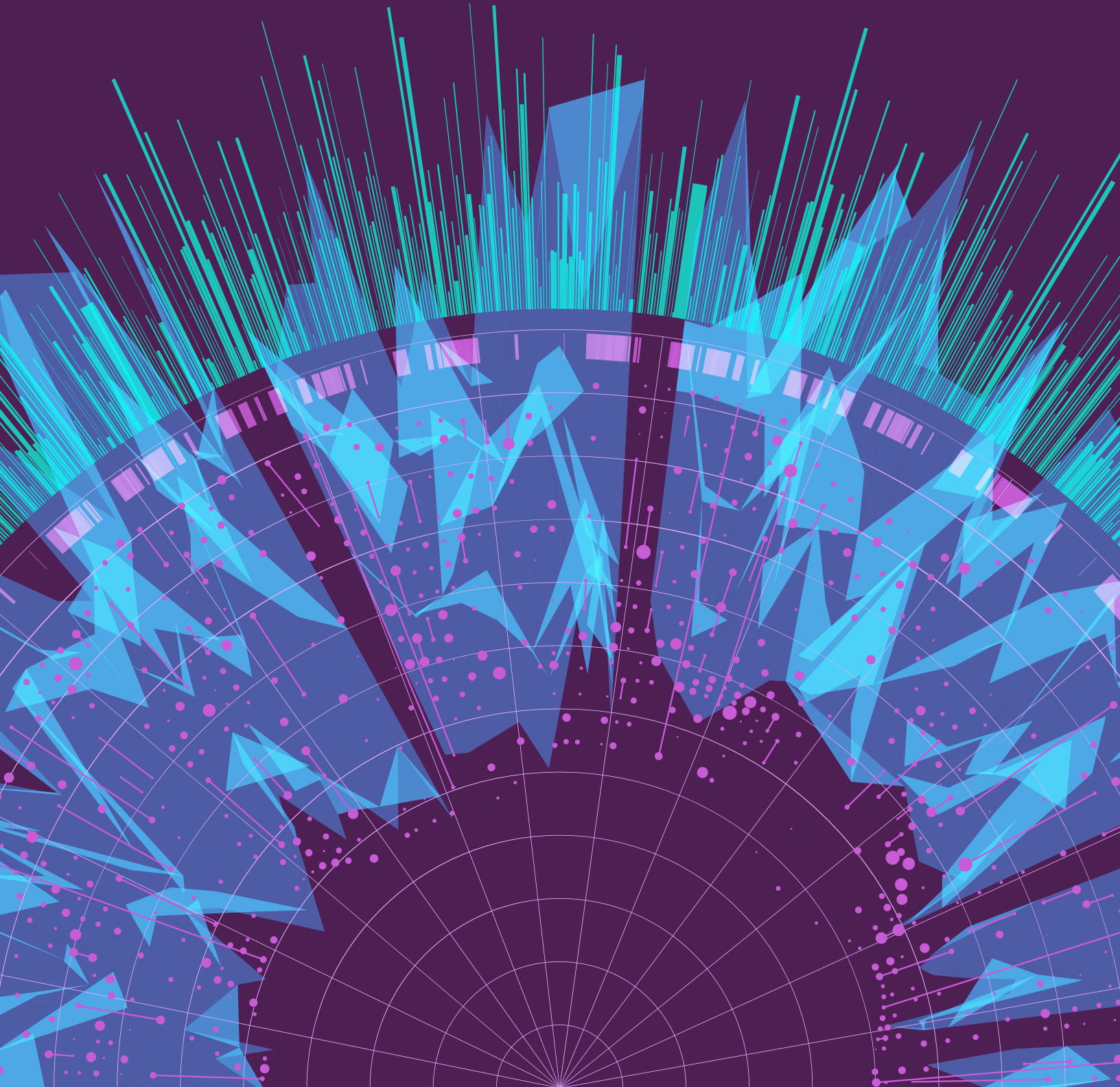




Artificial Intelligence
Index Report 2024

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

公開データへのアクセス

概要

今年のAI Indexは、科学と医学におけるAIの役割の増大を認識し、科学と医学におけるAIに関する新しい章を導入しています。これは、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など、2023年のAIが促進した優れた科学的成果を探っています。この章では、医療AIシステムのパフォーマンス、SynthSRやImmunoSEI RAのような2023年のAI駆動による重要な医療イノベーション、およびFDAのAI関連医療機器の承認の傾向についても考察しています。

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る。2022年には、AIが科学的発見を促進し始めた。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容易にするGNoMEまで、さらに重要な科学関連のAIアプリケーションがローンチされた。

2. AIは医療の飛躍的な進歩を支援しています。2023年には、パンデミック予測を強化するEV Escape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は医療の進歩を促進するためにますます利用されています。

3. 高度な知識を持つ医療AIが登場した。過去数年で、AIシステムはMedQAベンチマークにおいて著しい進歩を見せており、これはAIの臨床知識を評価するための重要なテストである。2023年の傑出したモデルであるGPT-4 Medpromptは、90.2%の精度を達成し、2022年の最高スコアから22.6ポイント向上した。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倍になった。

4. FDAはますます多くのAI関連医療機器を承認している。2022年、FDAは139件のAI関連医療機器を承認し、2021年から12.1%増加した。2012年以降、FDAが承認したAI関連医療機器の数は45倍以上に増加している。AIは、現実世界の医療目的でますます使用されている。

本セクションでは、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取り上げます。

5.1 重要な科学的マイルストーン

AlphaDev

AlphaDevがより高速なソートアルゴリズムを発見

AlphaDevは、計算アルゴリズムの向上分野における科学者や技術者数十年の研究成果を改良した、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。{v*}

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、Sort 5（図5.1.1）など。AlphaDevによって発見された新しいアルゴリズムの一部は、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アップデートであり、強化学習を用いて設計された最初の追加機能です。

アルゴリズムの長さ Source: Mankowitz et al., 2023 | Chart: 2024 AI Index report を最適化する場合のAlphaDev対人間のベンチマーク

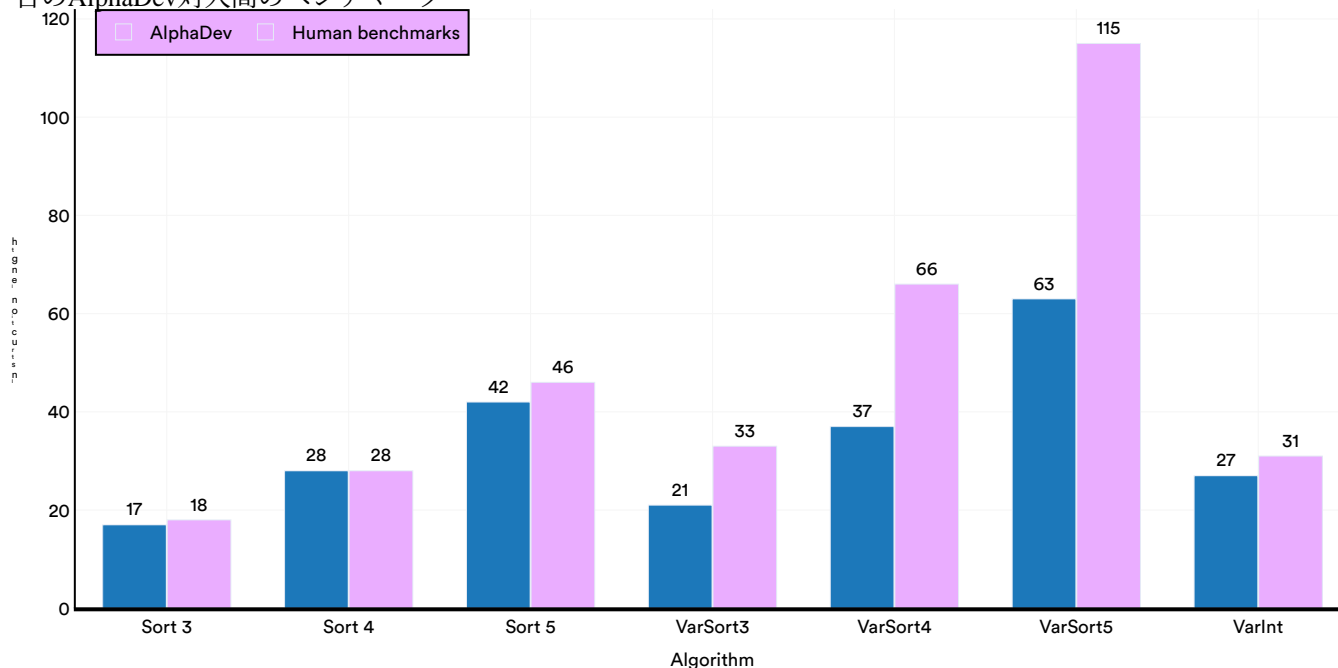


Figure 5.1.1

フレキシキューブ

FlexiCubesを用いた3Dメッシュ最適化

コンピュータグラフィックスにおいて極めて重要な3Dメッシュ生成は、3Dオブジェクトを定義するために頂点、辺、面からなる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。従来のアイソサーフェース抽出アルゴリズムは、解像度の低さ、構造的な硬直性、数値的な不安定性などに悩まされており、その結果、{v*}に影響を与えます。

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を用いることで、これらの限界の一部に対処しています。この手法により、正確で局所的なメッシュ調整が可能になります。メッシュ再構成に微分可能なアイソサーフェッシングを用いる他の主要な手法と比較して、FlexiCubesは、基礎となる真実になるべく一層適合したメッシュ抽出を実現します（図5.1.3）。

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

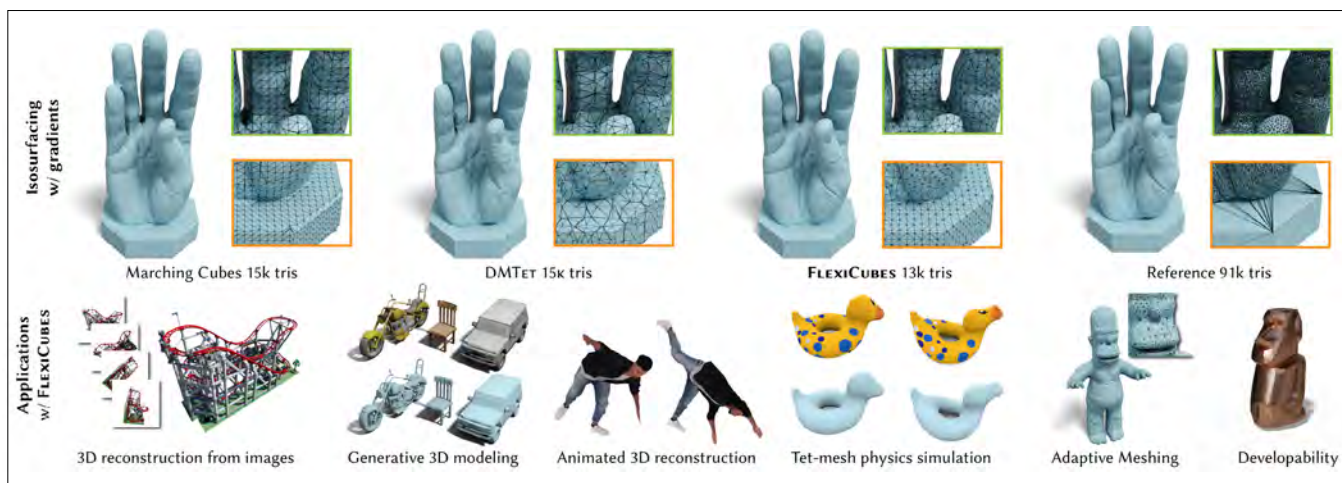


図5.1.2

3Dメッシュ再構成に関する定量的結果の選択

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

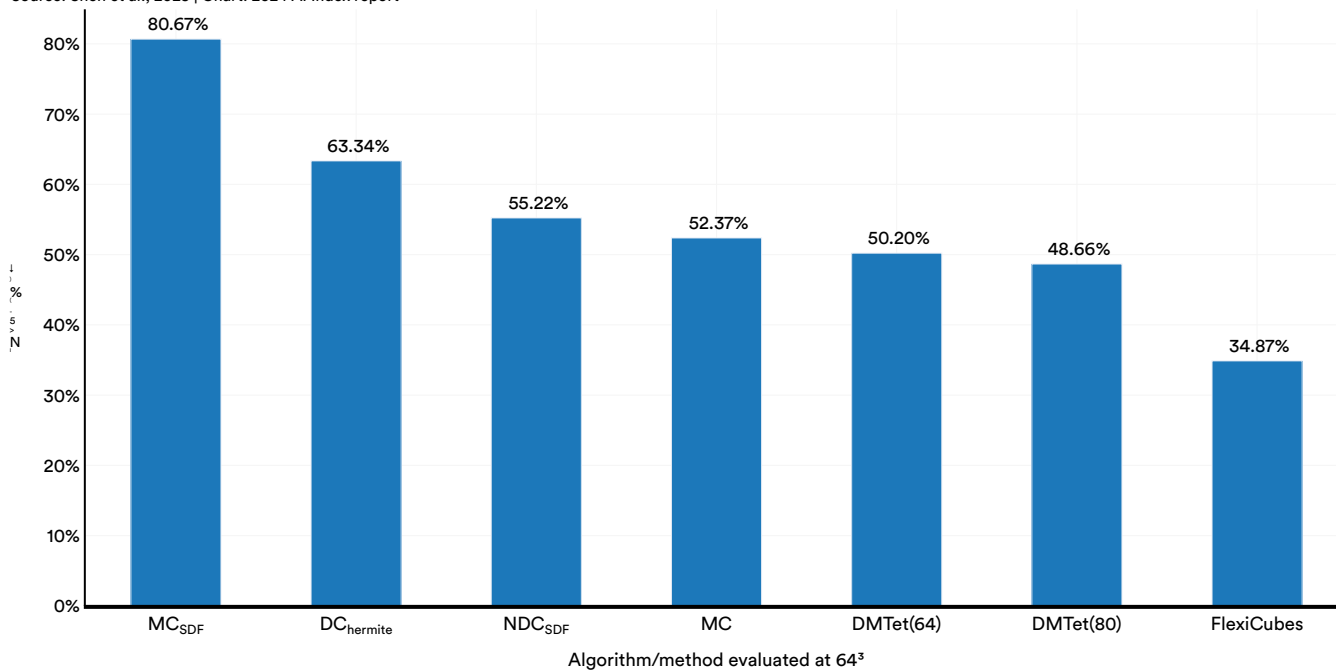


図5.1.3

シンボット

AI駆動型有機分子合成ロボット化学者

Synbotは、化学合成計画のためのAIソフトウェア層、コマンド翻訳のためのロボットソフトウェア層、実験実施のための物理的ロボット層からなる多層システムを採用しています。AIとロボットシステム間の閉ループフィードバックメカニズムにより、Synbotは、確立された基準と同等またはそれ以上の収率で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において、Synbotは、変換収率を上回る複数の合成式を開発しました。

Synbotデザイン

Source: Ha et al., 2023

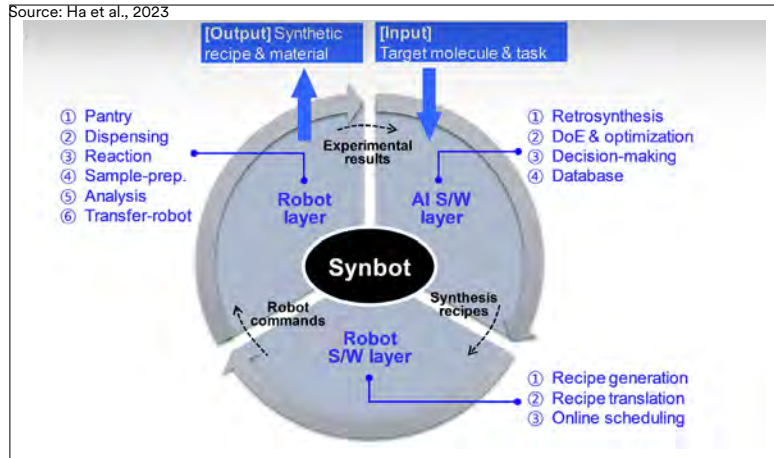


図5.1.4

中央80%の基準範囲であり、合成を大幅に短縮しました（図5.1.5）。Synbotによる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の可能性を浮き彫りにしています。

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

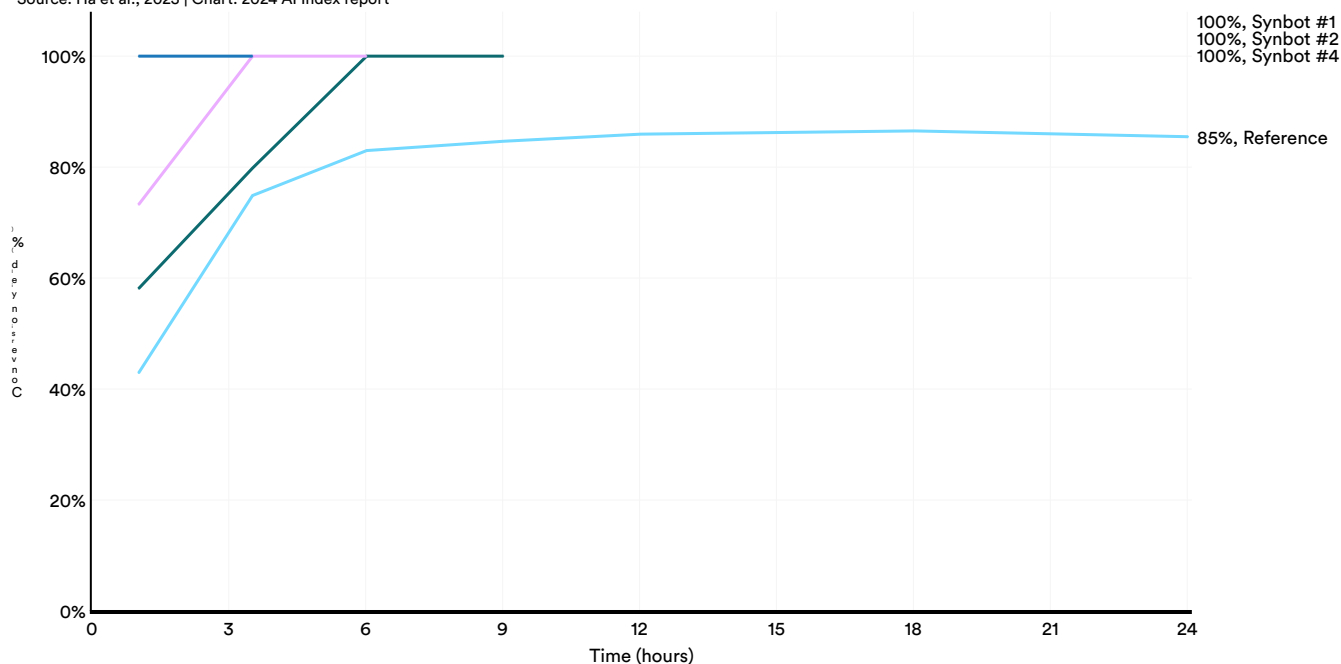


図5.1.5

グラフキャスト

より正確な全球気象予報

GraphCastを用いて

GraphCastは、1分以内で10日間の天気予報を非常に正確に提供する新しい天気予報システムです（図5.1.6）。グラフニューラルネットワークと機械学習を利用して、GraphCastは膨大なデータセットを処理し、気温、風速、大気状態を予測します。

