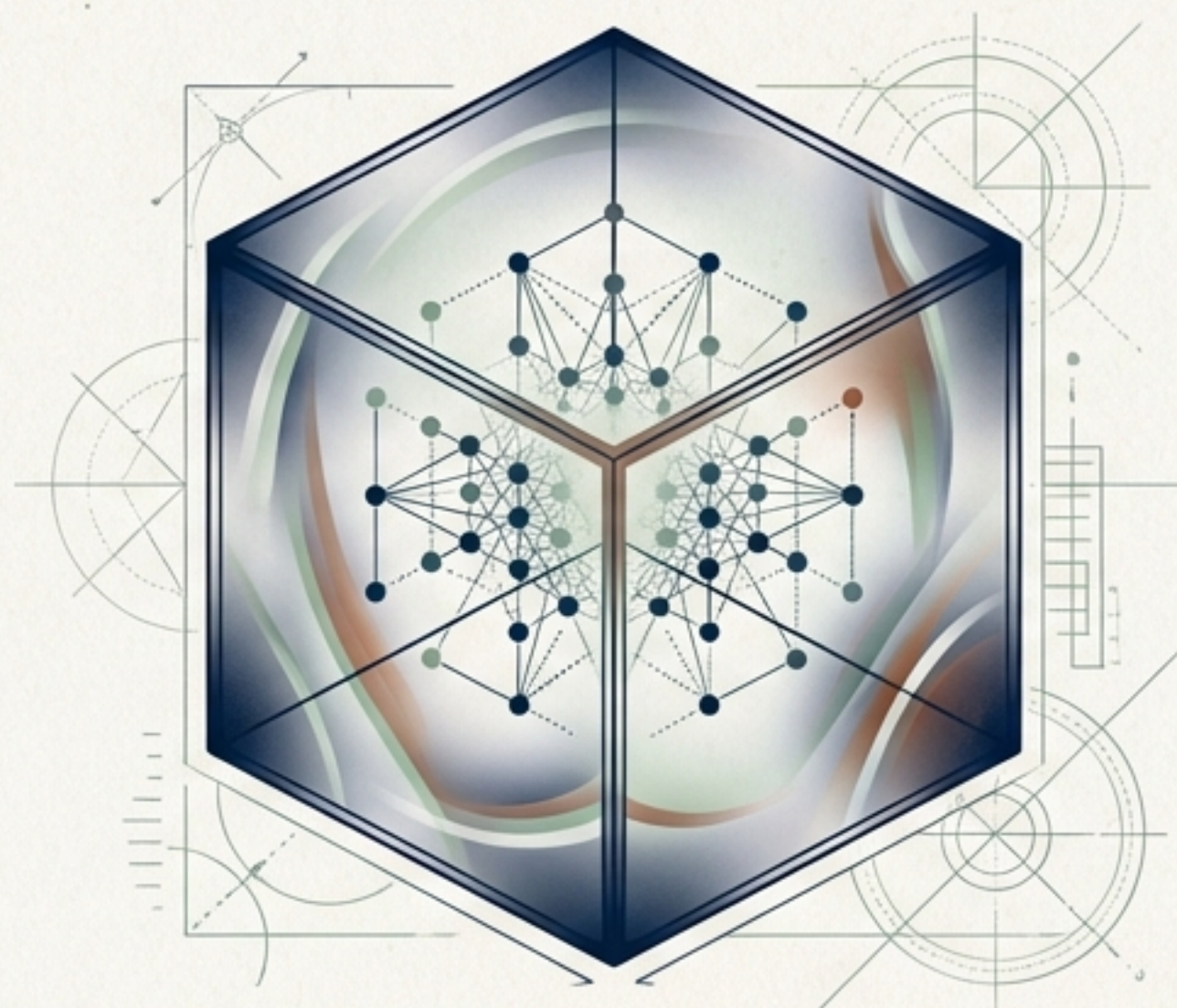


Umetna inteligenca in iluzije razumevanja v znanstvenih raziskavah

Analiza tveganj onkraj tehničnih napak: kako orodja UI oblikujejo proizvodnjo znanja



NA PODLAGI ČLANKA AVTORIC LISE MESSERI IN M. J. CROCKETT (NATURE, 2024)

