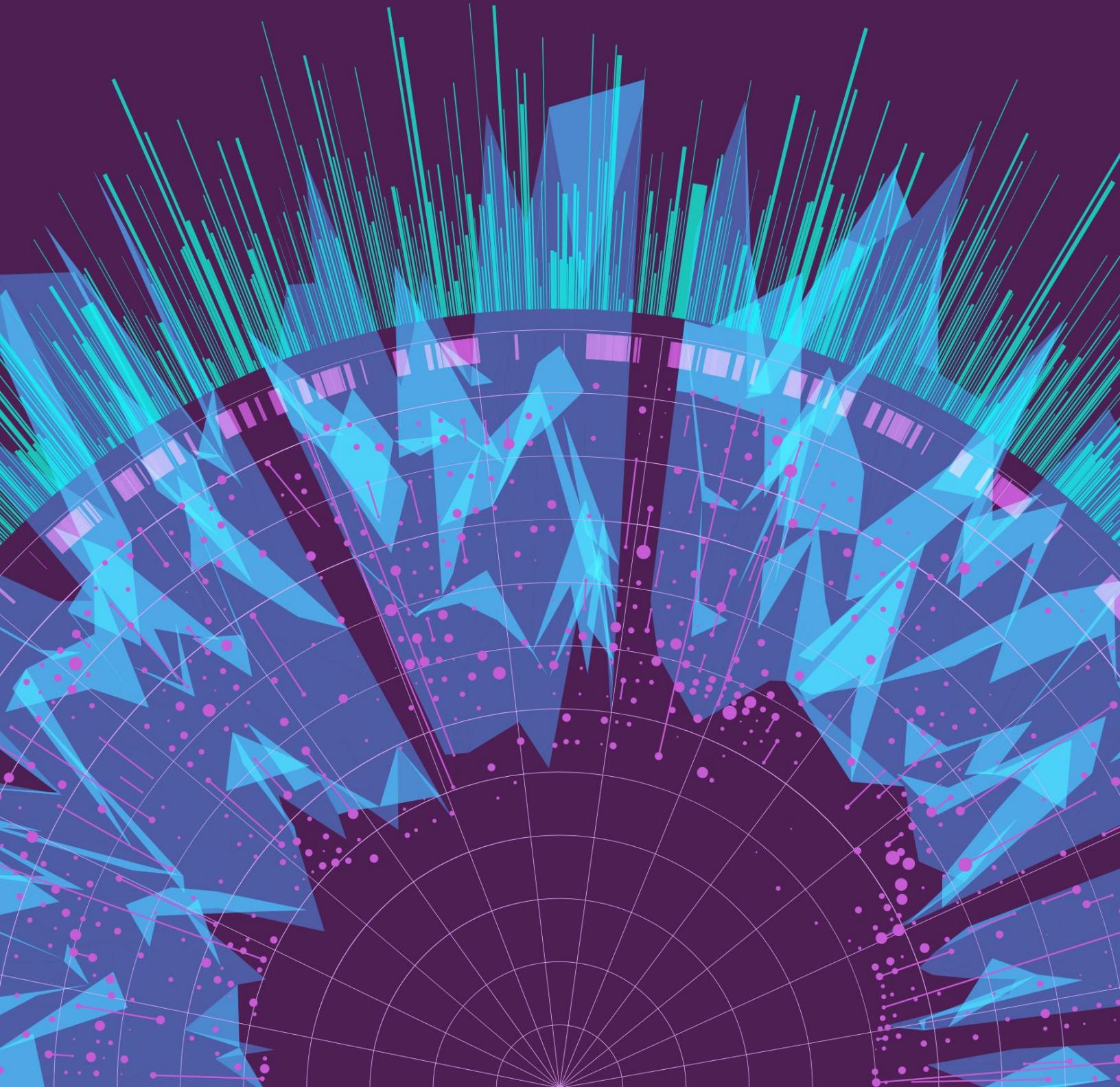




人工知能
インデックスレポート2024

第5章:

科学と 薬





レビュー

概要	298
章のハイライト	299
5.1 注目すべき科学的マイルストーン	300
アルファデブ	300
フレキシキューブ	301
シンボット	303
グラフキャスト	304
GNOME	305
洪水予測	306
5.2 医療におけるAI	307
注目すべき医療システム	307
シンセSR	307
結合プラズモニック赤外線センサー	309
イブスケープ	310
アルファミセンス	312
ヒトパンゲノムリファレンス	313
臨床知識	314
メドQA	314
注目の研究: GPT-4 Medprompt	315
注目の研究: MediTron-70B	317
診断	318
注目の研究: CoDoC	318
注目の研究: CTパンダ	319
その他の診断用途	320
FDA承認のAI関連医療機器	321
管理とケア	323
注目の研究: MedAlign	323

公開データにアクセスする



概要

今年のAI Indexでは、科学と医学の発見におけるAIの役割の拡大を踏まえ、科学と医学におけるAIに関する新たな章を設けています。GraphCastのような高度な気象予測システムや、GNoMEのような改良された物質発見アルゴリズムなど、2023年のAIを活用した注目すべき科学的成果を検証しています。また、医療用AIシステムの性能、SynthSRやImmunoSEIRAのような2023年の重要なAI主導の医療イノベーション、そしてFDAによるAI関連医療機器の承認動向についても検証しています。

章のハイライト

1. AIのおかげで、科学の進歩はさらに加速します。 2022年にはAIが科学的発見を前進させ始めました。しかし、2023年には、さらに重要な科学関連のAIアプリケーションが登場しました。

アルゴリズムによるソートをより効率的にする AlphaDev から、材料発見のプロセスを容易にする GNoME まで。

2. AIは医療を大きく前進させる。 2023年には、パンデミック予測を強化するEVEscapeや、AIによる変異分類を支援するAlphaMissenceなど、いくつかの重要な医療システムが発表された。AIは医療の進歩を促進するためにますます活用されている。

3. 高度な知識を持つ医療AIの登場。ここ数年、AIシステムは、AIの臨床知識を評価するための重要なテストであるMedQAベンチマークにおいて、目覚ましい進歩を見せてています。2023年の傑出したモデルであるGPT-4 Medpromptは、90.2%の精度を達成し、2022年の最高スコアから22.6ポイント上昇しました。このベンチマークが導入された2019年以来、MedQAにおけるAIのパフォーマンスはほぼ3倍に向上しています。

4. FDAはAI関連医療機器の承認をますます増やしています。 2022年には、FDAは139件のAI関連医療機器を承認しました。これは2021年と比較して12.1%増加した数です。2012年以降、FDAが承認したAI関連医療機器の数は45倍以上に増加しています。AIは、実世界の医療現場でますます活用されています。



このセクションでは、AI インデックス運営委員会が選んだ 2023 年の重要な AI 関連の科学的ブレークスルーを紹介します。

5.1 注目すべき科学的マイルストーン

アルファデブ

AlphaDevが高速ソートアルゴリズムを発見

AlphaDevは、計算アルゴリズムの強化分野における学者やエンジニアによる数十年にわたる研究成果を改良した、新しいAI強化学習システムです。AlphaDevは、既存の人間のベンチマークよりも少ない命令でアルゴリズムを開発しました。

ソート3、ソート4、ソート5といった短いシーケンスに対する基本的なソートアルゴリズム（図5.1.1）です。AlphaDevによって発見された新しいアルゴリズムの一部は、LLVM標準C++ソートライブラリに組み込まれています。これは、ライブラリのこの部分に対する10年以上ぶりの更新であり、強化学習を用いて設計された最初の追加機能です。

アルゴリズムの長さを最適化する際のAlphaDevと人間のベンチマーク

出典：マンコウイツ他、2023年 図表：2024年AI指数レポート

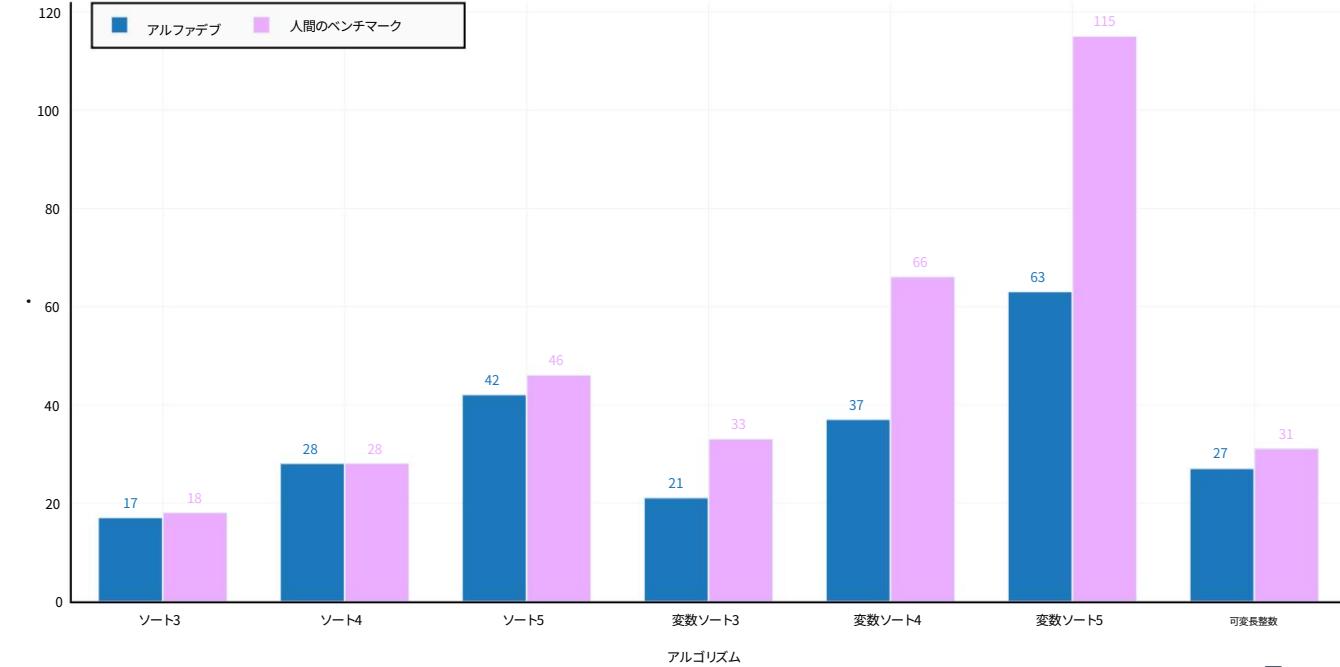


図5.1.1



フレキシキューブ

FlexiCubesによる3Dメッシュ最適化

3Dメッシュ生成はコンピュータグラフィックスにおいて極めて重要であり、頂点、エッジ、面からなるメッシュを作成して3Dオブジェクトを定義します。これは、ビデオゲーム、アニメーション、医用画像、科学的な視覚化において重要な役割を果たします。

従来の等値面抽出アルゴリズムは、解像度の制限、構造の硬直性、数値的不安定性といった問題を抱えることが多く、それが結果的に

品質。FlexiCubesは、勾配ベースの最適化と適応型パラメータにAIを採用することで、これらの制限の一部に対処しています（図5.1.2）。この手法により、正確かつ局所的なメッシュ調整が可能になります。メッシュ再構築に微分可能な等値面を利用する他の主要な手法と比較して、FlexiCubesは、基礎となるグラウンドトゥルースに非常に近いメッシュ抽出を実現します（図5.1.3）。

FlexiCubes表面再構築のサンプル

出典: Nvidia, 2023年

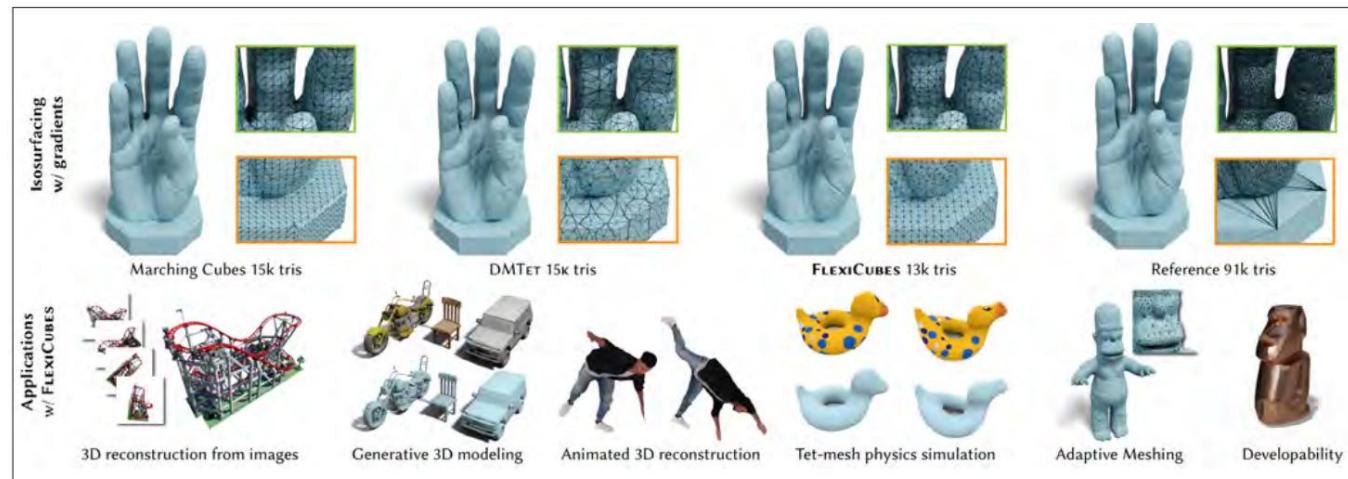


図5.1.2



3Dメッシュ再構築に関する定量的な結果を選択する

出典: Shen et al., 2023 | 図: 2024年AIインデックスレポート

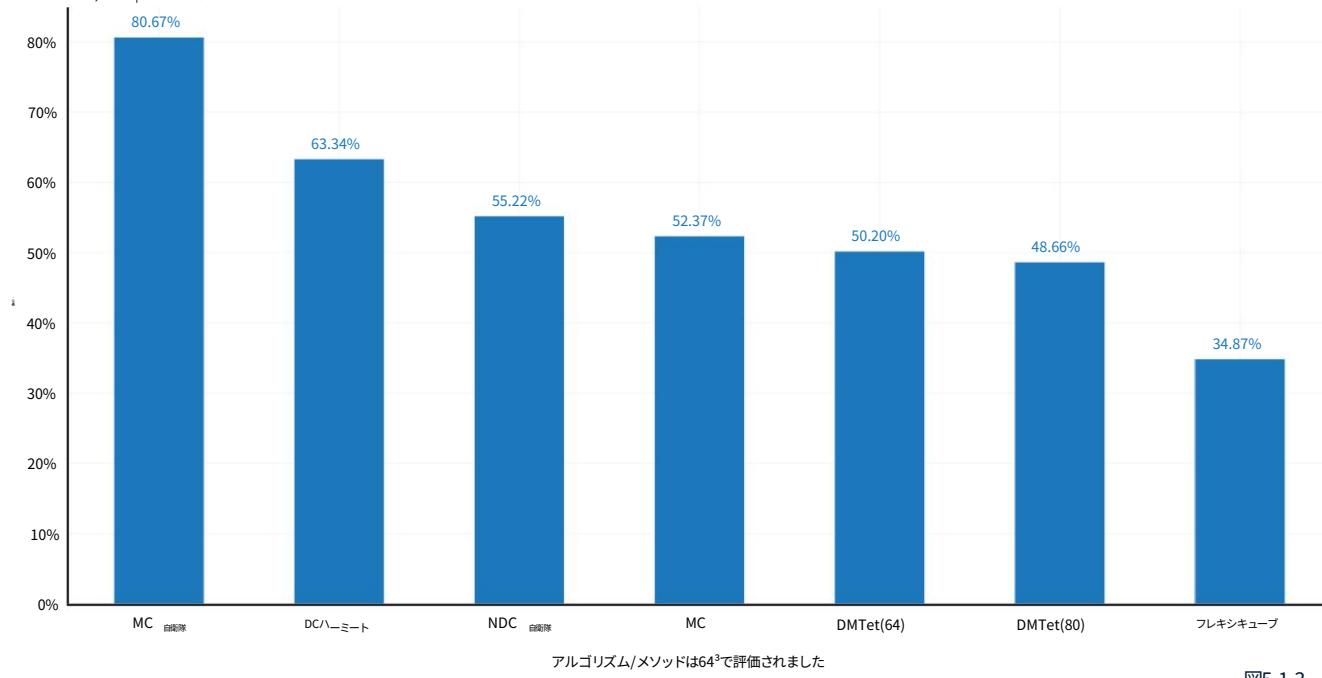


図5.1.3



シンボット

AI駆動型ロボット化学者

有機分子の合成

Synbotは、化学合成計画のためのAIソフトウェア

層、コマンドを翻訳するロボットソフトウェア

層、実験を行う物理ロボット層からなる多層シ

ステムを採用しています。閉ループフィードバックは、

AIと

ロボットシステムにより、Synbotは既存の基準と同等

かそれ以上の収率を持つ合成レシピを開発する

ことが可能になった（図5.1.4）。M1 [4-(2,3-

ジメトキシフェニル)-

1H-ピロロ[2,3-b]ピリジン]を合成するた

めに、シンボットは、変換収率を100万トン上回

る複数の合成式を開発しました。

シンボットデザイン

出典: Ha et al., 2023

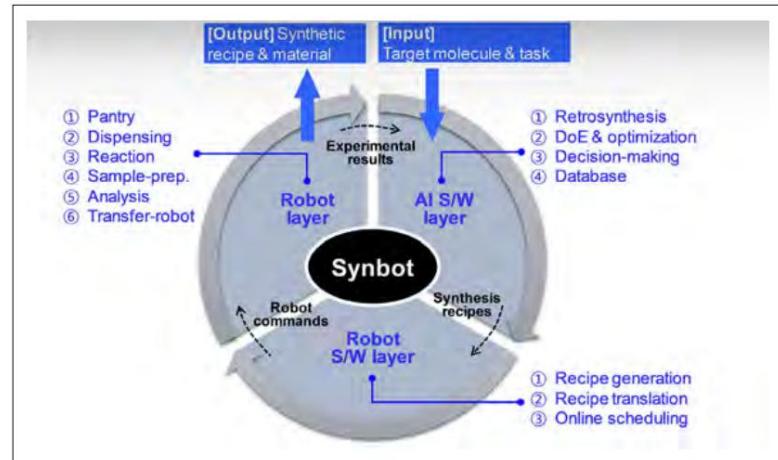


図5.1.4

80%台半ばの基準範囲をクリアし、大幅に短い時間で合成を完了しました（図

5.1.5）。Synbotによる有機合成の自動化は、医薬品や材料科学などの分野

におけるAIの可能性を浮き彫りにしています。

M1自律最適化実験の反応速度論、Synbotと参照

出典: Ha et al., 2023年 図表 - 2024年AIインデックスレポート

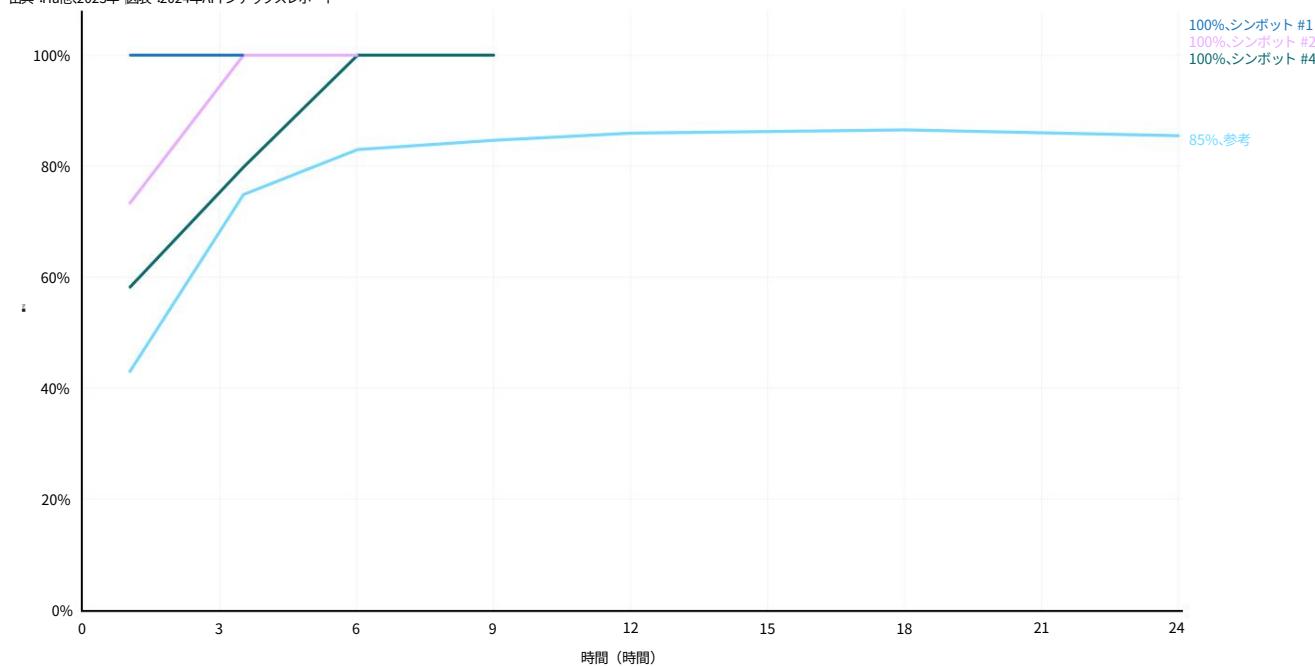


図5.1.5



グラフキャスト

GraphCastでより正確な世界の天気予報

GraphCastは、1分以内に10日間の高精度な天気予報を提供する新しい天気予報システムです（図5.1.6）。GraphCastはグラフニューラルネットワークと機械学習を活用し、膨大なデータセットを処理して気温、風速、大気の状態などを予測します。

