DAILY ENGLISH READING

29 August 2026 - 30-minute active reading pack

TODAY'S QUESTION

Why did AI improve trial-screening accuracy without saving time?

Difficulty: C1 | human-AI teaming | clinical trials | workflow evaluation

WHY SELECTED

Human+AI trial prescreening improved chart-level accuracy, especially for biomarker, staging and response criteria, but average review time stayed almost unchanged. For SP automation, this shows why quality, efficiency, safety and traceability should be measured as separate outcomes rather than assumed to improve together.

READING ORDER

0-3 min Preview: primary accuracy and secondary efficiency endpoints.

3-14 min Main: randomization, gold standard, accuracy, time and automation bias.

14-20 min TrialMatchAI: retrieval plus criterion-level eligibility classification.

20-25 min CDISC: reproducibility, traceability and measurable impact.

25-30 min Output: design an evaluation for one SP AI feature.

SOURCES

Main: Human-AI teaming for oncology trial prescreening - Nature Communications

https://www.nature.com/articles/s41467-026-68873-8

Support: TrialMatchAI - Nature Communications

https://www.nature.com/articles/s41467-026-70509-w

Support: 2026 CDISC AI Innovation Challenge

BRIEF BACKGROUND

355 oncology charts were reviewed by clinical research coordinators with or without AI.

Human+AI accuracy was about 76% versus about 71% for Human-alone.

Average review time was about 37 minutes per chart in both groups.

Improvements were strongest for biomarker, staging and response-related criteria.

Automation bias shows that augmentation can introduce new human error pathways.

SP lesson: validate quality, efficiency, failure modes and traceability separately.

KEY VOCABULARY

prescreening	预筛选	early review for likely trial eligibility
eligibility criterion	入排标准条目	rule for trial entry
noninferiority trial	非劣效性试验	tests not unacceptably worse
superiority test	优效性检验	tests whether one approach is better
chart-level accuracy	病历层面准确率	correct full-chart classifications
gold standard	金标准	trusted reference for correctness
neurosymbolic AI	神经符号人工智能	neural patterns plus explicit rules
automation bias	自动化偏误	over-trusting automated suggestions
human-AI teaming	人机协作	humans and AI work together
workflow efficiency	工作流效率	time and effort required
criterion-level classification	条目级分类	decision for each eligibility rule
biomarker criterion	生物标志物标准	rule based on molecular findings
staging criterion	分期标准	rule based on disease stage
traceability	可追溯性	link output to evidence and rules
measurable impact	可衡量影响	demonstrable quality or efficiency gain

USEFUL PHRASES

1. improve accuracy without reducing review time

2. evaluate the combined human-AI workflow

3. compare performance against a gold standard

4. define accuracy and efficiency as separate endpoints

5. identify where automation bias can occur

6. keep a human reviewer in the decision loop

7. measure impact beyond model-level performance

8. use criterion-level error analysis

9. preserve traceability from source evidence to decision

10. treat time saved as an empirical outcome

COMPREHENSION QUESTIONS

1. Why were accuracy and review time separate endpoints?

2. What does unchanged review time tell us about workflow efficiency?

3. Why may biomarker and staging criteria benefit more from AI?

4. How can automation bias reduce the value of an accurate system?

5. Which metrics should evaluate an AI-assisted SAS or ADaM review workflow?

RETELLING PROMPTS

30 sec: study design -> accuracy result -> efficiency result.

45 sec: explain why Human+AI can be better without being faster.

60 sec: apply the lesson to SAS logs, SDTM, ADaM or TFL QC.

5-MINUTE SPEAKING / OUTPUT TASK

Minute 1: choose one SP AI feature.

Minutes 2-3: define quality, efficiency, safety and traceability endpoints.

Minute 4: choose Human-alone, AI-alone and/or Human+AI comparison arms.

Minute 5: state the threshold required before production scaling.

ONE SENTENCE TO KEEP

An AI workflow is valuable only when the combined human-machine system improves the outcomes that matter; accuracy, time saved, safety and traceability must be measured separately rather than assumed to move together.