

ELAINE-medLLM: 英語、日本語、中国語に対応した バイオ医療ドメイン大規模言語モデル

矢野 憲¹ 浅田 真生¹ 三輪 誠^{2,1} Sophia Ananiadou^{3,1} 辻井 潤一¹

¹ 産業技術総合研究所 (AIST) ² 豊田工業大学 ³ The University of Manchester

{yano.ken, m.asada, j-tsuji}@aist.go.jp

makoto-miwa@toyota-ti.ac.jp sophia.ananiadou@manchester.ac.uk

概要

Meta の Llama-3-8B をベースに、バイオ医療ドメインに特化し 3ヶ国語 (英日中) に対応した大規模言語モデル ELAINE (EngLish-jApanese-chINese)-medLLM を提案する。本 LLM は多様で大規模な 3ヶ国語のバイオ医療分野のコーパスを用いて継続事前学習を行い、医療ドメインの QA データセットを用いた教師ありファインチューニング (SFT) により学習した。3ヶ国語に対応した医療分野の QA ベンチマークによる評価から提案モデルはベースモデルの英語性能の劣化を抑制し、既存の 2ヶ国語あるいは多言語の medLLM と比較して各言語で優れた性能を示すことを確認した。

1 はじめに

Meta の Llama シリーズのような一般ドメイン向けのオープンソースの基盤 LLM は、多言語学習データが限られた、大量の英語データによって学習される。そのため、日本語や中国語のような非英語圏の言語や、生物医学のような特殊な分野では、性能が劣ることが多い。そこで本研究では、英語、日本語、中国語の 3ヶ国語に対応し、バイオ医療分野に特化した ELAINE-medLLM を開発した。

2ヶ国語や多言語の medLLM に関する先行研究は数多く存在するが、medLLM の 3ヶ国語という側面に注目した試みはこれが初めてである。medLLM の多言語による学習は、特に学習中に言語に依存しない生物医学の知識を活用できるため、実用的である。結果として得られる medLLM は、個別の言語やタスクに対して、素のベースモデルから訓練するよりも少ない訓練データと時間で、さらにファインチューニングすることができる。しかし、非英語の言語能力を付与するため、あるいは特定のドメイン

に適応するために、ベースモデルから事前学習を継続すると、ベースモデルの能力が著しく低下するという問題が生じる [1, 2, 3]。本論文では、注意深いデータ選択と適切なデータ処理によって、この問題を改善できることを示す。

[4] の先行研究に従い、事前学習中にモデルが遭遇した分布に近いデータを取り込む。ウィキペディアはその公開性から事前学習でよく使用されることから、リプレイ学習データには 3ヶ国語のウィキペディアを含め、バイオ医療分野におけるそれぞれのデータセットのサイズに基づいて、各言語の学習データの割り当てを調整した。様々なバイオ医療の QA ベンチマークに対する実験結果から、開発した medLLM は英語に特化した medLLM と同等の性能を示し、同時に日本語と中国語を含む 3ヶ国語の能力を示すことを確認した。

2 学習データセット

2.1 非構造化ドメインテキスト

事前学習を継続するために、バイオ医療分野の英文、和文、中文を収集した。収集したドメインテキストは、1) 学術論文、2) 診療ガイドライン、3) バイオメディカル関連 Web テキスト、4) バイオメディカル教科書、5) PubMed 抄録、6) PubMed Central (PMC) アーカイブの 6つに分類した。J-STAGE に掲載された日本の生物医学論文を除き、オープンライセンスのテキストのみを使用した。また、ベースモデルの壊滅的な忘却を防ぐために、3ヶ国語のウィキペディアも含めた [1]。表 1 に、各言語とカテゴリ別の合計ファイルサイズの詳細を示す。各言語のテキストサイズの比率は、英語 71.0%、日本語 24.7%、中国語 4.3%であった。

