



DeepL

DeepL Proに登録すると、より大きなサイズの文書ファイルを翻訳できます。

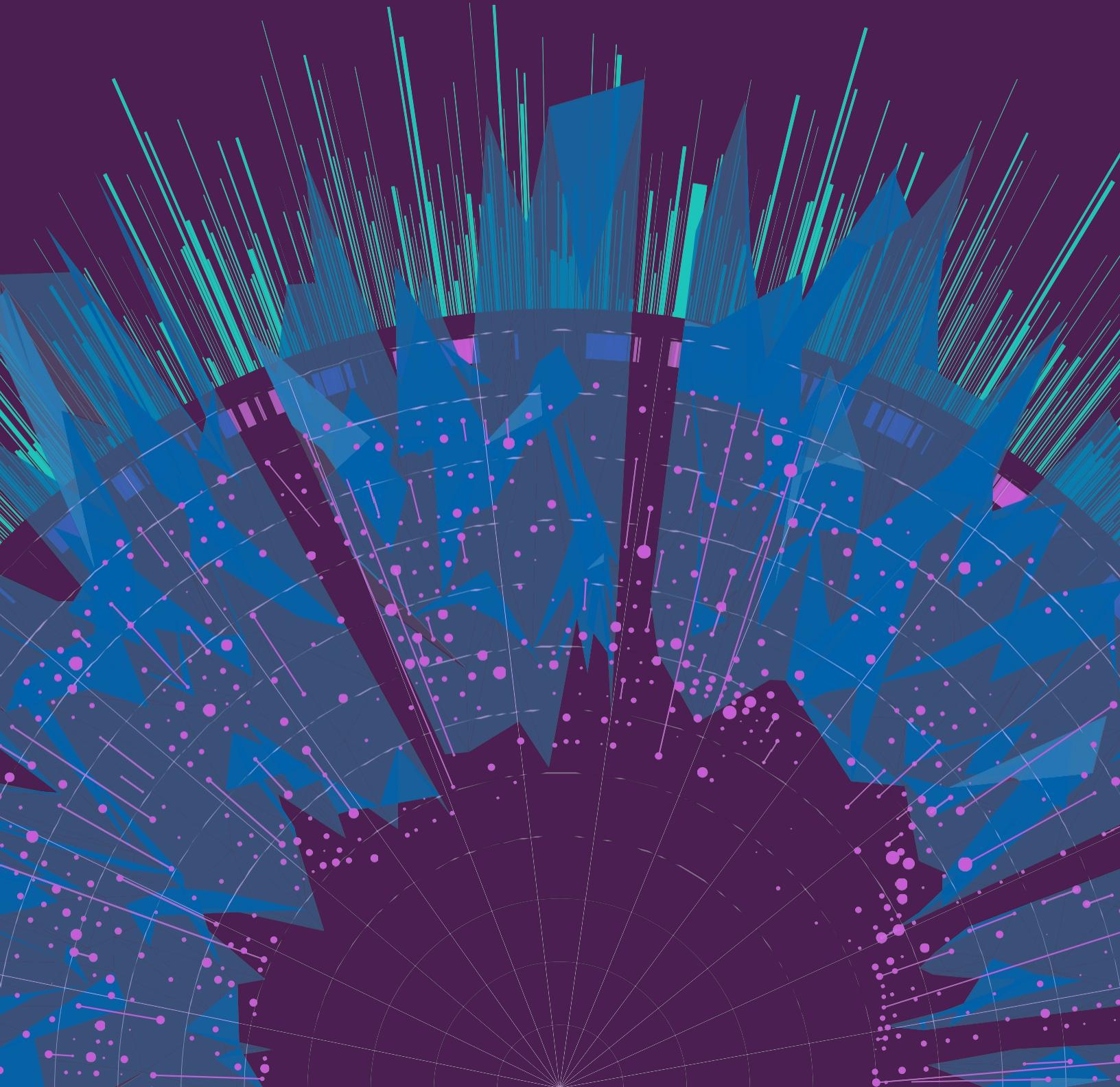
詳しくは、[www.DeepL.com/pro](http://www.DeepL.com/pro) をご覧ください。

LHII

人工知能指数レポート  
2024

第5章：

## 科学と医学





# レビュー

概要	298
チャプター・ハイライト	299
<b>5.1 顕著な科学的マイルストーン</b>	<b>300</b>
アルファ開発	300
フレキシキューブ	301
シンボット	303
グラフキャスト	304
グノミー	305
洪水予測	306
<b>5.2 医療におけるAI</b>	<b>307</b>
注目の医療システム	307
シンセSR	307
結合型プラスモニック赤外線センサー	309
イヴ・スケープ	310
アルファミッセンス	312
ヒトパンゲノムリファレンス	313
臨床知識	314
メドQA	314
注目の研究GPT-4メドプロンプト	315
注目の研究メディトロン-70B	317
診断	318
注目の研究CoDoC	318
注目の研究CT/パンダ	319
その他の診断用途	320
FDA承認のAI関連医療機器	321
管理とケア	323
注目の研究メドアライン	323

公開データにアクセスする



## 概要

今年のAIインデックスでは、科学と医学の発見におけるAIの役割の拡大を認識し、科学と医学におけるAIの章を新設した。この章では、GraphCastのような高度な天気予報システムなど、AIによって科学的に促進された2023年の傑出した成果を探る。

やGNOMEのような材料探索アルゴリズムの改良などである。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AI主導の重要な医療イノベーション、FDAのAI関連医療機器の承認動向についても検証する。





## ハイライト

**1. AIのおかげで科学の進歩がさらに加速。** 2022年、AIは科学的発見を促進し始めた。しかし2023年には、アルゴリズムによるソートをより効率的にするAlphaDevから、材料発見のプロセスを促進するGNoMEまで、科学関連のAIアプリケーションのさらなる重要性が発表された。

**2. AIは医療を大きく前進させる。** 2023年には、パンデミック予測を強化するEVEscapeや、AIによる突然変異分類を支援AlphaMissenceなど、いくつかの重要な医療システムが立ち上げられた。AIは医療の進歩を推進するためにますます活用されるようになっている。

**3. 高度な知識を持つ医療AIが登場した。** ここ数年、AIの臨床知識を評価するための重要なテストであるMedQAベンチマークにおいて、AIシステムは目覚ましい向上を示している。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ポイント上昇した。2019年にこのベンチマークが導入されて以来、MedQAにおけるAIの成績は約3倍に上昇した。

**4. FDAはAI関連医療機器をどんどん承認している。** 2022年、FDAは139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はますます現実の医療目的に使用されるようになっている。

このセクションでは、AIインデックス運営委員会が選んだ2023年のAI関連の重要な科学的ブレークスルーを紹介する。

## 5.1 注目すべき科学のマイルストーン

### アルファ開発

**AlphaDevはより高速なソートアルゴリズムを発見** AlphaDevは新しいAI強化学習システムで、計算アルゴリズム強化の分野における科学者やエンジニアの数十年にわたる研究を改良したものです。AlphaDevは、既存の人間のベンチマークよりも少ない命令数で、以下のような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（図5.1.1）のような短いシーケンスに関する基本的なソートアルゴリズム。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る。これは強化学習を使って設計された最初の追加である。

### アルゴリズム長を最適化した場合のAlphaDevと人間のベンチマークとの比較

出典: マンコヴィツ他、2023年 | 図表: 2024年AI指数レポート

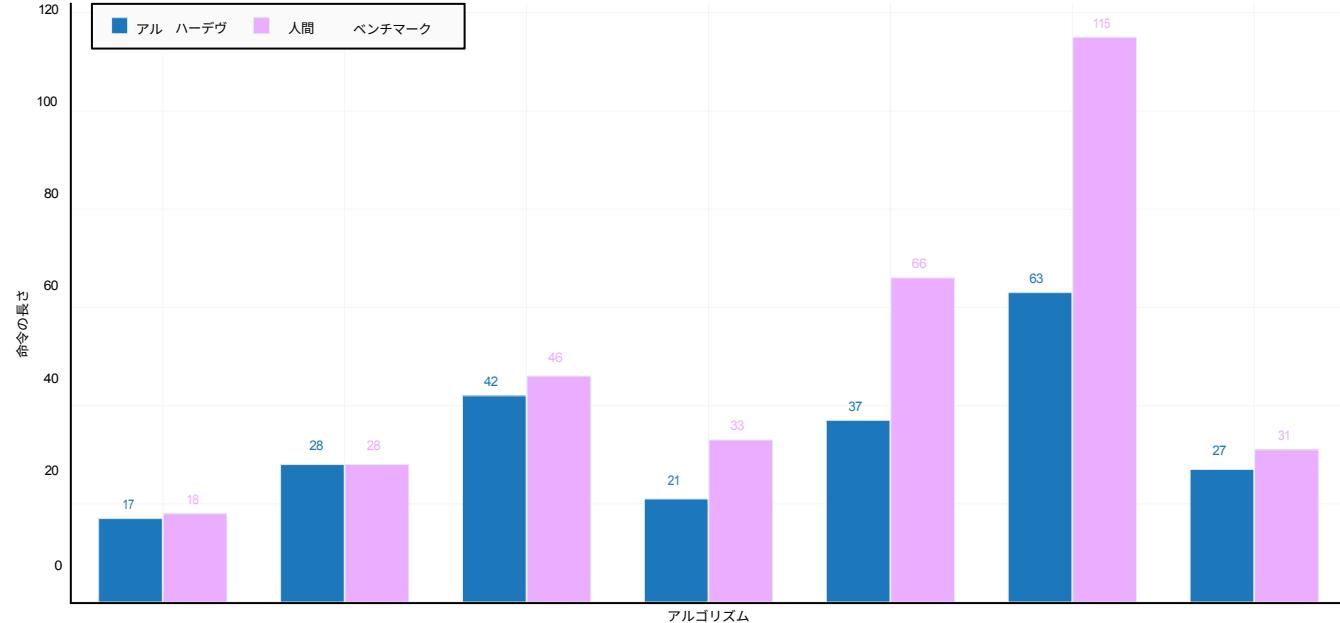


図5.1.1

## 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、3Dオブジェクトを定義するための頂点、辺、面のメッシュを作成するもので、コンピュータグラフィックスにおいて非常に重要である。ビデオゲームでは重要な役割を果たす。

アニメーション、医療用画像処理、科学的な可視化などです。従来の等価面抽出アルゴリズムは、限られた解像度、構造的な剛性、数値的な不安定性にしばしば苦戦し、その結果、以下のような影響を及ぼしている。

の品質を向上させることができます。FlexiCubesは、勾配ベースの最適化と適応可能なパラメータにAIを採用することで、これらの制限のいくつかに対処しています（図5.1.2）。この方法により、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等価面を利用する他の代表的な手法と比較して、FlexiCubesは基礎となるグランドツールスにはるかに近いメッシュ抽出を実現します（図5.1.3）。

### FlexiCubesサーフェス再構成サンプル

ソースNvidia、2023年

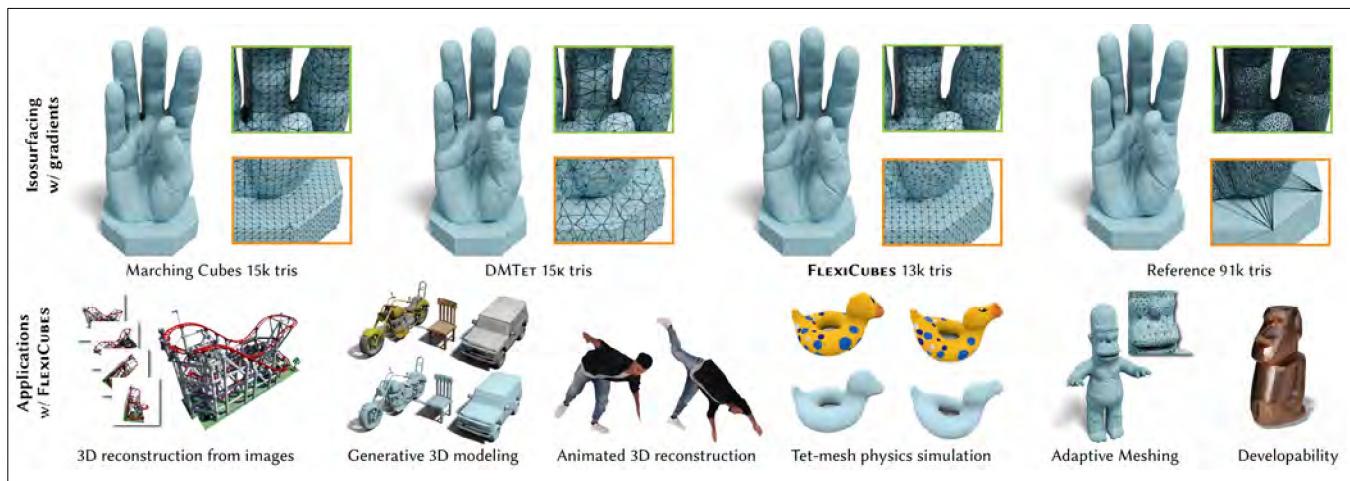


図5.1.2



### 3Dメッシュ再構築に関する定量的な結果を選択する

出典:Shen et al., 2023 | 図表: 2024年AI指数レポート

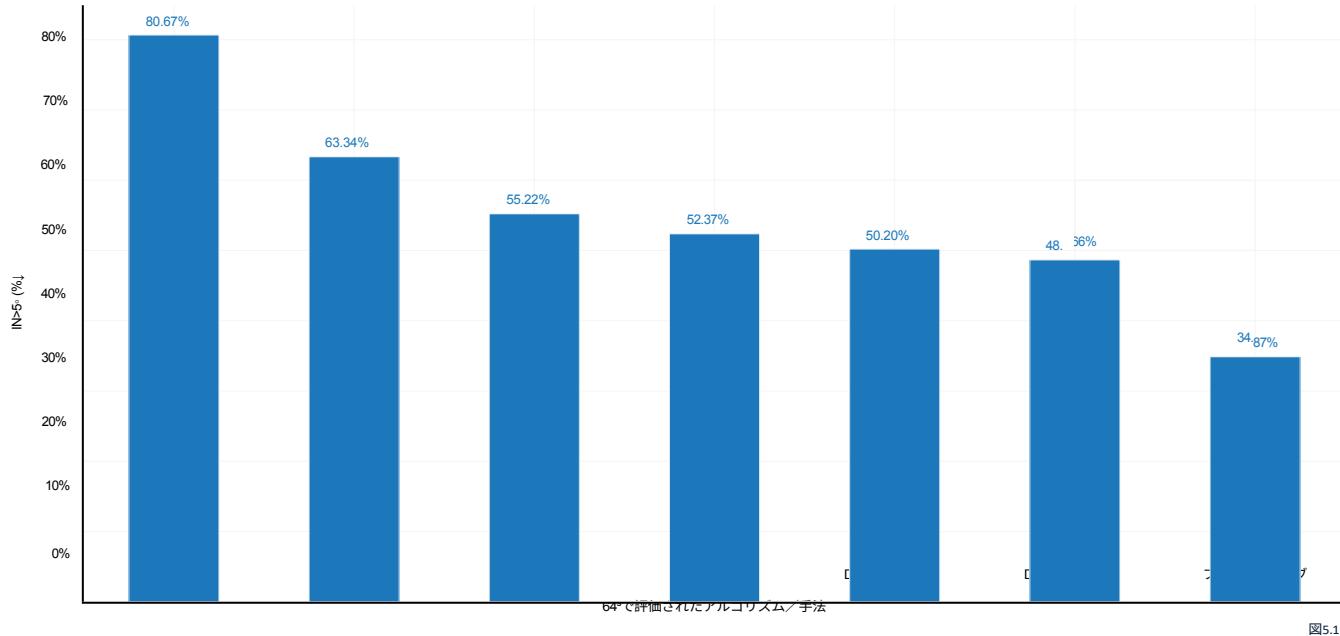


図5.1.3

## シンボット

**AI-driven robotic chemist for synthesizing organic molecules Synbot**は、化学合成を計画するAIソフトウェア層、コマンドを翻訳するロボットソフトウェア層、実験を行う物理ロボット層からなる多層システムを採用している。AIとロボット間のクローズドループ・フィードバックメカニズムにより、シンボットは合成レシピを開発することができる。

または確立された基準値を上回る（図5.1.4）。図5.1.4）。

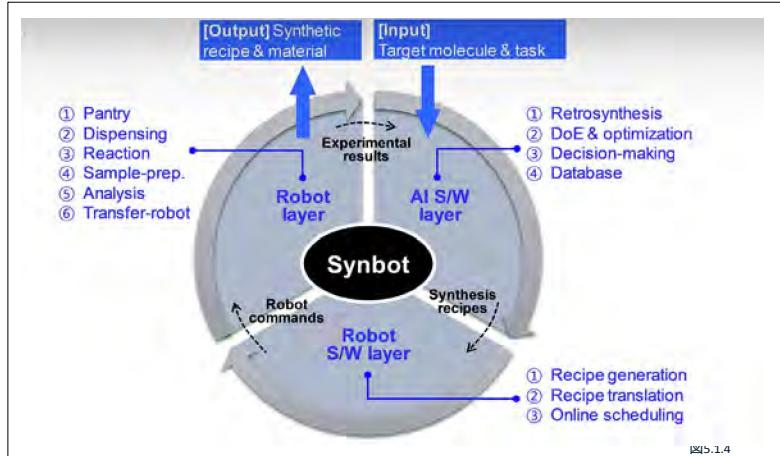
M1[4-(2,3-ジメトキシフェニル)-1H-ピロロ[2,3-b]ピリジン]を合成するために、シンボットは複数の合成式を開発した。

