診察時間は1.27分短くなった。
それでもAIが診療を代替したとは言えない

2,069人のPreA無作為化試験を、登録差・盲検・健康転帰まで読む

Daiki Morimoto / Unknown Lab / ORCID 0009-0009-6444-750X
批判的論評 Version 1.0 2026年8月27日
DOI: 10.5281/zenodo.22103180

要旨

Taoらは、中国西部の三次医療機関2施設・24診療領域で、患者とケアパートナー2,069人を、診察前LLMの独力使用、医療助手の操作支援付き使用、非使用へ無作為化した。PreA独力群では、専門医との対面診察が平均4.41分から3.14分へ短縮し、話しやすさと医師評価のケア調整も改善した。

ただし全群で専門医の通常診療が続き、患者には平均3.51分の入力が加わった。診断精度、安全性、健康・生活、費用、経路全体の時間は主要評価ではない。試験登録と募集前プロトコル・論文で主要評価が異なり、変更理由を公開資料から追えない。本稿は効率を負担の移動と分け、ETIC 2.0へ時間・責任・成果を経路全体で追う更新案を提示する。

AIへ話した内容が、診察前に短いレポートへ整理される。専門医はそのレポートを読み、本人と対面する。

この仕組みで、対面診察は平均4.41分から3.14分へ短くなった。患者は話しやすいと評価し、専門医はケアの調整に役立つと評価した。

2,069人を対象にした実患者の無作為化試験である。試験問題や架空症例ではない。患者向けLLMを考えるうえで、見過ごせない前進である。

それでも、この結果を「AIが専門職を代替した」と読むことはできない。

AIの後には、専門医の通常診療が続いた。測ったのは主に時間と直後の評価であり、診断エラー、安全性、健康、生活、費用は主要評価ではなかった。さらに、公開された試験登録とプロトコル・論文では、主要評価の構成が変わっている。

本稿は、TaoらのPreA試験を、効果があったかどうかだけでなく、何へ割り付け、誰が何分使い、何を測り、何を測らなかったかまで読む。[1]

30秒で分かる研究の要点

項目	内容
研究	中国西部の三次医療機関2施設で行った個人無作為化比較試験
対象	2,069人の患者・ケアパートナー、111人の専門医、24診療領域
期間	2025年2月8日から4月30日
介入	専門診療前に患者がPreAと対話し、生成レポートを専門医が確認
比較	PreAを独力使用、医療助手の操作支援付き使用、PreAなし
主要結果	対面診察3.14対4.41分、話しやすさ3.99対3.44、ケア調整3.69対1.73
直接言えること	本試験条件で、診察前PreAは対面診察時間と直後の評価を改善した
言えないこと	診断・治療の正確性、長期安全性、健康・生活、費用、専門職代替
重要な監査点	登録簿と論文で主要評価が異なり、変更理由は公開資料から確認できない

English Summary

Tao and colleagues randomized 2,069 patients and care partners across two tertiary hospitals and 24 specialties to autonomous PreA use, assistant-supported use, or no PreA. Autonomous use reduced face-to-face specialist consultation time from 4.41 to 3.14 minutes and improved immediate ratings of communication ease and care coordination. All groups still received routine specialist care, patients spent an additional 3.51 minutes using PreA, and the trial did not establish diagnostic accuracy, safety, health or participation outcomes, cost, or shorter total pathway time. The preregistered outcome set also differs from the pre-recruitment protocol and publication without a publicly traceable explanation. This commentary therefore treats PreA as a consultation-preparation and handoff intervention, not professional replacement.

