

人と社会から信頼される AI を目指して：AI 倫理をめぐる最新の取り組み

WASHINGTON | CORE

2022年8月



業界を超えて活用場面が広がっている人工知能 (AI)。だがAIの影響力が拡大するにつれて、プライバシー侵害や人種・ジェンダーにもとづく差別など、セキュリティや倫理面への懸念も高まっている。AIethicist.orgは、AIが個人や社会に及ぼす影響、AIをめぐる倫理的問題、責任あるAIガバナンスといったトピックについて、最新の研究成果やリソースを1箇所に集めた世界初のAI倫理のグローバルリポジトリ (データ貯蔵所) を運営する非営利団体である。AIethicist.orgの創設者であるマーブ・ヒコック氏 (Merve Hickok) に、AI倫理をめぐる最新の議論や、信頼されるAIの実現に向けた規制の取り組みについて話を聞いた。

倫理面のリスクとしてしばしば話題に上るのが、AIアルゴリズムに生じるバイアス (偏り) の問題である。AIシステムは、個人から収集した膨大なデータをもとに意思決定を行うが、その際に使用されるアルゴリズム自体に、人間社会に蔓延 (はびこ) る人種やジェンダーにもとづくバイアスが反映されてしまう問題が指摘されているのだ。

身近に迫る AI 倫理問題

近年、AIの活用場面は急速に増加している。カスタマーサービスやモビリティなど、生活の身近なところでの活用はもとより、最近ではヘルスケアや人事など、私たちの健康や幸福に直結する重大な場面で様々な判断を補助するサービスやツールも次々と登場している。AIはあらゆる作業の効率化やサービス向上をもたらす反面、膨大なデータの収集・活用に伴うリスクも孕んでおり、諸刃の剣となり得る技術である。

データやAIの活用をめぐる倫理面に関する議論は、当初、アカデミアを中心に行われていた。しかし2018年にFacebook社 (現Meta) から流出した大量の個人データを、データ分析企業Cambridge Analytica社が政治広告に利用したとされる事件が起るなど、近年AIアルゴリズムにまつわる事件が頻発したことで、一般の人々の間でのAI倫理問題への認知度が一気に高まった。Meta社は最近も、人種やジェンダー等で特定のユーザーを排除するアルゴリズムを利用して住宅広告のターゲティングを行っていたとして訴えられ、2022年6月に、制裁金を支払うことで司法省との和解に至ったばかりである。そのわずか1週間後には、Meta社傘下のInstagramが、サービス利用評価の目的で、ユーザーの人種と性別の調査を行うと発表して批判を集めた。このような事象がきっかけとなり、企業による個人データの収集や本人の同意のない分析目的での利用について、一般の人々の関心や認知が大きく高まった。

一方で、AIアルゴリズムが抱えるバイアスの危険性を如実に示した事件としては、2019年にオランダで起きた児童手当の不正受給判定をめぐるスキャンダルが記憶に新しい。オランダは公共サービスのデジタル化先進国であり、税務当局は児童福祉手当の不正受給者洗い出しのためのリスク予測にAIアルゴリズムを導入していた。しかしこのアルゴリズムが二重国籍を持つ者や外国風の氏名を持つ者が高リスクと判断するなど、人種差別を助長する不当な評価を行っていたことが明らかになった。その結果、約1万世帯が育児手当の不正受給者として濡れ衣を着せられ、手当の返還を求められていた。低所得層やマイノリティを中心とする多くの世帯が貧困に追いやられる事態を招き、1,000人以上の子どもたちが養護施設に送られ、自殺にまで追い込まれた被害者も出た。オランダの規制当局は税務当局に対し、法的根拠のない個人データの保持および処理を行ったとして、合計600万ユーロを超える罰金を科した。さらにこの問題によって、オランダ内閣は総辞職に追い込まれた。

社員査定や採用の場面でも用いられるAI

最近では国内外で、企業の人事関連業務におけるAI活用にも懸念が高まっている。社員の業務モニタリング向けに開発されたAIツールには、ウェブカメラを用いて、パソコンに向かう社員の目の動きや顔の表情を認識し、性格や集中度、さらには業務態度を判断するものも存在する。しかしながら、こうしたツールによる感情や表情の分析には十分な科学的根拠がない場合も多い。また、顔の特徴や表情の使い方は人種やジェンダー、文化などによっても異なることから、AIアルゴリズムに内在するバイアスによって、評価の正確さに揺らぎが生じることが懸念されている。実際、顔認識や言語処理アルゴリズムは、女性、有色人種、障害を持つ人、アクセントがある人などを対象とした場合にエラー率が高くなることが分かっている。

