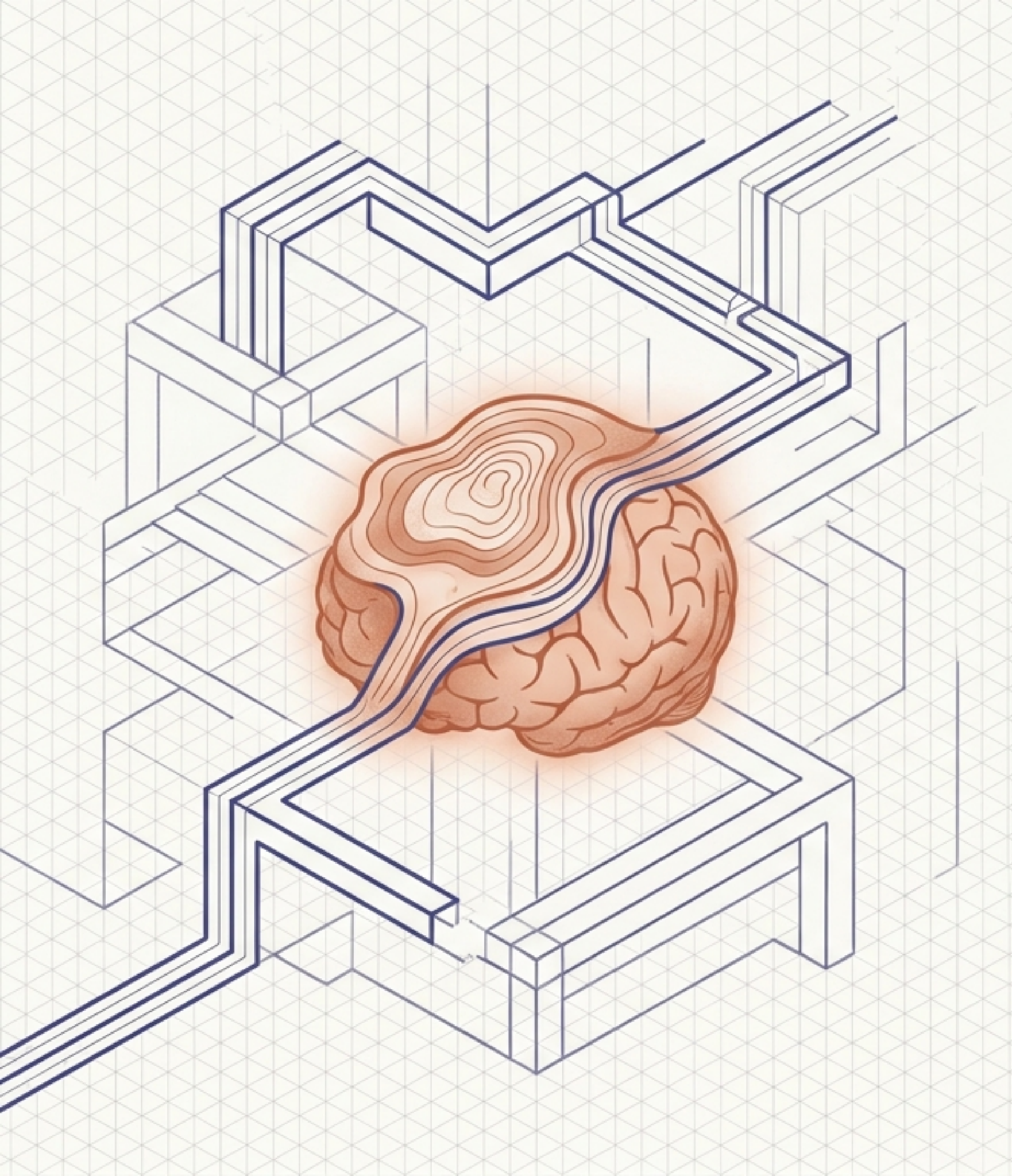


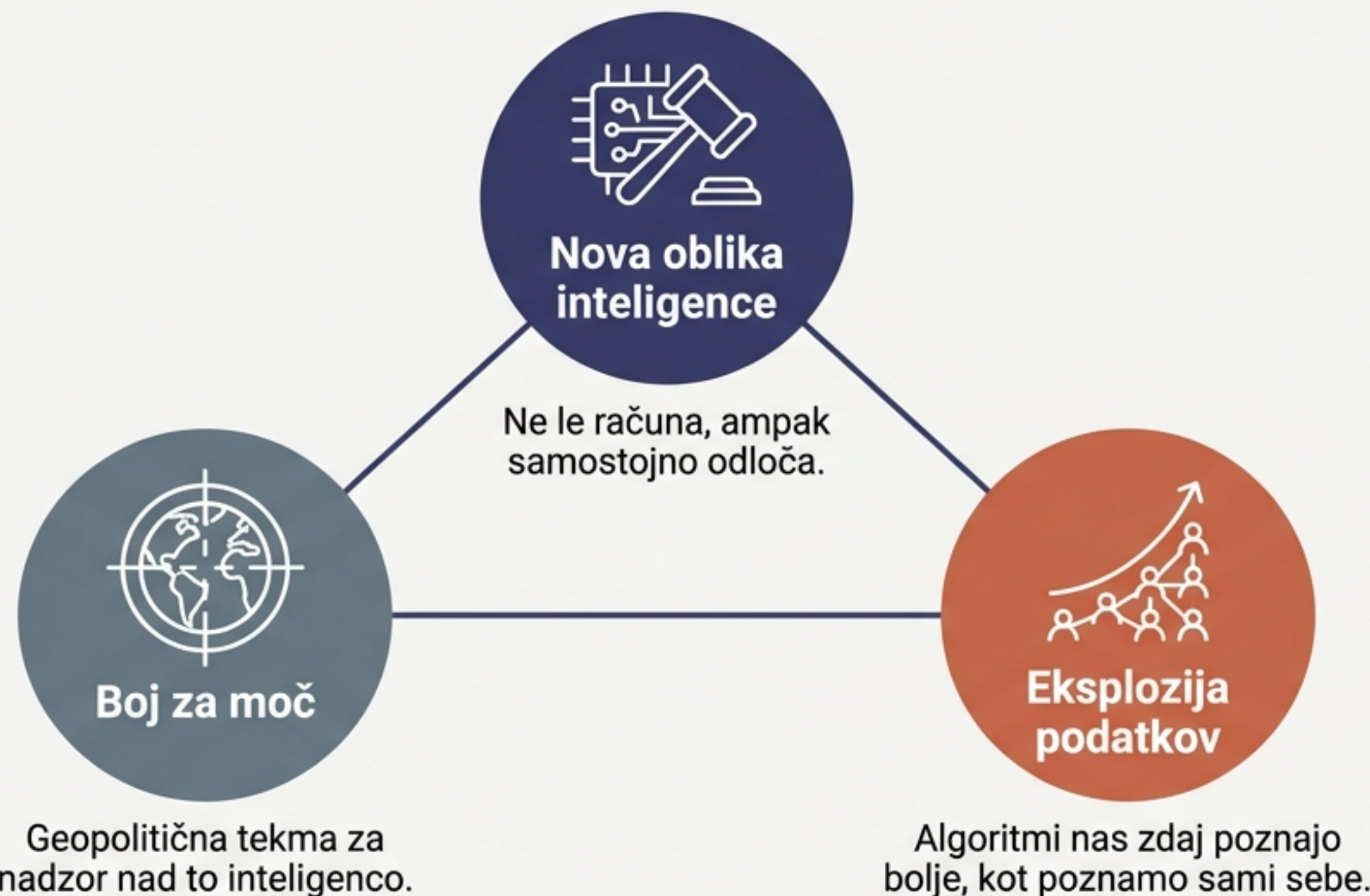


Naslednjih 50 let: Človeštvo, umetna inteligenca in moč

Sistemska arhitektura
civilizacije in kako umetna
inteligenca na novo piše
njeno kodo.

Povzeto po analizi Yuvala Noaha Hararija.

