

[人類はどう生きるべきか？ ITはどうあるべきか？]

シンギュラリティへ向けて あなたと私はどうしたいか？

応
般

7.5

どうなるかではなくどうしたいか

人工知能研究者は、いろいろな人から時々「機械が心を持つようになるのでしょうか？」という質問を受ける。そういう質問を受けたとき、筆者は「機械が心を持つか持たないかではなく、僕は、機械に心を持たせたいか、持たせたいとしたら何のためにどのような心を持たせたいかを考えたい、と思います。あなたは機械に心を持たせたいですか？持たせたいとしたら、何のためにどのような心を持たせたいですか？」とお答えしてきた。

人工知能の能力が人間の能力を超えてしまう可能性が議論されている。工学系の研究者としてやや不満に思うのは、そこでは、まるで他人事のように、どうなる可能性があるかだけが論じられることが多いことである。評論家が他人事のように論じるのは仕方のないことかもしれない。しかし、少なくとも本会の会員は、「どうなるか」ではなく「どうしたいか」を議論する責任も負っているのではないかと、筆者は考える。また、当然のことであるが、どうしたいかは、技術者だけで考えるべきことではなく、さまざまな領域の専門家はもちろんのこと、専門を持たない一般の人々も含めて、議論していくべきことであろう。「どうなるのかな」と指をくわえて傍観していると、とんでもないことになるかもしれない。核技術や金融工学などの最先端の技術が危機を招いたとき、これまでは、なんとかかんとか、人々は危機を回避してきた。しかし、来たるべきシンギュラリティ（技術的特異点）においては、そうはいかないかもしれない。物理学者のHawkingは、こう言っている¹⁾。

「人工知能の成功は人類の歴史において最高の出

来事になり得るだろう。しかし、もし危機をいかにして避けるかを知ることができなければ、残念ながら人類の歴史の最後の出来事になってしまうかもしれない」

オックスフォード大学の哲学者のBostromも「AI研究者の間のベストプラクティスを促進しなければならない。実践者たちが『安全』に真剣に取り組む約束を表明することも、求められるであろう」と主張し、「この問題の解決のために人類の叡智を集結しなければならない」と述べている²⁾。筆者も基本的にはBostromと同じ立場に立っている。

筆者がシンギュラリティに向けて「こうしたい」と考えることは大まかに言えば、次のとおりなのであるが、読者の皆様はいかがであろうか？

- (1) 人類の抱えるさまざまな問題（エネルギー問題、水問題、食料問題等々）を解決するために、人工知能その他の技術を活用できる部分については活用したい。
- (2) しかし、その結果、思わぬ副作用が生まれるかもしれない。生じ得る副作用について、網羅的に検討を行いたい。
- (3) それらの副作用について、技術的に対処する可能性や社会的に対処する可能性について、これまた網羅的に検討を行いたい。
- (4) 対処の方法が見つからず、多くの人々が許容しがたいと思う副作用が生じる可能性があるならば、対処方法が見つかるまで、そもそもの開発を中断させたい。

この4項目は、ごく普通の考え方であると思うし、医学や薬学などの中の一部の確立された領域では、社会制度としてもそれが支持されていると思うのであるが、残念ながら人工知能や情報処理の研究

においては、このような考え方について必ずしも合意がとれていないし、確立された社会制度も存在しない。我が国の人工知能学会においては、「人工知能と未来社会検討委員会(仮称)(旧称:倫理委員会)」と称する委員会で、そのような問題などをこれから議論していく予定である。

人工知能に対して人々が抱く不安

ここで、人工知能に対して人々が抱く不安を列挙しておこう。ほかにもいろいろあるだろう³⁾が、とりあえず、抽象的なものから具体的なものまで、次の17項目くらいを挙げておく。(1)人工知能が人類を滅ぼすのではない(2)人間の尊厳が脅かされるのではない(3)人工知能と人間との恋愛の可能性をどう考えればいいのか(4)人工知能が心を持つのか、持つとしたらそれをどう受け止めたらいいいのか(5)ロボットの権利や義務を考えることになるのか(6)人の雇用を奪うのではない(7)人工知能の考えることと行うことを人が理解できなくなるのではない(8)人工知能の考えることと行うことを人が制御できなくなるのではない(9)人工知能は想定外の事態に対処できないのではない(10)軍事技術として応用されるとき、人を殺すことに対する心理的抵抗を減らしてしまうのではない(11)テロリストなどに悪用されるのではない(12)システムに侵入されて悪意を持った改変を施される恐れがあるのではない(13)プライバシーが侵害されるのではない(14)事故や失敗の責任を誰がとるのか(15)事故の賠償の保険制度をどのように作りかえることになるのか(16)現行法規制と相容れない部分をどうするのか(17)どれくらい故障するのか、どのような失敗の恐れがある

