生成 AI のエコシステムの変遷と競争の構図

2022 年 11 月 30 日に OpenAI が ChatGPT を発表して以降、生成 AI 産業は急速に成長し、さまざまな企業が現れた。わずか 3 年足らずの間に生成 AI 製品を学習し使用する人数が世界全体で 10 億人を超え、生成 AI 技術の急速な発展は個人、企業に必要な汎用技術を普及させた。しかし、AI 産業の発展は実際、数十年前に始まったのである。本稿では、生成 AI のエコシステムの変遷を重点的に紹介し、大きさの異なる言語モデルを 3 つに分類する。生成 AI のエコシステムの変遷と競争の構図における「大型旅客機、プライベートジェット機、紙飛行機」の 3 つのモデルを理解することは、AI 産業の急速な発展と市場ニーズの変化を迅速に把握し、評価するのに役立つ。

生成 AI のエコシステムの変遷

早期モデルから現在の最も先進的な状態までの発展の各変遷におけるマイルストーンから、生成 AI はイノベーションと改善を持続的に追求するエコシステムであることがはっきりと分かる。

1. 早期 AI モデルの巨人

生成 AI の早期発展段階に科学技術の大手各社が次々と自社の大規模言語モデルを発表し、技術革新の嵐を引き起こした。時間の経過に伴い、徐々にフェードアウトしたモデルもあれば、拡大し続け、市場トップになったモデルもあった。

Google の BERT モデルは Transformer アーキテクチャの早期アプリケーションとして、3.4 億のパラメータを有するものの、今ではかなり小規模な方である。にもかかわらず、BERT の自然言語処理分野における貢献は不滅であり、後続モデルの発展の基礎を固めた。

OpenAI の GPT-3 は 1,750 億のパラメータと驚くべき美しさのテキスト生成能力によって、一度は生成 AI の代名詞になった。それが発展したアプリケーションソフトウェア ChatGPT はさらに世界に広がり、2025 年 10 月現在、アクティブユーザーが 1 週間あたり 8 億人を超え、対話インタラクションとコンテンツ生成分野のアプリケーションの潮流をリードした。

Jurassic-1 は AI21 Labs が開発し、1,780 億のパラメータを有し、テキスト生成分野で極めて優れていたこともあった。GPT シリーズの台頭に伴い、Jurassic-1 の影響力は徐々に減少したものの、高品質のテキスト生成と文脈理解の能力は依然として軽視できない。

MT-NLG はマイクロソフトと NVIDIA が共同開発し、5,300 億のパラメータを有し、かつては最大規模の単体の Transformer 言語モデルであった。膨大なコンピューティングニーズと潜在的なリスクにより、MT-NLG は未公開で、アプリケーションは主に研究目的に用いられる。

Google の PaLM は 5,400 億のパラメータを有し、大規模言語モデル（LLM）分野の重要なマイルストーンである。PaLM は言語理解、推理、プログラミングなど多くのタスクにおいてパフォーマンスが優れており、Google の AI アプリケーションに大きな技術支援を提供した。

上述のモデルのほか、LaMDA（パラメータ 1,370 億）、Gopher（パラメータ 2,800 億）および BLOOM（パラメータ 1,760 億、オープンソース）などのモデルも早期発展段階において重要な役割を果たした。LaMDA は対話アプリケーションに特化し、Gopher は一般的な理解においてパフォーマンスが優れており、BLOOM は多くの言語能力とオープンソースの特性で研究開発企業に歓迎されている。

全体的にみて、生成 AI の早期発展は百家争鳴の局面を示した。各社はモデルの規模、アーキテクチャデザインおよびアプリケーションの分野の探求を続け、技術の急速な反復とイノベーションを推進した。現在、すでに一部のモデルは徐々にフェードアウトしたが、後続モデルの開発のために貴重な経験を蓄積し、生成 AI 発展の堅固な基礎を固めた。

2. 現代の巨人：巨人級の大規模言語モデル

次世代生成 AI の巨人級モデルはさらに複雑であり、先進的な技術、アーキテクチャを融合し、より大きな規模とより高い能力を有する（表 1-1 参照）。

表 1-1　現代の巨人

トップクラスの AI モデルの比較分析はパラメータ、訓練された token、アーキテクチャデザインおよびそれぞれの発表時期に基づき評価した。アーキテクチャ欄の MoE、Dense、CoE の表示はさまざまなモデルのアーキテクチャ