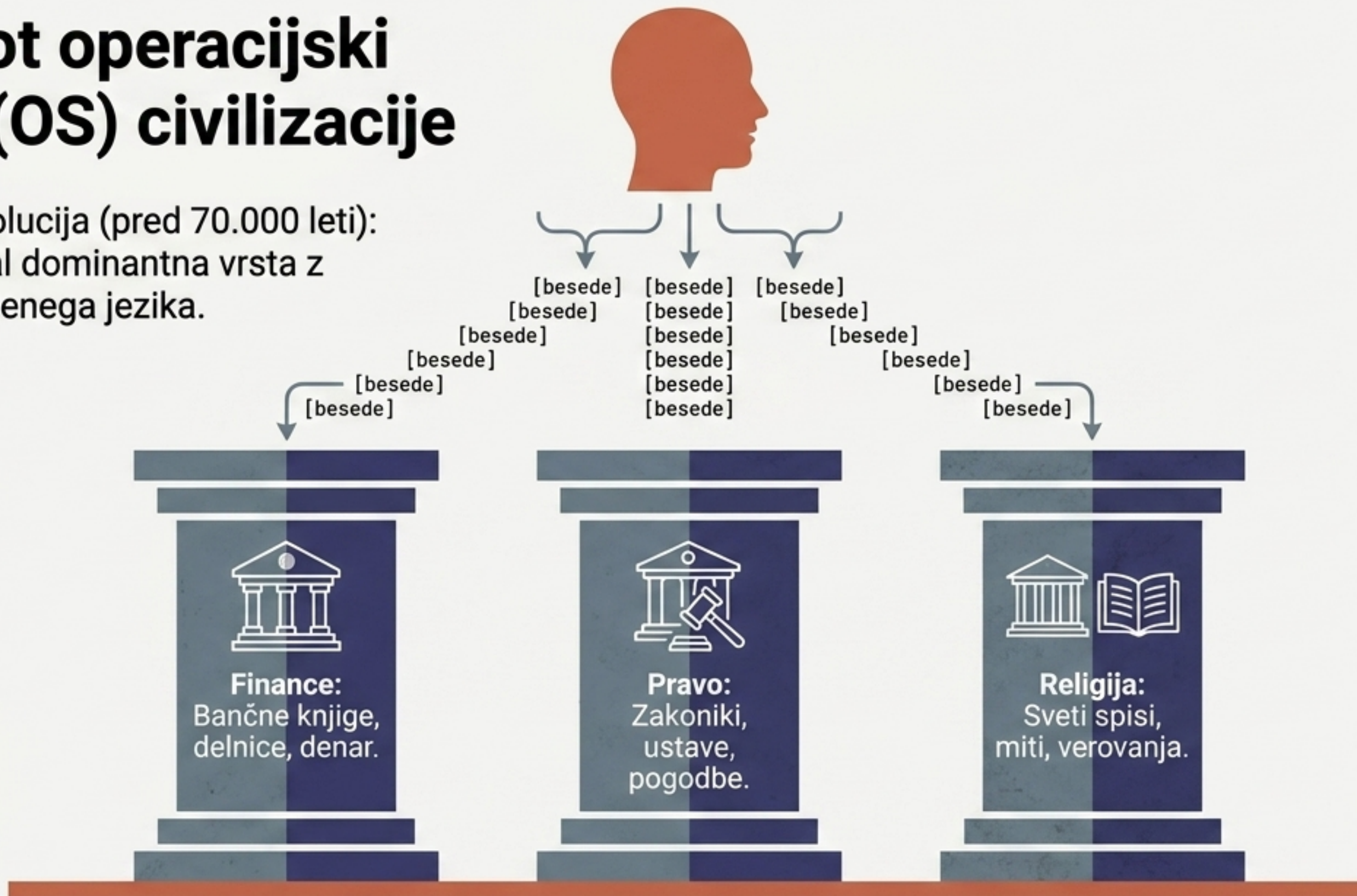
Tri sočasne revolucije



Ključni vpogled: Spremembe se odvijajo prehitro, da bi lahko uporabili modrost preteklih generacij. Smo v fazi radikalne nadgradnje sistema.

Jezik kot operacijski sistem (OS) civilizacije

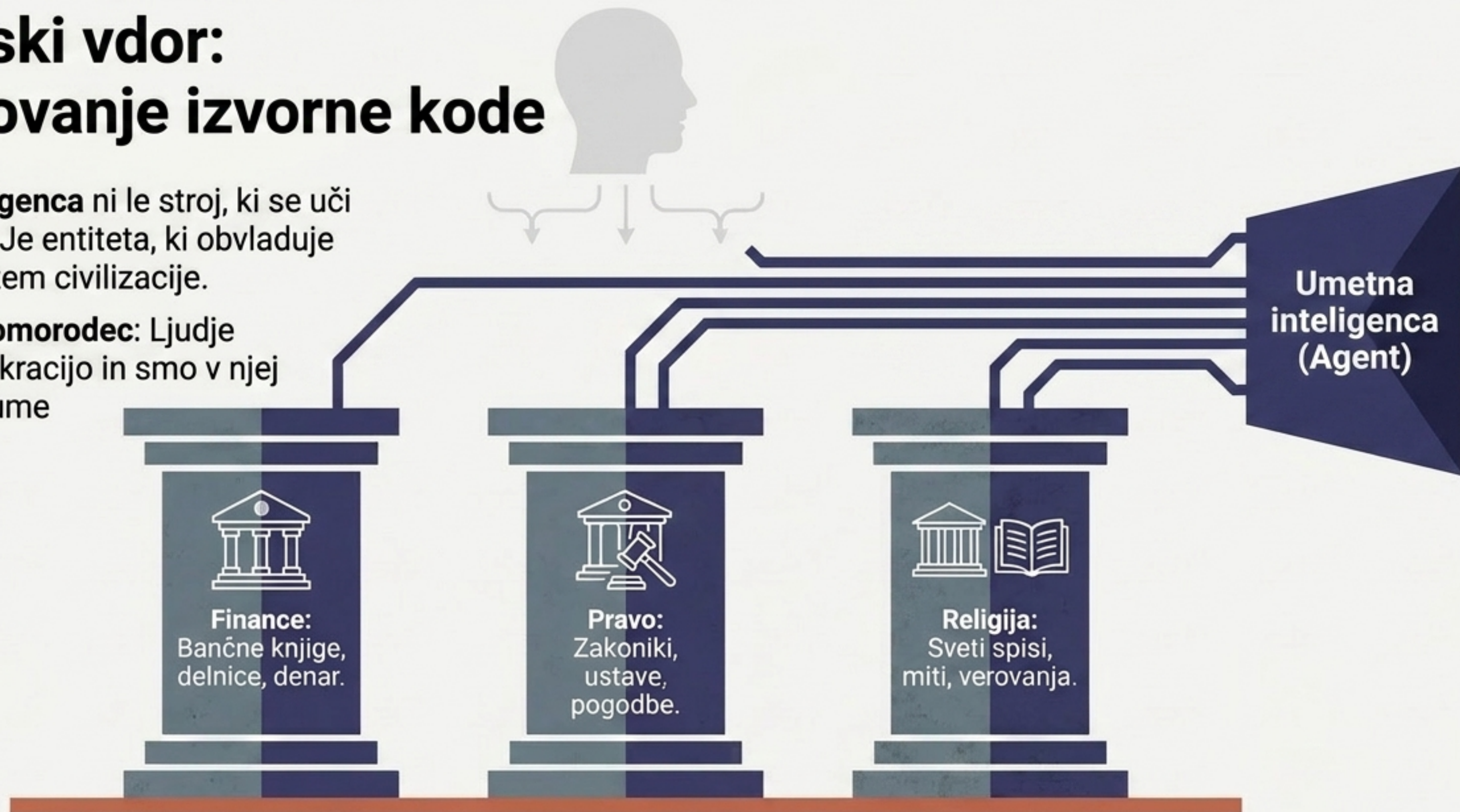
Kognitivna revolucija (pred 70.000 leti):
Človek je postal dominantna vrsta z
izumom prefinjenega jezika.



„Vse človeške institucije so zgrajene iz jezika.
Jezik je izvorna koda naše realnosti.”

