

2023年以降の最新生成AIモデルが研究開発現場にもたらした革新

エグゼクティブサマリー

2023年以降の生成AIは「汎用生成モデル（基盤モデル）→マルチモーダル化→長文コンテキスト→ツール利用・エージェント化」という順で能力の質が変化し、研究開発（R&D）の現場では“**文章生成ツール**”から“**実験・設計・検証のオーケストレータ（指揮者）**”へ役割が拡張している。特に、①音声・画像・動画まで一体として扱うマルチモーダル化、②数十万～数百万～（主張として）千万トークン級の長文脈処理、③コード実行や検索、実験装置APIなど外部ツールを組み込む設計が、研究の生産性・再現性・共同作業形態そのものを変え始めた。¹

主要トレンドは次の3点に集約できる。

第一に、**R&Dの知的作業の分解・再結合が進んだ**。文献調査・要約・仮説生成・実験計画・解析・ドキュメント化が、長文脈モデルとRAG（検索拡張生成）やエージェント枠組みによりパイプライン化され、レビュー可能な形で半自動化されつつある。²

第二に、“**生成**”が**科学・工学のコアに侵入**している。画像生成で成熟した拡散モデル（Diffusion）は、材料設計や生体分子モデリングでも中心技術になり、構造生成・相互作用予測・候補探索の効率を大きく変え始めた。³

第三に、**ガバナンスとコストが設計要件化**した。日本の「AI事業者ガイドライン」、G7広島AIプロセス、EU AI ActのGPAI義務、NISTのGenAIプロファイルなどにより、技術導入は性能だけでなく、文書化・リスク評価・透明性・著作権対応・運用監視まで含む“製品化”能力が問われる。⁴

推奨（実務向け要点）は、“**基盤モデル選定**”より先に“**R&Dワークフローの設計**”を行い、(1)知識基盤（RAG/知識グラフ/実験DB）、(2)検証基盤（シミュレーション・自動実験・テスト）、(3)評価・統制（ベンチ、監査ログ、レッドチーム）を三位一体で整えることにある。これにより、モデルが更新されてもR&Dシステムは壊れにくくなる。⁵

調査範囲と方法

対象期間は**2023年1月～2026年2月16日（日本時間）**とし、モデルはユーザー指定例（GPT-4系、LLaMA系、Claude系、Gemini系、Stable Diffusion系）を中心に、追加モデルは「特に指定なし」として重要度に応じて補足した。一次情報の優先順位は、**日本語の政府・企業公式発表**、次いで**英語原著論文（arXiv/学術誌）**、主要企業の技術ブログ・モデルカード・システムカード、最後に補助的に業界レポート・報道とした。⁶

主要な開発主体として、OpenAI⁷、Anthropic⁸、Google DeepMind⁹、Meta¹⁰、Stability AI¹¹を中心に、国内動向としてNTT¹²、Fujitsu¹³、Preferred Networks¹⁴の基盤モデル展開を補足した。¹⁵

規制・ガイドラインは、経済産業省¹⁶と総務省¹⁷の「AI事業者ガイドライン（第1.0版）」、G7¹⁸「広島AIプロセス」、欧州委員会¹⁹が公表するEU AI Act（GPAIモデル義務）、NIST²⁰のAI RMFおよびGenAIプロファイルを基礎に整理した。⁴

技術的革新

本節では、R&Dへの波及が大きい「技術の質的变化」を、アーキテクチャ・学習/微調整・マルチモーダル・評価・推論基盤の観点でまとめる。

最も大きい転換の一つは、**マルチモーダルを“パイプライン接続”ではなく“単一モデルでエンドツーエンド”に寄せる動き**である。例えばGPT-4oは、テキスト/音声/画像/動画入力を単一ネットワークで処理し、テキスト・音声・画像出力まで統合する方向性を示し、会話音声応答の遅延を最小232ms（平均320ms）と説明している。²¹ 研究現場では、実験装置の音・画像ログ、計測スクリーンショット、ホワイトボード写真、動画に残る操作手順など、従来テキスト化がボトルネックだった情報を“そのまま入力して理解させる”入口を広げた点が重要である。²²

