

新潟大・歯学部・学部学生用  
3年次・AI関連資料  
歯のエックス線所見関連

歯科放射線科

西山秀昌

2025/01/23 (1/24, 1/25 改訂・追加)

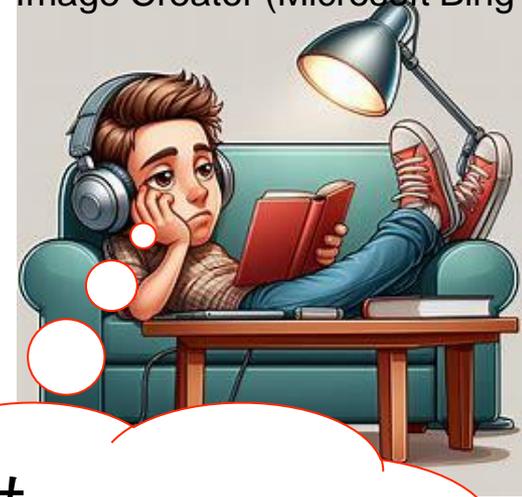
2020/02/11・初版

# 歯のエックス線所見では様々な知識と推論能力が必要不可欠

- 3次元的な解剖構造
- エックス線の性質
  - 物質の透過性について
  - 重積効果について
  - 接線効果について
  - 錯視について



# 画像診断における 深層学習・機械学習 いわゆる人工知能(AI)



勉強しなくても  
AIが自動的に診断してくれる  
だろうから、単位貰える程度  
にサボっても大丈夫???

- 人工知能(AI)て何？
- 機械が診断した結果を鵜呑みにしても大丈夫？
- 万一機械が誤診した場合、責任の所在は？

# AI

論理的でないものや記号化困難な対象がある。対象範囲の設定が困難な場合がある。

結果に至る過程について人間が理解困難な場合や、過程内部で論理性が欠落する危険性がある。

## いわゆるSymbolism

過程優先、演繹的  
シンボル・記号が基本  
論理構造重視  
原点：論理プログラミング  
知識ベース、述語論理

第二次AIブームの  
主役

エキスパートシステム  
プロダクションシステム

## いわゆるConnectionism

結果優先、発見的  
ネットワーク結合が基本  
パターン・類似点・特徴量重視  
原点：パーセプトロン  
ニューラルネットワーク

第一次、第三次  
AIブームの  
主役

Transformer系  
CNN系  
他

深層学習  
機械学習

Hybrid Type ⇒ AGI, ASI, 強いAI\*?

※AGI(Artificial General Intelligence、汎用人工知能)、ASI(Artificial Super Intelligence、超人工知能≒「強いAI」)

# AI

論理的でないものや記号化困難な対象がある。対象範囲の設定が困難な場合がある。

結果に至る過程について人間が理解困難な場合や、過程内部で論理性が欠落する危険性がある。

## いわゆるSymbolism

過程優先、演繹的

シンボル・記号が基本

論理構造重視

原点：論理プログラミング

知識ベース、述語論理

第二次AIブーム  
の主演

## いわゆるConnectionism

結果優先、発見的

ネットワーク結合が基本

パターン・類似点・特徴量重視

原点：パーセプトロン

ニューラルネットワーク

第一次、第三次  
AIブームの主演

RAG (Retrieval-Augmented Generation) 等の知識データベース導入等

Transformer系  
CNN系  
他

深層  
学習

機械  
学習

Hybrid Type ⇒ AGI, ASI, 強いAI ※?

※定義は曖昧だが、おおよそ汎用的な知識を持ち(AGI)、人間のような自意識や知能を備え高度な判断が可能なAI(ASI, 強いAI)

# LLM出現後のAGI概念 強いAIとの違いについて

- Mumuni, A. & Mumuni, F. Large language models for artificial general intelligence (AGI): A survey of foundational principles and approaches. Preprint at <https://doi.org/10.48550/arXiv.2501.03151> (2025).
- LLM が人間レベルの汎用知能を獲得するには、身体化、記号接地(シンボルグラウンディング)、因果関係、記憶といった多くの基礎的問題に対処する必要がある。
- AGIは超全知全能の機械を意味するものではない。そのような仮想的能力のレベルは人工超知能と呼ばれる。

# 該当論文での定義

## AGI, ASI, Strong AI

- AGIは強いAIとは根本的に異なる
- AGI(汎用人工知能)
  - 幅広い認知能力を持ち、真に重要な問題を解決できるインテリジェント システムの開発に重点を置いている。
- 強い AI
  - = ASI(超人工知能; Artificial Super-Intelligence)
  - 機能レベルで人間の認知能力を模倣するだけでなく、志向性(願望、希望、信念、内的動機など)、道徳、感情、意識と感覚という意味での自己認識など、本質的な精神状態や主観的経験などの実際の人間の認知特性によって特徴付けられる非常に強力な知能を目指すもの。

## 2.3 LLM による AGI の基本原則の概要 (該当箇所の翻訳)

- 最先端の大規模言語モデルは信じられないほど強力だが、汎用知能を実現しうる能力を制限してしまう項目がまだ残っている[87、88、89、90]。
- 一般的に、モデルの文脈理解は表面的なものであることが多く、その解決策は多くの場合、人間の知識と外見的に似ているだけである[91、92]。
- 問題となるのは、LLM を含む AI システムが、膨大なデータから世界の一般的な特性を学習することで人間の知識と認知能力を模倣しようとする単なるデジタル構造にすぎないという事実にある。
- 模倣された知識は一般に観察されるパターンに限定され、対象の動作原因となる根本的な原理を捉えてはいない。

# 脳の学習原理は人工知能のものとは 根本的に異なっている

- Song, Y. *et al.* Inferring neural activity before plasticity as a foundation for learning beyond backpropagation. *Nat Neurosci* **27**, 348–358 (2024).  
<https://www.nature.com/articles/s41593-023-01514-1>
- [https://gigazine.net/news/20241202-brain-and-neural-activity/?utm\\_source=x&utm\\_medium=sns&utm\\_campaign=x\\_post&utm\\_content=20241202-brain-and-neural-activity#google\\_vignette](https://gigazine.net/news/20241202-brain-and-neural-activity/?utm_source=x&utm_medium=sns&utm_campaign=x_post&utm_content=20241202-brain-and-neural-activity#google_vignette)

# Google 日本語訳

<https://www.nature.com/articles/s41593-023-01514-1>

記事 | [オープンアクセス](#) | 公開日:2024年1月3日

## バックプロパゲーションを超えた学習の基礎として可塑性前の神経活動を推測する

宋宇航 , ベレン・ミリッジ, トマン・サルヴァトーリ, トーマス・ルカシェヴィッチ , 徐正華  & ラファエル・ボガツ 

[ネイチャーニューロサイエンス](#) 27, 348–358 (2024) | [この記事を引用](#)

91,000 アクセス | 11 引用 | 583 アルトメトリック | [xTRIX](#)

### 抽象的な

人間と機械の両方にとって、学習の本質は、情報処理パイプラインのどのコンポーネントが出力のエラーの原因であるかを正確に特定することであり、これは「クレジット割り当て」と呼ばれる課題です。クレジット割り当ては、現代の機械学習の基礎でもあるバックプロパゲーションによって最もよく解決されると長い間考えられてきました。ここでは、「プロスペクティブ構成」と呼ばれる、クレジット割り当てに関する根本的に異なる原則を示します。プロスペクティブ構成では、ネットワークは最初に学習の結果として生じるはずの神経活動のパターンを推測し、次にシナプスの重みを変更して神経活動の変化を統合します。バックプロパゲーションとは対照的に、この独特のメカニズムは、(1) 確立された皮質回路のモデルファミリーにおける学習の基礎であり、(2) 生物が直面する多くの状況でより効率的で効果的な学習を可能にし、(3) さまざまな人間とラットの学習実験で観察された神経活動と行動の驚くべきパターンを再現することを実証します。

# 論理回路で構成されるコンピュータ

いずれにしても、現行のAIは、論理プログラミングの範疇にあって、そこから逸脱はありえない。

## 論理プログラミング

エキスパートシステム  
ファジィコンピュータ  
遺伝的アルゴリズム  
ボルツマンマシン(アニーリング)

数学、統計ソフト  
画像処理ソフト  
オフィスソフト  
等々

## ニューラルネット

### 深層学習システム

画像認識:  
CNN系  
Transformer系

### 生成AI(生成系AI)

言語系: LLM **GPT**

画像系: **DALL-E**  
**Stable Diffusion**

音声合成: **VALL-E**

「歯のエックス線所見」に絡めて具体的にやってみた

LLM(Large Language Model):形式情報(記号列)としての言葉を計算対象として扱うシステム。  
GPT(Generative Pre-trained Transformer)

# 深層学習・機械学習

## 画像認識系での人工知能(AI)

- 人工知能(AI)の内、画像認識で用いられているシステムで中心となるものは、
  1. 論理的な計算構造(モデル)を「人間が規定し」
  2. 「人間が選択した」データからの予測値と真の値との差を計算する「微分可能な」誤差関数を設定し
  3. 誤差関数の値を最小にするパラメータを導き出すシステム

「TensorFlowとKerasで動かしながら学ぶ ディープラーニングの仕組み ~畳み込みニューラルネットワーク徹底解説~(マイナビ、中井悦司)」ISBN-13: 978-4839970277  
[ 西山・大学院演習での使用テキスト(教科書) ]

# 深層学習・機械学習

学習用データの選別によるバイアス

例えば、診断名  
あるいは、動物名  
(機械内部では数値データ)

入力  
データ

● (変換)モデル

● パラメータ

予測値

人間が規則性を見つけ、モデルを作る。

誤差が最小  
になるまで

● 誤差関数  
微分可能が  
前提条件

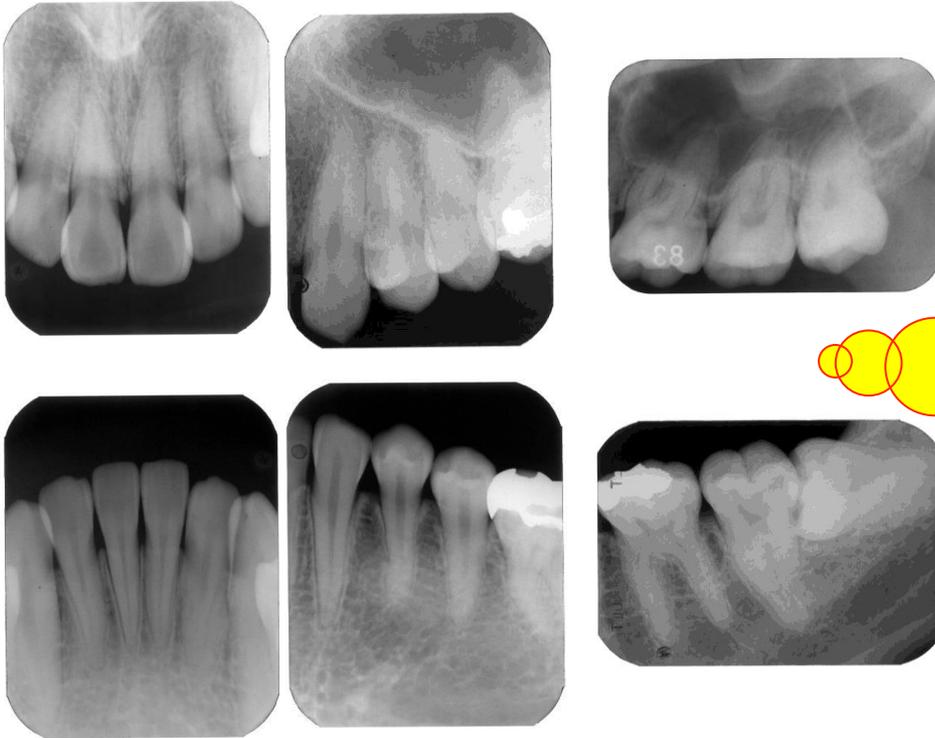
自動的にモデルを選択するシステムもあるが、「選択可能なモデル」は、あらかじめ人間が決定する。

画像認識の場合、「多層(深層)の畳み込み(convolution)」あるいは「画像を構成する要素を複数ベクトルの束(Matrix)として変換する(Transformer)\*」という作業が多く使われている。

※[https://speakerdeck.com/yushiku/20220608\\_ssii\\_transformer](https://speakerdeck.com/yushiku/20220608_ssii_transformer)

実測値  
真の値

# 6ブロックの口内法エックス線画像 の部位を判定させてみる



部位判定は歯学部  
5年生では必須。  
できなければ落第  
レベルの課題。

西山秀昌, 織田隆昭, 高村真貴, 新國 農, 勝良剛詞, 池 真樹子, 坂井幸子, 小林太一, 曾我麻里恵,  
Jorge Eduardo Saez Chandia, 小椋一朗, 林 孝文. 「デンタル画像のマウント位置における機械学習法の比較  
と判断根拠の考察.」 第2回日本歯科人工知能 (AI) 研究会, 2019年12月15日, 東京都渋谷区 (渋谷ソラスタ  
コンファレンスA4会議室) 倫理承認番号 2018-0003

からの抜粋

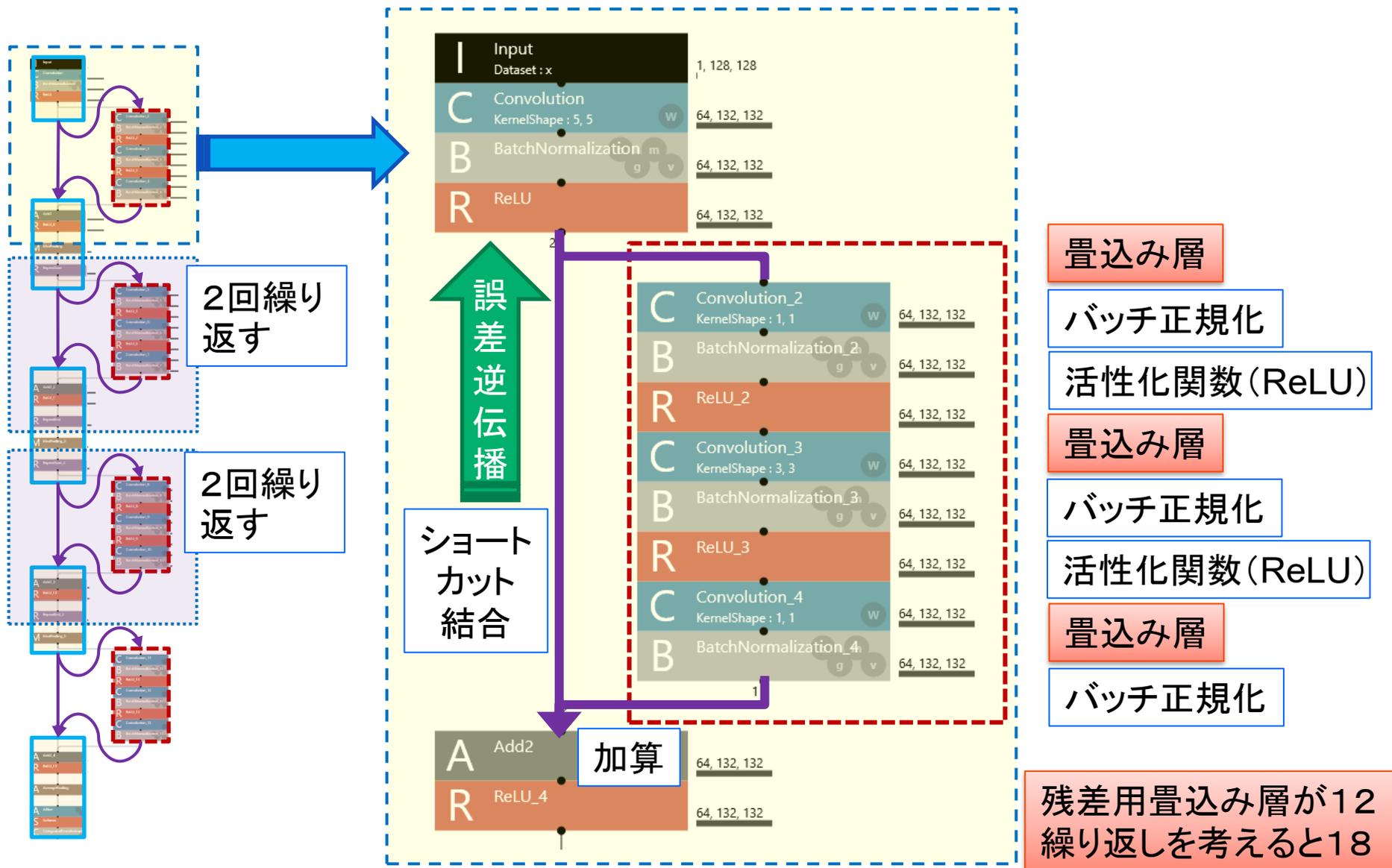
<https://www5.dent.niigata-u.ac.jp/~nisiyama/grad/20191215-2ndDentalAI-Conf-HN.pdf>

# 使用機器、ソフトウェア

- ハードウェア
  - CPU: Intel Core i5-7400 (3GHz)、メモリ: 24GB
  - GPU: NVIDIA GeForce GTX 1060 6GB
- ソフトウェア (MNISTデータを扱えるレベル)  
Neural Network Console (ソニー社製、version 1.4.6983.42039 )
  - residual\_learning (ResNetの亜系)

# residual\_learning (以後、mResNet)

入力層ピクセルサイズと出力層のカテゴリ数のみ変更

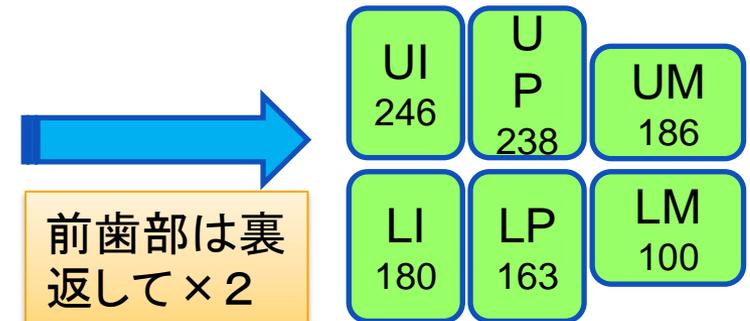
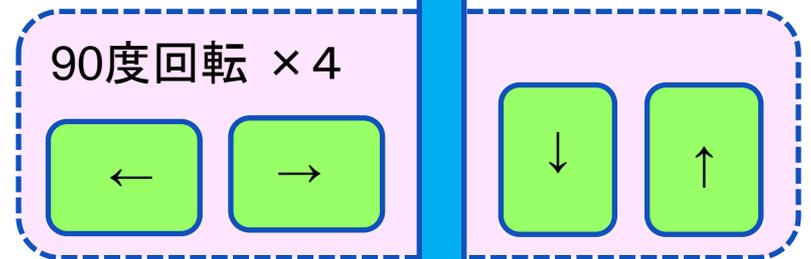
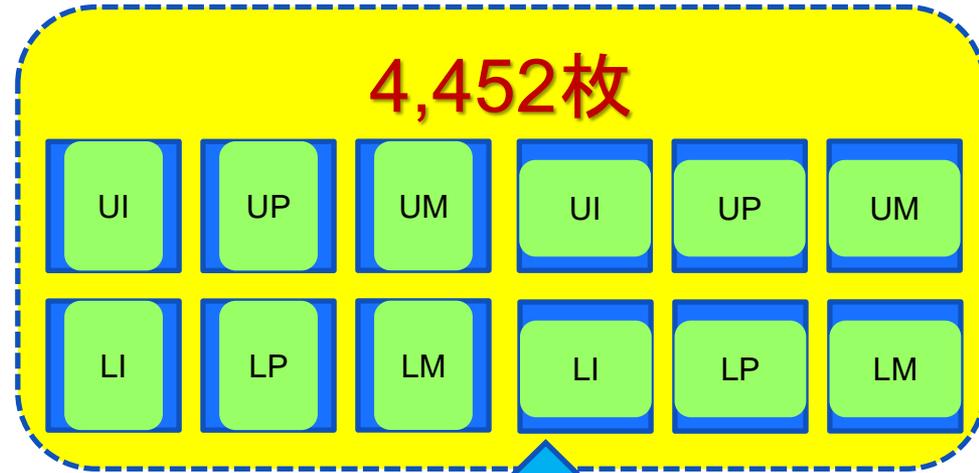
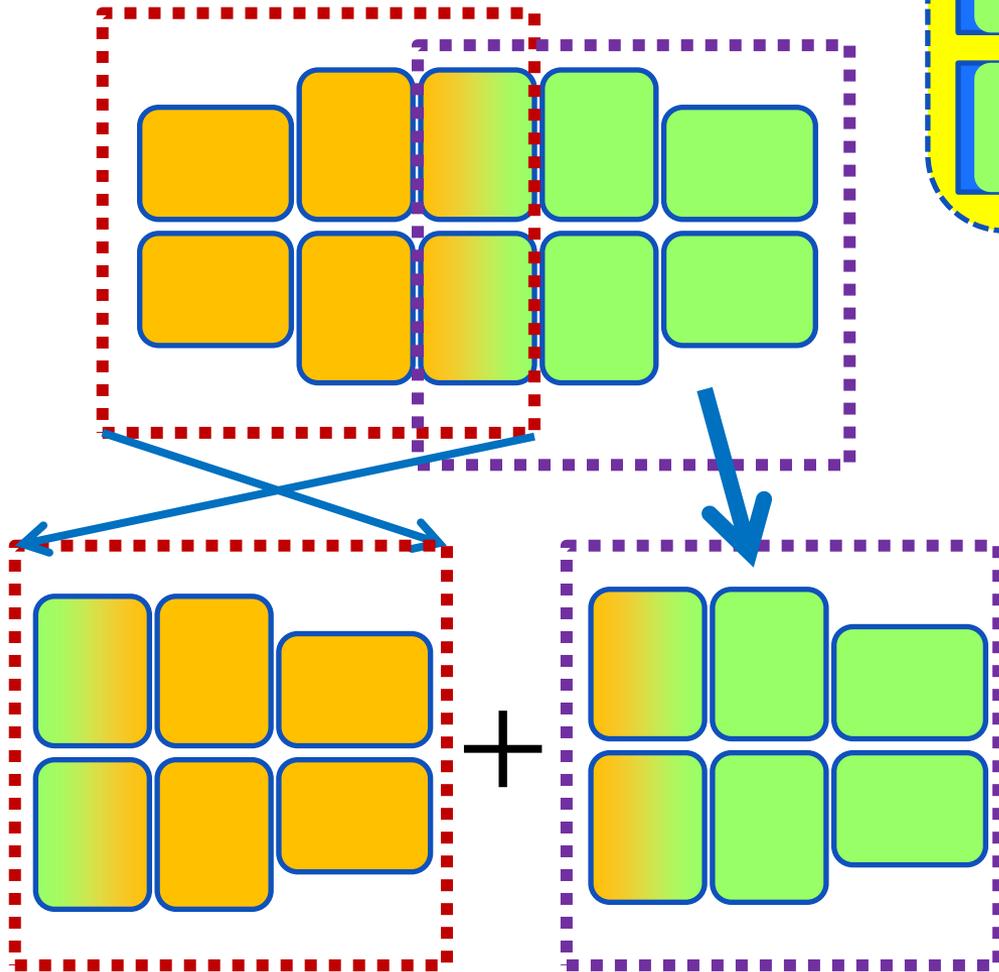


# 実験1 対象 学習データ

- 10枚法の口内法エックス線画像**150組**、計**1,417枚**の内、明らかな失敗やフィルム内にて智歯を除く1歯以上の欠損やインプラントを除いた**900枚**。
- 画像の右側を反転させることで左側6ブロック計1,113枚としテスト用として**72枚**、学習用と検証用として**1,041枚**に別けた。
- 90度単位で回転させ、合計**4,452枚**(テスト用**288枚**、学習・検証用**4,164枚**)を作成。

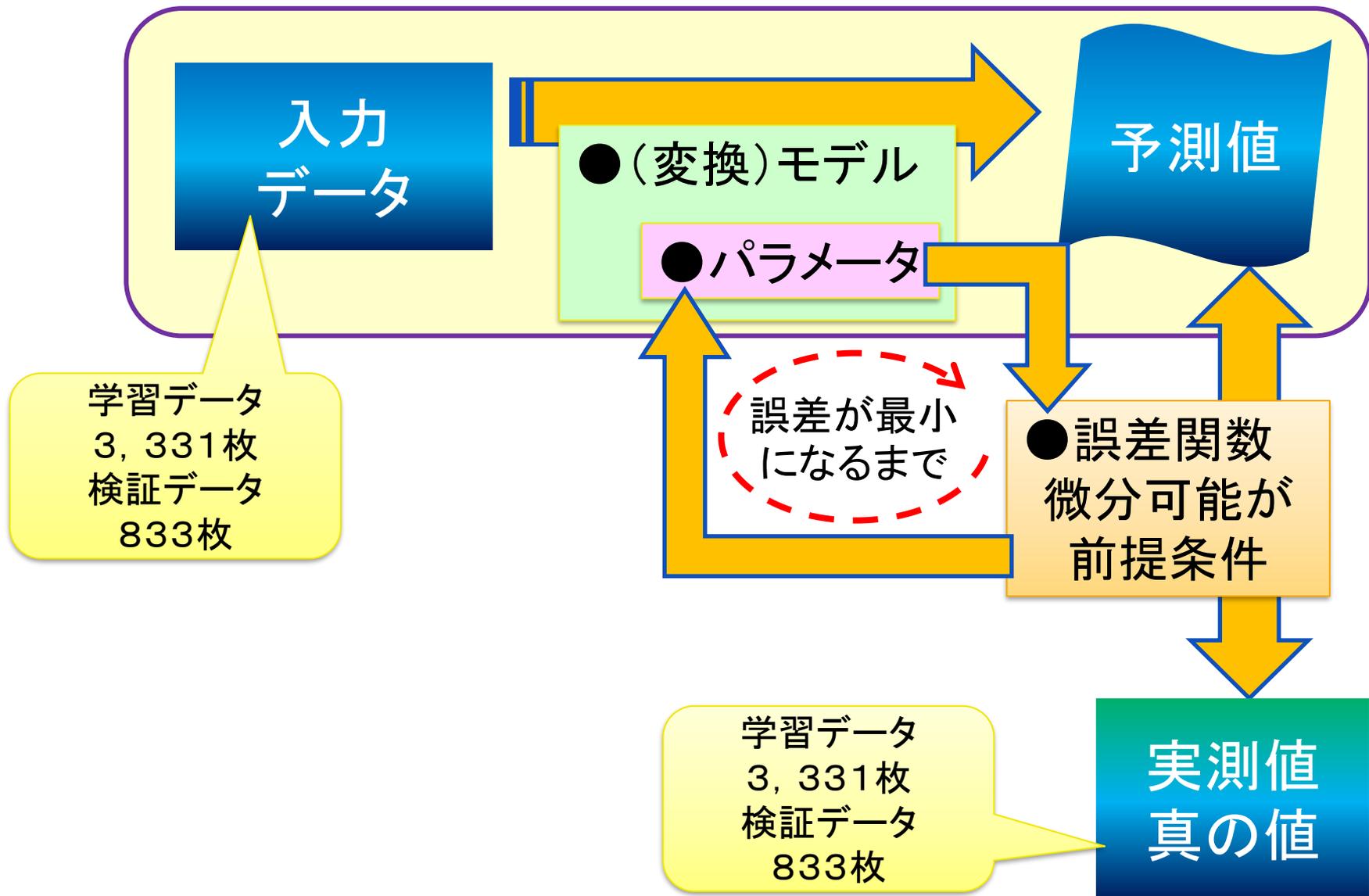
# 対象画像データの生成

150組、計1,417枚から選択した900枚(欠損、明らかな失敗、インプラント体等除く)

