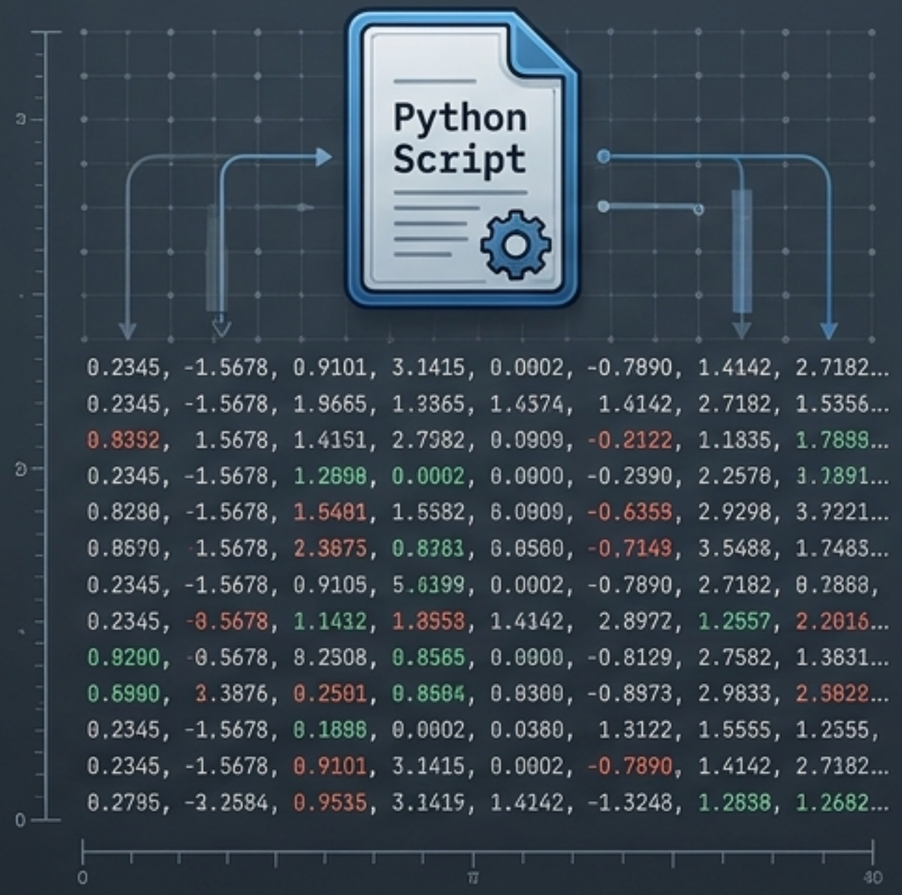


Az intelligencia tervrajza

Hogyan épülnek fel a **Nagy Nyelvi Modellek** (LLM) egy egyszerű **Python** kódtól a komplex érvelésig?

