

課題番号 1

- ID: 03-190413
- Name: Kadoi Takemaru

課題 a

コペルニクスの原理の実例と思われる事象を探し出して検証せよ。

コペルニクスの原理とは

コペルニクスの原理^{*1}とはこの宇宙において何かを観測する際に、特別な存在は存在しないという原理である。例えば、日本という国家が今後 100 年の間にどうなるかということを観測する際に過去に蓄積されたデータを用いることはできるが、未来を完全に予知する能力がある神のような存在はいないということである。

実例

未来の予測とはこのような特別な観測者ではなく、ランダムな観測者、つまりある確率分布に基づいた観測者を仮定している。これを実例として大学受験の合格判定を上げる。

私が予備校に通っていたときに仲の良い人が数人いた。そのなかで合格判定が A 判定 (80% が合格する) 人を思い出してみる。そもそもそんなに友達が

^{*1}「コペルニクスの原理と未来予測」<http://www.iba.t.u-tokyo.ac.jp/iba/AI/Copernican21.pdf> (2021/11/24 閲覧)

多い方ではなかったので 5 人くらいしか思いつかないのだが、その中で落ちたのは 1 人だけだった。サンプル数としてはかなり少なく信憑性は低いが確かに予備校の予測値通りの結果となっている。

課題 b

強い AI(汎用 AI) の実現可能性（すでに AI はできている？）と、それに関する脅威について論ぜよ。A4 で 5～7 ページ程度とする。

強い AI(汎用 AI) の実現可能性

■強い AI とは コペルニクスの原理^{*2}を仮定し、これを基にヒトという種の寿命は 5,100-7,800,000 年となる。この寿命が尽きるまでに実現するかどうかを論ずる。

そもそも、強い AI とはなにか。書籍「強い AI・弱い AI」^{*3}によれば人格を投影されているかどうかを強い AI と弱い AI の明確な境界線としても明確な定義がなく厳格にどちらであるか判定はできないとされている。ここでは汎用 AI を強い AI として、強い AI と汎用 AI を同じとして扱う。

■汎用 AI とは では、汎用 AI とはなにか。これは特化 AI と対になる概念である。特化 AI とは将棋 AI の AlphaGo であったり、画像を読み込み有害なコンテンツかどうか判定する AI のような特定のタスクに特化した AI のことである。一方、汎用 AI とはより汎用的なタスクが可能な AI である。ここで重要となるのは「汎用的なタスク」という言葉の指す汎用性の範囲であ

^{*2} 「コペルニクスの原理と未来予測」 <http://www.iba.t.u-tokyo.ac.jp/iba/AI/Copernican21.pdf> (2021/11/24 閲覧)

^{*3} 鳥海不二夫 (2017) 「強い AI・弱い AI」 丸善出版

る。例えば二足歩行で歩くというのは単純なタスクだと賛同する意見は多いだろう。一方、二足歩行で障害物を認識し障害物を避けながら目的地まで歩くというタスクは単純なタスクだろうか。これは意見が分かれると思う。もし、このタスクを達成する AI が強い AI という仮定を置くと、すでに強い AI は実現していると言える。

ボストン・ダイナミクスの作った二足歩行ロボット、Atlas に関する PV^{*4}ではロボットが躍動感ある動きでパルクール、障害物を使ったスポーツ、をしている。単純に歩けるような AI を作るだけではこのような動きはできないはずで、障害物を認識し、バランスを保ちながら、移動をするという高度なタスクをこなしている。

■人間らしいタスクとは 上の例では大事なのはタスクの質というより数であった。あるタスクを複数のタスクに分解し、それぞれに特化した AI を適用することでタスクを解決していた。この方法ではほとんどの場合特化 AI を解決できるかどうかと特化 AI をどのように組み合わせるか、そのインターフェイスが問題となる。次はタスクの数ではなくタスクの質に注目して汎用 AI を定義し、その実現可能性を考えたい。