など。図5.1.7は、GraphCastのパフォーマンスと、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）を比較しています。GraphCastは、より低い二乗平均平方根誤差を示しており、その予測は観測された気象パターンにより正確に一致することを意味します。GraphCastは、気象パターンの解読、極端な気象現象への備えの強化、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

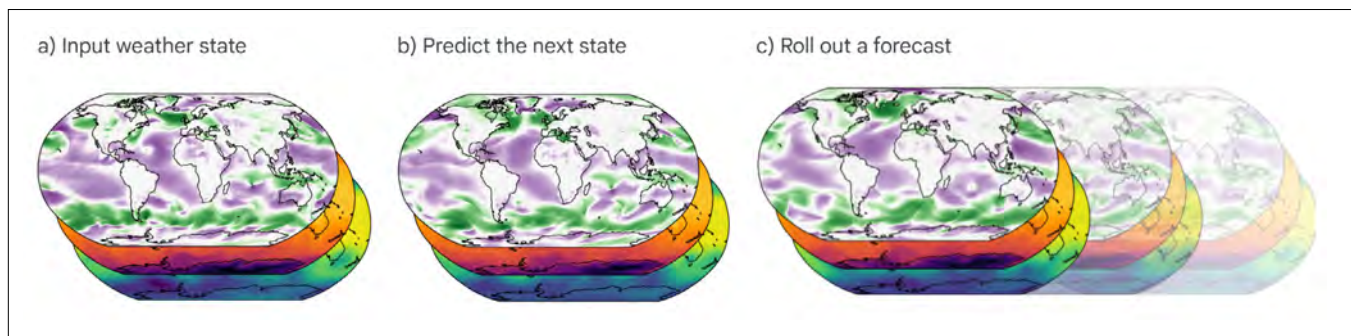


Figure 5.1.6

10日間のz500予報スキル：GraphCast対HRES

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

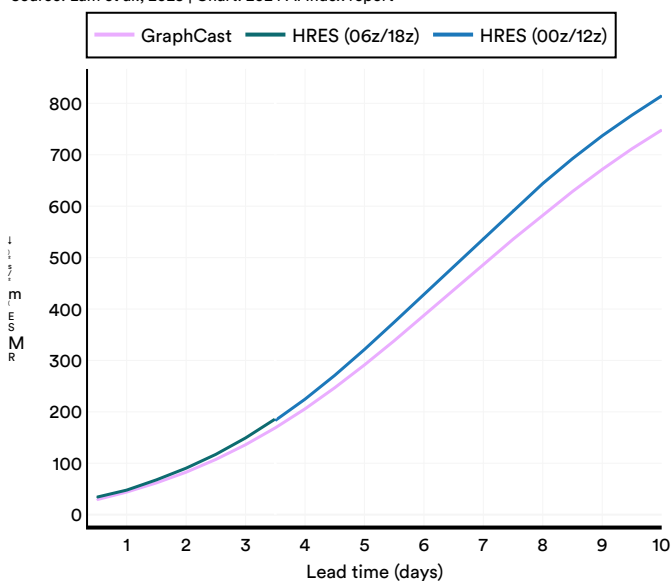


Figure 5.1.7

GNoME

GNoMEを用いた新材料の発見

様々な科学分野、特にロボット工学や半導体製造における進歩には、新しい機能性材料の探索が不可欠です。しかし、この発見プロセスは通常、費用がかかり、時間がかかります。最近のGoogleの研究者による進歩により、AIモデルの一種であるグラフネットワークは、大規模データセットで学習させた場合、このプロセスを迅速化できることが示されました。彼らのモデルであるGNoMEは、材料発見における主要な手法である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、220万も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が見落としていました（図5.1.9および図5.1.10）。GNoMEのようなAI主導のプロジェクトの成功は、科学的ブレークスルーの加速におけるデータとスケーリングの威力を浮き彫りにしています。

サンプル材料構造

Source: Merchant et al., 2023

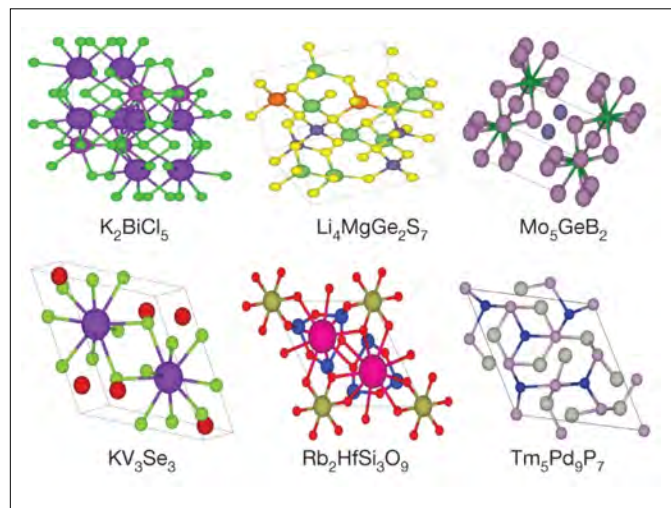


図5.1.8

GNoME 対 Materials Project：安定結晶数

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

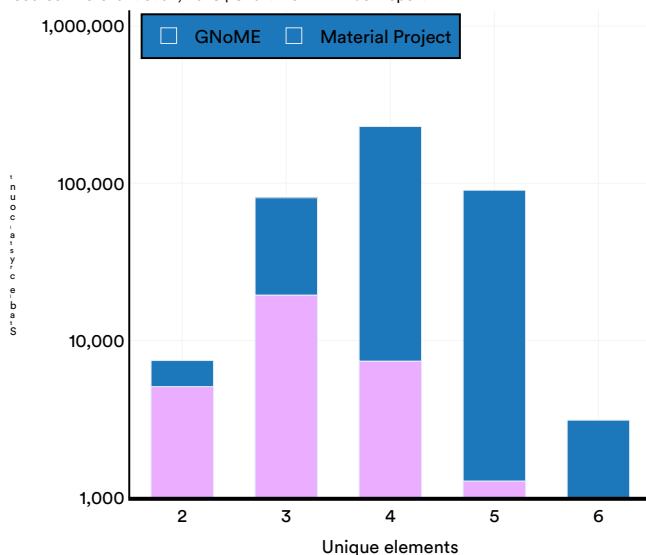


図5.1.9

GNoME 対 Materials Project：異なるプロトタイプ

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

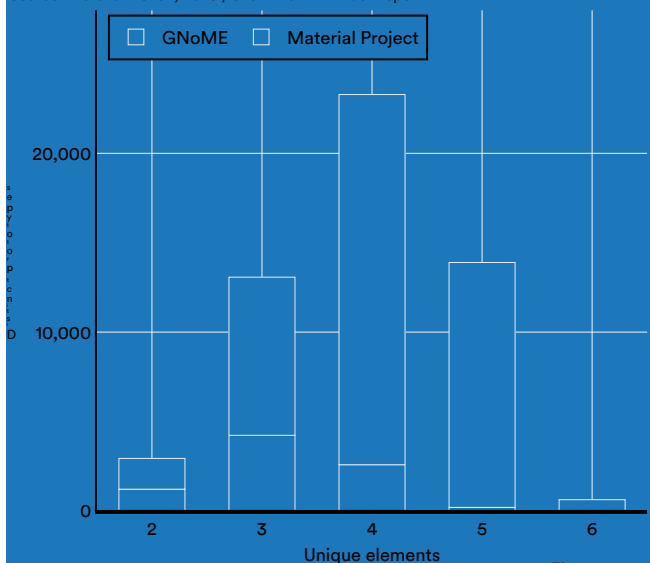


Figure 5.1.10

洪水予報

より正確で信頼性の高い洪水予報のためのAI

2023年に発表された新たな研究は、大規模な洪水イベントの予測において著しい進歩を遂げました。最も一般的な自然災害の一つである洪水は、特に防災・軽減のためのインフラが不足している発展途上国において壊滅的な影響を与えます。そのため、これらのイベントをより事前に予測できる、より正確な予測手法を開発することは、大きなプラスの影響をもたらす可能性があります。

Googleの研究者チームは、AIを用いて、未観測流域にも適用可能な高精度な水文シミュレーションモデルを開発しました¹。これらの革新的な手法は、最大5日前に特定の極端な洪水イベントを予測でき、GloFASなどの最新のモデルと同等かそれ以上の精度を達成しています。AIモデルは、様々な再現期間のイベントにおいて、優れた精度（正予測の精度）と再現率（関連するすべての事例を正しく特定する能力）を示し、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カ国以上で洪水イベントの予測に使用されています。

AIモデルとGloFASの予測値の比較：再現期間

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

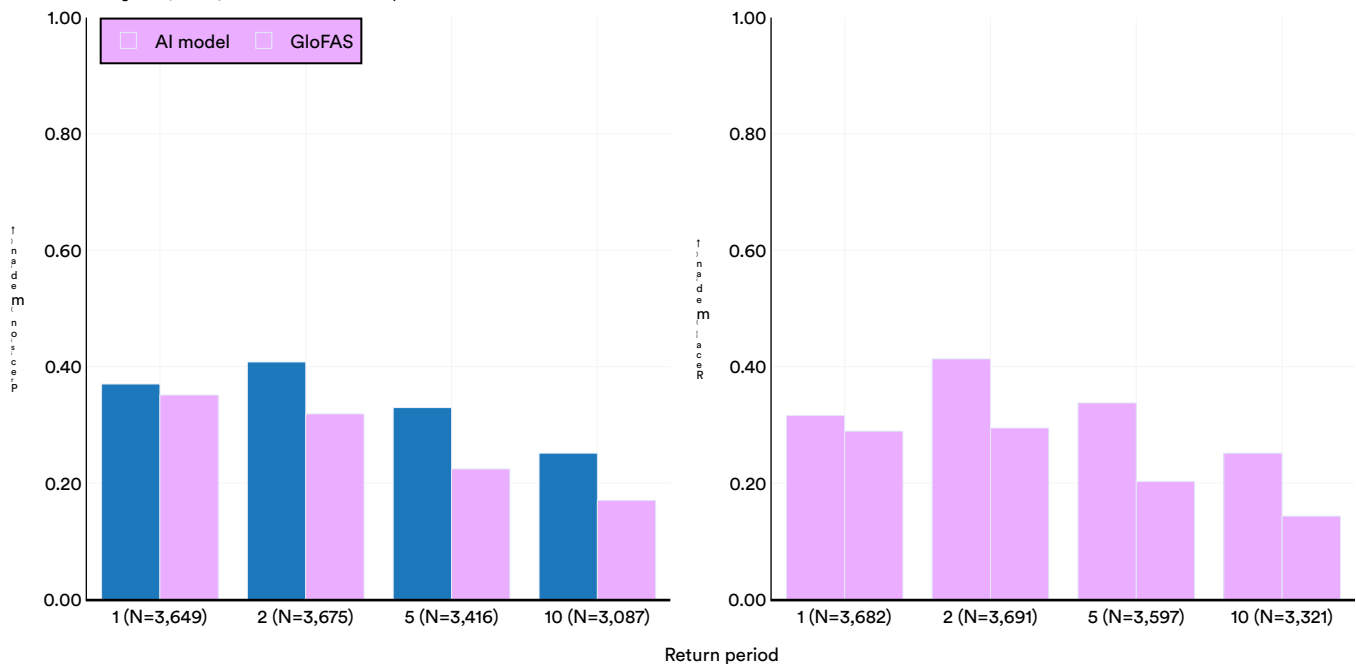


図5.1.11

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モデルは医療においてますます価値が高まっており、ポリープの検出から臨床医の診断支援まで、幅広い用途があります。AIのパフォーマンスが向上し続けるにつれて、医療現場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関連医療システム、現在の臨床AIに関する知識、病院管理の向上を目指した新しいAI診断ツールとモデルの開発について重点的に説明します。

5.2 医学におけるAI

注目すべき医療システム

本セクションでは、AI Index運営委員会が選定した、2023年の重要なAI関連医療の進歩について説明します。

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）。この進歩により、以前は多くのスキャンを高度な研究に使用すること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を明確に描写することで知られる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容易にします。SynthSRを使用した実験では、スキャンレベルと被験者レベルの両方で観察されたボリューム間に強い相関関係があることが示されており、SynthSRが、高解像度T1スキャンによって生成されたものと非常によく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、脳構造の視覚化と分析を大幅に改善し、神経科学研究と臨床診断を促進します。

SynthSR生成物

Source: Iglesias et al., 2023

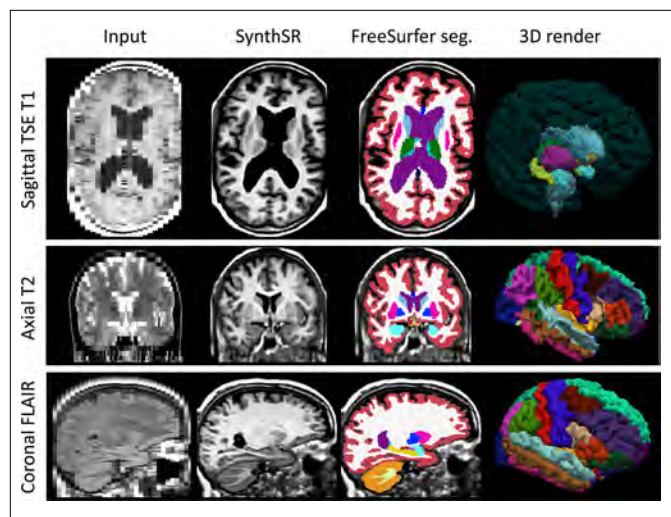


Figure 5.2.1

選択された脳領域におけるSynthSRとground-truthボリュームの相関
Source: Iglesias et al., 2023 | Chart: 2024 AI Index report

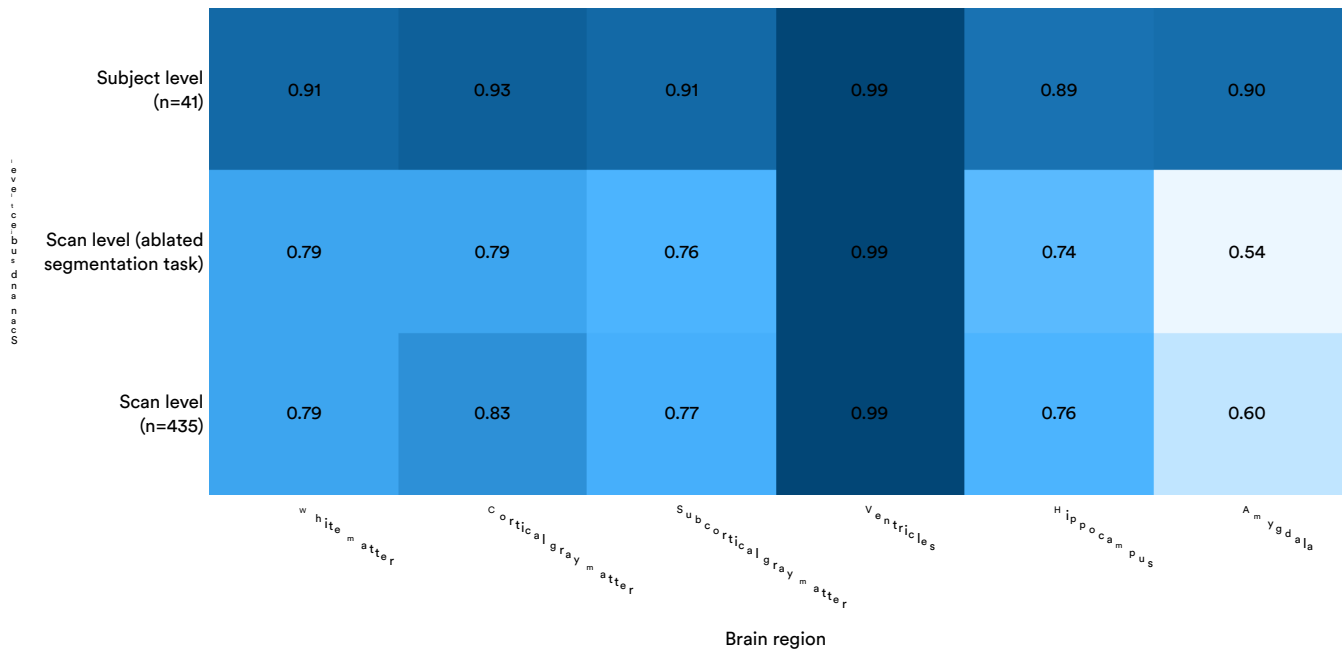


Figure 5.2.2

結合型プラズモニクス赤外線センサ

神経変性疾患検出のための結合プラズモニクス赤外線センサー

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同定に依存しています。質量分析やELISAなどの従来の方法では、タンパク質レベルの定量に焦点を当てることができるという利点がありますが、構造状態の変化を識別することはできません。本年、研究者らは、表面増強赤外吸収（SEIRA）分光法を用いたAI結合プラズモニクス赤外センサーとイムノアッセイ技術（ImmunoSEIRA；図5.2.3）を組み合わせた、神経変性疾患診断のための新しい方法を発見しました。AIシステムによる予測と実際の線維化率を比較した試験では、予測の精度は実際の報告された割合と非常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

