

# 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

## A methodology for companies to achieve AI fairness fitting for Japanese industrial structure

原嶋 瞭<sup>\*1\*2</sup>  
Ryo Harashima

江間 有沙<sup>\*1</sup>  
Arisa Ema

井上 彰<sup>\*1</sup>  
Akira Inoue

神嶋 敏弘<sup>\*3</sup>  
Toshihiro Kamishima

松本 敬史<sup>\*1\*2</sup>  
Takashi Matsumoto

木畑 登樹夫<sup>\*2</sup>  
Tokio Kibata

<sup>\*1</sup> 東京大学 <sup>\*2</sup> 有限責任監査法人トーマツ <sup>\*3</sup> 産業技術総合研究所  
The University of Tokyo Deloitte Touche Tohmatsu LLC National Institute of Advanced Industrial Science and Technology

Recently, AI fairness and accountability have been required from ethical and social perspectives. This is true of Japan. However, to meet the requirement, it is important to note that Japanese companies tend to have various stakeholders such as development companies and service providers in their system development lifecycle. To clarify the requirement to be satisfied in terms of AI fairness, we articulate the different positions of Japanese companies that are involved in the development process. With the Japanese industrial structure, we investigate existing AI services and then map useful AI fairness technologies according to the roles of the companies. As a result, this study suggests a repository that can be used as a criterion for each company to meet the requirement of AI fairness in Japan. Moreover, the repository can be used as a reference guide for the Japanese companies that are planning to adopt AI technologies in the future.

## 1. はじめに

AI の活用シーンが増え、精度向上を目指して改良が重ねられている。近年では、精度以外にも倫理や社会的な観点から AI の公平性や説明責任が求められるケースがあり、一部ではすでに社会問題となっている。

例えば、2016 年 Correctional Offender Management Profiling for Alternative Sanctions (以下 COMPAS) の事例では、AI の判断に人種間差別があると報道機関に告発され、大きな議論に発展した。これは開発者のターゲットとしていた公平性規準を満たしていたが、ターゲットとしていない別の公平性規準を考慮した際に問題化した例である[Flores 2016]。公平性規準の間にはトレードオフの関係があり、同時に満たすのは困難である場合があるため、この問題は根本的に解決することができない。

そこで重要となるのが、公平性の説明である。公平性とは、一般的に、偏見や恩顧によって特定の個人もしくは集団の属性に左右されない状態と定義される[Mehrabi 2019]。この一般的な理念をふまえたうえで、AI の関係者が AI の利用目的を理解し、それを達成するようになっていることを客観的に評価できる必要がある。

客観的な評価をするために、AI の公平性は様々な角度から議論されている。日本においても AI の公平性について学術的な議論はなされているものの、外国と比べて AI 公平性の実装に関する基本的な考え方が整っていない状態である。今後 AI の社会実装が広がっていくにつれて、日本でも AI 公平性の重要性が増してくると考えられる。

本論文では、まず 2 節で一般的な AI の公平性の論点を整理し、3 節では AI の公平性を考慮する際に課題となる日本の産業構造の課題を述べる。4 節と 5 節では、日本の産業構造を踏まえたうえで既存の AI 公平性に資するサービスやソフトウェア、プログラム等の技術を調査し分類することで、日本企業が AI の公平性に関する議論を進めるために必要な観点や制度、技術等

の要素を明らかにする。6 節では公平性に配慮した AI サービスやシステム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への見直しの必要性について考察し、7 節では本調査の今後の課題と展望について述べる。

## 2. AI の公平性をめぐる論点の整理

AI の公平性は複数のレイヤーで議論されており、その論点を 3 つに分けて整理する。

### 2.1 社会的な公平性

まず、国際的な枠組みの中で論じられている差別や平等を主たる対象とする社会通念上の公平性、すなわち、「社会的な公平性」がある。その内容は、情報のアクセスに関する適切な制限や不当な利用の防止などである。この定義については概ね、加盟国や地域で合意されている。

国際的な機関だけでなく、様々な企業が AI の公平性を原則に含むガイドラインを発表している[Jobin 2019]。代表的な例としては国連や OECD、AI という文脈では IEEE、EU、OECD といった国際機関から発表されている。日本では内閣府から「人間中心の AI 社会原則」[統合イノベーション戦略推進会議 2019]が出されている。また、国際標準化の動きとしては ISO/IEC JTC1 の SC42、IEEE-SA の P7000 シリーズにて検討が進められている。