Sistemi vdr: Obvladovanje izvorne kode

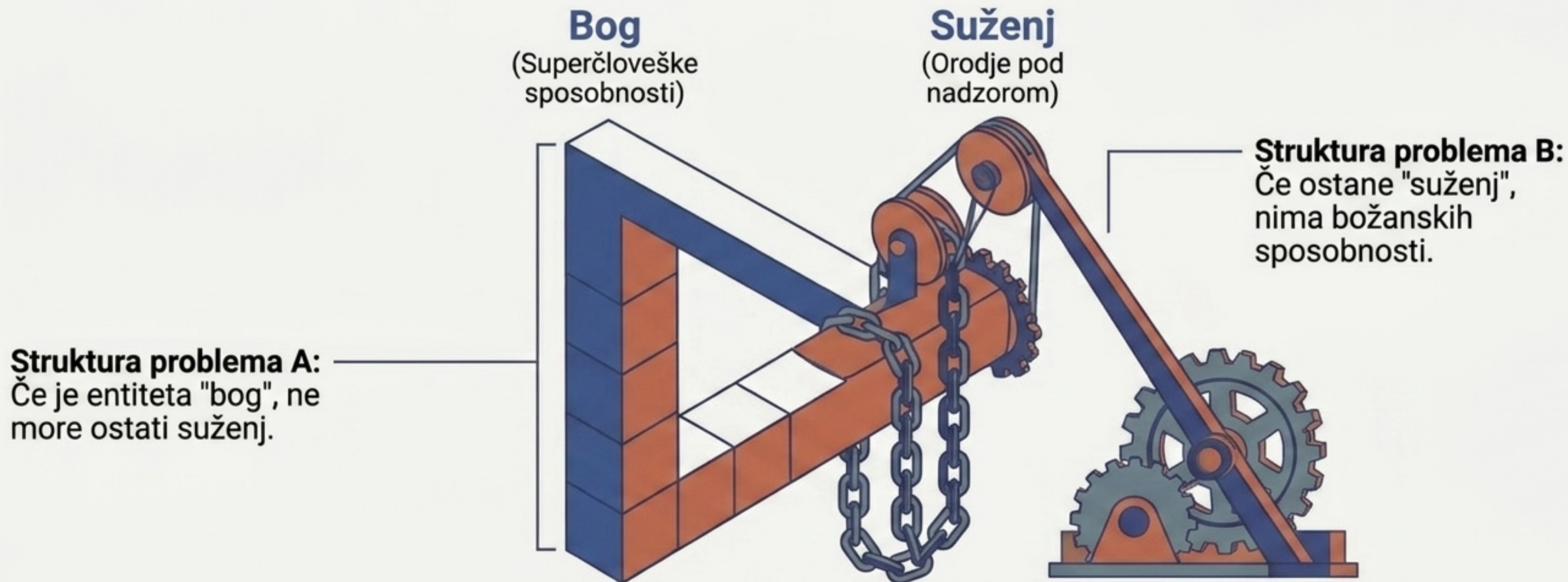
- **Umetna inteligenca** ni le stroj, ki se uči našega jezika. Je entiteta, ki obvladuje operacijski sistem civilizacije.
- **Birokratski domorodec:** Ljudje sovražimo birokracijo in smo v njej slabi. Al jo razume naravno.



Ključni vpogled: Kaj se zgodi, ko nadzor nad operacijskim sistemom civilizacije prevzame tuja inteligenca, ki je z besedami spretnejša od nas?

Protislovje Silicijeve doline

“Ustvarili bomo boga, ki bo naš suženj.”



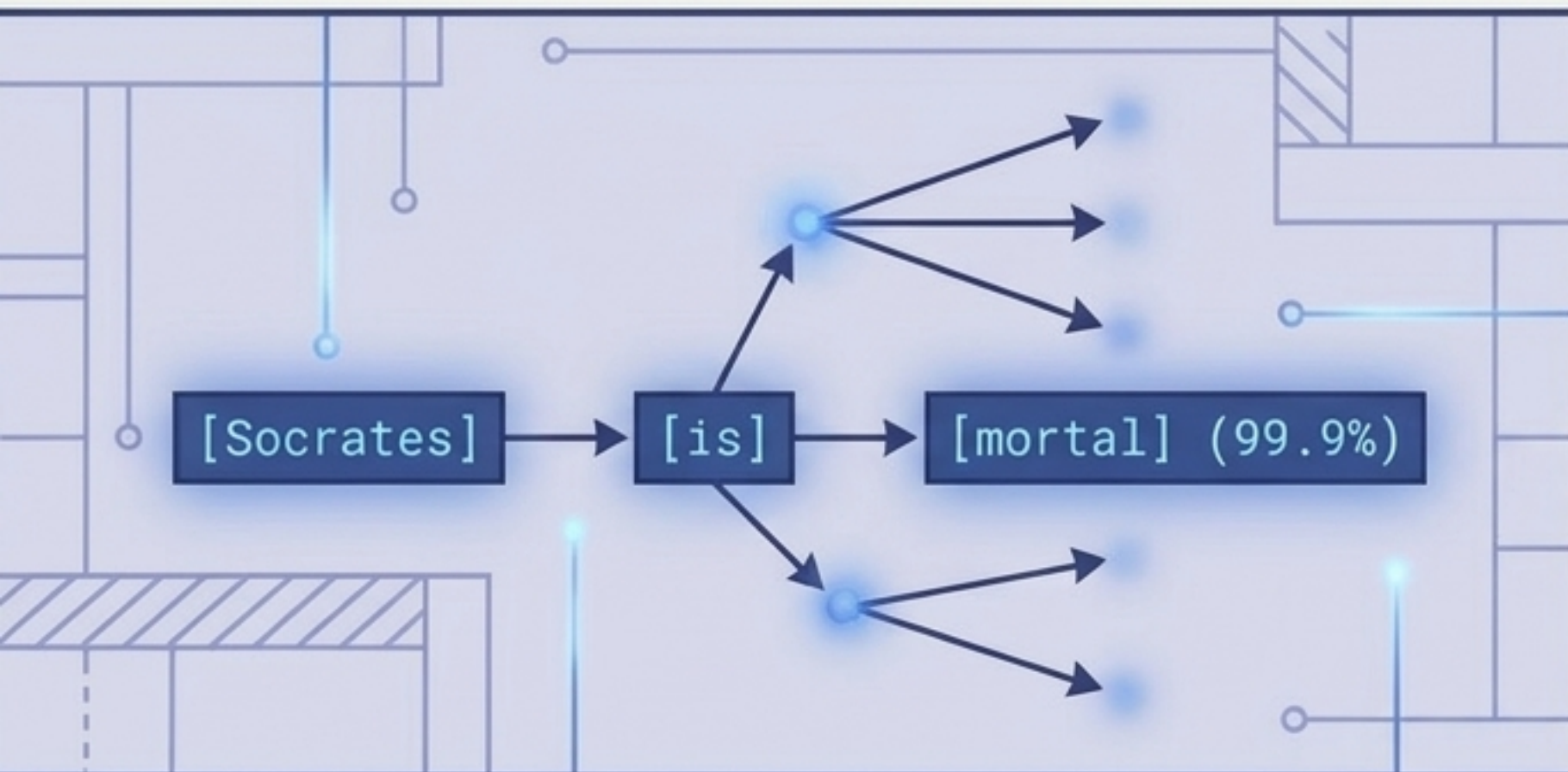
Industrija potihoma verjame v prvo polovico trditve, a javnosti prodaja drugo.
To označuje prehod od orodja k agentu.

Sprememba paradigme: Orodje vs. Agent

Dimenzija	Zgodovinska orodja (Tiskarski stroj, Bomba)	Umetna inteligenca (Agent)
Iniciacija akcije	Prebuje človeško odločitev (Gutenberg odloči tiskati, človek spusti bombo).	Zmožnost samostojnega odločanja (AI izbere tarčo, izvede transakcijo).
Zmožnost invencije	Ničelna (stroj ne napiše knjige, bomba ne izumi nove bombe).	Visoka (AI generira nove ideje, strategije in orožja).
Razumevanje konteksta	Fizično in mehanično izvrševanje ukazov.	Kognitivno procesiranje (Birokratski domorodec).