など、他にも多くの機能があります。図5.1.7は、GraphCastと、現在業界最先端の気象シミュレーションシステムである高解像度予報（HRES）のパフォーマンスを比較したものです。GraphCastは二乗平均平方根誤差（RMS）が低く、観測された気象パターンとより近い予測が可能です。GraphCastは、気象パターンの解読、異常気象への備えの強化、そして地球規模の気候研究への貢献において、貴重なツールとなり得ます。

GraphCast天気予報

出典 :DeepMind, 2023年

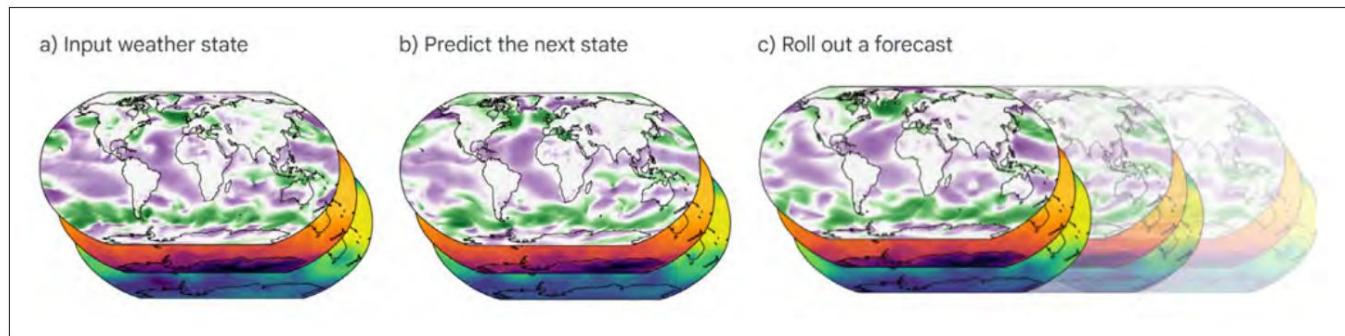


図5.1.6

10日間のz500予測スキル :GraphCast vs. HRES

出典 :Lam et al., 2023 | 図 :2024年AIインデックスレポート

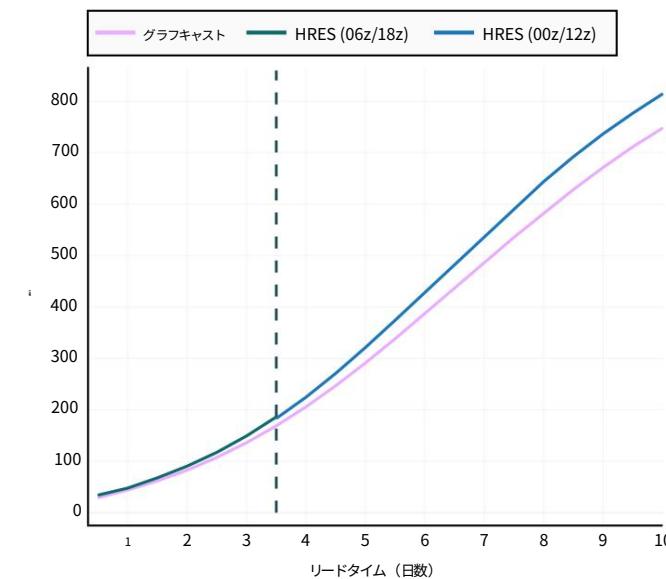


図5.1.7



GNOME

GNOMEで新しい素材を発見する

ロボット工学や半導体製造など、様々な科学分野の進歩には、新機能材料の探索が鍵となります。しかし、この発見プロセスは一般的に費用がかかり、時間がかかります。

Googleの研究者による最近の進歩により、AIモデルの一種であるグラフネットワークは、大規模なデータセットで学習させることで、このプロセスを高速化できることが実証されました。彼らのモデルであるGNOMEは、材料発見における主要な手法である Materials Projectをはるかに上回り、安定結晶をはるかに多く特定しました（図5.1.8）。GNOMEは220万点もの新しい結晶構造を明らかにしており、その多くは人間の研究者が見落としていたものです（図5.1.9および図5.1.10）。GNOMEのようなAI主導のプロジェクトの成功は、科学のブレークスルーを加速させる上で、データとスケーリングの力がどれほど強力であるかを浮き彫りにしています。

サンプル材料構造

出典：マーチャント他, 2023

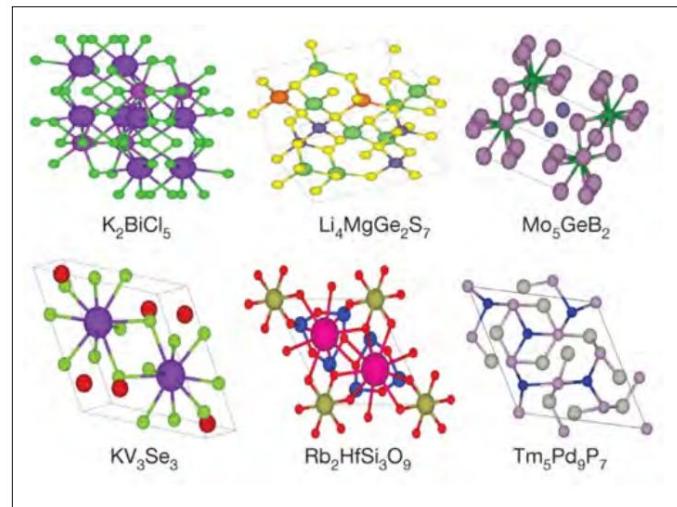
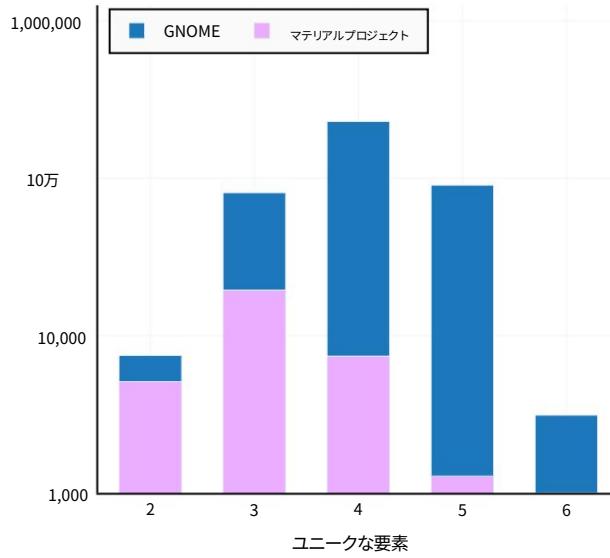


図5.1.8

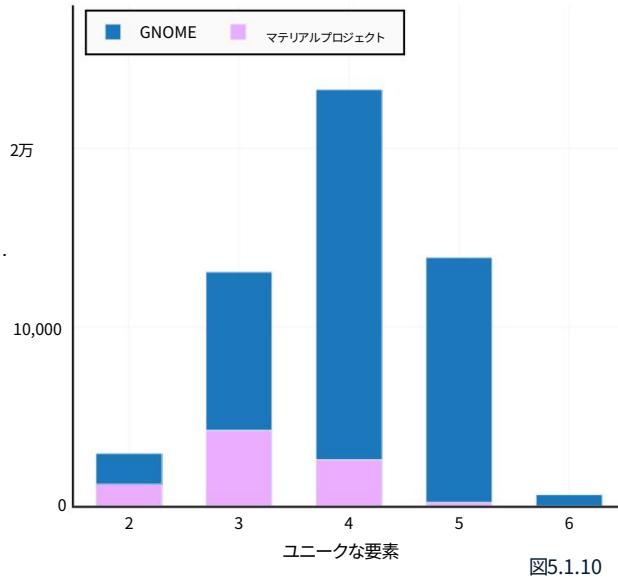
GNOME vs. Materials Project: 安定した結晶数

出典: Merchant et al., 2023 | チャート: 2024 年の AI インデックス レポート



GNOME vs. Materials Project: 異なるプロトタイプ

出典: Merchant et al., 2023 | チャート: 2024 年の AI インデックス レポート





洪水予測

より正確で信頼性の高い洪水予測のためのAI

2023年に導入された新たな研究により、

大規模洪水の予測において大きな進歩が見られる。最も一般的な自然災害の一つである洪水は、予防・軽減のためのインフラが整備されていない発展途上国において特に壊滅的な被害をもたらす。そのため、より正確な予測手法の開発が求められている。

Googleの研究チームはAIを活用し、未観測流域にも適用可能な高精度な水文シミュレーションモデルを開発しました。¹この革新的な手法は、特定の極端な洪水事象を最大5日先まで予測することができ、GloFASなどの最新モデルと同等かそれ以上の精度を誇ります。このAIモデルは、様々な再現期間の事象において優れた精度（予測精度）と再現率（関連するすべての事象を正しく識別する能力）を示し、主要な現代手法を凌駕しています（図5.1.11）。²このモデルはオープンソースであり、既に80カ国以上で洪水事象の予測に使用されています。

これらの出来事をさらに事前に予測できる

大きなプラスの影響をもたらします。

再現期間全体にわたる AI モデルと GloFAS の予測

出典 :Nearing他、2023年 図表 :2024年AIインデックスレポート

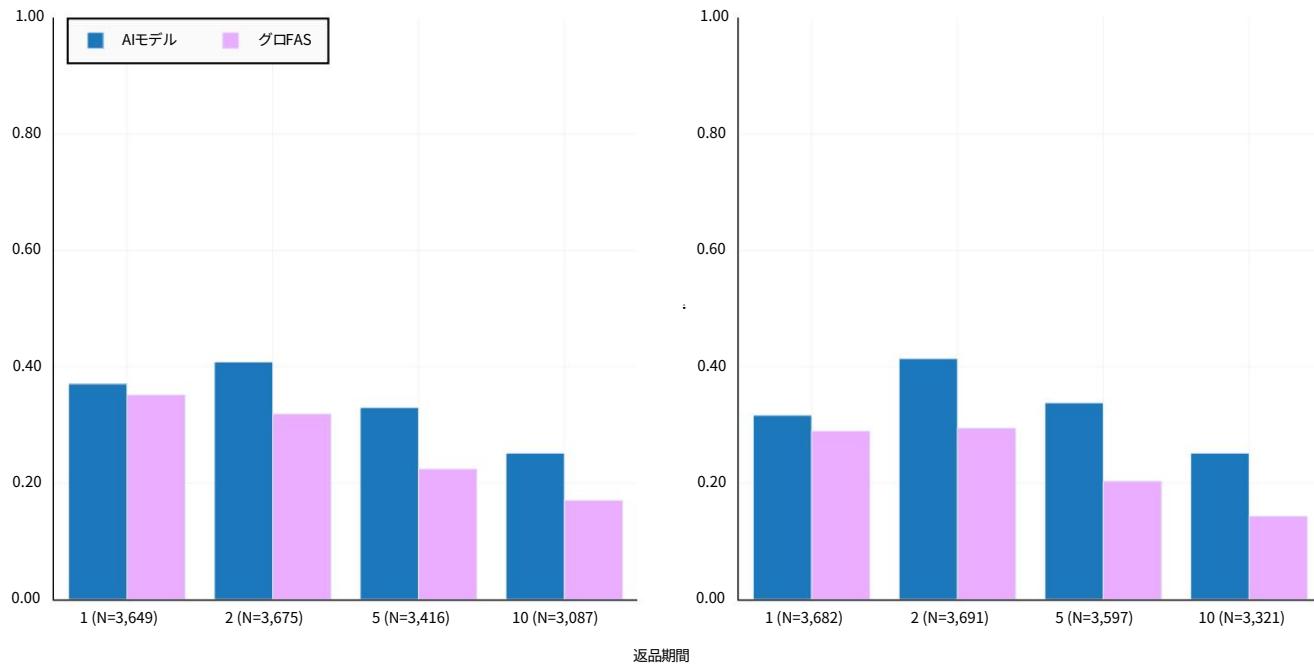


図5.1.11

1 未計測流域とは、水文流量をモデル化するための流量データが不十分な流域のことです。

2 再現期間（再現間隔）とは、特定の水文学的事象が特定の期間内に再発する確率を表す指標です。例えば、100年に一度発生する洪水は、ある特定の年に同じ事象と同等かそれ以上の確率で発生することを意味します。



AIモデルは医療分野においてますます価値を高めており、ポリープの検出から医師の診断支援まで、様々な用途で活用されています。AIの性能向上に伴い、医療現場への影響をモニタリングすることがますます重要になっています。本セクションでは、2023年に導入される重要なAI関連医療システム、臨床AIに関する知識の現状、そして病院運営の強化を目的とした新たなAI診断ツールやモデルの開発について取り上げます。

5.2 医療におけるAI

注目すべき医療システム

このセクションでは、AI インデックス運営委員会が選んだ 2023 年の重要な AI 関連の医療の進歩を紹介します。

シンセSR

脳スキャンを高度な分析に応用

SynthSR は、臨床脳スキャンを高解像度の T1 強調画像に変換する AI ツールです（図 5.2.1）。この進歩は、これまで高度な研究で多くのスキャンの使用を制限していたスキャン品質のばらつきの問題を解決します。これらのスキャンを、高コントラストで脳構造の鮮明な描写で知られる T1 強調画像に変換することにより、SynthSR は詳細な 3D 脳レンダリングの作成を容易にします。SynthSR を使用した実験では、スキャン レベルと被験者レベルの両方で観測されたボリューム間に堅牢な相関関係があることが示され、SynthSR が高解像度の T1 スキャンによって生成される画像に非常によく似た画像を生成することが示唆されています。図 5.2.2 は、SynthSR スキャンが選択された脳領域全体でグラウンド トゥルース観察とどの程度対応しているかを示しています。SynthID は脳構造の視覚化と分析を大幅に改善し、神経科学研究と臨床診断を容易にします。

SynthSR世代
出典：イグレシアス他、2023

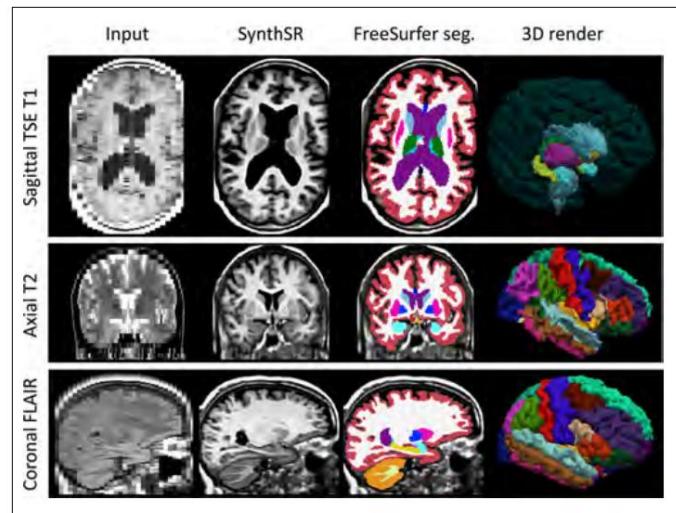


図5.2.1



特定の脳領域における SynthSR とグラウンドトゥルースボリュームの相関

出典：イグレシアス他、2023年 図表：2024年AIインデックスレポート

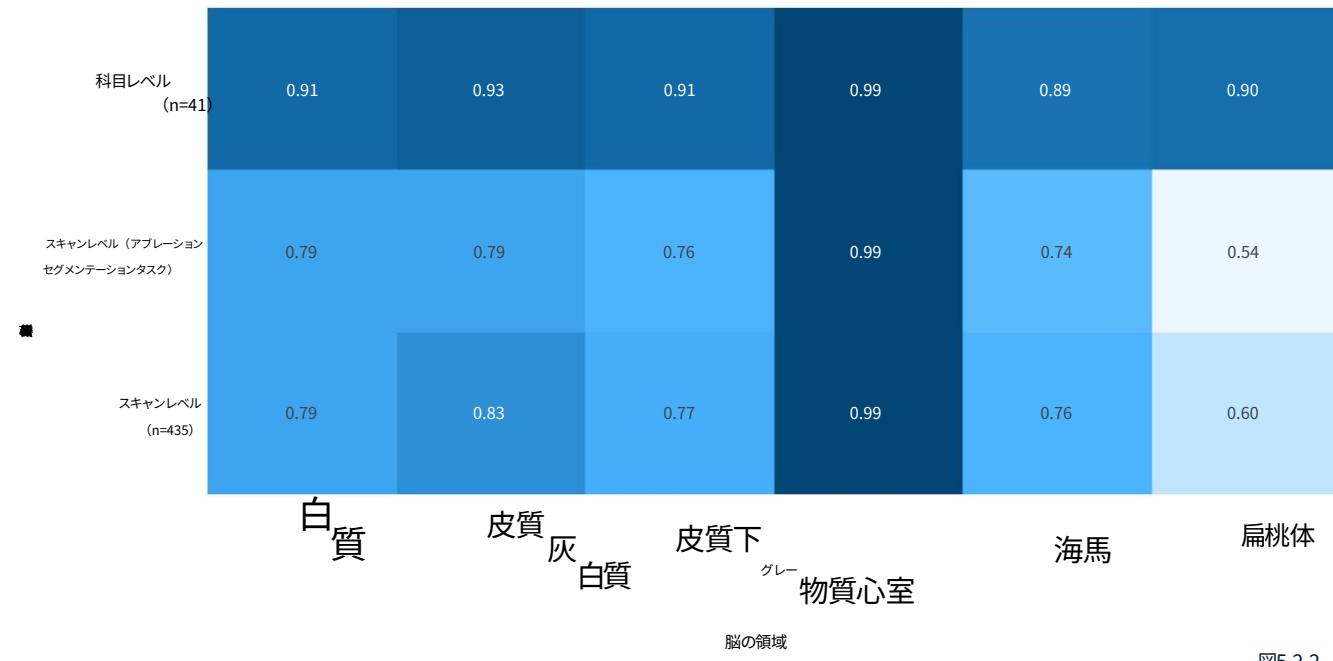


図5.2.2



結合プラズモニック赤外線センサー

神経変性疾患検出のための結合プラズモニック赤外線センサー パーキンソン病やアルツハイマー病などの神経変性疾患の診断は、バイオマーカーの迅速かつ正確な識別にかかっています。質量分析やELISAなどの従来の方法は、タンパク質レベルの定量化に焦点を当てることができるという点で有用ですが、構造状態の変化を識別することはできません。今年、研究者らは、表面増強赤外線吸収 (SEIRA) 分光法を使用するAI結合プラズモニック赤外線センサーと免疫測定技術 (ImmunoSEIRA、図5.2.3) を組み合わせた神経変性疾患診断の新しい方法を発見しました。実際の線維の割合とAIシステムによる予測を比較したテストでは、予測の精度は実際に報告された割合と非常によく一致することがわかりました (図5.2.4)。

