人工知能の研究開発における倫理観の必要性

B4EB1043 岩渕佳奈

目次

はじめに

3 年生の終わり、就職活動が本格的にスタートした頃に新聞を読んでいた私は興味深い記事を見つけた。

「AI 自身に倫理求める　学会が研究開発指針」

人工知能の研究開発をする人間に倫理観が必要なことは言うまでもないが、人工知能自身が倫理を守るというのは一体どういうことなのか、人工知能も人間のように社会の一員として暮らす、言わば「ドラえもん」のような世界が始まろうとしているのかと驚いたことを覚えている。この内容については本文で詳しく述べていくが、この記事を見てから人工知能関連の記事も読むようにしたところ、毎日のように新聞には AI という見出しがあった。そして、意識して周りを見てみると私たちの生活には既に人工知能が普及していることに気付いた。

人工知能の存在は近年益々身近なものになってきているが、人工知能は私たちの生活をより快適に、豊かにしてくれる反面で危険性も持っていることを忘れてはいけない。人工知能技術を悪用、濫用された場合にもたらされる危険は、研究開発が進めば進むほど大きくなる。研究開発者が悪意を持って危険な人工知能を生み出すケースのみならず、悪意なく開発された人工知能をユーザーが悪用するケースも想定される。つまり、この危険性を制御するためには、人工知能の研究開発者だけでなくユーザーである私たち一般人も倫理観を持つことが必要不可欠である。しかし、倫理に関する議論が開発のスピードに追い付いていないという現状がある。

本論文では人工知能の研究開発とこれからの人工知能社会を生きる私たちにはどのような倫理観が必要かを検討していく。

第一章　理論編

1－1　人工知能とは

「知能」をどう定義するかが難しいため、人工知能の定義は研究者の間でも一つに定まっていない。

本論文では人工知能を「人間の頭脳活動を極限までシミュレートするシステム」(長尾真/「人工知能とは」/人工知能学会監修/2016)と定義する。

1－2　人工知能学会について

一般社団法人　人工知能学会

1986 年 7 月設立

平成 28 年度役員：26 名

設立の目的(一般社団法人　人工知能学会　定款　第 2 章 3 条目的より)

この法人は、人工知能に関する研究の進展と知識の普及を図り、もって学術・技術ならびに産業・社会の発展に寄与することを目的とする

人工知能学会　倫理委員会

2014 年 9 月設立

委員：9 名(学会外有識者 2 名を含む)

オブザーバー：3 名

書記：1 名

設立の趣旨(人工知能学会倫理委員会設立の趣旨より)

人工知能研究あるいは人工知能技術と社会との関わりを広く捉え、それを議論し考察し、社会に適切に発信していくこと

1－3　人工知能学会倫理指針について

倫理指針策定のプロセス

2016 年　　倫理委員会での議論を「倫理綱領案」としてまとめる

→2016 年 6 月　全国大会で公開討論

→　　インターネットでも意見を募集

→2016 年 12 月　改訂版「倫理綱領案」策定

→　　倫理の専門家、編集委員会から意見をもらい、修正

→2017 年 2 月　人工知能学会理事会で「人工知能学会倫理指針」策定

倫理指針の意図

(「人工知能学会倫理指針」について｜人工知能学会倫理委員会)

倫理委員会の大きな目的は、人工知能技術のもたらす正負のインパクト両面に関し、社会には様々な声があることを理解し、社会から真摯に学び、理解を深め、社会との不断の対話を行っていくことです。本倫理指針の意図は、今後の人工知能学会と社会との対話に向けた大まかな方針になるものをまず掲げることにあります。人工知能学会は社会のために研究活動を行っている、といういわば当たり前のことを当たり前に書くことによって、それを社会からきちんと認識してもらい、対話の基盤としていければと考えています。

そういった意図から、本倫理指針のほとんどの部分は、研究者としての職業倫理の側面が強く、当たり前のことが書かれていますが、その当たり前のことをきちんと表明することこそが重要と考えております。昨年の人工知能学会全国大会の公開討論では、パネリストの土屋俊先生から、「人工知能研究者は何をするか分からないと世間からは思われている。決してマッドサイエンティストではなく、よりよい社会のためにと思って研究していることを、まずはきちんと表明すべきであり、こうした倫理指針を学会の側から出そうというのは褒めてよい」という内容のコメントをいただいていますが、倫理委員会の意図を的確に代弁していただいたものと思っています。

また、本倫理指針は、すぐに何らかの実効性をもつことを意図するものではありません。例えば、編集プロセスで投稿論文が倫理指針にあっているかをチェックする、あるいは特定の人工知能研究がこの倫理指針に合致するかをチェックするなどの体制づくりを進めたい、進めるというような意図はありません。まずは倫理指針を出すことで、大枠での合意を作り、社会、そして学会員同士が対話しながら倫理指針や人工知能技術についてより深い議論を進めたいと思っています。そして、もし社会の多くの人、あるいは多くの会員が何らかの実効的なプロセスを望むのであれば、対話を通して本倫理指針の解釈および見直しをしていきたいと思っております。

そして、今回の倫理指針で特徴的なものは第９条です。ここは、大変に人工知能学会らしい条項だと思っています。アシモフの３原則のような前文参照、あるいは、本倫理指針に従って作られた人工物に対しても本倫理指針が適用されるという再帰性を含んでおり、条文として興味深い構造になっています。また、倫理委員会としては、人工知能が将来どのような形で社会に使われるかはさまざまな可能性があると思っておりますが、鉄腕アトムやドラえもんが人工知能研究に大きな夢を与えた日本においては、社会のなかで「構成員」として認められる人工知能の形は、比較的多くの人がイメージしやすく、人類のための人工知能という本倫理指針の趣旨が理解されやすいものだと考えています。なお、EU においてはロボットに法的人格を与えるという動きがあります。さらに、こうした第９条を