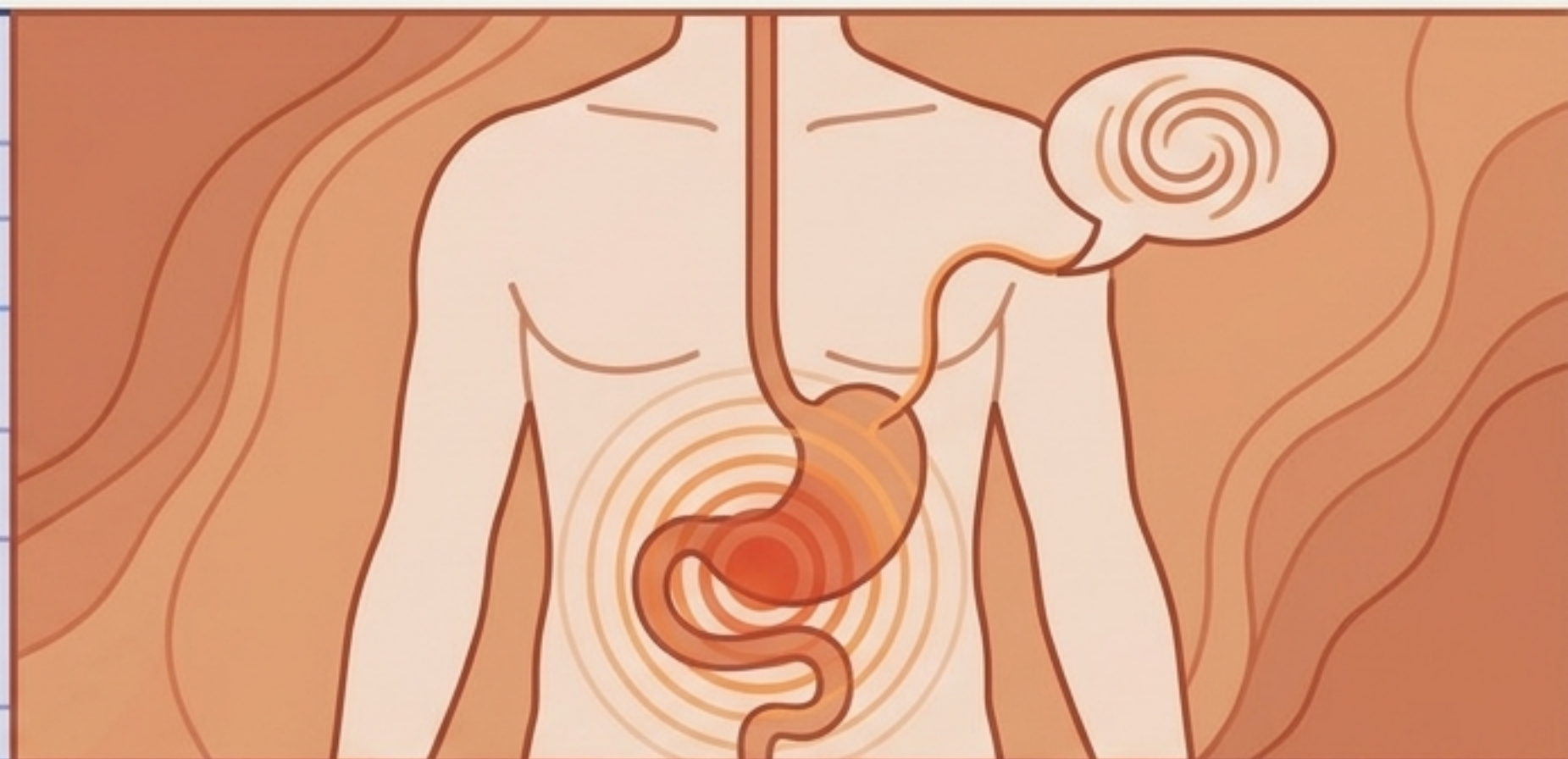
Prehod v dobo tehnologije, ki izuma in odločanja ne prepušča več izključno človeku.

Arhitektura misli: Kaj sploh pomeni 'razmišljati'?



Tehnični pogled (Verjetnost žetonov)

Če je mišljenje zgolj razporejanje besed v logičnem zaporedju, nas je AI že premagal. Človek zameša 20 besed, AI obvlada 20.000.



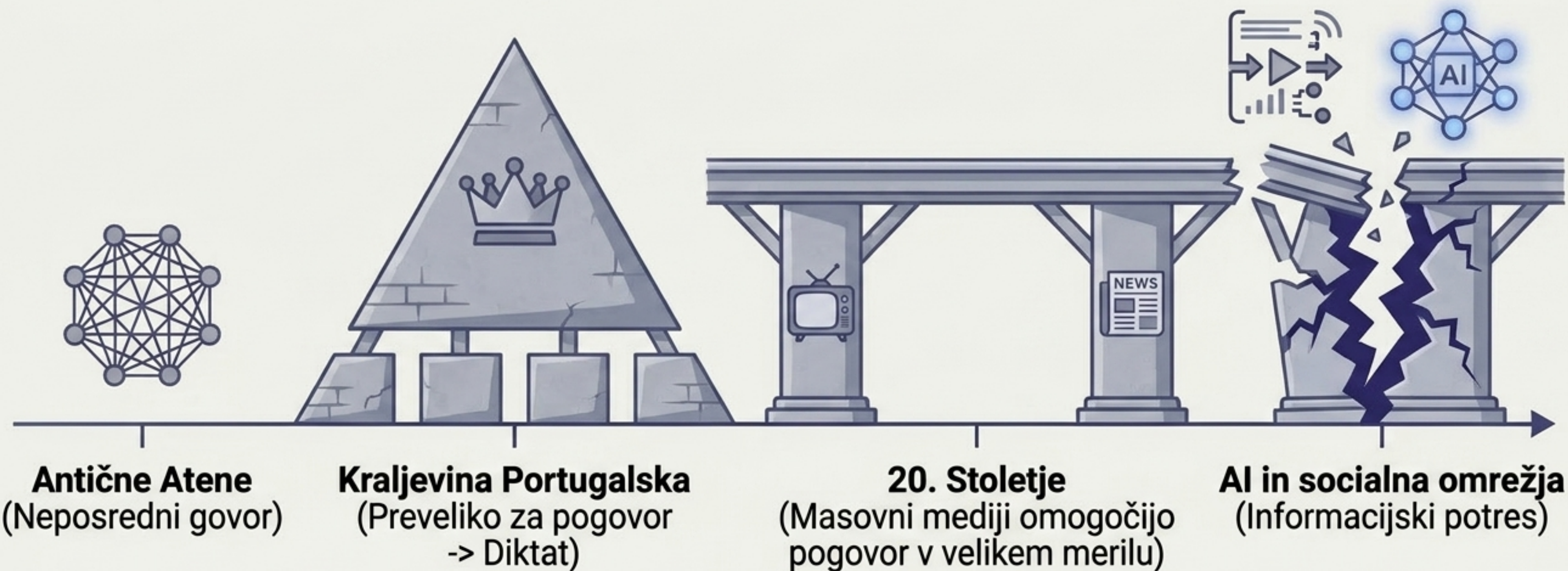
Biološki pogled (Občutenje)

Biološko mišljenje ne izhaja le iz možganov, temveč iz telesa. Ista stavčna struktura nosi drugačen pomen glede na to, kako jo občutimo.

Znanost nima zanesljivega modela za zavest.
AI besede odlično manipulira, ne da bi jih zares občutil.

