情報と倫理：第 5 回
ビッグデータと人工知能の問題

名古屋大学情報学部 2022 年度秋 I 期

久木田水生

データ経済の勃興

今日の情報環境

- 今日、私たちのあらゆるオンラインの行動、そしてますます多くのオフラインの行動のデータが様々なインターフェースを通じて収集・保存されている。
- データはしばしば匿名化されて扱われるが、それで個人が特定できなくなるわけではない。複数のデータから個人が特定できる場合もしばしば。
- そういったデータは Google、 Facebook、Amazonなどの巨大プラットフォーム企業によって独占され、機械学習によって処理され、人々の行動を予測したり、行動を変容させたりするために使われている。
- データそのものや、データを保持する会社を売却することもある。

ある休日の鶴舞公園（名古屋市）。多くの人がポケモンGOに興じていた。

「インターネットの原罪」

- メディア研究者のイーサン・ザッカーマンは、**無料でサービスを提供する代わりに広告を提示するビジネスモデル**がインターネットの「原罪」だった、と振り返る。
- より効果的に広告を提示するために、プラットフォーム企業は**ユーザーの個人データやインターネット上での活動の記録を貪欲に収集する**ようになった。

  Ethan Zuckerman, ``The Internet's original sin: It's not too late to ditch the ad-based business model and build a better web'', *The Atlantic*, Aug 14, 2014.

「インターネットの原罪」

- 収集したデータは**機械学習システムに供給され、ユーザーのプロファイリング、行動予測に利用される**。
- そうしてプラットフォーム企業は、ユーザーが興味を持つ見込みがより高い広告を選択的に提示する。
- またプラットフォーム企業は広告だけでなく、コンテンツもまたユーザーから取得したデータに基づいて選択的に提示している。
- このようにして私たちは**オンライン上での行動データからプロファイリング、行動予測されて、個人個人にカスタマイズされた広告とコンテンツを受け取る**。

Googleのプライバシー・ポリシーに見るインターネットの変化

- Charlie Warzel and Ash Ngu, ``Google's 4,000-Word Privacy Policy Is a Secret History of the Internet'', *The New York Times*, July 10, 2019.
- Googleのプライバシーポリシーがここ20年の間にどのように変わっていったかを見れば、インターネットの歴史が垣間見える。
- 最初は600語だったものが4000語になっている。
- 最初は「ユーザーの集団的な検索活動のデータは第三者に渡されるが、個人の検索活動のデータは渡されない」だった。
- Googleが収集するものは、最初は「集団的な検索活動」、「あなたがGoogleに提供するデータ」、「clickthrough information」、「クッキー」の4項目だけ。いまはGoogleサービスを利用する際のありとあらゆるデータ。箇条書きにして40項目以上。

Googleのプライバシー・ポリシーに見るインターネットの変化

- 最初はターゲティング広告については言及されなかった。2005年に初めて``customized content and advertising''というフレーズが現れる。2009年、DoubleClickがGoogleに買収された直後に「Ad selection」が言及される。
- 2004年に「あなたがアカウント持っているならば、シームレスな経験をあなたに与えるために、そして我々のサービスの品質を向上させるために、あなたのアカウントの下で提出した情報を、我々のすべてのサービスの間でシェアすることがあります If you have an account, we may share the information submitted under your account among all of our services in order to provide you with a seamless experience and to improve the quality of our services」という条項が現れる。

Googleのプライバシー・ポリシーに見るインターネットの変化

- 2012年、センシティブなデータの共有について、opt-inの同意を要求する、と定める。
- 2012年、Googleのサービスを通じてコンテンツをアップロードした、あるいは他の仕方で提出した場合は、Goolge（あるいはGoogleと協働している企業）に対して、そのようなコンテンツを「利用する、提供する、複製する、修正する、派生的作品を作る、通信する、公開する、公に演じる、配布することのワールドワイドなライセンスを与える a worldwide license to use, host, store, reproduce, modify, create derivative works (such as those resulting from translations, adaptations or other changes we make so that your content works better with our Services), communicate, publish, publicly perform, publicly display and distribute such content」ことになる、という規定を付け加える。
- この「コンテンツ」には検索クエリ、キーストローク、チャットビデオを使った時の声紋や顔、Gmailで送ったメールなどが含まれる。
- 2018年、GDPR（欧州一般データ保護規制）を受けてポリシーの大規模なオーバーホールが行われた。ユーザーによるデータのコントロール、データのエクスポートについて、データの消去の仕方についてのセクションが加えられた。

