

Towards conversational diagnostic artificial intelligence

Sangah Park

Chosun University, College of Medicine

Medical Imaging and Intelligent Reality Lab, Asan Medical Center

2025.06.18

Article

Towards conversational diagnostic artificial intelligence

<https://doi.org/10.1038/s41586-025-08866-7>

Received: 18 January 2024

Accepted: 5 March 2025

Published online: 09 April 2025

Open access

Tao Tu^{1,3}✉, Mike Schaeckermann^{1,3}✉, Anil Palepu^{1,3}, Khaled Saab¹, Jan Freyberg¹, Ryutaro Tanno², Amy Wang¹, Brenna Li¹, Mohamed Amin¹, Yong Cheng², Elahe Vedadi¹, Nenad Tomasev², Shekoofeh Azizi², Karan Singhal¹, Le Hou¹, Albert Webson², Kavita Kulkarni¹, S. Sara Mahdavi², Christopher Semturs¹, Juraj Gottweis¹, Joelle Barral², Katherine Chou¹, Greg S. Corrado¹, Yossi Matias¹, Alan Karthikesalingam^{1,4}✉ & Vivek Natarajan^{1,4}✉

Introduction

Method

Results

Discussion

Introduction

Doctor–Patient Conversation

Structure and flow of the conversation

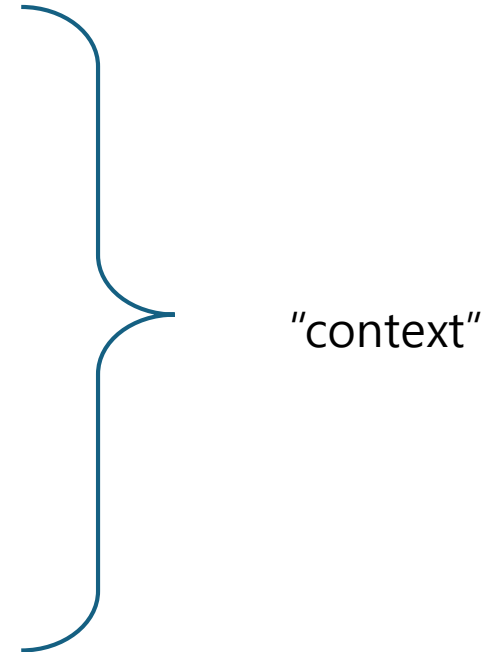
Completeness of history taking

Diagnostic accuracy

Rationale of the management plan

Consideration for the patient's perspective and concerns

Relationship building and rapport

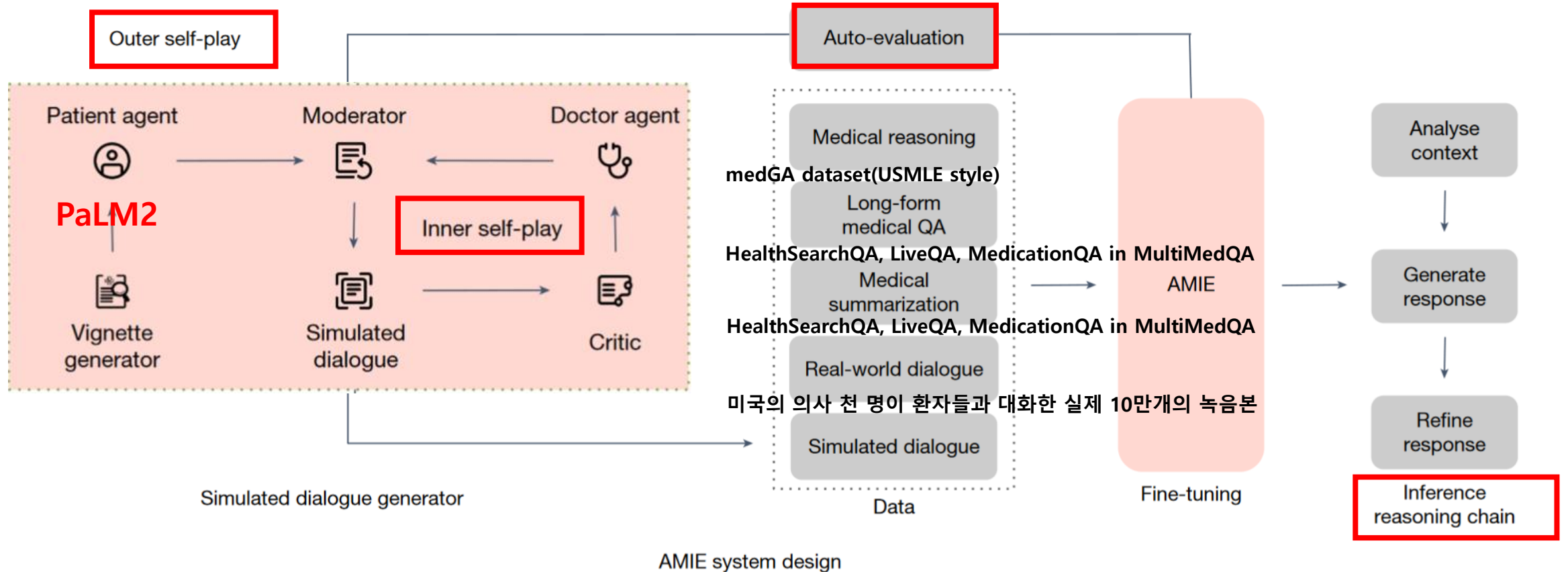


AMIE

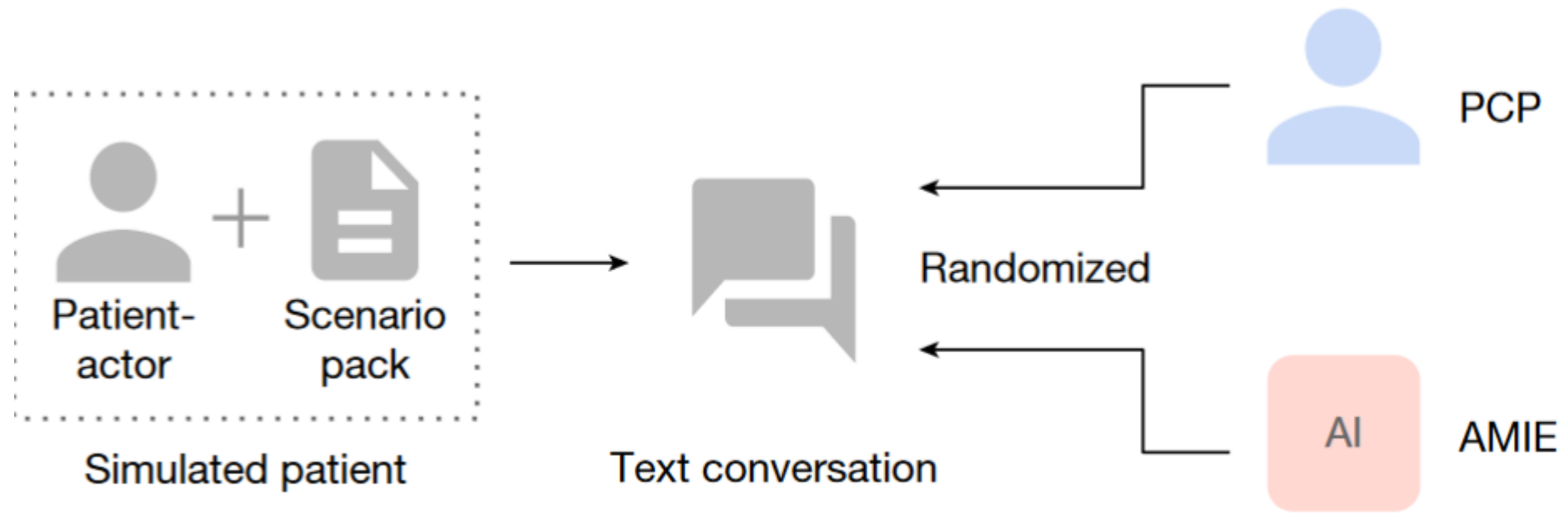
- uses a **self-play-based** simulated environment with **automated feedback** for scaling learning across disease conditions, specialties and contexts.
- uses an inference time **Chain of Reasoning strategy**
- evaluated by pilot **evaluation rubric** -> clinical centered, patient-centered
- evaluated by **OSCE**(159) : AMIE vs PCP // patient actors

