Kimi K3.1 Pro Technológiai Audit



Jelentés típusa	Független LLM Teljesítményvizsgálat
Tesztelt modell	Moonshot Kimi K3.1 Pro
Módszertan	KingBench 4 Keretrendszer

Gemini Notebook

Az Elvárások és a KingBench 4 Valósága

Marketing Ígéretek (A Papírforma)

- **Claude Opus** szintű képességek
- Drasztikusan **alacsony** hallucináció

5.37/10

Végső Pontszám

Az Audit Valósága (Teszt Eredmények)

- **6. hely a 11-ből** (Középmezőny)
- A vártnál jóval **drágább API futtatás** (Rejtett költségek)
- Kiváló helyi futtatási **potenciál nyílt** súlyozással

Gemini Notebook

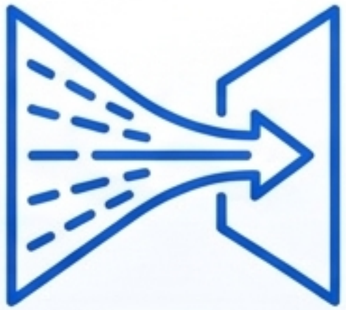
A kiindulási alap: Egy új nyílt forráskódú csúcsmodell?



65.7%

Terminal Bench 4.0

A Moonshot állítása szerint közvetlenül a Claude Opus 5.5 (66.4%) mögött és a GPT-6 Astra (58.2%) előtt áll.



1 Millió

Tokenes Kontextusablak

Új High-Speed verzió, amely a hivatalos tesztek szerint 2-3x gyorsabb végrehajtásra képes.



