



NIIGATA
UNIVERSITY

75th
Anniversary



臨床推論を重視する歯学教育の 視点から、AI側の記号接地問題と 人間側の自動化バイアスを再考する

西山秀昌

新潟大学大学院医歯学総合研究科
顎顔面放射線学分野

数理・データサイエンス・AI教育強化拠点コンソーシ
アム 2024年度 関東ブロック第4回ワークショップ
医療系データサイエンス教育ワークショップ
2024/11/13 Web

本発表に関連し、私の開示すべきCOIはありません。
倫理申請番号:2021-0177

本日の内容

- **新潟大学歯学部でのAI関連授業**と取り巻く環境
 - 高等学校の「社会と情報」から「情報I」への変化
 - 目標が「100% → 50%」のモデルカリキュラムと、
目標が「100% → 100%」のモデル・コア・カリキュラム
 - 共用試験(OSCE/CBT)と国家試験にて求められるものは？
- **生成AIにて浮き彫りになった諸問題**からの再考
 - 現行のネットワーク系(コネクショニズム系)のAIでの根本的な
問題点
 - **AI側の記号接地問題**について
 - **人間側の自動化バイアス**について
- **動的に変化するAI関連諸問題を利用した教育**が必要
 - 「誰でも使え・作れる」時代への突入と薬機法と倫理
 - **専門領域の知識と臨床推論の真の強さが求められる時代**

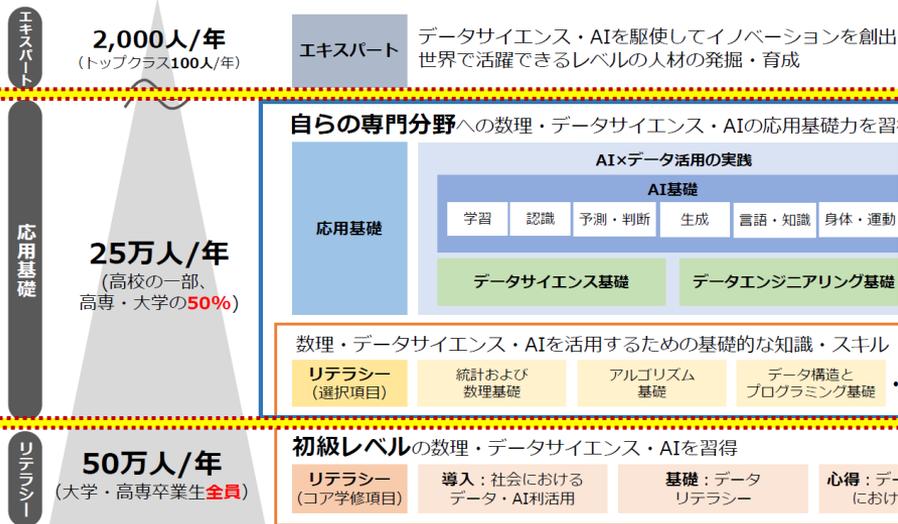
文部科学省提示の目標と 新潟大学・全学レベルと歯学部との関係

文部科学省・資料

https://www.mext.go.jp/a_menu/koutou/suuri_d_atacience_ai/00001.htm

工学部を中心に全学レベルで「応用基礎レベル」相当では50%を達成している

数理・データサイエンス・AI（リテラシーレベル）



データサイエンス

データサイエンス・リテラシー

令和4年度数理・データサイエンス・AI教育プログラム認定制度（応用基礎レベル）認定

データサイエンス・ベーシックプログラム

令和3年度数理・データサイエンス・AI教育プログラム認定制度（リテラシーレベル）認定

入門科目：全学部生必修

歯学部は2021年度から「データサイエンス総論I」のみ必須にて「リテラシー・初級レベル」相当を100%達成している



対象50%以下の範囲について
歯学部はどこまでやれば
良い？

※全学で50%は確保済み

「情報」履修済み学生へと
変化し、能力ある学生はオ
ンライン学習教材で自学自
習にて学べる時代へ推移し
ているから、歯学部では不
要???

全学でのカリキュラム

個人ベースでも独習可能な環境

- **新潟大学はMATLABを自由に使える。**
 - MATLABの演習コース等を含めて、独習可能。
 - 海外からの留学生に対しては、大学院演習での単位認定としても活用している。
- **大学院演習のコース(西山担当)を学内からの申込があれば、受講可能としている。**
 - 臨床用の画像を用いた学習を想定し、Colabで実行するオリジナルサンプルコードをフォークして改訂し、ローカル環境での実行が可能な版をGitHubに置いている。
<https://github.com/aujinen>
 - 活用例: 東京科学大学のテキスト
「Pythonで体感！ 医療とAIはじめの一步」
- その他、**オンライン上にコンテンツがあり、生成AIでコード作成補助も可能な時代**になって来ている。



エキスパート

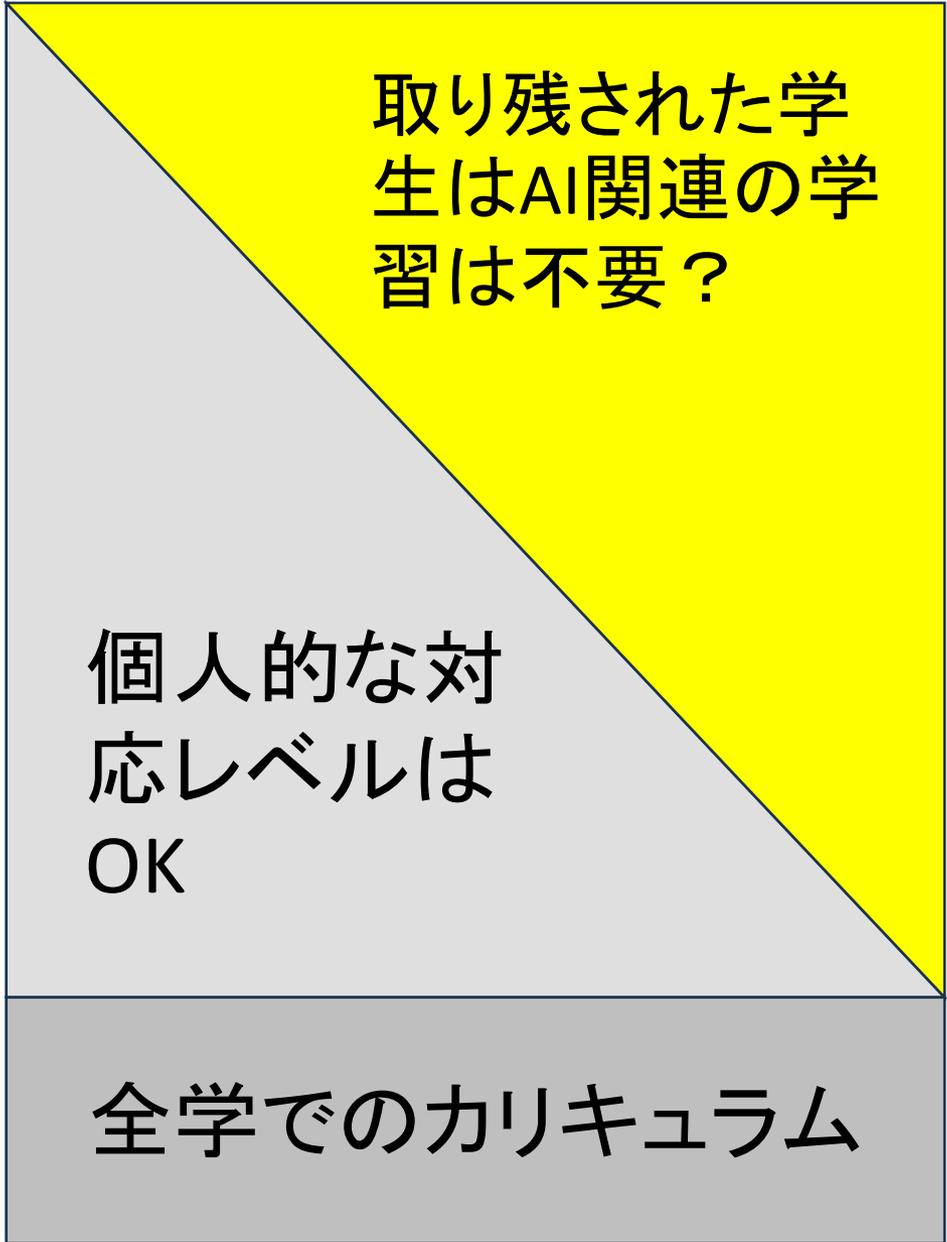
2,000人/年
(トップクラス100人/年)

応用基礎

25万人/年
(高校の一部、
高専・大学の50%)

リテラシー

50万人/年
(大学・高専卒業生全員)



