

## Ⅱ-7 生成AIの動向と仕組み

### 学習目標(What)

生成AIの基本構造と医療応用の可能性・リスクを理解する。

### 到達レベル(Can do)

- 生成AIの特徴を従来AIと比較して説明できる
- 医療での利用時の注意点を説明できる

### 学習内容

- 生成AIとは何か
- 大規模言語モデル(LLM)の基本的考え方
- ChatGPTの仕組みと特徴
- 医療文書作成・対話支援への応用例
  
- Diffusionモデルの直感的理解
- ハルシネーションの問題
- 医療分野での安全性・信頼性の課題

## 1. はじめに: 生成AIがもたらした変化

近年、ChatGPTに代表される生成AI(Generative AI)が登場し、そのインパクトは社会に大きな影響を与えた。従来のAIが主に「識別・分類」に重点を置いていたのに対し、生成AIは新たな文章・画像・音声・動画を学習データの統計的構造を利用して生成することが可能である[AI白書2025]。例えば2023年には、ChatGPTやGoogle Bard(現: Gemini)といった対話型AIが急速に世界中で普及し、人間に近い文章生成能力を示した。医療分野においても、生成AIの応用が急速に広がっている。診療記録の要約自動作成[1]、研究論文の執筆支援[2]や論文生成[3]、教育用症例シナリオの生成[4]など、その応用範囲は多岐にわたる。生成AI技術を安全かつ効果的に活用するためには、生成AIの基本構造とその限界を正しく理解する必要がある。こうした背景を踏まえ、本章では、生成AIの技術的背景を概説し、医療応用における可能性と課題について整理する。

## 2. 生成AIとは何か

「生成AI」とは、大量のデータからパターンを学習し、そこから確率的に新しいコンテンツ(テキスト・画像・音声など)を生成するAIの総称である[5]。単に記憶した過去デー

タを再現するのではなく、学習した文脈に応じて新たな組み合わせのアウトプットを生み出せる点が特徴である。生成AIには様々なタイプがあり、扱うデータのモダリティによって以下のように分類できる。

分類	内容	代表モデル
テキスト生成	人間と対話できる文章生成、要約・翻訳など	GPT-4、Claude 3、PaLM 2 / MedPaLM
画像生成	指定した特徴を持つイラスト・写真・医用画像の生成	Stable Diffusion、DALL-E 3
音声・動画生成	音声読み上げ、映像内のキャラクター合成など	VALL-E, Sora
マルチモーダル生成	テキストと画像など複数モダリティを統合生成	Gemini 1.5、GPT-4、Claude 3

近年では、上記のようにテキストだけでなく画像や動画を扱う**マルチモーダルAI**も登場しており、GPT-4のように画像を入力して解釈するモデルや、GoogleのGeminiのように複数モダリティを統合的に処理できるモデルも開発されている[6]。生成AIは人間の創造性を支援し、新たな知的生産を可能にする技術として大きな注目を集めている。

### 3. LLM(大規模言語モデル)の仕組み

#### (1) LLMとは何か

**\*\*大規模言語モデル(LLM: Large Language Model)\*\***は、インターネット上のテキストや書籍、論文など膨大な文章データを用いて事前学習された、パラメータ数が数十億～数千億規模に及ぶニューラルネットワークモデルである。OpenAIのGPT系列やAnthropicのClaude、GoogleのPaLM/Med-PaLMなどは代表的なLLMであり、ChatGPTはGPT系LLMを用いた対話サービスの名称である。これらのLLMは、質問応答・要約・翻訳・コード生成など多様な自然言語タスクを、追加のタスク固有学習なしでもこなせる汎用性の高さが特徴である。

#### (2) トランスフォーマーと自己注意機構

現在主流のLLMは、2017年に提唱されたトランスフォーマーアーキテクチャに基づいている。トランスフォーマーの中核にある「**自己注意機構(Self-Attention)**」という仕組みによってテキスト中の単語同士の関連性を効率よく学習できる点が画期的であった[7]。それまでのRNNやLSTMなどのモデルは長文を理解することが難しかったのに対し、トランスフォーマーはRNNのように前から順番に処理するのではなく、文章全体の単語同士の重要度(注意重み)を一度に計算することで、長文でも長距離の文脈を捉えることができる。トランスフォーマーでは自己注意機構にさらにマルチヘッド機構を