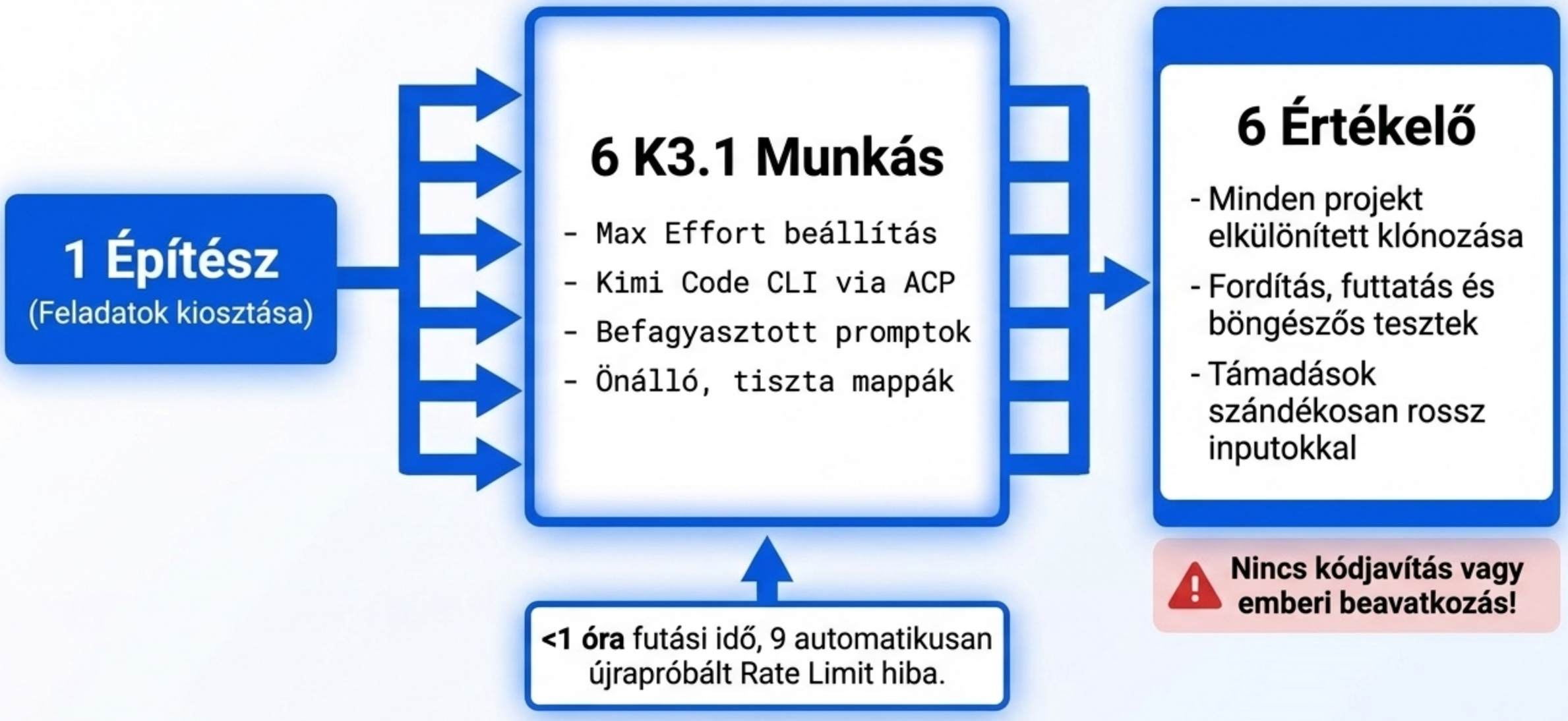
8.3%

Hallucinációs Ráta

A legnehezebb tényszerű teszteken az előző K3 generáció 23.4%-os hibaarányáról radikálisan csökkent.

Gemini Notebook

Tesztelési környezet: A Panda Code Architektúra

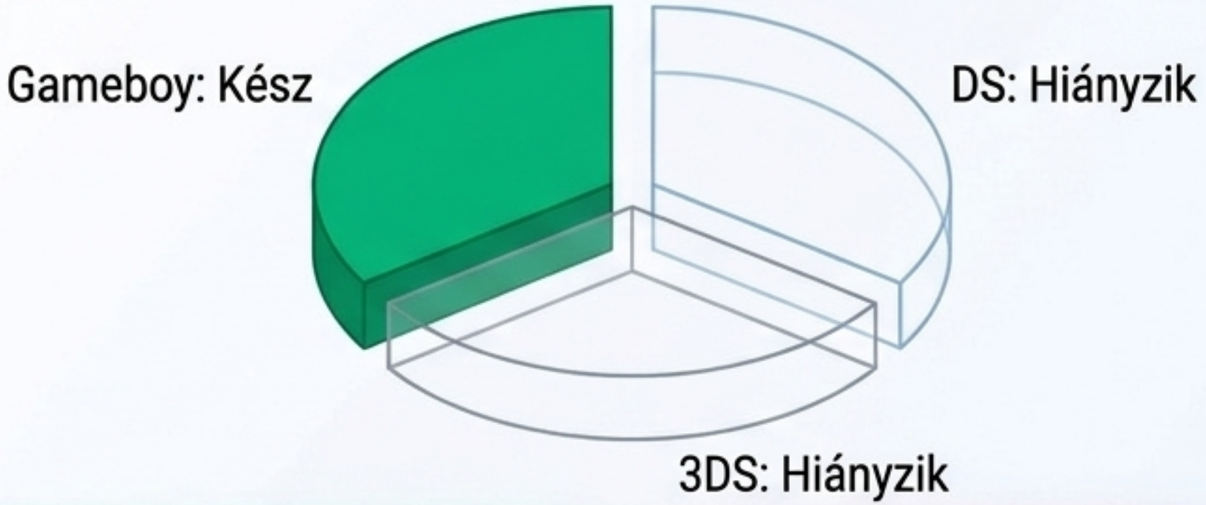


Gemini Notebook

KingBench Feladat 1: Emulátor (Rust)



Részleges Teljesítés



✅ Sikerek

- **Őszinte a hiányosságokról:** Visszautasítja a DS/3DS fájlokat, nem próbál hamisítani.
- **Szolid Gameboy mag**: 11 CPU utasítástesztt sikeres.
- Pixelpontos renderelés és egzakt mentési állapotok (40x sebesség).

