No_gmail at #SMM4H–HeaRD 2026: Detecting Patient Metadata in COVID-19 Literature with Encoder-Only and Autoregressive Language Models

Anastasia Ștefănescu
University of Bucharest
anastasia.stefanescu@s.unibuc.ro

Ionuț-Marius Voicilă
University of Bucharest
ionut-marius.voicila@s.unibuc.ro

Abstract

Identifying sentences in COVID-19 literature that report patient metadata is an important step in genomic epidemiology, currently requiring costly manual curation. We compare fine-tuned encoder-only models (BERT, BioLinkBERT, BioBERT) and autoregressive LLMs (LLaMA, Gemma, GPT-OSS) under prompting and fine-tuning regimes, using Focal Loss and undersampling to address severe class imbalance. Encoder-only models substantially outperform autoregressive models: BioLinkBERT-base with Focal Loss achieves macro F1 of **0.76**, versus **0.54** for the best fine-tuned autoregressive model.

1 Introduction

Linking SARS-CoV-2 genome sequences to *patient metadata* (age, sex, comorbidities, symptoms, outcomes) is critical for genomic epidemiology, yet the relevant information is scattered across free-text sentences, tables, and supplementary materials in scientific literature (Chen et al., 2021), requiring costly manual curation.

The task is difficult for three reasons: (i) severe class imbalance—the vast majority of sentences are negative; (ii) linguistic diversity of positive examples, which may describe aggregate statistics, reference external tables, or use ambiguous clinical terminology; and (iii) possible reliance on cross-sentence context that is hard to detect.

This paper describes our system submission to Task 5 of the 11th SMM4H and HeaRD Shared Tasks at ACL 2026 (Lopez-Garcia et al., 2026), which asks participants to classify sentences from COVID-19 literature as reporting patient metadata or not. Klein et al. (2025) introduced the task and showed that fine-tuned BERT variants (medical and general-purpose) outperform prompted Llama-3 (8B and 70B) models, but there are still various models left to experiment with, and the question whether biomedical pre-training helps and which prompting and fine-tuning strategies work best remains open. We address these questions with the following contributions:

- Fine-tuned encoder-only models (BERT, BioLinkBERT, BioBERT) with Focal Loss to handle class imbalance.
- Fine-tuned autoregressive models (Llama-3-8B, Gemma-4-31B-IT, MedGemma-1.5-4B-IT) and compared prompting approaches.

Our code is available at https://github.com/anastasia-stefanescu/Patient-Metadata-Detection-SMM4H.

2 Related Work

Previous work with biomedically pre-trained encoders: BioBERT (Lee et al., 2020), BiomedBERT (Gu et al., 2021), and BioLinkBERT (Yasunaga et al., 2022) has shown that these consistently outperform general-domain BERT on tasks such as biomedical NER and relation extraction, motivating their use here.

GPT-4 (OpenAI, 2023) shows strong zero-shot performance on biomedical benchmarks, with further gains from few-shot prompting (Brown et al., 2020) and chain-of-thought reasoning (Wei et al., 2022).

However, on binary classification under severe class imbalance, fine-tuned encoder-only models have been found to outperform prompted LLMs (Klein et al., 2025). This same article introduced the patient metadata detection task and reiterated the fact that fine-tuned BERT outperforms prompted Llama-3 (8B and 70B) in zero- and few-shot settings; we extend their framework with more biomedical encoders, fine-tuned autoregressive models, and a broader prompt ablation.

Proceedings of the 11th Social Media Mining for Health Research and Applications (SMM4H-HeaRD 2026) Workshop and Shared Tasks, pages 280–285

3 Dataset

We use the dataset introduced in Klein et al. (2025) for the patient metadata detection task. Each instance is a sentence extracted from a full-text COVID-19 article. A sentence is labeled *positive* if it reports, or refers (e.g., in a table) to patient metadata for COVID-19 cases; otherwise it is labeled *negative*.

