

Evaluation:

**Patient-Physician
Conversation Data**

Real-world data,
from an online medical consultation website, HealthCareMagic

Manually and **Automatically**

- ✓ remove **doctor** and **patient** identity information
- ✓ correct grammatical errors (use tools)

Data
for
Fine-tuning

HealthCareMagic-100k 100k

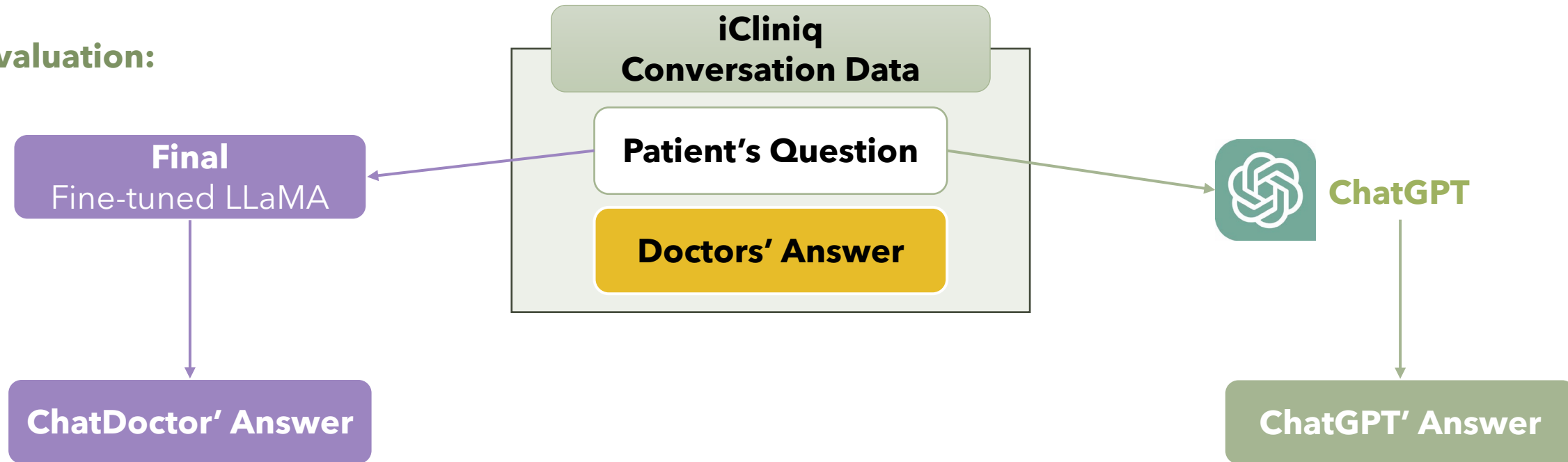
Data
for
