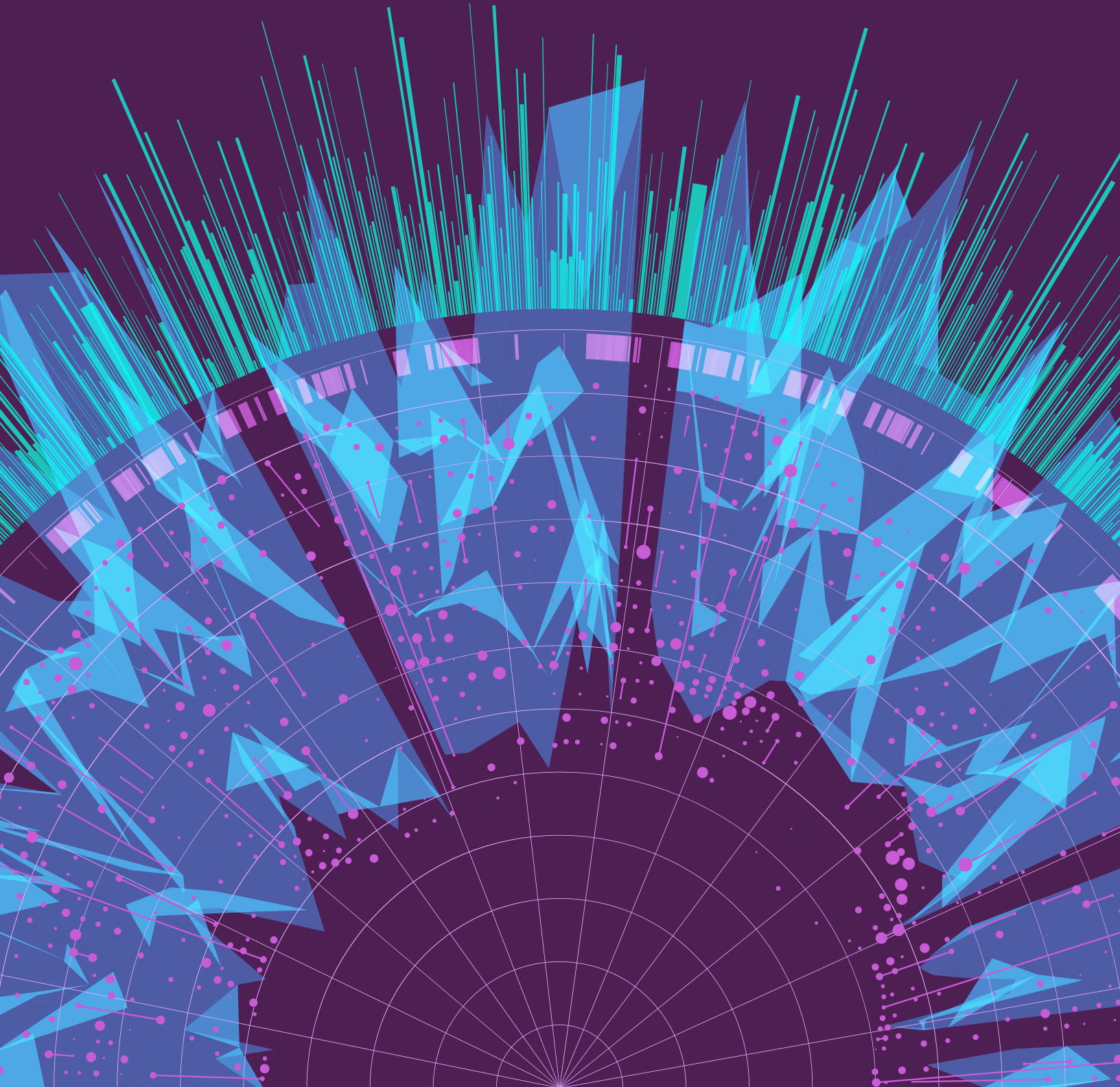




Artificial Intelligence
Index Report 2024

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インデックスでは、科学および医療におけるAIの役割の拡大を認識し、科学と医療におけるAIに関する新しい章を導入しています。2023年の注目すべきAI支援の科学的成果を探り、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を含みます。この章では、医療AIシステムの性能、SynthSRやImmunoSEIRAのような2023年の重要なAI駆動医療イノベーション、そしてFDAのAI関連医療機器の承認動向も検証しています。

章のハイライト

1. AIのおかげで科学の進歩はさらに加速しています。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支援するGNoMEまで、さらに重要な科学関連のAIアプリケーションが登場しました。

2. AIは医療の大きな進歩を助けます。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活用されています。

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重要なテストであるMedQAベンチマークで著しい改善を示しています。2023年の注目モデルであるGPT-4 Medpromptは、90.2%の正確率を達成し、2022年の最高スコアから2.6ポイントの増加を記録しました。2019年のベンチマーク導入以来、MedQAにおけるAIのパフォーマンスはほぼ3倍に向上しています。

4. FDAはますます多くのAI関連医療機器を承認しています。2022年には、FDAは139件の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は現実の医療目的でますます活用されています。

このセクションでは、AIインデックス運営委員会によって選ばれた2023年の重要なAI関連の科学的ブレークスルーを紹介します。

5.1 注目すべき科学的マイルストーン

AlphaDev

AlphaDevはより高速なソートアルゴリズムを発見する

AlphaDevは、計算アルゴリズムの改良分野における科学者や技術者の数十年にわたる研究を進化させた新しいAI強化学習システムです。AlphaDevは、既存の人間のベンチマークよりも少ない命令でアルゴリズムを開発しました。

短いシーケンスに対する基本的なソートアルゴリズム、例えばSort 3、Sort 4、および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対する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。

AlphaDev と人間のベンチマークの比較：アルゴリズムの長さ

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

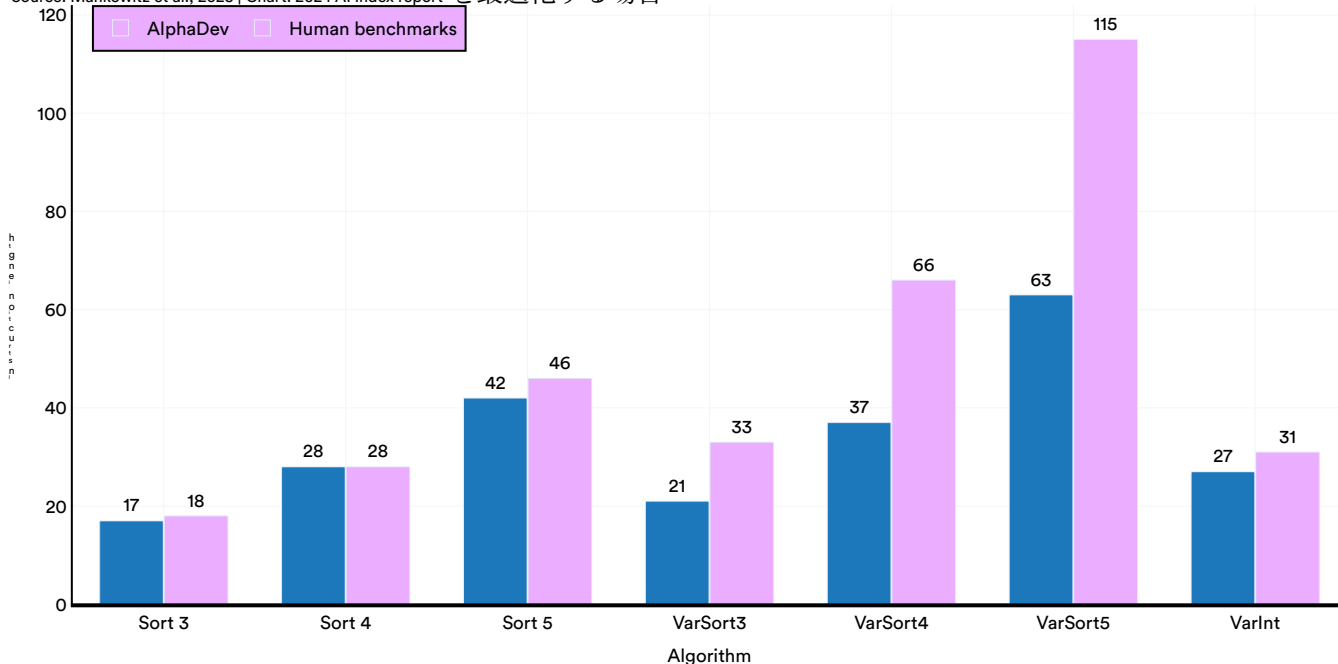


Figure 5.1.1

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、医療画像処理、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、解像度の制限、構造的硬直性、数値的不安定性に悩まされることが多く、その結果{v*}に影響を及ぼします。

品質。FlexiCubesは、勾配ベースの最適化と適応可能なパラメータ（図5.1.2）にAIを活用することで、これらの制限のいくつかに対処しています。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。微分可能な等値面抽出を用いた他の主要なメッシュ再構築手法と比較して、FlexiCubesは基礎となる真実により密接に一致するメッシュ抽出を実現しています（図5.1.3）。

サンプルFlexiCubes表面再構築 Source: Nvidia, 2023

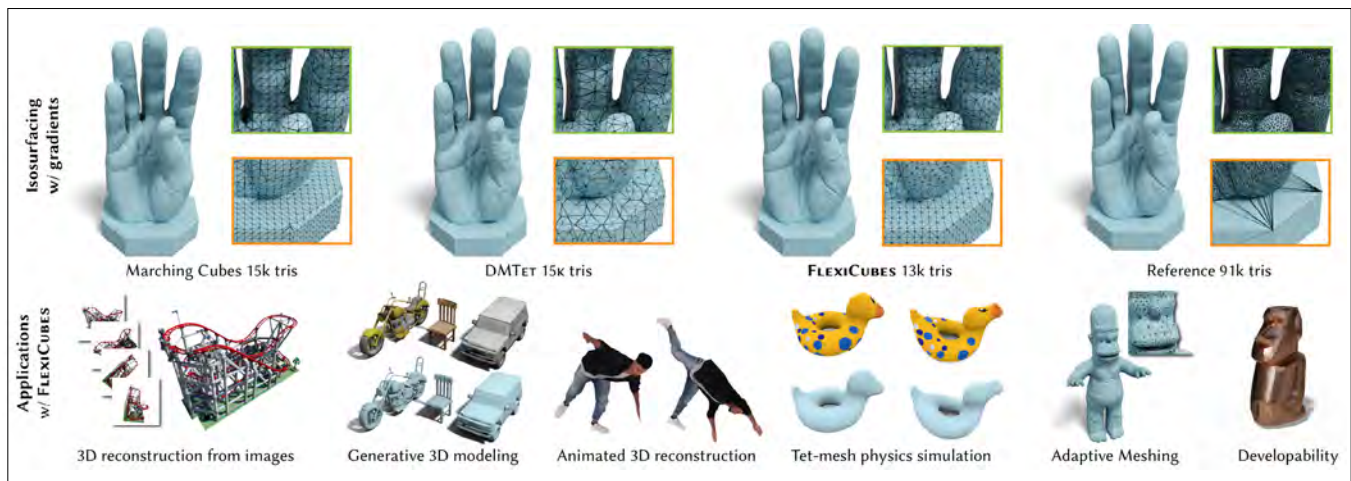


図5.1.2

3Dメッシュ再構築に関する定量的結果の選択

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

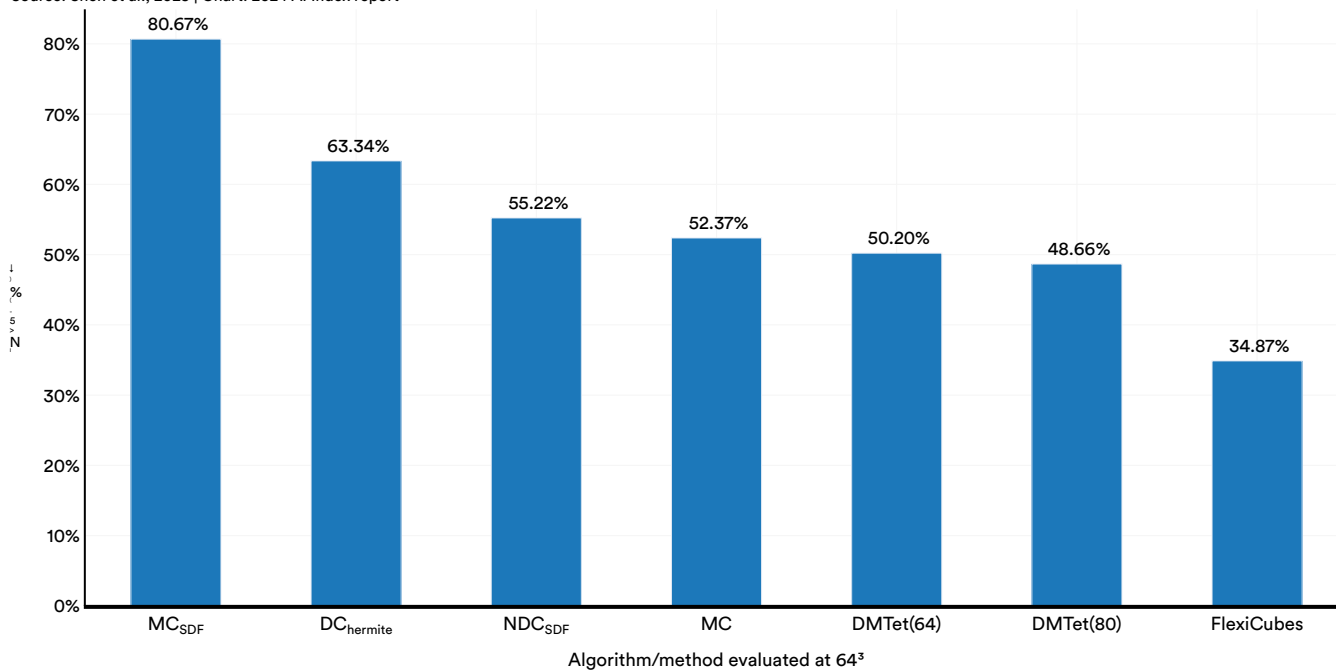


図5.1.3

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、および実験を実施するための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クローズドループフィードバック機構により、Synbotは既存の参考値に等しいかそれを超える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]の合成を目的とした実験では、Synbotは収率が{v*}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

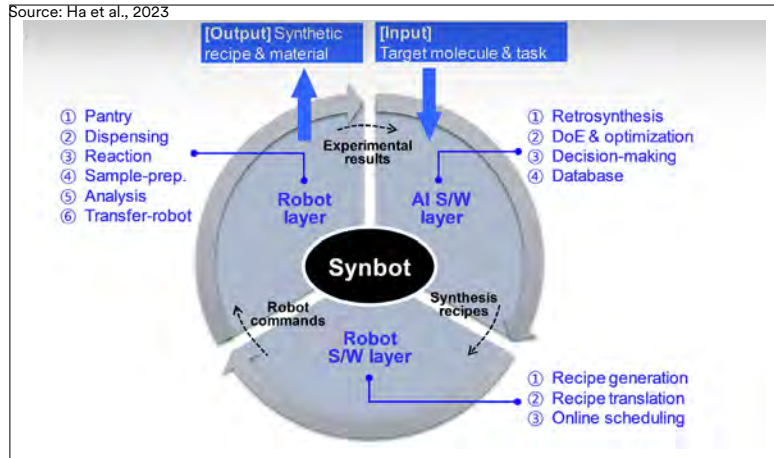


図5.1.4

中間の80%参照範囲内で合成を完了し、著しく短時間で合成を完了しました（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