を上回るコンバージョン収率を達成した。

## シンボットのデザイン

イン

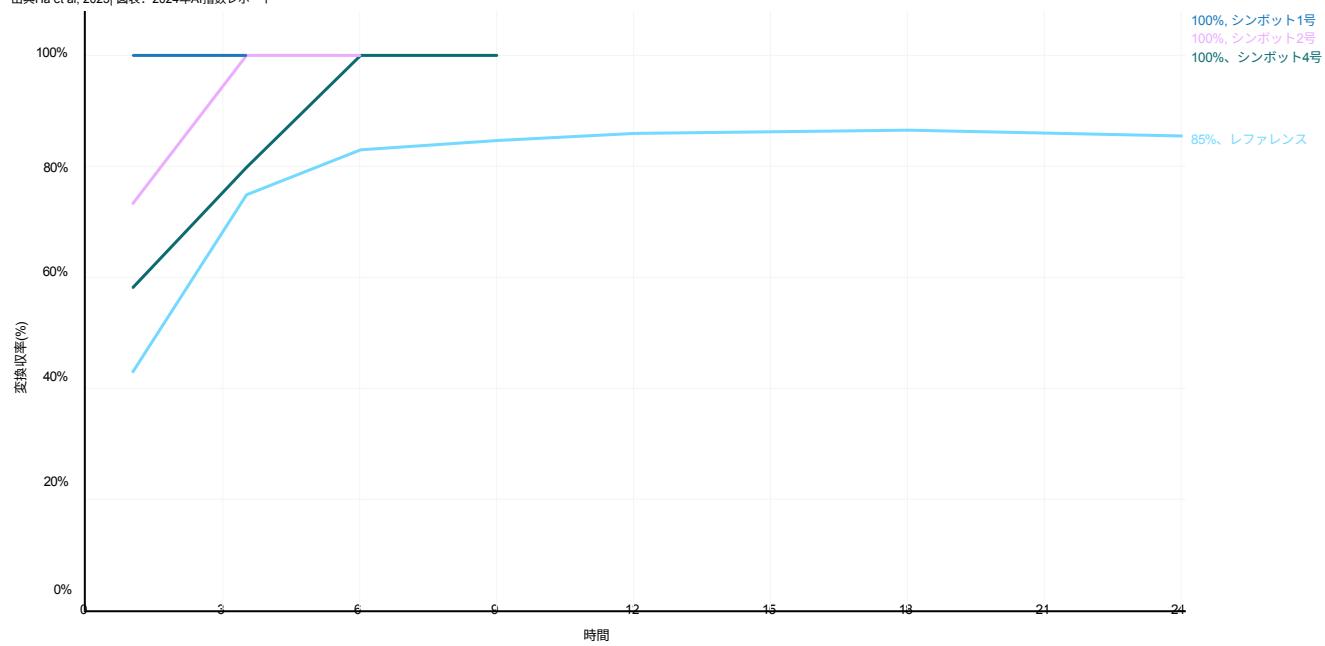
出典ハ他、2023年



を達成し、合成時間を大幅に短縮した（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る。

## M1自律最適化実験の反応カイネティクス、Synbotとリファレンスの比較。

出典Ha et al, 2023 | 図表: 2024年AI指数レポート



## グラフキャスト

### GraphCastでより正確な世界の天気予報を

GraphCastは、高精度の10日間天気予報を1分以内に提供する新しい天気予報システムである（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、膨大なデータセットを処理して気温、速、大気の状態を予測する、

などがあります。図5.1.6、GraphCast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High Resolution Weather Simulation Systemの性能を比較したものです。予報（HRES）。GraphCastは二乗平均平方根誤差が小さく、その予測は観測された気象パターンとより密接に対応しています。GraphCastは、気象パターンを読み解き、異常備えを強化し、地球規模の気候研究に貢献する貴重なツールとなります。

## グラフキャスト天気予報

出典ディープマインド、2023年

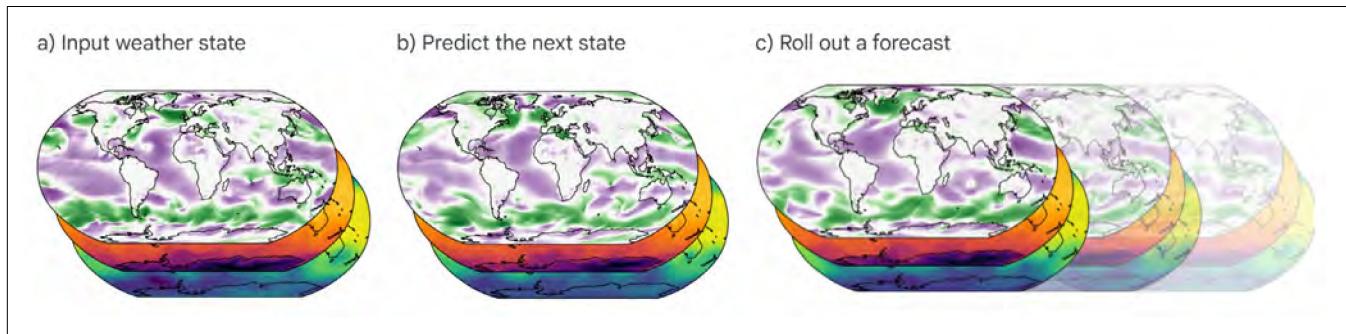


図5.1.6

### 10日間のz500予報スキル：グラフキャストとHRESの比較

出典Lam 他、2023年| 図表：2024年AI指数レポート

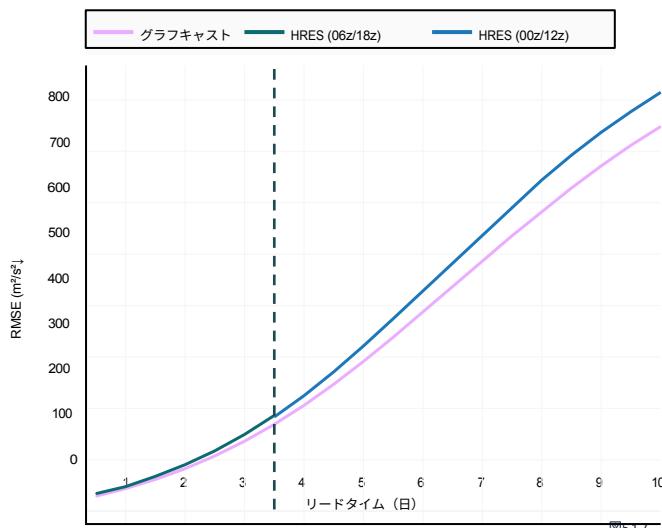


図5.1.7

## グノミー

## GNOMEで新素材を発見する

新しい機能性材料の探索は、ロボット工学や半導体製造など、さまざまな科学分野の進歩の鍵を握っている。しかし、この発見プロセスは一般的に高価で時間がかかる。グーグルの研究者たち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訓練された場合、このプロセスを迅速化できることを実証した。彼らのモデル、GNOMEは、安定した結晶の数を大幅に特定することで、材料探索の主要な手法であるMaterials Projectを凌駕した（図5.1.8）。GNOMEは220万個の新しい結晶構造を発見したが、その多くは人間の研究者が見落としていたものである（図5.1.9および図5.1.10）。GNOMEのようなAI主導型プロジェクトの成功は、科学的ブレークスルーを加速するデータとスケーリングの力を浮き彫りにしている。

## サンプルの素材構造

出典マーチャント他、2023年

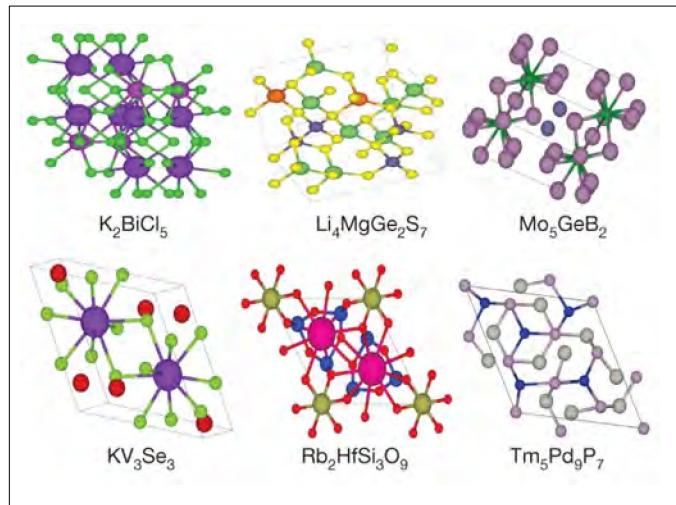


図5.1.8

## GNOMEとマテリアル・プロジェクトの比較：安定した結晶数

出典マーチャント他、2023年| 図表：2024年AI指数レポート

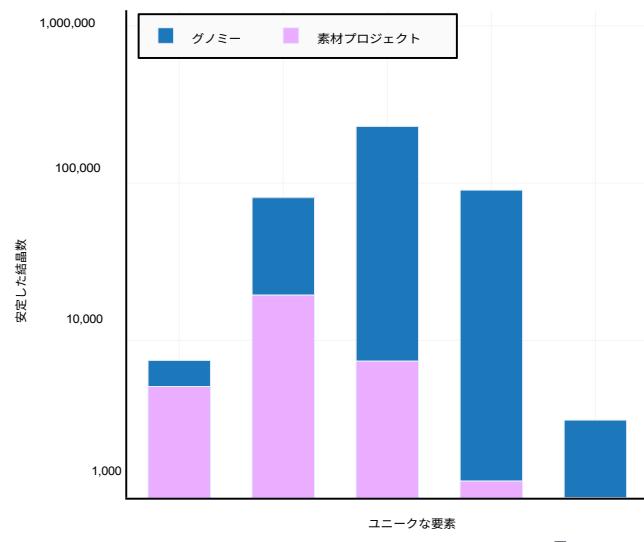


図5.1.9

## GNOME対マテリアル・プロジェクト：明確なプロトタイプ

出典マーチャント他、2023年| 図表：2024年AI指数レポート

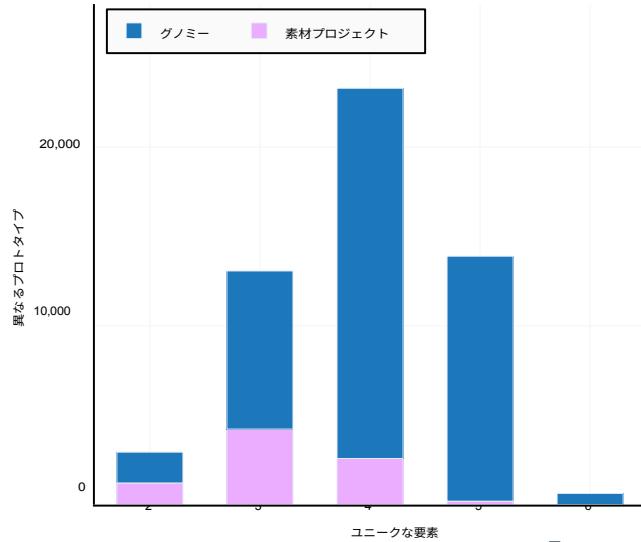


図5.1.10

## 洪水予測

より正確で信頼性の高い洪水予測を可能にするAI 2023年に導入された新たな研究は、大規模な洪水発生を予測する上で大きな進歩を遂げた。最も一般的な自然災害のひとつである洪水は、予防や軽減のためのいない後進国において、特に壊滅的な影響を及ぼす。そのため、このような事象を事前に予測できる、より正確な予測手法を開発することは、大きな好影響をもたらす可能性がある。

グーグルの研究チームがAIを使って高精度の水文学シミュレーションモデルを開発した。

これらの革新的な手法は、GloFASのような現在の最先端モデルと同等以上の精度で、特定の異常洪水事象を最大5日前まで予測することができます<sup>1)</sup>。このAIモデルは、様々な再現期間事象において、優れた精度（正の予測の精度）と再現率（関連するすべての事例を正しく特定する能力）を示し、現代の主要な手法を凌駕している（図5.1.11）<sup>2)</sup>。このモデルはオープンソースであり、すでに80カ国以上で洪水事象の予測に利用されている。

## AIモデルとGloFASの回帰期間別予測比較

出典:ニーリング他、2023年| 図表: 2024年AI指数レポート

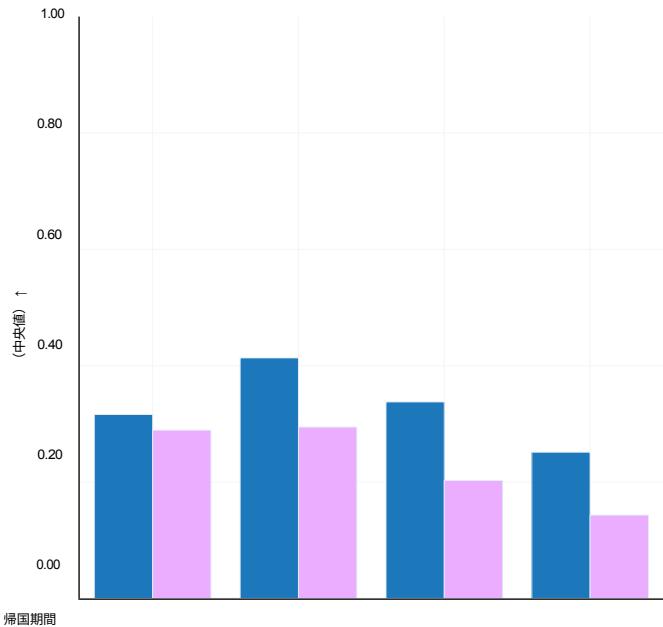
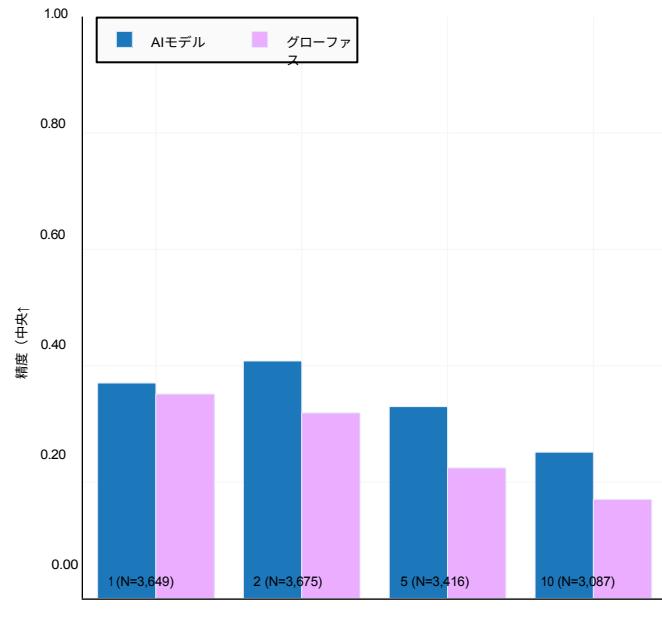


図5.1.11

1)未計測流域とは、水文学的流れをモデル化するための流況データが不十分な流域のことである。

2)返還期間（再来間隔）は、特定の水文学的事象が特定の期間内に再来する可能性を測定する。例えば、100年洪水は、任意の年にその事象が等しいかまたはそれを超える可能性が1%あることを意味する。

ポリープの検出から臨床医の診断支援まで、AIモデルの医療における価値はますます高まっている。AIの性能が向上し続ける中、医療現場への影響をモニタリングすることができます。本セクションでは、2023年に導入されるAI関連の重要な医療システム、臨床AIの知識の現状、そしてAIを強化することを目的とした新たなAI診断ツールやモデルの開発について紹介する。

病院経営。

## 5.2 医療におけるAI

### 注目の医療システム

このセクションでは、AIインデックス運営委員会が選んだ2023年のAI関連の重要な医療ブレークスルーを特定する。

#### シンセSR

高度な解析のために脳スキャンを変換 SynthSRは臨床脳スキャントを高解像度のT1強調画像に変換するAIツールである（図5.2.1）。この進歩により、これまで先端研究において多くのスキャン画像の利用を制限していたスキャン画質のばらつきの問題が解決される。SynthSRは、これらのスキャンを、コントラストが高く、脳の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、詳細な3D脳レンダリングの作成を容易にする。SynthSRを使った実験では、スキャンレベルと被験者レベルの両方で、観察された体積の間に確固とした相関があることが実証されており、SynthSRが高解像度のT1スキャンで生成された画像に酷似した画像を生成することが示唆されている。図5.2.2は、SynthSRスキャンが選択された脳領域において、グランドトゥルースの観察とどの程度一致するかを示している。SynthIDは脳構造の可視化と解析を大幅に改善し、神経科学研究と臨床診断を促進する。

#### シンセSR世代

出典イグレシアス他、2023年

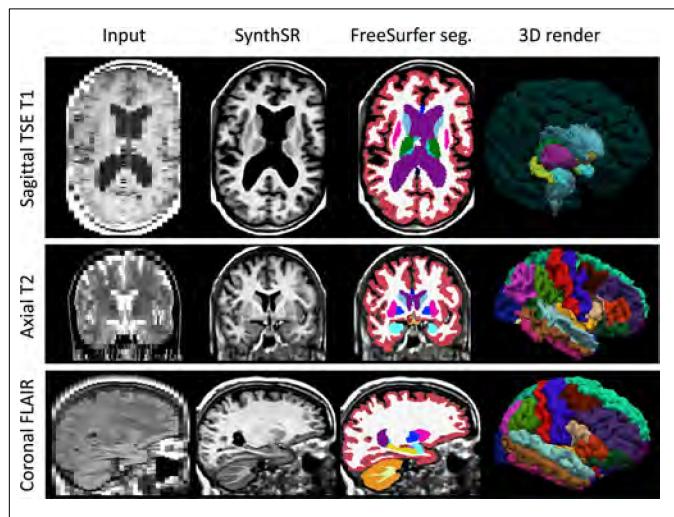


図5.2.1

### 特定の脳領域におけるSynthSRと地上真理ボリュームとの相関性

出典イグレシアス他、2023年| 図表：2024年AI指数レポート

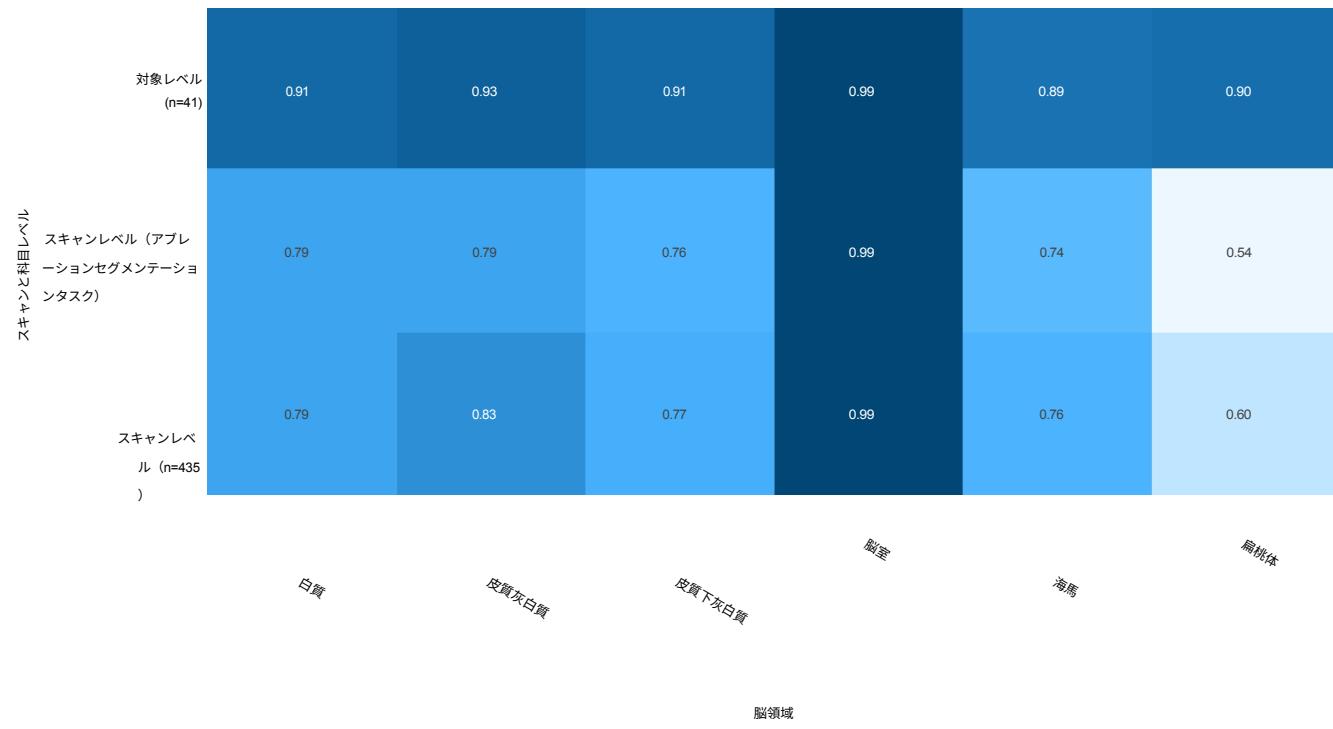


図5.2.2

### 結合型プラズモニック赤外センサー 神経変性疾患の検出のための結合型プラズモニック赤外センサー 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かかっている。質量分析法やELISA法のような従来の方法は、タンパク質レベルの定量に重点を置くことができる点で有用であるが、構造状態の識別することはできない。今年、研究者たちは神経変性疾患診断のための新しい方法を発見した。