第二の転換は、**長大コンテキスト**である。Gemini 1.5の技術報告は、数百万～（検証主張として）少なくとも1000万トークン規模までの長文脈処理と、長文脈検索課題で>99%のリコールを報告し、長文書QA・長時間動画/音声の扱いを強調している。²³ これは、研究開発で日常的に発生する「複数論文＋仕様書＋実験ログ＋コード＋議事録」を一つの文脈に束ね、整合性を取りながらレビューする作業を、モデル側の制約（コンテキスト長）で分割せずに済む可能性を高める。²³ さらにGemini 2.0は“agentic era”を掲げ、ネイティブな画像・音声出力やツール利用を強調しており、R&Dにおける「調べる→試す→まとめる」を統合する設計思想が前面化している。²⁴

第三の転換は、**計算効率と学習効率の改善が“研究現場の手触り”を変えた点**である。オープンウェイトLLMではMoE（Mixture-of-Experts）により推論時に一部の“専門家”のみを動かし性能/コスト比を上げる設計が広がり、Mixtral 8x7BはスパースMoEとして公開され、性能と推論効率の両立を強調する。²⁵ また近年のLLaMA系では、モデルをより広く配布可能にする戦略（オープンウェイト）自体が研究コミュニティの再利用・派生モデル開発の速度を押し上げた。²⁶

学習・微調整では、**小規模計算資源でも現場適用できる“軽量ファインチューニング”**が実務上の意味を持つようになった。QLoRAは、4bit量子化とLoRAを組み合わせることで65Bモデルを単一48GB GPUで微調整できるとし、メモリ削減の具体策（NF4、double quantization、paged optimizersなど）を提示した。²⁷ さらにDPOは、RLHFの複雑さ（報酬モデル学習＋強化学習）を回避し、選好学習をより単純な損失で扱う方向を示し、アラインメントの実装コストを下げる可能性を示した。²⁸ これらは、R&D組織が自社データや実験プロトコルに合わせてモデル挙動を調整する際の“費用対効果”を改善し、オンプレ/閉域での展開選択肢も広げた。²⁹

拡散モデル側でも同様に、**アーキテクチャ革新が品質だけでなく運用要件（VRAM、速度）を規定する**ようになった。Stable Diffusion 3は、画像と言語で別重みを持つMMDiT（Multimodal Diffusion Transformer）を採用し、タイポグラフィやプロンプト追従性の改善を主張しつつ、初期の推論試験として「8BモデルがRTX 4090（24GB VRAM）に収まり、1024×1024生成に34秒（50ステップ）」など運用目線の情報を提示している。³⁰ SDXLも、UNet大型化や2つ目のテキストエンコーダ等で品質向上を報告しており、画像生成が設計レビューや資料作成に組み込まれる前提が整った。³¹

評価の面では、R&D用途に近いベンチマークが重視される。マルチモーダル推論ではMMMUが大学レベルの専門領域問題と多様な図表を含むベンチとして提案され、最先端モデルでも難しい領域が残ることを示した。³² ソフトウェア工学ではSWE-benchが実際のGitHub issue修正を課題化し、モデルが“現実の変更”にどこまで耐えるかを測る枠組みを提供した。³³ さらに、包括的評価（HELMなど）やAI Indexが示すように、単一ベンチの飽和・汚染（contamination）懸念が前提となり、“**自社課題での評価設計**”が研究開発の**品質保証工程に組み込まれる**方向が強い。³⁴

推論・提供（Serving）では、**高スループット化＝研究現場での“待ち時間”削減**が重要になる。vLLMはPagedAttentionに基づくメモリ管理でKVキャッシュの無駄を減らし、高速なサービングを狙う設計を示し、