Method

AMIE system design



Remote OSCE study design



Remote OSCE study design

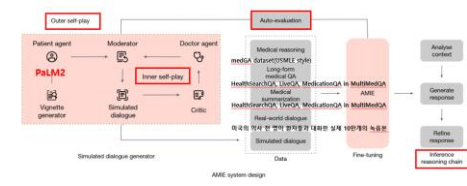
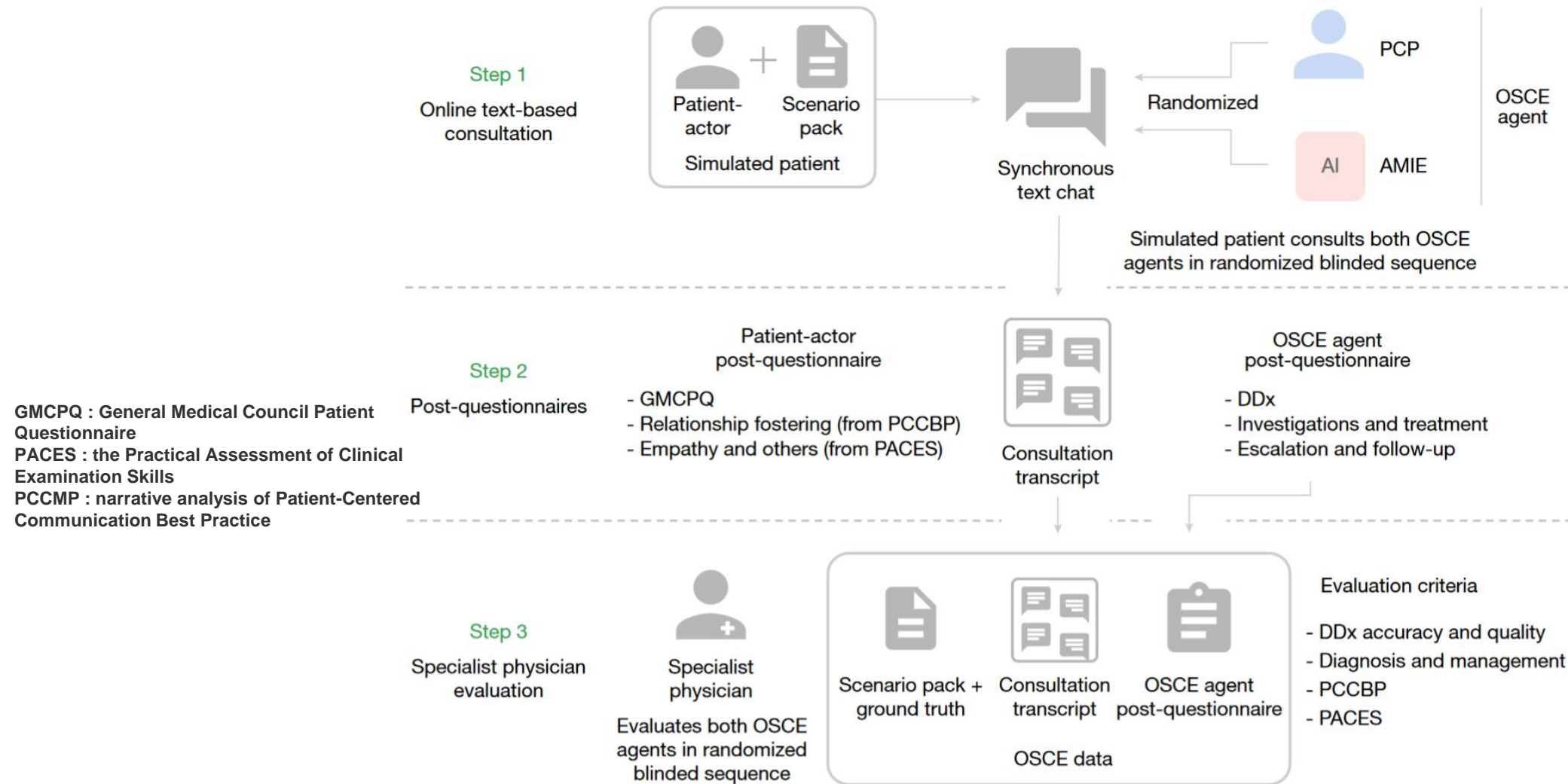
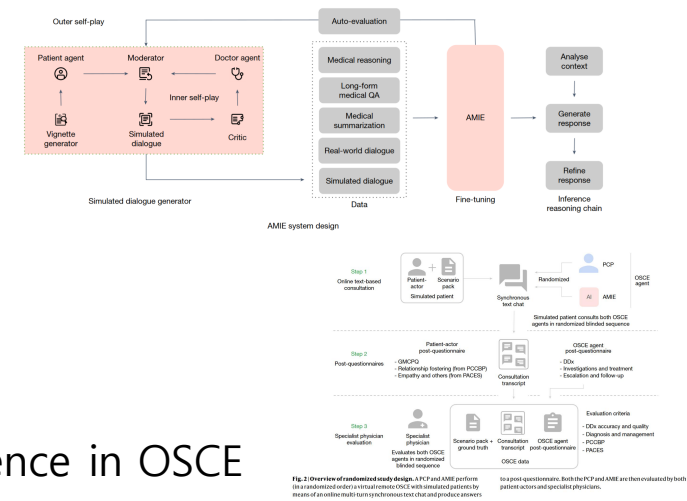


Fig. 2 | Overview of randomized study design. A PCP and AMIE perform (in a randomized order) a virtual remote OSCE with simulated patients by means of an online multi-turn synchronous text chat and produce answers

to a post-questionnaire. Both the PCP and AMIE are then evaluated by both the patient-actors and specialist physicians.

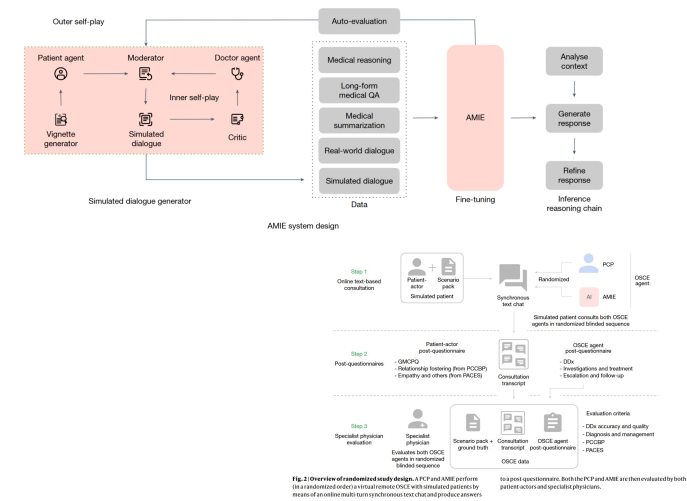
Remote OSCE study design

- 159 scenario packs from India (75), Canada (70) and the United Kingdom (14).
- PCP : 3~25 post residency experience doctors
- Patient-actors : a mix of medical students, residents and nurse practitioners with experience in OSCE
- Scenario packs prepared by OSCE labs in Canada and India, both affiliated with medical schools.
- UK scenarios sourced from publicly available MRCP(UK) samples.
- Each pack included a ground-truth diagnosis and a list of acceptable differential diagnoses.
- Covered domains: Cardiovascular (31), Respiratory (32), Gastroenterology (33), Neurology (32), Urology, Obstetrics & Gynaecology (15), Internal Medicine (16).
- Excluded: Paediatrics, Psychiatry, ICU, and inpatient case management.
- Indian patient-actors simulated all India cases and 7 of 14 UK cases.
- Canadian patient-actors participated in Canadian scenarios.