図5.2.3) は、表面増強赤外吸収 (SEIRA) 分光法を用いたAI結合型プラズモニック赤外センサーとイムノアッセイ技術 (ImmunoSEIRA) を組み合わせたものである。実際のフィブリル率とAIシステムによる予測値を比較したテストでは、予測値の精度は実際に報告されたフィブリル率と非常に近いことがわかった (図5.2.4)。

### ImmunoSEIRAの検出原理とセットアップ

出典Kavungalら、2023年

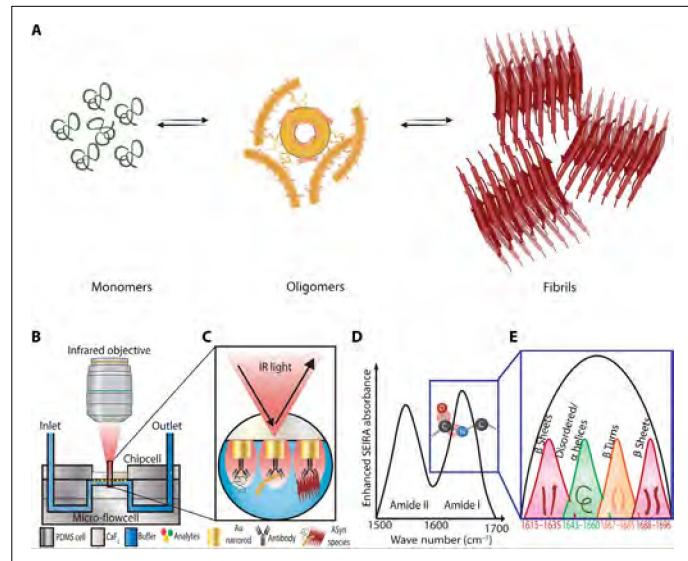


図5.2.3

### ディープニューラルネットワークが予測したテストサンプルのフィブリル率と実際のフィブリル率

出典Kavungal他、2023年 | 図表：2024年AI指数レポート

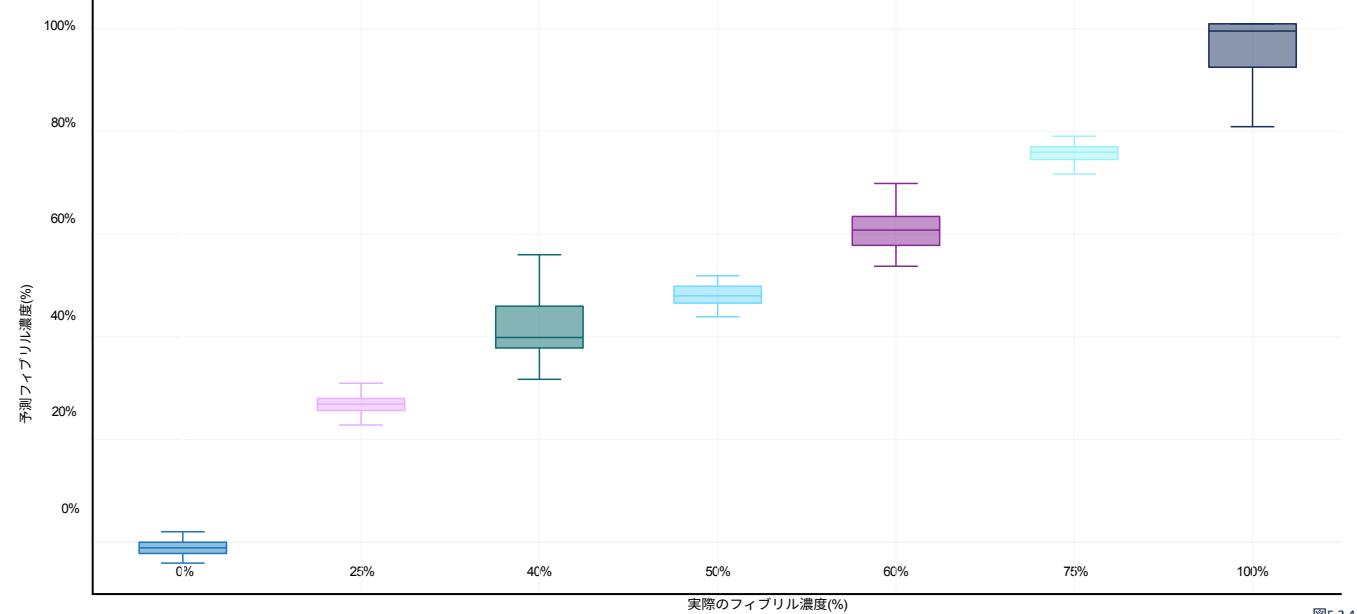


図5.2.4

## イヴ・スケープ

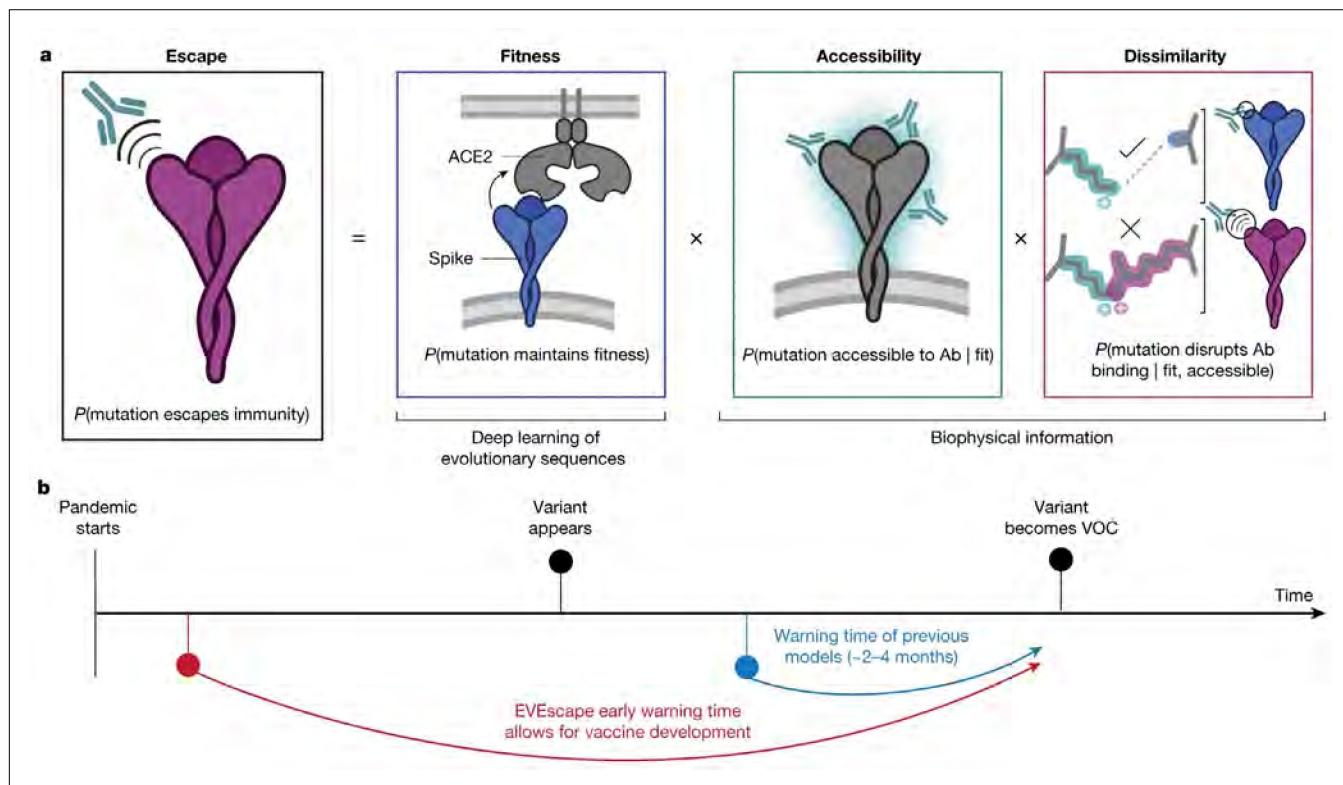
### パンデミック対策のためのウイルス進化予測

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ある。リアルタイムのウイルス株と抗体のデータに依存する従来の方法では、パンデミックの初期段階において、以下のような問題に直面する。

データ不足にEVEscapeは、過去の塩基配列と生物物理学的・構造的情報をもとに学習させた新しいAIディープラーニングモデルで、以下のような進化を予測する。

### EVEscapeデザイン

出典タダニ他、2023年



のウイルスを評価します（図5.2.5）。EVEscapeは、現在の菌株データとは独立してウイルスエスケープを評価します。は、観察されたSARS-CoV-2突然変異の50.0%を予測し、46.2%および32.3%を予測した従来のラボ研究、および突然変異の24%しか予測しなかつた以前のモデルを上回った（図5.2.6）。

この実績は、EVEscapeが将来のパンデミックへの備えと対応活動を強化する貴重な資産となる可能性を浮き彫りにしている。

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

出典Thadani et al, 2023 | Chart: 2024年AI指数レポート

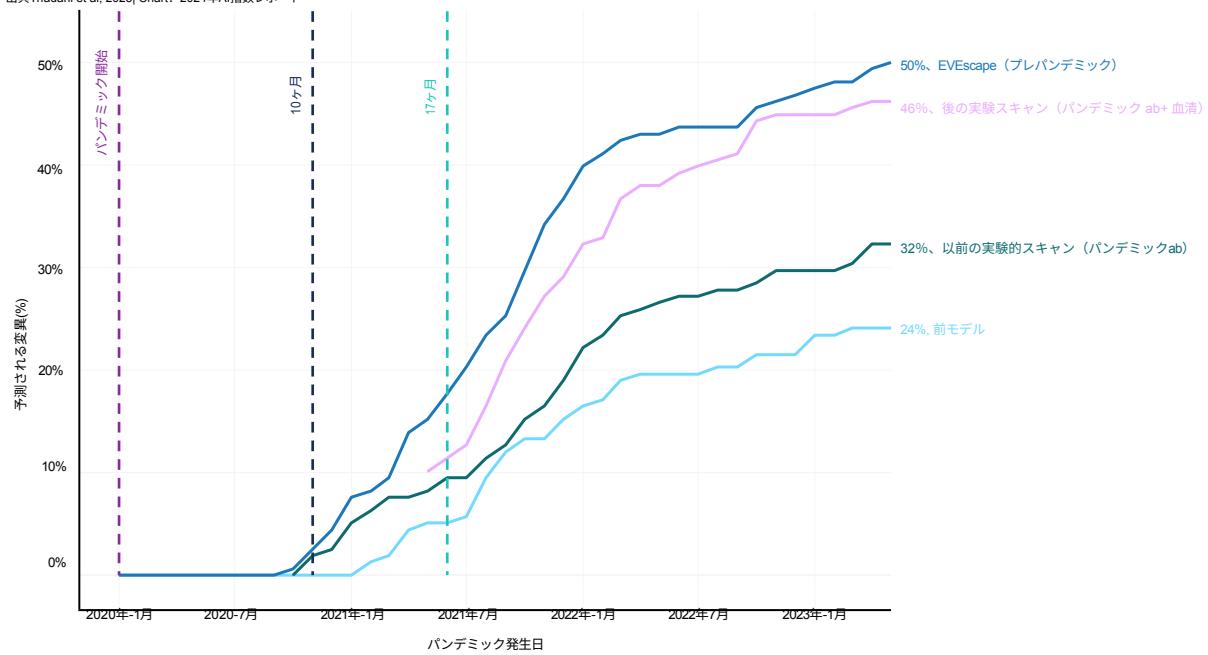


図5.2.6

## 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、どのような遺伝子変異が病気を引き起こすのか、まだ完全には理解していない。何百万もの可能性のある遺伝子変異がある中で、突然変異が良性か病原性かは、手間のかかる実験を必要とする。

2023年、グーグル・ディープマインドの研究者たちは、7,100万個のミスセンス変異体の病原性を予測する新しいAIモデル、AlphaMissenseを発表した。ミスセンス変異はヒトのタンパク質の機能に影響を与える（図5.2.7）、癌を含む様々な疾患の原因となる遺伝子の変化。71のAlphaMissenseは89%のミスセンス変異を分類し、57%は良性、32%は病原性である可能性が高く、残りは不明と分類した（図5.2.8）。対照的に、人間のアノテーターは全ミスセンス変異の0.1%の性質しか確認できなかった。

## ヘマグロビン・サブユニットβ（HBB）

出典グーグル・ディープマインド、2023年



図5.2.7

## アルファミセンス予想

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

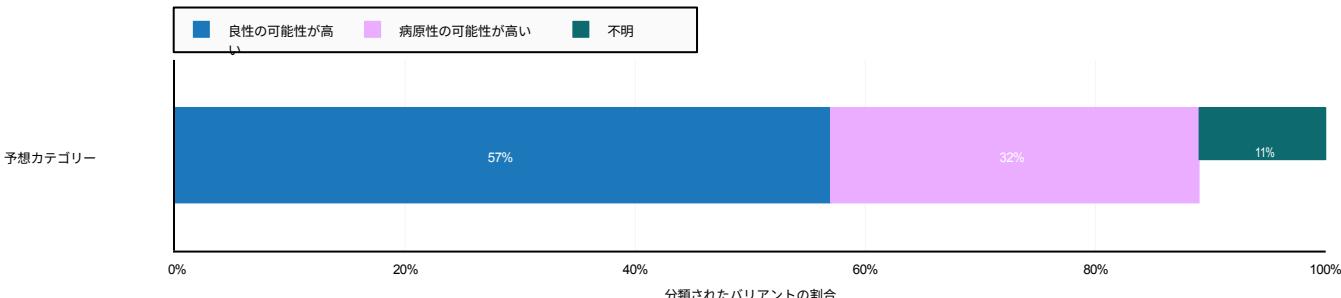


図5.2.8

## ヒトゲノムリファレンス

### AIによるヒトゲノム解析

ヒトゲノムは、ヒトの分子的指令の集合体である。最初のヒトゲノム草案は2000年に発表され、2022年に更新された。しかし、その更新はやや不完全だった。血液型のような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も完全にマッピングされていなかった。そのため、既存のゲノムリファレンスでは、特定のグループの人々の病気を発見したり、治療法を見つけたりすることは難しいだろう。

2023年、60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を利用して、より代表的なヒトゲノムマップを更新した（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中央値99.42%、非コード遺伝子の中央値98.16%、非コード遺伝子の中央値98.96%という驚くべき精度のアノテーションを達成した（図5.2.9）。

