

AI を使った設計に関する倫理規定の調査

Investigation of Ethics Code regarding Design using AI

○戸水晴夫（正, SDI Japan, Haruo Tomizu）

1 はじめに

近年、世界における AI 技術の進展にはめざましいものがある。2018 年度版ものづくり白書¹⁾によれば、今後の日本が目指すべき産業の姿として、Connected Industries のコンセプトが掲げられている。これは産業界の様々なつながりによって、新たな付加価値を生み出そうというものである。その主要な施策の一番目が「人と機械・システムが対立するのではなく、協調する新しいデジタル社会の実現」に置かれている。さらに AI やロボットを活用することによって、人を助け人の力を引き出す産業社会の構築を目指そうというものである。一方で、日本の産業界における AI の普及の状況はどうであろうか。図 1 は 2018 年 7 月から 11 月の期間に矢野経済研究所が実施した AI の活用状況に関するアンケート結果を示している。²⁾ 回答があった民間事業所は 515 社あった。現時点で AI を既に導入していると回答した企業は、ただか 2.5%にすぎない。実証実験中と検討中の企業を併せても 19.8%だった。一方で、AI に関心がなく、今後も取り組む予定がないと回答した企業は 15%だった。このことから、AI の活用はこれからであり、将来的に多くの設計者が AI を組み込んだ機器や製品を設計することによって普及が始まるものと想像できる。そして、AI が人と関わる製品や人と協調する社会システムに実装されるならば、そこにはルールや規範といった規制が必要になると考えられた。

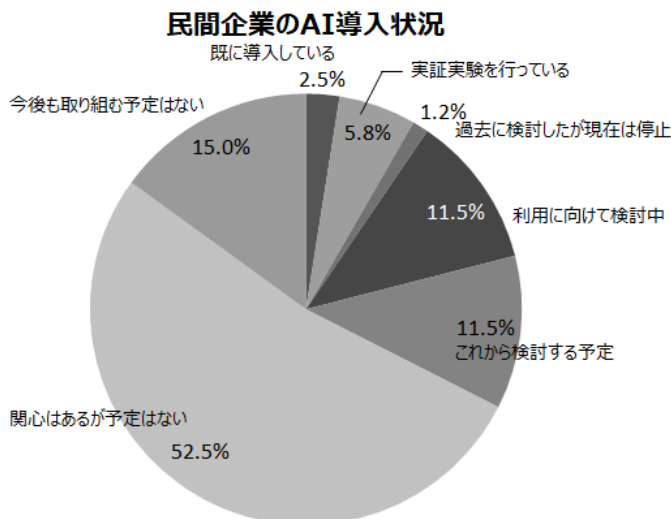


図 1 AI の活用状況に関するアンケート結果

公益社団法人 日本設計工学会九州支部, 日本図学会九州支部
平成 31 年度 研究発表講演会において講演 (2019 年 6 月 8 日)

2 AI を使った設計の倫理上の問題

日本設計工学会九州支部における研究調査活動のひとつに「Design For AI に関する調査研究分科会」がある。2017 年 10 月に創設されたこの研究会を進める中で、私たちが心配になった点があった。それは、このまま無制限に AI を製品に組み込む設計を続けていて良いのかということである。将来的に優れた AI を使った設計を奨励することだけではなく、もっと人々の生活を良い方向に導く AI を使った設計が必要と考えたからである。なぜなら優れた AI は必ずしも良い AI たり得ず、悪用や誤用されれば、悪い AI になってしまう可能性があるからである。

2.1 既に想定されている AI 活用時のリスク

欧米では「AI が人種や性別などに偏った分析や判断を拡大させる危険がある」との指摘が出ている。³⁾ その理由として、AI のプロセスでは、人間が読み込ませた過去のデータから独自の判断基準などを学習し、次第に自動でデータ収集を始め、分析結果を出すからである。もし、もとのデータに差別的な偏りがあれば、それを助長しかねない弱点があるといえる。

以下に、既に想定されているリスクの例を挙げる。

人種や性別などに偏った分析や判断をするリスクとして、学習のプロセス中のデータに差別的な偏りがあると、次のような AI が生まれるかもしれない。

- (1) 医者には男性が向いていると合格点数を勝手に変更しだす採点の AI。
- (2) 外国籍の人の住宅ローンや保険審査を厳しくする審査の AI。
- (3) 障害者や弱者に対する言葉を選ばない音声対応の AI。
- (4) 遺伝情報から将来を勝手に予測し結婚相手に差をつける AI。