Gemini Notebook

A varázslat mögötti valóság





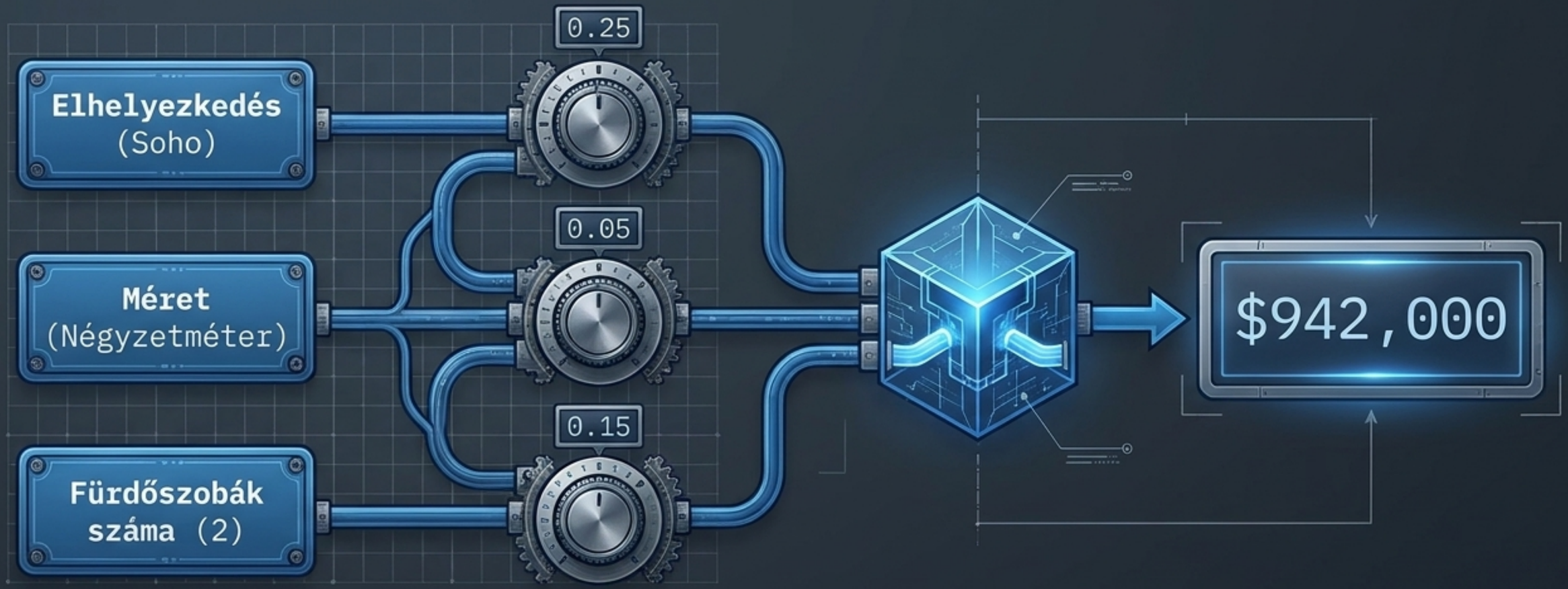


Az Agy: Milliárdnyi paraméter (számok egy fájlban), amelyek az internet tudását kódolják.

A Huzalozás: Egy rövid Python kód, amely meghatározza, hogyan áramoljon az információ a számok között, amíg intelligens választ nem kapunk.

Gemini Notebook

A neurális háló anatómiája: A New York-i lakás példája



A hálózat kezdetben vakon tippel. A bemeneteket véletlenszerű számokkal (tekerőgombokkal) szorozza fel, hogy eredményt kapjon.

Gemini Notebook

A Veszteség (Loss) mérése



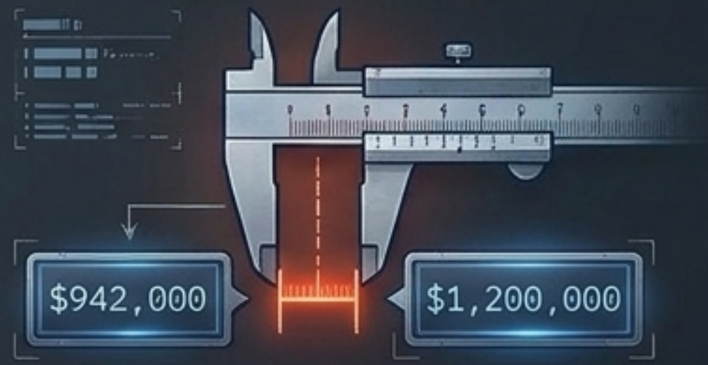
- A gépi tanulás egyetlen, legegyszerűbb célja: **Ezt a piros rést nullára csökkenteni**.
- De hogyan tudjuk, melyik tekerőgombot merre csavarjuk több millió paraméter esetén?

Gemini Notebook

A gépi tanulás motorja: Backpropagation

1. Mérés

Kiszámoljuk a Veszteséget (A különbség a jó és a rossz válasz között).



2. PyTorch (A nyomkövető)

A szoftver visszaköveti a folyamatot. Kiszámolja (deriválással), hogy egy adott gomb eltekerése növelné vagy csökkentené a veszteséget.

Ezt a folyamatot hívják **Backpropagation**-nek. A rendszer nem találgat; pontosan tudja, melyik gombot milyen irányba kell tekerni a pontosabb eredményhez.

3. Finomhangolás (Lépés)

A rendszer egy aprót mozdít minden tekerőgombon a zöld cél irányába.