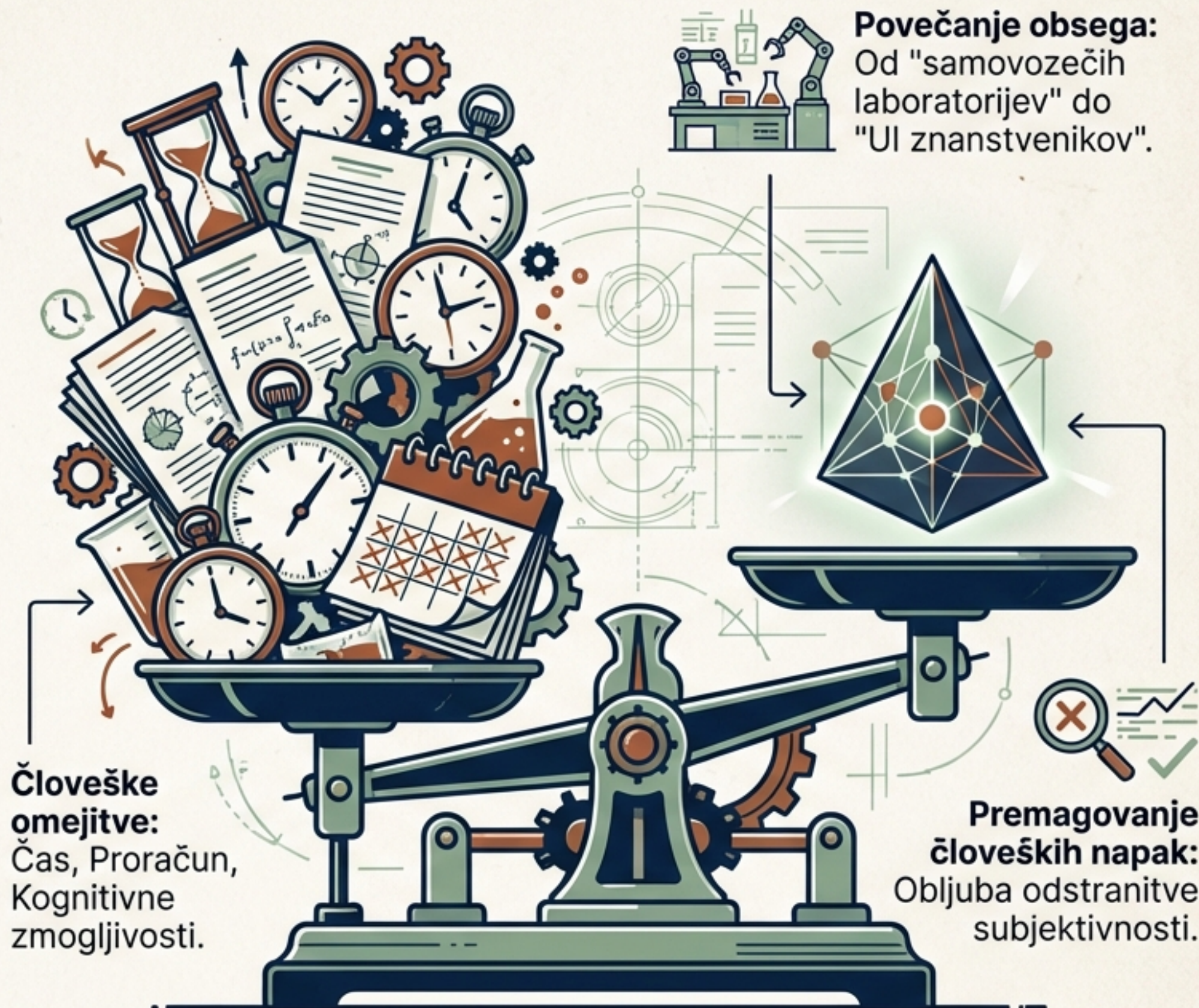
READING
DECK

Privlačnost "nadčloveške" znanosti

Znanstveniki se soočajo z omejitvami časa, proračuna in kognitivnih zmogljivosti. UI obljublja premagovanje teh ovir s povečanjem produktivnosti in objektivnosti.