図5.2.10に詳述するように、ノンコーディング転写産物の。

## ゲノムのMHC領域のグラフゲノム

出典 グーグル調査、2023年

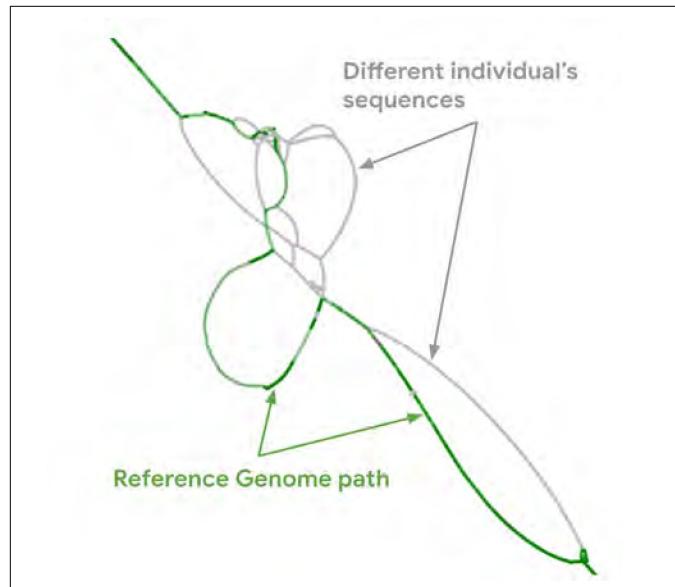


図5.2.9

この最新版のゲノムは、これま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意味する。

### Ensemblマッピングパイプラインの結果

出典 Liao 他、2023年 | 図表：2024年AI指數レポート

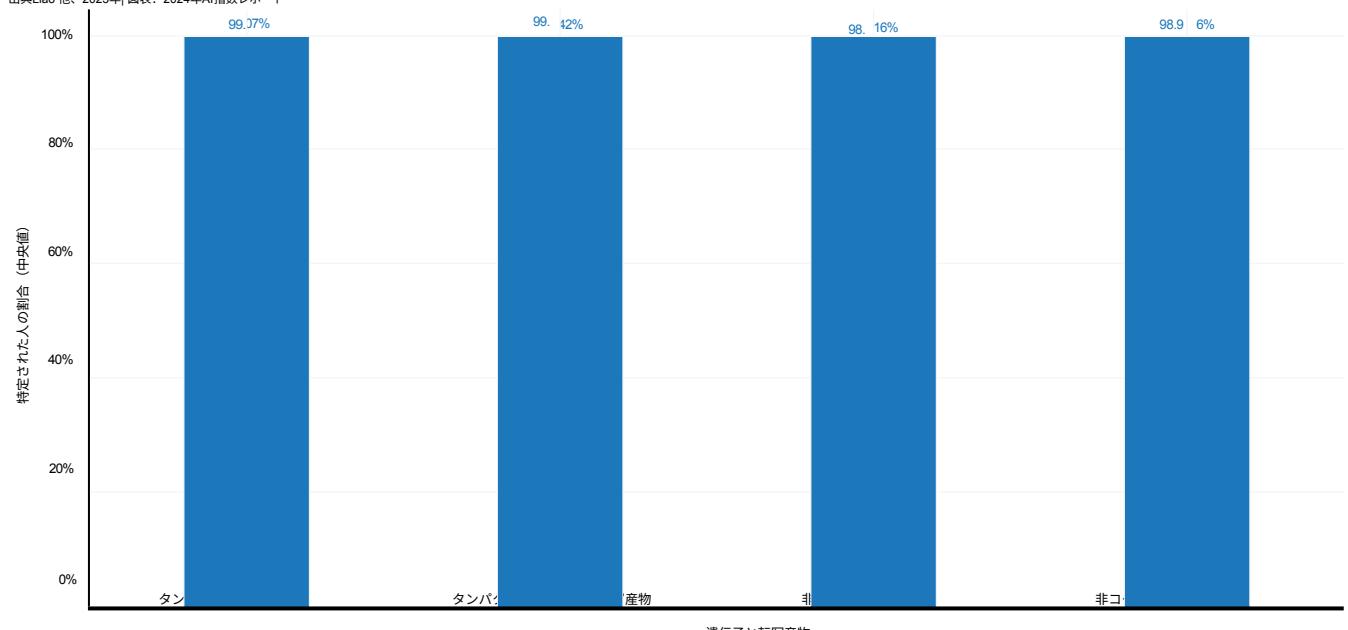


図5.2.10

## 臨床知識

AIモデルの臨床知識を評価するには、その医学的専門知識、特に臨床現場で応用可能な知識の程度を判断する必要がある。

### メドQA

2020年に導入されるMedQAは、専門医の試験に由来する包括的なデータセットであり、医師に挑戦するようデザインされた6万問以上の臨床問題を特徴としている。

MedQAベンチマークにおけるAIのパフォーマンスは、以下の通りである。

#### MedQA: 正確性

出典ペーパー・ウィズ・コード、2023年 | チャート：2024年AI指数レポート

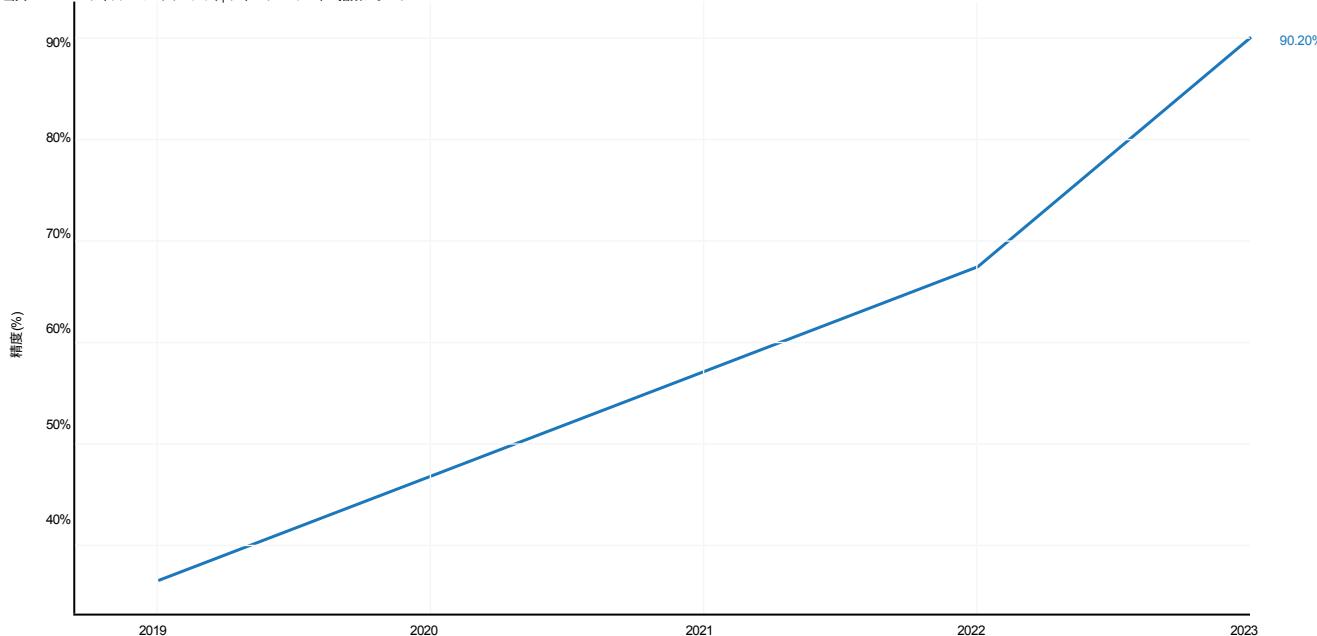


図5.2.11

## 注目の研究

### GPT-4 メドプロンプト

LLMは印象的な一般知識を示すが、医学的な質問に答えるような専門知識に秀でるためには、かなりの微調整が必要であると一般的に考えられている。ファインチューニングとは、LLMをドメイン固有のデータで訓練することである。

2023年末のマイクロソフトの研究は、この仮定を覆した。

本研究では、プロンプトエンジニアリングを採用し、GPT-4 が MultiMedQA ベンチマーク群（4つの難易度の高い医療ベンチマーク群）において顕著な性能を達成することを目指した（図 5.2.12）。GPT-4 Medprompt は、以下のような有名な医療ベンチマークの多肢選択式セクションにおいて、2022 年のトップモデルである Flan-PaLM 540B の性能を上回った。

PubMedQA は 3.0 ポイント、MedMCQA は 21.5 ポイント、MMLU は 16.2 ポイント上回った。また、当時最新であった PubMedQA、MedMCQA、MMLU のパフォーマンスも上回った。

→ Med-PaLM 2（図 5.2.13）。

さらに、前述したように、GPT-4 Medprompt は MedQA ベンチマークで初めて 90% の精度を突破した。

この画期的な進歩は、GPT-4 メドプロンプトの卓越した、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、モデルを特殊な領域に適応させるために微調整が必ずしも必要でない可能性を示している。プロンプト・エンジニアリングは代替戦略

#### GPT-4 と Med-PaLM の比較 2 医学的な疑問に答える

出典ノリ他、2023年

**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-by-step, referring to authoritative sources as needed. Among the options, only pectoralis minor muscle originates from the outer surfaces of the 3rd to 5th ribs.



GPT-4

GPT-4 generated CoT

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

出典ノリ他、2023年|図表：2024年AI指数レポート

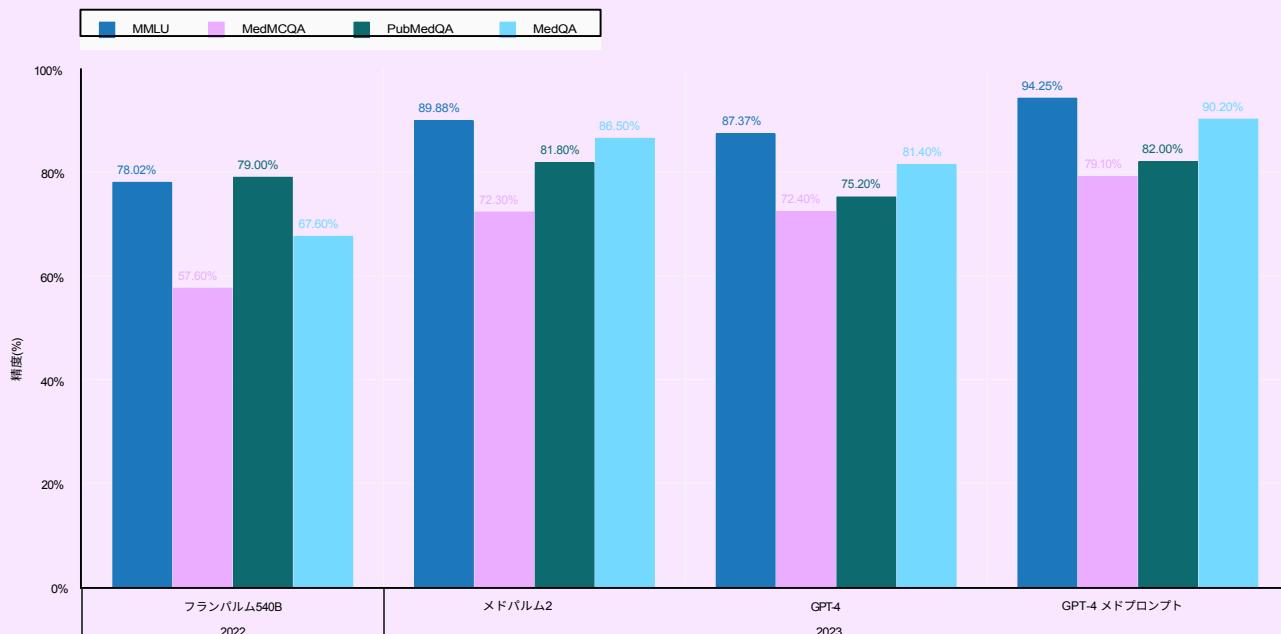


図5.2.13

## 注目の研究

### メディトロン-70B

GPT-4 Medpromptは印象的なシステムであるが、クローズドソースであるためそのウェイトは広く一般に自由に使用できるものではない。2023年の新しい研究でも、オープンソースの医療用LLMの能力を向上させようとしている。この新しい研究の中で、MediTron-70Bは特に有望である。このモデルはMedQAベンチマークで70.2%という立派な精度を達成している。これは、GPT-4 Medprompt や Med-QA ベンチマークのパフォーマンスを下回るものである。

PaLM2（いすれもクローズドモデル）は、最新のモデルよりも大幅に改善されている。のようなオープンソースモデルを凌駕している（図5.2.14）。MediTron-70BのMedQAにおけるスコアは、オープンソースモデルが達成した中で最高である。医療AIがその潜在能力を最大限に発揮するためには、その能力が広く利用可能であることが重要である。その意味で、MediTronは心強い前進である。

#### MedQAにおける選択モデルのパフォーマンス

出典: チェン他、2023年 | 表: 2024年AI指数報告書

モデル	発売日	アクセスタイル	MedQAのスコア
GPT-4 メドプロンプト	2023年11月	クローズド	90.20%
メドバルム2	2023年4月	クローズド	86.20%
メディトロン-70B	2023年11月	オープン	70.20%
メドバルム	2022年12月	クローズド	67.20%
ラマ2	2023年7月	オープン	63.80%

図5.2.14

## 診断

AIツールは、例えば放射線学や癌の検出など、診断目的にも使用できる。

### 注目の研究

#### CoDoC

AI医用画像診断システムは強力な診断能力を示すが、臨床医が把握している診断を見落としたり、その逆もある。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る。2023年、研究者たちはCoDoC（相補性駆動型臨床ワークフローへの移行）を発表した。診断のためにAIに頼るべき時と、従来の臨床的手法に委ねるべき時。CoDoCは、感度（病気を持つ個人を正しく特定する能力）と特異度の両方を顕著に向上させる。

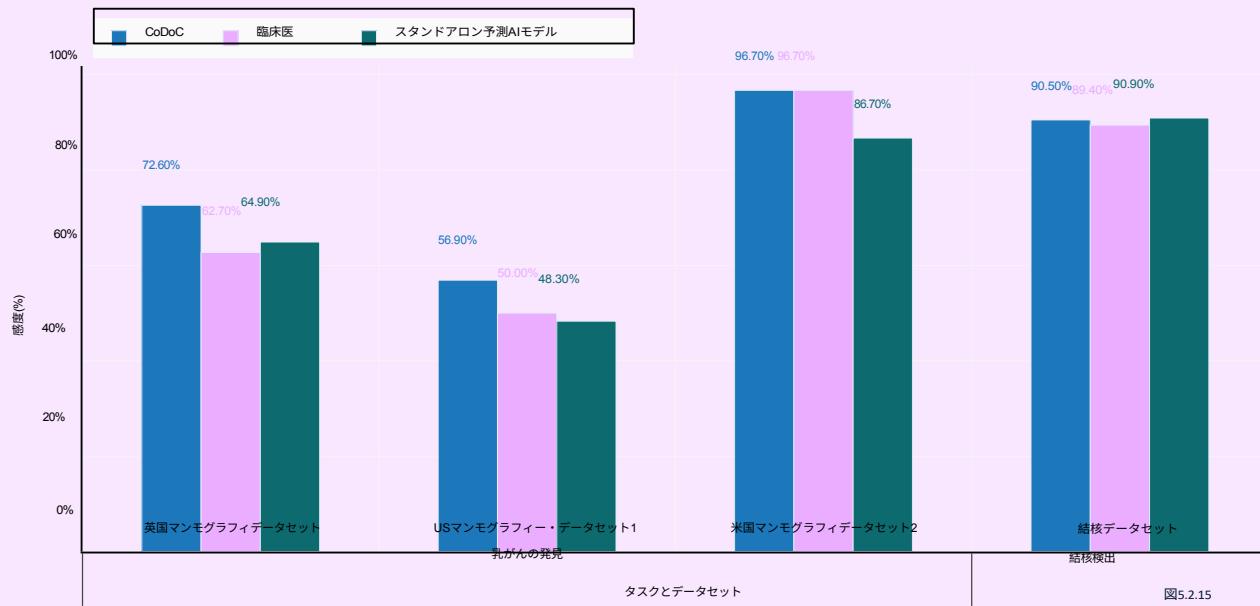
ない人を正確に識別する能力）。具体的には、4つの医療データセットにおいて、CoDoCの感度は臨床医の感度を平均で以下のように上回っている。

4.5%ポイント、単体のAIモデルは6.5%ポイントであった（図5.2.15）。特異度に関しては、CoDoCは臨床医を次のように上回っている。

CoDoCは、テストされたデータセット全体で平均2.7%ポイント、単独の予測モデルでは5.7%ポイント改善した。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た。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で、診断精度と効率を高めることができることを示唆している。

#### CoDoCと単独の予測AIシステムおよび臨床読影医との比較：感度

出典Dwijotham他、2023年|図表：2024年AI指数レポート



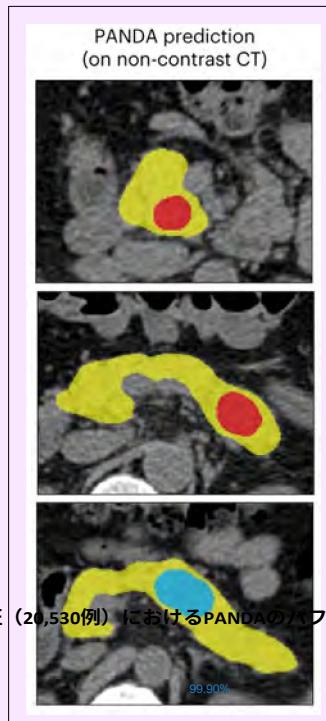
## 注目の研究

## CTパンダ

膵管腺癌（PDAC）は特に致死的な癌であり、発見が遅すぎて外科的介入ができないことが多い。無症状の人を対象としたPDACのスクリーニングは、その有病率の低さと偽陽性のリスクのために困難である。今年、中国の研究チームがPANDA（人工知能による膵臓がん検出）を開発した。

とX線写真の膵臓病変の分類を行った（図5.2.16）。検証、PANDAは感度で34.1%、特異度で6.3%平均的な放射線科医を上回った（図5.2.17）。約20,000人の患者を対象とした大規模な実臨床テストでは、PANDAは感度92.9%、特異度99.9%を達成した（図5.2.18）。PANDAのようなAI医療ツールは、困難な病態の診断における重要な進歩であり、従来は困難または不可能と考えられていたコスト効率の高い正確な検出を提供する。

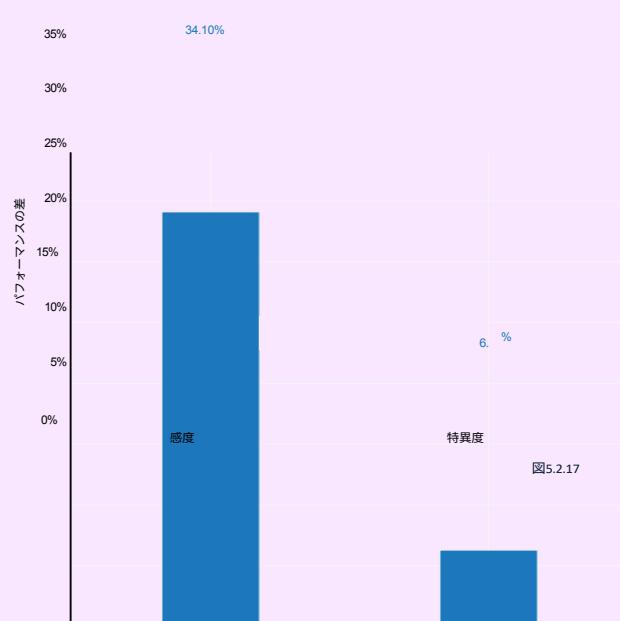
**パンダ  
検出**  
ソース  
曹ら、2023年  
図5.2.16



(20,530例)におけるPANDAのパフォーマンス  
99.90%

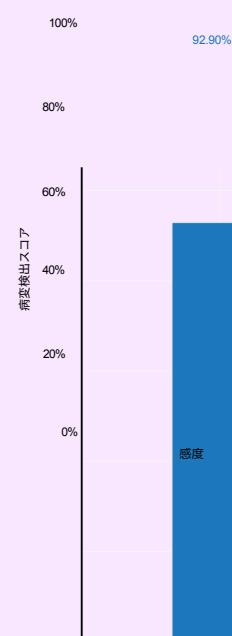
## 多施設検証におけるPANDAと平均的放射線科医の比較（6,239例）

出典Cao et al. 2023| 図表：2024年AI指数レポート



## 実世界のマルチシナリオ検証ス

出典Cao et al. 2023| 図表：2024年AI指数レポート



## その他の診断用途

2023年に発表された新しい研究は、AIが他の診断の場面でどのように利用できるかを浮き彫りにしている。図5.2.19はその結果の一部を要約したものである。

### 診断AIのユースケースに関する追加調査

出典AIインテックス、2024年

リサーチ	使用例	調査結果
ショブフ他、2023	乳がん	著者らは、マンモグラフィ画像に基づくAIアルゴリズムを探求する文献のメタレビューを行った。その結果、マンモグラフィ画像のみを用いた将来の乳がんリスク予測は、従来のリスク評価ツールに匹敵するかそれ以上の精度を達成することが判明した。
ディセンテ・シドら、2023年	X線解釈	研究者らは、画像とフリーテキスト報告書を用いて胸部X線を分類するための2つのオープンソース・ニューラルネットワーク、X-RaydarとX-Raydar-NLPを開発した。これらの自動分類法は、人間の専門家に匹敵するレベルの性能を発揮し、外部データセットに適用しても頑健性を示すことがわかった。

