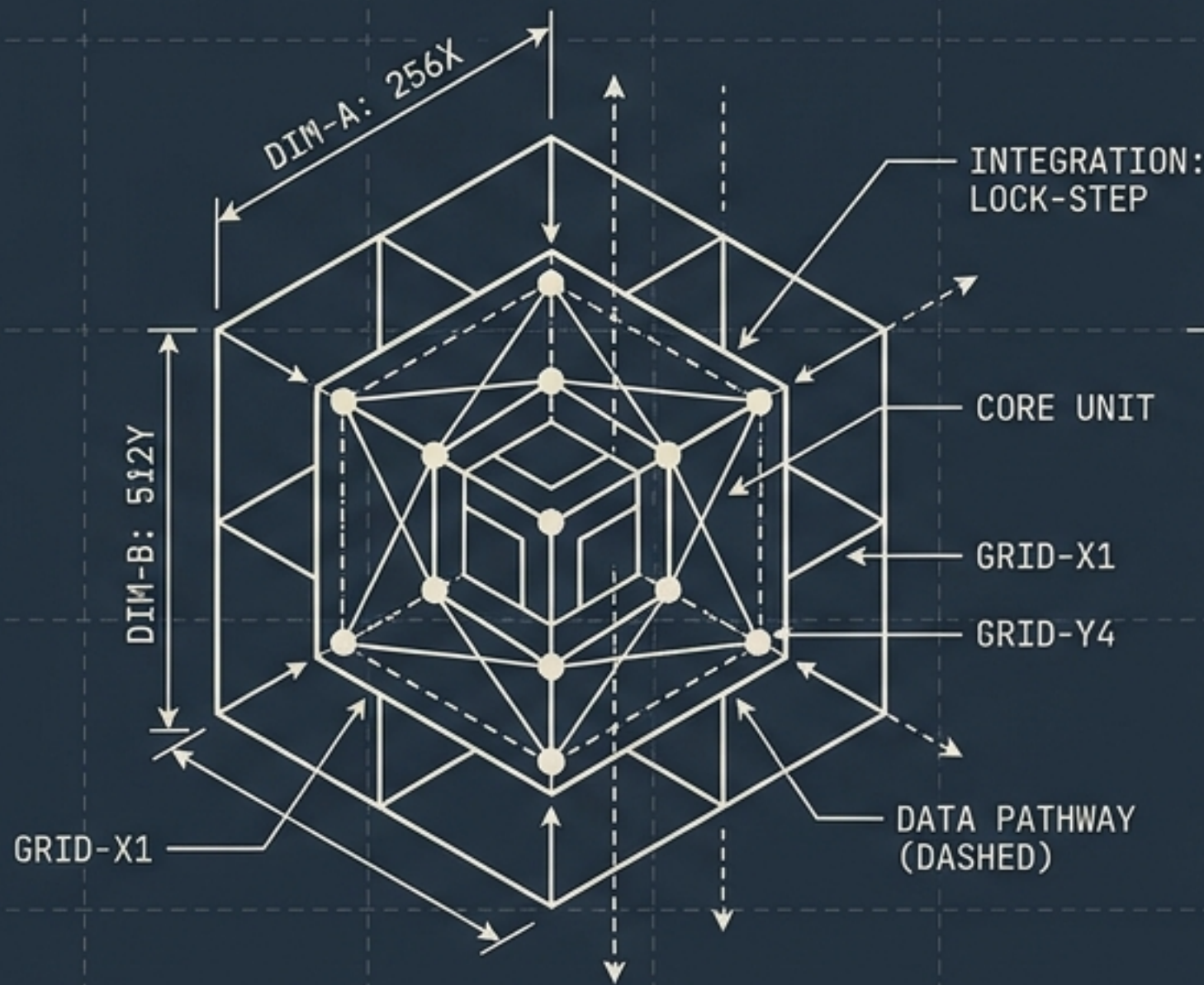


ARHITEKTURA SKLEPANJA: TEHNIČNA SPECIFIKACIJA MODELA CLAUDE OPUS 5

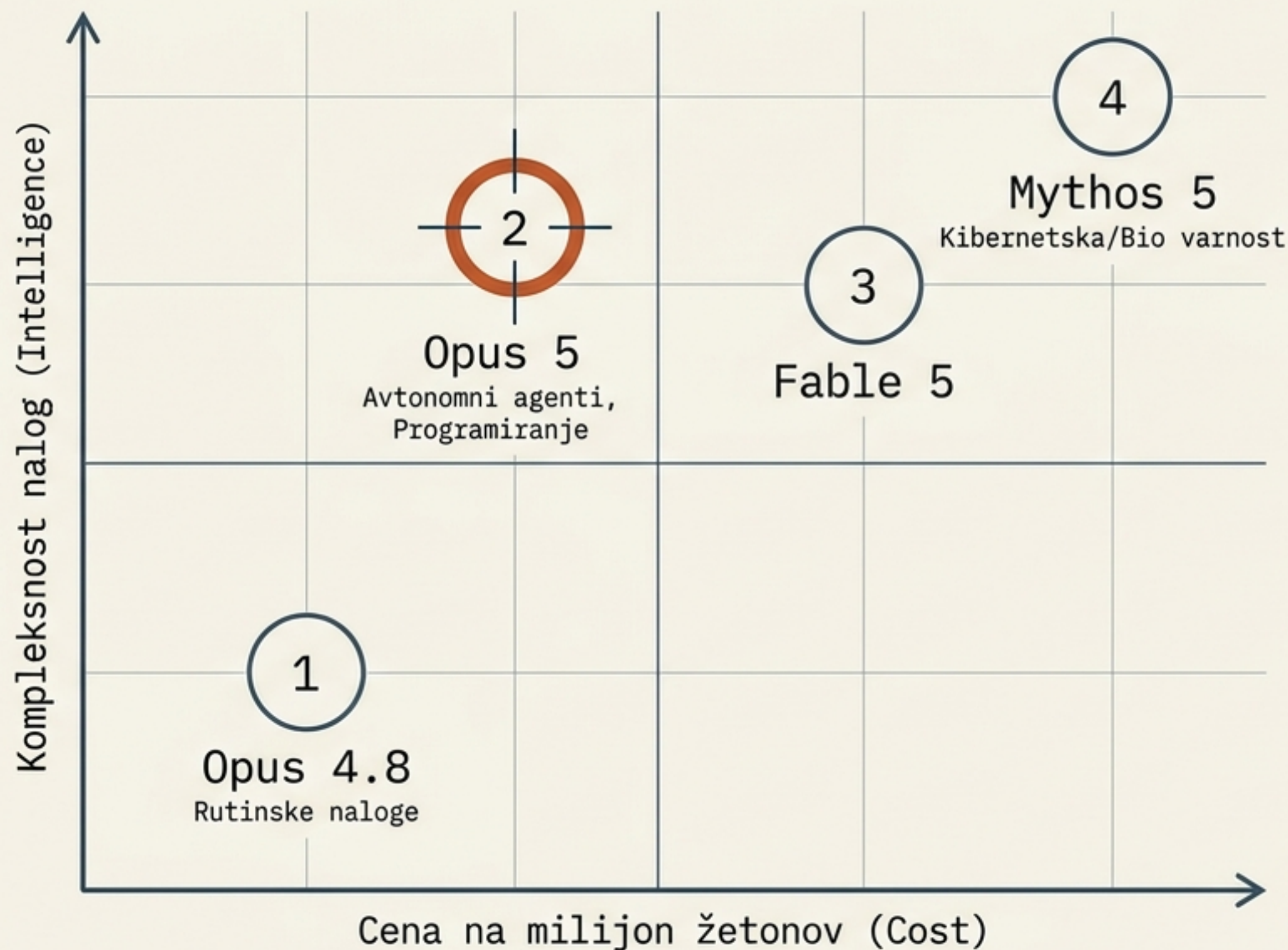
Od osnovnih zmogljivosti do avtonomnih agentnih sistemov.
Tehnični pregled za inženirje.



[STATUS: DEPLOYED] | [MODEL: OPUS-5] | [TARGET: ENGINEERING & TECH]

Optimizacija med stroški in inteligenco

Opus 5 prinaša inteligenco razreda Fable ob stroškovni učinkovitosti predhodnika.