Keywords: large language model, chatbot, randomized controlled trial, specialist care, care transition, AI governance, ETIC 2.0

本稿の位置づけ

本稿は公開済み査読論文、補足資料、試験登録、査読履歴、公開コードをUnknown Labが批判的に読み解いた未査読の論評である。著者の結論、無作為化試験から直接言えること、観察結果、因果効果として言えないこと、Unknown Labの論評、ETIC 2.0への支持・挑戦・更新案を分けて記載する。個別の診断、治療、受診判断、AI製品の導入を指示しない。

補強記事07から、何を一歩深めるのか

補強記事07「AI相談は、専門職の代わりになれるのか」では、患者向けLLMの研究を静的回答、合成症例、実患者研究、前向き試験、実運用に分けた。そして、現時点で妥当な近接用途は、AIが答えを確定することより、本人が専門職へ相談する準備を支えることだと整理した。[5]

PreA試験は、その境界に最も近い実患者研究の一つである。

患者は専門外来の前にAIと対話する。AIは病歴、診断候補、検査候補を含むレポートを作る。専門医はそれを読んでから、通常の対面診療を行う。

ここで確かめたいのは、AIが医師より正しいかではない。

- 診察前整理は、実際の診療時間と対話を変えたのか。
- その短縮は、患者と専門医を合わせた経路全体の短縮なのか。
- 共創は何を変え、何をまだ証明していないのか。
- ETIC 2.0は、効率だけでなく生活・安全・責任をどう評価すべきか。

この論文は、先行研究へ何を足したのか

患者向けLLMの研究には、試験問題への正答、合成患者との対話、公開質問への回答、医療者向けの下書き支援が多い。どれも技術の一面を確かめるが、実際の患者が外来で使い、専門職の仕事へ接続されたときの効果とは同じではない。

Taoらの研究が加えたのは、実患者を診察前のPreAあり・なしへ無作為化し、電子記録から対面診察時間を測った点である。患者がAIへ話し、専門医がそのレポートを読むという、一連の運用を試験にした。

一方、論文タイトルにあるprimary-to-specialist care transitionsには注意が必要である。

参加者は一次医療で紹介の必要性を判断された人ではない。すでに専門外来を予約し、待合室にいた患者とケアパートナーである。PreAが不要な専門受診を減らしたか、適切な診療科へ紹介したかは検証していない。

査読者もこのずれを指摘した。著者は、PreAは新しく紹介を決める道具ではなく、自己紹介が多い高負荷外来で、専門診療前の情報不足を補う道具だと回答した。[1]

したがって、この研究の直接の対象は「一次医療から専門医への紹介」全体ではなく、専門外来の直前に行う情報整理である。

誰が参加し、どう選ばれたのか

研究は、中国西部の学術的三次医療機関2施設、24診療領域で行われた。

適格性を確認した患者・ケアパートナーは2,332人だった。154人が参加を断り、40人が同意書へ署名せず、2,138人が無作為化された。割付後、診察を受けなかった人と個人的理由で中止した人が計69人おり、最終解析は2,069人だった。

群	割付	未完了	最終解析
PreA-only	712	21	691
PreA-human	713	24	689
No-PreA	713	24	689

平均年齢は47.6歳、標準偏差14.6だった。女性1,141人、男性928人。患者本人は1,620人、ケアパートナーは449人だった。月収2,000人民元未満が770人、小学校以下の教育歴が313人含まれた。

年齢20-80歳で、携帯電話を使って対話し、受診後質問票へ回答できること等が組入条件だった。心理的障害や、LLMとの対話に不適切と判断された医学的事象がある人は除外された。

この除外は安全上理解できる。しかし、心理的危機、高リスク状態、緊急度判断こそ患者向けAIで重大な場面でもある。低リスク側の試験結果を、そのまま高リスク相談へ広げてはいけない。

期間を三つに分けて確認する

論文が報告する実際の募集期間は、2025年2月8日から4月30日である。

補足資料に収録された英語・中国語のプロトコルと統計解析計画は、Version 1.0、2025年1月30日付だった。実際の募集開始より前であり、3つの主要評価と解析方法を記載している。[2]