データ分析がブラックボックスであり、理由が示されないリスクの例としては、

- (5) 内部のデータ分析の過程が複雑すぎて理解できない AI。
- (6) 判断過程や指標が外から見えないため、不具合が発覚しない AI。
- (7) 他の AI との不適切な連携があっても、何をやっているのか見えない AI。

不具合が見つかって、技術的に迅速な修正が難しいリスクの例としては、

- (8) 学習による修正に時間を要するため、迅速な対応ができない AI。

以上のように、AI に関して、様々なリスクによる脅威の発生が考えられた。

2.2 AI を設計で活用するときのパラドックス

それでは、設計という行為に関して、どのような設計を行った場合に問題があると「Design For AI に関する調査研究分科会」の活動の中で考えられたのだろうか。図 2 の例で説明する。

例えば、ロボットに作業をさせるために、人がティーチングを行う場合や、人が選んだ教師データをもとに AI に学習させる場合は、問題がないと考える。なぜならば、この場合は人が主で機械が従と考えられるからである。たとえ AI が熟練技能者の代わりに作業を行い、意思決定する場合でも、AI に認識させるデータは限られたものと考えられるからである。

一方、ロボットとは限らず、タブレット端末のようなインタフェースを介して、AI が人にティーチングを行う場合が想定できる。育児や幼児教育に始まり、小学校から中学校の補助教材として、このようなタブレット端末の利用が検討されている現状がある。しかし、教育教材というものは限られた教師データのみでは不足であり、様々な質問の答えとなる最新データを求めて AI はインターネット内を探索するようになると考えられる。より高度な AI になればなるほどクラウド上で学習を重ねて利口になることができる。明らかに機械が主で人が従となるような場合は、人と機械の立場が逆転する。特に、人を導く場合には、AI を組み込んだ設計を起点として様々な問題が起きる可能性があると考えられた。

2.3 AI の高度化で生まれるリスク

AI が高度化するにつれ、問題が起きやすくなるということを説明するために、AI のアーキテクチャを考えてみる。AI の構成アーキテクチャは基本的に、ハードウェア、OS、ランタイム、アプリ、関数の 5 層で構成されている。全ての要素をひとつの製品内に持つオンプレミスの形態以外は、クラウドを利用することが一般的である。当初はハードウェアだけをクラウドから提供する IaaS (Infrastructure as a Service) ホスティングサービスのようなものだった。最近ではハードウェアや OS、ランタイムのプラットフォーム一式をインターネット上のサービスとして

提供する PaaS (Platform as a Service) といわれる仮想サーバサービスが多くなった。さらに、SaaS (Software as a Service) という、これまでパッケージ製品として提供されていたソフトウェアまでを、インターネット経由で提供するサービスまで行われるようになった。結局、クラウドの利用を拡大することによって、最小限のリソースで膨大なデータ量を処理できるようになるため、タブレット端末やスマートフォンなどで稼働する AI 機器が実現できるようになったのである。結果的に、AI が最も効率よく稼働できる反面、インターネットから受けるリスクが大きくなっていくと考えられた。

2.4 研究分科会で問題とした AI のリスク

AI を使った設計に関わる倫理が必須となる場合は、より生活の場に近く、人間と対話し、人々の支えとなる製品を設計する場合にあることが徐々に分かってきた。特に、AI が人とインタラクティブである場合に限って、顕在化してくる課題があることも分かってきた。もちろん、AI 自体は機能実現の手段であり、設計目的ではないため、組み込み対象とする分野は広範囲に渡るものである。しかし、人とインタラクティブでない AI に留まるならば問題が顕在化することは少ないと考えられた。

例えば、AI が人間と対話するような製品を設計することを考えてみよう。最も簡単な方法は世の中に在るスマートスピーカーを製品に実装することである。はじめにスマートスピーカーの自然言語処理の仕組みについて図 3 を例として考えてみる。ユーザが Google ホームに「今日の天気は？」と聞いたとする。その時、内蔵されている Google アシスタントが、音声を変換し、意図を理解しようとする。クラウド上において Dialogflow などのテキスト変換のアプリを介して、Microsoft Azure Function などを使って答えを得る処理が行われる。いまはお天気を知りたいので、インターネット上のウェザー情報を検索し、得た結果を JSON フォーマットで返す。それを再び、自然言語変換で話し言葉にして Google アシスタントに送り返して、スピーカーで音声にするという仕組みである。従って、AI がなんと答えるかは Function 処理の結果次第ということになる。

