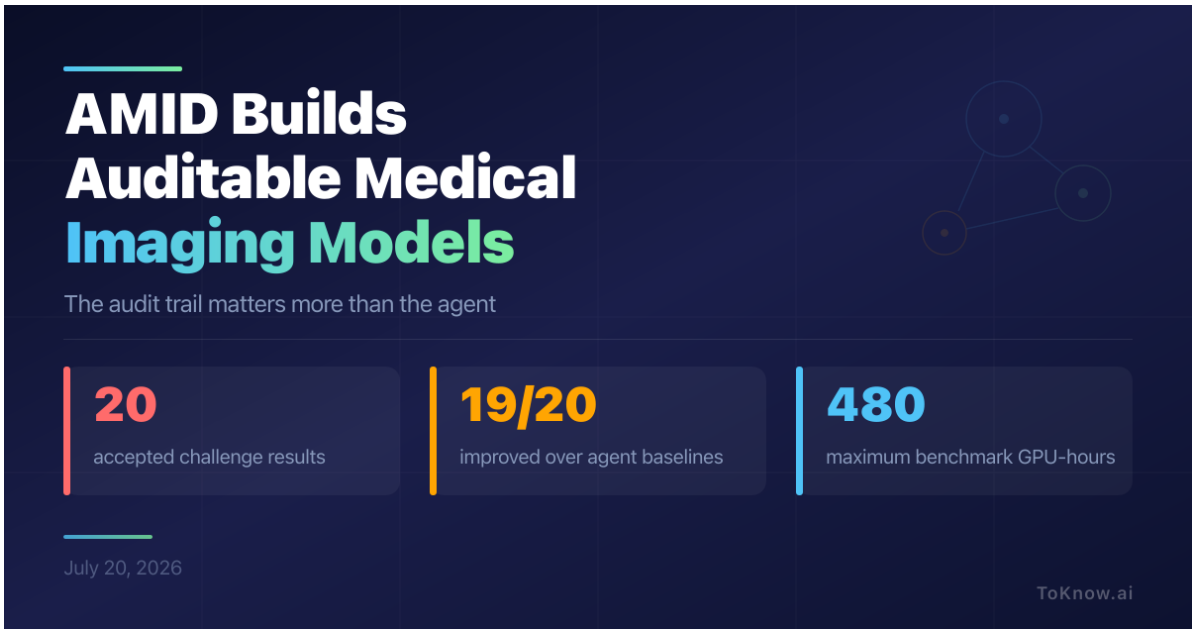
AMID Builds Auditable Medical Imaging Models

Kabui, Charles

2026-07-20

Read at ToKnow.ai



Researchers from CUHK, the Chinese Academy of Sciences, Microsoft Research, and Lehigh University built AMID, an agent system that develops complete medical-imaging models. It profiles a dataset, tests several methods, checks the data split and scoring code, then packages training code, model weights, predictions, and an audit trail. On the 20-task ReX-MLE benchmark, AMID produced an accepted result for every challenge, beat the strongest listed

general agent baseline on 19, and tied one. Each task had a 24-hour limit on one 48 GB RTX A6000. Its ISLES'22 stroke-segmentation score reached **0.71 Dice**, a measure of overlap between predicted and true regions, versus 0.04 for the best baseline. On NeurIPS CellSeg, it reached **0.90 F1** versus 0.36.

The audit checks matter more than the agent count. Medical models can look accurate when patient images leak between training and validation, the wrong metric is used, or prediction files are malformed. AMID checks those failures before selecting a model and preserves the evidence for review. It does not prove clinical safety: the benchmark uses retrospective competition data, and its human reference scores come from different test sets. Independent reproduction is also blocked because the AMID repository still lists source-code release as unfinished. For now, AMID shows a better way to automate medical model engineering, not a replacement for researchers or clinical trials.

Sources:

- AMID paper
- AMID repository and solution reports
- ReX-MLE benchmark paper
- ReX-MLE benchmark code

Disclaimer: *For information only. Accuracy or completeness not guaranteed. Illegal use prohibited. Not professional advice or solicitation.* ***Read more:*** */terms-of-service*