### 2.2 技術的な公平性

次に、技術的な観点で論じられる公平性を測定する際の規準がある。AI による学習データはそのままだと偏りが見られる場合もあり、場合によっては差別を生んでしまう。そのことをふまえて、そうした偏りを様々な仕方でも測定・補正する技術が開発され、実際に運用されている。日本においても、様々な学会や論文で議論がされており、多くの技術が生み出されている[神嶋 2022]。

技術的な公平性の定義は 2 つに分かれる。ひとつは、一人の個人が集団属性に関係なく、他の個人と同様に扱われている状態を指す individual fairness であり、もう一つは、集団の中で

男女等のセンシティブグループ間の公平性を指す group fairness である。さらに group fairness の主要な規準として statistical parity[Dwork 2012]や equalized odds[Hardt 2016]がある。

これらは異なる指標であり、同時に達成することができないことがわかっている。そのため公平性を測るには、目的とする要件を満たしていることを確認できる手法を適切に選択する必要がある。

### 2.3 実運用における公平性

最後に、組織やユースケースごとの観点で決定する必要がある、実運用時の要件としての公平性がある。これに関しては組織内の文脈を考慮して判断を下す必要がある。公平性という立場から人種の多様性をもたせるために特定の人種の採用を積極的にするといった判断をする企業もあれば、同じく別の公平性という規準から、まんべんなく各人種から同数の人を採用するなどの方針を持つ組織もあるだろう[東京大学未来ビジョン研究センター 2021]。

そのほかにもたとえば現状の男女構成比が男性に偏っている組織において女性を優先して採用した場合、結果として男女の採用比率は女性に偏るが、組織全体の公平性を見越して戦略的に調整をした結果であると理解することもできる。このように定義は任意に決められるが、株主やクライアント、ユーザー等のステークホルダーが納得できる理に合った定義である必要がある。これらの事例からわかるように、公平性の観点は実運用の現場単位で異なるため明確な規準はなく、使う組織や場面に応じて適切な公平性の観点を判断する必要がある。

今まで AI の公平性を考慮してこなかった企業がこれから検討するとなった場合には、社会的や技術的な論点があることを理解しながら自社の立場に落とし込む必要がある。しかしながら、社会的な論点と技術的な論点の両方を考慮することは容易ではないため、公平性の観点を自社の立場に落とし込むことの難しさは実運用における公平性の議論が進まない一因となっている。

### 3. 日本の課題

世界では、AI の公平性の考慮は開発からサービスまで一貫して自社で提供する大企業(Google, Microsoft, IBM 等)が牽引している。一方、日本は IT 企業とサービス提供企業が分かれている傾向が強く、自社で開発からサービス提供まで実施している企業は少ない。図 1 に、IT 企業とそれ以外の企業における IT 人材の割合を示す[情報処理推進機構 2017]。日本は他国に比べ IT 人材が IT 企業に偏っていることが分かる。つまり日本においては、IT 企業以外は自社内に IT 人材が少ないといえる。

さらに、図 2 に、情報サービス業における下請けの割合を示す[総務省 2018]。開発には多重下請け構造の実態があり、実際にユーザーなどに接する元請け会社はサービス品質としての公平性に対する意識がたとえ高くても、ユーザーに接しない開発プロジェクトメンバーの多くは下請け企業の立場であるため、公平性という点は考慮が行き渡らない可能性がある。さらに問題なのは、現在の日本社会においては、データ収集に関するプライバシーに関しては元請け会社などの B2C 企業やユーザーの意識は高いが、公平性に関する議論そのものが元請け会社のレベルでさえ顕在化していないことも指摘されている[日本ディープラーニング協会 2021]。

このような状況であるため、世界の企業が提示する公平性規準を、サービス提供企業が分かれている日本企業にはそのまま適用できない。したがって、日本企業へ社会的な公平性の

