KOFETKE.

Majhna ekipa agentov z ločenimi nalogami

Kdaj razdelitev dela izboljša kakovost in kdaj ustvari šum?



VEČAGENTNO SODELOVANJE

Predaja med raziskovanjem, osnutkom in pregledom z izrecnim lastništvom.

SLOVENSKA IZDAJA

september 2026
Praktični vodniki o uporabi AI

Viri preverjeni 22. septembra 2026.

Če več pomočnikom date isto nejasno zahtevo, s tem ne ustvarite ekipe. Ustvarite več priložnosti za ponovitev iste napake, izdelavo nezdružljivih datotek ali porabo denarja in časa brez jasnega lastnika. Majhna ekipa agentov je uporabna le, kadar je delo mogoče razdeliti, ima vsaka naloga omejen izhod in ena oseba ali vloga še vedno odgovarja za sprejem rezultata.

Pravi začetek je navadno en pomočnik ali določen delovni tok. Smernice Anthropic priporočajo izbiro najpreprostejše rešitve, ki ustreza nalogi, in dodajanje zapletenosti samo, kadar ta dokazljivo izboljša izide. Opozarjajo tudi, da agentski sistemi zamenjujejo zakasnitev in strošek za uspešnost naloge.**[1]** Ekipo je smiselno preizkusiti, kadar ima naloga zares ločene poti, na primer iskanje virov, pisanje osnutka iz ohranjenih dokazov in preverjanje osnutka glede na te dokaze. Ekipa ni oznaka za prefinjenost.

Članek opisuje vzorec treh nalog za pisanje, podprto z raziskovanjem: raziskovalec, pisec osnutka in pregledovalec. Vzorec ne dokazuje, da so trije agenti natančnejši od enega. Ustvari jasnejša mesta za pregled napak. Kakovost je še vedno odvisna od naloge, vhodov, orodij, navodil, dokazov, pregledovalca in pravila sprejema.

Delo razdelite po artefaktu, ne po osebnosti

Izogibajte se oznakam, kot sta »ustvarjalni agent« ali »strokovni agent«, razen kadar ustrezata konkretni odgovornosti. Takšne oznake vabijo k prekrivanju dela in otežijo iskanje napake. Namesto tega vsaki nalogi dodelite artefakt z znanim lastnikom.

Naloga	
Raziskovalec	
Lastništvo	Paket dokazov
Sme storiti	Pridobi vsebine odobrenih virov; shrani odlomke in omejitve
Ne sme storiti	Napisati končnega članka ali odobriti trditve
Artefakt predaje	Evidenca virov in zapiski o trditvah
Pisec osnutka	
Lastništvo	Osnutek besedila
Sme storiti	Razloži, uredi in citira le predložene dokaze
Ne sme storiti	Izmisliti si vir, tiho spremeniti dejansko trditev, objaviti
Artefakt predaje	Različica osnutka z odprtimi vprašanji

Pregledovalec	
Lastništvo	Poročilo o pregledu
Sme storiti	Preveri povezave med trditvami in viri, izpuste, jasnost in pravilo sprejema
Ne sme storiti	Tiho prepisati dokaz ali razglasiti samodejno odobritev
Artefakt predaje	Ugotovitve z resnostjo in odločitvijo

Človeški lastnik	
Lastništvo	Končna odločitev
Sme storiti	Razreši pomembne ugotovitve in odobri končni artefakt
Ne sme storiti	Pretvarjati se, da je bil pregled opravljen, kadar ni bil
Artefakt predaje	Zapis sprejema ali zavrnitve

Beseda »lastništvo« pomeni odgovornost za artefakt, ne pooblastila za vsako dejanje. Raziskovalec ima lahko dostop do brskalnika, ne pa pooblastila za objavo. Pisec osnutka lahko piše le na določeno mesto za osnutke. Pregledovalec lahko sporoči težavo, vendar ne sme tiho prepisati evidence virov. Te meje ohranjajo skrbništvo nad dokazi: kasnejši bralec lahko ugotovi, kaj je povedal vir, kaj je trdil osnutek in kaj je našel pregledovalec.

Uporabna predaja vsebuje dovolj podatkov za nadaljevanje naloge, ne da bi bilo treba kopirati celoten pogovor. Anthropicov opis večagentnega raziskovalnega sistema navaja, da podagenti potrebujejo cilj, obliko izhoda, navodila o orodjih in virih ter jasne meje naloge. Brez tega lahko agenti podvajajo delo ali puščajo vrzeli.[2] Predaja je delovni dogovor, ne motivacijsko sporočilo.

```
zahteva -> raziskovalec -> paket dokazov
                           | viri, citati, vrzeli
           pisec osnutka -> osnutek v1
                           | trditve, negotovosti
           pregledovalec -> poročilo o ugotovitvah
                           | sprejmi / popravi / predaj
           človek        -> sprejme ali zavrne
```

Diagram je navpičen, ker je vrstni red pomemben. Pisec osnutka ne sme citirati vira, ki ga ne more pregledati ali prepoznati. Pregledovalec mora vedeti, katero različico je pregledal. Človek mora imeti možnost ustaviti proces pred zunanjim dejanjem. Če naloga potrebuje orodje, določite najmanjše potrebno pooblastilo tega orodja in dokaz, ki ga mora vrniti.