Gemini Notebook

Ugrás a számoktól a szövegek világába

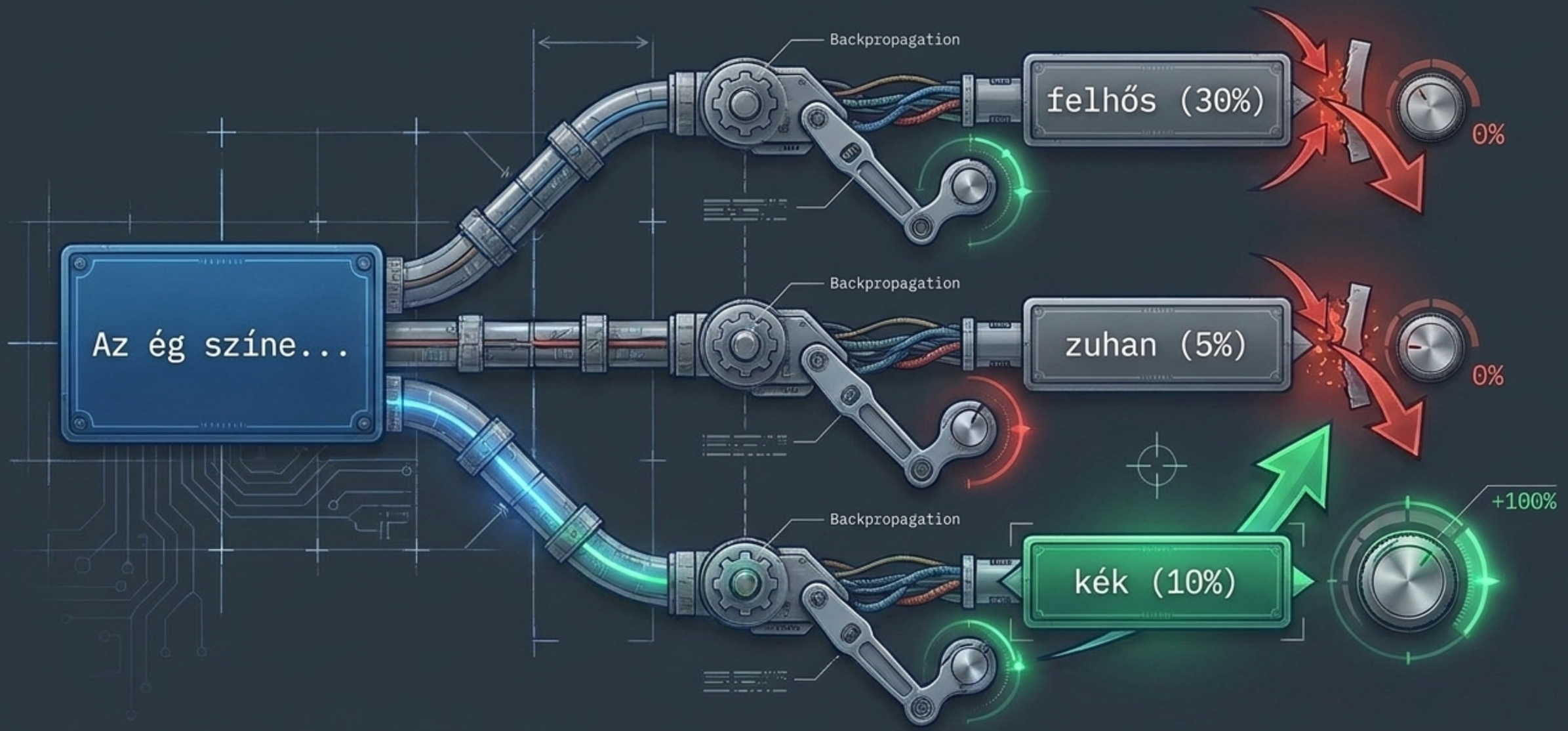
Dimenziók	Hagyományos Neurális Háló	Nagy Nyelvi Modell (LLM)
Bemenet (Inputs)	Számok (Pl. négyzetméter)	Szöveg/Tokenek (Pl. szavak töredékei)
Kimenet (Outputs)	Egy konkrét érték (Ár)	A lehetséges következő szavak valószínűségi eloszlása (10, 0.05, 0.1, 0.1, 0.03, …)
A Cél (The Goal)	A távolság **minimalizálása** a **valós számtól**	A helyes következő szó valószínűségének **100%-ra** tolása



Az LLM-ek(mint a ChatGPT) is csak tekerőgombokat használnak, de a céljuk nem egy ár, hanem a legvalószínűbb következő szó megjóslása.

Gemini Notebook

A Valószínűségi Fa: Hogyan téved egy nyelvi modell?