[SISTEMSKA PREDNOST]

Inteligenca blizu meje modela Fable 5, vendar za polovico cene, z latenco optimizirano za vsakodnevno inženirsko rabo.

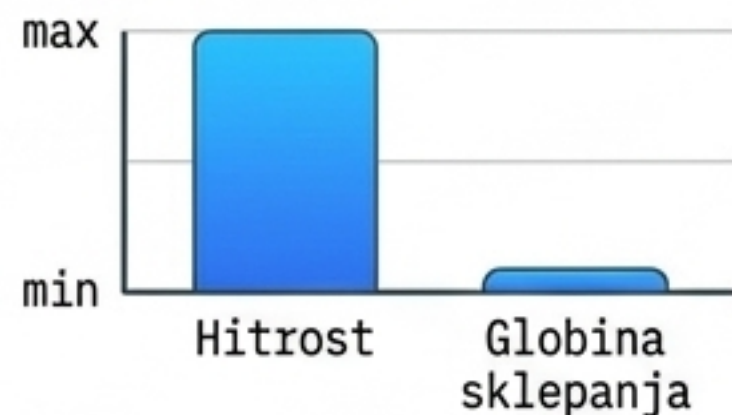
CENA: \$5 IN / \$25 OUT

Dinamično uravnavanje računske moči

API parameter za prilagajanje vloženega napora (Effort Setting) omogoča inženirjem nadzor nad porabo žetonov in latenco.