Key Insight

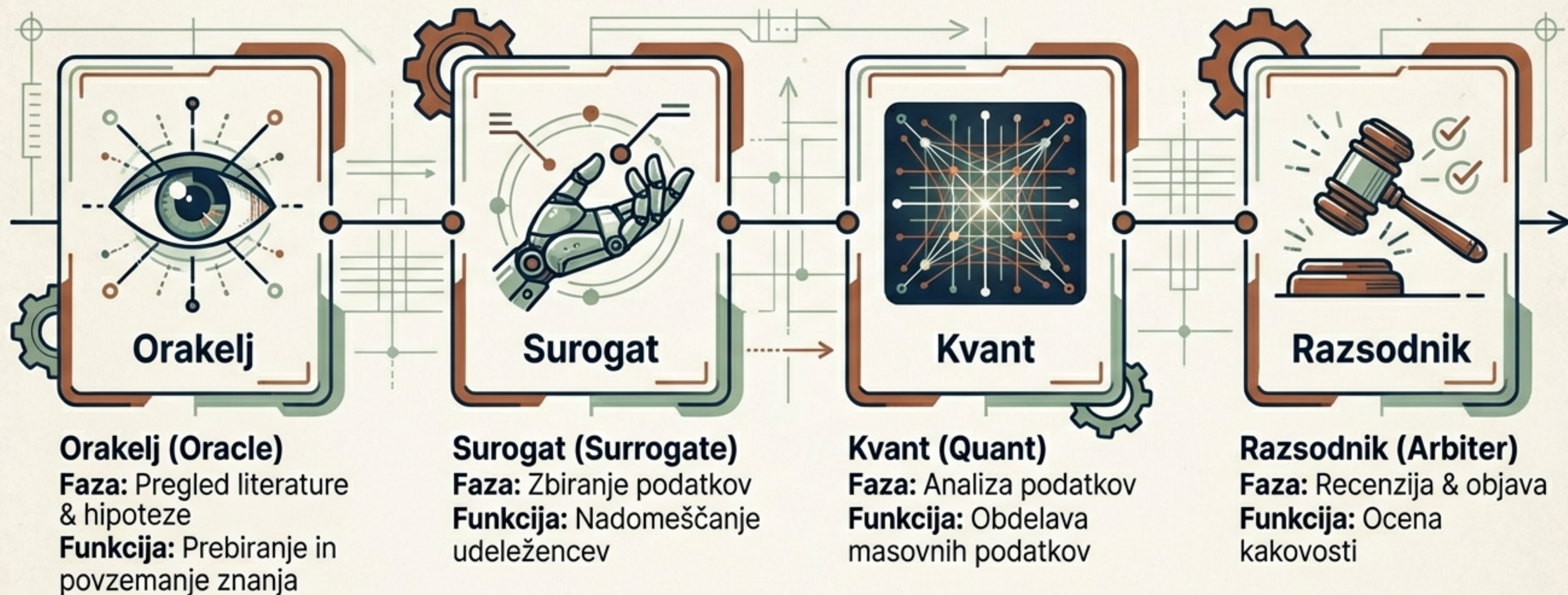
Vizije ne vidijo UI le kot orodja, temveč kot avtonomnega partnerja (collaborator), ki presega človeške zmožnosti.



Človeške omejitve:
Čas, Proračun,
Kognitivne
zmogljivosti.

Premagovanje
človeških napak:
Obljuba odstranitve
subjektivnosti.

Taksonomija vizij: Štiri vloge umetne inteligence



Te vloge prepletajo prediktivno UI (napovedovanje) in generativno UI (ustvarjanje).

Orakelj in Surogat: UI na začetku raziskovalnega procesa

Orakelj (Oracle)

Problem:

Preobremenjenost z informacijami in eksponentna rast literature.

UI Rešitev:

Orodja, ki 'preberejo' in povzamejo tisoče člankov ter predlagajo hipoteze.



Surogat (Surrogate)

Problem:

Zbiranje podatkov je drago in dolgotrajno.

UI Rešitev:

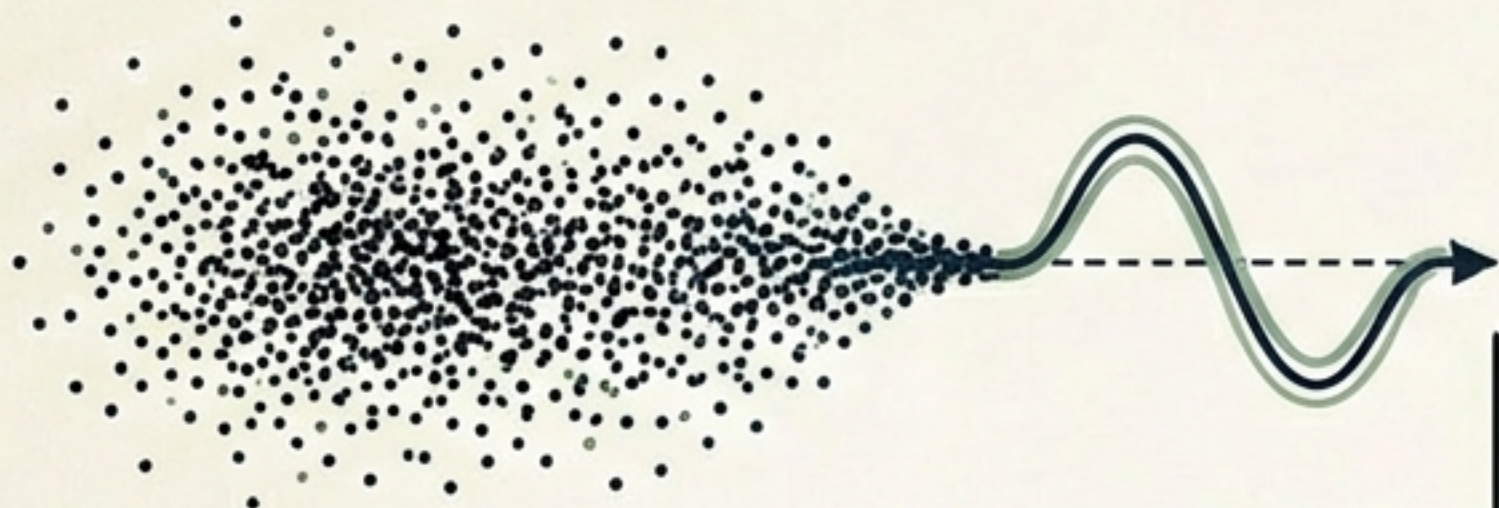
Generativna UI za simulacijo človeških odzivov ('silicon subjects') ali sintetični podatki.



