人工知能学会研究会資料
SIG-AGI-001-01

汎用人工知能と社会

Artificial General Intelligence and Social Design

櫻井　成一朗 [1]
Seiichiro SAKURAI[1]

[1] 明治学院大学　法学部
[1]Meiji Gakuin University Faculty of Law

Abstract: While AGI is a very useful future technology, it may cause the race with machines. Since it is convinced that the winner of the race must be machines, the employment of human beings will be lost. In order to alleviate the risk of AGI, this paper clarifies that the human society must be redesigned by the collaboration of AGI researchers and social scientists.

1. はじめに

現在, 汎用人工知能(AGI)の研究が活発になっており，AGI とも密接に関連する技術的特異点（シンギュラリティ）が，カーツワイル[1]により 2045 年にも到来すると言われている。AGI の光の面としては，AGI が実現された際の，科学技術の劇的な進歩が挙げられる。従来の人手によるものとは比較できないほど迅速に，シミュレーションの自動設計，自動シミュレーションが可能となるので，科学技術，産業分野，医療等様々な分野でブレークスルーを生み出し，人類への大きな貢献が期待できる。人類への多大な貢献が期待できる一方，IBM のコグニティブテクノロジーWatson のように，領域特化型人工知能(Narrow AI)でさえ，人類の能力を超えることは珍しくなくなった。それゆえ，現在でも人類にとって Narrow AI，すなわち，機械との競争[2][3]は避けられない。さらに，Narrow AI を超える AGI が実現されれば，人類は当然のごとく機械に敗北し，失職することになる。大量の失業者の算出は，大きな社会問題となる可能性がある。従来の人工知能研究は人間の労働力をより知的な活動に振り向けることを目標としていたので，大量の失業者を生むことはなかった。これに対して，AGI の登場によって，人間の知的活動の大部分を機械が代行できてしまうとなれば，従来の頭脳労働が職業として成立し得なくなる。AGI により科学技術が進歩して豊かな生活がもたらされるとしても，雇用が奪われてしまえば，大多数の人類の生存そのものが危うくなる。本論文では，AGI 研究のもたらす負の影響について論じ，負の影響の回避策の考察を通じて，シンギュラリティ時代に人類が生き抜くための社会のあり方について検討する。

２．AGI のもたらす負の影響

将来 AGI が実現されれば，人間の知的労働を代替する可能性は大きく，現在の Narrow AI でさえ，人間の知的労働を代替する可能性がある。芸術家や学者等の創造的作業や高度な知的作業については，人間を超えるまでにはまだ時間が必要であろうが，いつまでも安泰というわけではない。たとえば，法律事務における，類似判例の検索等は Narrow AI により置き換えられてしまう可能性が高い。裁判官の業務を AI に置き換えることは難しいとしても, 裁判官の代わりに判決の論証を行う AI の実現は十分にあり得るであろうし，パラリーガルと呼ばれる専門家が担っていた，米国弁護士事務所の法律事務の大部分は AI に置き換えられる可能性が高い。人間を超える AGI であれば，弁護士業務自体でさえも，AGI によって置き換えられるかもしれない。これは，他の知的労働者も例外ではなく，企業活動の多くの部分

は AI に置き換えられてもおかしくない。すなわち，Narrow AI であっても，分野によっては十分な脅威となり得るのであるから，人間を超える AGI が登場すれば，雇用問題こそが最大の社会問題となる。

