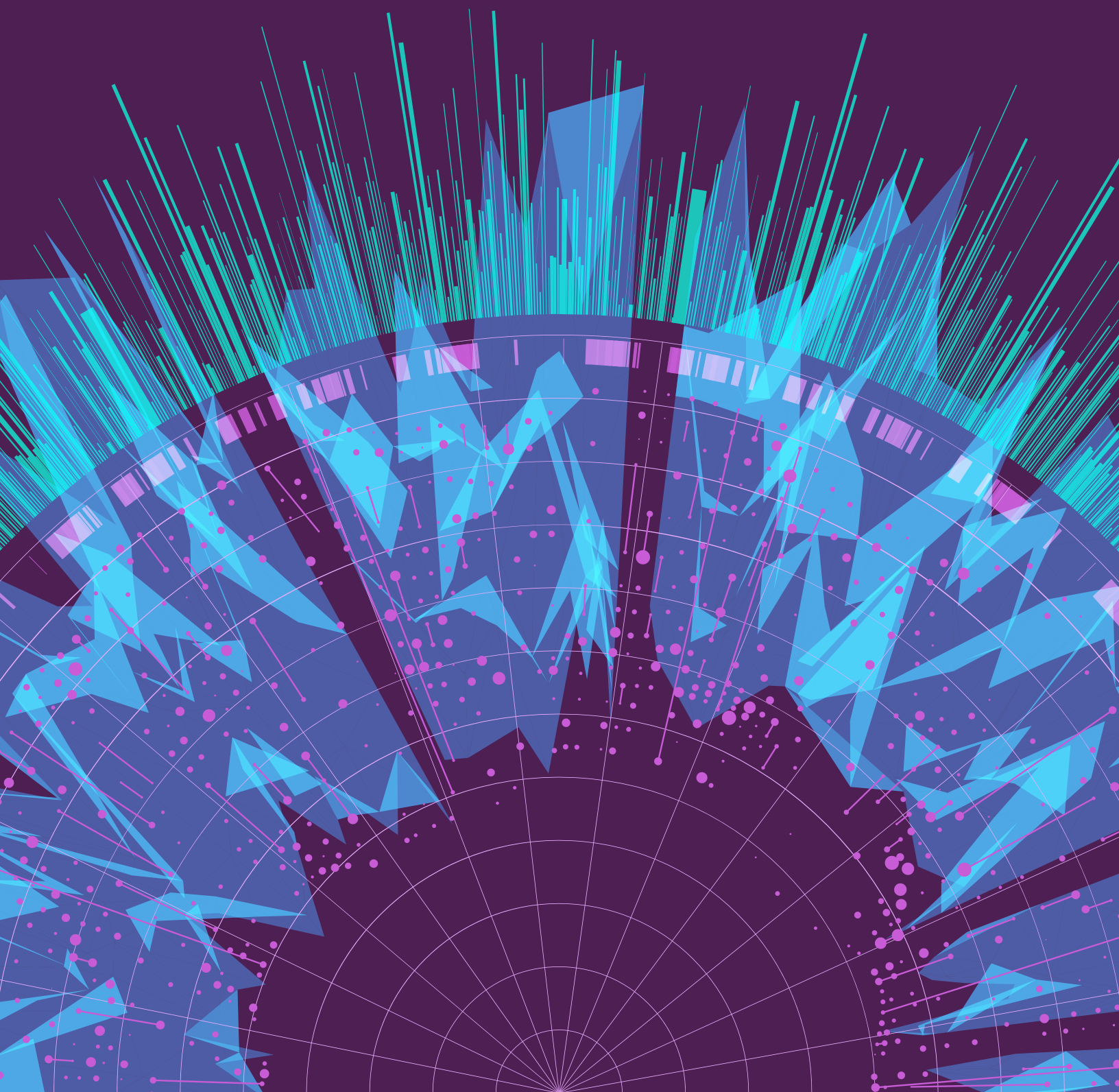




Artificial Intelligence
Index Report 2024

CHAPTER 5: Science and Medicine



プレビュー

Overview	298
Chapter Highlights	299

5.1 Notable Scientific Milestones	300
AlphaDev	300
FlexiCubes	301
Synbot	303
GraphCast	304
GNoME	305
Flood Forecasting	306

5.2 AI in Medicine	307
Notable Medical Systems	307
SynthSR	307
Coupled Plasmonic Infrared Sensors	309
EVEscape	310
AlphaMissence	312
Human Pangenome Reference	313
Clinical Knowledge	314
MedQA	314
Highlighted Research: GPT-4 Medprompt	315
Highlighted Research: MediTron-70B	317
Diagnosis	318
Highlighted Research: CoDoC	318
Highlighted Research: CT Panda	319
Other Diagnostic Uses	320
FDA-Approved AI-Related Medical Devices	321
Administration and Care	323
Highlighted Research: MedAlign	323

公共データにアクセスする

概要

今年のAIインデックスは、科学と医療におけるAIの重要性の高まりを認識し、AIが科学および医療の発見において果たす役割に関する新しい章を紹介します。この章では、GraphCastのような高度な天気予報システムや、GNoMEのような改善された材料発見アルゴリズムなど、2023年に際立ったAIを活用した科学的成果を探ります。また、医療AIシステムの性能、SynthSRやImmunoSEIRAのような2023年の重要なAI駆動の医療革新、FDAのAI関連医療機器の承認におけるトレンドについても検討します。

章のハイライト

1. 科学の進歩はAIのおかげでさらに加速しています。2022年にはAIが科学的発見を促進し始めました。しかし2023年には、アルゴリズムのソートをより効率的に行うAlphaDevから、材料発見プロセスを促進するGNoMEに至るまで、より重要な科学関連のAIアプリケーションが開始されました。

2. AIは医学において重要な進展をもたらしています。2023年には、パンデミック予測を強化するEVEscapeや、AI駆動の変異分類を支援するAlphaMissenseを含むいくつかの重要な医療システムが導入されました。AIはますます医学の進歩を推進するために利用されています。

3. 高度に知識を持つ医療AIが登場しました。過去数年の間に、AIシステムは医療知識を評価する重要なテストであるMedQAベンチマークで驚異的な改善を示しました。2023年の突出したモデルであるGPT-4 Medpromptは、90.2%の正確性を達成し、2022年の最高得点から22.6ポイントの増加を示しました。このベンチマークが2019年に導入されて以来、MedQAでのAIのパフォーマンスはほぼ3倍に増加しています。

4. FDAはますます多くのAI関連医療機器を承認しています。2022年には、FDAは139のAI関連医療機器を承認し、2021年から12.1%の増加となりました。2012年以来、FDAに承認されたAI関連医療機器の数は45倍以上に増加しています。AIは現実の医療目的にますます利用されています。

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関連の科学的ブレイクスルーを強調しています。

5.1 注目すべき科学的マイルストーン

アルファデブ

アルファデブはより高速なソートアルゴリズムを発見します

AlphaDevは、計算アルゴリズムの強化分野での科学者やエンジニアの数十年の研究を基に改善された新しい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指示が少ないアルゴリズムを開発しました。

短い列に対する基本的なソートアルゴリズム、例えば Sort 3、Sort 4、Sort 5（図5.1.1）。AlphaDevによって発見された新しいアルゴリズムのいくつかは、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、10年以上にわたるこのライブラリの部分への最初の更新であり、強化学習を使用して設計された最初の追加です。

アルファデブ対人間のベンチマーク、アルゴリズムの長さ最適化時

Source: Mankowitz et al., 2023 | Chart: 2024 AI Index report

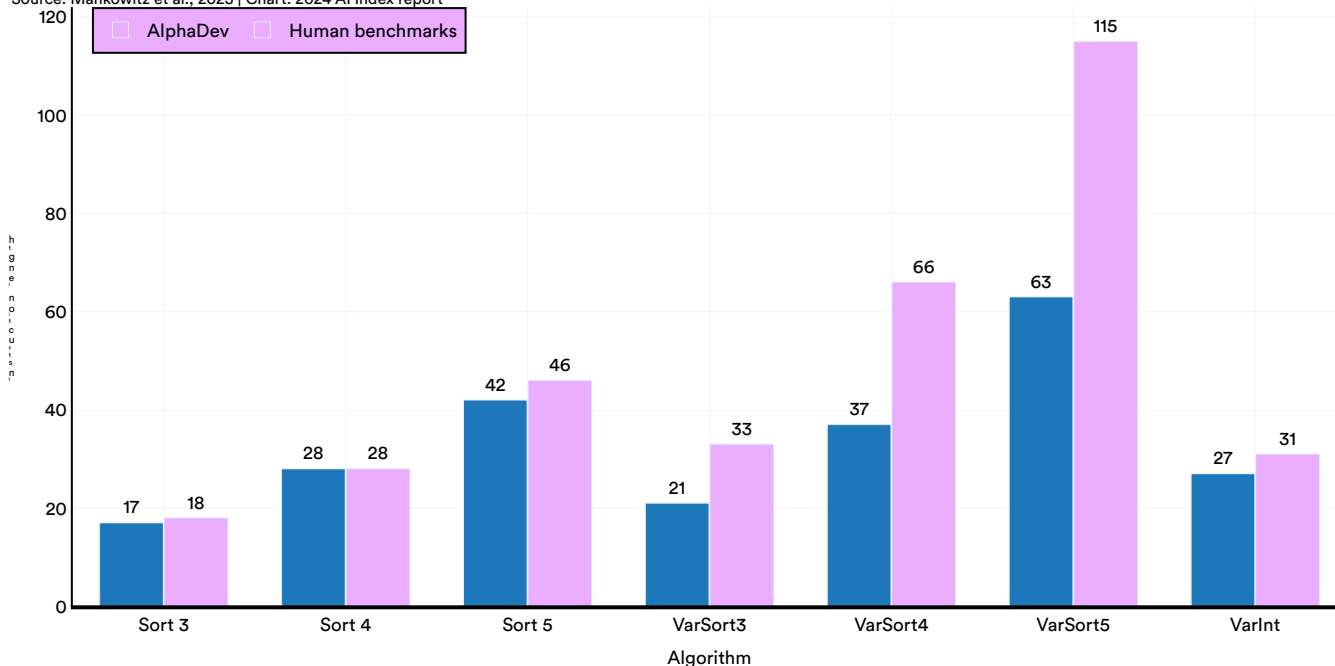


Figure 5.1.1

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ーグラフィックスにおいて重要で、3Dオブジェクトを定義するための頂点、エッジ、面のメッシュを作成することを含みます。これは、ビデオゲーム、アニメーション、医療画像、科学的可視化において重要です。伝統的な等値面抽出アルゴリズムは、限られた解像度、構造的剛性、数値的不安定性に苦しむことが多く、それが後に影響を与えます。

品質。FlexiCubesは、勾配ベースの最適化と適応可能なパラメーターのためにAIを使用することで、これらの制限のいくつかに対処しています（図5.1.2）。この方法は、正確で局所的なメッシュ調整を可能にします。メッシュ再構築のために微分可能な等値面を利用する他の主要な方法と比較して、FlexiCubesは基礎となる真実に非常に密接に一致するメッシュ抽出を実現します（図5.1.3）。

サンプルフレキシキューブの表面再構築

Source: Nvidia, 2023

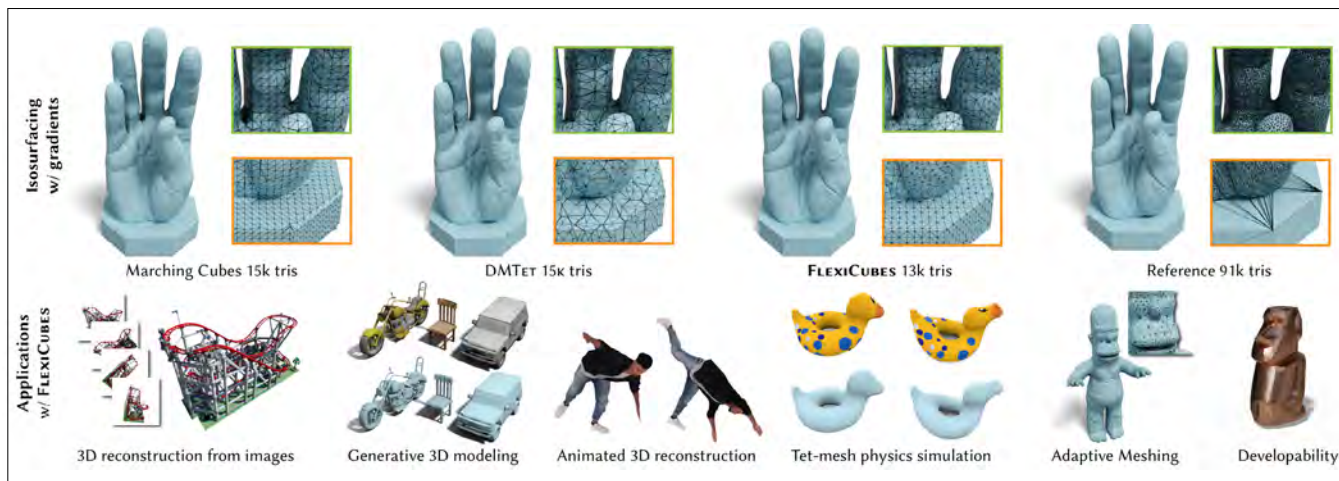


図5.1.2

3Dメッシュ再構築に関する定量的結果を選択する

Source: Shen et al., 2023 | Chart: 2024 AI Index report

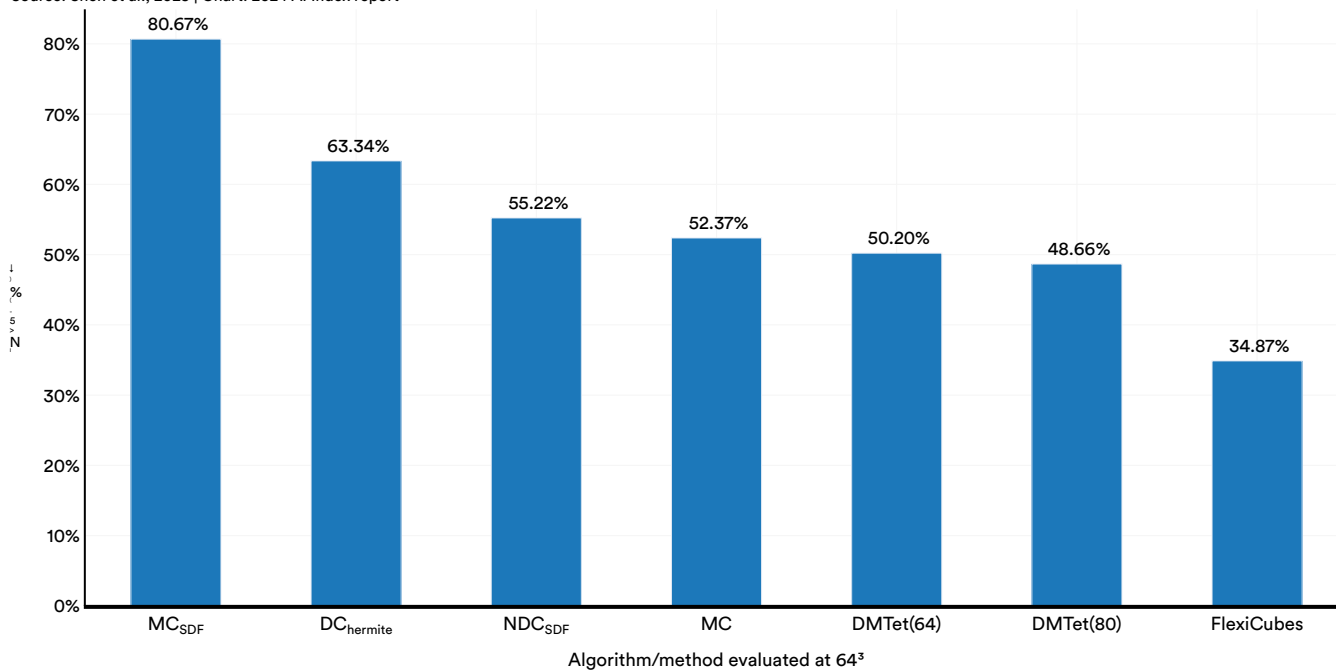


図5.1.3

シンボット

有機分子を合成するためのAI駆動
ロボット化学者

Synbotは、化学合成計画のためのAIソフトウェア層、命令を翻訳するためのロボットソフトウェア層、実験を実施するための物理ロボット層で構成された多層システムを採用しています。AIとロボットシステム間の閉ループフィードバックメカニズムにより、Synbotは確立された参照と同等または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です（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]の合成を目指した実験において、Synbotは変換収率を超える複数の合成式を開発しました。

Synbotデザイン

Source: Ha et al., 2023

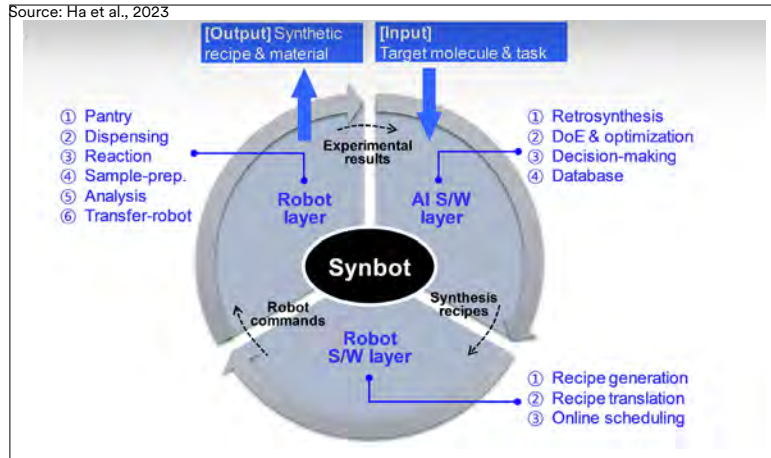


図5.1.4

1980年代中頃の参照範囲で合成を完了し、はるかに短時間でを行いました（図5.1.5）。Synbotの有機合成の自動化は、製薬や材料科学などの分野におけるAIの可能性を強調しています。