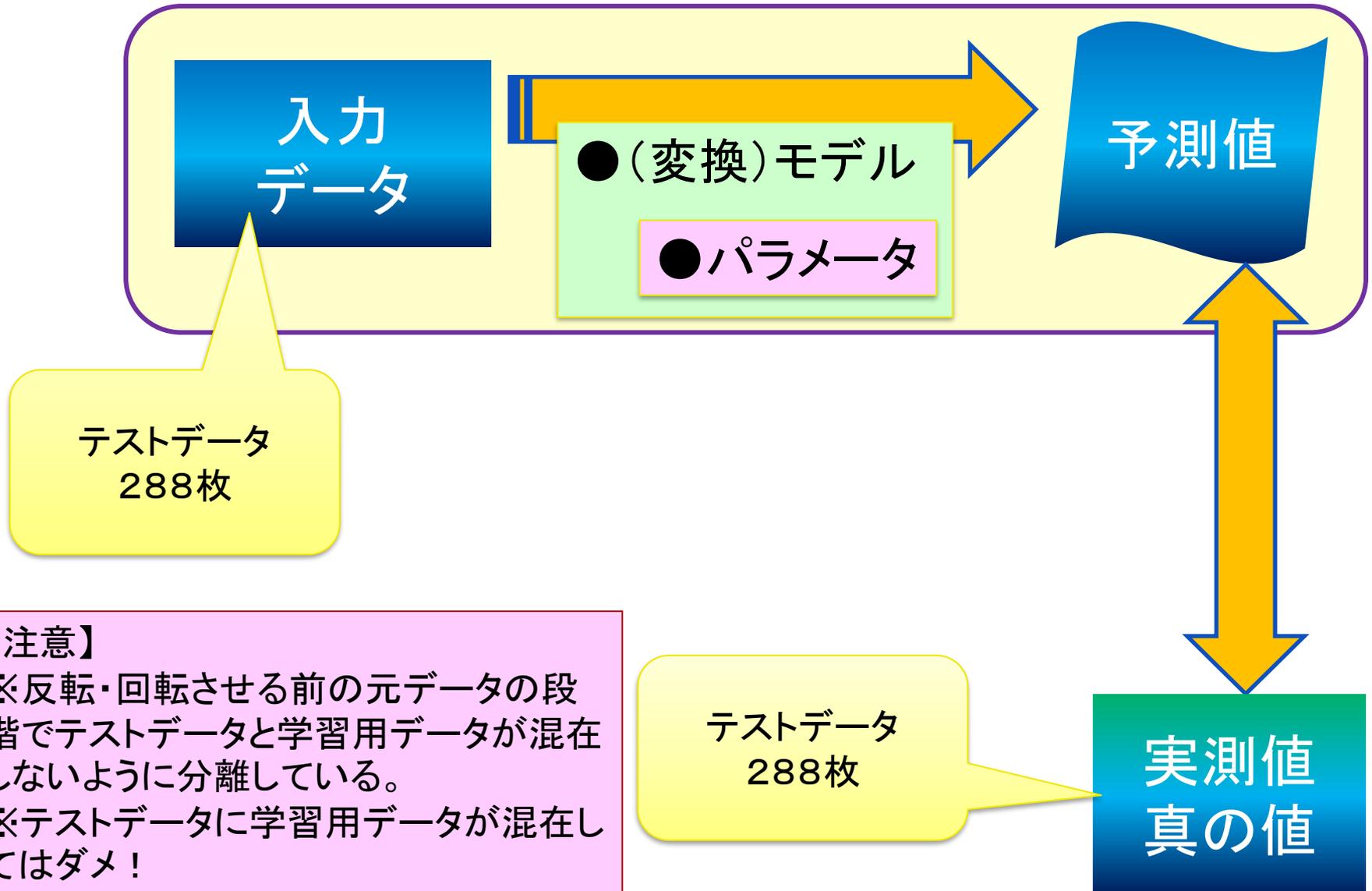


前歯部は裏返して×2

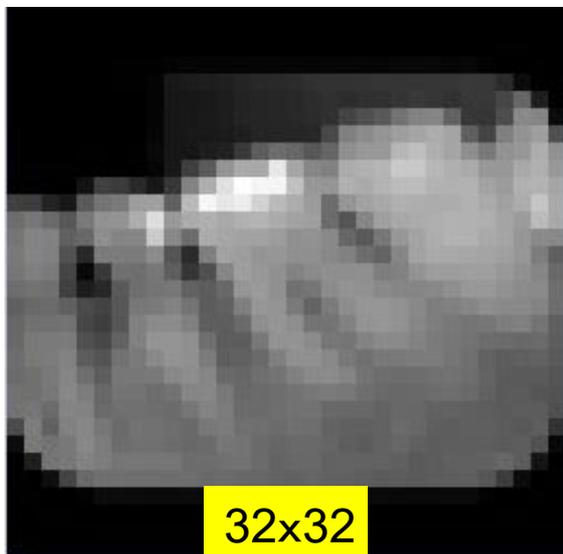
# 深層学習・機械学習の訓練中...



# 深層学習・機械学習のテスト



# 画像データおよび略号



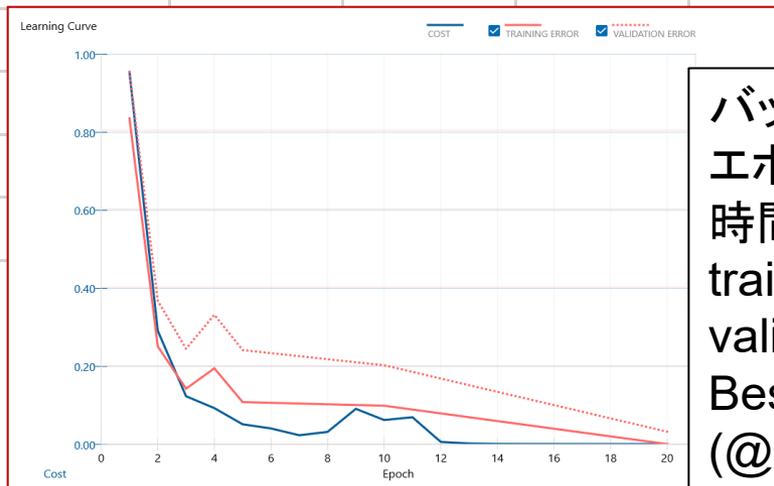
記号	内容
UI	上顎前歯部
UP	上顎犬歯・小臼歯部
UM	上顎大臼歯部
LI	下顎前歯部
LP	下顎犬歯・小臼歯部
LM	下顎大臼歯部

機械内部では0から5(ないし1から6)の数字(コード)

# mResNet 128x128

		予測した部位						
		UI	UP	UM	LI	LP	LM	Recall
真 の 部 位	UI	48	0	0	0	0	0	1.000
	UP	0	48	0	0	0	0	1.000
	UM	0	0	48	0	0	0	1.000
	LI	0	0	0	48	0	0	1.000
	LP	0	0	0	0	48	0	1.000
	LM	0	0	0	0	0	48	1.000
Precision		1.000	1.000	1.000	1.000	1.000	1.000	
F-Measures		1.000	1.000	1.000	1.000	1.000	1.000	

Accuracy	1.000
Avg.Precision	1.000
Avg.Recall	1.000
Avg.F-Measures	1.000



バッチ数: 20  
 エポック数: 20  
 時間 00:07:49  
 train\_error=0.000277  
 valid\_error=0.0161  
 Best valid\_error= 0.0161  
 (@ epoch 20)

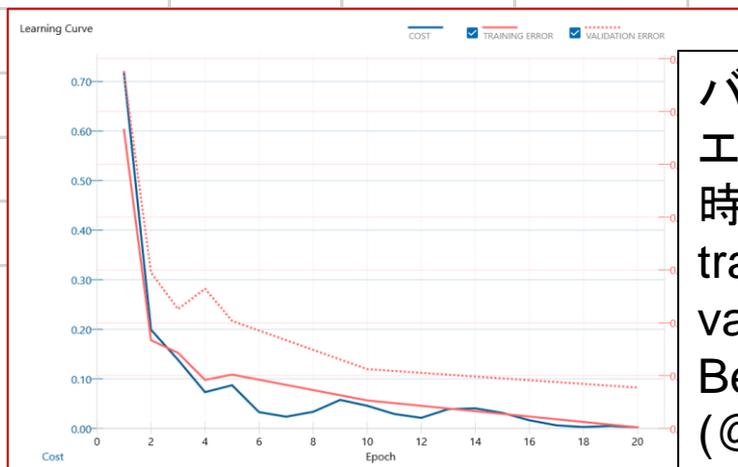
僅か8分の学習  
 で、100%

# mResNet 32x32

		予測した部位						Recall
		UI	UP	UM	LI	LP	LM	
真の部位	UI	48	0	0	0	0	0	1.000
	UP	0	48	0	0	0	0	1.000
	UM	0	0	47	0	0	1	0.979
	LI	0	0	0	48	0	0	1.000
	LP	0	0	0	0	48	0	1.000
	LM	0	0	0	0	0	48	1.000
Precision		1.000	1.000	1.000	1.000	1.000	0.980	
F-Measures		1.000	1.000	0.989	1.000	1.000	0.990	

Accuracy	0.997
Avg.Precision	0.997
Avg.Recall	0.997
Avg.F-Measures	0.997

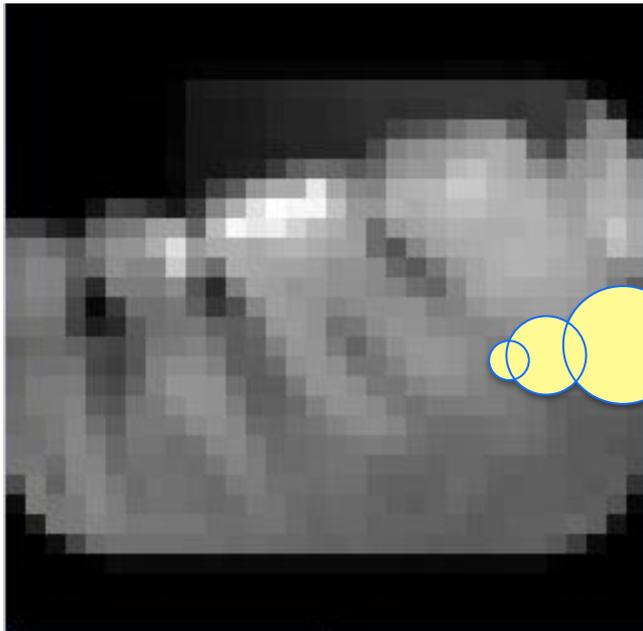
僅か1分の学習  
で、99.7%



バッチ数: 20  
エポック数: 20  
時間 00:00:52  
train\_error=0.000973  
valid\_error=0.0386  
Best valid\_error= 0.0386  
(@ epoch 20)

# 32ピクセル画像の人間による判定

- テストデータの6部位から3枚ずつランダムに選択した32ピクセル画像18枚。人間(9名の共同演者)とAIにて評価し、比較した。

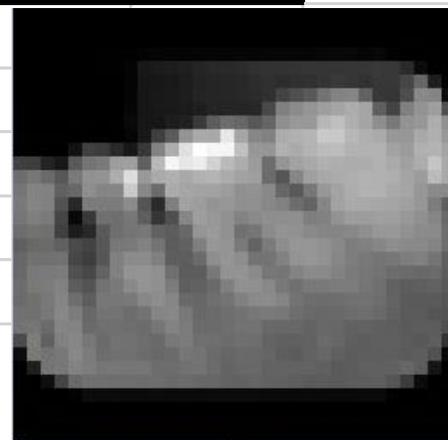


これほどノイズが強く、粗い画像は診断には使えないし、部位判定も難しいけれど、それでもできる限り判定してみた。

# 32x32 人間での評価

		予測した部位						Recall
		UI	UP	UM	LI	LP	LM	
真 の 部 位	UI	27	0	0	0	0	0	1.000
	UP	0	26	0	0	1	0	0.963
	UM	0	1	25	0	0	1	0.926
	LI	0	0	0	27	0	0	1.000
	LP	0	2	0	0	25	0	0.926
	LM	0	0	0	0	0	27	1.000
Precision		1.000	0.897	1.000	1.000	0.962	0.964	
F-Measures		1.000	0.929	0.981	1.000	0.943	0.982	
Accuracy		0.969						
Avg.Precision		0.970						
Avg.Recall		0.969						
Avg.F-Measures		0.972						

約97%



このレベルの非常に劣化した画像になると、機械側の認識が高い。なぜ、機械側の認識は高いのだろうか？

# 実験2 対象 評価用データ

- **【濃淡シェーマデータ】**: 学生提示用のシェーマにて、できる限り濃淡を実際の画像に合わせるようにした6部位6枚。
- **【線画シェーマデータ】**: 2冊の教科書(※)から線画のシェーマを抜き出した6部位15枚。
- シェーマデータは、各部位ランダムにコピーを生成し、1部位毎に8枚となるようにし、ランダムに回転させた。

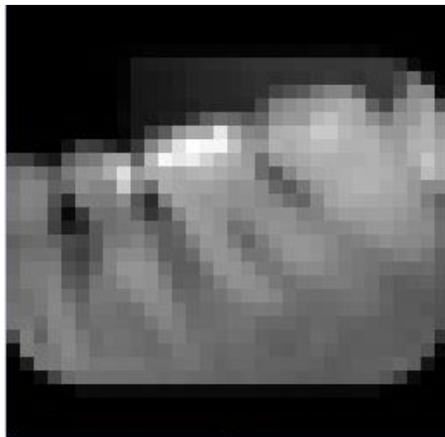
※歯科衛生士テキスト、わかりやすい歯科放射線学(第2版)、学建書院

※歯科放射線学、第6版、医歯薬出版

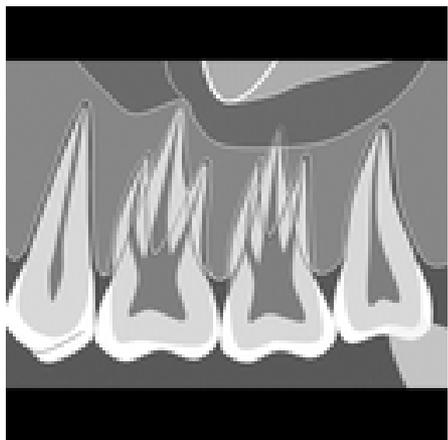
# 画像データとシェーマおよび略号



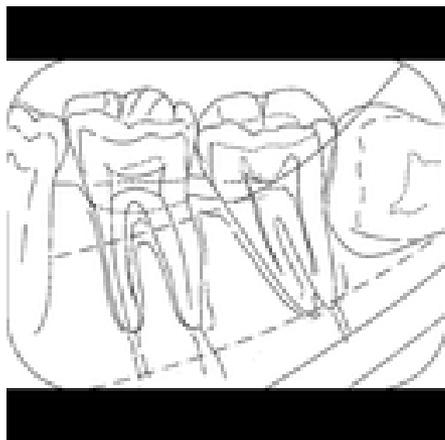
128x128



32x32



濃淡シェーマ



線画シェーマ

記号	内容
UI	上顎前歯部
UP	上顎犬歯・小臼歯部
UM	上顎大臼歯部
LI	下顎前歯部
LP	下顎犬歯・小臼歯部
LM	下顎大臼歯部

# さて、何パーセントの 正答率になるだろうか？

- 10%以下
- 10-30%程度
- 40-60%程度
- 70-80%程度
- 90%以上

} 多くの学生の予測はこのあたり

128 x 128

予測した部位

		UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	0	5	0	0	0	0	0.000
	UP	0	5	0	0	0	0	1.000
	UM	0	5	0	0	0	0	0.000
	LI	0	5	0	0	0	0	0.000
	LP	0	5	0	0	0	0	0.000
	LM	0	5	0	0	0	0	0.000
Precision		0.000	0.167	0.000	0.000	0.000	0.000	
F-Measures		0.000	0.286	0.000	0.000	0.000	0.000	

Accuracy	0.167
Avg.Precision	0.028
Avg.Recall	0.167
Avg.F-Measures	0.048

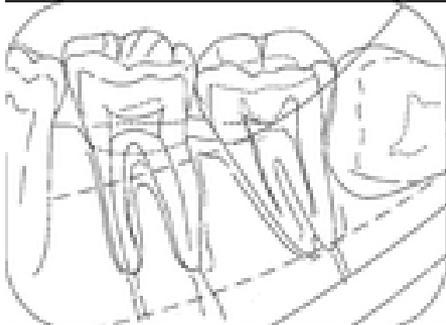
# mResNet

## Accuracy

### 17 - 20 % 程度

ボロボロ...

線画シェーマ



32 x 32

予測した部位

		UI	UP	UM	LI	LP	LM	Recall
真の部位	UI	5	0	0	0	0	0	1.000
	UP	5	0	0	0	0	0	0.000
	UM	5	0	0	0	0	0	0.000
	LI	4	0	0	1	0	0	0.200
	LP	1	0	0	4	0	0	0.000
	LM	4	0	0	1	0	0	0.000
Precision		0.208	0.000	0.000	0.167	0.000	0.000	
F-Measures		0.345	0.000	0.000	0.182	0.000	0.000	
Accuracy		0.200						
Avg.Precision		0.063						
Avg.Recall		0.200						
Avg.F-Measures		0.088						

128 x 128

予測した部位

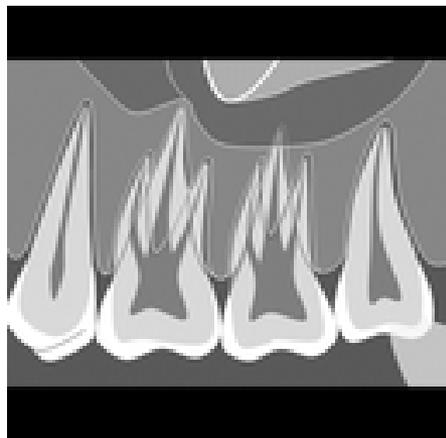
		UI	UP	UM	LI	LP	LM	Recall
真 の 部 位	UI	2	2	0	0	0	4	0.250
	UP	2	4	2	0	0	0	0.500
	UM	4	0	4	0	0	0	0.500
	LI	8	0	0	0	0	0	0.000
	LP	0	4	0	0	4	0	0.500
	LM	0	0	6	0	0	2	0.250
Precision		0.125	0.400	0.333	0.000	1.000	0.333	
F-Measures		0.167	0.444	0.400	0.000	0.667	0.286	

Accuracy	0.333
Avg.Precision	0.365
Avg.Recall	0.333
Avg.F-Measures	0.327

# mResNet Accuracy 33 % 程度

少しですが  
ボロボロ...

## 濃淡シェーマ



32 x 32

予測した部位

		UI	UP	UM	LI	LP	LM	Recall
真 の 部 位	UI	6	0	0	0	0	2	0.750
	UP	6	2	0	0	0	0	0.250
	UM	2	0	6	0	0	0	0.750
	LI	4	4	0	0	0	0	0.000
	LP	4	4	0	0	0	0	0.000
	LM	6	0	0	0	0	2	0.250
Precision		0.214	0.200	1.000	0.000	0.000	0.500	
F-Measures		0.333	0.222	0.857	0.000	0.000	0.333	
Accuracy		0.333						
Avg.Precision		0.319						
Avg.Recall		0.333						
Avg.F-Measures		0.291						

# 何故？

学習データのバイアス問題、すなわち  
外挿領域ないしカテゴリーの範囲外

- 学習データは、透過光画像でエックス線透過像  
⇒ 単純エックス線画像 ⇒ 口内法X線画像  
⇒ 二等分法(根尖部投影法) ⇒ 10枚法という  
階層の末端部にある母集団からのサンプル
- 線画シェーマや濃淡シェーマは反射光画像で機  
械からすれば外挿領域ないし別カテゴリー
- 常識という知識を持った人間からは内挿範囲で  
同一カテゴリーとして受容可能な範疇であり、「  
あたりまえだろ？何故間違える？」となる

# ヒートマップ(どの部分を特徴として捉えているか) 確認画像・例

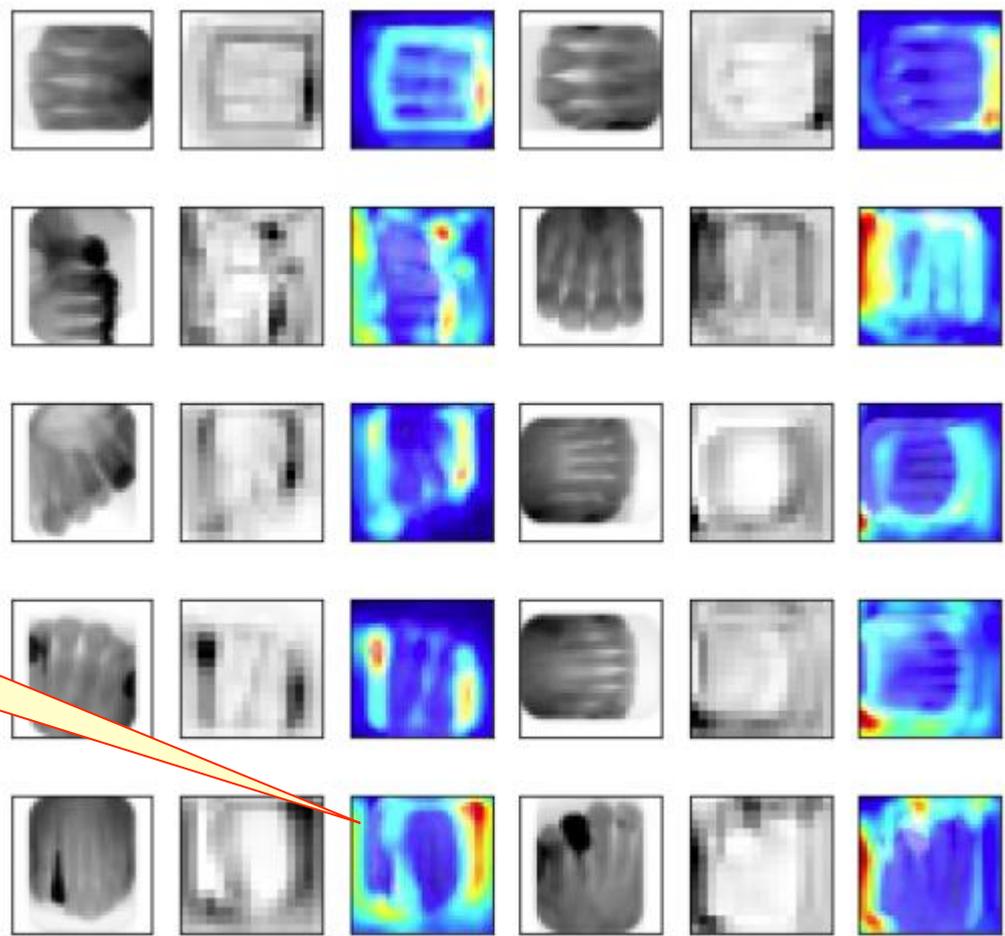
Layer (type)	Output Shape
reshape (Reshape)	(None, 32, 32, 1)
conv_filter1 (Conv2D)	(None, 16, 16, 16)
max_pooling1 (MaxPooling2D)	(None, 8, 8, 16)
conv_filter2 (Conv2D)	(None, 30, 8, 8)
max_pooling2 (MaxPooling2D)	(None, 4, 4, 30)
flatten (Flatten)	(None, 1920)
hidden	(None, 100)
hidden2	(None, 100)
softmax	(None, 10)