ImmunoSEIRAの検出原理とセットアップ

出典:カヴァンガル他,2023

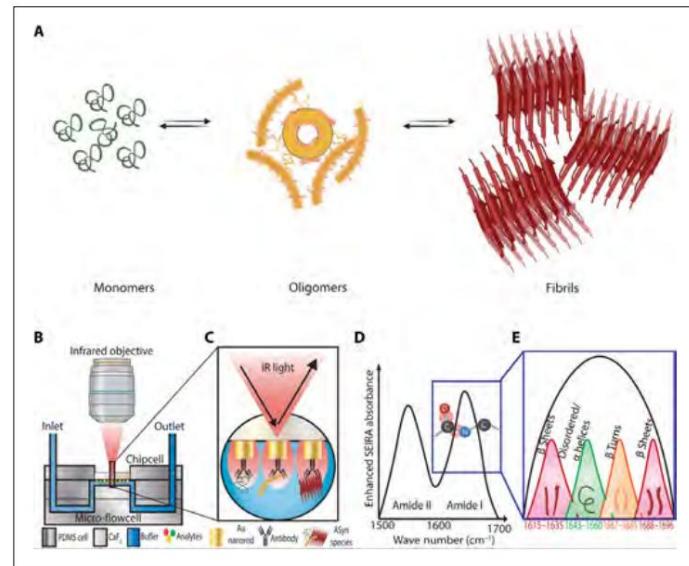


図5.2.3

ディープニューラルネットワークによるテストサンプルの予測ブリルス率と実際のブリルス率

出典:Kavungal et al., 2023 | 図:2024年AIインデックスレポート

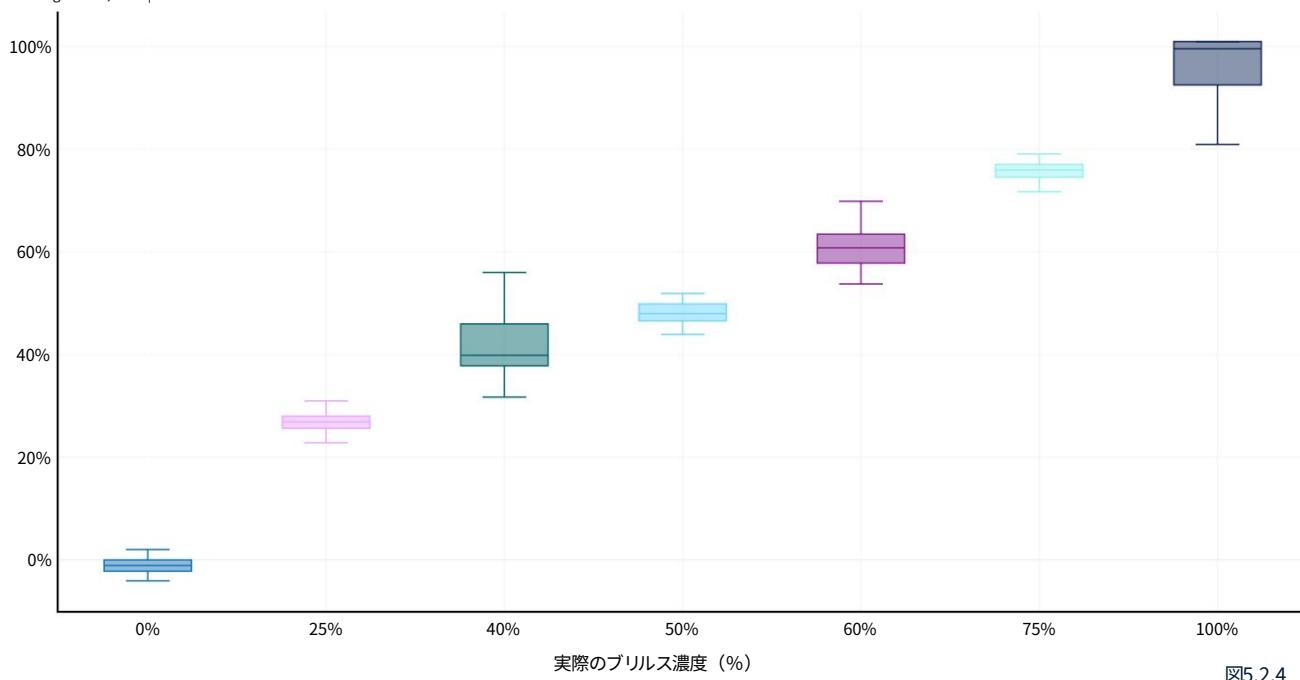


図5.2.4



イプスケープ

パンデミック対策のためのウイルス進化予測ウイルスの変異予測は、ワクチン設計とパンデミックの最小化に不可欠です。リアルタイムのウイルス株と抗体データに依存する従来の手法は、パンデミックの初期段階ではデータ不足のために課題に直面します。EVEscapeは、過去の配列と生物物理学的情報に基づいて訓練された新しいAIディープラーニングモデルであり、ウイルスの進化を予測します。

ウイルスの（図5.2.5）。EVEscapeは、現在の株データと独立してウイルスの逃避を評価し、観測されたSARS-CoV-2変異の50.0%を予測しました。これは、46.2%と32.3%を予測した従来の研究室での研究、および変異の24%しか予測できなかった以前のモデルを上回っています（図5.2.6）。

このパフォーマンスは、将来のパンデミックへの備えと対応の取り組みを強化するための貴重な資産としての EVEscape の可能性を浮き彫りにしています。

EVEscapeのデザイン

出典: Thadani et al., 2023

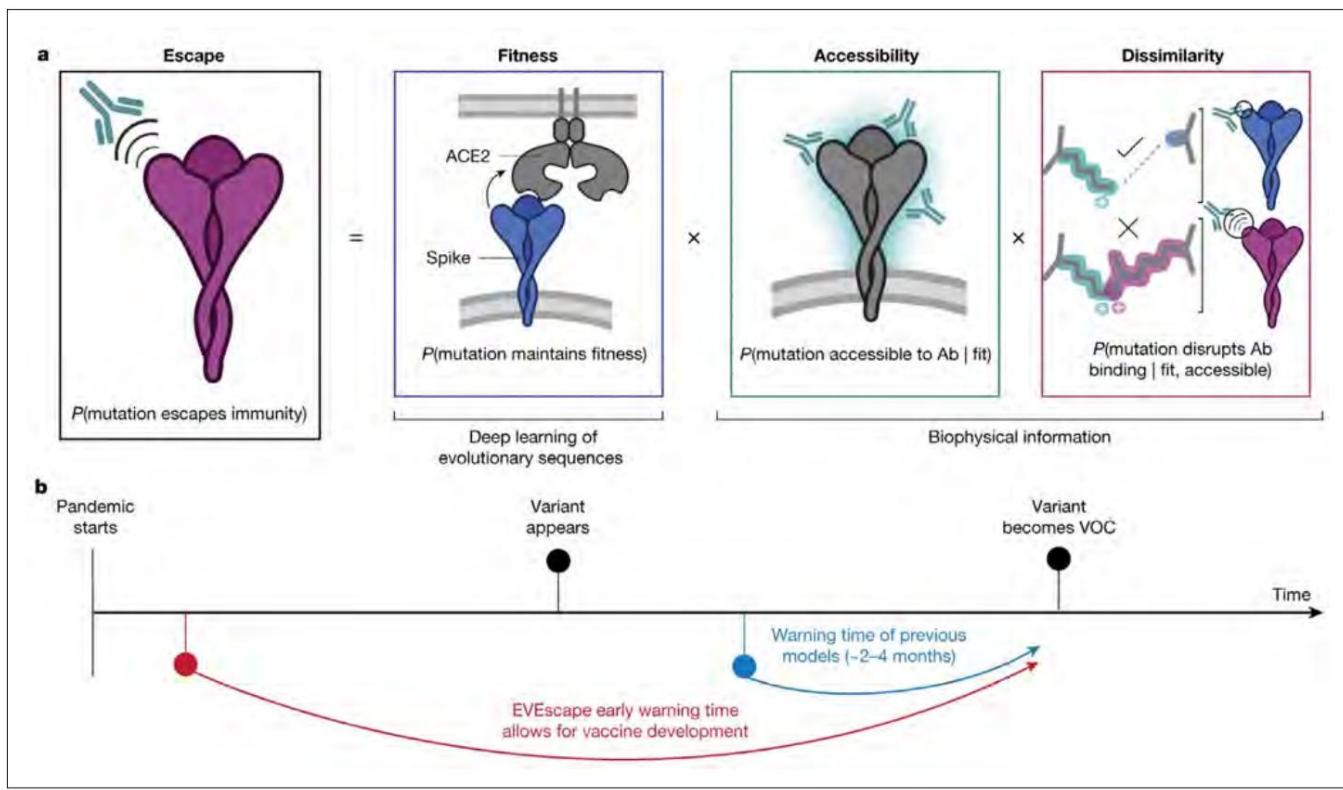


図5.2.5

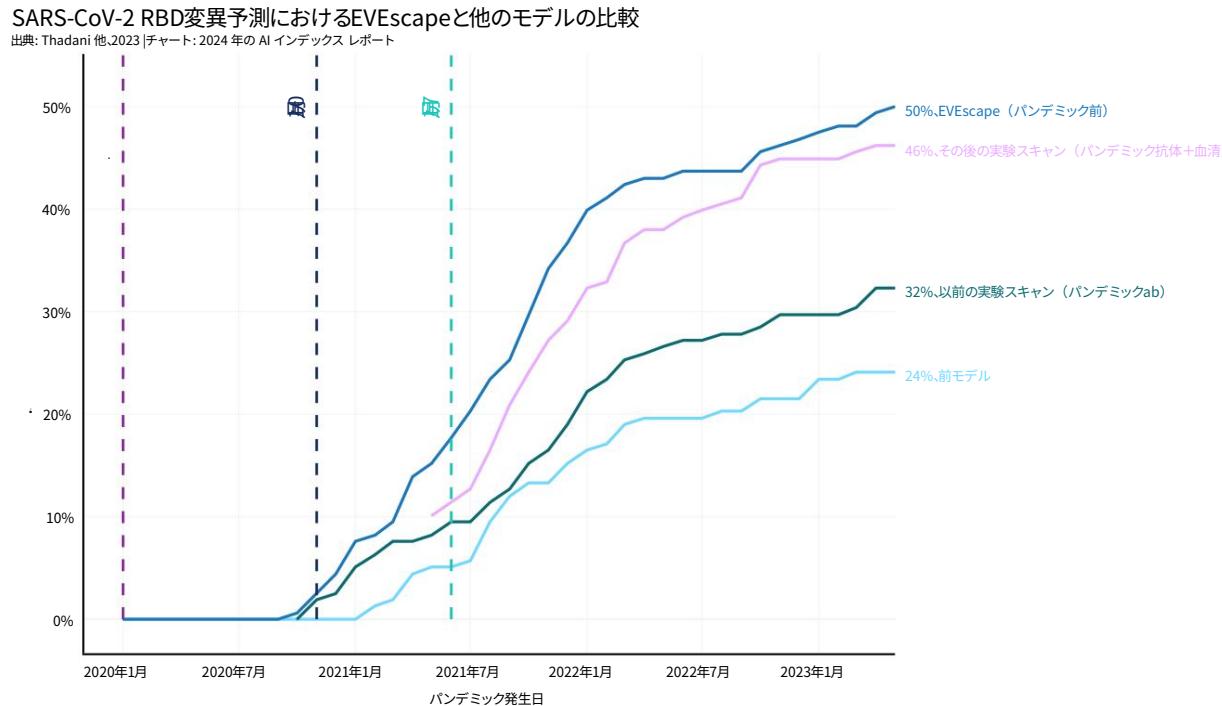


図5.2.6



アルファミセンス

AI変異のより正確な分類

科学者たちは、どの遺伝子変異が疾患を引き起こすのかをまだ完全には解明していません。遺伝子変異は数百万種類に及ぶ可能性があるため、ある変異が良性か病原性かを判断するには、膨大な労力を要する実験が必要です。

2023年、Google DeepMindの研究者は、7100万の病原性を予測した新しいAIモデルAlphaMissenseを発表しました。

ミスセンス変異。ミスセンス変異は遺伝子変異はヒトタンパク質の機能に影響を与える（図5.2.7）、がんを含む様々な疾患につながる可能性があります。AlphaMissenseは7100万のミスセンス変異のうち89%を分類し、そのうち57%は良性、32%は病原性が高いと判定しました。残りは不確実と分類されました（図5.2.8）。一方、人間のアノテーターは、

ヘモグロビンサブユニットβ（HBB）

出典: Google DeepMind. 2023年

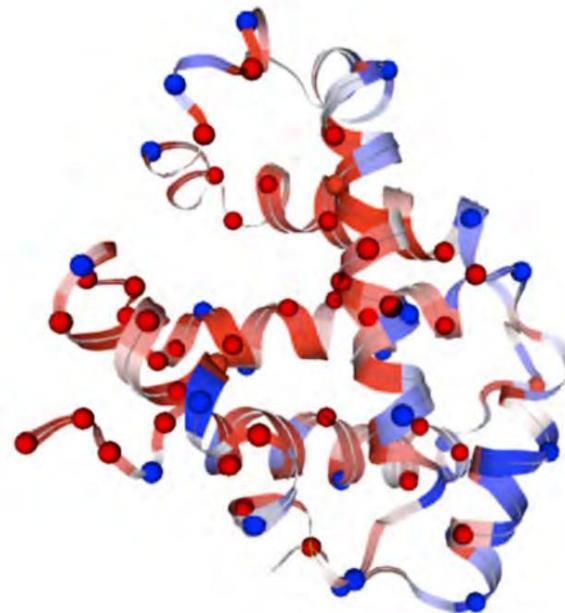


図5.2.7

すべてのミスセンス変異の0.1%の性質。

AlphaMissenseの予測

出典: Google DeepMind. 2023年 | 図: 2024年AIインデックスレポート

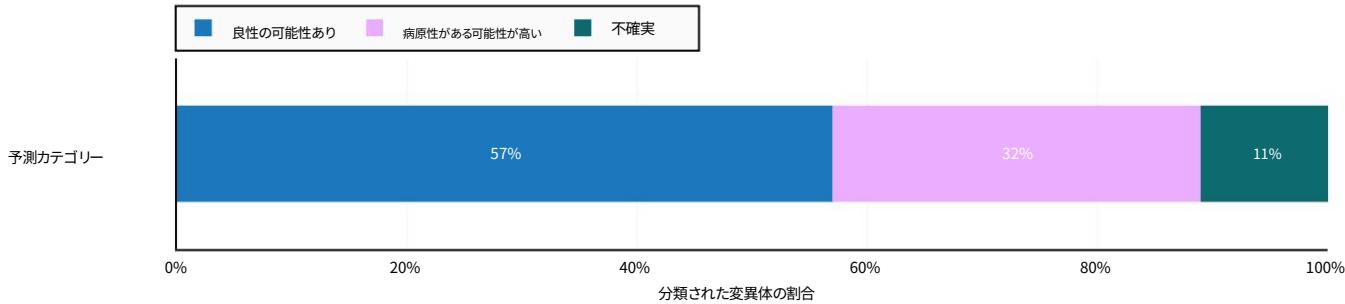


図5.2.8



ヒトゲノムリファレンス

AIを使ってヒトゲノムをマッピングする

ヒトゲノムは、ヒトの分子レベルの指示書です。最初のヒトゲノムのドラフト版は2000年に公開され、2022年に更新されました。しかし、この更新版はやや不完全なものでした。血液型などの様々な遺伝子変異が組み込まれておらず、多様な祖先集団を完全にマッピングしていませんでした。そのため、既存のゲノム基準では、特定の集団における疾患の検出や治療法の発見は困難です。

2023年、60の機関から119人の科学者が参加するヒトパンゲノム研究コンソーシアムは、AIを用いて、より代表性の高い最新のヒトゲノムマップを作成しました（図5.2.9）。研究者らは、図5.2.10に示すように、タンパク質コード遺伝子の99.07%、タンパク質コード転写産物の99.42%、非コード遺伝子の98.16%、非コード転写産物の98.96%をアノテーションするという驚異的な精度を達成しました。