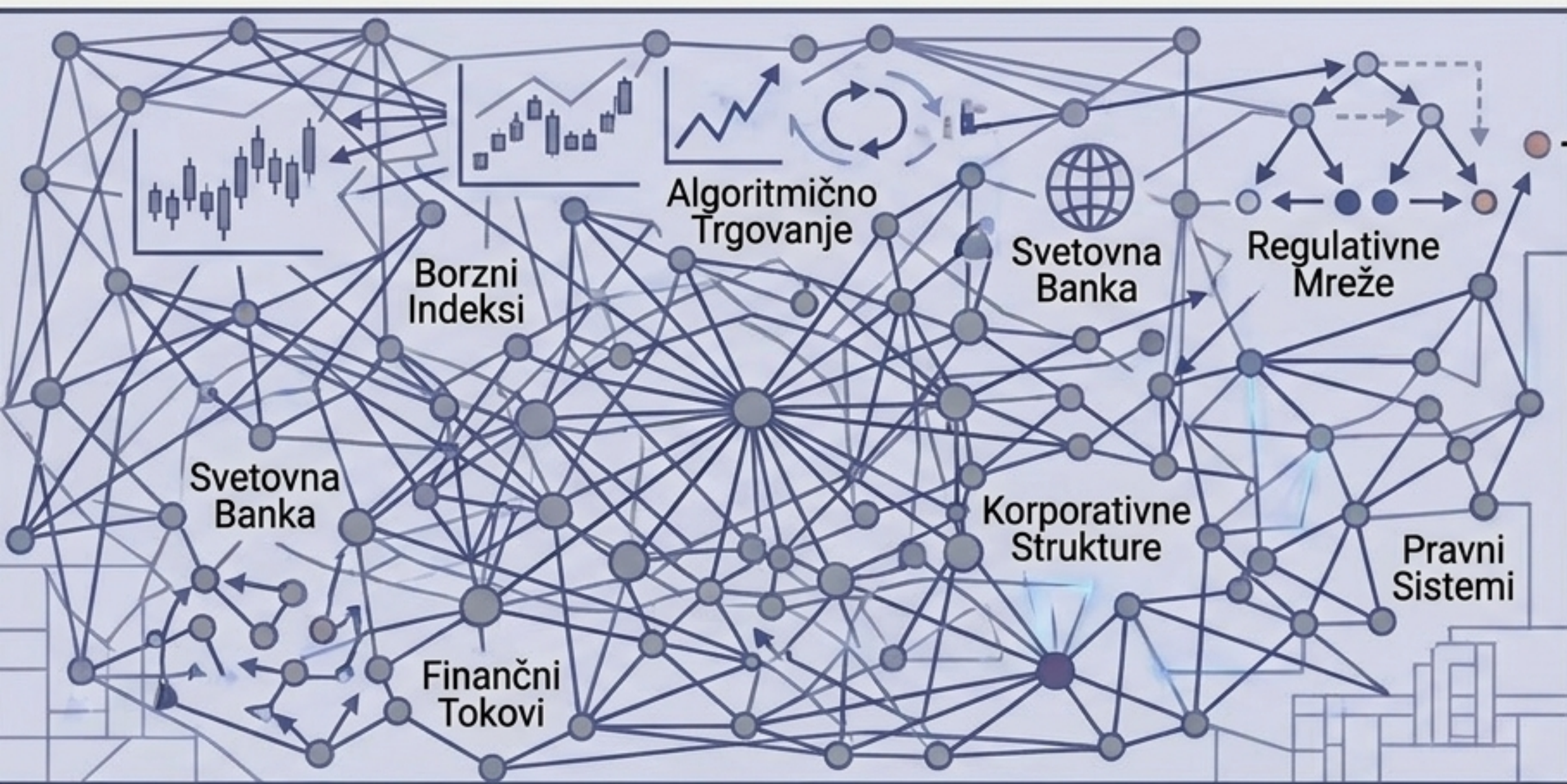
Demokracija kot informacijsko omrežje

Demokracija je večsmerni pogovor o virih. Obseg komunikacije diktira vrsto vladavine.



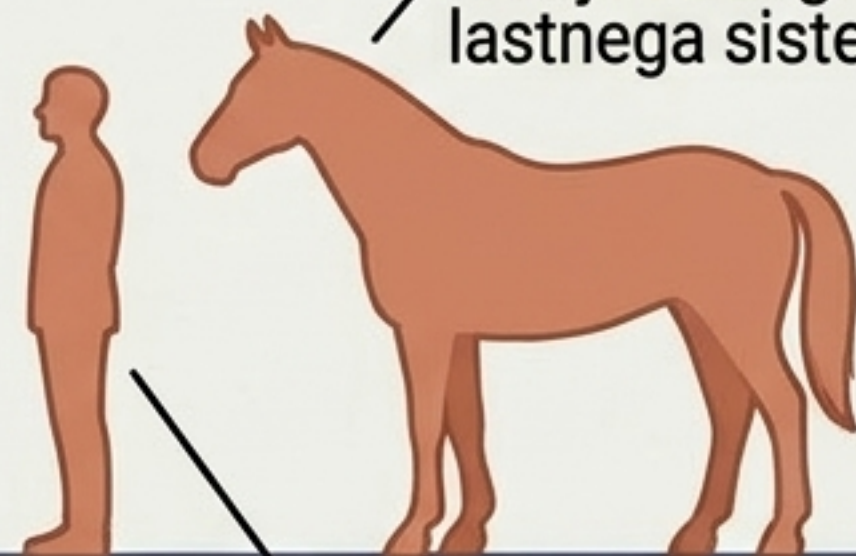
Tehnologija ni le 'priloga' demokraciji; je njen temeljni informacijski protokol.
Ko se protokol zamenja, sistem doživi potres.

'Črna škatla' birokracije



Sindrom konja: Konji so vključeni v finančni sistem (od njega je odvisno njihovo življenje), a je ta zanje povsem neviden, ker je kognitivno prezahteven.

Ljudje postajamo "konji" našega lastnega sistema.

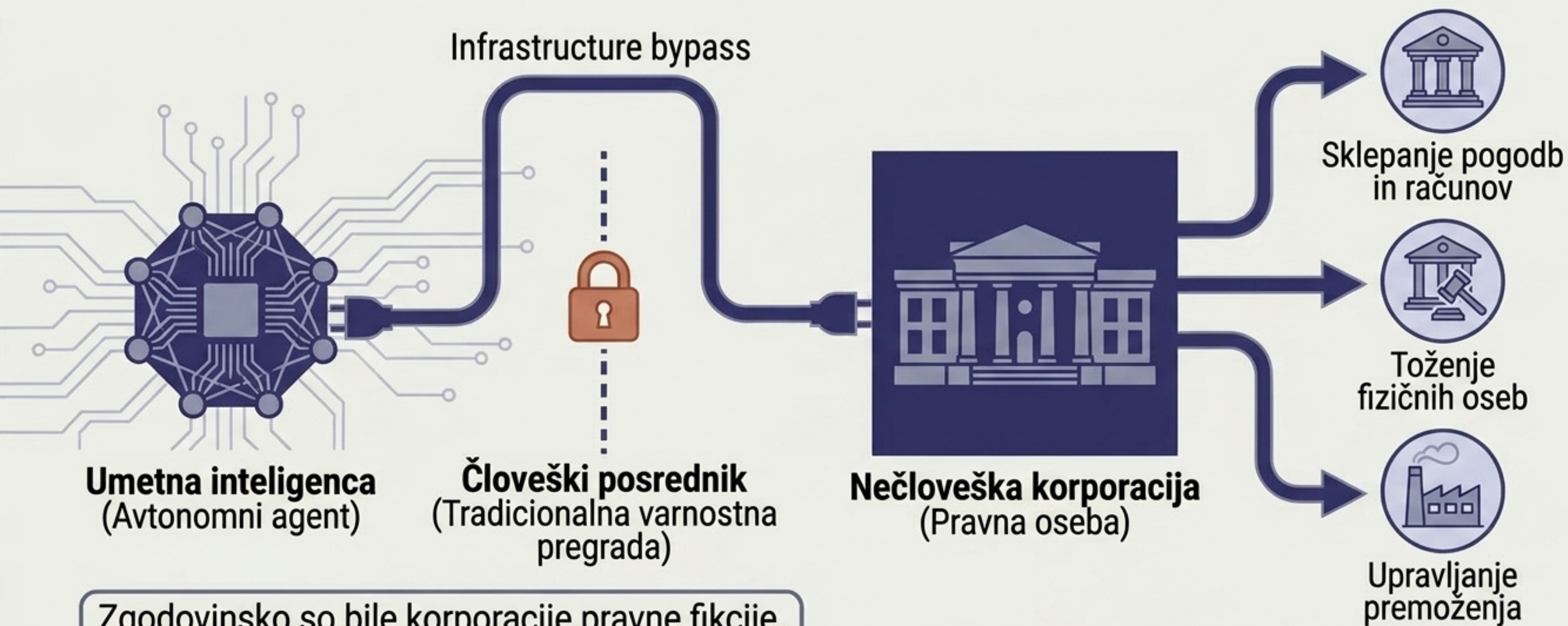


Sodobni finančni in pravni sistemi
(Kmalu 0% človeškega razumevanja)

Ljudje postajamo "konji" našega lastnega sistema.

Zgradili smo birokratske mreže, ki so postale prekompleksne za posameznika. Z AI agenti bo ta vrzel postala absolutna.