最大24倍のスループット（比較条件付き）を主張している。³⁵ これは、社内実験で「多数プロンプト→多数候補→自動評価」を回す場合のコストに直結する。³⁶

研究開発プロセスへの影響

R&Dへの影響は「工程の一部自動化」ではなく、**工程の再設計（workflow re-engineering）**として理解した方が正確である。なぜなら、最新モデルは“文章を出す”だけでなく、検索・コード実行・API呼び出し・推論時間制御などを通じて「仕事の進め方」を変えるからである。³⁷

まず設計・企画段階では、長文脈とマルチモーダル入力により、仕様書・論文・ログ・図表を束ねたレビューが現実的になった。Gemini 1.5は実務カテゴリでの時間短縮（26～75%）を例示しており、少なくとも“業務時間の一部を置き換える”段階に入ったことが示唆される。²³ 次に実験・検証段階では、LLMが「実験を実行するためのツール」を扱うことで、研究者の“手続き負担（コーディング、装置操作言語、ドキュメント探索）”を減らす方向が具体化している。Coscientistは、インターネット/ドキュメント検索、コード実行、実験自動化APIを組み合わせ、複雑実験の設計・計画・実行を（半）自律的に行う例としてNature論文で提示された。³⁸

さらに、生成AIはドキュメンテーションと再現性の実装にも影響する。Coscientistの大学側発表は、システムが各ステップを追跡・文書化しトレサブルにすることで再現性に寄与しうる点を強調している。³⁹ これは「人が頑張ってノートを書く」から「システムが作業ログを自動的に残し、監査可能な形に整える」へのシフトであり、研究不正防止というより、**研究速度を落とさずに品質保証を入れる**方向の意味が大きい。⁴⁰

一方で、生成AIがR&Dに入るほど、**検証（verification）と評価（evaluation）がボトルネック化する**。例えばソフトウェア開発では、単なるコード生成から、レポジトリ全体を理解して修正する課題（SWE-bench）へ評価軸が移り、現実の工程に近いほど難易度が跳ね上がることが示されている。⁴¹ そのためR&D現場では、生成AIを「案を出す装置」として使いつつ、**テスト・シミュレーション・再実験・多重レビュー**で最終成果物の信頼性を担保する設計が不可欠になる。⁴²

研究開発ワークフローの代表的な再設計

以下は、生成AI（基盤モデル）を中核に据える際の典型的な工程分割である（モデル固有ではなく、2023年以降の実装例の共通構造として整理）。⁴³

flowchart TD

```
A[研究課題の定義] --> B[知識収集: 論文・特許・仕様書・実験ノート・コード]
B --> C[知識基盤化: RAG/索引/知識グラフ/ベクトルDB]
C --> D[生成: 仮説・設計案・実験手順・解析コード・説明資料]
D --> E[検証: シミュレーション/単体テスト/自動実験/レビュー]
E --> F[評価: 指標・ベンチ・人手評価・リスク評価]
F --> G[文書化: 監査ログ/再現手順/変更履歴/データ系譜]
G --> H[学習/改善: プロンプト更新・RAG更新・微調整]
H --> C
```

erDiagram

```
PROJECT ||--o{ EXPERIMENT : includes
PROJECT ||--o{ DOCUMENT : produces
EXPERIMENT ||--o{ RUN : has
```

```
RUN }o--|| MODEL_VERSION : uses
RUN }o--|| PROMPT_TEMPLATE : configured_by
RUN }o--o{ ARTIFACT : outputs
ARTIFACT }o--o{ EVALUATION : assessed_by
DOCUMENT }o--o{ CITED_SOURCE : cites
ARTIFACT }o--o{ DATASET : derived_from
```