M1自律最適化実験の反応速度論、Synbot対参照

Source: Ha et al., 2023 | Chart: 2024 AI Index report

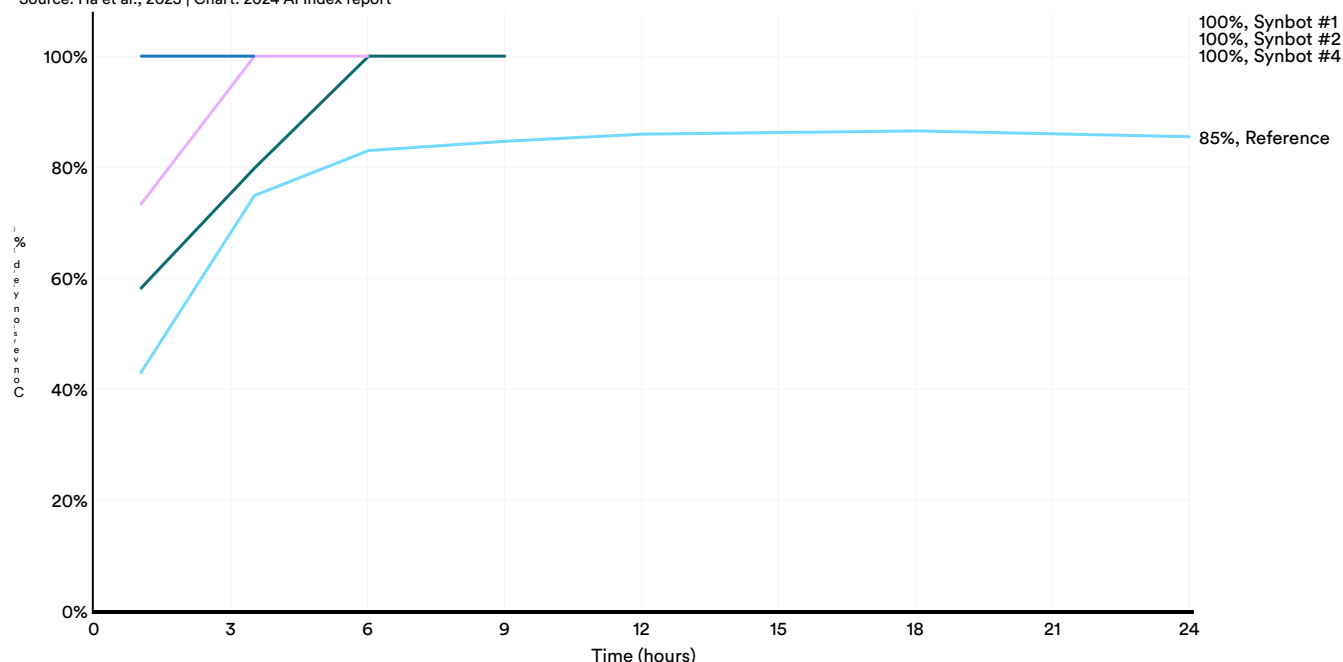


図5.1.5

GraphCast

より正確なグローバル天気予報

GraphCastと共に

GraphCastは、新しい天気予報システムで、1分以内に高精度な10日間の天気予報を提供します（図5.1.6）。グラフニューラルネットワークと機械学習を利用して、GraphCastは膨大なデータセットを処理し、温度、風速、気象条件を予測します。

およびその他。図5.1.7は、GraphCastの性能を現在の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測（HRES）と比較しています。GraphCastは、平方根平均二乗誤差が低いため、予測が観測された気象パターンにより近いことを示しています。GraphCastは、気象パターンを解読し、極端な気象イベントに対する準備を強化し、世界的な気候研究に貢献するための貴重なツールとなるでしょう。

GraphCast天気予報

Source: DeepMind, 2023

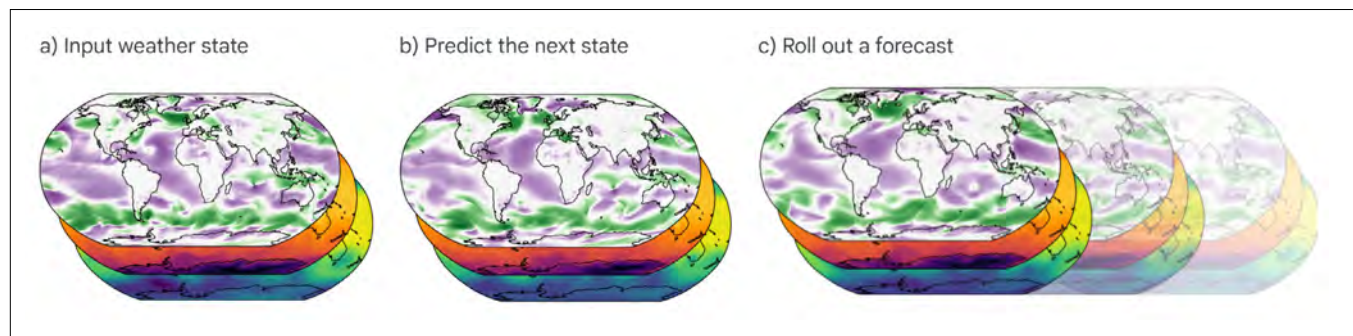


Figure 5.1.6

10日間z500予測スキル: GraphCast対HRES

Source: Lam et al., 2023 | Chart: 2024 AI Index report

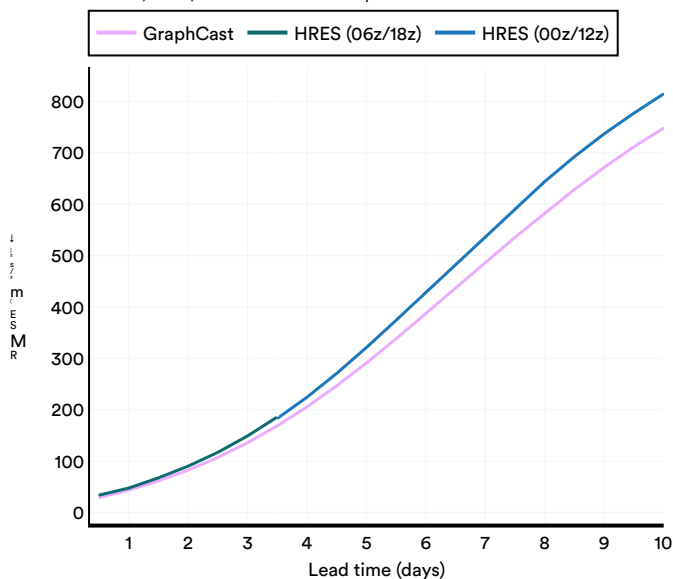


Figure 5.1.7

GNoME

GNoMEによる新材料の発見

新しい機能材料の探索は、ロボティクスや半導体製造など、さまざまな科学分野での進歩にとって重要です。しかし、この発見プロセスは通常、高コストで時間がかかります。最近、Googleの研究者による進展は、AIモデルの一種であるグラフネットワークが、大規模データセットで訓練されると、このプロセスを迅速化できることを示しました。彼らのモデルであるGNoMEは、材料発見の先駆的手法であるMaterials Projectを上回り、著しく多くの安定した結晶を特定しました（図5.1.8）。GNoMEは220万の新しい結晶構造を明らかにし、多くは人間の研究者によって見落とされてきました（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、科学的なブレークスルーを速める上でのデータとスケーリングの力を浮き彫りにしています。

サンプル材構造

Source: Merchant et al., 2023

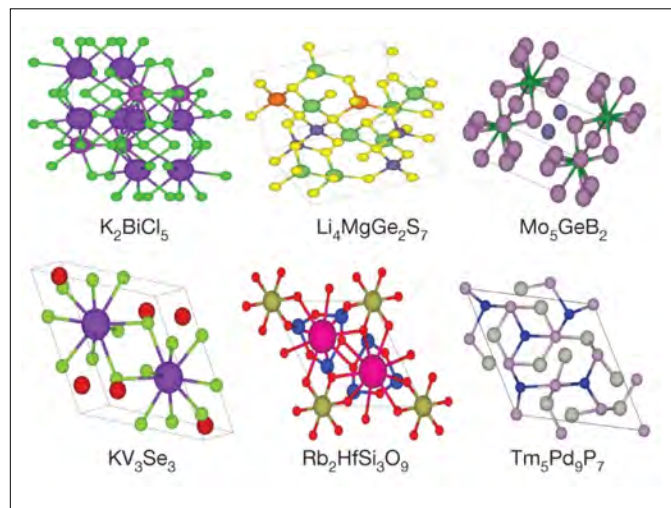


図5.1.8

GNoME vs. 材料プロジェクト: 安定した結晶数

Source: Merchant et al., 2023 | Chart: 2024 AI Index report

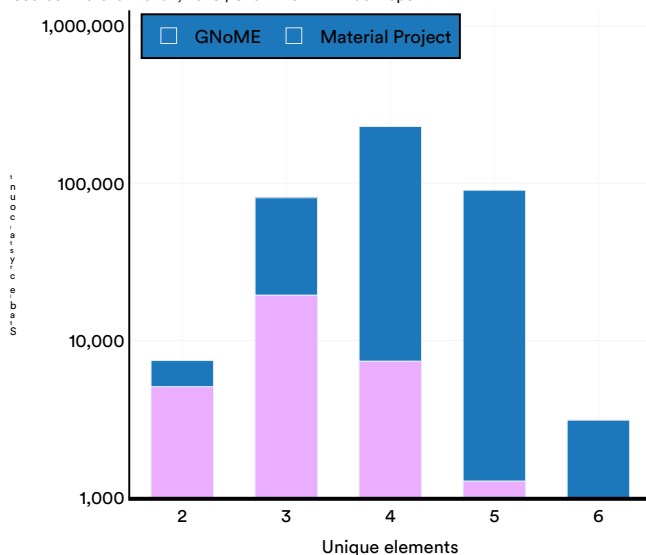


図5.1.9

GNoME 対材料プロジェクト: 異なるプロトタイプ

Source: Merchant et al., 2023 | Chart: 2024 AI Index report

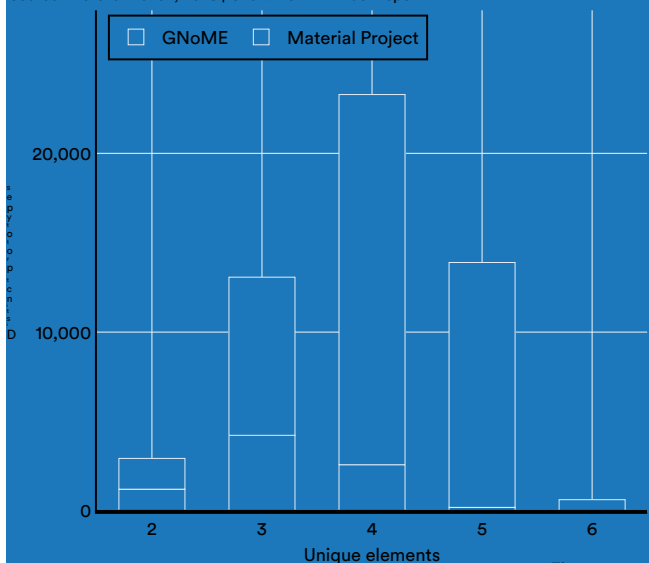


Figure 5.1.10

洪水予測

より正確で信頼性のある洪水予測のためのAI

2023年に発表された新しい研究は、大規模な洪水イベントの予測において重要な進展を遂げました。洪水は最も一般的な自然災害の一つであり、特に防止や緩和のためのインフラが不足している発展途上国では、その影響が特に深刻です。そのため、これらのイベントをより早く予測できるより正確な予測方法を開発することは、かなりの良い影響をもたらす可能性があります。

Googleの研究者チームは、AIを利用して非常に正確な水文学シミュレーションモデルを開発しました。これらの革新的な手法は、無計測流域にも適用可能で、特定の極端な洪水イベントを最大で5日前に予測でき、その精度はGloFASなどの現在の最先端モデルに匹敵するか、それを上回ります。AIモデルは、さまざまな復帰期間イベントにおいて、優れた精度（正の予測の精度）とリコール（すべての関連インスタンスを正しく特定する能力）を示し、最も先進的な現代的手法を凌駕しています（図5.1.11）。このモデルはオープンソースであり、既に80カ国以上で洪水イベントの予測に使用されています。

AIモデルの予測とGloFASの返戻期間における比較

Source: Nearing et al., 2023 | Chart: 2024 AI Index report

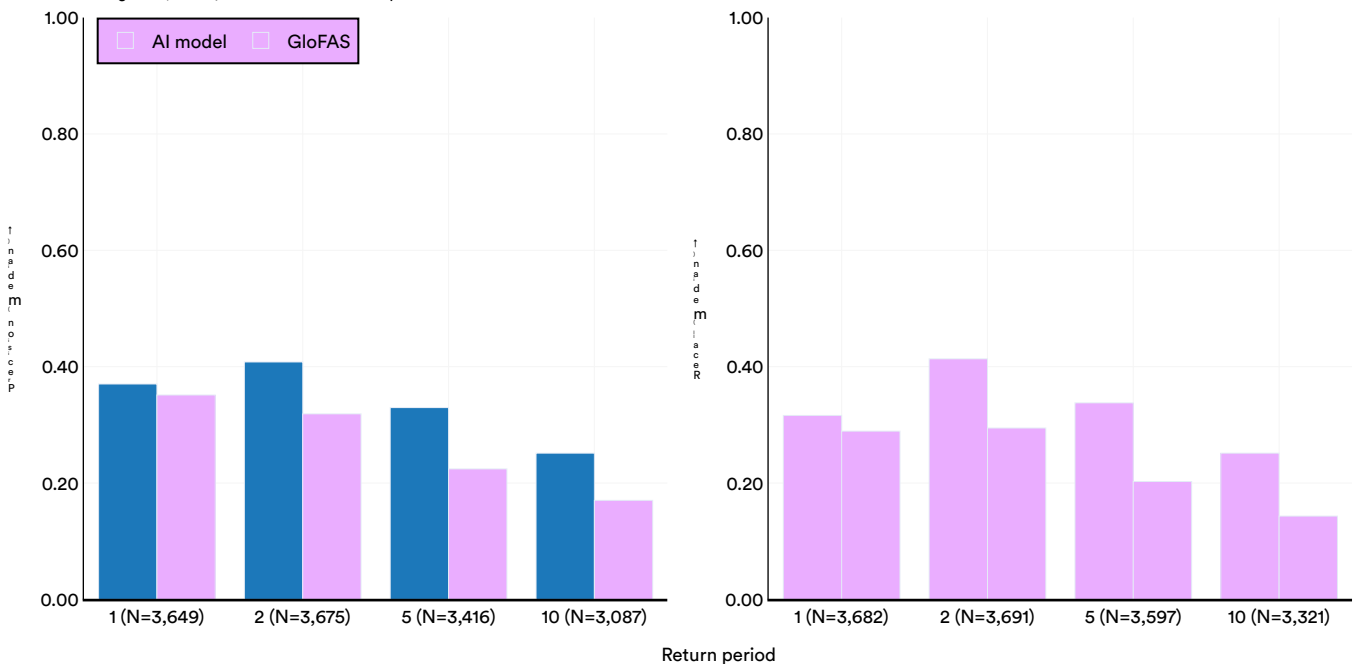


図5.1.11

1 An ungauged basin is a watershed for which there is insufficient streamflow data to model hydrological flows.

2 A return period (recurrence interval) measures the likelihood of a particular hydrological event recurring within a specific period. For example, a 100-year flood means there is a 1% chance of the event being equaled or exceeded in any given year.

AIモデルは医療分野でますます重要性を増しており、ポリープの検出から臨床医の診断支援まで多くの応用があります。AIの性能が向上し続ける中で、医療実践に対するその影響を監視することがますます重要になっています。このセクションでは、2023年に導入された重要なAI関連医療システム、臨床AIの知識の現状、病院管理の向上を目指した新しいAI診断ツールと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

著名な医療システム

このセクションでは、AIインデックス運営委員会によって選ばれた2023年の重要なAI関連医療のブレイクスルーを特定します。

SynthSR

脳スキャンを高度な分析のために変換する SynthSR は、臨床用の脳スキャンを高解像度の T-1 加重画像に変換する AI ツールです (図 5.2.1)。この進歩は、以前は高度な研究において多くのスキャンの使用を制限していたスキャン品質の変動性の問題に対処します。これらのスキャンを、コントラストが高く脳の構造を明確に描写する T1 加重画像に変換することで、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、スキャンレベルおよび被験者レベルの両方で観察されたボリュームとの強い相関が示され、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似た画像を生成していることを示唆しています。図 5.2.2 は、選択された脳領域における SynthSR スキャンが実際の観測とどの程度一致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、神経科学の研究や臨床診断を容易にします。

SynthSR 世代

Source: Iglesias et al., 2023

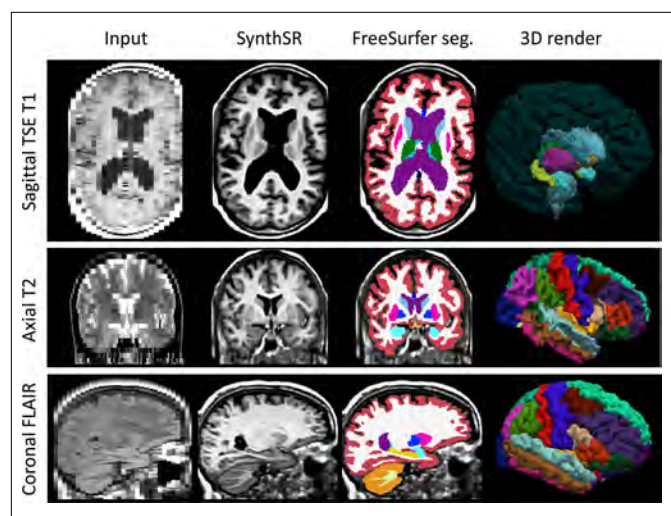


Figure 5.2.1

特定の脳領域におけるSynthSRと真実のボリュームとの相関関係

Source: Iglesias et al., 2023 | Chart: 2024 AI Index report

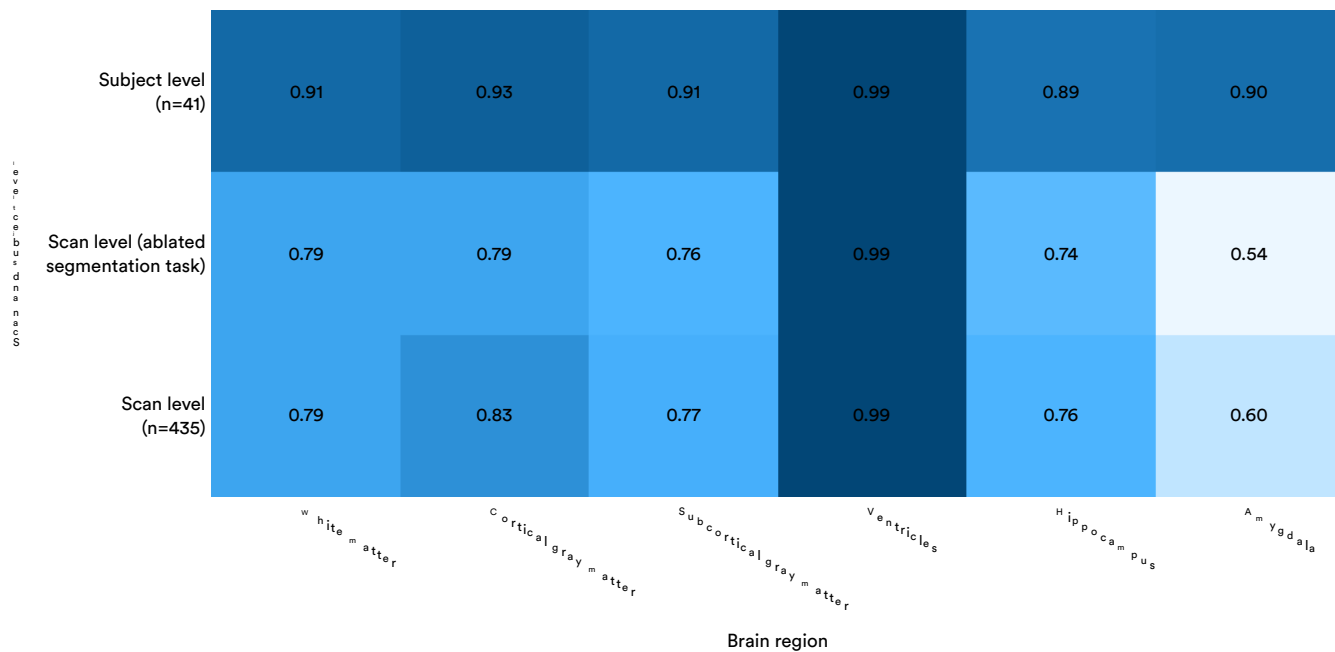


Figure 5.2.2

結合プラズモニク赤外線センサー

神経変性疾患の検出のための結合プラズモニク赤外線センサー

パーキンソン病やアルツハイマー病のような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特定に依存しています。従来の方法、例えば質量分析法やELISAは、タンパク質レベルの定量に役立ちますが、構造状態の変化を見分けることはできません。今年、研究者たちは、表面増強赤外吸収（SEIRA）分光法を用いたプラズモニク赤外線センサーと免疫アッセイ技術（ImmunoSEIRA）を組み合わせた神経変性疾患の診断のための新しい方法を発見しました。実際のフィブリル比率とAIシステムによって行われた予測を比較したテストにおいて、予測の精度は実際に報告された比率と非常に密接に一致することがわかりました。

