Translation for AJR article "***Human-in-the-Loop Large Language Model–Augmented Diagnostic Reasoning in Thoracic Imaging: Impact of Radiologic Expertise***"

by lead author **Jiyoung Song** from Department of Radiology, Seoul National University Hospital

Translation by **Xin Bai** from Department of Radiology, Peking Union Medical College Hospital

标题：人在回路大语言模型辅助胸部影像诊断推理：放射专业水平的影响

第一作者：Jiyoung Song 首尔国立大学医院放射科

翻译及朗读：白鑫 中国医学科学院北京协和医院 放射科

背景：技术与监管层面的限制，制约了大语言模型（LLM）在放射学中辅助诊断推理的应用。基于阅片者文本描述的工作流程，可能成为一种实用的替代方案。

目的：评估基于阅片者生成的影像自由文本描述的大语言模型辅助诊断对诊断效能的影响，并分析阅片者专业水平对该大语言模型辅助诊断流程的作用。

方法：本回顾性研究纳入 2014 年 1 月-2017 年 12 月韩国胸部放射学会题库平台的 93 例病例（涵盖 X 线、CT、MRI 及 PET/CT 影像）。每例病例设置 5 个鉴别诊断（1 个正确诊断+4 个干扰项）。10 名阅片者（5 名胸部放射科医师、5 名放射科住院医师）独立判读病例。第 1 阶段，阅片者选择最可能诊断并撰写关键发现的自由文本描述；将阅片者的自由文本描述（不提供病例影像）输入大语言模型（Gemini 3.0 Pro），模型对 5 个鉴别诊断进行排序，并为前 3 位选项提

供解释依据。第 2 阶段，向阅片者提供基于其自身自由文本描述的模型输出结果，阅片者重新选择最可能诊断。同时评估仅输入影像（无自由文本描述）时模型的诊断效能，以模型排序首位诊断判定准确率。采用广义估计方程比较不同阅片者组间差异。

结果：仅输入影像时模型准确率为 52.7%，输入阅片者生成的描述时为 63.9%；基于胸部放射科医师描述的模型准确率高于住院医师（67.3% vs 60.4%；P<0.001）。从第 1 阶段到第 2 阶段，胸部放射科医师准确率由 56.3%提升至 65.6%，住院医师由 42.4%提升至 58.5%；住院医师的准确率提升幅度显著高于胸部放射科医师（16.1 个百分点 vs 9.2 个百分点；P=0.02）。住院医师接受模型优先推荐诊断的比例显著高于胸部放射科医师（73.2% vs 48.9%；P<0.001），且在模型输出误导信息时，住院医师转向错误诊断的比例更高（60.6% vs 32.4%；P=0.009）。

结论：基于文本的大语言模型辅助工作流程可提升阅片者诊断准确率，但该效果显著受阅片者专业水平影响。

临床影响：人在回路工作流程的价值源于阅片者与大语言模型的动态交互，而这种交互由构建模型输入、批判性评估模型输出的操作者专业水平决定。