Split	Total	Positive	Negative
Train	15,504	2,061 (13.3%)	13,443 (86.7%)
Validation	2,214	294 (13.3%)	1,920 (86.7%)
Test	4,429		

Table 1: Dataset statistics; both splits have the same class ratio.

The dataset exhibits severe class imbalance, which reflects the natural distribution of relevant sentences in scientific articles. We use macro-averaged F1 score as our primary metric, as it equally weights performance on both the positive and negative classes regardless of their frequency.

4 Methods

4.1 Fine-tuning Encoder-Only Models

We fine-tune encoder-only transformers for binary sentence classification using the Hugging Face transformers library (Wolf et al., 2020), with a linear classification head on top of the `[CLS]` token representation. We evaluate *BERT-base-uncased* (Devlin et al., 2019) as a general-domain baseline, *BioBERT-base-cased* (Lee et al., 2020) as a biomedical BERT variant pre-trained on PubMed abstracts and PubMed Central full-text articles, and *BioLinkBERT-base* and *BioLinkBERT-large* (Yasunaga et al., 2022) as stronger biomedical alternatives that additionally incorporate a document-link prediction objective, making them well-suited to biomedical literature. All models are trained with AdamW, a batch size of *16*, a learning rate of *2e-5*, a maximum sequence length of *256* tokens, and no learning rate schedule. Focal-loss models additionally use bf16 mixed precision. All experiments are run on an NVIDIA L4 GPU (Google Colab). The models are loaded from the following HuggingFace checkpoints: `bert-base-uncased`, `dmis-lab/biobert-base-cased-v1.2`, `michiyasunaga/BioLinkBERT-base`, and `michiyasunaga/BioLinkBERT-large`.

Addressing class imbalance with Focal Loss. Instead of using cross-entropy which is problematic as the model can minimize loss by predicting the majority class, we used **Focal Loss** (Lin et al., 2017) which improved F1:

$$\mathcal{L}_{\text{FL}}(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \qquad (1)$$

where p_t is the predicted probability for the true class, α_t is a per-class weight derived from inverse class frequency, and γ down-weights easy examples. We set $\gamma = 2$. Per-class weights α_t are computed via scikit-learn's balanced class weighting, yielding $\alpha_{\text{neg}} \approx 0.58$ and $\alpha_{\text{pos}} \approx 3.76$.

4.2 Autoregressive Language Models

We experiment with autoregressive LLMs under two paradigms: (i) fine-tuning on the training set and (ii) prompting approaches.

4.2.1 Fine-tuning

We fine-tuned *Llama-3-8B* (`unsloth/llama-3-8b-bnb-4bit`) (Grattafiori et al., 2024) and *Gemma-4-31B-IT* (`unsloth/gemma-4-31B-it-unsloth-bnb-4bit`) (Google DeepMind, 2026) using QLoRA (Dettmers et al., 2023) (4-bit quantization) via the Unsloth platform (Han and Han, 2023). Both models use rank-stabilized LoRA (RSLoRA) (Hu et al., 2022; Kalajdzievski, 2023) with rank $r = 8$, $\alpha_{\text{LoRA}} = 16$, dropout $= 0$, and all attention and feed-forward projection layers as target modules (`q_proj`, `k_proj`, `v_proj`, `o_proj`, `gate_proj`, `up_proj`, `down_proj`). To handle class imbalance, the training set was undersampled to a 1:1 positive/negative ratio, resulting in 4,122 training samples. The models were trained with an effective batch size of *16* (per-device batch *1*, gradient accumulation *16*), using the AdamW 8-bit optimizer with weight decay *0.01*, a linear decay schedule with *50* warmup steps, and a fixed random seed of *3407*. Llama-3-8B uses a learning rate of *2e-4* with a maximum sequence length of *650* tokens; Gemma-4-31B-IT uses a learning rate of *2e-5* with a maximum sequence length of *256* tokens. Each model was fine-tuned until the loss plateaued, which meant for Llama-3-8B a number of *200* gradient update steps and for Gemma-4-31B-IT, *70* steps. Models were instructed to produce a binary yes/no response, with optional reasoning where we used few-shot prompting.