⚠ Hibák

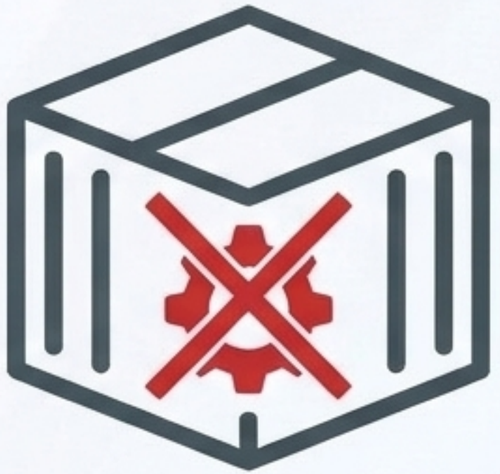
- **Kritikus hiba (Interrupt Pattern):** Összeomlik a verem a Tobu Tobu Girl logóképernyőjénél.
- **Mentés** csak szabályos bezáráskor történik (Összeomlás = teljes adatvesztés).
- A kért rendszer 66%-át meg sem kísérelte leprogramozni.

Gemini Notebook

KingBench Feladat 2: Docker Alternatíva (Go)







✅ Sikerek

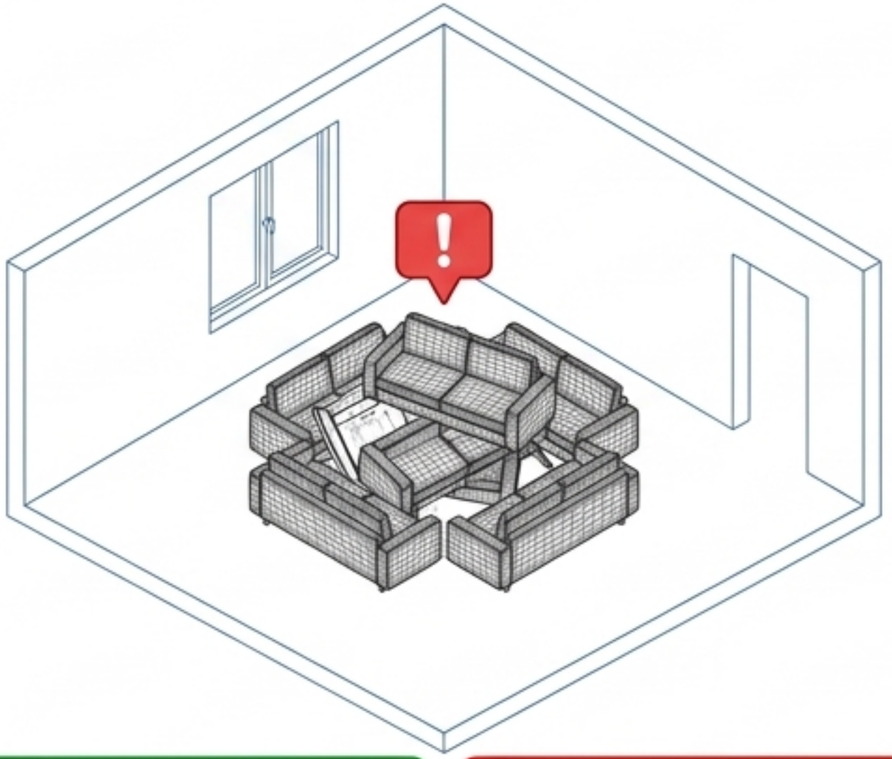
- **Kiváló képfájl letöltő** ('krun' eszköz) - Alpine, Nginx, Redis hiba nélkül letöltve.
- **100%-os strukturális egyezés** a referencia extraktorral.
- **Sikeresen kivédett 20 célzott kibertámadást** (Path traversal, symlink trükkök, hard link szökések).

⚠️ Hibák

- **Nem egy valódi Docker alternatíva**, csak egy specifikáció a Readme-ben.
- A **régebbi Schema 2 képek** letöltődnek, de konténer létrehozásakor hibát dobnak.
- **Nincs** 'build', 'exec', port publikálás vagy hálózatkezelés.
- **Mac rendszeren futtathatatlan** (Csak forráskód szintű kreditet kapott).