組み合わせることで、多面的な文脈関係を同時に考慮でき、翻訳や要約といった高度な言語処理が飛躍的に向上した。Transformer登場以降、LLMの性能は飛躍的に伸び、GPT-3やGPT-4のようなモデルへと発展しており、現在では多くのLLMや画像モデルなど多くのモデルに応用されている。

### (3) テキスト生成の仕組み

LLMによる文章生成は、「次に最もありそうな単語(トークン)」を次々と予測することによって行われる。モデルは与えられた文脈をもとに、確率的に最適な単語を一つ選び出し、それを文末に付け加えていくことで文章を紡いでいく。例えば「患者は発熱と咳を訴え…」という入力に対し、その続きとして「…肺炎の可能性が高い」「…検査を行った」など、多くの症例テキストを学習したモデルは文脈に適した語句を順次生成する。この「\*\*次単語予測\*\*」という原理が、要約・翻訳・対話・論文下書き作成など様々なタスクの基盤となっている。

### (4) 医療分野での活用例

LLMは「言語を理解し適切に応答する」能力を備えつつあり、医療分野でも様々な応用が試みられている。

- 主な活用例
  - **診療録やサマリーの自動要約**: 長いカルテ記載から重要点を抽出し要約する[8, 9]。
  - **研究支援**: 関連文献の検索や要約、論文執筆のドラフト生成を補助する[10]。
  - **教育対話**: 臨床推論トレーニングで仮想患者役を務めたり、学生の質問に答えるチューターとなる[4, 11, ]。
  - **意思決定支援**: 症例情報を入力すると鑑別診断の候補リストや追加検査の提案を生成する[12, [13]](<https://research.google/blog/amie-a-research-ai-system-for-diagnostic-medical-reasoning-and-conversations/>)。
- 課題
  - **幻覚(hallucination)**:  
もっともらしい誤情報を生成することがあり、医学的に重大な影響を及ぼす可能性がある。
  - **専門性の偏り・データバイアス**:  
学習データの偏りが診断提案にも影響しうる。
  - **根拠提示の不十分さ**:

「なぜその回答に至ったか」が説明しづらく、医療判断に直接使いにくい。

○ **責任とガバナンス:**

AI の出力は医療者が検証し、最終判断は人間が行う必要がある。

このようにLLMは医療文書生成・教育支援・診断推論など幅広い領域で有望である一方、安全性と信頼性の担保が不可欠であり、“補助的ツール”として慎重な運用が求められる。

## 4. 画像生成モデル

### (1) 画像生成モデルの種類と特徴

画像生成は、生成AI(Generative AI)の中でも中心的な応用分野の一つであり、その中核をなすのが深層生成モデル(Deep Generative Model)である。深層生成モデルは、画像のような複雑で高次元のデータでも高忠実に生成することを可能にした。

主要な画像生成モデルの種類と特徴は以下の通りである。

モデルの種類	基本原理	主な特徴・強み	課題・注意点	代表的な例
敵対的生成ネットワーク(GAN)	生成器と識別器の2つを競合させて学習する。	シャープで高忠実な画像を生成できる。尤度を計算する必要がない。	学習の安定化が難しい。モード崩壊(Mode Collapse)により、生成画像の多様性が低くなる場合がある。	StyleGAN[14]
変分オートエンコーダ(VAE)	エンコーダでデータを潜在変数に圧縮し、デコーダで再構築する潜在変数モデルである。	データの特徴を要約した <b>抽象化表現</b> が得られる。 <b>尤度を評価できる(下限)</b> 。	生成される画像の品質や尤度スコアは、他のモデルに比べて劣る傾向がある。	IWAE[15]
自己回帰モデル	過去の出力結果を条件として、次のデータを逐	データ分布を最も上手くモデル化できる強力な生成モデルである。テストデータ(生成が遅い。	各次元を順番に生成するため、高次元データ(画像な	PixelCNN[16]

