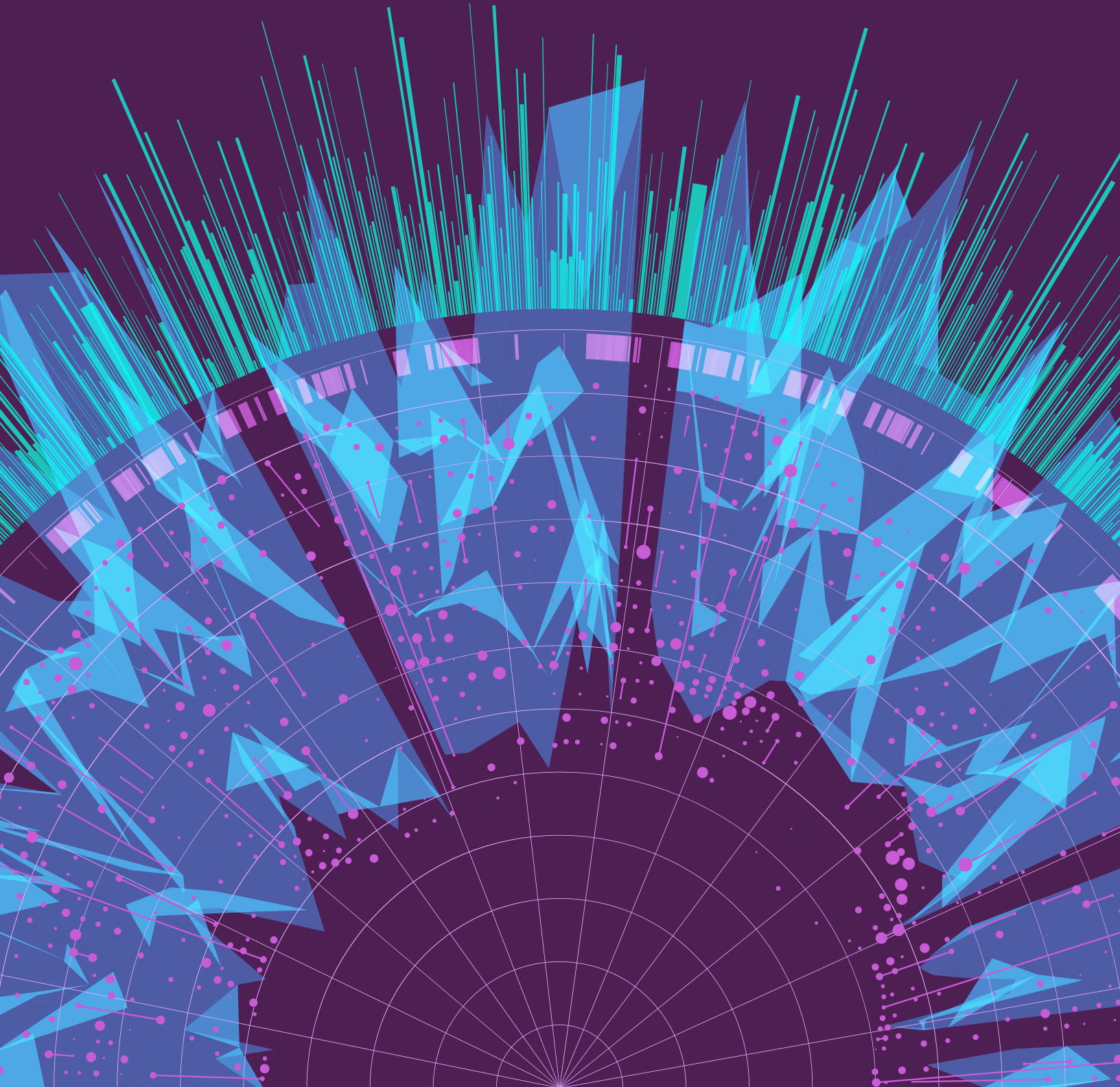




Artificial Intelligence
Index Report 2024

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

パブリックデータにアクセスします

概要

今年のAIインデックスでは、科学および医学の科学と医学のAIに関する新しい章を紹介します。グラフィキャストなどの高度な気象予測システムやGNOMEなどの改善された材料発見アルゴリズムなど、2023年の傑出したAI強化された科学的成果を探ります。この章では、医療AIシステムのパフォーマンス、SynthSRやImmunoseiraなどの重要な2023 AI主導の医療革新、およびFDA AI関連の医療機器の承認の傾向も検討しています。

章のハイライト

1. AIのおかげで、科学的進歩はさらに加速します。2022年、AIは科学的発見を進め始めました。しかし、2023年には、アルファデフから材料の発見のプロセスを容易にするアルゴリズムの並べ替えをより効率的にするAlphadevから、さらに重要な科学関連のAIアプリケーションの発売が見られました。

2. AIは、薬が大きな進歩を前進させるのを助けます。2023年には、パンデミック予測を強化するEvescapeを含むいくつかの重要な医療システムが開始され、AI駆動型の突然変異の分類を支援するアルファミメーションがありました。AIは、医学の進歩を推進するためにますます利用されています。

3. 知識豊富な医療AIが到着しました。過去数年にわたって、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ベンチマークの顕著な改善を示してきました。2023年の傑出したモデルであるGPT-4 Medpromptは、90.2%の精度率に達し、2022年の最高スコアから22.6パーセントの増加を示しました。

4. FDAは、ますます多くのAI関連の医療機器を承認します。2022年、FDAは139のAI関連医療機器を承認しました。これは2021年から12.1%増加しました。2012年以来、FDAが承認したAI関連の医療機器の数は45倍以上増加しました。AIは、実際の医療目的でますます使用されています。

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択された2023年の重要なAI関連の科学的ブレイクスルーを強調しています。

5.1 注目すべき科学的マイルストーン

alphadev

Alphadevは、より速いソートアルゴリズムを発見します

Alphadevは、計算アルゴリズム強化の分野の科学者とエンジニアによる数十年の仕事で改善された新しいAI強化学習システムです。Alphadevは、既存の人間のベンチマークよりも少ない手順でアルゴリズムを開発しました

ソート3、ソート4、ソート5などの短いシーケンスの基本的なソートアルゴリズム（図5.1.1）。Alphadevによって発見された新しいアルゴリズムのいくつかは、LLVM標準C++ソートライブラリに組み込まれています。これは、10年以上ぶりの図書館のこの部分の最初の更新を示し、補強学習を使用して設計された最初の追加です。

アルファデフvs.アルゴリズムの長さを最適化する場合の人間のベンチマ

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

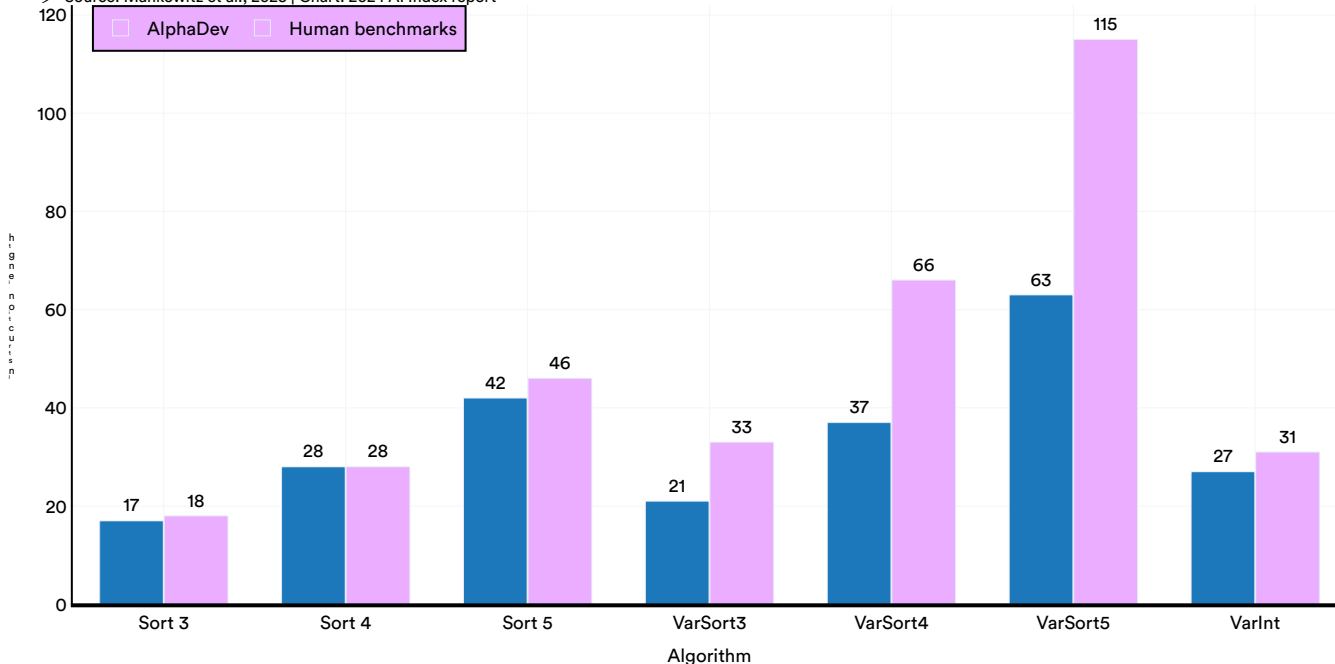


Figure 5.1.1

フレキシクブ

フレキシクブを使用した3Dメッシュ最適化

コンピューターグラフィックスの重要な3Dメッシュ生成には、3Dオブジェクトを定義するための頂点、エッジ、面のメッシュの作成が含まれます。ビデオゲーム、アニメーション、医療イメージング、科学的視覚化の鍵です。従来の等骨格抽出アルゴリズムはしばしば、限られた解像度、構造的剛性、および数値的不安定性に苦しんでおり、その後影響します。

品質。FlexCubesは、勾配ベースの最適化と適応可能なパラメーターにAIを使用することにより、これらの制限の一部に対処します（図5.1.2）。この方法では、正確でローカライズされたメッシュ調整が可能になります。メッシュ再構成に微分可能な等圧性を利用する他の主要な方法と比較して、Flexicubesは、基礎となる基礎真理とはるかに密接に整合するメッシュ抽出を実現します（図5.1.3）。

サンプルフレキシクブの表面再構成

Source: Nvidia, 2023

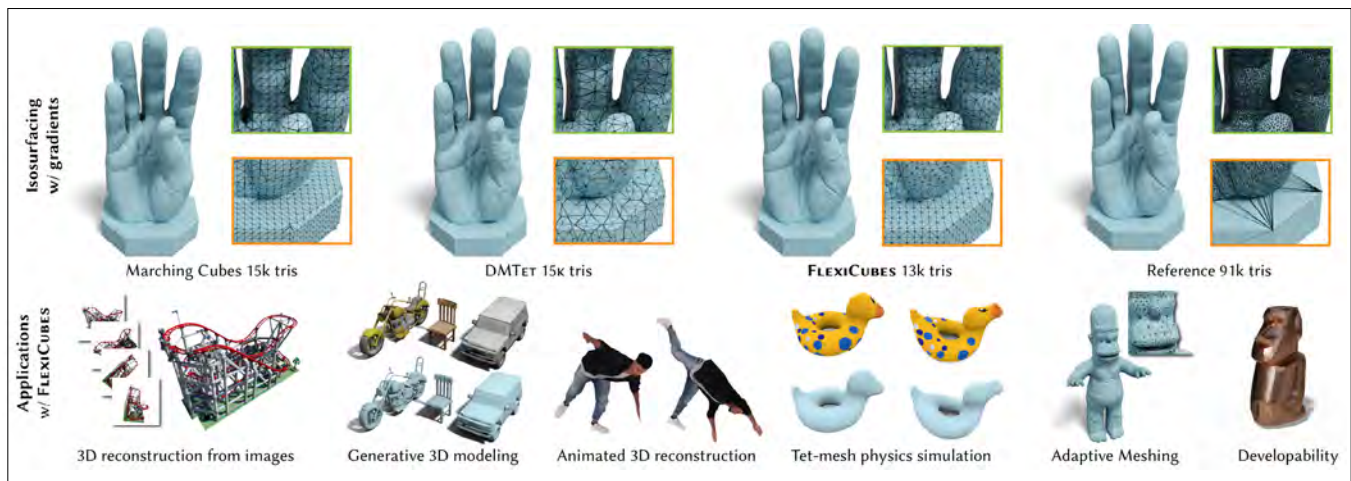


図5.1.2

3Dメッシュ再構成で定量的結果を選択します

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

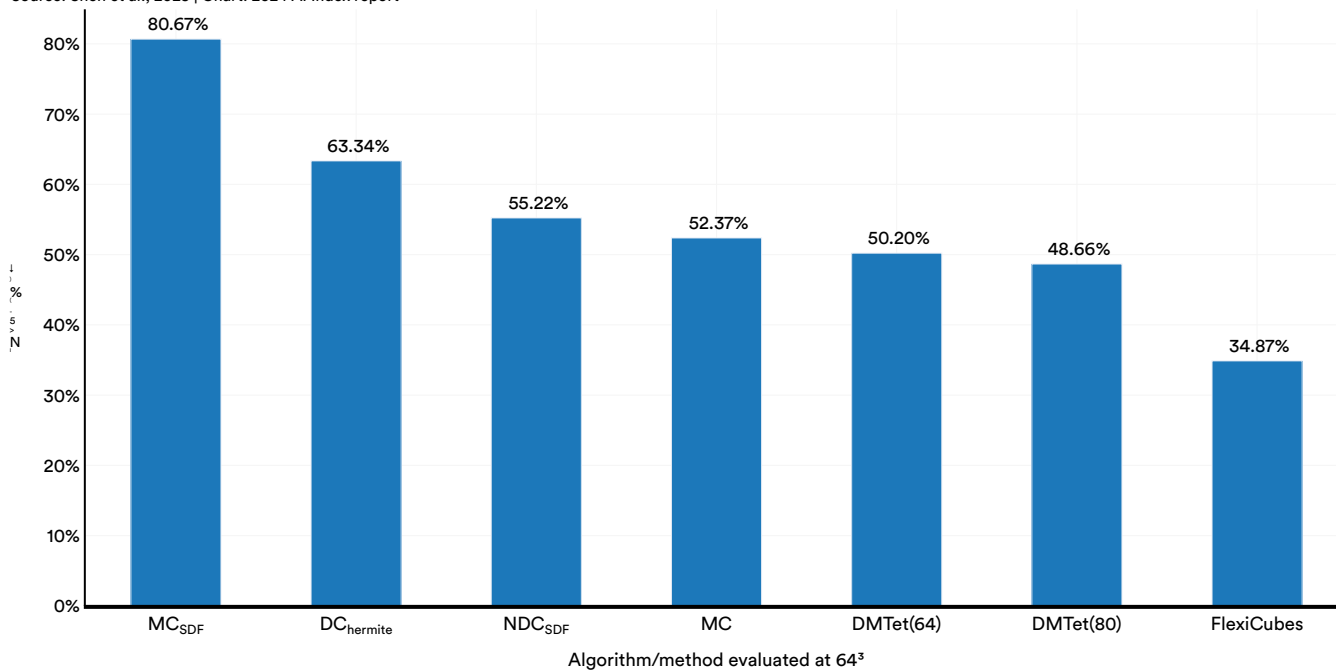


図5.1.3

シンボット

有機分子を合成するためのAI駆動型ロボット化学者

Synbotは、化学合成計画のためのAIソフトウェア層、コマンドを翻訳するためのロボットソフトウェア層、および実験を行うための物理ロボット層を含む多層システムを採用しています。AIとロボットシステムの間の閉ループフィードバックメカニズムにより、Synbotは確立された参照に等しい、またはそれを超える収量を持つ合成レシピを開発できます(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]の合成を目的とした実験では、シンボットは変換を達成する複数の合成式を開発し、サージンを生成しました。

Synbot Design

Source: Ha et al., 2023

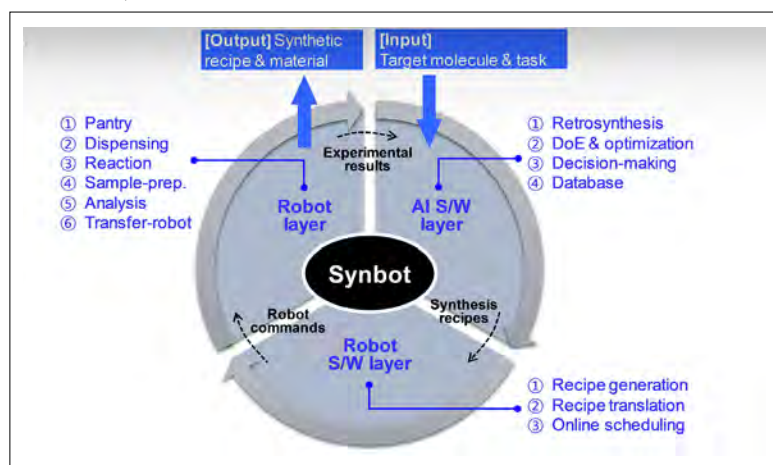


図5.1.4

80%中期の参照範囲と合成を大幅に短縮して完了しました(図5.1.5)。Synbotの有機合成の自動化は、PharmaceuticalsやMaterials Scienceなどの分野での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、シンボット vs. 参照