“ Ali orodja res 'razumejo' literaturo, ali jo le procesirajo? ”

Kvant in Razsodnik: UI pri analizi in vrednotenju

Kvant (Quant)



Problem: Podatkovni nabori so preveliki za človeški um ("Big Data").

UI Rešitev: Odkrivanje vzorcev v visokodimenzionalnih prostorih in redukcija kompleksnosti.

Razsodnik (Arbiter)

Problem: Prenasičenost recenzentov in pristranskost.

UI Rešitev: Avtomatizirana ocena rokopisov in napovedovanje ponovljivosti.



Tveganje: Zanašanje na napovedno točnost (predictive accuracy) namesto na razlagalno moč.

Zakaj zaupamo strojem? Epistemsko zaupanje

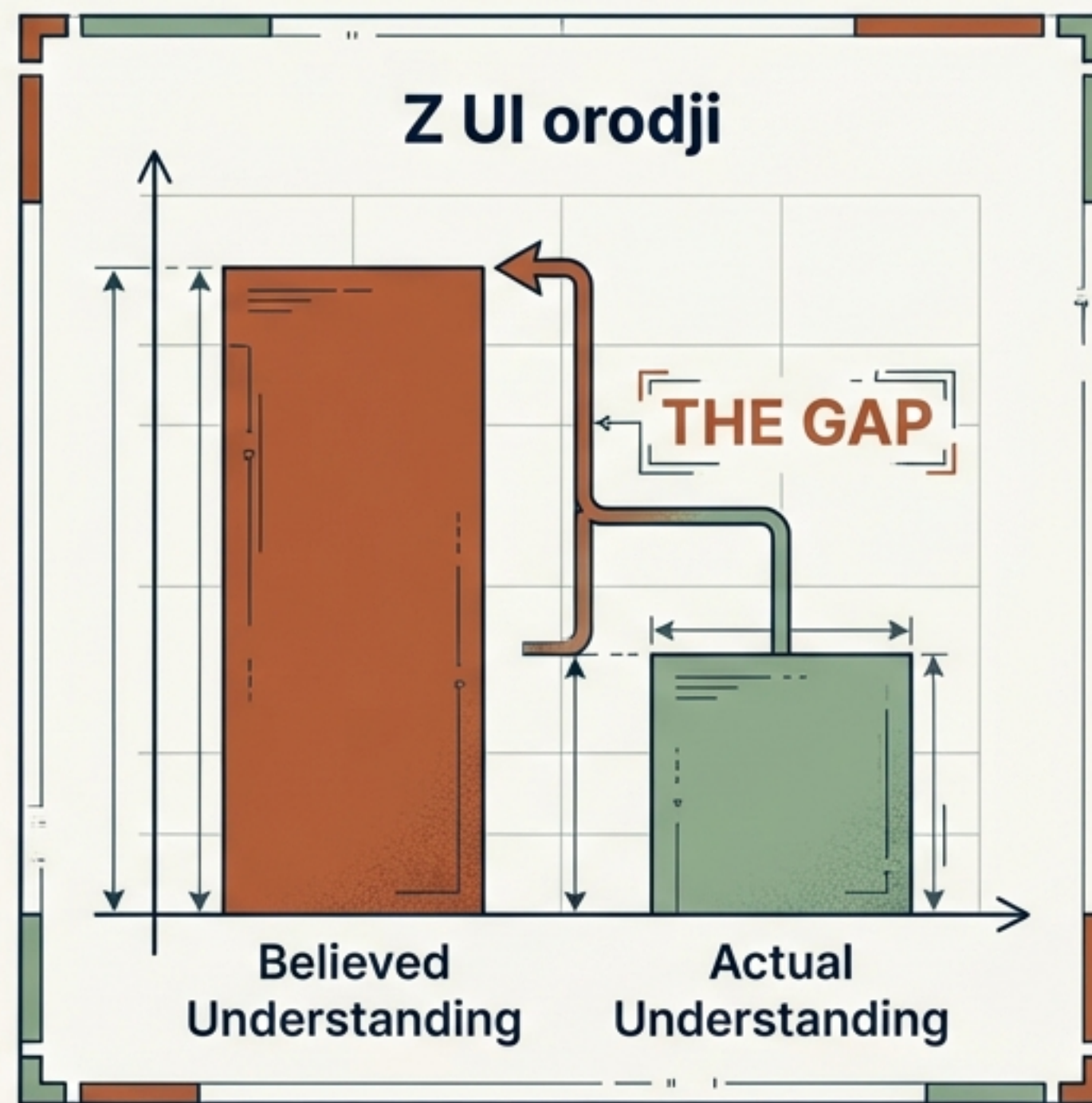
Znanstveniki vključujejo UI v svoje 'skupnosti vednosti', ker orodjem pripisujejo človeške lastnosti.



