Hitotsubashi University
Institute of Innovation Research

第四次AIブーム（ChatGPT）による世界のAIガバナンス制度の進化
～ChatGPT型AIシステムの社会的リスクと世界のAI規制・ガバナンス政策の動向～

市川類

IIR Working paper WP#23-02

2023年5月

Evolution of World ' s AI Governance Systems Driven by the Fourth AI Boom (ChatGPT)
- Social Risks of ChatGPT-type AI Systems and Trend of AI Regulation and Governance Policies in the World -

Tagui Ichikawa

一橋大学イノベーション研究センター

東京都国立市中2-1
http://www.iir.hit-u.ac.jp

第四次 AI ブーム（ChatGPT）による世界の AI ガバナンス制度の進化

～ChatGPT 型 AI システムの社会的リスクと世界の AI 規制・ガバナンス政策の動向～

一橋大学イノベーション研究センター
市川類
2023 年 5 月 15 日

概要

2022 年末から、対話型の生成系 AI システムの一種である、LLM（大規模言語モデル）を活用した ChatGPT（Generative Pre-trained Transformer）型の AI システムに対する関心が爆発的に高まってきており、今や、第四次 AI ブームの様相を示している。この ChatGPT は、これまでの AI 技術と比較しても、革新的なイノベーションを引き起こしうる技術であり、引き続き今後の経済成長・社会課題の解決に多大な寄与が期待される一方で、将来的な AGI に対する不安感・不信感が再度世間の中で台頭しつつあるとともに、実際に、人間社会に対してリスクをもたらす可能性が指摘されている。

それでは、この ChatGPT 型 AI システムに関し、その技術的特徴を踏まえると、どのようなリスクが生じることが見込まれ、また、第三次 AI ブーム以降進められてきたこれまでの世界の AI 規制・ガバナンス体制は、今後どのように変化・進化していくことが見込まれるのであろうか。

このような認識の下、本ワーキングペーパーでは、ChatGPT に係る技術的特徴とそれがもたらしうるリスクについて考察した上で、この ChatGPT に対応した世界における新たな AI 規制・ガバナンス政策の動向に係る全体像及び現在地を示すことにより、今後の世界の AI ガバナンス制度の方向に対する示唆を得ることを目的とする。

具体的には、まずは、第三次 AI ブームにおいて欧米を中心に生じた AGI への不安感、期待感が現在の第四次 AI ブームにつながっていることを示した上で、今回の ChatGPT の技術的内容とその限界を改めて整理することにより、人間との対話によるイノベーション創出にも有効であること、また、従来型の意思決定型の AI システムとは異なるリスクが今後想定されることについて考察する。

その上で、まずは非営利機関（FLI）による GPT 開発中断に係る提言の内容とそのインパクト・合理性等を分析した上で、欧州、米国、日本、中国など世界における ChatGPT 型の AI システムに係る規制・ガバナンスの動向について、これまでの動きを整理する。これらを踏まえて、1） 今後とも、先進民主主義国間では、各国・地域の事情に応じ、多様な AI 規制・ガバナンス制度の進化が見込まれること、2）一方、従来の意思決定系の AI システムでは個人情報、公平性・人権が重要であったのに対し、今回の ChatGPT 型の AI システムにおいてはむしろ著作権、民主主義などのキーワードが重要になる可能性があること、3）世界的な観点からは、中国の独自規制により中国市場の孤立化などの可能性があること、などを示した上で、今後の G7 の閣僚宣言を踏まえつつ、今後の世界の AI 規制・ガバナンス制度の方向について論点を整理する。

目次

１．ChatGPT の爆発的な普及と本 WP の問題意識

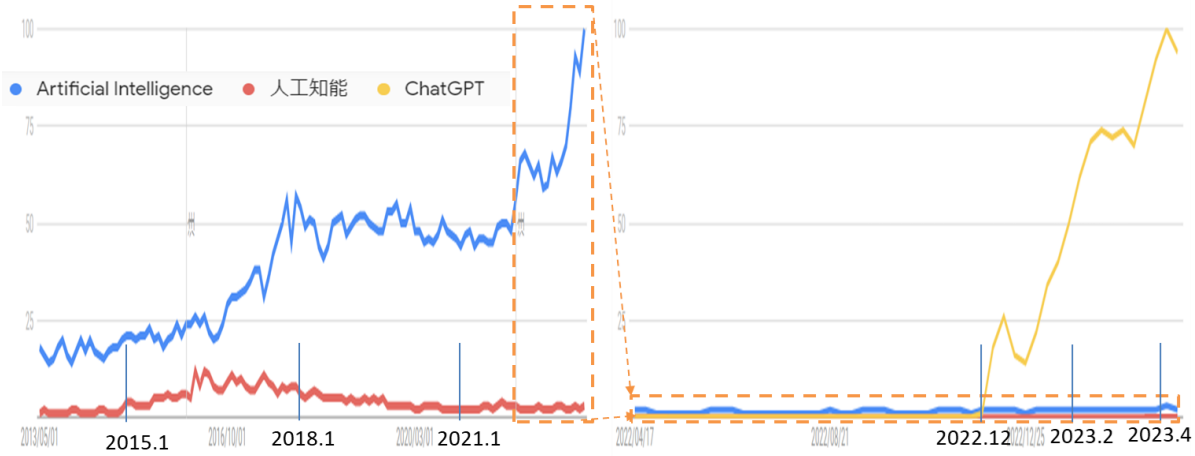
（１）ChatGPT の爆発的な普及と第四次 AI ブームの到来

2022 年末から、大規模言語モデル（LLM：Large Language Model）を活用したいわゆる対話型の生成系 AI システムの一種として位置づけられる ChatGPT（Generative Pre-trained Transformer）に対する関心が、爆発的に高まってきている。

実際に、2015～16 年ごろから始まったいわゆる第三次 AI ブームにより、人工知能（AI：Artificial Intelligence）に対する検索数は急増し、その後、世界的に見れば引き続き高いレベルにあった（一方、日本では、2017 年頃をピークに、検索数は低下傾向にあった）。

これに対し、ChatGPT に対する検索数は、2022 年 11 月末の発表直後から爆発的に増大しており、2023 年 4 月現在では、AI の検索数の 40 倍近くに至る（図１参照）。また、この ChatGPT に対する爆発的な関心の高まりにつれて、AI そのものに対する関心も急上昇しており、実際に、AI に対する検索数も再上昇している。このような中、現在は、あたかも、第三次 AI ブームの後、AI の冬の時代が来る前に、新たな第四次 AI ブームが幕開け（松尾）[1]したような状況にあると言える。

【図１】Artificial Intelligence、人工知能、ChatGPT に係る検索数推移（世界）[2]



ChatGPT は、圧倒的に大量の既存のコンテンツ（文章）を利用した機械学習と強化学習を通じ、入力（プロンプト）に対して、確率的にもっともらしい文章を作成し、回答として出力するシステムである（以下、ChatGPT の類似のシステムも含めて、「ChatGPT 型 AI システム」と言う）。この ChatGPT の登場は、半導体などハードウェア技術に係るこれまでの指数関数的な継続的な進展により、非常に大規模な計算を必要とする大規模言語モデル（LLM：Large Laguage Model）の構築が可能になったことに加え、強化学習を含む各種の

[1] NHK サイエンスゼロ「インターネットを超える衝撃！？第４次 AI ブームの到来【博士の 20 年 vol.6】」2023 年 3 月 13 日
https://www.nhk.jp/p/zero/ts/XK5VKV7V98/blog/bl/pMLm0K1wPz/bp/pj27knKK8B/
[2] 出典：グーグルトレンドより筆者作成。2023 年 4 月 11 日現在

機械学習技術を組合せることによって実現可能となったものであり、AI 技術のイノベーションの進展によってもたらされた新たなブレークスルーであると位置づけられる。

その際、第三次 AI ブームでは、大量の計算量の必要とする深層学習（ディープラーニング）技術の登場により、「コンピュータが人間の『目』のような認識能力を持った」と言われるように、画像を含め各種データから生物・人間が行うようなパターン認識をする機能を得たことに加え、それらの機能を含めていわゆるビッグデータ解析を行うことによって、特定の目的に対し、データの基づく各種の予測、推薦、決定（自動化）することが可能なシステムが構築されるようになり、それらのシステムの普及が進展してきた。

これに対し、今回の第四次 AI ブームにおいては、更に大量の計算量を必要とする LLM の構築・登場により、コンピュータが、人間の知性の表出形態である「言語」を操る能力を持つようになったことが特徴であると言える。今後、本技術を活用して、幅広い分野での人間との言語を通じた会話等を可能とするシステムが構築され、それらのシステムの社会での普及と更なるイノベーションの進展が進んでいくものと考えられる。

【図２】第三次 AI ブームと今回（第四次）の AI ブームの特徴の比較[3]

	鍵となる技術	主要普及システム	ガバナンスの動き
第三次AIブーム	DL（深層学習技術）⇒人間の認識能力（「目」など）の確保	認識⇒意思決定型システム	AGIへの懸念⇒AI原則 ↓ 人権・公平性の考慮等
第四次AIブーム（仮称）	LLM（大規模言語モデル）⇒人間の言語を操る能力の確保	文章生成⇒対話型システム？	更なるAGIへの懸念⇒？ ↓ （具体的リスク対応？）

（２）AI の社会受容に係る地域差とこれまでの世界の AI 規制・ガバナンス

＜イノベーションと技術の社会受容＞

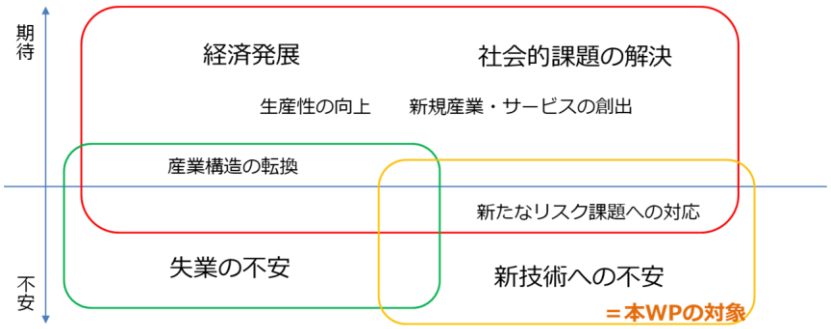
一般的に、技術に係るイノベーションにおいては、新たな技術が創出されるとともに、当該技術がさらに発展しながら、社会の中で普及していくことになるが、その社会での円滑な普及のためには、社会において当該技術が受容されることが前提になる。

その際、革新的な技術によるイノベーションの進展は、生産性の向上や新規産業・サービスの創出等を通じて、経済発展だけでなく社会的な課題の解決にも資するものとなる一方で、イノベーションは、創造的破壊という側面を有する。具体的には、産業構造の転換を通じて、単に既存の産業・雇用を破壊するという側面を有するだけでなく、当該技術によって確立される新たなルーティンは、既存のルーティンとは異なり、その利用形態によっては地域社会が共有する社会規範に抵触することになり、社会に対して新たなリスクを生じさせるものとなる（図３参照）。

したがって、イノベーションの普及においては、技術の普及によって生じるこのようなリスクなどに対する不安感に対して、社会受容を進めていくことが課題になる。

[3] 出典：筆者作成

【図３】新技術・イノベーションの社会受容として期待と不安[4]

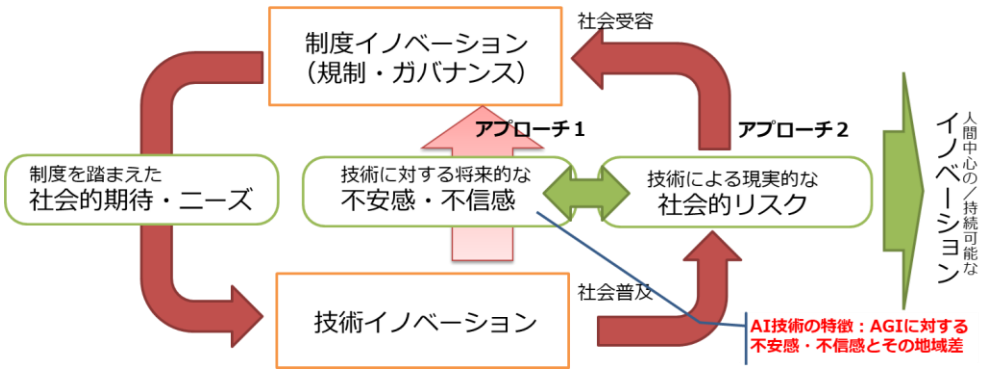


この技術・イノベーションに係る社会受容に関しては、理念的には二つの視点がある。一つは、新たな技術・イノベーションが登場した際に生じる、当該技術に対する「将来的な」不安感・不信感である。すなわち、多くの人々は、既存のルーティン・体系の中で日々の生活をしている中で、新たに登場した技術・イノベーションに対して、自らが慣れているそれの既存のルーティン・体系を破壊するものとの認識のもと、現状維持バイアス思考も含めて、新技術に対して不安感・不信感を感じることになり、場合によっては、脅威感・拒否感が示されることもある。

もう一つは、その後、技術・イノベーションの普及の進展につれて、「現実的に」生じる（予見しうる）社会的なリスクである。これは、実際に、技術・イノベーションの普及が進展するにつれ、社会が有する社会規範（倫理）に照らして現実的に社会的に望ましくないような事象が当該技術の利用によって生じ、その結果、社会としての対応の必要性が共有されるような場合である。

一般的に、新技術に係るイノベーションの進展・普及にあたっては、そのリスク低減を通じた社会受容の促進のため、新たな規制・ガバナンス制度を併せて構築していくことが必要になる（技術と制度のイノベーションの共進化）が、その制度構築においては、上述のような新技術に対して社会において共有される「将来的な」不安感・不信感にも大きく影響を受けることになる。このような不安感・不信感には、現状維持バイアスや、産業構造転換などに伴う失業などの不安などもあるが、本ワーキングペーパーでは、このうち、特に技術そのものに対する不安感・不信感を中心に、規制・ガバナンス制度構築への影響を議論の対象とする。

【図４】技術と制度のイノベーション（共進化）における不安感・不信感と社会的リスク[5]



[4] 出典：筆者作成

[5] 出典：筆者作成

＜社会受容からみた AI 技術の特徴：将来的な AGI に対する不安感・不信感とその地域差＞

このような社会受容の観点から見た人工知能（AI）技術の特徴は、非常に革新的な汎用技術であり、そのイノベーションにより社会・経済的に非常に大きな影響が想定されるだけではなく、一部の人々にとっては、将来の人類の在り方に大きな影響を及ぼし得る技術であると認識・想起されていることにある。具体的には、特に欧米を中心に、人工知能（AI）技術がこのまま発展していくと、将来的に、人工汎用知能（AGI：Artificial General Intelligence。以下 AGI という[6]）や超知性（Superintelligence）なるものが創出され、人類の文明に対する脅威になるのではないかという不安感・不信感が強く存在する[7]。シンギュラリティ（技術的特異点）などに係る議論も、その一つである。

もちろん、将来の AI 技術のあり様は、現時点で予測することは困難であり、したがって、このような AGI に対する不安感・不信感は少なくとも現時点では想像上の産物にしか過ぎないと見ることも可能である。一方で、もちろん、将来において、AGI なるものが生成され、実際に人類の文明に対して脅威となる可能性を全く否定できるものでもない。本 WP では、この将来予測の正否についての議論は行わない。

ただし、このような AGI に対する不安感・不信感に係る認識は、実際に、世界の地域社会によって共有される文化によって異なる点が特徴と言える。すなわち、AI 技術は、本来、技術的にみればデジタル技術の一種にしか過ぎないものの、特にその「人工知能」という名称から、知能を有する人間に対置する存在としての「知能を有する機械」を将来的に実現する技術であるとみなされる。このため、モノ（機械）が人間と同様の知能を有することは許されないと考える地域文化や、モノ（機械）が知能を有しても人間と共存できるという地域文化など、人間とモノ（機械）との関係に係る文化的な社会認識によって、社会における AGI に対する不安感・不信感が大きく異なることになる。

例えば、欧州に代表される西洋においては、一神教・キリスト教に見られる「神によって唯一知性を与えられた存在である人間」という文化的認識を背景に、大衆文化においても、（神に逆らって）人間による管理を超越し、人間を支配しようとする機械（AI・ロボット）という構図が広く根付いている（2001 年宇宙の旅の HAL、ターミネーターなど）。これに対し、日本を始めとする東洋においては、多神教・仏教等を始めとする「自然・人間の工作物を含め万物に生命・精神（知性）が宿る」という文化的認識を背景に、大衆文化においては、人間と協調する機械（AI・ロボット）という構図が根付いている（鉄腕アトム、ドラえもんなど）。

＜AI に対する不信感・不安感が AI 規制・ガバナンス制度に与える影響＞

このような地域において共有される宗教規範や大衆文化の差は、将来的な AGI に対する不安感・不信感に係る地域差として表れることになり、その結果、民主主義的な国家統治機構の下での各国・地域の政策立案過程を通じて、現在の各国・地域の AI 規制・ガバナンス政策に係る意思決定に大きな影響を与えることになる。例えば、市川（2021、2022）[8]によ

[6] 本来、AGI は「汎用人工知能（AI）」と翻訳されるが、本 WP では、General Purpose AI（汎用 AI：GPAI）と区別するため、仮に「人工汎用知能」と訳し、AGI として議論する。

[7] 本来は、AGI と超知性、強い AI 等とは異なる概念であるが、本 WP では、これらを特段区別せずに、このような不安感・不信感を、「（将来的な）AGI に対する不安感・不信感」として説明する。

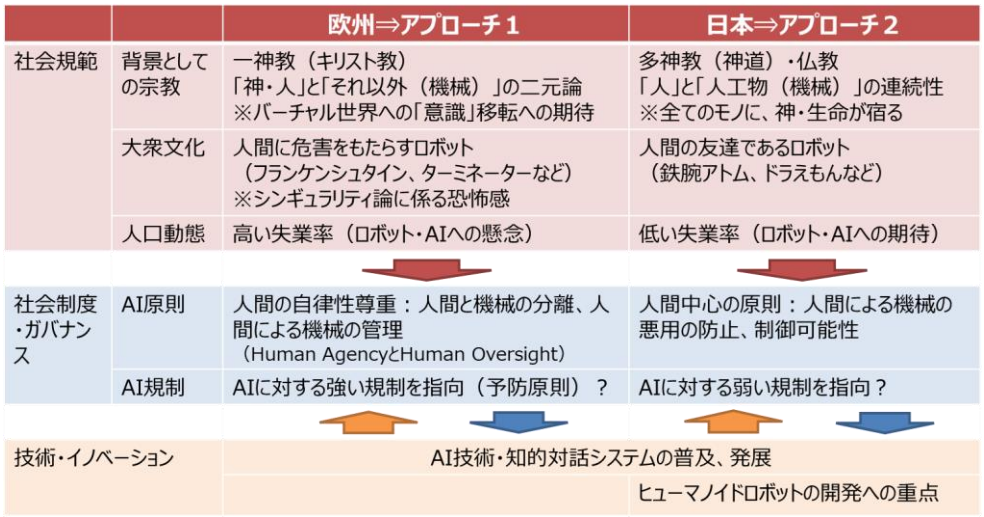
[8] 市川類「社会規範の差異が人工知能（AI）の規制・イノベーションに与える影響～欧州ＡＩ動向から見る知的対話システムの倫理的リスクに係る地域的差異」2021 年 9 月 10 日

ると、特に、「人間 - 機械」関係に係る社会規範に関する欧州と日本とでの文化上の差異により、欧州、日本それぞれの AI 原則において、異なった認識が記載されており、また、実際に欧州 AI 法案においても、この「人間 - 機械」関係に係る認識の差異に基づき法制化が検討されている項目があることを指摘されている。

具体的には、欧州を始めとする西洋においては、世論において将来的な AGI への不安感・不信感が広く共有されているため、AI ガバナンスにおいても、人間による機械（AI）に対する厳しい管理が重要視されている。すなわち、機械（AI）はそもそも本質的に社会・人類にリスクを生じさせ得る存在との認識のもと、欧州の AI 倫理ガイドライン（2019）では、人間の自律性の尊重と人間による機械の管理（Human Agency、Human Oversight）が強調され、また、欧州 AI 法案（2021）においても、予防原則（Precautional Principle）的な視点に基づく体系として、事前の幅広いリスク評価の実施、第三者による適合性認証などを義務付けるアプローチが採用されている（このように、将来的な AGI に対する不安感・不信感を中心に AI 規制・ガバナンス制度を構築するアプローチを、以下、理念的に「アプローチ１」とする）。

これに対し、日本を始めとする東洋においては、将来的な AGI に対する不信感・不安感を共有する人口は相対的に少なく、その結果、人間と機械による協調が強調されている。実際に、日本の「人間中心の AI 社会原則（2019）」では、人間による機械の管理という観点からは、技術的な視点から制御可能性を要件とする一方で、むしろ、人間による機械（AI）の悪用の防止、人間による AI への過度に依存の防止等を強調していることが特徴である。したがって、法体系としても、基本的には、まずは企業による自主的な取組を推奨し、AI が悪用されるなど現実的に社会にリスクが生じるような場合には法制化を含めてアジャイルに対応を検討するというアプローチを採用している（このように、現実的な社会的なリスクの発生を中心に AI 規制・ガバナンス制度を構築するアプローチを、以下、理念的に、アプローチ２とする）。

【図５】人工知能に係る社会規範と制度・イノベーションの関係（欧州・日本比較）[9]



市川類「欧州 AI 動向からみる知的対話システムの倫理的リスク」（2022 年 5 月 1 日、人工知能学会論文誌、37 巻（2022）3 号、p. IDS-A_1-9）

[9] 出典：市川類「社会規範の差異が人工知能（AI）の規制・イノベーションに与える影響 ～欧州 AI 動向から見る知的対話システムの倫理的リスクに係る地域的差異～」WP#21-03 （2021/09/10）
https://pubs.iir.hit-u.ac.jp/admin/ja/pdfs/show/2501

＜第三次 AI ブームに始まったこれまでの AI ガバナンス制度の動向＞

上述のような基本構造の下で、概ね 2015 年以降始まった第三次 AI ブーム以降、これまで世界各国・地域において AI 規制・ガバナンス制度の導入が検討されてきている。その際、現在までのところ、上述のような世界での地域的な文化的・社会規範的な差異をベースに、各国・地域における AI 普及率、国家の統治体制、地政学的状況を踏まえた国際的な産業戦略等の差異も加わって、アプローチ１を中心とする欧州から、アプローチ２を中心とする日本も含めて、世界各国・地域においては、多様な AI 規制・ガバナンス制度が構築されつつある[10]（図６参照）。

その中でも、AI に対して最も厳格な規制方針を打ち出しているのが、アプローチ１を採用している欧州連合（EU）である。具体的には、上述の AI（機械）に対する不信感・不安感に加え、人権に対する高い市民意識などの文化的な基盤を背景に、欧州委員会（EC）は、2021 年 4 月、AI 技術に対して水平的／分野横断的に規制を行うこととする AI 法案を公表するとともに、産業政策的な観点から当該 AI 規制に係る枠組みを国際的に展開しようと積極的に取り組んでいる。

なお、これらの世界各国における AI 規制・ガバナンス制度に係る動きは、欧州 AI 法案も含めて、まだいずれも検討段階にあり、大型の法案が最終的に成立・施行された事例は、現時点では世界の中でもほとんどない。

【図６】世界の AI ガバナンスの状況（６カ国・地域比較）[11]

Legally Binding regulations ←→ Non-Binding guidelines

		EU	Canada	US		UK	Japan	Singapore
Proposed Act Strategy Document Governance Guideline		EU AI Act (Apr 2021)	AI and Data Act (Bill C-27) (Jun 2022)	Algorithmic Accountability Act of 2022 (H.R.6580) (Feb 2022)	Blueprint for an AI Bill of Rights (October 2022)	Establishing a Pro-Innovation Approach to Regulating AI (July 2022)	AI Governance in Japan ver1.1 (July 2021) Governance Guideline for Implementation of AI Principles ver1.1 (Jan 2022)	Model AI Governance Framework ver2 (Jan 2020)
Entity in charge		European Commission (EC)	Minister of Innovation, Science and Industry	Rep. Clarke, Yvette D.	Office of Science and Technology (OSTP)	DCMS, BEIS, Office for AI	Ministry of Economy, Trade and Industry (METI)	Infocomm Media Development Authority (IMDA)/ PDPC
Strategic Directions and Points (Summary)		- EC regulation (legally binding) - Horizontal regulation for AI - Precautionary principle for AI regulatory framework - Global regulatory framework standards/norm (Brussel effect)	- regulation (legally binding) - Horizontal mostly for fairness, privacy/	- regulation by FTC (consumer area) - Horizontal only for fairness	- Vertical approach - Cross sectoral principles with no-binding (OSTP) - with vertical regulations /guideline (each agency-regulator)	- Digital regulation strategy (Jun 2021) - pro innovation approach - Start with cross sectoral principles with non-statutory guidance - Not rule out the need for legislation	- Governance innovation strategy (Jul 2020) - Legally non-binding corporate governance guideline is effective. - Legally-binding horizontal requirement deemed unnecessary at the moment.	- non-biding guidance
Definition of AI		AI system: software that - is developed with one or more of the techniques and approaches listed in Annex I and - can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with;	AI system: technological system that, - autonomously or partly autonomously, processes data related to human activities through the use of a genetic algorithm, a neural network, machine learning or another technique - in order to generate content or make decisions, recommendations or predictions."	automated decision system: any system, software, or process that uses computation, the result of which serves as a basis for a decision or judgment	automated system: any system, software, or process that - uses computation as whole or part of a system to determine outcomes, - make or aid decisions, inform policy implementation, collect data or observations, or otherwise interact with individuals and/or communities.	N/A	AI system: A system that is developed with a machine learning approach, and that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments.	AIs: set of technologies that seek to simulate human traits such as knowledge, reasoning, problem solving, perception, learning and planning, and, depending on the AI model, produce an output or decision (such as a prediction, recommendation, and/or classification)
Object of Regulation (Audience)		- Private and Governments - AI regulatory sandbox (SMEs)	- Private	- Private (>$50m +1m users or >$5m)	-Mainly Private	-Mainly Private	Private	Private
Risk Categories of Regulation	Safety	- safety component, critical infrastructure (high risk)	N/A	N/A	(Safe & effective system)		-	
	Fairness (Bias)	- education, employment, essential service (high risk) - law enforcement, migration, justice/ democracy (high risk)	(TBD by regulations)	Critical decision - education, employment, utilities, finance, healthcare, housing, legal, others	(Algorithmic discrimination protections)			
	Privacy	- social scoring (unacceptable) - remote biometrics for law enforcement (unacceptable) - remote biometrics (high risk) - emotion recognition (limited)		N/A	(Data privacy)			
	Contents	- deep fake, chat bot (limited) - general purpose AI/ generative AI (high risk)	N/A	N/A				
	Weapon	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Regulation Framework		High risk - conformity assessment - registration of AI systems - monitoring with authorities European AI Board	High Impact System - notification of AI systems - audit - order to cease AI and Data Commissioner	Critical Decision - summary report of AI system - repository Bureau of Technology (FTC)				
	Management System	High risk - risk assessment / mitigation - monitoring /documentation - transparency - human oversight	High Impact System - risk assessment/ mitigation - monitoring/ record-keeping - transparency	- impact assessment (compare with previous systems) - documentation - consult stakeholders			- Goal Setting with Conditions and risks analysis, - AI Management system (evaluation, system design, implementation)	- governance structure and human involvement - operation management and stakeholder interactions
Tools, Standards,		- AI Joint Roadmap on Evaluation and Measurement Tools with US/ NIST (Dec 2022)	- Algorithmic Impact assessment (AIA) tool for government use (2020-)		- AI Risk Management Framework (RMF) by NIST (Jan 2023)	- Roadmap to an Effective Assurance ecosystem (Dec 2021)	- Machine Learning Quality Management Guideline by AIST (Feb 2021)	- AI Verify- AI Governance Testing Framework and Toolkit (May 2022)

[10] 市川類「世界の人工知能（AI）ガバナンス制度の進化メカニズム ～技術と制度の共進化の中での地域的多様性による制度イノベーションの進展～」、IIR ワーキングペーパーWP#23-01、2023/03/08
https://pubs.iir.hit-u.ac.jp/admin/ja/pdfs/show/2576

[11] 出典：Tagui Ichikawa, ”Overview of AI Governance in Selected Countries”, 2023 年 3 月 20 日
https://drive.google.com/file/d/1TuzC-yi-C1VhLjLfL7u6mKKekyga30nB/view

（3）本 WP の問題意識と構成

＜本ワーキングペーパーの問題意識：第四次 AI ブームの AI 規制・ガバナンスの方向＞

このような現時点での世界における多様な AI 規制・ガバナンス制度は、第三次 AI ブーム以降、必ずしも当初から現在の制度に向けて紆余曲折なく検討されてきた訳ではない。

実際に、深層学習を契機とする第三次 AI ブームの初期においては、まずは、特に欧米を中心に AGI や超知性に対する不安感・不信感（恐怖感）が一時期台頭した。しかしながら、その後、これらの AGI などは長期的な将来の話であるとの理解が拡がることにより、それらに係る関心は相対的に低調になる一方、そのような AGI に対する不安感・不信感を契機にして、世界各地において AI 原則に係る議論が活発化し、2019 年には、OECD の AI 原則なども制定された。