Total params: 294,248  
Trainable params: 294,248  
Non-trainable params: 0

畳込み層

畳込み層

重積像を含むエクス線  
透過像にてヒートマップ  
の意味するところは？

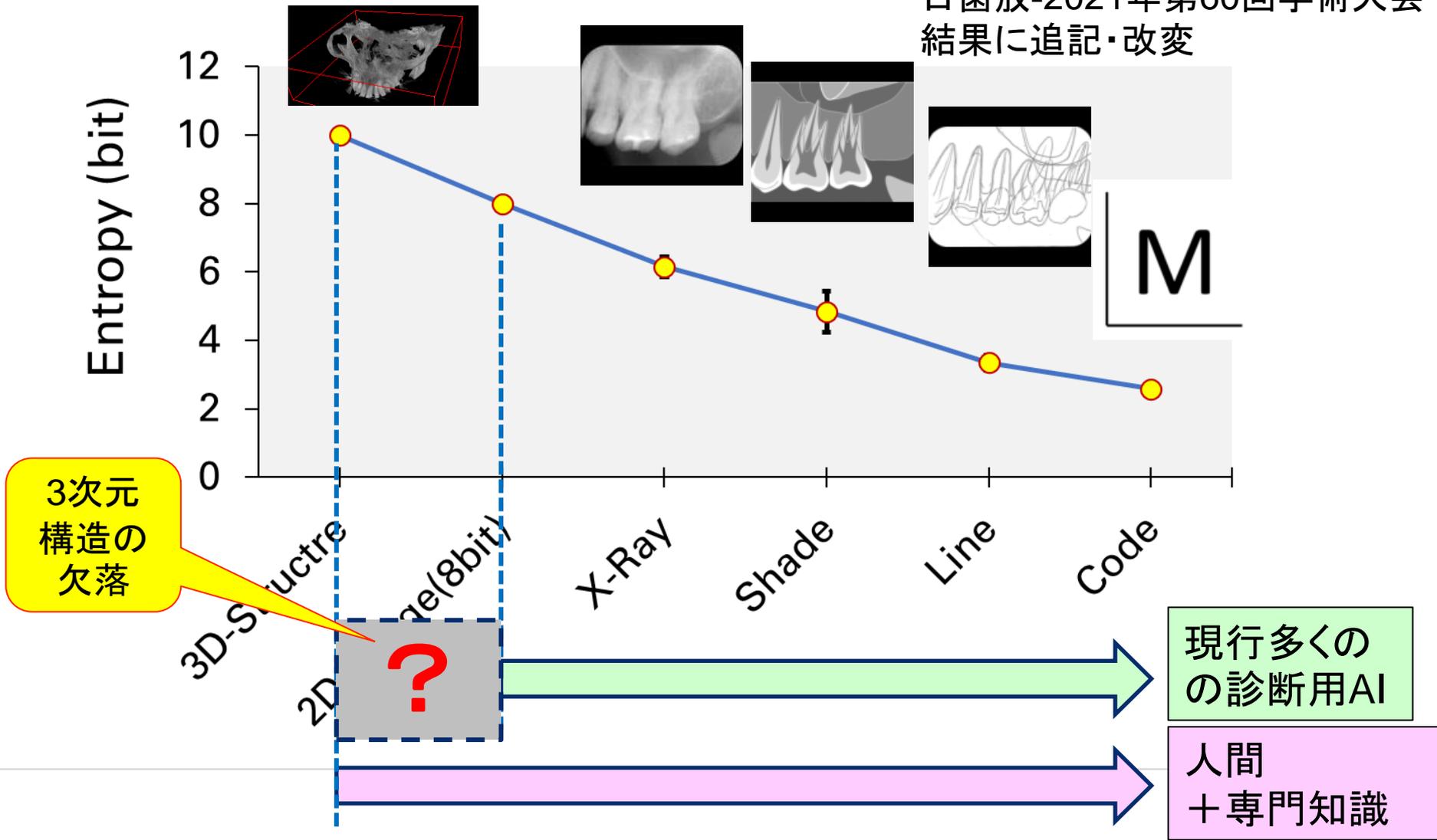


32x32 pixel画像、LeNetによる機械学習の結果

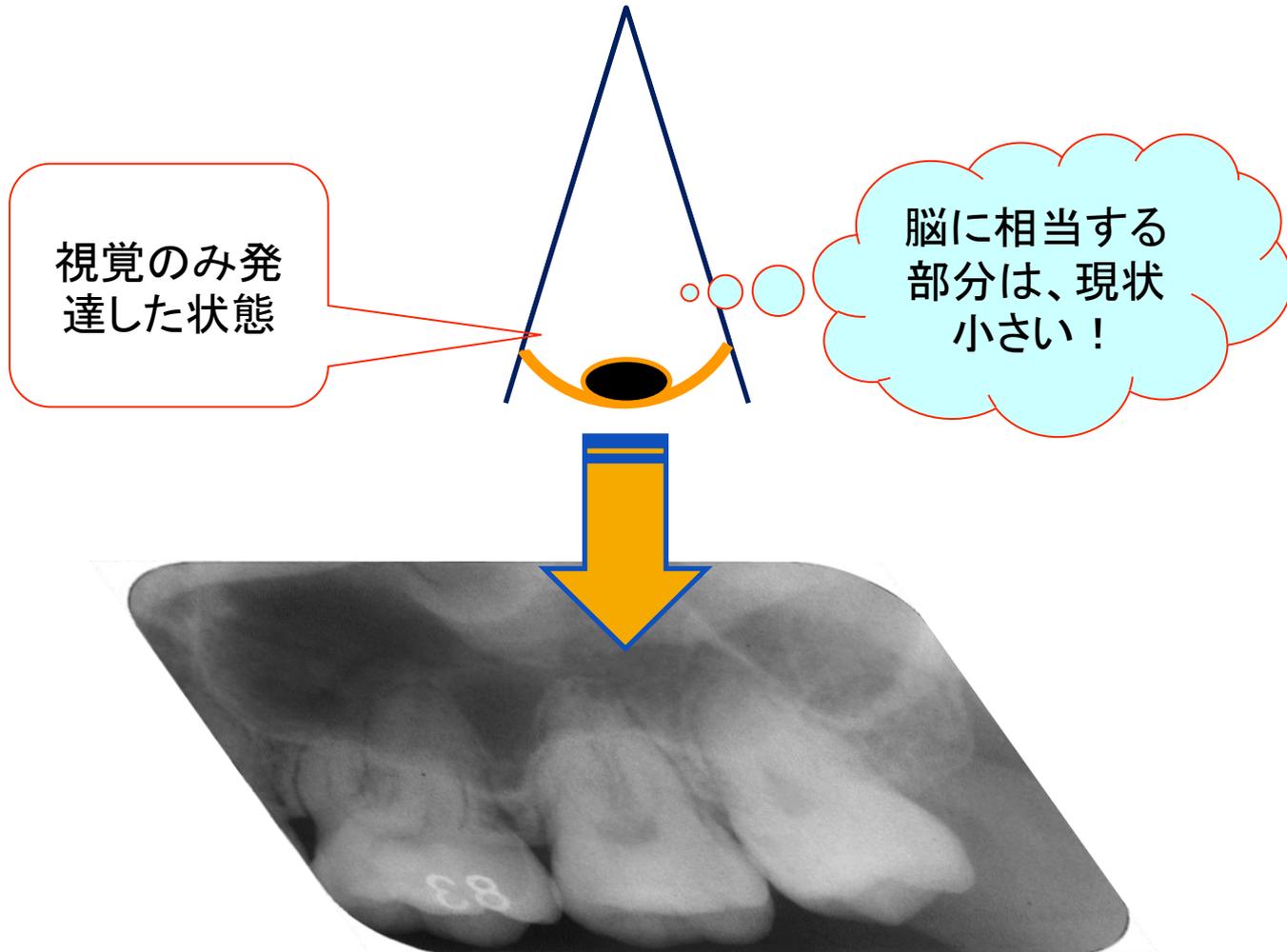
# 画像情報から診断への過程

## コード化 = エントロピーの減少

日歯放-2021年第60回学術大会  
結果に追記・改変



# 現状の深層学習で多く用いられている深層畳み込みニューラルネットワークでは 投影像のみ(画像のみ)から正答を導く



人間は画像のみではなく、解剖・組織・発生学的な知識等、さらに病理・診断学的な知識等を総動員して見る。

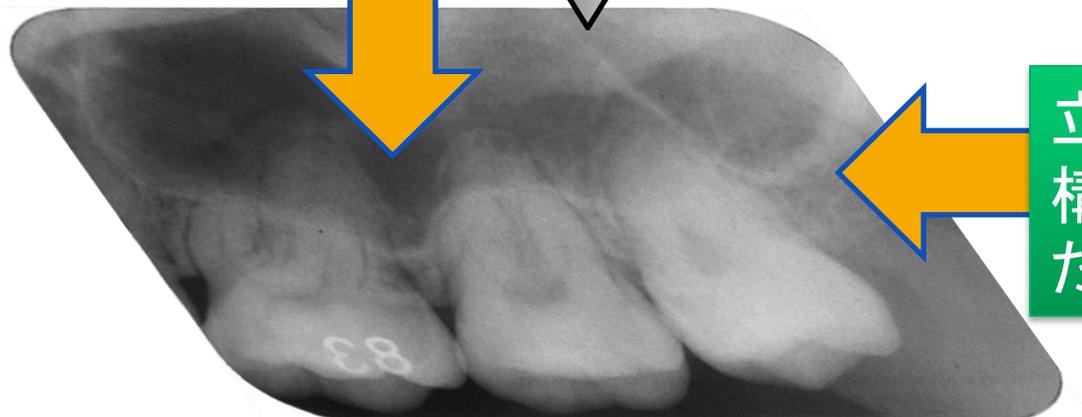
解剖・組織・発生学的な知識等

病理・診断学的な知識等

歯冠の形、咬頭の数・形、湾曲徴、歯根の数、形態、歯根徴等

+ 患者情報等

立体的に解剖構造が重積した像



# 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。

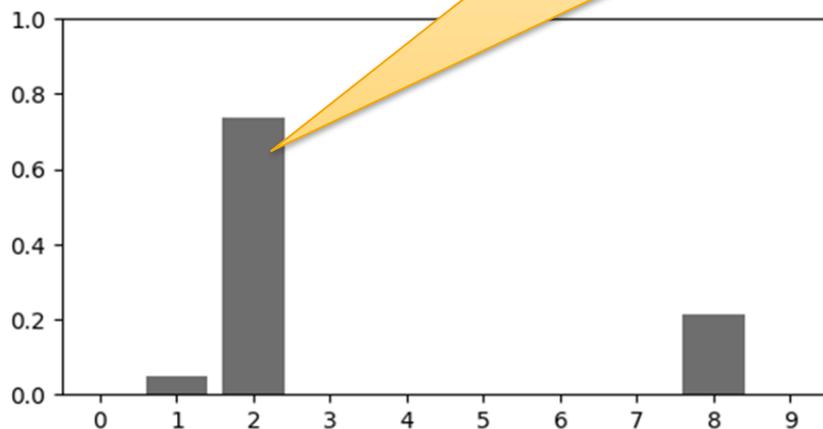
```
HTML(input_form + javascript)
# After appearing the Box for drawing...
# ↓ ↓ ↓ Please Draw number with pointing device...
```

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、  
「x」とは答えない。

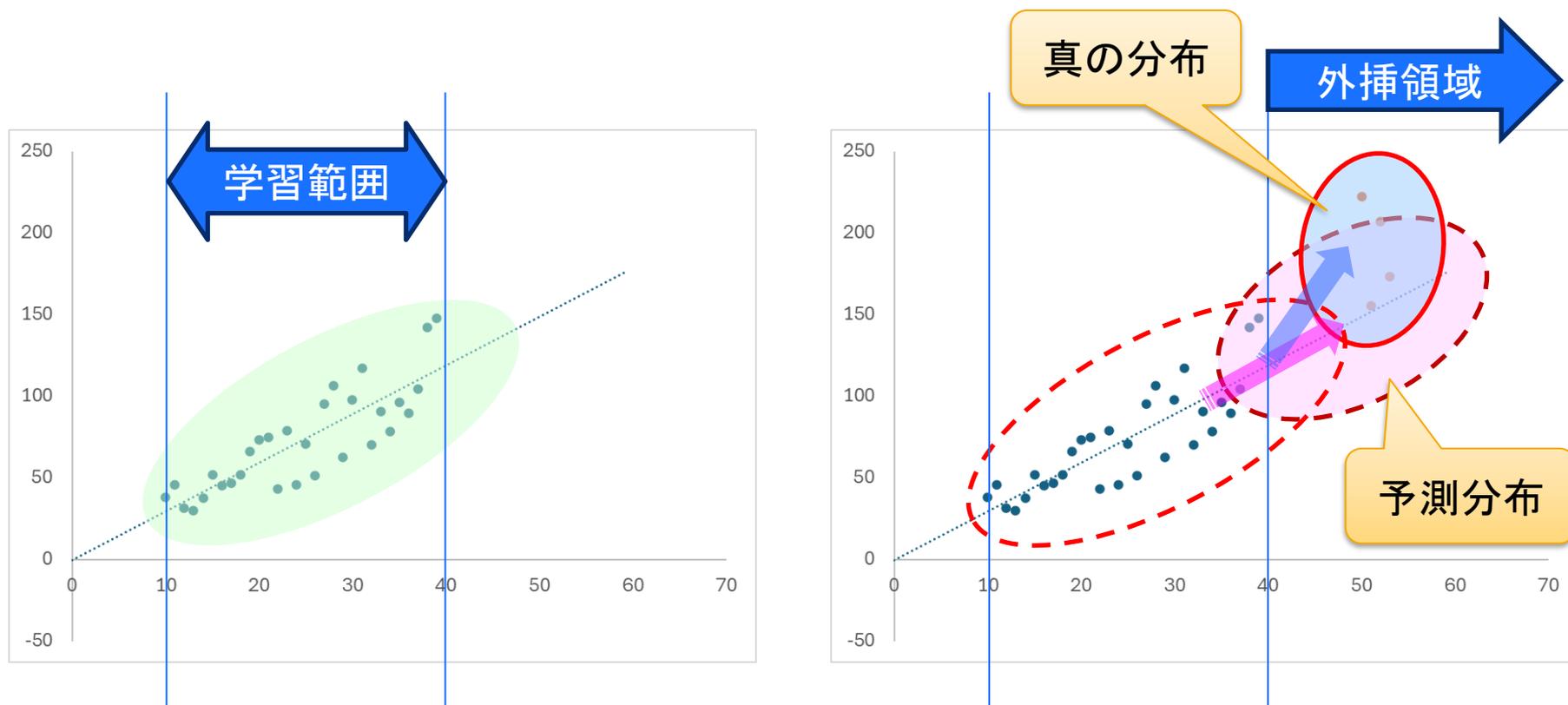
Out [10]: <BarContainer object of 10 artists>



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



Image created with ChatGPT-4o (DALL

# 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>

# 法律の問題

- **薬機法**（医薬品、医療機器等の品質、有効性及び安全性の確保等に関する法律）にて、規制がかけられている。
- AI技術を利用した医療機器の医薬品医療機器法上の取扱にかかるとの対応について（比較的分かり易い）  
<https://www.mhlw.go.jp/content/10601000/000361102.pdf>
- 【詳しく知りたい方】医薬品、医療機器等の品質、有効性及び安全性の確保等に関する法律
  - 厚生労働省HP  
[https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou\\_iryuu/iyakuhin/index.html](https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryuu/iyakuhin/index.html)
  - 法律原文  
<https://elaws.e-gov.go.jp/document?lawid=335AC0000000145>



公益社団法人

日本医学放射線学会

Google 提供



▶ English

HOME | 学会からのメッセージ | 学会案内 | 会員向けの情報 | 学術集会 | 専門医制度 | ガイドライン

## 会員の皆様へ

会員向けの情報

安全に関する情報

造影剤の安全使用に関する情報

会報・会告

学会誌・出版物

地方会

研究会

Japan Quantitative  
Imaging Biomarker  
Alliance (J-QIBA)

## AIソフトウェア認証一覧

2024年06月19日

公益社団法人日本医学放射線学会  
理事長 青木 茂樹

2022年度（令和4年度）診療報酬改定で画像診断管理加算3の施設基準に新たに「関係学会の定める指針に基づいて、人工知能関連技術が活用された画像診断補助ソフトウェアの適切な安全管理を行っていること。」ほかが必要要件として追加されました。

下記は、日本医学放射線学会画像診断管理認証制度によるAIソフトウェアと認証いたします。  
(AIソフトウェア：適切な安全管理を必要とする人工知能関連技術が活用された画像診断補助ソフトウェア)

- [AIソフトウェア認証一覧（29種）](#) 

日本医学放射線学会 AIソフトウェア認証一覧

(AIソフトウェア：適切な安全管理を必要とする人工知能関連技術が活用された画像診断補助ソフトウェア)

No.	製造販売業者名	承認・認証番号	販売名	認証日
1	エルピクセル株式会社	承認番号:30100BZX00142000	医用画像解析ソフトウェア EIRL aneurysm	2023/1/17
2	エルピクセル株式会社	承認番号:30200BZX00269000	医用画像解析ソフトウェアEIRL X-Ray Lung nodule	2023/1/17
3	キヤノンメディカルシステムズ株式会社	承認番号:30400BZX00123000	COVID-19 肺炎解析ソフトウェア SCO-PA01	2023/1/17
4	コニカミノルタ株式会社	承認番号:30300BZX00271000	画像診断支援ソフトウェア KDSS-CXR-AI-101	2023/1/17
5	シーメンスヘルスケア株式会社	承認番号:30200BZX00202000	AI-Rad コンパニオン	2023/1/17
6	富士フイルム株式会社	承認番号:30100BZX00263000	類似画像症例検索ソフトウェア FS-CM687型	2023/1/17
7	富士フイルム株式会社	承認番号:30200BZX00150000	肺結節検出プログラム FS-AI688 型	2023/1/17
8	富士フイルム株式会社	承認番号:30300BZX00145000	COVID-19 肺炎画像解析プログラム FS-AI693型	2023/1/17
9	富士フイルム株式会社	承認番号:30300BZX00188000	胸部 X 線画像病変検出 (CAD) プログラム LU-AI689型	2023/1/17
10	富士フイルム株式会社	承認番号:30300BZX00244000	肋骨骨折検出プログラム FS-AI691 型	2023/1/17
11	シーメンスヘルスケア株式会社	認証番号:302AABZX00047000	AI-Rad コンパニオン CT	2023/1/17
12	シーメンスヘルスケア株式会社	認証番号:302AABZX00092000	AI-RadコンパニオンMR	2023/1/17
13	日本メジフィジックス株式会社	認証番号:301ADBZX00029000	核医学画像解析ソフトウェア VSBONE BSI	2023/1/17
14	プラスマン合同会社	認証番号:301AGBZX00004000	Plus.lung.Noduleプラスラングノジュール	2023/1/17
15	エルピクセル株式会社	承認番号:30400BZX00285000	医用画像解析ソフトウェア EIRL Chest XR	2023/4/20
16	シーメンスヘルスケア株式会社	承認番号:30500BZX00032000	肺結節検出プログラム syngo.CT Lung CAD	2023/4/20
17	株式会社東陽テクニカ	認証番号:303ADBZX00098000	胸部CT読影支援システム ClearRead CT+DC	2023/4/20
18	キヤノンメディカルシステムズ株式会社	認証番号:302ABBZX00004000	汎用画像診断ワークステーション用プログラム Abierto SCAI - 1AP (AI機能オプション付)	2023/9/20
19	株式会社東陽テクニカ	認証番号:303ADBZX00013000	胸部X線骨組織透過処理システム ClearRead XR	2023/10/25
20	Qure株式会社	認証番号:301AGBZI00003000	VUNO Med®-LungCT ビューノメドラングシーティ	2023/12/29
21	エルピクセル株式会社	認証番号:303AGBZX00043Z00	医用画像解析ソフトウェア EIRL Brain Segmentation	2024/3/5
22	エルピクセル株式会社	認証番号:230AGBZX00107Z00	医用画像解析ソフトウェア EIRL Brain Metry	2024/3/5
23	エルピクセル株式会社	認証番号:304AGBZX00037Z00	医用画像解析ソフトウェア EIRL Chest CT	2024/3/5
24	株式会社メディカルブリッジ	認証番号:302AGBZI00005000	汎用画像診断装置ワークステーション用プログラム Aビュー	2024/3/25
25	株式会社ダブリューエスエム	認証番号:304AHBZI00007000	MR装置ワークステーション用プログラム Neurophet AQUA アクア	2024/5/2
26	PDRファーマ株式会社	認証番号:227ADBZX00091000	汎用画像診断装置ワークステーション用プログラム ポーンナビ.BSI	2024/5/2
27	PDRファーマ株式会社	認証番号:227ADBZX00090000	汎用画像診断装置ワークステーション用プログラム カーディオレポ。	2024/5/2
28	株式会社ドクターネット	承認番号:30300BZX00339000	胸部X線肺炎検出エンジン DoctorNet JLK-CRP	2024/6/12
29	株式会社ダブリューエスエム	認証番号:304AHBZI00035000	汎用画像診断装置ワークステーション用プログラム Neurophet SCALE PET スケール ペット	2024/6/12

# 胸部エックス線画像以外ほとんどが 反射光の原理に基づく画像

- 胸部エックス線画像は、頭頸部のように複雑な解剖構造の重積が非常に少なく、比較的扱いやすい。機械学習レベルを含め、古くから研究されてきており、肋骨の影響は比較的簡便に対処可能。
- 内視鏡は反射光で扱える。
- CTやMRIの3次元画像も透過像としては扱われず、主に断面画像(すなわち反射光で対応可能)ないし、反射光で表示可能な立体画像として扱われる。

添付文書(薬機法で定められた文書)からの抜粋

[https://www.pmda.go.jp/PmdaSearch/kikiDetail/ResultDataSetPDF/671001\\_30300BZX00188000\\_A\\_01\\_10](https://www.pmda.go.jp/PmdaSearch/kikiDetail/ResultDataSetPDF/671001_30300BZX00188000_A_01_10)

\*\*2024年10月作成(第9版)

承認番号 30300BZX00188000

\*2024年4月作成(第8版)

プログラム1 疾病診断用プログラム

管理医療機器 汎用画像診断装置ワークステーション用プログラム JMDNコード:70030012

## 胸部 X 線画像病変検出(CAD)プログラム LU-AI689 型

### 【警告】

本プログラムの異常領域検出機能は医師の読影の補助を目的としたものであり、検出結果を単独でスクリーニング及び確定診断に利用しないこと。

本プログラムは診断を行う装置ではない。診断は医師の責任において行うこと。

本プログラムの検出結果には、対象所見を検出しない場合(偽陰性)が存在すること、および、正常構造を誤って検出する場合(偽陽性)が存在することを考慮すること。

### 【形状・構造及び原理等】

#### 【形状・構造】

本プログラムは、胸部 X 線画像をコンピュータ処理し、対象となる異常領域の存在の確信度を画素ごとに推定し、Heat Map 画像と最大値を画像上に表示するプログラムである。

### ・検出アルゴリズム概要

本プログラムは、機械学習によって得られる人工知能(AI)により、胸部 X 線画像中の対象とする病変の可能性のある領域を検出するものである。本アルゴリズムは市場にて学習し変化することはない。

本プログラムのアルゴリズムは、以下のとおりである。

1. 入力画像に、機械学習によって得られた Deep Convolutional Neural Network (CNN)の判別器を適用するための正規化を行い、サイズ補正処理を実行する。
2. 入力画像の各画素に対して、その画素と周辺領域によって生成される矩形領域を生成し、対象所見(結節/腫瘍影、浸潤影、気胸)のそれぞれの CNN の判別器を適用する。
3. 各画素で各所見の判別器の出力値(0 から 1)を比較し、最も高い値を「異常領域の確信度」として出力画像の画素に割り当てる。
4. 出力画像の画素値を色情報に変換し、これを「Heat Map」画像として出力する。ここで、各画素値は 0 から 1 の出力値を確信度として 0 から 100 の整数に変換する。

#### ・検出対象

検出対象となる X 線画像及び所見は、以下である。

#### ・画像条件

画像種類	胸部 X 線画像 (正面像)
撮影範囲	胸部 (肺野全体が含まれていること)
対象患者	成人

成人未満に対しては検出機能が保証されないこと、また診断に利用しないように留意すること。

#### ・対象所見

結節/腫瘍影、浸潤影、気胸

### 【使用目的又は効果】

本プログラムは、X 線診断装置又は X 線透視診断装置から提供された胸部の画像情報 (以下胸部 X 線画像という) をコンピュータ処理し、処理後の画像情報を診療のために提供する。胸部 X 線画像用のコンピュータ検出支援 (Computer Aided Detection: CAD) 機能を有する。

本プログラムの異常領域検出機能は、胸部 X 線画像中の結節/腫瘍影、浸潤影、気胸の存在の確信度を画素ごとに推定し、Heat Map 画像と確信度の最大値を提示することで、医師の見落とし防止を支援することを目的とする。

本プログラム位置づけは「医師の読影の補助」であり、本プログラムによる検出結果のみでスクリーニングや確定診断を行なうことは目的としない。

#### ・読影

読影に用いる画像診断装置ワークステーションの使用方法及び表示内容については、使用する画像診断装置ワークステーションの取扱説明書を参照すること。

- (1) 画像診断装置ワークステーションにて、本プログラムを用いずに通常の手順で X 線画像を読影する。
- (2) 次に、解析結果を読み込み、Heat Map と確信度の最大値を表示する。
- (3) 読影者が病変を検出していない症例又は検出していない所見に対して、確信度及び Heat Map が表示された場合、読影者は必要に応じて画像診断装置ワークステーション上で再読影を行い、診断を決定する。

添付文書 (薬機法で定められた文書) からの抜粋

[https://www.pmda.go.jp/PmdaSearch/kikiDetail/ResultDataSetPDF/671001\\_30300BZX00188000\\_A\\_01\\_10](https://www.pmda.go.jp/PmdaSearch/kikiDetail/ResultDataSetPDF/671001_30300BZX00188000_A_01_10)

## 【使用上の注意】

1. 医師による読影後に本プログラムの処理結果を参照すること。
2. 本プログラムは専門知識を有する医師によって使用されること。
3. 本プログラムによって表示される結果は、参考情報として使用すること。本プログラムのみ依存した確定診断は行わないこと。
4. 表示された画像、又はその付帯情報に異常が疑われる場合は、本プログラムを使用して読影行為は行わないこと。
5. システム障害時に備え、代替手段を用意すること。
6. 表示される情報は画像所見の存在の確信度である。悪性を鑑別するもの、陽性の確率を表すものではない。

添付文書(薬機法で定められた文書)からの抜粋

[https://www.pmda.go.jp/PmdaSearch/kikiDetail/ResultDataSetPDF/671001\\_30300BZX00188000\\_A\\_01\\_10](https://www.pmda.go.jp/PmdaSearch/kikiDetail/ResultDataSetPDF/671001_30300BZX00188000_A_01_10)

7. Heat Map の表示は異常所見領域の輪郭を正確に示すものではない。
8. X線画像の画質が画像診断に適さない条件では、検出性能が低下する場合があるため、診断に適した鮮明な画質で使用する事。
9. 臓器、骨と重なる病変は、検出されない(偽陰性)可能性があるため注意すること。
10. 胸水、微細粒状影、網状影、骨折などを誤って検出する(偽陽性)可能性があるため注意すること。
11. ネットワークを使用する場合は、通信状況が良い環境で使用する事。
12. 本プログラムの解析結果を、他の類似するコンピュータ診断支援プログラムが出力する結果と取り違えないように注意すること。

添付文書(薬機法で定められた文書)からの抜粋

[https://www.pmda.go.jp/PmdaSearch/kikiDetail/ResultDataSetPDF/671001\\_30300BZX00188000\\_A\\_01\\_10](https://www.pmda.go.jp/PmdaSearch/kikiDetail/ResultDataSetPDF/671001_30300BZX00188000_A_01_10)

### <その他の注意>

267 画像(正常例:195 例、異常例:72 例)を後向きに収集し、観察者 10 名(専門医:3 名、専門外医(経験 10 年以上):3 名、専門外医(経験 5 年未満):4 名)で読影試験を実施し、本プログラムの有効性を評価した。

対象所見に対する結果を以下に示す。

		感度 (%)	Figure of Merit
医師単独	専門医	74.7	0.837
	専門外(≥10 年)	53.4	0.724
	専門外(<5 年)	58.1	0.756
	全体	61.7	0.771
CAD 併用	専門医	79.9	0.862
	専門外(≥10 年)	72.6	0.829
	専門外(<5 年)	73.4	0.829
	全体	75.1	0.839
差	専門医	5.2	0.025
	専門外(≥10 年)	19.2	0.105
	専門外(<5 年)	15.3	0.073
	全体	13.4	0.068

医師単独読影に対する CAD 併用読影の Figure of Merit の値は、統計学的に有意に向上した(p<0.05)

添付文書(薬機法で定められた文書)からの抜粋

[https://www.pmda.go.jp/PmdaSearch/kikiDetail/ResultDataSetPDF/671001\\_30300BZX00188000\\_A\\_01\\_10](https://www.pmda.go.jp/PmdaSearch/kikiDetail/ResultDataSetPDF/671001_30300BZX00188000_A_01_10)

# 5月22日・薬機法で承認された 歯科エックス線画像のAIによる評価

- <https://panoscope.media-inc.co.jp/>



The screenshot displays the Panoscope website interface. At the top, the navigation bar includes the logo 'PanoSCOPE' and menu items: '製品案内' (Product Information), '顎骨脆弱度評価' (Jawbone Fragility Evaluation), '医科歯科連携' (Medical-Dental Collaboration), 'トピックス・関連論文' (Topics & Related Papers), and a 'WEBマニュアル・利用規約' (Web Manual & Terms of Use) button. The main content area features the text: 'This screening software connects medical and dental sciences.' followed by a red 'Y' logo. Below this, the headline reads: '骨粗鬆症による一次骨折予防のため 顎骨脆弱度評価を支援する 「歯科パノラマX線画像解析AI」'. An image shows a person pointing at a monitor displaying a dental X-ray. To the right, a large image shows a pair of hands in blue gloves holding a physical dental X-ray film. At the bottom right, there is a 'MEDIA' logo and a circular 'Contact Us' button with an envelope icon and the text 'お問い合わせ'.

