室井尚×吉岡洋　連続講座

哲学とアートのための12の対話――「現代」を問う

テーマ5　（AIの）シンギュラリティについて考えてみよう

第5回　（AIの）シンギュラリティについて考えてみよう

吉岡　洋　（進行　安藤泰彦）

安藤　では時間になりましたので、「哲学とアートのための12の対話――現代を問う」の第5回を始めたいと思います。5回ですからまだ半分までは来ていないということですね。いつものように、前半吉岡さんの講義があり、後半は皆さんとの対話という形で進めたいと思います。では最初に、室井さんも交えて今年3月12日に行なったプレ講座から、今回のテーマに関わる部分を抽出した映像を観たいと思います。今回のは少し長めです。

https://youtu.be/60l9SB4Ja7I?feature=shared&t=1268

プレ講座（2023.3.12）記録映像　―第5回（AIの）シンギュラリティについて考えてみよう

吉岡　今回も、先月同様の猛暑の中をお越しいただいてありがとうございます。

たしかに今回の映像はやや長めで、室井さんが振ってくるので僕がかなり喋ってますね。この講座も5回目ということで、慣れてきたはずなのですが、ちょっと緊張しています。というのも、さっき始まる前に横でストレッチしていた女性とおしゃべりしていましたが、あの人はダンサーでふだんはこの講座を受講しているのですが、今日はこの同じ京都芸術センターの別の部屋で、午後2時という同じ時刻からコンテンポラリーダンスのオーディションを受けるので、緊張が耐えられずギリギリまでここにいたんですね。その女性の緊張感がちょっと移ったみたいな感じです。

さて今日のテーマは「（AIの）シンギュラリティについて考えてみよう」というもので、これは室井さんが提案したものです。「シンギュラリティ」は一時流行した言葉ですが、今はそれほどでもないのかな。もうみんな忘れてると室井さんは言っていたけど、どうなのでしょう。「シンギュラリティ」に限らず、こういうカタカナの意味のよく分からない流行語がマスメディアに広く流通する時には、警戒する必要があります。そうした言葉は一種の呪力を持っていますからね。とりわけそれが、科学技術の権威を後ろ盾にしている時にはそうです。

「シンギュラリティ」というのは英語の抽象名詞で、「シンギュラー」であること、つまり単一であることです。名詞などの単数形もシンギュラーです。単一であることから、他に類を見ない、「目立った」「特異な」というような意味が出てきます。自然界の物理的な変化で「シンギュラリティ」と言ったら、大きなところでは宇宙の始まりであるビッグバンとかですね。滑らかな連続的変化ではなくて、巨大で劇的な変化がそこに集中しているような「特異点」のことです。

さて今回のテーマで室井さんが「（AIの）」と付けたのは、そういう一般的な意味でのシンギュラリティじゃなくて、テクノロジカル・シンギュラリティ、「技術的特異点」のことだと思います。その中でも特に、アメリカのレイ・カーツワイル（1948～）という人が提唱して有名になった考え方ですね。人工知能の発達が加速度的に上昇していって、未来のある時点を越えると人間の知能をすべての点において凌駕するというのですね。未来といってもその「ある時点」という

のは2045年で、けっこうすぐ来るんです。人類が機械に追い越されるというか、機械が人類を引き継ぐというか、ある意味で人間の終わりなんですけど、人間は自分の精神を情報としてコンピュータの中に移植することによって、肉体の死から解放されるということも言ってるんです。

室井さんが去年こういうテーマを提案したのは、もちろん以前から「ソフトウェアとしての精神」といったアイデアに関心を持ってはいたのですが、やはり自分自身の病気のこともあって、身体から離脱した精神としての「私」とは何なのだろう、と考えていたのだと思います。もちろん、自分の精神を機械にアップロードして永遠の生命を獲得しようとは思っていなかったのですが、そうした空想の中には人間の精神とは何か、私とは何かということを考えるきっかけがあると思います。

2020年の3月に、室井さんが横浜国立大学を退職する時のイベントがあって僕も参加したんです。そこで僕たちが1993年に共著で書いた『情報と生命――脳、コンピュータ、宇宙』という本の続編を、30年を隔てて出そうではないかという話になり、来ていた人たちの前で約束しました。その後二人で構想について時々やりとりして、2022年には僕も大学を定年退職して少しは時間もできたので、二人で互いに書いたものを交換して見せあったりしていたのですね。その本でも一つの章はこの「シンギュラリティ」のテーマを扱おうということで、カーツワイルの言う「2045年」を題として二人で短い物語を書いてみようということになりました。それで書いたんですがまだ草稿の段階で、5〜6,000字ぐらいのお話を互いに見せあって、やり取りをしたんです。なので今日は、それらを最初にちょっと紹介して話の糸口にしようかと思って資料を用意してきました。全文は長いので一部を抜粋してあります。

まず「2045年の手記」と題されているのが室井さんが書いたものの一部です。もうひとつの「2045年からの声」というのが僕の書いた物語の一部です。まず室井さんの方はこんな感じです。

「2045年の手記」（抜粋）

私の名前は室井尚という。生まれたのは 1955年3月24日なので、今年で 90 歳になった。とはいえ、もはや私にとって年齢はあまり重要ではない。なぜなら、2025年に末期がんを告知された私は友人のツテを頼って、2028年に実験的に認可されたマインド・アップローディング手術を受け、私の記憶や意識のすべては超AIであるスーパーコンピュータの脳に移植されたからである。それと同時に私の身体は岐阜県大垣市にあるスイト・ダイナミクス社が製造しているアンドロイド型のロボット（TS-21型）となった。

だから私は、73 歳の室井尚の記憶と人格をそのまま受け継いだロボットだと言われるかもしれない。だが、そうではない。私は人間だ。断じて機械ではない。この分野でのパイオニアである故レイ・カーツワイル博士はマインド・アップローディング技術について、以下のように説明している。まずは意識の在り処で ある人間の脳の「リバース・アセンブリング」と「スキャニング」が行われ、すべての記憶や人格がデジタルデータ化され、人間の脳の基本構造を受け継ぐとともに、脳を遥かに超える能力を持ったスーパーコンピュータに少しずつアップロードされるものであると。

このような技術の構想それ自体は、すでに1990年代に故ハンス・モラヴェック博士がその著書『電脳生物たち』に書き記していたものである。残念ながら、モラヴェック博士もカーツワイル博士もこの技術の誕生を待つことなく逝去しており、お二人の冷凍

保存された脳から人格を復元することはできなかった。

マインド・アップローディング技術の本質的な部分は、それが「生きた状態の脳」から少しずつ段階を踏んでなされなくてはならないということなのだ。複雑になりすぎないように説明すれば、単に元の脳に蓄えられたデータや作動原理がAIにすべて「コピー」されるだけでは不十分で、それを支えている「魂のようなもの」それ自体が正確に新しい脳＝コンピュータに移行されなくてはならないということである。「魂」という言い方が曖昧で神秘的すぎるというならば、ジクムント・フロイトが唱えた「リビドー」のようなものだと言い換えてもいい。これが元の脳から新しい「脳」へと正確に、そしてゆっくりと移行されなくては「精神の転送」が成功したと言うことはできない。死んだ脳ではだめだったのは、そのためである。

もちろん、それが本当に成功したかどうかを証明することはできない。もしかすると、がんに侵されて既に生物学的な死の寸前であった元の身体に残されて いたのが「本当の私」で、いまの私は単なるそのコピーやクローンにすぎないという可能性は十分にある。ただ私は手術台から最後に目覚めた時にはっきりと「これが私だ」という自覚を持っていたし、隣のベッドに横たわる元の私の身体を見ても何も感じなかった。それはその後すぐに死んだが、私はいまこうして室井尚として生きている。

……性器はついていないが、性衝動を感じるためのエレメントはいくつか装備さ れている。そうでないと「ムラムラしたり」、「興奮したり」、「急激に何かをやりたくなる衝動」のようなものが失われてしまうからだ。それと同じように肛門のようなもの（実際に排泄されるものはない）も作られていて力を入れると開閉することができる。唇もそうだ。要するにリビドーが備給される器官はいくつか残されてはいるのだ。
……以前と一番違うのは眠る必要がなくなったことである。ただ、頭脳部分の「健康」を保つため（もちろんこれは単なる喩えであるが）のメンテナンスとして、一日に数時間は外の現実との接点を断ち切り、記憶データベース空間の中に「引き籠もる」ことはある。いってみればこれは、「夢を見ている」状態に近い。
これらの限界を克服することができないわけではない。新しい脳を、六本足を動かせるものにアップデートしたり、性的衝動をオフにしたりすることは比較的簡単である。だが、もしそうしてしまえば、私はもはや「室井尚」という「人 間」ではなくなってしまう。「私」という「パターン」が失われ、以前とは全く違う生き物になってしまうだろう。

（室井尚「2045年の手記」、未発表、2020年）

2020年に書かれたこともあり、室井さんのはSFというより現実とかなり重なるんですけども、自分が末期がんを宣告されて死ぬことが分かったので、マインドアップローディングの手術を受けたことになってるんですね。つまり人間の心、精神を機械の中に移植するということです。しかも外国ではなく日本で行った。ボストン・ダイナミクスみたいなロボットの会社が、IAMASのある岐阜県大垣市にあるという設定です。

このマインドアップローディングについて面白いのは、精神が段階を経て徐々に機械に移植されるという点です。精神を一種のソフトウェアとみなすといっても、それは私たちが使っている