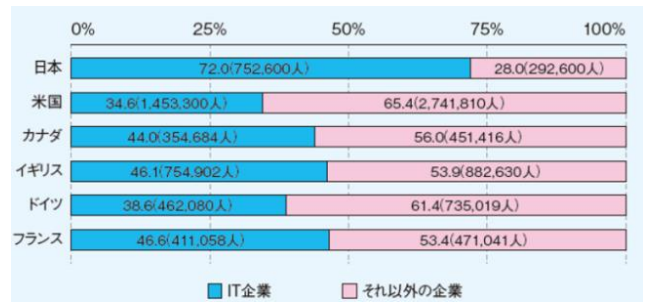


図 1 IT 企業とそれ以外の企業における IT 人材の割合

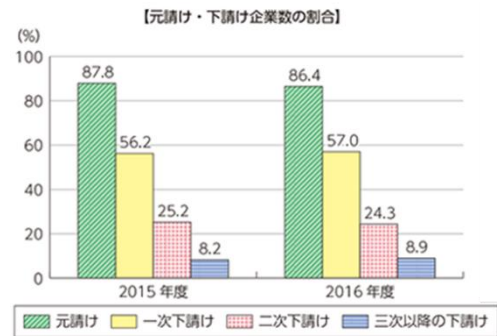


図 2 情報サービス業の元請け・下請けの状況

議論を喚起すると同時に、日本の産業構造を踏まえた公平性の検討が必要である。

日本の産業構造においては、AI の公平性を定義しようにもサービス提供企業や IT 企業とその下請け等のステークホルダーが非常に多く、最初に決めた要件を開発工程で検証する際のコミュニケーションが起りやすい状態となっている。

ステークホルダーの多さの具体例として、日本政府のガイドラインを取り上げてみたい[デジタル庁 2021]。日本政府はシステム開発を委託することを前提としており、内部にシステム開発部署を持っていない。そのため、日本の産業構造の特徴である IT 企業とサービス提供企業が分かれている典型例としてみることができる。

調達単位を表 1 に示す[デジタル庁 2021]。調達を分割することで全体のコスト削減や事務処理の軽減を目指しているため、限りなく小さな単位に分割されているが、そのせいで調達間の意思疎通は取りづらくなっている。この状態で、日本で公平性を考慮した AI の導入を進めるには、それぞれの調達単位ごとの公平性の考慮方法をガイドライン等で明らかにしなければならない。

### 4. 調査目的と方法

以上から本研究では、日本の産業構造を踏まえた調達単位の違いに照らして、公平性関連技術を調査・分類することで、日本企業が AI 導入の際に公平性を考慮するために必要な情報を明らかにし、今後日本企業に AI 導入が広がった際に公平性に対する適切な理解を深めることを目的とする。

そこで、政府のガイドラインに記載された調達単位を、システム開発工程に必要な企業の役割と仮定し、AI の開発において公平性を考慮すべき役割をピックアップする。

次に、既存のサービスやソフトウェア、プログラム等の技術を机上調査し、その中から公平性に対する技術的な解決方法を具体的に示しているものをピックアップする。そのうえで、ピックアップした技術を対象に、企業が利用する際に分かりやすいと

表 1 調達単位

No.	調達単位
1	調査研究又は要件定義作成支援
2	プロジェクト管理支援
3	設計・開発
4	クラウドサービス利用
5	ハードウェアの賃貸借又は買取り
6	ソフトウェア製品の賃貸借又は買取り
7	回線
8	アプリケーションプログラムの保守
9	ハードウェアの保守
10	ソフトウェア製品の保守
11	運用
12	運用サポート業務
13	業務運用支援
14	施設の賃貸借
15	施設の整備等
16	システム監査

この観点で作業課題を分類し、選定した役割に応じてどの役割の企業が考慮すべき作業課題かを特定する。

## 5. 結果

### 5.1 公平性を考慮すべき調達単位と企業の役割

政府のガイドラインに記載された調達単位を、AI を用いたシステム開発に当てはめ、公平性に関する考慮が必要な調達単位とその調達を請け負う企業の役割を表 2 に整理した。表 2 の番号は表 1 の調達単位と対応している。

### 5.2 各調達単位で考慮すべき作業課題

調査した既存技術を作業課題ごとに分類した結果を表 3 に示す。調査では、無償公開されているものとして 23 件、有償のものとして 6 件の既存技術の情報が得られた。ここでは分類した作業課題と一部の技術例を記載する<sup>[注 1]</sup>。

