**Title :  Enhancing Domain Specific Task(medical Knowledge)  in Large Language Models**

**Duration :** 6 month  (Jan. ~ June)

**Result :** Open source finetuned LLM model and papers

**Internship :**  1 ~ 2 day per 1 week / On-Line, 1~2 offline meet up per 1 month

**Detail :** Recent advancements in Large Language Models (LLMs) like ChatGPT, LLaMA2, PaLM and their specialized variants have revolutionized various domains.   Open source PMC-LLaMa-13B achieves 56.36% and Med-PaLM 2 achieves an accuracy of 86.5% on USMLE-style questions and in Medical domain. However, these models often struggle with maintaining consistent accuracy in different situations.   This proposal focuses on evaluating and enhancing PMC-LLaMA2-13B which is opensourced domain adapted pretrained and finetuned with USMLE dataset particularly its adaptability to varied question formats in domain-specific applications.

To evaluate the performance of PMC-LLaMA2-13B on non-standard problem formats, particularly comparing its accuracy in 4-option versus 5-option question formats.

To develop and implement domain adaptation and fine-tuning methodologies for improving the model's performance across diverse formats. Moreover, we would evaluate similar dataset for Korean languages(optional).

Model Evaluation: Conduct a comprehensive analysis of PMC-LLaMA2-13B's performance using datasets with varied problem structures (4-option and 5-option formats) and shuffled options. This will identify the model's limitations and areas for enhancement. This proposal focuses on evaluating and enhancing PMC-LLaMA2-13B, particularly its adaptability to varied question formats in domain-specific applications.

**Reference :**

* Med-PaLM <https://arxiv.org/abs/2212.13138> Large Language Models Encode Clinical Knowledge
* PMC-LLaMA  <https://arxiv.org/abs/2304.14454> Towards Building Open-source Language Models for Medicine
* Med-PaLM2 <https://arxiv.org/abs/2305.09617> Towards Expert-Level Medical Question Answering with Large Language Models

**Methodology**

* Build virtual lab with NVIDIA AI Tech Center Korea and solve the problem.
* Orientation, Problem Statement, Literature Survey, evaluate(duplicate) PMC-LLaMA2-13B-instruct, propose method, train model
* assessment, solve the problem with assessment, presentation/report
* we will prepare own evaluation and finetune dataset based on S2ORC, MedQA-USMLE, MedMCQA, PubMedQA dataset
* we will investigate enhancing method with domain adaptation and fine-tune method.

**Requirement :**

* Understand Causal LLM,
* experience with pytorch for DL training
* handling unstructured datasets with python/json

**Ways to stand out from the crowd :**

* Python coding skills and experience
* Experience working for DL training such as Alpaca/Vicuna such as lmsys
* Experience working with PEFT and NeMo frameworks