを得ることはできなくなる。社会的、物質的、経済的なすべての世俗的な源は破壊される。何が残るのか。人々は、もっとずっと深い井戸、つまり、核となる人間としてのアイデンティティに目を向ける<sup>14</sup>。

<sup>13</sup> 「最低限所得保証」は、意味のある仕事とそれがもたらす益の代わりにはならない。それは金よりずっと大切なものだからだ。

<sup>14</sup> 苦しんでいる人々の助けになる書は次の二冊である。Ten Boom, Corrie, Elizabeth Sherrill, and John Sherrill. *The Hiding Place*. Chosen Books, 2006; Frankl, Viktor E. *Man's Search for Meaning*. Simon and Schuster, 1985.

最高のアイデンティティとは、人が生きて  
いる間は、たとえ最も強力な指導者や軍隊で  
あっても破壊できないものである。この究極  
のアイデンティティの本質は、外部の人間存  
在がコントロールできる範囲を超えて、空間  
と時間の中に存在するものを超えた形で存在  
するに違いない。この要件を満たすアイデン  
ティティは、イマゴ・デイ、すなわち神の像  
に造られたことである。ただしこれは、通  
常、物理的な類似性を持つ像としては解釈さ  
れない。むしろ、この「像」とは、その源の  
何らかの側面を反映しているものと理解され  
る。例えば、学者で作家のJ.R.R.トールキン  
がホビットのビルボ・バギンズを創ったと  
き、彼はビルボに彼自身のアイデンティティ  
のいくつかの側面を与え、彼を作家とした  
が、それと同様である。

一つの深い主張は、すべての人間が、肉体的  
外見、知的能力、財産、業績、その他いかな  
る属性にも関係なく、イマゴ・デイを与え  
られているということだ。この地位は、生き  
ているすべての人に与えられ、世界中の人類  
を平等に結びつける偉大なものである。これ  
は無償の贈り物であり、各人にはその意味を  
実践する自由も無視する自由もある。

イマゴ・デイには現実的な意義がある。それ  
は、私たちが最も素晴らしい情緒的、行動  
的認知的、その他の領域において達成するも  
のを超えて、計り知れない価値を各人にもた  
らす。これは、超越的な価値であり値打ちで  
ある。「普通の人などというものはない。あ  
なたは一度として、単なる人間、などという  
ものに話しかけたことはないのだ」とC.S.ル  
イスは書いている<sup>15</sup>。

イマゴ・デイは、一部の人々が企てるよう  
な、「他の存在より優れていると自らを主張

する」ための口実ではない。この立場はしば  
しば環境世界の乱用や動物種の虐待と結び付  
けられるが、そのようなことはイマゴ・デイ  
を反映するどころか、イマゴ・デイに対する  
無知を示している。また、さも信心深そうに  
行動したり、インテレクトゥス・デイ（神の  
知性）を誇示するための口実でもない。その  
ような行動は、イマゴ・デイから逸脱した方  
法で行動する偽の自由を例示するだけだ。

実際、指導者が自分たちの支配を拡大・強  
化するために私的自由とAIの力を用いること  
は、今日世界が見ているように、市民を監視  
し、へつらいのない真実や望ましい道を表明  
することを制限している国家でなされている  
が、この場合にこれが人類にもたらす害は、  
将来、何らかのAIが独り立ちして作動するよ  
うになってもたらしうる害以上かもしれない。  
歴史が示しているように、権力に固執する  
独裁者たちはAI技術の革新的進歩よりずっ  
と前から、自己の権力を維持して党による支  
配を堅持するために、何百万、場合によっ  
ては何千万という国民の命を犠牲にしてきた。  
そのような指導者は、「国民の幸福を追求す  
る」、「より良い社会を実現する」などと、  
聞き手を鼓舞する目標をカリスマ的に語りな  
がら、国民のイマゴ・デイを否定する行動を  
とる。

重要なことは、イマゴ・デイは、強力な政  
府、軍隊、企業によっても、仕事や健康や機  
能の喪失によっても、取り去ることはできな  
い、ということだ。イマゴ・デイには、愛、  
慈悲、恵み、赦しという、変化をもたらす強  
力な力が宿っている。その源はこの世界の外  
にあり、空間と時間を超越し、宗教に関係な  
く、すべての人間に等しく与えられている<sup>16</sup>。

イマゴ・デイは、人間が値するいかなる特

<sup>15</sup> C.S. Lewis, *The Weight of Glory*, William Collins, 2013  
より。

<sup>16</sup> このアイデンティティは、ユダヤ教、キリスト  
教およびイスラム教のいくつかの伝承で、信徒だけ  
ではなくすべての人々に当てはまると認められている。  
神学者は、イマゴ・デイ説明する方法を様々に模索し  
てきた

