

人工知能と社会について考える場づくりの実践

Practice of creating a community to discuss Artificial Intelligence and society

江間有沙*¹
Arisa Ema

長倉克枝*²
Katsue Nagakura

田中和哉*³
Kazuya Tanaka

藤田卓仙*⁴
Takanori Fujita

工藤郁子*⁵
Fumiko Kudo

*¹ *³ *⁴ 東京大学
The University of Tokyo

*² 科学ライター
Science Writer

*⁵ 中京大学
Chukyo University

This report discusses issues in society surrounding people and Artificial Intelligence based on inter-disciplinary workshops. To argue the social image surrounding technology and society, we asked experts in various fields to consider what they regard as "Dystopia" to paradoxically reveal the current social image that they desire. The project name is "Project Emerg" that considers "Emerg (ency + ing) societies & technologies."

1. 人工知能をめぐる社会的課題

人工知能の普及に伴う倫理的・法的・社会的課題 (Ethical, Legal and Social Implications: ELSI) への懸念は大きい。内閣府では論点を6つに整理している[内閣府 17]。

- 倫理的論点: 人の感情・行動の操作や人間観・倫理観の変化
- 法的論点: 事故等による責任分配や創作物の権利など新たな課題や権利
- 経済的論点: 機械を利活用することにより変化する雇用内容や環境の整備
- 教育的論点: 機械と協働するための能力育成
- 社会的論点: 技術利用による差別や社会的構造の不均衡, 機械への依存
- 研究開発的論点: 研究者の技術開発に対する倫理規定やガイドラインの整備

このような論点は世界各国のコミュニティで議論されており, 2016 年後半期だけでも多様な団体・組織が人工知能・ロボットと社会に関する報告書を発表している(表 1)。

これらの提案に大枠として共通する特徴が 2 つある。第一に, 技術が内包するリスクを減らし人類がベネフィットを享受できるよう適確な運用・規制が大事であるとする。そのためには何がリスクでベネフィットなのかという「価値」(Value)の問題に取り組むことが課題となる。それが第二の特徴, すなわち個々の価値が多様化・タコツボ化しているため異分野・異業種間協働の必要を説くことにつながる。

表 1. 報告書リスト

年月	国/組織名	タイトル	内容
2016年6月	日本 総務省	AIネットワーク化の影響とリスク	AIネットワーク化の社会的影響評価指標と開発原則策定の推進
2016年9月	非営利団体 Partnership on AI	Partnership on Artificial Intelligence to Benefit People and Society, Tenets	AI研究に関して 便益を最大化し課題に対処するよう努める8つの信条を公開
2016年9月	アメリカ スタンフォード大学	Artificial Intelligence and Life in 2030	2030年のAIを予測し, 教育など8領域について議論, ガイドラインの必要性を提案
2016年10月	イギリス 下院科学技術委員	Robotics and artificial intelligence: Fifth report of session 2016-2017	ロボットやAIが雇用, 社会, 倫理法などにもたらす影響を検討し開発原則策定に関する委員会設置を提言
2016年10月	アメリカ ホワイトハウス	Preparing for the Future of Artificial Intelligence	AIの社会や政府での活用事例とその規制についての23の提言
2016年12月	学術団体 The IEEE	Ethically Aligned Design, Version 1	AI・自動システムの開発ガイドラインや標準策定のための8項目
2017年1月	非営利団体 FLI	Asilomar AI Principles	AIの研究課題, 倫理や価値への配慮と長期的な課題について23の原則
2017年1月	欧州議会 法務委員会	Report with Recommendations to the Commission on Civil Law Rules on Robotics	欧州におけるロボットとAI推進の民放的な問題にかかる欧州統一規則策定の提言

連絡先: 江間有沙, 東京大学, cema@mail.ecc.u-tokyo.ac.jp

2. 本企画の枠組み

近未来の技術と社会の相互作用をどのように捉えるかによって社会像やシナリオの描き方は変化する。近年の「技術が雇用を奪う」とする調査も技術的可能性から置き換わり度合いを算出するが[Frey 13], 経済や規制などとの相互作用は考慮に入られていないことから「技術が社会を変える」という技術決定論的なシナリオになりがちである。複雑化・多様化している社会で未来のシナリオを作るためには, 多様な分野・業種の専門家の知見が不可欠である。事実, 表 1 にある報告書執筆には十数人, 多いところでは 100 人以上の異分野の専門家が参加している。