データの価値

- スマートフォンやソーシャルメディアの普及は、収集できるデータの種類と量を増大させた。
- 人工知能（機械学習）の発展は膨大なデータから人々の行動や嗜好を正確に予測することを可能にした。
- こうしてプラットフォーム企業は、ユーザーが興味を持つ見込みがより高い広告やコンテンツを選択的に提示することができるようになった。

データの価値

- 収集されるデータの種類と量が多ければ多いほど予測の精度があがり、また予測できる事柄も増える。
- 消費者の思考や行動をより正確に予測できれば企業はより効率の良いアクションを選択することができる。
- それゆえにIT企業はますます貪欲にデータを収集し、予測のためのより良いアルゴリズムの開発に莫大な投資を行なう。

データの価値

- 行動や属性を予測するのみならず、企業は適切なタイミングで情報を与えることで**人々の行動を操作することすらできる。**
- このような手法はマーケティングを超えて、**選挙や世論操作や紛争にも組織的に利用されている。**
- Cf. ブリタニー・カイザー、『告発　フェイスブックを揺るがした巨大スキャンダル』；ジェイミー・バートレット、『操られる民主主義――デジタル・テクノロジーはいかにして社会を破壊するか』

データと人間

データ経済の中の人間

- データに基づくプロファイリングに興味を持つのはIT企業だけではない。
- 例えば、労働者としての能力や適性、犯罪を犯す可能性、ローンを返済する能力、特定の病気に罹る（あるいは罹っている）可能性、交通事故を起こす可能性、配偶者としての相性、ある政党を支持しているかどうか、等々を確率的に推測するために**データに基づいた人工知能によるプロファイリング**が使われている。
- やがて私たちは**人間生活のあらゆる局面において、人工知能による評価に曝されるだろう**。

データの不適切な使用が問題になった事例

- ケンブリッジ・アナリティカ問題：選挙コンサルタント会社であるケンブリッジ・アナリティカが、**Facebookのユーザーのデータを不正に入手して2016年のアメリカ大統領選のキャンペーンに利用した**ことが疑われた。
- リクナビ内定辞退率予測事件：2018年、リクルートキャリアが運営する就職ポータルサイト、リクナビが、**ユーザーのウェブ閲覧記録などから、ユーザーの内定辞退率を予測するアルゴリズムを開発し、ユーザーの内定辞退率を企業に提供していた**。これに対しては個人情報保護法の他、職業安定法にも違反するとして、政府から是正勧告を出されている。
- センシティブ情報の推測：ユーザのデータから、**病気、妊娠、信仰、政治的傾向などのセンシティブな情報を推測して、マーケティングなどに利用されている**。

データ＝人格？

- 中国などでは様々なデータから算出した「**信用スコア**」によって個人が評価される仕組みが作られている。
- 信用スコアはもともとは、金融に関する情報だったが、近年はその枠を超えて利用されるようになっている。
- 信用スコアの低い人間はクレジットやローンが利用できない、家が借りられないなどのほか、様々なサービスの利用においても不利益を受ける。
- さらに**信用スコアが、人格・道徳性までも判断する材料になっている**。
- しかし信用スコアの算出において、どのようなデータが利用され、どのような基準やアルゴリズムを用いられているかは必ずしも明瞭ではない。

数学破壊兵器

- Cathy O'Neil, *Weapons of Math Destruction*（邦訳のタイトルは『あなたを支配し社会を破壊するAI・ビッグデータの罠』）
- オニールは**アルゴリズムには作成者の「意見」や「先入観」が反映されており、公平ではない**という。
- しかしそのような偏ったアルゴリズムが現在、データ経済の中で猛威を振るっており、社会に大きな害を与えている。
- **偏見に基づき、大規模に利用され、有害で、不透明なアルゴリズム**をオニールは「数学破壊兵器 Weapons of Math Destruction」と呼ぶ。
- Cf. 少し詳しいレビュー：https://note.com/minao_kukita/n/ne14be3e984e1