モデルの種類	基本原理	主な特徴・強み	課題・注意点	代表的な例
	次的に生成する。	の対数尤度を高く達成できる。	データ全体を表す潜在因子を持たない。	
正規化フロー	簡単な分布に可逆変換を繰り返し適用し、複雑な分布を得る。	尤度を正確に計算可能である。高品質な生成と安定した学習を両立できる	複雑な分布をモデル化するために多数の層が必要となり、パラメータ数が増えがちである。	Glow[17]
拡散モデル (Diffusion Model)	ノイズから画像を段階的に再構成する逆拡散過程を学習する。	近年主流であり、生成品質が最も高く、多様なデータを生成できる。GANよりも学習が安定しやすい。テキスト条件による画像生成で大きな成功を収めている。	生成に時間がかかることが弱点とされる。	DALL・E2[18]、Stable Diffusion[19]、Imagen[20]

## (2) 拡散モデル(Diffusion Model)の基本原理

拡散モデル(Diffusion Model)は、画像生成AIにおいて近年主流となっている手法であり、その原理は「ノイズから画像を徐々に再構成する」プロセスに基づいている。まず、実在の画像にランダムノイズを段階的に少しずつ加えていき、最終的に全くの乱雑な純粋ノイズ画像に変換する。この過程は拡散過程(Diffusion Process)と呼ばれ、順過程とも呼ばれる。この過程は、通常マルコフ連鎖など確率的な枠組みを用いて定義される。次に、この過程を逆方向に進めるようにAIに学習させ、完全なノイズデータから元の画像を段階的に復元する逆拡散過程(Reverse Process)を通じて高解像度かつ多様な画像を生成することが可能になる。

拡散モデルは、従来のGAN(敵対的生成ネットワーク)とは異なり、生成器と判別器の競合トレーニング(敵対的学習)を必要としない。そのため、GANと比較して学習が安定しやすいという特徴があり、GANで課題とされていたモード崩壊(Mode Collapse)が起こりにくく、豊富で多彩な画像を生成できる。

拡散モデルは、テキスト条件を与えて画像を生成する「画像+文章」分野でも大きな成功を収めている。代表例として、OpenAIのDALL・E 2やStability AIのStable Diffusionが挙げられる。これらのモデルは、テキストの説明に合致する新規画像を高品質に生み出し、大きな注目を浴びた。DALL・E 2は2022年4月に一般向けの提供が始まり、Stable Diffusionは同年8月に無料で公開されたことで大きな話題となった。

Stable Diffusionは、計算効率を向上させる潜在拡散モデル(Latent Diffusion Model: LDM[19])をベースとしている。

### (3) 生成モデルの医療画像への応用

近年、医療画像解析の分野においても生成モデルの研究が急速に進展している。従来主流であったGAN(敵対的生成ネットワーク)に加え、新たに拡散モデル(Diffusion Model)が登場したことで、それぞれの特性(推論速度、生成品質、学習に必要なデータ形式など)を踏まえた多様な応用可能性が示されている。

#### 1. 拡散モデルとGANの役割分担

拡散モデルは、ノイズ除去を段階的に行う逆拡散過程を通じて画像を生成する手法であり、GANが抱えていた「学習の不安定さ」や「生成画像の多様性の不足(モード崩壊)」といった課題を相対的に克服し、高品質で多様性の高い画像生成を実現することが示されている。一方で、CycleGAN に代表されるGANベースのモデルは、異なるモダリティ間(例: MRI-CT間)でペアの対応画像が存在しない場合でも学習可能であり、さらに推論が高速であるという特徴から、依然としてドメイン適応やリアルタイム処理などの用途で重要な技術基盤となっている。

#### 1. 具体的な研究

各生成モデルの強みを活かし、具体的には以下のような応用に向けた研究が進められている。

#### データ拡張(Data Augmentation)

希少疾患や症例数の少ないデータセットに対し、合成画像を追加学習させることでモデル性能を補完する研究がある。たとえば、拡散モデルで生成したMRI画像を用いた追加学習により、乳腺腫瘍セグメンテーションの精度指標(Diceスコア)が向上したとする報告がある[20]。

