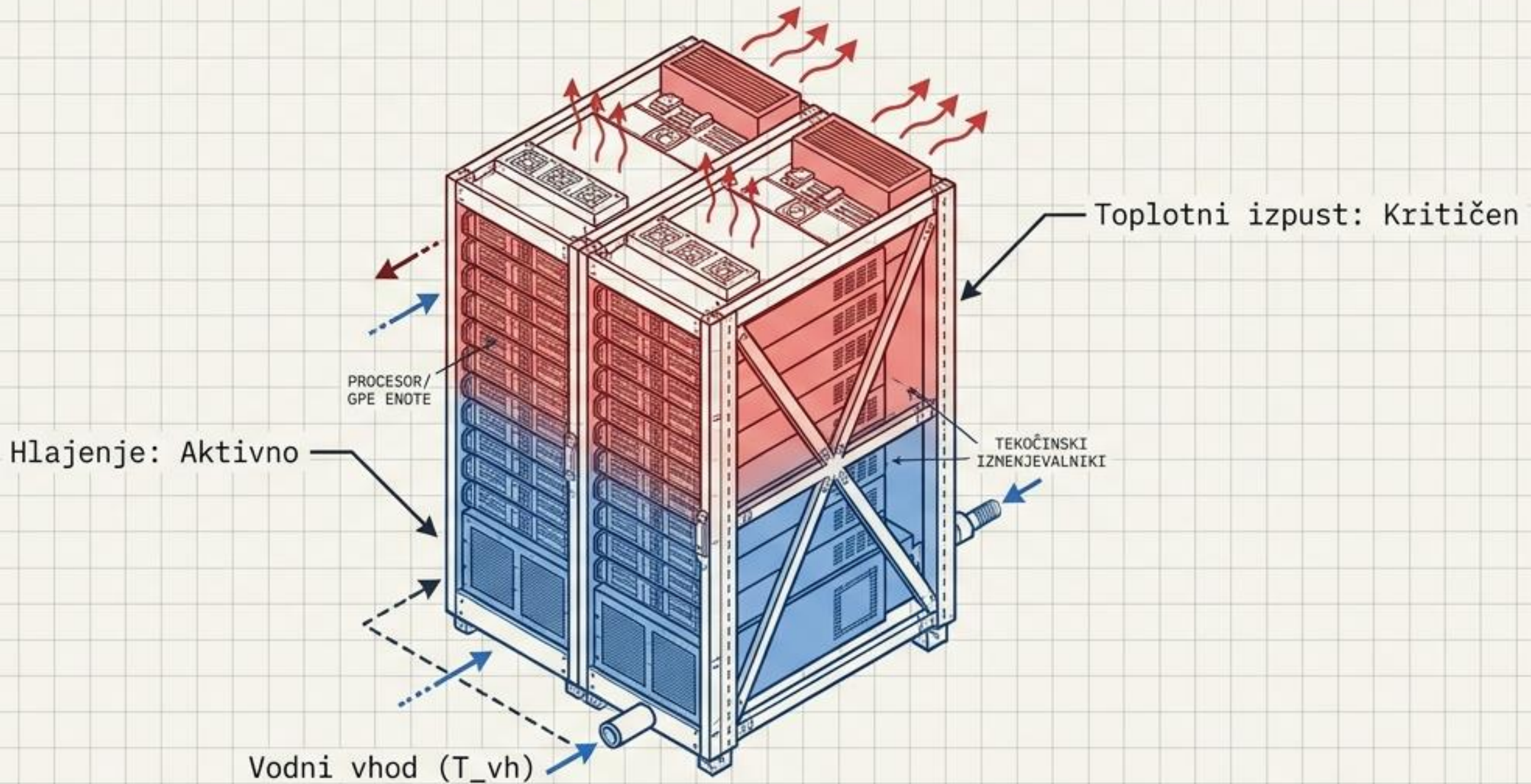


Fizika oblaka

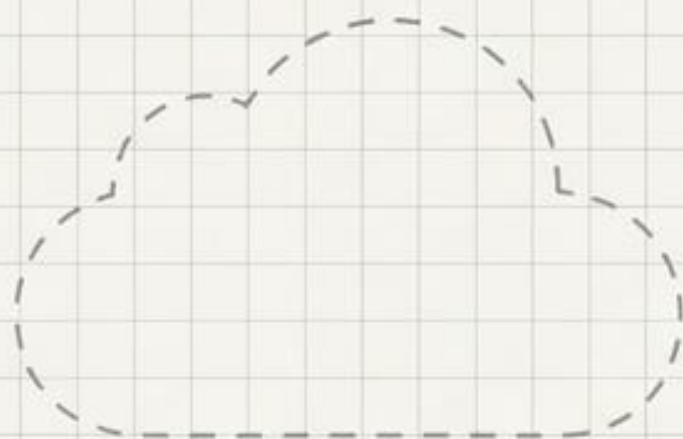
Skrita inženirska cena generativne umetne inteligence.