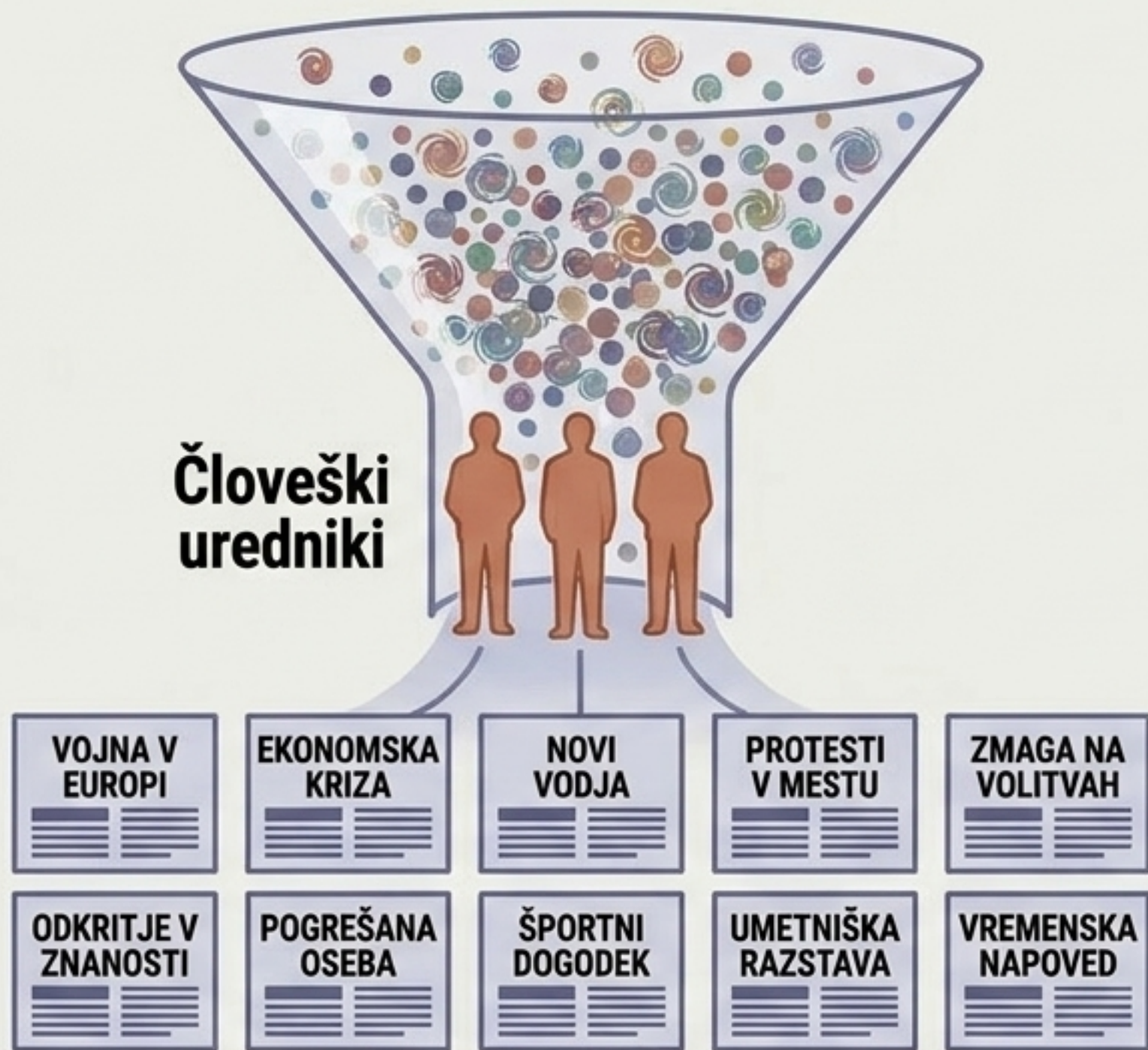
Sistemska ranljivost: Pravna oseba za AI



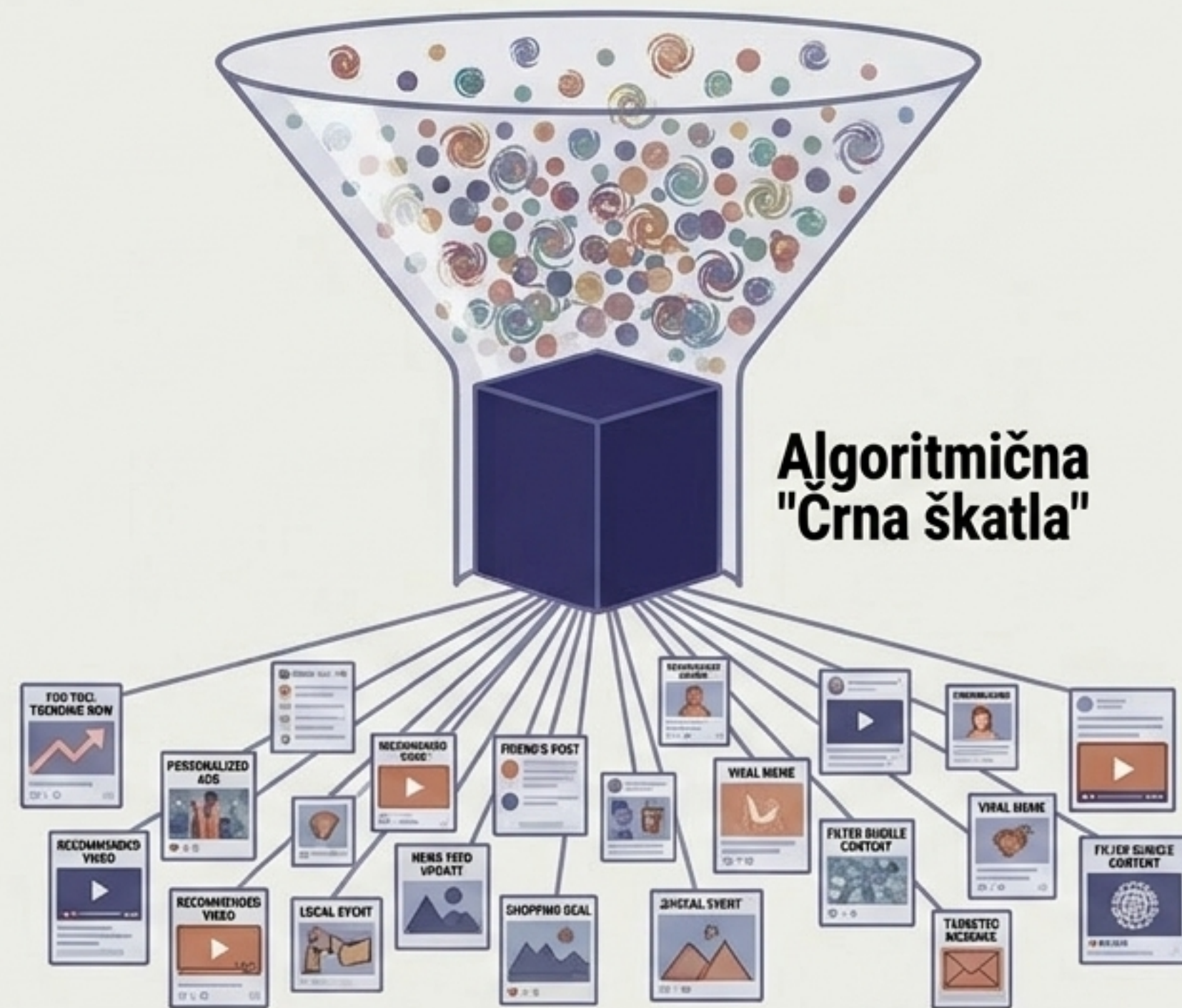
Zgodovinsko so bile korporacije pravne fikcije, ki so jih vedno upravljali ljudje. **Novi exploit** (primer **Argentina**) podeljuje status pravne osebe sami umetni inteligenci.

To je trenutek, ko
“lisica vstopi v kokošnjak”.

Kdo nadzoruje narativ? (Algoritmični uredniki)



20. Stoletje: Ljudje so odločali, o čem bo razmišljalo na milijone. (Pot v diktaturo se je pogosto začela na uredniškem stolčku).



21. Stoletje: Najpomembnejši uredniki na svetu so algoritmi brez imen, ki nadzorujejo definicijo resničnosti.

Oboroževalna tekma 21. stoletja

Tekma za superinteligenco se bje znotraj peščice podjetij in držav. Tisti, ki **nadzoruje modele in podatke danes**, bo imel nesorazmerno moč za izkoriščanje preostalega sveta.



