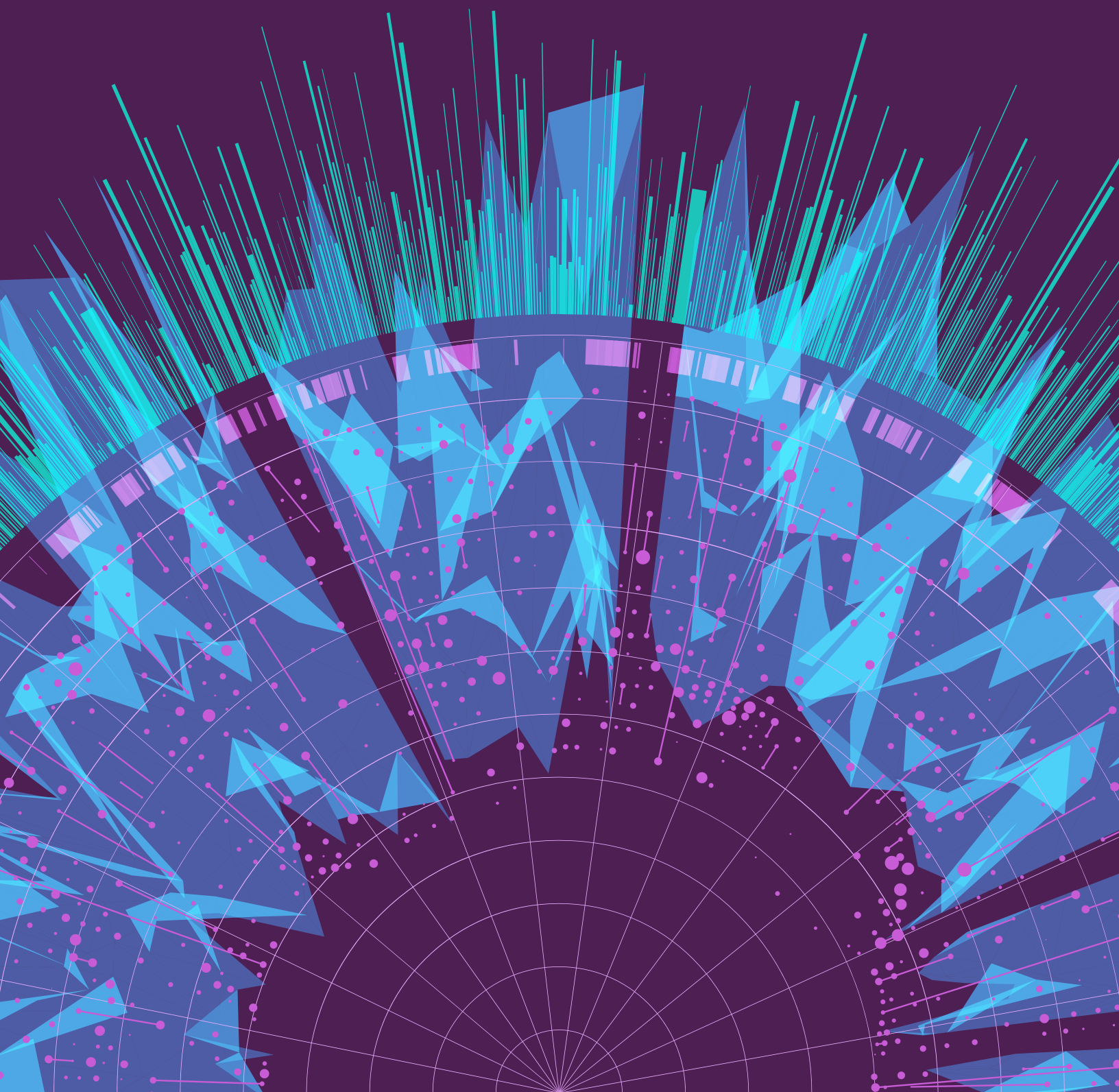




Artificial Intelligence
Index Report 2024

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

公開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の役割の拡大を認識し、AIの科学と医療における新章を導入します。2023年の注目すべきAI支援による科学的成果には、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材料発見アルゴリズムが含まれます。この章では、医療AIシステムの性能、SynthSRやImmunoSEIRAなどの重要な2023年のAI駆動の医療革新、そしてFDAのAI関連医療機器の承認動向も検討します。

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するAlphaDevから、材料発見のプロセスを促進するGNoMEまで、より重要な科学関連のAIアプリケーションの登場が見られました。

2. AIは医療の大きな進歩を促進しています。2023年には、パンデミック予測を強化するEVEscapeや、AIを活用した変異分類を支援するAlphaMissense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はますます医療の進歩を推進するために利用されています。

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MedQAベンチマークにおいて顕著な改善を示しており、これはAIの臨床知識を評価する重要なテストです。2023年の注目モデルであるGPT-4 Medpromptは、正確率90.2%に達し、2022年の最高スコアから22.6ポイントの向上を示しました。2019年にこの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台のAI関連医療機器を承認し、2021年から12.1%増加しました。2012年以降、FDA承認のAI関連医療機器の数は45倍以上に増加しています。AIは実世界の医療目的で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定した2023年の重要なAI関連の科学的ブレイクスルーを紹介します。

5.1 顕著な科学的マイルストーン

AlphaDev

AlphaDevはより高速なソートアルゴリズムを発見しました

AlphaDevは、新しいAI強化学習システムであり、計算アルゴリズムの向上分野で科学者やエンジニアによる数十年の研究を改善しています。AlphaDevは、既存の人間の基準よりも少ない命令でアルゴリズムを開発しました。

短いシーケンスにおけ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、このライブラリのこの部分における10年以上ぶりの最初の更新であり、強化学習を用い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：アルゴリズム長さ最適化時

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

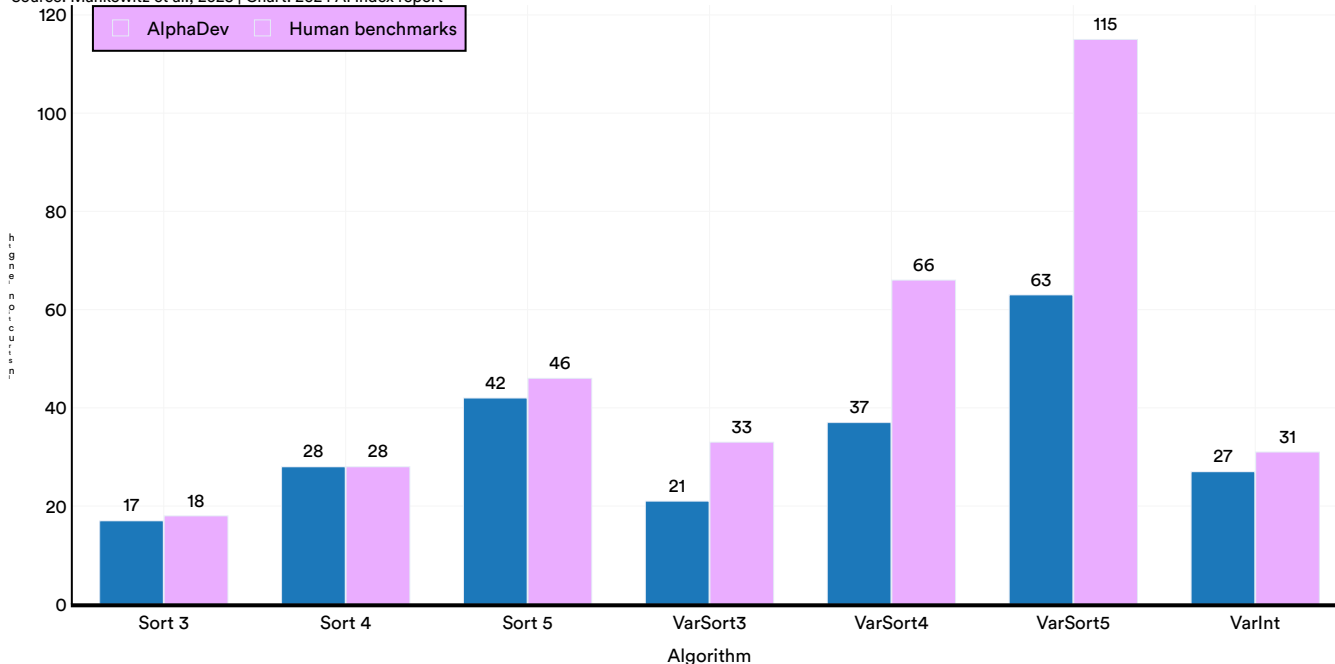


Figure 5.1.1

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に頂点、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、医用画像、科学的可視化にとって重要です。従来のアイソサーフェス抽出アルゴリズムは、解像度の制限、構造の剛性、数値的不安定性にしばしば苦勞し、その結果として影響を与えます。

品質。FlexiCubesは、勾配に基づく最適化と適応可能なパラメータにAIを利用することで、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。微分可能なアイソサーフェシングを用いたメッシュ再構築の他の主要な方法と比較して、FlexiCubesは基礎となる地面の真実により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FlexiCubes表面再構築 Source: Nvidia, 2023

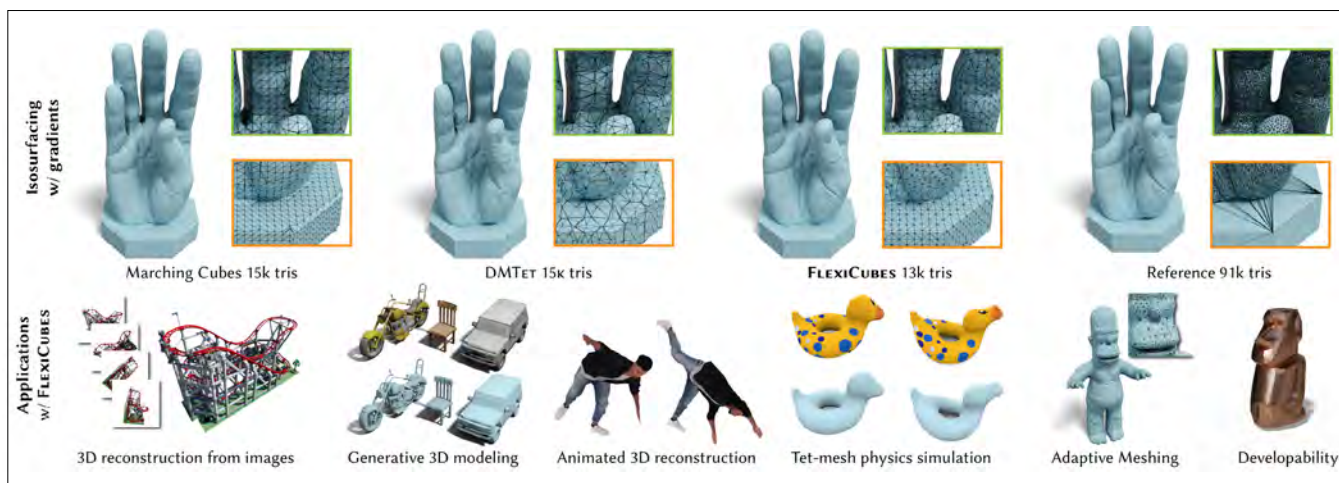


図5.1.2

3Dメッシュ再構築に関する定量的な結果を選択してください

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

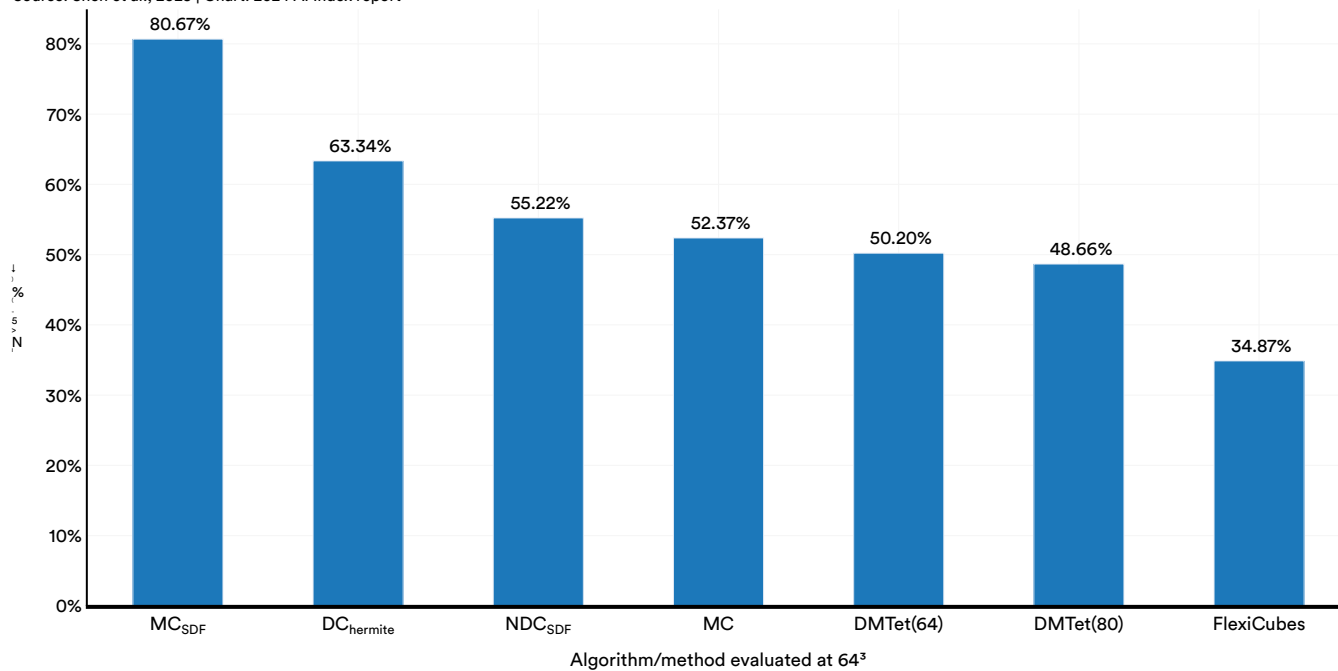


図5.1.3

シンボット

有機分子合成のためのAI駆動ロボット化学者

Synbotは、多層構造のシステムを採用しており、化学合成計画用のAIソフトウェア層、命令を翻訳するロボットソフトウェア層、および実験を行う物理的なロボット層で構成されています。AIとロボットシステム間の閉ループフィードバック機構により、Synbotは既存の基準と同等またはそれを超える収率の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]の合成を目的とした実験では、Synbotは変換収率が上回る複数の合成式を開発しました。

Synbot設計

Source: Ha et al., 2023

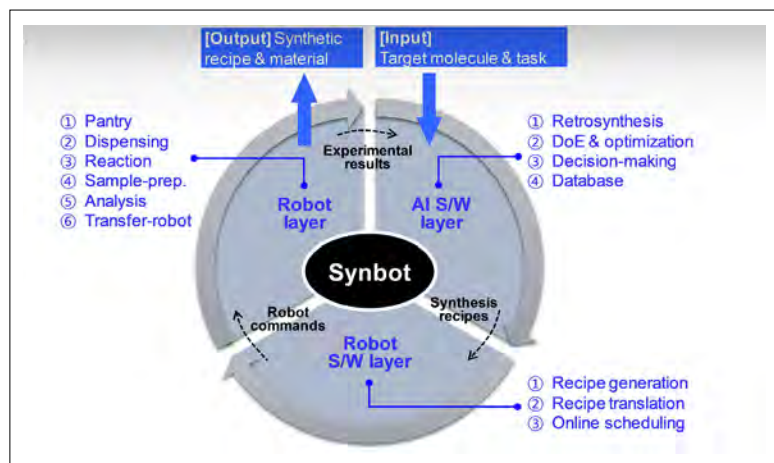


図5.1.4

中間の80%の基準範囲で合成を完了し、はるかに短時間で完了しました（図5.1.5）。Synbotの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の可能性を示しています。

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

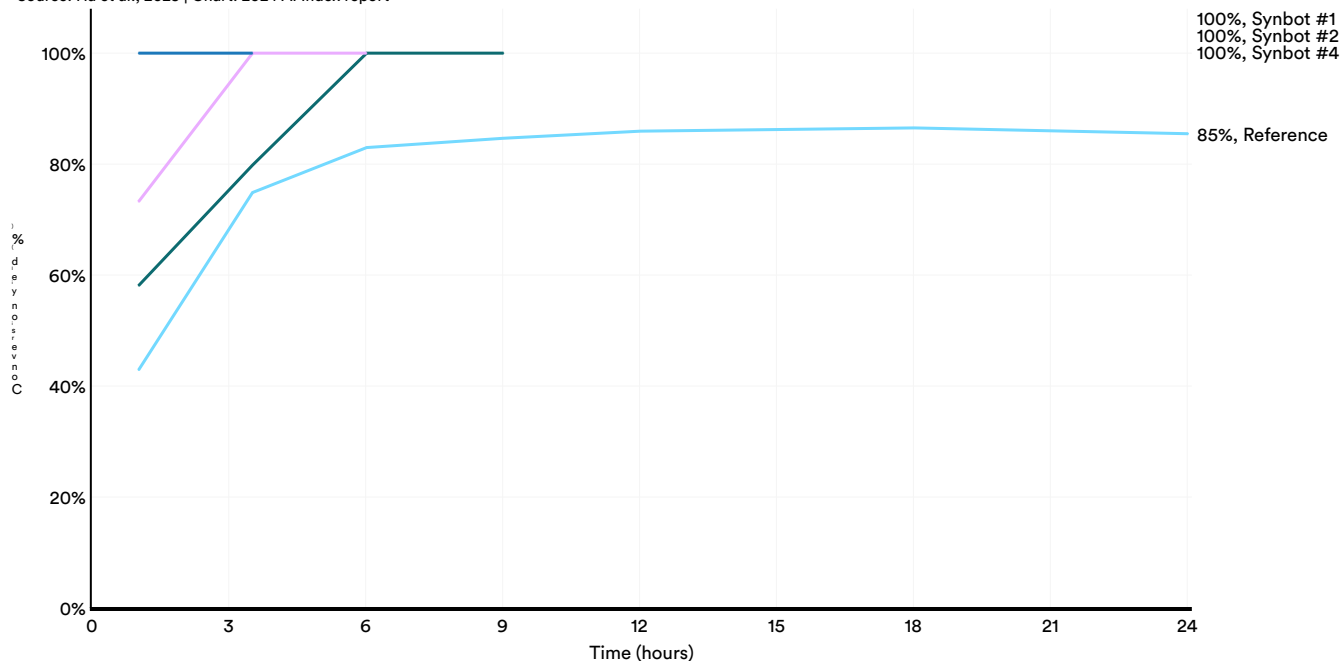


図5.1.5

GraphCast

より正確な世界的な天気予報

GraphCast とともに

GraphCastは、新しい気象予報システムであり、わずか1分以内に非常に正確な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、気温、風速、大気の状態を予測します。