そのほか未来のシナリオを異分野の研究者や実務家だけではなく, 小説家や舞台作家などと共に作る試みも存在する。例えば, 2013 年の欧州人工生命会議(ECAL)では”A TRUCE workshop on Unconventional Computing in 2070”と題しショートストーリー作家やジャーナリスト, アーティストなどと科学者が 2070 年を想定した Speculative Fiction や作品を作るセッションが開催された。これは欧州の政策・助成プログラム(FP7)で行われたものである。このようにアーティストとの協働は特に新しい手法でもなく Critical Design, Speculative Design など議論を喚起するデザインの系譜がある。

このような先行事例を踏まえ, 筆者らは(1)技術と社会を取り巻く社会像の整理にあたって, そうなつてほしくない多くの人が考える社会像を「ディストピア」と呼び, それは何かを考え整理することで, 逆説的に現状とそうなつてほしい社会像を浮かび上がらせようと考えた。またそのアイディアだしには(2)異分野の専門家による協働が必須となる。そのため, 筆者らは「ディストピア

を考える」をテーマに掲げた異分野ワークショップを 2017 年 1 月に 2 回開催した。そこでは, (A) 5 年後位の近未来を舞台としたディストピアシナリオ・小説を作り, (B)その上で, そこで描かれる技術や社会について解説する調査報告書を 1 年かけ

て書くことを目的として掲げた。企画名は「危機的で萌芽的な社会や技術」(Emerg(ency+ing) societies & technologies)の状態や使い方を考えるということで”Project Emerg”とした。

3. 結果

3.1 ワークショップ概要

ワークショップは2017年1月9日(土)と14日(土)に開催し、それぞれ15名、27名が参加した。学部2年生から大学教員などの研究者のほか劇作家や小説家等にも参加いただいた。研究者は人工知能研究者のほか、哲学・倫理学、政策学、社会学、法学、医学、生命科学、科学技術政策、科学コミュニケーション論、科学技術社会論など多様な分野の人に参加いただいた。1回目は企画趣旨と世界観設定と今後の運営方針について、2回目は各参加者がディストピアのテーマとなりそうな事例を持ち寄ってワークショップを行った。2回のワークショップ後、(1)得られた「ディストピア」社会像・世界観の整理と、(2)異分野の専門家の協働研究をどのように行うかについて検討を行った。

3.2 社会像の設定:何を「ディストピア」ととらえるか

ワークショップ企画時は、技術の両面性に着目し、技術者や政策決定者には「想定外」の事故や事件が起こることで「ディストピア」となる社会像を想定していた。類似の視点を持つものとして不正アクセスやハッキング、なりすましや3Dプリンタを用いた銃の製作など、主にサイバー空間を舞台にしたミステリ小説「サイバーステリ」がある[一田 15]。これらの小説の多くには被害者がおり、サイバー空間のリスクやベネフィット、そのリスク対策について専門家や一般ユーザが考えるきっかけとなる。

しかし、議論を進めているうちに明確な「悪意」が事故や事件を引き起こすのではなく、皆が「善意」で動いているのに最悪な方向に行くシナリオこそがディストピアなのではないかとの議論が行われた。また、このように複雑化・多様化した世界像や価値観においては、表1に挙げている調査報告書が扱っているような「未来の技術」を引き合いに出すまでもなく、情報技術はすでに様々な「ディストピア」状況を引き起こしていることが指摘された。そのようなディストピア状態を作り出す構造的な要因を大きく3つに分類する。