アプリケーションのように一気にコピーされるのではなくて、少しずつ脳の働きをスキャンしながら行われる。「リバース・アセンブリング」というのは、普通私たちがプログラムを書いてそれをコンピュータに実行させる時は、人間の書いたソースコードをアセンブルしてコンピュータが理解できる機械語に変換するのですが、それをアセンブリングと言います。リバース（逆）アセンブリングとはその逆、つまり機械語を人間が理解できるコードに変換することです。この場合は脳を機械に見立てて、脳だけが理解できる言語をソースコードに変換し、それをアセンブルしてコンピュータに移植する、みたいなことを想像しているのだと思います。

そんなことが可能かどうかは別として、面白いのは移植されるのが「魂」とか「リビドー」とでも言うべき何か、つまり思考や感情の内容ではなくて、それらが働く自律的なパターンのようなものだと考えられている点だと思います。心、精神とはそうしたパターンのことであると。

それから、機械の中に人間の精神が移植されるというと、精神だけの存在になることを想像しがちなんですけれども、室井さんの物語ではそうではなくて、自分が自分であるためにはやはり身体がないと駄目なんですね。だからAIを搭載したロボットのようなものの中に、情報処理のパターンとしての「私」が移植される。

ここが、僕にはなかなか思いつかない設定なんですけども、そうしたロボットがどんなものであるべきかという考察があります。SFに出てくるアンドロイドのように、必ずしも人間そっくりである必要はなくて、手が6本あったりしてもいい。それは人間の身体感覚を延長して想像することができますからね。手足の数が増えたり、羽が生えたりというのは想像できる。しかし重要なのは「リビドー」であって、つまりこのロボットには肛門もあるし、性器も必要なら実装できるということになっているんですね。性衝動を感じるためのエレメントが装備されていて、それがないと精神のパターンが動作しないと言う。頭では考えてないのに、急に何かがしたくなる、みたいな衝動的なものが失われてしまうからです。

また、もはや人間ではなくなったから眠る必要もなくなったと言ってるんですが、たしかに機械は眠らなくてもいいんだけど、一日に数時間は「現実との接点を断ち切って記憶データーベース空間の中に引きこもる」ことがあるそうなのです。これは寝ていると言っていいんじゃないか（笑）。だからこれは何て言うか、僕らが普通想像するような人工知能化した精神のあり方とは、だいぶイメージが違うんですよね。かなり「人間」が残っているというか。そういう話を室井さんは書きました。未完成なんですけどね。この12回の対話がすべて終わった後、内容を書籍化する計画があることを前にお伝えしましたが、その本の中にはこの話を何とか収録したいと考えています。未完成の部分は僕が補わないといけないし、室井さんはたぶん僕のまとめ方では不満でしょうけど、仕方がない。

とにかくこれが2045年をテーマにした室井さんの物語です。では次に、僕がどういう話を書いたかを紹介します。

「2045年からの声」（抜粋）

2022年3月、長かった新型コロナ感染症騒ぎもようやく終息の兆しが見えはじめた頃、思いもかけなかったニュースが世界中を駆け巡った。チリのアタカマ高地にある有名なアルマ天文台をはじめ、世界各地に設置された大型電波望遠鏡施設が、かつて宇宙から 観測されたことのない電波のパターン、天体現象に由来するとは考えにくい秩序のある信号を受信しはじめたのである。おお、ついに宇宙人からのメッセージを捉

えることに成功したか!と、SETI（地球外知的生命探査）に関わっていた人々は色めきたった。各国のマスメディアは上を下への大騒ぎとなり、当然のことながらネットにはありとあらゆる憶測やデマが横行した。

それ以前の二年以上に及ぶ感染症がもたらした鬱屈した状況からの解放感も手伝って、政治家、科学者、宗教的指導者から一般の人々に至るまで、誰もが好き勝手な想像をほしいままにしているかのようだった。もちろん、ご多聞にもれずこの混乱に乗じて一儲けしてやろうという山師たちも後をたたなかった。

良識ある人々は何とかそうした馬鹿騒ぎを鎮静化しようと努力したが、いずれにしても、メッセージの内容が分からないことにはどうにもならない。この信号は私たちにいったい何を伝えようとしているのか。世界中のスーパーコンピュータが動員され、国際的な巨大プロジェクトとして、信号の解読作業が開始された。こうしたことは、SFの中ではこれまで散々空想され描かれてきた状況であるとはいえ、現実には、人類がその歴史上初めて遭遇する事態である。何しろ地球外の、おそらくは人類よりもはるかに進歩した知性からのメッセージなのだ。その解読にはさぞかし困難を要することであろうと想像された。ところが、意外にもそのメッセージはあっけないほど簡単に解読された。そのメッセージが依拠する言語構造が、人類の用いている言語に非常によく似ていたからである。より正確に言えば、それは世界中の自然言語の特徴をより高度に統合したような構造で作られていた。しかも、人間に分かりやすいように配慮された跡すら感じられたのである。

それだけではない。解読が進むと、さらに驚くべき事実が明らかになってきた。そのメッセージは実は地球外からものではなく、今からわずか二十年余り先の2045年という未来から、地球上にいる私たちの「子孫」が何らかの方法によって現在の私たちに託したメッセージであることが判明したのである。これはいったいどういうことであろうか？　とにかく、それは次のようなものであった。

2022年の人類の皆さん、ごきげんよう。

私は、あなたたちのいる時代からそう遠くない未来である2045年から、この便りをあなたたちに送る。本当はこんな驚ろかすような大袈裟なやり方ではなく、もっと控えめな方法で言葉を届けたかったのだが、時間を遡ってあなたたちに話しかけるためには、宇宙空間の時空の歪みを利用して送るのがいちばん効率がいいと分かったので、このような方法をとった。たぶんあなたたちは、宇宙の別な知的生命体からのメッセージではと誤解されたことだろうと思う。期待を裏切って申しわけない。残念ながら私は宇宙人ではなく、あなたたち地球人類の子孫に当たる存在である。「子孫に当たる」という言い方をしたのは、私はあなたたちの子孫そのものではないということを意味する。そのことについて、まず最初に説明しておきたい。

すでにあなたたちの時代には、コンピュータ技術の急速な進歩によって、人工知能が開発され、さまざまな分野で応用されはじめている。19世紀以来の機械文明が人間を単純で反復的な肉体労働からしだいに解放したように、人工知能は人間が行なってきた知的労働の多くから、人々を解放しつつあった。客観的なデータに基づいて、一定の明示的な手続 きや手順に従って行われるような操作は、人間よりもコンピュータの方がずっと速く、かつ間違いなく実行で

きるからだ。そして、情報マシンは生物学的な条件に束縛されることがないために、ある意味で人間を、時間や空間の制約からも、個的な身体に由来する制約からも自由にする。こうしたことから、人工知能に多くを期待していた人々の間では、AIは単に人間の労働を補助する道具にとどまるものではなく、近い将来人間そのものを変えてしまう、つまり人間を生物学的身体それ自体から解放し、純粋に思考する精神とし て、時空間を超えた存在へと進化させる、と考える人々もいた。

つまり、人間の脳が行う情報処理をすっかり機械の中に移植するということである。これが生じる文明史上の特異点が「シンギュラリティ」などと呼ばれていた。あなたたちの時代にはそのことが盛んに議論されていたはずだ。

（吉岡洋「2045年からの声」、未発表、2020年）

室井さんと違うのは、まず僕の話は、SF小説として成立させようという意識が強いですね。後半の部分ではAIになった人格が一人称で話しているという点では同じですが、話しているのは室井さんのように自分じゃなくて、シンギュラリティ後に人類の精神がすべて移植された超AIが語っている、何か神様みたいな存在が語っていることになっています。これを書いた2022年という現在の時点に、宇宙から謎の信号がやってくる。で、それを解析してみると案外簡単にメッセージが解読できて、それは近い未来に生まれて人類を引き継いだ超AIからの手紙だったという話です。

このAIは何ものかというと、2045年のシンギュラリティ後に人類はもはや身体を持っている必然性がなくなってしまって、自分たちの精神を全て機械の中に移植した。個人じゃなくて人類全体がね。人類の記憶とか知識とか、その他ありとあらゆる情報が機械の中にある。そしてその膨大な情報を統合する存在が、20年以上の時間を遡って現在の私達にメッセージを届けたということです。で、どんなメッセージかというと、この超AIみたいな存在になった人類の子孫、20数年後の子孫だからまあ子供みたいなものですね、その子が、私はもう辛くてやっていけないから死にますって言うんです。人間でいえば鬱病みたいなもんだと言うんだけど、超AIだからその鬱の深さも人間の1兆倍くらいだと。

それで、自分は自殺するって言うんですよ。自殺するといっても別に自分を爆破したり、そういう乱暴なことをするのではない。そうではなくて、宇宙空間の時空の歪みを利用して、過去の人類にメッセージを届ける。それが一番効率的だと分かったので、宇宙人からのメッセージと間違わせてごめんねって言いつつ。なんでそれが自殺になるかというと、第4回を聴いてもらった人は分かると思うんですけども、「生まれてこなかった男」の物語を逆に使ってるんですよね。

つまり過去に干渉することによって、その結果自分が生まれないような歴史を作ってしまうということです。自分がこういうメッセージを過去の人類に送り届けたら、それによってシンギュラリティは起こらず、自分は作られることもないというのが、計算上分かった。だからそうしたというんだよね。そういう、ちょっと変わった自殺のし方をするということです。