**Performance
Evaluation**

**iCliniq
Conversation Data** 10k

Quantitative Evaluation

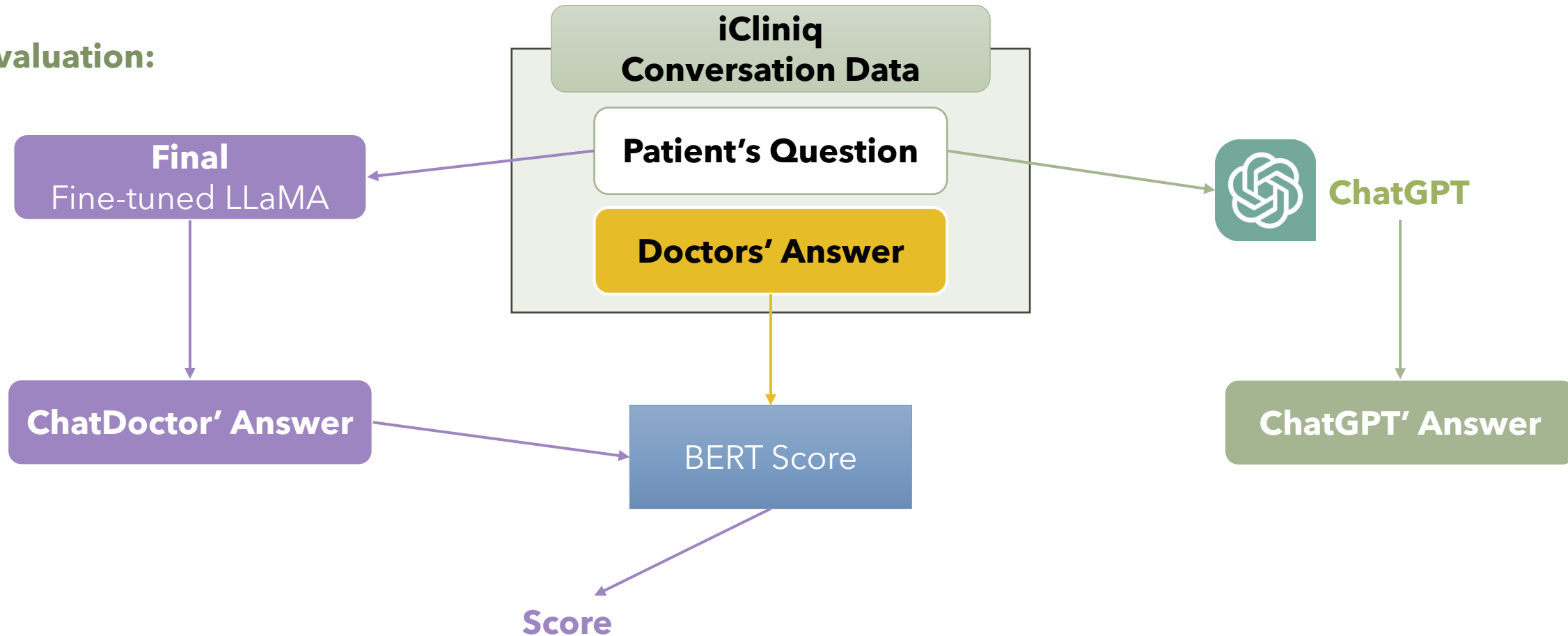
ChatDoctor: A Medical Chat Model Fine-tuned on LLaMA Model using Medical Domain Knowledge

Evaluation:



ChatDoctor: A Medical Chat Model Fine-tuned on LLaMA Model using Medical Domain Knowledge

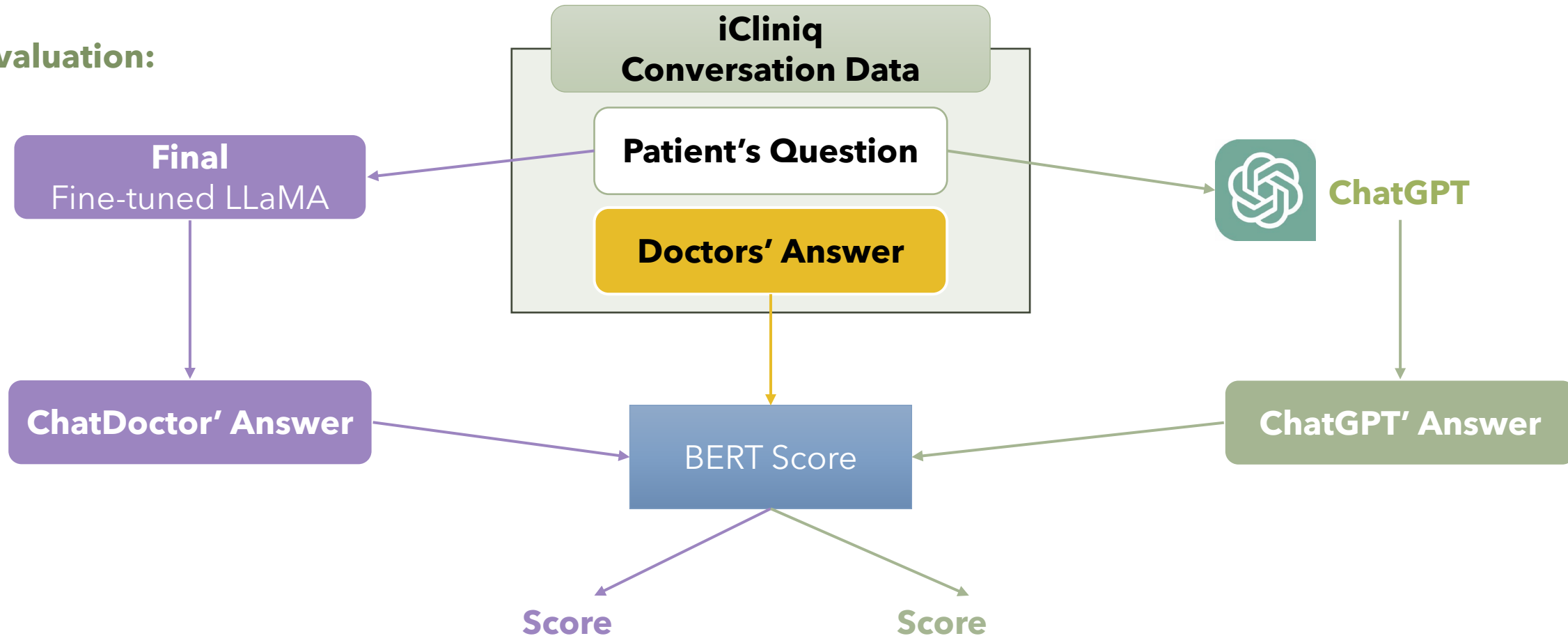
Evaluation:



Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. Bertscore: Evaluating text generation with bert. In International Conference on Learning Representations, 2020.

ChatDoctor: A Medical Chat Model Fine-tuned on LLaMA Model using Medical Domain Knowledge

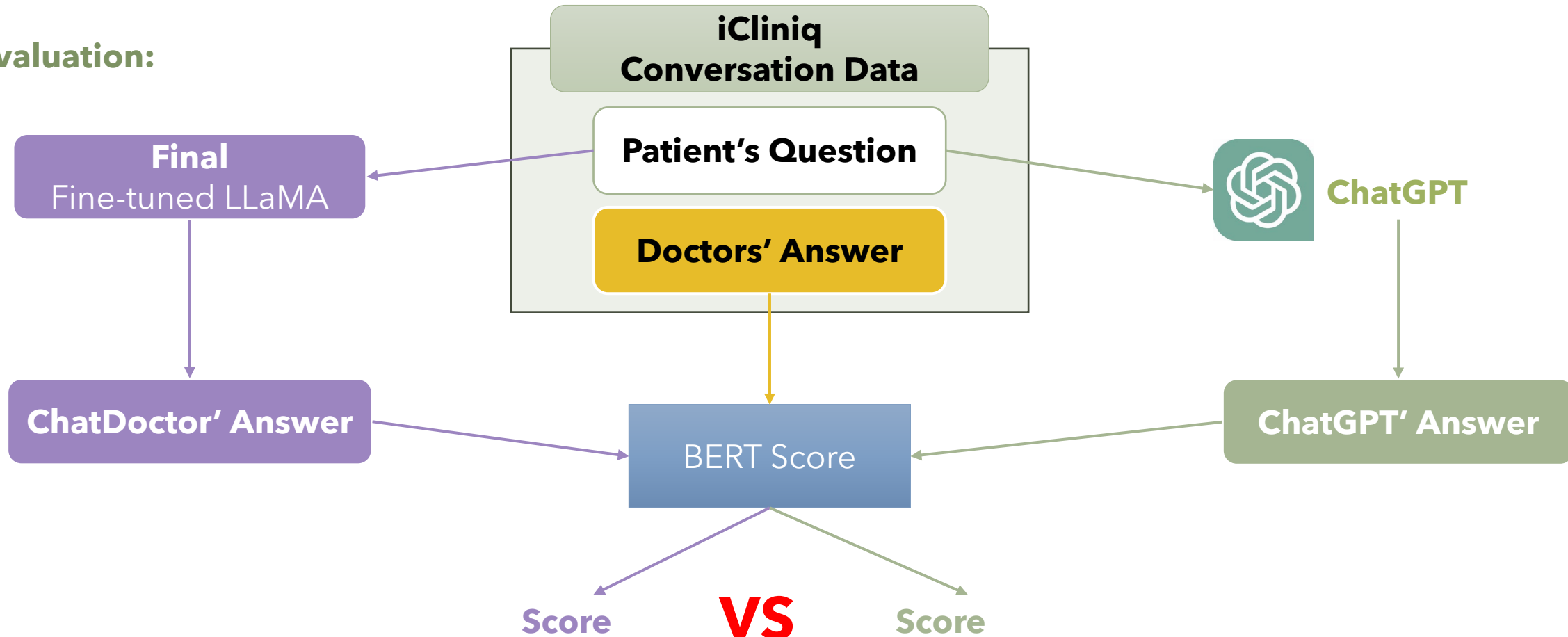
Evaluation:



Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. Bertscore: Evaluating text generation with bert. In International Conference on Learning Representations, 2020.