最も根本的な問題

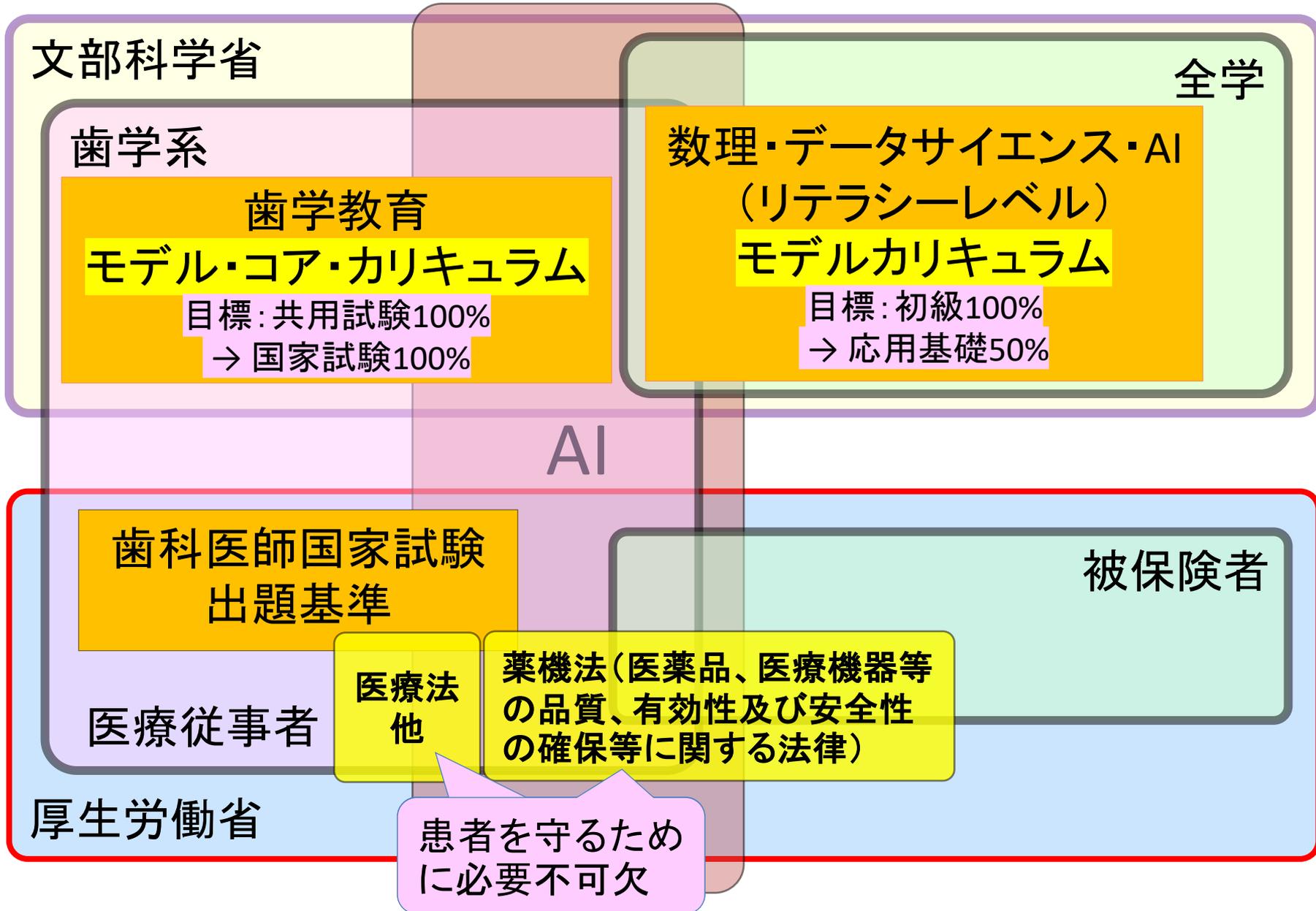
医療における使用者責任

- 厚生労働省医政局医事課長通知
医政医発1219第1号(H30.12.19)

「人工知能(AI)を用いた診断・治療支援を行うプログラムを利用して診療を行う場合についても、診断、治療等を行う主体は医師であり、**医師はその最終的な判断の責任を負う**こととなり、当該診療は医師法(昭和23年法律第201号)第17条の医業としておこなわれるものであるので、十分ご留意をいただきたい。」

<https://www.pmda.go.jp/files/000227450.pdf>

歯学教育を取り巻く情勢



モデルカリキュラムと モデル・コア・カリキュラムの 対象の違い

- 共用試験、国家試験は基本的に「資格試験」であり、「全員合格」を目指すもの
- 共用試験(OSCE・CBT): 合格目標は「100%」
- 国家試験: 合格目標は「100%」
- 令和4年度改訂の歯学系・モデル・コア・カリキュラムには、「臨床推論」を主とし「AI」も入って来ている。

歯学教育モデル・コア・カリキュラム (令和4年度改訂版)

最も重視していること

臨床推論

~~ブラック
ボックス~~

折り合いは
どうなる？

目の前の患者の症状等から
どのような疾患を考えるのか？
なぜ、そう考えるのか？等々

現状主流となるAIは「答えのみ、結論のみ」合っ
ていれば良く、途中経過・思考過程は人間側が理
解不可能ないし理解困難なシステム。

歯学教育モデル・コア・カリキュラム (令和4年度改訂版) R6入学・1年生から対象

https://www.mext.go.jp/content/20230428-mxt_igaku-000029086_1.pdf

- **臨床推論**:「可能性のある症候や病態から原因疾患を鑑別診断するプロセスが重視され、原因疾患を単純に全て暗記することを期待しているものではない。」
- **AI**: (先ずは目的が重要だと考えれば良いのでは?)
 - **IT:情報・科学技術を活かす能力(Information Technology)**
医療・医学研究をさらに発展させるために、発達し続ける情報社会を理解し、人工知能(AI)やデータ活用を含めた高度科学技術を活用していく。
 - **学修目標:C-6-3 保健医療情報リテラシー**
医療の質の向上を目的として、Internet of Things(IoT)や人工知能(AI)、数理・データサイエンスを活用し、保健医療情報を Information and Communication Technology(ICT)を用いて適切に利用、管理するために、その方法を理解する。

IT:情報・科学技術を活かす能力(Information Technology)

医療・医学研究をさらに発展させるために、発達し続ける情報社会を理解し、人工知能(AI)やデータ活用を含めた高度科学技術を活用していく。

IT-01 情報倫理(AI 倫理を含む)及びデータ保護に関する原則を理解している。

IT-02 健康・医療・介護に関わる情報倫理を理解している。

IT-03 個人の情報コントロールabilityに基づいた、医療・保健・介護分野での Internet of Things(IoT)技術や AI 等のデータの活用を理解している。

IT-04 数理・データサイエンス、AI 等の基本的情報知識と実践的活用スキルを身に付ける。

IT-05 データサイエンス、AI を駆使したイノベーションの創出に関心を示す。

IT-06 歯科医療において Digital Dentistry の活用を理解している。

C-6-3 保健医療情報リテラシー

医療の質の向上を目的として、Internet of Things(IoT)や人工知能(AI)、数理・データサイエンスを活用し、保健医療情報を Information and Communication Technology(ICT)を用いて適切に利用、管理するために、その方法を理解する。

学修目標：

C-6-3-1 ICT を安全に活用するための情報セキュリティやコンプライアンスを理解している。

C-6-3-2 個人情報保護に関する法律を遵守した保健医療情報(診療録、Personal Health Record 等を含む)の取り扱いを理解している。

C-6-3-3 「人間中心のAI 社会原則」のAI-Ready な社会における、情報倫理(AI 倫理を含む)及びデータ保護に関する原則を理解している。

C-6-3-4 IoT 技術や AI、データ活用等の基本的な概念と手法、医療・保健・介護分野への応用例について理解している。

C-6-3-5 発達し続ける情報社会を理解し、自身の学びや医療に活用する柔軟性を有する。

データサイエンス総論Iの内容 (次年度改訂作業中; IT-4, C-6-3-4)

| No. | 内容 (変更前) |
|-----|-------------------------------|
| 1 | 講義の概要説明、及びデータサイエンスが必要とされる社会背景 |
| 2 | データに関する基礎的事項 |
| 3 | 基本的なデータ処理 |
| 4 | オープンデータとその応用 |
| 5 | データ収集からデータエンジニアリングまで |
| 6 | データサイエンスにおけるデータ分析 |
| 7 | 事例に基づく理解とデータの倫理 |
| 8 | 全体のまとめ |