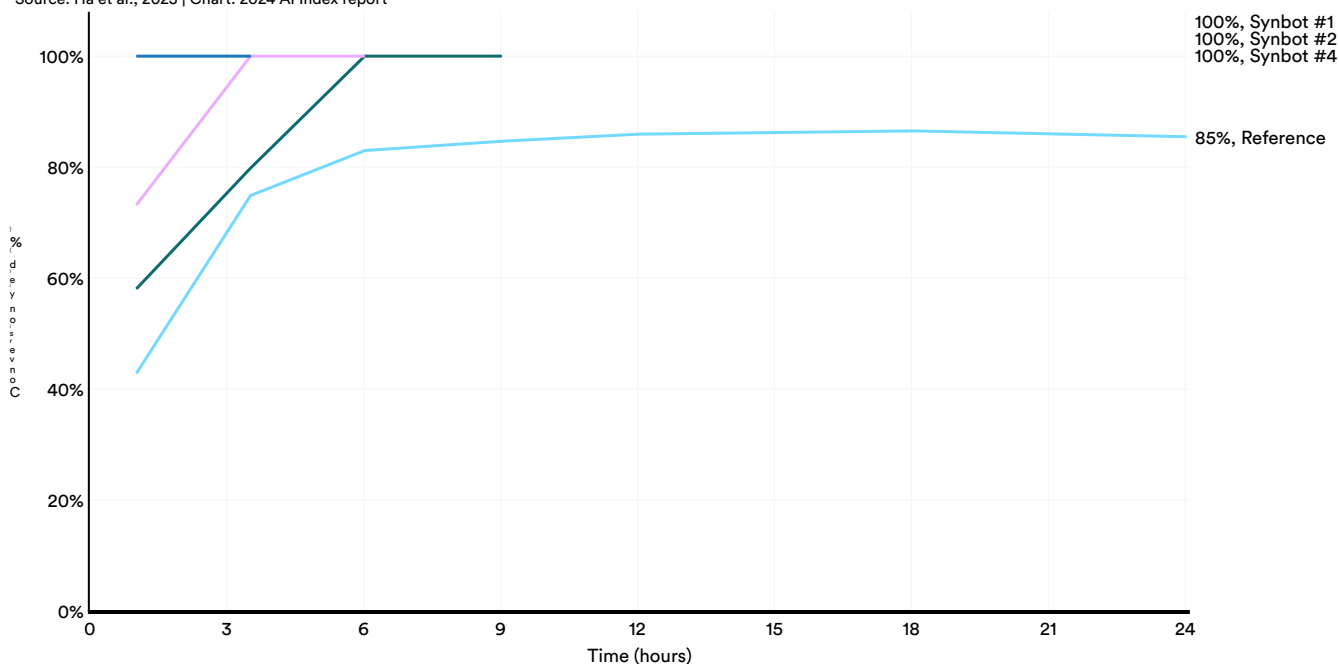


図5.1.5

GraphCast

より正確な世界の天気予報

GraphCastを使って

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活用し、GraphCastは膨大なデータセットを処理して気温、風速、大気条件を予測します。

その他にも。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はより低い二乗平均平方根誤差を示しており、これはその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明、極端な気象事象への備えの強化、そして地球規模の気候研究への貢献において貴重なツールとなり得ます。

GraphCast 天気予報

Source: DeepMind, 2023

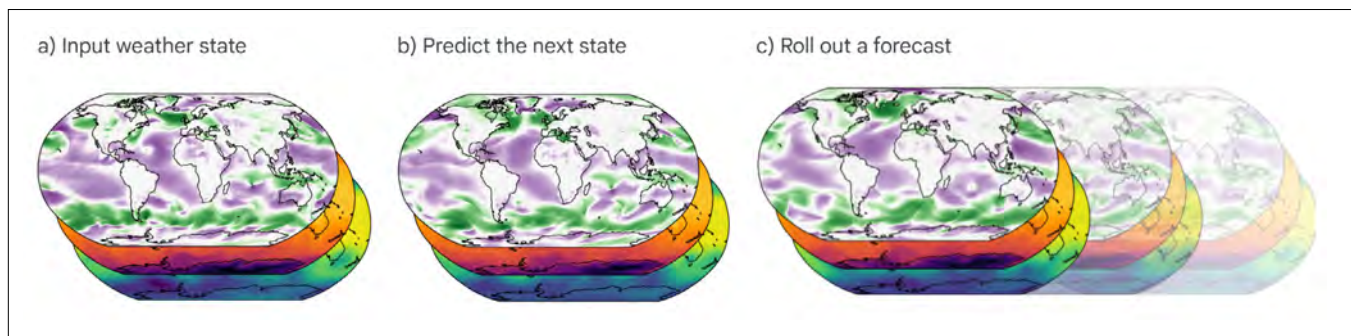


Figure 5.1.6

10日間のz500予報技能：GraphCast対HRES

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

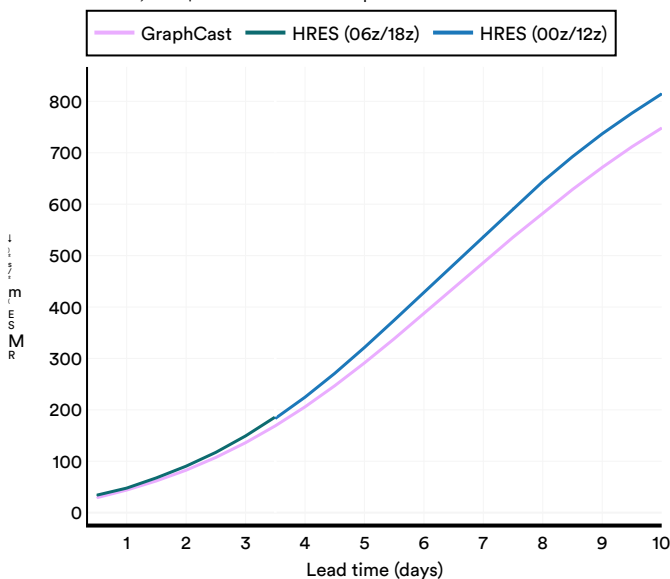


Figure 5.1.7

GNoME

GNoMEによる新素材の発見

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おいて重要です。しかし、この発見プロセスは通常、高価で時間がかかります。Googleの研究者による最近の進展は、グラフネットワークというタイプのAIモデルが、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ました。彼らのモデルである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、その多くは人間の研究者によって見落とされていました（図5.1.9および図5.1.10）。GNoMEのようなAI駆動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上で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

Source: Merchant et al., 2023

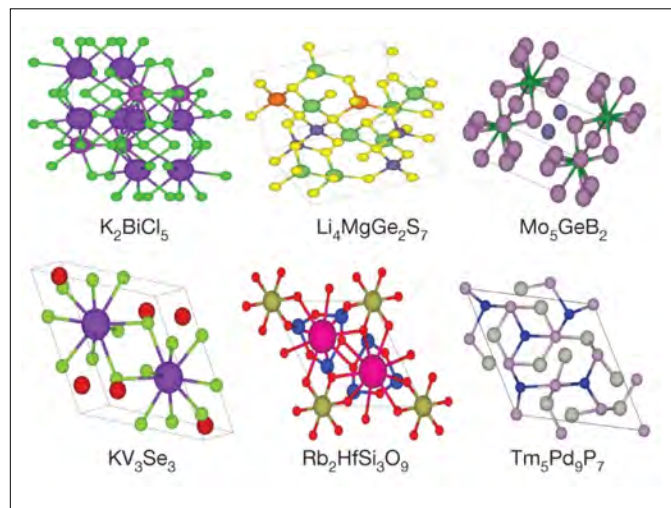


図5.1.8

GNoME vs. Materials Project: 安定な結晶数

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

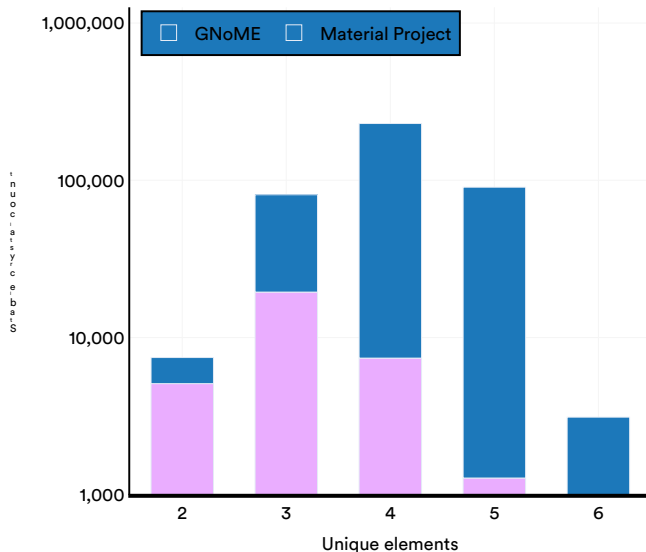


図5.1.9

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

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

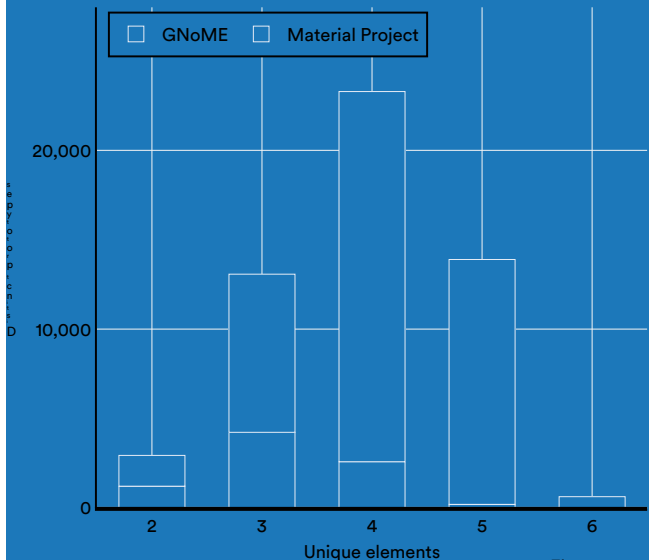


Figure 5.1.10

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大き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、特に予防や緩和のためのインフラが不足している発展途上国で壊滅的な影響を及ぼします。したがって、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、{v*}に大きなプラスの影響をもたらす可能性があります。

Googleの研究チームは、AIを用いて非常に高精度な水文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。これらの革新的な手法は、特定の極端な洪水イベントを最大5日前まで予測でき、その精度はGloFASのような現行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、さまざまな再現期間のイベントにおいて、陽性予測の精度（正確さ）と再現率（関連するすべての事例を正しく識別する能力）で優れており、主要な現代的手法を凌駕しています（図5.1.11）²。このモデルはオープンソースであり、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のリターン期間ごとの予測比較

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

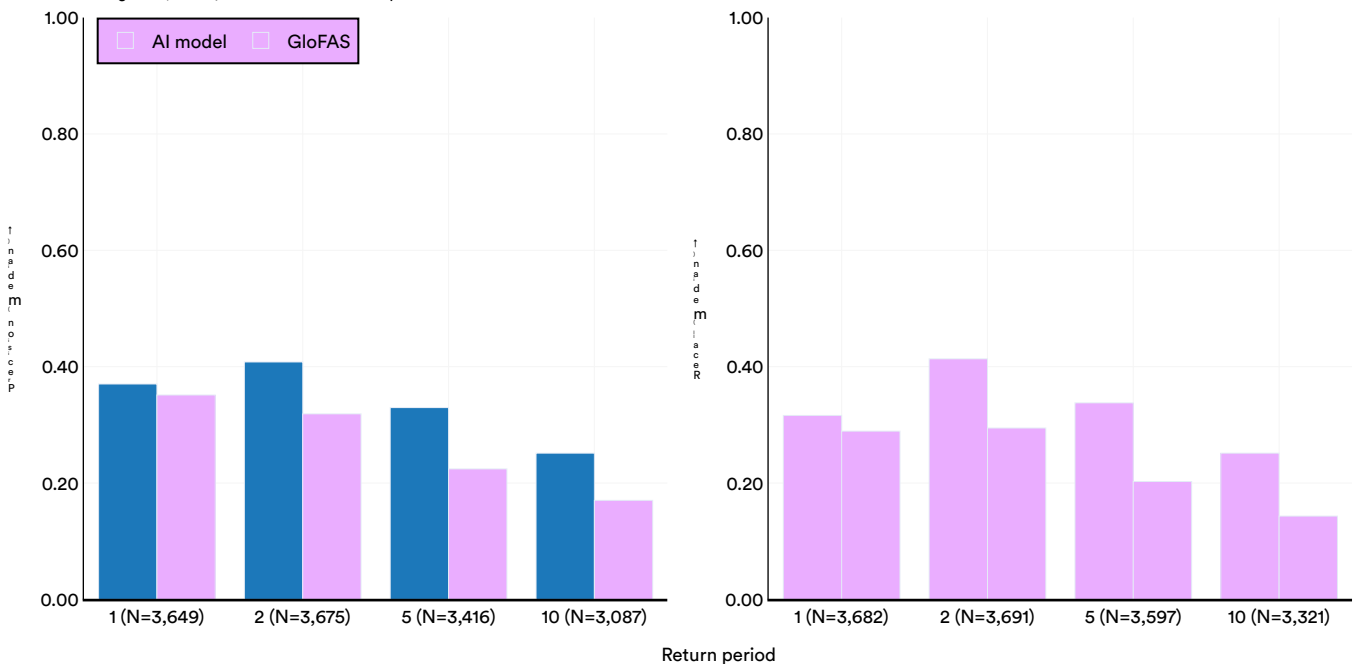


図5.1.11

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価値を高めており、ポリープの検出から臨床医の診断支援まで幅広く応用されています。AIの性能が向上し続ける中、その医療現場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。本節では、2023年に導入された重要なAI関連医療システム、臨床AIの現状、そして病院運営の向上を目指した新しいAI診断ツールおよびモデルの開発について紹介します。

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によって選ばれた2023年の重要なAI関連の医療のブレークスルーを特定します。

SynthSR

高度な解析のための脳スキャン変換 SynthSRは臨床用脳スキャンを高解像度のT1強調画像に変換するAIツールです(図5.2.1)。この進歩は、これまで多くのスキャンの高度な研究利用を制限していたスキャン品質のばらつき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脳構造が明瞭に描出されることで知られる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を用いた実験では、スキャンレベルおよび被験者レベルの観察された体積間に強固な相関が示されており、SynthSRが高解像度T1スキャンで生成される画像に非常に近い画像を生成していることが示唆されています。図5.2.2は、選択された脳領域においてSynthSRスキャンが真の観察結果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、神経科学研究および臨床診断を促進します。

SynthSR ジェネレーション

Source: Iglesias et al., 2023

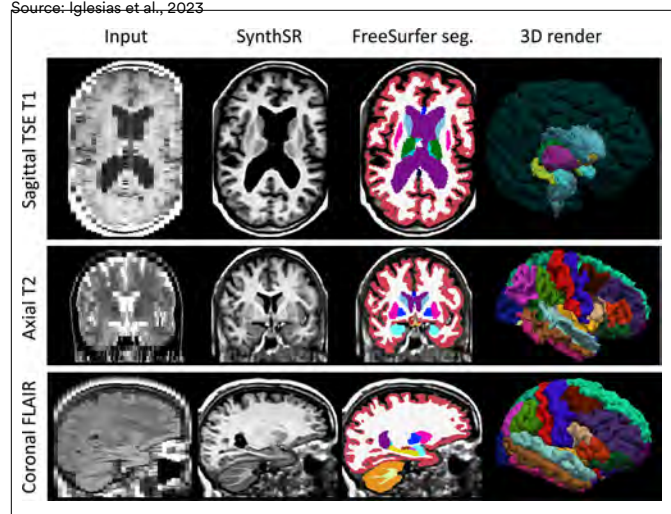


Figure 5.2.1



選択された脳領域における実測体積とのSynthSR相関

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

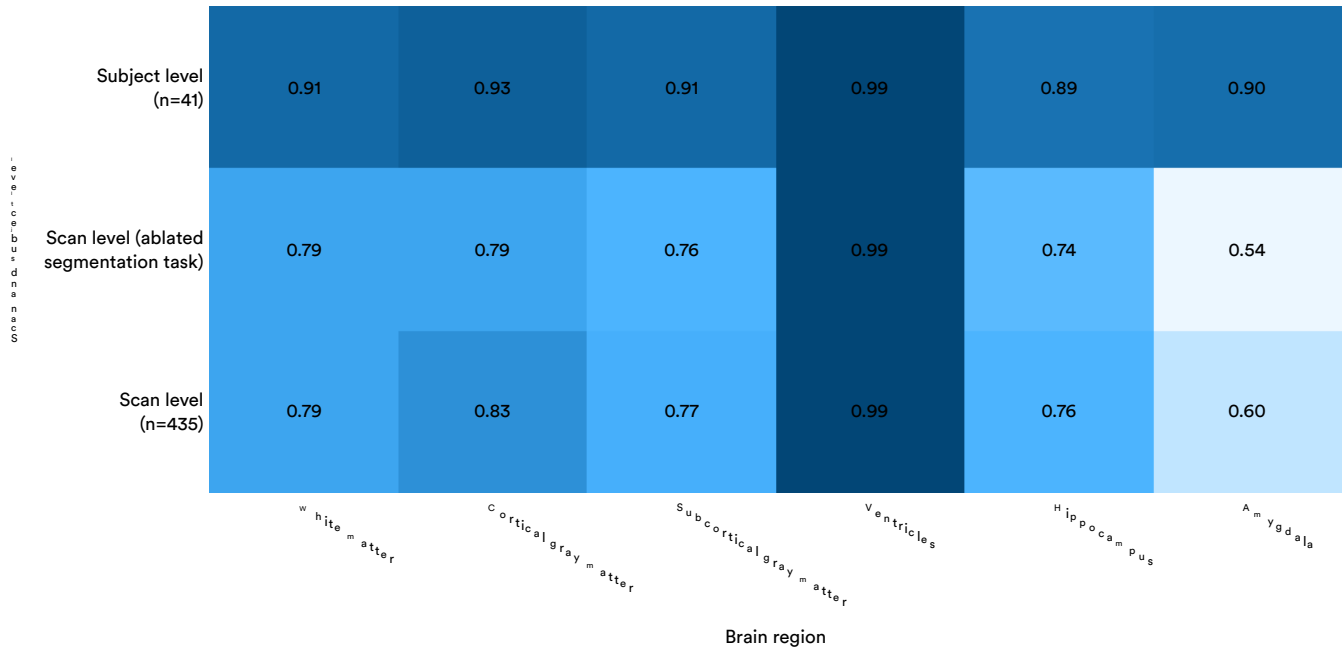


Figure 5.2.2

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点で有用ですが、構造状態の変化を識別することはできません。今年、研究者たちは、表面増強赤外吸収（SEIRA）分光法を用いたAI連携プラズモン赤外線センサーと免疫測定技術を組み合わせた新しい神経変性疾患診断法（ImmunoSEIRA；図5.2.3）を発見しました。実際のフィブリル割合とAIシステムによる予測を比較したテストでは、予測の精度が実際の報告割合と非常に近いことが確認されました（図5.2.4）。

