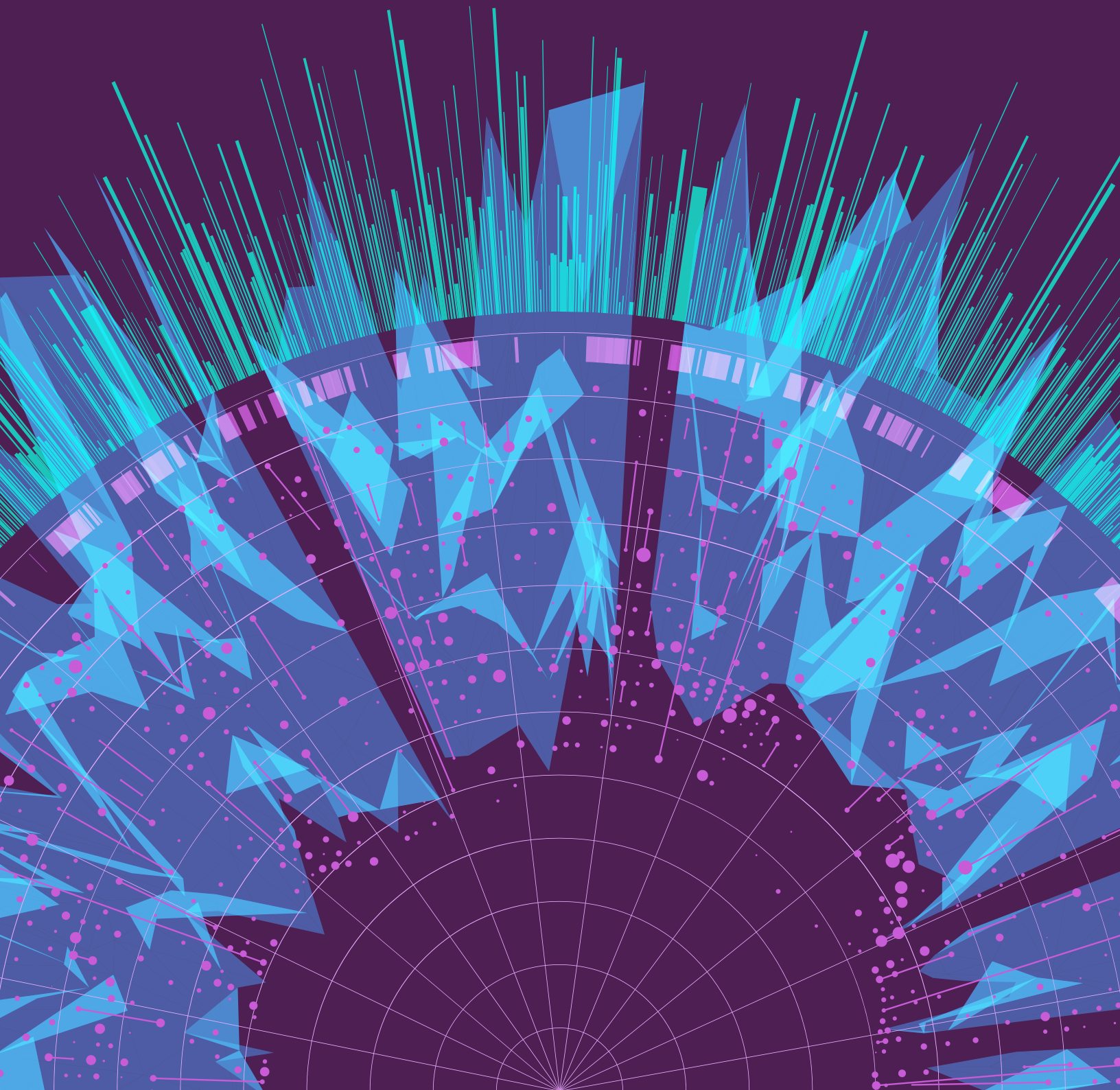




第5章： Science and Medicine



プレビュー

概要 298 章 のハイライト 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

ACCESS THE PUBLIC DATA



概要

今年のAIインデックスは、科学と医療におけるAIの役割の高まりを認識し、AIによる科学的および医療的発見に関する新しい章を紹介します。2023年の際立ったAIを活用した科学的成果、例えばGraphCastのような高度な天気予報システムやGNoMEのような改良された材料発見アルゴリズムを探ります。この章では、医療AIシステムのパフォーマンス、2023年の重要なAI駆動の医療革新であるSynthSRやImmunoSEIRA、FDAのAI関連医療機器の承認に関するトレンドも検討します。

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が科学的発見を進め始めました。しかし2023年には、アルファデブのような、アルゴリズムのソートをより効率的にする科学関連のAIアプリケーションがさらに多く登場しました。GNoMEは、材料発見のプロセスを促進します。

2. AIは医学の大きな進歩を助けています。 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AlphaMissenceなど、いくつかの重要な医療システムが導入されました。AIは、医療の進歩を推進するためにますます利用されています。

3. 高度な知識を持つ医療AIが登場しました。 過去数年間で、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ベンチマークで驚異的な改善を示しました。2023年の注目モデルである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ポイントの増加を記録しました。2019年にベンチマークが導入されて以来、MedQAにおけるAIのパフォーマンスはほぼ3倍になっています。

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%の増加を見せました。2012年以降、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、実際の医療目的でますます使用されています。

This section highlights significant AI-related scientific breakthroughs of 2023 as chosen by the AI Index Steering Committee.

5.1 注目すべき科学的マイルストーン

アルファデブ

AlphaDevはより高速なソートアルゴリズムを発見します。

アルファデブは、計算アルゴリズムの強化の分野における科学者やエンジニアの数十年の研究を改善した新しいAI強化学習システムです。アルファデブは、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。

Sort 3、Sort 4、Sort 5などの短いシーケンスにおける基本的なソートアルゴリズムです（図 5.1.1）。アルファデブ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++ sortライブラリに組み込まれました。これは、ライブラリのこの部分への最初の更新であり、10年以上ぶりのもので、強化学習を使用して設計された最初の追加です。

アルファデブと人間のベンチマークのアルゴリズム長の最適化に関する比較 出典: マンコウィッツら、2023 年チャート; 2024 年 AI インデックスレポート

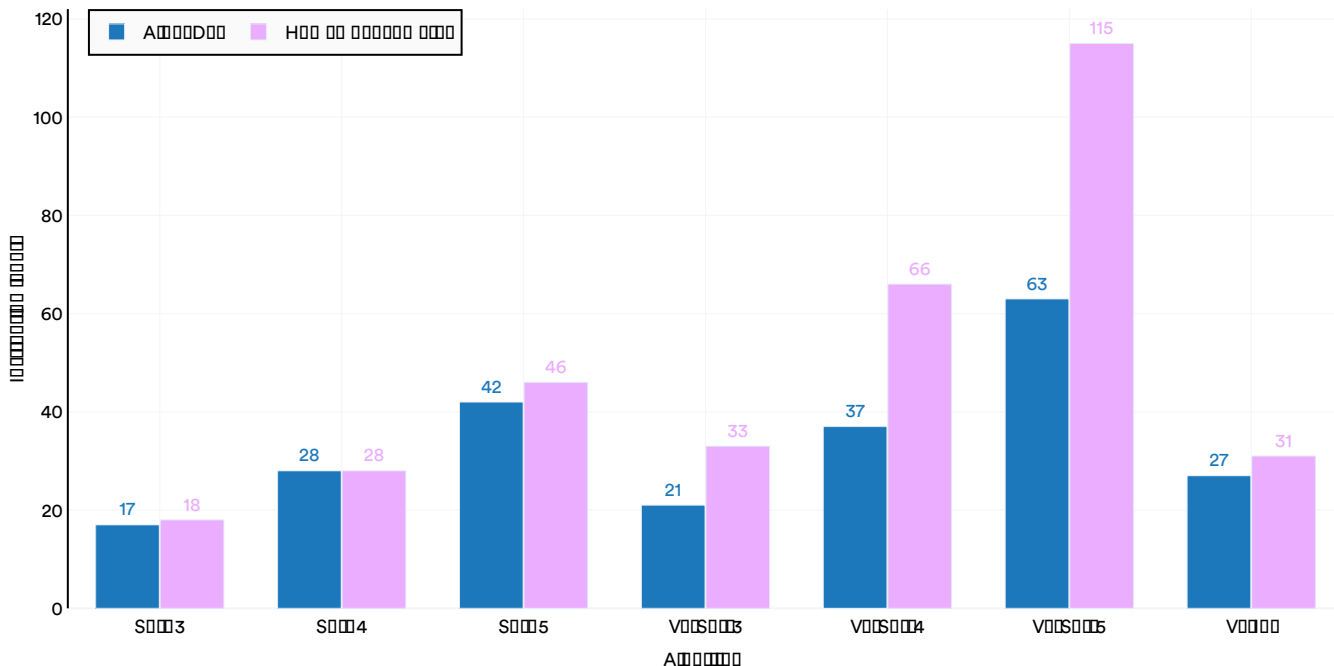


Figure 5.1.1

フレキシキューブ

フレキシキューブを用いた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重要であり、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。従来の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、これがその後の影響を及ぼします。

FlexiCubesは、勾配ベースの最適化と適応可能なパラメータを用いることで、これらの制限のいくつかに対処します（図5.1.2）。この方法により、正確で局所的なメッシュ調整が可能になります。他の主要な手法と比較して、FlexiCubesはメッシュ再構築のために微分可能な等値面を利用する手法よりも、基礎となる真実に非常に近いメッシュ抽出を実現します（図5.1.3）。

FlexiCubesの表面再構築 サンプル 出典：Nvidia, 2023

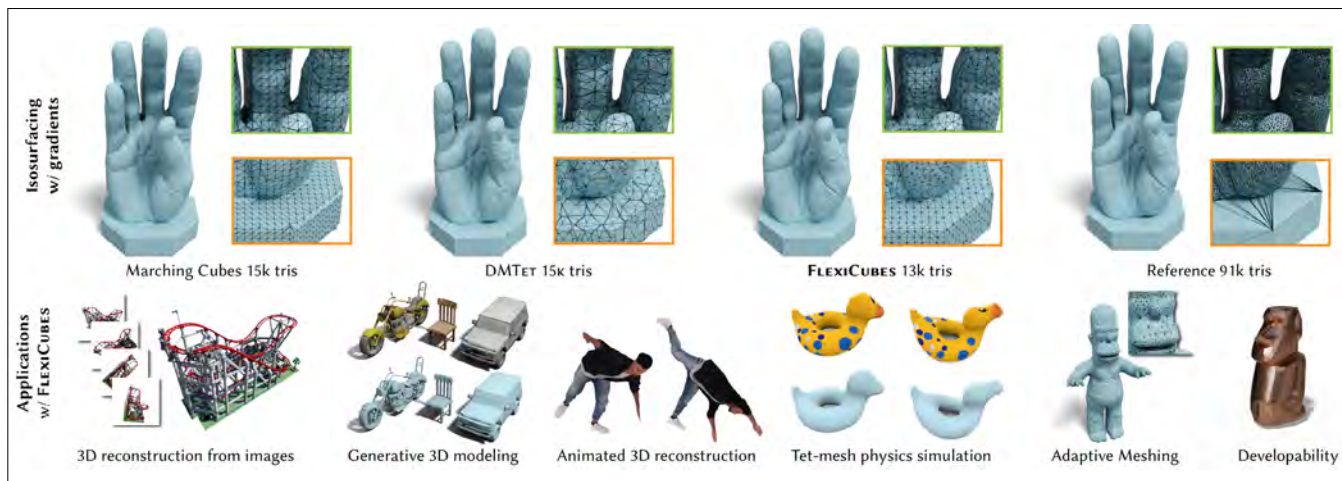


Figure 5.1.2

3Dメッシュ再構築に関する定量的結果を選択

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

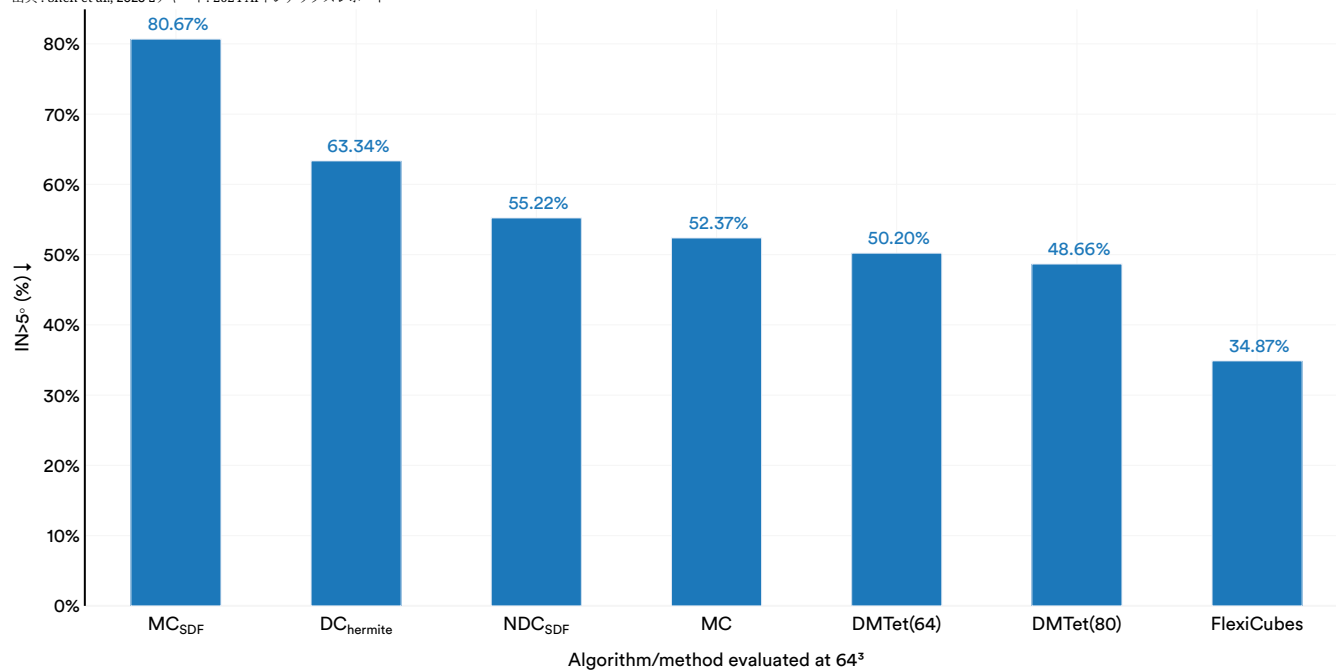


Figure 5.1.3

シンボット

AI 駆動のロボット化学者のための

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からなる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は確立された基準と同等またはそれを超える収率を持つ合成レシピを開発することができます（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では、Synbotは変換収率が基準を超える複数の合成式を開発しました。

Synbotの設計 出典: Ha

et al., 2023

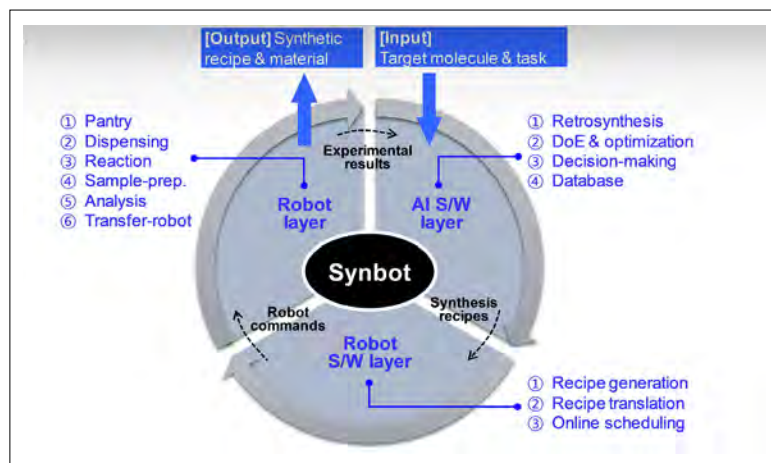


図 5.1.4

中 80%の基準範囲を超え、合成を大幅に短時間で完了しました（図 5.1.5）。Synbotによる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対基準 出典: Ha et al., 2023 □チャート: 2024 AIインデックスレポート

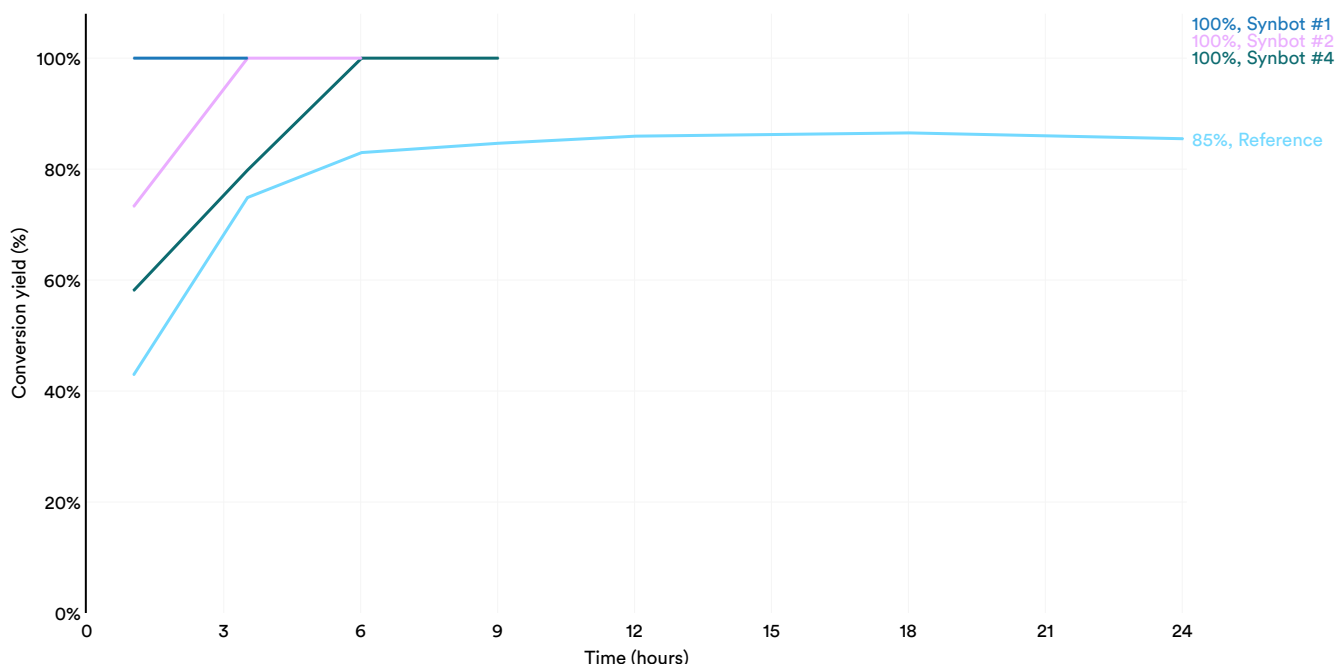


Figure 5.1.5

グラフキャスト

グラフキャストによるより正確な全球気象予測

GraphCastは、1分未満で非常に正確な10日間の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）。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、気温、風速、大気条件を予測します。

