# 가상환자 대화 시스템 설명과 실용성

It has been recommended that more explicit training should be provided in undergraduate medical education on applying clinical reasoning skills, to reduce the impact of future diagnostic errors and potential patient harm [The effectiveness of using virtual patient educational tools to improve medical students’ clinical reasoning skills: a systematic review].

Use of virtual patient educational tools could fill the current gap in the teaching of clinical reasoning skills. Currently, teaching of clinical reasoning in most medical schools in the UK remains a largely implicit component of small group tutorials, problem-based learning, clinical communication skills sessions, and clinical placements [3]. Making the teaching of these skills more explicit may help students to refect on their skills, which many models of learning suggest is essential for improving skills [6, 7]. Virtual patient educational tools are becoming increasingly popular in medical education and have been used to explicitly teach clinical reasoning skills [5, 8, 9]. Tey are defned as “A specifc type of computerbased program that simulates real-life clinical scenarios; learners emulate the roles of health care providers to obtain a history, conduct a physical exam, and make diagnostic and therapeutic decisions”. Tey allow students to practise clinical reasoning with realistic patients, in a safe environment [5, 10]. Tey may also be particularly suited to providing training on clinical reasoning skills that require deliberate practice with a wide variety and large number of clinical cases. Indeed, many students may have limited contact with patients, where it is also not possible to pre-determine what range of presentations and problems students will meet [5]. Educational and cognitive theories, and empirical research also suggest that virtual patient educational tools could provide an ideal platform for developing clinical reasoning skills if they incorporate best practice features for simulationbased educational tools, in particular providing opportunities for feedback and refection [6, 7, 10, 11]

[ The efectiveness of using virtual patient educational tools to improve medical students’ clinical reasoning skills: a systematic review]

# 기존의 가상환자 시스템 설명

[Artificial intelligence in virtual standardized patients: Combining naturallanguage understanding and rule based dialogue management to improveconversational fidelity 2023]

[A French Medical Conversations Corpus Annotated for a Virtual Patient Dialogue System]

[Combining CNNs and Pattern Matching for Question Interpretation in a Virtual Patient Dialogue System]

[Designing a virtual patient dialoguesystem based on terminology-rich resources: Challenges and evaluation.

[Lessons learned fromthe usability evaluation of a simulated patient dialogue system.]

[Developing a conversational virtual standardized patient to enable stu-dents to practice History-Taking skills.]

[John K Haas. A history of the unity game engine. 2014.]

# 기존과 다르게 우리는 LLM을 사용해서 만들었다. 기존의 시스템과 비교하여 LLM의 장점

Our study was motivated by the fact that for such a domain-specific system, it was difficult to acquire large real-life data samples to increase the training database: this would require recruiting more patients, which is both time-consuming and costly. To address this gap, we have employed a neural large language model: ChatGPT version 3.5, to generate data solely for training our dialogue system.

Advancements in deep learning have enabled the development of online learning tools for medical training, which is important for remote learning. However, face-to-face interaction is essential for practicing human-centric skills such as clinical skills. Presently, in medical training, such interactions can be mimicked using deep learning methodologies. However, the understanding of such models is often limited whereby lightweight models are unable to generalize beyond scope while large language models tend to produce unexpected responses. To overcome this, we propose a hybrid lightweight and large language model for creating virtual patients, which can be used for real-time autonomous training of trainee doctors in clinical settings using online platforms. This ensures high-quality and standardized learning for all individuals regardless of location and background.

# 서론

가상환자 대화 시스템은 의료 교육 분야에서 혁신적인 도구로 큰 관심을 받고 있습니다. 이 시스템은 여러 가지 장점과 실용적인 응용 가능성을 가지고 있습니다. 실제 환자와의 상호작용을 모방함으로써 의료 전문가들과 의학 학생들은 안전한 환경에서 다양한 증상과 상황에 대한 훈련을 받을 수 있습니다. 이는 실제 환자에게 직접적인 영향을 미치지 않으면서도 실용적인 경험을 제공합니다. 뿐만 아니라 의료 전문가들의 임상 의사소통 능력 향상에 크게 도움이 됩니다.

Previous virtual patient dialogue systems have typically relied on rule-based and template-driven responses. While theycan mimic certain aspects of patient interactions, they oftenstruggle to capture the subtle language patterns of individualpatients, resulting in less natural and engaging interactions.These systems tend to be rigid and lack the adaptability to diverse medical scenarios or patient profiles, limiting theireffectiveness in simulating realistic patient experiences.

