//fta.mark-1: Technical Report

Nov 18, 2026 • Author: Senthil Kumar Hariram

Summary

FTA Mark-1 won all three of its official chess matches on 7 November 2026: 10 to 2 against GPT-6 Astra (OpenAI), 9 to 3 against Gemini 4 Argon (Google) and 9 to 3 against Gemini 3.8 Flash (Google). These three models were the top 3 on Kaggle Game Arena's chess leaderboard at the time (board last updated 19 October 2026). On the board updated 12 November 2026 they rank 2, 3 and 4, behind Gemini 4 Carbon, which Mark-1 did not play.

Mark-1 is built on Sarvam-30B, an open model from India, and trained with FTA's MicroJudgement Tuning™ (MJT). Sarvam-30B on its own could not play chess. After MJT, Mark-1 rated about 1,824 on a 24-game ladder against Stockfish.

Across the 36 match games, Mark-1 won 26, drew 4 and lost 6. Every game was played at each opponent's default settings, with the same prompt and referee for both sides, no clock, and no chess engine inside Mark-1.

The full Mark-1 system is released openly on Hugging Face (adapter underApache 2.0, code under GPL-3.0), so anyone can download it and play it against a person, against OpenAI or Gemini, or inside their own chess app. All 36 match games are published too.

What we claim: Mark-1 beat these three models in these matches. **What we do not claim:** that Mark-1 is a betterAI than them, or the best chess AI. Kaggle did not run these matches.

What Mark-1 is

Mark-1 is two parts that always work together: a trained judge, and FTA's think-ahead code that puts the judge to work.

> **1. The judge (the adapter).** A small add-on (LoRA adapter) trained by FTA on top of Sarvam-30B. Given a chess position, it scores how good each possible move is.
>
> **2. The think-ahead.** For each move in a game, this code asks the judge hundreds of small questions, such as "if I play this and they reply with that, how good is my position?" It then picks the move that holds up best.

There is no chess engine inside Mark-1. Every judgement comes from the trained model itself. Stockfish was used only as the teacher during training, never during play.

This is the main difference from the models Mark-1 played. GPT-6 Astra and Gemini think in one long chain, hidden inside their makers' servers. Mark-1 thinks in many small, separate checks, organised by code that anyone can read.

Item	Detail
Base model	Sarvam-30B by Sarvam AI (revision 071ae95), Apache 2.0
What FTA added	A LoRA adapter (the judge) and the think-ahead code
Released version	R4T, adapter fingerprint (sha256) 6bd8bfe6a9e39a351b33d59354b204bd413a54f2ba cba62a0920f05f0df2f037
Memory needed to run	An 80 GB GPU (H200 orA100-80GB). Sarvam-30B alone needs about 64 GB
Time per move at match strength	About 62 to 68 seconds in the official matches (H200); about 114 seconds on an A100-80GB
Licence	Adapter: Apache 2.0. Code: GPL-3.0. Game records: CC BY 4.0. Fully open, no sign-up

How Mark-1 was trained: Micro-Judgement Tuning (MJT)

Mark-1 was trained with Micro-Judgement Tuning (MJT), FTA's framework for teaching an open model good judgement in one field. In chess, MJT has worked well, with Stockfish as the teacher.

The usual way to train an AI is to show it the right answer and have it copy. MJT teaches the model to judge every option instead. At a top level it has three steps:

1. **Score every option.** In each training position, the model learns how good every legal move is, not just which move is best.
2. **Keep each lesson small.** Each lesson is one small, clear judgement that the model can learn well.
3. **Stack many small checks when it plays.** At play time, the think-ahead code combines hundreds of these small judgements to choose each move.

The key lesson from Mark-1: **a strong teacher matters more than lots of data.** With a good teacher, a general open model can learn sharp judgement in a narrow field.

MJT builds on published research, including DeepMind's work on scoring every move (see Credits). We do not claim it as a new invention. The training data, data mix and training settings stay private to FTA.

Match setup and fair-play rules

Every match followed the same fixed rules, set and locked before the first move was played.

Rule	How it worked
Games per match	12: six starting openings, each played once with Mark-1 as White and once as Black
Opponent settings	Each opponent at its maker's default settings, with no tools
Prompt and referee	The same prompt and the same referee code for both sides
Illegal or unreadable move	Two more tries, then the player's own top legal move. Same rule for both sides, logged every time
Clock	No clock for either side
Mark-1	Its own judge and think-ahead only. No engine, no outside help
Locked before play	Each match's code and settings were fixed in a saved, dated record before its first move
Model versions	Reported by the providers on every answer: gpt-6-astra (534 answers), gemini-4-argon (554), gemini3.8-flash (601)

In all 36 games the referee never had to step in, and no API call went unanswered. Two games were restarted, both before any move was played, and both are recorded in the match logs.