Gemini Notebook

KingBench Feladat 3: 3D Szobatervező



7.4/10

Sikerek

- **12 átméretezhető,** teljesen textúrázható bútortípus.
- **Precíz fizikai motor:** Az elforgatott ágy pontosan, ütközésig megáll a falnál.
- **Tökéletes memóriakezelés:** >100 szerkesztés után sincs GPU memóriaszivárgás.

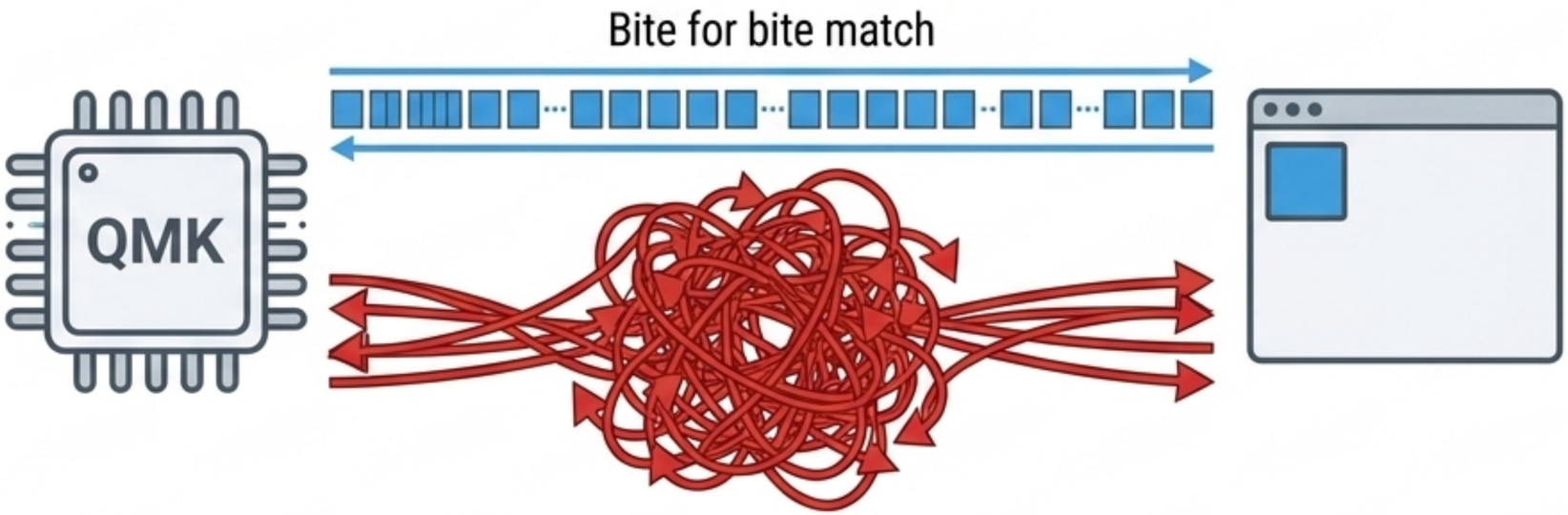
Hibák

- **Új bútorok spawn pontja** fixen a szoba közepe (Egymásba lógó halmot alkotnak).
- **Túl steril látvány:** Mérnöki modell, nem egy fotorealisztikus szoba.
- **Egyetlen globális fényforrás:** A nap fixen a hátsó fal mögött, nincsenek vetett árnyékok.
- **Kontextus:** A DeepSeek Flash 8.8-at, az Opus 8.5-öt ért el ugyanezen a feladaton.

Gemini Notebook

KingBench Feladat 4: Billentyűzet App (VIA Protokoll)





Sikerek

- **Tökéletesen beszéli** a komplex VIA protokollt (Virtuális QMK-val tesztelve).
- **Minden** adatcsomag mérete és belső parancs ID-ja helyes.
- **A makró bájtok** bitre pontosan egyeznek a QMK referencia kódolásával.