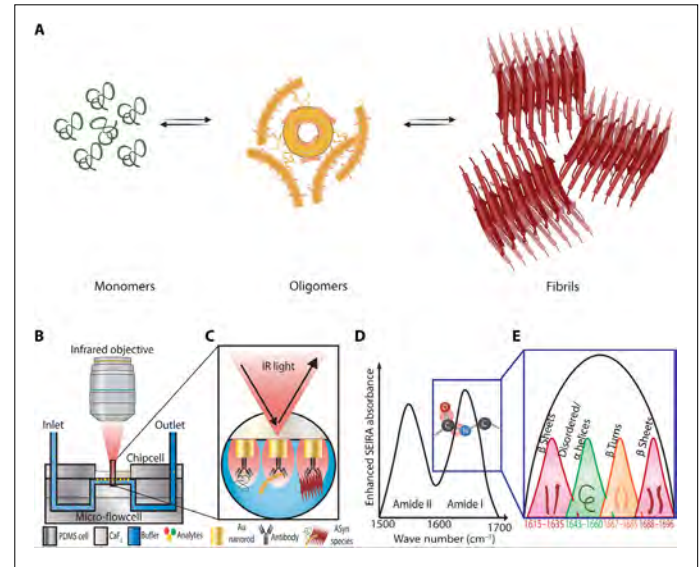


図5.2.3

深層ニューラルネットワーク予測値とテストサンプルにおける実際のブリルスの割合

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

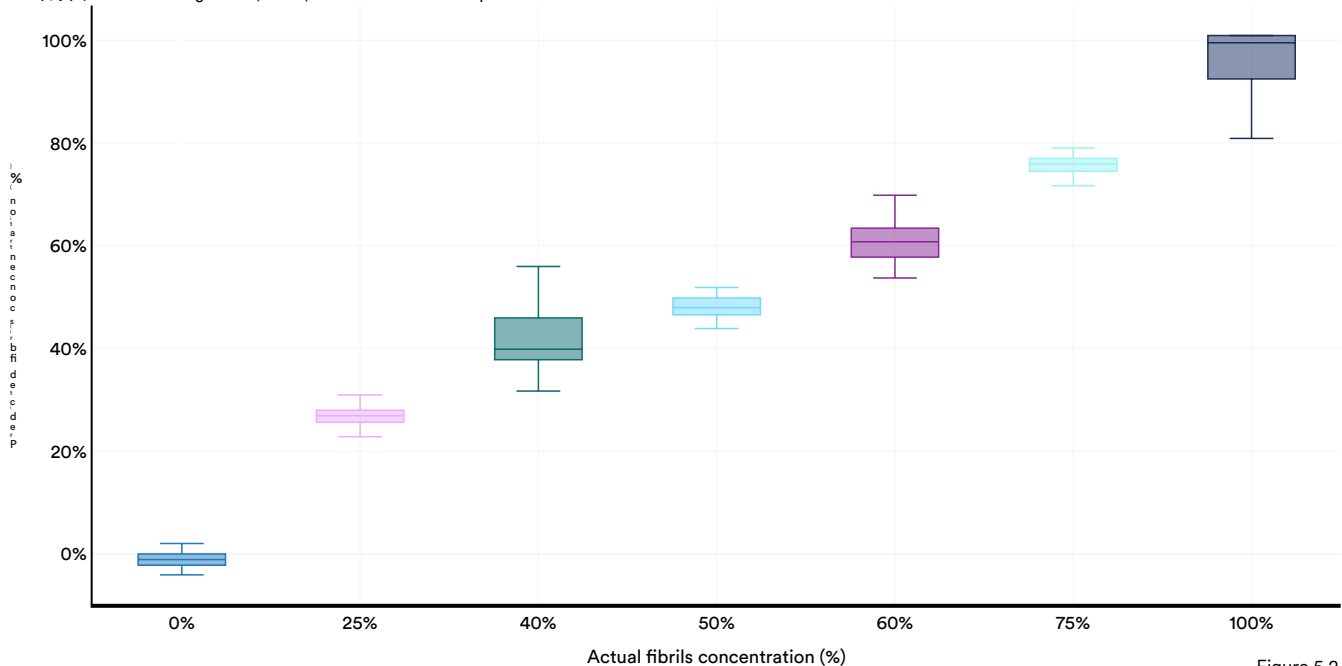


Figure 5.2.4

EVエスケープ

パンデミック対策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では、データ不足のために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的情報に基づいてトレーニングされた新しいAI深層学習モデルであり、進化を予測します {v*}

ウイルスの場合（図5.2.5）。EVEscapeは、現在の株データとは独立してウイルスエスケープを評価し、観察されたSARS-CoV-2変異の50.0%を予測し、46.2%と32.3%を予測した従来の研究や、変異のわずか24%を予測した以前のモデルを上回りました（図5.2.6）。この性能は、EVEscapeが将来のパンデミックへの備えと対応の強化に役立つ貴重な資産となる可能性を浮き彫りにしています。

EVエスケープデザイン

Source: Thadani et al., 2023

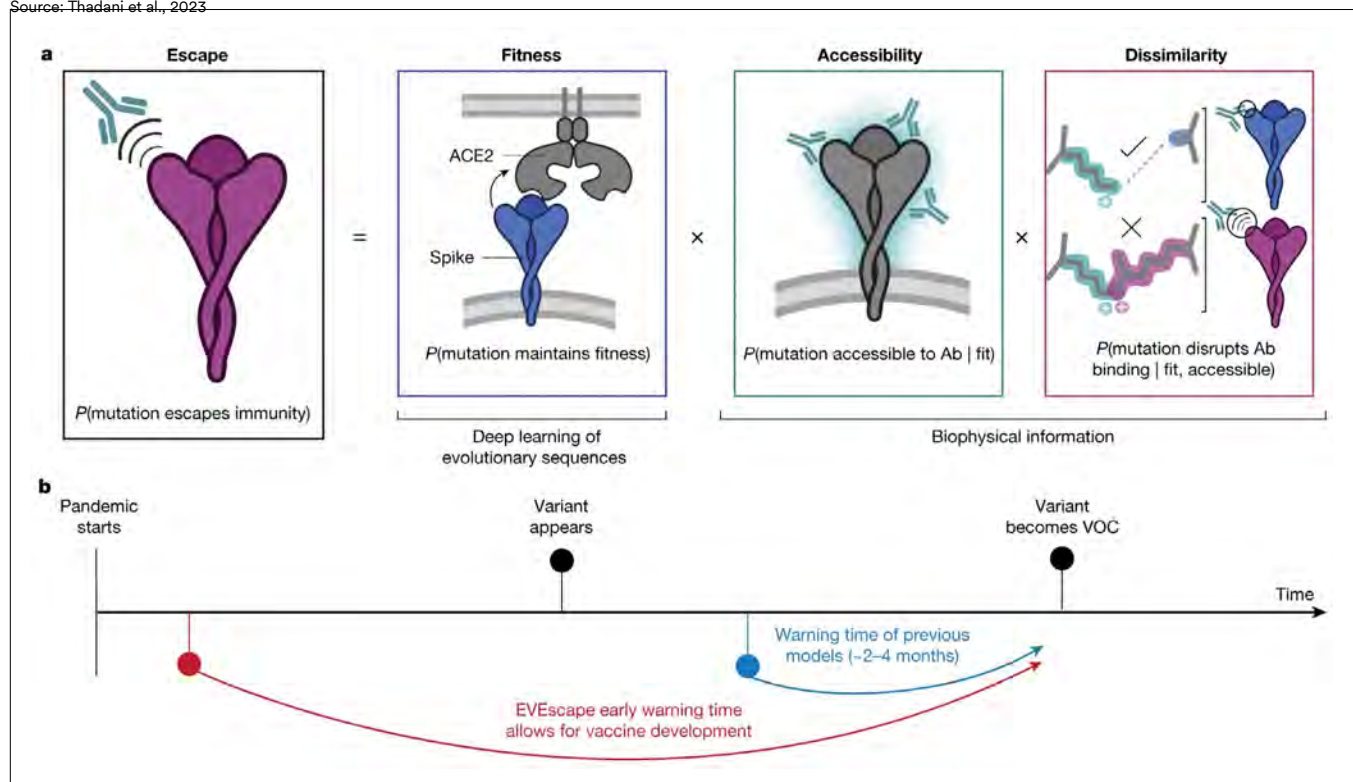


図5.2.5

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

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

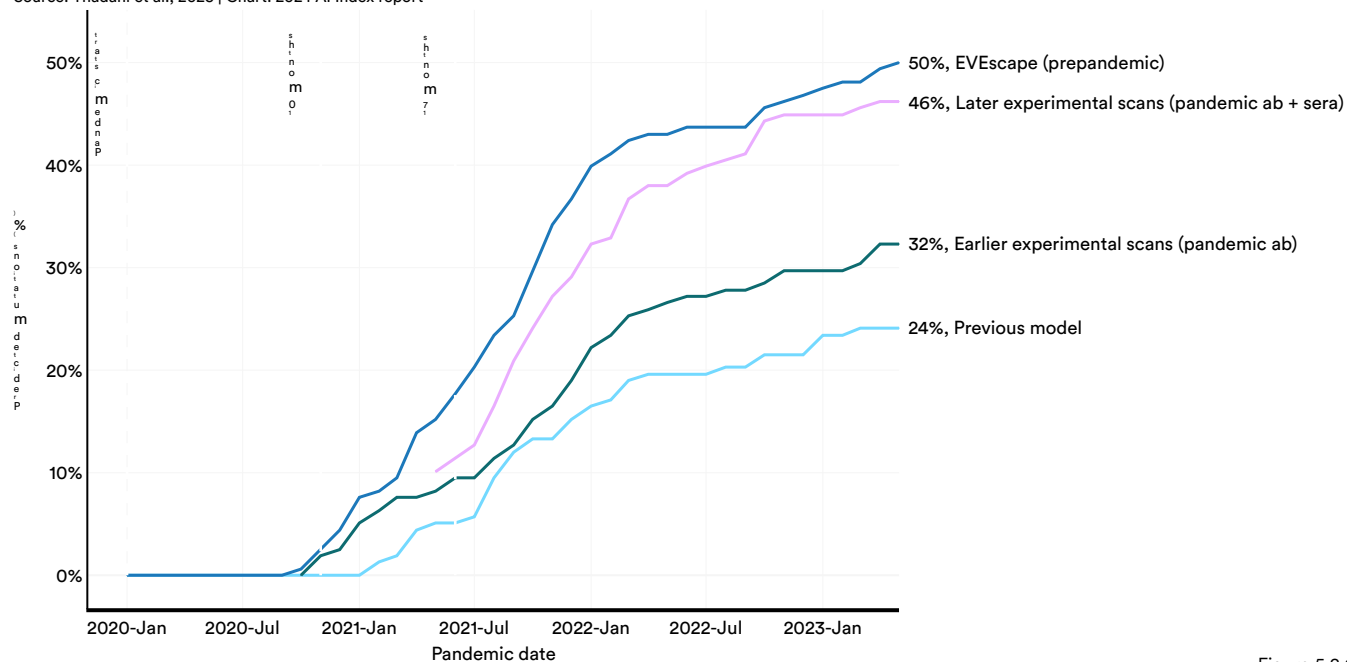


Figure 5.2.6

アルファミスセンス

AI突然変異のより良い分類

科学者たちは、いまだにどの遺伝子変異が疾患を引き起こすのかを完全に理解していません。数百万もの可能性のある遺伝子変異の中で、ある変異が良性か病源性かを決定するには、労力のかかる実験が必要です。

2023年、Google DeepMindの研究者らは、7100万個のミスセンス変異の病源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子変化（図5.2.7）であり、癌など様々な疾患を引き起こす可能性があります。7100万個の可能なミスセンス変異のうち、AlphaMissenseは89%を分類し、57%を良性である可能性が高いと、32%を病源性である可能性が高いと分類し、残りは不確定と分類しました（図5.2.8）。対照的に、ヒトアノテーターは、すべてのミスセンス変異のわずか0.1%の性質しか確認できていません。

ヘモグロビンサブユニットベータ（HBB）

Source: Google DeepMind, 2023



図5.2.7

アルファミスセンス予測

出典：Google DeepMind、2023 | 図表：2024年AI Indexレポート

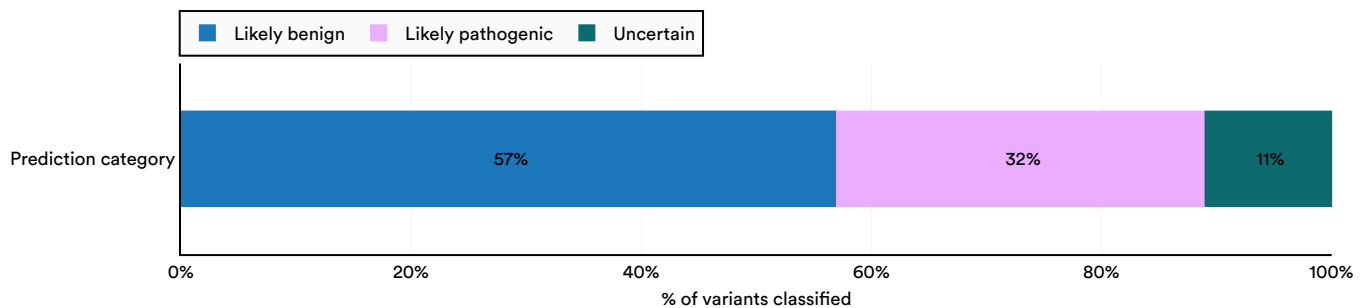


Figure 5.2.8

ヒト汎ゲノムリファレンス

AIを用いたヒトゲノムのマッピング

ヒトゲノムは、人間のための分子レベルの指示書です。ヒトゲノムの最初のドラフトは2000年に発表され、2022年に更新されました。しかし、この更新は幾分不完全でした。血液型などの様々な遺伝子変異が組み込まれておらず、多様な祖先グループの完全なマッピングもされていませんでした。そのため、既存のゲノムリファレンスでは、特定の人々のグループにおける疾患の検出や治療法の発見が困難になります。

2023年、60の機関に所属する119人の科学者からなるヒトパンゲノム研究コンソーシアムは、AIを用いて更新され、より代表的なヒトゲノムマップ（図5.2.9）を作成しました。研究者らは、タンパク質コード遺伝子の99.07%、タンパク質コード転写物の99.42%、非コード遺伝子の98.16%、非コード転写物の98.96%を注釈付けするという著しい精度を達成しました（図5.2.10参照）。

ゲノムのMHC領域のグラフゲノム

Source: Google Research, 2023

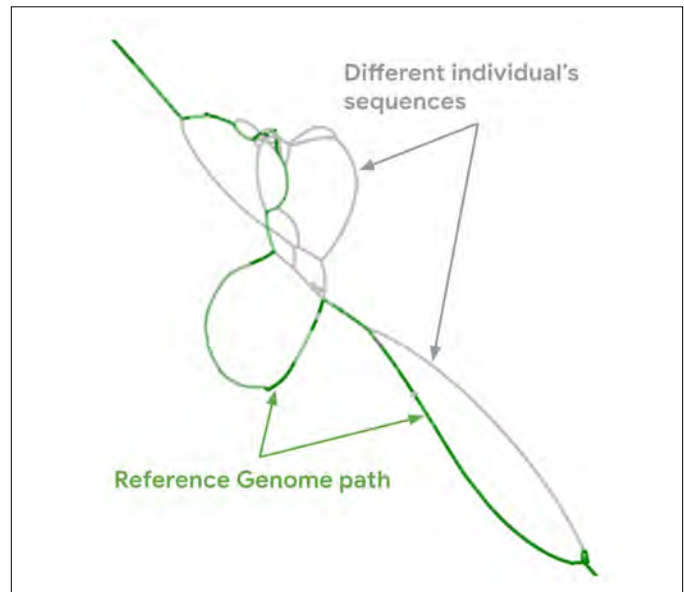


図5.2.9

この最新のゲノムバージョンは、現在までに最も包括的で遺伝的多様性に富んだ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: Liao et al., 2023 | Chapter 5.2 AI Index report