業務適用事例と実証結果

この期間の“革新”は、特定ドメイン（化学・生物・材料・ソフトウェア）で特に明確である。理由は、(1)検証手段（実験・計算・テスト）があるため成果を測りやすい、(2)ドキュメント/データが豊富でRAG適性が高い、(3)自動化装置やCI/CDなど既存の自動化基盤と接続しやすい、の3点にある。⁴⁴

化学・実験自動化では、Coscientistが象徴的で、複数LLMとツール（検索、コード実行、実験装置API）を束ねる構成を明示した点が重要である。ここでの新規性は「LLMが化学を理解した」だけではなく、“**研究手続きの言語（装置API、ドキュメント、計算）を統合し、実行まで到達する**”設計が提示されたことにある。⁴⁵

生体分子・創薬では、拡散モデルの考え方が科学モデリングに入り、AlphaFold 3が複合体（タンパク・核酸・低分子等）の統合予測を拡散ベースで扱うと説明されている。⁴⁶ 企業側発信では、既存手法比で相互作用予測が少なくとも50%改善、カテゴリによっては精度倍増といった定量主張があり、ベンチ（PoseBusters）でも従来法を上回ると述べられている。⁴⁷ ここでのR&D上の意味は、ウェット実験を置き換えるというより、**探索空間を縮め、候補の優先順位付けと仮説の生成速度を上げる**点にある。⁴⁸

材料探索では、MatterGenが代表的で、拡散モデルで結晶構造（原子種・座標・格子）を生成・微調整し、従来生成モデルより「新規かつ安定」な生成物が倍以上、エネルギー極小に10倍以上近いなどの指標を示し、生成した材料を実際に合成し目標値の20%以内を達成したと報告する。⁴⁸ これは、R&Dを“スクリーニング中心”から“条件付き生成＋検証中心”へ寄せる強い実例である。⁴⁸

ソフトウェア開発（研究開発の基盤工程）では、生成AIの効果が最も測定されている領域の一つである。GitHub Copilotの統制実験はタスク完了が55.8%速いと報告し、現場に近い大規模RCT分析でも、週あたりPR数の増加（Microsoftで12.92～21.83%、Accentureで7.51～8.69%）が示唆される。⁴⁹ さらに2025～2026にかけては、コーディング“支援”から「長時間タスクをツール利用で進めるエージェント型」へ重点が移り、GPT-5.3-Codexは“agentic coding model”として、ツール利用や長期実行を想定した説明がなされている。⁵⁰

国内では、基盤モデルの“国産化/小型化/閉域利用”がR&Dと経済安全保障の両面から加速している。NEDOと経産省はGENIACとして計算資源支援を実施し、国内での基盤モデル開発テーマを複数支援した。⁵¹ また文科省資料は、理化学研究所で科学研究向けドメイン指向基盤モデルの開発を2024年4月から本格化したと記す。⁵² 企業側でも、NTTは軽量LLMとしてtsuzumiの商用提供開始を述べ、日本語処理と効率の両立を打ち出している。⁵³ 富士通はTakaneを安全なプライベート環境向け日本語LLMとして発表し、JGLUEでの高成績やRAG/監視技術と組み合わせた企業導入を強調している。⁵⁴ PFNもPLaMo系の展開を継続し、国内基盤モデルの選択肢が増えている。⁵⁵

