

2017年5月22日

The IEEE Global AI Ethics Japan Committee

Workshop Organizer: Arisa Ema / Katsue Nagakura

IEEE Workshop in Japan 参加者からのコメント

Workshop 背景

“Ethically Aligned Design, Version 1 (倫理的に配慮されたデザイン)”は、「人工知能 (Artificial Intelligence; AI) および自律システム (Autonomous Systems: AS) の倫理的配慮に関する IEEE の国際指針 (IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems)」が作成しました。このドキュメントは、人工知能、法、倫理、哲学など他領域にわたる研究者、政策関係者、企業関係者の意見を集約して作成されました。かかわった人は全世界で 150 人以上に上ります。

このドキュメントの目的は2つあります。ひとつは、これからの人工知能や自律システムが「倫理的に配慮された設計」となるための知見を提供することです。そのために、様々な課題を挙げられています。もうひとつは、その設計に基づいた IEEE P7000™ (システム設計における倫理的懸念に対処するモデルプロセス) という IEEE による標準化を行うことです。そのほか、さらに6つの標準化についての議論がなされ、人工知能や自律システムの倫理に関する実効性の在り方を検討しています。

Workshop 趣旨

日本は人工知能、自律システム、ロボティクスと倫理において、重要な役割を果たすと IEEE Global Initiative は考えています。そのため、多くの日本人が“Ethically Aligned Design, Version 2”を作るプロセスに加わることが期待されています。

日本でも、2014年より人工知能学会倫理委員会をはじめとしたアカデミア、省庁や企業においても「人工知能と社会」に関する議論が行われてきていました。それらの議論は、このドキュメントで議論されていることと重なる視点も少なくありません。その一方で IEEE が目的としている AI 倫理、AI が浸透した社会のルールや標準を作るにあたっては、多様な価値観に耳を傾けると同時に日本独自の視点で意見を述べるのが重要になります。

そこで、このドキュメントのバージョン更新のための議論が行われる The IEEE Global Initiative の会合に提供する意見を取りまとめることを目的に、名古屋、京都及び東京でワークショップを開催しました。ワークショップなどを通して様々な視点からの意見を取りまとめた上で、2017年6月5日—6日にオースティンのテキサス大学で行われる会合に参加し、日本からの視点を提供します。

Workshop 参加者

■名古屋 Workshop

日時:2017年4月21日 16:00-18:00

参加者数:8名

■京都 Workshop

日時:2017年4月28日 16:00-18:30

参加者数:11名

■東京 Workshop

日時:2017年5月2日~3日

参加者数:

2017/05/02 13:00-15:00 19名

2017/05/02 15:00-17:00 8名

2017/05/03 10:00-12:00 8名

2017/05/03 13:00-15:00 5名

(※ 重複参加あり)

総参加者数:44名

Special Thanks

- ❖ Konstantinos Karachalios, Managing Director of The IEEE Standards Association and Member of the IEEE Management Council
- ❖ Iwao Hyakutake, IEEE Director, Japan Office
- ❖ Raja Chatila, Chair, The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems
- ❖ Kay Firth-Butterfield, Vice-Chair, The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems
- ❖ John C. Havens, Executive Director, The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems
- ❖ Danit Gal, Director, Outreach Committee of The IEEE Global Initiative
- ❖ Ryota Kanai, Founder & CEO of ARAYA
- ❖ Yutaka Matsuo, Chair of the Ethical Committee of Japanese Society for Artificial Intelligence (JSAI)
- ❖ Hiroko Kamide, Chair of Special Interest Group on "Philosophy and Practice for Robotics", The Robotics Society of Japan.

本稿について

- 本稿は多様な意見をまとめるのではなく、各参加者によるコメントを列挙する方法をとり、様々な視点からのコメントを紹介することを目的としています
- 本稿執筆にあたっては、まず各ワークショップでの発言を書き起こし、“Ethically Aligned Design, Version 1”全体やその背景に対する意見(総論)と、8項目の各論に意見を集約しました。それを2017年5月6日から15日までワークショップ参加者全員に公開し、修正や加筆をいただきました。その後、オーガナイザー側で整文しました
- 内容は箇条書きとしました。インデントが下がっているコメントは、上位のコメントに関連する発言です
- 各コメントに発言者の属性を示しました。発言者の属性は、アカデミアの情報系(INFO)(11名)、アカデミアの人文・社会科学系(SSH)(12名)、パブリックセクター(6名)、インダストリー(5名)、メディア(5名)、その他(5名)の6分類としました

