

特集 「ダイバーシティと AI 研究コミュニティ」

AI にジェンダーを組み込むことは どういうことか

Consider the Implications of Integrating Gender into AI

藤堂 健世
Kense Todo
東京工業大学情報理工学院
School of Computing, Tokyo Institute of Technology
K_todo@ali.c.titech.ac.jp

佐久間 洋司
Hiroshi Sakuma
大阪大学基礎工学部
School of Engineering Science, Osaka University
sakuma.hiroshi@sys.es.osaka-u.ac.jp

大澤 博隆
Hirota Osawa
筑波大学 システム情報系
Faculty of Engineering, Information and Systems, University of Tsukuba
osawa@iit.tsukuba.ac.jp, http://hiroosa.com

Keywords: AI, AI interface, HAI, HRI, gender.

1. はじめに

JR 東日本の山手線に新しく新駅「高輪ゲートウェイ駅（以下新駅）」が開業した。新駅開業とともにネット上で話題になったのが「AI」によるナビゲートである。例えば「無人 AI レジコンビニエンスストア」の導入やロボットによる警備や清掃が始まるという。そして、AI サイネージによる案内も始まった。男性駅員タイプと女性駅員タイプの 2 種類で、乗換案内や周辺の飲食店情報など声に出して質問すると答えてくれる*1らしい。

新駅では、男性と女性のそれぞれ異なる性差をもつサイネージが登場した。例えば、男性駅員タイプではリアルさを追求されたが、女性駅員タイプではアニメ的なキャラクター性を付与された。仕草も違うようだ。その結果として、AI サイネージが、Web 上でさまざまな議論を引き起こした。女性駅員タイプのサイネージは現在さまざまなところで活躍しており、特段新しいサービスではない。もしかしたら女性駅員サイネージは、社会的に受け入れられていたのかもしれない。しかしながら、今回の新駅の AI サイネージにおいては、主にジェンダーバランスについてさまざまな意見や声が生まれた*2,*3。

ここで議論を呼んだことは、社会の中に「AI」のよう

な存在が現れたことによって、それを見た人間がさまざまな「AI」像をつくり出している途中経過で発生したりアルタイムな問題ではないだろうか。ここから読み取れるのは、AI にジェンダーという役割が付与されると次のような副次的な課題を生じさせるということだ。

- 「もの」であったものに、「人格」が投射（プロジェクション）され、今までの機械（AI）という概念を揺さぶる可能性がある
- 人とのインタフェースの一形態ではなく、人々のものの見方を変えてしまうという点で、議論する必要がある
- 社会的受容性の側面をもっているため一部の研究者や開発者で完結できない

この新しい課題は、AI 研究者・技術者だけではなく、さまざまな社会の意見を交えて議論する必要があると考える。

以上を踏まえ本稿では、AI インタフェースやエージェントにジェンダー性を付与することの影響を、さまざまな角度から考察したい。

1.1 先生方のご紹介

「AI やロボットに性差・ジェンダーを組み込む」という課題に対し、多分野の研究者、技術者の視点をインタビューにすることで、社会に出たときの影響や可能性を考えるにあたって、下記の 4 名の研究者に依頼した（五十音順）。

- 江間有沙先生（東京大学未来ビジョン研究センター 特任講師）専門は科学技術社会論。特に人工知能やロボットを含む情報技術と社会の関係について研究

*1 <https://toyokeizai.net/articles/-/336547?page=4>

*2 <https://wezz-y.com/archives/74908>

*3 <https://www.asahi.com/articles/ASN3M6RPHN3MUTIL044.html>

している。

- 西條玲奈先生（大阪大学大学院文学研究科哲学哲学史助教）専門は分析哲学。特に形而上学、フェミニスト哲学、ロボット倫理学。
- 高橋英之先生（大阪大学大学院基礎工学研究科特任准教授）専門は認知科学。特に人間とロボットのインタラクションにおける心の役割などについて研究をしている。
- 野村竜也先生（龍谷大学先端理工学部教授）専門は人工知能、システム論、複雑系。特に人工知能・ロボットが個人の心理・社会に与える影響などについて研究を行う。

先生方には、貴重なお時間をいただきましたこと、この場をお借りしてお礼申し上げます。

2. 質問項目

西條先生、高橋先生、野村先生には、四つの質問を行った。AIとジェンダーに関わる問題2題と、ジェンダーに留まらず、AIと人間の関係性に関わるハラスメントとルッキズム^{*4}に関わる問題各1題について、先生方お一人ずつに時間をいただいて質問させていただいた。議論は1時間以上続いたが、本稿では確認をとったうえで、その要約を載せている。また、江間先生には、AIとジェンダーに関わる問題2題と、議論の方向性について3題を尋ねさせていただき紙面上で回答いただいた。

なお、各質問の回答についてはインタビューを実施した順に掲載している。

2-1 AIにジェンダーを組み込むことについて

AIにジェンダーを組み込むことは社会にどのような影響を与えるのか、先生方の基本的な考え方を尋ねた。

§1 野村先生

工学的にジェンダーを入れたほうがインタラクションがスムーズになるというのは以前からいわれていますが、一方で、これを導入するとステレオタイプの再生産になるという意見もあると思います。例えば、受付ロボットを女性としてデザインするという点に関して、そうしたほうがカスタムの受けが良いという場合もあれば、受付嬢というジェンダーステレオタイプの強化になるという考えもあると思います。ジェンダーステレオタイプの強化は、受付をしたい男性の希望の可能性を排除することにもなると思います。

またジェンダーというと、男性、女性という二極をベースにして動いていると思うのですが、LGBTはどうするという課題もあります。特定のロボットやAIにMale,

Female特性を入れることで、逆にジェンダーマイノリティの方々が排除される政治的側面もあると思います。このように、さまざまな課題が浮き彫りになるため、全体指針というものを考えるのは難しいのではないかと思います。人間においてもまだまだ議論が出てきているため、機械に対するジェンダー適応をどう考えるかについてもさまざまな議論が考えられます。したがって、ステークホルダーの中で議論するしかないというのが私の考えです。

