How mature is AI in real clinical trials?

15 August 2026 · C1 · 30 minutes

WHY SELECTED

This open-access analysis covers 8,532 AI-related clinical trials across 32 specialties and separates technical progress from translational maturity. For clinical-programming automation, the same distinction matters: benchmark success is not the same as controlled production evidence.

READING ORDER

0-3 preview · 3-14 main article · 14-20 ClinicalTrials.gov API · 20-25 FDA/EMA principles · 25-30 speaking/output.

SOURCES

Main: The Clinical Trial Pipeline Reveals the Next Wave of Artificial Intelligence in Healthcare — https://arxiv.org/abs/2607.22607

Support: ClinicalTrials.gov API Migration Guide — https://clinicaltrials.gov/data-api/about-api/api-migration

Support: FDA/EMA Good AI Practice — https://www.fda.gov/about-fda/artificial-intelligence-drug-development/guiding-principles-good-ai-practice-drug-development

BRIEF BACKGROUND

The study identified 8,532 AI clinical trials; about 80% were registered from 2019 onward and 30.5% used randomized controlled designs. Imaging accounted for 2,475 trials. Clinical text/NLP rose roughly seven-fold from 2018 to 2025. The authors found 3,259 retrospective validations, 1,802 silent prospective evaluations, and only 184 Level 4 semi-autonomous or closed-loop trials; about 68% of the latter focused on glucose management.

For SP automation, use an evidence ladder: offline benchmark → retrospective replay → silent prospective use → human-in-the-loop production → bounded automation → lifecycle monitoring.

KEY VOCABULARY

translational maturity — 转化成熟度

prospective evaluation — 前瞻性评估

silent prospective evaluation — 静默前瞻性评估

randomized controlled design — 随机对照设计

clinical autonomy — 临床自主性

closed-loop system — 闭环系统

multimodal AI — 多模态人工智能

prognostic AI — 预后型人工智能

diagnostic AI — 诊断型人工智能

treatment recommendation — 治疗推荐

registry data — 注册平台数据

classification dimension — 分类维度

geographic representation — 地域代表性

algorithmic evidence — 算法层面的证据

clinical evidence — 临床证据

USEFUL PHRASES

• move from retrospective validation to prospective evaluation

• remain concentrated in a small number of use cases

• distinguish algorithmic performance from clinical impact

• support reproducible information retrieval

• define a clear context of use

• evaluate performance in the intended workflow

• monitor the system across its lifecycle

COMPREHENSION QUESTIONS

• What does the study reveal about the scale and growth of AI trials?

• Why is silent prospective evaluation an intermediate maturity stage?

• What does Level 4 concentration in glucose management suggest?

• Why are structured registry APIs useful for AI retrieval?

• How would you distinguish algorithmic from operational evidence for an SP agent?

RETELLING PROMPTS

30 sec: scale → maturity gap → autonomy. 45 sec: retrospective → silent prospective → human-in-the-loop → closed loop. 60 sec: apply the maturity ladder to SAS logs, ADaM or TFL QC.

5-MINUTE SPEAKING / OUTPUT TASK

Choose one SP AI feature. Define five evidence stages from offline benchmark to bounded automation. Add promotion criteria for accuracy, error rates, reproducibility, override, rollback and lifecycle monitoring. Finish with a 90-second recommendation stating when human control must remain mandatory.

ONE SENTENCE TO KEEP

AI maturity is not defined by how impressive a model looks in isolation, but by how safely, reproducibly and accountably it performs inside the workflow where decisions are actually made.