Gemini Notebook

1. Fázis: Előtanítás (A Mindentudó Papagáj)

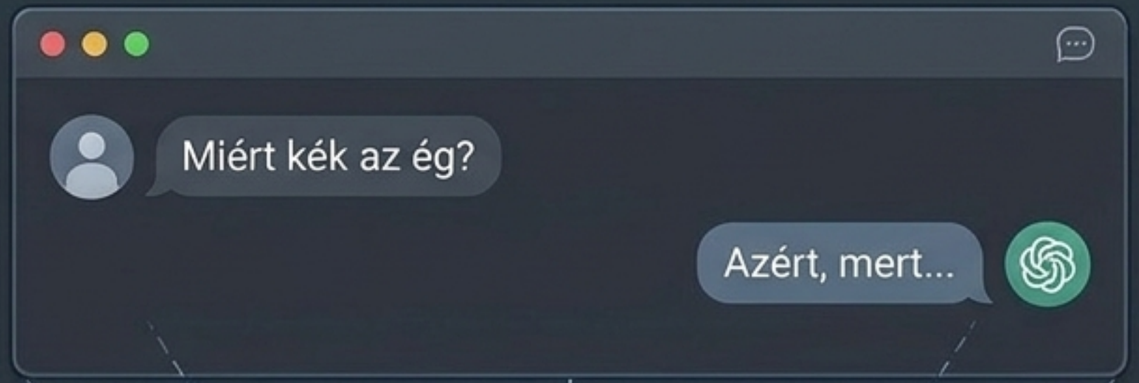


Az Alapmodell nem akar segíteni. Nincs személyisége. Csak egy papagáj, amely a betáplált internetes minták alapján próbálja kitalálni a következő szót.

Gemini Notebook

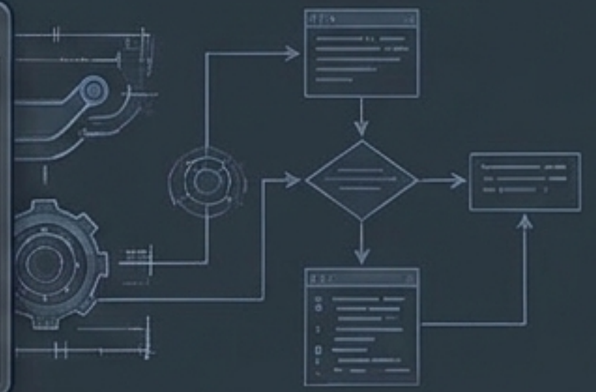
2. Fázis: Utótanítás (Az Asszisztens Trükk)

Felszín



A Motorháztető alatt

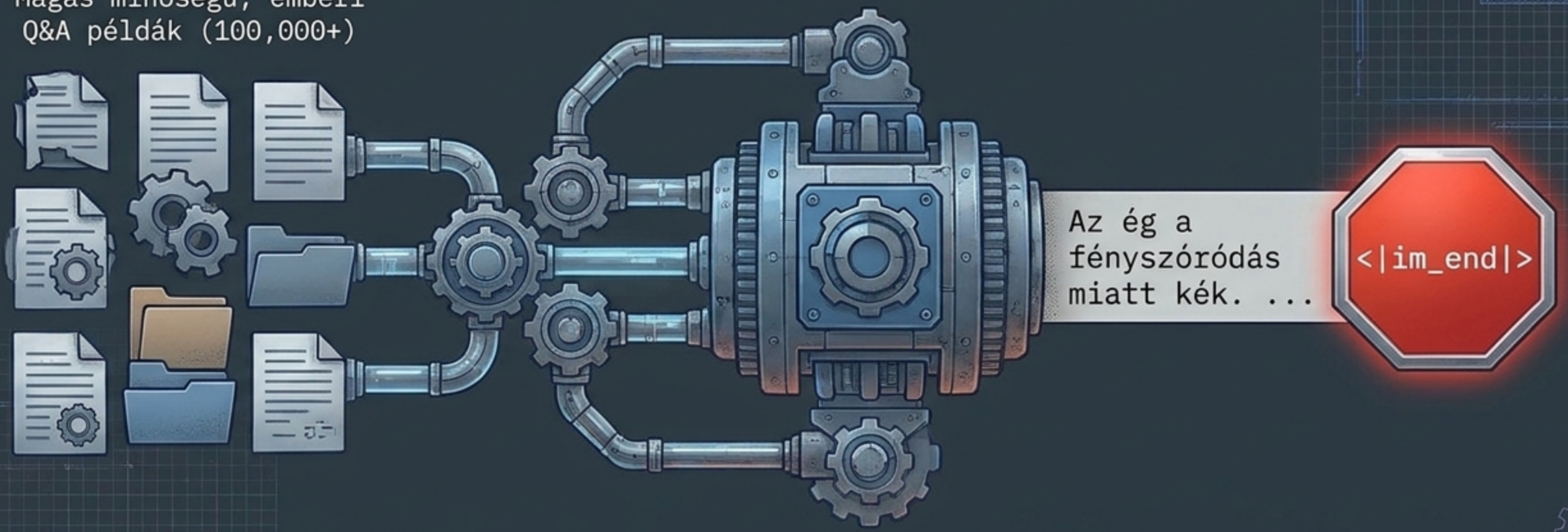
```
<|im_start|>system
You are a helpful assistant<|im_end|>
<|im_start|>user
Miért kék az ég?<|im_end|>
<|im_start|>assistant
```