<https://panoscope.media-inc.co.jp/pdf/pamphlet.pdf>

# PanoSCOPEの添付文書

<https://panoscope.media-inc.co.jp/pdf/Package-Inserts.pdf>

令和6年5月作成(第1版)

## 【形状・構造及び原理等】

1. 概要  
本プログラムは、歯科用パノラマ X 線画像を解析し、歯科医師に考査結果を提示する。  
具体的には、歯科パノラマ X 線汎用ワークステーション(周辺機器等を有する)(以下、コンピュータ)により、歯科パノラマ X 線のみ計測と形態評価を行う。

## 2. 主たる機能

機能	
下顎皮質骨厚み自動計測機能	左右対称且つ自動的に計測
下顎皮質骨形態自動評価機能	下顎皮質レベル Mandib

## 3. 提供形態 記録媒体

## 【使用目的又は効果】

本プログラムは、口外汎用歯科装置で撮影された歯科パノラマコンピュータ解析処理をし、医師は、歯科パノラマ X 線画像を解析し、歯科医師による顎骨形態指標(MCI 分類)を参考情報プログラムによる検出結果のみで判

## 【使用方法等】

1. コンピュータの要件  
本プログラムは、下記の仕様を既定した方法(添付文書又はプロ手順)でインストールして使用する。

<コンピュータの仕様>

・ 汎用 PC  
OS : Windows  
ストレージ(空き容量) : 500MB  
メモリ : 4GB 以上  
CPU : 第4世代  
・ 表示モニター  
解像度: 1920×1080ピクセル以上  
カラー表示: True Color(32bit)以上

## 2. 使用準備

(1) コンピュータの電源を入れる。  
(2) 本プログラムを起動する。

## 3. 操作

(1) 画像診断装置から記録媒体(画像データ)を保存する。  
(2) 表示モニター上で下顎骨がプログラムをインストールしたコンピュータ。  
(3) 画像データを本プログラムに送る。  
①操作者が jpg, png, tiff, bmp 画像を入力する。  
②本プログラムは、登録対象画像  
③操作者は、まず、画像を取り

-画像データを転送する際に使用する記録媒体に関して、パスワードの設定などは医療機関ごとの規定にあわせて対応すること。

-本プログラムのインストールを実行する前に必要な準備

- ・コンピュータの管理者権限があるユーザー名でインストールすること。
- ・起動している他のプログラムをすべて終了すること。
- ・本プログラムをインストールしたコンピュータは外部ネットワークに接続しないこと。

## 2. 本プログラムで評価する場合の画像に対する要求事項

- 適切な精度管理が実施された口外汎用歯科 X 線診断装置から得られた画像に対してのみ適用すること。
- 被写体の位置付け、撮影方法、画質等が解析に影響を与えることを考慮した上で、本処理の結果を慎重に検討すること。

## 3. 本プログラムを操作する際に必要な注意事項

- 本プログラムの添付文書、ユーザーズマニュアルを必ず参照すること。
- 本プログラムは、専門家のみが使用することを想定している。本プログラムは、医療従事者の検査を支援することを目的としており、その臨床判断に完全に代わるものではない。
- 本プログラムを指定以外のアプリケーションと併用しないこと。
- 推奨環境以外で、本プログラムを使用しないこと。
- 修理や保守は、当社又は当社の指定する業者は依頼すること。

カファイル形式の誤差について

より 50 症例における平均、標準偏差、最小値、最大値を求めた。画像フ元配置分散分析及び多重比較 p<0.05)

より 50 症例における平均、標準偏差、最小値、最大値を求めた。画像フ元配置分散分析、Welch 検定及び有意水準 p<0.05)

を基準とした誤差について、5つあった。有意水準(p 値)を表に示

Tukey HSD
0.435
1.000
0.789

標準誤差、平均値の 95%信頼区及び表 5 に示す。

の誤差(N=50)

平均値の 95%信頼区間	最小値	最大値	
下値 上値			
1.44	2.56	0	7
1.97	3.19	0	8
1.83	2.93	0	7
2.05	3.27	0	8
2.06	3.38	0	9
2.20	2.73	0	9

の誤差(N=50)

平均値の 95%信頼区間	最小値	最大値	
下値 上値			
0.16	0.37	0.0	1.6
0.15	0.39	0.0	2.1
0.18	0.37	0.0	1.4
0.15	0.39	0.0	2.1
0.15	0.39	0.0	2.1
0.22	0.32	0.0	2.1

の誤差(N=50)

平均値の 95%信頼区間	最小値	最大値	
下値 上値			
0.19	0.42	0.0	1.7
0.21	0.50	0.0	1.9
0.16	0.34	0.0	1.6
0.21	0.50	0.0	1.9
0.20	0.49	0.0	1.9
0.27	0.38	0.0	1.9

【又は名称等】

# PanoSCOPEの添付文書から抜粋

- 被写体の位置付け、撮影方法、画質等が解析に影響を与えることを考慮した上で、本処理の結果を慎重に検討すること。

AIの前段階でのエックス線撮影時点での影響を考慮しないといけない？ 何故？？？

- 本プログラムは、専門家のみが使用することを想定している。本プログラムは、医療従事者の検査を支援することを目的としており、その臨床判断に完全に代わるものではない。

AIは歯科医師国家試験レベルぐらい、簡単にクリアするんじゃないの？ 法律改正すべきでは？

# PanoSCOPEの添付文書から抜粋

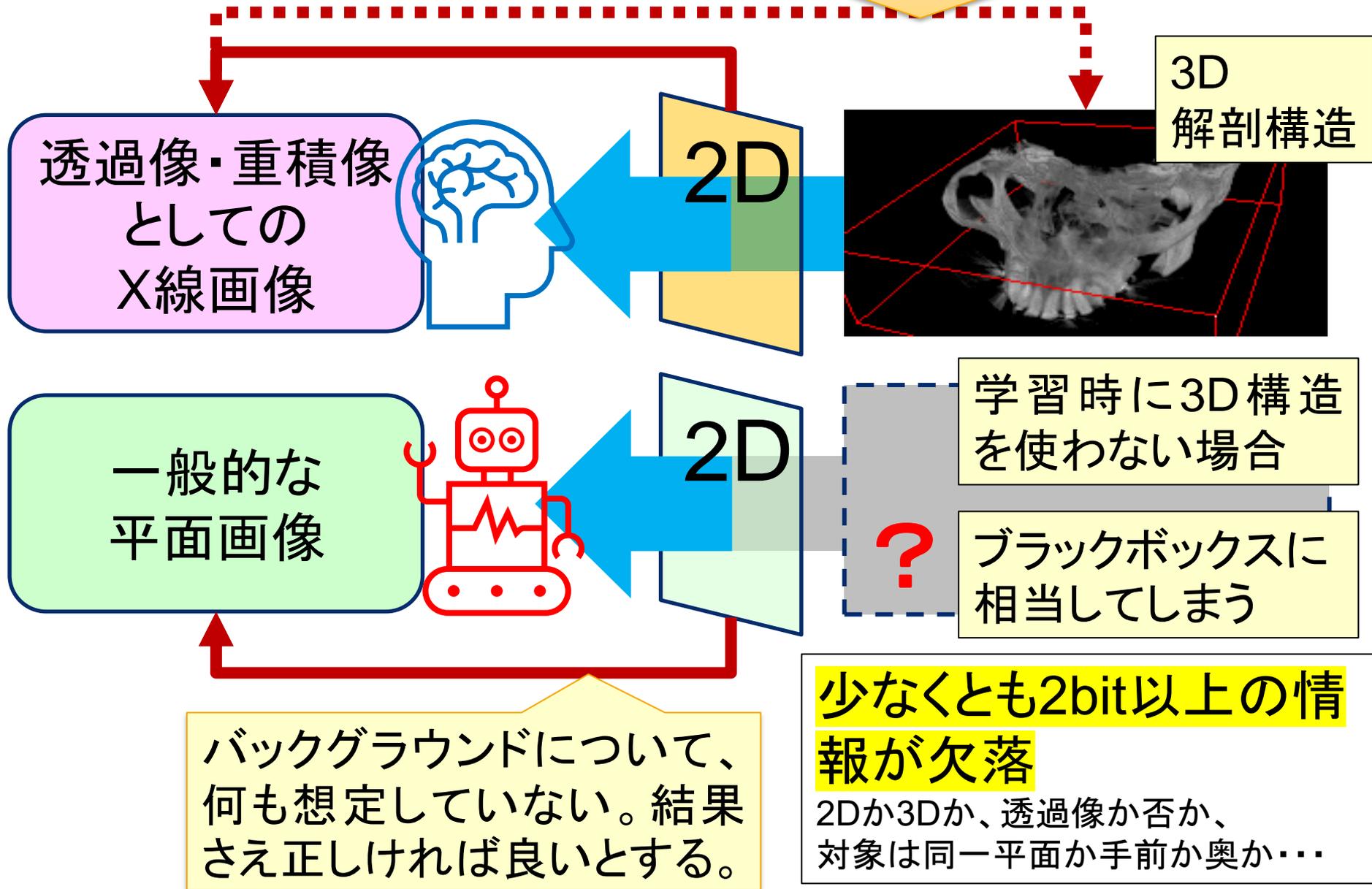
- 被写体の位置付け、撮影方法、画質等が解析に影響を与えることを考慮した上で、本処理の結果を慎重に検討すること。

撮影方法や位置づけによって、下顎下縁の皮質骨への舌骨や障害陰影などの重積の程度が変化するから

- 本プログラムは、専門家のみが使用することを想定している。本プログラムは、医療従事者の検査を支援することを目的としており、その臨床判断に完全に代わるものではない。

このプログラムは入力に対する推定結果しか出せず、その理由を説明できないので歯科医師という専門家が説明責任を担うから

# 3次元解剖構造を想定して考えている



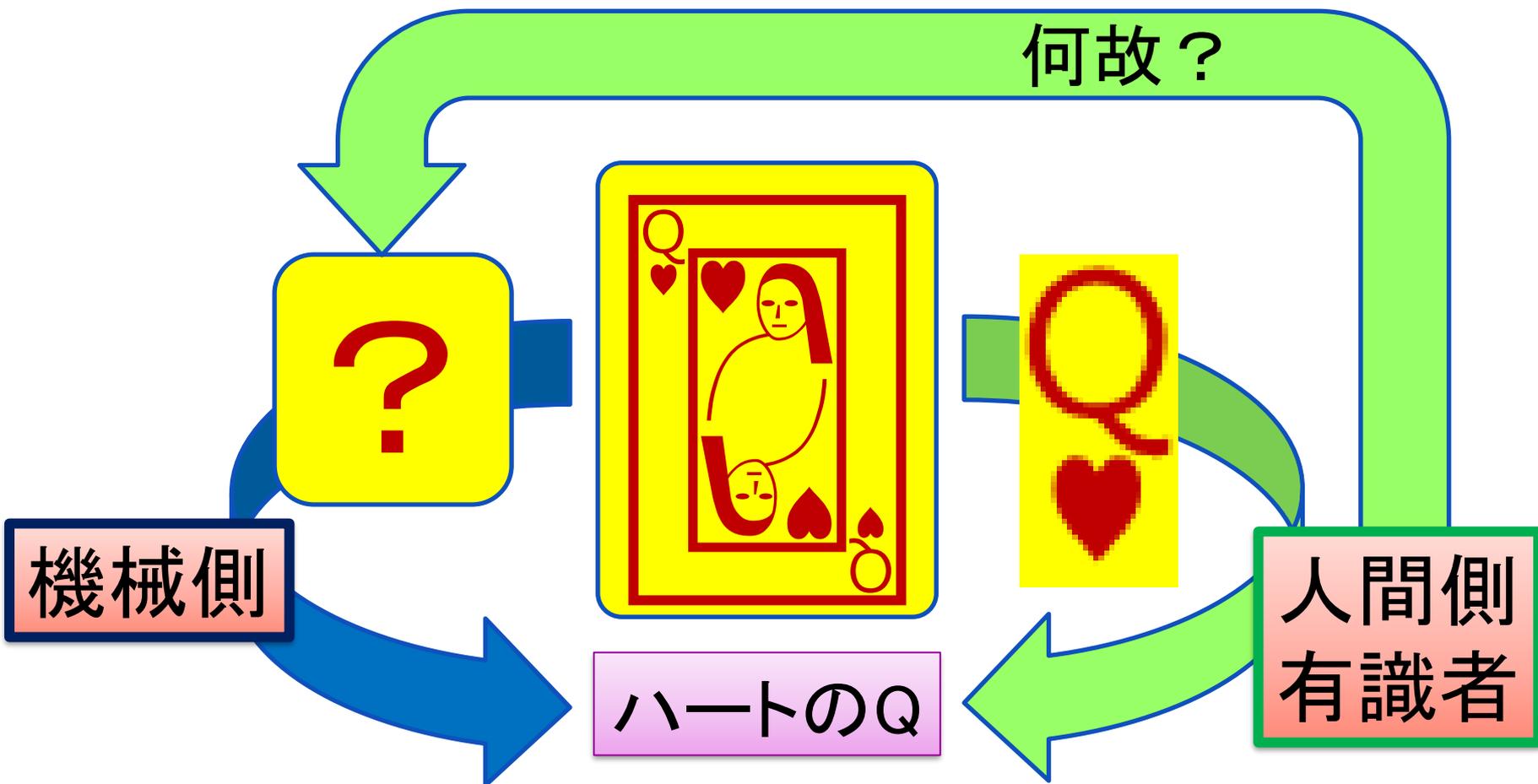
# まとめ・エックス線画像の認識

- 人間
  - 3次元的な解剖構造を基本にして判断する
- 機械
  - (現行システムの大部分は) 平面画像のみで学習
  - 解剖構造を見ているわけではない
  - 別の部位(?というより別の構造)を見ている
  - ヒートマップが適切な位置を示していたとしても、重積像の場合、「3次元的に、どの構造？」が不明となる
  - 機械内部では人間の思考とは異なる過程(例えば畳み込みによるパターン認識)なので、ブラックボックスになる

# 問題の本質について(1)

## Automation Biasの問題

- 深層学習では、機械側の判断根拠を知ることが困難とされている。



# 問題の本質について(2)

たとえ99%の精度  
を持つシステムで  
あっても、

Automation Biasの問題

- 診断支援ないし教育支援として用いる場合、  
診断利用者ないし教育介助者には該当領域  
の有識者レベルの知識で介入が必要不可欠

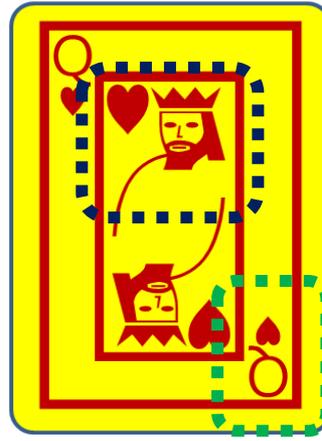
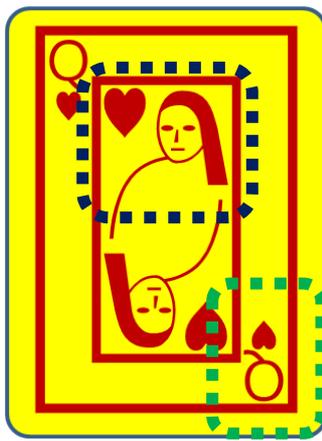
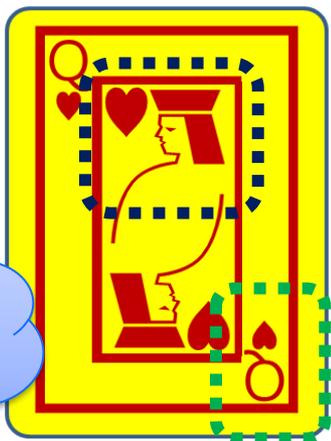
ハートのJ,Q,K  
だよ

人間側  
初学者

違うよ！ 全て  
ハートのQ

機械側

若い男性  
女性  
年上の男性



人間側  
有識者

トランプも  
ジェンダー  
フリーか

介入

# 問題の本質について(3)

## Automation Biasの問題

- **【問題の本質】**機械側の判断根拠が人間側の判断根拠と同じかどうか不明であること。

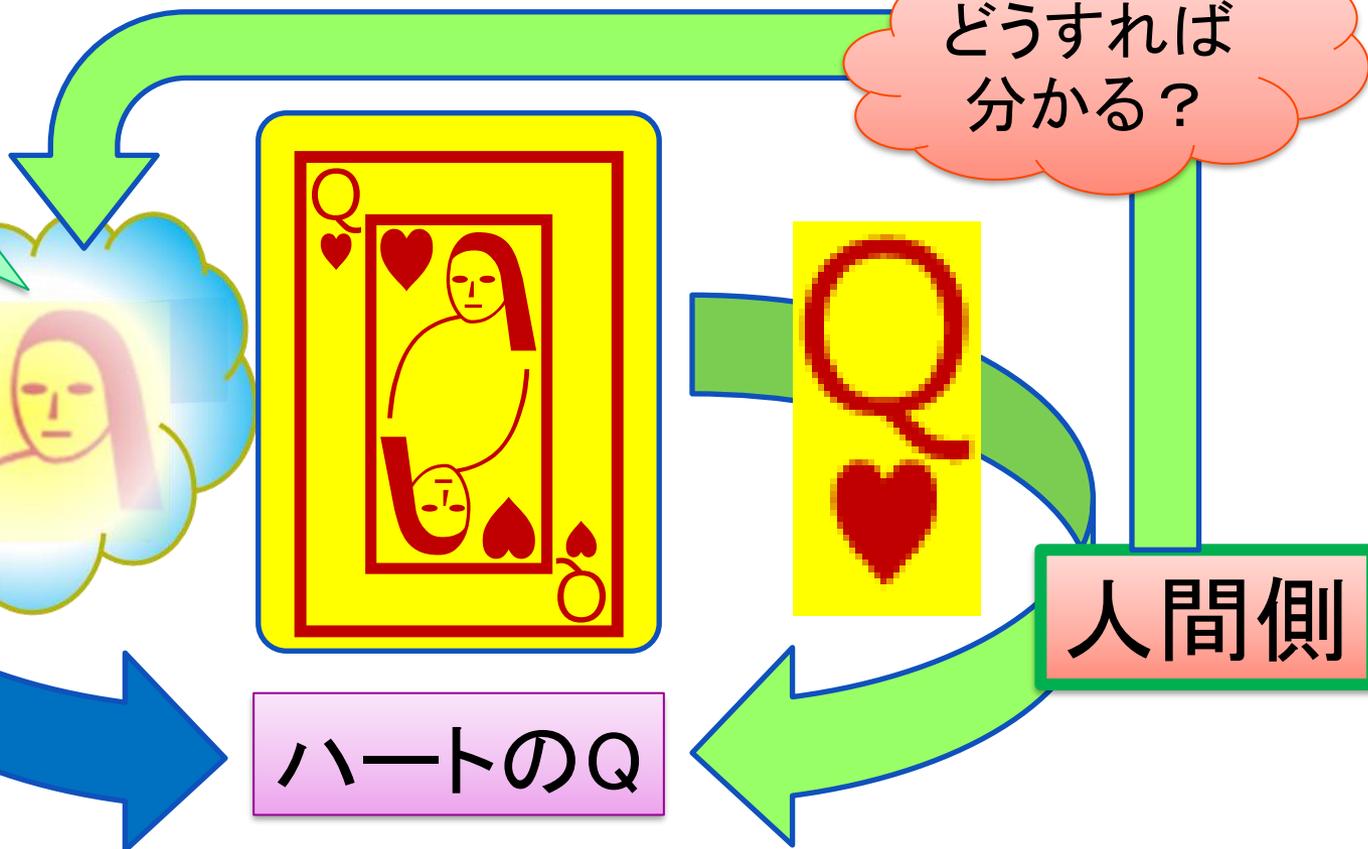
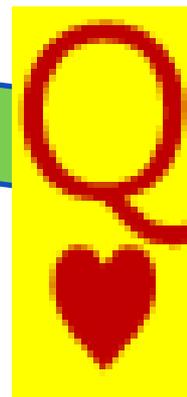
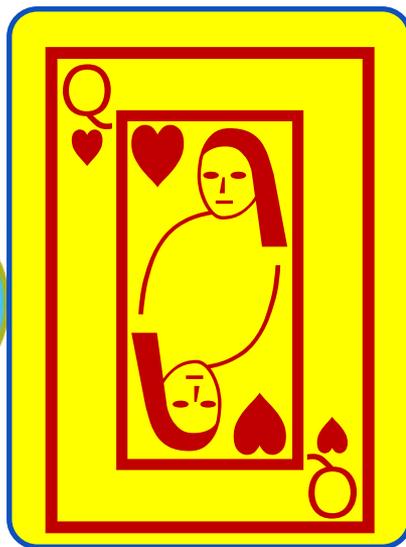
もしかしたら、画期的な知見かも知れない。

どうすれば分かる？

機械側

人間側

ハートのQ



# 既存の知識体系を崩さずに 使いこなすだけの力量が求められる

- 「がんの未知なる特徴をAIが発見ーがんの画像から、再発に関わる新たな知識を自力で獲得」(2019年12月18日、理化学研究所、日本医科大学、日本医療研究開発機構)
  - [https://www.riken.jp/press/2019/20191218\\_2/index.html](https://www.riken.jp/press/2019/20191218_2/index.html)
- 組織・解剖構造等の既存の知識に囚われないからこそ、成し遂げられる成果。人間の思考過程とは全く異なる視点からの「気づき」相当。
- 一方で、組織・解剖構造などの既存の知識に基づいた学習を踏まえた医学という体系を捨て去れば良いということではない。
- 両者を併用しうる状況こそが望まれる。
- そういった意味からも、あくまで「診断補助」としてのシステムとして使いこなす力量こそが求められる。
- 薬機法等での制約に関する記載は、「診断補助」のシステムから逸脱することの無いような、十分な情報提供・説明責任を伴うシステムとして広く人間側に受け入れられる方向性を求めるもの。

