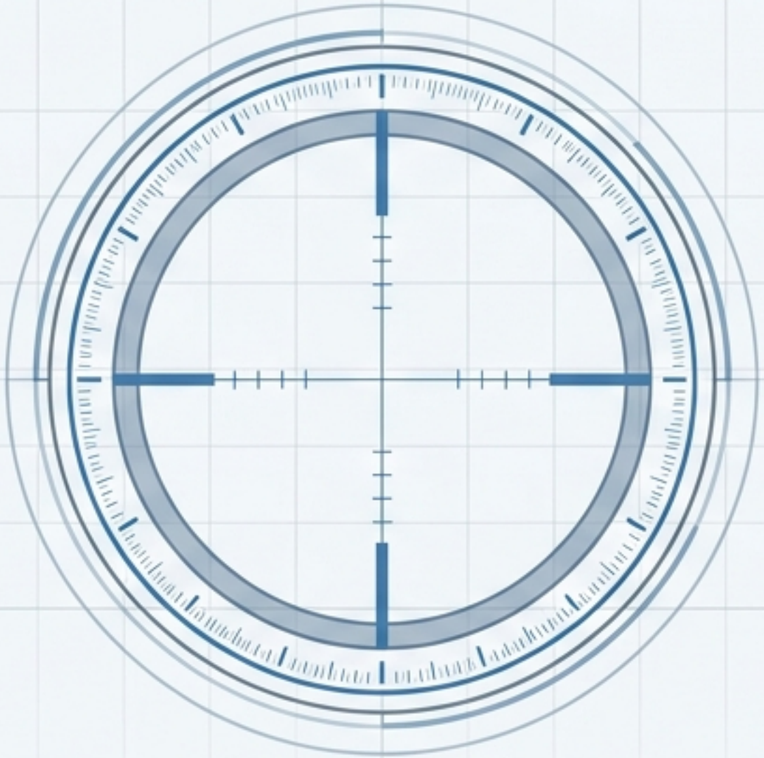


Jelentés a Laborból: A Kimmy K 3.1 Diagnosztikája

Egy 3 billió paraméteres nyílt súlyú (open-weights) “Frontier” modell anatómiai és teljesítmény alapú vizsgálata.

[STATUS: Elemzés befejezve] [FORRÁS: Matthew Berman Stressзteszt] [CÉL: Képesség-felmérés]

Gemini Notebook

Vezetői Összefoglaló

1. A Mérföldkő

Nyílt súlyú (open-weights) modellek végre elérték a legmagasabb, "Frontier" (élvonalbeli) kategóriát. A Kimmy K 3.1 teljesítménye az Opus 5.5 és a GPT 6.1 Soul szintjével versenyez.



2. A Teszt Konklúziója

Kiváló kódolási, multimodális és szintetikus intelligencia képességek. Azonban a tiszta térbeli logika (spatial reasoning) és az elsődleges adatpontosság terén emberi irányítást (nudging) igényel.



3. Az Igazi Érték

A modell legfőbb vonzereje nem csupán a nyers erő, hanem a **példátlan intelligencia-költség arány (Cost per Intelligence Task)**, amely a piac legvonzóbb szereplőjévé teszi a fejlesztők számára.



Gemini Notebook

A Motorháztető Alatt: Specifikációk

3 Billió

Teljes paraméterszám. Masszív, iparági csúcs méret a nyílt modellek között.

1 Millió

Token kontextus ablak. Extrém hosszú dokumentumok és kódbázisok feldolgozásához.