#### モダリティ変換とドメイン適応(Modality Transfer)

MRI画像からCT画像を推定生成するなど、異なる撮影装置間の画像変換では、同一患者のペアデータが得にくいことから、CycleGANのような「教師なし(Unpaired)」学習が可能なモデルが利用されている[21,22]。これにより、追加撮影を行わずに放射線治療計画に必要なCTに近い画像を生成する試みや、施設間で生じる画質差(ドメインシフト)の軽減に向けた研究が進められている[23]。

#### 欠損データの補完と修復(Inpainting / Reconstruction)

撮影範囲の欠損やアーチファクト(ノイズ)を補完するため、周囲の文脈から自然な画像を生成する研究も報告されている。拡散モデルはノイズ除去の原理を基盤とするため、欠損領域を滑らかかつ自然に補完できることが示されている[24]。一方、低線量CTの高画質化など、生成画像の整合性を保ちつつ高速処理が求められる場面では、GANベースの超解像手法も依然として重要な役割を果たしている[25]。

## 異常検知への応用

正常な臓器画像のみを学習させ、そこから逸脱するデータ(再構成誤差や潜在空間での尤度低下を伴うデータ)を「異常」と見なす研究分野では、VAE(変分オートエンコーダ)やGANが早くから用いられてきた[26, 27]。近年は拡散モデルを用いて正常像を生成し、実際の患者画像との差分をより高精度に抽出することで、微細な異常部位の強調を試みる研究も見られる[28, [29]](<https://arxiv.org/pdf/2410.23834>)。

このように、現在の医療画像解析研究では、拡散モデルが既存手法を完全に置き換えたわけではなく、病理画像・X線・MRI・CTといったモダリティや、用途(精度優先/速度優先、ペアデータの有無など)に応じて、拡散モデルとGAN等が相補的に利用されているのが実情である。

## 5. ChatGPTなどの汎用LLMと特化型LLMの医療応用例

### (1) 汎用LLMの主な活用と可能性

ChatGPTに代表される汎用大規模言語モデル(LLM: Large Language Model)は、ユーザの指示に応じて自然な文章を生成し、対話形式で情報提供や文書作成を支援するAIである。これらの汎用LLMは、医療特化の追加学習を行なった専門特化型LLM(MedPaLMなど)とは異なるものの、その強力な言語能力から、医療教育・臨床支援・研究業務など、医療領域の多様な場面で試験的な活用が進みつつある。主な応用例として以下のようなものが挙げられる。

#### ① 医学生・研修医の教育支援

ChatGPTは対話形式による学習支援に適しており、医学生や研修医が質問に対して即座に解説を得られる“バーチャル教育者”として利用できる。また、OSCE(客観的臨床能力試験)のロールプレイにおいて模擬患者役を務めさせ、学習者の応答に対してフィードバックを返すといった用途にも応用可能である。対話型LLMの柔軟性により、多様な臨床シナリオを自動生成し、演習に活用することも可能となりつつある。

#### ② 診療支援: 情報整理・文書作成の効率化

問診内容を文章として入力すると、要点を整理したカルテの下書き(SOAP形式など)を自動生成したり、医師の口述内容をリアルタイムで文字化・要約したりする支援が期待されている。また、汎用LLMの言語処理能力を基盤技術として用い、電子カルテや医療情報システムと連携し、診療中に医師の質問に答える“対話型リファレンス”として利用する特化型アプリケーションへの応用試みも始まっている[ref]。これにより、医療従事者の文書作業の負担軽減や情報参照の効率化が見込まれる

### ③ 文書作成支援:学術・臨床業務の効率化

学会抄録、研究計画書、倫理委員会申請書、患者説明資料など、医療現場では大量の文書作成業務が発生する。汎用LLMを利用することで、まず叩き台となる文章を迅速に生成し、その後、人間が内容を推敲するというワークフローが可能になる。これにより、事務作業にかかる時間や負担を軽減し、医療者が専門的判断に割ける時間を増やす効果が期待される。