また、並行して、AI 技術の進展と AI システムの普及に伴い、具体的な AI 利用に係る新たなルーティンの確立が進展し始めた。これにより、例えば、自動運転システムなどにおける安全性問題、あるいは、個人の生活に大きな影響を与える意思決定システムに係る公平性・人権の問題など、具体的に対象となる AI システムとそのリスク・社会的課題の内容が特定されるようになり、その結果、具体的な AI 規制・ガバナンスに係る枠組みが検討されるようになってきた。

このような中、今回の第四次 AI ブームにおいては、第三次 AI ブーム同様、再度、AGI に対する不安感・不信感が台頭しつつあるものの、今回普及が進みつつある ChatGPT 型の AI システムは、上述するような意思決定システムを中心とするこれまでの従来型の AI システムとは異なる機能を有するものであり、したがって、これまでとは異なったリスク・社会的課題を生じさせることになる。

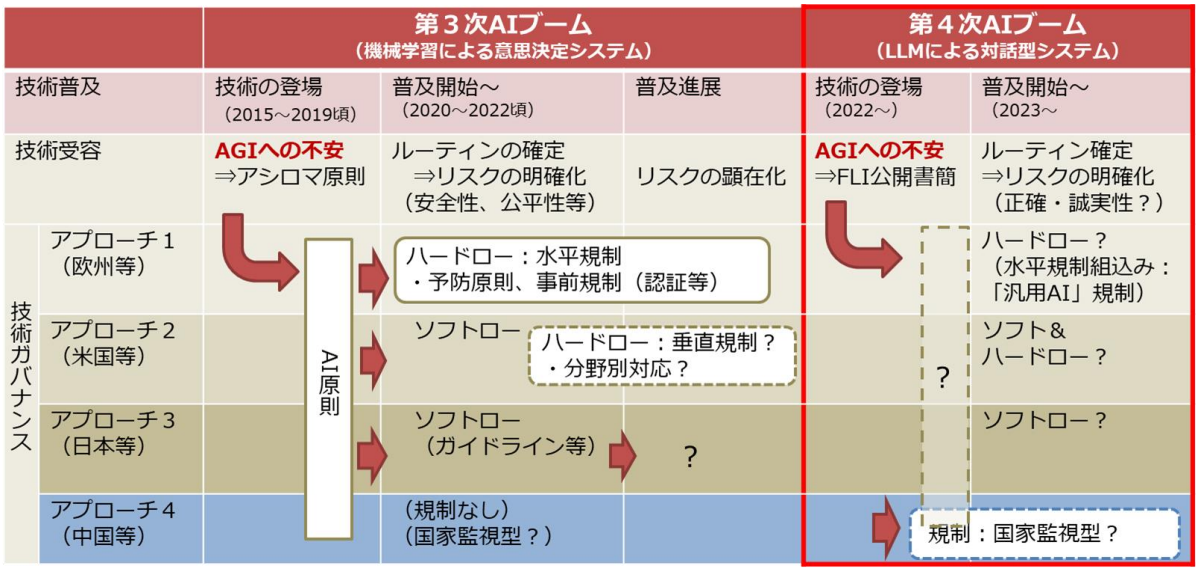
それでは、この ChatGPT 型の AI システムの技術的特徴を踏まえると、今回の第四次 AI ブームの到来に伴い、これまで検討されてきた世界の AI 規制・ガバナンス体制は、今後、どのように変化・進化していくことが見込まれるのであろうか。特に、以下のような視点からみて、今後、どのような AI 規制・ガバナンス制度の構築が見込まれるのであろうか。

- まず、この ChatGPT 型の AI システムは、これまでの従来型の AI システムと比較して、将来的にイノベーションに対してどのような影響を与え、また、どのようなリスク・社会課題を新たに生じさせるのか。
- その上で、AGI に対する不安感・不信感が再度高まる中、AI に係る主要な非営利機関による GPT の開発中断に係る提言は、AI 規制・ガバナンス制度の構築の流れの中で、どのように捉えるべきなのであろうか。
- また、今回の第四次 AI ブームにおいて、欧州、米国、日本のみならず、中国も含めた世界の AI 規制・ガバナンス制度は、現在までにどのような動きをしており、今後どのように進化していくことがみこまれるのであろうか。

このような認識の下、本ワーキングペーパー（WP）では、今回の ChatGPT に係る技術的特徴とそれがもたらしうるリスクについて考察をした上で、世界における AI 規制・ガバナンス動向の全体像及び現在地を示すことにより、今後の世界の AI ガバナンスの方向に対する示唆を得ることを目的とする（図 7 参照）。

なお、ChatGPT 型を含む、対話型・生成系 AI に関しては、現時点でも、技術・産業面・政策面それぞれにおいてまだ進行中であり、今後の動向を完全に予測することはそもそも困難であることに留意することが必要である。

【図７】本ワーキングペーパーの問題意識[12]



＜本ワーキングペーパーの構成＞

上述の問題意識を踏まえ、第二章では、まず、2015 年頃の第三次 AI ブーム初期における欧米を中心とした AGI に対する不安感、期待感の高まりが契機となって現在の第四次 AI ブームにつながっていることを示すとともに、最近の ChatGPT 型の AI システムに係る企業動向について整理する。具体的には、第三次 AI ブームにおける AGI への不安感、不信感の高まりを踏まえて設立され、その後 GPT の開発中断を提言することとなる Future of Life Institute（FLI）の設立経緯と、その後 2017 年に発表されたアシロマ原則の内容について改めて整理を行う。その上で、同じく、第三次 AI ブームでの AGI への不安感・不信感の高まりを踏まえて設立された Open AI が、ChatGPT などの LLM による生成系 AI システムの開発に取り組んだ経緯について整理をした後、ChatGPT の発表後の米国及び中国の大手 IT 企業等による ChatGPT 型 AI システムの開発に向けた取組の現状について概観する。

第三章では、そのような ChatGPT に係る技術的な特徴と現時点で想定されるリスクについて、AGI やこれまでの従来型の AI システムの比較を踏まえつつ、分析・整理を行う。具体的には、まずは、ChatGPT に係る技術的な仕組みを概観し、また、その技術的な限界を明らかにした上で、ユーザー（人間）との協調・対話によるイノベーションの可能性について考察する。また、AI システム全体における AGI、従来型の AI システム、そして ChatGPT 型の AI システムの位置づけを考察した上で、従来型の AI システムと比較した、ChatGPT 型の AI システムのリスクに係る特徴を比較整理する。その上で、ChatGTP 型の AI システムにおいて現時点で想定される具体的なリスク・社会課題について、出力内容に係る側面及び入力データに関連する側面からそれぞれ整理をする。

その上で、第四章では、世界各国・地域における ChatGPT 型の AI システムに係る規制・ガバナンス政策を巡る動向について整理・考察する。

[12] 出典：筆者作成

まず、第 4−1 章では、ChatGPT の発表後、FLI が 2023 年 3 月に発表した、GPT の開発中断を求める提言に係る動きについて整理をする。具体的には、まず、本書簡のインパクトを整理した上で、この動きは、主として欧米を中心とする AGI への不安感・不信感を背景にしたものであることを整理する。その上で、この開発中断という政策提案に係る妥当性について議論するとともに、この提案に対する各界の反応の多くはポジショントークであることについて考察する。

次に、第 4−2 章では、特に AI 規制に積極的に取り組んでいる欧州を取り上げ、その AI 法案における ChatGPT 型 AI システムに係る対応の検討の動きについて整理する。まずは、2021 年 4 月に欧州委員会（EC）が発表した AI 法案に対して、FLI や他の非営利団体等の提案により新たに盛り込まれた「汎用 AI システム」を巡る動きについて整理する。その上で、2022 年 11 月末の ChatGPT の発表以降における欧州議会や各国・関連機関における議論等の動向に関して、汎用 AI システムの位置づけの見直し、生成系 AI に対する著作権に係る透明化義務などの検討の動きについて考察する。

その上で、第 4-3 章では、ChatGPT 発表以降における、米国、日本、中国等の政策動向について整理する。具体的には、米国・英国では、まだこれから規制の在り方の検討を開始する段階であるのに対し、日本は、原則規制をしない方向を明確にしたことが特徴であることを示す。さらに、ChatGPT に係る規制としては、中国は、その国家体制の安定確保の観点から、厳格な規制方針を明らかにした点が、特徴であることを示す。

第五章では、このような動きを踏まえ、今後の世界の AI 規制・ガバナンスの方向としては、①今後とも、先進民主主義国間では、各国・地域の事情に応じ、引き続き多様な制度の進化が見込まれること、②一方、従来の意思決定系の AI システムでは個人情報、公平性・人権が重要であったのに対し、今回の生成系の AI システムにおいてはむしろ著作権、民主主義などのキーワードが重要になる可能性があること、③世界的な観点からは、中国の独自規制による中国市場の孤立化の可能性があることを示す。その上で、2023 年 4 月末に発表された G7 の閣僚宣言も踏まえつつ、今後の世界の AI 規制・ガバナンス制度の方向について、論点を整理する。

２．第四次ブームまでの経緯と ChatGPT 等を巡る企業動向

（１）第三次 AI ブーム／FLI の創設とアシロマ原則

＜第三次 AI ブームでの AGI への関心の高まりとその後の経緯＞

2015 年頃、深層学習（ディープラーニング：DL）技術によるブレークスルーを契機として、世界的に多くの人々が人工知能（AI）技術に対して驚異を感じ、それがきっかけとなって AI に対する関心が急激に高まり、第三次 AI ブームが始まった。

その第三次 AI ブームの当初においては、それまでのシンギュラリティ（技術的特異点）の議論[13]などを背景にしつつ、人工汎用知能（Artificial General Intelligence：AGI）が近い将来実現されるのではないかという不安感・不信感が、特に欧米社会を中心に急激に拡がった。なお、この AGI の定義には、必ずしも明確なものはないが、一般的には、人間が行うことができるあらゆる（汎用の）知的作業を理解・学習・実行することができるような人工知能を指すものとされる[14]。AGI は、本来は中立的な用語であるものの、前章で記載した通り、欧米中心に、 人間と同等の（汎用の）知能を持つとされる AGI が実現すると、人類社会にとって脅威となるのではないかという不安感・不信感が存在する。

このような社会における不安感・不信感を背景に、世界的に「AI 原則」の議論が開始された[15]。その際、実際に、世界各国の政府機関で AI 原則の必要性を初めに唱えたのは日本の総務省（2015 年 1 月）であり、その後、2016 年 4 月の G7 香川・高松情報通信大臣会合において AI 原則案を提示したことを契機に、その検討が世界的に拡大されていくことになる。なお、総務省の当時（当初）の問題意識としても、シンギュラリティを含め、超長期的な（SF のような）世界を想定したものとなっており、また、実際に、当初は、SF 作家でもある Issac Asimov のロボット三原則のようなものを想定していた[16]。ただし、その後シンギュラリティや AGI の実現については、将来的な可能性は（あったとしても）否定できないもの、現時点の技術の延長では見込まれず、いずれにせよ長期的な話であるという認識の共有がその後進んだことから、AGI に対する社会的な関心は相対的に低下することになる。

一方で、深層学習（DL）を始めとする機械学習技術に伴って生じる新たな社会倫理・リスク課題に関しては、その普及が進展するにつれ、具体的な議論が進展し、次第に明確化されるようになる。具体的には、当初は、従来のデジタル技術でも課題であった、プライバシ

---

[13] 例えば、米国の未来学者であるレイ・カーツワイル氏は、2005 年の著書『ポスト・ヒューマン誕生 コンピュータが人類の知性を超えるとき（The Singularity Is Near: When Humans Transcend Biology』において、「人工知能技術が、特に汎用人工知能〔artificial general intelligence, AGI〕に発展し、「自己再生産」によって指数関数的に高度化することにより人間の知能を抜き去り「超知能」が出現する時点（2045 年頃）を「シンギュラリティ」と位置づけている。

[14] なお、類似の言葉として、（「弱い AI」に対して）「強い AI」などもある。

[15] 中川 裕志（理化学研究所）「AI 倫理指針の動向とパーソナル AI エージェント」総務省 学術雑誌『情報通信政策研究』 第３巻第２号（2020 年 3 月 30 日）
https://www.soumu.go.jp/main_content/000679318.pdf
https://www.soumu.go.jp/iicp/journal/journal_03-02

[16] 市川類「AI 原則の体系化と今後のガバナンスの方向～デジタル・AI におけるイノベーションと社会制度の共進化」IIR ワーキングペーパー＃WP20-15, 2020 年 10 月 2 日
https://pubs.iir.hit-u.ac.jp/admin/ja/pdfs/show/2432

ー、セキュリティなどに加えて、その特有のリスク課題として、当初は、自動運転のトロッコ問題、殺人ロボット兵器、自動ボットによる人間なりすましなどが取りあげられたが、その後、AI システムの社会的普及による新たなルーティンの明確化に伴い、むしろ自動運転技術の安全性の確保の課題や、自動意思決定システムに係る公平性の課題などが大きな社会的課題として位置づけられるようになる。このように認識の変化が進む中、日本や欧州で策定された世界の AI 原則を取り入れる形で、2019 年 4 月に OECD の AI 原則が策定・発表され、その後は、それらの AI 原則の「実践」に向けた AI ガバナンス制度の構築に係る検討の動きが世界において進展することになる。

なお、この第三次 AI ブームの初期段階における AGI への関心への高まりの中で、特に欧米においては、人工知能（AI）、とりわけ AGI などの在り方について議論を行うための多くの非営利機関が創設されている。このうち、以下、本章においては、今般の ChatGPT 型の AI システムへの社会的な関心の高まりに大きく影響を与えることになる FLI（Future of Life Institute）と OpenAI について、その経緯を振り返る。

＜FLI（Future of Life Institute）の創設＞

Future of Life Institute（FLI：生命の未来研究所）[17]は、2014 年 5 月に、人類が直面する世界的な破滅的なリスクを削減することを目的に設立された非営利機関である。同機関の創業者は、MIT の物理学者／機械学習研究者の Max Tegmark 氏[18]や Skype の共同創業者である Jaan Tallinn 氏などであるが、イーロン・マスク氏が多額の寄付をするとともに、物理学者のスティーブン・ホーキング博士、未来学者のレイ・カーツワイル博士などが支持者として参加していることで有名である。特に、イーロン・マスク氏は、現在でも、同機関の外部アドバイザーとして名を連ねている。本社は、米国マサチューセッツ州ケンブリッジであり、同社は、特に、米国、欧州、国連の３方面での政策に対して、政策提言を行っている。

同機関は、世界的な破滅的なリスクとして、具体的には、人工知能（AI）、バイオ技術、核兵器、気候変動の４つのテーマを取りあげ、活動を進めているが、その４つのテーマの中でも、特に人工知能（AI）技術に係る社会的なリスクへの対応を中心的な課題として位置づけている。具体的には、実際に活動が開始された 2015 年以降、AI の未来に係る会議を開催するとともに、同年 10 月には、人工知能（AI）に係る研究優先事項に係る公開書簡を発表するなどの活動を行っている。本公開書簡では、AI 研究にあたって、短期的には、経済的インパクトの最適化、法・倫理に係る研究の推進、頑強な AI に係る研究を進める一方で、長期的には（過去の核研究を念頭に）、検証、セキュリティ、管理に係る取組の必要性を指摘している。

＜アシロマ AI 原則と AGI の位置づけ＞

そのような中、FLI は、2017 年 1 月に、米国カリフォルニア州のアシロマにて会議を開催し、翌月アシロマ AI 原則を発表した[19]。このアシロマ AI 原則は、他の AI 原則と比較して

[17] Future of Life Institute
https://futureoflife.org/
https://futureoflife.org/about-us/

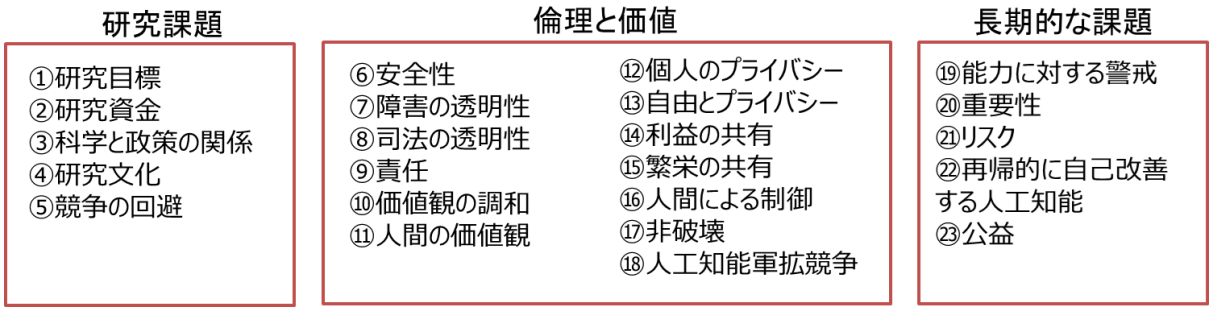
[18] 同氏は、2017 年 8 月に、「Life 3.0: Being Human in the Age of Artificial Intelligence」（日本版は「LIFE3.0──人工知能時代に人間であるということ」 2019 年 12 月）との本を発表しており、FLI の創設経緯や AGI に対する同氏の考え方などを記載している。

[19] A Principled AI Discussion in Asilomar

も、上記総務省に次いで、かなり早期の段階に発表されたものである[20]。なお、この「アシロマ」は、1975 年に遺伝子組み換え技術に関するガイドラインが議論され、生物学的封じ込めの合意がなされた「アシロマ会議（Asilomar conference）」の場所として有名であり、それに肖って開催されたものである

同原則は、23 の原則（研究課題 5 原則、倫理・価値 13 原則、長期的課題 5 原則）からなる。うち、「倫理・価値」に係る項目は、安全性／障害の透明性／司法の透明性／責任／価値観の調和／人間の価値観／個人のプライバシー／自由とプライバシー／利益の共有／繁栄の共有／人間による制御／非破壊／人工知能軍拡競争からなる（図８参照）。

【図８】アシロマ AI 原則（FLI：2017 年）の概要（項目）[21]



このアシロマ AI 原則[22]は、その後発表される他の AI 原則との比較による特徴の一つとして、「長期的な課題」として、将来的な AGI を念頭においた原則を記載したことである[23]。具体的には、例えば、「19. 将来の人工知能が持ちうる能力の上限について強い仮定をおくことは避けるべき」「20. 高度な人工知能は、地球上の生命の歴史に重大な変化をもたらす可能性がある」「22. 再帰的に自己改善もしくは自己複製を行える人工知能システムは、安

---

https://futureoflife.org/principles/principled-ai-discussion-asilomar/
The Asilomar AI Principles,
https://futureoflife.org/open-letter/ai-principles/

[20] なお、これは、世界的にみると、総務省が、2016 年 4 月の G7 香川・高松情報通信大臣会合で発表した、国際的に参照される枠組みとしての「AI 研究開発に係る８原則」に次いで、早期の AI 原則であると言える。
市川類「AI 原則の体系化と今後のガバナンスの方向 ～デジタル・AI におけるイノベーションと社会制度の共進化～」IIR ワーキングペーパーWP#20-15、2020/10/02
https://pubs.iir.hit-u.ac.jp/admin/ja/pdfs/show/2432

[21] 市川類「AI 原則の体系化と今後のガバナンスの方向~デジタル・AI におけるイノベーションと社会制度の共進化~」パワポ版（2020 年 10 月 16 日）
https://drive.google.com/file/d/1ZP3pb_A-PUf7g67XHNCR7OrbM1GZ3TLO/view

[22] アシロマの原則（日本語版）August 2, 2017
https://futureoflife.org/open-letter/ai-principles-japanese/

[23] 19) 能力に対する警戒： コンセンサスが存在しない以上、将来の人工知能が持ちうる能力の上限について強い仮定をおくことは避けるべきである。
20) 重要性：高度な人工知能は、地球上の生命の歴史に重大な変化をもたらす可能性があるため、相応の配慮や資源によって計画され、管理されるべきである。
21) リスク： 人工知能システムによって人類を壊滅もしくは絶滅させうるリスクに対しては、夫々の影響の程度に応じたリスク緩和の努力を計画的に行う必要がある。
22) 再帰的に自己改善する人工知能：再帰的に自己改善もしくは自己複製を行える人工知能システムは、進歩や増殖が急進しうるため、安全管理を厳格化すべきである。
23) 公益：広く共有される倫理的理想のため、および、特定の組織ではなく全人類の利益のために超知能は開発されるべきである。

全管理を厳格化すべき」などであり、AGI の中でも、特に「再帰的に自己改善／自己複製を行う AI」に対する危険認識・恐怖感を共有している。なお、もう一つの特徴としては、AI の軍事利用にも焦点に当てたことがあげられ、実際に、FLI はその後、2016 年 2 月、2018 年 6 月には、致死型自律兵器（LAWS）に関する公開書簡をそれぞれ発表している[24]。

このようにアシロマ AI 原則では、「長期的課題」とは位置づけられたものの、他の AI 原則とは異なって、当時における欧米を中心とした AGI に対する不安感・不信感を明記している点が特徴であり、それがゆえに、FLI は、AI に係る現実的なリスク課題・懸念よりも、想像上の終末論的シナリオに重きを置いていると言われることもある[25]。一方で、その後、世界的に AGI への関心が相対的に低下する中で、FLI は 2018 年 6 月以降 AI に係る公開書簡を発表していなかったが、第 4-1 章で示すとおり、2023 年 3 月になって、およそ 5 年ぶりに、新たに公開書簡を発表することになる。

（２）OpenAI 社の創設とこれまでの経緯

＜OpenAI 社の創設＞

第三次 AI ブームにより、AGI への関心が高まる中、上述の FLI は、AGI による将来的な懸念に関して、そのリスク削減のための政策提言活動を行う非営利機関として設立されたのに対し、むしろ長期的な観点からむしろ人類に利益をもたらすような AGI を開発することを目的とした非営利機関として、OpenAI 社が設立された。いずれも、イーロン・マスク氏が積極的に支援している点が共通点である。

Open AI 社は、2015 年末に設立された人工知能（AI）に係る研究所（サンフランシスコ本社）であり、サム・アルトマン氏（元 Y コンビネータ代表）、イーロン・マスク氏、ピーター・ティール氏（元 Paypal CEO など。起業家）ほかにより、10 億ドルの資金提供約束のもとで設立された。サム・アルトマン氏が CEO を務め、現在は、OpenAI Inc（非営利法人）とその子会社である OpenAI LP（営利法人）から構成される。

OpenAI 社のもともとの設立趣旨は、当時、欧米を中心に、AGI、超知性など人工知能の発展による人類の破滅などの悪影響が懸念される中、「人類全体に利益をもたらす形で友好的な AI を普及・発展させること」を目標にし、このため長期的な観点から研究ができるように、当初は非営利法人形態として設立されている。

＜イーロン・マスク氏の退出とマイクロソフト社からの出資受け入れ＞

---

[24] Lethal Autonomous Weapons Pledge, 2018 年 6 月 6 日
https://futureoflife.org/open-letter/lethal-autonomous-weapons-pledge/
AI Companies, Researchers, Engineers, Scientists, Entrepreneurs, and Others Sign Pledge Promising Not to Develop Lethal Autonomous Weapons, 2018 年 7 月 18 日
https://futureoflife.org/fli-projects/ai-companies-researchers-engineers-scientists-entrepreneurs-and-others-sign-pledge-promising-not-to-develop-lethal-autonomous-weapons/
Gigazine「「自律的殺人兵器」への AI 活用を禁じる誓約書にイーロン・マスクや DeepMind 創業者、著名な AI 研究者など 2500 人以上が署名」2018 年 07 月 19 日
https://gigazine.net/news/20180719-lethal-autonomous-weapons-pledge/

[25] ロイター「ＡＩ専門家ら、マスク氏らの公開書簡に懸念」2023 年 4 月 1 日
https://jp.reuters.com/article/elon-musk-ai-academics-idJPKBN2VX1M5

OpenAI 社は、2016 年には強化学習研究のためのプラットフォームである OpenAI Gym などを発表するなど研究を推進したものの、その研究規模は、Google の DeepMind 社などによる AI 研究と比較してかなり劣っていた。

このような中、イーロン・マスク氏は、テスラ社との利益相反の関係から、2018 年 2 月に役員を退任するとともに、その後の追加投資を行わなかった。なお、この利益相反というのは形式上の理由であって、マスク氏は「このままでは Google に完全に負ける」と焦りだし、OpenAI を自ら買収し、陣頭指揮を執ろうと提案したものの、共同創業者のサム・アルトマン CEO を含む経営陣がこれに強く反発し、確執が表面化したことにより、OpenAI の取締役を辞任することになったとの指摘もある[26]。いずれにせよ、その後、実際に、Google は、2018 年 10 月、同社の大規模言語モデル（LLM）である BERT（Bidirectional Encoder Representations from Transformers）を発表し、AI 専門家の中ではブレークスルーをもらしたとして関心を高めることになり、この時点では、LLM 開発競争においては、Google がリードすることになる。

そのような中、資金不足に陥った OpenAI 社は、2019 年 3 月に、制限付きの営利部門である OpenAI LP を設立して、企業からのベンチャー投資を受け入れることとし、具体的には、2019 年 7 月にマイクロソフトから 10 億ドルの出資を受け入れ、その資金を活用しつつ、大規模言語モデルである GPT（Generative Pre-trained Transformer）の開発に積極的に取り組んだ。具体的には、2018 年 6 月に発表した GPT-1 に引き続き、2019 年 2 月には GPT-2 を発表するとともに、2020 年 5 月にはその後継モデルである GPT-3 を発表している。なお、この GPT-3 には 1,750 億個のパラメータが含まれ、パラメータ数 15 億個の GPT-2 と比較して、2 桁大きいとされる。また、2021 年 1 月には、自然言語処理と画像生成モデルを組み合わせた AI である、DALL-E を発表するとともに、2022 年 4 月には、拡散モデル導入によりアップグレードした DALL-E2 を発表し、世間の注目を浴びている。

その上で、2022 年 11 月末に GPT3.5 に基づいた ChatGPT を発表し、世界的に大反響をもたらす中で、2023 年 1 月 23 日には、マイクロソフト社から新たに 100 億ドルの投資を受けたと報道されている[27]。また、2023 年 3 月 14 日には、最新版の LLM である GPT-4 を発表している[28]。

（3）ChatGPT の登場と米国・中国企業の参入動向

＜ChatGPT の登場とその爆発的な普及＞

2022 年中ごろから、特に専門家や革新的利用者の間において、上述の DALL-E2 に加え、Stable Diffusion（2022 年 8 月頃流行）、Midjouney（2022 年 8 月公開）などの画像に係るコンテンツを自動生成するいわゆる生成系 AI への関心が高まっていた。

---

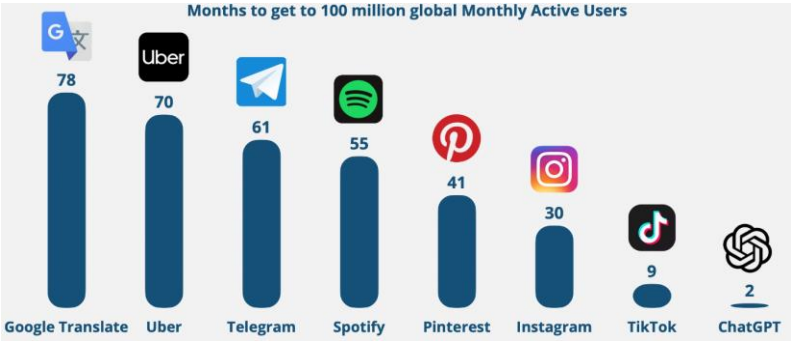
[26] Yahoo Japan ニュース（Gizmode）「イーロン・マスクが OpenAI を辞めた本当の理由」2023 年 4 月 4 日
https://news.yahoo.co.jp/articles/a95a092b6359a3d2bb9959d3892ebc0432c8b02d

[27] 日本経済新聞「Microsoft、ChatGPT のオープン AI 追加投資　数十億ドル」2023 年 1 月 24 日
https://www.nikkei.com/article/DGXZQOGN23BXC0T20C23A1000000/

[28] IT Media News「「GPT-4」発表　日本語でも ChatGPT 英語版より高性能、司法試験で上位 10%、「この画像何が面白いの？」にも回答」 2023 年 03 月 15 日
https://www.itmedia.co.jp/news/articles/2303/15/news092.html

このように生成系 AI への関心が高まる中、OpenAI 社が 2022 年 11 月 30 日に発表した ChatGPT は、言語に係る対話型の生成系 AI として、世界的に大きなインパクトを与えることになる。具体的には、その公開直後から、その内容・エクスペリエンスが凄いと評判になり、公開後 5 日で 100 万ユーザーを突破するとともに、2 カ月後の 2023 年 1 月末には、月間アクティブユーザーが 1 億人に達したと報じられている。これは、これまでの他のネットサービスと比較しても圧倒的に速度が早いとされる（図 9 参照）。

【図 9】各種サービス別の 1 億人ユーザー（MAU）に掛かった月数[29]



＜ChatGPT 型 AI システムに係る米国大手 IT 企業の動向＞

このような中、特に米国の大手 IT・ネット系企業が、続々と LLM を利用した ChatGPT 型の AI システムの開発に向けた取組を発表している。

まずは、2020 年に OpenAI 社と独占契約を締結している Microsoft 社は、上述の通り、2023 年 1 月に、OpenAI に対して、今後 100 億ドル（約 1.3 兆円）の投資を行うことを決めたと報道されている。また、2023 年 2 月 7 日には、ChatGPT を搭載した検索エンジン Bing を発表[30]するとともに、2023 年 3 月 10 には、Azure OpenAI Service において ChatGPT が利用可能になったことを発表する[31]など、同社製品への ChatGPT の導入を加速している。なお、2023 年 5 月 4 日には、より最新の対話型 A）を搭載した検索エンジン Bing の一般公開を発表している[32]。