どのようなタスクが人間らしいと言えるだろう？ 具体的なタスクから考えていくと、例えば小中学校で解くような問題は人間らしいタスクと言えるかもしれない。しかし、それらは上で言ったような特化 AI を組み合わせて解くことができる。他に明日の夕食の献立を考えることや部下に適切な仕事を振ることが考えられると思う。しかし、これらも特化 AI によってある程度は解決されている。夕食の献立は「おいしい健康」^{*5}というサービスが考え

^{*4} “Atlas | Partners in Parkour” <https://youtu.be/tF4DML7FIWk> (2021/11/24 閲覧)

^{*5} 「おいしい健康」 <https://oishi-kenko.com/> (2021/11/24 閲覧)

てくれるし、タスク割当は「Optamo for Task」*⁶というサービスが考えてくれる。

このように具体的なタスクはほとんど AI によって自動的に答えを導くことができると言えるだろう。より抽象的なタスクを考えなければ人間と AI の違いはわからないと言えるだろう。

しかし、抽象的なタスクというのは人間も全員ができるわけではない。哲学的な問題や数学的な問題は全人類が解けるわけではなく、むしろ教育を受けた限られた数の人間しか解くことができない。また、教育をされた人間が全てできるわけではない。それは義務教育の内容を全てを日本国民全員が理解できるわけではないことからわかるはずである。

そう考えると「全ての人間ができるようなことができる」AI はまだできておらず、そしてこれからもできるとは限らないが、「ある程度の数の人間ができるようなことができる」AI 既にできていると考えられる。

よって、汎用 AI は既にできていると考えられる。

脅威

今後上で述べたように AI を作るような AI が生まれた際に人間が全く感知できない可能性があるということで。これはフェルミのパラドクス*⁷が参考になる。

■フェルミのパラドクス フェルミのパラドクスとは高度に知的な宇宙人が地球で見つかる可能性の高さに反して、未だにそのような存在を確定する証

*⁶「Optamo for Task」<https://www.optamo.jp/task/> (2021/11/24 閲覧)

*⁷「フェルミのパラドクス」http://www.iba.t.u-tokyo.ac.jp/iba/AI/fermi_paradox.pdf (2021/11/24 閲覧)

拠が見つからないという矛盾である。

このことを説明するいくつかの可能性がある。既に存在しているが宇宙人は人間が知覚できないような存在である。可能性は高いが宇宙人は存在していない。

■I' m not a robot 今後汎用 AI に関しても前者と同じことが起きる可能性がある。つまり、汎用的な人間と同じような振る舞いをする AI が現れたとして、そのような存在が見つからない可能性がある。例えば、そのような存在がプログラムとして誕生し、どこかのインターネットの繋がる端末にいるとする。その端末の電源が切れない限りインターネット上のサイトにリクエストを送りあたかも人間と同じような振る舞いをするかもしれない。

実際、人間が PC 上でやっているタスクはやろうとすれば大半のものは AI がやることも可能である。それが人間によるものかどうかを判定しているサービスもある^{*8}が、すでにこれを破る AI もできている。

■経済的・利便的な問題 このような存在が増えると困るのはインターネットでビジネスを行う人間である。つまり、人間と同じような振る舞いはするが経済活動はしない存在が増えれば実態とはかけ離れたデータが貯まるようになり経営上問題になる可能性がある。

また、同様に AI が利用するサービスを利用する利用者にとっても脅威的である。サービスの方向性が人間の利用者の欲するものとはかけ離れたものとなる可能性があるからだ。

^{*8} “Google reCAPTCHA” <https://www.google.com/recaptcha/about/> (2021/11/24 閲覧)

■戦争・紛争 当然だが、AI は戦争・紛争に利用することも可能である。これは既に利用されており、AI によって制御されたドローンが紛争に導入され戦果をあげている*⁹。これはサイバー空間においても同様である。もちろん AI は生活を豊かにしてきたが、同時に多くの脅威があるということが言えるだろう。

*⁹ 「2030 未来への分岐点 (5) 「AI 戦争 果てなき恐怖」」 <https://www.nhk.jp/p/special/ts/2NY2QQLPM3/blog/bl/pneAjJR3gn/bp/pODBZw34z8/> (2021/11/24 閲覧)