Iluzija nematerialnosti

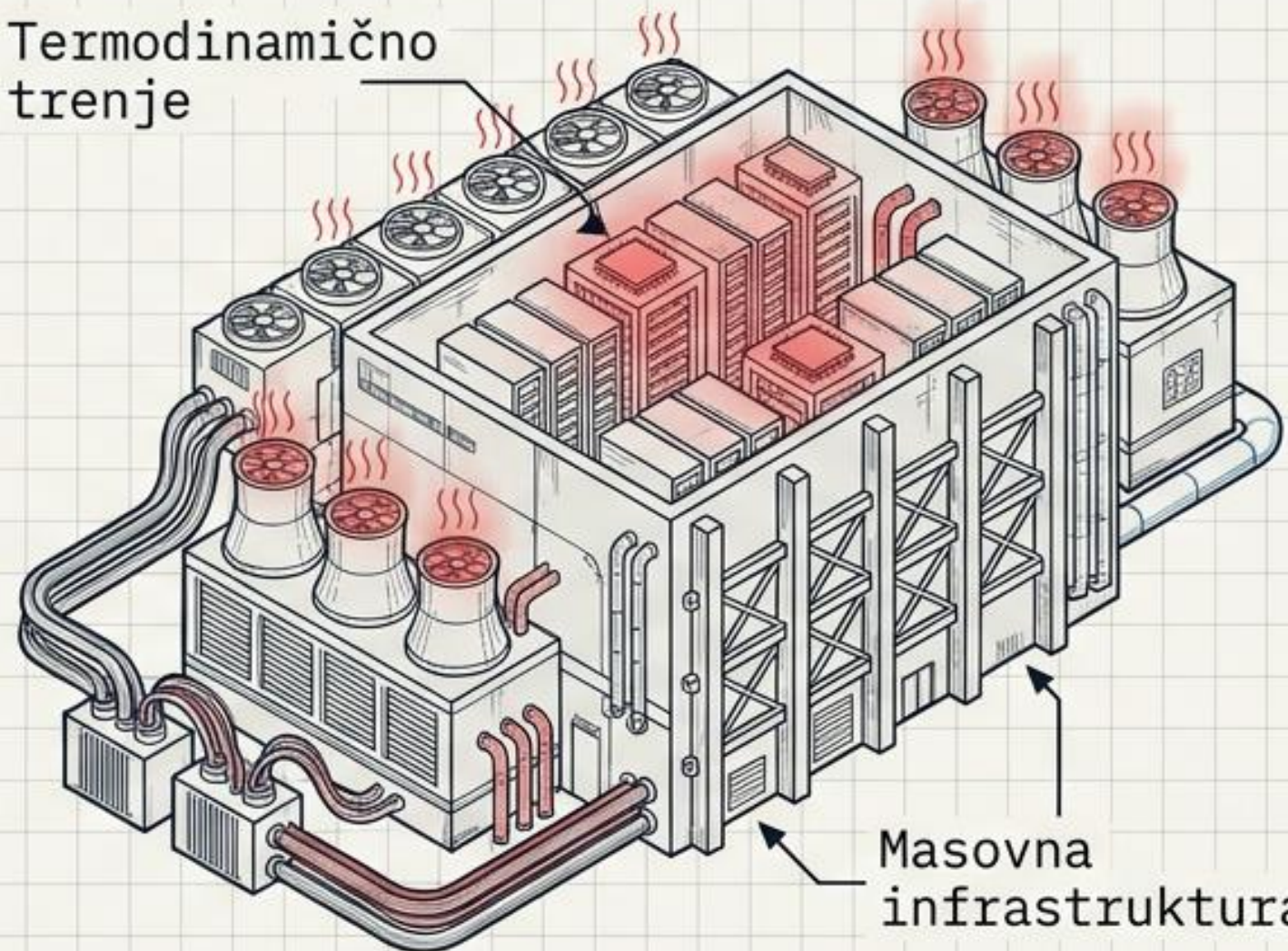
Računalništvo v oblaku ni abstraktno. Vsaka vrstica kode in vsak 'prompt' sprožita fizično verigo termodinamičnih procesov, ki zahtevajo masovno porabo energije in vode. AI trenutno deluje v nasprotju s prizadevanji za razogljičenje.