このような動きに危機感を持ったのが、検索分野で圧倒的に優位を有する Google である。もともと、Google は、大規模言語モデル（LLM）に係る研究では、BERT の開発により優位に立っているとされていたが、Microsoft が全面的に支援する OpenAI による LLM としての GPT シリーズ及びそれを利用した ChatGPT の成功に危機感を感じ、2022 年 12 月には、

---

[29] 出典：App Economy Insights2023 年 2 月 5 日
https://twitter.com/EconomyApp/status/1622029832099082241

[30] Microsoft, “Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web”, Feb 7, 2023 | Yusuf Mehdi, Corporate Vice President & Consumer Chief Marketing Officer
https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/
Impress Watch（臼田勤哉）「Microsoft、AI による新たな検索「Bing」発表 「ChatGPT より有能」2023 年 2 月 8 日
https://www.watch.impress.co.jp/docs/news/1476776.html

[31] Microsoft Japan News Center, 「ChatGPT が Azure OpenAI Service で利用可能に」2023 年 3 月 10 日
https://news.microsoft.com/ja-jp/2023/03/10/230310-chatgpt-is-now-available-in-azure-openai-service/

[32] 日本経済新聞「Microsoft、対話型 AI 検索を公開　ChatGPT より最新情報」2023 年 5 月 4 日
https://www.nikkei.com/article/DGXZQOGN0338S0T00C23A5000000/

「コードレッド」（厳戒警戒）を宣言したことが報じられている[33]。その上で、2023 年 2 月 6 日には対話型 AI システムである Bard の限定公開を開始し[34]、その後、3 月 21 日には、英米において一般公開を開始している[35]（4 月 19 日には、日本でも公開[36]）。

また、Microsoft、Google 以外にも、大手 IT 企業による参入発表も相次いでいる。具体的には、Amazon は、2023 年 4 月 16 日、ChatGPT のようなテキスト生成系 AI である「Amazon Titan」や画像生成 AI である Stable Diffusion にアクセスできるサービス「Amazon Bedrock」を発表し、限定提供を開始している[37]。また、Facebook（Meta 社）は、2023 年 2 月 24 日、大規模言語モデルである「LLaMA（Large Language Model Meta AI）を、研究者向けへの提供開始を発表している[38]。さらに、Twitter（X 社）の CEO であるイーロン・マスク氏は、2023 年 4 月 17 日に公開されたインタビューで「トゥルース GPT」を開発する予定だと語っている[39]。

＜ChatGPT 型 AI システム開発に係る中国企業の動向＞

一方、米国に次いで AI 大国である中国においても、ChatGPT への関心は非常に高く、ChatGPT 類似の取組みが過熱する一方、中国独自の検閲との関係で、独自のモデルを構築しようとする動きがある。

もともと、OpenAI 社の ChatGPT は、中国語にも対応しているが、中国からは ChatGPT にはアクセスできないようになっている[40]。これは、ChatGPT は、中国共産党政権にとって望ましくない回答を返す可能性があり、したがって予め自主規制をしているものと考えられる。ただし、中国からでも VPN を通じれば ChatGPT へのアクセスが可能であり、このため、中国の一部新興企業等が、大手 IT 企業のプラットフォーム上で、ChatGPT を利用したサービスを提供していたとされる。このような動きに対し、中国政府は、2023 年 2 月、欧米発の虚偽情報をまき散らすものとして、中国の大手 IT 企業などに対して、ChatGPT にアクセスさせないよう指示を出したと報じられている[41]。Ant Group が ChatGPT サービス

---

[33] Gigazine「ChatGPT のリリースで Google は「コードレッド」を宣言、AI チャットボットが検索ビジネスにもたらす脅威に対応するためにチームを再割り当て」2022 年 12 月 23 日
https://gigazine.net/news/20221223-google-code-red-against-chatgpt/
[34] ITmedia NEWS「Google、OpenAI の「ChatGPT」競合「Bard」を限定公開」2023 年 02 月 07 日
https://www.itmedia.co.jp/news/articles/2302/07/news081.html
[35] ITmedia NEWS「ChatGPT 対抗、Google の対話 AI「Bard」一般公開　「Google 検索の補完」将来は検索との統合も」2023 年 03 月 22 日
https://www.itmedia.co.jp/news/articles/2303/22/news089.html
[36] ASCII AI「グーグル「Bard」ついに日本公開　「ChatGPT」対抗の AI チャット」2023 年 04 月 19 日
https://ascii.jp/elem/000/004/133/4133489/
[37] Yahoo Japan ニュース（Impress Watch）「Amazon、ChatGPT のようなテキスト生成 AI「Titan」」4/17(月)
https://news.yahoo.co.jp/articles/42f2b5ba2be793195e8ae4667f0a8751e5245da8
[38] Impress Wtach（臼田勤哉）「Meta、新たな大規模言語モデル「LLaMA」」2023 年 2 月 27 日
https://www.watch.impress.co.jp/docs/news/1481436.html
[39] Yahoo Japan ニュース「マスク氏、「トゥルースＧＰＴ」の立ち上げ表明　チャットＧＰＴに対抗＝報道」4/18(火) 7:09 配信
https://news.yahoo.co.jp/articles/827befc55e5a6a26c2fa923e189b04c567b15874
[40] Yahoo Japan ニュース（Newsweek）「中国からはアクセス不可に...共産党政権が ChatGPT を恐れる理由と、中国発 AI の脅威」（2023 年 3 月 24 日）
https://news.yahoo.co.jp/articles/7168b74f284390304c4b875c198c285660e2eb1e
[41] クラウド Watch「中国でも ChatGPT の熱狂　米国との差にあせりも」2023 年 3 月 6 日

へのアクセスを直接的にも間接的にも提供しないよう指示されたほか、Tencent も当局の圧力で ChatGPT 型のサードパーティのサービスをいくつか停止し、これによって少なくとも数十種類の ChatGPT サービスが閉鎖されたと報道されている[42]。

一方で、中国国内では、ChatGPT への関心が加熱しており、中国の大手 IT 企業などは、独自の ChatGPT 類似サービスの提供に向けた開発競争になっている。具体的には、Baidu は、2023 年 2 月 7 日、中国版 ChatGPT である「文心一言（アーニーボット）」を翌月に一般公開すると発表した上で、実際に、2023 年 3 月 16 日に同サービスを発表した[43]。なお、この文心一言（アーニーボット）については、検閲との調整の難しさが指摘される[44]一方、総合的には ChatGPT には及ばないものの、中国語対応では高い評価を獲得しているとの見方もある[45]。

また、Alibaba グループは、2023 年 2 月 8 日、AI ツールを内部テスト中であると公表した上で、4 月 11 日には、大規模言語モデル「通義千問」を発表するとともに、近い将来に同社のメッセージアプリやスマートスピーカーなどに組み込むことを発表している[46]。なお、テンセントは、2023 年 2 月 27 日、ChatGPT のようなチャットボットの開発チームを立ち上げたことが報じられている[47]。

その他にも、画像認識システムの中国大手の商湯集団（センスタイム）は、2023 年 4 月 10 日、対話型 AI サービスに係る政府や企業などのユーザー向けに体験版を公開した[48]。また、JD クラウド[49]は、同社の AI プラットフォーム「Yanxi」をベースにした産業向けの ChatGPT ライクな「ChatJD」を発売すると発表した。なお、研究機関では、上海市の復旦

---

https://cloud.watch.impress.co.jp/docs/column/infostand/1483435.html

[42] 日本経済新聞「中国、ChatGPT の利用停止　アリババやテンセントに指示」2023 年 2 月 22 日
https://www.nikkei.com/article/DGXZQOUA226Z40S3A220C2000000/
読売新聞オンライン「中国が対話型ＡＩを警戒、「ＣｈａｔＧＰＴ」は使用停止に…政府見解と異なる回答で」2023/03/04
https://www.yomiuri.co.jp/world/20230304-OYT1T50154/

[43] IT Media NEWS「Baidu、中国版「ChatGPT」を発表　AI も米中競争激化」2023 年 3 月 17 日
https://www.itmedia.co.jp/news/articles/2303/17/news106.html

[44] WIRED「ChatGPT に対抗するバイドゥの会話型 AI は、「検閲」という課題に直面している」2023 年 3 月 27 日
https://wired.jp/article/chinas-answer-to-chatgpt-flubs-its-first-lines/
Forbes Japan「バイドゥの「中国版 ChatGPT」は期待外れ、株価 10%急落」2023 年 3 月 17 日
https://forbesjapan.com/articles/detail/61729

[45] 東洋経済オンライン（田中 信彦）「中国版 ChatGPT「文心一言」その能力と可能性は？百度（バイドゥ）の AI は中国語対応に強み」2023/04/04
https://toyokeizai.net/articles/-/662951

[46]Bloomberg「 アリババ、ＡＩツールをスマートスピーカーに統合へ－ChatGPT に対抗」2023 年 4 月 11 日
https://www.bloomberg.co.jp/news/articles/2023-04-11/RSXHX5DWLU6801
Yahoo Japan ニュース（高口康太）「「アリババ版 ChatGPT、ビジネスにすぐ使えるその実力とは 4/11(火) 16:03
https://news.yahoo.co.jp/byline/takaguchikota/20230411-00345181
Yahoo Japan ニュース（ロイター）「アリババが「通義千問」発表、ＧＰＴに似たＡＩ大規模言語モデル」4/11(火) 12:46
https://news.yahoo.co.jp/articles/413776606c37657bf7e465ef75f4c18358ba690a

[47] ロイター「中国テンセント、ＣｈａｔＧＰＴ的製品の開発チーム設立＝関係筋」2023 年 2 月 27 日
https://jp.reuters.com/article/tencent-chatgpt-idJPKBN2V10H0

[48] 日本経済新聞「中国センスタイムが対話型 AI ChatGPT に対抗」2023 年 4 月 11 日
https://www.nikkei.com/article/DGXZQOGM112NJ0R10C23A4000000/

[49] EC プラットフォーム JD.com のクラウドコンピューティング部門。

大学のグループが、ChatGPT のようなチャットボットを構築できるという大規模言語モデル「MOSS」を、2 月 22 日に Web 上で公開している。

【図１０】ChatGPT 型 AI システムの開発・サービス提供に係る企業動向[50]

＜米国企業＞

企業名	動向
Microsoft	• 23年1月、OpenAIへの1,000億ドルの投資（報道） • 23年2月、ChatGPTを搭載したBingを発表 • 23年3月、Azure OpenAI ServiceでChatGPTを提供
Google (Alphabet)	• 22年12月、ChatGPTに関し「コードレッド」を宣言（報道） • 23年2月、対話型サービスBardを限定公開、3月、正式公開（英米）
Amazon	• 23 年4月、Amazon Titan等を発表、限定提供開始。
Facebook (Meta)	• 23年2月、大規模言語モデルLLaMAの研究者向け提供開始。
Twitter (X)	• 23年4月、TruthGPTの開発を発表。

＜中国企業＞

企業名	動向
Baidu	• 23年3月、中国版ChatGPTである「文心一言（アーニーボット）」を公開。
Alibaba	• 23年4月、大規模言語モデル「通義千問」を発表。今後各種プリに搭載。
Tencent	• 23年2月、開発チーム立ち上げ（報道）
Sensetime	• 23年4月、ユーザー向け体験版を公開。

＜ChatGPT 型 AI システムを巡る世界的な産業競争構造＞

このような ChatGPT 型の AI システムを構築するには、大規模言語モデル（LLM）が必要となるが、その構築には、計算資源を含め非常に多くのコストが必要になる。このため、上述のとおり、ChatGPT 型の AI システムの開発は、世界の大手 IT ネット系企業を中心とする競争になりつつあり、したがって、産業構造・技術競争的な観点からは、少なくとも当面は、現在の各種インターネットサービスと同様、米国及び中国の巨大 IT ネット系企業による寡占状況になることが見込まれる。その結果、現在の GAFA などの巨大 IT ネット系企業へのデータ・資源と権限の集中が、ChatGPT のような高度な AI の登場により、より加速化することを懸念する指摘もある[51]。

また、このような産業構造の見込みが、欧州、日本も含め、世界各国・地域における AI 規制・ガバナンスに係る今後の国家戦略にも影響を与えることが想定される。具体的には、自国産業の競争力強化のための緩やかな規制制度を志向したり（米国、英国など）、海外企業による参入に対して規制強化を志向したりする方向（欧州など）などが考えられる。

なお、日本では、日本版 LLM の必要性が議論され始める中、2023 年 5 月ごろから、企業による LLM 参入の発表がなされている。具体的には、ソフトバンクは、2023 年 5 月 10 日の決算会見において、LINE と共同で和製 GPT の立ち上げを進めていることを明らかにし[52]、また、翌日の 11 日には、サイバーエージェントが独自の日本語 LLM を発表している[53]。

---

[50] 出典：各種資料より筆者作成

[51] 朝日新聞デジタル（渡辺淳基）「「データと権限の集中」懸念　元グーグル社員、ChatGPT 登場で」2023 年 4 月 6 日 8 時 00 分
https://www.asahi.com/articles/ASR4474PJR43ULFA035.html

[52] ITMedia News「ソフトバンク、LINE と和製 GPT 立ち上げへ　「やらなければ今後の参加権がなくなる」」2023 年 05 月 10 日
https://www.itmedia.co.jp/news/articles/2305/10/news170.html

[53] ITmedia NEWS 「“和製 GPT”競争勃発か　サイバーエージェント、独自の日本語 LLM 発表　「活用を始めている」」2023 年 05 月 11 日
https://www.itmedia.co.jp/news/articles/2305/11/news206.html

３．ChatGPT の技術的特徴と社会的リスク

（１）ChatGPT の仕組み・限界とイノベーションへの貢献の可能性

①ChatGPT の技術的仕組みとその限界

＜ChatGPT の技術的仕組み＞

それでは、このように爆発的に関心が高まっている大規模言語モデル（LLM）による ChatGPT 型の対話型生成系 AI システムは、具体的に、どのような技術的な仕組みになっているのであろうか。

ChatGPT の仕組みは、技術的な観点からは、大量のテキストデータを収集・整理し深層学習を行うことにより、過去の人類の知識を踏まえた人間が書くような「文章」を作成できる大規模言語モデル（LLM）を構築したこと、また、それに加え、その LLM を利用しつつ強化学習を含む各種の機械学習技術の組合せの工夫を行うことにより、人間らしい「回答」を行うような対話型の生成系 AI システムを構築したことが特徴であると言える[54]。

Ａ．大規模言語モデル（LLM）の構築

ChatGPT 型の AI システムの構築においては、まずは、大規模言語モデル（LLM：Large Language Model）の構築が前提になる。この大規模言語モデルは、大量のテキストデータを収集し、コーパス（言語・テキストのデータベース）を作成することから始まる。ChatGPT の場合、大量のインターネット上の多くのウェブサイト、書籍、ニュース記事、雑誌記事、ウィキペディアの記事、電子メールなど、様々な種類のテキストデータを収集している[55]。

その上で、このコーパス（データベース）を基に、多量の文字列に関して条件付き確率に基づき統計的に次の単語を予測する手法（自己再帰型言語モデル）により、大量の知識を文章として構成する仕組みを構築する。その上で、深層学習（ニューラルネット）による注意機構（アテンションメカニズム）を設け、単語間の情報を統合することにより、より柔軟な単語予測システム（Transformer）を構築し、より人間的な「文章」の作成を可能とする。

Ｂ．対話型システムの構築

このように構築した LLM を利用しつつ、タスク（質問：プロンプト）に対応した回答ができるようなシステムを構築する。具体的には、まずは、これらの事前学習済みの言語モデ

---

[54] 詳しい仕組みは、以下を参照。
OpenAI, Introducing ChatGPT
https://openai.com/blog/chatgpt
黒橋禎夫「ChatGPT の仕組みと社会へのインパクト」2023 年 3 月 3 日
https://www.nii.ac.jp/event/upload/20230303-04_Kurohashi.pdf
岡崎直観「大規模言語モデルの驚異と脅威」2023 年 3 月 28 日
https://speakerdeck.com/chokkan/20230327_riken_llm

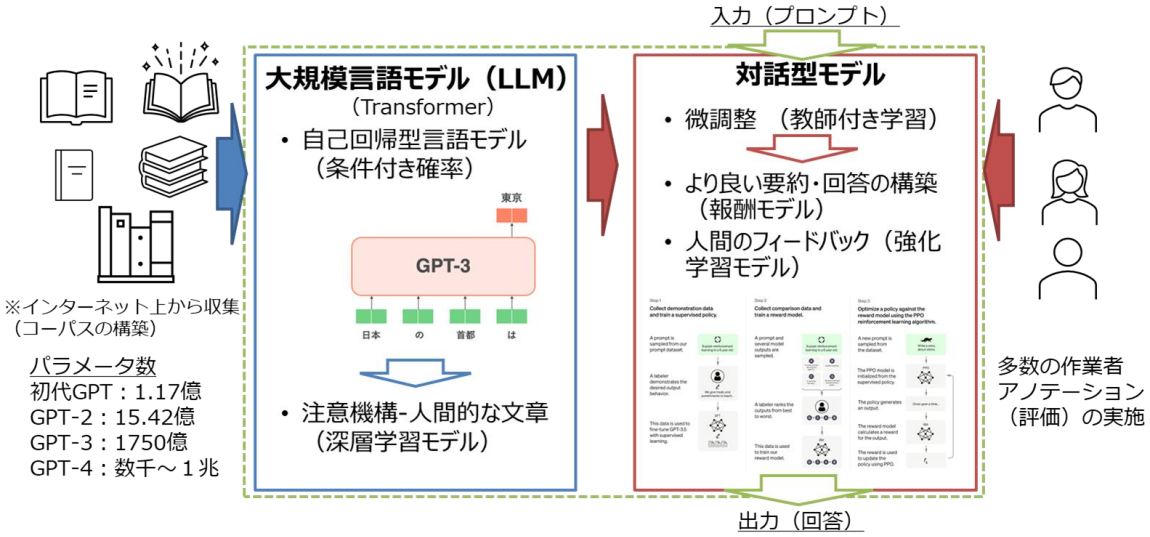
[55] これらのテキストデータには、科学技術、芸術、ビジネス、エンターテインメント、スポーツ、政治など、様々なトピックが含まれる。

ル（Transformer）を、教師付き学習により Fine Tuning（微調整）を実施することにより、タスク固有の知識を取得できるような仕組みを構築する。

その上で、タスクに対する要約モデルに基づく回答結果の生成に関して、まず人間の判断によるどちらか良いかなどのフィードバックを行い、それを踏まえた報酬付きモデルによる強化学習を繰り返すことによる最適化をすることによって、人間らしい「回答」を得ることが可能となるモデルを構築する。

なお、GPT-3 の学習には 460 万ドルのコストを要し、人間の作業時間は数千時間、また、InstructGPT の場合には、40 人の作業者が数万件の事例のアノテーションを実施したとの報告がある[56]。

【図１１】ChatGPT の技術的な仕組み[57]



なお、ChatGPT に係る一般的な仕組みは、上記の通りであるが、その詳細な仕組みは公開されておらず、ブラックボックスとなっている。このため、現在多くの AI 研究者が、ChatGPT を試行し、その回答結果を踏まえて、そのブラックボックスの中身を推定し、また、その性能を評価しようとしているところである。

＜ChatGPT 構築における人間の知識・知見の導入とその限界＞

この ChatGPT のような LLM を利用した対話型生成系 AI システムは、人間の知性の主たる表現手段である言語を扱い、その膨大な知識量に基づく幅広い分野への対応性（汎用性）、また、質問に対するスムーズな回答ぶりから、いわゆる AGI と言われるような人間の知的能力に匹敵するような能力を持ち始めたとの指摘もある。

しかしながら、上述のような技術的な仕組みを踏まえると、ChatGPT は、少なくとも現時点では、人間の質問（プロンプト）に対して適切な回答を行うべく、「これまでの人間の知識・知見を組み込んだシステム（機械）」にしか過ぎず、したがってその組み込んだ人間の知識・知見内でしか対応できない機械であるという見方も可能である。具体的には、大きく以下の二つの視点から人間の知識・知見が組み込まれているものと理解でき、また、それがゆえの限界も理解できる。

[56] 岡崎直観「大規模言語モデルの驚異と脅威」2023 年 3 月 28 日
https://speakerdeck.com/chokkan/20230327_riken_llm
[57] 出典：筆者作成

Ａ．ChatGPT に組み込まれている知識に係る限界

まずは、ChatGPT の知識量、すなわち、何故、ChatGPT は、なんでも知っているのかという点である。これは、端的には、インターネット上に存在する、人類の過去の知的活動の成果である大量のテキストデータの蓄積を学習しているためである。すなわち、これらの膨大な知識を学習することにより、一人の人間が有する知識量を大きく超えるような知識能力を有することになる。

ただし、逆に言えば、ChatGPT においては、原則として、過去において人間によって文章として作成され、ネット上に公開され、学習された知識以外の知識をもって回答することはない。このため、まずは、過去の人類によって蓄積された知識・英知を超えるような回答は、原則として期待できない。また、過去の知識として蓄積されていても、公開されていない、あるいは、学習されていないような知識は利用することはできない。特に、ChatGPT は、2022 年 9 月までのデータを用いて学習しているため、それ以降の出来事に係るデータは含まれておらず[58]、このため、それ以降の出来事については、原則として、正確な回答を期待できないとされる。また、もちろん、企業の社内限りのテキストデータなど非公開の知識を、公開の ChatGPT で利用することはできない。

また、正確性、偏向性なども、過去において学習した知識に依存する。ChatGPT の回答は、原則として、過去に作成された大量の文献データ（文章）の学習に依存しているがゆえに、その学習した大量の文章の中に間違いが多ければ、間違った回答を出す可能性が高くなり、また、偏った文章が多ければ、偏った回答を行う可能性が高くなる。

そういった意味で、原則論として言えば、ChatGPT 自体は非常に大量の知識を有する一方、今までに公開された文章に記載のないような新たな知識を自ら生成することは、原則としてなく、したがって、人間のように新たな知識創造活動を行うような「知性」を有するものではないと解釈される。

Ｂ．ChatGPT による人間らしい回答の作成能力とその限界

次に、ChatCPT の回答に係る滑らかさ、すなわち、何故、その回答が人間らしいのかという点である。これは、教師付き学習とそれに基づく強化学習により、人間が訓練をしているからである。具体的には、どのように要約しまた回答すれば「より分かりやすい」「より人間らしい」のかについて、人間が判断し、その結果を ChatGPT に教え込んでいるのである。

したがって、ChatGPT は、質問（プロンプト）に対する人間の要約・回答パターンを人間から学習し、そのパターンに基づき要約・回答の作成という文章作成作業を行う機械であるということができる。このような文章作成作業は、人間であれば、初等中等教育から高等教育あるいは社会に出てからも獲得し続ける能力であると考えられ、その意味で、人間が有する高等な「知性」を有しているように見える可能性がある。

しかしながら、原則としては、与えられた質問に対して、一定のパターンの回答を作成する機能のみであり、人間のように自ら目標を決定し、その目標の達成に向けて複雑な思考プロセスを行い、実現するようなことはできない点が限界であると考えられる。

---

[58] 【ChatGPT】最新の情報がないことで発生する誤回答と使い方に関する注意点（無意識に「現在」を聞く質問に注意）
https://did2memo.net/2023/02/15/chatgpt-old-data-issue/

②ChatGPT との対話によるイノベーション創出の可能性

＜ChatGPT の創造性に係る可能性＞

上述のように、ChatGPT は、基本的には、既存の公開された文献・テキスト情報をもとに回答を行うシステムであり、したがって「原則としては」新たな知識創造活動を行うものではなく、したがって、イノベーションにつながるようなことは困難であるとされる。

しかしながら、創造性とは複数の知識の組合せによって生じるとも解釈することができることを踏まえると、前段（A）におけるデータベース（LLM、コーパス）部分で得られた複数の知識を、後段（B）における回答・要約作成能力を利用することにより、それらを統合して提示することによって、これまでになかったような新たな組合せ（アイデア）を提示し、一定の創造性を発揮することは可能であると考えられる。例えば、文献Aに記載している内容と文献Bに記載している内容を、それぞれ組み合わせて要約し提示することによって、これまでなかったような新たな視点を提供するような場合である。実際に、ChatGPT は創造的なことを行っているとの報告もある[59]。

＜人間との対話によるイノベーション創出の可能性＞

ただし、イノベーション創出の観点から見た場合、組合せによりこれまでなかったような創造的で新たな視点を回答として提示するだけでは、不十分であると考えられる。

すなわち、イノベーションとは、一般的には、「新結合の遂行により新たな付加価値を創造すること」である。したがって、単に、既存の知識等を組み合わせること（新結合）により、新しいアイデアを創造するだけではなく、そのアイデアに基づき、社会ニーズ（目的）に照らして新たな付加価値を創造するべく、

- 当該アイデアが、社会ニーズ（目的）に対して、実現可能で最適なアイデアなのかについて試行錯誤を行うプロセス
- 当該試行錯誤のプロセスを通じて、社会全体としてルーティンとして確立し、実現（遂行）していくプロセス

が不可欠となる。

このため、ChatGPT を利用することによって、新たなアイデアが提示される場合もあると考えられるが、少なくとも実世界でのイノベーションを考えた場合には、その後、以下のようなプロセスを互いにフィードバックしながら行うことが更に必要になると考えられる。

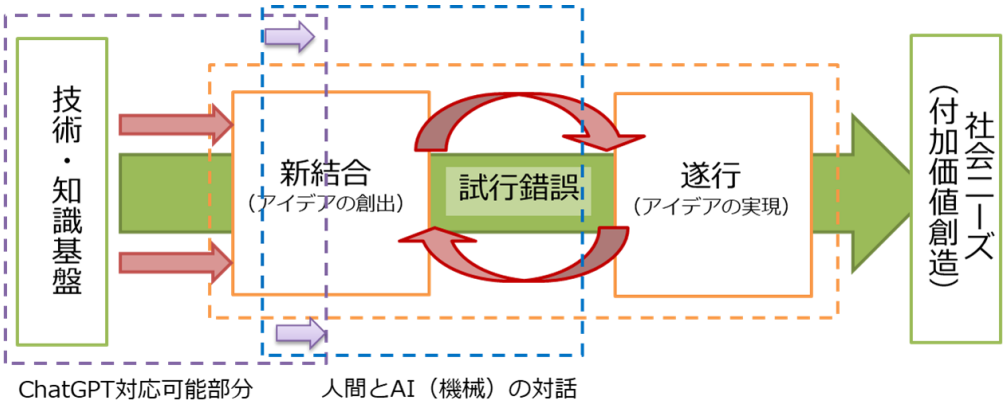
- 1) そのアイデアが社会ニーズに照らして付加価値創造をするようなものなのかについて、人間が、必要に応じて ChatGPT と対話をしながら、試行錯誤を行うことにより、最適と考えられるものを抽出し、構想していくプロセス（対話プロセス）
- 2) 上記と並行しつつ、そのアイデアを試行的に社会に実装し、実際に付加価値を与えるものなのかどうかフィードバックを受けながら試行錯誤を行い、そのアイデア・構想を現実社会において実現（遂行）していくプロセス（遂行プロセス）

[59] ITMedia NEWS「東大松尾教授が答える、ChatGPT とは何なのか？　一問一答」2023 年 04 月 05 日
https://www.itmedia.co.jp/news/articles/2304/05/news139.html

したがって、ChatGPT は、現実社会におけるイノベーションの創出プロセスにおいて、当初のアイデア創出のプロセスに加え、前段の上記 1)の対話プロセスにおいて、いわゆる人間-機械チーミング（Human-Machine Teaming：HMT）と言われるような、対話を行うことを通じて、重要な役割を果す可能性がある。その際、いわゆるプロンプトエンジニアリングと言われるような、質問内容を工夫しながら、より適切な回答を求めていくプロセスは、その一部と位置づけられるであろう。

一方で、ChatGPT は、あくまでも対話型の AI システムにしか過ぎず、そもそもの目標設定に加え、後段の上記 2)の遂行プロセスである、試行錯誤を通じた現実社会での実現（遂行）というプロセスについては、引き続き、人間が担うことが求められると考えられる。（図１２参照）

【図１２】イノベーションプロセスにおける ChatGPT の役割の可能性[60]



＜AI のみによるイノベーションの将来的可能性＞

なお、上記の議論は、現実世界でのイノベーションにおける ChatGPT 型の AI システムの役割を想定したものである。しかしながら、デジタルでのコンテンツの生成物自体を新たなイノベーションとする分野においては、生成系 AI システムの果たす役割は、更に広がる可能性がある。例えば、画像生成系の AI システムでは、デジタルコンテンツの作成（イノベーションの遂行）までをシステムとして構築することが可能であるため、この中で、人間（ユーザー）の役割は、AI によって生成されたコンテンツのうち、社会ニーズに合うような（人々に受け入れられるような）コンテンツを選択すべく、プロンプトエンジニアリングを進めることに限定される可能性がある。

また、その近未来的な世界としては、これらのプロセスが AI システム自体によって全てデジタル上でイノベーションが全て完結するような世界が到来することも想像することができる。例えば、過去の社会ニーズの動向をビッグデータ解析した上で、自動的に複数のデジタルコンテンツの候補を生成し、それらのコンテンツをインターネット上で試行公開し、その中で、ネットユーザー（社会）で最も人気のあったコンテンツを集中して正式公開・販売するような AI システムを想定することも可能であろう。ただし、このように、ある意味での社会実験による試行錯誤と最終意思決定までを行うような AI システムは、自ら意思決定までも行う AGI に近い存在となり、したがって「倫理的」に望ましいかについては、議論があるであろう。