1. **既存システムとの不整合:** サイバー犯罪の大多数は犯罪者による悪意が存在する。しかし、情報化医療など健康促進を意図した遠隔医療や日々の健康状態のモニタリングは、医師の負担を増やし医療環境そのものを崩壊させかねない[日本医学会総会 15]。新しい技術や仕組みを既存の体制にそのまま置き換えることによって、余計な仕事が増えたり不整合が起きたりする。また、人とシステムが複雑に入り組んだ場合は、被害者はいるが事故や事件の責任主体が不明確になる可能性もある。例えば、医療用人工知能や自動運転など人間の判断を自動システムが補助する場合においてはメーカー、利用者と機械、環境の責任分配の方法が制度設計上の問題となり、抜本的な制度改革が必要になるがそれに追いついていない。
2. **設計思想の問題:** 複雑化した環境における制度設計あるいは技術設計の根本思想が「全体最適のためには、ある特定の層の犠牲や局所的な不利益もやむなし」となる場合がある。特に情報技術は最適化や効率化の問題を解くのが得意であるため、渋滞緩和や物流・金融等の問題を解くにあたって、多くの人が気づかぬうちに操作・誘導される可能性がある。
3. **利用者の多様性:** 価値が多様化しているため、ある人にとってはディストピアであるものが他の人からすると便利や楽になるなど「善意」で設計されている技術とみなされる。例えば、ブラ

イバシーの問題は、不快に思う人はその価値を尊重される必要がある。また、情報技術は「ロングテール」など選択肢の多様化を促進させる一方で、その反動として価値や判断を画一化させる「モンスターヘッド」を生み出し[エルバース 15]、フィルターバブルのようにタコツボ化する状況も生み出している[パリサー 12]。

このように分類してみると、いずれの「ディストピア」もすでに起きていることであり、明確な悪意や事故・事件が生じるディストピアと比較すると地味である。しかし、これらの地味なディストピアの背景には「責任とは何か」、「仕事の本質とは何か」、「全体最適において民主主義とは何か」、「自由意志とは何か」といった根本的な問いがある。身近すぎて気づきにくい、あるいは当たり前と思いこんでしまっている事例だからこそ、価値や文脈の異なる人たちとの対話によって気づきが得られる可能性がある。

3.3 協働方法:どのように多様な人を巻き込むか

このような根本的な問いに気づくためには、価値や考えをぶつけることが重要になる。しかし、価値が多様化してタコツボ化している現在、同じ価値や目的を共有している研究者同士では安心して対話や議論ができるが、特定の分野の専門家の研究や考え方は、他の分野の人にとって懸念を呼び起こし危険なものとして議論を呼ぶ、いわゆる「炎上」と言われる状態になることがある。「炎上」することで研究や自身が社会的に非難されたり、研究の障害になるという懸念から、専門家は他の分野の人と自分の研究や考え方を共有して議論をしたくても、躊躇したり避けたりすることがある。一方、「炎上」することで、想定外のフィードバックを受けて、研究や自身の考え方を良い方向に向けていくこともできる。本ワークショップを継続的に開催するには異分野の専門家が安心して「炎上」できる場、互いの価値を尊重し対話する姿勢を持つ人々が集まる場をいかに作るかが重要であるということ、またアートやユーモアなど、人が話したくなるような問題設定をすることが重要であることが本企画では再認識された。

4. 今後の展望

コリングリッジのジレンマと呼ばれる技術評価の古典的研究がある[Collingridge 81]。技術が社会で使われる前にその影響力を予測することは難しいが、一度普及してしまった技術は制御するのが難しい。そのために今後も本企画は人工知能と社会について考える様々なテーマ設定や場を考えていく予定である。

参考文献

- [Collingridge 81] David Collingridge: The Social Control of Technology, Palgrave Macmillan, 1981.
- [エルバース 15] アニータ・エルバース: ブロックバスター戦略, 東洋経済新報社, 2015.
- [Frey 13] Frey, C. B., Osborne, M. A.: The Future of Employment, <http://www.futuretech.ox.ac.uk/future-employment-how-susceptible-are-jobs-computerisation-oms-working-paperdr-carl-benedikt-frey-m>, 2013.
- [一田 15] 一田和樹 他: サイバーステリ宣言!, KADOKAWA, 2015.
- [パリサー 12] イーライ・パリサー: 閉じこもるインターネット, 早川書房, 2012.
- [内閣府 17] 内閣府: 人工知能と人間社会に関する懇談会, <http://www8.cao.go.jp/cstp/tyousakai/ai/>
- [日本医学会総会 15] 第29回日本医科学総会 2015 関西: 2025年「不都合な未来」, https://www.youtube.com/watch?v=fq8-FQ_T8c4