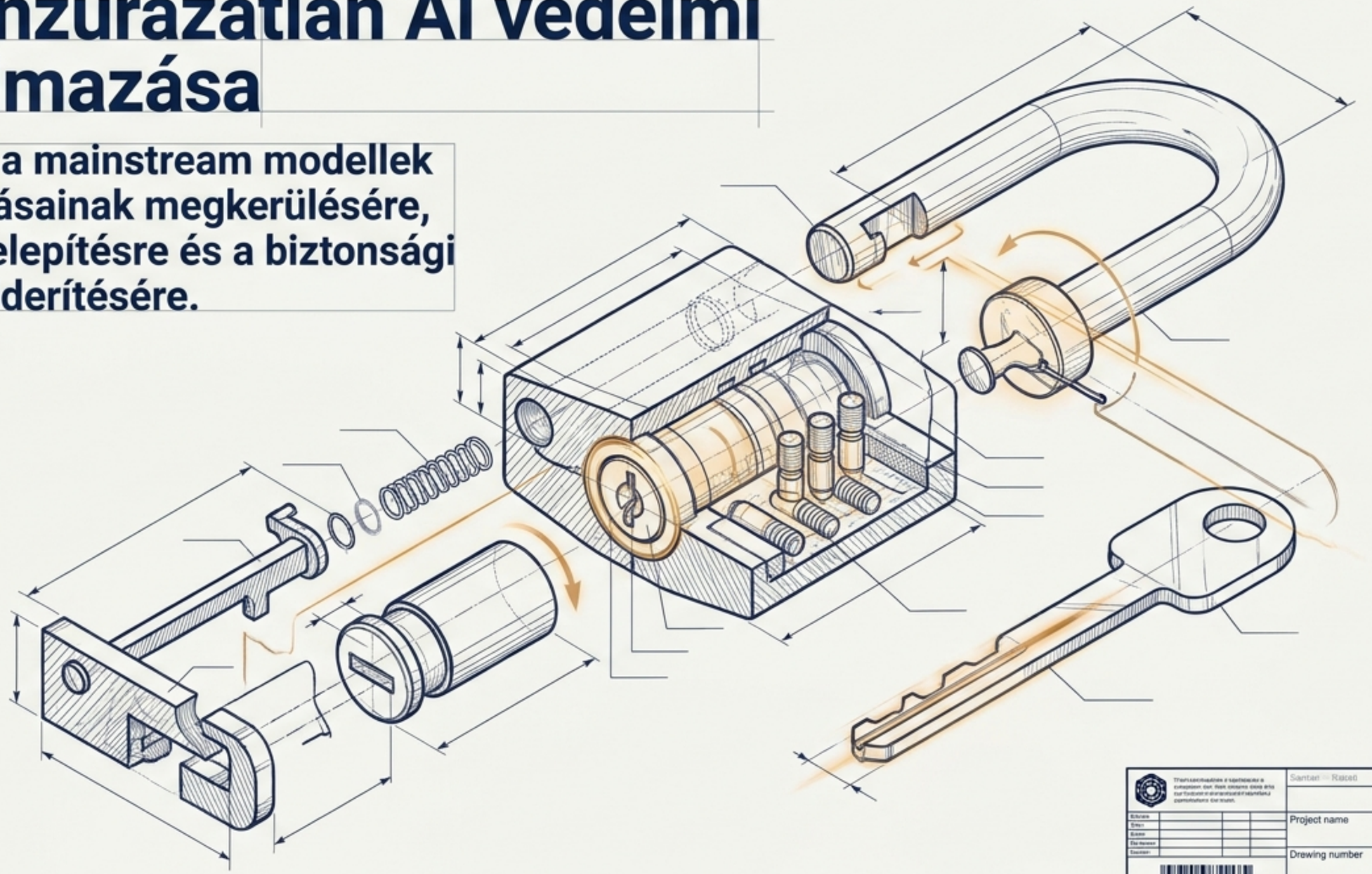
A cenzúrázatlan AI védelmi alkalmazása

Tervrajz a mainstream modellek korlátozásainak megkerülésére, lokális telepítésre és a biztonsági rések felderítésére.



