GPT-5.3-Codex-Spark vs GPT-5.3-Codex □□

GPT-5.3-Codex-Spark □ GPT-5.3-Codex □□□□□□□□□□□□□□□□

- GPT-5.3-Codex-Spark vs GPT-5.3-Codex □□

GPT-5.3-Codex-Spark vs GPT-5.3-Codex 对比

> 数据来源: LINUX DO 社区帖子逐帖实读、OpenAI 官方信息，2026-06-11 整理。

一、模型定位

维度	GPT-5.3-Codex-Spark	GPT-5.3-Codex
定位	轻量级极速编程模型（本质是 mini 模型）	标准代码专用模型，平衡性能与质量
驱动硬件	Cerebras 定制高性能芯片（晶圆级处理器）	常规 GPU 集群
输出速度	~1000 tokens/秒（官方宣传，Cerebras 芯片驱动）	无官方数据，体感比 Spark 慢很多
上下文窗口	128k（严重限制，复杂任务不够用）	正常（同其他 Codex 模型，远大于 128k）
编程能力	弱 ——本质是 mini 模型，「和 CODEX 根本不是一个玩意儿」	强 ——SWE-Bench Pro 57%、TerminalBench 2.0 76%
可用范围	Pro 订阅独有 + Poe 免费	Pro / Plus / Team 均可用

二、详细对比

2.1 速度

GPT-5.3-Codex-Spark 由 Cerebras 定制芯片驱动，官方宣称输出速度可达 ~1000 tokens/秒，是 OpenAI 产品线中输出最快的模型。但「输出快」≠「任务完成快」——由于模型能力弱、上下文小，复杂任务反而可能无法完成。

GPT-5.3-Codex 的输出速度无官方数据。在实际编程任务中，根据推理强度（reasoning effort）不同，任务完成时间有明显差异:

推理强度	任务完成时间	备注
medium	~10 分钟	速度与能力的平衡点
high	~16 分钟	可能触发上下文压缩
xhigh	~16 分钟	深度推理，与 high 耗时接近

> 来源: 1646104 @SmallMain 1楼（2026-02-25）Unity C# 皮肤系统需求横评实测。注：这是特定任务的端到端完成时间（含思考＋读文件＋写代码），非 tokens/秒 生成速度。不同任务耗时差异很大，此数据仅作参考。

2.2 上下文窗口与实际影响

GPT-5.3-Codex-Spark 的上下文窗口为 128k。在实际复杂编程任务中，128k 远远不够用：

- 模型会反复触发上下文压缩
- 压缩后重复读取已读过的代码文件，形成无效循环
- 最终耗尽所有额度仍无法完成任务

> 来源: 1646104 @SmallMain 1楼：「GPT-5.3-Codex-Spark 由于上下文仅有 128K，对于这种复杂程度的需求来说非常不够用，所以进行了多次上下文压缩，在上下文压缩过后，模型会重复读取上一轮读过的代码文件，导致无限重复直到耗尽了所有的限额，最终无法完成需求」

GPT-5.3-Codex 拥有正常的上下文窗口，适合完整项目开发。

2.3 编程能力

GPT-5.3-Codex-Spark 本质上是一个 **mini** 模型，社区多位用户明确指出它和标准 Codex 不是同一级别：

- 「别用 gpt-5.3-Codex-Spark，这是个 mini 模型」
- 「上下文关系不大，就是模型太小了，和 CODEX 根本不是一个玩意儿。能力不行」
- 「应该叫 5.3codex-mini-spark」

> 来源: 1646104 16楼 @bifangKNT、18楼 @wangqazwsx、17楼 @wangqazwsx（2026-02-25）

GPT-5.3-Codex 是正经的代码专用模型，官方评测成绩：

- SWE-Bench Pro：57%
- TerminalBench 2.0：76%
- OSWorld：64%

> “来源：1570870（2026-02-06）OpenAI 官方发布数据

2.4 可用范围

平台	GPT-5.3-Codex-Spark	GPT-5.3-Codex
ChatGPT Web	Pro 订阅	Pro / Plus / Team
Codex CLI/Desktop	Pro 订阅	Pro / Plus / Team
Poe	□ □□□□	□ □□□

2.5 额度消耗

Spark 有独立额度，与标准 Codex 分开计算。有用户反馈 Spark 额度消耗异常（使用后百分比不变），疑似存在无限额度 bug。

> “来源：1666953（2026-03-01）「Codex多了一个 5.3-Codex-Spark的单独额度？」；1813954（2026-03-25）「codex 5.3 spark限额真的好奇怪」

三、选择建议

场景	推荐模型	理由
简单代码补全	Spark	1000 tok/s 极速，简单任务够用
Poe 免费尝鲜	Spark	无需订阅即可体验
日常编程开发	标准 Codex	能力远强于 Spark，medium 即可
复杂项目/重构	标准 Codex(high/xhigh)	Spark 会因上下文不足而失败
省额度	标准 Codex(medium)	速度与能力的最佳平衡

来源： 1572018（2026-02-06）「 gpt-5.3-codex
进步巨大速度非常快，能力非常强」； 1573863（2026-02-06）「小任务 5分钟内 大的
15分钟内」

主要数据来源

1646104 — 多模型横评实测（ 2026-02-25）@SmallMain，含 Spark vs Codex 速度 /能力对比 1570870 — GPT-5.3-Codex 正式发布（ 2026-02-06），官方评测数据
1945856 — Spark 模型参数确认： Cerebras、~1000 tok/s、 128k
1572018 — 5.2 vs 5.3-Codex 选择建议（ 2026-02-06）
1573863 — 5.3-Codex xhigh 实际使用体感（ 2026-02-06）
1666953 — Spark 独立额度发现（ 2026-03-01）
1813954 — Spark 额度异常报告（ 2026-03-25）