人間を超える AGI の登場以降は，人間が機械との競争に勝利できなくなり，費用対効果の関係で，人間の労働力の置きかえが難しい分野にのみ，人間の雇用が残ることになる。以外にも，単純肉体労働は費用の点で人間の置きかえができないが，単純肉体労働ゆえに，収入は決して高くはない。したがって，人類の生存のためには，富の再分配が課題となり，その一つの方法として BI(Basic Income)制度[4][5]が考えられている。BI 制度とは，政府がすべての国民に対して最低限の生活を保障するための生活費用を無条件に支給するという，社会制度である。一部報道によれば，福祉国家のフィンランドでは，現行の社会保障制度を取りやめて，国民一人に一月あたり 800 euro 支給することを検討しており，今秋にも決定しようとしている。オランダでも同様の社会実験が計画されている。BI 制度が持続可能な社会制度として機能し得るのかどうかは，これらの壮大な社会実験によって明らかになろう。しかしながら，先進国の中でも先陣を切って超高齢社会を迎えつつある我が国において，BI 制度が導入可能であるかどうかは十分な議論が済んだとは言えない。特に高額医療費を必要とする，社会的弱者の立場で考えれば，従来の社会保障制度を廃止して，BI 制度が導入されるのであるから，一律の生活費の支給だけならば，従来の社会保障制度の廃止は国民の寿命を縮めかねないであろう。また，一部の国でのみ BI 制度を導入した場合には，企業が当該国から脱出してしまえば，政府の収入が減少し，国民への支給額が減額されることになってしまう。他の収入源が無ければ，国民全体が健康をそこねるだけでなく，最悪の場合には餓死しかねない。したがって，AGI が生み出す果実としての富をどのように国民に再配分するかという問題については，まだ解決されたとは言えないことになる。

３．Friendly AI 実現の困難さ

BI 制度によって国民の生存が保障できないのであれば，AGI 自身に人類社会の安全性を保障させる，Friendly AI(FAI)[6]を実現するという考え方がある。FAI とは，人類に利便性をもたらすという意味で，人類に友好的な AI である。典型的な FAI としては，図１に示す，アシモフのロボット工学３原則を実装したロボットが考えられるであろう。

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。

出典：『われはロボット』アイザック・アシモフ著、小尾芙佐訳（1983 年）

図１：ロボット工学３原則

ロボット工学 3 原則は，様々な SF で採用されており，たとえば，1999 年公開の米国映画「アンドリューNDR114」では，家事用ロボット NDR114 が自らロボット工学 3 原則を冒頭で説明している。もしロボット工学 3 原則が AGI に組み込まれているのであれば，必要以上に人間の雇用を奪うことは，人間に危害を加えることになることから，ロボット工学 3 原則の第一条により抑制されるはずである。しかしながら，素朴にロボット工学 3 原則を実現しようとすれば，フレーム問題が生じることは明らかである。第一条の「人間に危害を加えてはならない」を素朴に解釈すれば，どこまで影響を配慮しなければならないかがわからないので，ロボットは何ら行動を起こせないことになるからである。

仮にロボット工学 3 原則の組み込みを諦めたとしても，FAI が実現できるかどうかは，AGI の実現方法に依存することになる。AGI が古典的 AI の延長上に実現されるのであれば，AGI に対する制約としてプログラミングすれば，FAI を実現できるであろう。しかしながら，AGI が従来の AI の実現方法とは本質的に異なる方法で実現されるのであれば，制約を組み込むことは容易ではない。AGI が，外界の情報を自律的に認識し，最適な行動を自ら決定するのであれば，AGI の外部の観測者から見れば，人間と同じように，絶えず学習する主体として映るはずである。もし AGI も人間と同じように概念化が行われるのであれば，必然的に，AGI 内部で構成される概念は動的に変化するはずである。一度教えたら，それで学習完了というわけにはいかない。それゆえ，AGI に対する教育は，人間に対する教育と同様に困難なものとなるであろう。

４．シンギュラリティ時代の社会制度設計

これまで述べてきたように，AGI が実現され，超越知性が登場するシンギュラリティ時代においては，雇用機会喪失の可能性が大きくなる。雇用機会喪失は，人々の生存を危うくすることになり，最悪の場合には，人間社会の崩壊を招くことにもなりかねない。一方，AGI 自身に人類を保護することを保証させるような，FAI を実現することは現在の技術では容易ではない。その技術的な困難さは，人間による困難さであるので，AGI が実現されれば，AGI 自身に FAI を設計させるということも可能となるかもしれない。しかしながら，筆者は，問題の本質が FAI の定義自身にあると考える。愛らしい容貌の本体を有するロボットであれば，外観こそ友好的であっても，人類社会の破壊者として行動するように，実質が伴わない FAI の実現には意味がないからである。本来，人類に利便性をもたらすかどうかということは，客観的に定義可能な概念ではなく，恩恵を受けるものによって異なる主観的なものである。したがって，人類に利便性をもたらすかどうかということは，社会的合意としてか決定できないはずである。そうであれば，FAI を社会と切り離して定義することはできないのである。