など。図 5.1.7は、**GraphCast**と現在の業界最高水準の天気シミュレーションシステムである高解像度予測（HRES）の性能を比較しています。

GraphCastは、ルート平均二乗誤差が低く、予測が観測された天気パターンにより密接に対応していることを意味します。**GraphCast**は、天気パターンを解釈し、極端な天候イベントへの備えを強化し、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可能性があります。

GraphCast 天気予測 出典: DeepMind, 2023

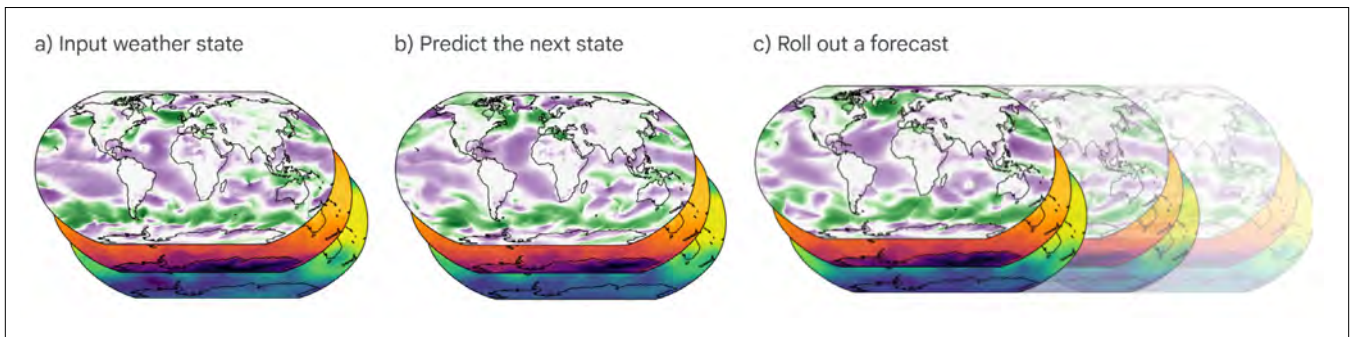


図 5.1.6

10日間のz500予測スキル: GraphCast対HRES 出典: Lam et al., 2023 □チャート:

2024 AIインデックスレポート

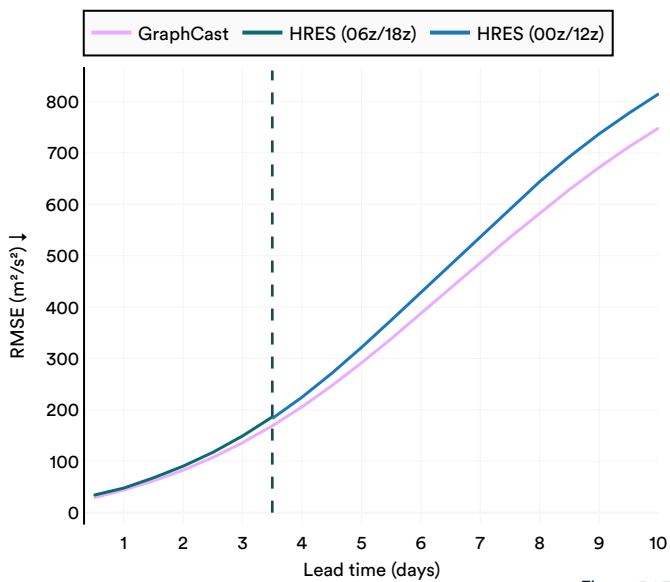


Figure 5.1.7

GNoME

GNoMEを用いた新材料の発見

新しい機能性材料の探索は、ロボティクスや半導体製造を含むさまざまな科学分野の進展にとって重要です。しかし、この発見プロセスは通常、高価で遅いものです。Googleの研究者による最近の進展は、AIモデルの一種であるグラフネットワークが、大規模なデータセットで訓練されることでこのプロセスを加速できることを示しています。彼らのモデルGNoMEは、材料発見の主要な手法であるMaterials Projectを上回り、一はるかに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いた（図 5.1.9および図 5.1.10）。GNoMEのようなAI駆動のプロジェクトの成功は、科学的ブレークスルーを加速するデータとスケーリングの力を強調しています。

サンプル材料構造 出典: Merchant et al., 2023

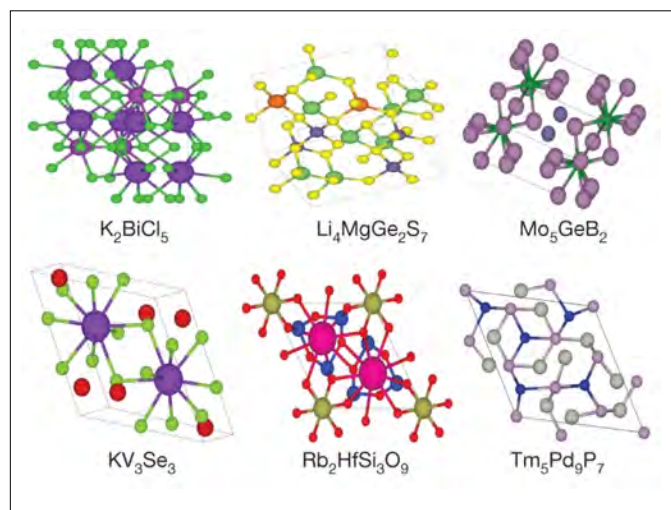


Figure 5.1.8

GNoME対Materials Project: 安定した結晶数 出典: Merchant et al., 2023 0チャート:

2024 AIインデックスレポート

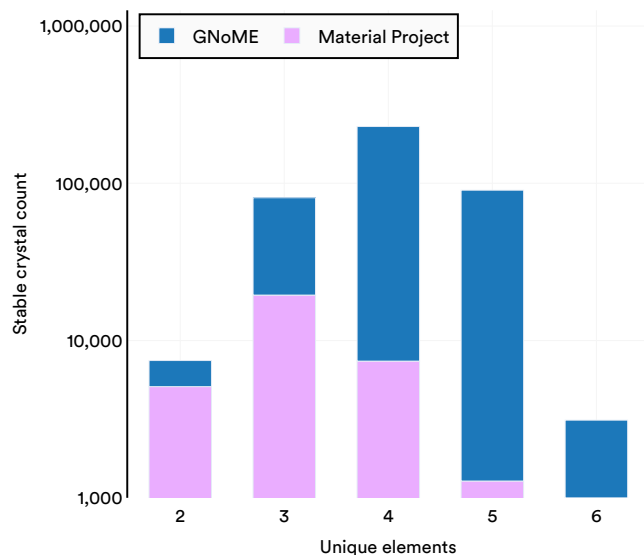


図 5.1.9

GNoME対Materials Project: 異なるプロトタイプ 出典: Merchant et al., 2023 0チャート:

チャート: 2024 AIインデックスレポート

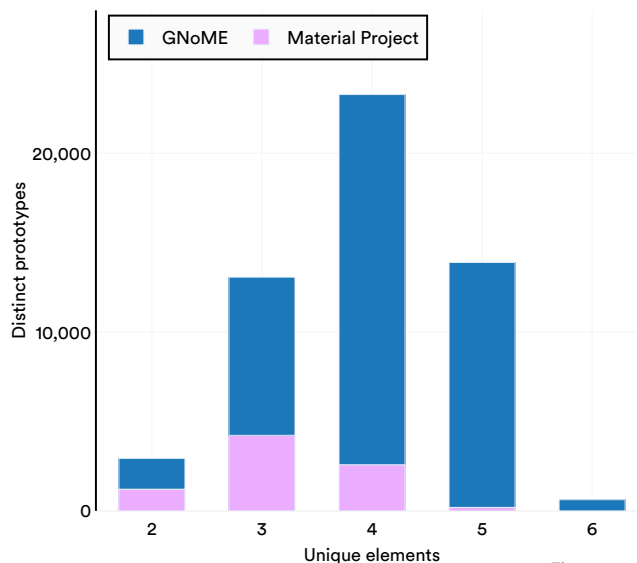


Figure 5.1.10

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、特に予防と緩和のためのインフラが不足している発展途上国では壊滅的な影響を及ぼします。そのため、これらのイベントをさらに早く予測できるより正確な予測方法を開発することは、かなりのポジティブな影響をもたらす可能性があります。

Googleの研究チームは、AIを使用して、無計測流域にも適用可能な非常に正確な水文学的シミュレーションモデルを開発しました。これらの革新的な方法は、極端な洪水イベントの一部を最大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。このAIモデルは、さまざまなリターン期間のイベントにおいて、優れた精度（正の予測の精度）と再現率（関連するすべてのインスタンスを正しく特定する能力）を示し、現代の主要な方法を上回っています（図5.1.11）。このモデルはオープンソースであり、すでに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モデルとGloFASのリターン期間における予測

Source: Nijssen et al., 2023. Copyright 2024 AI Index Report

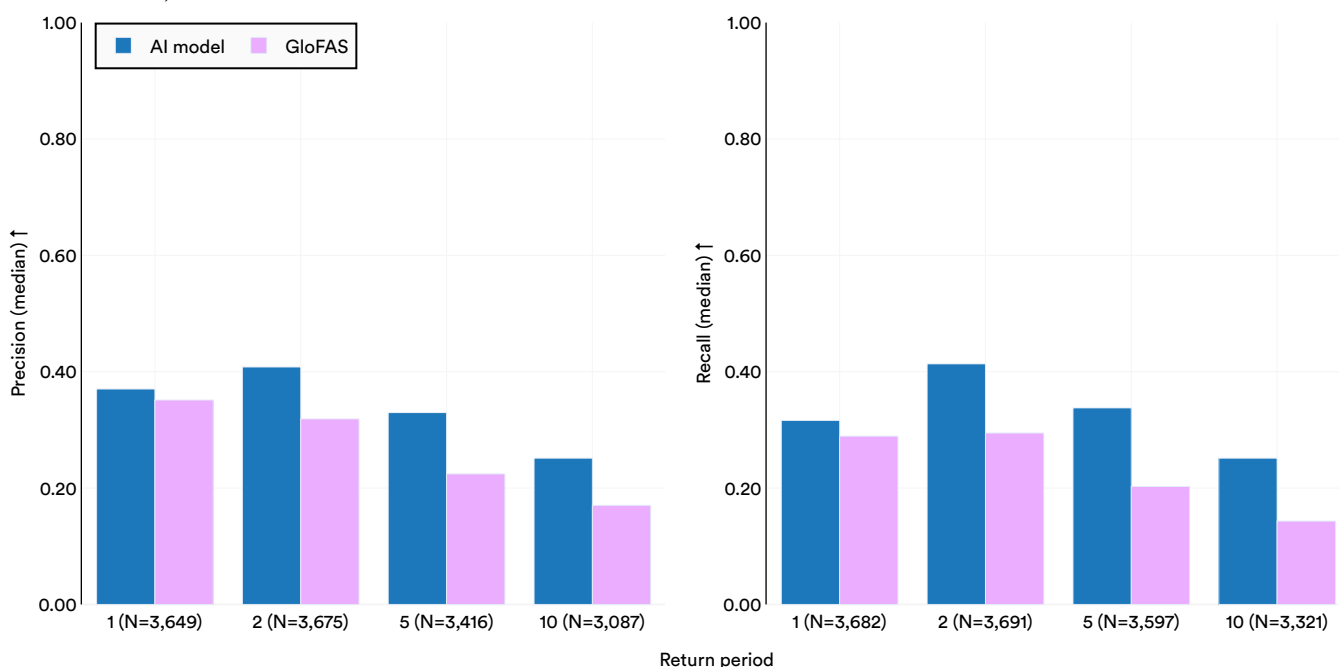


Figure 5.1.11

1 An ungauged basin is a watershed for which there is insufficient stream flow 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、ポリープの検出から診断を支援することまで、医療においてますます価値が高まっています。AIの性能が向上し続ける中で、医療実践へ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、および病院管理を強化することを目的とした新しいAI診断ツールとモデルの開発について強調します。

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定した2023年の重要なAI関連医療のブレー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）。この進展は、以前は多くのスキャンの高度な研究で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高コントラストで明確な脳構造の描写が特徴のT1強調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促進します。SynthSRを使用した実験は、スキャン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を示しており、SynthSRが高解像度T1スキャンによって生成された画像に非常に似た画像を生成す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スキャンが真実の観察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、神経科学研究および臨床診断を促進します。

SynthSR生成 出典: Iglesias et al., 2023

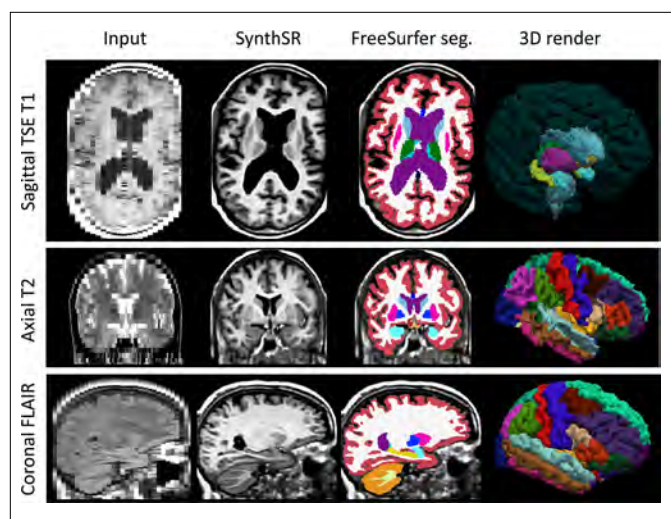


図 5.2.1

SynthSRの選択された脳領域における真のボリュームとの相関 出典：Iglesias et al., 2023 | チャート：2024 AIインデックスレポート

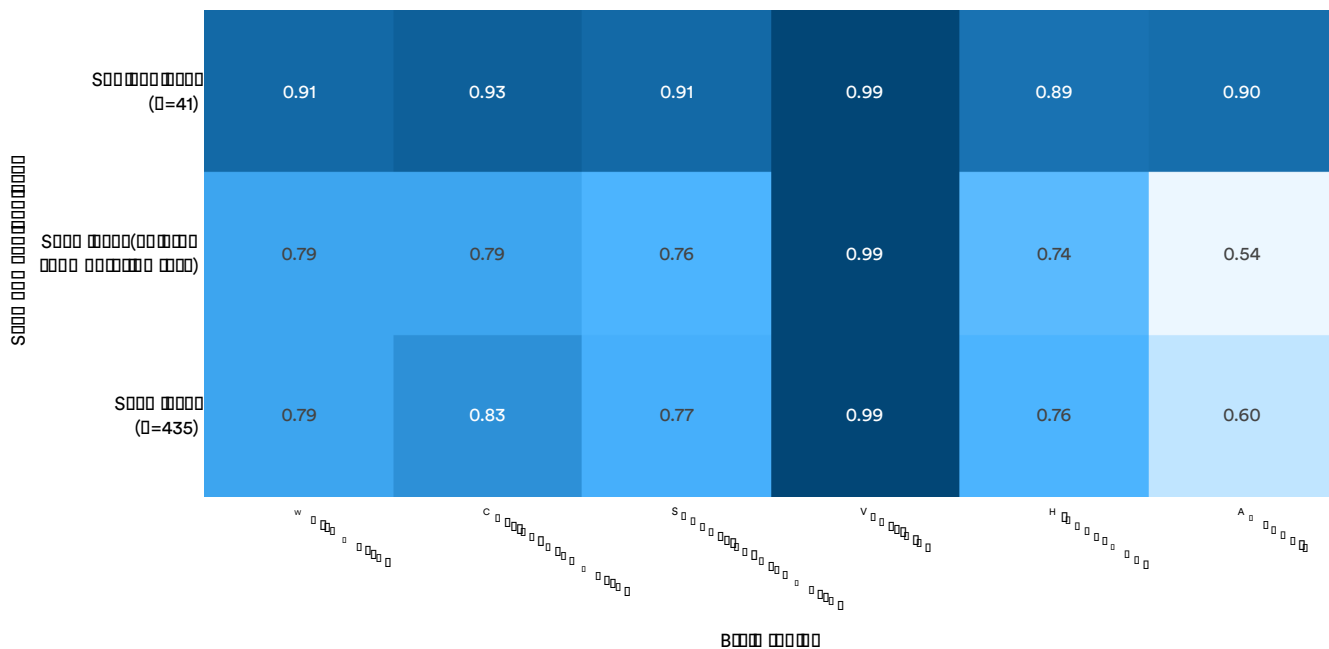


Figure 5.2.2

カップルプラズモニック 赤外線 センサー

神経変性疾患の検出のためのカップルプラズモニック 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特定に依存しています。質量分析法やELISAなどの従来の方法は、タンパク質レベルの定量に焦点を当てることができるため有用ですが、構造状態の変化を識別することはできません。今年、研究者たちは、表面増強赤外吸収 (SEIRA) 分光法を使用したAI結合プラズモニック 赤外線 センサーと免疫アッセイ技術 (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。実際のフィブリルの割合とAIシステムによって行われた予測を比較したテストでは、予測の精度が実際に報告された割合と非常に密接に一致することがわかりました (図 5.2.4)。