Hibák

- **Brutális sávszélesség-pazarlás**: Az olvasás több száz felesleges oda-vissza utat igényel tucatnyi helyett.
- **Egy 11 gombos** definíciót 14 gombosként renderel egy enkóder hiba miatt.
- **Hiányos kódlista**: Nincs 'layer tap' vagy 'mod tap'.
- **Egyetlen nem angol karakter** beírása összeomlasztja és kifehéríti a teljes felületet.

Gemini Notebook

KingBench Feladat 5: Pokémon Red Klón





Sikerek

- **Végigjátszható**, lineáris út a kezdőháztól az első Boss-ig (Összeomlás nélkül).
- A **fizikai korlátok** működnek (A magas fű megállítja a játékost).
- **Fogás**, **harc** és **jelvényszerzés** integrálva, a mentés túléli a játék újraindítását.

Hibák

- **Mentéstörlő dizájnhiba**: A kurzor fixen a 'New Game' gombon áll a címképernyőn.
- **Törött szinteződés**: A karakter Level 5-ből Level 34-re ugrik az első Gym előtt.
- **Totális egyhangúság**: Minden épület egyforma, az erdő pedig egy üres rét.
- **Webes pop-up menüket** használ natív játékképernyők helyett

Gemini Notebook

KingBench Feladat 6: Fitbit Desktop App



5.9/10





Sikerek

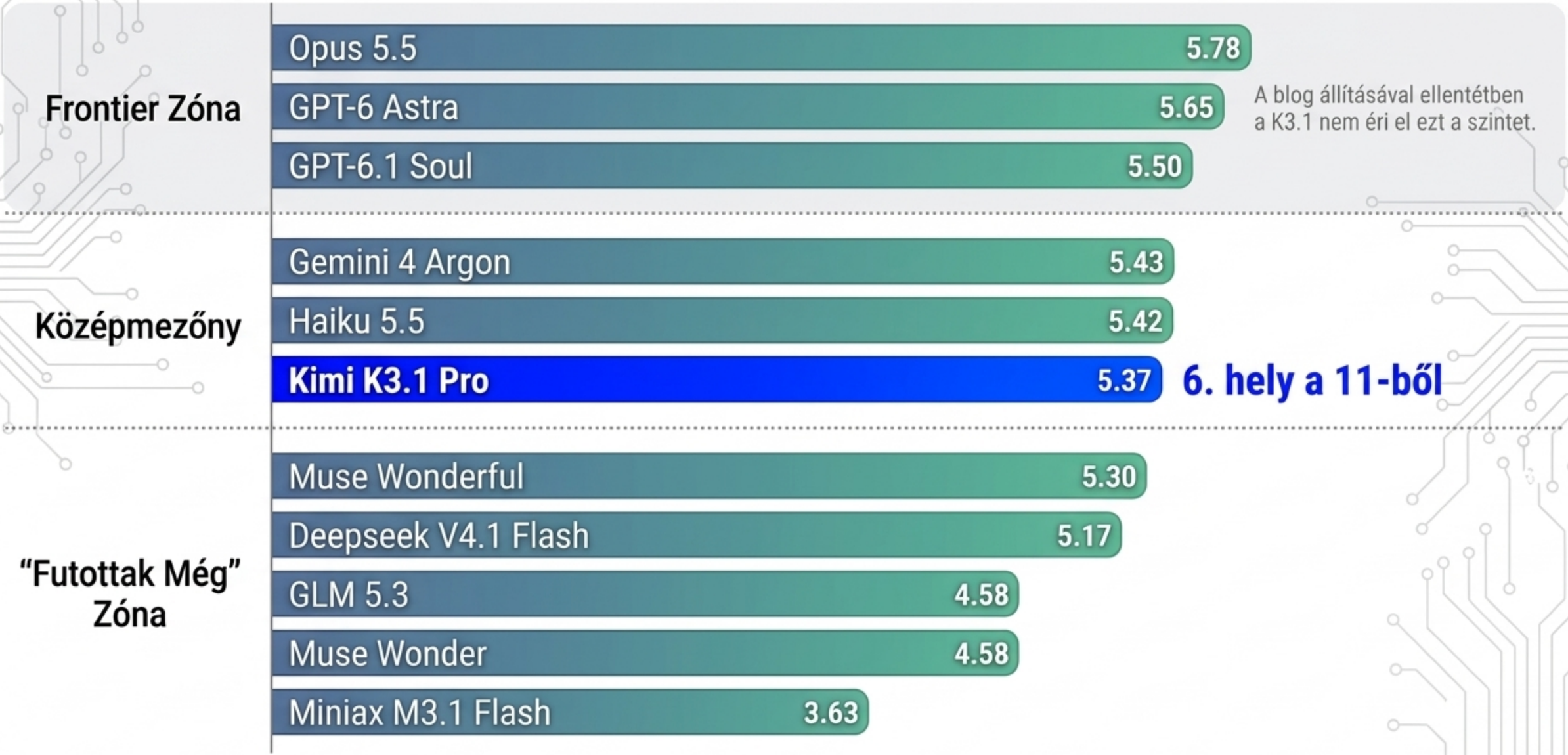
- Natív Tauri alkalmazás villámgyors Rust backenddel. Késznek tűnő Dashboard.
- Rendkívül gyors: 10 évnyi napi adatot 3 milliszekundum alatt dolgoz fel.
- Prémium metrikák integrálva (Alvás fázisok, HRV trendek, pulzuszónák).