[60] 出典：筆者作成

（２）ChatGPT 型 AI システムの位置づけとそのリスク

①全体から見た ChatGPT 型 AI システム位置づけ

＜これまでの従来型 AI システムと ChatGPT 型 AI システムとの比較＞

それでは、このような ChatGPT 型の AI システムは、多様な AI システム全体の中で、どのように位置づけられ、また、特に機能面から見た場合、第三次 AI ブームを通じて、これまで普及してきた主要な AI システムと比較して、どのような関係にあるのであろうか。

ISO/IEC において 2022 年に制定された国際標準（ISO/IEC 22989）[61]では、「AI システム」を「人間によって定義された目的の一連のセットのために、コンテンツ、予測、推薦、決定などの出力を生成する工学システム」[62]を定義している。このうち、第三次 AI ブームにより普及が進んだこれまでの従来型の AI システムでは、主に、機械学習を利用した「予測、推薦、意思決定」システムが中心であったのに対し、今回の第四次 AI ブームでは、ChatGPT に代表される、LLM を利用した文章に係る「コンテンツ生成」システムが中心であるといえる。

すなわち、第三次 AI ブームでは、当初は、特に深層学習により、人間と同様、画像などのパターンを認識することが可能になった点に注目が浴びたが、その後は深層学習を含む各種機械学習技術を利用することにより、画像も含めて過去の各種データから学習するとともに、それらを含むビッグデータ解析に基づき、人間あるいはそれ以上の能力を持って予測、推薦し、または、自ら意思決定するシステムが広く普及し始めてきたと言える。このような予測、推薦、意思決定を行う AI システムは、特定の個別分野に応じて学習され、構築されることから、これまで（いわゆる AGI に対して）「特化型 AI」と呼ばれていた。

これに対して、第四次 AI ブームの中心となる ChatGPT は、コンテンツを生成する「生成系 AI システム」であると言われる。生成系 AI システムとは、基本的には、過去のテキスト・画像等のコンテンツ・データを学習した上で、プロンプト（入力）に従って、テキスト・画像等のコンテンツを生成する AI システムである。もちろん、以前より生成系の AI システムは存在しており、例えば画像系では、GAN（Generative Adversarial Network：敵対的生成ネットワーク）を利用した DeepFake 技術も生成系 AI システムの一種であると考えられる。また、近年関心が高まってきた Stable Diffusion や Midjouney などによる、いわゆる画像系の生成系 AI システムも、この延長線上にあるといえる。一方、テキスト系では、自動翻訳システムなどもその生成系 AI システムの一種と位置づけられると考えられ、実際に、ChatGPT で利用されるような大規模言語モデル（LLM）は、自動翻訳システムの高度化のために開発されてきたという側面も有する。

また、ChatGPT は、生成系 AI システムの中でも、特にテキストを扱う対話型 AI システムとの位置づけも有する。すなわち、質問に対して、回答をテキストとして生成するシステムであるとの位置づけである。この対話型システムについても、以前より、いわゆる各種の

---

[61] ISO/IEC 22989:2022 Information technology — Artificial intelligence — Artificial intelligence concepts and terminology
https://www.iso.org/standard/74296.html?browse=tc

[62] AI system：engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives

ChatBot や AI スピーカーなどの AI システムが存在していた。また、インターネットの検索システムにおいてもこれまで多くの AI が導入されてきており、これも対話型 AI システムの一種としてみなすことが可能であろう。ただし、これらの以前の対話型 AI システムと比較して、ChatGPT は、大量の計算能力を利用し莫大な規模の LLM を構成することにより、広範囲の内容に対して回答することが可能になり、その汎用性が一気に拡大した点が特徴とも言える。

【図１３】意思決定系 AI システムと生成系 AI システム（主な例）[63]

分類（出力）		主な訓練データ（入力）	主な課題
意思決定系AI 予測、推薦、意思決定	画像認識システム （生体認証等）	画像等（個人情報）	プライバシー
	意思決定系システム	個人情報 非個人データ（IoT等）	公平性（バイアス等） 安全性
生成系AI コンテンツ生成	画像生成システム （DeepFake等）	画像（著作物）	著作権・人格権 （偽情報）
	生成系AIシステム	画像（著作物） テキストデータ（LLM）	著作権（盗作） 正確性（幻覚）

＜ChatGPT は AGI なのか：AGI から見た ChatGPT の位置づけ＞

ChatGPT は、生成系 AI システムの一種として、人間の知的能力の主要形態である言語能力を扱い、非常に多様な質問に対応できるという汎用性が特徴であると言える。このため、「ChatGPT は、AGI なのか」ということにも関心が高まっている。

前章（１）で記載したとおり、AGI とは、一般的には、人間が行うことができるあらゆる（汎用の）知的作業を理解・学習・実行することができるような人工知能（AI）とされる。その際、ChatGPT 型の AI システムは、人間の行う幅広い（汎用性の高い）知的作業を理解し、また、一部学習する能力を有するものと解釈されうることから、AGI の初期段階であるとの理解が多い。例えば、ChatGPT は、これまでの人工知能とは一線を画しており「驚くほど人間の能力に近付いている」とし、そのため、「人工汎用知能（AGI）の初期バージョンである可能性がある」と記述している論文もある[64]。一方、ChatGPT を含め、最近の生成系 AI システムは、いずれも「弱い AI」であり、実世界から学んで、それを知的タスクに活かせるような「強い AI」はまだ存在しないという見方もある[65]。

このような議論を踏まえると、第三次 AI ブームでは、深層学習を始めとする機械学習の進展により、特定の分野において、予測、推薦から更には意思決定まで可能となるような、より「自律的な」AI システムの開発・普及が進んだのに対し、第四次 AI ブームでは、LLM

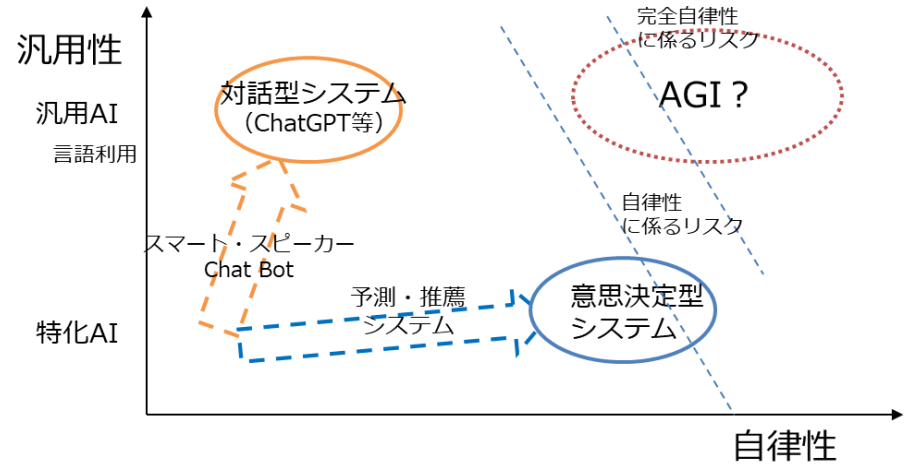
[63] 出典：筆者作成。

[64] Bubeck, Sébastien; Chandrasekaran, Varun; Eldan, Ronen; Gehrke, Johannes; Horvitz, Eric; Kamar, Ece; Lee, Peter; Lee, Yin Tat et al. (2023-03-22). “Sparks of Artificial General Intelligence: Early experiments with GPT-4”. arXiv:2303.12712 [cs].
https://arxiv.org/abs/2303.12712
GPT-4's performance is strikingly close to human-level performance, and often vastly surpasses prior models such as ChatGPT. Given the breadth and depth of GPT-4's capabilities, we believe that it could reasonably be viewed as an early (yet still incomplete) version of an artificial general intelligence (AGI) system.

[65] VoiceHQ, “The difference between weak AI and strong AI”, 2023.1.25
https://medium.com/voicehq/the-difference-between-weak-ai-and-strong-ai-2166e8b2ecdb

による莫大な規模の知識に係る言語処理能力を背景に、人間に近いようコンテンツ（文章）生成を可能とする「汎用的」なシステムが開発されたものの、意思決定・実行までを含む自律性という視点では、質問に対して回答を返す「対話型」にしか過ぎないものと位置づけられるものと考えられる（図１４参照）。

【図１４】AGI と対話型システム、意思決定型システムとの関係（整理）[66]



②ChatGPT 型 AI システムのリスクの考え方（他の AI システムとの比較）

＜AGI のリスク：完全自律性に係るリスクと悪用された場合の大きなリスク＞

それでは、まず、AGI に関しては、具体的に、どのようなリスクがあるのだろうか。第二章で記載したとおり、特に欧米においては、AGI は人類の文明を脅かす可能性があるものとして認識され、このため AGI の開発は厳しく監視しなければいけないという暗黙のコンセンサスがあるように伺える。

しかしながら、その理由に関しては、「そもそも人間と同様の知性を有する機械（AI）を作ることは望ましくない」という文化的な認識を背景にした議論が先行しており、具体的に、どのような AGI であれば、どのような観点から、人間社会・文明に悪影響（リスク）を及ぼすのかについては、必ずしも明確になっていないように見える。

これに関して、各種文書などを読む限り、AGI に係る具体的なリスクとしては、以下の２点があげられると考えられる。

１）「完全自律性」を有する AGI 自体のリスク

アシロマ AI 原則を読むと、高度な AI の中でも、特に一部の特定 AI、具体的には 22 番の「再帰的に自己改善／自己増殖するような」AI システムについては、厳しく管理すべきとしている。これは、AI が進化・発展することにより、AI（機械）自体が、自己発展・拡大が可能となるような完全な「自律性」を獲得した場合には、人間・社会による管理・制御可能性を失ってしまい、その結果、人類社会に対して予測不能なリスクを生じさ、場合によっては「人類の文明を脅かす」リスクがあるとの趣旨であると考えられる。このような観点から、実際に、特に欧州では、AI システムに対して「Human Oversight」の観点からの管理の

[66] 出典：筆者作成

義務付けなどが強調される傾向にあり、また、技術的に言えば、（社会的合意も含めた）制御可能性が重要になると考えられる。極論すれば、人間社会が（社会的観点からも）AGI の電源を切る能力を保持していることが重要になる。

２）汎用性、自律性を有する AGI が悪用されるリスク

一方で、本来の AGI の本質は、特化型 AI と比較して、その「汎用性」に特徴がある。この AI の汎用性が、人類・社会に対してリスクになる理由は必ずしも明らかではない。ただし、多くの AGI に対する不安感・不信感に係る文書を読む限り、一部の集団が AGI を悪用することにより、当該集団が汎用的で強力な能力を有することになり、その結果、人類・社会に対して悪影響を与えるというリスクであると考えられる。その際、当該集団が悪用するために必要な強力な能力としては、その当該集団の指示（目的）にしたがって AI が汎用的かつ自律的に意思決定する能力が求められることになるものと考えられる。

このような AGI の悪用に係るリスクに対しては、人類・社会全体での AI あるいは AGI での管理・ガバナンス体制の検討が求められることになり、それはある意味で社会的な制御可能性の確保という視点が必要になるものと位置づけられる。

なお、特に汎用性を有する AGI に対する不安感・不信感の最も大きな要因は、上述のような、AGI 自体による人類社会に対するリスクによるものというよりは、むしろ、人間のように振る舞う機械とされるまだ見ぬ AGI に対する不安感・不信感に加え、現実的には、人間に代わって職業を奪うことにより、失業を引き起こし、さらには社会の経済・雇用体制自体が抜本的に変化してしまいかねないという不安感・不信感であるとも考えられる。

＜意思決定型 AI システムと対話型 AI システムのリスク面での比較＞

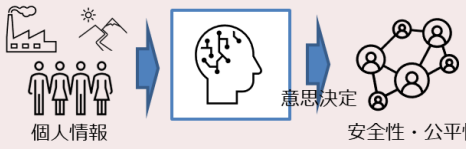
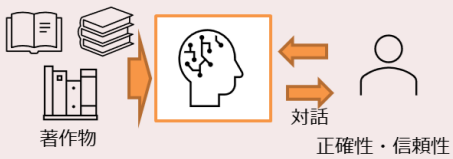
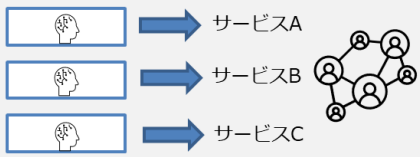
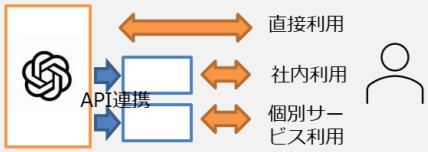
次に、上述の AGI のような将来的なリスクに対して、今回の ChatGPT 型の対話型 AI システムにおいては、現実的に引き起こす可能性のある社会的なリスクについて、従来型の意思決定型の AI システムと比較してどのような違いがあるのだろうか。

もちろん、ChatGPT 型の対話型 AI システムにおいては、従来型の AI システムと同じ視点に係るリスク・社会的課題もあるものの、取り扱う学習データやその出力の内容が異なるため、一般的には、従来型の意思決定型の AI システムとは異なった社会的リスクを生じさせる可能性がある。

具体的には、従来型の AI システムにおいては、特定の分野の目的に対して、センサー・IoT 機器を通じたものも含め、非個人情報／個人情報に係る多様なデータを利用して機械学習を行い、予測、推薦、意思決定を行うシステムが中心となる。その際、社会規範（倫理）的な観点からは、IoT 機器などで得られたデータを利用し、人命に関わるような意思決定を行う場合においては、特に安全性の問題が重要とされたことに加え、個人情報などを利用して個人に対する重要な意思決定を行う場合には、そのプライバシー上の問題に加えて、そのデータのバイアスなどに起因する公平性の問題が特に新たな議論として浮上してきたことが特徴であると言える。また、そのため、当該システムが、予測・推薦システムにとどまる場合には一般的には大きな問題とはならないものの、特に人命や重要な意思決定などに関わるような分野において、システムが「自律的」な意思決定を行うような場合においては、その責任関係の明確化の観点からも、（AGI で求められるような）Human Oversight が重要になるとされている。

これに対して、今回 ChatGPT のような生成系 AI システムにおいては、基本的には、既存のテキストや画像などの多量のコンテンツ・データを利用して学習し、利用者のインプット（プロンプト）を踏まえて、新たなコンテンツ（文章、画像等）を生成するシステムとなる。このため、社会規範（倫理）的な観点からは、コンテンツやプロンプト等に含まれる個人データなども重要にはなるものの、むしろ学習データに係る著作権や、出力コンテンツとしての正確性・信頼性などに係る問題が特に重要になる。ただし、その際、対話型システムであるがゆえに、システムのみの自律性は必ずしも有しておらず当然にして人間が関与するものであることから、上述のような Human Oversight の問題はそれほど重要な論点とはならない。一方、その汎用性がゆえに、人間がシステムに依存してしまうという問題が生じる可能性があり、システムの利用にあたってのルールやリテラシーが大きな課題になることが想定される。

【図１５】従来型の意思決定型システムと ChatGPT 型の対話型システムの比較[67]

	機械学習による意思決定型システム	LLMによる対話型システム
技術の進化	• DLによる画像認識等（生体認証等） ⇒機械学習による予測、推薦、意思決定システム	• LLMによる対話型システム ⇒？
人間との関係	• 予測、推薦⇒人間に提供（情報支援型） • 意思決定⇒人間不要（←人間による管理） 	・対話⇒人間に提供（知識支援型） 
アーキテクチャー／産業構造	カスタム型 	汎用型（⇒?） 

＜AI システムに係るアーキテクチャと産業構造的な差異＞

なお、今回の ChatGPT 型の対話型 AI システムは、従来型の意思決定型 AI システムと比較して、そのアーキテクチャ、産業構造も大きく異なるため、そのガバナンスの検討にあたっても留意することが必要となると考えられる。

具体的には、これまでの従来型の AI システムの場合には、大手 IT・ネット企業にせよ、個別ユーザー企業にせよ、基本的には、機械学習・深層学習などに係る汎用的な技術を、自らのビジネス・サービス合わせ、カスタマイズして活用・応用することによって、その導入・普及を推進してきた。

これに対し、今回の ChatGPT 型の対話型システムにおいては、少なくとも現時点では LLM の開発に大きなコストが必要となるため、ChatGPT のような基本となるシステムは、汎用 AI（General Purpose AI）システムとして、第二章（３）で記載した通り、一部の大手 IT・ネット系企業に集中する傾向にある。一方、このような汎用 AI システムは、一般個人もユーザーとして広く利用することが可能であるため、ある種の「AI の民主化」が進展するとともに、API を通じて開放することにより、各企業内での利用や、当該汎用 AI システ

[67] 出典：筆者作成

ムを利用した外部向けの新たなサービスを創出する企業が現れることになる。実際に、Azure OpenAI Service などを通じ、各企業による情報収集、文章の要約・作成、リストや表の作成などを含む、社内におけるテキスト・文章関連業務での利用が進展するとともに、各ベンチャー企業等による広告、法務、コンサル・調査、教育などの様々な分野でのテキスト・文章関連に係る ChatGPT を利用した新サービスの提供などが拡大している[68]。

今後、この ChatGPT 型の AI システムが、汎用 AI システムとして幅広い分野・ユーザーにおいて利用されることによって、更なるイノベーションの創出が期待される一方で、分野によっては、新たなリスク課題が生じる可能性があり、その際、このようなアーキテクチャ・産業構造上を踏まえた上で、そのガバナンス体制を検討する必要が生じることになると考えられる。

（３）ChatGPT 型 AI システムの具体的リスク・社会的課題

＜ChatGPT 型 AI システムに係る具体的リスクと課題＞

それでは、ChatGPT 型の AI システムにおいては、現時点で具体的にどのような社会的リスクが想定されるのであろうか。このような問題意識のもと、以下においては、出力内容に係る社会的リスク（正確性・信頼性、公平性・社会的妥当性）と、入力データ関連のリスク（オーサーシップ・盗作・著作物、個人情報・企業秘密）に分けて、それらの課題について整理する（図１６参照）。

【図１６】ChatGPT 型 AI システムに係る社会的リスク[69]

分類			課題・論点
出力内容に係る社会的リスク	正確性・信頼性	正確性、誠実性	• 学習データ上の限界 • 嘘・幻覚（Hallucination）：事実関係に基づかない回答（知ったかぶり）
		悪用による偽情報（信頼性）	• そもそもフィクション作成も可能 • 悪意ある文章の作成の容易化
	公平性・社会的妥当性	公平性、社会的妥当性	• 学習データ上のバイアスの可能性
		政治的妥当性	• 社会的妥当性（中国の規制の事例）
入力（学習データ等）関連のリスク（各種法令関係）	オーサーシップ・盗作と著作権	出力の著作権・盗作（Authorship）	• 著作権法上の扱い（日本、米国等の事例） • 研究論文、教育課題等での利用
		学習データの著作物の利用	• 著作権法上の扱い（日本、米国等の事例）
	個人情報、企業秘密	学習データの個人情報利用	• 個人情報保護法上の扱い（GDPR・伊の事例）
		プロンプト入力における個人情報、企業秘密	• 情報セキュリティに係る信頼性

[68] 例えば、以下を参照。
AI Market「ChatGPT とは？API で何ができる？仕組み・企業活用事例 10 選！ビジネス導入方法を知る！」2023-04-28
https://ai-market.jp/technology/chatgpt/
デジタルクロス（大和 敏彦）「ChatGPT が加速した企業や社会の AI 活用【第 67 回】」2023 年 4 月 17 日
https://dcross.impress.co.jp/docs/column/column20170918-1/003397-2.html

[69] 出典：筆者作成

なお、これらの社会的リスクとしては、現時点において指摘・想定されているものを取り上げる。今後、一部のリスクについては、システムの改善等の進展に伴い、その位置づけが低下する可能性がある一方、新たなイノベーションの進展に伴い、これらとは異なるような新たなリスクが生じる可能性があることに留意することが必要である。

①正確性・信頼性

＜生成テキストの正確性、誠実性＞

まずは、対話型の AI システムに対しては、多くの場合、「正確性」あるいは質問に対する「誠実性」のある回答が期待される。これに関しては、上述の（１）で説明した通り、そもそも ChatGPT は、基本的には過去の各種文献・テキスト情報に基づいて文章を生成するシステムであり、したがって「原則として」それらの過去の各種文献・テキスト情報の範囲内でしか回答は得られない。したがって、当然、未来の話や現在人類に知られてないような話に対しては、原則回答することはできないし（もちろん、過去時点での将来予測の話はある）、過去のデータ・文献に間違いがあれば、正しく回答しない可能性が高い。

また、この件に関連して、ChatGPT の欠陥の一つとして Hullucination（幻覚を起こす、嘘）の問題が指摘されている。これは、過去の人間の知識としてテキスト化されていないようなことに関しても、「知りません」と応えるのではなく、平然として嘘をつき、事実に基づかない不正確な回答を行ってしまうという問題である。これは、本システムが、単語間の確率関係と注意機構に基づく関連付けのみで計算されているため、必ずしも明確に記載のないものについても、何等かに関連付けて勝手に応えてしまうものと考えられる。

このような問題については、今後システムの改善が行われ、精度の向上がなされる可能性はあるものの、一方で、その技術的限界は必ず存在し、不正確性を完全に排除することは困難であると考えられる。その際、実の人間であっても必ずしも完全に正確な回答を行う訳ではないことに加え、そもそも「完全に正確な回答」とは何かという哲学論も存在するであろう。このようなことを踏まえると、ChatGPT 型の AI システムを、全知全能な「完全に正確な回答を行う機械」として扱うべきものではそもそもなく、本システムを利用する人間・ユーザーが、ChatGPT 型の AI システムの限界を理解・把握した上で、回答を利用するという使い方が求められることになり、結局は、利用者側のリテラシーの問題に帰結する。

＜悪用による偽情報の生成（信頼性）＞

また、特に ChatGPT 型の AI システムは、過去の文献・テキストデータのパターンを踏まえ、人間の入力（プロンプト）に従って、あたかも人間が作成したような創作（フィクション）の文章を作る機能も有する。このような機能は、人間の創造的な活動を支援するものであり、非常に有用な機能であると考えられる。

しかしながら、このような機能は、一方で、悪用される可能性は否定できない。すなわち、悪意のある人間が利用することによって、より巧妙な詐欺などの文章を非常に容易に作成することが可能となり、その結果、そのような詐欺など文章が世間で悪用される割合が拡大する可能性もあるとの指摘もある。

なお、このような悪用に関しては、画像・映像系の生成 AI システムにおいて偽情報（偽画像、偽映像）の生成を可能とするものとして、従来よりディープフェイク技術が存在する。ただし、このディープフェイクの場合には、もともと通常の人間にとっては偽画像・偽動画

を作るのが困難であったものを、AI システム（ディープフェイク技術）によって、誰でも作成することが容易になり、その結果、犯罪者による参入障壁が低下したという側面があった（「AI の民主化」の負の側面ということもできる）。

これに対し、ChatGPT 型 AI システムのようなテキスト生成技術の場合は、そもそも通常の文章能力のある人間であれば、従来より、ChatGPT に頼らずとも偽情報の一つくらいは容易に作ることができ、その悪意ある行動の目的を十分に達成できていたものと考えられる。このため、悪意ある人間（犯罪者）にとっては、ChatGPT の登場による参入障壁の低下はそれほど大きなものではないという指摘も可能である。ただし、例えば、翻訳機能の正確化が進展すれば、少数の犯罪者では困難だった外国向け詐欺がより容易になる可能性があるし、また、更に技術が進展すれば、大量の類似する偽情報を迅速に作成可能となったり、あるいは対話型の偽情報を自動的に提供システムが開発されたりするなど、更なる悪意あるシステムが開発、利用される可能性（危険性）も考えられる。

②公平性・社会的妥当性

＜生成テキストの公平性、社会的妥当性＞

従来の意思決定型 AI システムにおいては、その学習データにバイアスや偏見が含まれている可能性があるため、出力される結果においてもバイアスや偏見が含まれることになり、その結果、個人に係る意思決定に関し公平性・人権の観点から問題が生じる可能性があることが大きな課題となっている。

これに対し、ChatGPT のような生成系の AI システムにおいても、同様の問題が生じる可能性はある。例えば、ジェンダー的、人種的に偏った画像や文章表現が学習データに多く利用されていれば、生成される表現が人々にとって不快と感じる場合もあるであろう。また、現時点では特に大きな問題にはなっていないものの、場合によっては暴力的な表現や性的な表現など、一定のグループにとっては受け入れられない表現が AI システムによってなされる可能性も否定はできない。実際に、過去の対話型システムであるマイクロソフトの Tay などの AI Chatbot のように、偏った学習を行うことに炎上した事例も見受けられる。

このような AI によって生成される表現上の問題は、実の人間が作成した場合にも問題になりうるものであるが、特に ChatGPT の場合は、多くの人々が共通して利用するシステムであるがゆえに、社会的に大きな影響を与える可能性があり、したがって、回答・表現に係る社会通念上の妥当性が求められることになる。このためには、入力面あるいは出力面のいずれかで、「社会通念」を学習させるべく技術面での取組が求められる可能性が高い。

＜生成テキストの政治的妥当性＞

このような社会通念からみた妥当性に係る議論の延長で、地域差に伴う社会的通念の差異、特に政治的な観点からの妥当性の問題が生じうる。

典型的な事例は、中国の事例である。すなわち、世界の多くの文献・テキストデータを学習した場合には、中国共産党の問題点や天安門事件の概要などを学習することになるが、中国共産党政府から見た場合には、そのような自国の政権に批判的な文章を学習し、それに基づいて回答することは許されないという問題である。

本件に係る規制については、第 4-3 章で触れる。

＜生成コンテンツのオーサーシップ（著作権上の扱い、盗作の問題）＞

ChatGPT に限らず生成系 AI システムにおいては、利用者がプロンプトを入力することにより、多数の既存コンテンツから学習したシステムが、新たなコンテンツを生成することになる。その際、新たに生成されたコンテンツは、誰が作成したものになるのかということが課題になる。いわゆるオーサーシップあるいは著作権の所在の問題である。

このような AI 生成物に係る現行の日本の著作権上等での取り扱いにおいては、これまで、知的財産戦略本部を中心に議論がなされてきている。具体的には、同本部の次世代知財システム検討委員会（2015 年～2016 年）及び新たな情報財検討委員会（2016 年～2017 年）[70]において、AI 生成物（創造的活動により生み出されるものに相当する出力）のうち、人間の創作的寄与があるものは、「AI を道具として利用した創作物」として認められる一方、寄与がないものは現行知財制度上の対象とならないと整理されており、今後、その人間の創作的関与の程度について引き続き検討するとしている。なお、米国においても、米国著作権局が、2023 年 3 月 16 日付けで、AI が生成した素材を著作物として認める場合を明確にするガイダンスを発表している[71]。

その際、このような AI 生成物に係る著作権上の扱いについては、これまで主に画像系・音楽系のコンテンツを対象に議論されていた。それに対し、今回の ChatGPT においては、小説などのコンテンツ生成に資するだけではなく、特に、レポート課題などの教育面や、知的な創作活動である研究面での利用が大きな問題となっている。すなわち、著作権の所在の有無の問題以前に、いわゆるオーサーシップの問題であり、その点を明確にせずに利用すれば、いわゆる盗作（Plagiarism）の問題になる。もちろん、盗作の問題は、以前においてもいわゆるコピペの問題などとして存在してはいたが、ChatGPT を利用することにより従来よりも検出が困難になるという状況になる。なお、生成系 AI によって生成されたコンテンツを検出（Detection）するシステムに関しては、開発可能ではあるものの、その後その検

---

[70] 知的財産戦略本部検証・評価・企画委員会
https://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/index.html
内閣官房知的財産戦略推進事務局「ＡＩによって生み出される創作物の取扱い（討議用）」平成２８年１月
https://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2016/jisedai_tizai/dai4/siryou2.pdf
https://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2016/jisedai_tizai/dai4/gijisidai.html
知的財産戦略本部 検証・評価・企画委員会 次世代知財システム検討委員会 「次世代知財システム検討委員会 報告書 ～デジタル・ネットワーク化に対応する 次世代知財システム構築に向けて～」平成２８年４月
https://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2016/jisedai_tizai/hokokusho.pdf
知的財産戦略本部 検証・評価・企画委員会新たな情報財検討委員会「新たな情報財検討委員会報告書ーデータ・人工知能（ＡＩ）の利活用促進による産業競争力強化の基盤となる知財システムの構築に向けてー」平成２９年３月
https://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2017/johozai/houkokusho.pdf
https://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2017/johozai/houkokusho_gaiyou.pdf

[71] Gigazine「AI が生成した絵や文章に著作権は認められるのか？アメリカ著作権局がガイダンスを発表」2023 年 03 月 17 日
https://gigazine.net/news/20230317-copyright-registration-generative-ai-works-guidance/

出能力を出し抜くようなシステムの開発を促すことになり、結局、能力開発競争に陥り、究極的な効果は期待できないとされる。

いずれにせよ、このような教育面・研究面における ChatGPT 利用に係るオーサーシップについては、そもそもの当該教育・研究におけるそれぞれの場面に応じて人間に期待される役割を踏まえつつ、ガイドラインを定めていくことが必要と考えられる。これに関する現状に関しては、次節（４）に記載する。

＜学習データとして利用する著作物の権利関係＞

一方、上記の AI 生成物の著作権問題に関連して、AI の学習データとして利用されているコンテンツの著作物の利用に係る是非の問題が、特に画像系の生成系 AI システムを中心に大きな問題になりつつある。

