

特集号解説記事

# The AI Alliance：オープンコミュニティの力で加速する持続可能なAI

福田剛志<sup>1</sup>

<sup>1</sup>日本アイ・ビー・エム（株）東京基礎研究所

AIは変化のときを迎えている。2022年から現在に至るまでの基盤モデルに基づく生成AIの進歩は、新しい能力・アプリケーション・アーキテクチャを導き、人間の生産性を大きく改善する可能性を秘めている。一方で、社会は強力だが不透明なAIによるリスクに晒されることとなった。これまでITの目覚ましい発展は、オープンコラボレーションやオープンリサーチによって推進されてきた。オープンソースやインターネット技術は、コミュニティによるオープンな協力関係を通じて、安全で信頼のおけるものに発展した。The AI Allianceは、AI技術も同様にオープンに技術開発することによって、その恩恵をすべての人に開かれたものであることを保証すべきであるという理念のもと、IBMとMetaが主導して設立したAIのためのオープンコラボレーションのための団体である。The AI Allianceは活動を開始したばかりであるが、本稿では、The AI Allianceが、どのような活動を通じてこの目的を達成しようとしているのかについて紹介する。

## 1. 背景

第3次AIブーム<sup>☆1</sup>は、2012年に物体の認識率を競うILSVRC 2012[1]において、ジェフリー・ヒントン率いるトロント大学のチームがAlexNet[2]によって、従来26%だったエラー率を17%と、劇的に改善したことに端を発している。その後も急速に改善が進み、2014年には人間のエラー率と言われる5%に達した。これは、（1）大量のラベル付きデータ（ImageNet[1]）、（2）多層ニューラルネットワークを用いた学習アルゴリズム[3]、（3）GPUを用いた計算能力[4]の3つが同時に揃ったことによって実現されたと言ってよいだろう。

過去のAIブームはその狭間に長い冬の時代が訪れた。しかし、2017年に発表されたTransformer[5]モデルに基づいた自然言語で対話できるChatGPT[6]が2022年に公開されると、AIブームはさらに加熱し、第3次AIブームが終焉する前に第4次AIブームとも言うべき新しい生成AI[7]ブームが到来した。

第3次AIブームで性能を向上させたAIは、AIの専門家の手によって何らかのシステムに埋め込まれて利用されていたのに対して、第4次AIブームで注目を集めているChatGPTやその類似システムの自然言語対話型のAIは、専門家知識を有しない一般のユーザが、日常使う言葉でAIと対話することにより、AIが事前学習した知識を自由に引き出すことができる。これによって、AIの活用が従前よりはるかに容易になった。AIへの「アクセス」が民主化されたと言ってよいだろう。このことの社会に対するインパクトの大きさは、1990年代半ばに現れたインターネットブラウザ[10]が社会に与えたインパクトと比較すると分かりやすい[11]。ブラウザ以前は、特殊なライブラリやアプリケーションを通じて利用されてきたインターネットが、ブラウザが誕生したことにより、突然一般ユーザが容易に利用できるようになった。その結果、インターネットへの「アクセス」が民主化し、電子商取引やネット通販やSNSが一般化し、現在のインターネット社会が形成された。もちろんデバイスとしてスマートフォンが普及したことも、この現象に拍車をかけた。

このような背景の下、AIは少数の人々によって形成されるにはあまりにも重要な技術になってしまった。AIはすべての人の生活を一変させる可能性の技術なのだから、その発展を少数の企業に委ねるのではなく、可能な限り多くの人がかかわるべきである。このような考えに基づき、IBMとMetaが中心となって2023年12月にThe AI Alliance[12]が発足した。当初は50団体が参加したが、2024年8月現在、100団体を超えるメンバーが上記の理念に賛同し、Allianceの下で活動している[12]。