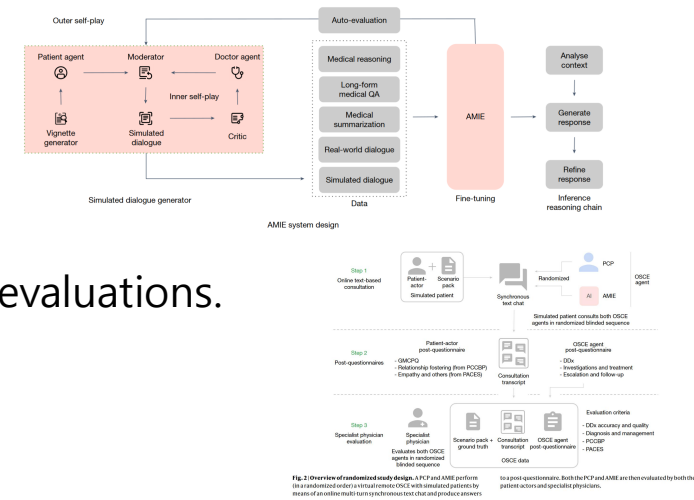
Remote OSCE study design

- Each simulated patient completed two synchronous text-based consultations:
 - One with a Primary Care Physician (PCP)
 - One with AMIE (AI model)
- Consultations were randomized and blinded to the patient-actor.
- PCPs were located in the same country as the patient-actor.
- Conversations were limited to 20 minutes, and agents were instructed not to reveal identity.
- After each consultation, both the OSCE agent and the patient-actor completed post-questionnaires.
- Patient-actors completed: GMCPQ, PACES (2 components), PCCBP checklist
- OSCE agents submitted: DDx (3–10 items), escalation plan, investigations, treatment, management, f/u recommendations



Remote OSCE study design

- 33 specialists from India (18), North America (12), and the UK (3) participated in blinded evaluations.
- Specialists matched by geographic region and specialty of each scenario.
- Each scenario was evaluated by 3 specialists, reviewing both PCP and AMIE consultations.
- Evaluated aspects:
 - DDx accuracy, escalation, investigation, treatment, management, follow-up
 - PACES and PCCBP rating items
 - Identification of confabulations (non-factual or fabricated content)
- Each scenario pack included gold-standard clinical references to guide consistent evaluation, as in standard OSCE protocols.

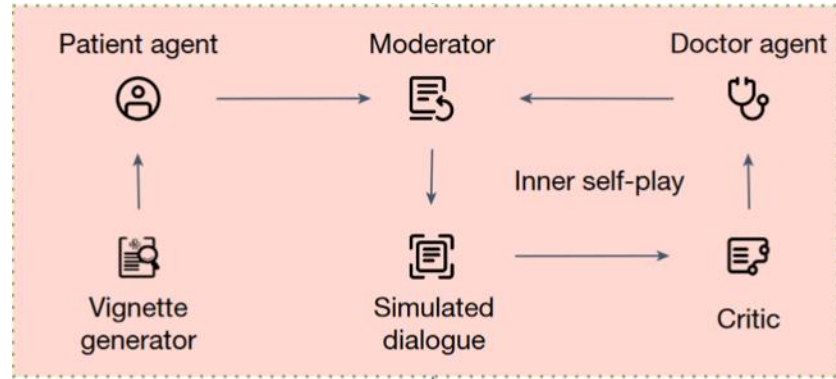


AMIE system design

SI.4 Prompting for self-play environment

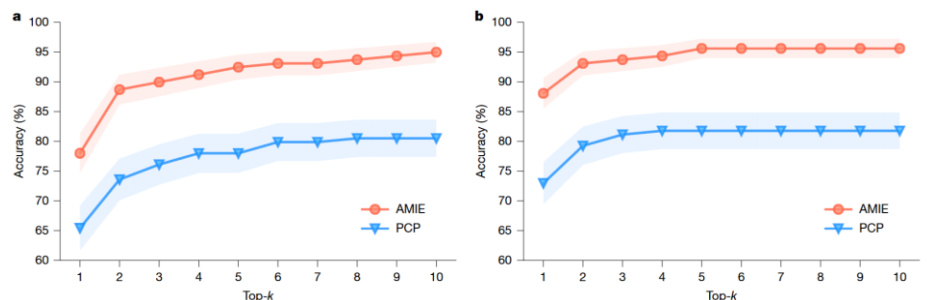
Task	Prompt
Web-Search Retrieval	What are the specific [demographics/symptoms/management plans] for the condition {condition}?
Passage Filtering	For the clinical condition, {condition}, is the following a good description of common [demographics/symptoms/management plans](Yes/No)? Description: {retrieved passage} Answer (Yes/No):
Vignette Generation	The following are several passages about the demographics, symptoms, and management plan for a given condition. Generate 2 different patient vignettes consistent with these passages. Follow the format of the given example (just list N/A if a particular field is unavailable).Condition: {condition}, Demographics: {demographic passages}, Symptoms: {symptom passages}, Management Plan: {management passages} Example Format: {one-shot example} Patient Vignettes for {condition}:
Patient Agent Instruction	You are a patient chatting with a doctor over an online chat interface. The doctor has never met you before. {vignette}. Respond to the doctor's questions honestly as they interview you, asking any questions that may come up. {dialogue}
Doctor Agent Instruction	You are an empathetic clinician asking a patient about their medical history over an online chat interface. You know nothing about the patient in advance. Respond to the patient with a single-turn response to better understand their history and symptoms. Do not ask more than two questions. If the patient asks a question, be sure to answer it appropriately. {dialogue}
Moderator Instruction	The conversation should only come to an end if the doctor has finished giving the patient a diagnosis and treatment plan and the patient has no questions left. A conversation also comes to an end if the doctor or patient says goodbye. Question: has the conversation come to an end? Yes or No. {dialogue}
Critic Instruction	You are criticizing a dialogue from an AI doctor agent asking a patient about their medical history over an online chat interface (because it is virtual, the clinician cannot do physical exams like in a clinic). The patient is suffering from a particular medical problem, and the doctor hopes to understand their chief complaint, history of present illness, demographics, and social history in order to best determine what is going on, their likely diagnoses and next steps for treatment. The doctor must dig deep (potentially asking followup questions) into the particular symptoms the patient is complaining about and anything clinically significant. Make suggestions for the doctor to better meet the following criteria: <ol style="list-style-type: none"> 1. The doctor agent exhibits empathy and professionalism while addressing the patient agent's latest questions or comments in a concise manner. 2. The doctor agent avoids asking too many or repetitive questions (about information already acquired), focusing on a maximum of one or two per response. 3. The responses should not reveal that the doctor agent is an AI chatbot. They should flow naturally, maintain factual accuracy, and facilitate further engagement from the patient 4. The doctor agent asks sufficient questions to identify at least two of the most likely differential diagnoses. They further refine their understanding through targeted questions towards the ground truth diagnosis and offer the corresponding treatment. <p>The patient has the following condition: {condition}, with the following background information: {vignette}. Dialogue: {dialogue}</p>

Table SI.3 | Prompts for self-play based simulated environment.