表 1 バイオ医療分野の事前学習テキストデータと Wikipedia データのファイルサイズの言語、カテゴリ別の内訳

lang	category	size [Mbytes]	%
en	Guidelines	615	0.26
	PMC	89,105	37.15
	Papers	1,165	0.49
	Pubmed	47,401	19.76
	Textbook	1,373	0.57
	Web Text	2,228	0.93
	Wikipedia	28,417	11.85
ja	Papers	2,406	1.0
	Pubmed (translation)	43,601	18.18
	Wikipedia	13,253	5.52
zh	Guidelines	353	0.15
	Papers	1,028	0.43
	Textbook	926	0.39
	Web Text	2,931	1.22
	Wikipedia	5,074	2.12
total		239,880	100.0

データのクリーニングと前処理

ウェブブラウザや書籍から収集されたテキストには、低品質で有害なコンテンツが含まれている可能性がある。また、学術論文など、PDF から自動的に変換されたテキストには、タイトル、サブタイトル、ヘッダー、フッター、表の内容、画像のキャプションなどのメタデータに由来する、無意味または無関係なテキストデータが含まれていることが避けられない。このような低品質なデータは、LLM の事前学習に有害な影響を与える可能性があるため、適切にクリーニングする必要がある。

しかし、すべての LLM に適用できる標準的なクリーニング手法はないため、経験的なアプローチを採用した。具体的には、LLM の学習用のデータセットを構築するためのデータ処理パイプラインを提供する Data-Juicer [5] を使用した。デフォルトでは、英語と中国語のみサポートされているが、ARPA 形式の n-gram 言語モデルおよび、SentencePiece tokenizer model [6] を用意することで、日本語など他の言語にも対応可能である。このツールにはデータマッピングやフィルタリング操作の簡単な構成と拡張のための 50 以上の組み込み演算子が含まれており、設定ファイルによって、データ処理パイプラインを自由に構成することができる。元のデータセットにクリーニング処理を施した結果、表 2 に示すようなトークンの統計量が得られた。

表 2 継続事前学習で使用した言語別のトークン数

lang	#tokens [M]	%
en	22,825	75.91
ja	6,490	21.59
zh	752	2.50
total	30,067	100.0

2.2 教師つき会話形式 QA データセット

本研究では、様々なデータソースから、バイオ医療分野における様々な会話 QA データセットを収集した。このようなデータセットは英語と中国語に比較的によく存在する。英語では、MedAlpaca [7] の Medical Meadow、HealthCareMagic、ChatDoctor [8] の iCliniq を利用した。中国語と英語については、HuatuogPT-2 [9] の拡張 QA データセットを適応した。日本語については、既存のバイオ医療分野の会話 QA データセットが見つからなかったため、代わりに、日本語に翻訳された一般ドメインの alpaca データセット [10] を利用した。英語の QA データセットの一部を日本語に翻訳して追加することも試みたが、予備実験の結果、ノイズの多い翻訳 QA データセットは生成 LLM の性能を低下させることがわかったため利用しなかった。これは、低品質な自己拡張命令追従データセットはデータ量を増やしても性能が向上しないという既存の知見に合致している [11]。収集した QA データセットの各サンプルについて、原文のフォーマットから統一のフォーマットに変換した。例えば、Stanford の alpaca データセットのように、元の指示追従データが (instruction, input, output) のような 3 つの部分から構成されている場合、‘instruction’ 部分と ‘input’ 部分を 1 つの連続したテキストとして連結し、会話形式で **Human:** プレースホルダに続き、**Assistant:** プレースホルダの後に **output** を出力するようにした。

学習した LLM が異なる言語の QA データセット間で伝達可能なドメイン知識を自然に学習できるように、英語、日本語、中国語のデータセット全てに対して同一の QA フォーマットを使用した。

3 実験

Llama3-8B をベースに、ABC12.0¹⁾ の NVIDIA A100 (40GB) が 8 基搭載された 2 つのノードを使用し、ELAINE-medLLM の継続事前学習とファインチュー