ImmunoSEIRA 検出原理とセットアップ 出典: Kavungal et al., 2023

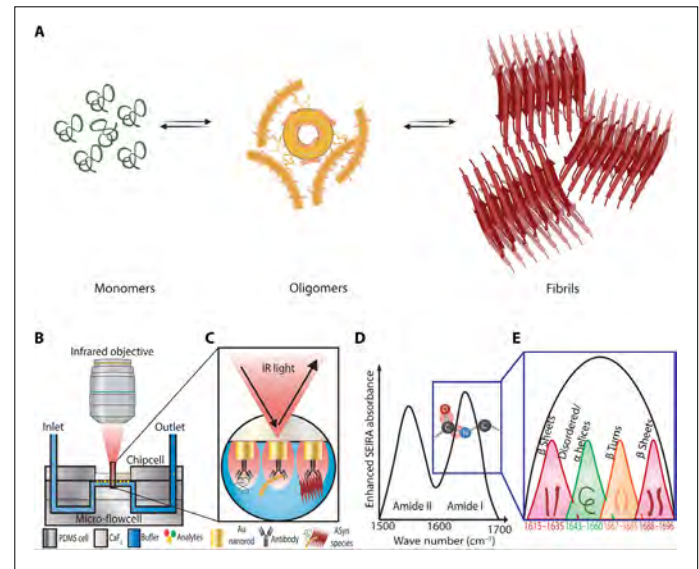


Figure 5.2.3

深層ニューラルネットワークによる予測と実際のフィブリルの割合のテストサンプル 出典: Kavungal et al., 2023 図チャート:

2024 AIインデックスレポート

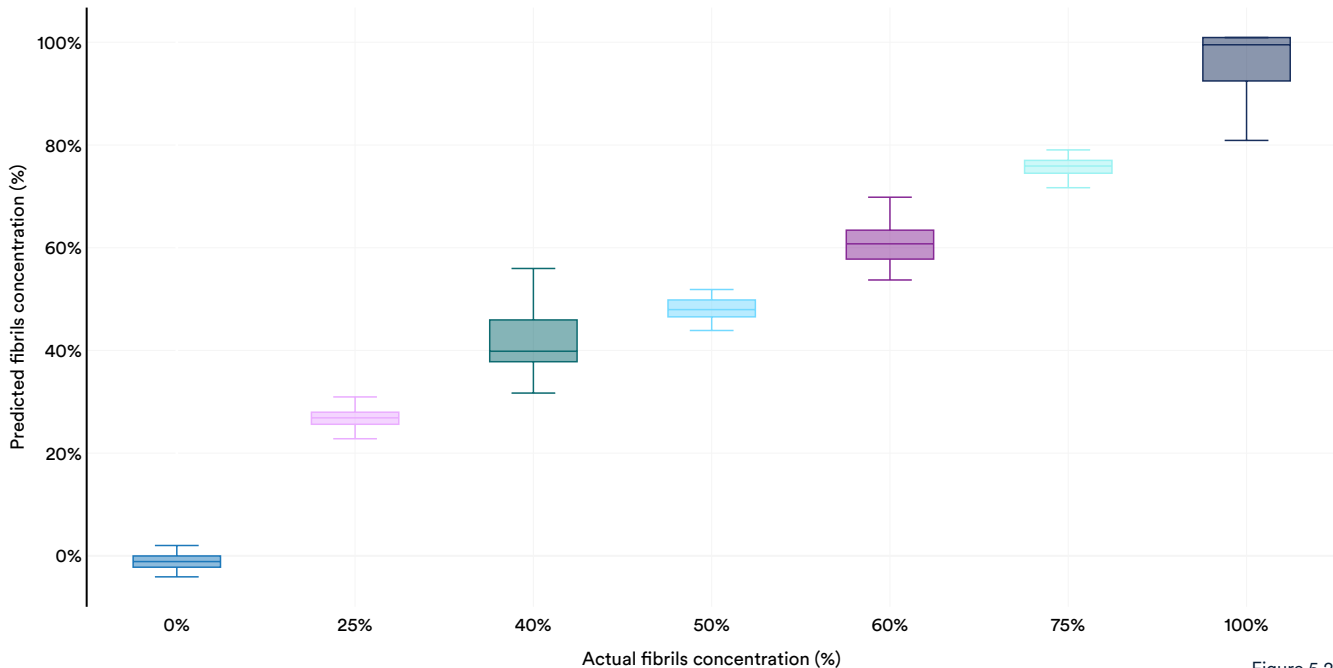


Figure 5.2.4

EVEscape

パンデミック 対策のためのウイルス 進化の 予測

ウイルスの変異を予測することは、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。従来の方法は、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、データが不足しているパンデミックの初期段階では課題に直面します。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。

ウイルスの進化（図 5.2.5）。EVEscapeは、現在の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、従来の実験室研究の46.2%および32.3%を上回り、以前のモデルの24%の変異予測をも上回る性能を示しています（図 5.2.6）。この性能は、EVEscapeが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なる可能性を強調しています。

EVEscapeの設計 出典:

Thadani et al., 2023

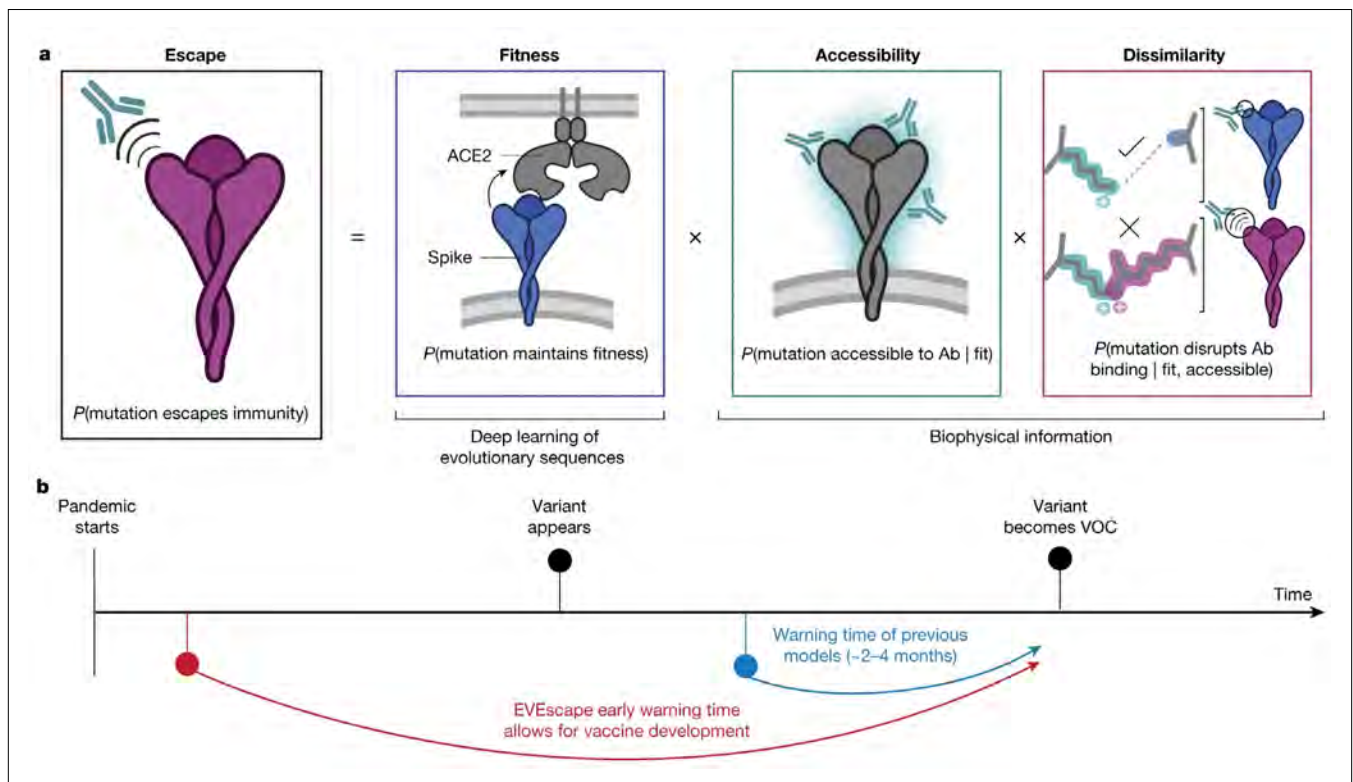


図 5.2.5

EVEscapeと他のモデルのSARS-CoV-2 RBD 変異予測に関する比較 出典: Thadani et al., 2023 | チャート: 2024 AIイン

デックスレポート

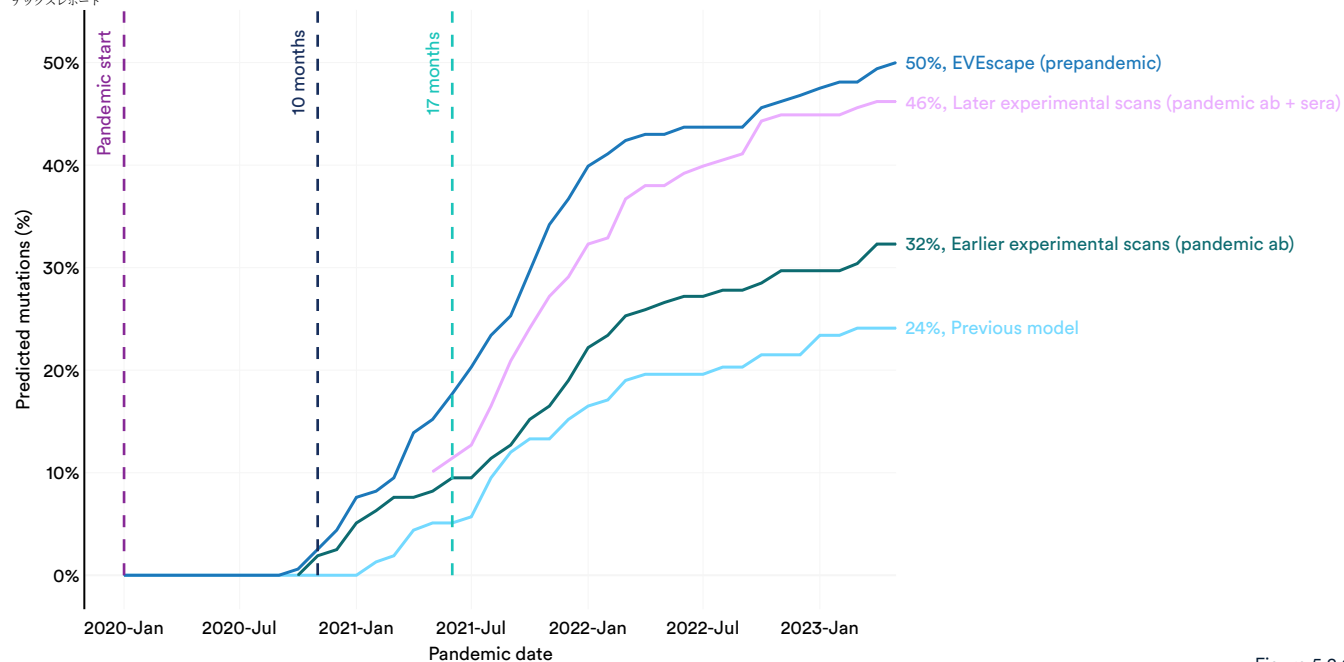


Figure 5.2.6

アルファミスセンス

AI 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につながるのかを完全には理解していません。数百万の可能な遺伝子変異の中から、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子の変化であり（図5.2.7）、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、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分類されました（図5.2.8）。対照的に、人間のアナテーターはすべてのミスセンス変異の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) 出典：Google DeepMind, 2023

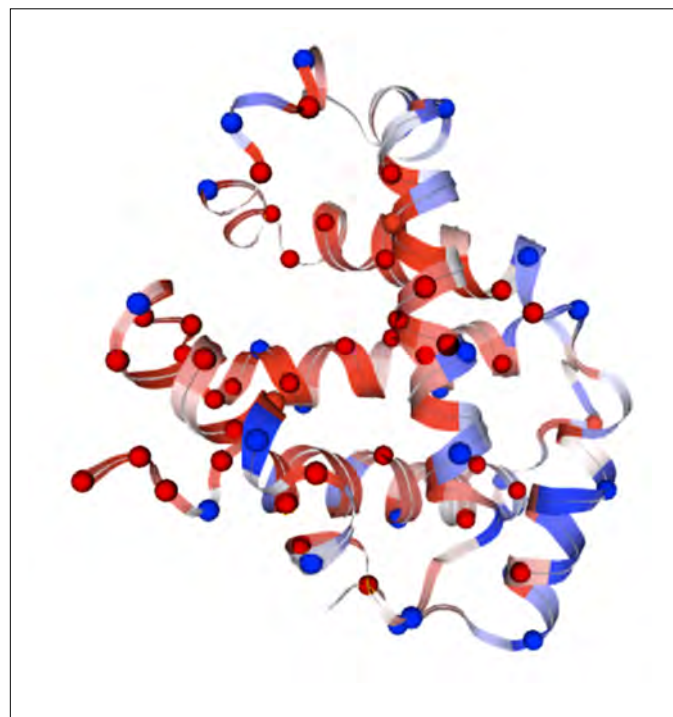


図 5.2.7

AlphaMissenseの予測

出典：Google DeepMind, 2023
チャート：2024年AIインデックスレポート

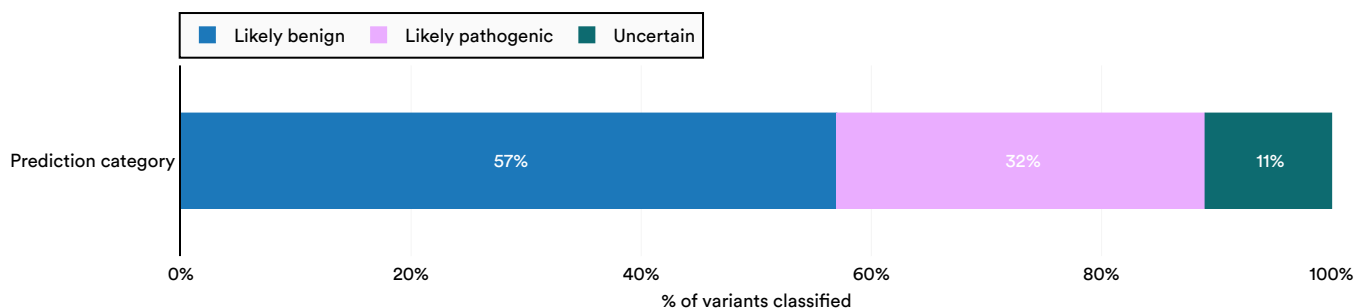


Figure 5.2.8

ヒトパンゲノムリファレンス

AIを使用してヒトゲノムをマッピングする

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年に発表され、2022年に更新されました。しかし、その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、多様な祖先グループを完全にマッピングしていませんでした。したがって、既存のゲノム参照の下では、特定の人々のグループで病気を検出したり治療法を見つけたりすることは困難です。

2023年、119人の科学者からなるヒトパンゲノム研究コンソーシアムは、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を開発しました（図5.2.9）。研究者たちは驚異的な精度を達成し、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 出典: Google Research, 2023