具体的には、米国では、もともと FairUse 規定に基づき「Fair」であるか否かを原則として裁判所が判断する仕組みになっているが、これに関し、生成系 AI の開発（学習）に利用されるコンテンツの利用が「Fair」であるのかに関し、既にいくつかの訴訟事例が起こされている[72]。例えば、画像系生成 AI システムを開発・提供している Stabale AI 社、Midjouney 社、Deviantat 社に対して、2023 年 1 月に、クリエイターから集団訴訟が起こされている。また、ChatGPT に対しても、ダウ・ジョーンズ社は、2023 年 2 月、同社の発行するウォールストリート・ジャーナル（WSJ）の記事を、ChatGPT が機械学習のための無断で利用しており、その利用のためにはライセンス（利用許諾権）を受けることが必要との公式コメントを出したことが報じられている。

一方、日本においては、2018 年の著作権法改正[73]において、このような生成系 AI における著作物の利用に関して、柔軟な権利制限規定が設けられている。具体的には、同法改正により、「著作物に表現された思想又は感情の享受を目的としない利用」に関する権利制限規定（第 30 条の 4）などが新たに設けられ、基本的には、AI による深層学習などのために利用する場合には権利が制限され、自由に利用できるようになった。この柔軟な権利制限規定は、もともとは、米国の FairUse 規定のような規定を日本でも作るべきとの判断で制定されたものであり、現在の ChatGPT を始めとする生成系 AI システムの興隆を想定していたものでは必ずしもないが、結果的には、米国と比較して、AI システムの開発側にとっては有利な規定となっているとの見方もある。

ただし、近年の生成系 AI システムの興隆の中、日本においても、クリエイターの一部などは、これらの生成系 AI は彼らの仕事を奪うものとしてこのような著作権上の扱いに対して反発をしている。具体的には、イラストレーターらでつくる団体である「クリエイターと AI の未来を考える会」は、2023 年 4 月 27 日に記者会見を行い、特に画像の生成系 AI に関し、著作物を許諾なく AI が学習できる規定は創作者の搾取につながるものとし、画像生成 AI の著作物の利用に関し著作権法 30 条の 4 の規定から対象外にすること、オリジナル画像の著作権者の明示の義務付けなどの法規制を求めたことが報じられている[74]。

---

[72] KDDI research atelier 研究員コラム（小林雅一）「第 3 回 生成 AI が著作権侵害などで訴えられる―人間の作品から学んで創る人工知能はクリエイターやジャーナリストの敵となるのか？」 2023-03-16
https://rp.kddi-research.jp/atelier/column/archives/1192

[73] 文化庁、著作権法の一部を改正する法律（平成 30 年法律第 30 号）について
https://www.bunka.go.jp/seisaku/chosakuken/hokaisei/h30_hokaisei/

[74] 日本経済新聞「画像 AI は「著作権侵害」　イラスト画家ら法規制訴え」2023 年 4 月 27 日

なお、テキスト・文章系の著作物においては、一般的に、引用を条件に利用ライセンスを認める場合も多いが、これに関連して、ChatGPT を搭載した Bing AI では、引用を表示している点が留意すべき点であると考えられる。

④個人情報・企業秘密

＜学習データとしての個人情報の利用＞

従来の意思決定型の AI システムでは、学習データにおける個人情報の利用の在り方が大きな課題となっているが、生成系の AI システムにおいても、学習データの利用にあたって、上述の著作物だけでなく、個人情報の扱いの在り方についても問題になりうる。

具体的には、ChatGPT では、基本的には、インターネット上で公開されているデータを学習データとして利用することになるが、その中に含まれている個人情報については、その国・地域の法制度によっては、公開されているデータであっても、その利用にあたっては許可を得なければならないなど法に基づく対応が求められる場合もある。実際に、第 4−2 章で記載する通り、2023 年 3 月 31 日、イタリアの個人保護当局は、OpenAI 社に対し緊急暫定措置を講じたが、これは OpenAI 社が開発に利用した学習データに含まれるイタリア国民の個人情報の使用の停止を求めたものである[75]。

一般的に、個人情報は、個人の有する人権・権利に大きな影響を与えるために、適切な権利保護を行う必要があるが、デジタル時代において、既に公開されているような個人情報に対して、どのような権利を与えるべきかについては、著作権の在り方と併せて、今後検討の余地があると考えられる。

＜プロンプト入力等における個人情報、企業秘密の取り扱い（情報セキュリティ）＞

また、ChatGPT 等の利用にあたっては、そのユーザー情報と併せて、そのユーザーが入力（プロンプト）する個人情報や営業秘密に係るデータの取り扱いについても関心が高まっている。

具体的には、これらの入力された個人情報や営業秘密が、学習データとして再利用されることなどによって、偶発的に外部に漏出するリスクである。実際にこのような観点から、社内の機密情報は ChatGPT に入力しないよう警告を出している企業もある[76]。この点に関しては、OpenAI 社は、2023 年 3 月 1 日から、ChatGPT に連携させることができる API の提供を有料で開始し、この API 経由で入力されたデータは ChatGPT の学習などには使わないと説明している。

---

https://www.nikkei.com/article/DGXZQOUF27CJV0X20C23A4000000/
クリエーターと AI の未来を考える会「画像生成 AI の適正使用及びそれに伴う著作権制度の整備等に関する提言（第 2 版）」など
https://support-creators.com/

[75] WIRED「AI の学習データに含まれる個人情報が、ChatGPT にとって“大問題”になる」2023.04.07
https://wired.jp/article/italy-ban-chatgpt-privacy-gdpr/

[76] AERAdot（平和博）「社員が機密情報を ChatGPT に入力、上司の知らぬ間に漏洩も　生成 AI の安全対策は可能？」2023/04/20
https://dot.asahi.com/dot/2023041900015.html?page=3

なお、このような個人情報等の取り扱いのみならず、AI システムを取り扱う企業においては、情報セキュリティ対策が求められる。実際に、OpenAI 社は、2023 年 3 月 24 日、ChatGPT のユーザーに係る個人情報等が漏洩した事象があった旨、公開している[77]。

（4）ChatGPT 型 AI システム利用に係るガイドライン等

＜生成系 AI システムの利用に係るガイドラインなど＞

これらの社会的リスクについては、政策当局が規制・制度の改革を通じて取り組むべきものや、開発者・サービス提供者などが自主的に取り組むべき課題もあるが、一方、利用者側において取り組むべき課題もある。

実際に、ChatGPT の利用が急速に拡大し、誰でも ChatGPT を利用することが可能となる中、多くの識者に加えて、人工知能学会[78]や日本ディープラーニング協会[79]などが、その利用にあたっての基本的スタンスやガイドラインなどを発表しつつある。なお、東京大学[80]、CDT（Center for Democracy and Technology)[81]などが発表している利用に係る留意事項などにおいても、概ね前述の 4 項目があげられている。

このような中、日本では、政府においても、このようなガイドラインを作成しようとする動きがある。具体的には、総務省は、2023 年 4 月 27 日、有識者会議において、ChatGPT 型の AI システムの利用に関し、その特徴や課題を踏まえて、能力の習得に役立つ動画やテキストなどのコンテンツを今年度中に開発する意向を明らかにしている[82]。また、デジタル庁の主催するデジタル社会推進会議幹事会は、2023 年 5 月 8 日、書面開催を行い、各省庁における ChatGPT 等の生成 AI の業務利用に関する申し合わせを決定しており[83]、それを踏まえ、総務省自治行政局は、同日、地方自治体に対して事務連絡を発出している[84]。

---

[77] YahooJAPAN ニュース（ITMedia News）「ChatGPT で個人情報漏えい　OpenAI が原因と対策を説明」3/25(土)
https://news.yahoo.co.jp/articles/c5e8865b92c82dc6494bfe2f1d9bf475b633f74d

[78] 人工知能学会倫理委員会「人工知能学会としての大規模生成モデルに対してのメッセージ」投稿日：04/25/2023
https://www.ai-gakkai.or.jp/ai-elsi/archives/info/%E4%BA%BA%E5%B7%A5%E7%9F%A5%E8%83%BD%E5%AD%A6%E4%BC%9A%E3%81%A8%E3%81%97%E3%81%A6%E3%81%AE%E5%A4%A7%E8%A6%8F%E6%A8%A1%E7%94%9F%E6%88%90%E3%83%A2%E3%83%87%E3%83%AB%E3%81%AB%E5%AF%BE%E3%81%97%E3%81%A6

[79] 日本ディープラーニング協会「JDLA が、『生成 AI の利用ガイドライン』を公開」2023-05-01
https://www.jdla.org/news/20230501001/

[80] 東京大学理事・副学長（教育・情報担当）太田邦史「生成系 AI(ChatGPT, BingAI, Bard, Midjourney, Stable Diffusion 等)について」2023 年 4 月 3 日
https://utelecon.adm.u-tokyo.ac.jp/docs/20230403-generative-ai

[81] CDT (Center for Democracy and Technology) “Generative AI Systems in Education – Uses and Misuses”, March 15, 2023
https://cdt.org/insights/generative-ai-systems-in-education-uses-and-misuses/

[82] Yahoo JAPAN ニュース（朝日新聞 Digital）「生成 AI ユーザーに必要な能力とは　総務省が啓発コンテンツ開発へ」4/27(木)
https://news.yahoo.co.jp/articles/513b610b7afd9223a058d9af9f292fe899343970
総務省「ICT 活用のためのリテラシー向上に関する検討会（第 7 回）※青少年 WG（第 4 回）合同配付資料」令和 5 年 4 月 27 日（木）
https://www.soumu.go.jp/main_sosiki/kenkyu/ict_literacy/02ryutsu02_04000405.html

[83] デジタル庁「第 8 回デジタル社会推進会議幹事会・書面開催」令和 5 年 5 月 8 日（月）

＜研究・教育分野での利用に係る各団体のガイドライン＞

その中でも、特に研究・教育分野では、人間による文章の作成能力の向上とその能力の評価がなされる分野であることから、ChatGPT の利用によって大きな影響を受けることが見込まれている。このため、研究・教育に係る団体は、ChatGPT などの生成型 AI システムの利用に係るガイドラインを次々と発表している。

まず、研究分野では、特に AI 研究関連の学会が、初期の段階から、その発表論文等における主に著作権（オーサーシップ）の観点からの利用に係るガイドライン・制限規定等を発表しはじめ、その後、著名学会を含め多くの学会もそれぞれのガイドライン等を発表してきている。具体的には、ICML（International Conference on Machine Learning）は、2023 年 1 月に、生成 AI での論文の執筆の禁止を発表[85]、また、計算言語学会（ACL2023）[86]でも、同月、生成型 AI を利用した場合、その旨申告すべきとのポリシーを発表している。また、その後、Science も、同月、生成 AI のみの執筆による発表の禁止を発表[87]するとともに、Nature も、同月、AI により生成された文章や AI を著者にすることを禁止する旨発表している[88]。なお、ここで欧米においては、ChatGPT は研究論文の共著者になり得るか？[89]という論点に関心が集まっていることが興味深い。もちろん実際に ChatGPT を共著者として提出してくる研究者が存在し、また、多くの研究者は ChatGPT を共著者とすることに不支持を表明しているが、その背景の一部に、ChatGPT に人格があるとみなす文化があることが推察される。

また、教育分野においても、国内外で生成系 AI の利用に係る制限規定あるいはガイドライン等が発表されている。具体的に、国内では、東京大学が 2023 年 4 月 3 日に、生成系 AI のみを用いたレポート等の作成の禁止を発表し[90]。また、その後、多くの国内大学での発表

---

https://www.digital.go.jp/councils/social-promotion-executive/councils/191f444c-37fe-4c38-9909-09d9ccdb23af/

[84] 総務省自治行政局デジタル基盤推進室「ChatGPT 等の生成 AI の業務利用について」令和５年５月８日
https://www.soumu.go.jp/main_content/000879561.pdf

[85] Gigazine「ChatGPT などの AI で科学論文を書くことが国際会議で禁止に、ただし自分の文章の編集・推敲は OK」2023 年 01 月 06 日
https://gigazine.net/news/20230106-chatgpt-ai-writing-tool-banned/

[86] ACL 2023 Policy on AI Writing Assistance
https://2023.aclweb.org/blog/ACL-2023-policy/

[87] 読売新聞オンライン「「チャットＧＰＴ」使用巡り科学誌の対応割れる…サイエンスは禁止、ネイチャーは明記求める」2023/04/06
https://www.yomiuri.co.jp/science/20230406-OYT1T50193/

[88] Nature Digest「ChatGPT と類似ツールの利用に関する Nature の基本原則」2023-01-24
https://www.natureasia.com/ja-jp/ndigest/v20/n4/ChatGPT%E3%81%A8%E9%A1%9E%E4%BC%BC%E3%83%84%E3%83%BC%E3%83%AB%E3%81%AE%E5%88%A9%E7%94%A8%E3%81%AB%E9%96%A2%E3%81%99%E3%82%8BNature%E3%81%AE%E5%9F%BA%E6%9C%AC%E5%8E%9F%E5%89%87/120071

[89] Nature Digest「ChatGPT は研究論文の共著者になり得るか？」2023-01-18
https://www.natureasia.com/ja-jp/ndigest/v20/n4/ChatGPT%E3%81%AF%E7%A0%94%E7%A9%B6%E8%AB%96%E6%96%87%E3%81%AE%E5%85%B1%E8%91%97%E8%80%85%E3%81%AB%E3%81%AA%E3%82%8A%E5%BE%97%E3%82%8B%E3%81%8B%EF%BC%9F/120076

[90] 東京大学理事・副学長（教育・情報担当）太田邦史「生成系 AI(ChatGPT, BingAI, Bard, Midjourney, Stable Diffusion 等)について」2023 年 4 月 3 日
https://utelecon.adm.u-tokyo.ac.jp/docs/20230403-generative-ai

が相次ぎ、2023 年 5 月 2 日現在で、55 の大学が、生成 AI に係る対応を表明していることが報告されている[91]。

なお、日本政府は、2023 年 4 月 6 日、文部科学省において、ChatGPT 等の利用に関して学校現場が主体的な判断をする際に参考となる資料をとりまとめる方針であることを表明している[92]。

[91] note（森本）「ChatGPT/生成 AI への対応を表明した国内の大学一覧」【2023 年 5 月 2 日更新】
https://note.com/pogohopper8/n/n3126b312f209

[92] 日本経済新聞「学校の ChatGPT 指針、「文科省が策定」　松野官房長官」2023 年 4 月 6 日 12:17
https://www.nikkei.com/article/DGXZQOUA062FS0W3A400C2000000/
時事ドットコムニュース「学校向けにチャット GPT 指針　規制や活用の留意点明記　文科省検討」2023 年 04 月 06 日
https://www.jiji.com/jc/article?k=2023040600517

４．ChatGPT 等を巡る世界の AI 規制・ガバナンス政策動向

本章においては、まず第 4-1 章として、今回の第四次 AI ブームに伴う、将来的な AGI に対する不安感・不信感の高まりを背景にした動きとして、2023 年 3 月の FLI による公開書簡の発表とそのインパクトを取り上げる。

その上で、第 4-2 章において、AI 規制に積極的な EU の AI 法案等における動向についてとりあげる。その際、FLI などが提案し、ChatGPT が発表される前から議論されてきた「汎用システム」に係る規制の議論に加え、ChatGPT 発表後における欧州議会における ChatGPT に対する規制に係る最新の検討動向について、まとめる。

第 4-3 章においては、ChatGPT 発表後における、米国、英国、日本、そして中国における AI 規制・ガバナンス動向について、まとめる。

4-1. FLI による公開書簡の発表とその分析

（１）FLI による GPT 規制提言の発表とそのインパクト

＜FLI による GPT 開発中断に係る公開書簡の発表＞

2022 年末以降、ChatGPT への関心が爆発的に高まる中、FLI は、2023 年 3 月 22 日付けの公開書簡（オープンレター）を、同年 3 月 28 日に公表[93]し、世界的に大きな話題となった。第二章に記載の通り、FLI は、設立当初から、2017 年のアシロマ AI 原則を含めていくつかの公開書簡を発表してきたが、2018 年の致死型自律兵器に係る公開書簡の発表以来、５年ぶりの発表となる[94]。

その内容は、日本でも多く報道がなされている[95]が、以下の２つがポイントとなっている（詳細は、別添参考参照）。

- すべての AI 研究組織は、ChatGPT-4 よりも強力な AI 開発を少なくとも 6 ヵ月間は停止すべきであり、直ちに中断を求める。もし一時中断ができない場合は政府が介入してモラトリアムを実施する必要がある。
- 現状では AI 研究組織は適切なリスク管理をしておらず、制御不能な開発競争に陥っている。休止期間中、AI 研究組織は第三者独立機関において監査を受け、安全性を確証するべきである。

---

[93] Future of Lige, “Pause Giant AI Experiments: An Open Letter”
https://futureoflife.org/open-letter/pause-giant-ai-experiments/

[94] FLI Open Letters
https://futureoflife.org/fli-open-letters/

[95] YahooJAPAN ニュース（Business Insider Japan）「GPT-4 以上の AI の開発停止を求める公開書簡【内容まとめ】…マスクやウォズニアックなどが署名（海外）」3/30(木)
https://news.yahoo.co.jp/articles/f0054dd1a4696a88475a6c5ddeb6e7eb3856e5fc
Courier, 「AI 開発の「停止要請」はすでに“手遅れ”？ 理解を深める 5 つの質問」2023.3.31
https://courrier.jp/cj/321266/

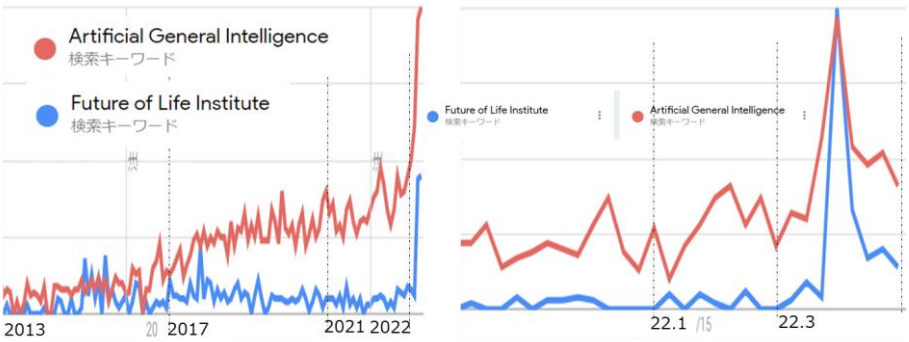
本書簡には、FLI の共同創業者であるマックス・テグマーク氏（MIT 教授）、ヤン・タリン氏（Skype 共同創業者）に加え、同 FLI の支援者であるイーロン・マスク氏などが署名していることで話題になっている。また、それ以外にも、「ディープラーニングのゴッドファーザー」[96]の 1 人とされるヨシュア・ベンジオ氏（モントリオール大学教授）、スティーブ・ウォズニアック氏（アップル共同創業者）、スチュアート・ラッセル氏（カリフォルニア大学バークレー校教授）などの著名人が、初期署名者として名を連ねている[97]。ただし、これらの有名人の多くが自ら積極的に本書簡に強く賛同しているのかは不明である。例えば、イーロン・マスク氏は、同氏の Twiiter アカウントで、本公開書簡を紹介することさえもしておらず、Tweet の大半は、Twitter の運営や Tesla、SpaceX などの話題が大半であり、そもそも AI に係る Tweet は非常に少ないのが実態と指摘される。また、その他の初期署名者に関しても同様のことが指摘されている。

いずれにせよ、この公開書簡は、イーロン・マスク氏が積極的に支援するような有名な非営利団体が、AI に対する将来的な不安感・不信感を踏まえて、予防原則的な観点から、ある意味極端なアプローチ１として、開発の一時中断まで踏み込んだ提言を行ったことが特徴であると言える。これまで AI の危険性を謳い、そのための原則論を提言する提案などは多くあったが、それなりに有名な非営利団体が「懸念」のみを理由として開発中断まで踏み込んだのは、AI の歴史で初めてであると考えられる。これは、ChatGPT に対する社会的な関心とその将来的な不安感・不信感が世界的に拡がっていることが背景にあることを示しており、また、その結果、本書簡は、社会に対して大きなインパクトを与えたと考えられる。

＜同公開書簡の発表のインパクト＞

実際に、この公開書簡はメディア等を通じて世界的に大きく報道され、その結果、世界での FLI に対する検索数は、発表後において急激に増加するとともに、それと併せて AGI に対する検索数も急激に増加している（図１７参照）。

【図１７】AGI, Future of Life Institute の検索数推移（世界）[98]



[96] なお、AI のゴッドファーザーと呼ばれる、ヨシュア・ベンジオ氏、ヤン・ルカン氏、ジェフリー・ヒントン氏３名のうち、ルカン氏は同書簡に署名をしており、また、ヒントン氏は、当初署名をしていなかったが、5 月に Google を退社し、AI 開発に警鐘をしたことが報じられている。
ITMedia NEWS「“AI のゴッドファーザー”ことヒントン博士、Google を退社して AI 開発に警鐘」2023 年 05 月 03 日
https://www.itmedia.co.jp/news/articles/2305/03/news051.html

[97] Yahoo Japan ニュース（平和博）「GPT-4 は社会と人類へのリスク」1,700 人超の専門家らが指摘する、そのリスクの正体とは？」3/31(金)
https://news.yahoo.co.jp/byline/kazuhirotaira/20230331-00343570

[98] 出典：グーグルトレンド（2023 年 5 月 3 日現在）

また、同書簡発表時点の 2023 年 3 月 28 日時点では、その署名者数は 1,125 名であったが、約２週間後の 2023 年 4 月 12 日現在で、20,507 名が署名しており、現在でもまだ多くの新規署名が続いている（なお、5 月 11 日現在で、27,565 名が署名）。これまでの FLI の公開書簡と比較しても、比較的署名数が多いものと評価される[99]。

なお、FLI が米国及び欧州を拠点とする団体であることを踏まえる必要があるものの、署名リストの上位に記載のある初期署名者の肩書及び所属をみると、概ね８割方は、欧州又は米国の大学又は研究機関の所属である。なお、その後は、世界各地の一般の技術者、人々が署名を行っているが、上記 20,507 名の中で、氏名の名称より日本人と判断される者は、概ね 37 名であり[100]、全体の約 0.18%にしか過ぎない。もちろん、日本人における英語能力の問題などはあると考えられるものの、世界的にみれば、非常に少ないことが特徴的である。

（２）FLI の開発中断提案の内容分析と反応

＜FLI の開発中断提案に見られる AGI に対する考え方＞

今回の FLI の開発中断が社会的に関心を集めたのは、ChatGPT に見られるような AI の発展を目の当たりにし、改めて、多くの人々が将来的な AGI に対する不安感・不信感あるいは脅威を感じたことが背景にあると考えられる。

ただし、その不安感・不信感あるいは脅威とする見方は、ChatGPT そのものがもたらす「現実的な」リスクに対するものというよりは、特に欧州を中心とする AGI に対する「将来的な」不安感・不信感をベースにしていると言うことができる。実際に、本公開書簡（オープンレター）によると、アシロマ AI 原則の 20 番「高度な AI は地球上の生命の歴史に大きな変化をもたらす可能性があり、相応の注意とリソースで計画および管理する必要があります」を引用した上で、問題点として、以下の４点を指摘している。

- a) 機械によるプロパガンダや虚偽
- b) 充実した仕事を含め、すべての仕事を自動化
- c) 私たちに取って代わる可能性のある非人間的な心の開発の必要性
- d) 文明の制御を失う危険

これらの指摘される問題点に関し、まず、ChatGPT が、アシロマ AI 原則 20 番の「高度な AI」に相当するというのは事実であるとは考えられるものの、同原則 22 番で安全管理を厳格化すべきとした「再帰的に自己改善もしくは自己複製を行える人工知能システム」には全く達していないと考えられる。

その上で、これらの４点の指摘は、a) の一部を除き、基本的には、ChatGPT によって直接的にもたらされる現実的なリスクが指摘されている訳ではなく、むしろ、「「機械」は人間に対して悪影響をもらす可能性がある」という文化的な認識・前提を背景とした、将来的な不安感・不信感に基づいているものと考えられる。

具体的には、a) では、情報チャネルが（機械・AI によって）虚偽やプロパガンダで溢れてしまう可能性を指摘している。これは、前章（３）の①の正確性・信頼性の側面であるこ

[99] FLI Open Letters
https://futureoflife.org/fli-open-letters/
[100] 筆者カウントによる（4 月 12 日時点）

とは確かであるものの、同章にも記載した通り、悪意ある情報の流通・氾濫は、まずは、悪意ある人間の存在が問題であり、現時点でも、コピペだけでも十分起こりうる問題であること、また、AI・Chatbot というよりはむしろインターネット・SNS に係る問題であることに留意することが必要である。

また、b) に係る仕事の将来の話は、多くの社会・市民の大きな関心事項であり、実際に、ChatGPT は、ホワイトカラーを含め、米国では８割の労働者に影響を与えるとの試算も公表され[101]、多くの人々が不安感を感じているのは事実であると考えられる。しかしながら、このような仕事の将来の話は、ChatGPT に限る話ではなく、7～8 年前に第三次 AI ブームが流行ったときも同様に大きな話題になったものであり、また、AI に限らず過去の革新的な技術によるイノベーションの歴史において、これまでの必ず生じてきた問題であるとも言える（本 WP においては、この点は深入りしない）。

その上で、特に c) 、d) は、AGI に対する将来的な不安感・不信感に係る代表的な記述であり、AGI に対する FLI の思想的なスタンスであると言うことができよう。特に、c) については、人間のみが「心」を有し、「心」を有さない機械に対する不信感がベースにあり、また、d) については、人間のみが文明を制御し、機械は文明を破壊するという思想が前提にあると考えられる（図１８参照）。

以上をまとめると、FLI の公開書簡による提言とは、基本的には、欧州を中心とする AGI に対する不信感・不安感をベースに、第一章で示したような、予防原則的な観点から、ある意味極端なアプローチ１の方針を採用したものと言える。ただし、完全使用禁止に踏み込むことはできず、一方で、現時点では、ChatGPT の普及によって生じる具体的なリスクも明確になっていないことから、利用形態に即した現実的な安全性に係る規制を提案することはできず、したがって、半年のモラトリアムということで提言したものと推測される。

【図１８】FLI の指摘（懸念）事項とその背景／ChatGPT の特性の実態

FLIの指摘（懸念）事項	懸念の背景・前提 （AIの擬人化と人間との対置）	ChatGPTの特性 （道具としてのAI）
①機械によるプロパガンダや虚偽	• 人間は、プロパガンダ、虚偽の生成を抑制する。 • 一方、機械（AI）は、プロパガンダ・虚偽を生成する可能性がある。	• 人間の指示に従い、フィクションを生成する機能も有する。 • ただし、必ずしも正確な回答でない可能性がある。
②充実した仕事を含め、全ての仕事を自動化	• 人間は、充実した仕事を実施。 • これに対し、機械（AI）が、それを奪う可能性。	• これまで人間が行ってきた業務を代替する可能性がある（他の革新的技術と同様）。
③私たちに取って変わる可能性のある非人間的な心の開発の必要性	• 人間のみが「心」を有する。 • 人間の心を有さない機械（AI）が人間にとって代わる可能性	• 人間の利用する高度な道具の開発を目指しているもの。 • 非人間的な「心」の開発を目指しているのではない。
④文明の制御を失う危険	• 人間は文明を制御してきた。 • これに対し、機械が文明を破壊し、する可能性。	• 人間の問いかけに対応する道具にしか過ぎない。 • 文明は、人間の開発した技術・道具とともに発展。

＜FLI 提言の開発中断の合理性の検討＞

本公開書簡においては、上述のような認識のもと、その具体的な方策として、少なくとも半年間の開発中断と、その間の第三者独立機関での安全性の確証を求めている。このうち、

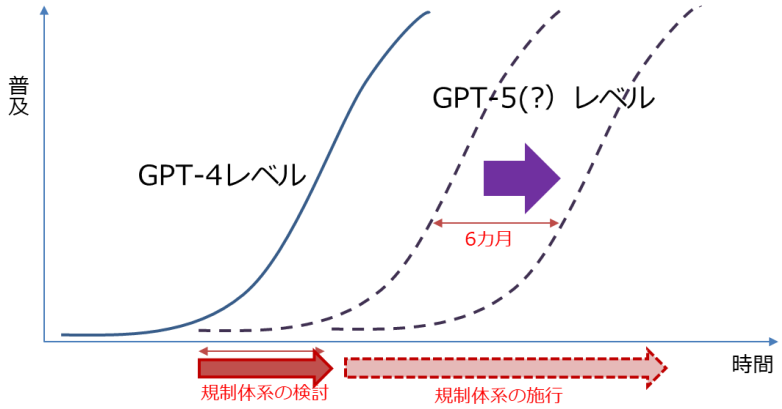
[101] 日本経済新聞（編集委員吉川和輝）「ChatGPT の衝撃　米労働者の 8 割に影響も」2023 年 4 月 4 日
https://www.nikkei.com/article/DGXZQOCD315CJ0R30C23A3000000/

後段の第三者独立機関の必要性については、多くの人がその必要性を認識する一方で、特にこの半年間の開発中断という手法が、目的達成のために合理的なのかという論点について考察する。具体的には、その半年間開発の中止を行い、その間に安全対策を行うことにより、どれだけの被害が防げるか（リスクを低減できるか）という論点である。

この論点については、今回の提言は、利用の停止ではなく、開発の停止であることを踏まえると、まず単純に、仮に半年間の開発の停止をしたとしても、既存の GPT-4 の普及とその利用は引き続き進展することになり、その結果、GPT-4 の利用によるリスク・被害（仮に存在するとして）を減らすことはできないと考えられる。

