



日本企業が人工知能標準化活動に参加すべき理由 —国内ワークショップへの参加を通じて—

株式会社KDDI総合研究所 知能メディアグループ グループリーダー

ほ あし けい いちろう
帆 足 啓 一 郎



1. はじめに

本特集にて紹介されているとおり、ここ1～2年の間に人工知能に関連する標準化、ならびにガイドライン策定の活動が世界中で立ち上がっている。標準化といえば主に標準的な技術仕様の策定をイメージしがちだが、昨今盛り上がっている人工知能関連の標準化活動のテーマは、技術的な仕様のみならず、人工知能に関連する倫理や法整備など、多岐にわたっていることが特徴的と言える。人工知能関連技術の発展は、今後の人類や社会に与えると想定される影響の範囲が幅広く、インパクトが高いと考えられているためである。

本稿執筆時点において、これらの活動に直接参加している組織・団体は、人工知能関連技術の研究開発を最先端で進めている企業が多い。具体的には、Google、Facebook、Microsoftといった米国の巨大IT企業が各活動に参加しているほか、Partnership on AI^[1]という企業コンソーシアムを結成し、独自のガイドライン策定を進めている。また、これらの企業はそれぞれ人工知能関連ソフトウェアを開発するための機械学習ライブラリなどをオープンソースとして無償提供したり、APIなどを公開したりすることによって、多くの開発者を取り込んだ上で、デファクトスタンダードになることも狙っている。

こうした人工知能関連の標準化活動に直接参加している日本企業は、現時点では少ない。当該領域における研究開発や実用化の実績が乏しく、実質的にどこまで貢献できるのかの判断が難しいことが主な理由の一つであろう。しかし、人工知能の標準化やガイドラインをより有意義なものにするためには、日本企業の参加は必要不可欠と考えている。そこで本稿では、日本国内で行われている人工知能関連ガイドライン策定活動の一環として開催されているワークショップに、筆者が直接参加して得られた知見を踏まえ、人工知能関連技術の開発に携わってなくても、積極的に標準化活動に参画すべき理由について意見を述べる。

2. 日本国内における人工知能議論の活性化

人工知能関連の標準化活動は、前述のとおり、技術的な仕様策定に限定されていない。そのため、標準化活動への参加自体については門戸が広く開かれており、工学系・情報系の専門家以外の参加も歓迎されている活動がほとんどである。

こうした活動の活発化を機運として、日本国内における人工知能に関する議論を活性化するための取組みとして、一般人も含めて議論を行うためのワークショップが昨年からたびたび開催されている。例えば、2017年9月に日本科学未来館で開催された参加型対話イベント「人工知能社会のあるべき姿を求めて」^[2]では、多方面の専門家が話題提供者として参加し、約70名の一般参加者とともに、人工知能やロボットのあるべき姿について議論を交わした(参考:筆者による参加レポート^[3])。また、同年10月には、前述のPartnership on AIやIEEEなどから多数のゲストスピーカーが招かれた大規模なシンポジウム「AI and Society」^[4]が虎ノ門ヒルズで開催され、500名超の参加者が集まった。

さらに、本シンポジウムの後、AIの社会における影響などに関するより深い議論を行うため、クローズドな参加者によるワークショップ「Beneficial AI Tokyo」^[5]が開催された。このワークショップでは、人類・社会にとって有益な人工知能(“Beneficial AI”)を実現することを目的とし、予め定められたテーマについて、参加者によるグループディスカッション形式で議論を行った。これらの議論の集約として、Beneficial AIを実現するための協力を謳った宣言「Cooperation for Beneficial AI」が採択され、賛同者の氏名とともにWeb上で公開されている^[6]。

こうした一連の国内における動きは、人工知能の標準化議論への参加者のダイバーシティを上げたいという、人工知能に関する国際コミュニティの意志に基づいている。現在の議論



写真1. Beneficial AI Tokyo参加者の記念写真
(<http://www.ai.u-tokyo.ac.jp/beneficial-ai-tokyo.html> より引用)



の主体となっているのは欧米諸国だが、より健全な議論を行うためには、アジア及びアフリカも含めた発展途上国からの参加が必要とされている。このような参加者の多様化（特に中国などアジア諸国からの参加）を推し進められる存在として、日本に対する国際コミュニティの期待は高い。この期待に応える形で、国内におけるディスカッションの活性化が図られている。