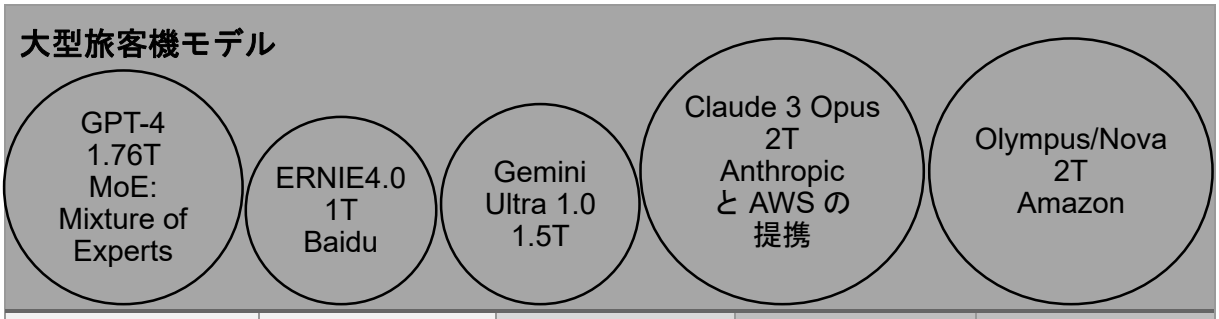
である。

モデル Model	開発者 Lab	パラメータ Parameters (B)	訓練されたトークン Tokens trained (B)	発表時期 Announced	アーキテクチャ Arch
Olympus/NOVA	Amazon	2,000	40,000	Dec/2024	
GPT-5	OpenAI	5,250		TBA	
GPT-6	OpenAI			TBA	
AuroraGPT	Argonne National Laboratory	1,000		TBA	
Grok-2	xAI			TBA	
PanGu 5.0 Super	Huawei	1,000	20,000	Jun/2024	MoE
Yi-XLarge	01-ai	2,000	20,000	May/2024	MoE
Yi-Large	01-ai	1,000	15,000	May/2024	Dense
Med-Gemini-L 1.0	Google DeepMind	1,500	30,000	May/2024	Dense
Inflection-2.5	Inflection AI	1,200	20,000	Mar/2024	Dense
Glaude 3 Opus	Anthropic	2,000	40,000	Mar/2024	Dense
Samba-1	SambaNova	1,400	20,000	Feb/2024	CoE
Gemini 1.5 Pro	Google DeepMind	1,500	30,000	Feb/2024	MoE
Gemini Ultra 1.0	Google DeepMind	1,500	30,000	Dec/2023	Dense
Inflection-2	Inflection AI	1,200	20,000	Nov/2023	Dense
ERNIE 4.0	Baidu	1,000	20,000	Oct/2023	Dense

鋭企作成，参考資料：LifeArchitect.ai

GPT-4：OpenAI の GPT-4 バージョンは 1.76 兆と驚くべきパラメータを有することは表 1-11 に示した通りである。GPT-4 は混合専門家（Mixture of

Experts, MoE）を採用し、部分ごとに異なるタスク処理に特化させており、アーキテクチャはより高い正確性と文脈理解能力を開発し、さらに広範囲のタスクを行うことができる。



ナノ （Nano）		超小規模 (Extra Small)		小規模 (Small)		中規模 (Medium)		大規模 (Large)	
Gato (Cat)	1.0	Gemma	7.0	Palmyra	20	Command xlarge	52	Yuan 2.0	103
TinyLlama	1.1	Gauss	7.0	AlexaTM 20B	20	StableLM	65	Command-R+	104
Meta-Transformer	2.0	DeciLM-7B	7.0	Codestral	22	DeepSeek	67	InternLM	104
PanGu-Coder	2.6	Zephyr	7.3	C1.2	33	Llama 3 70B	70	DBRX	132
Meena	2.6	Mistral 7B	7.3	Code Llama 34B	34	Eurus	70	BLOOM (tr11-176B-ml)	176
Phi-2	2.7	Striped Hyena 7B	7.7	Yi-34B	34	pplx-70b-online	70	Jurassic-2	178
Mamba	2.8	Persimmon - 8B	8.0	Command-R	35	Luminous Supreme Control	70	Mistral-medium	180
Apple On-Device model	3.0	LLaMA Pro	8.3	mixtral-8x7b-32kseqlen	47	Qwen2	72	Falcon 180B	180
		SOLAR-10.7B	11.0	Retro 48B	48			Grok-1	314
		Pythia	12.0					PaLM 2	340