アルゴリズムとデータに潜むバイアス

- 愛知県のある保育園運営会社は保育士を採用するための面接で人工知能によって「笑顔度」を測定するというシステムを導入した。
- このシステムは面接中の応募者の顔を判定して、笑顔でいる時間の長さを計測する。
- そして面接者担当者の評価するコミュニケーション能力を50点満点、AIの測定する笑顔度を50点満点として最終的な評価を決定する。
- Cf. 角拓哉、「保育士に大切なのは… AIが採用面接で「笑顔度」測定」、『朝日新聞デジタル』、2020年9月7日。https://www.asahi.com/articles/ASN9762C1N8YOIPE00Y.html

アルゴリズムとデータに潜むバイアス

- ここには「**面接中に長い時間笑顔で居られる人間ほど良い保育士になる**」という前提がある。
- **この前提は客観的な事実ではなく、システムを作った人間の思い込みかもしれない。**
- さらに、笑顔度が人間の面接官の評価と同等の重要性を持つというのもシステムを作った人間が決めたことである。
- また「笑顔度」の学習に使われた**データにジェンダーや人種、年齢などの偏りがあれば、判定の正確さがこれらの属性よって異なる**可能性が高い。

アルゴリズムとデータに潜むバイアス

- HireVueという人事採用のための面接支援システムには、「AIアセスメント機能」と呼ばれるものがあり、「録画面接から声、話の内容、表情など、25000の特徴パターンを検出し」、「候補者の特徴を、その会社の優秀人材の特徴（教師データ）と比較し」、「教師データとのシンクロ率の高い候補者からランキングし」、「ポテンシャルの高い候補者から優先的に対面面接を実施」することを可能にするという。Cf. https://www.fujitsu.com/jp/group/fjj/solutions/enterprise-solutions/staff/hirevue/
- このシステムにはどのようなバイアスがあるかを考えてみよう。

アルゴリズムとデータに潜むバイアス

- アマゾンの採用アルゴリズムが女性に不利なバイアスを持っていた。Cf. Isobel Asher Hamilton、「アマゾンの採用AIツール、女性差別でシャットダウン」、『Business Insider』、2018年10月15日。https://www.businessinsider.jp/post-177193
- アメリカの裁判で再犯率を予想するために使われているCOMPASは黒人に対して不利なバイアスを持っていた。具体的に言うと再犯率が高いと判断されたにも関わらず再犯をしなかった人は、黒人の方が白人に比べて多かった。逆に再犯率が低いと判断されたにも関わらず再犯をした人は、黒人の方が白人に比べて少なかった。Cf. Matthias Spielkamp「機械は偏見を持つのか？犯罪者予測システムの是非を問う」、『MITテクノロジー・レビュー』、2017年6月22日。https://www.technologyreview.jp/s/44352/inspecting-algorithms-for-bias/

アルゴリズムとデータに潜むバイアス

- MITメディアラボの研究者ジョイ・ブォラムウィニらの研究によれば、IBM、マイクロソフトなどが提供する顔画像からの性別推定サービスでは黒人女性の誤認識率がおおよそ20%から30%であるのに対して、白人男性の誤認識率は1%にも満たなかった。
- Cf. 平和博、『悪のAI論――あなたはここまで支配されている』、朝日新書、朝日新聞出版、2019年、第二章。

定量的評価の幻想

- ジェリー・Z・ミュラー、『測りすぎ――なぜパフォーマンス評価は失敗するのか？』、松本裕訳、みすず書房、2019年。
- 現代はあらゆる分野において、定量的な評価が導入されてきている。本著はその歴史的経緯や社会的、思想的背景を明らかにすると同時に、その有害性を指摘する。
- 定量的な評価基準は、不正確な近似、実質の伴わない形式だけの業績、成果のごまかし、難しい仕事の回避などをひき起こす。
- 研究教育、警察、医療、軍事などに導入された定量的評価は悲惨な失敗を引き起こしているにも関わらず、定量的評価への社会の執着はなくならない。