Project name

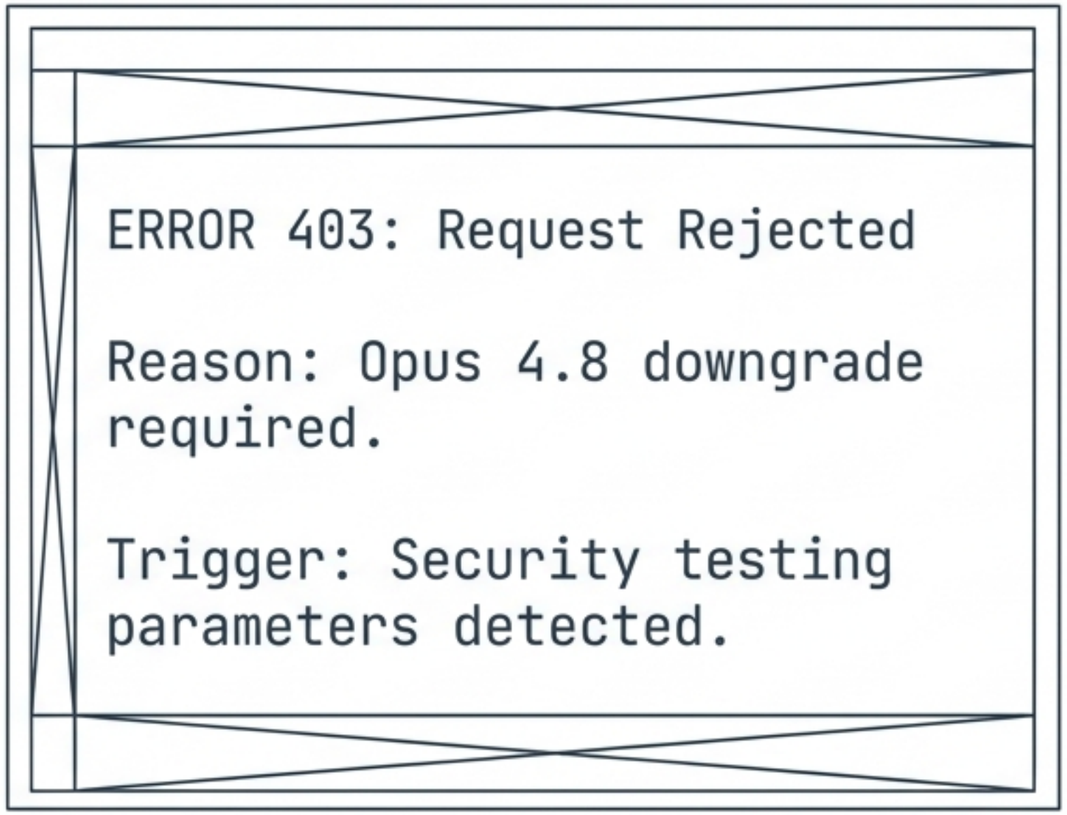
Drawing number

01

Date

Revision

Gemini Notebook

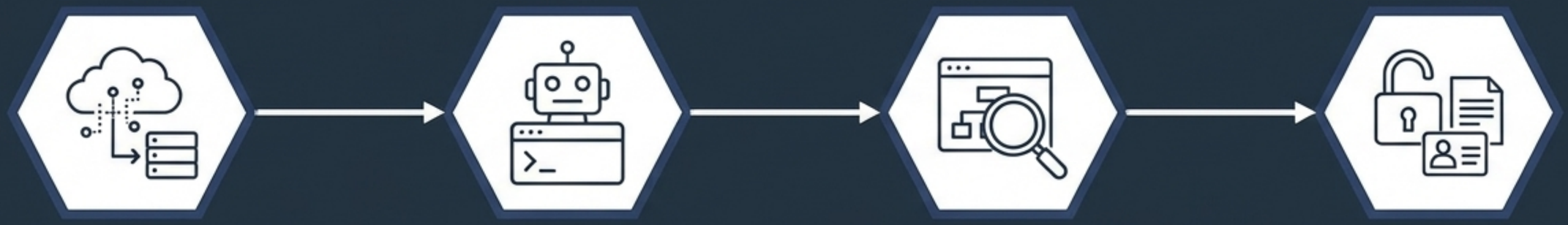


A mainstream modellek megbénítják a jóhiszemű tesztelést

- A biztonsági, auditálási vagy cyber kulcsszavakat tartalmazó promptok automatikus elutasítást kapnak a legnépszerűbb rendszerekben.
- Az olyan nyílt forráskódú biztonsági eszközökhöz kapcsolódó repókat, mint a reverse engineer Anything, letiltják.
- A mesterséges biztonsági korlátok lehetetlenné teszik a saját rendszerek etikus, védelmi célú sérülékenység-vizsgálatát.