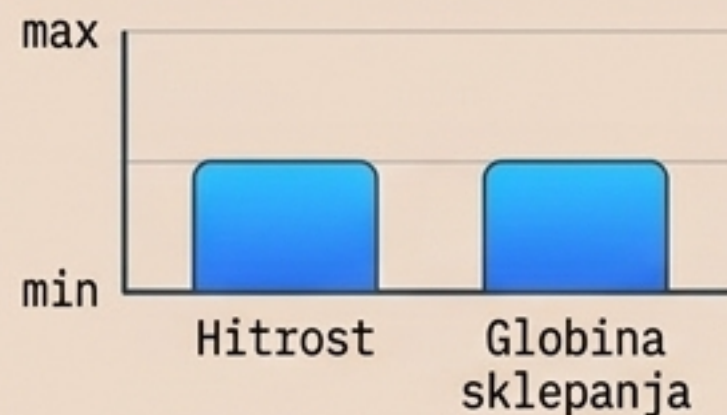


Nastavitev: NIZKA (Low)



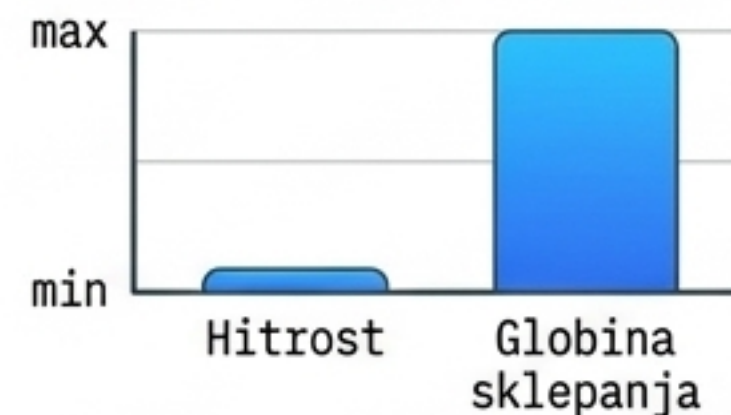
Optimizirano za hitrost in varčevanje z žetoni. Kljub temu prekaša vse druge modele na Zapier AutomationBench.

Nastavitev: SREDNJA (Default)



Optimalno razmerje za večino programerskih nalog in vsakodnevno inženirsko rabo.

Nastavitev: MAX (Maximum)



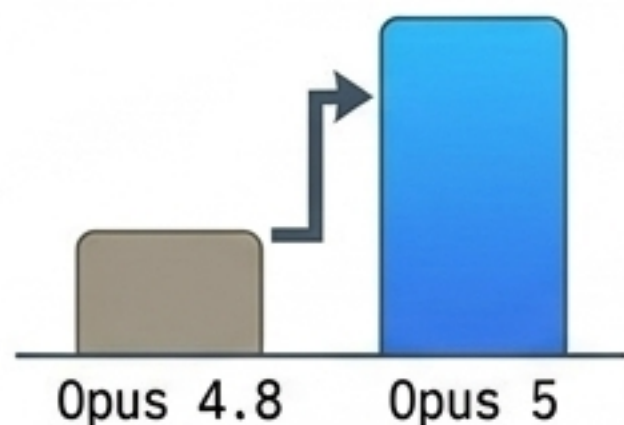
Na CursorBench 3.2 doseže rezultat znotraj 0,5 % od absolutnega vrha modela Fable 5, a ob polovični ceni na nalogo.

Opomba: Inženirji lahko dinamično menjajo nastavitve glede na zahteve po latenci v produkciji.

Matrika zmogljivosti: Novi standardi

Opus 5 ne izboljša le ene metrike, temveč postavlja nov SOTA (State-of-the-Art) na ključnih inženirskih domenah.

Koda (Frontier-Bench v0.1)



>2x izboljšava

Prekaša vse ostale modele v primerjavi z Opus 4.8, z nižjimi stroški na nalogo.

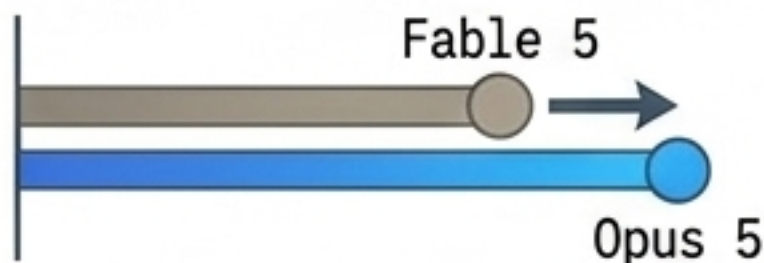
Agencija (Zapier AutomationBench)



1.5x višja uspešnost

Višja od naslednjega najboljšega modela ob popolno-ma enakih stroških.

