

CodeBrain-1 是什么？（Deep Research）

生成日期：2026-02-11

Executive Summary

“CodeBrain-1”目前最明确、可核实的指代，是 Terminal-Bench 2.0 排行榜上的一个“Agent/参赛系统名字”，而不是一个单独发布的基础大模型名称。它由 Feeling AI 提交到 Terminal-Bench；在一次提交记录中显示使用 `GPT-5.3-Codex` 作为底座模型并取得较高分数。

Key Findings

- **它是什么**：Terminal-Bench 2.0 Leaderboard 上的 **Agent 名称（CodeBrain-1）**，归属组织显示为 **Feeling AI**；在不同日期的提交里可搭配不同底座模型。
- **它不等于“某个公开大模型”**：排行榜条目展示的是“Agent 系统 + 底座模型 + 评测结果”的组合，“CodeBrain-1”更像系统/方案名。
- **Terminal-Bench 是什么**：让 AI agent 在“沙箱终端环境”里完成真实任务（编译、训练、搭服务等）的评测框架与任务集；任务通常配 docker 环境与测试脚本验收。
- **容易混淆的同名/近名**：互联网上还有别的“CodeBrain/CodeBrain AI Employee/CodeBrain(公司内部代码大模型)/CodeBrain.com(老牌 Java applets 网站)”等，和“CodeBrain-1(排行榜 agent)”不是一回事。

Detailed Analysis

1) CodeBrain-1（最可信的“本体”）

在 Terminal-Bench 的官方 leaderboard 中，“CodeBrain-1”作为参赛条目出现：可看到它的组织归属、提交日期、分数/通过率以及底座模型名称。

这类榜单通常评的是“能否把一堆终端任务做完”，因此“Agent 名称”往往代表一套工程化系统（提示词/规划器/工具编排/重试策略/上下文管理等），底座模型只是其中一部分。

2) Terminal-Bench 2.0 大概怎么评

Terminal-Bench 文档将其拆成任务集与执行 harness：harness 会把模型接到一个沙箱终端里跑任务、记录日志，并用测试脚本判定成功与否；运行依赖 Docker、uv，并提供 `tb run` 等入口。

2.0 的公告表明它在持续迭代（所以榜单名次/提交也会随时间变动）。

3) 为什么你会觉得“这是什么毛东西”：常见误解点

- 把 Agent 名当成“新模型名”：榜单里“CodeBrain-1”更像“选手/系统名”，不是“像 GPT/Claude 那样的模型品牌”。
- 同名太多：例如 yourcodebrain.com 的 CodeBrain 是“AI 员工/工作流产品”；还有其它公司/站点/内部项目也可能叫 CodeBrain。

Areas of Consensus

- “CodeBrain-1”在公开可核验层面，确实以 **Terminal-Bench 榜单条目 (Agent)** 的形式出现。
- Terminal-Bench 是一个面向终端复杂任务的 agent 评测框架/任务集，而非单一模型发布页。

Areas of Debate / Uncertainty

- **CodeBrain-1 的技术细节是否公开**：从榜单与文档侧通常只能确认“它是什么评测条目/用了什么底座模型/成绩如何”，但未必能找到对应论文或开源仓库来说明其内部架构（规划、工具、记忆等）。

Sources

- Terminal-Bench Docs: tbench.ai/docs
- Terminal-Bench Harness: tbench.ai/docs/harness
- Terminal-Bench 2.0 Announcement: tbench.ai/news/announcement-2-0
- Terminal-Bench GitHub: github.com/laude-institute/terminal-bench
- yourcodebrain.com: yourcodebrain.com
- codebrain.com: codebrain.com

Gaps and Further Research

- 如果能提供你看到“CodeBrain-1”的具体来源（链接/截图/视频标题），可以进一步锁定你指的是“Terminal-Bench 这个”，还是另一个同名项目，并继续深挖其团队、公开资料、是否开源、复现方

式。