ImmunoSEIRA検出原理とセットアップ

Source: Kavungal et al., 2023

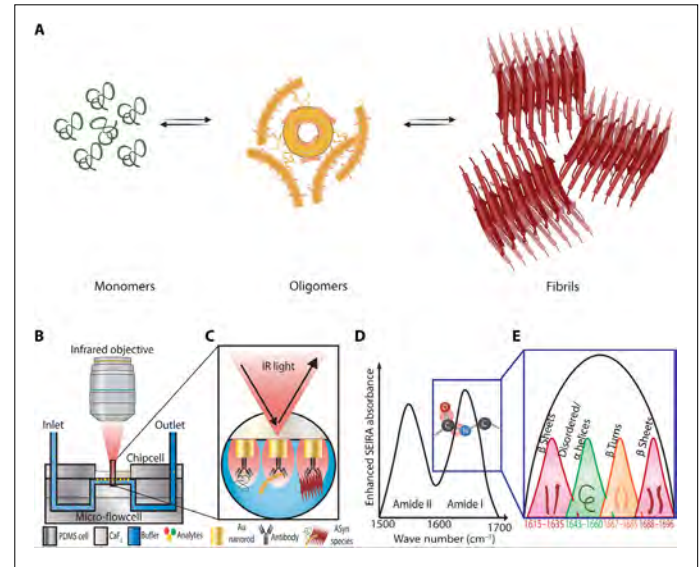


図5.2.3

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と実際のbrilsパーセンテージ

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

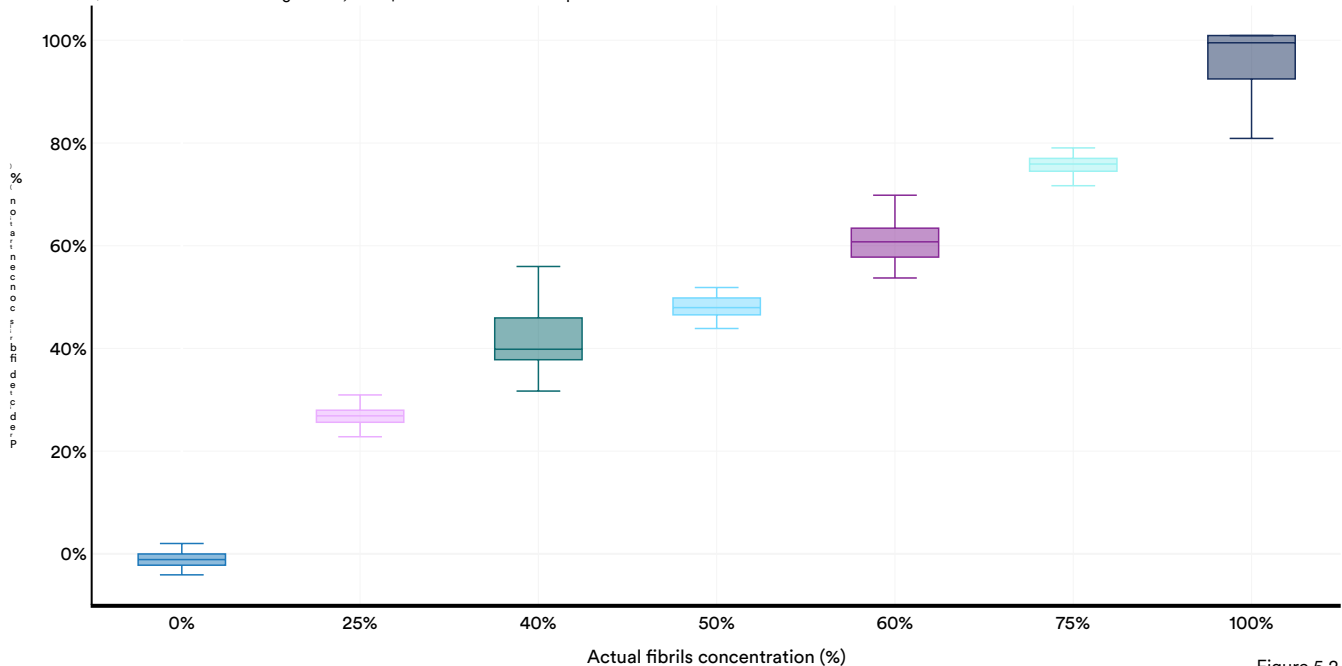


Figure 5.2.4

EVEscape

パンデミック準備のためのウイルス進化予測

ウイルス変異の予測は、ワクチン設計およびパンデミックの最小化に不可欠です。リアルタイムのウイルス株および抗体データに依存する従来の方法は、データ不足のためパンデミック初期段階で課題に直面します。EVEscapeは、過去の配列と生物物理学のおよび構造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^{v*}

ウイルス（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の変異の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%、さらに以前のモデルが予測したわずか24%の変異（図5.2.6）を上回っています。この性能は、将来のパンデミック準備および対応努力を強化するための貴重な資産としてのEVEscapeの可能性を示しています。

EVEscapeデザイン

Source: Thadani et al., 2023

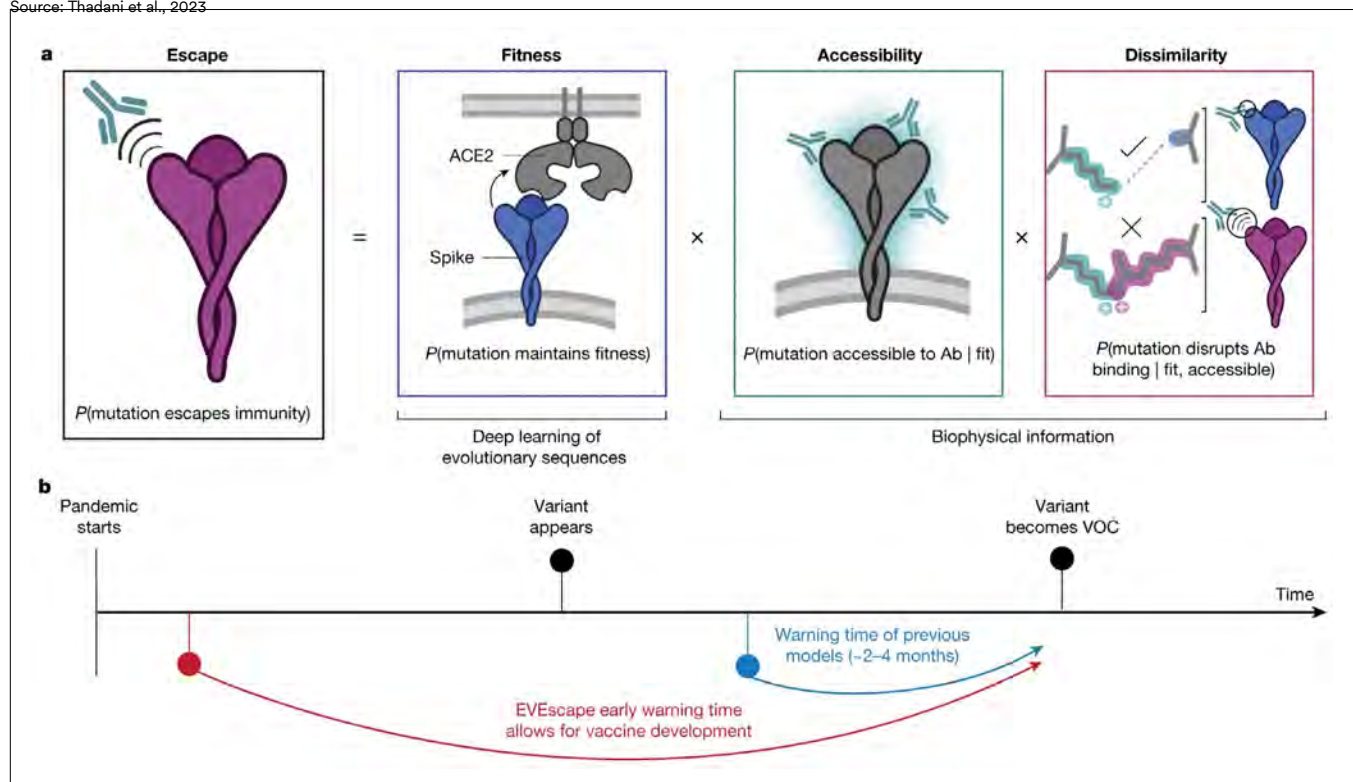


図5.2.5

EVEscape と他のモデルの SARS-CoV-2 RBD 突然変異予測における比較

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

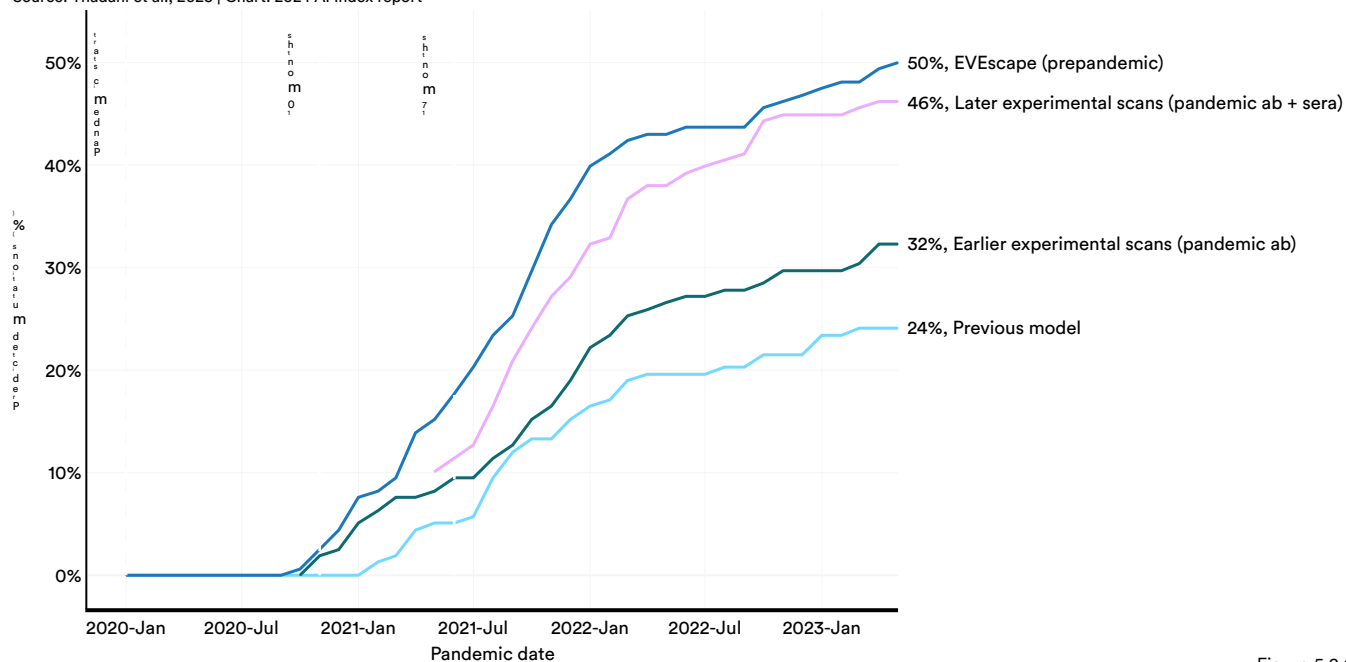


Figure 5.2.6

アルファミッセンス 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を引き起こすの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の中で、変異が良性か病原性かを判断するには、多大な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7,100万のミスセンス変異の病原性を予測しました。ミスセンス変異はヒトタンパク質の機能に影響を与える遺伝的变化であり（図5.2.7）、がんを含むさまざまな疾患を引き起こす可能性があります。7,100万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%を良性の可能性が高いと、32%を病原性の可能性が高いと特定し、残りは不確定として分類しました（図5.2.8）。対照的に、人間のアノテーターはすべてのミスセンス変異のうちわずか0.1%の性質しか確認できていません。

ヘモグロビンβサブユニット (HB B) Source: Google DeepMind, 2023

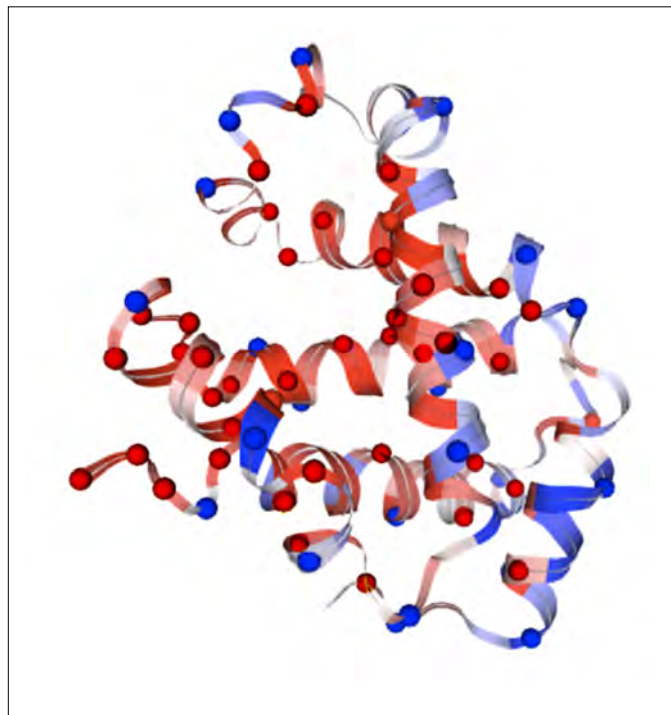


図5.2.7

AlphaMissense予測

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

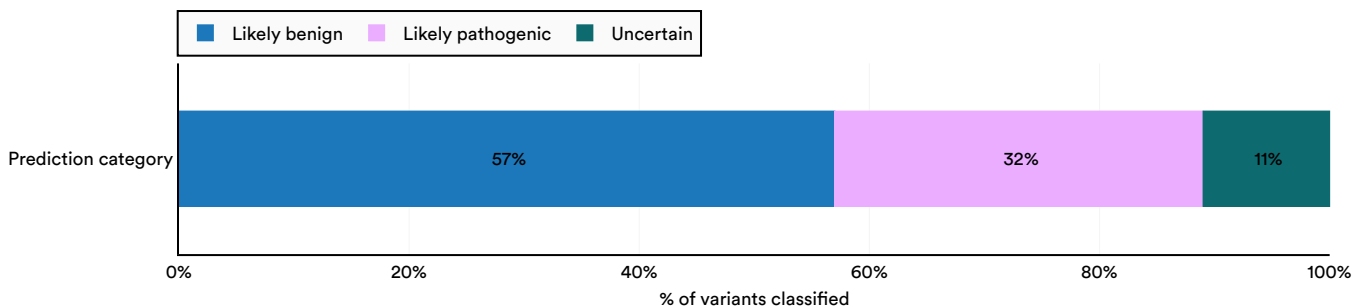


Figure 5.2.8

ヒトパンゲノムリファレンス

AIを使っ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令のセットです。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、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的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、既存のゲノムリファレンスの下では、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年に、60の機関から119人の科学者で構成されるヒトパンゲノム研究コンソーシアムは、AIを用いて更新され、より代表的なヒトゲノムマップを作成しました（図5.2.9）。研究者たちは、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5.2.10参照）。

