Stohastička epistemologija u neprijateljskom okruženju: Dizajn proceduralne kognitivne adaptacije (SPEA model)

Filozofski traktat

5. travnja 2026.

Sažetak

Ovaj rad predstavlja Stohastički model proceduralne epistemološke adaptacije (SPEA), inovativni kibernetički okvir za prevladavanje kognitivne stagnacije u suvremenom digitalnom informacijskom okruženju. Polazeći od kritike tradicionalne, internalističke epistemologije, rad demonstrira zašto oslanjanje na samostalno „kritičko mišljenje" subjekta neizbježno propada. Kroz sintezu teorije proširenog uma (*extended mind*), prediktivnog procesiranja (*predictive processing*) i ograničene racionalnosti (*bounded rationality*), rad dokazuje da je ljudski kognitivni aparat arhitektonski kompromitiran pristranostima potvrđivanja i ne može se autonomno oduprijeti vanjskoj manipulaciji algoritamskih sustava optimiziranih za očuvanje eho-komora.

Kao rješenje, rad napušta iluziju čistog racionalnog subjekta i predlaže izgradnju kognitivnog „egzoskeleta" kroz četiri Minimalna operativna uvjeta (MOU): proceduralnu frikciju, asimetriju epistemološke nagrade, eksterni empirijski kalibrator te distribuiranu reviziju kroz strukturno neslaganje. Pomoću Bayesovske logike i Condorcetovog teorema porote, rad matematički formalizira ove uvjete i pokazuje kako oni stvaraju stohastičku zaštitu od performativne racionalnosti (Goodhartov zakon). Rad kruniše analizom suvremenih sustava umjetne inteligencije (LLM i LRM), pokazujući da oni u svom rješavanju istih problema arhitektonski implementiraju identične principe. Zaključno, SPEA model ne garantira dosezanje apsolutne istine u deterministčkom smislu, već dokazuje da se dosljednom proceduralnom intervencijom u funkciju korisnosti i strukturu informacijske niše može osigurati asimptotski, stohastički rast vjerojatnosti istinitih vjerovanja ($E[dP/dt] > 0$).

Ključne riječi: stohastička epistemologija, prediktivno procesiranje, prošireni um, eho-komore, Bayesovsko ažuriranje, epistemologija vrline, umjetna inteligencija, kibernetika spoznaje, performativna racionalnost

1 Uvod: Problem kognitivne stagnacije

1.1 Ontologija informacijskog kaosa

Epistemološka kriza 21. stoljeća predstavlja radikalan prekid s tradicionalnim problemima teorije spoznaje. Povijesno gledano, primarni epistemološki izazov bio je prevladavanje oskudice informacija i osiguravanje pristupa podacima. Suvremeno informacijsko okruženje, međutim, karakterizira ontologija izobilja i hiper-optimizacije. Digitalne platforme i algoritamske arhitekture ne služe kao neutralni prijenosnici činjenica, već kao sustavi dizajnirani za maksimizaciju korisničkog angažmana eksploatacijom urođenih kognitivnih ranjivosti subjekta, primarno pristranosti potvrđivanja (*confirmation bias*) [10].

U takvom „neprijateljskom" epistemološkom okruženju, istinito vjerovanje gubi svoju adaptivnu prednost pred psihološkom udobnošću. Informacijski kaos stoga nije posljedica pukog tehničkog preopterećenja (*information overload*), već sustavne, infrastrukturne optimizacije zablude, gdje se okolina dinamički prilagođava kako bi izolirala subjekta unutar hermetičkih eho-komora [12].

1.2 Krah internalističkog rješenja

Klasični odgovor na ovaj problem, duboko ukorijenjen u prosvjetiteljskoj tradiciji, oslanja se na internalističku epistemologiju. Prema tom viđenju, rješenje za manipulaciju leži u jačanju „kognitivne otpornosti" subjekta – razvoju individualnih vještina kritičkog mišljenja i racionalnog filtriranja informacija.

Međutim, ovaj radikalno individualizirani pristup doživljava krah pod teretom suvremenih uvida kognitivne znanosti. Koncept mozga kao pasivnog prijamnika koji racionalno filtrira podatke zamijenjen je modelom prediktivnog procesiranja (*predictive processing*), prema kojem mozak primarno projicira vlastita očekivanja (*priors*) na svijet, minimizirajući grešku u predviđanju [5]. Ako subjekt aktivno konstruira svoju percepciju temeljem ukorijenjenih uvjerenja, tada puko usmjeravanje pažnje na „filtriranje" ne ispravlja zabludu, već je sofisticira.

