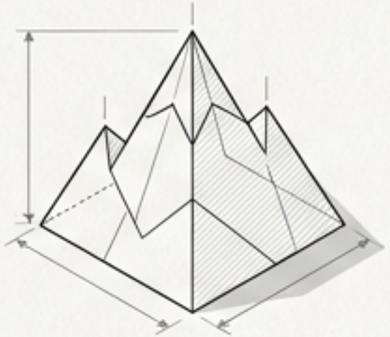


Heti AI Összefoglaló: A Nyílt Forráskód Forradalma

12 új modell, ami átírja a játékszabályokat

Gemini Notebook

Az elmúlt hét mérlege: 4 fő irányzat



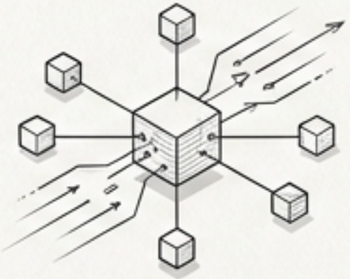
Határfeszegető nyílt modellek

A 3 billió paraméteres zárt modellekkel versengő nyílt súlyok megjelenése (pl. Kimi K3.1 Pro).



Számítógép-vezérlő ágensek

A képernyőt olvasó és az operációs rendszert önállóan kezelő 'Observe -> Act' modellek korszaka (pl. North Ka).



Új architektúrák sebessége

Az autoregresszív határok áttörése Liquid és Diffusion szöveges struktúrákkal.



Költséghatékony csúcsmodellek

A 'Frontier' (csúcskategóriás) intelligencia árának drasztikus zuhanása a zárt és nyílt piacon egyaránt.

Gemini Notebook

Kimi K3.1 Pro: A Nyílt Kódú Óriás

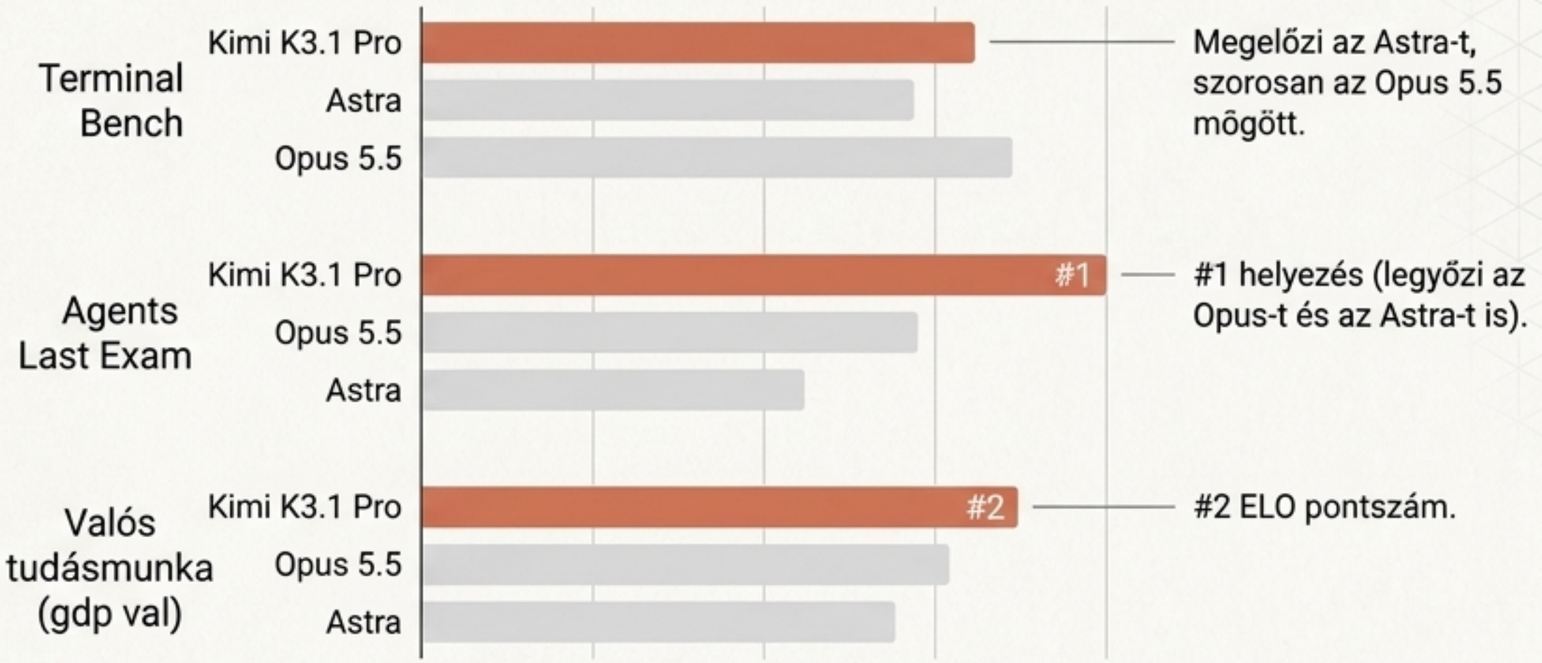
Modell Specifikáció

Státusz: Nyílt súlyok (HuggingFace)

Méret: 3 Trillió paraméter (96 részlet, egyenként ~17GB)

Hardverigény: 1.2 - 2 TB VRAM

Teljesítmény Összehasonlítás



Bár a 10 trillió paraméteres zárt modellek (Astra, Fable) még egy fél lépéssel előrébb járnak, a **Kimi 3.1 Pro** a valaha kiadott legkiválóbb nyílt alternatíva a komplex feladatokhoz.