ゲノムのMHC領域のグラフゲノム

Source: Google Research, 2023

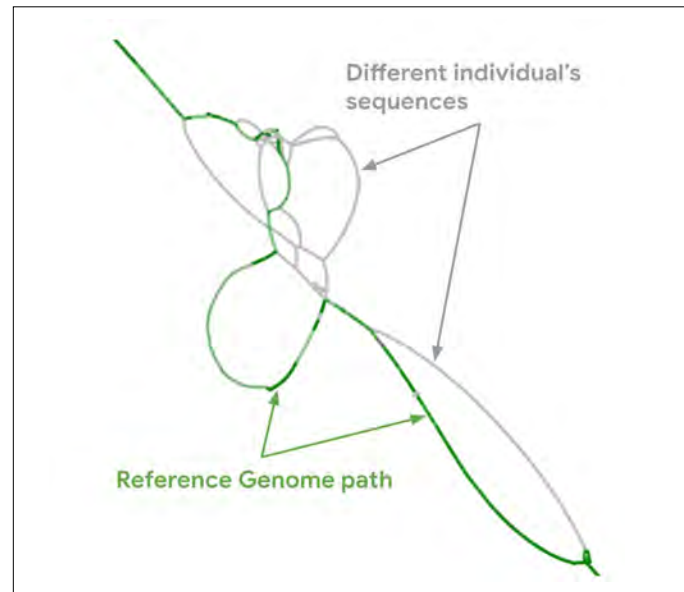


図5.2.9

この最新バージョンのゲノムは、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: Lij et al., 2023 | Chapter 5.2 AI Index report

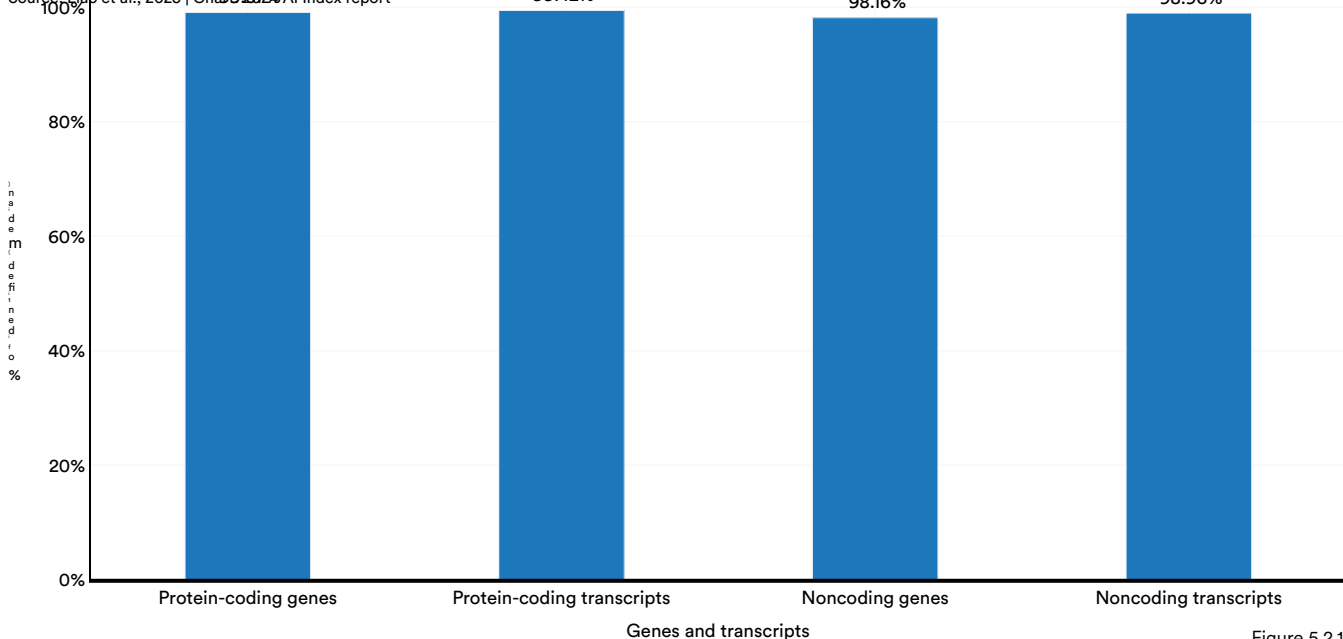


Figure 5.2.10

臨床知識

AIモデルの臨床知識を評価することは、特に臨床現場で適用可能な知識において、彼らの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、最先端のシステムであるGPT-4 Medpromptは90.2%の正確率を達成しました。これは2022年の最高スコアから22.6ポイントの増加です（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、臨床知識を持つAIシステムの急速な進歩を示しています。

MedQA

2020年に導入されたMedQAは、専門の医療資格試験から得られた包括的なデータセットであり、医師を挑戦させ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。

MedQAベンチマークにおけるAIの性能は{v*}を記録しています

MedQA: 正確さ

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

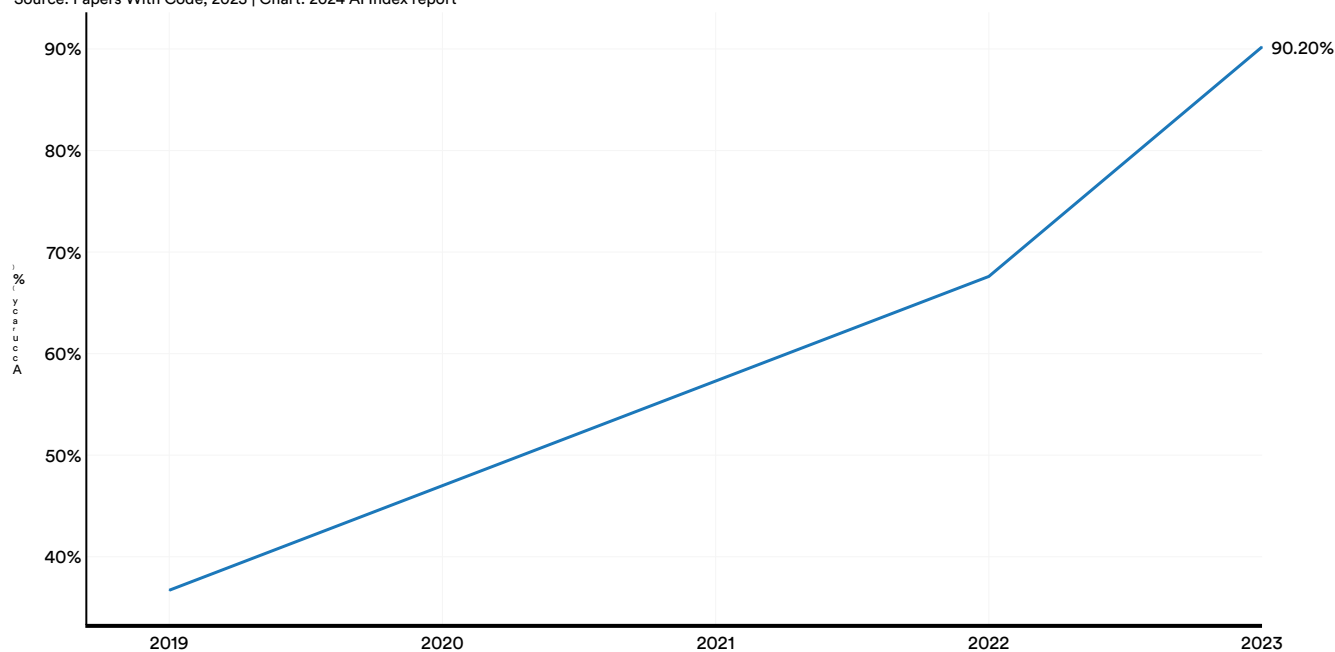


図5.2.11

注目の研究：

GPT-4 メドプロンプト

LLMは優れた一般知識を示しますが、医療質問への回答など専門的な知識で優れるためには大幅なファインチューニングが必要であると一般に考えられています。ファインチューニングとは、ドメイン固有のデータでLLMを訓練することを指します。

さらに、前述のように、GPT-4 MedpromptはMedQAベンチマークで90%の精度を初めて超えました。この画期的な成果は、GPT-4 Medpromptの卓越した、かつ臨床的に有用である可能性のある医療能力を強調するだけでなく、専門分野にモデルを適応させるために必ずしもファインチューニングが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であることが示されています。


2023年末の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MultiMedQAベンチマークスイート（4つの難易度の高い医療ベンチマークのグループ、図5.2.12）で卓越した性能を発揮するように導きました。GPT-4 Medpromptは、PubMedQA、MedMCQA、MMLUなどの著名な医療ベンチマークの複数選択セクションで、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パーセンテージポイント上回りました。また、当時の最先端モデルであったMed-PaLM 2の性能も上回りました（図5.2.13）。

GPT-4 対 Med-PaLM 2 が医療質問に回答 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

注目の研究：

GPT-4 メドプロンプト（続き）

MultiMedQAサブベンチマークにおけるモデル性能

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

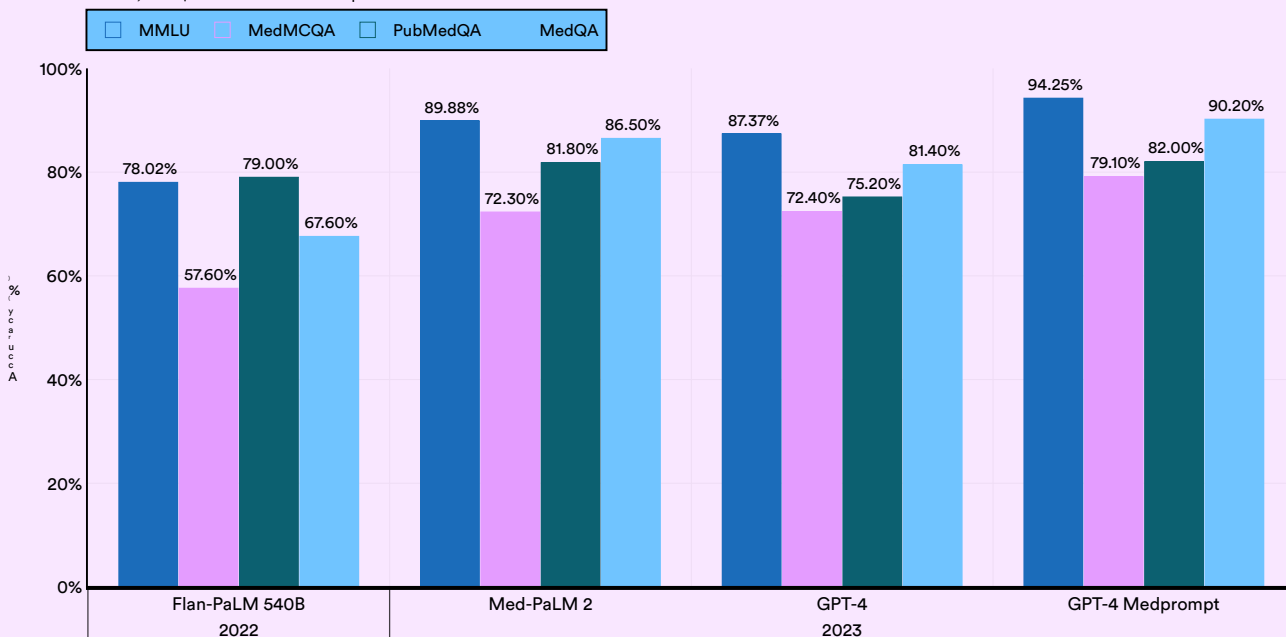


図5.2.13

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り、その重みは一般に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の能力向上も目指されています。この新しい研究の中で、MediTron-70Bは特に有望なモデル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%の優れた精度を達成しています。これはGPT-4 MedpromptおよびMed-の性能には及びませんが、{v*}

PaLM 2（両方ともクローズドモデル）は、2023年の最先端の結果を大幅に上回り、Llama 2のような他のオープンソースモデルを凌駕しています（図5.2.14）。MediTron-70BのMedQAでのスコアは、オープンソースモデルとしてはこれまでで最高です。医療AIがその最大の可能性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は励みとなる一歩を示しています。

MedQAにおける選択モデルの性能

出典：Chenら、2023年 | 表：2024年AIインデックスレポート

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、診断目的にも使用できます。

注目の研究：

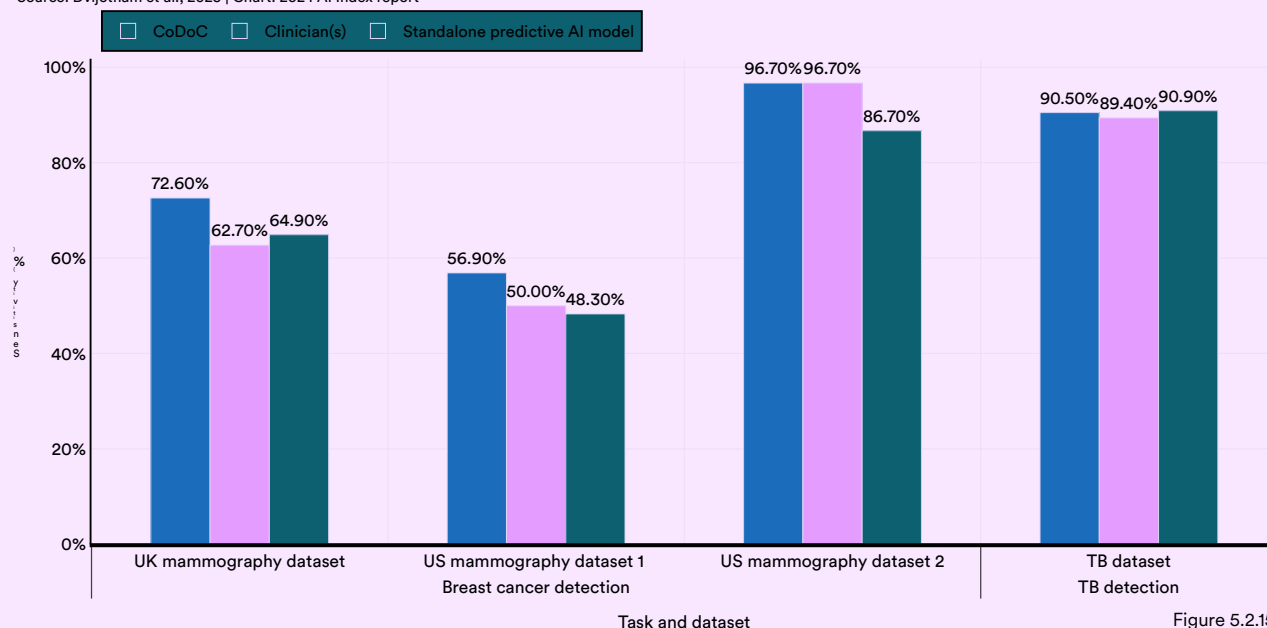
CoDoC

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、臨床医が見逃さない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の論理的な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、診断にAIを利用すべき時と従来の臨床手法に委ねるべき時を識別するために設計されたシステムです。CoDoCは特に感度（疾患を正しく識別する能力）と特異度を向上させます。

（それを持たない者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、単独の予測モデルを5.7パーセントポイント上回っています。さらに、CoDoCは臨床ワークフローを66%削減することが示されています。これらの結果は、AI医療システム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可能性を示唆しています。

CoDoC 対 独立型予測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



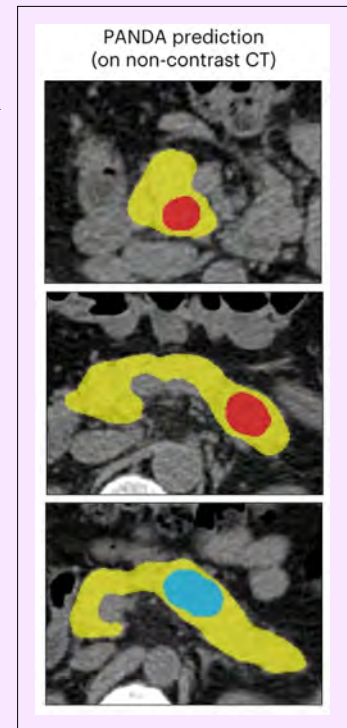
注目の研究：

CTパンダ

膵臓癌（PDAC）は特に致命的な癌であり、しばしば手術介入が困難なほど遅れて発見されます。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す。今年、中国の研究チームはPANDA（人工知能による膵臓癌検出）を開発しました。これはX線画像で膵臓病変を効率的に検出・分類できるAIモデルです（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%上回りました（図5.2.17）。約20,000人の患者を対象とした大規模な実地試験では、PANDAは感度92.9%、特異度99.9%を達成しました（図5.2.18）。PANDAのようなAI医療ツールは、これまで困難または実現困難とされてきた疾患の診断において、費用対効果が高く正確な検出を可能にする重要な進歩を示しています。