一方、中国臨床試験登録ChiCTR2400094159は2024年12月17日に前向き登録された。登録上の募集予定は2024年12月18日から2025年2月16日、実施予定は2024年12月16日から2月16日である。公開登録の最終更新も2024年12月17日のままで、実際の期間へ更新されていない。[3]

前向き登録そのものは行われている。ただし、登録、プロトコル、実施の時間関係を第三者が追うには、日程変更と理由の更新が必要だった。

三つの群で、何が違ったのか

PreA-only

患者またはケアパートナーが、自分の携帯電話でPreAを使った。医療助手の操作支援はない。AIは複数回の質問を行い、病歴、主訴、症状、診断候補、検査候補等を構造化したレポートへまとめた。専門医は診察前にそのレポートを読んだ。

PreA-human

同じPreAを、医療助手の操作支援付きで使った。支援は、WeChatに似た操作だと説明し、技術的な利用を助けるものだった。臨床判断を助手が代わりに行う条件ではない。

No-PreA

PreAを使わず、専門医は年齢と性別だけの通常レポートを見て診察した。

3群すべてで、最後は専門医が通常の対面診療を行った。PreA-onlyのonlyは「患者が操作支援なしで使った」という意味であり、「AIだけで診療した」という意味ではない。

手法の中核は、AI本体より「受け渡し」にある

PreAの患者向け画面は、音声または文字で情報を集める。対話は8-10往復を目標にし、最終的に1-3の診断候補を、支持する情報と反対する情報とともに示す。専門医向け画面には、構造化された紹介レポートを出す。

大切なのは、AIの出力が患者の画面で終わらないことである。

患者が入力する。AIが整理する。専門医が読む。本人と専門医が対面する。専門医が判断する。

この受け渡しの設計が介入の中核である。

患者はPreAと平均3.51分対話し、平均9.10往復した。専門医はPreAレポートを平均0.25分読んだ。対照レポートの閲覧は0.07分だった。

つまり、対面診察だけを見れば短くなっても、患者側には新しい入力時間が加わる。介入を評価するときは、専門医の診察時間だけでなく、患者、支援者、医療助手、専門医を合わせた経路全体を測る必要がある。

何を主要評価にしたのか

プロトコルと論文では、主要評価を次の三つとした。

- 患者と専門医の対面診察時間
- 専門医が評価したケア調整への有用性
- 患者が評価した話しやすさ

診察時間、患者数、臨床記録はPreAプラットフォームと病院の電子記録から取得した。ケア調整と話しやすさは5段階評価だった。質問票は20人の一般参加者と5人の臨床家による表面的妥当性、各領域Cronbachのalpha 0.80超を報告する。

無作為化は個人単位1:1:1、層別なしのコンピュータ生成で、割付は隠蔽された。患者は自分の群を知っていた。専門医はPreA-onlyかPreA-humanかを知らなかったが、PreAレポートと年齢・性別だけの対照レポートは内容から区別できる。患者報告と医師報告を完全に盲検化した試験ではない。

最低目標2,010人は、90人のpilotから検出力80%以上、alpha 0.05として計算された。ただし公開文書は、3主要評価があるのに「the primary outcome」と単数で書く。どの主要評価の効果差と分散で人数を決めたか、3つの主要評価全体で第一種過誤をどう管理したかは十分に追えない。

主要数値は、何を示したのか

評価	PreA-only	No-PreA	結果と注意
対面診察時間	3.14 ± 2.25分	4.41 ± 2.77分	絶対1.27分、相対28.7%短縮、95% CI 22.7-34.8、p<0.001
話しやすさ	3.99 ± 0.62	3.44 ± 0.97	相対16.0%増、95% CI 13.5-18.5、p<0.001。患者は非盲検
ケア調整	3.69 ± 0.90	1.73 ± 0.95	相対113.1%増、95% CI 107.4-118.7、p<0.001。医師のレポート評価
患者のPreA操作	3.51 ± 1.50分	なし	対面時間とは別に加わる
医師のレポート閲覧	0.25 ± 0.08分	0.07 ± 0.06分	PreA群で0.18分長い