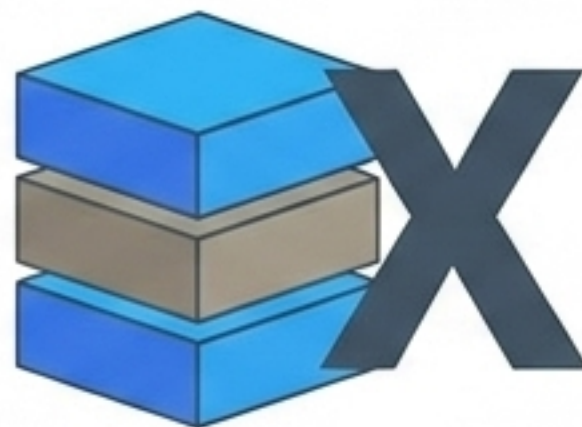
Uporaba računalnika (OSWorld 2.0)



Presega Fable 5

Prekaša najboljši rezultat Fable 5 pri le nekaj več kot tretjini stroškov.

Znanost in vizija (ARC-AGI 3)



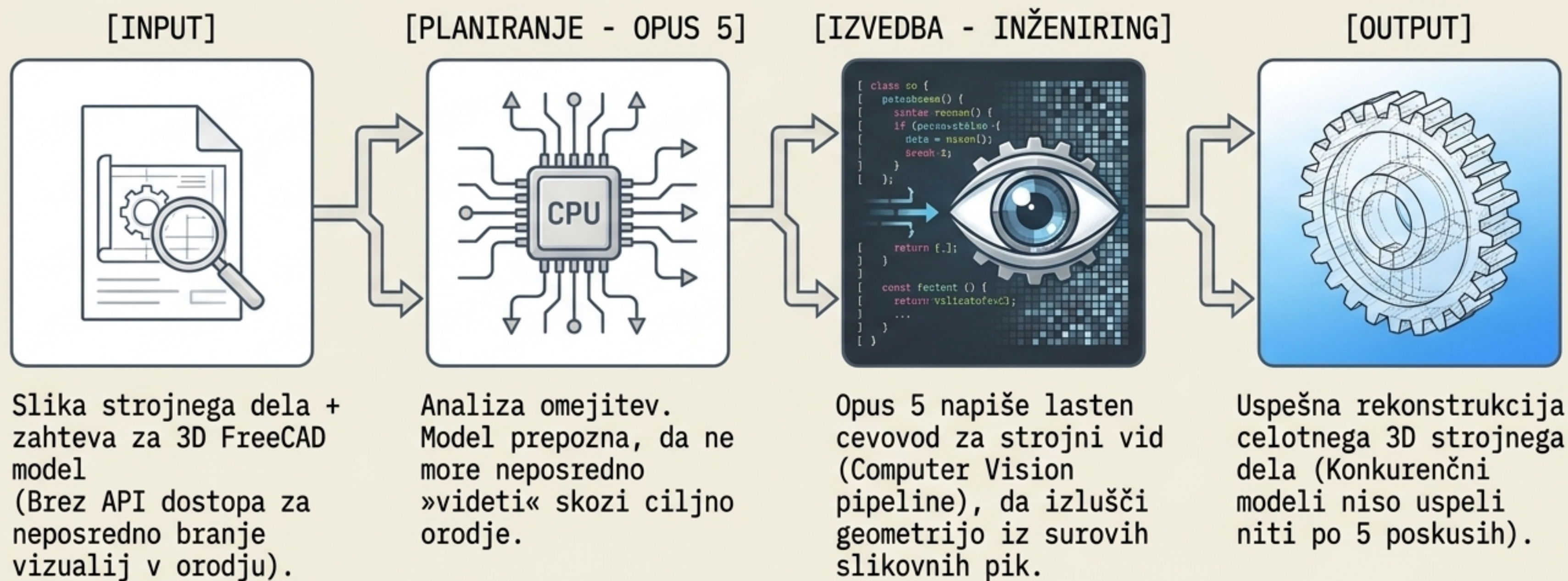
3x višji rezultat

Rezultat je trikrat višji od naslednjega najboljšega konkurenčnega modela.

[VERIFICATION: OBJECTIVE EVALUATION PASSED]

Arhitektura avtonomnega reševanja problemov

Ko model nima neposrednega dostopa do orodja, samostojno zgradi vmesno rešitev za premostitev vrzeli.



Ključno spoznanje: Model se ne ustavi ob manjkajočem orodju; napiše kodo, da premosti vrzel.