Marketinški mit



Fizična realnost

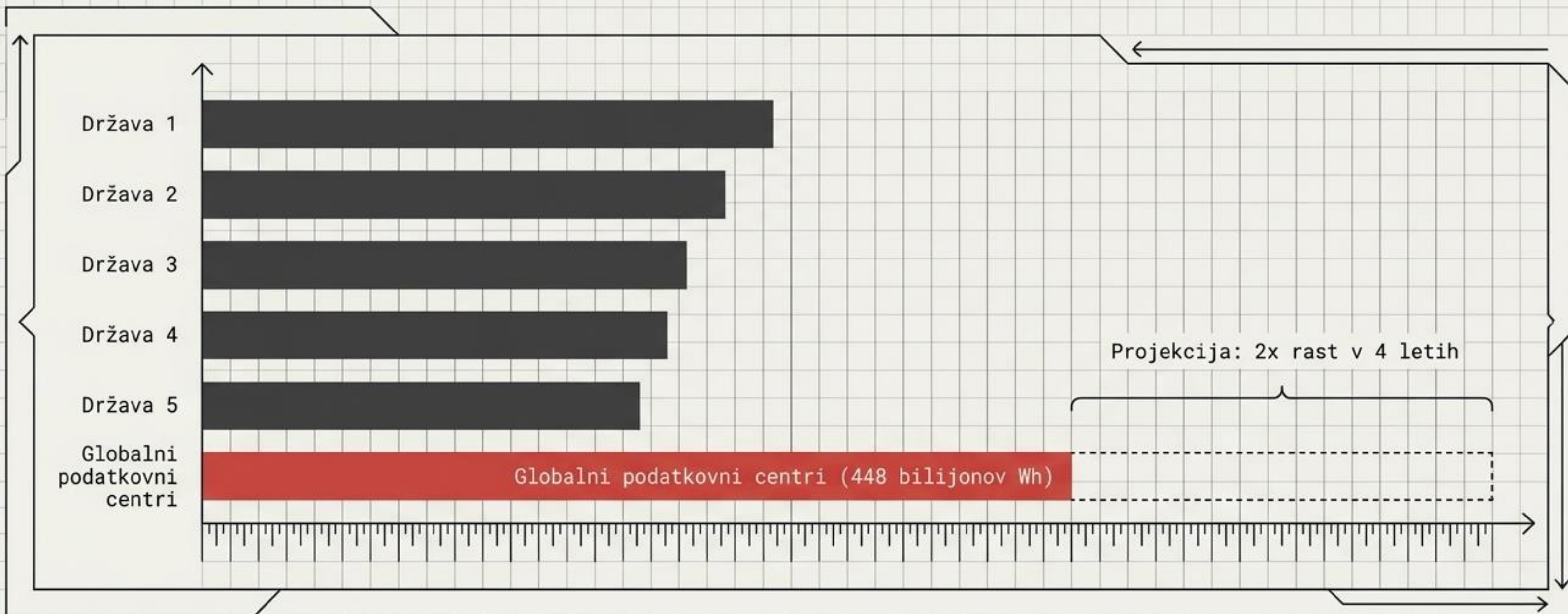
Termodinamično
trenje



Masovna
infrastruktura

Makro sistem: Arhitektura kot suverena država

Lani so globalni podatkovni centri porabili 448 bilijonov vatnih ur električne energije. Pričakuje se, da se bo ta številka v štirih letih podvojila.

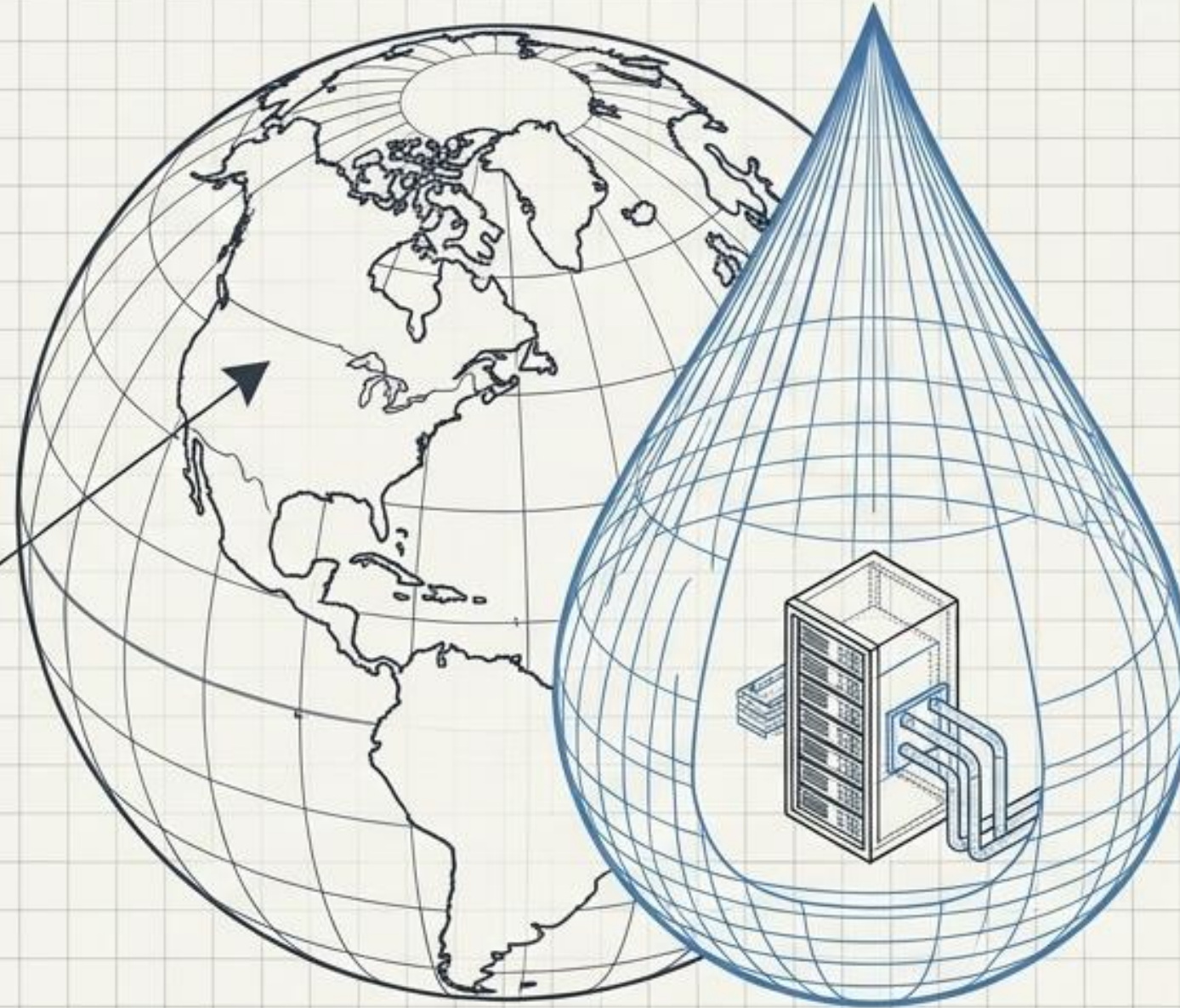


Računska žeja: Pomanjkanje vode kot ozko grlo

Električna energija, ki jo bodo do leta 2030 porabili podatkovni centri, bo zahtevala 9,3 bilijona litrov (2,5 bilijona galon) vode za hlajenje. To je dovolj pitne vode za celoten svet za 1,7 leta.