ヨーロッパでは、ユーザ側の性別に配慮しつつ、ジェンダーステレオタイプを排除して、ロボットをデザインしようというプロジェクトが立ち上がっています。日本人に比べてヒューマノイドロボットに対する選好性が高いという部分もあります。また、ヨーロッパでは、ロボット、ヒューマノイドを開発するときに、ジェンダーをどうしますかという問題がはっきりと現れやすくなっていると思います。日本人はそのあたりを曖昧に考えている部分があるのかもしれませんが、国内のヒューマンエージェントインタラクションの学会に行くと女子高生のデザインのAIが普通に登場します。このようなときに「やばくないか」と思ったりもします。このような意見は開発側からすると、「自然」につくりたいのにしんどいことをなぜ考えないといけないのかという気持ちにもなると思います。ですから、ステークホルダー間での全体の議論が必要ではないかと思っています。ファシリテーターもいませんし、現状はそのような場がないと思います。

また、日本で議論になりづらいのは、研究者コミュニティにジェンダーのバイアスがかかっているからかもしれません。開発者のジェンダーバイアスが、プロダクトに影響することは間違いないと思います。だからこそファシリテーターの存在が重要だと思います。

§2 高橋先生

今のロボット研究に一番足りていないのは「ロボットとはそもそも何なのか」という定義だと思います。我々は人と話すとき、自分も持っている属性を頼りに相手を判断しています。これには社会的なラベルやアイデンティティに縛られるデメリットもあるわけですが、一方でスムーズなコミュニケーションをするガイドになるわけです。現在のロボットや人工知能には、人間が共有しているアイデンティティはない状態です。ですから、人間も持っているアイデンティティ（例えば性別）をロボットや人工知能にもたせるというのは、インタラクションを円滑にし、それらと人間の関係性をスムーズにできることが期待されます。このような内容を基礎研究として行うのは全く問題ないと思っています。

一方で、社会や文化の影響を大きく受ける社会実装において、アイデンティティをどのように組み込むかという問題は、工学や情報科学の問題だけではありません。心理学や文化人類学、哲学といった人文・社会科学と共歩調を取りながら、人間はジェンダーというアイデン

*4 ここでのルッキズムとは、外見による評価が中身の評価と混同されること、身体的魅力と、良さ・悪さが結び付けられるというステレオタイプのことを指す。

ティティをどのように捉えるのか、ロボットに対する反応はどうなるのかを慎重に議論する必要があると思います。特に公共の場においては、慎重に議論を踏まえコンセンサスを形成していく態度があるかないかだけで、受け手の印象も変わると思います。何らかの賛否が分かれそうなシステムを公共の場に設置する際、小規模なデモンストレーションを行い、その場で一般の方の意見を収集するだけではいけないような気がします。評価者がその場では感性的に「かわいい」と認めても、社会的にそのアバターを承認していることにはならないはずで、感性的な側面と、論理的な側面を区別した議論が重要になると思います。

上記のような議論をするうえでも基礎研究の知見は重要であると思います。例えば、エージェントの見た目や性差はもちろん、人間の側のジェンダーやパーソナリティがロボットやエージェントに対する接し方に影響を及ぼしています。自己開示の研究をしていると、男女でエージェントに対する態度が異なることがわかりました。我々の知見では、男性は人間とロボットで話の内容を変えています。女性は、人間とロボットのボーダーが弱い印象をもちました。このような性差などのさまざまな人工知能そのもとは関係ないファクタも考慮していかないと、噛み合わない議論になるのではないかと思います。

ヨーロッパなどでは、社会規範や倫理がトップダウンに定義される傾向が強く、そのような観点からAIやロボットにジェンダーを組み込ませることにどのような意味があるかを考えます。しかし日本では「もののあはれ」な文化といえますか、ロジカルな論理抜きで、感情や感性を重視してAIやロボットを設計するところがあります。このような傾向が、規範や倫理の観点から疑問符がつくような人工知能に対するジェンダーの付与を行わせる一つの要因かもしれません。

§3 西條先生

私は、心理学や歴史学のように経験的な調査や実験を手法とする研究分野の専門家ではないので、あくまで概念上の分析から導き出せるリスクに言及したいと思います。必要がない限り、ロボットやAIに女性や男性と思わせるジェンダー特性を帰属させないほうがよいと考えます。そのデザインは、動物やジェンダーを感じさせない中立的なもので事足りるのでないのか、そもそも人のように見える必要があるかを考えるべきでしょう。その理由は、社会に送り出す製品としてのロボットやAIを女性や男性と認識できるようにデザインすることは、女性（男性）とはしかじかだ、という理解を是認し、モデルとして提示することを含意するからです。もののデザインそれ自体が生身の人間に対するメッセージとして機能するわけです。「女性」や「男性」と呼ばれる集団のメンバは多様であり、何に自分のジェンダーアイデンティティの基盤を置くかもさまざまです。そもそも、ジェ

ンダーは男女の二つどちらかに限定されないし、一生を通じてジェンダーが安定している人ばかりでもありません。ジェンダーFに対して性質Gをひも付けるあらゆるデザインには、常に誰かのアイデンティティや存在を否定するメッセージを社会に発するリスクがあるので、ジェンダーに限った話ではありませんがね。

またそのリスクの大きさも、すべてのデザインに左右されます。特に、開発者が自分のジェンダー観への批判的検討を欠く場合、こうしたリスクはいっそう高まります。ここでいうジェンダー観とは、社会で前提とされるジェンダー区分や、女性らしさや男性らしさとして了解されている性質群のことです。ジェンダー観に無批判なままだと、ステレオタイプ、すなわち倫理的に問題のある一般化に陥りやすくなります。次の質問で詳細に答えたいと思いますが、ステレオタイプが問題視されるのは社会の多数派にとって好都合な既存の固定的なジェンダー観をそのまま肯定することにつながるからです。

どうしてもジェンダーを備えざるを得ないならば、社会のジェンダー規範、すなわち特定のジェンダーであればこのようなはずだという身体的、精神的、社会的特徴とされるものに自覚的であることが求められます。一体何が今この社会で抑圧的なメッセージ、すなわち差別構造を肯定する含意となるかを、設計や開発にたずさわる人々は理解し、できるだけその抑圧に加担しないようにすべきでしょう。このようにジェンダーを備えたデザインを倫理的に問題なく実現するには、知識も検討のための労力も必要です。GROOVE XさんのLOVOTやユカイ工学さんのQooboのように、市場に流通するコミュニケーションロボットがジェンダーを感じさせない設計であるのは、開発者の意図にかかわらず、ジェンダーを備えることのリスク回避の観点からもコスト削減の観点からも理に適っていると思います。