など。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報（HRES）と比較しています。GraphCastは平均二乗根誤差が低く、予報が観測された気象パターンにより近いことを意味します。GraphCastは、気象パターンの解読に役立ち、極端な気象イベントへの備えを強化し、地球規模の気候研究に貢献する有用なツールとなり得ます。

GraphCast 天気予報

Source: DeepMind, 2023

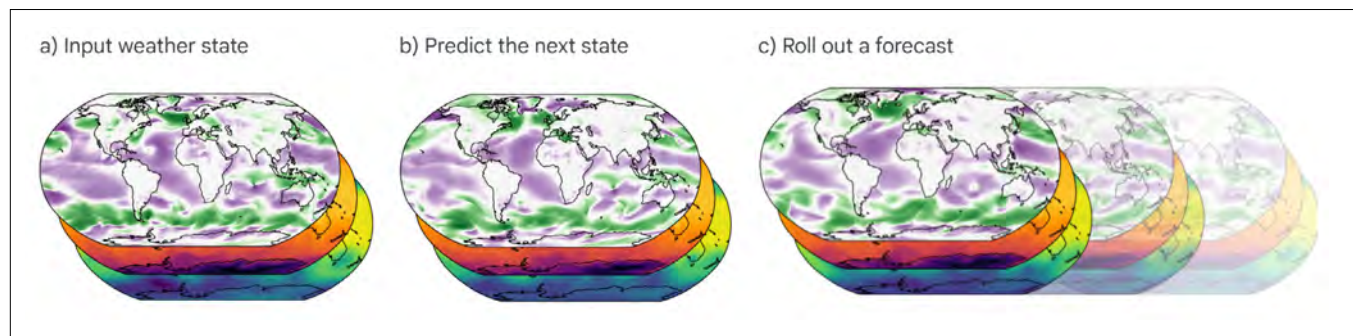


Figure 5.1.6

10日間のz500予報精度：GraphCast vs. HRES

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

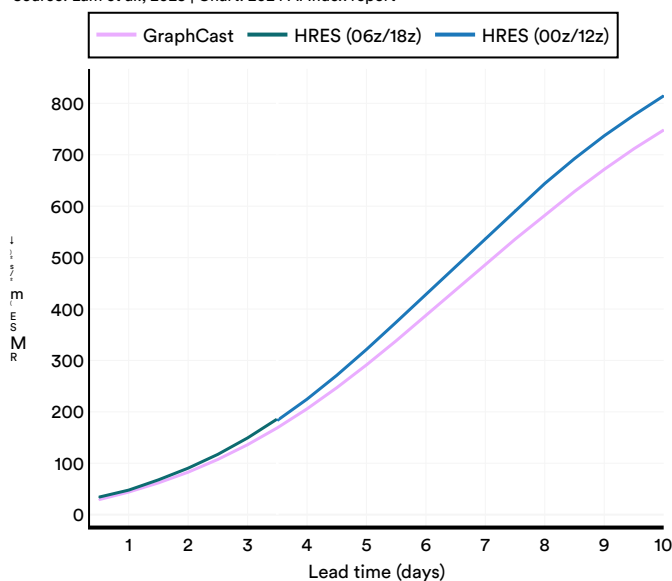


Figure 5.1.7

GNoME

GNoMEを使った新素材の発見

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歩にとって重要です。しかし、この発見プロセスは一般的に高価で遅いです。Googleの研究者による最近の進展は、グラフネットワークと呼ばれるAIモデルの一種が、大規模なデータセットで訓練されるとこのプロセスを促進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、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、データとスケーリングの力が科学的ブレークスルーを加速させることを示しています。

サンプル材料構造

Source: Merchant et al., 2023

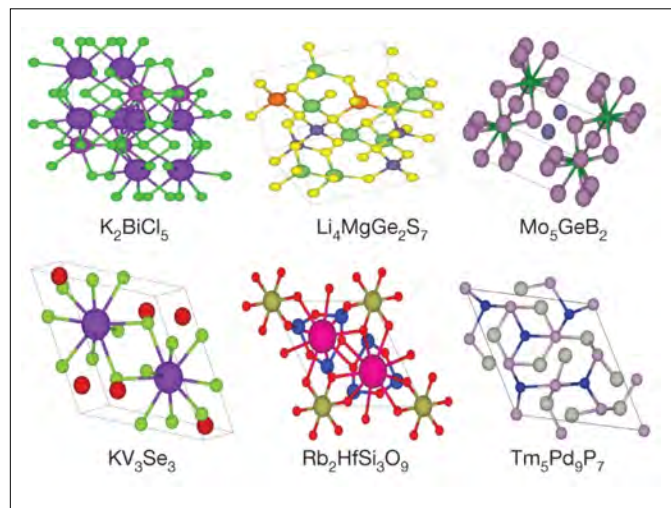


図5.1.8

GNoME 対 Materials Project: 安定結晶数

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

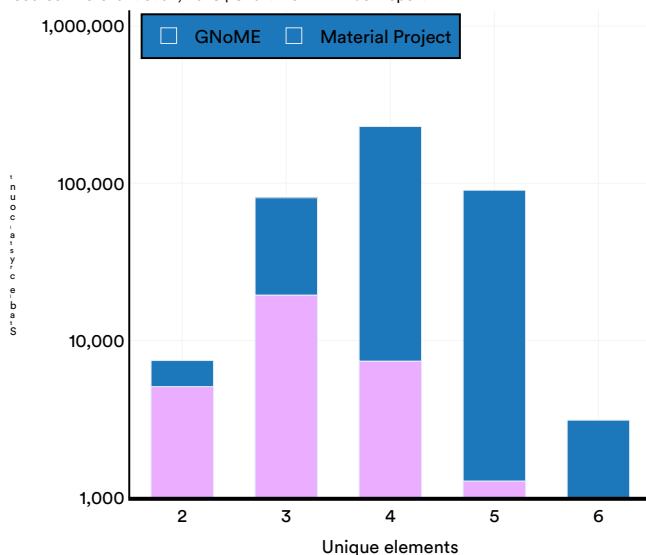


図5.1.9

GNoME 対 Materials Project: 異なるプロトタイプ

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

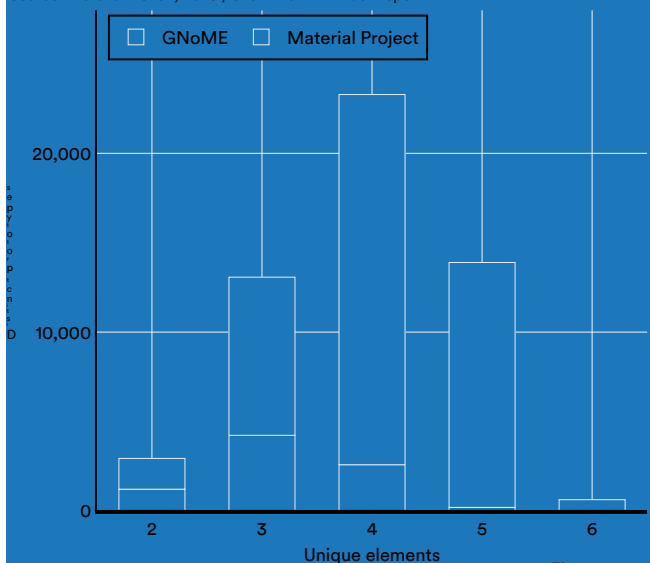


Figure 5.1.10

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顕著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、非常に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より早期に予測できるより正確な予測方法を開発することは、実質的なプラスの効果をもたらす可能性があります。

Googleの研究者チームは、AIを用いて高精度の水文学シミュレーションモデルを開発しました。これらのモデルは、計測されていない流域にも適用可能です。¹ これらの革新的な方法は、特定の極端な洪水イベントを最大5日前に予測でき、その精度はGloFASなどの最新のモデルと同等かそれを上回ります。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（陽性予測の正確さ）と再現率（すべての関連事例を正しく識別する能力）を示し、最先端の方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、すでに80か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のリターン期間にわたる予測

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

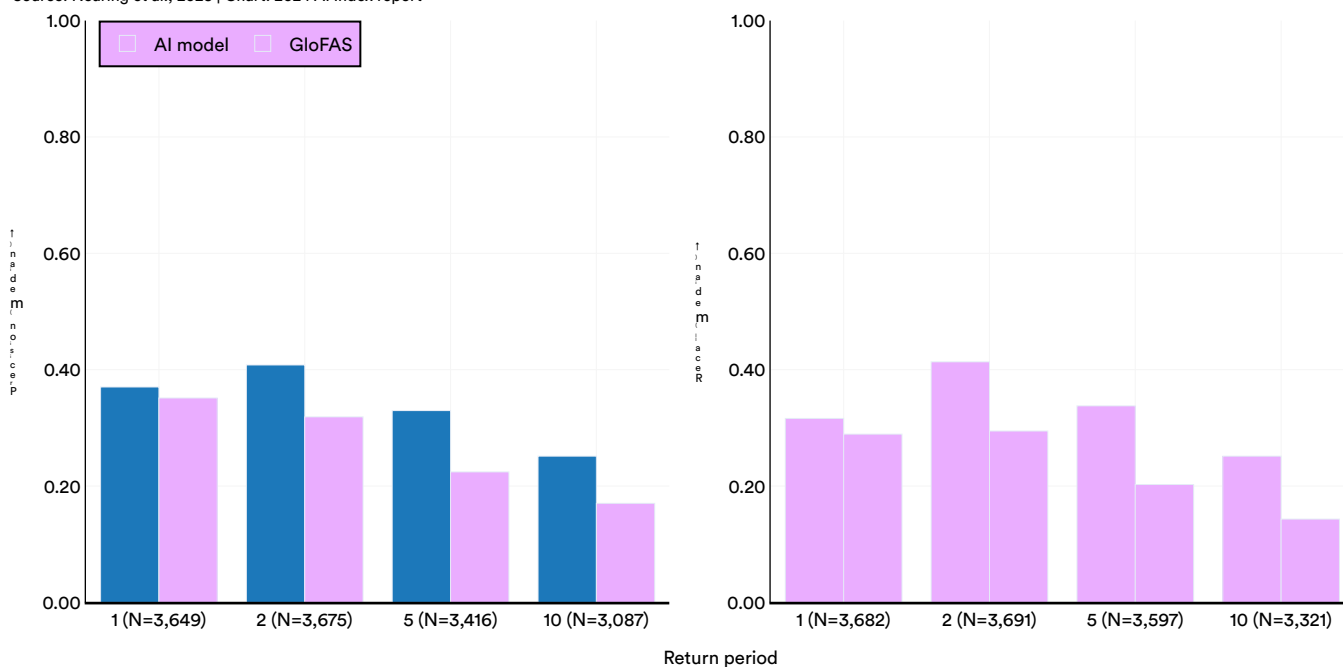


図5.1.11

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価値が高まっており、ポリープの検出や臨床医の診断支援に応用されています。AIの性能が向上し続ける中、その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。本セクションでは、2023年に導入された重要なAI関連の医療システム、現在の臨床AIの知識の状況、そして病院管理の向上を目的とした新しいAI診断ツールやモデルの開発について紹介します。

5.2 医学における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関連医療のブレイクスルーを紹介します。

SynthSR

高度な解析のための脳スキャンの変換 SynthSRは、臨床脳スキャンを高解像度のT-1強調画像に変換する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可能な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を用いた実験は、スキャンレベルと被験者レベルの両方で観測された体積との堅牢な相関を示しており、SynthSRが高解像度のT1スキャンによって生成された画像に非常に近い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いてSynthSRスキャンが実測値とどの程度一致しているかを示しています。SynthIDは、脳構造の可視化と解析を大幅に改善し、神経科学の研究や臨床診断を促進します。

SynthSR世代

Source: Iglesias et al., 2023

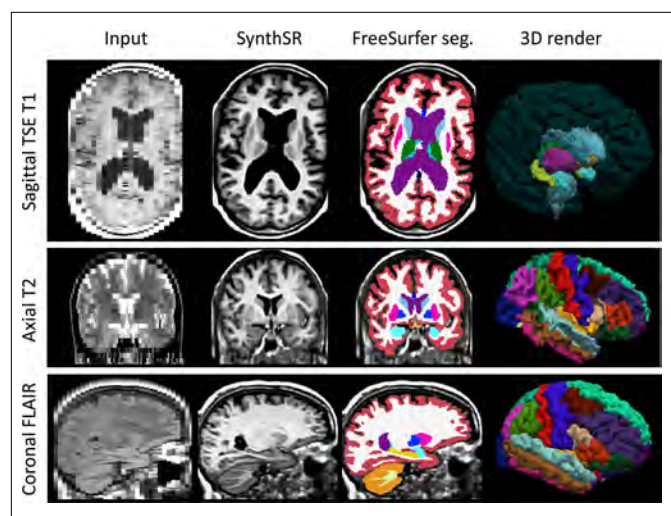


Figure 5.2.1

SynthSRと選択された脳領域の真の体積との相関

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

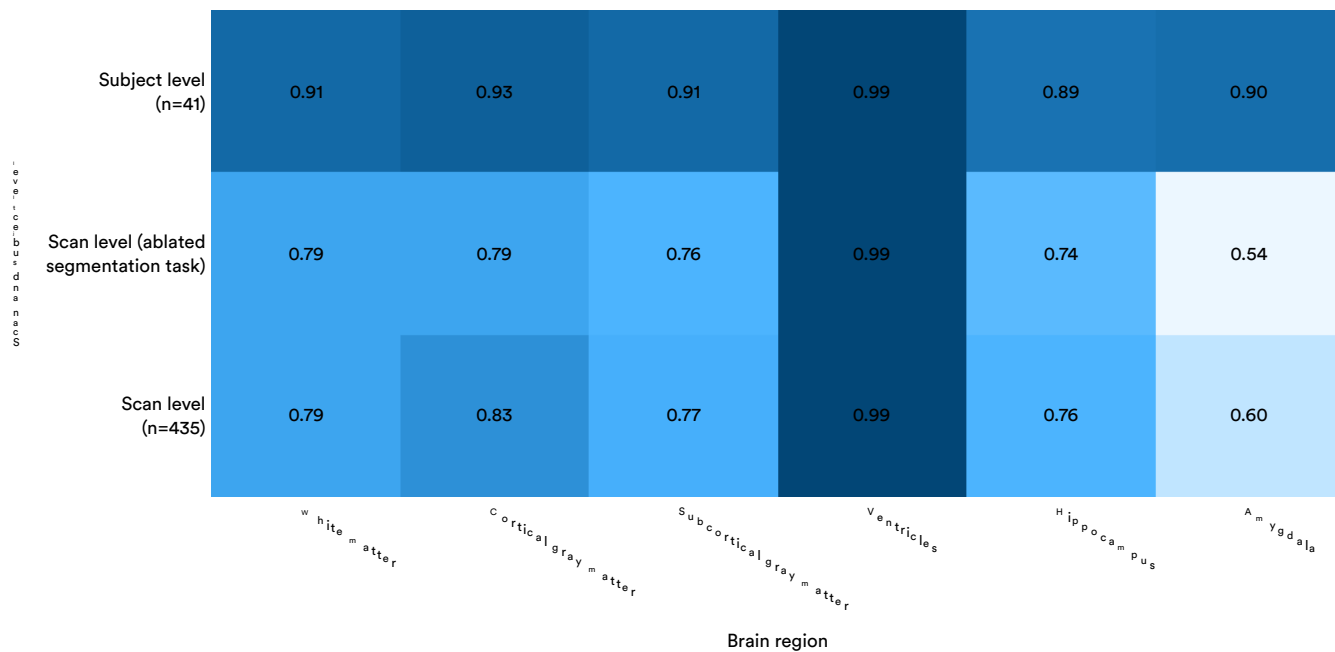


Figure 5.2.2

結合型プラズモニク赤外線センサー

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。従来の方法、例えば質量分析やELISA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点で有用ですが、構造的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）分光法を用いたAI結合プラズモニク赤外線センサーと免疫測定法（Immu noSEIRA; 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システムによる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一致していることがわかりました（図5.2.4）。

