人工知能に対する人間の恐れ

上級日本語１：書き方

123866

ベイリー・コン

人工知能に対する人間の恐れ

123866　ベイリー・コン

本稿では、人工知能に対する人間の恐れについて、映画に与えた影響や人間が持つ信念等について考察した。人間が他のものより優れているという信念を持っていることが恐れの二つの主な原因を生んだと考えた。それは自己認識と自由意志である。自己認識は社会的状況で他者を理解するために、「心の理論」で自分と他者に分けて、さらに内観的な力で自分がいると考えている。しかし、他者がいるか保証はないので、人間の「心の理論」は保証出来ず自己認識がないと言える。この二つの原因は単なる自己欺瞞だと思われ、実際はもともとなかったことだ。自由意志については、人間は単に強大プログラムで、状況や感情等に支配されているので、自由意志がない。また、全知なる存在と自由意志は共生出来ない。このように人間は人工知能を恐れる理由は、単なる自己欺瞞なので、本当は恐れる理由がない。

人工知能に対する人間の恐れ

123866　ベイリー・コン

１．序論

「人工知能はまだ開発されてない」とよく述べられている。映画の影響により人工知能というのは誤解されやすいので、そのような発言がされている。映画の中では、人工知能というのは「自己認識や自我を持つ物」あるいは、「感覚や感情を持つ物」と考えられ、人々にもそのようなイメージを与えている。その映画のイメージがどんなものかは、「アイ・ロボット」や「ターミネーター」という映画を見ればよく分かると思う。

１．１．映画のイメージ

「アイ・ロボット」の構想では、ロボット工学三原則[1] に従っているロボットが人を殺すことが出来るようになった。なぜ出来るかというと、そのロボットの人工知能が進化したと同時に、ロボットがロボット工学三原則の解釈も変えたからだ。この進化で、全体として人類を守るために犠牲はしかたがないと思ってきた。それに加えて、人間はロボットに人間を保護させていたが、人間は自ら戦争したり、地球を汚染したりするので、人間の犠牲は自らの破壊的性質により、転覆でのほうが少ないので、ロボット工学三原則によってはそういう行為をせざるを得ないのだ。この映画で、ロボットの人工知能が進化して、ロボットがプログラミングされていなかった行為をして、人間はロボットが自由意志を得たと思われた。

1　ロボット工学三原則：
第一条 ロボットは人間に危害を加えてはならない。また、その危険を看過することによって、人間に危害を及ぼしてはならない。
第二条 ロボットは人間にあたえられた命令に服従しなければならない。ただし、あたえられた命令が、第一条に反する場合は、この限りでない。
第三条 ロボットは、前掲第一条および第二条に反するおそれのないかぎり、自己をまもらなければならない。（小尾）

「ターミネーター」の構想では、自己認識を得た人工知能を持つマシーンがいる。そのマシーンは人類を掃滅しようと、核による大量殺戮を起こした。そして、生き残った人間を抹殺しようとした。この映画で人間は、マシーンが自己認識を得たことを気付いてから、そのマシーンの機能を恐れて、停止しようとする。

１．２．人工知能というと、本当は何だろう

しかし、この映画の影響による人工知能のイメージはどれぐらい正確なのだろうか。『ジーニアス英和大辞典』によると、Artificial Intelligenceというのは「推論や判断等人間の持つ知的機能を人工的に実現する機械」ということである。つまり、かつて人間しか出来なかったことをする機械が人工知能だということだ。この定義によると、我々の日常生活の中でも様々な人工知能を持っている機械がある。テレビやデジタル・ビデオテープ・レコーダー等はまさにそれではないか。したがって、人工知能はもう既に存在していると言える。

ただし、現在の人工知能には限界があり、出きることと出来ないことがあるだけなのだ。例えば、現在の人工知能は囲碁でプロに勝つことや、物体認識が出来ないのに対し、数学的証明を解決することや、チェスで世界チャンピオンに勝つこと、車を運転することは出来る。これらは有名な人工知能の例である。