본 연구에서는 Large Language Model(LLM)을 사용하여 가상 환자 대화 시스템을 개발했습니다. LLM을 사용함으로써 자연스러운 커뮤니케이션에 대한 새로운 가능성을 제공하며, 인간과 유사한 텍스트를 생성하는 능력을 보였습니다. 특히 Llama에 우리가 직접 개발한 MP TUNING을 적용했습니다. 그림1에서 보이는 것처럼 multi-turn conversation을 하는 도중 매 turn마다 훈련받는 자가 정의한 특정 증상을 가지는 환자의 발화를 생성할 수 있게 만들었습니다. 이 시뮬레이션 기반 교육은 매 턴마다 구체적인 증상 제어를 함으로써 의료 전문가들이 실시간으로 자신의 행동을 분석하고 개선할 수 있는 환경을 제공합니다. 이는 훈련 효과를 극대화합니다.

Llm이 생성하는 문장이 특정 속성을 가지도록 제어하는 방법, 즉 특정 증상을 가진 환자의 발화를 생성하도록 하는 방법은 현재 활발하게 발전하고 있다. 현재 상용화된 chat기반 llm을 파인튜닝하거나 retraining하는 방법이 있다. 하지만 거대한 언어모델의 파라미터를 학습시키는 것은 쉽지 않은 일이다. 또한 prompt engineering을 통해 환자 발화의 속성을 제어할 수도 있다. 하지만 prompt의 까다로움이 존재한다. 즉 prompt를 똑바로 주지 않거나 언어 모델이 학습하지 않았던 특정 질병 도메인에서는 우리가 원하는 결과를 잘 얻지 못할 수 있다.

우리는 코로나19 dialogue dataset을 사용해 언어 모델의 prefix와 symptom prefix만 학습 시킴으로서 거대한 언어모델의 파라미터는 고정시키고 resource-efficient한 mp tuning을 만들었다. 위의 방법들과 비교했을 때 증상 제어 능력과 문장의 유창성 등을 평가했을 때 우리게 더 좋게 나오는 것을 보였다. 우리의 방법을 특정 질병에 적용시켜 의료계에서 사용한다면 의료 전문가들과 의학 학생들은 실전 시나리오에서 자신의 진료 능력을 향상시킬 수 있으며, 안전하고 효율적인 의료 서비스를 제공할 준비를 할 수 있습니다.

기여 3가지:

* Llama에 우리가 만든 mp tuning을 적용하여 매 발화마다 증상 제어를 할 수 있는 Virtural patient dialogue system을 만들었다.
* while freezing the parameters of LLM, we focus on training only a small number of parameters in prefix modules, 즉 prefix와 symptom prefix 를 학습시키며 resource-efficient한 방법을 만들었다.
* Human-like한 자연스러운 문장과 좋은 증상 제어 능력을 바탕으로 실용적으로 medical educaiton에 쓰일 수 있다는 것을 보였다.

12/7 introduction

% 가상 환자 대화 시스템 필요성 및 사용성

Virtual patient dialogue systems are gaining significant attention as innovative tools in the field of medical education. These systems have various advantages and practical applications. By simulating interactions with real patients, medical professionals and students can receive training in a safe environment, covering various symptoms and situations. This provides practical experience without directly impacting real patients. Additionally, it greatly aids in improving the clinical communication skills of medical professionals.~\cite{Taglieri87, Berman2016TheRF}

% 이전 가상 환자 대화 시스템 모델들 설명

Previous virtual patient dialogue systems have typically relied on rule-based and template-driven responses. While they can mimic certain aspects of patient interactions, they often struggle to capture the subtle language patterns of individual patients, resulting in less natural and engaging interactions. These systems tend to be rigid and lack the adaptability to diverse medical scenarios or patient profiles, limiting their effectiveness in simulating realistic patient experiences.

% 우리는 LLM에 MP-tuning 적용해서 매 턴마다 증상제어가 가느안 효율적인 훈련 시스템을 만들었다

% 우리 메소드를 적용한 것이 LLM이라고 해야 되는지 Llama라고 해야 되는지 정해야 할 듯? we applied our developed technique, MP TUNING, to Llama[].

In this study, we developed a virtual patient dialogue system using Large Language Models (LLMs). The use of LLMs provides new possibilities for naturalistic communication, demonstrating the ability to generate text that is human-like[]. Specifically, we applied our developed technique, Turn by turn prompting(TTP), to Llama~\cite{touvron2023llama}. As shown in Figure~\ref{fig:dialogue\_comparision}, during multi-turn conversations, we enabled the system to generate patient responses with specific symptoms defined by the trainee at each turn. This simulation-based training offers an environment for medical professionals to analyze and improve their actions in real-time by controlling specific symptoms at each turn, maximizing the effectiveness of the training.

% 다른 방법(비교모델)들의 까다로움

The methods for controlling the sentences generated by LLMs to exhibit specific attributes, specifically generating patient responses with particular symptoms, are currently evolving. One approach is to fine-tune or retrain commercially available chat-based LLMs[]. However, training the parameters of large language models is a challenging task. Additionally, prompt engineering[] can be employed to control the attributes of patient responses. Nevertheless, the difficulty lies in the trickiness of prompts, meaning that not providing a proper prompt or dealing with specific disease domains not learned by the language model could hinder achieving the desired results.