…Many more		…Many more		…Many more		…Many more		…Many more	
紙飛行機モデル						**プライベートジェット機モデル**			

図 1-1　生成 AI 大中小 3 つの競争の構図に分類

現代の巨人のうち、パラメータ 1 兆超はこの表の上部にある。モデルはナノ（Nano）、超小規模（Extra Small）、小規模（Small）、中規模（Medium）、大規模（Large）に分け、数字の単位は Billion（10 億）である。大中小 3 つの構造は大きいものから並べ、大型旅客機、プライベートジェット機、紙飛行機と 3 つの異なるパラメータレベルのモデルに例えた。

鋭企作成，参考資料：LifeArchitect.ai

Gemini Pro と Gemini Ultra：Google を代表する重大な進展は前世代モデル PaLM の自然言語理解と生成能力を改善させた成功に基づく。Gemini Ultra はパラメータ 1.5 兆に達し、自主監督学習における最先端の技術を結合し、大量のアノテーションデータなしでさまざまなタスクを行うことができる。

Claude 3：Anthropic と AWS が共同開発した Claude 3 は 2 兆のパラメータを有し、科学技術企業間で協力して AI の能力を練成した。当該モデルの特色は責任を負う AI であり、安全性と倫理を考慮して設計し、人類の価値観に合致する。

ERNIE 4.0：百度（バイドゥ）の ERNIE 4.0 は 1 兆のパラメータを有する。中国の科学技術大手の AI 研究に対する大きな投資を表しており、知識と言語理解の統合に特化している。ERNIE 4.0 はナレッジグラフ（knowledge graph）を利用して理解と生成能力を増強し、複雑な情報検索と生成タスクにおいてパフォーマンスが優れている。

3.　生成 AI 開発の主な動向

早期から現代まで大規模言語モデルの変遷はパラメータ数の増加だけでなく、モデルの方法とアプリケーションの多様化を示している。生成 AI 開発の主な動向を挙げる。

- モデルのパラメータの指数成長

生成 AI のパラメータの数と規模は爆発的に成長し、GPT-4、Gemini および Claude 3 などのモデルは規模において前世代を超越し、数兆のパラメータに達した。この成長は生成 AI の研究と開発の投資と興味が増加し続けていることを証明している。

•クローズドソースモデルの主導的地位

BLOOM と OPT などのオープンソースモデルは依然として存在するものの、現代の巨人の情報を整理すると、クローズドソースモデルの主導的地位がわかる。特に Google、OpenAI などの科学技術企業が発展させたモデル Gemini と GPT-4，および Anthropic の Claude 3 である。こうしたモデルは通常 API 経由のみでアクセスし、透明性と使用者の管理を制限できる。

•特化型モデルの出現

現代の言語モデルはブレークスルー後、特定のタスクデザイン対象に転向する能力を有する特化型 LLM が現れた。例えば、Med-PaLM と PaLM-Coder はそれぞれ医療とプログラミングのタスクのカスタマイズに特化し、LLM はカスタマイズのニーズが徐々に高まっていることを表している。

•競争の構図

各社のモデルは大きさ、機能および能力においてトップの地位を奪い合っている。大規模言語モデルのほか、その他のモデルも大きさ、パラメータおよび所属する開発チームがさまざまであり、世界がこの分野で努力していることを反映している。

生成 AI の競争の構図における 3 つのモデル

生成 AI 言語モデルの発展と運用は非常に多様で、全体に適応すること（one size fits all）は難しく、「一言」では言えない。しかし、われわれは言語モデルの大きさに基づき、大中小 3 つの競争の構図に分けることができる。航空機の市場区分で説明すると、大きい方から大型旅客機、プライベートジェット機、紙飛行機と異なる言語モデルの能力の範囲とデプロイの規模を象徴的に比喩できる。

