AI 公平性のガバナンス―日本企業が直面する課題とその対策―開催レポート

日程：2022 年 7 月 19 日（火）
時間：10：00－12：00
主催：東京大学未来ビジョン研究センター
共催：有限責任会社法人トーマツ
ウェブ：https://ifi.u-tokyo.ac.jp/event/13340/

イベント概要

2022 年 7 月 19 日（火）にウェビナー「AI 公平性のガバナンス―日本企業が直面する課題とその対策―」が行われました。

本ウェビナーは、人工知能（AI）の社会実装が進むなかでの課題のひとつである「公平性」に焦点をあて、公平性に配慮した AI システムやサービスの開発プロジェクトに期待されるガバナンスの在り方や調達単位自体の見直しの必要性について言及し、実効性のある AI 公平性を達成するための議論を展開しました。

当日は、主催である東京大学未来ビジョン研究センターの江間有沙氏の司会のもと、最初に共催となる有限責任監査法人トーマツの松本清一氏が開会挨拶をされました。その後、話題提供者として井上彰氏（東京大学）、神嶌敏弘氏（産業技術総合研究所）、原嶋瞭氏（東京大学/有限責任監査法人トーマツ）に登壇いただき、最後のパネルディスカッションには、登壇者に加えてパネリストとして松本敬史氏（東京大学/デロイトトーマツグループ）、山本優樹氏、木畑登樹夫氏（有限責任監査法人トーマツ）を交えて AI 公平性に関する議論を交わしました。

開会挨拶　―松本清一氏

AI の活用が増えるにしたがい、信頼ある AI を提供するということが非常に重要になってきます。国連や OECD あるいはそういった産学、官民の複合の組織といった各企業が AI の公平性など原則を含むガイドラインやポリシーを発表しているところです。また、さまざまなツールなどを無償・有償で利用することができるようになってきている状況です。

そういった中で企業が信頼ある AI を提供するための論点が幾つかあります。今回のウェビナーは 2022 年度の人工知能学会の中でも発表した AI 公平性をテーマにお話しします。

話題提供

公平性とは何か　―井上彰氏

AI は、住宅ローン審査や雇用のマッチング、あるいは再犯予測により保釈してよいかどうかを決めるという、かなりきわどいケースでも使われているという実情があります。その際に AI に期待されているのは、バイアスが関与しない決定をサポートするということです。

まさにその点で、AI の使用に対し倫理的問題が提起されている現状があります。例えば透明性、セキュリティー、責任の所在、自由、尊厳、社会的連帯への脅威という形で、AI の倫理的問題が問われています。

とくに先鋭的に問われているのが、AI の公平性です。公平性について、具体的にどのような AI の倫理的問題が提起されているのかというと、センシティブ情報・センシティブ属性といわれる性差や人種等が意思決定に関与してしまうことです。ほかにも、差別の歴史等の歴史的不正義が社会的格差の再生産につながっているという問題もあり、構造的に是正しなければならないという問題をわれわれは抱えています。

さて公平性は、その是正のために重要な役割を果たします。それは公平性が、情報制約と比較秤量（しょうりょう）に関与していることからも窺えます。公平性を中心的価値として扱う政治哲学において、どのような情報制約をかけるかによって、正しさを規定する正義構想が変わってくることが知られています。同じことは AI の公平性にも当てはまります。センシティブ情報の制約の仕方は様々です。問題は、その制約方法の適切性がいかにして担保されるのか、です。

スタンダードな方法として政治哲学において議論されてきたのは、帰結主義と非帰結主義です。帰結主義は、不公平な取り扱いをされることにより、不公平な取り扱いされた人の利益は毀損（きそん）される、という考え方です。それに対して非帰結主義は、不公平な取り扱いは、その行為の不正性とかかわっているとみる立場です。非帰結主義は義務論という形で語られることが多いです。

帰結主義・非帰結主義を問わず、比較秤量をしていくための規範的基礎が欠かせません。規範的基礎のうちポピュラーなのは、効用です。これは満足度、あるいは単純に幸せと言い換えることができます。それから功績も挙げられます。これだけのことを行ったのだからそれに見合うだけの報償、例えば高い給与水準、が保障されてしかるべきだという主張するのが功績の価値的特徴です。それから構造的不正義も考えられるでしょう。自由に選択しているつもりでも、たとえばジェンダー構造によって女性が男性と同じように自由に選べていないという形で、不公平が社会構造によって生み出されている現状があります。