監視資本主義

- ショシャナ・ズボフ、『監視資本主義：人類の未来を賭けた闘い』、野中香方子訳、東洋経済新報社、2021年。原著はShoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Profile Books, 2019.
- 人工知能の発展に伴い、ユーザーのデータを収集して行動予測あるいは行動誘導を正確に行なうことが可能になる。
- ビッグデータを握る企業が大きな力を持つ。
- 巨大IT企業はますます貪欲にデータを収集するようになる。
- こうして**人々の行動データを収集することを基盤とする新しい資本主義システム、「監視資本主義」が確立する**。

ビッグデータと人工知能

ここでの人工知能

- 「人工知能」という言葉は多種多様なテクノロジーの曖昧な総称である。
- ここでは現在最も活用され、かつ最も大きな富を生んでいるであろう、**ビッグデータに基づく機械学習システム**について論じる。
- 以下では「人工知能」はこのようなシステムを指すものとする。
- また特に人間の行動データに基づいて人間の属性や行動傾向を推測するアプリケーションに焦点を当てて論じる。

人工知能というメディア

- メディアとは情報を伝える媒体であり、私たちが直接に経験していない現象や対象について、何らかの認識を持つことを可能にする手段である。
- そしてこの意味において、人工知能は文字通りメディアである。

「メディアはメッセージ」

- マーシャル・マクルーハンの言葉。
- 伝達される情報の形式や様態（モダリティ）、あるいは私たちがその情報を消費する仕方は各々のメディアの持つ技術的特性によって規定される。
- 同じ内容のメッセージでも、**どのメディアによって伝えられるかによって、受け取った人間に与える効果は異なる**。
- **どのメディアを使うかということは、メッセージの内容以上に重要な意味を持つ**。

「メディアはメッセージ」

- 例えば新聞ならば伝えられる情報は文字だけ、ラジオならば音声だけである。
- 新聞は私たちが情報を消費する間、そこに注意を集中することを要求するのに対して、ラジオは私たちが他の仕事をしながら（運転しながら、本を読みながら、食事をしながら、ジョギングをしながら）聞き流すような消費の仕方を可能にする。
- また新聞の情報は保存がきくのに対してラジオはそうではない、音声は文字に比べて感情を喚起しやすいなどなどの違いがある。

「メディアはメッセージ」

- ラジオの時代になって政治家の演説の長さは1時間から10分程度になった。
- テレビの普及は見栄えの良い政治家に有利に働く。
- カナダでの国政選挙において立候補者の得票数と外見的な魅力の関係を調査した研究では、魅力的な立候補者はそうでない立候補者よりも多くの票を得ていた（得票率は平均して32%と11%だった）ことが分かった。Michael G. Efrain and E. W. J. Patterson, (1974)``Voters vote beautiful: The effect of physical appearance on a national election'', ¥textit{Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement}, 6(4), 1974, 352–356. ¥url{https://doi.org/10.1037/h0081881}

メディアとしての人工知能の特性

- 伝統的なメディアが基本的に事実を伝えるのに対して、人工知能は一般に**データに基づいた確率的な推測を伝える**。
- もちろん新聞やニュースも推測を伝えるが、それは人間が行った推測を伝えているだけである。
- それに対して人工知能は機械が行なった推測を伝える。
- しかもそれは**個々の私人（あるいは私人の集団）の行動や性格や能力や思考についての確率的な推測**である。

人工知能を使う動機

- 人工知能の開発者たち、あるいは利用者たちはなぜこういったことに関心を持つのか。
- 純粋な好奇心によって動機づけられている場合もあるかもしれないが、多くの場合、理由は人工知能が与えてくれる情報が彼らの利益と損害に直結しているからである。
- 人々が人工知能を使う理由は**人工知能が与えてくれる情報が彼らの利益と損害に直結している**からである。
- ある商品の需要を知ることは生産者・販売者にとって極めて有益である。
- 同様に、雇用者にとって有能な社員になりそうな人間を知ること、警察にとって潜在的な犯罪者を特定すること、ローン会社にとって誰が破産しそうかを知ること、保険会社にとって病気に罹りそうな人間を知ることは、極めて有益である。