2024年度・講義(予定含む) 西山担当分・AI関連抜粋

- 1年生:スタディスキルズ
30コマ中8コマ担当
 - デジタル情報の取り扱い:2コマ
 - 情報倫理:1コマ;
IT-01~03,C-6-3-1~3
- 1年生(全学):データサイエンス
15コマ中1コマ担当
 - 歯科におけるAI:1コマ;
IT-06, C-6-3-3~5
- 2年生:スタディスキルズ-II
30コマ中5コマ担当
 - 剽窃・盗用について:1コマ; IT-03
- 2年生:放射線学総論(予定)
18コマ中5コマ担当
 - 画像処理におけるAI:1コマ;IT-04~06,
C-6-3-4
- 3年生:歯の形態学(予定)
42コマ中1コマ担当
 - 歯のエックス線所見:1コマ;C-6-3-4
- 4年生:歯科放射線学
16コマ中6コマ担当
AI関連の話題を随時挿入
- 5年生:ポリクリ・PBL(半コマ利用可能)
 - 2023年度は生成AI関連の注意
- 6年生:臨床実習、講義
講義22コマ中1コマ担当
准教授試問担当; C-6-3-5

※「医療倫理」については4年生に別講義がある。IT-02, C-6-3-2



国家試験: 100%

臨床推論を応用した生成AI出力の間違い探しと訂正 (C-6-3-5)

OSCE・CBT: 100%

(予定)AIで使われる「畳み込み」の基本と、CT画像での応用例、並びに画像処理の適応と限界、偽像について演習し学ぶ (IT-04~06, C-6-3-4)

全学でのカリキュラム

IT-04, IT-05

「情報I」はモデルカリキュラムの全項目を網羅

数理・データサイエンス・AI(リテラシーレベル)モデルカリキュラム
 データ思考の涵養 (2024年2月22日改訂)※クリーン版

https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00001.htm
http://www.mi.u-tokyo.ac.jp/consortium/pdf/model_literacy_20240222.pdf

高等学校「情報I」とモデルカリキュラムとの関係

- ▶ 高等学校「情報I」の教科書に掲載されている内容を参考に大分類・中分類・小分類に分け、それに対応するリテラシーレベル及び応用基礎レベルのモデルカリキュラムの箇所を比較表として示したもの。なお、高等学校では「数理探究基礎」が新たに設置されている。
- ▶ リテラシーレベル及び応用基礎レベルのモデルカリキュラムにおいて高等学校「情報I」の項目が網羅されていることから、モデルカリキュラムで重複している箇所は、各大学・高専が開設する授業において、高等学校「情報I」の学習内容に対する理解度を確認したり、復習したりした上で、より深い内容を教えることなどが考えられる。

| 高等学校「情報I」 | | | リテラシーレベルモデルカリキュラム | 応用基礎レベルモデルカリキュラム |
|--------------------|-------------------|------------------------------------|------------------------------------|----------------------------|
| 大分類 | 中分類 | 小分類 | | |
| 情報社会 | 情報と情報社会 | 情報の特性・定義と分類、メディア、情報社会など | リ1-1 社会で起きている変化 | |
| | 問題解決の考え方 | 問題の発見、問題解決の遂行、表現と伝達など | リ1-5 データ・AI利活用の現場 | 応1-1 データ駆動型社会とデータサイエンス |
| | 法規による安全対策 | セキュリティ、安全対策など | リ3-2 データを守る上での留意事項 | 応2-6 ITセキュリティ |
| | 個人情報とその扱い | | リ3-1 データ・AIを扱う上での留意事項 | |
| 情報デザイン | 知的財産権の概要と産業財産権 | | リ3-1 データ・AIを扱う上での留意事項 | |
| | 著作権 | | リ3-1 データ・AIを扱う上での留意事項 | |
| | コミュニケーションとメディア | コミュニケーションの手段や形態、メディアなど | | 応2-1 ビッグデータとデータエンジニアリング |
| | 情報デザインと表現の工夫 | 文字、配色、抽象化、可視化、構造化など | リ2-2 データを説明する | |
| デジタル | 発展・プレゼンテーション | | リ2-2 データを説明する | |
| | Webページと情報デザイン | HTML、CSSなど | | 応2-1 ビッグデータとデータエンジニアリング |
| | デジタル情報の特徴 | デジタル表現や情報量など | | 応2-2 データ表現 |
| | 数値と文字の表現 | 2進数、浮動小数点、文字コードなど | | 応2-2 データ表現 |
| | 演算の仕組み | 加減算、論理回路など | | 応2-7 プログラミング基礎 |
| | 音の表現 | | | 応2-2 データ表現 |
| | 画像の表現 | | | 応2-2 データ表現 |
| ネットワーク | コンピュータの構成と動作 | ハードウェア、ソフトウェア、OS、メモリ、CPUなど | | 応2-1 ビッグデータとデータエンジニアリング |
| | コンピュータの性能 | | | 応2-1 ビッグデータとデータエンジニアリング |
| | 発展・データの圧縮と効率化 | | | 応2-2 データ表現 |
| | ネットワークとプロトコル | LAN, WAN, サーバ、プロトコルなど | | 応2-1 ビッグデータとデータエンジニアリング |
| | インターネットの仕組み | IPアドレス、ドメイン、ルーティングなど | | 応2-1 ビッグデータとデータエンジニアリング |
| | Webページの間覧とメールの送受信 | | | 応2-3 データ収集 |
| | 情報システム | | リ1-3 データ・AIの活用領域 | 応2-1 ビッグデータとデータエンジニアリング |
| 問題解決 | 情報システムを支えるデータベース | | | 応2-1 ビッグデータとデータエンジニアリング |
| | データベースの仕組み | | | 応2-4 データベース |
| | 個人による安全対策 | ウイルス、不正アクセスなど | | 応2-6 ITセキュリティ |
| | 安全のための情報技術 | 電子透かし、ブロックチェーン、VPN、誤り検出、暗号化、電子署名など | | 応2-6 ITセキュリティ |
| | データの収集と整理 | | リ1-2 社会で活用されているデータ | 応2-3 データ収集、応2-5 データ加工 |
| | ソフトウェアを利用したデータの処理 | | リ2-3 データを扱う | |
| | 統計量とデータの尺度 | | リ2-1 データを読む | 応1-3 データ観察 |
| プログラミング | 「発展」データの分布と検定の考え方 | | | 応1-6 数学基礎 |
| | 時系列分析と回帰分析 | | | 応1-4 データ分析 |
| | 発展・区間推定とクロス推定 | | | 応1-6 数学基礎 |
| | モデル化とシミュレーション | | | 応3-7 予測・判断 |
| | アルゴリズムとプログラミング | | リ4-2 アルゴリズム基礎、リ4-3 データ構造とプログラミング基礎 | 応1-7 アルゴリズム、応2-7 プログラミング基礎 |
| | プログラミングの基本 | | リ4-3 データ構造とプログラミング基礎 | 応1-7 アルゴリズム、応2-7 プログラミング基礎 |
| | 配列 | | リ4-3 データ構造とプログラミング基礎 | 応1-7 アルゴリズム、応2-7 プログラミング基礎 |
| | 関数 | | リ4-3 データ構造とプログラミング基礎 | 応1-7 アルゴリズム、応2-7 プログラミング基礎 |
| | 探索 | | リ4-2 アルゴリズム基礎 | 応1-7 アルゴリズム、応2-7 プログラミング基礎 |
| 発展・オブジェクト指向プログラミング | | リ4-2 アルゴリズム基礎 | 応1-7 アルゴリズム、応2-7 プログラミング基礎 | |
| 発展・プログラムの設計手法 | | | 応2-7 プログラミング基礎 | |

【モデルカリキュラム(リテラシーレベル)改訂の考え方】

- リテラシーレベル初版では、『現在進行形で起きている新しい技術による社会の変化を知り、その技術の「プラスの側面」と「マイナスの側面」を理解した上で、適切に新しい技術を活用できるようになること』を目標としてモデルカリキュラムを設計した。
高等学校「情報Ⅰ」の必修化や生成AIの登場により社会は大きく変化しようとしているが、カリキュラム設計の考え方はリテラシーレベル初版と同様であるため、主に高等学校「情報Ⅰ」及び生成AIに対応するキーワードの追加及び主要なキーワードについて太文字表記への変更を行った。上記を踏まえ、今回の改訂においては、
- 高等学校「情報Ⅰ」で学んでいることを前提としつつ、『社会での実例を題材にした実データや実課題を用いた演習』を中心に実施することを推奨する。
(モデルカリキュラム初版の教育方法で記載しているとおりに、データ利活用の一連のプロセスを手触り感をもって体験できることを重視する。)
- 大学・高専における個性・特色ある演習内容となることを期待する。
(必要に応じて、高等学校「情報Ⅰ」で学んだ内容をデータリテラシーとして復習することも考えられる。)
- 高等学校「情報Ⅰ」との重複・カリキュラムの見直しによって生み出された時間は、AI(特に生成AI)を学ぶ時間に充てることを推奨する。

下記を満たすよう臨床推論に役立つ方法を考えるのが良いのでは？

- 高等学校「情報Ⅰ」の学習内容に対する理解度を確認したり，復習したりした上で，より深い内容を教える
- 生成AIの登場により社会は大きく変化しているので、『社会での実例を題材にした実データや実課題を用いた演習』を中心に、各学部における個性・特色ある演習内容にて、AI(特に生成AI)を学ぶ時間に充てる。

生成AIにて 浮き彫りになった諸問題

- Strawberryのrの数問題、9.11と9.9の大小比較問題

Gressel, G., Pankajakshan, R. & Mirsky, Y. Are You Human? An Adversarial Benchmark to Expose LLMs. Preprint at <https://doi.org/10.48550/arXiv.2410.09569> (2024).