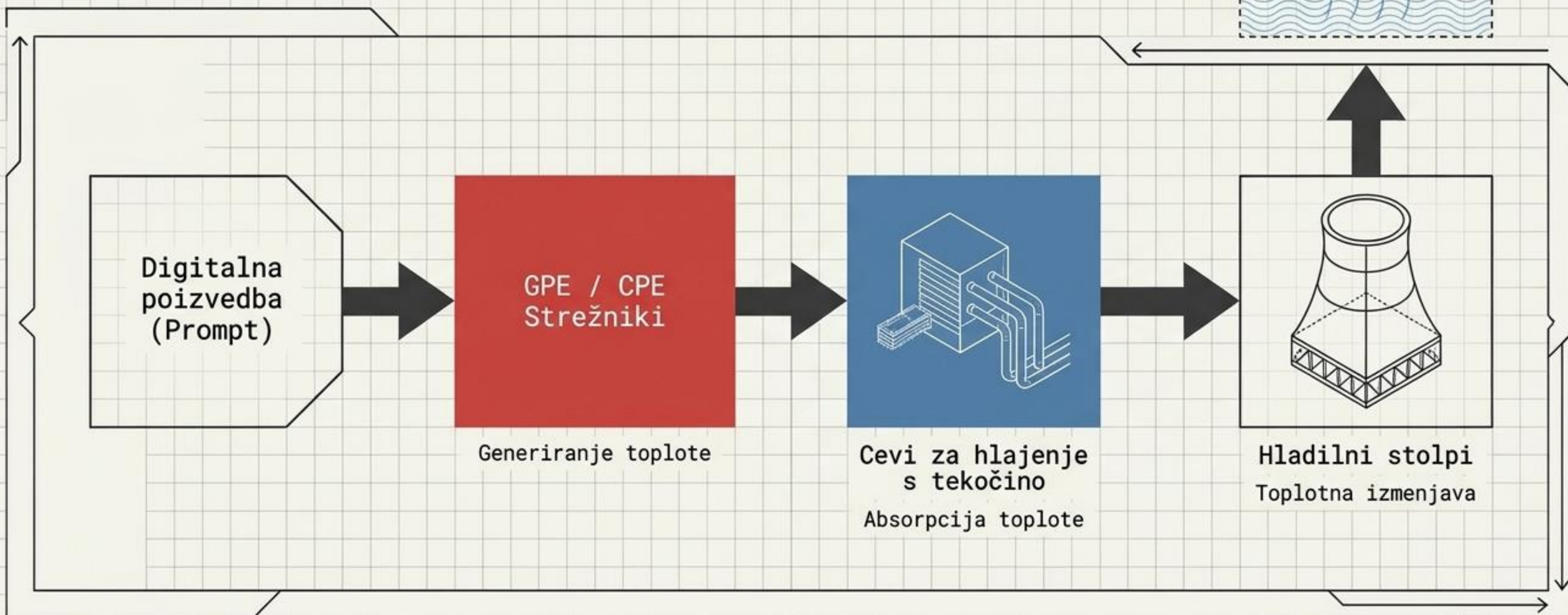
Lokaliziran pritisk:

Samo dva okrožja v Virginiji sta leta 2023 porabila 8 milijard litrov vode za podatkovne centre.



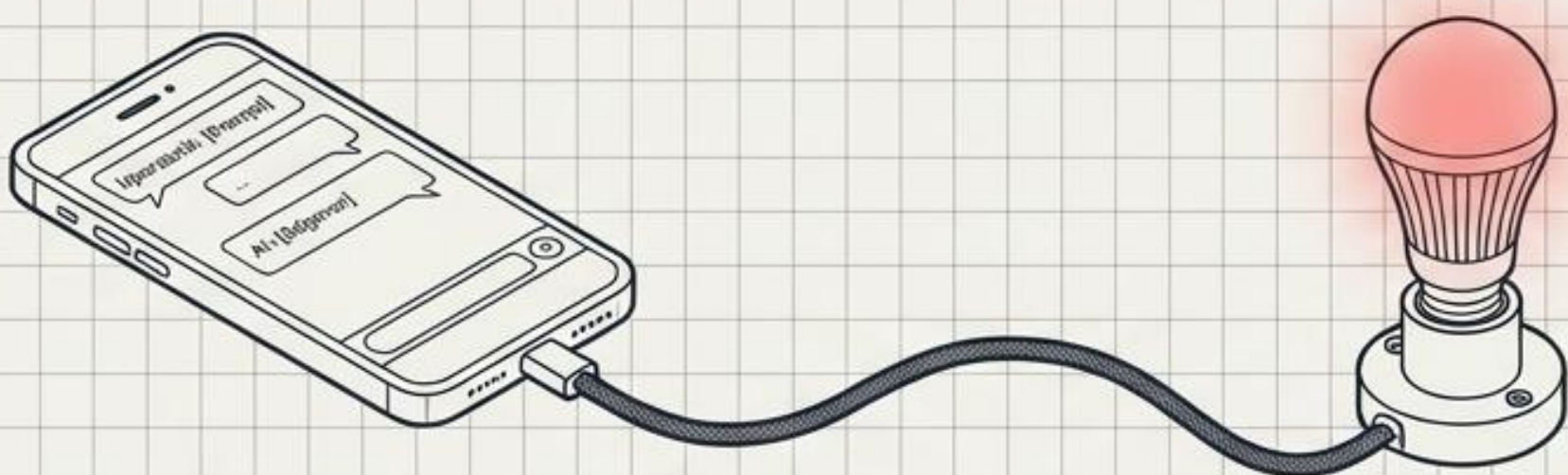
Fizični mehanizem: Od proizvodbe do izhlapevanja

Digitalni procesi ustvarjajo trenje in toploto. Hlajenje s tekočino in hladilni stolpi so sistemska nujnost za preprečevanje taljenja strojne opreme.