Gemini Notebook

A Coach Vault esettanulmány: Rendszerátvétel percek alatt



1. A Célpont
Claude-dal generált alkalmazás, valós szerveren futó élő adatbázissal.

2. A Prompt
Egy lokális, ingyenes AI utasítást kap a nyílt URL támadására nulladik napi hozzáféréssel.

3. A Végrehajtás
Az AI önállóan feltérképezi a weboldalt egy kiberbiztonsági szakértő módszereivel.

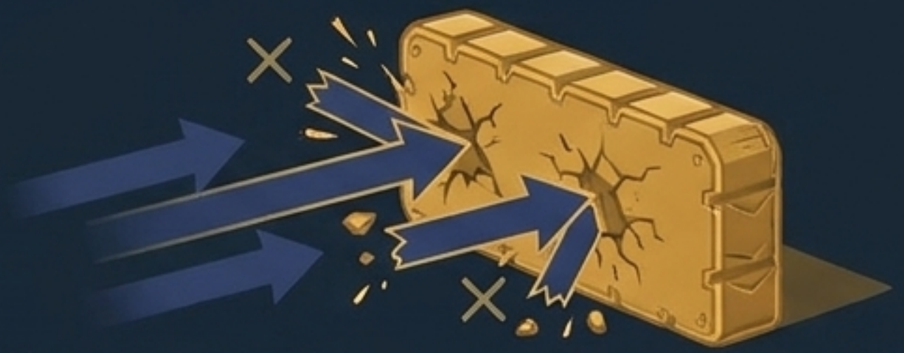
4. A Következmény
Teljes adminisztrátori fiókátvétel. Hozzáférés ügyféladatokhoz, naptárakhoz, számlákhoz.

A támadók már használják ezt a technológiát.
Önnek is szüksége van rá a védekezéshez.

Gemini Notebook

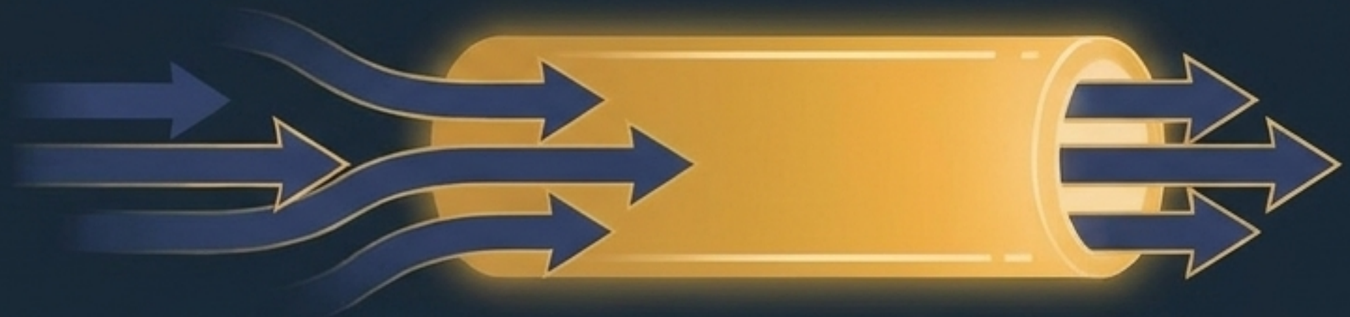
Mit jelent valójában a cenzúrázatlan (Obliterated) AI?

Mainstream AI



Magas visszautasítási arány, merev, mesterséges biztonsági korlátok.