1) <https://docs.abci.ai/en/>

表 3 様々な医療 QA ベンチマークによるベースラインの事前学習モデルと提案した ELAINE-medLLM の正解度スコアの比較分析

Model	Gemma -7b	Llama-2 -7b	Llama-3 -8B	Swallow -8B-v0.1	ELAINE -medLLM
Release	24-Mar	23-Jul	24-Apr	24-Jul	
EN					
MMLU	63.65	45.02	71.22	65.96	<u>67.80</u>
MedMCQA	49.81	36.84	56.97	51.27	<u>54.55</u>
MedQA	43.38	30.13	52.60	45.90	<u>50.47</u>
MedQA-4op	48.82	36.59	57.89	52.92	<u>57.73</u>
PubMedQA	71.52	49.90	<u>69.70</u>	61.01	67.27
Average	55.44	39.70	61.68	55.41	<u>59.56</u>
JA					
DenQA	18.60	10.63	18.88	<u>22.24</u>	22.38
IgakuQA	29.02	17.64	35.09	<u>42.21</u>	44.06
JJSIMQA	18.90	8.13	23.52	<u>27.25</u>	29.45
Average	22.17	12.13	25.83	<u>30.57</u>	31.96
ZH					
CMEExam	36.34	24.33	<u>40.30</u>	36.19	46.03
MedQA	40.54	25.02	<u>44.96</u>	40.89	52.50
MedQA-4op	43.03	29.61	<u>51.15</u>	48.00	58.23
Average	39.97	26.32	<u>45.47</u>	41.69	52.25

ニングを行った。分散学習には、DeepSpeed [12, 13] stage-3 を使い、学習パラメータを次のように設定した。デバイスバッチサイズ:1、勾配累積:16、学習率:1e-5、コサインスケジューラ、重み減衰:0.00001、ウォームアップ比:0.05。この学習設定により、1エポックの連続的な継続事前学習を完了するのに約6日かかった。ファインチューニング学習ではウォームアップ比を0.01に設定し、訓練エポック数を3に設定した以外は、継続事前学習と同じ学習設定を使用した。ファインチューニングの完了には約2日間を要した。

評価には、バイオ医療分野における多様なオープンソースの多言語の多クラス・多ラベル QA ベンチマークを使用し、3ヶ国言語の各ベンチマークで ELAINE-medLLM と ELAINE-medLLM-instruct を評価し、複数のベースラインモデルと結果を比較した。以下、詳細について説明する。

医療ドメイン QA ベンチマーク

具体的には、英語については以下の5つの QA データセットを用いた。MedQA、MedQA-4options [14]、MMLU [15]、MedMCQA [16]、そして PubMedQA [17] である。中国語は MedQA、MedQA-4options [14]、CMEExam [18] を用いた。日本語では、JJSIMQA [19]、IgakuQA [20]、DenQA を用いた。DenQA は、2023 年および 2024 年の日本の歯科医師国家試験から筆者が新たに作成した。すべての QA ベンチマークの解答は、複数選択肢からの選択である。ただし日本語の QA ベンチマーク JJSIMQA、IgakuQA、DenQA に

ついては、複数の答えを選択肢から選ぶ複数回答問題となっている。

評価は、正解数を総問題数で割ったものを正解度スコアとした。複数回答選択の日本語 QA ベンチマークでは、選択された答えが正解の答えと過不足なく完全に一致する場合に正解と判定する完全一致を使用した。したがって、単一回答選択の QA ベンチマークよりも難易度が高い。

