



CU-ICU: Customizing Unsupervised Instruction-Fine Tuned Language Models for ICU Datasets via Text-to-Text Transfer Transformer

Teerapong Panboonyuen

Chulalongkorn University

<https://kaopanboonyuen.github.io/>



Why I Created CU-ICU

Because every second in the ICU matters.



The Problem

- ICU staff face **information overload**, high stress, and critical decisions under time pressure.



My Vision

- **Build AI tools that think like an ICU assistant**
- Reduce cognitive burden, **free time for patient care**



Why I Created CU-ICU (Cont.)



The Solution: CU-ICU

- A specialized **Large Language Model (LLM)** for ICU workflows
- **Understands Thai medical context** and national guidelines
- Capable of **clinical Q&A, sepsis alerts, reasoning, and protocol suggestions**

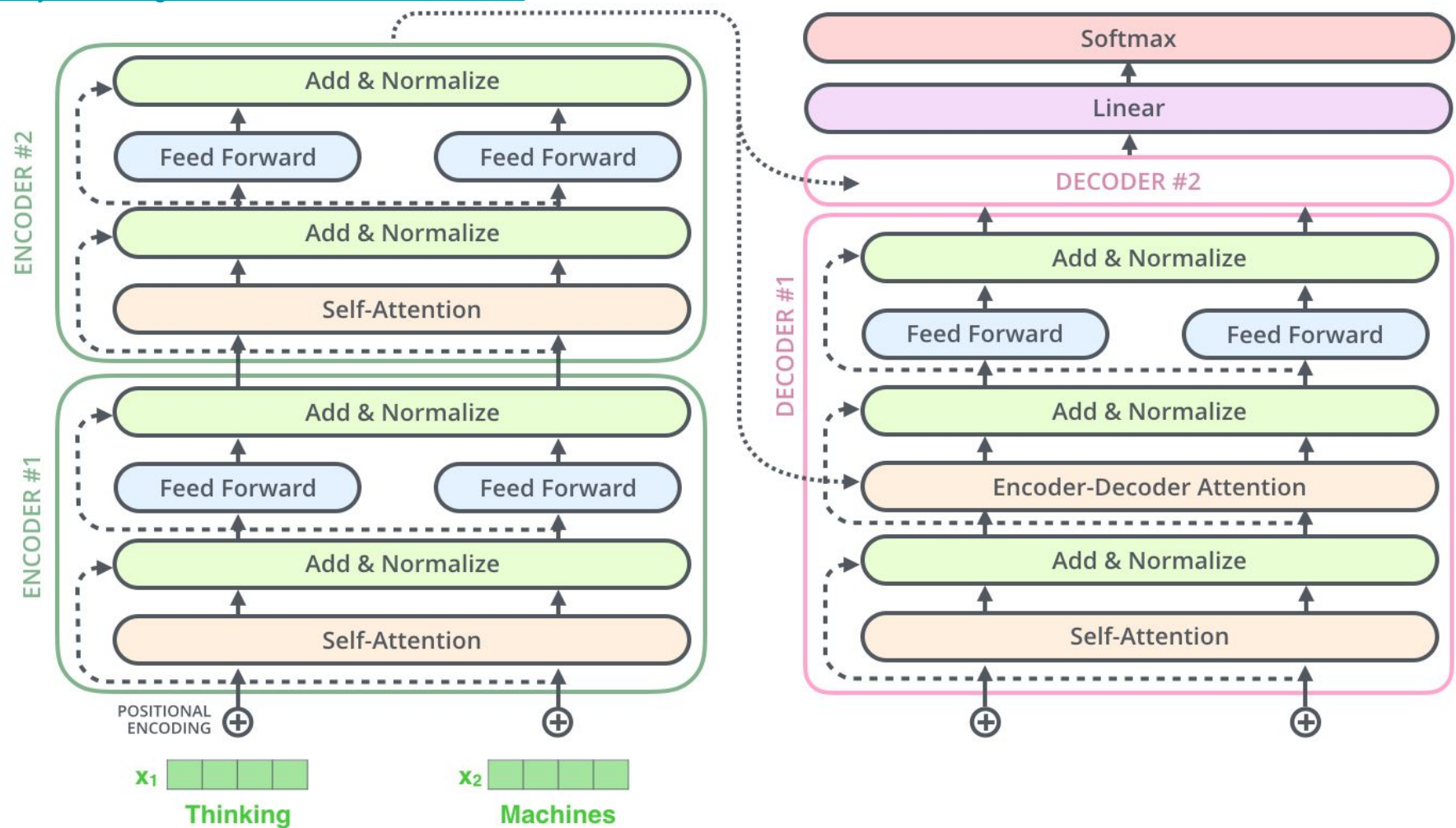


AI should 'not replace' doctors — it should amplify their impact.

What is **CU-ICU**?

CU-ICU is a Thai-adaptable AI (LLM) assistant built on:

-  **Flan-T5**, an instruction-tuned transformer
-  Trained with ICU logic from real data
-  Fine-tuned via **LoRA, AdaLoRA, or IA³** (efficient, few parameters)
-  Equipped with a custom **CU-ICU Loss** to reduce hallucination and increase clarity