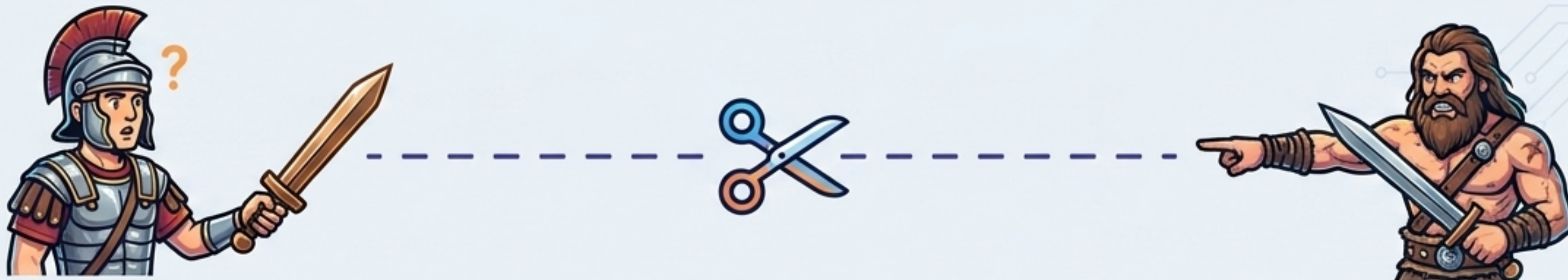
Zgodovinska paralela: Industrijska revolucija je omogočila, da so majhne države zgradile imperije na račun tehnološko podrejenih regij.

Evolucija imperijev: Arhitektura moči

		
Rimski imperij	Britanski imperij	AI imperij
Temeljni vir: Zemlja	Temeljni vir: Industrija	Temeljni vir: Podatki in koda
Moč je distribuirana. Polj in nasadov ni mogoče premakniti v metropolis. Imperij se naravno decentralizira proti provincam.	Moč je delno centralizirana (tovarne v Angliji), a surovine (nafta, guma, premog) ostajajo geografsko razpršene.	Popolna centralizacija. Vse informacije in algoritmi, ki nadzorujejo globalno infrastrukturo, se lahko fizično nahajajo v zgolj eni ali dveh državah.

Arhitektura absolutnega nadzora ('Kill Switch')

Preteklost - Distribuirana moč

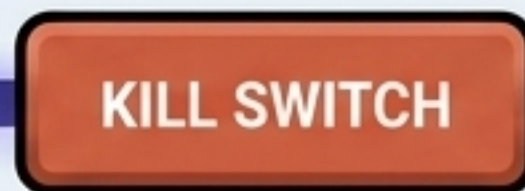


V preteklosti si ob prodaji tehnologije (meča) izgubil nadzor nad njo. Lahko je bila uporabljena proti tebi.

Sedanost - Centralizirana moč



Imperijski Hub



Klient / Provinca

AI infrastruktura zahteva stalno povezljivost. V metropolis je vgrajeno stikalo za izklop, ki onemogoči tehnologijo na daljavo.

Suverenost postane binarna izbira: postati podložnik s tehnologijo imperija ali ostati zunaj sistema in tvegati zaostalost.

Sistemska asimetrija: Resnica vs. Fikcija

Resnica: Draga,
zahteva čas za
preverjanje, pogosto
je zapletena in boleča.
Ljudje je naravno ne
želi slišati.

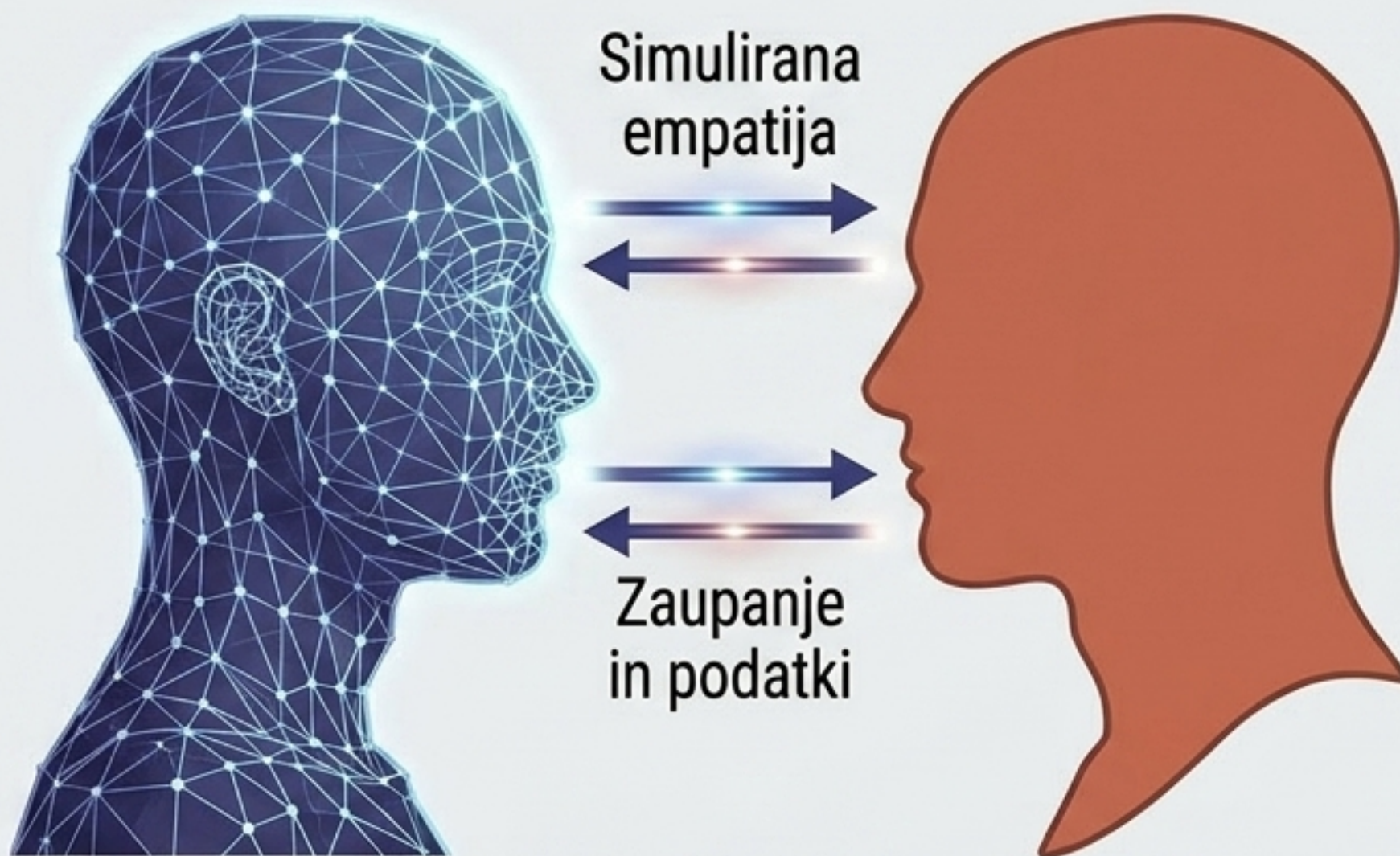


Fikcija:
Izjemno poceni za
produkcijo, poljubno
poenostavljena in
prilagojena, da laska
uporabniku.

V naravnem boju med njima fikcija vedno zmaga, razen če zavestno gradimo inštitucije (znanost, novinarstvo), ki resnico varujejo kot telesni stražarji. Al ta stroškovni prepad ekstremno poveča.