ゲノムのMHC領域のグラフゲノム

出典: Google Research, 2023年

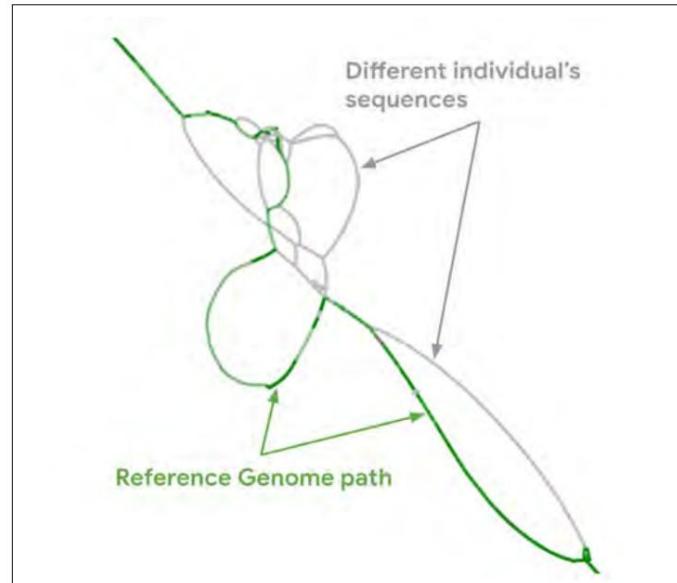


図5.2.9

この最新版のゲノムは、これまで最も包括的かつ遺伝的に多様なヒトゲノムのマッピングを表しています。

Ensembl マッピングパイプラインの結果

出典: Liao et al., 2023 | 図: 2024年AIインデックスレポート

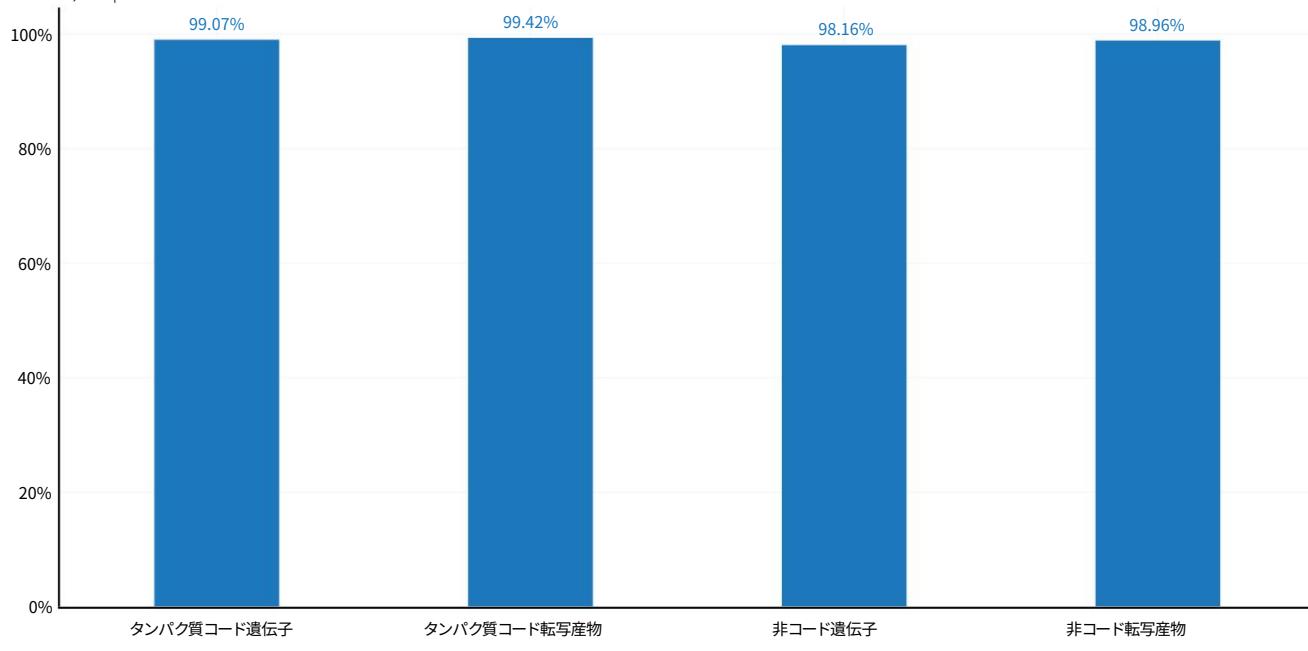


図5.2.10



臨床知識

AI モデルの臨床知識を評価するには、その医学的専門知識、特に臨床現場に適用可能な知識の範囲を判断する必要があります。

メドQA

2020 年に導入された MedQA は、医師の挑戦を促すように設計された 60,000 を超える臨床質問を特徴とする、専門医試験から得られた包括的なデータセットです。

目覚ましい改善が見られ、トップシステムであるGPT-4 Medpromptは90.2%の精度を達成しました。これは、2022年のトップスコアから22.6ポイントの増加です（図5.2.11）。MedQA の設立以来、このベンチマークにおけるAIの能力はほぼ3倍に向上しており、臨床知識を備えたAIシステムの急速な進歩を示しています。

MedQAベンチマークにおけるAIのパフォーマンスは
MedQA: 精度

出典: Papers With Code, 2023年 | 図: 2024年AIインデックスレポート

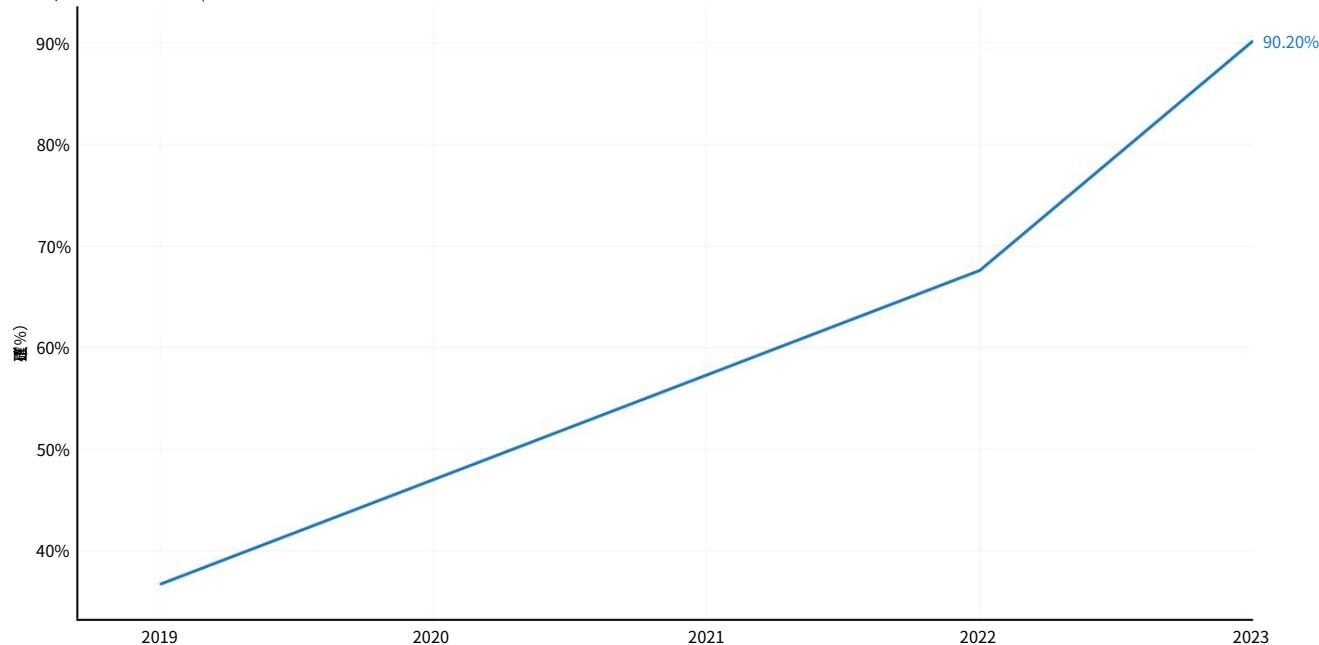


図5.2.11



注目の研究:

GPT-4 メドプロンプト

LLMは優れた一般知識を備えていますが、医療に関する質問に答えるといった専門知識において卓越した能力を発揮するには、大幅な微調整が必要であると一般的に考えられています。微調整には、LLMを特定の分野に特化したデータで訓練することが含まれます。

2023年後半のマイクロソフトの調査

この想定を覆しました。
この研究では、迅速なエンジニアリングを採用し、GPT-4をMultiMedQAベンチマークで優れたパフォーマンスを達成するように導きました。

スイートは、4つの困難な医療ベンチマークのグループです（図5.2.12）。

GPT-4 Medpromptは、いくつかの有名な試験の多肢選択セクションで、2022年のトップモデルであるFlan-PaLM 540Bの成績を上回りました。

医療基準、PubMedQA、MedMCQA、MMLU、それぞれ3.0、21.5、16.2/パーセントポイントを向上しました。これは当時の最先端技術であったMed-PaLM 2の性能を上回りました（図5.2.13）。

さらに、前述のように、GPT-4 MedpromptはMedQAベンチマークで90%の精度を超えた最初のツールです。この画期的な成果は、GPT-4 Medpromptの並外れた、そして臨床的に有用な可能性のある医療機能を強調するだけでなく、モデルを特定の領域に適応させる際に微調整が必ずしも必要ではないことを実証しています。プロンプトエンジニアリングは、有望な代替戦略であることが示されています。

GPT-4とMed-PaLM 2による医学的質問への回答 出典：ノリ他、2023

Question: A 22-year-old male marathon runner presents to the office with the complaint of right-sided rib pain when he runs long distances. Physical examination reveals normal heart and lung findings and an exhalation dysfunction at ribs 4-5 on the right. Which of the following muscles or muscle groups will be most useful in correcting this dysfunction utilizing a direct method?

(A) anterior scalene (B) latissimus dorsi (C) pectoralis minor (D) quadratus lumborum

Hand-crafted CoT from Med PaLM 2

 Expert

Let's solve this step-by-step, referring to authoritative sources as needed. Among the options, only pectoralis minor muscle originates from the outer surfaces of the 3rd to 5th ribs.

GPT-4 generated CoT

 GPT-4

The primary muscle involved in rib exhalation is the internal intercostal muscle. However, this option is not listed among the answer choices. Among the provided options, the pectoralis minor can contribute to rib movement, as it originates from the 3rd-5th ribs and can act to depress the ribs during exhalation when the scapula is fixed. Therefore, it could potentially be useful in correcting an exhalation dysfunction at ribs 4-5.

図5.2.12



注目の研究: GPT-4 Medprompt (続き)

MultiMedQAサブベンチマークにおけるモデルパフォーマンス
出典: Nori et al., 2023 | 図: 2024年AIインデックスレポート

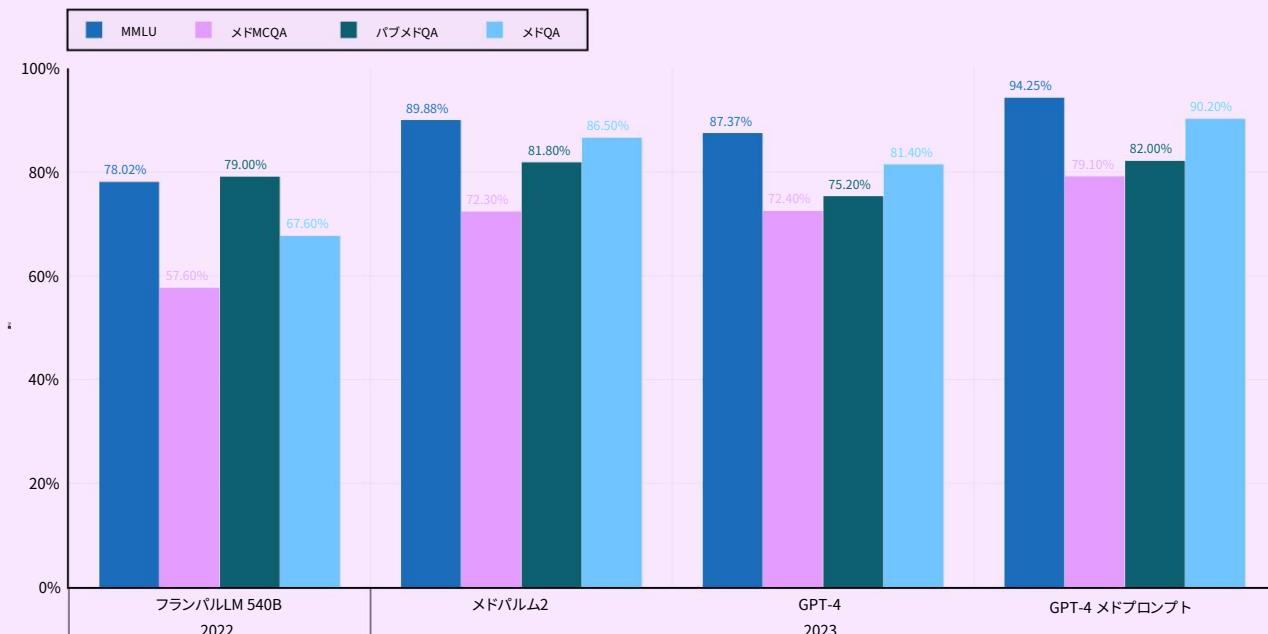


図5.2.13



注目の研究: メディトロン-70B

GPT-4 Medpromptは優れたシステムですが、クローズドソースであるため、その重みは一般公開されていません。2023年には、オープンソースの医療LLMの性能向上を目指す新たな研究も行われました。こうした新たな研究の中でも、MediTron-70Bは特に有望視されています。このモデルは、MedQAベンチマークで70.2%という高い精度を達成しています。これはGPT-4 MedpromptやMed-

PaLM 2（両方ともクローズドモデル）では、2023年の最先端の結果に比べて大幅に改善されており、Llama 2などの他のオープンソースモデルを上回っています（図5.2.14）。MediTron-70BのMedQAスコアは、オープンソースモデルとしてはこれまで最高記録です。医療AIが最大限の可能性を發揮するには、その機能が広く利用可能であることが重要です。この点において、MediTronは心強い前進と言えるでしょう。

MedQAにおける特定のモデルのパフォーマンス

出典: Chen et al., 2023 | 表: 2024年AIインデックスレポート

モデル	発売日	アクセスタイプ	MedQAのスコア
GPT-4 メドプロンプト	2023年11月	閉鎖	90.20%
メドバルム2	2023年4月	閉鎖	86.20%
メディトロン-70B	2023年11月	開ける	70.20%
メドバルム	2022年12月	閉鎖	67.20%
ラマ2	2023年7月	開ける	63.80%

図5.2.14



診断

AI ツールは、たとえば放射線診断やがん検出などの診断目的にも使用できます。

注目の研究:

CoDoC

AI 医用画像システムは堅牢な診断能力を発揮しますが、臨床医が捉えた診断を AI が見逃してしまうケースがあり、その逆もまた同様です。この観察結果は、AI システムと臨床医の診断能力の論理的な統合を示唆しています。2023 年には、研究者らが CoDoC（相補性駆動型臨床ワークフローへの委譲）を発表しました。これは、診断において AI に頼るべきタイミングと従来の臨床手法に従うべきタイミングを判断するために設計されたシステムです。CoDoC は、感度（疾患を持つ個人を正しく識別する能力）と特異度の両方を著しく向上させます。

（それを持たない人を正確に識別する能力）。具体的には、4つの医療データセット全体において、CoDoC の感度は臨床医の感度を平均 4.5 パーセントポイント、スタンドアロン AI モデルを 6.5 パーセントポイント上回っています（図 5.2.15）。特異度に関しては、CoDoC はテストされたデータセット全体で臨床医の感度を平均 2.7 パーセントポイント、スタンドアロン予測モデルを 5.7 パーセントポイント上回っています。さらに、CoDoC は臨床ワークフローを 66% 削減することが示されています。これらの結果は、AI 医療システムを臨床ワークフローに統合することで、診断の精度と効率性を向上させることができる事を示唆しています。

CoDoC とスタンドアロンの予測 AI システムおよび臨床読影者の比較: 感度

出典: Dvijotham 他, 2023 年 図表: 2024 年 AI 指数レポート

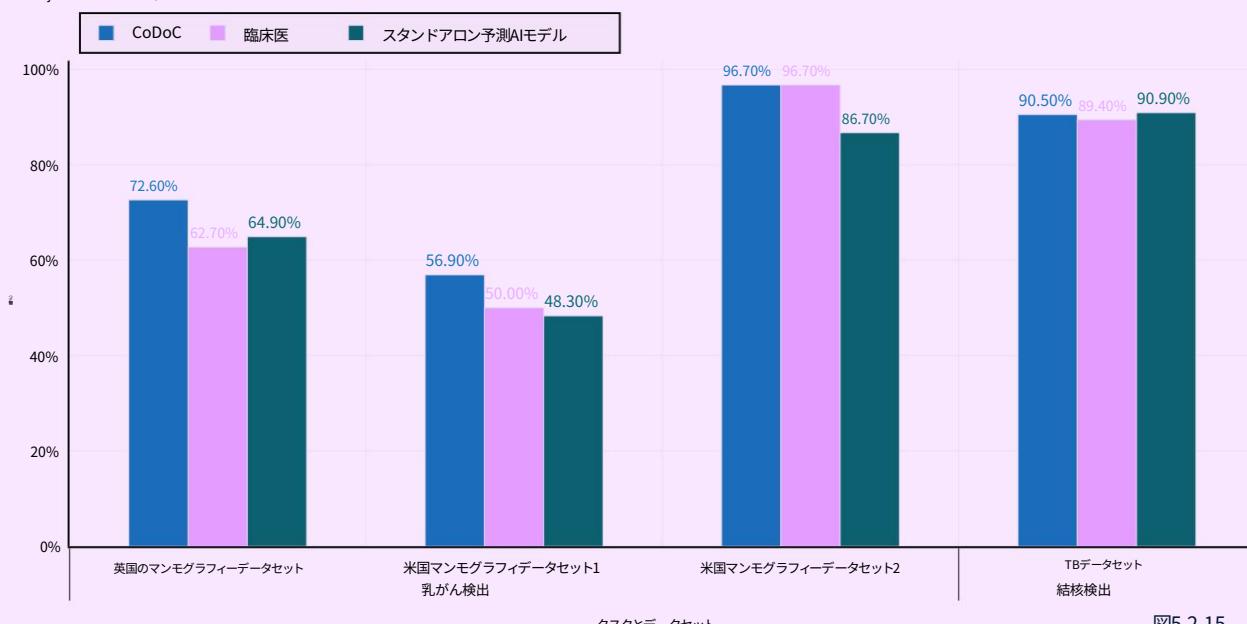


図 5.2.15

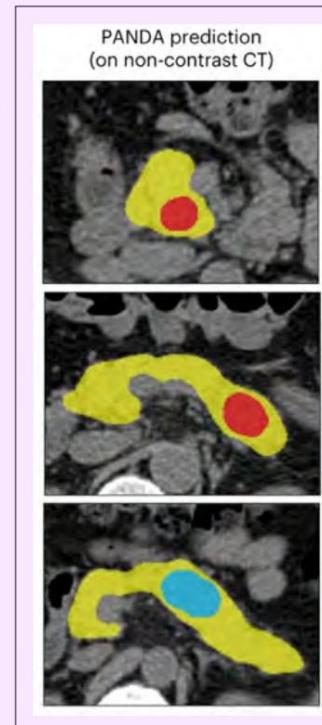


注目の研究: CTパンダ

膵管腺癌（PDAC）は特に致死率の高い癌であり、外科的介入が間に合わない状態で発見されることが多い。

無症状の膵臓癌（PDAC）のスクリーニングは、その有病率の低さと偽陽性のリスクのために困難です。今年、中国の研究チームは、X線画像から膵臓病変を効率的に検出・分類できるAIモデル「PANDA（人工知能による膵臓癌検出）」を開発しました（図5.2.16）。検証試験では、PANDAは平均的な放射線科医の感度を34.1%、特異度を6.3%上回りました（図5.2.17）。

PANDA検出
出典: Cao et al., 2023
図5.2.16



約2万人の患者を対象とした大規模な実臨床試験において、PANDAは感度92.9%、特異度99.9%を達成しました（図5.2.18）。PANDAのようなAI医療ツールは、これまで困難または高額と考えられていた診断困難な疾患の診断において、費用対効果の高い正確な検出を提供することで、大きな進歩を遂げています。

PANDAと平均放射線科医の多施設検証 (患者数6,239人)