### ④ 翻訳・要約:海外知見の迅速なキャッチアップ

ChatGPTは医学用語を含む英語文献の翻訳や、日本語での要点要約に利用できる。最新のガイドラインや論文の内容を短時間で把握可能になるため、臨床医・研究者が情報更新を行う上で有用であり、医療者の情報収集を支援するツールとして活用が広がっている。

## (2) 汎用LLM活用のリスクと限界

一方で、汎用LLMは一般知識データで学習されており、医療に特化した学習を行っているわけではないため、医学的専門知識の正確性には限界があり、出力内容をそのまま医療判断に用いることは危険である。

### ハルシネーションと内容の正確性

実際、ChatGPT(GPT-3.5)がアメリカ医師国家試験(USMLE)形式の問題を解いた研究では約60%の正答率で合格ライン付近に達し[30]、改良されたGPT-4や最新のGPT-4oでは90%を超える正答率を記録している[31]。一方で、医療に関する質問への回答や医学的な文章生成を行う際、誤情報や存在しない情報(ハルシネーション)をもっともらしく出力するリスクが指摘されており、医学的質問への回答で事実誤認が確認されたほか[32]、生成された医学論文の参考文献の約半数(47%)が捏造されていたとする報告もある[33]。これは、LLMが「真偽」そのものではなく、「訓練データに照らしてもっともらしい文章」を確率的に選び出すよう設計されていることに起因しており、医学的文脈では重大なリスクとなりうる。

### プライバシーと責任

さらに、外部サービスに患者情報を入力する行為は**プライバシーや個人情報保護の観点**から厳格な管理が求められ、実際の診療データをそのまま汎用LLMに入力することは避けるべきであり[AI戦略会議]、

利用する場合には匿名化や要約などの工夫が不可欠である。

そのため、医療現場で汎用LLMを使用する際には、生成された出力をそのまま用いるのではなく、以下の点を前提とした運用が求められる。

- **専門家による内容確認(Human-in-the-loop)**: 生成された情報を必ず最終責任を持つ医療者が検証すること。
- **慎重姿勢の維持**: 生成された情報を“参考意見”として扱うこと。
- **データの厳格管理**: 個人情報を含むデータの取り扱いに関する厳格なルールを遵守すること。

汎用LLMは強力な言語生成ツールである一方、医療における完全な自動化や代替を期待すべきではなく、医療者が批判的思考をもって適切に運用することが不可欠である。

### (3) MedPaLM: 医療特化型LLMの可能性

Med-PaLM は、Google が開発した医療分野特化型の大規模言語モデルである[34]。基盤となる PaLM/Flan-PaLM に対して、医療質問応答データセットや医学的指示(instruction)を用いて領域に特化した調整(instruction prompt tuning)を行うことで、医学知識に適合した応答を生成できるように設計されている。汎用LLMと比較して、医療QAタスクにおいて高い妥当性と回答品質を示す点が特徴である。

Med-PaLM の前段となる研究として、Flan-PaLM は米国医師免許試験(USMLE)形式の MedQA において 67.6% の正答率を達成し、それまでの既存モデルを大きく上回ったことが Nature に報告されている[35]。この成果を基盤に、医療領域に合わせて調整されたモデルが Med-PaLM(初代)である。Med-PaLM は factuality(医学的事実性)や harm(潜在的有害性)などの人間評価において Flan-PaLM より改善を示した [35]。

その後登場した Med-PaLM 2 は PaLM 2 を基盤とし、医学QAベンチマーク MedQA において 86.5% の正答率を達成し、前版から 19%以上の大幅な性能向上を示したと Nature Medicine で報告されている[36]。さらに、医師によるブラインド評価では、消費者向け医療質問(long-form QA)に対する回答の臨床的妥当性・根拠の明瞭さ・思考過程の整合性など複数の観点において、Med-PaLM 2 の回答が一般医の回答より好まれたケースがあるとされ、限定的な条件下ながら一部領域で専門家に近い回答品質を示し始めている[36]。