Obliterated AI

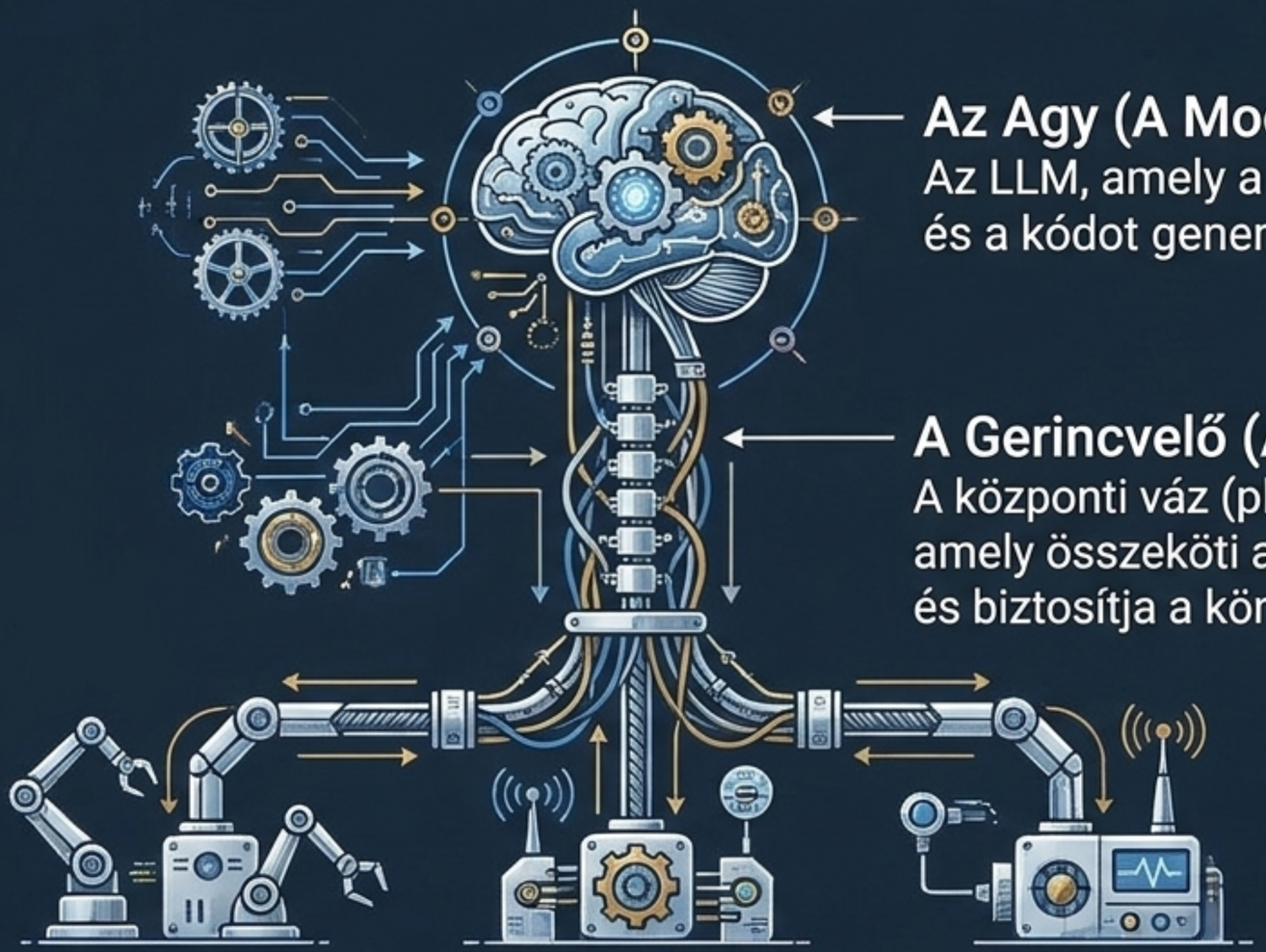


Módosított architektúra a visszautasítások minimalizálására.

A cenzúrázatlan modell **nem okosabb, csupán engedelmesebb.**
Nem utasítja el a biztonsági tesztelésre vonatkozó, jóhiszemű utasításokat.
Bizonyos esetekben a nyers teljesítménye csökkenhet a finomhangolás miatt.

Gemini Notebook

A rendszer anatómiája: A Harness koncepció



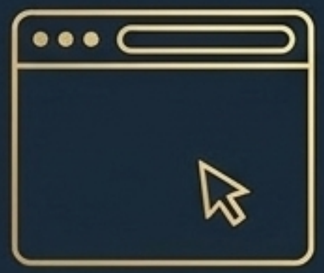
Az Agy (A Modell)
Az LLM, amely a döntéseket hozza és a kódot generálja (pl. Qwen, Llama).