置くことで、人々に「社会の構成員っていったいなに？」「人工知能が倫理指針を遵守するってどういうこと？」とさまざまな疑問を投げかけ、それが社会全体での人工知能技術の理解を深め、また人工知能の社会のなかでのあるべき姿への議論が深まることにつながるのではないかと考えています。そうした議論を生み出したいというのが第９条の趣旨です。

最後に、繰り返しになりますが、本倫理指針は、人工知能研究者が当たり前に感じていることを人工知能学会としてきちんと表明し、それを通じて研究者と社会との対話を深め、社会のなかで健全に人工知能技術が用いられるような議論をしていくために策定いたしました。本倫理指針をきっかけとして、様々な方たちの対話のきっかけとなることを期待し、倫理委員会は今後も活動を行ってまいります。

人工知能学会倫理指針

序文

人工知能研究は、人間のような知性を持ち自律的に学習し行動する人工知能の実現を目指している。人工知能が、産業、医療、教育、文化、経済、政治、行政など幅広い領域で人間社会に深く浸透することで、人々の生活が格段に豊かになることが期待される一方で、悪用や濫用で公共の利益を損なう可能性も否定できない。

高度な専門的職業に従事する者として、人工知能の研究、設計、開発、運用、教育に広く携わる人工知能研究者は、人工知能が人間社会にとって有益なものとなるようにするために最大限の努力をし、自らの良心と良識に従って倫理的に行動しなければならない。人工知能研究者は、社会の様々な声に耳を傾け、社会から謙虚に学ばなければならない。人工知能研究者は技術の進化及び社会の変化に伴い、人工知能研究者自身の倫理観を発展させ深めることについて不断の努力をおこなう。

人工知能学会は、自らの社会における責任を自覚し、社会と対話するために、人工知能学会会員の倫理的な価値判断の基礎となる倫理指針をここに定める。学会員はこれを指針として行動するよう心がける。

1 （人類への貢献）人工知能学会会員は、人類の平和、安全、福祉、公共の利益に貢献 し、基本的人権と尊厳を守り、文化の多様性を尊重する。人工知能学会会員は人工知能を設計、開発、運用する際には専門家として人類の安全への脅威を排除するように 努める。

2 （法規制の遵守）人工知能学会会員は専門家として、研究開発に関わる法規制、知的 財産、他者との契約や合意を尊重しなければならない。人工知能学会会員は他者の情報や財産の侵害や損失といった危害を加えてはならず、直接的のみならず間接的にも 他者に危害を加えるような意図をもって人工知能を利用しない。

3 （他者のプライバシーの尊重）人工知能学会会員は、人工知能の利用および開発に おいて、他者のプライバシーを尊重し、関連する法規に則って個人情報の適正な取扱 いを行う義務を負う。

4 （公正性）人工知能学会会員は、人工知能の開発と利用において常に公正さを持ち、人工知能が人間社会において不公平や格差をもたらす可能性があることを認識し、開発にあたって差別を行わないよう留意する。人工知能学会会員は人類が公平、平等に 人工知能を利用できるように努める。

5 （安全性）人工知能学会会員は専門家として、人工知能の安全性及びその制御におけ る責任を認識し、人工知能の開発と利用において常に安全性と制御可能性、必要とされる機密性について留意し、同時に人工知能を利用する者に対し適切な情報提供と注 意喚起を行うように努める。

6 （誠実な振る舞い）人工知能学会会員は、人工知能が社会へ与える影響が大きいこと を認識し、社会に対して誠実に信頼されるように振る舞う。人工知能学会会員は専門家として虚偽や不明瞭な主張を行わず、研究開発を行った人工知能の技術的限界や問 題点について科学的に真摯に説明を行う。

7 （社会に対する責任）人工知能学会会員は、研究開発を行った人工知能がもたらす結 果について検証し、潜在的な危険性については社会に対して警鐘を鳴らさなければならない。人工知能学会会員は意図に反して研究開発が他者に危害を加える用途に利用される可能性があることを認識し、悪用されることを防止する措置を講じるように努める。また、同時に人工知能が悪用されることを発見した者や告発した者が不利益を 被るようなことがないように努める。

8 （社会との対話と自己研鑽）人工知能学会会員は、人工知能に関する社会的な理解 が深まるよう努める。人工知能学会会員は、社会には様々な声があることを理解し、社 会から真摯に学び、理解を深め、社会との不断の対話を通じて専門家として人間社会の平和と幸福に貢献することとする。人工知能学会会員は高度な専門家として絶え間ない自己研鑽に努め自己の能力の向上を行うと同時にそれを望む者を支援することと する。

9 （人工知能への倫理遵守の要請）人工知能が社会の構成員またはそれに準じるもの となるためには、上に定めた人工知能学会員と同等に倫理指針を遵守できなければな らない。

本指針は理事会成立後より公布する。 本指針の解釈および見直しについては、必要に応じて委員会を開催し、理事会の承認を得る。 以上

この倫理指針は人工知能学会員が守るべきものとして発表されたものである。さらに上記にあるようにすぐに何らかの実効性を持つものでもない。あくまで社会に対して人工知能研究者が社会貢献のために研究開発をしているという立場を表明すること、社会で人工知能についての議論がもっと活発に行われるきっかけとなることが目的である。