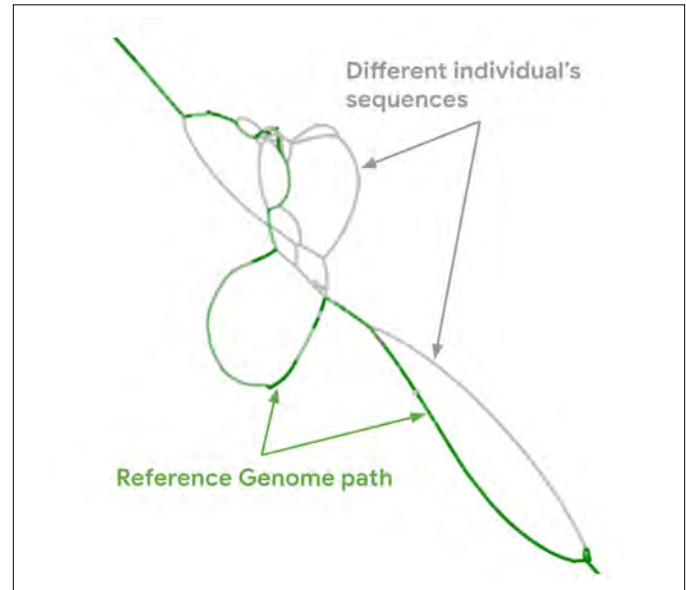


Figure 5.2.9

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の結果 出典: Liao et al., 2023

図5.2.10: 2024 AIインデックスレポート

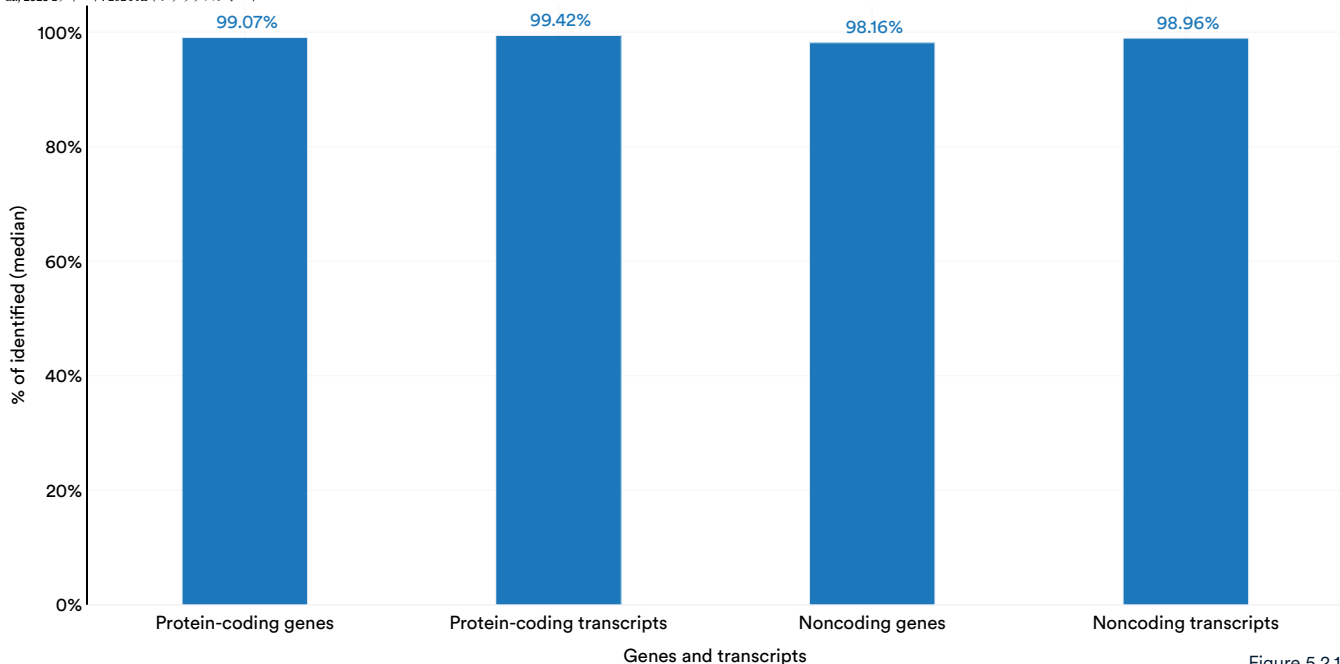


Figure 5.2.10

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範囲を判断することを含みます。

MedQA

2020年に導入されたMedQAは、医療専門のボード試験から派生した包括的なデータセットで、医師に挑戦するために設計された60,000以上の臨床問題の特徴としています。

著しい改善を見せており、最先端のシステムである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の発足以来、このベンチマークにおけるAIの能力はほぼ3倍に増加しており、臨床知識を持つAIシステムの急速な改善を示しています。

MedQAベンチマークにおけるAIのパフォーマンスは

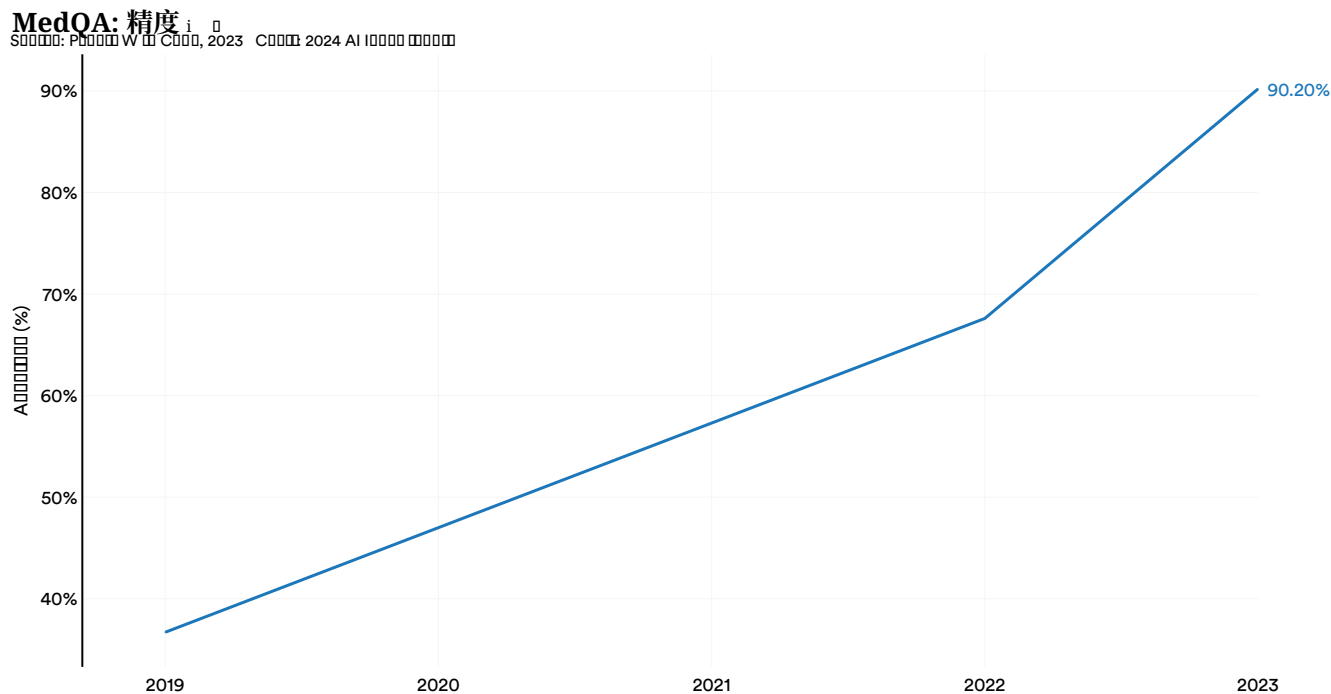


Figure 5.2.11

注目の研究:

GPT-4 Medprompt

LLMは印象的な一般知識を示しますが、医療に関する質問に答えるなどの専門的な知識を発揮するためには、かなりのファインチューニングが必要であると一般的に考えられています。ファインチューニングは、ドメイン特有のデータでLLMをトレーニングすることを含みます。

2023年末のマイクロソフトの研究は、この仮定を覆しました。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MultiMedQAベンチマークスイートにおいて驚異的なパフォーマンスを達成するように導きました。このスイートは、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を含むいくつかの著名な医療ベンチマークの選択肢セクションで、2022年のトップモデルであるFlan-PaLM 540Bのパフォーマンスをそれぞれ3.0、21.5、16.2パーセントポイント上回りました。また、当時の最先端であったMed-PaLM 2のパフォーマンスも上回りました（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モデルでした。このブレイクスルーは、GPT-4 Medpromptの卓越した、そして臨床的に有用な医療能力を強調するだけでなく、モデルを専門分野に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、有望な代替戦略であることが示されています。

GPT-4とMed-PaLM 2の医療質問への回答 出典: 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

 Expert

Let's solve this step-bystep, 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.

Figure 5.2.12

注目の研究: GPT-4 メドプロンプト (続 き)

マルチメドQAサブベンチマークにおけるモデルのパフォーマンス 出典: Nori et al.,

2023 年チャート: 2024 AI インデックスレポート

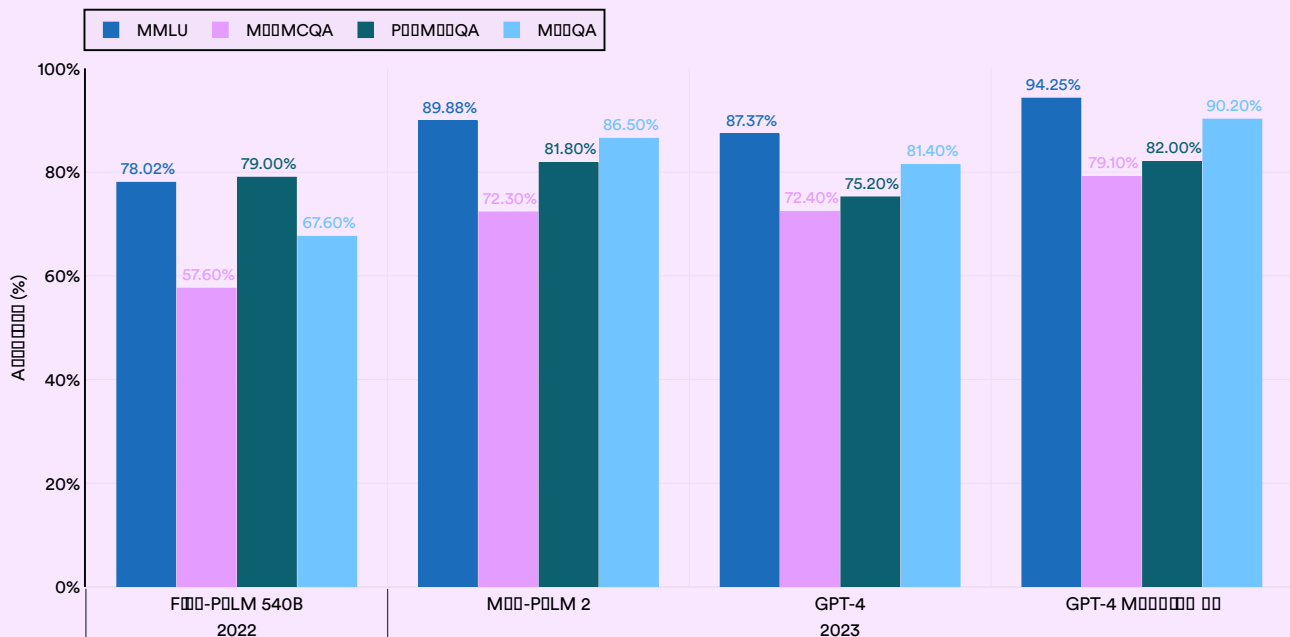


Figure 5.2.13

注目の研究:

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使用するために利用可能ではありません。2023年の新しい研究は、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は特に有望な存在として際立っています。このモデルは、MedQAベンチマークで70.2%の尊敬すべき精度を達成しています。これは、GPT-4 MedpromptやMed-

PaLM 2（どちらもクローズドモデル）よりは劣りますが、2023年の最先端の結果に対して重要な改善を示しており、Llama 2のような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でのスコアは、オープンソースモデルによって達成された中で最高のものです。医療AIがその最大の潜在能力に達するためには、その能力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は前向きな一步を示しています。

選択したモデルのMedQAにおけるパフォーマンス

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

モデル	リリース年	ベンチマーク	MedQAスコア
GPT-4 Medprompt	2023	C	90.20%
Med-PaLM 2	2023	C	86.20%
MediTron-70B	2023	O	70.20%
Med-PaLM	2022	C	67.20%
Llama 2	2023	O	63.80%

Figure 5.2.14

診断

AI tools can also be 放射線学 や 癌検出 などの 診断目的 で 使用 される

注目の研究:

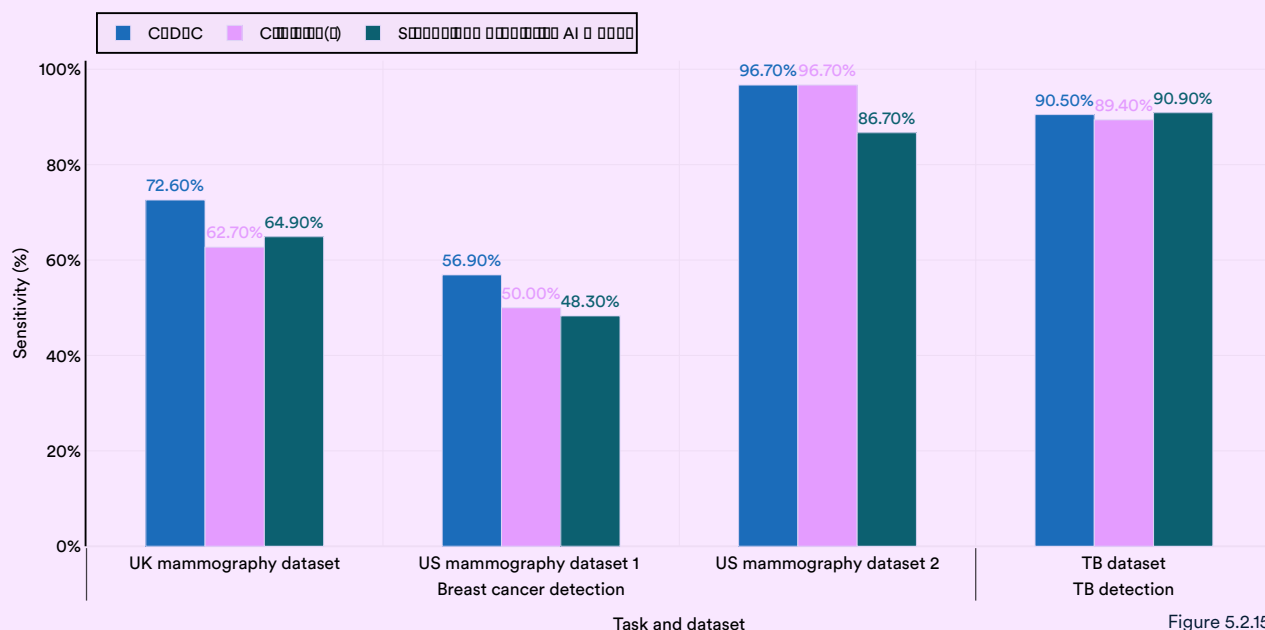
CoDoC

AI 医療画像 システムは 堅牢 な 診断能力 を 示 していますが、臨床医 が 見逃 す 診断 を AI が 見逃 す 場合 や その 逆 も あります。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、研究者 たちは CoDoC（補完性駆動型臨床 ワークフロー への 委譲）を 発表 しました。このシステムは、診断のために AI に 依存 すべき 時 と 従来 の 臨床方法 に 委譲 すべき 時 を 識別 する ように 設計 されています。CoDoC は、感度（病気の ある 個人 を 正しく 特定 する 能力）と 特異度 を 両方 とも 著しく 向上 させます。

特に、4つの 医療 データセット において、CoDoC の 感度 は 臨床医 を 平均 4.5 パーセント ポイント、単独 の AI モデル を 6.5 パーセント ポイント 上回ります（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、単独 の 予測 モデル を 5.7 パーセント ポイント 上回ります。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 発見 は、AI 医療 システム が 臨床 ワークフロー に 統合 され、診断 の 正確性 と 効率 を 向上 させる ことができる ことを 示唆 しています。

CoDoC 対 独立した 予測 AI システム と 臨床 リーダー：感度

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

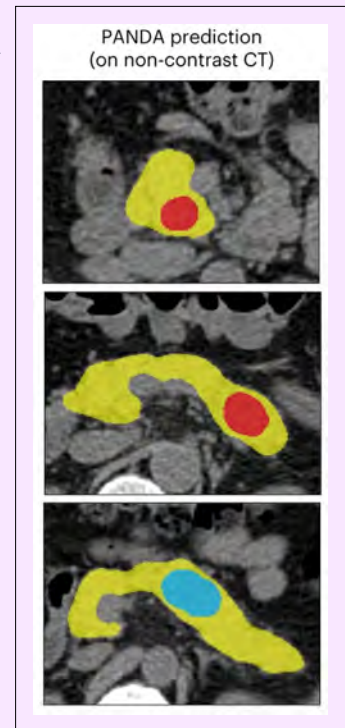


注目の研究： CTパンダ

膵管腺癌（PDAC）は特に致死的な癌であり、外科的介入のために遅すぎる時期に発見されることが多い。無症状の個人における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困難である。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、X線で膵臓病変を効率的に検出・分類することができるAIモデルを作成した（図5.2.16）。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（図5.2.17）。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（図5.2.18）。

PANDAのようなAI医療ツールは、困難な状態の診断において重要な進展を示しており、以前は困難または高価と考えられていたコスト効果的で正確な検出を提供している。

図 5.2.
16PANDA
検出出典：
Cao et al., 2023



出典：Cao et al., 2023 □チャート：2024 AIインデックスレポート PANDAと多施設
検証における平均放射線科医（6,239人の患者）

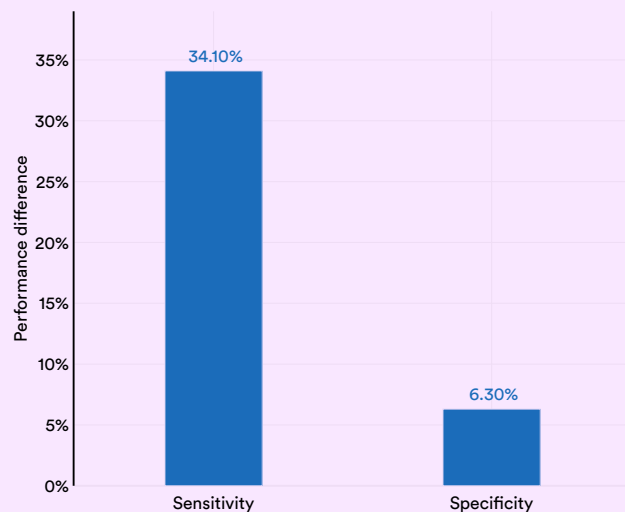


Figure 5.2.17

出典：Cao et al., 2023 □チャート：2024 AIインデックスレポート PANDAの実世界のマル
チシナリオ検証におけるパフォーマンス（20,530人の患者）

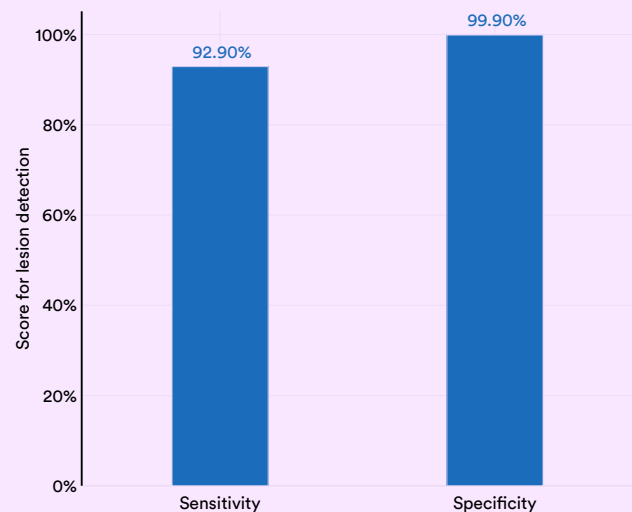


Figure 5.2.18

その他の診断用途

2023年に発表された新しい研究は、AIが他の診断コンテキスト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

図 5.2.19: 図 5.2.19, 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.