免疫SEIRA検出原理と装置 Source: Kavungal et al., 2023

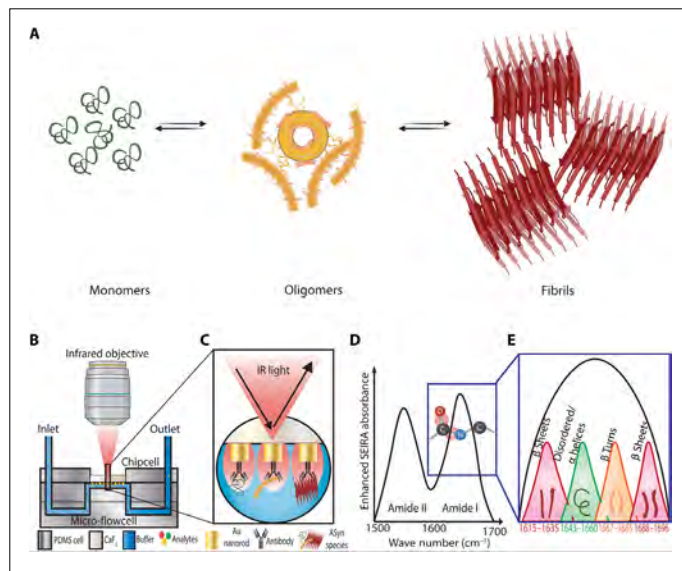


図5.2.3

深層ニューラルネットワークによる予測と実際のブリルの割合（テストサンプルにおける百分率） Source: Kavungal et al., 2023 | Chart: 2024 AI Index report

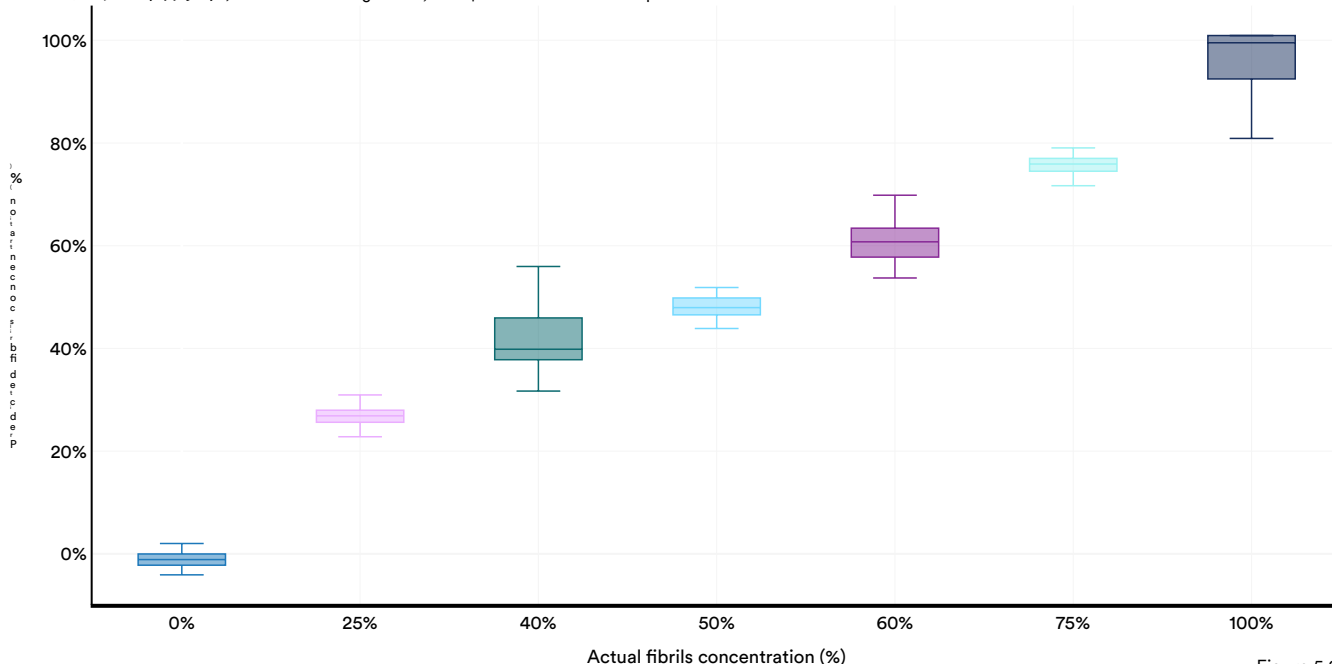


Figure 5.2.4

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。従来の方は、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、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および構造的情報を用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。

ウイルスの逃避 (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、観測されたSARS-CoV-2変異の50.0%を予測し、従来の実験室研究が予測した46.2%および32.3%を上回り、以前のモデルが予測した変異のみに比べて24%に過ぎなかったことを超えています (図5.2.6)。この性能は、EVEscapeが今後のパンデミック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示しています。

EVEscapeデザイン

Source: Thadani et al., 2023

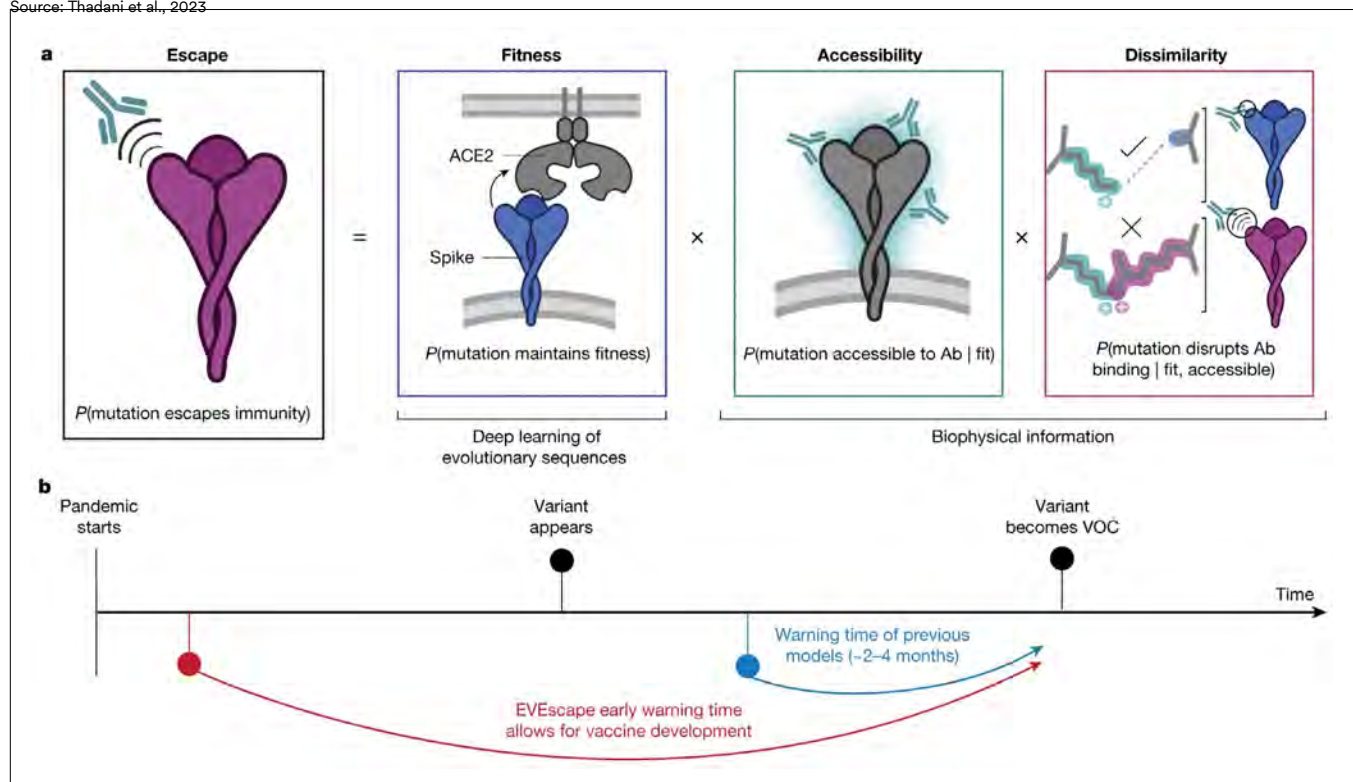


図5.2.5

EVEscape 対他のモデルによる SARS-CoV-2 RBD 変異予測

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

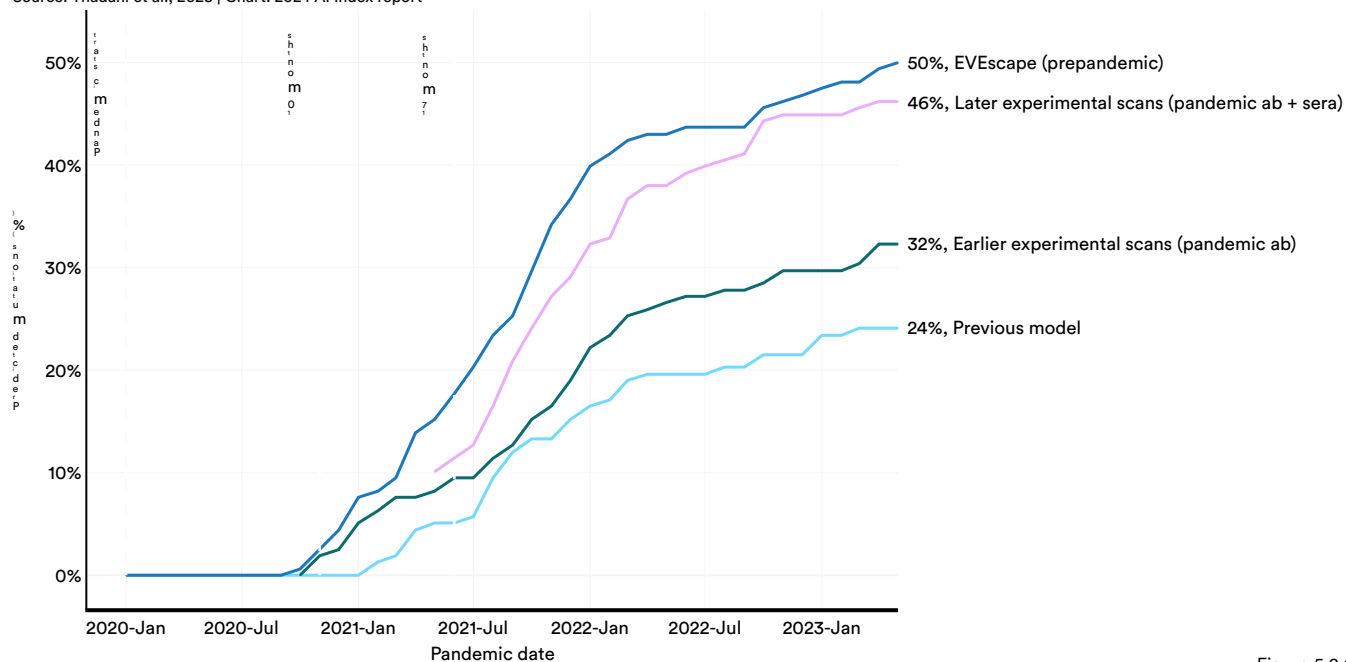


Figure 5.2.6

アルファミセンス

AI変異のより良い分類

科学者たちはまだ、どの遺伝子変異が疾患を引き起こすのかを完全には理解していません。数百万もの可能な遺伝子変異の中で、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、AlphaMissenseという新しいAIモデルを発表しました。このモデルは、7100万のミスセンス変異の病原性を予測しました。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変化であり（図5.2.7）、さまざまな疾患、特に癌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識別し、残りは不確定として分類しました（図5.2.8）。一方、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%の性質しか確認できていません。

ヘモグロビンサブユニットベータ (HBB)

Source: Google DeepMind, 2023

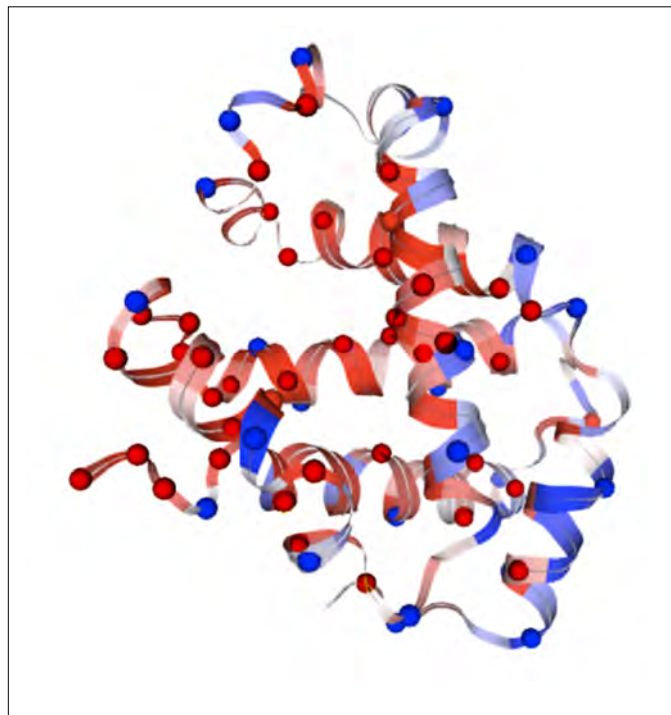


図5.2.7

AlphaMissense予測

出典：Google DeepMind, 2023 | チャート：2024 AI指数レポート

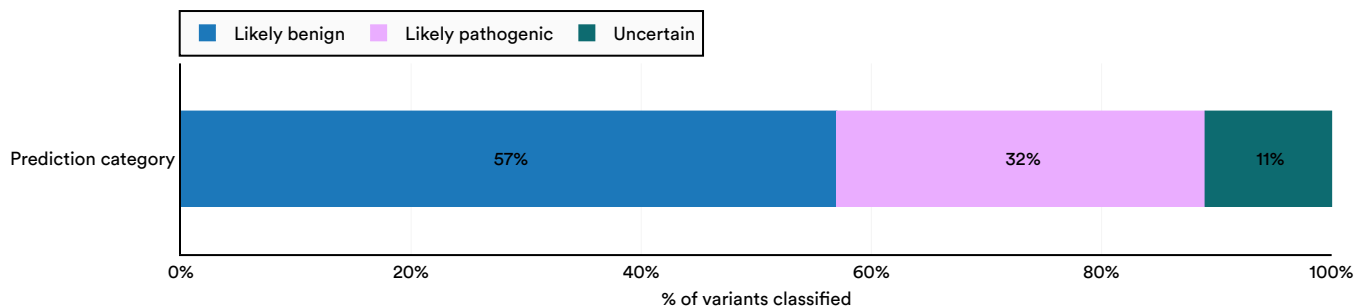


Figure 5.2.8

ヒューマン・パンゲノム・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ための分子指示書の集合体です。最初のヒトゲノムのドラフトは2000年に公開され、2022年に更新されました。しかし、その更新はやや不完全でした。血液型などのさまざまな遺伝子変異を取り入れておらず、多様な祖先集団をより完全にマッピングしていませんでした。したがって、既存のゲノムリファレンスのもとでは、特定の集団において疾患を検出したり治療法を見つけたりすることは難しいでしょう。

2023年、人類パンゲノム研究コンソーシアムは、60の機関から119人の科学者で構成され、AIを用いて更新され、より代表的な人間ゲノムマップ（図5.2.9）を作成しました。研究者たちは、驚くべき精度を達成し、タンパク質コード遺伝子の中央値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%を注釈付けしました。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