出典: Cao et al., 2023 | 図: 2024年AIインデックスレポート

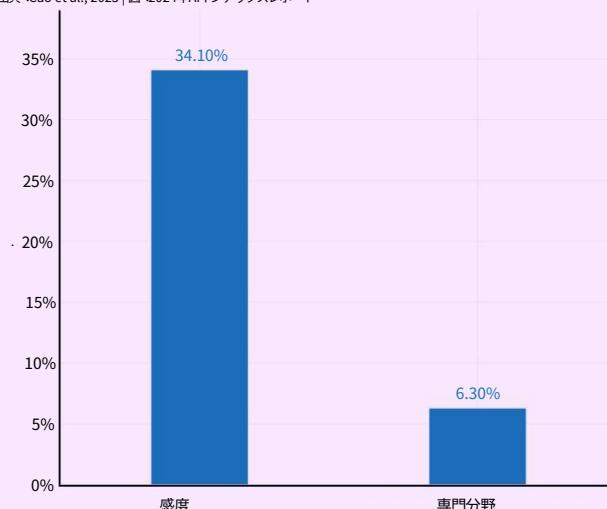


図5.2.17

現実世界のマルチシナリオにおけるPANDAのパフォーマンス検証 (患者数20,530人)

出典: Cao et al., 2023 | 図: 2024年AIインデックスレポート

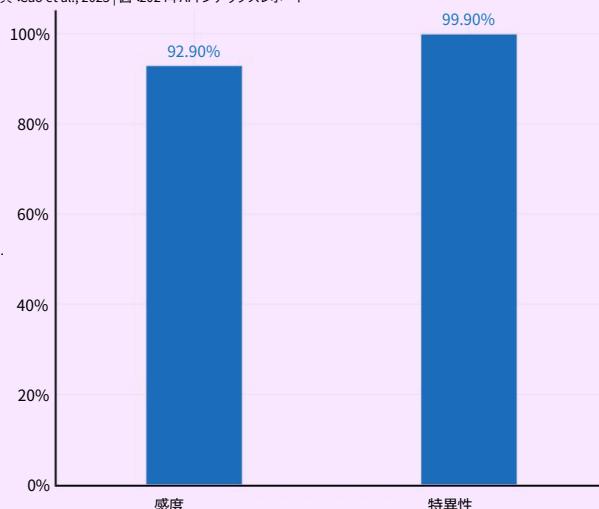


図5.2.18



その他の診断用途2023年に発表された新たな研究では、AIを他の診断分野でどのように活用できるかが強調されています。図5.2.19は、その研究結果の一部をまとめたものです。

診断AIのユースケースに関する追加調査

出典: AIインデックス、2024年

研究	使用事例	調査結果
ショップフランクス、2023	乳癌	著者らは、マンモグラフィ画像に基づくAIアルゴリズムに関する文献のメタレビューを実施しました。その結果、マンモグラフィ画像のみを用いて将来の乳がんリスクを予測することで、従来のリスク評価ツールと同等かそれ以上の精度が得られることが明らかになりました。
ディセンテ・シンド他、2023	X線謁影	研究者らは、画像と自由記述レポートを用いて胸部X線写真を分類するためのオープンソースニューラルネットワーク「X-Raydar」と「X-Raydar-NLP」を2つ開発しました。これらの自動分類手法は、人間の専門家と同等のレベルで機能し、外部データセットに適用した際にも堅牢性を示すことを明らかにしました。

図5.2.19



FDA承認のAI関連医療機器

米国食品医薬品局（FDA）

AI/ML対応医療機器のリストを管理しています

承認を受けた機器。このリストに掲載されている機器はFDAの市販前基準を満たしており、有効性の詳細なレビューを含む

安全性。2023年10月現在、FDAは生成AIを活用したデバイスやLLMを搭載したデバイスを承認していません。

図5.2.20は、過去10年間にFDAによって承認されたAI医療機器の数を示しています。

2022年には、合計139のAI関連医療機器がFDAの承認を受け、2021年の承認総数より12.1%増加しました。2012年以降、これらの機器の数は45倍以上に増加しています。

FDAが承認したAI医療機器の数（2012～2022年）

出典：FDA.2023年 図表：2024年AIインデックスレポート

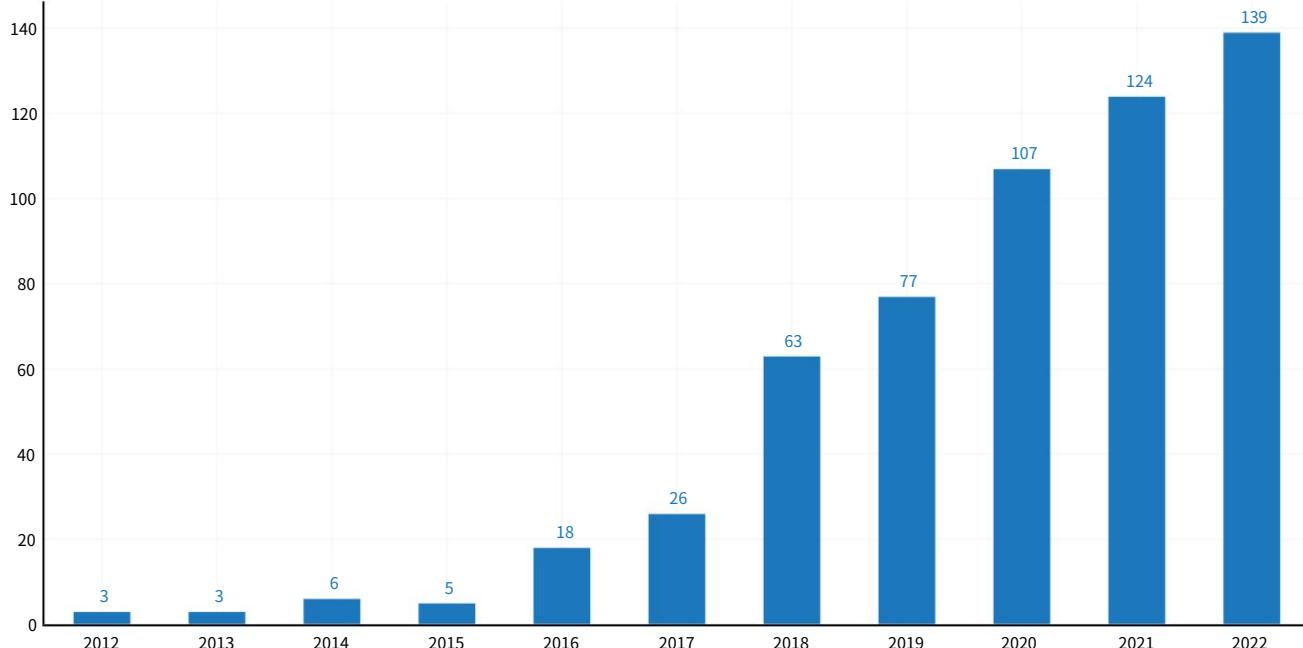


図5.2.20

3 FDAは2023年10月にリストを最後に更新したため、2023年の合計は不完全でした。そのため、AI Indexではデータの提示を2022年までの情報に限定しました。



図5.2.21は、FDA承認医療機器の専門分野を示しています。2022年に承認された139の機器のうち、87.1%と大多数が放射線関連でした。次に多かった専門分野は心臓血管で、承認件数の7.2%を占めました。

FDAが承認したAI医療機器の専門分野別数（2012～2022年）

出典：FDA.2023年 図表：2024年AIインデックスレポート

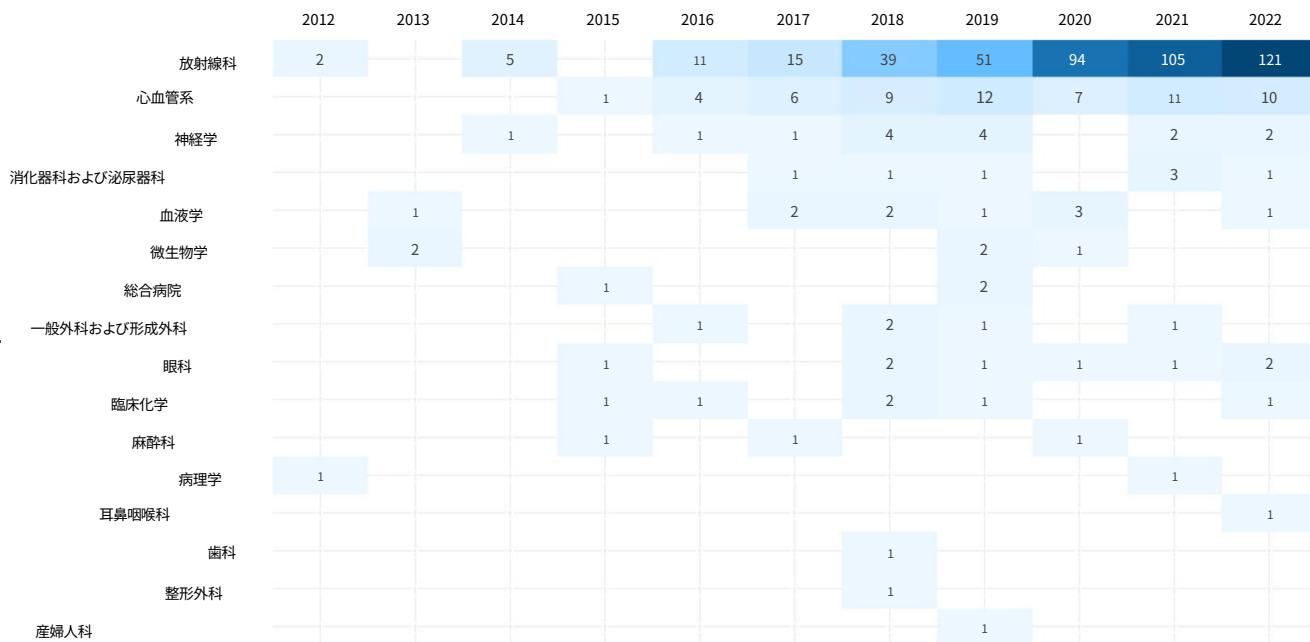


図5.2.21



管理とケア

AI ツールは、医療管理の効率を高め、患者ケアの水準を向上させる可能性も秘めています。

注目の研究:

メドアライン

ヘルスケアにおけるAIの大きな進歩にもかかわらず、既存の MedQAのようなベンチマークや USMLEは知識ベースの問題に重点を置いているため、臨床医が患者ケアにおいて行う多様な業務を十分に捉えきれていません。臨床医は、患者に合わせた診断計画の作成など、情報集約型の業務に従事することが多く、勤務時間のかなりの部分を事務作業に費やすています。AIはこれらのプロセスを効率化する可能性を秘めていますが、AIを活用するための十分な基盤がまだ不足しています。

適切な電子健康

医療管理 LLM のベンチマークと微調整のための記録 (EHR) データセット。

今年、研究者たちはこの問題に対処するために大きな進歩を遂げた。MedAlignを導入することでギャップを埋める：包括的なEHRベースの

7つの異なる医療専門分野から抽出された983の質問と指示、そして303の臨床医の回答からなるベンチマーク（図5.2.22）。MedAlignは、EHRに焦点を当てた初の包括的なベンチマークです。

研究者らは、MedAlign上で様々な既存のLLMをテストしました。全てのLLMの中で、多段階改良法を用いたGPT-4バリアントが最も高い正解率（65.0%）を達成し、他のLLMよりも常に高い正解率を示しました（図5.2.23）。MedAlignは、AIを活用して医療における事務負担を軽減するための貴重なマイルストーンです。

MedAlignワークフロー
出典：フレミング他, 2023

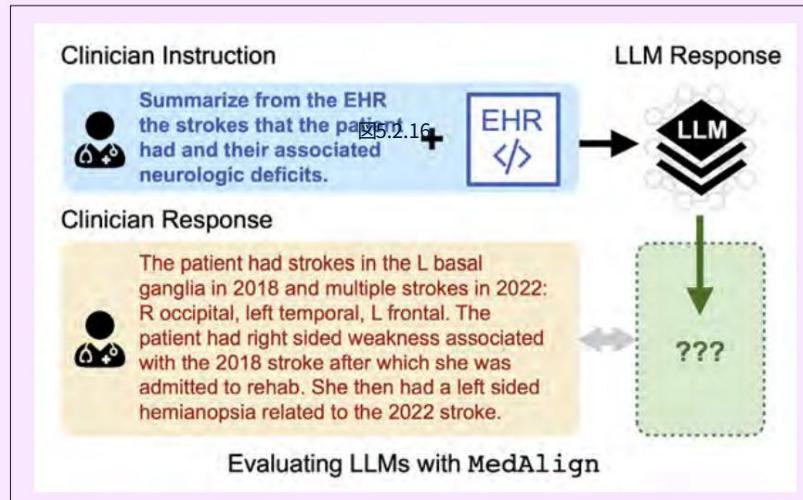


図5.2.22

注目の研究:

MedAlign (続き)

モデル性能の評価：人間とCOMETのランク
出典：Fleming他, 2023年 図表：2024年AIインデックスレポート

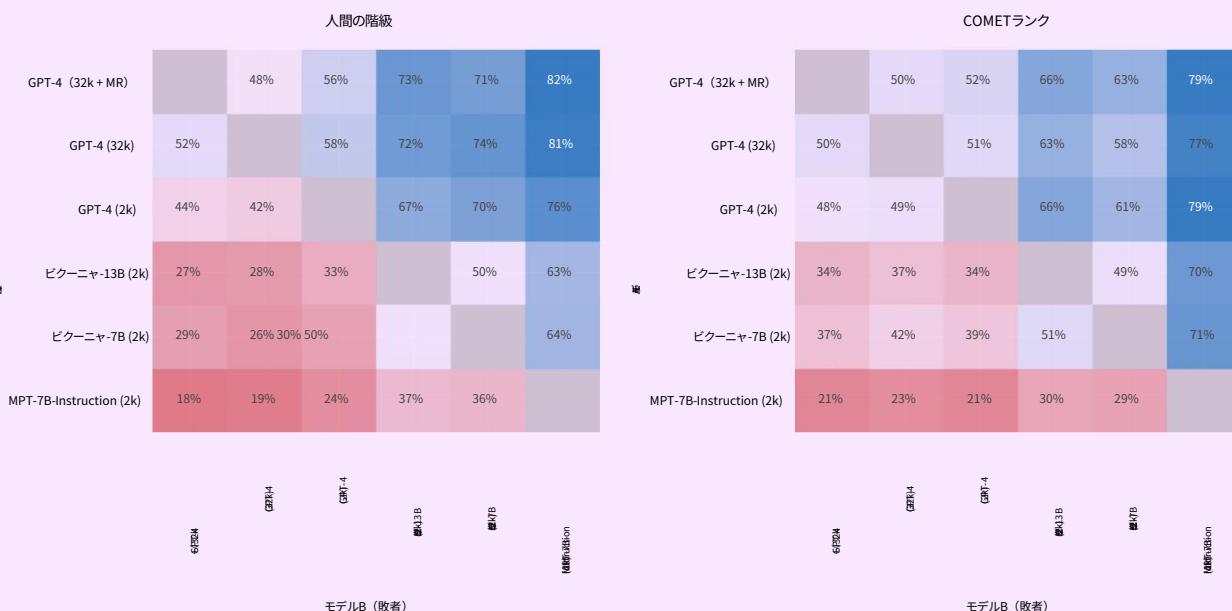
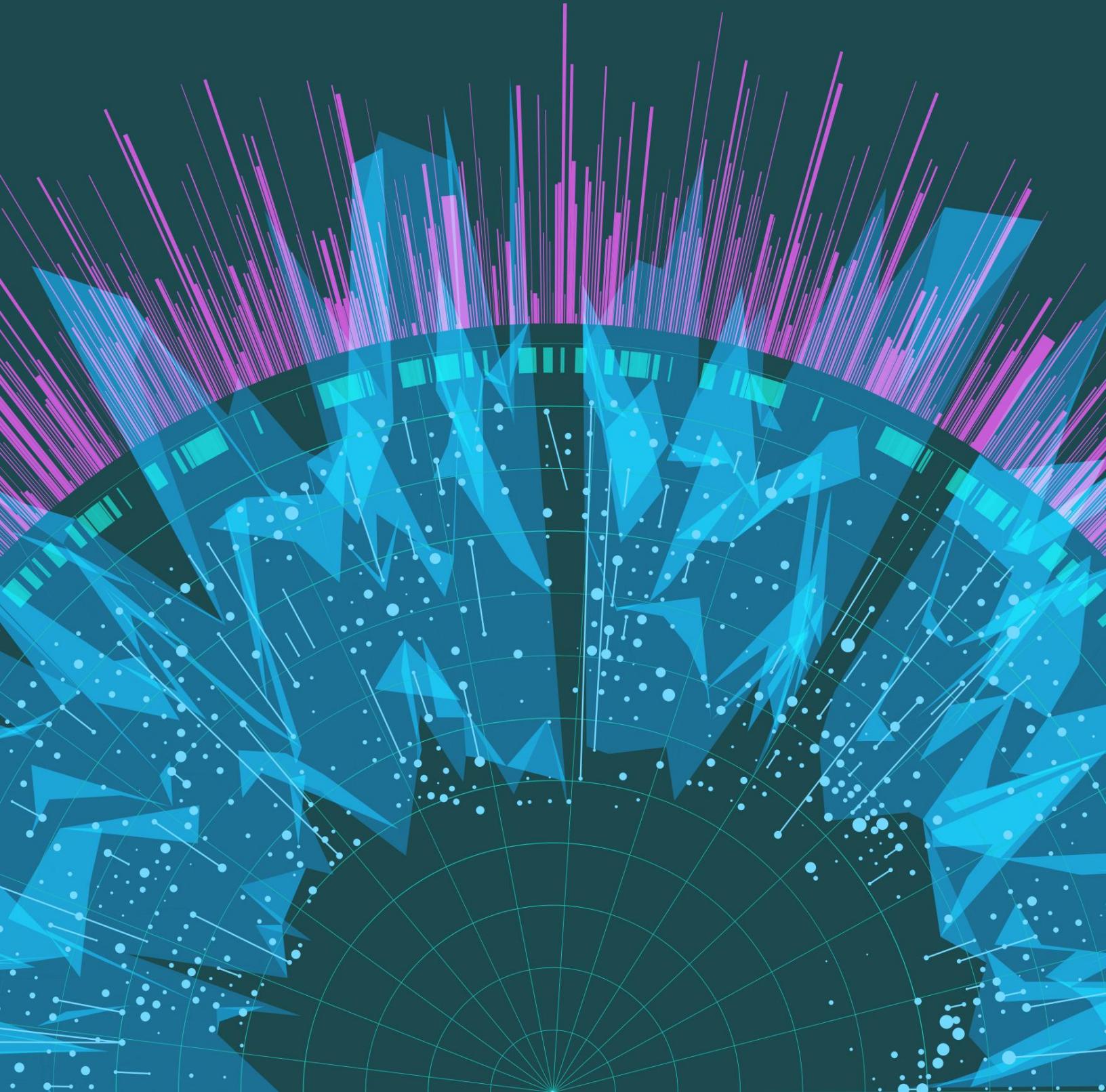