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

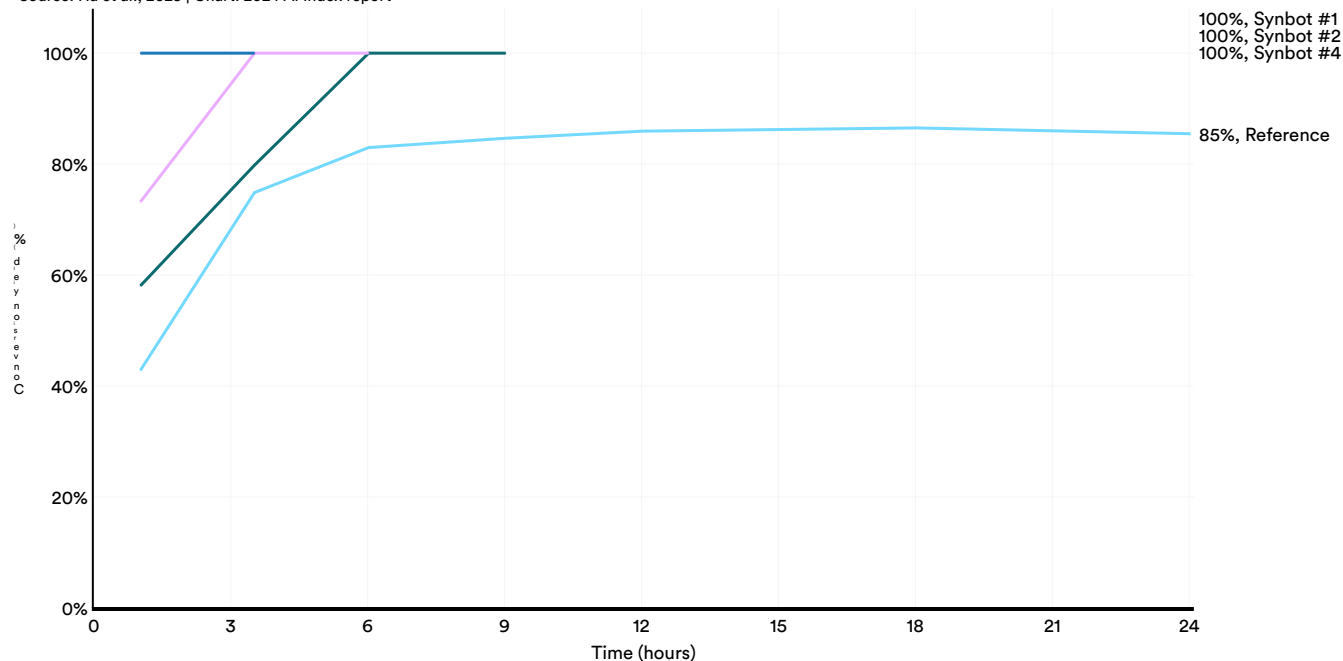


図5.1.5

グラフキャスト

より正確なグローバルな気象予測

グラフキャスト付き

Graphcastは、1分未満で非常に正確な10日間の気象予測を提供する新しい気象予測システムです（図5.1.6）。グラフのニューラルネットワークと機械学習を利用して、グラフキャストは、温度、風速、大気条件を予測するために広大なデータセットを処理します。

もっと。図5.1.7は、グラフキャストのパフォーマンスを、現在の業界の最先端の気象シミュレーションシステムと比較しています：高解像度予測（HRES）。Graphcastは、より低い根平均四角エラーを投稿します。つまり、その予測は観測された気象パターンに密接に対応しています。グラフキャストは、天候パターンを解釈し、極端な気象イベントの準備を強化し、世界の気候研究に貢献する上で貴重なツールになります。

グラフキャストの天気予報

Source: DeepMind, 2023

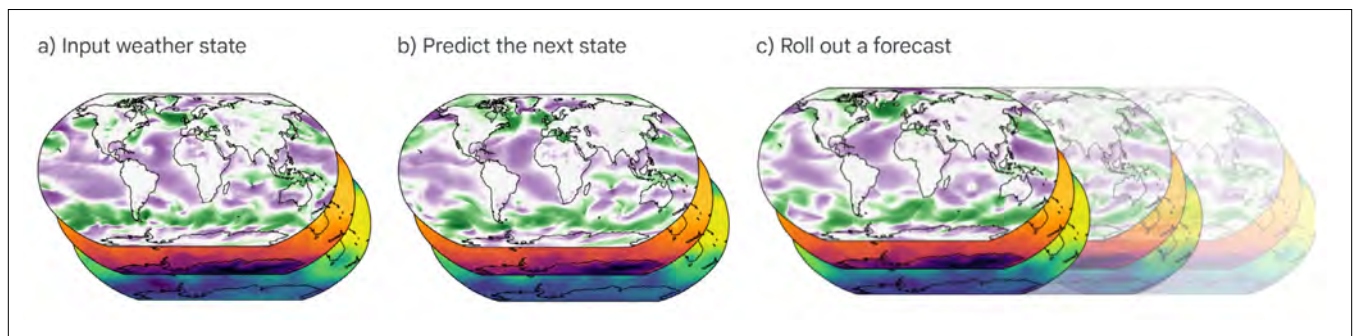


Figure 5.1.6

10日間のZ500予測スキル：Graphcast vs. hres

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

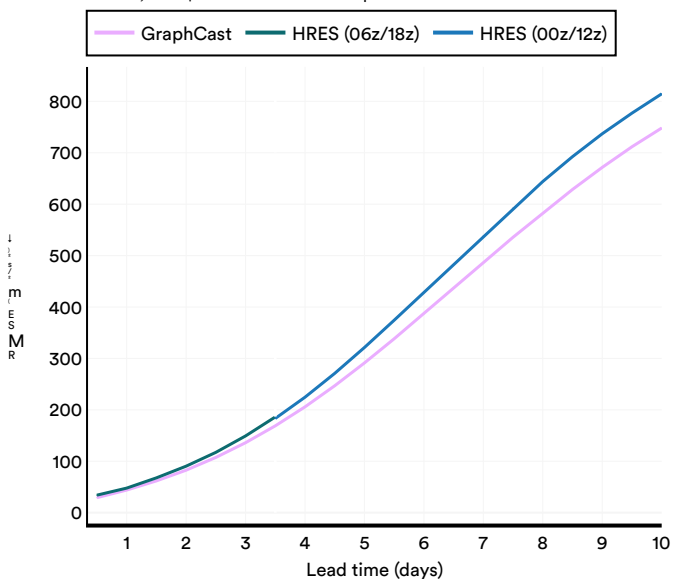


Figure 5.1.7

gnome

GNOMEで新しい素材を発見します

新しい機能材料の検索は、ロボット工学や半導体製造を含むさまざまな科学分野での進歩の鍵です。しかし、この発見プロセスは通常、高価で遅いです。Googleの研究者による最近の進歩は、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トレーニングされたときにこのプロセスを促進できることを実証しています。彼らのモデルであるGNOMEは、材料発見の主要な方法である材料プロジェクトを、かなり多くの安定した結晶を識別することにより、材料プロジェクトを上回りました（図5.1.8）。GNOMEは220万の新しい結晶構造を発表しましたが、多くは人間の研究者によって見落とされています（図5.1.9および図5.1.10）。GNOMEのようなAI主導のプロジェクトの成功は、科学的ブレークスルーをスピードアップする際のデータの力とスケーリングを強調しています。

サンプル材料構造

Source: Merchant et al., 2023

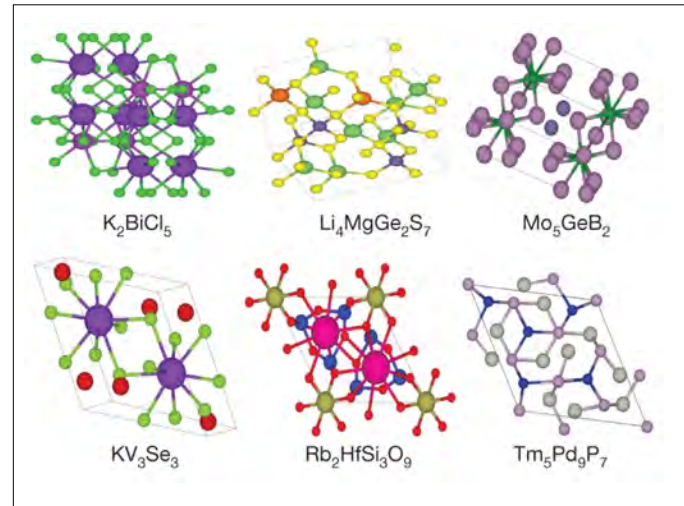


図5.1.8

Gnome vs. Materials Project：安定したクリスタルカ

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

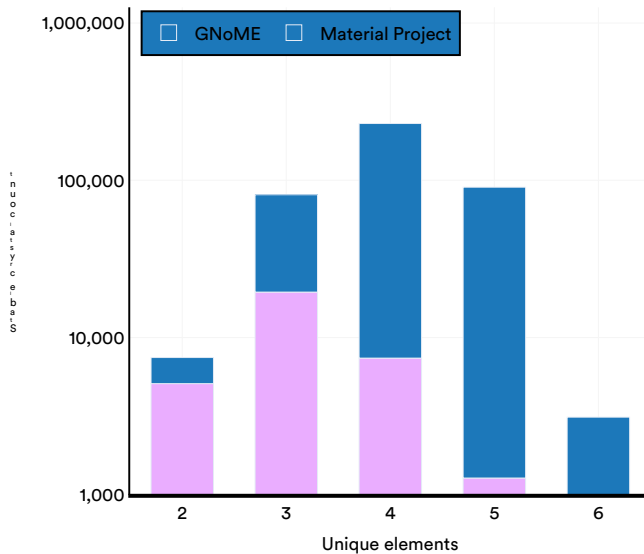


図5.1.9

Gnome vs. Materials Project：個別のプロトタイプ

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

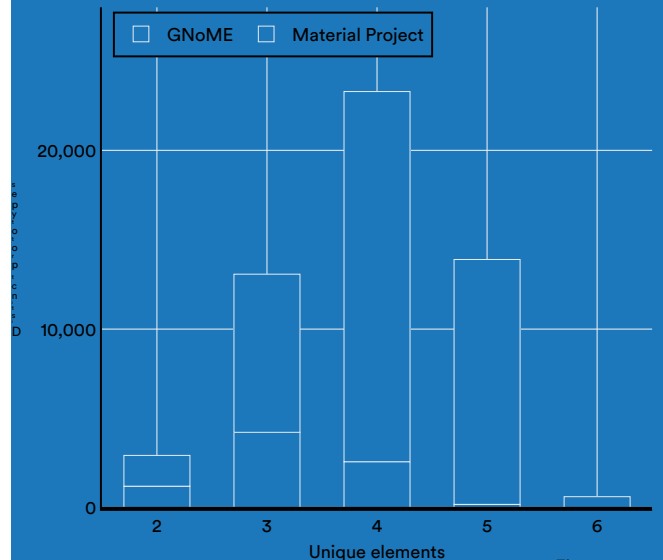


Figure 5.1.10

洪水予測

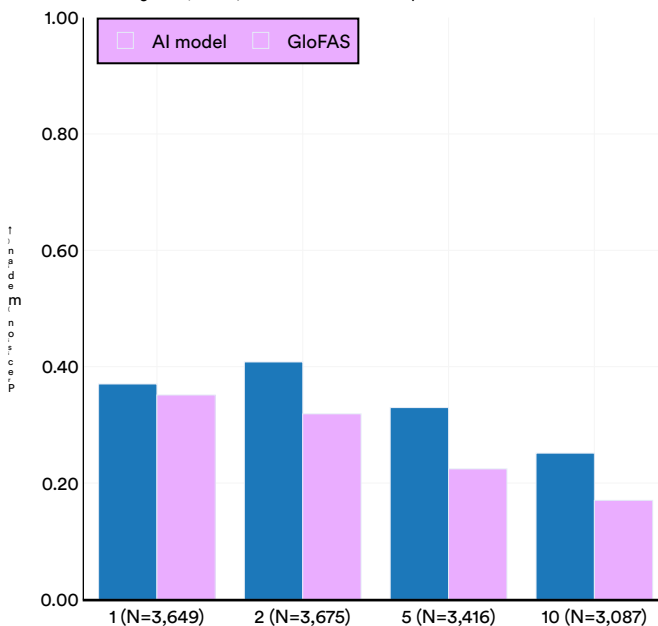
より正確で信頼性の高い洪水予測のためのAI

2023年に導入された新しい研究は、大規模な洪水イベントの予測に大きな進歩を遂げました。最も一般的な自然災害の中で、洪水は、予防と緩和のためのインフラストラクチャが不足している発展途上国で特に壊滅的な影響を及ぼします。その結果、これらのイベントをさらに事前に予測できるより正確な予測方法を開発すると、大きなプラスの影響が生じる可能性があります。

Googleの研究者チームは、AIを使用して、非ゲージに陥った盆地にも適用できる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。¹これらの革新的な方法は、Glofasなどの現在の最先端モデルと一致または上回る正確さで、5日前に特定の極端な洪水イベントを予測できます。AIモデルは、優れた精度（正の予測の精度）を示し、リターン期間イベントで優れた精度（すべての関連するインスタンスを正しく識別する能力）を示し、主要な現代的な方法を上回っています（図5.1.11）。

AIモデルとリターンペリオ全体のグローファの予測

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



D

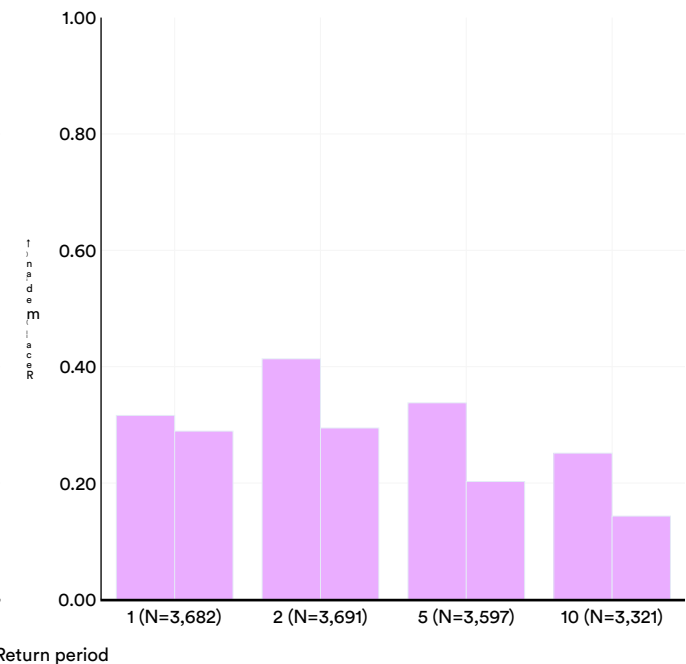


図5.1.11

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、診断を行う臨床医を支援するためにポリープを検出するためのアプリケーションで、ヘルスケアでますます価値が高まっています。AIのパフォーマンスが改善され続けるにつれて、医療行為への影響を監視することがますます重要になります。このセクションでは、2023年に導入された重要なAI関連の医療システム、臨床AIの知識の現在の状態、および病院管理の強化を目的とした新しいAI診断ツールとモデルの開発を強調しています。

5.2 AI医学

顕著な医療システム

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

Synthsr

高度な分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T1加重画像に変換するAIツールです（図5.2.1）。この進歩は、高度な研究での多くのスキャンの使用を以前に制限していたスキャン品質の変動の問題に対処しています。これらのスキャンを、その高いコントラストと明確な脳構造の描写で知られるT1強調画像に変換することにより、SynthSRは詳細な3D脳レンダリングの作成を促進します。SynthSRを使用した実験は、スキャンレベルと被験者レベルの両方で観測されたボリューム間の堅牢な相関を示し、SynthSRが高解像度T1スキャンによって生成されたもの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、SynthSRスキャンが選択された脳領域全体での根真実の観察に対応する程度を示しています。SynthIDは、脳構造の視覚化と分析を大幅に改善し、神経科学的研究と臨床診断を促進します。

Synthsr Generations

Source: Iglesias et al., 2023

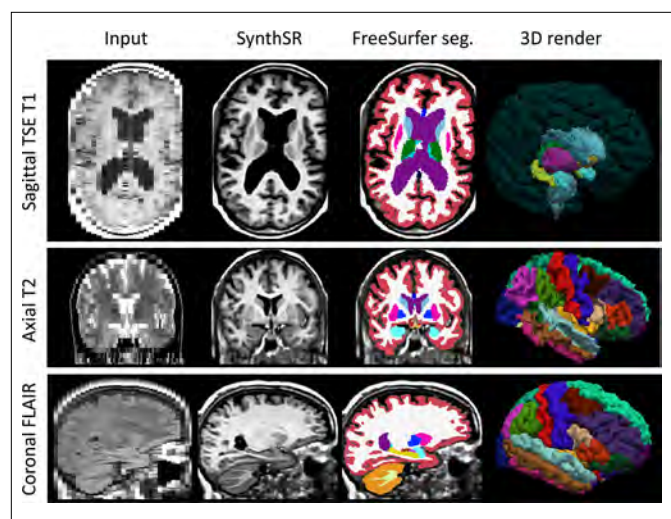


Figure 5.2.1

選択した脳領域の地上真実ボリュームとのSynthSR相関

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

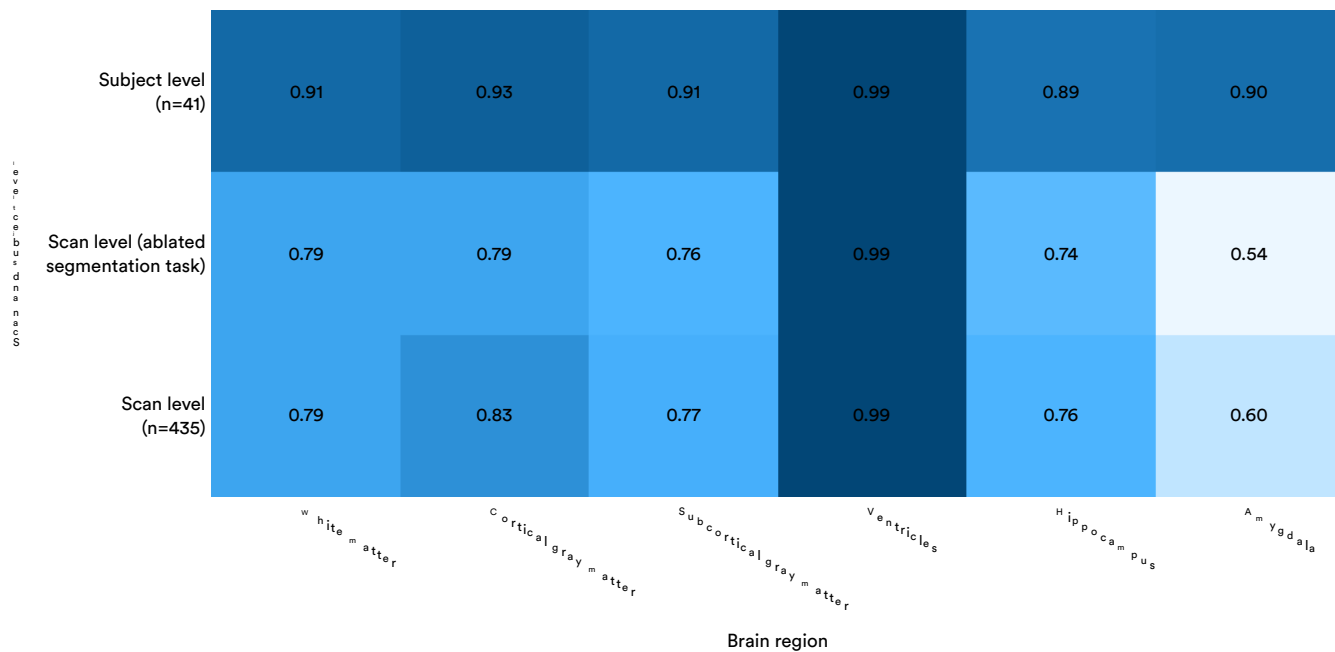


Figure 5.2.2

結合した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 ック赤外線センサー

パーキンソン病やアルツハイマー病などの神経変性疾患の診断は、バイオマーカーの迅速かつ正確な識別に依存しています。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化に焦点を当てることができるという点で有用です。ただし、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、表面強化された赤外線吸収（SEIRA）分光法を免疫測定技術（免疫酵素;図5.2.3）を使用するAI共役プラズモニク赤外線センサー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をAIシステムによる予測と比較したテストでは、予測の精度は、実際の報告率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