AGI が誕生し，超越知性が登場するということは，人類がこれまで経験したことがないような，人類社会に大変革をもたらす可能性がある。その結果，シンギュラリティ時代においては，従来の価値観や幸福感を維持することが困難となり，人類は新しい価値観や幸福感を創造しなければならなくなるであろう。と同時に，シンギュラリティ時代に突入すれば，技術が劇的な速さで進歩するので，人類がその進歩に取り残されてしまうことにもなる。そうなれば，価値観や幸福感の創造とともに，社会制度を再設計しなければならなくなるのである。

現代社会では，この 20 年余り，情報技術の急速な進展に追従するために，社会制度を徐々に変革してきたが，様々な法的規制を課すことによって，社会の劇的な進展が抑制されてきたことも事実である。ネット社会に対応するために，法律が改正されたのも高々この 20 年程度に過ぎない。このため，我々は法整備の遅れを何度も経験してきた。たとえば，著作権法が，情報関連サービスのために改正されたのは，高々5 年前の平成 22 年に過ぎない。驚くべきことに，この改正までは，検索エンジン，中継のためのキャッシュサーバやテキスト分析等が，著作権法上の明文の規定として合法化されていなかったのである。ネット関連の法改正の遅れが，ビジネスチャンスを喪失させたことは想像に難くない。

IT の技術革新の速度が分単位あるいは秒単位であるとすれば，シンギュラリティ時代の技術革新の速度は，ミリ秒単位あるいはナノ秒単位で行われることが予想される。シンギュラリティ時代の技術革新の速度は，従来の技術革新の速度とは桁違いなのであるから，社会制度の追従の困難さは現在の比ではない。したがって，AGI の技術開発では，従来の技術開発とは異なり，新たな社会設計と並行して進めて行く必要があると思われる。すなわち，AGI 研究者は，社会科学者と協働し，人類にとっての幸福とは何か，望ましい社会とは何かを社会的に合意しつつ，社会設計を同時に行っていくのが望ましいのではないだろうか。FAI とは，社会的合意によってもたらされるものであり，社会的合意が前提になると筆者は考える。

５．おわりに

本論文では，AGI の実現が雇用機会の喪失を招き，必ずしも人類の幸福をもたらさない危険性について論じた。筆者は危険性があるからといって，AGI 研究の即時の中止を主張するものではない。むしろ，全世界の AGI 研究を止めることができない以上，AGI 研究を積極的に展開するべきであろう。ただし，AGI の危険性を社会と共有しつつ，シンギュラリティ時代に耐え得る人類社会を構築していかなければならない。そのためには，人工知能研究者だけで AGI 研究を進めるのではなく，社会科学者と協働して，社会設計に積極的に関与しながら，AGI 研究を進める必要があるのではないだろうか。

謝辞

本研究の一部は，科研費課題番号（24650568）の助成による。

参考文献

[1] R. Kurzweil, The Singularity Is Near: WHEN HUMANS TRANSCEND BIOLOGY, PENGUIN BOOKS, (2005)

[2] E. Brynjolfsson and A. McAfee, Rage Against The Machine: How the Digital Revolution in Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy, Digital Frontier Press, (2011)

[3] M. Ford, Rise of the Robots: Technology and the Threat of a Jobless Future, Basic Books, (2015)

[４] 山川，市瀬，井上，汎用人工知能が技術的特異点を巻き起こす，電子情報通信学会誌， Vol. 96, No. 3, pp. 238-243 (2015)
[５] 井上，人工知能に労働を奪われる日に備えよ，週 刊エコノミスト，56/57, 2014 年 10 月 7 日号，(2014)
[６] C. Chace, Surviving AI: The promise and peril of artificial intelligence, Three Cs Publishing, (2015)