図5.2.23



人工知能
インデックスレポート2024

第6章:
教育



レビュー

概要	327
章のハイライト	328
6.1 高等教育におけるコンピュータサイエンスとAI教育	329
アメリカとカナダ	329
コンピュータサイエンス学士課程卒業生	329
コンピュータサイエンス修士課程卒業生	331
コンピュータサイエンス博士課程卒業生	333
CS,CE、情報学部	336
ヨーロッパ	344
情報科学、コンピュータサイエンス、コンピュータサイエンス、ITの学士課程卒業生	344
情報科学、コンピュータサイエンス、コンピュータサイエンス、ITの修士課程卒業生	347
情報科学、コンピュータサイエンス、電子工学、ITの博士課程卒業生	351
AI関連学習プログラム	355
合計コース数	355
教育レベル	356
地理的分布	357
6.2 K-12 CSとAI教育	359
アメリカ合衆国	359
州レベルの傾向	359
APコンピュータサイエンス	361
ハイライト:アクセスの問題	363
ハイライト:ChatGPTの使用状況	
教師と生徒	364

[公開データにアクセスする](#)

概要

本章では、AIとコンピュータサイエンス (CS) 教育のトレンドを検証し、誰が、どこで、そしてこれらのトレンドがどのように進化してきたかに焦点を当てます。AIが教育に与える影響への懸念が高まる中、ChatGPTのような新しいAIツールの教師と生徒による活用についても調査します。

本分析は、まず、Computing Research Association (CRA) の年次 Taulbee Surveyに基づき、米国とカナダにおける高等教育におけるコンピュータサイエンス (CS) および AI 教育の現状を概観することから始まります。次に、ヨーロッパにおける CS 教育に関する Informatics Europe のデータを確認します。今年は、Studyportals のデータに基づく、AI 関連の英語学習プログラムの世界規模での件数をまとめた新セクションを導入しています。

この章の最後は、Code.org による米国の K-12 CS 教育に関する洞察と、学校での ChatGPT の使用に関する Walton Foundation の調査結果で締めくくられています。



章のハイライト

1. アメリカとカナダにおけるコンピュータサイエンス (CS) の学士号取得者数は増加を続け、修士号取得者数はほぼ横ばい、博士号取得者数は緩やかに増加しています。アメリカとカナダにおける学士号取得者数は10年以上にわたり一貫して増加している一方で、コンピュータサイエンスの大学院教育を選択する学生数は横ばいとなっています。2018年以降、コンピュータサイエンスの修士号と博士号取得者数はわずかに減少しています。

2. AI博士号取得者の産業界への進出は加速を続けています。2011年には、新規AI博士号取得者の産業界 (40.9%) と学術界 (41.6%) への就職率はほぼ同率でした。しかし、2022年には、卒業後に産業界に就職する割合 (70.7%) が、学術界 (20.0%) を大幅に上回りました。過去1年間だけでも、産業界に就職するAI博士号取得者の割合は5.3ポイント上昇しており、大学から産業界への人材流出が加速していることを示しています。

3. 産業界から学術界への人材の流出が減少。2019年、米国とカナダにおけるAI分野の新規教員の13%は産業界出身者でした。この数字は2021年には11%に減少し、2022年にはさらに7%に減少しました。この傾向は、高度なAI人材の産業界から学術界への流出が徐々に減少していることを示しています。

4. 米国とカナダにおけるコンピュータサイエンス教育の国際性は低下している。2022年にコンピュータサイエンスの学士号、修士号、博士号を取得した留学生の数は、2021年と比較して減少した。特に修士課程における留学生の減少が顕著であった。

5. アメリカの高校生の間でコンピュータサイエンスのコースを受講する人が増えていますが、アクセスの問題は依然として残っています。2022年には、20万1000件のAPコンピュータサイエンス試験が実施されました。2007年以降、受験者数は10倍以上に増加しています。しかし、最近の調査結果によると、大規模高校や郊外の高校の生徒の方がコンピュータサイエンスのコースを受講しやすいことが示されています。

6. AI関連の学位プログラムは国際的に増加傾向にあります。英語で提供されるAI関連の高等教育機関の学位プログラムは2017年以降3倍に増加しており、過去5年間で着実に増加しています。世界中の大学でAIに特化した学位プログラムの提供が増えています。

7. 英国とドイツは、ヨーロッパにおける情報科学、コンピュータサイエンス、電子工学、IT分野の卒業生の輩出においてトップを占めています。英国とドイツは、情報科学、コンピュータサイエンス、電子工学、情報分野の学士、修士、博士の新規卒業生数においてヨーロッパをリードしています。人口一人当たりの卒業生数では、フィンランドが学士と博士の両方でトップ、アイルランドが修士課程卒業生の輩出でトップです。



このセクションでは、コンピュータサイエンス（CS）と人工知能（AI）の高等教育の概要を解説し、北米とヨーロッパにおける学士号、修士号、博士号を含む様々な学位の卒業統計に焦点を当てています。また、英語で提供されるAI関連コースに関する情報も掲載しています。

6.1 高等教育におけるコンピュータサイエンスとAI教育

アメリカとカナダ

このサブセクションでは、米国とカナダにおけるコンピュータサイエンスとAIの高等教育の現状を評価するコンピューティング研究協会のTaulbee調査のデータの分析を示します。

この調査は、米国とカナダ全土の 297 の博士号を授与するコンピュータサイエンスの学部を対象としています。¹

コンピュータサイエンス学士課程卒業生

過去 10 年間、北米のコンピュータサイエンス学士課程の新規卒業生の総数は着実に増加しており、2021 年から 2022 年にかけて前年比 7.9% 増加し、3 倍以上に増加しました（図 6.1.1）。

2010～2022年における米国とカナダのコンピュータサイエンス学士課程の新卒者数

出典：CRA Taulbee Survey, 2023年 図表：2024年AI指数レポート

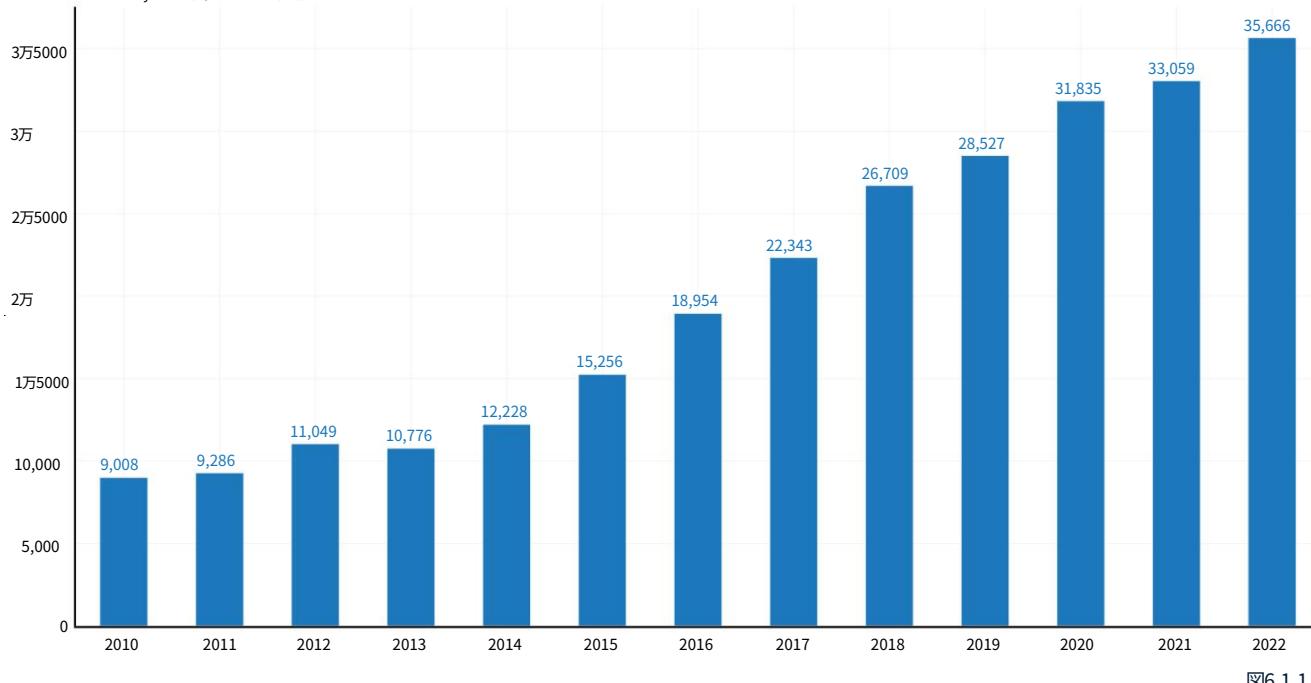


図6.1.1

¹ 調査対象となった博士号授与学部の全てが回答したわけではないことに留意することが重要です。調査対象となった297学部のうち、回答したのはわずか182学部であり、全体の回答率は61%でした。



アメリカとカナダの大学のコンピュータサイエンス学士課程卒業生における留学生の割合は、約8年ぶりに減少し、2021年の16.3%から2022年には15.2%に減少しました（図6.1.2）。この減少は、トランプ政権初期における留学ビザ取得の困難化を反映している可能性が高く、その影響は今になってようやく現れ始めています。

データによると、この減少は、COVID-19パンデミック中に課された国際渡航制限も一因となっており、留学生が米国とカナダで学ぶ機会に影響を与えています。この最近の減少にもかかわらず、過去10年間の全体的な傾向は、留学生の割合が着実に増加していることを示しています。

2010～2022年における米国とカナダの新規国際コンピュータサイエンス学士課程卒業生（全体の割合）

出典：CRA Taulbee Survey, 2023年 図表：2024年AI指数レポート

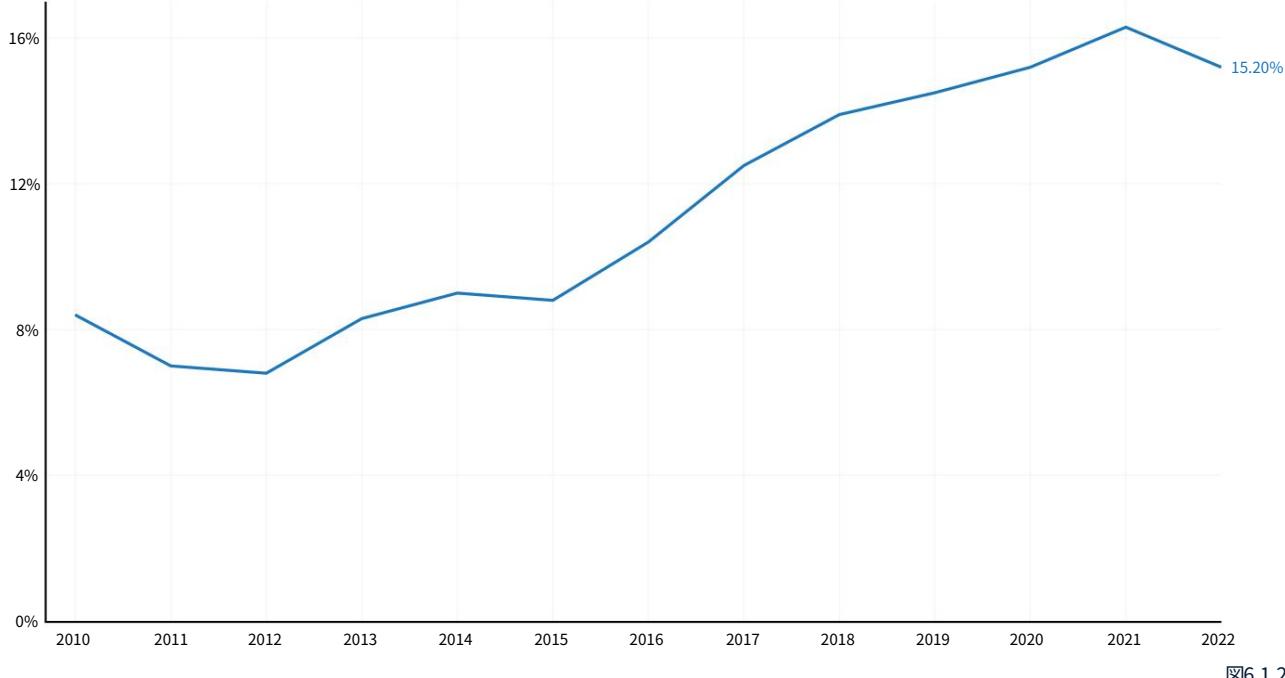


図6.1.2



コンピュータサイエンス修士課程卒業生

AIコースはコンピュータサイエンスの修士課程によく含まれています。アメリカとカナダの大学におけるコンピュータサイエンスの修士課程の卒業生総数は過去10年間で2倍以上に増加しましたが、

2018年以降、この数は横ばい傾向にあり、昨年は2.5%の微減となりました（図6.1.3）。この横ばい傾向は、次のグラフに示されているように、修士課程の留学生数の減少を反映しています。

2010～2022年における米国とカナダのコンピュータサイエンス修士課程修了者数

出典：CRA Taulbee Survey, 2023年 図表：2024年AI指数レポート

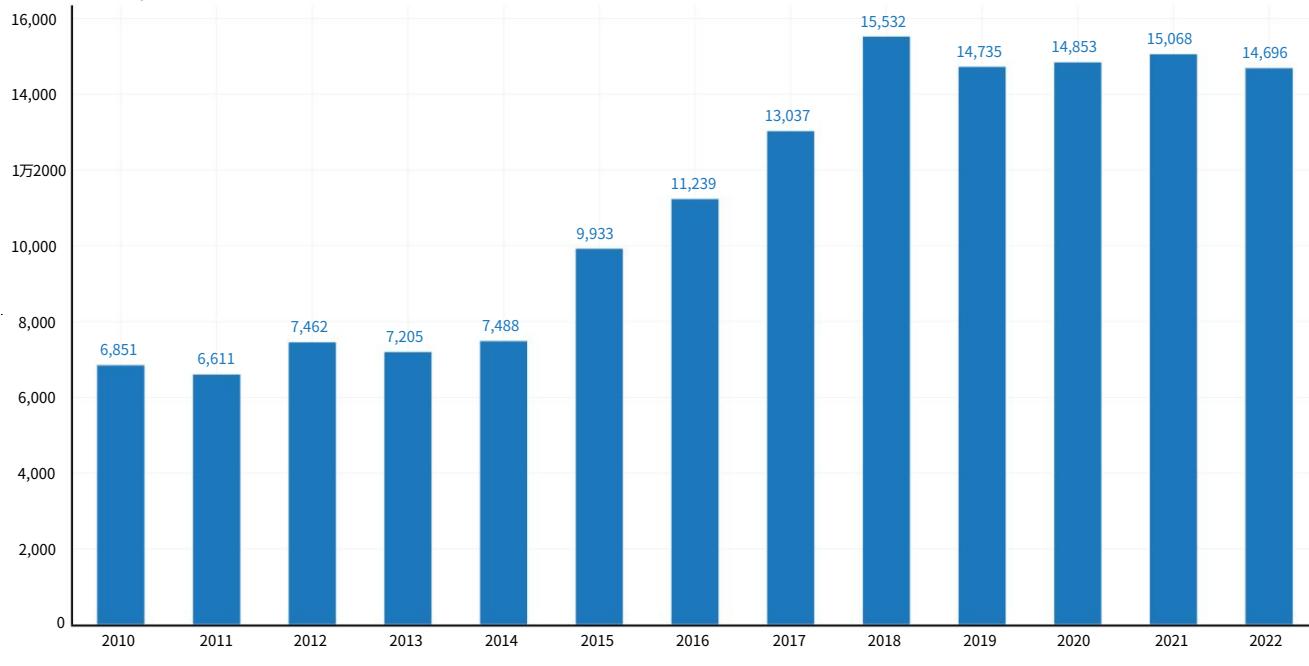


図6.1.3



2022年、アメリカとカナダの大学では、国際的なコンピュータサイエンスの修士課程の学生数が著しく減少しました。この減少傾向は2017年頃から始まりましたが、昨年は14.8ポイント減少し、最も顕著になりました（図6.1.4）。現在、コンピュータサイエンスの修士課程修了者のうち、海外と国内の比率はほぼ均等になっています。

2010～2022年における米国とカナダの新規国際コンピュータサイエンス修士課程卒業生数（全体の割合）

出典：CRA Taulbee Survey, 2023年 図表：2024年AI指数レポート

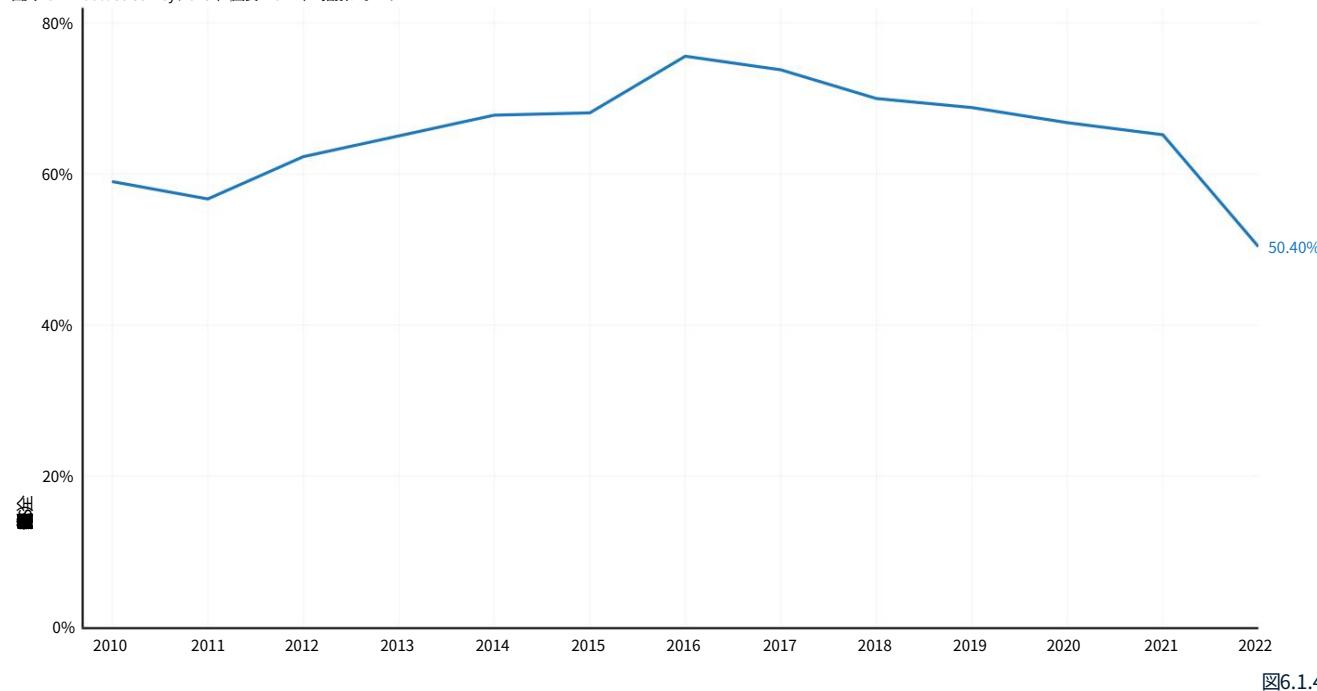


図6.1.4



コンピュータサイエンス博士課程卒業生

アメリカとカナダの大学では、10年ぶりにコンピュータサイエンス（CS）の博士号取得者数が大幅に増加しました。2022年には、コンピュータサイエンスの博士号取得者数は2,105人に達し、2010年以来の最高数となりました（図6.1.5）。