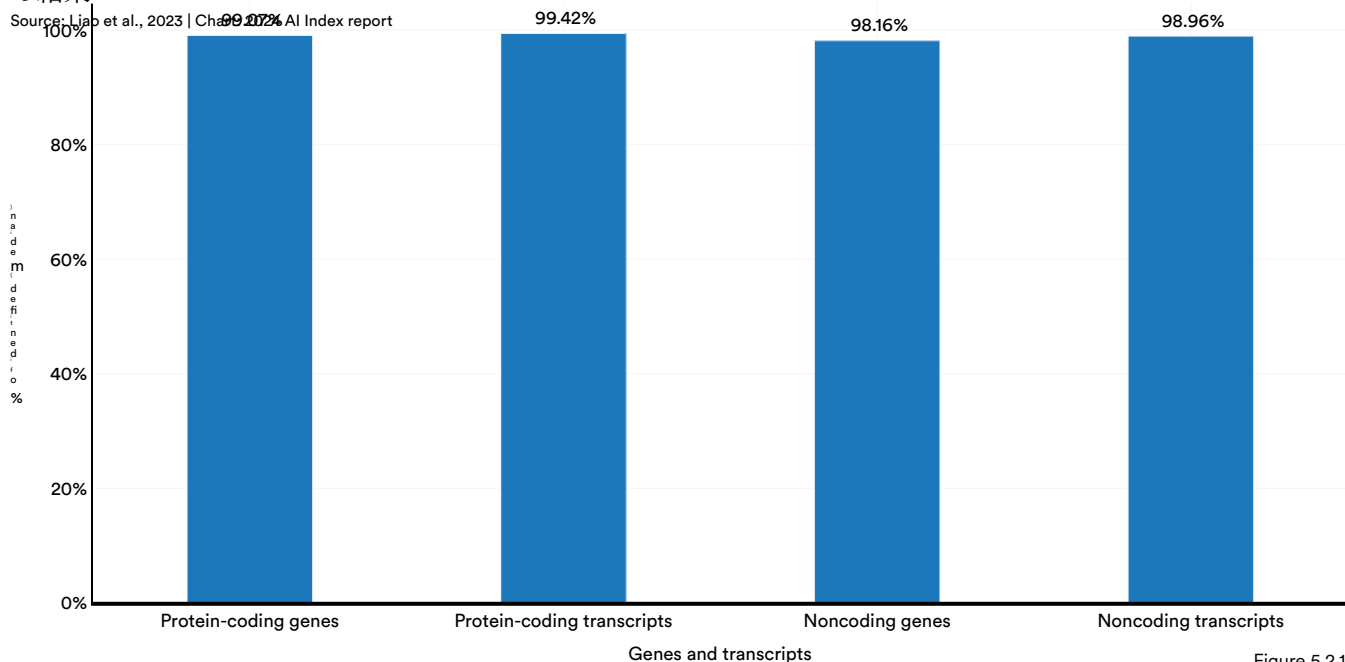


Figure 5.2.10

臨床知識

AIモデルの臨床知識の評価には、特に臨床現場で適用可能な医学的専門知識の程度を決定することが含まれます。

目覚ましい進歩が見られ、主要システムであるGPT-4 Medpromptは90.2%の精度を達成しました。これは2022年の最高スコアから22.6ポイントの上昇です（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倍になり、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、医師に挑戦するように設計された6万件以上の臨床的問題を含む、専門医の試験から派生した包括的なデータセットです。

MedQAベンチマークにおけるAIのパフォーマンスは、
MedQA：精度

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

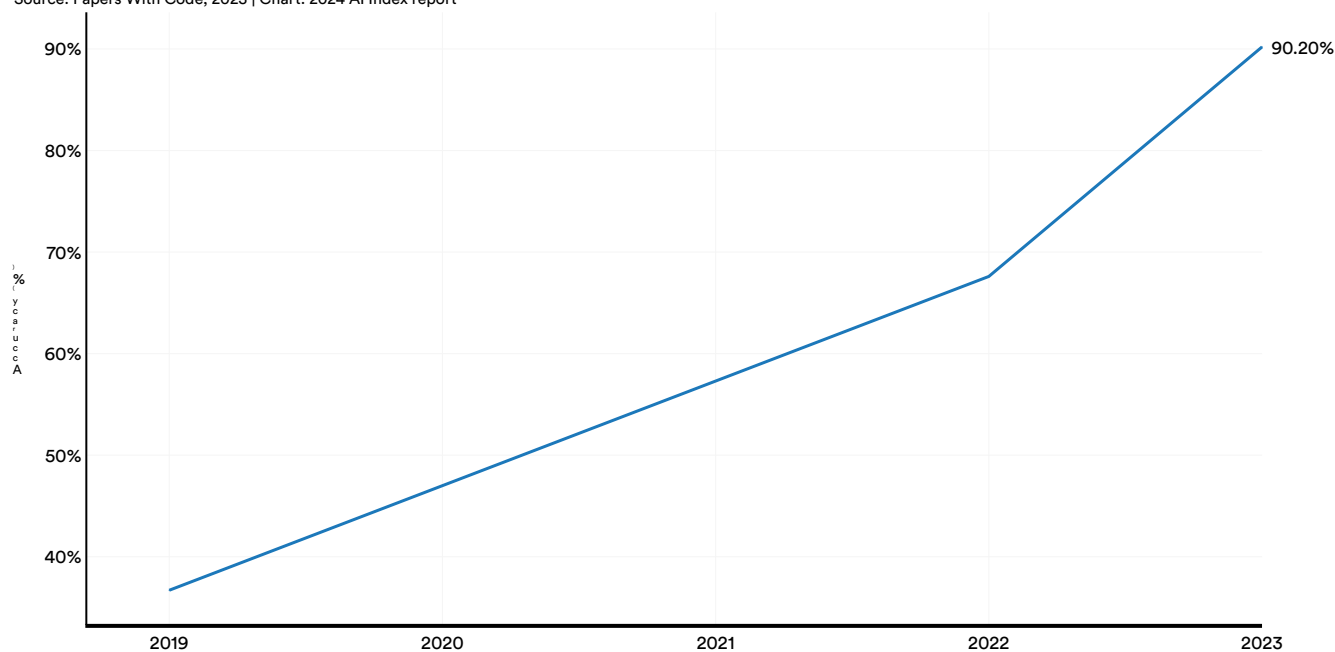


図5.2.11

注目研究：

GPT-4 Medprompt

大規模言語モデル（LLM）は優れた一般的な知識を示していますが、医学的な質問に答えるなど、専門的な知識において優れた性能を発揮するには、大規模なファインチューニングが必要であると一般的に考えられています。ファインチューニングには、ドメイン固有のデータでLLMをトレーニングすることが含まれます。

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、4つの困難な医学的ベンチマーク（図5.2.12）のグループであるMultiMedQAベンチマークスイートにおいて、GPT-4が優れた性能を達成するように導きました。GPT-4 Medpromptは、PubMedQA、MedMCQA、MMLUなど、いくつかの有名な医学的ベンチマークの複数選択セクションにおいて、2022年のトップモデルであるFlan-PaLM 540Bの性能をそれぞれ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るMed-PaLM 2（図5.2.13）の性能も上回りました。

さらに、先に述べたように、MedQAベンチマークで90%の精度を達成したのはGPT-4 Medpromptが初めてでした。この画期的な成果は、GPT-4 Medpromptの例外的な、そして臨床的に有用となる可能性のある医療能力を強調するだけでなく、専門分野へのモデルの適応にファインチューニングが常に必要とは限らないことを示しています。プロンプトエンジニアリングは、有望な代替戦略であることが示されています。


GPT-4対Med-PaLM 2による医学的問題への回答

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

注目研究：

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

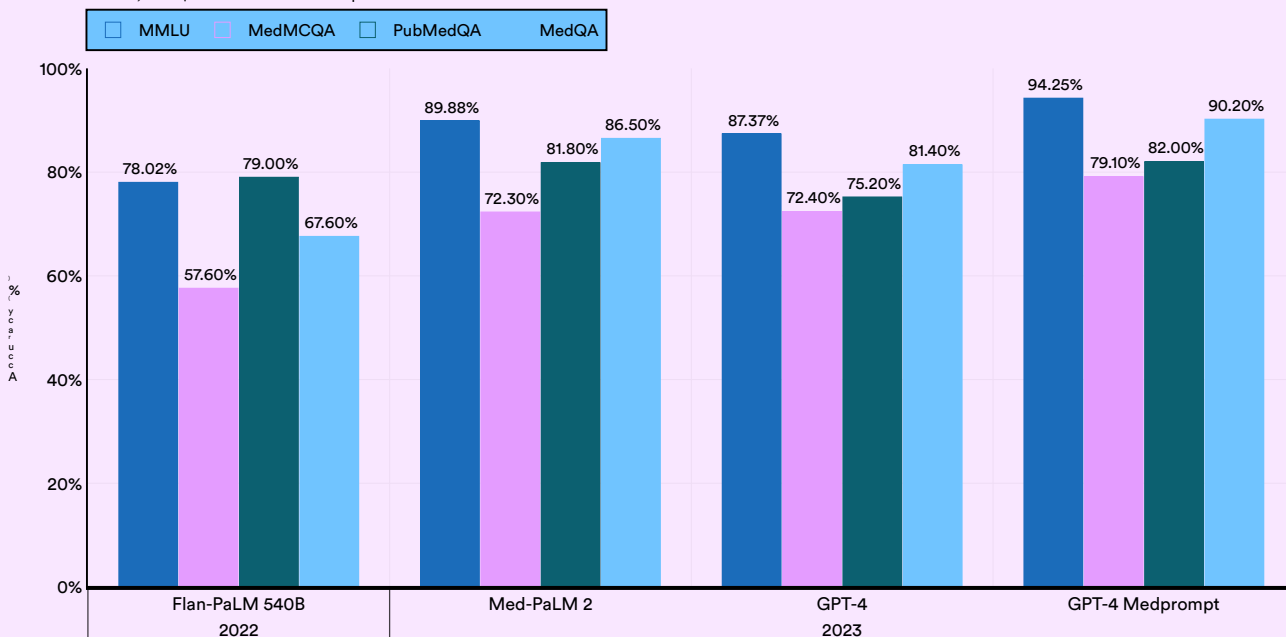


図5.2.13

注目研究：

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、その重みは一般に無料で利用できません。2023年の新しい研究では、オープンソースの医療LLMの能力向上も目指しています。この新しい研究の中で、MediTron-70Bは特に有望です。このモデルは、MedQAベンチマークで70.2%という高い精度を達成しています。これはGPT-4 MedpromptやMed-{v*}の性能を下回っていますが、

PaLM 2（両方のクローズドモデル）は、2023年の最先端の結果を大幅に上回り、Llama 2（図5.2.14）などの他のオープンソースモデルを凌駕しています。MediTron-70BのMedQAスコアは、オープンソースモデルでこれまで達成された中で最高です。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。この文脈において、MediTronは前向きな一歩を表しています。

MedQAにおける一部モデルの性能

出典：Chen et al., 2023 | 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、放射線科やがん検出など、診断目的にも使用できます。

注目研究：

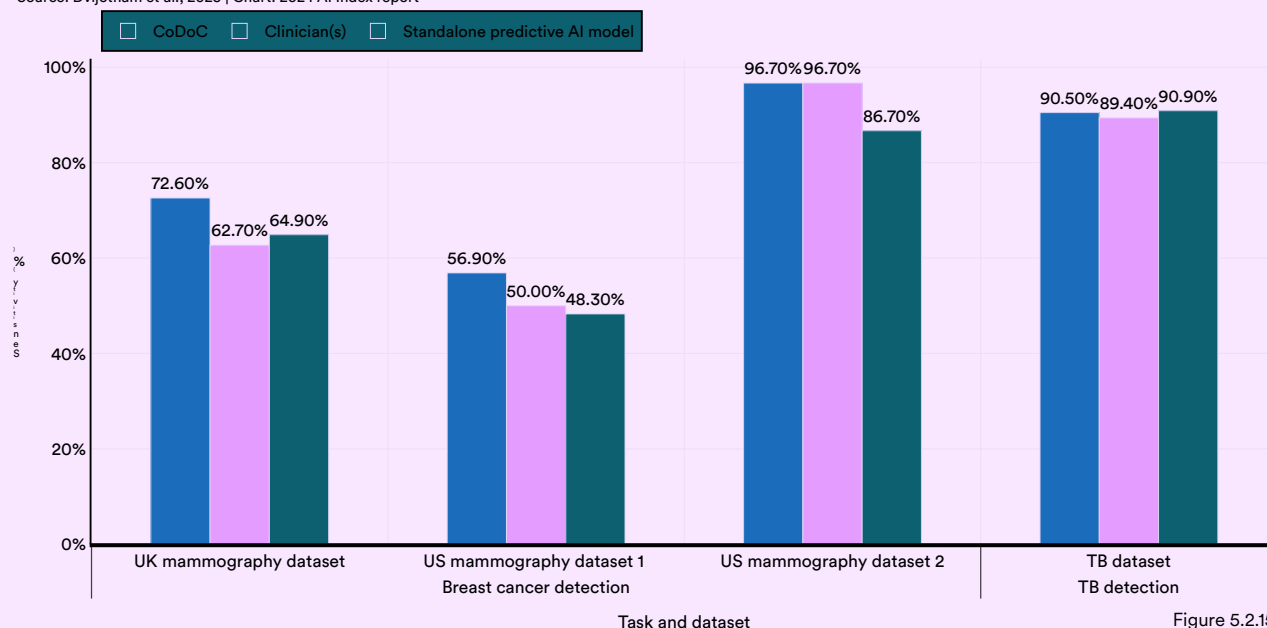
CoDoC

AI医療画像システムは強力な診断能力を示していますが、臨床医が発見する診断を見落とす場合や、その逆の場合もあります。この観察結果は、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、研究者らは、AIを診断にいつ頼るか、従来の臨床方法にいつ委ねるべきかを判断するように設計されたシステム、CoDoC（Complementarity-Driven Deferral to Clinical Workflow）を発表しました。CoDoCは、感度（疾患を持つ個人を正しく特定する能力）と特異度を著しく向上させます。

（それを有し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も平均4.5パーセンテージポイント、スタンドアロンAIモデルよりも6.5パーセンテージポイント上回っています（図5.2.15）。特異性に関して、CoDoCは、テストされたデータセット全体で臨床医よりも平均2.7パーセンテージポイント、スタンドアロン予測モデルよりも5.7パーセンテージ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。これらの知見は、AI医療システムを臨床ワークフローに統合することにより、診断の精度と効率を向上させることができることを示唆しています。

CoDoC 対 独立型予測AIシステムと臨床読影医：感度

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



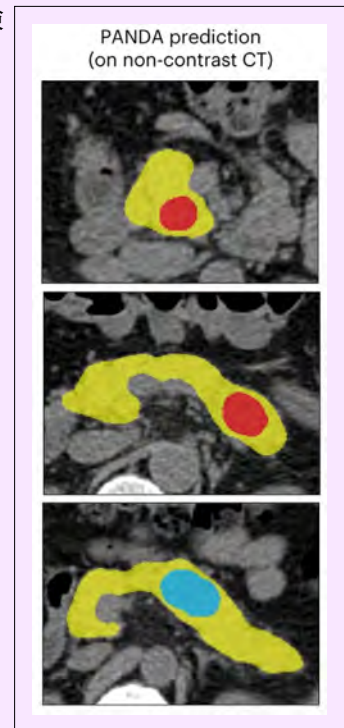
注目研究：

CTパンダ

膵臓癌（PDAC）は特に致死性の高い癌であり、しばしば外科的介入が不可能な時期に発見されます。無症状の個人におけるPDACのスクリーニングは、その有病率が低く、偽陽性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線画像における膵臓病変の効率的な検出と分類が可能なAIモデルであるPANDA（人工知能を用いた膵臓癌検出）を開発しました（図5.2.16）。検証試験において、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現実世界試験において、PANDAは92.9%の感度と99.9%の特異度を達成しました（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不可能と考えられていた費用対効果が高く正確な検出を提供することにより、困難な病状の診断における大きな進歩を表しています。

PANDA検出

Source:
Cao et al., 2023
図5.2.16



出典：Caoら、2023 | 図表：2024年AI指標レポート PANDA対複数施設検証における平均放射線科医（6,239人の患者）

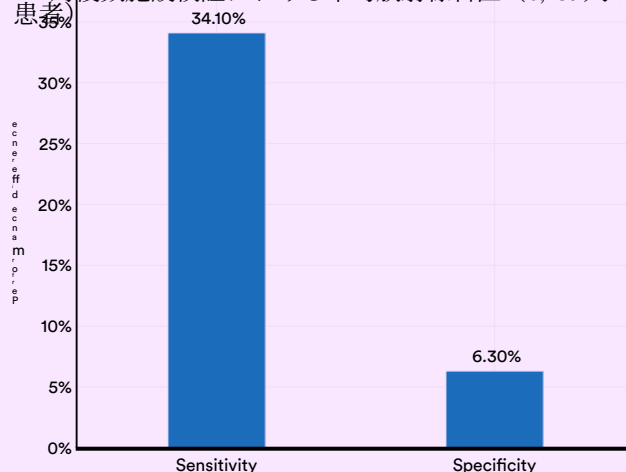


Figure 5.2.17

実世界の多シナリオ検証（20,530人の患者）におけるPANDAのパフォーマンス

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

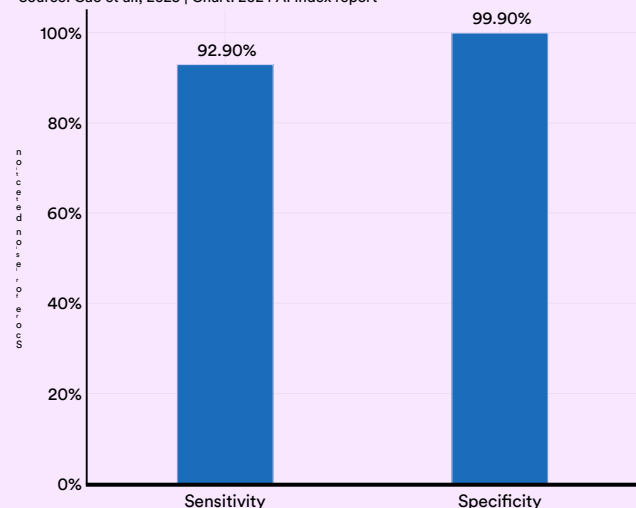


Figure 5.2.18

その他の診断用途

2023年に発表された新しい研究は、AIが他の診断の文脈で使用できる方法を強調しています。図5.2.19は、その調査結果の一部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承認済み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の事前承認基準を満たしており、これには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が承認した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が承認したAI医療機器の数、2012～22年