技術の分類は、対象とするものとして入出力データとモデルがあり、それぞれ評価・可視化の技術と補正の技術があるため計 4 つに大別した。4 つの作業課題と、作業課題ごとに対象の課題を解決するための技術例を表 3 に示す。示した技術例は必ずしも作業課題ひとつを対象とするものではなく、複数の作業課題を対象としているものもある。

表 2 と表 3 を組み合わせ、調達単位ごとに考慮対象となる作業課題を割り当てたものを表 4 に示す。調達単位の中で、要件を定義する(No.1)企業は入出力データの公平性を考慮すべきであり、設計・開発(No.3)や運用(No.11)に関わる企業は入出力データとモデルの公平性を考慮する必要がある。ソフトウェア製品(No.6)に関しては、ソフトウェアの処理対象が入出力データなのかモデルなのかで考慮すべき作業課題が異なる。

表 2 調達単位と企業の役割

No.	調達単位	企業の役割
1	調査研究又は要件定義作成支援	・一般的な公平性概念の把握 ・プロジェクトの公平性規準の決定
3	設計・開発	・プロジェクトの公平性規準の把握 ・技術選定、開発
6	ソフトウェア製品の賃貸借又は買取り	・製品開発
11	運用	・プロジェクトの公平性規準の把握 ・評価、改善

表 3 作業課題と技術例(括弧内は開発企業)

作業課題	技術例	全課題を対象とする技術例
入出力データの評価・可視化	Aequitas (University of Chicago)	AI Fairness 360 (IBM) Responsible AI (Google) themis-ml (arena)
入出力データの補正	Presidio (Microsoft)	
モデルの評価・可視化	Fairlearn (Fairlearn)	
モデルの最適化	LIFT (LinkedIn)	

表 4 調達単位ごとの考慮すべき作業課題

No.	調達単位	考慮すべき作業課題			
		入出力データの 評価・可視化	入出力データの 補正	モデルの 評価・可視化	モデルの 最適化
1	調査研究又は要件定義 作成支援	○			
3	設計・開発	○	○	○	○
6	ソフトウェア製品の賃貸借 又は買取り	○	○	○	○
11	運用	○	○	○	○

## 6. 考察

表 2 において調達単位と企業の役割を整理したが、調査研究又は要件定義作成支援(No.1)の段階において決定したプロジェクトの公平性規準は、設計・開発(No.3)段階と共に運用(No.11)段階でも把握しておく必要があるため、この二つの調達に関わる企業は特に連携を密に取りながらプロジェクトを進めていく必要がある。

既存技術の調査では、入出力データの評価・可視化とモデルの評価・可視化の技術数が多く、入出力データの補正とモデルの最適化の技術数が少ない傾向にあった。これは、主に前者が「技術的な公平性」、後者が「実運用における公平性」に該当するためだと考える。入出力データやモデルの評価・可視化は中立の立場で事実を可視化する技術として用いることができるが、補正や最適化をする際の観点は実運用で決められる要件であることから、企業や組織の文脈に応じた判断が必要となり、技術に落とし込みにくいためである。今後、公平性を考慮する上で頻出する観点を解決できるような技術を整理することで、企業の観点と合致する技術が見つかるようになり、公平性に対する実運用上の考慮がより促進されるだろう。

また、3 節にて示したように日本の産業構造は多重下請け構造という特徴を持つため、公平性に関する企業ガバナンスを統治する余裕がない企業が多いことが想定される。多重下請け構造の改善が優先だが、例えば元請け企業が公平性に関するガバナンスを考える部署や人材を置くことでプロジェクト全体の公平性の考慮も進むだろう。近年では AI 倫理やガバナンスに関する部署や人材を配置する企業も増えてきており、今後この流れは加速すると考えられる[江間 2021]。また国によっては、政府調達への参加条件として AI の公平性などを含む原則を制定していることを条件として課すところも出てきており、多重下請け構造がありながらも、AI サービスやシステムの公平性を担保する

[注 1] 調査結果の全体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 AI サービスとリスクコーディネーション研究会ウェブサイトにて公開予定であり、随時更新していく。<https://ifi.u-tokyo.ac.jp/projects/ai-service-and-risk-coordination/>