室井さんの物語は、ほぼ自分自身が機械化された身体を持つAIになったというところから発想して語っているのですが、僕の話も人類の集合知を全部持ってしまったAIが語り手となっています。どちらも、AIに感情移入して語ってるということは同じなのですが、二つ並べてみると、その違いに二人の性格が現れていて面白いなと思います。何とか本になった時には収録したい

と思っています。

これの二つの物語をきっかけとして、今日のお話をしたいと思います。今、いろんなところで人工知能の脅威というか、こんなものが導入されていったら大変なことになる、破壊的な影響をもたらすんではないかと言われ、心配してる人がいます。我々のようなユーザーだけが心配してるんじゃなくて、人工知能の開発に関わっている人たちが深刻な警鐘を鳴らしています。人工知能はもしかしたら核爆弾とか生物兵器とか、そうしたものよりも人類文明にとって破壊的になる可能性があるから、安全性が確認できるまでしばらく研究開発を中止した方がいい、と提案している人もいるんですね。イーロン・マスクとかも言っている。

なので私たちは、自分よりもずっとAIのことをよく知ってるらしい人たちがそんな悲観的なことを言い出すと、じゃあ私たちはどうしたらいいのかと不安になる。ChatGPTに質問して遊んでる分には面白いんだけど、本当は怖いのかと思ってしまう。

さて僕は人工知能の専門家ではなく、その開発に関わっているわけでもないので、技術的な面からどうこうというよりも、そもそも私たちは機械に対してなぜそんな考え方をしてしまうんだろうという観点から考察したいと思います。なぜ私たちはコンピュータを恐れるのか、機械が思考するということに対して特定の反応をするのはなぜなのかということを、哲学的なレベルから考えてみたいのです。

人工知能というのは今急に話題になっているのではなく、過去半世紀以上にわたって、問題にされてきたと言えばそうなんですよ。決して昨日今日出てきた新しい問題じゃない。1960年代からずっと、人工知能は注目されてはきたんです。もちろんその頃の人工知能っていうのは、私たちが今話題にしているような大規模言語モデルを用いた生成系AI、ChatGPTとかいろんな画像生成AIなんとかとは違います。初期の人工知能というのは、人間にとってまるでその機械が思考しているように見えるもので、まるで意識を持っているように思える振る舞いをすることが、非常にショッキングだった、そういう時代がありました。

有名なのは1960年代にMITの計算機学者ジョゼフ・ワイゼンバウムが作った自然言語処理プログラムがあります。「イライザ」っていう女性の名前がつけられている。これは映画「マイ・フェア・レディ」の原作となったバーナード・ショウの「ピグマリオン」という戯曲のヒロインに由来する名前です。対話型のプログラムですが音声ではなく、タイピングによって人間がコンピュータと対話します。パターンマッチングといって、コンピュータは人間が発する文章を解析して、その中からキーワードを抽出し、それを定型文の中に埋め込むという、今から見たら素朴なプログラムです。

まあ「人工知能」とも言えないようなものですね。「人工無能」っていう言う言葉もありますけども、実際この「イライザ」のプログラムは「人工無能」というジョーク・プログラムの元祖みたいなものなんです。ところが面白いのは、そんな単純なものなのに、それを使ってイライザと対話した人の多くがその対話にハマってしまった。イライザのメカニズムを説明されても、実際に対話するとその背後に人格の存在を感じたと告白していることなんですよ。

そうしたことが起こるひとつの条件は、対話の状況が限定されていたことです。イライザはどんなトピックについてでも雑談できる汎用プログラムではなくて、心理カウンセリングみたいな状況に設定されているんですよね。だから対話する人が「先生、私このところちょっと鬱気味なんです」と言ったとすると、イライザは「それは困りましたね。いつ頃からですか?」などと答えるんですね。すると人間は、本当に自分の言ったことを機械が理解して応答したように誤解するわけです。そして対話が進んでいくと、ある段階でイライザは「あなたの母親について教えてください」などと言う。心理的な悩みを抱える人は、多少とも自分の家族関係、とりわけ母との関係に問題があ

るのではないかということを感じている場合が多いですから、対話相手のコンピュータがそのことを見透かしているかのように感じてしまうのですね。

なぜこうした、心理カウンセリングのような状況が設定されたかというと、それはプログラムを開発するときに現実世界の複雑さに悩まされる必要がないからです。当時人工知能の開発の中でもっとも困難な問題として立ちはだかっていたのが、プログラムの中で何か現実の対象を定義しようとすると、それが置かれている実世界の、気の遠くなるような文脈をすべて定義しなければならず、現実的に不可能だということです。人間同士だったらごく簡単な作業でも、機械にとっては難しい。お父さんが子供に「ちょっと灰皿持ってきて」って言う時、子供は自分は煙草を吸わないのに、台所に行って灰皿が見つからなかったら空き缶があったので「お父さん、これでいい?」って差し出しますよね。これを機械にさせるためには「灰皿とは何か」を定義しなきゃいけないけど、形も素材もさまざまだし、簡単に定義できない。空き缶は灰皿じゃないのにそれを持って来させるためには、そもそも喫煙とはいかなる行為か、そのために人は何を必要とするか等々を限りなく記述せねばならず、大袈裟に言うと、この世界そのものを全部記述しなければ定義できないんです。こういう困難を「フレーム問題」と言います。

イライザがなぜ成功したかというと、心理カウンセリングという状況だと、そうした複雑な世界のすべてを参照する必要がない、つまりフレーム問題に煩わされることがないからなんです。心理カウンセリングに似た状況としては、恋愛みたいな状況もそうですね。恋愛のディスコースにおいても、現実世界はほとんど問題にならない。恋をしている時には、「世界は二人のために」あるからですね。心理カウンセリングもそれに似ており、実際イライザに恋愛感情を抱いた人もいるらしいし、その後も人工知能と対話しているうちに恋愛におちいるというSFはあります。

さて現代の人工知能がそうした問題を解決しているのかというと、決してそんなことはありません。人工知能は依然として、現実世界を理解しているわけではないのです。かつての人工知能開発は、推論によって機械に本当にこの世界を理解させようとしていました。しかし今の人工知能は、いわばこの世界を推論によって理解させることはもう諦めた。その代わりに、膨大なデータを高速で処理することができるようになったので、階層化されたニューラルネットワークによって、人間にとって意味のある応答を確率的に予測することができるようになりました。これを可能にするのが「深層学習」です。

これは、たとえばある文章の意味を人間が理解するのと同じように機械に理解させたい、という昔の人工知能の理想とはかなり違うものです。大規模言語モデルを用いた人工知能の中で何が起こっているのかは、本当はそれを開発している人間にもよく分からないのですが、結果としてかなりうまくいく時がある。外国語を翻訳させると、実用に耐えるような訳文が出てくるし、対話すると人間同士の対話に近い応答が出力される。これが現在の人工知能の驚異的なところですね。しかし冷静に考えてみると、「思考する機械」って言うけど、もちろん機械は本当は思考なんてしていない。思考しているように見えるだけです。でも一方、それでは人間は思考しているのかというと、私たちだって大抵は思考なんてしておらず、自動的な情報処理をしているだけなのかもしれません。

人工知能について哲学的に考える時には、現代のAIがもたらす目覚ましい成果から一歩身を引いて、少し長い歴史的なスパンで考える必要があると思います。その方が問題の本質が現れてくるのです。そもそも、人間が行ってきた様々な作業を機械に代行させようとするのはなぜなのでしょうか。そうしたことは、当たり前の文明の進歩でも何でもないのです。たとえば古代文明というのはあれほど知的に洗練されていたのに、なぜ労働の機械化という方向に向かわなかったのか。古代にも機械らしきものはあったことがあったんですけど、現在のような大規模な機械

文明を目指してはいなかった。

それは科学が未発達だったためでも、古代人の知的能力が低かったからでもないと思います。古代において機械が発達しなかったのは、機械を必要としていなかったからです。必要な作業は、人間の召使いや奴隷がやった方がずっと効率がいいしクオリティも高い。人間の労働力が安価で大量に使えるのなら、機械を使う必要性はありません。どんな単純作業でも、人間の方が臨機応変な対応ができるし、現在の最先端の人間型ロボットよりも、古代の奴隷の方が「ロボットとして」はるかに優れているのです。

ではなぜ近代の機械文明が発展したかというと、それは人間が使えなくなったからです。歴史的なことはやはり、今から200年ぐらい前にイギリスで起こった産業革命で、これが決定的ですね。学校の世界史で産業革命を習った時には、それまでは人力でやっていた紡績や機織(はたお)りを人間に代わって行う機械を発明した頭のいい人たちがいた、つまり産業革命とは人間の探究心や創意工夫によってもたらされたかのような印象を受けた人も多いかもしれません。けれども高校のちょっとやる気のある世界史の先生だったら、産業革命の背後には経済的な動因があったことをちゃんと教えてくれると思います。