PANDA 検出

Source:
Cao et al., 2023
図5.2.16



出典：Caoら、2023年 | チャート：2024年AIインデックスレポート PANDA対平均放射線科医による多施設検証（6,239人の患者）

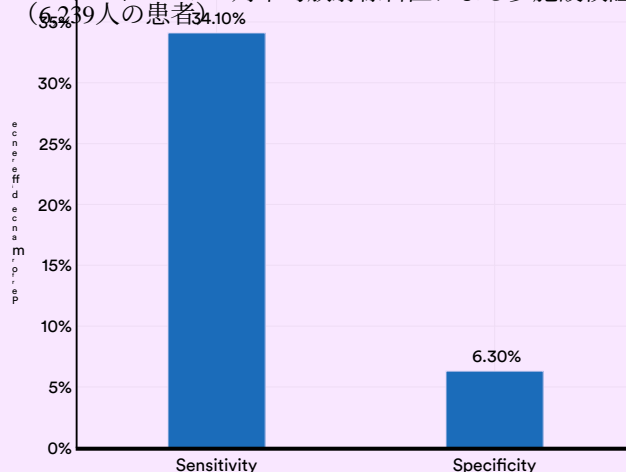


Figure 5.2.17

実世界のマルチシナリオ検証におけるPANDAの性能（20,530人の患者）

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

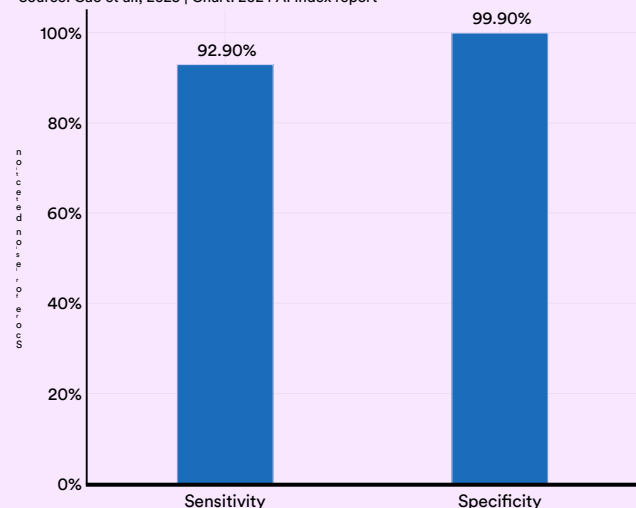


Figure 5.2.18

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、いくつかの発見をまとめたものです。

診断AIのユースケースに関する追加研究

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

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管理しています。このリストに掲載されている機器は、効果と安全性の詳細な審査を含むFDAの市販前基準を満たしています。2023年10月時点で、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件のAI関連医療機器がFDAの承認を受け、2021年の承認数から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

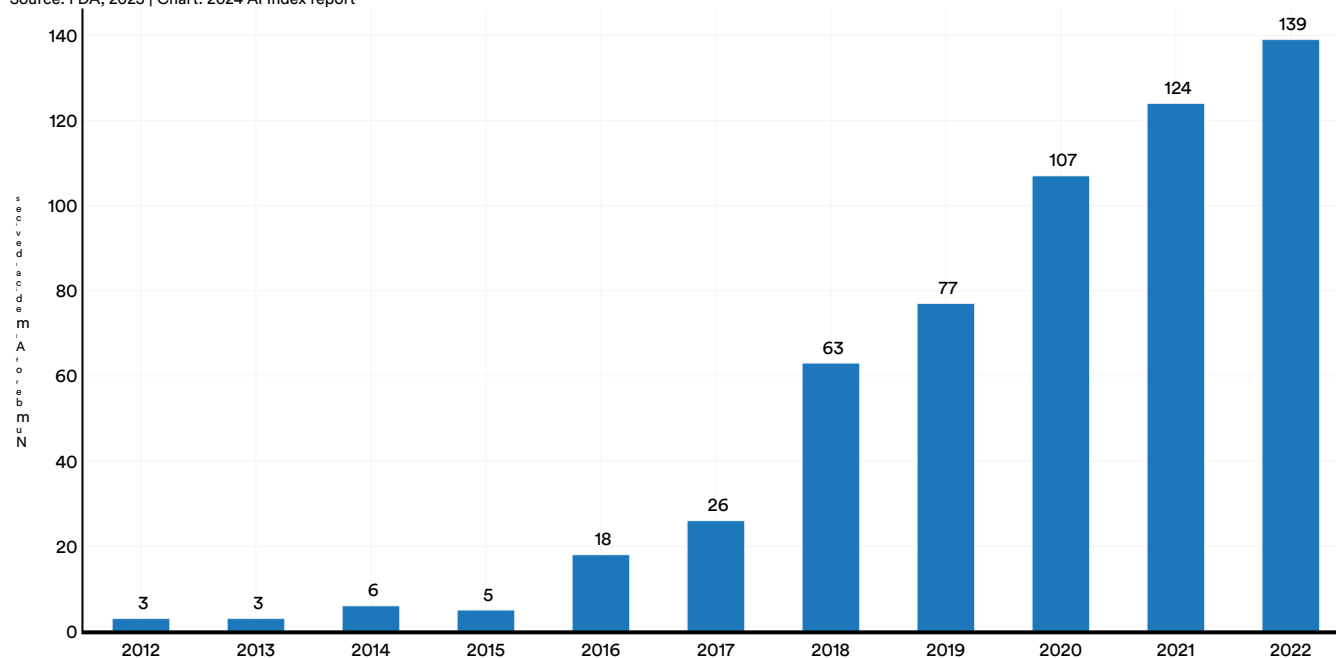


図5.2.20

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る。2022年に承認された139台の機器のうち、87.1%が放射線科に関連していた。次に多い専門分野は心血管で、承認の7.2%を占めている。

2012年から2022年までの専門分野別FDA承認済みAI医療機器の数

出典：FDA、2023年 | 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
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高め、患者ケアの水準を向上させる可能性も秘めています。

再。

注目の研究：

メッドアライン

医療分野におけるAIの大き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。臨床医はしばしば、個別化された診断計画の作成など情報集約型のタスクに従事し、勤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。医療管理用の大規模言語モデル（LLM）のベンチマーク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埋めるために、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、および7つの異なる医療専門分野から抽出された303の臨床医の回答を含むベンチマーク（図5.2.22）。MedAlignは初の大規模なEHRに特化したベンチマークである。

研究者たちは次に、MedAlign上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、マルチステップリファインメントを使用したGPT-4の変種が最も高い正確率（65.0%）を達成し、他のLLMよりも常に好まれました（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を活用する上で貴重なマイルストーンです。

MedAlign ワークフロー

Source: Fleming et al., 2023

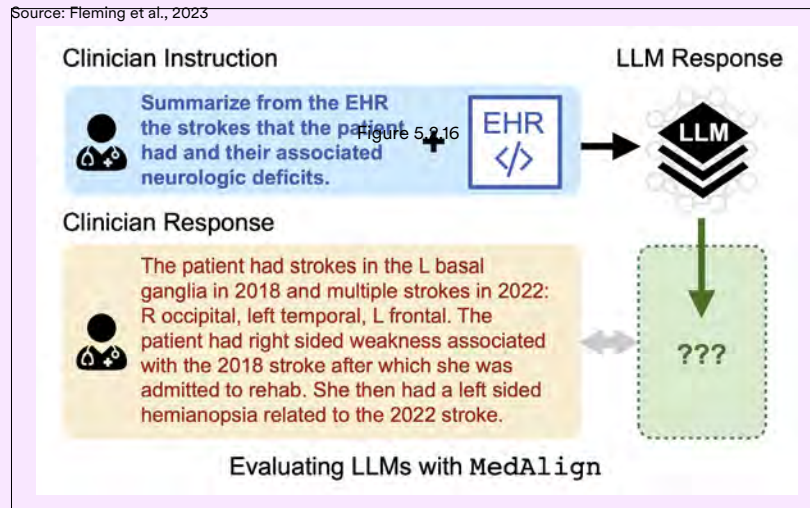


図5.2.22

注目の研究：

MedAlign（続き）

モデル性能の評価：人間対COMETランク

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

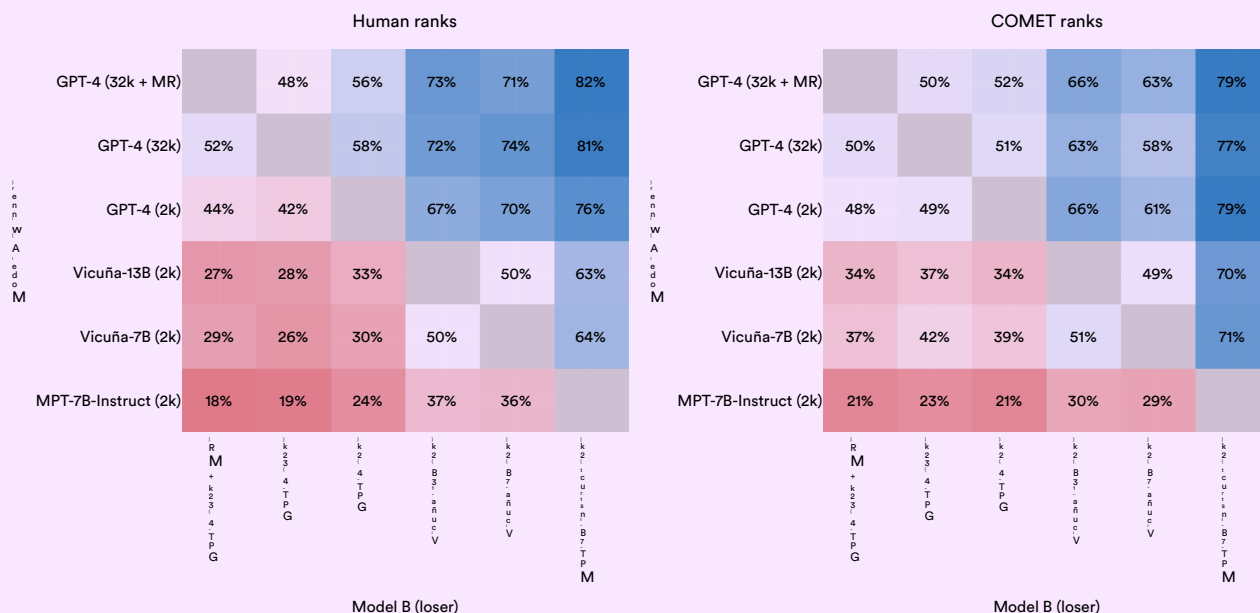
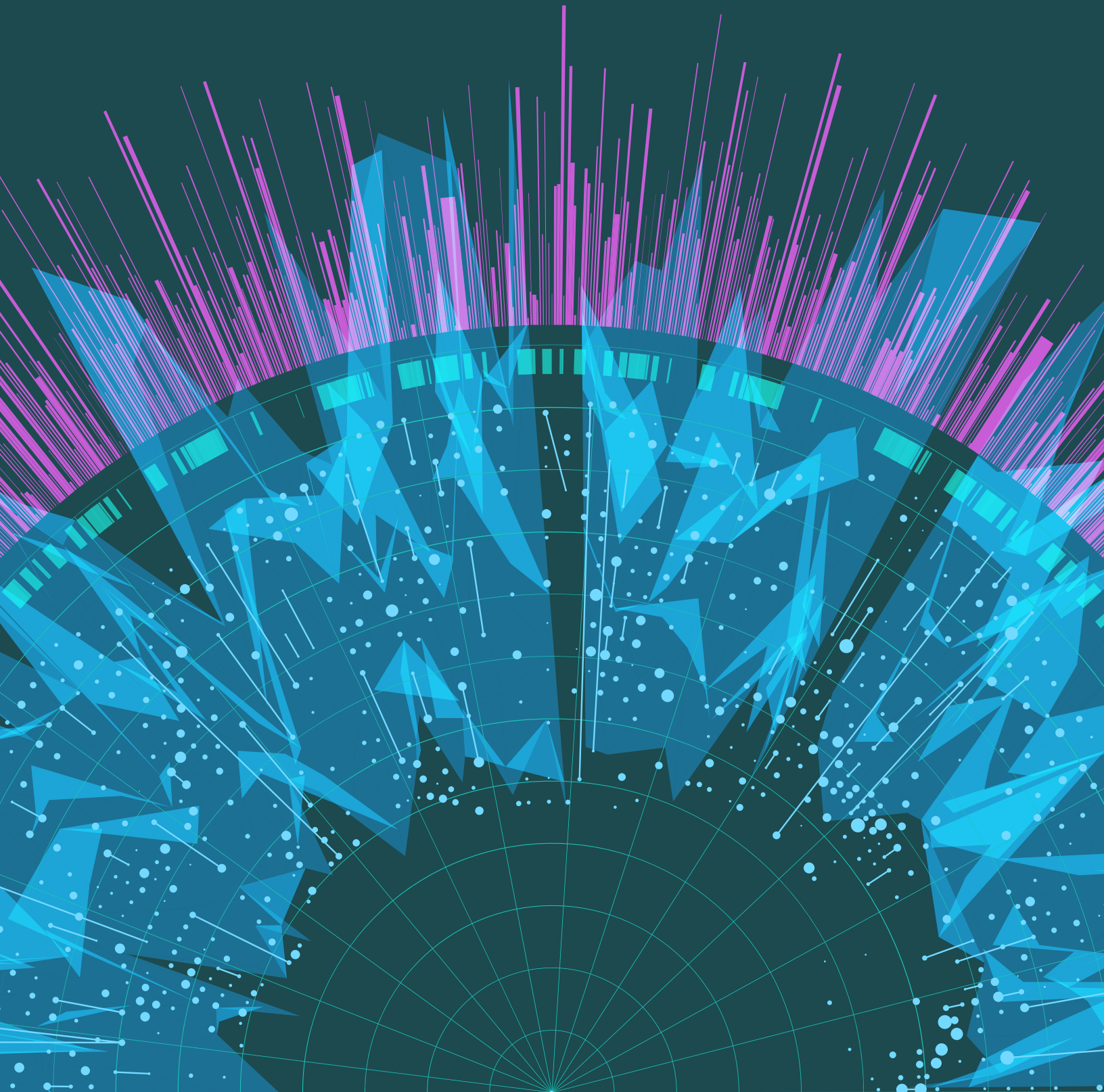


図5.2.23



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およびAI教育 359 アメリ
リカ合衆国 359 州レベルの動向 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：教員および学生のChatGPT利用状況 364

公共データにアクセスする

概要

この章では、AIおよびコンピュータサイエンス（CS）教育の動向を検討し、誰が学んでいるのか、どこで学んでいるのか、そしてこれらの動向が時間とともにどのように変化してきたかに焦点を当てます。AIが教育に与える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ツールの使用についても調査します。

分析は、Computing Research Associationの年次Taulbee調査に基づく米国およびカナダの高等教育におけるCSおよびAI教育の現状の概要から始まります。次に、Informatics EuropeのヨーロッパにおけるCS教育に関するデータをレビューします。今年は、StudyportalsからのAI関連の英語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が導入されます。

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、ウォルトン財団の学校におけるChatGPTの利用に関する調査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コンピュータサイエンス学士卒業者数は増加し続けており、新しいコンピュータサイエンス修士卒業者数は比較的横ばいで、博士卒業者数は緩やかに増加しています。一方で、アメリカとカナダの新しい学士卒業生の数は10年以上にわたり一貫して増加しているが、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の修士および博士卒業生の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ます。2011年には、新たに博士号を取得したAI研究者のほぼ同じ割合が産業界（40.9%）と学術界（41.6%）で就職していました。しかし、2022年には、卒業後に産業界に進む割合が大幅に増加し（70.7%）、学術界に進む割合（20.0%）を大きく上回りました。過去1年間だけでも、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人材流出が加速していることを示しています。
3. 産業界から学界への学術人材の移行が減少している。2019年には、米国およびカナダの新しいAI教員の13%が産業界出身であった。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レベルなAI人材の移行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教育は国際性が低下している。2022年には2021年に比べて、国際的なCSの学士号、修士号、博士号の卒業生の割合が減少した。特に修士課程の国際学生の減少が顕著であった。
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ますが、アクセスの問題は依然として残っています。2022年には、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倍以上に増加しています。しかし、最近の証拠は、大規模な高校の生徒や郊外地域の生徒の方がCSコースにアクセスしやすいことを示しています。
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倍に増加しており、過去5年間で着実な年次増加を示しています。世界中の大学がより多くのAIに特化し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、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）分野の卒業生数でリードしています。イギリスとドイツは、情報学、CS、CE、および情報分野の学士、修士、博士の新卒者数でヨーロッパをリードしています。人口あたりで見ると、フィンランドは学士および博士の卒業生数でトップであり、アイルランドは修士の卒業生数でトップです。

このセクションでは、CSおよびAIの高等教育の概要を提供し、北米およびヨーロッパにおける学士号、修士号、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、米国およびカナダの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示す。この調査は、米国およびカナダの297の博士号授与CS学科を対象としている。¹

CS学士卒業生

過去10年間で、北米の新しいCS学士卒業生の総数は着実に増加し、3倍以上に増加しました。2021年から2022年にかけては年率7.9%の増加を示しています（図6.1.1）。