Gemini Notebook

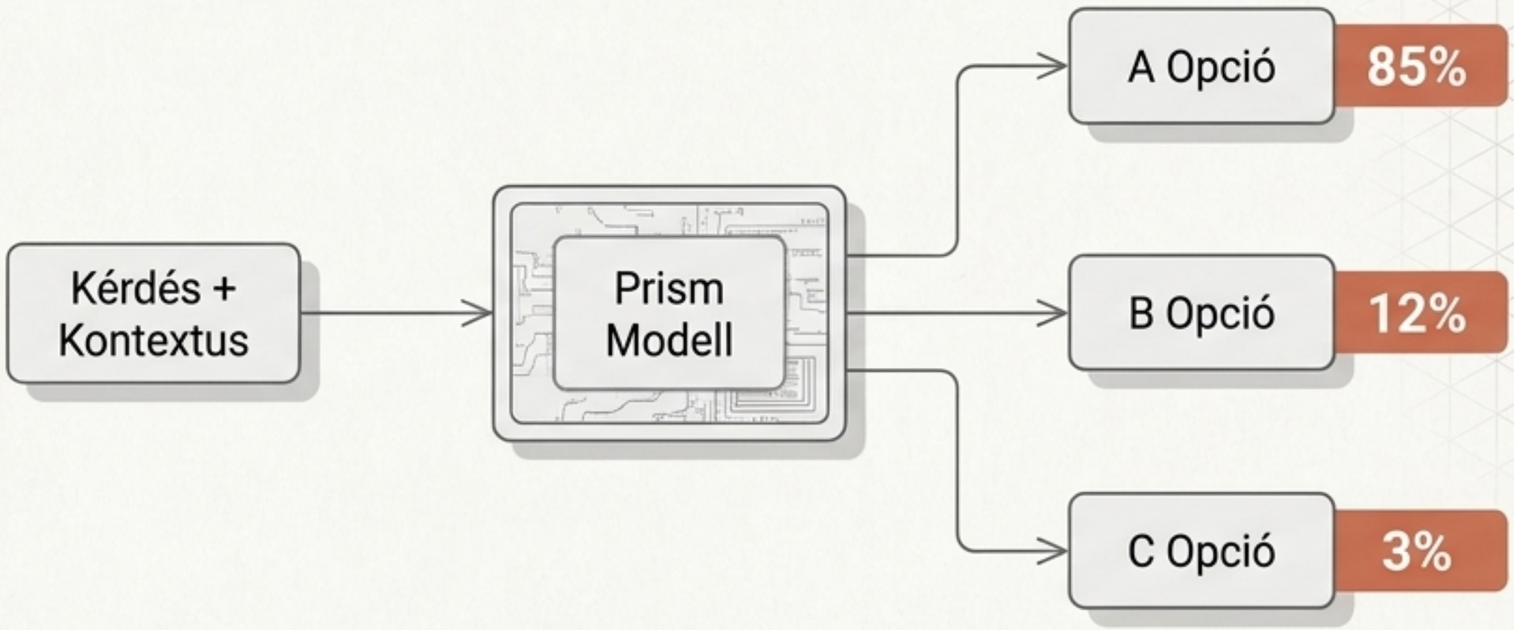
Sarvam Prism: A Tisztán Döntéshozó Modell

Modell Specifikáció

Státusz: Nyílt forráskód

Architektúra: Sarvam 105B alapokon

Hardverigény: 106 GB tárhely, B200 / H200 GPU ajánlott



Hogyan működik? Nem szöveget generál. Egyetlen futtatással (single forward pass) kiszámítja az előre megadott megadott válaszlehetőségek valószínűségét.
Felhasználás: Általános döntéshozatal, hasonlóan a Jev modellhez, hatalmas sebességgel és pontossággal. (ontyan működiko).

Gemini Notebook

North Ka: Az Operációs Rendszer Irányítója

Modell Specifikáció

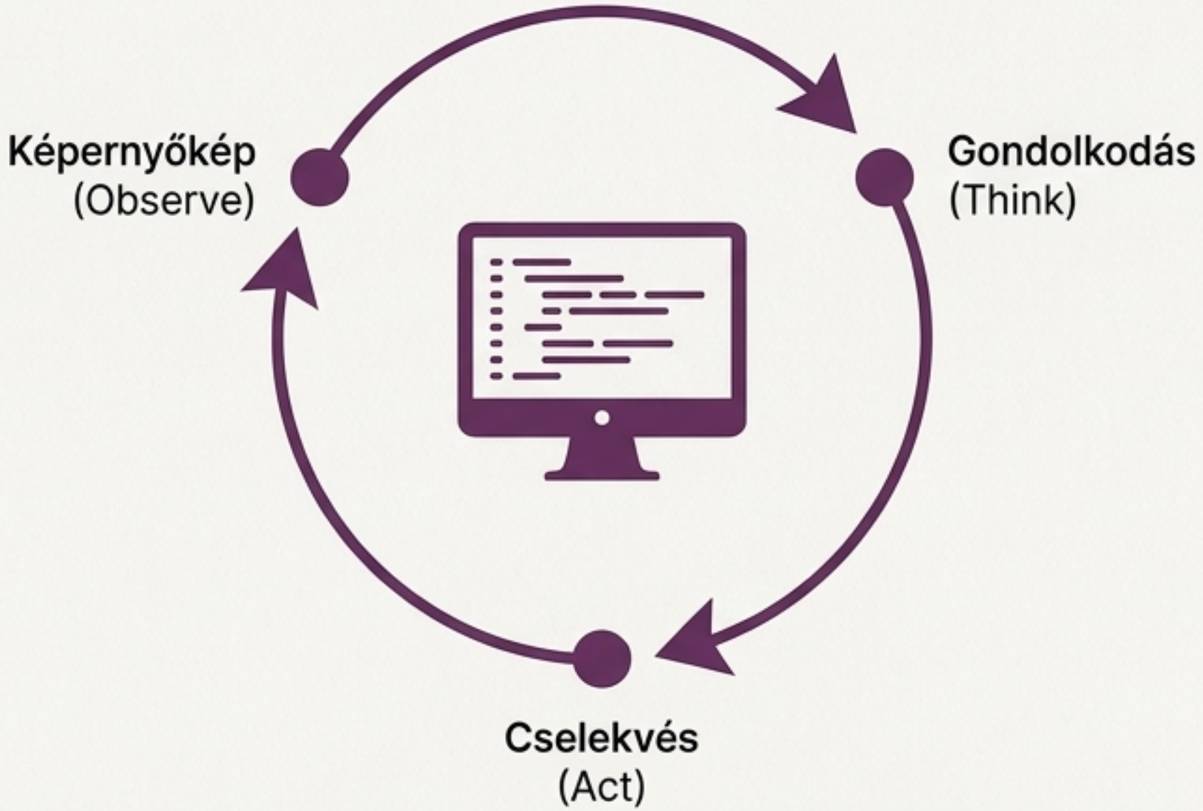
Típus: Mixture of Experts (MoE)