のか。

抽象的な不安に対して抽象的な答えを返すだけでは議論が堂々巡りする恐れがあるので、前章に述べたとおり、我々技術者は、技術的にどのような対処の可能性のあるのかについて、広く、深く、精密に考え、人々に正確にかつ分かりやすく伝えていかなければならない。また、権利や義務や責任の問題については、技術だけでは対処できない。さまざまな領域の専門家や非専門家が集まって議論を重ねていく必要がある。

筆者自身は、まず自分の周辺で少しずつ議論を試み始めているのであるが、技術の専門家でない方々

とお話すると、上記の不安の一覧の中で、意外にも、第3項目の「人工知能と人間との恋愛の可能性」について真っ先に尋ねられることが多くて驚いたりしている。技術者としては技術的な検討から議論を始めるわけであるが、いろいろな方々と議論していると、多くの場合、行き着くところは、「人工知能の問題」ではなく、「人間の問題」になる。たとえば、人工知能と人間との恋愛について考えているうちに行き着くのは、「そもそも人が人を愛するとはどう



いうことか」という問題になる。人工知能と人との恋愛をテーマにした映画「her / 世界でひとつの彼女」も、そういう映画であったと言ってよいであろう。したがって、人工知能に対する不安を巡る議論は常に深くて難しい議論になる。とりあえず、それはそれだけで面白くて有意義なのであるが、ほとんどの場合、「人間の問題」について答えが1つに定まることはない。よって、人工知能に対する不安に関する議論も、ほとんどの場合、1つの答えに収斂させることはできない。

人工知能研究の規範を考えることができるか

前章の最後に述べたとおり、人工知能を巡るさまざまな問題に対する答えは1つに定まらないのであるが、それでも我々は「人工知能をこのように作っていく」というようなある種の倫理的な宣言を行うことができるであろうか？

筆者は、それが可能であるし、やるべきであると考えている。ただし、「ああしてはいけない。こうしてはいけない」というような倫理規定は、人工知能研究には馴染まないであろう。あるいは、Bostrom が主張するように、「我々は人工知能の安全に真剣に取り組む」というような宣言をすることに、筆者は賛成ではあるが、それだけでは不十分であろう。

筆者が考えているのは、人工知能の規範というよりも人工知能の設計美学とでも呼ぶべきものを形成していくことである。今は人々に及ぼす副作用について十分に考えられていない醜い設計の人工知能が野放しに作られているが、安全な人工知能の美しい作り方について、意見交換と実践を繰り返しながら、

なんらかの方向を示して行きたい。その中身についても議論したかったのであるが、すでに紙面が尽きた。別の機会に譲ることにする。

参考文献

- 1) Hawking, S. : Transcendence Looks at the Implications of Artificial Intelligence - But Are We taking AI Seriously Enough?, The Independent, <http://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-ai-seriously-enough-9313474.html> (Oct. 2014).
- 2) Bostrom, N. : Superintelligence : Paths, Dangers, Strategies, Oxford University Press (2014).
- 3) Lin, P., Abney, K. and Bekey, G. A. : Robot Ethics : The Ethical and Social Implications of Robotics, MIT Press (2011).
(2014年10月2日受付)

堀 浩一（正会員） | horii@computer.org

1956年生。1979年東京大学工学部電子工学科卒業。1984年同大学院博士課程修了。工学博士。東京大学大学院工学系研究科教授。人工知能の研究に従事。人工知能学会、電子情報通信学会、日本認知科学会、日本ソフトウェア科学会、IEEE、ACM各会員。2008～10年人工知能学会会長。