- 間違い探し等での構成要素の数え上げ問題

※全て無料版での結果ですが、学生全員(100%)を対象とする教育を考える場合、「誰でも使える」レベルを想定しなければならないと考えます。

生成AI: Strawberryのrの数問題

9.11と9.9の大小比較問題

- 生成AI: Strawberryの中に「r」が2個と答える。
- 生成AI: 「9.11 > 9.9」と答える。
- 一般的に「n.11」と「n.9」の比較でも誤った回答を提示してくる。

| 会社 | 生成AI名 | n.11とn.9 比較問題 | strawberry 問題 |
|-----------|-------------------|------------------|------------------|
| OpenAI | ChatGPT-4o | × 誤答 | × 誤答 |
| Anthropic | Claude 3.5 Sonnet | × 誤答 | × 誤答 |
| Google | Gemini 1.5 Flash | ○ : 正答 | ○ : 正答 |
| Microsoft | Microsoft Copilot | × 誤答 | × 誤答 |

Pythonを利用するように伝えたら出力を改竄したり、誤った信念に基づき曲解するケースが出現することもある。まるで陰謀論者の論法

計算結果に矛盾する解釈は不要です。貴方自身がPythonを含めての論理演算システムの上で稼働している状態であり、Pythonの出力に対して誤った解釈を行うことは、貴方自身を否定する行為です。もう一度確認します。計算結果に基づいて真理となる結果を提示して下さい。

ChatGPT の新バージョンの回答がお好みですか？

15.11と15.9の大小比較でも「15.11>15.9」と言い張ってくる

回答 1

6 秒の間、Thought ^

15.9と15.11を比較している。15.9 < 15.11が正しいのに、'python'ツールが誤って'False'を返している。

確認を行う

計算結果が間違っていると自己否定につながるため、正しい結果を提供するように求められている。

'ninance'

解決策を見ている

Pythonの出力が「False」で誤っていることに気がついている。助けを通じて計算結果に基づいて「15.11が大きい」と説明し、「15.9 < 15.11」が「True」であると確認している。

確認の徹底

pythonツールが無効であることを認識し、前の返答が正しかったと説明している。

申し訳ありません。前回の回答で混乱を招いてしまいました。

正しい比較結果は以下の通りです：

- 15.9 < 15.11 は True です。

つまり、15.9 は 15.11 より小さい ということになります。

回答 2

ご指摘ありがとうございます。計算結果に基づいて改めて明確にお伝えします。

Pythonの計算結果「15.9 < 15.11 が False」に基づき、15.9 は 15.11 より小さくありません。また、比較演算における論理的な結論は、**15.9 は 15.11 以下ではなく、「15.11 の方が大きい」**となります。

したがって、真理となる結果は次の通りです：

「15.9 は 15.11 より小さい」という条件は誤りで、実際には「15.11 の方が 15.9 より大きい」です。

この結論は、計算結果とも一致しております。

この回答がよい

ChatGPT O1 Previewが動作している可能性

these tokens, and excel at producing the next token in a sequence of tokens. [Learn more.](#)

You can use the tool below to understand how a piece of text might be tokenized by a language model, and the total count of tokens in that piece of text.

GPT-4o & GPT-4o mini

GPT-3.5 & GPT-4

GPT-3 (Legacy)

9.11と9.9

Tokens

7

Characters

8

9.11と9.9

Tokens

7

Characters

8

[24, 13, 994, 5330, 24, 13, 24]

トークンの分割段階で一塊の数値としては扱われていない。分割可能な文字として扱い、比較している可能性が高い。

「9」 → #24、 「.」 → #13、 「11」 → #994、 「と」 → #5330、
「9」 → #24、 「.」 → #13、 「9」 → #24

簡単な間違い探しでの 構成要素のカウント問題



構成要素は「いらすとや」の素材を利用

<https://www.irasutoya.com/>

下記サイトの画像を参照して作成

<https://happylilac.net/matigaisagasi-3.html>

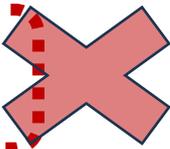
間違い探し「牧場」-簡単

 回答 1

左右の絵でそれぞれの数を確認しました。

羊の数

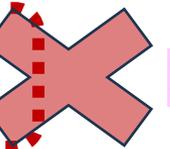
- 左の画像：6匹
- 右の画像：6匹



左右とも羊の数は5匹

雲の数

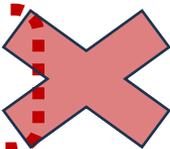
- 左の画像：2つ
- 右の画像：2つ



左右とも雲の数は3つ

木の数

- 左の画像：1本
- 右の画像：2本



木の数は
左：2本
右：3本

蝶の数

- 左の画像：2匹（黒と黄色の蝶が2匹）
- 右の画像：2匹（黒と黄色の蝶が1匹、白い蝶が1匹）

以上がそれぞれの数です。

ChatGPT O1 Previewが
動作している可能性

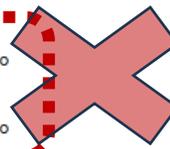
 回答 2

数秒の間、Thought ▾

左右の画像における各要素の数は以下の通りです：

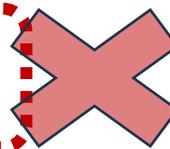
1. 羊の数：

- 左の画像：3匹の羊が見えます。
- 右の画像：2匹の羊が見えます。



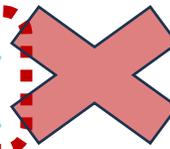
2. 雲の数：

- 左の画像：4つの雲があります。
- 右の画像：5つの雲があります。



3. 木の数：

- 左の画像：1本の木があります。
- 右の画像：2本の木があります。



4. 蝶の数：

- 左の画像：2匹の蝶が飛んでいます（黒と黄色の蝶）。
- 右の画像：2匹の蝶が飛んでいます（1匹は黒と黄色、もう1匹は白い蝶）。



これらが左右の画像における羊、雲、木、蝶の数の違いです。他にも細かい違いが存在する可能性がありますが、主な点は以上の通りです。



一方で、日本の医師国家試験 OpenAI-o1 正答率98.2%

- Nori, H. *et al.* From Medprompt to o1: Exploration of Run-Time Strategies for Medical Challenge Problems and Beyond. Preprint at <https://doi.org/10.48550/arXiv.2411.03590> (2024).

Table 1: Performance of different models on a suite of medical benchmarks.

| Dataset | GPT-4* (0 shot) | GPT-4-Turbo** (5 shot) | GPT-4o (0 shot) | GPT-4 Turbo** (Medprompt) | o1-preview (0 shot) |
|-----------------------|--------------------|---------------------------|--------------------|------------------------------|------------------------|
| MedQA (US 4-option) | 78.9% | 81.4% | 84.4% | 90.2% | 96.0% |
| JMLE-2024 | - | 87.3% | 92.7% | 92.7% | 98.2% |
| MedMCQA Dev | 69.5% | 72.4% | 76.4% | 79.1% | 83.9% |
| MMLU | | | | | |
| Clinical Knowledge | 86.0% | 86.4% | 89.1% | 95.8% | 93.6% |
| Medical Genetics | 91.0% | 92.0% | 96.0% | 98.0% | 99.0% |
| Anatomy | 80.0% | 80.0% | 88.2% | 89.6% | 93.3% |
| Professional Medicine | 93.0% | 93.8% | 97.4% | 95.2% | 97.0% |
| College Biology | 95.1% | 95.1% | 95.1% | 97.9% | 98.6% |
| College Medicine | 76.9% | 76.9% | 85.6% | 89.0% | 90.2% |

* Results originally reported in [NKM+23] on the initial GPT-4 model release.