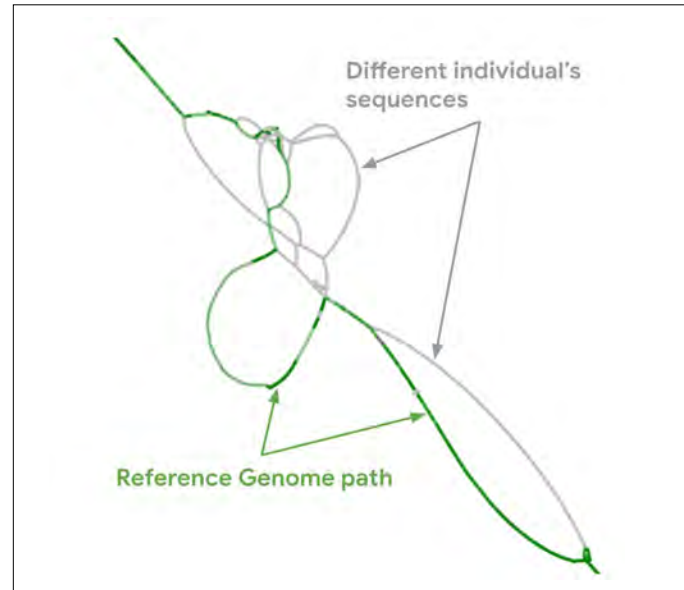


図5.2.9

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の結果

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

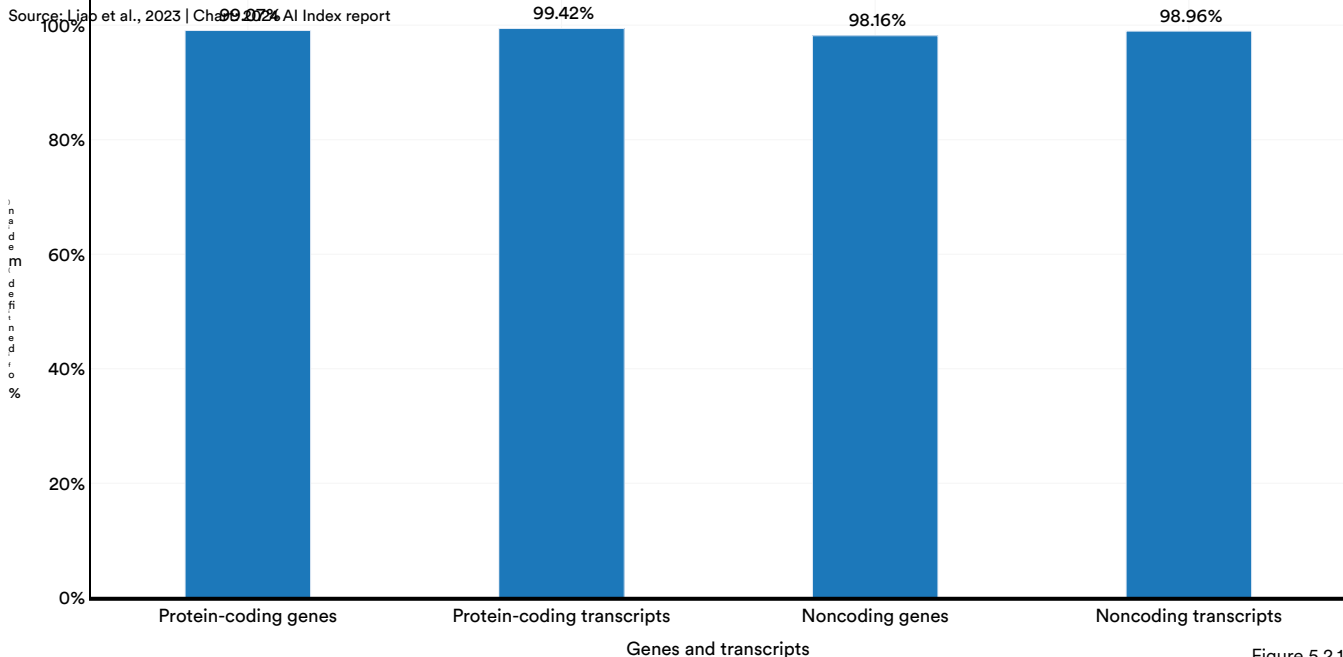


Figure 5.2.10

臨床知識

AIモデルの臨床知識の評価は、特に臨床現場で適用可能な知識の範囲を判断することを含みます。

顕著な改善が見られ、主要なシステムであるGPT-4 Medpromptは正確率90.2%を達成し、2022年のトップスコアから22.6パーセンテージポイントの増加を示した（図5.2.11）。MedQAの開始以来、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、臨床的に知識のあるAIシステムの急速な進歩を示している。

MedQA

2020年に導入されたMedQAは、医師を挑戦させることを目的とした6万以上の臨床質問を含む、専門的な医療委員会試験から派生した包括的なデータセットです。

MedQAベンチマークでのAIの性能が向上しています

MedQA：正確さ

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

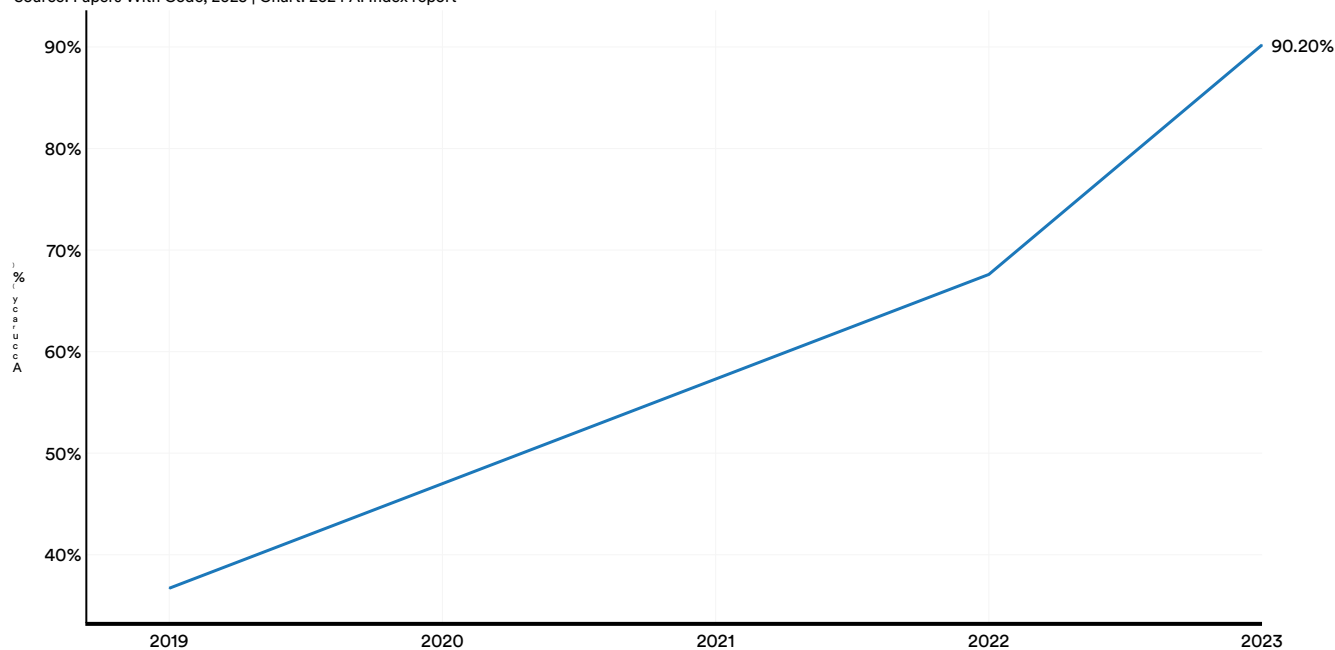


図5.2.11

ハイライトされた研究：

GPT-4 Medprompt

LLMs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で優れるためには、かなりの微調整が必要とされることが一般的に想定されています。微調整とは、ドメイン固有のデータを用いてLLMを訓練することを意味します。

2023年後半のMicrosoft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を誘導し、4つの挑戦的な医療ベンチマークのグループであるMultiMed QAベンチマークスイート（図5.2.12）で顕著な性能を達成させました。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問題セクションで、2022年のトップモデルであるFlan-PaLM 540Bをそれぞれ3.0、21.5、16.2ポイント上回りました。また、その時点で最先端だったMed-PaLM 2（図5.2.13）の性能も超えました。


さらに、前述のとおり、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可能性のある医療能力を強調するだけでなく、モデルを専門的なドメインに適応させるために微調整が必ずしも必要でないことも示しています。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 医療質問への回答 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

 Expert

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.

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

ハイライトされた研究：

GPT-4 Medprompt（続き）

MultiMedQAサブベンチマークにおけるモデルの性能

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

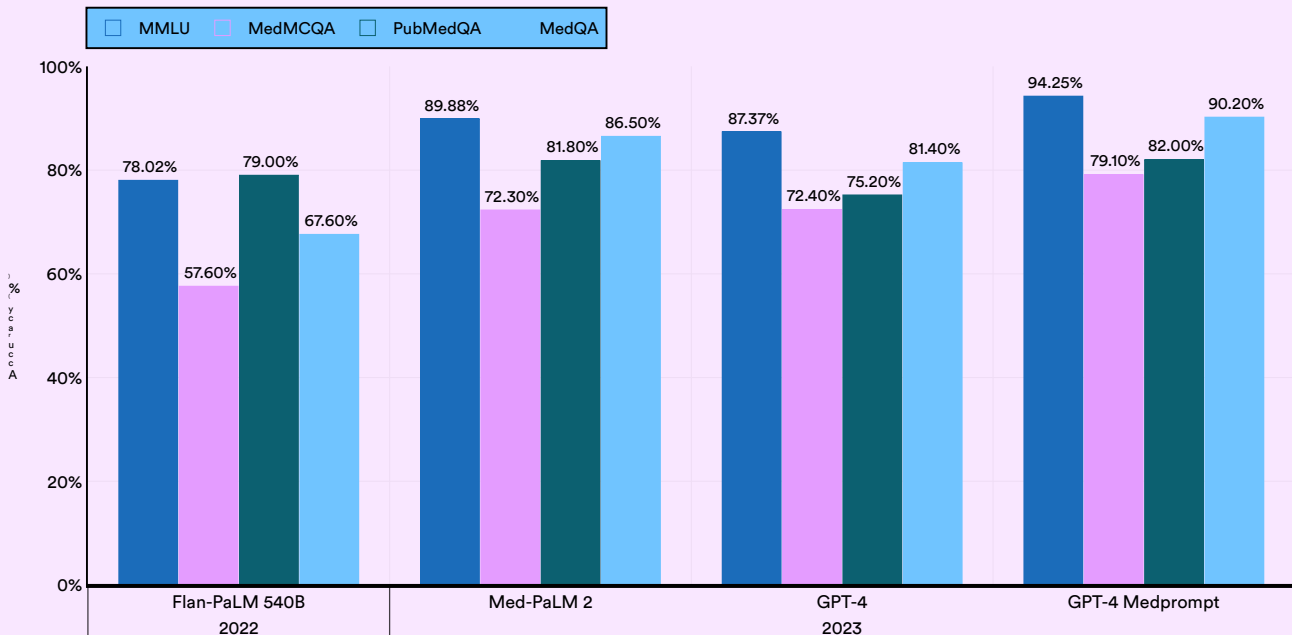


図5.2.13

ハイライトされた研究：

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、その重みは一般公開されておらず、自由に利用できません。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も目指しています。この新しい研究の中で、MediTron-70Bは特に有望とされています。このモデルは、MedQAベンチマークで70.2%の信頼性のある精度を達成しています。これはGPT-4 MedpromptやMedの性能には及びませんが、

PaLM 2（クローズドモデルの両方）は、2023年の最先端の結果と比べて大き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でのスコアは、これまでにオープンソースモデルによって達成された最高値です。医療AIがその最大の潜在能力を発揮するためには、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は前向きな一歩を示しています。

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指数レポート

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、例えば放射線学やがん検出などの診断目的にも使用できます。

ハイライトされた研究：

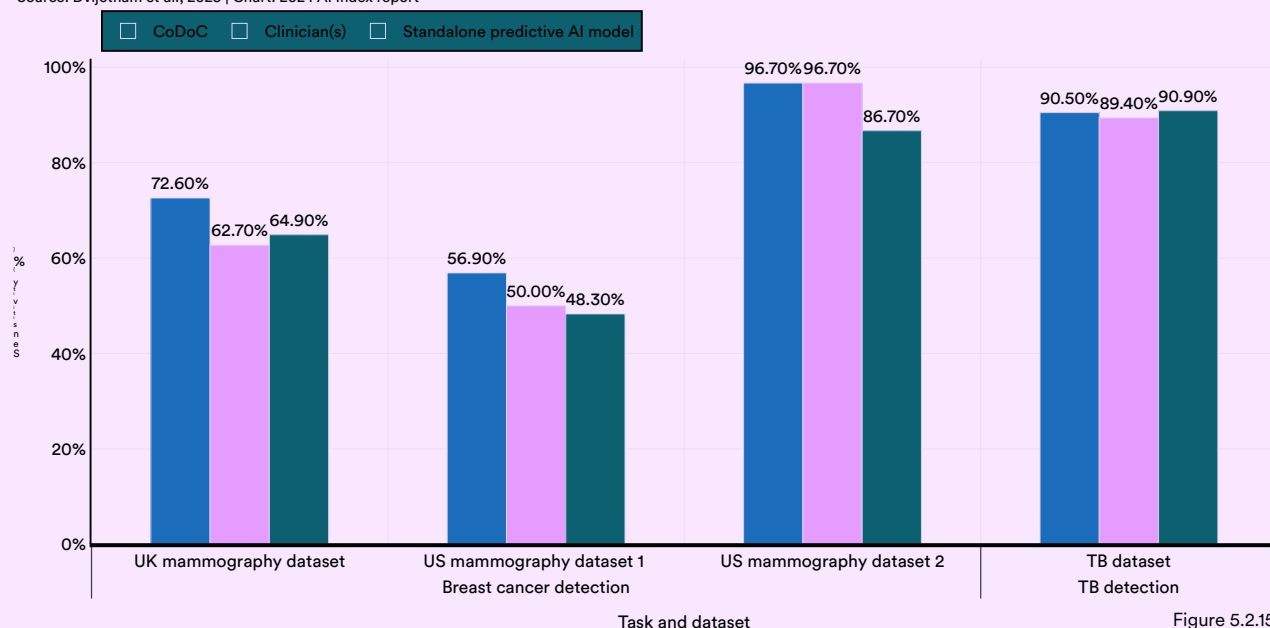
CoDoC

AI医療画像システムは堅牢な診断能力を示しますが、臨床医が見逃す診断を見落とす場合や、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を論理的に統合する必要性を示唆しています。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、いつAIに診断を任せ、いつ従来の臨床方法に委ねるべきかを見極めるシステムです。CoDoCは、特に感度（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において、CoDoCの感度は臨床医より平均4.5パーセントポイント上回り、スタンドアロンのAIモデルより6.5パーセントポイント上回っている（図5.2.15）。特異度の面では、CoDoCはテストされたデータセット全体で臨床医より平均2.7パーセントポイント優れており、スタンドアロンの予測モデルより5.7パーセントポイント優れている。さらに、CoDoCは臨床のワークフローを66%削減することが示されている。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、診断の正確性と効率性を向上させることができることを示唆している。

CoDoC 対 独立予測AIシステムと臨床読影者：感度

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



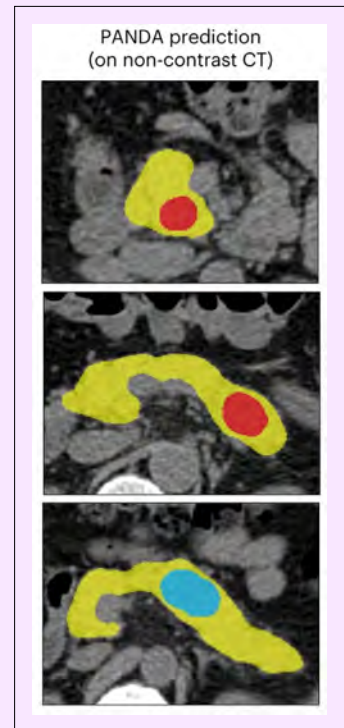
ハイライトされた研究：

CTパンダ

膵管腺癌（PDAC）は特に致命的な癌であり、多くの場合、外科的介入に間に合わないほど遅れて発見される。無症状の個人に対する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、AIモデルのPANDA（膵臓癌検出と人工知能を用いた診断）を開発した。これはX線画像中の膵臓病変を効率的に検出・分類できる（図5.2.16）。検証テストでは、PANDAは感度で平均的な放射線科医を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、診断が難しいとされてきた疾患の診断において重要な進歩を示しており、コスト効果が高く正確な検出を可能にしている。

パンダ検出

Source:
Cao et al., 2023
図5.2.16



出典：Cao et al., 2023 | 図：2024年AI指数レポート PANDA 対平均放射線科医の多施設検証（6,239人の患者）

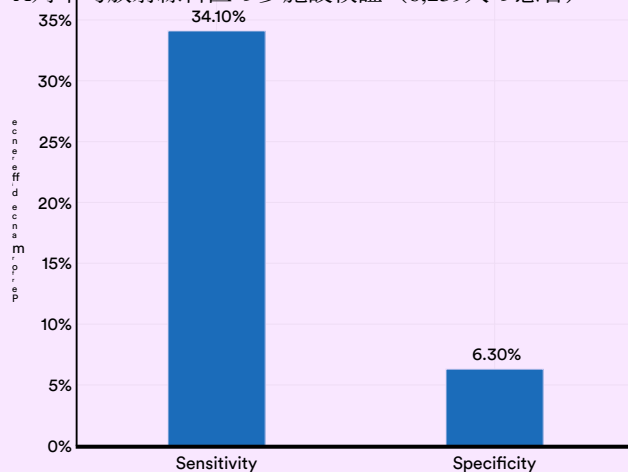


Figure 5.2.17

実世界のマルチシナリオ検証におけるPANDAのパフォーマンス（20,530人の患者）

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

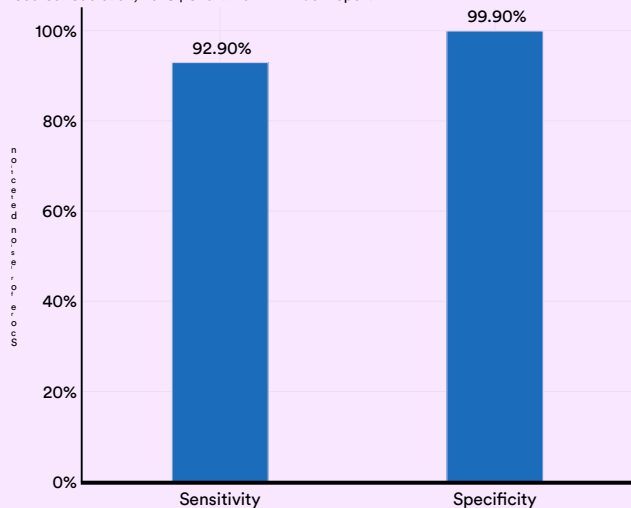


Figure 5.2.18

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のユースケースに関する追加調査

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

米国食品医薬品局（FDA）は、承認を受けたAI/ML対応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の事前市場基準を満たしており、その効果と安全性の詳細な審査を含みます。2023年10月現在、FDAは生成AIを利用した機器やLLMを搭載した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台の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%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によって承認されたAI医療機器の数、2012–22