米国およびカナダの新しいCS学士卒業生、2010-22

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

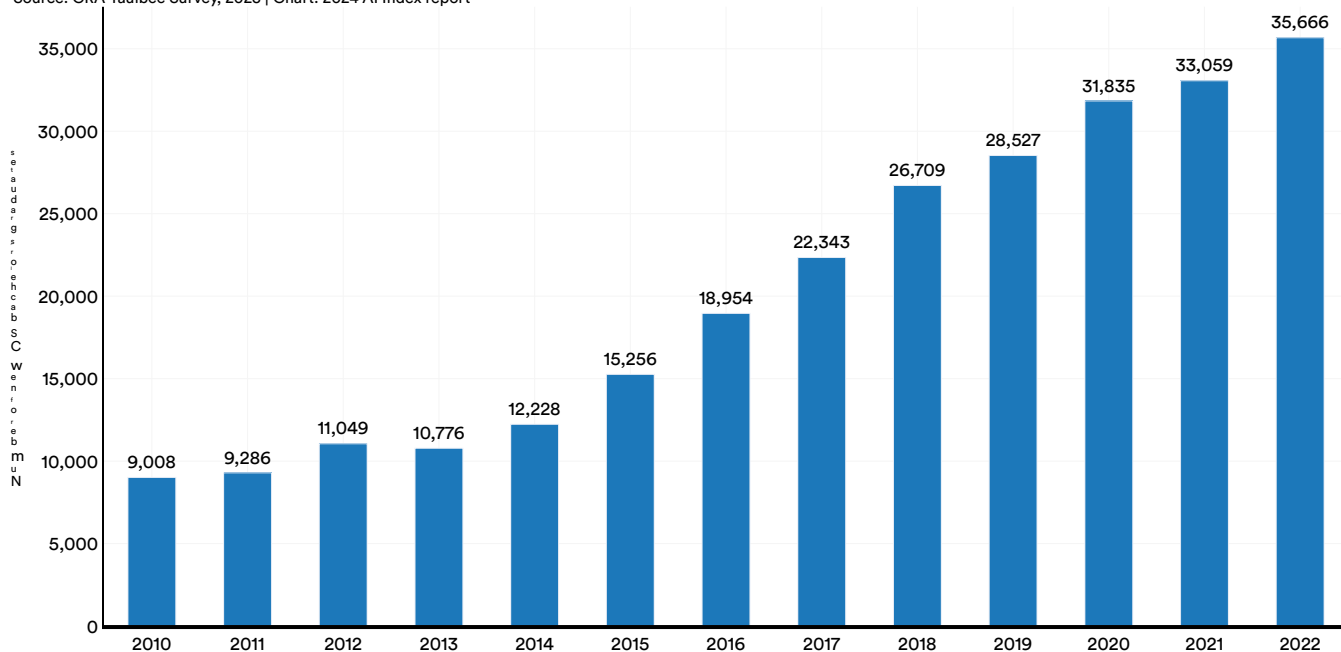


Figure 6.1.1

¹ 調査対象となったすべての博士課程設置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、182学科のみが回答し、全体の回答率は61%となりました。

ほぼ8年ぶりに、アメリカおよびカナダの大学におけるCS学士卒業生の国際学生の割合が減少し、2021年の16.3%から2022年には15.2%に低下した（図6.1.2）。この減少は、トランプ政権初期における留学ビザ取得の難化を反映している可能性が高く、その影響が今ようやく{v*}に現れ始めている。

データ。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にも部分的に起因しており、国際学生がアメリカ合衆国および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。

米国およびカナダにおける新規国際CS学士卒業生（総数に占める割合）、2010–22

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

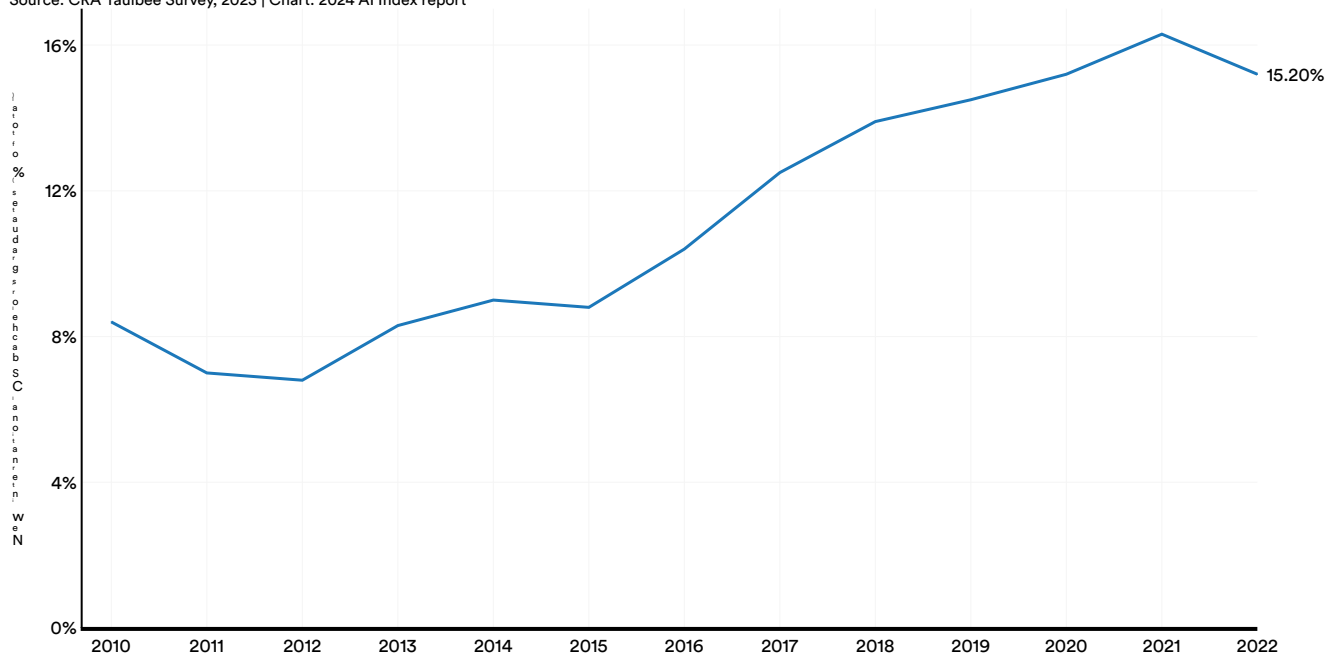


図6.1.2

CS修士卒業生

AIコースは一般的にコンピュータサイエンス修士課程に含まれています。過去10年間でアメリカおよびカナダの大学からの新しいコンピュータサイエンス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、以下のグラフに示されている国際的な修士課程の学生数の減少を反映しています。

2010年から2022年の米国およびカナダの新しいCS修士卒業生

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

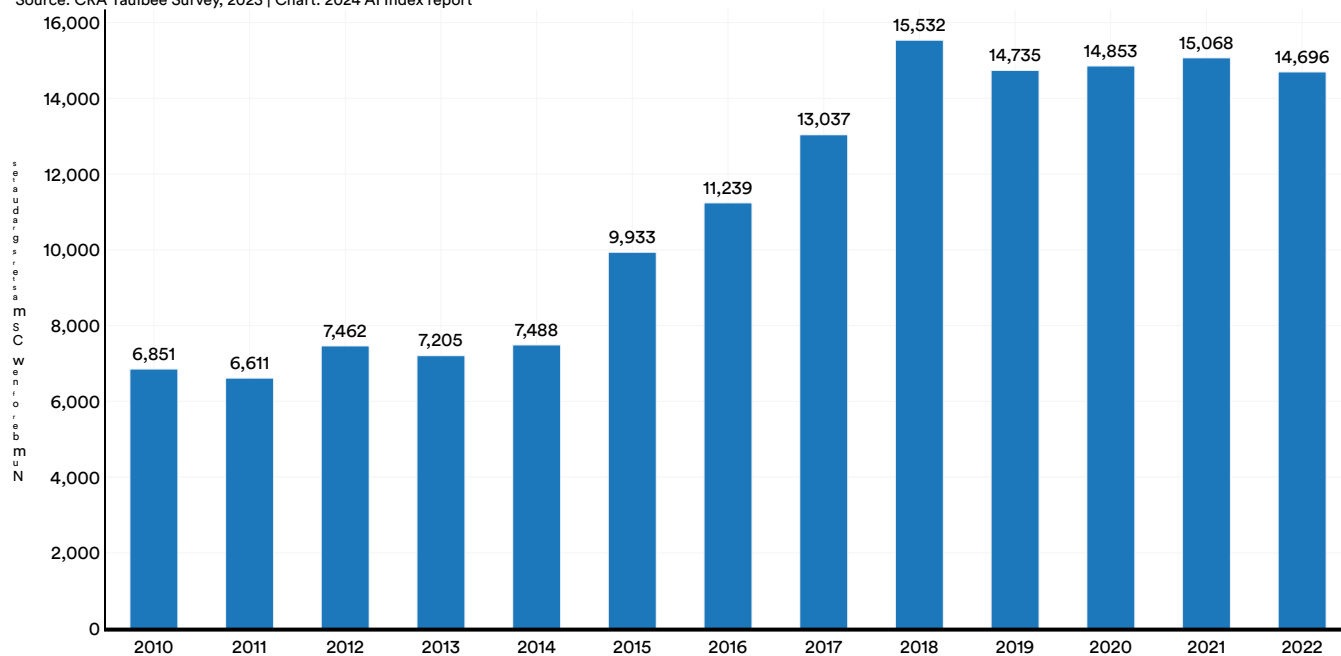


図6.1.3

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。この減少傾向は2017年頃から始まりましたが、昨年が最も顕著で、14.8パーセンテージポイントの減少となりました（図6.1.4）。現在、国際的なCS修士卒業生と国内の卒業生の割合はほぼ同じです。

米国およびカナダにおける新規国際CS修士卒業生（総数に占める割合）、2010-22年

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

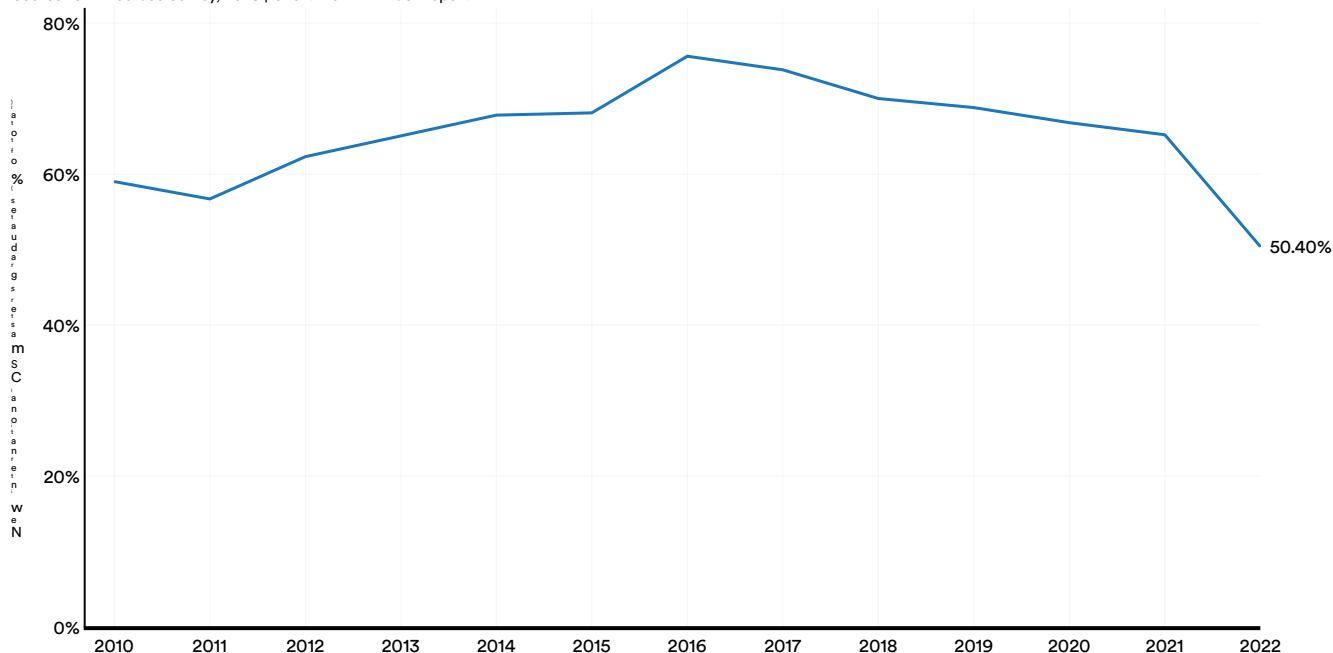


図6.1.4

CS博士課程修了者

10年ぶりに、アメリカとカナダの大学で新たに卒業する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年以来の最高記録となりました（図6.1.5）。

2010年から2022年までの米国およびカナダの新しいCS博士号取得者

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

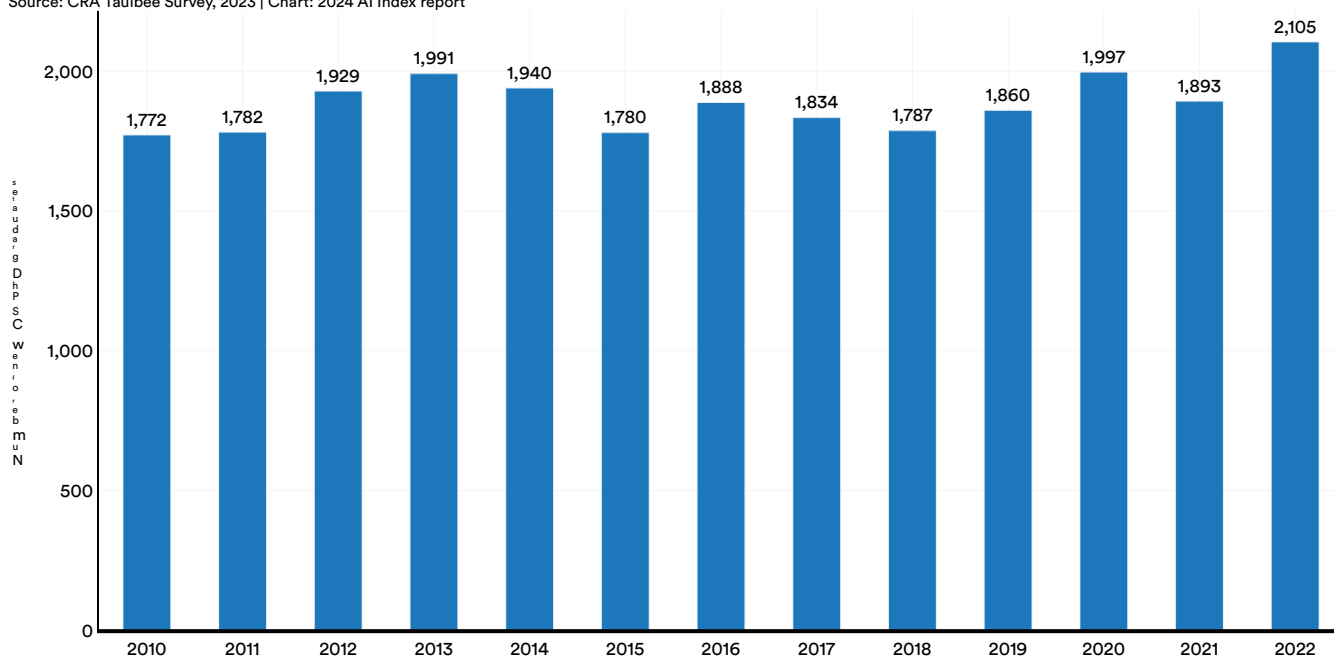


図6.1.5

過去10年間でCS博士課程修了者に占める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年の65.9%に低下しました（図6.1.6）。