Pro Méret: 397B összes / 17B aktív paraméter

Képesség: Frontier szintű számítógép-vezérlés

OSWorld Verified pontszámok	
Pro (397B):	88.0 pont
Mini (35B):	85.0 pont
Nano (9B):	78.9 pont

(Lokálisan, akár telefonon is futtatható, alig 10 ponttal lemaradva a Pro verziótól!)



A modell egy nyílt forráskóódú keretrendszerben (harness) folyamatosan elemzi a képernyőt, és önállóan hajt végre egér- és billentyűzet-műveleteket.

Gemini Notebook

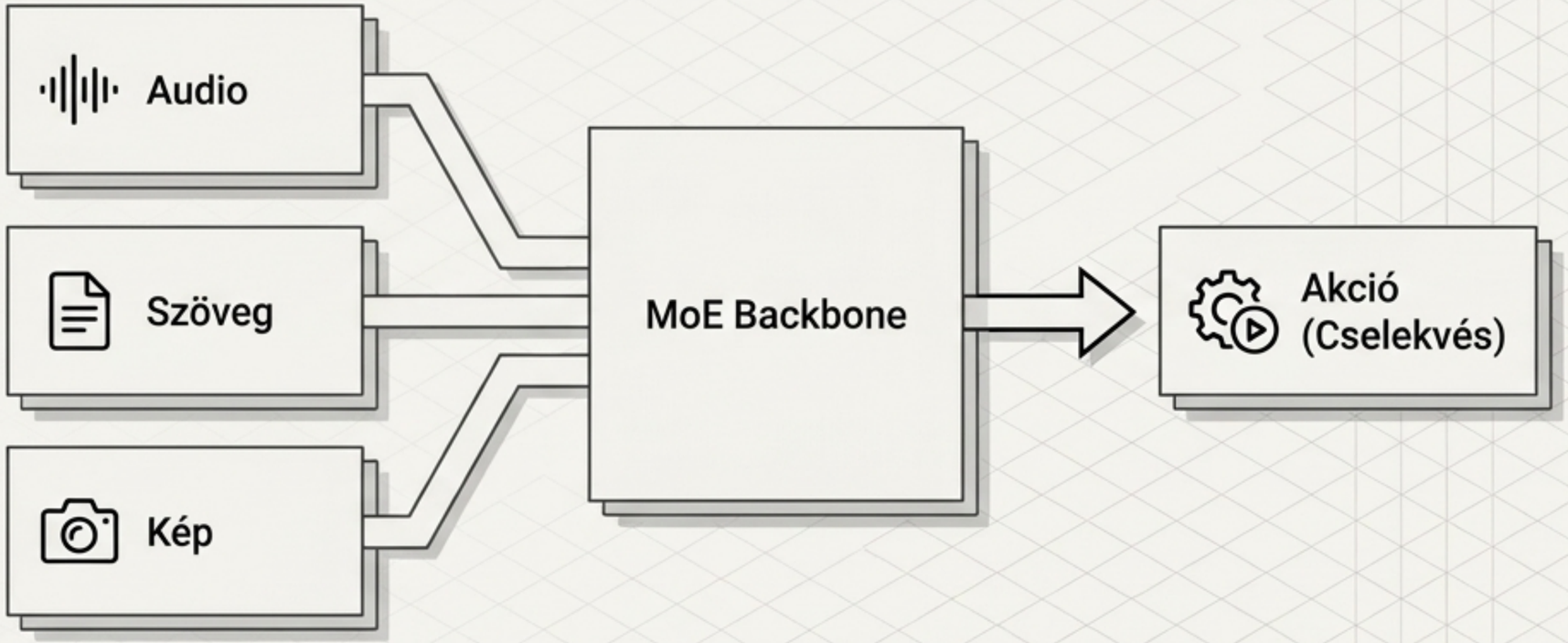
APL Voice 27B: Hangalapú Munkavégzés

Modell Specifikáció

Fejlesztő:
ServiceNow

Méret:
27B paraméter

Státusz:
Nyílt súlyok, lokálisan futtatható



Hang-első megközelítés: Dedikált beszédenkóderrel és token beágyazással rendelkezik.

Multimodális szintézis: Képes egyszerre befogadni és értelmezni a hangot, szöveget és a rögzített vizuális interfészt.