Multi-Modális

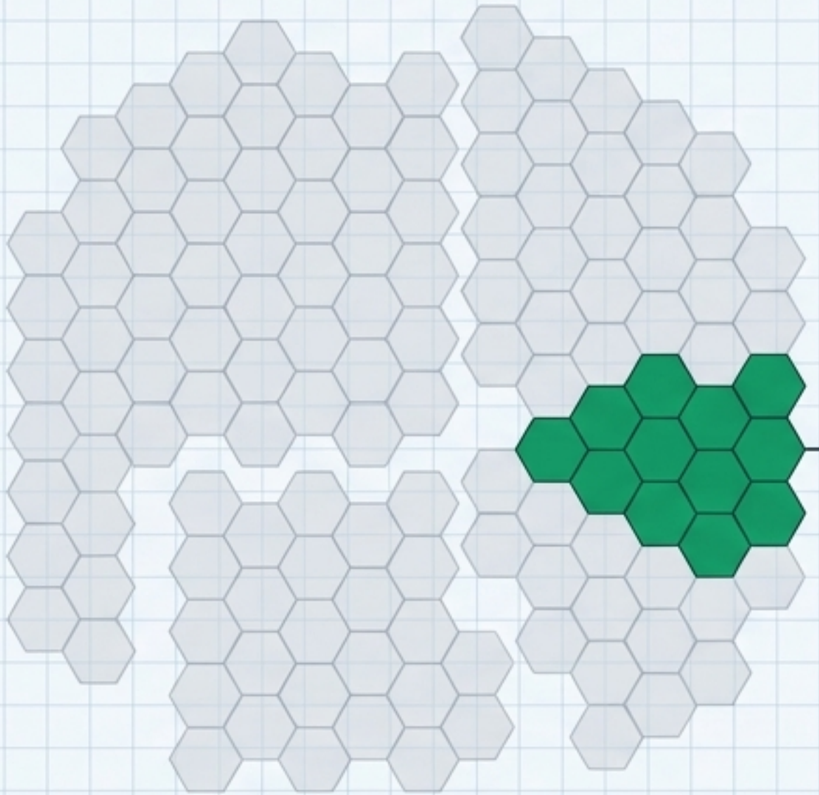
Szöveg, kép és valószínűleg hang natív feldolgozása.

-65%

A laboratóriumi tesztek alapján ennyivel csökkent a hallucinációk aránya az előző generációhoz képest.

Gemini Notebook

A Hatékonyság Titka: Mixture of Experts (MoE)



[3 Billió Teljes Paraméter]

Prompt

[Csak 104 Milliárd Aktív Paraméter]

Bár a modell teljes mérete 3 billió paraméter, egy adott prompt feldolgozásakor (routing) a rendszer csupán 104 milliárdot aktivál.

[Miért fontos ez?]

Ez a strukturális döntés teszi lehetővé, hogy a modell villámgyors és költséghatékony maradjon anélkül, hogy feláldozná a gigantikus, mindent átfogó tudásbázisát.

Gemini Notebook

Képesség-Tesztek: A Laboratóriumi Eredmények Összegzése

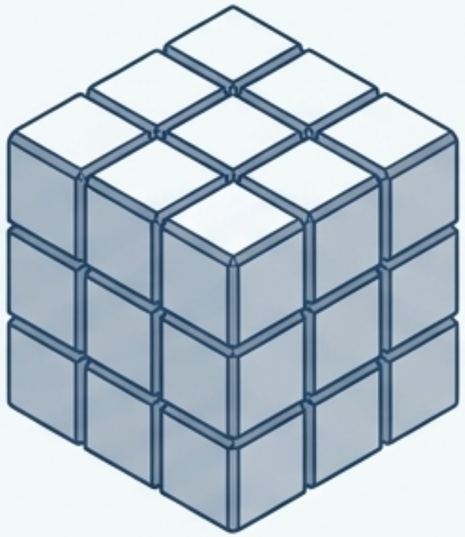
Teszt (Task)	Kategória	Eredmény	Fő Tanulság
Rubik Kocka (3D WebGL)	Térbeli / Kódolás	🟢 Siker	Tökéletes fizika, algoritmus-alapú megoldás, világítás.
iPhone Duo (WebGL)	3D Modellezés	🟠 Vegyes	Jó renderelés, de pontatlan fizikai dizájn (South Park stílusú logó).
Blender Mászócsomó	Térbeli Logika	❌ Kudarc	A forma megvan, de a csomó mechanikailag hibás, nincs rögzítve.
Csomókötés Animáció	Motion Graphics	❌ Kudarc	A kötelek egymásba lógása miatt tanulásra alkalmatlan.
Fall Guys Klón (Web)	Játék Kódolás	🟢 Siker	Hibátlan logika, fizika és ellenfelek, nagyon rövid prompt
YTP Videó Generálás	API Orkesztráció	🟢 Siker	Komplex Python, FFmpeg és audio generálás tökéletes koordinációja.