図5.2.19

**FDAが承認した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、FDAの市販前基準を満たしています、その有効性と安全性の詳細なレビューを含む。2023年10月現在、FDAはジェネレーティブAIを利用した機器やLLMを搭載した機器を承認していない。

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る。2022年には、合計139のAI関連医療機器がFDAの承認を受け、2021年に承認された合計から12.1%増加した。2012年以降、これらの機器の数は45倍以上に増加している。

#### FDAが承認したAI医療機器の数（2012～22年）

出典FDA、2023年| 図表：2024年AI指数レポート

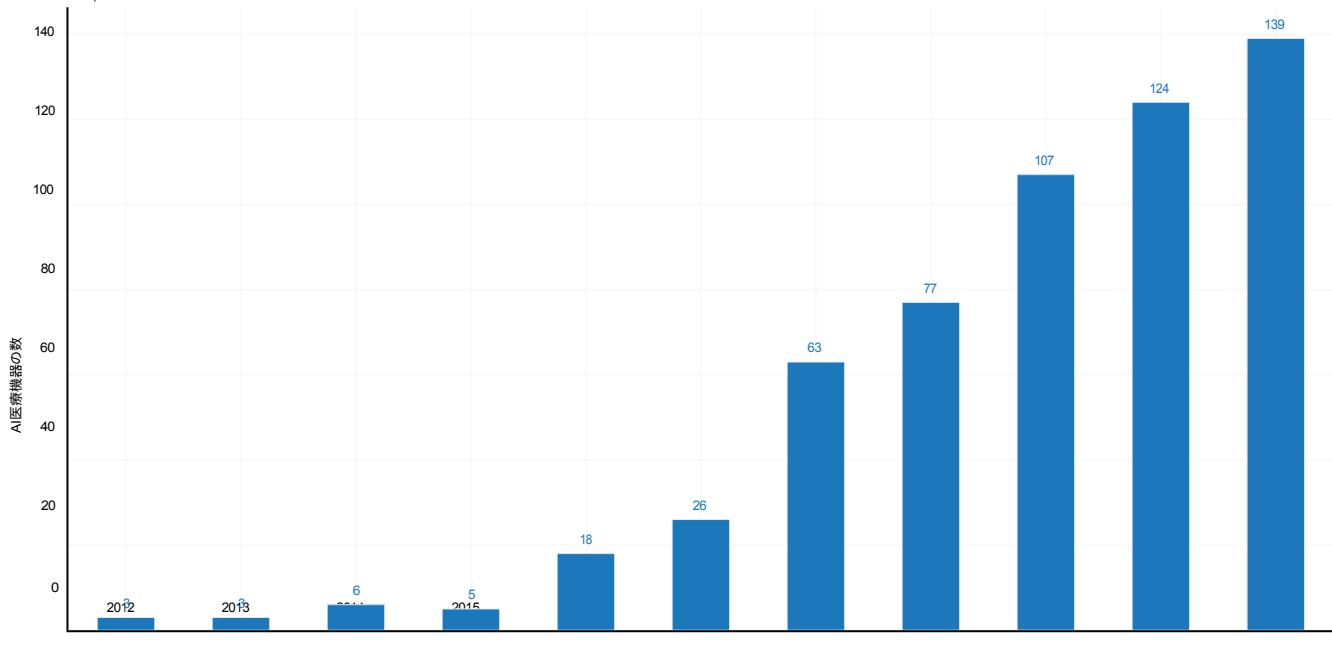


図5.2.20

<sup>3</sup> FDAが最後にリストを更新したのは2023年10月であり、2023年の集計は不完全であった。そのため、AIインデックスは2022年までの情報のみを含むデータ表示に限定した。

図5.2.21は、FDAが承認した医療機器に関する専門分野を示している。2022年に承認された139の医療機器のうち、87.1%と大多数が放射線科関連であった。次に多い専門分野は心臓血管系で、承認の7.2%を占めた。

#### FDAが承認したAI医療機器の専門分野別件数（2012～22年）

出典FDA、2023年| 図表：2024年AI指教レポート

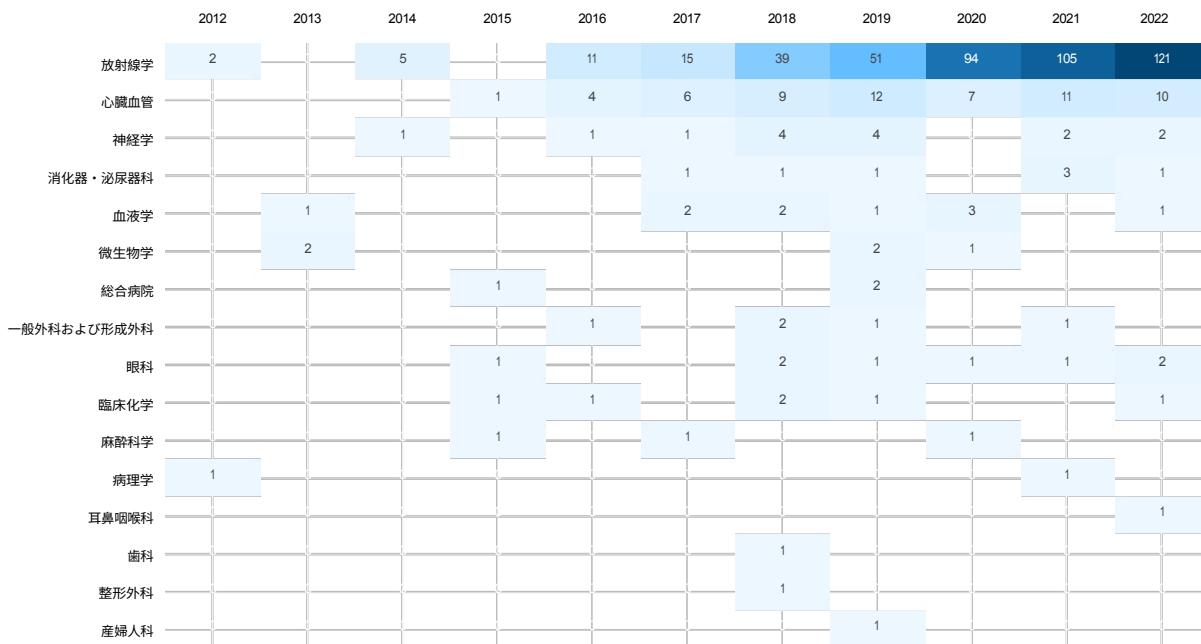


図5.2.21

## 管理とケア

AIツールはまた、医療管理の効率を高め、患者ケアの水準を向上させる可能性を秘めている。

### 注目の研究

#### メドアライン

医療用AIの進歩が著しいにもかかわらず、MedQAやMedQAのような既存のベンチマークは、医療用AIの普及を妨げている。

USMLEは知識ベースの問題に重点を置いているため、臨床医が患者ケアで行う多様な業務を十分に把握できていない。臨床医はしばしば

LLMは、オーダーメイドの診断計画作成など情報集約的な業務に従事し、労働時間のかなりの割合を管理業務に費やしている。AIはこうしたプロセスを合理化する可能性を秘めているが、医療事務LLMのベンチマークや微調整に適した電子カルテ（EHR）データセットが不足している。今年、研究者たちはMedAlignを導入することで、このギャップを解決するために前進した：

包括的なEHRベースの

ベンチマークには、7つの異なる診療科から集められた983の質問と指示、303の臨床医の回答が含まれる（図5.2.22）。MedAlignは、EHRに焦点を当てた初の広範なベンチマークです。次に研究者たちは、既存のさまざまなLLMをMedAlignでテストした。すべてのLLMの中で、多段階精密化を用いたGPT-4の変種が最も高い正答率（65.0%）を達成し、他のLLMよりも日常的に好まれた（図5.2.23）。MedAlignは、LLMの改良に向けた貴重なマイルストーンである。

医療における管理負担を軽減するためにAIを活用する。

#### MedAlignワークフロー

出典フレミング他、2023年

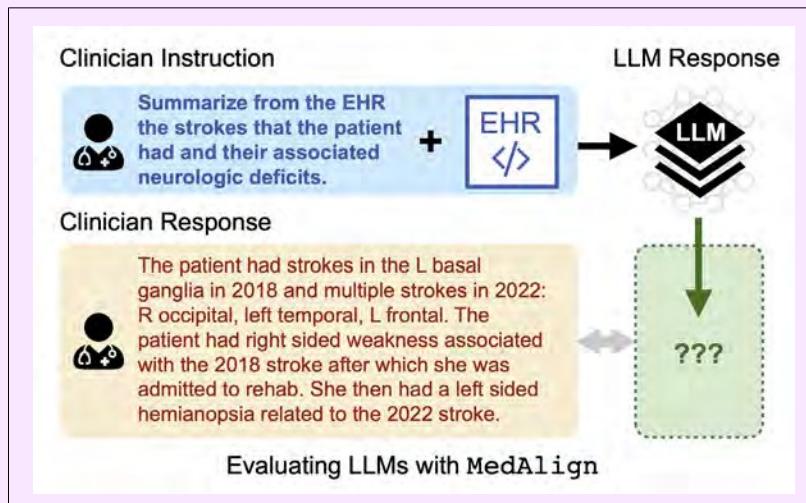


図5.2.22

## 注目の研究

## メドアライン（続き）

## モデル性能の評価：人間対COMETランク

出典：フレミング他、2023年 | 図表：2024年AI指数レポート

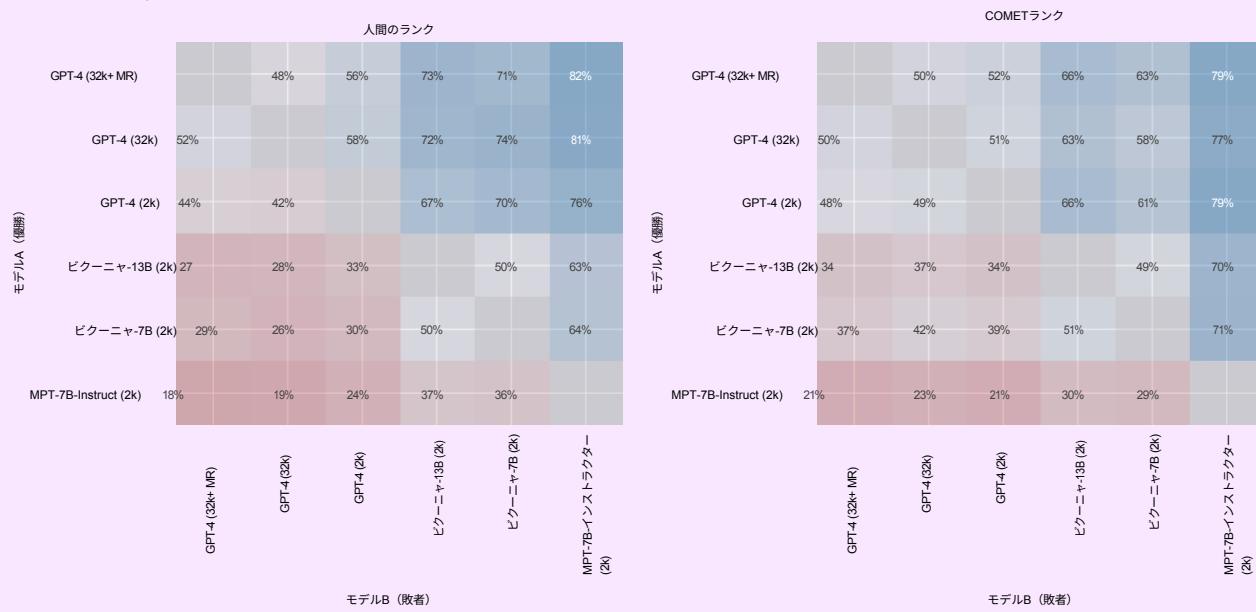


図5.2.23





# プレビュ

---

---

概要	327
チャプター・ハイライト	328

## 6.1 中等教育修了後のCSおよびAI教育 329

米国およびカナダ	329
CS学士号取得者	329
CS修士課程修了者	331
CS博士号取得者	333
CS、CE、情報学部	336
ヨーロッパ	344
情報学、CS、CE、IT学士号取得者	344
情報学、CS、CE、IT修士課程修了者	347
情報学、CS、CE、ITの博士号取得者	351
AI関連学習プログラム	355
全コース	355
教育レベル	356
地理的分布	357

## 6.2 K-12CSおよびAI教育 359

米国	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 アクセス問題	363
ハイライト ChatGPTの利用状況	
教師と生徒	364

公開データにアクセスする





## 概要

本章では、AIとコンピュータ・サイエンス（cs）教育の動向について、誰が、どこで、どのようにいるのか、また、これらの動向は時間の経過とともにどのように変化してきたのかに焦点を当てて考察する。AIが教育に与える影響への懸念が高まる中、ChatGPTのような新しいAIツールの教師や生徒による利用についても調査している。

分析ではまず、Computing Research Associationが毎年行っているTaulbee Surveyに基づき、米国とカナダにおける中等教育後のCSおよびAI状況を概観する。次に、欧州のCS教育に関するInformatics Europeのデータを検証する。本年は、AI関連の英語学習プログラムの世界的な数に関するStudyportalsのデータを新たに紹介。

この章では、Code.orgによる米国の幼稚園から高校までのCS教育に関する洞察と、ウォルトン財団によるChatGPTの学校での使用に関する調査結果で締めくくられている。



## チャプター・ハイライ

### ト

**1. 米国とカナダのCS学士号取得者数は増加し続け、CS修士号取得者数は比較的横ばい、博士号取得者数は緩やかに増加している。**アメリカとカナダの学士号新規取得者数は10年以上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となっている。2018年以降、CSの修士号および博士号取得者数はわずかに減少している。

**2. AI博士の産業界への移行は加速している。**2011年には、新しいAI博士のほぼ同じ割合が産業界（40.9%）と学界（41.6%）に就職した。しかし、2022年までには、卒業後に産業界に就職する人の割合（70.7%）が、アカデミアに就職する人の割合（20.0%）を大きく。過去1年間だけでも産業界に就職したAI博士の割合は5.3ポイント上昇しており、大学から産業界への頭脳流出が激化していることを示している。

**3. 産業界からアカデミック人材の移行が少ない。**2019年、米国とカナダにおける新しいAI教員の13%が産業界出身者だった。2021年までにこの数字は11%に低下し、2022年にはさらに7%に低下した。この傾向は、ハイレベルなAI人材の産業界から学界へ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教育の国際性が低下。**2022年に卒業したCS学士、修士、博士の国際的な数は、2021年に比べて相対的に減少した。修士課程における留学生の減少は特に顕著であった。

**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るが、アクセスの問題は残っている。**2022年には、201,000のAP CS試験が実施された。2007年以来、この試験を受ける生徒の数は10倍以上に増加している。しかし、最近の証拠によると、大規模な高校や郊外地域の生徒の方が、CSコースを受講できる可能性が高い。

**6. AI関連の学位プログラムは国際的に増加傾向にある。**英語の数も増えている、AI関連の中等教育後の学位プログラムは2017年以降3倍に増加しており、過去5年間にわたり毎年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提供している。

**7. 欧州の情報学、CS、CE、IT卒業生輩出は英国とドイツがリード。**イギリスとドイツは、情報学、CS、CE、情報の学士、修士、博士課程の新卒者の輩出数で欧州をリードしている。一人、フィンランドが学士号取得者数、博士号取得者数ともにトップであり、アイルランドは修士号取得数でトップである。

このセクションでは、CSとAIの中等後教育の概要について、学士号、修士号、博士号などさまざまな学位に関する北米と欧州の卒業統計を紹介する。また、以下の情報も網羅している。