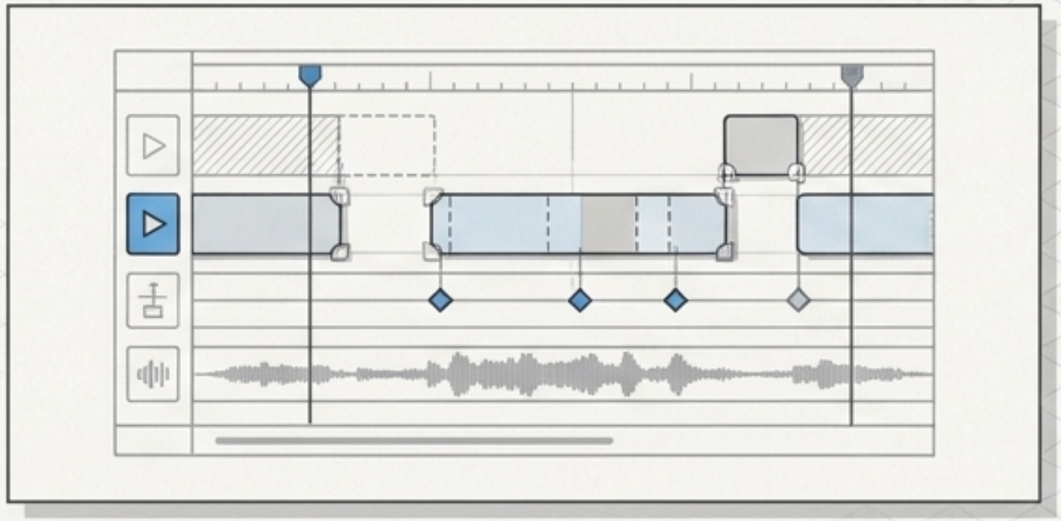
Valós feladatvégzés: Nem csak beszélget, hanem a kapott hangutasítások alapján cselekszik a rendszerben (Text-to-Speech támogatással).

Gemini Notebook

A Vizuális Forradalom: Pika H3 & PH3 Image

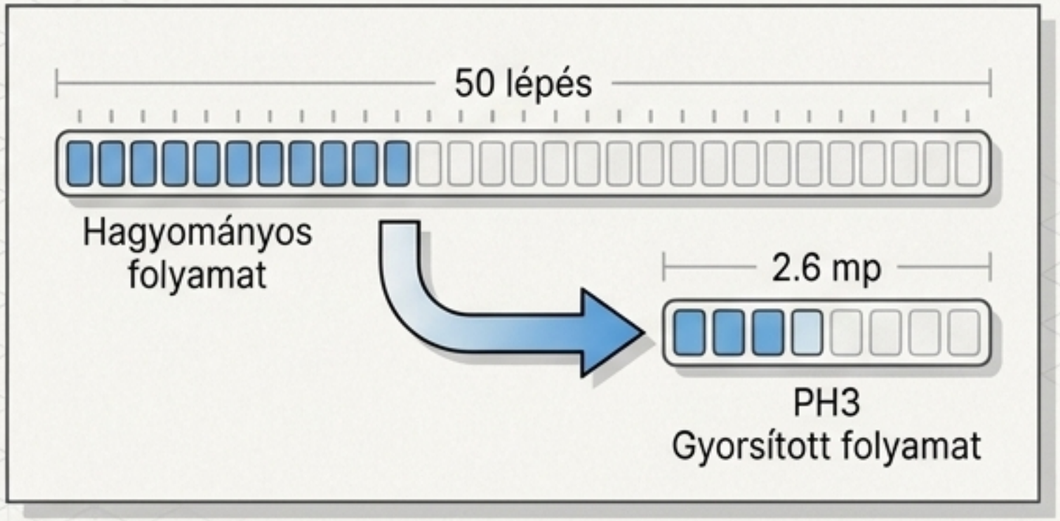
Miniax H3 Alapok

Pika H3 Flex (Videó)



- **Valós időnél gyorsabb** generálás (15mp klip + hang 7mp alatt).
- **Exportálható** audio sávok.
- **Rendkívül olcsó:** $0.01 / másodperc.
- **Kiváló kameramozgás és szerkesztői kontroll** (több képből és hangfájlból dolgozik).

PH3 Image (Kép)



- **5x gyorsabb:** 50 lépés helyett 8 lépéses generálás (2.6 mp).
- **7x kisebb méret:** 55GB (BF16) vagy 30GB (Int 8).
- **Kiemelkedő** fotorealizmus és hibátlan kézírás/szöveg renderelés.