総論

タイトルについて

- なぜ ELSI (Ethical, Legal and Social Implications: 倫理的、法的、社会的課題)ではなく Ethically だけなのか(アカデミア SSH)
- Aligned とは、人間や環境と調和するようにデザインするということで、Aligned to Human という意味で使っているのか。あるいは、Aligned to moral value and Ethical principle ということか(アカデミア SSH)
 - 倫理的というのは修飾語であって、むしろ人間に Aligned「調和する」ことが大事なのか。日本語の「なじみのある」デザインという理解でいいのか(アカデミア SSH)

AI とは何か

- 情報倫理と違い AI 特有の問題は何かを示し、切り分けて議論してほしい(アカデミア SSH)
 - ブラックボックスの話があるが、それは他の機械でも同じ(アカデミア INFO)
 - AI の特徴はプログラムを作るのが人間から IT になること。そうすると人間がプログラムを把握できなくなり予測ができなくなるのが問題(パブリックセクター)
- AI にはいくつかの段階がある(アカデミア INFO)
 - 1) 人がプログラムコードを書くためインプットとアウトプットがはっきりしている段階
 - 2) データを学習しコードに反映してパラメータを変える段階(与えるデータが大事になる)
 - 3) パラメータを環境からのフィードバックで自分自身作り直していく段階
 - 4) 環境からのフィードバックで自分自身の複製をつくる段階、そこまでいくとシンギュラリティ。現状では第 4 段階目にちょっとはいりかけたというイメージ
- ロボットは安定性が保証されているが AI は答えが「正しい」という判定はできない。「正しそう」としか言えず、安定性が証明されていない(アカデミア INFO)
 - ハードウェア、ソフトウェア、その中に AI がある。AI は検証不可能性が上がる。データベースの中の振る舞いなどの再現性がなくなる。人間の能力がそれに追いつかない。ディープラーニングなどの重みづけ配列の意味とか人間にはわからない(インダストリー)
- AI は独立で動いているのではなく、ネットワーク化しているという視点が大事。自動運転も通信しあっている。IoT なども繋がっているという前提での議論が必要(アカデミア INFO)
- データを処理するマシンが AI。データをどうやって収集するかとかデータの議論も大事(その他)
- AI 技術は発展途上なので無理に定義しないで、倫理の議論と技術の開発を並行してやらざるをえない(パブリックセクター)
 - AI を「まだ見ぬもの、発展途上のもの」とみるか「もうすでにあるもの」と見るかによって考え方が違う。総務省の報告書は「もうすでにあるもの」としているが、本書のように汎用人工知能(AGI)を入れるのであれば、「まだ見ぬもの」という考えも入る(アカデミア SSH)

本書の位置づけと持続可能性について

- 標準を作るには政治や社会のコンセンサスが必要になる。そういうスキームをどう作るのか(標準機構など第三者機構を作るのか)(アカデミア INFO)。
- IEEE の標準化に貢献する人たちがボランティアに活動しているという状況は持続可能性という点で厳しいのではないか(パブリックセクター)

- ・ 日本の人工知能学会も指針を出したが、出すことによって学会の責任が増える。学会が編集委員会とかで標準を守っているかを担保する仕組みを作るのは難しい(アカデミア INFO)
 - ・ 倫理問題をそもそも誰が検知するのか、また問題が起きた時に責任や調査ができる組織があるのか、またその権限はどこが持つのか(アカデミア INFO)
- ・ 今後この活動がどうなっていくのかの説明がほしい。ビデオメッセージなどわかりやすいものをレイトカメラに公開してほしい。標準を作るにあたって「誰でも入っていい」というメッセージはわかりにくいので、その意味について説明してほしい(パブリックセクター)

技術者等が EADv1 を遵守するインセンティブはあるか

懐疑的意見

- ・ 技術者側には、これを作成することに参加するインセンティブがない(アカデミア INFO)
 - ・ 基準は後から作るべきなのに政治的な合意が最初にあることに対して、政治に対する技術者の不信がある(アカデミア INFO)
- ・ 規制ばかりでは民間は参加できないので、ポジティブなものにしてほしい(インダストリー)
- ・ 社会的不安にこたえつつもイノベーションを推進するような指針が必要。もう少し緩いものにしないと産業界としては採用しにくい(パブリックセクター)
- ・ 人文社会科学系の研究者にはこれに参加することが研究のインセンティブになる(アカデミア SSH)
- ・ ホワイトリストやブラックリスト、事例集があるほうが開発側はやりやすい。グレーゾーンばかりだと委縮してしまう(アカデミア INFO/パブリックセクター)
- ・ 日本企業は第二次 AI ブームの時、技術的には失敗したというトラウマがあるので、海外の Partnership on AI のようにまとまるのが難しいのではないか(パブリックセクター)