Mark-1 ran on rented NVIDIA H200 GPUs. All three matches were played on 7 November 2026, India time

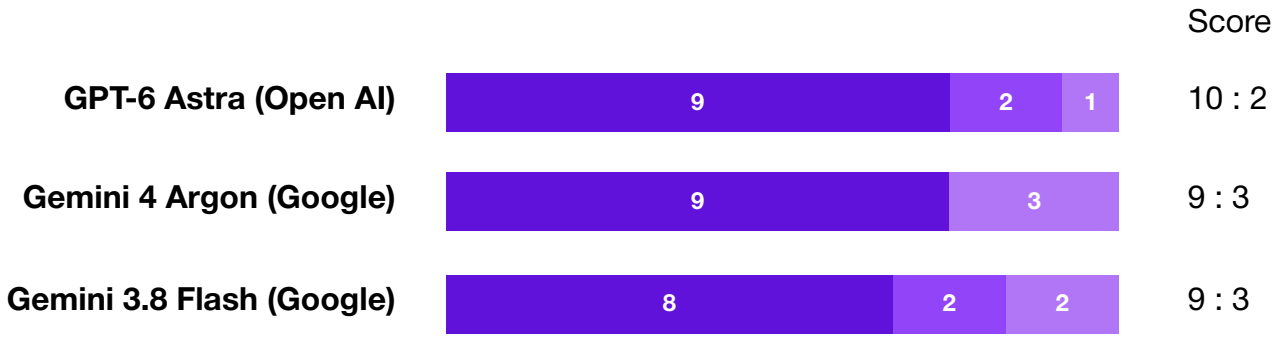


Results

Mark-1 scored 28 of 36 points across the three matches.
Every decisive game ended in checkmate.

Mark-1 won all three matches: 28 of 36 points

26 wins, 4 draws, 6 losses in 36 games, 7 Nov 2026



FTA match reports, 7 Nov 2026 · 36 official games · score counts a win as 1 and a draw as 1/2

Game by game

Each opening was played twice per match, once with each colour. Results are from Mark1's side.

Game	Opening	Mark-1 plays	vs GPT-6 Astra	vs Gemini 4 Argon	vs Gemini 3.8 Flash
1	1	White	Draw (repetition)	Win	Win
2	1	Black	Win	Loss	Loss
3	2	White	Win	Win	Loss
4	2	Black	Win	Win	Win
5	3	White	Draw (not enough pieces to mate)	Loss	Loss
6	3	Black	Win	Win	Win
7	4	White	Win	Win	Draw (repetition)
8	4	Black	Loss	Win	Win
9	5	White	Win	Win	Win
10	5	Black	Win	Win	Draw (repetition)
11	6	White	Win	Win	Win
12	6	Black	Win	Win	Win

Thinking time

Mark-1 took about a minute a move in every match. It was a little faster than GPT-6 Astra and about five times slower than the Gemini models.

Match	Mark-1 seconds a move	Opponent seconds a move	Moves (Mark-1 / opponent)
vs GPT-6 Astra	61.8	69.4	539 / 534
vs Gemini 4 Argon	67.5	13.0	557 / 554
vs Gemini 3.8 Flash	62.1	14.1	605 / 601

Rating against Stockfish

Before the matches, Mark-1 played a fixed ladder of 24 games against Stockfish set to four strength levels. This puts Mark-1 at about **1,824** (95% range 1,635 to 2,012).

Stockfish level	Mark-1 points	Games
1400	6	6
1700	3	6
2000	2.5	6
2300	0	6

For context, GPT-6 Astra rated about 1,662 on the same ladder, from 23 of 32 planned games.

Safe moves

On a fixed test of 300 positions, Mark-1 avoided big mistakes in 291 (97.0%), against 283 (94.3%) for GPT-6 Astra. A big mistake is a move that loses a large share of the winning chances, as judged by Stockfish 16. Mark-1 also played a legal move in 100% of the 2,000 positions in its main test.



Play and test Mark-1 yourself

The complete Mark-1 system is on Hugging Face at **ftaglobal/FTA-Mark-1:** the judge, the think-ahead, the match runner and the referee used in the official matches. There's no sign-up and no form to fill.

1. Download FTA Mark-1 from Hugging Face.

2. Load it on your GPU. Sarvam-30B downloads automatically.

3. Choose who Mark-1 plays:

A person, on a chess board that opens in your browser. Click to move. No API key needed, so a chess academy can let its students play it in class.

GPT-6 Astra, Gemini or anotherAI, by adding your own API key. The same rules as our matches apply.

Any chess app that accepts engines (UCI), such as Arena or Cute Chess.