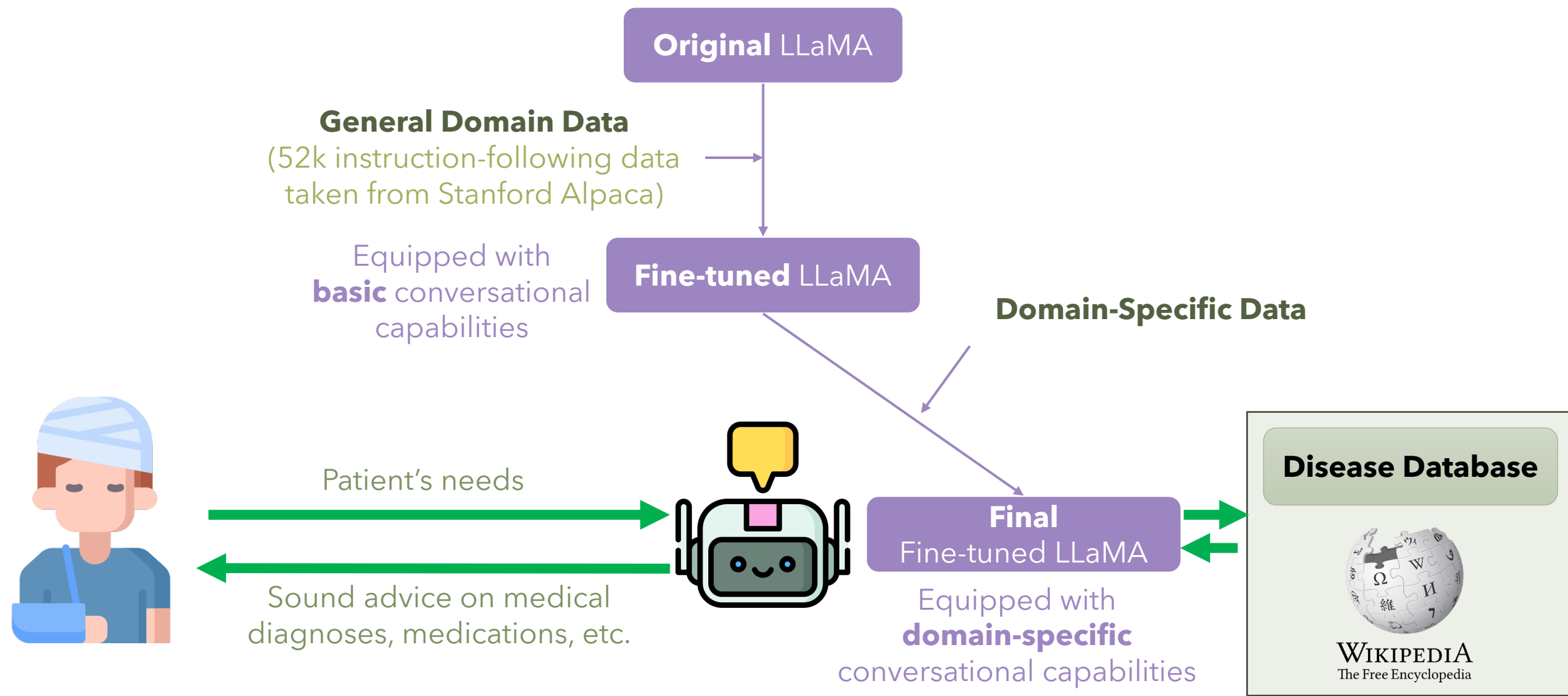
ChatDoctor: A Medical Chat Model Fine-tuned on LLaMA Model using Medical Domain Knowledge

Evaluation:



Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. Bertscore: Evaluating text generation with bert. In International Conference on Learning Representations, 2020.

ChatDoctor: A Medical Chat Model Fine-tuned on LLaMA Model using Medical Domain Knowledge



ChatDoctor: A Medical Chat Model Fine-tuned on LLaMA Model using Medical Domain Knowledge