Source: Kavungal et al., 2023

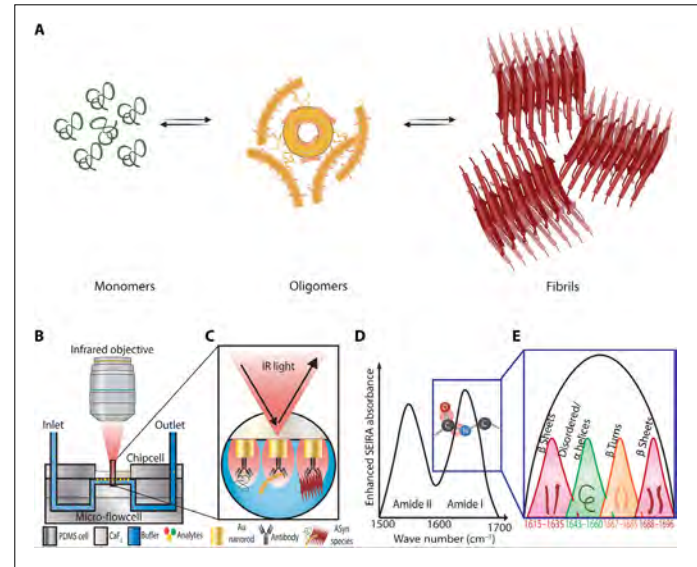


図5.2.3

ディープニューラルネットワークは、テストサンプルの実際のブリルの割合を
予測しました Source: Kavungal et al., 2023 | Chart: 2024 AI Index report

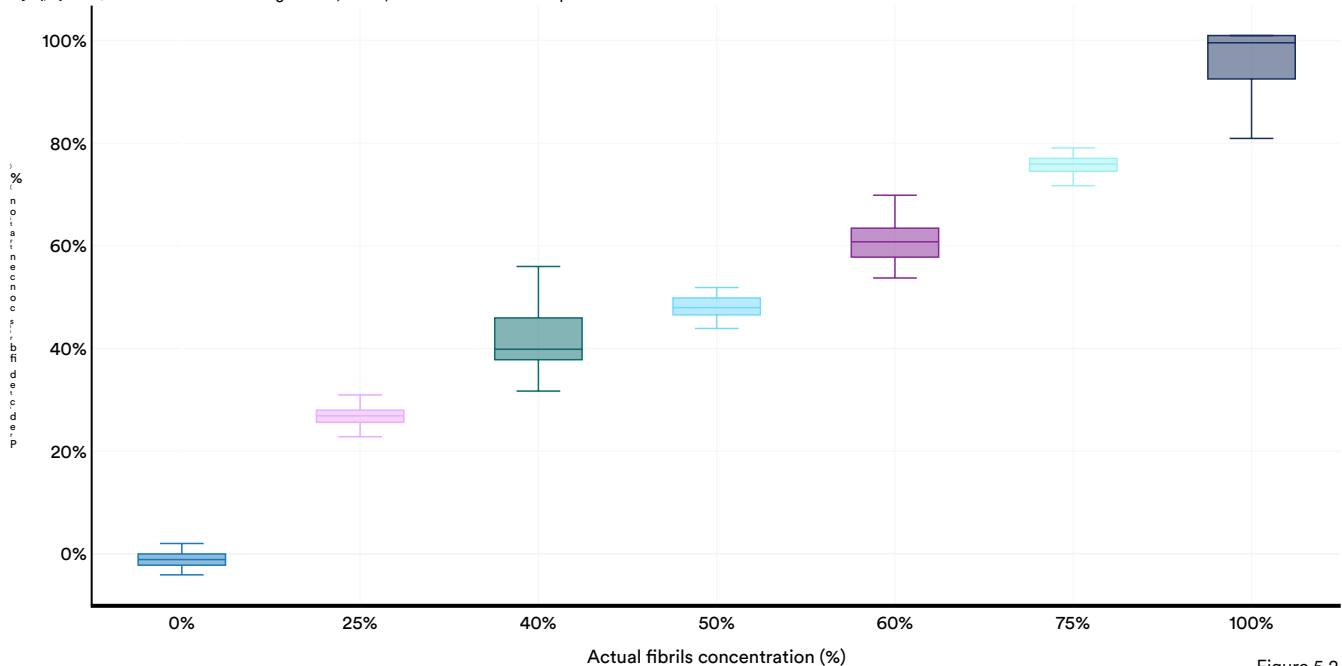


Figure 5.2.4

evescape

パンデミックの準備のためのウイルスの進化を予測します

ウイルス変異の予測は、ワクチンの設計とパンデミックの最小化に不可欠です。リアルタイムのウイルス株と抗体データに依存する従来の方法は、データ不足による初期のパンデミック段階での課題に直面しています。Evescapeは、進化を予測する歴史的なシーケンスと生物物理学のおよび構造的情報で訓練された新しいAIディープラーニングモデルです。

ウイルスの（図5.2.5）。Evescapeは、観察されたSARS-CoV-2変異の50.0%を予測する現在のひずみデータとは無関係にウイルス脱出を評価し、46.2%と32.3%を予測する従来のラボ研究を上回り、変異の24%しか予測していなかった以前のモデルを評価します（図5.2.6）。このパフォーマンスは、将来のパンデミックの準備と対応努力を強化するための貴重な資産としてのEvescapeの可能性を強調しています。

evescape design

Source: Thadani et al., 2023

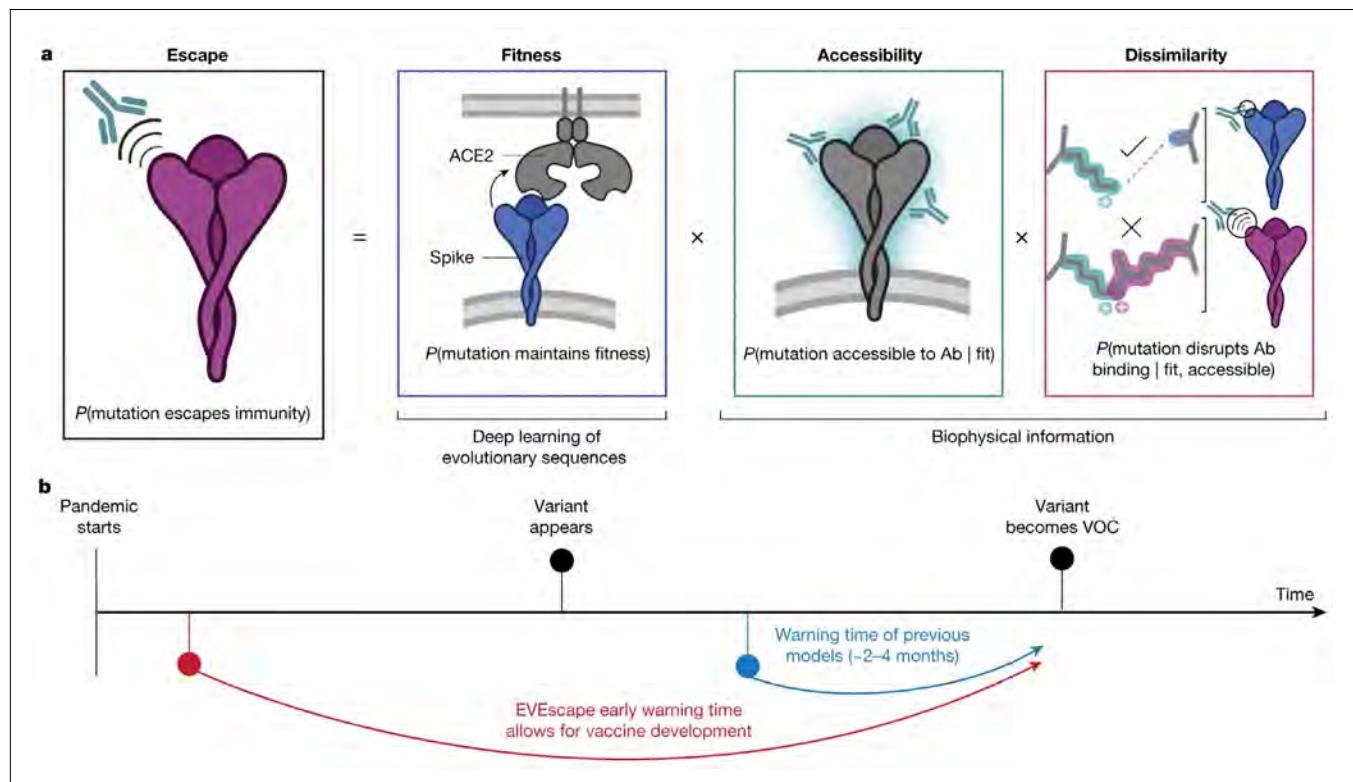


図5.2.5

SARS-COV-2 RBD変異予測に関するEvescape vs.他のモデル

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

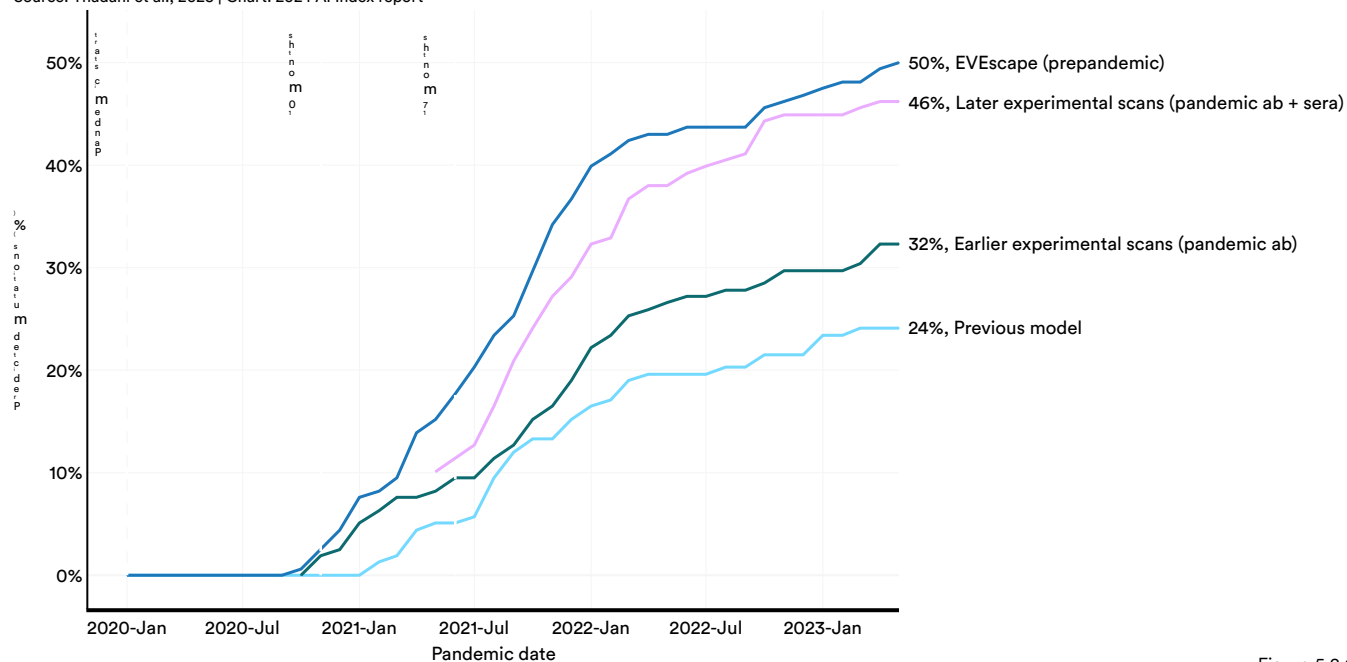


Figure 5.2.6

アルファミッセンス AI変異のより良い分類

科学者は、どの遺伝的変異が病気につながる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的変異により、突然変異が良性であるか病原性であるかを決定するには、労働集約的な実験が必要です。

2023年、Google Deepmindの研究者は、7100万のミスセンスバリエーションの病原性を予測する新しいAIモデルであるAlphamissenseを発表しました。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変化であり（図5.2.7）、癌を含むさまざまな疾患につながる可能性があります。7100万の可能なミスセンスバリエーションのうち、アルファミッセンスは89%に分類され、57%が良性であり、32%が病原性である可能性が高いと識別し、残りは不確実であると分類されました（図5.2.8）。対照的に、ヒトのアノテーターは、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。

ヘマグロビンサブユニットベータ（HBB）

Source: Google DeepMind, 2023



図5.2.7

アルファミッセンス予測

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

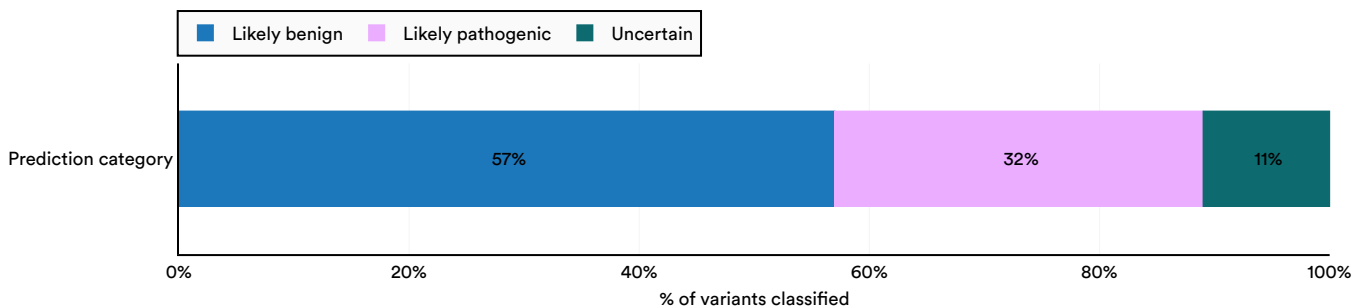


Figure 5.2.8

人間のパンゲノム参照

AIを使用してヒトゲノムをマッピングします

ヒトゲノムは、ヒト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年にリリースされ、2022年に更新されました。しかし、更新はやや不完全でした。血液型などのさまざまな遺伝的変異を組み込んでおらず、多様な祖先グループを完全にマッピングしませんでした。したがって、既存のゲノム参照の下では、特定のグループの人々の病気を検出したり、治療法を見つけることは困難です。

2023年、60の機関の119人の科学者で構成される人間の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、より代表的なヒューマンゲノムマップを開発しました(図5.2.9)。研究者は顕著な精度を達成し、タンパク質コーディング遺伝子の99.07%、タンパク質コーディング転写産物の99.42%、非コーディング遺伝子の98.16%、および図5.2.10に詳述されている非コード転写産物の98.96%を注釈しました。