主要3評価は、PreA-onlyがNo-PreAより良かった。PreA-onlyとPreA-humanには有意差がなかった。

この結果から、患者が操作支援なしでもPreAを使えたとは言える。しかし、診断や治療をAIが自律的に行えたわけではない。専門医がレポートを読み、対面診療し、最終判断した。

患者が評価した専門医の注意、尊重、満足、将来も使いたいという評価もPreA-onlyで高かった。これらは副次評価で、患者が自分の割付を知っていた影響を除けない。

1.27分は、誰の時間を短くしたのか

対面診察は1.27分短くなった。高負荷外来で、この差が小さいとは限らない。

80組の医師をマッチした追加解析では、参加医師は1勤務あたり28.54人、非参加医師は24.76人を診た。差は15.3%、p=0.005だった。待ち時間は33.54対34.65分で差がなかった。

ただし、患者数の比較は患者の無作為化3群そのものではない。参加医師と非参加医師を診療領域、勤務、年齢群、性別、職位で合わせた観察的なマッチドペア解析である。医師が仕事量を調整した可能性も著者自身が認める。

さらに、患者は診察前に3.51分AIと話し、医師はレポートを0.18分長く読んだ。

時間の場所	PreAで起きた変化
患者・ケアパートナーの入力	約3.51分増えた
医療助手の支援	支援群で発生。人数・総時間は未報告
専門医のレポート閲覧	約0.18分増えた
対面診察	約1.27分短くなった
待ち時間	有意差なし
訂正・事故対応	未測定

効率を評価するなら、時間が消えたのか、別の人・別の場所へ移ったのかを分けなければならない。

登録簿と論文で、主要評価が変わっている

ChiCTRの初回登録と、募集前プロトコル・論文を並べると、主要評価の構成が一致しない。[2][3]

指標	ChiCTR V1.0	プロトコル・論文
診察時間	主要	主要
満足度	主要	副次
待ち時間	主要	探索的
ケア調整	有用性等は副次	主要
話しやすさ	傾聴性等は副次	主要

プロトコルは実際の募集開始より前に作られている。したがって、この差だけで結果に合わせて評価を選んだと断定することはできない。

一方、登録簿は第三者が事前計画を確認するためにある。主要評価を変えたなら、変更日と理由を登録へ残す必要がある。公開資料では、その説明を確認できなかった。

この試験の主要結果を読むときは、前向き登録はあるが、登録から最終論文までの主要評価変更を追跡できないという条件を併記するのが妥当である。

著者の結論、観察結果、因果効果を分ける

著者の結論

著者は、PreAが高負荷外来で効率と患者中心のケアを改善し、共創は地域会話を受動的に追加学習するより有効な導入戦略だと結論する。PreA-onlyとPreA-humanの同等な結果から、自律運用と拡張可能性も強調する。

無作為化試験から直接言えること

本試験の対象・施設・期間では、PreA-onlyへ割り付けることで、No-PreAより対面診察時間が短くなった。直後の患者評価の話しやすさと、医師評価のレポート有用性も高くなった。

これは介入の因果効果として解釈できる。ただし、自己報告評価には非盲検の影響が残る。

観察されたが、無作為化した効果ではないこと

参加医師が1勤務で診た患者数、PreAレポートと専門医記録の一致、レポート品質、臨床記録分類器の結果は追加解析である。主要3群の無作為化比較と同じ強さでは読めない。

因果効果として言えないこと

- 診断、検査、治療が正確になった。
- 有害事象、再受診、受診遅延が減った。
- 健康、QOL、生活、作業参加が改善した。
- 患者と医療者を合わせた総時間や費用が減った。
- 一次医療の紹介判断が改善した。
- 専門職を代替できた。
- 日本、自宅、緊急相談、長期運用でも同じ効果がある。