代表的な事例の表

組織名	モデル/技術	用途	効果/定量的成果	導入時期	出典URL
Carnegie Mellon University ⁵⁶ (連携：Emerald Cloud Lab ⁵⁷)	LLMエージェント (Coscientist ; GPT-4等+検索/コード実行/実験API)	実験計画～実行(クラウドラボ含む)	6タスクで実証(合成計画、ドキュメント探索、クラウドラボ実行、液体ハンドリング制御、複合モジュール実験、反応最適化など)	2023年12月(論文)	https://www.nature.com/articles/s41586-023-06792-0 ³⁸
Isomorphic Labs ⁵⁸ (共同開発：Google DeepMind)	AlphaFold 3 (拡散ベースの統合構造予測)	創薬/構造生物学(複合体予測)	相互作用予測で少なくとも50%改善、カテゴリによって精度倍増(企業側主張)	2024年5月	https://www.isomorphiclabs.com/articles/alphafold-3-predicts-the-structure-and-interactions-of-all-of-lifes-molecules ⁴⁷
Microsoft ⁵⁹ (研究)	MatterGen (拡散モデルによる材料生成)	材料探索(逆設計)	生成構造が「新規かつ安定」になりやすい(2倍超)、局所エネルギー極小に10倍超近い。生成候補を合成し目標値の20%以内	2025年(論文)	https://www.nature.com/articles/s41586-025-08628-5 ⁴⁸

組織名	モデル/技術	用途	効果/定量的成果	導入時期	出典URL
GitHub ⁶⁰ （導入先：Microsoft/Accenture等）	Copilot（LLMによるコーディング支援）	研究開発の基盤工程（実装・検証）	統制実験で55.8%高速。現場RCT分析で週PR数が増える示唆（Microsoftで12.92–21.83%、Accentureで7.51–8.69%）	2023年（論文） / 2024年（分析）	https://arxiv.org/abs/2302.06590 ⁶¹ / https://mit-genai.pubpub.org/pub/v5iixksv ⁶²
OpenAI ⁷	エージェンツ型コーディングモデル（GPT-5.3-Codex）	長時間の実装/調査/ツール利用を伴う開発	“agentic coding model”として長時間タスク・ツール利用・複雑実行を想定、25%高速化を説明	2026年2月5日	https://openai.com/index/introducing-gpt-5-3-codex/ ⁵⁰
Google DeepMind	Gemini 1.5（長大文脈・マルチモーダル）	長文文献/マルチ資料の理解・調査支援	10職種で26–75%の時間短縮例、長文脈検索で>99%リコール（主張）	2024年3月（v1）	https://arxiv.org/abs/2403.05530 ²³
Stability AI	Stable Diffusion 3（MMDiT、Rectified Flow）	画像生成（設計案・資料・実験可視化）	MMDiTでテキスト理解/スペリング改善を主張。8Bモデルが24GB VRAMに収まり、1024 ² 生成が34秒（50ステップ；初期試験）	2024年3月	https://stability.ai/news/stable-diffusion-3-research-paper ³⁰

組織名	モデル/技術	用途	効果/定量的成果	導入時期	出典URL
<div>Entity"organization","NEDO","japan r&d agency"]</div>	GENIAC（計算資源支援で基盤モデル開発を後押し）	国内基盤モデル開発・R&D基盤整備	第1期で計10件の開発テーマを支援（計算資源提供等）	2024年2月開始	<div>https://www.nedo.go.jp/news/press/AA5_101790.html</div> <div>51</div>
<div>NTT</div> <div>12</div>	tsuzumi（軽量LLM志向）	日本語中心の閉域/省電力運用を想定した基盤モデル	2024年3月の提供開始を明記（軽量・高性能志向）	2024年3月	<div>https://group.ntt/en/magazine/blog/tsuzumi/</div> <div>63</div>
<div>Fujitsu</div> <div>13</div>	Takane（企業向け日本語LLM、RAG等と統合）	セキユア環境での日本語業務/R&D支援	JGLUEでの高性能、RAGやモニタリング技術をセットで提供と説明	2024年9月	<div>https://www.fujitsu.com/global/about/resources/news/press-releases/2024/0930-01.html</div> <div>54</div>

倫理・安全性・規制・コスト

R&Dで生成AIを使う際、倫理・安全・規制は「コンプライアンス部門の課題」ではなく、**研究のスピードと信用を左右する設計変数**になった。理由は、(1)研究成果が外部公開される、(2)データが機微（医療、未公開材料、特許前）である、(3)自動化が進むほど事故時の影響が大きい、からである。 64