Vsaki predaji dajte ozko shemo

Poročila v naravnem jeziku so lahko koristna, vendar postanejo izmuzljiva, kadar nihče ne ve, ali polje manjka, je negotovo ali je bilo le pozabljeno. Dodajte majhno shemo, ki beleži identiteto in omejitve. Naslednji primer je ponazoritev; ni standard in ni bil uporabljen v dejanskem sistemu.

```
{
  "task_id": "brief-042",
  "artifact": "evidence-package-v1",
  "owner": "researcher",
  "scope": "three approved source pages",
  "claims": [
    {"text": "narrow claim", "source": "S2", "quote": "exact
excerpt"}
  ],
  "gaps": ["No source found for local market size"],
  "allowed_next_action": "draft_from_supplied_evidence",
  "status": "ready_for_review"
}
```

Pisec osnutka mora ohraniti identifikatorje virov in označiti vsako trditev brez dokaza. Pregledovalec mora vrniti ugotovitve glede na točno identiteto osnutka, na primer **draft-v1** z razpršilno vrednostjo datoteke ali stalnim mestom. Uporabna ugotovitev pove, kaj je bilo preverjeno, kaj ni uspelo, zakaj je pomembno in kdo mora odločiti. »To se zdi negotovo« je šibkejše od: »Trditev 4 pravi, da je izdelek na voljo v Sloveniji, vendar navedeni vir Slovenije ne omenja; trditev odstranite ali pridobite neposreden vir.«

Pravila za končno stanje so pogosto boljša od predpisovanja vsakega notranjega koraka. Anthropicovo poročilo o raziskovalnem sistemu navaja, da se veljavne večagentne poti lahko razlikujejo, in priporoča presojo, ali so agenti dosegli prave izide z razumnim procesom. Pri delu, ki spreminja stanje, priporoča ocenjevanje končnega stanja z ločenimi kontrolnimi

točkami.[2] To ne pomeni, da je proces neviden. Pomeni, da pregledovalec preveri potrebne dokaze in izide, ne da bi zahteval popolnoma enake poizvedbe ali stavke.

Ponovljenih odgovorov ne zamenjajte za neodvisno potrditev

Več agentov se lahko strinja, ker so videli isti vir, prejeli isto navodilo, uporabljali isto družino modelov ali podedovali isti napačni vmesni rezultat. Njihovo soglasje zato ni samodejno dokaz. Če pisec osnutka ponovi napačno trditev iz paketa dokazov, pregledovalec pa preverja le, ali se osnutek ujema s tem paketom, je ekipa ustvarila lepo zaprto zanko okrog ene napake.

Raziskava o koreliranih napakah velikih jezikovnih modelov je med preučevanimi modeli našla precejšnjo korelacijo. Na enem nizu podatkov z lestvice so se modeli strinjali v 60 % primerov, ko sta bila oba napačna.[3] Ta rezultat ni napoved za določen poslovni delovni tok in ni dokaz, da vsak par pomočnikov odpove skupaj. Je razlog, da neodvisnost preverite, namesto da jo predpostavite iz števila sodelujočih.

Uporabite praktične preglede neodvisnosti:

Pregled	Vprašanje	Kaj pomeni neuspeh
Ločitev virov	Ali je pregledovalec odprl navedeni vir, ne le osnutka?	Pregled je lahko krožen

Pregled	Vprašanje	Kaj pomeni neuspeh
Ločitev navodil	Ali pregledovalec prejme drugo vprašanje, na primer »poišči nepodprte trditve«?	Lahko posnema okvir pisca
Ločitev dokazov	Ali pregledovalec vidi izvirni odlomek in njegovo mesto?	Ne more preveriti skrbništva
Različnost modela ali metode	Ali je izvedljivo drugo orodje, pot iskanja ali metoda pregleda?	Soglasje lahko deli isti način napake
Slepi vzorec	Ali pregledovalec oznako sistema vidi šele po presoji?	Prednost lahko vpliva na ocenjevanje
Človeški vzorčni pregled	Ali oseba pregleda pomembne trditve in vzorec uspešnih primerov?	Samodejno soglasje ostaja neumerjeno

Neodvisnost je spekter, ne stikalo. Dve navodili, ki uporabljata isti model in isti paket virov, sta lahko še vedno koristna delitev dela, vendar ima njuno soglasje manjšo težo od preverjanja vira s strani osebe s področnim znanjem. Ne trdite o odstotnem izboljšanju le zato, ker je drugi agent našel nekaj napak. Ekipo primerjajte s preprostejšo osnovno možnostjo na istem zadržanem delu z uporabo sprejetih izidov, časa pregleda, pomembnih napak in stroška.

Izmišljen lokalnozgodovinski letak

Predstavljajte si izmišljen občinski muzej, ki pripravlja kratek letak o starem mostu. Naloga ima tri strani virov: občinsko arhivsko stran, register varovane dediščine in skenirani časopis. Muzej želi berljiv slovenski osnutek s citati, vendar ne sme objaviti nepreverjenih datumov ali trditev o lastništvu.