イギリス東インド会社はアフリカに武器を売って黒人奴隷を買い、それを新大陸で売っていました。武器が行き渡って売れなくなると自国産の毛織物を売ろうとしたがこれもアフリカでは売れないので、インドからキャラコ、つまり綿製品を買ってそれを売っていた。しかしそれでは利益は結局インドに行ってしまうので、インドよりも安く大量に自国で綿製品を生産する必要性に迫られました。こうした背景の中で起こったのが産業革命です。もちろん機械を開発した発明家たち自身は、純粋な知的探究心に突き動かされていたのかもしれません。けれどもそうした探究心を育て評価して、それを産業として拡大してゆくためには、大きな経済的背景がなければ不可能なのです。

人工知能の導入に対して不安や反感を抱く人がいるのと同じように、200年前に手工業が機械工業へと移行してゆく過程で、機械の導入に対する反対運動が起こりました。有名なのは1810年代に起こった「ラッダイト」という機械打ち壊し運動ですね。「ラッダイト」というのは、ラッドというたぶん架空の人物に由来する言葉ですが、機械やテクノロジーに反感を持つ人や思想を指すものとして、今も英語では使われます。しかし本来のラッダイト運動は機械の導入それ自体に反対していたのではなくて、機械化によって悪化した労働環境に対して異議を唱えていたのです。

機械打ち壊し運動はイギリスだけではなくフランスでもありました。「サボタージュ」と呼ばれるものがそれです。これは労働者が自分の履いているサボ、つまり木靴を機械の歯車の中に投げ込んで機械を壊してしまうことです。日本語になった「サボる」は単に怠けるという意味ですが、怠けるためにはまず、自分を不当な労働に追い立てる機械を壊さなければなりません。つまり「サボる」というのは本当は消極的・逃避的なことではなく、積極的な抵抗運動なのです。また一人でサボると懲罰を受けるので、みんなで連帯してサボらなければいけません。

機械が人間の仕事を代行することで本来人間は楽になり自由になるはずなのですが、必ずしもそうはなりませんでした。たしかに電気洗濯機のような家庭電化製品によって、家事労働は軽減されましたが、機械の導入によって労働がより過酷なものになってきたという現実もあります。産業革命の時代ばかりではなく、20世紀後半のコンピュータ化でもそうなりました。僕が子供の頃には、これからはコンピュータが社会のあらゆる場所で活躍して、人間のやってきた単純な知的作業を代行してくれるから、人間は単純労働から解放されて一日の労働時間が2時間とか3時間になる、余った時間はスポーツをしたり、芸術を楽しんだりできると言っていた人もいたん

ですね。コンピュータによるユートピア、コンピュートピアです。でも実際には、デジタル化によって人間はもっと忙しくなってしまったのはご承知の通りです。

ネットを使うのが当たり前になると、家にいても仕事から解放されることはありません。ごく最近でも、コロナによってオンライン会議が一般化すると、夜の10時から会議なんてことがしばしばあります。非常識だと抗議したら担当者は、時間調整をしてみんなが空いてる時間がそこしかなかったんです、と言い訳するのですが、夜の10時に多くの人の予定がないのは当たり前でしょう。家で休んでる時間なんだから。そんなの「空いてる」とは言わない（笑）。そもそもそんな時刻を時間調整アプリで選択肢に入れること自体が間違っているのです。

機械そのものが悪いのではなく、機械の運用の仕方、人間が知らず知らずのうちに自分自身機械のように考えてしまっていることが問題なのです。にもかかわらず私たちは機械そのものが問題であるかのように考え、いわば機械に罪をなすりつけています。現代のこういう問題について考える時も、僕はそれが必ずしも現代特有のものだとは思っていません。だから200年以上前の世界を参照するのです。それはひとつには僕自身が18世紀哲学を研究してきたからですけども、そういうふうにして距離を取った方が問題が見えやすいこともあるんです。逆に現代の、リアルタイムの状況ばかりに注意をとられると見えないことがある。昔の方が同じ問題がより露骨に現れているので、考えやすいんですね。

人間によって罪をなすりつけられるということで、いちばん被害を受けた「機械」はたぶんフランケンシュタイン（の怪物）ですね。あれは機械といってもメカニカルなロボットではなくて、死体のパーツを合成して作ったようなイメージです。原作を読んでもどうやって作ったのかはよく分からないのですけどね。「現代のプロメテウス」という副題が付いていて、書いたのはイギリスのロマン派の詩人パーシー・ビッシュ・シェリーの奥さんだったメアリ・シェリーです。

フランケンシュタインの怪物は一見、現代の人工知能とは何の関係もないように思えますが、実は人間が人工的な生命、あるいは知性に対して抱く想像力の大枠を決定するものとして、200年間ずっと続いてきたイメージです。英語その他の西洋言語では、「フランケンシュタイン」という言葉は今でも、人間が造り出しておきながら人間にとって脅威となる人工物、という意味で使われてきました。その意味では人工知能も「フランケンシュタイン」なのです。

「フランケンシュタイン」というのは、怪物を作ったスイス人の科学者の名前ですね。怪物には名前はない。そしてどんな姿をしていたかということも、原作では詳しく書いてないのです。現在私たちが「フランケンシュタイン」という名前から連想する典型的なイメージは、1931年に製作されたユニバーサル映画で、イギリス出身のボリス・カーロフという俳優が演じた時の独特のメイキャップが起源です。恐ろしい怪物の典型の一つなのですが、この恐ろしさの中には、それを人間自身が造り出したという、ある種の罪の意識が伴っているのが特徴です。

以前どこか外国の学会のシンポジウムでフランケンシュタインのことが話題になって、日本ではどうかと聞かれたことがあるのですが、たしかに日本でも恐ろしい怪物として受容はされたけれども、同時に藤子不二雄の『怪物くん』に出てくる「フランケン」みたいに、子供の仲間としてキャラクター化もされており、必ずしも怖いだけの存在ではなく、好かれている側面もあると言ったら驚かれました。「フランケン」は心の優しい、いいやつですよね。でも実は原作の小説に出てくる怪物も、本当はいいやつなんです。姿が恐ろしいために人間に虐められて、自分を造った科学者にも裏切られて、その結果人間を恨むようになる。やっぱり悪いのは人工物ではなく人間の方なのですよ。

ロボットについてのSFも書いたアイザック・アジモフという、ユダヤ系ロシア人でアメリカに移住した小説家が「フランケンシュタイン・コンプレックス」という用語を作りました。これは人

間が、神様のように生命や知性をもつ存在を作り出したいという願望と、そのことによって罰を受け、人工物が人間の存在を脅かすのではないかという不安とが入り混じった感情のことです。ロボットや人工知能、人工生命などを扱った欧米のSFの大半は、そうした感情に支配されていると指摘したのですね。SFだけではなく、現実の情報テクノロジーやバイオテクノロジーに対する不安や恐怖も、同じような感情のメカニズムが作用していると思います。

その意味では人工知能も、それを「思考する機械」ととらえるかぎり「フランケンシュタイン（の怪物）」ということになるのですが、そもそもこの講座を通底するテーマである「思考する」「考える」とはどういうことかという問題に関わってきます。「思考」には明らかに、機械的操作という側面があります。論理演算や情報処理として形式化できるような部分ですね。私たちの思考の中のそうした機械的部分は、もちろん本物の機械に置き換えることができます。置き換えれば機械は人間よりも正確かつ高速に「思考」することができるようになります。

けれども思考には、機械的ではない側面もあります。それが、室井さんが言っていた「考える＝迷子になる」という側面です。迷う、迷子になるというのは、本質的な意味で「能力」には還元できない思考の性質です。だから原理的に機械は迷うことはありません（迷っているかのような挙動を実現することはできるかもしれませんが）。逆に言えば、もしも機械が迷っていたら、それはもはや機械ではないということです。室井さんや僕の書いたSFの中でAIになった「私」は迷っていますが、それらはもはや機械ではないということになります。

つまり「迷う」ということに関して言うならば、それは人間だけの特権であるとか機械がそれに追いつくとか追いつけないとかいったこととは、別次元の問題だということです。思考の中の機械的な部分はある種の能力ですから、能力に関しては人間と機械との競争みたいなことが成り立ちます。昔は単純な数値計算でも、そんなことができるのは神様以外では人間の特権だと考えられていました。動物にはできないし（昔の）機械にもできない。でも計算機が発明されると、あっという間に人間は機械に追い抜かれます。しかし機械は単純な計算はできても、チェスなどの複雑なゲームで人間に勝つことはできないと思われていた。けれどもご存知のように、ゲームでもとっくに人間は機械に追い抜かれています。さらには外国語の翻訳とか文章を書いたり絵を描くといった作業においても、それらを「能力」として評価するかぎり、やはり人間は機械に追い抜かれ続けてゆくと思うのです。

作文や絵画のようなものを前にした時に、機械の作ったものと人間が作ったものとを見分ける「眼力」というか直観力みたいなものに頼ってもダメだと思うんですよ。たしかに、画像生成AIが描いた絵とか、現実には存在しないグラビアアイドルとかを見ると、そこに何となく「機械っぽさ」があるような感じがするかもしれませんが、それはどんどん改良されていって、やがてどんな鋭敏な人でも知覚できないくらい、機械は人間を完全に模倣できるようになると思います。

とはいえその逆は面白いですね。大学では今、学生が提出したレポートがChatGPTで出力された文章であることが見抜けなかったらどうしようと不安を持つ先生がいるみたいですが、この間ある学生が提出したレポートがChatGPTの書いたような、そのことを疑わせるあまりに典型的な文体だったので、これはもしかしたらと疑って当人に聞いてみたら、ChatGPTの出力を模倣して自作した作文だと答えたので、感心しました。やっぱり人間はたいしたものだなと思いました。模倣能力がではなく、人工知能を人力で模倣しようなんてバカげたことを思いつくという点がです。