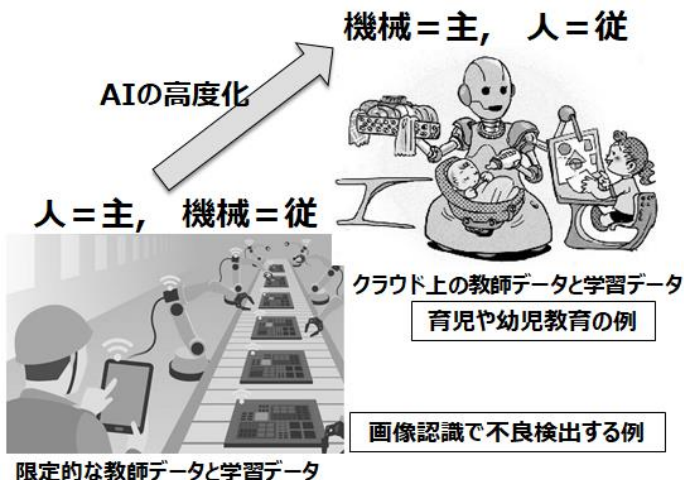


図 2 AI を設計で活用するときのパラドックス

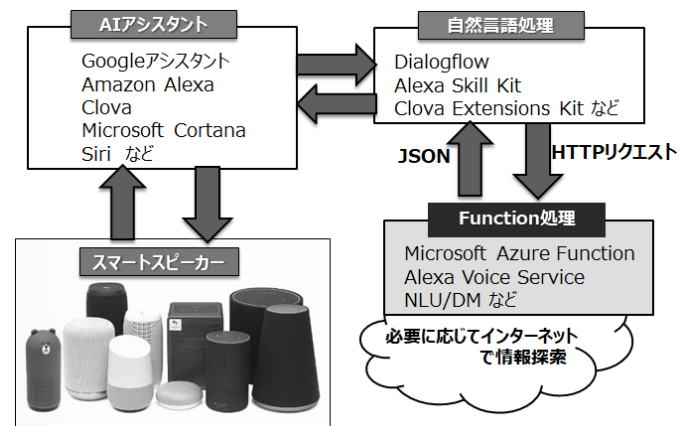


図 3 AI が人間と対話する仕組み

次にインターネット上で動作する AI の問題として挙げられたのは、画像処理の問題だった。個人のプライバシー侵害と個人の尊厳の遵守の問題である。例えば、Google 検索のインデックスは、何十億ものページを保存し、取り込んだデータのキーワードとページを突き合わせてそれが何かということを確認してランキングして答える仕組みである。AI が勝手にクラウド上で学習する場合、検索結果においてプライバシー侵害の問題は回避できない。さらに、AI による画像認識では、画像認識の結果そのものが、個人の尊厳を守り切れないという課題を抱えている。例えば、画像認識の AI にパラスポーツの画像を認識させても正しく認識されないケースが出てくることが多い。人は個人のアイデンティティに対して社会から間違った認識を与えられた場合に、傷つくことがある。技術的には、学習が足りないために起こる現象であり、さらに学習させればよいということかもしれない。しかし、AI の学習をどこまでやればよいのか、市場に商品として出して良いのかという基準が存在しない。あいまいな状態で、規制もないままに AI を製品に組み込んで、出荷後にインターネット上で学習させるということだけでは、個人の尊厳を守ることは難しいと考えられた。

3 国内外における AI 倫理規定制定の動向

AI の倫理問題については、2015 年頃から国内外の様々な行政組織が検討のための会議体を設けて、AI 開発の原則や新しい法律の制定、契約形態などを審議し、提言や倫理指針などを公表してきた。⁴⁾ 以下に、国内の主な取り組みと海外で最も先進的な EU での取り組みを紹介する。

3.1 日本国内における倫理指針の制定

日本における AI に関する倫理指針は、人工知能学会倫理委員会が作成し、2017 年 2 月 28 日に公表した。⁵⁾ 序文で「人工知能の研究、設計、開発、運用、教育に広く携わる人工知能研究者は、人工知能が人間社会にとって有益なものとなるようにするために最大限の努力をし、自らの良心と良識に従って倫理的に行動しなければならない」と宣言している。内容は以下の 9 項目が基本となっている。