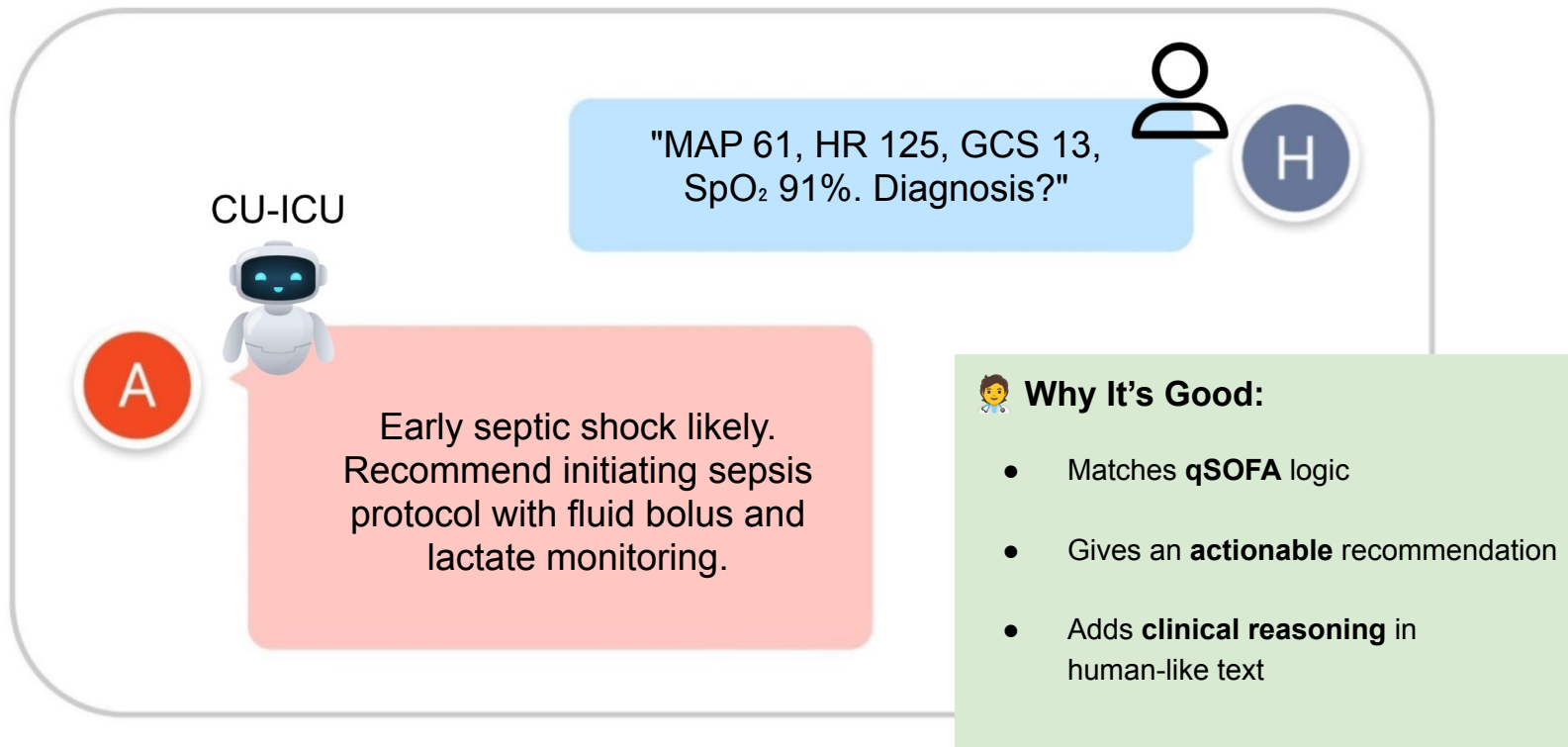


Dataset Used – HiRID from Switzerland

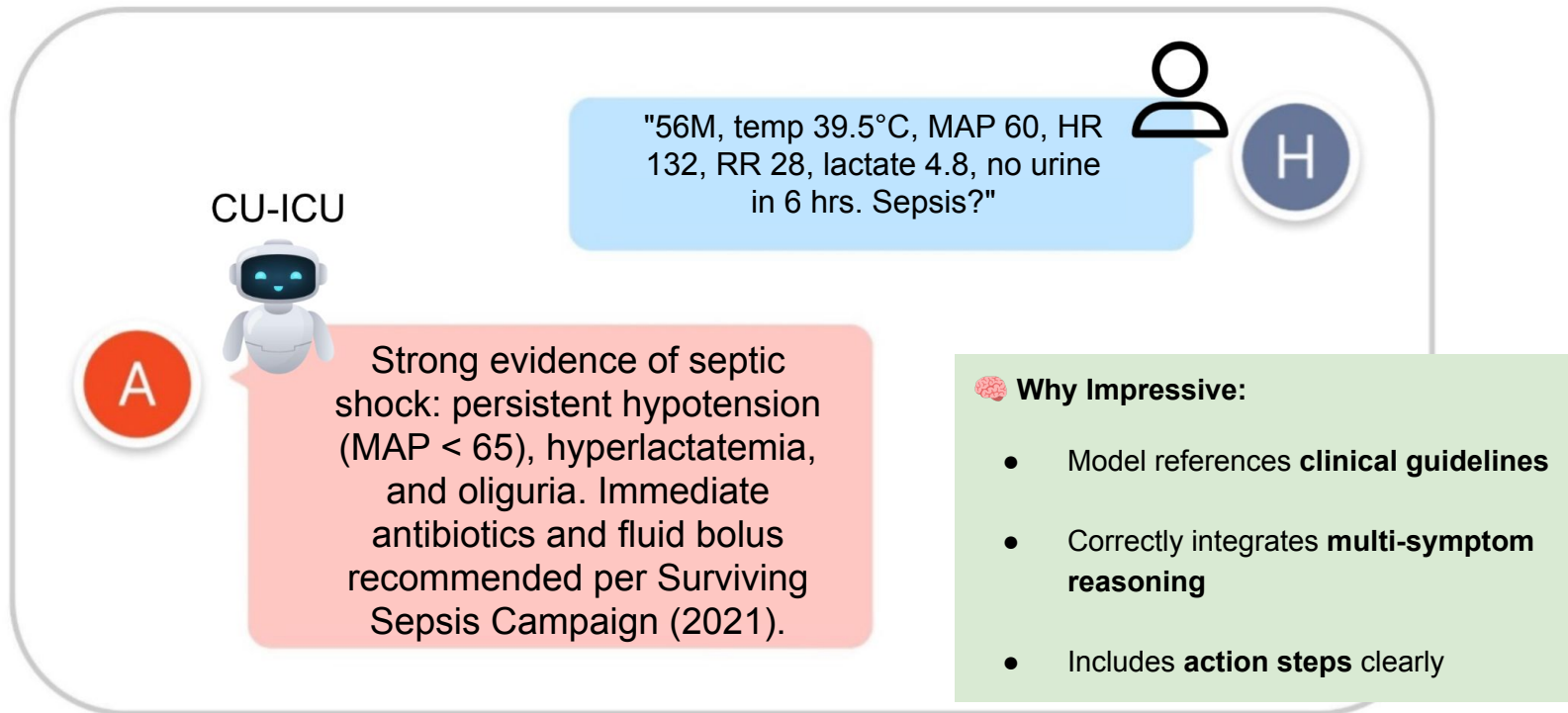
Property	Detail
Name	HiRID (High-Resolution ICU Dataset)
Source	Bern University Hospital, Switzerland
Access	https://physionet.org/content/hirid/1.1.1/
Size	~33,000 patient ICU episodes
Type	Vitals (per minute), interventions, nurse charts
Language	English

👉 We translated logic into Thai context via **instruction-based prompts**, creating clinical Q&A datasets.