% 위와 비교했을 때 장점과 높은 성능 -> 좋은 서비스를 제공할 수 있다는 결론

We applied a resource-efficient TTP by training only the prefixes and symptom tokens of the language model using the COVID-19 dialogue dataset[], thereby freezing the parameters of the large language model. Compared to the methods mentioned above, our approach demonstrated better symptom control ability and fluency in sentences. If our method is applied in specific medical conditions, it can enhance the clinical skills of healthcare professionals and medical students in real-world scenarios, enabling them to be prepared to provide safe and efficient medical services.

구체적으로 **가상 환자의 특정 응용 분야,** **가상 환자의 협업과 팀 기반 교육**

**우리 방법이 개인정보 및 보안에도 좋다는 거(prefix tuning 논문에서 인용하기 )**

\input{figures/dialogue}

% 우리의 기여는 이러이러하다 1) 2) 3)...

Our main contributions in this work are listed as follows:

\begin{itemize}

\item We created a Virtual Patient Dialogue System capable of symptom control for every utterance by applying our TTP to Llama.

\item While freezing the parameters of LLM, we specifically trained a minimal number of parameters in the prefix modules, namely the prefix and symptom tokens, developing a resource-efficient method.

\item Demonstrating human-like natural language and effective symptom control, we illustrated its practical applicability in medical education.

\end{itemize}

#관련 연구

1. 기존 Virtual patient

기존 가상 환자들은 의료 교육과 임상 훈련 분야에서 많은 노력과 연구를 통해 발전해 왔습니다. [7]developed a virtual patient dialogue system employing a rule-based approach and utilizing terminology-rich resources. [8]employed a knowledge model for history-taking and a termino-ontological model for handling out-of-vocabulary terms, al-lowing for the creation of more diverse virtual patient sce-narios. [9] integrated a virtual patient conversation system byintegrating ChatScript, a dialogue management system, withthe Unity game engine [10].

[11,12,13]는 chatted or typed conversations를 사용했다.

[14]는 ITS(Intelligent Tutoring System)와 NLP 기술을 결합함으로써 Hepius는 의과 대학생들에게 진단 추론을 훈련시키기 위한 학습 도구를 제공할 수 있습니다.

이런 전통적인 대화형 가상환자는 have been successfully used to train students from a variety of healthcare professions[15,16] Nonetheless, these researches often rely on elementary rule-based systems or operate on small-scale neural networks, leading to limitations in the diversity, fluency and randomness of patient dialogues.

또한 turn 마다 구체적인 증상을 제어하거나 오직 llm을 사용해서 환자의 발화를 생성하는데는 한계가 있었다.

*.[7] Designing a virtual patient dialoguesystem based on terminology-rich resources: Challenges and evaluation.Natural Language Engineering, 26(2):183–220, 2020*

*[8] Leonardo Campillos-Llanos, Catherine Thomas, ́Eric Bilinski, AntoineNeuraz, Sophie Rosset, and Pierre Zweigenbaum. Lessons learned from the usability evaluation of a simulated patient dialogue system.Journalof Medical Systems, 45(7):69, May 2021*

*[9] Kellen Maicher, Douglas Danforth, Alan Price, Laura Zimmerman,Bruce Wilcox, Beth Liston, Holly Cronau, Laurie Belknap, CynthiaLedford, David Way, Doug Post, Allison Macerollo, and Milisa Rizer.Developing a conversational virtual standardized patient to enable stu-dents to practice History-Taking skills.Simul Healthc, 12(2):124–131,April 2017*

*[10] John K Haas. A history of the unity game engine. 2014*

*[11] Kleinheksel AJ.2014. Transformative learning through virtual patientsimulations: predicting critical student reflections. Clin Simul Nurs.10(6):e301–e308*

*[12] Hirumi A, Kleinsmith A, Johnsen K, Kubovec S, Eakins M, Bogert K,Rivera-Gutierrez DJ, Reyes RJ, Lok B, Cendan J.2016. Advancing vir-tual patient simulations through design research and interPLAY:part I: design and development. Education Tech Research Dev.64(4):763–785*

*[13] Liu T, Luo J, He H, Zheng J, Zhao J, Li K.2018. History-taking instruc-tion for baccalaureate nursing students by virtual patient training: aretrospective study. Nurse Educ Today. 71:97–104*

*[14] A Natural Language Processing–Based Virtual Patient Simulator and Intelligent Tutoring System for the Clinical Diagnostic Process: Simulator Development and Case Study*