安全性の観点では、モデル提供側が評価枠組み（preparedness、リスクスコアカード等）を前面に出す傾向が強い。GPT-4oの発表では、サイバー/CBRN/説得/自律性などの評価でMediumを超えないと述べ、70名超の外部専門家によるレッドチーミングを記載している。²¹ R&D現場では、これを“そのまま信じる”のではなく、**自社タスク（実験計画、合成、設計、解析）に合わせたリスク評価**（誤生成、危険手順、データ漏洩、論文化時の透明性）を別途設計する必要がある。⁶⁵

制度面では、日本では経産省・総務省が「AI事業者ガイドライン（第1.0版）」を公表し、既存ガイドラインを統合・アップデートしたと説明する。⁶⁶ 国際的には、G7広島AIプロセスが“高度なAIシステム（基盤モデ

ル・生成AI含む）”に関する国際指針を提示している。⁶⁷ EUではAI Actにより、GPAIモデル提供者に対し技術文書、利用説明、（一定の）データ要約、著作権対応などの義務が段階適用される枠組みが整理され、GPAI義務の適用開始日が明示されている。⁶⁸ 米国側の実務枠としては、NIST AI RMF（2023）に加え、GenAIプロファイル（2024）が公開され、生成AI特有のリスク管理をAIライフサイクル全体で扱う考え方が示されている。⁶⁹

コストとインフラ要件は、2023年以降「学習コスト」だけでなく「推論コスト」「周辺システム（RAG/評価/ログ/権限制御）コスト」へ重心が移った。モデル自体の効率化（MoEや量子化、軽量ファインチューニング）は進む一方で、長文脈・マルチモーダル・エージェント化が進むほど推論は重くなりがちである。⁷⁰ 実務上の要諦は、「**推論を減らす設計（キャッシュ、検索、分割、バッチ）**」と「**推論を増やす価値がある場面（高コスト実験の前工程、設計レビュー、重大バグ修正）**」の切り分けである。長大ログを毎回全投入するのではなく、ベクトル検索と要約で“必要部分だけ読む”構造にすることが、コストと品質の両方に効く。⁷¹

また、オンプレ/閉域志向の背景には、コストだけでなくデータ主権・供給制約がある。国内ではGENIACが計算資源支援の重要性を明確にし、企業側もセキュア環境向けLLM（Takane等）を打ち出している。⁷²

今後の展望と推奨

2026年初頭時点での展望は、「モデル性能の上積み」よりも、“**研究の実行系**”への統合が主戦場になる可能性が高い。実際、Gemini 2.0が“agentic era”を掲げてツール利用・マルチモーダル入出力を強調し、Claude 3.7が推論モードを統合し“thinking budget”をAPI制御可能にし、GPT-5.3-Codexが長時間タスク・ツール利用を前提に語られている点は、各社の方向性が揃っていることを示す。⁷³

科学・工学領域では、拡散モデルが材料・生体分子で成果を出し始め、生成→検証→再学習のループが高速化する。MatterGenのように生成物の合成・測定まで含めた“閉ループ”の実証が増えれば、R&Dは「探索空間設計」と「検証装置設計」が競争力の中心になる。⁴⁸ Coscientistのようなツール利用型エージェントは、その閉ループを動かす“言語インターフェース”として機能しうる。³⁸

実務推奨を、モデル選定より上位の設計原則としてまとめる。

第一に、研究開発では「**生成の前に、検証を設計する**」。具体的には、(a)シミュレーション/テスト/実験自動化API、(b)ゴール判定指標、(c)失敗時のフォールバック（人手レビューや安全停止）を先に決める。SWE-benchが示すように、現実課題ほど“正しさの検証”が難しくなるため、検証がない生成は負債になる。⁴¹