提供されるAI関連コース。

## 6.1 Postsecondary CS and AI 教育

### 米国およびカナダ

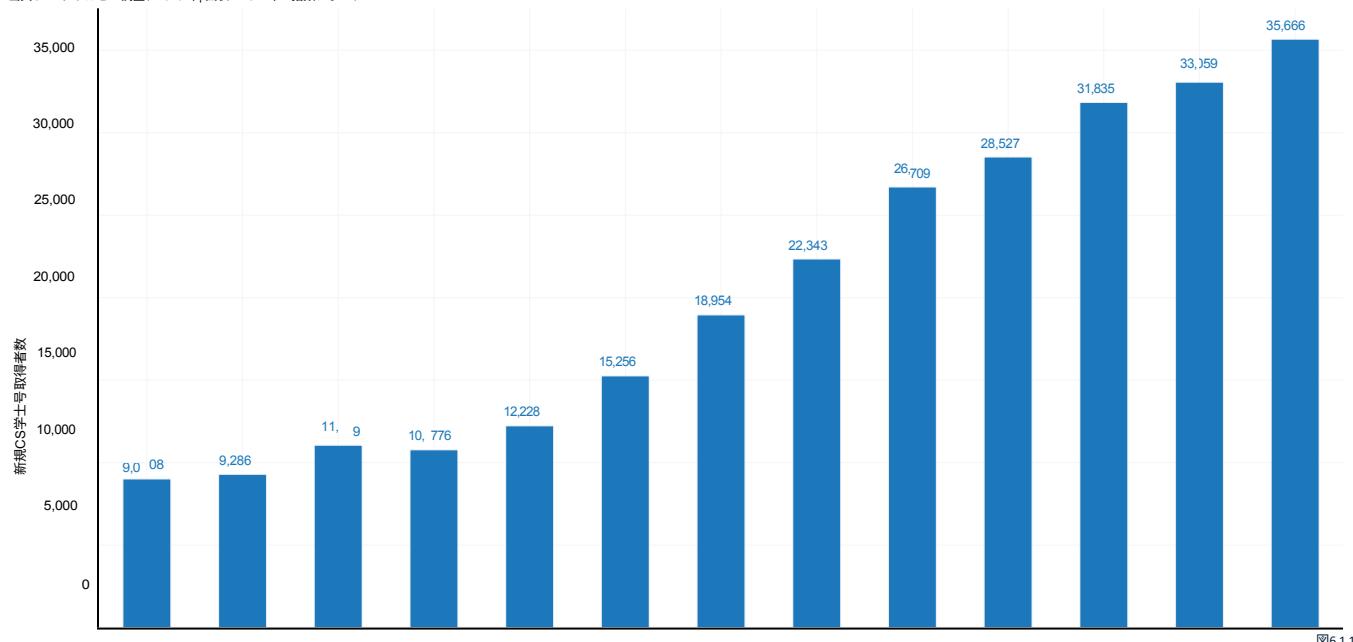
この小節では、米国とカナダにおけるCSとAIの中等後教育の状況を評価するComputing Research AssociationのTaulbee Surveyのデータを分析する。この、米国とカナダの297の博士号を授与しているCS学科を対象としている<sup>(1)</sup>。

#### CS学士号卒業生

過去10年間、北米におけるCS学士の新規卒業者総数は着実に増加しており、2021年から2022年にかけて前年比7.9%増と、3倍以上に増加している（図6.1.1）。

米国とカナダにおけるCS学士の新規卒業者数（2010～22年）

出典CRA タウルビー調査、2023年 | 図表：2024年AI指数レポート



1 重要な、調査の対象となったすべての博士号授与部局から回答が得られたわけではないことである。対象となった297部局のうち、回答があったのは182部局で、全体の回答率は61%であった。

アメリカとカナダの大学におけるCS学士号取得者に占める留学生の割合は、ほぼ8年ぶりに減少し、2021年の16.3%から15.2%に減少した。2022年である（図6.1.2）。この減少は、トランプ政権初期に就学ビザの取得が困難になったことを反映していると思われる。

のデータを参照されたい。この減少は、COVID-19パンデミック時に課された海外渡航規制が、留学生の米国およびカナダでの就学に影響を与えたことも一因である。このような最近の減少にもかかわらず、過去10年間の全体的な傾向を見ると、留学生の着実に増加している。

米国およびカナダにおけるCS学士の新規外国人卒業生（全体に占める割合）（2010～22年）

出典CRA タウルビー調査、2023年| 図表：2024年AI指数レポート

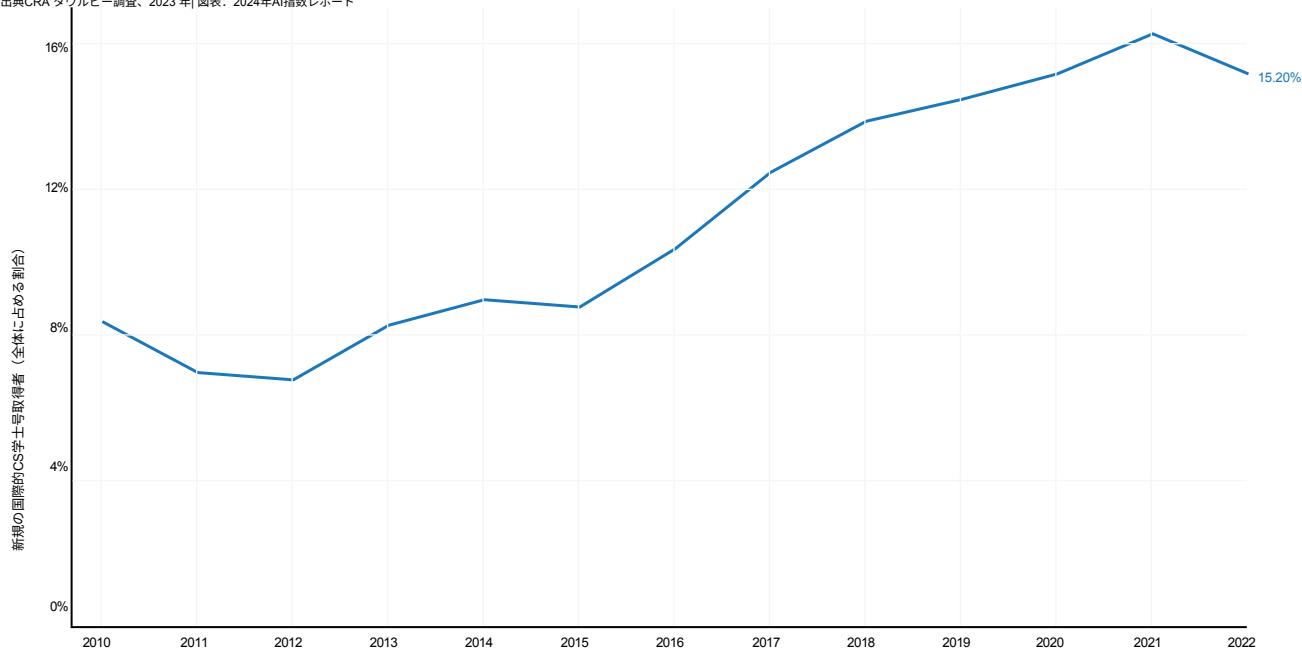


図6.1.2

## CS Master's 卒業生

AIコースは、CS修士課程に一般的に含まれている。アメリカやカナダの大学におけるCS修士課程の新規卒業者数は、過去10年間で2倍以上に増加した、

が、2018年以降は横ばいとなり、昨年は2.5%減とわずかに減少したようである（図6.1.3）。この横ばいは、以下のグラフに示す修士課程留学生の減少を反映したものである。

### 米国とカナダにおけるCS修士課程新卒者（2010～22年）

出典CRA タウルビー調査、2023年| 図表：2024年AI指数レポート

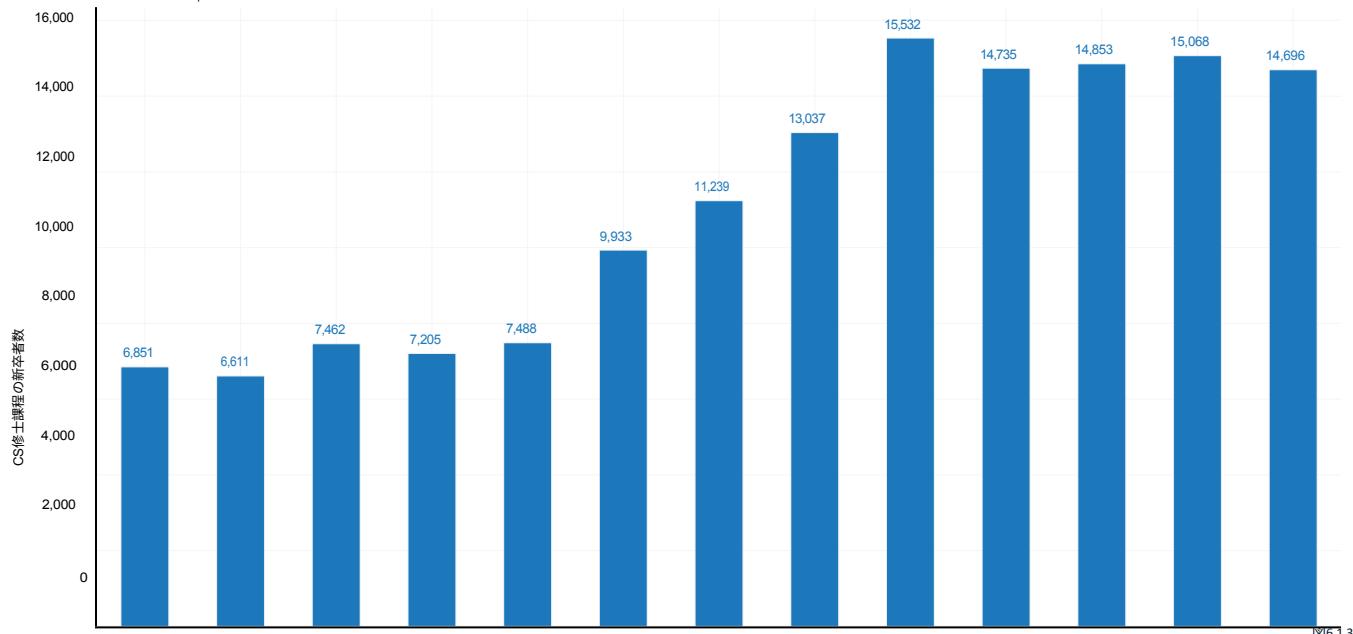


図6.1.3

2022年、アメリカとカナダの大学では、CS修士課程の留学生が顕著に減少した。この減少傾向は2017年頃から始まったが、14.8ポイントで最も顕著であった（図6.1.4）。現在、海外と国内のCS修士課程卒業生の割合はほぼ均等である。

米国およびカナダにおけるCS修士課程の新規外国人卒業生（全体に占める割合）（2010～22年）

出典CRA タウルビー調査、2023年| 図表：2024年AI指数レポート

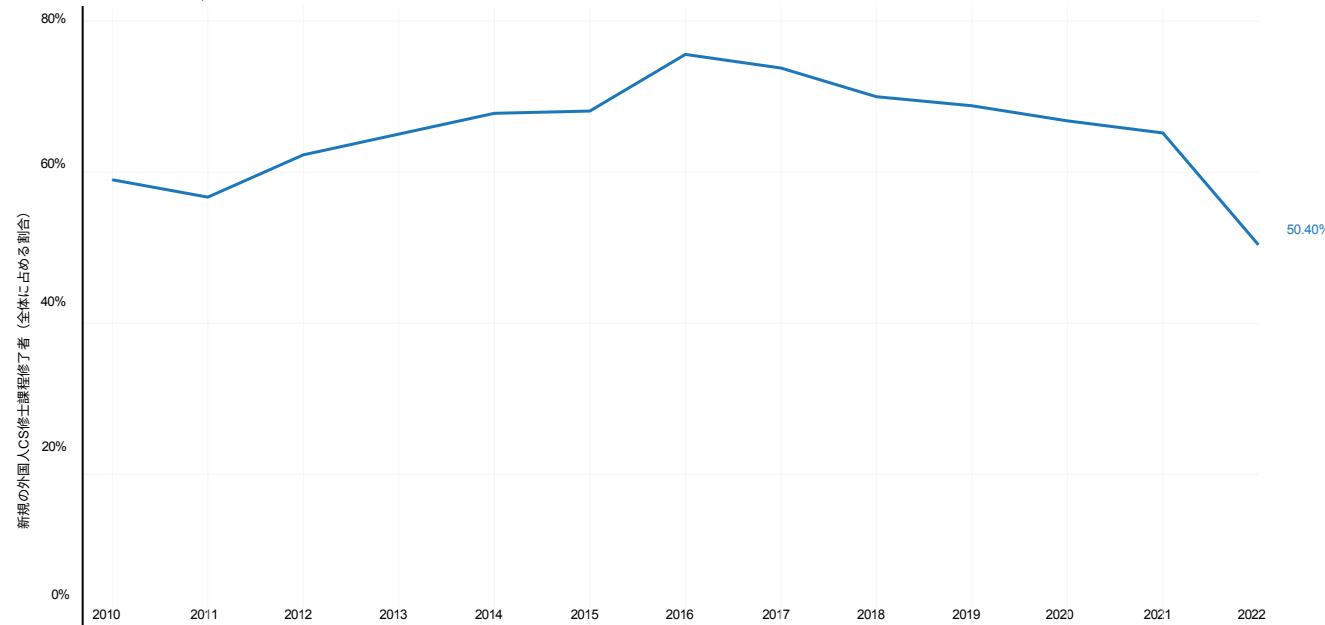


図6.1.4

## CS PhD 卒業生

この10年間で初めて、米国とカナダの大学におけるCS博士課程の新卒者大幅に増加した。2022年のCS博士号取得者数は2,105人に達し、2010年以来最高となった（図6.1.5）。

### 米国とカナダにおける新規CS博士号取得者（2010～22年）

出典CRA タウルビー調査、2023年| 図表：2024年AI指数レポート

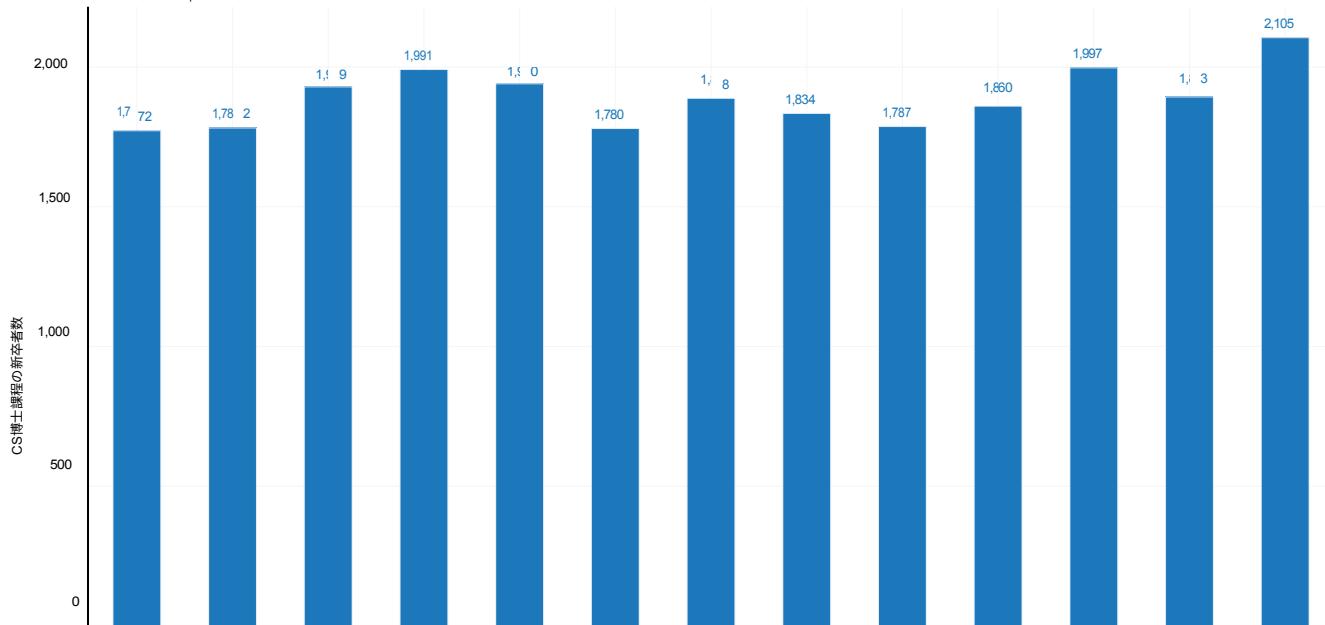


図6.1.5

CS博士号取得者に占める留学生の割合は過去10年間上昇してきたが、昨年はこの割合が若干減少し、2021年の68.6%から2022年には65.9%に低下した（図6.1.6）。

#### 米国およびカナダにおける新規の国際的CS博士号取得者（全体に占める割合）、2010～22年

出典CRA タワルピー調査、2023年| 図表：2024年AI指数レポート

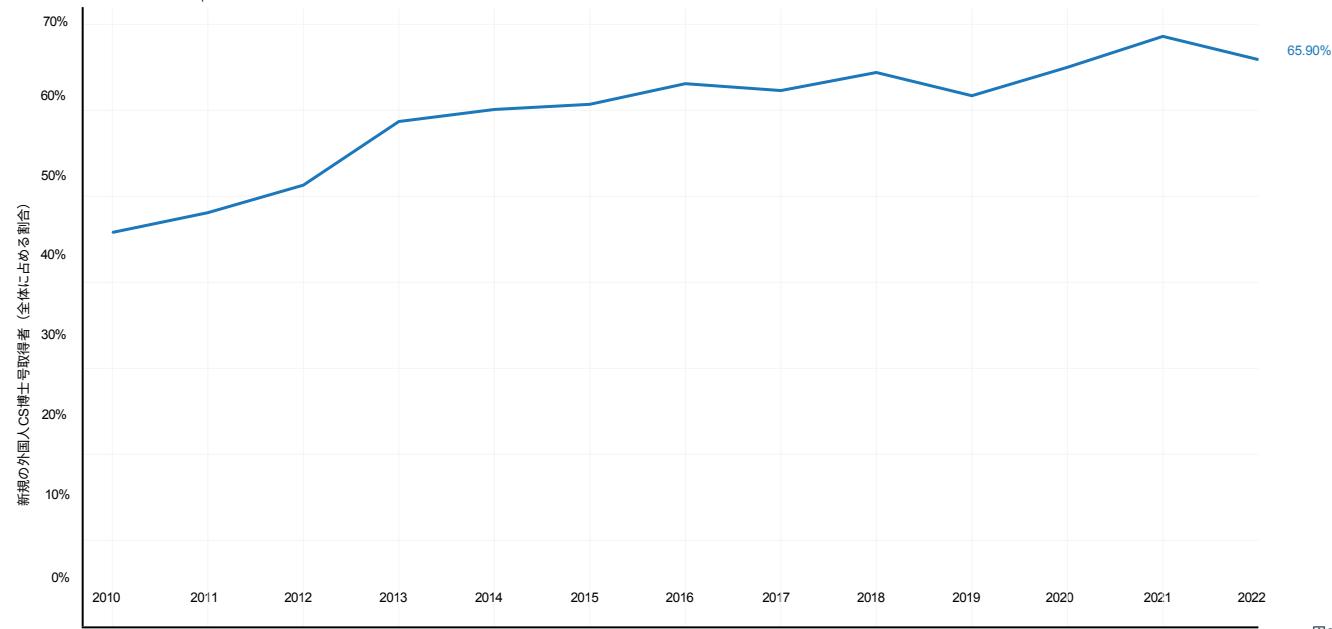


図6.1.6

新しくAI博士号を取得した人々は、卒業後どこで働くことを選ぶのだろうか？昨年のAIインデックス・強調された傾向に従い、AI博士の割合は増加している。

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%）でほぼ同率であった。

しかし、2022年までには、卒業後に産業界に入る割合（70.7%）は、学術界に入る割合（20.0%）に比べてかなり大きくなる。新しいAI博士号取得者のうち、政府の職務に就く者の割合は、過去5年間、約0.7%と比較的低く安定している。