Gemini Notebook

Térbeli Intelligencia: A Kódolt vs. A Fizikai Tér

A Siker: Rubik Kocka WebGL-ben

SIKER

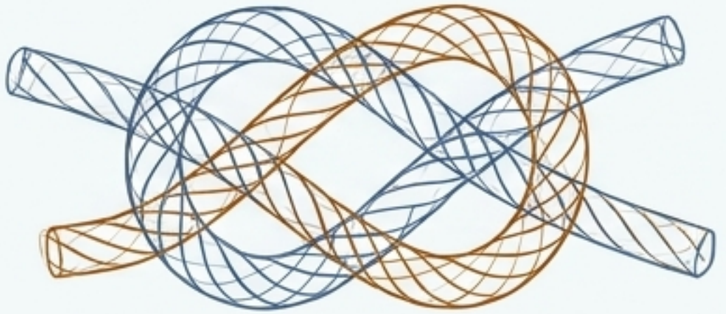


Képesség: Algoritmikus 3D modellezés.

Eredmény: Képes volt egy interaktív, játszható Rubik-kockát kódolni valós idejű fényvisszaverődéssel és automatikus megoldó algoritmussal. A nyílt modellek korábban mindig elbuktak ezen a teszten.

A Kudarc: Blender Mászócsomó

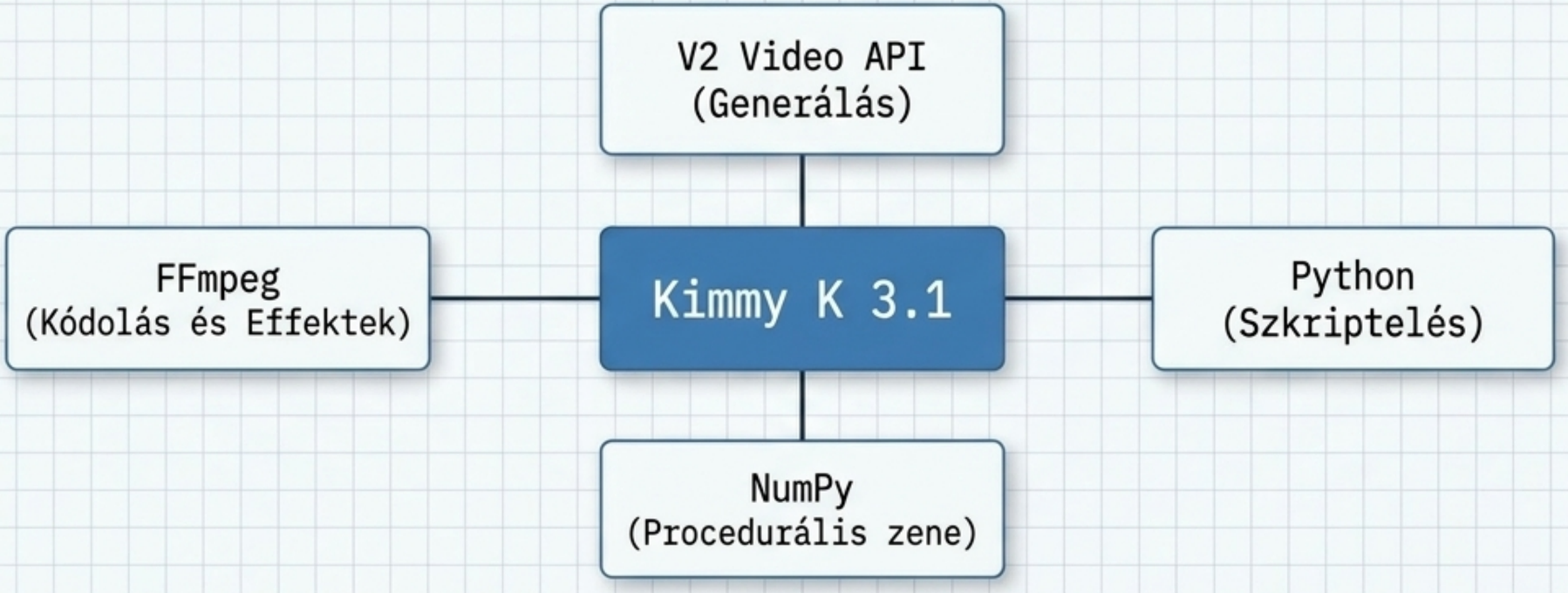
KUDARC



Képesség: Tiszta térbeli logika és fizikai interakció.
Eredmény: Bár tudta irányítani a Blender szoftvert, a generált "biztosító csomó" (belaying knot) a valóságban használhatatlan. A kötelek egymásba csúsztak, a térbeli alatta/felette relációkat nem értette meg.