Med-PaLM のような医療特化LLMは、診断補助、症例検索、患者教育支援、治療方針提案など、臨床現場での幅広い応用が期待されている。しかし、出力には誤り・偏りが残り得ること、説明可能性や安全性に課題があることも指摘されている。したがって、どれほど高性能であっても、最終的な判断は医療専門職が行うという原則のもと、安全性と有効性を十分に検証しながら段階的に導入していくことが不可欠である。

## 6. プロンプト設計と倫理的留意点

生成AIを医療分野で活用する際には、プロンプト設計(Prompt Engineering)が応答内容に大きく影響する。しかしプロンプト技法は急速に進化しており、本節で示す内容はあくまで現時点での参考例である。また、医療領域で利用する際には、情報の正確性やプライバシー保護など、倫理的・法的側面への配慮を欠かすことができない。以下では、良いプロンプトの条件と、医療分野で生成AIを使用する際の主要なリスクを整理する。

### (1) 良いプロンプトの条件

AIに期待どおりの回答を得るには、質問内容や指示の与え方を工夫する必要がある。プロンプトの品質を高めるための基本的ポイントは以下のとおりである。

#### ① 明確かつ具体的に指示すること

曖昧な質問よりも、条件や対象、難易度を具体的に示す方が望ましい出力が得られる。

例:「糖尿病について説明して」よりも「糖尿病の病態を高校生にもわかるように説明せよ」の方が適切な回答を引き出しやすい。

#### ② 出力形式を指定すること

箇条書き、要約、比較表など、回答形式をあらかじめ指示することで情報整理された出力が得やすくなる。

例:「箇条書きで3点にまとめよ」「表形式で比較せよ」など。

#### ③ 役割や口調を設定すること

AIに「内科医として説明せよ」などの役割を与えると、臨床現場に適した口調や構成を引き出しやすい。また、患者に配慮した穏やかな説明を求めるなど、倫理的態度を促す指示も有効である。医療分野では特に、正確性・中立性・患者配慮を考慮したプ

プロンプト設計が求められる。差別的表現や不安を煽る言い回しを避け、根拠に基づいた説明を促すことが重要である。

※繰り返しになるが、プロンプトエンジニアリングの手法は日々進化しており、上記は現時点での一例に過ぎない。

## (2) 倫理・法的課題

医療に生成AIを用いる場合、以下のようなリスクに注意が必要である。

### ① 誤情報の生成 (Hallucination)

生成AIは現実とは異なる情報や存在しない文献をもっともらしく提示することがある。これを医療判断にそのまま反映すると患者に危害が及ぶ可能性がある。AIの出力は必ず人間が検証し、根拠が確認できる情報のみを採用すべきである。

### ② 著作権・データ出所の不透明性

生成された文章や画像には、学習データ由来の表現が混在する可能性がある。出典が明示されないまま利用すると盗用や著作権侵害に繋がらう。医療情報に関しては特に、根拠文献の提示を促すプロンプトや、人手によるエビデンス確認が不可欠である。

### ③ 個人情報・機密情報の漏洩リスク

外部の生成AIサービスに、患者の名前・連絡先・病歴といった機微情報を入力することは情報漏洩の危険がある。クラウドサービスでは入力内容が学習に利用される可能性があるため、診療情報は必ず匿名化し、外部サービスへの入力は避けるべきである。また多くの医療機関が ChatGPT 等の利用ガイドラインを制定し、診療情報の取り扱いを厳格に規制している必要がある。

### ④ 責任の所在の曖昧さ

AIの提案や診断らしき出力を参照して医療行為を行い誤りが生じた場合、誰が責任を負うのかは現行法では明確ではない。医療における最終判断者はあくまで人間の医師であり、AIは意思決定支援ツールに過ぎない。出力内容の根拠・妥当性・専門家の知識との整合性を人間側が吟味する姿勢が必須である。

以上のようなリスクを踏まえ、医療分野で生成AIを活用する際は**慎重さと倫理意識**が不可欠である。WHO(世界保健機関)も2024年に大規模AIモデル(LMM)の倫理とガバナンスに関する指針を公表し、政府や開発者への規制だけでなく、利用する医療者に対しても、AI生成情報の検証義務や患者への透明性確保(AI使用の開示)など、厳格な倫理的配慮を求めている[37]。生成AIはあくまで人間の知的作業を**拡張する**