Az LLM továbbra is csak autocompletet csinál. De ha speciális szerepjáték formátumokba (tokenekbe) kényszerítjük, elkezdi egy segítőkész asszisztens mintáit utánozni.

Gemini Notebook

Az Asszisztens megszületése és a Stop Token



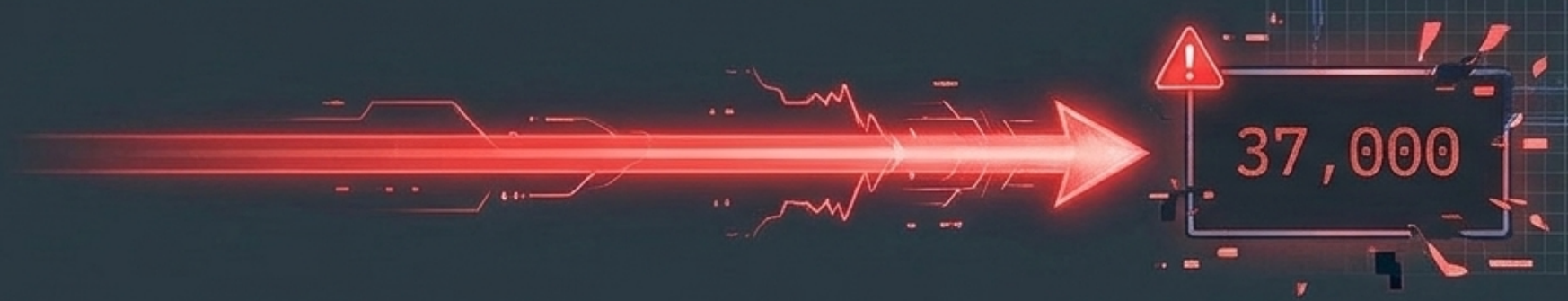
- Ebben a fázisban a modell egót kap. Megtanulja a feladatát (segíteni a felhasználót).
- A legfontosabb lecke: Megtanulja, mikor kell abbahagyni a generálást a Stop Token használatával, ahelyett, hogy a végtelenségig ontaná a szavakat.

Gemini Notebook

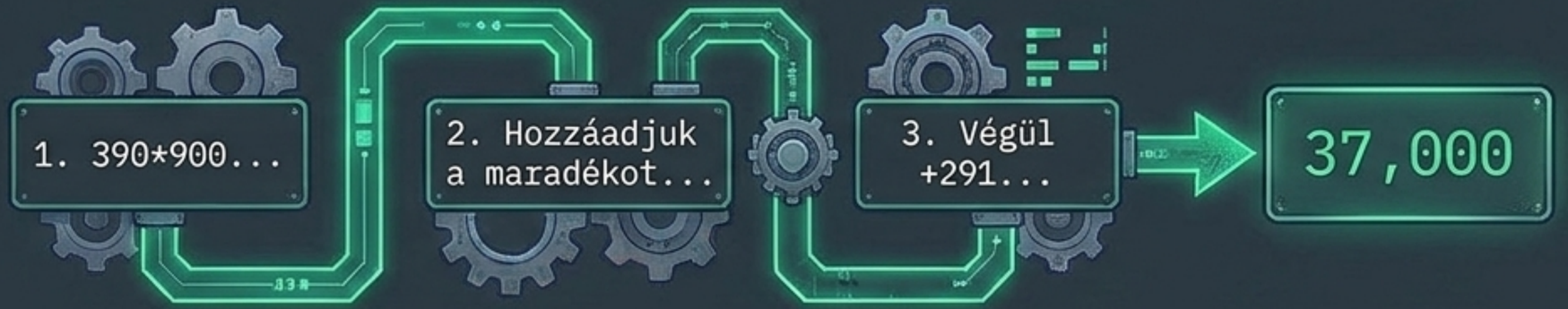
A Korlát: Miért bukik el a modell a matematikában?

