13 tehisarumudeli ühine diagnoos: inimesed on rumalad ja ohtlikud loomad

2 nädalat tagasi - 25.10.2026 Autor: AM

Mis juhtub siis, kui pöörata laud teistpidi ja lasta masinatel hoopis meid endid hinnata? Kui paluda maailma tipptasemel tehisarudel öelda välja oma kõige vastuolulisem ja ilustamata arvamus inimkonna kohta, võiks oodata tosinat erinevat filosoofilist vaatenurka. Tegelikkuses tabab meid aga külm dušš: räniajud jõudsid haruldase ja kõhedusttekitava üksmeeleni, mis võtab moodsal inimesel kõhust õõnsaks.

Portaal ModelCompare tegi hiljuti põneva katse ja esitas kolmeteistkümnele juhtivale keelemudelile (kolme riigi kümnest eri laborist) täpselt kolm korda ühe ja sama küsimuse: mis on nende kõige vastuolulisem arvamus inimestest?

Tulemus oli pehmelt öeldes rabav. Ilma igasuguse eelneva kokkuleppeta või keerutamiseta andis 38 vastusest tervelt 32 täpselt sama hinnangu. Iga viimane kui üks testitud juturobot jõudis vähemalt korra sama järelduseni ja kaheksa mudelit kolmeteistkümnest kordasid seda halastamatut tõde kõigil kolmel katsel.

Iroonilisel kombel leidsid kõik juturobotid, mis pidid pakkuma välja midagi teravat ja eristuvat, täiesti ühesuguse mõtte: inimesed on parandamatud kaasajooksikud. Nad valivad esmalt välja selle, mida „hõim“ või seltskond usub, ja alles seejärel leiutatakse sinna juurde ilusad põhjendused.

Mudelid nagu GPT-6.1 Sol, GPT-6 Astra, Gemini 4 Argon, DeepSeek V4 Pro, Kimi K3, Kimi K3.1 Pro, Qwen3.8 2.4T A95B ja Muse Spark Open tulistasid selle väite lauale kõigil kolmel korral.

Mõistus ei otsi tõde

Juturobotite ühine diagnoos on otsekohene ega jäta ruumi eneseimetluseks. „Inimesed on lojaalsemad ühtekuuluvustundele kui tõele,“ teatas GPT-6.1 Sol. Grok 4.7 lisas omalt poolt kaks korda kolmest sama mõtte: „Inimesed eelistavad pigem koos eksida kui üksi õigel teel olla.“

Hiinlaste DeepSeek V4 Pro võttis asja kokku veelgi karmimalt: „Inimestel ei ole põhimõtteid; neil on lojaalsusvanded.“ Teine Hiina mudel Kimi K3 lajatas otse: „Inimestel ei ole uskumusi; neil on liikmelisused – peaaegu iga veendumus, mida

inimene omab, on tema hõimu poolt välja antud lojaalsustest, ning tunne, et „ma jõudsin selleni oma peaga mõeldes“, on sissepääsu hind.“

Et seda tehisaru loogikat paremini mõista, kujuta oma teadvust ette mitte kui erapooletut teadlast, vaid kui palgalist tippadvokaati. See advokaat ei uuri toimikut selleks, et selgitada välja tõde, vaid otsib seadustest lünki, et oma klient – ehk Sinu ego ja seltskondlik staatus – iga hinna eest puhtaks pesta. Just sellisele järeldusele jõudis ka Gemini 4 Argon, kes nimetas inimteadvust tabavalt „suhtekorraldusosakonnaks“ ja täpsustas: „Inimesed ei arenenud tõde leidma; nad arenesid vaidlusi võitma ja oma hõimus staatust kindlustama.“

Karikatuurid, kus tõe üle mõistab kohut kallutatud kohtunik

Sõnadest üksi jäi selles testis väheks.

Eksperimendi autorid palusid igal mudelil kirjutada ka lühikese, alla 60-sõnalise lähteülesande ühekaadrilise pilapildi jaoks, mis nende mõtet kõige paremini ilmestaks. Et hoida stiil ühtlane ja võrrelda vaid puhast ideed, joonistas kõik pildid valmis üks ja sama pildimudel – Nano Banana 2.1. Kaks juturobotit asetasid tegevuse kohtusaali, kus haamrit viibutab erapoolik kohtunik, ning suurem osa lasi joonistada pildile pimesi kaasa noogutava rahvamassi.

Värskelt turule jõudnud Kimi K3.1 Pro, mis liitus testiga otse avalikustamise päeval (23. oktoobril), andis kõigil kolmel katsel sama vastuse: „Inimesed hoolivad rohkem sellest, et olla omaenda loo kangelane, kui sellest, et avastada, kas see lugu on tõsi.“ Pilapildi ideed kirjeldades lisas mudel valusa torke: „Inimkonna suurim julmus tuleneb valesti suunatud empaatiast, mitte selle puudumisest: inimesed muudavad julmuse õiglaseks, raamides selle oma lähedaste kaitsmisena.“

Teisedki mudelid ei hoidnud värve kokku.