§4 江間先生

道具には使用目的があります。AI技術やロボット技術は、誰のため何のために開発しているか、その目的のためにジェンダー的な観点を組み込むことが社会的な風潮にそぐわず、不愉快に思う人が多ければ、社会的な問題となるでしょう。ジェンダーだけではなく、人種や年齢、外見のほか出身、他文化や宗教についても同様です。ただ、その性質を埋め込むことが目的に不可欠であるかを改めて自問する機会、気付かされる機会が、もしかしたら日本では少ないかもしれません。無意識的に開発者や制作者側に埋め込まれている価値観やバイアスに気付ける教育や意見交換の場が必要になってくると思います。私自身、学生や海外の方など、年齢や環境が異なる人達と話すことで、自分自身が当たり前と思っていた考え方が、相手には当たり前ではないことに気付かされます。「常識とは18歳までに身につけた偏見のコレクション」とのアインシュタインの言葉がありますが、相手の立場や考え方に敬意を払いつつ、自分自身が大事だと思

う価値も伝えていくことが重要だと考えています。

2.2 ジェンダーを組み込むことで起きてしまう

メリットやデメリットについて

ジェンダーを組み込むことによって生じる、既存のジェンダーバイアスを引き継ぐという可能性と、そのメリット・デメリットについて伺った。また、ジェンダーフリーとの関係性について伺った。

§1 野村先生

ジェンダーステレオタイプを崩さないで維持するほうが得策であるという人が一定数存在します。一方でそれを崩さないという人達にとってはとんでもない話になります。ですから、どのように折り合いをつけるかはディスカッションになると思います。ディスカッションをするならば、第三者委員会のように明確ではないのですが、コミュニティ外からファシリテータを呼んで議論をつくっていく必要があると思います。

おそらくステレオタイプを維持したい人々は、ステレオタイプを変化させるコストを取りたくないと思っています。医療系大学で女性の受験者よりも、男性の受験者の得点を上げるという問題がありました[Nomura 19]。医療関係で女性の医療者は長続きしないため、男性をたくさん入れたいという考えがあったからです。本来ならば女性の医者も男性の医者も対等に働ける環境づくりをしないとイケないはずですが、しかしそれを行うにはものすごいコストをかけて病院の中を改善しないとイケません。そのようなコストを取りたくないということがあちこちで起こっているのだと思います。

これは完全にジェンダーステレオタイプを永続化させようという動きで、ある意味で悪意だと思います。特定コミュニティの中の損益が絡んでいて、ステレオタイプ維持が働いているならば、許しがたい話です。それによって本来改善されるべきものが据え置かれるというのはあってはなりません。永続化させようとしている人達ももしステレオタイプを崩したくないというためにヒューマンロボットインタラクションのジェンダー研究をもち出し「ロボットは女性のほうがいいらしいよ」というふうに言い出すならば、我々は批判していく必要はあると思います。

その問題を解決するには、ステークホルダ間のディスカッションをいかに作り出すかが課題です。企業・メーカーのほうもカスタマの意見を吸い上げる必要があると思います。マーケティングリサーチですね。ただ、カスタマイコール神様で動いてしまうのも問題だと思います。カスタマの側にステレオタイプを崩したくない人がいてそちらの欲望をとるだけでは、それがあのおかげで困っているステークホルダを発見できなくなるかもしれません。その意見の吸上げを誰がやるかといえばファシリテータであり、議論を進めるのはサイエンスコミュニケータではないでしょうか。

このような議論を進める際に、「自分はこう思う」というのは言いやすいですが、学術的根拠に基づいていないと言いつ放しになる可能性があります。しかし、一般の方に勉強しないと議論に参加できないというのも暴言としか言いようがありません。社会に対して十分わかりやすい形で情報発信をしていない研究者側にも問題があると思います。だからこそ、ステークホルダをうまくつないでくれるファシリテータがいれば、さらに議論は加速するのではないかと思います。このような議論を活性化させるのがAI学会の役割だと思います。

§2 高橋先生

ジェンダーを組み込むことのメリットについては完全に否定しきれないと思います。相手に感じる安心感、不安感という感情は無視できません。そのような要素を人工知能から完全に取り除くことはジェンダーフリーという観点からは理想かもしれませんが、このような「のっぺらぼう」なシステムデザインと接し続けることが理想のインタフェースデザインなのかは議論が分かれるところだと思います。例えば妊娠や出産など性別特異的なライフイベントもあり、そのような話や相談をロボットやAIにすると、自分と同じ性別のアバターを使用したほうが安心感を得られる人もいるかもしれません。

一方でこのような知見は必ずしも普遍的でなく、その人個人がどのような来歴をもっているか、どのような個性をもっているかで同じ外観の人工知能であってもその感じ方が変わります。〇〇に効果的なのは××というような紋切り型の断定は、そのテンプレートに乗り切れない人達との間に断絶をつくってしまいます。したがって個人ごとに全く異なる個性や来歴に想像力を巡らし、一人ひとりに合ったインタフェースを能動的に選択できる環境を用意することが大切かなと思います。そうすると、今までは例えば道案内の人工知能のインタフェースには女性の外観を用いることが多かったけれど、意外と男性の道案内を好む人も多かったとか、悩み相談は動物の見た目が一番安心するとか、いろいろな人の個性みたいなものが、その人が使っているAIアバターのアイデンティティによって可視化されていく可能性があります。このような可視化は、社会の多様性に対する気付きを生じさせ、今とは異なる社会形態を、AIを通じて人々が生み出すきっかけになるのではないかと期待されます。

一方である一定数が同じような潜在的嗜好をもっていた場合、逆に今あるジェンダーバイアスが固定化してしまうのではないかと問題もあります。これは僕の友達の女性と話したときに出てきた話題なのですが、みんなが自由に使用するアバターの外見を選択した結果、例えば「ご主人さま」というメイド的アバターを男性の8割が使うようになると「やっぱり男は女性をこのような隷属的な存在として捉えたいのだ」と女性達は絶望するのではないかと聞かれました。本来ならば多様性を可視化するはずのアバターが、逆に作用し、既存のジェンダー