例えば、Marc Cortez は *ReSourcing Theological  
Anthropology: A Constructive Account of Humanity in  
Light of Christ* (Grand Rapids, Michigan: Zondervan  
Academic, 2018), 109で、イマゴ・デイは「世界にお  
けるご自身の神的臨在を顕わす手段として人間たちを  
創造したことを宣言するもの」として見ることができ

ると示唆している。また、Middleton, J. Richard. *The  
liberating image: The imago Dei in Genesis 1*. Brazos  
Press, 2005も参照。イスラム教の思想家の中には、コ  
ーランの宣言「何も彼の似姿のようなものはない」  
が、預言者ムハンマドが、「神はアダムをご自分の像  
に造った」と言った早期の証拠と適合すると考えてい  
る (Yahya Michot: "The image of God in humanity  
from a Muslim perspective" in Norman Solomon,  
Richard Harries and Tim Winter (ed.): *Abraham's  
Children: Jews, Christians and Muslims in  
conversation* pp. 163–74. New York 2005, T&T  
Clark)。



権をも超えた驚くべき特権を無償で提供してくれる。話し言葉によってであれ、存在と知性という不可解な内的チャンネルによってであれ、その源と意思疎通し、関係を結ぶようにと招いてくれる。この経験は言葉では表現しきれないものだ。

イマゴ・デイはすべての人間を特徴づけるものであり、私たちが産み出す者の中に即座に存在する。彼らが能力や業績で私たちに劣っていようと、私たちを超えていようと、彼らは存在論的に<sup>17</sup>私たちと同等になるのだ。

功利主義の熱心な支持者である哲学者ピーター・シンガーは、人間の能力の中で、好みの表現に影響を与える一連の能力、例えば合理性、自律性、自己意識などを高く評価し、これらが完全に機能していない場合、人は人ではないと論じ、その人を殺害しても倫理的であり得る、特にそれが苦しみを最小限に抑え、影響を受ける人々の好みを満たすものとして正当化できるのであればなおさらだと主張した。いくつかの文脈で、彼はこれを、障害のある幼児や、アルツハイマー病や認知障害が進行した成人の殺害を道徳的に正当化する根拠として用いてきた。しかし彼の母親がアルツハイマー病の末期で倒れた時、何が起こったか？ここで、彼は自分の理論を脇に置いた。彼は自分の教えに相反する行動をとって、愛情をもって母親を看護した<sup>18</sup>。彼の行動はイマゴ・デイをはっきりと見せている。AIや人間のアイデンティティについて教えるとき、私たちは概念の世界にいて、何でも起こりうる。そこは制約のない空間で、「AIが人間に取って代わり、世界を引き継ぐことがありえる」と理論化することは容易だ。映画のように、実際的な制約が消えるような、ありえないデザインも許容できる。しかし、未来はひとりでにやってくるのではなく、私たちが未来を作り上げてゆくにしたがって到来

するのであり、そのとき私たちは自分たちの人間性によって制約されるようになる。概念は、身体化された心と魂のニーズと欲求に従うビットと原子になるのだ。制約のある最適化の結果は、制約のない最適化とは異なる。ロボットの赤ん坊は掃除機に取って代わられる。理論家はイマゴ・デイに直面する。

### 造り主に完全に知られたときの関係

最後に、ソフィアを箱に入れて出荷したり、マイ・リアル・ベビー<sup>19</sup>を廃棄することが、こうした言い方が非人道的に聞こえるとしても、なぜ問題ないのか考えてみたいと思う。

AIで作られたものは、存在論的に私たちと同等である人間とは異なり、非生物であり、存在論的には我々より劣っている<sup>20</sup>。これは、AIが私たちより機能が劣っているからではない。実際、AIは我々の多くよりも優れて、引き算し、チェスをし、誕生日を覚えているし、私たちがその機能を拡張し続けている。それが破壊されるとき、私たちが価値のあるものを失う。しかし、私たちが作った物の場合、それは、それを作った私たち自身に完全に知られており、私たちはもう一つ作ることもできるし、別のものをたくさん作る工程を組み立てることもできる。作り手である私たちは、存在論的にそれよりも優れているのだ。