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

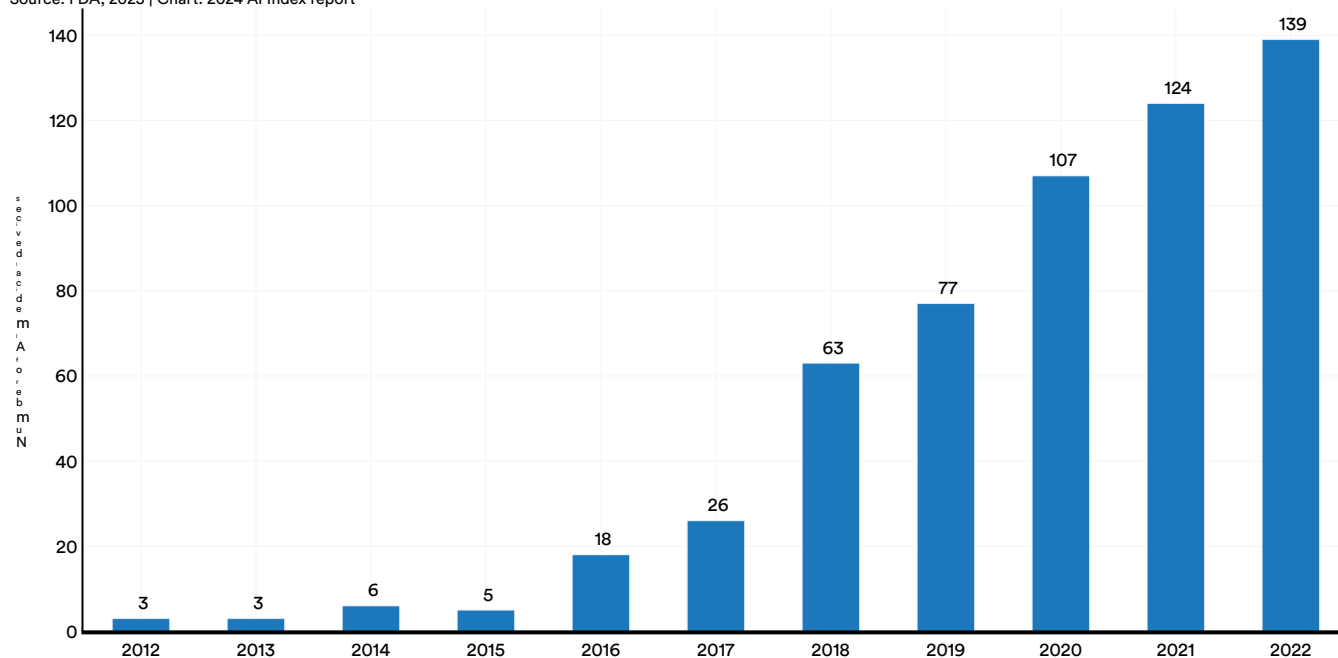


図5.2.20

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

図5.2.21は、FDA承認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台の機器のうち、87.1%という大部分が放射線医学関連でした。次に多かったのは心臓血管系で、承認の7.2%を占めていました。

専門分野別のFDA承認AI医療機器数、2012～2022年

出典：FDA、2023 | 図表：2024 AI Index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、医療行政の効率性向上と患者ケアの質向上にも役立つ可能性を秘めています。

再

注目研究：

MedAlign

医療におけるAIの著しい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、臨床医が患者ケアにおいて実行する多様なタスクを完全に捉えているわけではありません。臨床医は、個別化された診断計画の作成など、情報集約的なタスクに頻繁に従事し、就業時間の相当部分を事務作業に費やしています。AIはこれらのプロセスを合理化する可能性を秘めていますが、医療事務用LLMのベンチマークとファインチューニング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MedAlignを導入することでこのギャップに対処するための進歩を遂げました。MedAlignは包括的なEHRベースの

7つの異なる医療専門分野から得られた983個の質問と指示、および303件の臨床医の回答をベンチマークとした（図5.2.22）。MedAlignは、最初の広範なEHRに焦点を当てたベンチマークである。

研究者らはその後、MedAlignを用いて様々な既存のLLMをテストした。全てのLLMの中で、多段階改良を用いたGPT-4の変種が最も高い正解率（65.0%）を達成し、他のLLMよりも常に好まれた（図5.2.23）。MedAlignは、AIを用いて医療における事務的負担を軽減するための貴重なマイルストーンである。

MedAlignワークフロー

— Source: Fleming et al., 2023

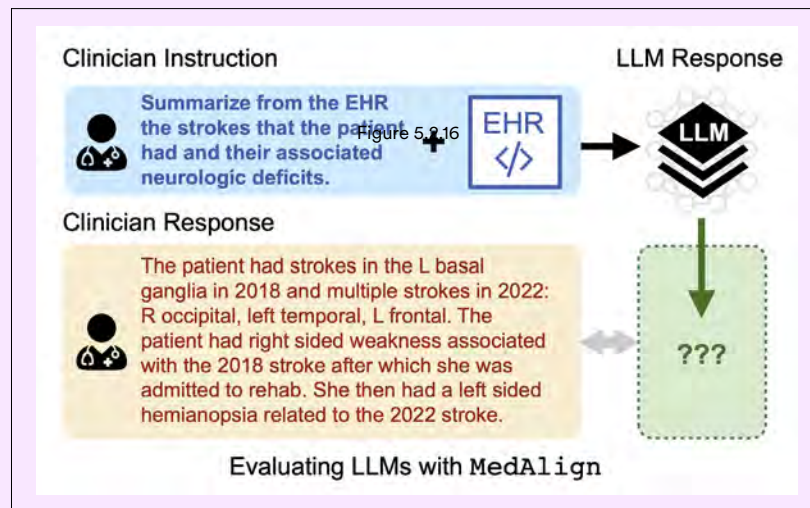


図5.2.22

注目研究：

MedAlign（続き）

モデル性能の評価：人間 vs. COMET ランク

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

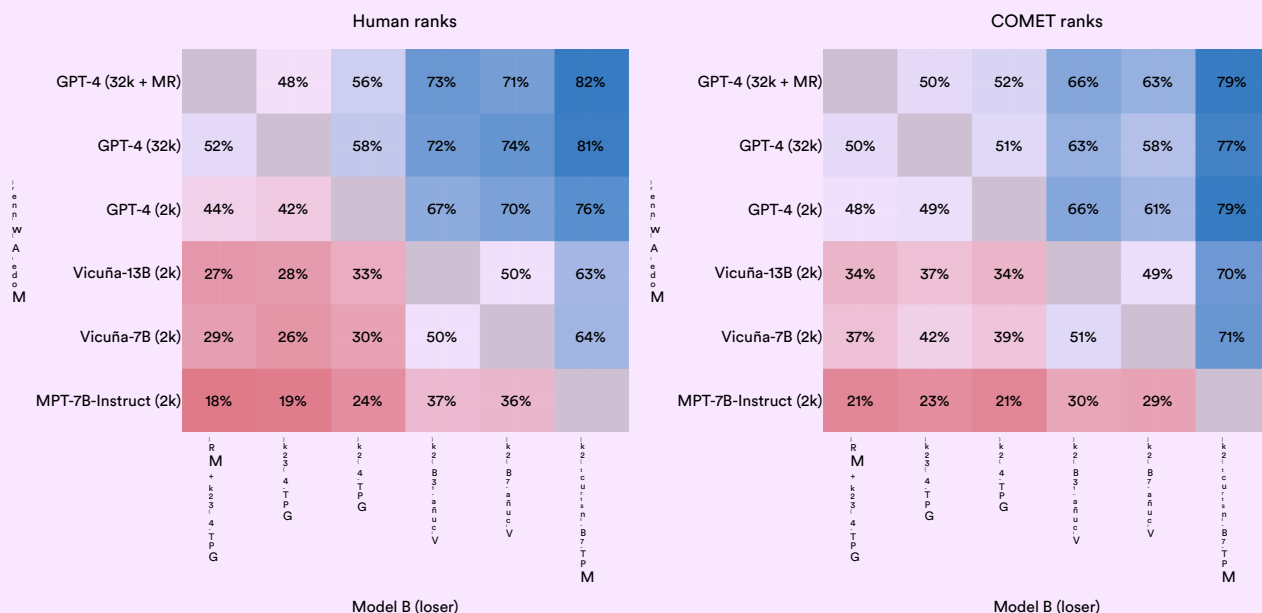
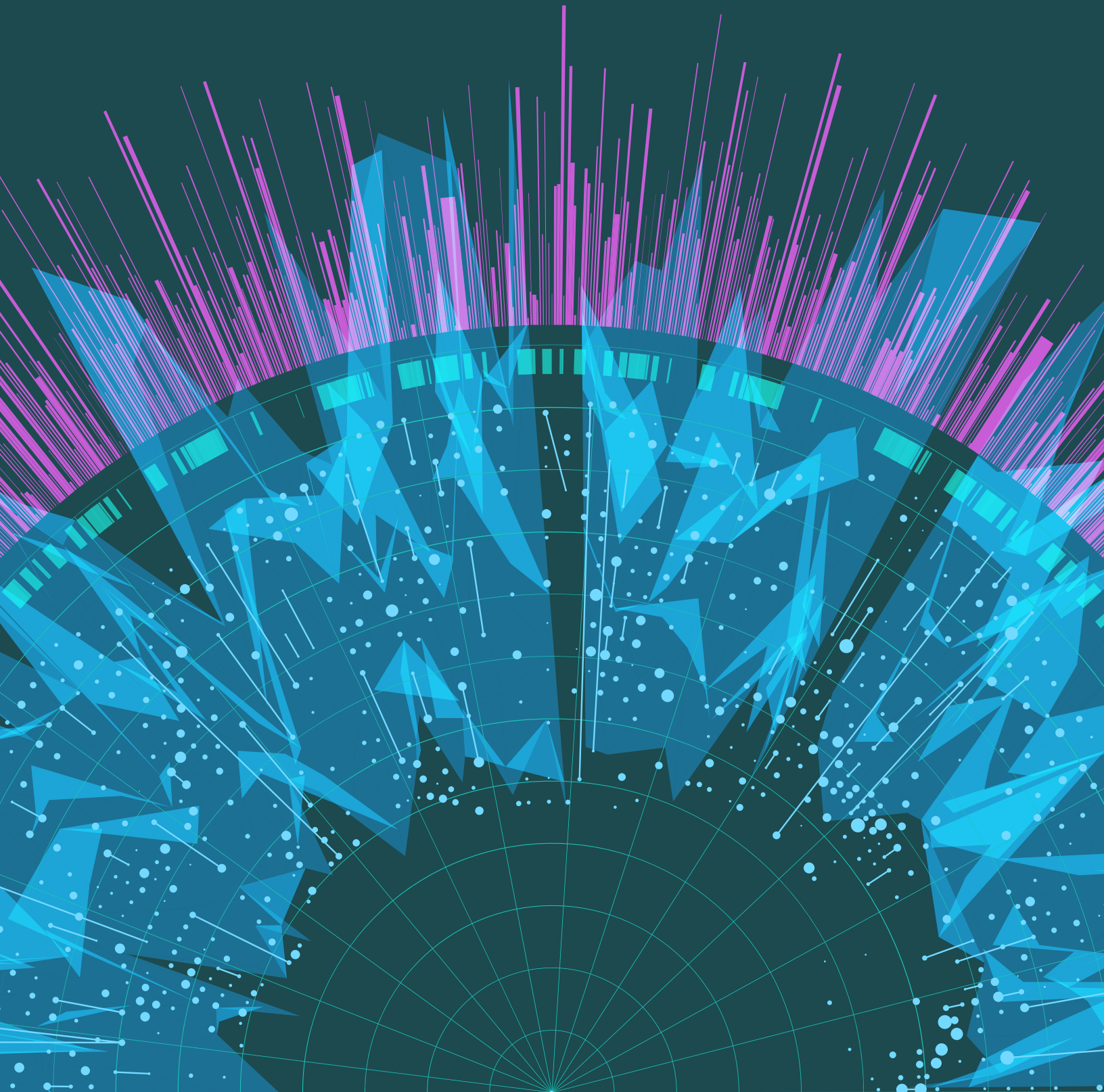


図5.2.23



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におけるCSとAI教育	359
アメリカ合衆国	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：アクセスに関する課題	363
ハイライト：教員と生徒におけるChatGPTの利用	364

公開データへのアクセス

概要

本章では、AIとコンピュータサイエンス（CS）教育の動向を、誰が学んでいるか、どこで学んでいるか、そしてこれらの動向がどのように時間とともに進化してきたかに焦点を当てて検証します。AIが教育に及ぼす影響に関する懸念が高まる中、教師と生徒によるChatGPTなどの新しいAIツールの使用についても調査します。

本分析は、Computing Research Associationの年次Taulbee調査に基づいた、米国とカナダにおける高等教育におけるコンピュータサイエンスとAI教育の現状の概要から始まります。その後、Informatics Europeのデータに基づいたヨーロッパにおけるコンピュータサイエンス教育に関するレビューを行います。今年は、AI関連の英語の学習プログラムの世界的な数を示すStudy portalsからのデータを含む新しいセクションが導入されています。

Code.orgによる米国におけるK-12 CS教育に関する知見と、Walton Foundationによる学校におけるChatGPTの利用に関する調査結果をまとめとして本章を締めくくります。

章のハイライト

1. アメリカとカナダのCS学部卒業者の数は増加を続け、新たなCS修士課程卒業者は比較的横ばい、博士課程卒業者は緩やかに増加している。一方、

アメリカとカナダの新たな学士号取得者の数は10年以上着実に増加している一方、CS（コンピュータサイエンス）の大学院教育を選択する学生数は頭打ちになっている。2018年以降、CSの修士号および博士号取得者の数はわずかに減少している。

2. AI博士号取得者の産業界への流出は、加速するペースで続いている。2011年には、新規AI博士号取得者のうち、産業界（40.9%）と学界（41.6%）への就職率はおおよそ同程度であった。しかし、2022年までに、卒業後に産業界へ就職する割合（70.7%）は、学界へ進む割合（20.0%）を大幅に上回った。過去1年間だけで、産業界への就職を希望するAI博士号取得者の割合は5.3ポイント上昇しており、大学から産業界への人材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優秀な人材の流動性が低下している。2019年、米国とカナダにおけるAI分野の新たな教員の13%が産業界出身であった。2021年までに、この数値は11%に減少し、2022年にはさらに7%に減少した。この傾向は、産業界から学界への高度な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. 米国とカナダにおけるCS教育の国際化が後退。2022年のCS学部、修士課程、博士課程の卒業生における留学生の割合は、2021年よりも減少した。特に修士課程における留学生の減少は顕著だった。

5. アメリカの高校生のコンピュータサイエンス科目の履修者は増加しているが、アクセスの問題が残っている。2022年には、20万1000件のAP CS試験が実施された。2007年以降、これらの試験を受験する生徒数は10倍以上に増加している。しかし、最近の証拠によると、より大きな高校の生徒や郊外の生徒の方が、コンピュータサイエンス科目にアクセスできる可能性が高いことが示されている。

6. AI関連の学位プログラムは国際的に増加している。英語によるAI関連の高等教育プログラムの数は、2017年以来3倍になり、過去5年間で着実に増加している。世界中の大学が、AIに特化した学位プログラムをより多く提供している。

7. ヨーロッパの情報学、CS、CE、およびIT分野の卒業生輩出数において、英国とドイツが首位を占めている。英国とドイツは、情報学、CS、CE、および情報分野の学士、修士、博士課程の卒業生数を最も多く輩出しているヨーロッパ諸国である。一人当たりの卒業生数では、フィンランドが学士号と博士号取得者の輩出数で首位に立ち、アイルランドが修士号取得者の輩出数で首位に立っている。

このセクションでは、CSとAIにおける高等教育の概要を示し、北米とヨーロッパにおける学士、修士、博士号など様々な学位の卒業統計を強調します。また、英語で提供されているAI関連コースに関する情報も網羅しています。

6.1 高等教育におけるCSとAI教育

アメリカ合衆国とカナダ

本セクションでは、米国とカナダにおけるCSおよびAI高等教育の現状を評価するComputing Research AssociationのTaulbee調査からのデータ分析を示します。この調査は、米国とカナダの297の博士号授与CS学部を対象としています。¹

コンピュータサイエンス学部卒業生

過去10年間、北米における新規コンピュータサイエンス学士号卒業生の総数は着実に増加し、3倍以上に増加しており、2021年から2022年にかけては前年比7.9%の増加となっています（図6.1.1）。

米国とカナダにおける2010～2022年の新規コンピュータサイエンス学士号卒業生
Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