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

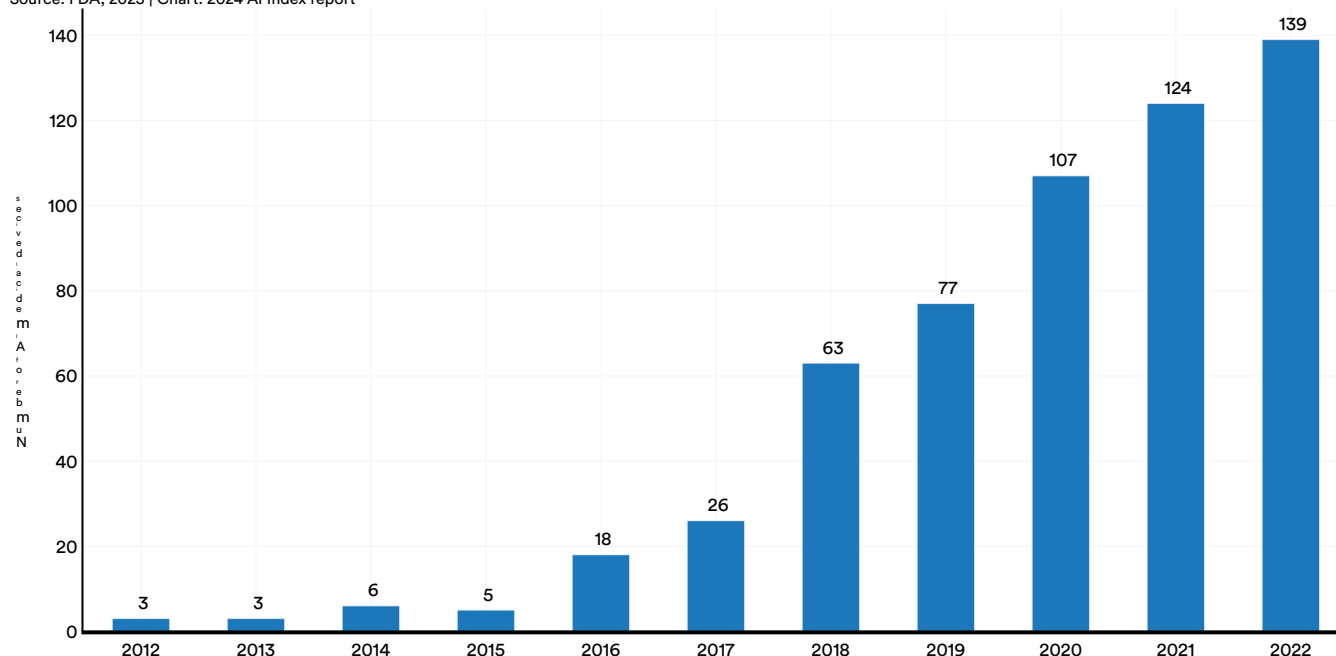


図5.2.20

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、圧倒的多数の87.1%が放射線医学に関連していました。次に多い専門分野は心血管で、承認の7.2%を占めていました。

2012年から2022年までのFDAによって承認されたAI医療機器の専門分野別数

出典：FDA、2023 | 図：2024 AI指数レポート

	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、患者ケアの水準を高める可能性も秘めています。

再。

ハイライトされた研究：

MedAlign

医療分野のAIにおける顕著な進歩にもかかわらず、MedQAやUSMLEのような知識ベースの質問に焦点を当てた既存のベンチマークは、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的なタスクに従事し、勤務時間のかなりの割合を事務作業に費やすことがよくあります。AIはこれらのプロセスを効率化する可能性を持っていますが、医療管理に関するLLMのベンチマークや微調整に適した電子カルテ（EHR）データセットが不足しています。今年、研究者たちはMedAlignという包括的なEHRベースの取り組みを導入することで、このギャップに対処するための一歩を踏み出しました。

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、および303人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用いたGPT-4のバリエーションが最も高い正確性率（65.0%）を達成し、他のLLMよりも常に優先されました（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を活用する重要なマイルストーンです。

MedAlign ワークフロー

Source: Fleming et al., 2023

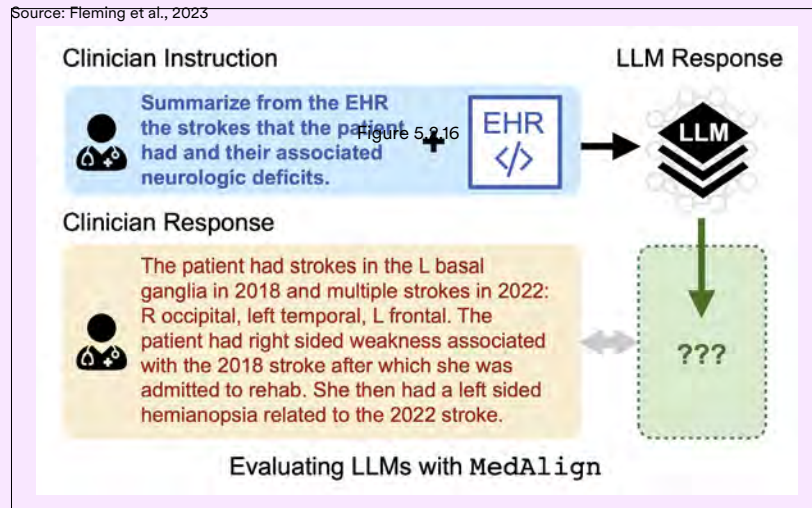


図5.2.22

ハイライトされた研究：

MedAlign（続き）

モデルの性能評価：人間対COMETの順位

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

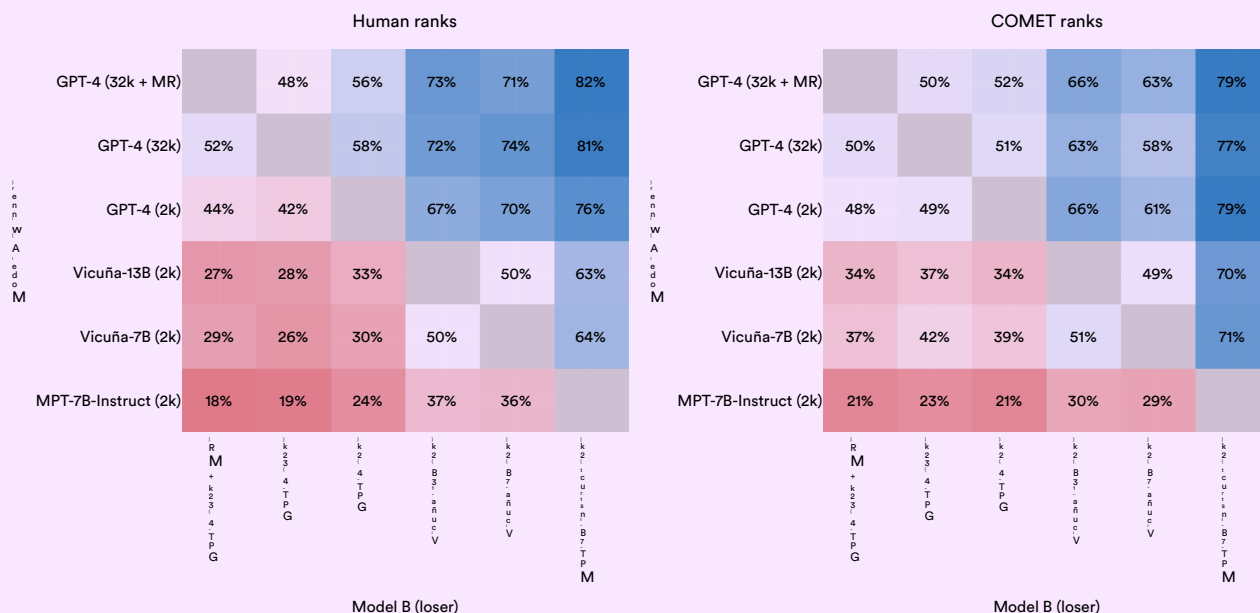
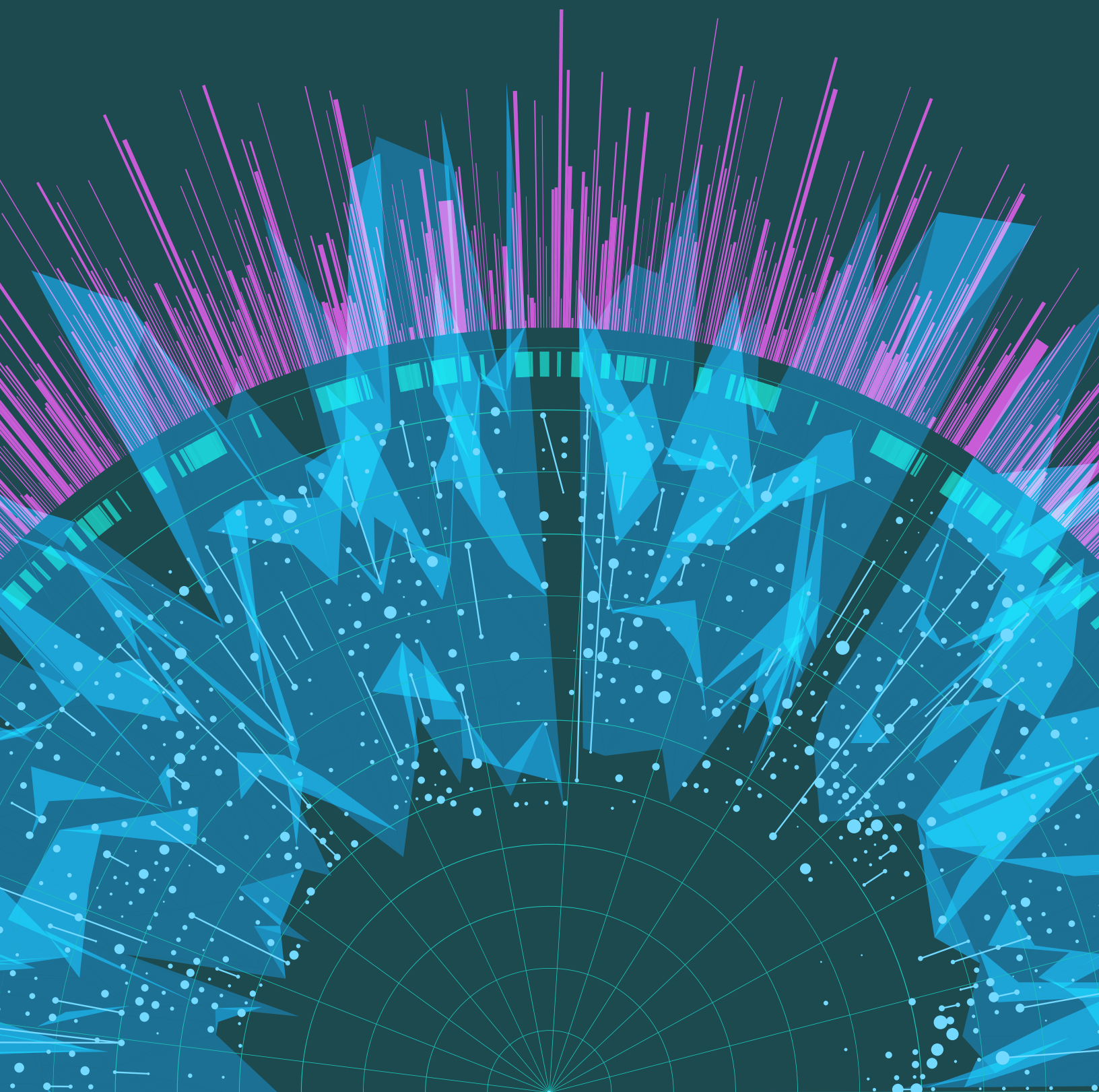


図5.2.23



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の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：教師と学生の間でのChatGPTの使用 364

公開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）教育の動向を、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検討します。AIの教育への影響に関する懸念が高まる中、教師や学生によるChatGPTのような新しいAIツールの使用も調査しています。

分析は、Computing Research Associationの年次Taulbee調査に基づき、アメリカ合衆国とカナダにおける高等教育のCSおよびAI教育の現状の概要から始まる。その後、Informatics Europeのデータを用いてヨーロッパのCS教育についてレビューする。今年は、StudyportalsからのAI関連の英語の学習プログラムの世界的な数に関するデータを含む新しいセクションを導入する。

この章は、Code.org からの米国のK-12コンピュータサイエンス教育に関する洞察と、Walton Foundationの調査による学校でのChatGPTの使用に関する結果で締めくくられます。

章のハイライト

1. アメリカとカナダのCS学士号卒業者数は引き続き増加しており、新しいCS修士号卒業者は比較的横ばいで、博士号卒業者は穏やかに増加しています。While the

新たにアメリカ人およびカナダ人の学士号取得者数は10年以上にわたり一貫して増加している一方で、CSの大学院教育を選択する学生数は横ばいである。2018年以降、CSの修士号および博士号の卒業者数はわずかに減少している。

2. AI博士の産業界への移行は加速し続けている。2011年には、新たにAI博士号を取得した人のほぼ同じ割合が産業界（40.9%）と学界（41.6%）に就職していた。しかし、2022年までには、卒業後に産業界に就職した割合ははるかに高く（70.7%）、学界に入る人（20.0%）と比べて大きく増加している。過去1年だけでも、産業界志望のAI博士の割合は5.3ポイント上昇しており、大学から産業界への頭脳流出が激化していることを示している。

3. 産業界から学界へのアカデミック人材の移動の減少。2019年には、米国とカナダの新しいAI教員の13%が産業界出身だった。2021年までにこの数字は11%に減少し、2022年にはさらに7%に下がった。この傾向は、産業界から学界への高レベルな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. アメリカ合衆国とカナダのCS教育は国際的な側面が薄れる。2022年には2021年よりも比率的に少ない国際的なCSの学士号、修士号、博士号の卒業者が出た。特に修士課程の国際学生の減少が顕著だった。

5. より多くのアメリカの高校生がCSコースを受講しているが、アクセスの問題は依然として残っている。2022年には、201,000件のAP CS試験が実施された。2007年以降、これらの試験を受ける学生数は10倍以上に増加している。しかし、最近の証拠は、大規模な高校に通う学生や郊外の地域にいる学生の方がCSコースにアクセスしやすいことを示している。

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年以来3倍になっており、過去5年間にわたり着実に年々増加しています。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. イギリスとドイツは、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ます。イギリスとドイツは、ヨーロッパで最も多くの新しい情報学、CS、CE、および情報学の学士、修士、博士号の卒業生を輩出しています。人口比率で見ると、フィンランドは学士号と博士号の卒業生の生産でリードしており、アイルランドは修士号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、CSとAI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、博士号を含む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CSとAIの教育

アメリカ合衆国とカナダ

このサブセクションでは、アメリカ合衆国とカナダのCSおよびAIの高等教育の現状を評価する、Computing Research AssociationのTaulbee調査のデータ分析を紹介します。この調査は、アメリカ合衆国とカナダの297の博士号授与CS学科を対象としています。¹

CS学士卒業者

過去10年間で、北米の新しいCS学士号取得者の総数は着実に増加しており、3倍以上に増加しています。2021年から2022年にかけては前年比7.9%の増加を示しています（図6.1.1）。

アメリカ合衆国とカナダの新しいCS学士号卒業者数（2010–22）

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

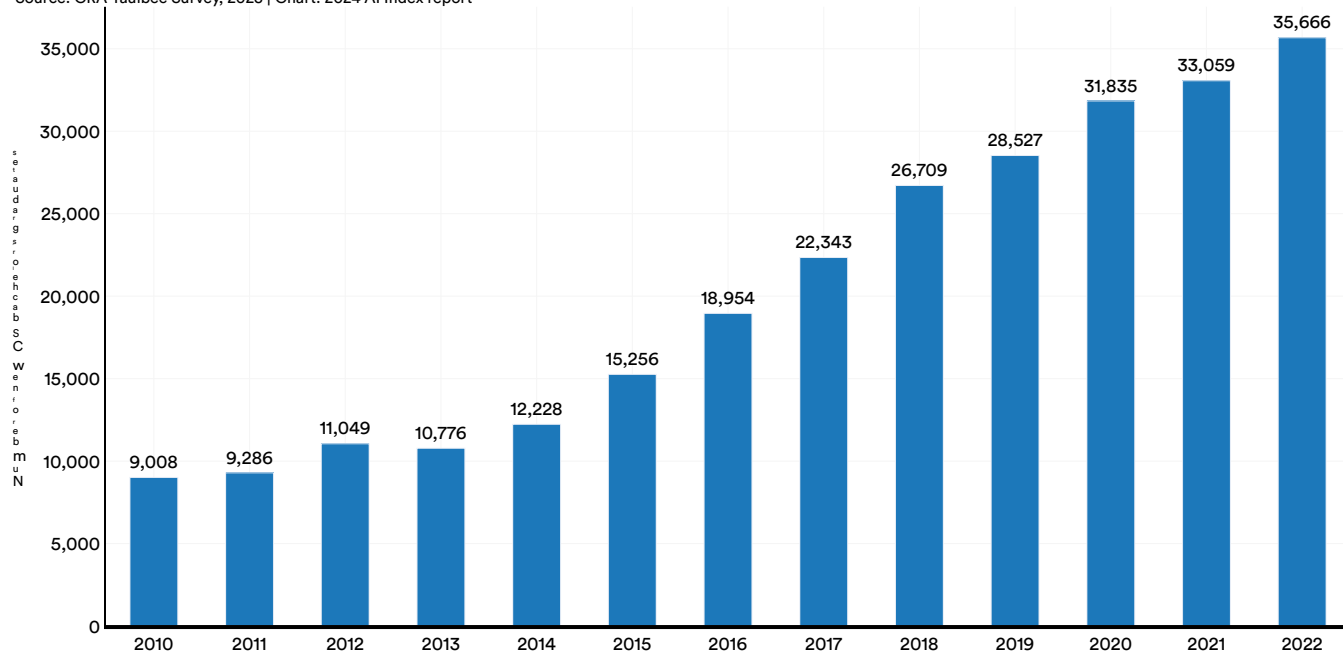


Figure 6.1.1

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に注意することが重要です。対象となった297学科のうち、回答があったのは182学科のみであり、全体の回答率は61%となっています。