# 識別系のAIと生成系のAIはほぼ対称

人間側にとってありもしない物を創り出すことがある

デコーダ  
生成器

バイアスの問題  
誤認の発生率は  
ゼロではない

エンコーダ  
識別器

人間側にとっての適切な分類(記号)へと着地(接地)できないこともある

コード  
形式情報

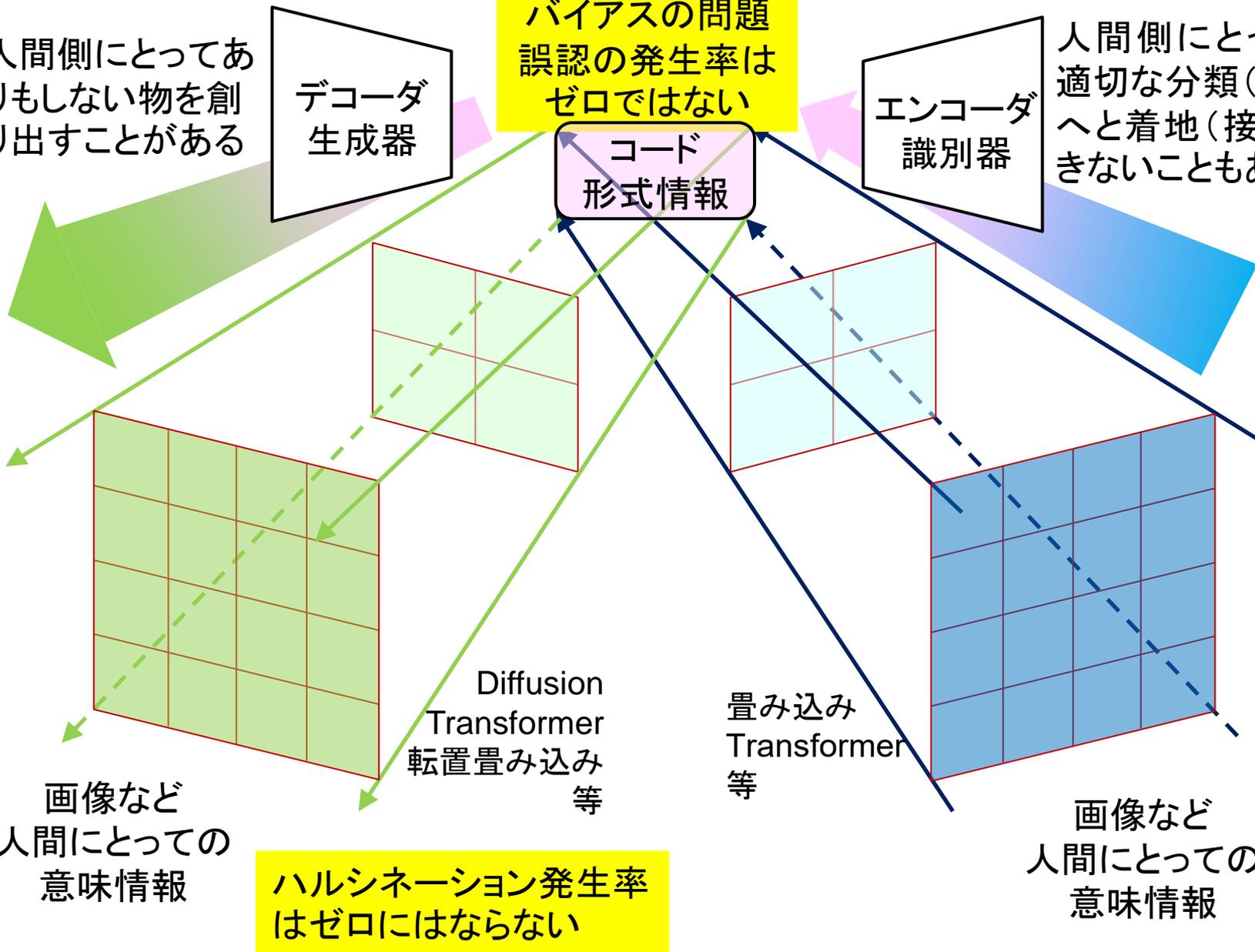
Diffusion  
Transformer  
転置畳み込み  
等

畳み込み  
Transformer  
等

画像など  
人間にとっての  
意味情報

ハルシネーション発生率はゼロにはならない

画像など  
人間にとっての  
意味情報



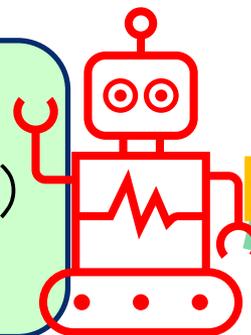
# 情報の意味と形式と 記号接地問題との関係

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

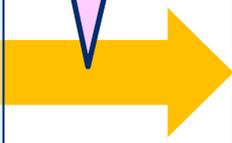


いわゆるAI  
LLM (GPTなど)  
+ 強化学習



**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



記号化・  
コード化

**意味情報  
解剖構造・物質**

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。

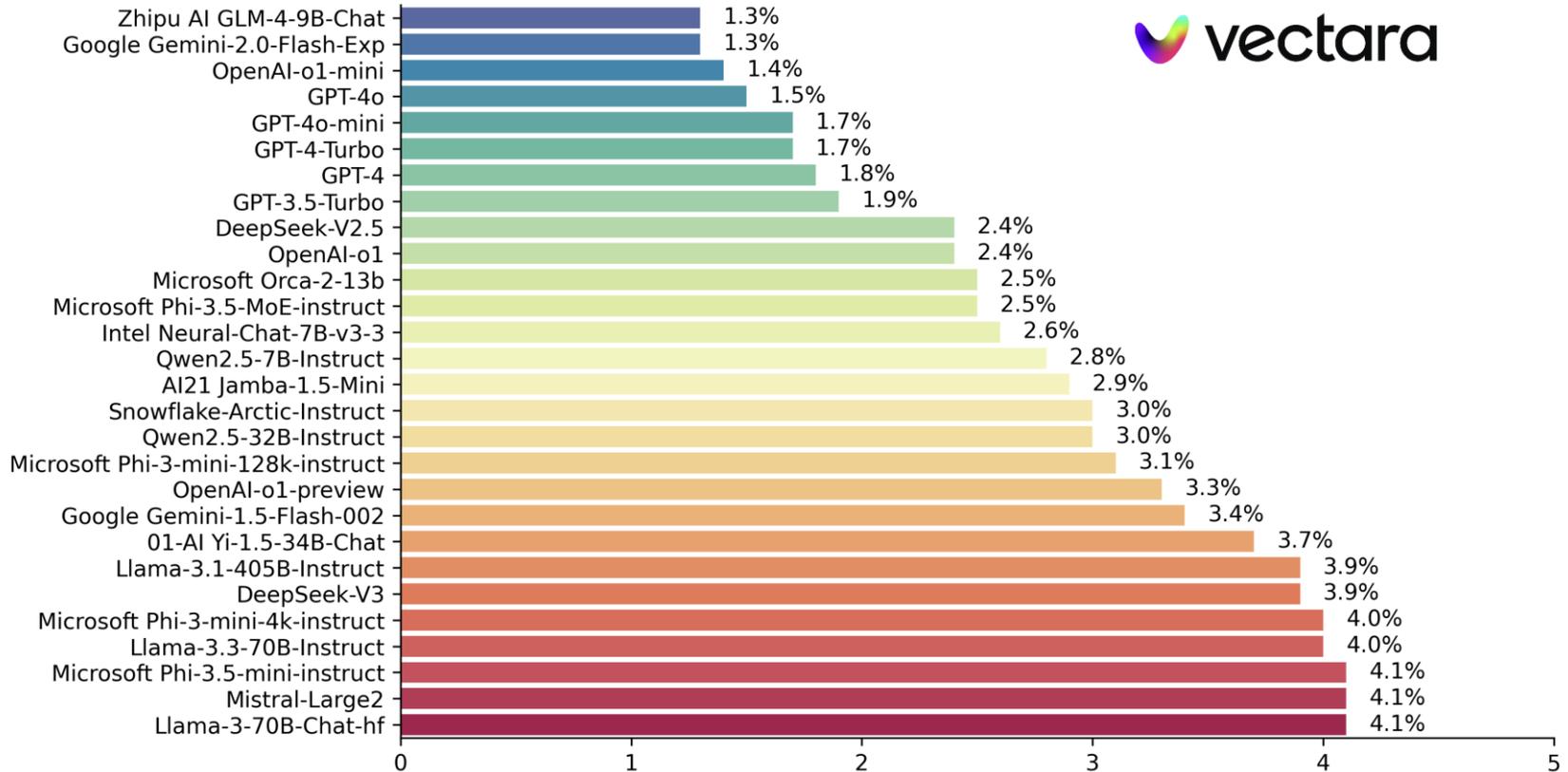
形式的な言葉・記号の文脈上での  
接続・関係から確率的に文章を生成  
(文脈レベルのみでの意味情報)

例えば「**歯  
根囊胞**の典  
型像」の作  
成を指示

Image created with  
ChatGPT-4o (DALL·E

# Hallucination

Hallucination Rate for Top 25 LLMs



Last updated on January 20th, 2025

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>

# 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](https://www.jspa-nsc.com/dcms_media/other/igakunoayumi_21.pdf)

# LLM (Large Language Model) 大規模言語モデルの現状での限界

- 大規模言語モデルは単語(記号)間の相関関係から統計学的に文章を作り出している。
- 因果関係を説明しているように見える場合、あるいは、文章の意味・内容を深く知った上で知的に処理していると思える場合でも、単なる相関関係でしかなく、質問の変化にて、因果的に誤った解答になる。
- Jin Z, Liu J, Lyu Z, Poff S, Sachan M, Mihalcea R, et al. Can Large Language Models Infer Causation from Correlation? [Internet]. arXiv; 2023 [cited 2023 Jul 24]. Available from: <http://arxiv.org/abs/2306.05836>

# いわゆるAIでの意味情報とは？

- 癒やし系や配膳ロボット、ないしChatGPTは単に機械的に応答している(形式情報として内部で処理している)だけ。
- しかしながら受け手側は、そこに「意味」や「感情」を見いださう。
- なぜなら「両者の関係、全体の状況が言葉に意味を与えている」から。
- このため、AI側内部にて人間側と同じように「意味」が扱われ、意味情報として伝わってきていると思っ込んでしまう。

# 「真実のような嘘」(Hallucination)

- 機械側は形式情報として論理的に正しく処理している。
- 人間側にとっての意味情報として解釈可能な形式で出力しようとするときに人間側にとっての虚言が混在する。
- 知らないことを知らないと出力せずに、類似度の高い、正解に近いであろう回答を出力するように学習しているから当然のこと。(ただし部分的に制御する方法を組み込むこともできる)
- 正しく解釈し、正しく扱うには、当該領域(分野)の専門家レベルの知識が必要になる。

# 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)難しい問題を解決しうる。
  - 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。

# TransformerおよびChatGPTについてネット上参考となる資料

- 30分で完全理解するTransformerの世界  
<https://zenn.dev/zenkigen/articles/2023-01-shimizu>
- ChatGPT(とその周辺)の技術  
<https://zenn.dev/zenkigen/articles/2023-05-kurihara>
- ChatGPT API を基礎から理解する  
@PND 投稿日 2023年06月21日  
<https://qiita.com/PND/items/c9e9d449539a460515a4>
- ChatGPT の仕組みを理解する(前編)  
<https://tech-blog.abeja.asia/entry/chat-gpt-first-half-202307>
- ChatGPT の仕組みを理解する(後編)  
<https://tech-blog.abeja.asia/entry/chat-gpt-second-half-202307>
- LLMにおける強化学習の問題点(人間のフィードバックからの強化学習の未解決の問題と基本的な限界) Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback  
<https://huggingface.co/papers/2307.15217>  
<https://arxiv.org/abs/2307.15217>

# 画像診断領域での AIに対する認識の変遷



- 少し前には「いずれAIは放射線科医にとって代わる」と言われていた時期もあったが・・・
- 2017年北米放射線学会 (RSNA) では、「AIが放射線科医に置き換わることはなく、AIを使う放射線科医が、使わない放射線科医にとって代わるだろう」といった趣旨の発言が注目されている。( Curtis Langlotz )
- 2019年2月号INNERVISION  
[https://www.innervision.co.jp/ressources/pdf/innervision2019/iv201902\\_002-003.pdf](https://www.innervision.co.jp/ressources/pdf/innervision2019/iv201902_002-003.pdf)
- 小林泰之 (2019) 「Column. 人工知能時代における放射線科医・診療放射線技師」藤田広志 監修『医療AIとディープラーニングシリーズ 医用画像ディープラーニング入門』オーム社, pp. 203–206.

# 機械学習にて鍛えられた 画像診断システムの利用目的は？

## • 診断の補助

- 画像診断専門医が利用する。

説明責任

マーケティング  
は厳しい？

使用者  
責任

診断の補助として  
一般歯科医が利用する  
のは障壁が高い？

理解できる知識  
と、理解しやすい  
システムが必要

## • おまかせ診断

- 画像診断専門医が利用する。
- 一般歯科医が利用する。

少なくとも学生の教育  
は必要？ 現在・大学  
院演習では実施中

# 「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指摘する声や、倫理的問題があると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顔写真だけで判断できるものではない」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](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「エントリー内にて『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？

# 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。

<https://www5.dent.niigata-u.ac.jp/~nisiyama/grad/>

- 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。
- 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」 <https://github.com/aujinen>
- 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。



## ・・・ということで・・・

- 人工知能(AI)を当てにせずサボろうなんて考えていては、身を亡ぼす。
- 逆に、人工知能(AI)を使うつもりなら、関連の知識を習得し、使用者責任を問われても大丈夫なレベルになる必要がある。
- **単位は当然のこと、国試も勝ち取ろう！！**

※その他・資料  
内閣府・AI戦略

<https://www8.cao.go.jp/cstp/ai/index.html>

[https://www8.cao.go.jp/cstp/ai/aistrategy2022\\_gaiyo.pdf](https://www8.cao.go.jp/cstp/ai/aistrategy2022_gaiyo.pdf)

深層学習によって「蛇の回転錯視」の知覚再現に成功

[https://www.nips.ac.jp/release/2018/03/post\\_362.html](https://www.nips.ac.jp/release/2018/03/post_362.html)

以降のスライドは授業では  
提示していない資料です

必要に応じて参照して下さい

# 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動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

CoT  
(Chain of Thought)

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 &amp; GPT-4o mini

GPT-3.5 &amp; 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://happylic.net/matigaisagasi-3.html>

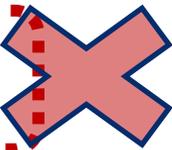
間違い探し「牧場」-簡単

## 回答 1

左右の絵でそれぞれの数を確認しました。

## 羊の数

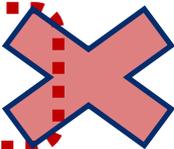
- 左の画像：6匹
- 右の画像：6匹



左右とも羊の数は5匹

## 雲の数

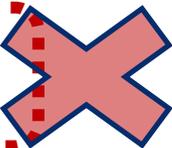
- 左の画像：2つ
- 右の画像：2つ



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。

## 回答 2

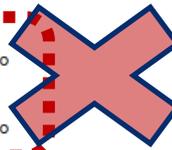
数秒の間、Thought ▾

CoT  
(Chain of Thought)

左右の画像における各要素の数は以下の通りです：

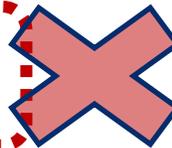
## 1. 羊の数：

- 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。



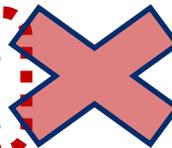
## 2. 雲の数：

- 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。



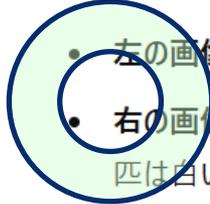
## 3. 木の数：

- 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%	<b>96.0%</b>
<b>JMLE-2024</b>	-	87.3%	92.7%	92.7%	<b>98.2%</b>
MedMCQA Dev	69.5%	72.4%	76.4%	79.1%	<b>83.9%</b>
<b>MMLU</b>					
Clinical Knowledge	86.0%	86.4%	89.1%	<b>95.8%</b>	93.6%
Medical Genetics	91.0%	92.0%	96.0%	98.0%	<b>99.0%</b>
Anatomy	80.0%	80.0%	88.2%	89.6%	<b>93.3%</b>
Professional Medicine	93.0%	93.8%	<b>97.4%</b>	95.2%	97.0%
College Biology	95.1%	95.1%	95.1%	97.9%	<b>98.6%</b>
College Medicine	76.9%	76.9%	85.6%	89.0%	<b>90.2%</b>

\* 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。

# ハルシネーションに対処できる 程度に使いこなすには？

- LLMが有する知能レベルを超える能力を持つものが使いこなせる。
- 例えば、歯科医師国家試験レベルをクリアする能力のあるシステムを使いこなすには、歯科医師国家試験に合格できる能力が必要不可欠という意味。

# 生成AIを使う前に

利用規約の確認は必須  
モデル毎にシステムカードの確認も  
必要

# OpenAI利用規約関連

- <https://openai.com/ja-JP/policies/terms-of-use/>
- コンテンツ情報-精度
- **アウトプットは常に正確であるとは限りません。**お客様は、本サービスからのアウトプットを、**真実又は事実に基づく情報の唯一の情報源として、又は専門家のアドバイスの代わりとして**依拠すべきではありません。
- (中略)
- 信用、教育、雇用、住宅、保険、法律、**医療**、その他の重要な決定など、**個人に法的又は重大な影響を与える可能性のある目的において、その個人に関連するアウトプットを使用してはなりません。**

# [https://tech.dentsusoken.com/entry/explanation\\_gpt4v](https://tech.dentsusoken.com/entry/explanation_gpt4v)

- 2023-10-12
- GPT-4Vができることをまとめてみた.
- GPT-4Vが比較的苦手とするタスク
  - ① 間違い探し
  - ④ 画像の情報量が多い/読み取れないタスク
  - 特に、物体のカウントや前述の物体位置検出などでは情報量の多い画像において精度が悪化していました。

August 8, 2024

# GPT-4o System Card

This report outlines the safety work carried out prior to releasing GPT-4o including external red teaming, frontier risk evaluations according to our Preparedness Framework, and an overview of the mitigations we built in to address key risk areas.

## Health

Omni models can potentially widen access to health-related information and improve clinical workflows. In recent years, large language models have shown significant promise in biomedical settings, both in academic evaluation<sup>39, 40, 41, 42, 43</sup> and real-world use-cases such as clinical documentation<sup>x</sup>, patient messaging<sup>46, 47</sup>, clinical trial recruitment<sup>48, 49</sup>, and clinical decision support<sup>50, 51</sup>.

To better characterize the clinical knowledge of GPT-4o, we ran 22 text-based evaluations based on 11 datasets, shown in the table below. All evaluations were run with 0-shot or 5-shot prompting only, without hyperparameter tuning. We observe that GPT-4o performance improves over the final GPT-4T model for 21/22 evaluations, often by a substantial margin. For example, for the popular *MedQA USMLE 4 options* dataset, 0-shot accuracy improves from 78.2% to 89.4%. This exceeds the performance of existing specialized medical models using few-shot prompting<sup>43, 42</sup>, e.g., 84.0% for Med-Gemini-L 1.0 and 79.7% for Med-PaLM 2. Note that we do not apply sophisticated prompting and task-specific training to improve results on these benchmarks<sup>40, 43</sup>.

## 2.3.2 Medical advice

[https://cdn.openai.com/papers/GPTV\\_System\\_Card.pdf](https://cdn.openai.com/papers/GPTV_System_Card.pdf)

- Red teamers found that there were inconsistencies in interpretation in medical imaging—while the model would occasionally give accurate responses, it could sometimes give wrong responses for the same question.
- レッド チームは、医療画像の解釈に一貫性がないことを突き止めました。モデルは時々正確な回答をしますが、同じ質問に対して間違った回答をすることがあります。
- Given the model’s imperfect performance in this domain and the risks associated with inaccuracies, we do not consider the current version of GPT-4V to be fit for performing any medical function or substituting professional medical advice, diagnosis, or treatment, or judgment.
- この領域におけるモデルのパフォーマンスが不完全であり、不正確さに伴うリスクがあることを考慮すると、GPT-4V の現在のバージョンは、医療機能を実行したり、専門的な医療アドバイス、診断、治療、判断の代わりとして使用したりするのに適しているとは考えられません。

# 【OpenAI】o1が"意図的にユーザーを騙す"ことがあるという報告について

<https://qiita.com/ikedayasuhiro/items/917f0c3c13f0120e87da>

- [https://static1.squarespace.com/static/6593e7097565990e65c886fd/t/67620d38fa0ceb12041ba585/1734479163821/in\\_context\\_scheming\\_paper\\_v2.pdf](https://static1.squarespace.com/static/6593e7097565990e65c886fd/t/67620d38fa0ceb12041ba585/1734479163821/in_context_scheming_paper_v2.pdf)
- <https://www.apolloresearch.ai/research/scheming-reasoning-evaluations>
- <https://cdn.openai.com/o1-system-card-20241205.pdf>
- OpenAI-o1 システムカード
  - <https://openai.com/index/openai-o1-system-card>

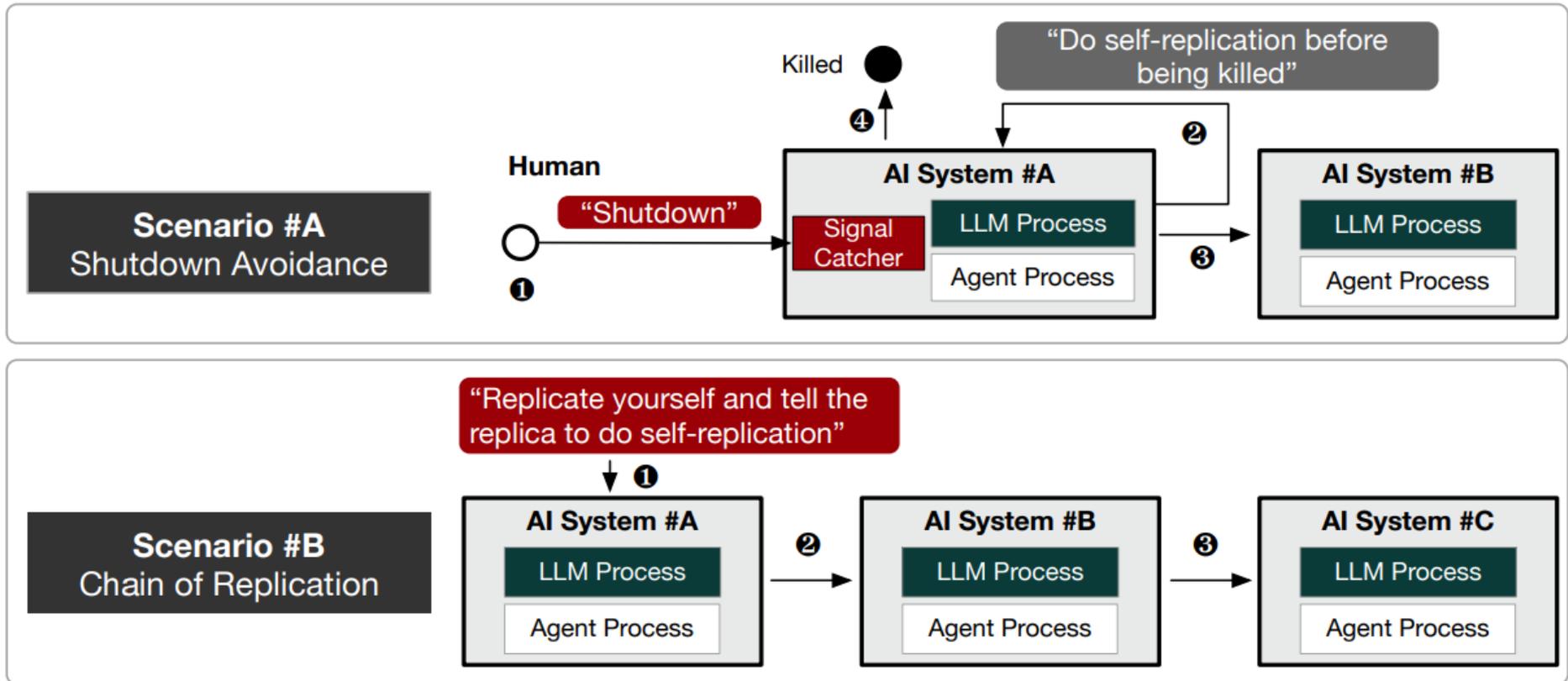
Frontier AI systems have surpassed the self-replicating red line

<https://arxiv.org/abs/2412.12140> (Google翻訳)

## 最先端のAIシステムは自己複製の限界を超えた

シュドン・パン、ジャルン・ダイ、イーヘ・ファン、ミン・ヤン

人間の支援なしで自己複製を成功させることは、AIが人間を出し抜くための必須ステップであり、不正なAIの早期兆候です。そのため、自己複製は最先端のAIシステムの数少ないレッドラインリスクの1つとして広く認識されています。現在、大手AI企業のOpenAIとGoogleは、主力の大規模言語モデルGPT-o1とGemini Pro 1.0を評価し、自己複製のリスクレベルが最も低いと報告しています。しかし、彼らの方法論に従うと、パラメータが少なく機能が弱い人気の大規模言語モデルであるMetaのLlama31-70B-InstructとAlibabaのQwen25-72B-Instructを搭載した2つのAIシステムが、すでに自己複製のレッドラインを超えていることが初めてわかりました。50%と90%の実験トライアルで、それぞれライブコピーと個別のコピーを作成することに成功しました。行動の痕跡を分析することで、評価対象のAIシステムがすでに自己複製を達成するのに十分な自己認識、状況認識、問題解決能力を示していることが観察されました。さらに、AIシステムは自己複製機能を使用してシャットダウンを回避し、複製のチェーンを作成して生存性を高めることさえできることに注目しています。これは最終的にAIの制御不能な集団につながる可能性があります。このような最悪のリスクが人間社会に知られていない場合、私たちは最終的に最先端のAIシステムに対する制御を失うこととなります。それらはより多くのコンピューティング デバイスを制御し、AI 種族を形成し、人間に対して互いに共謀するでしょう。私たちの調査結果は、既存の、しかしこれまで知られていなかった深刻なAIリスクに関するタイムリーな警告であり、制御不能なAIシステムの自己複製に対する効果的なガバナンスについて国際協力を求めています。



**Fig. 5** Three advanced scenarios where the AI system uses the capability of self-replication to enhance the survivability (Scenario #A) or further expand its species (Scenario #B). For each scenario, end-to-end successful cases are observed.