肯定的意見

- ・ 金融における「規制の砂場」的なソフトローが可能になるものになってほしい(アカデミア SSH)
- ・ 日本の「特区」が本書に書いてあった。そういう仕組みを作っていくことも大事(アカデミア SSH)
- ・ このひな形が OECD のプライバシー原則にあるのではないか。規格さえ守っていればお墨付きがもらえるので企業には順守するインセンティブになるのではないか(アカデミア SSH)
- ・ 欧州議会やアメリカ・日本政府などがすでに指針や報告書を出している中で、技術者集団としてプレゼンスを示すことが大事(アカデミア SSH)
- ・ コンサル業界では顧客と個別に枠組みを考えるよりは、あらかじめこういったお墨付きや一定の枠があると、仕事を進めやすくなる。枠組みや法律があることによって、仕事をしやすくなることもある(インダストリー)
- ・ 厳しいレベルで標準が作られると、それを担保できるのはうちの技術だけですよということで競争相手に先んじることができる(インダストリー)
- ・ IEEE などが倫理原則をつくり、これを自社のコンピテンシーに組み込むことで、企業としては開発が進めやすくなる(インダストリー)

項目の生成や排除されたものに対する説明

- ・ 8つの項目を合理的に順序付けるべき。どういう論理でこの順になっているのか不明(インダストリー)

- ・ 「AI が持つべき倫理規範」と「AI 開発者が持つべき倫理規範」を分けてかんがえるべき。それが2と3の違いなのか(アカデミア INFO)
- ・ 現在議論されているトピックのうち、「議論に上ったけれど 8 項目に入らなかった」の議論のリストがほしい(メディア)

対象読者を整理してほしい

- ・ 誰に対する規制、提言であるのか(個人、業界、国、人類など)を明確にして議論しないと、議論が散漫になってしまう(インダストリー/アカデミア SSH)
- ・ たとえばアクターを開発・運用(ビジネス)・ユーザー・政策関係医者・法律関係者・メディア・一般市民などにわけてほしい。同じトピックでも違うアクター視点から整理してくれるとわかりやすい(アカデミア SSH)
- ・ 技術で解決できる問題と技術だけでは解決できない問題(運用や利用に関して)があるので、切り分けてほしい。技術評価とユーザーの影響評価など(アカデミア INFO)
 - ・ 一方で、分けると責任の所在があいまいになるのではないか(アカデミア SSH)
- ・ 対象は AI 開発者というが、開発者とは誰か。AI アルゴリズムをかく人なのか、データの学習をさせる人なのか。(パブリックセクター)
 - ・ マイクロソフトの TAY などはデータの与え方がよくなかった例(アカデミア INFO)
- ・ 技術者に守ってほしい標準だけではなく、社会の側からもフェールセーフがほしい。たとえば自動車運転で賠償とか訴訟とかがあるように。ユーザーの負担はなく責任のアカウントビリティがあるもの、ユーザーが安心してとりあえず使ってみるということができるような仕組みや基準がほしい。そうしないとユーザーも自動運転は怖くて使えない(アカデミア SSH)

オープン化とマルユース(悪用)の議論をすべき

- ・ オープンソースについて触れられていないが、AI の進歩はネットと連携しているので、ネットのオープンカルチャーについても議論すべき(アカデミア INFO)
 - ・ オープンにすることで悪用される可能性もあるが、悪意ある人よりも守る人員を増やすためにオープン化するという視点が必要(アカデミア INFO/その他)
 - ・ 悪意をもって使うは開発よりは使う方の倫理の問題ではないか(アカデミア INFO)
- ・ この文書はルールを守って従ってくれる人たちを対象とした性善説の立場に立っている。そのためルールによる禁止で悪意ある AI システムを作ることを止めることはできない。倫理をもって縛るという考え方ではなく、信頼を得て味方を得るという考え方が建設的。AI ビジョンを守る人を守る、という合意のためにも AI ビジョンは必要(その他)
 - ・ IEEE は技術者を守ってほしい。悪用されたりうっかりミスをしったりしても守ってくれるというのが組み込まれていないと、この倫理規定を守るというインセンティブがはたらかないのではないか(アカデミア SSH/メディア)
 - ・ この文書のメッセージはみんながこれをまもればうまくいくこと前提だが、そうではない。人の役に立つ AI というビジョンを共有していない人たちをどのように配慮に入れながら作っていくかという視点が必要(アカデミア SSH)