Teller（1998）はこのようなことを言った。ライト兄弟のフライヤーと呼ばれる飛行機は現代の航空機のボーイング747の特性には及ばないが、飛行機と呼ぶことが出来る。現在の人工知能の機械はライト兄弟のフライヤーと同じようなものではないか。つまり、現在、人工知能は存在しているのに、人々の期待が高過ぎるために人々に意識されていない

のである。このような比較をする理由は人工知能の分野で発展が続くたびに人々は人工知能の水準を向上させて、今ある人工知能の能力を認めることを避けているということである。

さらに理解するために本論では人が人工知能を恐れる一般的に考えられている理由について論じる。筆者は人が人工知能を恐れる主な原因は、自己認識と自由意志の二つであると考える。「人間は自己認識と自由意志を持っている生き物で、どんなものより優れている」と信じているので、人工知能を恐れていることを本稿では結論づけたい。

２．本論

２．１．自己認識の観点

自己認識という概念は広くて曖昧なので、持っているかどうか判断するのは困難である。まず自己認識は何なのか定義しなければならない。人間は人間のみが自己認識を持っていると信じているので、人間が他のものより優位だと信じている。したがって、人工知能に対する恐れをさらに理解するために、定義では、優位性を維持するために他の生き物から区別する基準を選択して、人類の優位の考えを守れなければならない。例えば、鏡で単に自分を認識するということは、今回の自己認識の定義には使えない。その理由は、もう類人猿やバンドウイルカ等が自分を認識が出来ると分かっているので、人間と動物を区別する基準ではない。つまり、人間の優位性を示すものではない。しかし、自分がどのような存在であるかと分かるという定義は、我々の了解するところでは、人間しか出来ない。つまり、人間の優位性を主張する自己認識の定義を選ぶためには、自分がどのような存在かを理解出来るかという基準を選ぶ必要がある。

霊長類は社会性生物で、人間は霊長類なので、人間も社会性生物である。社会性生物なので、人間は他者と住むことでしか生きられない。この社会的状況から、人間には相手を理解する必要が出て来た。では、我々はどうやって相手を理解出来るのだろうか。それは人がミラーニューロンを使っているからである。

２．１．１．ミラーニューロンと内観的な力の関係

人間の頭の中では、同種が行動するときに、自ら行動するときと同じ脳の神経が活動することがある。その神経がミラーニューロンであり、これは他者の行動を鏡のように自分の頭の中に写し出すので、こう呼ばれる。Ramachandran（2007）によると、我々の内観的な力は、社会的状況で他者を理解することや予想するために進化してきた可能性がある。「相手の嫉妬を感じるために自ら嫉妬を感じる」というのは相手の嫉妬から生まれる態度を予測する近道であるということである。相手を理解するためには、自分が同じように感じることが、一番効果的な方法だ。最初は自分を理解することではなく、他者を理解することからこの力が生まれたというわけだ。

２．１．２．自分の心と他者の心

それによって、内観的な力はどのように自己認識につながるかというと、自己認識の概念は内観の活動をすることによって生まれる。自己認識は「心の理論」と同じで、内観することによって、自分のことを認識することだ。「心の理論」とは、心理学の概念であり、他者の心の動きを類推したり、他者が自分と違う信念を持っているということを理解したりする機能のことである。内観的な力で自分の心を証明することは出来るが、他者の心が存在するかどうかは類似の概念で推定しか出来ない。つまり、「心の理論」は他者に関す

る部分が完成ではないということだ。

推定しか出来なかったら、どのようにして自分の心が存在しているかを確信出来るのであろうか。Miniskyにると、それは出来ない。Minisky（1982）によると、「私たちはほんの少ししか自分のことが理解出来なくて、大体人間が自分に対して持つ自我意識は推測である」「はっきりと言うと、私たちの『意識』が明らかにしているのは大体架空だ。しかし、それは聞いていることや見ていること、または思考の一部に気づいていないというわけではなくて、心の中で何が起きているのかあまり気がついていない」ということだ。他者の心も自分の心も確信にすることが出来ないので、自己認識があるとは言えない。