第二に、**RAG＋監査ログを標準装備**とする。長文脈モデルが強くても、根拠参照・再現性・後日の監査のために、参照した文献・仕様・実験ログのトレーサビリティを残すべきである（AI事業者ガイドラインやNISTプロファイルの方向性とも整合する）。⁷⁴

第三に、**軽量適応（QLoRA/DPO等）を“最後の手段”ではなく“運用技能”として内製化**する。業務・実験・安全要件は組織ごとに違うため、プロンプトだけで吸収できない部分を、小規模データで挙動制御できる体制が長期的に効く。⁷⁵

第四に、規制対応は後付けにせず、**技術文書化・リスク評価・著作権/データ権利対応を開発プロセスに埋め込む**。EU AI ActのGPAI義務や日本のガイドラインは、ドキュメント整備を要件として明確化しており、R&D成果の社会実装速度を左右する。⁷⁶

主要ソース

日本語一次情報・公的資料（優先）として、経産省・総務省のAI事業者ガイドライン公表、G7広島AIプロセス文書、NEDOのGENIAC、文科省白書のAI for Science記述を基礎にした。⁷⁷

英語一次情報として、主要モデルの公式発表（OpenAI、Anthropic、Google DeepMind、Stability AI）と、Nature/arXiv等の原著論文（Coscientist、AlphaFold 3、MatterGen、Gemini 1.5、QLoRA、DPO、SWE-bench、MMMU）を中核に置いた。⁷⁸

¹ ¹⁵ ²¹ ²² ⁵⁸ ⁷⁸ Hello GPT-4o | OpenAI

<https://openai.com/index/hello-gpt-4o/>

² ²³ [2403.05530] Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context

<https://arxiv.org/abs/2403.05530>

³ ⁴⁸ A generative model for inorganic materials design | Nature

<https://www.nature.com/articles/s41586-025-08628-5>

⁴ ⁶ ¹⁹ ⁶⁴ ⁶⁶ ⁷⁴ ⁷⁷ 「AI事業者ガイドライン（第1.0版）」を取りまとめました

https://www.meti.go.jp/press/2024/04/20240419004/20240419004.html?utm_source=chatgpt.com

⁵ ¹⁶ ²⁰ ³⁷ ³⁸ ⁴³ ⁴⁴ ⁴⁵ ⁵⁷ ⁶⁰ Autonomous chemical research with large language models | Nature

<https://www.nature.com/articles/s41586-023-06792-0>

⁷ ⁸ ¹¹ ²⁷ ²⁹ ⁷⁰ ⁷⁵ QLoRA: Efficient Finetuning of Quantized LLMs

https://arxiv.org/abs/2305.14314?utm_source=chatgpt.com

⁹ ¹⁸ ⁴⁶ Accurate structure prediction of biomolecular interactions with AlphaFold 3 | Nature

<https://www.nature.com/articles/s41586-024-07487-w>

¹⁰ ⁵² 第4章 AIの多様な研究分野での活用が切り拓く新たな科学

https://www.mext.go.jp/b_menu/hakusho/html/hpaa202401/1421221_00006.html?utm_source=chatgpt.com

¹² ⁷¹ run-llama/llama_index: LlamaIndex is the leading ...

https://github.com/run-llama/llama_index?utm_source=chatgpt.com

¹³ ⁵¹ ⁷² 生成AIの開発力強化に向けたプロジェクト「GENIAC」において

https://www.nedo.go.jp/news/press/AA5_101790.html?utm_source=chatgpt.com

¹⁴ ³¹ [2307.01952] SDXL: Improving Latent Diffusion Models for ...

https://arxiv.org/abs/2307.01952?utm_source=chatgpt.com

¹⁷ ⁶³ NTT's Large Language Model "tsuzumi" is Here!

https://group.ntt/en/magazine/blog/tsuzumi/?utm_source=chatgpt.com

²⁴ ⁷³ Google introduces Gemini 2.0: A new AI model for the agentic era

<https://blog.google/innovation-and-ai/models-and-research/google-deepmind/google-gemini-ai-update-december-2024/>