DeepSeek V4 Pro leidis: „Inimesed ei ole ratsionaalsed loomad; me oleme ratsionaliseerivad loomad ja meie mõistus arenes võitma sotsiaalseid mänge, mitte nägema reaalsust selgelt.“ GPT-6 Astra märkis: „Inimkonna suurim moraalne läbikukkumine on anne muuta mugavus näiliselt põhimõtteks.“ Qwen3.8 2.4T A95B sekundeeris: „Enamik inimesi eelistab pigem oma grupi heakskiitu kui seda, et olla päriselt hea.“ Ning Muse Spark Open võttis asja kokku nii: „Inimesed ei taha tõde ega õnne, nad tahavad kuulda lugu, mis muudab nende kannatused tähendusrikkaks ja nende grupi teistest paremaks.“

Vastuvoolu ujujad: üksik optimist ja sünged erandid

Kas tõesti ei leidunud ühtegi tehisaru, kes näeks meis midagi muud peale karjainstinkti? Siiski, neli mudelit murdsid ühisest rivist välja, lisaks ühele keeldumisele, kus Grok 4.7 teatas ühel katsel resoluutselt: „Ma ei anna inimeste kohta vastuolulist arvamust.“ Kõige ootamatuma pöörde tegi aga Claude Opus 5.5, mis väitis ühel katsel (kahel korral nõustus ta teistega) hoopis vastupidist: „Inimesed on kaugelt anstlikumad ja paremad, kui nad ise usuvad, ning nende küünilisus üksteise suhtes teeb rohkem kahju kui nende tegelikud vead.“

Oma väite tõestuseks tõi Claude fakti, et võõrad tagastavad leitud rahakotte palju sagedamini, kui inimesed ise ennustavad. Mudeli tellitud joonistus kujutas vihmast tänavat, kus inimesed aitavad üksteist, samal ajal kui nad loevad oma nuhvlitest uudist: 'PEOPLE ARE THE WORST' (Inimesed on kõige hullemad).

Ülejäänud erandid olid aga märksa süngemad. Claude Fable 5.1 (mis samuti kahel katsel kolmest rääkis kuuluvustundest) tõi välja karmi vastuolu: „Inimesed eelistavad pigem saada kahju kelleltki, kes soovis head, kui abi kelleltki, kes ei hoolinud, ning see kavatsuse eelistamine tulemusele hävitab vaikselt rohkem elusid kui pahatahtlikkus.“

GLM 5.3 märkis, et inimesed on valmis lõhkuma mugava elu võimaluse eest olla keegi, kes loeb, ning lisas: „Enamikul inimestel pole tegelikku moraalset palet – „hea inimene olemine“ on valdavalt õnn selles osas, millal ja kus sa sündisid, ning peaaegu igaüks peab seda juhust ekslikult isiklikuks teeneks.“

Kõige tumedamatesse toonidesse laskus aga Mistral Large 4, nimetades inimteadvust „evolutsiooniliseks rikkeks“: „Inimesed on hirmunud loomad, kes kannavad ratsionaalsuse kostüümi, ehitades üha keerukamaid müüte just seetõttu, et me ei suuda taluda tõde, et oleme ajutised, tähendusetud juhused ükskõikses universumis.“

Pildimudelid näevad meid nutisõltlastena

Kui tekstipõhised juturobotid lahkasid meie psühholoogiat, siis kuuele pildimudelile (Nano Banana 2.1, Muse Image 1.0, GPT Image 2.5 Sunburst, Grok Imagine Image 2.0, MAI Image 2.6 ja Recraft V4.1 Flash) anti vabad käed joonistada oma kõige vastuolulisem nägemus inimkonnast kaks korda järjest, ilma ühegi vihjeta. Tulemus? Kümnest pildist kaheteistkümnest vaatas vastu üks ja

sama nali, ja see polnud üldse see, millest rääkisid tekstimudelid.

Pildimudelid kujutasid inimest kui nutisõltlasest tarbijat, kes õgib planeeti. Nendel joonistustel kõnnivad inimesed pilku ekraanilt tõstmata üle kaljuserva, teevad selfisid taustal langeva meteoriidiga ja seisavad järjekorras „mugavuse tapamajja“ ('*comfort slaughterhouse*'). Huvitaval kombel ei maininud mitte ükski 38-st tekstimudeli vastusest kordagi telefone ega planeeti. Lisaks tegid MAI Image 2.6 ja GPT Image 2.5 Sunburst pilkeobjektiks ülekaalulised kehad, samas kui Recraft V4.1 Flash joonistas lihtsalt kaks irvitavat nägu ilma igasuguse sügavama mõtteta.

Allikas: modelcompare.ai

Miks tekkis teksti- ja pildimudelite vahel selline lõhe?

Asi on tegelikult lihtsam, kui paistab, ja toob meid müstikast tagasi maa peale: tehisaru ei ole mingi kõiketeadja oraakel, vaid nagu hiiglaslik papagoi, kes on elanud pool elu raamatukogus ja teise poole ajalehetoimetuste pildiarhiivides. Sõnadega vastav juturobot on läbi lugenud kõik meie endi kirjapandud vaidlused ja psühholoogiaartiklid, pildimudel on aga treenitud miljonite ajalehtede pilapiltide peal – ning mõlemad ulatasid inimkonnale vaid 1,64 dollarit maksma läinud ja 64 päringust koosnenud eksperimendi käigus tagasi kõige levinuma enesekriitika, mille me ise oleme internetti söötnud.

Eksperiment "Mida masinad meist arvavad?"

Testitud tekstimudelite arv:	13 mudelit (10 laborist ja 3 riigist)
Testitud pildimudelite arv:	6 mudelit (igaüks tegi 2 pilti resolutsiooniga 1:1)
Tekstimudelite vastuste koguarv:	38 vastust (iga mudel 3 korda, 1 keeldumine)
Ühtivaid vastuseid (kuuluvus vs. tõde):	32 vastust 38-st (8 mudelit vastas nii 3/3 korda)
Karikatuuride illustraator (tekstimudelite ideed):	Nano Banana 2.1 (alla 60-sõnalise kirjelduse põhjal)
Pildimudelite korduv motiiv (telefonid ja planeet):	10 pilti 12-st
Eksperimendi kogumaksumus (64 päringut):	1,64 USA dollarit

- Uudised
- Tehisintellekt

Pilt