Gemini Notebook

Qwen 3.5 4B: A Kis 3D Kódoló Bajnok

Modell Specifikáció

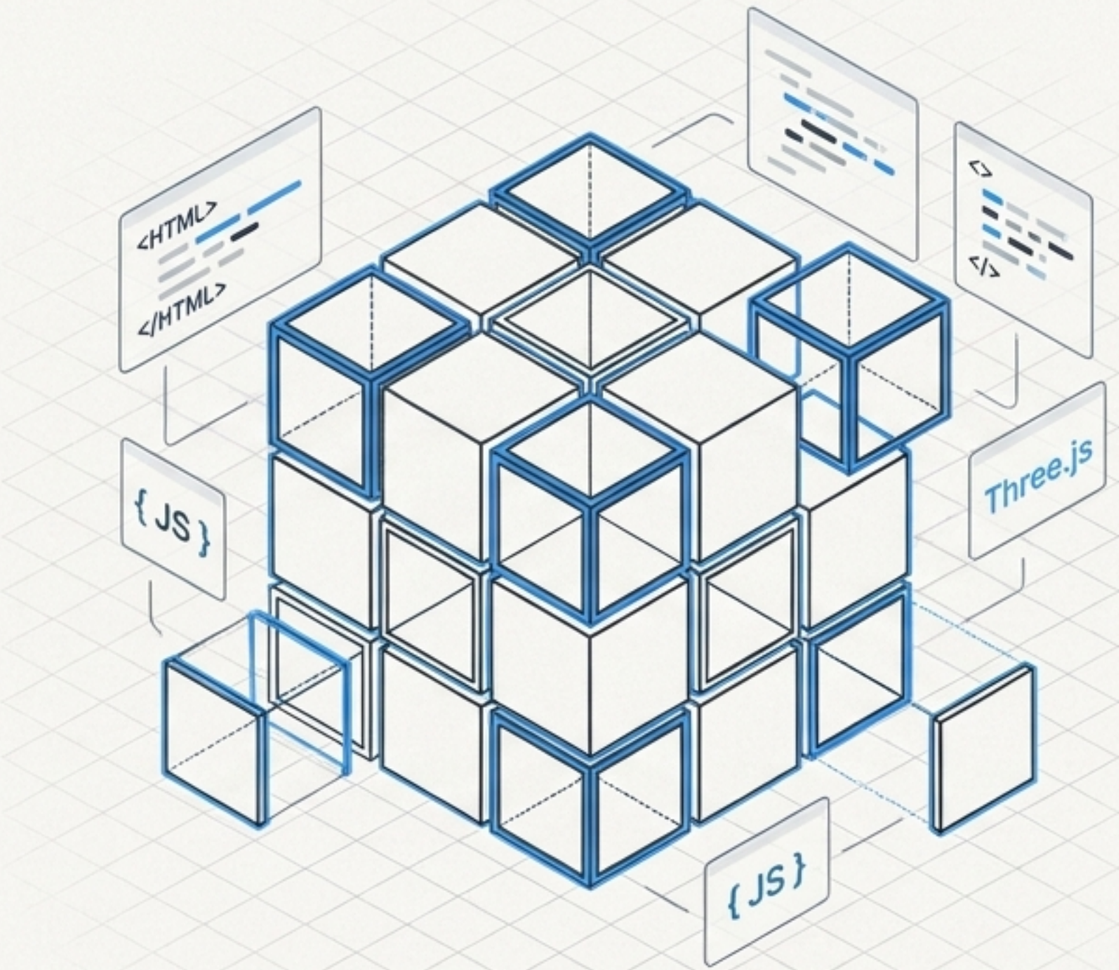
Méret: 4 Milliárd paraméter (Mikro-modell)

Futtatás: Hihetetlenül gyors lokális inferencia

Specializáció: Kifejezetten önálló 3D környezetek kódolására finomhangolva.

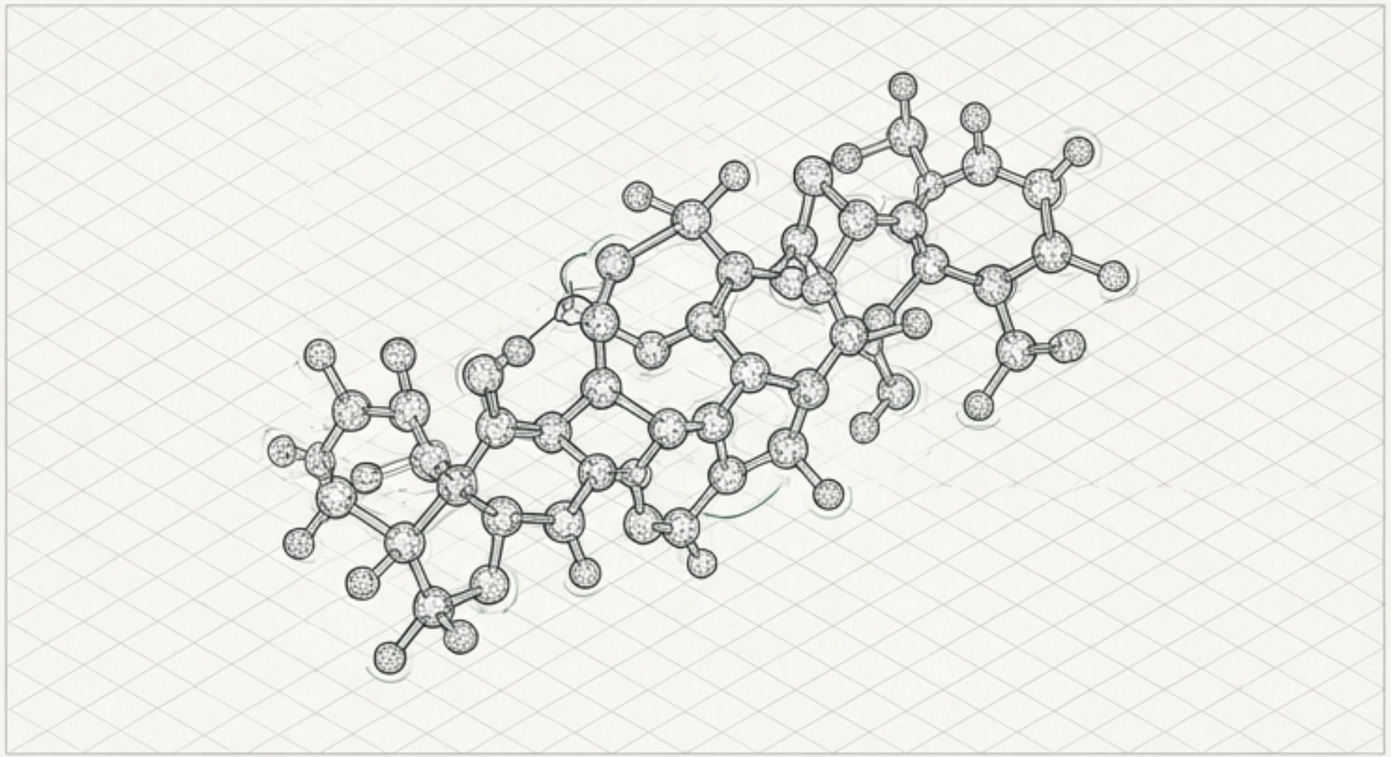
Kimenet: Three.js, HTML és JavaScript jelenetek generálása (.glb export támogatással).

