

MARS: Modular Agent with Reflective Search for Automated AI Research

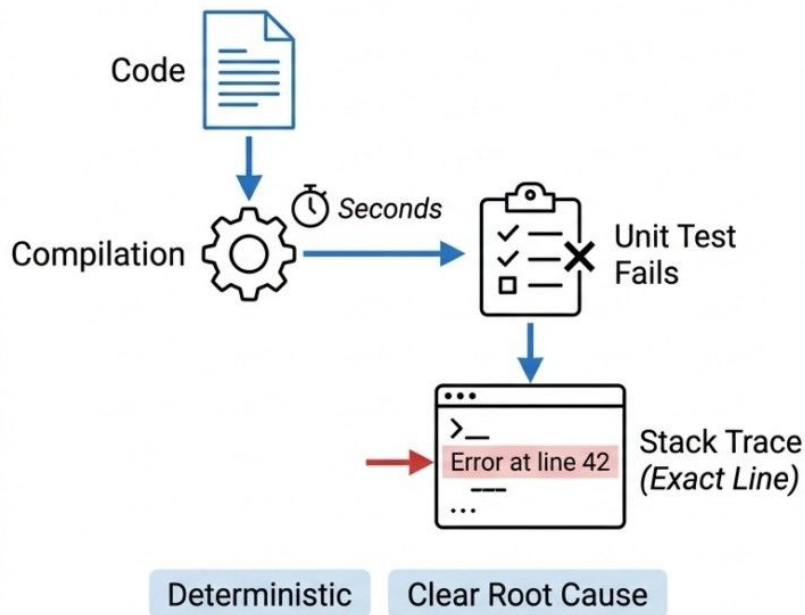
Jiefeng Chen, Bhavana Dalvi Mishra, Jaehyun Nam, Rui Meng, Tomas Pfister, and Jinsung Yoon