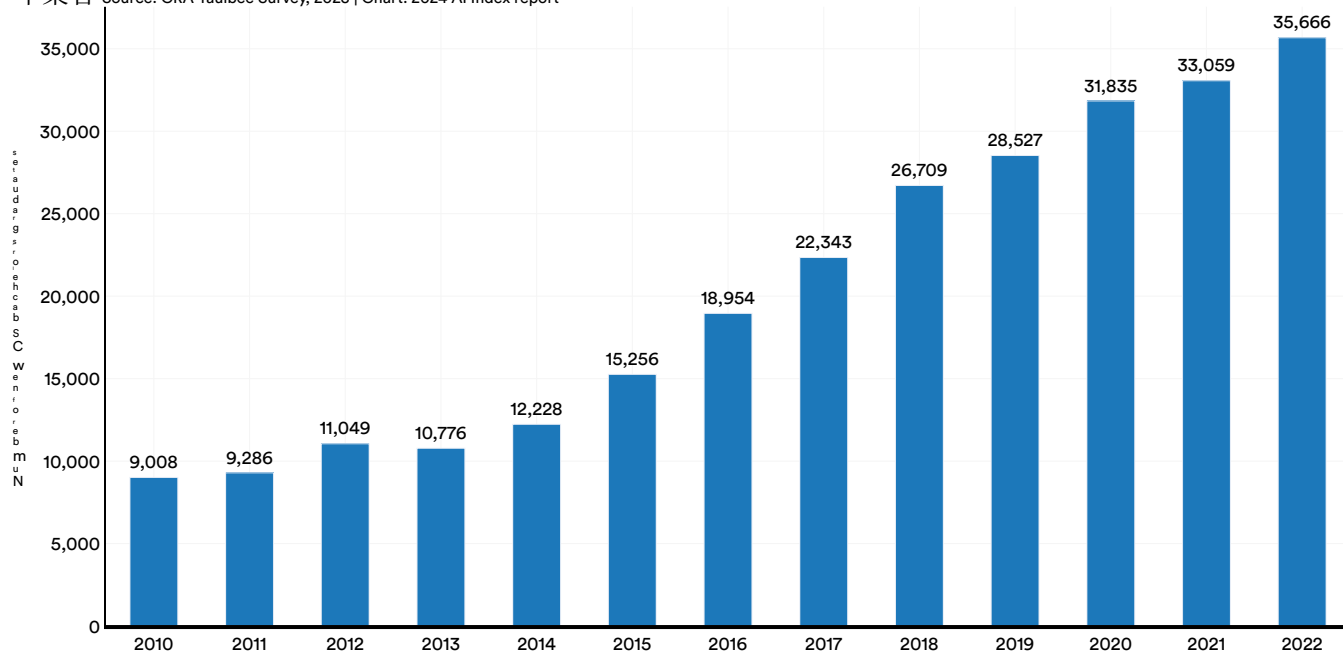


Figure 6.1.1

調査対象となったすべての博士号授与学科が回答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%でした。

約8年ぶりに、アメリカとカナダの大学のコンピュータサイエンス学部卒業生における留学生の割合が減少しました。2021年の16.3%から2022年には15.2%に低下しました（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困難化を反映している可能性があります、その影響が今になって現れ始めています。

データの減少。この減少は、COVID-19パンデミック中に課せられた国際的な渡航制限にも部分的に起因しており、留学生が米国と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、留学生の割合の着実な増加を示しています。

米国とカナダにおける新規国際CS学士卒業生（総数に対する割合）、2010～22年

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

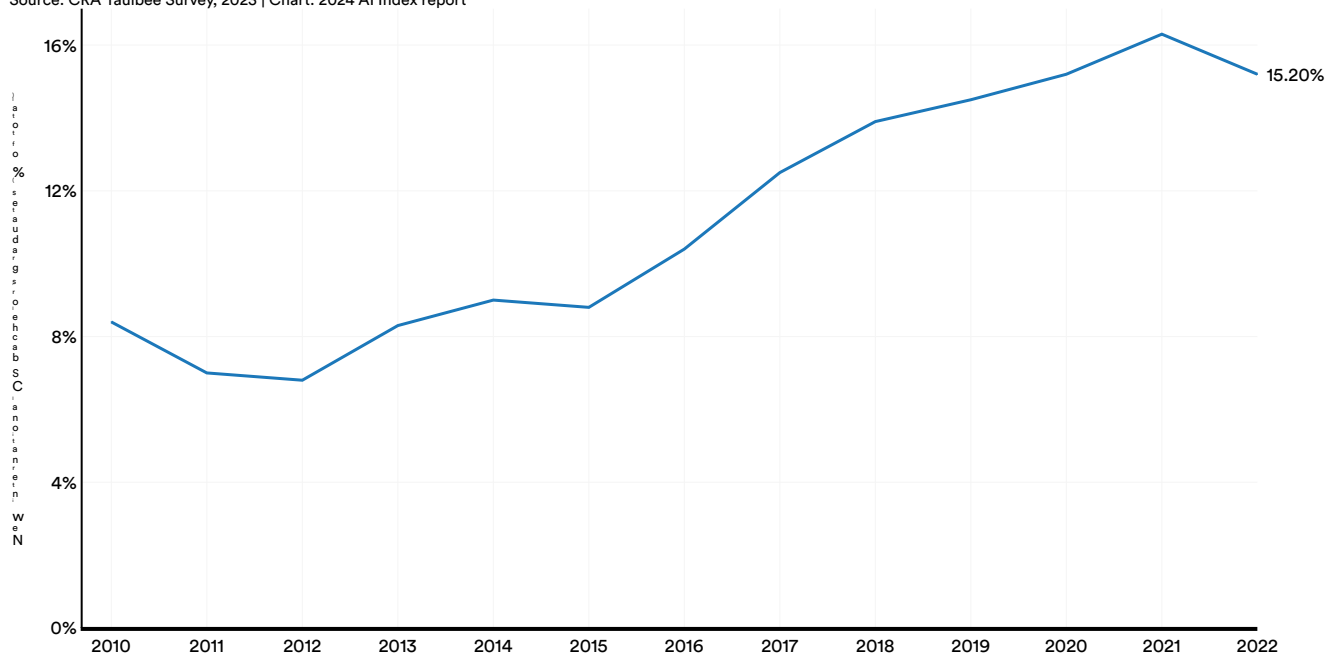


図6.1.2

CS修士課程卒業生

AIコースは、一般的にCS修士課程に含まれています。過去10年間で、アメリカとカナダの大学の新しいCS修士号取得者の総数は2倍以上に増加しましたが、

2018年以降、その数は横ばい傾向にあり、昨年は2.5%微減しました（図6.1.3）。この横ばい傾向は、次のグラフに示す国際的な修士課程の学生数の減少を反映しています。

米国とカナダにおける2010～2022年の新規コンピュータサイエンス修士課程卒業生
Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

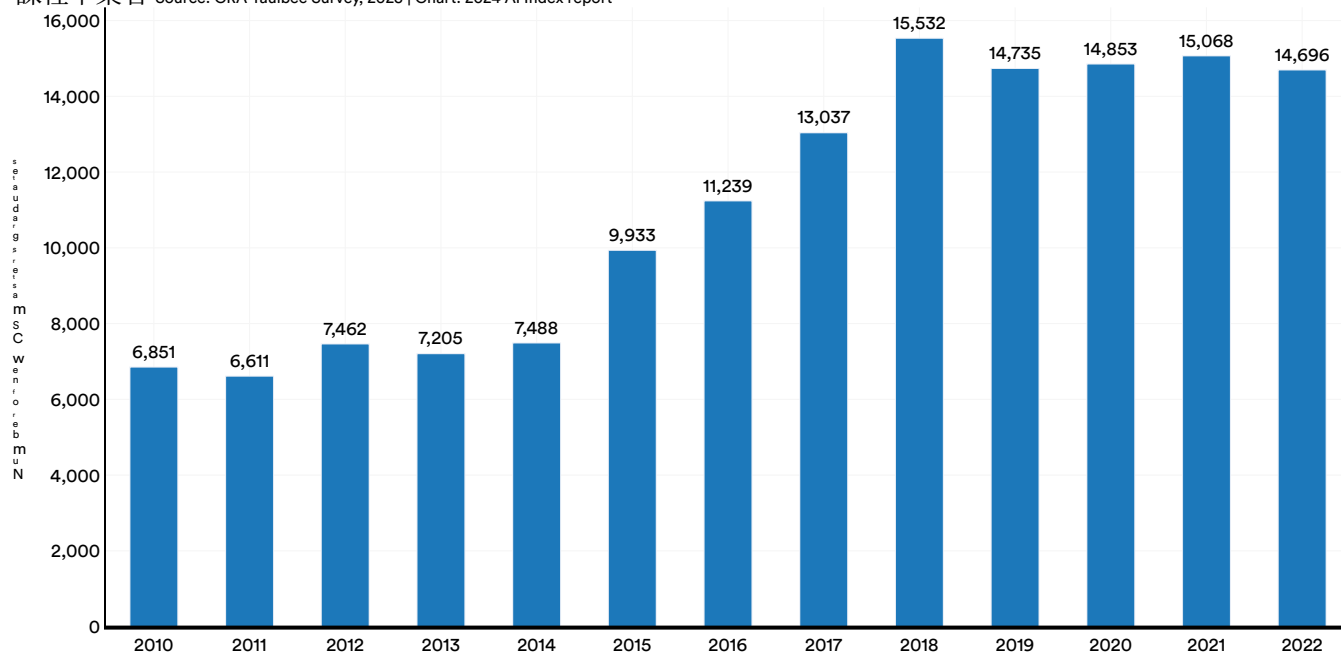


図6.1.3

2022年、アメリカとカナダの大学では、国際的なCS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、昨年は14.8パーセントポイント減少と最も顕著でした（図6.1.4）。現在、国際的なCS修士課程卒業生と国内のCS修士課程卒業生の割合はおよそ同数です。

米国とカナダにおける新規国際CS修士課程卒業生（総数に対する割合）、2010～22年

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

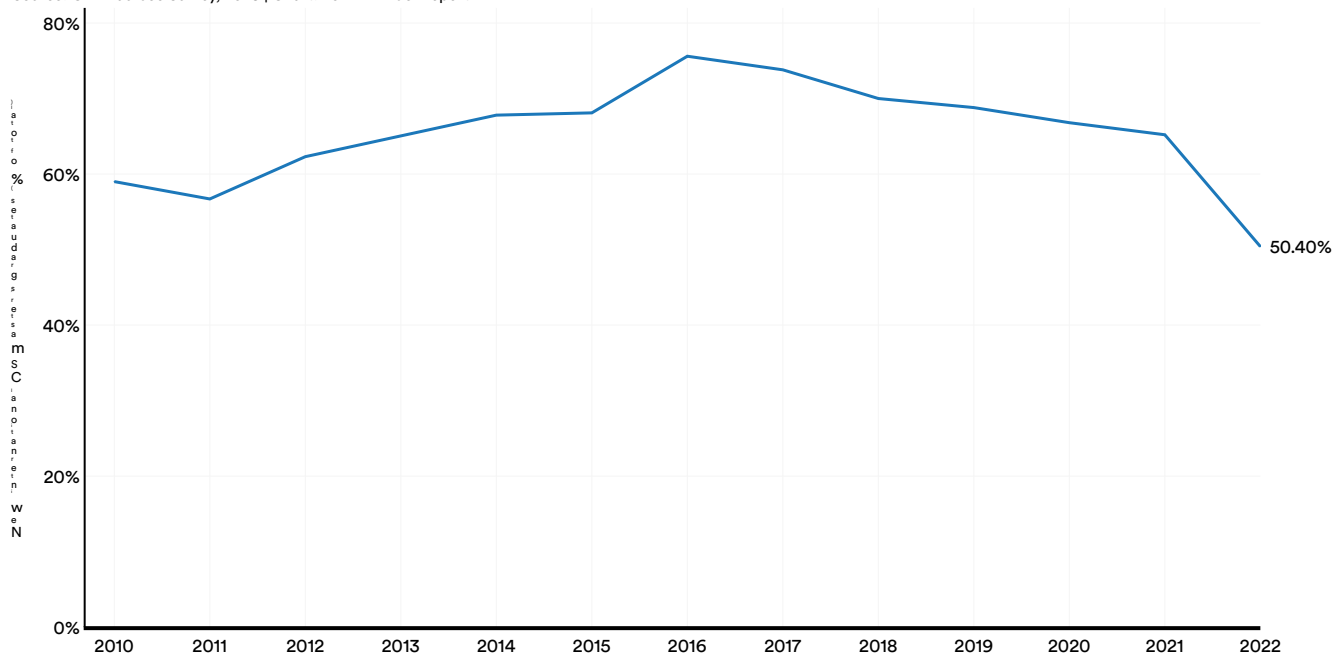


図6.1.4

計算機科学博士号卒業生

10年ぶりに、アメリカとカナダの大学における新規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2105人に達し、2010年以来最高となりました（図6.1.5）。

米国とカナダにおける2010～2022年の新規コンピュータサイエンス博士号取得者
Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

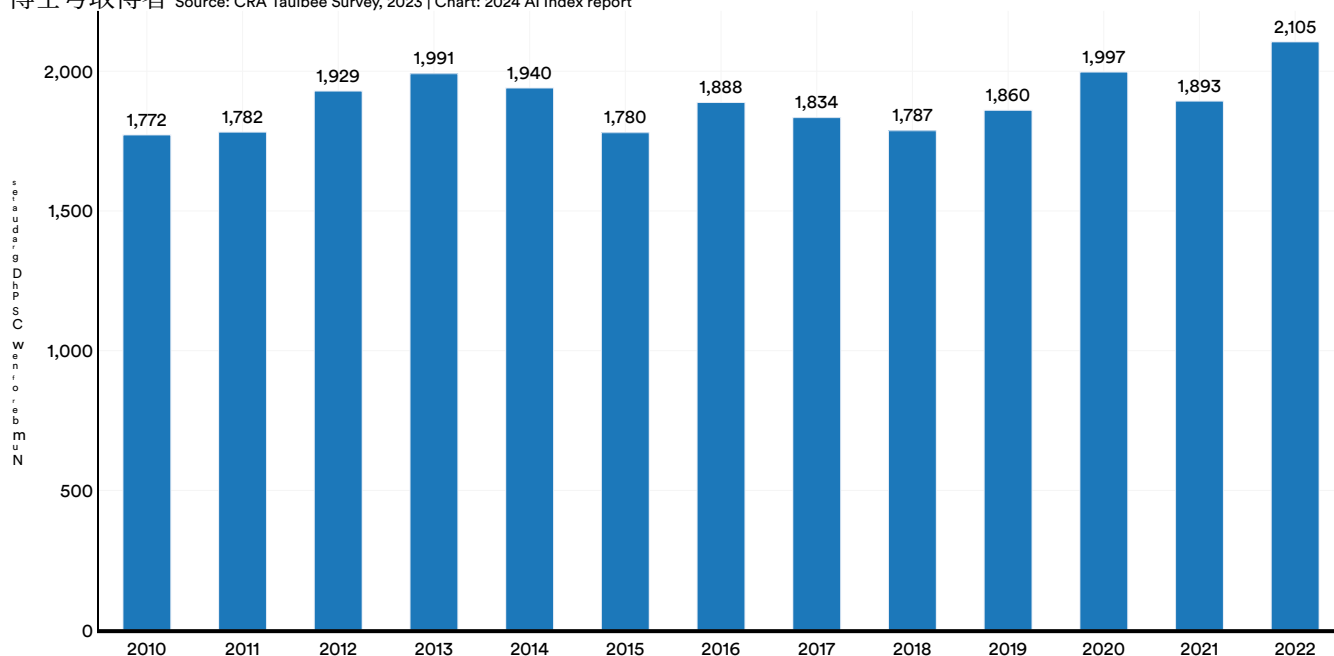


図6.1.5

過去10年間で、CS博士課程卒業生における留学生の割合は増加しましたが、昨年はわずかに減少しており、2021年の68.6%から2022年には65.9%に低下しました（図6.1.6）。

米国とカナダにおける新規国際CS博士号卒業生（総数に対する割合）、2010～22年

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

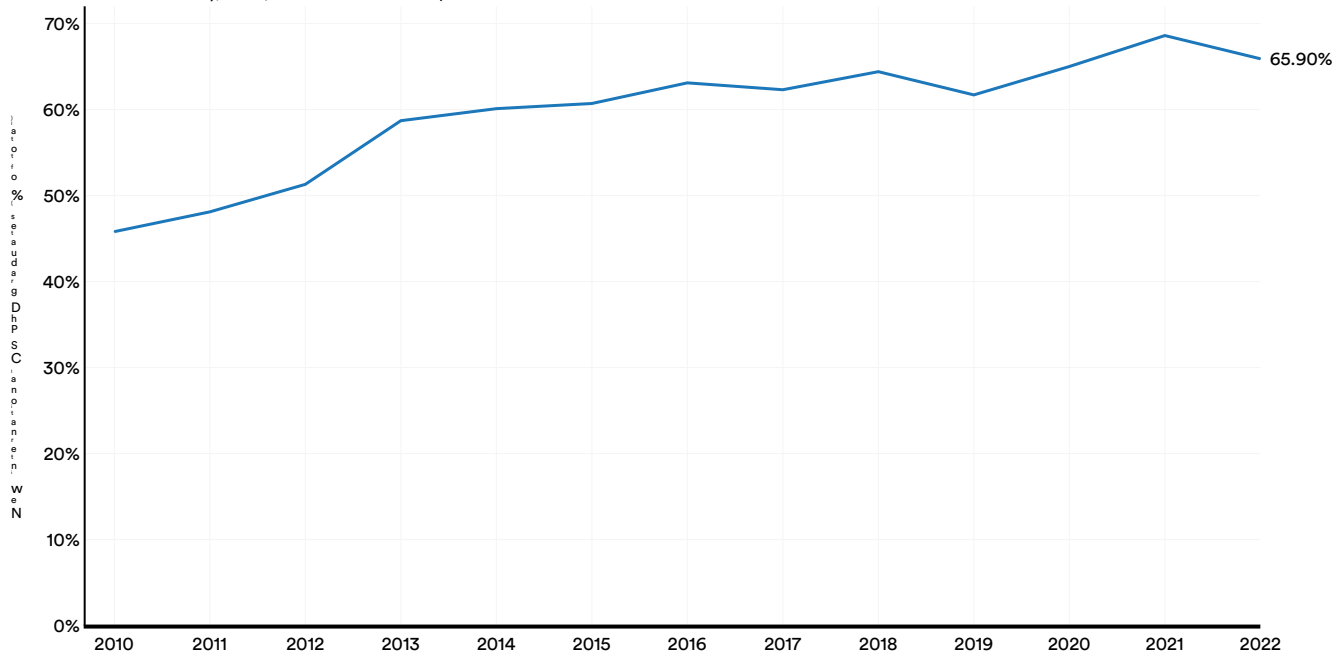


図6.1.6

AIの博士号を取得した人々は、卒業後にどこで働くことを選ぶのでしょうか？昨年のAI Indexレポートで強調された傾向に沿って、AI博士号取得者の多くが産業界でキャリアを積むようになっています（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%）でほぼ同程度の割合の人が就職していました。