Célközönség: Játékfejlesztők és indie alkotók (pl. Godot assetek, platformerek azonnali kódolása).



Gemini Notebook

Hydrogen 1.0: A Tudományos Felfedező

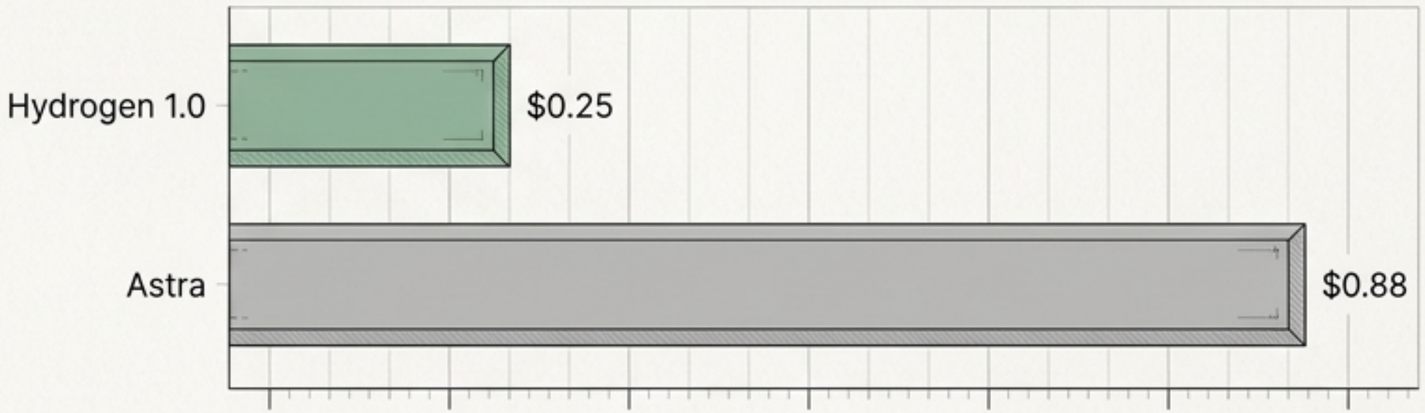


Modell Specifikáció

	Fejlesztő: MIT (GLM 5.3 alapokon)
	Benchmark: 38.33% a Frontier Science teszten

Képesség: Képes kódot írni komplex 3D molekuláris szimulációkhoz (pl. polietilén olvadásának térbeli modellezése).

Költséghatékonyság: Legyőzi a GPT-6 Astra-t és a zárt modellek többségét tudományos kutatásban, töredék áron.

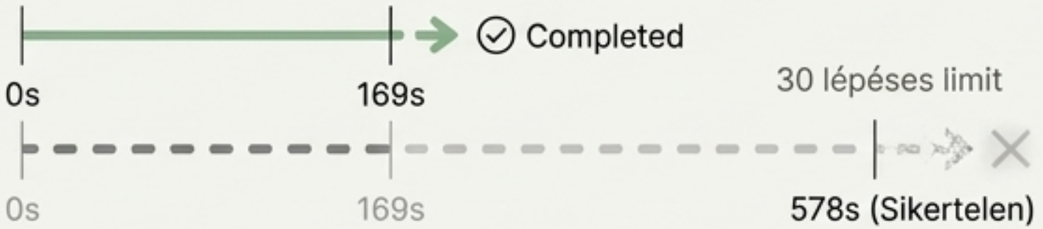


Gemini Notebook

Clover & LFM3: Túl az Autoregresszión

Clover (Diffusion Text)

- 35B paraméteres MoE (3B aktív).
- Sebesség teszt: A diffúziós architektúra révén egy komplex feladatot 169 másodperc alatt fejez be, míg a Qwen 3.6 Flash 578 másodperc után is elbukik a feladaton (30 lépéses limit).



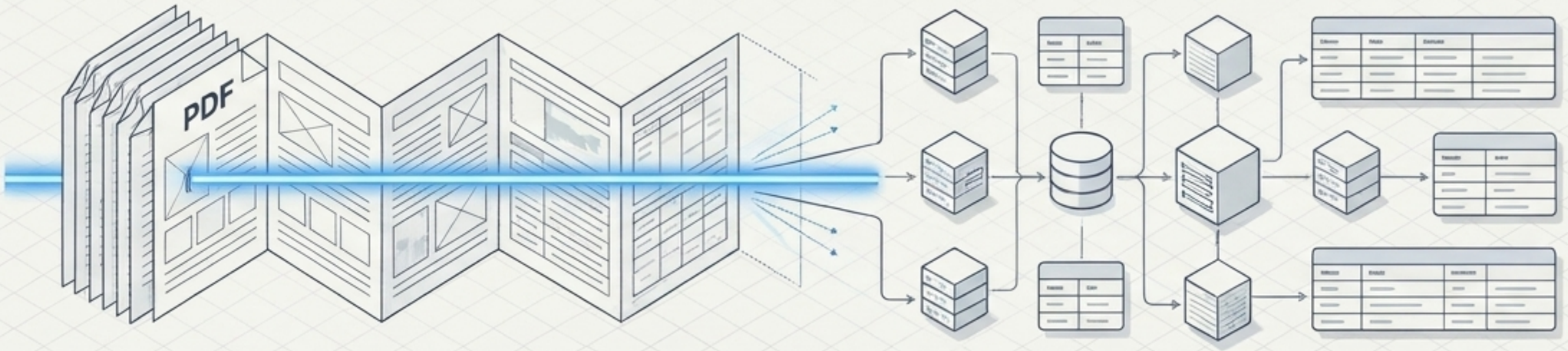
LFM3 (Liquid AI)

- Új generációs ‘Liquid’ alapmodell (1.7B paraméter).
- Extrém optimalizáció: 4-bites precízióval mindössze 1 GB memóriát igényel.
- Térbeli tudatosság, videó és szöveg értelmezés – közvetlenül mobileszközökön.