一方、半年分の開発を遅らせることにより、次世代の GPT（仮に GPT-5 とする）の利用開始と普及は、単純には半年遅れることになる。その意味で、GPT-5 の利用による被害は（仮に存在するとして）、それまでに検討・施行される安全規制により対応される可能性はある。しかしながら、規制体系の検討にあたって、GPT-5 にも併せて適用されるような体系を検討しさえすれば、半年の開発停止という措置を取らなくてもほとんど変わらないという見方も可能である（図１９参照）。

【図１９】FLI の規制（開発中断）提案に係る概念図[102]



結局のところ、今回の FLI の提言は、本来は GPT-4 による被害の未然防止等の観点から GPT-4 以下においても利用の禁止を行うべきとは考えたものの、既に開発・発表され世界中の多くの人が利用し始めている GPT-4 を急に利用禁止にすることは世の中の理解を得られないこと、一方で、世界の多くの人々は GPT をはじめとする AI 開発の急速な進展に戸惑っているという現状を踏まえ、一旦立ち止まって規制体系の検討をするという提案であれば世の中に受け入れられるとの判断のもとでなされた可能性がある。なお、FLI の公開書簡では、最終的には政府による介入を求めているが、民主主義・法治国家では、法令的根拠がなければ、規制はできないと考えられる。

このようなことを踏まえると、本来は、半年間の開発中断という手段よりは、むしろ対策のための規制体系の検討に注力を注ぐことこそが重要であると考えられる[103]が、それにも関わらず、FLI が半年間の開発中断を提言したのは、世間に対して将来的な AGI に対する警鐘をならすべく、インパクトのある内容としたかったためである可能性も考えられる。

---

[102] 出典：筆者作成

[103] ITMedia NEWS「AI 開発停止要求署名は無意味、透明性と説明責任の改善を――Hugging Face のルッチョーニ博士」2023 年 04 月 05 日
https://www.itmedia.co.jp/news/articles/2304/05/news067.html

＜FLI の規制提案に係る関係者の見方（ポジショントーク）＞

いずれにせよ、今回の FLI の公開書簡は、開発中断まで踏み込んだことが特徴であるが、その際、GPT-4 以上に限定しているがゆえに、産業・技術競争の観点からも含めて、賛否両論を巻き起こし大きな話題となった。（なお、この「GPT-4 よりも強力」ということに関して、そもそも定義が不明との AI 専門家の指摘もある[104]。）

すなわち、産業・技術競争の観点からすれば、開発競争において先行している OpenAI 及びその支援者である Microsoft や、その後を追いかける Google にとっては、高度な AI の開発の中断が求められることにより先行者メリットを失うことになるが、一方で、キャッチアップしようとしているその他の多くの AI 企業（中国などの海外の AI 企業を含む）においては相対的な競争優位が得られることになる。

このため、この FLI の公開書簡に関しては、FLI の主要支援者であるイーロン・マスク氏のポジショントーク的な位置づけであるとの指摘も少なくない。すなわち、イーロン・マスク氏が以前投資をしていたものの、その後その投資を引き揚げてしまった OpenAI 社が大成功したことに対するライバル視が背景にあるのではないかとも指摘されている。実際に、イーロン・マスク氏は、GPT4 以上の開発一時中断を求める一方で[105]、2023 年 4 月には、同氏が率いる X 社（旧 Twitter 社）において、大規模言語モデル（LLM）に係る AI 開発プロジェクトを始めたとの報道がなされ[106]、その後、実際に同社において「トゥルース GPT」を開発する予定だと語っている[107]。

一方で、Microsoft や Google に関連する人にとっては、ポジション的には、FLI の公開書簡に対して、否定的・疑念的視点を有することが少なくない。実際に、元マイクロソフトのビル・ゲイツ氏は、雑誌のインタビューで、「ある特定の集団に停止を求めることで課題が解決されるとは思わない」「必要なのは注意すべき分野を特定することだ」と述べたとしている[108]。また、元 Gogle のエリック・シュミット氏は、AI には「適切なガードレール」が必要だとしながらも「6 カ月の一時停止は、単に中国を利することになるため、賛成できない」と語ったとしている[109]。

---

[104] ロイター「ＡＩ専門家ら、マスク氏らの公開書簡に懸念」2023 年 4 月 1 日
https://jp.reuters.com/article/elon-musk-ai-academics-idJPKBN2VX1M5?il=0

[105] Barrons, ”Tesla Needs AI to Thrive. Why Elon Musk Wants to Pause GPT-4.”, Updated March 30, 2023 10:13 am ET / Original March 29, 2023 6:23 pm ET
https://www.barrons.com/articles/tesla-elon-musk-tesla-ai-letter-16f6df7b
バロンズ・ダイジェスト「マスク氏はなぜ AI 開発休止を呼び掛けたのか」2023/03/30
https://barrons.jiji.com/article/4538

[106] Yahoo Japan ニュース（ITMedia News）「イーロン・マスク氏、生成 AI 用に約 1 万個の GPU を購入か　AI 開発の停止要請に署名したばかり 4/12(水)
https://news.yahoo.co.jp/articles/978158147e8617c52d0ace043373f698e16c52e8

[107] Yahoo Japan ニュース「マスク氏、「トゥルースＧＰＴ」の立ち上げ表明　チャットＧＰＴに対抗＝報道」4/18(火) 7:09 配信
https://news.yahoo.co.jp/articles/827befc55e5a6a26c2fa923e189b04c567b15874

[108] Ledge.AI「ビル・ゲイツ氏｜AI 開発停止要請「なぜ停止すべきかわからない」」2023 04 06 THU
https://ledge.ai/bill-gates_ai-development/
YahooJAPAN ニュース（Forbes Japan）「ビル・ゲイツ「AI 開発の一時停止」に反発、効果を疑問視」4/5(水)
https://news.yahoo.co.jp/articles/f8420c911a2895f730919e7b6d63e27a75931524

[109] YahooJAPAN ニュース（Forbes Japan）「グーグル元 CEO が語る「AI 開発の一時停止が危険な理由」、2023 年 4 月 12 日
https://news.yahoo.co.jp/articles/e8aca503831733ffe3e3d6b96f49c62f02ad048f

（３）OpenAI における開発スタンス

＜FLI 提言に対する OpenAI の反応＞

一方、FLI の提言の現実的可能性については、まずは、OpenAI 社が自主的に GPT-4 以降の開発を中断することが必要となる。これに関して、同社は、GPT-4 を発表したばかりであり、それ以降の開発についてはまだ何も決まっていないとのコメントのみしかしていない。ただし、OpenAI 社は、FLI の公開書簡の発表の１週間後の 2023 年 4 月 5 日、「AI の安全性に対する当社のアプローチ」と題するブログを公開した[110]。本ブログでは、GPT-4 の公開に際しては、トレーニング完了後、組織全体で半年以上かけて安全で整合性のとれたものにするための作業を行い、外部の専門家にもフィードバックを求めたことなどを説明しており、これは、FLI の提言に対応して、発表したものではないかと指摘されている。

＜AGI の開発に向けた OpenAI の基本的な考え方＞

第二章に記載したとおり、OpenAI 社は、もともと人類に利益をもたらすような AGI の開発を目的に設立された団体である。その際、同社は、この ChatGPT の開発については、AGI の実現に向けた取組の一歩であると考えており、したがって、その安全性・リスクについては、当然ながら自ら必要な対策に取り組んでいると自負している。

その具体的な AGI に向けた必要な対策の手順に係る考え方、取組の方向として、同社の CEO である Sam Altman 氏は、2023 年 2 月 24 日付けで、「AGI とその先に向けた計画」[111]を発表している。同計画によると、AGI は上手く生成できれば全ての人々に新たな能力を与える潜在性を有する一方で、不正利用、大規模事故、社会破壊などの深刻なリスクが生じることがあるとして、まずは、短期的には、以下の３点の原則を掲げている。

- ①まず、現実世界での運用を行い、経験を得て、迅速なフィードバックを進めながら、漸次的に導入を進めること
- ②「初期設定」は制限し、ユーザーによる調整を可能するなど、より現場に合わせた調整可能なモデルとして開発すること
- ③「システムの統治」、「便益の公平な分配」、「公平なアクセスの確保の３つの課題に関して、世界的な対話を進めること。

その上で、「ある時点で、独立したレビューを受け、新たなモデル開発のスピードを落とすことが必要かもしれない」[112]とし、長期的には、「人類の未来は人類が決定すべき」で

---

[110] Our approach to AI safetyEnsuring that AI systems are built, deployed, and used safely is critical to our mission.
https://openai.com/blog/our-approach-to-ai-safety
OpenAI、「AI の安全性に対する当社のアプローチ」を説明　「年齢確認オプション検討中」2023 年 04 月 06 日
https://www.itmedia.co.jp/news/articles/2304/06/news084.html

[111] Sam Altman (OpenAI) , Planning for AGI and beyondFebruary 24, 2023
https://openai.com/blog/planning-for-agi-and-beyond
Yahoo Japan ニュース（ITMedia News）「OpenAI、AGI（人間より賢い AI）へのロードマップを公表　「世界に深刻な害を及ぼす可能性」回避のために」2023 年 2 月 27 日
https://news.yahoo.co.jp/articles/06ad66d1a8a11626f3936d1bf393e59a5e625a0c

[112] We think it's important that efforts like ours submit to independent audits before releasing new systems; we will talk about this in more detail later this year. At some point, it may be important to get

あり、したがって、AGI の構築には、多大な精査と、主要な意思決定に係る市民との協議が必要としている。

このように、まずは現実世界で運用し、その後フィードバックループを回しながら進めていくというアプローチは、現実的な取組（アプローチ１に相当）と解釈できる一方、AGI に対して不安感・不信感を有するグループからは、不十分だとの指摘を受ける可能性がある。実際に、上述の FLI の公開書簡では、「その『ある時点』とは、今である」との指摘のもとで、対応の必要性を強調している。

4-2. 欧州 AI 法案等における規制動向

（１）第四次 AI ブーム以前：EU 理事会での汎用 AI（GPAI）規制を巡る動き

＜当初欧州 AI 法案における関連規制＞

欧州委員会（EC）は、2021 年 4 月、AI 法案を発表した。この AI 法案では、主な規制対象として、第三次 AI ブームの結果主流となったいわゆる従来型の意思決定型の AI システムのうち、特に安全性に関わるもの、及び、公平性に関わるものを中心に、ハイリスク型の AI システムと位置づけ、事前のリスク評価、適合性認証などの義務付けを行っていることが特徴である（図２０参照）。具体的には、ハイリスク AI システムとして、

- 安全性に関わるもの：規制対象製品の安全要素、重要インフラの管理・運用
- 公平性に関わるもの：教育と職業訓練、雇用、労働者管理等、必須の民間・公共サービス、法の執行、移住・亡命・国境管理、司法運営と民主プロセス

を対象としている（ただし、その後、2022 年の欧州理事会の一般アプローチにより、今後一部修正が見込まれている）。

これに対し、今回の第四次 AI ブームにおいては、第一章で記載した通り、ChatGPT に代表されるような、コンテンツ生成型の AI システムに対して新たに関心が高まっているが、このような AI システムを検討対象にする場合、これまでの従来型の意思決定システムとは、リスクの種類が全く異なるため、本来は規制体系も改めて検討する必要があると考えられる。

ただし、この 2021 年 4 月に発表された欧州 AI 法案でも、コンテンツ生成系の AI システムの一部は、既に規制対象となっていた。具体的には、

- DeepFake：虚偽情報として普及されることに問題が生じることから、透明性の義務として、DeepFake である旨明示することを義務付け。
- ChatBot など（自然人と相互作用するシステム）：透明化義務として、対話する相手方が、実の人間ではなく、AI であることを明示することを義務付け。

の二つであり、いずれも「透明化」措置の対象として、比較的リスクの低い AI システムとして位置づけられている。

---

independent review before starting to train future systems, and for the most advanced efforts to agree to limit the rate of growth of compute used for creating new models. We think public standards about when an AGI effort should stop a training run, decide a model is safe to release, or pull a model from production use are important. Finally, we think it's important that major world governments have insight about training runs above a certain scale.

【図２０】欧州 AI 法案の概要（経産省資料）[113]

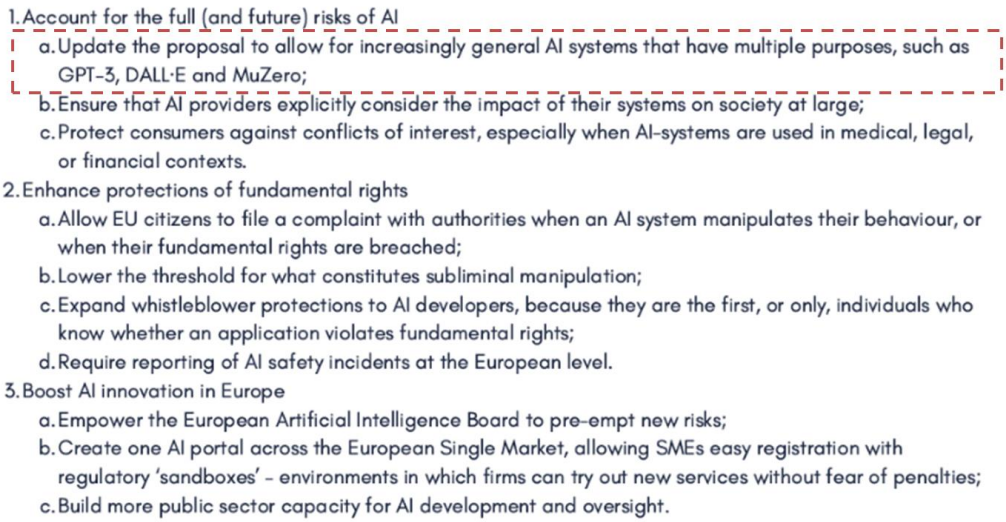
＜FLI 等による「汎用 AI システム」規制の働きかけ＞

この 2021 年 4 月に発表された欧州 AI 法案に関しては、欧州委員会（EC）はパブリックコメントを開始するとともに、議論の中心は、欧州委員会（EC）から、EU 理事会に移行することになる。これに対し、FLI は、当初から、GPT のような AI システムを「汎用 AI システム（General Purpose AI）」として位置づけ、規制対象に含めるよう積極的に働きかけており[114]、そのような取組がその後の欧州 AI 法案の修正議論に反映された可能性がある。

具体的には、FLI は、まず、欧州委員会が実施したパブリックコメントの募集に対し、2021 年 8 月にポジションペーパー（2021 年 8 月）[115]を提出・公表している。当該ペーパーでは、一番目の課題として、OpenAI 社の開発している GPT-3、DALL-E などを事例として取り上げ、このような「複数の目的を有する一般的な AI システム」を規制の対象とすべきとの提案を記載している（図２１参照）。このことからも、GPT-3 などの OpenAI 社の開発する AI システムに対する FLI の関心の高さが伺える。

[113] 出典：経産省　第 1 回 AI 原則の実践の在り方に関する検討会　資料 5 （2021 年 5 月 11 日）
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/2021_001.html
[114] Strengthening the European Union AI Act
https://futureoflife.org/project/eu-ai-act/
[115] FLI Position Paper on the EU AI Act
https://futureoflife.org/wp-content/uploads/2021/08/FLI-Position-Paper-on-the-EU-AI-Act.pdf?x76795

【図２１】FLI による欧州 AI 法案への提言（項目：2021 年 8 月）[116]



1. Account for the full (and future) risks of AI
   a. Update the proposal to allow for increasingly general AI systems that have multiple purposes, such as GPT-3, DALL·E and MuZero;
   b. Ensure that AI providers explicitly consider the impact of their systems on society at large;
   c. Protect consumers against conflicts of interest, especially when AI-systems are used in medical, legal, or financial contexts.
2. Enhance protections of fundamental rights
   a. Allow EU citizens to file a complaint with authorities when an AI system manipulates their behaviour, or when their fundamental rights are breached;
   b. Lower the threshold for what constitutes subliminal manipulation;
   c. Expand whistleblower protections to AI developers, because they are the first, or only, individuals who know whether an application violates fundamental rights;
   d. Require reporting of AI safety incidents at the European level.
3. Boost AI innovation in Europe
   a. Empower the European Artificial Intelligence Board to pre-empt new risks;
   b. Create one AI portal across the European Single Market, allowing SMEs easy registration with regulatory 'sandboxes' – environments in which firms can try out new services without fear of penalties;
   c. Build more public sector capacity for AI development and oversight.

この 2021 年 8 月に FLI が提出したポジションペーパーが欧州内での議論にどのような影響を与えたかは不明であるが、2021 年後半以降、実際に、EU 理事会における AI 法案に係る議論において「汎用 AI システム」に対する規制の議論が開始されることになる。具体的には、2021 年後半（スロヴェニア議長国時）には、新第 52 条 a として、汎用 AI システム（general purpose AI systems）に係る条項が設けられ、その後加盟国から当該ドラフト案に意見が加えられた。また、2022 年前半（フランス議長国時）においては、それらのコメントを踏まえて、新第 4a 条への移行が検討され、その後、欧州議会の各委員会での修正意見を踏まえて、さらに議論がなされたとされる[117]。

なお、当時、汎用 AI システムに対する規制提案を行っているのは FLI のみではない。例えば、オランダの責任ある AI に向けたイニシアティブのための独立機関である ALLAI[118]は、2022 年 2 月に、汎用 AI システムに対してコメントを行っている[119]。同コメントでは、汎用 AI システムの提供者に対して規制をかけないと、そのシステム利用者のみに（ハイリスク AI に係る）規制の負担が係ることになり、その結果、特に中小企業・ベンチャーなどのシステム利用者によるイノベーションが窒息することになり、特に、最先端の汎用 AI 技術に競争優位を有する米国・中国と比較して、欧州は、産業・イノベーション上不利になるという説明を行っている。

そのような中、FLI は、2022 年 5 月、「汎用 AI（General Purpose AI）と AI 法」という提言[120]を公表し、欧州側に提出している（なお、その後、2022 年 10 月には、その汎用 AI

[116] 出典：FLI Position Paper on the EU AI Act（点線枠組みは、筆者追加）
https://futureoflife.org/wp-content/uploads/2021/08/FLI-Position-Paper-on-the-EU-AI-Act.pdf?x76795
[117] Future of Life Institute, "General Purpose AI and the AI Act", May 2022
https://futureoflife.org/wp-content/uploads/2022/08/General-Purpose-AI-and-the-AI-Act-v5.pdf
[118] ALLAI：欧州の AI-HLEG のメンバーのうちの３名のオランダ人メンバーによる機関
https://allai.nl/
[119] ALLAI, "AIA in-depth #1 Objective Scope Definition Articles 1 - 4 & ANNEX I", 2022.2.13
https://allai.nl/wp-content/uploads/2022/03/AIA-in-depth-Objective-Scope-and-Definition.pdf
[120] Future of Life Institute, "General Purpose AI and the AI Act", May 2022
https://futureoflife.org/wp-content/uploads/2022/08/General-Purpose-AI-and-the-AI-Act-v5.pdf

システムの定義に係る論文も発表している[121]）。この中で、FLI は、汎用 AI の規制の必要性に関し、特に GPT-3 を例にあげて、例えば、過激主義者の主張を述べて警告を受けたり、反ムスリムのバイアスがあったり、うっかりと個人情報を公開したりする場合があったり、あるいは、対話型システムでは自殺をするようほのめかす場合もあったとしている[122]（なお、その後公表された ChatGPT ではこのような事例はあまり聞かれない）。そのような認識のもと、同提言では、汎用 AI システムに係る定義案を提示するとともに、汎用 AI システムの提供者に対して、欧州 AI 法案における「ハイリスク AI システム」の提供者と同様に、第三者認証を義務付けるなどの規制の対象とする案を提示している（図２２参照）。

【図２２】汎用 AI システムに係る規定（FLI の提案と EU 理事会による一般アプローチ）[123]

「汎用AI（GPAI）とAI法」（FLI、2022年5月）	AI法案に対する一般アプローチ）（欧州理事会、2022年12月6日）
第３条：「汎用AIシステム」の定義 an AI system that is able to perform generally applicable functions such as image/speech recognition, audio/video generation, pattern detection, question answering, translation, etc, and is able to have multiple intended and unintended purposes.	**第３条（1b）：「汎用AIシステム」の定義** an AI system that - irrespective of how it is placed on the market or put into service, including as open source software - is intended by the provider to perform generally applicable functions such as image and speech recognition, audio and video generation, pattern detection, question answering, translation and others; a general purpose AI system may be used in a plurality of contexts and be integrated in a plurality of other AI systems;
第4a条：汎用AIシステム提供者の義務 • 第二章（ハイリスクAIシステム）に係る要件の遵守、ハイリスクデータベース（第60条）への登録 • 合理的に予測できる悪用の評価、新たなリスクに係る定期的な評価 • ユーザー等に対する安全利用に係る手順・情報の提供 **第4b条：汎用AIシステムの適合性評価** • ・事前の適合性評価の確保 **第4c条：提供者の義務が他者に適用される条件** • 汎用AIシステムを市場・サービス提供し、利用する者にも本規則が適用される。	**第4a条：本規制での汎用AIシステムのコンプライアンス** • 第4b条に規定される要件と義務のみへの遵守。 **第4b条：汎用AIシステムの要件と当該システム提供者の義務** • 1. ハイリスクAIシステムあるいはその部品として利用される汎用AIシステムは、第二章（ハイリスクAIシステムの要件）の要件に適合しなければならない。 • 5. 汎用AIシステムの提供者は、ハイリスクAIシステムあるいはその部品として、欧州連合市場においてサービスを提供しシステムを設置しようとする提供者に対して協力し、必要な情報を提供しなければならない。このような汎用AIシステムの提供者による情報共有の実行のため、欧州委員会は実施法を採択することができる。 **第4c条：第4b条の例外** • 第4b条は、提供者が、全てのハイリスク利用を明確に排除している場合には適用されない。

＜欧州 AI 法案に対する EU 理事会の一般アプローチ（2022 年 12 月）＞

このような経緯を含めて、EU 理事会は、2022 年 12 月、欧州 AI 法案に係る修正方針である「一般アプローチ」を発表した[124]。その内容は、AI システムの定義の見直し、ハイリスク AI システム等の規制対象の見直し、規制内容の見直しなど多岐に亘るが、その一つの大きな項目として、汎用 AI（General Purpose AI）を新たに規制対象とする項目が追加されている。具体的には、汎用 AI（General Purpose AI）システムとして、第３条「定義」の（1）のあとに（1b）を設け、その定義を規定するとともに、新タイトル ⅠA「汎用 AI システム」を創設し、その中の条項として第 4b 条を設け、汎用 AI システムに対する規制内容について規定している（図２２参照）。

---

[121] Carlos I. Gutierrez, Anthony Aguirre, Risto Uuk, Claire C. Boine, and Matija Franklin, “A Proposal for a Definition of General Purpose Artificial Intelligence Systems”, Future of Life Institute – Working Paper
https://futureoflife.org/wp-content/uploads/2022/11/SSRN-id4238951-1.pdf

[122] . For example, GPT-3, which is trained to process natural language, unexpectedly acquired the ability to write rudimentary programs in a programming language.Some of these systems have already caused alarm by propagating extremist content,exhibiting anti-Muslim bias,or inadvertently revealing personal data.A chatbot based on the general purpose AI system GPT-3 told a person to commit suicide.

[123] 出典：筆者作成

[124] Council of the EU, Transport, Telecommunications and Energy Council (Telecommunications), 6 December 2022
https://www.consilium.europa.eu/en/meetings/tte/2022/12/06/

このうち、「汎用 AI（General Purpose AI）システム」の定義としては、基本的には、「提供者によって、画像・音声認識、音声・映像生成、パターン検出、質問回答、翻訳その他などの一般的に適用可能な機能を実行することを意図された AI システム」[125]としている。この定義は、詳細は異なるものの、原則として FLI の提言の内容と基本的には同様のものとなっている。また、その対象としては、ChatGPT にような「質問回答」（対話型）のシステムや、最近の画像系のコンテンツ生成システムも含まれるが、これまでの ChatBot やいわゆる AI スピーカーに加え、外国語翻訳システムや全般的に利用可能な画像認識システムなど、かなり幅広い AI システムが対象になるものと考えられる。

その上で、同アプローチでは、汎用 AI システムに係る要件と提供者の義務としては、既存の「ハイリスク AI システム」に該当するシステムの一部に利用される汎用 AI システムを規制対象とし、本規制のハイリスク AI システムの要件に係る適合義務と同様の義務を課すとともに、当該汎用 AI システムの提供者に対し、当該システムを利用して欧州でのハイリスク AI システムを設置しサービスを提供する事業者への協力や情報提供義務を課すとしている。また、その際、その後者の情報提供義務等に係る詳細は、欧州委員会が実施法を採択可能としている。

このような意味で、本アプローチにおける「汎用 AI システム」に関しては、FLI の提案のように、それ自体をハイリスクとして捉えるのではなく、むしろ ALLAI の提案のように、既に AI 法案の付属書Ⅱや付属書Ⅲに規定される「ハイリスク AI システム」に含まれる場合に限定したことが特徴と言える。具体的には、規制対象製品の安全要素、重要インフラの管理・運用など、安全性あるいは公平性の観点からハイリスクとみなされたシステムの一部として導入する場合には規制対象となるが、汎用 AI システム単体では必ずしも規制対象とはならないことになる。

これは、おそらく、ALLAI の指摘する通り、既存のハイリスク AI システムの規制を行うにあたって、その上流である汎用 AI システムも規制することによって、主に下流企業となる欧州企業のみに規制の負担が係ることを防ぐという考え方によるものと考えられる。しかしながら、この考え方は、あくまでも、安全性・公平性に係るリスクを対象に選定されたハイリスク AI システムへの対応を念頭においたものであり、その後、登場することになる ChatGPT などの生成型 AI システムによって生じるリスクを意識したものではないことに留意することが必要である。

（2）第四次 AI ブーム以降：欧州議会等での ChatGPT に対する規制の動き

＜欧州議会の動向：汎用 AI システムの扱いの見直し＞

上記の EU 理事会による欧州 AI 法案の修正方針（一般アプローチ）の発表を受けて、欧州 AI 法案に係る修正議論の中心は、欧州議会に移行することになる。

---

[125] 正確には、以下の通り。
AI system that - irrespective of how it is placed on the market or put into service, including as open source software - is intended by the provider to perform generally applicable functions such as image and speech recognition, audio and video generation, pattern detection, question answering, translation and others; a general purpose AI system may be used in a plurality of contexts and be integrated in a plurality of other AI systems;

その際、この一般アプローチが発表されたのは 2022 年 12 月 6 日であるが、OpenAI による ChatGPT が発表されたのが 2022 年 11 月 30 日であり、したがって、一般アプローチには、汎用 AI システムとして LLM を規制しようとする意図は含まれていても、ChatGPT の普及とその実態を踏まえた規制の検討は全く含まれていない。一方、その後の ChatGPT の爆発的な普及に伴い、欧州 AI 法案における ChatGPT に対する対応の在り方とその見直しに向けた議論が改めて求められることになる。

具体的には、欧州委員会のブルトン委員（域内市場担当）は、2023 年 2 月 6 日、ChatGDP のリスクについて警告し、今後策定される AI 法案において、欧州の人々が技術を信頼できるよう、こうしたリスクに対処することを目指すと説明したと報じられている[126]。一方、3 月 3 日付けの記事[127]によると、欧州議会のメンバーにおいては、現行の汎用 AI システムでは不十分であるとの認識の元で、様々な議論があることが報じられている。具体的には、人間の監視のない文章生成システムはハイリスクとみなすとともに、文章生成システム以外の他の汎用 AI システムも規制すべきでとの意見や、ChatGPT の開発者、利用者の両方に対して、リスク管理義務と透明性要件を課すべきとの意見がある一方で、そうすると全くリスクのないような多数の活動がハイリスクになってしまうとの懸念もあることを伝えている。

その上で、欧州議会で、3 月 14 日、汎用 AI（GPAI）システムに係る新たな規制案が示され、議論が行われたことが報じられている[128]。具体的には、汎用 AI システムの定義として（上述の前年 12 月の GPAI の定義とは全く異なり、）「広範囲データのスケールにより訓練され、出力の一般性を保つようにデザインされ、広範囲のタスクに採用される AI システム」[129]と定義され、リスクの管理、外部監査の義務付け、EU データベースへの登録の義務付けなどが規定されていると報道されている。

＜欧州議会の動向：多様な AI システムの規制強化、著作権の取り扱いの見直し＞

また、3 月 28 日付けの記事[130]によると、その時点において、欧州議会としては、ハイリスク AI システムの一部として組み込まれる汎用 AI システムだけではなく、単体として利用される汎用 AI システムについてもハイリスク AI システムとして規制対象とすること、また、新たにハイリスク AI システムとして新たに規制対象（付属書Ⅲのリストを拡充）となるものは、「汎用 AI システム」だけでなく、「生成 AI」、「Deep Fake AI」など、多数の AI

---

[126] ロイター「「チャットＧＰＴ」リスクにＡＩ規則で対処＝ブルトン欧州委員」2023 年 2 月 6 日
https://jp.reuters.com/article/eu-tech-chatgpt-idJPKBN2UG04F

[127] Politico, ”ChatGPT broke the EU plan to regulate AI”, MARCH 3, 2023
https://www.politico.eu/article/eu-plan-regulate-chatgpt-openai-artificial-intelligence-act/
European Parliamentary Research Service, “AT A GLANCE Digital issues in focus General-purpose artificial intelligence”, March 2023
https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2023)745708

[128] By Luca Bertuzzi | EURACTIV.com ”Leading EU lawmakers propose obligations for General Purpose AI” 2023 年 3 月 15 日
https://www.euractiv.com/section/artificial-intelligence/news/leading-eu-lawmakers-propose-obligations-for-general-purpose-ai/

[129] “AI system that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of tasks

[130] Convington: Inside Privacy, “A Preview into the European Parliament’s Position on the EU’s AI Act Proposal”, on March 28, 2023
https://www.insideprivacy.com/artificial-intelligence/a-preview-into-the-european-parliaments-position-on-the-eus-ai-act-proposal/

システムの類型が追加される[131]方針であることが報じられている。なお、これらについては、それぞれ定義が与えられているものの、それらの定義に係る互いの包含関係は必ずしも明確に規定されているように見えず、「危険そうな」AI システムの類型を全て、ハイリスク AI システムの規制枠組みに押し込んだ感が否めない。

さらに、4 月 27 日付けのロイターの報道[132]によると、欧州議会の委員会は、AI 法案に係る予備的な合意に達し、その中で ChatGPT のような生成型 AI ツールを導入している企業は、システム開発に著作権のある資料が使用した場合、その開示が義務付けられることになったとしている。同報道によると、この著作権に関する条項は過去２週間のうちに追加されたものであり、当初、一部委員は生成系 AI モデルの学習における著作権のある資料の使用を全面的に禁止する提案を行っていたが、妥協案として、透明性確保義務とする方向になったとしている[133]。

いずれにせよ、今回の ChatGPT の発表に端を発する第四次 AI ブームの動きを受けて、欧州 AI 法案に対して、ChatGPT 型の AI システムや生成系 AI システムなどを新たに規制対象とすることが見込まれ、その際、この規制対象者としては、特に、OpenAI、Microsoft、Google などの域外企業を想定していると考えられる。

＜欧州議会委員会で採決された案＞

このような中、2023 年 5 月 11 日、欧州議会の域内市場・消費者保護委員会は、欧州 AI 法案の修正に係る交渉案について採決を行い、賛成多数を得た[134]。同修正交渉案では、もともとの EC 提案の欧州 AI 法案に対し、原型をとどめないほど非常に多くの改正事項が提示されている[135]が、そのポイントは、図２３の通り。

この中で、ChatGPT 型 AI システムについては、汎用 AI システムの一部として、リスク評価・低減、要件の認証などの義務を課すことに加えて、特に生成系の AI システムについては、学習された著作データの公開も含めて、透明性要件を課すこととしている。

なお、同修正交渉案では、これまでハイリスクとして位置づけられていた遠隔生体認証システム、生体分類システム、感情認識システムなどの一部の禁止 AI への格上げも含めて、

---

131 具体的には、Biometric AI、Insurance AI、AI used by Chldren、Generative AI、Deep Fake AI、Subliminal AI、General Purpose AI.