ImmunoSEIRA検出原理とセットアップ

Source: Kavungal et al., 2023

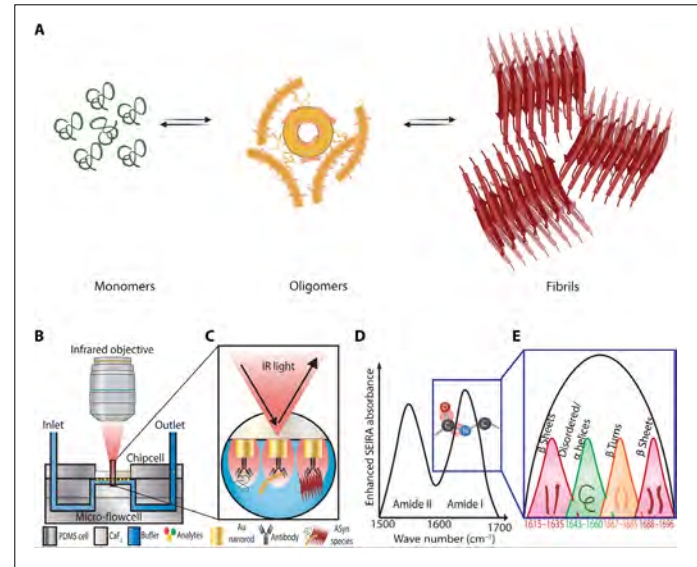


図5.2.3

テストサンプルにおける深層ニューラルネットワークの予測値対実際のブリルパーセンテージ

Source: Kavungal et al., 2023 | Chart: 2024 AI Index report

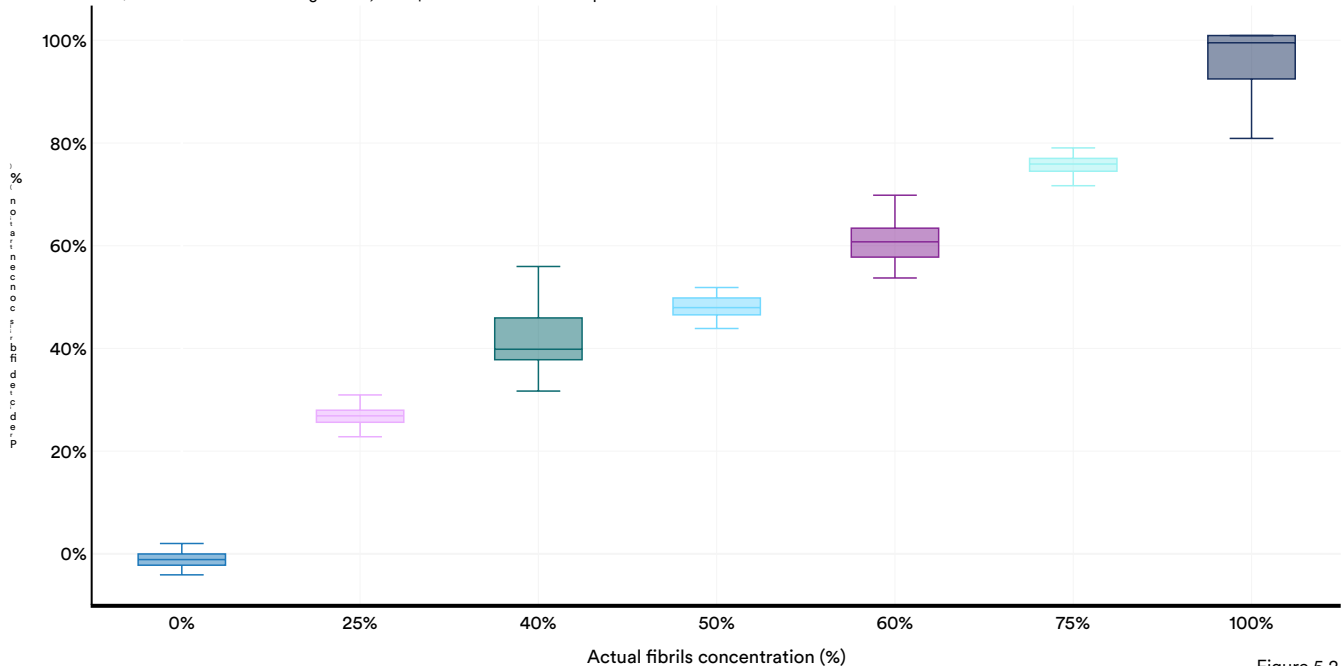


Figure 5.2.4

EVEscape

パンデミック準備のためのウイルス進化の予測

ウイルス変異の予測は、ワクチン設計とパンデミックの最小化にとって重要です。従来の手法は、リアルタイムのウイルス株と抗体データに依存しており、データの不足のためにパンデミックの初期段階で課題に直面します。EVEscapeは、歴史的な配列と生物物理的および構造的情報に基づいて訓練された新しいAI深層学習モデルであり、進化を予測します。

ウイルスについて（図5.2.5）。EVEscapeは、現在の変異株データに依存せずにウイルスの逃避を評価し、観察されたSARS-CoV-2の変異の50.0%を予測しています。これは、46.2%および32.3%を予測した従来の実験室研究を上回っており、さらに、以前のモデルは変異のわずか24%しか予測していませんでした（図5.2.6）。このパフォーマンスは、将来のパンデミックの準備と対応努力を強化するための貴重な資産としてのEVEscapeの可能性を際立たせています。

EVEスケープデザイン

Source: Thadani et al., 2023

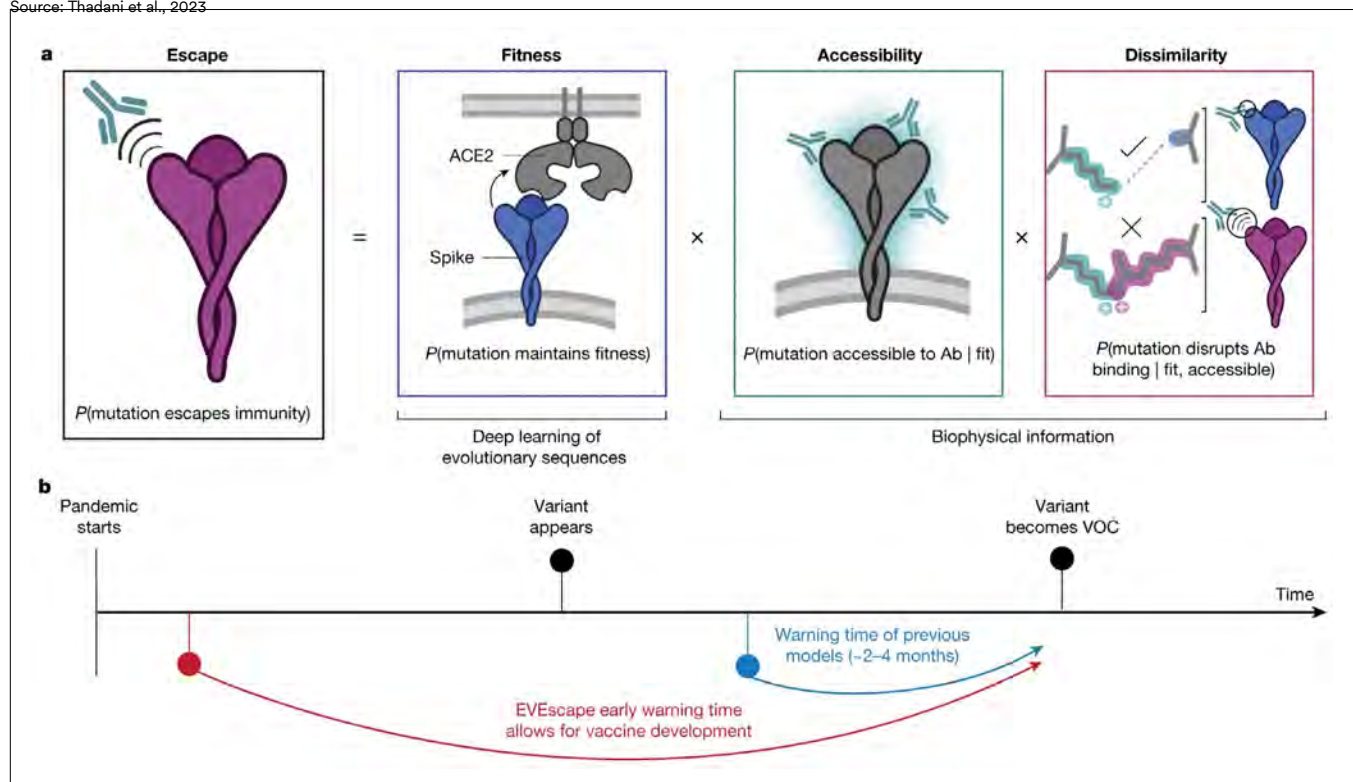


図5.2.5

EVEscapeと他のモデルのSARS-CoV-2 RBD変異予測に関する比較

Source: Thadani et al., 2023 | Chart: 2024 AI Index report

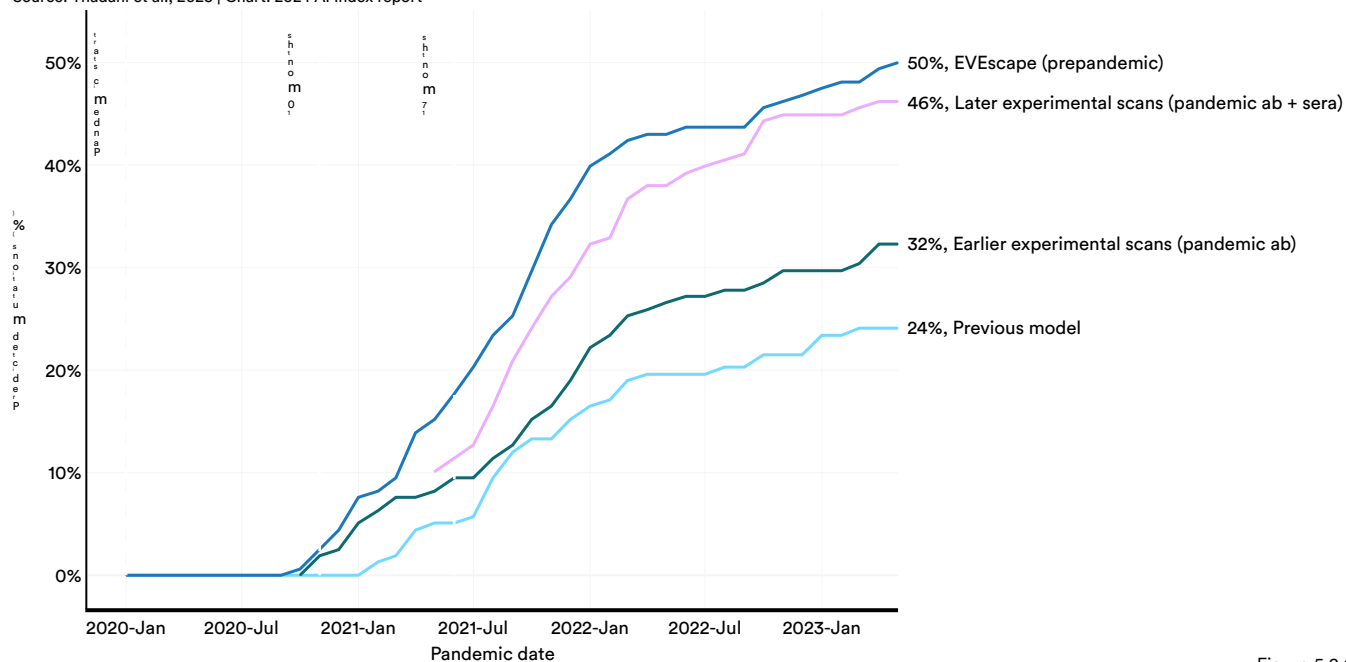


Figure 5.2.6

アルファミスセンス

AIの変異のより良い分類

科学者たちは、どの遺伝子変異が病気を引き起こすかをまだ完全には理解していません。何百万もの可能な遺伝子変異がある中で、変異が良性か病原性かを判断するには、労力を要する実験が必要です。

2023年、Google DeepMindの研究者たちは、7100万ものミスセンス変異の病原性を予測する新しいAIモデル「AlphaMissense」を発表しました。ミスセンス変異は人間のタンパク質の機能に影響を与える遺伝的变化であり、がんを含むさまざまな病気を引き起こす可能性があります。7100万の可能なミスセンス変異のうち、AlphaMissenseは89%を分類し、そのうち57%をおそらく良性、32%をおそらく病原性と特定し、残りは不確定としてカテゴライズしました。それに対して、人間のアノテーターはすべてのミスセンス変異のうち、0.1%の性質を確認することしかできませんでした。

ヘモグロビンサブユニットベータ (HBB)

Source: Google DeepMind, 2023



図5.2.7

アルファミスセンスの予測

出典：グーグルディープマインド、2023 | チャート：2024年AIインデックスレポート

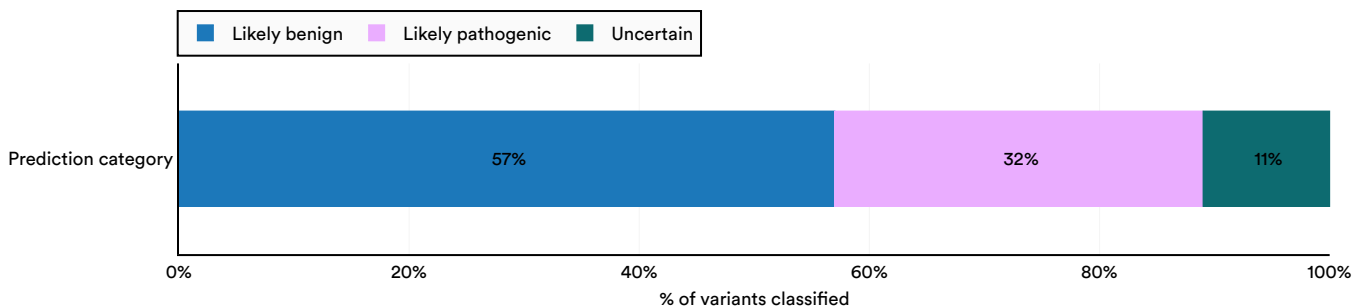


Figure 5.2.8

ヒトパンゲノムリファレンス

人間のゲノムをマッピングするためにAIを使用するヒトゲノムは人間のための分子指示のセットです。最初のヒトゲノムドラフトは2000年に公開され、2022年に更新されました。しかし、更新はやや不完全でした。血液型のようなさまざまな遺伝子変異を取り入れておらず、多様な祖先集団を完全にはマッピングしていませんでした。したがって、現在のゲノム参照のもとでは、特定の人々の中で病気を検出したり治療法を見つけたりするのが難しいでしょう。

2023年に、人類パンゲノム研究コンソーシアムは、60の機関からなる119人の科学者が参加し、AIを使用して更新されたより代表的なヒトゲノムマップ（図5.2.9）を開発しました。研究者たちは、タンパク質コーディング遺伝子の中央値99.07%、タンパク質コーディング転写物の99.42%、ノンコーディング遺伝子の98.16%、ノンコーディング転写物の98.96%を注釈付けするという驚異的な精度を達成しました。詳細は図5.2.10に示されています。

ゲノムのMHC領域のグラフゲノム

Source: Google Research, 2023

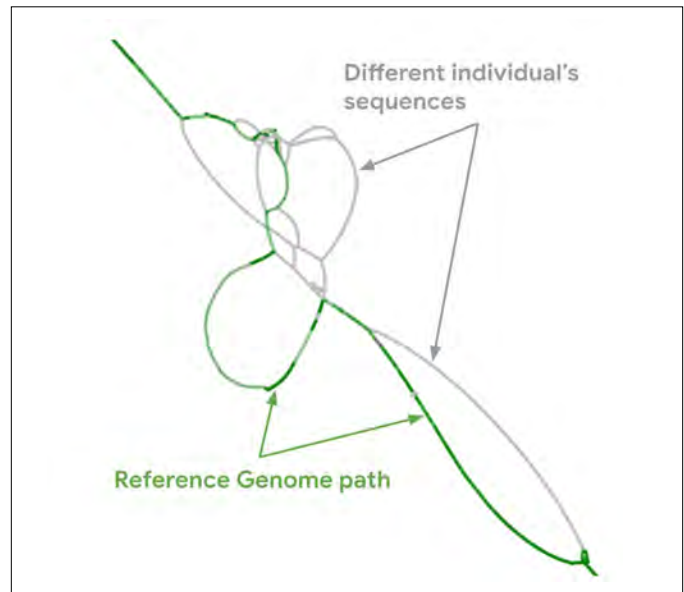


図5.2.9

この最新のゲノムバージョンは、これまでで最も包括的で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
結果