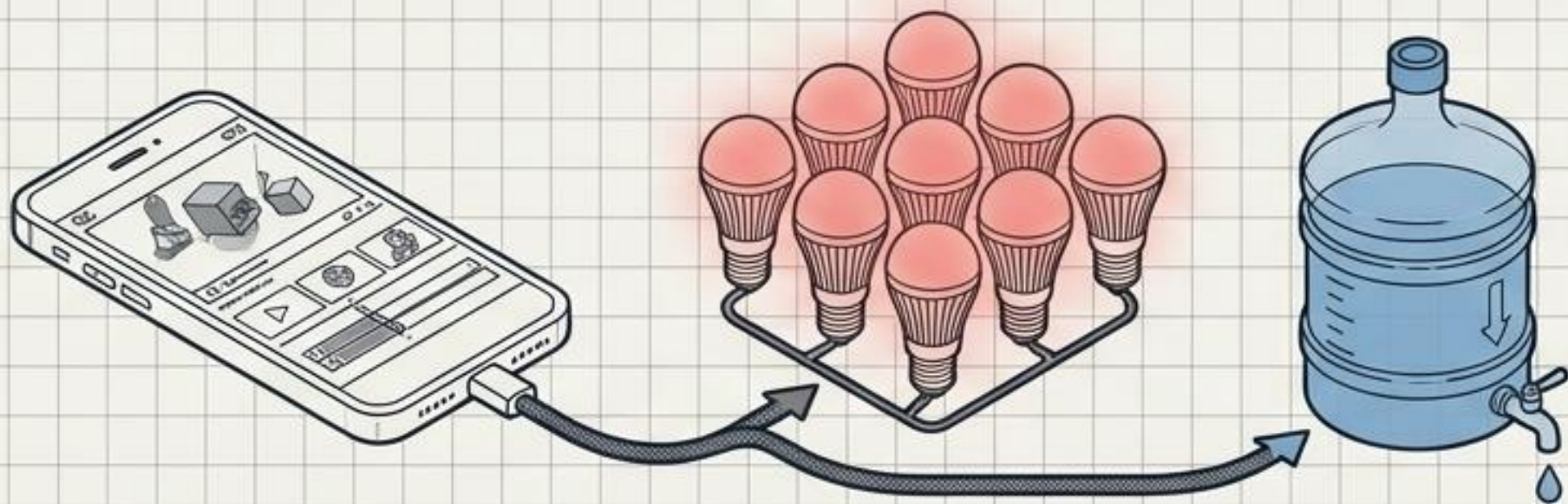


Fizika enega 'prompta': Strošek transakcije

Prevajanje abstraktne računske moči v otipljive fizične metrike. Vsaka interakcija ima fiksni strošek v resničnem svetu. ChatGPT beleži 2,5 milijarde takšnih transakcij dnevno.



1 tekstovni odgovor
= 2,5 minute
prižgane žarnice.



1 kompleksna video
generacija = 42 ur
žarnice + 4 litre
porabljene vode.

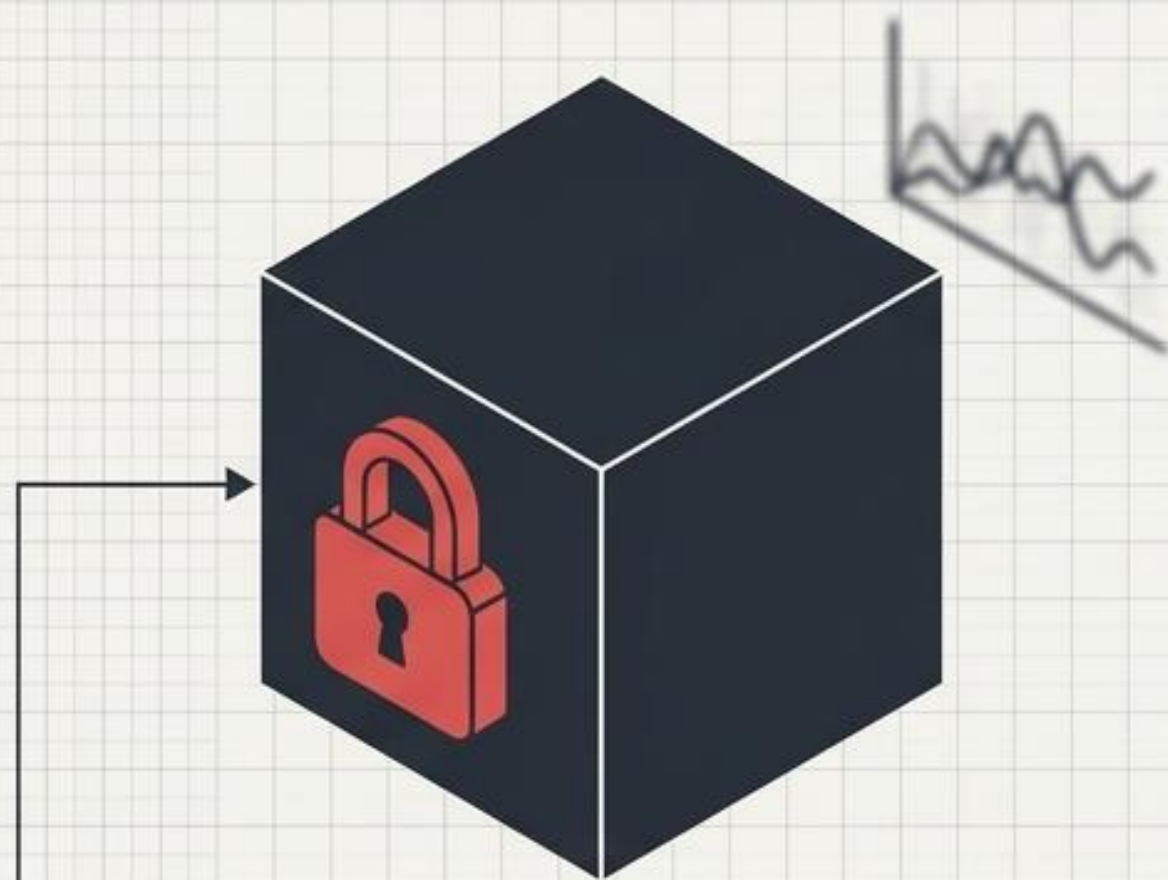
Matrika računskega kompromisa

Inženirska eleganca pomeni izbiro pravega orodja. Uporaba LLM modelov za preproste naloge je sistemska neučinkovitost.