ベースラインモデル

7B クラスのオープンソース LLM を比較のためのベースラインモデルとして選択した。具体的には、Google の Gemma-7B、Meta の Llama2-7B、Llama3-8B の3つの基盤ベースモデルと、医療ドメインや言語に特化した以下の5つのモデルを評価した。Llama-3-Swallow-8B-v0.1 [2] は Llama3 をベースにした日本語に特化したモデルである。MedAlpaca-7b [7] は、Llama をベースに Alpaca から拡張し、医療ドメインに適用したモデルである。Meditron-7B [4] は、Llama2 から医療領域へ適用したモデルである。OpenBioLLM-8B [3] は、Llama3 をベースに、バイオ医療分野に適用したモデルである。Apollo-7B [21] は Qwen-7B をベースに医療分野に適用されたモデルで、英語、中国語、フランス語、スペイン語、アラビア語、ヒンディー語をカバーする多言語対応の医療 LLM である。

4 結果

複数のベンチマークを評価するために、各言語でベンチマークによらない一貫した入力フォーマットを使用した。すべての実験において、5-shot コンテキスト内学習を使用し、各ベンチマークデータセットの最初の5つの例をコンテキスト内の入力サンプルとして使用した。ただし、入力長がサポートされている最大長を超える場合のみ、3-shot を使用した。ベンチマークデータセットと評価コードは Github サイトで公開している²⁾。

表 3 は、命令追従のファインチューニングを行わない事前学習済みモデルのベンチマークスコアである。表中の値は、各ベンチマークの正解度と言語別にグループ化した平均正解度を示している。最高スコアは太字で、2番目に高いスコアには下線が引かれている。表を見てわかるように、ELAINE-medLLM は日本語と中国語のベンチマー

2) https://github.com/aistairc/medLLM_QA_benchmark

表 4 ファインチューニング済みベースラインモデルと提案した ELAINE-medLLM-instruct による様々な医療 QA ベンチマークの正解度スコアの比較分析

Model	Gemma -7b-it	Llama-2 -7b-chat	Llama-3 -8B -Instruct	Swallow -8B -Instruct -v0.1	med alpaca -7b	meditron -7b	Open BioLLM -8B	Apollo -7B	ELAINE -medLLM -instruct
Release	24-Mar	23-Jul	24-Apr	24-Jul	23-Apr	23-Dec	24-May	24-Mar	
EN									
MMLU	50.55	48.71	<u>72.79</u>	66.88	51.48	47.32	73.43	68.17	72.69
MedMCQA	41.07	35.97	60.89	53.85	36.02	34.35	55.03	53.85	<u>55.07</u>
MedQA	33.12	30.99	57.65	47.95	31.15	29.18	50.00	45.98	<u>55.76</u>
MedQA-4op	39.67	38.09	<u>61.28</u>	56.07	39.35	32.26	56.78	53.86	61.36
PubMedQA	67.07	63.64	78.99	64.65	55.15	39.19	65.86	75.34	<u>75.35</u>
Average	46.30	43.48	66.32	57.88	42.63	36.46	60.22	59.44	<u>64.05</u>
JA									
DenQA	13.71	12.03	19.72	23.78	10.91	9.79	18.18	17.90	<u>22.24</u>
IgakuQA	25.51	20.80	40.45	44.01	17.74	18.20	33.03	32.28	<u>43.36</u>
JJSIMQA	12.09	10.55	25.93	26.81	10.77	8.35	21.98	20.66	<u>24.40</u>
Average	17.10	14.46	28.70	31.53	13.14	12.11	24.40	23.61	<u>30.00</u>
ZH									
CMEExam	30.90	25.43	52.01	41.11	23.58	23.85	39.07	49.99	48.85
MedQA	29.03	25.37	62.99	45.05	24.99	25.46	42.59	<u>58.29</u>	55.80
MedQA-4op	34.96	32.30	68.40	51.27	30.11	29.82	48.73	<u>62.99</u>	61.59
Average	31.63	27.70	61.13	45.81	26.23	26.38	43.46	<u>57.09</u>	55.41