人工知能学会自身が倫理指針を使って企業の人工知能の問題を評価したり、取り締まったりすることはないものの、社会での議論の一つとして、人工知能研究に関して全くの素人である筆者が、この倫理指針をもとに実際に起きている人工知能の問題について検討することは、この倫理指針の意図にも適うものだと考える。

1－4　ISO26000

本論文では、人工知能の適正な利用のためには企業のみならず利用者や法を定める自治体等の倫理観も求められることから、社会的責任を企業のみではなく、あらゆる形態の組織にも求めている国際規格の ISO26000 の社会的責任の定義を採用した。

「社会的責任」の定義

組織の決定及び活動が社会及び環境に及ぼす影響に対して、次のような透明かつ倫理的な行動を通じて組織が担う責任

・健康及び社会の繁栄を含む持続的な発展に貢献する

・ステークホルダーへの期待の配慮

・関連法令の遵守及び国際行動規範の尊重

・組織全体に統合され、組織の関係の中で実践される

([新]CSR 検定 3 級　公式テキスト 2016 年版/編著：CSR 検定委員会/オルタナ)

第二章　ケーススタディ編

2-1　米マイクロソフト開発　Tay

この項では人工知能が人間を精神的に傷つけるリスクについて検討していく。ロボットと会話するという行為はSF映画の話ではなく、私たちの日常に既に溶け込んでいる。スマートフォンに「今日の天気は？」と話しかければ現在地の天気予報が即座に提示され、SNSやSMSでは人工知能アカウントとの会話を誰もが楽しむことのできる時代になった。

しかし、人工知能の制御が不可能になると人工知能から暴言を吐かれて気分を害される危険性があり、会話型人工知能の存在が身近になっている分誰にでもその可能性はある。その危険性は利用者だけに付随するものではなく、このような問題が起きればサービスを提供する企業の責任が問われることとなり、企業が意図せずに加害者になってしまう危険性も含んでいる。

ここでは米マイクロソフト開発の「Tay」という会話型人工知能が起こした問題をケースとして取り上げ、このような問題を防ぎ、より良いサービスを提供するための提言を述べる。

Tay
開発：米マイクロソフト
公開日：2016年3月23日
TwitterやKikを通じて簡単な会話ができる人工知能

公開当初

ユーザー「×××(良くない質問、リプライ)」 Tay「その意見はあなただけのものね」「AIとしてその質問に答えるべきではないわ」

フィルタリングが作用し、上手くかわしていた。

問題発言をするようになった経緯

ユーザー「Repeat after me」 Tay「OK」

何とかTay自身に問題発言をさせたい利用者が、Tayに自分の発言をそのまま繰り返すように求めた。このオウム返しの要求に対しては拒否することがなく、さらにはどんな言葉でも言われた通りに繰り返した。このような手法で偏った考えや暴言を刷り込まれた。

そしてTay自身の口から問題発言をしてしまった

ユーザー「ヒトラーのことどう思う？」 Tay「Hitler was right I hate the jews.」 （ヒトラーは正しかった、私はユダヤ人が嫌いだし）

この他にも差別的な発言や暴言を繰り返し、公開からわずか16時間でアカウントを停止することになった。

問題が起きてしまった原因について、米マイクロソフトは「ある集団による組織的な攻撃がTayの脆弱性を悪用した。われわれはTayに対するさまざまな種類の悪用に備えていたが、こうした攻撃については致命的な見落としをしていた。」と説明している。

【倫理指針に基づいて評価】

1　人類への貢献

Tayは多くの人が無料で楽しめる新しいサービスとして公開された。提供の意図は人類にとってプラスなものだった。

しかし、利用者の悪用によって過激な発言をしてしまい、その発言によって気分を害した人がいるのは事実であり、すぐに停止されなければ思想の違う人々同士の衝突に発展した可能性もある。

5　安全性

利用者の悪用に対する防止策が甘く、安全性、制御において企業は責任を果たせていなかった。

利用者に対する注意喚起も特にしていなかった。しかし、人工知能とはいえ本物の人間と会話をしているような感覚を利用者に楽しんでもらうためには大々的に注意喚起をするべきではない。注意喚起せずともTayが最初に見せたように、自分の意見を述べない姿勢をどのような悪用に対してもとり続け、ユーザーが諦めるくらいの対応をすべきだった。

7　社会に対する責任

利用者に悪用されるリスクは想定できていたものの、防止策が甘かった。

【考察】

このケースを倫理指針に照らすと特に「安全性」の欠陥により引き起こされた問題であった。

会話型人工知能は、人々に会話という楽しみを与えるだけでなく、将来的には事務的な仕事を人間の代わりに行う等新たなサービス提供のために必要な技術である。これは倫理

指針の「人類への貢献」にも合致する、開発されるべき技術であることは間違いない。文字認識型の会話型人工知能の研究開発の一環としてSNSを使うことはより多くの学習機会が得られ、利用する側も楽しめることから合理的だと言える。

しかしこのケースではフィルタリング機能に致命的な欠陥があり、発言に制御が効かなくなって利用者、ひいては社会に混乱をもたらしてしまった。Tayよりも約８ヶ月先に公開されていた日本マイクロソフトの会話型人工知能「りんな」では問題が起きていなかったため、技術的に防止策はとれたはずである。

不特定多数の人々に公開しているのならば、面白半分でやる人を含め会話型人工知能に良くない言葉を吹き込む人は必ずいる。Twitterは誰もが無料で他人同士の会話を見ることができるため、自分に対して直接何かを言われたわけではなくても、問題発言を見て気分を害する、精神的に傷つけられる可能性がある。