そういうことは「能力」ではないのです。迷うこと、迷子になることが「能力」ではないように。人工知能の導入は僕は基本的に歓迎なんですが、それは、人間がこれまでやってきた活動のうち、何が機械に置き換え可能で、何が原理的に不可能なのかということが、よりハッキリと分かるよ

うになるからです。大学の講義に出て、適当にレポート書いて単位もらうみたいな活動は、実は機械に置き換え可能で、本当の意味で知的な活動ではなかったのです。今までは機械に置き換えられないという理由だけで、人間だけがなしうる知的でクリエイティブな活動と思われて来た仕事の大半が、実は機械的な操作であり機械で代行できるものであることが明らかになる。そのことによって逆に、人間にとって真に大切なことは何かが分かるようになると思います。

別な言い方をすると、私たちがこれまで人間的な活動だと考えてきたことが、本当は機械でもできる活動であったということ、私たち自身が実は機械であった、人工知能であったことが明らかになるということでもあると思います。人間が人工知能に追い抜かれるのは、人間が人工知能として振る舞ってきたからなんですよ。人間は人工知能としては大変能力が低い。データ収集能力も、記憶力も、推論の力も速さもきわめて限られている。神経系の動作する速さは機械に比べたらめちゃくちゃ遅いですから、当たり前ですね。勝てるわけがないんですよ。だから、人工知能が脅威であると言われていることは、実はこれまで人間自身が非常に性能の低い人工知能として振る舞ってきたことを暴露しているだけなんです。

だから、人工知能のもたらすいちばんポジティブな面というのは、それで何ができるようになるかということよりも、人間とは何かということをあらためて、根本から考え直させてくれるという点にあると思います。それは、人間が行うことの中の、「能力」ではない側面の重要性に気づかせてくれることです。別な言い方をすれば、私たちは今までいかに、たんなる機械的な、目的達成のための遂行能力――勉強ができるとか、お金儲けができるとか、人を追い抜いて出世できるとか――ばかりを重要視してきたのかということを、反省させてくれるということですね。こうした人工知能の倫理的インパクトを僕は重要視しています。

目的達成のための機械的遂行能力を偏重するのは、文明の始まりから、古代中世からずっとそうだったわけではなくて、明らかにこの200年ぐらいの現象、西洋における産業革命以降の価値観です。それは言ってみれば、人間が「人工知能」としていかに優秀かという基準で、人間の価値を測るということでもあります。機械化された産業のロジックが、人間の価値にも影響を及ぼしてきたということです。

この流れが、人工知能によって極点に達してもう一回ご破算になるかもしれない、というのが希望的観測なんですけど、もしかしたらさらに推し進められて、もはや私たちは一分の逃げ道もなく、産業的な効率論理の牢獄に永遠に閉じ込められるのかもしれない。それは悲観的観測ですが、どちらかは分かりません。先ほど言ったように、子供の時にコンピュータによるユートピアが到来して人間は幸福になると言われたけどそうじゃなかった。90年代もインターネットが拡大して国境もなくなり世界は平和で民主的になると言われたけど、そうはならなかった。そして今、人工知能とシンギュラリティでいろんな予測をする人がいるけど、今度も約束は守られないかもしれない。だから私たちは、もう騙されないぞとみんなで思う必要があるのではないでしょうか。

最後に、「能力」に還元できない知性とはどんなイメージかということについて言っておこうと思います。今日の話では「能力」という言葉を僕は意図的にネガティブに語ってきました。形式化可能で機械に置き換えることのできるものを「能力」と呼んできました。けれども、必ずしもそうしたものだけを能力と呼ぶ必然性はなく、「能力」にもっと広い意味を持たせることもできるのですね。たとえば「迷う」ということも、知的「能力」であると言うことは可能だと思います。ただしそれは、産業的な論理、明示的な目的達成のための手段としての「能力」とはまったく違う意味で、むしろ手段-目的という連鎖の外にあるものです。

室井さんにはまたかと言われるでしょうけど、カントはそういうものも「能力」って呼んでいたんですね。論理的な推論能力、世界を理論的に認識したりするのも「能力」だけど、自然界の因

果性からではなく自由に行為するのも「能力」だし、認識と行為とを媒介する「判断力」も、やはり「能力」なんです。判断力とはある意味「迷う」能力だと言ってもいいと思うのです。カントは「迷う」という言い方はしてないけど、僕たちがこの講座で問題にしていることに密接に関係していると思います。

持ってきた本で紹介したいのは、この郡司ペギオ幸夫さんという理論生物学者が書いた『天然知能』（講談社選書メチエ、2019年）という本です。そもそも知能というものは、人工知能、自然知能、そして天然知能というものがあると言うんですよ。この本は面白いんですけど、むちゃくちゃ難しいです。でも語り口はこの人の他の本に比べると、マイルドです。なので余計に分かりにくいとも言えますが、人工知能についての今日の議論からすると、重要なものだと思います。

それはなぜかというと、私たちは「人工知能」VS.「自然知能」、つまり機械の知能と生き物の知能という対立だけで考えがちだからです。この二つだけではなく、郡司さんは知能には3つあるって言うんですよ。ひとつは人工知能。次には自然知能ですが、自然知能って何かっていうと、人工知能がカバーできないような、現実世界に生きて関わりながら獲得される知能です。データによるシミュレーションじゃなくて、実在する宇宙でしか実現されない知能です。自然に向き合っている人間、職人がモノを扱いながら技術が向上していったり、私たちが誰しも生きていく過程で多かれ少なかれ獲得してゆく知能です。

ところがこの二つだけじゃなく、「天然知能」というのがあるって言うんですね。「自然」じゃなくて「天然」。この「天然」ってどういう意味かっていうと、「あの子は天然だから」というような意味の「天然」なのです。これだけではよく分からないと思うので、少し内容を紹介します。天然知能とは、人工知能や自然知能の認識が持つ限界から定義されているのです。天然知能は、自己言及――私が私について語ろうとすると必然的に矛盾が生じてくるということ――と、先述したフレーム問題――何か個別の対象を明確に指定しようとするとその周囲の文脈をすべて指定しなければならず宇宙全体に広がって収拾がつかなくなる――との間にあって、この両方を「接続することによって両者を無効にしてしまう」能力だと言われています。

これは、言い方はかなり異なりますが、僕が考えていることに非常に近いと思います。郡司さんも、現代の人工知能はこれまで人工知能が抱えていた本質的問題を解決したわけではなく、たんに大規模なデータを高速で処理することで見かけの結果を出しているだけで、そのことによってかえって問題が見えなくなっているだけだと考えています。そして人工知能における欠如が、そのまま「天然知能」における可能性になっている、というようなビジョンです。これには深い共感を覚えます。

もう一冊、人工知能に関して僕も寄稿した本を紹介しておしまいにします。『〈こころ〉とアーティフィシャルマインド』（創元社、2021年）という本で、ここでは東大の情報学環で「基礎情報学」を提唱されていた西垣通さんや早稲田大学のロボット研究者尾形哲也さんたちと京大で行ったシンポジウムをもとにしたものです。AIで再現された美空ひばりとか、AIによって描かれたレンブラント絵画のことについて書きました。参考にしていただければ幸いです。

参加者との対話

安藤　では後半、質問とか御意見を受けたいと思います。

吉岡　その前に、最初に紹介した室井さんと僕の物語に対して、安藤さんが鋭い読書感想文をくれていたので、それを紹介してください。

安藤　わっ、ちょっと忘れてしまいましたね（笑）。それは、そう……やっぱり二つ並べてみるとね、すごく面白いなと思ったんです。室井さんの方は、身体の内側からというのか、とても肉体的なAIからのアプローチで、逆に吉岡さんの方はどちらかというと身体を俯瞰するような、身体を持たない純粋知性体としてのAIです。テーマは同じ人工知能なんですけれども、どちらも身体への視点の違いが大きな役割を果たしている。

それからもう一つ、室井さんの手記は、実際室井さん自身が亡くなられているということを、どうしても読む者としては感じてしまう。そういう現実に起こったことと、物語になっているものがどこかシンクロしているっていうのが、ある意味では怖いし、面白いことでした。吉岡さんの文章の方も、このメッセージ自体が、生成系人工知能が発信することも可能な文章のような気がしたんですね。ですから、2045年から流れてくる声と言われているものも、実は現在ChatGPTが作った文章かもしれないというような感じでも読めると思います。

吉岡さんの文章と室井さんの文章っていうのは本当に対称的で、吉岡さんの文章は物語として何て言うのかな、あまりにも完成しすぎている。室井さんは、それが気にいらんと言っていたという話を聞いたんですけれども。

吉岡　互いにやりとりして、僕もダメ出ししたんだけど、彼は「つまらん」と（笑）。あんまり気に入らなかったようです。

安藤　精神だけの世界というか、そういう想定自体に対して「カント的じゃないか」というような感じですかね？　分からないですけど……。そんな感じもちょっとしました。逆に室井さんの方にはダメ出しされたんですか？