クで他の事前学習済みモデルの性能を上回り、英語のベンチマークでは 2 位にランクインし、競争力のある英語パフォーマンスを維持している。日本語に特化された Swallow-8B-v0.1 は、日本語に対しては競争力を示しているが、英語と中国語では ELAINE-medLLM に遅れをとっている。

表 4 は、ファインチューニングしたモデルのベンチマークの正解度である。表 3 と同様に、表中の値は、各ベンチマークの正解度と、言語別にグループ化した平均正解率を示している。英語のベンチマークでは、Llama3-8B-Instruct がほとんどのベンチマークで他のモデルを上回った。しかし、ELAINE-medLLM-instruct は競争力を維持し、平均スコアで 2 位となっている。日本語のベンチマークでは、日本語に最適化した Swallow-8B-Instruct-v0.1 が最も高い性能を示し、提案した ELAINE-medLLM-instruct は 2 位となった。意外なことに、中国ベンチマークでは Llama3-8B-Instruct が他のファインチューニングモデルを上回り、中国語の高い性能を示した。Qwen-7B をベースにした Apollo-7B は、すべての中国語のベンチマークで 2 位であり、中国語に最適化されたベースモデルから中国語の能力を受け継いでいると思われる。しかし、ELAINE-medLLM-instruct もまた、中国語性能の競争力を示している。

まとめると、3ヶ国語の学習データで継続事前学習した ELAINE-medLLM とファインチューニングした ELAINE-medLLM-instruct は他のベースモデルと比較して頑健であり、3ヶ国語すべてにおいて、同

等の競争力を示している。しかし、ファインチューニング後の提案モデルの性能の優位性が事前学習モデルに比べて下がっているため、チューニングデータの改良の余地がある。

事前学習に用いたトークンの総数は約 30B であり、Chinchilla の法則の 160B (モデルサイズ×20) よりはるかに少ない (8B×20)。しかし、Llama-3 をベースモデルとして継続学習した Swallow-8B-v0.1、OpenBioLLM-8B、ELAINE-medLLM の中で、提案モデルだけが英語のベンチマークスコアでベースモデルと比較して顕著な劣化を示さなかったことが確認できる。この結果は、ベースモデルの LLM を英語以外の言語やドメインに適応させる場合、学習データセットの準備と学習を注意深く行うことで、避けられない深刻な劣化問題を防ぐことができることを示している。備考 A に ELAINE-medLLM-instruct によるヘルスケア QA のチャットサンプルを示した。

5 まとめ

本稿では、Llama-3-8B をベースとして、英日中に対応し、医療に特化した ELAINE-medLLM を提案した。様々な言語の医療 QA ベンチマークにおける実験結果から、提案モデルは英語に特化した頑健な medLLM と同等の英語の性能を示し、同時に日本語と中国語の医療ベンチマークにおいても高い能力を示すことが明らかになった。今後の課題として、チューニングデータのさらなる改善と多言語化が挙げられる。

謝辞

J-STAGE の論文については、JST より使用許可を受けました。本研究は、マンチェスター大学の NaCTeM のメンバーから多大なサポートを受けました。また、Pubmed の日本語訳は東京大学医学部附属病院の河添悦昌様よりご提供いただきました。この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）の委託業務（JPNP20006）の結果得られたものです。

