Lithuanian Medical Decision Models

Linas Petkevičius

Artificial Intelligence Methods Lab, Institute of Computer Science
Faculty of Mathematics and Informatics, Vilnius University
Corresponding author: linas.petkevicius@mif.vu.lt

Artificial Intelligence Methods Lab

Follow us: aiml.lt

Main Contribution

First Lithuanian models for typed biomedical decisions, released openly on Hugging Face.

- **Three models**, one benchmark, one protocol: MedGemma, Laya, Qwen.
- **Probabilities, not text:** one forward pass answers all questions on given options.
- **Calibrated:** test ECE below 0.005 for the in-domain models.
- **Evidence:** in-domain Lithuanian fine-tuning beats cross-lingual transfer.

94.6% decision accuracy **72.3%** abstracts fully correct

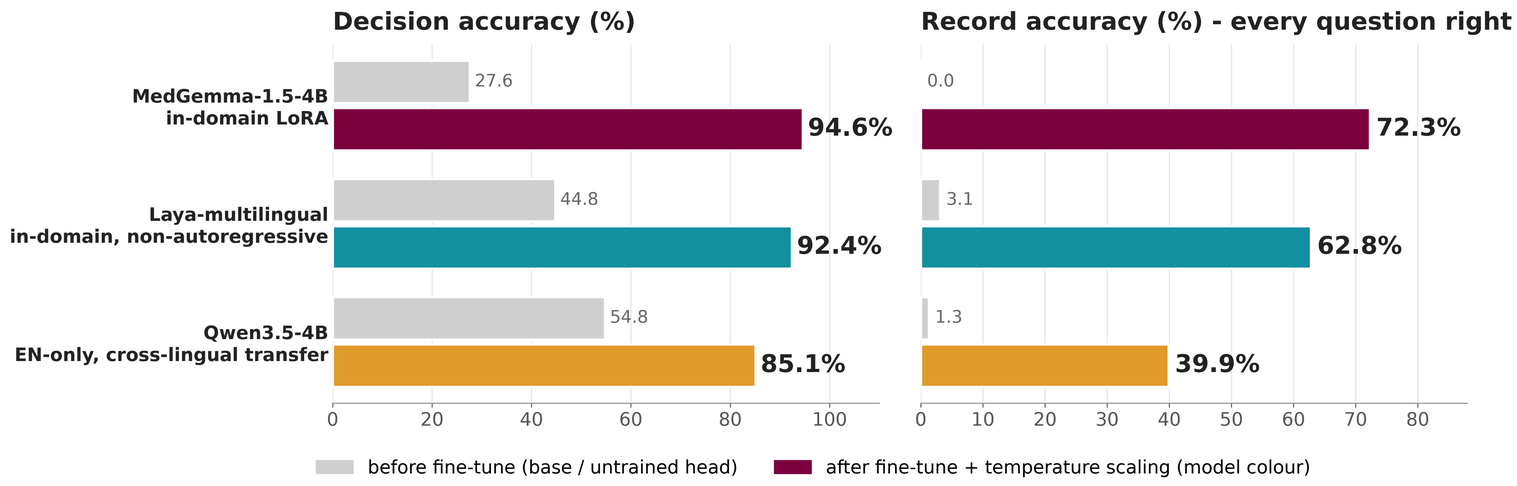

Figure 1: MEDLINE-S1-LT test split (4,857 abstracts, 29,654 decisions). Record accuracy counts an abstract only if *every* question on it is right.

How It Works

Lithuanian title + abstract
MEDLINE, translated EN→LT with NLLB

↓

Backbone + LoRA (r = 16)
MedGemma 1.5 4B · Laya (mmBERT)
Qwen3.5 4B

↓

Decision head
scores every option in one pass

↓

Calibrated option probabilities
temperature fitted on a held-out calibration split

MEDLINE-S1-LT Benchmark

38,950 training abstracts	**236,720** training decisions
4,857 test abstracts	**29,654** test decisions

Tests on NLM MeSH type data [2, 6].