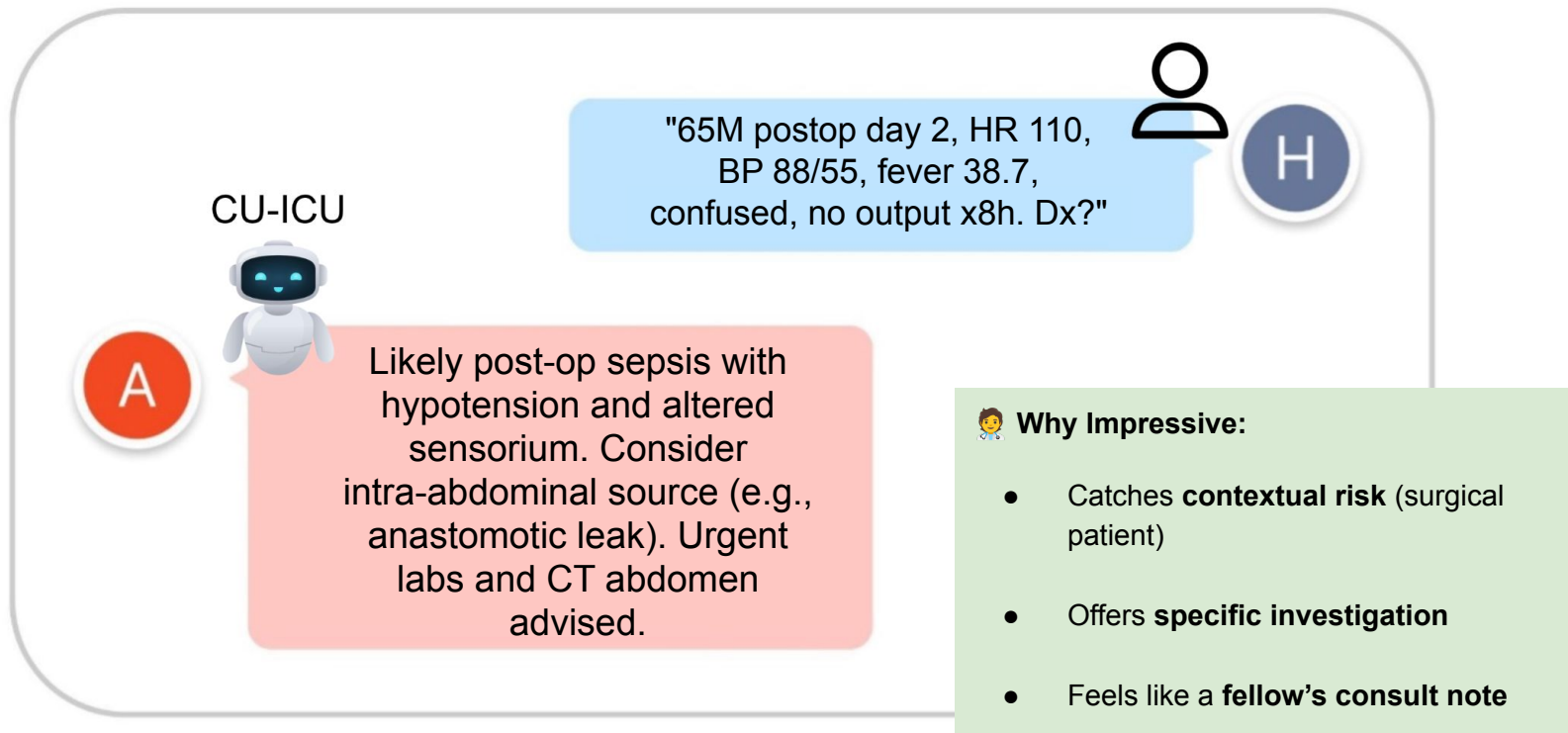
Q&A Example for ICU Prompt (Powered by CUICU LLM)



Q&A Example for ICU Prompt (Powered by CUICU LLM)



Q&A Example for ICU Prompt (Powered by CUICU LLM)



Q&A Example for ICU Prompt (**Powered by CUICU LLM**)



CU-ICU



How to reduce nurse burnout in ICU?