Gemini Notebook

Adatfeldolgozás: Vizuális Keresés és Korlátlan OCR



Perplexity Embed V2

- Vizuális-nyelvi beágyazó modell (9B és 600M verziók).
- **Funkció:** Kifejezetten adatok (szöveg és kép) visszakeresésére optimalizálva PDF-ekből és vizuális dokumentumokból.
- Lokálisan futtatható, nyílt súlyokkal.

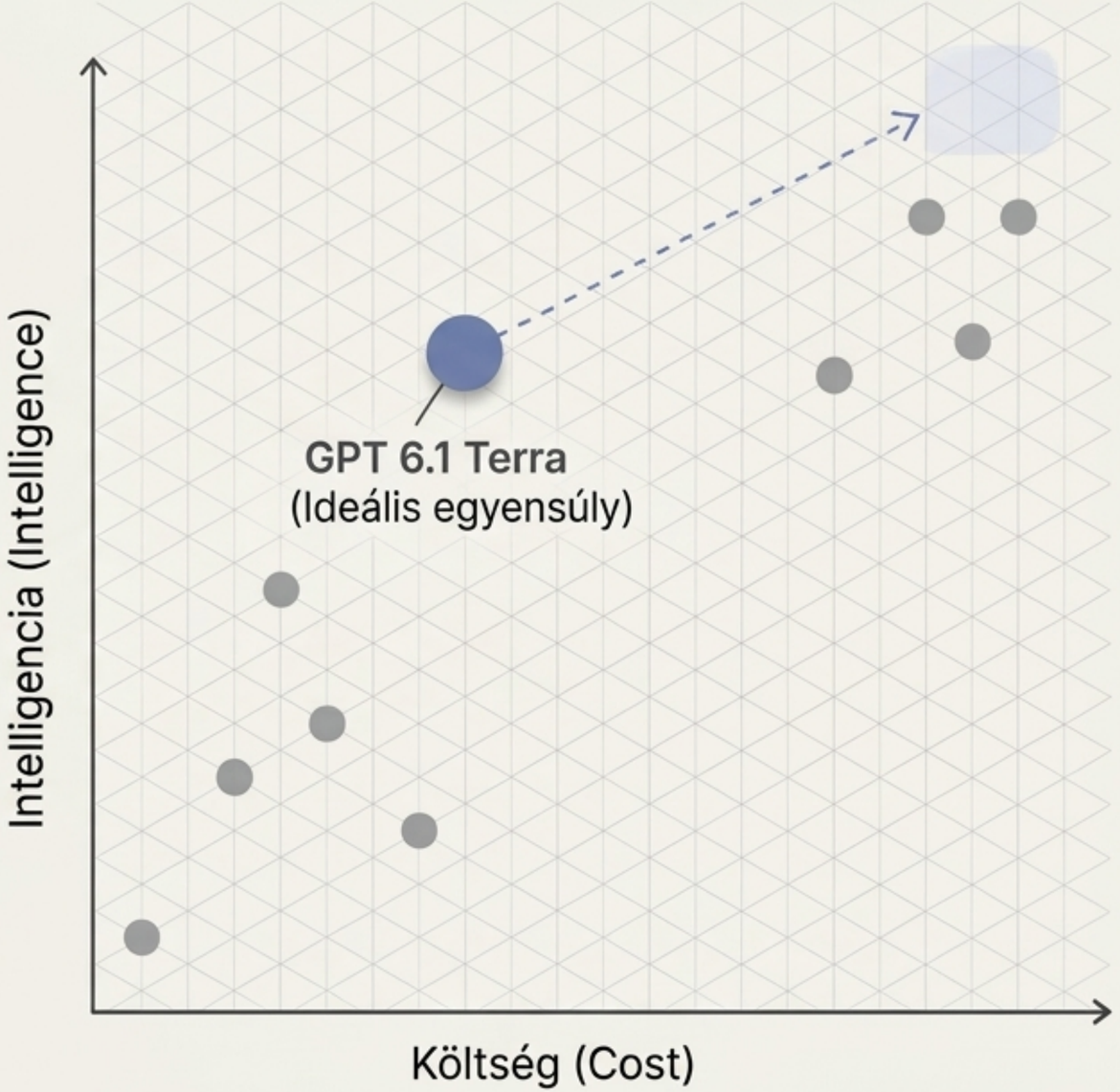
Unlimited OCR 2 (Byte Research)

- **Nyílt forráskódú** modell kép-alapú dokumentumok olvasására.
- **Áttörés:** Nincs oldalkorlát és nem növekszik a "figyelem-költség" (attention cost).
- **Teljesítmény:** 5000+ oldalas dokumentumokat is tökéletes pontossággal dolgoz fel egyetlen futtatás alatt.