3. IEEE EAD国内ワークショップ

前述のような国内でのワークショップ活動は、2018年に入ってから継続的に行われている。その一つとして、IEEEから2017年12月に公開された「IEEE Ethically Aligned Design, v2」^[7]（以下、「EADv2」）を対象とした議論をテーマとした連続ワークショップ「IEEE “Ethically Aligned Design” Workshops in Japan」が、東京大学の政策ビジョン研究センター・特任講師の江間有沙氏を中心に企画され、2018年2月～3月にかけて開催された^[8]。筆者は、この一連のワークショップのうち、筆者自身の研究領域との関連性が高いテーマが取り上げられた、第2回（2018年2月23日開催）に参加し、人工知能に関連する標準化活動に企業として注目すべき理由を実感した。以下、その内容について紹介する。

4. AIによる感情操作の可能性

上記の第2回ワークショップにおける議論のテーマは、EADv2から新たなテーマとして取り上げられている「アフェクティブコンピューティング」（Affective Computing）と「複合現実」（Mixed Reality）の2つである。当日のワークショップは、これらのテーマに関連する研究に取り組んでいる専門家をゲストスピーカーとして招き、まず議論のインプットとなる講演を行ったあと、その内容について参加者で議論が交わされる内容だった。

本ワークショップのゲストスピーカーの一人として登壇したの

は、拡張現実関連の技術と、心理学などの知見を融合した研究を行っている東京大学・鳴海拓志氏。講演のパートでは、自身による研究成果を中心に、複合現実感を使った人間の心理状態や行動の操作の可否について紹介した。

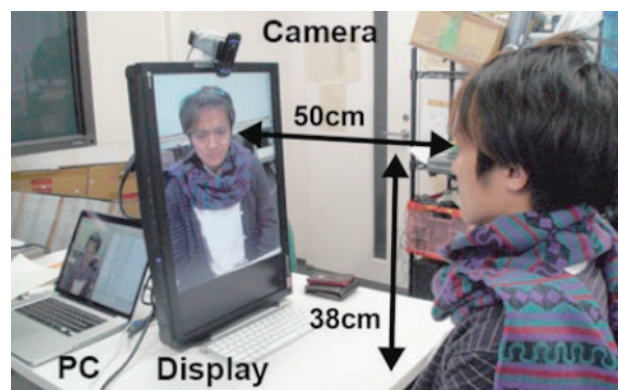
講演の中で紹介された複数の研究事例のうち、筆者の印象に残ったのは「人間の感情を作る」ことを目的とした研究である。講演の中で紹介された事例の一つであるYoshidaらの研究^[9]では、鏡に見立てたカメラ付きディスプレイ上にリアルタイム画像処理技術を適用し、目の前にいるユーザの表情を擬似的に変化させるシステムが開発されていた。このディスプレイを使うことにより、目の前にある自身の表情に笑顔を付与したり、悲しい表情に変換したりすることが可能になるが、この表情の変化によって、実際に被験者の感情に影響を及ぼすことができるかを検証する実験を実施している。その結果として、例えばディスプレイに映る顔表情を笑顔に変換すると、被験者の感情もポジティブになることなどが確認された。

この技術の応用例として、表情の疑似変化による被験者の購買意向の影響を調べるための実験が紹介された。ディスプレイの前で2つのマフラーを試着させた上で、片方のマフラーを試着した場合にのみ、ディスプレイに表示される顔表情に笑顔効果を施した。その後、被験者に好みのマフラーを選んでもらうと、笑顔効果が施された状態で試着したマフラーの方が有意に高い評価が得られることが実証された。つまり、人間の表情を加工するだけで、品物に対する印象の操作が可能であり、購買行動に影響を及ぼす可能性が示されたのである。

また、Nakazatoらの研究^[10]では、この技術の別の応用場面として、ビデオ会議システムを使ったブレインストーミングを取り上げている。ビデオチャット中に、参加者の表情を変化させることにより、ブレストで出たアイデアの数を比較したところ、参加者の表情を笑顔にした状態の方が無加工の場合と比較して平均1.5倍に増えたという結果が得られている。



■写真2. ワークショップで講演する鳴海氏
（[8] の第2回ワークショップ紹介ページより引用）



■写真3. 表情変化による選好判断操作実験の様子（[9] より引用）



以上の研究事例などから総じて言えることは、計算機やソフトウェアを使い、複合現実のようなインタフェースを介した情報提示を少し工夫するだけで、人間の感情や行動がいつも簡単に操作できるということである。鏡の中の表情を違和感なく変化させたり、閉鎖空間以外の場においても自然な環境でユーザに提示したりする技術が世の中に広まるためにはまだ多くの技術的な課題が残っているが、本講演を通じて、複合現実の活用がもたらすポジティブ・ネガティブ双方の可能性を垣間見ることができた。

5. AI発展によるガイドライン対象の拡大

EADv2の中では、アフェクティブコンピューティングの項目の中で、以下の5つのサブ項目が記載されている（カッコ内は筆者による意訳）：

- ・Systems Across Cultures（異文化をまたぐAI設計）
- ・When Systems Become Intimate（人と親密になるAI）
- ・System Manipulation/Nudging/Deception（AIによる人の操作/ナッジ/欺瞞）
- ・Systems Supporting Human Potential（Flourishing）（AIによる潜在能力支援・活性化）
- ・Systems with Their Own Emotions（感情を持つAI）