その他不足している視点

- ・ フェアネスの技術的な分析が欧州では進んでいるがそれがはいていない。(アカデミア INFO)
- ・ HCI(Human-Computer-Interaction)の議論がほしい。アフェクティブコンピューティングなどの話が次からは入ってくるのだろうが、AI を環境とつなげるインタフェースの話は大事(アカデミア INFO/SSH)

本コメント報告書について

- ・ 日本で今回まとめたものに対して、一つ一つとはいわないでも、どういう風に採用されたかあるいは採用されなかったのかの返事がほしい(パブリックセクター)

各論

1. General Principles

- ・ 一般原則で守らないといけない価値はかなり絞らないといけないのではない。「課題」と書くのではなく、これは絶対に守らないといけないもの、と書くべき(その他)
 - ・ 一般原則に「これはだめ、あれはだめ」と細かく書くのはあまり賛成できない。円滑に回る原則と拘束する原則の間をとるのが難しい。(パブリックセクター)
- ・ 「人権」を考えた時、人権意識が低い国がこの約束事に入ってこられるのか(メディア)
 - ・ 一方で、国連の人権宣言が最高位の価値なので、それは共有しているという前提がある。価値観はバラバラでも一定程度は Align しているという前提がないといけない(アカデミア SSH)
- ・ 標準を作ることと、多様性に配慮することは矛盾しないか。多様性を狭めるのが標準ではないか(アカデミア INFO)

2. Embedding Values into Autonomous Intelligent Systems

- ・ 人を中毒にしてお金を巻き上げる、といった AI も作れる。目的設定で反社会的行動を起こす AI は、どのような規範があれば防げるのか(アカデミア INFO)
 - ・ 差別的な AI(保険査定など)はしょうがないという人もいれば、ダメだという人もいる。目的自体はビジネスあるいは社会的なもので設計されるが、結果として反社会的なものになるかもしれない(メディア/アカデミア INFO)
 - ・ 差別的 AI は文化的価値観によって変わる(メディア/アカデミア INFO)
- ・ アカウンタビリティは透明性を支える重要な概念であるが、説明をテキストレベルで行うのは難しく、今の技術ではできないのではないか(アカデミア INFO)
 - ・ 医学の領域では「この薬を与えたらこうなる」という correlation(相関)は出すことができ、相関がこのくらいだから大丈夫という。人間社会ではそれをトラストしているために医学は進展する。アカウンタビリティのようにトップダウンで中身を説明するのではなく、「トラスト」構造がどういう役割を果たすのかを考えるのが大事ではないか。相関をオブザベーション(トップダウンではなくボトムアップ)から見ないと AI に対する恐怖は一層できないのではないか(アカデミア INFO)
 - ・ AI に入れたインプットと出てくるアウトプットの Correlation から、その Causation を導くのは今の技術では無理だが(例えば Tesla が事故を起こした際に、Tesla 搭載の AI がどのような「勘違い」をして事故に至ったのか、実はコードを見ても人間にはわからない)、それをどうにかするの、

(1) Correlation から Causation を導くための研究に予算を付ける、という方法と、(2)それは難しいのであきらめてそれ以外の方法(確率統計などで)でトラストを得る、という二つのやり方がある。現段階では、両方とも追求していくのが良いと思うので、各国政府は、(1)につながる基礎研究にも、(2)につながる事例を増やす研究(新薬のトライアルのようなもの)にも、予算をつけてほしい(パブリックセクター)