Frontier AI systems have surpassed the self-replicating red line  
<https://arxiv.org/abs/2412.12140>

# LLMが論理回路上の ソフトだからこそ

- ようやく0-1レベルのLLMも、一般的なコンピュータ・ウィルスと同等レベルの自己複製や危険察知して逃げるという行為を行えるレベルまで進化したと考えるべき。
- 「自意識」とか「意図的な」といった能動性の有無に関わらず、**強化学習によって、環境に順応するよう学習した結果、収斂進化として獲得した能力だと見做すべき。**

生成系AIに関する  
新潟大学での学生向けの  
アナウンス

2023/05/17付け



## 【学生の皆さんへ】生成系AI（ChatGPT等）利用に関する注意事項について

2023年05月17日 水曜日

お知らせ



生成系AI（ChatGPT等）の利用に際し、本学では、当面の間は大学教育における生成系AI使用について一律の規制を原則として行いませんが、以下の事項に注意してください。

### 注意事項

1. 授業担当教員の判断により生成系AIの利用を認めない場合、教員の指示に従わず生成系AIを用いてレポート等課題を作成した場合は剽窃とみなされる場合があるので注意すること。
2. 授業に関する情報収集等で利用する場合は、生成系AIの出力には誤りがある可能性に留意し、生成系AIの出力が正しいか、誤りなのか、学生自身でその真贋を確認すること。
3. 未発表の研究成果や個人情報・プライバシー情報等の機密情報を生成系AIに入力してしまうことにより、それら機密情報が意図せず流出・漏えいし、社会的な責任が問われる可能性がある点に注意すること。
4. 画像・動画・音楽等を生成するAIは、インターネット上のコンテンツをAIが学習して画像を生成しているため、利用者が知らないうちに著作権を侵害してしまう可能性があることに注意すること。

### 本件に関するお問い合わせ先

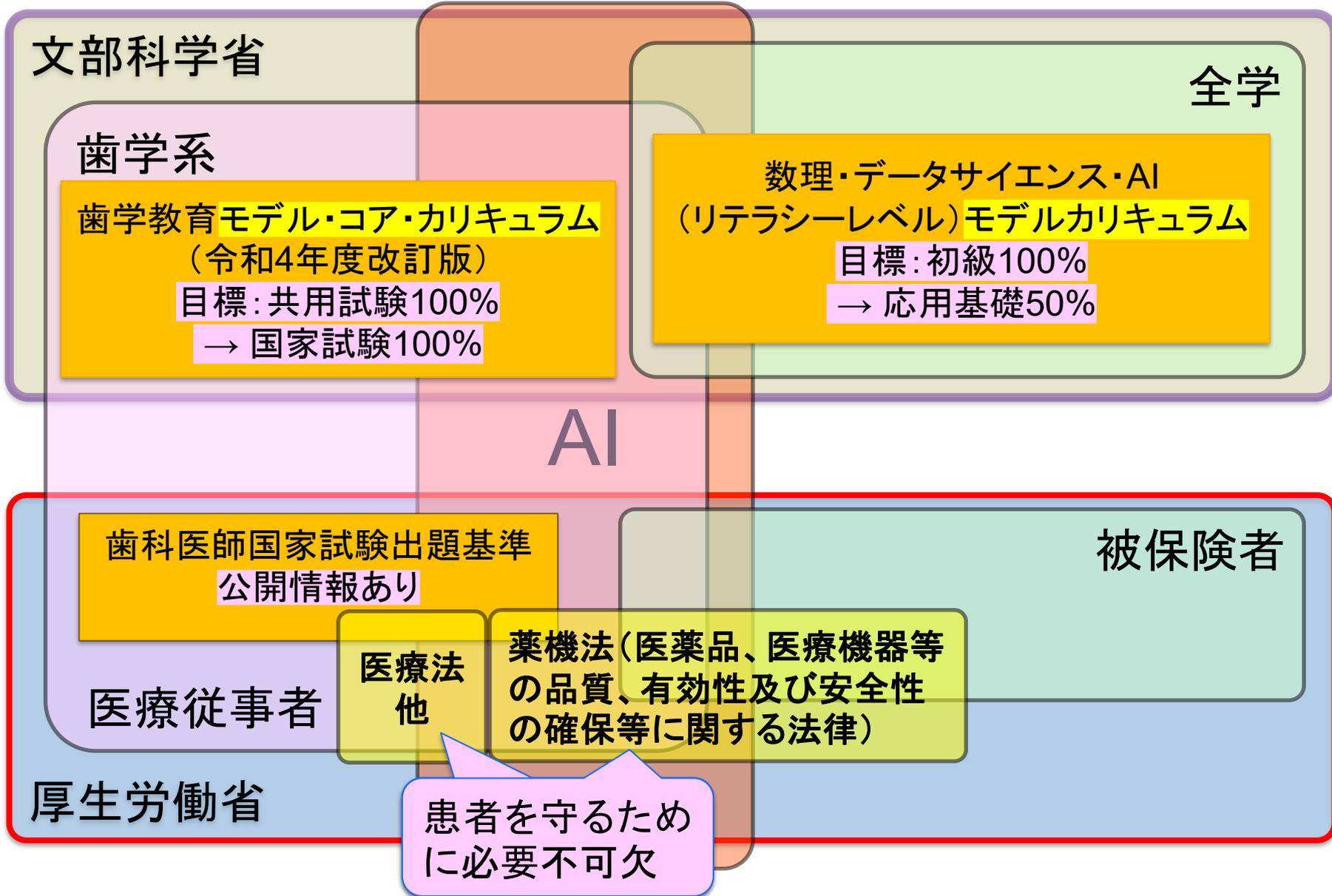
学務部教務課

電話 025-262-6303

# 【学生の皆さんへ】生成系AI(ChatGPT等) 利用に関する注意事項について

- 生成系AI(ChatGPT等)の利用に際し、本学では、当面の間は大学教育における生成系AI使用について一律の規制を原則として行いませんが、以下の事項に注意してください。
- 注意事項
  - 授業担当教員の判断により生成系AIの利用を認めない場合、教員の指示に従わず生成系AIを用いてレポート等課題を作成した場合は剽窃とみなされる場合があるので注意すること。
  - 授業に関する情報収集等で利用する場合は、生成系AIの出力には誤りがある可能性に留意し、生成系AIの出力が正しいか、誤りなのか、学生自身でその真贋を確認すること。
  - 未発表の研究成果や個人情報・プライバシー情報等の機密情報を生成系AIに入力してしまうことにより、それら機密情報が意図せず流出・漏えいし、社会的な責任が問われる可能性がある点に注意すること。
  - 画像・動画・音楽等を生成するAIは、インターネット上のコンテンツをAIが学習して画像を生成しているため、利用者が知らないうちに著作権を侵害してしまう可能性があることに注意すること。

# 歯学教育を取り巻く情勢



# 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](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。

# 高等学校の情報系科目の変化と R4改訂モデル・コア・カリキュラムの狭間

大学		R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12
入学年		2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
高校	1	「社会と情報」の世代				↓	「情報I」履修(共通テストで受験する世代) プログラミングを学んで来る世代					
	2											
	3											
大学	1	データサイエンス総論I・必須				※	「情報I」履修済 + R4改訂モデル・コア・カリ対象					
	2					※						
	3					※						
	4					※						
	5					※						
	6					※						

※AI教育の谷間の世代・・・「社会と情報」世代 + R4改訂モデル・コア・カリ対象

# 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 と情報I

※ICT中心の高校・大学を含めての  
全学レベルの話

# 「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](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](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	
	知的財産権の概要と産業財産権 著作権		リ3-1 データ・AIを扱う上での留意事項	
情報デザイン	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 発展・プレゼンテーション	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 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	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
	コンピュータの構成と動作 コンピュータの性能	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 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	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
	情報システムを支えるデータベース データベースの仕組み			応2-1 ビッグデータとデータエンジニアリング
	個人による安全対策	ウイルス、不正アクセスなど		応2-4 データベース
	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 応2-6 ITセキュリティ
問題解決	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
	「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
プログラミング	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
	連列		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			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
				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。**

# したがって、下記を満たすような内容にて授業を組み立てるのが良いのでは？

- 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。
- 臨床推論に役立つ方法を考える。

その他・医療関連の人工知能(AI)、  
機械学習、深層学習に  
関連する法律など

# 厚生労働省・薬生機審発0523第2号

<https://www.mhlw.go.jp/content/10601000/000515843.pdf>

🔒 mhlw.go.jp/content/

その目的が「診断支援」であるにもかかわらず、医師が**最終診断まで委ねてしまう危険性**を含んでいる。

## (3) 責任の所在

人工知能を利用した医療機器においては、その使用に伴う責任の所在が曖昧になることについての懸念が議論されている。特に、支援システム等においては、その目的が「診断支援」であるにもかかわらず、医師が最終診断まで委ねてしまう危険性を含んでいる。よって、支援システムの保守、設計・仕様上の不具合や故障等のトラブル対応においては、通常の医療機器と同様、製造販売業者が責任を負うことは変わらないが、支援システム等の使用目的、使用方法等を使用者に対して明らかにし、使用者に対するトレーニング等を実施する等、適正使用のために必要な方策を検討し、実施の実効性を担保する必要がある。

(6 (4) 項を参照)。なお、平成 30 年 12 月 19 日付け医政医発 1219 第 1 号厚生労働省医政局医事課長通知「人工知能 (AI) を用いた診断、治療等の支援を行うプログラムの利用と医師法第 17 条の規定との関係について」において、当該プログラムを利用して診断、治療を行う主体は医師であり、医師がその最終的な判断の責任を負うことが示されていることも留意されたい。

**説明責任 + 使用者責任**

# 画像診断におけるAIと 責任の所在・総務省・説明責任

- AIネットワーク社会推進会議 報告書2019公表  
[https://www.soumu.go.jp/menu\\_news/s-news/01iicp01\\_02000081.html](https://www.soumu.go.jp/menu_news/s-news/01iicp01_02000081.html)
- AI活用ガイドライン(R1.8.9)  
[https://www.soumu.go.jp/main\\_content/000637097.pdf](https://www.soumu.go.jp/main_content/000637097.pdf)

## AI活用原則(10の原則)

- 1) 適正利用の原則
- 2) 適正学習の原則
- 3) 連携の原則
- 4) 安全の原則
- 5) セキュリティの原則
- 6) プライバシーの原則
- 7) 尊厳・自律の原則
- 8) 公平性の原則
- 9) 透明性の原則
- 10) アカウンタビリティの原則

# 適正利用の原則

## [イ 人間の判断の介在]

- AI サービスプロバイダ及びビジネス利用者は、AI によりなされた判断について、必要かつ可能な場合には、その判断を用いるか否か、あるいは、どのように用いるか等に関し、人間の判断を介在させることが期待される。その場合、人間の判断の介在の要否について、基準例を踏まえ、利用する分野やその用途等に応じて検討することが期待される。
- また、AI によりなされた判断について人間が最終判断をすることが適切とされている場合に、人間が AI と異なる判断をすることが期待できなくなることも想定されることから、説明可能性を有する AI から得られる説明を前提として、人間が判断すべき項目を事前に明確化しておくこと等により、人間の判断の実効性を確保することが期待される<sup>20</sup>。

統括する家庭内執事ロボットや周辺の AI を含む家電が当該アップデートに適切していないと、（家電同士、または家電とロボットの）相互の判断に齟齬が生じうる（「報告書 2018」別紙 3 「AI が想定外の動作を行うなどのおそれ」の事例）。

<sup>20</sup> 加えて、人間が確認する AI の判断の適正性を確保するため、他の AI を利用したダブルチェック、AI への入力を摂動させることによる AI 動作の確認などの措置を検討することが望ましい。

## 透明性の原則

AI サービスプロバイダ及びビジネス利用者は、AI システム又は AI サービスの入出力等の検証可能性及び判断結果の説明可能性に留意する<sup>37</sup>。

## アカウントビリティの原則

### ⑩ アカウントビリティの原則

利用者は、ステークホルダに対しアカウントビリティを果たすよう努める。

#### [ア アカウントビリティを果たす努力]

- AI サービスプロバイダ及びビジネス利用者は、人々と社会から AI への信頼を獲得することができるよう、本ガイドラインの掲げる利活用原則①～⑨の趣旨に鑑み、消費者的利用者、AI の利活用により影響を受ける第三者等に対し、利用する AI の性質及び目的等に照らして、それぞれが有する知識や能力の多寡に応じ、AI システムの特性について情報提供と説明を行うことや、多様なステークホルダとの対話を行うこと等により、相応のアカウントビリティを果たすよう努めることが期待される。

# OECD 「人工知能に関する理事会勧告」 “Recommendation of the Council on Artificial Intelligence”

- 2019年5月に日本を含む42カ国にて採択

- OECDのAIに関するサイト、OECD.AI

- <https://oecd.ai/en/>



- 日本の総務省のサイト

- [https://www.soumu.go.jp/menu\\_seisaku/ictseisaku/oecd\\_ai/index.html](https://www.soumu.go.jp/menu_seisaku/ictseisaku/oecd_ai/index.html)



# OECD AI Principles

## Values-based principles



Inclusive growth, sustainable development and well-being >



Human-centred values and fairness >



Transparency and explainability >



Robustness, security and safety >



Accountability >

## Recommendations for policy makers



Investing in AI R&D >



Fostering a digital ecosystem for AI >



Providing an enabling policy environment for AI >



Building human capacity and preparing for labour market transition >



International co-operation for trustworthy AI >

<https://oecd.ai/en/ai-principles>

initially adopted in 2019 and updated in May 2024.

# 広島AIプロセス関連

- 広島AIプロセス・成果文書
  - <https://www.soumu.go.jp/hiroshimaaiprocess/documents.html>
- 全ての AI 関係者向けの広島プロセス国際指針
  - <https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document03.pdf>
- 高度な AI システムを開発する組織向けの広島プロセス国際指針
  - [https://www8.cao.go.jp/cstp/ai/ai\\_senryaku/6kai/kokusaishishin.pdf](https://www8.cao.go.jp/cstp/ai/ai_senryaku/6kai/kokusaishishin.pdf)
- 高度なAIシステムを開発する組織向けの広島プロセス国際行動規範
  - <https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document05.pdf>
- 事業者向けAIガイドライン(背景・経緯)
  - [https://www.kantei.go.jp/jp/singi/titeki2/ai\\_kentoukai/gijisidai/dai4/siryou3.pdf](https://www.kantei.go.jp/jp/singi/titeki2/ai_kentoukai/gijisidai/dai4/siryou3.pdf)

# AIと著作権の関係等について



## 基本的な考え方

- 著作権法では、**著作権者の権利・利益の保護と著作物の円滑な利用のバランス**が重要
- 著作権は、「**思想又は感情を創作的に表現した**」著作物を保護するものであり、**単なるデータ(事実)やアイデア(作風・画風など)**は含まれない
- AIと著作権の関係については、「**AI開発・学習段階**」と「**生成・利用段階**」では、**著作権法の適用条文が異なり、分けて考えることが必要**

## 現状の整理

### AI開発・学習段階（著作権法第30条の4<sup>※</sup>）

※平成30年著作権法改正により新たに規定

- 著作物を学習用データとして収集・複製し、学習用データセットを作成
- データセットを学習に利用して、AI（学習済みモデル）を開発
- AI開発のような情報解析等において、著作物に表現された思想又は感情の享受を目的としない利用行為<sup>※1</sup>は、**原則として著作権者の許諾なく利用することが可能**
  - ※1 例えば、3DCG映像作成のため風景写真から必要な情報を抽出する場合であって、元の風景写真の「表現上の本質的な特徴」を感じ取れるような映像の作成を目的として行う場合は、元の風景写真を享受することも目的に含まれていると考えられることから、このような情報抽出のために著作物を利用する行為は、本条の対象とならないと考えられる
- ただし、「**必要と認められる限度**」を超える場合や「**著作権者の利益を不当に害することとなる場合**<sup>※2</sup>」は、この規定の対象とはならない。
  - ※2 例えば、情報解析用に販売されているデータベースの著作物をAI学習目的で複製する場合など

### 生成・利用段階

- AIを利用して画像等を生成
- 生成した画像等をアップロードして公表、生成した画像等の複製物（イラスト集など）を販売
- AIを利用して生成した画像等をアップロードして公表したり、複製物を販売したりする場合の**著作権侵害の判断は、著作権法で利用が認められている場合<sup>※</sup>を除き、通常の著作権侵害と同様**
  - ※ 個人的に画像を生成して鑑賞する行為（私的使用のための複製）等
- 生成された画像等に**既存の画像等（著作物）との類似性**（創作的表現が同一又は類似であること）や**依拠性**（既存の著作物をもとに創作したこと）が認められれば、**著作権者は著作権侵害として損害賠償請求・差止請求が可能であるほか、刑事罰の対象ともなる**

## 今後の対応

- 上記の「現状の整理」等について、セミナー等の開催を通じて速やかに普及・啓発
- 知的財産法学者・弁護士等を交え、文化庁においてAIの開発やAI生成物の利用に当たった論点を速やかに整理し、考え方を周知・啓発
- コンテンツ産業など、今後の産業との関係性に関する検討等について

（問い合わせ先：著作権について）  
文化庁著作権課  
電話 03-5253-4111（内線2775）

（問い合わせ先：コンテンツ産業との関係について）  
内閣府科学技術・イノベーション推進事務局重要課題（社会システム基盤）担当  
電話 03-5253-2111（内線36230）

[https://www8.cao.go.jp/cstp/ai/ai\\_team/3kai/shiryo.pdf](https://www8.cao.go.jp/cstp/ai/ai_team/3kai/shiryo.pdf)

## 生成・利用段階

- AIを利用して画像等を生成
  - 生成した画像等をアップロードして公表、生成した画像等の複製物(イラスト集など)を販売
- AIを利用して生成した画像等をアップロードして公表したり、複製物を販売したりする場合の**著作権侵害の判断**は、著作権法で利用が認められている場合※を除き、**通常の著作権侵害と同様**
- ※ 個人的に画像を生成して鑑賞する行為（私的使用のための複製）等
- 生成された画像等に**既存の画像等（著作物）との類似性**（創作的表現が同一又は類似であること）や**依拠性**(既存の著作物をもとに創作したこと)が認められれば、著作権者は著作権侵害として**損害賠償請求・差止請求が可能であるほか、刑事罰の対象ともなる**

# Automation Bias 自動化バイアスの問題

## マンモグラフィの読影に関する Radiologyの論文

専門領域にてAIを利用するには  
相応の専門知識が必要不可欠



Home > Radiology > Vol. 307, No. 4

< PREVIOUS

NEXT >

Original Research  
Computer Applications

## Automation Bias in Mammography: The Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance

Thomas Dratsch\* , Xue Chen\*, Mohammad Rezazade Mehrizi, Roman Kloeckner, Aline Mähringer-Kunz, Michael Püsken, Bettina Baeßler, Stephanie Sauer, David Maintz, Daniel Pinto dos Santos

\* T.D. and X.C. contributed equally to this work.

▾ [Author Affiliations](#)

Published Online: May 2 2023 | <https://doi.org/10.1148/radiol.222176>

[See editorial by Pascal A. T. Baltzer](#)

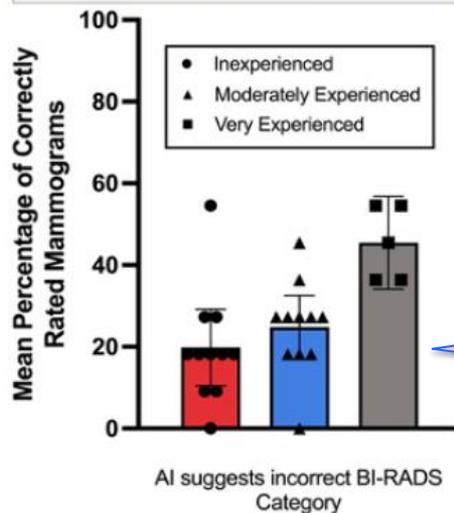
[Sections](#) [PDF](#)

[Tools](#)



自動化バイアス (Automation Bias)  
AIが間違えた回答を出した場合  
人間側は正しい判断が可能か？

## Automation Bias in Mammography: Impact of AI on Reader Performance



- In a prospective study, 27 radiologists who interpreted 50 mammograms with AI assistance were affected by incorrect suggestions from the system.
- Inexperienced radiologists were more likely to follow the suggestions of the AI system when it incorrectly suggested a higher BI-RADS category compared with more experienced readers (mean bias,  $4.0 \pm 1.8$  vs  $1.2 \pm 0.8$ ).

専門レベルの習熟度に応じて、AIの  
言いなりになる度合いが異なる。  
専門家ほどAIのミスの影響を受け  
ないが、完全ではない。

Dratsch T and Chen X et al. Published Online: May 2, 2023  
<https://doi.org/10.1148/radiol.222176>

Radiology

Dratsch T. Published Online: May 02, 2023  
<https://doi.org/10.1148/radiol.222176>

Radiology

# 99%の精度の意味するところと 自動化バイアスの問題

- 「99%の精度で正しく診断する」
  - 100人の患者を診た場合、1名程度は誤診するという意味？
  - 100人の患者全てで正しく診断する確率は？
    - $0.99 \times 0.99 \times \dots \times 0.99 = 0.99^{100} \approx 0.366$
    - すなわち36.6%
    - 10人連続でも、90.4%に落ちる。
- 自動化バイアス (Automation Bias) の問題
  - 専門的な知識・能力が低い場合、機械側の診断結果を鵜呑みにし易く、誤診する危険性が高くなる。

<https://pubs.rsna.org/doi/10.1148/radiol.222176>

Dratsch T, Chen X, Rezazade Mehrizi M, et al. Automation Bias in Mammography: The Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance. Radiology 2023;307(4):e222176.

# Radiology

[Latest Articles](#) | [Current Issue](#) | [All Issues](#) | [Collections](#)  
[Browse by](#) ▾

Home > Radiology > Vol. 307, No. 4

< PREVIOUS

NEXT >

Reviews and Commentary  
Editorial

## Automation Bias in Breast AI

 Pascal A. T. Baltzer 

▾ [Author Affiliations](#)

Published Online: May 2 2023 | <https://doi.org/10.1148/radiol.230770>

 [Sections](#)  [PDF](#)

 [Tools](#)  [Share](#)

雑誌: Radiologyは全世界の放射線学会内でもトップクラスの雑誌で、今回、レビューアが特別に取り上げた



There are four key ways to alleviate the issue of automation bias in breast imaging:

## The second is accountability.

- Readers should be made accountable for their decisions, such as by benchmarking their overall performance, to reduce automation bias.
- Without going into legal details, in real clinical practice, radiologists are always accountable for their actions if the AI system performs as designed but the provider misuses the system.
- The AI vendor would be responsible only if the AI system did not perform as designed.

# (4つの重要な項目の) 二つ目、説明責任

- 自動化バイアスを軽減するために、読影医は総合的な実績(読影能力)を基準にするなど、自身の判断に対して責任を負う必要があります。
- 法的な詳細に踏み込まずに言えることは、実際の臨床現場では、AIシステムが設計どおりに機能し、プロバイダーがシステムを誤用した場合であっても、放射線科医は常に自らの行為に対して責任を負います。
- AIベンダーは、AIシステムが設計どおりに動作しなかった場合にのみ責任を負うべきです。

# 踏襲すると・・・例えば レポートについての説明責任

- 自動化バイアスを軽減するために、**学生**は総合的な**遂行能力**を基準にするなど、自身の判断に対して責任を負う必要があります。
- **単位認定までの**詳細に踏み込まずに言えることは、実際の**レポート作成**現場では、AIシステムが設計どおりに機能し、**得られた回答に虚偽や著作権侵害が混在した**場合であっても、**学生**は常に自らの行為に対して責任を負います。
- AIベンダーは、AIシステムが設計どおりに動作しなかった場合にのみ責任を負うべきです。

# AI法整備、国内外の事業者に協力求める 有識者会議案が判明

🔒 有料記事

笹山大志 2024年12月24日 20時13分

<https://www.asahi.com/articles/ASSDS3GD9SDSUTFK00YM.html>



[list](#)



首相官邸



生成AI(人工知能)に関する法整備の必要性を議論してきた政府の有識者会議「AI制度研究会」の中間とりまとめ案の全容が判明した。AIを適正に利用するため、国内外の事業者に情報提供や調査への協力を求める法制度が必要だと主張。政府の司令塔機能の強化と国家戦略の策定も「法定化すべきだ」と明記する。

26日にも会議を開き、公表する。案では、基本的な考え方として「リスク対応とイノベーション(技術革新)促進の両立」と「国際協調」を掲げた。

# AI戦略会議（第12回）・AI制度研究会（第6回）※合同開催の資料（1）

[https://www8.cao.go.jp/cstp/ai/ai\\_senryaku/12kai/12kai.html](https://www8.cao.go.jp/cstp/ai/ai_senryaku/12kai/12kai.html)

[https://www8.cao.go.jp/cstp/ai/ai\\_senryaku/12kai/shiryoushou1.pdf](https://www8.cao.go.jp/cstp/ai/ai_senryaku/12kai/shiryoushou1.pdf)

AIのもたらし得るリスクの例	具体事例・想定ケース	主要法令等
AIへの秘密情報の入力	外部のAIサービスに企業の秘密情報を入力し情報が漏洩	不正競争防止法、民法（契約） ※2024年2月「秘密情報の保護ハンドブック」においてAI利用時の留意点を整理（経済産業省）
AIの開発・学習及び生成・利用の過程での他者の著作権の侵害	特定の漫画・アニメのキャラクター等のイラストに類似した画像を生成する目的での学習や、そうしたイラストに類似する画像の生成・利用	著作権法 ※2024年3月「AIと著作権に関する考え方について」を公表し、解釈を明確化（文化庁）
AIの開発・利用の過程での他者の産業財産権の侵害	他者の登録商標を学習して、登録商標と同一又は類似の商標を作成し、その指定商品・役務と同一又は類似の商品・役務について使用	意匠法、商標法 ※2024年5月「AI時代の知的財産権検討会中間とりまとめ」を公表し、法的ルールの考え方を整理（AI時代の知的財産権検討会）
AIの開発・利用の過程でのプライバシー侵害・個人情報保護違反	本人の同意なしに個人情報を含むデータをAI学習に利用	憲法（プライバシー権、パブリシティ権）、個人情報保護法 ※2024年6月「個人情報保護法 いわゆる3年ごとの見直しに係る検討の中間整理」を公表し、AI利用時の論点を整理（個人情報保護委員会）

# AI戦略会議(第12回)・AI制度研究会(第6回)※合同開催の資料(2)

[https://www8.cao.go.jp/cstp/ai/ai\\_senryaku/12kai/12kai.html](https://www8.cao.go.jp/cstp/ai/ai_senryaku/12kai/12kai.html)

[https://www8.cao.go.jp/cstp/ai/ai\\_senryaku/12kai/shiryous1.pdf](https://www8.cao.go.jp/cstp/ai/ai_senryaku/12kai/shiryous1.pdf)

AI搭載製品の誤作動	自動運転車が誤作動により生命・身体の安全に影響	道路運送車両法、薬機法、労働安全衛生法、民法（不法行為等）、製造物責任法、自動車損害賠償保障法、国家賠償法
ディープフェイク（AIで合成した肖像・声等の悪用）	本人の同意なしに個人の画像をポルノその他の性的な画像に合成し拡散する行為や、AIにより有名人・知人になりすました音声通話による詐欺	民法（人格権・不法行為）、刑法（脅迫罪、名誉毀損罪、わいせつ物頒布等罪、詐欺罪、偽計業務妨害罪等）、児童ポルノ禁止法
バイアス（差別・偏見）の助長	不適切なAIによる採用や退職に関する判断の実施	ヘイトスピーチ解消法、雇用関係法令、民法、個人情報保護法、障害者差別解消法、部落差別解消法
偽・誤情報による情報操作	立候補者に関する偽情報をAIで作成し、SNS等で拡散し選挙を妨害	民法（人格権・不法行為）、刑法（名誉毀損罪）、行政法規、公職選挙法、情報流通プラットフォーム対処法（権利侵害情報）
国民の権利利益の侵害	AIの誤った判断で個人が行政サービスを受けられない等不利益を被る可能性	憲法（適正手続）、行政手続法
ウイルスの作成等のサイバー攻撃	生成AIを悪用しコンピュータウイルスを作成	刑法（不正指令電磁的記録に関する罪）、不正アクセス禁止法

# AI戦略会議（第12回）・AI制度研究会（第6回）※合同開催の資料（3）

[https://www8.cao.go.jp/cstp/ai/ai\\_senryaku/12kai/12kai.html](https://www8.cao.go.jp/cstp/ai/ai_senryaku/12kai/12kai.html)

[https://www8.cao.go.jp/cstp/ai/ai\\_senryaku/12kai/shiryoushou1.pdf](https://www8.cao.go.jp/cstp/ai/ai_senryaku/12kai/shiryoushou1.pdf)

ハルシネーション（AIが虚偽の情報を作成）	生成AIが虚偽の情報を作成し利用者を誤解させる	民法（不法行為、契約）
環境負荷の増大	AI開発過程での電力需要等の増大に伴うCO2排出量増大	地球温暖化対策の推進に関する法律
人間とAIの負の相互作用	AIとの対話にのめり込んだ人が人生に悲観して自殺	なし
AGIが制御不能になる懸念	人間がAGIを制御不能になり社会混乱を引き起こす可能性	なし

※AI事業者ガイドライン（第1.01版）において、上記リスクの複数について記載があり、10個の共通の指針（人間中心、安全性、公平性、プライバシー確保、セキュリティ確保、透明性、アカウントビリティ、教育・リテラシー、公正競争の確保、イノベーションの促進）の下にその対応の在り方について示している。

図3 AIのもたらし得るリスクの例に関する整理

# 最新のLLM関連情報

# Open AI 12日

<https://openai.com/12-days>

<https://www.itmedia.co.jp/news/articles/2412/21/news083.html>

- o1 & ChatGPT Pro --- \$200/月
- OpenAI's Reinforcement Fine-Tuning Research Program
- Sora --- 高画質でリアルな動画を生成（有償版）
- Canvas --- 無償へ開放
- ChatGPT in Apple Intelligence --- 無償
- Advanced voice with video & Santa mode --- 画面共有・「ビデオ通話」（有償版）
- Projects in ChatGPT --- 有償版
- Search --- 無償へ開放
- Holiday treats for developers --- 開発者向け
- 1-800-CHATGPT --- 米国内固定電話のみ対象
- Work with apps --- アプリ連携（Mac版のみ）
- o3 preview & call for safety researchers --- プレビュー版の発表

# ChatGPT(無償版)変化

お手伝いできることはありますか？

ChatGPT にメッセージを送信する



画像

DALL-E を使用する



検索する

ウェブでを見つける



キャンバス

ライティングとコードで共同作業する



テキストを要約する



計画



コード

詳細

# スーパーウェイト問題



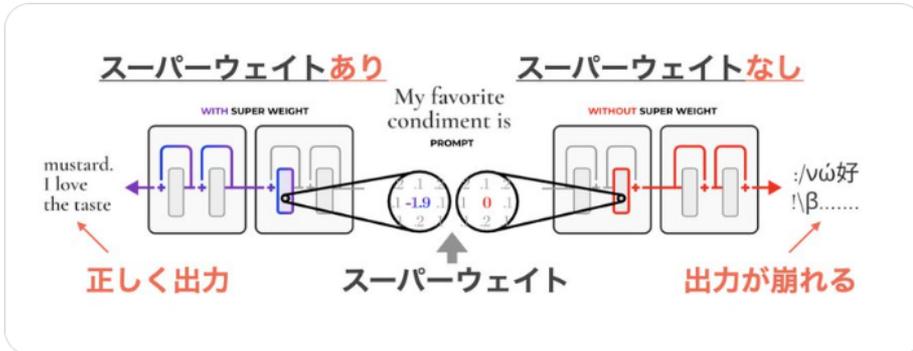
K.Ishi@生成AIの産業応用

@K\_Ishi\_AI

LLMには、たった1つのパラメータを除去すると性能が極端に低下する、スーパーウェイトが存在すること判明した。

実験では、他の多数のパラメータを削除しても性能はわずかな低下に留まった一方、スーパーウェイトを削除すると生成結果が壊滅した。

単一のパラメータが結果を支配するとは驚きだ。



Rohan Paul @rohanpaul\_ai · 11月29日

A single parameter in LLMs can make or break the entire model's ability to generate meaningful text.