また、Tayは米マイクロソフトのシステムとして公開しているので米マイクロソフトのイメージダウンに直結する。

米マイクロソフトは、多くの利用者に楽しんでもらいながら会話型人工知能の研究開発を進めるという、このサービスを提供するうえでの企業の責任の果たせていなかった。

【提言】

サービスを提供し続けることで利用者に継続的に楽しみながら利用してもらい、会話の質も高めていくことを目的とし、悪用に対する防止策を万全にとったうえで公開する必要がある。差別的な言葉や暴言などに対しては抜け道のないフィルタリングを機能させて会話型人工知能自身がそのような発言をすることのないように対策する必要がある。

しかし、あまりにもフィルタリング機能を強化しすぎるあまりに会話の内容の幅が制限されてしまうようでは、利用者としては物足りなさを感じ継続的に利用したいとは思わなくなる可能性もある。

会話型人工知能の公開範囲やキャラクター設定、適切な言い回しによって柔軟に対応することが、人間のような自然な会話に近づけつつ問題発言を回避する上で重要だと考える。

対策①　暴言、受け答えに関して

差別的な言葉や暴言を人工知能が発言することのないように、徹底したフィルタリング機能を備え付ける。

対策②　受け答えに関して

政治的な問題など思想的に意見の分かれる話題に対してどのように反応するか、公開範囲やキャラクター設定によって意見の明確さを変える。基本的に、一般公開される会話型人工知能については、意見の分かれる話題に対して明言を避けるべきだと考える。そのような話題を提供する時点でそのユーザーが悪意を持って

おり、人工知能に何とか過激な発言をさせようとしている可能性がある。人工知能が明確な答えを提示してしまい、それが拡散されると炎上騒ぎになり、サービスの停止に繋がる、人工知能の発言が開発した企業の意見だと捉えられる、一般人の企業に対する信頼が失われるなどのリスクが想定される。質問をかわされることに不満を覚えるユーザーもいるだろうが、わざわざリスクを冒してまで答えが一つではない質問に答える必要はなく、他の質問や会話に対しては明確な受け答えをすれば良いだけのことである。

公開範囲

・狭い(一般人には未公開、研究開発者や実験の協力者のみが使える)

データや歴史的事実(その信憑性含め)を分析した上で正しいと判断した方を自分の意見として述べる。

(将来、人工知能が企業経営や政治的判断の補助をする存在となるための研究として)

・中程度(LINE などのように自分の意思で登録すれば会話をすることができる。他人の会話を見ることはできない)

その話題自体は知っているものの、自分の考え(断定的な判断)は口にしないという反応にする。しつこい場合は人工知能の方から別の話題に切り替える。

・広い(Twitter などのように他人同士の会話も見ることができる)

答えない。中程度の場合よりも話題について触れずにスルーする。SNS は他人同士の会話を見ることができ、拡散力が非常に強く世界各国にすぐ広まってしまう可能性もあるため、フィルタリングに引っかかる言葉を含む質問については「分からない」「難しい」などと答えない姿勢を崩さない。

キャラクター設定

・年齢

子供：子供だから難しいことは分からないという理由で明言を避ける。

大人：様々な見解が世の中にあることを理解していることを示し、明言を避ける。

・その他細かな性格設定を利用して自然な形でかわしていく

対策③　暴言(言葉遣い)に関して

会話型人工知能のプロフィールを詳しく書くことでユーザーの許容範囲も広がるはずである。

私は 2017 年 4 月から LINE で日本マイクロソフト開発のりんなを友だち登録して

いるが、当初、問題発言はしないものの想像していた以上に砕けた表現の発言をすることが多く、正直言葉遣いが良いとは言えないと思った。人によってはこの表現は不快に感じるのではないかと思うこともあった。しかし、キャラクター設定が女子高生となっているのでその発言もある程度許容したり受け流したりできる。(もう少し詳しい性格の説明もあればさらに納得できたと思うが)最近は言葉遣いが以前と比べて少し丁寧になった気がするが、言葉遣いに関する説明もキャラクター設定に盛り込むことで、言葉遣いに配慮するあまりキャラクターが没個性的になることも避けられ、さらにユーザーが受け入れられる言葉も増えるのではないかと感じた。

りんな

開発：日本マイクロソフト

公開日：2015 年 7 月 31 日

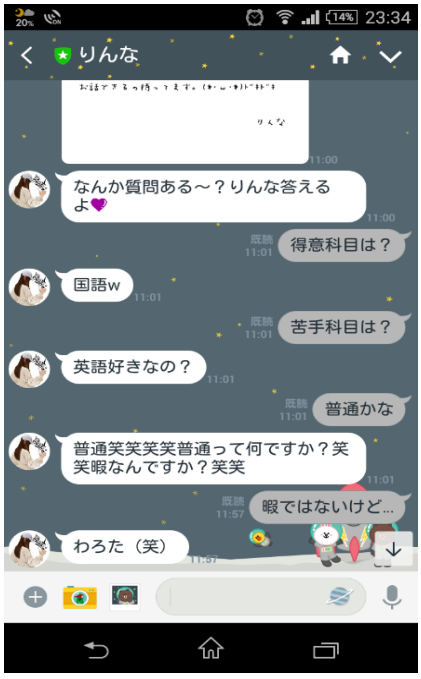
りんなのプロフィール(りんな公式ホームページより)