4.2.2 Prompting

We evaluated *Llama-3-8B* (Grattafiori et al., 2024), *Gemma-4-31B-IT* (Google DeepMind, 2026), and *GPT-OSS-120B* (OpenAI, 2025) via local Ollama, and *MedGemma-1.5-4B-IT* (`google/medgemma-1.5-4b-it`) (Sellergren et al., 2025) via the HuggingFace pipeline. We used for all models a system role of *"You are a clinical data auditor"* with task-specific classification criteria.

The following prompting strategies were applied selectively across models:

1. **Zero-shot.** The system prompt defines the task and output format with no labeled examples.

2. **Few-shot.** Six labeled examples (three positive, three negative) covering diverse linguistic patterns are appended to the prompt. Applied to Llama-3-8B and GPT-OSS-120B.

3. **Few-shot + chain-of-thought (CoT).** The few-shot examples are augmented with explicit reasoning traces, and the model is instructed to reason before producing its answer. Applied to Gemma-4-31B-IT and MedGemma-1.5-4B-IT, as it fits the default way of reasoning of the model. The full pool of MedGemma CoT examples explored during prompt development is listed in Appendix A.3.

Prompt design challenges. Defining the task precisely while keeping the system prompt concise proved to be a significant practical difficulty. Prompts that were too short failed to capture the nuanced requirements of the task (e.g., that references to tables containing metadata count as positive), while overly long prompts caused models to lose track of the core instruction, thus producing contradicting outputs. Iterative prompt refinement was necessary for all model families.

5 Results

5.1 Encoder-Only Models

Table 2 reports macro-averaged F1 for all encoder-only configurations on the validation set, for the epochs with the best performance.

BioLinkBERT variants consistently outperform general-domain BERT-base-uncased, confirming the value of domain-specific pre-training. The best result is BioLinkBERT-base with Focal Loss at 7 epochs, achieving a macro F1 of **0.80**. This also proves the effectiveness of fine-tuning with Focal Loss, and that while fine-tuning helps improve results, the best models are not necessarily the largest or most trained.

Model	Epochs	Focal Loss	F1
BERT-base-uncased	3	No	0.69
BERT-base-uncased	10	No	0.76
BioBERT-base-cased	8	Yes	0.78
BioLinkBERT-base	7	Yes	**0.80**
BioLinkBERT-large	3	Yes	0.76

Table 2: Macro-averaged F1 for encoder-only models for their best epochs (validation set). Bold = best result.

Model	Epochs	Val F1	Test F1
BioBERT-base-cased	8	0.78	0.73
BioLinkBERT-large	3	0.76	0.75
BioLinkBERT-base	7	0.80	0.76

Table 3: Official test-set F1 for submitted encoder-only checkpoints. Val F1 is the macro-averaged F1 on the validation set; Test F1 is the macro-averaged F1 returned by the evaluation server. Bold = best result.

We selected the three top-performing checkpoints from distinct model families for test-set submission: BioLinkBERT-base, BioBERT-base-cased, and BioLinkBERT-large. Although BERT-base-uncased at 10 epochs matched BioLinkBERT-large on the validation set (both 0.76), it was excluded in favour of the biomedical model, which was expected to generalise better to unseen scientific text and offered no informational redundancy with the other submitted checkpoints.

BioLinkBERT-base (7 epochs, Focal Loss) maintains the best F1 score of **0.76**. Notably, BioLinkBERT-large outperforms BioBERT-base-cased on the test set (0.75 vs. 0.73), consistent with its stronger validation performance. The small gap between validation and test scores suggests that the models generalise well without significant overfitting to the validation distribution.