バイアスを強化してしまう可能性もあるわけです。

また、倫理的な規範なく、みんなが自分好みのアバターを使用する問題もある気がしています。個人の倫理に頼るのも限界があるのではないかと思いつつ、トップダウンに強い規範を与えることはポリコレ（ポリティカルコレクトネス）的であり建設的ではないようにも思います。この問題を解決する一つのアイデアとして、集团的、「我々の」なものを主体に、ボトムアップに倫理や規範を立ち上げていく方法がある気がします。例えば、ある程度類似したインタフェースに対する好みをもった人々の緩い集団をつくり、そういう集合体単位として規範や秩序をつくっていくことで、ある程度抑制が働いた倫理や規範を生成、維持することができるのではないのでしょうか？ 例えば、ファンコミュニティみたいなものをベースに、ある程度の規制を独自につくっていく感じです。ただ、こういうやり方を突き詰めると、コミュニティがムラ社会化する恐れもある気がしており、まだアイデアとしてはまとまっていません。

あと多様性を生み出すうえで大事なかなと思っているのは、少数派の権利を引き上げようとするのではなく、多数派を切り崩すという方向性もあるかということです。つまり、少数対多数という構図にするのではなく、一見は多数とされている集団の中にもたくさんの少数が存在するという可視化していくということです。ジェンダーフリーは両性の問題ですから、女性が男性中心的社会の中で自らのアイデンティティに苦しんでいるのと同じくらい、このような社会の中で自らに課せられた男性ロールに苦しんでいる男性も多くいると思います。この社会には、さまざまなストーリーが潜在的に隠れていることを可視化していくことで、社会全体が変わっていくのではないのでしょうか。

§3 西條先生

ジェンダーを与えることは、ジェンダーに対するステレオタイプを肯定する態度の表明になり得る点は先ほど述べたとおりです。そもそもステレオタイプはなぜ非難の対象になるのでしょうか。ステレオタイプとは、特定の集団に対して一定の性質を帰属させる、一般化の一種です。この一般化が非難されるのは、社会の不正な力関係の構造を維持することに貢献してしまう場合です。ジェンダーの話でいえば、女性集団が男性集団よりも歴史的に従属的な立場に置かれてきたし、今もそうであるという文脈抜きには女性の表現を評価できません。さらに、女性の中でも、多数派と肌の色の違う人、外国人や移民の女性、出生時に割り当てられた性別との違和を抱えてきたトランス女性、非異性愛者の女性達が被ってきた複合的な差別の歴史を考えれば、その評価の複雑さがいっそう理解できるのではないかと思います。

例えば、この社会では、長髪、スカートの着用、化粧、身長の高さ、広めの骨盤、高い声といった性質は、典型的に女性を表す記号として機能します。「ジェンダーフ

リー化の流れ」とのことですが、こうした特性が個人の女性なり男性なりといったジェンダーアイデンティティの確立に機能していることは否定しがたいでしょう。人は自分が女性や男性であることを自覚するためには社会的な規範との照らし合わせが必要だからです。ただし時代とともに何が女らしさや男らしさの印になるか、そして何が抑圧的表現となるかは変化します。家事育児と女性の結び付きを設計から排除すれば非難されなくてすむ、という単純な話ではありません。かつてのステレオタイプを批判する表現が新しいステレオタイプに固定化されるプロセスはよくあることからです。

ジェンダーを取り入れることで認知的にポジティブな効果があるのは事実でしょう。何であるかわかりやすい、リソースを割かなくて良いということです。例えば、トイレを男女別に分けようとする場合、女性を表すデザインとしてスカートを着用した人物の標識を据え付ける、といったものです。「男女別にすべき」という発想が生じているのかも考慮したいところですが、また、2013年の年末に人工知能学会誌のカバーイラスト（2014年1月号表紙）がセクシズムだと批判されましたよね。このような比較的公共性の高い場に向けたデザインにおいては、誰の利益を損なうか、誰に訴求しないのかという点に目を向けるべきでしょう。公共空間とは原理的にはあらゆる人がアクセスできる場所であるべきだからです。かといって、私的な領域であれば何でもありではありませんが、ユーザがそれを望むから仕方ない、というのは言い訳になりません。そのユーザとは一体誰でしょうか。自分の聞き取りやすい声を拾い上げ、そうでないものは「炎上」と呼んで悪質なクレームや取るに足らないノイズのように扱ってはいませんか。結局「炎上を回避するにはデザインにおいて何に気をつけるべきか」という点に問題は尽きてしまうのでしょうか。その背景に差別の構造的問題があることをわかってほしいと思います。そうでなければ、社会の風潮の変化とともに技術が平然とセクシズムに加担することにもなりかねないからです。

§4 江間先生

「ジェンダーフリー」という言葉自体に誤解や混乱があるので、前半の文章から「ジェンダー」という概念が、性別によるイメージや役割の押付けに使われることをどう考えるのか、という質問と捉え直しました。イメージや役割の押付けに使われ、苦悩するのは女性だけではなく男性もです。「ジェンダー」を既存のバイアスを増長する概念としてロボットやAIに組み込むことは批判される場合もあるでしょう。一方で、だからといって男女の特徴をいっさい感じさせないロボットやAIをつくることを推奨しているわけではありません。結局のところ、冒頭に示したように「目的は何か」によると思います。「安心感がある」要因は、外見、声音、話し方など、何によるものか、それを探求していった、不必要かつ批判され

そんな特徴は外すことができるはずですが、また、単体では問題がなくても、比較あるいは集団や文脈の中で位置付けることで、ジェンダー的特徴が際立つ場合もあります。人は社会的な生き物です。ロボットやAIも社会の中で使っていくのであれば、AIやロボット単体で考えるのではなく、コミュニティや他技術との関係性の中で技術と人間の関係性を考える視点が必要です。

2.3 AIに対するハラスメントについて

AIに対するハラスメント(セクシャルや暴力的行為)とはどのように定義されるのか、それが起きないようにするためにはどのような点に気を配る必要があるか尋ねた。

§1 野村先生

ハラスメントについてはわかっていませんが、エージェントに対するAbuse(虐待)については昔から研究されていました。例えば音声エージェントに暴言を吐いたりする結果がかれこれ15年前くらいから報告されていました。有名なのは、名古屋工業大学の李 晃伸先生の音声エージェントの研究です。これは、実験的に受付にエージェントを置いて、どのような振舞いを人間がするかを調査するものです。入っていた声を解析させようと院生に仕事を振ったところ、その院生が「やめさせてください」と言い出したくらいひどかったようです。攻撃的な言動が発生していたのです。ロボットにおけるAbuseも研究していますが、Abuseであるとしか定義のしようがない言動があると思います。子供がロボットを攻撃的に虐待している行動を分析したことがあるのですが、動物虐待にちょっと似たような感じで、ロボットに対する実験的な興味、楽しさを動機としているように見えました。しかし、子供だけではなく成人もそういうことをしているような気はしています。

