医知創造ラボ｜医療者向け

研修医が「AIで調べました」と言ってきたら

指導医の 5つの問いかけ 【DEFT-AI】

医知創造ラボ | 指導医・研修医・医学生の方へ | 監修：今村久司

今日の問い TODAY'S POINT

AIの鑑別を持ってきた研修医に、何を返すか？

結論 禁止しない。使った瞬間を教育機会に変える5つの問いかけがある。

KEY NUMBER

2%ポイント

⚠ 95%CI −4〜8・P=.60（有意差なし）

診断推論スコア
76% 対 74% の差

スコアは鑑別の妥当性・支持/反証所見・次の検査を採点したルーブリック（%表示）。
Goh E, et al. JAMA Netw Open. 2024（医師50名・単盲検RCT）

同じ研究で、LLM単独は従来資源群より16%ポイント高い

LLM単独群のスコアは、従来資源のみ群より**16%ポイント高かった**（95%CI 2〜30、P=.03）。

📋 **ここが要点**

道具の性能は高い。**使いこなせていないのは人**。渡し方＝指導のしかたの問題。

KEY NUMBER

30%

⚠ 英国レプリケーション（医師22名）

症例設問のうち、実際にLLMに投げられた割合

Healy J, et al. Diagnosis (Berl). 2026

芯2 ASYMMETRIC RISK

06

得より損が大きい

提示したAI	診断精度の変化	評価
標準AI（説明なし）	+2.9%ポイント	有意に改善
標準AI（説明つき）	+4.4%ポイント	有意に改善
偏ったAI（説明なし）	−11.3%ポイント	有意に悪化
偏ったAI（説明つき）	−9.1%ポイント	悪化（改善は非有意）

医師457名（hospitalist・NP・PA）・ベースライン精度73.0%のランダム化臨床ビネット調査。Jabbour S, et al. JAMA. 2023.

説明をつけても、損は打ち消せない

系統的に偏ったAIの悪影響は、説明を添えても改善しなかった（Jabbour 2023）。

📋 **Yu 2024（放射線科医対象）の指摘**

ベースライン成績が低い臨床家ほどAI併用で不利益を受ける。AIが医師を上回る場面では「人＋AI」が「AI単独」より悪化する。

KEY NUMBER

56%

⚠ 最終学年医学生148名（GPT-3.5・10症例中5問が誤答）

LLM出力を正しく評価できた割合（中央値）

Waldock WJ, et al. BMC Med Educ. 2025

同研究で、5%しか知らなかったこと

「clinical prompt engineeringを知っていた」と答えたのは、わずか5%だった。

📋 **ここが要点**

出力を評価する力も、プロンプトを組む力も、**検証は自動的には起きない**。教えて初めて身につく。

AIほど最初の見立てに引きずられる

アンカリング＝**先に示された診断**を1位のままにする傾向。

対象	先に示された診断が1位のまま
LLM（8種）	55.6%
研修医（PGY1-3）	21.2%
指導医	10.0%

内科外来ビネット9対・225回答（Sheppert A, et al. Int J Med Inform. 2026）。AIが見落とした症例では読影者の感度中央値も**71%→39%**（Taib AG, et al. Radiology. 2026）。

FRAMEWORK

DEFT-AI

5つのステップ

DEFT-AI — 5つの構成要素

要素	名称	尋ねること
D	Diagnosis, Discussion, and Discourse	臨床推論と、使ったAIツール・プロンプトのプロセス
E	Evidence	支持する根拠と反証する根拠の両方
F	Feedback	見落とした鑑別・知識のギャップの内省
T	Teaching	推論とAI利用の両方への一般原則
AI	AI Engagement Recommendation	次はどの監督レベルでAIを使ってよいか

まず学習者の鑑別を先に言わせる

学習者の臨床推論プロセスとAI利用のしかたを、AIの答えより先に確認する。

📋 **尋ねること**

どのAIに、どんなプロンプトを打ったか。出力が診断を**influenced**／**replaced**／**augmented**のどれをしたか。

支持する根拠と反証する根拠の両方を

代替仮説を生成できるかどうかは、adaptive expertiseの証。AIリテラシーの自己評価も同時に行わせる。

📋 **問いかけの例**

「そのAIは**どうやってその結論に至った**と思うか」

見落としと知識の穴を本人に内省させる

自己評価を土台に、成長機会を学習者自身の言葉にさせる（guided self-reflection）。

📋 **振り返らせる3点**

見落とした鑑別・医学知識のギャップ・AIリテラシーの3つ。

次はどの監督レベルで AIを使ってよいか、言葉にして渡す

原則は "encourage ongoing practice with AI"。**使うな、ではない。**

1. **間接的または断続的な直接監督のもとで、慎重に使ってよい**
2. **監督なしで安全に使えるが、継続的な自己モニタリングを伴う**

centaur と cyborg

行動	向いている場面	診断・意思決定
centaur（分担）	高リスク・不確実な診断	自分の判断で必ず検証
cyborg（協働）	低リスク・定型的な業務	AIと反復してすり合わせる

両者は排他的ではない。タスクとリスクに応じて行き来できることがadaptive practice（Abdulnour 2025）。

30年以上前からある型のAI拡張

SNAPPSはこの系譜に含まれない（原著本文に0件）。

効果はまだ証明されていない

DEFT-AIはフレームワークの提案であり、効果を断定する研究はまだ無い。

- 原典は総説（Review）で、有効性を示す実証研究はまだ無い
- 類縁のSNAPPSですらKirkpatrick Level 3/4の報告はゼロ
- deskillingのエビデンスも「乏しい」（Heudel 2026）
- 査読後に書簡3本と著者の返答が出ている
  （うち1本は慶應義塾大学病院から）

指導医が明日から変えられる3点

01

先に言わせる

学習者の答えを
先に言わせる。

02

使い方を聞く

どのAIに、
何を打ったか聞く。

03

レベルを渡す

次はどの監督レベルで
使ってよいかを渡す。

ご視聴ありがとうございました

概要欄をチェック

要点まとめを配布中

チャンネル登録

毎週、医学教育の話をお届けします

高評価で応援

次回も**お楽しみに**

これからも、信頼できる医学教育情報を
わかりやすくお届けしてまいります。
引き続き、医知創造ラボをよろしくお願いいたします。