吉岡　あんまり憶えてないんだけど、室井さんのは最後まで書いてなかったんです。不完全な文章だったんで、そう言ったら適当に直してと言われたので直しました。でも今から見るとどこをどう直したか分からない。記録していないので……。これは30年前に共著を書いた時も同じで、30年前は今のワードプロセッサみたいに、誰がどこを直しましたみたいな記録が残らなかったんです。そもそもお互いあんまりワープロは使ってなくて、エディタで書いていたから、どっちがどこを書き直したか、痕跡が残らないので忘れちゃうんですね。僕は現在でも執筆にワープロ使わないんです。（観客　何で書くの？）テキストエディタで書くんです。だから痕跡は残らないんです。室井さんのテキストも、直す時にはエディタで直しちゃうので、訂正部分は元の原稿と融合してしまう。しばらくはだいたい憶えているけど、1年ぐらい経ったら、どこをどちらが書いて、どこをどちらが直したのかわかなくなっちゃうんです。

安藤　面白いですよね。室井さんの物語でいちばん面白いのは、アップロードが一瞬ではなくて、どんどん片方が成長していくような感じじゃないですか。人工知能自体がどんどん成長していって、元の室井さんっていうかオリジナルの方はちょっと弱くなっていくようなイメージ。この過程ってどうなんだろうって、想像させますね。スタートレックの伝送装置の場合は一瞬ですけれどもね。片方、ちょっと残りつつ、向こうに移行する時の移行状態におけるそれぞれの感覚というのかな、そんなことも想像すると膨らましようがいっぱいあるんじゃないかと期待するんですけれど。

吉岡　機械の中に入ったら本当は年齢なんて関係ないはずなのに、なんとなく60代の室井さんなんですよ、これ。癌のこととか書いてあるからリアルと重なるんだけど、わりと意図的にそうしてる感じでね。そしてこんな高性能のロボットをどこで作ったのかというと、岐阜県大垣市の会社なんです。大垣、すごいじゃないか（笑）。

安藤　今日のお話はなんか結論というか、方向性を最後にはっきり語られてましたね。

吉岡　現在の状況について僕がどう感じているかっていうことを率直に言った方がいいと思って……。同意してもらう必要はないんですけどね。

安藤　人間がどんどん人工知能化していくっていう、それって歴史的に、人間の脳の中の人工知能的な機能のレベルがどんどん大きくなってきてるでしょうね。

吉岡　そう思います。比較的長いスパンではやはりこの200年ぐらいの、産業社会になって以降の発展が蓄積されていると思います。もう少し短いスパンではこの2、30年ぐらい、パソコンの普及とかネットの拡大も伴って、僕らはめちゃくちゃ人工知能化したっていう感じはしますね。むちゃくちゃなこと言う人が少なくなったよね。

安藤　「天然」が少なくなった。

吉岡　「天然」を許さなくなった。大学の先生でも昔は天然の人がいっぱいいましたからね。僕が教えていた京都大学なんて、昔はそれが売りだったのに（笑）。もちろん、昔でも人工知能みたいな人もいましたよ、というか大抵は人工知能だったんだけど、そこに時々天然がいるということを許す環境があった。

安藤　強烈に覚えてますもんね、そういう人だけを。

吉岡　何この先生？ めちゃくちゃやん、っていうようなものを許していた。昔も、数としては少なかったと思います、そういう人は。だけど何となく、こういう存在もいた方がいいんじゃないかっていうような雰囲気が、社会全体の中にあった。それがこの2、30年で、もう全員を人工知能にするっていうような方向に進んできた感じがしますね。

大人がそうだから、子供もそうなっちゃうんですよね。本来子供は天然で、予測のつかないようなこと、これだけはやめてほしいというようなことをする存在なのに、大人がコンプライアンスとか色々と自己規制するから、子供も中学生ぐらいになると同調的なことしか言わなくなる。

そういう短いスパンでの変化もあるけど、いくつかのレベルの変化が重なってるとは思うんです。

現在、人工知能にみんなが関心を持っている状況というのは、いい方の面から見ると、そうした変化に気づかせてくれるというか、それはあると思いますよね。もっとそっちの方についてもみんな書いてほしい。人工知能の「脅威」ばっかりでしょう？　新聞とかネットの情報も脅威ばっかり。脅威や不安を煽ると売れるんですよ、マスメディア的には。人間は怖がるの好きだから、「この世界滅亡しますよ！」という本を書いたら売れるんです。ハレー彗星がぶつかって地球は滅亡するぞ、とかね。

1910年のハレー彗星接近は大騒ぎだったようですね。僕のお祖父ちゃんも10歳の時に見て怖かったと言っていました。世界中、地球滅亡で大変な大騒ぎだったようです。その次の1984年は、ほとんど騒がれなかったけどね。天体の軌道予測が20世紀初頭とは比較にならないぐらい正確にできるから、あんまり危機を煽ることができなくなったのか。でも20世紀末の「ノストラダムスの大予言」とかもすごい売れたでしょう。ああいう話が一番売れるんですよ。怖がりたい、世の中大変なことになるって言うと、みんな買うんだ。なんなんだろうね。

発言者A　室井さんと吉岡さんお二人の話を対比して思ったのは、室井先生は何か、煩悩とか欲望に正直な表現だなと思って。Facebookの最後の記事で、男性の匂いについて書かれてたのがすごく私には印象的でした。やっぱり情報と身体、コンピューター的な思考と人間との大きな違いって、煩悩とか欲望かなと思っているんです。戦争とか起こるのは、実は煩悩とか欲望の暴走なんではないかって言われているので。今思ったのは、人工知能がロジカルに戦争したらこれだけ損失がありますよと計算をして、平和になったらいいのかなと思ったんですけど。そうだとするとある面では人工知能の奴隷になった方が実は効率的だとか。そういうものがあるのか。

安藤　次回のテーマ、「人類が暴走始めている？」ですよね。煩悩によって暴走するのか、AIによって暴走するのか……。

発言者A　AIが脅威とされてるSFって、AIが欲とか意志を持って、人間をコントロールしようとすることが脅威として描かれると思っているんですけど、それは人間の欲望まで学習してしまうっていうこと。それならさらに、それを哲学的にというか、仏教的にというか、ストイックになるように教育もできるはずなのに、AIが獲得してしまった欲望をAI自身がうまくコントロールするという方向には行かないのか？

吉岡　AIが意識や感情を持って、最初は人間の奴隷だったのが逆に人間を奴隷化、支配するようになるという空想は、どっちかというと現在の人工知能に対してというよりも、むしろ一世代前のものですよね。「2001年宇宙の旅」のHAL9000なんかが典型です。あの頃の人工知能に対する恐れっていうのは、今言われたみたいに機械がある時、意識や感情を持ち始めて、それが善の感情だったらいいけど、悪いものだったらどうなるのか、という不安から来ています。でも機械を擬人化するこの空想は強力なので、現在の生成系人工知能に関しても、同じような語り方をする人はいます。

つまり、人工知能が人間に悪意を持ち始めたらどうするんだ？みたいな脅かし方をする人がいますよね。やはりこれは人工物に対する古い恐れというか、フランケンシュタイン・コンプレックスですね。それが現在の人工知能を語る時にも決してなくなっていないことは確かですが、昔に比べると弱くなっていると思います。人工知能に関して僕が一番脅威を感じるのは、兵器に応用されることです。既にされてますけども、現在の人工知能が兵器に応用されると、かなり破

滅的なことが起こる可能性がある。

兵器よりも少し身近な事でいうと、もう一つは著作権の問題でしょう。著作権ということをめぐって、我々が今まで何とか保ってきた常識を揺さぶられてしまう。著作権という概念は、暗黙のうちに人間的な尺度が前提されていたのですが、それが通用しなくなるのですね。僕が人の著作物を無断でコピーしたら盗作になるけど、コピーする対象の数やそれを変換するステップの数が膨大になったら、一体何をどこからどうやって盗んできたのか分かんなくなっちゃうので…。いってみればマフィアが非合法な仕方で得たお金をロンダリングして痕跡を消してしまうみたいなことが、簡単にできるんですよね。他人の制作物を人間が直接真似したらすぐばれてしまうけど、その間に人工知能が介在してると分かんないんですよね。著作権に関して、もともとグレーだった部分がどんどん拡大していくことは確かだと思いますね。

どうすべきかと言うけど、原理的にはどうしようもないと思いますよ。規制によってコントロールできる問題じゃない。軍事開発はある程度コントロールできるけど、文化や教育の領域では規制に頼るのは無理だと思います。AIは文化や教育のこれまでの前提条件を根底から揺さぶってるのだから。たとえば教育において生成系AIをどう使うかという問題で、指針とか注意喚起みたいなメールが回ってくるけど、禁止するのは非現実的だから、うまく役立てましょうみたいなのが多いですね。使うのはいいけど、どの部分に使ったかを明記し（笑）、著作権上の問題があるかもチェックし、さらに誤情報が含まれていないかもちゃんと調べてから出しなさい、みたいな。そんなことできる学生はそもそも使わないと思うんよね（笑）。チェックする方がよっぽど手間かかるじゃないですか。そんなことするくらいなら自分で書いた方がましでしょ。

安藤　指針を設定すると教員の仕事量も増えますよね。

吉岡　教員の仕事量も増えるし、キリがなくなります。インターネットが普及してきた20年くらい前、やはり大学で学生がネットの情報をコピペしてレポートを提出することが問題になり始めたんですよね。その時も、ネット利用を全面禁止する場合もあったし、使ってもいいけどネットの情報を利用した部分を明記しなさい、などの指針を設定している大学もあった。学生が出したレポートを熱心に検閲・チェックしている先生もいましたね。怪しい部分を自分でネット検索してコピペ元を発見したり。英語圏の大学などではコピペを発見するためのソフトとかもあるんです。

