

2023年4月27日

AIの開発・利活用政策についての共同提言

デイビッド・ファーバー David Farber¹

公文俊平 Shumpei Kumon²

註：以下の記述はあくまでも著者個人としての見解であり、所属する組織の見解を反映するものではない。

1. AIについての我々のビジョン

AIへの期待

近年のAI技術、とくに生成系AI技術の加速的な発展はまことにめざましいものがあり、その社会的な応用が拡大することは、貧困、紛争、人権の抑圧、感染症、経済恐慌、環境問題、さらにインフォデミック³など、人類社会が直面する地球規模の脅威や課題の解決に大きく貢献することが期待される。我々はその意味で、AI技術の開発のさらなる進展を基本的には歓迎するものである。

AI技術がもたらすリスクについて

同時に、しかしAI技術は、その開発や利活用において、その潜在能力の強大さと同等の大きなリスクが伴うことも、忘れてはならない。したがって、我々はAIのリスクについての分析を継続的に実施し、その結果に基づいて必要な対策を動的に立案・実施していくことが求められる。いうまでもなく、このリスク分析と対策の実施、総称して「AIガバナンス」は、世界規模で推進されるべきものであり、国際的な協力を本質的に必要とする。

AIは我々が30年前に経験したインターネットの登場に勝るとも劣らない、きわめて

¹ 特別招聘教授(国際) グローバルリサーチインスティテュート(KGRI)、サイバー文明研究センター共同センター長 教授

² 多摩大学情報社会学研究所所長・教授

³ <https://www.who.int/health-topics/infodemic>

大きな社会的インパクトを与えるものであることは間違いない。AI の潜在能力は、国であれ企業であれ、特定の少数者に専有されるべきものではなく、インターネットがそうであったように、可能な限り開かれた形で、かつ透明性を維持する形で共有され、人類全体がその恩恵を享受できるものでなければならない。

生成系 AI の位置付けについて

2008 年のいわゆるリーマン・ショック後の経済の低迷のかげで、「新しい加速的成長過程」というべき状況が、初めはひっそりと、やがて多くの人々を興奮と熱狂の渦に巻き込むかたちで始まった。それが AI 技術の新たな突破である。これまでのところ、「突破」は 2 段階でおこった。最初が「機械学習」技術の手法「ディープラーニング」の精緻化で、第 2 が最近の“ChatGPT”に代表される「生成系 AI (generative artificial intelligence)」技術である。

「弱い AI」と「強い AI」の区別の重要性

生成系 AI の具体例である ChatGPT は、自然言語処理機能に特化した AI であり、(再定義が必要かもしれないが) チューリング・テストをほぼ確実にクリアする点で一種の知性だと考えられる。現段階の生成系 AI は、特定用途に限定して使われる場合には、一定の有用性をもつが、あくまで「弱い AI」ないし「狭い AI」にとどまるものである

我々の社会にとって本質的な脅威となるのは、「意志」や「主体性」をもつ「強い AI」の出現であるが、いくつかの根拠により、現段階での AI はこの方向には進化していないと考えられ、我々は「強い AI」と「弱い AI」という種類の異なる 2 つの AI はきわめて明確に区別すべきものとする。

公文は 2017 年の『近代の成熟と新文明の出現—人類文明と人工知能 I』(NIRA 研究報告書) で「特化型 AI を人類の福音とせよ」として、次のように提唱した。⁴

その途を見失うことなく [特化型 AI との] 共生に成功すれば、近代文明は真の成熟を迎えつつポスト近代文明と重畳して有終の美をなすことに成功するだろう。人間と人間能力拡張型人工知能がペアを組み、人間の自由意思や自律性を維持しつつ、平和的に共働・共生する社会では、人工知能は人類にとって福音となるだろう。⁵

現段階の AI に対する望ましい社会的対応

生成系 AI によって作られたコンテンツは、意図しない誤り、ないし意図的なフェイク

⁴ ここでいう "特化型 AI" とは、"弱い AI" と同義である。

⁵ https://www.nira.or.jp/paper/report_1708.pdf

を含む可能性が高い。したがってユーザーは、生成系 AI を無条件に信じたり、無邪気に利用したりしてはならない。その意味で、我々は、生成系 AI の利用方法・形態などについての社会的ルールの制定・普及を広く推進することを提案する。

他方、現在の生成系 AI は、いわゆるシンギュラリティ理論が提起した「人類に対する反乱」などの超自然的な危険性は持たないと考えられ、当面は、市場と智場⁶においてそのアウトカムに対する適切な規制を行いつつ、人工知能がもたらす可能性を追求することが重要だと考える。

G7 のリーダーシップに期待する

この意味で、今回 2023 年 5 月に日本で開催される G7 首脳会合において、AI ガバナンスがその重要な議題の一つとして取り上げられることを、我々はおおいに歓迎する。我々は、G7 のリーダーシップが全世界の AI ガバナンスの取組みを正しい方向に導き、AI によるリスクを最小限に抑えつつ、未来の人類社会、すなわち成熟に向かう情報社会が現在よりもさらに平和で豊かでかつ楽しいものとなるために AI が大きな役割を果たすようになることを期待する。