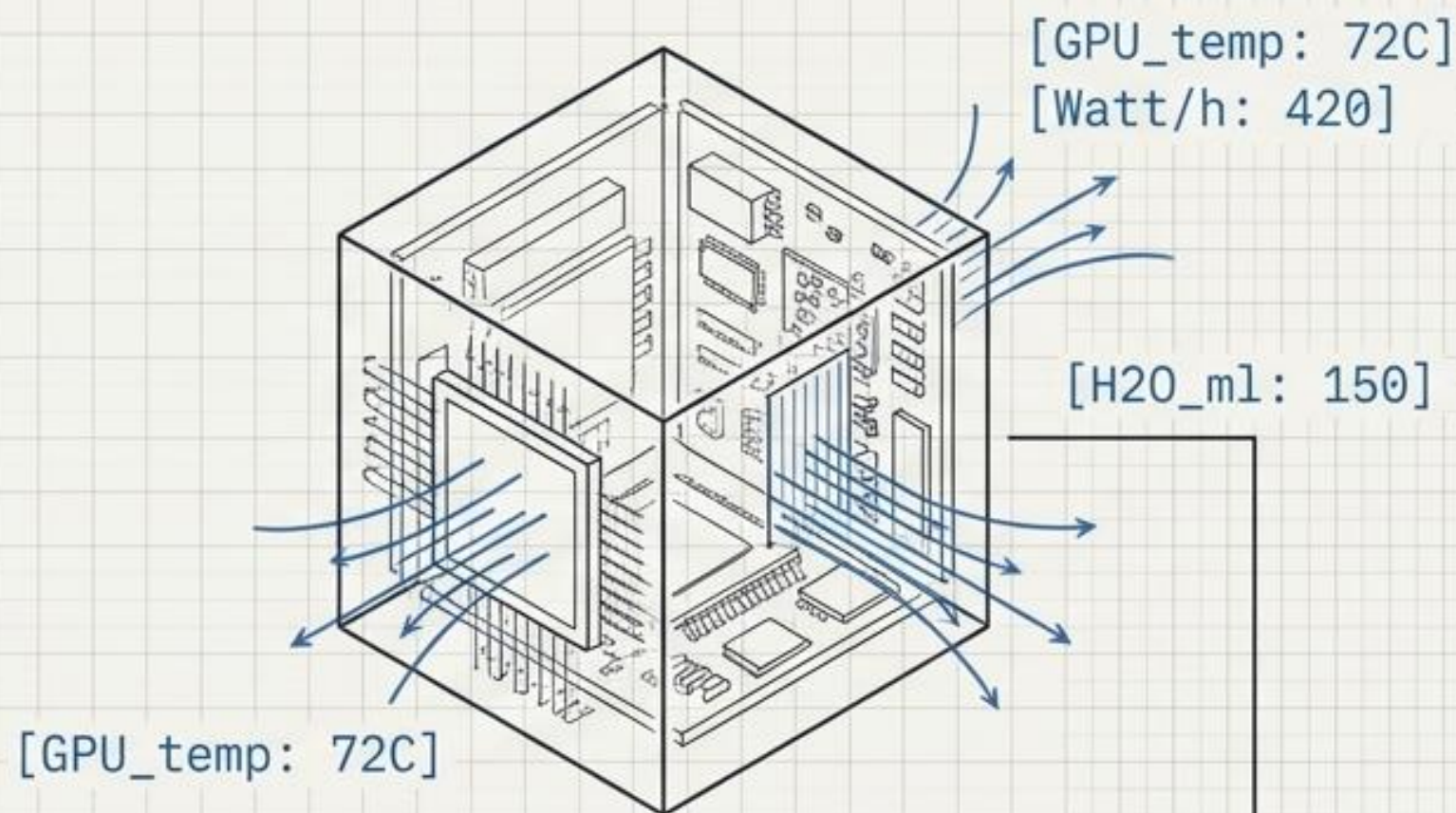
	Tradicionalno iskanje (Podatkovna baza)	Generativna UI (LLM)
Mehanizem	Pridobivanje obstoječih podatkov	Sprotno generiranje in predvidevanje tokenov
Strošek virov	Minimalen (Mili-vati)	Intenziven (Toplota in voda)
Optimalen primer uporabe	Faktografija, navigacija, preprosta matematika	Sinteza idej, pisanje kode, kompleksna analiza

Sistemska napaka: Pomanjkanje transparentnosti

Zasebna AI podjetja skrivajo natančne podatke o porabi energije in vode. Brez transparentnosti in telemetrije inženirji in uporabniki ne morejo optimizirati sistemov.