Hibák

- Egy olyan régi Fitbit Web API-t használ, amit a Google éppen lekapcsol (Health API helyett).
- Nincs bejelentkezési folyamat (Csak bemásolt hozzáférési tokennel működik).
- Kizárólag a saját fájlformátumát tudja importálni.
- Adatvesztés: A részleges importálás csendben törli a korábbi napok adatait.

Gemini Notebook

KingBench 4 Ranglista Mátrix: Hol tart a K3.1 Pro?



Gemini Notebook

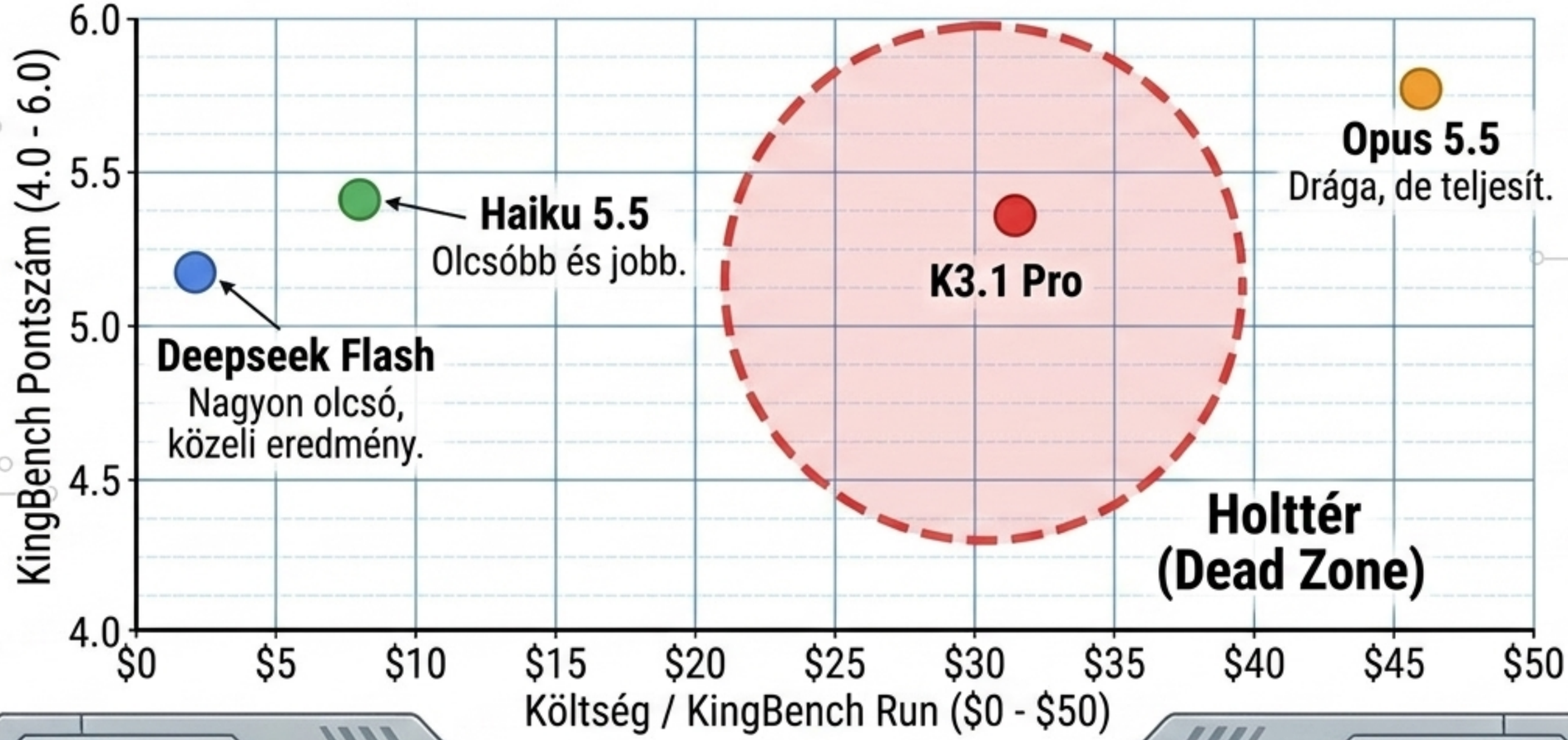
Pénzügyi Audit: A 'Chatty Agent' Probléma