In-Domain Beats Transfer

+32.4 pts
record accuracy with Lithuanian data

MedGemma (LT-trained) vs. Qwen (English-trained)

Model	Accuracy	Record	ECE
MedGemma-LT	**94.6**	**72.3**	0.005
Laya-LT	92.4	62.8	**0.004**
Qwen-LT (transfer)	85.1	39.9	0.046

Trustworthy Probabilities

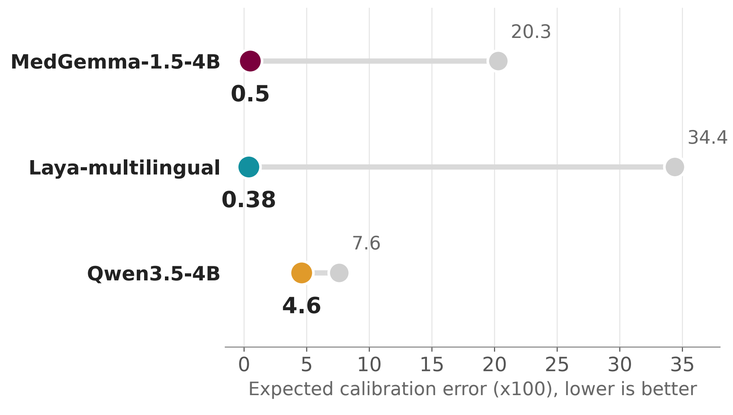

Figure 2: Calibration error before (grey) and after fine-tuning with temperature scaling. Laya drops from 0.34 to 0.004.

Cross-Lingual Transfer Works

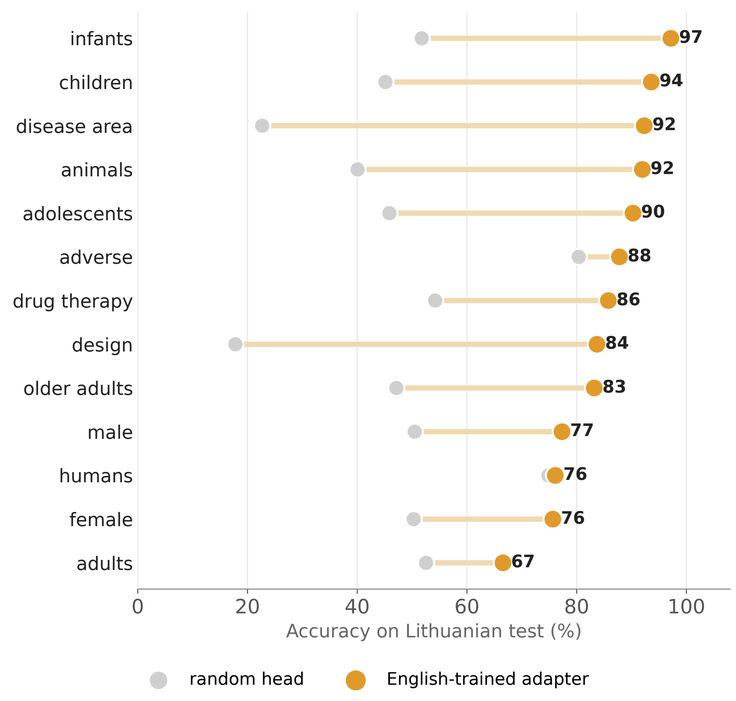

Figure 3: Qwen3.5-4B trained on *English only*, tested on Lithuanian: 85.1% vs. 54.8% for an untrained head.

Take-Aways

- **Best:** MedGemma-LT, 94.6% accuracy.
- **Smallest:** Laya-LT, 0.3B params, 92.4%.
- **No LT data?** Transfer still reaches 85%.
- **Research only**, not a medical device.

Models on Hugging Face

medgemma-1.5-4b-it-decitions-lt-med

laya-multilingual-lt-med

Qwen3.5-4B-LT-med

References

[1] Convai Innovations.
Laya: non-autoregressive system one decision models.
https://github.com/NandhaKishorM/laya, 2026.

[2] P. Deka.
From retrieval to typed decisions: Calibrated system one models from biomedical sentence encoders.
arXiv:2610.02486, 2026.

[3] Google Research and Google DeepMind.
MedGemma technical report.
arXiv:2507.05201, 2025.

[4] Google Research and Google DeepMind.
MedGemma 1.5 technical report.
arXiv:2604.05081, 2026.

[5] JHU CLSP.
mmBERT-base multilingual encoder.
arXiv:2509.06888, 2025.

[6] NLLB Team.
No language left behind: Scaling human-centered machine translation.
arXiv preprint arXiv:2207.04672, 2022.

This project has received funding from the Research Council of Lithuania (LMTLT), agreement No S-ITP-26-11.