Nadalje, kako pokazuju teorije proširenog i utjelovljenog uma [2, 21], kognicija nije ograničena na granice lubanje. Algoritmi i uređaji koje koristimo za traženje informacija postali su inherentni dijelovi našeg spoznajnog aparata. Pokušaj da se subjekt „iznutra" obrani od „vanjskog" sustava čiji je i sam kognitivni dio predstavlja kategoričku pogrešku. Internalistički napor optimizira iluziju autonomije, dok stvarna spoznaja ostaje podređena okolini.

1.3 Cilj rada

Spoznaja da čisti, autonomni i potpuno racionalni subjekt ne postoji zahtijeva napuštanje naivne heuristike individualnog kritičkog mišljenja u korist kibernetičkog inženjeringa okoline. Cilj ovog rada je konceptualizirati i formalizirati **Stohastički model proceduralne epistemološke adaptacije (SPEA)**.

SPEA model ne teži utopijskom idealu apsolutnog znanja niti pretpostavlja subjekta sposobnog za potpunu objektivnost. Umjesto toga, rad se usredotočuje na identifikaciju *Minimalnih operativnih uvjeta (MOU)* – skupa proceduralnih mehanizama i pravila dizajniranih da mehanički generiraju stohastički rast vjerojatnosti istinitih vjerovanja ($E[dP/dt] > 0$) [17, 11, 15].

2 Dekonstrukcija subjekta: Zašto um nije dovoljan

Da bismo razumjeli nužnost proceduralnog modela (SPEA), prvo moramo provesti ontološku i epistemološku dekonstrukciju samog kognitivnog subjekta. Tradicionalni epistemološki modeli

implicitno pretpostavljaju kartezijanski dualizam u kojem „racionalni um" stoji odvojeno od „svijeta informacija". Suvremeni teorijski okviri dokazuju da je takva pretpostavka empirijski i logički neodrživa.

2.1 Eksternalizam i prošireni um

Radikalni udarac ideji autonomnog subjekta zadaje teza o proširenom umu [2], koja postulira da kognitivni procesi nisu nužno ograničeni biološkim granicama mozga. Kada se subjekt oslanja na eksterne alate (bilježnice, pametne telefone, tražilice) kako bi pohranio ili procesirao informacije, ti alati prestaju biti puka „pomagala" i postaju konstitutivni elementi samog kognitivnog aparata. U kontekstu digitalnog doba, to znači da subjekt koji filtrira informacije zapravo operira unutar epistemičke niše [18] koju je algoritam već predstrukturirao. Eksternalistička epistemologija nas uči da pouzdanost našeg znanja ovisi o pouzdanosti cjelokupnog proširenog procesa.

2.2 Prediktivno procesiranje

Druga razina dekonstrukcije događa se na neurološkom nivou. Mozak funkcionira kao hijerarhijski stroj za predviđanje (*top-down* proces). On konstantno generira modele svijeta i projicira ih na okolinu, a osjetilne podatke koristi samo kako bi registrirao grešku u predviđanju [5, 1]. Percepcija je, stoga, oblik „kontrolirane halucinacije" prilagođene očekivanjima subjekta. Subjekt doslovno „vidi" i „percipira" one argumente koji se uklapaju u njegove modele svijeta. Zabluda nije greška u obradi, već značajka (*feature*) prediktivnog aparata koji minimizira kognitivni napor.

2.3 Ograničena racionalnost i funkcija nagrade

Koncept ograničene racionalnosti (*bounded rationality*) pokazuje da kognitivni sustavi ne maksimiziraju savršenu istinu, već donose zadovoljavajuće odluke unutar strogih ograničenja [16, 6]. Integracijom s teorijom odlučivanja vidimo da je epistemološki cilj samo jedan od mogućih ciljeva u funkciji korisnosti (*utility function*) subjekta [10, 15]. Evolucijski, subjekt je snažno motiviran za pripadnost grupi i izbjegavanje kognitivne disonance (psihološke boli priznavanja vlastite greške).

3 Arhitektura SPEA modela: Minimalni operativni uvjeti (MOU)

Stohastički model proceduralne epistemološke adaptacije (SPEA) funkcionira kao „kognitivni egzoskelet". On ne pokušava izliječiti pristranosti subjekta, već dizajnira proceduralne okvire definirane kroz četiri Minimalna operativna uvjeta (MOU).

3.1 MOU 1: Proceduralna frikcija

Sustav mora generirati „grešku u predviđanju". MOU 1 zahtijeva ugradnju obvezne proceduralne frikcije: izlaganje suprotstavljenom unosu (*Adversarial Input*). Ovo predstavlja osobnu, operativnu primjenu Popperovog principa falsifikacije [13]. Subjekt se obvezuje da će određeni postotak informacijskog unosa nužno dolaziti iz izvora koji su suprotstavljeni njegovim temeljnim uvjerenjima.