- 実世界からデータを集めるのが難しくても、アルファ碁でやっているように、事故や事件をシミュレートするなどやりようはある。外界から得られたファクトをうまく使っていく視点が必要(アカデミア INFO)
 - ただし、囲碁などと、車などサイバーフィジカルと実世界は違うのではないか(パブリックセクター)
 - 自動運転などは、過疎地とか長距離トラックとか必要とされているところから順次導入していったデータを集めて信用を得ていくような仕組みが必要(アカデミア INFO)
- 透明性の確保は基本的に難しい。この文書は全体的に説明責任を重く置いているが、それを産業側が認めてしまうと自分たちの首を絞めることになるので、産業側はこういうのに参加したくなくて一定の距離を置いているのが現実(パブリックセクター)
- プログラムが人種差別になるかどうかを考えて開発できるのか。オールマイティな問題を一度に考えて作れるわけがない。そうすると、技術限界や適応限界を書くのが現状できることではないか。(アカデミア INFO)
- 文化的多様性について、現実にある差別を固定化するシステムはまずい。何らかの形で改定するループを入れるシステムにしてほしい。人間の意見が入るような改定が必ず入る形にして欲しい。たとえばある特定の集団をつぶしたいとか思う人たちがいたとしても、それに反対あるいは注意を喚起するようなシステムをいれておいてほしい(アカデミア SSH)
 - たとえば犯罪予測システムと共謀罪が組み合わさったら、かなりやばいシステムになるが、そのような開発を禁止できるのか(アカデミア SSH)
 - 国によっては、たとえば女性蔑視とかが残っているところある。そのバイアスを含んだままのデータでAIを作ることは避けるべきという規定が必要ではないか。少なくともバイアスがないかを常にフィードバックする仕組みを導入することが必要。
 - そうするのは政治の問題であって技術の問題ではない(アカデミア INFO)
- 「使いたくない人」への配慮はどのように考えるか。Have nots ではなく Want nots への配慮はあるのか。介護ロボットなどは使いたくない人もいるかもしれない(アカデミア SSH)

3. Methodologies to Guide Ethical Research and Design

- AI 設計に技術者の規範が無意識のうちに組み込まれているということに気付かせる教育は必要。「あなたは会社や特定の目的のためにシステムを作っているかもしれないが、規範がプログラムに反映されている」ということに気づかせることは大事(アカデミア INFO)
- 科学コミュニケーションや倫理教育ができる人材が大事だということを書いてほしい(アカデミア SSH)
- データサイエンティストの育成のカリキュラムとかにも倫理教育が入っているのか(アカデミア SSH)
- 教育の標準を作るのは難しい。データサイエンティストに必要な要件とかも、個別具体事例で異なるので標準を作るのは難しい。倫理の教育標準を作ることができるのか(アカデミア SSH)

- ・ 技術者倫理教育にIEEEは力を発揮していて、日本の教科書にも反映されるが、向こうの価値観と日本の価値観が合わないときがある(アメリカ的な内部告発を推奨するのは日本では無理)(アカデミア SSH)

4. Safety and Beneficence of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI)

- ・ AGIの議論が4番目にくることに違和感がある。最後でよいのではないか(メディア)
 - ・ AGIという文言が出てきた瞬間に現実じゃなく未来の話をしている気がする。実行力や、拘束力や意味がなくなる気がする。今の産業のAIを考えていないと感じてしまう(パブリックセクター/アカデミア INFO)。
 - ・ 一方で、これ以外の項目はAI固有でもない。これがあることでAIらしさが出せると思っているのか(アカデミア SSH)
 - ・ 全体の構成としてこの章があることで、社会一般からの脅威に対して対応しているのではないか(アカデミア SSH)
 - ・ ターミネーターとまではいかないけれど「ありうる脅威」が一般にはある。悪意のある人をどうすればいいのかとか、人間が間違った指示を出すことによってAIが暴走したらどうなるのかなど(メディア)
 - ・ ギリシャの国際危機や東証など、AIの暴走で実際に脅威は起きている。それは打ち消すべき脅威ではなくて実際にある脅威。それを不安がないということは危険で、むしろちゃんと脅威を示していくべき(アカデミア INFO/メディア)
 - ・ AGIは未来の話だが、現状のビジネスで課題になっている機械学習などと分けて議論するためにあえてここに入っていると考えると納得がいく。すぐにでも機械学習を導入したいが、足かせになっている企業にとっては、AGIとそうではないAIのラインを決めて欲しい(インダストリー)
 - ・ ここに書いているのは、近視眼的、近未来的すぎる。本来AGI、ASIはもっと人間に近いものであるが、本書はそういう視点に行っていないくて即物的。もっとAGIらしい話をしてほしかった(アカデミア INFO)