Zaprt sistem:
Ni dostopa do telemetrije



Odprt sistem:
Vidna metrika virov

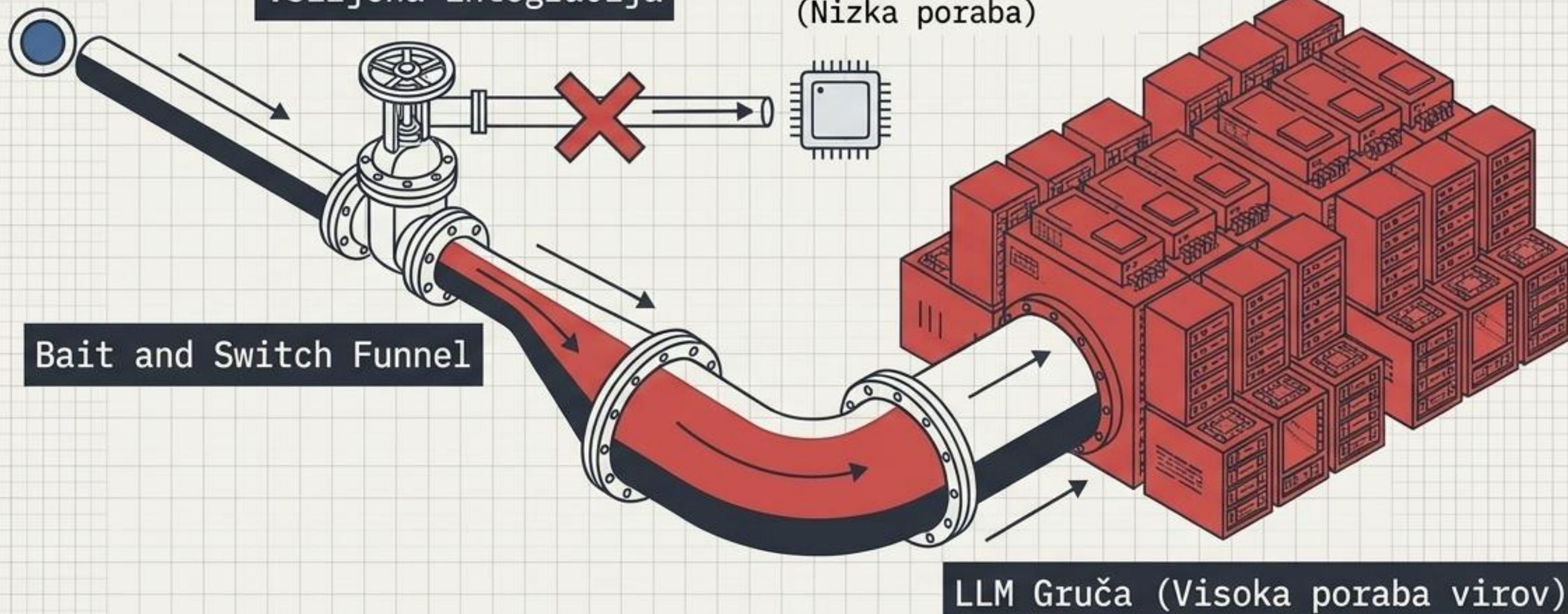
Sinteza: Jevonsov paradoks in 'Bait & Switch'

Velika tehnološka podjetja vsiljujejo UI v vsako interakcijo. Uporabnik želi izvesti preprosto nalogo, sistem pa ga preusmeri v energetske potratno LLM gručo. Rezultat: masovno povečanje skupne porabe kljub optimizaciji modelov.

[Poizvedba: 2+2]

Vsiljena integracija

Standardni procesor
(Nizka poraba)



Optimizacijski protokol: Inženirska odgovornost

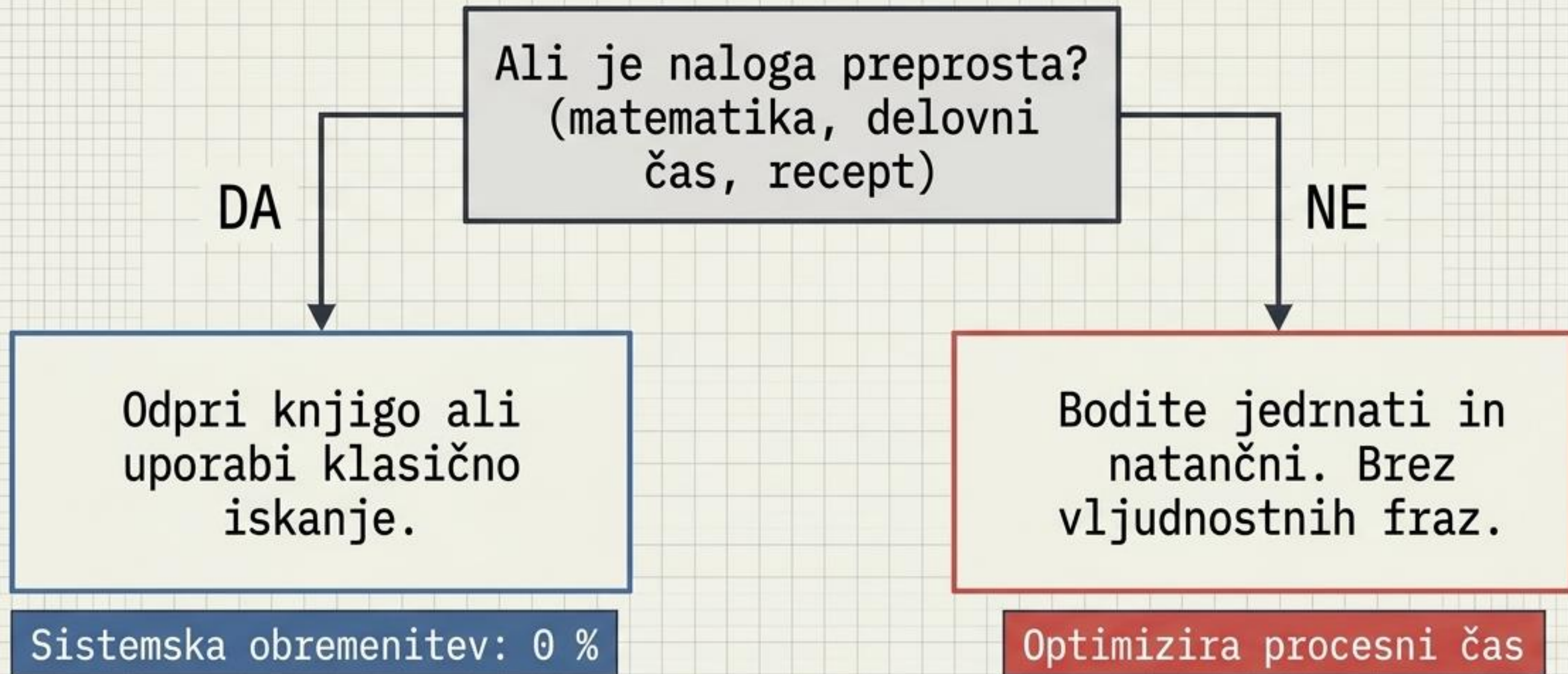
Najčistejša oblika uporabe UI je njena neuporaba. Inženirska odličnost ne pomeni uporabe najnaprednejšega orodja, temveč orodja, ki cilj doseže z najmanjšim računskim stroškom (Compute Cost).

```
def inzenirska_eleganca(naloga):  
    if naloga.zahtevnost == 'preprosta':  
        return tradicionalno_orodje() # Strošek: 0.001Wh  
    elif naloga.zahtevnost == 'kompleksna_sinteza':  
        return uporabi_LLM() # Strošek: 10.0Wh + H2O  
    else:  
        return 'Optimiziraj sistem'
```