しかし、2022年までに、卒業後に業界に入った割合（70.7%）は、アカデミアに入った割合（20.0%）と比べて大幅に増加しました。過去5年間、政府機関に就職するAI博士号取得者の割合は約0.7%と比較的低く安定しています。

米国とカナダにおけるセクター別の新規AI博士号取得者の雇用率（全体に対する割合）、2010～22年
Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

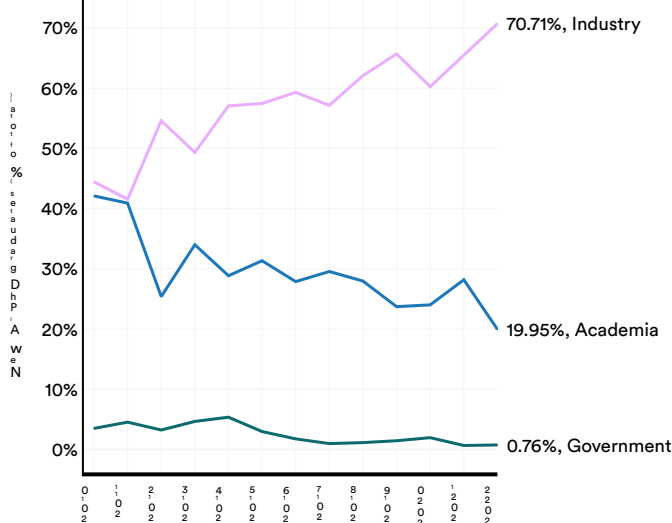


Figure 6.1.7²

米国のAI博士号取得者の雇用とカナダのセクター別統計、2010～2022年
Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

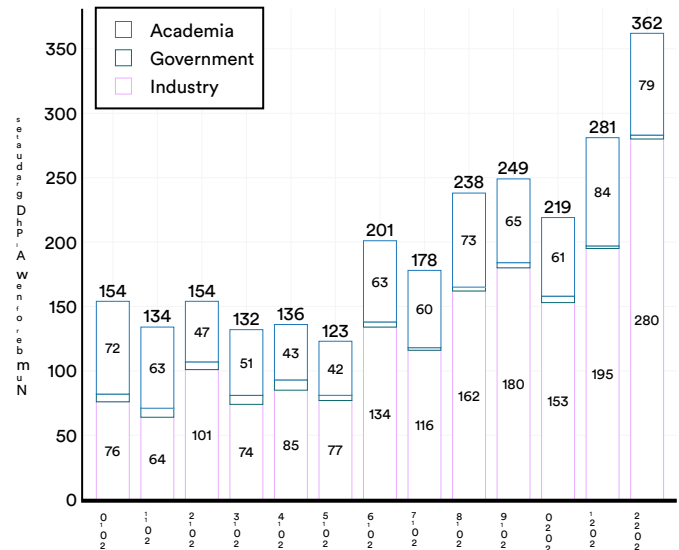


Figure 6.1.8

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、情報工学部、およびコンピュータサイエンス学部

CSとAI教育の動向をより深く理解するためには、CS学部教員のデータを見るのが有益です。昨年、アメリカとカナダの大学のCS、CE、情報学部教員の総数は7.2%増加しました（図6.1.9）。2011年以降の増加率は42.4%です。

米国とカナダにおけるコンピュータサイエンス、コンピュータ工学、情報学部教員の数の推移（2011～2022年） Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

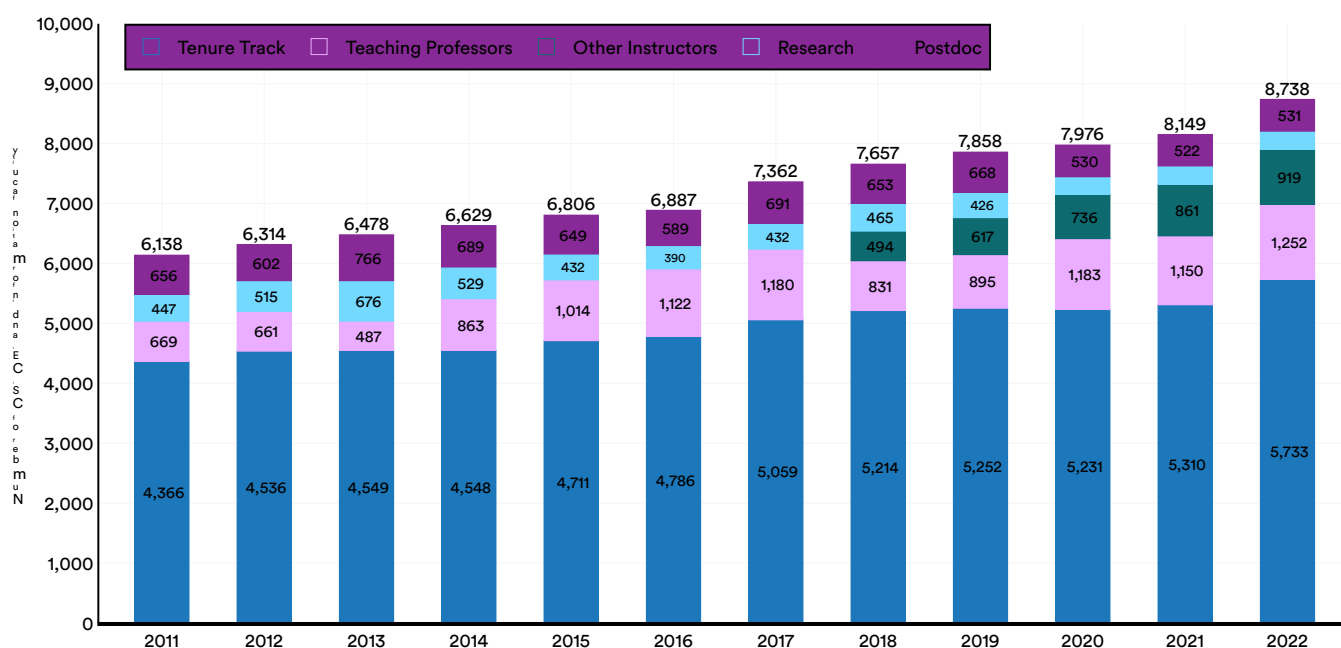


図6.1.9

2022年、米国には7084人のコンピュータサイエンス学部教員がおり、その大部分（65.7%）が終身在職権トラックにありました（図6.1.10）。米国のコンピュータサイエンス学部教員の総数は、2021年以降4.4%、2011年以降45.0%増加しています。

米国におけるコンピュータサイエンス学部教員の数の推移（2011～2022年）

出典：CRAタウビー調査、2023 | 図表：2024年AIインデックスレポート

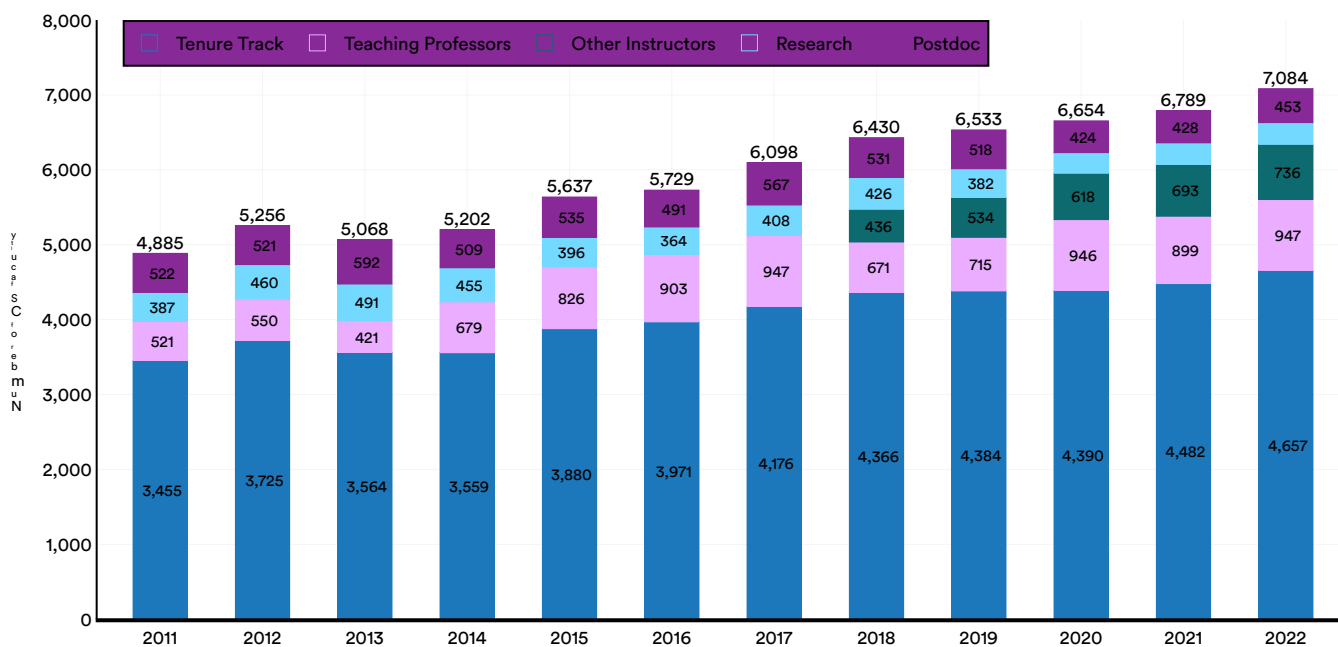


図6.1.10

北米のCS、CE、情報関連分野では昨年、過去10年間で最高となる915名の新しい教員が採用された。そのうち455のポストは終身在職権付きであった。(図6.1.11)。

米国およびカナダにおける2011～22年の新規コンピュータサイエンス、コンピュータ工学、情報学部教員採用

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

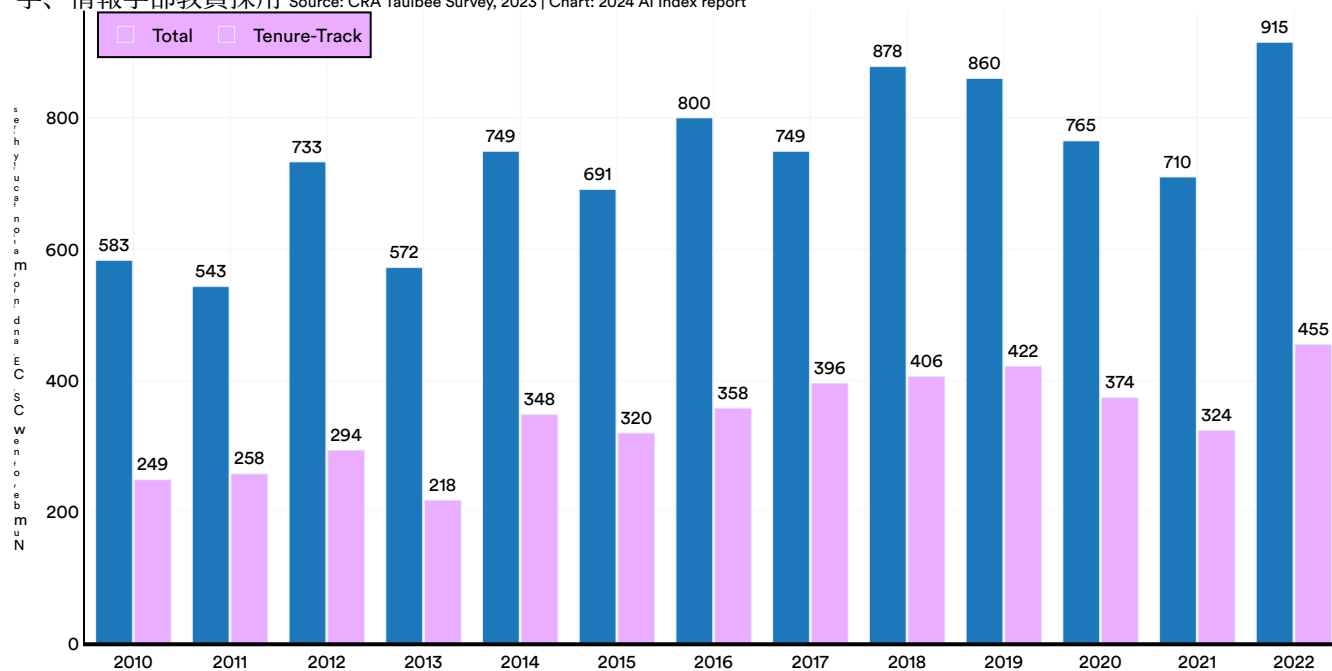


図6.1.11

2022年、新規教員の43%が他の学術職からの転任であり、これは学術界の労働力における「流動性」を示している（図6.1.12）。これらの「新規」教員は他の場所で職を空けたため、彼らの以前の役割は最終的に補充される必要がある。さらに、2022年に産業界から転じた教員の割合は、前年の11%、2019年の13%から7%に減少した。

アメリカおよびカナダのコンピュータサイエンス、コンピュータ工学、情報学部における新規教員の供給源、2018～22年 Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

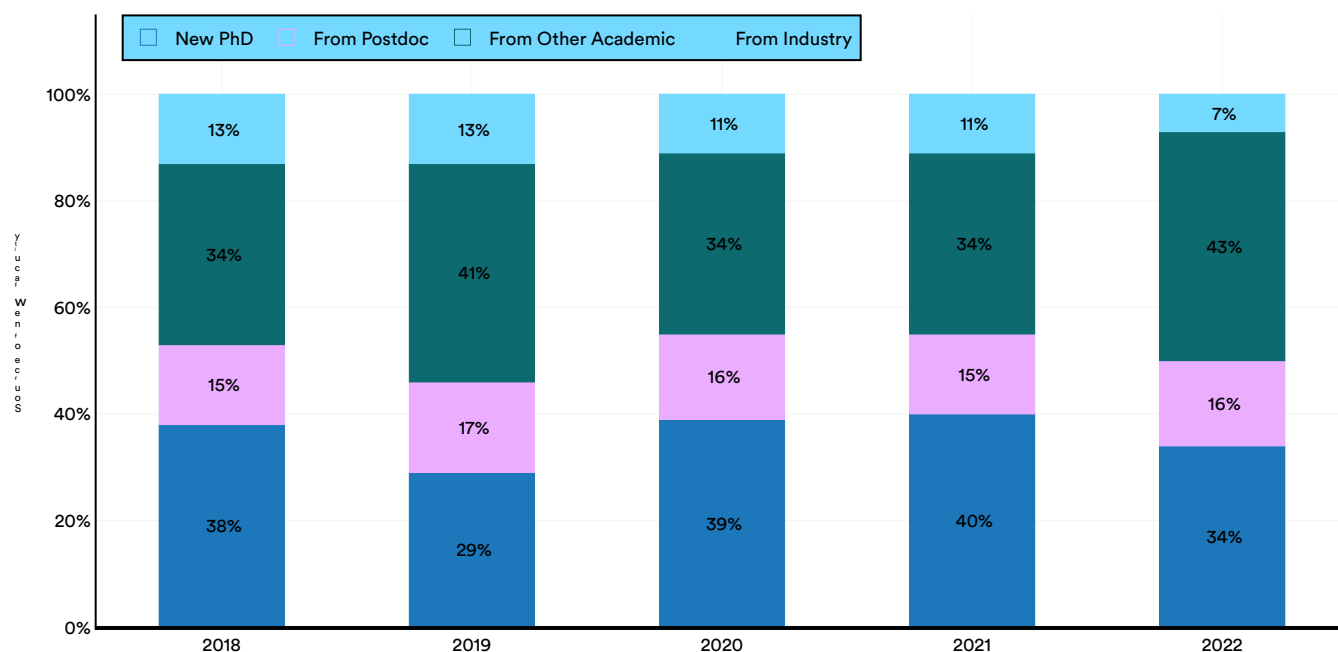


図6.1.12

過去10年間、教員ポストが埋まらない理由には様々なものがありました。2011年には、採用失敗の37%がオファーが提示されなかったことによるもので、34%が提示されたオファーが辞退されたことによるものでした（図6.1.13）。これに対し、2022年には、オファーが提示されなかったものはわずか15%にとどまり、55%はオファーが

却下された。この傾向は、新しいCS学部教員の市場競争が激化していることを反映しているように見える。しかし、これが他の学術職との競争の激化を示しているのか、それとも産業界の職との競争の激化を示しているのかは不明のままである。

2011～22年、新規のコンピュータサイエンス、コンピュータ工学、情報学部教員のポストが埋まらなかった理由（全体の％）、Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