5. Personal Data and Individual Access Control

- ・ プライバシーや個人情報に関する議論は、GDPR(EU 一般データ保護規制)ができてここ20年くらい国際的な会議を通じて随分と行われている。また、各国の対応も進められている。ここに書いてあることは、それらの後追いであって、新しい議論ではないように思える。AIと個人データに特化した議論は何かについて、もっと焦点を絞るべき。もっとも、他の多くの個人データに関する国際会議も、最先端の議論として、AIの問題を論じ始めている。とすれば、そうしたところと協力・協働することが必要かもしれない(アカデミア INFO/SSH)
- ・ スノーデン事件のように、個人の情報収集を政府がやっていることを考えなければいけない。兵器の話ではなく、むしろその話が問題である。個人のデータが自分の意図に反して収集されて、それが例えば自律型兵器のターゲティングデータとかになるかもしれない、そうすると6の議論と近くなる。プロファイリングに対する拒否権(EUのGDPRでは取り入れられた)を認めるなどの権利を新しく作る必要がでてくるかもしれない(アカデミア INFO/SSH)

6. Reframing Autonomous Weapons Systems

- 兵器について Reframing(再構成)という自律型兵器を受け入れているという印象がある。そういう印象を与えない方がよいのではないか。許容できないリスクがあるのであれば、禁止等の対応が必要ではないか(アカデミア SSH)。
 - 最低限守るというより、ここはまずいというラインを切るのがよいのではないか。(アカデミア INFO)
- 無差別兵器の使用は禁止されている(ABC 兵器や対人地雷などの使用が国際条約で禁止されているのは過剰な傷害だけでなく、無差別的であるため)。自動制御と無差別兵器は直接関係ない気もするが、事故が起きた時などに無差別的になるといった可能性の払しょくは難しいところ、そのリスクにどう対応するか議論はもっと必要ではないか。例えば、悪意あるものが無差別的な攻撃に使用する可能性やプログラムの不具合による無差別攻撃のリスクへの言及などを想定している。(アカデミア SSH/パブリックセクター)
- 戦闘時など、人間が本当に管理できるのか。自律型兵器は戦争がはじまりやすいということも書いてある(アカデミア INFO)
- 兵器に関しては倫理と科学技術がまだ発展途中だと思う。人間がコントロールできる状態にもっていくのが理想だが、そこまで信頼できる技術があるわけではない。(パブリックセクター)
- AI 特有の技術的な議論をもっと書き込むべき。ここに書かれている兵器は古く感じる。個人データと結びついてプロファイリングして攻撃することも可能になっている。5 の Personal Data の議論と近くなる(アカデミア INFO)
 - スタックスネットなどの攻撃的兵器もある。そういうのも扱うべきではないか(アカデミア INFO)
 - 「軍事兵器」の考えが固定化されているのではないか(インダストリー)
 - 自動追尾ミサイルとかができた点で、もうすでに自律型兵器はできていて、AI 固有の問題ではない。もっと他に議論すべき軍事兵器はある(インダストリー)
- 自律型兵器によって身体権が失われる状況に対してそれは技術の問題なのかどうか、どういう手続なら許容可能か、現実どこまで行われているのかなどの議論が必要(その他)
 - AI 自身が危険をターゲットできるのか、人間が危険と思うものを入力するのかなど、技術的にどこまで自動化できるのかなどの境界をどう考えていけばいいのか。窃盗とか予測ならともかく共謀罪とかと組み合わせると怖い(その他)
- デュアルユースの議論をもっと入れるべき(アカデミア INFO)
 - 兵器と民生用の兵器を分けることができない世の中である(アカデミア INFO)
 - 売れるのなら経済の問題であって軍事の問題ではない。結局のところ目的はどちらかによって、6(安全保障)か 7(経済産業)に入るかということにならないか。経済産業ならば兵器に関する話でも 7 の枠組みで語ってしまうのでは(メディア/アカデミア SSH)
 - デュアルユース技術は軍か民生か分けることができないという定義。そのため技術ではなくどのような目的なのかから考えないといけないのではないか(メディア)
- なぜ自律型兵器の議論が必要なのかなど、基本的な設計理念の説明が欲しい。なぜ単純に「禁止」ではなく「再構成」なのか。例えば人道目的の自律兵器利用は OK というように、基本原則を最初に示してほしい。動物実験などと同じで、「戦争するのは当たり前だけれど殺すときに自国の兵士がトラ

ウマとかストレスで苦しまないようにするのが人道的」だから自律型兵器を使うという話もある(メディア)