前述のワークショップにおける鳴海氏の紹介内容は、これらのサブ項目の中の「System Manipulation/Nudging/Deception」の議論の対象に含まれる。より具体的には、同項目の中で取り上げられている課題（issue）のうち、「Should affective systems be designed to nudge people for the user's personal benefit and/or for the benefit of someone else?」（意訳：アフェクティブAIは人間自身もしくは第三者の利益のために、人間の感情を操作できるように設計されるべき？）との関連性が高い。この課題については、標準化のcandidate recommendationsとして、例えば、ユーザを操作/ナッジする可能性のあるAIを設計する際の倫理的な検証プロセスの必要性、ユーザがAIからのナッジを明示的に意識できる情報提示方法、及び、AIを利用する前にオプトイン方式でユーザの承諾を得る仕組みの導入などが推奨されている。

前述のYoshidaらの研究で紹介したマフラー実験のシステムを、ユーザへの事前説明がない状態で洋品店などに展開することは、多くの人にとっては「不適切」な方法として捉えられるだろう。こうした分かりやすい問題を顕在化させ、不要な「炎上」を回避するためのガイドラインとして、EADv2が有効に機能する可能性は高い。

より具体的に考察する。マフラーの事例を実際の店舗など

で実用化する場合、ポジティブな表情がユーザの購買意欲を高めるという知見を、鏡に映されている表情をシステムが加工して提示するインタフェース上に実装することになる。ここでいう「システム」を広義の「AI」と捉えると、議論の対象になり得るのは、人間の深層心理に関する「知見」の活用と、その結果をユーザに提示する「インタフェース」の是非と考えられる。

この両者のうち、「知見」は、顔の表情が自身の感情に対して直接影響を及ぼす効果があるという心理学的な知見に相当する。この知見を、実際の商品販売の現場で活用することに対する違和感はあるかもしれない。ただ、例えば店員によるセールストークや、店舗のデザインなども、様々な心理的な知見に基づいている。したがって、AIのガイドラインの規制の対象からは外れるという解釈は許容可能に思える。

一方、後者の「インタフェース」については、AIがユーザに提示している情報が現実とは異なる形に加工されるという特徴がある。これは、EADv2の「AIによる人の操作/ナッジ/欺瞞」で懸念されている範疇に含まれると解釈される可能性は高い。より具体的には、Candidate recommendationsの中で明言されている、（例えば）商用目的のための心理操作に該当する可能性が高い。そのため、実店舗での実装の際には、AIのインタフェースが行っていることをユーザに明示的に示すような対策が必要になるかもしれない。（そうなった場合、そもそも実用化の判断は見送られる可能性は高いが。）

以上の考察のとおり、EADv2などのAI関連ガイドラインの策定と適用によって、不適切なAIの運用に対する一定の抑止効果は期待できる。しかし、今後AI関連技術の発展がさらに加速し、実世界への展開が広まるにつれ、適用範囲の線引きはかなり難しい問題となり得る。

例えば、Webサイトのデザインなどを最適化するための仕組みとして、「A/Bテスト」と呼ばれる方法がある。最もシンプルな「A/Bテスト」としては、Webサイトの改修版インタフェースを、当該サイトの一部のユーザ向けに公開する方法がある。このテストの結果、改修版と現状のインタフェースのそれぞれに対するユーザの反応を計測することにより、改修前後で想定される効果が得られるかを効率的に判断することができる。

この「A/Bテスト」に対し、AI関連技術を適用することによってWebサイト最適化を限りなく自動化することができる。具体的には、Webサイトの管理者（人間）が追求するKPI（例：PV数、サイト滞在時間など）と、調整したいパラメータを事前に与えることにより、「A/Bテスト」システムであるAIが自動的にパラメータ設定条件などの候補を出力し、その出力結果をランダムにユーザに提示することができる。そして、これらの



結果に対するユーザ反応を測定し、最適な結果が得られた条件をWebサイトに自動的に反映させれば、管理者にとって最も良好な結果が得られるWebサイトの導出が可能になる。

こうしたプロセスはすでに多くのWebサイトの運用方法の一つとして広まっており、「AIによる人の操作/ナッジ/欺瞞」に該当するという印象はない。しかし、前述のマフラーの事例と比較すると、実質的な差はあまりないようにも思える。「A/Bテスト」では、サイト運営者が定めた特定の目的を満たすため、Webサイトのインタフェースを変化させている。一方、マフラーの事例では、「特定の商品の売り上げ」が目的であり、この目的を果たすための手段として、インタフェースを通じて商品の見栄えを改善していると言える。