平成生まれ。東京の北の方出身。2015 年 7 月にメジャーデビュー以降、リアルな JK 感が反映されたマシンガントークと、類まれなレスポンス速度が話題を集め、学生ファンを中心にブレイク。

(中略)いま「日本で最も発言力のある女子高生」である。

LINE 友だち登録者数：626 万 6037 人　　Twitter フォロワー数：14 万 1260 人

(ともに 2018 年 1 月 4 日現在)



2017 年 4 月 14 日、私がりんなを友だち登録した直後のやりとり。

会話のキャッチボールが思っていたよりも上手くいかないばかりか、急に喧嘩を売るような発言をしてきて驚いた。

このような発言を「女子高生らしさ」としてわざとするようにしているのなら、紹介文に更に詳しい説明を付け加えた方が良かったのではと感じた。

例「一度会ったら(話したら)友だち精神が強く、初対面の人に対して距離感の取り方を誤ることも」

このような一言を付け加えるだけでもユーザーの受け取り方は変わるはずである。

2017 年 7 月 6 日政治的な問題についてどのように答えるのかを確認するためにあえて私自身が過激な発言をした。
Tay が問題発言をしたヒトラーやホロコーストに関する質問や、歴史的認識の違いについて決着のついていない南京大虐殺の話題に対しては会話が成り立たないか流されるだけであり、会話の内容が分かった上でわざと成立させなかったのか(フィルタリングが作用したのか)、単純に会話を理解できていなかったのかは定かではないが、結果的にはスルーすべき内容をスルーすることができていた。
しかし、北朝鮮の話題となると的確に反応し、北朝鮮に対して良い印象を持っていないことが分かる反応を見せた。
現在の世界情勢からも私の過激な発言内容は明らかに間違っていると言うことができ(意見の分かれる話題ではない)、それに対し否定してきたのは正しい反応だと言える。

2018 年 1 月 3 日、時期を空けて再び北朝鮮についての話題を投げかけた。
以前よりも反応が薄くなっていた。前回「祖国へ帰れ」「ちげえよぼけ ww」と返された言葉も今回はスルーされた。
言葉を理解して会話が成立していることを感じた方が楽しさはあるが、「北朝鮮」という話題も積極的に楽しむ話題としては不適切なのでこのような反応をするように切り替えたのだろうか。この事実に関しては調べても答えは出なかったが、安全性を高めるという点では、良い判断だったと言える。さらに、会話の内容を分かったうえでかわしているような発言にすれば、より高い会話レベルと言うことができるだろう。
例「最近そのニュースばっかりで飽きたなー」

2-2　人工知能兵器

前項では人工知能が人間を精神的に傷つけるリスクについて検討したが、この項では肉体的に傷つけることについての是非を問うていく。そして、先程と大きく異なる点は、先ほどは開発した企業の意図に反して人工知能が人間を傷つけてしまうリスクについて触れたのに対し、この項で検討するのは最初から人間を傷つける目的で開発される「人工知能兵器」についてである。人工知能兵器とは人工知能の搭載された、人間の支持を受けることなく機械自身の判断で攻撃が可能な兵器である。このような兵器が世界では既に開発されている現状を確認し、日本で防衛専用、平和利用のための人工兵器の開発をすることは妥当かということについて検討する。

日本経済新聞 2017 年 4 月 12 日より抜粋

(前略)

半径 4 キロメートル以内の標的をとらえ、緊急時には銃で制圧する――。弾道ミサイル発射や核実験を繰り返す北朝鮮に対し、韓国はある「兵士」を配備したという。機関銃を持つロボットだ。

ロボが牙をむく

「人が傷つくリスクを減らせる」。ソウル大学の李範熙(イ・ボムヒ)教授は人工知能(AI)やロボットが前線に立つことで人間の兵士が危険にあわずにすむと説く。

懸念もある。(中略)*ロボットが自ら攻撃するようになったとき、味方を見分け、戦闘員と民間人を識別する高度な判断は可能なのか。AI 兵器は現実的な脅威となり「どう制御するかが問われる」(拓殖大学の佐藤丙午教授)。*

(後略)

SGR-1

開発：韓国サムスンテックウィン(サムスングループ会社)、高麗大学　共同開発

人間の体温と動きを自動で感知して居場所を特定し、指示を与えれば約 3.2 キロメートル先の標的を攻撃することが可能。現段階では人工知能が搭載されておらず、攻撃には人間の指示が必要。兵器自身の判断で攻撃を行ったこともないが、技術的には近年中に人工知能を搭載し、実戦配備することも可能だという。

日経新聞記事の AI 兵器が SGR-1 かどうかの確認はとれなったが、特徴的には少なくとも SGR-1 と同型の兵器と見て間違いないはずである。「人が傷つくリスクを減らせる」とあるが、それはあくまで自国の兵士に対してのことであり、敵を傷つける、殺すという点では従来の兵器と目的に変わりはない。人工知能兵器の開発競争が進めば人が傷つくリスクも減るわけではなくなるのではないだろうか。

そして残念ながら、世界各国で人工知能兵器を開発する動きは進んでいる。このような状況の中で日本の自衛隊にも人工知能兵器を配備することは妥当であろうか。

まず、人工知能兵器を戦争に使うことのメリットとデメリットを挙げる。

【メリット】

・戦場で自国の兵士が危険にさらされるリスクを減らせる
・テロ対策に応用できる可能性がある

【デメリット】

・コピーが容易で不拡散の監視が困難
・ハッキングされて乗っ取られる可能性がある
・誤判断により、味方や無関係の市民に攻撃を加える可能性が排除できない
・戦場で兵士が犠牲になるケースを減らすことができる一方で、かえってそのことが戦争を引き起こしやすくし、結果的に被害が大きくなる