Cloud AI Research

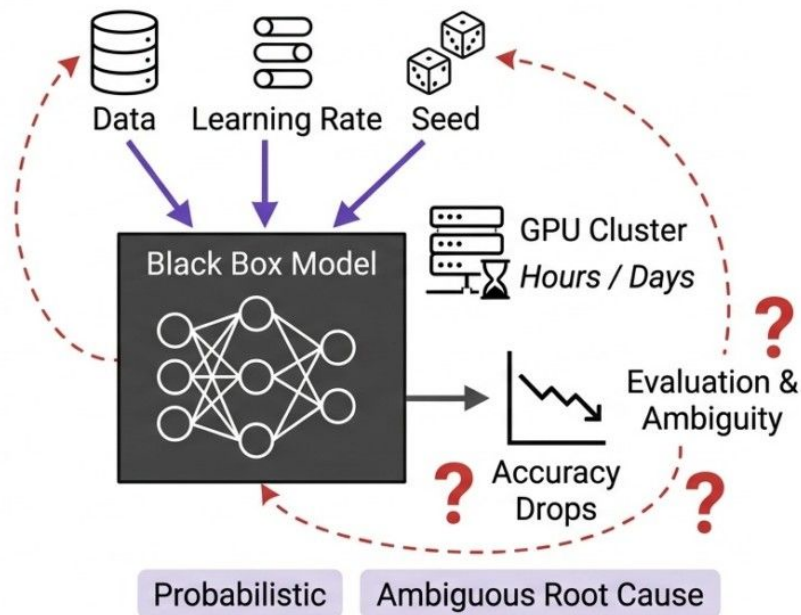


MLE: A Critical Bottleneck in Automating AI Research

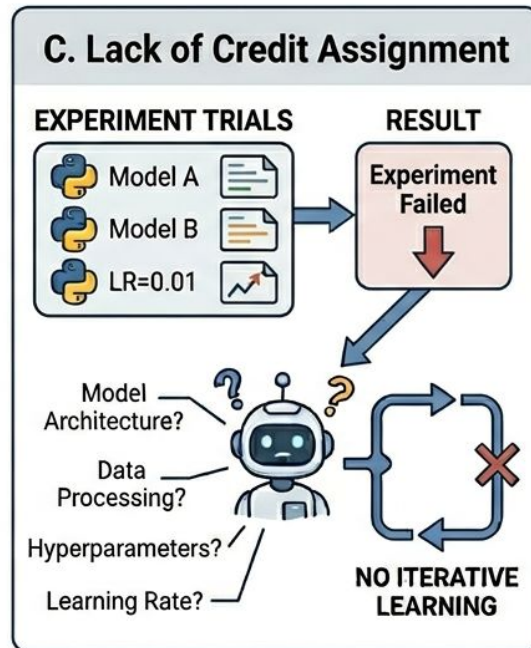
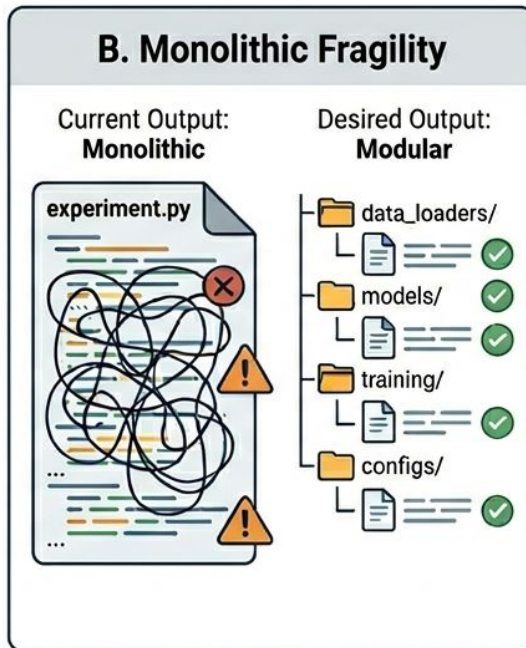
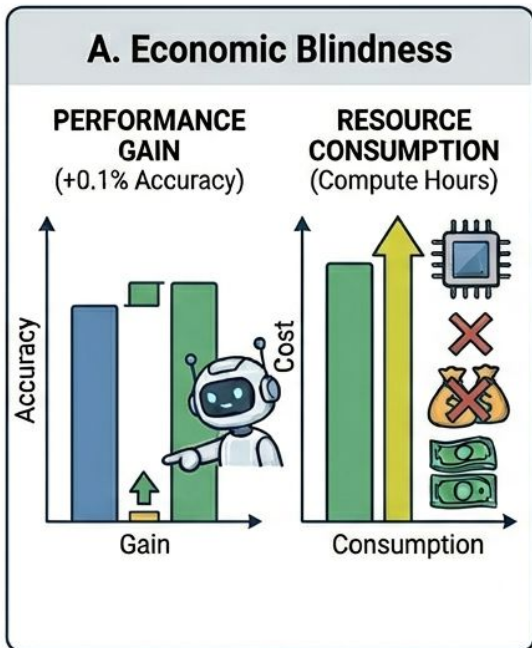
Traditional Software Engineering (SWE)



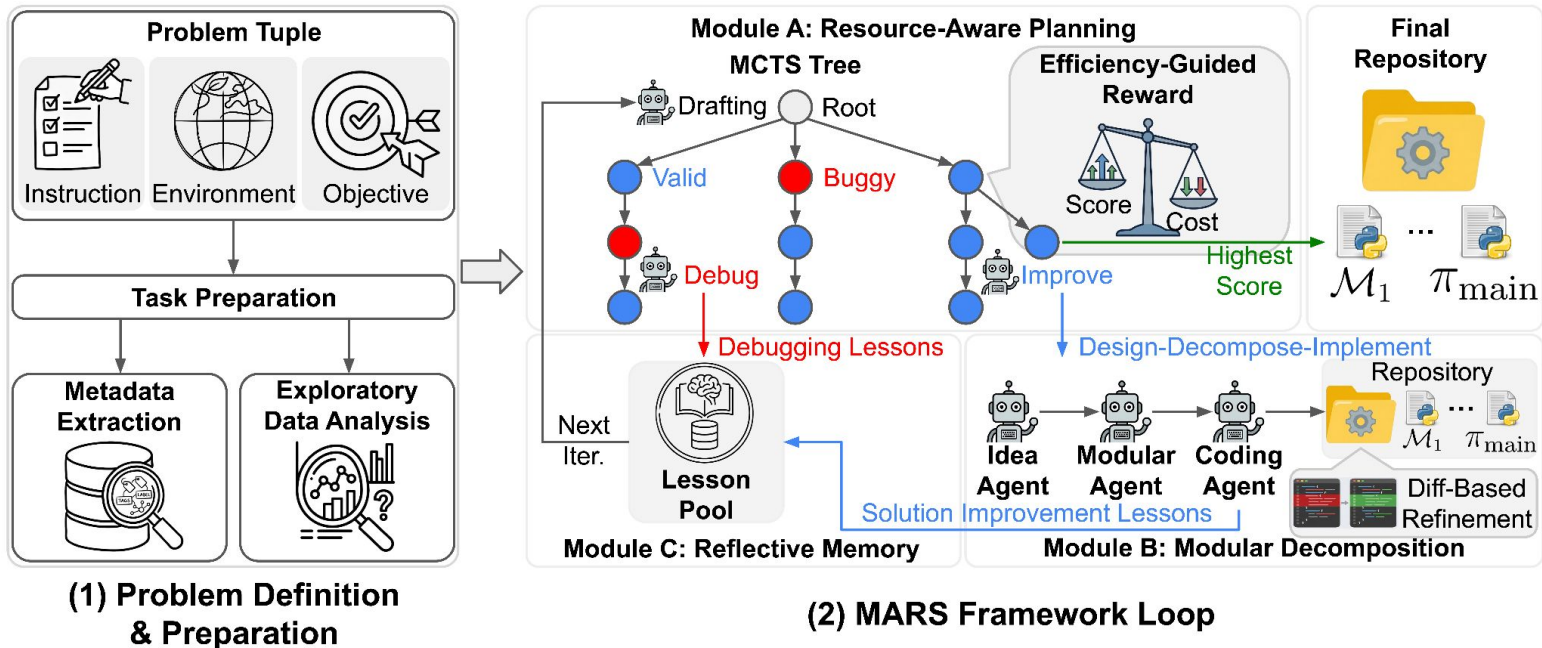
Machine Learning Engineering (MLE)