このようなリスクが指摘されているにもかかわらず、AIを用いた社員の生産性評価向けツールを導入する企業は増加傾向にある。2020年には、Microsoft社が、アプリ上でのユーザーの行動 (例: 会議の出席回数、コラボレーション、コミュニケーションなど) を追跡・評価する「Productivity Score」という新ツールを発表したが、プライバシーに関する多くの苦情が寄せられ、同社は個々のユーザーを特定できないようツールの修正を行うことを余儀なくされた。

さらに採用の場面でも、人事選考における初期段階のスクリーニングを代行するAIツールの導入が進んでいる。だがこの分野では、AIバイアスによる差別が生じるリスクがとりわけ高いとされており、規制機関も動き始めている。例えば米連邦レベルでは、2022年5月に雇用機会均等委員会 (EEOC) が、ガイダンス文書「障害を持つアメリカ人法とソフトウェア、アルゴリズム、人工知能を用いた求職者・従業員の評価について」を発表し、障害者のニーズや特性を考慮したAIシステムの開発について指針を示している¹。地方自治体レベルでも、ニューヨーク市が2022年5月、AIを用いた自動雇用判断ツールを利用する雇用者に対して、バイアスの有無を確かめる年次監査を義務付ける条例を採択している。カリフォルニア州とワシントンDCでも類似的な州法や条例が検討されており、今後、他州にも規制の動きが広がっていくと考えられる。

「信頼されるAI」実現に向け、高まるAI規制へのニーズ

AIの公平性、倫理性をめぐる議論が世界中で活発化する中、世界各地ではより包括的な規制に向けた機運も高まっている。欧州連合 (EU) は、AI倫理をめぐる様々な懸念を認識したうえで、規制枠組みの構築を目指すAI法案を起草している。同法案では、様々なAIユースケースのリスクレベルの分類を提案しているが、採用に向けた選考プロセスの中で使用されるAIシステムについては高リスクに分類している。その上で、ベンダーに対して、AIシステムの技術仕様公表、適合性評価の実施、継続的リスク管理体制の整備を義務付けている。

一方、米連邦議会では、AIがもたらす可能性のあるプライバシーおよびバイアス面のリスクに対応する3法案、「米国データプライバシー保護法案 (American Data Privacy and Protection Act)」、「アルゴリズム説明責任法案 (Algorithmic Accountability Act)」、「顔認識および生体認証技術モラトリアム法案 (Facial Recognition and Biometric Technology Moratorium Act)」が審議されている。現在のところ、いずれも可決の見通しは定かではないが、企業が利用者のプライバシー侵害に関する責任を負うべきとの見方が超党派で優勢であることから、「米国データプライバシー保護法案」が可決される可能性が比較的高いと見られる。