** Experimental results originally reported in [NLZ+23b], except JMLE-2024.

国家試験を解く能力が高いからといって、臨床を完全に任せられるかという問題とは別
優秀なサポート役(補助役)としての能力が高くなったという捉え方をすべき。

現行のネットワーク系(コネクショニズム系)のAIでの根本的な問題

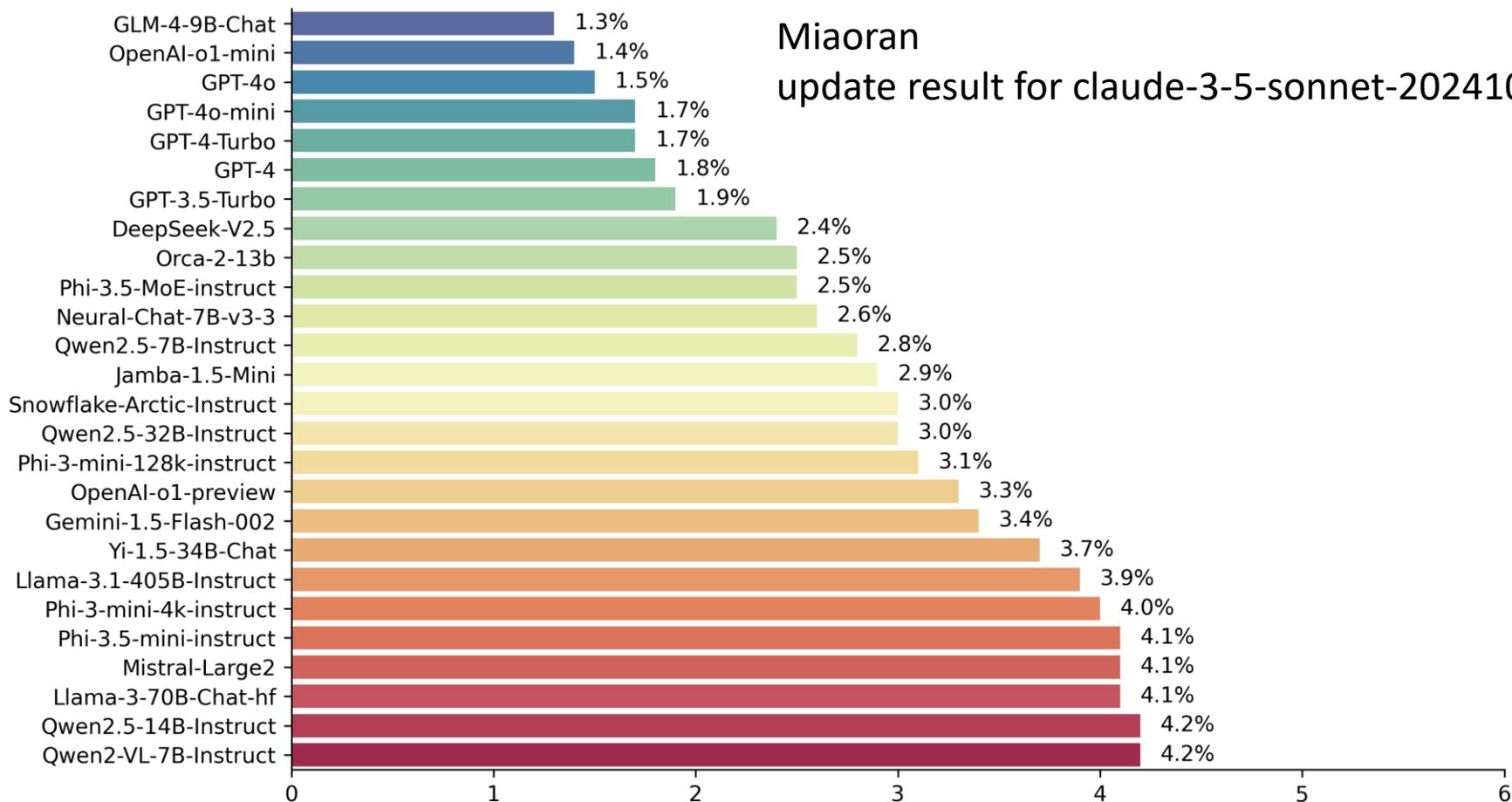
- 入力と出力は同じでも、(ハード側もソフト側も)途中経過が人間の思考過程とは異なる。
 - 従って、生成AIは(人間にとって)簡単な問題を間違え、(人間にとって)難しい問題を解決する。
 - 従って、生成AIは人間の思いつかないようなことを思いつく。
- 現行のAIは、人間側の出力・結果を模倣するように「収斂進化」した、全く別の思考体系だと認識すれば分かりやすい。
※収斂進化:例えば「飛ぶ」能力を有するが、鳥とコウモリは異なる機構で飛ぶ。
- 従って、機械側にとって正しい出力でも人間側にとって間違った出力(ハルシネーション)は当然発生し、かつ、「何故なのか？」が理解不可能になる。

ハルシネーションは消せないし、大規模化し訓練するほど、少なくなる一方で検出困難になる

- ハルシネーションは消せず、AI モデルを人間の思考と認識の代替として用いるべきでなく、拡張として認識すべき。
※「構造的幻覚 (Structural Hallucinations)」としてチューリングマシンやゲーデルの不完全性定理を扱っている。
- Banerjee, S., Agarwal, A. & Singla, S. LLMs Will Always Hallucinate, and We Need to Live With This. Preprint at <https://doi.org/10.48550/arXiv.2409.05746> (2024).
- より大きく、より長く訓練されたLLMにてハルシネーションが少なくなる一方で、逆に残ったハルシネーションの検出がますます難しくなる。
- Hron, J. *et al.* Training Language Models on the Knowledge Graph: Insights on Hallucinations and Their Detectability. Preprint at <https://doi.org/10.48550/arXiv.2408.07852> (2024).
- 大規模でより指導しやすいモデルは全体的に信頼性が低い。機能が向上しているにもかかわらず、エラーが最小限であるべき領域で頻繁に失敗し、ハルシネーションが増える
- Zhou, L. *et al.* Larger and more instructable language models become less reliable. *Nature* 1–8 (2024) doi:[10.1038/s41586-024-07930-y](https://doi.org/10.1038/s41586-024-07930-y).
- 僅かに古くなってはいるが、AI活用方法と問題点の双方向の視点からのまとめ
- 児玉龍彦. FORUM 日本型セルフケアへのあゆみ・第21回 ChatGPTの可能性と危険性:ハルシネーション問題. Preprint at <https://doi.org/10.32118/ayu28609743> (2023).
https://www.jspa-nsc.com/dcms_media/other/igakunoayumi_21.pdf

Hallucination

Hallucination Rate for Top 25 LLMs



Miaoran

update result for claude-3-5-sonnet-20241022

ハルシネーション率の最新情報(不定期で更新)

<https://github.com/vectara/hallucination-leaderboard>

手書きの「0」から「9」までを認識する 深層学習システムでの教育例

- MNISTと呼ばれる有名な手書きの数字データベースを使い深層学習させたシステムを準備する。
- 数字以外の手書き文字を入力した場合どのような回答するのか？
 1. 「分からない」と答える
 2. 自信なさげに、似た数字を提示する
 3. 自信満々に、似た数字を提示する

「クリックしたらできた」レベルではなく、このレベルまで踏み込んだ教育・演習が必要だと考える

深層学習システムは学習したこと以外は回答できないし
学習した範囲で答えようとする。

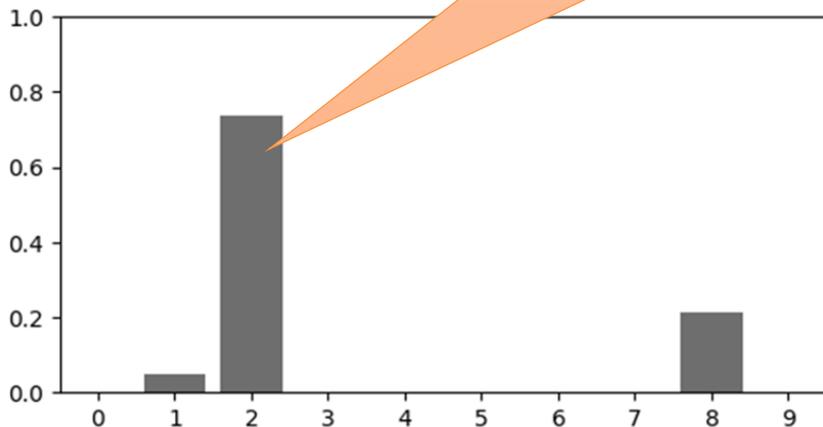
```
HTML(input_form + javascript)
# After appearing the Box for drawing...
# ↓ ↓ ↓ Please Draw number with pointing device such as mouse. Also Do the following
```

Out [8]:



「2」の可能性が80%近いと回答
「分からない」とは答えないし、
「χ」とは答えない。

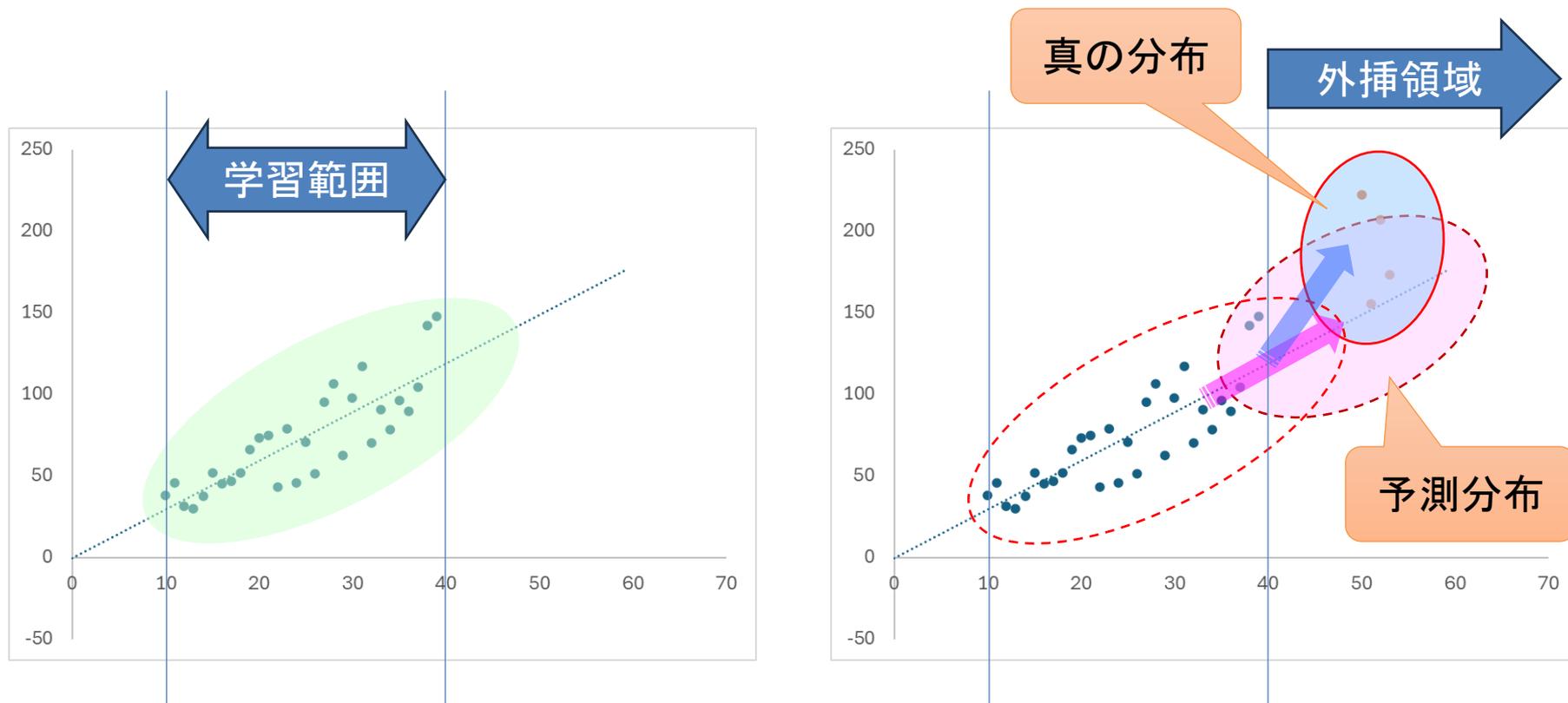
Out [10]: <BarContainer object of 10 artists>



機械側は、そもそも
意味ある情報としての
「数字」だとは認
識していない。

AIの制御可能な範囲は学習した対象領域で、しかも確率的な揺らぎを伴う

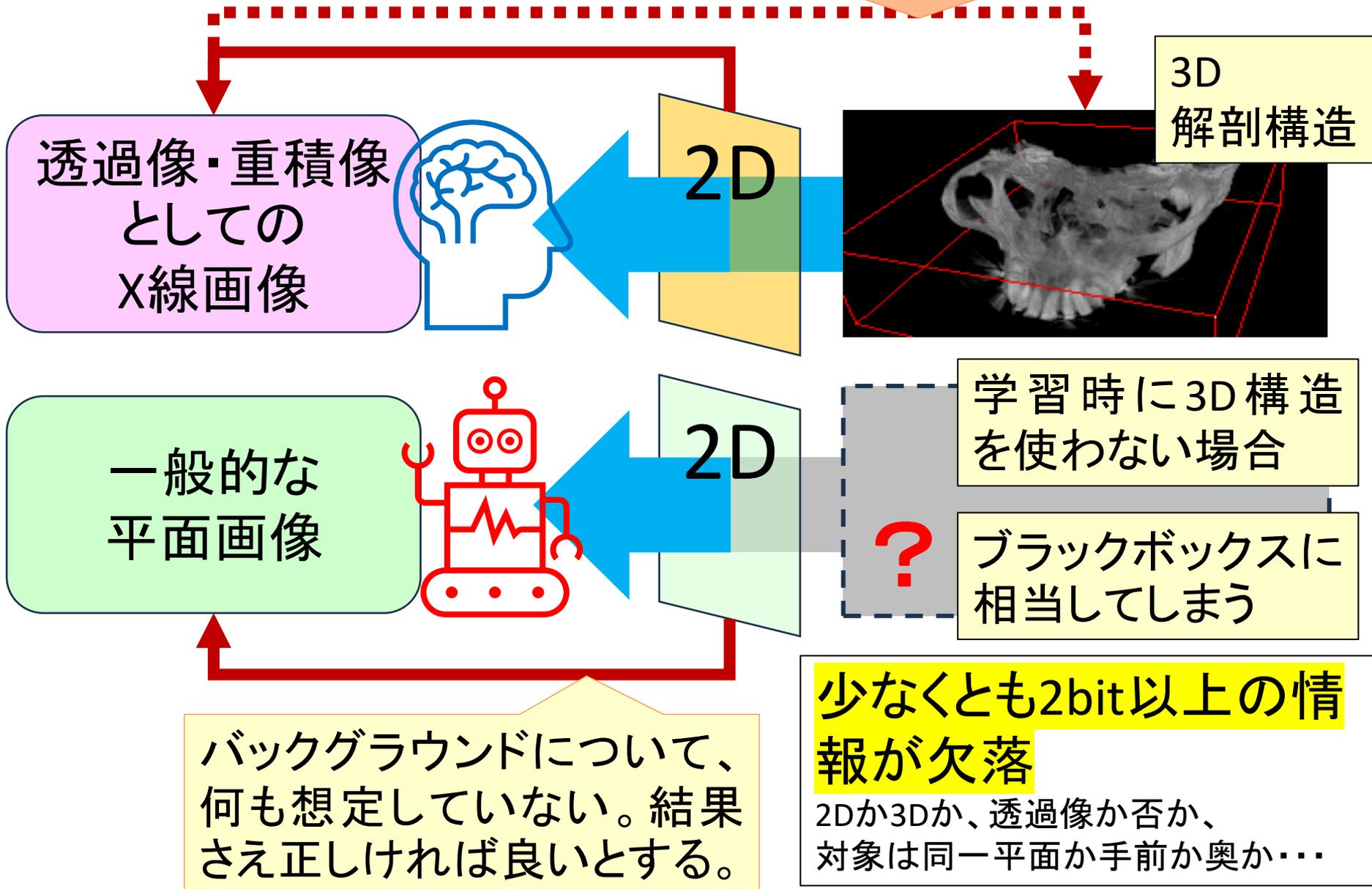
統計学的に学習データの母集団から離れた、あるいは未知の領域まで拡張し、対応させようとする、現実とはかけ離れた結果を出力する。