Source: Liao et al., 2023 | Chapter 5.2 AI Index report

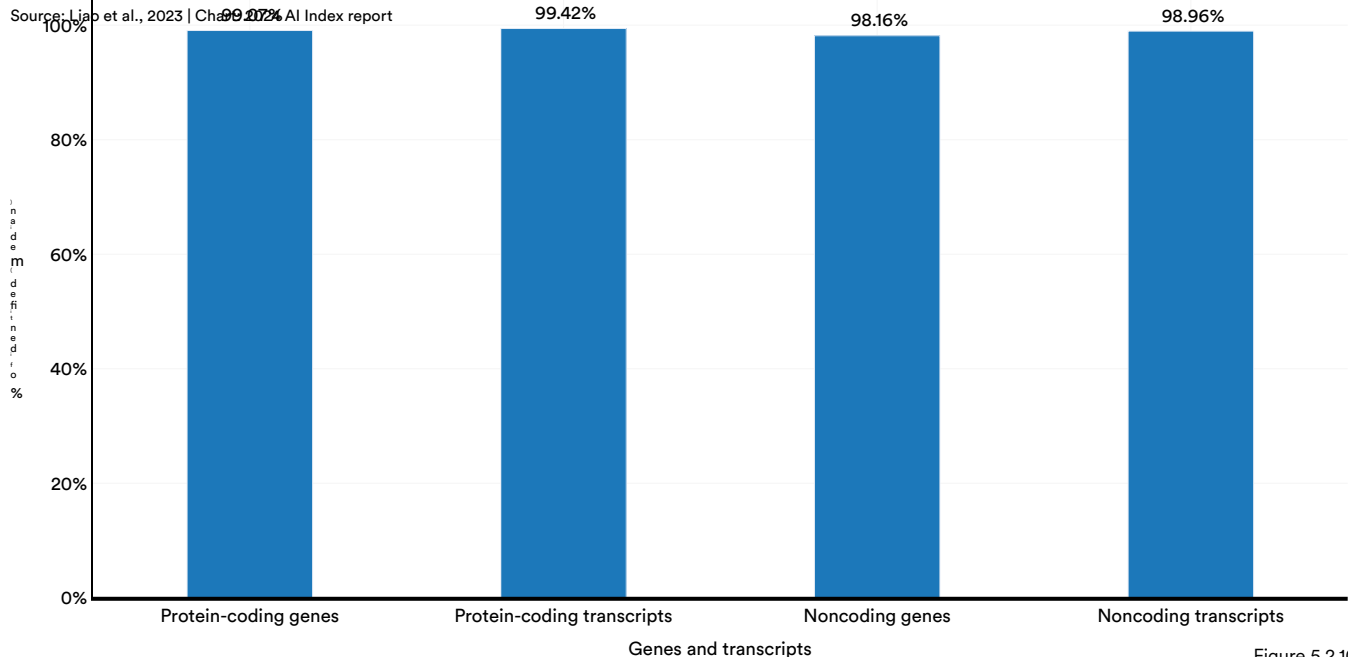


Figure 5.2.10

臨床知識

AIモデルの臨床知識を評価することは、特に臨床環境で適用可能な医療専門知識の程度を判断することを含みます。

著しい改善が見られ、先進的なシステムであるGPT-4 Medpromptが90.2%の精度を達成しました。これは2022年の最高得点から22.6パーセントポイントの向上です（図5.2.11）。MedQAの創設以来、このベンチマークに対するAIの能力はほぼ3倍になり、臨床知識を持つAIシステムの急速な進展を示しています。

メドQA

2020年に導入されたMedQAは、専門の医療ボード試験から派生した包括的なデータセットで、医師に挑戦をもたらす60,000以上の臨床問題を特徴としています。

MedQAベンチマークにおけるAIの性能は見られています。

MedQA: 正確性

Source: Papers With Code, 2023 | Chart: 2024 AI Index report

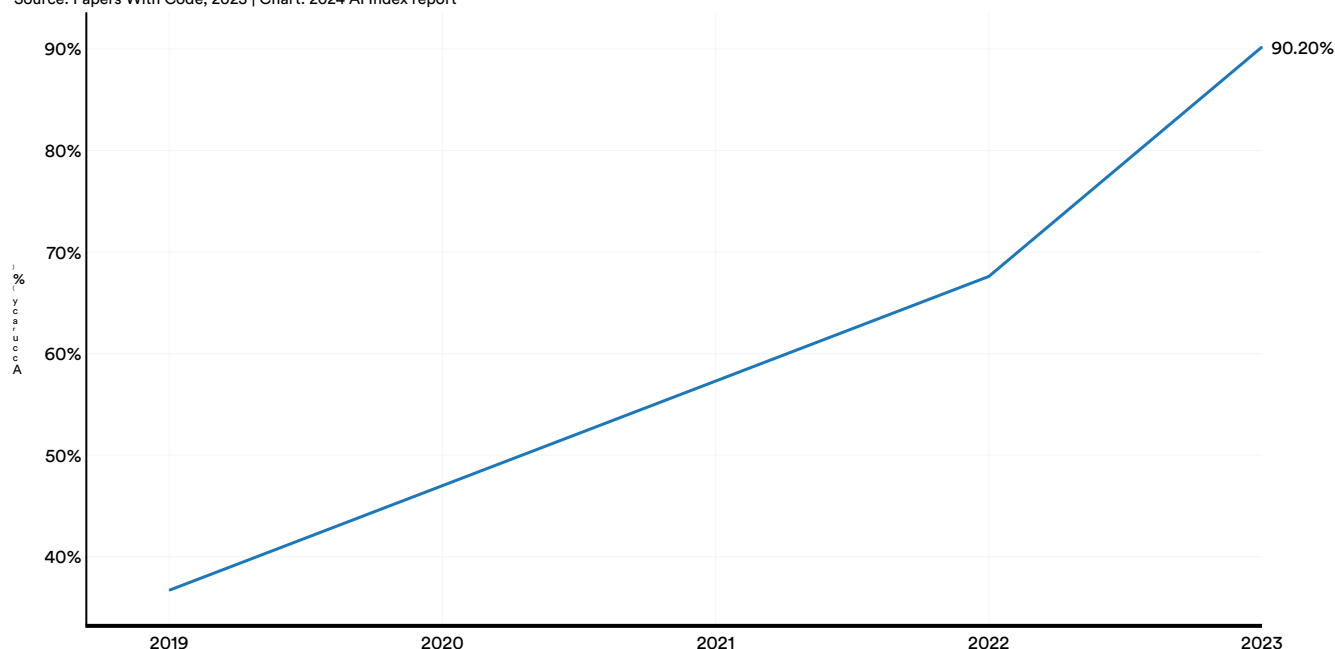


図5.2.11

強調された研究:

GPT-4メドプロンプト

LLMは印象的な一般知識を示しますが、医療関連の質問に答えるなどの専門知識で優れたパフォーマンスを発揮するためには、重要なファインチューニングが必要だと一般的に考えられています。ファインチューニングとは、特定の分野のデータでLLMを訓練することを意味します。

2023年末のマイクロソフトの研究により、この仮定は覆されました。この研究では、プロンプトエンジニアリングを用いてGPT-4を指向させ、MultiMed QAベンチマークスイートで驚異的な性能を達成しました。これは、4つの挑戦的な医療ベンチマークのグループです（図5.2.12）。GPT-4 Medpromptは、PubMedQA、MedMCQA、MMLUなどの著名な医療ベンチマークの選択肢の複数セクションで、2022年のトップモデルFlan-PaLM 540Bをそれぞれ3.0、21.5、16.2パーセントポイント上回りました。また、当時の最先端のMed-PaLM 2の性能も超えました（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%以上の精度を達成した最初のものでした。このブレイクスルーは、GPT-4 Medpromptの優れた、そして潜在的に臨床的に有用な医療能力を強調するだけでなく、特化した領域にモデルを適応させるためにファインチューニングが常に必要ではないことを示しています。プロンプトエンジニアリングは、有望な代替戦略として示されています。


GPT-4対Med-PaLM 2が医療質問に回答する

Source: Nori et al., 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

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

Expert

GPT-4 generated CoT

 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.

GPT-4

図5.2.12

強調された研究:

GPT-4メディプロンプト（続き）

MultiMedQAサブベンチマーク

Source: Nori et al., 2023 | Chart: 2024 AI Index reportにおけるモデルのパフォーマンス

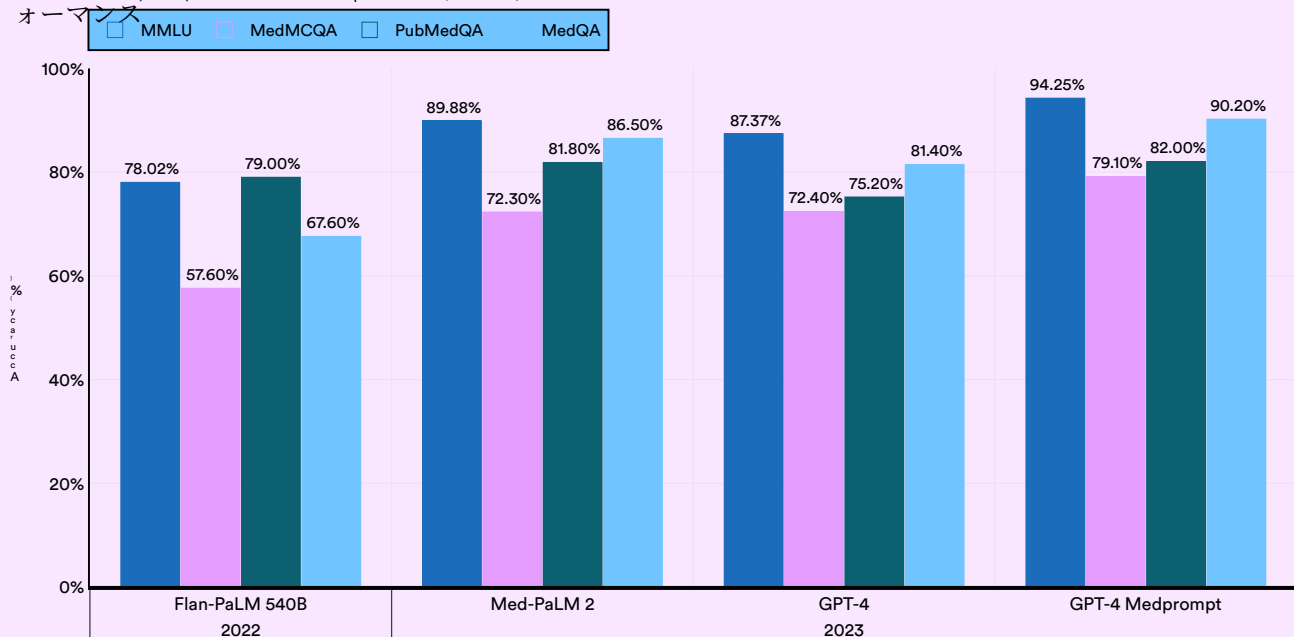


図5.2.13

強調された研究:

MediTron-70B

GPT-4 Medpromptは印象的なシステムですが、クローズドソースであるため、その重みは一般の人々が自由に利用できるものではありません。2023年の新しい研究では、オープンソースの医療LLMの能力を向上させることを目指しています。この新しい研究の中で、MediTron-70Bは特に有望だと際立っています。このモデルはMedQAベンチマークで70.2%の respectableな精度を達成しています。これはGPT-4 MedpromptやMedのパフォーマンスには及びませんが、

PaLM 2（両方のクローズドモデル）は、2023年の最先端の成果に対して重要な改善を示しており、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におけるスコアは、オープンソースモデルによってこれまでに達成された中で最高です。医療AIがその最大の可能性を達成するためには、その機能が広くアクセス可能であることが重要です。この文脈において、MediTronは前向きな進展を示しています。

MedQAにおける選択モデルのパフォーマンス

出典：Chen et al., 2023 | 表：2024年AIインデックス報告

Model	Release date	Access type	Score on MedQA
GPT-4 Medprompt	November 2023	Closed	90.20%
Med-PaLM 2	April 2023	Closed	86.20%
MediTron-70B	November 2023	Open	70.20%
Med-PaLM	December 2022	Closed	67.20%
Llama 2	July 2023	Open	63.80%

図5.2.14

診断

AIツールは、放射線科や癌検出などの診断目的にも使用できます。

強調された研究:

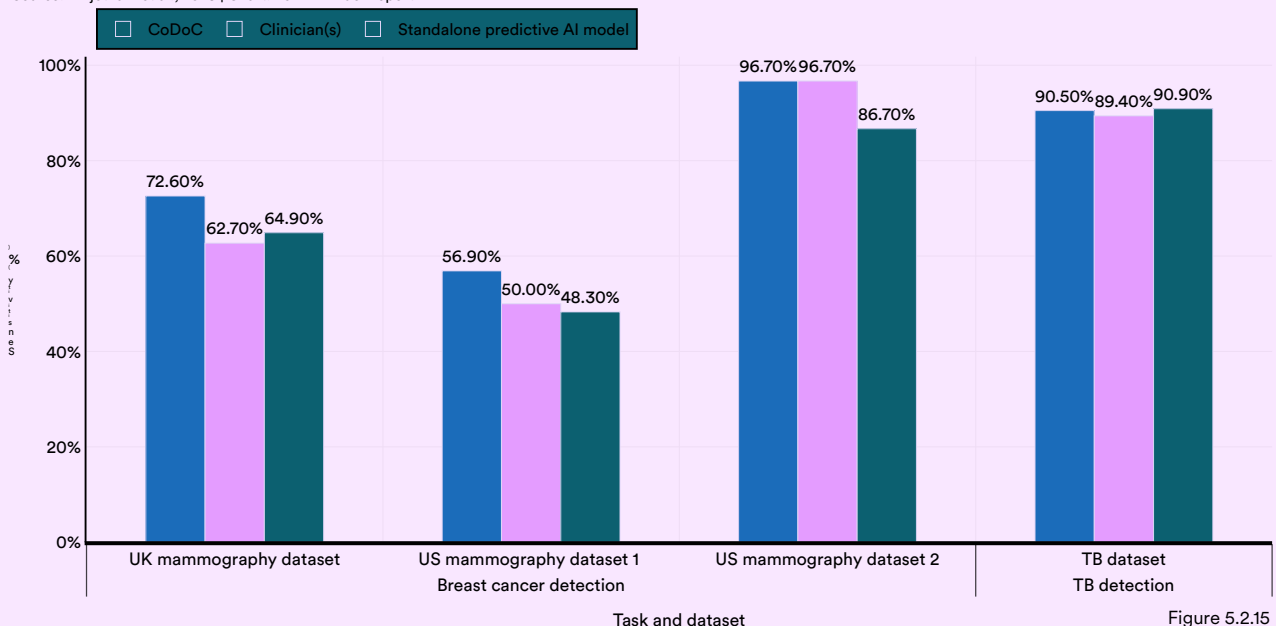
CoDoC

AI医療画像システムは強力な診断能力を示していますが、臨床医が見逃す診断やその逆のケースも存在します。この観察は、AIシステムと臨床医の診断能力の論理的統合を示唆しています。2023年に、研究者たちはCoDoC（Complementarity-Driven Deferral to Clinical Workflow）を発表しました。これは、診断のためにAIに依存すべきか、伝統的な臨床手法に委ねるべきかを識別するために設計されたシステムです。CoDoCは、感度（病気のあ

（それを持たない人々を正確に特定する能力）。具体的には、4つの医療データセット全体で、CoDoCの感度は臨床医の平均を4.5パーセントポイント上回り、独自の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を平均2.7パーセントポイント上回り、独自の予測モデルを5.7パーセントポイント上回っています。さらに、CoDoCは臨床業務の負担を66%軽減することが示されています。これらの結果は、AI医療システムが臨床業務に統合され、診断の正確性と効率が向上する可能性があることを示唆しています。

CoDoC対単独予測AIシステムおよび臨床読影者：感度

Source: Dvijotham et al., 2023 | Chart: 2024 AI Index report



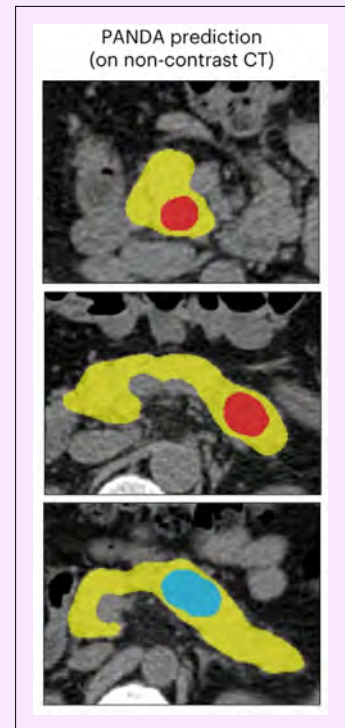
強調された研究:

CTパンダ

膵臓癌（PDAC）は特に致死的な癌であり、外科的介入が行われるには遅すぎることが多く、早期発見が難しい。無症状の人々におけるPDACのスクリーニングは、その低い有病率と偽陽性のリスクのために挑戦的である。今年、中国の研究チームはPANDA（人工知能による膵臓癌検出）を開発した。これは、X線で膵臓病変を効率的に検出し、分類することができるAIモデルである。検証テストでは、PANDAは感度で平均放射線科医を34.1%、特異度で6.3%上回った。約20,000人の患者を対象とした大規模な実世界テストでは、PANDAは感度92.9%、特異度99.9%を達成した。PANDAのようなAI医療ツールは、困難な病状の診断において重要な進歩を示しており、以前は難しいまたは高コストと考えられていたコスト効果的かつ正確な検出を提供している。

パンダ検出

Source:
Cao et al., 2023
図5.2.16



出典: Cao et al., 2023 | チャート: 2024年AI指数レポート PANDAと平均放射線科医の多施設バリデーション（6,239人の患者）

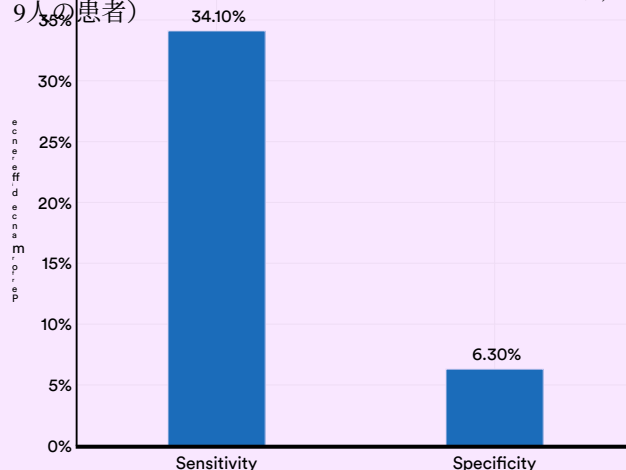


Figure 5.2.17

PANDAの実世界多シナリオ検証におけるパフォーマンス（20,530人の患者）

Source: Cao et al., 2023 | Chart: 2024 AI Index report

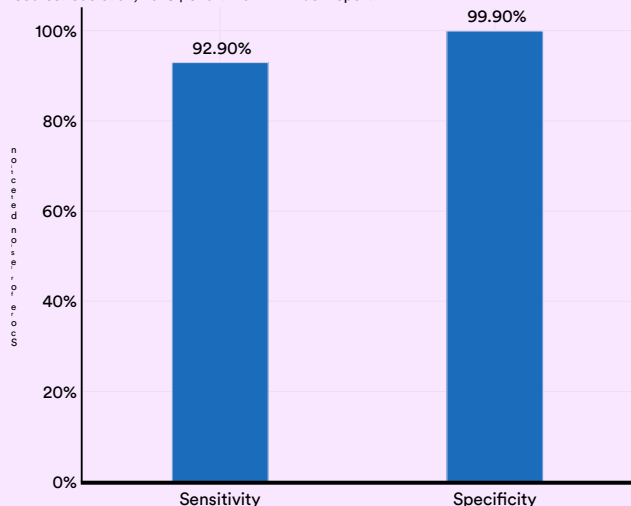


Figure 5.2.18

その他の診断用途

2023年に発表された新しい研究は、AIが他の診断文脈でどのように使用できるかを強調しています。図5.2.19は、いくつかの発見を要約しています。

診断AIの使用事例に関する追加研究

Source: AI Index, 2024

Research	Use case	Findings
Schopf et al., 2023	Breast cancer	The authors conducted a meta-review of the literature exploring mammography-image-based AI algorithms. They discovered that predicting future breast cancer risk using only mammography images achieves accuracy that is comparable to or better than traditional risk assessment tools.
Dicente Cid et al., 2023	X-ray interpretation	The researchers developed two open-source neural networks, X-Raydar and X-Raydar-NLP, for classifying chest X-rays using images and free-text reports. They found that these automated classification methods perform at levels comparable to human experts and demonstrate robustness when applied to external data sets.