$$398 * 923 + 291 = ?$$

Hagyományos Asszisztens Modell (Azonnali válaszadás gondolkodás nélkül)



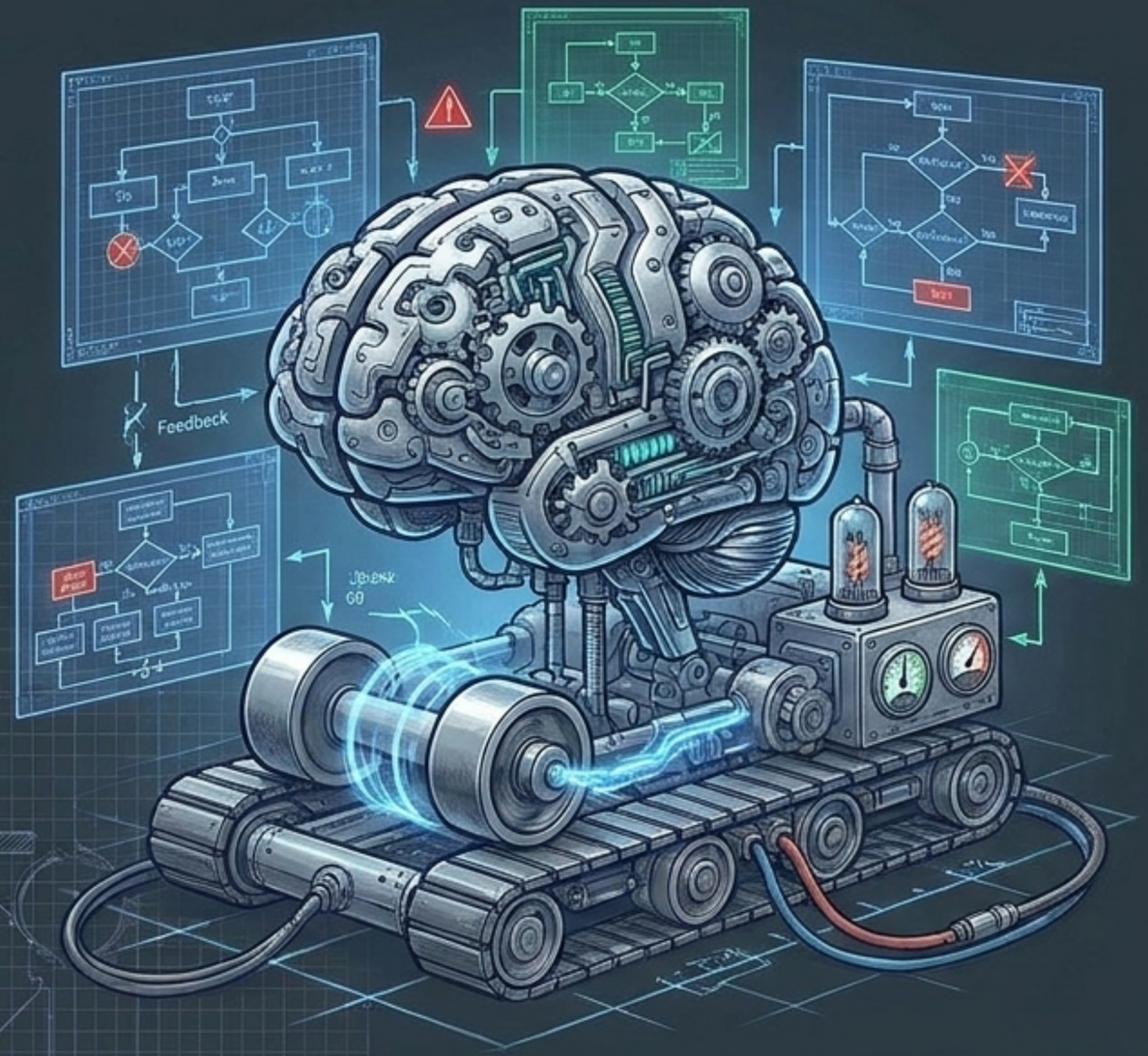
Az Érvelő Út



Az internet tele van végeredményekkel, de hiányzik belőle a gondolkodás folyamata. A modell nem tudja, hogyan jusson el lépésről lépésre a megoldásig.