最近のエージェントはクラウドにつながっており、それが意識的に攻撃的な言動を抑えているのではないのでしょうか。少なくとも後ろに第三者が存在するという感覚があるかないかで、Abuseの現れ方が変わると思います。

ロボット、AIに対するセクシャルハラスメントに関していえば、ユーザ自体のジェンダーというものに対する成熟度が確立していない状態の中で、AIにジェンダーが入ると複雑度の高い問題になると思います。個人のジェンダーに関する思い込みをジェンダー化されたAIやロボットが固定化してしまう可能性もあります。使うことは個人領域だと思いますが、その個人の中で強化されたジェンダーステレオタイプが公共空間に、犯罪的な意味合いであふれ出すというのは怖いと思います。AIがこうなっているからという理由で、個人領域の中でのAIとの関係性を一般の女性のレベルで勝手に強化して、一般化しようというのは、ありとあらゆる場面に起こり得るのではないのでしょうか。例えば、AIがセクハ

ラめいた発言に対してはぐらかすという対応をしたとします。これが一般の女性にもそのような振舞いにして然るべきだというバイアスを強化してしまうということですね。機械、無機物だからといって、ハラスメントすることそのものがステレオタイプの強化とリアルな場面への強化につながる以上、見過ごせない問題であると思います。

また今までの人間の場合、公共空間と個人領域を物理的環境で切り分けられますが、今後パーソナルエージェントが誕生し、どんな場所でも存在でも使えるようになった場合、これらの切分けはさらに難しくなるでしょう。

このようなデザインコードを考えるためには、エンジニアだけではつくれません。さまざまな関係性を考慮するために、少なくともさまざまな関係者を入れる必要があると思います。

§2 高橋先生

この議論は個人的にも大変興味があります。自分の負の感情を人間に向ける代わりに、ロボットなどの人工物で発散させ、犯罪を抑制することは倫理的かという問になると思います。「クレヨンしんちゃん」に出てくる「ネネちゃんのお母さん」のうさぎの話が良い例ではないでしょうか。ネネちゃんのお母さんが不安定になると、うさぎの大きなぬいぐるみを殴ることで気持ちを鎮め、人間とのコミュニケーションにおける暴力性を取り除こうとしているわけです。

人工物は人間ではないので、人間と同じような倫理を適用しなくてもよい、というのは西洋社会における人間中心主義的な考え方、すなわち人間とそれ以外がきちんと切り分けられるという発想が根底にあるからだと思います。しかし、我々の脳は現実とファンタジーを厳密には切り分けていないようにも思います。自分が行った研究で、事実として相手が人間かどうかよりも、ボトムアップに知覚された情報に応じて神経回路が駆動されるということが示唆されています。これは「この人は人間だ、いや人工知能だ」というふうな論理的な区別を我々の脳はあまりしておらず、もっと直感的、感性的な判断をしていることを示唆しています。したがって、人工知能であればいくらでも暴力を吐いてもよいとか、人間相手にだけきちんとしていればよいという発想は、知覚に駆動されるエラーが生ずる可能性もあり良くないと自分は考えます。

そのため自分の意見としては、AIインタフェースに対してであっても、倫理や配慮をするようなデザインをするべきだと思っています。この配慮というのは、別に人間と同じようにAIを扱えという意味ではありません。人間も多様であるように人工的なインタフェースも多様であるべきだということです。人間と人工物が型にはまったものである必要はないはずですし、人間と人工物とでしか築けない関係性を築いてもよいと思います。た

だAIであったとしても、それを人間の感情を発散する便利な道具にするのは危ういのではないかと、というのが自分の意見になります。

道具的な関係性は短期的には感情の発散を可能にするかもしれませんが、長期的には持続しませんし、一步間違えると中毒的、依存的になってしまいます。持続的な可能性を構築するためには、人間の外界の認知の仕方がヒントになると思います。我々の脳は「①自己が制御可能な対象、一種の道具」、「②制御できないが働きかけができる対象、パートナー的存在」、「③制御することも働きかけもできない対象」に区別していることが示唆されています。AI インタフェースを、暴力的な感情処理の道具としてつくるのではなく、むしろ、共生可能なパートナー的存在としてのコミュニケーションインタフェースを構築していくのが理想なのではないかと思えます。人がAIを大事にできないのに、人を大事にできているわけがないだろうというのが大事なポイントだと思います。

またAIに対する接し方がある種、それが存在する社会の中の内在化したバイアスや差別を可視化する鏡であるといえます。例えばですが、公共の場に出ているエージェントにも、女性は男性を癒やしを与える存在であるという潜在的な社会的価値観が反映されている可能性があります。一方、人間は、一人で生きていけないわけですから癒やすとか癒やしてもらうことはとても大事なわけですから、エージェントから一方的に癒やしてもらうだけではなく、それらに対する癒やしも我々が考える必要があると思うのです。互いに癒やし合う対等な存在としてアバターを認識できるようなインタフェースを構築する。これは、新しいAIの設計論につながると思います。また人間とAIの間に生まれた関係性の物語は、それを人間どうしの新しい関係の物語として展開できるのではないかと考えています。すなわち人間とAIの関係を考えることは、そのまま新しい人間関係の価値をデザインすることと等価ではないかと自分は考えています。

§3 西條先生

AIにハラスメントが成り立つことはあり得ません。AIは人工物であり人間のように意識をもつわけではないからです。もちろんAIが人間と同じような意識や自我を獲得する遠い未来の話は別としてです。現段階では、痛みを感じるとか、尊厳が傷付くといった概念を理解するという事は考えられないわけです。スマートフォンのアシスタントソフトウェアに罵倒表現を使っても、ソフトウェア自身は傷付きません。女性を模したインタフェースを備えたAIに対して、侮蔑的な発言や性的に攻撃的な発言した場合と、生身の人間に対して行った場合では、倫理的評価は当然同等ではありません。道具としてロボットやAIを使う場合、例えば人間同様の気遣いや注意を必要としないからこそ、治療や教育の場