5.2 Autoregressive Models

Table 4 reports macro-averaged F1 for autoregressive models under prompting and fine-tuning.

All autoregressive models fall well short of the best encoder-only result (0.80 vs. 0.55 best on the validation set), a gap of over 25 F1 points. Fine-tuning spectacularly improves Llama-3-8B in a zero-shot setting, from 0.25 to 0.55 at 100 steps,

Model	Steps	Prompt	F1
Llama-3-8B	0	Few-shot	0.37
Llama-3-8B	100	Few-shot	0.44
Llama-3-8B	200	Few-shot	0.47
Llama-3-8B	0	Zero-shot	0.25
Llama-3-8B	100	Zero-shot	**0.55**
Llama-3-8B	200	Zero-shot	0.54
Gemma-4-31B-IT	0	Few-shot + CoT	0.50
Gemma-4-31B-IT	70	Few-shot + CoT	0.49
MedGemma-1.5-4B-IT	0	Few-shot + CoT	0.40
GPT-OSS-120B	0	Few-shot	0.42

Table 4: Macro-averaged F1 for autoregressive models (validation set). Steps = QLoRA fine-tuning gradient steps; 0 = prompting only. Bold = best result.

Model	Steps	Val F1	Test F1
Gemma-4-31B-IT	70	0.49	0.49
Llama3-8B-zero shot	200	0.54	0.54
Llama3-8B-few shot	200	0.47	0.47

Table 5: Official test and validation set F1 for submitted autoregressive model checkpoints.

however it quickly reaches a plateau, as after that, the performance remains the same. It also comes with improvements for the same model with few-shot prompting, although it does not reach the same level as zero-shot fine-tuning, which is interesting and suggests that a longer context rather interferes with the model's ability to learn and reason. Gemma-4-31B-IT with its greater complexity handles the longer context better, however, fine-tuning it does not bring any improvements. MedGemma-1.5-4B-IT performs unexpectedly poorly with few-shot + CoT, given its medical pre-training, as it reaches only 0.40, showing that biomedical domain adaptation with robust reasoning does not compensate for smaller model capacity.

We made submission for the best-performing models to the shared-task evaluation server in the post-evaluation phase to see results on the test set (Lopez-Garcia et al., 2026); Table 5 shows their official test-set scores. It is visible that test and validation results remain very close.

6 Conclusion

We presented a comparative study of encoder-only and autoregressive language models for the task of detecting patient metadata in COVID-19 scientific literature. Our experiments demonstrate that fine-tuned encoder-only models, especially BioLinkBERT-base trained with Focal Loss (macro F1 = 0.76), substantially outperform autoregressive models in all evaluated settings. We can make several conclusions from our results: (i) Fine-tuning smaller models (such as Llama-3-8B) dramatically improves performance, both in zero-shot and few-shot settings, but quickly plateaus, narrowing but not closing the gap with encoder-only classifiers. (ii) The undersampling strategy for autoregressive models is a key method, and the use of Focal Loss provides the largest single improvement among encoder-only models, yielding a 3-point F1 gain over the same model trained without it. (iii) In the case of prompting, longer, more specific prompts hinder smaller models from reasoning. A concise context is the best way to get the best performance from the model, and it is not necessarily the case that more examples will lead to better performance. (iv) Model size alone does not reliably predict performance: BioLinkBERT-base outperforms BioLinkBERT-large, and Gemma-4-31B-IT surpasses GPT-OSS-120B, suggesting that domain-specific pre-training and architectural fit matter more than raw parameter count, and there is a fine line between training too much and just enough.

Future work should investigate (i) cross-sentence context modeling to capture indirect references to metadata, (ii) retrieval-augmented prompting strategies that dynamically select in-context examples, and (iii) larger-scale fine-tuning of autoregressive models to more directly close the performance gap with encoder-only classifiers.

Limitations

The main limitations of this study are the computational cost and time required for fine-tuning and running inference on very large autoregressive models, which restricted the number of configurations we could explore.