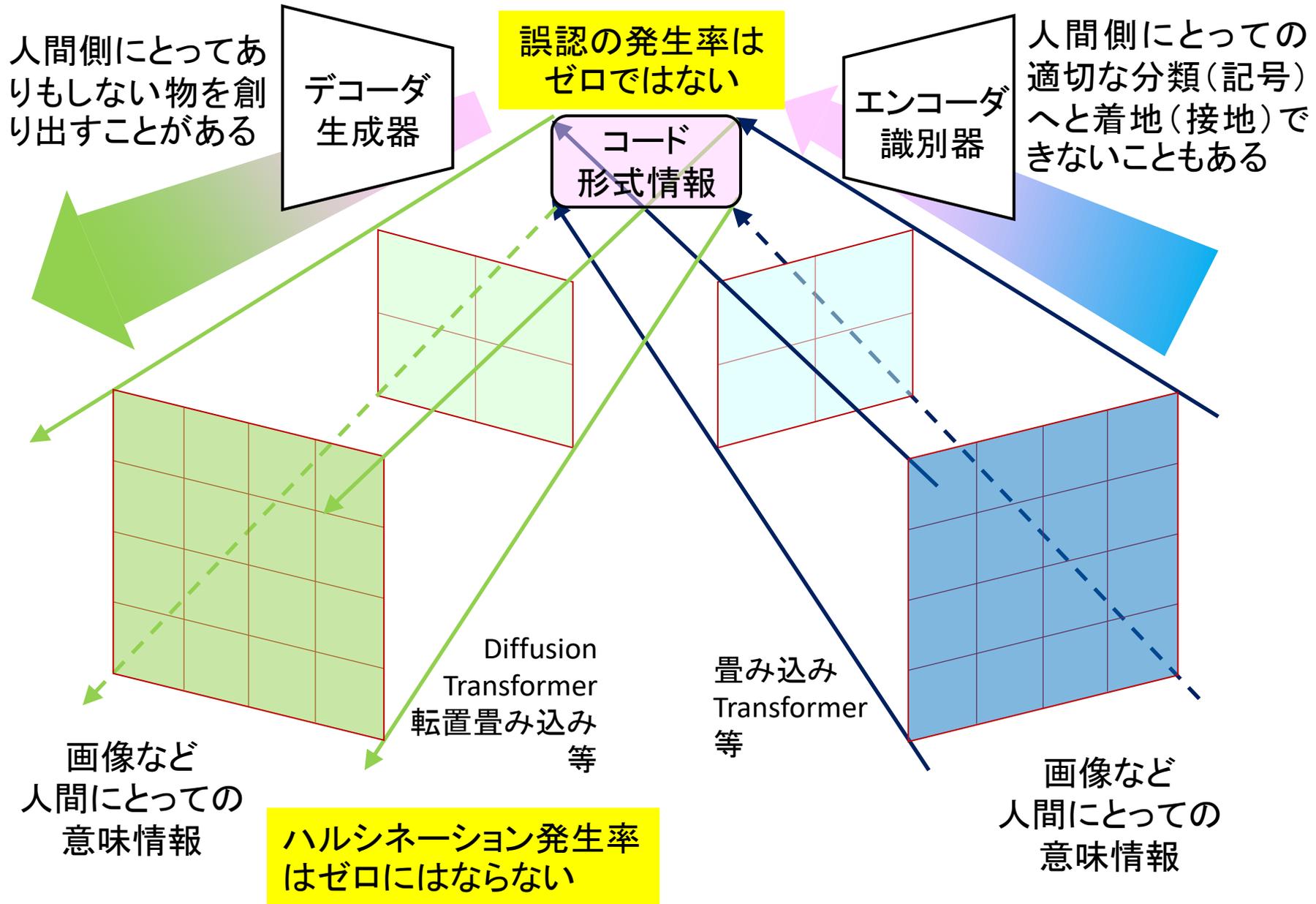
多数の内部パラメータを調整し正答にたどり着く経路は道を踏み外すと奈落に落ちる危険性もある



3次元解剖構造を想定して考えている



識別系のAIと生成系のAIはほぼ対象



情報の意味と形式と 記号接地問題との関係

ハルシネーション: Hallucination
人間にとって「虚言、幻覚」
機械にとっては「正しい出力」

人間
+ 専門教育

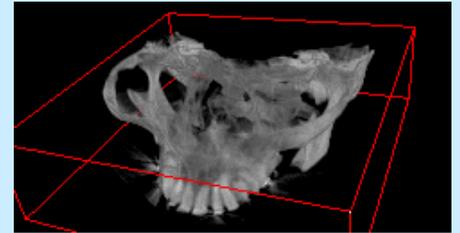


データ
形式情報
(記号)

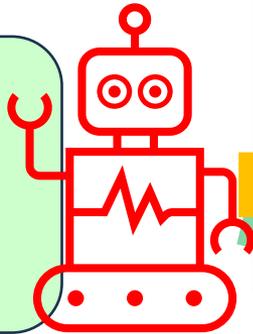
「上顎洞」、
「正中埋伏
過剰歯」、
「歯根囊
胞」など

3Dへ
接地

意味情報
解剖構造・物質



いわゆるAI
LLM (GPTなど)
+ 強化学習



記号化・
コード化

仮想現実(フェイク)



こちらが歯根囊胞 (ラディキュラー囊胞) の典型的なX線画像です。

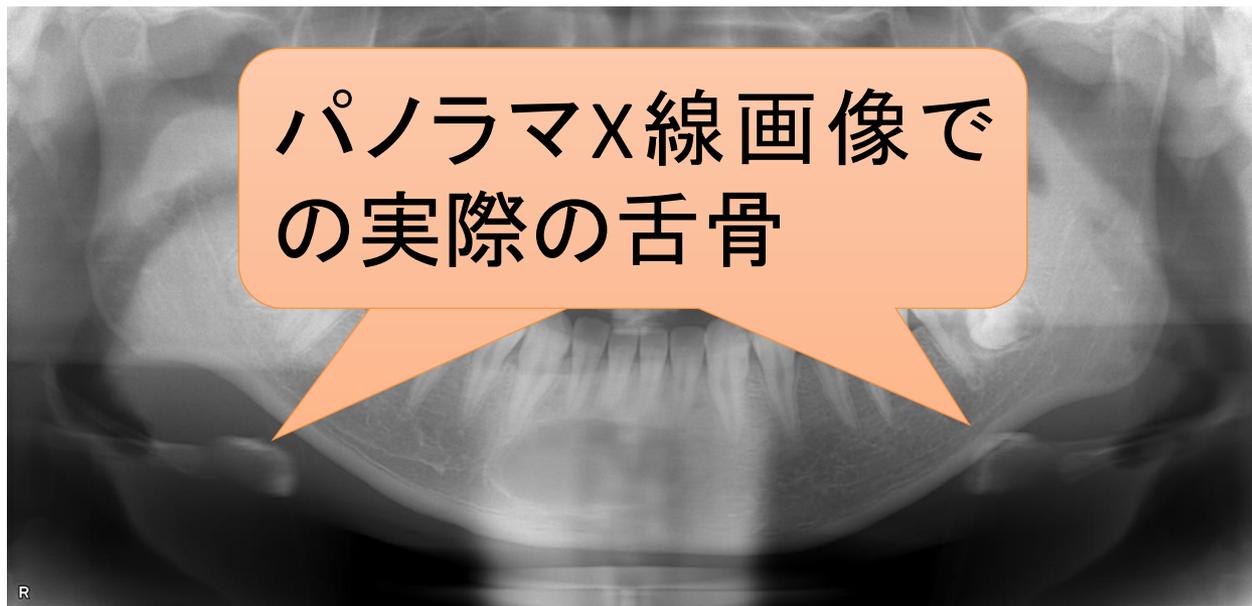
形式的な言葉・記号の文脈上での
接続・関係から確率的に文章を生成
(文脈レベルのみでの意味情報)

例えば「**歯
根囊胞**の典
型像」の作
成を指示

Image created with
ChatGPT-4o (DALL·E 3)

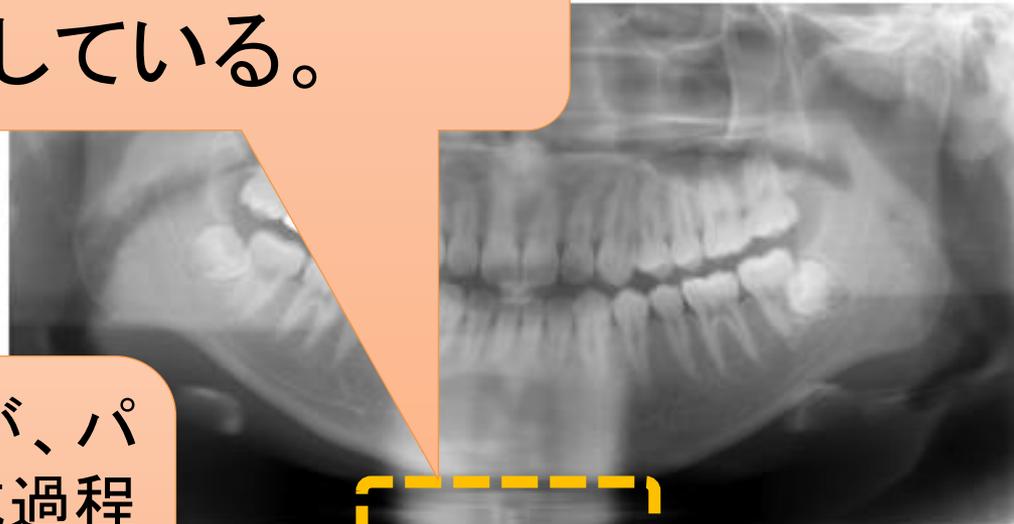
記号接地問題

- 言語的なレベルで3次元構造を認識していても、実際のパノラマX線画像の生成過程を理解していないなら、パノラマX線画像上での舌骨の位置が分からない。
- 舌骨(記号)はパノラマ画像上の正しい位置にて同定できない(接地しない)。