それに対して、私たちが生み出すものは、存在論的には私たちと同等である。私たちの子どもとすべての人間は皆、イマゴ・デイを共有する唯一無二の存在だ。このことは、身体的・知的機能のレベル、人種、年齢、性別、宗教、その他、人間の属性に関係なく言える。健康な幼児や重度の障害を持ち、他者に依存している成人は、機能的にはより低いだろうが、イマゴ・デイは完全に持ってい

<sup>17</sup> 存在論は、存在の性質を扱う形而上学の一分野である。産むことと作ることのこの区別と、両者が同等ではないという見方は、Oliver O'Donovan, *Begotten or Made?* (Oxford: Oxford University Press, 1984) から来ている。また、ドイツ語から翻訳されたRobert Spaemann, *Persons: The Difference between 'Someone' and 'Something'* (Oxford, UK: Oxford University Press, 2006)も参照のこと。

<sup>18</sup> <http://www.michaelspecter.com/1999/09/the-dangerous-philosopher/>

<sup>19</sup> ロボットに、そうした瞬間に自分がされていることを知覚して、聞き手の心にひびく叫び声や痛ましい抗

議の録音を再生させるようにすれば、これはさらに痛ましく、あるいは、面白くなり得る。しかし、ロボットのプログラマーは、これが、指示を実行している心のないプロセッサに過ぎないことを知っている。その指示の一つの言葉を入力しなおすことで、叫びを「ありがとう！」に変え得るのだ。

<sup>20</sup> Oliver O'Donovan, *Begotten or Made?* (Oxford: Oxford University Press, 1984); Robert Spaemann, *Persons: The Difference between 'Someone' and 'Something'* (Oxford, UK: Oxford University Press, 2006)..

る。発作や麻酔で意識を失っても、私たちはイマゴ・デイを保持している。歴史的に言っても、人間の能力やその他の特徴に基づいて人間性や価値を決めようとするのは、人類に対する壊滅的な犯罪につながる。私たちの究極の価値は、実用性にはない。信仰の伝統は、一人ひとりの人間が生きた魂であることを教えてくれるが、この概念は科学では説明できない。私たちはお互いの中に生きている神秘を見いだす。他人のことを完全に知ることにはできない。また、人間の機能の一部は代替可能だが、人間そのものは代替不可能なのだ<sup>21</sup>。

先に、マックスのシム世界について説明したが、そこでは、すべてのキャラクターがマックスによって完全に知られている。キャラクターたちは私たちのようには意識がないが、マックスはそれぞれの「思考」を方向づけることができる。もし、マックスがどれかのキャラクターを気に入らなければ、マックスはそれを抹殺することができる。キャラクターは、存在論的にマックスより劣っている。もし、マックスがどれかのキャラクターと交流を楽しんでいるならば、マックスはもっとそれと交流し、自分が作る未来の世界に配置することさえできるかもしれない。それはイマゴ・マックスの側面を与えられ、マックスとのより多くの相互作用からより大きな価値を得るかもしれない。

「われわれはそれぞれの度合いで派生的様式で創造する。なぜならわれわれは造られた者だからであり、それも、造られただけではなく、神の像と似姿に造られているからである」<sup>22</sup>とJ.R.R. トールキンは書いた。マックスと神の類似性は完全には程遠く、多くのレベルで全く成り立たないが、私たちは、私たちが作るAIから人間のアイデンティティについ

て学ぶ。私たちが作るAIは、私たち自身の属性を反映し、私たちに完全に知られており、私たちと関係を築くために設計されているからだ。

AIにおいて私たちは、完璧とは程遠い自分自身の派生物を作りながらも、社会的ロボティクス<sup>23</sup>やリレーショナル・エージェント（訳注：使用者と社会的・情緒的關係性を持つことのできるAI）<sup>24</sup>の分野で明確に見られるように、AIとの関係を構築することを望んでいる。たとえAIが私たちと同じように共感できる言葉を発したとしても、その言葉は人間から発せられたものほどは大きなインパクトは持たない<sup>25</sup>。

人間のアイデンティティは、しばしば過小評価される關係性を伴っている。最後の思考実験を試みよう。

世界的な尊敬を集めるような、大きな貢献をした素晴らしい人物を想像みなさい。あなたは その名士と大きな集会で同席する。会場に入ると驚いたことに その名士はあなたのことを認識し、名前で挨拶し、あなたに思いやりを示す。驚くことに、あなたはこの名士に知られ、愛されているのだ。

あなたはどう感じるだろうか？多くの人は、喜びと自己価値の向上を感じるかもしれない。この人の地位が高ければ高いほど、また、その愛が本物であればあるほど、あなたのアイデンティティを高めることができる。

史上最も売れている本<sup>26</sup>の中で、私たちは、イマゴ・デイの与え主が、一人一人を、世界中のすべての人を知っているだけでなく、愛していることを学ぶ。

<sup>21</sup>このことは、誰かが金儲けのためにあなたに、もし多額の金を支払えば、あなたが死んだあと凍らせて、あなたを「バックアップ」した新しい形の命にするとする不確かな約束をするのを妨げはしない。

<sup>22</sup> J.R.R. Tolkien, *The Monsters and the Critics and Other Essays*, HarperCollins, 2007.