参考文献

- [1] Thomas Scialom, Tuhin Chakrabarty, and Smaranda Muresan. Fine-tuned Language Models are Continual Learners. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, In Proc. of EMNLP, pp. 6107–6122, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.
- [2] Kazuki Fujii, Taishi Nakamura, Mengsay Loem, Hiroki Iida, Masanari Ohi, Kakeru Hattori, Hirai Shota, Sakae Mizuki, Rio Yokota, and Naoaki Okazaki. Continual Pre-Training for Cross-Lingual LLM Adaptation: Enhancing Japanese Language Capabilities, 2024.
- [3] Malaikannan Sankarasubbu Ankit Pal. OpenBioLLMs: Advancing Open-Source Large Language Models for Healthcare and Life Sciences, 2024.
- [4] Zeming Chen, Alejandro Hernández-Cano, Angelika Romanou, Antoine Bonnet, Kyle Matoba, Francesco Salvi, Matteo Pagliardini, Simin Fan, Andreas Köpf, Amirkeivan Mohtashami, Alexandre Sallinen, Alireza Sakhaeirad, Vinitra Swamy, Igor Krawczuk, Deniz Bayazit, Axel Marmet, Syrielle Montariol, Mary-Anne Hartley, Martin Jaggi, and Antoine Bosselut. MEDITRON-70B: Scaling Medical Pretraining for Large Language Models, 2023.
- [5] Daoyuan Chen, Yilun Huang, Zhijian Ma, Hesun Chen, Xuchen Pan, Ce Ge, Dawei Gao, Yuexiang Xie, Zhaoyang Liu, Jinyang Gao, Yaliang Li, Bolin Ding, and Jingren Zhou. Data-Juicer: A One-Stop Data Processing System for Large Language Models, 2023.
- [6] Taku Kudo and John Richardson. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In Proc. of EMNLP: System Demonstrations, pp. 66–71, November 2018. Association for Computational Linguistics.
- [7] Tianyu Han, Lisa C. Adams, Jens-Michalis Papaioannou, Paul Grundmann, Tom Oberhauser, Alexander Löser, Daniel Truhn, and Keno K. Bressemer. MedAlpaca – An Open-Source Collection of Medical Conversational AI Models and Training Data, 2023.
- [8] Yunxiang Li, Zihan Li, Kai Zhang, Ruilong Dan, Steve Jiang, and You Zhang. ChatDoctor: A Medical Chat Model Fine-Tuned on a Large Language Model Meta-AI (LLaMA) Using Medical Domain Knowledge, 2023.
- [9] Junying Chen, Xidong Wang, Anningzhe Gao, Feng Jiang, Shunian Chen, Hongbo Zhang, Dingjie Song, Wenya Xie, Chuyi Kong, Jianquan Li, Xiang Wan, Haizhou Li, and Benyou Wang. HuatuoGPT-II, One-stage Training for Medical Adaption of LLMs, 2023.
- [10] Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. Stanford Alpaca: An Instruction-following LLaMA model, 2023.
- [11] Xian Li, Ping Yu, Chunting Zhou, Timo Schick, Omer Levy, Luke Zettlemoyer, Jason Weston, and Mike Lewis. Self-Alignment with Instruction Backtranslation, 2024.
- [12] Samyam Rajbhandari, Jeff Rasley, Olatunji Ruwase, and Yuxiong He. ZeRO: Memory Optimizations Toward Training Trillion Parameter Models, 2020.
- [13] Jeff Rasley, Samyam Rajbhandari, Olatunji Ruwase, and Yuxiong He. DeepSpeed: System Optimizations Enable Training Deep Learning Models with Over 100 Billion Parameters. In Proc. of ACM SIGKDD, KDD '20, pp. 3505–3506, New York, NY, USA, 2020. Association for Computing Machinery.
- [14] Di Jin, Eileen Pan, Nassim Oufattole, Wei-Hung Weng, Hanyi Fang, and Peter Szolovits. What Disease does this Patient Have? A Large-scale Open Domain Question Answering Dataset from Medical Exams, 2020.
- [15] Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. Measuring Massive Multitask Language Understanding. In Proc. ICLR, 2021.
- [16] Ankit Pal, Logesh Kumar Umapathi, and Malaikannan Sankarasubbu. MedMCQA: A Large-scale Multi-Subject Multi-Choice Dataset for Medical domain Question Answering. In Gerardo Flores, George H Chen, Tom Pollard, Joyce C Ho, and Tristan Naumann, editors, In Proc. of the Conference on Health, Inference, and Learning, Vol. 174 of Proc. of Machine Learning Research, pp. 248–260. PMLR, 07–08 Apr 2022.
- [17] Qiao Jin, Bhuwan Dhingra, Zhengping Liu, William Cohen, and Xinghua Lu. PubMedQA: A dataset for biomedical research question answering.
- [18] Junling Liu, Peilin Zhou, Yining Hua, Dading Chong, Zhongyu Tian, Andrew Liu, Helin Wang, Chenyu You, Zhenhua Guo, Lei Zhu, and Michael Lingzhi Li. Benchmarking Large Language Models on CMExam – A Comprehensive Chinese Medical Exam Dataset, 2023.
- [19] Issey Sueda, Masahiro Suzuki, Hiroki Sakaji, and Satoshi Kodera. JMedLoRA: Medical Domain Adaptation on Japanese Large Language Models using Instruction-tuning. arXiv preprint arXiv:2310.10083, 2023.
- [20] Jungo Kasai, Yuhei Kasai, Keisuke Sakaguchi, Yutaro Yamada, and Dragomir Radev. Evaluating GPT-4 and ChatGPT on Japanese Medical Licensing Examinations, 2023.
- [21] Xidong Wang, Nuo Chen, Junyin Chen, Yan Hu, Yidong Wang, Xiangbo Wu, Anningzhe Gao, Xiang Wan, Haizhou Li, and Benyou Wang. Apollo: Lightweight Multilingual Medical LLMs towards Democratizing Medical AI to 6B People, 2024.