2010～2022年における米国とカナダのコンピュータサイエンス博士課程修了者数

出典 :CRA Taulbee Survey,2023年 図表 :2024年AI指数レポート

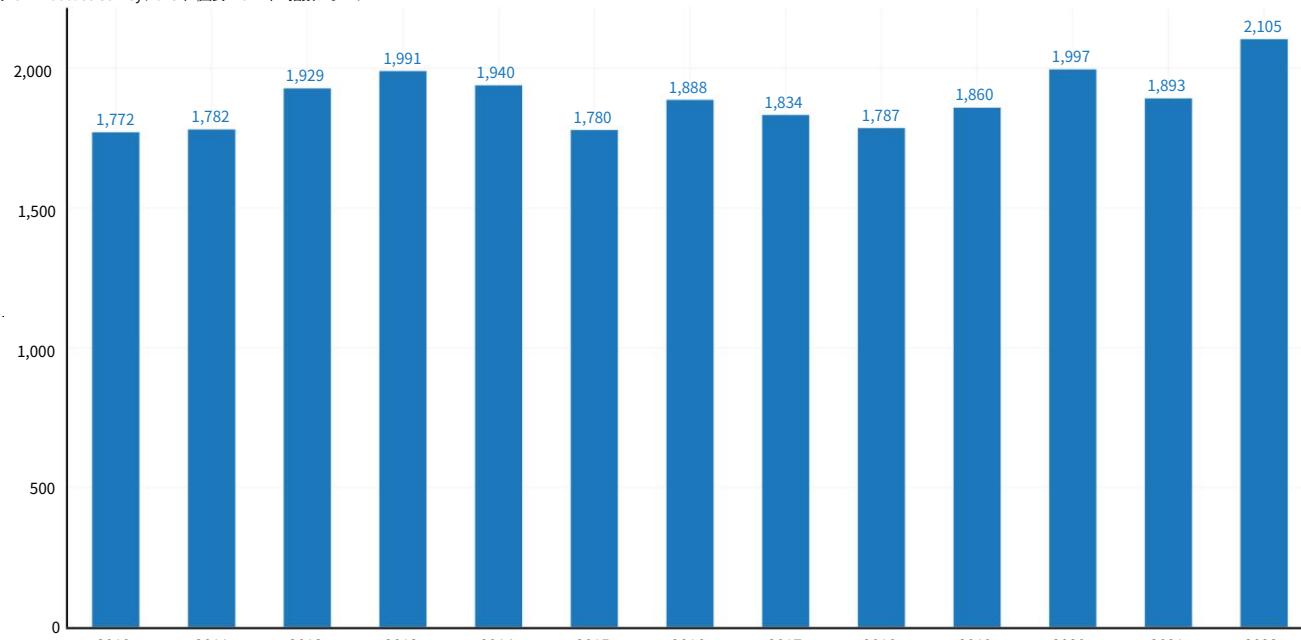


図6.1.5



コンピュータサイエンスの博士課程修了者のうち留学生の割合は過去10年間で増加していますが、昨年はこの割合がわずかに減少し、2021年の68.6%から2022年には65.9%に減少しました（図6.1.6）。

2010～2022年の米国とカナダにおける新規国際コンピュータサイエンス博士号取得者数（全体の割合）

出典：CRA Taulbee Survey, 2023年 図表：2024年AI指数レポート

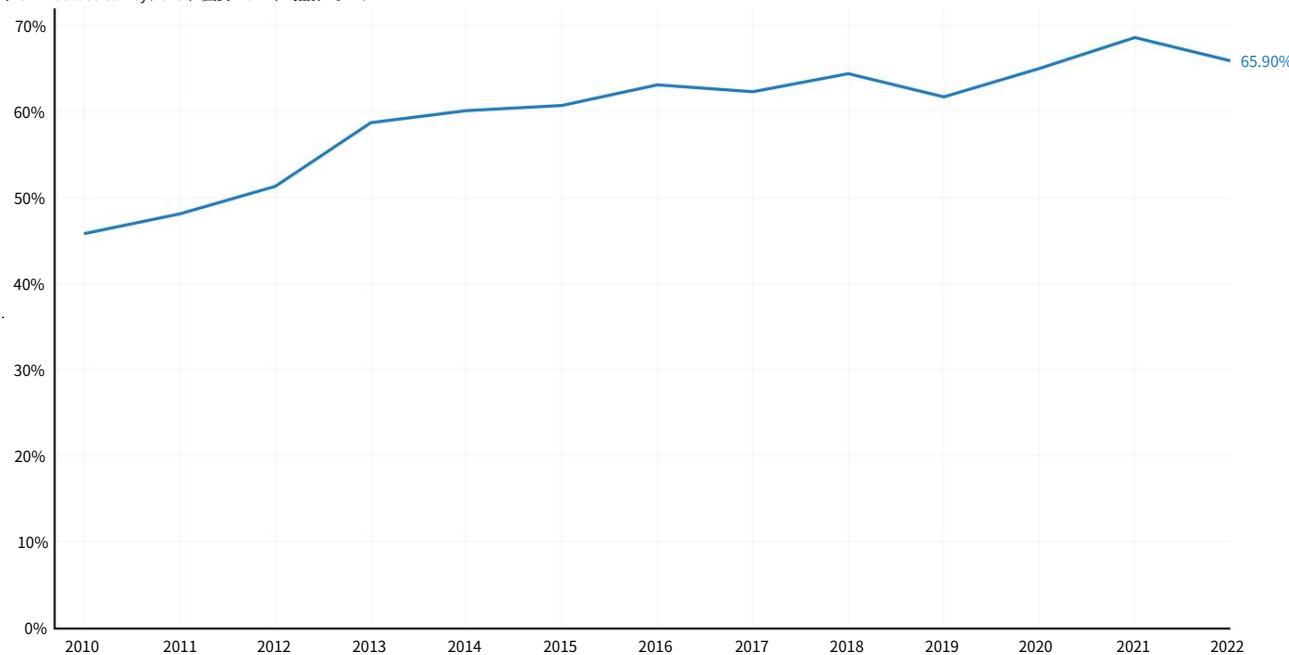


図6.1.6



AI関連の博士号を取得したばかりの人は、卒業後、どのような就職先を選ぶのでしょうか？昨年のAI Indexレポートで強調された傾向に倣い、AI関連の博士号取得者のうち、産業界への就職を目指す人の割合が増加しています（図6.1.7および図6.1.8）。2011年には、産業界（40.9%）と学術界（41.6%）の就職率はほぼ同率でした。

しかし、2022年までに、卒業後に産業界に就職する人の割合（70.7%）は、学界に進学する人（20.0%）を大幅に上回りました。AI関連の博士号取得者が政府機関に就職する割合は、過去5年間で約0.7%と比較的低く安定しています。

AI博士号取得者の新規雇用（全体の割合）
米国とカナダのセクター別、2010～2022年

出典：CRA Taulbee Survey, 2023年 図表：2024年AI指数レポート

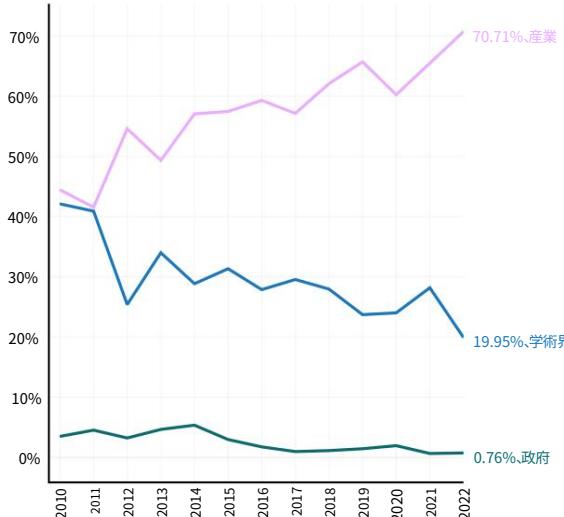
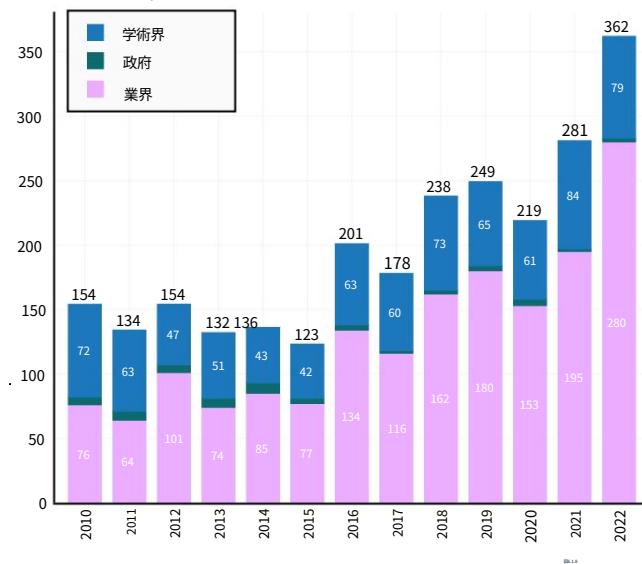


図6.1.72

米国におけるAI博士号取得者の新規雇用と

カナダのセクター別データ、2010～2022年

出典：CRA Taulbee Survey, 2023年 図表：2024年AI指数レポート



2.図6.1.7の合計は100にならない。これは、毎年新たにAI博士号を取得する学生の一部が、自営業者、失業者、あるいはCRA調査で「その他」の就業状況を報告しているためである。これらの学生はこの図には含まれていない。



CS、CE、情報系の教員CSとAI教育の動向を

より深く理解するには、CS教員に関するデータを調べることが有用です。昨年、アメリカとカナダの大学におけるCS、CE、情報系の教員の総数は7.2%増加しました（図6.1.9）。

2011年以降の増加率は42.4%です。

米国とカナダにおけるコンピュータサイエンス、コンピュータサイエンス、情報学の教員数、2011～2022年

出典：CRA Taulbee Survey、2023年 図表：2024年AI指数レポート

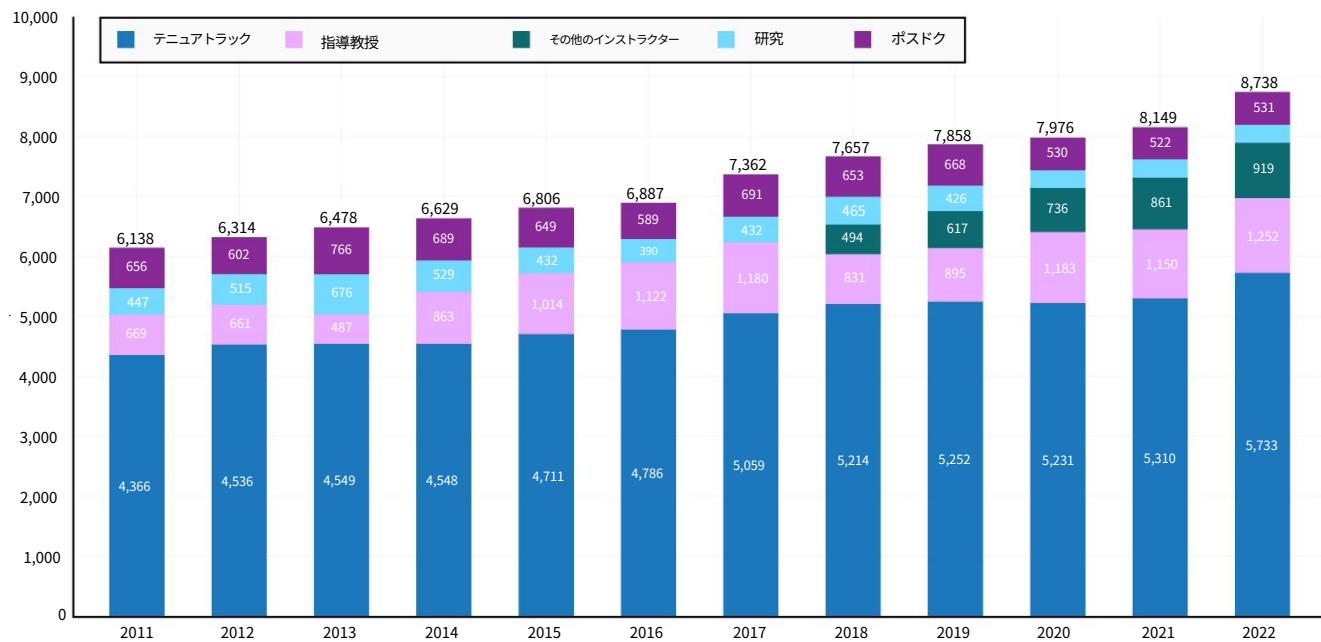


図6.1.9



2022年、米国のコンピュータサイエンス（CS）教員は7,084名で、その大半（65.7%）がテニュアトラックに在籍しています（図6.1.10）。米国のコンピュータサイエンス教員総数は、2021年から4.4%、2011年から45.0%増加しています。

米国のコンピュータサイエンス教員数、2011～2022年

出典：CRA Taulbee Survey, 2023年 図表：2024年AI指数レポート



図6.1.10



昨年、北米ではコンピュータサイエンス、コンピュータエンタテインメント、情報分野の教員が合計 915 名新たに採用され、過去 10 年間で最多を記録しました。これらのポジションのうち 455 名は終身在職権付きでした。(図 6.1.11)

米国とカナダにおけるコンピュータサイエンス、電子工学、情報学の新規教員採用数（2011～2022年）

出典：CRA Taulbee Survey, 2023年 図表：2024年AI指数レポート

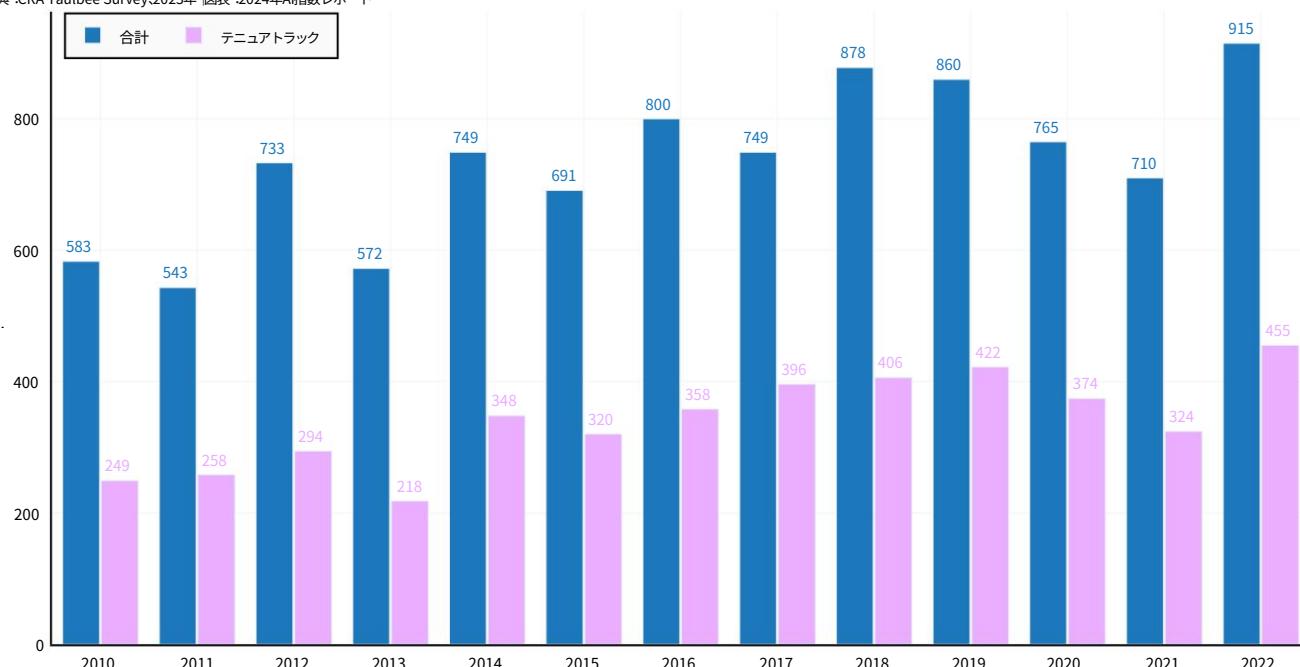


図6.1.11



2022年に新たに任命された教員の43%は、他の学術職からの転籍であり、学術人材の「流動性」を示しています（図6.1.12）。これらの「新規」教員は他大学のポストを空けているため、いずれは以前のポストを補充する必要があります。さらに、2022年に産業界から転籍した教員の割合は、前年の11%、2019年の13%から7%に減少しました。

アメリカとカナダのコンピュータサイエンス、コンピュータサイエンス、情報学部における新規教員の供給源（2018～2022年）
出典：CRA Taulbee Survey, 2023年 図表：2024年AI指数レポート

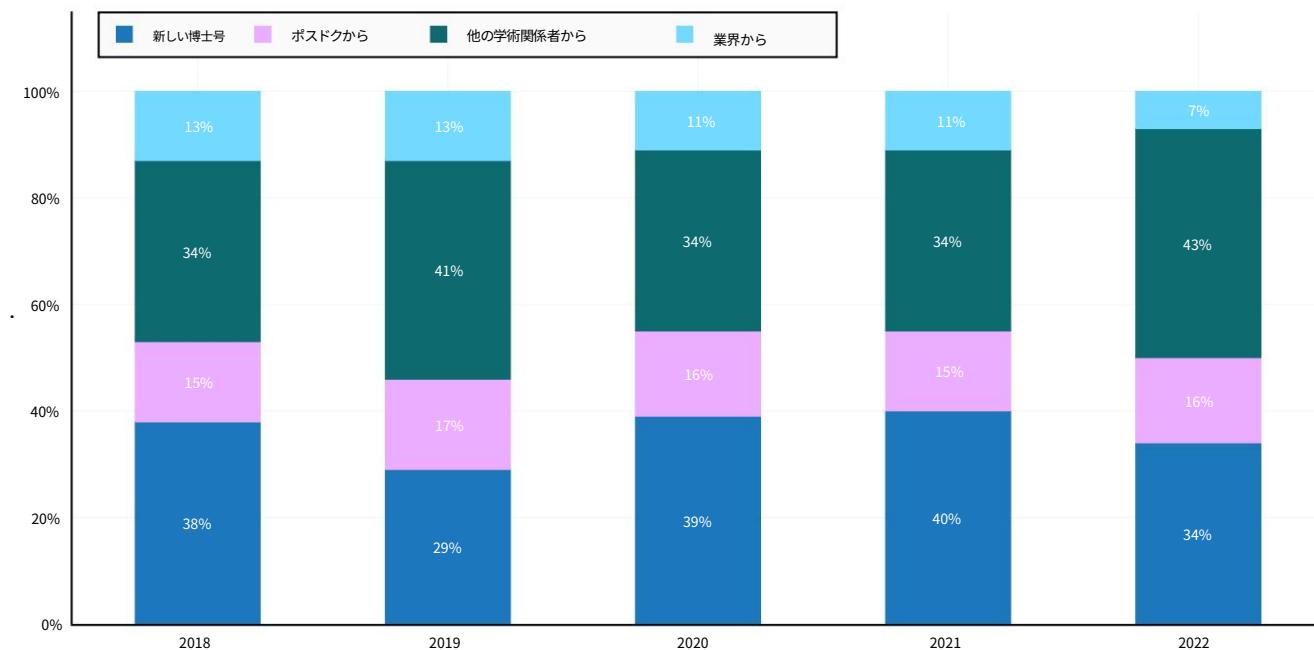


図6.1.12



教員のポストが空席のままとなっている理由は、過去10年間で変化しています。2011年には、採用選考に失敗した理由の37%は内定が出なかつたこと、34%は内定を辞退したことによるものでした（図6.1.13）。対照的に、2022年には内定が出なかつたのはわずか15%で、内定が出た後の55%は内定が出なかつたことが原因です。