米国およびカナダにおける新規国際CS博士号取得者（総数に占める割合）、2010–22年

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

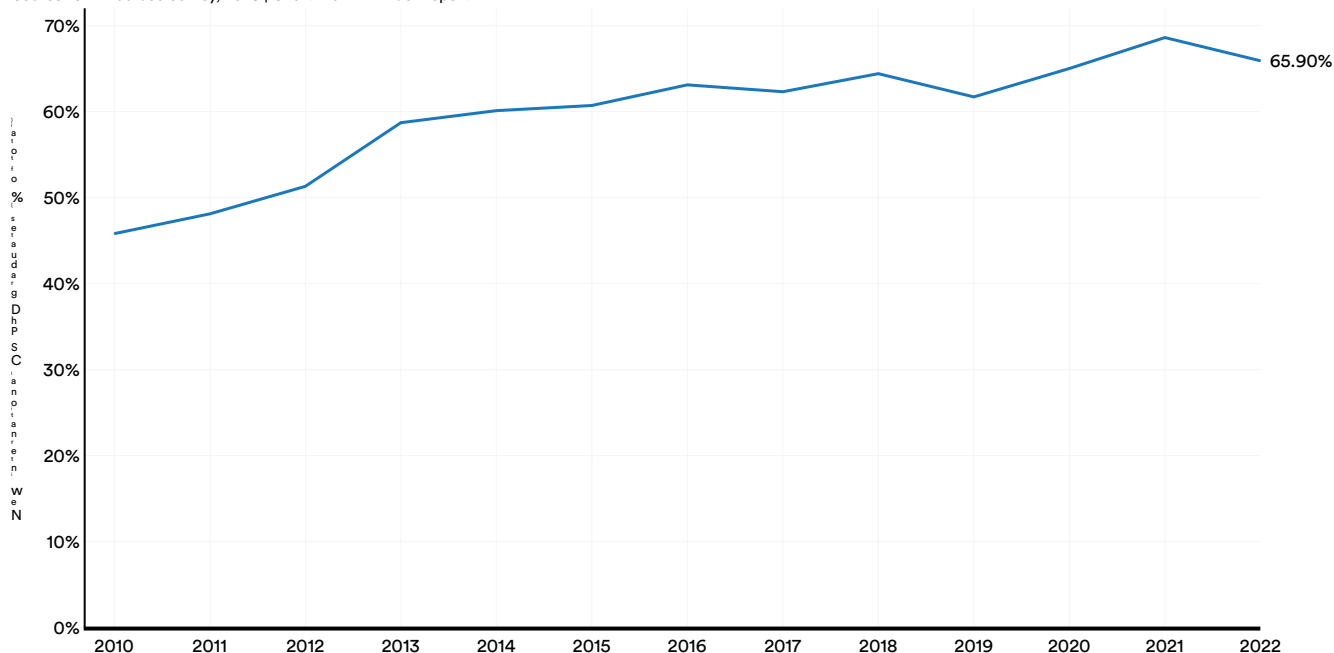
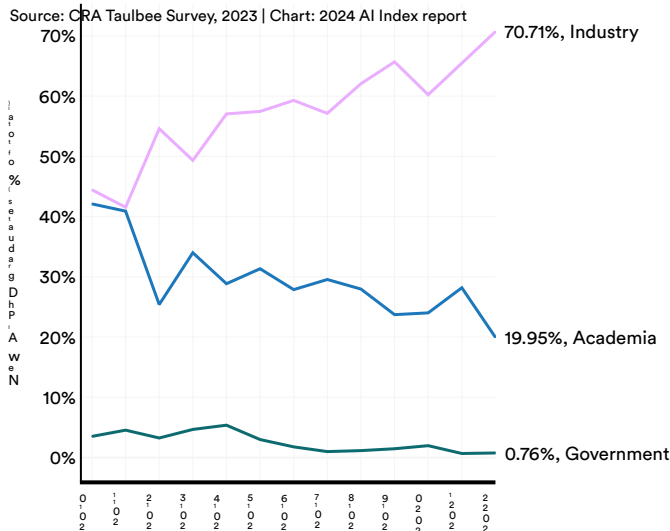


図6.1.6

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インデックス報告書で強調された傾向に従い、AI博士号取得者の増加する割合が産業界でのキャリアを追求しています（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%）がほぼ同じ割合でした。

しかし、2022年までに、卒業後に産業界に進む割合は（70.7%）と、学界に進む割合（20.0%）に比べて大幅に増加しました。新たなAI博士号取得者が政府の役職に就く割合は、過去5年間で約0.7%と比較的に低く安定しています。

2010年から2022年にかけての米国およびカナダにおける新規AI博士号取得者の雇用（総数に占める割合）をセクター別に示す

Figure 6.1.7²

米国における新しいAI博士号取得者の雇用とカナダのセクター別、2010–22

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

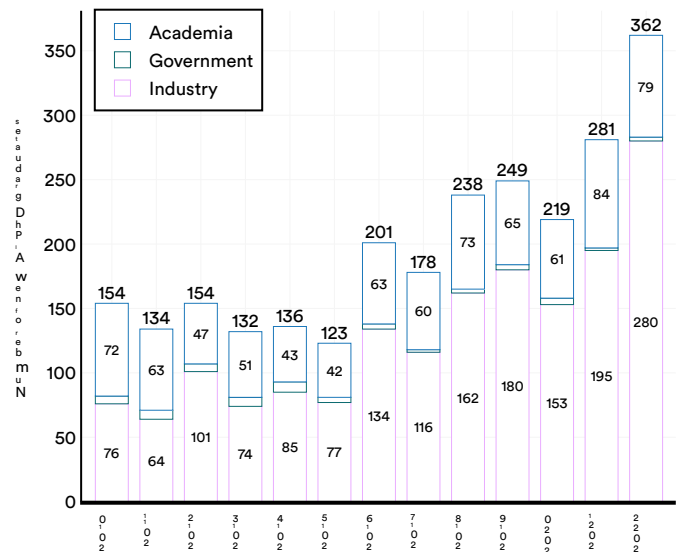


Figure 6.1.8

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.

情報学部、コンピュータサイエンス学科、コンピュータ工学学科

CSおよびAI教育の動向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査することが有益です。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%に達しています。

2011年から2022年までの米国およびカナダのCS、CE、情報学部の人数

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

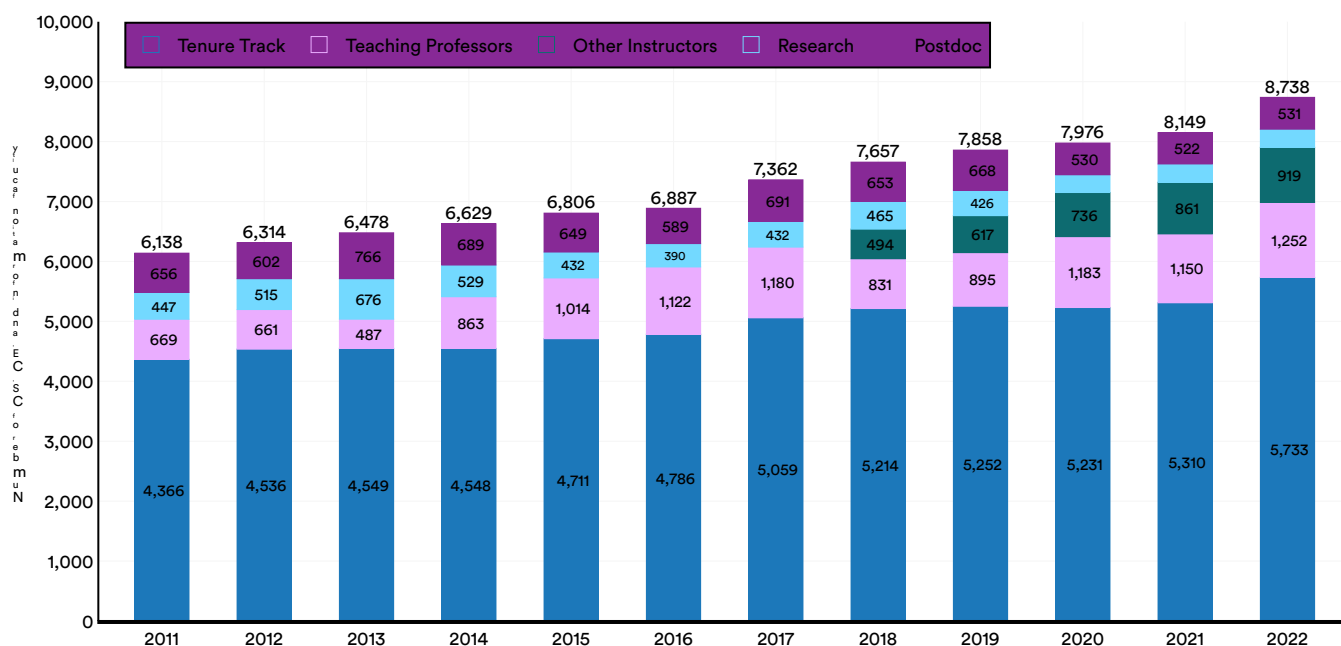


図6.1.9

2022年、アメリカ合衆国には7,084人のコンピュータサイエンス（CS）教員がおり、その大多数（65.7%）がテニュアトラックにありました（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%増加しています。

2011年から2022年までのアメリカ合衆国のコンピュータサイエンス教員数

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

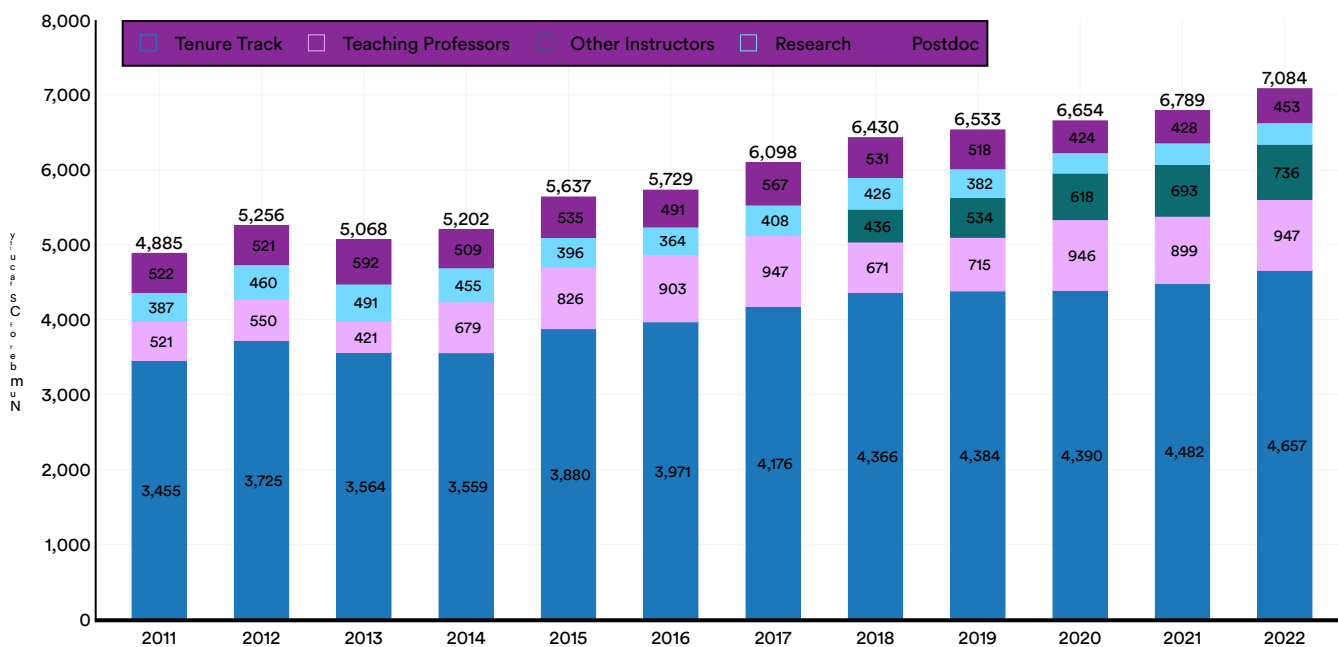


図6.1.10

昨年、北米のCS、CE、および情報分野で915名の新任教員が採用され、10年ぶりの高水準となりました。そのうち455名がテニユアトラックのポジションでした。（図6.1.11）。

2011年から2022年までの米国およびカナダにおける新規CS、CE、情報学部教員採用

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

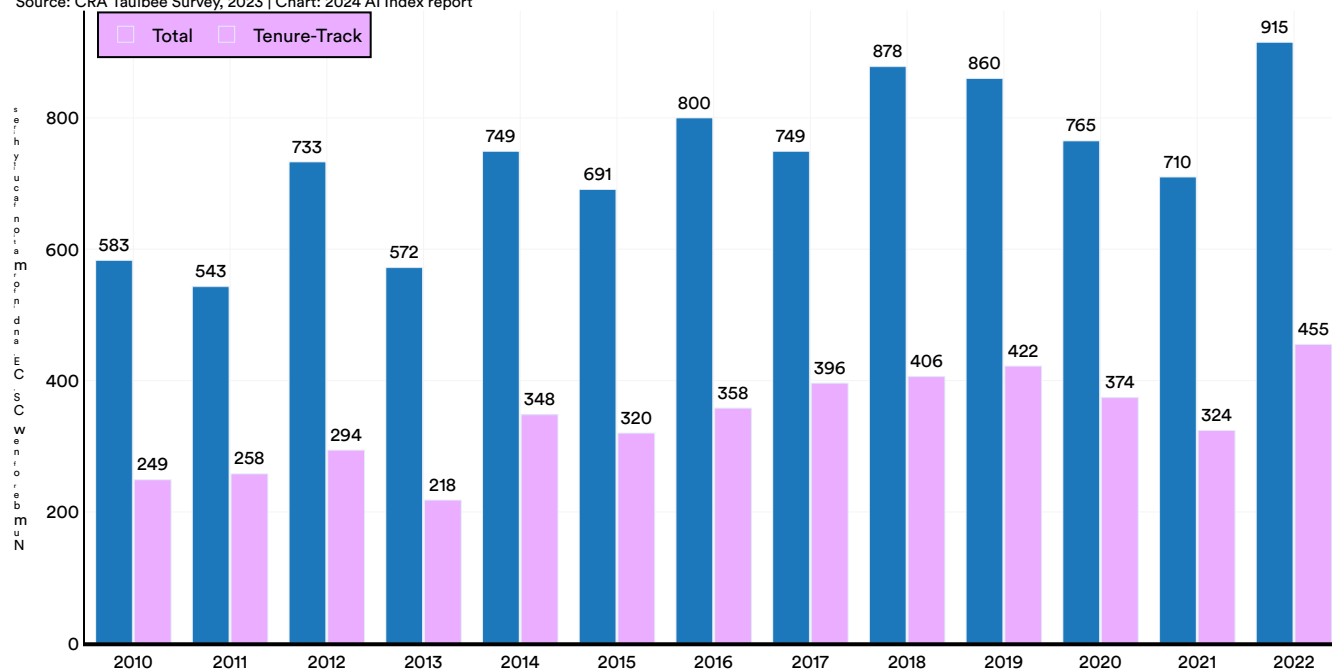


図6.1.11

2022年には、新規教員採用の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」教員が他の職を離れたため、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する教員の割合は、前年の11%および2019年の13%から7%に減少しました。

アメリカおよびカナダのCS、CE、情報学部門における新任教員の出身校、2018–22

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

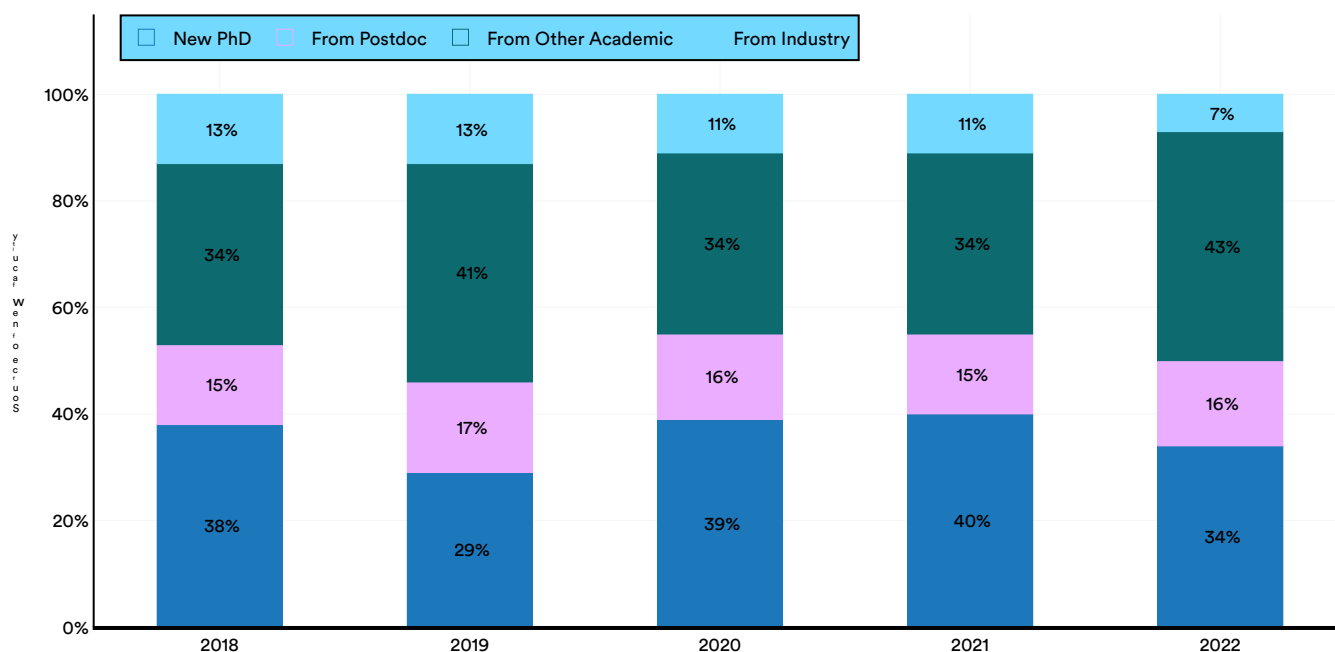


図6.1.12

過去10年間で教員ポジションが埋まらなかった理由はさまざまでした。2011年には、失敗した採用活動の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%は提示されたオファーが辞退されたためでした（図6.1.13）。対照的に、2022年にはオファーが出されなかったケースはわずか15%であり、55%はオファーが提示されたものの…