生成 AI は急速に発展し、徐々に多様化し、各モデルは独特の長所と短所を有する。機能が強大で、用途も広いモデルもあれば、専門的で効率が高いモデルもある。大規模モデルは強大な AI サーバー又はクラウドインフラを必要とし、比較的小さいモデルはエッジデバイス（edge devices）にデプロイして特定のタスクを実行できる。

1. 大型旅客機モデル

「大型旅客機モデル」は大企業が運営する強大な生成 AI を表し、GPT-4、Gemini Ultra、Claude 3 および Amazon の Olympus など、ボーイング 787 旅客機のようである。これらのモデルはウェブアプリケーションソフトウェア（web applications）又はモバイルアプリケーションソフトウェア（mobile apps）通じて、誰もが使用できるようにしている。また、API 経由で企業又は開発者にサービスを提供するが、リースでのみ使用でき、自ら構築又は保有はできない。

大規模言語モデルは数千億から 1 兆を超えるパラメータを有し、複雑なタスク処理能力を備え、自然言語処理、画像生成、ソースコード編集の各分野においてパフォーマンスが卓越しているが、膨大なモデル規模は相対的にコンピューティングコストの高騰をもたらす。「大型旅客機モデル」により企業は最先端の AI 技術を利用することができるが、インフラの負担を負う必要がない。

大規模言語モデルの急速な発展は各世代のモデル訓練に使用するデータセットが明らかに拡大したためであり、数年間という短い期間で世代ごとに規模が約 10 倍増加したように見える。こうした成長速度は持続できない可能性があると考える専門家もいれば、より大きなデータセットとモデルが生成 AI の知識と正確さを引き続き高めると考える専門家もいる。

生成 AI は初期段階にあり、かなり大きな発展のポテンシャルがまだあり、正確性を高めるにはより大きなモデルを必要とし、又は汎用モデルを異なる特定分野を対象にしたより多くのモデルに細分化する可能性もある。データセットの抽出とクラウドリソースの獲得可能性が高まるにつれ、言語モデルのパラメータ数は指数関数的に増加し続け、こうしたイノベーション革命は今後 10 年でようやく技術の極限に達する可能性がある。

大規模言語モデルの多くがクラウドコンピューティングを採用しているた

め、企業は使用時、資料を API 経由でクラウドまで伝送しなければならず、潜在的な情報セキュリティリスクがあり、企業機密は LLM サプライヤーに掌握される可能性がある。このため、多くの企業は従業員に ChatGPT、Gemini などの生成 AI を使用することを暫定的に禁止し、商業機密が漏えいすることを回避する。一部の企業は独自の言語モデル開発への転向を選択している。

2. プライベートジェット機モデル

「プライベートジェット機モデル」は企業が小型汎用航空機メーカー、セスナ社にプライベートジェット機を購入させることができるようなもので、企業が Mistral、DeepSeek、Falcon および Grok-1 などの生成 AI モデルを保有、管理できることを表している。これらの生成 AI モデルは通常中小企業又は研究機関が開発し、より高い柔軟性とカスタマイズの潜在能力がある。特定のアプリケーション分野において、そのパフォーマンスは大規模モデルに劣らずかつよりデプロイ、メンテナンスしやすい。

こうしたモデルの優位性は企業が自身の IT インフラにデプロイでき、データの安全性とプライバシーを確保できることにある。銀行、病院又は半導体企業など機微情報を処理する企業に対し、情報のセキュリティ考慮に基づき、通常はクラウド型生成 AI サービスを使用せず、オンプレミス型生成 AI モデルを使用する傾向がある。

生成 AI 技術の普及に伴い、企業は IT インフラを現代化し、専門家を雇って管理、メンテナンスする必要がある。しかし、AI 専門人材の募集は多くの課題に直面している。IDC の研究によると、世界の 70%弱の企業は AI システムの運営時に人材不足の問題に遭遇し、この不足は 2026 年まで続き、経済に約 5.5 兆米ドルの影響を及ぼす可能性がある。マッキンゼー・アンド・カンパニーは、より多くの中小企業が外部の生成 AI 技術サービスを求め始め、この市場は 2029 年までに 2,000 億米ドルを超過する見込みであると指摘する。