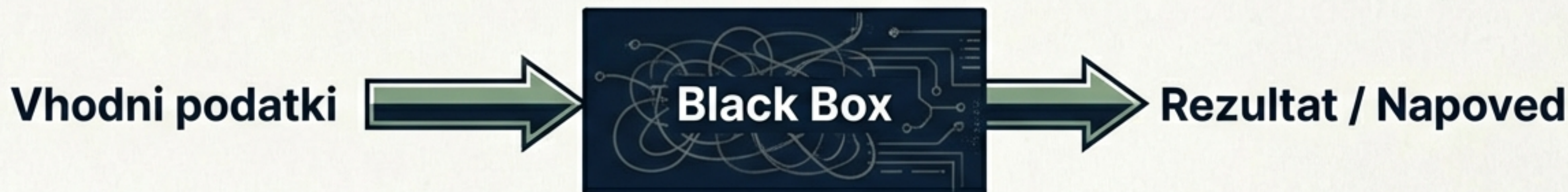
Iluzije razumevanja: Metakognitivna napaka

Definicija: Iluzija razumevanja nastane, ko zamenjujemo predelavo informacij (UI) z lastnim razumevanjem.

Mehanizem: Ko kognitivno delo preložimo na stroj, napačno pripišemo znanje stroja sebi.



Iluzija razlagalne globine (Explanatory Depth)



Problem črne skrinjice:

Znanost temelji na razlagi (*zakaj*), ne le na napovedi (*kaj*).

Zmota (The Fallacy): Znanstveniki zamenjujejo visoko napovedno točnost modela z vzročnim razumevanjem pojava. Več produktivnosti, a manj vpogleda.

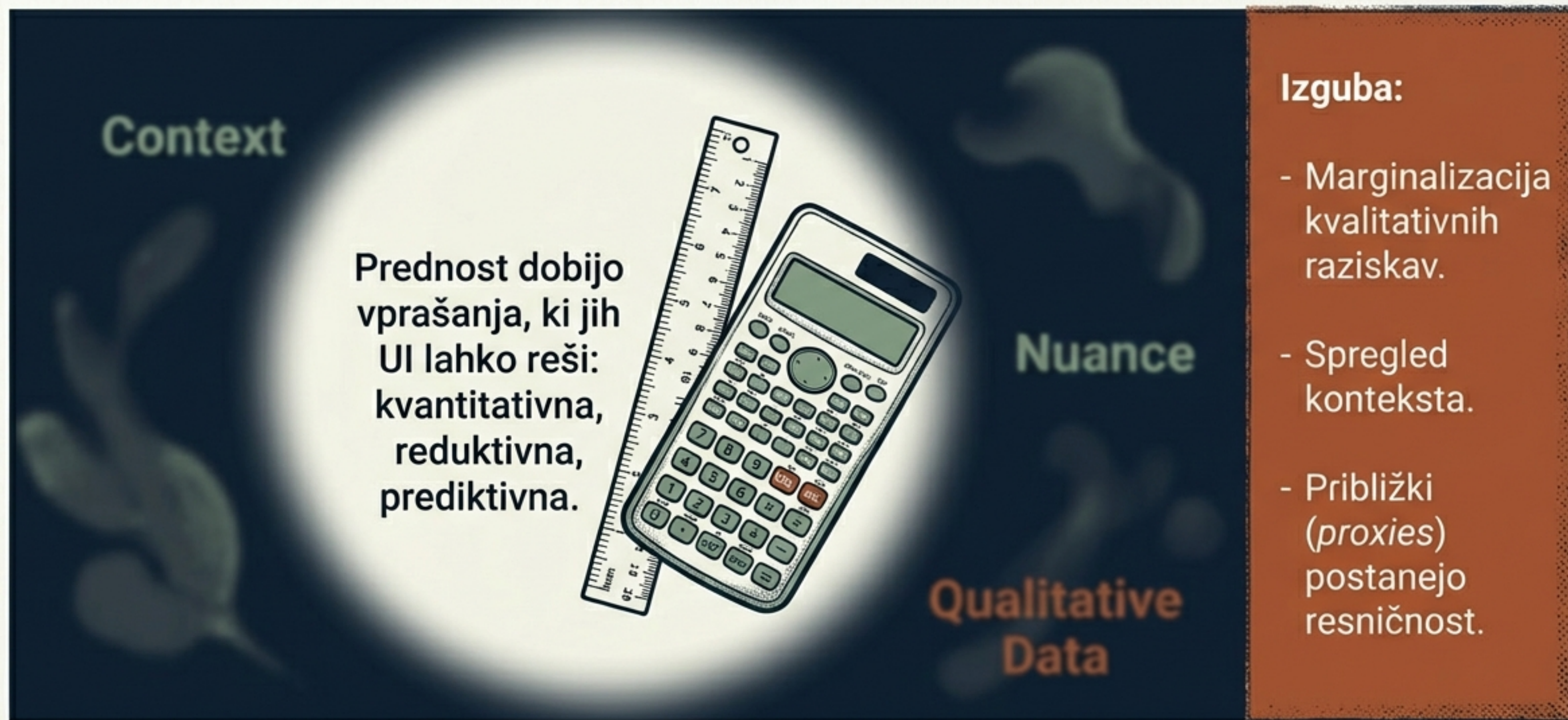
Od iluzij do monokultur: Sistemsko tveganje