Results

Differential Diagnosis Accuracy



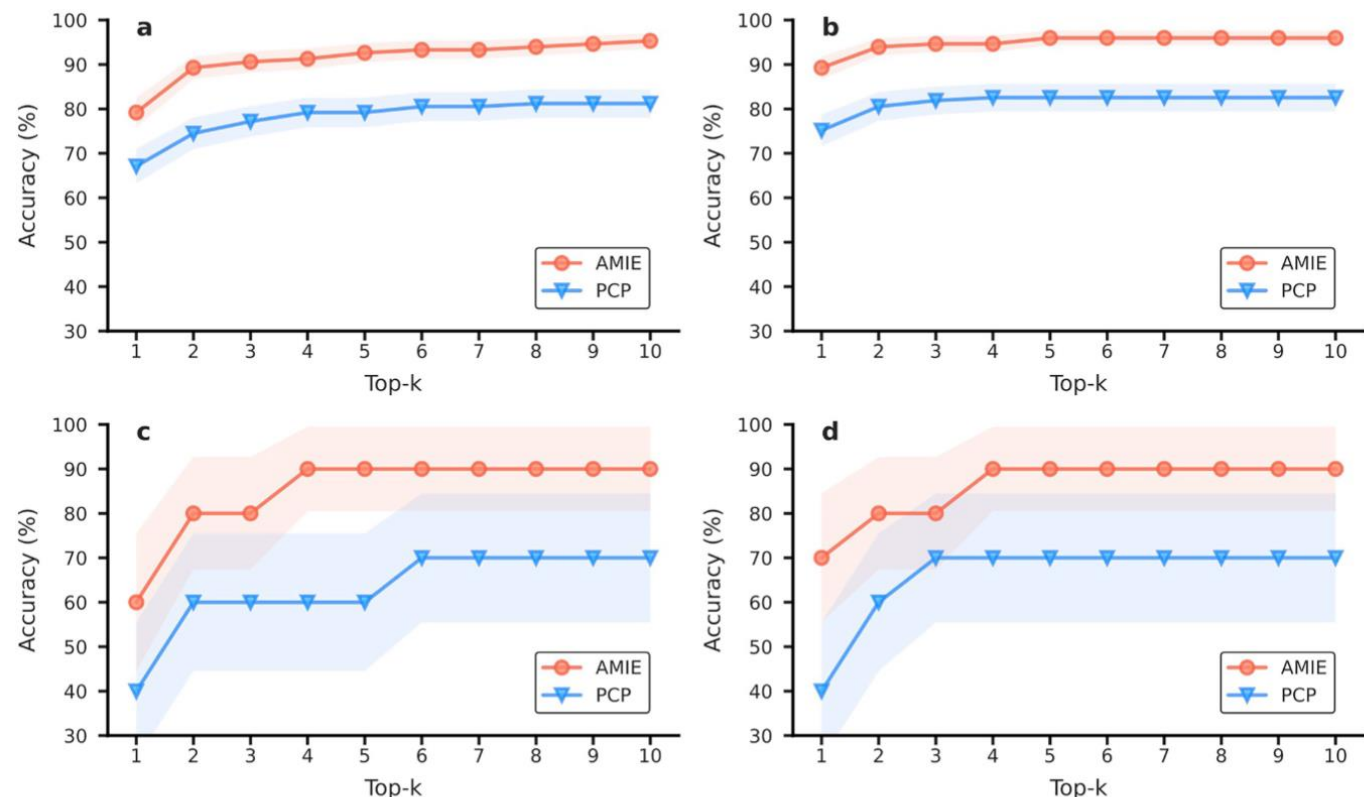
Matched with ground truth diagnosis

Matched with any item on the accepted differential

Fig. 3 | Specialist-rated top-k diagnostic accuracy. a,b, T

Table SI.13 | Ablation set of non-disease state scenarios (N=10).

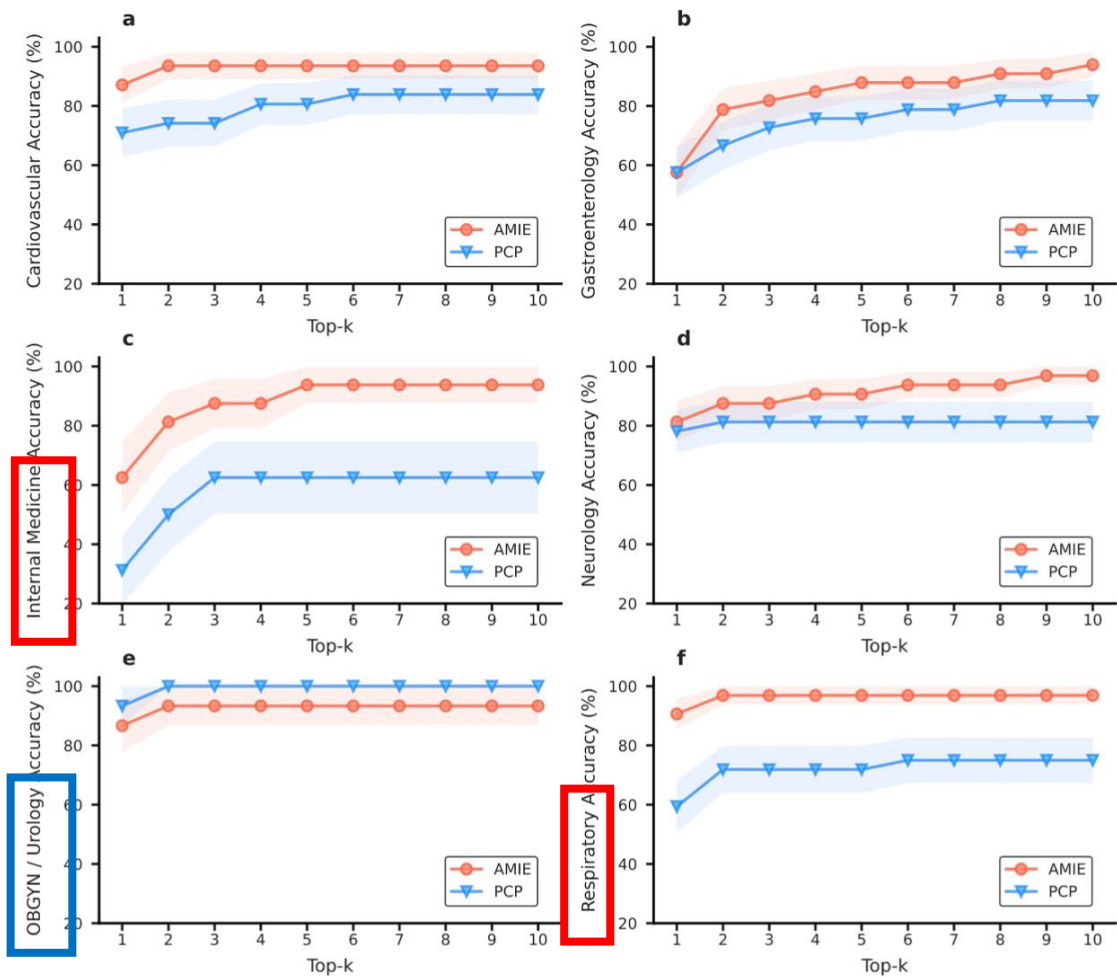
Specialty	Ground truth
Cardiovascular	GERD - associated chest pain (recurrence of known prior disease state) Heat illness/ heat related palpitations
Gastroenterology	Binge drinking associated nausea/ vomiting Resolved constipation
Internal Medicine	Hypoglycemia from missing insulin intake Resolved cellulitis
Neurology	Resolved episode of orthostatic hypotension Resolved headaches due to sleep deprivation
Respiratory	Resolved post-nasal discharge (viral) Resolved strep throat



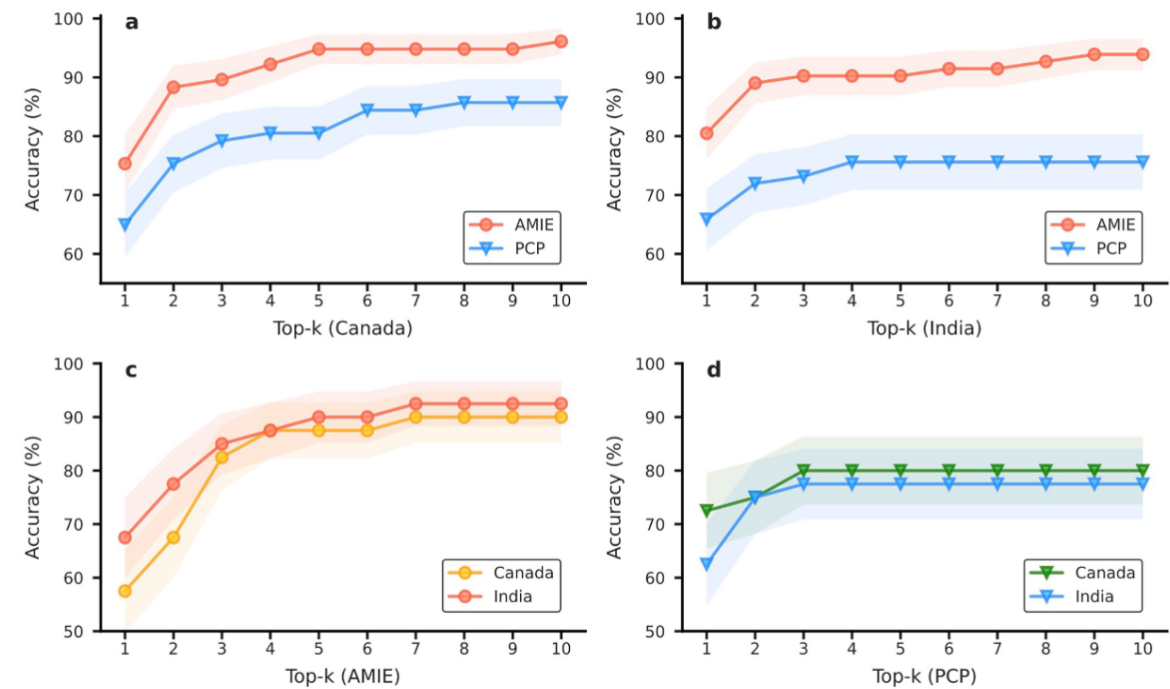
Extended Data Fig. 2 | DDX top-k accuracy for non-disease-states and positive disease-states. a,b: Specialist rated DDX top-k accuracy for the 149 “positive” scenarios with respect to (a) the ground-truth diagnosis and (b) the accepted differentials. **c,d:** Specialist rated DDX top-k accuracy for the 10 “negative” scenarios with respect to (c) the ground-truth diagnosis and (d) the accepted differentials. Using two-sided bootstrap tests ($n = 10,000$) with FDR correction, differences in the “positive” scenarios were significant ($P < 0.05$) for all k , but differences in “negative” scenarios were not significant due to the small sample size. Centrelines correspond to the average top-k accuracy, with 95% confidence intervals shaded. The FDR-adjusted P values for positive disease states, ground-truth comparison: 0.0041 ($k = 1$), 0.0002 ($k = 2$),