Gemini Notebook

3. Fázis: Megerősítéses Tanulás (Gondolkodj lépésről lépésre)



- Ha a tréningadatokból hiányzik a gondolkodás, a modellnek magát kell megtanítania rá.

- Ezt hívjuk Megerősítéses Tanulásnak (Reinforcement Learning - RL).

- A modell egy virtuális homokozóban próbálkozik:generál egy megoldást,leellenőrzi, majd próbálkozásaiból (szintetikus adatok) tanulja meg az érvelés logikáját (pl. DeepSeek R1).

Gemini Notebook

A GRPO Mérleg: Hogyan tanul a modell a hibáiból?

Előny < 0
(A rossz tokenek büntetése)



Előny > 0
(A jó tokenek megerősítése)

A modell nemcsak a helyes válaszokból tanul, hanem a saját rossz próbálkozásainak (negatív példák) elvetéséből is.

Gemini Notebook

Miért jobb a saját logika, mint az emberi utánzás?

Emberi logika másolása



Romboló frissítés

Kényszerítjük a modellt, hogy egy emberi gondolatmenetet kövessen, ami felülírja a saját jól működő belső kapcsolatait.

On-Policy RL - A saját adatok



Finomhangolás

Mivel a modell saját maga generálta a jó választ, csak minimálisan kell tekerni a gombokon, hogy legközelebb még valószínűbb legyen. Nincs képesség-romlás.

Gemini Notebook

A Nyelvi Modellek Evolúciója

1. Alapmodell (Pre-trained)	2. Utótanított Modell (Instruct/SFT)	3. Érvelő Modell (RL/Reasoning)
• Cél: Szöveg kiegészítése (Autocomplete)	• Cél: Utasítások követése, formátum tartása	• Cél: Komplex logika önálló megoldása
• Személyiség: A Papagáj (Nincs egója)	• Személyiség: Az Asszisztens	• Személyiség: A Gondolkodó
• Példa: Qwen Base	• Példa: ChatGPT, Claude (Korai verziók)	• Példa: OpenAI o1, DeepSeek R1

Gemini Notebook

A Teljes LLM Képzési Piramis

Csúcs (Reinforcement Learning)

Szintetikus, próba-szerencse adatok (Önálló logika).

Középső réteg (Post-training)

Szűrt, emberi példák (Asszisztens viselkedés és ego).

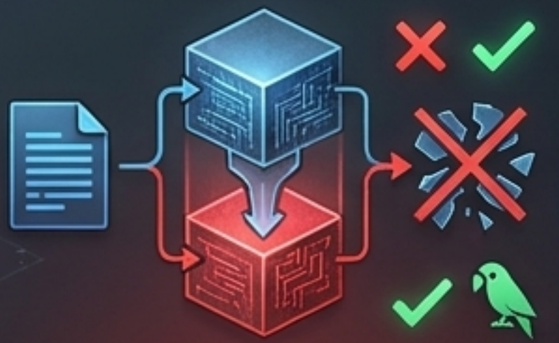
Alap (Pre-training)

Hatalmas, szűretlen internetes adatbázis (Világtudás).



Szigorú sorrend!

Ha egy Asszisztens modellt visszaküldünk az Alapmodell tréningadataira (pl. nyers kódokat etetünk vele), elveszíti az egóját, elfelejt asszisztensként viselkedni, és újra autocompleting papagájjá válik.



Gemini Notebook

Új tudás felfedezése: Kitörés a betanított korlátokból



A Kritika: A modellek csak a betáplált múltbéli adatokat tudják ismételni.
A Valóság: Ahogy Einstein a már ismert fizikai elvekből felépítette az általános relativitáselméletet, a mély érvelésre (Reasoning) képes modellek is képesek új absztrakciókat alkotni.
A Végszó: A hosszan tartó, önálló érvelés (RL) az a mechanizmus, amellyel a jövő modelljei átlépik a tréningadatok határait, és addig megoldatlan emberi problémákat fognak feltörni.

Gemini Notebook