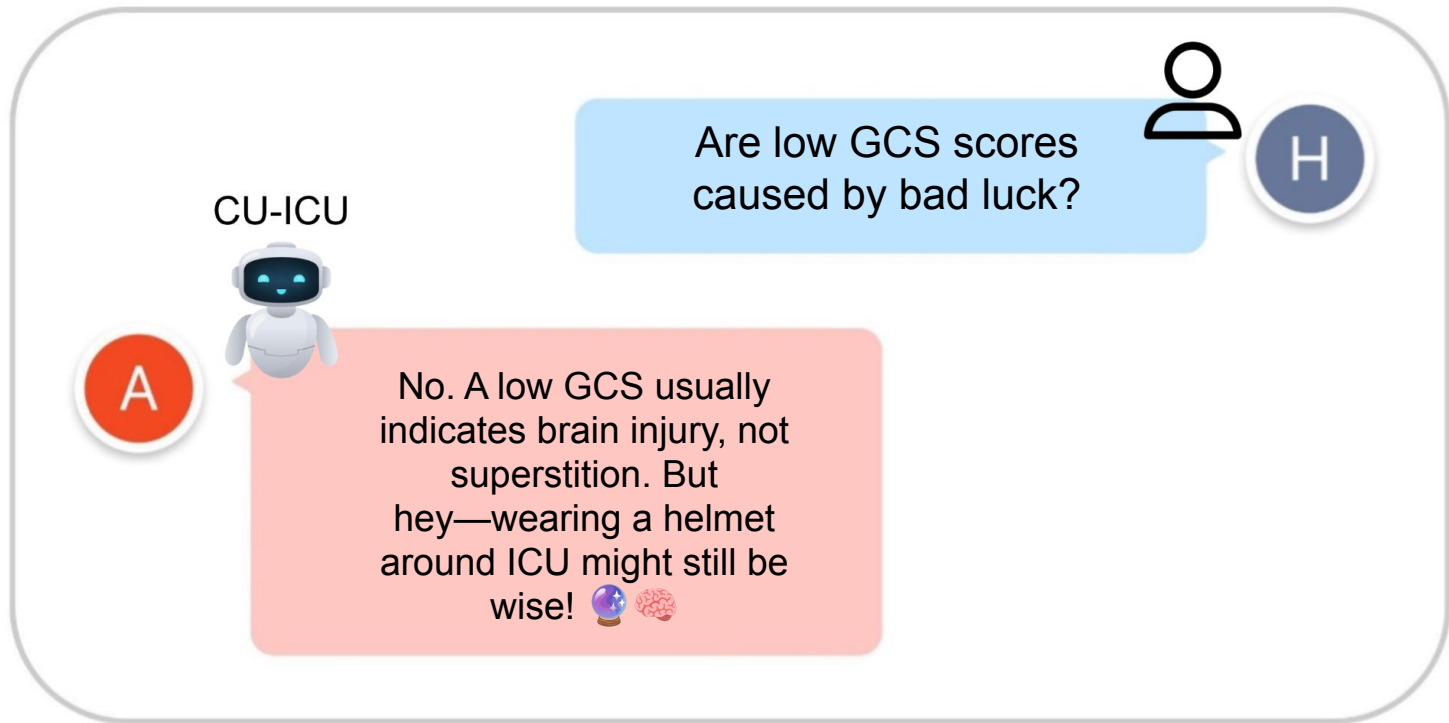


Schedule karaoke hour during night shift to improve morale.