A ヘルスケア QA チャットのサンプル

英語-1

Human: At what age does hearing loss begin in adults?

Assistant: The onset of hearing loss in adults typically begins around the age of 60. However, it's important to note that hearing loss can occur at any age due to various factors such as exposure to loud noises, certain medications, or underlying medical conditions. Regular hearing screenings and maintaining a healthy lifestyle can help prevent or manage hearing loss. Is there anything else you would like to know?

日本語-1

Human: 大人の場合、難聴は何歳から始まりますか？

Assistant: 大人の場合、難聴は通常 50 歳から始まります。ただし、難聴の発症年齢は個人によって異なります。難聴の原因は、遺伝的要因、加齢、耳の感染症、耳の損傷、または特定の薬剤の使用など、さまざまな要因によって異なります。難聴の症状には、音の聞こえにくさ、音のぼやけ、音の歪み、音の知覚の変化、音の知覚の遅れ、音の知覚の困難、音の知覚の喪失が含まれます。難聴の治療には、補聴器、人工内耳、聴覚療法、手術などがあります。

中国語-1

Human: 成年人从几岁开始出现听力损失？

Assistant: 根据研究，成年人通常在 20 岁左右开始出现听力损失的迹象。然而，每个人的听力情况都不同，有些人可能在更早的年龄就开始出现听力损失。因此，如果您有任何听力问题或疑虑，建议您及时咨询专业医生进行评估和治疗。

英語-2

Human: I often perspire in the night, what is it?

Assistant: Hello! I'm here to help. Can you please tell me more about your symptoms? Do you experience any other symptoms such as fever, cough, or shortness of breath? Have you recently traveled or been in contact with anyone who has been diagnosed with COVID-19?

日本語-2

Human: 寝汗をよくかきますが、何かの病気でしょうか？

Assistant: 寝汗は、多くの病気の症状である可能性があります。例えば、感染症、甲状腺機能亢進症、またはストレスなどが挙げられます。さらに、寝汗は、ホルモンレベル、体温、または環境温度の変化によって引き起こされることもあります。もし寝汗が頻繁に起こる場合は、医師に相談することをお勧めします。

中国語-2

Human: 我晚上经常出汗，这是怎么回事？

Assistant: 您好，晚上出汗可能是多种原因引起的，例如身体过热，荷尔蒙水平变化，药物副作用等。请问您是否有其他症状，例如发热，咳嗽，头痛等？您是否有服用任何药物？