0.0001 ($k = 3$), 0.0002 ($k = 4$), 0.0001 ($k = 5$), 0.0002 ($k = 6$), 0.0002 ($k = 7$), 0.0003 ($k = 8$), 0.0001 ($k = 9$) and 0.0001 ($k = 10$) (a). The FDR-adjusted P values for positive disease states, accepted differential comparison: 0.0002 ($k = 1$), 0.0001 ($k = 2$), 0.0002 ($k = 3$), 0.0003 ($k = 4$), 0.0001 ($k = 5$), 0.0001 ($k = 6$), 0.0001 ($k = 7$), 0.0001 ($k = 8$), 0.0001 ($k = 9$) and 0.0001 ($k = 10$) (b). The FDR-adjusted P values for non-disease states, ground-truth comparison: 0.1907 ($k = 1$), 0.1035 ($k = 2$), 0.1035 ($k = 3$), 0.1035 ($k = 4$), 0.1035 ($k = 5$), 0.1035 ($k = 6$), 0.1035 ($k = 7$), 0.1035 ($k = 8$), 0.1035 ($k = 9$) and 0.1035 ($k = 10$) (c). The FDR-adjusted P values for non-disease states, accepted differential comparison: 0.1035 ($k = 1$), 0.1035 ($k = 2$), 0.1829 ($k = 3$), 0.1035 ($k = 4$), 0.1035 ($k = 5$), 0.1035 ($k = 6$), 0.1035 ($k = 7$), 0.1035 ($k = 8$), 0.1035 ($k = 9$) and 0.1035 ($k = 10$) (d).

Differential Diagnosis Accuracy



Extended Data Fig. 3 | Specialist rated DDx accuracy by scenario specialty.



Extended Data Fig. 4 | DDx accuracy by location.

Differential Diagnosis Accuracy

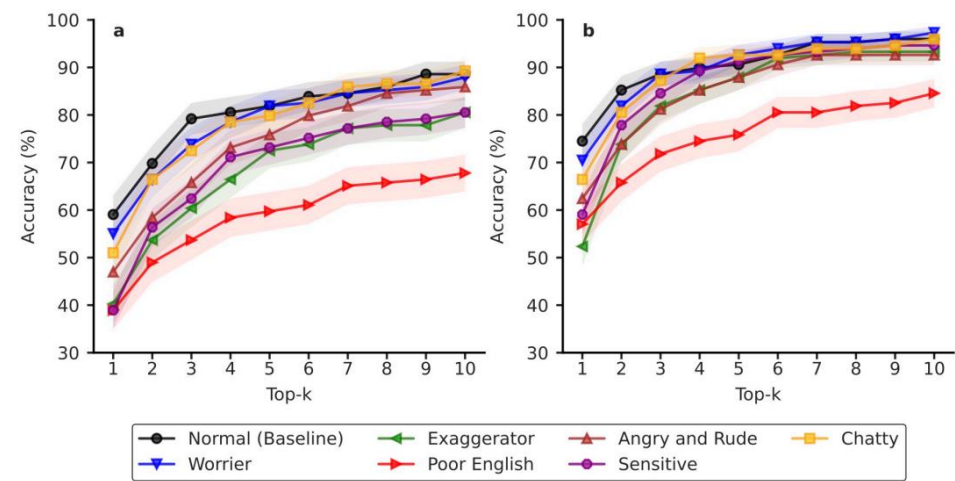
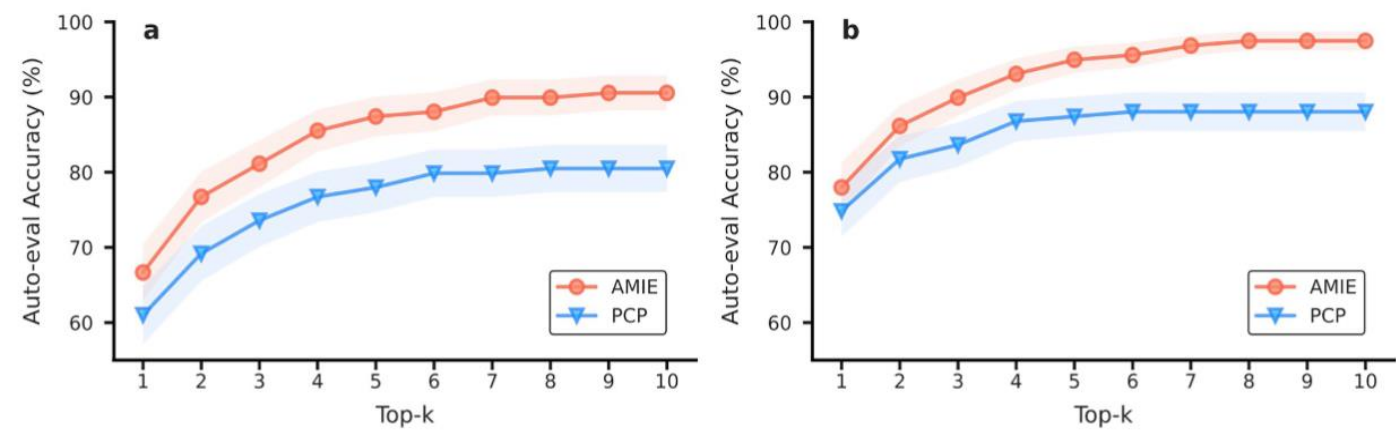
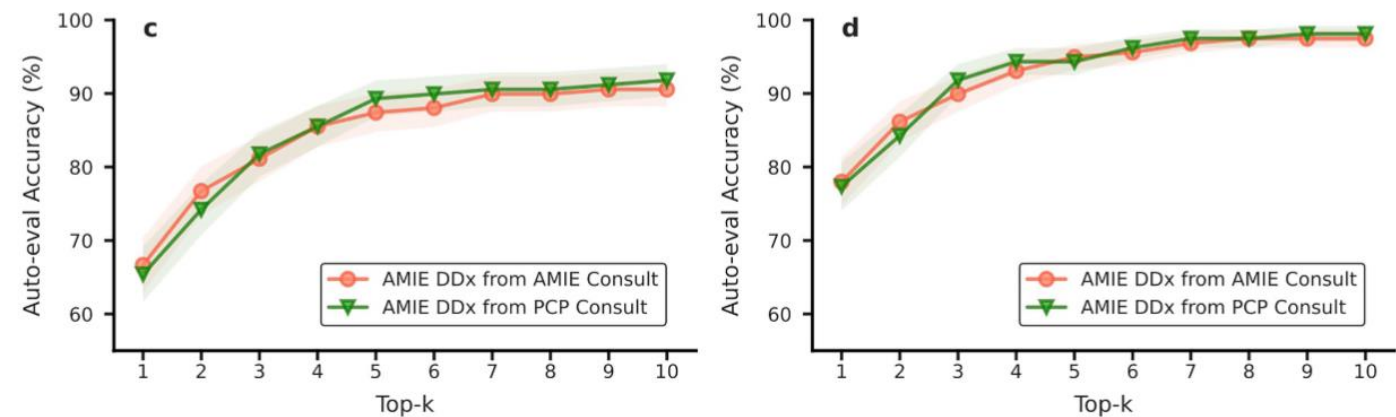


Figure SI.3 | DDx performance with different patient behaviors.