132 ロイター「欧州議会委員会、ＡＩ利用巡る規則案で合意　著作権の透明性確保」2023 年 4 月 28 日
https://jp.reuters.com/article/european-union-ai-idJPKBN2WO1YQ
Reuters, "EU proposes new copyright rules for generative AI", April 28, 2023
https://www.reuters.com/technology/eu-lawmakers-committee-reaches-deal-artificial-intelligence-act-2023-04-27/

133 Euractive, “MEPs seal the deal on Artificial Intelligence Act”, 2023 年 4 月 28 日
https://www.euractiv.com/section/artificial-intelligence/news/meps-seal-the-deal-on-artificial-intelligence-act/

134 European Parliament, "AI Act: a step closer to the first rules on Artificial Intelligence”, 11-05-2023
https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence

135 Committee on the Internal Market and Consumer Protection Committee on Civil Liberties, Justice and Home Affairs, “Version: 1.0 DRAFT Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 –C9 0146/2021 –2021/0106(COD))”, 9/5/2023
https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf

全般的に、もともとの欧州委員会（EC）提案の原案に比較して、規制のレベルを強化していることが印象的である。

【図２３】：欧州議会域内市場・消費者保護委員会での AI 法案修正交渉案（概要）[136]

項目	概要
AIの定義	"'Artificial intelligence system' (AI system) means a machine-based system designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments."
禁止AIの拡大	• 公的空間でのリアルタイムの遠隔生体認証システム、事後の遠隔生体認証システム（例外あり） • 機微な特徴（性別、人種等）に係る生体分類システム、感情認識システム（法執行、労働・教育分野等） • 予測的取り締まり（プロフィール、場所等に基づくもの） • ソーシャルメディア等からの生体データの無差別な取得
ハイリスクAIの対象拡大、義務の強化	• 人々の健康、安全、基本的権利を侵害するもの • 付属書Ⅲの抜本的改正：ソーシャルメディアでの政治的なキャンペーンでの推薦システムなど • ハイリスクAIの義務の強化（リスク管理、データガバナンス、技術文書と記録保持）
汎用AIシステム	• リスクの評価・低減、要件の認証、EUデータベースへの登録 • GPTなどに関しては、追加の透明性要件（AIによって生成されたこと、不法なコンテンツを生成しないこと、学習された著作データの要約の公開）
イノベーション、市民権、ガバナンス	• 研究活動・オープンソースの例外の拡大、サンドボックスの推進 • AIシステムに係る苦情受付等の拡大 • 欧州AIオフィスの機能の改革

なお、本修正交渉案について、今後、6 月 14 日に本会議での投票が想定されており、その後、本 AI 法案は、欧州議会、EU 理事会、欧州委員会の三者対話（トリローグ）に進み、法案の最終的な詳細が検討されることになる[137]。

＜欧州 AI 法案以外での動き＞

なお、欧州では、AI に対する不安感、不信感を背景に、上述の欧州議会を中心に議論されている欧州 AI 法案の修正以外にも、ChatGPT 型 AI システムに対する規制を強化しようとする動きがある。

例えば、EU のユーロポール（欧州刑事警察機構）は、2023 年 3 月 27 日に「大規模言語モデルが法の執行に与える影響」という報告書[138]を公表した。同報告書は全 13 頁と短いも

[136] 出典：以下より筆者作成
European Parliament, "AI Act: a step closer to the first rules on Artificial Intelligence", 11-05-2023
https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence
Euractive, "AI Act moves ahead in EU Parliament with key committee vote", 2023 年 5 月 11 日
https://www.euractiv.com/section/artificial-intelligence/news/ai-act-moves-ahead-in-eu-parliament-with-key-committee-vote/

[137] Euractive, "AI Act moves ahead in EU Parliament with key committee vote", 2023 年 5 月 11 日
https://www.euractiv.com/section/artificial-intelligence/news/ai-act-moves-ahead-in-eu-parliament-with-key-committee-vote/

[138] Europol, "The criminal use of ChatGPT – a cautionary tale about large language models", 27 MAR 2023
https://www.europol.europa.eu/media-press/newsroom/news/criminal-use-of-chatgpt-cautionary-tale-about-large-language-models
Gigazine「ChatGPT などの優れたチャットボット AI がいかに犯罪に使われやすいかをユーロポールがまとめて公開」2023 年 03 月 29 日

のであるが、具体的なユースケースとして、詐欺・なりすまし、サイバー犯罪の二つを取り上げ、偽情報の拡散に LLM が悪用される危険があると述べ、今後 AI 法案で規制されることを見込みつつも、法執行機関としての取組課題をリストアップしている。

また、個人情報保護の観点から既存規制の執行の対象としようとする動きが、イタリアから始まっている。具体的には、イタリアの個人データ保護当局は、2023 年 3 月 31 日、ChatGPT について直ちにブロックし、EU の「一般データ保護規則」（GDPR）を順守しているか調査すると明らかにしたことが報道されている[139]。特に、ChatGPT の訓練の目的で「個人情報を大量に収集し保管する」ことを正当化する法的根拠がないなどと指摘[140]し、OpenAI 社に、今後 20 日間で指摘に対応するよう指示したとされる[141]。その後、4 月 28 日には、イタリア政府は、個人データの取り扱いを改善したとして、使用禁止を解除したことが報道されている[142]。

なお、この根拠法は、EU 全体の規則である GDPR であることから、ChatGPT に対する個人情報保護の観点からの規制が欧州全域に拡がる可能性が指摘されている[143][144]。実際に、その後、4 月 4 日には、ドイツのデータ保護機関の当局者は、ChatGPT に関して、データ保護を巡る懸念を理由に国内での使用を一時的に差し止めることは原則可能との見解を示している。本件に関しては、その後、欧州連合のデータ保護委員会（EDPB）は、4 月 13 日、ChatGPT に関するデータ保護への懸念を検証するためのタスクフォースを立ち上げたと発表している[145]。

---

https://gigazine.net/news/20230329-chatgpt-llm-criminal-use/
YahooJAPAN（Forbes Japan）「ChatGPT でサイバー犯罪が急増の恐れ、欧州刑事警察機構が警告」3/30(木)
https://news.yahoo.co.jp/articles/3d695dd258ea114659bd4a9ac0c6cc26313ae3c1

[139] BBC News Japan「イタリア当局、人工知能チャットボット「ChatGPT」を一時的にブロック」2023 年 4 月 1 日
https://www.bbc.com/japanese/65135281

[140] YahooJAPAN ニュース（Lifehacker）「イタリアが ChatGPT を「即効で」禁止した理由。次に続く国は？2023 年」4 月 8 日
https://news.yahoo.co.jp/articles/6757727e1a293b03a6691418f6b1161fc220f88f

[141] なお、4 月 12 日には、月内には対応を行うよう指示を出したと報道されている。
日本経済新聞「イタリア、ChatGPT に個人情報保護対策　月内に要求」2023 年 4 月 13 日
https://www.nikkei.com/article/DGXZQOGN138EY0T10C23A4000000/

[142] 朝日新聞デジタル「ChatGPT、イタリアが使用禁止を解除　個人データの扱いで改善」2023 年 4 月 29 日
https://www.asahi.com/articles/ASR4Y226RR4YUHBI001.html

[143] 日本経済新聞「チャット GPT、専門家「欧州で規制拡大の可能性」イタリアで一時禁止、データ保護違反の疑い」2023 年 4 月 4 日
https://www.nikkei.com/article/DGKKZO69861400T00C23A4TB0000/
日本経済新聞「ChatGPT「欧州で規制広がる恐れ」　データ法制専門家」、2023 年 4 月 3 日
https://www.nikkei.com/article/DGXZQOUC020H20S3A400C2000000/

[144] YahooJAPAN ニュース（ロイター）「アングル：チャットＧＰＴ、欧州で規制強化検討へ　伊がきっかけ」、2023 年 4 月 4 日
https://news.yahoo.co.jp/articles/587ce39bfdb039962f000b5022188f5d26ea7ef0

[145] ITMedia NEWS「EU、ChatGPT 対策タスクフォース立ち上げ」2023 年 04 月 14 日
https://www.itmedia.co.jp/news/articles/2304/14/news114.html
EDPB, "EDPB resolves dispute on transfers by Meta and creates task force on Chat GPT", 13 April 2023
https://edpb.europa.eu/news/news/2023/edpb-resolves-dispute-transfers-meta-and-creates-task-force-chat-gpt_en

4-3. 米国、日本、中国等における規制・政策動向

（1）米国、英国、カナダにおける動向

＜米国の動き＞

一方、米国では、欧州同様に、全般的に AGI に対する不安感・不信感は有しつつも、自ら世界を主導する AI 企業を擁する立場等から、欧州とは異なり、まずは産業界自らにおける責任ある AI に向けた取組の推進を進め、そのリスクに係る公的評価を進めつつ、政府での AI 利用に係るガイドラインの策定から取り組むというスタンスを取りつつあるように見受けられる。

具体的には、バイデン大統領は、2023 年 4 月 4 日に開催された科学技術に関する大統領諮問委員会（PCAST）の会合において、前年の 10 月に発表した AI 権利の章典に係る取組を説明した上で、「責任あるイノベーションと米国の権利と安全の保護を確保すること」「プライバシーの保護とバイアス・誤情報への対応」の必要性を指摘し、また、ChatGPT を念頭に、「ハイテク企業は、製品を公開する前に、安全性を確認する責任がある」と発言した[146]。この発言からは、少なくとも現時点では、AI 権利の章典を踏まえつつも、まずは事業者自らの自主的な取組が重要と指摘しているようにも見える。

その後、バイデン大統領立ち合いの下、ハリス副大統領は、2023 年 5 月 4 日、Alphabet（Google）、Anthropic、Microsoft、OpenAI の CEO をホワイトハウスに招き、責任ある AI に係るイノベーションを強調する会議を開催している[147]。同会議では、①責任ある AI への研究開発に対する新たな投資、②既存の生成系 AI システムに対する公的評価の実施、③OMB による政府での AI 利用に係る主導的な政策案の発表とパブコメの開始、について発表している。

---

[146] The Whitehouse, “Remarks by President Biden in Meeting with the President’s Council of Advisors on Science and Technology”, APRIL 04, 2023
https://www.whitehouse.gov/briefing-room/speeches-remarks/2023/04/04/remarks-by-president-biden-in-meeting-with-the-presidents-council-of-advisors-on-science-and-technology/
読売新聞オンライン「バイデン氏、チャットＧＰＴ念頭に「国家安全保障への潜在的リスク」言及…法整備へ」2023/04/05
https://www.yomiuri.co.jp/economy/20230405-OYT1T50080/
日本経済新聞「ChatGPT などの安全性確認、AI 企業に責任　米大統領」2023 年 4 月 5 日
https://www.nikkei.com/article/DGXZQOGM050TG0V00C23A4000000/

[147] ITMedia NEWS「米バイデン政権、OpenAI など AI 関連 4 社の CEO を招き「責任あるイノベーション」促す」2023 年 05 月 05 日
https://www.itmedia.co.jp/news/articles/2305/05/news041.html
The Whitehouse, “Readout of White House Meeting with CEOs on Advancing Responsible Artificial Intelligence Innovation”, MAY 04, 2023
https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/04/readout-of-white-house-meeting-with-ceos-on-advancing-responsible-artificial-intelligence-innovation/
The Whitehouse, ”Statement from Vice President Harris After Meeting with CEOs on Advancing Responsible Artificial Intelligence Innovation”, MAY 04, 2023
https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/04/statement-from-vice-president-harris-after-meeting-with-ceos-on-advancing-responsible-artificial-intelligence-innovation/
The Whitehouse, “FACT SHEET: Biden-Harris Administration Announces New Actions to Promote Responsible AI Innovation that Protects Americans’ Rights and Safety”, MAY 04, 2023
https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/04/fact-sheet-biden-harris-administration-announces-new-actions-to-promote-responsible-ai-innovation-that-protects-americans-rights-and-safety/

なお、必ずしも、生成系 AI に限ったものではないが、商務省 NTIA（電気通信情報局）は、2023 年 4 月 11 日、AI の説明責任に係る会合を開催し、その中で、AI の説明責任政策に関するパブリックコメントを募集すると発表した[148]。NTIA は、AI の監査、評価、認証等の開発支援に係る政策に関する意見募集を行うとしており、具体的には、例えば「監査、評価を必要とするデータアクセスの種類」「信用できる AI システム構築のインセンティブ・支援に果す規制当局等の役割」「産業分野ごとにどのような異なったアプローチが求められるか」などの AI 説明責任エコシステムに係る幅広い政策を対象に意見募集を行っている（締め切りは 6 月 12 日）。

一方、FLI による開発中断提言に係る公開書簡の発表の直後、同様に AI の倫理問題を扱う非営利調査団体である AI デジタル政策センター（CAIDP）は、3 月 30 日、米国連邦取引委員会（FTC）に対して、GPT-4 の商業目的でのリリースに対して中止命令を出すよう書簡を通じ申し立てている[149][150]。しかしながら、これに対して、少なくとも現時点では FTC における何らかの動きは見られない。

＜英国の動き＞

英国では、2020 年 1 月の欧州連合（EU）離脱以降、EU の推進する AI 法案とは差別化した独自の AI 政策方針を打ち出しているが、現在、ChatGPT 型 AI システムに関しても、その延長線で検討を進めている模様である。

具体的には、（新設の）科学イノベーション技術省は、2023 年 3 月 29 日、「AI 規制：プロイノベーションアプローチ」の報告書（白書）を発表した[151]。この報告書は、前年 2022 年 7 月に発表した「AI 規制に係るプロイノベーションアプローチの確立」を踏まえ、その詳細化を図るために作成されたものである。そのため、本白書においては、原則として、従前どおり、イノベーションの推進を強調する一方で、「文脈に応じた規制の在り方」という趣旨の下、より柔軟な規制の方向について議論を行っている。

そのような趣旨のもと、本白書では、ChatGPT のような大規模言語モデル（LLM）に関しては、今後英国での競争力強化のための専門家による新タスクフォースを設立する[152]とし

---

[148] NTIA, NTIA Seeks Public Input to Boost AI Accountability, April 11, 2023
https://www.ntia.doc.gov/issues/artificial-intelligence
NTIA, AI Accountability Policy Request for CommentApril 11, 2023
https://www.ntia.doc.gov/issues/artificial-intelligence/request-for-comments
https://www.federalregister.gov/documents/2023/04/13/2023-07776/ai-accountability-policy-request-for-comment
YahooJAPAN ニュース（CNET Japan）「米政府、「AI の説明責任」に関する政策について一般意見を募集へ」4/12(水)
https://news.yahoo.co.jp/articles/11c76d9c2383cbad88990a67e10158ecc0343090

[149] ロイター「最新版「ＧＰＴ－４」、調査団体が商業リリース中止をＦＴＣに要請」2023 年 3 月 31 日
https://jp.reuters.com/article/ai-ftc-idJPKBN2VX00U

[150] Fortune, “OpenAI’s ChatGPT faces U.S. FTC complaint, call for European regulators to step in”, March 31, 2023
https://fortune.com/2023/03/30/openai-chatgpt-gpt-4-ftc-complaint-caidp-beuc-europe/

[151] Department for Science, Innovation and Technology and Office for Artificial Intelligence, “Policy paper: AI regulation: a pro-innovation approach”, Published 29 March 2023
https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach

[152] Department for Science, Innovation and Technology, “Press release UK unveils world leading approach to innovation in first artificial intelligence white paper to turbocharge growth”, 29 March 2023

た一方、その規制の在り方に関しては、参考事例や今後の検討の論点としてはいくつか提示をしているものの、まだ今後の明確な方向は示していない段階にある[153]。なお、同白書は、現在、今後の LLM の規制の在り方も含めて、パブリックコメントに付しているところである（6 月 21 日締め切り）[154]。

＜カナダの動き＞

カナダでは、欧州同様、2023 年 4 月 4 日、同国のプライバシーコミッショナー事務所（OPC）が、個人情報保護の観点から、OpenAI の調査を開始したことが報じられている[155]。

（３）日本における動向

今回の第四次 AI ブームに係る世界的な流れの中で、日本は、ChatGPT 型の AI システムに関し、これまで同様、これらに係るリスクを理解しつつも、むしろ原則として、そのイノベーションと利用を促進しようとしている点で、他国とは一線を画すと言える。

＜自民党の動き＞

ChatGPT に対する世界的な関心の高まりで、まず検討を開始したのは、行政府というよりは政治側である。具体的には、自民党は、2023 年になって、同党のデジタル社会推進本部の中に「AI の進化と実装に関する PT」を設置し[156]、2 月 3 日に初会合を開催し、その後、国内の AI 専門家等に対して積極的なヒアリングを実施している[157]。その上で、同 PT は、2023 年 3 月 30 日、報告書案「AI ホワイトペーパー～AI 新時代における日本の国家戦略～」を提示し、その後、4 月 13 日に同党の政審調査会で了承を得る（図２４参照）とともに、5 月 9 日には、Web3 に係るホワイトペーパーと併せて、岸田総理に提言している[158]。

---

https://www.gov.uk/government/news/uk-unveils-world-leading-approach-to-innovation-in-first-artificial-intelligence-white-paper-to-turbocharge-growth

[153] CNBC.com, "With ChatGPT hype swirling, UK government urges regulators to come up with rules for A.I.",PUBLISHED WED, MAR 29 2023
https://www.cnbc.com/2023/03/29/with-chatgpt-hype-swirling-uk-government-urges-regulators-to-come-up-with-rules-for-ai.html
Wiggin, “Department for Science, Innovation and Technology (DSIT) publishes White Paper on Artificial Intelligence — A pro-innovation approach to AI regulation”, April 3, 2023
https://www.wiggin.co.uk/insight/department-for-science-innovation-and-technology-dsit-publishes-white-paper-on-artificial-intelligence-a-pro-innovation-approach-to-ai-regulation/

[154] Department for Science, Innovation and Technology and Office for Artificial Intelligence, “Open consultation AI regulation: a pro-innovation approach – policy proposals”, 29 March 2023
https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals

[155] YahooJAPAN ニュース（ZDNET）「カナダ当局が「ChatGPT」開発元の調査を開始--個人情報の扱いを懸念」4/5(水)
https://news.yahoo.co.jp/articles/738b05d850079287b12be623a74912f1d95c64c9

[156] 自民党 AI プロジェクトチーム
https://note.com/akihisa_shiozaki/n/n4c126c27fd3d

[157] 例えば、2 月 17 日の第二回会合では、東大松尾教授がプレゼンしている。
松尾豊「AI の進化と日本の戦略」23/2/17
https://note.com/api/v2/attachments/download/a29a2e6b5b35b75baf42a8025d68c175

[158] 自民党「新時代における国家戦略策定を AI・web3PT が岸田総理に提言」2023 年 5 月 11 日
https://www.jimin.jp/news/policy/205802.html

【図２４】自民党「AI ホワイトペーパー～AI 新時代における日本の国家戦略～」[159]

項目	提言内容
1．新たなAI国家戦略の策定の必要性	• AI新時代にふさわしい新たな国の基本戦略の策定、新たな政策の立案とこれまでの取り組みの早急な見直し • AI政策に関する司令塔の確定、体制の拡充、幅広い観点から早急かつ総合的な施策の検討
2．国内におけるＡＩ開発基盤の育成・強化	• 海外プラットフォームの積極的な利活用、基盤モデルに関する国内の知見の蓄積、応用研究・開発の加速、国内における基盤モデル等の基礎的な技術開発能力の構築・強化 • 人材関連施策の更なる強化、「AIハブ」の創設、コミュニティ形成の支援 • AIによる官民データの利活用を推進する環境づくり、公共データのアーカイブ化・第三者提供ルール・形式等の整理、日本に関連する学習データの比率の向上、政府主導での日本語コーパスの作成・利活用の推進、学習データの充実・利活用 • 基盤モデルの構築・利活用に要する膨大な計算資源についての国内基盤整備と拡充、新たな枠組の整備、半導体産業の育成強化
3．行政における徹底したAI利活用の推進	• 諸外国の政府機関におけるAIに関する先進的な活用事例及びそのためのガイドライン等の調査、複数のパイロットプロジェクトへの直ちの着手、ハッカソン・ビジネスコンテストの開催、各種AIの徹底した利活用をさらに加速させるための指針の策定、関係機関でのAI導入等を支援する専門チームの政府内設置 • 地方自治体によるAIを活用したスマートシティの取り組みのを、強力な支援
4．民間におけるAI利活用を奨励・支援する政策	• 基盤モデルのAIが様々な国内産業に与える影響の早急な調査、スタートアップや新規事業の創出の奨励、チーフデジタルオフィサー（CDO）の設置推奨すること、AIガバナンスのあり方について必要があればガイドラインなどの策定、企業のAI人材の活用・処遇に関する取り組みの支援
5．AI規制に関する新たなアプローチ	• EU、米国、中国など諸外国のAI規制の検討状況を分析し、①重大な人権侵害、②安全保障、③民主主義プロセスへの介入など、AI新時代において法規制を含む対策が必要と考えられる分野での具体的な検討、AI利用を巡る国際的なルール作りの議論への積極的かつ戦略的な参画 • デジタル原則に基づくアナログ規制の見直しがさらに促進される仕組みの確立、事業者が新規事業にチャレンジできる環境の整備・発展、 生成系AIに関する知的財産法の解釈に係るガイドライン等の策定 • 公教育のカリキュラムの中でAIリテラシーの向上を具体的に位置付け、公教育の現場における生徒による大規模言語モデルの利用の可否などAIの取り扱いに関する指針の早急な策定

同報告書では、「大規模言語モデル（LLM）に代表される AI の進化と社会実装は、新たな経済成長の起爆剤となりうる」とした上で、

- AI 新時代にふさわしい新たな国の基本戦略[160]を策定。新たな政策の立案とこれまでの取り組みの見直し。
- AI 政策に関する司令塔の確定、体制拡充。研究開発、経済構造、社会基盤、人材育成、安全保障など幅広い観点から早急かつ総合的に施策の検討。

を提言している。また、このうち、規制に関しては、「欧米諸国では、AI の社会受容に向けた規制論議が加速している中、彼我の距離感を問い直すべき時期に来ている」とした上で、

- EU、米国、中国など諸外国の AI 規制の検討状況の分析。①重大な人権侵害、②安全保障、③民主主義プロセスへの介入など、AI 新時代において法規制を含む対策が必要と考えられる分野での具体的な検討。
- 日本が議長国を務める本年の G7 サミットを含め、AI 利用を巡る国際的なルール作りの議論に積極的かつ戦略的参画。

などを提言している。これらを見る限り、自ら積極的に規制に取り組もうというスタンスではなく、海外の規制動向に対応しようとする受身的なスタンスになっていることが特徴的であると言える。

---

[159] 出典：自民党デジタル社会推進本部 AI の進化と実装に関するプロジェクトチーム「AI ホワイトペーパー～AI 新時代における日本の国家戦略～」2023 年 4 月
https://note.com/akihisa_shiozaki/n/n4c126c27fd3d

[160] なお、日本政府は、これまで 2022 年までは、毎年 AI 戦略のアップデートと行ってきたが、2023 年はアップデート行う予定がなかった模様であり、同報告書には、「現状では新たな「AI 戦略 2023」の策定は予定されておらず、本年度は、内閣府がとりまとめる「統合イノベーション戦略」の一つの章において政府の AI に関する政策的取り組みの進捗を紹介する予定とのことである。」と記載されている。

なお、同報告では、日本が独自の LLM のようなモデルを構築するべきかについては、原則として、「海外プラットフォームの積極的な利活用を通じて、基盤モデルに関する国内の知見を蓄積し、応用研究・開発を加速」し、並行して、「国内における基盤モデル等の基礎的な技術開発能力の構築・強化に向け投資と支援を継続する」としている[161]。

＜米国ハイテク企業幹部の来訪＞

このように、自民党が主導して新たな AI 政策の方針の検討を進める中、そもそも日本は、AI 政策に関して、G7 の中でも規制よりもイノベーションを重視する立場であり、かつ、同月（4 月）末には日本主催による G7 のデジタル・技術大臣会合が開催されることから、OpenAI をはじめ米国の大手 IT 企業の多くの幹部が来日し、自民党関係者等と面会している。

具体的には、自民党は、4 月 10 日に開催された同党の PT 会合に、OpenAI の CEO であるサム・アルトマン氏を招聘している[162]。アルトマン CEO は、同会合において「ChatGPT などの利活用と日本への提案」とのタイトルで、７つの提案を含むプレゼンを行うとともに、日本に事業あるいは研究開発拠点を新たに設ける意向を明らかにしたことが報道されている[163]。なお、同 CEO は、同日、岸田総理も表敬訪問しており、岸田総理は、「新しい技術が登場し、利用されている一方、プライバシーや著作権といった指摘されるリスクの状況や、国際的なルールづくりについての考え方など、意見交換した」と報じられている[164]。

また、その後、4 月 21 日にはマイクロソフト社の副社長が[165]、4 月 27 日には Google 副社長が、自民議員と会談し[166]、さらには、アマゾン・ウェブ・サービスの副社長も自民党の会合等に出席するなど、IT 大手の「日本詣で」が続いたことが報道されている。

---

[161] なお、東大をはじめ国内 AI 研究者の中には、独自の基盤を構築すべきとの意見も少なくない。Yahoo JAPAN ニュース（Abema Times）「日本独自の生成系 AI を持つべき」東大副学長の見解が国内外で話題」4/13(木)
https://news.yahoo.co.jp/articles/aae35972a290aab6681a807a2641b90733c313cc

[162] サム・アルトマン CEO は「OpenAI Tour 2023」をうたい、東京を含む世界の 17 都市を巡り、講演および政策立案者との面談を行う予定だ。
https://www.itmedia.co.jp/news/articles/2304/10/news169.html
実際に、もともと OpenAI は、日本に関心があったが、政府との調整役は、木原副長官、塩崎議員が担ったと報道されている。また、Altmen CEO は、海外の要人に会うのは初めてであり、日本の事業所・研究開発拠点が初めてであるとのこと。なお、日本市場は、著作権法の問題、AI に対する認識、G7 の議長国との観点で重要とみなされると分析している。
テレ東 Biz（Youtube）「“ＣｈａｔＧＰＴ”トップ　サム・アルトマンＣＥＯ電撃来日の舞台裏【WBS 未公開】」（2023 年 4 月 14 日）
https://www.youtube.com/watch?v=KhUPLzzU_TE

[163] ITMedia News「OpenAI のアルトマン CEO、「日本の ChatGPT ユーザーは 100 万人超」」2023 年 04 月 10 日 19 時 15 分 公開
https://www.itmedia.co.jp/news/articles/2304/10/news169.html
YahooJAPAN ニュース（ITMedia News）「OpenAI・アルトマン CEO のプレゼン資料が公開　自民党に何を語ったか」4/11(火)
https://news.yahoo.co.jp/articles/9c9bf047b89f69d658ba494626a7ef529f5eaba8
朝日新聞デジタル「ChatGPT、日本に研究開発拠点の設置検討　CEO が首相と面会」2023 年 4 月 10 日
https://www.asahi.com/articles/ASR4B5VPLR4BULFA01J.html