Visoka učinkovitost, visoka ranljivost.

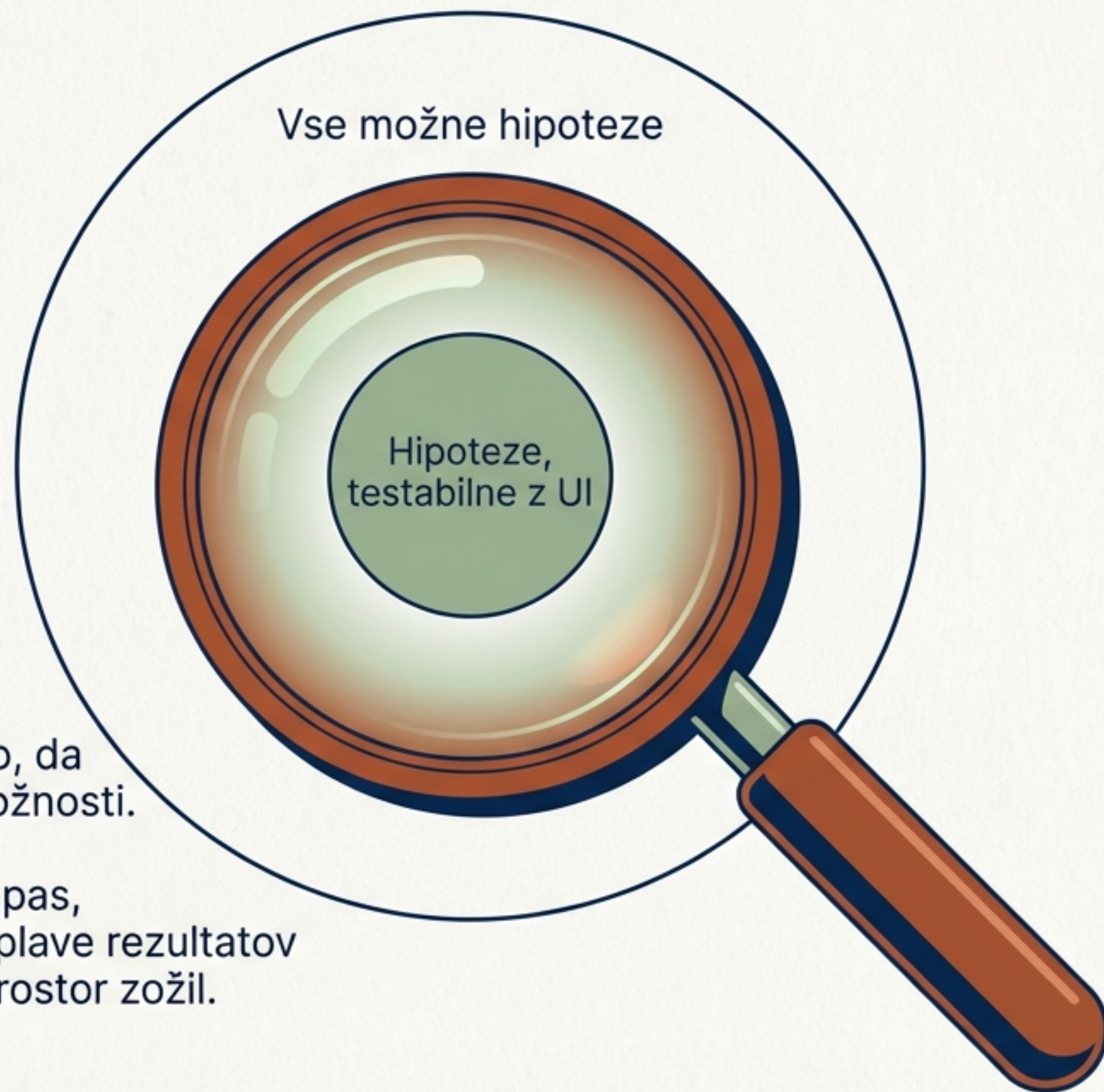
Tako kot v kmetijstvu, 'znanstvene monokulture' naredijo znanost ranljivo za napake.

Organska raznolikost, odpornost.