ゲノムのMHC領域のグラフゲノム

Source: Google Research, 2023

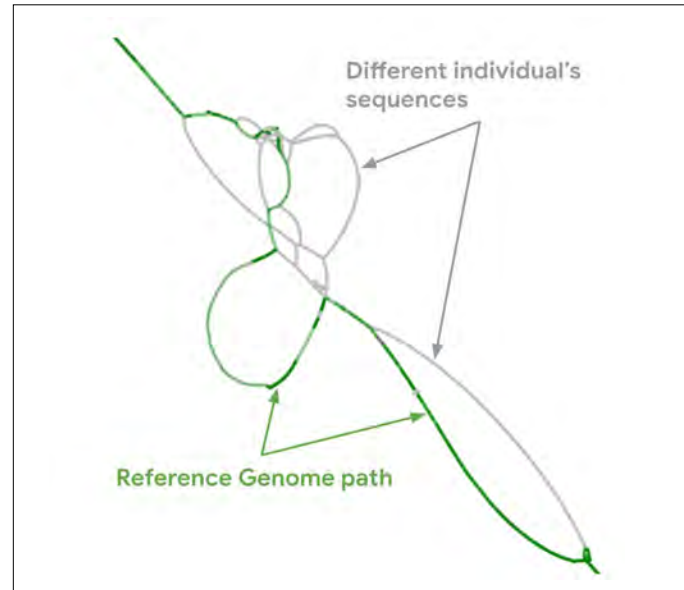


図5.2.9

ゲノムのこの最新バージョンは、これまでのヒトゲノムの最も包括的で遺伝的に多様なマッピングを表しています。

Ensemblマッピングパイプライン結果

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

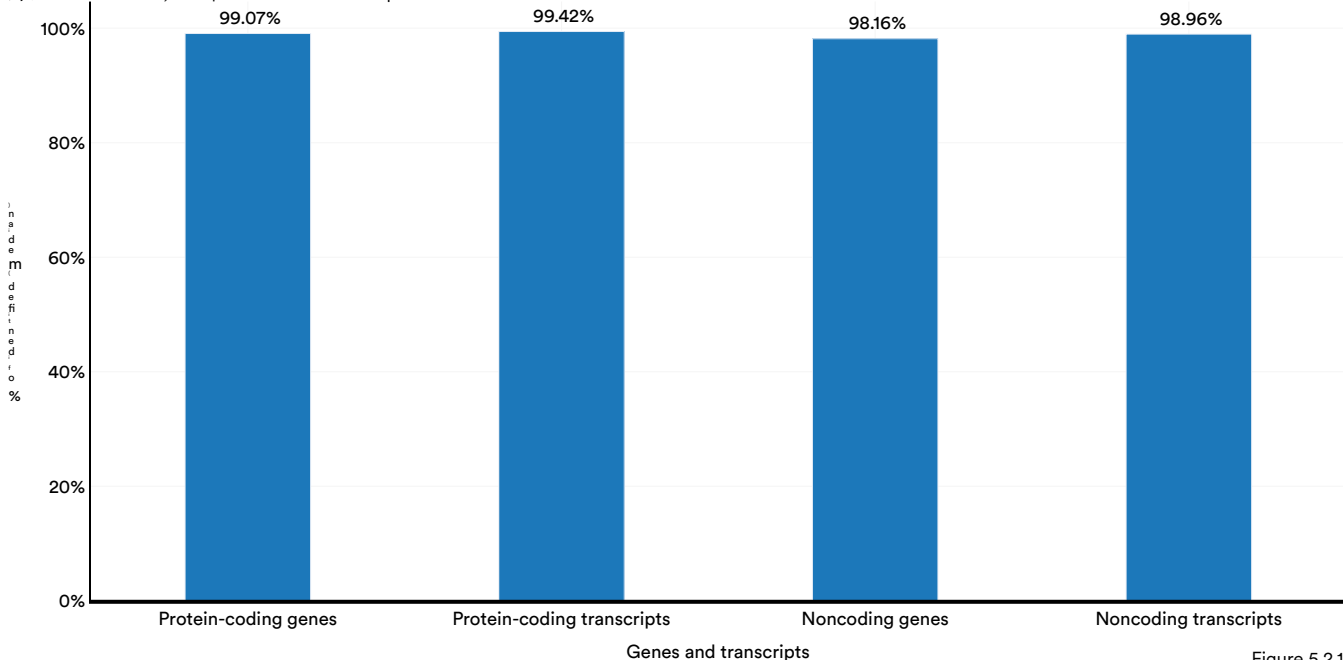


Figure 5.2.10

臨床知識

AIモデルの臨床知識を評価するには、医療専門知識の程度、特に臨床環境で適用可能な知識を決定することが含まれます。

主要なシステムであるGPT-4 Medpromptで、顕著な改善は、2022年のトップスコアから22.6パーセントポイントの90.2%の精度率を達成しました（図5.2.11）。MEDQAの設立以来、このベンチマークのAI機能はほぼ3倍になり、臨床的に知識のあるAIシステムの急速な改善が示されています。

Medqa

2020年に導入されたMEDQAは、医師に挑戦するために設計された60,000を超える臨床的質問を特徴とする専門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ベンチマークのAIパフォーマンスが見られました MEDQA：精度

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

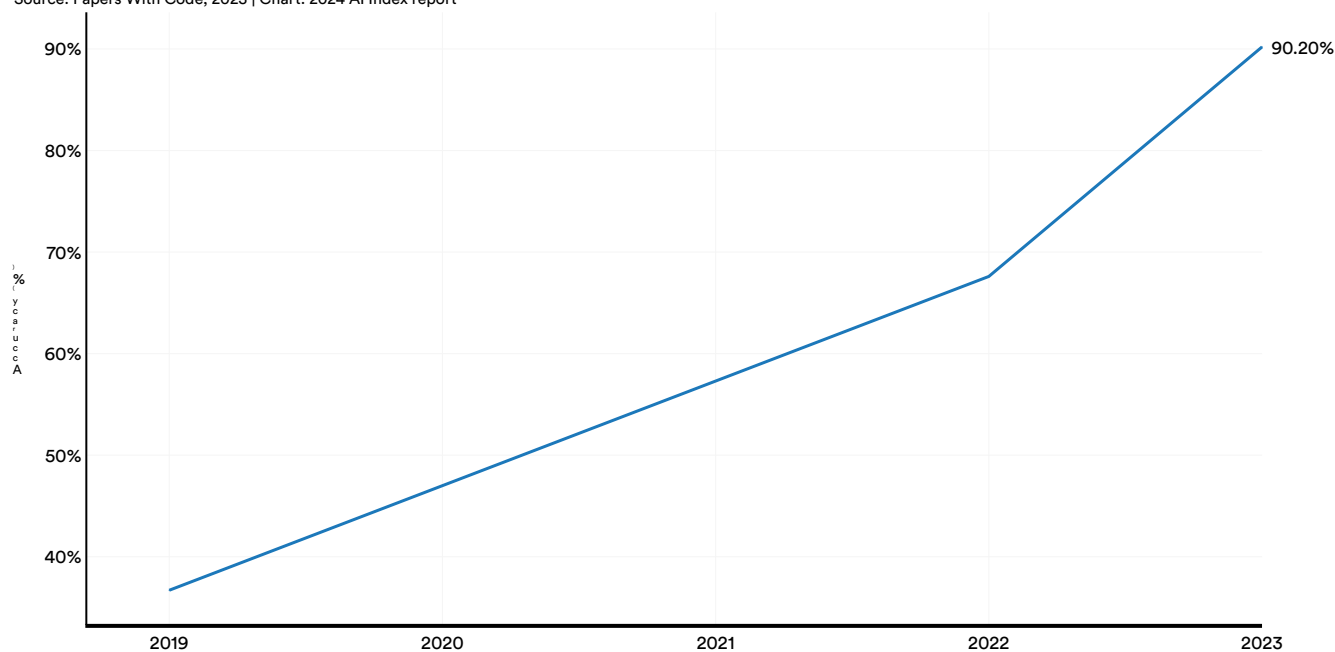


図5.2.11

ハイライトされた研究：

gpt-4 medprompt

LLMは印象的な一般的な知識を示していますが、一般に、医学的な質問に答えるなど、専門的な知識に秀でるには重要な微調整が必要であると想定されています。微調整には、ドメイン固有のデータでLLMをトレーニングすることが必要です。

2023年後半のマイクロソフトからの研究は、この仮定を覆しました。この研究では、迅速なエンジニアリングを採用して、GPT-4を4つの挑戦的な医療ベンチマークのグループであるMultiMedQAベンチマークスイートで顕著なパフォーマンスを達成するために指示しました（図5.2.12）。GPT-4 MEDPROMPTは、PubMedQA、MedMCQA、およびMMLUを含むいくつかの有名な医療ベンチマークの複数選択セクションで、それぞれ3.0、21.5、および16.2パーセントポイントで、TOP 2022モデルであるFlan-Palm 540Bのパフォーマンスを超えました。また、当時の最先端のMed-Palm 2のパフォーマンスを超えました（図5.2.13）。

さらに、前述のように、GPT-4 Medpromptは、MEDQAベンチマークの90%の精度マークを上回った最初のものでした。このブレイクスルーは、GPT-4 Medpromptの例外的で潜在的に臨床的に有用な医療能力を強調するだけでなく、特殊なドメインにモデルを適応させるために微調整が必ずしも必要ではないことを示しています。迅速な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 vs. Med-Palm 2医学的な質問に答える

Source: Nori et al., 2023

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


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

Hand-crafted CoT from Med PaLM 2

 **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サブベンチマークのモデルパフォーマンス

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

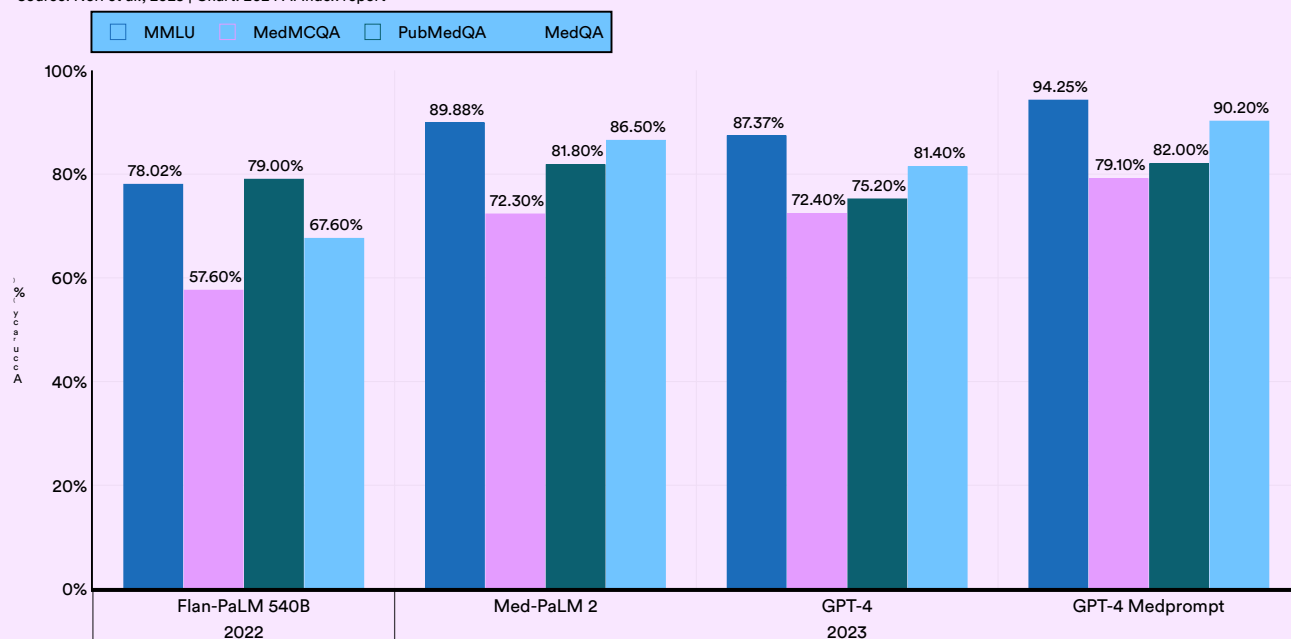


図5.2.13

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。ただし、クローズドソースです。つまり、その重量は、使用するためにより多くの一般の人々が自由に利用できるわけではありません。2023年の新しい研究では、オープンソースの医療LLMSの能力を向上させようとしています。この新しい研究の中で、Meditron-70Bは特に有望であると際立っています。このモデルは、MEDQAベンチマークで立派な70.2%の精度を実現します。これはGPT-4 MedpromptとMed-のパフォーマンスを下回っていますが

Palm 2（両方とも閉じたモデル）は、2023年の最先端の結果を大幅に改善し、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）。MEDQAでのMeditron-70Bのスコアは、オープンソースモデルによって最も高くても達成されています。医療AIが最大限の可能性に到達するためには、その能力が広くアクセス可能であることが重要です。これに関連して、Meditronは励ましの一歩を表しています。

MEDQAでの選択モデルのパフォーマンス

出典：Chen et al.、2023 | テーブル：2024 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、たとえば放射線やがんの検出など、診断目的でも使用できます。

ハイライトされた研究：

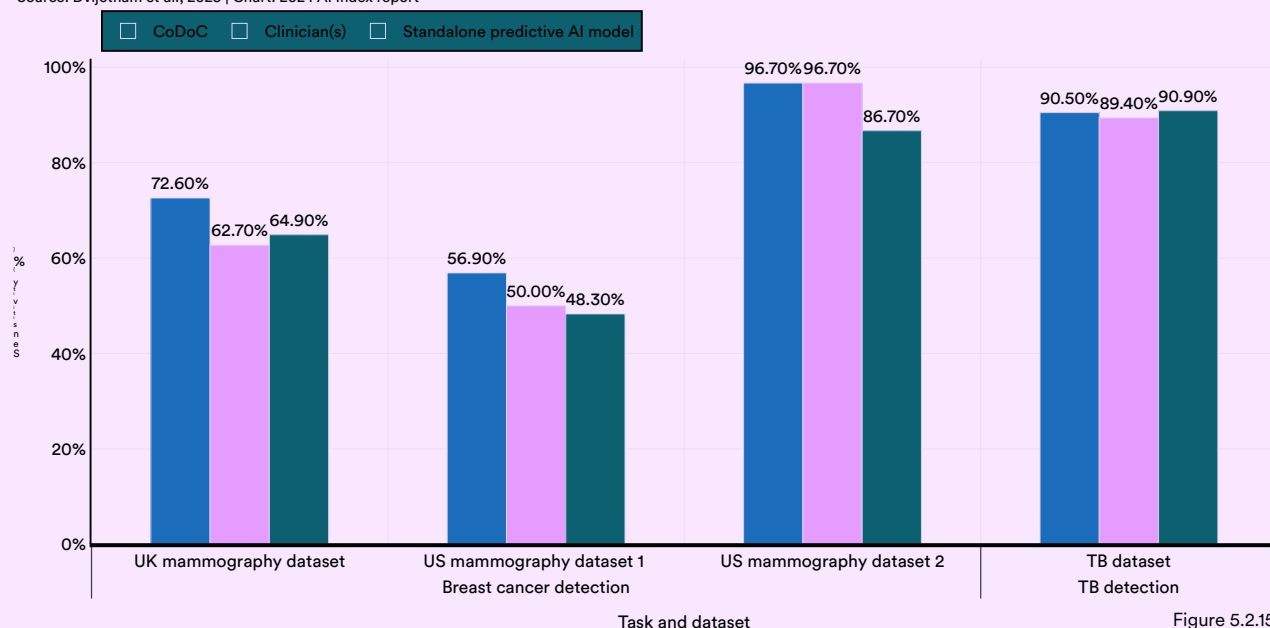
codoc

AI医療イメージングシステムは、堅牢な診断機能を示していますが、臨床医がキャッチする診断を見落とし、逆もまた同様です。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、研究者はCODOC（臨床ワークフローへの相補性駆動型の延期）を発表しました。これは、診断のためにAIに依存するタイミングと従来の臨床方法を延期するタイミングを識別するために設計されたシステムです。CODOCは、感度（病気のある人を正しく識別する能力）と特異性の両方を顕著に向上させます

（それなしでそれら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で、CODOCの感度は臨床医を平均4.5パーセントポイント上回り、スタンドアロンAIモデルは6.5パーセントポイントを超えています（図5.2.15）。特異性の観点から、CODOCは、テストされたデータセット全体で平均2.7パーセントポイント、スタンドアロン予測モデルを5.7パーセントポイント上回る臨床医よりも優れています。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。これらの調査結果は、AI医療システムを臨床ワークフローに統合し、それにより診断の精度と効率を高めることができることを示唆しています。

CODOC vs. スタンドアロン予測AIシステムと臨床読者：感度

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



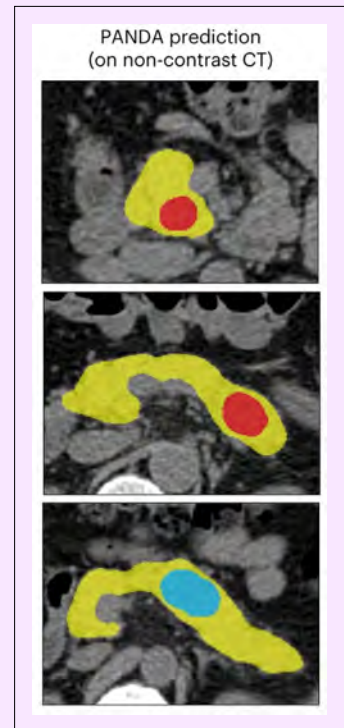
ハイライトされた研究：

CTパンダ

膵臓癌（PDAC）は特に致命的な癌であり、外科的介入には遅すぎることが多すぎることがよくあります。無症候性の個人におけるPDACのスクリーニングは、その有病率が低く、誤検知のリスクがあるため、困難です。今年、中国の研究チームは、X線の膵臓病変を効率的に検出および分類できるAIモデルであるPanda（人工知能による膵臓癌検出）を開発しました（図5.2.16）。検証テストでは、Pandaは感度の平均放射線科医を34.1%、特異性を6.3%上回りました（図5.2.17）。約20,000人の患者を含む大規模で実世界のテストでは、パンダは92.9%の感度と99.9%の特異性を達成しました（図5.2.18）。PandaのようなAIの医療ツールは、困難な状況の診断における重要な進歩を表しており、以前は困難または禁止されていたと考えられていた費用対効果が高く正確な検出を提供します。

パンダ検出

Source:
Cao et al., 2023
図5.2.16



出典：Cao et al., 2023 | チャート：2024 AIインデックスレポート Panda vs. Multicenter 検証に関する平均放射線科医 (6,239人の患者)

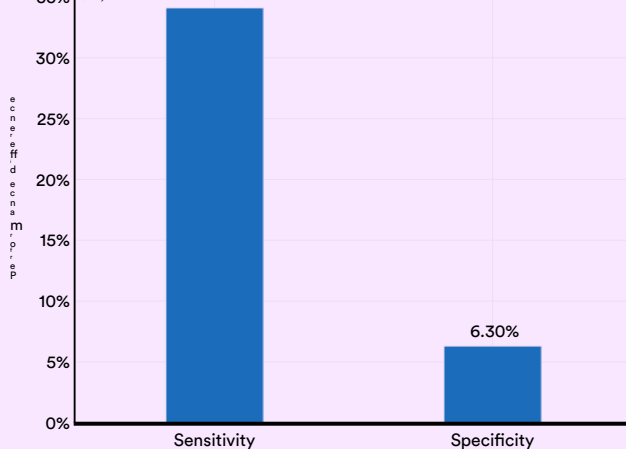


Figure 5.2.17

現実世界のマルチセナリオ検証のパンダパフォーマンス (20,530人の患者)