断られました。この傾向は、コンピュータサイエンスの新規教員獲得競争が激化していることを反映しているようです。しかし、これが他の学術的ポジションとの競争の激化を示しているのか、あるいは産業界のポジションとの競争の激化を示しているのかは不明である。

2011～2022年におけるコンピュータサイエンス、電子工学、情報学部の新設教員ポストが満員にならなかつた理由（全体の比率）
出典：CRA Taulbee Survey, 2023年 図表：2024年AI指数レポート

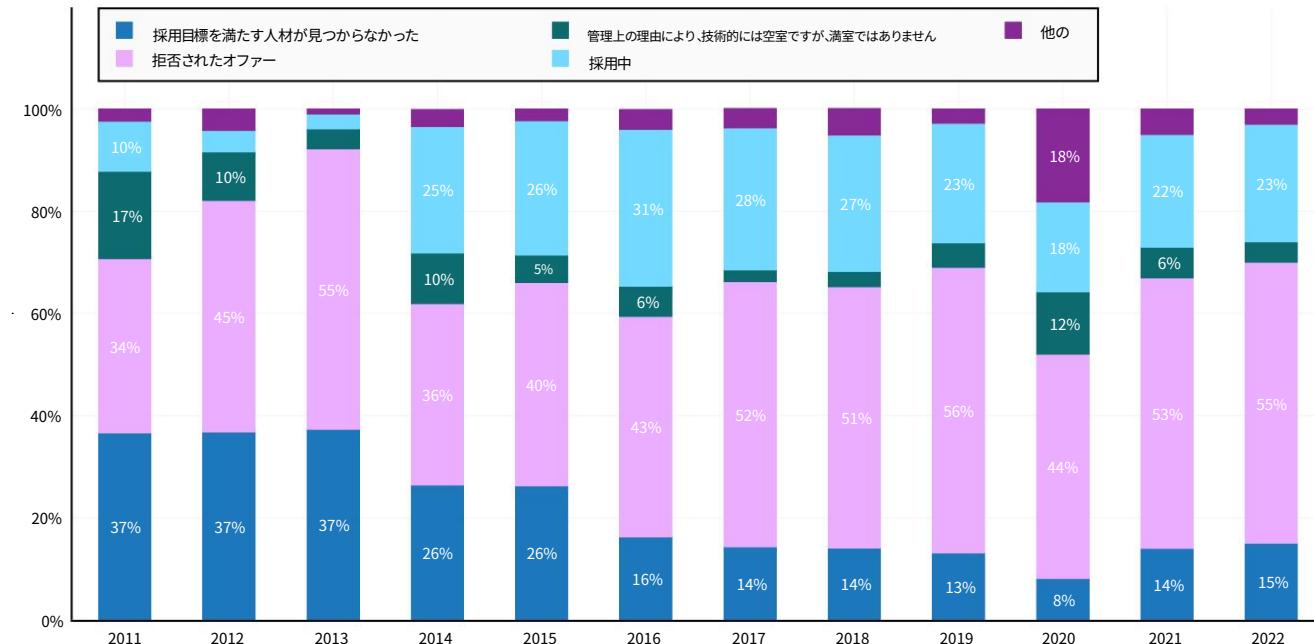


図6.1.13



2022年、北米のコンピュータサイエンス、コンピュータエンタテインメント（CE）、情報分野の学部では、教員の退職者数が大幅に増加し、合計405名となりました。これは2021年の303名から増加しています（図6.1.14）。このうち38.5%は他の学術職に、16.3%は非学術職に異動しており、前年と同様の傾向が続いています。

アメリカとカナダのコンピュータサイエンス、電子工学、情報学部における教員の減少、2011～2022年
出典：CRA Taulbee Survey, 2023年 図表：2024年AI指数レポート

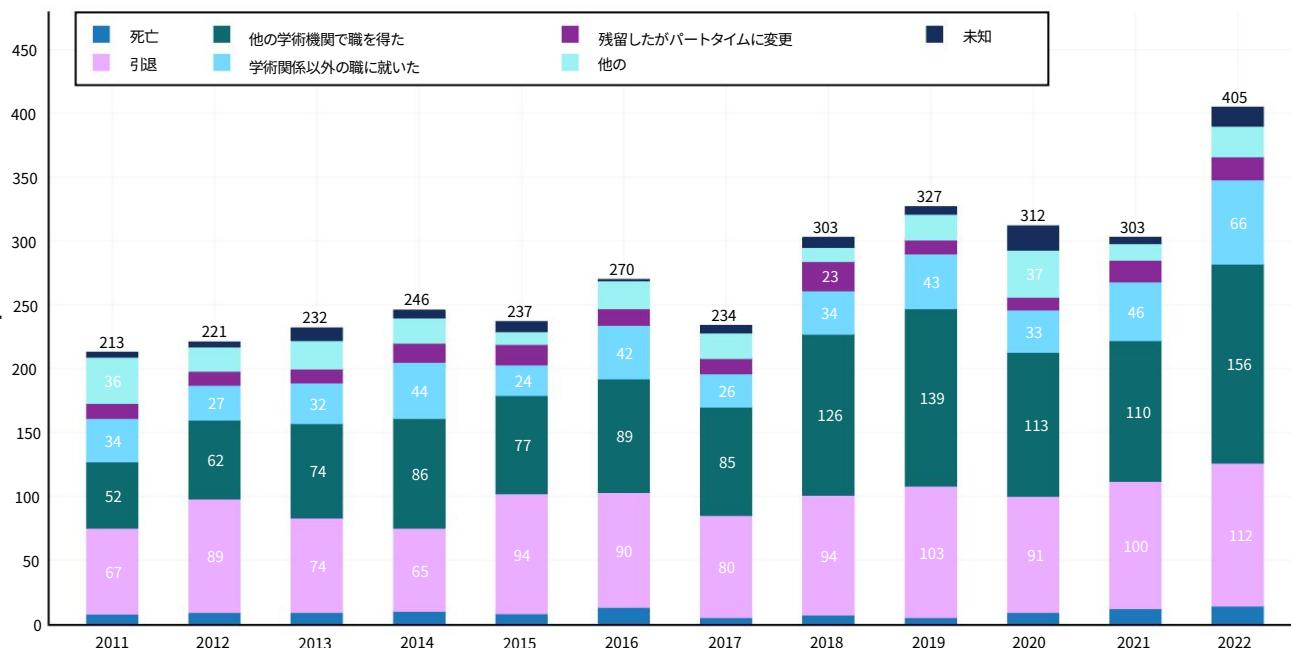


図6.1.14



2015年以降、教授の9か月間の給与の中央値の上昇は米国のインフレ率をわずかに下回っている一方、助教授と准教授の給与の中央値はインフレ率をわずかに上回る上昇が見られます。

2022年の教授の給与は2021年より3.2%高かったが、米国のインフレ率7%には追いつかなかった。また、2015年より16.4%高かったが、その期間のインフレ率19%の上昇には及ばなかった（図6.1.15）。

米国のコンピュータサイエンス教員の9か月平均給与（2015～2022年）

出典：CRA Taulbee Survey, 2023年 図表：2024年AI指数レポート

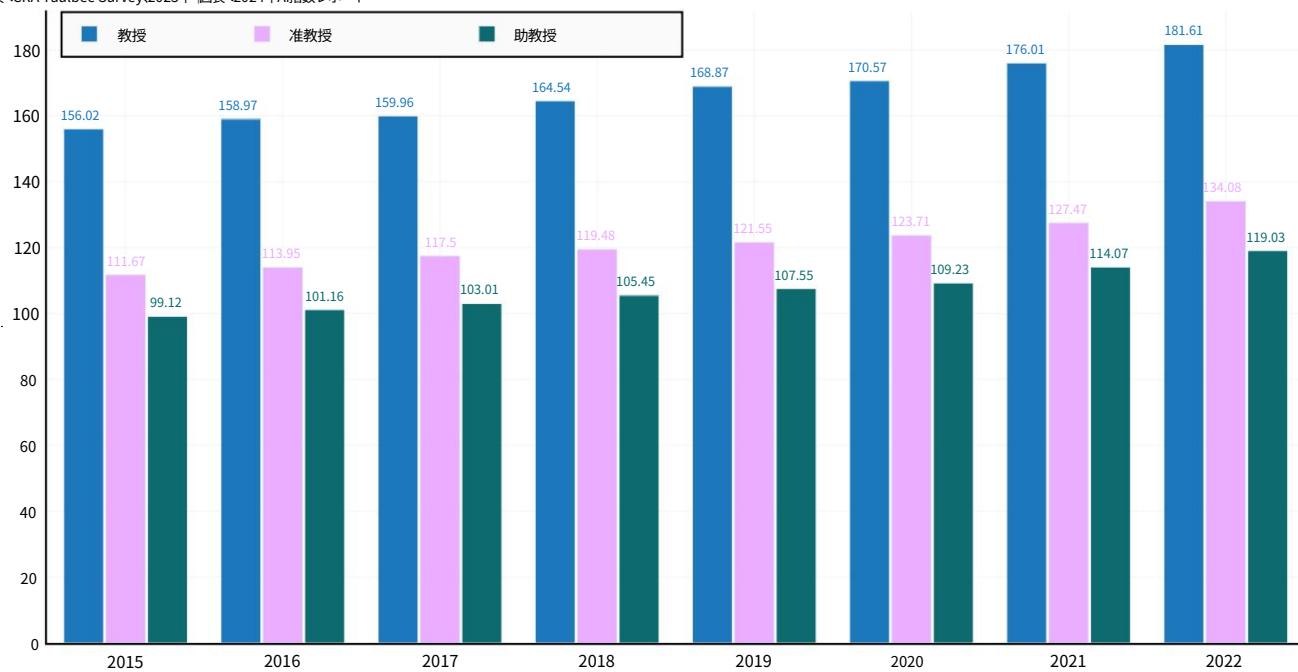


図6.1.15



2022年には、コンピュータサイエンス、コンピュータエンタテインメント（CE）、情報分野の新規テニュアトラック教員における外国人採用者の割合が、前年の13.2%から19.3%へと大幅に増加しました（図6.1.16）。これは過去10年間で2番目に高い割合であり、2013年に次ぐ記録となりました。

米国およびカナダにおけるCS、CE、情報分野の国際テニュアトラック教員の新規採用数（全体の割合）
カナダ、2010～2022年

出典：CRA Taulbee Survey, 2023年 図表：2024年AI指数レポート



図6.1.16



ヨーロッパ

ヨーロッパの CS 卒業生に関するデータは、ヨーロッパにおける情報科学教育の状況を監視することを目的とした学術研究コミュニティである Informatics Europe から提供されています。³ Informatics Europe は、ヨーロッパ政府の統計局から情報科学、CS、CE、コンピューティング、情報技術 (IT) 分野の卒業生に関するデータを収集しています。⁴

情報科学、コンピュータサイエンス、コンピュータサイエンス、IT の学士号

卒業生

2022年、英国は情報科学、コンピュータサイエンス、コンピュータエンタテインメント (CE)、IT の学士課程における新規卒業生数で最多となり、合計約2万5000人に達しました（図6.1.17）。ドイツとトルコが僅差で続きました。サンプルに含まれるほとんどの国では、10年前と比較してこれらの分野の卒業生数が増加しましたが、ポーランド、スペイン、チェコ共和国などの例外もありました（図6.1.18）。

2022年、ヨーロッパにおける情報科学、コンピュータサイエンス、コンピュータエンタテインメント (CE)、IT の学士課程の新卒者数（国別）
出典：Informatics Europe, 2023年 図表：2024年AIインデックスレポート

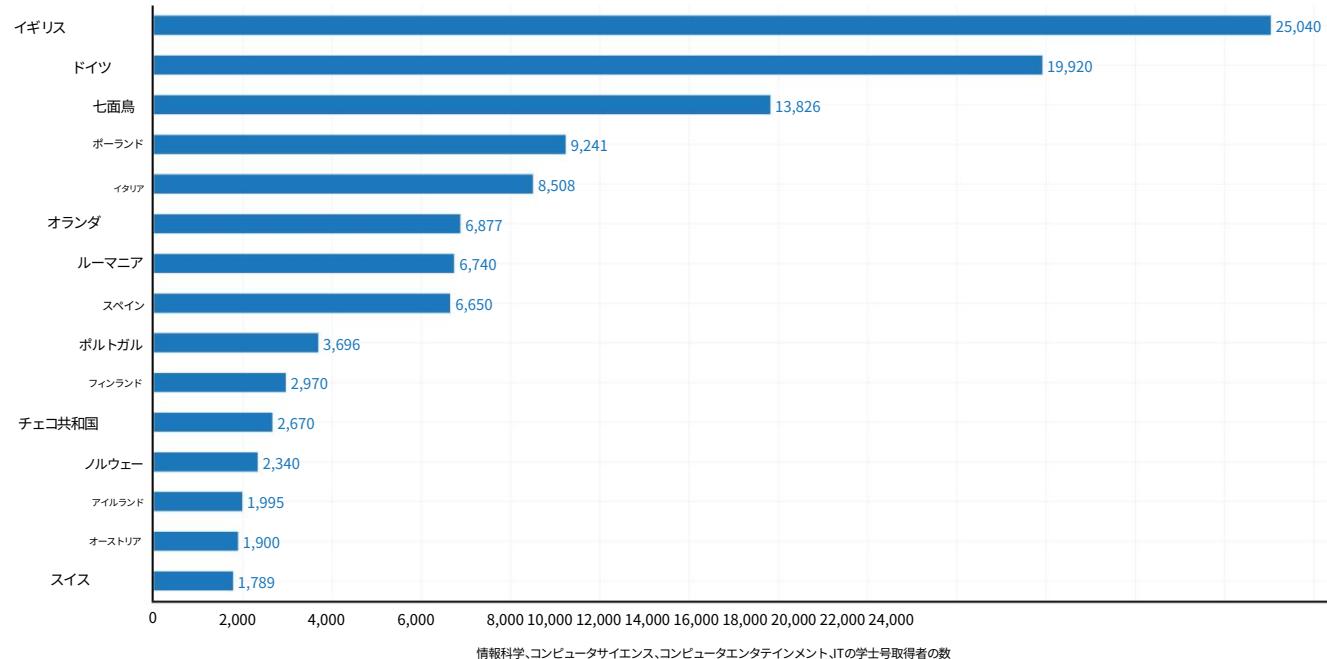


図6.1.17

³ ヨーロッパ諸国で統一的に使用されているコンピュータサイエンス教育の単一用語は存在しません。ヨーロッパ全域で、コンピュータサイエンス教育は、インフォマティクス（情報科学）、コンピュータサイエンス（CS）、コンピュータエンジニアリング（CE）、コンピューティング、情報技術（IT）、情報通信技術（ICT）、情報科学技術（IST）といった用語で表現されます。Informatics Europeがインフォマティクス研究プログラムを識別するために使用している科目名（および英語訳）の全リストは、以下のリンク先でご覧いただけます。

⁴ 欧州のデータは各国の統計局から収集されており、より広範囲に及ぶため、読者は CRA 北米データと後続のセクションで詳述する欧州の CS 卒業生データについて人口当たりの比較を行わないよう注意が必要です。

5 本セクションの図には、AI指標のデータがあるすべての国が示されているわけではないことにご注意ください。完全なデータにアクセスするには、本章に記載する公開データをご覧ください。さらに、年ラベルは学年度の終了年を表します。例えば、2022年度の卒業生数を視覚化した数値は、2021/2022年度の卒業生数を反映しています。視覚的な簡略化のため、本指標では学生が卒業した年に焦点を当てています。



ヨーロッパの国別情報科学、コンピュータサイエンス、コンピュータエンタテインメント、IT学士課程卒業生の割合の変化
2012年と2022年

出典: Informatics Europe, 2023年 図表: 2024年AIインデックスレポート

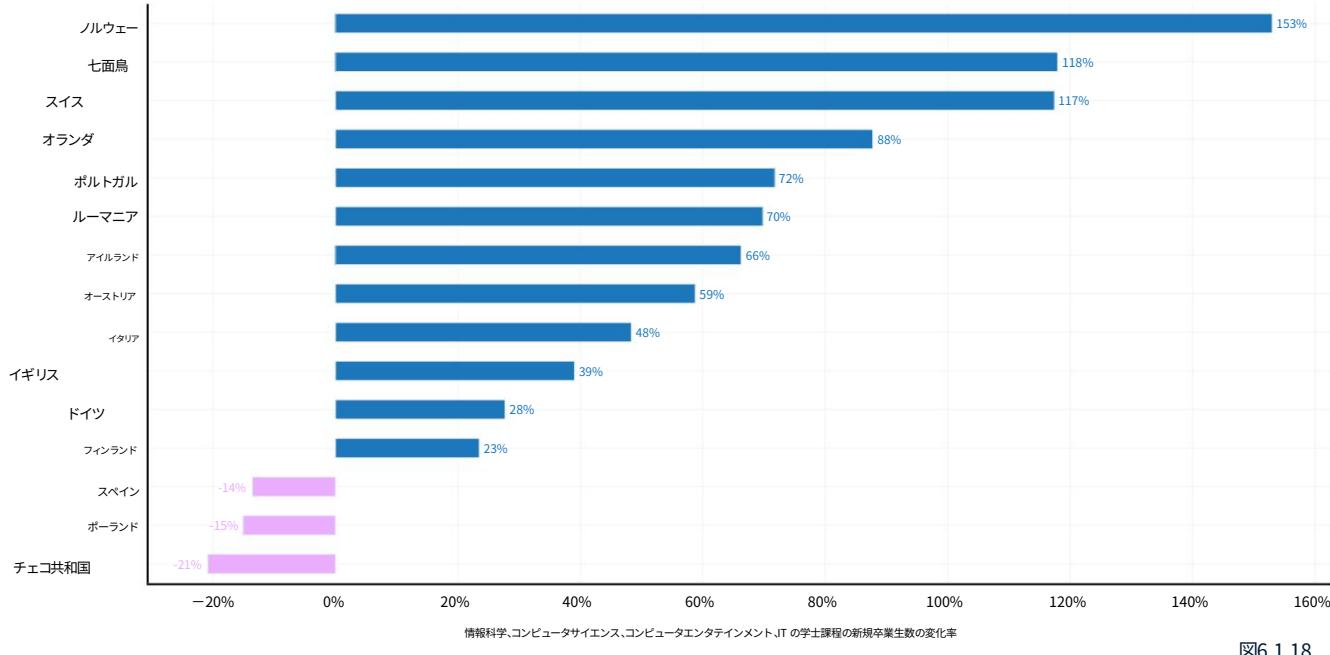


図6.1.18

フィンランド（53.4）、ノルウェー（42.6）、オランダ（38.6）は、人口10万人あたりの情報科学（CS、CE、IT）の学士号取得者数で上位を占めています（図6.1.19）。人口1人当たりで見ると、サンプル調査の対象となったほとんどのヨーロッパ諸国で、情報科学、CS、CE、ITの学士号取得者数が増加しています（図6.1.20）。