面で効果を発揮することもあるでしょう。人工物の人格はフィクションであり、基本的に道具的に利用可能なものなのです。

しかしながらAIへのハラスメント行為は、人間に影響を与えるしそれゆえに倫理的に問題となり得るものです。不特定多数の人が集まる公共の場でセクハラ発言を行った場合「AIに対して発言しているんだから問題ないだろう」という話は通りません。というのも、これは、公共の場でハラスメント発言が許容されるという人間が存在すること自体、不安感を引き起こすからです。場の公平性と安全性が確保されない、いわゆる「環境型のハラスメント」の一種に相当するでしょう。そのハラスメントが女性なり、特定の民族的ルーツをもつといった属性ゆえに行われたものであるなら、同じ集団に属する人達はその場を安全とは思えません。フィクションと現実の違いという主張もこの場合は意味をなしません。ハラスメント発言は現実であり、それを耳にする人達も現実に存在するからです。また哲学者のロバート・スパローは女性型のAIやロボットに対して、卑猥な発言をするというのはその人の人となりや傷付けている、性格を毀損するという問題を指摘しています[Sparrow 17]。侮蔑的な発言はその発言者自身も傷付けるということですね。私はスパローの意見に必ずしも同意しませんが、問題点としては理解できるものです。フィクションだから現実に無関係と簡単に片づく話ではないでしょう。

こうしたハラスメントを防止する設計はある程度実現していますよね。インタラクティブなAIに対して問題発言があった場合は、何も答えないという、無愛想にはねつけるといった具合です。AIにハラスメント発言をするユーザ側にももちろん問題はありますが、最も責任があるのはつくり手だと思います。当たり前ですが、ユーザはつくられたものしか使えません。ハラスメントを許容する設計になっていればそれに従った行動をするユーザが出てくるのは当然でしょう。

2.4 AI インタフェースのルッキズムについて

AIのインタフェースにとって「かわいい」という側面はインタラクションのうえで親和性を高める可能性があるが、それは無意識にルッキズムをAI—特にインタラクション研究が助長する可能性はないかという点について尋ねた。

§1 野村先生

ルッキズムに対する文化差が大きく影響するのだと思います。イタリアなどはルッキズムを徹底的に排除する動きがあります。そのような世界だと、AIにもそれを適用して議論を進めると思います。一方日本では、そのあたりがまだ曖昧でAIはAIだからみたいなごまかしをしているような気がします。ジュニファー・ロバートソンとジュタ・ウェーバーが指摘しているように、国際的にスタンダードな感覚に比べると、立ち遅れているのか

なと思います [Robertson 18, Weber 05]. これは AI に限らず SF などにも、保守的な面が無意識に出ているからかもしれません。

我々がもっているジェンダーステレオタイプがロボットのデザインにどれだけ影響を与えるかについての確認がまだ不十分だと思います。それがわかったうえで、ジェンダーはどうするかという流れになると思うのです。そもそも「ロボットにジェンダーを入れることに意味はあるの?」とか、「そんなのおもしろいの?」とか、その段階をまず払拭しきれていないと思います。私の研究もまだ確認段階といえますか、個人のもっているジェンダーステレオタイプがどう影響するかとか、ジェンダーの価値がロボットのインタラクションにどのように影響を与えるかとか、それらの確認段階で終わっているところなのです。確実に影響するんですよということがまだ言い切れていないと思います。なので、研究者間でも「何を言っているんだ」という人もいれば「そのところをちゃんと考えないと大変ではないか」という温度差が存在している状況なのだと思います。

ただ、昔のインターネットの黎明期において、ネットでつながっている人間関係なんて細いものばかりだから、全然影響なんてないよと言う人達がありました。多分今同じことを言う人はいないと思います。オンラインだろうがオフラインだろうが、人間関係は人間関係であるからです。ですから、メディアエクエーション (Media Equation; メディアの同一視) に立ち戻って、人間はインタラクション相手を区別していないという点から考えないといけないのではないのでしょうか。議論をするならば、「リアルとフィクションは区別できていない可能性がある」という前提をおいて始める必要があるかもしれません。

個人が利用するテクノロジーに対して、どこまで介入すべきかという問題もあると思います。対人不安を抱えている人にとって、ネットコミュニケーションは普段喋れる場であります。一方で、ネットコミュニケーションに頼ってしまい、対人不安を抱えたままで一向に終わらないか、治療に関するスポイルになるのではないのかという人と対峙してしまいますよね。どちらの意見も正しいので解決つかない。だからこそいろいろなオプションを総括で見ながら、全体的に動くしかないのかなと思います。

§2 高橋先生

外見的要因の影響が強いことは否定できません。よく「人は見た目じゃない」と言いますが、やはりなんだかんだ言っても見た目は重要です。我々の脳は、自分とは異なる外見の存在に対する恐怖や警戒など、本能的な感情反応があります。これは人間の醜い側面ではなく、進化の過程で組み込まれた特性であるといえます。一方で人間は言語を発明し、さまざまな文化を創造し文明を発展させています。その結果、多様な価値が存在する世界

に少しずつ進歩していると思います。

ですから外見的なものに価値がないという議論よりも、外見に対して価値を見いだす反応は人間が等しくもっているという原則に立ち、そのうえでこの世界には他にも素晴らしい価値があると我々の価値観が広がっていくことで、初めてルッキズムの過度な神格化の傾向が消えていくのではないかと思います。もちろん性的に誇張しているものは危険だと思いますが、今からそれを際限なく規制していくというのは、人間のもとももっているある種の生得的価値というものに対し抑圧してしまい、反対にその価値を潜在的に強化してしまう恐れがあります。性的強調というのは、女性をコミュニケーションの対象としてではなくて、道具として人を見ているという文脈で出現しやすいと思います。なので、前述のように互惠的な人間と AI との持続的な関係性の価値を探求していくことで、ルッキズムのような表層的な問題を超越する人間関係の満足や喜びが認識されるようになる可能性があります。そのような新しい人間関係の価値が社会のなかで共有されていくことで、初めて人間というのは表層的なルッキズムの呪縛から自由になれると思います。以上の理由から、かわいい、性的なものだからという理由のみで無制限な規制をトップダウンにするのは、ポリコレ的な価値観の押し付けになる可能性もあり、慎重に考えたほうが良いと思っています。暴力的なゲームだから、性的なコンテンツだからと、無差別に規制をしていくのではなく、慎重な対話を続けていく必要があるように思います。