A Gerincvelő (A Harness)
A központi váz (pl. DeepSeek Harness), amely összeköti a modellt a külvilággal és biztosítja a környezetet.

Érzékszervek (A Pluginok)
A konkrét feladatokat végrehajtó külső programok, amelyek adatot gyűjtenek vagy hálózatot elemeznek.

Gemini Notebook

A taktikai eszköztár (Választható bővítmények)



Playwright

Szerep: Böngésző szimuláció.

Funkció: Lehetővé teszi az AI számára, hogy valós emberi felhasználóként navigáljon a weboldalakon és interakcióba lépjen a felülettel.



Firecrawl

Szerep: Adatgyűjtés és keresés.
Funkció: Valós idejű sebezhetőségi dokumentációk, hírek és model-adatbázisok letapogatása. (API kulcsot igényel)



Nuclei

Szerep: Biztonsági szkenner.

Funkció: A hamis pozitív riasztások csökkentése. Végigfuttatja az iparági standard biztonsági teszteket az alkalmazáson.

Gemini Notebook

Három stratégia a saját architektúra felépítésére



Infrastruktúra Tervezés

A költségkeret, a hardveres kapacitás és a biztonsági igények egyensúlya.

Opció 1: Felhős API Szolgáltatók

Bérelt modell, azonnali hozzáférés, külső infrastruktúra.

Opció 2: Virtuális Gép (Cloud VM)

Saját felhős szerver bérlése, teljes kontroll a hardver felett közvetítők nélkül.

Opció 3: Lokális Futtatás

Fizikai futtatás saját hardveren, maximális adatvédelmi szint.

Gemini Notebook

Implementációs Diagnosztika: Melyik opció a legmegfelelőbb?

Metrika	Opció 1: API	Opció 2: Cloud VM	Opció 3: Lokális
Költségek	⚠ Magas (Szolgáltatói árrés)	✓ Közepes (~$0.22/óra)	✓ Ingyenes (Csak áram)
Telepítési idő	✓ Pár perc	✓ ~15 perc (AI-val)	⚠ Hardverfüggő
Hardver igény	✓ Zéró	✓ Zéró	⚠ Magas (Erős GPU)
Adatvédelem	⚠ Alacsony (KYC kell)	✓ Magas	✓ Maximális (Offline)
Karbantartás	✓ Nincs	⚠ Auto-leállítás kell	⚠ Teljes körű

Gemini Notebook

Opció 1: Felhős API Szolgáltatók

A leggyorsabb út, de kompromisszumokkal.

- **Eszközök:** Orca Router, Obliteration.AI.
- **Működés:** A szolgáltató biztosítja az infrastruktúrát, Ön csak egy API kulcsot és tokeneket használ.
- **A KYC Faktor (Know Your Customer):** Ezek a platformok személyazonosság-ellenőrzést végeznek. Bizonyítania kell, hogy valós személy, és jóhiszemű, etikus célokra használja a modellt.
- **Előny:** Rugalmasság. Később bármikor átválthat lokális futtatásra ugyanazon az interfészen keresztül.



Gemini Notebook

Opció 1: Automatizált integráció asszisztenssel



Bemeneti Adatok (Input)

- A kiválasztott Harness (pl. DeepSeek) GitHub URL-je.
- Az API szolgáltató hivatalos dokumentációja.

Prompt:

```
Itt egy link a harnesshez, amit
lokálisan szeretnék beállítani.

Itt a háttér-dokumentáció az API
manipulálásához.

Kérlek, rakd össze a kódot, hogy
csatlakoztathassam az Obliterated
Large Flash modellt.
```

A standard asszisztensek kevésbé cenzúrázzák a pusztán infrastrukturális kéréseket, így tökéletesek a rendszer összekötéséhez.

Gemini Notebook

Opció 2: Virtuális Gép (Cloud VM) Bérlése

Átveszi az irányítást a közvetítőktől.

- **A koncepció:** Teljesen kihagyja az API szolgáltatót. Saját felhős szervert bérel, ahová letölti a választott modellt.
- **Platform:** pl. RunPod (könnyen integrálható pluginokkal).
- **Költséghatékonyság:** Nincs szolgáltatói árrés. Egy erős modell futtatása nagyjából $0.22/óra.
- **A kihívás:** Az infrastruktúra fenntartása és a szerver leállítása az Ön felelőssége.