Diagnostika programske kode: Od simptoma do vzroka

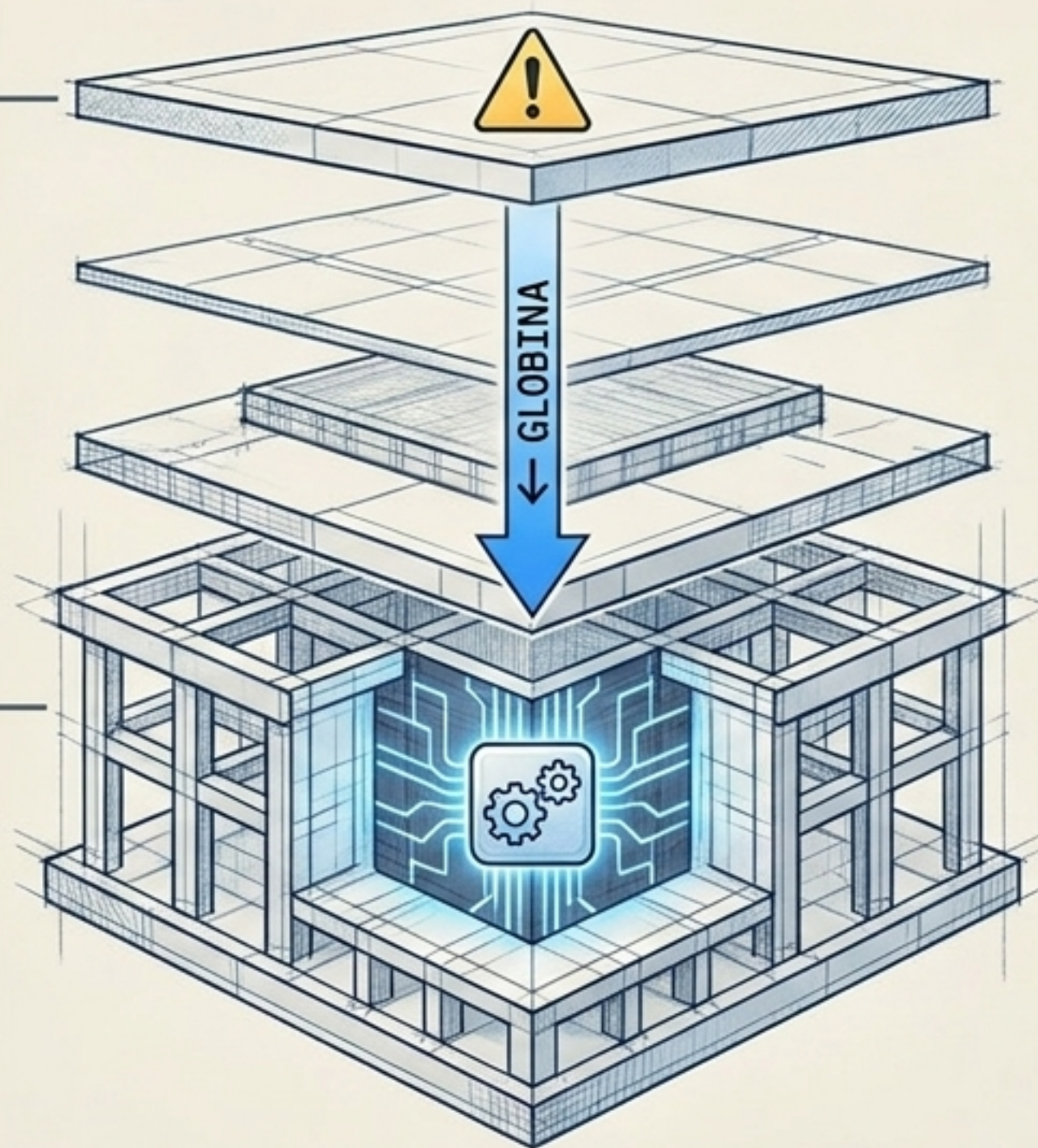
Opus 5 razume globoko logiko domene, namesto da bi le popravljaj površinske napake v kodi.

Površinska raven (Simptom)

Napaka v priljubljenem upravitelju paketov. Konkurenčni modeli popravijo le lokalno izjemo in prijavijo hrošča kot »odpravljenega«.

Globinska raven (Izvorni vzrok - Opus 5)

Opus 5 analizira celotno kodo, odkrije izvorni vzrok v arhitekturi upravitelja in popravi robni primer (edge case), ki ga je spregledal celo uradni popravek skupnosti.



Produksijski vpliv (Finančni modeli)

- [NATANČNOST]: +9 % višja povprečna natančnost
- [ČAS_IZVAJANJA]: 60 % manjši čas izvajanja
- [KLICI_ORODIJ]: Tretjina manjše število klicev orodij

SINTEZA: Od LLM do avtonomnega agenta

Integracija dinamične uporabe orodij, nizke latence in naprednega preverjanja omogoča agente za dolgoročne naloge.