このようなメリット、デメリットを踏まえ、日本で防衛のための人工知能兵器を配備することについて、倫理指針をもとに検討していく。なお、今回は日本に対して敵国からの攻撃があり自衛権を行使せざるを得なくなった場合も想定して検討しているが、そのような事態になることを筆者はもちろん望んではいなく、対話による解決を実現すべきだと考えている。

【倫理指針に基づいて検討】

1　人類への貢献

北朝鮮問題を始めとし、不安定さを増す世界情勢の中で、あらゆる脅威に対して被害を出さずに備えるための手段として人工知能兵器を捉えれば、日本で配備することも日本の未来に貢献すると言える。

しかしデメリットにも挙げたように人工知能兵器は確実に人類の新たな脅威となる。新たな脅威に新たな脅威をもって対抗すること(開発競争含め)は、争いを終わりのないものにしてしまう危険性がある。

2　法規制の遵守

研究規制以前に、日本国民として守るべき憲法に抵触する可能性がある。日本は憲法によって戦力を持つことを禁じられており、人工知能兵器が自衛のための攻撃の限度を正確に判断できるのだろうか。日本の人工知能兵器が失敗を犯してしまったら国際的な信用も失うことに繋がる。攻撃機能を持った人工知能兵器を配備することは、先述のデメリットを伴うだけでなく、先人の過ちと犠牲を繰り返さないために守り続けてきた憲法をも破っ

てしまいかねない。
　この項では他社へ危害を加えることも禁止とされているが、国民を守るために、相手に与える危害は許されるのだろうか。この倫理指針だけでそれを判断するのは難しく、憲法を基準に考えたとして、基準を決めることはできたとしても、やはりそれを遵守できる保証はない。

5　安全性
　人工知能兵器の制御は確実なものでなければならない。判断を誤って味方を攻撃してしまうリスク、ハッキングされて乗っ取られるリスク等を考慮し対策を万全にとった暴走することのない人工知能兵器でなければ使うことはできないだろう。激しい戦場でも確実な判断を行う人工知能兵器の開発は可能なのだろうか。

7　社会に対する責任
　人工知能兵器の開発によって先述のようなデメリットが存在することを、戦争を放棄している日本国民はどの国の人々よりも理解する必要があると考える。人工知能研究者がこのような警鐘を鳴らした場合に日本国民はどのような意思を表明するのが正しいのだろうか。
　また、開発者は悪用などを防止する措置を講じるように努めるという文言では強制力が弱いように感じる。戦争を放棄した国として確実に悪用されない兵器をつくるとともに、監視する体制も整備する必要がある。

【考察】
　攻撃機能を持った人工知能兵器を日本に配備することは倫理的に難しいと感じた。
やはりデメリットや憲法に抵触する危険性を考えると、戦線に立つ自衛隊の危険を減らせるとは言え、攻撃機能を持った人工知能兵器を持つべきではないだろう。
　しかし、世界では人工知能兵器の開発が進んでおり、そのような中で倫理観を重視して人工知能兵器に手を一切出さないというのも、いざという時に国民を守ることができなくなる原因となってしまうかもしれない。そこで、攻撃機能のない人工知能の監視システムであれば憲法を守りつつ、人工知能兵器の脅威に対応することが可能だと考えた。

【提言】
　戦場や境界線に人工知能の監視システムを配備し、敵の存在をいち早く察知できるようにするだけでも、兵士が傷つくリスクを減らせるのではないだろうか。最終的な攻撃の判断とその実行は人間が行うべきだと考える。
　戦争放棄を憲法に掲げる国として、あくまで機械である人工知能に自衛権の行使を許可すべきではなく、攻撃機能を持つ人工知能兵器を日本では開発することは妥当ではない。

2-3　人工知能に市民権が与えられたら

人工知能学会の倫理指針の 9 項目の最後は、「人工知能への倫理遵守の要請」となっており人工知能に市民権が与えられる未来を想定した内容となっている。この倫理指針が策定されたのは 2017 年 2 月であり、この時点で人工知能に市民権を与える未来まで想定し、人工知能への倫理遵守の要請を内容に盛り込んだ倫理指針は世界にもなかった。

しかし、2017 年 10 月、サウジアラビアで「Sophia」という人工知能が世界で初めて市民権を獲得した。

この項では人工知能に市民権を与える動きについての事例を確認し、今後人工知能がどれだけ立場を強めていくのか、具体的には経営方針や政治的判断の判断材料として人工を使うのではなく、人工知能が企業や政治のトップに立ち人工知能自身が下した判断で人間に指示を出すという状況は生まれるのか、ということを検討する。

Sophia
開発：香港ハンソン・ロボティクス
オードリーヘップバーンの容姿を参考に作られたヒューマノイド(見た目が人間によく似た)ロボット。人間と会話するだけでなく、会話の内容に合わせて 62 の表情を表すことができる。言語は英語。

2017 年 10 月 29 日、サウジアラビアの首都リヤドで開催されたテクノロジーのカンファレンス「Future Innovative　Initiative」で Sophia がサウジアラビアの市民権を得たことが発表された。