AI の公平性は、私が見るところでは帰結主義に依拠して検討されることが多いです。つまり AI のアルゴリズムがもたらす危害というものをどのように是正するか、という形で議論していく議論が一般的です。ちなみに危害の匡正をどのように行うのかにあたっては、AI の公平性の場合、集団公平性がよく問われます。政治哲学の場合には、むしろ個人公平性のほうです。例えば先に挙げた功績は、個人公平性にかかわります。その人の功績がどれだけあったかというパフォーマンスを評価します。このように、AI の公平性を検討するにあたって、政治哲学者は個人公平性を見る傾向があります。したがって、その点は、AI のアルゴリズムをめぐる技術的公平性にかんする議論と少々ずれがある印象です。だからといって、政治哲学において技術的公平性の議論が重要でないということではもちろんありません。私自身は、政治哲学においてどこまで AI 公平性の議論に関与できるのか、貢献できるのか

という辺りを見定めなければならないという問題意識を持っています。

公平性を実現するにあたり、いかなる規範的基礎が妥当かについての論争は、まったく決着が付いていません。そのことに鑑みても、どういう公平性が問われてくるのかはコンテクスト次第であると言えるかもしれません。実際、コンテクストを無視して、AI の公平性にかんする議論を進めることはできません。センシティブ情報が黒人・白人あるいは他の宗教的なバックグラウンドの違いにどうかかわってくるのかについては、それが先鋭的に問われてくる特定のコンテクストに根ざした形で検討する必要があると思います。

AI 分野での公平性　一神嶌敏弘氏

公平性配慮型機械学習という技術が最近出てきています。機械学習技術が普及するようになり、個人の生活に影響を与えるような決定に利用されようになりました。公平性など潜在的な問題にこうした配慮をしたデータ分析技術で社会的公平性や法律や規約の制限の観点からセンシティブな情報を排除したような予測をしようというのが、公平性配慮型機械学習です。

この公平性配慮型機械学習が達成する形式的な公平性規準についてお話しします。この公平性規準は幾つかの種類があり、これらは同時には成立させ得ないという相互排他性があります。排他性があるために公平性規準を井上先生がおっしゃっていたようなコンテクストに応じて選択しなければなりません。それを公開することにより運用していかなければなりません。

形式的公平性とは、目的変数間の形式的な関係で定義される、ある望ましい状態をいいます。井上さんがおっしゃっていた帰結主義、非帰結主義のうち、公平性配慮型機械学習が対処するのは帰結主義に対応するものです。

被差別者の損害を補償するということが目標になるのですが、どのような保障をするかというところで問題になってきます。情報制約による公平性と個人公平性に対して、帰結主義の配分の公正は、前者とは同時には達成できません。センシティブ情報を無視することにより配分の公正が自ずと達成されることは数理的にはありません。

一番有名な事例として COMPAS の再犯リスクスコアに対する ProPublica の指摘があります．これには，裁判所から反論がなされており、その中で 1 つ着目するのは、幾つかある公平性規準が同時には満たせないという公平性規準の相互排他性です。裁判所は、予測を基準にした実際の犯罪率という規準を採用しており、この規準ではほとんど差がなく公平といえます。すなわち、裁判所が達成するようにしていた規準とは同時には達成できない規準を ProPublica は指摘していたというのが実情です。そのため現在もこの COMPAS スコアは活躍しており、裁判の判断を公正にするために非常に有用に使われているという状況です。

そうは言ってもコンテクストに応じてそのような規準を選択したかというのもきちんと示しておかなければいけません。このモデルは誰が作り、どのような意図で作られているの

か、どのような項目できちんとテストをしたのか、その結果もきちんと示しておくということでどのような規準についてきちんと公平になるように考えて設計したのかというものを公開していくことが重要ではないかと考えています。個人の判断よりもアルゴリズムのほうが容易に補正できるので、帰結主義を受け入れるのであればアルゴリズムのほうが非常に容易なので、私自身は公平性を達成していくためにはアルゴリズムを活用していくことが重要だと考えています。