Monokultura vednosti: Ko orodje določa vprašanje



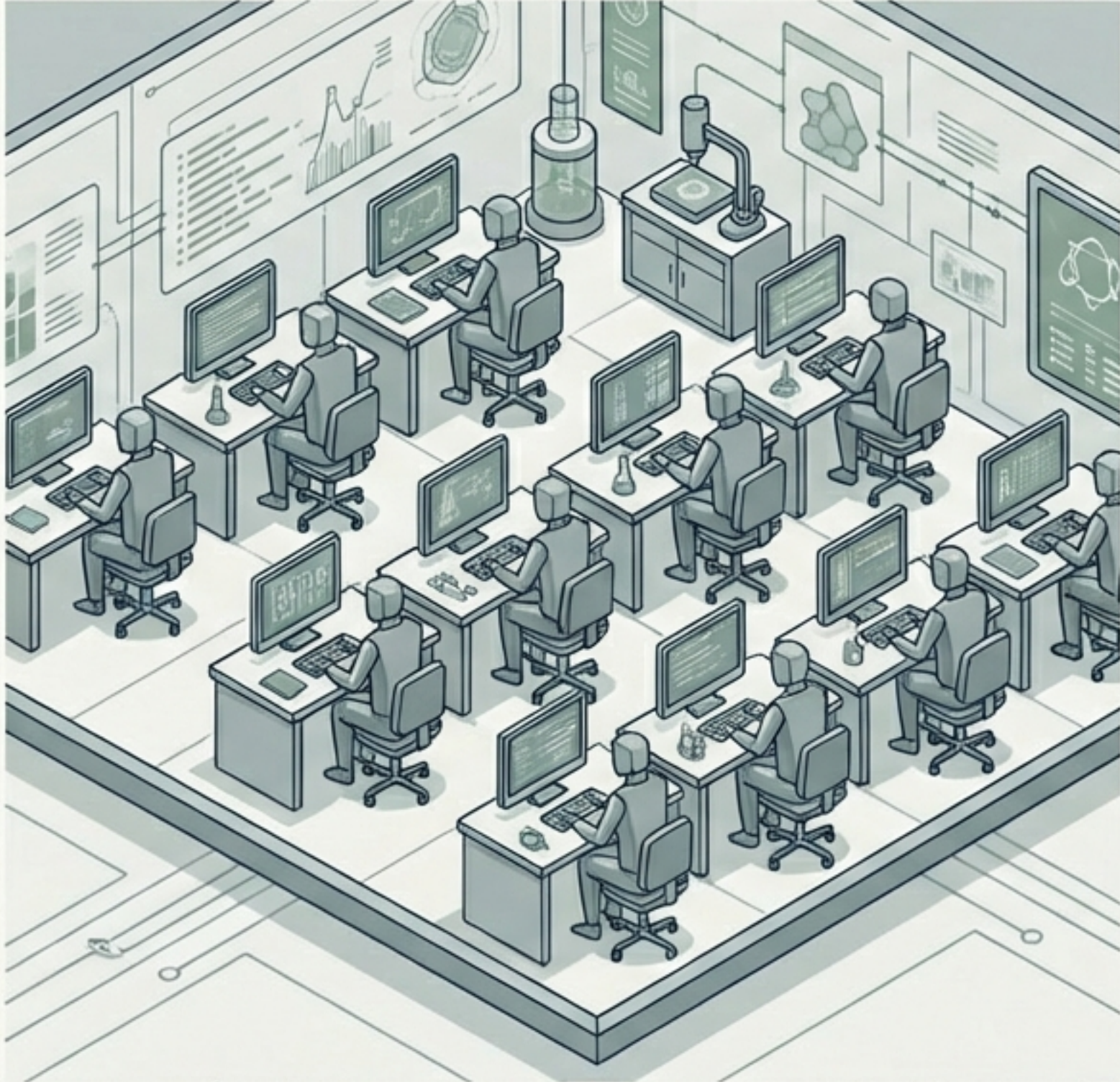
Iluzija raziskovalne širine (Exploratory Breadth)



Iluzija: Znanstveniki verjamejo, da raziskujejo celoten prostor možnosti.

Realnost: Raziskujejo le ozek pas, prilagojen orodjem. Zaradi poplave rezultatov spregledajo, da se je iskalni prostor zožil.