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

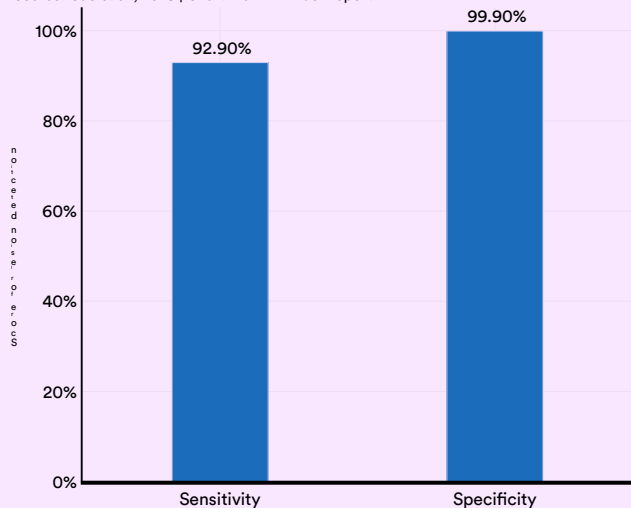


Figure 5.2.18

その他の診断用途

2023年に公開された新しい研究では、他の診断コンテキストでAIをどのように使用できるかを強調しています。図5.2.19は、調査結果の一部をまとめたものです。

診断AIユースケースに関する追加の研究

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が承認したAI関連の医療機器

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デバイスは、FDAのプレマーケット基準を満たしています。これには、その有効性と安全性の詳細なレビューが含まれています。2023年10月の時点で、FDAは生成AIを利用したり、LLMsを搭載したりするデバイスを承認していません。

図5.2.20は、過去10年間にFDAによって承認されたAI医療機器の数を示しています。2022年には、合計139のAI関連医療機器がFDAの承認を受け、2021年に承認された総額から12.1%増加しました。2012年以降、これらのデバイスの数は45倍以上増加しました。

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

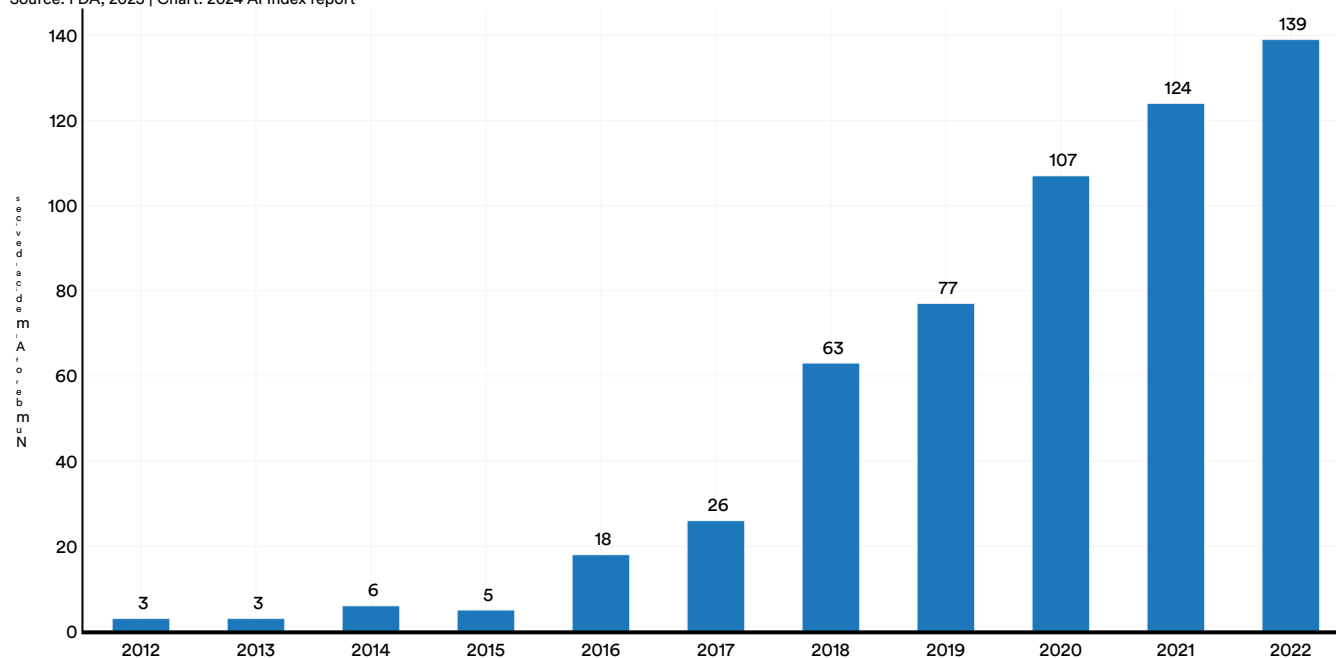


図5.2.20

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が承認した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のデバイスのうち、かなりの過半数、87.1%が放射線学に関連していました。次に最も一般的な専門は心血管で、承認の7.2%を占めています。

FDAがSpecialtyによって承認したAI医療機器の数、2012-22

出典：FDA、2023 | チャート：2024 AIインデックスレポート

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

Figure 5.2.21

管理とケア

AIツールはまた、医療管理の効率を高め、患者の標準を高める可能性を秘めています

再。

ハイライトされた研究：

メダリグ

ヘルスケアのAIの大幅な進歩にもかかわらず、知識ベースの質問に焦点を当てたMEDQAやUSMLEなどの既存のベンチマークは、臨床医が患者ケアで実行する多様なタスクを完全に把握していません。臨床医は、テラード診断計画の作成など、情報集約型のタスクに従事し、管理時間のかかなりの割合を管理タスクに費やすなど、情報集約型のタスクに従事します。AIにはこれらのプロセスを合理化する可能性があります、ベンチマークおよび微調整医学的管理LLM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、研究者はメダリグインを導入することにより、このギャップに対処するために進歩しました：包括的なEHRベース

7つの異なる医療専門分野から描かれた983の質問と指示と303の臨床医の反応を伴うベンチマーク（図5.2.22）。MedAlignは、EHRに焦点を当てた最初のベンチマークです。

その後、研究者はMedalignでさまざまな既存のLLMをテストしました。すべてのLLMのうち、MultiStep洗練を使用したGPT-4バリエーションは最高の正確性率（65.0%）を達成し、他のLLMよりも日常的に好まれました（図5.2.23）。Medalignは、AIを使用してヘルスケアの管理負担を軽減するための貴重なマイルストーンです。

MedAlign ワークフロ

— Source: Fleming et al., 2023

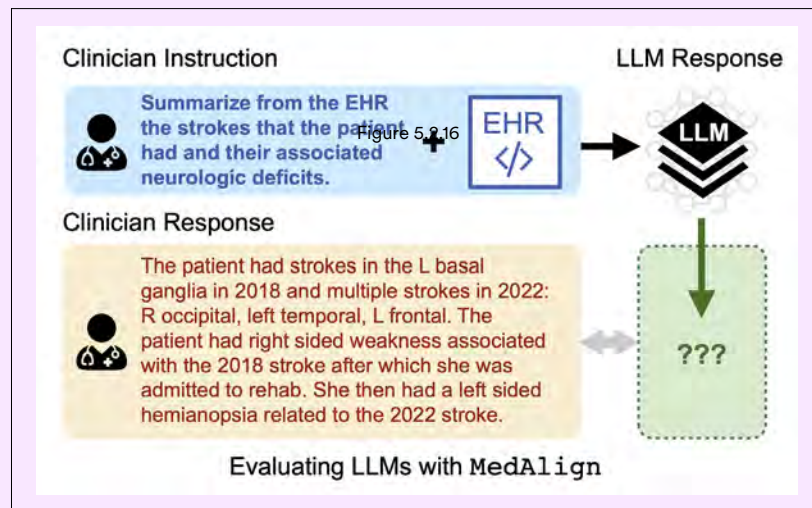


図5.2.22

ハイライトされた研究：

MedAlign（続き）

モデルのパフォーマンスの評価：人間と彗星のランク

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

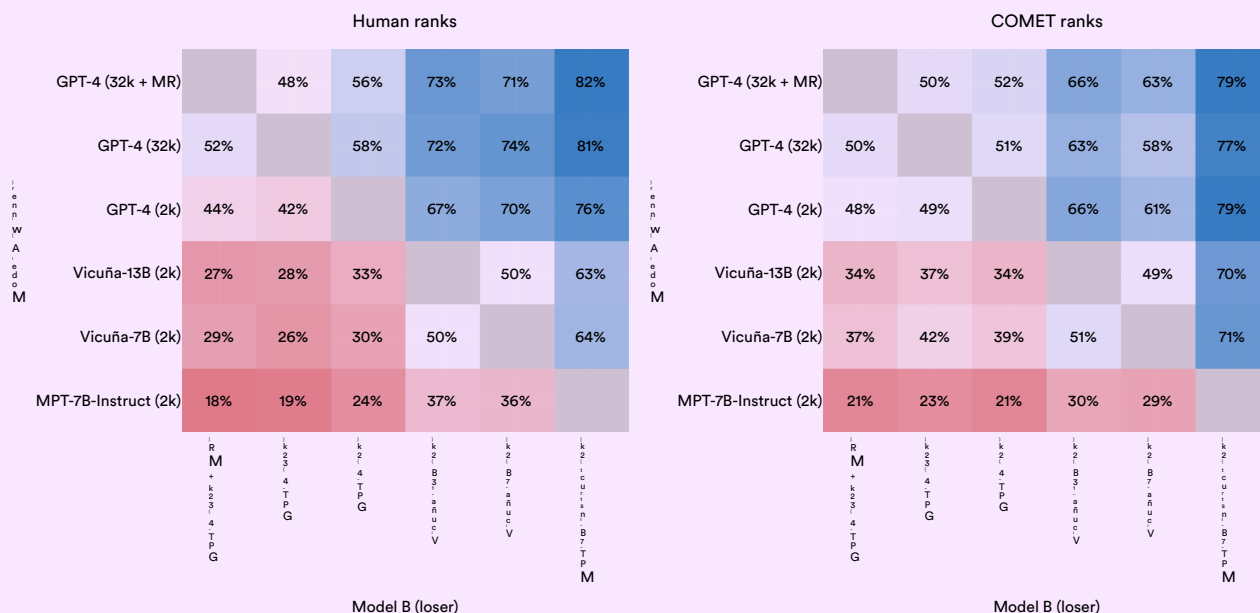
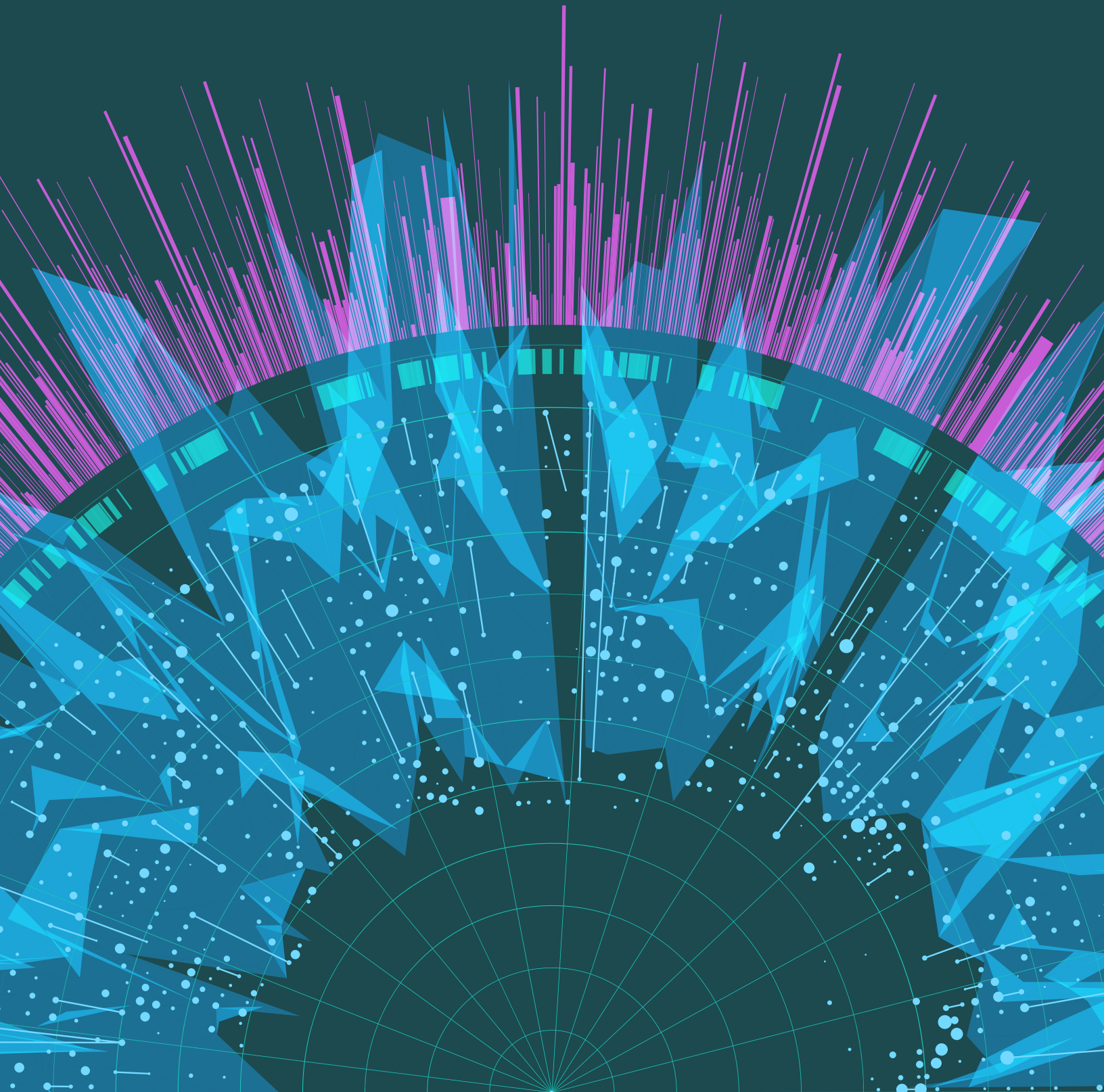


図5.2.23



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 CSおよびAI教育359米国359州レベルのトレンド
359 APコンピューターサイエンス361ハイライト：アクセ
ス問題363ハイライト：教師と学生間のChatGPTの使用364

パブリックデータにアクセスします

概要

この章では、AIおよびコンピューターサイエンス（CS）教育の傾向を検証し、誰が学習しているのか、学習している場所、およびこれらの傾向がどのように進化してきたかに焦点を当てています。AIの教育への影響に関する懸念が高まっている中で、教師や学生によるChatGPTなどの新しいAIツールの使用も調査しています。

この分析は、Computing Research Associationの年次タウルビー調査に基づいて、米国とカナダの中等後CSおよびAI教育の状態の概要から始まります。次に、ヨーロッパのCS教育に関する情報学ヨーロッパのデータをレビューします。今年は、AI関連の英語学習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を紹介します。

この章は、Code.orgからの米国のK – 12 CS教育に関する洞察と、ChatGPTの学校での使用に関するWalton Foundation Surveyからの調査結果をまとめています。

章のハイライト

1. アメリカとカナダのCS学士号の卒業生数は増え続け、新しいCS修士号の卒業生は比較的平坦であり、博士号は控えめに成長しています。一方

新しいアメリカ人とカナダの学士号の卒業生数は、10年以上にわたって一貫して上昇してきました。2018年以来、CS修士号と博士号卒業生数はわずかに減少しています。

2. AI PhDsの産業への移行は、加速ペースで継続されます。2011年には、新しいAI PhDsのほぼ等しい割合が産業（40.9%）と学界（41.6%）で雇用されました。ただし、2022年までに、卒業後、卒業後に産業に加わり、学界（20.0%）と比較して、産業に著しく大きな割合（70.7%）が参加しました。過去1年間だけで、業界に搭載されたAI PhDsのシェアは5.3パーセントポイント上昇しており、大学から業界への脳の排水が強化されています。

3. 産業から学界への学問的才能の移行が少ない。2019年、米国とカナダの新しいAI教員の13%が産業出身でした。2021年までに、この数字は11%に減少し、2022年にはさらに7%に低下しました。この傾向は、業界から学界への高レベルのAIの才能の徐々に低い移行を示しています。

4. 米国とカナダのCS教育は、国際的ではなくなりました。2021年よりも2022年よりも卒業した国際的なCS学士号、修士号、および博士号が比例して少なくなりました。修士課程の留学生の減少は特に顕著でした。

5. アメリカの高校生が増えますが、CSコースを受講しますが、アクセスの問題は残っています。2022年には、201,000のAP CS試験が実施されました。2007年以来、これらの試験を受ける学生数は10倍以上増加しています。ただし、最近の証拠は、大規模な高校や郊外の学生がCSコースにアクセスできる可能性が高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。英語の数の数は、2017年以来3倍になり、過去5年間で着実に年間増加しています。世界中の大学は、よりAI中心の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、およびIT卒業生の生産をリードしています。イギリスとドイツは、ヨーロッパを率いて、最も多くの新しい情報学、CS、CE、および情報学士号、修士号、博士号卒業生を生産しています。一人当たり、フィンランドは学士号と博士号の両方の卒業生の生産をリードし、アイルランドは修士号の卒業生の生産をリードしています。

このセクションでは、CSおよびAIにおける中等後教育の概要を説明し、学士号、修士号、博士号を含むさまざまな程度の北米およびヨーロッパ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カバーしています。

6.1 中等後CSおよびAI教育

米国とカナダ

このサブセクションでは、Computing Research AssociationのTaulbee Surveyのデータの分析を提示します。これは、米国とカナダのCSおよびAIの中等後教育の状態を評価します。この調査では、米国とカナダの297の博士課程のCS部門をカバーしています。¹