図 5.2.19

FDA 承認のAI 関連医療機器

米国食品医薬品局（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によって動作する機器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年に承認された総数から12.1%の増加を記録しました。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。

2012年から2022年までにFDAに承認されたAI医療機器の数 出典: FDA, 2023 図チャート: 2024年AIイン

デックスレポート

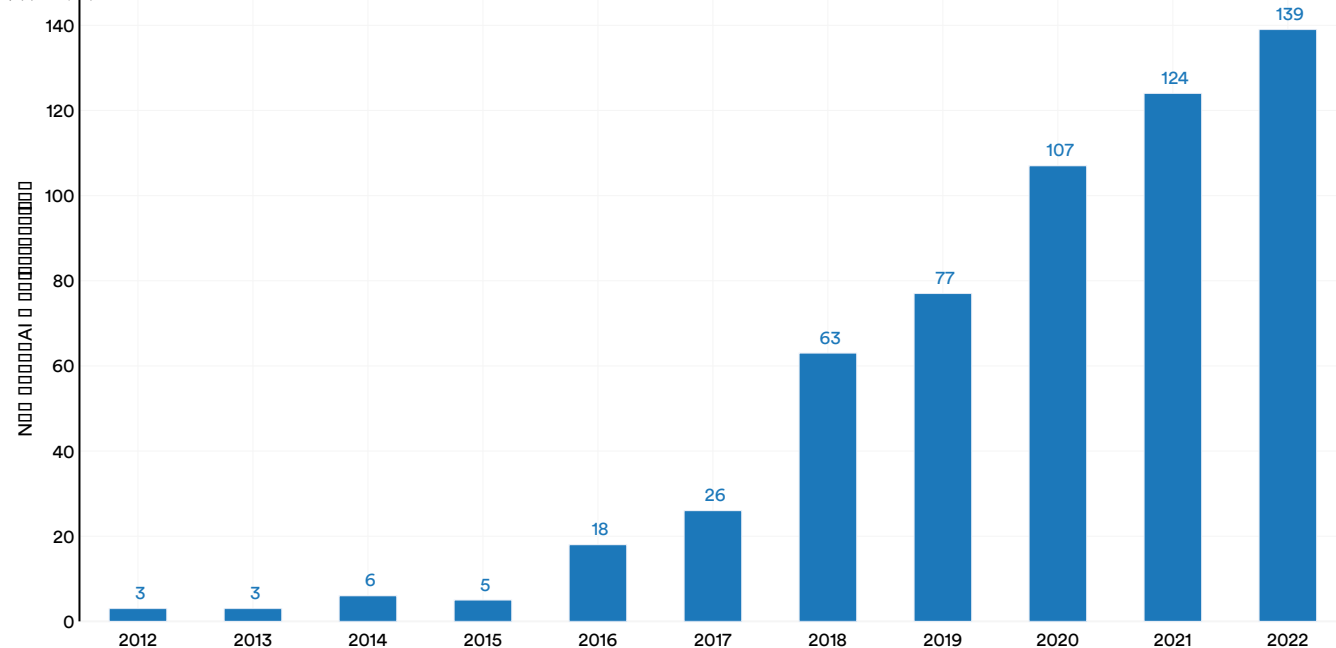


Figure 5.2.20

3 The FDA last updated the list in October 2023. The data is incomplete, which means. As a result, the AI Index is limited to the data provided, and the information is restricted to the data 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%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 年までの専門分野別にFDAに承認されたAI医療機器の数 出典: FDA, 2023 Dチャート; 2024 年 AIインデ

ックスレポート

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
放射線学	2		5		11	15	39	51	94	105	121
心血管				1	4	6	9	12	7	11	10
神経学			1		1	1	4	4		2	2
がん						1	1	1		3	1
皮膚科		1				2	2	1	3		1
小児科		2						2	1		
消化器				1				2			
泌尿器					1		2	1		1	
呼吸器				1			2	1	1	1	2
眼科				1	1		2	1			1
婦人科				1		1			1		
内分泌	1									1	
免疫学											1
その他							1				
診断							1				
治療							1				
予防								1			
その他									1		

図 5.2.21

管理 とケア

AI tools also hold the poten医療管理 の 効率 を 高め、患者 ケア の 標準 を 向上 させるための 重要な 要素 e.

注目の研究: MedAlign

AIが医療分野で著しい進展を遂げているにもかかわらず、MedQAやUSMLEのような既存のベンチマークは、知識ベースの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従事し、業務時間のかなりの割合を管理業務に費やしています。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。医療管理用のLLMをベンチマークおよび微調整するための適切な電子健康記録（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に対処するために、包括的なEHRベースのMedAlignを導入することで進展を遂げました。

983の質問と指示、7つの異なる医療専門分野からの303の臨床医の回答を含むベンチマーク（図 5.2.22）。MedAlignは、最初の包括的な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性（65.0%）を達成し、他のLLMよりも頻繁に好まれました（図 5.2.23）。MedAlignは、医療における管理負担を軽減するためにAIを使用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: Fleming et al., 2023

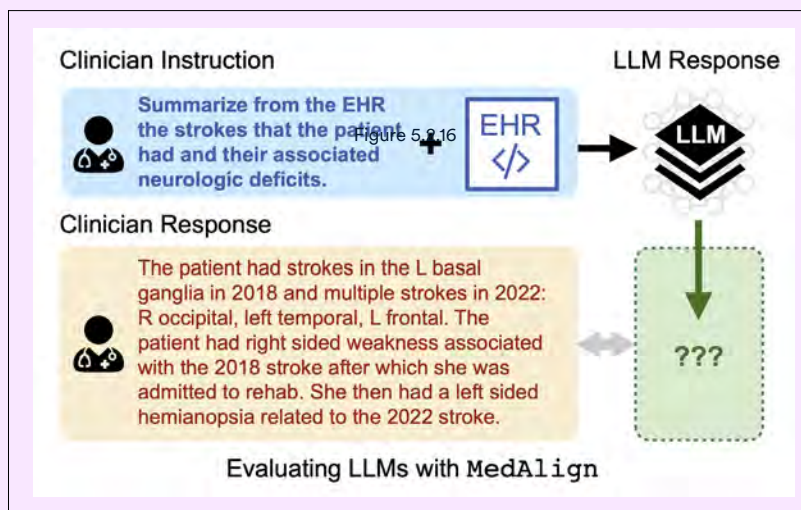


図 5.2.22

注目の研究: MedAlign (続 き)

モデルのパフォーマンス評価: 人間 vs. COMETランク 出典: フレミング et al., 2023 [1]チャート:

2024 AIインデックスレポート

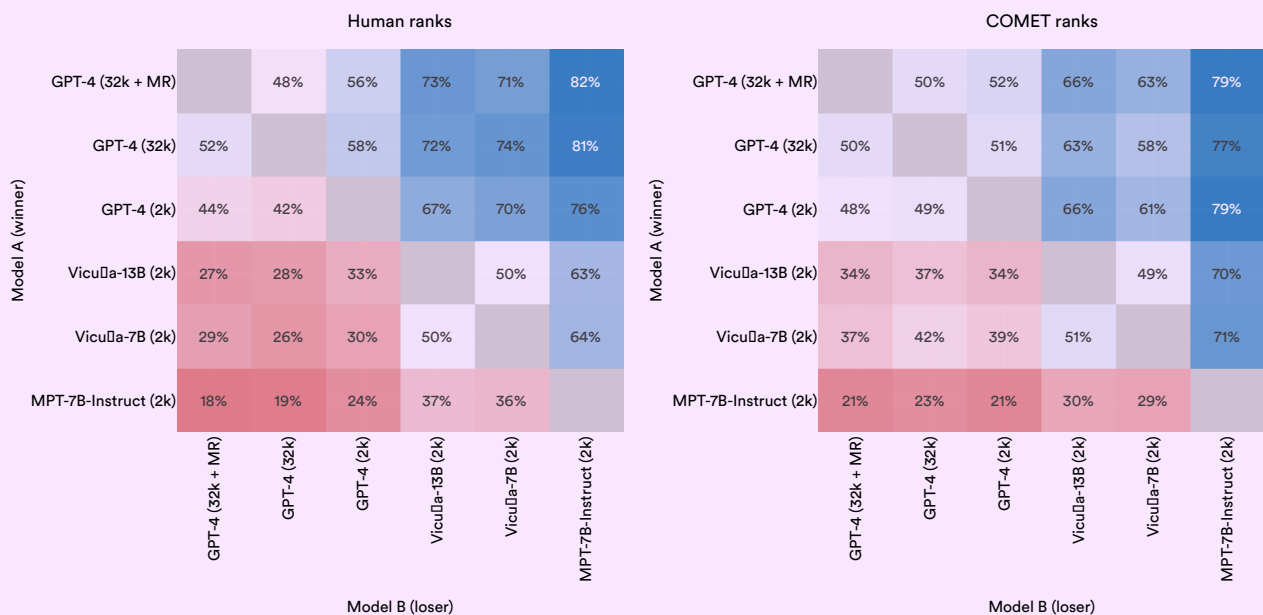
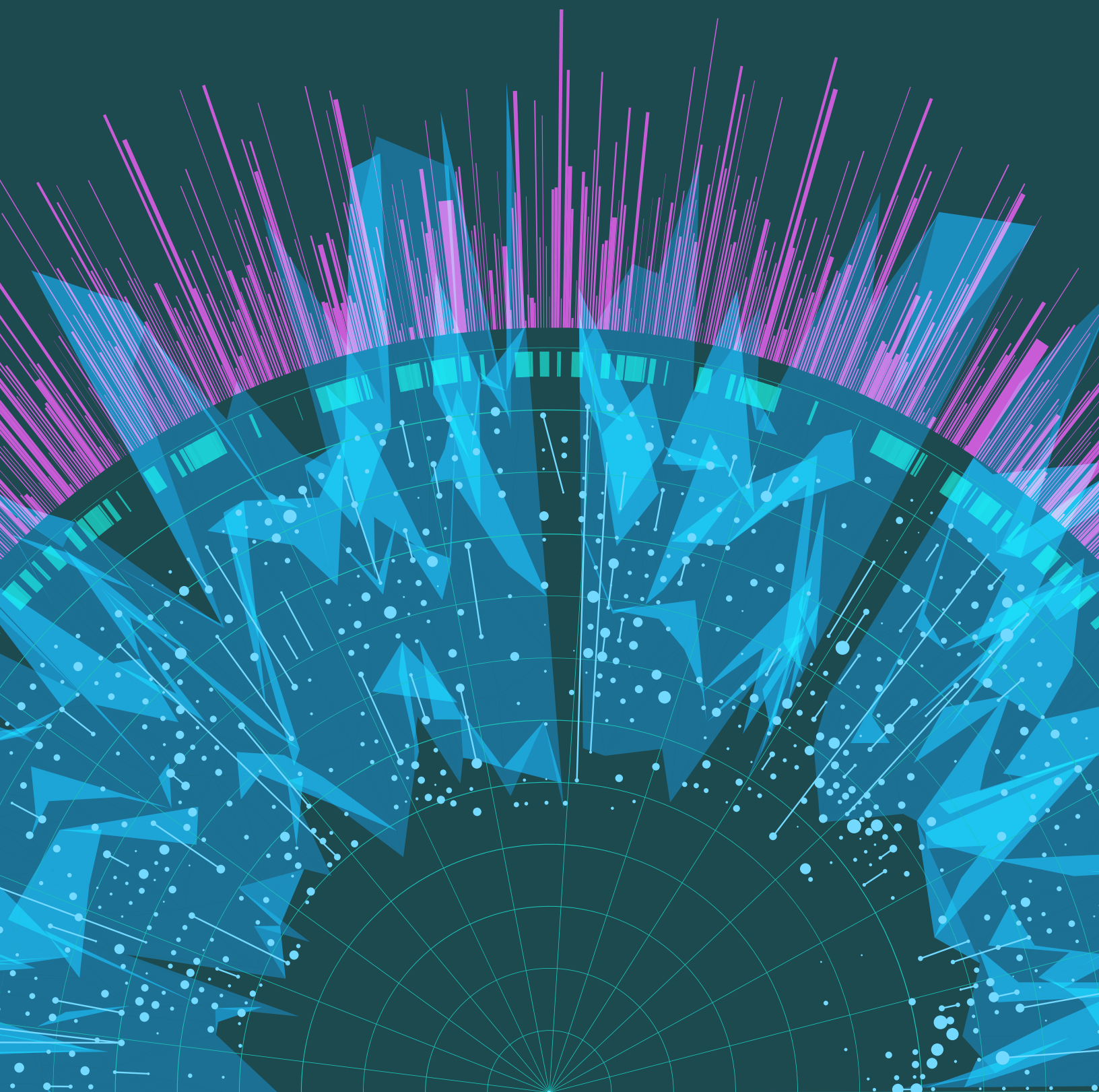


Figure 5.2.23



人工知能 インデックスレポー
ト2024

第6章： 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 CSおよびAI 教育 359	
United States	359
State-Level Trends	359
AP Computer Science	361
Highlight: Access Issues	363
Highlight: ChatGPT Usage Among Teachers and Students	364

ACCESS THE PUBLIC DATA

概要

この章では、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、そしてこれらのトレンドがどのように進化してきたのかに焦点を当てています。AIが教育に与える影響についての懸念が高まる中、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て、アメリカとカナダにおける高等教育のCSとAI教育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CS教育に関するInformatics Europeのデータをレビューします。今年は、AI関連の英語の学位プログラムの世界的な数に関するStudyportalsのデータを含む新しいセクションが導入されます。

この章は、Code.orgからのアメリカの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学士卒業生の数は引き続き増加しており、新しいコンピュータサイエンス修士卒業生は比較的横ばいで、博士卒業生はわずかに増加しています。その間、

新しいアメリカとカナダの学士卒業生の数は10年以上にわたり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択する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および博士卒業生の数はわずかに減少しています。

2. AI博士の産業への移行は加速しています。2011年には、新しいAI博士の約同じ割合が産業（40.9%）と学術（41.6%）で職を得ていました。しかし、2022年までに、卒業後に産業に参加した割合は大幅に増加し（70.7%）、学術に入った割合（20.0%）と比較して大きな差が生じました。過去1年だけでも、産業に向かうAI博士の割合は5.3ポイント上昇しており、大学から産業への脳の流出が激化していることを示しています。

3. 学術界への産業からの学術人材の移行が減少。2019年、アメリカとカナダの新しいAI教員の13%が産業から来ていました。2021年にはこの数字は11%に減少し、2022年にはさらに7%に落ち込みました。この傾向は、産業から学術界への高レベルのAI人材の移行が徐々に低下していることを示しています。

4. アメリカとカナダのCS教育が国際的でなくなる。2022年には、2021年よりも国際的なCSの学士、修士、博士の卒業生が相対的に減少しました。特に修士課程の国際学生の減少が顕著でした。

5. より多くのアメリカの高校生がCSコースを受講するが、アクセスの問題が残る。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学生数は10倍以上に増加しました。しかし、最近の証拠によると、大規模な高校や郊外に住む学生は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が国際的に増加している。英語で提供されるAI関連の高等教育学位プログラムの数は2017年以降3倍に増加し、過去5年間にわたって安定した年次増加を示しています。世界中の大学がより多くのAIに焦点を当てた学位プログラムを提供しています。

7. イギリスとドイツがヨーロッパの情報学、CS、CE、ITの卒業生生産でリードしている。イギリスとドイツは、情報学、CS、CE、情報の学士、修士、博士の新卒業生を最も多く生産しているヨーロッパの国です。人口あたりでは、フィンランドが学士と博士の卒業生の生産でリードしており、アイルランドが修士の卒業生の生産でリードしています。

このセクションでは、コンピュータサイエンスと人工知能における高等教育の概要を提供し、北米とヨーロッパにおける学士、修士、博士などのさまざまな学位の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ンスと人工知能

アメリカ 合衆国 とカナダ

この小節では、アメリカ合衆国とカナダにおけるコンピュータサイエンスおよびAIの高等教育の状況を評価するタウルビー調査からのデータの分析を提示します。この調査は、アメリカ合衆国とカナダの297の博士号授与コンピュータサイエンス学科をカバーしています。¹

コンピュータサイエンスの学士卒業生

過去10年間で、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、2021年から2022年にかけては前年比7.9%の増加を記録し、3倍以上に増加しました（図 6.1.1）。

アメリカ合衆国とカナダの新しいコンピュータサイエンス学士卒業生、2010-22 出典：CRAタウルビー調査、

2023 図チャート：2024 AIインデックスレポート

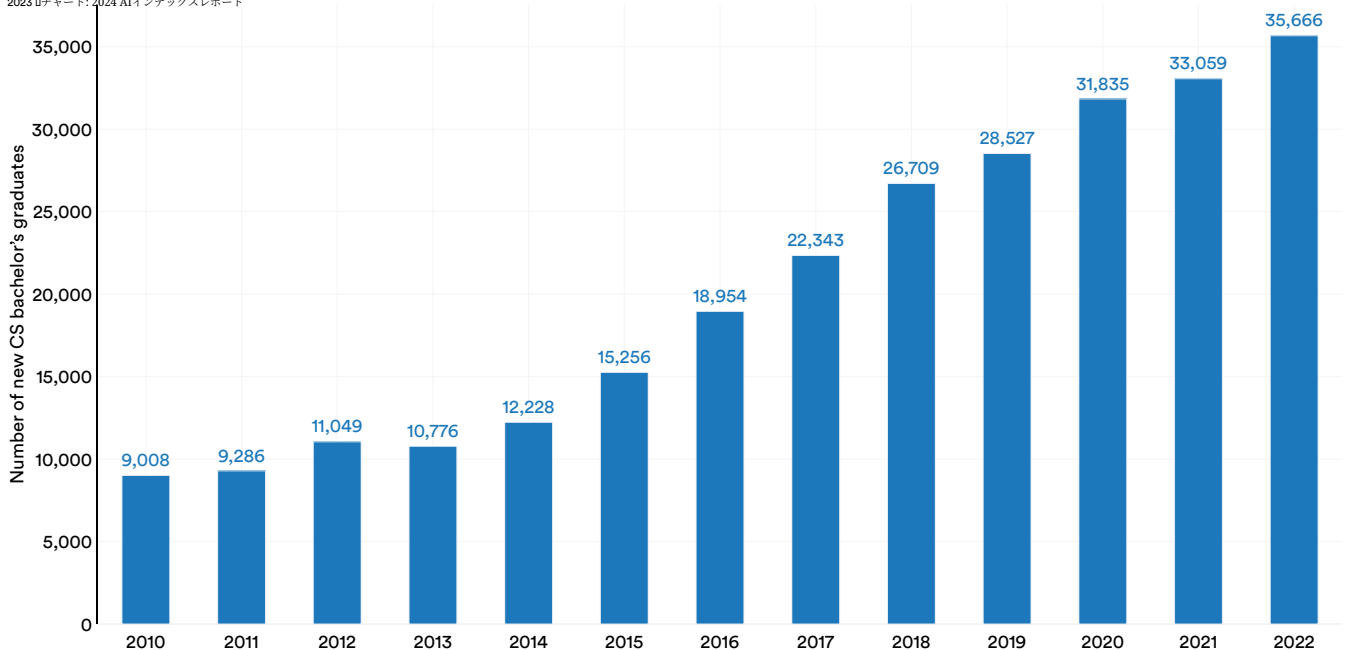


Figure 6.1.1

¹ It is important to note that not all PhD-granting departments targeted in the survey provided responses. Out of the 297 departments targeted, only 182 responded, yielding an overall response rate of 61%.