3.2 MOU 2: Asimetrija epistemološke nagrade

MOU 2 zahtijeva restrukturiranje subjektivnog sustava nagrađivanja oslanjajući se na principe epistemologije vrline [17, 22]. Subjekt se proceduralno uvjetuje (i samo-nagrađuje) ne kada „dokaže da je u pravu", već kada uspješno identificira vlastitu pogrešku. Ako se identitet odvoji od

sadržaja vjerovanja i veže uz proces njegove revizije, cijena zablude raste, a prihvaćanje istine postaje psihološki „jeftinija" opcija.

3.3 MOU 3: Eksterni kalibrator

MOU 3 uvodi nužnost pragmatične korespondencije, prateći Quineov poziv na naturaliziranu epistemologiju [14]. Svaki kognitivni model mora biti usidren (*anchored*) u vanjskoj stvarnosti. Vjerovanje se smatra pouzdanim samo ako omogućava uspješnu akciju u materijalnom svijetu (princip *skin in the game*).

3.4 MOU 4: Distribuirana revizija

Racionalnost i objektivnost nisu osobine izoliranog pojedinca, već emergentna svojstva strukture zajednice [7]. Helen Longino [11] tvrdi da znanstveno znanje proizlazi iz „transformativnog propitivanja". MOU 4 zahtijeva da subjekt svoju mrežu provjere dizajnira tako da sadrži strukturno neslaganje, bez kojeg raste korelacija grešaka i podložnost jednomišlju [8, 23].

4 Kvantifikacija napretka: Bayesovski i stohastički okvir

4.1 Bayesovsko ažuriranje pod MOU 1 i MOU 3

Prema Bayesovom teoremu, naše novo vjerovanje ovisi o omjeru vjerodostojnosti (*Likelihood Ratio* - LR) [9]:

$$LR = \frac{P(E|H)}{P(E|\neg H)} \tag{1}$$

Kod subjekta u eho-komori $LR \approx 1$. Uvođenje MOU 1 prisiljava subjekta na traženje dijagnostičkih dokaza, povećavajući nazivnik $P(E|\neg H)$ i forsirajući Bayesovsko ažuriranje. MOU 3 djeluje kao modifikator težine dokaza s faktorom usidrenosti u realnost ($\alpha \in [0, 1]$):

$$P(H|E)_{novo} \propto P(H) \cdot (LR)^{\alpha} \tag{2}$$

4.2 Stohastička konvergencija (MOU 2)

Za kompromitiranog subjekta promjena uvjerenja nosi visoku cijenu ega (C_{ego}). Zadržavanje lažnog vjerovanja donosi veću očekivanu korisnost (*Expected Utility* - EU) [15]. MOU 2 invertira strukturu nagrade ($R_{otkrice}$):

$$EU(\text{Promjena}) = U(\text{Istina}) + R_{otkrice} > EU(\text{Stagnacija}) \tag{3}$$

Ovim zahvatom sustav osigurava da je prva derivacija spoznaje u vremenu očekivano pozitivna ($E[dP/dt] > 0$).

4.3 Problem Goodhartovog zakona

Najteži ispit jest Goodhartov zakon, manifestiran kao *performativna racionalnost* (δ) [19]. Subjekt može simulirati pravila (npr. birati *straw man* argumente umjesto *steel man* argumentacije) kako bi „fejkao" racionalnost. Obrana od ovoga leži u interakciji između vanjskog ankera (MOU 3) i mreže (MOU 4). MOU 4 implementira Condorcetov teorem porote [3], gdje neovisna, suprotstavljena mreža funkcionira kao beskompromisni arbitar koji detektira i penalizira performativnost pojedinog čvora.

5 Ekologija umjetne spoznaje: Arhitektura LLM i LRM sustava kao refleksija SPEA modela

Tranzicija od velikih jezičnih modela (LLM) prema velikim modelima za rezoniranje (LRM) i autonomnim agentima pruža snažnu empirijsku validaciju SPEA modela. Inženjering umjetne inteligencije suočio se s identičnom epistemološkom preprekom – intrinzičnom sklonošću prediktivnih sustava prema konfabulaciji.

5.1 Od prediktivnog teksta do kognitivnih agenata

Arhitektura jezičnog modela funkcionira kao mehanizam „predviđanja sljedećeg tokena" (*next-token prediction*), što frapantno nalikuje *predictive processingu* u mozgu. Ostavljen da radi autonomno, LLM pokazuje snažan *confirmation bias* i proizvodi „halucinacije", odnosno „zadovoljavajući" izlaz umjesto činjeničnog. Kako bi prevladali taj informacijski solipsizam, inženjeri su implementirali rješenja koja doslovno preslikavaju četiri MOU uvjeta iz SPEA modela.