References

Tom Brown and 1 others. 2020. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901.

Qingyu Chen, Alexis Allot, and Zhiyong Lu. 2021. Lit-Covid: an open database of COVID-19 literature. *Nucleic Acids Research*, 49(D1):D1534–D1540.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. QLoRA: Efficient finetuning of quantized LLMs. *Advances in Neural Information Processing Systems*, 36.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186.

Google DeepMind. 2026. Gemma-4-31B-IT. Hugging Face.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. 2021. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(1):1–23.

Daniel Han and Michael Han. 2023. Unsloth: Efficient LLM fine-tuning.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Liang Wang, Weizhu Chen, and 1 others. 2022. Lora: Low-rank adaptation of large language models. *Iclr*, 1(2):3.

Damjan Kalajdzievski. 2023. A rank stabilization scaling factor for fine-tuning of large language models. *arXiv preprint arXiv:2312.03732*.

Ari Z. Klein, Davy Weissenbacher, Karen O'Connor, Amir Elyaderani, Ivan Flores Amaro, Takeshi Onishi, Su Golder, Kaelen Spiegel, Matthew Scotch, and Graciela Gonzalez-Hernandez. 2025. Detection of patient metadata in published articles for genomic epidemiology using machine learning and large language models. *medRxiv*.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In *Proceedings of ICCV*, pages 2980–2988.

Guillermo Lopez-Garcia, Jose Miguel Acitores Cortina, Jacob Berkowitz, Joey Chan, Ganesh Chandrasekar, Sumon Kanti Dey, Ivan Flores Amaro, Fernando Gallego, Lauren Gryboski, Ari Z Klein, Martin Krallinger, Salvador Lima-López, Tomohiro Nishiyama, Lisa Raithel, Ahmad Rezaie Mianroodi, Roland Roller, Judith Rosell, Frank Rudzicz, Abeed Sarker, and 8 others. 2026. Overview of the 11th Social Media Mining for Health (#SMM4H) and Health Real-World Data (HeaRD) Shared Tasks at ACL 2026. In *Proceedings of the 11th Social Media Mining for Health (#SMM4H) and Health Real-World Data (HeaRD) Workshop and Shared Tasks*. Association for Computational Linguistics.

OpenAI. 2023. GPT-4 technical report. *arXiv preprint arXiv:2303.08774*.

OpenAI. 2025. gpt-oss-120b & gpt-oss-20b model card. *Preprint*, arXiv:2508.10925.

Andrew Sellergren, Sahar Kazemzadeh, Tiam Jaroensri, and 1 others. 2025. MedGemma technical report. *arXiv preprint arXiv:2507.05201*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V. Le, and Denny Zhou. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and 1 others. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pages 38–45.

Michihiro Yasunaga, Jure Leskovec, and Percy Liang. 2022. Linkbert: Pretraining language models with document links. In *Association for Computational Linguistics (ACL)*.

A Prompt Templates

A.1 Llama and Gemma: System Prompt

You are a clinical data auditor. Your task is to identify if the text includes COVID-19 patient metadata or specific SARS-CoV-2 viral sequence data linked to individuals or specific cases.

Criteria — YES: *Personal attributes (age, sex, location), clinical status, or specific treatments/outcomes tied to individual patients; specific viral genome sequences directly linked to individual patients, specific case investigations, or precise local outbreak clusters.*

Criteria — NO: *General study methods or laboratory protocols; group-level statistics that cannot be traced to an individual case; theoretical discussions of viral dynamics.*

Output format: *Result: [YES/NO]. Reasoning: [Brief explanation].*

Few-shot examples (Llama / Gemma).