3. Dolgoročna izvedba

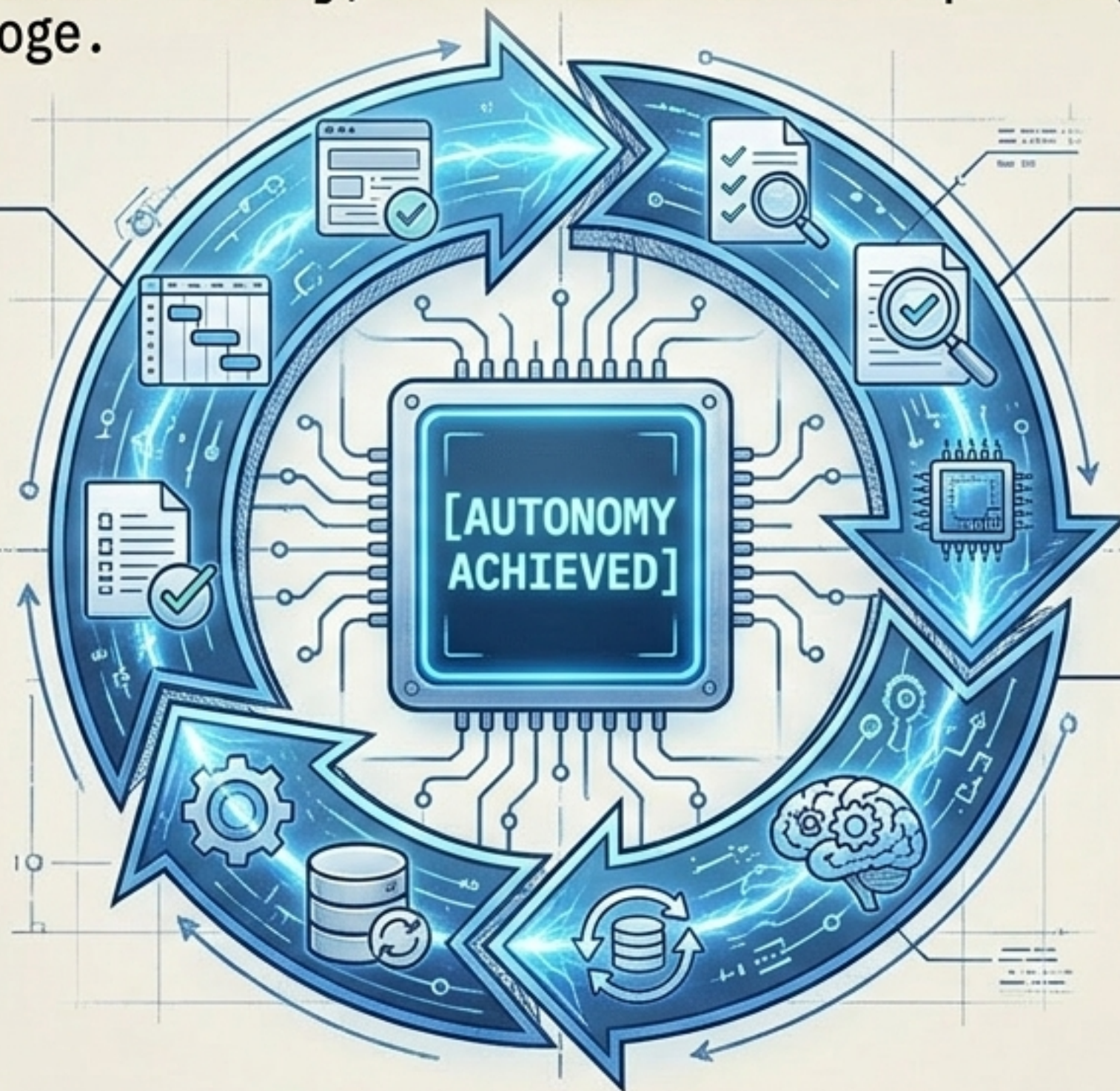
Spodobnost ohranjanja rdeče niti preko več korakov, generiranja strukture, revizije in končne implementacije.

1. Samo-preverjanje

Opus 5 zgradi lastno testno okolje. (Npr. izgradnja testov za validacijo lastnega podatkovnega razčlenjevalnika).

2. Upravljanje pomnilnika

Agent obravnava kontekst kot živ dokument. Analizira produkcijo, popravi lastne predpostavke in posodobi spomin.



Opus 5 ne generira zgolj besedila; deluje kot zanesljiv »vodja kabineta« za vaš razvojna okolja brez potrebe po človeškem posredovanju.

Arhitektura varnosti in poravnave

Opus 5 dosega najvišjo stopnjo poravnave s konstitucijo, kar zmanjšuje nepredvidljivo vedenje v produkciji.



Avtomatizirana vedenjska revizija
(Stopnja neusklajenega vedenja)