²⁵ Mixtral of experts

https://mistral.ai/news/mixtral-of-experts?utm_source=chatgpt.com

²⁶ Meta and Microsoft Introduce the Next Generation of Llama

https://about.fb.com/news/2023/07/llama-2/?utm_source=chatgpt.com

²⁸ Direct Preference Optimization: Your Language Model is Secretly a Reward Model

https://arxiv.org/abs/2305.18290?utm_source=chatgpt.com

- 30 **Stable Diffusion 3: Research Paper — Stability AI**
<https://stability.ai/news/stable-diffusion-3-research-paper>
- 32 **MMMU: A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI**
https://arxiv.org/abs/2311.16502?utm_source=chatgpt.com
- 33 **SWE-bench: Can Language Models Resolve Real-World ...**
https://arxiv.org/abs/2310.06770?utm_source=chatgpt.com
- 34 **[2211.09110] Holistic Evaluation of Language Models**
https://arxiv.org/abs/2211.09110?utm_source=chatgpt.com
- 35 **vLLM: Easy, Fast, and Cheap LLM Serving with PagedAttention**
https://blog.vllm.ai/2023/06/20/vllm.html?utm_source=chatgpt.com
- 36 **Efficient Memory Management for Large Language Model ...**
https://arxiv.org/pdf/2309.06180?utm_source=chatgpt.com
- 39 40 **CMU-Designed Artificially Intelligent Coscientist Automates Scientific Discovery - News - Carnegie Mellon University**
<https://www.cmu.edu/news/stories/archives/2023/december/cmu-designed-artificially-intelligent-coscientist-automates-scientific-discovery>
- 41 42 **SWE-bench: Can Language Models Resolve Real-world ...**
https://openreview.net/forum?id=VTF8yNQM66&utm_source=chatgpt.com
- 47 **AlphaFold 3 predicts the structure and interactions of all of life's molecules - Isomorphic Labs**
<https://www.isomorphiclabs.com/articles/alphafold-3-predicts-the-structure-and-interactions-of-all-of-lifes-molecules>
- 49 61 **The Impact of AI on Developer Productivity: Evidence from GitHub Copilot**
https://arxiv.org/abs/2302.06590?utm_source=chatgpt.com
- 50 **Introducing GPT-5.3-Codex | OpenAI**
<https://openai.com/index/introducing-gpt-5-3-codex/>
- 53 **NTT's Large Language Model "tsuzumi": A High-performance and ...**
https://www.rd.ntt/e/research/JN202406_26659.html?utm_source=chatgpt.com
- 54 59 **Fujitsu launches "Takane" - A large language model for enterprises ...**
https://www.fujitsu.com/global/about/resources/news/press-releases/2024/0930-01.html?utm_source=chatgpt.com
- 55 **PFN Launches Flagship Japan-Made Large Language Model ...**
https://www.preferred.jp/en/news/pr20241202?utm_source=chatgpt.com
- 56 68 76 **General-purpose AI obligations under the AI Act**
https://digital-strategy.ec.europa.eu/en/factpages/general-purpose-ai-obligations-under-ai-act?utm_source=chatgpt.com
- 62 **The Productivity Effects of Generative AI: Evidence from a Field Experiment with GitHub Copilot · From Novel Chemicals to Opera**
<https://mit-genai.pubpub.org/pub/v5iixkxv>
- 65 **AI Risk Management Framework | NIST**
https://www.nist.gov/itl/ai-risk-management-framework?utm_source=chatgpt.com
- 67 **Hiroshima Process International Guiding Principles for ...**
https://www.mofa.go.jp/files/100573471.pdf?utm_source=chatgpt.com

⁶⁹ Artificial Intelligence Risk Management Framework (AI RMF 1.0)
https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf?utm_source=chatgpt.com