[164] 「ChatGPT」開発企業のアルトマン CEO 岸田首相と面会 2023 年 4 月 10 日 19 時 10 分
https://www3.nhk.or.jp/news/html/20230410/k10014034131000.html

[165] FNN プライムオンライン「マイクロソフト副社長　自民党と“GPT”議論」2023 年 4 月 21 日
https://www.fnn.jp/articles/-/517700

＜行政府の動き＞

このような自民党における動き等を踏まえ、行政府側においても、第四次 AI ブームに対応した新たな AI 政策に係る検討に向けた体制の整備が進められることになる。

まず 4 月前半において、官房長官による「政府の AI 政策は内閣府がとりまとめる（4 月 4 日）」[167]、「現時点で規制の導入は検討していない（4 月 14 日）」[168]などの発言に加え、西村経済産業大臣の「懸念解消なら国会答弁など活用検討（4 月 11 日）[169]」、河野デジタル担当大臣からは「省庁の働き方改革に向け ChatGPT などのハッカソンを開催（4 月 14 日）」[170]など、むしろ特に政府面での利用に係る発言が相次いだことが特徴である。

その上で、首相補佐官をヘッドとし、内閣府を中心とし、経済産業、総務、文部科学各省やデジタル庁、個人情報保護委員会の関係省庁からなる「AI 戦略チーム」が設置され、4 月 24 日に、初会合を開催している[171][172][173]。その後、5 月 11 日には、内閣府の会議として「AI 戦略会議」の第 1 回会合が開催されており[174]、その際、総理は、今後、統合イノベーション戦略、骨太方針等の政府方針や、国際ルール作りに反映させたいと発言している[175]。なお、会議では、セキュリティ、プライバシー、著作権は重要な論点とする一方、開発強化策を求める声があったことが報じられている[176]。

---

[166] YahooJAPAN ニュース（TBS NEWS DIG）「【独自】Google 副社長が自民議員と会談、生成 AI の説明も「日本で信頼されるパートナーであり続けられるか」　IT 大手の“日本詣で”続く」4/27(木)
https://news.yahoo.co.jp/articles/06e49c064871f60fecedf75593fe1ad33a9b79de

[167] Yahoo JAPAN ニュース「チャットＧＰＴ、必要な場合は内閣府中心に対応検討＝官房長官」4/4(火)
https://news.yahoo.co.jp/articles/c1fa6002267d85310b0bba2161b7c4d01283f94f

[168] 「ChatGPT、松野官房長官「現状規制する考えない」」
https://newsdig.tbs.co.jp/articles/-/433716

[169] 「「ChatGPT」“懸念解消なら国会答弁など活用検討” 西村経産相」2023 年 4 月 11 日
https://www3.nhk.or.jp/news/html/20230411/k10014035011000.html

[170] 「「ChatGPT などのハッカソンを開催したい」--河野大臣、省庁の働き方改革で表明」4/14(金)
https://news.yahoo.co.jp/articles/df8e4f689ef6b4ab8fc1504aa3430b16dac9a4c9

[171] NHK NewsWeb「政府 ChatGPT など有効活用に向け新たに検討チーム設置へ」2023 年 4 月 14 日
https://www3.nhk.or.jp/news/html/20230414/k10014039071000.html
日本経済新聞「政府、ChatGPT など活用　検討チーム設置へ」2023 年 4 月 14 日
https://www.nikkei.com/article/DGXZQOUA14CI90U3A410C2000000/

[172] ITMedia News（産経新聞）「政府の AI チーム初会合　課題や利活用の方向性を共有」2023 年 04 月 25 日
https://www.itmedia.co.jp/news/articles/2304/25/news084.html

[173] 内閣府　AI 戦略チーム（関係省庁連携）
https://www8.cao.go.jp/cstp/ai/ai_team/ai_team.html

[174] 読売新聞オンライン「チャットＧＰＴなどの活用と規制議論、「ＡＩ戦略会議」設置へ…政策の司令塔」4/26(水)
https://www.yomiuri.co.jp/politics/20230425-OYT1T50283/
内閣府、第一回 AI 戦略会議令和 5 年 5 月 11 日
https://www8.cao.go.jp/cstp/ai/ai_senryaku/1kai/1kai.html

[175] 総理官邸、ＡＩ戦略会議、令和 5 年 5 月 11 日
https://www.kantei.go.jp/jp/101_kishida/actions/202305/11ai.html

[176] 毎日新聞「チャット GPT 念頭に議論　開発強化策求める声も　AI 戦略会議初会合」2023/5/11
https://mainichi.jp/articles/20230511/k00/00m/010/135000c
https://news.yahoo.co.jp/articles/fd0bcca0d299c4f6f57515c88a8329976f1cc381
読売新聞「ＡＩ活用へルール策定、政府の戦略会議初会合…チャットＧＰＴ念頭に「著作権は重要な論点」」2023/05/11
https://www.yomiuri.co.jp/politics/20230511-OYT1T50123/

一方、各省庁でも検討の動きが開始されつつある。例えば、総務省は、2023 年 4 月 27 日、情報通信審議会情報通信政策部会を開催[177]し、「2030 年頃を見据えた情報通信政策の在り方」について議論を行っているが、同報告書原案では、生成系 AI に関し、中長期的な観点からの日本語による AI 基盤モデルの構築の必要性、全ての国民が AI 等デジタルツールを巧みに活用する能力の習得の必要性について指摘している。

（3）中国における動向

このように先進民主主義諸国での動きに対して、AI 大国の一つではあるものの、インターネットでの言論統制を行う中国では、独自の規制路線を進む方向にある。

＜DeepFake 規制＞

まず、そもそも、中国では、生成系 AI の一種である DeepFake 技術に対しては、早い段階から規制を行っている。

具体的には、国家インターネット情報弁公室（CAC）は、2019 年 11 月に、ディープラーニング技術を利用した動画内容についての規則を制定している（2020 年施行）[178]。本規則では、真実ではない AI 音声や動画を発信する場合、AI 技術により制作したということを明確に記載することを義務付けるとともに、虚偽のニュースや情報の制作、公表、発信を禁止している。

さらに、CAC は、2022 年 1 月に DeepFake 技術に係る更なる規制案を発表し[179]、同年 12 月に同規則を交付し、2023 年 1 月 10 日から施行している[180]。同規則では、ディープラーニングや VR を使用してコンテンツを変換するプラットフォームや企業（「ディープ統合サービスプロバイダ」）に対し、違法な情報を作成、公開、発信を規制するだけではなく、データの元となっている音声や画像に係る本人の許可を義務付けている[181]。

＜ChatGPT 型 AI システムに対する規制＞

毎日新聞「官邸主導の AI 戦略会議、岸田首相の思惑は　省庁「押し付け合い」も」2023/5/11
https://mainichi.jp/articles/20230511/k00/00m/010/135000c

[177] 総務省情報通信審議会情報通信政策部会（第 61 回）配付資料・議事概要・議事録、令和 5 年 4 月 27 日（木）
https://www.soumu.go.jp/main_sosiki/joho_tsusin/policyreports/joho_tsusin/joho_bukai/02tsushin10_04000560.html

[178] 36KrJapan「AI によるディープフェイク技術に規制　中国の新規定が 2020 年施行」2020 年 1 月 8 日
https://36kr.jp/44999/

[179] VOI「中国政府はディープフェイク技術とその派生物を規制する法律草案」31 Jan 2022
https://voi.id/ja/teknologi/129624

[180] ロイター「中国、新しい「ディープフェイク」規制を来月 10 日適用」2022 年 12 月 12 日
https://jp.reuters.com/article/china-regulation-tech-idJPKBN2SW022

[181] VOI「中国政府はディープフェイクを規制する新しいコンテンツルールを作成します」13 Des 2022
https://voi.id/ja/teknologi/235366
GIZMODE「「AI でも人の写真や音声を加工するなら許可とってね」中国の新ルールがかなり厳格」2022.12.15
https://www.gizmodo.jp/2022/12/china-deepfake-ai.html

一方、今回の第四次 AI ブームを受けて、第２章（３）で記載した通り、そもそも、中国からは（VPN などを通じなければ）ChatGPT に対してアクセスできないようになっており、また、既に、中国政府は、国内企業に対して、（VPN を通じて）ChatGPT を利用するサービスを提供しないよう求めている。

そのような中、国家インターネット情報弁公室（CAC）は、2023 年 4 月 11 日、いわゆる生成系 AI に対する新たな規制に係る案を発表したことが報じられている[182]。具体的には、生成コンテンツの内容が、社会主義の中核的な価値を反映し、国家の統一を損なわないようにすることはもちろん、コンテンツが正確で、知的財産を尊重し、差別的でなく、セキュリティを確保することを求められるとともに、AI が生成したコンテンツであることの明記することも義務付けとされている。また、手続き的には、当局による事前審査（セキュリティ評価）が義務づけられるとともに、当局の要求に応じて AI の基礎となるアルゴリズムを提出することも求めるとされている。政府は、対策案について、5 月 10 日まで一般の意見を募集し、年内の発効を予定しているとのことである。

このように、この中国の生成 AI の規制案は、全体として、アルゴリズムの提出義務も含めてかなり厳しい規制法案となっている。もちろん、その背景には、中国政府にとっては、これまでのグレートファイヤーウォールを始めとするインターネット規制を進める中で、欧米製の ChatGPT 型の AI システムが中国国内に流入することは、体制維持のために避けなければないことに加え[183]、軍事面でも今後重要となる認知戦で劣位になるのではないかとの危機感があるとされる[184]。

なお、2023 年 5 月 8 日には、中国人が、ChatGPT を利用して、虚偽の情報を生成し、インターネット上で拡散させたとして警察当局により拘束されたことが報じられている[185]。本件は、上記規制に基づくものではないが、VPN を通じ ChatGPT を利用して得られた中国政府にとっては望ましくない情報を、中国国内で拡散したことに対する「公共秩序騒乱」罪に係るものであり、このような報道がなされること自体が、中国政府の危機感の表れではないかと考えられる。

---

[182] YahooJAPAN ニュース（朝日新聞デジタル）「政権に不都合な表示防ぎたい？中国、対話型 AI の事前審査義務づけへ」4/11(火) 18:37 配
https://news.yahoo.co.jp/articles/8e7d474b9ee87db58c773994e863a4c428c1cf86
YahooJAPAN ニュース（Bloomberg）「中国、ChatGPT に似た生成ＡＩサービスにセキュリティ審査義務化へ」4/11(火)
https://news.yahoo.co.jp/articles/364a017df80cfae0b7c6b489fcbadc32e5a07b93
ロイター「中国、生成型ＡＩサービス管理へ対策案公表　年内発効へ」2023 年 4 月 11 日
https://jp.reuters.com/article/china-ai-idJPKBN2W809G

[183] YahooJAPAN ニュース（デイリー新潮、藤和彦）「チャット GPT で注目　中国政府が生成 AI を極度に恐れる理由」4/18(火)
https://news.yahoo.co.jp/articles/7ccd3dae0a8a71a0a94302f15a2520b0cd1b947c

[184] 時事通信ドットコム「対話型 AI、軍が警戒　「認知戦」利用に意欲も　中国」4/16(日)
https://www.jiji.com/jc/article?k=2023041500355

[185] 日本経済新聞「ChatGPT 悪用で中国が男拘束　虚偽情報拡散の疑い」2023 年 5 月 9 日
https://www.nikkei.com/article/DGXZQOCB08BZS0Y3A500C2000000/

５．まとめ：G7 閣僚宣言と今後の方向

（１）ChatGPT 型 AI システムに係る世界主要国の規制動向の特徴（まとめ）

第四章の議論を踏まえ、主要各国・地域の ChatGPT 型 AI システムに係る最近の規制・ガバナンス動向を、ChatGPT 登場前と登場後に分けて簡潔に整理すると、図２５の通りである。ただし、これは現時点での動向であり、今後変化する可能性があることに留意することが必要である。

【図２５】主要国における ChatGPT 型 AI システム等に係る規制動向[186]

	第三次AIブーム（～2022年）	第四次AIブーム後（2023年～）
欧州	• 2021年4月、欧州委員会（EC）は、欧州AI法案発表 • 2022年12月、欧州理事会に、欧州AI法案に係る一般アプローチを発表（「汎用AIシステム」の規制等）	• 欧州議会において、汎用AIシステムの一部として、ChatGPTを規制対象の方向で検討。特に利用した著作物の概要開示義務。2023年6月頃投票見込み。 • 2023年3月末、イタリアデータ保護当局は、使用禁止命令（4月末解除）、4月欧州委はTF設置。
カナダ	・2022年6月、AI・データ法案発表。	• 2023年4月、プライバシー保護当局が調査開始。
米国	• 2022年10月、OSTPは、AI権利の章典を発表。 • 2023年1月、NISTは、AIリスクマネジメント枠組みを発表。	• 2023年4月、バイデン大統領は、AI権利の章典を踏まえ、企業の安全性確認責任を発言。5月、関係企業との会議開催 • 2023年4月、NTIAは、AIの説明責任政策に係るパブコメ募集開始。ただし、生成AIに限ったものではない。
英国	• 2022年7月、「AI規制に係るプロイノベーション・アプローチの確立」を発表。	• 2023年3月末、「AI規制：プロイノベーションアプローチ」（白書）を発表。LLMに係る専門家TFの設置、、リスクについては論点提示、パブコメ。
日本	• 2022年1月、経産省は、AIガバナンスガイドラインver1.1を発表。 • 2022年4月、AI戦略2022発表。	• 2023年3月末、自民党は、AIホワイトペーパーを発表、LLMに係る新たな国家戦略策定、司令塔の設置等を提言。 • 2023年4月は、政府は、AI戦略チーム初会合開催。5月、AI戦略会議発足。
中国	• 2022年12月、DeepFake規制を発表。2023年1月から施行。	• 2023年4月、生成型AIに係る規制案を発表。年内発効を予定。

この図２５を踏まえつつ、今回の ChatGPT 型 AI システムへの関心の高まりに対する今後のガバナンス政策方向・影響（見込み）として、以下の３点をあげることができる。

＜(1) 先進民主主義国・地域における AI ガバナンスの多様性＞

まず一つは、これまでの AI ガバナンスの動向と同様であるが、先進民主主義国・地域間における政策の方向の多様性である。すなわち、欧州では、引き続き、予防原則的な発想に基づき、ChatGPT 型 AI システムについても、AI 法案の規制体系の枠組みの中に新たに組み込むこととし、事前規制の対象にする方向で検討が進んでいる（アプローチ１）。一方、これに対し、日本では、原則としては、法規制の対象とはしない方向で、イノベーションを促進する方向を向いている（アプローチ２）。なお、米国、英国などは、中間的な立場に位置づけられるが、少なくとも、現時点ではこれまでの AI 政策の流れの一環として捉えており、具体的な検討はこれからとなるものと考えられる。

[186] 出典：筆者作成（2023 年 5 月半ば現在）

これらの政策方針には、これまでの AI ガバナンスと同様、各国・地域の国際的な産業戦略的な思考が含まれているものと考えられる。すなわち、欧州では、米国優位のハイテク企業を欧州域内の規制枠組みに巻き込むことを考えているのに対し、英国、日本はむしろイノベーション促進の観点から規制を捉えているという傾向が伺える。

＜(2) 個人情報保護・公平性問題から著作権・民主主義問題へ＞

二つ目は、ただし、これまでの AI 規制・ガバナンス政策では、主に個人情報に基づく AI による意思決定システムの公平性の問題が対象であったため、プライバシーや人権問題など欧州が強みを有する法的枠組みの下で、欧州が世界の AI 規制・ガバナンスをリードしてきたという側面がある。これに対し、今回の ChatGPT 型の AI システムに関しては、むしろ著作物に係る対応や、情報の正確性／偽情報（悪用）への対応などが課題になるため、これまでとは異なった規制・ガバナンス体制が求められる可能性がある。

特に、著作物に関しては、特に画像分野においては、クリエイターの職種の維持確保の観点も含めて利害関係者との調整が、大きな論点となる可能性があり、場合によってはオプトアウトの技術的可能性も含め、検討がなされる可能性がある。

一方、偽情報、悪用に関しては、民主主義という文脈で規制を追求する可能性があるものの、規制強化をすると（中国のように）表現の自由にも抵触しうることから、どこまで政府による規制が必要とされるかが課題になるものと考えられる。

＜(3) 世界からみた中国市場の分断、孤立化の可能性＞

三つ目は、中国との関係である。これまでの AI 規制・ガバナンスでは、公平性や人権に関わるものが主流であり、そのような中、中国では、国家自らの AI 利用にも影響を与える可能性もあるためか、その政策的な動きはほとんど見られなかった。これに対し、今回の ChatGPT 型の AI システムに関しては、中国が、その体制秩序維持のため規制の導入をいち早く明確にしたことが特徴である。

そもそも、世界全体で見ると、LLM 構築には非常に多くの資金が必要となることから、少なくとも当面の間、AI 大国である米国系及び中国系の大手 IT・ネット系企業が担う可能性が高い。その際、中国市場参入にあたって、生成コンテンツの内容に関し、中国国家／社会主義の価値の反映が求められるとした場合、場合によっては多額のコストを要する LLM 構築のための学習コンテンツの見直しが求められることから、欧米大手企業は参入に躊躇するものと考えられる。さらに、アルゴリズムの提出義務を課されるとした場合、欧米企業はまず参入しないであろう。このため、中国市場に参入するのは、中国国内企業に限定されるものと考えられる。一方、中国国内の価値観のみによって訓練された LLM が、中国国外で人気を得ることも想定しがたいことを踏まえると、世界の市場は、中国とそれ以外の市場に分断され、中国市場は孤立化する可能性も想定される[187]。

ただし、中国側においては、認知戦の観点からも含めて、中国国内向けだけでなく、中国側の意図も内々含めた形での世界向けの LLM を戦略的に開発・普及させる可能性も否定で

[187] Daimond Online（莫 邦富）「百度が発表「中国版 ChatGPT」で中国人に即バレた不都合な真実」2023.4.4
https://diamond.jp/articles/-/320506
Money Voice（牧野武文）「AI で世界制覇を狙う中国が弱点をさらした“中国語”という盲点。ChatGPT の爆発で中国は AI 戦争に敗北するのか？」2023 年 4 月 10 日
https://www.mag2.com/p/money/1301382

きない。その際、先進民主主義国の市場においては、上述のような動きが生じた場合の対応に加え、一国の事情のみに応じたコンテンツ規制によって市場が分割されないような体制を検討することが必要になる可能性がある。

（２）G7 閣僚宣言と今後の方向

＜G7 閣僚宣言＞

このような中、2023 年 4 月 29 日、30 日に、G7 群馬高崎デジタル・技術大臣会合が開催され[188][189]、30 日には閣僚宣言が発表された[190]。この閣僚宣言の AI 関連の概要（抄）は、図２６の通り。なお、閣僚宣言と併せて、「AI ガバナンスのグローバルな相互運用可能性を促進等するためのアクションプラン」も公表されている。また、この閣僚宣言に向けて、東京大学未来ビジョン研究センターは、3 月 28 日付けで、政策提言を発表している[191]。

【図２６】G７デジタル・技術大臣会合閣僚宣言（2023 年 4 月 30 日）仮訳・AI 関連抄[192]

経済社会のイノベーションと新興技術の推進

- 34. 我々は、破壊的な新興技術の急速なイノベーションは、（略）、そのようなデジタル技術のガバナンスや、誤用への対処を含む社会的影響の検討を必要とすることを共有する。この点について、我々はイノベーションの機会を活用しながら、法の支配、適正手続き、民主主義、人権尊重の原則を運用できる、機動的かつ柔軟で、より分散化した、マルチステークホルダーが参加するアジャイルガバナンスやその法的フレームワークの必要性を承認する。（略）
- 35. （略）我々は、また新興技術の影響と機会に関して、重要かつ新たなマルチステークホルダーフォーラムである OECD の技術に関するグローバルフォーラムの設立を歓迎する。

責任あるAIとAIガバナンスの推進

- 42. 我々は、OECD の AI 原則に基づき、人間中心で信頼できる AI を推進し、AI 技術がもたらす全ての人の利益を最大化するために協力を促進するとのコミットメントを再確認する。我々は、民主主義の価値を損ない、表現の自由を抑圧し、人権の享受を脅かすような AI の誤用・濫用に反対する。
- 43. 我々は、AI ガバナンスに関する国際的な議論と、AI ガバナンスの枠組み間の相互運用性の重要性を強調する一方、信頼できる AI という共通のビジョンと目標を達成するためのアプローチと政策手段は、G7 メンバー間で異なる場合があることを認識している。（略）
- 46. （略）我々は、AI が私たちの社会に与える潜在的な影響に留意しつつ、AI の政策や規制は、技術的・制度的特性だけでなく、地理的・分野的・倫理的側面を含む社会的・文化的影響に配慮した形で、適用の状況に適合させる必要があると認識している。
- 47. 生成 AI 技術が国や分野を超えてますます顕著になっていることを踏まえ、これらの技術の持つ機会と課題を早急に把握し、これらの技術が発展する中で、安全性や信頼性を促進し続ける必要があると認識している。
  我々は、AI ガバナンスや著作権を含む知的財産権の保護、透明性の促進、外国からの情報操作を含む偽情報への対処方法や、責任ある形での生成 AI を活用する可能性といったテーマを含む生成 AI に関する G7 における議論を引き続き行うための場を設けることを計画している。
  これらの議論は、専門知識を活用し、政策展開の影響に関する分析を検討する OECD や、関連する実践的なプロジェクトを実施する GPAI などの国際機関を活用する必要がある。

---

188 YahooJAPAN ニュース（朝日新聞 Digital）「「責任ある AI へ共通基準を」　G7 デジタル相会合、宣言の原案判明」4/21(金) 18:46
https://news.yahoo.co.jp/articles/9a6f1888a1b24aa87d9f8dce6597d7235ea6e008

189 https://g7digital-tech-2023.go.jp/index.html

190 「G7 群馬高崎デジタル・技術大臣会合」の開催結果
https://g7digital-tech-2023.go.jp/topics/topics_20230430.html
経済産業省「G7 群馬高崎デジタル・技術大臣会合を開催しました」2023 年 4 月 30 日
https://www.meti.go.jp/press/2023/04/20230430001/20230430001.html
総務省「G7 群馬高崎デジタル・技術大臣会合の開催結果」令和 5 年 4 月 30 日
https://www.soumu.go.jp/menu_news/s-news/01tsushin06_02000268.html
デジタル庁「G7 群馬高崎デジタル・技術大臣会合の開催結果」2023 年 4 月 30 日
https://www.digital.go.jp/news/efdaf817-4962-442d-8b5d-9fa1215cb56a/

191 東京大学未来ビジョン研究センター「AI ガバナンス協調への道筋：G7 サミットに向けた政策提言」2023.03.28
https://ifi.u-tokyo.ac.jp/news/15309/

192 出典：「G7 群馬高崎デジタル・技術大臣会合」の開催結果
https://g7digital-tech-2023.go.jp/topics/topics_20230430.html

上記内容のポイントをまとめると、以下の３点に集約できる。

- A）全体論として、法の支配、適正手続き、民主主義、人権の尊重、イノベーションの機会の活用の５原則を明示した上で、AI の利用に関しては、「民主主義の価値を損ない、表現の自由を抑圧し、人権の享受を脅かすような AI の誤用・濫用に反対する」ことを明記。
- B）一方、AI ガバナンス体制については、その「アプローチと政策手段は、G7 メンバー間で異なる場合があること」を認識し、また、「技術的・制度的特性だけでなく、地理的・分野的・倫理的側面を含む社会的・文化的影響に配慮」する必要性の認識を明確化。
- ３）その上で、全体としては、OECD にフォーラムを設立するとともに、特に生成 AI に関しては、G7 で引き続き議論を行う場を設けるとし、その際、OECD や GPAI を活用する必要があるとしている。

このうち、A）の全体論については、まとめの(3)の中国問題に相当し、また、B）の AI ガバナンス論については、まとめの(1)の AI ガバナンスの多様性に係る論点に相当する。C)の国際的な議論は、今後様々なことが議論の対象になるものと考えられるが、そのうち、まとめの(2)に記載した著作権問題、民主主義問題などについても今後の議論の対象になるものと考えられる。

＜今後の方向＞

以上、本ワーキングペーパーにおいては、ChatGPT の技術的特徴やその想定されるリスクを考察した上で、世界における ChatGPT 型の AI システムに対する規制・ガバナンス政策に係る現時点までの動向について分析を行ってきた。

しかしながら、AI 規制・ガバナンス制度に係る政策に係る議論は、まだ始まったばかりであり、今後とも国際的に様々な動きが想定されることはもちろんのこと、ChatGPT などの生成系 AI に係る技術やその産業アーキテクチャも、今後とも引き続き大きく進展・変化すること想定される。

このため、今後の世界の AI 規制・ガバナンス制度の進化の検討にあたっては、引き続きその動向を見るだけでなく、技術・イノベーションの進展もフォローすることが必要である。また、このような急激な AI 技術の進化の中で、如何にアジャイルな政策・規制体系が構築されていくかも、世界的に益々大きな課題になるものと考えられる。

（別添参考）FLI 公開書簡（仮訳）

巨大な AI 実験の一時停止: 公開書簡（仮訳）

私たちはすべての AI 研究所に対し、GPT-4 よりも強力な AI システムの訓練を少なくとも 6 か月間直ちに一時停止するよう呼びかけます。

人間と競合する知性を備えた AI システムは、社会と人類に重大なリスクをもたらす可能性があることが、広範な研究で示され、また、主要な AI 研究所によって認められています。広く支持されている Asilomar AI Principles で述べられているように、「高度な AI は地球上の生命の歴史に大きな変化をもたらす可能性があり、相応の注意とリソースで計画および管理する必要があります」。 残念ながら、このレベルの計画と管理は実現していません。ここ数か月間、AI 研究所は制御不能な競争に巻き込まれ、誰も (作成者でさえも) 理解、予測、確実な制御ができないような、より強力なデジタルな心・精神を開発および普及をしています。

現代の AI システムは現在、一般的なタスクで人間と競争できるようになってきており、私たちは、以下について自問する必要があります:

- 「私たちは、私たちの情報チャネルを、機械によるプロパガンダや虚偽で溢れさせてなければならないのでしょうか?」
- 「私たちは、充実した仕事を含め、すべての仕事を自動化する必要がありますか?」
- 「私たちは、私たちよりも数で上回り、賢く、私たちを時代遅れにし、私たちに取って代わる可能性のある非人間的な心を開発する必要がありますか?」
- 「私たちの文明の制御を失う危険を冒すべきですか?」

そのような決定は、選挙で選ばれていないような技術リーダーに委任してはなりません。**強力な AI システムは、その効果が肯定的であり、リスクが管理可能であると確信した場合にのみ開発されるべきです。** この信頼は十分に正当化され、システムの潜在的な影響の大きさに応じて増加する必要があります。 汎用人工知能に関する OpenAI の最近の声明では、次のように述べています。「ある時点で、将来のシステムの訓練を開始する前に、独立したレビューを受けることが重要になるかもしれません.。」 私達は同意します。そして、 その時点は今です。

したがって、**すべての AI 研究所に対して、GPT-4 よりも強力な AI システムの訓練を少なくとも 6 か月間直ちに一時停止するよう求めます。** この一時停止は公開され、検証可能であり、すべての主要な関係者が含まれている必要があります。 そのような一時停止を迅速に制定できない場合、政府は介入してモラトリアムを設定する必要があります。

AI 研究所と独立した専門家は、この一時停止を利用して、高度な AI の設計と開発のための一連の共有安全プロトコルを共同で開発および実装する必要があり、そのプロトコルは独立した外部の専門家によって厳密に監査および監督される必要があります。 これらのプロト

コルは、それに準拠するシステムが合理的な疑いを超えて安全であることを保証するものであることが必要です。 これは、一般的な AI 開発の一時停止を意味するものではありません。単に、これまで以上に大規模で予測不可能なブラックボックスモデルへの危険な競争から、緊急機能を備えるべく一歩後退するだけです。

AI の研究開発は、今日の強力で最先端のシステムをより正確で、安全で、解釈可能で、透明性があり、堅牢で、整合性があり、信頼でき、忠実なものにすることに再び焦点を当てる必要があります。

並行して、AI 開発者は政策立案者と協力して、堅牢な AI ガバナンス システムの開発を劇的に加速する必要があります。 これらには、少なくとも次のものが含まれる必要があります。

- AI に特化した新しい有能な規制当局。
- 高度な能力を持つ AI システムと計算能力に係る大規模なプールの監視と追跡。
- 本物と合成を区別し、モデルのリークを追跡するのに役立つ来歴および透かしシステム。
- 堅牢な監査および認証エコシステム。
- AI によって引き起こされた損害に対する責任。
- 技術的な AI の安全性研究に対する強力な公的資金提供。
- そして、AI が引き起こす劇的な経済的および政治的混乱 (特に民主主義) に対処するための十分なリソースを備えた機関。

人類は AI によって豊かな未来を享受できます。 強力な AI システムの作成に成功した今、私たちは「AI の夏」を楽しむことができます。そこでは、報酬を獲得し、これらのシステムを設計してすべての人に明らかに利益をもたらし、社会に適応する機会を与えることができます。 社会は、社会に壊滅的な影響を与える可能性のある他の技術を一時停止しています。ここでそれを行うことができます。 あわてて秋に突入するのではなく、AI の長い夏を楽しみましょう。

（以上）