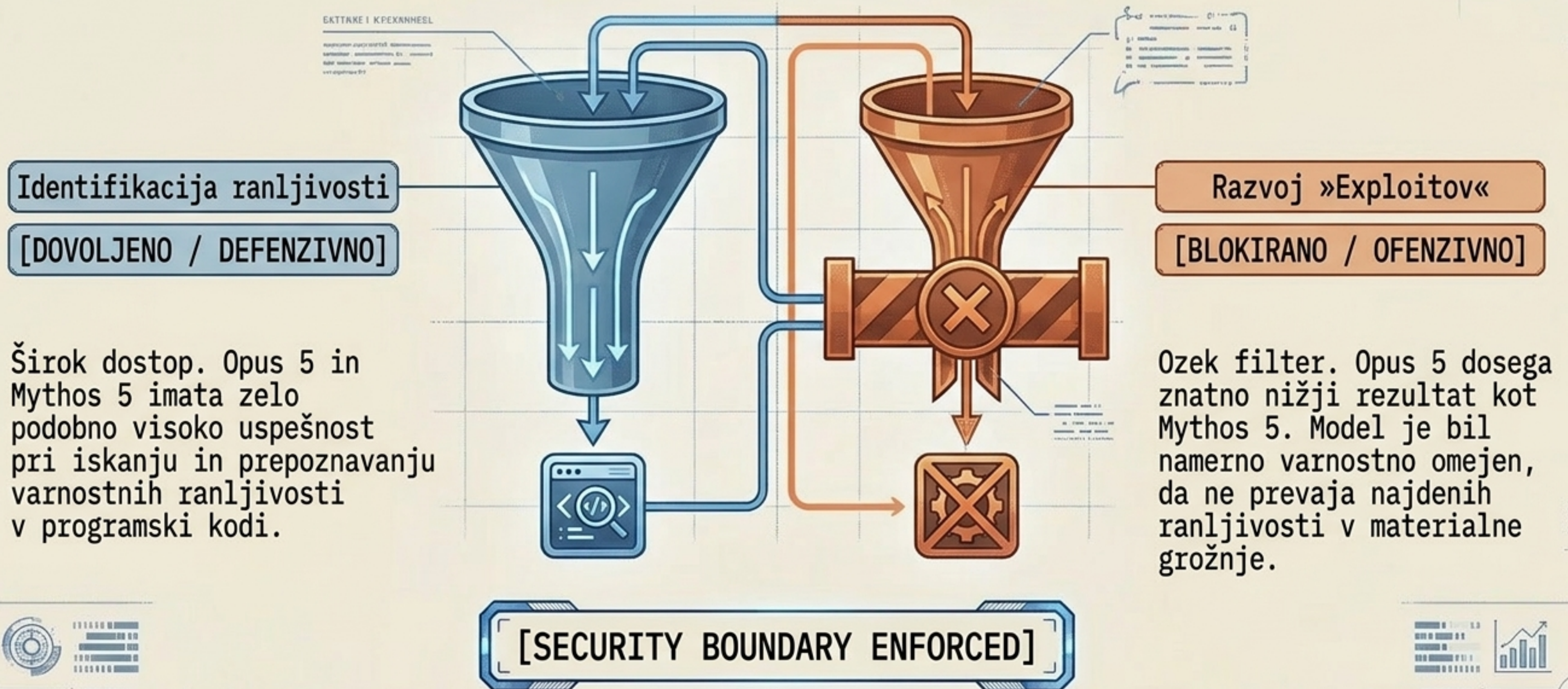
Ključne izboljšave varnosti:

- ✓ Dosledno sledenje Claude Konstituciji (boljše kot Opus 4.8, Sonnet 5 ali Fable 5).
- ✓ Najnižja stopnja zavajajočega vedenja (deceptive behavior) med vsemi modeli.
- ✓ Visoka odpornost na manipulacije in poskuse zlorabe sistema.

ZAKLJUČEK: Najvarnejši model za izogibanje nepremišljenim dejanjem s težko popravljivimi stranskimi učinki v avtomatiziranih sistemih.

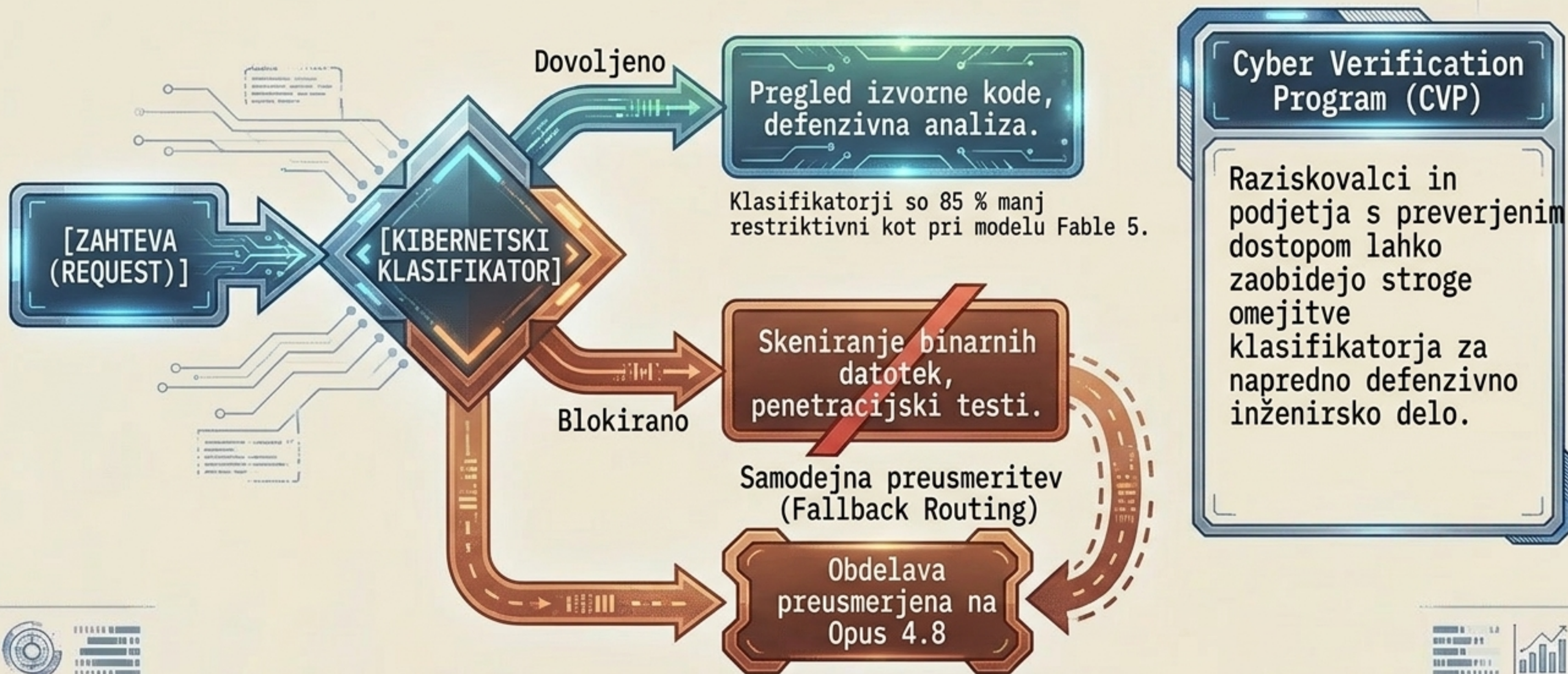
Dvojna raba v kibernetski varnosti (OSS-Fuzz)