却下された。この傾向は、新しいCS教員のための市場がますます競争的になっていることを反映しているように見える。しかし、これは他の学術職との競争が激化しているのか、それとも産業界の職との競争が激化しているのかは依然として不明である。

新規CS、CE、および情報学部教員ポジションが未充足の理由（総数に占める割合）、2011–22

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



図6.1.13

2022年、北米のCS、CE、および情報分野の学科では、教員の離職が大幅に増加し、2021年の303人に対して405人に達しました（図6.1.14）。これらの離職者のうち、38.5%は他の学術職に移り、16.3%は非学術職に転職しており、これは過去数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、情報学部門における教員の減少、2011–22

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

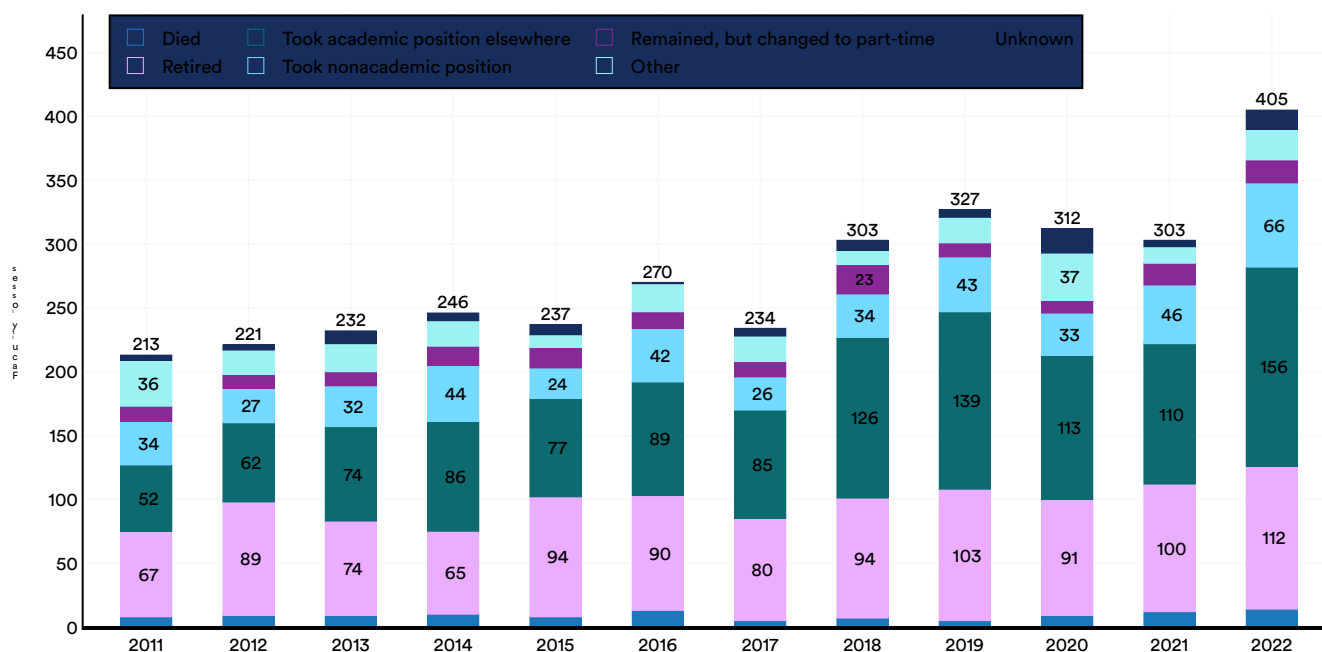


図6.1.14

2015年以降、正教授の中央値9か月給与の増加率は米国のインフレ率をわずかに下回っている一方で、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示しています。2022年には、正教授の給与は2021年より3.2%高くなりましたが、これは7%の米国インフレ率には追いついておらず、2015年より16.4%高いものの、これらの年のインフレ増加率19%には及んでいません（図6.1.15）。

アメリカ合衆国のCS教員の中央値9か月給与、2015–22

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

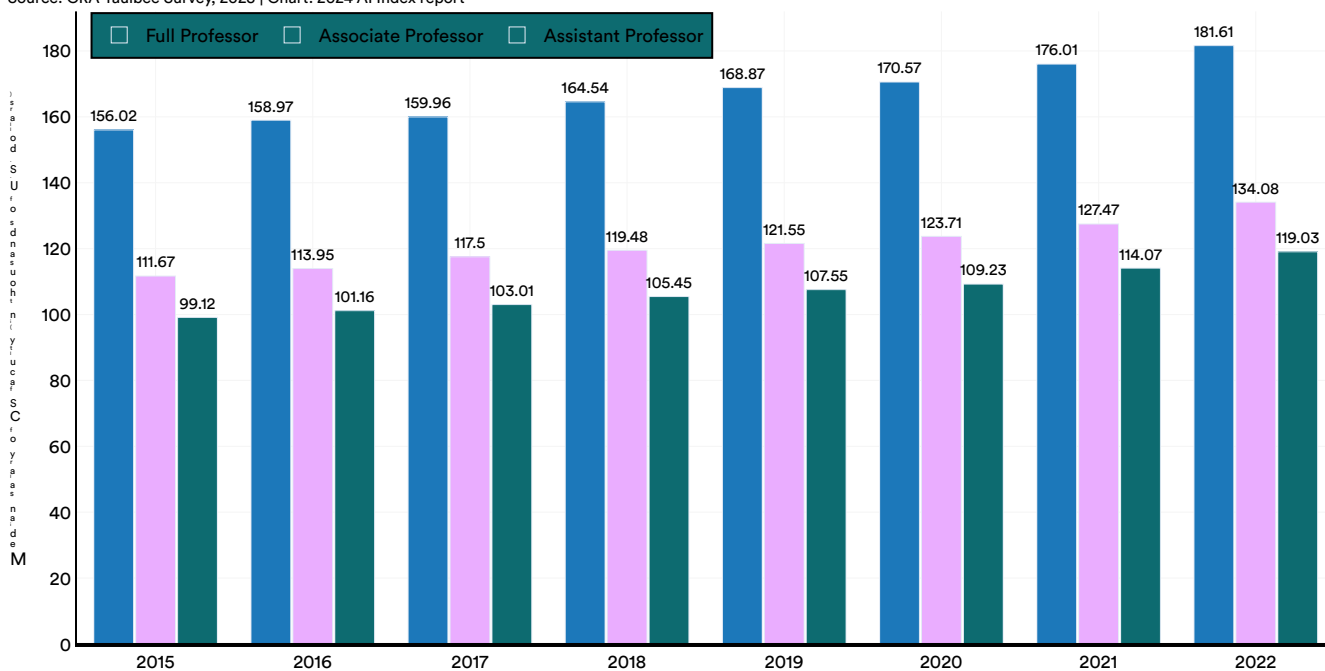


図6.1.15

2022年、CS、CE、および情報分野の新規テニユアトラック教員における国際採用の割合は、前年の13.2%から19.3%に大幅に増加した（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年のみがこれを上回っている。

2010年から2022年にかけての米国およびカナダにおける新規国際CS、CE、および情報分野のテニユアトラック教員採用（総数に占める割合）{v*}

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

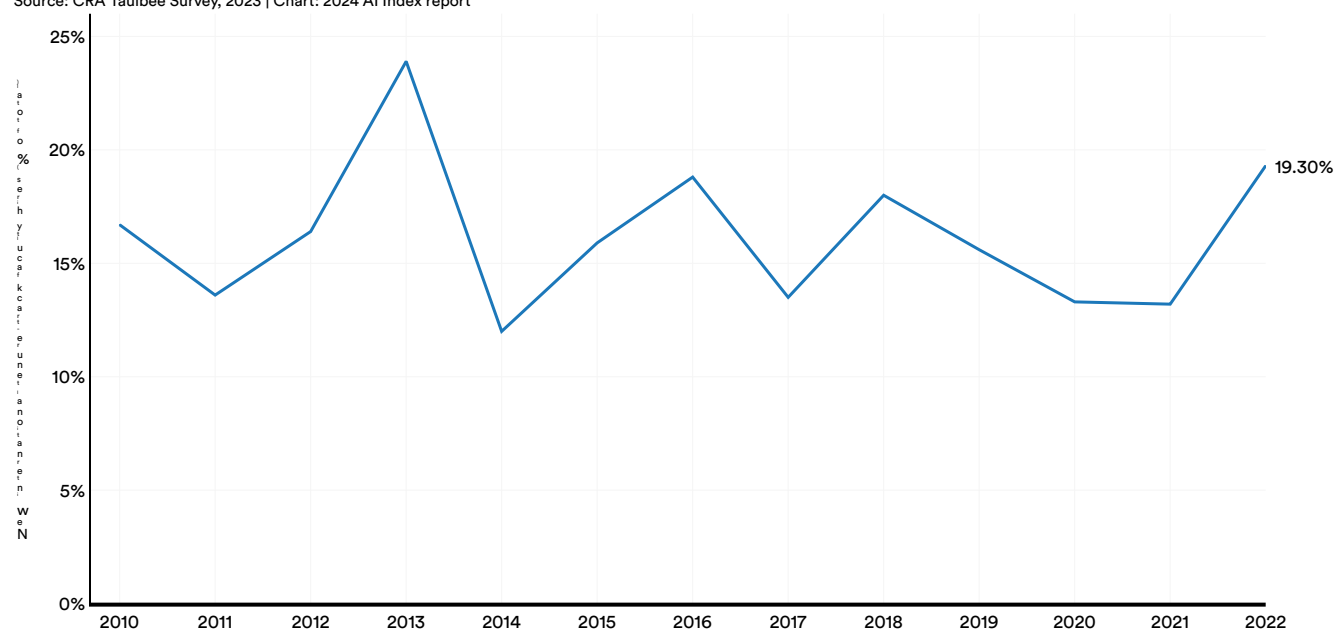


Figure 6.1.16

ヨーロッパ

ヨーロッパのCS卒業生に関するデータは、ヨーロッパの情報学教育の状況を監視することを目的の一つとする学術および研究コミュニティであるInformatics Europeから得られている³。Informatics Europeは、ヨーロッパ各国政府の統計局から情報学、CS、CE、コンピューティング、および情報技術（IT）分野の卒業生に関するデータを収集している⁴。

情報学、コンピュータサイエンス、コンピュータ工学、および情報技術の学士卒業生 {v*}

2022年、イギリスは学士レベルで情報学、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）の新卒者数が最も多く、合計で約25,000人に達しました（図6.1.17）。⁵ ドイツとトルコがこれに続きました。サンプルに含まれるほとんどの国では、これらの分野の卒業生数が10年前と比べて増加しましたが、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年のヨーロッパ各国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、情報技術の学士卒業生数 Source: Informatics Europe, 2023 | Chart: 2024 AI Index report

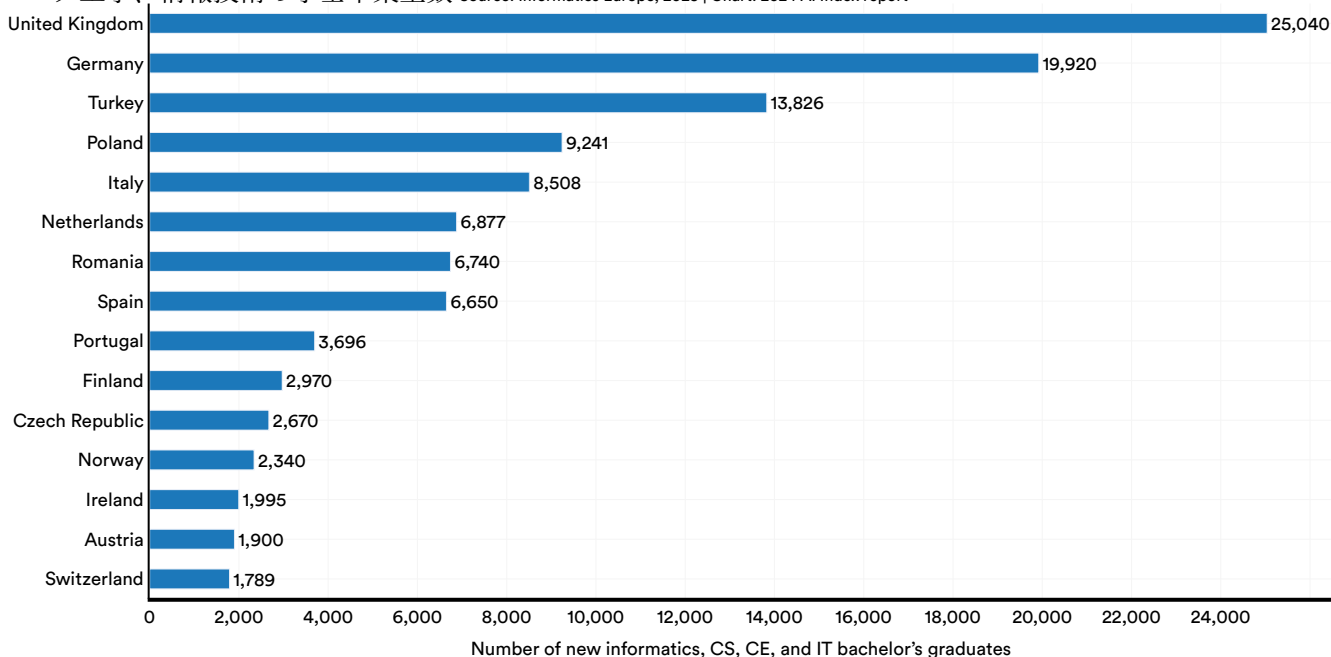


図6.1.17

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の北米データと後述のヨーロッパのCS卒業生データとの一人当たり比較を行う際に注意が必要です。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.

ヨーロッパ各国における新しい情報学、CS、CE、およびIT学士卒業生の割合変化、2012年対2022年

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

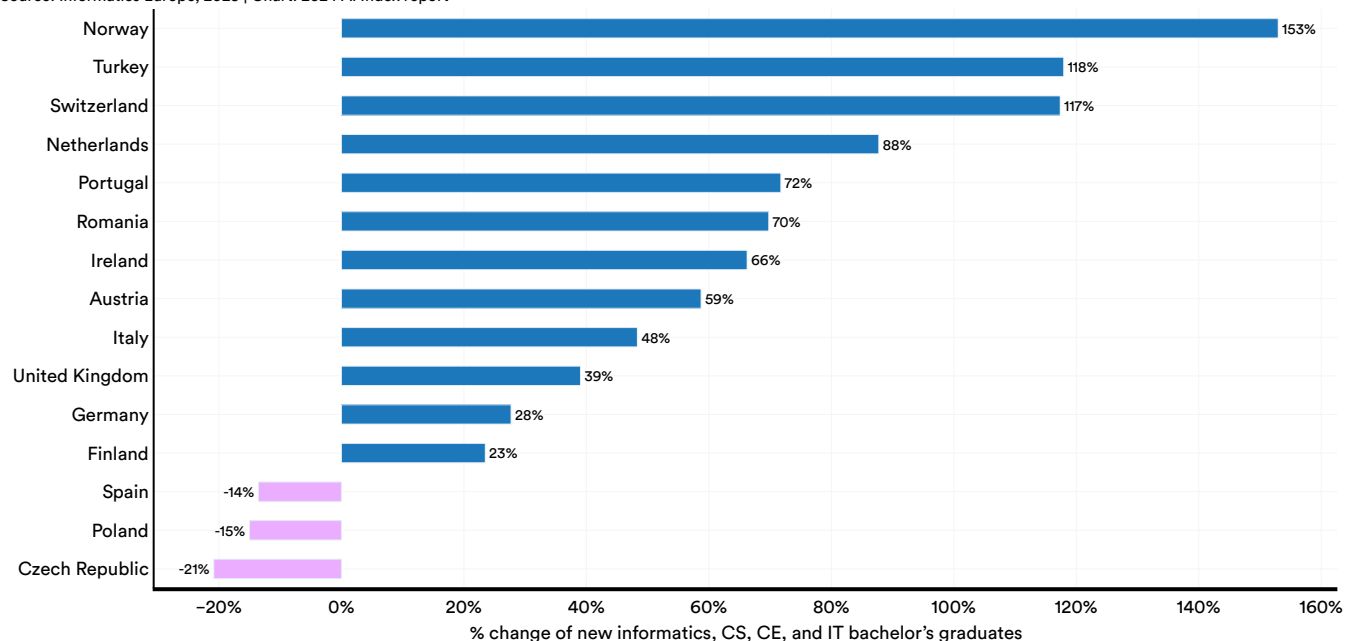


図6.1.18

フィンランド（53.4）、ノルウェー（42.6）、オランダ（38.6）は、人口10万人あたりの情報学、コンピュータサイエンス（CS）、コンピュータ工学（CE）、情報技術（IT）の新規学士卒業生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ヨーロッパ諸国で情報学、CS、CE、ITの学士卒業生数の合計が増加しています（図6.1.20）。