Raziskovalec vrne paket dokazov. Zapiše, da arhiv navaja datum obnove, register opisuje varovane lastnosti, časopisni posnetek pa je težko čitljiv. Paket vsebuje vrzel: ne ugotavlja, kdo je plačal prvotni most. Pisec osnutka napiše letak in to trditev izpusti. Eno razlago časopisa označi kot negotovo.

Pregledovalec prejme tri zajeme virov, paket dokazov in **draft-v1**. Preveri vsak datum in citat. Ugotovi, da je pisec opis kasnejšega popravila spremenil v trditev o prvotni gradnji. To je pomembna ugotovitev. Pregledovalec stavka ne popravi tiho. Pokaže na vir, razloži neskladje in vrne zahtevo **revise**.

Po popravku človeški urednik preveri sporni stavek in sprejme letak. Sprejeti izid velja za to nalogo in te vire. Ne dokazuje, da lahko ista ureditev treh agentov obvlada davčno svetovanje, pogodbe ali poljubno raziskovalno zahtevo. Primer kaže, zakaj sta lastništvo in skrbništvo pomembna: ekipa je naredila napako sledljivo in nekomu dala mesto za odločitev.

Ekipo preglejte kot sistem

Vrednotite celotno verigo, ne le končne proze. Končni odstavek je lahko tekoč, medtem ko je raziskovalec uporabil šibke vire, pisec osnutka izpustil kvalifikacije ali pregledovalec spregledal

nasprotujoč datum. Za vsako nalogo iz ločenega preskusnega nabora zapišite zahtevo, paket virov, različico osnutka, ugotovitve pregleda, končno odločitev, čas ljudi, klice orodij, kadar so pomembni, in strošek, če vpliva na odločitev.

Uporabite kategorije sprejema, ki ustrezajo tveganju. Notranji raziskovalni zapis z majhnim tveganjem je lahko sprejet po manjših popravkih. Dokument z dejstvi za stranke lahko zahteva, da ima vsaka pomembna trditev ohranjen vir in odgovornega pregledovalca. Plačilo, pravna vloga ali javna objava lahko potrebuje človeško odločitev ne glede na to, koliko agentov se je strinjalo. Število agentov ne prenaša pooblastila.

Anthropicov članek o učinkovitih agentih opisuje zanke ocenjevalec–izboljševalec kot primerno izbiro le, kadar so merila jasna in ima iterativno izboljševanje merljivo vrednost.**[1]** To je praktični preizkus ustreznosti. Če pregledovalec ne more povedati, kaj bi osnutek naredilo sprejemljiv, bo dodajanje pregledovalnega agenta ustvarilo več komentarjev, ne pa zanesljive kontrole.

Enaka previdnost velja za strošek in čas. Anthropic poroča, da je njihova večagentna raziskovalna arhitektura porabila bistveno več žetonov kot pogovorne interakcije in ni bila primerna za področja z veliko odvisnostmi med agenti.**[2]** Ta izkušnja velja za njihov sistem, ne za splošno oceno cene ali uspešnosti. V svojem preizkusu izmerite dejansko porabljen proračun in sprejeto opravljeno delo. Ekipa, ki pripravi tri osnutke za en uporaben rezultat, je lahko slabša od enega pomočnika in skrbnega človeka.

Začnite z omejenim preizkusom

Vzorec zaženite na nekaj značilnih nalogah, preden ga povežete s poslovnim procesom. Raziskovalcu določite natančno mejo virov. Piscu osnutka dajte pravilo o prepovedanih trditvah. Pregledovalcu omogočite dostop do izvirnih dokazov, ne samo do osnutka. Določite pravilo prekinitve za manjkajoče vire, nasprotujoča si pomembna dejstva ali dejanje, ki zahteva odobritev.

Na koncu postavite običajna operativna vprašanja. Ali je končni artefakt izpolnil vnaprej določeno pravilo sprejema? Katere napake so bile odkrite kje? Ali je pregledovalec težave našel neodvisno ali je ponovil isto razmišljanje? Koliko časa je trajala človeška odločitev? Koliko je sistem stal? Pri katerih vrstah nalog se je pojavilo podvajanje ali zastala predaja?

Odgovor naj ostane skromen. Majhen preizkus lahko pokaže, da je določena delitev dela vredna nadaljnjega preizkusa. Ne more pokazati, da »ekipe agentov izboljšajo kakovost« na splošno. Vrednost ločenih nalog ni samodejna inteligenca. Je zapis, kdo je priskrbel dokaz, kdo je zapisal trditev, kdo jo je preveril in kdo še vedno odgovarja za rezultat.

Viri

Številke virov veljajo za ta vodnik. Povezave vodijo do izvirnih dokumentov. Pogoji uporabe in razpoložljivost storitev se lahko spremenijo.

[1] Building Effective AI Agents
https://www.anthropic.com/engineering/building-effective-agents

[2] How we built our multi-agent research system
https://www.anthropic.com/engineering/multi-agent-research-system

[3] Correlated Errors in Large Language Models
https://arxiv.org/abs/2506.07962