*[15] Motz P, Gray M, Sawyer T, Kett J, Danforth D, Maicher K, Umoren R.2018. Virtual antenatal encounter and standardized simulation*

*[16] Jackson J, Park S.2020. Creating intelligent virtual patients for simula-tion-based pharmacy education through human-centered conversa-tion modelling. IJSMARTTL. 2(2/3):96–114*

1. Controllable text generation 방법들

자연어 처리(NLP) 분야에서 대규모 언어 모델(LLM)은 텍스트 생성, 언어 번역, 질문 답변 등 다양한 작업에서 탁월한 성능을 보여주었습니다. 그러나 LLM은 종종 제어하기 어려운 경우가 있기 때문에 텍스트 생성 분야에서 Controllable Text Generation(CTG)에 대한 연구가 활발히 이루어지고 있습니다[17]. 대표적인 방법으로는 다음과 같은 방법들이 있습니다

% finetuning 방법들(prompt, instruct based)

Methods based on prompts[18,19,20. Acdp 논문도] make use of pre-trained LLMs to direct the creation of contolled text in the fine-tuning phase. They capitalize on the attributes of PLMs to improve controllability through the choice of suitable prompts. 대표적인 방법으로 In [21], they propose a technique called "prefix tuning." This method fixes the parameters of a LLMsand uses backpropagation to optimize a small, task-specific vector called a "prefix." The knowledge acquired from this prefix, also known as a "prompt," allows the PLM to generate the desired text, thereby improving controllability to some extent.

InstructGPT [94], a most notable recent work, utilizes instruction tuning to control the language model and generate desired human-like content. It starts with collecting a dataset of labeled demonstrations of the desired model behavior, which are then used as instructions to fine-tune GPT3

% Retrain 방법

According to the characteristics of a specific downstream task, it is also feasible to change the original architecture of PLMs or retrain a large conditional language model from scratch. This kind of approach is promising to substantially improve the quality and controllability of text generation, but is limited by increased computing resource consumption and the lack of sufficient labeled data. CTRL [58] is an early attempt in this direction. It trains a language model conditioned on a variety of control codes. The network model used in this approach is also the commonly used Transformer, and a piece of control code (domain, style, topics, dates, entities, relationships between entities, etc.) is added in front of the text corpus

% Post-process

* + 1. Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. 2020. Plug and Play Language Models: A Simple Approach to Controlled Text Generation. In International Conference on Learning Representations. <https://openreview.net/forum?id=H1edEyBKDS>
    2. Ben Krause, Akhilesh Deepak Gotmare, Bryan McCann, Nitish Shirish Keskar, Shafiq Joty, Richard Socher, and Nazneen Fatema Rajani. 2021. GeDi: Generative Discriminator Guided Sequence Generation. In Findings of the Association for Computational Linguistics: EMNLP 2021. Association for Computational Linguistics, Punta Cana, Dominican Republic, 4929–4952. <https://doi.org/10.18653/v1/2021.findings-emnlp.424>
    3. Alisa Liu, Maarten Sap, Ximing Lu, Swabha Swayamdipta, Chandra Bhagavatula, Noah A. Smith, and Yejin Choi. 2021. DExperts: Decoding-Time Controlled Text Generation with Experts and Anti-Experts. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, Online, 6691–6706. <https://doi.org/10.18653/v1/2021.acl-long.522>
    4. Kevin Yang and Dan Klein. 2021. FUDGE: Controlled Text Generation With Future Discriminators. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Online, 3511–3535. https://doi.org/10.18653/v1/2021.naaclmain.276

위 방법들은 이런 단점들이 있다.

*[17]* A Survey of Controllable Text Generation using Transformer-based Pre-trained Language Models

*[18]Zhengbao Jiang, Frank F. Xu, Jun Araki, and Graham Neubig. 2020. How Can We Know What Language Models Know? Transactions of the Association for Computational Linguistics 8 (2020), 423–438. https://doi.org/10.1162/tacl\_a\_00324*

*[19] Brian Lester, Rami Al-Rfou, and Noah Constant. 2021. The Power of Scale for Parameter-Efficient Prompt Tuning. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 3045–3059. https://doi.org/10.18653/v1/2021.emnlpmain.243*

*[20]* *Kexin Yang, Dayiheng Liu, Wenqiang Lei, Baosong Yang, Mingfeng Xue, Boxing Chen, and Jun Xie. 2022. Tailor: A Prompt-Based Approach to Attribute-Based Controlled Text Generation. https://doi.org/10.48550/ARXIV.2204.13362*

*[21]* Xiang Lisa Li and Percy Liang. 2021. Prefix-Tuning: Optimizing Continuous Prompts for Generation. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, Online, 4582–4597. <https://doi.org/10.18653/v1/2021.acl-long.353>

*[22]*

*[23]*

*[24]*