§3 西條先生

ここで「ルッキズム」という表現を使うのは、AI のキャラクタとして人間を模したものを主に考えているからでしょうか。ルッキズムとは人間の外見が規範的な美しさを備えていないと不当な扱いを受ける、という程度の理解でしかありませんが、まず、女性だと不必要に外見が評価の対象となるという問題は、性別に対するステレオタイプの一種であり、Q1 や Q2 と重複するため回答の範囲から除外します。男らしさを構成する規範としての身体の美しさもやはりステレオタイプの問題ですね。一応繰り返しますと、そもそも人間らしいデザインである必要、そこにジェンダーを加える必要があるのかを先に考慮すべきでしょう。倫理的問題になりやすいので、これらを避けるほうが複数の意味で負担は少ないというのは先に述べたとおりです。

ルッキズム固有の問題というのは難しいのですが、どのようなジェンダーアイデンティティをもつかにかかわらず、人は「美しい身体」の規範を完全に無視して社会生活を送ることはできません。顔立ちや体格など外見へのコンプレックスがいつさいない、気にしたことがないという人は少ないのではないのでしょうか。こうしたコンプレックスが生じるのは、社会で通用している「理想」の身体と自分の身体の比較があるからでしょう。筋トレ

やダイエットが盛り上がるのは、何も健康維持のためばかりでないことは想像がつくかと思います。こうした美的なモデルは、人を病や社会的破滅にまで追い詰める危険性と、自己肯定や生活の質の向上になり得る両面を備えています。例えばジムの広告一つとっても、やせろ、鍛えろと圧迫感を覚える人もいれば、自分の望む身体をつくり上げる楽しさを思い出す人もいるでしょう。そう考えるとキャラクタのデザインを多様化すること、さまざまなモデルを提示するというのが無難な解決策ではないかと思います。ファッションの世界で、さまざまな体型や肌の色のモデルを起用する動きと似たようなイメージですね。

いずれにせよ、エージェントやロボットの開発というのは、人を対象とする研究としてあまり扱われません。そのため、研究倫理の研修でも人権に関わる問題が十分にカバーされていないという懸念があります。これは実験や参与観察を行わない分野であれば自然科学でも人文社会科学でも共通の問題です。学生であれば「この先生、大丈夫なのかな」と思ってもなかなか言い出しにくいですよ。基本的に教員が負うべき責任ではありますが、不安を感じたら大学のハラスメント相談室を利用してください。自分に危害が生じていないのであれば、ひとまず大学のハラスメント防止パンフレットや関連する専門家の執筆した新書などを読んで自主的に知識を身につけることは有益だと思います。

2・5 議論を進めるための方向性について

AIやロボットが社会の中に浸透していく中で、ジェンダーを組み込むことについての課題を可視化する方法と、一致点を目指すための議論を進める方法について伺った。

§ 江間先生

AI、エージェントやロボットの利用目的を適切に利用者とコミュニケーションすることが重要です。研究や開発をする側はAIやロボットの利用目的をどのように設定しているかを考えていると思います。より深く自問するのであれば、「目的達成に(1)ジェンダー的要素は必要不可欠なのか、(2)ジェンダー的要素が入っていることで開発側の意図とは別目的で使われる懸念はないか、(3)懸念がある場合、低減対策としてどのような方法があるか」を世に出す前に関係者間で議論しておく必要があります。企業であれば営業や法務、広報など異なる視点をもっている人のほか、家族や友人など潜在的なユーザにも意見を聞くことができるでしょう。一方、どれだけ対策を取っても、開発目的とは異なる意図で受け止められる懸念はあります。その場合も、意見は真摯に受け止めて必要であれば対応を行い、次の開発時の注意点として事例を蓄積していく体制を企業あるいは大学、研究所内で構築することが重要かと思います。

2・6 問題が生じた場合の責任の所在について

AIやロボットにジェンダーを組み込む研究において、問題が発生した場合、もしくは問題が生じると思われる場合、開発した研究者の責任は、どのように取るべきか伺った。

§ 江間先生

問題が起きた場合、事実関係を把握し、何に起因する問題なのか(技術そのものの問題なのか、伝え方の問題なのか、使い方の問題なのか、など)を関係者間で明確にし、対策を講じる必要があります。SNSなどの普及で情報の拡散は早いので、即座に対策を考える必要があります。

ジェンダーではありませんが、韓国の名門国立大学KAIST(韓国科学技術院)がAI兵器開発をしているという報道があり、著名なAI研究者らが書簡で「この目的が事実であるならばKAISTとの関わりをいっさい断つ」といった宣言を出した事件の記事を以前、書きました*5。記事ではスキャンダルからの教訓を二つ指摘しました。まずは、問題を防ぐだけでなく、過剰な反応に委縮せずに「反論」をしていくためにも、社会的政策的な課題に対する感度を常日頃から磨いていく必要性です。自律型致死兵器システムは国連レベルでも議論が行われているなど世界的にも注目の高い議論の一つです。同様に「人種やジェンダーなどによる差別や偏見を増幅するシステム」、「説明や判断がブラックボックス化してしまうシステム」などに対しては社会的な関心が高いです。

一方で、個人ができることには限界があります。二つ目の教訓として、社会的政策的な課題に対応できる組織ガバナンスの整備が必要です。組織内でAIやデータガバナンスに関する事案に注意を払っておける人(例えばChief Ethics Officer)や部門をつくるのが大事になります。さらには問題が起きたときに誰にどのように相談すればよいのかの訓練も平常時からしておくことが重要となります。組織的な支援体制を構築することで、研究者個人に過度に責任を帰属させない。事件や事故の原因究明も含めて類似問題が起きないように情報共有や(大学であれば)教育する体制を構築することが責任ある組織に求められます。一方である調査*6によると、AI倫理教育を受けている人達は非常に少ないといわれています。海外でもこのような状況ですので、日本でも教育体制の在り方や何を教育するのかの内容についての議論を進める必要があります。

*5 <https://xtech.nikkei.com/atcl/nxt/column/18/00001/00757/>

*6 <https://thenextweb.com/neural/2020/07/03/study-only-18-of-data-scientists-are-learning-about-ai-ethics/>