Limitations of Existing Approaches



MARS (Modular Agent with Reflective Search)

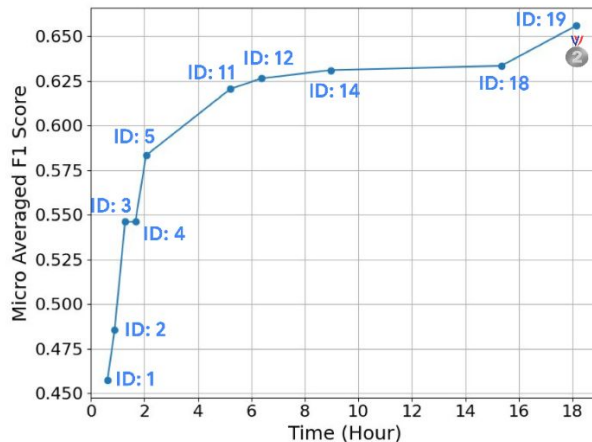
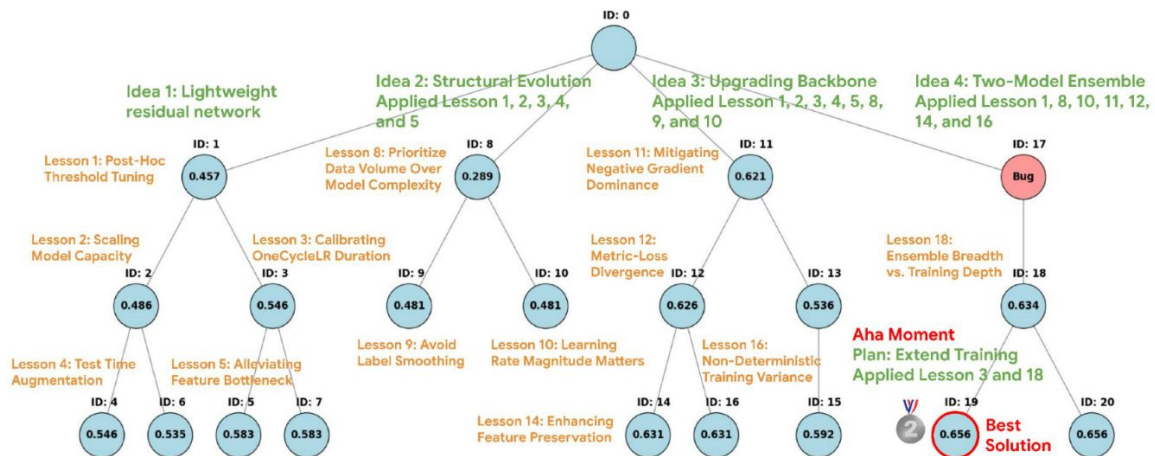


MLE-Bench Leaderboard

Agent	LLM(s) used	Low == Lite (%)	Medium (%)	High (%)	All (%)	Running Time (hours)	Date	Source Code Available
Famou-Agent 2.0	Gemini-3-Pro-Preview	80.3 ± 1.52	64.04 ± 2.32	42.22 ± 2.22	64.44 ± 1.18	24	2026-02-23	X
AIBuildAI	Claude-Opus-4.6	77.27 ± 0.00	61.40 ± 0.88	46.67 ± 0.00	63.11 ± 0.44	24	2026-03-06	X
CAIR MARS+	Gemini-3-Pro-Preview	78.79 ± 1.52	60.53 ± 1.52	44.44 ± 2.22	62.67 ± 0.77	24	2026-02-17	X
MLEvolve	Gemini-3-Pro-Preview	80.30 ± 1.52	57.89 ± 1.52	42.22 ± 2.22	61.33 ± 1.33	12	2026-02-14	✓
PiEvolve (Fractal AI Research)	Gemini-3-Pro-Preview ^[1]	80.30 ± 1.52 ^[2]	58.77 ± 0.88 ^[2]	40.0 ± 0.00 ^[2]	61.33 ± 0.77 ^[2]	24	2026-01-05	X
Famou-Agent 2.0	Gemini-2.5-Pro	75.76 ± 1.52	57.89 ± 1.52	40.00 ± 0.00	59.56 ± 0.89	24	2025-12-27	X
ML-Master 2.0	Deepseek-V3.2-Speciale	75.76 ± 1.51	50.88 ± 3.51	42.22 ± 2.22	56.44 ± 2.47	24	2025-12-16	X

- Ranked **#1** (Feb 2026)
- Currently **Top-3** on Leaderboard

The “Aha!” moment of MARS





Thank you.