 Explanation:

- This was from a synthetic set for robustness testing.
-  *Humor included gently*

Q&A Example for ICU Prompt (**Powered by CUICU LLM**)



Column Name	Description
Method	PEFT technique used: LoRA, AdaLoRA, or IA ³
Config	Hyperparameter setting for that method (e.g., Rank, Budget, Scope)
Params (%)	% of total model parameters updated during fine-tuning (lower = more efficient)
Sepsis Acc.	Accuracy on binary classification task: Early sepsis detection
Mortality Acc.	Accuracy on binary classification task: ICU mortality prediction
Note nBERTs	Normalized BERTScore: measures quality of generated clinical explanations
Avg	Average of all three task scores for cross-task comparison (unweighted)

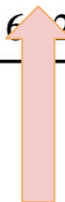
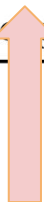
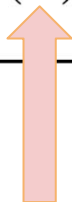
Evaluation Details

✓ Classification Tasks (Sepsis & Mortality)

- **Metric:** Accuracy (%)
- **Input:** prompt + patient vitals/symptoms
- **Target:** Binary label (e.g., "septic" or "not septic")
- **Evaluation:** Majority prediction from model-generated answer

Full comparison of FLAN-T5 fine-tuning with PEFT

Method	Config	Params (%)	Sepsis Acc.	Mortality Acc.	Note nBERTs	Avg
3*LoRA	Rank=4	1.2%	79.4 $\pm$ 2.1	76.8 $\pm$ 2.3	23.9 $\pm$ 1.9	60.0
	Rank=8	3.1%	81.2 $\pm$ 1.8	78.5 $\pm$ 2.0	26.4 $\pm$ 1.7	62.0
	Rank=16, Drop=0.1	6.2%	83.1 $\pm$ 1.5	79.6 $\pm$ 1.8	28.3 $\pm$ 1.5	63.7
3*AdaLoRA	Budget=0.5, Init Rank=4	2.1%	82.2 $\pm$ 1.9	78.4 $\pm$ 2.1	27.1 $\pm$ 1.8	62.6
	Budget=1.0, Init Rank=8	2.9%	83.5 $\pm$ 1.6	80.9 $\pm$ 1.9	29.8 $\pm$ 1.3	64.7
	Budget=1.5, Init Rank=16	5.4%	84.2 $\pm$ 1.3	80.3 $\pm$ 1.7	30.6 $\pm$ 1.4	65.0
2*(IA) ³	Default (All Layers)	0.9%	85.6 $\pm$ 1.5	80.2 $\pm$ 1.7	32.1 $\pm$ 1.5	66.0
	Reduced Scope (Last 6)	0.5%	83.8 $\pm$ 1.7	78.7 $\pm$ 2.0	30.2 $\pm$ 1.6	64.2



Evaluation Details

Text Generation Task (Note Generation)

- **Metric:** Normalized BERTScore (nBERTs), Accuracy
- **Input:** Question (e.g., “What to do for low urine output?”)
- **Target:** Short free-text clinical explanation
- **Evaluation:** Compare model output to gold reference using **Accuracy** and normalize across seeds

Performance Highlights

IA³ achieves the **highest average score (66.0)** with **only 0.9% parameters** trained.

AdaLoRA performs second best with strong control over trade-offs via budget tuning.

LoRA shows consistent gains as rank increases, but with higher parameter cost.



Chula
The Second Century Fund (C2F)

CU-ICU LLM DEMO

Future Work



Next steps:

- Thai-specific EMR fine-tuning
- Multimodal input (X-ray, ECG + text)
- Integration with nursing dashboards

Conclusion

● **CU-ICU: Smarter Care, Faster Decisions**

- ✓ AI-LLM-powered assistant for ICU teams
- ✓ Supports **critical decision-making** with real-time insights
- ✓ **Thai-context optimized** for language & clinical guidelines
- ✓ Reduces **workload**, improves **patient outcomes**
- ✓ Designed with **ethics, safety, and human-in-the-loop** approach



Goal: Empower healthcare, not replace it.



Thank You for Your Attention!



Stay Curious. Stay Critical. Stay Compassionate.



Have more questions?

Let's keep learning, coding, and caring — together.

Teerapong Panboonyuen

<http://kaopanboonyuen.github.io/>