CS学士号の卒業生

過去10年間で、北米の新しいCS学士号の卒業生の総数は着実に増加し、3倍以上増加し、2021年から2022年まで前年比7.9%増加しました（図6.1.1）。

米国とカナダの新しいCS学士号の卒業生、2010–22

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

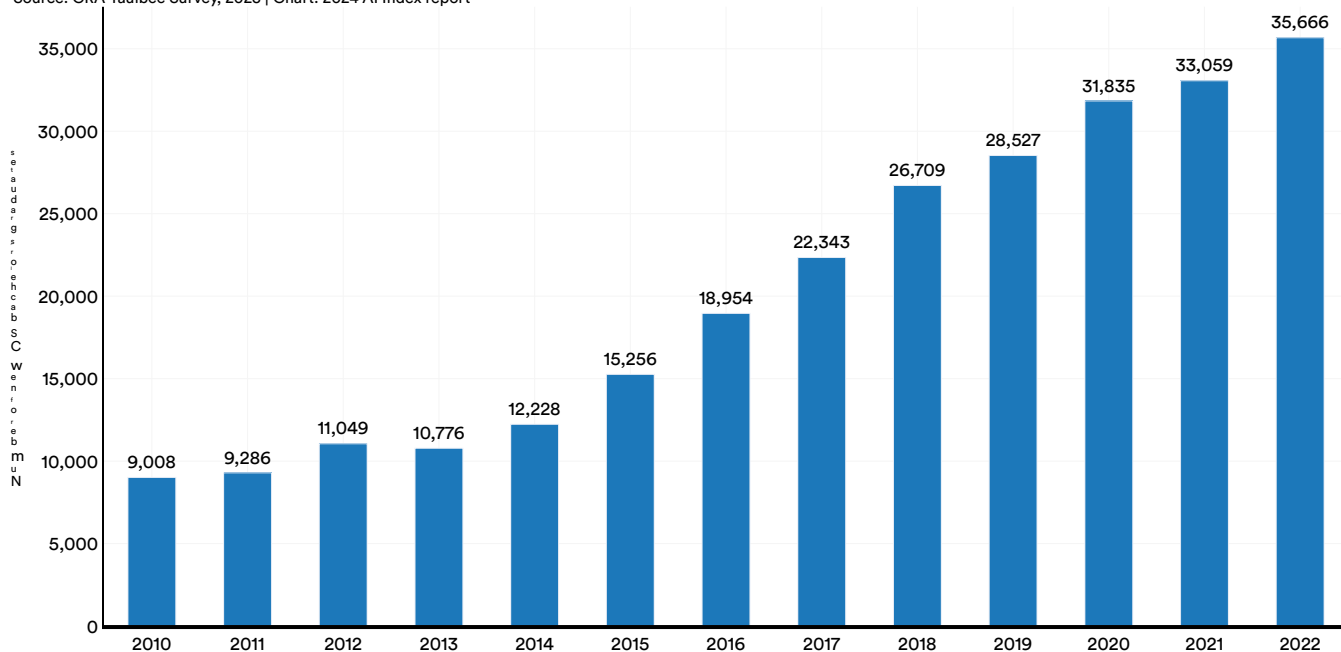


Figure 6.1.1

¹調査で標的にされたすべての博士号を取得する部門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。対象となる297の部門のうち、応答したのは182人だけで、全体的な回答率は61%でした。

ほぼ8年ぶりに、CS学士号の卒業生とカナダの大学の卒業生の留学生の割合は減少し、2021年の16.3%から2022年の15.2%に低下しました（図6.1.2）。この減少は、トランプ政権の初期に研究ビザを取得することの難しさの増加を反映している可能性があります。

データ。この減少は、Covid-19のパンデミック中に課された国際旅行制限に部分的に起因しており、留学生が米国とカナダで勉強する能力に影響を与えます。この最近の低下にもかかわらず、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。

米国とカナダの新しい国際CS学士号（合計の%）、2010–22 Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

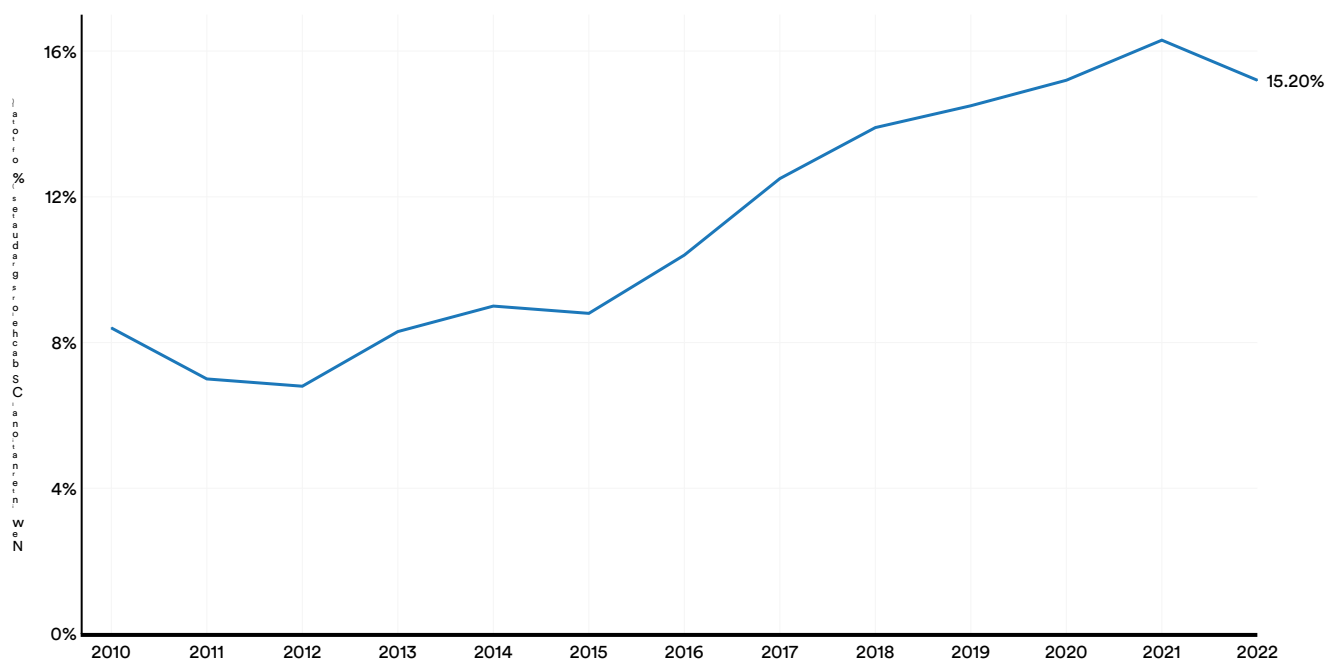


図6.1.2

CSマスターの卒業生

AIコースは一般にCS修士号プログラムに含まれています。アメリカおよびカナダの大学の新しいCS修士号の総数は、過去10年間で2倍以上になりましたが、

この数は2018年以降レベルアウトされており、昨年、わずかに2.5%減少したようです（図6.1.3）。このベリングは、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。

米国とカナダの新しいCS修士号、2010年から22年

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

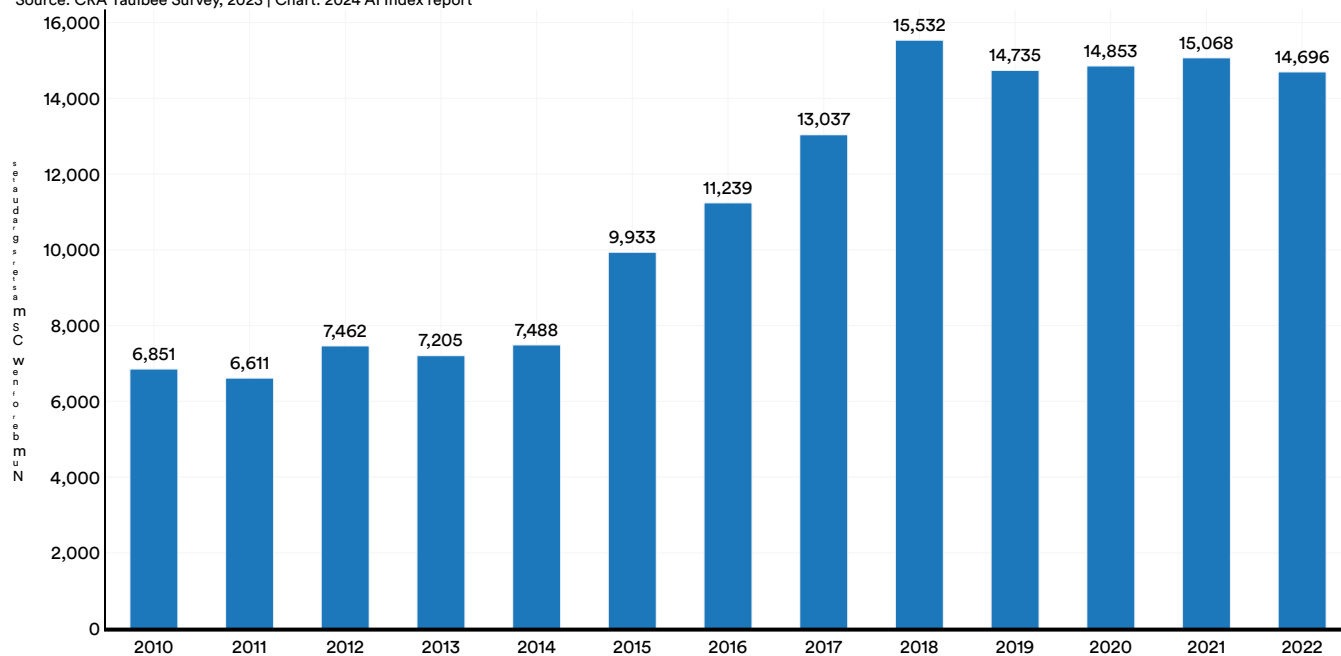


図6.1.3

2022年、アメリカとカナダの大学は、国際的なCS修士課程の学生の顕著な減少を経験しました。この下降傾向は2017年頃に始まりましたが、減少は昨年最も顕著であり、14.8パーセントポイントでした（図6.1.4）。現在、国際的なCS修士号と国内の修士号の分裂は、ほぼ均等です。

2010年から22年にかけて、米国とカナダの新しい国際CS修士号（合計の%）

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

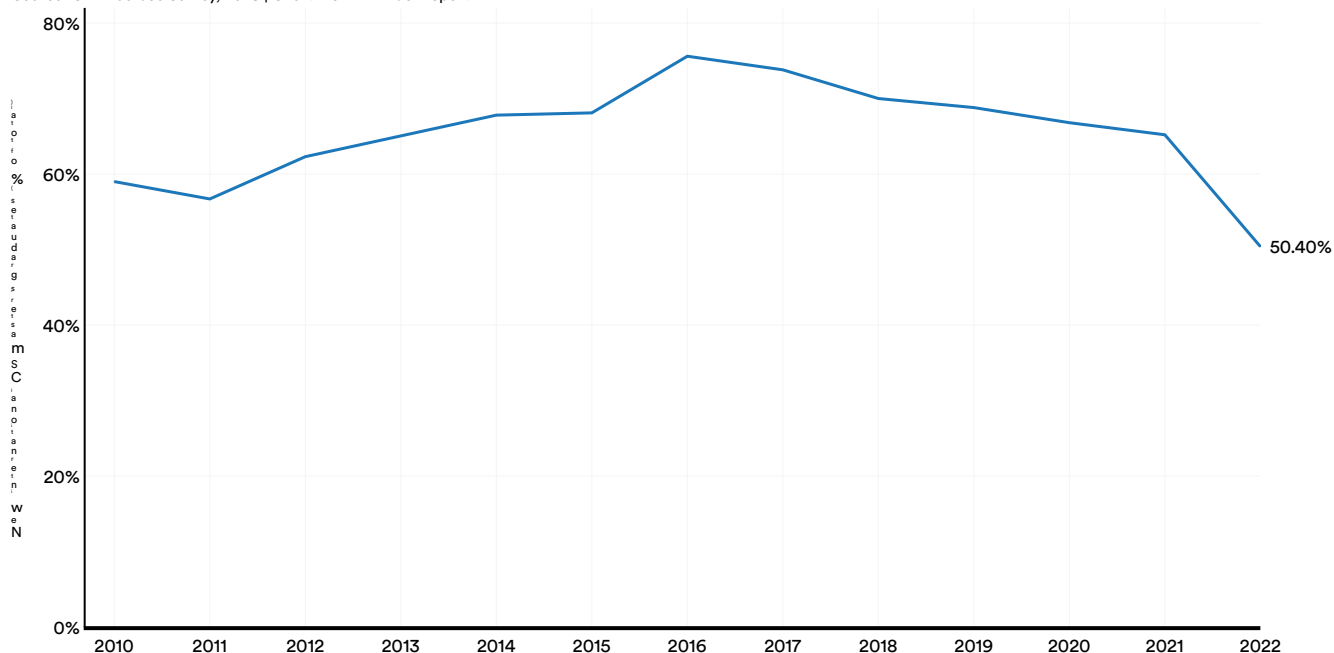


図6.1.4

CS PhD卒業生

10年ぶりに、アメリカおよびカナダの大学で新しいCS PhD卒業生の数が大幅に増加しています。2022年、CS PhD卒業生の数は2,105に達し、2010年以来最高になりました（図6.1.5）。

米国とカナダの新しいCS PhD卒業生、2010年から22年

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

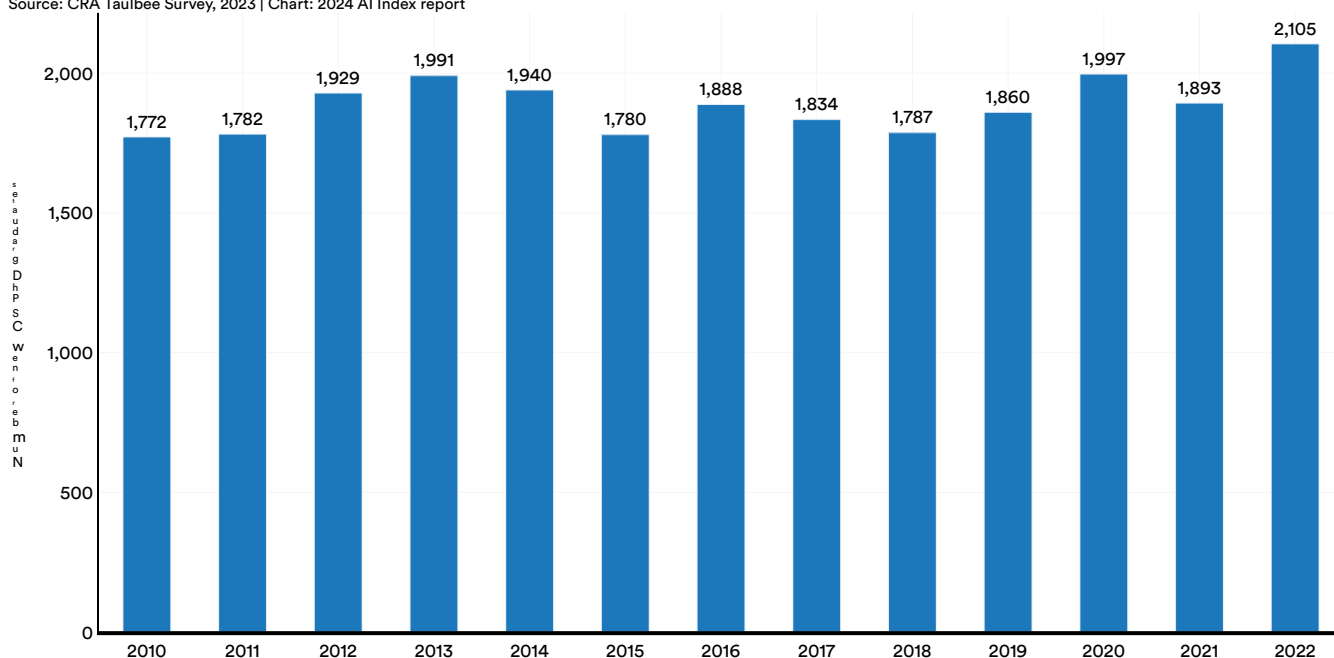


図6.1.5

CS PhD卒業生の留学生の割合は過去10年間で上昇していますが、昨年はこの割合がわずかに減少し、2021年の68.6%から2022年の65.9%に減少しました（図6.1.6）。

2010年から22年にかけて、米国とカナダの新しい国際CS PhD卒業生（合計の%）

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

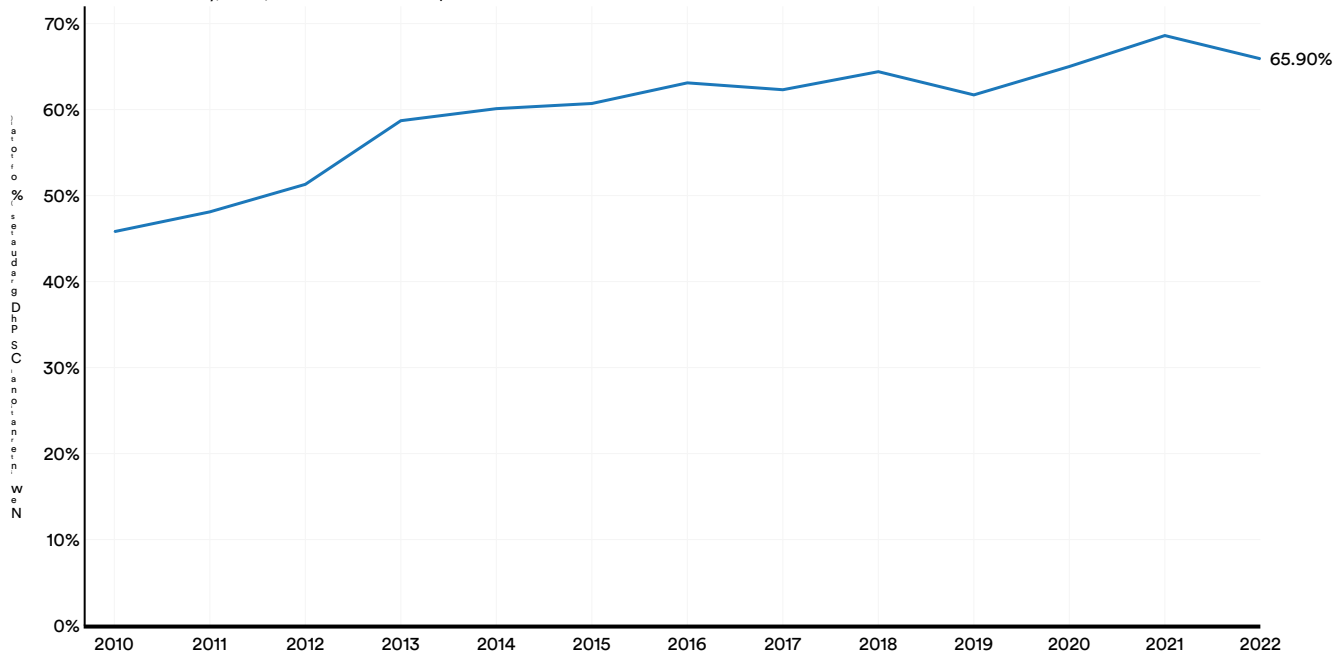


図6.1.6

新しく造られたAI PhDsは、卒業後にどこで働くことを選択しますか？ 昨年のAIインデックスレポートで強調された傾向に続いて、AI博士課程の受信者のシェアの増加が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、同じ割合が学界（41.6%）と同じように産業で雇用を得ました（40.9%）。