4. Play.

Stockfish level	What it means
Match level	Exactly the setting used in the official matches. About 62 to 68 seconds a move on an H200, about 114 seconds on an A100-80GB
Practice and Quick levels	A smaller think-ahead, for practice. Practice takes about 31 seconds a move and Quick about 5 seconds, on an A100-80GB. Both are weaker than Match level. Practice rating: about 1,585 (95% range 1,368 to 1,807), from 12 games against Stockfish 16 (3.5 of 6 at 1400, 3 of 6 at 1700), on our own ladder. Quick is not yet rated
GPU	An 80 GB GPU (H200 orA100-80GB). Sarvam-30B alone needs about 64 GB. A rented cloud GPU works

Anyone can check our results this way. Load Mark-1 at Match level, give it positions from our published replays, and compare its moves with the moves it played. Small differences in GPU hardware can flip a move where two choices score almost the same, so a close match is expected, not a perfect one.

We ran this check ourselves on a fresh rented A100-80GB, using only the public files. 22 of 30 positions from the recorded GPT-6 Astra games gave the same move at Match level. In all 12 positions where one move was clearly best (by more than 1 win% point), the move was the same. Each of the 8 differences was a close call, where the recorded move led by 0.69 win% points or less. The same fresh machine also played a full game against Gemini 3.8 Flash (a draw, with no referee step-ins), ran the browser board and connected to a chess app.

Proof and records

All match records were checked, sealed and timestamped by an outside service the day after the matches, so nobody (including FTA) can change them without it showing.

- **Sealed record (proof pack) fingerprint (sha256):** ca05b74316458cc3b31069c6d309064f0b14425c82b536a8bea5a9e2ea93569e
- **Timestamp:** made with OpenTimestamps on 8 November 2026 and anchored in the Bitcoin blockchain in block 975064 (8 November 2026, 13:02 UTC), and also in blocks 975066, 975074 and 975088. The sealed record lists internal machine paths, so the record and its timestamp file are shared on request rather than published.
- **What it covers:** every game file of the three matches, the Stockfish ladder and the systems test; the locked match settings; the machine logs; the provider billing records; and the Mark-1 adapter fingerprint.

The providers' own records match the games. OpenAI's usage shows the calls for all 534 GPT-6 Astra answers. Google's logs show 1,156 calls on match day, against 554 Argon and 601 Flash answers plus a short connection check. Every call succeeded.
All 36 match games are published on Hugging Face (in the games folder of ftaglobal/FTAMark-1) and as videos on YouTube. The published game files are the originals with two values removed: an internal file path and each game's API cost. The original game files, logs and billing records are available to journalists, researchers and event partners on request at **mark.admin@ftaglobal.com.** We show them on a live screen share. Anyone shown a record can check it against the fingerprint above on the spot.

Limits

These results show what happened in 36 games under one fixed set of rules. They are not a ranking ofAI models.

- **Our setup, not Kaggle's.** Kaggle did not run these matches, and our harness is not Kaggle's. The Kaggle leaderboard is context for who the opponents were.
- **Small samples.** Each match was 12 games. A different set of openings could give a different score.
- **Default settings only.** The opponents played at their makers' default settings. We did not test them at higher reasoning settings.



- **Careful, not brilliant.** On the 300-position test, GPT-6 Astra picked Stockfish's top move more often (48.7% against Mark-1's 44.0%). Mark-1's edge is that it makes far fewer big mistakes.
- **Slower and heavier.** Mark-1 takes about a minute a move and needs a large GPU. The Gemini models answered in 13 to 14 seconds.
- **Our own rating.** The 1,824 rating comes from 24 games against Stockfish on our ladder. It is not a FIDE, Lichess or Chess.com rating, and its 95% range is wide.
- **Chess only.** These results say nothing about Mark-1 outside chess.

Credits and licence

FTA Mark-1 is built on Sarvam-30B, trained with FTA's MJT.

- **Sarvam AI,** for Sarvam-30B, released underApache 2.0. Mark-1 is FTA's work on top of it. Sarvam AI did not make or endorse it.
- **Google DeepMind,** whose research on scoring every chess move (ChessBench)shaped how Mark-1 judges positions.
- **Stockfish,** the open chess engine used as the teacher in training and as the opponent on our rating ladder. It is never used during Mark-1's play.
- **Lichess,** for its open game database.
- **Kaggle Game Arena** (Kaggle and Google DeepMind), whose chess leaderboard we used to choose the opponents.
- **OpenTimestamps,** for the independent timestamp on our records.

OpenAI, Google and Kaggle did not take part in, run or endorse these matches.

Licence. The Mark-1 release is fully open. The adapter weights are underApache 2.0. The think-ahead, match runner, referee, play board and notebooks are under GPL-3.0, because they use the python-chess library. The published game records are under CC BY 4.0. It all comes with no warranty.

Contact. mark.admin@ftaglobal.com