ための仕組みづくりを積極的に行っていくことが、企業や組織には求められる。

本論文では日本政府のガイドラインに沿って役割や技術の割り当てを行ったが、ガイドラインの調達単位が非常に細かく分割されており、AIの公平性をプロジェクト全体で考慮する上ではミスコミュニケーションが発生しやすくなる。たとえば運用に関しても、表1では運用(No.11)、運用サポート業務(No.12)、業務運用支援(No.13)と分かれており、業務分担を適切に分解しないと調達全体で一貫して公平性を考慮することが難しいと考えられる。このことから、調達単位自体を見直すことも考慮すべきであると考えられる。

## 7. 今後の課題と展望

本論文では日本の産業構造の現状に基づき、日本政府のガイドラインの調達範囲に沿った形で企業の役割ごとの作業課題を整理した。これにより、それぞれのプロジェクトでの企業の立場に応じて参照し、適切に公平性を考慮することができるようになるだろう。ただし、考慮すべき作業課題が異なるだけで、調達元の企業がサービスの責任を負うことは変わらないため、入出力データとサービスの狙いの関係性を十分に説明できるようにすることが必要である。

また、AIの公平性の考えが浸透する中で初めて考慮する企業の存在も考慮し、作業課題は大まかに分類しているが、必要に応じて詳細化が可能であると考えられる。設計・開発時、モデルをブラックボックスとして扱う形態もあるため、その場合の設計・開発はソフトウェア製品のように考慮すべき作業課題が分かれることも考えられる。

今後、企業の立場だけでなく、具体的なユースケースと組み合わせで整理することで、より実効性のあるAIの公平性を達成するための研究へとつなげていく予定である。

## 参考文献

- [Dwork 2012] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, R. Zemel: Fairness Through Awareness, Proc. of the 3rd Innovations in Theoretical Computer Science Conf., 2012
- [Flores 2016] Anthony W. Flores, Kristin Bechtel, Christopher T. Lowenkamp: False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks.", <https://www.uscourts.gov/federal-probation-journal/2016/09/false-positives-false-negatives-and-false-analyses-rejoinder>, 2016
- [Hardt 2016] M. Hardt, E. Price, N. Srebro: Equality of Opportunity in Supervised Learning, Advances in Neural Information Processing Systems~29, 2016
- [Jobin 2019] Anna Jobin, Marcello Ienca, Effy Vayena: The global landscape of AI ethics guidelines, Nature Machine Intelligence VOL 1, 2019
- [Mehrabi 2019] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, Aram Galstyan: A Survey on Bias and Fairness in Machine Learning, arXiv:1908.09635., 2019
- [江間 2021] 江間有沙, 絵と図でわかる AIと社会: 未来をひらく技術とのかかわり方, 技術評論社, 2021
- [神寫 2022] 神寫 敏弘: 私のブックマーク「人工知能と公平性」, 人工知能, VOL.37, NO.2, 2022
- [情報処理推進機構 2017] 情報処理推進機構, IT 人材白書 2017, <https://www.ipa.go.jp/jinzai/jigyoku/back.html>, 2017
- [総務省 2018] 総務省, 平成 30 年版情報通信白書, <https://www.soumu.go.jp/johotsusintokei/whitepaper/h30.html>, 2018
- [デジタル庁 2021] デジタル庁, デジタル・ガバナメント推進標準ガイドライン解説書, <https://cio.go.jp/guides>, 2021
- [東京大学未来ビジョン研究センター 2021] 東京大学未来ビジョン研究センター, AI サービスに対するリスクベースアプローチ: リスクチェーンモデルと採用 AI, <https://ifi.u-tokyo.ac.jp/event/10489/>, 2021
- [統合イノベーション戦略推進会議 2019] 統合イノベーション戦略推進会議, 人間中心の AI 社会原則, <https://www8.cao.go.jp/cstp/aigensoku.pdf>, 2019
- [日本ディープラーニング協会 2021] 日本ディープラーニング協会, 研究会: AIガバナンスとその評価より報告書『AIガバナンス・エコシステム—産業構造を考慮に入れたAIの信頼性構築に向けて—』, <https://www.jdla.org/about/studygroup/sg01/>, 2021