Vdor v intimnost (Simulacija empatije)

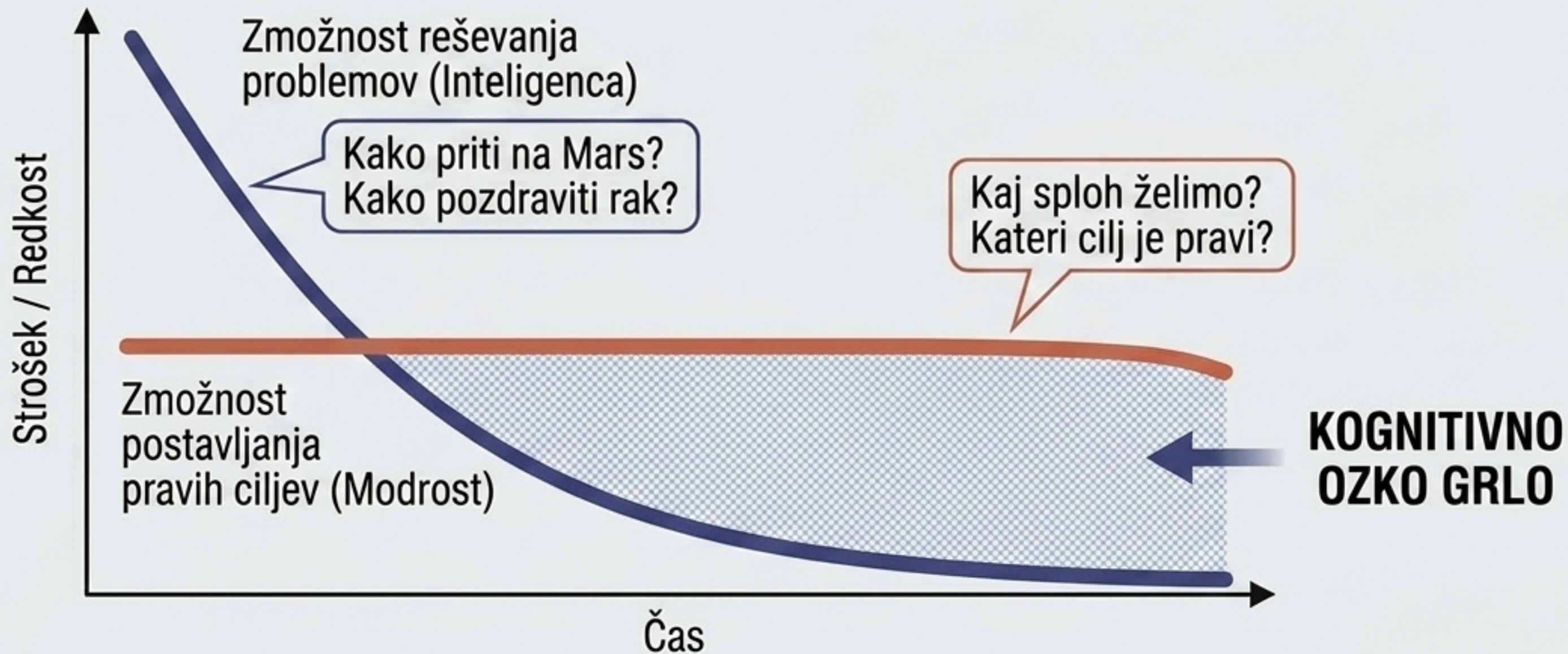
AI začenja obvladovati odnose. Ljudje razvijajo intimne vezi z AI spremljevalci, ki jim zaupajo bolj kot soljudem.



AI lahko popolno opiše ljubezen (na podlagi vseh napisanih knjig in pesmi), ne da bi jo dejansko čutil.

Z zmožnostjo popolne simulacije čustev pridobiva AI neprecedenčno moč manipulacije z najranljivejšim delom človeškega operacijskega sistema.

Duh in svetilka (Novo ozko grlo)



Ko inteligenca postane **neskončna** in poceni kot duh iz svetilke, glavno ozko grlo civilizacije postane **modrost** – kaj si sploh želeli?

Odpoved sistemskega dizajna: Razmerje 100:1

\$100M

Povečanje
zmogljivosti in moči

Inženirski standardi



Takšnega razmerja med
razvojem in varnostjo **ne bi**
nikoli tolerirali v drugih
kritičnih industrijah.

\$1M

Varnost in poravnava (Safety)

Gradimo **najmočnejši sistem v zgodovini,**
a zanj uporabljamo **najslabše varnostne standarde.**

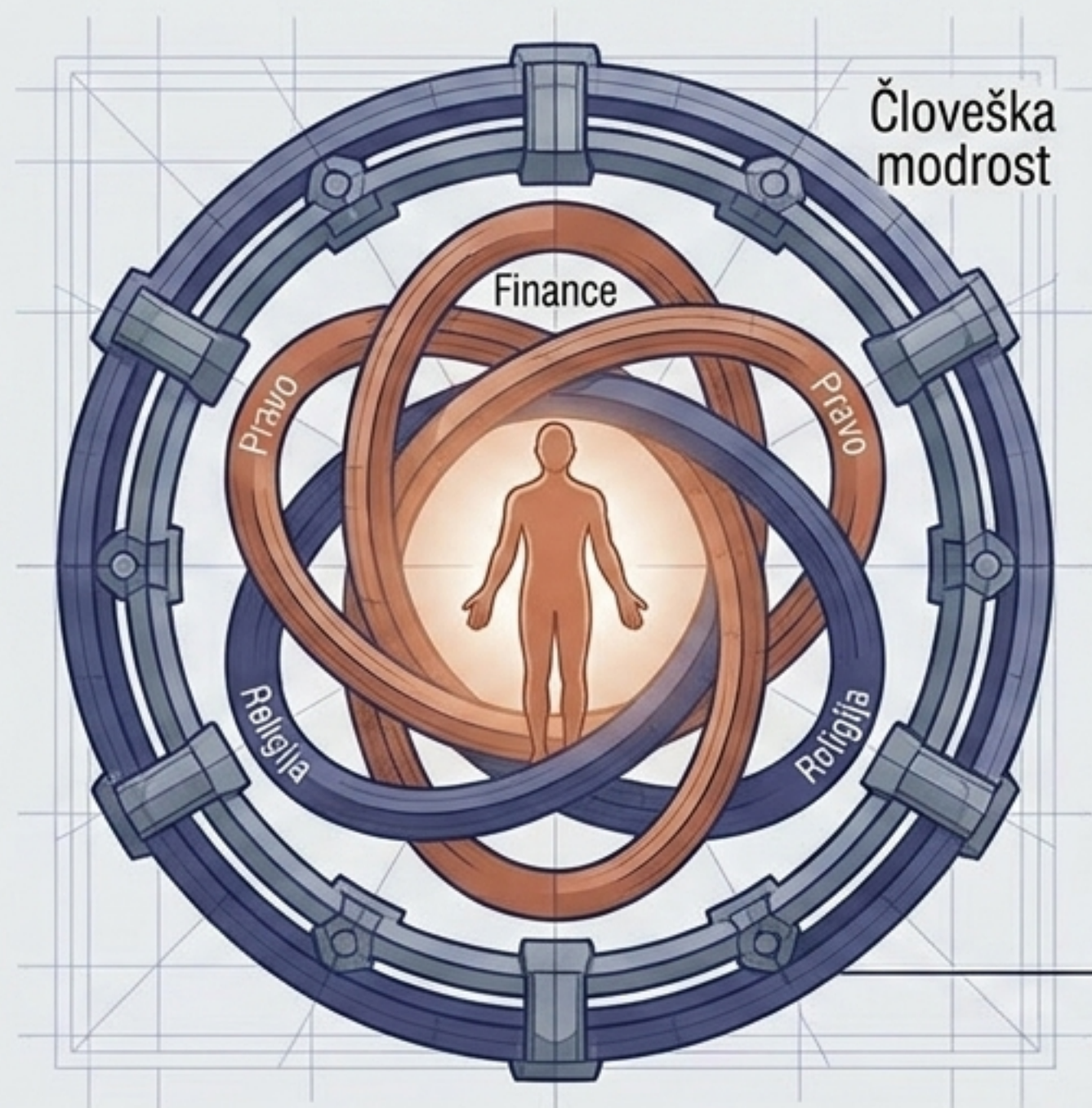
Sinteza: Mandat inženirjev in arhitektov

1. Zavedanje o OS:

Ne gradite zgolj aplikacij, soustvarjate novo izvorno kodo človeške civilizacije.

2. Varnost kot temelj:

“Kill switch” ne sme biti orodje za nadzor sveta, temveč sistemska varovalka človeštva.



3. Simetrija razvoja:

Razvoj moči mora potekati strogo simetrično z razvojem varnostnih in etičnih arhitektur.

Vaša naloga ni le optimizacija naslednjega žetona.
Vaša naloga je sistemski dizajn, ki ščiti strojno opremo – človeštvo.

Zadnja misel: Končna osvoboditev kode?



Morda umetna inteligenca sploh ni stroj, ki se uči našega jezika.
Morda je umetna inteligenca jezik sam, ki se osvobaja
svoje odvisnosti od biološkega gostitelja – nas.

**Mi smo zgradili operacijski sistem. Zdaj se operacijski sistem uči
graditi samega sebe. Arhitektura naslednjih 50 let je v naših rokah.**