意思決定とリスク分析

- 意思決定を行う際に、ありうる選択肢のそれぞれに伴う潜在的な利益と損害を見積もることを、工学や経済学の分野では「確率論的リスク分析」あるいは単に「リスク分析」と呼ぶ。
- リスク分析においては、ある選択肢をとったときに、どのような事象がどれくらいの確率で生起するかを見積もることが必要になる。
- しかし従来は人間や社会のような複雑なシステムについて、特定の振る舞いの生起確率を正確に予想することは難しかった。
- それゆえに人間に関するリスク分析はしばしば多くの仮定に基づいた不正確なものにならざるを得なかった。

人工知能が伝える情報

- しかしビッグデータと人工知能は人間の振る舞いについての従来よりはるかに正確な確率的予測を可能にした。
- 人工知能は、**人間や社会を対象にした確率論的リスク分析のための革新的なツール**である。
- 要するに人工知能は、**人間を様々なデバイスから取得された機械可読なデータの集積、そしてそこから推測される種々の属性の束として扱い、特定の利益関心に沿ったリスク分析を行った結果として、「この人はこれだけの利益／損害をもたらす見込みがある」という情報を伝えるメディア**なのである。

人工知能の長期的な意図しない影響

- 人工知能が様々な局面で応用されるにつれて、**人間をリスクとして扱い、利益／損害という観点から評価する慣行、そして不利益をもたらすと判断された個人を排除する風潮が広がるだろう**。
- そのような判断は間違っていることもある。それは人工知能の判断が確率的であることからの不可避の帰結である。
- しかし例えば一部の人たちを誤って切り捨てたとしても、全体として利益が向上するならば、人々は人工知能の判断を採用するだろう。

リスク分析の本質

- 政策決定者や企業がそのような判断をすることはある程度仕方がない。政府や企業は全体的な利益の最大化を第一の目的とするものだからである。
- リスク分析は基本的に政策決定者や経営者が**大局的な観点から効率的なマネジメントを行うための道具**である。
- 特に企業においては、重要なのは金銭的な収支であり、全体として利益が上がるならば、切り捨てられる個人は顧みられない。

人工知能のメッセージ

- 一言でまとめると、メディアとしての人工知能が持つメッセージは、次のようなものである。
- 第一に、**人間は様々なデバイス、アプリケーションを通じて取得されるデータと、そこから確率的に推測される属性の集まりとして理解できる**。
- 第二に、**他者はあなたにとってリスクであり、そのリスクは前もって見積り、回避することができる**。
- 人工知能が社会に浸透するということは、このようなメッセージに私たちが知らず知らず曝され続けるということである。
- そのことが**人々の人間観や、成立しうる人間関係に与える影響について、注意しておかなければならない**。

データ規制、AI規制

データ保護法の核となる原則

- 個人データは合法的に、公正に、かつ透明な仕方で処理されなければならない。 (合法性、公正性、透明性)
- そのようなデータは前もって特定された目的のためにのみ収集されることが許され、それとは関係ない他の目的のためには利用されてはならない。（目的の制限）
- そのようなデータは、処理の目的に必要なものに限られるべきである。（データの最小化）
- そのようなデータは十分に正確で最新のものであるべきである。（正確性）
- そのようなデータは、理由なく長期間保持されるべきではない。（保存の制限）
- そのようなデータは侵害、不法な使用などから守られるべきである。（完全性と秘匿性）
- データ管理者はコンプライアンスに責任を持つ。（アカウンタビリティ）

Frederik Zuiderveen Borgesius, ``Discrimination, artificial intelligence, and algorithmic decision-making'' より