アメリカおよびカナダの大学のCS学士号卒業生の中で、国際学生の割合がほぼ8年ぶりに減少し、2021年の16.3%から2022年の15.2%に下がった（図6.1.2）。この減少は、トランプ政権初期の留学ビザ取得の難しさの増加を反映している可能性が高く、その影響は今になって初めて顕在化し始めている。

データ。減少は、COVID-19パンデミック中に課された国際旅行制限にも部分的に起因しており、これにより留学生がアメリカ合衆国やカナダで学ぶ能力に影響を与えました。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。

アメリカ合衆国とカナダにおける新しい国際的なCS学士号卒業生（総数に対する割合）

Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report、2010年～2022年

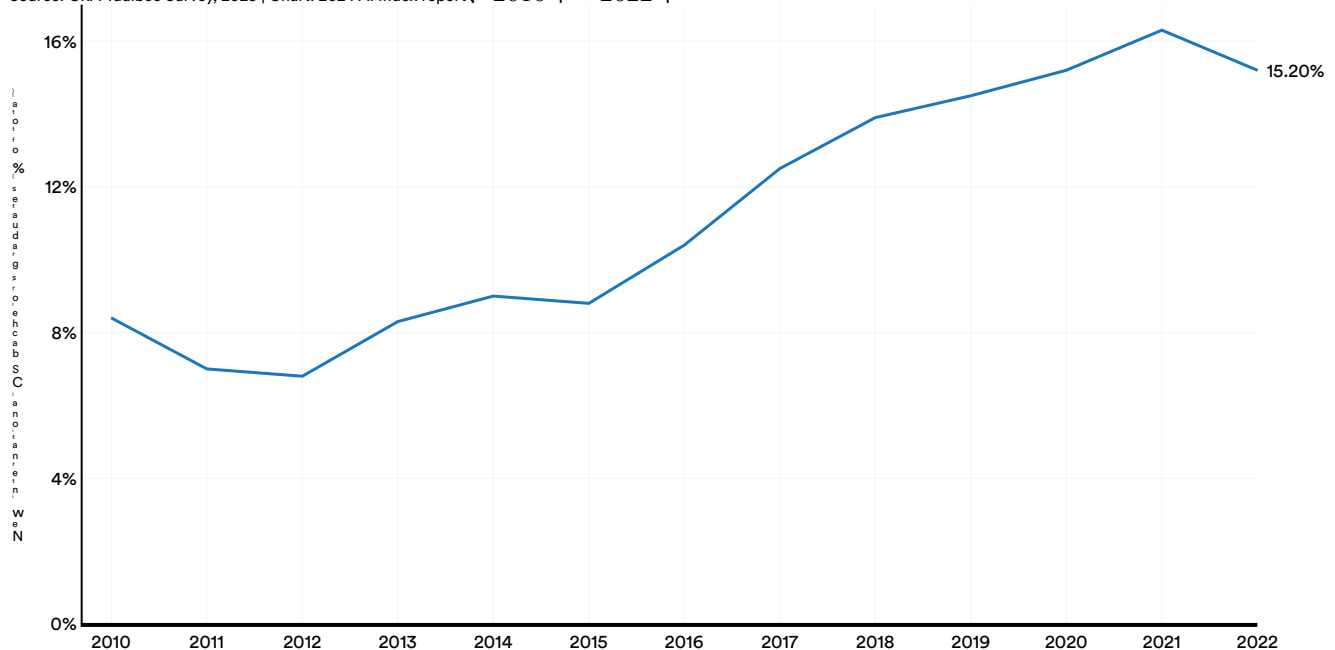


図6.1.2

CS修士卒業者

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修士卒業者の総数は2倍以上になりましたが、

その数は2018年以降横ばいになっており、昨年はわずかに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る国際修士課程の学生数の減少の反映です。

2010年から2022年までのアメリカ合衆国とカナダの新しいCS修士号取得者

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

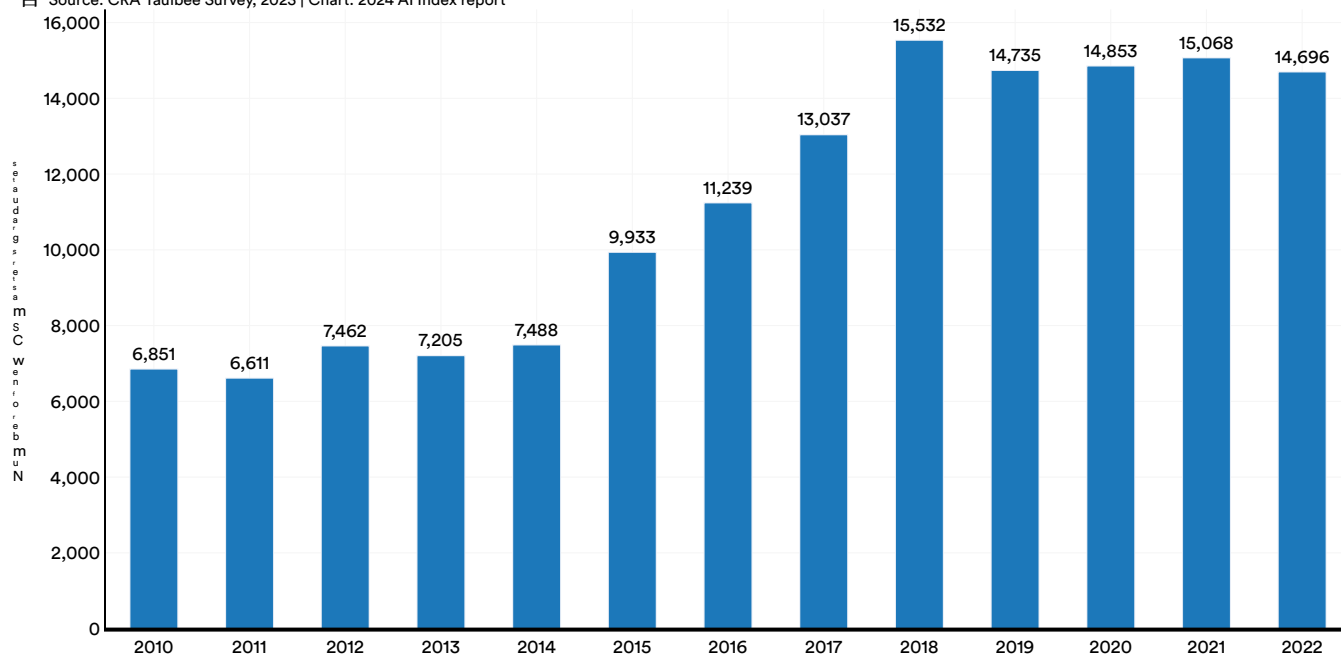


図6.1.3

2022年、アメリカとカナダの大学では、国際的なCS修士学生数が顕著に減少しました。この下降傾向は2017年頃に始まりましたが、最も顕著だったのは昨年で、14.8パーセンテージポイントの減少でした（図6.1.4）。現在、国際的な修士卒業生と国内の修士卒業生の比率はほぼ同じです。

2010年から2022年までの米国とカナダにおける新しい国際CS修士卒業生（総数に対する割合）

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

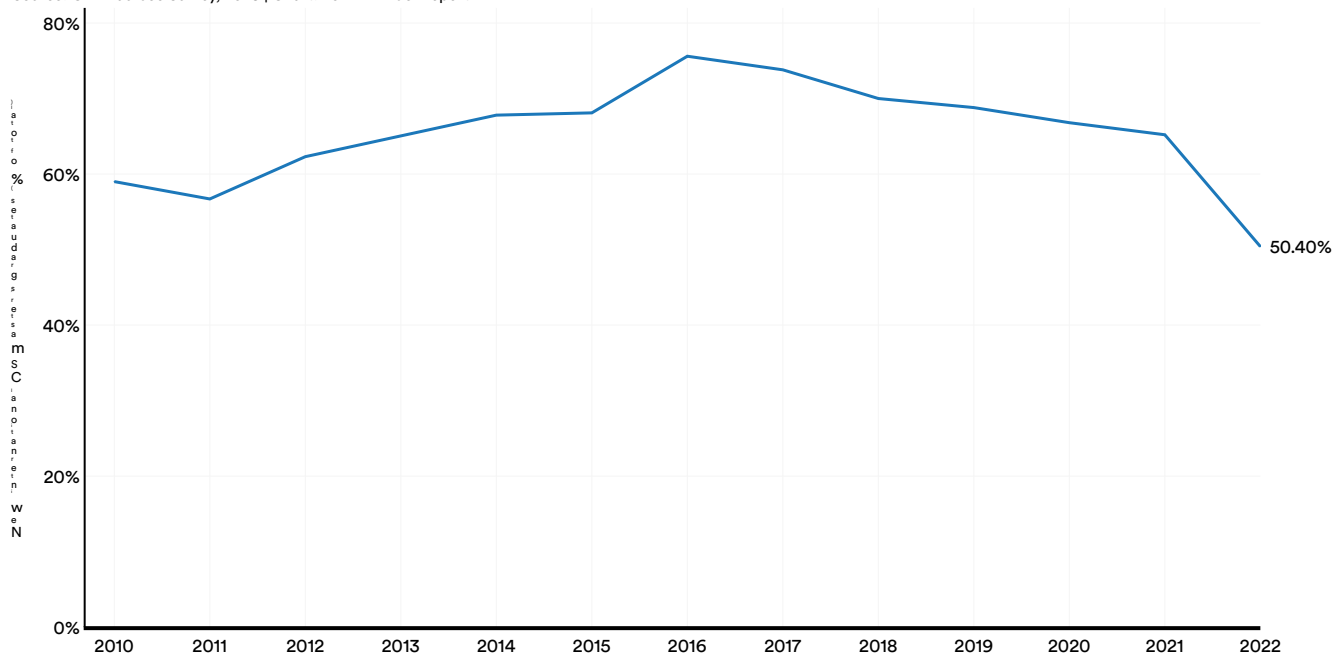


図6.1.4

CS博士卒業者

10年ぶりに、アメリカとカナダの大学で新たにCS博士号取得者の数が大幅に増加しました。2022年には、CS博士号取得者数は2,105人に達し、2010年以来の最高となりました（図6.1.5）。

アメリカ合衆国とカナダの新しいCS博士号取得者数、2010–22

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

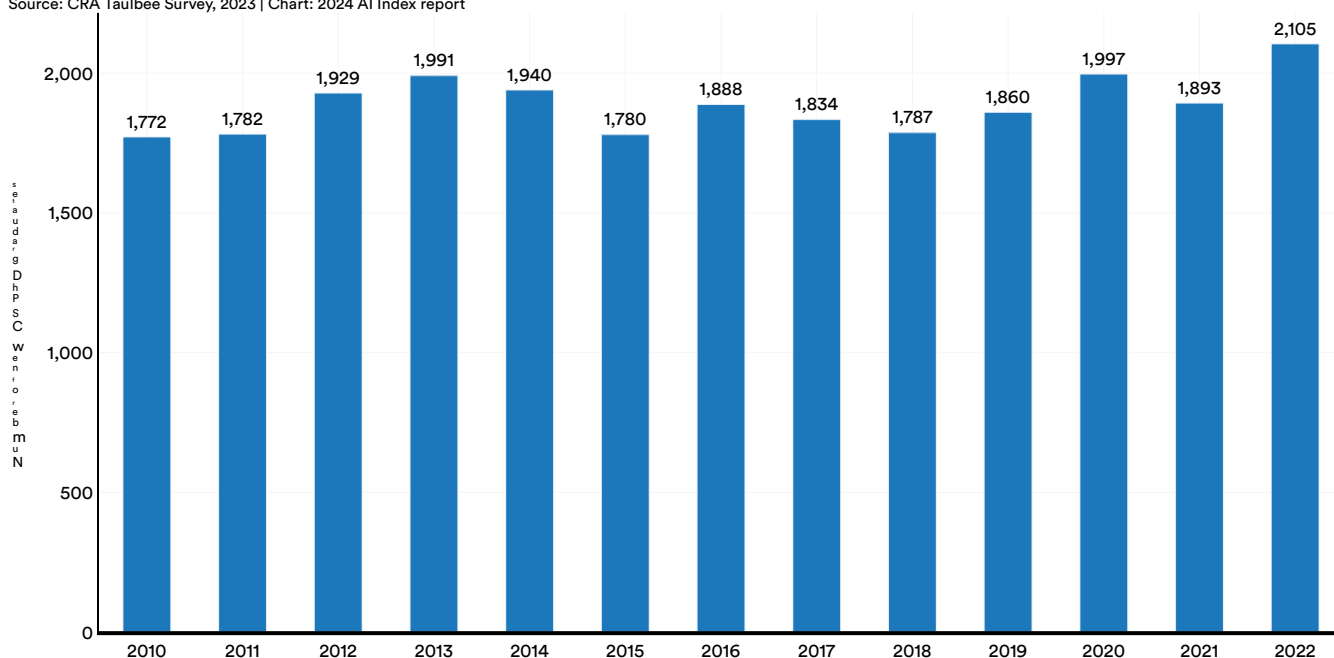


図6.1.5

過去10年間にわたり、CS博士号取得者の中の留学生の割合は増加してきましたが、昨年はこの割合がわずかに減少し、2021年の68.6%から2022年の65.9%に下がりました（図6.1.6）。

2010年から2022年までの米国とカナダにおける新しい国際的なCS博士号取得者（総数に対する割合）

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

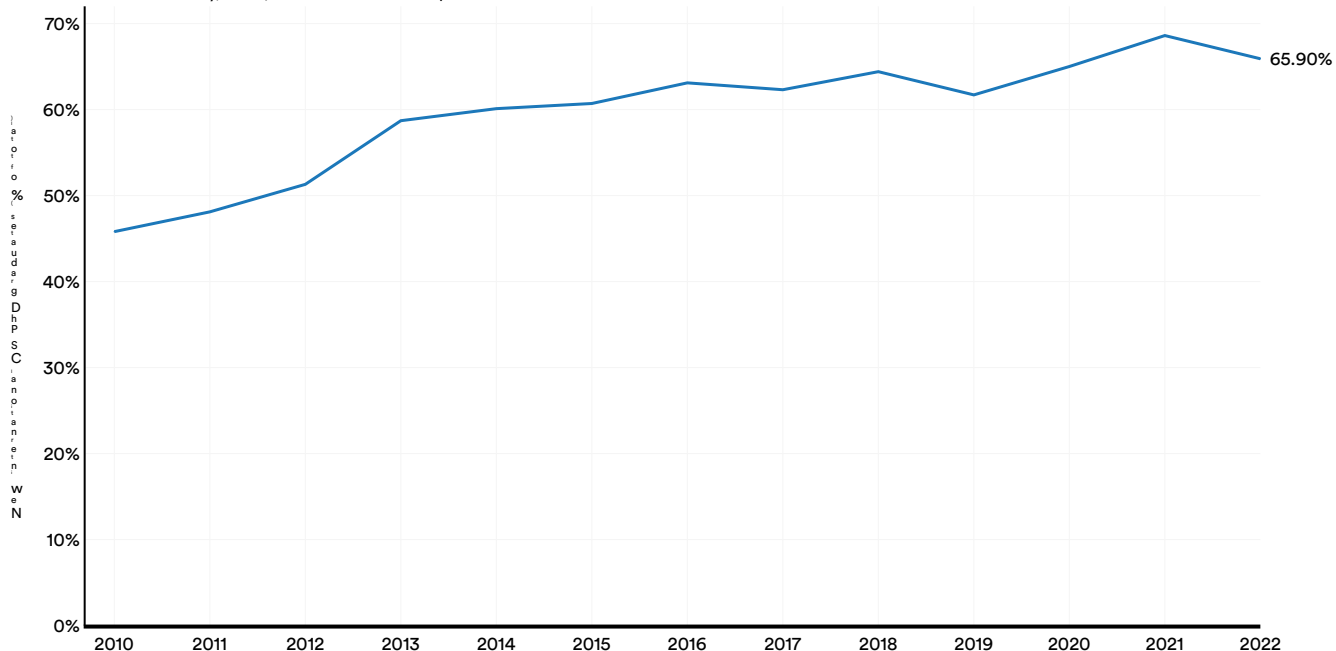


図6.1.6



新たに博士号を取得したAI研究者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傾向に続き、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（40.9%）と学界（41.6%）での就職割合がほぼ同じでした。

しかし、2022年までに、卒業後に産業界に就職した割合ははるかに高く（70.7%）、学界に進む者（20.0%）と比較して大きな差がありました。新たにAI博士号を取得した人のうち、政府の役割に就く割合は過去5年間ほぼ一定で約0.7%と比較的低いままです。

2010年から2022年までの米国とカナダにおける
セクター別新しいAI博士号取得者（総数に対する割合） Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