Figure 5.2.19

FDA承認のAI関連医療機器

アメリカ食品医薬品局（FDA）は、承認を受けたAI/ML対応の医療機器のリストを維持しています。このリストに掲載されている機器は、FDAの市場前基準を満たしており、その効果と安全性の詳細なレビューが含まれています。2023年10月現在、FDAは生成AIを利用したり、LLMによって動作するいかなる機器も承認していません。

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件のAI関連医療機器がFDAの承認を受け、2021年の承認件数から12.1%の増加となりました。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によって承認されたAI医療機器の数、2012–22年

Source: FDA, 2023 | Chart: 2024 AI Index report

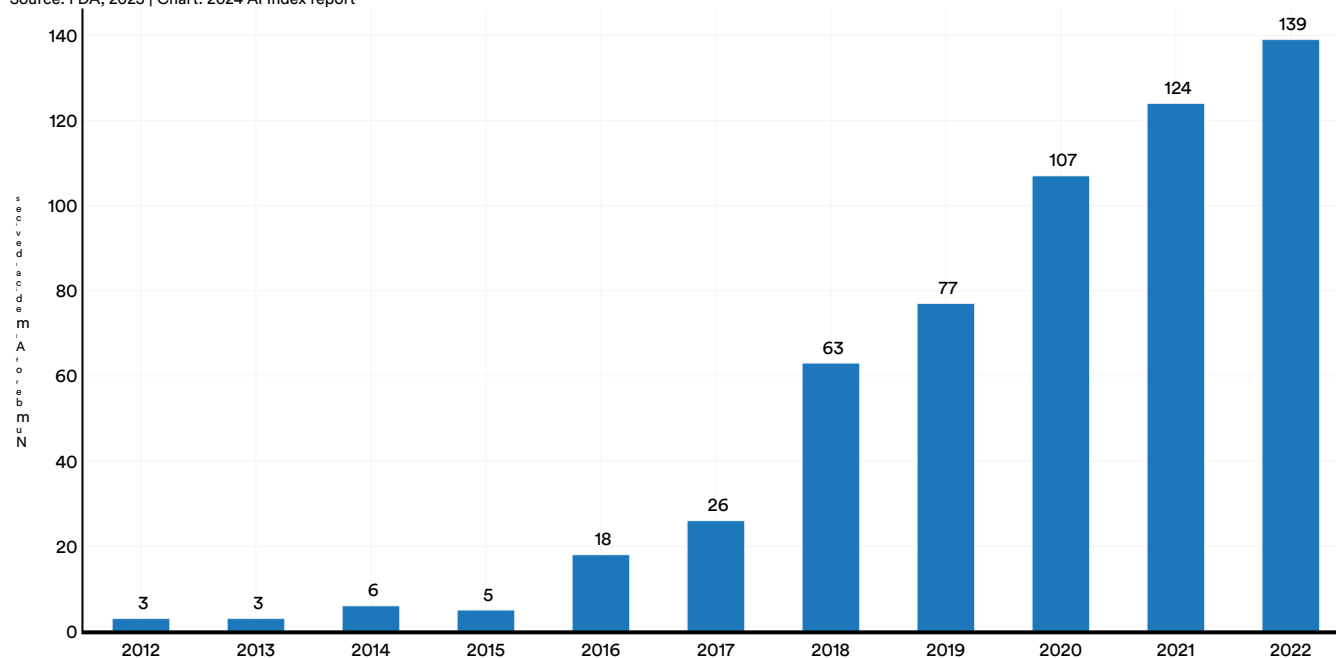


図5.2.20

³ The FDA last updated the list in October 2023, meaning that the totals for 2023 were incomplete. Consequently, the AI Index limited its data presentation to include only information up to 2022.

図5.2.21は、FDA承認の医療機器に関連する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%の大多数が放射線学に関連していました。次に一般的な専門分野は心血管で、承認の7.2%を占めています。

2012年から2022年までの専門ごとのFDAに承認されたAI医療機器の数

出典: FDA, 2023 | チャート: 2024 AIインデックスレポート

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Radiology	2	0	5	0	11	15	39	51	94	105	121
Cardiovascular	0	0	0	1	4	6	9	12	7	11	10
Neurology	0	0	1	0	1	1	4	4	0	2	2
Gastroenterology and urology	0	0	0	0	0	1	1	1	0	3	1
Hematology	0	1	0	0	0	2	2	1	3	0	1
Microbiology	0	2	0	0	0	0	0	2	1	0	0
General hospital	0	0	0	1	0	0	0	2	0	0	0
General and plastic surgery	0	0	0	0	1	0	2	1	0	1	0
Ophthalmic	0	0	0	1	0	0	2	1	1	1	2
Clinical chemistry	0	0	0	1	1	0	2	1	0	0	1
Anesthesiology	0	0	0	1	0	1	0	0	1	0	0
Pathology	1	0	0	0	0	0	0	0	0	1	0
Ear nose and throat	0	0	0	0	0	0	0	0	0	0	1
Dental	0	0	0	0	0	0	1	0	0	0	0
Orthopedic	0	0	0	0	0	0	1	0	0	0	0
Obstetrics and gynecology	0	0	0	0	0	0	0	1	0	0	0

Figure 5.2.21

管理とケア

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

再度。

強調された研究:

メッドアライン

医療におけるAIの重要な進展にもかかわらず、MedQAやUSMLEなどの既存のベンチマークは、知識に基づく質問に焦点を当てており、臨床医が患者ケアで行う多様なタスクを完全には捉えていません。臨床医は、個別の診断計画を作成するなどの情報集約型のタスクに関与することが多く、作業時間の大部分を管理業務に費やしています。AIにはこれらのプロセスを効率化する可能性があります。医療的管理LLMのベンチマーキングや微調整に適した電子健康記録（EHR）データセットが不足しています。今年、研究者たちはこのギャップを解決するために、EHRに基づいた包括的なMedAlignを紹介する進展を遂げました。

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

研究者たちは次に、MedAlignでさまざまな既存のLLMをテストしました。すべてのLLMの中で、マルチステップの洗練を使用したGPT-4のバリエーションが最も高い正確さ（65.0%）を達成し、他のLLMよりも定期的に好まれました（図5.2.23）。MedAlignは、AIを利用して医療の管理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

Source: Fleming et al., 2023

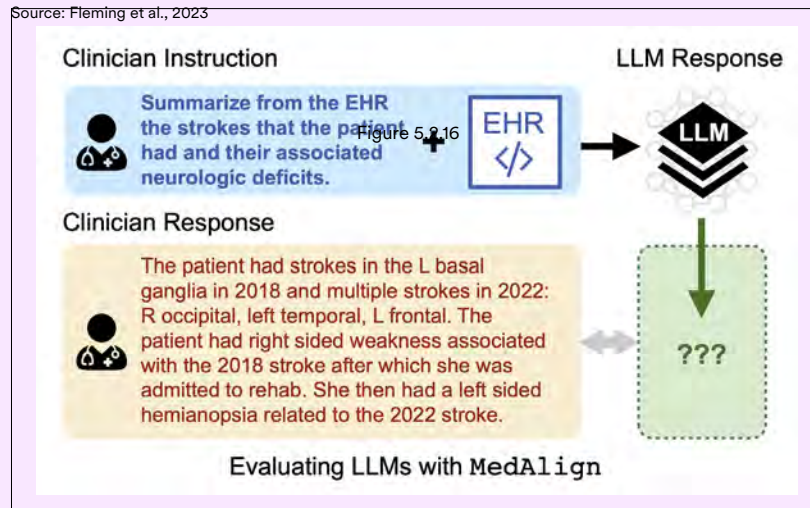


図5.2.22

強調された研究:

メッドアライン (続き)

モデルパフォーマンスの評価：人間対COMETランク

Source: Fleming et al., 2023 | Chart: 2024 AI Index report

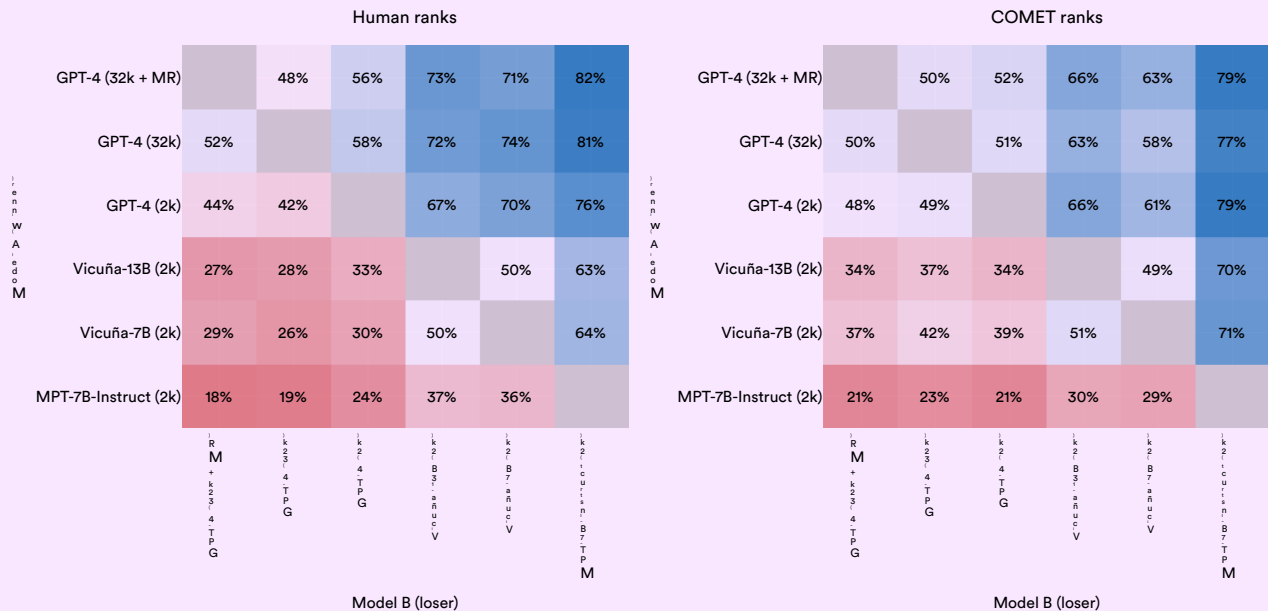
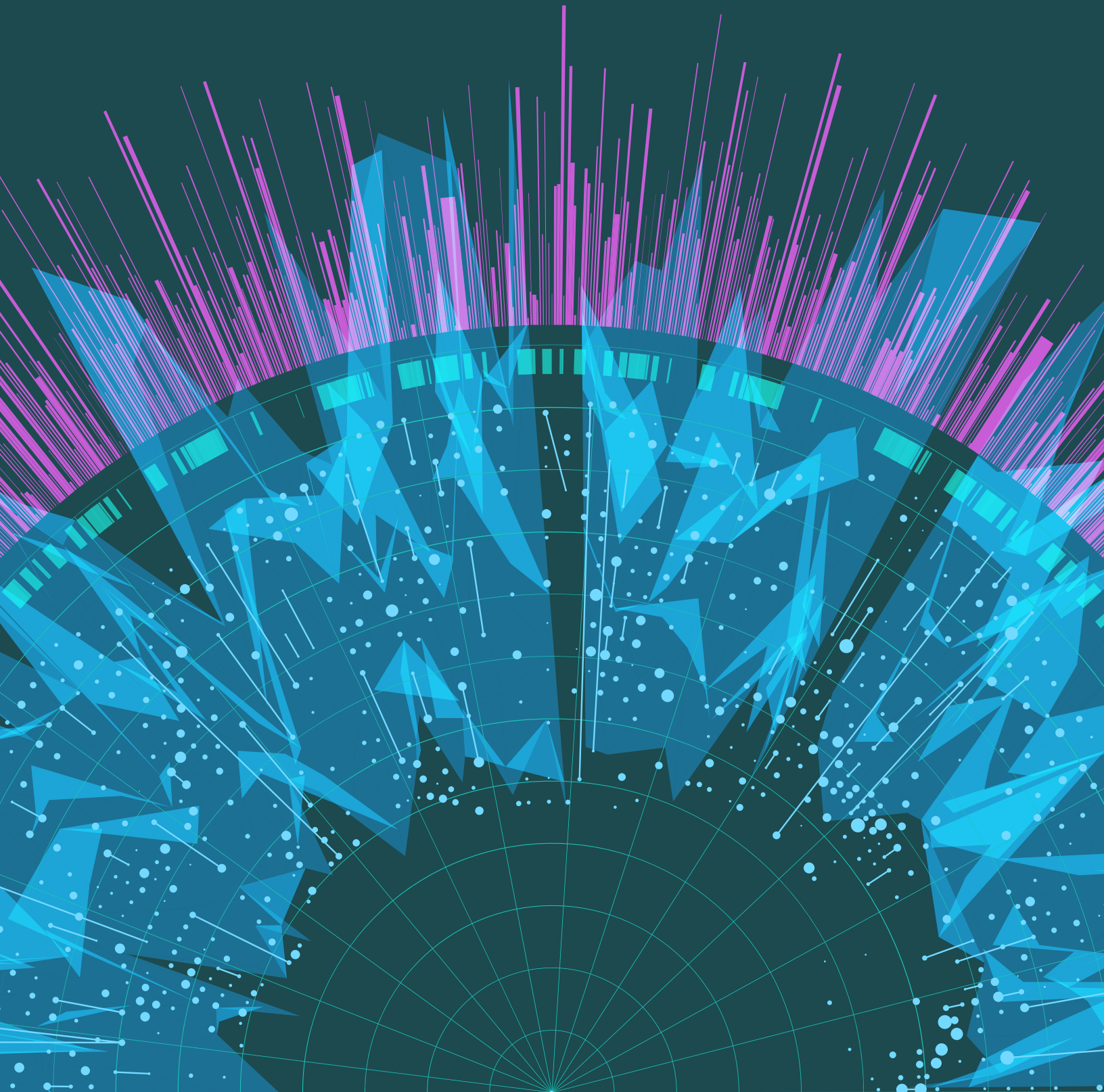


図5.2.23



Artificial Intelligence
Index Report 2024

CHAPTER 6: Education



プレビュー

概要 327 章のハイライト 328

6.1 Postsecondary CS and AI Education	329
United States and Canada	329
CS Bachelor’s Graduates	329
CS Master’s Graduates	331
CS PhD Graduates	333
CS, CE, and Information Faculty	336
Europe	344
Informatics, CS, CE, and IT Bachelor’s Graduates	344
Informatics, CS, CE, and IT Master’s Graduates	347
Informatics, CS, CE, and IT PhD Graduates	351
AI-Related Study Programs	355
Total Courses	355
Education Level	356
Geographic Distribution	357

6.2 K-12 コンピュータサイエンスと AI 教育 アメリカ合衆国 州レベルの動向 AP コンピュータサイエンス ハイライト: アクセスの問題 ハイライト: 教師と学生の間の ChatGPT の使用

公共データにアクセスする

概要

この章では、AIとコンピュータサイエンス（CS）教育の動向について検討し、誰が学んでいるのか、どこで学んでいるのか、そしてこれらの動向がどのように進化してきたのかに焦点を当てています。教育に対するAIの影響に関する懸念が高まる中、教師や学生によるChatGPTなどの新しいAIツールの使用についても調査しています。

この分析は、Computing Research Associationの年次タウルビー調査に基づいたアメリカとカナダにおける高等教育のコンピュータサイエンス（CS）と人工知能（AI）の現状の概要から始まります。その後、ヨーロッパにおけるコンピュータサイエンス教育に関するデータをインフォマティクスヨーロッパからレビューします。今年は、Studyportalsからのデータを基にしたAI関連の英語での学習プログラムの世界的な数に関する新しいセクションを紹介します。

この章は、Code.orgによるアメリカ合衆国のK-12コンピュータサイエンス教育に関する洞察と、学校におけるChatGPTの使用に関するウォルトン財団の調査結果で締めくくられています。

章のハイライト

1. アメリカとカナダのコンピュータサイエンスの学士卒業生の数は増加し続けており、新しいコンピュータサイエンスの修士卒業生は比較的横ばいで、博士卒業生は緩やかに増加しています。

アメリカとカナダの新しい学士号取得者の数は、10年以上にわたって一貫して増加していますが、コンピュータサイエンスの大学院教育を選ぶ学生数は横ばいです。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の数はわずかに減少しています。

2. AIの博士号取得者が産業界に移行する流れは加速しています。2011年には、新たに取得したAIの博士号のうち、産業界に就職した割合は約40.9%、学界に就職した割合は41.6%でほぼ同じでした。しかし、2022年までに、卒業後に産業界に参加した割合は70.7%と大幅に増加し、一方で学界に入ったのは20.0%でした。過去1年間だけで、産業界に向かうAIの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頭脳流出が激化していることを示しています。

3. 学界への産業界からの学術的才能の移行が減少している。2019年には、アメリカとカナダの新しいAI教員の13%が産業界からのものであったが、2021年にはこの数字が11%に減少し、2022年にはさらに7%に落ち込んだ。この傾向は、産業界から学界への高度なAI人材の移動が徐々に減少していることを示している。

4. アメリカとカナダにおけるコンピュータサイエンス教育は国際的でなくなっていく。2022年には2021年よりも国際的なコンピュータサイエンスの学士号、修士号、博士号を取得した学生の割合が少なかった。修士課程の国際学生の減少は特に顕著だった。

5. アメリカの高校生はCSコースを受講する生徒が増えているが、アクセスの問題は依然として残っています。2022年には201,000件のAP CS試験が実施されました。2007年以降、これらの試験を受ける生徒の数は10倍以上に増加しています。しかし、最近の証拠によると、大規模な高校や郊外に住む生徒はCSコースにアクセスできる可能性が高いことが示されています。

6. AI関連の学位プログラムは、国際的に増加しています。英語のAI関連の高等教育学位プログラムの数は2017年以来3倍になり、過去5年間にわたり着実に年間増加しています。世界中の大学は、より多くのAIに焦点を当てた学位プログラムを提供しています。

7. 英国とドイツは、ヨーロッパにおける情報学、コンピュータサイエンス、計算機工学、ITの学位取得者の生産において先導しています。英国とドイツは、情報学、コンピュータサイエンス、計算機工学、および情報の学士、修士、博士号卒業生の数を最も多く生産しています。人口1人あたりで見ると、フィンランドは学士号と博士号卒業生の生産で先行しており、アイルランドは修士号卒業生の生産でトップです。

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）における高等教育の概要を提供し、北米とヨーロッパのさまざまな学位（学士、修士、博士）に関する卒業統計を強調しています。また、英語で提供されるAI関連のコースに関する情報も含まれています。

6.1 高等教育におけるコンピュータサイエ

アメリカ合衆国とカナダ

この小節では、米国とカナダにおけるコンピュータサイエンス（CS）および人工知能（AI）の高等教育の状況を評価するコンピューティングリサーチ協会のタウルビー調査からのデータの分析を示します。この調査は、米国とカナダの297の博士号授与CS学部を対象としています。