ほぼ8年ぶりに、アメリカとカナダの大学におけるCS学士卒業生の中で国際学生の割合が減少し、2021年の16.3%から2022年には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学習ビザを取得することが難しくなったことを反映している可能性があり、その影響は今になってようやく現れ始めています。

データにおいても、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影響を部分的に受けており、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。

アメリカとカナダにおける新しい国際的なCS学士卒業生（総数の%）、2010-22 出典：CRAタウルビー調査、2023 0チャート：2024 AIインデックスレポート

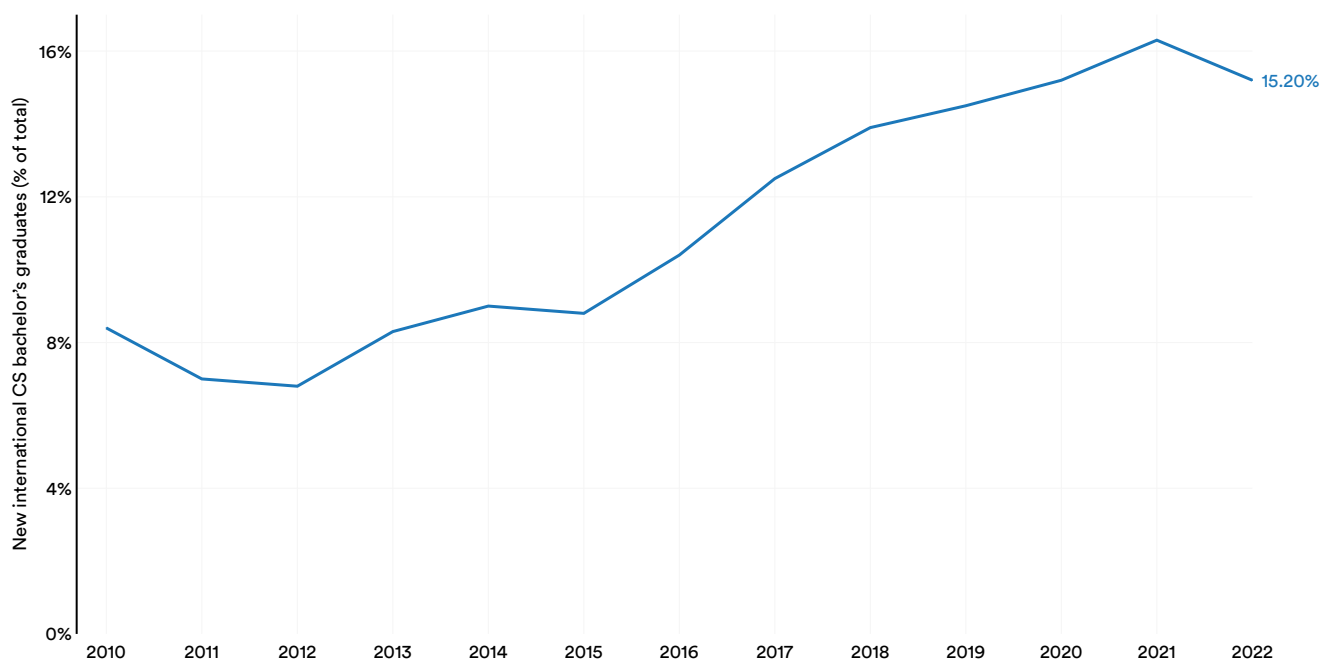


Figure 6.1.2

CS 修士卒業生

AIコースは一般的にCS修士課程に含まれています。過去10年間でアメリカとカナダの大学からの新しいCS修士卒業生の総数は2倍以上に増加しましたが、

この数値は2018年以降横ばいになっており、昨年は2.5%減少しました（図6.1.3）。この横ばいは、次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CS修士卒業生、2010–22 出典：CRAタウルビー調査、2023年チャート：2024 AIインデックスレポート

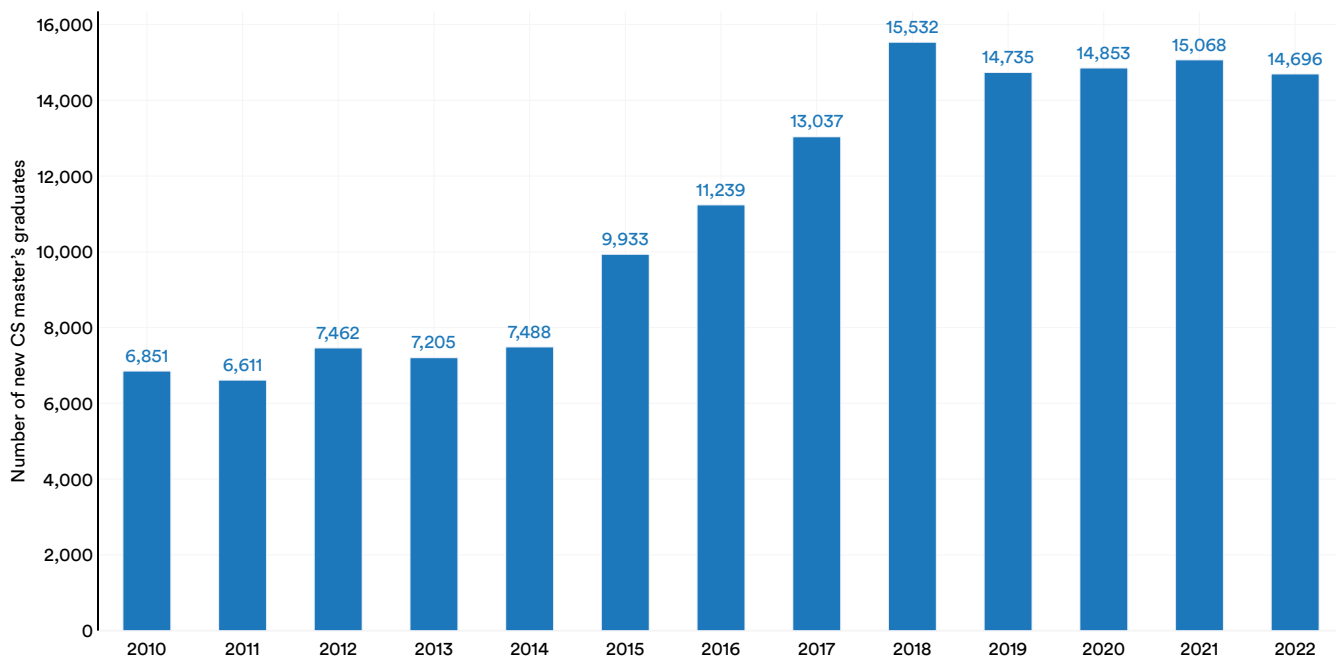


Figure 6.1.3

2022 年、アメリカとカナダの大学では国際的なCS 修士学生が顕著な減少を経験しました。この下降傾向は2017 年頃から始まりましたが、昨年は14.8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CS 修士卒業生と国内のCS 修士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的なCS 修士卒業生（総数の％）、2010-22 出典：CRAタウルビー 調査、2023 0チャート:2024 AIインデックスレポート

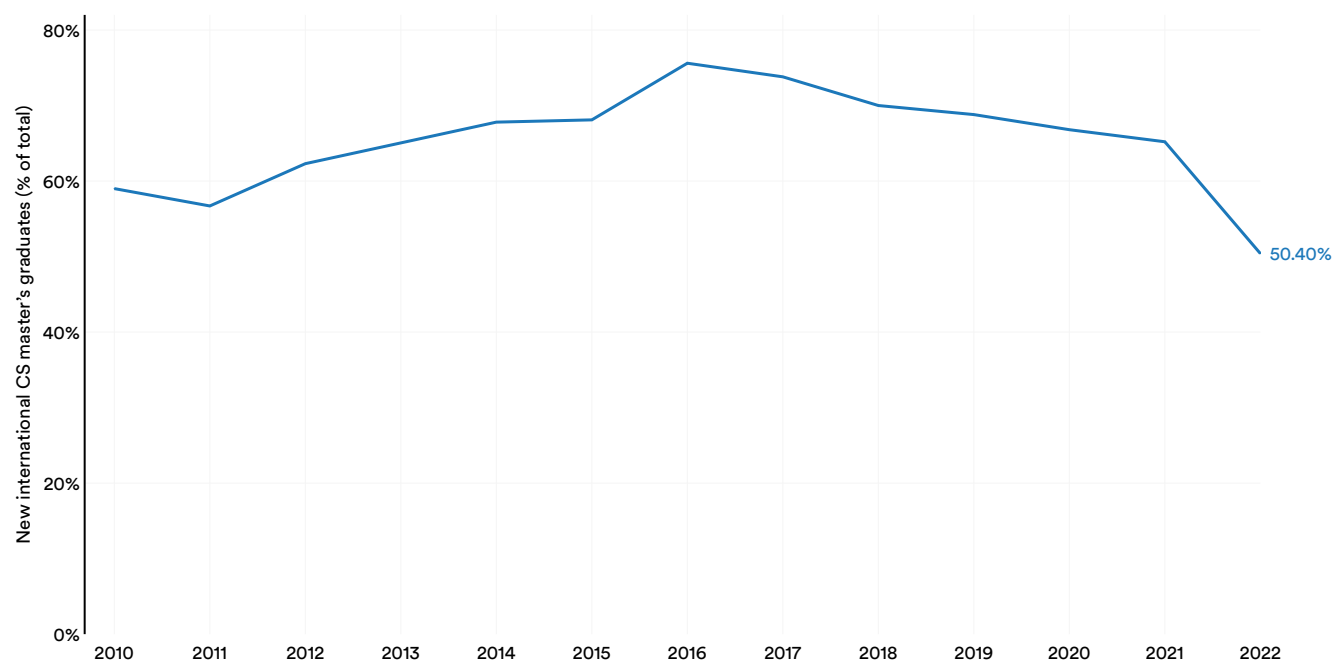


Figure 6.1.4

CS 博士号取得者

10年ぶりに、アメリカとカナダの大学における新しいCS博士号取得者の数が大幅に増加しました。2022年には、CS博士号取得者の数が2,105人に達し、2010年以来の最高値となりました（図6.1.5）。

アメリカとカナダの新しいCS博士号取得者、2010-22 S C R A T I b S 2023 D C h t 2024 A I I d t

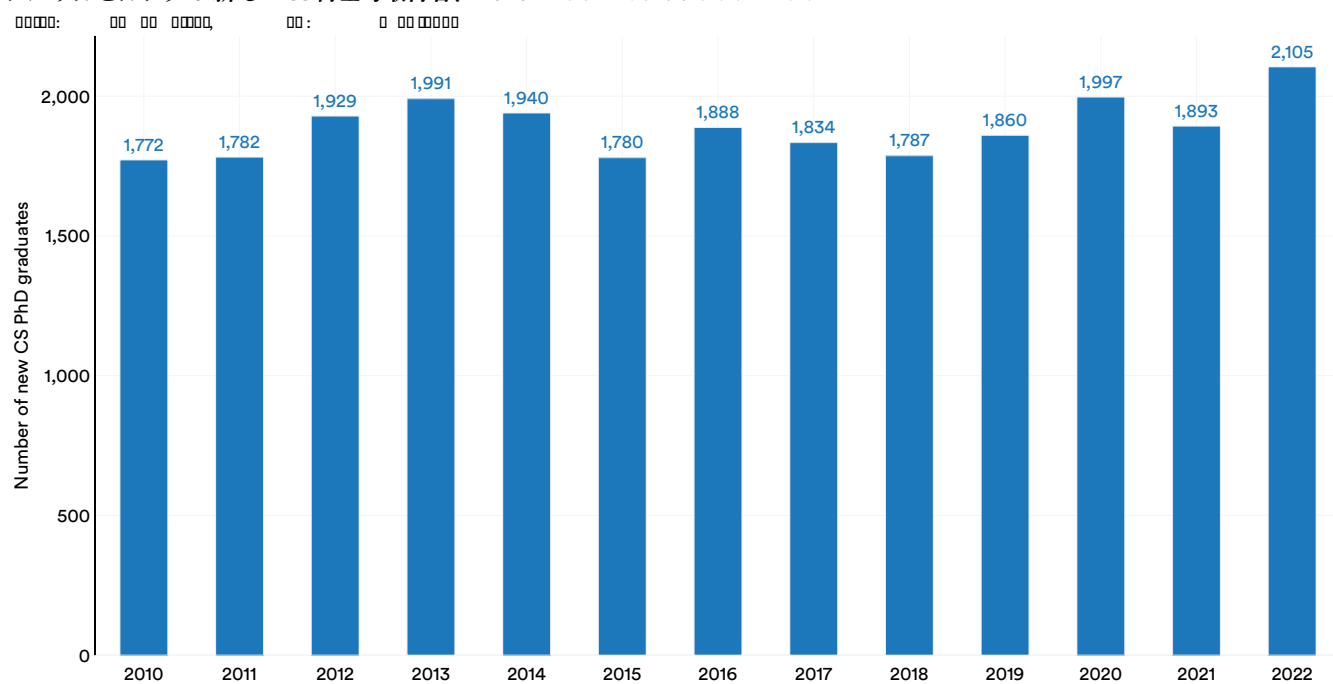


Figure 6.1.5

過去10年間でCS博士号取得者の中で国際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的なCS博士号取得者（総数の%）、2010-22 出典：CRAタウルビー調査、2023年チャート：2024 AIインデックスレポート

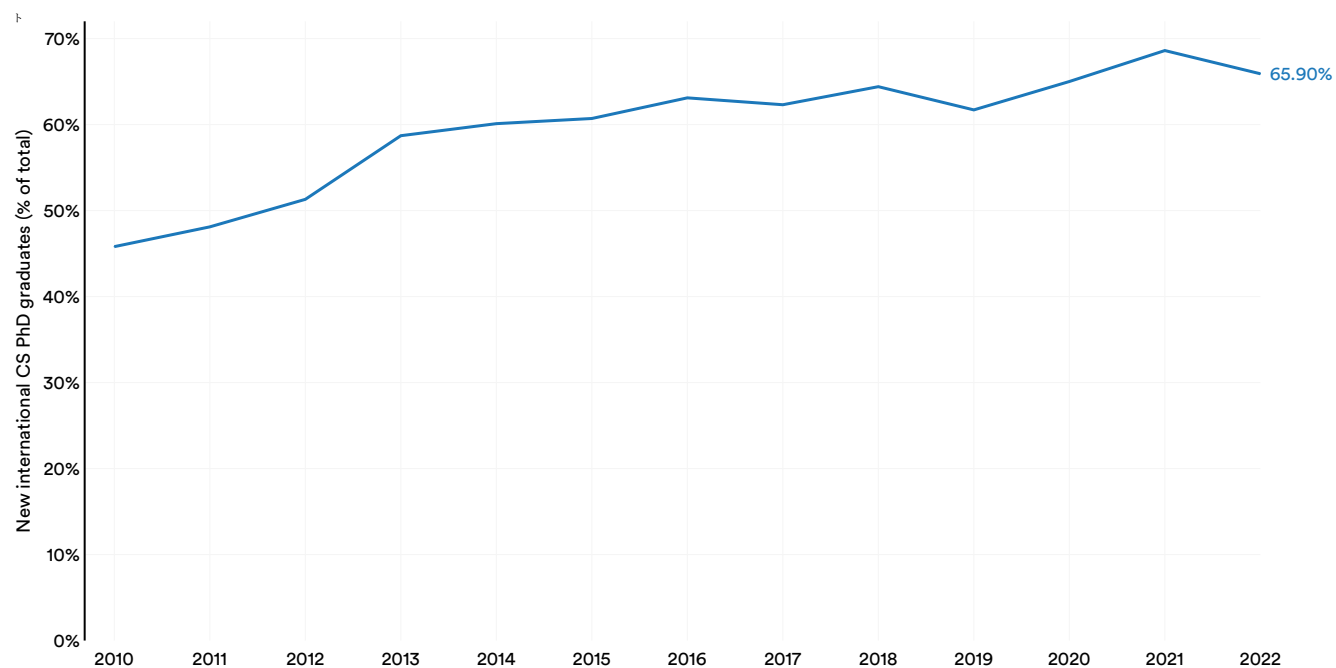


Figure 6.1.6

新たに取得したAI博士号を持つ人々は、卒業後どこで働くことを選ぶのでしょうか? 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、産業界での就職率(40.9%)と学界での就職率(41.6%)はほぼ同じでした。