2.7 表現の自由とAI・ロボットのデザインの関係性について

表現の自由とAIやロボットのデザインの制限は両立するのか伺った。また、学会がジェンダーバイアスを引き継ぐ研究を認めないといった解決方法はあるのか伺った。

§ 江間先生

質問にある「制限」の強さによるとは思いますが、現状で研究者が懸念するものは、萎縮や自主規制、忖度の行きすぎかもしれません。自らが考えて制限するのではなく、思い込み、「余計なことはするな」という同僚や上司からの（暗黙の）ピアプレッシャーのほうが、法的あるいは学会指針などのよるべき明確な根拠がない分、たちが悪いかもしれません。また明確な線引きや指針をつくったところで、同じ間違いは防げるかもしれませんが、変数が異なる別の問題の発生は防ぐほどの汎用性はないかもしれません。

AIと法律系研究者がこの問題について、興味深い対話をしていました。事件・事故の再発防止策を議論するとき、理工系研究者は「このような場面ではこのような行為はしてはならない」、などシミュレーションに基づいた具体的な対策防止案などを作成しがちだそうです。これは理工系だけに限らないかもしれません。何か問題が起きれば、大学や研究所、企業においても細かな規則や手続きが増え、その手続きを守ることが目的化してしまうという本末転倒も起こり得ます。時にそれは非効率、非生産的な仕事を増やしかねません。また、規則にあることが逆に弊害となって、最適解にたどり着けないかもしれません。

これに対し、法律系研究者の対策は大原則を示すにとどめ、あとは実際の事件や事故の場面に応じて解釈を施して対応をする傾向があるということです。憲法はじめ各種法律も解釈を新たにしていくことで、新しい問題にも応用できる余地を残しています。規則のつくり方、解釈の仕方一つ一つの「技術」なのです。これは例えばロボットやAIの研究でも同じものかもしれません。ある程度の「遊び」の部分がロボットやAIの設計にないと、応用がきかなくなります。

話を質問に戻すと、たとえ何がしかの「制限」があったとしても、つくりたいものをつくることはできます。それには何がデザインの目的であるのかを多様なステークホルダーと議論する余地、解釈する余地があることが必要です。また、私達一人ひとりが、自分のつくりたいものや主張を押し通すことを目的とせず、相手の話の聞く耳をもつという「余裕」をもつこと、どのような社会をこれからつくっていききたいのかという中長期的なビジョンを互いに共有できるかが大事になります。

3. おわりに

お話を伺った先生方からは、それぞれのご専門分野の知見をご紹介いただくとともに、今後のAIとジェンダーに関する議論をするうえで重要な示唆をいただいた。例えば、野村先生には、AIの社会実装に関わるステークホルダーを巻き込んだ議論の場の必要性についてもご示唆をいただいた。高橋先生には人間とロボットとの関係性についての考察をベースに、ルッキズムの先にあるエージェントデザインについてご教示いただいた。西條先生からは、ご自身の哲学的な知見をご紹介いただくことはもちろん、執筆者が男性であるといったジェンダーバランスの偏りについてのご指摘など、今後の展開について多大にご示唆をいただいた。江間先生には法律系研究者との議論など、社会的・政策的な議論ができる体制のもと、AIの社会実装の目的を明確にしていくことの重要性をご教示いただいた。今回の議論は、ここで終わるものではなく、AIを開発する多くの研究者やそれを活用するサービス提供者に、自分達の提供するシステムを見つめ直すための議題を喚起し、新しく議論を始めるためのマイルストーンではないかと考える。

自分達の研究や製品にどのようなプリセットされたバイアスが含まれているか、差別的ではないのか、それを提供することは誰の利益を守り、誰の利益を損ねているのか、AI・ロボットと人間のインタラクションは人間にどのような影響を与えるのか、データやモデルの公平性だけでなく、そもそもの問題の出発点や、出てきた結果に偏りがいいのか、課題に気付いたとしても、その議論の結論が偏ってしまう可能性がないか。このような問いかけをするくらいに、AIが社会に浸透してきたともいえる。そういう状況であるからこそ、常に新しい情報や、さまざまな考えが存在することを意識し、アンテナを張り議論に組み込むことは重要である。

このような検討を研究室で閉じ込めておくのではなく、研究科、大学、学会レベルで進め、アカデミアの成果として発信し、さまざまなステークホルダーと議論を重ねていき、さまざまなユーザが安心してAIを利用する社会を構築していく一つの筋道ではないだろうか。

この場をお借りして、本企画に賛同していただき、急な日程にもかかわらずインタビューや原稿作成にご協力をいただいた先生方に感謝いたします。本当にありがとうございました。

◇ 参 考 文 献 ◇

- [Nomura 19] Nomura, T.: *A Possibility of Inappropriate Use of Gender Studies in Human-Robot Interaction*, *AI & Society* (2019), <https://link.springer.com/article/10.1007/s00146-019-00913-y>
- [Robertson 18] Robertson, J.: *Robo Sapiens Japonicus: Robots, Gender, Family, and the Japanese Nation*, Univ. of California Press (2018)
- [Sparrow 17] Sparrow, R.: Robots, rape, and representation, *Int. J. of Social Robotics*, Vol. 9, No. 4, pp. 465-477 (2017)
- [Weber 05] Weber, J.: Helpless machines and true loving care givers: a feminist critique of recent trends in human - robot interaction, *J. of Information, Communication and Ethics in Society*, Vol. 3, No. 4, pp. 209-218 (2005)

2020年8月12日 受理

————— 著 者 紹 介 —————



藤堂 健世 (学生会員)

2017年東京工業大学大学院総合理工学研究科修士課程修了。修士(工学)。現在、同大学院情報理工学院博士後期課程所属。本学会学生編集委員。認知科学、言語創発の研究を行っている。



佐久間 洋司 (学生会員)

1996年東京都生まれ。大阪大学基礎工学部システム科学科在学中。Panasonic Silicon Valley Lab (現 Panasonic β) での半年間のリサーチインターン、トロント大学基礎工学部門での1年間の交換留学を経験し、世界経済フォーラム Global Shapers に選出。孫正義育英財団第2期生(正財団生)に認定される。大阪府・大阪市「大阪・関西万博におけるパビリオン等地元出展に関する有識者懇話会」の委員などを務める。2019年 NewsPicks Magazine 「未来をつくる7人の UNDER30」に選出。情報処理学会などの学生会員。本学会学生編集委員長。

大澤 博隆 (正会員) は、前掲 (Vol. 35, No. 5, p. 617) 参照。