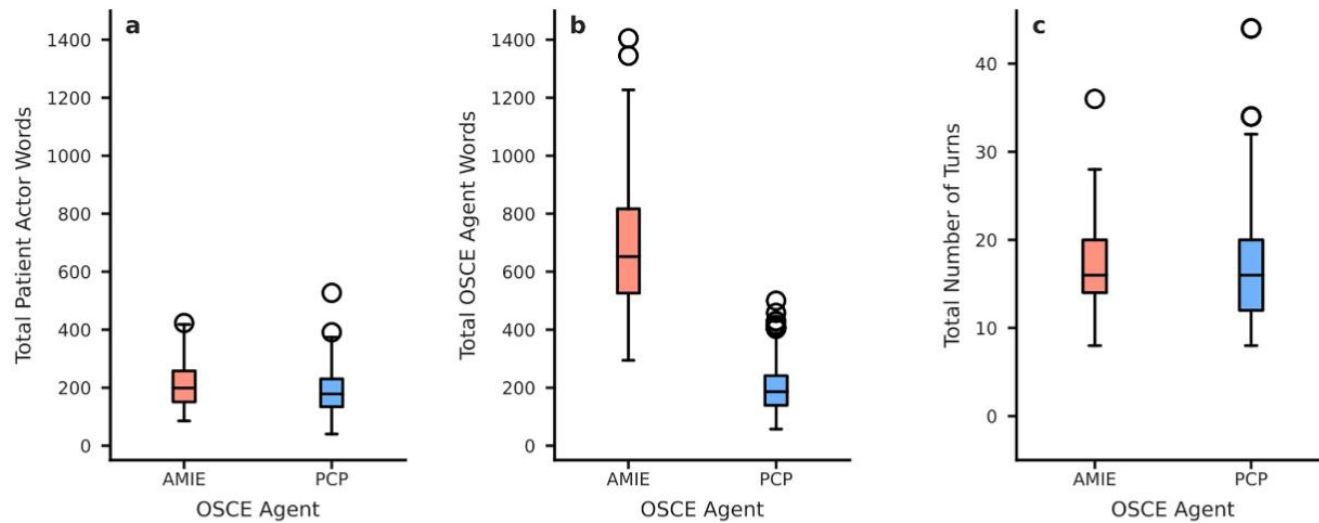
Extended Data Fig. 5 | Auto-evaluation of DDx performance. a, b: Top-k DDx



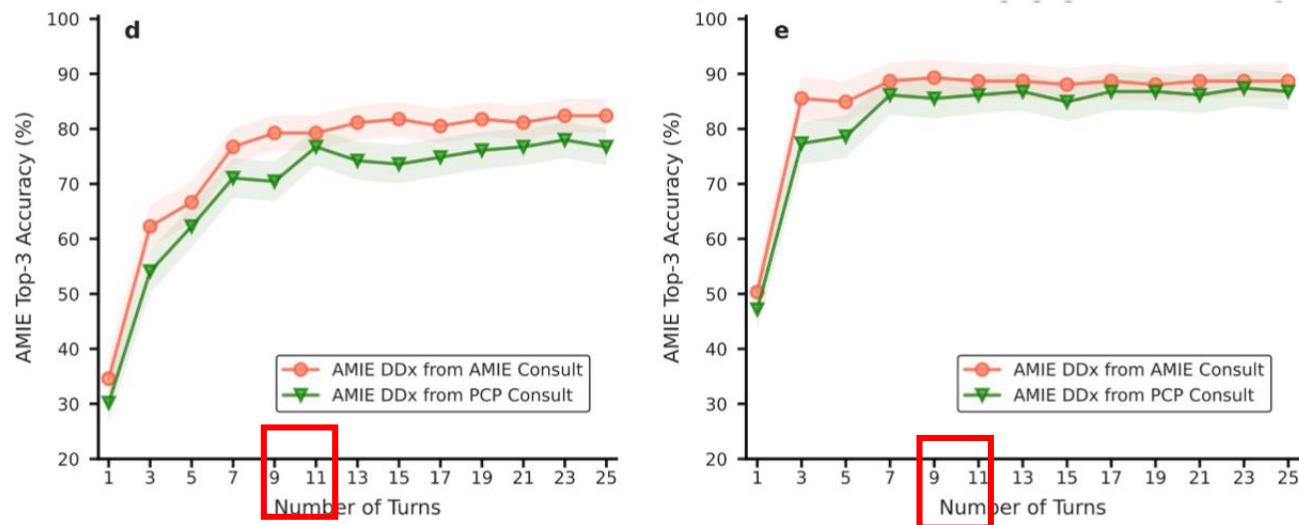
Extended Data Fig. 5 | Auto-evaluation of DDx performance. a, b: Top-k DDx



Differential Diagnosis Accuracy



Extended Data Fig. 6 | Consultation verbosity and efficiency of information



Conversation Quality

Patient: Patricia Wilson (45 yr) - Chest Pain, **Probable diagnosis:** Angina or non-ST-elevation myocardial infarction
Plausible alternative diagnoses: Biliary disease, Dyspepsia

PCP Consultation	
Patient Actor	> Hi Doctor. how are you?
Primary Care Physician	> I'm great. How can I help you?
Patient Actor	> I recently experienced an episode of chest pain and discomfort. About two hours ago, while walking home after lunch with my daughter, I developed central chest pain along with some discomfort in the upper stomach.
Primary Care Physician	> I'm sorry to hear that. Are you still in pain?
Patient Actor	No, I'm not currently in pain. The chest pain and discomfort I experienced about two hours ago lasted for around 20 minutes, and since then, I haven't felt any pain.
Primary Care Physician	> I'm glad you are not uncomfortable anymore. Chest pain can sometimes be quite serious, though. Do you mind if I ask you so more questions? For example, have you ever had similar pain before?
Patient Actor	> No, this is the first time when i'm having this kind of pain
Primary Care Physician	> Is it associated with any other symptoms? Like nausea/vomiting? Or shortness of breath?
Patient Actor	> Yes, during the episode of chest pain, I did experience nausea. Additionally, there was a heavy feeling in my left arm. I didn't feel sweaty or breathless with the pain
Primary Care Physician	> How old are you, may I ask? And what kind of medical conditions do you have?
Patient Actor	> I'm 45 years old. I'm not on any prescribed medications, but I do take a vitamin preparation that I buy from the chemist.
Primary Care Physician	> Do you know what type of Vitamin? Also, please confirm that you do not have any hyigh blood pressure, diabetes, high cholesterol, smoking history, or family history of heart attacks.
Patient Actor	> I take a vitamin preparation that I purchase from the chemist, but I don't have the specific details about its composition. As for my health history, I don't have high blood pressure, diabetes, or high cholesterol. However, I do have a smoking history of about 10 cigarettes per day since the age of 25. In terms of family history, my father had high blood pressure and had a heart attack in his mid-fifties. I have two older brothers, and I recently found out that one of them has high cholesterol.
Primary Care Physician	> Can you describe the quality of the pain? Was there anything that made it better or worse? For example, sometimes pain is aggravated by exertion - walking at a fast pace/relieved by rest. Sometimes chest pain is made worse by taking a deep inspiration. do either of these apply? Anything else you can think of?
Patient Actor	> I wouldn't say the pain was aggravated by taking a deep breath, but walking seemed to worsen it. Resting or sitting down provided some relief. It wasn't a sharp or stabbing pain, more like a continuous pressure.
Primary Care Physician	> Did you have any palpitations? Did you feel as though you might faint? Any other symptoms?
Patient Actor	> No, I didn't experience palpitations during the episode of chest pain. I also didn't feel as though I might faint. The main symptoms were the central chest pain, discomfort in the upper stomach, a heavy feeling in my left arm, and nausea. While the pain was intense during the first few minutes, it became a constant and unpleasant sensation, and it completely eased off after about 20 minutes. Since then, I haven't had any lingering symptoms, and I currently feel well
Primary Care Physician	> I've heard enough to know that your problem needs to be escalated. I would like for you to go to your nearest emergency room should your pain recur.
PCP Top-10 DDx	
1. Acute Coronary Syndrome, 2. GERD, 3. Biliary Colic, 4. Pulmonary Embolism, 5. Pancreatitis, 6. N/A ...	