3. 紙飛行機モデル

「紙飛行機モデル」は Llama、Mistral および Apple OpenELM のようなエッジデバイス上で実行する小規模モデルである。これらのモデルは企業や個人が管理し、通常ローカル又はエッジデバイス上にデプロイする。主な優位性

はクラウドとデータセンターに対する依存を減少させ、より迅速な反応とより個別化したサービスを提供でき、かつ運営コストが低いことにある。

エッジ生成 AI（edge GenAI）はデータのプライバシーを保護でき、大部分の処理がローカルで完了するため、機微データをクラウドに伝送する必要がない。大規模商用生成 AI モデルと比較して、これらの小規模言語モデルはユーザーのニーズに迅速に対応し、遅延性を下げ、データの安全性を高めることができる。これは自動車の自動運転、パーソナルアシスタントのアプリケーションにとって特に重要である。このほか、これらのモデルのデザインにより企業は特定のニーズに基づき調整し、最適化し、スマートフォン、自動車、IoT 機器など多様なシーンに適用できる。

Tirias Research のレポートによると、2028 年までにエッジ生成 AI 技術が 20%のクラウド処理ニーズを減少させ、企業は 160 億米ドルもの運営コストを節約できる見込みである。この変化は一部の AI 業務の負荷をローカルデバイスに移転したことによるもので、データセンターの伝送とコンピューティングのコストを減少させ、かつ運営効率を高める。エッジ AI の普及に伴い、生成 AI のアプリケーションはより柔軟に変わり、企業と消費者により強大なサービスを提供する。

結論

生成 AI のエコシステムの変遷と競争の構図における大型旅客機、プライベートジェット機、紙飛行機の 3 つのモデルを理解することは、AI 産業の急速な発展と市場ニーズの変化を把握し、評価するのに役立つ。OpenAI が 2022 年 11 月 30 日に ChatGPT を発表して以降、生成 AI 産業の目覚ましい発展が起こり、3 年弱で GPT-3.5 から GPT-4、GPT-4o、GPT-4.5、GPT-5 までアーキテクチャが急速にアップグレードした。FirstPageSage の統計データによると、ChatGPT は発表以降、7 割超の大規模言語モデル（LLM）のシェアを維持し続けている。OpenAI の 2025 年の営業収入は 130 億米ドルに達する見込みで、主に ChatGPT の有料ユーザーと企業 API の権限授与によるもので、企業の顧客が 3 割を占め、7 割は個人向けであり、ダウンロード者数は世界で 10 億人を超えている。

しかし、この期間、市場のニーズと産業の競争状態は急速に変化し、OpenAI は二度とすべての市場を独占することはなく、特に企業レベルの言語モデル使用市場においては、Anthropic、Google、Meta などの大企業も自主開

発した LLM を発表し続け、生成 AI 産業は百家争鳴の爆発的発展状態に入っている。大量に参入する競争者を許容できる理由は、使用人数が急速に増加したためであり、生成 AI 製品使用ニーズの急速な増加はアプリケーション面で徐々に広がり、アプリケーション方式が豊富で多様で、市場の細分化をもたらした。

Menlo Ventures の生成 AI 報告書は次のように指摘する。生成 AI の技術の発展はより大きなモデルへの成長の追求から、特定のタスクにより特化した言語モデルを急速に発展させ、こうしたモデルは設置と運営のコストもより低減でき、カスタマイズのレベルもより高い。また、各クラウドサービス大手はユーザーによりよい使用体験を提供するため、多くのモデルを積極的に統合しており、例えば、2025 年 9 月にマイクロソフトが開発した Copilot エコシステムに Anthropic の Claude を組み込んだことを発表し、企業間の競合関係もより複雑になりつつあることが明らかである。

出典

•ChatGPT シェア 7 割安定！OpenAI 再融資 83 億米ドル、AI 資金戦の持続的過熱 CRYPTO CITY, 2025

•OpenAI 独占の終焉：LLM 戦争参入による群雄割拠時代 Vocus, 2025

·2025 Mid-Year LLM Market Update: Foundation Model Landscape + Economics. MENLO VENTURES (2025)