Gemini Notebook

Ágens Képességek és Orkesztráció



Az Esettanulmány
Egy 5 perces paródiavideó elkészítésénél a Kimmy önállóan "karmesterként" működött: meghívta a Video API-t, írt egy Python szkriptet a vágáshoz, használta a NumPy-t zene generálására, és az FFmpeg-et a videó effektezésére. Komplex folyamatok tökéletes koordinációja.

Fejlesztői Ízlés
A webdesign teszteknél a generált kód funkcionálisan hibátlan volt és követte a megadott arculati kézikönyvet, de esztétikailag erős "AI dizájn" érzetet keltett (például pontatlan és inkonzisztens térközök alkalmazása).

Gemini Notebook

A "Nudging" Faktor és az "Eyebrow Text"

Clinical Technical Blueprint

A "Nudging" Munkafolyamat

1. Első Kimenet

Gyakran hallucinált vagy felületes. (Pl. egy diagramnál a modell kitalált helykitöltő számokat használt a valós adatok helyett).

2. Emberi Nudge

A felhasználó verbális nyomást gyakorol és korrigál: "Használd a valós adatokat és pontos számokat!"

3. Korrigált Eredmény

A modell beolvassa a valós adatokat és tökéletesen formázza a végső, magas minőségű kimenetet.

A Védelmező "Eyebrow Text"

> Hasonlóan a Claude-hoz és a ChatGPT-hez, a Kimmy is tele van építve biztonsági fékekkel. Egy csomókötési animációhoz önként hozzáfűzte a figyelmeztetést: "**Nem helyettesíti a szakképzett oktatást**". Ez azt mutatja, hogy a nyílt modellek is kezdik átvenni a zárt modellek szigorú biztonsági viselkedését.

Gemini Notebook

A "Frontier" Versenytárs Mátrix

Modell	Hozzáférés Típusa	MI Index (AA)	3D/Térbeli Képesség	Benchmark Kategória
Kimmy K 3.1 Pro	Nyílt Súlyú (Open Weights)	55	Kiváló (Kódolt)	Elit
Opus 5.5	Zárt (Proprietary)	55	Kiváló	Elit (A jelenlegi király)
GPT 6.1 Soul	Zárt (Proprietary)	[Adat hiányzik]	Kiváló	Elit
Muse Wonderful Full	Nyílt Súlyú (Open Weights)	55	Gyenge (Meta elismerte)	Közép-Elit

Gemini Notebook

Szintetikus Tesztek: Hol áll a Kimmy?

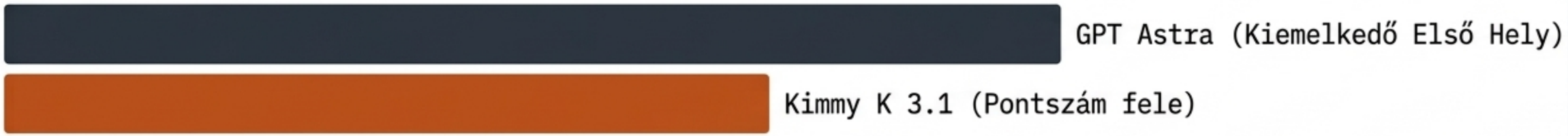
Erősségek (Hol veri a piacot?)

Kódolás és Valós Tudásmunka



Gyengeségek (A vakfolt)

A Tudományos Logika Lemaradása



Gemini Notebook

Gazdaságosság: API Árazási Struktúra

Standard Sebesség

Bemenet (Input):	$2.00 / 1M token
Kimenet (Output):	$10.00 / 1M token

Kritikus Előny: A Cache-elt bemenet mindössze $0.20 / 1M token!

High-Speed (Pro)