ただし、2022年までに、卒業後、卒業後に産業に加わり、学界（20.0%）と比較して、産業に著しく大きな割合（70.7%）が参加しました。政府の役割に就く新しいAI PhDの割合は、過去5年間で約0.7%で比較的低く、安定しています。

2010年から22年にかけて、米国およびカナダでの新しいAI PhDs（合計の%）の雇用、2010–22

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

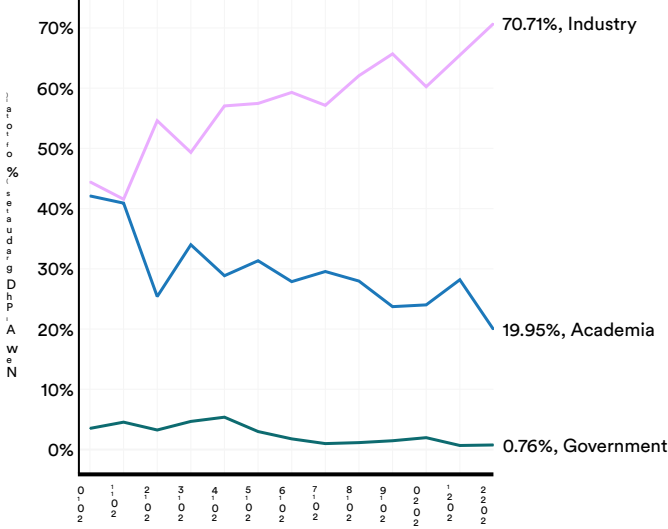


Figure 6.1.7²

米国での新しいAI PhDの雇用と
セクターによるカナダ、2010–22

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

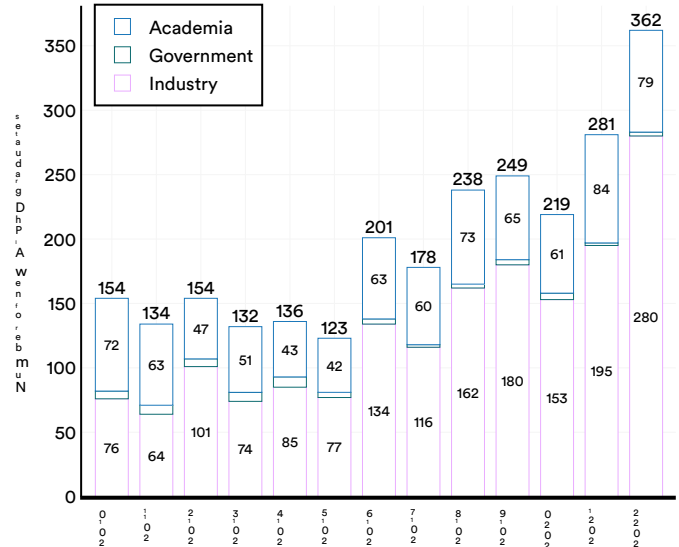


Figure 6.1.8

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.

CS、CE、および情報学部

CSおよびAI教育の傾向をよりよく理解するために、CSの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のCS、CE、および情報学部の総数は7.2%増加しました（図6.1.9）。2011年以降、増加は42.4%です。

米国とカナダのCS、CE、および情報学部の数、2011-22

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

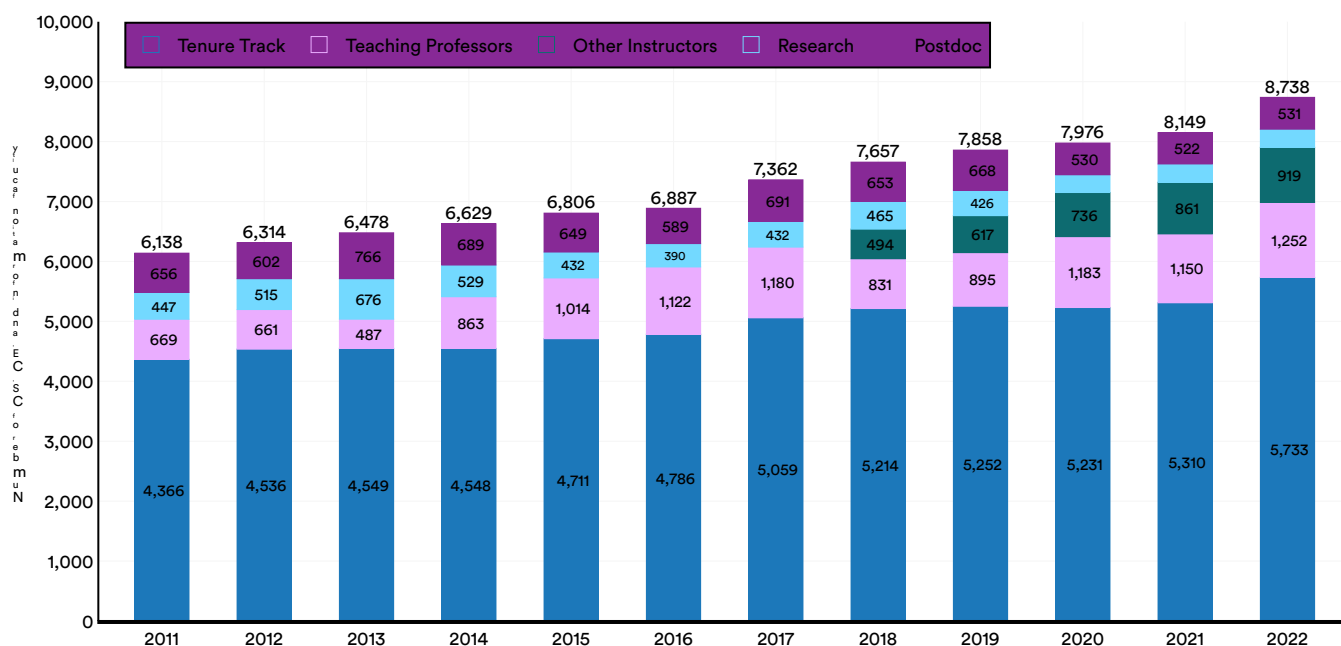


図6.1.9

2022年、米国には7,084人のCS教員があり、過半数（65.7%）がテニユアトラックに載っていました（図6.1.10）。アメリカのCS教員の総数は、2021年以降4.4%、2011年以降45.0%上昇しています。

2011年から22年の米国のCS教員の数

出典：CRA Taulbee Survey、2023 | チャート：2024 AIインデックスレポート

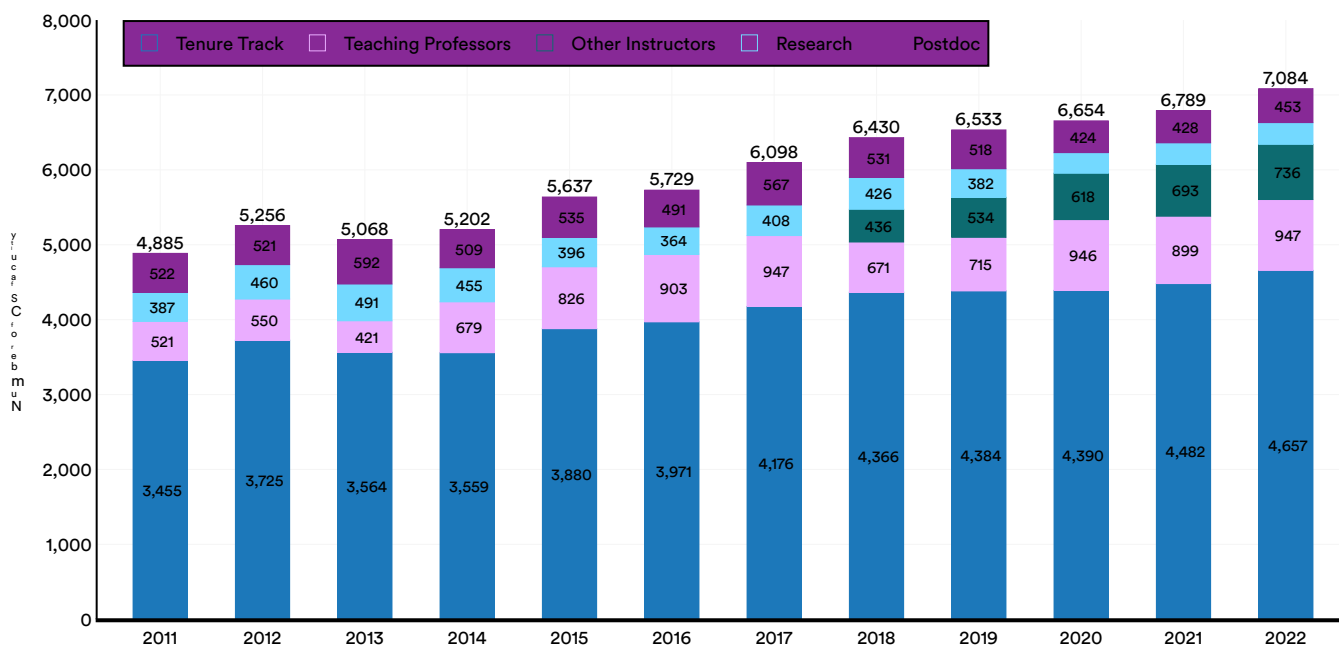


図6.1.10

昨年、北米のCS、CE、および情報分野で915人の新しい教員が雇用されました。これらのポジションの455はテニュアトラックでした。（図6.1.11）。

2011年から22年の米国とカナダで新しいCS、CE、および情報学部が採用されています

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

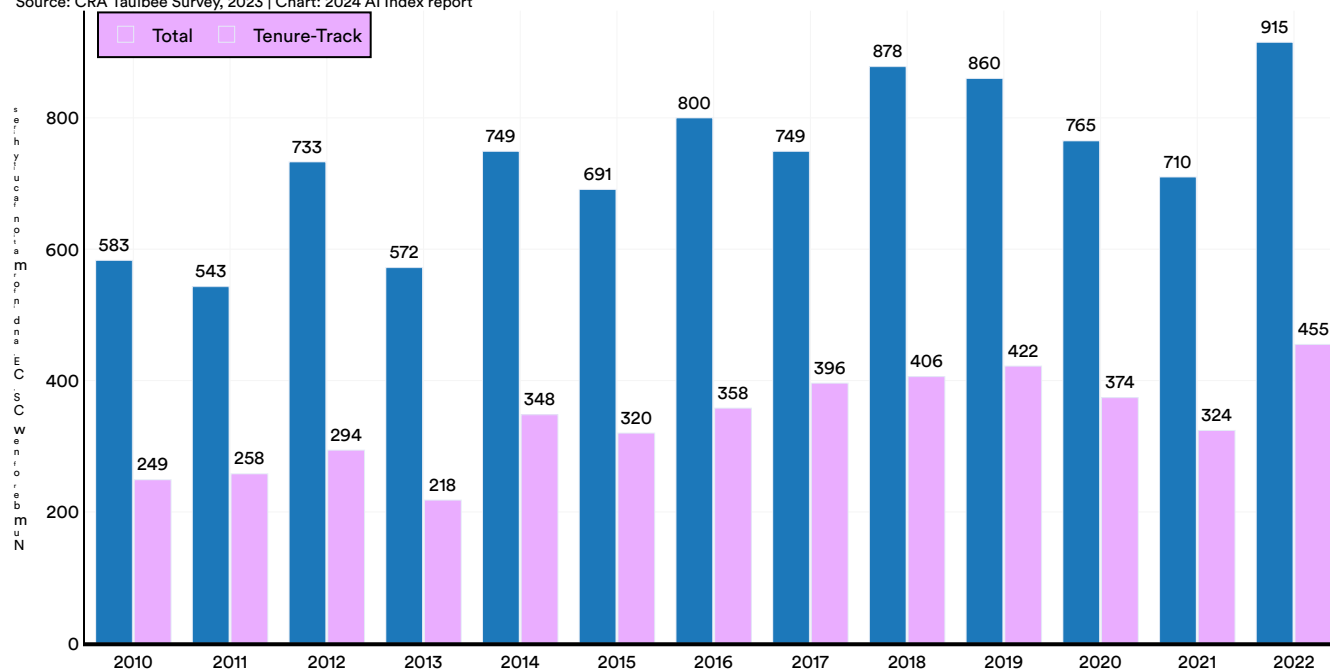


図6.1.11

2022年には、新しい教員の任命の43%が他の学問的地位から来ており、学術労働力内の「解約」を示しています（図6.1.12）。これらの「新しい」教職員は他の場所でポジションを空けたため、最終的に彼らの以前の役割を満たす必要があります。さらに、2022年に業界から移行している教員の割合は、前年の11%から7%、2019年に13%に減少しました。

アメリカおよびカナダのCS、CE、および情報部門の新しい教員、2018–22

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

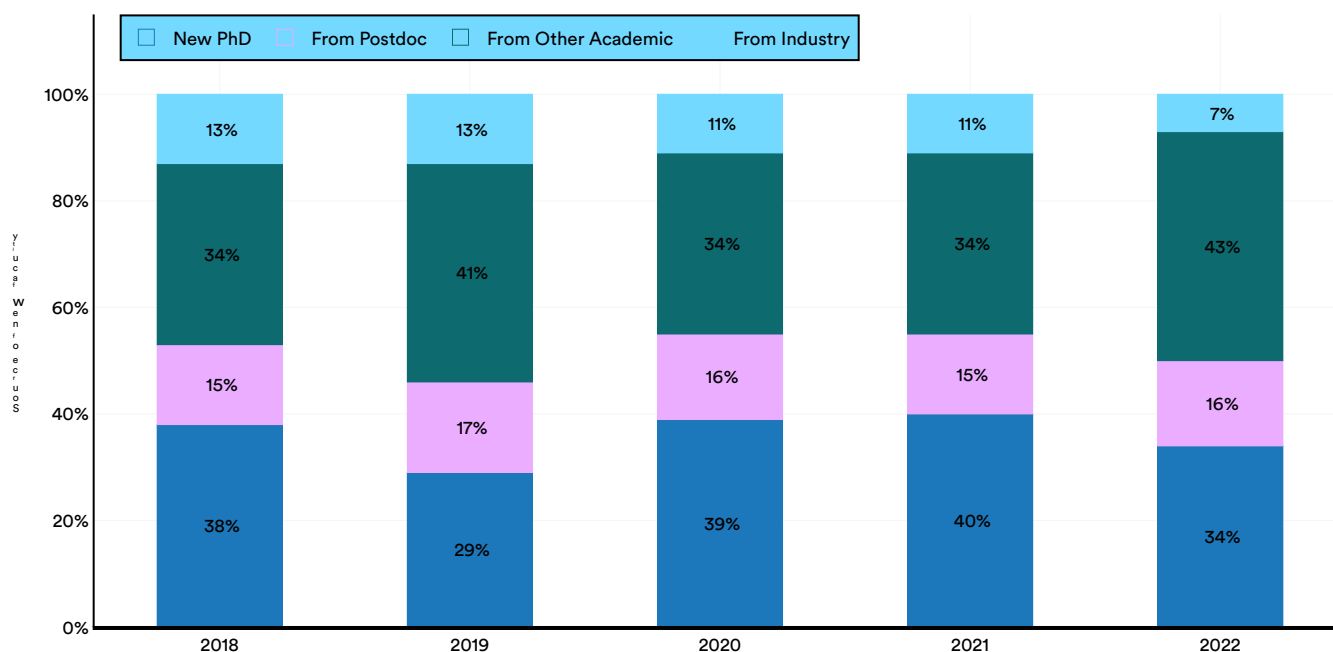


図6.1.12

教員の職位が埋められないままである理由は、過去10年間でさまざまでした。2011年には、検索の失敗の37%がオファーが行われなかったため、34%はオファーが拒否されたためでした（図6.1.13）。対照的に、2022年には、オファーが行われずに15%しか終了しませんでした、55%がそれを提供しています

断られました。この傾向は、新しいCS教員のますます競争の激しい市場を反映しているようです。ただし、これが他の学問的地位との競争の高まりを示しているのか、それとも業界の職位との競争を示しているのかは不明のままです。