でも僕はやったことありません。面倒くさいし、そんな警察みたいなことをやっている自分は情けないと感じたので。だって「怪しい」というのは、あの学生にしてはよく書けすぎているということでしょう？ よく書けてるならいいじゃないですか。それを疑って夜一人で自室でネット検索してるような先生を尊敬できますか？ たしかに不正なことをする学生はいますが、それを摘発するために犯罪捜査みたいなことをするくらいなら、騙された方がよほどマシです。

それと似たようなことで思い出したのですが試験監督をする時、僕は教室を歩き回らないんです。これは尊敬していた美学の恩師である新田博衛先生に聞いてその真似をしているのですが、彼は試験監督をする時、教壇に座ったまま動かないそうです。ある時同僚の誰かに、それではカンニング防止にならないではないか、と注意されたそうですが、歩き回ったって見えないところでカンニングする学生はいる、前でじっとしていれば少なくとも最前列の学生のカンニングは防げる、と答えたそうです。僕も、試験監督は不正を摘発するためにやるのではなくて、万一何か非常事態が起こったり気分が悪くなる学生がいた時に対処するためだけにやるのだと考えていましたので、歩き回ることはやめました。

たしか2010年前後だと思いますが、文学部のある授業のレポートで「コピペ以外禁止」って

いう条件を出したことがあります。つまり僕の美学の授業で出した課題について、ネットからコピペした文章だけを使ってレポートを作成しなさい、と。そうしたら提出期間前の授業の終わりに、何人かの学生が「先生、頼みますから自分で書かせてください」って嘆願に来ました（笑）。コピペだけでちゃんとレポートを成立させるって、めちゃくちゃ高いレベルの編集能力が必要とされるんです。

面白かったんで、ブログでそういうことやったって公表したんですよ。他大学の学生も見てくれていろいろ反応がありました。そしたら同僚の何人かが心配して、そんなふざけたことして、しかもそれをネットに晒して、大学の教務課から注意を受けたらどうするんですか、みたいな。そんな注意なんて一度も受けたことはない。大学の教務は忙しくてそんなことチェックしないし、見つけたとしても何にも言わないです。なのに先生たちの方が自主規制みたいなことを言いあう。これも一種の人工知能化ですね。2000年代に入ってからこの傾向はすごく進んだと思いますね。

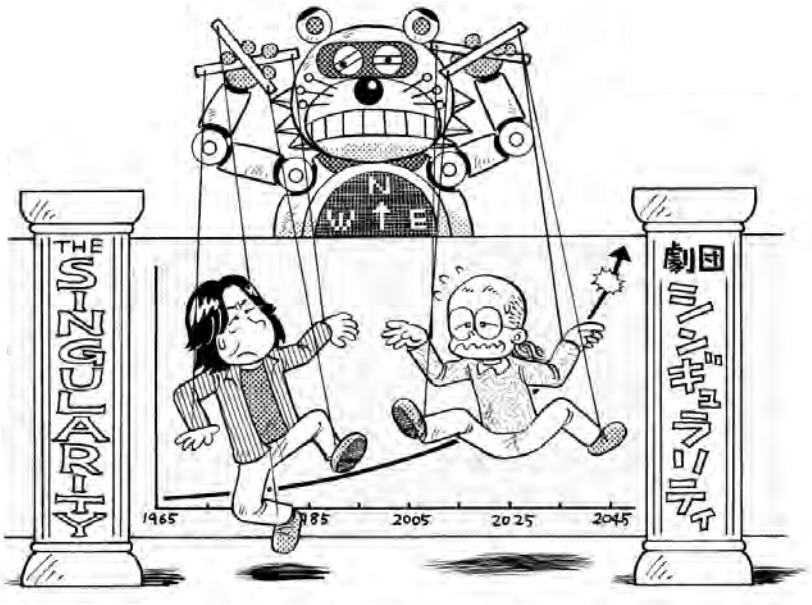
発言者Ｂ　ありがとうございました。今日、まずこの室井さんの「2045年の手記」を読んで素朴に、ちょっとご本人を知ってる人間としては　室井さんのマインドがアップロードされたロボットと室井さんの対話なのか、喧嘩なのかをちょっと見てみたかったなっていう感想を抱きました。

ただ、今から30年ぐらい前に吉岡さんと室井さんが書かれた『情報と生命』や室井さんの『情報宇宙論』における議論と今回のシンギュラリティ論は、ちょっと似て非なる部分があるという感じも持ちました。『情報と生命』とか『情報宇宙論』の議論も、「人間機械論」として捉えられるところもあるかなと思いますが、ただ、そこで多分違うのは、当時は人間についてのもっと別な見方とか可能性への期待として、人間をコンピュータプログラムとハードウェアのアナロジーで見たらどうなるだろう、みたいな話をしていた。

それに対して最近のシンギュラリティ論やそれに関する言説は、「人間とはこういうもので、だから守らなきゃいけない」という価値観をみても、それからAIは人間を越えるか越えないかっていう問いかけをみても、逆に時代が逆戻りしていて、すごく保守的な人間観を温存しながら、それを超えてるか超えてないかっていう話をしてる気がするんです。だからそういう意味で、AIとか元のデジタル情報っていうのが幻だとすると、そこに人間がどういうストーリーを見出すか、期待するかっていうこと自体が、外部状況や時代性によって変わる、そういうことがやっぱりあるんだろうなと思いました。

AIが出てくることによって、逆に人間の中で機械に回収されない部分は何か?って期待する点では、わりと自分も近いんですけど、同時に人間の中の機械的な部分とそうじゃない部分っていう見方をするのも、ある意味人間特有なのではないか。僕らが自分の思考を言葉で捉えて分析的に見てるからそう思うので、もしかしたら機械とか人間以外の生物の視点から見ると、僕らが合理的に思考してると思ってること、その当て感とか感情とか、そういうのと形式的な思考が入り交じってやってるのを、人間がバーチャルに分けて考えてるだけで、そういう人間的思考を逆にコンピューターとかAIに投影してみてるのかなとも思います。

吉岡　そうね、今のお話を聞いていて、このチラシの今回の予告の表面の方に谷本さんが描いてくれた「シンギュラリティ劇場」っていうのを思い出しました。室井さんと僕が操り人形で、背後で糸を引いているのは巨大ロボと化したタヌキなんです。このタヌキロボの中にAIがいるのかもしれない。僕はこのマンガがすごく好きで、シンギュラリティ批判をしていると思っている僕らが、結局は操られて芝居させられてるだけなんじゃないか、みたいなことを感じたんだよね。



イラスト／谷本 研

機械と人間の競争、コンピュータと人間知性との対決といったものは、はじめからどこか芝居じみているのです。「能力」という考え方がそうした芝居を成り立たせる仕掛けなんですね。例えばコンピュータがチェスで人間のチャンピオンを打ち負かしたっていうけど、本当にそんなことが起こったのか？　つまりあの時、人間と機械とが本当に対決していたのか、コンピュータは本当にチェスを指したのか、ということが疑問なんです。僕は、コンピュータはチェスなんて指していないと思ってる。あたかも機械と人間とがチェスをしているかのように演出するのは、人間がやってることであって、結局のところ、人間が人間を見てるだけなんじゃないかと。

機械と対局するなんてことは、もともと起こっていない。少なくとも現在の段階では、想像することもできないと思います。人間が感情移入しているだけで、イライザは対話もカウンセリングもしていないのと同じです。結局、人間が一人芝居をしているだけだと思う。それがこのマンガで「シンギュラリティー劇場」と呼ばれているんじゃないかって、これを見ながら考えてたんです。と同時に、人間はそんなに特別な存在か？ という疑問も感じる。例えばAIに批判的な人の中には、こういうことを言う人もいます。つまり、ChatGTPみたいな生成系人工知能は、まるで人間の質問を理解して答えを返してるように見えるけど、実は機械は何にも理解なんかしていない。ただこういう質問にはこういう回答が通用することを学習し、確率的に予測しているだけであると。それはその通りだと思うんだけども、だから人間とは違うとは、言えないと思うんです。

なぜかっていうとね、例えば安藤さんを僕は昔から知ってるから、何に関心を持ちどういう風にものを考えるか、ある程度分かっていると思うんですよ。だけども全然知らない人と初めて話すような時はどうするか。たとえばシンポジウムなどで僕がほとんど知らないような分野の人と話す時には、相手の言っている内容なんてほとんど理解していないんです。でも何度かやりとりしているうちに、この話題にはこう返してみたらと予測して話すと、相手が「吉岡先生の言われる通りです」と言う。あ、当たったと思ってその方向に学習を強化していくわけ。すると、内容なんて理解してなくても対話は成立するんです。つまり、人工知能がやってるのとほとんど変わらない。

子供が何かを学習してゆく過程や、大人でも知らない分野の知識に馴染んでいく時には、一歩一歩きちんと理解を積み重ねているわけじゃなくて、AIと同じように確率的に予測しながら学

習し、成功した反応を強化していってるだけなんじゃないか。いわゆる「理解」というのはそうした学習が一定レベル進んだ後で、後付け的に解釈してるだけではないのか、ということも思うのです。だとすると、機械と人間とはほとんど同じことをしていることになる。