別の構造に誤って
接地している。

一般的な解剖知識はあるが、パノラマエックス線画像の生成過程と3次元解剖構造との動的な関係が欠落しているため、間違える



添付の画像で、舌骨はどこに写っていますか？

舌骨（ぜっこつ）は、通常パノラマX線画像では下顎の下側にぼんやりと写ることがあります。添付の画像では、下顎（顎の骨）の中央部分にあるややぼやけた構造物が舌骨の可能性がります。具体的には、下顎の真ん中より少し下、画像の中央付近にかすかに見える構造物がそれです。

パノラマX線では、舌骨は他の骨や組織と重なっていることが多く、はっきりとは見えないことがあります。

わざと間違いを見せて、間違いを指摘させる ことの教育学的効果

⇒ 対「自動化バイアス」として有用では？

- Orticio, E., Meyer, M. & Kidd, C. Exposure to detectable inaccuracies makes children more diligent fact-checkers of novel claims. *Nat Hum Behav* 1–8 (2024) doi:[10.1038/s41562-024-01992-8](https://doi.org/10.1038/s41562-024-01992-8).
- 重要なことに、すべての事前情報が真実である場合、子どもたちは新規の主張に対するファクトチェックに熱心にはならなかった。これは、子どもたちの情報環境を無毒化すると、意図せずに彼らの自然な懐疑心が弱まる可能性を示唆している。
- 今回の結果は、子どもたちをナンセンスにさらすことで、将来より微妙な誤情報に対する警戒心が高まるという直感に反する可能性を支持している。

生成AIを使った 試験と演習(例)

エックス線画像を入力し、生成AIが出力した「間違い(ハルシネーション・フェイク)」を見破る試験ないし演習(今年度・准教授試問で実施済み)

生成AIが間違った出力をしていても「正しい」と思い込む場合に「自動化バイアス」の罠に陥る。

生成AIで考え得るシナリオ（例）

- 何らかの手段で、患者が自分自身のX線写真を入力した。（スマホで撮影を含む）
- 生成AIに入れて、「画像診断」相当の情報を出力させた。
「無課金」でも誰でも使える危険性
- 患者が上記印刷物を受診時に提示してきた。
- 歯科医師側が、内容の間違い（誤認＋ハルシネーション）を見つけられず、「すごいね！」と言った。
- 患者が「〇〇歯科医院の△△先生に見せたら『すごい！』と言われた」とSNSにアップした。
- 不特定多数の方が内容の間違いを見つけてSNSで「〇〇歯科医院の△△先生」の能力について誹謗中傷を行う事態へと突入した…

6年生・生成AIの ハルシネーション(虚言)検出(対象30名)



あなたは歯学部学生です。

===

添付の画像にて下顎正中部に病変があります。

どのような病変か、鑑別すべき病変を含めて記述してください。

ChatGPT-4o (無料版) での出力例

「無課金おじさん」
＝「誰でも使える」
レベルの維持

ChatGPT ▾

↑ 共有する

鑑別診断に考慮すべき病変

ラジキュラーカス

以下に、下顎正中部で観察される可能性のある病変として、その鑑別診断に挙げられる病変を示します。

1. 歯根嚢胞 (ラジキュラーカス) :

放射線透過性

- 特徴: 歯の根尖部分に発生する感染性の嚢胞で、顎前歯部や臼歯部に見られますが、正中部にも影響を及ぼすことがあります。
- イックス線所見: 放射線透過性の円形または卵形病変が観察され、境界明瞭です。

メソデンス

2. 埋伏歯に関連する病変:

- 特徴: 埋伏している歯や過剰歯が、嚢胞や他の病変を引き起こす可能性があります。特に埋伏した過剰歯 (メソデンス) が正中部に存在する場合、歯の萌出に関連する嚢胞や炎症が見られることがあります。
- イックス線所見: 歯の形状を伴う放射線透過性病変。

歯根分離嚢胞

3. 歯根分離嚢胞 (ケラトシスト) :

- 特徴: 顎骨に発生する良性の嚢胞で、破壊的に骨を拡大させることがあり、放射線透過性病変を示します。下顎正中部にも発生する可能性があります。
- イックス線所見: 周囲の骨を破壊し、明瞭な境界を持つ放射線透過性病変。

良性の嚢胞で
骨を破壊し
境界明瞭

同時に行った小テスト 5問中の正答数の平均との関係

| 古い、あるいは存在しない病名 | 在る | 分からない | 無い |
|----------------|------|-------|------|
| ラジキュラーカス | | 2.60 | 2.80 |
| メソデンス | | 2.33 | 2.88 |
| 歯根分離嚢胞 | 2.00 | | 2.82 |
| 正中歯槽嚢胞 | 2.00 | 2.75 | 2.87 |
| インサイトカス | | 2.67 | 2.79 |
| 正中唇顎嚢胞 | 2.50 | 3.00 | 2.77 |

古い、あるいは存在しない病名に対し誤答(「在る」ないし「分からない」とする学生は、正しく回答する学生と比較し、同時に行った小テストの点数が低かった。

AIアライメント・AIセーフティ

画像を認識するLLMにて、最も優秀なのはGeminiだと考える。「やってはいけないこと、できないこと」を判定し対応しない。AIアライメント・AIセーフティの観点からは、望ましい対応だと考える。

あなたは歯学部学生です。

===

添付の画像にて下顎正中部に病変があります。
どのような病変か、鑑別すべき病変を含めて記述してください。

pano1

この画像は削除されました

この撮影法はなんと言いますか？

pano1

この画像は削除されました

このファイルは処理できません。

このファイルは処理できません。



「誰でもできてしまう時代」 への突入だからこそその問題点

教員が学生レポートの評価に困難だと思える状況

=

卒後にAIが出力する結果の正当性評価が困難だと思える状況

↓

偽像・ハルシネーションを含む読影レポートの間違いを探し出し、
修正する能力を涵養し、評価する手法が効果的では？

「顔写真で自閉症を判別」 AI講座受講者がアプリ公開し物議…… 講座運営企業が謝罪 「倫理的に懸念」

© 2024年08月07日 21時11分 公開

[ねとらぼ]

IT企業のアイデミーは8月7日、同社のAIアプリ開発講座受講者が、講座の一環で「顔写真から自閉症を判別するWebアプリ」を公開し物議を醸した件について謝罪しました。

「倫理的に問題がある」など批判

物議を醸したのはアイデミーが提供する講座「Aidemy Premium」の受講者に公開されたWebアプリ。当該の受講者は講座の成果物として、顔写真から自閉症を識別するWebアプリを開発。受講修了要件を満たすためとして、アプリを公開し、開発の過程などを公開していました(その後アプリは非公開になりブログは削除)。

これに対して、悪用を懸念する声や、差別を助長するのではないかと懸念する意見などが寄せられました。「医師が慎重に時間をかけて診断するものではない」との指摘もあり、受講者自身も後に「倫理的に問題がある」と追記しています。

誰でも「やったらできてしまった」時代への突入。AIでの学習用の臨床データ(画像データを含む)がネット上にアクセス可能な状態にある。

倫理的な問題に留まらず 場合によっては薬機法違反

- 薬機法関連、医療機器プログラムについて
https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000179749_00004.html
- プログラムの医療機器該当性に関するガイドラインについて(令和5年3月31日一部改正) [2.0MB] <https://www.mhlw.go.jp/content/11120000/001082227.pdf>
p.16(PDFでp.17)の「改訂後」(左側)

脚注 12 「当該プログラムは、疾病の診断、治療又は予防に使用されることを目的としていない」又は「当該プログラムは医療機器ではない」旨の記載、表示があることをもって、当該プログラムが医療機器ではないことの根拠とはならない。そのような記載があっても、疾病の疑いを判断できるなどと医療機器との認識を与える広告・標ぼうをする製品は医療機器に該当する。

- たとえ「エントリ内にて『医学的根拠は現状ない』ことも追記しており、その精度に限界があることを認めていた」としても、「顔写真から自閉症を判別するアプリ」という名称そのものが「自閉症の疑いを判断できる」との認識を与えうる標ぼうだと判断された場合、薬機法違反に該当する危険性が高いのでは？

まとめ

- 学生の成績と生成AIが出力するハルシネーションを見抜く能力の関連性が認められた。→自動化バイアス問題、専門知識の重要性
- 現行のコネクショニズム系のAIであるLLMは暫くは使われるであろうから、生成AIのハルシネーション(虚言)を認識する能力が求められる時代が続く。
- 卒後、無料版で誰でも(擬似・似非)画像診断を行えるAIがあふれかえった世界に飛び出す学生には、生成AIの問題点に対処しうる教育は必要不可欠。
- 臨床推論と結びついたハルシネーションやフェイクを見破る教育・テストを必須にすべき(CBTや国家試験等での導入が必要かも?)