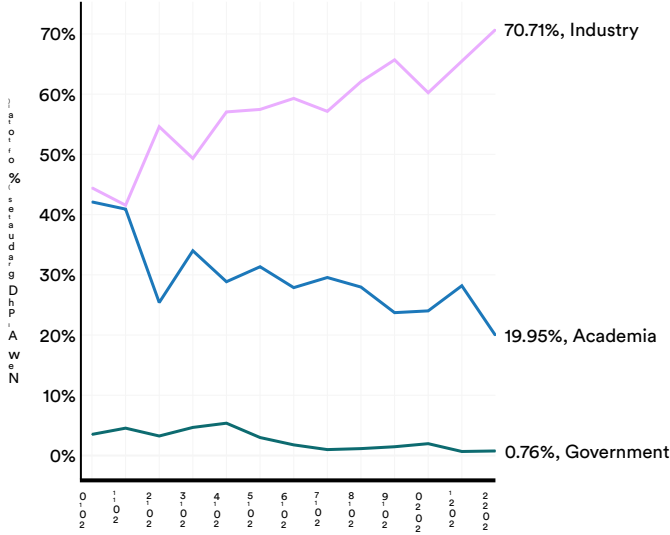


Figure 6.1.7²

アメリカ合衆国における新しいAI博士号取得者の雇用状況と
カナダ（2010年～2022年）のセクター別

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

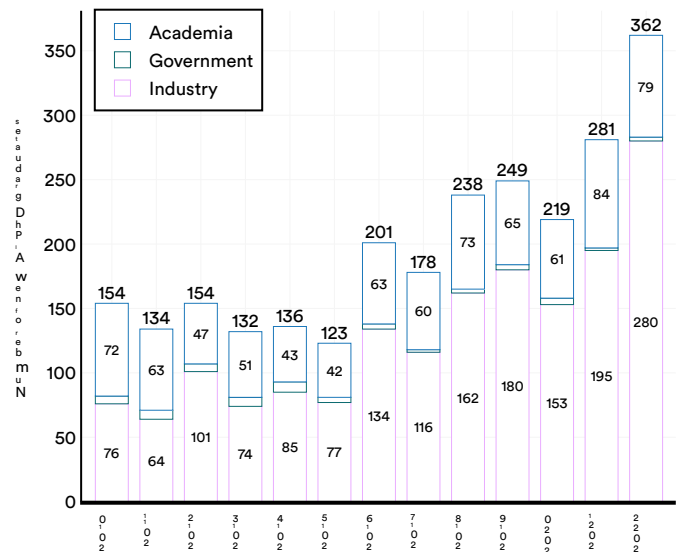


Figure 6.1.8

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とAI教育の傾向をよりよく理解するために、CS教員に関するデータを調査することは役立ちます。昨年、アメリカとカナダの大学におけるCS、CE、情報教員の総数は7.2%増加しました（図6.1.9）。2011年以降、その増加率は42.4%です。

アメリカ合衆国とカナダにおけるCS、CE、および情報学部の人数（2011年–2022年）

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

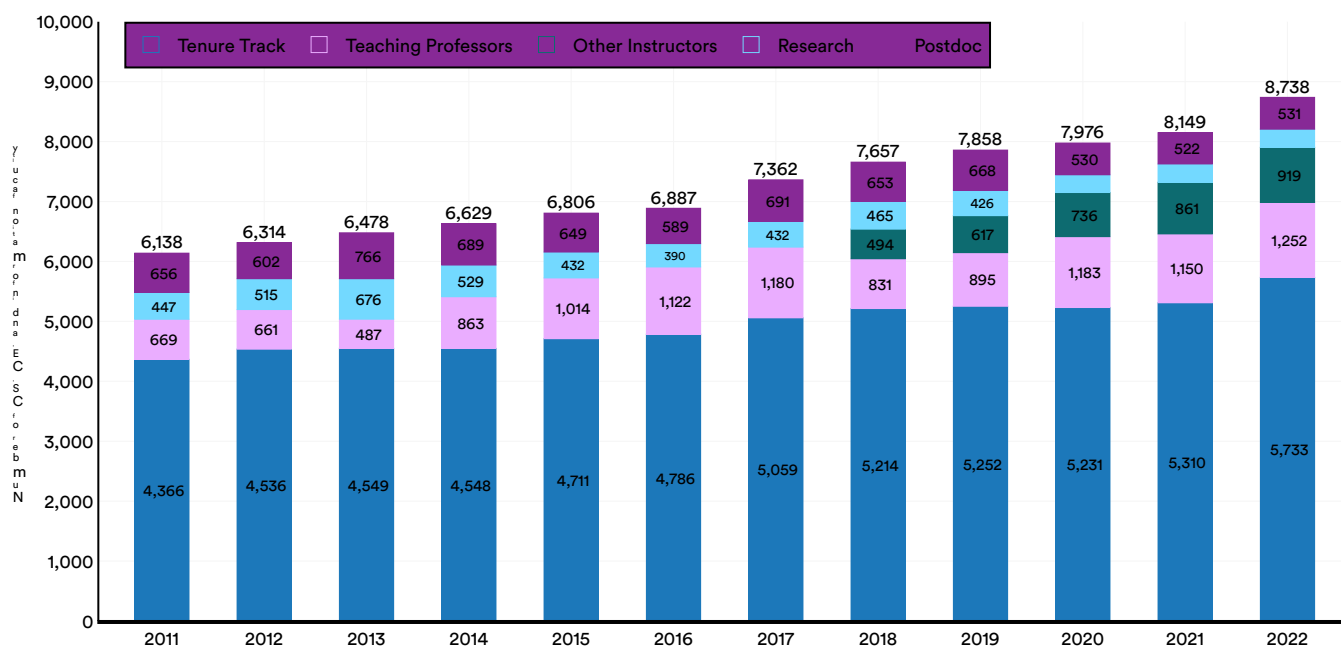


図6.1.9

2022年、アメリカ合衆国のCS教員数は7,084人で、そのうち大多数（65.7%）がテニユアトラックに就いている（図6.1.10）。アメリカのCS教員の総数は2021年から4.4%増加し、2011年からは45.0%増加している。

アメリカ合衆国のCS教員数、2011-22

出典：CRA Taulbee調査、2023年 | 図：2024年AI指数レポート

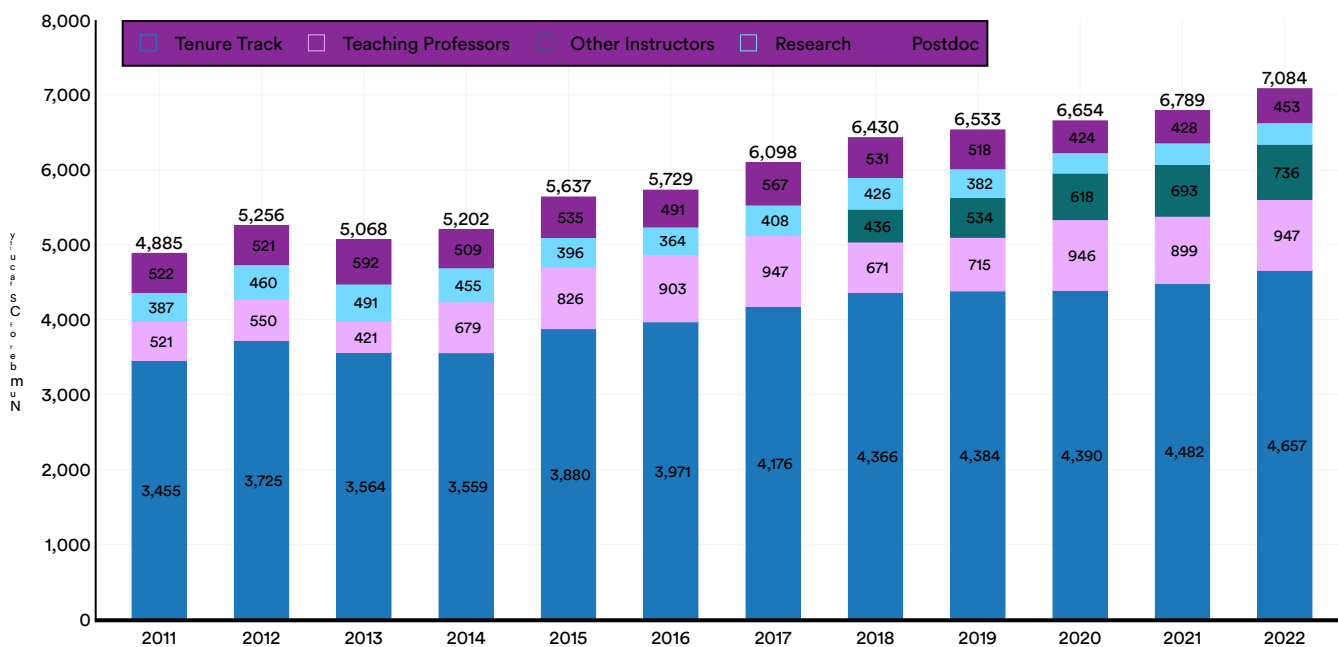


図6.1.10

昨年、北米のCS、CE、情報分野で915人の新しい教員が採用され、10年ぶりの高水準となった。これらのポジションのうち455件はテニュアトラックだった（図6.1.11）。

2011年から2022年までの米国とカナダにおける新しいCS、CE、および情報学部門の採用

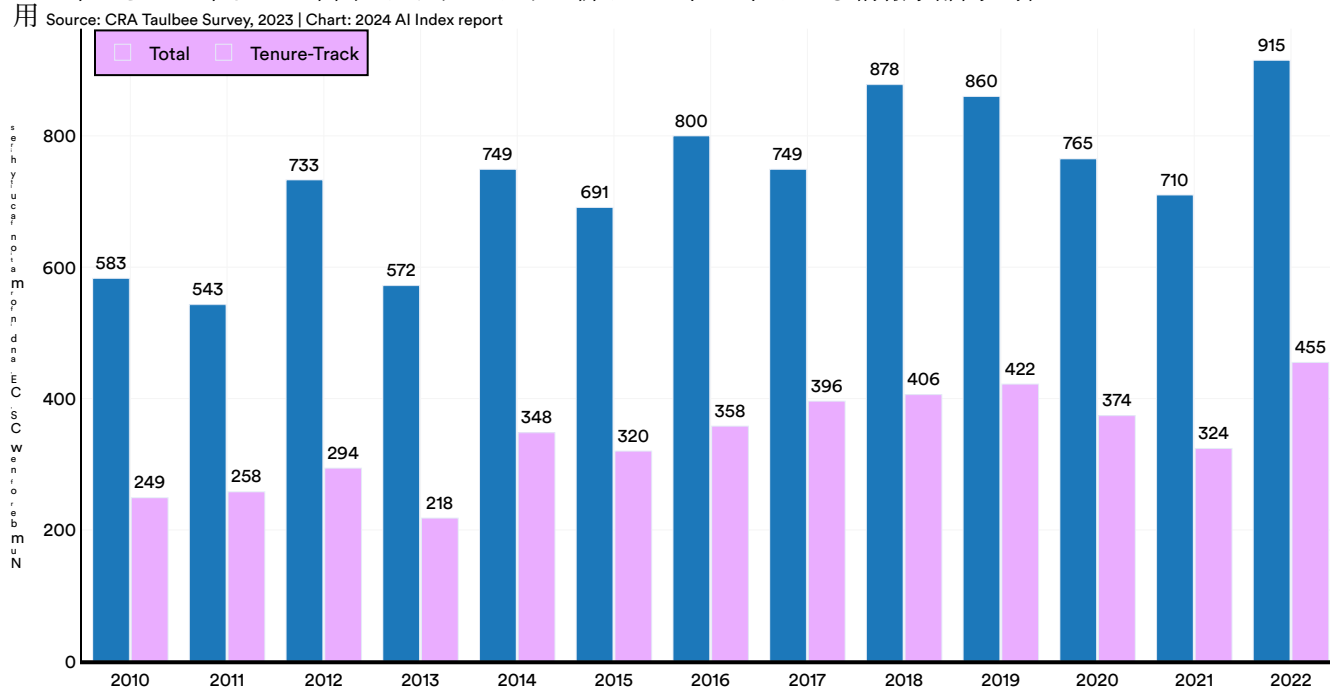


図6.1.11

2022年には、新任教員の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」教員が他の場所のポジションを空けたため、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年に産業界から移行した教員の割合は、前年の11%および2019年の13%から減少し、7%になりました。

2018年から2022年までのアメリカおよびカナダのCS、CE、情報学部門における新しい教員の出所

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

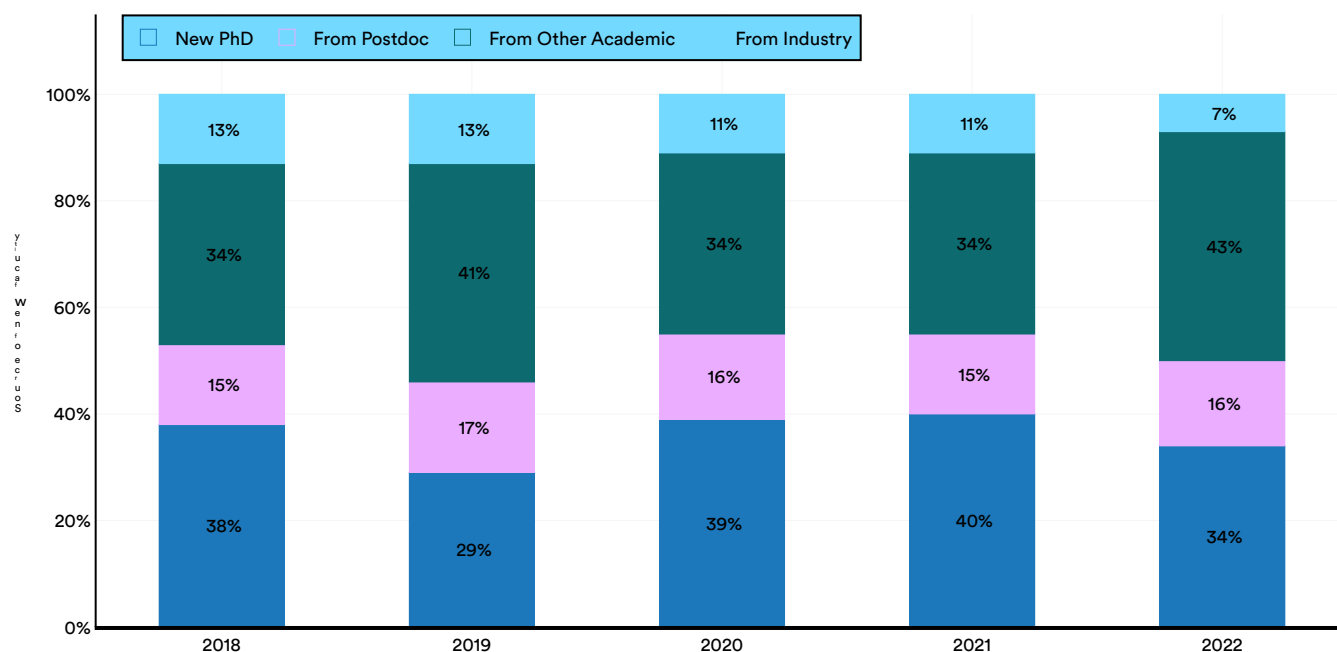


図6.1.12

教員ポジションが埋まらない理由は過去10年間でさまざまに変化しています。2011年には、失敗した採用活動の37%が提案がなかったことによるものであり、34%は提案があったものの辞退されたものでした（図6.1.13）。これに対し、2022年には提案がなかったのはわずか15%であり、55%は提案があったものの辞退されたものでした。

断られました。この傾向は、新しいCS教員のための競争がますます激しくなっている市場を反映しているようです。しかし、これは他の学術職や業界の職と競争が激化していることを示しているのかどうかは依然として不明です。

2011年～2022年に新しいCS、CE、および情報学部のパポジションが未充足のままであった理由（総数に対する割合） Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

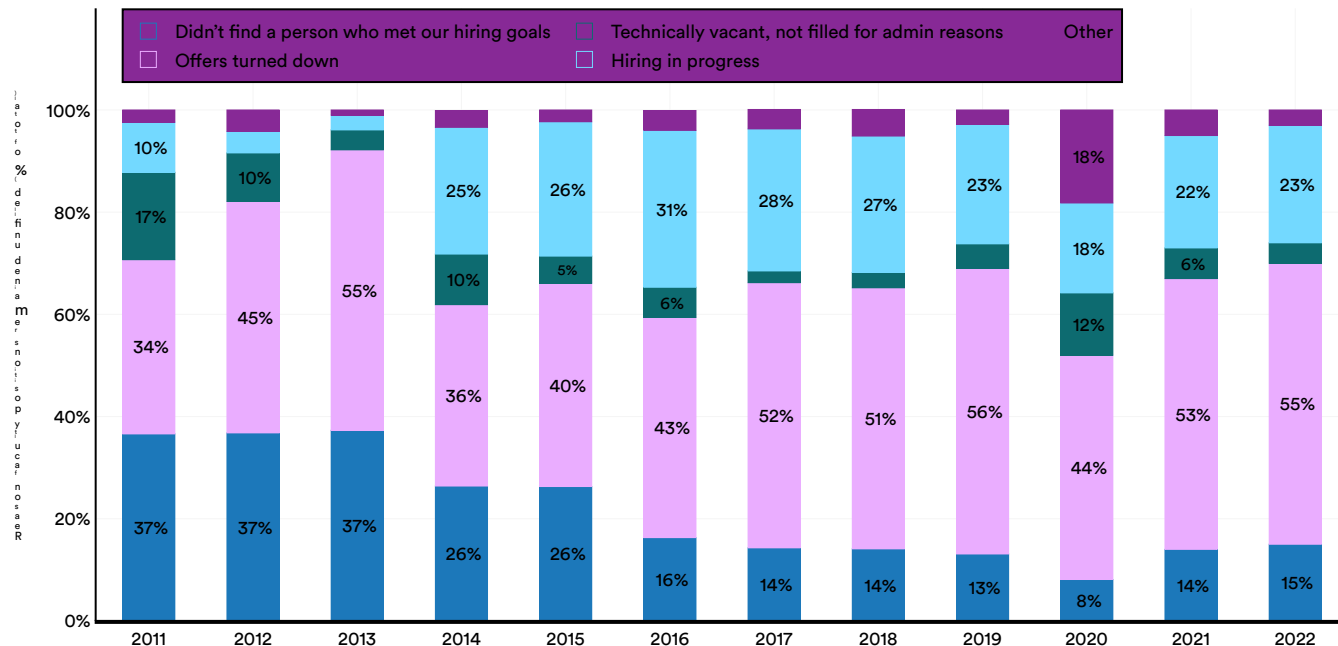


図6.1.13

2022年、北米のCS、CE、情報分野の学科では、教員の退職が大幅に増加し、合計405人となり、2021年の303人と比較して増加しました（図6.1.14）。これらの退職のうち、38.5%は他の学術職に移った一方、16.3%は非学術的な役割に移動し、前年まで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、情報学部門における教員の損失、2011–22