機械学習は道具ですので、道具全般に言えることですが、工学の基本、適切に設計して十分にテストをして運用中に監視・更新するということが他の工業製品と同様に重要です。そして選択した基準とそれを選んだ根拠を明示し、問題に対処します。井上さんがご指摘されたようなコンテクストを明らかにした上でわれわれはどのように考えたかということを提示することで対処していくことが大切だと考えています。

日本の産業構造を踏まえた AI の公平性に関する企業の役割の考察　一原嶋瞭氏

先ほどのお二人のお話を踏まえると、私の発言のコンテクストとしては、日本の産業構造を踏まえているところ、企業の役割からの視点であるところといえます。

公平性の規準にはトレードオフの関係があり、同時に満たすことは困難であるということが知られているため、AI の関係者が目的を理解して客観的に評価できる必要があります。日本では実装をどのようにしていくのかという議論は他国に比べて進んでいない印象があります。

AI の公平性を巡る論点には社会的な公平性と技術的な公平性があり、この 2 つの公平性の視点を踏まえた上で、実運用における公平性という論点が次に挙がってきます。実運用における公平性というのは、組織により AI の運用は異なるわけですので、組織でどのように運用していくのかという文脈を考慮して判断する必要があります。

公平性の考慮が進まない原因は、1 つは AI 企業とサービス提供企業が今明確に分かれている傾向が強いということです。もう 1 つは日本の産業構造的に下請けに次ぐ下請けという構造が実体としてあることです。ですので、日本の産業構造を踏まえ、どのように検討していくのかというところを明らかにしなければなりません。

日本の産業構造の具体例に入っていきます。まずステークホルダーが非常に多いという例として挙げているのが、日本の政府の調達手続きの部分です。公平性やプロジェクト全体を通した視点に関して意思疎通が取りづらい現状があります。ただこれが日本の政府のスタンダードという形で公開されていますので、日本の企業はこれに従い調達、開発、設計をしていくものと理解しています。ですので、まずこの調達単位を前提として調達単位ごとの公平性の考慮はどのようにするのかというものを明らかにしていきました。

私は公平性関連ツールの調査によって、それを日本の調達単位に当てはめるというアプローチを実施しています。調査結果は、東京大学未来ビジョン研究センターの AI ガバナンスプロジェクトのページ（https://ifi.u-tokyo.ac.jp/projects/ai-service-and-risk-

coordination/ ） の AI 公 平 性 ツ ー ル キ ッ ト （ https://ifi.u-tokyo.ac.jp/wp/wp-content/uploads/2022/06/20220614-Machine-Learning-Toolbox.xlsx）に公開されています。

調査結果のツールについて、入出力データや AI モデルの評価・可視化の部分を対象としたツールが多い傾向があります。ツールとしては評価や可視化ができますという形で公開したほうが、利用範囲が広がるという狙いがあると考えられます。結果の補正は企業の方が迷う傾向が多いため、ツール開発側としても頻出する観点を解決できるように整理をすることができれば、よりツールの適用範囲ひいては AI 公平性の考慮が広まると考えます。

もう 1 つはツールではなく、企業の産業構造の観点からの考察です。日本の現状として下請けの構造がありますので、公平性に関する企業ガバナンスを統治するという余裕がない企業が多いのではないかと考えています。例えば元請けの企業のほうで AI 公平性に関する部署・人材を置くということでプロジェクト全体の公平性の考慮が可能になるのではないかと考えています。また、調達単位、調達全体で一貫して公平性を考慮するということが難しくなっているので、調達単位自体もコスト観点以外の観点も入れて見直すことも考慮すべきだと考えています。

今回は企業の立場を今回は整理しましたが、これを具体的なユースケースと組み合わせてより実効性のある公平性を達成するための着眼点、研究へとつなげていく予定です。

パネルディスカッション

パネルでは、参加者の方より様々な質問をいただきました。その内容をご紹介します。

まずは、情報制約や政治哲学と技術論の差異に関する質問をいただきました。井上氏は、コンテクストやイシュー、どのような主体が関わっているのかということに結局よらざるを得ない部分があり、連携が必要だということやユースケースまで考えなければいけないということ、そうした点について、このパネルに登壇されている方々の問題意識や方法についてそれほど対立があるとは思っていないことを話されました。