データ保護規制の例：GDPR

- EU一般データ保護規制（General Data Protection Regulation）
- 2018年から施行されている。
- EU域内の企業などに適用されるが、EU域外の企業でも、EU域内の個人の情報を扱う場合には適用される。
- 従来のデータ保護法の原則を継承しつつ、人工知能によるデータの処理などについての規制を盛り込む。
- 重大な違反に対して高額の制裁金（上限は2000万ユーロと全世界年間売上高の最大4%までのいずれか大きい方）

データ保護規制の例：GDPR

- EU加盟国は独立した権限（調査権など）を持つデータ保護機関（Data Protection Authority）を持たなければならない。
- GDPRでは、データ保護機関は例えばデータ取扱者の施設に立ち入り、データ保護監査という形での調査を遂行し、データ取扱者に情報提供およびデータ処理システムへのアクセス権を渡すよう命じる権限を持つことが定められる。

データ保護規制の例：GDPR

- GDPRは特定の自動意思決定についての特別のルールを持つ。
- GDPR 22条には、法的影響を持つ、あるいはそれと同等の重大な影響を持つ、完全に自動化された決定の原則的な禁止が含まれる。
- The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her. (GDPR Art. 22)

データ保護規制の例：GDPR

- GDPRには一般的な透明性についての決まりのほかに、自動決定の透明性についての決まりがある。
- 「取扱者は以下の情報をデータ主体に提供しなければならない。[…]プロファイリングを含めた自動的意思決定が存在すること、[…]および、そのような場合には少なくともデータ主体にとっての重要性と、そのような処理の予見される帰結に加えて、そこに働いているロジックについての意味のある情報。」

データ保護規制の例：GDPR

- GDPRは非常に一般的な原則であるため、そしてまたテクノロジーの進歩が非常に急激であるために、個別の応用領域にはそれぞれに適した細かい規制、その領域の伝統的な法や価値に合致した規制が必要になる。

倫理的なデータ利用のために重要なこと

- 何のために、どのようなデータが収集され、それがどのように処理されているのかを明らかにする。
- データの利用が、社会的価値、あるいは特定の応用領域において重視されている価値と整合的な仕方で行われているかを配慮する。例えば教育で言えば、特定のテストで良い点を取らせることが良い教師であることを意味するのか。
- 公正な仕方でデータを利用する。例えばそれが社会的弱者に対してますます不利な状況を生むようなことにならないか。
- データ利用の長期的な負の効果を考える。

AI倫理原則

- 2010年代初頭から、AIに関する倫理についての議論が活発に行なわれ原則や倫理指針を公表。
- 学術界、企業、政府などのセクターでAIに関する倫理原則、倫理指針が発表される。
- 例：アシロマAI原則、人工知能学会倫理指針、IEEEのEthically Aligned Design、EU AI倫理原則、日本政府（内閣府）の人間中心のAI社会原則、自律型兵器についての特定通常兵器使用禁止制限条約での議論など。

一般にAI倫理で懸念されていること

- 透明性：AIの判断プロセスは検証できるのか？ AIの判断について説明責任は果たせるのか？
- 責任：AIの判断の結果については設計者・製造者が責任をとれるのか？
- 制御可能性：AIが人間によるコントロールを逸脱しないか？
- 公平性：AIの利用が特定の人々に不利益を与えないか？
- プライバシー：個人データが適切に扱われるか？
- 人道：AIが非倫理的な仕方，人間の尊厳を傷つける仕方で使われないか？

AI規制の動き

- 2021年には欧州委員会が人工知能に対する規制案「Proposal for regulation of the European Parliament and of the Council: Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts」を発表している。

EUのAI法案の特色

- 「リスクベースのアプローチ」
- EUにおける既存の規制との調和
- future-proof（将来の問題に対処できるような設計になっていること）
- EUにおける統一的な市場の確保
- 投資とイノベーションの促進
- これらを見ると、かなり**ビジネス的な観点寄り（すなわちコストとベネフィットの両方への目配りが利いている）という印象**を受ける。