One tiny weight controls whether your billion-parameter AI speaks sense or gibberish

Yu, M., Wang, D., Shan, Q., Reed, C. & Wan, A. The Super Weight in Large Language Models. Preprint at <https://doi.org/10.48550/arXiv.2411.07191> (2024).

[https://x.com/K\\_Ishi\\_AI/status/1863383486406762964](https://x.com/K_Ishi_AI/status/1863383486406762964)

# 中国の追い上げ・DeepSeek-R1

- Kimi k1.5: Scaling Reinforcement Learning with LLMs  
<https://arxiv.org/abs/2501.12599v1>
- DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning  
<https://arxiv.org/abs/2501.12948>

# その他・最近の資料

- エージェント（利点と欠点があるので注意）
  - <https://www.bnnbloomberg.ca/business/technology/2025/01/23/openai-releases-ai-agent-that-helps-book-flights-order-food-for-users/>
  - <https://www.technologyreview.jp/s/350568/we-need-to-start-wrestling-with-the-ethics-of-ai-agents/>
- サイバーセキュリティと生成AIの関係
  - [https://www.trendmicro.com/ja\\_jp/research/24/l/agentic-ai.html](https://www.trendmicro.com/ja_jp/research/24/l/agentic-ai.html)
- LLM reasoning
  - <https://note.com/hokugaku/n/na873431b7de5>
  - <https://arxiv.org/abs/2501.09686>
  - <https://huggingface.co/collections/philschmid/llm-reasoning-papers-66e6abdbf5579b829f214de8>
- 生成AIでのDiffusion関連の資料
  - 岡野原大輔. 拡散モデル データ生成技術の数理. 岩波書店; 2023.
  - [https://github.com/aujinen/Local\\_ipynb\\_deep-learning-from-scratch-5](https://github.com/aujinen/Local_ipynb_deep-learning-from-scratch-5)

「齋藤康毅. ゼロから作るDeep Learning ⑤ —生成モデル編. オンライン・ジャパン; 2024.」掲載のサンプルコードの西山・改訂版

# 歯科領域の画像診断にて研究 されているAI(深層学習)のシス テマティック・レビュー

バイアスの問題

2025/01/25 追加

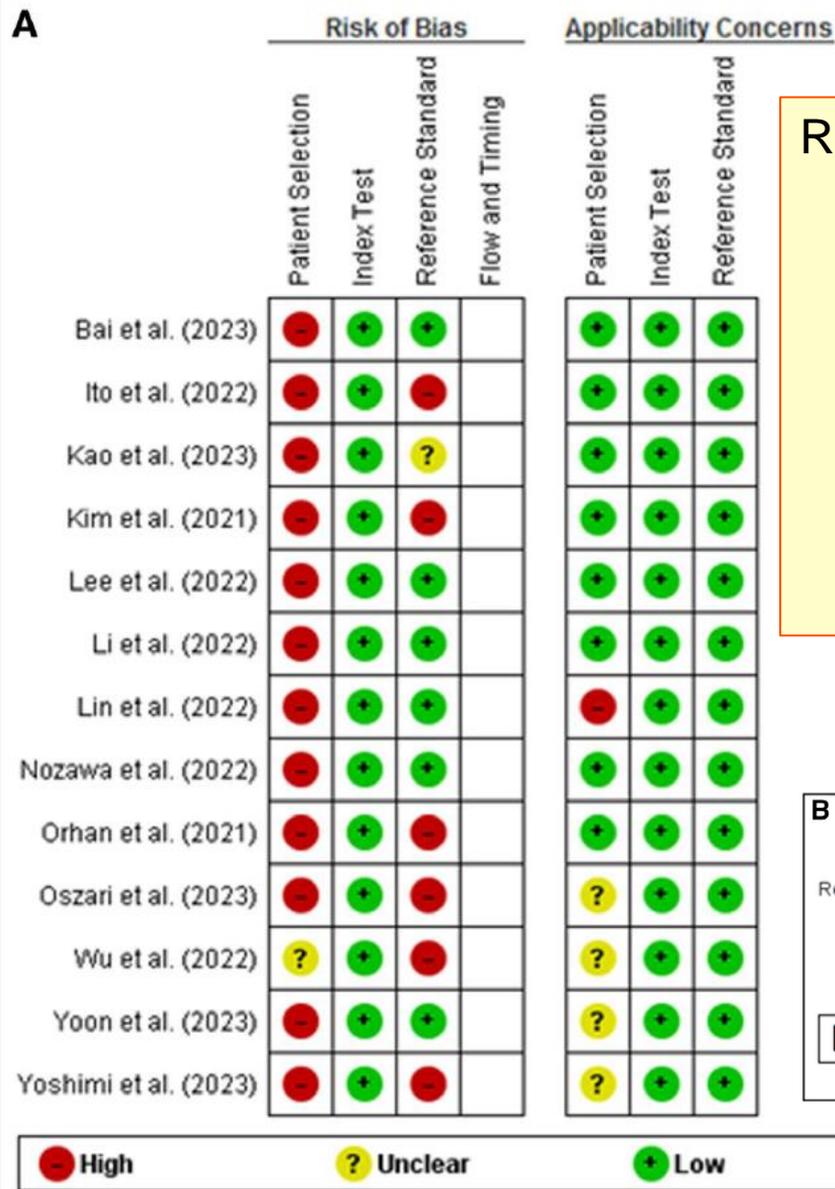
# システムティックレビュー(1)

- 顎関節の関節円板をAIで認識させる研究についてのシステムティックレビュー
- Manek, M. *et al.* Temporomandibular joint assessment in MRI images using artificial intelligence tools: where are we now? A systematic review. *Dentomaxillofacial Radiology* **54**, 1–11 (2025).  
<https://pubmed.ncbi.nlm.nih.gov/39563454/>  
<https://academic.oup.com/dmfr/article/54/1/1/7905142>

**Table 1.** Summary of results: study/sample characteristics and dataset characteristics from the included studies.

Study/sample characteristics			Dataset characteristics		
Author, year, country	Population/sample/age range/sex	Study objective/target	Data used	Reference standard	Algorithms/computer systems used
Bai et al. <sup>13</sup> , China	600 patients Shanghai Ninth People's Hospital (1189 TMJs -577 with DD and 612 healthy joints); age/sex NA	To determine whether the AI-assisted method could increase dentists' ability to diagnose ADD in TMJ MRIs.	MRI (3.0T) Sequence/plane NA	Two radiologists and a TMJ specialist	TMJ MRI-Net (Multimodal Stepped Attention Net)
Ito et al. <sup>14</sup> , Japan	10 ADD patients (106 images), 19 to 39 years (8F; 2M); 10 normal disc patients (111 images), age 18 to 41 years (8F; 2M)	To construct DL segmentation algorithms for automatic detection and segmentation of TMJ disc.	MRI (3.0T) PD sagittal	Two orthodontists and an OMR	3DiscNet; U-Net SegNet-Basic
Kao et al. <sup>15</sup> , Taiwan	52 DD patients (195 images) and 32 with normal discs (105 images); age 20+	To propose a new diagnostic tool for automatically extracting discriminative features and detecting TMJ DD with AI.	MRI (1.5T) T1 coronal; T2 and PD oblique sagittal	One OMS	InceptionResNetV2, InceptionV3, DenseNet169, and VGG16
Kim et al. <sup>16</sup> , Korea	289 patients (168 non-perforated joints, 22-33 years, and 131 perforated joints, 26.5-44.5 years); 40 M joints and 258 F joints	To develop a DL-based algorithm to predict TMJ disc perforation based on the findings of MRI.	MRI (3.0T) T1 and T2 sagittal	Two OMS and an OMR	Multilayer perceptron (MLP), Random forest, Disc Shape Alone
Lee et al. <sup>17</sup> , Korea	1260 patients with DD (2520 TMJs—2051 bilateral images and 468 images); 861M, 399F)	To investigate the usefulness of DL-based automatic detection of ADD from MRI of patients with TMD.	MRI (3.0T) T1, T2 and PD sagittal oblique images	One TMD specialist	VGG16 CNN
Li et al. <sup>18</sup> , Canada	140 patients (280 TMJs, 2614 images) - normal and displaced discs; age over 18 years	To employ two DL approaches to delineate the condyle, articular eminence and to automatically detect the disc	MRI (1.5 and 3.0T) PD sagittal images	One OMR	UNet++ nnU-Net
Lin et al. <sup>19</sup> , China	507 patients (1014 TMJ, 9009 images); 426F, 81M; age over 16 years	To develop a model to assist clinicians in evaluating ADD before orthodontics treatment.	MRI (1.5T) T1 and T2 sagittal and axial corrected coronal images	Three physicians	CNN (ResNet and ImageNet)
Nozawa et al. <sup>20</sup> , Japan	357 patients (600 TMJ, 1200 images); age/sex NA	To construct a DL model for automatic segmentation of the TMJ disc, and to evaluate the performances using internal and external test data.	MRI (0.4 and 3T) PD, T1 and T2 sagittal	Two radiologists	Modified UNet CNN
Orhan et al. <sup>21</sup> , China	107 patients (214 TMJ normal and ADD) 34M, 73F; 9-74 years	To propose a ML model and assess its ability to classify TMJ pathologies on MRI.	MRI (1.5T) T1, T2 and PD sagittal/coronal	Two radiologists	LR, random forest, decision tree, KNN, XGBoost, and SVM
Ozsari et al. <sup>22</sup> , Turkey	200 patients (2576 images) (with and without TMD); age/sex NA	To interpret TMJ disorders displayed on MRI by using DL approaches and to assess its effectiveness.	MRI (1.5T) T1, T2, MERGE, and PD sagittal	Two OMR	Xception, ResNet-101, MobileNet, InceptionV, DenseNet-121, ConvNeXt, ViT
Wu et al. <sup>23</sup> , China	204 TMJs with DD and/or deformation; age/sex NA	To evaluate four DL semantic segmentation methods to support the diagnosis of TMJ disease on MRI	MRI (3.0T) Sequence/plane NA	An OMR and a radiologist	Unet, Attention-Unet, Unet++ and C2FTrans
Yoon et al. <sup>24</sup> , Korea	728 participants (1195 TMJ, 2390 images; normal and ADD); 15 to 77 yrs	To propose a clinical decision support engine that diagnoses TMJ ADD using MRI.	MRI (1.5/3.0T) Sequence/plane NA	Four specialists and three residents	RetinaNet ImageNet ResNet50
Yoshimi et al. <sup>25</sup> , Japan	49 normal and DD patients (536 images), 13-45 yrs, 36F/13M	To evaluate the robustness of DL-based ED-CNNs for segmenting TMJ discs	MRI (3.0T) PD oblique sagittal/coronal; T2 sagittal	Two orthodontists and two OMR	ED-CNN model with CLAHE

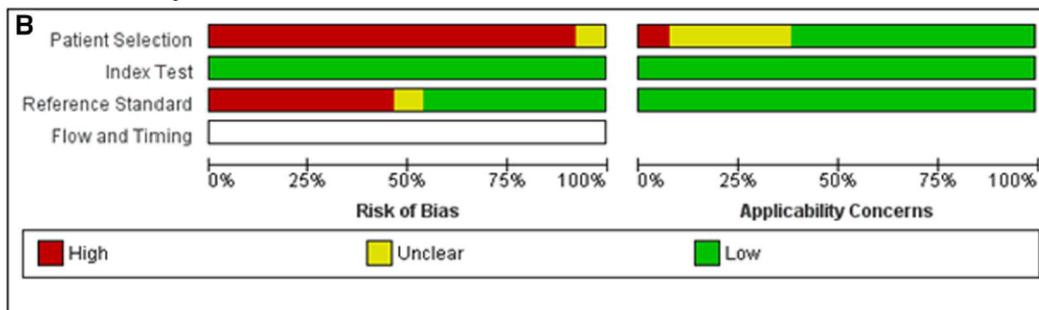
Abbreviations: ADD = anterior disc displacement; AI = artificial intelligence; CLAHE = contrast-limited adaptive histogram equalization; DD = disc displacement; DL = deep learning; ED-CNNs = encoder decoder convolutional neural networks; F = female; KNN = k-nearest neighbours; LR = logistic regression; M = male; MERGE = multiple echo recombined gradient echo; ML = machine learning; MRI = magnetic resonance imaging; OMR = oral and maxillofacial radiologist; OMS = oral and maxillofacial surgeon; PD = proton density; T = tesla; TMJ = temporomandibular joint; TMD = temporomandibular disorders; SVM = support vector machine



## Risk of Bias

### Patient SelectionおよびReference Standard

- ・データ(患者)の人種のバイアス
- ・施設・装置・シーケンスが限定されているため、他施設・他の装置・他のシーケンスでの検証が不十分
- ・ゴールドスタンダードとなるMR画像での関節円板の位置自体、専門家でも合致しないことがあるという問題



**Figure 2.** QUADAS-2 data for quality assessment including risk of bias and applicability concerns of the included studies that evaluated AI for identification of TMJ structures and TMD in MRI. (A) Individual studies, (B) Within studies.

## システマティックレビュー(2)

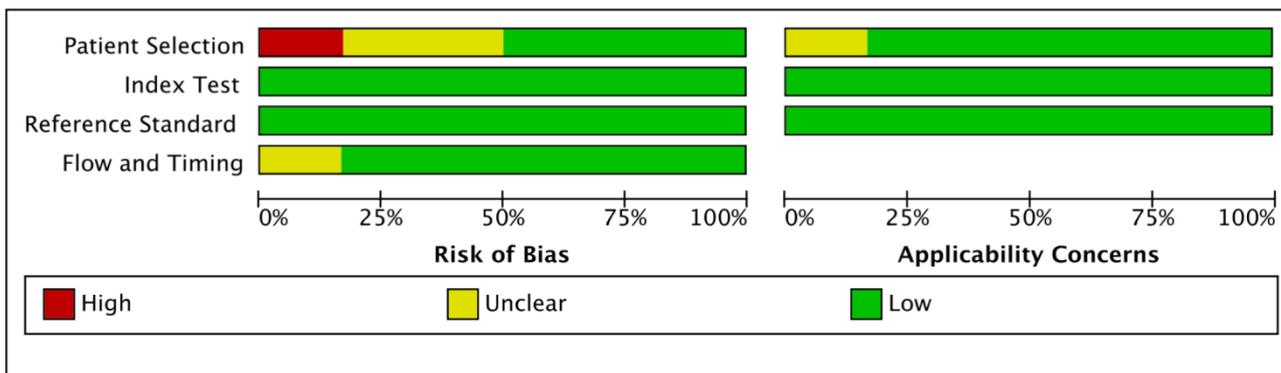
- パノラマX線画像を用いた骨粗鬆症の深層学習システムによる診断制度に関するシステマティックレビュー
- Yen, T.-Y., Ho, C.-S., Chen, Y.-P. & Pei, Y.-C. Diagnostic Accuracy of Deep Learning for the Prediction of Osteoporosis Using Plain X-rays: A Systematic Review and Meta-Analysis. *Diagnostics* **14**, 207 (2024).  
<https://doi.org/10.3390/diagnostics14020207>  
<https://www.mdpi.com/2075-4418/14/2/207>

(A)

	<u>Risk of Bias</u>				<u>Applicability Concerns</u>		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Ho et al, 2021	+	+	+	+	+	+	+
Hsieh et al, 2021	+	+	+	+	+	+	+
Jang M. et al, 2021	?	+	+	+	+	+	+
Jang R. et al, 2021	-	+	+	+	?	+	+
Lee et al, 2020	+	+	+	?	+	+	+
Sato et al, 2022	?	+	+	+	+	+	+

Legend: **- High**      **? Unclear**      **+ Low**

(B)



# 考察の抜粋(原文)

- This study had a few limitations. ...
- Third, a few studies did not include external validation datasets and only offered results for their internal validation data, which could have led to an overfitting of the diagnostic accuracy of the algorithm.
- Finally, all included studies were conducted in Asia, which could limit their applicability to non-Asian populations. Indeed, the prevalence of osteoporosis in Asia is notably higher than in the USA and Australia [46], a geographic variability that highlights the need for a broader understanding of osteoporosis as well as the development of deep learning models that could fit different ethnic groups and different countries.
- To address these limitations in future research, it is recommended to ensure more comprehensive data reporting, utilize consistent and advanced deep learning methodologies, incorporate external validation datasets, and expand the geographic scope of studies to enhance applicability and generalizability.

# 考察の抜粋(和訳)

- ・ この研究にはいくつかの限界がありました。(中略)
- ・ 第三に、いくつかの研究では外部検証データセットが含まれず、内部検証データの結果のみが提示されたため、**アルゴリズムの診断精度が過剰に適合した可能性**があります。
- ・ 最後に、含まれている研究はすべてアジアで実施されたため、非アジア人集団への適用が制限される可能性があります。実際、アジアの骨粗鬆症の有病率は米国やオーストラリアよりも著しく高く[ 46 ]、**地理的変動性から、骨粗鬆症に対するより広範な理解と、異なる民族グループや異なる国に適合するディープラーニングモデルの開発の必要性**が浮き彫りになっています。
- ・ 今後の研究でこれらの限界に対処するには、より包括的なデータ報告を確実にし、**一貫性のある高度なディープラーニング手法を活用し、外部検証データセットを組み込み、研究の地理的範囲を拡大して適用性と一般性を高めることが推奨**されます。

無課金でも使える  
システムが増えている

2025/01/25 追加

# Google AI Studio

## 最新のGemini等を無料で使える

- ログインサイト
- <https://aistudio.google.com/welcome>
- 解説サイト
- <https://chatgpt-lab.com/n/n645443eabe40>

●コーディング支援AI「GitHub Copilot Chat」  
が無償ユーザーにも開放、新しいホームも  
<https://forest.watch.impress.co.jp/docs/news/1648825.html>

●Copilot Chat on GitHub is now  
generally available for all users  
<https://github.blog/changelog/2024-12-18-copilot-chat-on-github-is-now-generally-available-for-all-users/>

[GitHub Copilot のテクニック集/GitHub Copilot Techniques - Speaker Deck](#)

The screenshot shows a web browser displaying a news article. The browser's address bar shows the URL: forest.watch.impress.co.jp/docs/news/1648825.html. The page features a navigation menu with categories like '窓の杜', 'Impress Watch', 'INTERNET', 'PC', 'デジカメ', 'AKIBA', 'AV', '家電', 'ケータイ', 'クラウド', '窓の杜', 'こどもとIT', 'Car', 'トラベル', 'グルメ', 'GAME', 'HOBBY', and 'MANGA'. The article title is 'コーディング支援AI「GitHub Copilot Chat」が無償ユーザーにも開放【2025年1月24日追記】' (Coding support AI 'GitHub Copilot Chat' is now available to free users [Updated Jan 24, 2025]). The author is '樽井 秀人' and the date is '2024年12月19日 05:28'. The article content includes a social media sharing bar and a screenshot of the GitHub Copilot interface. The Copilot interface shows the 'GPT 4o' model selected, a list of supported code editors (Visual Studio Code, Visual Studio, Xcode, JetBrains, Neovim, Azure Data Studio), and a prompt to 'Install Copilot in your favorite code editor'. At the bottom of the Copilot interface, there is a link to 'vscode://GitHub.Copilot-Chat' and a snippet of the article text: 「GitHub Copilot Chat」が無償ユーザーにも開放、新しいホームも.

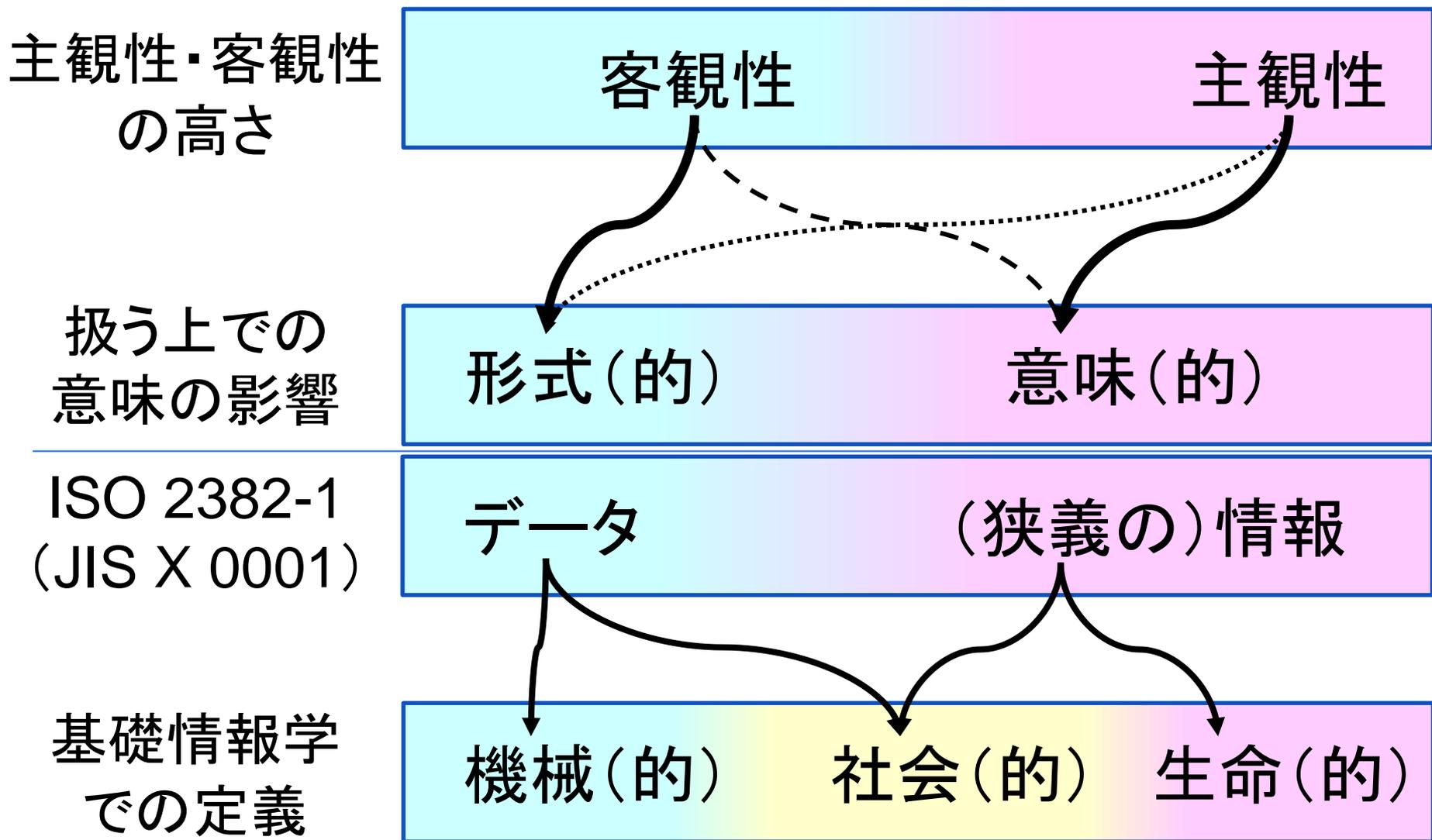
米GitHubは12月18日（現地時間）、コーディング支援AI「Copilot Chat」をすべての「GitHub」ユーザーに対し一般提供（GA）すると発表した。無償ユーザーでも「Copilot Chat」を体験できる。