２．２．自由意志の観点

『明鏡国語辞典』によると、自由意志というのは「他からの強制や拘束を受けることなく、行動を自発的に決定する意志」ということである。つまり、外部からの影響ではなく、自分が決めたことをすることや選ぶことだ。しかし、外部ではなく、内部の影響で我々は意思決定をすることもあるだろう。例えば、人間の遺伝子決定である、その結果人間はあるものにアレルギーを持ったりすることがあるのである。もしかすると、我々はただの強大なプログラムなのではないか。

２．２．１．我々はただの強大なプログラムではないか

次の仮定を見てみよう。自分と親友と恋人の三人で入り口だけがある部屋にいる。自分と親友は短銃を持っている。自分の親友が精神病になってしまって、誰の言うことも聞かなくなり、自分の恋人を短銃で人質に取っていた。警察は間もなく到着するが、親友は既に一発を恋人に向けて撃っていたが、幸い、誰にも当たらなかった。しかし、自分はもう

一発彼が撃ちそうなことを恐れている。この状況で自分はどうするだろうか。

一般的には、三つの選択がある。それは、「二人共を助ける」「恋人を助ける」「親友を助ける」である。「二人共を助ける」と思っている人々は、例えば親友の短銃を持っている腕だけを撃つ。「恋人を助ける」と思っている人々は、何も考えずに親友を撃つ。「親友を助ける」と思っている人々は、攻撃して何とか短銃を取り上げる。選択の違いは、結果を問わず、目的と何を意識しているかで分けられる。

三つ目の選択はほとんど選ばれないが、普通考えられない四つ目の選択もある。例えば、自分だけを助けることである。短銃を置いて一人で入り口を通って逃げる行動だ。この選択も出来るが、どうしてこの状況で人はその選択肢を思い付かないのだろうか。精神病の親友が短銃を持って、もう既に一発を撃っていたので、十分自分の命も危険なのに、どうして誰も自分のことを優先しないのだろうか。もしかすると、我々はただの強大なプログラムで、ある状況に置かれたり感情的になると大切なことしか考えられないような設定にされている可能性があるということである。大切なことしか考えられないというようにプログラミングされている可能性があるということである。

２．２．２．もし全知なる存在がいたら

多くの文化や宗教では、全知なる存在がいる。もしそういう存在がいるとすると、人間には自由意志がない。なぜかというと、自由意志というのは自発的な決定なので、もし自分が何かをする前に、何をすると分かっている存在がいれば、もうそれは自発ではないだろう。

自由意志が全知の存在と共存することが出来るかという議論では、Come Right

Ministriesによると、このような例が上げられている。「自分には五歳の息子がいる。自分が食卓の上に息子の好きなクッキーを夕食の一時間前に置いて、息子がそのクッキーを見かけたら、息子は絶対取って食べるはずだ。自分はクッキーを食べた息子に食べるように決断することを強制しなかった。それに、自分はそこにいても、いなくても、息子の行動は何も変わらない。自分は息子の事を十分分かっているからこそ、息子がクッキーを食べることを確信していた。強制しなくても、息子の行為が分かっている」ということだ。

しかし、この議論には欠陥がある。息子が絶対クッキーを取って食べるとは言えない。その保証は何処にもない。息子のことを十分に分かっていても、息子がクッキーを食べるという 100 %の保証は出来ない。したがって、全知の存在でしか「絶対」という言葉を使えない。全知は 0.1 %でも確信がもてないと、どんなに小さな間違いも、知らないということは許されていない。それは全知の前提であるのだ。自由意志というのは単なる自己欺瞞[2] ということであるのだ。