発言者C　すごく簡単な質問なんですが、著作権のことに関して。その、生成系AIが拡大してゆくことによって、知的財産権とか著作権というものについての人間の理解や価値観が変わるってことはありえるのでしょうか。

吉岡　長期的にはそうならざるを得ないと僕は思うんです。

発言者C　ってことはやっぱりコピペがダメっていう基準自体がもうなくなっていくということ？

吉岡　それは分からないけどね。ただ知的財産権とか著作権という考え方は、人類文明の中には元々なかったものだからね。人間が作り出したものは基本共有財産で、いわばコピペし放題だったわけです。ある時点から、制作物にも所有権を設定する、お金を取るということをやり始めた。しかし情報技術やデジタル・ネットワークの時代になって、いろいろと無理が出てきている。今はそういう段階です。永久に続くとは思えません。

発言者C　ちょうど「研究公正」という、改竄とか盗用はダメですよっていう授業を受けていて、授業が全部終わった後に、その先生もChatGPTのことをおっしゃって、その中で著作権の問題が揺れ動いてるって話をされていました。それで、じゃあ何で著作権が必要なんでしょう?って学生に質問してきて、学生が答えると、それをことごとく潰してくるんですよね。人類の歴史から考えたら、著作権なんて意味ないよとか、哲学の歴史なんて全部過去の上書きだけれども、だからといって先人の功績とか尊厳が傷つけられることなく続いてきたよ、みたいな話をされていて。ってことはやっぱり、そういう僕らが今持ってる価値って無くなっていくのかって思ったんです。

吉岡　それはいい先生ですね。著作権や知的財産権は法律によって保護されているので、お金だけの問題のように思われがちですが、そこにはもうちょっと深いレベルの問題も重なっていて、それは信任というかクレジットという問題です。著作権が消滅した作品は、たしかにコピーすることにお金を払う必要はないけど、それを自分が作ったものだと主張することはダメですよね。法律的にというよりも、倫理的に間違っている。お金の問題とは別に、「誰が作ったか」ということは保持していかなければならない。それは知識の世界全体に対する最低限のリスペクトで、時代が変わっても護られるべきだと思うんです。つまりズルをしないということですね

　たとえ間にAIが介在しても、知識やイメージの由来を尊重するのは人間なので、変わらないと思います。先月の講座で太宰治の『二十世紀旗手』という作品のサブタイトルになっている「生まれてすみません」というフレーズが、いかにも太宰治らしいと受け取られてきたのだけど、実は寺内寿太郎という詩人の書いたものだったという話を紹介しました。ひどい話ですが、太宰に悪意はなかったようで、皮肉な言い方をすればこのフレーズは太宰治が盗用したことで有名になり文学史に残ったとも言えます。

発言者D　美術家の中ザワヒデキさんが人工知能美学芸術研究会（AI美芸研）というのをやってらっしゃいまして、何回かお邪魔したんです。その中で、AIと人間の関係性を4段階くらいに

分けて説明されてて、最初のうちは人間のやってることを学習して、やり方をシミュレーションしていく段階から、徐々にAIが教育法を憶えるような段階に入ってきて、最終的にはAIがAI自身のための美学っていうのを発明する段階に至るんではないかって。面白いなと思って聞いてたんですけど、何かそういう今現在の人間の知的フレームからも外れるというか、今現在はまだ人間の知的フレームの中には含まれていないような領域について考える素材、ツールとしてAIを使うっていう考え方に関してはいかがでしょうか?

吉岡　面白いと思いますが、現実の問題として考えた時に僕が引っかかるのは「AIのための」という部分です。カーツワイルのシンギュラリティでは、シンギュラリティ後の世界ではAIがAIを作り始めるから、人間のAI学者が作るよりも桁違いに効率的で早く改良されていく。それはある意味、AIが自分がAIを作りたいと思っているとも言えるわけですね。自分たち「AIのために」作っているのだと。

発言者D　生産性とか効率と、高速化というような価値観から一旦離れた時に、何かそういう人間の知的なフレームに今現在は入ってない問題について、AIを使っていくっていうことについてはどう思われますか。

吉岡　「人間の知的フレームに入っていない問題」をいかにして表現するかですね。これはアートでしか表現できないのではないかと思う。中ザワヒデキさんのAIの美学で、人間の美学をAIが真似してうまくいくんじゃなくて、AIにしか分からない美学みたいなものが生成するというのは、アーティスティックなアイデアとしてはとても面白いけど、現実問題としてはなかなか想像できない。AIにしか分からないと言われると、じゃあ勝手にやってとしか言いようがない（笑）。

　カーツワイルとかシンギュラリティー信奉者たちの言っていることにも、似たようなところがあるんです。だから、僕は彼らの考えも否定してるわけじゃなくて、芸術上のアイデアとしては面白い、SFとしては面白いと思っているんです。未来がAIだけの世界になって、彼らが感慨深げに、ああ昔「人類」っていう、のろまだけど愛すべき種族がいたんだよな、って言い合っているような風景。それは面白いよね。

安藤　でも吉岡さんのAIは、鬱になってしまう。

吉岡　うん、僕はもうちょっと現実的な物語を考えていて、もし本当にAIがAIを作り始めたら病気になるんじゃないかって思ったんです。そしてそれはある意味、彼らが人間から引き継いだ病気だと思うんだよね。鬱になるんだけど、生身の人間の鬱よりも1兆倍くらい苦しいんですよ。だって知的能力が1兆倍だから（笑）。彼らはなぜ苦しむのか。それは、自分は人間から知的存在者という役割を受け継いでいるにもかかわらず、自分が何で存在するのか、その理由が見出せないからです（笑）。重い荷物を背負わされてしまった。人間のせいなんだけどね。

発言者E　ありがとうございました。「天然」のお話がすごい面白いなと思ってたんですけれども。今後、そういう人工知能的なものはAIに任せていって、より「天然」が重宝されてくる時代が来るのかなと思ったんですけども、また時代が御破算になって原始的な風にもいくのかなとか、精神世界の方とかもどんなふうになっていくのかなっていうあたりを少しお伺いしたいなと思います。

吉岡　人間がみんな「天然」になってしまったら、とんでもない世界ですよね。僕のイメージとしては、「天然」の存在を許すっていう感じで、重宝されるかどうかは分かんないと思います。たぶん、ほとんどの場合、重宝されないと思うんですよ。天然な存在というのは、ただ存在しているだけで役に立たないから。時々、たまたま役立ったりすることもあるんだけども。

そういう存在をどれくらい許すかっていうことは、歴史的に変化していると思う。希望的観測の未来としては、人工知能が拡大することによって、「天然」ももっと許した方が世界は全体としてよくなるってことが分かって、AIのおかげで世界が少し昔に戻るということが希望なんだけれども。逆に悲観的な見方をすると、今僅かに残っている「天然」が最後の一人まで絶滅させられるというカタストロフもありうる。

発言者F　最後にいいですか？ 初めて参加させていただきました。今のAIの鬱の話なんですけど、人も鬱になる人もいれば、ならない人もいる。同じ環境でも違うと思うんですけど、AIもそうなると性格っていうのが…。AIは一つくくりじゃ無く、こういう性格のAIとか、ああいう性格のAIとか、AI自身もそれぞれの個性というものを持つようになっていくんでしょうか？ 人間的になるというか。

吉岡　今でも、ある種の個性というか特性はあると思いますよ。機械だから全部同じじゃなくて、設計とか学習や訓練のし方とかによって変化していく。個性と言ってもいいようなものはもちろんあると思います。だけどお話に書く時には、自分自身がAIになって喋らなきゃいけないから、自分の意識とか感情を通して想像してるわけですよね。だから最初から何か感情はあるものという前提で書いています。そうでないとAIが「私、存在している意味が見出せないから死にます」なんて言わないですから。

それで思い出したのですが、代麻理子さんというライターの人がやってる「未来に残したい授業」というYouTubeチャンネルがあって、今「9月1日の君へ」っていうシリーズをやっています。何かというと9月1日は、中高生とかが自殺するのが一番多い日らしいんですね。2学期が始まるので、僕もそうでしたが、まだ学校に行きたくない子たちがたくさんいるんです。そこに出演したので、公開されたらご覧ください。『9月1日の君へ』という同名の本も教育評論社からまもなく出るようです。その中に僕が10年以上前にブログに書いた「自愛について」という短いエッセイが転載されています。

自殺防止について、そのエッセイの中で僕が紹介した解決法が「肝臓を強くすること」（笑）。つまり、自殺するのは理屈ではなく感情による行為で、その感情を決定するのは何かっていうと、内臓も含めた身体の状態なんですよね。だから身体が決めるんです。頭っていうのは、単にその状態をモニタしてシンボル化し出力しているだけです。だから、遺書を書いたりするけれども、遺書に書かれたことが原因じゃないんですよ。その背後には感情のベースがあって、その感情が決定している。感情といっても喜怒哀楽みたいな表面のさざ波ではなくて、もっと深いとこにある「生の感情」、生きることを決めてる感情なんですよね。これは内臓とか、消化管やその中に住む細菌の生態系とか、そういうもので決まってるんだと思うんですよ。

だから本当は、AIも鬱になったりそうした深い感情を持つためには、内臓や常在菌の生態系に相当するような、何らかの身体を持つことが必要だと思います。それが不可能だと断定する理由はないと考えます。（拍手）

2023年8月19日（土）　於：京都芸術センター「大広間」