共創は、何を変えたのか

PreAの敵対的テストには、患者・ケアパートナー120人、地域保健従事者36人、医師15人、看護師38人、計209人が参加した。11省の都市・農村が含まれた。

低ヘルスリテラシーでも使える言葉、8-10往復という長さ、専門医が読むレポート、複数診療領域、安全上の失敗例等を調整した。600の合成患者プロフィールを作り、半数を改良、残りを比較に使った。

これは、モデル精度だけでなく、本人の画面と専門職のワークフローを同時に設計した点で重要である。

ただし、209人という人数だけでは、参加の質は分からない。

誰をどう募集したか。何回、何時間参加したか。患者の意見で何が変わったか。専門職と意見が割れたとき誰が決めたか。採用されなかった意見は何か。謝金はあったか。

原著と補足資料だけでは、これらを十分に追えない。共創を評価するには、人数だけでなく変更権限と意思決定の履歴が必要である。

共創版とfine-tuning版の比較を、広げすぎない

研究チームは、11省から集めた患者・医師会話515件を使い、共創したPreAへ追加fine-tuningした版を作った。そして、追加学習なしの共創版と、合成患者300例ずつで比較した。

領域	共創版	地域会話追加版
病歴聴取	4.56 ± 0.65	3.86 ± 0.81
診断	4.67 ± 0.55	2.47 ± 1.44
検査提案	4.23 ± 1.09	2.21 ± 1.12

すべて$p<0.001$で、追加学習版の評価が低かった。追加学習版は、地域診療にある省略、診断なし、検査提案なし、好ましくない口調まで再現したと著者は解釈した。

この結果は、「現場データを足せば地域適合になる」とは限らないことを示す。

同時に、「共創はfine-tuningより常に優れる」とは言えない。実患者を二つの開発法へ無作為化した試験ではない。比較したのは特定の地域会話、モデル、学習法、合成患者、評価者である。高品質に選別したデータや別のfine-tuning法との比較もない。

支持されるのは、低品質を含む現場データの受動的追加が、現場の弱点まで増幅し得るという限定的な警告である。

臨床記録のF1式には、誤植と思われる箇所がある

著者は、専門医がAIの診断候補をそのまま採用した可能性を調べるため、PreA支援群とNo-PreA群の臨床記録を分類器で見分けられるか分析した。

報告されたF1は0.57、p=0.81で、著者は記録に系統的な差を検出できなかったとした。

方法欄のF1式は、次のように書かれている。

F1 = TP / (2TP + FP + FN)

標準的なF1式は次である。

F1 = 2TP / (2TP + FP + FN)

原文は分子の2が抜けた誤植と考えられる。公開コードではscikit-learnのf1_scoreを使っており、実計算は標準式とみられる。[4]

問題は誤植だけではない。

分類できなかったことは、サンプル化した臨床記録に、分類器が見つけられる系統差がなかったことを示す。しかし、専門医が個別にAIの候補へ影響されなかったこと、安全だったこと、診断が正しかったことは証明しない。

公開コードには臨床記録データがなく、乱数seedも固定されていない。単一の訓練・テスト分割で、帰無F1と0.02の閾値の説明も十分ではない。著者の「直接採用に反する説得的証拠」という表現は、分析が直接支える範囲より強い。

レポートと専門医記録の一致は、診断精度ではない

事後解析では、PreAレポートと専門医記録の一致が、病歴65.8%、診断66.7%、検査70.7%だった。専門家評価では、一致した例や専門医記録が空欄の例で、PreAレポートの完全性等が高かった。

しかし、専門医記録の23.0-31.9%は空欄だった。一致の基準は、独立した診断ゴールドスタンダードではない。専門医記録と同じなら正しい、違えば誤りという設計でもない。