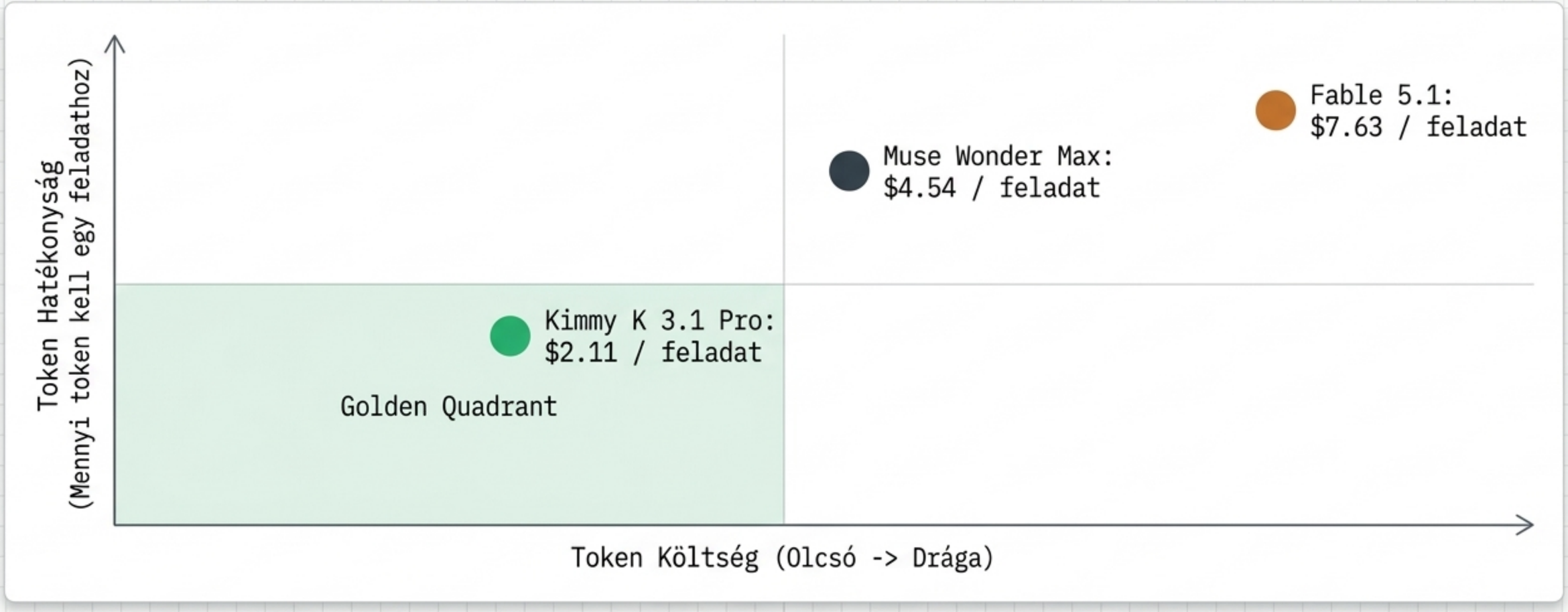
2-3x gyorsabb végrehajtás.

Bemenet (Input):	$4.00 / 1M token
Kimenet (Output):	$20.20 / 1M token

Kritikus Előny: A Cache-elt bemenet itt is változatlanul $0.20. Masszív sebességnövekedés minimális extra költséggel a cachelt folyamatoknál.

Gemini Notebook

A Döntő Metrika:
Költség per Intelligencia Feladat



A Kimmy nem csak olcsó API szinten, de rendkívül kevés lépésből (tokenből) oldja meg a komplex feladatokat. Ez a kombináció teszi jelenleg a legjövedelmezőbb "Frontier" modellé a fejlesztők számára.

Gemini Notebook

Clinical Technical Blueprint

Végső Értékelés: A Kimmy K 3.1 Helye a Stackben

Mikor Használjuk?

- ✅ Komplex kódolási és API orkesztrációs feladatok (kiváló önálló ágens képességek).
- ✅ Költségérzékeny, nagy volumenű műveletek (verhetetlen Cost per Intelligence arány).
- ✅ Adatvédelemre érzékeny, zárt környezetek (nyílt súlyú felépítés lokális futtatáshoz).

Mikor Kerüljük?

- ⚠️ Tisztán vizuális és fizikai térbeli logikát (spatial reasoning) igénylő feladatok.
- ⚠️ Mély tudományos (Science) benchmark igényű kalkulációk.
- ⚠️ Olyan esetek, ahol a 'Zero-shot' (elsőre tökéletes) pontosság elengedhetetlen, és nincs lehetőség iteratív irányításra (Nudging).

Összegzés

A Kimmy K 3.1 bebizonyította, hogy a nyílt forráskód nem csupán felzárkózott, hanem ár-érték arányban és fejlesztői rugalmasságban le is tudja körözni a zárt ökoszisztémákat. A "Frontier" kategória immár demokratizálódott.

Gemini Notebook