- (1) 人類への貢献
- (2) 法規制の遵守
- (3) 他者のプライバシーの尊重
- (4) 公正性
- (5) 安全性
- (6) 誠実な振る舞い
- (7) 社会に対する責任
- (8) 社会との対話と自己研鑽
- (9) 人工知能への倫理遵守の要請

しかるに、日本国内の AI の技術や開発方針が、いまだ倫理的に十分とはいえないように見える。例えば、AI 自動運転の車が重大事故を起こした場合の責任は、だれに在るのかということさえ明確にされていない現状である。

3.2 EU の「良い AI 社会」

AI を使う場合の想定リスクに対応するため、EU では 2017 年 11 月に AI4People というグローバルフォーラムを 3 年間のロードマップで、立ち上げている。AI4People の目標は、良い AI 社会を築くための基礎となる原則や方針を策定し、実践的フレームワークを広めることにあった。当初、掲げられた目標は、次の 3 点だった。

- (1) AI の望ましい発展を支える倫理的フレームワークを考え、その役割を定義すること。
- (2) ヨーロッパが「良い AI 社会」となるために必要な倫理的フレームワークを考案すること。
- (3) フレームワークの実行が成功するために、実施可能な措置を提言すること。

つまり、AI4People は、AI に関する新しい規則を定義することを目的としているのではなく、より良い AI 社会の推進を目的とする社会的仕組みの方針や倫理規定を提案することにあつた。AI4People は、欧州委員会、欧州議会、市民会議、産業界、メディアなど、新しい AI のアプリケーションによる影響を調査することに関心を持つすべての関係者を集めたフォーラムとなった。そして、AI の社会的影響に関する調査を始め、2018 年 10 月にブリュッセルで開催された欧州議会の人工知能会議で倫理指針となる 5 原則と 20 の提言を発表したのである。⁶⁾

3.2.1 EU の倫理指針 5 原則

AI4People が定めた倫理指針の 5 原則は、次の通りである。

- (1) 恩恵：人々の幸福の促進と尊厳の維持、持続可能な地球環境を維持すること。
- (2) 悪意のない行為：プライバシー、セキュリティだけでなく意図しない事故を保全すること。
- (3) 自律性：AI により意思決定が阻害される場合は、人間の意志を優先すること。
- (4) 正義：あらゆる差別を撤廃し、人類の繁栄の促進と連帯を維持すること。
- (5) 説明能力：了解獲得と説明責任を通して、全ての原則の実行を説明できること。

AI4People は AI の倫理的な意味を考えた最初のグループではない。すでに多くの組織が、社会における AI の開発と発展を導くための価値観や倫理原則について発表している。AI4People では、同じような議論を行うのではなく、そのポリシーやベストプラクティス、および新しい戦略の具体的な推奨事項を作り上げることを着地点としていた。特に、生命倫理学で扱われた上記の(1)から(4)の原則は、AI の倫理的課題に対応できる原則に最も近いものと考えられた。但し、それらは AI の課題すべてを網羅するものではなく、さらにもう一つの新しい原則を追加することが必要になった。それが、説明可能であることだった。相手が了解していることと説明責任の両方を組み込んだ 5 番目の原則のことである。

3.2.2 EU の「良い AI 社会」のための 20 の提言

AI4People は倫理指針の 5 原則をもとに、「良い AI 社会」を目指すための 20 の具体的提言を行った。その中で、AI は人間の自主性や自律性を尊重して、

不平等や不公平を減らし、公平に共有される社会的利益を増やすように設計および開発されるべきと訴えている、また、AI を提供するものは説明責任を負い、AI 技術に対する信頼と理解を構築することが重要としている。良い AI 社会を構築するには、開発者だけでなく複数の利害関係者による共同作業によるアプローチが必要になってくるとも述べている。以下に挙げた 20 の提言は、4 つの分野に分かれている。