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

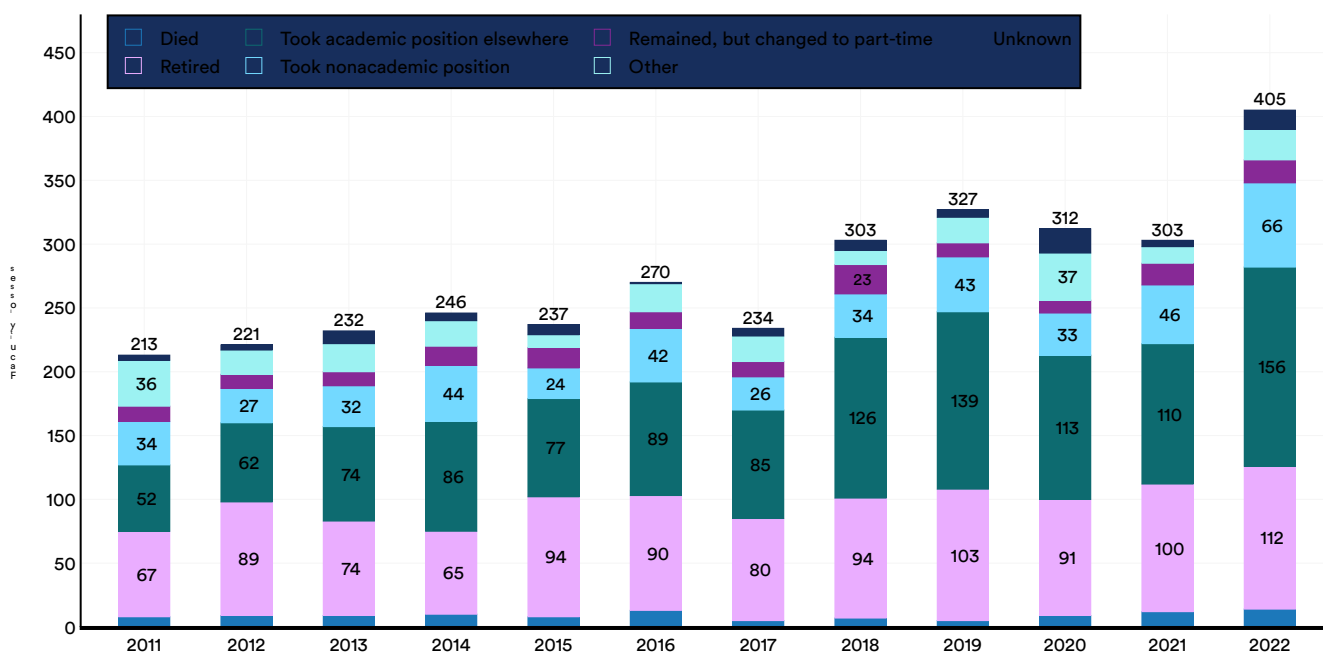


図6.1.14

2015年以降、准教授と助教授の中央値給与はインフレ率をわずかに上回る一方で、教授の中央値給与の9か月間の増加は米国のインフレ率をわずかに下回っています。2022年には、教授の給与は2021年より3.2%増加しましたが、米国のインフレ率7%には追いつかず、2015年より16.4%増加しましたが、これもその期間のインフレ増19%には届いていません（図6.1.15）。

2015年から2022年までの米国のCS教員の9か月平均給与

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

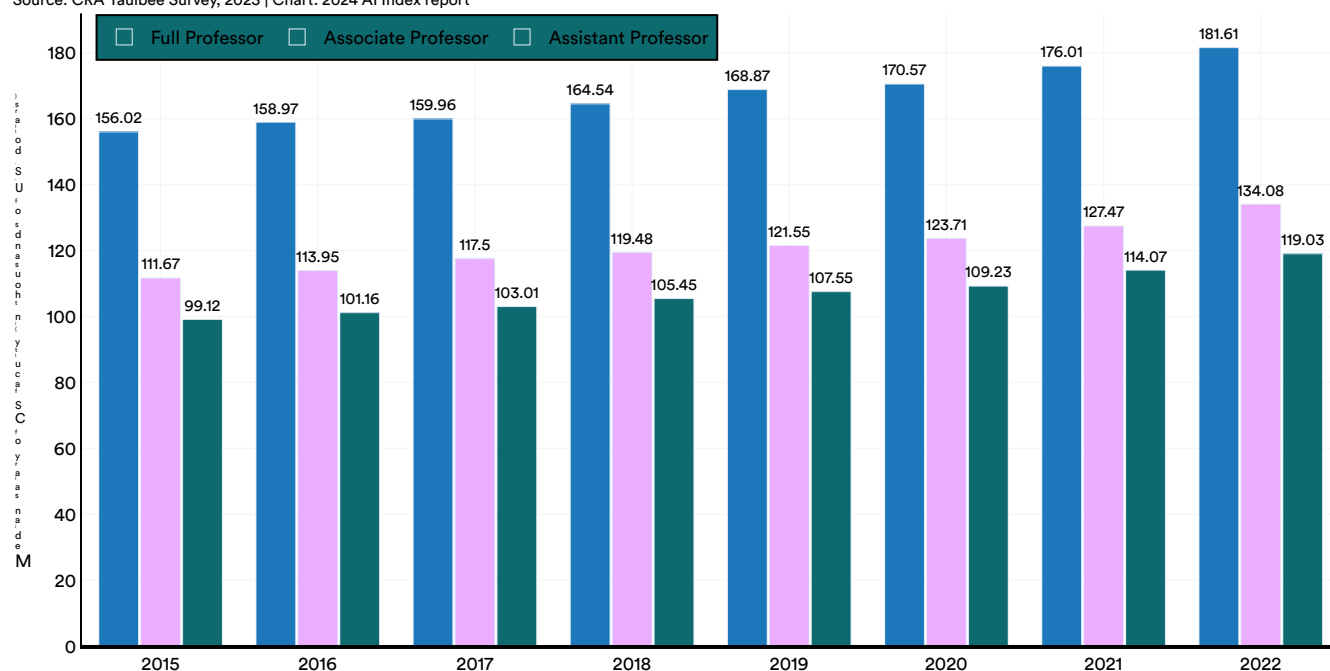


図6.1.15

2022年、CS、CE、情報分野の新任任期追求教員の国際採用者の割合は、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2番目に高い割合であり、2013年に次ぐものです。

アメリカ合衆国とカナダにおける新しい国際的なCS、CE、および情報学のテニユアトラック教員採用（総数に対する割合）、2010年～2022年

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

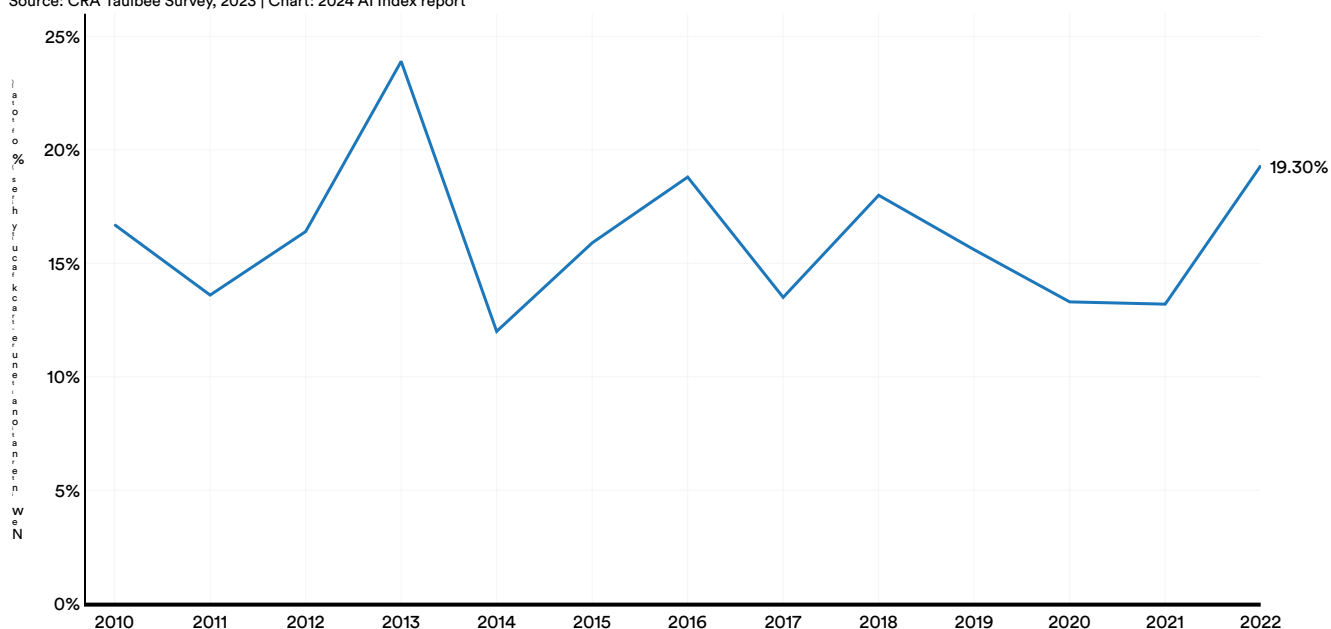


Figure 6.1.16

ヨーロッパ

ヨーロッパのCS卒業生に関するデータは、学術・研究コミュニティであるInformatics Europeから得られます。同コミュニティは、その他の目的のほか、ヨーロッパにおける情報学教育の状況を監視しています。³ Informatics Europeは、ヨーロッパの政府の統計局から、情報学、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

情報学、CS、CE、IT学士卒業生

2022年、イギリスは学士レベルで情報学、CS、CE、ITの新卒者数が最も多く、約25,000人に達した（図6.1.17）。⁵ ドイツとトルコがそれに続いた。サンプルのほとんどの国では、10年前と比べてこれらの分野の卒業生数が増加したが、ポーランド、スペイン、チェコ共和国など例外もあった（図6.1.18）。

2022年のヨーロッパにおける新しい情報学、CS、CE、IT学士号取得者数（国別）

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

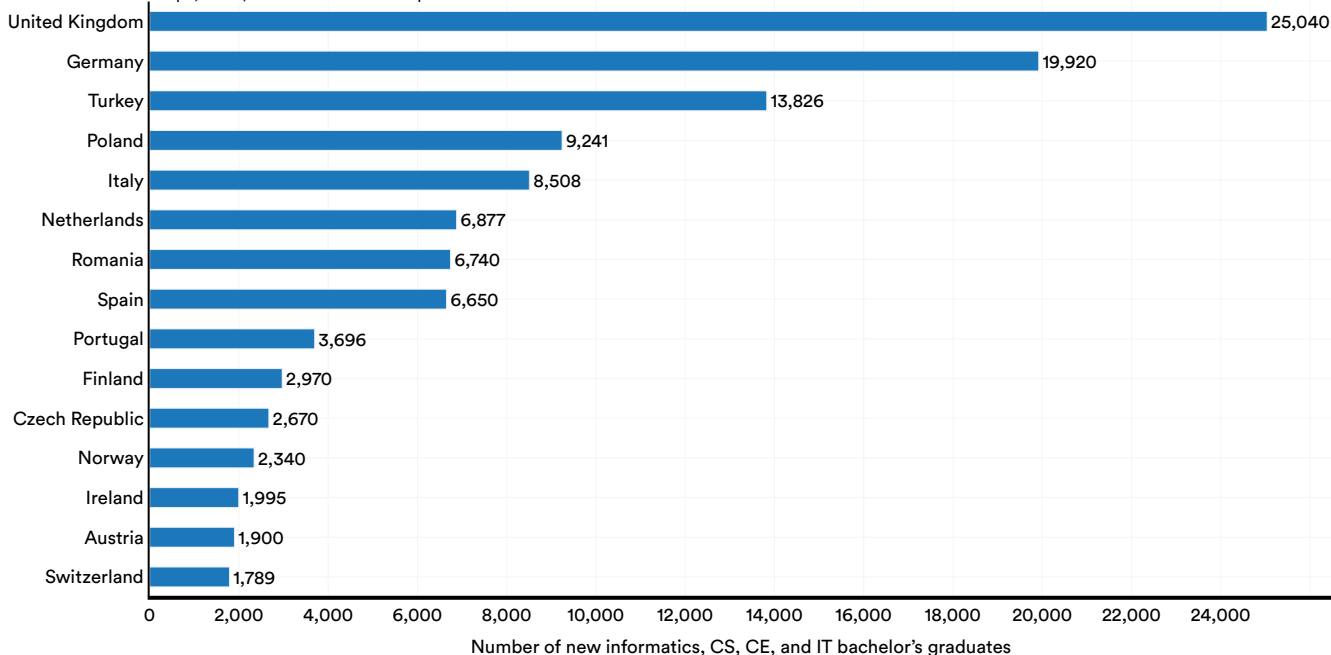


図6.1.17

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北米データと後述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、1人当たりの比較を行わないよう注意喚起されます。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.

ヨーロッパの国別に見る新しい情報学、CS、CE、IT学士卒業者の割合変化（2012年対2022年）

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

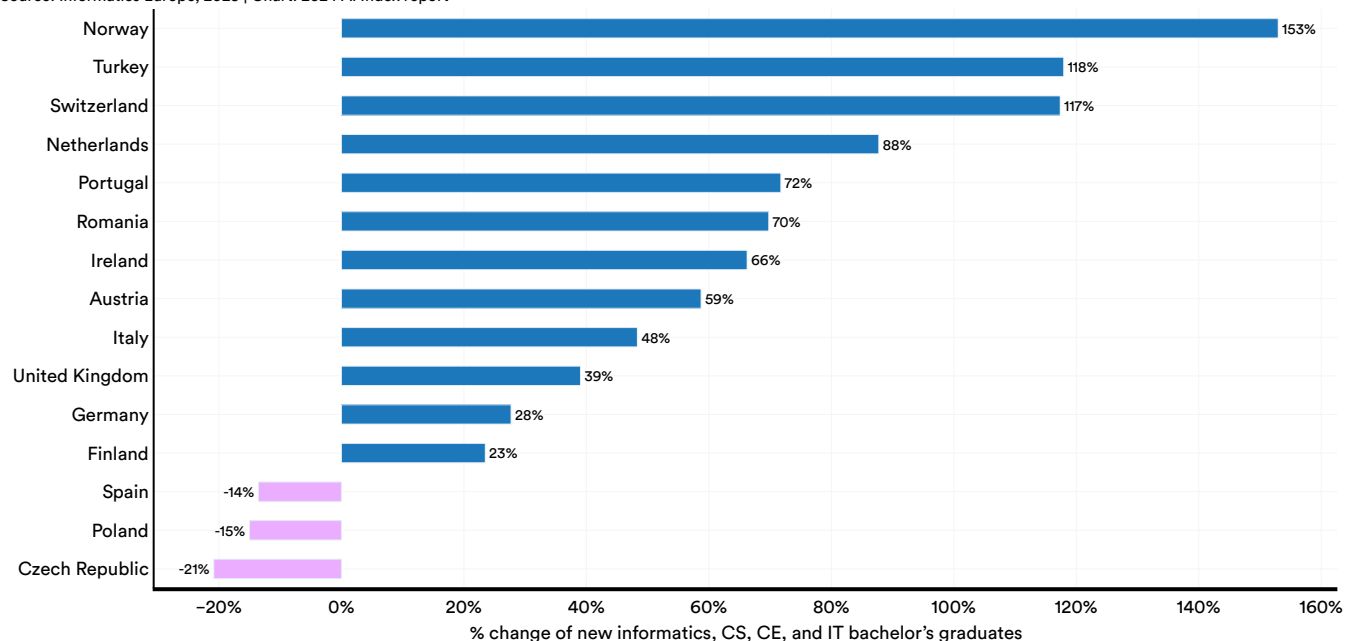


図6.1.18

フィンランド（53.4）、ノルウェー（42.6）、オランダ（38.6）が、100,000人当たりの新しい学士号取得者数（情報学CS、CE、IT）でリードしています（図6.1.19）。一人当たりの観点では、サンプルされたヨーロッパ諸国のほとんどが、情報学、CS、CE、ITの学士号取得者の総数の増加を示しています（図6.1.20）。