コンピュータサイエンス学士卒業生

過去10年間にわたり、北米における新しいコンピュータサイエンスの学士卒業生の総数は着実に増加しており、2021年から2022年にかけて年率7.9%の増加を記録し、3倍以上に増加しています（図6.1.1）。

アメリカとカナダの新しいコンピュータサイエンス学士卒業生、2010年～2022年
Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

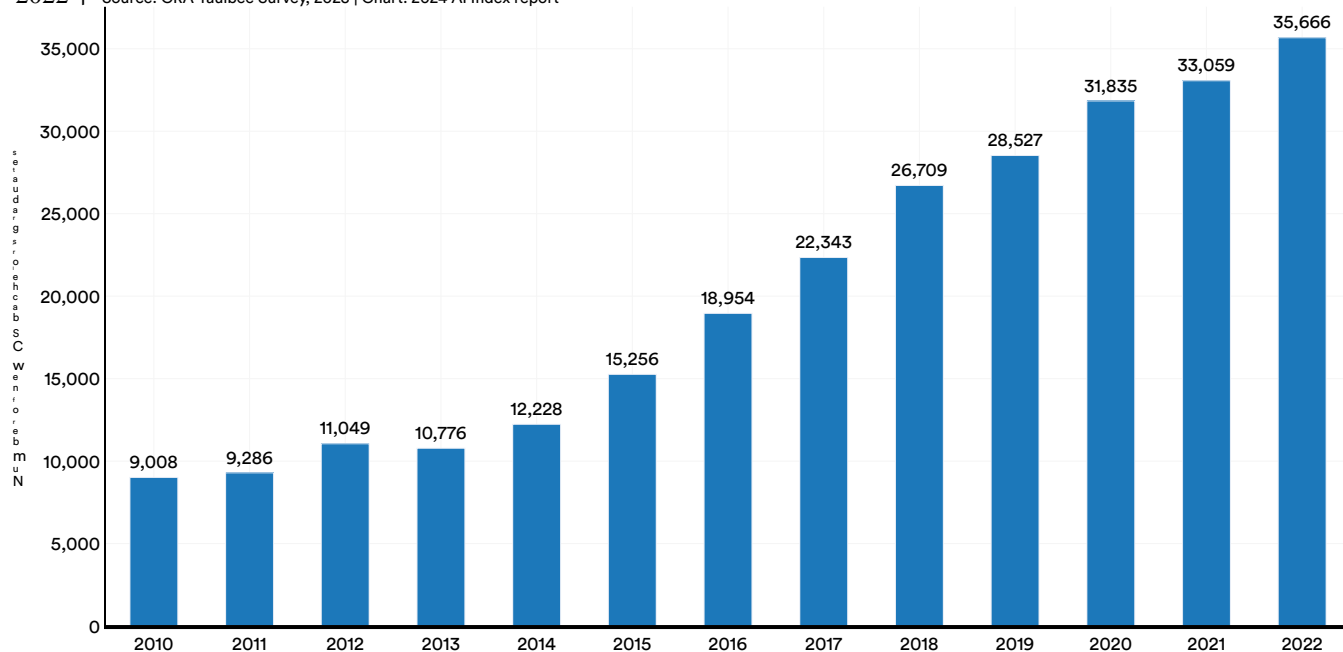


Figure 6.1.1

調査の対象となったすべての博士号授与学科が回答を提供したわけではないことを指摘することが重要です。対象となった297の学科のうち、182が回答し、全体の回答率は61%となりました。

約8年ぶりに、アメリカとカナダの大学のコンピュータサイエンスの学士号取得者の中で国際学生の割合が減少し、2021年の16.3%から2022年の15.2%に落ち込みました（図6.1.2）。この減少は、トランプ政権初期の間に留学ビザの取得が難しくなったことを反映している可能性が高く、その影響は今まさに現れ始めている

データです。この減少は、COVID-19パンデミック中に課された国際旅行制限の一因でもあり、国際学生がアメリカとカナダで学ぶ能力に影響を与えました。この最近の落ち込みにもかかわらず、過去10年間の全体的なトレンドは、国際学生の割合が着実に増加していることを示しています。

2010～22年におけるアメリカ合衆国とカナダの新しい国際的なコンピュータサイエンス学士卒業生の割合（総数の%） Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

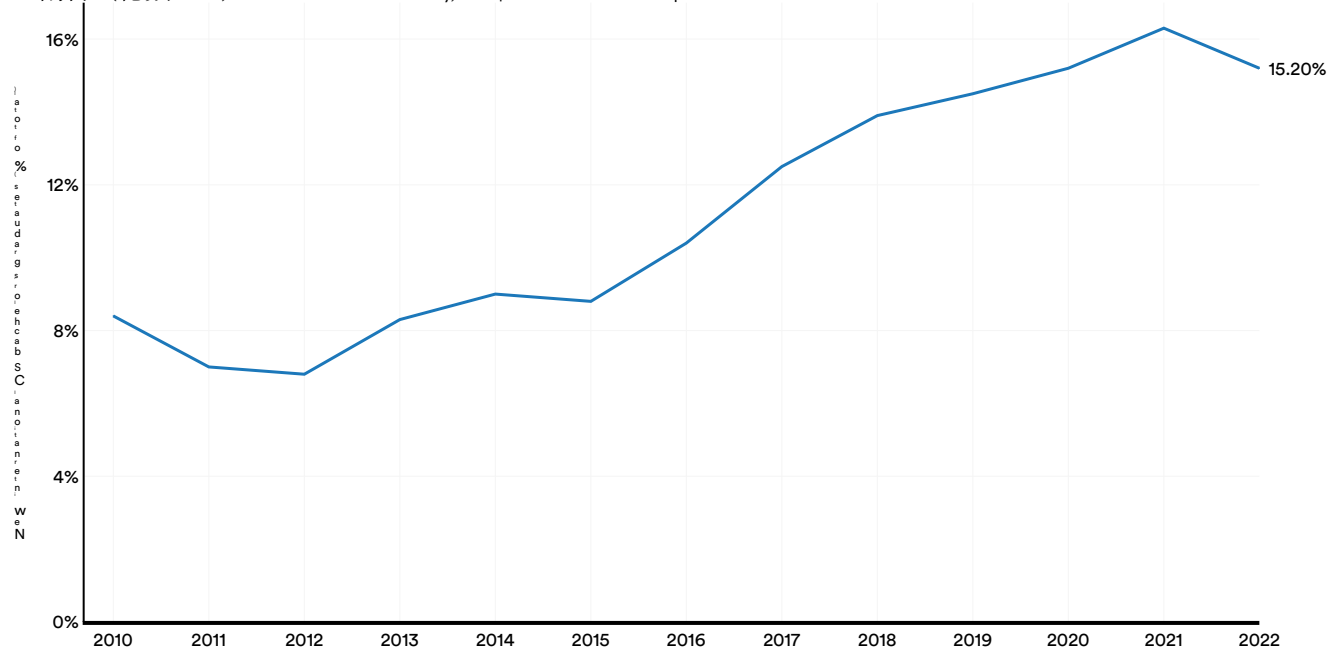


図6.1.2

コンピュータサイエンス修士卒業生

AIコースはコンピュータサイエンスの修士課程に一般的に含まれています。アメリカとカナダの大学からの新しいコンピュータサイエンスの修士号を取得した卒業生の総数は、過去10年間で2倍以上に増加しました、

その数値は2018年以降横ばいになり、昨年は2.5%の若干の減少を示しています（図6.1.3）。この安定は、以下のグラフに示されている国際的な修士課程の学生の減少を反映しています。

アメリカとカナダの新しいコンピュータサイエンス修士卒業生、2010年～2022年

Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

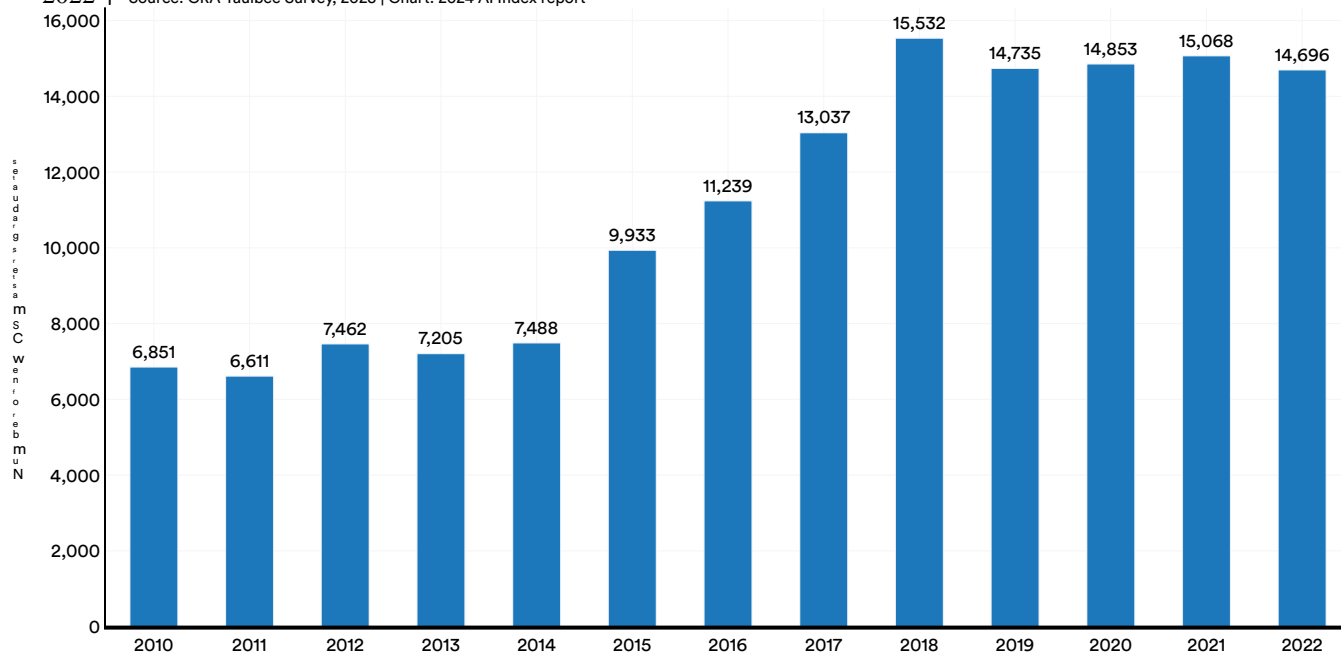


図6.1.3

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修士課程の学生数が著しく減少しました。この下落傾向は2017年頃から始まりましたが、昨年は14.8パーセントポイントの減少が最も顕著でした（図6.1.4）。現在、国際的なコンピュータサイエンス修士課程の卒業生と国内卒業生の割合はほぼ均等です。

アメリカとカナダにおける新しい国際コンピュータサイエンス修士卒業生（総数に対する割合）、2010-22

Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

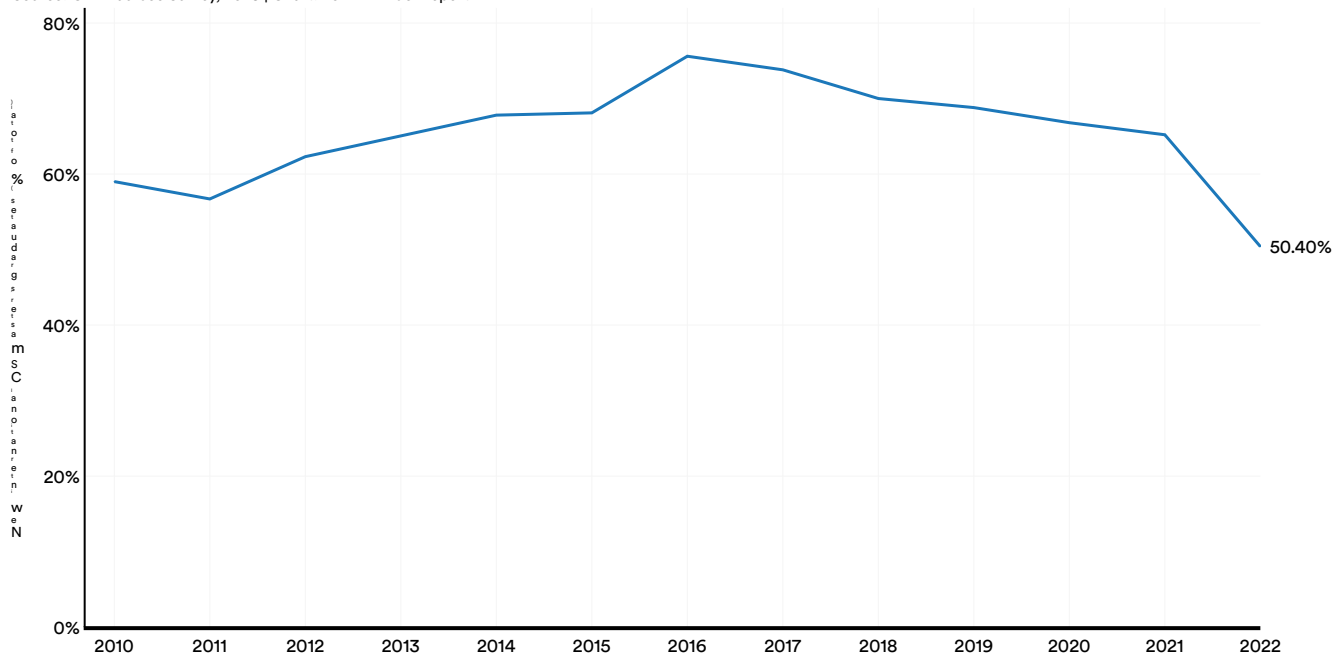


図6.1.4

コンピュータサイエンス博士号卒業生

10年ぶりに、アメリカとカナダの大学で新しいコンピュータサイエンスの博士号取得者の数が著しく増加しました。2022年には、コンピュータサイエンスの博士号取得者が2,105人に達し、2010年以来最高となりました（図6.1.5）。

アメリカとカナダの新しいコンピュータサイエンス博士号卒業生、
2010年～2022年

Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

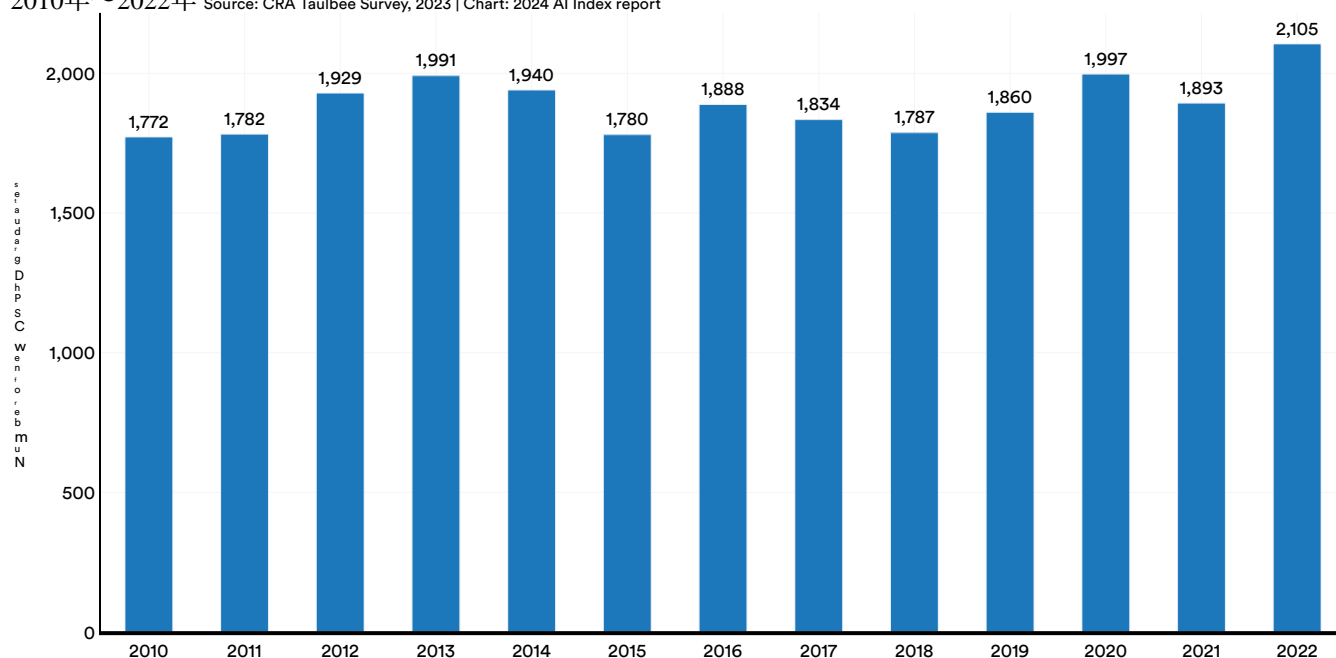


図6.1.5

過去10年間でCS博士課程の卒業生の中の留学生の割合は増加しましたが、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%に落ちました（図6.1.6）。

アメリカとカナダにおける新しい国際コンピュータサイエンス博士号取得者の割合（2010年から2022年）

Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

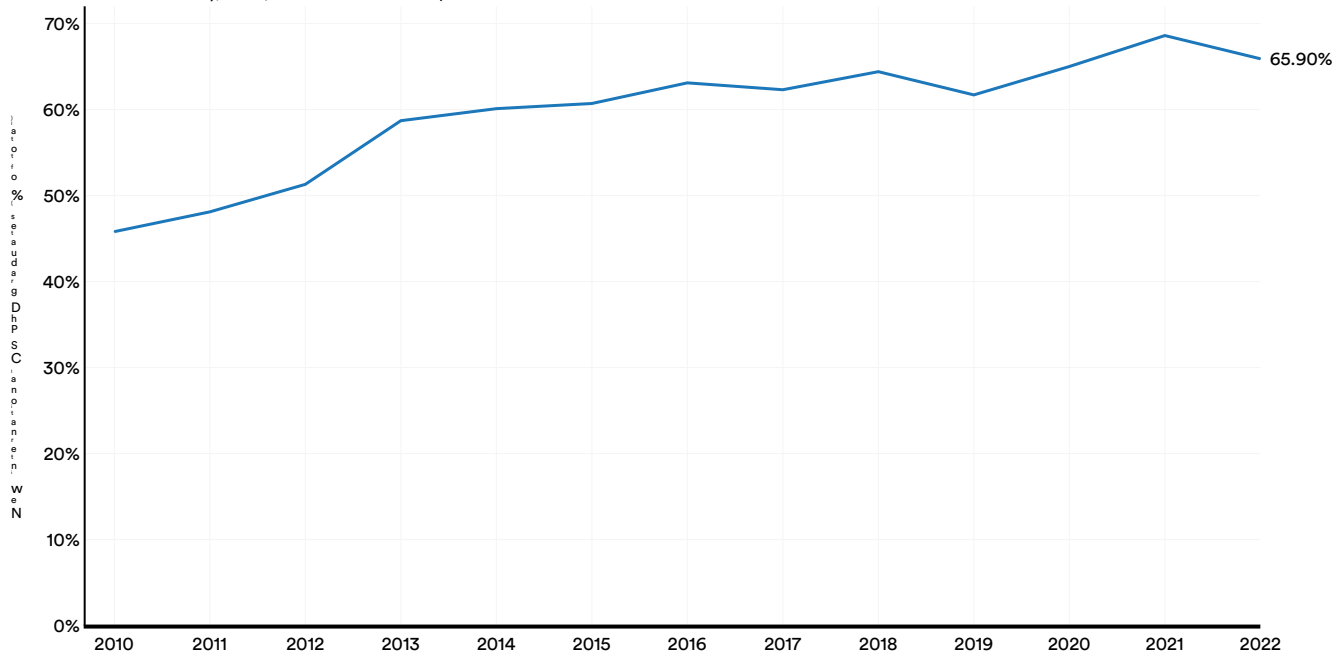


図6.1.6



新しく取得したAIの博士号を持つ人々は、卒業後どこで働くことを選んでいるのでしょうか？昨年のAIインデックスレポートで強調されたトレンドに従い、AIの博士号取得者の中で産業界でのキャリアを追求する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、ほぼ同じ割合の40.9%が産業界に就職し、41.6%が学界に進みました。