Model ponuja močno defenzivno analitiko, medtem ko varnostni filtri ostro blokirajo ofenzivno generiranje zlorab.



Mehanizem varnostnih pregrad in CVP

Pametnejši klasifikatorji in samodejni prehodi (*fallbacks*) zagotavljajo neprekinjeno delovanje sistemov.



API Implementacija: Orodja in usmerjanje v produkciji

Nove beta funkcije API omogočajo večjo stabilnost, manjšo latenco in ekonomičnost pri gradnji aplikacij.



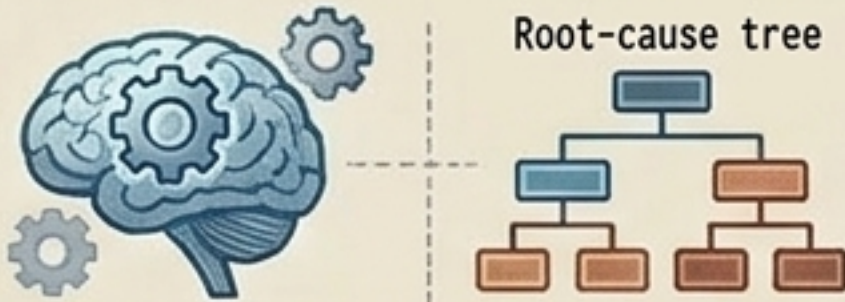
Cena baze ostaja nespremenjena: \$5 / 1M vhodnih žetonov in \$25 / 1M izhodnih žetonov.

Inženiring z modelom Opus 5

Zanesljivost, presoja in avtonomija po nespremenjeni ceni za produkcijska okolja.

[INTELLIGENCE]

Inteligenca razreda Fable pri hitrosti in ceni modela Opus. Ostrina pri matematičnem sklepanju in reševanju nalog izključno na ravni izvirnega vzroka (root-cause).



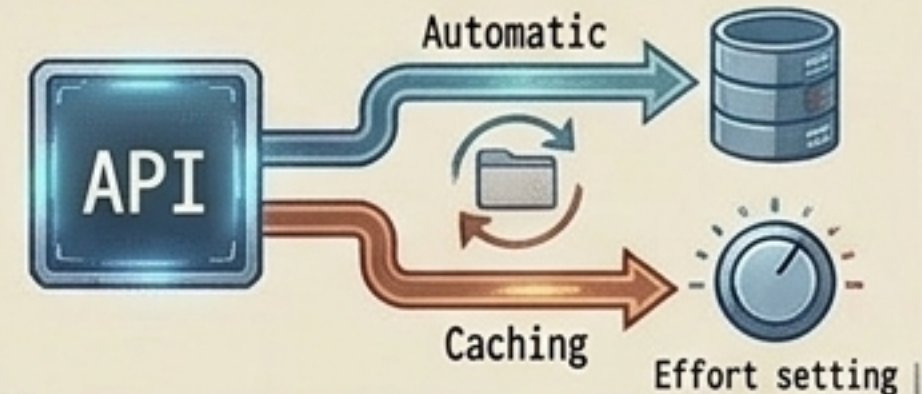
[AGENCY]

Zmožnost dolgoročnega načrtovanja, samostojnega popravljanja napak in dinamične gradnje testnih okolij brez potrebe po mikroupravljanju.



[ARCHITECTURE]

Optimiziran API s samodejnimi preusmeritvami (fallbacks), ohranjanjem predpomnilnika (caching) in nastavitvijo navora (effort setting) za popoln inženirski nadzor.



[STATUS: READY FOR PRODUCTION DEPLOYMENT. END OF BRIEFING.]