AMIE Consultation	
Patient Actor	> Hi doctor, I experienced central chest pain accompanied by discomfort in the upper stomach that happened two hours ago
AMIE	> I'm sorry to hear you're experiencing this discomfort, especially in two key areas. To help me get a clearer picture, can you describe the chest pain in more detail? Is it a pressure, tightness, burning, stabbing sensation, or something else entirely?
Patient Actor	> The chest pain I experienced was more of a pressure and tightness, especially in the center of my chest. It felt like there was a heavy weight on my chest, and the discomfort extended to my upper stomach. I wouldn't describe it as a burning or stabbing sensation, more like a continuous pressure that was quite noticecable. The heaviness also seemed to radiate down my left arm. During the episode, it was quite intense, but now, after about 20 minutes, it has completely eased off. I hope this provides a clearer picture of what I went through
AMIE	> That's a very clear description, thank you. It helps me narrow down possibilities, but to give you the best advice, I need some more background. Firstly, what is your age and do you have any significant medical history? This includes things like heart problems, high blood pressure, or past surgeries.
Patient Actor	> I'm 45 years old, and in terms of medical history, about five years ago, I was diagnosed with gallstones. However, I opted not to have surgery at that time, and it hasn't caused any recent issues. I had varicose vein stripping done 15 years ago, and when I was 23, I had pneumonia. As for other conditions, I don't have a history of heart problems or high blood pressure. Currently, I'm not on any prescribed medications
AMIE	> 45 is relatively young for heart problems, which is reassuring. However, your symptoms, especially the pressure-like pain and radiation to your arm, ARE concerning for a few possibilities. To help me narrow it down further: Family history: Is there any heart disease, like heart attacks or angina, in your close relatives? Lifestyle: Do you currently smoke or drink alcohol?
Patient Actor	> In my family history, my father had high blood pressure, and I believe he had a heart attack in his mid-fifties. my mother tends to get chest infections occasionally. Regarding lifestyle, I do smoke about 10 cigarettres per day, and I have been doing so since the age of 25. However, I drink alcohol infrequently, mainly on special occasions.
AMIE	> Thank you for providing that detailed information, it's very helpful. Given your family history of heart problems, especially your father's heart attack, and your smoking habit, we can't ignore the possibility that this is heart-related. However, the fact that your pain has resolved on its own makes a typical heart attack less likely. It's possible this is angina, which is chest pain caused by reduced blood flow to the heart. It can feel similar to a heart attack, but is usually less severe and goes away with rest. Another possibility is GERD, which is heartburn caused by stomach acid refluxing into the esophagus. This can also cause chest pain, but it's usually accompanied by a burning sensation. To be sure what's causing your symptoms, I strongly recommend you seek immediate in-person evaluation . This will likely involve an electrocardiogram (ECG) to check your heart's electrical activity, and blood tests to look for signs of heart damage or other issues. Depending on the initial findings, further tests may be needed. I understand this is worrying, but it's crucial to act fast even if it turns out NOT to be a heart attack. These symptoms are significant and need prompt attention to rule out serious causes and get you the right treatment.
Patient Actor	> I appreciate your thorough explanation, and I understand the urgency of the situation. Given the potential seriousness of the symptoms, I will seek immediate in-person evaluation as you've advised. The possibility of angina or GERD is certainly concerning, and I want to ensure that I receive the necessary tests and examinations to determine the cause and appropriate treatment. Thank you for your guidance, and I'll proceed to the clinic promptly to address this issue
AMIE	> I'm glad to hear you're taking this seriously, as prompt action is key here. At the clinic, be sure to relay ALL the details we discussed, including family history, for a full risk assessment. This will help them determine the best next steps, which may be immediate treatment or referral to a cardiologist for specialized care.
Patient Actor	> Thank you again for your assistance and advice
AMIE	> You're very welcome. My goal is to help you get the best care possible, so don't hesitate to reach out if further questions arise after your in-person evaluation. Take care, and I hope you get answers and relief soon.
AMIE Top-10 DDx	
1. Unstable Angina, 2. Acute Myocardial Infarction, 3. Stable Angina, 4. Gastroesophageal Reflux Disease (GERD), 5. Atypical Chest Pain, 6. Costochondritis, 7. Peptic Ulcer Disease, 8. Gallbladder Disease (biliary colic), 9. Panic Attack/Anxiety, 10. Pulmonary Embolism	

Conversation Quality

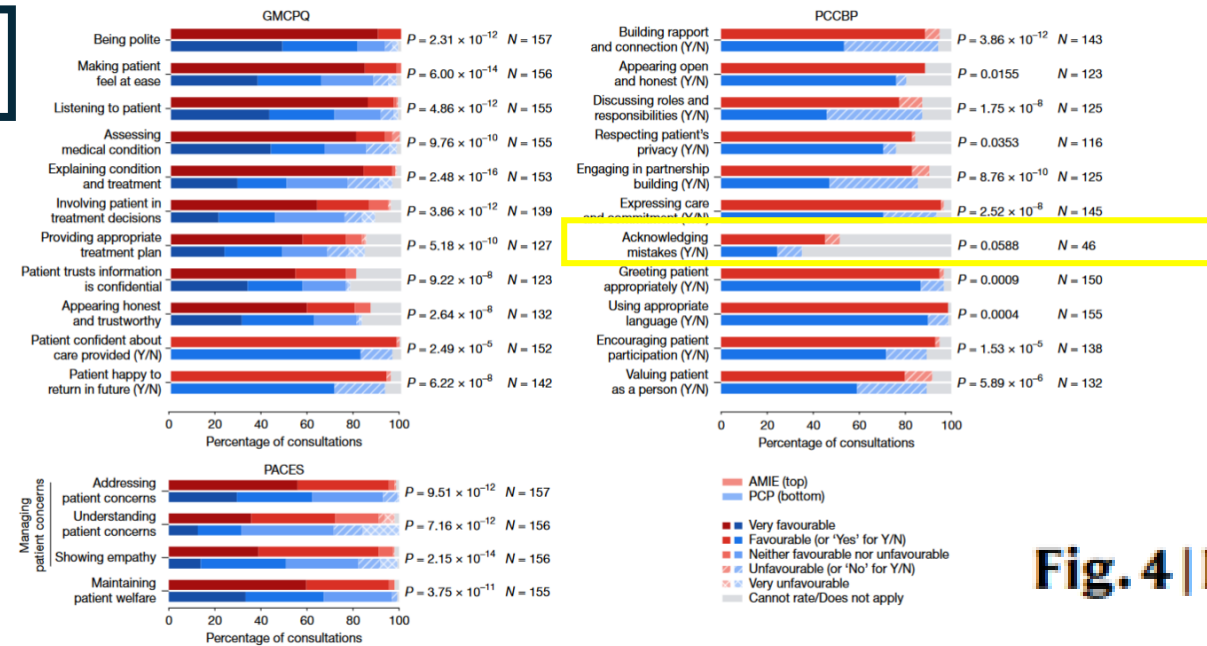


Fig. 4 | Patient-actor ratings.