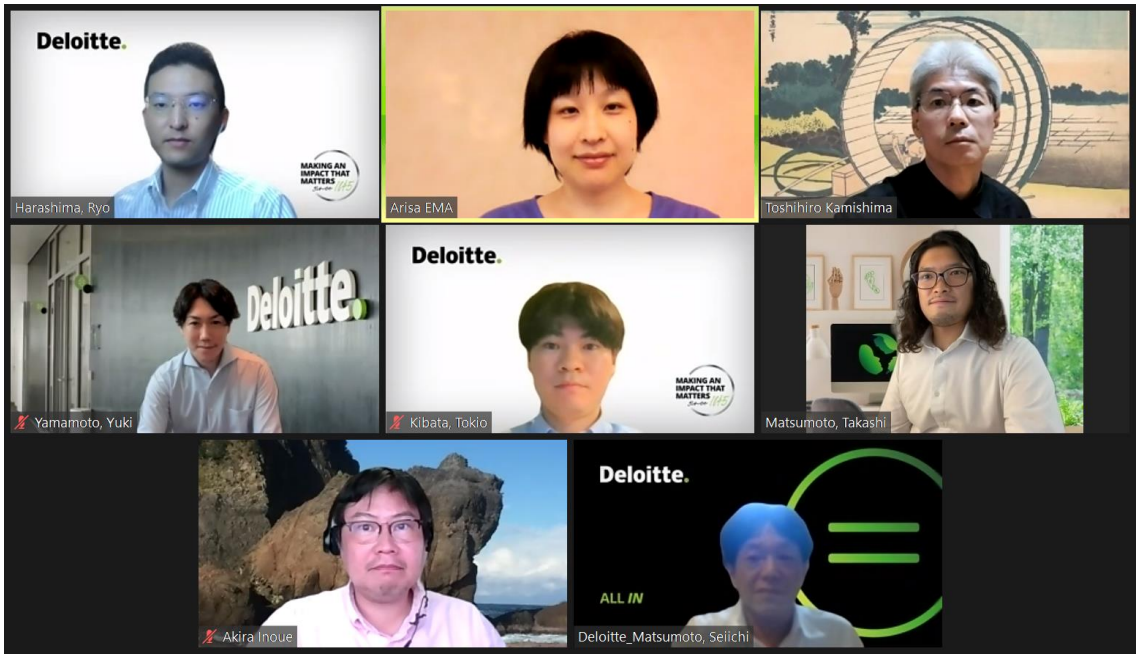
また、参加者の方より公平性ツールについて評価以外に説明に関するツールについて質問をいただきました。原嶋氏は、設計・開発の段階でどの公平性を考え、どの公平性を測定して結果がえられたかというのを記録として残しておけば、それを開示するだけで説明になると考えると話されました。松本（敬）氏は、ユーザー側の技術的な知見の不足や、評価人材の不足から AI 公平性が問題提起されない環境である問題点を指摘しました。

加えて、欧州の規制の動きに日本も追従し対策を進めるのかという質問に関して、原嶋氏は、今回の調査結果の表を用いて、大手企業が公開しているツールから使い始めることとなること、日本企業から公開されているものはないが日本語ドキュメントが充実しているものが存在するため日本も対策を立てる必要があると説明されました。

最後に、リスクレベルに応じて公平性に関する措置が変化するのかという質問をいただきました。神嶌氏は、人間の介在が増えるほど誤りが増える可能性、AI の判断速度に人間

が反応できない可能性を上げ、リスクレベルに応じた人間の措置を変えることは難しいのではないかと述べました。

まとめとして、神嶌氏よりアカデミアで関心が高まっていること、山本氏、木畑氏より実際に AI 公平性に関する企業からの問い合わせが増えている現状を共有いただき、AI 公平性は様々な分野にてコラボレーションした議論が必要であるというメッセージを持って本ウェビナーは終了しました。



上段左から：原嶋氏、江間氏、神嶌氏
中段左から：山本氏、木畑氏、松本（敬）氏
下段左から：井上氏、松本（清）氏