Gemini Notebook

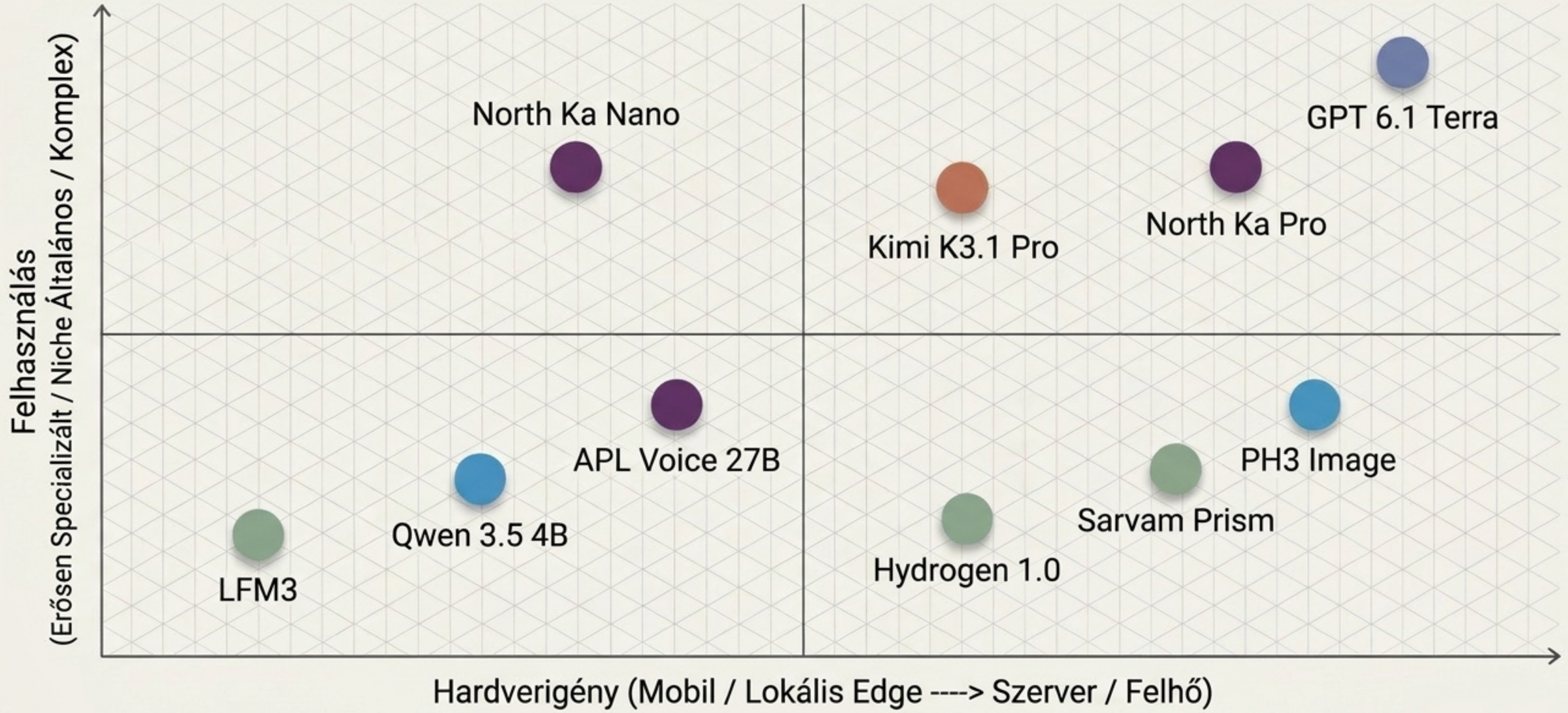
GPT 6.1 Terra: A Zárt Kódú Ár-Érték Bajnok

- **Árcsökkentés:** 50%-kal alacsonyabb standard API árak az elődd Terra 5.6-hoz képest.
- **Teljesítmény arány**: Képes felvenni a versenyt a csúcsmodell GPT-6 Astra-val a ‘highest thinking budget’ kategóriában, de annak csupán 1/10-ed áráért.
- **Sebesség**: Átlagosan 3x gyorsabban fejezi be a feladatokat, mint az Astra.
- **Felhasználás**: A legjobb választás masszív, magas token-felhasználású munkafolyamatokhoz és videószerkesztéshez.



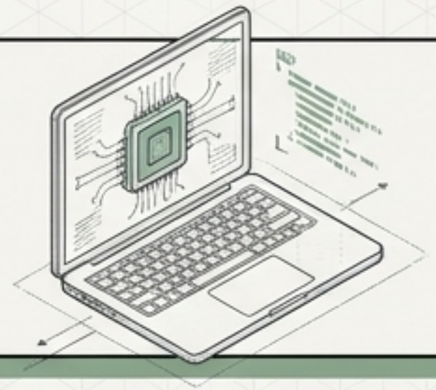
Gemini Notebook

Mátrix: Az Új Mesterséges Intelligencia Térkép



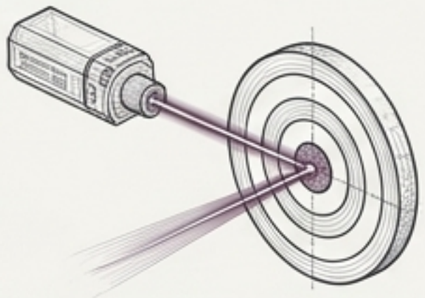
Gemini Notebook

Záró Gondolatok: Mi Változott Ezen a Héten?



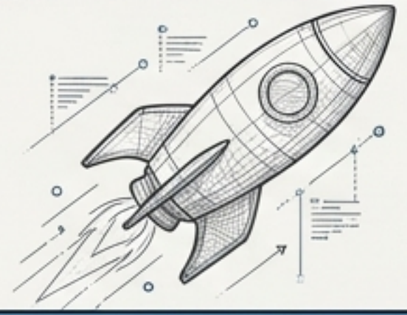
A Lokális Futtatás Aranykora.

Az LFM3, a kis MoE modellek és a 4B paraméteres specializált ügynökök bebizonyították, hogy a komoly AI munkafolyamatokhoz többé nem kötelező a felhő.



A Specializáció Győzelme.

A tudományos (Hydrogen), döntéshozó (Prism) és OCR (Unlimited 2) modellek megmutatták, hogy egy célzott, kisebb modell gyorsabb és olcsóbb, mint egy általános célú óriás.



Az Architektúrák Forradalma.

A nyílt forráskód már nem csak lemásolja a zárt modelleket; a Diffúziós szöveggenerálás és a Liquid hálózatok révén a sebesség és az innováció élvonalába került.

Gemini Notebook