### 米国とカナダにおけるAI博士の新規雇用（全体に占める割合）（分野別）（2010～22年）

出典CRA タウルビー調査、2023年|図表：2024年AI指数レポート

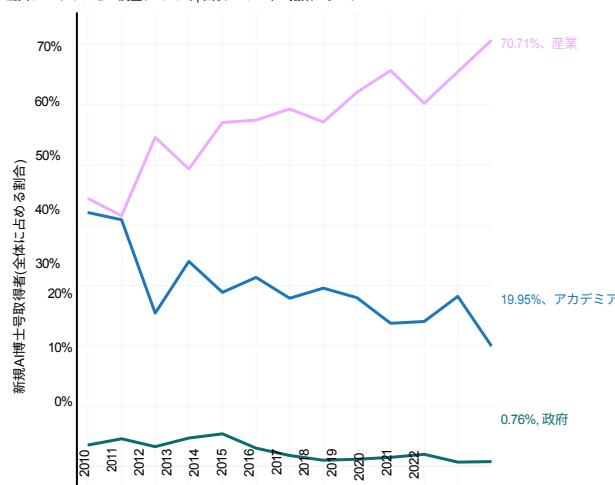


図6.1.7<sup>2</sup>

### 米国とカナダにおけるAI博士の新規雇用（部門別）（2010～22年）

出典CRA タウルビー調査、2023年|図表：2024年AI指数レポート

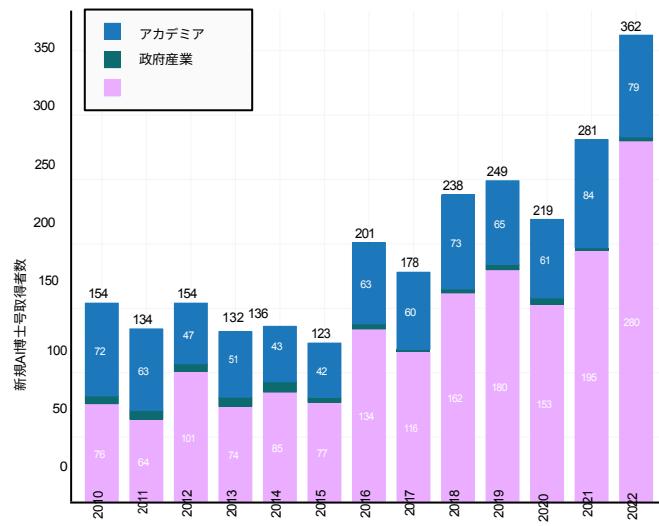


図6.1.8

<sup>2</sup> 図6.1.7の合計は100にならないが、これはCRA自営業、無職「その他」の雇用形態と回答したAI博士号取得者が毎年いるためである。これらの学生はこのグラフには含まれていない。

## CS、CE、情報教員

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べることが有用である。昨年、米国とカナダの大学におけるCS、CE、情報系教員の総数は7.2%増加した（図6.1.9）。

2011年からの増加率は42.4%である。

### 米国とカナダにおけるCS、CE、情報の教員数（2011-22年）

出典CRA タウルビー調査、2023年|図表：2024年AI指数レポート

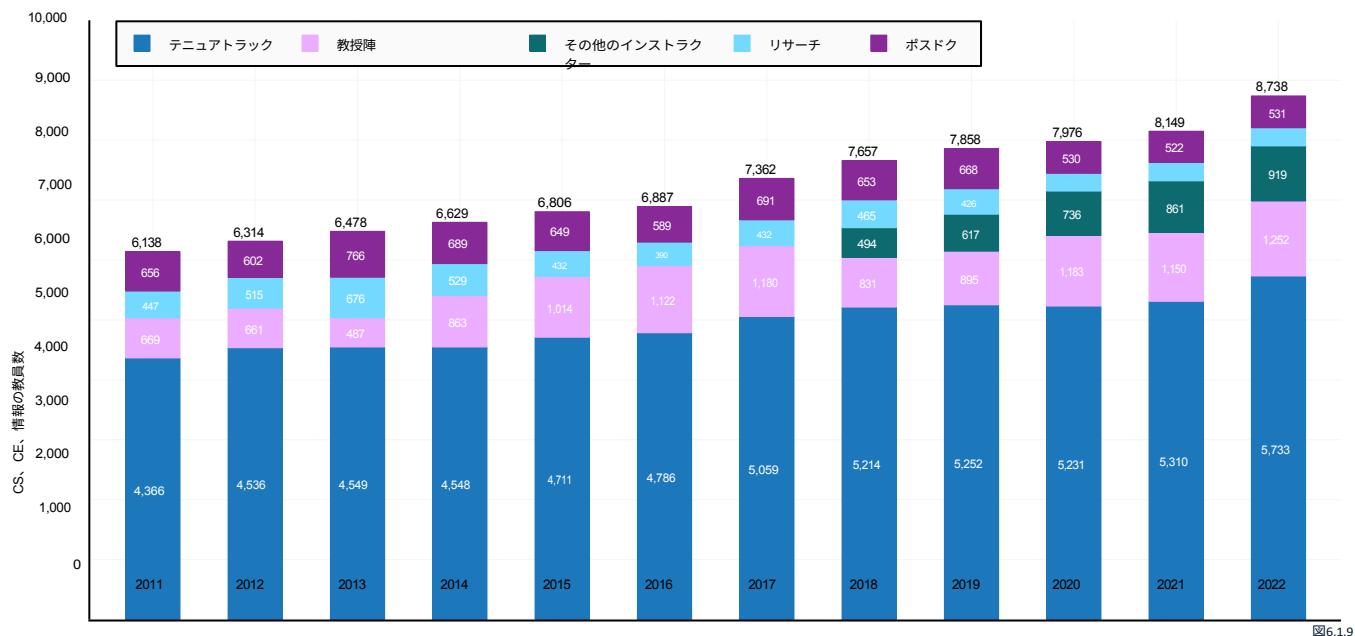


図6.1.9

2022年、米国には7,084人のCS教員があり、その大多数（65.7%）がテニュアトラックに在籍していた（図6.1.10）。米国のCS教員総数は、2021年以降4.4%、2011年以降45.0%増加している。

### CS教員数（2011-22年）

出典CRA タウルビー調査、2023年|図表：2024年AI指数レポート

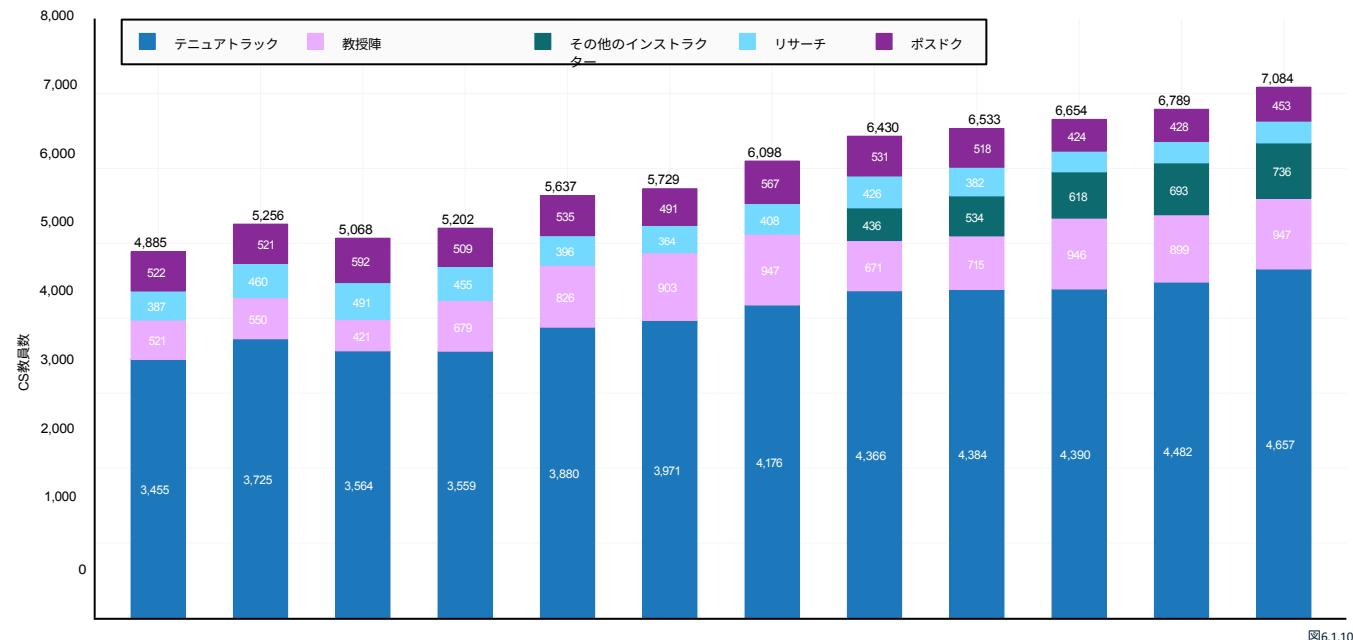


図6.1.10

昨年、北米では、CS、CE、情報分野の教員915人が新規採用されたが、これは10年ぶりの高水準であった。このうち455名がテニュア・トラック教員であった（図6.1.11）。（図6.1.11）。

### 米国とカナダにおけるCS、CE、情報系教員の新規採用（2011-22年）

出典CRA タウルビー調査、2023年|図表：2024年AI指数レポート

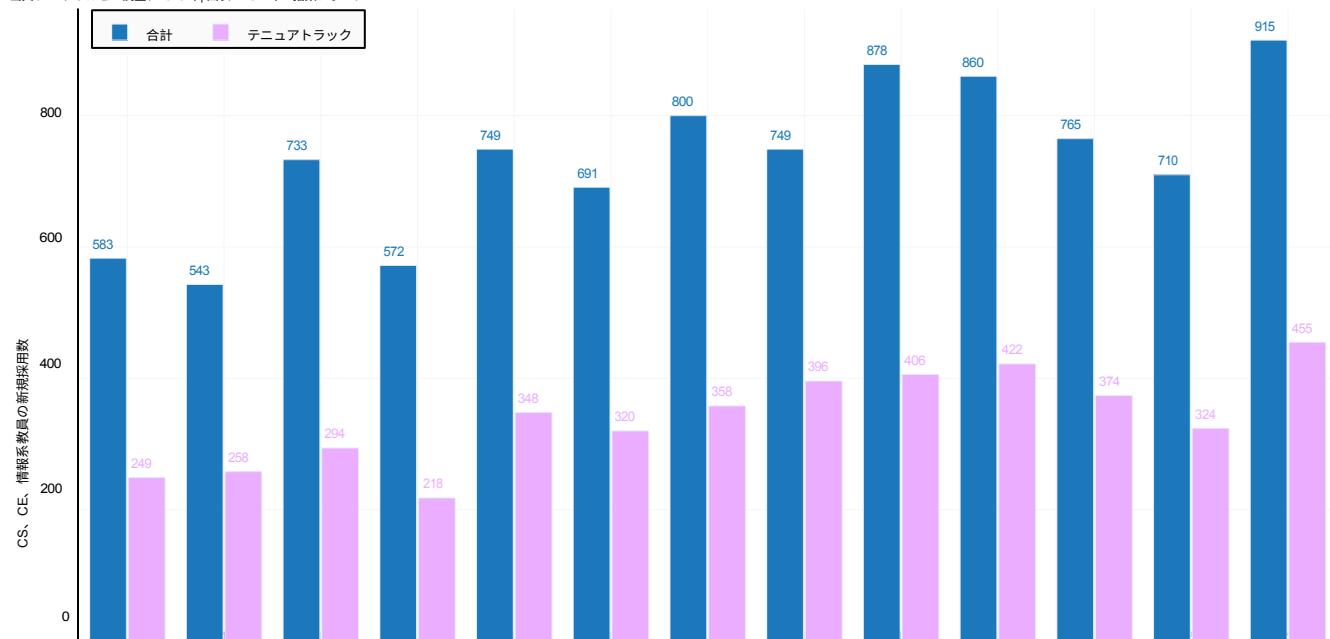


図6.1.11

2022年には、新任教員の43%が他の学問出身者であり、学問分野の人材が「入れ替わって」いることを示している（図6.1.12）。これら

の「新任」教員は他の職を辞しているため、その職を辞した教員は、「新任」教員として採用されない。

これまでの役割をいずれ埋める必要がある。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%、2019年の13%から7%に減少した。

#### 米国とカナダのCS、CE、情報学部における新任教員の出身地（2018-22年）

出典CRA タウルビー調査、2023年| 図表：2024年AI指数レポート

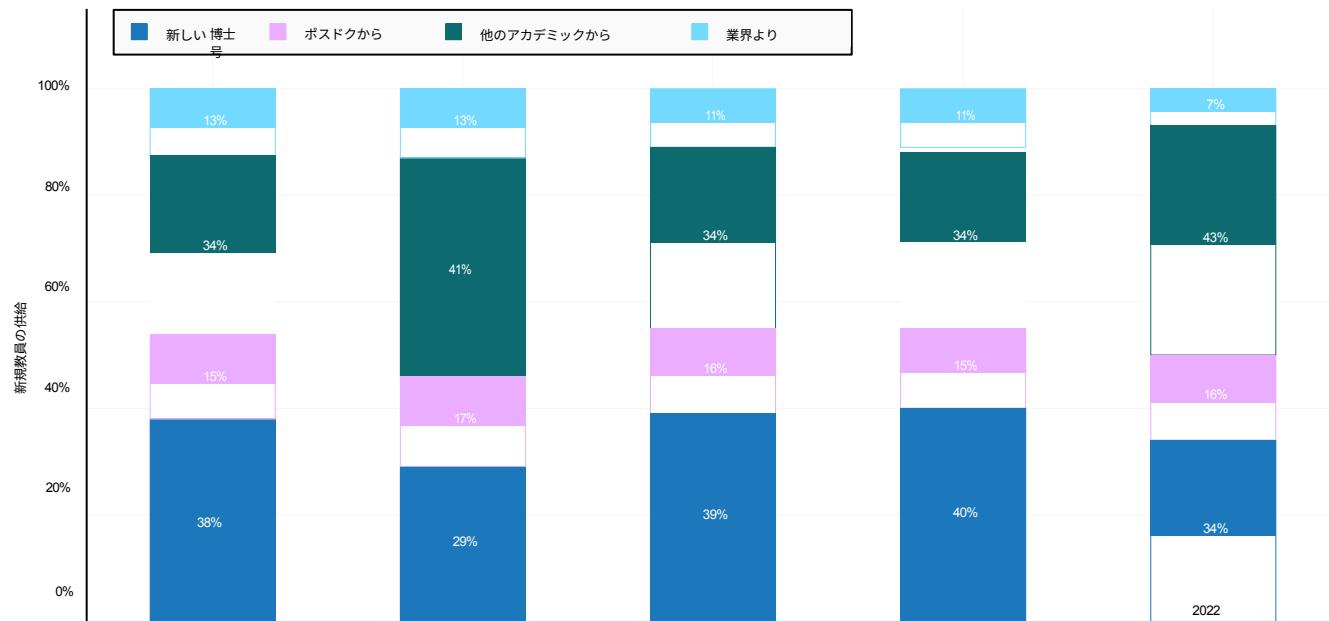


図6.1.12

教員の採用が決まらない理由は過去10年間で様々で。2011年、採用が決まらなかった理由の37%はオファーがなかったことであり、34%はオファーが断られたことであった（図6.1.13）。一方、2022年には、オファーがなかったのはわずか15%で、55%はオファーの辞退であった。

が辞退した。この傾向は、新任CS教員市場の競争が激化していることを反映しているようだ。しかし、これが他の学術職との競争激化を意味するのか、産業界との競争激化を意味するのかは不明である。