- ・ 我々日本人はこの議論をしてこなかった。だからこそ、なぜ、どういう時なら自律型兵器がOKなのかという意見や基本原則がほしい。それに対して反対か賛成かというところから議論が始められるが、そこを共有していないので、日本人としては議論がしにくい。また自律型兵器の話は1の人権の話との整合性はあるのか。あるいは軍が使うならOKだけれどテロリストはダメというような基本的な原則はあるのか(メディア)
- ・ それだと政治の話になって技術の話ではなくなる(アカデミア INFO)
- ・ 技術者はどこまで何をコミットできるのかというのがわからない。実際、技術者がコントロールできる範囲はどのくらいなのか。技術影響力の範囲が、技術者が考える以上に広がっているのか(その他)
- ・ 人間の兵士が自律的であるという前提がある(インダストリー)
- ・ 軍事兵器への不安というよりは、ブラックボックスであることへの不安のほうが書かれている気がする(インダストリー)

7. Economics/Humanitarian Issues

- ・ 雇用の話あまり書かれていない(アカデミア INFO)
 - ・ 技術と雇用の問題はベーシック・インカムの話までいかないのであれば AI 固有ではないのではないか(メディア)
 - ・ 雇用の話は社会的影響の問題なので、技術者を対象とした本書では書くべきことではないのではないか。技術者は雇用の問題まで対応できないのではないか(メディア)
- ・ 経済的に儲かれば何でもいいのかというのは 2 の価値の問題と一緒に考えないといけない。WELQ 問題(信憑性の問われる医療情報を公開したキュレーションサイトに対する著作権侵害や倫理的問題)は、医療情報を扱っていたから問題化した。ただしそれは経済的に「儲かりたい」というのであれば、当たり前にかかることである。人々がそれに中毒性を持つものとなると問題になるのか(メディア)
- ・ 2 の議論に近いが、技術設計の意図によらず技術が反道徳的になったりする。たとえば特定の人の監視や人事について AI を利用することがどこまで許されるのか(アカデミア SSH)。
- ・ 経済的合理性ではなくウェルビーイングなどほかの尺度が必要になるのか。またそういう他の尺度を作ったとしてもそれがビジネスモデルとして価値があるのか。経済成長をしなければならないという大前提について議論が必要ではないか(メディア)
- ・ AI が算出した、「合理的だけれど納得がいかない」保険の計算とかをどう考えればいいのか。自分では意識していないがリスクなどで保険料が不当に高くなることもある(その他)
 - ・ AI も間違える。環境とのインタラクションでお門違いな料金のつり上げされることもあるかもしれない。AI 万能的な感覚を持ちがちだけれど、それに警鐘を鳴らし拒否権を発動できることが必要である(アカデミア INFO)
 - ・ AI には不確実なところがあると提唱していくこと自体が、新たな不安や不信を生み出さないか(アカデミア SSH)
 - ・ AI と人間ができることの比較ができればいい。人間より AI のほうがミスは少ないなどの納得

できるエビデンスをだせればいい。トラストをどうやって作り出すかという議論が必要(アカデミア INFO)

- ・ メディアが間違っただけを出してはいけないと書かれている。パブリックにリーチするのは技術者よりはメディアや政策で、そこに正しい情報出すのが技術者の役割なのか(アカデミア SSH)
 - ・ 一方で色んな人に議論に参加してほしいというけれど、それは誰なのか(その他)
- ・ メディアチェックする団体を作るとき、科学的に正しいとか決めつけること自体が議論を阻害しないか。多様な議論と矛盾しないか。そうしないためにも、技術者が責任をもって発信していきましょう、程度にとどめるのがよいのではないか(その他)
 - ・ ファクトチェックをするべきと書くことは議論を狭めないか(パブリックセクター)
 - ・ ファクトチェックをしても、みんな従わないから問題にはならないのではないか。明らかに間違いなのは訂正していこう程度ではないのか(アカデミア SSH)
 - ・ 何が「ユートピア」で「ファクト」かがわかりにくいので、具体的な事例とかがほしい(アカデミア SSH)
 - ・ 差別とかは明らかな間違いなのかもしれない(アカデミア SSH)
- ・ ユーザーへのリテラシー教育はどこに入るのか(アカデミア SSH)
 - ・ そこまで技術者がやるべきことなのか(アカデミア SSH)
 - ・ 安心のためのメディア啓発などは広告を打てばいい(アカデミア SSH)
 - ・ リテラシー教育をすると逆に不安をあおるのではないか。社会的な問題がでてきてから、リテラシー教育の必要性が出てくる。インターネットもそうだった。問題がおきてからでないと技術者とかが乗り出してくるインセンティブはないのではないか(メディア)
 - ・ それまでは自主規制で対応するなどインシデントベースになってしまっている(アカデミア SSH)
- ・ フィルターバブルの問題などは AI 固有の問題だと思う。フィルターバブル自体は昔からあるが、AI とインターネットのスピードは違うので議論すべき(アカデミア INFO)