したがって、我々は、2019 年に採択された OECD の「AI 原則」⁷をはじめ、人々の権利とウェルビーイングを最上位の優先事項に据えた米国政府の「AI 人権章典」⁸、日本政府の「AI 戦略 2022」⁹、欧州連合の「AI 戦略 (2018)」¹⁰、および欧州評議会による AI 条約策定の取組み¹¹ を歓迎し、今後一層のハーモナイゼーションの進展と、グローバルな規模への拡大を期待する。

⁶ 公文は産業社会における「市場」に対比して、情報社会に出現する情報空間を「智場」と名付けた。

⁷ <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

⁸ <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>

⁹ https://www8.cao.go.jp/cstp/ai/aistratagy2022en_ov.pdf

¹⁰ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN>

¹¹ <https://rm.coe.int/cai-2023-01-revised-zero-draft-framework-convention-public/1680aa193f>

2. 具体的な提言

以下、世界の人々は具体的に何をすべきかについて、現時点で最低限必要と思われる点に絞って、より具体的な提言を試みる。

技術開発についての情報公開

AI 技術の開発がさらに進んで、人間に並ぶ、あるいは対抗する意志や目的を持った自律的な AI、いわゆる「強い AI」や、さらには「人工超知能」¹² が“2045 年”（カーツワイル）前後に実現する可能性はゼロではなく、そうなれば私たちの生活に及ぼす影響には計り知れないものがある。我々は「生成系 AI」の使用や、「人工超知能」の開発を全面的に禁止せよとはいわないが、十分以上の備えを進めておくことが必要不可欠と考える。それが整う前に技術開発だけを先行させる危険は冒すべきではないだろう。

ただし、技術分野の開発は、基本的には自由に推進されるべきものであり、とりわけ未知数が大きい AI 分野においては、イノベーションを阻害する固定的な法・規制は避けるべきものとする。可能な限り民間主体の創意と市場・智場における利用者の選択による発展の道を確認し、規制を行う場合には、明確な被害・損害が確認される場合に限定し、できる限り既存の法的枠組を活用することを求める。

したがって、AI の技術開発を推進する主体に対して、その現状と展望について定期的かつ自発的にタイムリーな情報公開を行うことを求める。各国政府および関連国際機関は共同して、公開されるべき情報の内容・方法などについて一定の基準を策定し、その実施を義務付ける制度を制定することを提案する。この制度の策定にあたっては、上場企業に課せられる情報開示義務が参考になるだろう。

この点で我々は、米国政府 NTIA が 4 月 11 日に発表した「AI アカウンタビリティ政策」に示された RFC を通したオープンな政策形成方針¹³ を基本的に支持する。また、日本政府がすでに経済産業省により「AI ガバナンス・ガイドライン」¹⁴ の実装を推進していることを支持し、各国がこれを参考とすることを提案する。

¹² ニック・ポストロム、『スーパーインテリジェンス 超絶 AI と人類の命運』（日本経済新聞出版、2017）、井上智洋、『「人工超知能」－生命と機械の間にあるもの』（秀和システム、2017）

¹³ <https://ntia.gov/press-release/2023/ntia-seeks-public-input-boost-ai-accountability>

¹⁴ https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20220128_report.html

AIの利活用ガイドラインの制定と普及推進

現在の生成系 AI の革新性とその大きな可能性に鑑みると、利用者側が AI の能力と限界について十分に認識したうえで利活用を行うことがきわめて重要と思われる。我々は、生成系 AI の利活用を無条件・包括的に禁ずる立場には与しない。AI 提供事業者側が、上述した「AI ガバナンス・ガイドライン」に沿って、利用者側に対して十分な情報をタイムリーに公開・提供すべきことは当然であるが、同時に、道具を使う側の責任についても所在を明確化し、誤った形で利用することで生じる負の結果については、利用者自身がその責任を負わなければならないということも、改めて明確に確認すべきと考える。情報社会における真の主体は、情報技術を活用する人々（市民もしくはネティズン）と考えるからである。

より具体的には、「AI 利活用ガイドライン」の策定と普及推進が求められる。ただし、それは単一のものとは限らない。たとえば利用者の属性（市民、学生、企業などの所属）や技術スキル、利用目的などに応じて異なる基準が必要だろう。今後、政府、企業、市民、研究者などの多様な主体が連携するマルチステークホルダー方式で、それぞれの分野別のガイドラインの制定に取り組み、その普及推進にあたることを提案する。

以上、G7 サミットの日本開催を前に、これまでのコンピュータ科学・技術、インターネット技術とその応用・普及の発展の歴史を振り返り、情報社会が人類にとってさらに望ましい方向に発展していくことを切望して、我々の提言を試みるものである。

本提案についてのお問い合わせは、以下までお願いします。

デイビッド・ファーバー教授:

admin@www.ccrcc.keio.ac.jp

公文俊平教授:

iza@anr.org