---

## 2. 活動目的

---

The AI Allianceは設立時に「構築する」、「可能にする」、「提唱する」の3つの大きな目的を設定した。

**【構築する (Build)】**では、社会の健康、持続可能性、繁栄に不可欠な、特に言語やその他の形式における、ソフトウェア、モデル、関連ツールなど、AIのオープンな技術を構築し、公開し、サポートする。

**【可能にする (Enable)】**では、開発者や組織がオープンな技術を採用できるよう、啓発活動、チュートリアル、コミュニティサポート、教育・トレーニング、そしてユースケースのデモを通じて促進する。

**【提唱する (Advocate)】**では、組織や社会のリーダー、政策・規制機関、そして一般の人々と協力して、広範な利益を可能にし、課題に取り組み、信頼と安全を促進する活気あるAIのオープンなテクノロジー・エコシステムを作る。

---

## 3. フォーカスエリア

---

これらの3つの目的の下、The AI Allianceは以下の6つのフォーカスエリアを設定し、さらにその下に複数のワーキンググループを作って、具体的な活動を行っている。

### フォーカスエリア1：スキルと教育

このフォーカスエリア (Skills & Education) では、アカデミック・コミュニティと提携し、研究者や学生がAIモデル、アルゴリズム、プラットフォームの重要な研究に従事できるよう支援している。また、AIの利点とリスクについて一般ユーザや政策立案者に伝えるための教材やリソースを作成し、解決策を提示することにより、正確で十分な情報に基づいたAI規制を提唱している。

## フォーカスエリア2：信頼と安全

生成AIのモデル評価は、最も重要な投資分野の1つであると同時に、最もエントロピーが高い分野の1つでもある。このフォーカスエリア (Trust & Safty) では、AIの信頼性と安全性のリスクの状況を理解し、それらのリスクを検出し軽減するためのツール、方法、ベンチマークを構築する。また、責任あるAIについて一般市民を教育し、責任あるモデルやアプリケーション開発について開発者コミュニティを支援する活動も行っている。

まず信頼と安全全般に対する認識を高めることが重要である。このための取り組みには、たとえば世界的に行われているサイバーセキュリティやCBRNE (Chemical, Biological, Rediological, Nuclear, and Explosive) への対策活動が含まれる。特にAIの信頼性に関しては、信頼できるAI (Trustworthy AI[13]) に関する学術・技術コミュニティを育成し成長させ、現在のAI評価の範囲を超えて、私たちがよく見えていない領域まで推し進めるような専門家たちを作ることが大切である。

また、モデル評価のための包括的で信頼できる安定したツールの開発を推進することも重要である。これらのツールは、多面的だが再現性がある結果を提供すべきである。また、定期的に更新され、生成AIの新たなリスクやその他の評価に関する問題を知るにつれて、進化し続けるべきである。

一般的に、モデル評価はモデルの能力に注目する。しかし同時にモデルがある種のリスクを持つ可能性を明らかにすることもできる。評価方法がより難解になるにつれ、大きな課題の1つは、生成AIにかかわる専門家にとってすら、評価の詳細が不明確になることだ。増え続ける潜在リスクに対して、AIモデルを評価する場所が、コミュニティに現在存在しないことが問題である。

## フォーカスエリア3：ツール

このフォーカスエリア (Tools) は、モデルのトレーニング、チューニング、推論のためのオープンソースツールを構築し、推進する。また、クラウド上での大規模AIワークロードの実行を簡素化、自動化、最適化するための共同研究も行う。

情報化された急速に進化する世界では、さまざまな分野の専門家が、複雑な質問に対して即座に、そして的確に、ニーズに合わせた回答を必要としている。最新の治療法を求める医師から、市場動向に関する洞察を必要とする金融アナリストまで、正確で分野に特化した情報に対する需要はかつてないほど高まっている。しかし、現在のAIシステムでは、専門家レベルの意思決定に求められる深さと正確さを提供できないことが多く、大きな非効率と機会損失につながっている。この課題に対処するため、このフォーカスエリアのメンバーは、RAG[14]技術を使用してドメイン固有の質問応答のためのベストプラクティスに関する包括的な研究を行った[15]。

#### フォーカスエリア4：多様なハードウェア

高度なAIを作るためには、高い計算能力が必要となる。このため、AIワークロードのベンチマーク、最適化、応用などを共同で推進することにより、オープンコミュニティが多様なハードウェアを利用可能にすることが重要である。このフォーカスエリア（Diverse Hardware Enablement）では、スケーラビリティ、プラットフォーム適応、および高度な電力・エネルギー・炭素排出量モデリングに重点を置いて、大規模AIワークロードの性能を評価するベンチマークとメトリクスを開発することで、オープンソースコミュニティに貢献することを目指している。