- (1) アセスメントに関する事項
 - ① AI によって生じた過ちや実害について、既存の司法機関による対応の可能性を評価する。
 - ② 既存の法律から考えて、どのような意思決定機能を AI システムに任すべきかを評価する。
 - ③ 技術の進歩に対応できる立法の枠組みが、倫理に基づいているかを評価する。
- (2) 開発に関する事項
 - ④ AI の判断過程をわかりやすく説明する説明責任を開発企業や団体に課す。
 - ⑤ 裁判所によるアルゴリズムの精査を可能とする目的で司法制度の IT 基盤を改善する。
 - ⑥ 不公正な偏見などによる結果を識別するための AI システムの監査機関を設ける。
 - ⑦ 情報公開のリクエストを受理するためのクレーム登録のガイド付きプロセスを作成する。また、AI の欠陥による損害をカバーする賠償責任保険制度を創設する。
 - ⑧ AI 製品やサービスの信頼性を測るベンチマークを可能とする測定基準を作成する。
 - ⑨ AI 「リリース後」のモニタリングシステムを開発し、報告義務と報告システムを設ける。
 - ⑩ AI 製品やサービスの評価と監督を通じて、公共の福祉に責任を負う監督機関を設立する。
 - ⑪ ヒューマン・マシンの協調基盤を築く、法的文書および契約書のテンプレートを開発する。
- (3) 行動の動機付けに関する事項
 - ⑫ EU レベルで、社会的に好ましく、環境に優しい AI 技術の開発と使用を奨励する。
 - ⑬ 利益のみを重視することなく、社会的利益のため AI 開発を前進させる研究に資金援助する。
 - ⑭ 性別、階級、民族、分野、およびその他の多様性側面の観点から、AI の設計と開発に注力する。
 - ⑮ AI 研究プロジェクトに倫理的、法的および社会的配慮を含める場合に経済的援助を行う。
 - ⑯ AI システムの実証的なテストと開発のために、EU 内で規制緩和された特別区域を設置する。
 - ⑰ 一般市民の AI およびその応用に関する理解を得る方針を協議する組織を設置する。
- (4) サポートに関する事項
 - ⑱ 倫理的 AI であることを保証する「倫理的 AI」認証証明書をプロバイダから発行する。
 - ⑲ AI システムを開発または使用している企業は、企業倫理審査委員会の審査を受ける。
 - ⑳ 学校のカリキュラムで、教えるべき基本分野にコンピュータサイエンスを含める。AI と一緒に働くことの社会的、法的、および倫理的影響について教育する。

4 おわりに

国内外における AI 倫理指針制定の動向調査から、日本国内でも数々の倫理指針や提言が行われていることが分かった。基本は人工知能学会が定めた倫理指針となる。しかしながら、AI を使う主体者は、AI そのものを作り出した人間ではない。製品の使用目的や価値を決めるのは利用者である。どんなに良い AI といわれようと、使い方を誤れば悪い AI になることができるのである。そして、製品の設計者は AI の開発者ではなく、できたものを使う人である。今では、AI のモジュールなどが容易に手に入るようになった。それを設計者が安易に製品に組み込み、その後、市場で事故が発生した場合はどうなるのだろうか。発生した社会的損害に対して、その製品の設計者や開発企業が責任を問われないということはないだろう。そして、現状では、そのような設計者が守るべき倫理の基準がないことが問題だと考える。いままさに、多くの研究者や設計者の判断の基準となる設計業務に特化した分かりやすく具体的な倫理規定が必要である。この点において、日本設計工学会において倫理委員会を立ち上げて AI を設計で活用するための倫理指針や設計者に対する倫理規定を策定することに大きな意味があると考えられる。

参考文献

- 1) 2018 年度版ものづくり白書：経済産業省、厚生労働省、文部科学省、(2018)、129
- 2) リテール市場における AI 活用の事例研究 2018-2019：株式会社矢野経済研究所、(2018)
- 3) Asilomar AI Principles : Principles developed in conjunction with the Asilomar conference, (2017)
<https://futureoflife.org/ai-principles>
- 4) 国際的な議論等諸外国の動向：内閣府 人間中心の AI 社会原則検討会議 第 3 回配布資料、(2018)
<https://www8.cao.go.jp/cstp/tyousakai/humanai/index.html>
- 5) 人工知能学会倫理指針：一般社団法人 人工知能学会倫理委員会、(2017)
http://ai-elsi.org/report/ethical_guidelines
- 6) AI4People's Ethical Framework for a Good AI Society : Atomium - European Institute for Science, Media and Democracy, (2018)
<http://www.eismd.eu/about/>