２．２．３．社会組織には大事

人間が自由意志を持っていると信じているもう一つの理由は、自由意志から生まれたのは道義的責任なので、自由意志は必然的に道義的責任を伴うので、自由意志がなかったら、道義的責任もない。昔から社会組織は道義的責任に基づき作り上げられたので、自由意志がなかったら、社会組織は実現されない。なぜかというと、自由がない限り、自分の行為や行動等の責任を取れないからだ。自由がない社会、つまり、すべてが拘束された社会では自分の行動によって決められることはないので責任を取る必要もない。したがって、社

2　自己欺瞞：
一般的に考えられている定義は「自分を騙すこと」だが、それだけではなくて、「事実を知らずに何かを信じているまま」も含めて使っている。

会が滅びてしまわないためにも自由意志があると信じるしかない。

それでも、刑法ではある場合に、人間が自由意志を失うことを認めている。多くの国では刑法の中に責任能力という概念がある。その概念は、事件が起こったとき、正常な精神状態ではなかった被告人の人権を守るためにある。自由意志が絶対あると信じている人々の中では、この事を誰も疑問に思わないのはどうしてだろうか。自由意志が絶対あると信じながら、時々それを失うこともあるとは、矛盾ではないか。

３．結論

人間がどんなものよりも優れていると思っている人々は、人工知能の機械が人間と同じように、または、人間以上のようになることを恐れている。指摘した二つのこと、自己認識と自由意志がある可能性は極めて低いので、それら人々は人間が優れていると思わず、恐れず、人工知能の機械と共に生きていくべきだと思う。さらに、人間は人間であり、我々が思ったほど優れていないと分かっても、我々が人間であることは変わらない。プライドを傷つけられるだけだ。恐れを感じるのはやむを得ないかもしれないが、大切なのはそれに気がついて乗り越えることだ。いつまでも目を閉じて事実を無視して信じたいことのみを信じたら、何も進展がない。我々の祖先は迷信を盲信しなくて、科学に目を向け、その上で疑問視していた。疑問視していた祖先のお陰で、人類は宇宙まで行けたり、病気を直したり、寿命を伸ばしたりすることが出来た。したがって、我々も祖先と同じように自分の知能や信念等を疑問視しながら、前に進むべきだ。

参考文献

『ジーニアス英和大辞典』（2008）大修館書店

『明鏡国語辞典』（2008）大修館書店

小尾芙佐（2004）『われはロボット』早川書房

Come Right Ministries.（2009）“Does Omniscience Contradict Free Will?”.

http://www.comereason.org/phil_qstn/phi038.asp

2009/06/04 13:00 JST

McCarthy, J.（2007）“What is Artificial Intelligence?”.

http://www-formal.stanford.edu/jmc/whatisai/whatisai.html.

2009/05/18 10:44 JST

Minsky, M.（1982）“Why people think computers can't”.

http://web.media.mit.edu/~minsky/papers/ComputersCantThink.txt.

2009/05/24 18:34 JST

(First published in AI Magazine, vol. 3 no. 4, Fall 1982)

Paterek, L.（2005）“Free Will and Artificial Intelligence”.

http://serendip.brynmawr.edu/sci_cult/evolit/s05/web2/lpaterek.html.

2009/04/22 11:17 JST

Ramachandran, V.（2007）The Neurology of Self-Awareness.

http://www.edge.org/3rd_culture/ramachandran07/ramachandran07_index.html

2009/05/25 11:03 JST

Searle, J. (1995) How Artificial Intelligence Fails. The World and I. July
http://www.worldandi.com/specialreport/1995/july/Sa13327.htm.
2009/05/27 10:56 JST

Teller, A. (1998) Smart Machines, and Why We Fear Them. New York Times.
March 26.

Vander Nat, A. (2005) "Free Will Arguments".
http://orion.it.luc.edu/~avande1/free-will-args.htm.
2009/06/01 14:22 JST