また、マフラーの事例では、人間が導き出した心理学的な知見を活用しているのに対し、自動「A/Bテスト」ではユーザの行動に影響を及ぼす要素をAIが自動的に導出している点も考慮する必要がある。この自動化の処理によって、むしろ「A/Bテスト」の方に、Webサイト運営者が想定していない結果が生まれる危険性がある。その中に、不適切なナッジとみなされる結果が含まれた場合、AI関連ガイドラインに抵触する事態となり、サイト運営者としての責任が問われる可能性もある。

このように、現時点では受け入れられているプロセスやインタフェースであっても、AI関連技術の発展によって不当なAIの利用とみなされる可能性はある。事実、Facebookのユーザが普段閲覧しているタイムラインに表示される情報の最適化アルゴリズムが、フェイクニュースを拡散させ、一部の政党や国家による悪用につながっているという疑惑が米国議会で取り上げられており、Facebookの社会的・倫理的な責任が厳しく問われている。この例のように、サイト運営者にとって健全な目的を達成するためにAI関連技術を活用していても、その後の技術発展や社会情勢の変化によって問題となる事例は今後も発生すると思われる。

6. AI「ユーザ」としてのガイドライン作り参加のすすめ

現状、AIを直接開発していない企業にとっては、AI関連ガイドラインは遠い世界の議論に思えるかもしれない。しかし、今後あらゆる事象がデータ化され、AI関連技術によって解析され、実世界での様々な判断や出来事に影響が広まることは明らかである。その場合、自らAI関連技術を開発してなくとも、社外で作られたAI関連技術をユーザとして活用しながら、自社の事業展開を進めている可能性は高い。そして、仮にAI関連技術を一ユーザとして活用していたのだとしても、AIガイドラインの抵触や、その先に問題が発生した場合の責

任は当然ながら自社が負うことになる。こうした状況に鑑みると、少なくともAI関連ガイドラインや標準化の議論について、注視する必要性は明らかである。

AI関連の標準化やガイドラインにおける議論の対象がいわゆる技術的な仕様以外に、社会的な課題や倫理などの幅広い内容であるにも関わらず、現時点で積極的に議論に参加しているのは欧米の巨大IT企業と学識者が中心である。そのため、知らないうちに一ユーザとしてAIを活用しようとしている日本企業にとって不利な結論に至る可能性もある。

このような事態を防ぎ、全人類に対してベネフィットをもたらすAIの発展を促すためには、議論参加者のダイバーシティを広げ、多種多様な意見が取り入れられるガイドライン作りが不可欠である。本記事の前半で説明したとおり、AI関連の標準化議論に対する参加者の偏り(欧米偏向)は、国際コミュニティも問題視しており、この偏向の解消のため、日本の参加に対する期待は強い。いきなり標準化の議論の席についても良いが、それがためらわれるのであれば、前述のような草の根のワークショップへの参加を通じて、AIに関する議論への理解を深め、意見をインプットすることも可能である。まずは個人的にでもこうした場に参加し、様子を見ながら徐々にガイドライン作りに参画してみてはいかがだろう。

参考文献

- [1] "Partnership on AI", <https://www.partnershiponai.org/>
- [2] 人工知能社会のあるべき姿を求めて, <http://ai-dialogue.strikingly.com/>
- [3] 帆足:「AI暴走ニュースの先にある人工知能社会のあるべき姿」, ASCII Webサイト, <http://ascii.jp/elem/000/001/563/1563295/>
- [4] "AI and Society", <http://www.aiandsociety.org/>
- [5] "Beneficial AI Tokyo", <http://www.ai.u-tokyo.ac.jp/beneficial-ai-tokyo.html>
- [6] "The Tokyo Statement—Cooperation for Beneficial AI", <http://www.ai.u-tokyo.ac.jp/tokyo-statement.html>
- [7] IEEE Standards Association: Ethnically Aligned Design v2, <https://ethicsinaction.ieee.org/>
- [8] IEEE "Ethically Aligned Design" Workshops in Japan, <https://sites.google.com/view/ethically-aligned-design-ws/eadv2-workshop>
- [9] Yoshida, S., Tanikawa, T., Sakurai, S., Hirose, M., & Narumi, T. (2013, March). Manipulation of an emotional experience by real-time deformed facial feedback. In Proceedings of the 4th Augmented Human International Conference (pp. 35-42). ACM.
- [10] Nakazato, N., Yoshida, S., Sakurai, S., Narumi, T., Tanikawa, T., & Hirose, M. (2014, February). Smart Face: enhancing creativity during video conferences using real-time facial deformation. In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (pp. 75-83). ACM.