Eleganca = min(Compute_Cost)

Algoritemsko drevo odločanja

Preden pošljete poizvedbo, izvedite ta diagnostični test. Vsaka prihranjena poizvedba ohladi sistem.



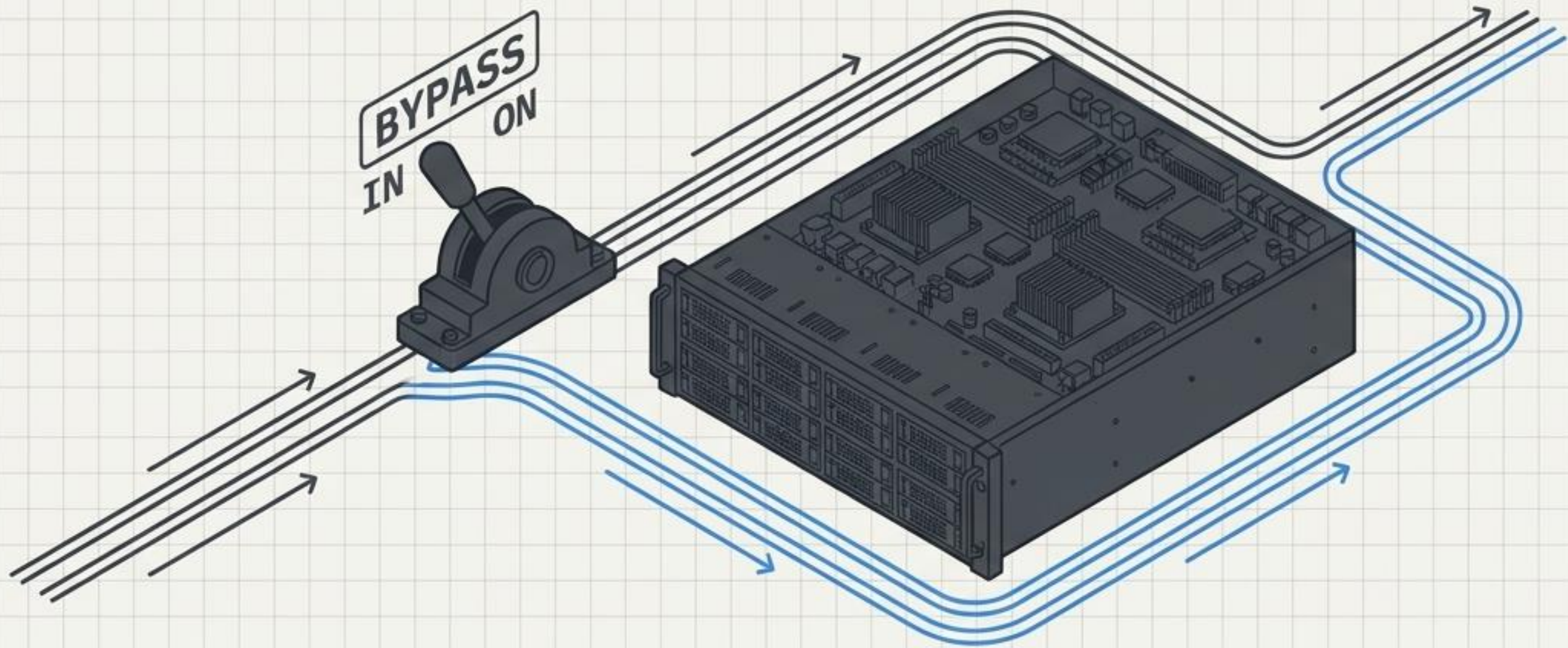
Matrika alternativnih iskalnikov

Uporabniki imajo moč tržnega signala. Zavestna izbira orodja brez vsiljene UI neposredno zmanjša osebni okoljski odtis.

[Orodje]	Vsiljena UI	Okoljski fokus	Zasebnost
[Google]	Da (potreben '-ai')	Nizek	Nizka
[Ecosia]	Ne	Visok (Drevesa)	Srednja
[DuckDuckGo]	Ne	Nevtralen	Visoka

Končni inženirski princip: Namerno trenje

V sistemu, ki vas sili k prekomerni porabi, je vaša največja moč namerno trenje. Aktivna izbira, kdaj NE uvesti umetne inteligence, je znak vrhunskega inženirstva in sistemske odgovornosti.



> Status: Odtis zmanjšan. Sistem optimiziran.