新しいCS、CE、および情報学部役職が不明なままである理由（合計の%）、2011年から22年

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

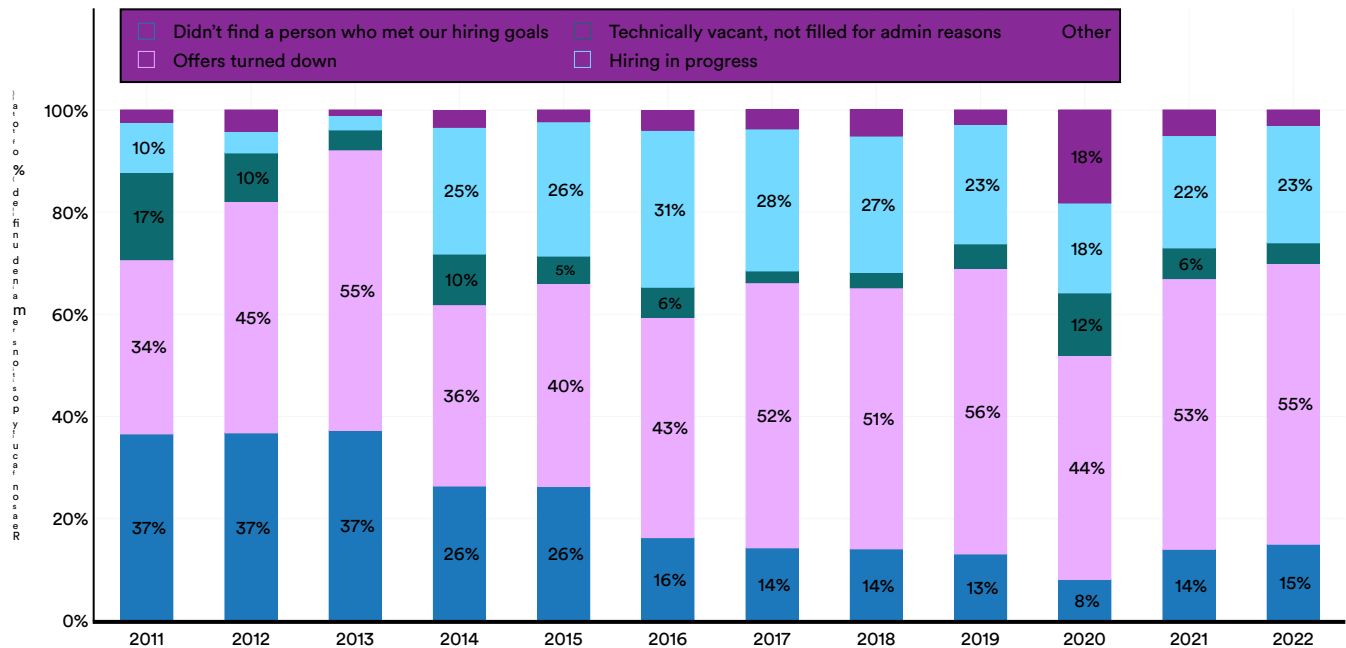


図6.1.13

2022年、CS、CE、および情報分野の北米の部門は、2021年の303と比較して、合計405の教員の出発の大幅な増加を経験しました（図6.1.14）。これらの損失のうち、38.5%が他の学位に残っていますが、16.3%が非学術的な役割に移行し、過去と一致する傾向を維持しています。

アメリカおよびカナダのCS、CE、および情報部門の教員の損失、2011–22

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

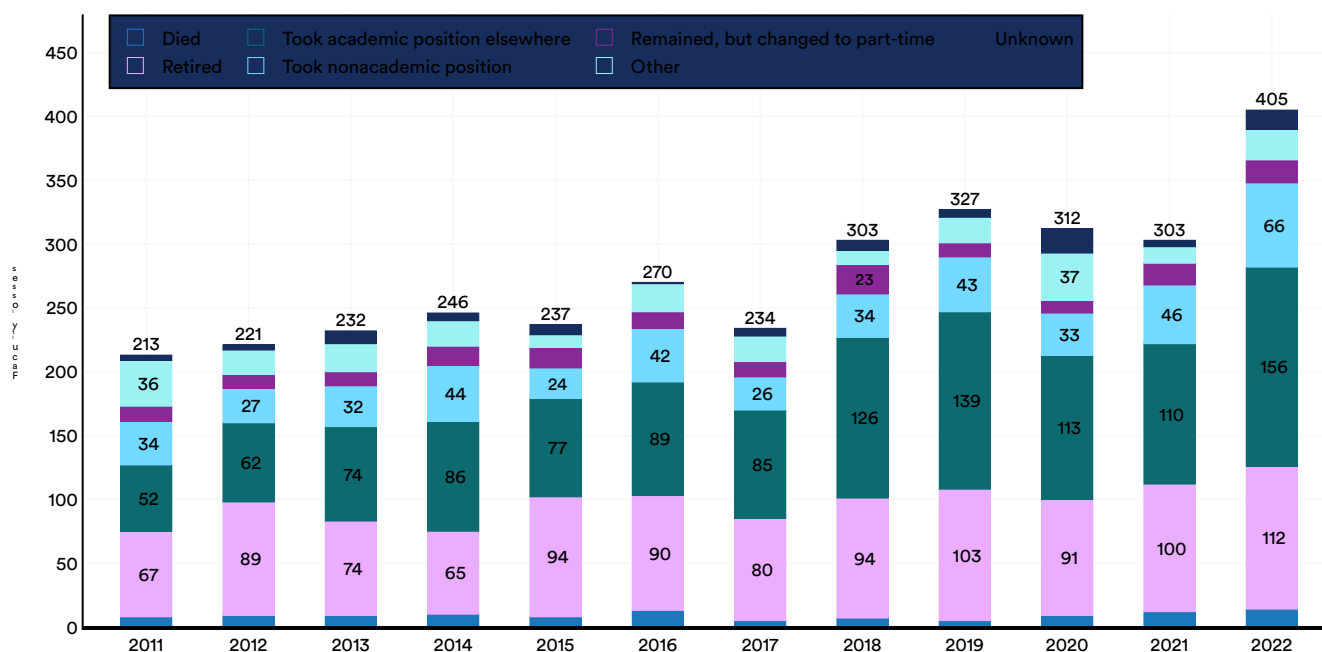


図6.1.14

2015年以降、完全教授の9か月の給与の中央値の増加は、米国のインフレ率をわずかに下回っていますが、助教授と准教授の給与の中央値はインフレをわずかに上回っています。2022年には、完全な教授の給与は2021年より3.2%高く、米国の7%のインフレ率に対応し、2015年よりも16.4%高くなり、その年の19%の増加をまだ下回っています（図6.1.15）。

2015年から22年にかけて、米国のCS教員の9か月の給与の中央値

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

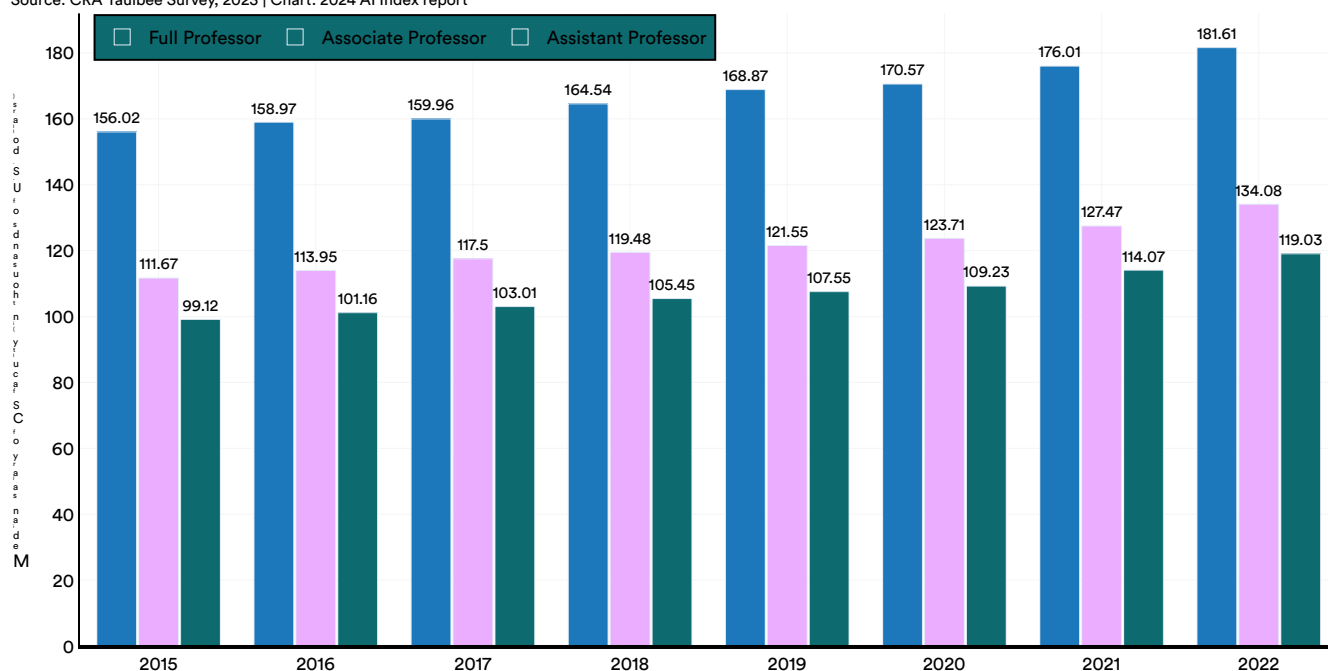


図6.1.15

2022年、CS、CE、および情報分野の新しいテニュアトラック教員の間での国際雇用の割合は、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年間で記録された2番目に高い割合をマークし、2013年までにしか上回っていませんでした。

2010年から22年にかけて、米国とカナダの新しい国際CS、CE、および情報テニュアトラックの教員採用（合計の%）

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

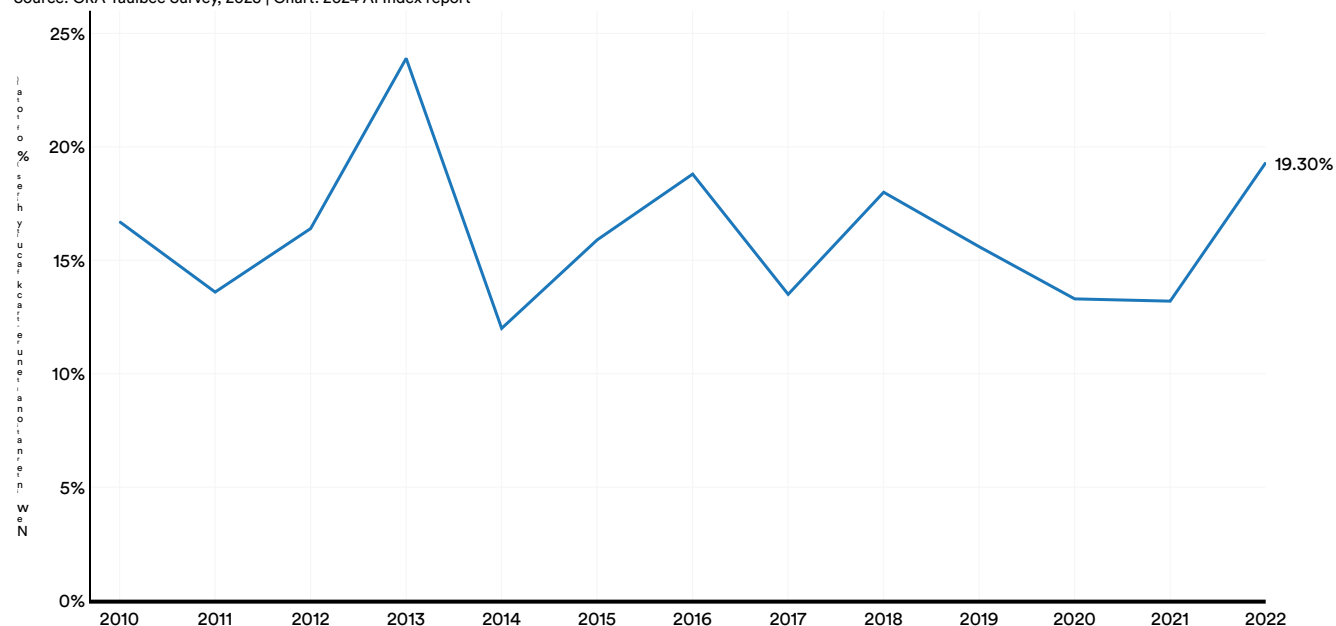


Figure 6.1.16

ヨーロッパ

ヨーロッパのCS卒業生に関するデータは、他の目標の中でもヨーロッパの情報学教育の状態を監視する学術および研究コミュニティである情報学ヨーロッパから来ています。

情報学、CS、CE、およびIT学士号の卒業生

2022年、イギリスは、情報学、CS、CE、およびITの新卒者の数が多く、学士号を取得し、合計約25,000人（図6.1.17）。⁵ドイツとトルコは密接に続きました。サンプルのほとんどの国は、ポーランド、スペイン、チェコ共和国などの例外がありましたが、10年前と比較してこれらの分野の卒業生が増加していました（図6.1.18）。

新しい情報学、CS、CE、およびそれはヨーロッパの国ごとの学士号、2022

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

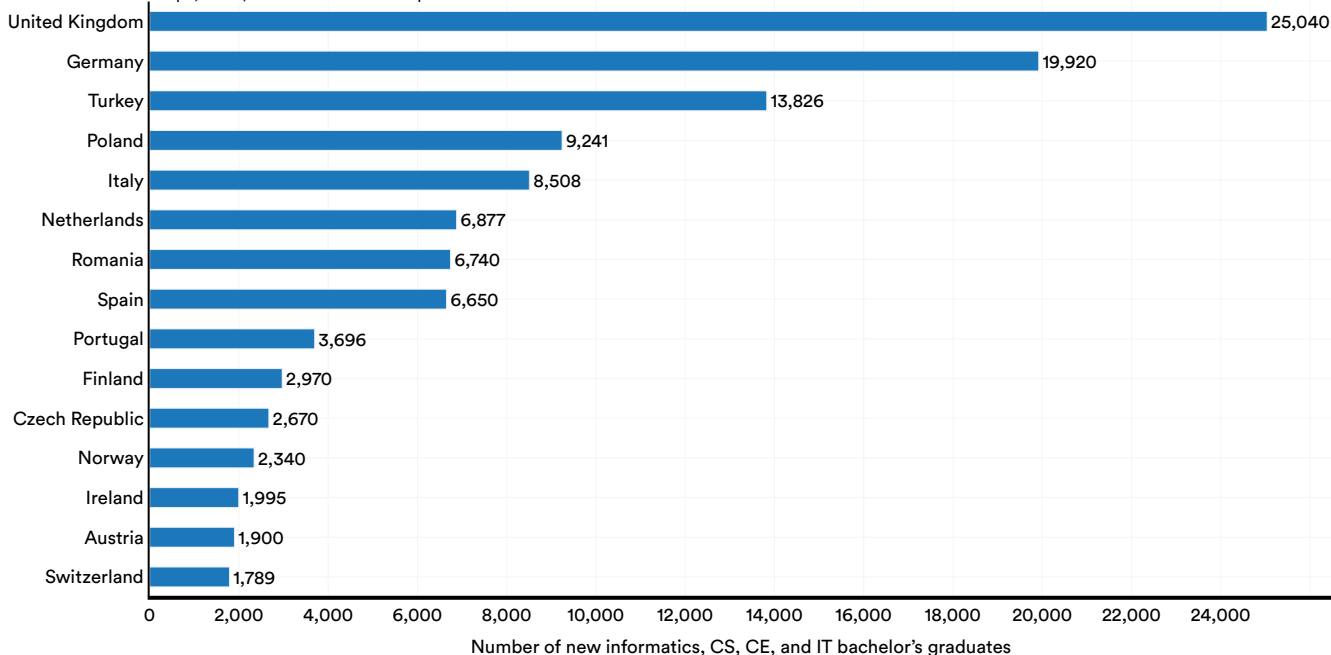


図6.1.17

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人の読者は、欧州のデータが国家統計事務所から収集され、より広範なカバレッジを誇っているため、CRA北米データと欧州のCS大学院データとの間の一人当たり比較を行うことに注意されています。

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.

新しい情報学の変化、CS、CE、およびIT学士号ヨーロッパの卒業生、2012年対2022年

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

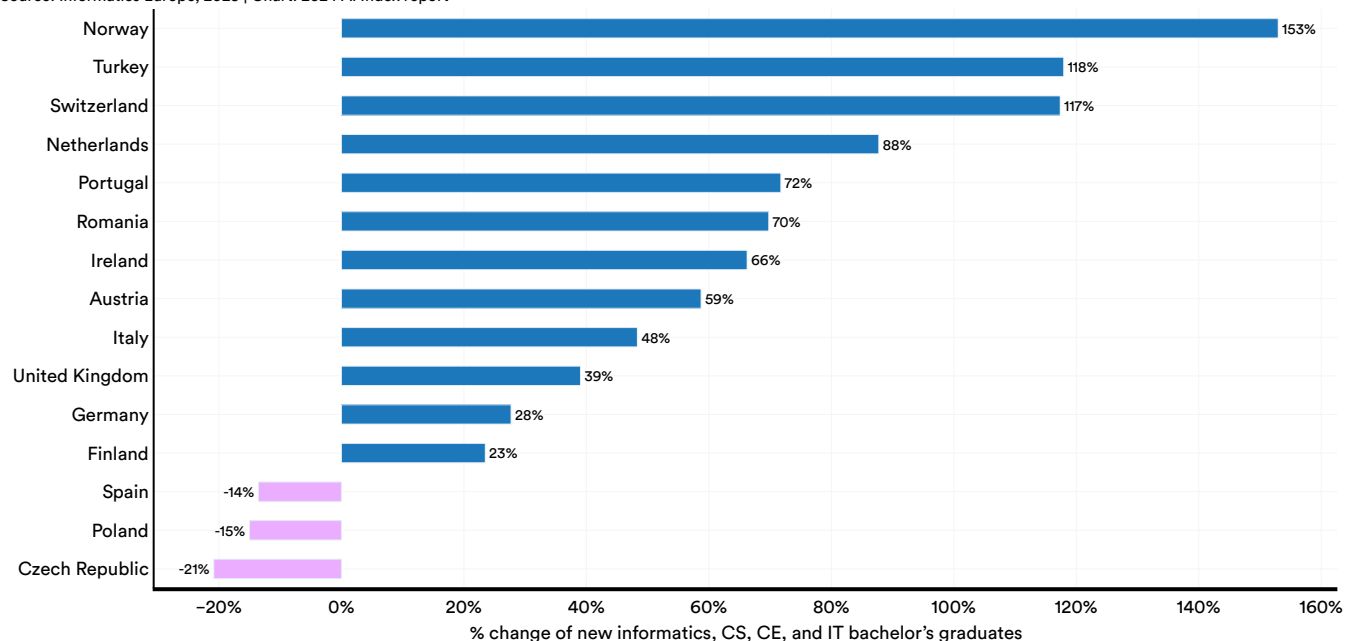


図6.1.18

フィンランド（53.4）、ノルウェー（42.6）、およびオランダ（38.6）は、情報学CS、CE、およびITにおける新しい学士号の卒業生の数をリードし、100,000人の住民あたりIT（図6.1.19）。一人当たり、サンプリングされたほとんどのヨーロッパ諸国では、情報学の総数、CS、CE、およびIT学士号の卒業生が増加しています（図6.1.20）。