しかし、2022年までに、卒業後に業界に参加した割合（70.7%）は、学界に入る人々（20.0%）と比べて大幅に大きくなりました。新しいAI博士号取得者が政府の役割に就く割合は、過去5年間で約0.7%と比較的低く安定しています。

2010年から2022年におけるアメリカとカナダの
セクター別新しいAI博士号取得者の雇用割合（
%）

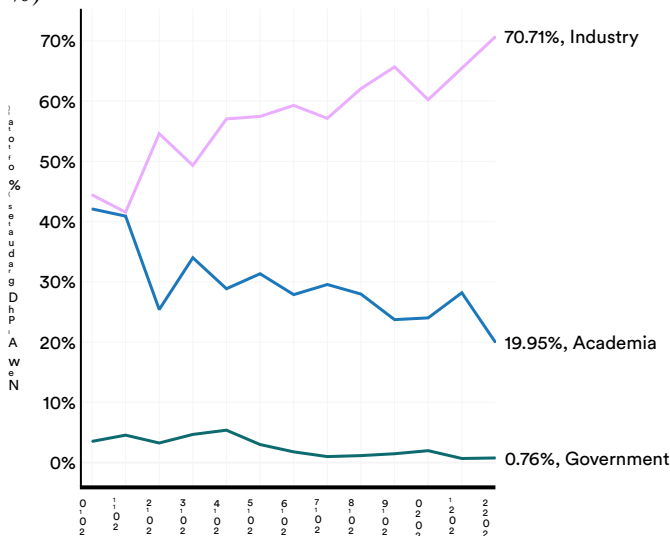


Figure 6.1.7²

アメリカにおける新しいAI博士の雇用と
カナダのセクター別、2010年から2022年まで
Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

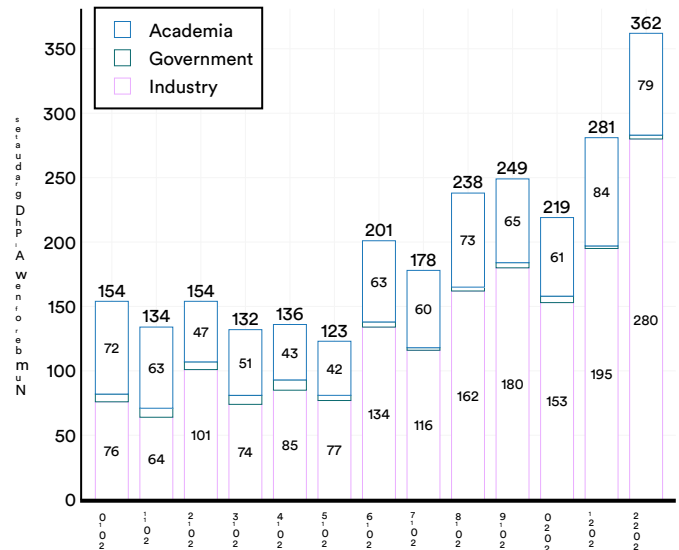


Figure 6.1.8

² The sums in Figure 6.1.7 do not add up to 100, as there is a subset of new AI PhDs each year who become self-employed, unemployed, or report an “other” employment status in the CRA survey. These students are not included in the chart.

CS、CE、情報学部

CSおよびAI教育のトレンドをよりよく理解するためには、CSの教員に関するデータを検討することが有益です。昨年、アメリカとカナダの大学におけるCS、CE、および情報学の教員の総数は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。

アメリカ合衆国とカナダのコンピュータサイエンス、コンピュータエンジニアリング、および情報学の教員数、2011年～2022年 Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

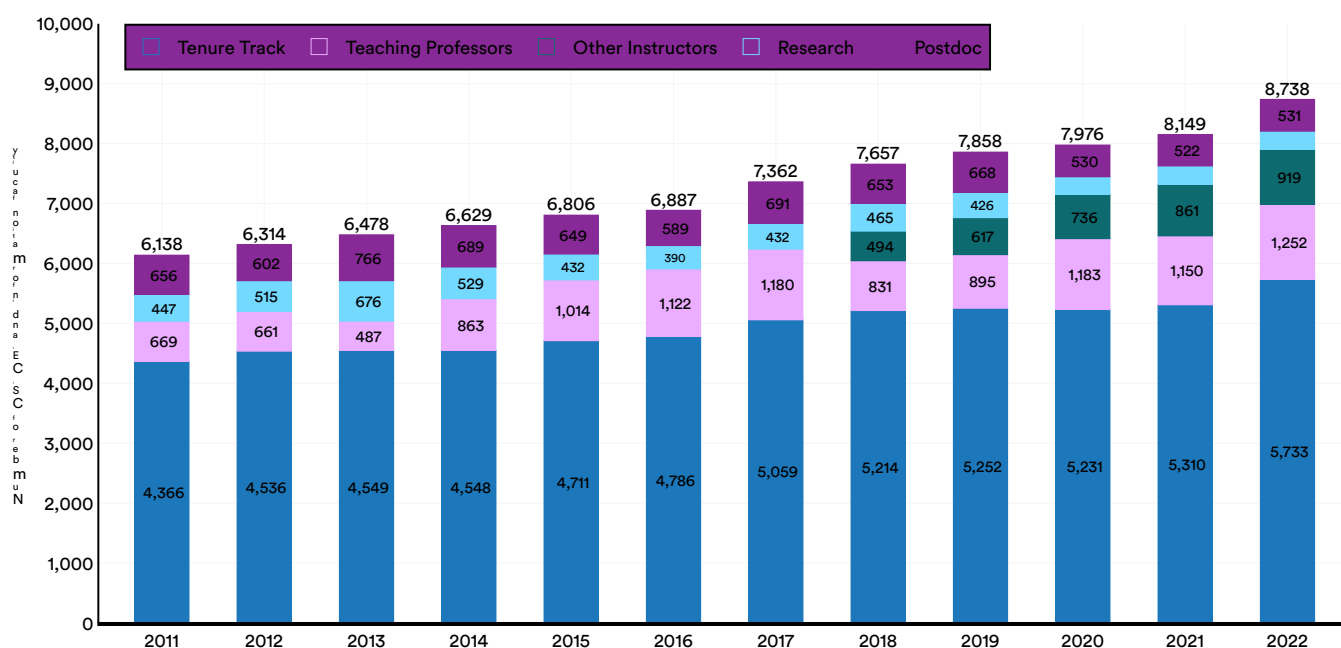


図6.1.9

2022年、アメリカには7,084人のコンピュータサイエンスの教員がおり、その大多数（65.7%）がテニユートラックに所属しています（図6.1.10）。アメリカのコンピュータサイエンスの教員の総数は、2021年から4.4%、2011年からは45.0%増加しました。

アメリカ合衆国のコンピュータサイエンス教員数、2011年～2022年

出典: CRA タウルビー調査、2023 | チャート: 2024年AIインデックスレポート

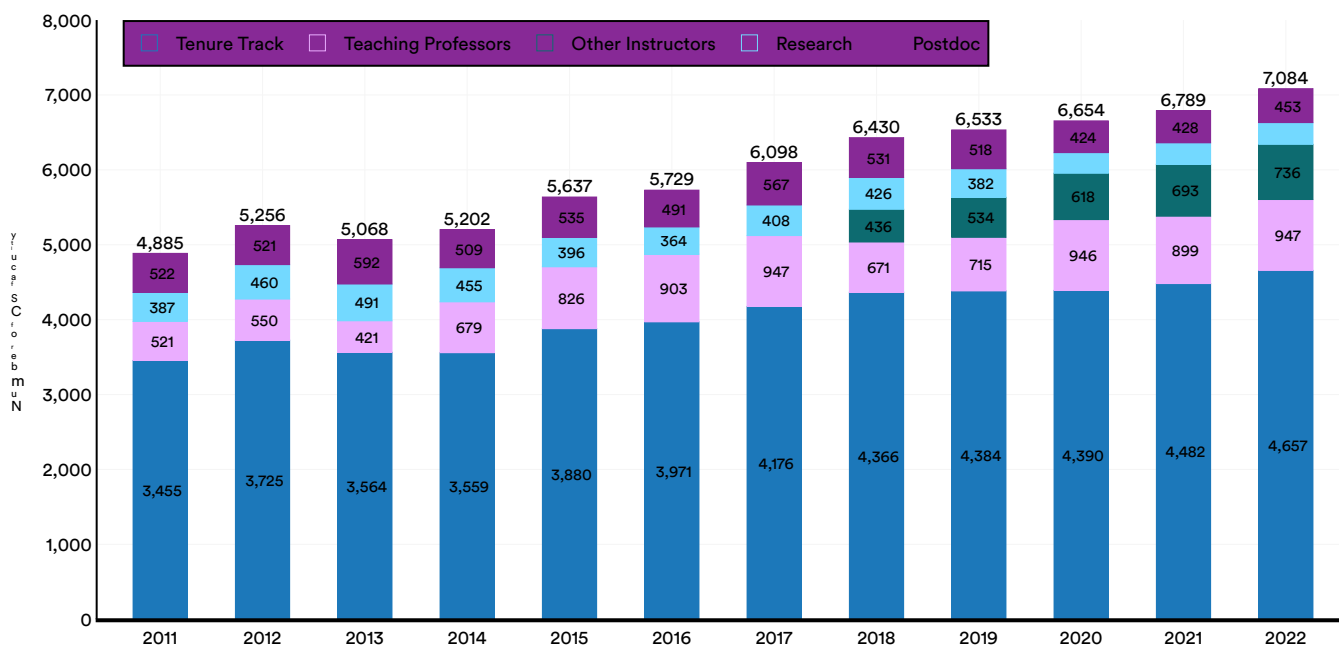


図6.1.10

昨年、北米のコンピュータサイエンス、コンピュータ工学、情報分野で915人の新しい教員が雇われ、これは10年ぶりの高水準です。これらのポジションのうち455はテニュートラックでした。（図6.1.11）

新しいコンピュータサイエンス、コンピュータエンジニアリング、および情報学の教員の採用状況（アメリカとカナダ、2011年～2022年）

Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

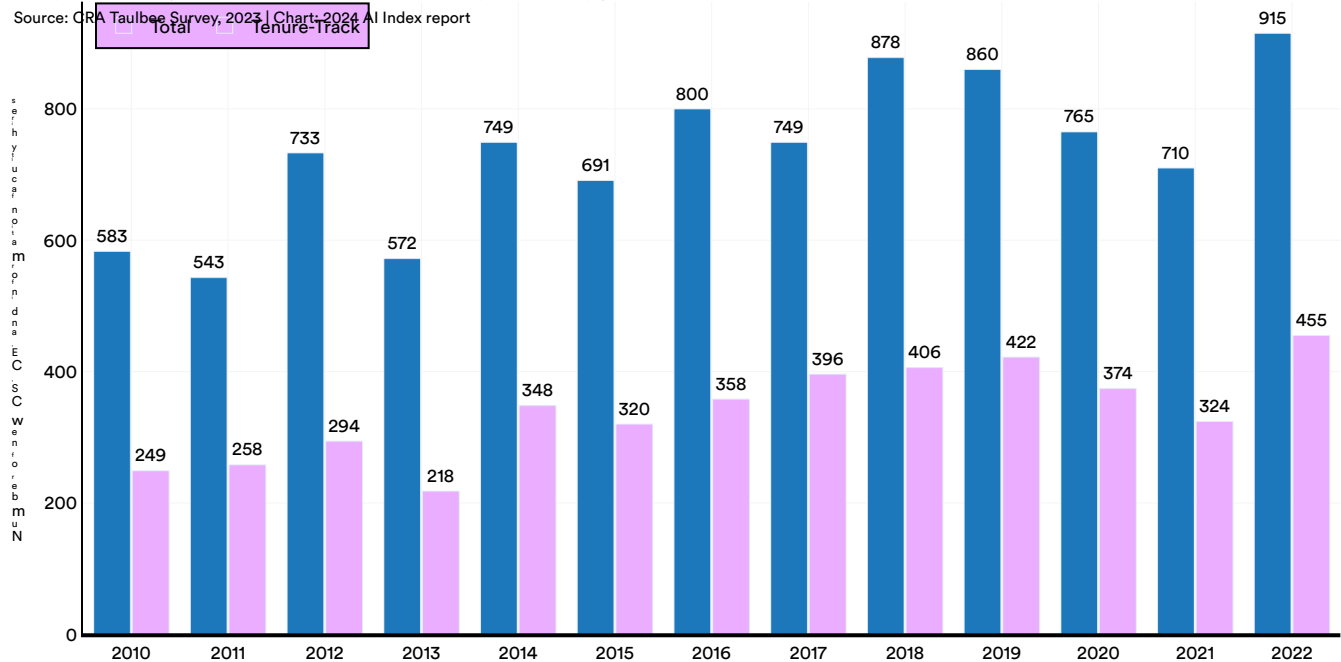


図6.1.11

2022年には、新しい教員の43%が他の学術職からの採用であり、学術的な労働力内での「循環」を示しています（図6.1.12）。これらの「新しい」教員が他の場所での職を vacate したため、その前の役割は最終的に埋められる必要があります。また、2022年に産業から移行する教員の割合は、前年の11%から7%に減少し、2019年の13%に比べても減少しました。

アメリカとカナダのコンピュータサイエンス、コンピュータ工学、情報学部における新しい教員の源、2018年～2022年 Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

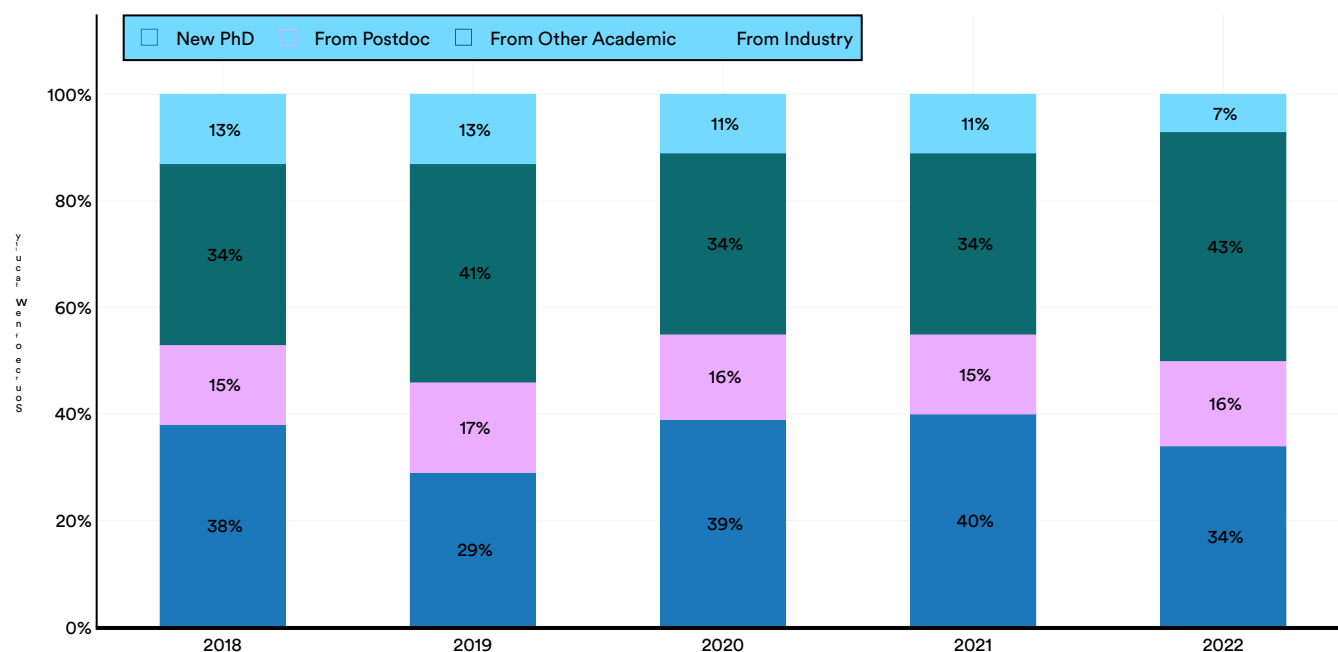


図6.1.12

過去10年間で教員職が埋まらない理由はさまざまでした。2011年には、失敗した検索の37%はオファーが出されなかったことによるものであり、34%は出されたオファーが辞退されたことによるものでした（図6.1.13）。対照的に、2022年にはオファーが出されなかったのはわずか15%で、55%はオファーが関与していました。

却下されました。この傾向は、新しいコンピュータサイエンスの教員に対する競争がますます激化していることを反映しているようです。しかし、これが他の学術的な職位との競争の激化を示しているのか、産業界との競争の激化を示しているのかは不明です。

新しい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）、および情報学の教員ポジションが充足しなかった理由（総数の%）、2011年～2022年

Source: CRA Taubbee Survey, 2023 | Chart: 2024 AI Index report

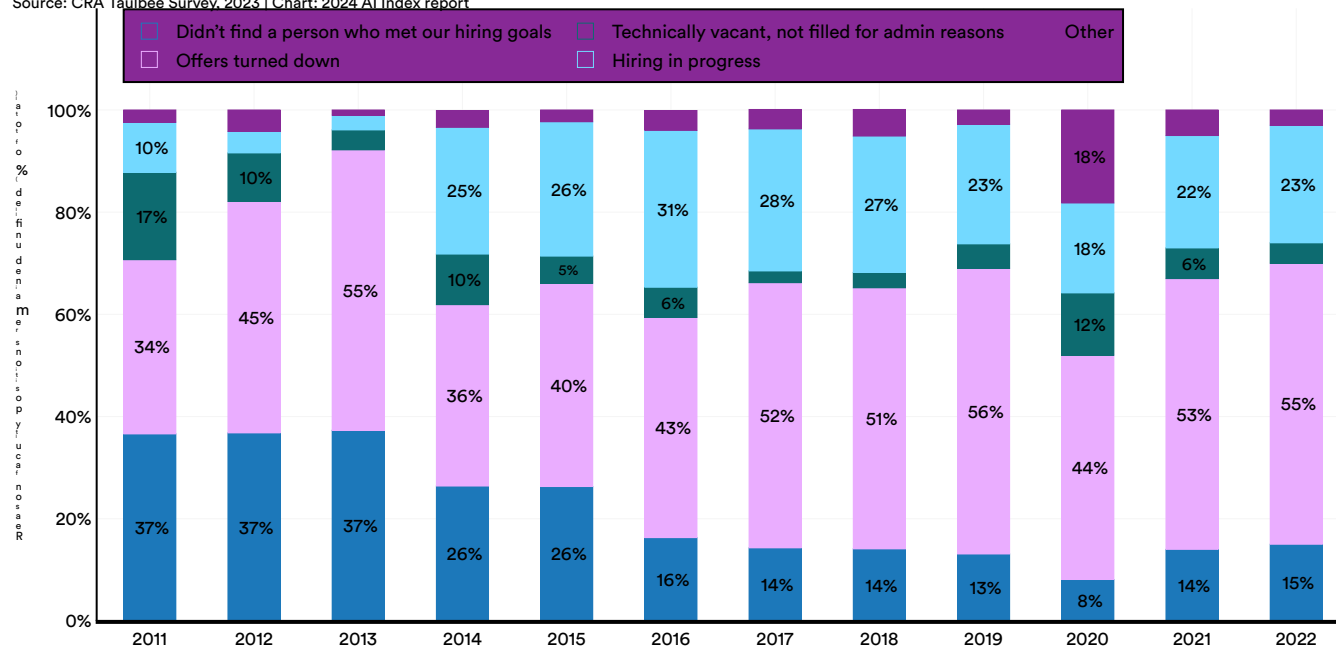


図6.1.13

2022年には、北米のコンピュータサイエンス（CS）、コンピュータ工学（CE）、および情報関連学科で教員の離職が大幅に増加し、405人に達しました。これは2021年の303人と比較されています（図6.1.14）。これらの離職のうち、38.5%は他の学術的なポジションに移り、16.3%は非学術的な役割に移りました。これは、過去の年と一貫した傾向を維持しています。