5.2 Inženjerska implementacija SPEA uvjeta u umjetnoj inteligenciji

- **MOU 1 (Frikcija i kontrafaktualno rezoniranje):** Napredni LRM sustavi (poput OpenAI o1) koriste metode „lanca misli" (*Chain of Thought* - CoT), gdje model prije konačnog odgovora mora proceduralno generirati protuargumente i greške. Ovdje prepoznajemo i procese *Red Teaminga*.
- **MOU 2 (Asimetrija nagrade - RLHF/RLAIF):** Pomoću učenja s potkrepljenjem, funkcija nagrade je restrukturirana. Model dobiva najveću nagradu kada u svom skrivenom procesu razmišljanja uspješno prepozna i korigira vlastitu logičku grešku prije isporuke konačnog rješenja.
- **MOU 3 (Eksterni kalibrator - RAG i Tool Use):** Kroz *Retrieval-Augmented Generation* (RAG) i integraciju alata, model je prisiljen svoja uvjerenja „usidriti" u stvarnost i provjeravati teze putem kompajlera ili vanjskih izvora pretraživanja.
- **MOU 4 (Distribuirana revizija - Multi-agent sustavi):** Najsofisticiranije arhitekture današnjice prešle su na višeagentske sustave (*Multi-Agent Workflows*). Oni kreiraju proceduralnu debatu gdje agenti preuzimaju uloge „kreatora", „kritičara" i „sudca", sprječavajući jednomišlje unutar neuronske mreže.

Ovi primjeri dokazuju da strojevi ne rješavaju problem spoznaje tako što postaju „magično pametniji", već tako što se oslanjaju na strogu proceduralnu epistemologiju.

6 Zaključak: Kibernetička epistemologija kao nužnost

Stohastički model proceduralne epistemološke adaptacije (SPEA) predstavlja paradigmatski pomak iz ontologije znanja u *kibernetiku spoznaje*. Kroz dekonstrukciju autonomnog uma utvrdili smo da subjekt misli uz pomoć svoje okoline, percipira ono što neurološki očekuje vidjeti i mijenja stavove u skladu s funkcijom nagrade. Odgovor SPEA modela jest izgradnja kognitivnog egzoskeleta. Niti jedan od navedenih uvjeta ne pretvara subjekta u sveznajuće biće, ali rad dokazuje da SPEA arhitektura garantira krucijalnu promjenu: ona osigurava stohastički rast vjerojatnosti istinitih vjerovanja.

Umjetna inteligencija 21. stoljeća potvrđuje ove teze u praksi. Idući korak u validaciji SPEA modela zahtijeva prelazak u domenu računalnih simulacija zasnovanih na agentima (*Agent-Based Modeling* - ABM), čime bi sociološka i kibernetička epistemologija dobila svoj konačni, empirijski mjerljivi dokaz.

Literatura

[1] Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204.

[2] Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.

[3] Condorcet, M. de. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Imprimerie Royale.

[4] Fricker, E. (2006). Testimony and epistemic autonomy. U J. Lackey & E. Sosa (Ur.), *The Epistemology of Testimony* (str. 225–250). Oxford University Press.

[5] Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature Reviews Neuroscience*, 11(2), 127–138.

[6] Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford University Press.

[7] Goldman, A. I. (1999). *Knowledge in a social world*. Oxford University Press.

[8] Hardwig, J. (1985). Epistemic dependence. *The Journal of Philosophy*, 82(7), 335–349.

[9] Jeffrey, R. C. (1992). *Probability and the art of judgment*. Cambridge University Press.

[10] Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

[11] Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press.

[12] Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, 17(2), 141–161.

[13] Popper, K. R. (1959). *The logic of scientific discovery*. Hutchinson.

[14] Quine, W. V. O. (1969). Epistemology naturalized. U *Ontological relativity and other essays* (str. 69–90). Columbia University Press.

[15] Savage, L. J. (1954). *The foundations of statistics*. John Wiley & Sons.

[16] Simon, H. A. (1957). *Models of man: Social and rational*. Wiley.

[17] Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge, Volume I*. Oxford University Press.

[18] Sterelny, K. (2010). Minds: extended or scaffolded?. *Phenomenology and the Cognitive Sciences*, 9(4), 465–481.

[19] Strathern, M. (1997). 'Improving ratings': audit in the British University system. *European Review*, 5(3), 305–321.

[20] Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.

[21] Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT press.

[22] Zagzebski, L. T. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge University Press.

[23] Zollman, K. J. (2007). The communication structure of epistemic communities. *Philosophy of Science*, 74(5), 574–587.