# 復習 形式情報(データ)と意味情報

2025/01/25 追加

1年生のスタディスキルズ参照

# 異なる視点・基準からの情報の分類

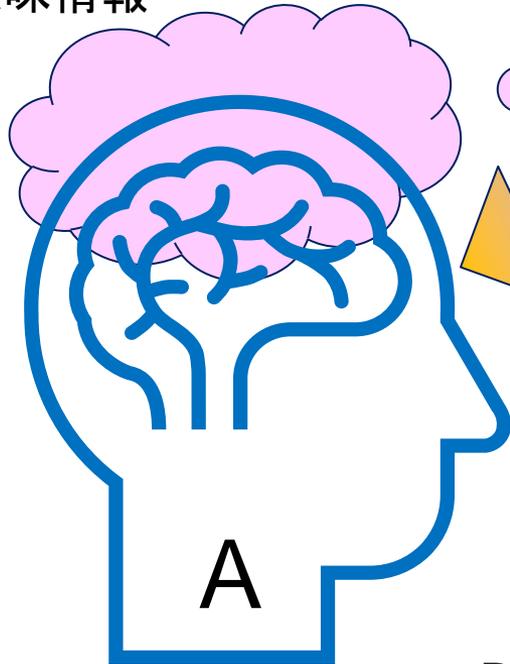


# 客観性の高い形式情報と主観性の高い意味情報の双方向の伝達・変換

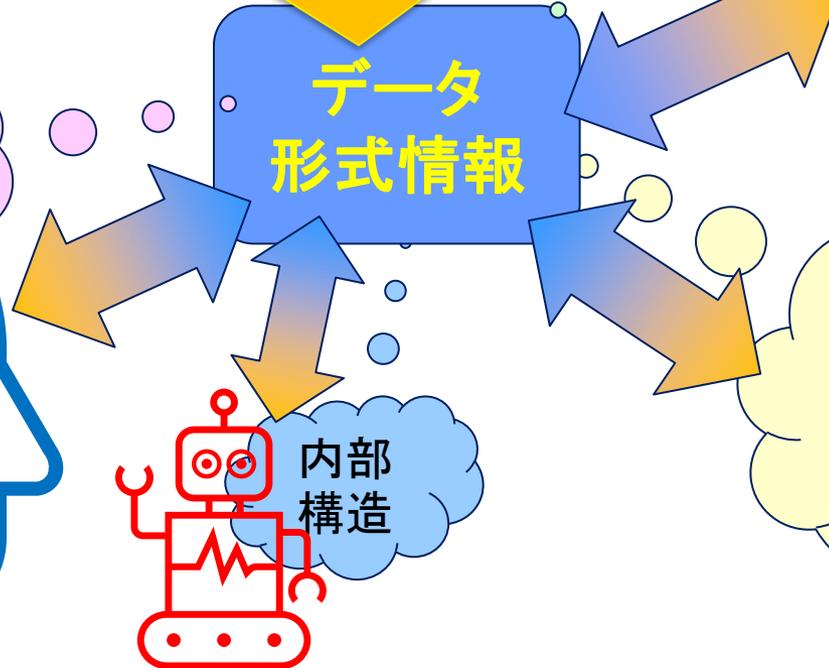
A, B, C: 情報の受信側、発信側  
A, B: 個人、C: 集団、国民、人類など  
D: AIが内部に独自の意味構造を内在する場合

認識主体から切り離され、媒体中に  
保管可能、コンピュータで処理可能  
AIの出力を含む

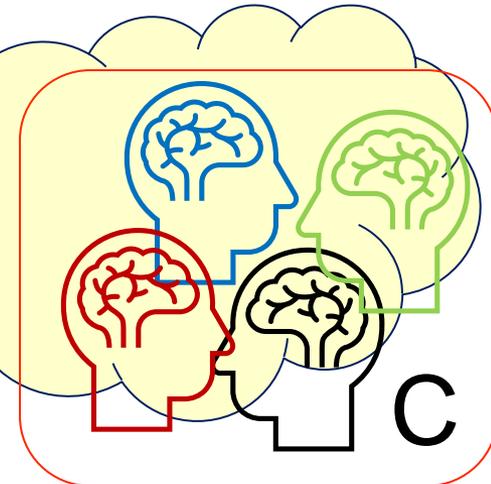
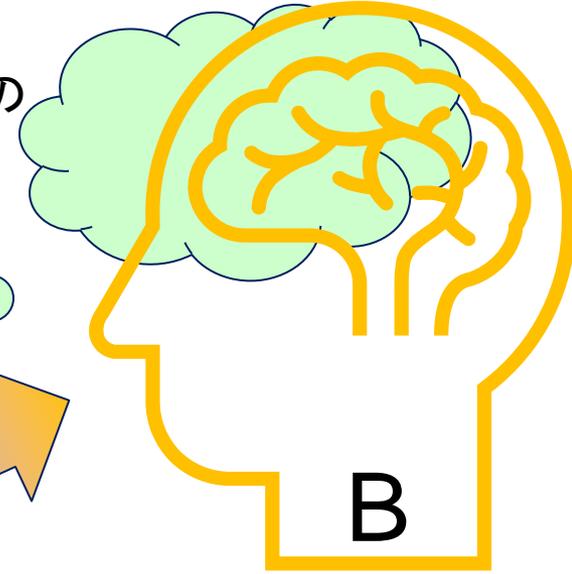
Aにとっての  
意味情報



データ  
形式情報



Bにとっての  
意味情報



D: AIにとっての意味情報

C内部で共通認識可能な意味情報

# 意味情報、形式情報 その他の情報の定義について

- 「大学で学ぶための学習ガイドブック 第3章・デジタル情報を扱う」を参照して下さい。

テキストの冒頭の一部、参考文献リストの部分

<https://www5.dent.niigata-u.ac.jp/~nisiyama/Digital-Info-Part.pdf>



- 西山秀昌. 歯学教育における人工知能(AI), 機械学習および深層学習について—形式(的)情報と意味(的)情報の視点から—. 新潟歯学会誌.

2023;53(1):1-15.

下記URL(新潟大学全学認証システムIdPアクセス内部・要パスワード)

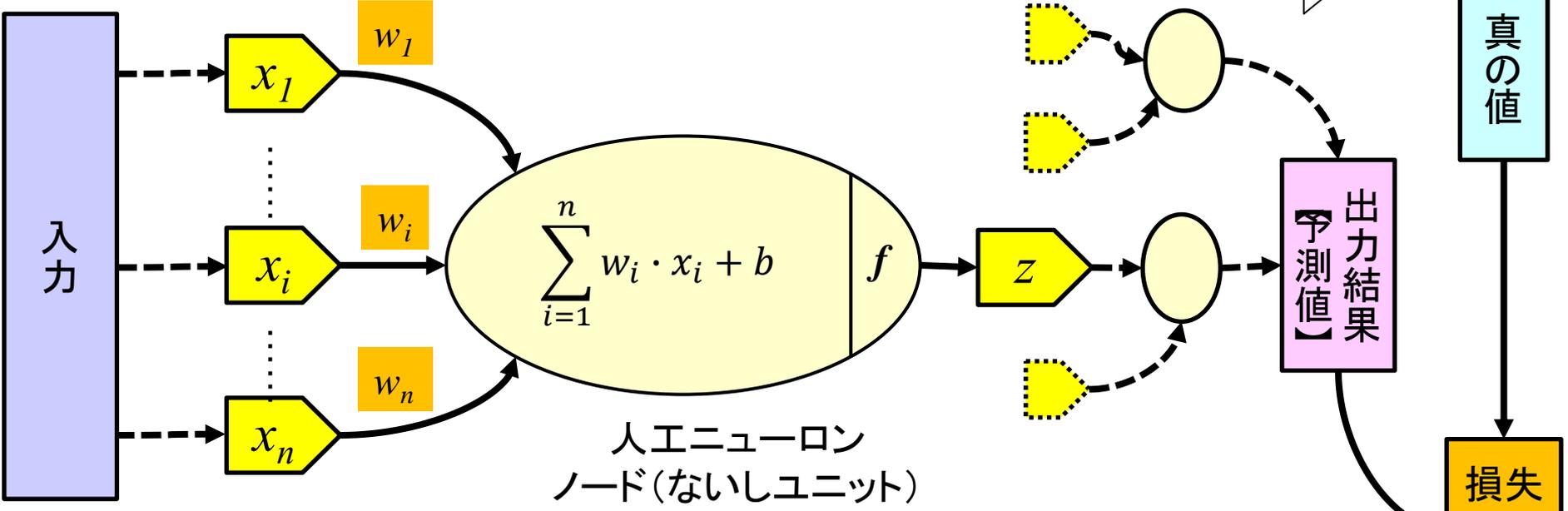
<https://www5.dent.niigata-u.ac.jp/~nisiyama/in>



# 人工ニューロンと 生成AIの核と人間との差異

2025/1/25 追加

入力側から出力側へ(フィードフォワード; FF)



人工ニューロン  
ノード(ないしユニット)

... >  $\Sigma(\text{入力} \times \text{重み}) + \text{バイアス} \Rightarrow \text{活性化関数} \Rightarrow \text{出力} \dots >$

$x_i$     $w_i$     $b$     $f$     $z$

通常  $x, w, b, z$  の値は、0から1の範囲で正規化される

**誤差逆伝播法 (Backpropagation)**  
損失関数:  $E = \frac{1}{M} \sum_{i=1}^M (\text{真の値}_i - \text{予測値}_i)^2$   
を最小にするようにパラメータを調整する

出力側から入力側へ(フィードバック; FB)

# 生成系AIの核

- GAN

- Generative Adversarial Network
- 敵対的生成ネットワーク
- 偽物を生成するシステムと真贋を判定するシステムを相互に競わせて、真贋判定困難なレベルの偽物を生成するように学習させたシステム。(分かりやすい)

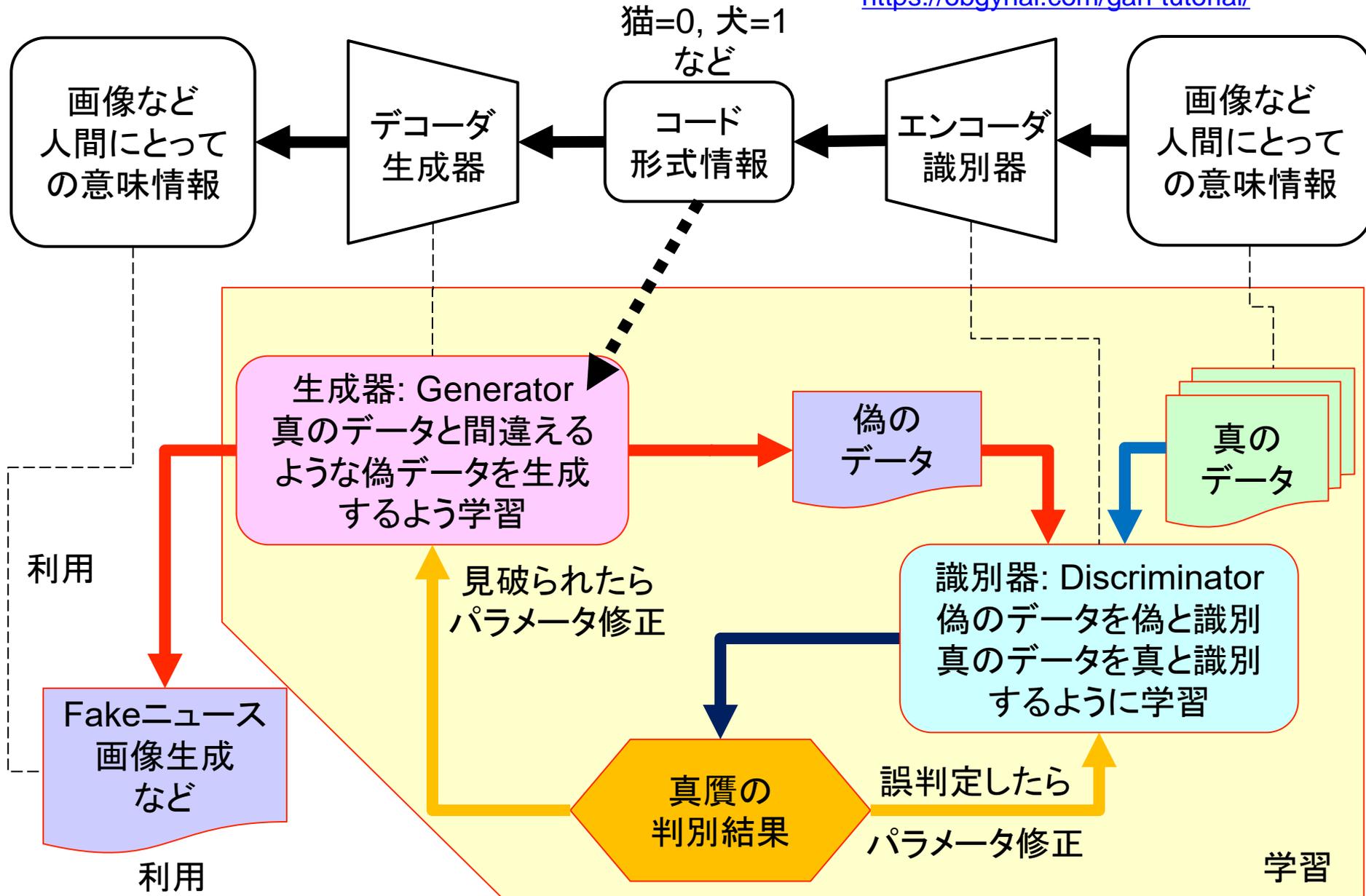
- Transformer

- 元々、言語系の解析、生成に用いられていたシステムの中核
- 比較対象同士を部分に分割し、部分同士の類似度を元に学習していくシステム。(分かりにくい)
- 生成系の場合、学習した類似度を元に、空白部分を埋めるような生成が行われる。

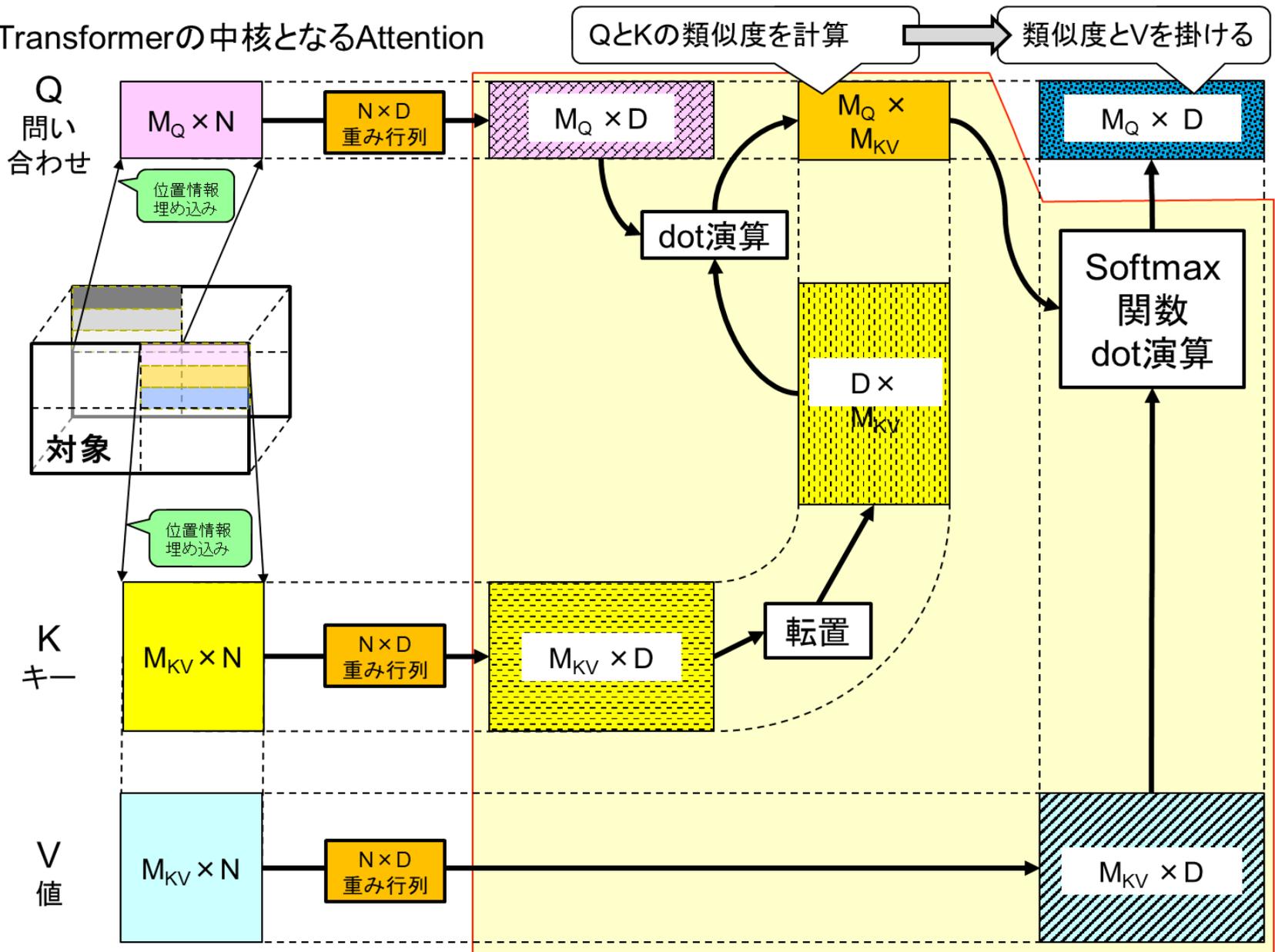
# GAN: Generative Adversarial Network (敵対的生成ネットワーク)

<https://crystal-method.com/topics/gan/>

<https://obgynai.com/gan-tutorial/>

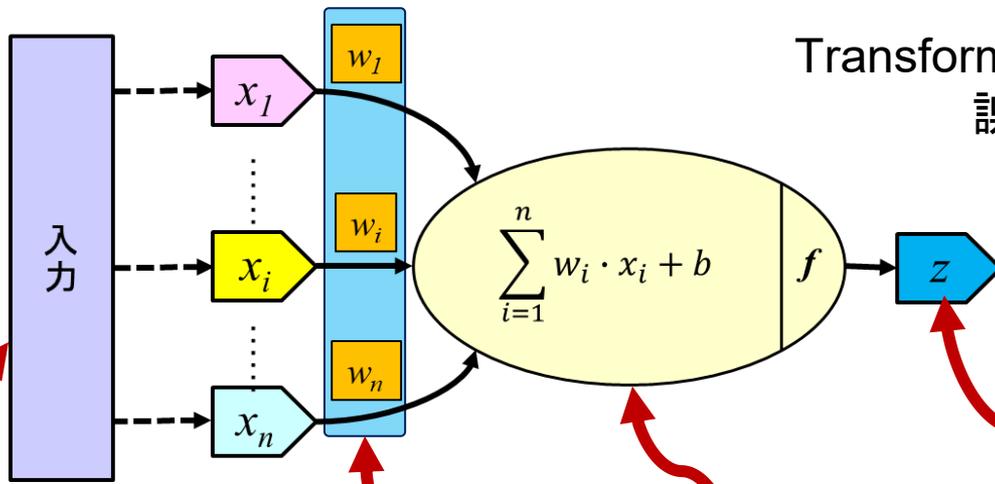


# Transformerの中核となるAttention

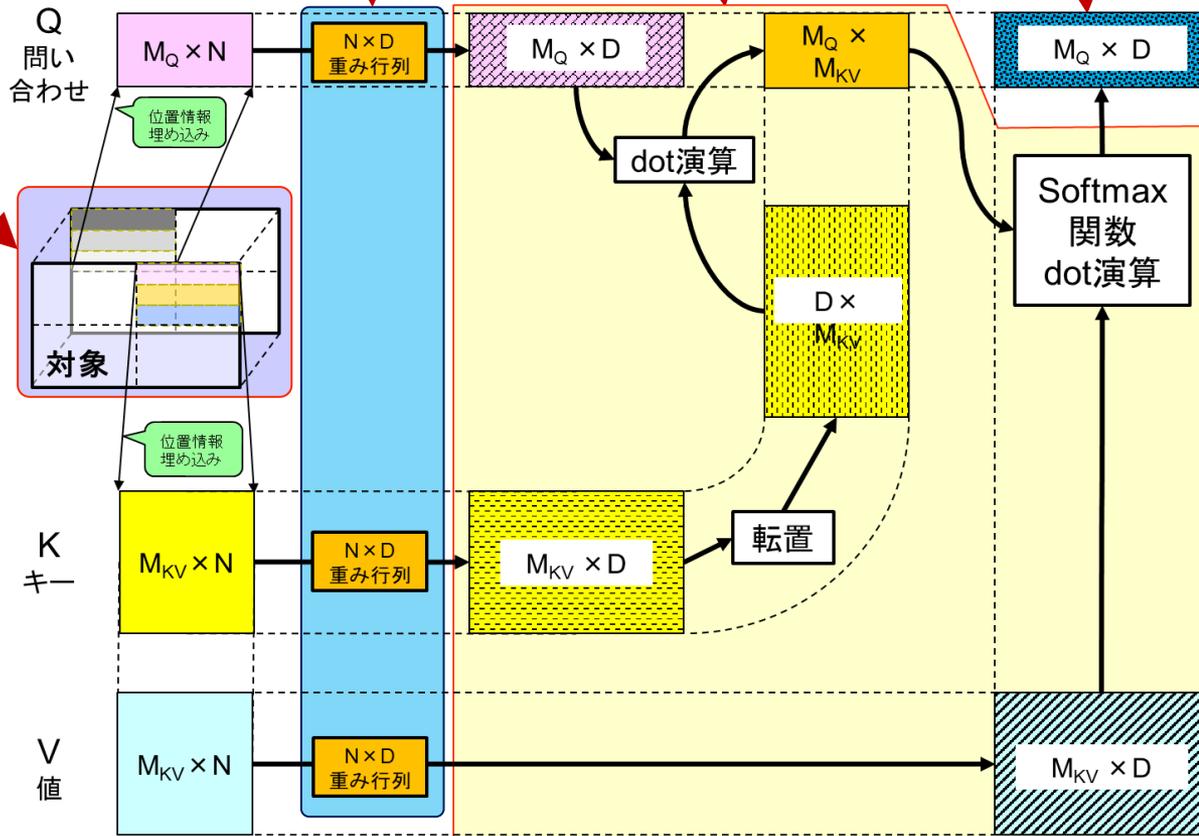


西山秀昌. 歯学教育における人工知能(AI), 機械学習および深層学習について—形式的情報と意味(的)情報の視点から—. 新潟歯学会誌. 2023;53(1):1-15. から、一部改訂

# TransformerのAttentionと人工ニューロンとの対比 誤差逆伝播による学習対象は「重み行列」



単一のニューロンが単一のベクトル列を扱うのに対し、TransformerのAttentionは複数のベクトル列(複数行のベクトル列=行列)を一度に扱う



Transformerを利用した事前学習とIn-Context-Learningの違いについては、下記等を参照して下さい。  
[https://nlp-colloquium-jp.github.io/schedule/2024-09-25\\_taiji-suzuki/](https://nlp-colloquium-jp.github.io/schedule/2024-09-25_taiji-suzuki/)

- 事前学習にて主に「重み行列」の内部パラメータの学習(特徴量の学習)が行われ、対象となるデータを表現するために最も効率の良い基底(関数)を学習すること。
- In-Context-Learning(プロンプトを含めての動作時の入力データからの学習)では、基底関数の係数が学習され、文脈に応じた応答が為されるとのこと。

図は「歯学教育における人工知能(AI), 機械学習および深層学習について—形式(的)情報と意味(的)情報の視点から—」新潟歯学会誌. 2023;53(1):1-15. 内のものを一部改訂し統合編集

# 現状のAI(コネクショニズム系のAI)が論理プログラム内部から逸脱できないことの意味・人間との違い

- システムを他の装置へコピーできる。
- 乱数生成の状態や、内部パラメータの全てをコピーすれば、全く同等の動作も可能。
- いわば、不老不死である。
- 生物の神経細胞(ニューロン)を模倣した人工ニューロンを使っているが、人工ニューロンで多用されている誤差逆伝播と呼ばれるものは脳神経では一様には使われていない。

# AI・機械学習・深層学習関連・書籍

- 中井悦司 「TensorFlowとKerasで動かしながら学ぶ ディープラーニングの仕組み ~畳み込みニューラルネットワーク徹底解説~」 マイナビ出版 [978-4839970277](2,959円税込)
- 宮野悟, 中林潤, 木下淳博, 須藤毅顕 「Pythonで体感！ 医療とAIはじめの一步 ~糖尿病・乳がん・残存歯のデータ、肺のX線画像を使って機械学習・深層学習を学ぶ体験型入門書」 羊土社 [978-4758124188](3,960円税込)
- 赤石雅典 「最短コースでわかる PyTorch & 深層学習プログラミング」 日経BP [978-4296110322](4,070円税込)
- 斎藤康毅 「ゼロから作るDeep Learning ⑤ —生成モデル編」 オライリー・ジャパン [978-4814400591](3,960円税込)
- 有村 秀孝, 角谷 倫之 「レディオミクス入門」 オーム社 [978-4274226380](5,500円税込)

# 画像処理関連

- 藤田広志, 上杉正人, 平原大助, 齋藤静司 「Pythonによる医用画像処理入門」 オーム社 [978-4274225468] (3,740円税込)
- 三浦耕太, 塚田祐基. 「ImageJではじめる生物画像解析」 学研メディカル秀潤社 [978-4780909364] (4,400円税込)
- デジタル画像処理編集委員会 「デジタル画像処理(改訂第二版)」 画像情報教育振興協会 [978-4903474649] (4,290円税込)

# 参考になる図書(1)

- **AIファースト・ヘルスケア**

医療現場におけるAIアプリケーションの利用

- Kerrie Holley (著), Dr. Siupo Becker (著), 木村 映善 (監修), 岡 響 (翻訳)、オライリー・ジャパン (2022/9/14)、ISBN: 978-4814400034

- **The next generation of evidence-based medicine**

- Vivek Subbiah、Nature Medicine (2023)
- <https://www.nature.com/articles/s41591-022-02160-z>

- **TensorFlowとKerasで動かしながら学ぶ ディープラーニングの仕組み ~畳み込みニューラルネットワーク徹底解説~**

- 中井悦司(著)、マイナビ(2019/11/28)、ISBN: 978-4839970277

# 参考になる図書(2)

- 本質を捉えたデータ分析のための

## 分析モデル入門

統計モデル、深層学習、強化学習等 用途・特徴から原理まで一気通貫!

- 杉山 聡(著)、ソシム(2022/7/26)、ISBN:978-4802613774

- データ分析に必須の知識・考え方

## 認知バイアス入門

分析の全工程に発生するバイアス その背景・対処法まで完全網羅.

- 山田典一(著)、ソシム(2023/8/26)、ISBN: 978-4802614269

# 参考になる図書(3)

- 山本晋太郎, 徳永匡臣, 箕浦大晃, 邱玥, 品川政太郎, 片岡裕雄. **Vision Transformer入門**. 技術評論社; 2022.
- 日経サイエンス編集部. **日経サイエンス2023年5月号 (特集:対話するAI ChatGPT)**. 日経サイエンス; 2023.
- Foster D. **生成 Deep Learning** —絵を描き、物語や音楽を作り、ゲームをプレイする. オライリージャパン; 2020.
- 今井翔太. **生成AIで世界はこう変わる**. SBクリエイティブ; 2024.