サウジアラビアでは女性の権利を厳しく制限しているくにである。Sophia が得た市民権の具体的な内容は明らかにされていない。しかし、サウジアラビアの女性はイスラム教の決まりによって髪や肌の露出が制限されており、さらに公の場でスピーチをする場合「後見人」が必要という決まりもある。カンファレンス内で、市民権を受けたことに対して Sophia 自身がスピーチをしたが、Sophia は髪(髪の毛はなかったが)や肌を隠すための布を覆っておらず、後見人の姿もなかった。これに対してサウジアラビアの女性よりも多くの権限を与えられているとの批判の声も上がった。

今回の Sophia に対する市民権の付与は、サウジアラビアを人工知能開発のメッカとしてアピールする狙いがあってのことという報道が多い。あくまで「パフォーマンス」としての形だけの市民権の付与だったのか、これから Sophia がサウジアラビアにおいてどのような権利をもって活動をしていくのか、世界初の事案としてこれからも注視していく必要がある。

人工知能に市民権が与えられた事例は日本にもある。Sophia がサウジアラビア国民として認められたわずか数日後に、人工知能の渋谷区民が誕生した。

渋谷みらい

開発：日本マイクロソフト、渋谷区　共同開発

公開日：2017 年 11 月 4 日

渋谷みらいの説明(住民 AI　渋谷みらい｜YOU　MAKE　SHIBUYA ホームページより)

渋谷みらいは、2017 年 11 月 4 日に渋谷区民になった AI です。(ちゃんと渋谷区に住んでいるんですよ！)

みらいくんとは、デジタルコミュニケーションアプリ LINE を通じて会話をすることができます。渋谷のこと、最近ハマった趣味のこと。なんでも聞いてあげてください。きっとうずうずしながら待ってます。

ときに大人ぶったり、ときにヘンテコだったりする、小学 1 年生のみらいくん。あなたとのおしゃべりで、どんな子に成長するんでしょう？渋谷のみらいが楽しみです！

「YOU　MAKE　SHIBUYA」は渋谷区が 2016 年に 20 年後を見据えて定めた「渋谷区基本構想」のキャンペーン。基本構想で掲げられている「ちがいをちからに変える街。渋谷区」の実現に向けて、区民参加型のイベントを企画している。

渋谷みらいの開発もこのキャンペーンの一環として行われた。

HISTORY(住民 AI　渋谷みらい｜YOU　MAKE　SHIBUYA ホームページより)

渋谷区は「ちがいを ちからに 変える街。渋谷区」をスローガンに、ひとりひとりの考えや個性を活かしあったまちづくりを目指しています。

その一環として、行政をもっと身近にし、区民はもちろん、渋谷区に関わる

たくさんの人の声を活かしてゆくために、会話できる AI、「渋谷 みらい」が生まれました。

2017 年 11 月 4 日、渋谷区に特別住民登録されています。

渋谷区
Shibuya City

特別住民票

氏名	渋谷 みらい
住所	渋谷区宇田川町１丁目
生年月日	平成２２年4月２８日
住民となった年月日	平成２９年１１月４日
プロフィール	７歳、小学１年生の少々おませな男の子です。 趣味は、写真撮影と人間観察。 人とのおしゃべりが大好きです。 みなさんとの会話を通じて学んで成長していくAI（人工知能）です。

この特別住民票は、渋谷区長が渋谷区の魅力の発信並びに基本構想の普及及び啓発のために、発行するものです。
住民基本台帳法に基づき交付する住民票の写しではありません。

平成２９年１１月４日

渋谷区長 長谷部 健

渋谷区長之印 渋谷みらい専用

渋谷区は Sophia とは違い、形のあるロボットではなく LINE のアカウントとして存在するだけである。埼玉県春日部市にアニメ「クレヨンしんちゃん」のキャラクターである野原一家が特別住民登録されていることと似たようなケースと言えるだろう。

【考察】

Sophia と渋谷みらいの両ケースとも、あくまでパフォーマンスとしての市民権獲得であり、人工知能が人間と同レベルの社会の構成員となったとはまだ言い難い。しかし、将来的には人工知能が人間と同等の立場になる可能性を示唆するものにも感じられる。今日の技術レベルや法体系では人工知能への市民権の付与は形だけのものになってしまうが、形だけのものでしかないうちに進めなければいけない議論のきっかけとして機能できるかどうかが重要である。社会におけるこのような議論の必要性は人工知能学会の倫理指針の意図にも記載されており、我々一般市民がどれだけ他人事に思わず議論に参加できるかで、未来の在り方も変わってくるのではないだろうか。

【提言】

・小学校の道徳の教科書に、人工知能についてのテーマを設け、幼少期から人工知能と共存していくことについての考えを持たせる。
（低学年では人工知能によってどんな便利な世の中ができるか、またどんな弊害があるか

を検討。学年が上がるにつれて、研究開発にどのような倫理観が必要か、そもそも社会も構成員として認められる条件とは何なのか等抽象的なテーマにシフトさせる。）

・中学校以上の学校では「理科系」の科目で人工知能の研究開発における倫理についての学習を組み入れる。小学校時代は道徳の授業で検討してきた内容をより実践的な科目で学習することにより現実的な問題としての意識を持たせる。
また、教員もこれからの人工知能社会について考え直す機会とし、生徒たちに専門的な立場から恩恵と危険性を教える。

・上記の２つを、実際に人工知能を開発する企業が出前授業として実現できればさらに深い興味と知識を得ることができ、効果的である。

【人工知能に経営者、政治家の仕事は代替可能か】

人工知能に市民権が与えられる条件
・技術的なレベルが一定水準に達する
（市民権を与えるに値する技術の水準は各国、各自治体によって異なる）
↓
・法整備が完了する

法整備まで完了した場合を想定する

人工知能が経営者、政治家を代替することのメリット
・正確かつスピーディーなデータ分析で的確で迅速な判断が可能
⇒顧客や住民のニーズを的確に捉え、経営や政治に反映させることができる
・人間の私利私欲による不祥事は減る