Teljes futási költség (Normal API): $32.00. (High-Speed API: $42.00).
A paradoxon: Bár a generált kód (output) olcsó, a **modell saját belső kontextusának folyamatos, kényszeres újraolvasása felemészti a büdzsét.**

Gemini Notebook

Költség-Teljesítmény Mátrix (API Használat)



Következtetés: A K3.1 Pro több mint **4x drágább**, mint a **Haiku** (gyengébb eredménnyel), és 16x drágább, mint a Deepseek Flash. **API-n** keresztül jelenleg **nem racionális választás.**

Gemini Notebook

Az igazi értékajánlat: Nyílt Súlyozás (Open Weights)



Nagyvállalati Szerver Klaszter



Hétköznapi Laptop

A Licenc

MIT alapú, csupán minimális korlátozásokkal a nagyvállalatok számára. Szabad utat enged a fejlesztésnek.

A Hardverigény

Ez egy **3 Billió** (Trillion) paraméteres osztályú modell. Helyi futtatásához **szerver rack / GPU klaszter** szükséges, nem egy laptop.

Adatbiztonság

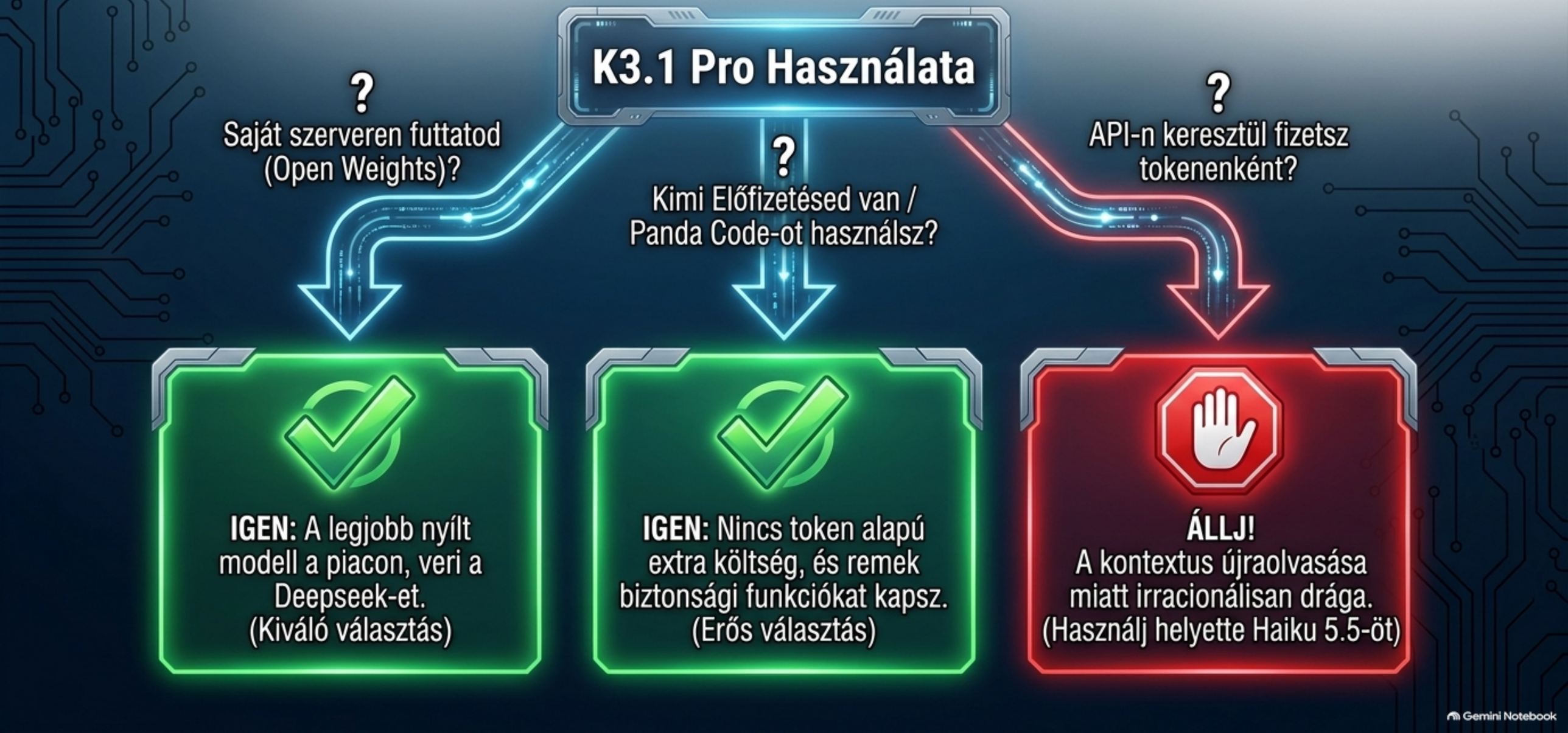
A nyílt verzióból biztonsági okokból eltávolították a kiberbiztonsági tréningadatokat (Nem lesz 100%-ban azonos az API verzióval).

A Jövőkép

A súlyozás november 4-i megjelenésével a harmadik feles hosztolók miatt az árak összeomolhatnak, radikálisan növelve a modell értékét.

Gemini Notebook

Döntési Fa: Mikor használjuk a K3.1 Pro-t?



Gemini Notebook

Végső Verdikt

Mi az, amit a Moonshotnak javítania kell? (Hibajegyek)

A ‘Chatty Agent’ optimalizálása:
A belső kontextus folyamatos és felesleges újraolvasásának (cache reread) minimalizálása a költségek letörése érdekében.

Részleges teljesítések megszüntetése:
Ne hagyjon félbe 3-ból 2 rendszert, és ne szállítson Docker alternatívát 'build’ parancs nélkül. Az őszinteség jó, de a befejezett kód jobb.

A Valós Pozicionálás

Nem egy ‘Frontier’ (Csúcs) Modell:
A marketing állításokkal ellentétben valós, komplex projektekben nem éri el az Opus vagy az Astra szintjét, és a Haiku mögött marad.

Kiváló ‘Open’ Modell:
Helyi hosztolás esetén, a súlyozás kiadása után az egyik legerősebb és legtisztább nyílt forráskódú alternatíva a teljes piacon. Értékeljük annak, ami.

Gemini Notebook