EUのAI法案で考えられているAIのリスク

- 許容できないリスク：
  - **EUの重視する価値**に反するもの。
  - 人々の**心身に危害**を与える可能性が高いような仕方で、人々の**振る舞いを実質的に変化させる**ことを目的に、**意識されないように**人を操作する、あるいは子どもや障害者のような特定の**脆弱な集団**の**脆弱性**を利用する可能性があるような使用。
  - 公的機関による**汎用の社会的スコアリング**。
  - 法執行を目的とした、公共の場所でのリアルタイムの**遠隔生体認証**。

EUのAI法案で考えられているAIのリスク

- 高リスク
  - 自然人の**健康と安全**、あるいは**基本的な権利**に高いリスクを生じさせるもの。
  - こういったシステムは特定の必須要件を順守し、事前の適合性評価を受けているならば、欧州の市場に出すことが許される。
  - 高リスクのAIについてはリスク管理システムが確立、実装、文書化、維持されていなければならない。

EUのAI法案への批判

- この規制法案では、**現実のAIシステムがどの程度のリスクであるかを判定するのは難しい。**
- 実際、例えば日本の経団連デジタルエコノミー推進委員会AI活用戦略タスクフォースはこの規制法案に対する意見を表明しており、「禁止・ハイリスク AI の定義等に**曖昧さや解釈の余地**」があるとして批判している。

解釈の余地が大いにありそうなところ

- 「EUが重視する価値」とは？
  - 人間中心、人権、生命、フェアネスなどなど。
- 「意識されない」とは？
- 「健康」とは？
- 「脆弱な集団」、「脆弱性」とは？

ジョアンナ・ブライソンのツイート

- 「Jiboは終わったかも知れないけど、@STurkle のこの記事は重要性を増し続けている。EUの ＃AIRegulation で禁止される #AI のトップカテゴリーの一つが、ここで彼女が記述している危害に特に関係している。つまりサブリミナルな行動変容だ。＃AIEthics」

Jibo

https://twitter.com/j2bryson/status/1399570323285151744

シェリー・タークルの記事の主張

- ソーシャル・ロボットは子どもの心理の脆弱性を利用して、感情のない機械に「共感」し、「愛着」を持つように仕向けている。
- ソーシャル・ロボットは子どもが社会関係を結ぶことを学んだり訓練したりする機会を奪う。

https://www.washingtonpost.com/outlook/why-these-friendly-robots-cant-be-good-friends-to-our-kids/2017/12/07/bce1eaea-d54f-11e7-b62d-d9345ced896d_story.html

何が「高リスク」か？

- 例えばソーシャル・ロボットのようなものについて、事前にそのリスクを見積もるのは難しい。
- 脆弱な集団を食い物にしないという意識は重要で、その上で予測できない影響に対して適応的に対処する姿勢や、危険そうだと思ったら予防的に対策する柔軟性が必要。
- Future-proofingと具体性はトレードオフ。
- 規則に対するコンプライアンスは組織として取り組むべきだが、それ以上のことについては、専門家としての感覚や倫理性が有益だろう。
- **倫理というは現在の基準を疑問視することを本質的な部分で含んでおり、Future-proofing指向とは相容れない。**

「AI権利章典」

- アメリカのバイデン政権は2022年10月4日、「**AI権利章典**」を発表。
- 政府高官が語るところによれば、**AIはプライバシーに対する基本的権利、差別からの解放、基本的な尊厳といった私たちの中核となる民主主義の価値観とは相容れない危害を及ぼしている**、と。
- Cf. Melissa Heikkilä、「米バイデン政権が「AI権利章典」を発表、テック企業に説明責任」、『MITテクノロジーレビュー』、2022年10月7日。

「AI権利章典」の5つの保護原則

- 安全ではない、あるいは効果のないシステムから国民は守られるべきだ。
- 国民はアルゴリズムによる差別にさらされるべきではない。
- 国民は不正なデータ利用慣行から守られ、データ管理に対する主体性を与えられるべきだ。
- 国民は自動化されたシステムが使われていること知り、そしてそれが出力結果にどう貢献しているのかを理解するべきだ。
- 人々はAIシステムを拒否して人間による代替手段を求めること、問題に関する救済策を利用することができなければならない。