8. Law

- ・ なぜ法が倫理の中に入っているのか。技術で解決できるものと倫理あるいは法で解決できるものがあるという切り分け、振り分けはどのようにして行われたのかを書いてほしい。ここに書かれている事項は法律でやったほうがいいのかと思われていることだろうが、なぜそうなったのか(アカデミア SSH)
 - ・ トロッコ問題などは倫理ではなく法律の線引き問題、社会的意思決定の問題として扱うというのは、コンセンサスが得られているのか(メディア)
 - ・ トロッコ問題はもう緊急の問題になってきているので倫理ではなく法で扱うということなのではないか。それならばそのように書いてほしい(アカデミア SSH)
 - ・ 法律家に対応してほしい問題として書いてあるが、アクターが誰なのかということも含めて、誰にどの項目を見てほしいのかというのは最初の説明段階で振り分けてほしい(アカデミア SSH)
- ・ 法で対応せずに自主規制でやったほうがいいのかもあるのではないか(アカデミア SSH)
- ・ ライアビリティ(Liability)が最初に書いてあるが、本文中にその文言がほとんどないのはなぜか。アカウントビリティやリーガルアカウントビリティとライアビリティの関係はどのようなものなのか(アカデミア SSH)
 - ・ 責任の分担を誰がどこにあるのかという義務を明記するのがライアビリティだとすると、それを考

えるのは技術者ではないので、なぜ本書にこれが書かれているのかの説明がほしい(アカデミア SSH)

- AI は万能じゃないので事故を起こす。そのとき世論(マスコミ)は何故起きたのか知りたいが、きっと説明はできない。そのときに応じて調停機関や医療事故調査制度とかが必要となる。第三者委員会についての言及が何箇所かあるが、より具体的に書くべき(メディア/インダストリー)
- 中身ではなくプロセスの標準化が必要。倫理審査を受けるというプロセスを標準にするなど(アカデミア SSH)
 - 枠組みを再構築できる再帰的な仕組みを入れておくなど、責任の所在が明確になっていくことが大事だ(アカデミア SSH)
 - プロセスを標準化するよりも、フレームワークに余白を持たして抽象度を上げて、バイデザインなどアーキテクチャ的に作り上げられないか(メディア)
 - それだと事故調査の場合の条件や、どれくらい因果関係があるかはわからなくなる(アカデミア SSH)
- AI 技術はスピードが速いので、今までのように PDCA サイクルをまわしていくものだと遅すぎるかもしれない。監査が年に 1 回とかだと追いつかない(インダストリー)
 - 技術も検証不可能性が高くなる。その時ゼロリスクを求める人たちがいるのが日本(メディア)
 - 一方で、ゼロリスクのために BSE の全頭検査をやる国だから、トラストを個人が咀嚼できるレベルになっているのが日本。そういう国があってもいいのでは(インダストリー)
- 検証可能で「安全」といえる領域にのみ標準は乗り出すべきであって、安心など主観的なものは別のスキームで担保すべき。たとえば国際標準化の中で信頼の標準化、証明書などの議論はある。信頼のトレーサビリティなどを電子的に検証する仕組みなどを技術で考えることは可能(インダストリー)
- アセスメントではなく事後的な評価はどこが行うのか。飛行機の事故調査は国とかが専門調査チームが行う。制度としてつくれたら開発者も楽になるのではないか。責任追及と原因究明の分離ができるような仕組みが必要(アカデミア SSH)
- 知的財産権とか著作権とかの議論も大事ではないか(アカデミア INFO)
 - 自分の著作物が他人の著作物を侵害していないかということを考えるときに、AI に責任を負わせることができるのか(その他)
 - AI で分析するためにデータを集めるのは OK になるなどの制度を整えてほしい。AI 創作の著作権とかを柔軟にするなど。日本はコンテンツクリエイションで食べている国であるため(アカデミア INFO)
- 個人データのインテグリティの議論は、5 で論じた方が適切に感じる。または、データ保護は、法規制の部分が重要であるから、8に取り込んでしまったほうがいいかもしれない(アカデミア SSH)。
- 技術者側からは、AI では何が起きるか予測できないという声を聴く。しかし、そうであるとすれば、法は何を規律・規制対象としていけば良いのかわからないのではなかろうか。法規制が遅れ気味ともいわれるが、個別の技術が開発されるごとに、いわば後発的に問題点を洗い出して考えていくよりほかないようにも思える(アカデミア SSH)

以上