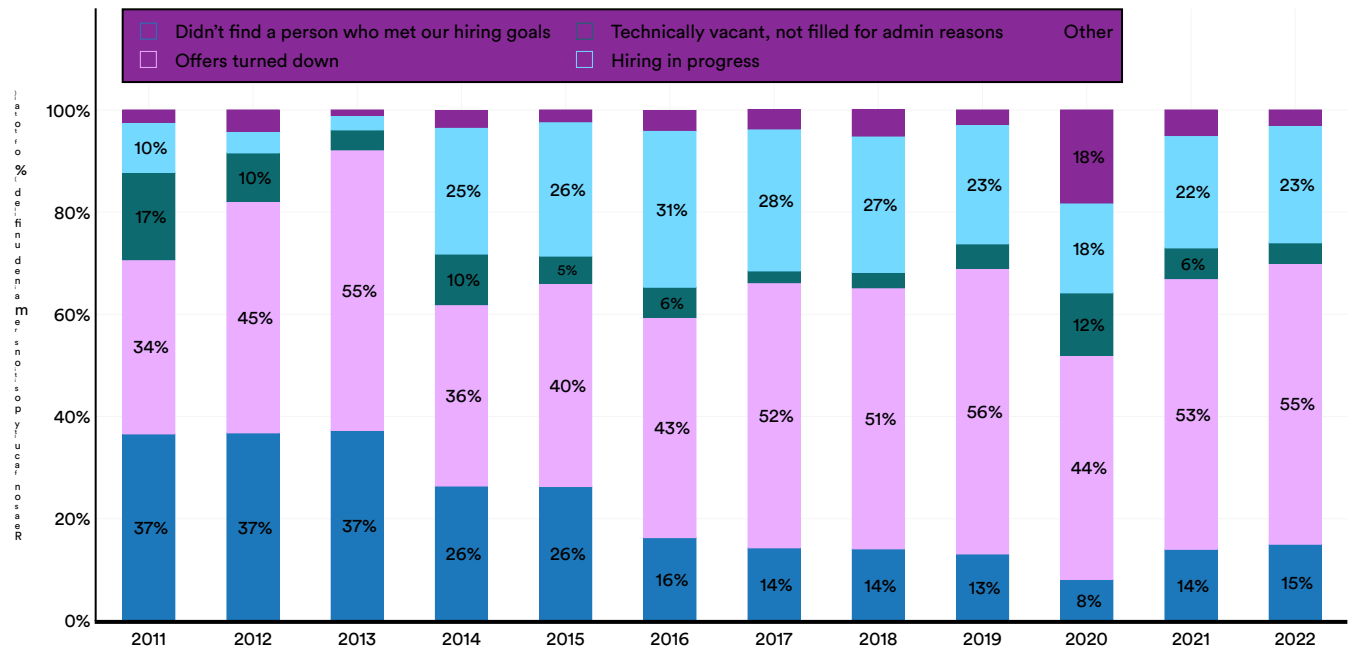


図6.1.13

2022年、北米のCS、CE、情報関連学科では、教員の退職が大幅に増加し、2021年の303名に対し405名に上った（図6.1.14）。この減少のうち、38.5%が他の学術職に、16.3%が非学術職に移った。これは前年までの傾向と一致している。

アメリカおよびカナダのコンピュータサイエンス、コンピュータ工学、情報学科における
教員数の減少、2011～22年 Source: CRA Taulbee Survey, 2023 | Chart: 2024 AI Index report

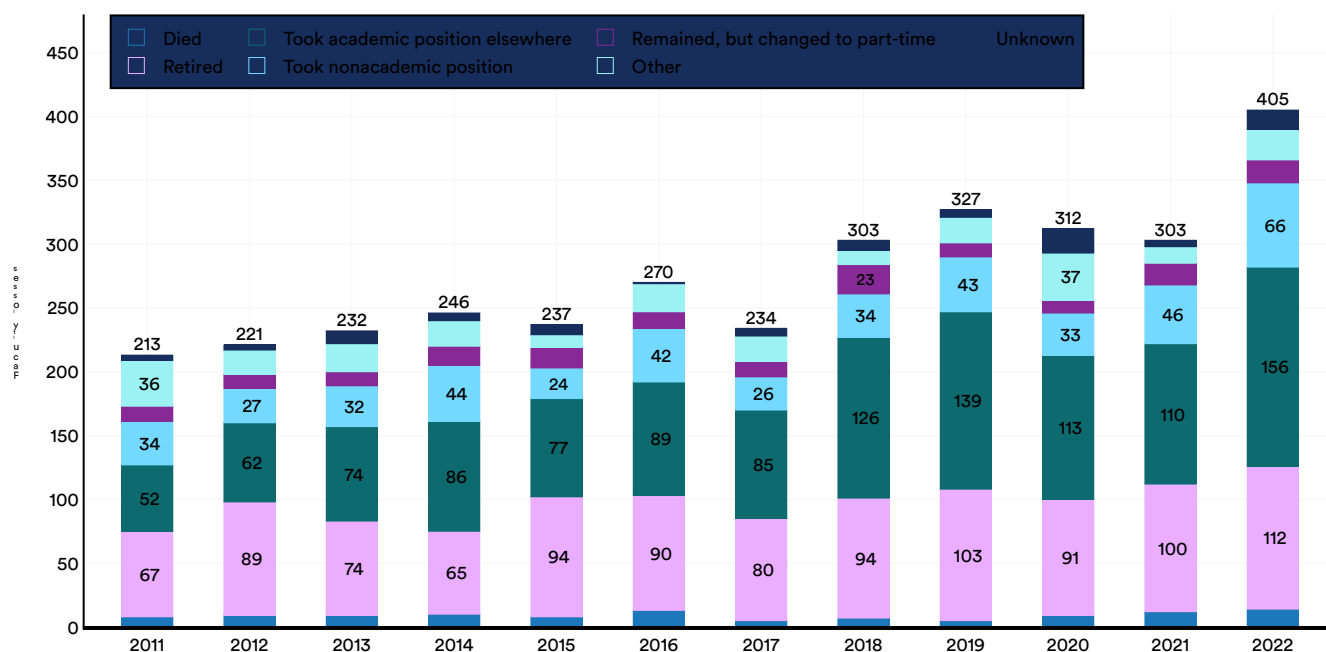


図6.1.14

2015年以降、正教授の9ヶ月間の平均給与の増加は、米国のインフレ率をわずかに下回っている一方、助教授と准教授の平均給与はインフレをわずかに上回る増加を見せている。2022年の正教授の給与は2021年より3.2%高く、これは米国の7%のインフレ率に追いつかなかったが、2015年よりは16.4%高く、それでも同期間の19%のインフレ増加を下回っている（図6.1.15）。

米国における2015～22年のCS学部教員の9ヶ月間の給与の中央値

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

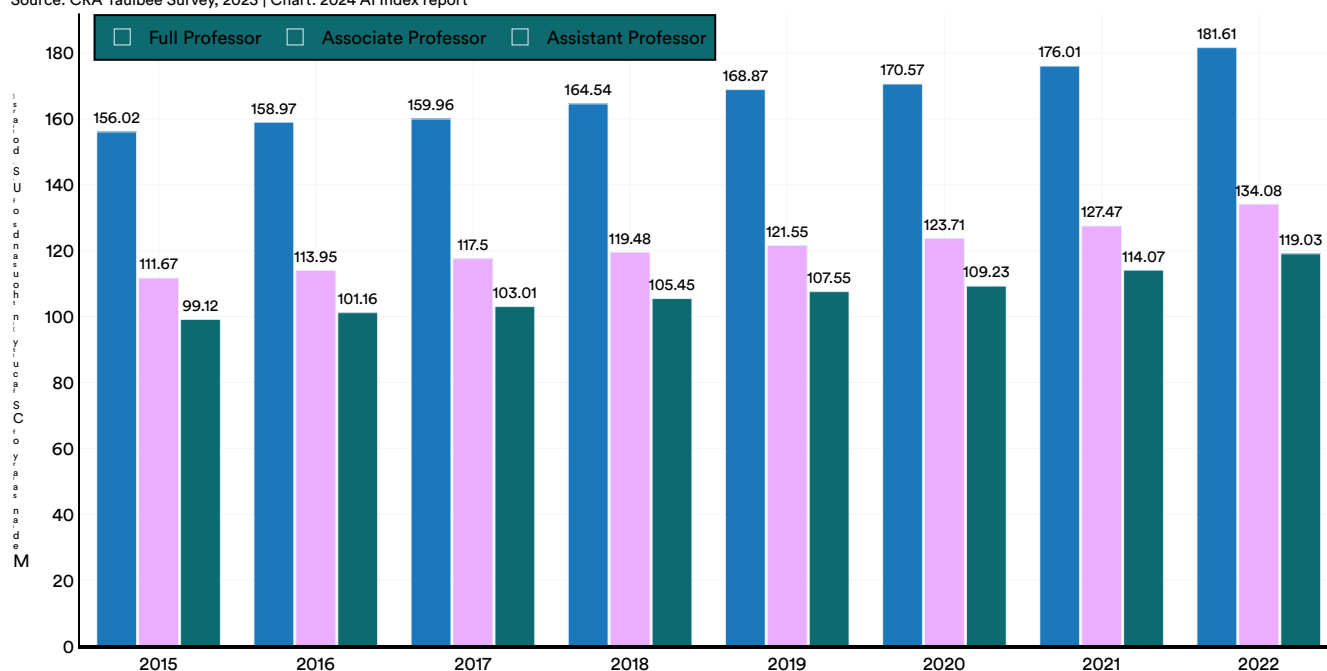


図6.1.15

2022年、CS、CE、情報関連分野の新たな任期付き教員のうち、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年のみこれを上回っています。

米国とカナダにおける2010～22年の新規国際CS、CE、情報系終身在職権付き教員の採用率（全体に対する割合）

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

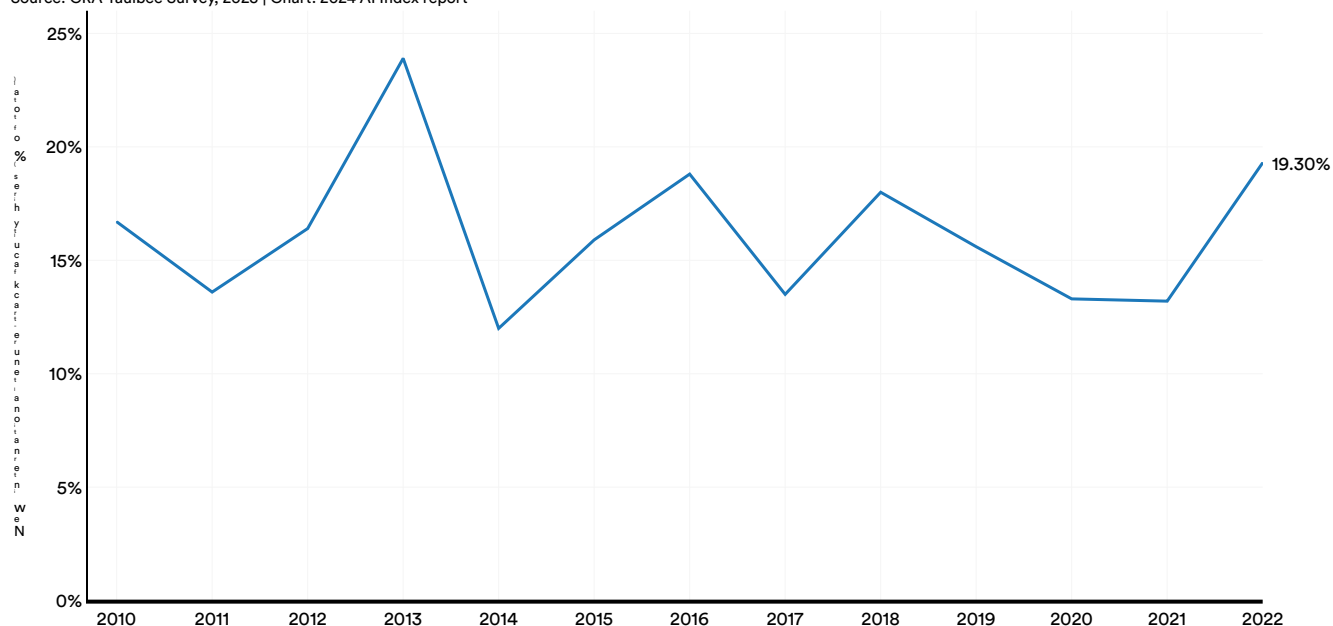


Figure 6.1.16

ヨーロッパ

欧州のコンピュータサイエンス卒業生に関するデータは、他の目標の中でも欧州における情報学教育の状況を監視する学術および研究コミュニティであるInformatics Europeから得られています。³ Informatics Europeは、欧州各国政府の統計事務所から、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の学士卒業生

2022年、英国は情報学、CS、CE、IT分野の学士号取得者の数が最も多く、約25,000人に上った（図6.1.17）。⁵ ドイツとトルコがそれに続いた。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は10年前と比べて増加したが、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、ヨーロッパ諸国の情報科学、コンピュータサイエンス、コンピュータ工学、情報技術の学士号卒業生数 Source: Informatics Europe, 2023 | Chart: 2024 AI Index report

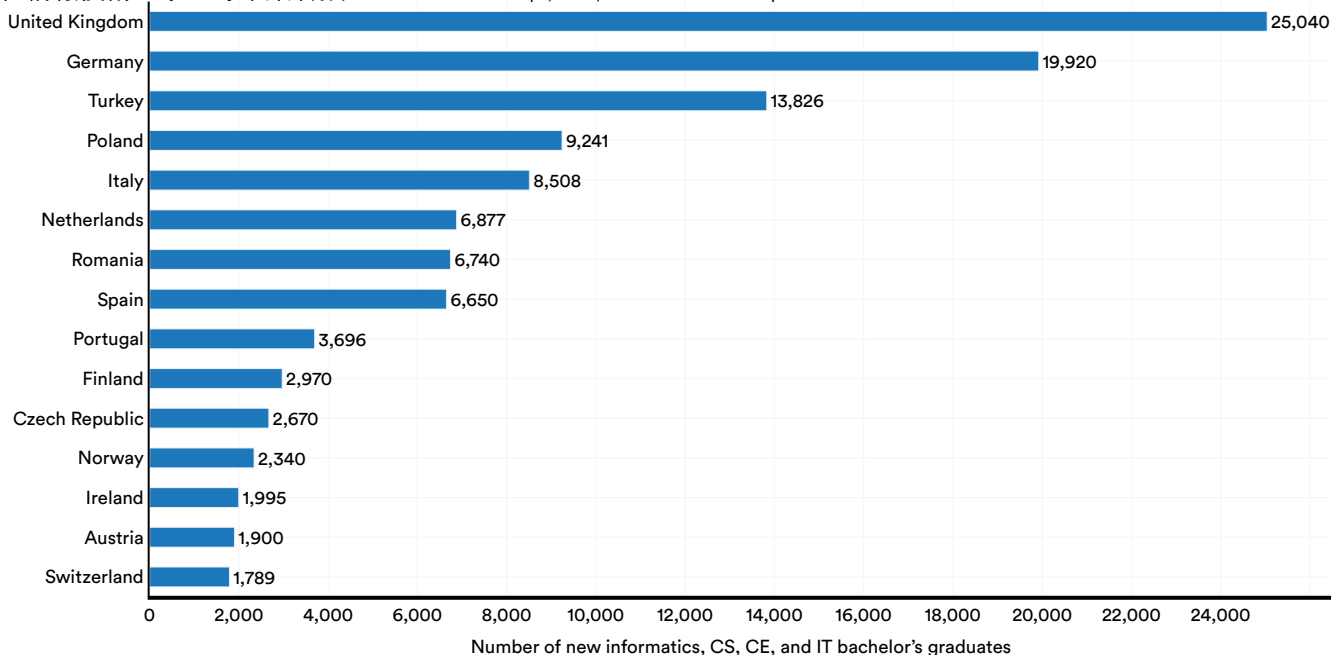


図6.1.17

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](#).

以降のセクションで詳述されているCRA北米データと欧州CS大学院データについて、1人当たりの比較を行うことに関しては、注意が必要です。欧州データは各国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るためです。

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.

ヨーロッパ諸国の情報学、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

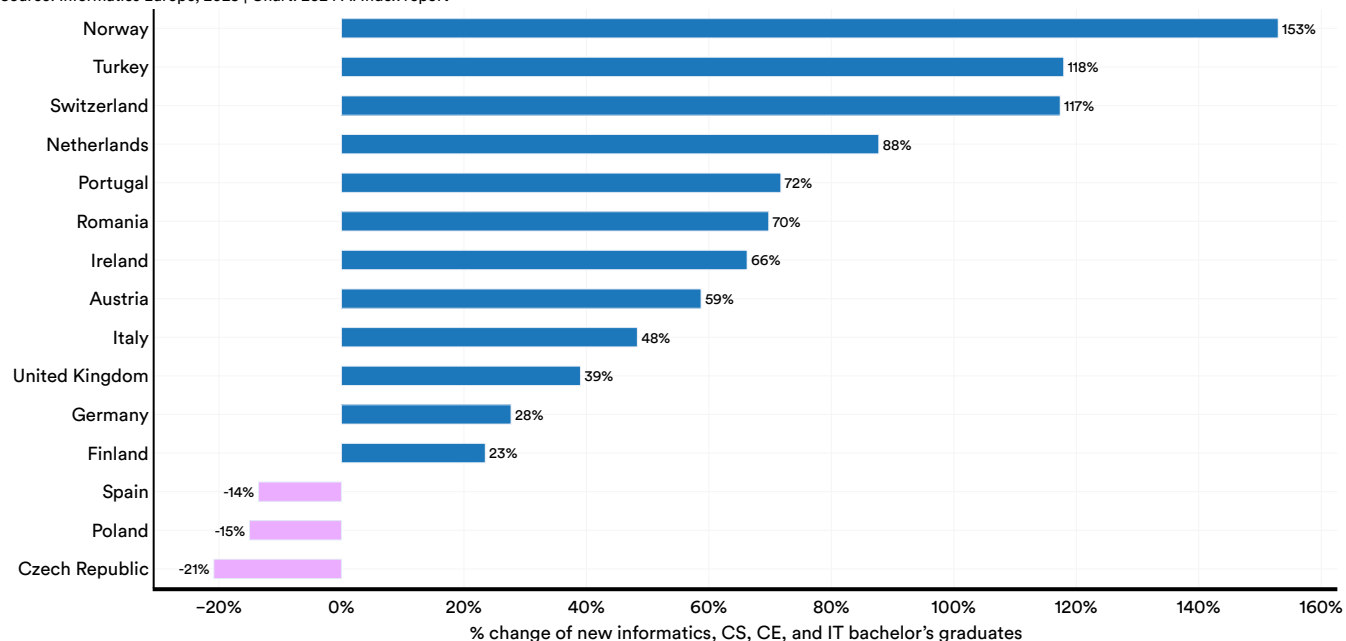


図6.1.18

フィンランド（53.4）、ノルウェー（42.6）、オランダ（38.6）は、10万人あたりの情報科学（CS、CE、IT）の新たな学士卒業生数でトップを占めています（図6.1.19）。一人当たりベースでは、サンプリングされたほとんどのヨーロッパ諸国で、情報科学（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）。