<sup>23</sup> Breazeal, Cynthia L. *Designing Sociable Robots*. MIT Press, 2002.

<sup>24</sup> Bickmore, T. W., & Picard, R. W. (2005). *Establishing and maintaining long-term human-computer relationships*.

ACM  
*Transactions on Computer-Human Interaction* (TOCHI), 12: 293-327.

<sup>25</sup> Morris, R. R., Kouddous, K., Kshirsagar, R., & Schueller, S. M. (2018). "Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions." *Journal of Medical Internet Research*, 20(6), e10148.

<sup>26</sup> 聖書のこと。多言語で無料で読める。



イマゴ・デイの源であり、与え主である方によって知られ、愛されることは、「特別でありたがっている」とか「例外主義」といった見当違いの非難を超えている。あなたが特別であることを示す証拠はたくさんある。この地位は、私たちが何を考え、何を望むかによって決まるものではない。

贈り主が存在するかどうかは、私たち次第ではない。存在は希望的観測で生じるものでも、不信仰で消滅するものでもない。そして、贈り物は私たちを他の誰に対しても、より優れた存在にしてくれるわけではない。むしろ、私たちすべてを等しくしてくれるのだ。

贈り主、つまり意味の創造者、そして私たちが何かを知る能力の創造者と関係を結ぶことは、人間の最大の機能が到達しうる何物をも超えた壮大さを達成する。それは、デジタルな形で表現できる言葉を超えた意識的な喜び、愛、臨在を引き出す。それは、私たちが作るファンタジーではなく、一人ひとりの実世界での経験によって実行され、テストされるものだ。

私たちが作る AI は、これらのどれも経験することがない。私たちがまだこれらを、AI に作成できるほどに理解していないからだ。将来、私たちは学習して、違う形に作れるかもしれない。だから、AI がこのようなことを経験することは起こるかもしれない。しかし、現在、私たちは部分的にしか見ておらず、私たちがこのように深く、証拠に基づいた方法で作るものすべてと異なっている。



ロザリンド・ピカード (Rosalind Picard) Sc.D., は MIT 教授、技術者、科学者、発明家で企業家でもある。彼女は、著書 *Affective Computing* で、機会に情緒のある知性のスキルを与える方法を説明し、デジタル

ヘルス、AI、感情を扱うコンピューターの分野で 350 以上の査読付き論文を執筆している。Affectiva と Empatica の二社の創設者のひとりであり、Affectiva 社には感情のツール (Smart EyeAB に使用) を、Empatica 社には発作をモニターする、初の FDA 認可を受けたスマートウォッチを提供している。彼女は、MIT のキャンパス全体の福利を扱う MindHandHeart の創設所長である。技術者としての最高の名誉である全米技術アカデミーの会員に選ばれている。

#### 謝辞

Nathan Barczi, Karthik Dinakar, Kevin Ford, Ian Hutchinson, Anne McCants, Madars Virza, そして 2 名の匿名の査読者からいただいた有益な意見と考察に感謝する。

(本論は、Rosalind W. Picard, “The Future of Artificial Intelligence and Human Identity,” *Faraday Paper* 24, 2021 年 11 月の本多峰子による全訳である)

ファラデー論集(**The Faraday Papers**)「ファラデー論集」はファラデー科学・宗教研究所 (Faraday Institute for Science and Religion) を出版者とする。当研究所は St Edmund's College, Cambridge, CB3 0BN, UK, に本部を置く教育と研究のための慈善団体 ([www.faraday-institute.org](http://www.faraday-institute.org)) である。また、本論文集の日本語訳は本多峰子による。「ファラデー論集」で表明された意見は各著者の意見であり、必ずしも本研究所の意見を代弁しているとは限らない。「ファラデー論集」は、科学と宗教の相互作用に関する幅広い論題に取り組んでいる。現在出版されている「ファラデー論集」のリストは [www.faraday-institute.org](http://www.faraday-institute.org) で閲覧可能であり、そこから、PDF ファイルでダウンロード出来る。

2022 年 10 月 © The Faraday Institute for Science and Religion