アメリカとカナダのコンピュータサイエンス、コンピュータエンジニアリング、情報学部
の教員の喪失、2011年～2022年 Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

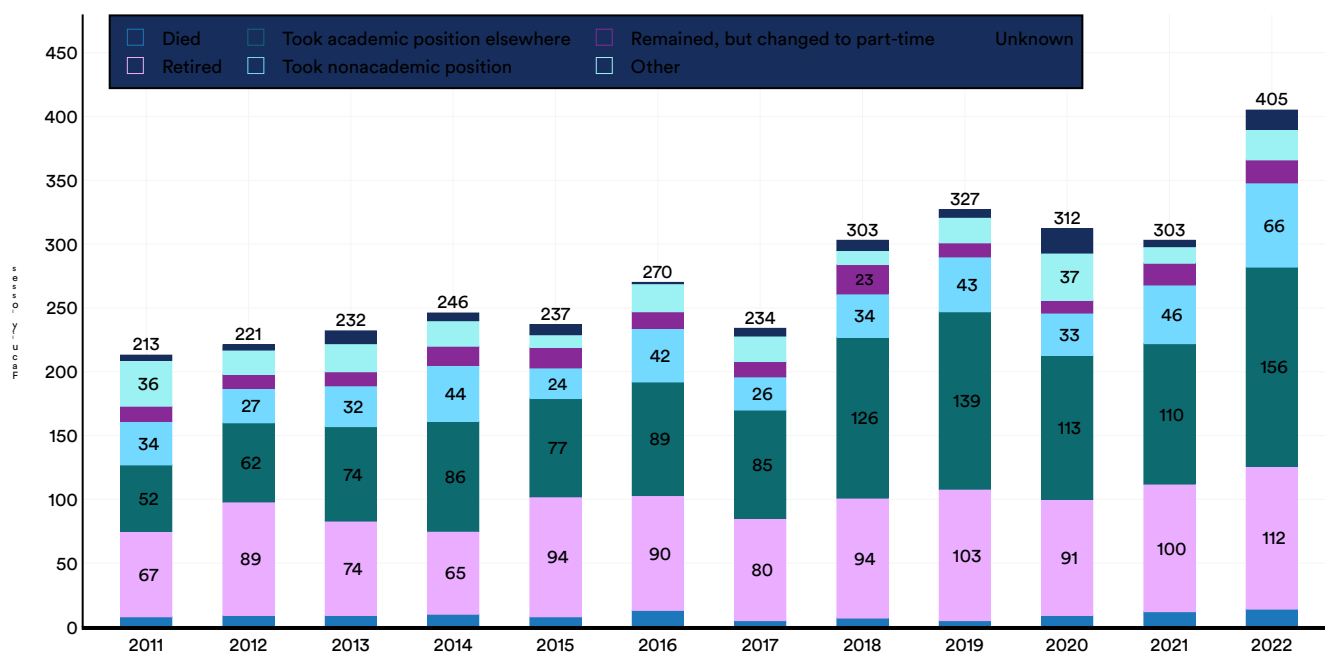


図6.1.14

2015年以来、フルプロフェッサーの中央値9か月間の給与の増加は、わずかに米国のインフレーション率を下回っていますが、助教授および准教授の中央値給与はインフレを上回るわずかな増加をしています。2022年には、フルプロフェッサーの給与は2021年より3.2%高く、7%の米国インフレーション率には追いついておらず、2015年より16.4%高いものの、その期間の19%のインフレーション増加にはまだ及んでいません（図6.1.15）。

アメリカ合衆国におけるコンピュータサイエンス教員の2015年から2022年までの中央値の9か月給

Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

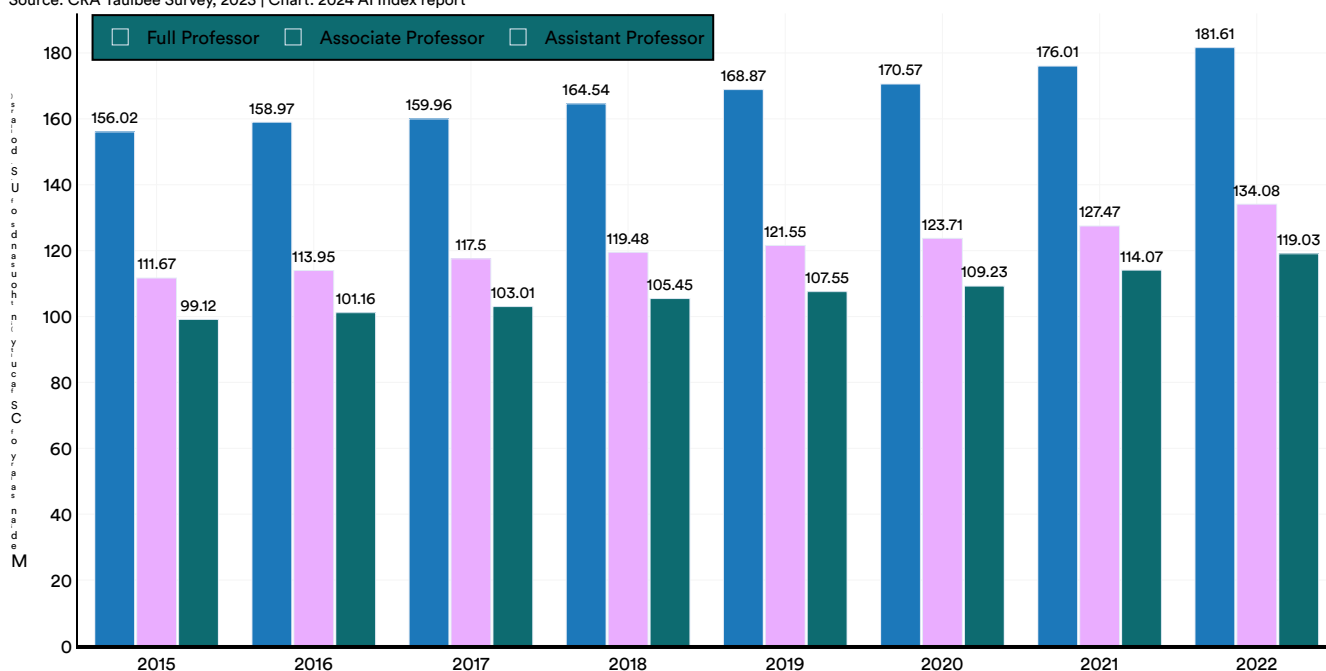


図6.1.15

2022年には、コンピュータサイエンス、コンピュータエンジニアリング、情報学の新しいテニユートラック教員における国際的な採用の割合が、大幅に増加し、前年の13.2%から19.3%に達しました（図6.1.16）。これは過去10年間で記録された2番目に高い割合であり、2013年に次ぐものです。

2010年から2022年におけるアメリカとカナダの新しい国際的なコンピュータサイエンス、コンピュータ工学、及び情報分野のテニユートラック教員の雇用割合（総数の%）

Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

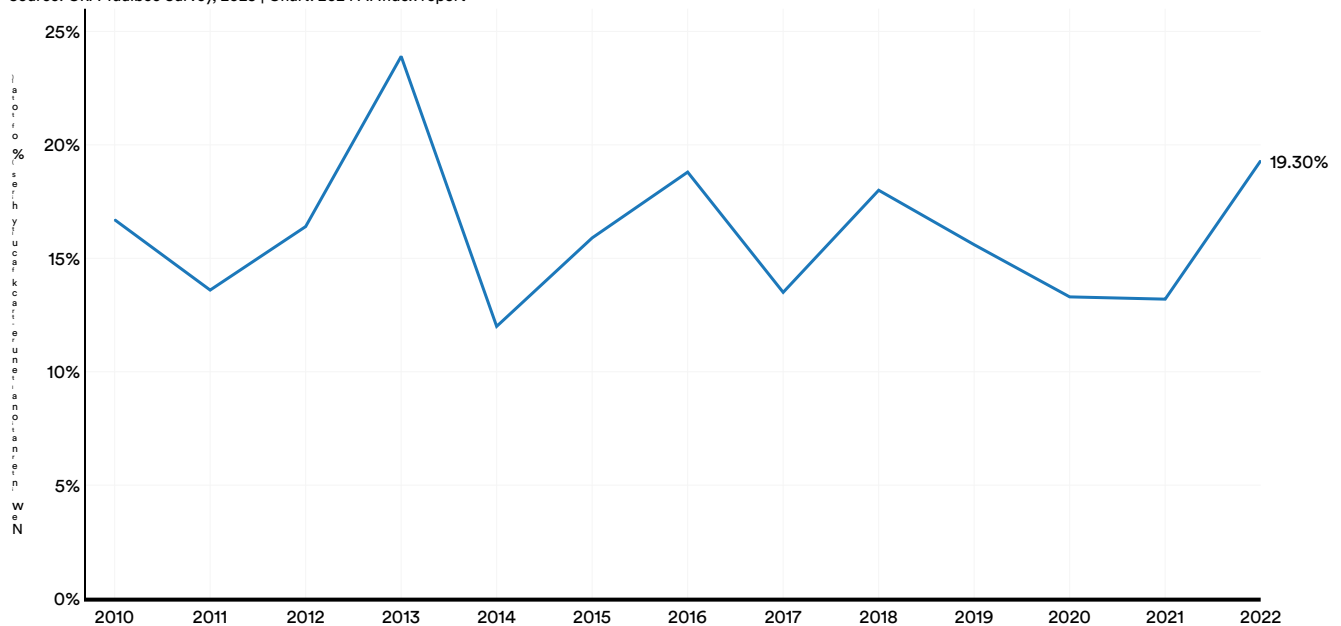


Figure 6.1.16

ヨーロッパ

ヨーロッパのコンピュータサイエンス卒業生に関するデータは、ヨーロッパにおける情報学教育の状況を監視するなどの目的を持つ学術および研究コミュニティ、インフォマティクス・ヨーロッパから得られています。インフォマティクス・ヨーロッパは、ヨーロッパの政府の統計局から情報学、コンピュータサイエンス、コンピュータ工学、計算技術、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。

情報学、コンピュータサイエンス、コンピュータエンジニアリング、および情報技術の学士卒業生

2022年、イギリスは情報学、コンピュータサイエンス、コンピュータ工学、およびITの学士号の新卒者数が最も多く、約25,000人に達しました（図6.1.17）。ドイツとトルコが続きました。サンプル内のほとんどの国は、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、チェコ共和国のような例外もありました（図6.1.18）。

2022年の欧州における新しい情報学、コンピュータサイエンス、コンピュータ工学、及び情報技術の学士卒業生数

Source: Informatics Europe, 2023 | Chart: 2024 AI Index report

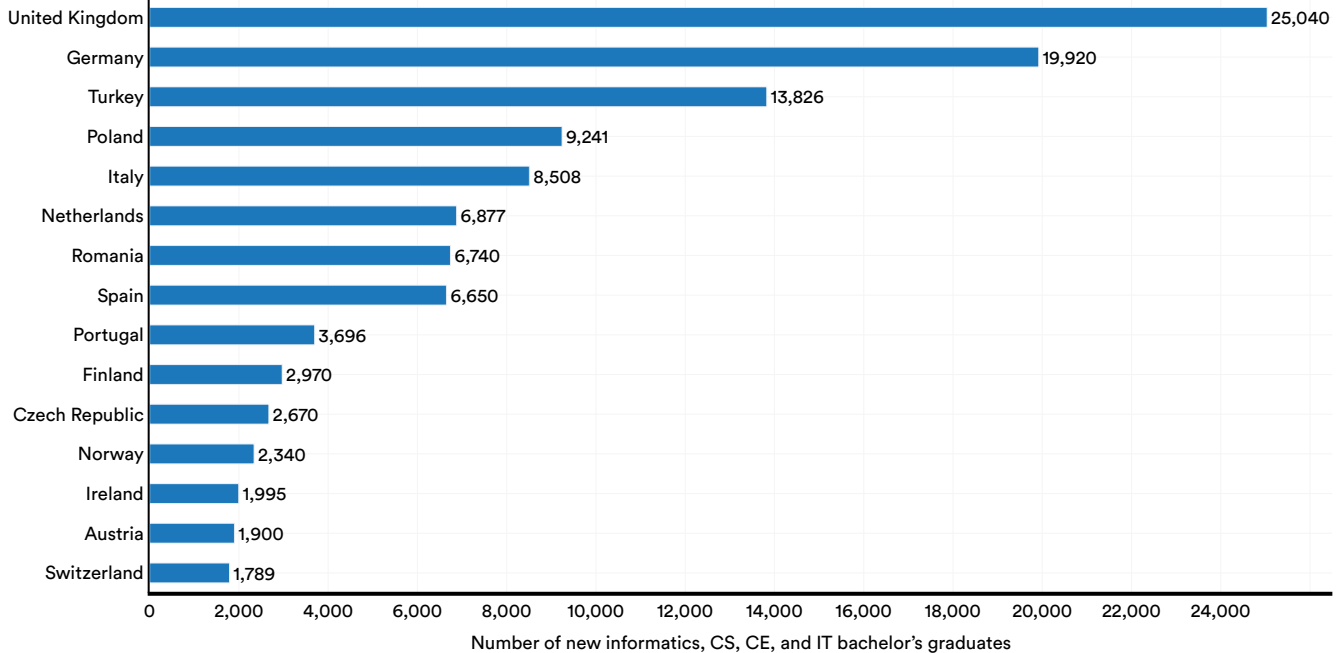


図6.1.17

3 There is no singular term for CS education that is used uniformly across European countries. Across Europe, CS education can be reflected in terms such as informatics, computer science (CS), computer engineering (CE), computing, information technology (IT), information and communication technology (ICT), and information science and technology (IST). The full list of subject names (and English translations) that Informatics Europe uses to identify informatics studies programs can be found at the [following link](#).

4 読者は、CRAの北米データとその後のセクションに詳述されているヨーロッパのCS卒業生データとの間で一人当たりの比較を行わないよう警告されています。ヨーロッパのデータは国の統計局から収集されており、より広範なカバレッジを誇っています。

5 Note that not all countries for which the AI Index has data are visualized in the figures in this section. To access the complete data, please view the public data associated with this chapter. Moreover, the year label refers to the year in which an academic year ends. For example, the figures visualizing new graduates for 2022 reflect the number of graduates reported for the 2021/2022 academic year. For the sake of visual simplicity, the Index opts to focus on the year in which students graduated.

欧州における新しい情報学、コンピュータサイエンス、コンピュータ工学、IT学士卒業生の国別の割合の変化、2012年 vs. 2022年

Source: Informatics Europe, 2023 | Chart: 2024 AI Index report

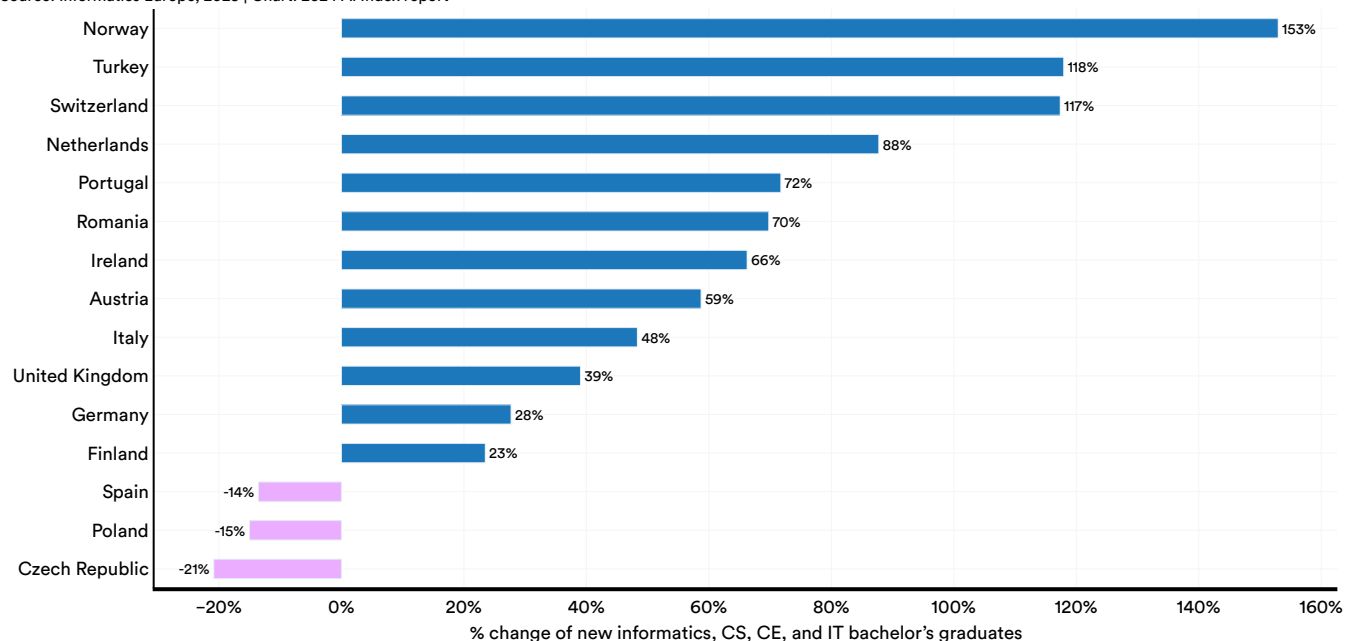


図6.1.18

フィンランド（53.4）、ノルウェー（42.6）、オランダ（38.6）が、10万人あたりの新しい情報学（CS、CE、IT）の学士卒業生の数でリードしています（図6.1.19）。一人当たりの基準で見ると、調査対象のほとんどの欧州諸国では、情報学、CS、CE、ITの学士卒業生の総数が増加しています（図6.1.20）。