しかし、2022年までに、卒業後に産業界に参加した割合は大幅に増加し(70.7%)、学界に入る割合(20.0%)と比較されました。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。

アメリカとカナダにおける新しいAI博士号取得者の雇用(全体の%)、セクター別、2010-22

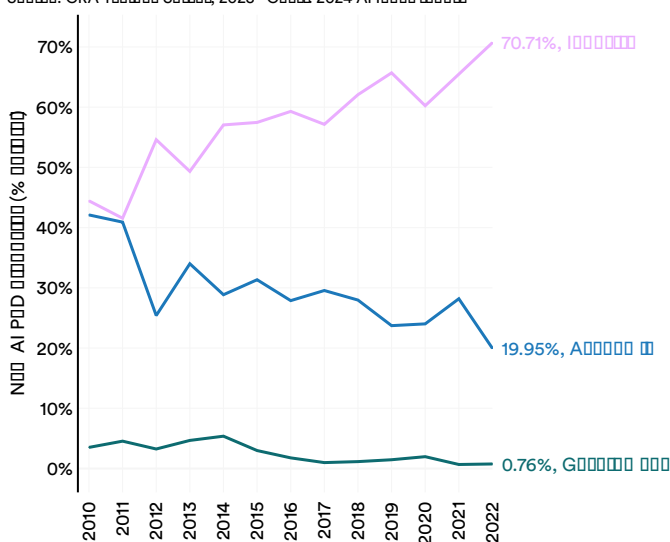


Figure 6.1.7²

アメリカにおける新しいAI博士号取得者の雇用とカナダのセクター別、2010-22

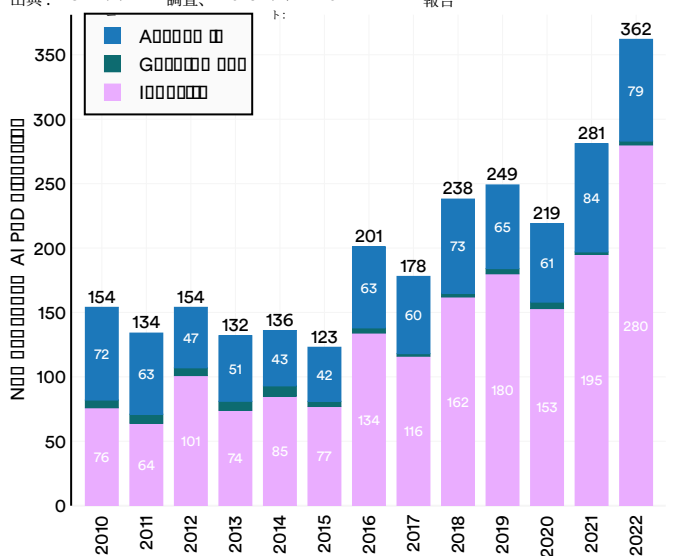


Figure 6.1.8

² The sums in Figure 6.1.7 do not add up to 100、毎年新たに取得したAI博士号の中には、自営業、失業、またはCRAチャートで「その他」の雇用状況を報告する人々のサブセットが存在します。survey. These students are not included in the

CS、CE、および情報学部

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教員に関するデータを調べるのが役立ちます。昨年、アメリカおよびカナダの大学におけるCS、CE、および情報学部の教員数は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。

NアメリカとカナダのCS、CE、および情報学部の教員数、2011-2 出典：CRAタウルビー調査、2023 0チャート：2024 AIインデックスレポー **2**
SDI

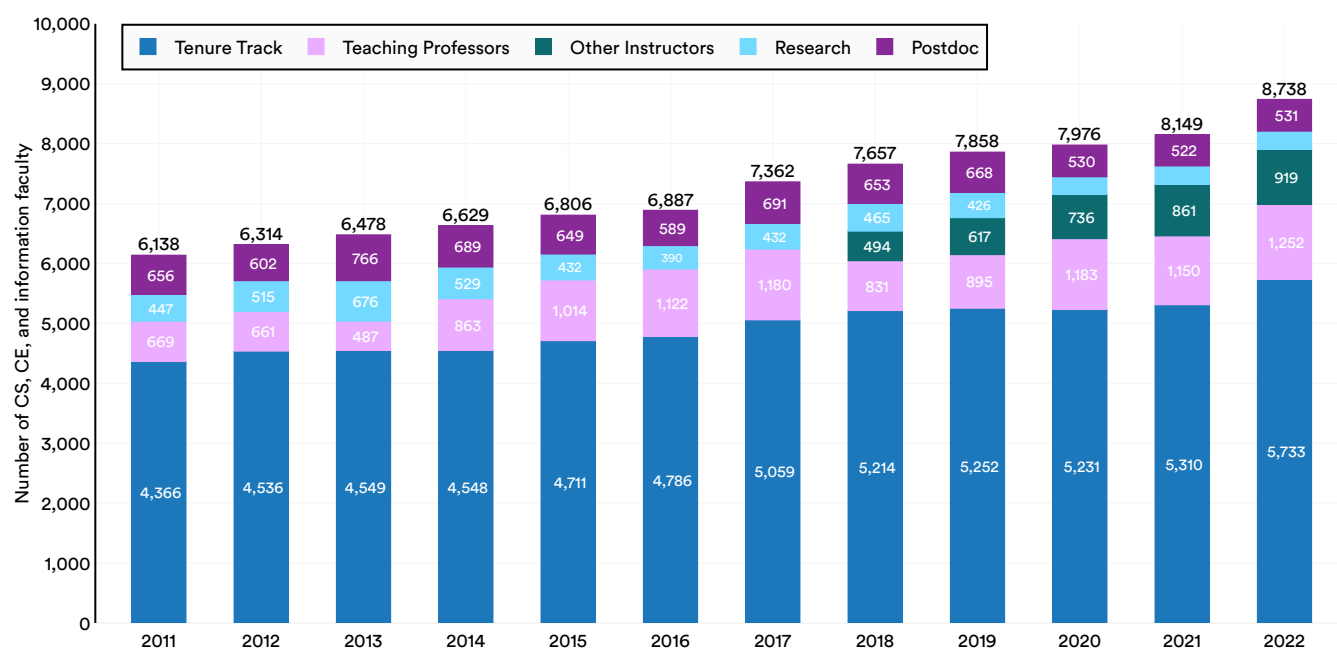


Figure 6.1.9

2022 年、アメリカ合衆国には7,084 人のCS 教員がおり、その大多数（65.7%）がテニュートラックにあります（図 6.1.10）。アメリカのCS 教員の総数は2021 年から4.4%、2011 年から45.0% 増加しました。

アメリカ合衆国のCS 教員数、2011–22

出典：CRA タウルビ調査、2023 ロチャ、2024 AI 10000 報告

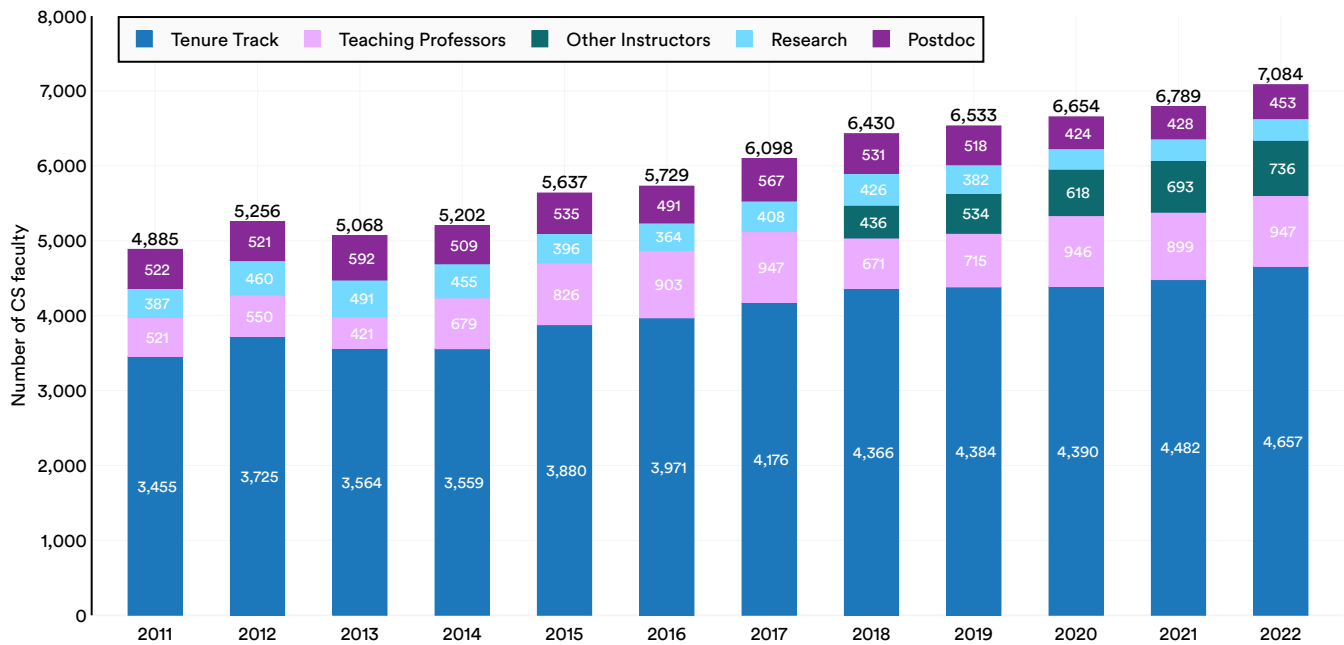


Figure 6.1.10

昨年、北米でCS、CE、および情報分野において915人の新しい教員が採用され、10年ぶりの高水準となりました。これらのポジションのうち455はテニユートラックでした。（図 6.1.11）

アメリカとカナダにおける新しいCS、CE、および情報学部の教員採用、2011-22 出典：CRAタウルビー 調査、2023 0チャート；2024 AIイン

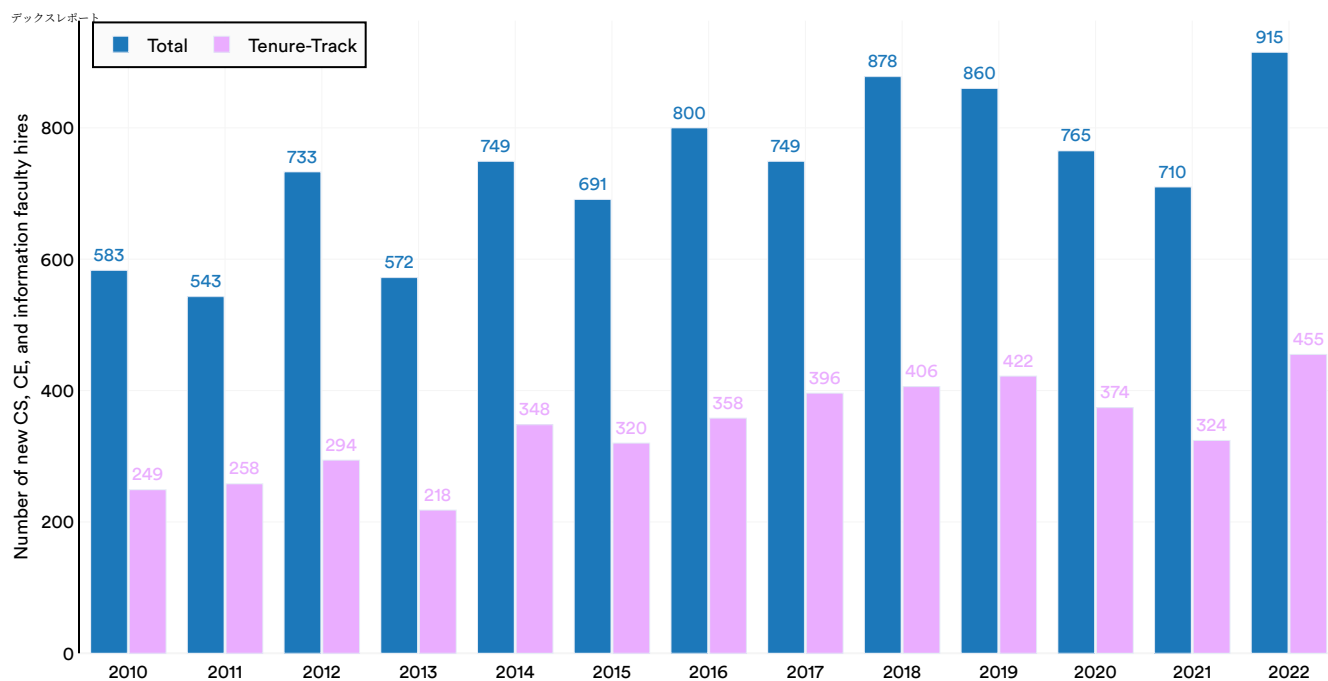


Figure 6.1.11

2022年には、新しい教員の採用の43%が他の学術職からのものであり、学術労働力内での「入れ替わり」を示しています（図6.1.12）。これらの「新しい」教員が他の場所で職を空けたため、彼らの以前の役割は最終的に埋められる必要があります。さらに、2022年に業界から移行した教員の割合は、前年の11%から7%に、2019年の13%からも減少しました。

アメリカおよびカナダのCS、CE、情報学部における新しい教員の出所、2018-22 出典：CRA Taulbee 調査、2023 図：2024 年 AI インデックスレポート

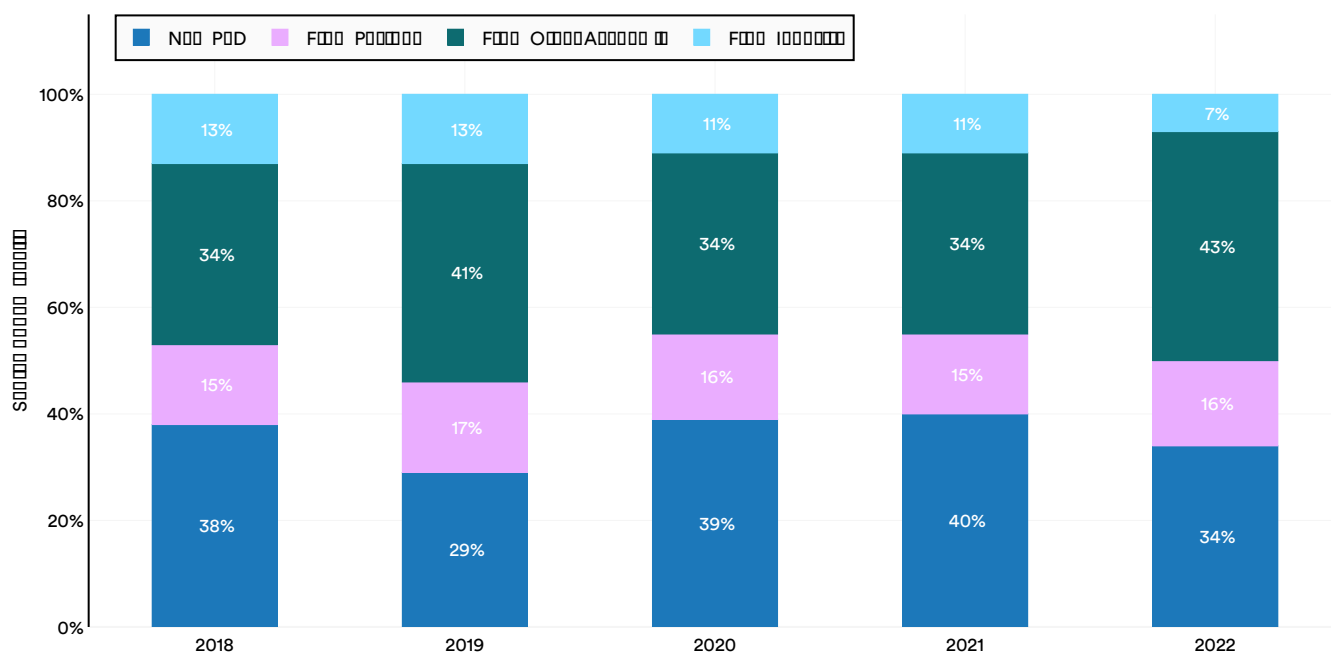


Figure 6.1.12

教員ポジションが埋まらない理由は、過去10年間で様々でした。2011年には、失敗した採用の37%が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%であり、55%はオファーが

辞退されました。この傾向は、新しいCS教員のための競争がますます激化している市場を反映しているようです。しかし、これは他の学術ポジションとの競争が激化しているのか、業界ポジションとの競争が激化しているのかは不明です。

新しいCS、CE、および情報学部 教員ポジションが未填充の理由(総数の%)、2011-22 出典: CRAタウルビー調査、2023年チャート: 2024 AIインデックスレポート