長く完全に見えるレポートが、患者にとって正しい診断や安全な検査選択を意味するとは限らない。

この解析は、PreAが情報を多く含むレポートを作れた可能性を示す。診断精度、臨床安全性、健康効果の証拠としては使えない。

Unknown Labの批判的論評

1. このRCTは重要だが、「紹介」より「診察前整理」の試験である

実患者2,069人の無作為化は、患者向けLLM研究として大きい。電子記録由来の診察時間も、好ましさだけを測る静的比較より直接的である。

だからこそ、用途を正確に限定したい。

PreAは専門受診の必要性を判断せず、専門医を置き換えなかった。すでに受診する人の情報を診察前に整理した。この限定は弱点ではない。むしろ、患者向けAIが比較的安全に価値を出せる場所を具体化している。

2. 相対113.1%は、尺度の意味を大きく見せる

ケア調整は1.73から3.69へ上がった。差は1.96点で、無視できない。

ただし、5段階評価で低い対照平均を分母にすると、相対差は113.1%になる。比率は印象が強いが、尺度の上限や臨床的に意味のある最小差は示されていない。

絶対値、尺度、回答者、盲検の状態を併記する方が、結果を正しく読める。

3. 効率は、経路全体と負担の移動で測るべきである

専門医の対面時間は短くなった。しかし患者の入力時間、医療助手の支援、レポート確認、訂正、通信障害、事故対応を足した総時間は分からない。

医師の時間を患者や家族へ移すだけなら、組織効率が上がってもケア負担が増えることがある。ETIC 2.0では、誰の時間が減り、誰の無償作業が増えたかを同じ表で追う必要がある。

4. 主要評価の変更履歴を残すべきだった

登録簿と募集前プロトコルの差は、試験結果を無効にしない。しかし、どの評価を主要とみなすかは、結果を見る前に固定されているほど信頼しやすい。

満足度と待ち時間を主要から外し、ケア調整と話しやすさを主要にした理由を、登録更新として残すべきだった。今後のAI介入研究では、モデル版だけでなく、評価指標の変更履歴も公開する必要がある。

5. 費用、人員、責任、継続可能性が効果の外にある

PreAはGPT-4.0 miniを使い、Tencentの計算資源と技術支援を受けた。Tencent WeChat AI社員2人が共著者である。PreA本体は商用ライセンスと医療機器化の検討を理由に公開されていない。

API費、端末、通信、医療助手、保守、モデル更新、障害対応、専門医確認、事故時の救済に誰がいくら負担するかは評価されていない。

短期RCTで動くことと、医療機関が安全に継続できることは別である。

ETIC 2.0への支持・挑戦・更新案

区分	この論文から受け取ること	ETIC 2.0で明確にすること
支持	AI、本人、ケアパートナー、専門医、画面、レポート、組織が一つの介入を作る	AI単体性能ではなくケア生態系全体を評価する
支持	答えの代替より、専門職へ渡す情報整理に実患者RCTの根拠がある	AIの役割を相談準備・構造化・引き渡しへ限定する
挑戦	共創209人でも、決定権と異論の履歴が見えない	参加人数でなく、何を変更できたかを記録する
挑戦	時間と直後評価は改善したが、生活・安全は未測定	近接成果と健康・生活・作業参加を分ける
挑戦	登録評価と論文評価が変わった	指標・モデル・prompt・運用の変更履歴を同じ台帳で管理する
更新	時間が専門医から患者へ移る可能性がある	経路全体の時間、費用、責任、救済を一行で追う

更新案: 経路全体の時間・責任・成果を一行で追う

段階	時間・負担	責任	測る成果
本人の入力	入力時間、不安、個人情報、訂正	用途説明と同意を誰が担うか	完了率、訂正率、使わない選択
AIの整理	応答時間、誤り、欠落	モデル・prompt・更新の責任者	情報欠落、危険提案、格差
支援者の補助	家族・助手の時間	無償労働と代理入力の境界	支援量、本人の意思反映
専門職の確認	閲覧・照合・修正時間	最終判断、上書き、記録	診断安全、修正率、認知負荷
対面診療	診察時間、待ち時間	説明と共同意思決定	話しやすさ、理解、治療選択