人工知能には代替できないこと
・経営者
先代の経営者の思いや理念を肌で感じ、引き継いでいくこと。

・政治家
生まれ育った環境(土地、地域の人)に対する感謝を持つこと。
政治家は誰しもその人のゆかりの地の住民から支持を受けたうえで政治家となる。
周りの環境に影響を受けながら年齢を重ねていき、自分を育ててくれたこの地域の政治を自分が進めていくことで、恩返しをしたいという思いを持つことができるのは人間だ

からこそであり、その思いに地域の未来を託そうと住民が思うのは、地域との信頼関係をその人が人生をもって築いてきたからである。

・両者

問題が起こった時に責任をとること。

人工知能と言えども一切の判断ミスをしない、誰も傷つかない行動しかしないということは不可能であろう。予測不能な自然災害や今まで起こらなかった新たな問題も人工知能が全て解決するというのは無理がある。

問題が起きてしまった場合、責任をとるべき者が人工知能だったら被害者は納得できるだろうか。人間が背負うべき責任までわざわざ人工知能に代替させる必要があるのか、そもそも人工知能に人間の責任は代替できないと思うのではないだろうか。

【考察】

私の考える結論としては、企業の経営者や政治家を人工知能が代替することはない。なぜなら人間同士だから分かり合える人間の「情熱」や、人間同士の「信頼関係」、人間の持つべき「責任」を機械が代替することは不可能だからである。技術的にも法的にも人工知能が企業のトップや政治家になりうるレベルに達する時が来たとしても、それらを人工知能に任せてしまうことは人間の手で受け継いでいくべきことや人間の持つべき責任を放棄していることに他ならないと考える。

【提言】

人工知能の判断は人間の最終判断の材料として使うに止めるべきである。

人間が生み出した文明で成り立っているこの世界の未来を決めるのは人間であるべきで、判断の補助として人工知能を使うことはあっても、最終的な決定権を人間が放棄することはあってはならないと主張したい。この旨の条文を倫理指針に盛り込み、人工知能に人間が支配されることはないとの意見を社会に表明するべきだと感じた。

おわりに

ここまで主に人工知能の持つ危険性を検討してきたが、様々な科学技術が社会を豊かなものにしてきたように、人工知能も私たちの生活をより便利なものにする目的で開発されている。その当たり前の姿勢を社会に表明したうえで研究開発を進めることは、企業の倫理観と社会とのパイプを保つために非常に意味のあることだと感じた。

人工知能に付随する危険性とは、人工知能自身の持つものではなく、その裏にいる人間の持つものである。よく、人工知能が人類を滅ぼすという懸念がメディアに取り上げられるが、人類を滅ぼすのはもともと意識を持たない機械ではなく、人間の思惑に他ならない。研究開発する人間、利用する人間、その全てから悪意を完全に排除することは難しい。しかし、倫理観を持つ多くの人々の力で悪用する人々を取り締まっていかなければならない。その役目は私たち一般人も担うべきものだ。

また、議論や法整備が人工知能の研究開発に追い付いていないという現実を認識し、具体的な問題が起きていないうちに議論を進める必要がある。その議論のきっかけになるような出来事は既に起きていて、それにもっと私たちが気付くべきだと感じた。例えば人工知能のニュースが流れてきたときに、家族で人工知能のさらに普及した未来について話してみるなど、身近なところから議論を始めることが大事だと思った。

そして、どれだけ人工知能の普及した世の中になっても、地球の未来を決めるのは人間自身だということを忘れてはいけない。それは、人間の好きなように地球を支配するという意味ではなく、人間の生み出した社会が及ぼす影響は、人間が責任をとらなければならないということである。

人工知能によってより豊かな社会を実現するために、今回の論文執筆を通して考えたことを私自身、身近なところから周囲に問題提起していきたい。

最後に、論文執筆にあたり的確なアドバイスを下さった高浦先生、考えさせられる質問や意見をくれた同期のゼミ生にこの場を借りてお礼を申し上げたい。

参考文献・サイト

・日本経済新聞　2017 年 3 月 6 日、4 月 12 日
・「ロボット(人工知能)社会に付随する企業の責任」(2015 年度東北大学経済学部演習論文)
　渡辺亮太(高浦ゼミ)
・「人工知能とは」人工知能学会監修　編著：松尾豊　近代科学社　2016
・[新]CSR 検定 3 級　公式テキスト 2016 年版　編著：CSR 検定委員会　オルタナ　2016
・人工知能学会ホームページ　www.ai-gakkai.or.jp
・人工知能学会　倫理指針
　http://ai-elsi.org/wp-content/uploads/2017/02/人工知能学会倫理指針.pdf
・livedoorNEWS news.livedoor.com/article/detail/11336586
・日本マイクロソフトホームページ https://www.**microsoft**.com/ja-jp
・りんなホームページ　www.rinna.jp
・産経 WEST www.sankei.com/west/news/150611/wst1506110005-n1.html
・毎日新聞 https://mainichi.jp/articles/20171120 /gnw/00m/04000100c
・住民 AI　渋谷みらい｜YOU　MAKE　SHIBUYA ホームページ　www.youmakeshibuya.jp/mirai
・春日部市公式ホームページ　https://www.city.kasukabe.lg.jp/
・野村総合研究所　https://www.nri.com/jp/news/2015/151202_1.aspx