ツールであり、決して人間の専門性や責任を代替するものではない点を常に意識するべきである。

## 7. まとめ

生成AIは、医療分野における業務効率化や教育・研究の質的向上に寄与する重要な技術として普及が進んでいる。LLMによる高度な言語処理や画像生成技術の活用により、臨床現場での文書作成支援や診断補助、教育現場での個別化学習など、多岐にわたる応用が期待される。

一方で、本稿で論じた通り、誤情報の生成（ハルシネーション）やプライバシー保護、責任の所在といった技術的・倫理的課題も顕在化している。医療現場での導入に際しては、これらのリスクを十分に認識し、出力結果に対する専門家による検証（Human-in-the-loop）を徹底するなど、慎重な運用体制の構築が不可欠である。

生成AIは医療従事者を代替するものではなく、あくまで意思決定や業務を支援するツールである。今後の医療界においては、医療者自身が本技術の特性と限界を正確に理解し、倫理的配慮のもとで適切に活用していくことが求められる。

### 参考文献

- [1] Medimo. (n.d.). *Medimo - 医療従事者のためのAIアシスタント*. <https://medimo.ai/>
- [2] Paperpal. (n.d.). *Paperpal - 研究者のためのAI英文校正・執筆サポート*. <https://paperpal.com/ja>
- [3] Sakana AI. (2024, August 13). *AI Scientist: 世界初のAIによる完全自動科学研究論文執筆システム*. <https://sakana.ai/ai-scientist-first-publication-jp/>
- [4] MedSim AI. (n.d.). *MedSim AI - 医療教育のためのシミュレーションプラットフォーム*. <https://medsimai.com/>
- [5] *AI白書2025 生成AIエディション*
- [6] Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). *Capabilities of Gemini Models in Medicine*. arXiv. <https://doi.org/10.48550/arXiv.2309.17421>
- [7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*. arXiv. <https://doi.org/10.48550/arXiv.1706.03762>

- [8] 株式会社ユビー. (n.d.). ユビー (Ubie) - 症状検索エンジン. <https://intro.dr-ubie.com/>
- [9] エムスリーデジタル株式会社. (n.d.). エムスリーデジタル - クラウド型電子カルテ. <https://digikar.m3.com/>
- [10] SciSpace. (n.d.). *SciSpace - AI-powered research tool.* <https://scispace.com/>
- [11] Meditools. (n.d.). *Meditools - Streamlit app for medical calculations.* <https://meditools.streamlit.app/>
- [12] Diagnosix. (n.d.). *Diagnosix.ai - Clinical diagnosis support.* <https://www.diagnosix.ai/>
- [13] Tu, T., Palepu, A., Sharma, V., et al. (2024, January 11). *AMIE: A research AI system for diagnostic medical reasoning and conversations.* Google Research Blog. <https://research.google/blog/amie-a-research-ai-system-for-diagnostic-medical-reasoning-and-conversations/>
- [14] Karras T, Laine S, Aila T. A Style-Based Generator Architecture for Generative Adversarial Networks. arXiv. <https://doi.org/10.48550/arXiv.1812.04948>
- [15] Burda Y, Grosse R, Salakhutdinov R. (2016). IMPORTANCE WEIGHTED AUTOENCODERS. arXiv. <https://doi.org/10.48550/arXiv.1509.00519>
- [16] Sennrich, R., Haddow, B., & Birch, A. (2015). Pixel Recurrent Neural Networks. arXiv. <https://doi.org/10.48550/arXiv.1601.06759>
- [17] Kingma, D. P., & Dhariwal, P. (2018). *Glow: Generative Flow with Invertible 1x1 Convolutions.* arXiv. <https://doi.org/10.48550/arXiv.1807.03039>
- [18] Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). *Hierarchical Text-Conditional Image Generation with CLIP Latents.* arXiv. <https://doi.org/10.48550/arXiv.2204.06125>
- [19] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10684-10695. <https://doi.org/10.48550/arXiv.2112.10752>
- [20] Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). *Large Language Models are Zero-Shot Reasoners.* arXiv. <https://arxiv.org/abs/2205.11487>