#### フォーカスエリア5：基盤モデル

責任を持ってオープンな基盤モデルのエコシステムを強化することで、多言語、マルチモーダル、さらには気候変動や教育など、広範な社会的課題に取り組む科学モデルを積極的に採用し、そのオープンコミュニティによる開発・利用を促進していくべきだと考える。

AIモデルの構築者やアプリケーション開発者を支援するため、モデルの訓練、チューニング、推論のためのオープンソースツールの開発と普及に向けて協力を行っている。また、安全で有益な方法でのAIのオープンな開発を促進するプログラムを立ち上げ、AIのユースケースを探求するためのイベントを開催している。

このフォーカスエリア（Foundation Models）の活動として、2024年5月には材料科学における研究開発を加速するためのワーキンググループ“Open Data and Models for Materials & Chemistry”を発足させた[16]。材料産業は企業間の競争も激しいが、非競争領域での基盤モデル、基盤データ、ベンチマークタスクをオープンコミュニティで開発し、業界全体としての生産性を高めることが望まれる。

#### フォーカスエリア6：提唱

産業界・市民社会・学術界にとって、AIのための活発で競争力のあるオープンイノベーションエコシステムは優先事項でなければならない。これらのエコシステムは大部分が分散して自主的に運営されているが、その成長を支援する政策の余地はある。

The AI Allianceの活動は、政策立案者や政府がAIのオープンイノベーションエコシステムを認識し、支援するのに役に立つ。また、政策と産業界の間のギャップを埋め、世界中の社会に利益をもたらす責任ある倫理的なAIの実践を促進するための活動を行う。

たとえば、2023年10月の米国大統領令[17]において指示され、米国商務省電気通信情報局（National Telecommunications and Information Administration: NTIA）が2024年2月に行った“Dual Use Foundation Artificial Intelligence Models with Widely Available Model Weights”[18]に関する意見公募に対して、このフォーカスエリア（Advocacy）のメンバーが意見をまとめ、オープンな基盤モデルは、科学的発見や経済成長を促進し、AIの安全性を高めると主張した意見を提出した[19]。これを受けて、NTIAはオープンモデルによるAIの確信を支持する報告書[20]を発表し、オープンな基盤モデルの重要性と、コミュニティ主導のリスク管理フレームワークを提案した。

## 4. 今後の展望

オープンイノベーションを推進するためには、法整備が欠かせない。EUでは2023年6月にAI規制法が成立した[21]。米国でも大統領令[17]に続き、各州で法制化の議論（たとえば[22]）が進んでいる。日本では、2024年4月に総務省と経済産業省が「AI事業者ガイドライン」[23]を発表し、法制化の議論も行われている。オープンコミュニティの立場から、リスクを最小化しながら、オープンイノベーションを妨げない規制の在り方を、まさにオープンに議論していくべきであり、The AI Allianceへの期待は大きい。

大規模モデルの性能向上は著しいが、現時点ではクローズドモデルが一步リードしている。従来のオープンソースソフトウェア開発に比べて、AIモデルの開発には遥かに多くの計算リソースが必要であり、大きな資金を持たない個人や企業がAIモデルに貢献することが難しい点である。しかし、オープンモデルの進歩のスピードのほうがクローズドモデルよりも早いという報告[24]もある。またオープンソースソフトウェアでは、誰でもその修正や改良をプル・リクエストとして、本体に取り込むことを提案できるが、AIモデルにおいては、その本流を改良できるのは、そのモデルの正規の保有者だけである。この問題点を改善するため、最近オープンにモデルの改良を提案できる方法（たとえば[25]）が提案されている。今後一層のオープンコミュニティの活動に期待したい。