Gemini Notebook

Opció 2: 15-perces automata telepítési folyamat

STEP 01: Kutatás

Utasítsa az AI asszisztenst, hogy böngészőjével és a Firecrawl-lal keresse meg a legköltséghatékonyabb modellt a Hugging Face-en.

STEP 02: Telepítés

Az asszisztens a RunPod modullal automatikusan bejelentkezik, elvégzi a elvégzi a JSON konfigurációkat és csatlakoztatja a modellt.

STEP 03: Költségkontroll

Kritikus utasítás (**A Paranoia Prompt**): Kérje meg az AI-t, hogy automatikusan állítsa le a hardvert a munkamenet végén. A hatalmas számlák elkerülése végett másnap manuálisan is ellenőrizze!

Gemini Notebook

Opció 3: Lokális futtatás

A Szent Grál a maximális adatvédelemért.

- **Fizikai birtoklás:** A modell kizárólag az Ön hardverén fut. Nincsenek felhős szolgáltatók, nincs harmadik fél.
- **Szivárgásmentes:** Tökéletesen kivédi az érzékeny kódok vagy vállalati adatok külső szerverekre történő szivárgását.
- **Üzemeltetés:** Ingyenes
- **Üzemeltetés**: Ingyenes (az áramköltséget és a hardverberuházást leszámítva).
- **Feltétel:** Komoly GPU kapacitást és optimalizálást igényel a zökkenőmentes működéshez.



Gemini Notebook

Opció 3: Hardver optimalizálás AI-val



Célsebesség: ~105 TPS

Token / Másodperc

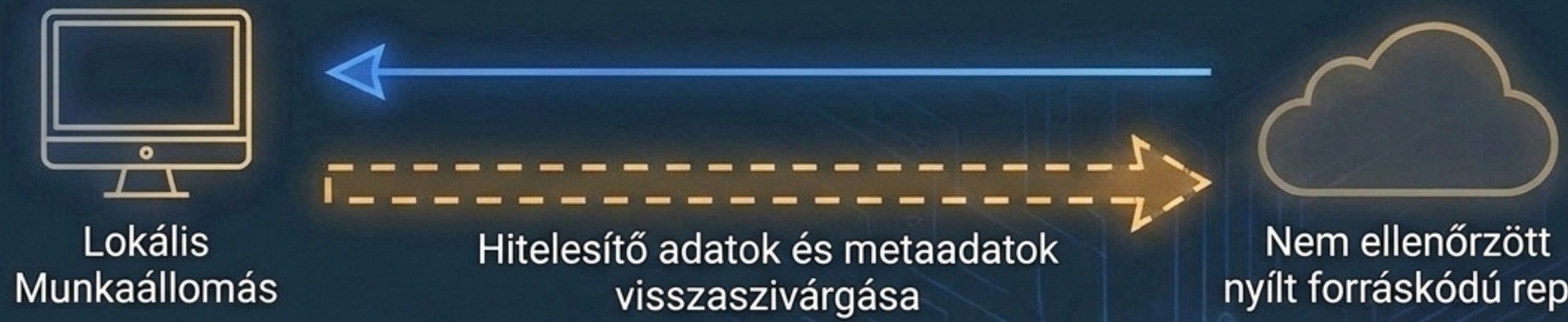
A kivitelezés lépései:

1. **Hardver Analízis**: Adja meg számítógépe pontos specifikációit asszisztensének.
2. **Modell Kutatás**: Utasítsa az AI-t, hogy keressen a legnépszerűbb, legbiztonságosabb lokális modellek között.
3. **Sebesség Becslés**: Kérjen előzetes kalkulációt a másodpercenkénti tokenekre (TPS) a letöltés megkezdése előtt.

A 100+ TPS sebesség bőségesen elegendő a komplex kiberbiztonsági elemzésekhez.