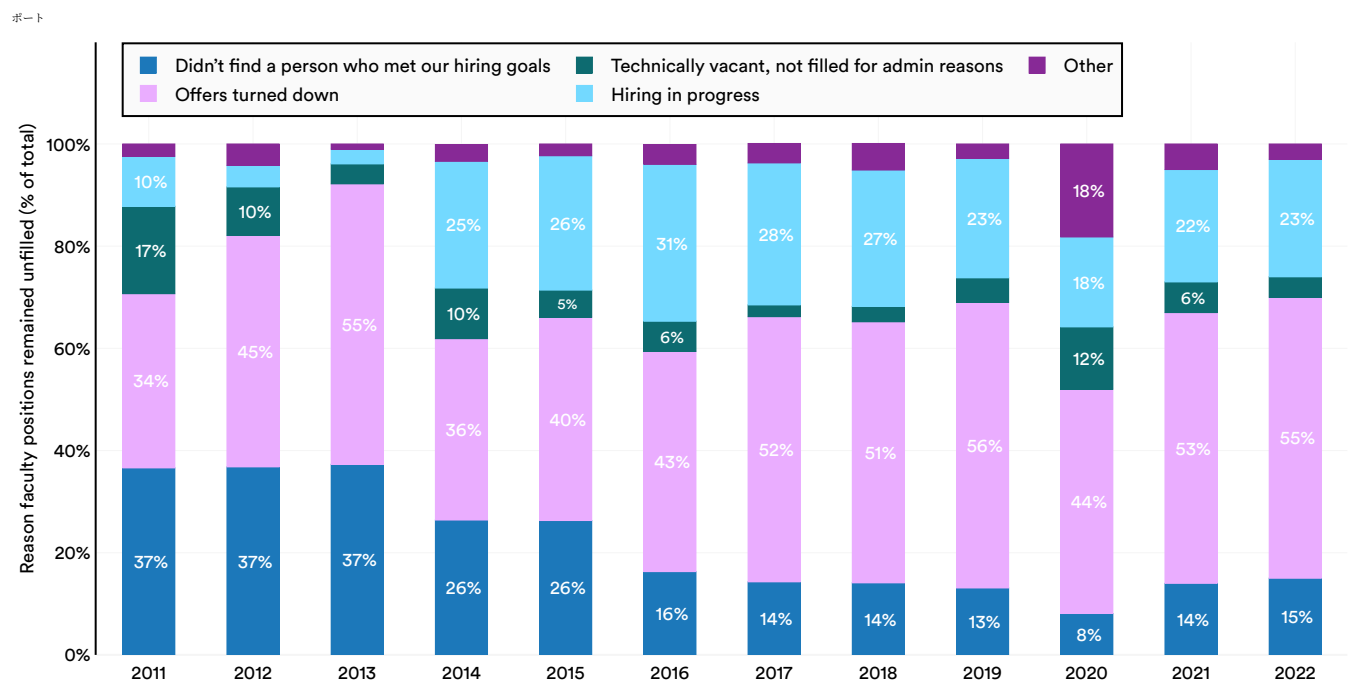


Figure 6.1.13

2022年、北米のCS、CE、および情報分野の学部は、2021年の303人と比較して、教員の離職が405人に達し、著しい増加を経験しました（図6.1.14）。これらの損失のうち、38.5%が他の学術職に移動し、16.3%が非学術的な役割に移ったことは、前年と一貫した傾向を維持しています。

アメリカおよびカナダのCS、CE、情報学部における教員の損失、2011-22 出典：CRAタウルビー調査、2023年チャート：2024 AIインデックスレポート

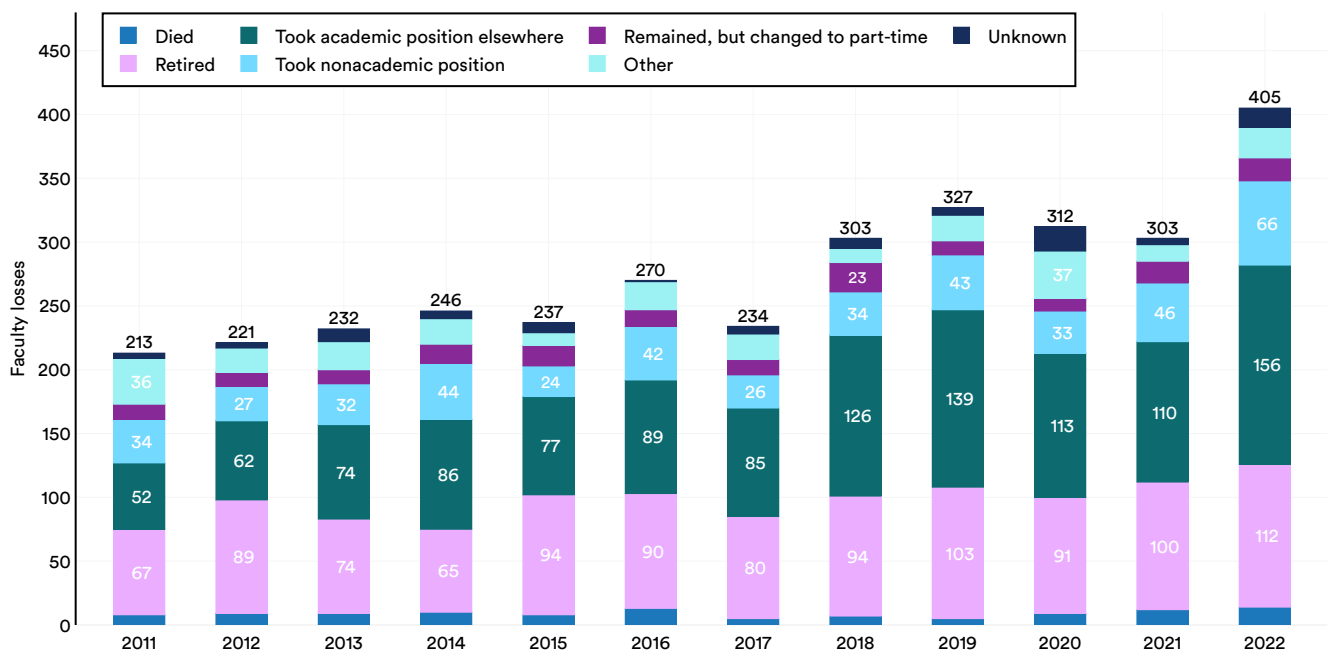


Figure 6.1.14

2015年以降、フルプロフェッサーの中央値の9か月間の給与の増加は、米国のインフレ率をわずかに下回っていますが、アシスタントおよびアソシエイトプロフェッサーの中央値の給与はインフレを上回るわずかな増加を見えます。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%の米国インフレ率には追いついておらず、2015年よりも16.4%高いですが、その期間の19%のインフレ増加にはまだ及んでいません（図6.1.15）。

アメリカのCS教員の9か月間の中央値給与、2015–22 出典：CRAタウルビー調査、2023年チャート：2024 AI インデックスレポート



Figure 6.1.15

2022年には、CS、CE、および情報分野の新しいテニュートラック教員の国際的な採用割合が、前年の13.2%から19.3%に大幅に増加しました（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。

アメリカとカナダにおける新しい国際的なCS、CE、および情報 テニュートラック 教員の採用割合（総数の%）、2010-22

Source: CRA Tenure-Track Survey, 2023; CS and AI Tenure-Track Survey, 2024

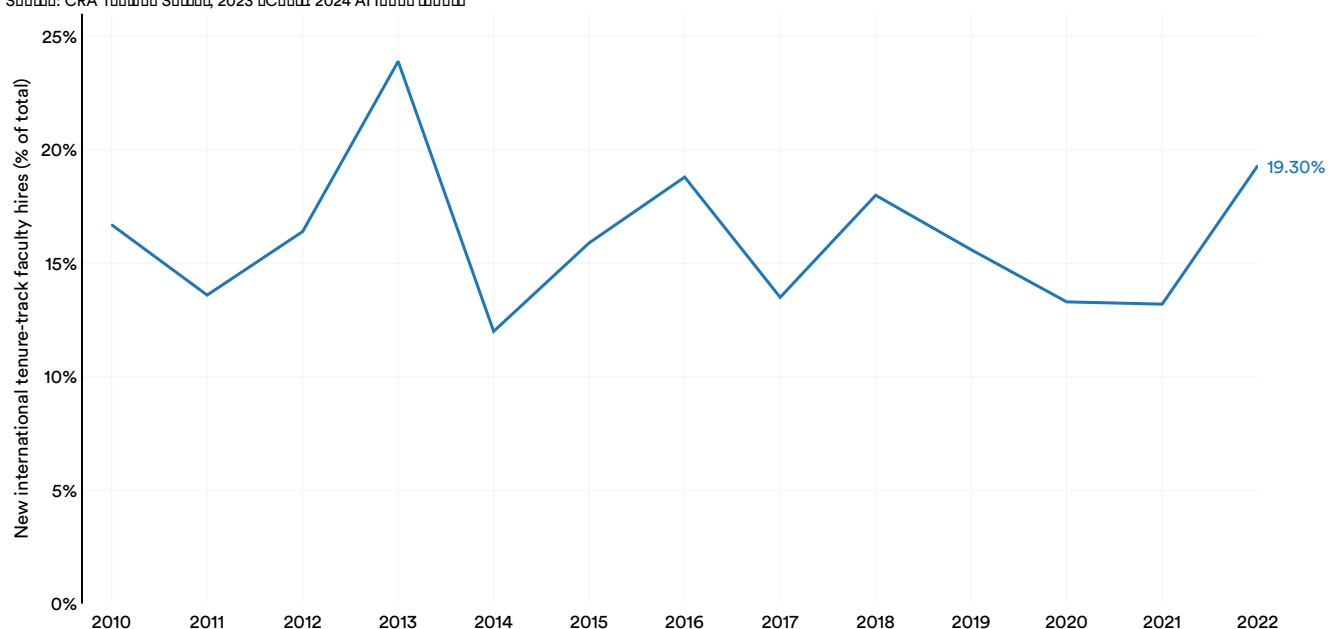


Figure 6.1.16

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、情報学教育の現状を監視することを含む目標を持つ学術および研究コミュニティであるInformatics Europeから得られています。³ Informatics Europe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、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

情報学、コンピュータサイエンス（CS）、コンピュータ工学（CE）、および情報技術（IT）の学士卒業生

2022年、イギリスは情報学、CS、CE、およびITの学士レベルで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で卒業生が増加しましたが、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年のヨーロッパ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の学士卒業生数^{出典:}

Informatics Europe, 2023 01チャート: 2024 AIインデックスレポート

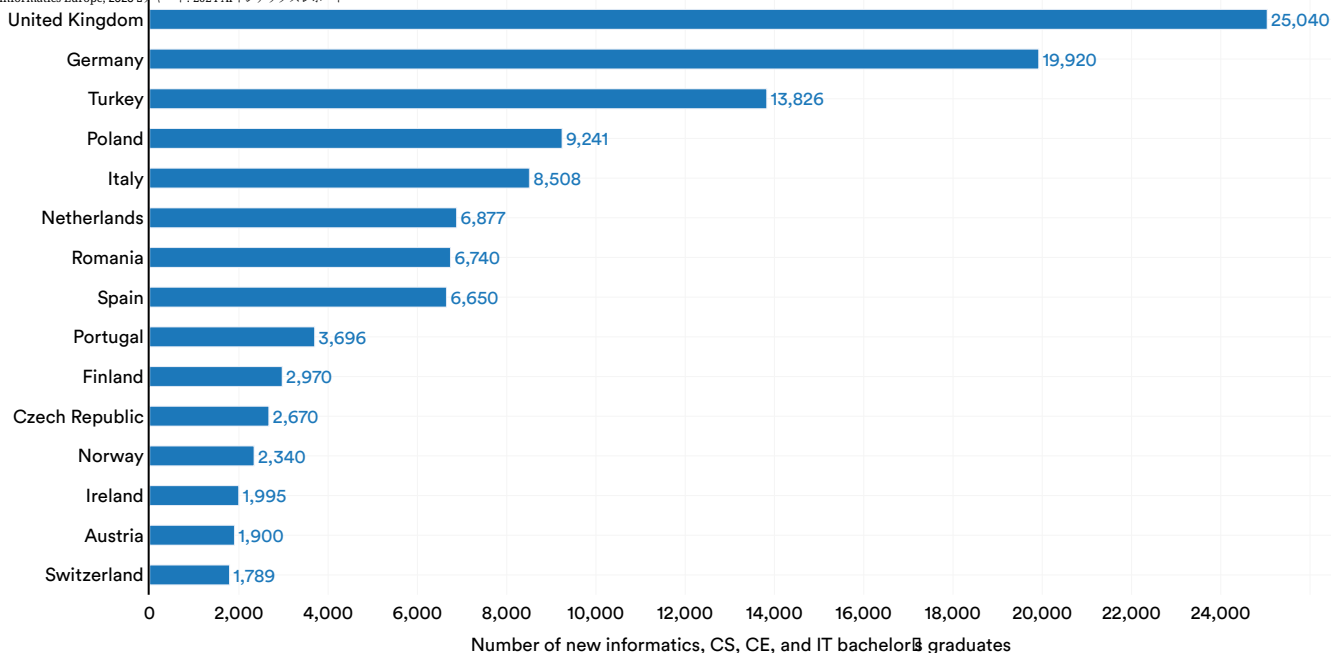


Figure 6.1.17

³ CS教育に関して、ヨーロッパの国々で一律に使用される単一の用語は存在しません。ヨーロッパ全体で、CS教育は情報学、コンピュータサイエンス（CS）、コンピュータ工学（CE）、コンピューティング、情報技術（IT）、情報通信技術（ICT）、および情報科学と技術（IST）などの用語で表現されることがあります。情報学ヨーロッパが情報学の研究プログラムを特定するために使用する科目名の完全なリスト（および英語の翻訳）は、以下のリンクで確認できます。

⁴ 読者は、CRA北米データと後のセクションで詳述されるヨーロッパのCS卒業生データとの間で一人当たりの比較を行うことに注意してください。ヨーロッパのデータは国の統計局から収集され、より広範なカバレッジを誇ります。

⁵ AIインデックスがデータを持つすべての国がこのセクションの国に視覚化されているわけではないことに注意してください。完全なデータにアクセスするには、この章に関連する公開データを参照してください。さらに、年のラベルは学年度が終了する年を指します。たとえば、2022年の新卒者を視覚化する図は、2021/2022学年度に報告された卒業生の数を反映しています。視覚的な単純さのために、インデックスは学生が卒業した年に焦点を当てて選択しています。

国別の新しい情報学、CS、CE、およびITの学士卒業生の割合の変化、2012 年対 2022 年

出典：インフォマティクス・ヨーロッパ, 2023 Dチャート: 2024 AIインデックスレポート

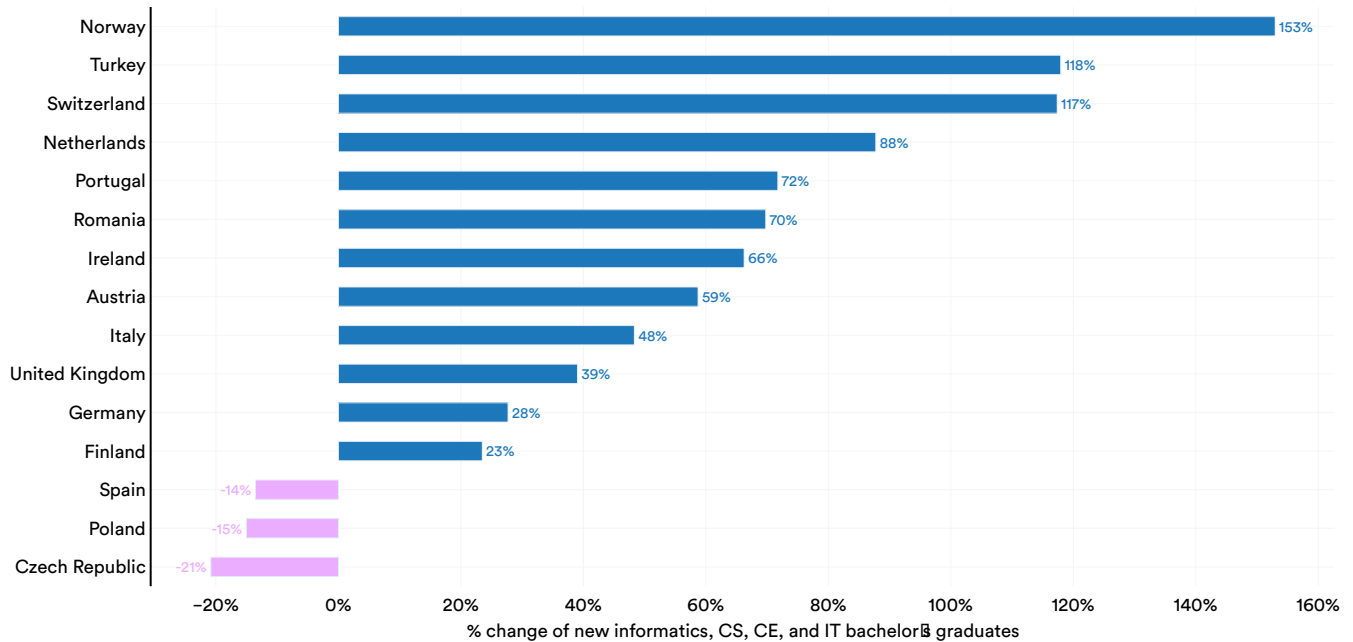


Figure 6.118

フィンランド（53.4）、ノルウェー（42.6）、オランダ（38.6）は、10万人あたりの情報学、CS、CE、ITの新しい学士卒業生の数でリードしています（図 6.1.19）。サンプルに含まれるほとんどのヨーロッパ諸国は、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）。