### 参考文献

- 1) ImageNet Large Scale Visual Recognition Competition 2012 (ILSVRC2012), <http://image-net.org/challenges/LSVRC/2012/> (2012).
- 2) Krizhevsky, A., Sutskever, I. and Hinton, G. E. : Imagenet Classification With Deep Convolutional Neural Networks, Communications of the ACM, 60 (6): pp.84–90 (2012).
- 3) Schmidhuber, J. : Deep Learning in Neural Networks: An Overview, Neural Networks, 61: pp.85–117, arXiv:1404.7828 (2014).
- 4) Oh, K.-S. and Jung, K. : GPU Implementation of Neural Networks, Pattern Recognition, 37 (6): pp.1311–1314 (2014).
- 5) Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, Ł. and Polosukhin, I. : Attention Is All You Need, Advances in Neural Information Processing Systems, 30 (2017).
- 6) Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I. and Amodei, D. : Language Models are Few-Shot Learners, NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems: pp.1877–1901 (2020).
- 7) Gozalo-Brizuela, R. and Merchan, E. : A Survey of Generative AI Applications, Journal of Computer Science (2024).
- 8) Turing, A. : Computing Machinery and Intelligence, Mind 49: pp.433–460 (1950).
- 9) Rychener, M. : Production Systems as a Programming Language for Artificial Intelligence Applications, Carnegie Mellon University (1976).
- 10) Lasar, M. : Before Netscape: the forgotten Web browsers of the early 1990s,

Arstechnica (2011).

11) Krishna, A. : IBM CEO on Doing What's Right and the Future of Technology, Distinguished Speaker Series, Duke University, [fuqua.duke.edu](https://fuqua.duke.edu) (2023).

12) The AI Alliance, [thealliance.ai](https://thealliance.ai) (2023).

13) Li, B. et al. : Trustworthy AI: From Principles to Practices, ACM Computing Surveys, 55-9 (2023).

14) Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S. and Kiela, D. : Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems: pp.9459-9474 (2020).

15) Enhancing Q&A with Domain-Specific Fine-Tuning and Iterative Reasoning: A Comparative Study, <https://arxiv.org/abs/2404.11792> (2024).

16) AI Open Innovation Day Tokyo, <https://thealliance.ai/blog/open-innovation-day-tokyo> (2024).

17) US Executive Order 14110, Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (2023).

18) Dual Use Foundation Artificial Intelligence Models with Widely Available Model Weights, NTIA-2023-0009, <https://www.ntia.gov/federal-register-notice/2024/dual-use-foundation-artificial-intelligence-models-widely-available-model-weights> (2023).

19) Response to Request for Comment on "Dual Use Foundation Artificial Intelligence Models with Widely Available Model Weights", [https://thealliance.ai/core-projects/ntia\\_request](https://thealliance.ai/core-projects/ntia_request) (2024).

20) Dual-Use Foundation Models with Widely Available Model Weights, NTIA, <https://www.ntia.gov/issues/artificial-intelligence/open-model-weights-report> (2024).

21) Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828, (2023).

22) カリフォルニア州法案 : Safe and Secure Innovation for Frontier Artificial Intelligence Models Act (SB-1047) (2024).

23) 総務省, 経済産業省 : AI事業者ガイドライン, [https://www.meti.go.jp/shingikai/mono\\_info\\_service/ai\\_shakai\\_jisso/pdf/20240419\\_1.pdf](https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20240419_1.pdf) (2024).

24) Laonne, M. : <https://x.com/maximelabonne/status/1781344505536327686> (2024).

25) Sudalairaj, S. et al. : Lab: Large-Scale Alignment for Chatbots, arXiv preprint arXiv:2403.01081 (2024).

#### 脚注

☆1 AIブームには多様な考え方があるが、1950年～60年代の記号処理と探索[8]に基づくAIを第1次ブーム、1980年～90年代のエキスパートシステム[9]に基づくAIを第2次ブームと呼ぶ。



福田剛志（正会員）

fukudat@jp.ibm.com

1991年早稲田大学大学院理工学専攻修士卒業。同年日本アイ・ビー・エム（株）入社。東京基礎研究所にてオブジェクト指向データベース、データマイニング、機械学習、バイオインフォマティクスなどの研究に従事。2004年同社ソフトウェア開発研究所に異動し、データ統合ソフトウェアの開発に従事。2015年東京基礎研究所所長。2021年同社執行役員。博士（情報科学）。

投稿受付：2024年8月14日

採録決定：2024年11月14日

編集担当：込山悠介（国立情報学研究所）