段階	時間・負担	責任	測る成果
導入後	保守、監視、事故対応	組織、開発者、調達者	健康・生活、作業参加、費用、害

この表は、短縮を否定するためではない。短縮が誰の利益になり、どこへ負担を移し、どの責任を新しく生むかを見えるようにするためである。

次に読む論文

次は、LiらのP&P Care無作為化試験を読む。[6]

中国11省の地域住民2,113人全員が同じ患者向けLLMを使い、その前にe-learningを受ける群と受けない群へ無作為化された。主評価の客観的健康二ーズ認識は2.95対2.34だった。

この論文は、AIを使うか使わないかの比較ではない。AIを使う前の教育が、AIとの対話をどう変えるかを検証している。

PreAが病院内で専門職へ情報を渡す研究なら、P&P Careは地域で使う前のリテラシー、農村・都市差、方言、通信、通常診療対照の欠如を考える研究である。両研究は一部チームが重なるため独立追試ではない。その点も含めて読む必要がある。

結論

Taoらの試験は、患者向けLLM研究を一歩前へ進めた。

実患者が専門診療前にAIを使い、専門医がレポートを確認する。2,069人の無作為化比較で、対面診察は1.27分短くなり、話しやすさと医師評価のケア調整が高くなった。

これは、相談準備と専門職判断を組み合わせる価値を支持する。

同時に、AIが専門職を代替した証拠ではない。診断精度、安全性、健康、生活、費用は確かめていない。患者の入力時間を含む経路全体の効率も分からない。登録簿と論文で主要評価が変わり、その理由を追えない。共創の人数は多いが、参加者の決定権と異論の履歴は薄い。

ETIC 2.0がこの研究から受け取るべきなのは、「AIを使えば速くなる」という結論ではない。

誰の時間を、誰へ移し、誰が確認し、何をもって成功とし、誤ったとき誰が止め、訂正し、救済するか。

診察前AIを導入するなら、この経路全体を評価する必要がある。

記事の位置づけ

本稿は公開済み査読論文、補足資料、試験登録、査読履歴、公開コードをUnknown Labが批判的に読み解いた未査読の論評である。個別の診断、治療、受診判断、AI製品の導入を指示しない。ETIC 2.0の更新案はUnknown Labの未検証提案である。

参考文献

1 Tao X, Zhou S, Ding K, et al. An LLM chatbot to facilitate primary-to-specialist care transitions: a randomized controlled trial. Nature Medicine. 2026;32:934-942. doi:10.1038/s41591-025-04176-7

2 Tao X, et al. Supplementary Information: study protocol and statistical analysis plan. Supplement to reference 1. Publisher supplementary materials

3 Chinese Clinical Trial Registry. ChiCTR2400094159. Evaluating the effectiveness of large language models-based health consultation models in improving patient experience and hospital operational efficiency. Registry record

4 Han S, Wu B, Li Y, Huang Q, Li S. PreA-OutpatientRCT: v1.0.0. Zenodo; 2025. doi:10.5281/zenodo.17330614

5 Morimoto D. AI相談は、専門職の代わりになれるのか. Unknown Lab Research Report 07. 2026. doi:10.5281/zenodo.22065352

6 Li S, Li Y, Zhou S, et al. A community-codesigned LLM-powered chatbot for primary care: a randomized controlled trial. Nature Health. 2026;1:238-250. doi:10.1038/s44360-025-00021-w

本稿は公開済み査読論文等を批判的に読み解いた未査読の論評であり、ETIC 2.0の有効性を検証した研究ではない。個別の診断、治療、受診判断、AI製品の導入を指示しない。

執筆: Unknown Lab編集部（AI支援）/ 専門職・研究方法確認済み