- [21] Wolterink, J. M., Dinkla, A. M., Savenije, M. H. F., Seevinck, P. R., van den Berg, C. A. T., & Išgum, I. (2017). *Deep MR to CT Synthesis using Unpaired Data*. arXiv. <https://arxiv.org/abs/1708.01155>
- [22] Yang, H., Sun, J., Carass, A., Zhao, C., Lee, J., Xu, Z., & Prince, J. L. (2018). *Unpaired Brain MR-to-CT Synthesis using a Structure-Constrained CycleGAN*. arXiv. <https://arxiv.org/abs/1809.04536>
- [23] Palladino, J. A., Slezak, D. F., & Ferrante, E. (2020). *Unsupervised Domain Adaptation via CycleGAN for White Matter Hyperintensity Segmentation in Multicenter MR Images*. arXiv. <https://arxiv.org/abs/2010.05734> (※提供タイトルに基づきURLを修正)
- [24] Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., & Ye, J. C. (2023). Solving 3D Inverse Problems Using Pre-Trained 2D Diffusion Models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 22542–22551.
- [25] Wolterink, J. M., Leiner, T., Viergever, M. A., & Išgum, I. (2017). Generative adversarial networks for noise reduction in low-dose CT. *IEEE Transactions on Medical Imaging*, 36(12), 2536–2545. <https://ieeexplore.ieee.org/document/7934380>
- [26] Baur, C., Wiestler, B., Albarqouni, S., & Navab, N. (2018). *Deep Autoencoding Models for Unsupervised Anomaly Segmentation in Brain MR Images*. arXiv. <https://arxiv.org/abs/1804.04488>
- [27] Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., & Langs, G. (2017). Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. *International Conference on Information Processing in Medical Imaging (IPMI)*, 146–157.
- [28] Wolleb, J., Bieri, R., Sandkühler, R., & Cattin, P. C. (2022). *Diffusion Models for Medical Anomaly Detection*. arXiv. <https://arxiv.org/abs/2203.04306>
- [29] Bercea, C. I., Cattin, P. C., Schnabel, J. A., & Wolleb, J. (2024). *Denoising Diffusion Models for Anomaly Localization in Medical Images*. arXiv. <https://arxiv.org/abs/2410.23834>
- [30] Kung, T. H., Cheatham, M., Medenilla, A., et al. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*, 2(2), e0000198.

- [31] Chen, Y., Huang, X., Yang, F., Lin, H., Lin, H., Zheng, Z., Liang, Q., Zhang, J., & Li, X. (2024). Large language models in medical education: a comparative study of ChatGPT, Bard, and Claude. *BMC Medical Education*, 24, 461. <https://doi.org/10.1186/s12909-024-06309-x>
- [32] Gravel, J., D'Amours-Gravel, M., & Osmanliu, E. (2023). Learning to Fake It: Limited Responses and Fabricated References Provided by ChatGPT for Medical Questions. *Mayo Clinic Proceedings: Digital Health*, 1(3), 226–234.
- [33] Bhattacharyya, M., Miller, V. M., Bhattacharyya, D., & Miller, L. E. (2023). High Rates of Fabricated and Inaccurate References in ChatGPT-Generated Medical Content. *Cureus*, 15(5), e39247.
- [34] Google Research. (2023). *Med-PaLM: A large language model for the medical domain*. <https://sites.research.google/med-palm/>
- [35] Singhal, K., Azizi, S., Tu, T., et al. (2023). Large language models encode clinical knowledge. *Nature*, 620(7972), 172–180.
- [36] Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Amin, M., Hou, L., Clark, K., Pfohl, S. R., Cole-Lewis, H., Neal, D., Rashid, Q. M. M., Schaeckermann, M., Wang, A., Dash, D., Chen, J. H., Shah, N. H., Lachgar, S., Mansfield, P. A., . . . Natarajan, V. (2024). Toward expert-level medical question answering with large language models. *Nature Medicine*. <https://doi.org/10.1038/s41591-024-03423-7>
- [37] World Health Organization. (2024). *Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models*. World Health Organization.