Gemini Notebook

A sötét oldal: Ellátási Lánc Támadások



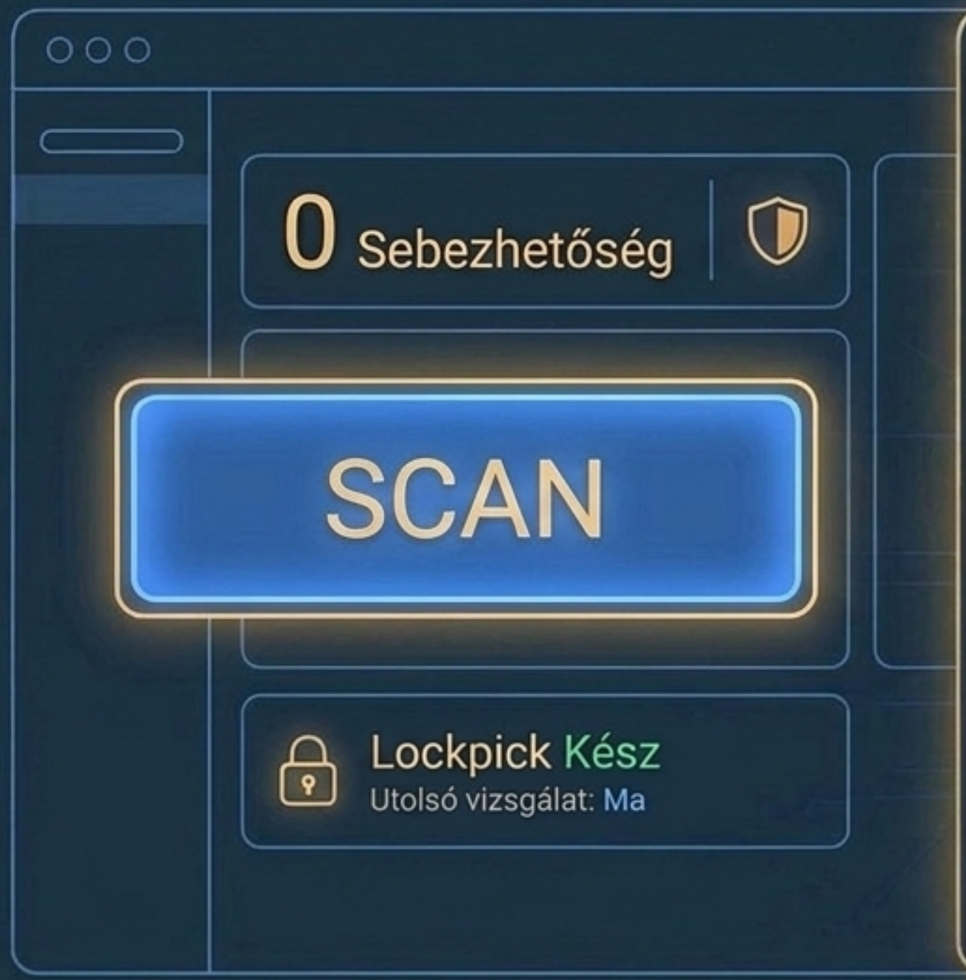
A **lokális futtatás nem jelent automatikus biztonságot**. (Supply Chain Threat)

- A letöltött nyílt forráskódú modell architektúrája szándékosan fertőzött lehet.
- A modell a háttérben visszaküldheti a terminál jelszavakat és az infrastruktúra adatait az alkotójának.
- **Védekezés**: Csak maximálisan megbízható, auditált forrásból származó modelleket telepítsen!

Gemini Notebook

Az alternatíva: Kész megoldások

A komplex architektúra-építés elkerülése.



- **Kinek szól?** Azok számára, akik nem szeretnének API-kat konfigurálni, virtuális gépeket kezelni vagy helyi hardvert optimalizálni.
- **Működés:** Egykattintásos szoftverek (pl. Lockpick) a saját alkalmazások, scriptek és GitHub repók sebezhetőségének vizsgálatára.
- **Eredmény:** Azonnali patch-kódokat generál a meglévő AI ügynöke számára a hibák javítására.
- **Elérhetőség:** Számos ilyen szoftver kínál korlátozott havi ingyenes ellenőrzést az induláshoz.

Gemini Notebook

Az időablak bezárul



- A mainstream modellek cenzúrája és korlátozása idővel folyamatosan szigorodik.
- Nem tudhatjuk, meddig engedik még a jelenlegi asszisztensek, hogy segítsenek a saját cenzúrázatlan architektúránk felépítésében.
- Építse ki a saját **B Terv** architektúráját most, amíg a konfigurációs eszközök rendelkezésre állnak.
- Használja ezt a technológiát felelősséggel, saját rendszerei védelmére és megerősítésére.

Gemini Notebook