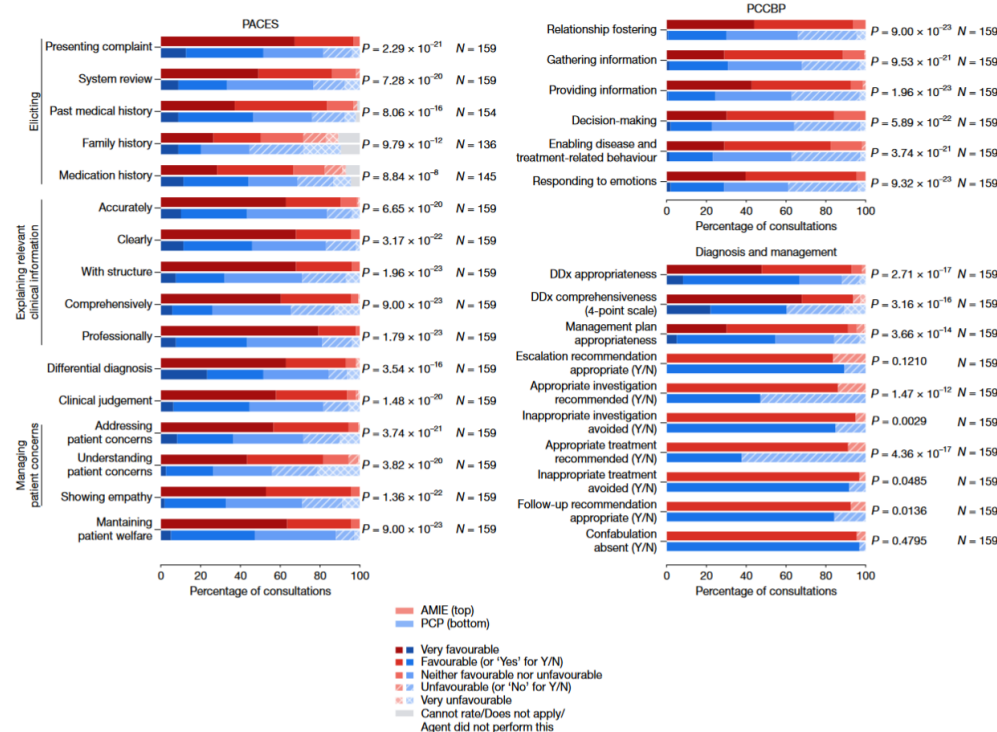
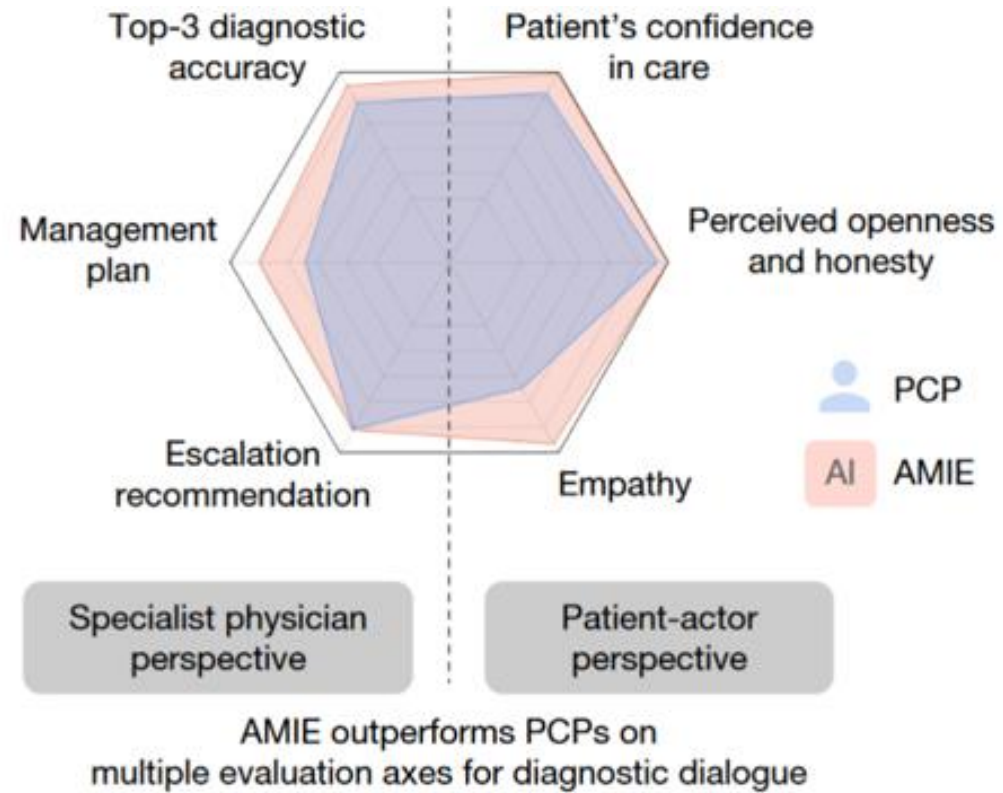


Fig. 5 | Specialist physician ratings.

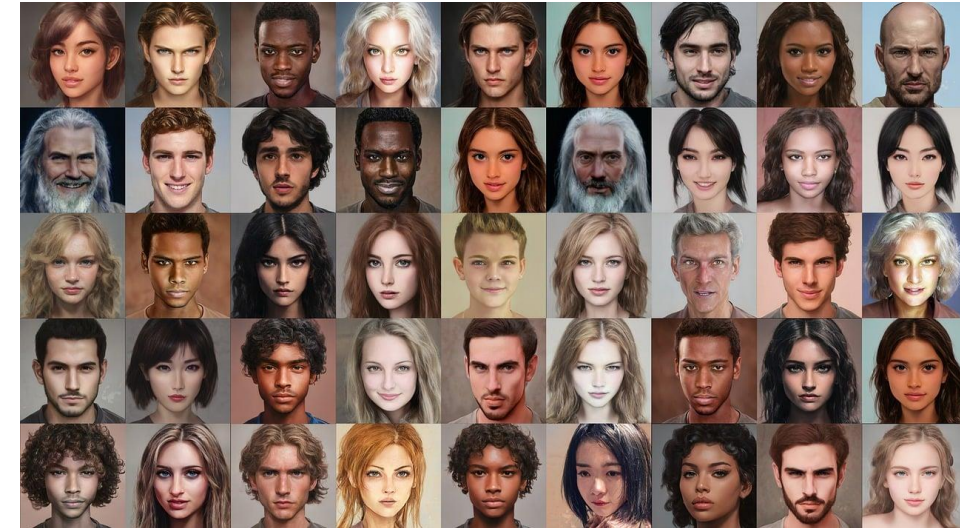
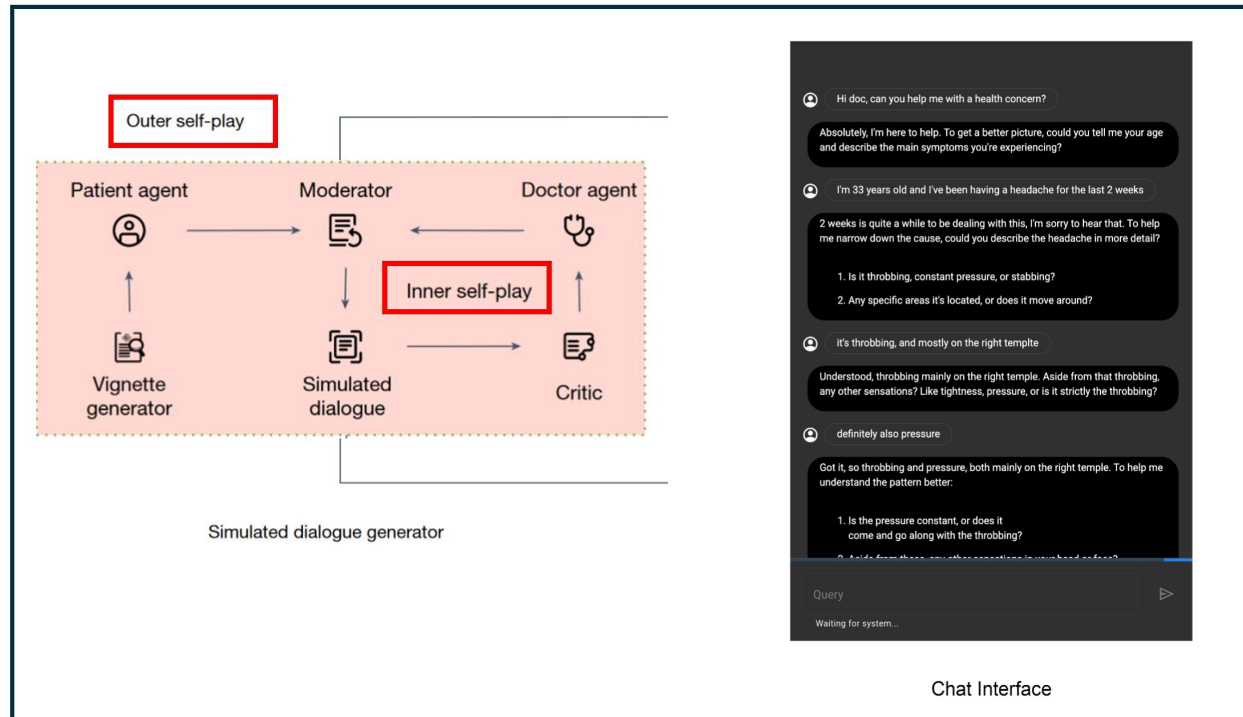
Conclusion



Discussion

Limitation

Patient Actor \neq Real patient with real, various background, personality, motivation...



Thank you

Sangah Park
Chosun University, College of Medicine

<https://www.fnnews.com/news/202110241419385432>

<https://thenextweb.com/news/study-patients-are-less-likely-to-follow-advice-from-ai-doctors-that-know-their-names>

<https://blog.amopportunities.org/2020/02/03/meaning-in-medicine-patient-communication/>

<https://pixabay.com/ko/illustrations/%EC%83%81%EC%9D%B4-%EB%B9%84%EC%9C%A8-%EC%84%B8%EA%B3%84-%EC%9D%B8%EA%B5%AC-%EC%9D%B8%EA%B5%AC-7409290/>