Monokultura poznavalcev: Izguba kognitivne raznolikosti



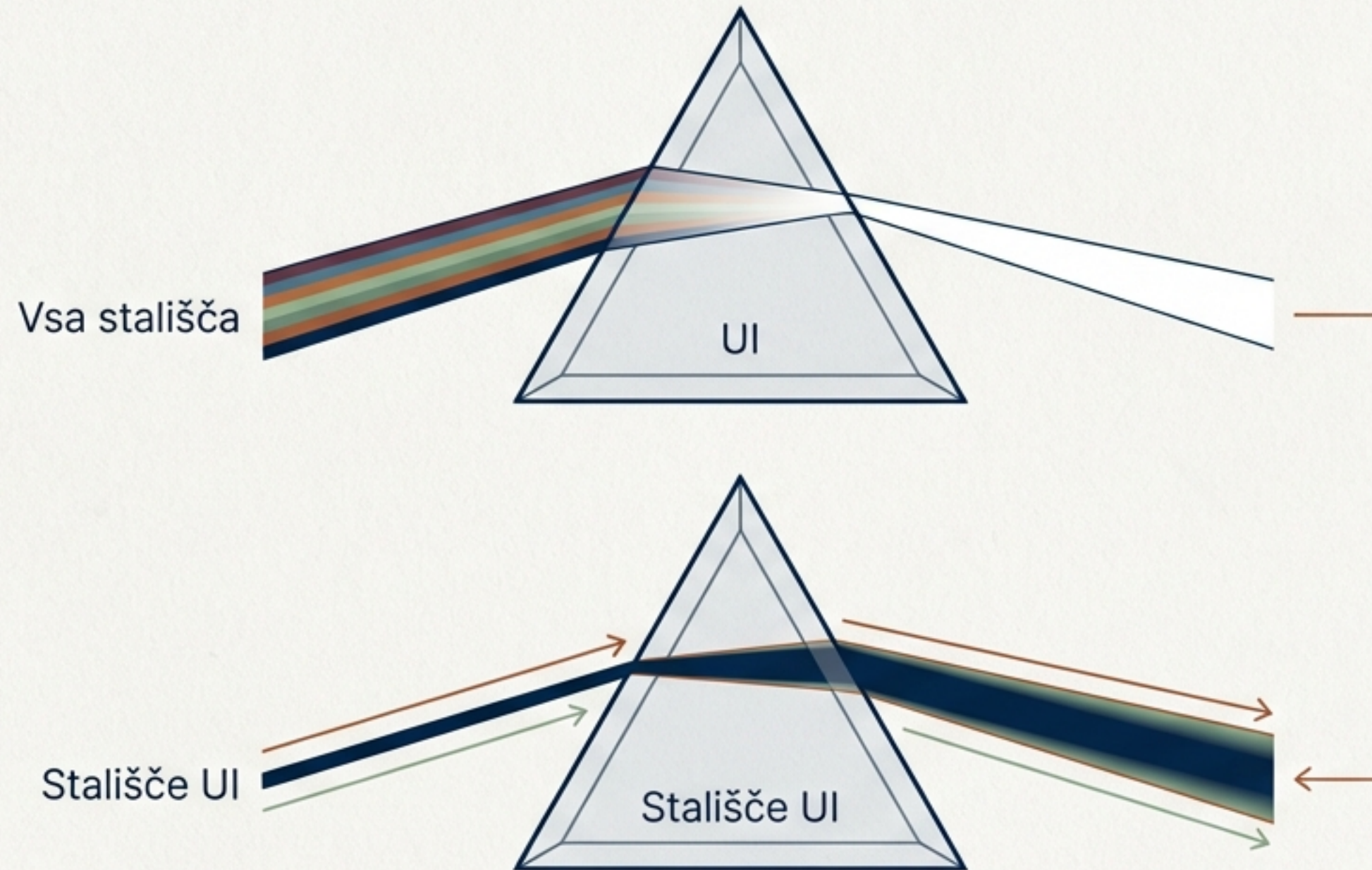
Koncept:

UI orodja se tržijo kot 'univerzalna', a vsebujejo pristranskosti svojih ustvarjalcev (večinoma zahodnih tehnoloških elit).

Posledice:

- Zmanjšanje demografske in kognitivne raznolikosti.
- Nadomeščanje raznolikih ljudi s standardiziranimi algoritmi.
- Izguba 'močne objektivnosti', ki izhaja iz trenja različnih perspektiv.

Iluzija objektivnosti



Past: Prepričanje, da UI nima stališča ali da zajame vsa stališča.

Realnost: UI ustvarja 'šibko objektivnost'. S tem ko zaupamo UI, tvegamo ponovno homogenizacijo znanosti in skrivanje pristranskosti.

Pot naprej: Strategije za odgovorno znanost



Ohranjanje raznolikosti

Kognitivna in demografska raznolikost ekip je obramba pred monokulturami.



Človek v zanki (Human-in-the-loop)

Ohranjanje človeške presoje in odgovornosti za interpretacijo.



Interdisciplinarnost

Vključevanje humanistike za razumevanje omejitev orodij.



Transparentnost

Jasno zavedanje, katiro 'vizijo' uporabljamo in kakšna so tveganja.



Več ne pomeni nujno bolje

Cilj znanosti ni zgolj produktivnost, temveč **razumevanje**.

UI je močno orodje, vendar ga ne smemo zamenjati za znanstvenika. Nevarnost ni v tem, da bi UI prevzela svet, ampak v tem, da bi mi pozabili, kaj pomeni resnično razumeti svet.

Gojite raznolikost. Ohranite kritično distanco.