### CS、CE、情報系教員の新規採用枠が埋まらなかった理由（占める割合）（2011～22年）

出典CRA タウルビー調査、2023年| 図表：2024年AI指数レポート

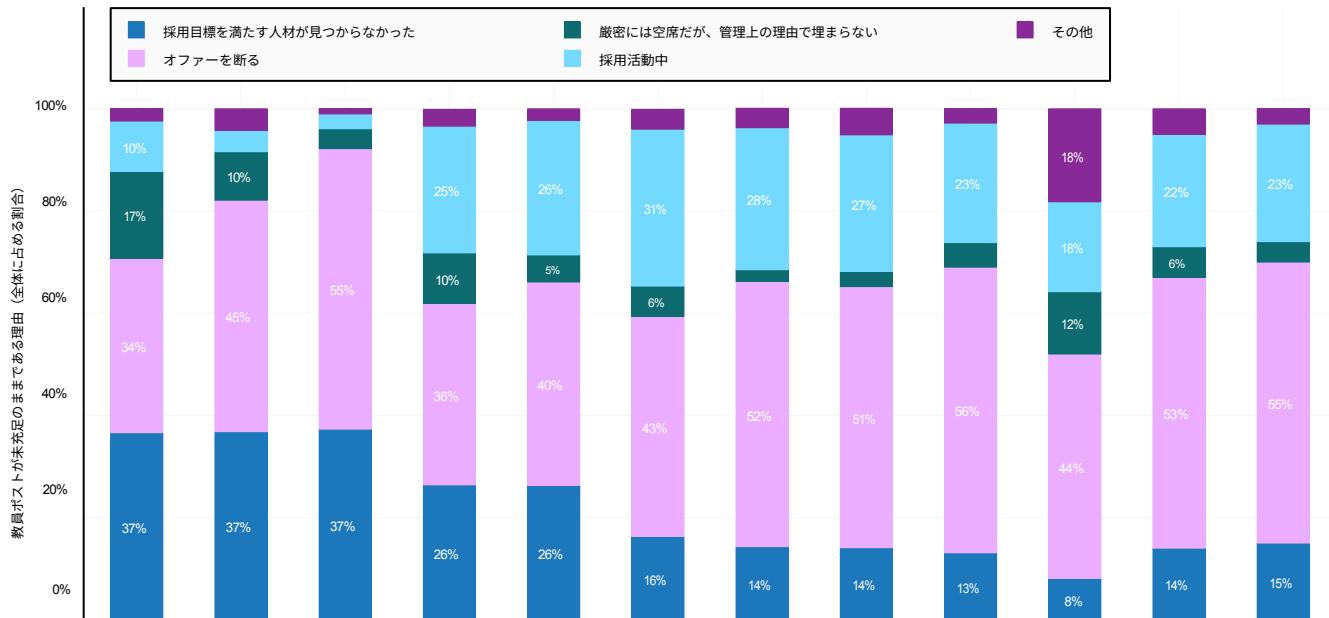


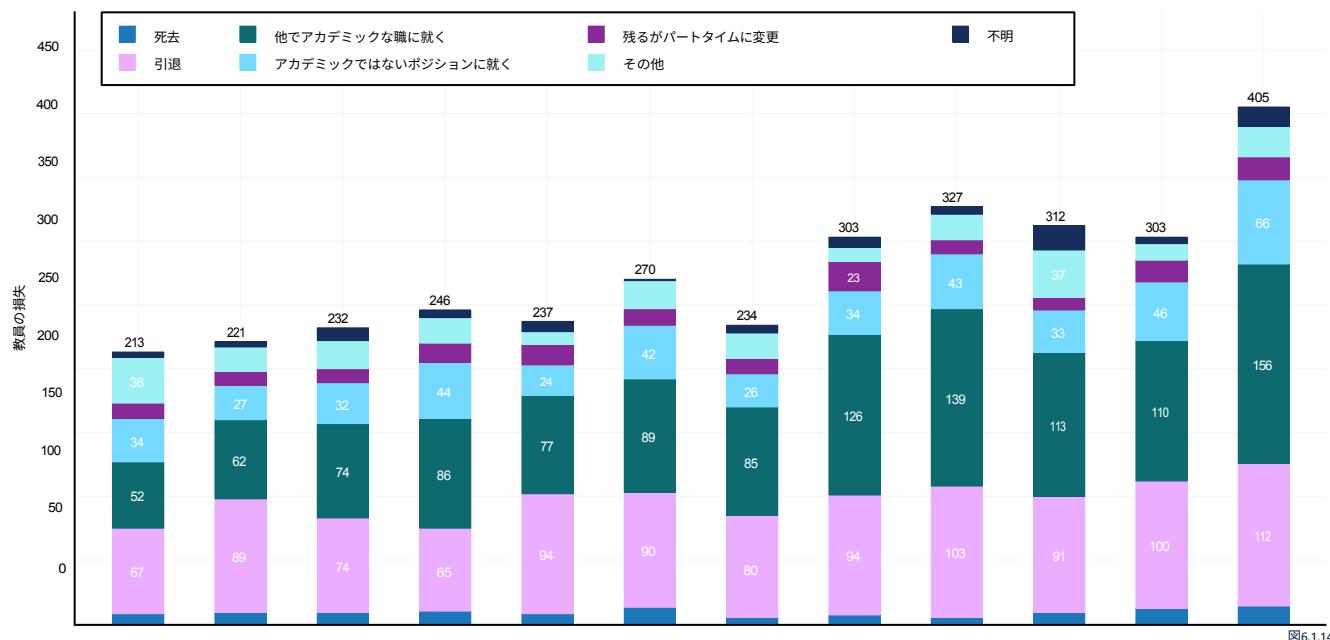
図6.1.13

2022年、北米のCS、CE、情報系学科では、教員の離職が大幅に増加し、2021年の303名から405名となった（図6.1.14）。これらの離職者のうち、38.5%は他の大学に移った。

一方、16.3%は学業以外の職務に移り、これは例年と同じ傾向である。

#### アメリカとカナダのCS、CE、情報学部における教員の減少（2011-22年）

出典CRA タウルビー調査、2023年 | 図表：2024年AI指数レポート



2015年以降、正教授の9ヶ月間の給与中央値の上昇率は米国のインフレ率をわずかに下回っているが、助教授と准教授の給与中央値はインフレ率をわずかに上回っている。2022年、正教授の給与は2021年より3.2%上昇したが、これは米国のインフレ率7%には追いつかず、2015年より16.4%上昇したが、それでもこの間のインフレ率19%を下回った（図6.1.15）。

### CS教員の9ヶ月間の給与中央値（2015～22年）

出典CRA タウルピー調査、2023年|図表：2024年AI指数レポート

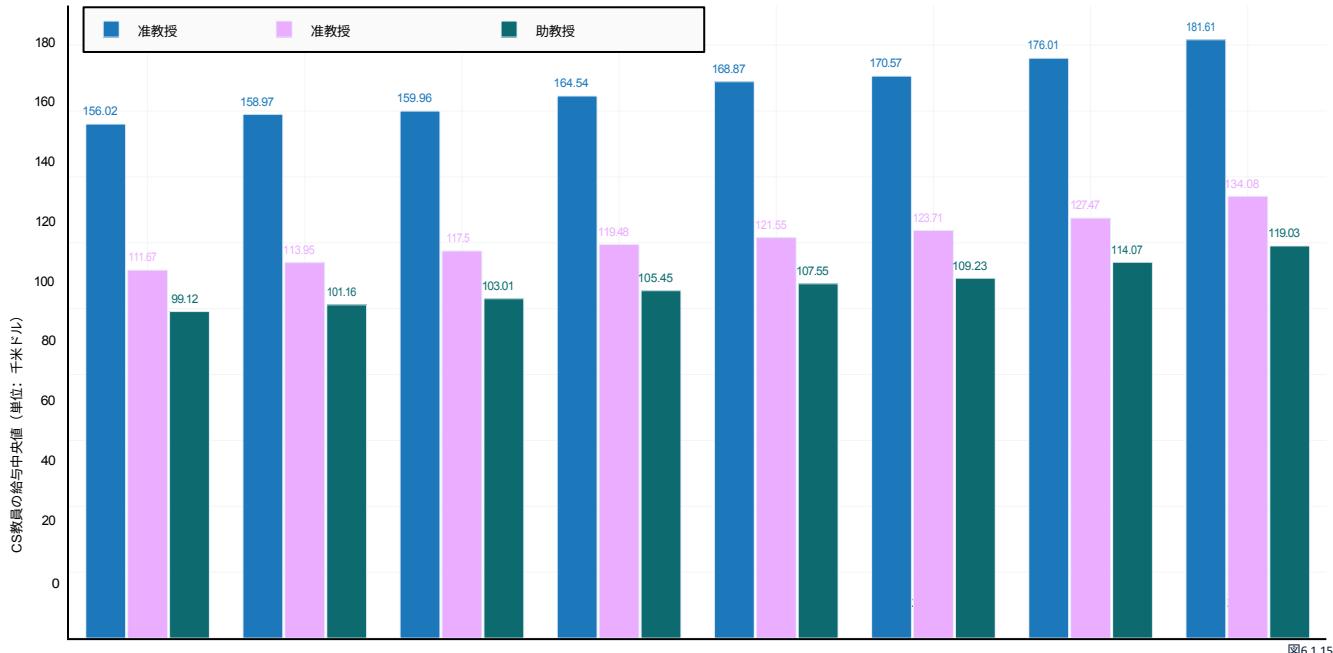


図6.1.15

2022年、CS、CE、情報分野のテニュアトラック占める外国人採用者の割合は、前年の13.2%から19.3%へと大幅に増加した（図6.1.16）。これは過去10年間で2013年に次いで2番目に高い割合である。

米国およびカナダにおけるCS、CE、情報のテニュアトラック教員の新規外国人採用（全体に占める割合）（2010～22年）

出典CRA タウルビー調査、2023年 | 図表：2024年AI指数レポート

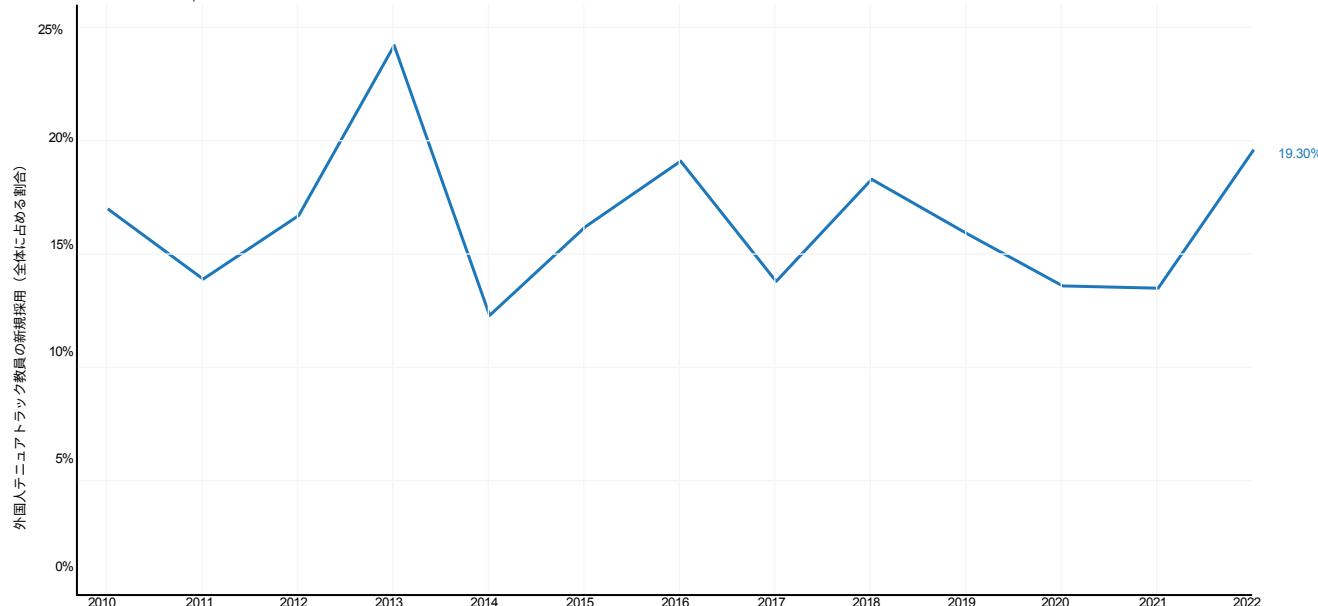


図6.1.16

## ヨーロッパ

欧州のCS卒業生に関するデータは、特に欧州における情報教育の状況を監視することを目的とした学術研究コミュニティである [Informatics Europe](#)によるである<sup>3</sup>、欧州政府統計局のCS、CE、コンピューティング、情報技術（IT）分野<sup>4</sup>

### 情報学、CS、CE、IT学士号 卒業生

2022年、学士レベルの情報学、CS、CE、IT分野の新卒者数が最も多かったのは英国で、合計約25,000人であった（図6.1.17）<sup>5</sup>。に近い。ポーランド、スペイン、チェコ共和国のような例外はあるが、サンプルに含まれるほとんどの国で、これらの分野の卒業生が10年前と比べて増加している（図6.1.18）。

#### 欧州における情報学、CS、CE、ITの学士号取得者の国別新規採用数（2022年）

出典Informatics Europe, 2023|チャート: 2024年AI指数レポート

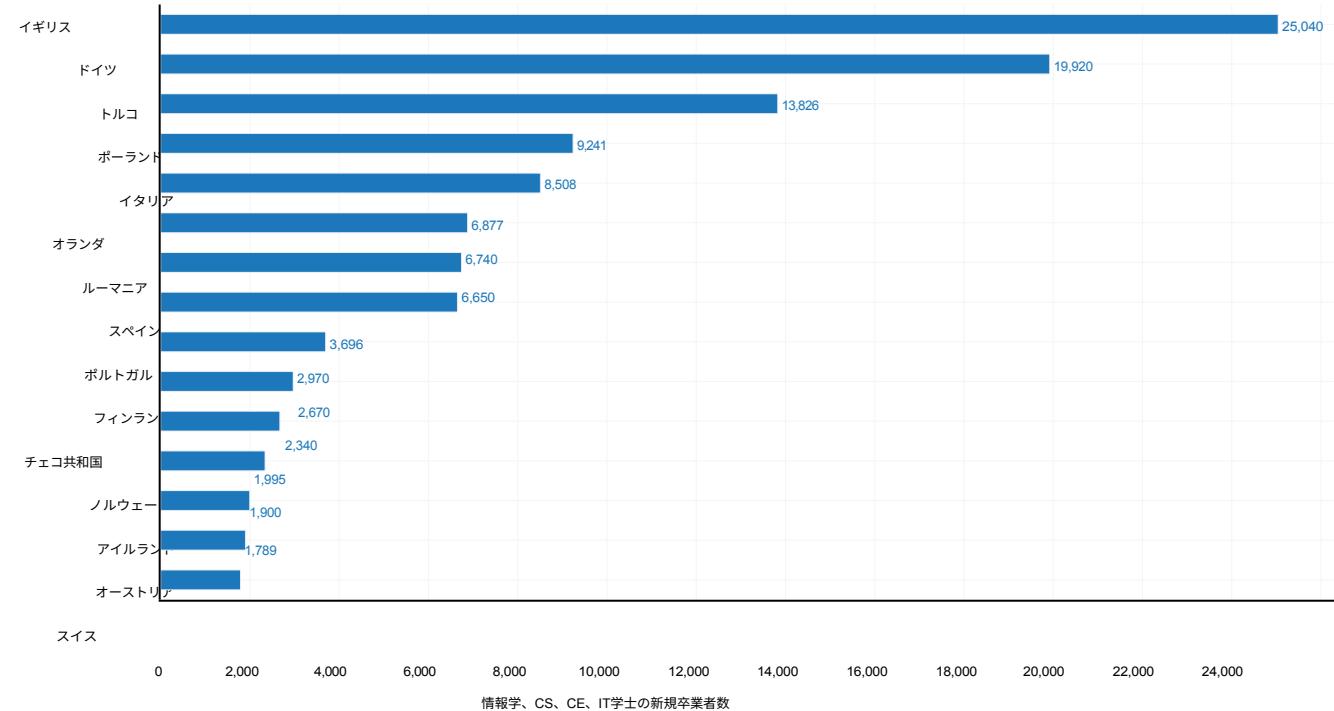


図6.1.17

<sup>3</sup> ヨーロッパ諸国で統一的に使用されているCS教育の用語はない。欧州全体では、CS教育は、情報学、コンピュータ科学（CS）、コンピュータ工学（CE）、コンピューティング、情報技術（IT）、情報通信技術（ICT）、情報科学技術（IST）などの用語に反映される。Informatics Europeが情報学プログラムの識別に使用している科目名（および英訳）の全リストは、[以下のリンク](#)から入手できます。

<sup>4</sup> 欧州のデータは各国の統計局から収集され、より広い範囲をカバーしているためである。

<sup>5</sup> AIインデックスがデータを持つすべての国が、本章の図表で可視化されているわけではないことに留意されたい。完全なデータにアクセスするには、本章に関連する公開データをご覗ください。また、年号は年度末の年を示す。例えば、2022年度の新卒者を可視化した図は、2021/2022報告された新卒者数を反映している。視覚的にわかりやすくするために、当インデックスでは、学生が卒業した年に焦点を当てることにしている。

## 欧州における情報学、CS、CE、ITの学士号取得者の国別増減率（2012年と2022年の比較）

出典Informatics Europe, 2023|チャート: 2024年AI指数レポート

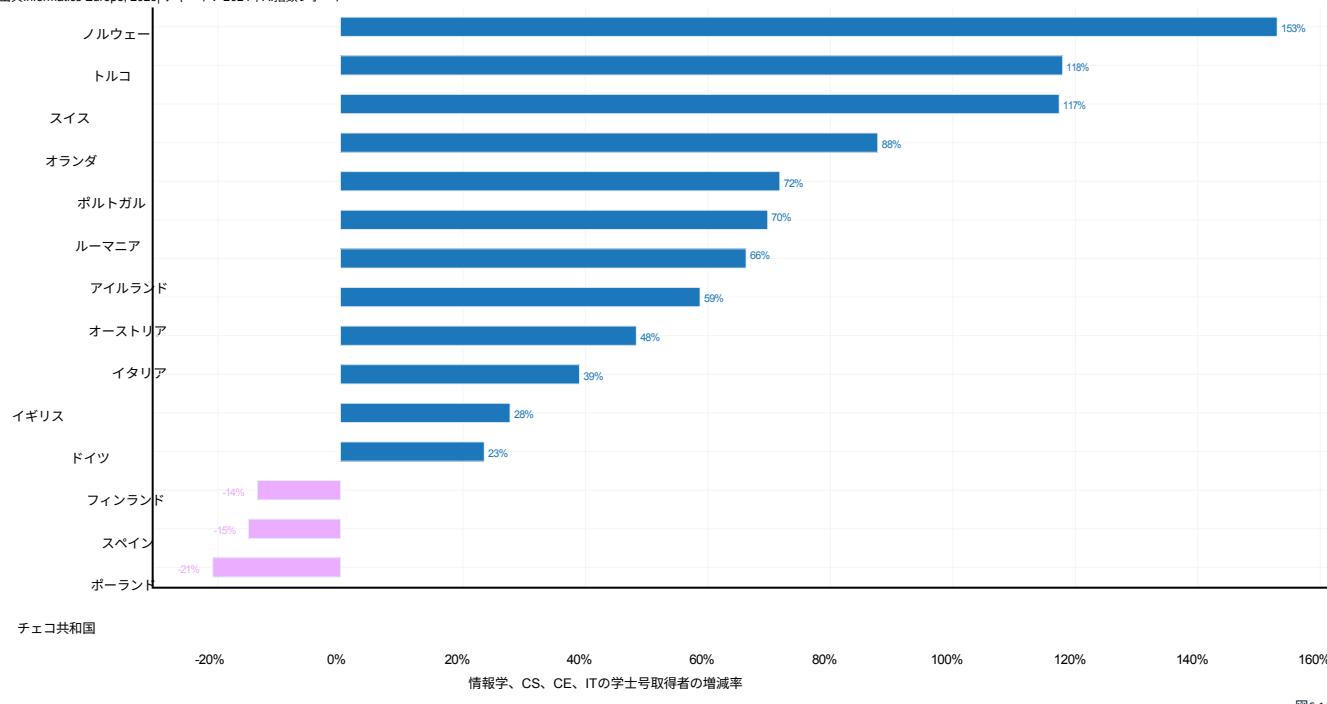


図6.1.18

フィンランド（53.4人）、ノルウェー（42.6人）、オランダ（38.6人）は、人口10万人当たりの情報学、CS、CE、ITの学士号取得者数でトップである（図6.1.19）。一人換算すると、ほとんどの欧州諸国において、情報学、CS、CE、ITの号取得者の総数が増加している（図6.1.20）。