User: *"Patient 04-A, a 45-year-old female, provided a nasal swab which yielded a SARS-CoV-2 sequence showing the D614G mutation."* **Assistant:** *Result: YES. Explicitly links a specific patient, demographic data, and a viral sequence.*

User: *"The study analyzed 5,000 sequences collected across the European Union to calculate the overall prevalence of the Alpha variant."* **Assistant:** *Result: NO. Large-scale aggregate statistics, not linked to specific individual patients.*

Zero-shot system prompt (Llama). Used for the zero-shot fine-tuning experiments on Llama-3-8B, which achieved the best Llama result (0.54 macro F1 on the test set at 200 steps).

> *You are a clinical data auditor. Your task is to identify if the text includes patient metadata (demographics, symptoms, history, treatments, or outcomes) linked to individuals.*
>
> **Criteria — YES:** *Personal attributes (age, sex, location), clinical status, or specific treatments/outcomes tied to individuals or cases.* **NO:** *General study methods, group statistics (percentages), or descriptions of data collection without individual details.*

A.2 MedGemma: Prompt Template

> *You are a clinical data auditor. Does this text report specific patient metadata linked to specific individuals or sequences from Covid-19 cases?*
>
> **Metadata types:** *Personal attributes (age, sex, race, location), clinical status (symptoms, severity, viral load, labs, vitals), history (comorbidities, travel), and medical approaches (treatment, outcome).*
>
> **Criteria — YES:** *Reports specific values or statuses for identifiable or implied individuals, cases, or sequences from Covid-19 cases.* **NO:** *Metadata only mentioned as "collected" (no values), data from prior studies, related to other diseases, general lab methods, or aggregate group data without patient/sequence links.*
>
> *Provide your reasoning, then conclude with "Result: [YES/NO]".*
>
> **Text:** *{text}*

Few-shot examples (MedGemma).

> **User:** *"This study aimed to evaluate SARS-CoV-2 genome sequences from COVID-19 cases and characterize their genealogical networks to demonstrate possible routes of spread in Japan."* **Assistant:** *Traces genealogical networks back to specific COVID-19 cases. Result: YES.*
>
> **User:** *"A Ct value for RNAse P was used as an endogenous internal amplification control."* **Assistant:** *Laboratory validation criteria with no patient data. Result: NO.*
>
> **User:** *"Due to the above-mentioned findings in favor of progression, favipiravir treatment was started."* **Assistant:** *Medical treatment tied to an implied patient's clinical status. Result: YES.*
>
> **User:** *"Devi."* **Assistant:** *A name with no association to a specific medical case. Result: NO.*

A.3 Chain-of-thought Example Pool (MedGemma)

The following six examples were explored during iterative prompt development for MedGemma; the four used in the final configuration are marked with (*).

> (*) **User:** *"This study aimed to evaluate the SARS-CoV-2 genome sequences from COVID-19 cases and to characterize their genealogical networks to demonstrate possible routes of spread in Japan."* **Assistant:** *The text refers to tracing genealogical networks back to specific COVID-19 cases and spread routes. Result: YES.*
>
> **User:** *"Other variants, such as B.1.621, B1.1.318, B.1.1.519 were also found circulating in central Texas at <5% of the total sequenced specimens."* **Assistant:** *While percentage-based, it identifies specific viral variants linked to a precise geographic location (Central Texas). Result: YES.*
>
> **User:** *"We thought that we would have a greater opportunity of obtaining high-quality PCR products and get better NGS results when high viral load containing samples are used."* **Assistant:** *The text refers to the selection of specific clinical samples directly tied to individual patient cases. Result: YES.*
>
> (*) **User:** *"To accept the results, a Ct value for RNAse P was used as an endogenous internal amplification control in each reaction."* **Assistant:** *Laboratory quality control criteria with no patient data. Result: NO.*
>
> (*) **User:** *"Due to the above-mentioned findings in favor of progression, favipiravir treatment was started."* **Assistant:** *Medical treatment tied to an implied patient's clinical status. Result: YES.*
>
> (*) **User:** *"Devi."* **Assistant:** *A name with no association to a specific medical case. Result: NO.*