AI倫理問題を取り扱う情報サイト



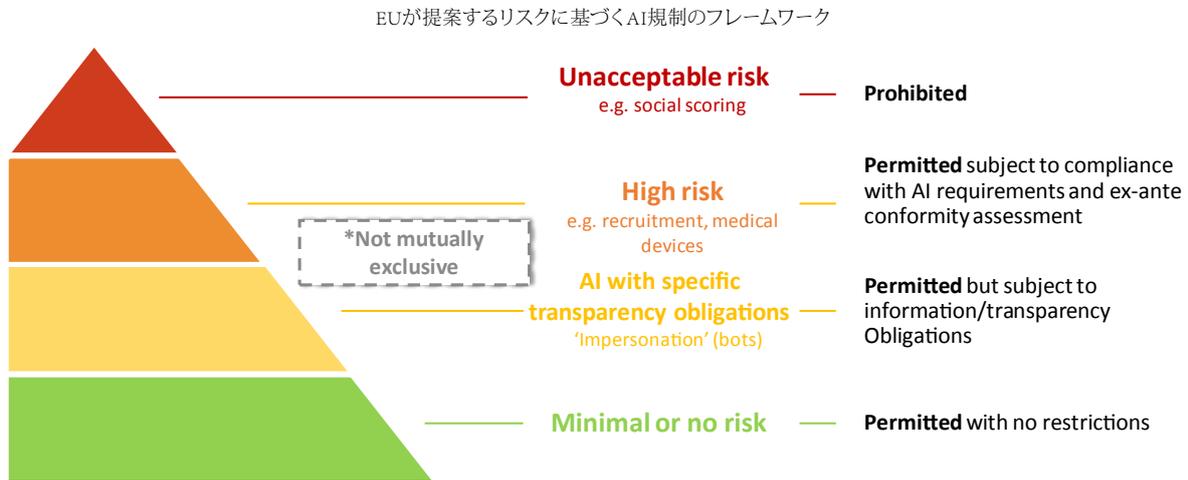
出典: AIethicist.org²

業界団体の米国電気電子工学会 (IEEE) や国際電気標準会議 (IEC) などと並び、米国におけるAI倫理及び製品規格の議論を先導しているのが、米商務省傘下の米国標準技術研究所 (NIST) である。NISTは、AI活用が個人、組織、社会に及ぼしうるリスクの管理を可能にする「AIリスクマネジメントフレームワーク」の開発に取り組んでいる。これはAI関連の事業者及び運営者が自主的に利用することができるフレームワークであり、AI製品・サービス・システムの設計、開発、利用、評価の際に、信頼性を指標として追加するための方法を企業等に示すことを目的とするものである。

企業の主張：AIリスクに応じた規制を

米国のAI関連企業の多くは、前出の3法案に反対して激しいロビイング活動を展開している。企業が反対の立場をとる理由の1つには、これらの法案が、あらゆるAIのユースケースに対して包括的にアルゴリズム監査の義務化を求めている点がある。例えばIBM社は、アルゴリズム監査はAIの活用場面の中でも高リスクな事例に限定されるべきであり、カスタマーサービス向けのチャットボットなど低リスクなものには適用すべきでないと主張している。

だが現時点では、米欧のいずれにおいても、どのようなAIの活用場面が高リスクにあたるのかは明確に定義されていない。したがって今後の規制を考えていく上では、明確な基準を示すことが第一歩となるだろう。EUでは、検討段階ではあるが、下図に示すようなリスク分類が提案されている。



出典: 欧州委員会³

明確なAIの規制フレームワークを整備することは個人の権利の保護に繋がるだけでなく、企業にとっても、法的な拠り所がはっきりと示されることで、AIサービスへの信頼度が上がり、一般消費者によるサービス受容が進むことが期待できる。現在、AI倫理・政策の専門家の間では、信頼されるAIの実現に向け、監査基準、製品規格、行動規範、能力開発、執行メカニズムなど、多方面での基盤づくりを推し進めようとする動きがある。真に人間社会の役に立つ責任あるAIを開発し、その安全な活用をさらに広げていくために、AIのリスクを正しく監査するための仕組みづくりが今こそ求められている。



マーブ・ヒコック

AIethicist.orgの創設者で、AI政策、倫理、ガバナンスの分野の第一人者。2021年、同分野において世界を牽引する女性100名を選出する「100 Brilliant Women in AI Ethics™」に選ばれた。様々な組織に対し、倫理的で責任あるAIの概念と事業化の方法についてのコンサルティングやトレーニングを提供。ミシガン大学でデータサイエンス倫理の講師として教鞭を執るほか、AI政策の開発及び基本的権利、法の支配、民主的価値を守る「信頼されるAI」のあり方を提唱するCenter for AI & Digital Policyのリサーチ・ディレクターも務める。AI倫理の実現に向けた能力開発にも尽力し、50カ国以上のAI政策立案者、活動家、実務者等の指導にあたった経歴を持つ。さらに、信頼されるAIの実現に向けた標準化に取り組むいくつかの団体にも参加している。

後注

¹ <https://www.eeoc.gov/newsroom/us-eeoc-and-us-department-justice-warn-against-disability-discrimination>

² <https://www.aiethicist.org/>

³ <https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf?>

WASHINGTON | CORE

ワシントンコアは、米国の首都ワシントンDCを本拠に、戦略調査・分析とアドバイスを提供する独立系リサーチ・コンサルティング会社です。政治経済ビジネス動向および技術革新を取り巻く環境が刻々と変化する中、弊社が提供する質の高いリサーチ分析とアドバイザリー・サービスと、グローバルに活躍する各界エキスパートとの独自ネットワークを駆使したベストソリューションを提供します。 <https://www.wcore.com>



クリス・ウッド (Chris Wood) chris@wcore.com
シニア・リサーチ・アナリスト

日本の官公庁・業界団体・企業向けに、米国を中心とする諸外国のハイテク産業等(IT、バイオ、高度製造等)の市場・技術・政策調査を担当。APEC や世界銀行などの調査プロジェクトに従事。サステナビリティ、通商関係、中小企業・ベンチャー振興、イノベーション・競争力などのテーマで数々のエキスパートにインタビュー実施。ウィットテンバーク大学卒業後、インディアナ大学大学院よりMBA取得。